Revisiting the Capacity of Noncoherent Fading Channels in mmWave System

Guido Carlo Ferrante, *Member, IEEE,* Tony Q. S. Quek, *Senior Member, IEEE,* and Moe Z. Win, *Fellow, IEEE*

***Abstract*—Millimeter wave communications use large transmission bandwidth and experience severe propagation conditions. This forces the communication system to operate in the so-called *wideband regime*, where signals must be increasingly "peaky" in order to attain a large fraction of the peak unconstrained wideband capacity. This paper investigates the capacity of noncoherent channels as a function of bandwidth for signals with average and peak power constraints operating in the millimeter spectrum. Upper and lower bounds on capacity are provided, and the impact of features peculiar to millimeter wave channels is investigated. Numerical results for a scenario based on recent experimental campaigns are provided. It is shown that the rate achievable by a typical user in a millimeter wave cell with "non-peaky" signaling can be bounded away from the peak unconstrained wideband capacity. This suggests to reconsider the role of signaling "peakedness" in future millimeter wave communications.**

***Index Terms*—Millimeter wave communication, Fading channels, Dispersive channels, Information theory**

I. Introduction

A. Background and Motivation

WIDEBAND FADING CHANNELS have been of interest for more than five decades, particularly since it was shown that the capacity of a fading channel is the same as that of an additive white Gaussian noise (AWGN) channel in the infinite-bandwidth limit under an average power constraint on the transmitted signal [1]. Although frequency-shift keying (FSK) was shown to approach capacity in this limit, it was demonstrated that a rate penalty is incurred by constraining the power spectral density of the received signal [2]. More recently, different fading models have been investigated in connection to ultra-wideband (UWB) systems [3]–[5] and code-division multiple-access (CDMA) systems [6]–[8]. The robustness of UWB signals in a multipath environment was demonstrated experimentally in [9]–[11]. The effect of duty cycle on the performance of single-user UWB systems was studied in [12]. The capability of wideband signals to resolve multipath was investigated in connection to locally generated reference systems [13]–[15] and transmitted reference systems [16]–[19]. Multipath has been also studied in connection to sequence acquisition and synchronization [20]–[23].

The wideband regime refers to the regime of vanishing signal-to-noise ratio (SNR) per symbol [24]. It has also been referred to as low-SNR regime [25], [26]. For a fixed transmitted power, the SNR per degree of freedom decreases with increasing bandwidth; the so-obtained channel is energy-limited (in the terminology of [27]) rather than bandwidth-limited. A critical phenomenon occurs in the wideband regime: if the input signal is constrained to allocate energy *evenly* over a scattering channel, then capacity scales with the inverse of transmission bandwidth [6], [28]. A similar behavior holds for multipath channels with a finite number of paths: the mutual information attained by "white-like" signals scales with the inverse of the number of paths [7]. The class of capacity-achieving input distributions in the wideband regime was derived and the signaling based on such distributions was called *flash signaling* in [24]. The general guideline arising from the above studies is that capacity scales as the inverse of the number of independent branches of the channel when the signaling is "non-peaky," i.e., when the energy of the transmitted signal is allocated evenly over the degrees of freedom of the channel. On the contrary, a signaling that concentrates energy over a small subset of degrees of freedom, which is referred to as "peaky" signaling, becomes asymptotically optimal as the SNR per degree of freedom vanishes. However, from a practical viewpoint, it was shown in [29] that "non-peaky" signaling is sufficient to achieve a significant fraction of the wideband capacity in current wireless scenarios, in particular Wi-Fi and cellular systems.

The proposal of millimeter wave (mmWave) communications as one enabling technology for 5G [30] stimulates a renewed interest in the study of wideband channels. In particular, the investigation of the wideband regime is of inherent interest for mmWave communications, where the bandwidth is much larger than that used in Wi-Fi and cellular systems. Several portions of the spectrum, namely segments accounting for up to 1 GHz of total bandwidth in the 28–38 GHz band, the license-free 57–64 GHz band, and E-bands at 71–76 GHz, 81–86 GHz, and 92–95 GHz, are available for potential use [30], [31]. A bandwidth of 1 GHz has already been considered in recent experimental campaigns [31]–[33]. Peculiar

Manuscript received December 14, 2015; revised March 18, 2016; revised December 23, 2016; revised March 23, 2017; accepted March 24, 2017. The associate editor coordinating the review was Prof. Vincent Y. F. Tan. This work was supported, in part, by the National Science Foundation under Grant CCF-1525705, the SUTD-ZJU Research Collaboration under Grant SUTD-ZJU/RES/01/2014, the MOE ARF Tier 2 under Grant MOE2015-T2-2-104, and the SUTD-MIT Postdoctoral Fellowship. This paper was presented at the 2016 IEEE International Conference on Communications (ICC), Kuala Lumpur, Malaysia.

Guido Carlo Ferrante is with the Laboratory for Information and Decision Systems (LIDS), Massachusetts Institute of Technology, Cambridge, MA 02139 USA, and with the Information Systems Technology and Design (ISTD) Pillar, Singapore University of Technology and Design, 487372 Singapore (e-mail: gcfer@mit.edu).

Tony Q. S. Quek is with the Information Systems Technology and Design Pillar, Singapore University of Technology and Design, 487372 Singapore (e-mail: tonyquek@sutd.edu.sg).

Moe Z. Win is with the Laboratory for Information and Decision Systems (LIDS), Massachusetts Institute of Technology, Cambridge, MA 02139 USA (e-mail: moewin@mit.edu).

propagation phenomena, such as blockage, line-of-sight (LOS) propagation, and frequency-selective power absorption, require to revisit theoretical studies in order to draw conclusions and design insights valid for mmWave communications.

In this work we derive upper and lower bounds on the capacity of wideband channels under both peak and average power constraints. We consider a doubly-dispersive block-fading channel model, which allows us to account for peculiar features of mmWave channels while remaining analytically tractable.

B. Related Work

A first classification of recent work on fading channels can be drawn on the basis of the channel model. The so-called standard block-fading model [25], [26], [34]–[36] and the stationary fading model [6], [37]–[41] represent the two extreme models corresponding to "discontinuous" and "continuous" fading correlation across signal space dimensions, respectively. Variations and special cases are treated in [42]–[45]. A second classification can be based on the SNR regime of interest: with the exception of [34], [41], capacity bounds are typically derived with a focus on either the high-SNR regime [38], [39] or the low-SNR regime [6], [24]–[26], [40]. Finer classification includes the type (average vs. peak) of power constraint and the specific description of the fading process (absence or presence of specular component, fading distribution, etc.). A survey on fading channels and the effect of signaling constraints on capacity can be found in [8]. The connection between block and stationary models is discussed in [46]. We detail below the above literature in terms of block and stationary models in both low- and high-SNR regimes. A partial summary of the discussed literature is presented in Table I.

1) Block-fading model: The standard block-fading model was introduced in [47]. In this model, channel coefficients are divided into blocks called *coherence blocks*: the fading coefficients within each block are identical, while different blocks fade independently. The size of each block depends on the physical properties of the channel. Certain structures of the optimal capacity-achieving input distribution as well as the bounds on the capacity of a Rayleigh flat fading channel were shown for several practical scenarios in [34]. In [35] it was shown that the capacity of the Rayleigh block-fading channel grows logarithmically as a function of SNR in the high-SNR regime. The Rayleigh block-fading channel in the low-SNR regime was studied in [25], where the sublinear term of capacity was characterized as a function of the SNR per degree of freedom and the coherence block size. The extension to the MIMO setting can be found in [26]. Capacity of the Rician block-fading channel in the low-SNR regime was studied in [48], [49] using the general framework proposed in [24]: their results are in line with our analysis for a channel with no blockage and absorption in the limit of infinite bandwidth. In [36], a block model accounting for sparsity in the delay-Doppler plane is proposed, and an analysis of mutual information in the wideband regime is presented.

2) Stationary fading model: The wide-sense stationary uncorrelated scattering (WSSUS) model was proposed in [50] and further investigated in [51]. It was shown in the '60s that the wideband capacity limit of a Rayleigh fading channel is equal to that of an AWGN channel, and that the limit is achievable via FSK [1], [52]. The extension to multipath fading channels was derived later in [7]. The bandwidth scaling was studied in [6], [28], where it was shown that the rate achieved by a spread-spectrum signal transmitted over a WSSUS fading channel scales as the inverse of the signal bandwidth. The adopted channel model assumes a diffuse scattering accounting for a continuum of infinitesimal paths. Spread-spectrum signals were defined as those signals with second and fourth moments scaling as $1/W_{\text{tot}}$ and $1/W_{\text{tot}}^2$, respectively, where W_{tot} denotes the signal bandwidth; such scaling implies that energy is evenly spreading on signal space dimensions. The conclusion that uneven energy spreading is required to achieve a nonvanishing information rate in the wideband regime was further reinforced in [37]. Capacity per unit energy, which is strictly related to capacity at low-SNR, is investigated in [40]. Further investigations on the low-SNR capacity of MIMO systems were conducted in [53]. A similar behavior was observed in [7] for "white-like" signals transmitted over a multipath channel with finite number of paths. It was shown that capacity scales as the inverse of the number of resolvable paths when the receiver has side information on path delays; without such information, capacity approaches zero with increasing bandwidth even when the channel has a single path with time-varying delay. The high-SNR regime was characterized in [38], where the high-SNR asymptotic expansion of capacity up to the second order term, called *fading number*, was derived. It was shown that capacity grows double-logarithmically as a function of SNR if the fading process is *regular*. The analysis was extended to *non-regular* fading processes in [39] to close the gap between the double-logarithmic behavior observed with stationary fading channels and the logarithmic behavior observed with the block-fading model. In [41], bounds on the capacity of Rayleigh fading WSSUS channels as a function of bandwidth were proposed under both average and peak power constraints.

C. Main Contributions

In this paper we derive upper and lower bounds on the capacity of noncoherent channels that take into account essential mmWave propagation features. The main contributions can be summarized as follows:

- We consider a channel model that describes peculiar mmWave propagation properties on the basis of recent experimental campaigns [31]–[33]. We consider a doubly-dispersive underspread channel, and use a block-fading model that can describe three mmWave propagation phenomena: blockage, LOS propagation, and frequency-selective power absorption due to oxygen and water vapor.
- We derive upper and lower bounds on the capacity of the noncoherent channel as a function of bandwidth. Together with the average power constraint we set a peak constraint

TABLE I: A (partial) summary of the previous literature.

<table>
<tr><th colspan="2"></th><th colspan="2">SNR regime</th></tr>
<tr><th colspan="2"></th><th>Low-SNR regime</th><th>High-SNR regime</th></tr>
<tr><td rowspan="4">Channel model</td><td rowspan="2">Block fading</td><td colspan="2">Rayleigh, MIMO capacity bounds for arbitrary SNR [34]
Rayleigh, MIMO, capacity lower bounds via training for arbitrary SNR [54]</td></tr>
<tr><td>Rayleigh, SISO [25] and MIMO [26] wideband
Rician, wideband [48], [49]
Sparse delay-Doppler, wideband [36]</td><td>Rayleigh, MIMO, multiplexing gain [35]</td></tr>
<tr><td rowspan="2">Stationary fading</td><td colspan="2">Rayleigh, capacity bounds for arbitrary SNR [41], [55]
Multipath, capacity bounds for arbitrary SNR [56]</td></tr>
<tr><td>Rayleigh, FSK [1], [52]
Rayleigh, bandwidth scaling [6], [28]
Multipath channels [7]
Rayleigh, capacity per unit-cost [37], [40]
Rayleigh, SISO [57] and MIMO [53]</td><td>Rayleigh, “regular” channels [38]
Rayleigh, “non-regular” channels [39]</td></tr>
</table>

(in the form of amplitude or fourth moment constraint), which restricts the system to use “non-peaky” signaling. Upper bounds are derived by providing side information about fading and blockage processes, and extending a technique that we refer to as *supremum splitting* from [41], [57], which accounts for the peak constraint. Lower bounds are derived either by considering a worst-case rate penalty due to the peak constraint or by assuming a training-based scheme to estimate the channel. In the former case, we adapt a previous bound proposed in [41] for WSSUS channels to the block-fading channel; in the latter case, we modify the analysis in [54] to accomodate the proposed channel model and specialize the result to truncated-Gaussian inputs. In both cases, bounds are tightened by using a time-sharing argument [58].

- We evaluate the bounds by considering a cellular scenario where system parameters are chosen in accordance with experimental campaigns [31]–[33]. We demonstrate that the system can operate in the wideband regime and discuss the detrimental effect on the achievable rate caused by the presence of a peak constraint as well as the importance of increased “peakedness” of the signal transmitted by a typical user.

The remainder of this paper is organized as follows. Section II describes the system model, including the channel model. Section III and IV present upper and lower bounds on the capacity of the noncoherent channel. Numerical results are discussed in Section V, and Section VI concludes the paper.

Notations: Random variables are displayed in sans serif, upright fonts; their realizations in serif, italic fonts. Vectors and matrices are denoted by bold lowercase and uppercase letters, respectively. For example, a random variable and its realization are denoted by x and x; a random vector and its realization are denoted by $\boldsymbol{\mathsf{x}}$ and $\boldsymbol{x}$; a random matrix and its realization are denoted by $\boldsymbol{\mathsf{X}}$ and $\boldsymbol{X}$, respectively. The element (k, l) of a matrix is denoted using brackets, e.g. $[\boldsymbol{X}]_{kl}$; $\odot$ denotes the Hadamard (elementwise) product; and $\mathrm{Vec}(\cdot)$ denotes the vectorization of the matrix in the argument. Proper Complex Normal distribution with mean m and variance σ^2 is denoted by $\mathcal{CN}(m, \sigma^2)$, and its probability measure is denoted by μ_{m,σ^2}. Bernoulli distribution with probability of success p is denoted by $\mathcal{Bern}(p)$. The nonnegative part of a real number x is defined as follows: $(x)^+ := \max\{0, x\}$. The Frobenius norm of a matrix $\boldsymbol{X}$ is denoted $\|\boldsymbol{X}\|_{\mathrm{F}}$. Expectation and variance of a random variable are denoted by $\mathbb{E}\{\cdot\}$ and $\mathbb{V}\{\cdot\}$, respectively.

II. System Model

This section is organized as follows. We present in Section II-A the signal model. Section II-B describes the channel model that accounts for some of the features of mmWave channels. Finally, the problem statement is formulated in Section II-C.

A. Signal Model

We adopt the following discrete signal model [8]

$$\boldsymbol{\mathsf{Y}} = \boldsymbol{\mathsf{H}} \odot \boldsymbol{\mathsf{X}} + \boldsymbol{\mathsf{N}} \tag{1}$$

where $\boldsymbol{\mathsf{X}}$ and $\boldsymbol{\mathsf{Y}}$ are the matrices of transmitted and received symbols, respectively, $\boldsymbol{\mathsf{H}}$ is the channel matrix, $\boldsymbol{\mathsf{N}}$ is the random matrix, representing additive noise, with i.i.d. elements drawn according to $\mathcal{CN}(0, N_0)$. Without loss of generality, we can assume $N_0 = 1$. Matrices in (1) have dimension $K \times L$, where K and L are the number of dimensions of the signal space in the time and frequency domains, respectively. Therefore, symbol received, symbol transmitted, channel coefficient

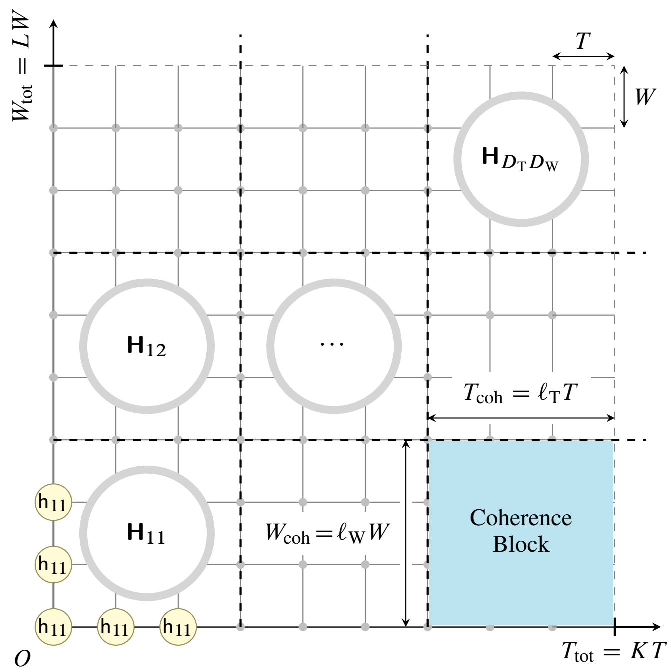


FIG. 1: Signal and channel models. Each degree of freedom is depicted as a small block of sides T and W. Each channel coherence block comprises a subset of blocks and has sides T_{coh} and W_{coh}. Fading coefficients within the coherence block (i, j) are collected in the matrix $\mathbf{H}_{ij}$, which is a constant matrix with generic element equal to h_{ij}.

and noise at time epoch k on the frequency subband l are $\mathsf{y}_{kl} := [\mathbf{Y}]_{kl}$, $\mathsf{x}_{kl} := [\mathbf{X}]_{kl}$, $\mathsf{h}_{kl} := [\mathbf{H}]_{kl}$ and $\mathsf{n}_{kl} := [\mathbf{N}]_{kl}$, respectively. This model can be derived from the discretization of a continuous model through transmission and reception over a time interval $[0, T_{\text{tot}}]$ and frequency interval $[0, W_{\text{tot}}]$ of a Weyl-Heisenberg set $\{g_{kl}(t) := g(t - kT)e^{i2\pi lW}\}_{(k,l)\in\mathbb{Z}^2}$, where $g(t)$ is a baseband pulse with effective duration T and effective bandwidth W; the reader is referred to [41], [55], [59]–[62] for a detailed exposition of the mapping from the continuous to the discrete representations.

B. Channel Model including mmWave Features

We adopt the standard block-fading model, where time and frequency resources are partitioned into coherence blocks of duration T_{coh} and bandwidth W_{coh}, and fading coefficients within each coherence block are identical. The channel is assumed to be doubly-dispersive with delay spread T_d and Doppler spread W_D, which are related to the coherence bandwidth W_{coh} and coherence time T_{coh}, respectively, as $W_{\text{coh}} = 1/T_\text{d}$ and $T_{\text{coh}} = 1/W_\text{D}$ (see Fig. 1). The quantity $\Delta_{\mathbf{H}} = T_\text{d}W_\text{D} = 1/(T_{\text{coh}}W_{\text{coh}})$ determines whether the channel is underspread ($\Delta_{\mathbf{H}} < 1$) or overspread ($\Delta_{\mathbf{H}} > 1$). As a matter of fact, most wireless channels are highly underspread, i.e., $\Delta_{\mathbf{H}} \ll 1$. This continues to hold for mmWave channels, as detailed in the following. Most channels experience delay spread between $T_\text{d} = 50$ ns and $T_\text{d} = 500$ ns, corresponding to coherence bandwidth 2 MHz $\leqslant W_{\text{coh}} \leqslant$ 20 MHz . Doppler bandwidth is mainly due to transmitter and receiver relative radial motion, and it is given by $W_\text{D}/2 = f_0 v_\text{r}/c$, where f_0 is the carrier frequency, v_r is the velocity of the radial motion, and c is the speed of light. Suppose $f_0 = 60$ GHz, that is in mmWave spectrum, and consider two scenarios, corresponding to $v_\text{r} = 1$ m/s (walking speed) and $v_\text{r} = 100$ m/s (high-speed train top speed): Doppler bandwidth is $W_\text{D} = 0.4$ KHz and $W_\text{D} = 40$ KHz, respectively, and thus coherence time is bounded as follows: 25 µs $\leqslant T_{\text{coh}} \leqslant$ 2.5 ms. In the above scenarios, $\Delta_{\mathbf{H}} = T_\text{d}W_\text{D} = 1/(T_{\text{coh}}W_{\text{coh}}) \leqslant 10^{-2}$, showing that channels operating at mmWave frequencies are significantly underspread even with high mobility and long delay spread.

Denote $D_\text{T} = T_{\text{tot}}/T_{\text{coh}}$ and $D_\text{W} = W_{\text{tot}}/W_{\text{coh}}$ the number of coherence blocks in time and frequency, respectively, and $\ell_\text{T} = K/D_\text{T} = T_{\text{coh}}/T$ and $\ell_\text{W} = L/D_\text{W} = W_{\text{coh}}/W$ the number of dimensions of each coherence block in time and frequency, respectively. Since the channel is underspread, signals transmitted over each coherence block have $\ell := \ell_\text{T}\ell_\text{W} \gg 1$ degrees of freedom that experience the same fading. For each $1 \leqslant i \leqslant D_\text{T}$ and $1 \leqslant j \leqslant D_\text{W}$, denote h_{ij} the common fading coefficient, i.e.

$$\mathsf{h}_{ij} := [\mathbf{H}]_{(i-1)\ell_\text{T}+k,(j-1)\ell_\text{W}+l} \tag{2}$$

which holds for all $0 \leqslant k \leqslant \ell_\text{T} - 1$ and $0 \leqslant l \leqslant \ell_\text{W} - 1$.

Three essential properties of mmWave propagation channels are taken into account [31]:

(i) blockages occur with probability p_B, i.e., there exists a fraction p_B of time during which the transmitted signal cannot reach the receiver;
(ii) the channel can be highly LOS, highly NLOS, or blocked;
(iii) there exists an average power absorption profile due to atmospheric gases and water vapor that is frequency-dependent (e.g. values are tabulated in [63]).

We model (i)–(iii) by the following generic coefficient h_{ij} in (2),

$$\mathsf{h}_{ij} = \mathsf{a}_i v_j \mathsf{g}_{ij} \tag{3}$$

where a_i and g_{ij} are random variables modeling blockage and fading at coherence time block i and coherence band j, respectively, while v_j is a deterministic attenuation modeling absorption at coherence band j. In particular, property (i) is modeled by assuming $\mathsf{a}_i \sim \mathcal{B}ern(1 - p_\text{B})$, i.e., the distribution of a_i is $P_{\mathsf{a}_i} = p_\text{B}\delta_0 + (1 - p_\text{B})\delta_1$, where δ_x denotes a Dirac measure with single atom at x. Property (ii) implies that, conditioned on $\mathsf{a}_i = 1$, the distribution of $\mathsf{h}_{ij} \propto \mathsf{g}_{ij}$ models both LOS and NLOS environments. To this end, g_{ij} is modeled as a Gaussian random variable with nonzero mean (Rician fading) and fixed second moment, i.e., $\mathsf{g}_{ij} \sim \mathcal{CN}(\eta, 1 - |\eta|^2)$. Note that $\mathbb{E}\{\mathsf{g}_{ij}\} = \eta$ is constant irrespective of the block, which models a channel whose specular component is dominated by a single path. Property (iii) is modeled by means of a deterministic sequence $(v_1, v_2, \ldots, v_{D_\text{W}})$ that accounts for pathloss and absorption of different frequency bands. Denote μ_j the Gaussian measure with mean $v_j\eta$ and variance $v_j^2(1 - |\eta|^2)$. From the above, h_{ij} is distributed as

$$\begin{aligned} P_{\mathsf{h}_{ij}} &= p_\text{B}\delta_0 + (1 - p_\text{B})\mu_j \\ P_{\mathsf{h}_{ij}|\mathsf{a}_i=1} &= \mu_j \\ P_{\mathsf{h}_{ij}|\mathsf{a}_i=0} &= \delta_0. \end{aligned} \tag{4}$$

In particular, the second moment of h_{ij} is

$$\mathbb{E}\{\,|\mathsf{h}_{ij}|^2\,\} = (1 - p_\mathrm{B})$$

and the variance of h_{ij} is

$$\sigma^2_{\mathsf{h}_{ij}} = \upsilon_j^2(1 - p_\mathrm{B})[1 - (1 - p_\mathrm{B})|\eta|^2]. \tag{5}$$

The channel block structure induces a natural partition of matrices in (1). Let $\mathbf{Y}_{ij}$ be the $\ell_\mathrm{T} \times \ell_\mathrm{W}$ matrix of the received signal in coherence block (i, j), and similarly define $\mathbf{H}_{ij}$, $\mathbf{X}_{ij}$, and $\mathbf{N}_{ij}$. From (1) it follows that

$$\mathbf{Y}_{ij} = \mathbf{H}_{ij} \odot \mathbf{X}_{ij} + \mathbf{N}_{ij} = \mathsf{h}_{ij}\mathbf{X}_{ij} + \mathbf{N}_{ij}.$$

Vectorizing each block, i.e., defining $\mathbf{y}_{ij} = \mathrm{Vec}(\mathbf{Y}_{ij})$ and using similar notations for $\mathbf{H}_{ij}$, $\mathbf{X}_{ij}$, and $\mathbf{N}_{ij}$, yields $\mathbf{y}_{ij} = \mathbf{h}_{ij} \odot \mathbf{x}_{ij} + \mathbf{n}_{ij} = \mathsf{h}_{ij}\mathbf{x}_{ij} + \mathbf{n}_{ij}$. Stacking vectors with respect to index $j \in \{1, 2, \ldots, D_\mathrm{W}\}$ yields

$$\mathbf{y}_i = \mathbf{h}_i \odot \mathbf{x}_i + \mathbf{n}_i\,. \tag{6}$$

Then, stacking vectors with respect to index $i \in \{1, 2, \ldots, D_\mathrm{T}\}$ yields

$$\mathbf{y} = \mathbf{h} \odot \mathbf{x} + \mathbf{n} \tag{7}$$

which is an equivalent vector form of (1).

In general, we assume $T_\mathrm{tot} \gg T_\mathrm{coh}$ and $W_\mathrm{tot} \gg W_\mathrm{coh}$; hence, $\mathbf{x}$ is transmitted over a large number of coherence blocks that fade independently conditioned on $\mathsf{a}_i = 1$, and are zero otherwise. The latter case is when blockage occurs. Blockages are modeled as independent events[1] with time scale T_coh. Therefore, a random subset of symbols is nulled by the channel; the average fraction of nulled symbols is p_B. Note that blockage events such that $\mathbf{Y} = \mathbf{N}$ are exponentially unlikely as a function of D_T since $\mathbb{P}\{\,\mathsf{a}_1 = \mathsf{a}_2 = \cdots = \mathsf{a}_{T_\mathrm{tot}/T_\mathrm{coh}} = 0\,\} = p_\mathrm{B}^{D_\mathrm{T}}$.

C. Problem Statement

The noncoherent capacity of the fading channel (1) is [40]

$$C = \frac{1}{T_\mathrm{tot}} \sup_{P_\mathbf{X}} I(\mathbf{X};\mathbf{Y}) \quad \text{nats/s}. \tag{8}$$

The supremum in (8) is over a family of input distributions satisfying an **average power constraint** [41]

$$\frac{1}{T_\mathrm{tot}} \sum_{k=0}^{K-1}\sum_{l=0}^{L-1} \mathbb{E}\{\,|[\mathbf{X}]_{kl}|^2\,\} \leqslant P_\mathrm{t} =: P\,W_\mathrm{tot} \tag{9}$$

and a **peak power constraint** in the form of either **amplitude constraint**

$$|[\mathbf{X}]_{kl}|^2 \leqslant \beta\frac{P_\mathrm{t}}{W_\mathrm{tot}} = \beta P \;\text{ a.s.} \tag{10}$$

or **fourth moment constraint**

$$\mathbb{E}\{\,|[\mathbf{X}]_{kl}|^4\,\} \leqslant \beta P^2 \tag{11}$$

for $0 \leqslant k \leqslant K-1$, $0 \leqslant l \leqslant L-1$. We note that:

- Parameters P_t and P are expressed in units of energy (normalized with respect to the noise spectrum level N_0) per unit time and per symbol, respectively. Since $T_\mathrm{tot}W_\mathrm{tot} = KL$, the average power constraint is equivalent to $\|\mathbf{X}\|_\mathrm{F} \leqslant KLP$.
- The amplitude constraint can be compactly rewritten as $\|\mathbf{X}\|^2_\infty \leqslant \beta P$ a.s., and it can be also regarded as a constraint on the support of the input distribution.
- Both peak power constraints are set over time and frequency dimensions, therefore on a per-symbol basis.
- The fourth moment constraint is similar to the constraint considered in [6], where energy is spread evenly over degrees of freedom while not bounding the input distribution support.
- A peak constraint related to the fourth moment constraint is the kurtosis constraint $\mathrm{Kurt}\{\,[\mathbf{X}]_{kl}\,\} = \mathbb{E}\{\,|[\mathbf{X}]_{kl}|^4\,\}/\,\mathbb{E}\{\,|[\mathbf{X}]_{kl}|^2\,\}^2 \leqslant \beta$. Kurtosis and fourth moment constraints are not equivalent: in fact, the former is stricter than the latter.

In the rest of the paper, we denote $\mathcal{P}_\mathrm{a}$, $\mathcal{P}_\mathrm{m}$ and $\mathcal{P}_\mathrm{f}$ the family of distributions satisfying average power constraint, modulus (amplitude) constraint, and fourth moment constraint, respectively. We denote joint constraints via set intersection, e.g. the subset of distributions satisfying both average power constraint and modulus (amplitude) constraint is denoted by $\mathcal{P}_\mathrm{a} \cap \mathcal{P}_\mathrm{m}$. We denote $\mathcal{P}_\mathrm{a}^\mathrm{B}$ and $\mathcal{P}_\mathrm{a}^\mathrm{S}$ the set of distributions with average power constraint set over blocks and symbols, respectively.

III. Capacity Upper Bounds

In this section, we present three upper bounds on capacity in Theorem 1, 2 and 3. The three bounds are specialized to the case of no absorption in Corollary 1, 2 and 3.

We derive in Theorem 1 an upper bound by neglecting the peak constraint and providing the channel knowledge to the receiver. Thus we compute the coherent capacity with average power constrained inputs, which will prove to be useful in the high-SNR regime. Then we specialize the result to the case of no absorption ($\upsilon_1 = \upsilon_2 = \cdots = \upsilon_{D_\mathrm{W}}$) in Corollary 1. We denote $J(P) = \mathbb{E}\{\,\log(1 + |\mathsf{g}|^2P)\,\}$, which is the coherent capacity of a fading channel with coefficient g when transmitted power is P.

Theorem 1. Consider a block-fading channel with D_T time blocks of dimension ℓ_T and D_W frequency blocks of dimension ℓ_W. Let the absorption profile be $(\upsilon_1, \upsilon_2, \ldots, \upsilon_{D_\mathrm{W}})$. Under the average power constraint, capacity is upper bounded as follows:

$$\sup_{P_\mathbf{X}\in\mathcal{P}_\mathrm{a}} \frac{1}{KL} I(\mathbf{X};\mathbf{Y}) \leqslant (1 - p_\mathrm{B})\frac{1}{D_\mathrm{W}} \sum_{j=1}^{D_\mathrm{W}} J(\upsilon_j^2 P_j^\star) \tag{12}$$

where $\{P_j^\star\}$ is given by statistical waterfilling, i.e. $P_j^\star = (p_j)^+$ being p_j the solution of $dJ(\upsilon_j^2 p_j)/dp_j + \lambda p_j = 0$ and λ such that $\sum_{j=1}^{D_\mathrm{W}} P_j^\star = D_\mathrm{W}P$.

Proof: See Appendix A. □

In general, J is not available in closed-form, and the power allocation problem has to be numerically solved. Note that $P_j^\star$ is a function of υ_j^2 and λ, and in turn of $\{\upsilon_1, \upsilon_2, \ldots, \upsilon_{D_\mathrm{W}}\}$

[1]The assumption that blockage events are independent can also be justified by assuming the presence of an interleaver, which artificially introduces time diversity.

and $D_{\mathrm{W}}P$. The optimal power allocation is similar to statistical waterfilling in MIMO channels without channel state information at the transmitter [64].

Corollary 1. Let assumptions be the same as in Theorem 1 with no absorption. Capacity is upper bounded as follows:

$$\sup_{P_{\mathbf{X}}\in\mathcal{P}_{\mathrm{a}}} \frac{1}{KL} I(\mathbf{X};\mathbf{Y}) \leqslant C_{\mathrm{coh}} \tag{13}$$

where $C_{\mathrm{coh}} = (1-p_{\mathrm{B}})J(P)$.

Proof: In the case of no absorption, the average power constraint implies $P_1^\star = P_2^\star = \cdots = P_{D_{\mathrm{W}}}^\star = P$, and the result follows from Theorem 1. □

In Theorem 2 we derive an upper bound by taking into account the amplitude constraint via *supremum splitting* [41], [57]. Then we specialize the result to the case of no absorption in Corollary 2.

Theorem 2. Consider a block-fading channel with D_{T} time blocks of dimension ℓ_{T} and D_{W} frequency blocks of dimension ℓ_{W}. Let the absorption profile be $(\upsilon_1, \upsilon_2, \ldots, \upsilon_{D_{\mathrm{W}}})$ and the blockage probability be p_{B}. Under average power and amplitude constraints, capacity is upper bounded as follows:

$$\begin{aligned}\sup_{P_{\mathbf{X}}\in\mathcal{P}_{\mathrm{a}}\cap\mathcal{P}_{\mathrm{m}}} \frac{1}{KL} I(\mathbf{X};\mathbf{Y}) &\leqslant (1-p_{\mathrm{B}})\frac{1}{D_{\mathrm{W}}} \\ &\sup_{q\in[0,1]}\Bigg\{\sum_{j=1}^{D_{\mathrm{W}}} \log(1+\upsilon_j^2 P_j^\star) \\ &- \frac{qP}{\ell P_{\mathrm{sum}}^\circ}\log(1+\upsilon_j^2\ell P_j^\circ(1-|\eta|^2))\Bigg\}\end{aligned} \tag{14}$$

where

$$\{P_i^\circ\}_{i=1}^{D_{\mathrm{W}}} = \underset{\substack{P_j\in\{0,\beta P\}\\ j=1,2,\ldots,D_{\mathrm{W}}}}{\operatorname{argmin}} \frac{\sum_{j=1}^{D_{\mathrm{W}}} \log(1+\upsilon_j^2\ell P_j(1-|\eta|^2))}{\sum_{j=1}^{D_{\mathrm{W}}} P_j} \tag{15}$$

$$P_{\mathrm{sum}}^\circ = \textstyle\sum_{j=1}^{D_{\mathrm{W}}} P_j^\circ \tag{16}$$

and $P_j^\star = (\lambda - 1/\upsilon_j^2)^+$ with λ such that $\sum_{j=1}^{D_{\mathrm{W}}} P_j^\star = qD_{\mathrm{W}}P$.

Proof: See Appendix B. □

Corollary 2. Let assumptions be the same as in Theorem 2 with no absorption. Capacity is upper bounded as follows:

$$\sup_{P_{\mathbf{X}}\in\mathcal{P}_{\mathrm{a}}\cap\mathcal{P}_{\mathrm{m}}} \frac{1}{KL} I(\mathbf{X};\mathbf{Y}) \leqslant C_{\mathrm{UB,MC}} \tag{17}$$

where

$$\begin{aligned}C_{\mathrm{UB,MC}} = (1-p_{\mathrm{B}}) \sup_{q\in[0,1]} \Bigg\{&\log(1+qP) \\ &- \frac{q}{\beta\ell}\log\left(1+\beta\ell P(1-|\eta|^2)\right)\Bigg\}.\end{aligned} \tag{18}$$

Proof: The optimum power allocations in Theorem 2 are given by $P_j^\star = qP$ and $P_j^\circ = \beta P$ for all j. The former is derived as in the proof of Corollary 1. The latter is derived by solving (15) as follows. For notational simplicity, let $b = \beta P$, $\alpha = \ell(1-|\eta|^2)$, and $n = D_{\mathrm{W}}$. The right-hand side of (15) can be simplified as

$$F(P_1, P_2, \ldots, P_n) := \frac{\sum_{j=1}^n \log(1+\alpha P_j)}{\sum_{j=1}^n P_j} \tag{19}$$

where $P_j \in \{0, b\}$. The minimum of (19) under the constraints $P_j \in \{0, \beta P\}$ for all j can be found by brute force. However, we can exploit the symmetries of F and check a subset of $n+1$ sequences only. Indeed, the value of F depends on the number of zero arguments, but not on their position. For example, $F(0,0,b,b,b,\cdots,b) = F(0,b,0,b,b,\cdots,b)$. A natural choice is to check the set of sequences $\{(\mathbf{0}_k, \boldsymbol{b}_{n-k}) : 0 \leqslant k \leqslant n\}$, where $\mathbf{0}_k$ denotes a length-k sequence of zeros and $\boldsymbol{b}_{n-k}$ a length-$(n-k)$ sequence of b's. The sequence $\mathbf{0}_n$ can be discarded since the function is decreasing in its neighborhood. The remaining sequences all yields the same function value:

$$F(\mathbf{0}_k, \boldsymbol{b}_{n-k}) = \frac{(n-k)\log(1+\alpha b)}{(n-k)b} = \frac{\log(1+\alpha b)}{b}. \tag{20}$$

The proof is completed by choosing $(P_1^\circ, P_2^\circ, \ldots, P_n^\circ) = \boldsymbol{b}_n$. □

In Theorem 3 we consider peak constraint in terms of fourth moment constraint, and in Corollary 3 the result is specialized to the case of no absorption.

Theorem 3. Consider a block-fading channel with D_{T} time blocks of dimension ℓ_{T} and D_{W} frequency blocks of dimension ℓ_{W}. Let the absorption profile be $(\upsilon_1, \upsilon_2, \ldots, \upsilon_{D_{\mathrm{W}}})$ and the blockage probability be p_{B}. Under average power and fourth moment constraints, capacity is upper bounded as follows:

$$\begin{aligned}\sup_{P_{\mathbf{X}}\in\mathcal{P}_{\mathrm{a}}\cap\mathcal{P}_{\mathrm{f}}} \frac{1}{KL} I(\mathbf{X};\mathbf{Y}) &\leqslant (1-p_{\mathrm{B}})\frac{1}{D_{\mathrm{W}}} \\ &\sup_{q\in[0,1]}\Bigg\{\sum_{j=1}^{D_{\mathrm{W}}}\Bigg[\log(1+\upsilon_j^2 P_j^\star) - \frac{1}{D_{\mathrm{T}}}\sum_{i=1}^{D_{\mathrm{T}}}\varphi(\acute{P}_{ij}, q)\Bigg]\Bigg\}\end{aligned} \tag{21}$$

where

$$\{\acute{P}_{ij}\} = \underset{\substack{P_{ij}:0\leqslant P_{ij}\leqslant\sqrt{\beta}P\\ \sum_{ij}P_{ij}=D_{\mathrm{T}}D_{\mathrm{W}}P}}{\operatorname{argmin}} \sum_{i=1}^{D_{\mathrm{T}}}\sum_{j=1}^{D_{\mathrm{W}}} \varphi(P_{ij}, q), \tag{22}$$

$$\varphi(P_{ij}, q) = \frac{q^2 P_{ij}^2}{\ell\beta P^2}\log\left(1+\upsilon_j^2(1-|\eta|^2)\frac{\ell\beta P^2}{qP_{ij}}\right) \tag{23}$$

and $P_j^\star = (\lambda - 1/\upsilon_j^2)^+$ with λ such that $\sum_{j=1}^{D_{\mathrm{W}}} P_j^\star = qD_{\mathrm{W}}P$.

Proof: See Appendix C.

Corollary 3. Let assumptions be the same as in Theorem 3 with no absorption. Capacity is upper bounded as follows:

$$\sup_{P_{\mathbf{X}}\in\mathcal{P}_{\mathrm{a}}\cap\mathcal{P}_{\mathrm{f}}} \frac{1}{KL} I(\mathbf{X};\mathbf{Y}) \leqslant C_{\mathrm{UB,FMC}} \tag{24}$$

where:

$$\begin{aligned}C_{\mathrm{UB,FMC}} = (1-p_{\mathrm{B}}) \sup_{q\in[0,1]} \Bigg\{&\log(1+qP) \\ &- \frac{q^2}{\ell\beta}\log\left(1+(1-|\eta|^2)\frac{\ell\beta P}{q}\right)\Bigg\}.\end{aligned} \tag{25}$$

Proof: The result follows from Theorem 3 by deriving the optimal power allocations when $\upsilon_j = 1$ for all $j \geqslant 1$. As in Corollary 2, we have $P_j^\star = qP$. Optimum power allocation for $\acute{P}_{ij}$ in (22) is uniform, i.e., $P_{ij} = P$ for all i and j. In fact, the problem in (22) is convex because the domain is convex (a parallelepiped) and the function to be minimized (cf. the right hand side of eq. (22)) is convex because it is the sum of convex functions (cf. $\varphi(P, q)$ in (23), which is convex with respect to P). Dropping the inequality constraints in (22) and using a Lagrange multiplier yield $P_{ij} = P$; since this solution satisfies the inequality constraints, it is also a solution for the original problem. □

We note that all upper bounds that are presented do not explicitly depend on ℓ_{T}, D_{T} and ℓ_{W} because the statistics of the channel are different across frequency blocks but remain the same across time blocks.

Remark 1 (Infinite-bandwidth capacity limit). Results in Corollary 2 and 3 have the same expansion (as $P \to 0$) up to the first order, which is equal to $(1 - p_{\mathrm{B}})|\eta|^2 P$. Since $P = P_{\mathrm{t}}/W_{\mathrm{tot}}$, capacity is equal to $(1 - p_{\mathrm{B}})|\eta|^2 P_{\mathrm{t}}$ nats/s as $W_{\mathrm{tot}} \to \infty$. Furthermore, the second order expansion (with respect to W_{tot}) shows that capacity is proportional to $1/W_{\mathrm{tot}}$ (nats/s). Both results are in line with [6], [41], [48], [49], which analyze the case $p_{\mathrm{B}} = 0$.

IV. Capacity Lower Bounds

We consider independent inputs over different coherence blocks, which implies $I(\mathbf{X}; \mathbf{Y}) \geqslant \sum_{i=1}^{D_{\mathrm{T}}} \sum_{j=1}^{D_{\mathrm{W}}} I(\mathbf{X}_{ij}; \mathbf{Y}_{ij})$ upon application of the chain rule for mutual information and the Kolmogorov inequality. The average power allocated to each block can be optimized by following a statistical water-filling on the basis of the sole knowledge of $\{v_1, v_2, \ldots, v_{D_{\mathrm{W}}}\}$ [64]. Achievable rates are derived for the generic block. Subscripts denoting block indeces are dropped for simplicity, and we adopt the following compact notation: $\boldsymbol{\xi} = \mathrm{Vec}(\mathbf{X}_{ij})$; $\boldsymbol{\psi} = \mathrm{Vec}(\mathbf{Y}_{ij})$; $\boldsymbol{\nu} = \mathrm{Vec}(\mathbf{N}_{ij})$. We denote ξ the generic element of $\boldsymbol{\xi}$, and similar notation is adopted for $\boldsymbol{\psi}$ and $\boldsymbol{\nu}$. Hence, we study the quantity $I(\boldsymbol{\xi}; \boldsymbol{\psi})$ where $\boldsymbol{\psi} = \mathsf{h}\boldsymbol{\xi} + \boldsymbol{\nu}$ and $\mathsf{h} = \mathsf{a} v \mathsf{g}$.

This section is organized as follows: we present general achievable rates derived via information-theoretical considerations in Section IV-A; then we derive the rate achievable with a practical training-based scheme using truncated-Gaussian inputs in Section IV-B.

A. General Achievable Rates

We start with the following lemma, which presents an achievable rate.

Lemma 1. Let input symbols in each coherence block satisfy average power constraint $\mathbb{E}\{\|\boldsymbol{\xi}\|_2^2\} \leqslant \ell P$ and fourth moment constraint $\mathbb{E}\{|\xi|^4\} \leqslant \beta P^2$. Capacity is lower bounded as follows:

$$\sup_{P_{\boldsymbol{\xi}} \in \mathcal{P}_{\mathrm{a}}^{\mathrm{B}} \cap \mathcal{P}_{\mathrm{f}}} \frac{1}{\ell} I(\boldsymbol{\xi}; \boldsymbol{\psi}) \geqslant R(P, \kappa) \tag{26}$$

where κ denotes the kurtosis of input symbols (cf. Section II-C) and

$$R(P, \kappa) := (1 - p_{\mathrm{B}})[h(v \mathsf{g} \xi + \nu | \mathsf{g}) - \log(\pi e)] - \tfrac{1}{\ell} \log(1 + \sigma_{\mathsf{h}}^2 \ell P). \tag{27}$$

Proof: Using the chain rule and discarding one mutual information term yields

$$I(\boldsymbol{\xi}; \boldsymbol{\psi}) \geqslant I(\boldsymbol{\xi}; \boldsymbol{\psi} | \mathsf{h}) - I(\mathsf{h}; \boldsymbol{\psi} | \boldsymbol{\xi}). \tag{28}$$

The first term in the right-hand side of (28), $I(\boldsymbol{\xi}; \boldsymbol{\psi} | \mathsf{h})$, is the mutual information for the coherent channel. Assuming independent inputs yields

$$I(\boldsymbol{\xi}; \boldsymbol{\psi} | \mathsf{h}) \overset{\text{(a)}}{=} \ell(1 - p_{\mathrm{B}}) I(\xi; v \mathsf{g} \xi + \nu | \mathsf{g}) \tag{29}$$

$$\overset{\text{(b)}}{=} \ell(1 - p_{\mathrm{B}})[h(v \mathsf{g} \xi + \nu | \mathsf{g}) - \log(\pi e)], \tag{30}$$

where (a) follows from taking the expectation with respect to a and using the assumption on independent inputs and (b) follows from the definition of mutual information and translation invariance of differential entropy. The second term in the right-hand side of (28), $I(\mathsf{h}; \boldsymbol{\psi} | \boldsymbol{\xi})$, can be bounded as follows:

$$I(\mathsf{h}; \boldsymbol{\psi} | \boldsymbol{\xi}) = h(\mathsf{h}\boldsymbol{\xi} + \boldsymbol{\nu} | \boldsymbol{\xi}) - h(\boldsymbol{\nu})$$

$$\overset{\text{(a)}}{\leqslant} \mathbb{E}\{\log(1 + \sigma_{\mathsf{h}}^2 \|\boldsymbol{\xi}\|_2^2)\} \tag{31}$$

$$\overset{\text{(b)}}{\leqslant} \log(1 + \sigma_{\mathsf{h}}^2 \ell P) \tag{32}$$

where (a) follows when h, and in turn $\mathsf{h}\boldsymbol{\xi} + \boldsymbol{\nu}$ conditioned on $\boldsymbol{\xi}$, are distributed according to a Gaussian distribution, and (b) follows from Jensen's inequality. □

In the following theorem, we apply Lemma 1 and a time-sharing argument [58], where transmission over a fraction θ of coherence blocks only is allowed, while keeping the fourth moment of input symbols constrained. Note that blocks can be chosen in time as well as frequency domain.

Theorem 4. Let the assumptions be the same as in Lemma 1. Capacity is lower bounded as follows:

$$\sup_{P_{\boldsymbol{\xi}} \in \mathcal{P}_{\mathrm{a}}^{\mathrm{B}} \cap \mathcal{P}_{\mathrm{f}}} \frac{1}{\ell} I(\boldsymbol{\xi}; \boldsymbol{\psi}) \geqslant \sup_{\kappa/\beta \leqslant \theta \leqslant 1} \theta \, R(P/\theta, \kappa/\theta). \tag{33}$$

Proof: We replace the fourth moment constraint with a kurtosis constraint $\mathrm{Kurt}\{\xi\} \leqslant \beta$, which is stricter. Denote P_ξ the distribution of ξ. In order to derive a lower bound on capacity, we further restrict ξ to have zero mean. Consider ξ' distributed according to $P_{\xi'} = (1 - \theta)\delta_0 + \theta P_{\theta^{-1/2}\xi}$ for some $\theta \in (0, 1]$. As a result $\mathbb{V}\{\xi'\} = \mathbb{V}\{\xi\} = P$ irrespective of θ and $\mathrm{Kurt}\{\xi'\} = \mathrm{Kurt}\{\xi\}/\theta = \kappa/\theta$. Following Lemma 1, if the rate achievable with P_ξ is $R(P, \kappa)$, the one achievable with $P_{\xi'}$ is $R' = \theta R(P/\theta, \kappa/\theta)$. The kurtosis constraint is $\mathrm{Kurt}\{\xi'\} \leqslant \beta$, hence $\theta \geqslant \kappa/\beta$ in (33). □

In the following corollary we specify the bound of Theorem 4 when $\beta \geqslant 2$. In this case, Gaussian inputs maximize the achievable rate in Lemma 1.

Corollary 4. Let assumptions be the same as in Lemma 1 and $\beta \geqslant 2$. Capacity is lower bounded as follows:

$$\sup_{P_{\boldsymbol{\xi}} \in \mathcal{P}_{\mathrm{a}}^{\mathrm{B}} \cap \mathcal{P}_{\mathrm{f}}} \frac{1}{\ell} I(\boldsymbol{\xi}; \boldsymbol{\psi}) \geqslant C_{\mathrm{LB}} \tag{34}$$

where:

$$C_{\mathrm{LB}} = \sup_{2/\beta \leqslant \theta \leqslant 1} \theta \, R(P/\theta) \tag{35}$$

$$\begin{aligned}R(P) &= \sup_{\kappa\leqslant\beta} R(P,\kappa)\\ &= (1-p_{\mathrm{B}})\mathbb{E}\{\log(1+\upsilon^2|\mathrm{g}|^2P)\} - \frac{\log(1+\sigma_{\hat{\mathrm{h}}}^2\ell P)}{\ell}.\end{aligned} \tag{36}$$

Proof: The differential entropy $h(\upsilon\mathrm{g}\xi+\nu|\mathrm{g})$ in Lemma 1 is maximized by $P_{\xi} = \mathcal{CN}(0,P)$ for $\beta \geqslant 2$. The result follows from Theorem 4 with $\kappa = 2$. □

Remark 2 (Geometric interpretation of the effect of time-sharing on the achievable rate). Since $\theta R(P/\theta)$ nats/symbol can be rewritten as $\theta W_{\mathrm{tot}} R(P_{\mathrm{t}}/(\theta W_{\mathrm{tot}}))$ nats/s, the supremum in (34) with respect to $\theta \in [\theta_0, 1]$, where $\theta_0 \in (0,1)$, can be interpreted as follows: any rate achievable with bandwidth $W'_{\mathrm{tot}} \in [\theta_0 W_{\mathrm{tot}}, W_{\mathrm{tot}}]$ is also achievable with bandwidth W_{tot} via time-sharing. In particular, with no peak constraint the rate achievable with bandwidth W_{tot} is the “running maximum” rate, i.e., the maximum rate achievable up to bandwidth W_{tot}.

B. Training-based Scheme with Truncated-Gaussian Inputs

In this section, we focus on the rate achievable via training with amplitude-constrained inputs. In Subsection *1)*, we adapt the exposition in [54] to the model of this paper. Then in Subsection *2)* we derive the rate achievable in the special case of amplitude-constrained inputs distributed according to a truncated-Gaussian law, which we define below.

1) Rate achievable with training-based schemes: Following [54], encoding and decoding are split into a training phase and a data transmission phase. The following partitions of signal vectors are considered: $\boldsymbol{\xi} = [\boldsymbol{\xi}_\tau^\mathsf{T}, \boldsymbol{\xi}_{\mathrm{d}}^\mathsf{T}]^\mathsf{T}$, $\boldsymbol{\psi} = [\boldsymbol{\psi}_\tau^\mathsf{T}, \boldsymbol{\psi}_{\mathrm{d}}^\mathsf{T}]^\mathsf{T}$, and $\boldsymbol{\nu} = [\boldsymbol{\nu}_\tau^\mathsf{T}, \boldsymbol{\nu}_{\mathrm{d}}^\mathsf{T}]^\mathsf{T}$, where subscripts τ and d denote training and data phases, respectively. Denote ℓ_τ and $\ell_{\mathrm{d}} = \ell - \ell_\tau$ the number of symbols devoted to training and data transmission phases, respectively. Denote ξ_{d} and ψ_{d} the generic element of $\boldsymbol{\xi}_{\mathrm{d}}$ and $\boldsymbol{\psi}_{\mathrm{d}}$, respectively. Denote $\sigma_{\mathrm{d}}^2 := \mathbb{V}\{\xi_{\mathrm{d}}\}$ and $r = \|\boldsymbol{\xi}_{\mathrm{d}}\|_\infty$. An achievable rate is derived in the following theorem.

Theorem 5. Let input symbols in each coherence block satisfy average power constraint $\mathbb{E}\{\|\boldsymbol{\xi}\|_2^2\} \leqslant \ell P$ and amplitude constraint $|\xi|^2 \leqslant \beta P$ almost surely. A training-based schemes using i.i.d. input symbols yields the following capacity lower bound:

$$C_\tau \geqslant \sup_{\substack{P_{\xi_{\mathrm{d}}}\\ \sigma_{\mathrm{d}}^2\leqslant P_{\mathrm{d}},\, r^2\leqslant\beta P,\, P_\tau\leqslant\beta P\\ \ell_\tau P_\tau+(\ell-\ell_\tau)P_{\mathrm{d}}\leqslant\ell P}} \left(1-\frac{\ell_\tau}{\ell}\right) I(\xi_{\mathrm{d}};\psi_{\mathrm{d}}|\hat{\mathrm{h}}). \tag{37}$$

Proof: See Appendix D. □

In general, $I(\xi_{\mathrm{d}};\psi_{\mathrm{d}}|\hat{\mathrm{h}})$ depends on the average and peak powers per symbol σ_{d}^2 and r^2, respectively, and the estimated channel distribution, which depends on η, υ, p_{out} (cf. Section II-B) and $E_\tau = \ell_\tau P_\tau$. Although (37) may be difficult to tackle analytically because of the form of $I(\xi_{\mathrm{d}};\psi_{\mathrm{d}}|\hat{\mathrm{h}})$, it can be easily solved numerically.

2) Rate achievable with truncated-Gaussian inputs: We specify below the rate achievable when truncated-Gaussian inputs are used in the data transmission phase. The truncated-Gaussian distribution has density

$$p_r(x) = \frac{1}{V_{r,\gamma}}\phi(x)\mathbb{1}_{\{|x|\leqslant r\}} \tag{38}$$

where $\phi(x)$ is the density of $\mathcal{CN}(0,\gamma)$ and $V_{r,\gamma}$ is the integral of $\phi(x)$ in $\mathcal{B}_r(0) = \{z \in \mathbb{C} \colon |z| \leqslant r\}$, given by:

$$V_{r,\gamma} := \int_{|x|\leqslant r} \phi(x)\,dx = 1 - e^{-r^2/\gamma}. \tag{39}$$

We denote $\mathcal{CN}_r(0,\gamma)$ the truncated-Gaussian distribution in $\mathcal{B}_r(0)$. Hence, ξ_{d} is distributed according to $\mathcal{CN}_r(0,\gamma)$. Note that σ_{d}^2 is not equal to γ: equality holds asymptotically as $r \to \infty$, while for fixed r it results

$$\sigma_{\mathrm{d}}^2 = \int_{|x|\leqslant r} |x|^2 p_r(x)\,dx = \gamma - \frac{r^2 e^{-r^2/\gamma}}{V_{r,\gamma}}. \tag{40}$$

Therefore, σ_{d}^2 is a monotonically increasing function of γ with supremum given by $r^2/2$. The amplitude constraint (10) implies that $r^2 = \sup|\mathrm{x}|^2 < \beta P = \beta\sigma_{\mathrm{d}}^2 < \beta r^2/2$: As a consequence, $\beta < 2$ is not achievable with any choice of $(r^2, \sigma_{\mathrm{d}}^2, \gamma)$. Moreover, it follows from the amplitude constraint that $r^2 \leqslant \beta P$, hence γ, r, and σ_{d}^2 are not independent. In particular, by fixing r, we will select γ to obtain a given value of σ_{d}^2. It can be shown that $p_r(x)$ is the density of the maximum differential entropy distribution among distributions with support $\mathcal{B}_r(0)$ that satisfy a variance constraint, with differential entropy given by

$$h(\xi_{\mathrm{d}}) = \mathbb{E}\{-\log p_r(x)\} = \log(\pi\gamma e^{\sigma_{\mathrm{d}}^2/\gamma} V_{r,\gamma}). \tag{41}$$

Special cases of (41) lead to two well known maximum-entropy distributions:

- $r \to \infty$: Since $V_{r,\gamma} \to 1$ and $\sigma_{\mathrm{d}}^2 \to \gamma$, (41) reduces to $h(\xi) = \log(\pi e \sigma_{\mathrm{d}}^2)$, i.e., the differential entropy of a Complex Normal distribution with variance σ_{d}^2, which is the maximum differential entropy distribution under an average power constraint only.
- $\gamma \to \infty$: Since (41) is monotonically increasing with respect to γ for fixed r, it results $\sup_{\gamma>0} h(\xi) = \lim_{\gamma\to\infty} h(\xi) = \log(\pi r^2)$, i.e., the differential entropy of a uniform random variable within the ball of radius r, which is the maximum differential entropy distribution under an amplitude constraint only.

We derive the rate achievable with truncated-Gaussian inputs in the following theorem.

Theorem 6. Let input symbols in each coherence block satisfy average power constraint $\mathbb{E}\{\|\boldsymbol{\xi}\|_2^2\} \leqslant \ell P$ and amplitude constraint $|\xi|^2 \leqslant \beta P$ a.s. with $\beta \geqslant 2$. The following rate is achievable via training and truncated-Gaussian inputs:

$$R_{\mathrm{TG}}^{\mathrm{tr}} = \sup_{0\leqslant\ell_\tau\leqslant\ell} \left(1-\frac{\ell_\tau}{\ell}\right) I_{\mathrm{LB}}(\xi_{\mathrm{d}};\psi_{\mathrm{d}}|\hat{\mathrm{h}}) \tag{42}$$

where:

$$I_{\mathrm{LB}}(\xi_{\mathrm{d}};\psi_{\mathrm{d}}|\hat{\mathrm{h}}) = \mathbb{E}_{\hat{\mathrm{h}}\sim P_{\hat{\mathrm{h}}}}\left\{\log\left(\frac{\gamma}{\hat{\gamma}}\cdot\frac{e^{\sigma_{\mathrm{d}}^2/\gamma}}{e^{\hat{\sigma}_{\mathrm{d}}^2/\hat{\gamma}}}\cdot\frac{V_{r,\gamma}}{V_{r,\hat{\gamma}}}\right)\right\} \tag{43}$$

TABLE II: List of parameters defining different SNR regimes

Notation	Variable	Case 1	Case 2	Case 3	Unit
R	Cell radius	200	200	200	m
r	User distance from center	50	150	200	m
p_{B}	Blockage probability	0	0.1	0.7	(dimensionless)
η	Line-of-sight fraction ($0 \leqslant \eta \leqslant 1$)	0.8	0.2	0	(dimensionless)
A	Additional attenuation w.r.t. free-space	15	25	30	dB
f_{c}	Carrier frequency	73	73	73	GHz
G	Antenna gain	15	15	15	dB
$P_{\mathrm{t}}^{\mathrm{tx}}$	Transmitted power	30	30	30	dBm
F	Receiver noise figure	7	7	7	dB
ℓ	Coherence block dimensions	500	500	500	(dimensionless)

$$\sigma_{\mathrm{d}}^2 = P_\tau = P_{\mathrm{d}} = P \tag{44}$$

$$\hat{\sigma}_{\mathrm{d}}^2 = \sigma_{\mathrm{d}}^2 \left(1 - \frac{|\hat{\mathsf{h}}|^2 \sigma_{\mathrm{d}}^2}{|\hat{\mathsf{h}}|^2 \sigma_{\mathrm{d}}^2 + \sigma_{\xi_{\mathrm{d}}}^2}\right) \tag{45}$$

with $\sigma_{\xi_{\mathrm{d}}}^2 = \sigma_{\tilde{\mathsf{h}}}^2 \mathbb{V}\{\xi_{\mathrm{d}}\} + 1$, $\sigma_{\tilde{\mathsf{h}}}^2 = \sigma_{\mathsf{h}}^2/(1+\sigma_{\mathsf{h}}^2 E_\tau)$, and γ and $\hat{\gamma}$ satisfying (40) for $\sigma = \sigma_{\mathrm{d}}$ and $\sigma = \hat{\sigma}_{\mathrm{d}}$, respectively.

Proof: See Appendix E. □

Remark 3 (Asymptotics as $r \to \infty$ and $\sigma_{\tilde{\mathsf{h}}} \to 0$). The above mutual information reduces to known results in some special cases. In particular, as $r \to \infty$, it results $\gamma \to \sigma_{\mathrm{d}}^2$, $\hat{\gamma} \to \hat{\sigma}_{\mathrm{d}}^2$, $V_{r,\gamma} \to 1$ and $V_{r,\hat{\gamma}} \to 1$, therefore $I_{\mathrm{LB}}(\xi_{\mathrm{d}}; \psi_{\mathrm{d}}|\hat{\mathsf{h}})$ becomes

$$\begin{aligned}\lim_{r\to\infty} I_{\mathrm{LB}}(\xi_{\mathrm{d}}; \psi_{\mathrm{d}}|\hat{\mathsf{h}}) &= \mathbb{E}\left\{\log\left(\frac{\sigma_{\mathrm{d}}^2}{\hat{\sigma}_{\mathrm{d}}^2}\right)\right\} \\ &= \mathbb{E}\left\{\log\left(1 + \frac{|\hat{\mathsf{h}}|^2\sigma_{\mathrm{d}}^2}{\sigma_{\tilde{\mathsf{h}}}^2\sigma_{\mathrm{d}}^2 + 1}\right)\right\}\end{aligned} \tag{46}$$

as was directly shown in [56]. Moreover, in the limit $\sigma_{\tilde{\mathsf{h}}}^2 \to 0$, the channel is perfectly estimated, $\hat{\mathsf{h}} = \mathsf{h}$, and (46) specializes to (the order of limits is inessential):

$$\lim_{\sigma_{\tilde{\mathsf{h}}}\to 0} \lim_{r\to\infty} I_{\mathrm{LB}}(\xi_{\mathrm{d}}; \psi_{\mathrm{d}}|\hat{\mathsf{h}}) = \mathbb{E}\{\log(1 + |\mathsf{h}|^2\sigma_{\mathrm{d}}^2)\} \tag{47}$$

which coincides with the coherent capacity.

V. Numerical Examples

We consider three cases characterizing different SNR regimes on the basis of recent mmWave experimental campaigns [31]:

- **Case 1** (User near BS): $P_{\mathrm{t}}/N_0 = 2.09 \cdot 10^9 \mathrm{s}^{-1}$;
- **Case 2** (User in *typical*[2] location): $P_{\mathrm{t}}/N_0 = 2.32 \cdot 10^7 \mathrm{s}^{-1}$;
- **Case 3** (User near cell edge): $P_{\mathrm{t}}/N_0 = 4.13 \cdot 10^6 \mathrm{s}^{-1}$.

We collect in Table II parameters that are used to produce the above listed figures of P_{t}/N_0. We assumed $N_0 = k_{\mathrm{B}}T = 4.14 \cdot 10^{-21}$ joules, where k_{B} is the Boltzmann constant and $T = 300$ kelvin. Values of P_{t}/N_0 can be directly interpreted as achievable rates in AWGN and fading channels with peak unconstrained inputs, which have infinite-bandwidth capacity equal to $C_\infty = P_{\mathrm{t}}/N_0$ nats/s.

[2]We will define later in the section the precise meaning of *typical*.

In the presence of blockages, the maximum achievable rate without peak constraint is $(1-p_{\mathrm{B}})C_\infty$, which numerically corresponds to approximately 3 Gb/s (Case 1), 30 Mb/s (Case 2), and 1.8 Mb/s (Case 3). In order to keep the number of parameters as low as possible, we focus on the case of no absorption. We consider cell radius and user distance from the center, where the base station (BS) is located, equal to R and r, respectively. Pathloss PL is derived by adjusting the free-space pathloss at distance r and carrier frequency f_{c} with a further attenuation $A = A(r, f_{\mathrm{c}})$ to match experimental data [31]:

$$\mathrm{PL\ (dB)} = 20\log_{10} r + 20\log_{10} f_{\mathrm{c}} + 20\log_{10}(4\pi/c) + A\ \mathrm{(dB)}$$

where c is the speed of light. Denote $P_{\mathrm{t}}^{\mathrm{tx}}$ the transmitted power, G (dB) the antenna gains, and F the receiver noise figure. The received power P_{t} is

$$P_{\mathrm{t}}\ \mathrm{(dBm)} = P_{\mathrm{t}}^{\mathrm{tx}}\ \mathrm{(dBm)} + G\ \mathrm{(dB)} - F\ \mathrm{(dB)} - N_0/(1\mathrm{J})\ \mathrm{(dB)}.$$

Due to propagation and pathloss model, numerical results plotted on figures are to be intended valid for $W_{\mathrm{tot}} \ll f_{\mathrm{c}}$, e.g. $W_{\mathrm{tot}} \leqslant f_{\mathrm{c}}/10 \approx 7.3$ GHz (see Table II). In any case, $W_{\mathrm{tot}} \leqslant f_{\mathrm{c}}$ from physical considerations (nonshaded areas on figures).

Figures 2 and 3 show rates (Mb/s) as a function of bandwidth W_{tot} (Hz) for Case 1 and 2, respectively, with $\beta = 3$. Curves on figure refer to: AWGN upper bound C_{AWGN}; coherent upper bound C_{coh} from eq. (13); capacity upper bounds with amplitude and fourth moment constraints from eqs. (17) and (25), respectively, both indicated with C_{UB} since the difference is not appreciable on plots; capacity lower bound C_{LB} from eq. (34); rates achievable via training along with truncated-Gaussian inputs $R_{\mathrm{TG}}^{\mathrm{tr}}$ from eq. (42). In Fig. 2, the communication system operates in a relatively high-SNR regime per degree of freedom (e.g. SNR ≈ 3 dB when $W_{\mathrm{tot}} = 1$ GHz). The curve corresponding to the capacity lower bound is almost overlapped with that of the coherent capacity upper bound. Moreover, the rate achievable by using training and truncated-Gaussian inputs is shown to be close to the upper bound. Therefore, it is shown that dense signaling schemes are sufficient to achieve rates in the order of Gb/s. In Fig. 3, the user approximately lies on the boundary of a circle that partitions the cell in two regions with equal area: when

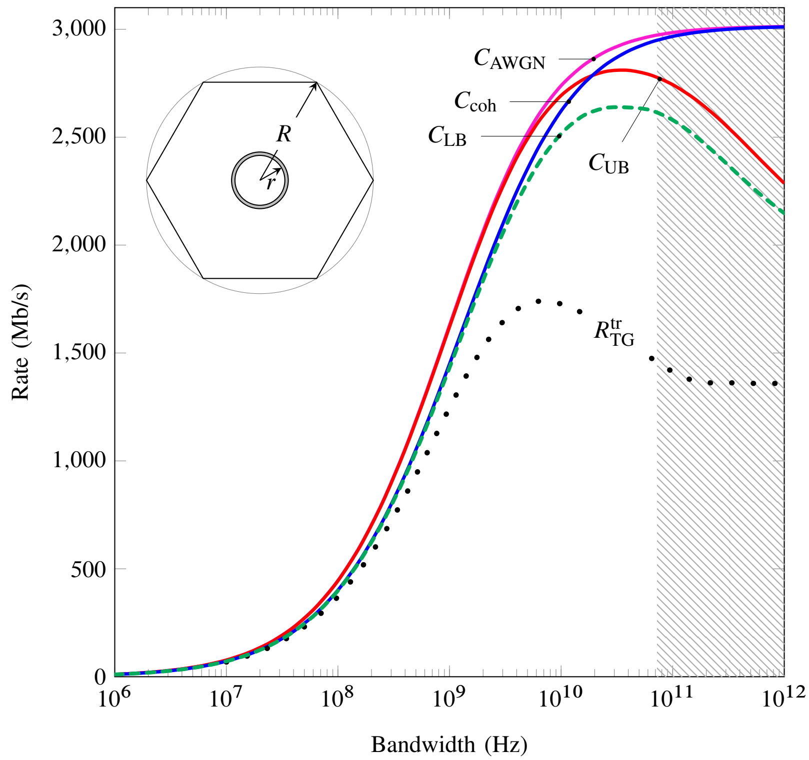


FIG. 2: Case 1: User near cell center (BS). Rate (Mb/s) as a function of bandwidth W_{tot} (Hz). Scenario parameters are specified in column "Case 1" of Table II.

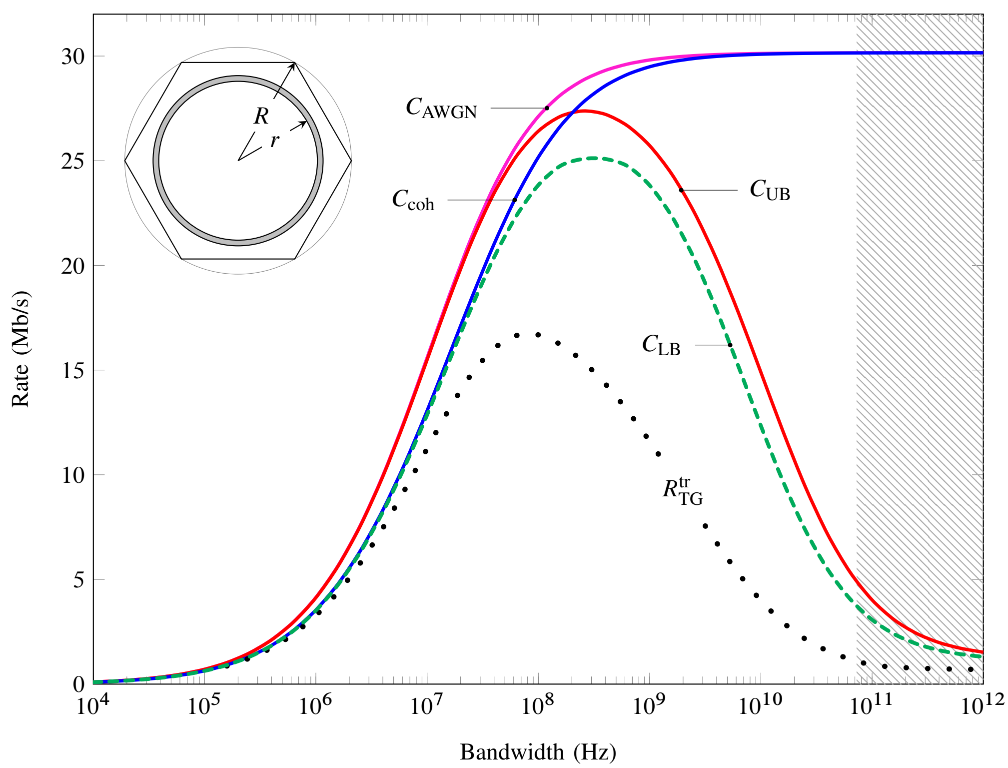


FIG. 3: Case 2: User in *typical* location. Rate (Mb/s) as a function of bandwidth W_{tot} (Hz). Scenario parameters are specified in column "Case 2" of Table II.

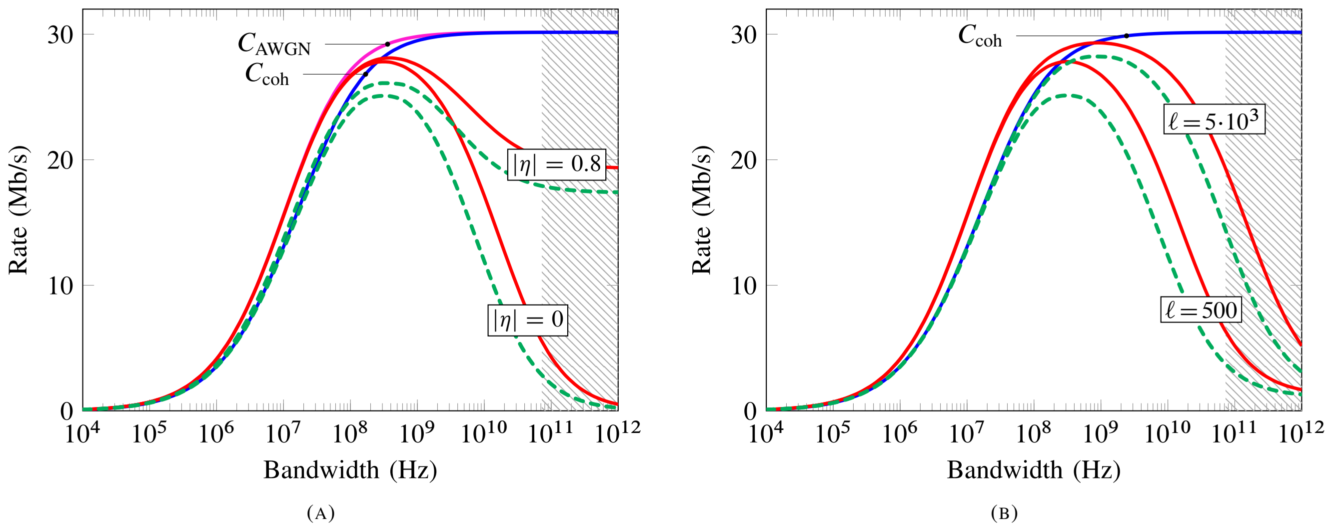


FIG. 4: Capacity upper bounds (solid line) and lower bounds (dashed line) as a function of bandwidth. (A): different values of LOS component $|\eta| \in \{0, 0.8\}$. (B): different block sizes $\ell \in \{500, 5\cdot 10^3\}$.

users are placed uniformly at random, approximately half of the users is closer to and the other half farther from the BS than the user on the boundary. We referred to this location on the boundary as "typical." In this case the user is no longer in the high-SNR regime at bandwidths of interest (e.g. SNR ≈ -16 dB when $W_{\mathrm{tot}} = 1$ GHz), and the system enters the wideband regime where the slope of upper and lower bounds on capacity is negative. Although amplitude-constrained inputs using training experience a significant rate degradation, upper and lower bounds are close to each other, and also close to the wideband capacity. Farther users, such as those described by Case 3, experience more severe degradation. We do not provide a figure for Case 3, which is qualitatively similar to Fig. 3: in this case, the user is located at a distance comparable to the typical mmWave cell radius, and communication occurs in the deep low-SNR regime. The wideband capacity is reduced due to strong attenuation and blockages. Further rate degradation is experienced by dense signaling schemes: in particular, the capacity upper bound is much lower in this case than the wideband capacity without peak constraint.

Fig. 4A shows capacity upper bounds (from (25)) and lower bounds (from (34)) as a function of bandwidth for NLOS ($|\eta| = 0$) and LOS ($|\eta| = 0.8$) scenarios. While insensitive for sufficiently small bandwidths, rate depends on the strength of the specular component in the wideband region. Consistently with the curves on figure, eq. (36) predicts an achievable rate in the infinite-bandwidth limit equal to $(1 - p_{\mathrm{B}})^2 |\eta|^2 (\log_2 e) P_{\mathrm{t}}/N_0$ bits/s, which corresponds to 0 Mb/s for $|\eta| = 0$ and to approximately 17.4 Mb/s for $|\eta| = 0.8$. Similarly, eqs. (17) and (25) predict an upper bound equal to $(1 - p_{\mathrm{B}}) |\eta|^2 (\log_2 e) P_{\mathrm{t}}/N_0$ bits/s, which corresponds to 0 Mb/s for $|\eta| = 0$ and to approximately 19.3 Mb/s for $|\eta| = 0.8$.

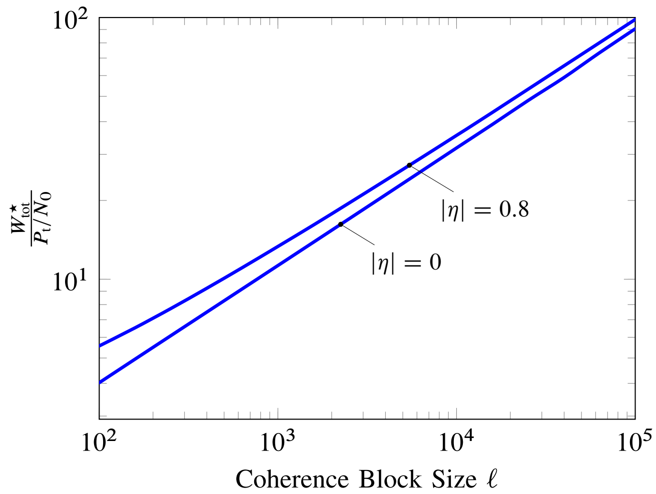
Fig. 4B illustrates upper and lower bounds on the capacity as a function of bandwidth for different block sizes $\ell \in \{500, 5\cdot 10^3\}$. It is shown that, as the block size ℓ grows, the wideband regime shifts towards higher bandwidth, and the width of the transition between the narrowband and the wideband regimes increases. Bounds are approximately parallel for a large portion of the wideband regime, until reaching a value close to the wideband limit, which is not affected by the block size (upper and lower bounds tend to 1.20 Mb/s and 1.08 Mb/s, respectively).

Fig. 5A shows the critical bandwidth $W^\star_{\mathrm{tot}}$ (normalized with respect to P_{t}/N_0) corresponding to the maximum rate achievable with Gaussian inputs as a function of the block size, for NLOS ($|\eta| = 0$) and LOS ($|\eta| = 0.8$) scenarios. It is shown that curves are approximately linear in both cases. This behavior can be qualitatively explained by noting that the critical bandwidth $W^\star_{\mathrm{tot}}$ lies in the frequency interval where the penalty term in eqs. (17), (25), and (36) becomes non-negligible as bandwidth increases, e.g. when $\ell P = \ell P_{\mathrm{t}}/(W_{\mathrm{tot}} N_0) \approx 1$, and thus $W^\star_{\mathrm{tot}}$ is linear with respect to ℓ; it can be interpreted as the bandwidth on the boundary between high SNR and low SNR per degree of freedom.

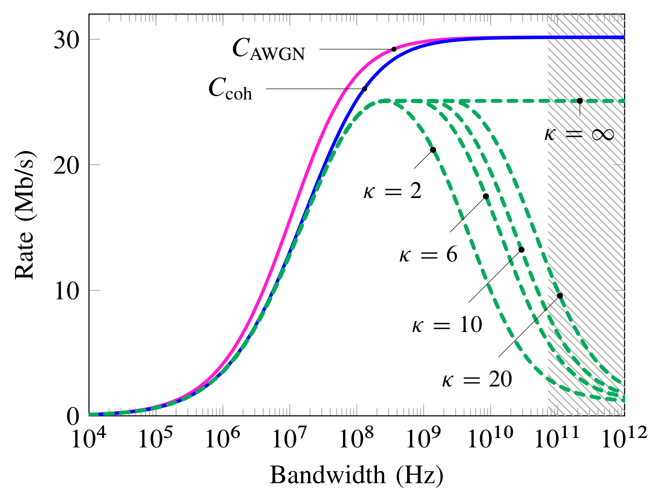
Fig. 5B shows achievable rates with inputs having different kurtosis, according to (34), as a function of bandwidth. It is shown that it is possible to push the wideband regime towards larger bandwidths by using time-sharing, i.e., by transmitting inputs over a subset of blocks only. We note that the shift is relatively slow. As kurtosis increases, it is possible to hold the "running maximum" rate (cf. Remark 2), hence time-sharing (with no peak constraint) provides a method to achieve a rate that is a nondecreasing function of bandwidth even when the channel is noncoherent.

VI. CONCLUSION

This paper investigated the noncoherent capacity of doubly-dispersive block-fading channels that include essential features of mmWave propagation, and presented upper and lower bounds on the capacity as a function of bandwidth, LOS strength, blockage probability, and coherence block size.



(A)



(B)

FIG. 5: (A): Normalized critical bandwidth corresponding to the maximum of the mutual information with Gaussian inputs as a function of block size, for NLOS ($\eta = 0$) and LOS ($\eta = 0.8$) scenarios. (B): Achievable rates with Gaussian inputs used on a fraction of the blocks as a function of the bandwidth. The fraction of blocks where transmissions occur is such that the input distribution has kurtosis κ.

Bounds were derived by leveraging on previous investigations, in particular [6], [41], which focused on Rayleigh fading. We showed that rate behaves differently from the Rayleigh fading case, especially in the wideband regime. In particular, LOS propagation guarantees a nonzero rate with traditional, "non-peaky" signaling. We showed that large bandwidth and severe attenuation can force mmWave systems to operate in the wideband regime and experience severe rate degradations from the wideband limit achievable without peak constraint.

We presented numerical results based on recent experimental campaign accounting for a scenario where an outdoor user is located at different distances from a BS. Aside from the technological issues of implementing OFDM or similar systems with several thousand subcarriers [12], "non-peaky" signaling was confirmed to be nearly capacity-achieving in the high-SNR regime. This corresponds to the case of a user that is sufficiently close to the BS to experience a channel with strong LOS and relatively low attenuation. However, a large fraction of users is not typically sufficiently close to the BS to experience such favorable propagation and high SNR. In the typical setting, NLOS propagation paired with "non-peaky" signaling was shown to significantly reduce the achievable rate from the wideband capacity without peak constraint, hence showing the importance of signaling "peakedness" (sparsity) for farther users.

Since users can, therefore, experience both high- and low-SNR per degree of freedom within cells of relatively small size such as mmWave cells, mmWave communications may have to support both "non-peaky" (dense) and "peaky" (sparse) signaling in order to maximize the achievable rate for a given power expenditure. Indeed, "non-peaky" signaling achieves rate close to capacity in the strong LOS or high-SNR scenario, but it is outperformed in the NLOS, low-SNR scenario by "peaky" signaling. Therefore, the rationale for the signaling scheme is that the farther the user, the sparser the signaling must become in order to avoid a significant rate degradation with respect to the wideband capacity without peak constraint.

APPENDIX A
PROOF OF THEOREM 1

Since inputs are independent on the channel realization, the derivation at the top of next page holds, where: (a) follows by considering Gaussian inputs with correlation $\boldsymbol{R}_{\mathsf{xx}} := \mathbb{E}\{\, \mathbf{xx}^\dagger \,\}$; (b) follows from applying Hadamard's inequality to $\mathbf{hh}^\dagger \odot \boldsymbol{R}_{\mathsf{xx}}$, which holds with equality when $\boldsymbol{R}_{\mathsf{xx}}$ is diagonal, i.e., when $\mathbf{x}$ has independent elements; (c) follows from averaging with respect to the blockage process and denoting $P_{ijkl} := \mathbb{E}\{\, |[\mathbf{x}_{ij}]_{kl}|^2 \,\}$; (d) follows from the optimal power allocation $\{P_1^\star, P_2^\star, \ldots, P_{D_{\mathrm{W}}}^\star\}$. In particular, writing the functional

$$J(P_1, P_2, \ldots, P_{D_{\mathrm{W}}}) = \sum_{j=1}^{D_{\mathrm{W}}} \mathbb{E}\{\, \log(1 + v_j^2 |\mathrm{g}|^2 P_j) \,\} + \lambda \sum_{j=1}^{D_{\mathrm{W}}} P_j \tag{48}$$

and differentiating with respect to P_j, yields $\partial J / \partial P_j = \mathbb{E}\{\, (v_j^2 |\mathrm{g}|^2)/(1 + v_j^2 |\mathrm{g}|^2 P_j) \,\} + \lambda = 0$. Denote $\psi(\xi) = \mathbb{E}\{\, 1/(1 + \xi |\mathrm{g}|^2) \,\}$ and let $p = p(v, \lambda)$ be the solution of $p^{-1}[1 - \psi(vp)] + \lambda = 0$. It follows $P_j^\star = P_j^\star(\lambda) = (p(v_j, \lambda))^+$ with λ such that $\sum_{j=1}^{D_{\mathrm{W}}} P_j^\star(\lambda) = D_{\mathrm{W}} P$.

APPENDIX B
PROOF OF THEOREM 2

When the receiver has the side information of those coherence times where the channel is blocked, mutual information

$$
\begin{aligned}
\sup_{\substack{P_{\mathsf{x}} \\ \mathbb{E}\{\|\mathsf{x}\|_2^2\}\leqslant KLP}} I(\mathsf{x};\mathsf{y}) &\leqslant \sup_{\substack{P_{\mathsf{x}} \\ \mathbb{E}\{\|\mathsf{x}\|_2^2\}\leqslant KLP}} I(\mathsf{x};\mathsf{y}|\mathbf{h}) \\
&\overset{\text{(a)}}{=} \sup_{\substack{\boldsymbol{R}_{\mathsf{xx}} \\ \operatorname{tr}\{\boldsymbol{R}_{\mathsf{xx}}\}\leqslant KLP}} \mathbb{E}_{\mathbf{h}\sim P_{\mathbf{h}}}\{\log\det(\boldsymbol{I}+\mathbf{h}\mathbf{h}^\dagger\odot\boldsymbol{R}_{\mathsf{xx}})\} \\
&\overset{\text{(b)}}{=} \sup_{\mathbb{E}\{\|\mathsf{x}\|_2^2\}\leqslant KLP} \mathbb{E}_{\mathbf{h}\sim P_{\mathbf{h}}}\left\{\sum_{i=1}^{D_{\mathrm{T}}}\sum_{j=1}^{D_{\mathrm{W}}}\sum_{k=1}^{\ell_{\mathrm{T}}}\sum_{l=1}^{\ell_{\mathrm{W}}}\log(1+|\mathsf{h}_{ij}|^2\,\mathbb{E}\{\,|[\mathsf{x}_{ij}]_{kl}|^2\,\})\right\} \\
&\overset{\text{(c)}}{=} (1-p_{\mathrm{B}}) \sup_{\sum_{ijkl}P_{ijkl}\leqslant KLP}\sum_{i=1}^{D_{\mathrm{T}}}\sum_{j=1}^{D_{\mathrm{W}}}\sum_{k=1}^{\ell_{\mathrm{T}}}\sum_{l=1}^{\ell_{\mathrm{W}}}\mathbb{E}_{\mathsf{g}\sim P_{\mathsf{g}}}\{\log(1+v_j^2|\mathsf{g}|^2P_{ijkl})\} \\
&\overset{\text{(d)}}{=} (1-p_{\mathrm{B}})\ell_{\mathrm{T}}\ell_{\mathrm{W}}D_{\mathrm{T}}\sum_{j=1}^{D_{\mathrm{W}}}J(v_j^2P_j^\star)
\end{aligned}
$$

is upper bounded as follows:

$$
\begin{aligned}
I(\mathbf{X};\mathbf{Y}) &\leqslant I(\mathbf{X};\mathbf{Y}|\mathsf{a}_1,\mathsf{a}_2,\ldots,\mathsf{a}_{D_{\mathrm{T}}}) \\
&= \sum_{i=1}^{D_{\mathrm{T}}} I(\mathbf{x}_i;\mathbf{y}_i|\mathsf{a}_i) \\
&= \sum_{i=1}^{D_{\mathrm{T}}}(1-p_{\mathrm{B}})I(\mathbf{x}_i;\mathbf{y}_i|\mathsf{a}_i=1) \\
&= (1-p_{\mathrm{B}})I(\mathbf{x};\acute{\mathbf{y}})
\end{aligned} \tag{49}
$$

where $\acute{\mathbf{y}} = \acute{\mathbf{h}}\odot\mathbf{x}+\mathbf{n}$, and $\acute{\mathbf{h}} = \boldsymbol{v}\odot\mathbf{g}$ denotes the channel vector when $\mathsf{a}_1=\mathsf{a}_2=\cdots=\mathsf{a}_{D_{\mathrm{T}}}=1$. Applying the chain rule for mutual information yields

$$
\begin{aligned}
&\sup_{\substack{P_{\mathsf{x}}\in\mathcal{P}_{\mathrm{m}} \\ \mathbb{E}\{\|\mathsf{x}\|_2^2\}\leqslant KLP}} I(\mathbf{x};\acute{\mathbf{y}}) \\
&\overset{\text{(a)}}{=} \sup_{\substack{P_{\mathsf{x}}\in\mathcal{P}_{\mathrm{m}} \\ \mathbb{E}\{\|\mathsf{x}\|_2^2\}\leqslant KLP}} \left\{I(\acute{\mathbf{y}};\acute{\mathbf{h}},\mathbf{x})-I(\acute{\mathbf{y}};\acute{\mathbf{h}}|\mathbf{x})\right\} \\
&\overset{\text{(b)}}{=} \sup_{q\in[0,1]}\sup_{\substack{P_{\mathsf{x}}\in\mathcal{P}_{\mathrm{m}} \\ \mathbb{E}\{\|\mathsf{x}\|_2^2\}=qKLP}} \Big\{I(\acute{\mathbf{h}}\odot\mathbf{x}+\mathbf{n};\acute{\mathbf{h}}\odot\mathbf{x}) \\
&\qquad\qquad -\mathbb{E}_{\boldsymbol{x}\sim P_{\mathsf{x}}}\{\,I(\acute{\mathbf{h}}\odot\boldsymbol{x}+\mathbf{n};\acute{\mathbf{h}})\,\}\Big\}
\end{aligned} \tag{50}
$$

where (a) results from the chain rule, and (b) follows from $\acute{\mathbf{h}}\odot\mathbf{x}$ being a sufficient statistic of $(\acute{\mathbf{h}},\mathbf{x})$ for $\acute{\mathbf{y}}$ and "splitting" the supremum. The two terms in eq. (50) are bounded below.

1) First term:

$$
\begin{aligned}
&\sup_{\substack{P_{\mathsf{x}}\in\mathcal{P}_{\mathrm{m}} \\ \mathbb{E}\{\|\mathsf{x}\|_2^2\}=qKLP}} I(\acute{\mathbf{h}}\odot\mathbf{x}+\mathbf{n};\acute{\mathbf{h}}\odot\mathbf{x}) \\
&\overset{\text{(a)}}{\leqslant} \sup_{\substack{P_{\mathsf{x}} \\ \mathbb{E}\{\|\mathsf{x}\|_2^2\}=qKLP}} I(\acute{\mathbf{h}}\odot\mathbf{x}+\mathbf{n};\acute{\mathbf{h}}\odot\mathbf{x}) \\
&\overset{\text{(b)}}{\leqslant} \sup_{\substack{P_{\mathsf{u}} \\ \operatorname{tr}\{\boldsymbol{R}_{\mathsf{xx}}\}=qKLP \\ \boldsymbol{R}_{\mathsf{uu}}=\boldsymbol{R}_{\acute{\mathsf{h}}\acute{\mathsf{h}}}\odot\boldsymbol{R}_{\mathsf{xx}}}} I(\mathbf{u}+\mathbf{n};\mathbf{u}) \\
&\overset{\text{(c)}}{=} \sup_{\substack{\boldsymbol{R}_{\mathsf{xx}},\mathbb{E}\{\mathsf{x}\}=\mathbf{0}, \\ \operatorname{tr}\{\boldsymbol{R}_{\mathsf{xx}}\}=qKLP}} \log\det(\boldsymbol{I}+\boldsymbol{R}_{\mathsf{xx}}\odot\boldsymbol{R}_{\acute{\mathsf{h}}\acute{\mathsf{h}}}) \\
&\overset{\text{(d)}}{=} \sup_{\sum_{ijkl}P_{ijkl}=qKLP}\sum_{i=1}^{D_{\mathrm{T}}}\sum_{j=1}^{D_{\mathrm{W}}}\sum_{k=1}^{\ell_{\mathrm{T}}}\sum_{l=1}^{\ell_{\mathrm{W}}}\log(1+P_{ijkl}\,\mathbb{E}\{\,|\acute{\mathsf{h}}_{ij}|^2\,\}) \\
&\overset{\text{(e)}}{=} D_{\mathrm{T}}\ell\sum_{j=1}^{D_{\mathrm{W}}}\log(1+P_j^\star v_j^2)
\end{aligned} \tag{51}
$$

where: (a) follows from removing the amplitude constraint; (b) results from taking the optimum over $\mathbf{h}\odot\mathbf{x}$, rather than $\mathbf{x}$ alone, with a given covariance structure; (c) follows from assigning the differential entropy maximizing distribution to $\mathbf{u}$, which is multivariate complex Gaussian with $\mathbb{E}\{\,\mathbf{x}\,\}=\mathbf{0}$; (d) follows from Hadamard's inequality, which holds with equality with diagonal $\boldsymbol{R}_{\mathsf{xx}}$; (e) results from the optimum water-filling solution to the power allocation problem, i.e., $P_{ijkl}^\star := (\nu-1/v_j^2)^+ =: P_j^\star$ with ν such that $\sum_{ijkl}P_{ijkl}^\star = \ell D_{\mathrm{T}}\sum_j P_j^\star = qKLP$.

2) Second term: Denote $\acute{\mathbf{h}}_i$ the channel vector equal to $\mathbf{h}_i$ in (6) with $\mathsf{a}_i=1$. The argument of the expectation of the second term in (50) can be simplified since $\{\acute{\mathbf{h}}_i\colon 1\leqslant i\leqslant D_{\mathrm{T}}\}$ are independent, and $\acute{\mathbf{h}}_i$, and in turn $\acute{\mathbf{h}}_i\odot\boldsymbol{x}_i+\mathbf{n}_i$, is Gaussian. Thus

$$
\begin{aligned}
I(\acute{\mathbf{h}}\odot\boldsymbol{x}+\mathbf{n};\acute{\mathbf{h}}) &= \sum_{i=1}^{D_{\mathrm{T}}} I(\acute{\mathbf{h}}_i\odot\boldsymbol{x}_i+\mathbf{n}_i;\acute{\mathbf{h}}_i) \\
&= \sum_{i=1}^{D_{\mathrm{T}}}\sum_{j=1}^{D_{\mathrm{W}}}\log(1+v_j^2\|\boldsymbol{x}_{ij}\|_2^2(1-|\eta|^2)).
\end{aligned} \tag{52}
$$

We will find a lower bound of the above expression as a function of the parameters in the power constraints. We adapt the idea of [41, eq. (31)] to the block-fading channel, and bound (52) as follows:

$$
\begin{aligned}
&I(\acute{\mathbf{h}}\odot\boldsymbol{x}+\mathbf{n};\acute{\mathbf{h}}) \\
&\overset{\text{(a)}}{\geqslant} \sum_{i=1}^{D_{\mathrm{T}}}\|\boldsymbol{x}_i\|_2^2 \inf_{\|\boldsymbol{u}_i\|_\infty^2\leqslant\beta P}\frac{\sum_{j=1}^{D_{\mathrm{W}}}\log(1+v_j^2\|\boldsymbol{u}_{ij}\|_2^2(1-|\eta|^2))}{\|\boldsymbol{u}_i\|_2^2}
\end{aligned}
$$

$$\stackrel{\text{(b)}}{=} \sum_{i=1}^{D_{\mathrm{T}}} \|\boldsymbol{x}_i\|_2^2 \frac{\sum_{j=1}^{D_{\mathrm{W}}} \log(1+v_j^2 \ell P_j^\circ(1-|\eta|^2))}{\ell P_{\mathrm{sum}}^\circ}$$
$$\stackrel{\text{(c)}}{=} \|\boldsymbol{x}\|_2^2 \frac{\sum_{j=1}^{D_{\mathrm{W}}} \log(1+v_j^2 \ell P_j^\circ(1-|\eta|^2))}{\ell P_{\mathrm{sum}}^\circ} \tag{53}$$

where: (a) follows from taking the infimum over time blocks; (b) results from $\{P_1^\circ, P_2^\circ, \ldots, P_{D_{\mathrm{W}}}^\circ\}$ being the power allocation that achieves the infimum, where $P_{\mathrm{sum}}^\circ = \sum_{j=1}^{D_{\mathrm{W}}} P_j^\circ$; (c) follows from noting that the infimum does not depend on time index. It turns out [40] that P_j° is on the boundary of the possible allocation interval, i.e., either $P_j^\circ = 0$ or $P_j^\circ = \beta P$, and it depends on the profile $(v_1, v_2, \ldots, v_{D_{\mathrm{W}}})$ only.

3) Completing the bound: Using (53) and (51) into (50), and (50) in (49) yields (14).

Appendix C
Proof of Theorem 3

The following lemma is used in the proof of Theorem 3.

Lemma 2. Let a be a nonnegative random variable with mean m_1, and $\gamma > 0$ and $m_2 > m_1^2$ fixed constants. Then

$$\inf_{\substack{\mathbb{E}\{\mathsf{a}\}=m_1 \\ \mathbb{E}\{\mathsf{a}^2\}\leqslant m_2}} \mathbb{E}\{\log(1+\gamma \mathsf{a})\} = \frac{m_1^2}{m_2}\log\left(1+\gamma\frac{m_2}{m_1}\right) \tag{54}$$

where the infimum is achieved by the two-mass distribution $P_A^\star = (1-m_1^2/m_2)\delta_0 + (m_1^2/m_2)\delta_{m_2/m_1}$.

Proof: Note that the set of distributions satisfying the constraints is convex and compact. Since $\mathbb{E}\{\log(1+\mathsf{a})\}$ is linear in the distribution of a, denoted by P_{a}, the minimum is achieved at the extreme points, hence P_{a} is a two-mass distribution [65], i.e., $P_{\mathsf{a}}^\star = (1-\theta)\delta_0 + \theta\delta_{m_1/\theta}$, $\theta \in [0,1]$. Hence the problem reduces to the minimization of $\mathbb{E}\{\log(1+A)\} = \theta\log(1+\gamma m_1/\theta)$ with respect to θ subject to $\mathbb{E}\{\mathsf{a}^2\} = m_1^2/\theta \leqslant m_2$, i.e., $\theta \geqslant m_1^2/m_2$. Since the objective function is increasing in θ, the optimal solution is achieved by $\theta = m_1^2/m_2$. □

Proof of Theorem 3: The proof is similar to that of Theorem 2, the different step lying in the lower bound of $I(\acute{\mathbf{y}};\acute{\mathsf{h}}|\mathbf{x})$ under the different peak constraint, as specified on the top of the next page, where (a) follows from splitting the infimum and noticing that the fourth moment constraint is set elementwise, (b) is obtained by relaxing the fourth moment constraint on a per-block basis as detailed below, and (c) follows from Lemma 2. The relaxation is as follows:

$$\mathbb{E}\{\|\mathbf{x}_{ij}\|_2^4\} = \sum_{kl}\sum_{k'l'} \mathbb{E}\{|[\mathbf{x}_{ij}]_{kl}|^2 |[\mathbf{x}_{ij}]_{k'l'}|^2\}$$
$$\stackrel{\text{(a)}}{\leqslant} \sum_{kl}\sum_{k'l'} \mathbb{E}\{|[\mathbf{x}_{ij}]_{kl}|^4\}^{1/2}\,\mathbb{E}\{|[\mathbf{x}_{ij}]_{k'l'}|^4\}^{1/2}$$
$$\stackrel{\text{(b)}}{\leqslant} \sum_{kl}\sum_{k'l'} \beta P^2 = \ell^2\beta P^2$$

where (a) follows from the Cauchy-Schwarz inequality, and (b) from the fourth moment constraint. Finally, since $\mathbb{E}\{|[\mathbf{x}_{ij}]_{kl}|^2\}^2 \leqslant \mathbb{E}\{|[\mathbf{x}_{ij}]_{kl}|^4\} \leqslant \beta P^2$, then $\mathbb{E}\{\|\mathbf{x}_{ij}\|_2^2\} = \sum_{kl}\mathbb{E}\{|[\mathbf{x}_{ij}]_{kl}|^2\} \leqslant \ell\sqrt{\beta}P$. □

Appendix D
Proof of Theorem 5

The average power constraint is $\mathbb{E}\{\|\boldsymbol{\xi}\|_2^2\} = \ell_\tau P_\tau + \ell_{\mathrm{d}} P_{\mathrm{d}} \leqslant \ell P$. We detail below the two phases.

1) **Training phase:** Receiver estimates the channel coefficient h from the observations of the signal received upon transmission of the sequence $\boldsymbol{\xi}_\tau$:

$$\boldsymbol{\psi}_\tau = \mathsf{h}\boldsymbol{\xi}_\tau + \boldsymbol{\nu}. \tag{59}$$

The LMMSE estimate of h is [54]:

$$\hat{\mathsf{h}} = \boldsymbol{w}_\tau^\dagger(\boldsymbol{\psi}_\tau - \mathbb{E}\{\boldsymbol{\psi}_\tau\}) + \mathbb{E}\{\mathsf{h}\}, \tag{60}$$

where:

$$\boldsymbol{w}_\tau^\dagger = \boldsymbol{C}_{\mathsf{h}\boldsymbol{\psi}_\tau}\boldsymbol{C}_{\boldsymbol{\psi}_\tau}^{-1}, \tag{61}$$
$$\boldsymbol{C}_{\mathsf{h}\boldsymbol{\psi}_\tau} = \sigma_{\mathsf{h}}^2\boldsymbol{\xi}_\tau^\dagger, \tag{62}$$
$$\boldsymbol{C}_{\boldsymbol{\psi}_\tau} = \sigma_{\mathsf{h}}^2\boldsymbol{\xi}_\tau\boldsymbol{\xi}_\tau^\dagger + \boldsymbol{I}. \tag{63}$$

In general, h is nonzero-mean, which makes the estimator expression different from that in [54]. Rearranging terms in $\boldsymbol{w}_\tau^\dagger$ and using the matrix inversion lemma yields

$$\boldsymbol{w}_\tau^\dagger = \frac{\sigma_{\mathsf{h}}^2}{1+\sigma_{\mathsf{h}}^2 E_\tau}\boldsymbol{\xi}_\tau^\dagger \tag{64}$$

where $E_\tau = \|\boldsymbol{\xi}_\tau\|^2$. We assume that $\boldsymbol{\xi}_\tau$ is a sequence of antipodal symbols with amplitude $\sqrt{P_\tau}$, hence the peak power constraint during the training phase reads $P_\tau \leqslant \beta P$. Denote $\mathsf{h}_0 := \mathsf{h} - \mathbb{E}\{\mathsf{h}\}$. The estimate (60) can be rewritten as

$$\hat{\mathsf{h}} = \frac{\sigma_{\mathsf{h}}^2}{1+\sigma_{\mathsf{h}}^2 E_\tau}\boldsymbol{\xi}_\tau^\dagger(\mathsf{h}_0\boldsymbol{\xi}_\tau + \boldsymbol{\nu}_\tau) + \mathbb{E}\{\mathsf{h}\}$$
$$= \alpha\mathsf{h}_0 + \zeta_\tau + \mathbb{E}\{\mathsf{h}\} \tag{65}$$

where $\zeta_\tau \sim \mathcal{CN}(0,\sigma_{\zeta_\tau}^2)$, with $\sigma_{\zeta_\tau}^2 = \alpha^2/E_\tau$ and $\alpha = \sigma_{\mathsf{h}}^2 E_\tau/(1+\sigma_{\mathsf{h}}^2 E_\tau)$. Let the true channel coefficient h be the sum of the estimated channel $\hat{\mathsf{h}}$ and the error $\tilde{\mathsf{h}}$:

$$\mathsf{h} = \hat{\mathsf{h}} + \tilde{\mathsf{h}}. \tag{66}$$

Given the distribution of the true channel (cf. (4)), it follows that:

$$P_{\hat{\mathsf{h}}|\mathsf{a}=0} = \mu_{(1-\alpha)\mathbb{E}\{\mathsf{h}\},\sigma_{\zeta_\tau}^2} \tag{67}$$
$$P_{\hat{\mathsf{h}}|\mathsf{a}=1} = \mu_{\frac{1-(1-\alpha)p_{\mathrm{B}}}{1-p_{\mathrm{B}}}\mathbb{E}\{\mathsf{h}\},\alpha^2 v^2(1-|\eta|^2)+\sigma_{\zeta_\tau}^2} \tag{68}$$
$$P_{\hat{\mathsf{h}}} = p_{\mathrm{B}}P_{\hat{\mathsf{h}}|\mathsf{a}=0} + (1-p_{\mathrm{B}})P_{\hat{\mathsf{h}}|\mathsf{a}=1}. \tag{69}$$

It can be shown that the variance of $\hat{\mathsf{h}}$ and $\tilde{\mathsf{h}}$ are related as follows: $\sigma_{\hat{\mathsf{h}}}^2 + \sigma_{\tilde{\mathsf{h}}}^2 = \sigma_{\mathsf{h}}^2$, $\sigma_{\tilde{\mathsf{h}}}^2 = \sigma_{\mathsf{h}}^2/(1+\sigma_{\mathsf{h}}^2 E_\tau)$.

2) **Data transmission phase:** The received signal during this phase is

$$\boldsymbol{\psi}_{\mathrm{d}} = \mathsf{h}\boldsymbol{\xi}_{\mathrm{d}} + \boldsymbol{\nu}_{\mathrm{d}} = \hat{\mathsf{h}}\boldsymbol{\xi}_{\mathrm{d}} + \boldsymbol{\zeta}_{\mathrm{d}}, \tag{70}$$

where $\boldsymbol{\zeta}_{\mathrm{d}} = \tilde{\mathsf{h}}\boldsymbol{\xi}_{\mathrm{d}} + \boldsymbol{\nu}_{\mathrm{d}}$. The following lower bound to the capacity of any training scheme is derived in [54]:

$$C_\tau = \sup_{\substack{P_{\boldsymbol{\xi}_{\mathrm{d}}} \\ \mathbb{E}\{\|\boldsymbol{\xi}_{\mathrm{d}}\|_2^2\}\leqslant \ell_{\mathrm{d}}P_{\mathrm{d}} \\ \|\boldsymbol{\xi}_{\mathrm{d}}\|_\infty^2 \leqslant \beta P}} \frac{1}{\ell} I(\boldsymbol{\xi}_{\mathrm{d}};\boldsymbol{\psi}_{\mathrm{d}}|\hat{\mathsf{h}}) \tag{71}$$

$$\inf_{\substack{P_{\mathbf{x}}\in\mathcal{P}_{\mathrm{f}}\\ \mathbb{E}\{\|\mathbf{x}\|_2^2\}=D_{\mathrm{T}}D_{\mathrm{W}}\ell q P}} \mathbb{E}\{\, I(\acute{\mathbf{h}}\odot\mathbf{x}+\mathbf{n};\acute{\mathbf{h}}\,|\,\mathbf{x})\,\} \tag{55}$$

$$\overset{(a)}{=} \inf_{\substack{P_{ij}\geqslant 0\\ \sum_{ij}P_{ij}=D_{\mathrm{T}}D_{\mathrm{W}}P}} \sum_{i=1}^{D_{\mathrm{T}}}\sum_{j=1}^{D_{\mathrm{W}}} \inf_{\substack{P_{\mathbf{x}_{ij}}\in\mathcal{P}_{\mathrm{f}}\\ \mathbb{E}\{\|\mathbf{x}_{ij}\|_2^2\}=\ell q P_{ij}}} \mathbb{E}\{\,\log(1+\upsilon_j^2(1-|\eta|^2)\|\mathbf{x}_{ij}\|_2^2)\,\} \tag{56}$$

$$\overset{(b)}{\geqslant} \inf_{\substack{0\leqslant P_{ij}\leqslant\sqrt{\beta}P\\ \sum_{ij}P_{ij}=D_{\mathrm{T}}D_{\mathrm{W}}P}} \sum_{i=1}^{D_{\mathrm{T}}}\sum_{j=1}^{D_{\mathrm{W}}} \inf_{\substack{P_{\mathbf{x}_{ij}}\\ \mathbb{E}\{\|\mathbf{x}_{ij}\|_2^2\}=\ell q P_{ij}\\ \mathbb{E}\{\|\mathbf{x}_{ij}\|_2^4\}\leqslant\ell^2\beta P^2}} \mathbb{E}\{\,\log(1+\upsilon_j^2(1-|\eta|^2)\|\mathbf{x}_{ij}\|_2^2)\,\} \tag{57}$$

$$\overset{(c)}{=} \inf_{\substack{0\leqslant P_{ij}\leqslant\sqrt{\beta}P\\ \sum_{ij}P_{ij}=D_{\mathrm{T}}D_{\mathrm{W}}P}} \sum_{i=1}^{D_{\mathrm{T}}}\sum_{j=1}^{D_{\mathrm{W}}} \frac{q^2P_{ij}^2}{\beta P^2}\log\left(1+\upsilon_j^2(1-|\eta|^2)\frac{\ell\beta P^2}{qP_{ij}}\right) \tag{58}$$

expressed in bits per symbol.

The result follows from (71) by assuming i.i.d. inputs.

Appendix E
Proof of Theorem 6

We write $\psi_{\mathrm{d}} = \hat{\mathrm{h}}\xi_{\mathrm{d}} + \zeta_{\mathrm{d}}$, where ζ_{d} is a noise that is uncorrelated with $\hat{\mathrm{h}}\xi_{\mathrm{d}}$ and

$$\mathbb{V}\{\zeta_{\mathrm{d}}\} = \sigma_{\tilde{\mathrm{h}}}^2\mathbb{V}\{\xi_{\mathrm{d}}\}+1 =: \sigma_{\zeta_{\mathrm{d}}}^2.$$

Rewrite mutual information $I(\xi_{\mathrm{d}};\psi_{\mathrm{d}}|\hat{\mathrm{h}})$ as

$$I(\xi_{\mathrm{d}};\psi_{\mathrm{d}}|\hat{\mathrm{h}}) = \mathbb{E}_{\hat{\mathrm{h}}\sim P_{\hat{\mathrm{h}}}}\{\,h(\xi_{\mathrm{d}})-h(\xi_{\mathrm{d}}|\psi_{\mathrm{d}},\hat{\mathrm{h}})\,\} \tag{72}$$

where $h(\xi_{\mathrm{d}}|\hat{\mathrm{h}}) = h(\xi_{\mathrm{d}})$ since ξ_{d} and $\hat{\mathrm{h}}$ are independent. Since the distribution of ξ_{d} conditioned on ψ_{d} and $\hat{\mathrm{h}}$ is difficult to find, we bound $h(\xi_{\mathrm{d}}|\psi_{\mathrm{d}},\hat{\mathrm{h}})$ with the maximum entropy over distributions with support $\mathcal{B}_r(0)$ and variance $\mathbb{V}\{\,\xi_{\mathrm{d}}\,|\,\psi_{\mathrm{d}},\hat{\mathrm{h}}\,\}$. Since (41) is a monotonically increasing function of the variance, any upper bound to the variance yields an upper bound on the differential entropy. It is a general fact [54] that

$$\mathbb{V}\{\,\xi_{\mathrm{d}}\,|\,\psi_{\mathrm{d}},\hat{\mathrm{h}}\,\} \leqslant \mathbb{E}\{\,|\xi_{\mathrm{d}}-\hat{\xi}_{\mathrm{d}}|^2\,|\,\hat{\mathrm{h}}\,\} =: \hat{\sigma}_{\mathrm{d}}^2 \tag{73}$$

for any estimation $\hat{\xi}_{\mathrm{d}}$ of ξ_{d}. Choosing the LMMSE yields (45). Therefore, $h(\xi_{\mathrm{d}}|\psi_{\mathrm{d}},\hat{\mathrm{h}})$ is upper bounded by the differential entropy of the distribution $\mathcal{CN}_r(0,\hat{\gamma})$:

$$h(\xi_{\mathrm{d}}|\psi_{\mathrm{d}},\hat{\mathrm{h}}) \leqslant \log\left(\pi\hat{\gamma}\,e^{\hat{\sigma}_{\mathrm{d}}^2/\hat{\gamma}}\,V_{r,\hat{\gamma}}\right). \tag{74}$$

Introducing (41) and (74) in (72) results in:

$$I(\xi_{\mathrm{d}};\psi_{\mathrm{d}}|\hat{\mathrm{h}}) \geqslant I_{\mathrm{LB}}(\xi_{\mathrm{d}};\psi_{\mathrm{d}}|\hat{\mathrm{h}}). \tag{75}$$

Acknowledgment

The authors wish to thank the Associate Editor and the anonymous reviewers for helpful suggestions. They would also like to thank W. Dai, Z. Liu, M. Mohammadkarimi, L. Ruan, F. Wang and T. Wang for careful reading of the manuscript.

References

[1] J. Pierce, "Ultimate performance of M-ary transmissions on fading channels," *IEEE Trans. Inf. Theory*, vol. 12, no. 1, pp. 2–5, Jan. 1966.

[2] A. J. Viterbi, "Performance of an M-ary orthogonal communication system using stationary stochastic signals," *IEEE Trans. Inf. Theory*, vol. 13, no. 3, pp. 414–422, July 1967.

[3] D. Cassioli, M. Z. Win, and A. F. Molisch, "The ultra-wide bandwidth indoor channel: From statistical model to simulations," *IEEE J. Sel. Areas Commun.*, vol. 20, no. 6, pp. 1247–1257, Aug. 2002.

[4] M. Z. Win and R. A. Scholtz, "Ultra-wide bandwidth time-hopping spread-spectrum impulse radio for wireless multiple-access communications," *IEEE Trans. Commun.*, vol. 48, no. 4, pp. 679–689, Apr. 2000.

[5] ——, "Impulse radio: how it works," *IEEE Commun. Lett.*, vol. 2, no. 2, pp. 36–38, Feb. 1998.

[6] M. Médard and R. G. Gallager, "Bandwidth scaling for fading multipath channels," *IEEE Trans. Inf. Theory*, vol. 48, no. 4, pp. 840–852, Apr. 2002.

[7] E. Telatar and D. Tse, "Capacity and mutual information of wideband multipath fading channels," *IEEE Trans. Inf. Theory*, vol. 46, no. 4, pp. 1384–1400, July 2000.

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

[9] M. Z. Win and R. A. Scholtz, "On the robustness of ultra-wide bandwidth signals in dense multipath environments," *IEEE Commun. Lett.*, vol. 2, no. 2, pp. 51–53, Feb. 1998.

[10] ——, "On the energy capture of ultra-wide bandwidth signals in dense multipath environments," *IEEE Commun. Lett.*, vol. 2, no. 9, pp. 245–247, Sept. 1998.

[11] ——, "Characterization of ultra-wide bandwidth wireless indoor communications channel: A communication-theoretic view," *IEEE J. Sel. Areas Commun.*, vol. 20, no. 9, pp. 1613–1627, Dec. 2002.

[12] D. Porrat, D. Tse, and S. Nacu, "Channel uncertainty in ultra-wideband communication systems," *IEEE Trans. Inf. Theory*, vol. 53, no. 1, pp. 194–208, Jan. 2007.

[13] M. Z. Win, G. Chrisikos, and N. R. Sollenberger, "Performance of Rake reception in dense multipath channels: Implications of spreading bandwidth and selection diversity order," *IEEE J. Sel. Areas Commun.*, vol. 18, no. 8, pp. 1516–1525, Aug. 2000.

[14] D. Cassioli, M. Z. Win, F. Vatalaro, and A. F. Molisch, "Low-complexity Rake receivers in ultra-wideband channels," *IEEE Trans. Wireless Commun.*, vol. 6, no. 4, pp. 1265–1275, Apr. 2007.

[15] W. C. Lau, M.-S. Alouini, and M. K. Simon, "Optimum spreading bandwidth for selective Rake reception over Rayleigh fading channels," *IEEE J. Sel. Areas Commun.*, vol. 19, no. 6, pp. 1080–1089, Jun. 2001.

[16] Y.-L. Chao and R. A. Scholtz, "Ultra-wideband transmitted reference systems," *IEEE Trans. Veh. Technol.*, vol. 54, no. 5, pp. 1556–1569, Sept. 2005.

[17] T. Q. S. Quek and M. Z. Win, "Analysis of UWB transmitted-reference communication systems in dense multipath channels," *IEEE J. Sel. Areas Commun.*, vol. 23, no. 9, pp. 1863–1874, Sept. 2005.

[18] T. Q. S. Quek, M. Z. Win, and D. Dardari, "Unified analysis of UWB transmitted-reference schemes in the presence of narrowband

interference," *IEEE Trans. Wireless Commun.*, vol. 6, no. 6, pp. 2126–2139, Jun. 2007.

[19] A. Rabbachin, T. Q. S. Quek, P. C. Pinto, I. Oppermann, and M. Z. Win, "Non-coherent UWB communication in the presence of multiple narrowband interferers," *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3365–3379, Nov. 2010.

[20] L.-L. Yang and L. Hanzo, "Serial acquisition of DS-CDMA signals in multipath fading mobile channels," *IEEE Trans. Veh. Technol.*, vol. 50, no. 2, pp. 617–628, Mar. 2001.

[21] W. Suwansantisuk, M. Z. Win, and L. A. Shepp, "On the performance of wide-bandwidth signal acquisition in dense multipath channels," *IEEE Trans. Veh. Technol.*, vol. 54, no. 5, pp. 1584–1594, Sept. 2005.

[22] W. Suwansantisuk and M. Z. Win, "Multipath aided rapid acquisition: Optimal search strategies," *IEEE Trans. Inf. Theory*, vol. 53, no. 1, pp. 174–193, Jan. 2007.

[23] ——, "Method and apparatus for signal searching," U.S. Patent 8,565,690, Oct. 22, 2013.

[24] S. Verdú, "Spectral efficiency in the wideband regime," *IEEE Trans. Inf. Theory*, vol. 48, no. 6, pp. 1319–1343, June 2002.

[25] L. Zheng, D. N. C. Tse, and M. Médard, "Channel coherence in the low-SNR regime," *IEEE Trans. Inf. Theory*, vol. 53, no. 3, pp. 976–997, Mar. 2007.

[26] S. Ray, M. Médard, and L. Zheng, "On noncoherent MIMO channels in the wideband regime: Capacity and reliability," *IEEE Trans. Inf. Theory*, vol. 53, no. 6, pp. 1983–2009, June 2007.

[27] R. G. Gallager, "Energy limited channels: Coding, multiaccess, and spread spectrum," MIT, Cambridge, MA, USA, Tech. Rep. LIDS-P-1714, Nov. 1987.

[28] R. G. Gallager and M. Médard, "Bandwidth scaling for fading channels," in *Proc. IEEE Int. Symp. Inf. Theory*, June 1997, p. 471.

[29] A. Lozano and D. Porrat, "Non-peaky signals in wideband fading channels: Achievable bit rates and optimal bandwidth," *IEEE Trans. Wireless Commun.*, vol. 11, no. 1, pp. 246–257, Jan. 2012.

[30] F. Boccardi, R. W. Heath, A. Lozano, T. L. Marzetta, and P. Popovski, "Five disruptive technology directions for 5G," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 74–80, Feb. 2014.

[31] M. R. Akdeniz, Y. Liu, M. K. Samimi, S. Sun, S. Rangan, T. S. Rappaport, and E. Erkip, "Millimeter wave channel modeling and cellular capacity evaluation," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1164–1179, June 2014.

[32] T. S. Rappaport, S. Sun, R. Mayzus, H. Zhao, Y. Azar, K. Wang, G. Wong, J. K. Schulz, M. Samimi, and F. Gutierrez, "Millimeter wave mobile communications for 5G cellular: It will work!" *IEEE Access*, vol. 1, pp. 335–349, May 2013.

[33] C. Gustafson, K. Haneda, S. Wyne, and F. Tufvesson, "On mm-wave multipath clustering and channel modeling," *IEEE Trans. Antennas Propag.*, vol. 62, no. 3, pp. 1445–1455, Mar. 2014.

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

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

[36] V. Raghavan, G. Hariharan, and A. M. Sayeed, "Capacity of sparse multipath channels in the ultra-wideband regime," *IEEE J. Sel. Topics Signal Process.*, vol. 1, no. 3, pp. 357–371, Oct. 2007.

[37] V. Subramanian and B. Hajek, "Broad-band fading channels: Signal burstiness and capacity," *IEEE Trans. Inf. Theory*, vol. 48, no. 4, pp. 809–827, Apr. 2002.

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

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

[40] V. Sethuraman and B. Hajek, "Capacity per unit energy of fading channels with a peak constraint," *IEEE Trans. Inf. Theory*, vol. 51, no. 9, pp. 3102–3120, Sept. 2005.

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

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

[43] J. Chen and V. V. Veeravalli, "Capacity results for block-stationary Gaussian fading channels with a peak power constraint," *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4498–4520, Dec. 2007.

[44] G. Taricco and M. Elia, "Capacity of fading channel with no side information," *Electron. Lett.*, vol. 33, no. 16, pp. 1368–1370, July 1997.

[45] I. C. Abou-Faycal, M. D. Trott, and S. Shamai, "The capacity of discrete-time memoryless Rayleigh-fading channels," *IEEE Trans. Inf. Theory*, vol. 47, no. 4, pp. 1290–1301, May 2001.

[46] N. Jindal and A. Lozano, "A unified treatment of optimum pilot overhead in multipath fading channels," *IEEE Trans. Commun.*, vol. 58, no. 10, pp. 2939–2948, Oct. 2010.

[47] L. H. Ozarow, S. Shamai, and A. D. Wyner, "Information theoretic considerations for cellular mobile radio," *IEEE Trans. Veh. Technol.*, vol. 43, no. 2, pp. 359–378, May 1994.

[48] M. Gursoy, H. Poor, and S. Verdú, "The noncoherent Rician fading channel–Part I: Structure of the capacity-achieving input," *IEEE Trans. Wireless Commun.*, vol. 4, no. 5, pp. 2193–2206, Sept. 2005.

[49] ——, "Noncoherent Rician fading channel–Part II: Spectral efficiency in the low-power regime," *IEEE Trans. Wireless Commun.*, vol. 4, no. 5, pp. 2207–2221, Sept. 2005.

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

[51] R. Kennedy, *Fading dispersive communication channels*. New York: Wiley-Interscience, 1969.

[52] I. Jacobs, "The asymptotic behavior of incoherent M-ary communication systems," *Proc. IEEE*, vol. 51, no. 1, pp. 251–252, 1963.

[53] V. Sethuraman, L. Wang, B. Hajek, and A. Lapidoth, "Low SNR capacity of noncoherent fading channels," *IEEE Trans. Inf. Theory*, vol. 55, no. 4, pp. 1555–1574, Apr. 2009.

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

[55] G. Durisi, V. Morgenshtern, and H. Bolcskei, "On the sensitivity of continuous-time noncoherent fading channel capacity," *IEEE Trans. Inf. Theory*, vol. 58, no. 10, pp. 6372–6391, Oct. 2012.

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

[57] V. Sethuraman and B. Hajek, "Low SNR capacity of fading channels with peak and average power constraints," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, July 2006, pp. 689–693.

[58] 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)*, Sept. 2005, pp. 515–519.

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

[60] G. Matz and F. Hlawatsch, "Time-varying communication channels: Fundamentals, recent developments, and open problems," in *Europ. Signal Process. Conf.*, Sept. 2006, pp. 1–5.

[61] G. Taubock, F. Hlawatsch, D. Eiwen, and H. Rauhut, "Compressive estimation of doubly selective channels in multicarrier systems: Leakage effects and sparsity-enhancing processing," *IEEE J. Sel. Topics Signal Process.*, vol. 4, no. 2, pp. 255–271, Apr. 2010.

[62] G. Taubock, M. Hampejs, P. Svac, G. Matz, F. Hlawatsch, and K. Grochenig, "Low-complexity ici/isi equalization in doubly dispersive multicarrier systems using a decision-feedback lsqr algorithm," *IEEE Trans. Signal Process.*, vol. 59, no. 5, pp. 2432–2436, May 2011.

[63] Recommendation ITU-R P.676-9, "Attenuation by atmospheric gases," 2012.

[64] A. M. Tulino, A. Lozano, and S. Verdú, "Capacity-achieving input covariance for single-user multi-antenna channels," *IEEE Trans. Wireless Commun.*, vol. 5, no. 3, pp. 662–671, Mar. 2006.

[65] A. F. Karr, "Extreme points of certain sets of probability measures, with applications." *Maths. Oper. Research*, vol. 8, no. 1, pp. 74–85, Feb. 1983.

Guido Carlo Ferrante (S'11–M'15) received a double Ph.D. degree in Electrical Engineering from Sapienza Università di Roma, Rome, Italy, and CentraleSupélec, Gif-sur-Yvette, France, in April 2015. Previously, he received both M.Sc. and B.Sc. degrees (*summa cum laude*) in Electrical Engineering from Sapienza Università di Roma, Rome, Italy.

He is currently a Postdoctoral Fellow in the Wireless Information and Network Sciences Laboratory at Massachusetts Institute of Technology (MIT), Cambridge, MA, USA. Prior to joining MIT, he was a Postdoctoral Fellow at Singapore University of Technology and Design (SUTD), Singapore. His research interests include application of communication theory, information theory and statistical inference to complex systems and networks.

Dr. Ferrante received the Italian National Telecommunications and Information Theory Group Award for Ph.D. Theses in the field of Communication Technologies (2015) and the SUTD-MIT Postdoctoral Fellowship (2015–2017).

Tony Q.S. Quek (S'98-M'08-SM'12) received the B.E. and M.E. degrees in Electrical and Electronics Engineering from Tokyo Institute of Technology, respectively. At MIT, he earned the Ph.D. in Electrical Engineering and Computer Science. Currently, he is a tenured Associate Professor with the Singapore University of Technology and Design (SUTD). He also serves as the Associate Head of ISTD Pillar and the Deputy Director of the SUTD-ZJU IDEA. His main research interests are the application of mathematical, optimization, and statistical theories to communication, networking, signal processing, and resource allocation problems. Specific current research topics include heterogeneous networks, wireless security, internet-of-things, and big data processing.

Dr. Quek has been actively involved in organizing and chairing sessions, and has served as a member of the Technical Program Committee as well as symposium chairs in a number of international conferences. He is serving as the Workshop Chair for IEEE Globecom in 2017, the Tutorial Chair for the IEEE ICCC in 2017, and the Special Session Chair for IEEE SPAWC in 2017. He is currently an elected member of IEEE Signal Processing Society SPCOM Technical Committee. He was an Executive Editorial Committee Member for the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, an Editor for the IEEE TRANSACTIONS ON COMMUNICATIONS, and an Editor for the IEEE WIRELESS COMMUNICATIONS LETTERS. He is a co-author of the book "Small Cell Networks: Deployment, PHY Techniques, and Resource Allocation" published by Cambridge University Press in 2013 and the book "Cloud Radio Access Networks: Principles, Technologies, and Applications" by Cambridge University Press in 2017.

Dr. Quek was honored with the 2008 Philip Yeo Prize for Outstanding Achievement in Research, the IEEE Globecom 2010 Best Paper Award, the 2012 IEEE William R. Bennett Prize, the IEEE SPAWC 2013 Best Student Paper Award, the IEEE WCSP 2014 Best Paper Award, the 2015 SUTD Outstanding Education Awards – Excellence in Research, the 2016 Thomson Reuters Highly Cited Researcher, and the 2016 IEEE Signal Processing Society Young Author Best Paper Award.

Moe Z. Win (S'85-M'87-SM'97-F'04) received both the Ph.D. in Electrical Engineering and the M.S. in Applied Mathematics as a Presidential Fellow at the University of Southern California (USC) in 1998. He received the M.S. in Electrical Engineering from USC in 1989 and the B.S. (*magna cum laude*) in Electrical Engineering from Texas A&M University in 1987.

He is a Professor at the Massachusetts Institute of Technology (MIT) and the founding director of the Wireless Information and Network Sciences Laboratory. Prior to joining MIT, he was with AT&T Research Laboratories for five years and with the Jet Propulsion Laboratory for seven years. His research encompasses fundamental theories, algorithm design, and experimentation for a broad range of real-world problems. His current research topics include network localization and navigation, network interference exploitation, intrinsic wireless secrecy, adaptive diversity techniques, and ultra-wideband systems.

Professor Win is an elected Fellow of the AAAS, the IEEE, and the IET, and was an IEEE Distinguished Lecturer. He was honored with two IEEE Technical Field Awards: the IEEE Kiyo Tomiyasu Award (2011) and the IEEE Eric E. Sumner Award (2006, jointly with R. A. Scholtz). Together with students and colleagues, his papers have received numerous awards, including the IEEE Communications Society's Stephen O. Rice Prize (2012), the IEEE Aerospace and Electronic Systems Society's M. Barry Carlton Award (2011), the IEEE Communications Society's Guglielmo Marconi Prize Paper Award (2008), and the IEEE Antennas and Propagation Society's Sergei A. Schelkunoff Transactions Prize Paper Award (2003). Highlights of his international scholarly initiatives are the Copernicus Fellowship (2011), the Royal Academy of Engineering Distinguished Visiting Fellowship (2009), and the Fulbright Fellowship (2004). Other recognitions include the IEEE Communications Society Edwin H. Armstrong Achievement Award (2016), the International Prize for Communications Cristoforo Colombo (2013), the *Laurea Honoris Causa* from the University of Ferrara (2008), the Technical Recognition Award of the IEEE ComSoc Radio Communications Committee (2008), and the U.S. Presidential Early Career Award for Scientists and Engineers (2004).

Dr. Win was an elected Member-at-Large on the IEEE Communications Society Board of Governors (2011–2013). He was the Chair (2005–2006) and Secretary (2003–2004) for the Radio Communications Committee of the IEEE Communications Society. Over the last decade, he has organized and chaired numerous international conferences. He is currently serving on the SIAM Diversity Advisory Committee and the advisory board of the IEEE COMMUNICATION LETTERS. He served as Editor-at-Large (2012–2015) for the IEEE WIRELESS COMMUNICATIONS LETTERS, as Editor (2006–2012) for the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, and as Area Editor (2003–2006) and Editor (1998–2006) for the IEEE TRANSACTIONS ON COMMUNICATIONS. He was Guest-Editor for the PROCEEDINGS OF THE IEEE (2009) and for the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS (2002).