

The SIMO Block Rayleigh Fading Channel Capacity Scaling with Number of Antennas, Bandwidth and Coherence Length

Felipe Gomez-Cuba¹

Abstract—This paper studies the capacity scaling of non-coherent Single-Input Multiple-Output (SIMO) independent and identically distributed (i.i.d.) Rayleigh block fading channels versus bandwidth (B), number of receive antennas (N) and coherence block length (L). In non-coherent channels (without Channel State Information –CSI) capacity scales as $\Theta(\min(B, \sqrt{NL}, N))$. This is achievable using Pilot-Assisted signaling. Energy Modulation signaling rate scales as $\Theta(\min(B, \sqrt{N}))$. If L is fixed while B and N grow, the two expressions grow equally and Energy Modulation achieves the capacity scaling. However, Energy Modulation rate does not scale as the capacity with the variable L . The coherent channel capacity with a priori CSI, in turn, scales as $\Theta(\min(B, N))$. The coherent channel capacity scaling can be fully achieved in non-coherent channels when $L \geq \Theta(N)$. In summary, the channel coherence block length plays a pivotal role in modulation selection and the capacity gap between coherent and non-coherent channels. Pilot-Assisted signaling outperforms Energy Modulation’s rate scaling versus coherence block length. Only in high mobility scenarios where L is much smaller than the number of antennas ($L \ll \Theta(\sqrt{N})$), Energy Modulation is effective in non-coherent channels.

Index Terms—Massive MIMO, wideband communications, non-coherent channel, energy modulation

I. INTRODUCTION

Internet traffic demand is expected to continue to grow exponentially [1]. Technologies such as the recent 5G cellular [2] and WiFi6 Wireless Local Area Network [3], [4] standards are upgraded regularly. The increase of the transmission bandwidth and number of antennas in each new standard generation are major factors in the increase of capacity of wireless systems [5], [6]. In addition, the new standards support “flexible waveforms” that can dynamically modify the transmit bandwidth and number of antenna ports used by different users. As more spectrum and antennas are added to wireless standards, efficient channel estimation and the impact of channel state information (CSI) uncertainty has received a lot of attention. On the one hand, CSI limitations are one key factor in the performance of coherent receivers in *massive MIMO* [6]. On the other hand, recent results for the wideband channel have sparked new interest in non-coherent signaling techniques [7]. Specifically, there has been significant recent literature proposing practical modulation schemes based on non-coherent Energy Detection (see [8]

and references therein). Overall, as wireless standards evolve, it is very important to understand the relation between the waveform dimensions (specifically bandwidth and number of antenna ports), the time-varying channel uncertainty, and the engineering choice between Pilot Assisted (PA) or non-coherent signaling schemes.

From an Information Theoretic point of view, the ultimate impact of channel uncertainty can be revealed by comparing the capacity of the *coherent channel*, with the assumption that CSI is a priori revealed to the receiver; and the *non-coherent channel*, in which CSI is not a priori available. On the transmitter side, transmitter CSI or feedback may also be assumed or not. We leave the topic of transmitter CSI out of the scope of this paper, in which we focus on comparing the capacity with and without receiver CSI, without feedback in both cases. In the non-coherent channel, it is possible to either use pilot sequences to estimate the channel, or to employ non-coherent signaling schemes. Thus, note that the non-coherent channel capacity also upper bounds the maximum rate of coherent receivers that rely on channel estimation (the name “non-coherent channel” is a bit counter-intuitive in this sense). While the coherent channel capacity is well-known [9], the non-coherent channel capacity is not yet fully characterized. Important results have obtained the Degrees of Freedom (DoF) and Diversity-Multiplexing Trade-off (DMT) in the high-SNR regime [10]–[13], the wideband capacity in the limit as bandwidth goes to infinity [7], [14]–[21], some properties of the probability density of the optimal channel input [16], [22]–[27], and optimal pilot strategies to maximize the “achievable rate” of specific PA receivers [28]–[30].

In this paper we compare the capacity *scaling* for coherent and non-coherent wideband Single-Input Multiple-Output (SIMO) block Rayleigh fading channels. Within the scope of achievable rates in a non-coherent channel, we also compare two signaling strategies: channel estimation with pilots vs energy detection. We focus on the single transmit antenna case for better comparison with closely related results in [27], detailed in the next paragraphs. The assumption of a wideband block fading model represents a reasonable abstraction for some modern wireless systems that use multi-carrier modulations and time-frequency-slotted frame allocations [2]–[4]. With the aim to inform the selection of waveform parameters, spectrum policies and antenna array dimensions, our analysis studies the effect on capacity growth of the following three major variables:

1) **The Channel Coherence Time:** refers generally to

¹AtlanTTic, Universidade de Vigo, Spain. Email: gomezcuba@gts.uvigo.es

the interval of time during which the random channel displays correlation. This concept is subject to different definitions under different channel models. In i.i.d. block-fading models, we define the **channel coherence length** as the length of time, in channel uses, that the channel remains constant. The block-fading channel coherence length has been shown to affect the non-coherent channel capacity-achieving input distribution [22], [23], DMT [13], and optimal pilot schemes [28]. The intuition is that, if the channel remains constant for a longer time, the relative size of the pilot overhead compared to the total data codeword length decreases. Thus, the non-coherent channel capacity is expected to converge to the coherent channel capacity as the channel coherence length increases. This holds for all existing results such as the optimal input [22] and DMT [13].

- 2) **The Bandwidth:** allows to increase the discrete time symbol rate, but it also decreases the Signal to Noise Ratio (SNR) when the thermal noise has constant power spectral density. The wideband non-coherent channel capacity is closely related to the low-SNR asymptotic analysis of capacity [15]–[19]. Médard and Gallager [7] showed that “overspreading” occurs when the channel input has finite fourth moment and bandwidth is too large, causing the achievable rate in the non-coherent channel to begin decreasing, rather than increasing monotonically as in the coherent case, if bandwidth is increased excessively. For Rayleigh fading, the “critical bandwidth” can be explicitly calculated [20]. The use of “peaky” input signals with an infinite fourth moment has been proposed to overcome this limitation [16]–[18]. Nevertheless, it can be shown that peaky signaling does not actually remove the critical bandwidth limit, but rather moves it to the time domain under a constant “critical bandwidth occupancy” limitation [21], [26]. The bandwidth overspreading threshold has been shown to grow with the square root of the coherence block length [18], [20], [21], showing that both parameters are closely related and that also in this sense the non-coherent channel capacity converges to the coherent channel capacity when the coherence length increases. The recent popularity of mmWave frequency bands increases interest in communications with very large bandwidth and antenna arrays [31], [32]. Although in these bands the assumption of rich scattering with a Rayleigh fading distribution may not be applicable, concern about these phenomena in sparse multipath mmWave channels is justified, as overspreading occurs in other channel models, including sparse multipath channels [14]. We remark that the results in our paper apply to the i.i.d. Rayleigh channel, and while mmWave channels are subject to the problem of overspreading, engineers must be cautious about the channel model difference before applying our conclusions in a mmWave context. A discussion of literature on multiple mmWave and ultra-wideband channel modeling frameworks and their relation to our result is provided in Appendix A.
- 3) **The Number of Antennas:** has been shown to increase the DMT of the channel [10]–[13], enabling dramatic

increases in the array gain or spatial multiplexing. In recent years, *Massive MIMO* proposed the use of a large number of antennas [6], [33]–[35]. Massive MIMO enables a very large array combining gain or diversity gain, mitigating pathloss and fading. Moreover the asymptotic properties of large matrices permit to implement efficient transceivers [36]. The DMT and optimal input distribution results in [12], [22] imply that, assuming there is no feedback, increasing the number of transmit antennas above the channel coherence length does not increase the non-coherent channel capacity. For this reason, analyses of the capacity scaling assuming a fixed coherence length have to focus on studying a large number of **receive** antennas. Chowdhury *et al* [24] showed that, as the number of receive antennas goes to infinity, an Energy Modulation (EM) constellation in a frequency-flat non-coherent SIMO channel can achieve the same capacity scaling as a coherent channel. This result was extended to the frequency selective case with a large bandwidth in [27]. Unlike in this paper, the scaling result in [27] considered scaling for both the bandwidth and the number of antennas, but assumed a fixed coherence length. This means that the result in [27] side-stepped one key parameter in the study of the capacity gap between coherent and non-coherent channels. Note that in our paper we assume that the coherence block length scales as well, and thus we could potentially support a scaling number of transmit antennas too. However, we adopt the SIMO channel model to provide a much more clear comparison with [27], as well as for the sake of clarity of the model and problem tractability. We leave the extension of our result to MIMO for future work.

Most Massive MIMO analyses assume a fixed bandwidth [6], [12], [13], [24] and most wideband analyses assume a fixed number of antennas [14], [15], [18]–[21]. However, in practice these two parameters, as well as the coherence length, display a non-trivial interplay between them. The DMT is related to high-SNR capacity analyses, whereas wideband capacity calculations focus on low-SNR regimes, making the combination of results from both families of literature challenging. The analysis in [27] studied capacity scaling when *both* the bandwidth and the number of antennas grow jointly with a fixed coherence length [27]. In brief words, the main principle in [27] and our model is a two-regime rate analysis that combines the DoF-limited narrowband and the power-limited wideband rate scaling: if the bandwidth is below a certain threshold, capacity scaling is in a high-SNR-like regime in which rate scales with the symbol rate and the DoF of the channel. Conversely, when the bandwidth exceeds the overspreading threshold, capacity scaling adopts a low-SNR-like regime in which rate scales with the total energy arriving at the receiver (possibly featuring the receive array gain). Moreover, this bandwidth threshold itself changes with the number of antennas and coherence block length. The result in [27] showed that, when the coherence length is fixed, the non-coherent channel capacity cannot scale faster than the minimum between the bandwidth and the square root of the

number of receive antennas, failing to match the linear scaling with the number of receive antennas of the coherent channel capacity in the wideband regime [27]. Moreover, [27] shows that EM achieves the joint capacity scaling in the non-coherent channel. As the analysis in [27] assumes a constant coherence length, the main goal of our paper is to complete the model by fully taking into account the channel coherence length. This will show that changes in the channel coherence length play a very important role in capacity scaling that was not captured by the result of [27].

In this paper we characterize the capacity of coherent and non-coherent channels as a **three dimensional** function of bandwidth, number of receive antennas, and channel coherence length. To characterize the capacity scaling, we consider that these three parameters grow to infinity jointly. We revise the achievable rate of the EM scheme proposed in [27], and introduce an alternative PA signaling scheme that outperforms EM in terms of achievable rate scaling. In consequence, our result shows that EM is no longer generally optimal, and depending on the channel coherence length, PA schemes can provide a greater rate. In addition, we compute an upper bound to non-coherent capacity scaling, which shows that our PA scheme always scales optimally, i.e. it can always achieve the capacity scaling, whereas EM cannot.

Our main result shows that, when the bandwidth is smaller than a certain threshold, the coherent channel and the non-coherent channel can achieve the same capacity scaling. In general, this is the “high-SNR-like” regime in which capacity grows linearly with the “pre-log” term, i.e. in our case the bandwidth. The critical bandwidth threshold scales with the minimum between *i*) the square root of the product of the number of antennas and the coherence length, or *ii*) just the number of receive antennas. When the bandwidth is greater than the critical bandwidth threshold, the non-coherent channel enters the “low-SNR-like” regime in which the capacity grows linearly with the received energy, i.e. in our case with the minimum between the square root of the product of the number of antennas and the coherence length, or the number of receive antennas. Therefore, for large bandwidths, when the channel coherence length is large in comparison to the number of antennas, in a scaling sense, the conditions of the converse proof in [27] are no longer applicable and EM signaling cannot achieve the non-coherent channel capacity scaling. Moreover, when the channel coherence length scales faster than the number of antennas, the non-coherent wideband massive SIMO channel achieves the same capacity scaling as the coherent channel, as should be expected, and PA channel estimation achieves the capacity scaling in all regimes.

If we assume a fixed coherence length in our result, the critical bandwidth threshold scales with the square root number of antennas and the non-coherent channel capacity scaling reproduces the result of [27]. Moreover, the EM scheme of [27] can achieve the non-coherent capacity in this particular case. However, our result generalizes [27] showing that, as the coherence length increases, the non-coherent capacity scaling approaches that of the coherent channel, PA schemes can achieve this rate scaling, and EM schemes remain limited to the same rate scaling as in [27], failing to keep up with the

capacity.

The rest of this paper is structured as follows. Section II describes the system model. Section III describes a selection of prior known results relevant to our analysis. Section IV contains the main results of our analysis. Section V presents the capacity scaling upper bound that proves the converse of the main result. Section VI presents the PA scheme that achieves the capacity scaling result. Section VII presents our revision of the EM encoding scheme. Section VIII demonstrates the main results in simulations. And finally, Section IX provides the conclusions.

A. Notation

Lowercase bold letters \mathbf{a} denote vectors and uppercase bold letters denote matrices \mathbf{A} . We use \mathbf{A}^* , \mathbf{A}^T and \mathbf{A}^H to denote the conjugate, transpose and Hermitian matrices of \mathbf{A} , respectively. The Kroënecker product between two matrices \mathbf{A} and \mathbf{B} is denoted as $\mathbf{A} \otimes \mathbf{B}$. We use a_i to denote the i^{th} element of vector \mathbf{a} and A_{ij} for the i, j -th element in a matrix \mathbf{A} . $\text{vec}(\mathbf{A})$ denotes the “vector stacking” of matrix \mathbf{A} , defined as a vector containing all the coefficients of \mathbf{A} in column-first order, i.e. the (i, j) -th coefficient of an N -row matrix \mathbf{A} is the $i + (j - 1)N$ -th element of $\text{vec}(\mathbf{A})$. We use calligraphic letters to denote sets, and $|\mathcal{A}|$ is the cardinality of set \mathcal{A} . We use $\{\mathbf{A}[i]\}_{i=n}^m$ to represent a set that contains the collection of matrices $\mathbf{A}[i]$ for i from n to m . We use $f(n) = \Theta(g(n))$ to denote that there exist constants $c_1, c_2, n_0 > 0$ such that for all $n > n_0$, $c_1 < \frac{f(n)}{g(n)} < c_2$ [37].

II. SYSTEM MODEL

We assume a frequency selective, block fading, i.i.d. Rayleigh SIMO wideband channel with a single-antenna transmitter and N antennas at the receiver. The transmitted signal bandwidth is divided in B independent subcarriers separated by a bandwidth Δf each, that experience independent frequency-flat Rayleigh distributed channel coefficients. We assume a block-fading channel in which, for each subcarrier, the channel coefficient remains the same for L consecutive symbols, and takes a new i.i.d. value in each *coherence slot*. We remark that the symbol period or channel use time in seconds is $\frac{1}{\Delta f}$, the channel coherence time in seconds is $\frac{L}{\Delta f}$, and Δf may be interpreted as the “coherence bandwidth” as well as the subcarrier separation. Our channel model can be represented as a resource grid depicted in Fig. 1. Among others, this model is representative of contemporary multicarrier technologies, in which the OFDM symbol duration is shorter than the channel coherence time and pilots are transmitted once in each frame of several consecutive OFDM symbols.

In each receive antenna n the received signal coefficient on subcarrier b at instant ℓ is given by

$$\begin{aligned} y_{n,\ell}[b] &= h_n[b]x_\ell[b] + z_{n,\ell}[b], \\ b &\in \{0 \dots B - 1\}, \\ n &\in \{0 \dots N - 1\}, \\ \ell &\in \{0 \dots L - 1\}. \end{aligned} \tag{1}$$

where the fading is normalized and independent for each n and b , $h_n[b] \sim \mathcal{CN}(0, 1)$. The Additive White Gaussian Noise

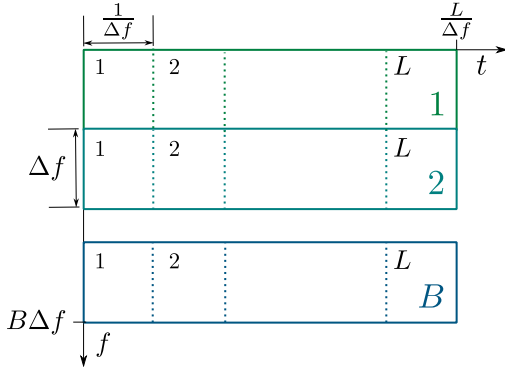


Fig. 1: Block-fading frequency-selective channel output with bandwidth $B\Delta f$, coherence length L and codeword time $L/\Delta f$.

(AWGN) is also normalized as $z_{n,\ell}[b] \sim \mathcal{CN}(0, 1)$, and the input average power constraint is

$$\mathbb{E} \left[\frac{1}{L} \sum_{b=1}^B \sum_{\ell=0}^{L-1} |x_{\ell}[b]|^2 \right] \leq P.$$

Here, we use the single letter P to represent a normalized power limitation. In practice P is computed based on the actual transmitter power P_T , the channel mean gain G , the noise power spectral density N_o and subcarrier bandwidth Δf as $P = \frac{P_T G}{\Delta f N_o}$. Thus, $\frac{P}{B} = \frac{P_T G}{B \Delta f N_o}$ is the average SNR in each receive antenna.

Since channel inputs and outputs are independent for each frequency bin b , we can write the channel as a collection of B frequency-flat SIMO subchannels

$$\mathbf{Y}[b] = \mathbf{h}[b] \mathbf{x}^H[b] + \mathbf{Z}[b], \forall b \in \{0 \dots B-1\}. \quad (2)$$

where $\mathbf{Y}[b], \mathbf{Z}[b] \in \mathbb{C}^{N,L}$, $\mathbf{h}[b] \in \mathbb{C}^N$, and $\mathbf{x}[b] \in \mathbb{C}^L$ with some probability density function (p.d.f.) $p(\{\mathbf{x}[b]\}_{b=0}^{B-1})$.

Even though we study the limit $L \rightarrow \infty$, we assume the ergodic mutual information is achieved by using the channel for a sufficiently long period of time of length $\gg L/B$, and we only apply constraints on the average power of the distributions. This permits to encode information over a large number of realizations of the block fading process and achieving the ergodic capacities defined as follows:

Definition 1. The *coherent ergodic capacity* of the block fading Rayleigh channel as a function of N , B and L is

$$C_c(N, B, L) = \sup_{p(\{\mathbf{x}[b]\}_{b=0}^{B-1})} \frac{\mathbb{E}_{\{\mathbf{h}[b]\}_{b=0}^{B-1}} [\mathbb{I}(\{\mathbf{x}[b]\}_{b=0}^{B-1}; \{\mathbf{Y}[b]\}_{b=0}^{B-1} | \{\mathbf{h}[b]\}_{b=0}^{B-1})]}{L/\Delta f} \quad (3)$$

Definition 2. The *non-coherent ergodic capacity* of the block fading Rayleigh channel as a function of N , B and L is

$$C_n(N, B, L) = \sup_{p(\{\mathbf{x}[b]\}_{b=0}^{B-1})} \frac{\mathbb{E}_{\{\mathbf{h}[b]\}_{b=0}^{B-1}} [\mathbb{I}(\{\mathbf{x}[b]\}_{b=0}^{B-1}; \{\mathbf{Y}[b]\}_{b=0}^{B-1})]}{L/\Delta f} \quad (4)$$

Here, the division by $L/\Delta f$ provides the unit conversion to express capacity in units of bits per second rather than bits

per block. In this paper, we will study the asymptotic scaling when B , N and L all grow to infinity. Therefore, the division by the block length will play an important role in the scaling result. Generally, the capacity definition must consider joint encoding in all B subcarriers, so the supremum is over the joint distribution $p(\{\mathbf{x}[b]\}_{b=0}^{B-1})$. However, in the case with no feedback and no CSIT we will show that, aside from power allocation, independent encoding in each subcarrier is sufficient.

In this paper we characterize the scaling of the functions $C_c(N, B, L)$ and $C_n(N, B, L)$ as the parameters N , B , and L grow. Since this is a multi-dimensional asymptotic, the point $(N, B, L) \rightarrow (\infty, \infty, \infty)$ can be approached from infinite “directions”. We specify the relative growth speed of N , B , and L as they jointly increase with scaling exponents.

Definition 3. We define the *bandwidth exponent* as

$$\epsilon = \lim_{B, N \rightarrow \infty} \frac{\log(B)}{\log(N)}, \quad (5)$$

for a non-negative constant ϵ . This may equivalently be expressed as $B = \Theta(N^\epsilon)$.

Definition 4. The *block length exponent* is defined as

$$\tau = \lim_{L, N \rightarrow \infty} \frac{\log(L)}{\log(N)}, \quad (6)$$

for a non-negative constant τ , or equivalently $L = \Theta(N^\tau)$.

Using these definitions, we can compare the scaling of $C_c(N, B, L)$ and $C_n(N, B, L)$ as a function of N , ϵ and τ .

III. KNOWN RESULTS

The comparison of the coherent and non-coherent channel capacities in [27] showed that with a fixed L , i.e. $\tau = 0$, EM achieves the capacity scaling of the non-coherent channel. However, for $\epsilon > \frac{1}{2}$ overspreading occurs, and the non-coherent channel capacity can only scale at most as \sqrt{N} , whereas the coherent channel capacity can reach linear scaling with N if $\epsilon \geq 1$.

Lemma 1. The capacity scaling of a Coherent Block Fading Rayleigh i.i.d. wideband SIMO channel with N receive antennas, bandwidth $B = \Theta(N^\epsilon)$ and $L = \Theta(N^\tau)$ is

$$C_c(B, N, L) = \Theta(N^{\min(\epsilon, 1)})$$

Proof. Proof of [27, Lemma 1] holds also for $\tau > 0$. \square

Lemma 2. The capacity scaling of a Non-Coherent Block Fading Rayleigh i.i.d. wideband SIMO channel with N receive antennas, bandwidth $B = \Theta(N^\epsilon)$ and constant L ($\tau = 0$) is

$$C_n(B, N, L) = \Theta(N^{\min(\epsilon, \frac{1}{2})}) \quad (7)$$

Proof. The full proof is given in [27]. \square

The interest in EM techniques can be associated with the seminal work by Marzetta and Hochwald [22], which shows that the capacity-achieving input in the non-coherent channel is a product of an EM of unspecified density function, times an Isotropically Distributed Unitary Vector (IDUV) distribution.

Lemma 3. [22, Theorem 2] *The optimal distribution for a frequency-flat Rayleigh block fading $1 \times N$ SIMO (sub)channel is $\mathbf{x}[b] = \sqrt{a[b]}\mathbf{u}[b]$ where $a[b]$ and $\mathbf{u}[b]$ are independent, $a[b]$ is a non-negative real number distribution and $\mathbf{u}[b]$ is IDUV.*

Remark 1. *Comparing Lemmas 2 and 3, it would appear that in non-coherent channels of large dimensions, information is mostly carried by the EM part of the input, $a[b]$. In our work we will show that this interpretation may be an artifact caused by the assumption of a constant block length L in the previous works. Indeed, this is the length of the IDUV $\mathbf{u}[b]$, and when L is relatively large compared to B and N the role of the “input shape”, $\mathbf{u}[b]$, in capacity scaling is not negligible.*

The achievable scheme for Lemma 2 is the EM technique proposed in [24], [27]. EM assumes that $M \leq B$ sub-channels are used, each with an average transmitted power $E[|\mathbf{x}[b]|^2] = LP/M$. The transmitter uses a real positive energy constellation to modulate the energy of the input, selecting a symbol from the set

$$a[b] \in \mathcal{C}_{\mathcal{M}} = \left\{ 0, 2d, 4d, \dots, \frac{2}{M} \right\},$$

and creating a transmitted signal by repeating the symbol L times

$$\mathbf{x}[b] = \sqrt{a[b]}\mathbf{1}_L.$$

To perform detection, the receiver computes the statistic

$$v[b] = \sum_{n=0}^{N-1} \left| \frac{1}{L} \sum_{\ell=0}^{L-1} y_{n,\ell}[b] \right|^2 \sim (a[b]^2 + \frac{1}{L})\chi^2(2N)$$

and decides the symbol $\hat{a}[b]$ that is nearest to $v[b] - \frac{1}{L}$. As the transmitted symbol is repeated L times, it is trivial to adjust the analysis of [27, Theorem 2] to the case with non-constant $L = \Theta(N^\tau)$ as follows:

Lemma 4. *In EM with $M = \Theta(N^{\min(\epsilon, \frac{1}{2} + \tau)})$ and $d = \Theta(N^{-t})$ satisfying $\epsilon < t < \frac{1}{2} + \tau$, the probability of error vanishes as $N \rightarrow \infty$ and rate scales as $\Theta(N^{\min(\epsilon - \tau, \frac{1}{2})})$.*

Proof. Since one data symbol is repeated L times, the error probability upper bound computed in [27, theorem 2] vanishes for any inter-symbol distance that satisfies $t < \frac{1}{2} + \tau$ (this was $t < \frac{1}{2}$ with $\tau = 0$ in the original proof). The constellation size is $\frac{N^t}{M}$, there are M subchannels and a $1/L$ repetition code, so the rate is $\frac{M}{L} \log_2(\frac{N^t}{M})$. For any $\epsilon > 0$ a rate scaling of $\Theta(N^{\min(\epsilon - \tau, \frac{1}{2})})$ is achievable, by choosing $M = \Theta(N^{\min(\epsilon, \tau + \frac{1}{2})})$ subchannels. \square

Remark 2. *For $\tau = 0$, the EM scheme achieves the non-coherent channel capacity scaling (Lemma 2), and if we also assume $\epsilon < \frac{1}{2}$, the coherent channel capacity scaling is achieved (Lemma 1).*

The main motivator to introduce an extension of Lemma 2 featuring a variable coherence block length is the exact calculation of the critical bandwidth for Rayleigh channels with fixed values of B, N, L , introduced in [20], [21], reproduced by the following lemma

Lemma 5. *In a Non-Coherent Block Fading Rayleigh i.i.d. wideband SIMO channel with constant N, B , and L , there is a “critical bandwidth” number B_{crit} , contained in the interval*

$$2P\sqrt{\frac{\log \pi}{1+N} \frac{L}{\log L}} \leq B_{crit} \leq 2P\sqrt{(1+N) \log \pi \frac{L}{\log L}} \quad (8)$$

and such that the achievable rate with non-peaky signaling is maximum when $B = B_{crit}$ and decreases for $B > B_{crit}$ (overspreading).

Proof. Lemma 4 of [21] reproduced with our notation. \square

By observing the right hand side of (8) in Lemma 5, in comparison to the two capacity scaling regimes of (7) in Lemma 2, it is clear that both results indicate that the maximum bandwidth scaling is proportional to $\Theta(\sqrt{N})$ when L is fixed. By observing the role of L in (8), it is therefore natural to expect that, for non-constant L that increases as $L = \Theta(N^\tau)$, an extension of Lemma 2 in which the critical bandwidth is proportional to $\Theta(\sqrt{NL})$ should exist. Although this is intuitive, the proof will be the main result of our paper. Significant mathematical differences versus [20], [21], [27] are required, described in Appendix C.

IV. MAIN RESULT

A. Capacity Scaling

The main result in our paper introduces the effect of L into the non-coherent capacity, fully characterizing its scaling.

Theorem 1. *The capacity of a Non-Coherent Block Fading Rayleigh i.i.d. wideband SIMO channel with N receive antennas, bandwidth $B = \Theta(N^\epsilon)$ and block length $L = \Theta(N^\tau)$ scales as*

$$C_n(B, N, L) = \Theta(N^{\min(\epsilon, \frac{1+\tau}{2}, 1)})$$

Proof. This rate scaling is achievable using the PA scheme described in Sec. VI. Moreover, the same exponents are observed in the upper bound in Sec. V. Therefore, the capacity scaling exponent is fully characterized. \square

Remark 3. *When $\tau \geq 1$, the PA scheme can fully achieve the same capacity scaling as the coherent channel for any value of ϵ . For $\tau < 1$, the capacity scaling of the coherent channel is only matched by the non-coherent channel if its bandwidth is not too large, i.e. $\epsilon < \frac{1+\tau}{2}$.*

Remark 4. *The comparison of Theorem 1 with Lemma 4 indicates that EM does not achieve the non-coherent capacity scaling for any value of ϵ if $\tau > 0$. Even if both the EM and PA schemes can admit bandwidth exponents up to $\epsilon \leq \frac{1}{2} + \tau$ and $\epsilon \leq \frac{1+\tau}{2}$, respectively, the EM scheme only achieves the rate exponent $\min(\epsilon - \tau, \frac{1}{2})$. Therefore, when $\tau > 0$ the EM scheme displays poor spectral efficiency and does not match the rate scaling of the PA scheme, nor the non-coherent channel capacity.*

The interpretation of the main result is more intuitive by imagining the growth of rates in a channel with a very large but finite number of receive antennas (N), with a fixed coherence length (L), and with a very large bandwidth ($B \rightarrow \infty$) so that

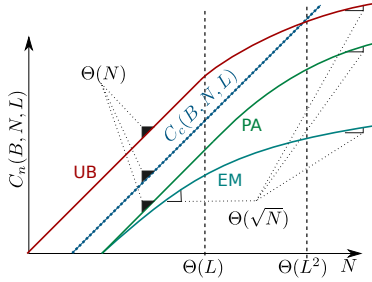


Fig. 2: Scaling of rate vs N , with fixed L and very large B

rate is always in the wideband “power limited” regime with overspreading. In essence, the main result warns us that the term “scaling with N ” must be always considered *in comparison to L* . We illustrate in Fig. 2 the rate versus the number of receive antennas for the special case of our result with $\epsilon \geq 1$. In this special case the coherent capacity scales as $\Theta(N)$, the non-coherent capacity upper bound and the achievable rate with PA both scale as $\Theta(\min(\sqrt{NL}, N))$, and the EM rate scales as $\Theta(\sqrt{N})$. We can thus see three regions for the non-coherent capacity scaling: First, the rates grow nearly linearly with N when $N \ll L$, nearly matching the coherent channel capacity scaling. Second, when $L \ll N \ll L^2$, the non-coherent channel capacity cannot continue to match the coherent channel, but the PA rate and the upper bound grow more than \sqrt{N} . In this regime PA achieves the non-coherent capacity scaling and outperforms EM. Finally, when $L^2 \ll N$, both the upper bound and PA achievable rate grow as \sqrt{N} , and the EM scheme achieves capacity scaling too. The prior scaling analysis assuming L was a constant in [27] reflected only this third regime. Nevertheless, in many channels L is very large and the first two regimes have practical importance. For example, reasonably practical values of the coherence time $\frac{L}{\Delta f} = 1ms$ and delay spread $\frac{1}{\Delta f} = 1\mu s$, would lead to $L = 10^3$, which means that the third regime where EM is optimal is not achieved until $N > 10^6$. On the other hand, in a practical scenario with very high speed mobility where $L = 2$ the third regime would appear quite soon, at $N \geq 4$. For $L = 1$, the first two regimes would not exist at all.

B. The Capacity with Only Energy Detection

The energy-shape separation of the input in Lemma 3, and the ability of EM to achieve the non-coherent channel capacity scaling when $\tau = 0$ in Lemma 2, have arisen interest in encoding schemes that rely solely on encoding information using the instantaneous energy per transmitted block ($a[b]$ in lemma 3).

Definition 5. We define an Only Energy Detection (OED) input as any input distribution that only encodes information in the input energy variable $a[b] \triangleq \|\mathbf{x}[b]\|^2$ according to the input decomposition $\mathbf{x}[b] = \sqrt{a[b]}\mathbf{u}[b]$ proposed by Lemma 3. This definition comprehends both EM and also other modulations that satisfy this pattern, such as for example a Frequency Shift Keying (FSK) that holds the same frequency in the entire interval $\ell \in \{0 \dots L - 1\}$

As an intermediate step of the converse proof in Theorem 1, we have shown that no OED schemes can achieve a rate scaling greater than that of the EM scheme in Lemma 4.

Theorem 2. In a Non-Coherent Block Fading Rayleigh i.i.d. wideband SIMO channel (2), the OED Capacity defined as

$$C_E(B, N, L) \triangleq \sup_{p(\{a[b]\}_{b=0}^{B-1})} \frac{\mathbb{E}_{\{\mathbf{h}[b]\}_{b=0}^{B-1}} [\mathbb{I}(\{\{a[b]\}_{b=0}^{B-1}; \{\mathbf{Y}[b]\}_{b=0}^{B-1})]}{L/\Delta f},$$

for N receive antennas, bandwidth $B = \Theta(n^\epsilon)$ and $L = \Theta(N^\tau)$, scales as

$$C_E(B, N, L) = \Theta(N^{\min(\epsilon-\tau, \frac{1}{2})}) \quad (9)$$

Proof. The upper bound to the scaling of $C_E(B, N, L)$ is given in Appendix B. The achievable rate is in Lemma 4. \square

Remark 5. By the chain rule, it is clear that $C_E(B, N, L) \leq C_n(B, N, L)$. By comparing Theorems 1 and 2, they display equal scaling when $\tau = 0$. Our PA achievable scheme has shown that the scaling factor \sqrt{N} can be exceeded for $\tau > 0$. However, Theorem 2 is stronger than Lemma 2 in the sense that it proves that EM alone cannot exceed the square root limitation **even if we allow** $\tau > 0$, and that no other OED schemes can either. The capacity scaling in Theorem 1 for $\tau > 0$ cannot be achieved by simply adopting a different OED modulation other than EM in Lemma 4. The rate scaling differences display a fundamental gap between OED Capacity and Non-Coherent Capacity scaling.

A final question arises if non-coherent energy modulations can be improved at all. To this, we can answer favorably, but to achieve this we must not associate the concept of “energy modulation” with the variable $a[b]$ in Lemma 3. Indeed, in Section VII we propose a Fast Energy Modulation (FEM) scheme that transmits an energy constellation of L independent i.i.d. symbols in the variable $\mathbf{x}[b]$. This allows to significantly improve the spectral efficiency, however a rate scaling limitation proportional to the square root of the number of antennas still arises.

Lemma 6. The FEM scheme achieves a vanishing error probability as $N \rightarrow \infty$ with rate

$$R_{FEM} = \Theta(N^{\min(\epsilon, \frac{1}{2})}) \quad (10)$$

Proof. The FEM scheme is analyzed in Section VII. \square

Remark 6. Theorem 2 is valid for any OED scheme according to Definition 5, following from the Marzetta-Hochwald energy-shape separation of the input in Lemma 3. The rate upper bound converse in Theorem 2 does not apply to any non-coherent encoding scheme that does not meet Definition 5. As in Lemma 6, examples of non-OED schemes can be proposed for other non-coherent modulation techniques. For example, Theorem 2 would not apply to an FSK scheme that allows to change the frequency in each instant ℓ .

We give a visual summary of all of our results in Fig. 3. For $\tau < 1$, the capacity scaling of a coherent channel can be matched in a non-coherent channel only if the bandwidth

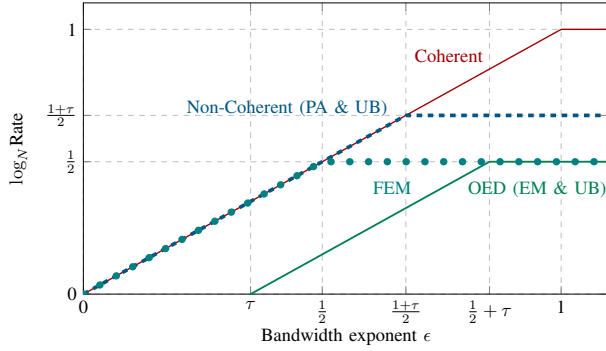


Fig. 3: Achievable rate exponents with $0 < \tau < 1$.

is below a certain threshold, using either *a*) a FEM scheme when $\epsilon < \frac{1}{2}$ or *b*) a PA scheme when $\epsilon < \frac{1+\tau}{2}$. The latter is also the capacity scaling of the non-coherent channel. For $\tau \geq 1$ the capacity scaling of the non-coherent channel equals that of the coherent channel for all bandwidths. The OED schemes designed with a strict interpretation of the energy-shape separation by Marzetta and Hochwald, such as the canonical EM scheme, cannot achieve the capacity scaling of the non-coherent channel at all $\forall \tau > 0$.

V. NON-COHERENT CAPACITY UPPER BOUND

A. Subchannel Decomposition

Each subchannel output depends only on the corresponding subchannel input, so we can separate the mutual information using the chain rule. Moreover the subchannels are i.i.d. and there is no CSI nor feedback, so time-diversity is exploited by the ergodic capacity average and frequency diversity is unnecessary. Therefore, we can fully decompose the rate as a sum of subchannel rate contributions:

$$\begin{aligned}
 C_n(B, N, L) &= \sup_{\substack{p(\{\mathbf{x}[b]\}_{b=0}^{B-1}) \\ \text{s.t. } \sum_{b=0}^{B-1} \mathbb{E}[|\mathbf{x}[b]|^2] \leq PL}} \frac{\mathbb{E}_{\{\mathbf{h}[b]\}_{b=0}^{B-1}} [\mathbb{I}(\{\mathbf{x}[b]\}_{b=0}^{B-1}; \{\mathbf{Y}[b]\}_{b=0}^{B-1})]}{L/\Delta f} \\
 &= \frac{\Delta f}{L} \sup_{\substack{p(\{\mathbf{x}[b]\}_{b=0}^{B-1}) \text{ s.t.} \\ \sum_{b=0}^{B-1} \mathbb{E}[|\mathbf{x}[b]|^2] \leq PL}} \sum_{b=0}^{B-1} \mathbb{E}_{\mathbf{h}[b]} [\mathbb{I}(\mathbf{x}[b]; \mathbf{Y}[b])] \\
 &= \frac{\Delta f}{L} \sup_{\{\rho_b\}_{b=0}^{B-1} \text{ s.t. } \sum_{b=0}^{B-1} \rho_b \leq PL} \sum_{b=0}^{B-1} f_C(\rho_b)
 \end{aligned} \tag{11}$$

where ρ_b is the power allocated to subchannel b , and

$$f_C(\rho_b) \triangleq \sup_{p(\mathbf{x}[b]) \text{ s.t. } \mathbb{E}[|\mathbf{x}[b]|^2] \leq \rho_b} \mathbb{E}_{\mathbf{h}[b]} [\mathbb{I}(\mathbf{x}[b]; \mathbf{Y}[b])]$$

is the subchannel capacity contribution in each of the identical subchannels as a function of ρ_b .

Since all subchannels are identical and $f_C(\rho)$ is an increasing function of ρ , by the symmetry of the problem (11) has a solution of the form

$$C_n(B, N, L) = \frac{\Delta f}{L} \sup_{M \in \{1, \dots, B\}} M f_C(PL/M). \tag{12}$$

In summary, to compute the capacity we must solve two problems: find the input distribution to achieve the maximum in $f_C(PL/M)$ for one subchannel, given a value of M ; and then find the optimum number of subchannels that should be actively used, with equal power allocations to each. We reiterate that in this paper we assume a non-coherent channel without feedback in which the optimality of this equal power allocation is reasonable. However, in the cases with feedback or CSIT that we leave for future work, equal power allocation would likely be suboptimal.

B. Energy-Shape Decomposition

Since subchannels are i.i.d., hereafter we omit the index $[b]$ to simplify notation. Using Lemma 3, we need only to find the best energy p.d.f. $p(a)$, whereas the shape distribution is known to be uniform in the hyper sphere $p(\mathbf{u}) = \frac{(\pi)^L}{(L-1)!} \delta(\|\mathbf{u}\|^2 - 1)$. Using the chain rule we can write

$$f_C(\rho) = \sup_{\substack{p(a) \text{ s.t.} \\ \mathbb{E}[|a|^2] \leq \rho}} \mathbb{E}_{\mathbf{h}} [\mathbb{I}(a; \mathbf{Y})] + \mathbb{E}_{\mathbf{h}, a} [\mathbb{I}(\mathbf{u}; \mathbf{Y}|a)] \Big|_{\mathbf{u} \sim \text{IDUV}}. \tag{13}$$

We recall that we must further maximize over the number of active subchannels M in (12). By adding the chain rule and Lemma 3, this results in

$$\begin{aligned}
 C_n(B, N, L) &= \sup_{\substack{1 \leq M \leq B \\ p(a) \text{ s.t.} \\ a > 0, \mathbb{E}[a] \leq \frac{PL}{M}}} \frac{M \Delta f}{L} \left(\mathbb{E}_{\mathbf{h}} [\mathbb{I}(a; \mathbf{Y})] + \mathbb{E}_{\mathbf{h}, a} [\mathbb{I}(\mathbf{u}; \mathbf{Y}|a)] \Big|_{\mathbf{u} \sim \text{IDUV}} \right)
 \end{aligned} \tag{14}$$

Moreover, we may upper bound the capacity (12) by performing separate optimizations in the first and second terms. Thus, we define the ‘‘Only Energy Detection Capacity’’ and the ‘‘Shape Encoding Capacity’’ such that their sum forms an upper bound to the non-coherent capacity as follows:

$$C_n(B, N, L) \leq C_E(B, N, L) + C_S(B, N, L)$$

with:

$$\begin{aligned}
 C_E(B, N, L) &= \sup_{\substack{1 \leq M \leq B \\ p(a) \text{ s.t.} \\ a > 0, \mathbb{E}[a] \leq \frac{PL}{M}}} \frac{M \Delta f}{L} \mathbb{E}_{\mathbf{h}} [\mathbb{I}(a; \mathbf{Y})] \\
 C_S(B, N, L) &= \sup_{\substack{1 \leq M \leq B \\ p(a) \text{ s.t.} \\ a > 0, \mathbb{E}[a] \leq \frac{PL}{M}}} \frac{M \Delta f}{L} \mathbb{E}_{\mathbf{h}, a} [\mathbb{I}(\mathbf{u}; \mathbf{Y}|a)] \Big|_{\mathbf{u} \sim \text{IDUV}}
 \end{aligned} \tag{15}$$

C. Scaling of the Sum of Upper Bounds

We have upper bounded $C_E(B, N, L) \leq \Theta(N^{\min(\epsilon - \tau, \frac{1}{2})})$ in Theorem 2, which is proven in Appendix B. For the second term we produce the following upper bound:

Lemma 7. *In a Non-Coherent Block Fading Rayleigh i.i.d. wideband SIMO channel (2) with N receive antennas, bandwidth $B = \Theta(N^\epsilon)$ and $L = \Theta(N^\tau)$, the Shape-Encoding Capacity defined in (15), scales at most as*

$$C_S(B, N, L) \leq \Theta(N^{\min(\epsilon, \frac{1+\tau}{2}, 1)}) \quad (16)$$

Proof. The proof is given in Appendices C-E. \square

Since for all $\epsilon > 0$ we have that

$$\min(\epsilon - \tau, \frac{1}{2}) < \min(\epsilon, \frac{1+\tau}{2}, 1),$$

we observe that $C_E(B, N, L) \ll C_S(B, N, L)$ for sufficiently large N , and we can conclude that the non-coherent channel capacity scaling is upper bounded by

$$C_n(B, N, L) \leq 2C_S(B, N, L) = \Theta(N^{\min(\epsilon, \frac{1+\tau}{2}, 1)}).$$

This concludes the proof of the converse of Theorem 1.

VI. PILOT-ASSISTED COHERENT SCHEME

We propose a simple pilot-assisted communication scheme on the channel (1). The design of the optimal number of pilot symbols and their allocated power in narrowband Rayleigh block fading MIMO channels was studied in [28]. We assume M subchannels are used simultaneously, with equal power allocation $\rho = \frac{PL}{M}$ each. Since we consider the SIMO case, one symbol in each subchannel b must be used as a pilot [28]. We assume the first symbol in each coherence block of each subchannel is the pilot $x_p[b] = x_0[b]$, and the remaining $L-1$ symbols are used to convey data $\{x_\ell[b]\}_{\ell=1}^{L-1}$. Moreover, we assume the power budget per subchannel is divided between pilots and data as $\mathbb{E}[|x_p|^2] = \alpha\rho$ and $\mathbb{E}\left[\frac{1}{L-1} \sum_{\ell=1}^{L-1} |x_\ell[b]|^2\right] = \frac{(1-\alpha)\rho}{L-1}$, for some fraction $\alpha \in [0, 1]$.

We assume an independent estimation of each i.i.d. $h_n[b]$ using the first channel output in each antenna as

$$y_{p,n}[b] = h_n[b]x_p[b] + z_n[b]$$

where $z_n[b] \sim \mathcal{CN}(0, 1)$. To maximize the achievable rate, the Minimum Mean Squared Error (MMSE) estimation $\hat{h}_n[b] = \frac{x_p^*}{|x_p|^2 + 1} y_{p,n}[b]$ is employed [28]. Using a constant amplitude pilot $|x_p[b]| = \sqrt{\alpha\rho}$ minimizes the variance of the MMSE estimator error, defined as $\tilde{h}_n[b] = h_n[b] - \hat{h}_n[b] \sim \mathcal{CN}(0, \varepsilon)$, where $\varepsilon = \frac{1}{1+\alpha\rho}$ [28].

In the data part, for $1 \leq \ell < L-1$, we write

$$y_{n,\ell}[b] = \hat{h}_n[b]x_\ell[b] + \tilde{h}_n[b]x_\ell[b] + z_{n,\ell}[b],$$

where we assume i.i.d. symbols, so $\mathbb{E}[|x_\ell[b]|^2] = \frac{(1-\alpha)}{L-1}\rho$.

In addition, we assume the receiver employs Maximum Ratio Combining (MRC) treating $\hat{h}_n[b]$ as the true channel, producing

$$\begin{aligned} r_\ell[b] &= \hat{\mathbf{h}}^H[b] \mathbf{y}_\ell[b] \\ &= \|\hat{\mathbf{h}}[b]\|^2 x_\ell[b] + \left(\sum_{n=0}^N \hat{h}_n^*[b] \tilde{h}_n[b] \right) x_\ell[b] + z_{n,\ell}[b]. \end{aligned} \quad (17)$$

Using a nearest neighbor decision is not optimal, but is sufficient to achieve at least the same rate as in an AWGN channel with an equivalent noise variance equal to the power of the second plus third terms of $r_\ell[b]$ [28]. Since $(\mathbf{h}_n[b]|\hat{\mathbf{h}}_n[b])$ is distributed as $\mathcal{CN}(\hat{\mathbf{h}}_n[b], \varepsilon\mathbf{I})$, the effective noise variance is $\varepsilon \frac{(1-\alpha)}{L-1} \rho + 1$, and the effective signal energy is $\frac{(1-\alpha)}{L-1} \rho \|\mathbf{h}[b]\|^2 (1-\varepsilon)$. Thus the achievable rate as a function of the perceived SNR at the symbol decider is

$$R_{PA}(\alpha) \triangleq \frac{L-1}{L} M \log \left(1 + \frac{\frac{(1-\alpha)}{L-1} \rho \|\mathbf{h}[b]\|^2 (1-\varepsilon)}{1 + \varepsilon \frac{(1-\alpha)}{L-1} \rho} \right) \quad (18)$$

In order to study the scaling as $N \rightarrow \infty$ we introduce the approximation $\|\mathbf{h}[b]\|^2 \approx N$ for large N . We also define $\kappa \triangleq \frac{\rho}{L-1}$, and replace $\varepsilon = 1/(1+\alpha\rho)$. With some rearrangement, this produces

$$\begin{aligned} R_{PA}(\alpha) &\simeq \frac{L-1}{L} M \log \left(1 + \frac{(1-\alpha)\kappa N \rho \alpha}{1 + \alpha\rho + (1-\alpha)\kappa} \right) \\ &= \frac{L-1}{L} M \log \left(1 + \frac{N\kappa\rho}{1 + \kappa} \underbrace{\frac{\alpha - \alpha^2}{\alpha \frac{\rho - \kappa}{1 + \kappa} + 1}}_{\gamma(\alpha)} \right) \end{aligned} \quad (19)$$

which we can maximize by selecting α to maximize the decider-perceived SNR inside the logarithm, $S_o\gamma(\alpha)$. We perform a standard maximization of the function of the form $\gamma(\alpha) = \frac{\alpha - \alpha^2}{\alpha C_1 + 1}$, where the constant is $C_1 = \frac{\rho - \kappa}{1 + \kappa}$. This produces the following solution

$$\alpha^* = \frac{\sqrt{1 + C_1} - 1}{C_1}.$$

For convenience, we can rewrite the same result in the following forms as well

$$\alpha^* = \frac{1}{\sqrt{1 + C_1} + 1} = \frac{1}{\sqrt{\frac{1+\rho}{1+\kappa}} + 1}.$$

In the third form, we can observe that, since $\rho > 0$ and $\rho > \kappa > 0$, one unconstrained global solution always satisfies $0 \leq \alpha^* \leq 1$. Thus there is no need to introduce a Lagrangian multiplier to enforce such a constraint, and the global solution may be adopted directly.

Next we define a second constant $C_2 = \sqrt{\frac{1+\rho}{1+\kappa}} = \sqrt{C_1 + 1}$. This allows to write, $\alpha^* = \frac{1}{C_2 + 1}$, $1 - \alpha^* = \frac{C_2}{C_2 + 1}$, and $\alpha^* C_1 + 1 = C_2$. Substituting we get

$$\begin{aligned} S_o\gamma(\alpha^*) &= S_o \frac{(1 - \alpha^*)\alpha^*}{\alpha^* C_1 + 1} \\ &= S_o \frac{1}{(C_2 + 1)^2} \\ &= \frac{N\kappa\rho}{1 + \kappa} \frac{1}{\left(\sqrt{\frac{1+\rho}{1+\kappa}} + 1\right)^2} \\ &= \frac{N\kappa\rho}{(\sqrt{1 + \rho} + \sqrt{1 + \kappa})^2} \end{aligned} \quad (20)$$

Let $M = \Theta(N^\mu)$ with $0 \leq \mu \leq \epsilon$. We study the scaling of the decider-perceived SNR, $S_o\gamma(\alpha^*)$, as $N \rightarrow \infty$ for different

values of τ and μ . We recall that $\rho = \Theta(N^{\tau-\mu})$ and $\kappa = \Theta(N^{-\mu})$. Thus, we must distinguish two different cases

- 1) For $\mu \leq \tau$, we have $\rho \gg 1 \gg \kappa$. Therefore, $S_o\gamma(\alpha^*) \simeq N\kappa = \Theta(N^{1-\mu})$. For all ϵ we can choose $\mu \leq \min(\epsilon, \tau)$ and, if $\mu \leq 1$, (18) scales as $R_{PA} = \Theta(N^\mu)$. Conversely, if $\tau > \mu > 1$, $R_{PA} = \Theta(N)$.
- 2) For $\mu > \tau$ we have $1 \gg \rho \gg \kappa$. This means that, $S_o\gamma(\alpha^*) \simeq N\kappa\rho = \Theta(N^{1+\tau-2\mu})$. For all ϵ we can choose $\mu \leq \min(\epsilon, \frac{1+\tau}{2})$ and (18) scales as $R_{PA} = \Theta(N^\mu)$. On the other hand, if we choose $\mu > \frac{1+\tau}{2}$, R_{PA} scales as $\Theta(N^{1+\tau-\mu}) \leq \Theta(N^{\frac{1+\tau}{2}})$. Clearly, making μ increase above $\frac{1+\tau}{2}$ is not beneficial, and the bandwidth overspreading constraint is $M \leq \Theta(\sqrt{NL}) = \Theta(N^{\frac{1+\tau}{2}})$.

When $\tau \leq 1$, $\frac{1+\tau}{2}$ is greater than τ , with equality at $\tau = 1$. Therefore, we can summarize all the PA achievable rate scaling with the following expression

$$R_{PA} = \Theta(N^{\min(\epsilon, \frac{1+\tau}{2}, 1)}),$$

concluding the proof of the achievability of Theorem 1.

VII. FAST ENERGY MODULATION (FEM)

We introduce an improved FEM scheme in order to compensate for the poor spectral efficiency of EM, which stems from its use of a $L \times$ repetition code. Although we propose FEM as a minor modification of EM, these two modulations echo the abundant literature on energy detection non-coherent schemes (see [8] and references therein). To improve the spectral efficiency, we simply remove the repetition coding component as follows: In each block of L symbols, the transmitter selects L independent energy symbols

$$a_0[b] \dots a_{L-1}[b] \in \mathcal{C}_M^L,$$

and the transmitted signal is created as

$$x_\ell[b] = \sqrt{a_\ell[b]}.$$

As in EM, we introduce an energy statistic for decoding, which now must take the form

$$v_\ell[b] = \sum_{n=0}^{N-1} |y_{n,\ell}[b]|^2 \sim (a_\ell[b]^2 + 1)\chi^2(2N).$$

Here, we note that there is no repetition anymore, an independent statistic $v_\ell[b]$ is calculated and an independent nearest neighbor $\hat{a}_\ell[b]$ is decided for each value of ℓ and $[b]$. Since the channel remains constant for the duration of the block of L symbols, yet the decoder is making symbol by symbol decisions, this decoder is clearly suboptimal. Nevertheless, this scheme is sufficient to guarantee that the error probability vanishes as $N \rightarrow \infty$ for the desired rates.

To show this we need only to introduce a minor modification to the error analysis of [27, theorem 2] by considering the union bound over L independently decoded symbols. Since a $\times L$ repetition code is no longer present as in EM, the inter-symbol distance requirement for FEM is that $d = \Theta(N^t)$ must now satisfy $t < \frac{1}{2}$. The constellation size is thus $\log_2(\frac{N^t}{M}) \leq \log_2(N^{\frac{1}{2}-\epsilon})$. As a result, the total rate of FEM is $\Theta(M \log_2(N))$ with $M \leq \Theta(N^{\min(\epsilon, \frac{1}{2})})$.

For $\tau = 0$, EM and FEM schemes achieve the same scaling; hence the use of EM in [24] was sufficient to achieve the upper bound and prove the capacity scaling result. On the other hand, for $\tau > 0$, EM cannot properly achieve the coherent capacity scaling, whereas FEM can for $\epsilon \leq \frac{1}{2}$.

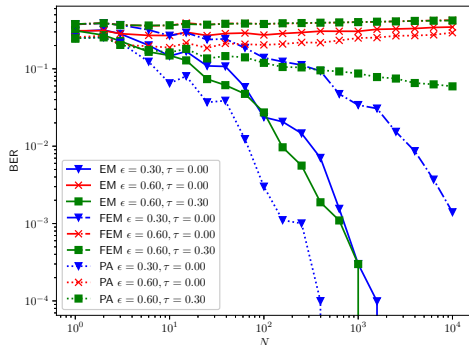
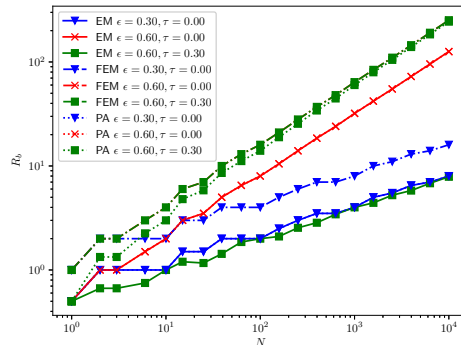
We remark that the FEM scheme we propose does **not** coincide with the Marzetta-Hochwald energy and shape decomposition of the input in Lemma 3. To prove the achievability of Theorem 2, in which we ultimately concluded that both terms in the Marzetta-Hochwald input decomposition are relevant, we have considered the EM scheme. FEM always achieves better rate exponent than EM, and it can achieve the capacity scaling for $\epsilon \leq \frac{1}{2}$. Nonetheless FEM does, in fact, encode information in the vector $\mathbf{u}[b]$ of the decomposition $\mathbf{x}[b] = a[b]\mathbf{u}[b]$, and is not a canonical OED scheme in the sense of Lemma 3.

VIII. SIMULATION RESULTS

We demonstrate our results by simulating the progression of the Bit Error Rate (BER) and rate in the channel (2) as N increases for different practical schemes. We simulate three achievable schemes: EM, FEM and PA, for different values of ϵ and τ . To implement the simulation, the number of subchannels and block length must be integers, so we adopt $B = \lceil N^\epsilon \rceil$ and $L = \lceil N^\tau \rceil$. For EM and FEM, we adopt a binary constellation with symbols $\mathcal{C}^{EM} = \{0, \frac{2}{M}\}$. Recall that EM repeats the same symbol L times in each channel coherence block, whereas FEM transmits sequences of L i.i.d. symbols. For PA, we adopt a scaled BPSK constellation $\mathcal{C}^{PA} = \{-1, +1\}$ where the first symbol is a pilot scaled by a factor $PL\alpha^*$ and the remaining data symbols are scaled by a factor $\frac{PL}{L-1}(1-\alpha^*)$ according to the results in Section VI. We set $P = 2$ so that the initial SNR per receive antenna is 3 dB when $B = 1$.

We depict in Fig. 4 the BER for all schemes. We recall that the EM requires that $\epsilon < \frac{1}{2} + \tau$, PA requires $\epsilon < \frac{1+\tau}{2}$ and FEM requires the strictest condition $\epsilon < \frac{1}{2}$. For the case $\epsilon = 0.3$ and $\tau = 0$ (blue), all three conditions are met and all schemes are able to achieve a decreasing error probability as N increases. It must be noted that 10^4 symbols were simulated and the ‘‘saw tooth’’ shape of the curve is due to the use of the ceiling function to compute M and not to a lack of sufficient bits for Monte Carlo approximation. For the case $\epsilon = 0.6$ and $\tau = 0$ (red), all conditions are unmet and all schemes experience a large error probability that does not decay as N increases. Finally, when $\epsilon = 0.6$ and $\tau = 0.3$ (green), FEM experiences a large error probability, as the condition $\epsilon < \frac{1}{2}$ is not satisfied. On the other hand, the conditions $\epsilon < \frac{1}{2} + \tau$ and $\epsilon < \frac{1+\tau}{2}$ are both satisfied, and both EM and PA schemes display a decaying error probability as N increases.

Next, we look at the **nominal** bit rates of each scheme in Fig. 5, defined simply as the number of transmitted bits per second. For EM, the nominal rate is $R_b^{EM} = B \log \|\mathcal{C}^{EM}\|/L$. For FEM, the nominal rate is $R_b^{FEM} = B \log \|\mathcal{C}^{EM}\|$. And for PA, the nominal rate is $R_b^{EM} = B \log \|\mathcal{C}^{PA}\|(L-1)/L$. For the case $\epsilon = 0.3$ and $\tau = 0$ (blue), we can see that all schemes achieve rates that scale as $N^{0.3}$. Moreover, as we

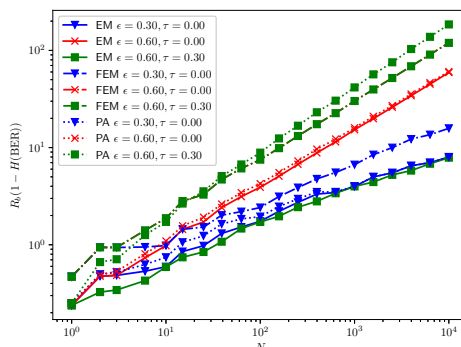
Fig. 4: BER vs N for different schemes.Fig. 5: Bit rate vs N for different schemes.

saw in the BER curves, these rates correspond to a “reliable communication” in the sense that the BER vanishes even as the rate grows. For the case $\epsilon = 0.6$ and $\tau = 0$ (red), Fig. 5 indicates that all schemes transmit with rates that scale as $N^{0.6}$. However, as we saw in the BER curves, these are capacity-exceeding transmission rates in the sense that the communication at these rates is not reliable. Finally, when $\epsilon = 0.6$ and $\tau = 0.3$ (green), FEM transmits with a rate that scales as $N^{0.6}$ that is not reliable, EM communicates reliably but its rate scales only as $N^{0.3}$, and only PA can simultaneously guarantee that the communication is reliable and that the rate scales as $N^{0.6}$.

The comparison of nominal rates and error probability in Figs. 4 and 5 can be unified informally by using an ad-hoc metric we call the **Binary Symmetric Channel (BSC) equivalent reliable rate**, defined as $R_{BSCeq} = R_b(1 - H(BER))$. This is, we compute the capacity of a BSC in which the error probability is equal to the empirical BER, and the bit rate is equal to the nominal bit rate of each modulation. Although this definition is not very rigorous, this ad-hoc metric allows to visualize the fraction of the nominal rate that is reliable in the simulation. We show the BSC equivalent reliable rates in Fig. 6. Again, for $\epsilon = 0.3$ and $\tau = 0$ (blue), all schemes achieve reliable rates that scale as $N^{0.3}$. We can also see that in the case with $\epsilon = 0.6$ and $\tau = 0$ all schemes achieve reliable rates that scale as $N^{0.5}$. Finally, for the case $\epsilon = 0.6$ and $\tau = 0.3$, EM achieves only a rate scaling of $N^{0.3}$, FEM achieves a reliable rate scaling of $N^{0.5}$, and only the PA scheme can achieve a reliable rate scaling of $N^{0.6} = N^\epsilon$. Despite the non-rigorous definition of R_{BSCeq} , the concordance of all these results with our capacity analyses is complete and quite remarkable.

IX. CONCLUSIONS

In this paper we have studied the scaling of the capacity gap between coherent and non-coherent wideband SIMO Rayleigh block-fading channels, as a function of the bandwidth, number of receive antennas and channel coherence block length. The main insight of our paper is that the channel coherence block length plays a critical role in the capacity gap between coherent and non-coherent channels. Moreover, the maximum bandwidth allocation that can be used to avoid “overspreading”

Fig. 6: Equivalent BSC capacity vs N for different schemes.

is also a key factor. In some prior formulations of the study of non-coherent channel capacity, Energy Detection techniques such as Energy Modulation have arisen as a promising encoding technique due to its capacity-like scaling with the number of antennas when the coherence block length is fixed. On the other hand, many practical technologies employ Pilot Assisted channel estimation and coherent receivers. Our result shows that Pilot Assisted schemes are always capacity-scaling-achieving in non-coherent channels, whereas Energy Detection schemes only achieve the capacity scaling in a more restricted sense that does not take into account the channel coherence block length.

Two important engineering messages emerge from our analysis. One is that as new technology standards continue to incorporate more spectrum, attention must be paid to the excess channel uncertainty caused by bandwidth overspreading, in which the capacity enters the wideband regime with power limited rates and the addition of even more degrees of freedom is not beneficial. The threshold to enter this regime scales as $\Theta(\min(\sqrt{NL}, N))$, which means that the limitation is more severe in high mobility scenarios. In addition, increasing the number of receive antennas can actually help push up the critical bandwidth threshold and permit to keep increasing wireless network rates by purchasing more spectrum.

The second key engineering lesson of our analysis is that the optimal choice between channel estimation or non-coherent

encoding schemes seems to favor the first option in a significant number of cases. Only in very high mobility scenarios, which result in very short channel coherence length, the scaling analysis suggests that general Energy Detection schemes, and Energy Modulation in particular, are competitive in a rate-scaling sense. And even in those cases, both methods achieve equal scaling and a finer comparison of the rates beyond the scaling analysis should be performed before making the decision.

APPENDIX A DISCUSSION OF THE CHANNEL MODEL

The results in this paper are proven for an i.i.d. Rayleigh block-fading model. This classical fading model is extremely valued in wireless communications research for its excellent mathematical tractability while the results maintain some degree of proximity to the physical truth. Nevertheless, when considering a very large bandwidth or antenna array, engineers should be specially wary about the difference between this popular channel model and the physical reality of multipath propagation. The usual justification of i.i.d. Rayleigh fading assumptions is rich scattering, that is, assuming a number of multipath reflections that is much higher than the number of taps, frequency bins, array elements, or both. This would motivate using the Central Limit Theorem (CLT) to characterize the coefficients, as a sum of many unknowns, as i.i.d. Gaussian. However, when the bandwidth is very large, the delays of multi-path reflections in the channel can be resolved individually. Likewise, the Angle of Arrival can be resolved in large antenna arrays with elements separated half a wavelength. These phenomena induce frequency correlation (delay sparsity) and spatial correlation (angular sparsity), respectively. In mmWave channels, advanced Saleh-Valenzuela sparse multipath models are more accurate [38]. In fact, many results show that in mmWave the CLT does not even justify the modeling of each reflection coefficient as Gaussian [39]–[43]. The full-fledged physically-realist mmWave channel models in [40], [41] are, in fact, too cumbersome for analytical tractability. Theoretical analyses have resorted to different degrees of simplification ranging from sparse-Gaussian multipath [14], [44] to almost-rich-scattering modified Rician or Rayleigh models that capture specific aspects of mmWave [31] such as oxygen absorption [45]. More recently, there has been interest in “holographic MIMO” models that assume a very large number of antennas packed into a spatially constrained volume and spaced closer than $\lambda/2$ [46], [47], in contrast to earlier mmWave array models with elements spaced $\lambda/2$ [40], [44], in which the antenna aperture goes to infinity with N .

The compromise between tractability and validity of the channel model is a common occurrence in capacity analyses also outside mmWave. Lozano and Pórrat indicate that their seminal critical bandwidth calculation holds for the “wide-band” channel model, but may lose applicability as the bandwidth increases and the channel behaves as in the “ultra-wideband” model [20]. Özgür, Leveque and Tse pointed out a similar weakness in network capacity scaling analysis [48]. However, these authors have argued that the assumption

of i.i.d. coefficients is justifiable when there is a range of the scaling analysis parameters in which simultaneously the number of elements is large and the channel assumptions hold. In our results, the argument is similar: our analysis is relevant for the range of values of the parameters B , N , and L that are large-but-finite, and that are simultaneously sufficiently large for the big-O scaling results to dominate capacity, but not large enough that the channel model needs to be replaced. Abundant literature suggests that such a range of parameters exists. The seminal work by Medard and Gallager [7] characterized the problem of overspreading without our assumptions on the channel model. Telatar and Tse showed that even a sparse multipath channel with a single path experiences overspreading when the reflection delay is not known a-priori [14]. Raghavan *et al* found similar results for capacity scaling with bandwidth under sparse multipath channels [49]. And the revision of the critical bandwidth results for mmWave by Ferrante *et al* [45] simplified the model to a multiplicative combination of rich Ricean fading, Bernoulli blockage and oxygen absorption.

In addition, the 5G [2] and WiFi6 [3], [4] standards employ OFDMA waveforms that treat each “frame” as an independent channel block realization with separate pilot estimation. Therefore, even when these standards operate on top of an underlying physical propagation that does not respond to the i.i.d. Rayleigh block fading model, the *channel model assumption* that is implicit to the waveform design and engineering of many practical devices is indeed consistent with our model. In summary, while readers ought to be cautious about the channel model, our result insights are reasonably applicable to current wireless systems, and give analysts an important starting step toward more general results with advanced channel models in future work.

APPENDIX B OED CAPACITY UPPER BOUND

Here we present an upper bound of the first term of (14), $C_E(B, N, L)$, defined in (15). This also proves the converse of Theorem 2.

Let us define the per-subchannel OED Capacity contribution function

$$f_E(\rho) \triangleq \sup_{p(a) \text{ s.t. } a > 0, E[a] \leq \rho} \mathbb{E}_{\mathbf{h}} [\mathbb{I}(a; \mathbf{Y})]$$

so that the OED Capacity satisfies

$$C_E(B, N, L) = \sup_{1 \leq M \leq B} \frac{M \Delta f}{L} f_E\left(\frac{PL}{M}\right).$$

We can use $\mathbf{x}^H = \sqrt{a} \mathbf{u}^H$ as defined in Lemma 3 to define the auxiliary random variable

$$\mathbf{v} = \sqrt{a} \mathbf{h} + \mathbf{z}_a, \quad (21)$$

where $\mathbf{z}_a = \mathbf{Z} \mathbf{u}$ is i.i.d. AWGN of dimensions $N \times 1$. Since $\mathbf{Y} = \mathbf{v} \mathbf{u}^H + \mathbf{Z}(\mathbf{I} - \mathbf{u} \mathbf{u}^H)$, we verify that $p(\mathbf{Y} | \mathbf{v}, a) = p(\mathbf{Y} | \mathbf{v})$ and therefore $a \rightarrow \mathbf{v} \rightarrow \mathbf{Y}$ is a Markov chain satisfying

$$f_E(\rho) \leq \sup_{p(a) \text{ s.t. } a > 0, E[a] \leq \rho} \mathbb{I}(a; \mathbf{v}) \triangleq g_E(\rho).$$

Finally, we have that (21) is a SIMO Rayleigh fading channel with block length of $L = 1$ and with power constraint $E[a] \leq \frac{PL}{M}$, so we can apply [27, Lemma 3] to $g_E(\rho)$

Lemma 8. (*[27, Lemma 3]*) For $\rho = \frac{PL}{M} = \Theta(N^{\frac{1}{2}+\alpha})$ with any $\alpha > 0$, the upper bound $g_E(\rho) \leq \Theta(\frac{1}{N^{2\alpha}})$ is satisfied.

Proof. Same as [27] with power constraint $\rho = \frac{PL}{M}$. \square

As a result, when $B < \Theta(N^{\frac{1}{2}+\tau})$, $f_E(\frac{P}{M})$ is upper bounded by the coherent capacity of the channel (21), and the optimal number of active subchannels is $M = B$. When $B \geq \Theta(N^{\frac{1}{2}+\tau})$, $g_E(\frac{P}{M})$ is upper bounded by Lemma 8 and the transmitter must not overspread the transmitted power in more than $M = \Theta(\sqrt{NL})$ subchannels. Thus in general the optimal M is $\Theta(N^{\min(\epsilon, \frac{1}{2}+\tau)})$ and, substituting into $\frac{M}{L}g_E(\frac{PL}{M})$, we get (9), completing the proof of Theorem 2.

APPENDIX C

SHAPE ENCODING CAPACITY UPPER BOUND

Here we present an upper bound of the second term in (14), $C_S(B, N, L)$, defined in (15). We recall the definition of the conditional mutual information

$$E_a [I(\mathbf{u}; \mathbf{Y}|a)] = E_a \left[E_{\mathbf{Y}, \mathbf{u}} \left[\log \frac{p(\mathbf{Y}|\mathbf{u}, a)}{p(\mathbf{Y}|a)} \right] \right].$$

Since we cannot compute this information exactly, we will proceed by first upper bounding the information function defined as $I(\mathbf{u}; \mathbf{Y}|a) \leq \Phi(a)$ for given values of a . And finally we will maximize $E_a[\Phi(a)]$ to compute an upper bound of the scaling of $C_S(B, N, L)$.

Thanks to the fact that \mathbf{u} is unitary and \mathbf{u}, a are known, we compute the distribution $p(\mathbf{Y}|\mathbf{u}, a)$ explicitly in Appendix D, resulting in

$$p(\mathbf{Y}|\mathbf{u}, a) = (1+a)^{-N} (2\pi)^{-NL} \exp \left[-(\|\mathbf{Y}\|^2 - \frac{a\|\mathbf{Y}\mathbf{u}\|^2}{1+a}) \right]$$

For the denominator, we settle for simply computing a function that lower bounds $p(\mathbf{Y}|a)$ as a function of a (Appendix E). This function is **not** a p.d.f., and is expressed as

$$p(\mathbf{Y}|a) \geq (1+a)^{-N} (2\pi)^{-NL} \exp \left[-(\|\mathbf{Y}\|^2 - \frac{a}{1+a} \frac{\|\mathbf{Y}\|^2}{L}) \right]$$

Combining the above, we have the following upper bound for the mutual information

$$I(\mathbf{u}; \mathbf{Y}|a) \leq \frac{a}{1+a} \left(E_{\mathbf{Y}, \mathbf{u}} [\|\mathbf{Y}\mathbf{u}\|^2] - \frac{E_{\mathbf{Y}} [\|\mathbf{Y}\|^2]}{L} \right) \quad (22)$$

To continue we need to calculate the average energies $E[\|\mathbf{Y}\mathbf{u}\|^2]$ and $E[\|\mathbf{Y}\|^2]$. Noting that $\mathbf{Y}\mathbf{u} = \sqrt{a}\mathbf{h}\mathbf{u}^H\mathbf{u} + \mathbf{Z}\mathbf{u} = \sqrt{a}\mathbf{h} + \mathbf{z}'$, where $\mathbf{z}' \sim \mathcal{CN}(0, I_N)$, we get that $E[\|\mathbf{Y}\mathbf{u}\|^2] = N(1+a)$. With similar reasoning we can deduce that $E[\|\mathbf{Y}\|^2] = N(L+a)$ from the fact that $\|\mathbf{h}\mathbf{u}^H\|^2 = \|\mathbf{h}\|^2$ and that $E[\|\mathbf{Z}\|^2] = LE[\|\mathbf{z}'\|^2]$. Thus we arrive at the following upper bound to the mutual information

$$I(\mathbf{u}; \mathbf{Y}|a) \leq \Phi_1(a) \triangleq \frac{a^2}{1+a} N \left(1 - \frac{1}{L} \right). \quad (23)$$

Since $\Phi_1(a)$ is convex, unless a is subject to a peak constraint, introducing the upper bound $\Phi_1(a)$ into the average and maximizing with regard to $p(a)$ obtains an upper bound to the Shape Encoding Capacity that is too loose. In particular

$$\sup_{p(a)} E_a [I(\mathbf{u}; \mathbf{Y}|a)] < \sup_{p(a)} E_a [\Phi_1(a)] = \rho N \left(1 - \frac{1}{L} \right).$$

where the supremum is achieved by the peaky distribution

$$p(a) = \lim_{\delta \rightarrow 0} \begin{cases} \rho/\delta & \text{w.p. } \delta \\ 0 & \text{otherwise} \end{cases}.$$

In other words, while $\Phi_1(a)$ can upper bound $I(\mathbf{u}; \mathbf{Y}|a)$ for small values of a , we cannot use this upper bound to study the optimal distribution of $p(a)$ when there are no peak constraints. Therefore, as a correction factor to $\Phi_1(a)$, let us use the capacity of an AWGN channel of power a as a second upper bound to the mutual information

$$\begin{aligned} I(\mathbf{u}; \mathbf{Y}|a) &\stackrel{a}{\leq} \sup_{p(\mathbf{x}): E[\|\mathbf{x}\|^2] \leq a} I(\mathbf{x}; \mathbf{Y}) \\ &\stackrel{b}{\leq} L \log(1 + aN) \\ &\stackrel{c}{\triangleq} \Phi_2(a), \end{aligned} \quad (24)$$

where a stems from the fact that \mathbf{u} is an IDUV, so get that $E[\|\mathbf{u}\mathbf{u}\|^2] = a$, and therefore the inequality with the supremum holds. Next, b) upper bounds the non-coherent mutual information by the coherent capacity with power constraint a . And c) simply writes this second upper bound as a function depending on a .

Finally, we can properly conduct an input distribution optimization with regard to the ‘‘useful’’ upper bound defined as the minimum of $\Phi_1(a)$ and $\Phi_2(a)$

$$\begin{aligned} I(\mathbf{u}; \mathbf{Y}|a) &\leq \Phi(a) \\ &\triangleq \min [\Phi_1(a), \Phi_2(a)] \\ &= \min \left[\frac{a^2}{1+a} N \left(1 - \frac{1}{L} \right), L \log(1 + aN) \right] \end{aligned} \quad (25)$$

Leading to

$$C_S(B, N, L) \leq \sup_{M \in \{1 \dots B\}} \frac{M\Delta f}{L} \sup_{p(a)} E_a [\Phi(a)].$$

The shape of $\Phi(a)$ is very particular. Let us define the point where the two sides of the minimum are equal as a_o satisfying

$$\frac{a_o^2}{1+a_o} N \left(1 - \frac{1}{L} \right) = L \log(1 + a_o N). \quad (26)$$

For $a < a_o$, $\Phi(a)$ is convex, and for $a > a_o$, $\Phi(a)$ is concave. Therefore, we can solve the optimization separately in the concave and convex regions. To this end, we define the auxiliary variable $\Upsilon = \int_0^{a_o} p(a) da$, and the auxiliary distributions $q(a) = \frac{p(a|a \leq a_o)}{\Upsilon}$, and $g(a) = \frac{p(a|a > a_o)}{1-\Upsilon}$. These auxiliary variables enable the following separation of the original problem:

$$\begin{aligned} \sup_{p(a)} E_a [\Phi(a)] &= \sup_{\Upsilon} \left(\Upsilon \sup_{q(a)} E_{a \leq a_o} [\Phi(a)] + \right. \\ &\quad \left. + (1 - \Upsilon) \sup_{g(a)} E_{a > a_o} [\Phi(a)] \right). \end{aligned} \quad (27)$$

Here, we separated the original problem in three subproblems, where one is purely convex and another is purely concave. Therefore we can consider Jensen's inequality to find the optimum distributions $q(a)$ and $g(a)$, and then complete the final result by finding the optimum value of Υ and constructing the final density function as $p^*(a) = \Upsilon^* q^*(a) + (1 - \Upsilon^*) g^*(a)$. This leads to the following results:

1) If $a_o \geq \rho$, then $\Upsilon^* = 1$ and

$$p^*(a) = q^*(a) = \begin{cases} a_o & \text{w.p. } \frac{\rho}{a_o} \\ 0 & \text{otherwise} \end{cases} \quad (28)$$

producing

$$\sup_{p(a)} \mathbb{E}_a [\Phi(a)] = \rho \frac{a_o}{1 + a_o} N \left(1 - \frac{1}{L}\right) \quad (29)$$

2) If $a_o < \rho$, then $\Upsilon^* = 0$, and

$$p^*(a) = g^*(a) = \delta(a - \rho), \quad (30)$$

producing

$$\sup_{p(a)} \mathbb{E}_a [\Phi(a)] = L \log(1 + \rho N) \quad (31)$$

Since solving (26) can be difficult, to complete our analysis we obtain a simple scaling characterization of a_0 .

Lemma 9. *If $\tau \leq 1$ the solution to (26) scales as*

$$a_o = \Theta\left(\sqrt{\frac{L}{N}}\right)$$

And if $\tau > 1$ the solution to (26) scales as

$$a_o = \Theta\left(\frac{L}{N}\right)$$

Proof. By contradiction, let us define $a'_o = \Theta(N^\alpha)$ and $\Phi_2(a'_o) = \theta(N^\tau \log(N))$.

If $\alpha \leq 0$, then $\Phi_1(a'_o) = \theta(N^{2\alpha-1})$. Thus for $\alpha < \frac{\tau-1}{2}$ we have $\Phi_1(a'_o) \ll \Phi_2(a'_o)$ and if $\alpha > \frac{\tau-1}{2}$ then $\Phi_1(a'_o) \gg \Phi_2(a'_o)$. Therefore, a'_o cannot be a solution to (26) if $\alpha \neq \frac{\tau-1}{2}$, where $\alpha < 0 \Leftrightarrow \tau < 1$.

Likewise, if $\alpha > 0$, then $\Phi_1(a'_o) = \theta(N^{\alpha-1})$, and a'_o cannot be a solution to (26) if $\alpha \neq \tau - 1$, where $\alpha > 0 \Leftrightarrow \tau > 1$. \square

Equipped with the scaling of a_o , recalling that $\rho = \frac{PL}{M}$, and using (29) and (31), we get the following cases:

- For $\tau \leq 1$,

- if $M \leq \Theta(\sqrt{NL})$ we get $a_o \ll \rho$, which leads to $\frac{M}{L} \mathbb{E}_a [\Phi(a)] = \frac{M}{L} \Phi_2(\rho) = M \log(1 + \rho N)$.
- if $M \geq \Theta(\sqrt{NL})$ we get $a_o \gg \rho$, leading to $\frac{M}{L} \mathbb{E}_a [\Phi(a)] = \frac{M}{L} \frac{\rho}{a_o} \Phi_1(\rho) = \frac{M}{L} a_o \rho N \left(1 - \frac{1}{L}\right) = \Theta(\sqrt{LN})$.

Thus, for $\epsilon < \frac{1+\tau}{2}$, the optimum number of subchannels satisfies $M = B$. And for the case $\epsilon \geq \frac{1+\tau}{2}$, all choices of M satisfying $M \geq \Theta(N^{\frac{1+\tau}{2}})$ are equally optimal in a scaling sense. Although different choices of $M \geq \Theta(N^{\frac{1+\tau}{2}})$ can overspread the bandwidth, this neither increases nor decreases the rate. Increasing M above $\Theta(N^{\frac{1+\tau}{2}})$ merely increases the peakyness of the optimal distribution $p^*(a)$ (28) in such a way that the time-frequency spreading of power remains constant and the rate scaling is unchanged.

- Finally, when $\tau > 1$ the bandwidth overspreading threshold condition associated with $a_o \ll \rho$ scales as $M \leq \Theta(N)$, and the rate is $\Theta(N^{\min(\epsilon, 1)})$.

Putting everything together produces the final Shape Encoding Capacity scaling upper bound

$$C_S(B, N, L) \leq \Theta(N^{\min(\epsilon, \frac{1+\tau}{2}, 1)}).$$

APPENDIX D

DISTRIBUTION OF $p(\mathbf{Y}|\mathbf{u}, a)$

In this section we compute the probability density $p(\mathbf{Y}|\mathbf{u}, a)$. The Kroënecker product and matrix vector stacking function satisfy the relation $\text{vec}(\mathbf{h}\mathbf{u}^H) = \mathbf{u}^* \otimes \mathbf{h}$. When \mathbf{u} is a known vector, $(\text{vec}(\mathbf{h}\mathbf{u}^H)|\mathbf{u})$ is a NL -dimension Gaussian-distributed vector with a singular covariance matrix $\Sigma_{\text{vec}(\mathbf{h}\mathbf{u}^H)} = a\mathbf{u}^*\mathbf{u}^T \otimes \mathbf{I}_N$, and its p.d.f. is undefined. Fortunately, the addition of the noise term in $(\mathbf{Y}|a, \mathbf{u}) = a\mathbf{h}\mathbf{u}^H + \mathbf{Z}$ introduces independent noise terms and, therefore, $p(\mathbf{Y}|\mathbf{u}, a)$ can be written as the p.d.f. of the NL -dimension Gaussian-distributed vector $(\text{vec}(\mathbf{Y})|a, \mathbf{u}) \sim \mathcal{CN}(0, \Sigma_{\text{vec}(\mathbf{Y})|\mathbf{u}, a})$, where the covariance matrix is full rank with value $\Sigma_{\text{vec}(\mathbf{Y})|\mathbf{u}, a} = \mathbf{I}_{NL} + a\mathbf{u}^*\mathbf{u}^T \otimes \mathbf{I}_N = (\mathbf{I}_L + a\mathbf{u}^*\mathbf{u}^T) \otimes \mathbf{I}_N$.

For convenience, we write a more compact expression of the p.d.f. in (32), where (a) comes from the distributive properties of the determinant and inverse of Kroënecker products, (b) substitutes the determinant and inverse of the identity matrix, (c) applies the Sylvester rule for the determinant, and the Sherman-Morrison formula for the matrix inverse, and (d)

$$\begin{aligned} p(\mathbf{Y}|\mathbf{u}, a) &= |(\mathbf{I}_L + a\mathbf{u}^*\mathbf{u}^T) \otimes \mathbf{I}_N|^{-NL} (2\pi)^{NL} \exp \left[-\text{vec}(\mathbf{Y})^H \left((\mathbf{I}_L + a\mathbf{u}^*\mathbf{u}^T) \otimes \mathbf{I}_N \right)^{-1} \text{vec}(\mathbf{Y}) \right] \\ &\stackrel{(a)}{=} |\mathbf{I}_N|^{-N} |\mathbf{I}_L + a\mathbf{u}^*\mathbf{u}^T|^{-L} (2\pi)^{-NL} \exp \left[-\text{vec}(\mathbf{Y})^H \left((\mathbf{I}_L + a\mathbf{u}^*\mathbf{u}^T)^{-1} \otimes (\mathbf{I}_N)^{-1} \right) \text{vec}(\mathbf{Y}) \right] \\ &\stackrel{(b)}{=} |\mathbf{I}_L + a\mathbf{u}^*\mathbf{u}^T|^{-L} (2\pi)^{-NL} \exp \left[-\text{vec}(\mathbf{Y})^H \left((\mathbf{I}_L + a\mathbf{u}^*\mathbf{u}^T)^{-1} \otimes \mathbf{I}_N \right) \text{vec}(\mathbf{Y}) \right] \\ &\stackrel{(c)}{=} (1+a)^{-N} (2\pi)^{-NL} \exp \left[-\text{vec}(\mathbf{Y})^H \left(\left(\mathbf{I}_L - \frac{a}{1+a\mathbf{u}^T\mathbf{u}^*} \mathbf{u}^*\mathbf{u}^T \right) \otimes \mathbf{I}_N \right) \text{vec}(\mathbf{Y}) \right] \\ &\stackrel{(d)}{=} (1+a)^{-N} (2\pi)^{-NL} \exp \left[-\left(\|\mathbf{Y}\|^2 - \frac{a}{1+a} \|\mathbf{Y}\mathbf{u}\|^2 \right) \right] \end{aligned} \quad (32)$$

comes from $\mathbf{u}^T \mathbf{u}^* = 1$, $\text{vec}(\mathbf{Y})^H \text{vec}(\mathbf{Y}) = \|\mathbf{Y}\|^2$, and a careful reorganization of the matrix product $\text{vec}(\mathbf{Y})^H (\mathbf{u}^* \mathbf{u}^T \otimes \mathbf{I}_N) \text{vec}(\mathbf{Y}) = \text{tr}\{(\mathbf{Y}\mathbf{u})^H \mathbf{Y}\mathbf{u}\} = \|\mathbf{Y}\mathbf{u}\|^2$.

APPENDIX E LOWER BOUND OF $p(\mathbf{Y}|a)$

The p.d.f. of the channel output may be computed from the conditional as $p(\mathbf{Y}|a) = \mathbb{E}_{\mathbf{u}} [p(\mathbf{Y}|\mathbf{u}, a)]$. We note that for the particular case $L = 1$ we have that \mathbf{u} is just a random phase, producing $\|\mathbf{Y}\mathbf{u}\|^2 = \|\mathbf{Y}\|^2$, this would lead to $p(\mathbf{Y}|a)|_{L=1} = p(\mathbf{Y}|\mathbf{u}, a)|_{L=1}$ and to $\mathbf{I}(\mathbf{u}; \mathbf{Y}|a) = 0$. We therefore focus on the case $L > 1$, when $p(\mathbf{Y}|a)$ is not Gaussian distributed and we are forced to come up with a lower-bounding strategy on $p(\mathbf{Y}|a)$ in order to upper bound $\mathbf{I}(\mathbf{u}; \mathbf{Y}|a)$.

The p.d.f. of \mathbf{u} is constant in the L -dimensional sphere, with value $\Gamma(L)/\pi^L$, and thus remains unchanged when \mathbf{u} is transformed by multiplication with any unitary matrix $p(\mathbf{V}\mathbf{u}) = p(\mathbf{u})$. Particularly, we use the SVD of $\mathbf{Y} = \mathbf{U}^H \mathbf{S} \mathbf{V}$ where \mathbf{U}, \mathbf{V} are unitary, \mathbf{S} is the diagonal matrix containing the singular values of \mathbf{Y} , and $\mathbf{u}^H \mathbf{Y}^H \mathbf{Y} \mathbf{u} = \mathbf{u}^H \mathbf{V}^H \mathbf{S}^2 \mathbf{V} \mathbf{u}$. Therefore we get (33), shown on the bottom of this page.

Now, for the purpose of just upper bounding the mutual information, we apply Jensen's inequality in the exponential $\mathbb{E}_{\mathbf{u}} \left[\exp \left[\frac{a}{1+a} \|\mathbf{S}\mathbf{u}\|^2 \right] \right] \geq \exp \left[\frac{a}{1+a} \mathbb{E}_{\mathbf{u}} [\|\mathbf{S}\mathbf{u}\|^2] \right]$. With this we obtain a lower bound of the function $p(\mathbf{Y}|a)$ that is not necessarily a p.d.f.

$$p(\mathbf{Y}|a) \geq (1+a)^{-N} (2\pi)^{-NL} \exp \left[-\left(\|\mathbf{Y}\|^2 - \frac{a \mathbb{E}_{\mathbf{u}} [\|\mathbf{S}\mathbf{u}\|^2]}{1+a} \right) \right] \quad (34)$$

Since \mathbf{S} is diagonal, $\|\mathbf{S}\mathbf{u}\|^2 = \sum_{\ell=1}^L |s_{\ell} u_{\ell}|^2$, producing

$$\mathbb{E}_{\mathbf{u}} [\|\mathbf{S}\mathbf{u}\|^2] = \sum_{\ell=1}^L s_{\ell}^2 \mathbb{E}_{\mathbf{u}} [|u_{\ell}|^2] = \frac{1}{L} \sum_{\ell=1}^L s_{\ell}^2 = \frac{\|\mathbf{Y}\|^2}{L}.$$

Therefore

$$p(\mathbf{Y}|a) \geq (1+a)^{-N} (2\pi)^{-NL} \exp \left[-\|\mathbf{Y}\|^2 \left(1 - \frac{a}{1+a} \frac{1}{L} \right) \right] \quad (35)$$

REFERENCES

- [1] Cisco Systems Inc., "Cisco Annual Internet Report (2018–2023)," *White Paper*, 2020. [Online]. Available: <https://www.cisco.com/c/en/us/solutions/collateral/executive-perspectives/annual-internet-report/white-paper-c11-741490.html>
- [2] 3GPP, "3rd Generation Partnership Project; Technical Specification Group Radio Access Network; NR; NR and NG-RAN Overall Description; Stage 2 (Release 16)," *ETSI TS 38.300*, no. 16.5.0, 2021.
- [3] E. Khorov, A. Kiryanov, A. Lyakhov, and G. Bianchi, "A tutorial on IEEE 802.11ax high efficiency WLANs," *IEEE Communications Surveys Tutorials*, vol. 21, no. 1, pp. 197–216, 2019.
- [4] P. Zhou, K. Cheng, X. Han, X. Fang, Y. Fang, R. He, Y. Long, and Y. Liu, "IEEE 802.11ay-based mmWave WLANs: Design challenges and solutions," *IEEE Communications Surveys Tutorials*, vol. 20, no. 3, pp. 1654–1681, 2018.
- [5] M. Dohler, R. W. Heath, A. Lozano, C. B. Papadias, and R. A. Valenzuela, "Is the PHY layer dead?" *IEEE Communications Magazine*, vol. 49, no. 4, pp. 159–165, apr 2011.
- [6] T. L. Marzetta, "How much training is required for multiuser MIMO?" in *IEEE 40th Asilomar Conference on Signals, Systems and Computers (ACSSC)*, Pacific Grove, CA, USA, 2006, pp. 359–363.
- [7] M. Médard and R. G. Gallager, "Bandwidth scaling for fading multipath channels," *IEEE Transactions on Information Theory*, vol. 48, no. 4, pp. 840–852, 2002.
- [8] R. K. Mallik and R. Murch, "Channel capacity of an asymmetric constellation in Rayleigh fading with noncoherent energy detection," *IEEE Transactions on Wireless Communications*, vol. 20, no. 11, pp. 7375–7388, 2021.
- [9] A. Goldsmith, *Wireless communications*. Cambridge University Press, 2005.
- [10] L. Zheng and D. N. C. Tse, "Sphere packing in the Grassmann manifold: A geometric approach to the noncoherent multi-antenna channel," in *2000 IEEE International Symposium on Information Theory*, 2000, pp. 364–364.
- [11] L. Zheng and D. Tse, "Optimal diversity-multiplexing trade-off in multi-antenna channels," in *Proc. of the 39th Allerton Conference on Communication, Control and Computing*, 2001, pp. 835–844.
- [12] L. Zheng and D. N. C. Tse, "Communication on the Grassmann Manifold: A Geometric Approach to the Noncoherent Multiple-antenna Channel," *IEEE Transactions on Information Theory*, vol. 48, no. 2, pp. 359–383, 2002.
- [13] —, "The diversity-multiplexing tradeoff for non-coherent multiple antenna channels," in *Proceedings Of The 40th Annual Allerton Conference On Communication Control And Computing*, no. 2, 2002, pp. 834–843.
- [14] İ. E. Telatar and D. N. Tse, "Capacity and mutual information of wideband multipath fading channels," *IEEE Transactions on Information Theory*, vol. 46, no. 4, pp. 1384–1400, 2000.
- [15] S. Verdú, "Spectral efficiency in the wideband regime," *IEEE Transactions on Information Theory*, vol. 48, no. 6, pp. 1319–1343, 2002.
- [16] M. Medard, "On approaching wideband capacity using multitone FSK," *IEEE Journal on Selected Areas in Communications*, vol. 23, no. 9, pp. 1830–1838, 2005.
- [17] L. Zheng, D. N. C. Tse, and M. Medard, "Channel coherence in the low-SNR regime," *IEEE Transactions on Information Theory*, vol. 53, no. 3, pp. 976–997, 2007.
- [18] S. Ray, M. Medard, and L. Zheng, "On noncoherent MIMO channels in the wideband regime: capacity and reliability," *IEEE Transactions on Information Theory*, vol. 53, no. 6, pp. 1983–2009, 2007.

$$\begin{aligned}
 p(\mathbf{Y}|a) &= \int_{|\mathbf{u}|^2=1} p(\mathbf{Y}|\mathbf{u}, a) p(\mathbf{u}) d\mathbf{u} \\
 &= \int_{|\mathbf{u}|^2=1} (1+a)^{-N} (2\pi)^{-NL} \exp \left[-\left(\|\mathbf{Y}\|^2 - \frac{a}{1+a} |\mathbf{Y}\mathbf{u}|^2 \right) \right] \frac{\Gamma(L)}{\pi^L} d\mathbf{u} \\
 &= (1+a)^{-N} (2\pi)^{-NL} \exp [-|\mathbf{Y}|^2] \int_{|\mathbf{u}|^2=1} \frac{\Gamma(L)}{\pi^L} \exp \left[\frac{a}{1+a} |\mathbf{Y}\mathbf{u}|^2 \right] d\mathbf{u} \\
 &= (1+a)^{-N} (2\pi)^{-NL} \exp [-|\mathbf{Y}|^2] \mathbb{E}_{\mathbf{u}} \left[\exp \left[\frac{a}{1+a} |\mathbf{S}\mathbf{u}|^2 \right] \right]
 \end{aligned} \quad (33)$$

- [19] V. Sethuraman, L. Wang, B. Hajek, and A. Lapidath, "Low-SNR capacity of noncoherent fading channels," *IEEE Transactions on Information Theory*, vol. 55, no. 4, pp. 1555–1574, 2009.
- [20] A. Lozano and D. Porrat, "Non-peaky signals in wideband fading channels: Achievable bit rates and optimal bandwidth," *IEEE Transactions on Wireless Communications*, vol. 11, no. 1, pp. 246–257, 2012.
- [21] F. Gómez-Cuba, J. Du, M. Médard, and E. Erkip, "Unified capacity limit of non-coherent wideband fading channels," *IEEE Transactions on Wireless Communications*, vol. 16, no. 1, pp. 43–57, jan 2017.
- [22] T. L. Marzetta and B. M. Hochwald, "Capacity of a mobile multiple-antenna communication link in Rayleigh flat fading," *IEEE Transactions on Information Theory*, vol. 45, no. 1, pp. 139–157, 1999.
- [23] R. R. Perera, T. S. Pollock, and T. D. Abhayapala, "Non-coherent Rayleigh fading MIMO channels: Capacity and optimal input," in *IEEE International Conference on Communications (ICC)*, 2006, pp. 4180–4185.
- [24] M. Chowdhury, A. Manolakos, and A. J. Goldsmith, "Design and performance of non coherent massive SIMO systems," in *IEEE Annual Conference on Information Sciences and Systems (CISS)*, 2014, pp. 1–6.
- [25] M. Chowdhury and A. Goldsmith, "Capacity of block Rayleigh fading channels without CSI," in *2016 IEEE International Symposium on Information Theory (ISIT)*, 2016, pp. 1884–1888.
- [26] M. Katz and S. Shamai, "On the capacity-achieving distribution of the discrete-time noncoherent and partially coherent AWGN channels," *IEEE Transactions on Information Theory*, vol. 50, no. 10, pp. 2257–2270, 2004.
- [27] F. Gómez-Cuba, M. Chowdhury, A. Manolakos, E. Erkip, and A. J. Goldsmith, "Capacity scaling in a non-coherent wideband massive SIMO block fading channel," *IEEE Transactions on Wireless Communications*, vol. 18, no. 12, pp. 5691–5704, 2019.
- [28] B. Hassibi and B. M. Hochwald, "How much training is needed in multiple-antenna wireless links?" *IEEE Transactions on Information Theory*, vol. 49, no. 4, pp. 951–963, 2003.
- [29] A. Lozano, "Interplay of spectral efficiency, power and Doppler spectrum for reference-signal-assisted wireless communication," *IEEE Transactions on Wireless Communications*, vol. 7, no. 12, pp. 5020–5029, 2008.
- [30] N. Jindal and A. Lozano, "A unified treatment of optimum pilot overhead in multipath fading channels," *IEEE Transactions on Communications*, vol. 58, no. 10, pp. 2939–2948, 2010.
- [31] J. Du and R. A. Valenzuela, "How much spectrum is too much in millimeter wave wireless access," *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 7, pp. 1444 – 1458, 2017.
- [32] G. C. Ferrante, T. Q. S. Quek, and M. Z. Win, "Revisiting the capacity of noncoherent fading channels in mmWave system," *IEEE International Conference on Communications, ICC*, vol. 65, no. 8, pp. 3259–3275, 2016.
- [33] E. Björnson, E. G. Larsson, and T. L. Marzetta, "Massive MIMO: ten myths and one critical question," *IEEE Communications Magazine*, vol. 54, no. 2, pp. 114–123, feb 2016.
- [34] E. Björnson, J. Hoydis, and L. Sanguinetti, "Massive MIMO has unlimited capacity," *IEEE Transactions on Wireless Communications*, vol. 17, no. 1, pp. 574–590, 2018.
- [35] T. L. Marzetta, G. Caire, M. Debbah, I. Chih-Lin, and S. K. Mohammed, "Special issue on massive MIMO," *Journal of Communications and Networks*, vol. 15, no. 4, pp. 333–337, 2013.
- [36] A. M. Tulino and S. Verdú, "Random matrix theory and wireless communications." *Foundations and Trends in Communications and Information Theory*, vol. 1, no. 1, 2004.
- [37] D. E. Knuth, "Big Omicron and big Omega and big Theta," *ACM SIGACT News*, vol. 8, no. 2, pp. 18–24, 1976.
- [38] A. Saleh and R. Valenzuela, "A statistical model for indoor multipath propagation," *IEEE Journal on Selected Areas in Communications*, vol. 5, no. 2, pp. 128–137, feb 1987.
- [39] F. Gómez-Cuba and A. J. Goldsmith, "Compressed sensing channel estimation for OFDM with non-Gaussian multipath gains," *IEEE Transactions on Wireless Communications*, vol. 19, no. 1, pp. 47–61, 2020.
- [40] M. K. Samimi and T. S. Rappaport, "3-D Millimeter-wave statistical channel model for 5G wireless system design," *IEEE Transactions on Microwave Theory and Techniques*, vol. 64, no. 7, pp. 2207–2225, jun 2016.
- [41] 3GPP, "Study on channel model for frequencies from 0.5 to 100 GHz," TR 38.901, v15.0.0.
- [42] A. J. Coulson, A. G. Williamson, and R. G. Vaughan, "A statistical basis for lognormal shadowing effects in multipath fading channels," *IEEE Transactions on Communications*, vol. 46, no. 4, pp. 494–502, apr 1998.
- [43] S. E. Foo, M. A. Beach, and A. G. Burr, "Wideband outdoor MIMO channel model derived from directional channel measurements at 2 GHz," in *7th International Symposium on Wireless Personal Multimedia Communications (WPMC)*, sep 2004.
- [44] J. Mo, P. Schniter, and R. W. Heath, "Channel estimation in broadband millimeter wave MIMO systems with few-bit ADCs," *IEEE Transactions on Signal Processing*, vol. 66, no. 5, pp. 1141–1154, dec 2018.
- [45] G. C. Ferrante, T. Q. S. Quek, and M. Z. Win, "Revisiting the capacity of noncoherent fading channels in mmWave system," *IEEE Transactions on Communications*, vol. 65, no. 8, pp. 3259–3275, 2017.
- [46] T. L. Marzetta, "Spatially-stationary propagating random field model for massive MIMO small-scale fading," in *IEEE International Symposium on Information Theory (ISIT)*, 2018, pp. 391–395.
- [47] A. S. Y. Poon, R. W. Brodersen, and D. N. C. Tse, "Degrees of freedom in multiple-antenna channels: a signal space approach," *IEEE Transactions on Information Theory*, vol. 51, no. 2, pp. 523–536, 2005.
- [48] A. Ozgur, O. Lévêque, and D. Tse, "Hierarchical cooperation achieves linear capacity scaling in ad hoc networks," *IEEE Transactions on Information Theory*, vol. 53, no. 10, pp. 3549–3572, 2007.
- [49] V. Raghavan, G. Hariharan, and A. M. Sayeed, "Capacity of sparse multipath channels in the ultra-wideband regime," *IEEE Journal of Selected Topics in Signal Processing*, vol. 1, no. 3, pp. 357–371, oct 2007.



Felipe Gómez-Cuba [M'11] received the Ingeniero de Telecomunicación degree in 2010, M.Sc in Signal Processing Applications for Communications in 2012, and a PhD degree in 2015 from the University of Vigo, Spain. He worked as a researcher in the Information Technologies Group (GTI), University of Vigo, (2010–2011), the Galician Research and Development center In Advanced Telecommunications (GRADIANT), (2011–2013), the NYUWireless center at NYU Tandon School of Engineering (2013–2014) and in University of Vigo with the FPU grant from the Spanish MINECO (2013–2016). He received a Marie Curie Individual Fellowship - Global Fellowship with the Dipartimento d'Ingegneria dell'Informazione, University of Padova, Italy (2016–2019) and he was a postdoctoral scholar with the Department of Electrical Engineering, Stanford University, USA (2016–2018). He has been awarded a Distinguished Researcher position with the Beatriz Galindo fellowship in the Departamento de Tecnología Electrónica y de las Comunicaciones (TSC), University of Vigo (2020–2024). His main research interests are new paradigms in wireless networks such as cooperative communications, ultra high density cellular systems, wideband communications and massive MIMO.