



**HAL**  
open science

## On MIMO phase noise channels at high SNR

Sheng Yang, Shlomo Shamai Shitz

► **To cite this version:**

Sheng Yang, Shlomo Shamai Shitz. On MIMO phase noise channels at high SNR. 2016 IEEE Information Theory Workshop (ITW), Sep 2016, Cambridge, United Kingdom. 10.1109/ITW.2016.7606845 . hal-01433715

**HAL Id: hal-01433715**

**<https://centralesupelec.hal.science/hal-01433715>**

Submitted on 10 Jun 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# On MIMO Phase Noise Channels at High SNR

Sheng Yang

L2S, CentraleSupélec

91190 Gif sur Yvette, France

Email: sheng.yang@centralesupelec.fr

Shlomo Shamai (Shitz)

Technion-Israel Institute of Technology

Haifa, Israel

Email: sshlomo@ee.technion.ac.il

**Abstract**—We consider the point-to-point multi-input-multiple-output (MIMO) channel with phase uncertainty (MIMO phase noise channel) at high SNR. With phase noise on the individual paths of the channel (model A), we show that the multiplexing gain is  $\frac{1}{2}$ , which implies that the capacity *does not* scale with the channel dimension at high SNR. With phase noise at both the input and output of the channel (model B), the multiplexing gain is upper-bounded by  $\frac{1}{2} \min\{n_t, (n_r - 2)^+ + 1\}$ , and lower-bounded by  $\frac{1}{2} \min\{n_t, \lfloor \frac{n_r+1}{2} \rfloor\}$ , where  $n_t$  and  $n_r$  are the number of transmit and receive antennas, respectively. The multiplexing gain is enhanced to  $\frac{1}{2} \min\{n_t, n_r\}$  without receive phase noise, and to  $\frac{1}{2} \min\{2n_t - 1, n_r\}$  without transmit phase noise. Our main results rely on the derivation of non-trivial upper and lower bounds on the capacity of such channels.

## I. INTRODUCTION

The capacity of a point-to-point multiple-input-multiple-output (MIMO) Gaussian channel is still open in the non-coherent case, i.e., when the channel state is unknown. Nevertheless, the capacity in asymptotic regimes, e.g., at high signal-to-noise ratio (SNR), has been obtained in some important cases.

In the seminal paper [1], Lapidoth and Moser proposed a powerful technique, called the duality approach, that can be applied to a large class of fading channels and derived the exact high SNR capacity up to an  $o(1)$  term. In particular, when the differential entropy of the channel matrix is finite, i.e.,  $h(\mathbf{H}) > -\infty$ , it was shown in [1] that the pre-log (a.k.a. multiplexing gain), of the capacity is 0 and the high-SNR capacity is  $\log \log \text{SNR} + \chi(\mathbf{H}) + o(1)$  where  $\chi(\mathbf{H})$  is the so-called fading number of the channel. In [2], Zheng and Tse showed that for block fading MIMO Rayleigh channels, the pre-log is  $M^*(1 - M^*/T)$  where  $T$  is the channel coherence time and  $M^* \triangleq \min\{n_t, n_r, \lfloor \frac{T}{2} \rfloor\}$ . In this work, we are interested in the MIMO phase noise channels in which the phases of the channel coefficients are not perfectly known.

Applying the duality approach and the “escape-to-infinity” property of the channel input, Lapidoth characterized the high-SNR capacity of the discrete-time phase noise channel in the single-antenna case [3]. It was shown in [4] that the capacity-achieving input distribution is in fact discrete. Recently, capacity upper and lower bounds of the single-antenna channels with Wiener phase noise have been extensively studied in the context of optical fiber and microwave communications (see [5], [6], [7] and the references therein). In these works, the upper bounds are derived via duality and lower bounds are computed numerically using the auxiliary channel technique

proposed in [8]. In particular, in [7], Durisi *et al.* investigated the MIMO phase noise channel with a common phase noise, a scenario motivated by the microwave link with centralized oscillators. The SIMO and MISO channels with common and separate phase noises are considered in [9]. The  $2 \times 2$  MIMO phase noise channel with independent transmit and receive phase noises at each antenna was studied in [10], where the authors showed that the multiplexing gain is  $\frac{1}{2}$  for a specific class of input distributions. For general MIMO channels with separate phase noises, to the best of our knowledge, even the multiplexing gain is unknown.

In this work, we make some progresses in this direction. We consider two classes of discrete-time stationary and ergodic MIMO phase noise channels: model A with individual phase noises on the entries of the channel matrix, and model B with individual phase noises at the input and the output of the channel instead. The phase noise processes in both models are assumed to have finite differential entropy rate. For model A, we obtain the exact multiplexing gain  $\frac{1}{2}$  for any channel dimension, which implies that the capacity *does not* scale with the channel dimension at high SNR. For model B with both transmit and receive phase noises, we show that the multiplexing gain is upper-bounded by  $\frac{1}{2} \min\{n_t, (n_r - 2)^+ + 1\}$ , and lower-bounded by  $\frac{1}{2} \min\{n_t, \lfloor \frac{n_r+1}{2} \rfloor\}$ , where  $n_t$  and  $n_r$  are the number of transmit and receive antennas, respectively. The upper and lower bounds coincide for  $n_r \leq 3$  or  $n_r \geq 2n_t - 1$ . Further, when receive phase noise is absent, the multiplexing gain is improved and we obtain the exact value of  $\frac{1}{2} \min\{n_t, n_r\}$ . If the transmit phase noise is absent instead, the multiplexing gain becomes  $\frac{1}{2} \min\{2n_t - 1, n_r\}$ . The main technical contribution of this paper lies in the upper and lower bounds on the capacity of MIMO phase noise channels. The novelty of the upper bound is the finding of a suitable auxiliary distributions with which we apply the duality upper bound proposed by Lapidoth and Moser [1]. In this paper, we introduce a class of *multi-variate* Gamma distributions that, combined with the duality upper bound, allows us to obtain a complete pre-log characterization for model A and partially for model B. The second contribution is the derivation of the capacity lower bounds for model B, based on the remarkable property of the differential entropy of the output vector in this channel.

The remainder of the paper is organized as follows. The system model and main results are presented in Section II. The upper bounds are derived in Section III. We prove the lower bound for model B in section IV. Concluding remarks

are given in Section V. Due to the lack of space, proofs of the lemmas and some technical steps are omitted and can be found in the full version of the paper [11].

## II. SYSTEM MODEL AND MAIN RESULTS

Throughout the paper, we use the following notational conventions. For random quantities, we use upper case letters, e.g.,  $X$ , for scalars, upper case letters with bold and non-italic fonts, e.g.,  $\mathbf{V}$ , for vectors, and upper case letter with bold and sans serif fonts, e.g.,  $\mathbf{M}$ , for matrices. Deterministic quantities are denoted in a rather conventional way with italic letters, e.g., a scalar  $x$ , a vector  $\mathbf{v}$ , and a matrix  $\mathbf{M}$ . Logarithms are in base 2. The Euclidean norm of a vector and a matrix is denoted by  $\|\mathbf{v}\|$  and  $\|\mathbf{M}\|$ , respectively. The transpose and conjugated transpose of  $\mathbf{M}$  are  $\mathbf{M}^T$  and  $\mathbf{M}^H$ , respectively.  $\mathbf{H}^\dagger$  is the pseudo-inverse of a *tall* matrix  $\mathbf{H}$ . We use  $\mathbf{A} \circ \mathbf{B}$  to denote the Hadamard (point-wise) product between vectors/matrices. We also use  $c_0$  to represent a bounded constant whose value is irrelevant but may change at each occurrence. Similarly,  $c_H$  is a constant that may depend on  $\mathbf{H}$  but the value is irrelevant.

### A. Channel model

In this paper, we are interested in a class of discrete-time MIMO phase noise channels with  $n_t$  transmit antennas and  $n_r$  receive antennas, defined by

$$\mathbf{Y}_t = (\mathbf{H} \circ e^{j\boldsymbol{\Theta}_t}) \mathbf{x}_t + \mathbf{Z}_t, \quad t = 1, 2, \dots, N, \quad (1)$$

where the deterministic channel matrix  $\mathbf{H}$  belongs to a set of generic matrices<sup>1</sup>  $\mathcal{H} \subset \mathbb{C}^{n_r \times n_t}$ ;  $\mathbf{x}_t \in \mathbb{C}^{n_t \times 1}$  is the input vector at time  $t$ , with the average power constraint  $\frac{1}{N} \sum_{t=1}^N \|\mathbf{x}_t\|^2 \leq P$ ; the additive noise process  $\{\mathbf{Z}_t\}$  is assumed to be spatially and temporally white with  $\mathbf{Z}_t \sim \mathcal{CN}(0, \mathbf{I}_{n_r})$ ;  $\boldsymbol{\Theta}_t$  is the matrix of phase noises on the individual entries of  $\mathbf{H}$  at time  $t$ ; the phase noise process  $\{\boldsymbol{\Theta}_t\}$  is stationary and ergodic, and is independent of the additive noise process  $\{\mathbf{Z}_t\}$ . Both  $\{\mathbf{Z}_t\}$  and  $\{\boldsymbol{\Theta}_t\}$  are unknown to the transmitter and the receiver. Since the additive noise power is normalized, the transmit power  $P$  is identified with the SNR throughout the paper. The end-to-end channel is captured by the random channel matrix  $\mathbf{H} \triangleq [h_{ik} e^{\boldsymbol{\Theta}_{ik}}]_{i,k}$ . We consider two types of phase noise according to the spatial structures.

- Model A refers to channels with phase uncertainty on the individual paths of the channel (path phase noise), such that the sequence  $\{\boldsymbol{\Theta}_t\}$  has finite entropy rate  $h(\{\boldsymbol{\Theta}_t\}) > -\infty$ . It corresponds to the case where the phase information of the channel cannot be obtained accurately, e.g., in an optical fiber channel.
- Model B refers to channels with phase noises at the input and/or output of the channel, i.e.,  $\boldsymbol{\Theta}_{ik} = \boldsymbol{\Theta}_{R,i} + \boldsymbol{\Theta}_{T,k}$ . The vector  $\boldsymbol{\Theta}_T \triangleq [\boldsymbol{\Theta}_{T,i}]_{i=1}^{n_t}$  contains the  $n_t$  phase noises at the transmit antennas, and  $\boldsymbol{\Theta}_R \triangleq [\boldsymbol{\Theta}_{R,k}]_{k=1}^{n_r}$  is the vector of the  $n_r$  phase noises at the receive antennas. This model captures the phase corruption at both the transmit and

<sup>1</sup>It means that the channel matrix  $\mathbf{H}$  does not lie on any algebraic hypersurface. If we draw the entries of  $\mathbf{H}$  i.i.d. from a continuous set, then  $\mathbf{H}$  is generic almost surely.

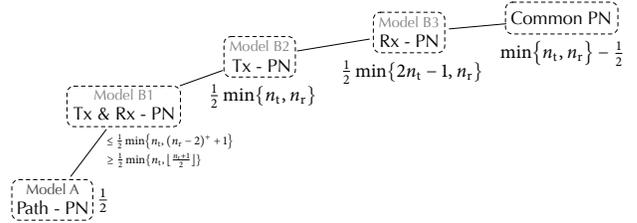


Fig. 1. Multiplexing gain of the MIMO phase noise channels.

receive RF chains, e.g., caused by imperfect oscillators. We consider three cases of model B: i) model B1 with both transmit and receive phase noises such that  $h(\{\boldsymbol{\Theta}_{T,t}, \boldsymbol{\Theta}_{R,t}\}) > -\infty$ ; ii) model B2 with only transmit phase noise such that  $h(\{\boldsymbol{\Theta}_{T,t}\}) > -\infty$ ; iii) model B3 with only receive phase noise such that  $h(\{\boldsymbol{\Theta}_{R,t}\}) > -\infty$ .

Note that model B1 covers the case where both the transmitter and receiver use separate (and imperfect) oscillators for different antennas, whereas models B2 and B3 correspond to the case with centralized oscillators at one side and separate oscillators at the other side.

The capacity of such a stationary and ergodic channel is  $C(P) \triangleq \lim_{N \rightarrow \infty} \sup \frac{1}{N} I(\mathbf{X}^N; \mathbf{Y}^N)$ , where the supremum is taken over all distributions with the average power constraint  $\frac{1}{N} \sum_{k=1}^N \mathbb{E} [\|\mathbf{X}_k\|^2] \leq P$ . Our work focuses on the multiplexing gain  $r$  of such a channel, defined as the pre-log of the capacity  $C(P)$  as  $P \rightarrow \infty$ ,  $r \triangleq \lim_{P \rightarrow \infty} \frac{C(P)}{\log P}$ .

### B. Main results

The main results of this work are summarized as follows, and are illustrated in Fig. 1. First, the case with common phase noise is rather straightforward from [7].

**Proposition 1.** *With common phase noise, i.e.,  $\boldsymbol{\Theta}_t = \boldsymbol{\Theta}_t \mathbf{1}_{n_r \times n_t}$  and  $h(\{\boldsymbol{\Theta}_t\}) > -\infty$ , the multiplexing gain is  $\min\{n_t, n_r\} - \frac{1}{2}$ .*

Then, our new results are on channels with separate phase noises either on the individual paths (model A) or at the input/output (model B) of the channel.

**Theorem 1.** *The multiplexing gain of model A is  $\frac{1}{2}$ .*

The above result shows that extra transmit and receive antennas do not improve the multiplexing gain of a channel with phase uncertainty on each path of the channel. The achievability of the single-antenna case was shown in [3]. Our main contribution is the converse, proved in Section III.

**Theorem 2.** *The multiplexing gain of model B is*

- *upper-bounded by  $\frac{1}{2} \min\{n_t, (n_r - 2)^+ + 1\}$ , and lower-bounded by  $\frac{1}{2} \min\{n_t, \lfloor \frac{n_r + 1}{2} \rfloor\}$  with both transmit and receive phase noises, the upper bound is achievable when  $n_r \leq 3$  or  $n_r \geq 2n_t - 1$ ;*
- *$\min\{\frac{n_r}{2}, \frac{n_t}{2}\}$ , with only transmit phase noise;*
- *$\min\{\frac{n_r}{2}, n_t - \frac{1}{2}\}$ , with only receive phase noise.*

Interestingly, the multiplexing gain of model B depends on the number of transmit and receive antennas differently, which

is rarely the case for previously studied point-to-point MIMO channels.

**Remark II.1.** *As shown in Fig. 1, transmit phase noise is more detrimental than receive phase noise, and strictly so when  $n_r > n_t > 1$ . Intuitively, with transmit phase noise, each transmitted symbol is accompanied by a different phase noise symbol, which means that no more than half of the total spatial degrees of freedom is available for useful signal. On the other hand, with receive phase noise, although half of the received signal dimension is occupied by phase noises, it is enough to increase the number of receive antennas to recover almost all transmitted symbols.*

**Remark II.2.** *Obviously, the multiplexing gain of model B1 is upper-bounded by that of models B2 and B3. Such a “trivial” upper bound is given by  $\min\{\frac{n_t}{2}, \frac{n_r}{2}, n_t - \frac{1}{2}\} = \min\{\frac{n_t}{2}, \frac{n_r}{2}\}$ . When  $n_r \leq n_t$ , the optimal multiplexing gain is  $\frac{n_r}{2}$  with phase noises at either side of the channel, whereas no more than  $\frac{(n_r-2)^++1}{2}$  is achievable with phase noises at both sides. These are the cases for which model B1 is strictly “worse” than both models B2 and B3. When  $n_r \geq 2n_t - 1$ , with transmit phase noise, the optimal multiplexing gain is  $\frac{n_t}{2}$  regardless of the presence of receive phase noise.*

The remainder of the paper is dedicated to outlining the proof of the main results. Due to space limitation, we only consider the memoryless case, which can be shown to be without loss of optimality as far as the multiplexing gain is concerned.

### III. CAPACITY UPPER BOUND

#### A. Capacity Upper Bound for Model A

The main ingredients of the proof are the genie-aided bound and the duality upper bound. In the following, we detail the five steps that lead to the desired result of Theorem 1.

1) *Genie-aided bound:* Let us define the auxiliary random variable  $U$  as the index of the strongest input entry, i.e.,<sup>2</sup>

$$U = \arg \max_{1 \leq i \leq n_t} |X_i|. \quad (2)$$

Thus, we use  $X_U$  to denote the element in  $\mathbf{X}$  with the largest module. It is obvious that  $U \leftrightarrow \mathbf{X} \leftrightarrow \mathbf{Y}$  form a Markov chain, and that  $U$  does not contain more than  $\log n_t$  bits. Assuming that a genie provides  $U$  to the receiver, we have

$$I(\mathbf{X}; \mathbf{Y}) \leq I(\mathbf{X}; \mathbf{Y}, U) \quad (3)$$

$$\leq I(\mathbf{X}; \mathbf{Y} | U) + \log n_t. \quad (4)$$

2) *Canonical form:*

**Definition 1** (Canonical channel). *We define the canonical form  $u$ ,  $u = 1, \dots, n_t$ , of the channel  $\mathbf{H}$  as*

$$\mathbf{G}_{(u)} \triangleq \underbrace{\text{diag}(h_{1,u}^{-1}, \dots, h_{n_r,u}^{-1})}_{\mathbf{A}_u} \mathbf{H}. \quad (5)$$

Note that the elements in the  $u$ th column of  $\mathbf{G}_{(u)}$  has normalized modules. Now, with the information  $U$  from the

<sup>2</sup>When there are more than one such elements, we pick an arbitrary one.

genie, the receiver can convert the original channel into one of the canonical forms, namely, the form  $U$ . Thus,

$$I(\mathbf{X}; \mathbf{H}\mathbf{X} + \mathbf{Z} | U) = I(\mathbf{X}; \mathbf{A}_U \mathbf{H}\mathbf{X} + \mathbf{A}_U \mathbf{Z} | U) \quad (6)$$

$$\leq I(\mathbf{X}; \mathbf{A}_U \mathbf{H}\mathbf{X} + a\mathbf{Z} | U) \quad (7)$$

$$= I(a^{-1}\mathbf{X}; a^{-1}\mathbf{G}_{(U)}\mathbf{X} + \mathbf{Z} | U) \quad (8)$$

$$= I(\tilde{\mathbf{X}}; \mathbf{G}_{(U)}\tilde{\mathbf{X}} + \mathbf{Z} | U), \quad (9)$$

where  $a \triangleq \min_{i,u} |h_{i,u}^{-1}|$ ; (7) is due to the fact that reducing the additive noise can only increase the mutual information; we define  $\tilde{\mathbf{X}} \triangleq a^{-1}\mathbf{X}$ , and accordingly,  $\mathbf{W} \triangleq \mathbf{G}_{(u)}\tilde{\mathbf{X}} + \mathbf{Z}$ . In the following, we derive an upper bound on  $I(\mathbf{X}; \mathbf{W} | U)$ . From now on, we focus on the mutual information

$$I(\tilde{\mathbf{X}}; \mathbf{W} | U) = h(\mathbf{W} | U) - h(\mathbf{W} | \tilde{\mathbf{X}}, U) \quad (10)$$

$$= h(\mathbf{W} | U) - h(\mathbf{W} | \tilde{\mathbf{X}}), \quad (11)$$

where the last equality comes from the fact that  $U$  is a function of  $\mathbf{X}$  and thus a function of  $\tilde{\mathbf{X}}$ , since  $\tilde{\mathbf{X}}$  is simply a scaled version of  $\mathbf{X}$ . Next, we bound  $h(\mathbf{W} | \tilde{\mathbf{X}})$  and  $h(\mathbf{W} | U)$  separately.

3) *Lower bound on  $h(\mathbf{W} | \tilde{\mathbf{X}})$ :*

**Lemma 1.** *For model A, we have*

$$h(\mathbf{W} | \tilde{\mathbf{X}}) \geq n_r \mathbb{E}[\log^+ |\tilde{X}_U|] + n_r \mathbb{E}[\log^+ |\tilde{X}_V|] + c_H, \quad (12)$$

where  $\tilde{X}_U$  and  $\tilde{X}_V$  have the largest and second largest modules in  $\tilde{\mathbf{X}}$ , respectively.

Remark that the above bound depends (not only on the strongest but also on the second strongest input of the channel.

4) *Upper bound on  $h(\mathbf{W} | U)$ :* Upper-bounding  $h(\mathbf{W} | U)$  as a non-trivial function of the input distribution is hard in general. A viable way for that purpose is through an auxiliary distribution, also called the duality approach [1]. Namely, for any<sup>3</sup> pdf  $q(\mathbf{w})$ , we have

$$h(\mathbf{W} | U) = \mathbb{E}[-\log q(\mathbf{W})] - \mathbb{E}_U[D(p_{\mathbf{W}|U=u} \| q)] \quad (13)$$

$$\leq \mathbb{E}[-\log q(\mathbf{W})] \quad (14)$$

due to the non-negativity of the Kullback-Leibler divergence  $D(p_{\mathbf{W}|U=u} \| q)$ . Hence, the key is to choose a proper auxiliary pdf  $q(\mathbf{w})$  to obtain a tight upper bound on the capacity of our channel. The commonly used distributions for MIMO channels are related to the class of isotropic distributions [1], [3], [7]. Unfortunately, the isotropic distributions are not suitable in our case. To see this, let us assume that an isotropic output  $\mathbf{W}$  was indeed close to optimal. On the one hand, the pdf of an isotropic output  $\mathbf{W}$  would only depend on the norm  $\|\mathbf{W}\|$  which is essentially dominated by the largest input entry  $X_U$ . Therefore, the value of  $\mathbb{E}[-\log q(\mathbf{W})]$  would be insensitive to the number of active input entries. On the other hand, the lower bound on the conditional entropy  $h(\mathbf{W} | \tilde{\mathbf{X}})$  is increasing with

<sup>3</sup>Formally, we should say that the probability measure  $Q$  corresponding to the density  $q(\mathbf{w})$  is such that  $P(\cdot | U = u)$  is absolutely continuous with respect to  $Q$ . Throughout the paper, for brevity, we implicitly make the assumption to avoid such formalities.

both of the largest input entries  $X_U$  and  $X_V$ , according to (12). Therefore, the capacity upper bound  $\mathbb{E}[-\log q(\mathbf{W})] - h(\mathbf{W} | \tilde{\mathbf{X}})$  would become larger when the second strongest input went to zero, i.e., only one transmit antenna is active. But this would be in contradiction with the isotropic assumption, since if only one transmit antenna was active, then the output entries would be highly correlated and the output distribution would be far from being isotropic.

In light of the above discussion, we are led to think that a good choice of  $q(\mathbf{w})$  should reflect not only the strongest input entry, but also the weaker ones. We adopt the following pdf built from the multivariate Gamma distribution from [12],

$$q(\mathbf{w}) = \frac{g\boldsymbol{\alpha}}{n_r!} |\hat{w}_1|^{2(\alpha_1-1)} \prod_{i=2}^{n_r} (|\hat{w}_i|^2 - |\hat{w}_{i-1}|^2)^{\alpha_i-1} \cdot \exp(-\mu |\hat{w}_{n_r}|^2) \mu^{\alpha_1+\dots+\alpha_{n_r}}, \quad \mathbf{w} \in \mathbb{C}^{n_r}, \quad (15)$$

where  $\hat{w}_1, \dots, \hat{w}_{n_r}$  are the ordered version of  $w_i$ 's with increasing modules. Essentially, we let each  $W_i$  be circularly symmetric and let the ordered version of  $(|W_1|^2, \dots, |W_{n_r}|^2)$  follow the multivariate Gamma distribution. Applying change of variables and the order statistics (whence the term  $n_r!$ ), we can obtain the pdf of  $\mathbf{W}$  as written in (15). Remarkably, the differences between  $|W_i|^2$  and  $|W_j|^2$ ,  $i \neq j$ , are introduced into the upper bound, which is crucial for bringing in the impact of individual input entries  $\tilde{X}_i$ 's other than the strongest entry as will be shown in the following.

**Lemma 2.** *By choosing  $0 < \alpha_i < 1$ ,  $i = 1, \dots, n_r$ , and  $\mu = \min\{P^{-1}, 1\}$ , we have for model A,*

$$\begin{aligned} & \mathbb{E}[-\log q(\mathbf{W})] \\ & \leq \sum_{i=1}^{n_r} \alpha_i \log^+ P + \left( (1 - \alpha_1) + \sum_{i=1}^{n_r} (1 - \alpha_i) \right) \mathbb{E}[\log^+ |\tilde{X}_U|] \\ & \quad + \sum_{i=2}^{n_r} (1 - \alpha_i) \mathbb{E}[\log^+ |\tilde{X}_V|] + c_H, \end{aligned} \quad (16)$$

where  $\tilde{X}_U$  and  $\tilde{X}_V$  are the strongest and second strongest elements in  $\tilde{\mathbf{X}}$ , respectively.

5) *Upper bound for model A:* Combining (11), (12), (14), and (16) from the previous steps, we get

$$I(\tilde{\mathbf{X}}; \mathbf{W} | U) \leq \left( \sum_{i=1}^{n_r} \alpha_i + \frac{1}{2} \right) \log^+ P + c'_H. \quad (17)$$

Finally, we conclude from (4) and (9) that, for model A,

$$I(\mathbf{X}; \mathbf{Y}) \leq I(\tilde{\mathbf{X}}; \mathbf{W} | U) + c_0 \quad (18)$$

$$\leq \left( \sum_{i=1}^{n_r} \alpha_i + \frac{1}{2} \right) \log^+ P + c_H \quad (19)$$

which implies that the multiplexing gain is upper-bounded by

$$r_A \leq \sum_{i=1}^{n_r} \alpha_i + \frac{1}{2}, \quad \forall \boldsymbol{\alpha} \in (0, 1)^{n_r}. \quad (20)$$

By taking the infimum over  $\boldsymbol{\alpha}$ , we have  $r_A \leq \frac{1}{2}$ .

## B. Capacity Upper Bound for Model B

We only provide the proof for model B1, with both transmit and receive phase noises. The other cases are proved in the full paper [11]. Note that the multiplexing gain of this case is upper-bounded by that of case B2 and case B3, since we can enhance the channel by providing the information on the transmit or receive phase noises to both the transmit and receiver. In other words, the upper bound  $\min\{\frac{n_r}{2}, \frac{n_t}{2}, n_t - \frac{1}{2}\} = \min\{\frac{n_r}{2}, \frac{n_t}{2}\}$  is still valid for this case. In the following, we show that we can tighten the upper bound  $\frac{n_r}{2}$  to  $\frac{(n_r-2)^++1}{2}$  with the duality upper bound using the multi-variate Gamma distribution. This new upper bound turns out to be crucial for the case  $n_t = n_r = 2$  where  $\min\{\frac{n_r}{2}, \frac{n_t}{2}\} = 1$  is not tight and is strictly larger than  $\frac{(n_r-2)^++1}{2} = \frac{1}{2}$ . The proof is in the same vein as the proof for model A. Specifically, the first four steps are exactly the same as for model A, except for Step 3 in which the conditional entropy has a different lower bound, as shown below.

**Lemma 3.** *For model B1, we have*

$$h(\mathbf{W} | \tilde{\mathbf{X}}) \geq n_r \mathbb{E}[\log^+ |\tilde{X}_U|] + \mathbb{E}[\log^+ |\tilde{X}_V|] + c_H, \quad (21)$$

where  $\tilde{X}_U$  and  $\tilde{X}_V$  have the largest and second largest modules in  $\tilde{\mathbf{X}}$ , respectively.

Finally, in Step 5, we apply (11), (14), (16), and (21), and after some manipulations

$$I(\tilde{\mathbf{X}}; \mathbf{W} | U) \leq \left( \frac{(n_r - 2)^+ + 1}{2} + \sum_{i=1}^{n_r} \alpha_i \right) \log^+ P + c'_H. \quad (22)$$

Therefore, we conclude from (4), (9), and (22) that,

$$I(\mathbf{X}; \mathbf{Y}) \leq \left( \sum_{i=1}^{n_r} \alpha_i + \frac{(n_r - 2)^+ + 1}{2} \right) \log^+ P + c'_H \quad (23)$$

from which we obtain the multiplexing gain upper bound  $r_B \leq \frac{(n_r-2)^++1}{2}$ , after taking the infimum over  $\boldsymbol{\alpha}$ .

## IV. CAPACITY LOWER BOUND FOR MODEL B

In this section, we derive a lower bound on the capacity of model B using the class of memoryless Gaussian input distributions. Although the optimal input distribution has been proved to be discrete in [4], the use of a simple Gaussian input provides tight lower bounds on the pre-log, which is enough for our purpose here.

### A. Case B1: Transmit and receive phase noise

In this case, we set  $\mathbf{X} \sim \mathcal{CN}(0, \frac{P}{n_t} \mathbf{I}_{n_t})$ . For convenience, let us rewrite the received signal as

$$\mathbf{Y} = \sqrt{\frac{P}{n_t}} e^{j\tilde{\Theta}_R} \circ (\mathbf{H}(e^{j\tilde{\Theta}_T} \circ \mathbf{X}_0)) + \mathbf{Z} \quad (24)$$

$$= \sqrt{\frac{P}{n_t}} e^{j\tilde{\Theta}_R} \circ \hat{\mathbf{Y}} + \mathbf{Z} = \sqrt{\frac{P}{n_t}} \tilde{\mathbf{Y}} + \mathbf{Z}, \quad (25)$$

where  $\mathbf{X}_0 \sim \mathcal{CN}(0, \mathbf{I}_{n_t})$  is the normalized version of  $\mathbf{X}$ ;  $\tilde{\Theta}_R \triangleq \Theta_R + \Theta_{T,1}$  and  $\tilde{\Theta}_T \triangleq \Theta_T - \Theta_{T,1}$ . Note that  $\tilde{\Theta}_{T,1} = 0$

by definition and  $h(\tilde{\Theta}_R) > -\infty$ . The mutual information of interest can be written as

$$I(\mathbf{X}; \mathbf{Y}) = I(\mathbf{X}, \tilde{\Theta}_T; \mathbf{Y}) - I(\tilde{\Theta}_T; \mathbf{Y} | \mathbf{X}) \quad (26)$$

$$= h(\mathbf{Y}) - h(\mathbf{Y} | \mathbf{X}, \tilde{\Theta}_T) - I(\tilde{\Theta}_T; \mathbf{Y} | \mathbf{X}). \quad (27)$$

First, the following lemma, which provides a lower bound on  $h(\mathbf{Y})$  in (27), is crucial for the achievability proof.

**Lemma 4.** *With receive phase noise such that  $h(\Theta_R) > -\infty$ ,*

$$h(\mathbf{Y}) \geq \left( \frac{n_r}{2} + \frac{1}{2} \min\{n_r, 2n_t - 1\} \right) \log^+ P + c_H. \quad (28)$$

Next, we derive upper bounds on the two negative terms in (27) as follows. The conditional differential entropy can be upper-bounded as

$$h(\mathbf{Y} | \mathbf{X}, \tilde{\Theta}_T) \leq \sum_{k=1}^{n_r} h(Y_k | \mathbf{X}, \tilde{\Theta}_T) \quad (29)$$

$$\leq \frac{n_r}{2} \log^+ P + c_H, \quad (30)$$

where the intuition behind this bound is that given  $(\mathbf{X}, \tilde{\Theta}_T)$  the dominating random part of  $\mathbf{Y}$  comes from the  $n_r$  receive phase noises and scales as  $\frac{n_r}{2} \log P$  when the signal is strong. And

$$I(\tilde{\Theta}_T; \mathbf{Y} | \mathbf{X}) \leq I(\tilde{\Theta}_T; \mathbf{Y}, \tilde{\Theta}_R | \mathbf{X}) \quad (31)$$

$$= I(\tilde{\Theta}_T; \mathbf{Y} | \mathbf{X}, \tilde{\Theta}_R) + I(\tilde{\Theta}_T; \tilde{\Theta}_R) \quad (32)$$

$$= I(\tilde{\Theta}_T; e^{j\tilde{\Theta}_T} \circ \mathbf{X} + \mathbf{H}^+ \mathbf{Z} | \mathbf{X}, \tilde{\Theta}_R) + c_0 \quad (33)$$

$$\leq I(\tilde{\Theta}_T; e^{j\tilde{\Theta}_T} \circ \mathbf{X} + \tilde{\mathbf{Z}} | \mathbf{X}, \tilde{\Theta}_R) + c_0 \quad (34)$$

$$= h(e^{j\tilde{\Theta}_T} \circ \mathbf{X} + \tilde{\mathbf{Z}} | \mathbf{X}, \tilde{\Theta}_R) - h(e^{j\tilde{\Theta}_T} \circ \mathbf{X} + \tilde{\mathbf{Z}} | \mathbf{X}, \tilde{\Theta}_R, \tilde{\Theta}_T) + c_0 \quad (35)$$

$$\leq \frac{n_t - 1}{2} \log^+ P + c_H, \quad (36)$$

where  $\tilde{\mathbf{Z}} \sim \mathcal{CN}(0, \sigma_{\min}^2(\mathbf{H}^+) \mathbf{I}_{n_t})$ , with  $\sigma_{\min}(\mathbf{H}^+)$  being the minimum singular value of  $\mathbf{H}^+$ ; to obtain the last inequality, we use the fact that  $\tilde{\Theta}_{T,1} = 0$  and apply the same reasoning as for (30) on the rest of the  $n_t - 1$  entries of  $\tilde{\Theta}_T$ .

Plugging (30), (36), and (28) into (27), we obtain

$$I(\mathbf{X}; \mathbf{Y}) \geq \frac{1}{2} \min\{n_r - n_t + 1, n_t\} \log^+ P + c_H. \quad (37)$$

Note that the above lower bound holds when we substitute  $n_t$  by any  $n'_t \leq n_t$ , i.e., by activating only  $n'_t$  transmit antennas. It is clear that when  $n_r - n_t + 1 \geq n_t$ , i.e.,  $n_r \geq 2n_t - 1$ , we should let  $n'_t = n_t$ . Otherwise, we should decrease  $n'_t$  to balance between  $n_r - n'_t + 1$  and  $n'_t$ , which gives  $n'_t = \lfloor \frac{n_t + 1}{2} \rfloor$ . This completes the proof of the lower bound for model B1.

### B. Case B2: Transmit phase noise

In this case, we use  $n'_t \triangleq \min\{n_t, n_r\}$  input antennas and deactivate the remaining ones. The active inputs, denoted by  $\mathbf{X}'$ , are i.i.d. Gaussian, i.e., follow  $\mathcal{CN}(0, \frac{P}{n'_t} \mathbf{I}_{n'_t})$ . We rewrite the output vector as  $\mathbf{Y} = \mathbf{H}'(e^{j\Theta'_T} \circ \mathbf{X}') + \mathbf{Z}$  where  $\mathbf{H}'$  is the  $n_r \times n'_t$  submatrix of  $\mathbf{H}$  corresponding to the active inputs, and  $\Theta'_T$  is similarly defined. It follows that

$I(\mathbf{X}'; \mathbf{Y}) = I(\mathbf{X}'; (\mathbf{H}')^\dagger \mathbf{Y})$ . Then, we have  $h((\mathbf{H}')^\dagger \mathbf{Y}) = h(e^{j\Theta'_T} \circ \mathbf{X}' + (\mathbf{H}')^\dagger \mathbf{Z}) = n'_t \log^+ P + c_H$  and  $h((\mathbf{H}')^\dagger \mathbf{Y} | \mathbf{X}') = h(e^{j\Theta'_T} \circ \mathbf{X}' + (\mathbf{H}')^\dagger \mathbf{Z} | \mathbf{X}') \leq h(e^{j\Theta'_T} \circ \mathbf{X}' + \sigma_{\max}((\mathbf{H}')^\dagger) \mathbf{Z} | \mathbf{X}')$ .

The latter is further upper-bounded by  $\frac{n'_t}{2} \log^+ P + c'_H$  according to the same reasoning as for (30). This shows the lower bound  $\frac{1}{2} \min\{n_t, n_r\}$  on the multiplexing gain.

### C. Case B3: Receive phase noise

As in Case B1, we let  $\mathbf{X} \sim \mathcal{CN}(0, \frac{P}{n_t} \mathbf{I}_{n_t})$ . First,  $h(\mathbf{Y})$  is lower-bounded in Lemma 4. Next, following the same steps (29) and (30), we can show that

$$h(\mathbf{Y} | \mathbf{X}) \leq \frac{n_r}{2} \log^+ P + c_H, \quad (38)$$

since we are in the same situation as in Case B1 when  $\tilde{\Theta}_T$  is known. Finally, combining (28) and (38), we get

$$I(\mathbf{X}; \mathbf{Y}) = h(\mathbf{Y}) - h(\mathbf{Y} | \mathbf{X}) \quad (39)$$

$$\geq \frac{1}{2} \min\{n_r, 2n_t - 1\} \log^+ P + c_H. \quad (40)$$

## V. CONCLUSIONS

In this work, we obtained the exact multiplexing gain of the discrete-time stationary and ergodic MIMO phase noise channel in some important cases. Interesting future works include the refined capacity analysis (beyond pre-log) of such channels, as well as the extension of the results to multi-user channels.

## REFERENCES

- [1] A. Lapidoth and S. Moser, "Capacity bounds via duality with applications to multiple-antenna systems on flat-fading channels," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2426–2467, Oct 2003.
- [2] L. Zheng and D. Tse, "Communication on the Grassmann manifold: A geometric approach to the noncoherent multiple-antenna channel," *IEEE Trans. Inf. Theory*, vol. 48, no. 2, pp. 359–383, Feb 2002.
- [3] A. Lapidoth, "On phase noise channels at high SNR," in *Proc. IEEE Information theory workshop*, Oct 2002, pp. 1–4.
- [4] M. Katz and S. Shamai, "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.
- [5] L. Barletta, M. Magarini, and A. Spalvieri, "The information rate transferred through the discrete-time Wiener's phase noise channel," *Journal of Lightwave Technology*, vol. 30, no. 10, pp. 1480–1486, 2012.
- [6] H. Ghozlan and G. Kramer, "Models and information rates for Wiener phase noise channels," *arXiv preprint arXiv:1503.03130*, 2015.
- [7] G. Durisi, A. Tarable, C. Camarda, R. Devassy, and G. Montorsi, "Capacity bounds for MIMO microwave backhaul links affected by phase noise," *IEEE Trans. Commun.*, vol. 62, no. 3, pp. 920–929, March 2014.
- [8] D. M. Arnold, H.-A. Loeliger, P. O. Vontobel, A. Kavčić, and W. Zeng, "Simulation-based computation of information rates for channels with memory," *IEEE Trans. Inf. Theory*, vol. 52, no. 8, pp. 3498–3508, 2006.
- [9] M. R. Khanzadi, G. Durisi, and T. Eriksson, "Capacity of SIMO and MISO phase-noise channels with common/separate oscillators," *IEEE Trans. Commun.*, vol. 63, no. 9, pp. 3218–3231, 2015.
- [10] G. Durisi, A. Tarable, and T. Koch, "On the multiplexing gain of MIMO microwave backhaul links affected by phase noise," in *Proc. IEEE International Conference on Commun. (ICC)*, 2013, pp. 3209–3214.
- [11] S. Yang and S. Shamai (Shitz), "On the multiplexing gain of MIMO phase noise channels," *IEEE Trans. Inf. Theory*, 2016, submitted. [Online]. Available: <http://arxiv.org/pdf/1603.05584v1.pdf>
- [12] A. M. Mathai and P. G. Moschopoulos, "A form of multivariate Gamma distribution," *Annals of the Institute of Statistical Mathematics*, vol. 44, no. 1, pp. 97–106. [Online]. Available: <http://dx.doi.org/10.1007/BF00048672>