The Impact of Phase Fading on the Dirty Paper Channel

Stefano Rini* and Shlomo Shamai (Shitz)†
* Stanford University, Stanford, CA, USA
E-mail: stefano@wsl.stanford.edu
† Technion-Israel Institute of Technology, Haifa, Israel
E-mail: sshlomo@ee.technion.ac.il

Abstract—The impact of phase fading on the classical Costa’s dirty paper coding channel is studied. We consider a variation of this channel model in which the amplitude of the interference sequence is known at the receiver, and the phase is known at the transmitter. Although the capacity of this channel has already been established, it is expressed using an auxiliary random variable and as the solution of a maximization problem. To circumvent the difficulty evaluating capacity, we derive alternative inner and outer bounds and show that the two expressions are to within a finite distance. This provides an approximate characterization of the capacity which depends only on the channel parameters. We consider, in particular, two distributions of the phase fading: circular binomial and circular uniform. The first distribution models the scenario in which the transmitter has minimal uncertainty over the phase of the interference while the second distribution models complete uncertainty. For circular binomial fading, we show that binning with Gaussian signaling still approaches capacity, as in the channel without phase fading. In the case of circular uniform fading, instead, binning with Gaussian signaling is no longer effective and novel interference avoidance strategies are developed to approach capacity.

I. INTRODUCTION

With the increase in network traffic and density, Base Station (BS) cooperation is becoming a common feature of modern cellular communication systems. BS cooperation offers many advantages: for instance, coordinated multipoint transmissions provide crucial coherent combining gains for users on the cell edge. Another advantage provided by BS cooperation is interference pre-cancellation: having knowledge of the interference created by neighboring BSs at the intended receiver, a BS can pre-code its transmission against such interference. The information theoretic model which characterizes the limiting performance of interference pre-cancellation is the Gel’fand-Pinsker (GP) problem [1]. Although well understood in the literature, the GP problem is rarely considered in practical systems. The difficulty in translating this theoretical results into practical transmission strategies partially lies in the idealized assumption that the transmitter has perfect knowledge of the communication channel. Channel knowledge at the BS is particular hard to obtain for different reasons, the main of which is perhaps fading. In this paper, we address the effect of partial transmitter channel knowledge in the presence of phase fading and characterize the optimal transmission strategies for different distributions of the fading realizations.

In the GP channel, a transmitter communicates to a receiver over a channel subject to both noise and state: the state is known non-causally at the transmitter but is not known at the receiver. A variation of the classical GP problem [1] is the model in which the channel state is partially known at the transmitter and partially known at the receiver. The capacity of this more general channel is established by Cover and Chiang in [2]. The GP problem in which the channel output is obtained as a linear combination of the input, the state which models interference and a white Gaussian noise is considered by Costa [3]. For this channel, it is shown that the presence of the interference does not reduce capacity: this celebrated result is known as “writing on dirty paper”. The variation of the writing on dirty paper channel in which fading is added to the interference sequence is known as “writing on fading dirt”. The capacity of this channel is a special case of [2] but its expression contains an auxiliary random variable and is obtained as the solution of a maximization problem. For these reasons, neither closed form expressions nor numerical evaluations of the capacity for the writing on fading dirt problem are not known. Outer and inner bounds to the capacity of the writing on fading dirt channel are derived in [4], [5] while achievable rates under Gaussian signaling and lattice strategies are derived in [6]. An outer bound for the vector writing on fading dirt problem was recently derived in [7].

In the following, we focus on the writing on fading dirt problem for the case in which only phase fading is considered. Additionally, the phase fading process is assumed to be known at the receiver but not at the transmitter. We study the capacity of this channel for two distributions of the phase fading: the circular binomial and circular uniform distribution. The first distribution represents the case in which the uncertainty over the fading process is minimal while the second distribution the case in which it is maximal. In both cases, we derive new inner and outer bounds and show that they lie to within finite additive gap which does not depend on the channel parameters. For the binomial circular distribution, the scheme which approaches capacity relies on binning with Gaussian signaling as in the channel without fading. For the circular uniform distribution, a novel transmission strategy is developed in which the transmitter only uses one dimension to send
The Dirty Paper Channel with Phase Fading (DPC-PF) model. The single line indicates real values while the double line indicates complex values.

**II. DIRTY PAPER CHANNEL WITH PHASE FADEING**

In Dirty Paper Channel with Phase Fading (DPC-PF), also depicted in Fig. 1 the channel output is obtained as

\[ Y_j = X_j + e^{j\theta}S_{Rj} + Z_j, \]

for \( i = \sqrt{-1}, j \in [1 \ldots N] \) and where \( X_j \) is the channel input, \( S_j \) the interference, \( Z_j \) the additive noise and \( \theta_j \) the fading realization. The channel input \( X_j = X_{Rj} + jX_{Ij} \) is subject to the power constraint

\[ E[|X_j|^2] = E[X_{Rj}^2 + X_{Ij}^2] \leq P, \]

and the interference \( S_{Rj} \) is a normal Random Variable (RV) with zero mean and covariance \( Q \), also indicated as \( \mathcal{N}(0, Q) \). The noise term \( Z_j = Z_{Rj} + jZ_{Ij} \) is an iid circular symmetric complex normal RV with zero mean and unitary covariance, also indicated as \( \mathcal{C}\mathcal{N}(0,1) \). The interference sequence \( S_N \) is assumed to be anti-causally available at the transmitter while the phase fading sequence \( \theta_N \) is known at the receiver. The term \( \theta_j \) represents the effect of phase fading on the interference sequence \( S_R \) and is iid draw from the circular distribution \( P_\theta \). In the following we focus on two distributions for \( P_\theta \):

- a circular binomial distribution
  \[ P_\theta(t) = \frac{1}{2} (1_{\{t+\Delta\}}(t) + 1_{\{t-\Delta\}}(t)), \quad \Delta \in [0, \pi/2] \]
- a circular uniform distribution
  \[ P_\theta(t) = \frac{1}{2\pi}, \quad t \in [0, 2\pi) \]

Note that any channel with a circular binomial phase fading can be reduced to the distribution in (3) without loss of generality by pre-rotating the channel input and rotating the channel output. The fading model in (3) is usually referred to as ergodic fading or fast fading, since the fading realization changes at each channel use in a memoryless fashion. This model represents a worst-case scenario: models is which the fading process has memory over the channel uses and vary with less randomness can be obtained from (3) by providing the transmitter with a genie-aided side information on the fading process.

**III. RELATED RESULTS**

The capacity for the DPC-PF in (1) is a special case of the result in (2).

**Theorem III.1. Capacity of the DPC-PF [2, Th. 1]**

The capacity of the channel in (1) is obtained as

\[ C = \max_{R \in X \cup S} \min \left\{ I(Y;U|\theta) - I(U;S_R) \right\} \]

The result in Th. III.1 holds for the general Gelf’and-Pinsker problem with partial channel state information at either the receiver or the transmitter but is stated in Th. III.1 only for the model in (1).

Equation (2) contains the auxiliary RV \( U \) and is expressed as the maximization over \( P_{U|X|S} \). This expression is concave in \( P_{U|S_R} \) for a fixed \( P_{X|S_R,U} \) and convex in \( P_{X|S_R,U} \) for a fixed \( P_{U|S_R} \), which implies that \( X \) can be chosen to be a deterministic function of \( U \) and \( S \). Given the fact that (2) contains an auxiliary RV and given its convexity properties, it is not easy to obtain an expression of \( C \) which depends solely on the channel parameters or to numerically approximate it. For this reason, alternative inner and outer bounds have been derived in the literature. In (9), the RHS (9) is optimized for the case in which \( U \) and \( X \) are restricted to be Gaussian.

**Theorem III.2. Achievability with Gaussian signaling [9, Sec. IV],[6, Th. 1]**

Let \( \varphi = (\rho_{xs}, \rho_{us}, \rho_{ux}) \) and let \( \mathcal{A} \) denote the region

\[ \mathcal{A} = \left\{ \begin{array}{l} \rho_x < 1, \quad t \in \{xs, us, ux\} \\ 1 + 2\rho_{as} \rho_{ux} - \rho_{as}^2 - \rho_{ux}^2 = 0 \end{array} \right\} \]

then, any distribution of \( P_\theta \), the following rate is achievable

\[ R \leq \max_{\rho \in \mathcal{A}} \mathbb{E}_{\theta}[R_1(\rho, a)|\theta = a], \]

for

\[ R_1(\rho, t) = \frac{1}{2} \log \left( (P + Q + 2\rho |\rho_{as}| \sqrt{PQ} + 1)(1 - |\rho_{as}|^2) \right) \]

\[ - \frac{1}{2} \log \left( (P(1 - |\rho_{xx}|^2) + Q(1 - |\rho_{xx}|^2) + 2|\rho_{as} - \rho_{as} | \sqrt{PQ} + 1 \right), \]

**Proof:** The proof can be obtained from [6, Th. 1] by noticing that the realization of \( \theta \) which corresponds to the lowest achievable rate is \( \theta = \angle(X) \).

An outer bound for the case where the fading is uniformly distributed among two values can be obtained from the “carbon copying onto dirty paper” [10] problem.

**Theorem III.3. Outer Bound for the Circularly Binomial Fading Dirt Channel [10, Th. 5]**
The capacity of the DPC-PF with the distribution of $\theta$ in $[\alpha, \beta]$ is upper bounded as
\[
C \leq \frac{1}{2} \log(1 + P) + \frac{1}{2} \log \left(1 + (\sqrt{P} + \sqrt{Q})^2\right) - \frac{1}{4} \log(4\sin(\Delta)^2Q).
\] (9)

Proof: This result is a variation of the result in [10, Th. 5] for $S_1 = e^{\Delta S_R}$ and $S_2 = e^{-\Delta S_R}$. The full proof is provided in Appendix A.

The result [10, Th. 5] was originally developed for the case in which the fading coefficient is fixed through successive channel uses. The result in Th. IV.3 is obtained by adapting the derivation in [10, Th. 5] to the case of in which the fading changes at each channel use.

IV. Circular Binomial Phase Fading

We begin by analyzing the scenario in which the phase fading takes only two values. Since the uncertainty on the fading realization is limited, the encoder can efficiently cope with the interference through binning and Gaussian signaling as in the channel without fading. We begin by introducing the outer bound inspired by the “carbon copying onto dirty paper” of [10]. The derivation is improved upon through a genie aided side information and by optimizing the outer bound over the power of the interference.

Theorem IV.1. Genie Aided Outer Bound

The capacity of the DPC-PF with the distribution of $\theta$ in $[\alpha, \beta]$ is upper bounded as
\[
C \leq R^{\text{OUT-B}} = \min_{\gamma \in [0,1]} \{Q, \rho, c_+, c_-, c_5 \in \mathbb{R} \} \left(\frac{1}{2} \log(T_1T_2) + \frac{1}{4} \log(T_1T_4) + 1, \right.
\]
with $\mathcal{A} = \{Q \leq Q, \rho \in [-1,1], c_+, c_-, c_5 \in \mathbb{R} \}$, and for
\[
T_1 = \left(1 + \gamma \sqrt{\frac{P}{Q}}\right)^2 Q + 1
\]
\[
T_2 = P(1 - \gamma^2) + 1 - \frac{(c_+ - c_-)^2(1 - \rho)^2}{c_+^2 + c_-^2 + 2\rho c_+ c_- + 1}
\]
\[
T_3 = 4\sin(\Delta)^2Q' + 2(1 - \rho) + \frac{(2\sin(\Delta)c_5Q' + (c_+ - c_-)(1 - \rho))^2}{c_5^2Q' + c_+^2 + c_-^2 + 2\rho c_+ c_- + 1}
\]
\[
T_4 = 2(1 + \rho) - \frac{(c_+ + c_-)^2(1 + \rho)^2}{c_5^2Q' + c_+^2 + c_-^2 + 2\rho c_+ c_- + 1}.
\]

Proof: The proof follows the same line as [10] but with two further refinements. The receiver is provided with a genie-aided side information which is obtained as a linear combination of the interference and the channel noise and an additional noise term, independent from all the other RV. Moreover, the outer bound is optimized over the power of the interference in the range $[0, Q]$. This is possible since the capacity of the channel increases as the power of the interference decreases. The full proof is provided in Appendix B.

The outer bound in Th. IV.1 is expressed as the optimization over multiple parameters and as the union over all $\gamma$. We now derive a simpler outer bound which is expressed only as a function of the channel parameters.

Lemma IV.2. Simpler Outer Bound

If $\pi/4 \leq \Delta \leq \pi/2$, the outer bound in Th. IV.1 can be further upper bounded as
\[
C \leq R^{\text{OUT-APP-B}} = \begin{cases} \log(P + 1) + 2 & \sin(\Delta)^2Q \leq 1 \\ \log(P + 1) + 1 & \sin(\Delta)^2Q \geq P + 1 \\ \frac{1}{2} \log(1 + (\sqrt{P} + \sin(\Delta)\sqrt{Q})^2) & 1 < \sin^2(\Delta)Q < P + 1 \end{cases}
\] (12)

Proof: The proof is similar to the proof in Th. IV.1 but does not consider the genie aided side information. The fundamental improvement from the proof in [10] is the optimization over the power of the interference. While the capacity of the channel increases as $Q$ decreases, the outer bound is not monotonically decreasing in $Q$. For this reason, optimizing the outer bound over $Q$ in the range $[0, Q]$ produces a tighter outer bound than [10].

We now derive an inner bound to the capacity region based on binning and Gaussian signaling as in the original DPC channel. If the transmitter disregards the uncertainty over the phase fading and codes as in the DPC channel, it can pre-code successfully against the interference only half of the time on average. Alternatively, the encoder can disregard the partial interference knowledge and transmit as if the interference were additional additive noise. A scheme that combines the above two choices can be obtained by using two codewords to produce the channel input: one codeword is pre-coded against one realization of the interference while another codeword treats the interference as noise. The performance of this scheme can then be optimized over the power allocated to the two codewords.

Theorem IV.3. Interference as Noise and Binning Inner Bound

The capacity of the DPC-PF with the distribution of $\theta$ in $[\alpha, \beta]$ is lower bounded as
\[
C \geq R^{\text{INOMIAL}} = \frac{1}{2} \log \left(1 + \frac{\alpha \beta P}{1 + \alpha \beta P + \sin(\Delta)^2Q}\right) + \frac{1}{4} \log \left(1 + \frac{\alpha \beta P + 1}{\alpha \beta P + \sin(\Delta)^2Q + 1}\right) + \frac{1}{4} \log \left(1 + \frac{\alpha \beta P + 1}{\alpha \beta P + 2\sin(\Delta)^2Q + 1}\right),
\] (13)
for any $\alpha, \beta \in [0, 1]$ and $\alpha = 1 - \alpha, \beta = 1 - \beta$.

Proof: On the real dimension, the interference sequence is always $\cos(\Delta)S_R$, so the encoder can pre-code against this
interference sequence as in the GP problem. On the imaginary dimension, the interference is \( \sin(\Delta)S_R \) half of the time and \( -\sin(\Delta)S_R \) the other half of the time. On the imaginary dimension then, the encoder transmits two codewords, one which pre-codes against \( \sin(\Delta)S_R \) while another codeword treats \( \sin(\theta)S_R \) as additional interference. Since \( \sin(\theta) \) is uniformly distributed over \( \{-\sin(\Delta), +\sin(\Delta)\} \), \( \sin(\theta)S_R \) is Gaussian distributed. This transmission scheme can be optimized over two parameters: \( \beta \), the ratio of the power used in the real versus imaginary dimension and \( \alpha \), the ratio of the power assigned to each codeword in the imaginary dimension. The full proof is provided in Appendix F.

The inner bound in Th. IV.3 is a function of two parameters, \( \alpha \) and \( \beta \): a simpler inner bound expression can be obtained by carefully choosing the values of these two parameters.

**Lemma IV.4. Simpler Inner Bound**
The inner bound of Th. IV.3 can be further lower bounded as

\[
C \leq R^{\text{IN-APP-B}} = \frac{1}{2} \log \left( 1 + P \right) + \frac{1}{2} \log \left( 1 + P + Q + 2\sqrt{PQ} \right) - \frac{1}{2} \log (Q + 1) + 3/2. \tag{15}
\]

**Proof:** The proof relies in adding a conditioning of a negative entropy term over the phase of the channel input. This conditioning is used to divide the channel output in two components: one affected by the channel input, interference and noise, and another one only affected by the interference and noise. This latter term can be easily evaluated since it is composed of a random mixture of iid Gaussian components. The complete proof can be found in Appendix H.

**Remark V.2.** The outer bound in Th. IV.1 can be tightened in the spirit of Th. IV.1 by providing a genie aided side information. With this approach one obtains an outer bound expression in the spirit of (10) which can be optimized over the parameters in the side information.

We next turn to the derivation of an inner bound for the circular uniform phase fading. When the phase fading values are uniformly distributed over a large set, binning with Gaussian signaling provides only marginal advantages. In particular, consider the Costa’s dirty paper channel \( Y = X + aS + Z \) with the assignment \( U = X + \lambda S \). The achievable rate as a function of \( \lambda \) is

\[
R \leq \log \left( \frac{P + a^2Q + 1}{1 + \frac{2a^2}{\lambda^2}P(a - \lambda)^2} \right). \tag{16}
\]

Let’s assume that the transmitter has an incorrect estimate of \( a \) and thus performs dirty paper coding for the gain \( a + \epsilon \), instead of \( a \). In this case the attainable rate is

\[
R \leq \log \left( \frac{P + a^2Q + 1}{P + a^2Q + 1 + QP\epsilon} \right), \tag{17}
\]

so the achievable rate goes quickly to zero as the product \( PQ\epsilon^2 \) increases. For this reason, binning with Gaussian signaling is not beneficial when the phase is circular uniform distributed, as even a small uncertainty over the exact channel realization drastically decreases the rate when \( P \) and \( Q \) are large. Binning achieves capacity for this channel model, but only for Gaussian signaling it is possible to easily evaluate the achievable region. Instead of focusing on determining a good assignment for \( P_{U|X}^{\text{IN}} \), we consider a different achievable scheme in which the imaginary dimension of the channel output is used to estimate the interference sequence and subtract it from the real dimension. By combining the estimate of the interference dimension over the imaginary axes and the information transmitted over the real axe, the transmitter obtains an equivalent channel output which corresponds to a real fading channel.

**Theorem V.3. Real Transmission Inner Bound**
The capacity of the DPC-PF with the distribution of \( \theta \) in H is lower bounded as

\[
C \geq R^{\text{OUT-C}} = \frac{1}{2} \log (1 + Q + a\lambda) + \frac{1}{2} \log (\lambda P + 1) - 3, \tag{18}
\]

for any \( \alpha \in [0,1] \).

**Proof:** The transmitter sends a codeword which threatens the interference as noise on the imaginary dimension. After this

\[
\frac{1}{2} \log \left( 1 + P + Q + 2\sqrt{PQ} \right) - \frac{1}{2} \log (Q + 1) + 3/2.
\]

We next focus on the case where the phase fading is uniformly distributed over the unitary circle. For this scenario, both inner and outer bound in Sec. IV are no longer effective and new results are necessary to characterize capacity. We begin by deriving an outer bound in which a genie provides the receiver with the phase of the channel input.

**Theorem V.1. Outer Bound**
The capacity of the DPC-PF with the distribution of \( \theta \) in H is upper bounded as

\[
C \leq R^{\text{OUT-C}} = \frac{1}{2} \log (1 + P) +
\]

\[
\frac{1}{2} \log \left( 1 + P + Q + 2\sqrt{PQ} \right) - \frac{1}{2} \log (Q + 1) + 3/2. \tag{15}
\]
codewords has been decoded, an estimate of the interference is produced. More specifically, an estimate of \( \sin(\theta^N)S^N_R \) is obtained from \( \sin(\theta^N)S^N_R + Z_I \). Successively, the receiver estimates the real part of the channel input \( X^N_R \) from

\[
\bar{Y}^N = \sin(\theta^N)Y^N_R - \cos(\theta^N)(\sin(\theta^N)S^N_R + Z_I) \tag{19a}
\]

\[
= \sin(\theta^N)X^N_R + \sin(\theta^N)Z_R - \cos(\theta)Z_I \tag{19b}
\]

\[
= \sin(\theta^N)X^N_R + \bar{Z} \tag{19c}
\]

for \( \bar{Z} \sim A(0,1) \). The sequence \( \bar{Y}^N \) in therefore equivalent to the output of a real fading channel with fading coefficient \( \sin(\theta) \). The complete proof is provided in Appendix I.

We next show a gap between inner and outer bounds.

**Theorem V.4. Finite Gap between Inner and Outer Bounds**

The gap between the inner bound in Th. IV.3 and in the outer bound in Th. V.1 is at most 5.5 bits/Hz.

**Proof:** The difference between the expression in (15) and the expression in (13) for \( \alpha = 1/2 \) is \( 3 + 5/2 = 5.5 \) bits/Hz.

The result of Th. V.4 clearly implies that interference precancellation is no longer useful when the transmitter has complete ignorance on the phase of the interference. Instead, in the high interference regime, capacity is achieved by sacrificing half of the signal space to the estimation of the interference sequence.

VI. NUMERICAL SIMULATIONS

We now numerically simulate results of the previous sections to provide some further insight on the problem at hand. In Figure 2 we plot the inner bound in Th. IV.3 and Lem. IV.4 and the outer bounds of Th. IV.1 in Lem. IV.2 for different values of \( P \) and \( Q = 10P \). In the figure we also plot the trivial outer bound \( R \leq \frac{1}{2} \log(1 + P) \), which is obtained by providing the interference sequence \( S^N_R \) to the receiver, and the trivial inner bound \( R \leq \frac{1}{2} \log(1 + P + Q) + \frac{1}{2} \log(1 + Q) \), which is obtained by treating the interference as noise. The constant gap result in Th. IV.3 is obtained comparing Lem. IV.4 and Lem. IV.2 but numerical simulations actually show a much smaller gap between more general inner and outer bounds. In Figure 3 we plot the inner and outer bound in Th. IV.3 and Th. IV.1 for increasing \( P \) and different scaling \( Q \).

VII. CONCLUSION

In this paper, we analyze the effect of phase fading in the classical Costa’s dirty paper channel. We consider a variation of the original setting in which the amplitude of the interfering sequence is known at the transmitter while its phase in known at the receiver. Although derived in the literature, the capacity of this channel is hard to characterize in closed form or through numerical simulations. We, therefore, derive the approximate characterization of capacity for the case in which the phase of the interference has a circular binomial and a circular uniform distribution.

REFERENCES

[1] S. Gel’fand and M. Pinsker, “Coding for channel with random parameters,” Problems of control and information theory, vol. 9, no. 1, pp. 19–31, 1980.

[2] T. M. Cover and M. Chiang, “Duality between channel capacity and rate distortion with two-sided state information,” IEEE Trans. Inf. Theory, vol. 48, no. 6, pp. 1629–1638, 2002.

[3] M. Costa, “Writing on dirty paper,” IEEE Trans. Inf. Theory, vol. 29, no. 3, pp. 439–441, 1983.

[4] P. Grover and A. Sahai, “On the need for knowledge of the phase in exploiting known primary transmissions,” in IEEE International Symposium on the New Frontiers in Dynamic Spectrum Access Networks (DySPAN), 2007, pp. 462–471.

[5] ——, “Writing on rayleigh faded dirt: a computable upper bound to the outage capacity,” in Proc. IEEE International Symposium on Information Theory (ISIT), Nice, France, 2007, pp. 2166–2170.

[6] Y. Avner, B. M. Zaidel, S. Shamai, and U. Erez, “On the dirty paper channel with fading dirt,” in Proceedings of the 26th Convention of IEEE Israel (IEEEI), Eliat, Israel, Nov. 17-20, 2010, pp. 525–529.

[7] D. T. Kao and A. Sabharwal, “An upper bound on the capacity of vector dirty paper with unknown spin and stretch,” arXiv:1305.3934, 2013.

[8] S. Rini and S. Shamai, “The impact of phase fading on the dirty paper channel,” arXiv: UPDATE, 2014.

[9] A. Bennatan and D. Burshtein, “On the fading-paper achievable region of the fading mimo broadcast channel,” IEEE Trans. Inf. Theory, vol. 54, no. 1, pp. 100–115, 2008.

[10] A. Khisti, U. Erez, A. Lapidoth, and G. W. Wornell, “Carbon copying onto dirty paper,” IEEE Trans. Inf. Theory, vol. 53, no. 5, pp. 1814–1827, 2007.
Appendix

A. Proof of Th. III.4

From Fano’s inequality we have
\[ R \leq I(Y^N;W|\theta^N) \]
\[ = \sum_{j=1}^{N} I(Y_j;W|\theta^N,Y^{j-1}) \]
\[ \leq \sum_j (H(Y_j|\theta_j) - H(Y_j|W,\theta^N,Y^{j-1})). \]

For the term \( H(Y_j) \) we have
\[ H(Y_j) = H(X_j + e^{i\theta_j}S_{Rj} + Z_j|\theta_j) \]
\[ = \frac{1}{2} \left( H(X_j + e^{i\Delta}S_{Rj} + Z_j) + H(X_j + e^{-i\Delta}S_{Rj} + Z_j) \right) \]
\[ = \frac{1}{2} \left( H(e^{i\Delta}X_j + S_{Rj} + Z_j) + H(e^{-i\Delta}X_j + S_{Rj} + Z_j) \right) \]
\[ \leq \frac{1}{2} \left( H(\text{Re}(e^{i\Delta}X_j) + S_{Rj} + Z_{Rj}; \text{Im}(e^{i\Delta}X_j) + Z_{Rj}) \right) \]
\[ \leq \frac{1}{2} \log(1 + P) + \frac{1}{2} \log \left( \frac{1}{\sqrt{2}} + \sqrt{Q} \right)^2. \]

Similarly, for the term \( H(Y_j|W,\theta^N,Y^{j-1}) \) we have
\[ H(Y_j|W,\theta^N,Y^{j-1}) = \frac{1}{2} \left( H(X_j + e^{i\Delta}S_{Rj} + Z_j|W,\theta^{j-1},\theta^N,\theta_{j+1}^{N-1},Y^{j-1}) \right) \]
\[ \leq \frac{1}{2} H(X_j + e^{i\Delta}S_{Rj} + Z_j, X_j + e^{-i\Delta}S_{Rj} + Z_j|W,\theta^{j-1},\theta^N, \theta_{j+1}^{N-1},Y^{j-1}) \]
\[ = \frac{1}{2} H \left( \frac{1}{\sqrt{2}} 2 \sin(\Delta)S_{Rj}, \frac{1}{\sqrt{2}} (2X_j + 2 \cos(\Delta)S_{Rj} + 2Z_j)|W,\theta^{j-1},\theta^N, \theta_{j+1}^{N-1},Y^{j-1} \right) \]
\[ \leq \frac{1}{2} H \left( \sqrt{2} \sin(\Delta)S_{Rj} \right) + \frac{1}{2} H \left( \frac{1}{\sqrt{2}} (2X_j + 2 \cos(\Delta)S_{Rj} + 2Z_j)|W,\theta^{j-1},\theta^N, \theta_{j+1}^{N-1},Y^{j-1},S_{Rj} \right) \]
\[ \leq \frac{1}{2} H \left( \sqrt{2} \sin(\Delta)S_{Rj} \right) + \frac{1}{2} H \left( \frac{1}{\sqrt{2}} (2X_j + 2 \cos(\Delta)S_{Rj} + 2Z_j)|W,\theta^{j-1},\theta^N, \theta_{j+1}^{N-1},S_{R}^{N-1} \right) \]
\[ \leq \frac{1}{4} \log 2\pi e Q + \frac{1}{2} H \left( \frac{1}{\sqrt{2}} (2X_j + 2 \cos(\Delta)S_{Rj} + 2Z_j)|W,\theta^{j-1},\theta^N, \theta_{j+1}^{N-1} \right) \]
\[ \leq \frac{1}{4} \log 2\pi e Q + \frac{1}{2} H \left( \sqrt{2} Z_j \right) \]
\[ \leq \frac{1}{4} \log 2\pi e(4Q). \]

By combining the terms in (21f) and (22f), we obtain (9).

B. Proof of Th. IV.1

There are three components to the outer bound which we separately before the actual proof
• the capacity of the channel is decreasing in \( Q \), the power of the interference \( S_R \),
• the correlation among the noise terms can be chosen as a function of \( \theta_j \),
• the receiver is provided with a genie aided side information \( U_j \).

The capacity is decreasing in \( Q \)
Consider two sequences \( S_{1Rj}^N \) and \( S_{2Rj}^N \) for two independent \( S_{mRj} \sim i.i.d. \mathcal{N}(0,Q_m), m \in \{1,2\} \), \( j \in [1 \ldots N] \) with \( Q = Q_1 + Q_2 \).
The interference sequence \( S_R^N \) can be equivalently written as:
\[ S_{Rj} = S_{1Rj} + S_{2Rj}, \quad j \in [1 \ldots N]. \]
Providing $S^N_1$ to both the transmitter and receiver can only increase the capacity, since they can both disregard this extra information. The capacity of the channel in which $S^N_1$ is provided to both encoder and decoder follows in the class of channels studied in [2, Th. 1]. Capacity is thus obtained as

$$C = \max_{X,U|S^N_2} I(X + S_1e^{i\theta} + Z, \theta, S_2^R; U) - I(U; S, S^R_2)$$  

(24a)

$$= \max_{X,U|S^N_2} I(X + S_1e^{i\theta} + Z, \theta, S_2^R; U) - I(U; S_1, S_2^R)$$  

(24b)

$$\leq \max_{X,U|S^N_2} I(X + S_1e^{i\theta} + Z, \theta; U|S_2^R) - I(U; S_2^R; S_1)$$  

(24d)

$$= \max_{X,U|S^N_2} I(X + S_1e^{i\theta} + Z, \theta; U) - I(U; \bar{U})$$  

(24e)

where, in (24d), we have used the independence of $S_1$ and $S_2^R$ and for $\bar{U} = [U, S_2^R]$ in (24e). Since $S_2^R$ does no longer appear in (24e), we conclude that it can be dropped from the maximization.

The expression in (24e) corresponds to the capacity of the channel in (1) in which the interference has power $Q_1$ instead of $Q$. This shows that the capacity of the channel in (1) is decreasing in $Q$.

**Correlation among the noise terms**

As in [10], we notice that the joint distribution among the noise term $Z_j$ in (1) can be chosen to depend on the realization of $\theta_j$, that is

$$P_{Z_j, \theta_j} = P_{\theta_j} \left( \frac{P_{Z_j|\theta_j = +\Delta} + P_{Z_j|\theta_j = -\Delta}}{2} \right)$$  

(25)

for two Gaussian RV with zero mean and unitary variance $Z_j|\theta_j = +\Delta$ and $Z_j|\theta_j = -\Delta$ that have any desired correlation. This holds since the channel transition probability

$$P_{Y_j|Z_j} = P_{\theta_j}P_{Y_j|\theta_j}$$  

(26)

is unaffected by the correlation between the RVs $Z_j|\theta_j = +\Delta$ and $Z_j|\theta_j = -\Delta$.

For the sake of convenience we use the notation

$$Z_j^+ = Z_j|\theta_j = +\Delta$$  

(27a)

$$Z_j^- = Z_j|\theta_j = -\Delta.$$  

(27b)

and indicate with $\rho$ the correlation between the terms. In general $\rho$ can be taken complex. In the following we focus on the case where $\rho$ is real, in which case imaginary parts are independent from the real parts.

**Genie aided side information**

In the outer bound, the receiver is provided with a genie aided side information $U^n$ which is obtained as a linear combination of $S_{R_j}, Z_{+j}$ and $Z_{-j}$, that is

$$U_j = c_S S_{R_j} + c_{+Z} Z_{+j} + c_{-Z} Z_{-j} + Z_j,$$  

(28)

for some iid $Z_j \sim \mathcal{N}(0, 1): U^n$ and for some $c_S, c_Z \in \mathbb{R}$.

We now proceed with the derivation of the actual outer bound, starting from Fano’s inequality:

$$N(R - \varepsilon_N) \leq I(Y^n, \theta^n; W)$$  

(29a)

$$\leq H(Y^n, U^n; W|\theta^n)$$  

(29b)

$$= H(Y^n; W|\theta^n, U^n)$$  

(29c)

$$= \int (H(Y^n|U^n, \theta^n = \phi^n) - H(Y^n|U^n, \theta^n = \phi^n)) dP_{\phi}^N,$$  

(29d)

where $\phi^n \in \{+\Delta, -\Delta\}^N$.

For the positive entropy term $H(Y^n|U^n, \theta^n = \phi^n)$ in (29d), we have

$$H(Y^n|U^n, \theta^n = \phi^n) = \sum_{j=1}^{N} H(Y_j|U^n, \theta^n = \phi^n, Y^{j-1})$$  

(30a)

$$\leq \sum_{j=1}^{N} H(Y_j|U_j, \theta_j = \phi_j)$$  

(30b)
where (30b) follows from the conditioning reduces entropy property of the mutual information, (30c) is obtained by choosing the \( m \) which maximizes the term \( H(Y_j|U_j, \theta_j = \phi_j) \) over all \( j = [1 \ldots N] \).

In the following we drop the subscript \( m \) for ease of notation.

\[
H(X + e^{\Delta S} R + Z_+ | U) = H(e^{-\Delta X} + S_R + e^{-\Delta Z_+} | U) \\
= H(-\sin(\Delta)X + S_R + Z_{R+} | U) \\
= H(-\sin(\Delta)X + S_R + Z_{R+}, \cos(\Delta)X + Z_{t+} | U)
\]

where in (31b) we have used the fact that the noise is circularly symmetric. The choice of \( X \) which maximizes (31c) is of the form

\[
X_G = e^\Delta \left( \gamma \sqrt{\frac{P}{Q}} S_R + \bar{X}_G \sqrt{P(1 - \gamma^2)} \right)
\]

for some \( \gamma \in [0, 1] \) of some \( \bar{X}_G \sim \mathcal{C}(0, 1) \). With this choice we have

\[
H(-\sin(\Delta)X + S_R + Z_{R+}, \cos(\Delta)X + Z_{t+} | U) = H \left( \left( 1 + \gamma \sqrt{\frac{P}{Q}} \right) S_R + Z_{R+} | U \right) + H \left( \sqrt{P(1 - \gamma^2)} \bar{X} + Z_{t+} | U \right) \\
= \frac{1}{2} \log(2\pi eT_1) + \frac{1}{2} \log(2\pi eT_2)
\]

where \( T_1 \) is obtained as

\[
T_1 = \text{Var} \left[ (1 + \gamma \sqrt{\frac{P}{Q}}) S_R + Z_{R+} | U \right] \\
= \text{Var} \left[ (1 + \gamma \sqrt{\frac{P}{Q}}) S_R + Z_{R+} | X_S + cS_R + c_R Z_{R+} + c_{-} Z_{R-} + \bar{Z}_R \right] \\
= (1b)
\]

where (34b) follows from the fact that \( \rho \) is positive and real and imaginary parts are independent. The term \( T_2 \) is obtained as

\[
T_2 = \text{Var} \left[ \sqrt{P(1 - \gamma^2)} \bar{X}_G + Z_{t+} | U \right] \\
= \text{Var} \left[ \sqrt{P(1 - \gamma^2)} \bar{X}_G + Z_{t+} | Z_{t+} + Z_{t-} + \bar{Z}_I \right] \\
= (11b)
\]

The term \( H(X_m + e^{-\Delta S} R_m + Z_{-m} | U_m) \) is bounded in an analogous manner to yield the same expression.

Let’s now focus on the negative entropy term:

\[
- \int H(Y^N|W, U^N, \theta^N = \phi^N) dP^N \phi \\
= - \frac{1}{2} \int \left( H(Y^N|W, U^N, \theta^N = \phi^N) + H(Y^N|W, U^N, \theta^N = -\phi^N) \right) dP^N \phi \\
\leq - \frac{1}{2} \int \left( H(X^N + e^{i\phi} S_R^N + Z_+^N \phi + X_N + e^{-i\phi} S_R^N + Z_-^N \phi | W, U^N) \right) dP^N \phi \\
= - \frac{1}{2} \int \left( H(2i\sin(\phi^N) S_R^N + Z_+^N \phi - Z_-^N \phi, 2X_N + 2 \cos(\phi^N) S_R^N + Z_+^N \phi - Z_-^N \phi | U^N, W) \right) dP^N \phi - N \log 2 \\
\leq - \frac{1}{2} \int \left( H(2i\sin(\phi^N) S_R^N + Z_+^N \phi - Z_-^N \phi | U^N) \\
- H(2X_N + 2 \cos(\phi^N) S_R^N + Z_+^N \phi - Z_-^N \phi | U^N, W, 2 \sin(\phi^N) S_R^N + Z_+^N \phi + Z_-^N \phi) \right) dP^N \phi - N \log 2 \\
\leq - \frac{1}{2} \int \left( NH(2i\sin(\phi) S_R j + Z_+ \phi j - Z_- \phi j | U_j) - H(2X_N + \cos(\phi) S_R j + Z_+ \phi j - Z_- \phi j | U^N, W, S_R^N) \right) dP^N \phi - N \log 2 \\
= - \frac{N}{2} \int \left( H(2\sin(\phi) S_R j + Z_+ \phi j - Z_- \phi j | U_j) - H(Z_+ \phi j + Z_- \phi j | U_j) \right) dP_{\phi j} - N \log 2 \\
= - \frac{N}{2} \int \left( H(2\sin(\Delta) S_R + Z_+ - Z_- | U) - H(Z_+ \phi + Z_- \phi | U) \right) dP_{\phi} - N \log 2
\]
where (36b) is obtained by paring each sequence $\phi^N$ with the complement sequence $-\phi^N$, in which each $+\Delta$ is replaced by a $-\Delta$ and each $-\Delta$ by a $+\Delta$. In (36c) we define $Z^N_\phi$ as the sequence of noise terms associated with the sequence of phase fading values $\phi^N$. The passage in (36d) is obtained with the transformation

$$H(U_1, U_2) = H(U_1 + U_2, U_1 - U_2) - \frac{1}{2}.$$  

(37)

In (36e) we have used the fact that the state and the noise are iid for the first term and the “conditioning reduces entropy” property of the entropy for the second term. In (36f) we have used the fact that the expression on longer depends on $j$ or $\theta_j$.

The terms $T_3$ and $T_4$ are obtained as

$$T_3 = \text{Var}[2i\sin(\Delta)S_R + Z_+ - Z_-|U] = 11e$$

(38a)

$$T_4 = \text{Var}[Z_+ - Z_-|U] = 11d$$

(38b)

C. Proof of Lem. IV.2

The proof we consider is a variation of the proof of Th. IV.1 in App. B.

We consider the case where no side information is provided at the receiver but still optimize over the correlation between the noise terms and over the power of the interference.

For the positive entropy term in $H(Y^N|U^N, \theta^N = \phi^N) = H(Y^N|\theta^N = \phi^N)$ in (29d) we have

$$H(Y^N|\theta^N = \phi^N) \leq NH(X_j + e^{\theta_j}S_{Rj} + Z_j|\theta_j = \phi_j)$$

(39a)

$$\leq \frac{N}{2} \left( H(X_j + e^{\theta_j}S_{Rj} + Z_j) + H(X_j + e^{-\theta_j}S_{Rj} + Z_j) \right).$$

(39b)

When $\pi/4 \leq \Delta \leq \pi/2$, we have that $\cos(\Delta) \leq \sin(\Delta)$, we can write

$$H(X + e^{\theta}S_R + Z) = H(X_j + \sin(\theta)S_R + Z_j; X_j + \cos(\theta)S_R + Z)$$

(40a)

$$\leq \frac{1}{2} \log \left( 1 + P + \sin(\Delta)^2Q + 2\sqrt{PQ} \right) + H(X_R + \cos(\theta)S_R + Z_R|X_j + \sin(\theta)S_R + Z)$$

(40b)

$$= \frac{1}{2} \log \left( 1 + P + \sin(\Delta)^2Q + 2\sin(\Delta)\sqrt{PQ} \right) + H(X_R + \cos(\theta)S_R + Z_R|X_j + \sin(\theta)S_R + Z)$$

(40c)

$$= \frac{1}{2} \log \left( 1 + P + \sin(\Delta)^2Q + 2\sin(\Delta)\sqrt{PQ} \right) + H(\frac{\cos(\theta)}{\sin(\theta)}X_j + \frac{\cos(\theta)}{\sin(\theta)}Z_j|X_R + \sin(\theta)S_R + Z_R)$$

(40d)

$$\leq \frac{1}{2} \log \left( 1 + P + \sin(\Delta)^2Q + 2\sin(\Delta)\sqrt{PQ} \right) + H(\frac{\cos(\theta)}{\sin(\theta)}X_j + \frac{\cos(\theta)}{\sin(\theta)}Z_j)$$

(40e)

$$\leq \frac{1}{2} \log \left( 1 + P + \sin(\Delta)^2Q + 2\sin(\Delta)\sqrt{PQ} \right) + \frac{1}{2} \log (4P + 2),$$

(40f)

$$\leq \frac{1}{2} \log \left( 1 + P + \sin(\Delta)^2Q + 2\sin(\Delta)\sqrt{PQ} \right) + \frac{1}{2} (P + 1) + 1.$$  

(40g)

$$\leq \frac{1}{2} \log \left( 1 + P + \sin(\Delta)^2Q + 2\sin(\Delta)\sqrt{PQ} \right) + \frac{1}{2} (P + 1) + 1.$$  

(40h)

For the negative entropy term $-H(Y^N|W, U^N, \theta^N = \phi^N)$ in (29d) we set $U^N = Z^N$. The overall outer bound is now

$$R^{\text{OUT-B}} = \frac{1}{2} \log (1 + P) + \min_{\rho, Q'} \left( \frac{1}{2} \log (1 + P + \sin(\Delta)^2Q' + 2\sqrt{PQ'}) - \frac{1}{4} \log ((2\sin(\Delta)^2Q' + 1 - \rho)(1 + \rho)) \right) + 2.$$  

(41)

We can now optimize this expression in (41) over $\rho$ and $Q'$. The optimal assignment for $\rho$ is

$$\rho^* = \min \{1, \sin(\Delta)^2Q\},$$

(42)

and, for $\rho = \rho^*$, the optimal value of $Q'$ is

$$\sin(\Delta)^2Q^* = \min \{P + 1, \sin(\Delta)^2Q\}.$$  

(43)

With this assignment, we obtain the bound in (12).
D. Proof of Th. IV.3

We analyze the performance on the real and the imaginary dimension separately. The transmitter can decide how to assign power in the two dimensions: in the following we assume that
\[
E[X_R^2] = \beta P \tag{44}
\]
\[
E[X_I^2] = \beta P \tag{45}
\]
for \( \beta = 1 - \bar{\beta} \).

Real Dimension

On the real dimension, the interference sequence is always \( \cos(\Delta)S_R^N \) and therefore in is possible to code as in the classical GP problem [1]. In particular, from the classical “writing on dirty paper” result [3] we have that the assignment
\[
X_R, \sim \mathcal{N}(0, \beta P) \tag{46}
\]
\[
U_R = X_R + \frac{\beta P}{\beta P + 1} \cos(\Delta)S_R, \tag{47}
\]
attains the rate achievable
\[
R_R = \frac{1}{2} \log(1 + \beta P), \tag{48}
\]
on the real dimension.

Imaginary Dimension

The imaginary channel input, \( X_I \), is composed of two codewords:
- a first codeword \( X_{IN} \)' (I for “Imaginary”, \( N \) as in “interference as Noise”) which treats the interference as noise while
- a second codeword \( X_{IP} \) (I for “Imaginary”, \( P \) as in “Pre-coded against the interference”) is pre-coded against the sequence \( + \sin(\Delta)S_R^N \). This pre-coding offers full interference pre-cancellation half of the time while only partial interference pre-coding the rest of the time.

The codeword \( X_{IN} \) is decoded first and removed from the channel output and, successively, the codeword \( X_{IP} \): this strategy attains the rate
\[
R_{IN} \leq I(Y;X_{IN}|\theta) \tag{49a}
\]
\[
R_{IP} \leq I(Y;U_{IP}|X_{IN}) - I(U_{IP};\sin(\theta)S_R), \tag{49b}
\]
with \( R = R_{IN} + R_{IP} \). We consider, in particular, the assignment
\[
X_{IN} \sim \mathcal{N}(0, \alpha P/2) \tag{50a}
\]
\[
X_{IP} \sim \mathcal{N}(0, \bar{\alpha} P/2) \tag{50b}
\]
\[
\alpha \in [0, 1], \; \bar{\alpha} = 1 - \alpha \tag{50c}
\]
\[
X_I = X_{IN} + X_{IP} \tag{50d}
\]
\[
U_{IP} = X_{IP} + \frac{\bar{\alpha} P}{\bar{\alpha} P + 1} S_R. \tag{50e}
\]
This assignment attains
\[
R_{IN} = \frac{1}{2} \log \left( 1 + \frac{\alpha \beta P}{1 + \alpha \beta P + \sin(\Delta)^2 Q} \right), \tag{51}
\]
and
\[
R_{IP} = \frac{1}{2} H(U|Y, \theta = \Delta) + \frac{1}{2} H(U|Y, \theta = -\Delta) + H(X_{IP}) \tag{52a}
\]
\[
= \frac{1}{4} \log (1 + \alpha \beta P) + \frac{1}{4} \log \left( \min \left\{ 1, \frac{(\bar{\alpha} \beta P + 1)(\bar{\alpha} \beta P + \sin(\Delta)^2 Q + 1)}{\bar{\alpha} \beta P + 2 \sin(\Delta)^2 Q \bar{\alpha} \beta P + \sin^2(\Delta^2 Q + 1) + 1} \right\} \right). \tag{52b}
\]

E. Proof of Lem. IV.4

Consider the inner bound of Th. IV.3 for \( \beta = 1/2 \) and disregard the last term in (13). The inner bound is then further lower bounded by
\[
R^{IN-B} \geq \frac{1}{2} \log \left( 1 + \frac{P}{2} \right) + \frac{1}{2} \log \left( 1 + \frac{\alpha P}{2 + \alpha P + 2 \sin^2(\Delta)^2 Q} \right) + \frac{1}{4} \log \left( 1 + \frac{\bar{\alpha} P}{2} \right). \tag{53}
\]
The derivative of RHS of (53) in $\alpha$ is
\[
D = -\frac{P}{2} \frac{1 + \alpha P - \sin(\Delta)^2 Q}{(1 + \alpha P)(1 + \alpha P + \sin(\Delta)^2 Q)},
\] (54)
therefore, if $\sin(\Delta)^2 Q > P + 1$, then $\alpha = 1$ is optimal. If $0 \leq \sin(\Delta)^2 Q - 1 < P$, the optimal $\alpha$ is $\frac{\sin(\Delta)^2 Q - 1}{P}$ while, if $\sin(\Delta)^2 Q < 1$, the optimal value is $\alpha = 0$.

F. Proof of Th. IV.5
Consider first the case $\sin(\Delta)^2 Q < 1$: in this case by treating the interference as noise in the imaginary dimension we attain
\[
R_I = \frac{1}{2} \log \left( 1 + \frac{P}{2 + 2\sin(\Delta)^2 Q} \right) \geq \frac{1}{4} \log \left( \frac{1}{4} + \frac{P}{4} \right) \geq \frac{1}{2} \log(1 + P) - 1,
\] (55) (56) (57)
while, using Costa pre-coding on the real axe, we attain
\[
R_R = \frac{1}{2} \log(1 + P/2) \geq \frac{1}{2} \log(1 + P) - \frac{1}{2}.
\] (58)
The gap between inner and outer bound when $\sin(\Delta)^2 Q \leq 1$ is therefore 1.5 bits/s/Hz. Let’s now compare inner and outer bound expression for $\sin(\Delta)^2 Q > 1$ by considering the case $\sin(\Delta)^2 Q \geq P + 1$ and $\sin(\Delta)^2 Q < P + 1$.

By comparing the outer bound in (61e) and the inner bound in (61d) for the case $\sin(\Delta)^2 Q > 1$ and $\sin(\Delta)^2 Q \geq P + 1$ we see that the two bounds differ by 3 bits/s/Hz.

For the case $\sin(\Delta)^2 Q < P + 1$ we have
\[
R^{\text{OUT-APP-B}}_{\text{IN-APP-B}} = \frac{1}{2} \log \left( \frac{2P + 2}{P + 1} \right) + \frac{1}{2} \log \left( \frac{3 + 4P}{\frac{P + 1}{\sin(\Delta)^2 Q}} \right) = \frac{1}{2} \log(1 + P) - \frac{1}{2},
\] (59a) (59b)
and we see that the distance between inner and outer bound is at most 3 bits/s/Hz.

G. Proof of Th. V.I
By applying Fano’s inequality we obtain
\[
N(R - \varepsilon_N) \leq I(Y^N;W|\theta^N) \leq H(Y^N|\theta^N) - H(Y^N|W,\theta^N),
\] (60a) (60b)
For the positive entropy term in (70b) we have
\[
H(Y^N|\theta^N) = \sum_j H(Y_j|\theta^N, Y^{j-1}) \leq \sum_j H(Y_j|\theta_j) \leq NH(Y_m|\theta_m) = NH(X + S\Re\theta + Z|\theta) = NH(X e^{-i\theta} + S + Z|\theta) = NH(\operatorname{Im} \{ e^{i\theta} X \} + Z_1; \Re \{ e^{i\theta} X \} + Z_2 + Z_R) \leq \frac{N}{2} \log(P + 1) + \frac{N}{2} \log \left( 1 + P + Q + 2\sqrt{PQ} \right), \leq \frac{N}{2} \log 2\pi e(P + 1) + \frac{N}{2} \log 2\pi e(1 + P + Q) + \frac{1}{2},
\] (61a) (61b) (61c) (61d) (61e) (61f) (61g)
where (61c) follows from choosing $m$ that maximizes $H(Y_m|\theta_m)$. In the following passages the index $m$ is dropped for convenience. In (61d) we have used the fact that the noise is circularly symmetric and thus rotations do not affect its distribution. In (71d) we have used the fact that
\[
P + Q + 2\sqrt{2PQ} + 1 \leq 2P + 2Q + 2.
For the positive entropy term in (70b) we provide the phase of the channel input $X^N$ as a genie aid side information to the receiver, Let
\[ \angle X^N = \psi^N \] (62)
so that we can write
\[-H(Y^N|W, \theta^N) \leq -H(Y^N|W, \phi^N, \psi^N) \leq -H(|X|^N e^{i\psi^N} + S_R e^{i\theta^N} + Z^N|W, \phi^N, \psi^N) \]
\[-H(|X|^N + e^{i(\theta - \phi)} S_R + Z^N|W, (\phi - \psi)^N) \]
\[-H(\sin((\phi - \psi)^N) S_{R}^N + Z_{N}^N, |X|^N + \cos((\phi - \psi)^N) S_{R}^N + Z_{R}^N, |W, (\phi - \psi)^N) \]
\[-H(\sin((\phi - \psi)^N) S_{R}^N + Z_{N}^N)(\phi - \psi)^N) - H(|X|^N + \cos((\phi - \psi)^N) S_{R}^N + Z_{R}^N, |W, (\phi - \psi)^N, S_{R}^N) \]
Note now, that regardless of the distribution of $\psi^N$, $(\phi - \psi)^N$ is iid and uniformly distributed over $[0, \pi)^N$ since $\phi^N$ is iid and uniformly distributed over $[0, \pi)^N$. For this reason we can write
\[-H(Y^N|W, \theta^N) \leq -NH(\sin(\phi_j - \psi_j) S_{R} + Z_{j} \theta_j - \psi_j) - H(|X|^N + \cos((\phi - \psi)^N) S_{R}^N + Z_{R}^N, |W, (\phi - \psi)^N, S_{R}^N) \]
\[-NH(\sin(\phi_j - \psi_j) S_{R} + Z_{j} \theta_j - \psi_j) \leq - \frac{N}{2} \log \left( \frac{2 \pi e}{2} \right) \]
\[-\frac{1}{2 \pi} \int_{0}^{2 \pi} \frac{1}{2} \log (\sin(t)^2 Q + 1) dt - \frac{N}{2} \log(2 \pi e) \]
\[-\frac{1}{2 \pi} \int_{0}^{2 \pi} \frac{1}{2} \log (\sin(t)^2 Q + 1) dt - N \log(2 \pi e). \]
We now use the fact that
\[ \sin(t) \geq \frac{2}{\pi} t, \quad t \in [0, \pi/2] \]
\[ \log(\sin(t)^2 Q + 1) \geq \log \left( \frac{4}{\pi^2} t^2 Q + 1 \right), \quad t \in [0, \pi/2] \]
so to obtain
\[-\frac{1}{2 \pi} \int_{0}^{2 \pi} \frac{1}{2} \log (\sin(t)^2 Q + 1) \leq - \frac{1}{2} \log(Q^2 + 1) + 1 - \frac{\pi \arctan(\sqrt{Q})}{\sqrt{Q}} \]
\[-\frac{1}{2} \log(Q^2 + 1) + 1. \]
This concludes the proof.

\textbf{H. Proof of Th. WJ}

By applying Fano’s inequality we obtain
\[ N(R - \epsilon_N) \leq I(Y^N;W|\theta^N) \leq H(Y^N|\theta^N) - H(Y^N|W, \theta^N). \]
For the positive entropy term in (70b) we have
\[ H(Y^N|\theta^N) \leq NH(Y_j|\theta) \]
\[-NH(X e^{i\theta} + S_R + Z) \]
\[ = NH(\text{Im}\{e^{i\theta} X\} + Z; \text{Re}\{e^{i\theta} X\} + S_R + Z_R) \]
\[ = \frac{N}{2} \log (P + 1) + \frac{N}{2} \log \left( (1 + P + Q + 2 \sqrt{PQ}) \right) \]
To bound the negative entropy term $-H(Y^N|W, \theta^N)$ in (70b) we introduce the a conditioning over the phase of the channel input $X^N$. For ease of notation let
\[ \angle X^N = \psi^N \]
so that we can write
\[-H(Y^N|W, \theta^N) \leq -H(Y^N|W, \phi^N, \psi^N) \]
\[ -H(|X|^N e^{i\psi^N} + S_{RC} e^{i\theta^N} + Z^N|W, \theta^N, \psi^N) \leq -H(|X|^N + e^{i(\theta-\phi)^N} S^N_R + Z^N|W, (\phi - \psi)^N) \]

\[ = -H(|X|^N + e^{i(\theta-\phi)^N} S^N_R + Z^N|W, (\phi - \psi)^N) \]

\[ = -H(\sin((\phi - \psi)^N) S^N_R + Z^N |X|^N + \cos((\phi - \psi)^N) S^N_R + Z^N|W, (\phi - \psi)^N) \]

Note now, that regardless of the distribution of \( \psi^N, \theta^N = (\phi - \psi)^N \) is iid and uniformly distributed over \([0, \pi]^N\) since \( \phi^N \) is iid and uniformly distributed over \([0, \pi]^N\). For this reason we can write

\[ -H(Y^N|W, \theta^N) \leq -H(\sin(\theta^N) S^N_R + Z^N |X|^N + \cos(\theta^N) S^N_R + Z^N|W, \theta^N) \]

\[ = -H(\sin(\theta^N) S^N_R + Z^N |\theta^N) - H(|X|^N + \cos(\theta^N) S^N_R + Z^N|W, \theta^N, \sin(\theta^N) S^N_R + Z^N) \]

\[ \leq -NH(\sin(\theta^N) S^N_R + Z^N |\theta^N) - H(|X|^N + \cos(\theta^N) S^N_R + Z^N|W, \theta^N, S^N_R) \]

\[ \leq -NH(\sin(\theta) S_R + Z^N |\theta) - N \frac{1}{2} \log(2\pi e) \]

\[ = -\frac{1}{2\pi} \int_0^{2\pi} \frac{1}{2} \log(\sin(t)^2 Q + 1) dt - \frac{N}{2} \log(2\pi e) \]

We now use the fact that

\[ \sin(t) \geq \frac{2}{\pi} t, \quad t \in [0, \pi/2] \]

\[ \log(\sin(t)^2 Q + 1) \geq \log\left(\frac{4}{\pi^2} t^2 Q + 1\right), \quad t \in [-\pi/2, \pi/2] \]

so to obtain

\[ -\frac{1}{2\pi} \int_0^{2\pi} \frac{1}{2} \log(\sin(t)^2 Q + 1) \leq -\frac{1}{2} \log(Q + 1) + 1 - \frac{\pi \arctan(\sqrt{Q})}{\sqrt{Q}} \]

\[ \leq -\frac{1}{2} \log(Q + 1) + 1. \]

Combining (71d) and (76) we obtain the expression in (15).

I. Proof of Th. 1.2

The transmission scheme can be described as:

- a first codeword \( X^N_{IN} \) (I for “Imaginary”, N as in “interference as Noise”) which treats the interference \( e^{i\theta} S_R \) as noise while
- a second codeword, \( X^N_{RC} \), (R for “Real” and for “interference Cancellation”) is transmitted only on the real axe and is decoded after \( X^N_{IN} \) and after having subtracted from the real channel output the estimate of the interference \( S_R \) obtained from the imaginary channel output.

More specifically, consider the following transmission scheme:

**Codebook Generation:** The message \( W \) is split into two sub-messages \( W_N \) and \( W_C \). The sub-message \( W_N \) is encoded in \( X^N_{IN} \). I for imaginary and N for “treating the interference as Noise” and the codebook for \( X^N_{IN} \) is generated by drawing \( 2^{NR_{IN}} \) sequences of length \( N \) with iid draws from the distribution \( \mathcal{N}(0, \alpha P) \). The sub-message \( W_C \) is encoded in \( X^N_{RC} \) which is generated by iid drawings \( 2^{NR_{RC}} \) sequences of length \( N \) with iid draws from the distribution \( \mathcal{N}(0, \beta P) \). Each codeword is indexed as \( X^N_{RC}(j), j \in [1 \ldots 2^{NR_{RC}} - 1] \).

**Encoding:** Each channel input \( X_j \) is obtained as

\[ X^N = X^N_{RC} + iX^N_{IN}, \]

which is a complex Gaussian (although not circularly symmetric) with covariance \( P \).

**Decoding:** The decoder first decodes \( X^N_{IN} \): this can be done as long as

\[ NR_{IN} \leq I(Y^N; X^N_{IN}|\theta) \]

\[ \leq NI(X_i + \sin(\phi) S_R + Z_i; X_{IN}|\phi), \]

which can yields

\[ R_{IN} = \int \frac{1}{2} \log(1 + \frac{\alpha P}{1 + \sin(t) Q}) dP_\theta(t). \]
A close form evaluation of $R_{IN}$ is possible only through inequality similar to (75) since

$$\sin(t) \geq t, \quad t \in [0 \ldots \pi/2]$$

(80a)

$$\log \left( 1 + \frac{\alpha P}{1 + \sin(t)^2 Q} \right) \geq \log \left( 1 + \frac{\alpha P}{1 + i^2 Q} \right),$$

(80b)

so that we obtain

$$R_{IN} \geq 4 \frac{1}{2\pi} \int_0^{\pi/2} \log \left( 1 + \frac{\alpha P}{1 + \frac{t^2}{4} Q} \right) dt$$

(81a)

$$= \frac{1}{2} \log \left( 1 + \frac{\alpha P}{\frac{\pi^2}{4} Q + 1} \right) - \frac{\arctan \left( \sqrt{\frac{\pi^2}{4} Q} \right)}{\sqrt{\pi^2/4 Q}}$$

(81b)

(81c)

(81d)

since (81c) is monotonically decreasing in $Q$ and

$$\lim_{Q \to 0} \frac{\arctan \left( \sqrt{\frac{\pi^2}{4} Q} \right)}{\sqrt{\pi^2/4 Q}} = 2$$

(82)

we conclude that

$$R_{IN} \geq \frac{1}{2} \log \left( 1 + \frac{\alpha P}{\frac{\pi^2}{4} Q + 1} \right) - 2$$

(83)

$$\geq \frac{1}{2} \log \left( 1 + \frac{\alpha P}{\frac{\pi}{Q} + 1} \right) - 3$$

(84)

(85)

After $X_{IN}^N$ has been decoded, it is subtracted from $Y_{IN}^N$ to estimate $S_{IN}^N$ through using the knowledge of $\theta^N$:

$$\tilde{Y}_{IN}^N = Y_{IN}^N - X_{IN}^N = \sin(\theta^N) S_{IN}^N + Z_{IN}^N,$$

(86)

The codeword $X_{RC}^N$ is estimated from the vector

$$\tilde{Y}_{R}^N = \sin(\theta^N) Y_{R}^N - \cos(\theta^N) \tilde{Y}_{IN}^N$$

(87a)

$$= \sin(\theta^N) X_{RC}^N + \sin(\theta^N) Z_{IN}^N - \cos(\theta^N) Z_{R}^N$$

(87b)

$$= \sin(\theta^N) X_{RC}^N + \tilde{Z}_{R},$$

(87c)

for

$$\tilde{Z}_{R} = \sin(\theta^N) Z_{IN}^N - \cos(\theta^N) Z_{R}^N \sim \mathcal{N}(0, 1).$$

(88)

This corresponds to the equivalent channel without interference whose capacity is

$$R_{RC} = I(\tilde{Y}_{R}; X_{RC}|\theta),$$

(89)

for which, as in (76) we have

$$R_{RC} \geq \frac{1}{2} \log \left( \alpha P + 1 \right) - 1,$$

(90)

as already evaluated in (76) (for $Q$ instead of $\alpha P$).