Capacity bounds for MIMO microwave backhaul links affected by phase noise

Giuseppe Durisi, *Senior Member, IEEE,* Alberto Tarable, *Member, IEEE,* Christian Camarda, *Student Member, IEEE,* Rahul Devassy, Guido Montorsi, *Senior Member, IEEE*

Abstract—**We present bounds and a closed-form high-SNR expression for the capacity of multiple-antenna systems affected by Wiener phase noise. Our results are developed for the scenario where a single oscillator drives all the radio-frequency circuitries at each transceiver (common oscillator setup), the input signal is subject to a peak-power constraint, and the channel matrix is deterministic. This scenario is relevant for line-of-sight multiple-antenna microwave backhaul links with sufficiently small antenna spacing at the transceivers. For the 2×2 multiple-antenna case, for a Wiener phase-noise process with standard deviation equal to $6°$, and at the medium/high SNR values at which microwave backhaul links operate, the upper bound reported in the paper exhibits a $3 \, \mathrm{dB}$ gap from a lower bound obtained using 64-QAM. Furthermore, in this SNR regime the closed-form high-SNR expression is shown to be accurate.**

I. INTRODUCTION

A cost-effective solution to the problem of guaranteeing backhaul connectivity in mobile cellular networks is the use of point-to-point microwave links in the Q and E bands [1], [2]. The current terrific rate of increase in mobile data traffic makes these microwave radio links a potential bottleneck in the deployment of high-throughput cellular networks. This consideration has stimulated a large body of research aimed at the design of high-capacity backhaul links [3]–[6]. One design challenge is that the use of high-order constellations to increase throughput (512 QAM has been recently demonstrated in commercial products) makes the overall system extremely sensitive to *phase noise*, i.e., to phase and frequency instabilities in the radio-frequency (RF) oscillators used at the transmitter and the receiver.

A fundamental way to characterize the impact of phase noise on the throughput of microwave backhaul links is to study their Shannon capacity. Unfortunately, the capacity of the phase-noise channel is not known in closed form even for simple channel models, although capacity bounds and asymptotic results in the limiting regime of high SNR have been reported in the literature. Lapidoth characterized the capacity of the general class of stationary phase-noise channels (the widely used Wiener model [7] belongs to this class) in the high-SNR regime [8]. Specifically, he showed that whenever the phase-noise process has finite differential-entropy rate, the high-SNR capacity is equal to half the capacity of an AWGN channel with the same SNR, plus a correction term that accounts for the memory in the phase-noise process. This result has been recently extended to the *waveform* phase-noise channel in [9], [10]. The high-SNR capacity of the block-memoryless phase-noise channel (a non-stationary channel) has been characterized in [11], [12].

For the non-asymptotic regime of finite SNR, only capacity bounds are available in the literature. Katz and Shamai [13] provided tight upper and lower bounds on the capacity of the memoryless phase-noise channel. They also established that the capacity-achieving distribution is discrete with an infinite number of mass points. Some of the bounds reported in [13] have been extended to the block-memoryless phase-noise case in [11], [12]. For the Wiener phase-noise model, an upper bound on the rates achievable with PSK constellations has been proposed in [14]. Capacity lower bounds obtained by numerically computing the information rates achievable with various families of finite-cardinality independent and identically distributed (i.i.d.) input processes (e.g., QAM, PSK, and APSK constellations) have been reported e.g., in [14], [7], [15]. The numerical evaluation of these bounds is often based on the algorithm for the computation of the information rates for finite-state channels proposed in [16] (see [17] for different approaches based, e.g., on particle filtering). Although the effect of phase noise in the *measurement* of MIMO channels has been extensively investigated in the literature (see, e.g., [18]), little is known about the impact of phase noise on the MIMO capacity. In [6], it was shown that different RF circuitries configurations (e.g., independent oscillators at each antenna as opposed to a single oscillator driving all antennas) yield different spatial multiplexing gains. The impact of phase noise on the achievable rates of massive MIMO systems has been recently discussed in [19], [20]. However, the observations reported in [19], [20] are based only on capacity lower bounds and are not conclusive.

Contributions

We study the capacity of multiple-antenna systems affected by phase noise. Specifically, we consider the scenario where a single oscillator drives all RF circuitries at each transceiver. We present a non-asymptotic capacity upper bound for the case of Wiener phase noise and the practically relevant scenario when the transmit codewords are subject to a peak-power constraint, which is more stringent than the average-power constraint analyzed so

G. Durisi and R. Devassy are with the Department of Signals and Systems, Chalmers University of Technology, Gothenburg, Sweden (e-mail: {durisi,devassy}@chalmers.se).

A. Tarable is with IEIIT-CNR, Turin, Italy (e-mail: alberto.tarable@ieiit.cnr.it).

C. Camarda and G. Montorsi are with the Department of Electronics and Telecommunications, Politecnico di Torino, Turin, Italy (e-mail: ({christian.camarda,guido.montorsi}@polito.it)

far in the phase-noise literature. This upper bound improves on the one recently reported in [21], which was derived under the assumption of codewords subject to an average-power constraint. When particularized to constant-modulus constellations and to single-antenna systems, our bound recovers the upper bound obtained in [14]. We compare our upper bound with lower bounds obtained by evaluating numerically the information rates achievable with QAM constellations. For the case of a Wiener phase-noise process with standard deviation of the phase increments equal to $6°$, the gap between our upper bound and the information rates achievable with 64-QAM is about $3\,\mathrm{dB}$ for medium/high SNR values.

We also provide a capacity characterization in the high-SNR regime that is accurate up to a term that vanishes as SNR grow large. This characterization yields a capacity approximation that turns out to be accurate already at moderate SNR values.

II. SYSTEM MODEL

A. The input-output relation

We consider the following $M \times M$ MIMO phase-noise channel with memory

$$\mathbf{y}_k = e^{j\theta_k}\mathbf{H}\mathbf{x}_k + \mathbf{w}_k, \quad k = 1, 2, \ldots, n. \quad (1)$$

Here, $\mathbf{x}_k$ denotes the M-dimensional input vector at discrete time k; $\mathbf{H}$ is the MIMO channel matrix, which we assume deterministic, full-rank, and known to the transmitter and the receiver; $\{\theta_k\}$ is the phase-noise process; and $\{\mathbf{w}_k\}$ is the additive Gaussian noise, which we assume independent and identically distributed (i.i.d.) circularly symmetric with zero mean and covariance matrix $\mathbf{I}_M$, i.e., $\mathbf{w}_k \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_M)$. The input-output relation (1) describes accurately line-of-sight (LoS) MIMO backhaul links in which the antenna spacing at the transceivers is sufficiently small for the RF circuitries at each antenna to be driven by the same oscillator [6]. We elaborate further on the accuracy of the model (1) in Section II-B.

We assume that the phase-noise samples $\{\theta_k\}$ form a Wiener process [7], i.e.,[1]

$$\theta_{k+1} = \theta_k + \Delta_k \quad (2)$$

where $\{\Delta_k\}$ is a sequence of i.i.d. zero-mean Gaussian random variables with variance σ_Δ^2, i.e., $\Delta_k \sim \mathcal{N}(0, \sigma_\Delta^2)$ and the sum in (2) is modulo 2π. The i.i.d. assumption on $\{\Delta_k\}$ implies that $\{\theta_k\}$ is a Markov process. Specifically,

$$f_{\theta_k \mid \theta_{k-1}, \ldots, \theta_0} = f_{\theta_k \mid \theta_{k-1}} = f_\Delta \quad (3)$$

where

$$f_\Delta(\delta) \triangleq \sum_{l=-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma_\Delta^2}} \exp\left(-\frac{(\delta - 2\pi l)^2}{2\sigma_\Delta^2}\right), \quad \delta \in [0, 2\pi]. \quad (4)$$

In words, f_Δ is the probability density function (pdf) of the innovation Δ_k modulo 2π.

Under the additional assumption that the initial phase θ_0 is uniformly distributed in the interval $[0, 2\pi]$, i.e., $\theta_0 \sim \mathcal{U}[0, 2\pi]$, the process $\{\theta_k\}$ is stationary. Let $\Delta \sim f_\Delta$ (defined in (4)). The

[1]See [9] for a discussion on the limitations of this model.

differential entropy rate of a stationary Wiener process is then given by

$$h(\{\theta_k\}) = h(\Delta). \quad (5)$$

B. LoS MIMO

The idea behind LoS MIMO is to achieve a full-rank channel matrix $\mathbf{H}$ over a LoS link by a careful placement of the antennas at the transceivers [22]–[24]. Indeed, when the antenna spacing d at the transmitter and the receiver satisfies

$$d \approx \sqrt{\lambda R/M} \quad (6)$$

where λ is the wavelength and R denotes the distance between the transmitter and the receiver, the channel matrix $\mathbf{H}$ can be made not only full-rank, but also *unitary* [23], [24].

We next discuss some implications of (6) on the design of microwave backhaul links. Consider a microwave backhaul link operating in the E band at 80 GHz. Assume that the transceivers are equipped with 2 antennas each and are 500 m apart. According to (6), the antenna spacing that results in a unitary channel matrix is about 97 cm, which is compatible with the assumption of using a single oscillator to drive the RF circuitries of both antennas. In some cases, it may be convenient to locate the two antennas closer than what (6) prescribes. Then, $\mathbf{H}$ ceases to be unitary, although it can still be made full rank [24]. For a microwave backhaul link operating at 20 GHz over a 3 Km link, (6) results in an antenna spacing of about 3.8 m, which calls for a distributed oscillator solution.

As pointed out in Section II-A, in this paper we will focus exclusively on the single-oscillator scenario. We will consider both the case of $\mathbf{H}$ unitary (our results will be somewhat sharper under this assumption), and the more general case of $\mathbf{H}$ full-rank but not necessarily unitary. The distributed oscillator setup will not be analyzed. Preliminary results concerning this scenario can be found in [6], [19], [20].

C. Peak-amplitude constraint

The results currently available on the capacity of phase noise channels [8], [13], [11], [12], [21] were derived under the assumption that each transmit codeword $(\mathbf{x}_1, \ldots, \mathbf{x}_n)$ is subject to the average-power constraint [25, Eq. (9.2)]

$$\frac{1}{n}\sum_{k=1}^{n}\|\mathbf{x}_k\|^2 \leq \rho. \quad (7)$$

In practice, each codeword entry $\mathbf{x}_k$ must obey a given *peak-power constraint* to avoid distortions due to nonlinearities and saturation effects at the high-power amplifier [26]. To account for this and obtain capacity results that are more relevant in practice, in this paper we substitute (7) with the more stringent peak-power constraint

$$\|\mathbf{x}_k\|^2 \leq \rho, \quad k = 1, \ldots, n. \quad (8)$$

The peak-power constraint (8) has been considered previously in the information-theoretic literature, but not in the contest of phase-noise channels. Smith [27] proved that the capacity-achieving distribution of an AWGN channel subject to (8) is discrete with a finite number of mass point (in contrast, the

capacity-achieving distribution under (7) [25, Ch. 9] is Gaussian). More recently, Lapidoth [28] characterized the high-SNR capacity of single-antenna stationary fading channels subject to (8) in the setting where no *a priori* channel-state information is available at the receiver.

D. Channel Capacity

We are interested in computing the capacity of the MIMO phase-noise channel (1), which—under the peak-amplitude constraint (8)—is given by

$$C(\rho) = \lim_{n\to\infty} \frac{1}{n} \sup I(\mathbf{y}^n; \mathbf{x}^n) \qquad (9)$$

where $\mathbf{x}^n = (\mathbf{x}_1, \ldots, \mathbf{x}_n)$ and, similarly, $\mathbf{y}^n = (\mathbf{y}_1, \ldots, \mathbf{y}_n)$. Here, the supremum is over all probability distributions on $\mathbf{x}^n$ that satisfy (8) with probability one (w.p.1). In Section III, we analyze $C(\rho)$ for the case of $\mathbf{H}$ being unitary. The general full-rank case will be discussed in Section IV.

III. THE UNITARY CASE

A. Capacity Upper Bound

We next present an upper bound on $C(\rho)$ that improves on the one reported in [21] for the average-power constrained case. With some minor adjustments, the bound turns out to be tight in the high-SNR regime (see Section III-B).

Before presenting our upper bound, two observations are in order.

i) As $\mathbf{H}$ is known to transmitter and receiver, $C(\rho)$ depends on $\mathbf{H}$ only through its singular values. Since $\mathbf{H}$ is unitary, all singular values are equal to one. Hence, we can (and will) assume without loss of generality that $\mathbf{H} = \mathbf{I}_M$.

ii) In the following proposition, we establish that the capacity-achieving input process $\{\mathbf{x}_k\}$ can be assumed isotropically distributed, a property that will be useful in our analysis.

Proposition 1: The input process $\{\mathbf{x}_k\}$ that achieves the capacity of the channel (1) when $\mathbf{H}$ is unitary can be assumed isotropically distributed. Specifically, if $\{\mathbf{x}_k\}$ achieves $C(\rho)$ in (9), then $\{\mathbf{U}_k\mathbf{x}_k\}$, where the matrix-valued random process $\{\mathbf{U}_k\}$ is i.i.d. and each $\mathbf{U}_k$ is uniformly distributed on the set of $M \times M$ unitary matrices, achieves $C(\rho)$ as well.

Proof: The proof, which exploits that $\mathbf{U}_k\mathbf{w}_k \sim \mathbf{w}_k$, follows the same steps as the proof of [29, Prop. 7]. ∎

Our upper bound on $C(\rho)$ is constructed by extending to the MIMO case the method used in [8] to derive an asymptotic bound on the capacity of stationary single-antenna phase-noise channels. We also use the approach proposed in [13], [12] to make the bound non-asymptotic, and some of the tools developed in [30] to account for the presence of the peak-power constraint (8).

For convenience, we introduce the following notation: for every $a > 0$, we let

$$\phi_l(a) \triangleq \underline{/1 + z_l/\sqrt{a}} \qquad (10)$$

where the random variables $\{z_l\}_{l\in\mathbb{Z}}$ are i.i.d. $\mathcal{CN}(0,1)$-distributed and $\underline{/x}$ denotes the phase of the complex number x. Roughly speaking, $\phi_l(a)$ is the noise level in the estimation of the phase-noise sample θ_l from the channel output $\mathbf{y}_l$ given that the input vector $\mathbf{x}_l$ is known and $\|\mathbf{x}_l\|^2 = a$.

Theorem 2: The capacity of the channel (1) under the peak-power constraint (8) can be upper-bounded as $C(\rho) \leq U(\rho)$, where[2]

$$U(\rho) \triangleq \min_{\alpha > 0} \left\{ \alpha \log \frac{\rho + M}{\alpha} + d_\alpha + \log(2\pi) \right.$$
$$\left. + \max_{0 \leq \xi \leq \sqrt{\rho}} g_\alpha(\xi, \rho) \right\}. \qquad (11)$$

Here,

$$g_\alpha(\xi, \rho) \triangleq (M - \alpha) \mathbb{E}\left[\log\left(|\xi + z_1|^2 + \sum_{j=2}^M |z_j|^2 \right) \right]$$
$$+ \alpha \frac{\xi^2 + M}{\rho + M} - h(|\xi + z_0|^2)$$
$$- h\left(\theta_0 + \phi_0(\xi^2) \mid \{\theta_l + \phi_l(\rho)\}_{l=-\infty}^{-1}, |\xi + z_0| \right) \qquad (12)$$

where $\{z_l\}$ are i.i.d. $\mathcal{CN}(0,1)$-distributed random variables and

$$d_\alpha \triangleq \log \frac{\Gamma(\alpha)}{\Gamma(M)} - M + 1 \qquad (13)$$

with $\Gamma(\cdot)$ standing for the Gamma function.

Proof: Because of Proposition 1, we can restrict ourselves to isotropically distributed input processes. Specifically, we consider $\{\mathbf{x}_k\}$ of the form $\{\mathbf{x}_k = s_k\mathbf{v}_k\}$, where $s_k = \|\mathbf{x}_k\|$ and $\mathbf{v}_k = \mathbf{x}_k/s_k$, with $\mathbf{v}_k$ uniformly distributed on the unit sphere in $\mathbb{C}^M$ and independent of s_k. We start by using chain rule as follows

$$I(\mathbf{y}^n; \mathbf{x}^n) = \sum_{k=1}^n I(\mathbf{y}_k; \mathbf{x}^n \mid \mathbf{y}^{k-1}). \qquad (14)$$

By proceeding similarly to [21, Eq. (10)], but accounting for the peak-power constraint,[3] we next upper-bound each term on the right-hand side (RHS) of (14). We first note that

$$I(\mathbf{y}_k; \mathbf{x}^n \mid \mathbf{y}^{k-1}) = h(\mathbf{y}_k \mid \mathbf{y}^{k-1}) - h(\mathbf{y}_k \mid \mathbf{y}^{k-1}, \mathbf{x}^n) \quad (15)$$
$$\leq h(\mathbf{y}_k) - h(\mathbf{y}_k \mid \mathbf{y}^{k-1}, \mathbf{x}^n) \qquad (16)$$
$$= h(\mathbf{y}_k) - h(\mathbf{y}_k \mid \mathbf{y}^{k-1}, \mathbf{x}^{k-1}, \mathbf{x}_k). \quad (17)$$

Here, in (16) we used that conditioning reduces entropy, and in (17) that $\mathbf{y}_k$ and $(\mathbf{x}_{k+1}, \ldots, \mathbf{x}_n)$ are conditionally independent given $(\mathbf{y}^{k-1}, \mathbf{x}^k)$. We next focus on the conditional differential entropy (the second term) on the RHS of (17). Intuitively, the past inputs $\mathbf{x}^{k-1}$ and the past outputs $\mathbf{y}^{k-1}$ can be used to obtain noisy estimates of the past phase-noise samples $\{\theta_l\}_{l=1}^{k-1}$. These estimates help us to guess the value of the current phase-noise sample θ_k. We next use this intuition to obtain a lower bound on $h(\mathbf{y}_k \mid \mathbf{y}^{k-1}, \mathbf{x}^{k-1}, \mathbf{x}_k)$, and, hence, an upper bound on $I(\mathbf{y}_k; \mathbf{x}^n \mid \mathbf{y}^{k-1})$ in (14). For each pair $(\mathbf{y}_l, \mathbf{x}_l)$, $l = 1, \ldots, k-1$, we compute the phase of the projection of $\mathbf{y}_l$ onto $\mathbf{x}_l$. This projection is distributed as $\theta_l + \phi_l(s_l^2)$. Since $\mathbf{y}_k$ and $\mathbf{y}^{k-1}$ are

---

[2]Throughout the paper, log stands for the natural logarithm.
[3]Recall the the upper bound developed in [21] holds for the average-power constraint case.

conditionally independent given both $\mathbf{x}^k$ and $\{\theta_l + \phi_l(s_l^2)\}_{l=1}^{k-1}$, we obtain

$$h(\mathbf{y}_k \,|\, \mathbf{y}^{k-1}, \mathbf{x}^{k-1}, \mathbf{x}_k) \tag{18}$$

$$= h\Big(\mathbf{y}_k \,\big|\, \{\theta_l + \phi_l(s_l^2)\}_{l=1}^{k-1}, \mathbf{x}^{k-1}, \mathbf{x}_k\Big) \tag{19}$$

$$\geq h\Big(\mathbf{y}_k \,\big|\, \{\theta_l + \phi_l(\rho)\}_{l=1}^{k-1}, \mathbf{x}_k\Big). \tag{20}$$

In the last step, we used that the best noisy estimate of the past phase-noise samples $\{\theta_l\}_{l=1}^{k-1}$ is achieved by transmitting inputs at peak power, i.e., $s_l^2 = \rho$, $l = 1, \ldots, k-1$. Substituting (20) into (17), we obtain

$$I(\mathbf{y}_k; \mathbf{x}^n \,|\, \mathbf{y}^{k-1}) \leq h(\mathbf{y}_k)$$
$$- h\Big(\mathbf{y}_k \,\big|\, \{\theta_l + \phi_l(\rho)\}_{l=1}^{k-1}, \mathbf{x}_k\Big) \tag{21}$$

$$= I\Big(\mathbf{y}_k; \{\theta_l + \phi_l(\rho)\}_{l=1}^{k-1}, \mathbf{x}_k\Big) \tag{22}$$

$$= I\Big(\mathbf{y}_0; \{\theta_l + \phi_l(\rho)\}_{l=-(k-1)}^{-1}, \mathbf{x}_0\Big) \tag{23}$$

$$\leq I\Big(\mathbf{y}_0; \{\theta_l + \phi_l(\rho)\}_{l=-\infty}^{-1}, \mathbf{x}_0\Big) \tag{24}$$

$$= I(\mathbf{y}_0; \mathbf{x}_0)$$
$$+ I\Big(\mathbf{y}_0; \{\theta_l + \phi_l(\rho)\}_{l=-\infty}^{-1} \,\big|\, \mathbf{x}_0\Big). \tag{25}$$

Here, (23) follows because $\{\theta_k\}$ is a stationary process. Substituting (25) into (14) and then (14) into (9), we obtain

$$C(\rho) \leq \sup\Big\{ I(\mathbf{y}_0; \mathbf{x}_0)$$
$$+ I\Big(\mathbf{y}_0; \{\theta_l + \phi_l(\rho)\}_{l=-\infty}^{-1} \,\big|\, \mathbf{x}_0\Big)\Big\}. \tag{26}$$

The supremum in (26) is over all probability distributions on $\mathbf{x}_0 = s_0 \mathbf{v}_0$ such that s_0 and $\mathbf{v}_0$ are independent, $\mathbf{v}_0$ is uniformly distributed on the unit sphere in $\mathbb{C}^M$, and $s_0^2 \leq \rho$ w.p.1.

We next upper-bound the first term on the RHS of (26), which corresponds to the mutual information of a memoryless phase-noise channel with uniform phase noise using a method similar to the one used in [12], [13]. Specifically, we use the duality approach [31, Th. 5.1] and choose an output probability distribution for which $\mathbf{y}_0$ is isotropically distributed and $r = \|\mathbf{y}_0\|^2$ follows a Gamma distribution with parameters α to be optimized later and $\beta \triangleq (\rho + M)/\alpha$. To summarize the probability density function (pdf) of r is given by

$$q_r(r) = \frac{r^{\alpha-1} e^{-r/\beta}}{\beta^\alpha \Gamma(\alpha)}. \tag{27}$$

This output distribution is optimal at high SNR (i.e., it achieves capacity up to a term that vanishes as SNR grows large) for the average-power constraint case [21]. However, it is not optimal for the peak-power constraint case, as we shall discuss in the Appendix. Nevertheless, it leads to a bound that is accurate for medium SNR values (see Section V).

Using (27), we upper-bound $I(\mathbf{x}_0; \mathbf{y}_0)$ as follows (see [21]):

$$I(\mathbf{y}_0; \mathbf{x}_0) \leq \alpha \log \frac{\rho + M}{\alpha} + d_\alpha$$
$$+ (M - \alpha)\,\mathbb{E}\left[\log\left(|s_0 + z_1|^2 + \sum_{j=2}^{M} |z_j|^2\right)\right]$$
$$- h\Big(|s_0 + z_0|^2 \,\big|\, s_0\Big) + \alpha \frac{\mathbb{E}[s_0^2] + M}{\rho + M}. \tag{28}$$

Here, d_α is the constant defined in (13) and $z_0, z_1, \ldots, z_M$ are i.i.d. $\mathcal{CN}(0,1)$-distributed random variables.

The second term on the RHS of (25) can be evaluated as follows:

$$I\Big(\mathbf{y}_0; \{\theta_l + \phi_l(\rho)\}_{l=-\infty}^{-1} \,\big|\, \mathbf{x}_0\Big)$$

$$= I\Big(e^{j\theta_0} s_0 + z_0; \{\theta_l + \phi_l(\rho)\}_{l=-\infty}^{-1} \,\big|\, s_0\Big) \tag{29}$$

$$= I\Big(e^{j\theta_0}(s_0 + z_0); \{\theta_l + \phi_l(\rho)\}_{l=-\infty}^{-1} \,\big|\, s_0\Big) \tag{30}$$

$$= I\Big(|s_0 + z_0|, \theta_0 + \phi_0(s_0^2); \{\theta_l + \phi_l(\rho)\}_{l=-\infty}^{-1} \,\big|\, s_0\Big) \tag{31}$$

$$= I\Big(\theta_0 + \phi_0(s_0^2); \{\theta_l + \phi_l(\rho)\}_{l=-\infty}^{-1} \,\big|\, |s_0 + z_0|, s_0\Big) \tag{32}$$

$$= h\big(\theta_0 + \phi_0(s_0^2) \,\big|\, |s_0 + z_0|, s_0\big)$$
$$- h\Big(\theta_0 + \phi_0(s_0^2) \,\big|\, \{\theta_l + \phi_l(\rho)\}_{l=-\infty}^{-1}, |s_0 + z_0|, s_0\Big) \tag{33}$$

$$= \log(2\pi)$$
$$- h\Big(\theta_0 + \phi_0(s_0^2) \,\big|\, \{\theta_l + \phi_l(\rho)\}_{l=-\infty}^{-1}, |s_0 + z_0|, s_0\Big). \tag{34}$$

Here, (29) follows because $\mathbf{v}_0^H \mathbf{y}_0 \sim e^{j\theta_0} s_0 + z_0$ is a sufficient statistics for $\{\theta_l + \phi_l(\rho)\}_{l=-\infty}^{-1}$; (30) follows because z_0 is circularly symmetric; (32) holds because $|s_0 + z_0|$ and $\{\theta_l + \phi_l(\rho)\}_{l=-\infty}^{-1}$ are independent; finally, (34) holds because $\theta_0 \sim \mathcal{U}[0, 2\pi]$.

We substitute (28) and (34) into (26), upper-bound the supremum over all probability distributions on s_0 satisfying $s_0 \leq \sqrt{\rho}$ w.p.1 with the supremum over all deterministic $\xi \in [0, \sqrt{\rho}]$, and tighten the resulting bound by minimizing it over the optimization parameter $\alpha > 0$. This concludes the proof. ∎

Remarks: The coarser upper bound provided in [21, Th. 2] can be obtained from $U(\rho)$ in (11) by assuming perfect knowledge of the past phase-noise samples. This results in the following cruder lower bound on $h(\mathbf{y}_k \,|\, \mathbf{y}^{k-1}, \mathbf{x}^{k-1}, \mathbf{x}_k)$ (cf., (20))

$$h(\mathbf{y}_k \,|\, \mathbf{y}^{k-1}, \mathbf{x}^{k-1}, \mathbf{x}_k) \geq h(\mathbf{y}_k \,|\, \theta^{k-1}, \mathbf{x}_k) \tag{35}$$
$$= h(\mathbf{y}_k \,|\, \theta_{k-1}, \mathbf{x}_k). \tag{36}$$

An even coarser bound can be obtained by assuming perfect knowledge of the additive noise $\phi_0(s_0^2)$ affecting the current phase-noise sample (see (31)–(34)). This results in the following simple capacity upper bound

$$C(\rho) \leq \sup \{I(\mathbf{y}_0; \mathbf{x}_0)\} + \log(2\pi) - h(\Delta) \tag{37}$$

where Δ is distributed as in (4). The inequality (37) can be interpreted as follows: the capacity of a Wiener phase-noise channel is upper-bounded by the capacity of a memoryless phase-noise channel with uniform phase noise, plus a correction term that accounts for the memory in the channel and does not depend on the SNR ρ.

If we now specialize (37) to single-antenna systems and we add the additional constraint that $|s| = \sqrt{\rho}$ w.p.1 (which holds, for example, if a PSK constellation is used), the first term on the RHS of (37) vanishes and we recover the upper bound previously reported in [14, Th. 2].

The last term in (12) can be computed by using a slightly modified version of the algorithm described in [16].

B. Asymptotic Behavior

In Theorem 3 below, we present an asymptotic characterization of $C(\rho)$ that generalizes to the MIMO case and to the case of peak-power-constrained inputs the asymptotic characterization reported in [8] for the single-antenna case and average-power-constrained inputs.

Theorem 3: In the high-SNR regime, the capacity of the Wiener phase-noise channel (1) behaves as

$$C(\rho) = \left(M - \frac{1}{2}\right)\log\rho - \log\left(M - \frac{1}{2}\right) - \log\Gamma(M)$$
$$+ \frac{1}{2}\log\pi - \left(M - \frac{1}{2}\right) - h(\Delta) + o(1) \quad (38)$$

where $o(1)$ indicates a function of ρ that vanishes in the limit $\rho \to \infty$.

Proof: The proof, which is rather technical, is relegated to the appendix. ∎

C. Average power versus peak power

By comparing the asymptotic capacity expansion provided in Theorem 3 with the one reported in [21, Th. 3] for the case of average-power-constrained input signals, we can assess the throughput loss at high SNR due to the presence of the more stringent peak-power constraint (8). Specifically, let $C_{ap}(\rho)$ denote the capacity of the channel in (1) when the input signal is subject to (7) instead of (8). Furthermore, let $C(\rho)$ as in (9). Then

$$\lim_{\rho\to\infty}\{C_{ap}(\rho) - C(\rho)\} = \log\Gamma\left(M - \frac{1}{2}\right)$$
$$- \left(M - \frac{3}{2}\right)\log\frac{1}{M - 1/2} + \left(M - \frac{1}{2}\right). \quad (39)$$

For the single-antenna case (i.e., $M = 1$) this asymptotic capacity loss is about 1 bit/s/Hz.

IV. THE NON-UNITARY CASE

As mentioned in Section II-B, practical considerations may force the channel matrix $\mathbf{H}$ to be non-unitary. In this section, we derive non-asymptotic upper and lower bounds on $C(\rho)$ for the general case of full rank $\mathbf{H}$, which are function of the capacity for the case of unitary $\mathbf{H}$. This allows us to extend the results reported in Section III to the non-unitary case.

Let $\lambda_{\min}$ and $\lambda_{\max}$ denote the smallest and the largest eigenvalue of $\mathbf{H}^H\mathbf{H}$, respectively. The following theorem gives upper and lower bounds to $C(\rho)$ for arbitrary full-rank matrix.

Theorem 4: Let $C_{unitary}(\rho)$ be the capacity of the channel in (1) for the case of unitary $\mathbf{H}$. The capacity for the case of an arbitrary full-rank matrix $\mathbf{H}$ with smallest and largest singular values given by $\sqrt{\lambda_{\min}}$ and $\sqrt{\lambda_{\max}}$, respectively, can be bounded as follows:

$$C_{unitary}(\lambda_{\min}\rho) \le C(\rho) \le C_{unitary}(\lambda_{\max}\rho). \quad (40)$$

Proof: Since $\mathbf{H}$ has full rank and is known at both sides, precoding at the transmitter can be done in order to invert the channel. Precisely, set $\mathbf{x}_k = \mathbf{H}^{-1}\tilde{\mathbf{x}}_k$, so that (1) becomes

$$\mathbf{y}_k = e^{j\theta_k}\tilde{\mathbf{x}}_k + \mathbf{w}_k, \quad k = 1, 2, \ldots, n. \quad (41)$$

The peak-power constraint (8) forces $\tilde{\mathbf{x}}_k$ within the hyperellipsoid

$$\tilde{\mathbf{x}}_k^H\left(\mathbf{HH}^H\right)^{-1}\tilde{\mathbf{x}}_k \le \rho, \quad k = 1, \ldots, n \quad (42)$$

w.p.1. By definition:

$$C_{unitary}(\lambda_{\max}\rho) = \lim_{n\to\infty}\frac{1}{n}\sup I(\tilde{\mathbf{x}}^n; \mathbf{y}^n) \quad (43)$$

where the supremum is over all distributions on $\tilde{\mathbf{x}}^n$ that satisfy

$$\|\tilde{\mathbf{x}}_k\|^2 \le \lambda_{\max}\rho, \quad k = 1, \ldots, n, \quad \text{w.p.1.} \quad (44)$$

The peak-power constraint in (44) is looser than (42). Indeed,

$$\|\tilde{\mathbf{x}}_k\|^2/\lambda_{\max} \le \tilde{\mathbf{x}}_k^H\left(\mathbf{HH}^H\right)^{-1}\tilde{\mathbf{x}}_k \le \|\tilde{\mathbf{x}}_k\|^2/\lambda_{\min}. \quad (45)$$

Hence, if (42) holds, then (44) holds as well. This implies that

$$C(\rho) \le C_{unitary}(\lambda_{\max}\rho). \quad (46)$$

In the same way, (45) and the definition of $C_{unitary}(\rho)$ allow us to conclude that

$$C(\rho) \ge C_{unitary}(\lambda_{\min}\rho). \quad (47)$$
∎

As a consequence of (40), bounds for the case of unitary $\mathbf{H}$ can be transformed into bounds for the case of full-rank $\mathbf{H}$ at the cost of a power offset. Using the asymptotic expression for $C_{unitary}(\rho)$ reported in (38), we see that in the high-SNR regime the gap between the upper and lower bounds in (40) is equal to

$$C_{unitary}(\lambda_{\max}\rho) - C_{unitary}(\lambda_{\min}\rho)$$
$$= \left(M - \frac{1}{2}\right)\log\frac{\lambda_{\max}}{\lambda_{\min}} + o(1) \quad (48)$$

which tends to a constant as SNR increases.

Note that both the upper and the lower bound are obtained by neglecting the actual structure of (42). In order to take this structure into account, a non-isotropic distribution of $\tilde{\mathbf{x}}_k$, with power allocated according to a waterfilling strategy, may result in a tighter lower bound.

V. NUMERICAL RESULTS

In this section, we numerically compute the upper bound in (11) and compare it with the asymptotic expression in (38), for a standard deviation of the phase-noise increments equal to $\sigma_\Delta = 6°$, in the two cases $M = 1$ (single-antenna system) and $M = 2$. In Fig. 1, the curves for the single-antenna case are displayed. The bound approaches the asymptotic expression as SNR grows large, although it remains below it for all the SNR values considered. In the figure, we also show the upper bound from [21, Th. 2] and its asymptotic expansion [21, Eq. (17)]. Although these results were derived for an average-power constraint, they serve as upper bounds for the capacity under a peak-power constraint, since the latter is more stringent than the former. Finally, we also plot an upper bound that is obtained from (11) by substituting the conditional differential entropy

$$h\left(\theta_0 + \phi_0(\xi^2) \mid \{\theta_l + \phi_l(\rho)\}_{l=-\infty}^{-1}, |\xi + z_0|\right) \quad (49)$$

with

$$h\left(\Delta + \phi_0(\xi^2) \mid |\xi + z_0|\right). \quad (50)$$

This bound, which we refer to as $U_\mathrm{s}(\rho)$ (where the letter "s" stands for "simplified") is much simpler to evaluate numerically than $U(\rho)$. Furthermore, its computational complexity does not scale with the number of antennas (on the contrary, the computational complexity of $U(\rho)$ increases exponentially with the number of antennas). Unfortunately, $U_\mathrm{s}(\rho)$ is less tight than $U(\rho)$ because

$$h\Big(\theta_0 + \phi_0(\xi^2) \mid \{\theta_l + \phi_l(\rho)\}_{l=-\infty}^{-1}, |\xi + z_0|\Big)$$
$$\geq h\big(\Delta + \phi_0(\xi^2) \mid |\xi + z_0|\big). \quad (51)$$

The newly derived bounds improve on the previous ones by 6-7 dB at moderate and high SNR values, in accordance with what reported in Subsection III-C. Finally, the numerically computed mutual information for the case of 64-QAM is also shown.[4] A gap ranging from $2\,\mathrm{dB}$ to $3\,\mathrm{dB}$ is observed between (11) and the 64-QAM curve, depending on the SNR. From the plot, we see that the asymptotic capacity expression, which, differently from both upper and lower bounds, is trivial to compute, accurately describes the behavior of the capacity for SNR larger than $16\,\mathrm{dB}$.

In Fig. 2, the curves for the case $M = 2$ are shown. As in Fig. 1, we also depict the upper bound from [21, Th. 2] together with its asymptotic version [21, Eq. (17)], the simplified upper bound $U_s(\rho)$, and the mutual information achieved by 64-QAM. The newly derived bounds improve on the previous ones by about $3\,\mathrm{dB}$ in the high-SNR region. For the MIMO case, the gap between (11) and the QAM curve is about $3.5\,\mathrm{dB}$ in the high-SNR region and larger for smaller SNR values. In this case, the asymptotic capacity expression seems to describe accurately the capacity behavior for SNR values as small as $4\,\mathrm{dB}$.

It is appropriate to point out that there is no guarantee that our upper bound $U(\rho)$ converges to the asymptotic capacity expression (38) as SNR grow large. In fact, the output distribution used in the duality step in the two cases is different. Obtaining a tighter non-asymptotic bound based on the output distribution that is optimal asymptotically remains an open problem.

Fig. 1. The upper bound $U(\rho)$ in (11), its simplified version $U_\mathrm{s}(\rho)$, the asymptotic capacity approximation (38), the upper bound from [21, Th. 2] and its asymptotic version [21, Eq. (17)], and the rates achievable with 64-QAM. In the figure, $\sigma_\Delta = 6°$.

Plots for the case of a non-unitary matrix $\mathbf{H}$ can be obtained directly from Fig. 2 by shifting the upper bound to the left by

<hr>

[4]Specifically, we use the algorithm for the computation of the information rates for finite-state channels proposed [16]. We choose 200 levels for the discretization of the phase-noise process, and average over a block of 2000 channel uses.

Fig. 2. The upper bound $U(\rho)$ in (11), its simplified version $U_\mathrm{s}(\rho)$, the asymptotic capacity approximation (38), the upper bound from [21, Th. 2] and its asymptotic version [21, Eq. (17)], and the rates achievable with 64-QAM. In the figure, $\sigma_\Delta = 6°$.

λ_max (expressed in dB), and shifting the lower bound to the right by λ_min (expressed in dB). The gap between the resulting upper and lower bounds increases proportionally to the logarithm of the ratio between λ_max and λ_min in accordance to (48).

VI. CONCLUSIONS

We presented an asymptotic (high-SNR) characterization, as well as nonasymptotic bounds, on the capacity of MIMO microwave backhaul links affected by Wiener phase noise. Our results are developed for the case of common oscillator at the transceivers, and under the practically relevant assumption that the transmit signal is subject to a peak-power constraint. By numerical simulations, we showed that our asymptotic capacity expression, which—differently from the capacity upper and lower bounds—is trivial to compute, is accurate at the SNR values typically encountered in microwave backhaul links ($15\,\mathrm{dB}$ or higher). In the regime where our asymptotic capacity formula is tight, QAM constellations exhibit a gap of about $3\,\mathrm{dB}$. The gap to the capacity upper bound may be reduced by replacing QAM with suitably optimized constellations. Furthermore, the upper bound could be further tightened by substituting the Gamma distribution (27) used in the duality step with the asymptotically optimal output distribution (85), and then by optimizing over the parameter ϵ. This, however, may further increase the computational complexity associated to the numerical evaluation of the upper bound.

APPENDIX

A. Proof of Theorem 3

The asymptotic characterization (38) is obtained by proving that the upper bound (37) matches up to a $o(1)$ term the lower bound we shall report in Appendix A1 below.

1) Lower bound: We take $\{\mathbf{x}_k\}$ i.i.d. and isotropically distributed according to Proposition 1. Specifically, we let $\mathbf{x}_k \sim \sqrt{\rho}\mathbf{z}_{\rho,\rho_0}$ where for a given $\rho_0 > 0$ the random variable $\mathbf{z}_{\rho,\rho_0}$ has the following pdf:

$$f_{\mathbf{z}_{\rho,\rho_0}}(\mathbf{a}) = \frac{f_\mathbf{z}(\mathbf{a})}{\Pr\{\|\mathbf{z}\|^2 \geq \rho_0/\rho\}} \mathbb{1}\left\{\frac{\rho_0}{\rho} \leq \|\mathbf{a}\|^2 \leq 1\right\}. \quad (52)$$

Here, $\mathbb{1}\{\cdot\}$ denotes the indicator function and the pdf $f_{\mathbf{z}}$ of the random variable $\mathbf{z}$ is given by

$$f_{\mathbf{z}}(\mathbf{a}) = \left(M - \frac{1}{2}\right)\frac{\Gamma(M)}{\pi^M}\frac{1}{\|\mathbf{a}\|}\mathbb{1}\{\|\mathbf{a}\|^2 \le 1\}. \quad (53)$$

Note that for every $\rho_0 > 0$ the pdf $f_{\mathbf{z}_{\rho,\rho_0}}$ converges pointwise to $f_{\mathbf{z}}$ as $\rho \to \infty$. The rationale behind the choice of $f_{\mathbf{z}}$ is that it turns out to maximize $h(\mathbf{z}) - \mathbb{E}[\log\|\mathbf{z}\|]$ under the constraint that $\|\mathbf{z}\| \le 1$ w.p.1. The pdf $f_{\mathbf{z}_{\rho,\rho_0}}$ is constructed from $f_{\mathbf{z}}$ so as to guarantee that $\|\mathbf{x}_k\|^2 \ge \rho_0$, a property that will be useful in the remainder of the proof.

To obtain the desired lower bound, we first proceed as in [21] and use chain rule for mutual information and that mutual information is nonnegative to obtain

$$I(\mathbf{x}^n; \mathbf{y}^n) = \sum_{k=1}^{n} I(\mathbf{x}_k; \mathbf{y}^n \mid \mathbf{x}^{k-1}) \quad (54)$$

$$\ge \sum_{k=2}^{n} I(\mathbf{x}_k; \mathbf{y}^k \mid \mathbf{x}^{k-1}). \quad (55)$$

Fix now $k \ge 2$ and set

$$\epsilon_k \triangleq I(\mathbf{x}_k; \theta_{k-1} \mid \mathbf{y}_k, \mathbf{y}_{k-1}, \mathbf{x}_{k-1}).$$

We have

$$I(\mathbf{x}_k; \mathbf{y}^k \mid \mathbf{x}^{k-1}) = I(\mathbf{x}_k; \mathbf{y}^k, \mathbf{x}^{k-1}) \quad (56)$$

$$\ge I(\mathbf{x}_k; \mathbf{y}_k, \mathbf{y}_{k-1}, \mathbf{x}_{k-1}) \quad (57)$$

$$= I(\mathbf{x}_k; \mathbf{y}_k, \mathbf{y}_{k-1}, \mathbf{x}_{k-1}, \theta_{k-1}) - \epsilon_k \quad (58)$$

$$= I(\mathbf{x}_k; \mathbf{y}_k, \theta_{k-1}) - \epsilon_k \quad (59)$$

$$= I(\mathbf{x}_k; \mathbf{y}_k \mid \theta_{k-1}) - \epsilon_k \quad (60)$$

$$= I(\mathbf{x}_2; \mathbf{y}_2 \mid \theta_1) - \epsilon_2. \quad (61)$$

Here, (56) follows because the $\{\mathbf{x}_k\}$ are independent; in (57) we used chain rule for mutual information and that mutual information is nonnegative; (59) follows because $\mathbf{x}_k$ and the pair $(\mathbf{y}_{k-1}, \mathbf{x}_{k-1})$ are conditionally independent given $(\theta_{k-1}, \mathbf{y}_k)$; (60) holds because $\mathbf{x}_k$ and θ_{k-1} are independent; finally, (61) follows from stationarity. Substituting (61) into (55) and then (55) into (9), we obtain

$$C(\rho) \ge I(\mathbf{x}_2; \mathbf{y}_2 \mid \theta_1) - \epsilon_2. \quad (62)$$

We next investigate the two terms on the RHS of (62) separately. We shall show that the first term has the desired asymptotic expansion, while the second term can be made arbitrarily close to zero by choosing ρ_0 sufficiently large.

The first term on the RHS of (62)*:* We write

$$I(\mathbf{x}_2; \mathbf{y}_2 \mid \theta_1) = h(\mathbf{y}_2 \mid \theta_1) - h(\mathbf{y}_2 \mid \mathbf{x}_2, \theta_1) \quad (63)$$

and bound the two terms separately. For the first term, we have that

$$h(\mathbf{y}_2 \mid \theta_1) \ge h(\mathbf{y}_2 \mid \mathbf{w}_2, \theta_1) \quad (64)$$

$$= h(e^{j\theta_2}\mathbf{x}_2 \mid \theta_1) \quad (65)$$

$$= h(\mathbf{x}_2) \quad (66)$$

$$= M\log\rho + h(\mathbf{z}_{\rho,\rho_0}) \quad (67)$$

$$= M\log\rho - \log\frac{(M-1/2)\Gamma(M)}{\pi^M \Pr\{\|\mathbf{z}\|^2 \ge \rho_0/\rho\}}$$
$$+ \mathbb{E}[\log\|\mathbf{z}_{\rho,\rho_0}\|]. \quad (68)$$

Here (66) follows because $\mathbf{x}_2$ is isotropically distributed and (68) holds because of (52) and (53). For the second term on the RHS of (63), we proceed as follows. Let $\mathbf{x}_2 = s_2\mathbf{v}_2$, with $s_2 = \|\mathbf{x}_2\|$ and, hence, $s_2^2 \sim \rho\|\mathbf{z}_{\rho,\rho_0}\|^2$. Furthermore, let $z_2 \sim \mathcal{CN}(0,1)$. Then, proceeding as in [12, Eq. (10)]

$$h(\mathbf{y}_2 \mid \mathbf{x}_2, \theta_1) = h(\mathbf{y}_2 \mid s_2, \mathbf{v}_2, \theta_1) \quad (69)$$

$$= h(e^{j\theta_2}s_2 + z_2 \mid s_2, \theta_1) + \log(\pi e)^{M-1}. \quad (70)$$

The first term on the RHS of (70) can be bounded as follows

$$h(e^{j\theta_2}s_2 + z_2 \mid s_2, \theta_1) \quad (71)$$

$$= h(e^{j\theta_2}(s_2 + z_2) \mid s_2, \theta_1) \quad (72)$$

$$= h(e^{j\Delta}(s_2 + z_2) \mid s_2) \quad (73)$$

$$= h(|s_2 + z_2|^2 \mid s_2)$$
$$+ h(\phi_2(s_2^2) + \Delta \mid |s_2 + z_2|, s_2) - \log 2 \quad (74)$$

$$\le \frac{1}{2}\mathbb{E}\left[\log\left(2\pi e\left[1 + 2\rho\|\mathbf{z}_{\rho,\rho_0}\|^2\right]\right)\right]$$
$$+ h(\phi_2(s_2^2) + \Delta \mid s_2) - \log 2. \quad (75)$$

Here, in (73) we used (2) and denoted by Δ a random variable distributed as in (4); in (74) we evaluated the differential entropy in polar coordinates using [31, Lemma 6.15 and Lemma 6.16]. Finally, (75) follows because the Gaussian distribution maximizes differential entropy under a variance constraint and because conditioning reduces entropy. Note that

$$h(\phi_2(s_2^2) + \Delta \mid s_2) \le \max_{\xi \ge \sqrt{\rho_0}} h(\phi_2(\xi^2) + \Delta)$$
$$= h(\phi_2(\rho_0) + \Delta). \quad (76)$$

This term can be made arbitrarily close to $h(\Delta)$ by choosing ρ_0 in (52) sufficiently large. Summarizing, we have shown that

$$I(\mathbf{x}_2; \mathbf{y}_2 \mid \theta_1) \ge M\log\rho - \log\frac{(M-1/2)\Gamma(M)}{\pi^M \Pr\{\|\mathbf{z}\|^2 \ge \rho_0/\rho\}}$$

$$+ \mathbb{E}[\log\|\mathbf{z}_{\rho,\rho_0}\|] - \frac{1}{2}\mathbb{E}\left[\log\left(2\pi e\left[1 + 2\rho\|\mathbf{z}_{\rho,\rho_0}\|^2\right]\right)\right]$$
$$- h(\phi_2(\rho_0) + \Delta) + \log 2 - \log(\pi e)^{M-1} \quad (77)$$

$$= \left(M - \frac{1}{2}\right)\log\rho - \log\left(M - \frac{1}{2}\right) - \log\Gamma(M)$$

$$+ \frac{1}{2}\log\pi - \left(M - \frac{1}{2}\right) - h(\phi_2(\rho_0) + \Delta) + o(1). \quad (78)$$

Here, the last step follows because

$$\mathbb{E}\left[\log(1 + c\rho\|\mathbf{z}_{\rho,\rho_0}\|^2)\right] = \log(c\rho)$$
$$+ \mathbb{E}\left[\log\|\mathbf{z}_{\rho,\rho_0}\|^2\right] + o(1) \quad (79)$$

for all $c > 0$, and

$$\lim_{\rho \to \infty} \Pr\{\|\mathbf{z}\|^2 \ge \rho_0/\rho\} = 1. \quad (80)$$

The second term on the RHS of (62)*:* Let $\mathbf{x}_1 = s_1\mathbf{v}_1$ and $z_1 \sim \mathcal{CN}(0,1)$. Proceeding similarly as in [31, App. IX], we obtain (see [21, Eq. (25)])

$$I(\mathbf{x}_2; \theta_1 \mid \mathbf{y}_2, \mathbf{y}_1, \mathbf{x}_1)$$
$$= h(\theta_2 \mid e^{j\theta_1}(\sqrt{\rho_0} + z_1)) - h(\theta_2 \mid \theta_1). \quad (81)$$

As claimed, the RHS of (81) can be made arbitrarily close to zero by choosing ρ_0 in (52) sufficiently large.

2) Upper Bound: We exploit the property that the high-SNR behavior of $C(\rho)$ does not change if the support of the input distribution is constrained to lie outside a sphere of arbitrary radius. This result, known as *escape-to-infinity* property of the capacity-achieving input distribution [31, Def. 4.11], is formalized in the following lemma.

Lemma 5: Fix an arbitrary $\xi_0 > 0$ and let $\mathcal{K}(\xi_0) = \{\mathbf{x} \in \mathbb{C}^M : \|\mathbf{x}\| \geq \xi_0\}$. Denote by $C^{(\xi_0)}(\rho)$ the capacity of the channel (1) when the input signal is subject to the peak-power constraint (8) and to the additional constraint that $\mathbf{x}_k \in \mathcal{K}(\xi_0)$ almost surely for all k. Then

$$C(\rho) = C^{(\xi_0)}(\rho) + o(1), \quad \rho \to \infty \tag{82}$$

with $C(\rho)$ given in (9).

Proof: The lemma follows directly from [30, Th. 8] and [31, Th. 4.12]. ∎

Fix $\xi_0 > 0$. By proceeding as in (37), we obtain[5]

$$C^{(\xi_0)}(\rho) \leq \sup\{I(\mathbf{y}; \mathbf{x})\} + \log(2\pi) - h(\Delta) \tag{83}$$

where, this time, the supremum is over all probability distributions on $\mathbf{x}$ that satisfy $\|\mathbf{x}\|^2 \in [\xi_0^2, \rho]$ w.p.1. We next upper-bound $I(\mathbf{y}; \mathbf{x})$ by using duality as in (28), i.e., we exploit that

$$I(\mathbf{y}; \mathbf{x}) \leq -\mathbb{E}[\log q_{\mathbf{y}}(\mathbf{y})] - h(\mathbf{y} \mid \mathbf{x}) \tag{84}$$

for every output distribution $q_{\mathbf{y}}(\mathbf{y})$. We choose a different $q_{\mathbf{y}}(\mathbf{y})$ than the one resulting in (27). Roughly speaking, we want $q_{\mathbf{y}}(\mathbf{y})$ to be the output distribution induced by the input distribution (52) we used for the lower bound. When constructing $q_{\mathbf{y}}(\mathbf{y})$, we shall ignore the additive noise over the support of the input distribution, and consider the effect of the additive noise only outside an ϵ-neighborhood of the set $\{\mathbf{x} \in \mathbb{C}^M : \|\mathbf{x}\|^2 \leq \rho\}$.[6] Specifically, we shall set $\mathbf{r} \triangleq \mathbf{y}/\sqrt{\rho}$, $\mathcal{S}_\epsilon \triangleq \{\mathbf{r} \in \mathbb{C}^M : \|\mathbf{r} - \mathbf{x}'\| \leq \epsilon$ for some $\|\mathbf{x}'\| \leq 1\}$ and choose the following probability distribution for $\mathbf{r}$

$$q_{\mathbf{r}}(\mathbf{r}) = \begin{cases} \dfrac{(M - 1/2)}{\pi^M K_{\rho,\epsilon}} \dfrac{\Gamma(M)}{\|\mathbf{r}\|}, & \text{if } \mathbf{r} \in \mathcal{S}_\epsilon \\[2mm] \dfrac{\rho^M}{\pi^M K_{\rho,\epsilon}} e^{-\rho\|\mathbf{r}\|^2}, & \text{if } \mathbf{r} \notin \mathcal{S}_\epsilon \end{cases} \tag{85}$$

where

$$K_{\rho,\epsilon} = \underbrace{\int\limits_{\mathbf{r} \in \mathcal{S}_\epsilon} \frac{(M - 1/2)}{\pi^M} \frac{\Gamma(M)}{\|\mathbf{r}\|} d\mathbf{r}}_{\triangleq K_{\infty,\epsilon}} + \int\limits_{\mathbf{r} \notin \mathcal{S}_\epsilon} \frac{\rho^M}{\pi^M} e^{-\rho\|\mathbf{r}\|^2} d\mathbf{r}. \tag{86}$$

This yields

$$-\mathbb{E}[\log q_{\mathbf{y}}(\mathbf{y})] = M \log \rho - \mathbb{E}[\log q_{\mathbf{r}}(\mathbf{r})] \tag{87}$$

[5]To keep notation compact, we write $\mathbf{x}_0$ simply as $\mathbf{x}$; same convention for $\mathbf{y}_0$.

[6]This choice is inspired by [32], where the rates achievable with dense constellations over an AWGN channel (no phase noise) are analyzed.

where

$$\begin{aligned} -\mathbb{E}[\log q_{\mathbf{r}}(\mathbf{r})] = &-\left[\log \frac{(M - 1/2)\Gamma(M)}{\pi^M K_{\rho,\epsilon}}\right] \Pr\{\mathbf{r} \in \mathcal{S}_\epsilon\} \\ &+ \mathbb{E}[(\log\|\mathbf{r}\|)\mathbb{1}\{\mathbf{r} \in \mathcal{S}_\epsilon\}] \\ &-\left[\log \frac{\rho^M}{\pi^M K_{\rho,\epsilon}}\right] \Pr\{\mathbf{r} \notin \mathcal{S}_\epsilon\} \\ &+ \mathbb{E}[\rho\|\mathbf{r}\|^2 \mathbb{1}\{\mathbf{r} \notin \mathcal{S}_\epsilon\}]. \end{aligned} \tag{88}$$

We next characterize each term on the RHS of (88) in the limit $\rho \to \infty$.

The first term: By construction (see (86)), we have that

$$\lim_{\epsilon \to 0} \lim_{\rho \to \infty} K_{\rho,\epsilon} = \lim_{\epsilon \to 0} K_{\infty,\epsilon} = 1. \tag{89}$$

Furthermore, let $\mathbf{w} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_M)$. Then

$$\Pr\{\mathbf{r} \notin \mathcal{S}_\epsilon\} = \Pr\{\|\mathbf{w}\|/\sqrt{\rho} \geq \epsilon\} \tag{90}$$

$$= \frac{\Gamma(M, \epsilon^2 \rho)}{\Gamma(M)} \tag{91}$$

$$= \frac{(\epsilon^2 \rho)^{M-1} e^{-\epsilon^2 \rho}}{\Gamma(M)} + o(\rho^{M-1} e^{-\epsilon^2 \rho}) \tag{92}$$

for $\rho \to \infty$. Here, (90) follows because $\mathbf{r} \sim \mathbf{x}' + \mathbf{w}/\sqrt{\rho}$ for some $\|\mathbf{x}'\| \leq 1$; in (91), the function $\Gamma(\cdot, \cdot)$ is the upper incomplete Gamma function [33, Eq. 6.5.3]; finally, (92) follows from [33, Eq. 6.5.32]. Using (92), we conclude that the first term on the RHS of (88) admits the following asymptotic expansion:

$$\begin{aligned} \left[\log \frac{(M - 1/2)\Gamma(M)}{\pi^M K_{\rho,\epsilon}}\right] &\Pr\{\mathbf{r} \in \mathcal{S}_\epsilon\} \\ &= \log \frac{(M - 1/2)\Gamma(M)}{\pi^M K_{\infty,\epsilon}} + o(1). \end{aligned} \tag{93}$$

Furthermore, (89) implies that $K_{\infty,\epsilon}$ can be made arbitrarily close to 1 by choosing ϵ sufficiently small.

The second term: Note that

$$\begin{aligned} \mathbb{E}[(\log\|\mathbf{r}\|)&\mathbb{1}\{\mathbf{r} \in \mathcal{S}_\epsilon\}] \\ &= \frac{1}{2} \mathbb{E}[(\log\|\mathbf{y}\|^2)\mathbb{1}\{\mathbf{y}/\sqrt{\rho} \in \mathcal{S}_\epsilon\}] - \frac{1}{2} \log \rho. \end{aligned} \tag{94}$$

Assume without loss of generality that $\rho > 1$. Then $(\mathbf{y}/\sqrt{\rho}) \notin \mathcal{S}_\epsilon$ implies that $\|\mathbf{y}\|^2 > \rho > 1$. Hence, we conclude that $\log\|\mathbf{y}\|^2 > 0$ whenever $(\mathbf{y}/\sqrt{\rho}) \notin \mathcal{S}_\epsilon$. As a consequence, we can upper-bound (94) by adding

$$\frac{1}{2} \mathbb{E}[(\log\|\mathbf{y}\|^2)\mathbb{1}\{\mathbf{y}/\sqrt{\rho} \notin \mathcal{S}_\epsilon\}] \tag{95}$$

and obtain

$$\mathbb{E}[(\log\|\mathbf{r}\|)\mathbb{1}\{\mathbf{r} \in \mathcal{S}_\epsilon\}] \leq \frac{1}{2} \mathbb{E}[\log\|\mathbf{y}\|^2] - \frac{1}{2} \log \rho. \tag{96}$$

The third term: It follows from (92) that

$$\left[\log \frac{\rho^M}{\pi^M K_{\rho,\epsilon}}\right] \Pr\{\mathbf{r} \notin \mathcal{S}_\epsilon\} = o(1). \tag{97}$$

The fourth term: We have that

$$\mathbb{E}\left[\rho\|\mathbf{r}\|^2\mathbb{1}\{\mathbf{r}\notin\mathcal{S}_\epsilon\}\right]\leq\rho\sqrt{\mathbb{E}[\|\mathbf{r}\|^4]\Pr\{\mathbf{r}\notin\mathcal{S}_\epsilon\}}\quad(98)$$
$$=o(1).\quad(99)$$

Here, (98) follows from Chaucy-Schwarz inequality and (99) follows from (92).

We next substitute (93), (94), (97), and (99) into (88) and then (88) into (87) and obtain

$$-\mathbb{E}[\log q_{\mathbf{y}}(\mathbf{y})]=\left(M-\frac{1}{2}\right)\log\rho-\log\frac{(M-1/2)\Gamma(M)}{\pi^M K_{\infty,\epsilon}}$$
$$+\frac{1}{2}\mathbb{E}\left[\log\|\mathbf{y}\|^2\right]+o(1).\quad(100)$$

Set now $s=\|\mathbf{x}\|$ and $z\sim\mathcal{CN}(0,1)$. By proceeding as in [21, Eq. (33)], we can rewrite the conditional differential entropy $h(\mathbf{y}\,|\,\mathbf{x})$ on the RHS of (84) as

$$h(\mathbf{y}\,|\,\mathbf{x})=h(|s+z|^2\,|\,s)+\log\pi^M+M-1.\quad(101)$$

Substituting (100) and (101) into (84) and using that

$$\|\mathbf{y}\|^2\sim|s+z_1|^2+\sum_{j=2}^M|z_j|^2\quad(102)$$

where $z_j\sim\mathcal{CN}(0,1)$, $j=1,\ldots,M$, we obtain

$$I(\mathbf{y};\mathbf{x})\leq\left(M-\frac{1}{2}\right)\log\rho-\log\frac{(M-1/2)\Gamma(M)}{K_{\infty,\epsilon}}$$
$$+\frac{1}{2}\mathbb{E}\left[\log\left(|s+z_1|^2+\sum_{j=2}^M|z_j|^2\right)\right]$$
$$-h(|s+z|^2\,|\,s)-(M-1)+o(1)\quad(103)$$
$$\leq\left(M-\frac{1}{2}\right)\log\rho-\log\frac{(M-1/2)\Gamma(M)}{K_{\infty,\epsilon}}-(M-1)$$
$$+\max_{\xi_0\leq\xi\leq\sqrt{\rho}}\left\{\frac{1}{2}\mathbb{E}\left[\log\left(|\xi+z_1|^2+\sum_{j=2}^M|z_j|^2\right)\right]\right.$$
$$\left.-h(|\xi+z|^2)\right\}+o(1).\quad(104)$$

Substituting (104) into (83) and using that

$$\lim_{\xi\to\infty}\left\{\frac{1}{2}\mathbb{E}\left[\log\left(|\xi+z_1|^2+\sum_{j=2}^M|z_j|^2\right)\right]-h\left(|\xi+z|^2\right)\right\}$$
$$=-\frac{1}{2}\log(4\pi e)\quad(105)$$

which follows by [8, Eq. (9)] and by proceeding similarly to the proof of [31, Lemma 6.9], we conclude that we can make the bound on $C^{(\xi_0)}(\rho)$ just derived to be arbitrarily close to (38) in the high-SNR regime by choosing ϵ sufficiently small and ξ_0 sufficiently large.

REFERENCES

[1] M. Coldrey, H. Koorapaty, J. Berg, Z. Ghebretensaé, J. Hansryd, A. Derneryd, and S. Falahati, "Small-cell wireless backhauling: A non-line-of-sight approach for point-to-point microwave links," in *Proc. IEEE Veh. Technol. Conf. Fall (VTC-Fall)*, Quebec City, Canada, Sep. 2012, pp. 1–5.

[2] J. Hansryd and J. Edstam, "Microwave capacity evolution," *Ericsson review*, vol. 1, 2011.

[3] H. Mehrpouyan, A. Nasir, S. Blostein, T. Eriksson, G. Karagiannidis, and T. Svensson, "Joint estimation of channel and oscillator phase noise in MIMO systems," *IEEE Trans. Signal Process.*, vol. 60, no. 9, pp. 4790–4807, Sep. 2012.

[4] S. Chinnici, C. Decanis, A. Quadrini, and D. Weinholt, "High order M-QAM transceiver for Gigabit radio microwave transmission: FPGA test chipset and ASIC design," in *IEEE Int Conf. New Circ. Syst. (NEWCAS)*, Bordeaux, France, Jun. 2011, pp. 297–300.

[5] A. Tarable, G. Montorsi, S. Benedetto, and S. Chinnici, "An EM-based phase-noise estimator for MIMO systems," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Budapest, Hungary, Jun. 2013, pp. 3215–3219.

[6] G. Durisi, A. Tarable, and T. Koch, "On the multiplexing gain of MIMO microwave backhaul links affected by phase noise," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Budapest, Hungary, Jun. 2013, pp. 3209–3214.

[7] G. Colavolpe, "Communications over phase-noise channels: A tutorial review," in *Advanced Satellite Multimedia Systems Conference (ASMS) and 12th Signal Processing for Space Communications Workshop (SPSC)*, Sep. 2012, pp. 316–327.

[8] A. Lapidoth, "On phase noise channels at high SNR," in *Proc. IEEE Inf. Theory Workshop (ITW)*, Bangalore, India, Oct. 2002, pp. 1–4.

[9] H. Ghozlan and G. Kramer, "On Wiener phase noise channels at high signal-to-noise ratio," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Istanbul, Turkey, Jul. 2013, pp. 2279–2283.

[10] ——, "Multi-sample receivers increase information rates for Wiener phase noise channels," in *Proc. IEEE Global Telecommun. Conf. (GLOBECOM)*, Atlanta, GA, Dec. 2013, to appear.

[11] R. Nuriyev and A. Anastasopoulos, "Capacity and coding for the block-independent noncoherent AWGN channel," *IEEE Trans. Inf. Theory*, vol. 51, no. 3, pp. 866–883, Mar. 2005.

[12] G. Durisi, "On the capacity of the block-memoryless phase-noise channel," *IEEE Commun. Lett.*, vol. 16, no. 8, pp. 1157–1160, Aug. 2012.

[13] M. Katz and S. Shamai (Shitz), "On the capacity-achieving distribution of the discrete-time noncoherent and partially coherent AWGN channels," *IEEE Trans. Inf. Theory*, vol. 50, no. 10, pp. 2257–2270, Oct. 2004.

[14] A. Barbieri and G. Colavolpe, "On the information rate and repeat-accumulate code design for phase noise channels," *IEEE Trans. Commun.*, vol. 59, no. 12, pp. 3223 –3228, Dec. 2011.

[15] L. Barletta, M. Magarini, and A. Spalvieri, "The information rate transferred through the discrete-time Wiener's phase noise channel," *J. Lightw. Technol.*, vol. 30, no. 10, pp. 1480–1486, May 2012.

[16] D. Arnold, H.-A. Loeliger, P. Vontobel, A. Kavcic, and W. Zeng, "Simulation-based computation of information rates for channels with memory," *IEEE Trans. Inf. Theory*, vol. 52, no. 8, pp. 3498–3508, Aug. 2006.

[17] L. Barletta, M. Margarini, and A. Spalvieri, "Tight upper and lower bounds to the information rate of phase noise channel," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Istanbul, Turkey, Jul. 2013, pp. 2284–2288.

[18] D. S. Baum and H. Bölcskei, "Information-theoretic analysis of MIMO channel sounding," *IEEE Trans. Inf. Theory*, vol. 57, no. 11, pp. 7555–7577, Nov. 2011.

[19] E. Björnson, J. Hoydis, M. Kountouris, and M. Debbah, "Massive MIMO systems with non-ideal hardware: Energy efficiency, estimation, and capacity limits," Jul. 2013. [Online]. Available: http://arxiv.org/abs/1307.2584

[20] A. Pitarokoilis, S. K. Mohammed, and E. G. Larsson, "Effect of oscillator phase noise on uplink performance of large MU-MIMO systems," in *Proc. Allerton Conf. Commun., Contr., Comput.*, Monticello, IL, U.S.A., Oct. 2012.

[21] G. Durisi, A. Tarable, C. Camarda, and G. Montorsi, "On the capacity of MIMO Wiener phase-noise channels," in *Proc. Inf. Theory Applicat. Workshop (ITA)*, San Diego, CA, U.S.A., Feb. 2013.

[22] P. Driessen and G. Foschini, "On the capacity formula for multiple input-multiple output wireless channels: a geometric interpretation," *IEEE Trans. Commun.*, vol. 47, no. 2, pp. 173–176, Feb. 1999.

[23] D. Gesbert, H. Bölcskei, D. Gore, and A. Paulraj, "Outdoor MIMO wireless channels: models and performance prediction," *IEEE Trans. Commun.*, vol. 50, no. 12, pp. 1926–1934, Dec. 2002.

[24] F. Bohagen, P. Orten, and G. E. Oien, "Design of optimal high-rank line-of-sight MIMO channels," *IEEE Trans. Wireless Commun.*, vol. 6, no. 4, pp. 1420–1425, Apr. 2007.

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

[26] G. Wunder, R. F. H. Fischer, H. Boche, S. Litsyn, and J.-S. No, "The PAPR problem in OFDM transmission: New directions for a long-lasting problem," *IEEE Signal Process. Mag.*, vol. 30, no. 6, pp. 130–134, Nov. 2013.

[27] J. G. Smith, "The information capacity of amplitude and variance-constrained scalar gaussian channels," *Information and Control*, vol. 1, pp. 203–219, 1971.

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

[29] S. M. Moser, "The fading number of multiple-input multiple-output fading channels with memory," *IEEE Trans. Inf. Theory*, vol. 55, no. 6, pp. 2716–2755, Jun. 2009.

[30] A. Lapidoth and S. M. Moser, "The fading number of single-input multiple-output fading channels with memory," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 437–453, Feb. 2006.

[31] ——, "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.

[32] T. Koch, A. Martinez, and A. Guillen i Fabregas, "The capacity loss of dense constellations," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Boston, MA, U.S.A., Jul. 2012, pp. 135–139.

[33] M. Abramowitz and I. A. Stegun, Eds., *Handbook of Mathematical Functions*, 10th ed.   Government Printing Office, 1972.