On Capacity of the Writing Onto Fast Fading Dirt Channel

Stefano Rini, *Member, IEEE*, and Shlomo Shamai (Shitz), *Life Fellow, IEEE*

***Abstract*— The Writing onto Fast Fading Dirt (WFFD) channel is investigated to study the effect of partial channel knowledge on the performance of interference pre-cancellation. The WFFD channel is the Gel'fand–Pinsker channel in which the channel output is the sum of the channel input, white Gaussian noise, and a fading-times-state term. The fading-times-state term is obtained as the product of the channel state sequence, known only at the transmitter, and a fast fading process, known only at the receiver. We consider the case of Gaussian-distributed channel states and derive an approximate characterization of capacity for different classes of fading distributions, both continuous and discrete. In particular, we prove that if the fading distribution concentrates in a sufficiently small interval, then capacity is approximately equal to the AWGN capacity times the probability of such interval. We also show that there exists a class of fading distributions for which having the transmitter treat the fading-times-state term as additional noise closely approaches capacity.**

***Index Terms*— Gel'fand-Pinsker channel, writing on fading dirt channel, fast fading, partial channel side information, Costa pre-coding, interference pre-cancellation.**

I. Introduction

THE classic "Writing on Dirty Paper" (WDP) channel capacity result [1] establishes that full state pre-cancellation can be attained in the Gel'fand-Pinsker (GP) channel with additive state and additive white Gaussian noise, regardless of the distribution of the state sequence. Albeit very promising, this result assumes that perfect channel knowledge is available at the users: this assumption is not valid in many communication scenarios in which channel conditions vary over time and with limited feedback between receiver and transmitter. For this reason, we investigate the effects of partial channel knowledge on the performance of state pre-cancellation. More specifically, we study the capacity of the "Writing onto Fast Fading Dirt" (WFFD) channel, a variation of Costa's WDP channel in which the state sequence, known only at the transmitter, is multiplied by a fast fading process, known only at the receiver. The WFFD channel models the downlink scenario in which a transmitter communicates to a receiver in the presence of a known interferer. The transmitter acquires the message sent by the interferer through the network architecture while the receiver learns the channel toward the interferer from the pilot tones broadcasted by the interferer. Due to rate limitations in the control and feedback channels, the transmitter and the receiver are unable to exchange each other's knowledge. This results in the situation in which the transmitter knows the interfering message but not the interfering channel, while the receiver knows the interfering channel but not the interfering message. For this scenario, one wishes to determine the limiting interference pre-cancellation performance that is attainable despite the partial and asymmetric knowledge at the transmitter and the receiver.

Related Results: The GP channel [2] is the point-to-point channel in which the output is obtained as a random function of the input and a state sequence which is non-causally known at the transmitter. The capacity of the GP channel is expressed in [2] as the maximization of a non-convex function for which the optimal solution is not easily determined, either explicitly or through numerical evaluations. For this reason, very few closed-form expressions of the GP channel capacity are available in the literature. One of the few models for which capacity is known in closed-form is the WDP channel: in [1] Costa shows that the capacity of the WDP channel is equal to the capacity of the Gaussian point-to-point channel. This result implies that it is possible for the encoder to fully pre-code its transmissions against the known channel state.

In the literature, few authors have investigated extensions of the result in [1] to include fading and partial channel knowledge. The "Carbon Copying onto Dirty Paper" (CCDP) channel [3] is the M-user compound GP channel in which the output at each compound receiver is obtained as the sum of the input, Gaussian noise and one of M possible state sequences, all non-causally known only at the transmitter. When the state sequences at each receiver are scaled versions of the same sequence, the CCDP channel models the WDP channel in which the channel state is multiplied by a slow fading process [4]. The WDP channel in which both the input and the state sequence are multiplied by the same fading realization is studied in [5]. The authors consider both the case of fast and slow fading and evaluate the achievable rates using the

Manuscript received April 25, 2017; revised April 19, 2018; accepted August 12, 2018. Date of publication September 24, 2018; date of current version November 9, 2018. The work of S. Rini was supported by the Ministry of Science and Technology (MOST) under Grant 103-2218-E-009-014-MY2. The work of S. Shamai was supported by the European Union's Horizon 2020 Research and Innovation Programme under Grant 694630. The associate editor coordinating the review of this paper and approving it for publication was R. K. Ganti. *(Corresponding author: Stefano Rini.)*

S. Rini is with the Department of Electrical and Computer Engineering, National Chiao-Tung University, Hsinchu 300, Taiwan (e-mail: rini.stefano@gmail.com; stefano@nctu.edu.tw).

S. Shamai is with the Department of Electrical Engineering, Technion–Israel Institute of Technology, Haifa 3200003, Israel (e-mail: sshlomo@ee.technion.ac.il).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TWC.2018.2866841

pre-coding of [1], showing that the rate loss from full state pre-cancellation is vanishing in both scenarios as the transmit power grows to infinity. Lattice strategies for this model are investigated in [6]. The WDP channel in which slow fading affects only the state sequence is first studied in [7] for the case of phase fading. In [8], we show the approximate capacity of this model for some classes of the fading distributions. Achievable rates under Gaussian signaling are derived in [9] for the case of Gaussian-distributed fast fading and compared to lattice coding strategies. The performance of lattice coding strategies for this channel model is further studied in [10]–[12].

Contributions: We investigate the capacity of the WFFD channel in which the state sequence is a white Gaussian sequence.[1] We consider separately the case of discrete and continuous fading distribution:

- **Sec. IV – Discrete fading distribution.** We begin by determining capacity to within a constant gap for the case of uniform antipodal fading. For this simple fading distribution capacity can be approached by transmitting the superposition of two codewords: the bottom codeword treats the fading-times-state as additional noise while the top codeword is pre-coded against one of the fading realizations times the channel state. This result is extended to two classes of fading distributions: distributions with mode larger than a half and uniform distributions with exponentially spaced points in the support. In both cases, capacity is approached to within a small gap by a combination of superposition coding and state pre-coding as in the case of uniform antipodal fading.
- **Sec. V – Continuous fading distribution.** We first show simple conditions under which capacity is at most half of the AWGN capacity, we then derive the approximate capacity for the case of a continuous fading distribution which concentrates around a sufficiently narrow interval. The converse proof is shown by relating the capacity of the model with continuous fading to the capacity of the model in which the fading distribution is a quantized version of the original distribution. Finally, we show that there exists a heavy-tailed fading distribution for which the capacity of the WFFD channel is approximatively equal to the capacity of the channel without transmitter statke knowledge.

The main theoretical contributions of the paper consist in the development of new outer bounding techniques to characterize the capacity of a model comprising both channel states and partially known fast fading.

Although the results in the paper are derived for white Gaussian-distributed channel states, they can be generalized to the case of any i.i.d. distribution of the state sequence. We consider the case of Gaussian channel states as we focus on deriving computable performance bounds that provide insights on the performance loss due to the lack of perfect channel knowledge. As such, fixing state distribution to Gaussian allows us to express the results in the paper solely as a function of the transmit power, the interference power and the fading distribution.

[1]In the literature this channel has also been referred to as "dirty paper channel with fading dirt", "writing on fast faded dirt" and "dirty paper coding channel with fast fading". We prefer the term "writing on fast fading dirt" for both brevity and clarity.

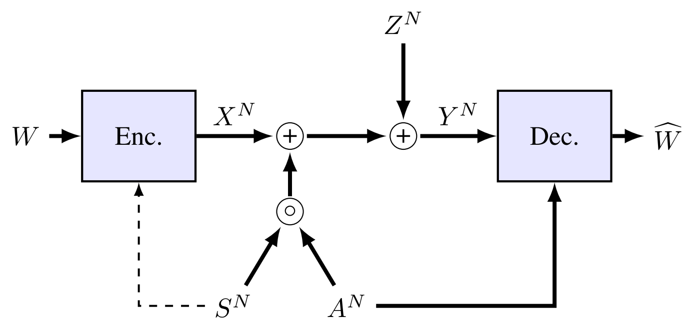


Fig. 1. The "Writing onto Fast Fading Dirt" (WFFD) channel.

Paper Organization: In Sec. II we introduce the channel model under consideration while Sec. III presents relevant results available in the literature. Sec. IV considers the case of discrete fading distributions while Sec. V studies the case of continuous distributions. Finally, Sec. VI concludes the paper.

II. Channel Model

The WFFD channel, also depicted in Fig. 1, is the GP channel in which the channel output is obtained as

$$Y^N = X^N + cA^N \circ S^N + Z^N, \tag{1}$$

where X^N denotes the channel input, S^N the channel state, A^N the fading sequence and Z^N the additive noise while $\circ$ indicates the Hadamard, or element-wise, product.[2] The channel input X^N is subject to the second moment constraint $\mathbb{E}\left[|X_i|^2\right] \leq P, \ \forall\ i \ \in [1\ldots N]$. Both the channel state and the additive noise are white Gaussian sequences, i.e. Z^N, $S^N \sim i.i.d.\ \mathcal{N}(0,1)$ while the fading sequence A^N is an i.i.d. sequence from the distribution $P_A(a)$, with support $\mathcal{A}$, either continuous or discrete. Without loss of generality we further assume $\mathbb{V}\mathrm{ar}[A] = 1$ and $c \in \mathbb{R}^+$. Having knowledge of the channel state S^N, the encoder wishes to reliably communicate the message $W \in \mathcal{W} = [1\ldots 2^{NR}]$ to the receiver through the channel input X^N. Upon receiving the channel output Y^N and the fading realization A^N, the receiver produces the estimate $\widehat{W} \in \mathcal{W}$ of the transmitted message.[3]

In the study of the WFFD channel, standard definitions of code, achievable rate and capacity are employed.

Definition 1 (Code, Probability of Decoding Error and Achievable Rate): A $(2^{NR}; N)$ code for the WFFD channel consists of an encoding and a decoding function, $X^N = f(W, S^N)$ and $\widehat{W} = g(Y^N, A^N)$ respectively. The probability of error for a $(2^{NR}; N)$ code, $P_e(2^{NR}; N)$, is defined as

$$P_e(2^{NR}; N) = \mathbb{P}\left[\widehat{W}(Y^N, A^N) \neq W\right], \tag{2}$$

where the probability in the RHS of (2) is averaged over all possible fading and state sequences and transmitted message. A rate $R \in \mathbb{R}^+$ is said to be achievable if there exists a sequence of codes such that the probability of error $P_e(2^{NR}; N)$ goes to zero as N goes to infinity.

[2]More explicitly, $A^N \circ S^N = [A_1\ S_1, A_2\ S_2 \ldots A_N S_N]^T$.

[3]The fading sequence A^N can be seen as an additional channel, together with Y^N in (1), as shown in Fig. 1.

Definition 2 (Capacity and Approximate Capacity): The capacity $\mathcal{C}$ is the supremum of all the achievable rates. An inner bound $R^{\rm IN}$ and an outer bound $R^{\rm OUT}$ to $\mathcal{C}$ for which

$$R^{\rm OUT} - R^{\rm IN} \leq \Delta, \tag{3}$$

for some constant $\Delta \in \mathbb{R}^+$, are said to characterize the capacity to within an additive gap of Δ bits–per–channel–use (bpcu) or, for brevity, to determine the approximate capacity to within Δ bpcu.

The WFFD channel is a special case of the GP channel and the capacity of the GP channel is obtained as

$$\mathcal{C} = \max_{P_{U,X|S}} I(Y;U|A) - I(U;S). \tag{4}$$

The expression in (4) is convex in $P_{X|S,U}$ for a fixed $P_{U|S}$ which implies that X can be chosen to be a deterministic function of U and S. On the other hand, (4) is neither convex nor concave in $P_{U|S}$ for a fixed $P_{X|S,U}$: for this reason, determining a closed-form solution for the maximization in (4) is generally challenging. Additionally, the lack of tight bounds on the cardinality of the auxiliary random variable U in (4) further complicates the task of obtaining numerical approximations of the optimal solution. For these reasons, in the following, we provide alternative inner and outer bounds to capacity which are expressed only as a function of P, c and P_A. We also determine the approximate capacity for some class of distributions, focusing on those instances in which a simple combination of known achievable strategies is sufficient to approach capacity.

In the remainder of the paper, we refer to the term $cA^N \circ S^N$ as the "fading-times-state" term and the parameter c is used to normalize the variance of both the state and the fading distributions to one.

Lemma 3: For the model in (1)*, the variance of the state and the fading distributions are taken unitary without loss of generality. Also, the channel state is taken to have zero mean and $c \in \mathbb{R}^+$ without loss of generality.*

The proof of Lem. 3 is omitted for brevity. The following lemma is useful in the tightening of certain outer bounds.

Lemma 4: The capacity of the WFFD channel is decreasing in c.

Proof: The proof is presented in App. A. ■

Remark 5: Generally speaking, the results derived in the remainder of the paper can be generalized to the case of any i.i.d. state distribution. We focus on the case of a unitary variance, Gaussian-distributed i.i.d. channel state as the corresponding capacity bounds are expressed as only as a function of the transmit power P, the gain of the fading-times-state term c, and the fading distribution P_A. Such expressions, in most instances, provide clearer insights on the effect of partial channel knowledge on the capacity of the WFFD channel. Generalizations of our results to other channel state distributions are, for the most part, rather straightforward.

III. Related Results

This section briefly introduces the results available in the literature which are relevant in the study of the WFFD channel.

• **The "Writing on Dirty Paper" (WDP) channel.** One of the few GP channel models for which the maximization in (4) is known in closed-form is the WDP channel [1]. For this model, the optimal assignment in (4) is

$$X \sim \mathcal{N}(0,P), \quad X \perp S, \; U = X + \frac{P}{P+1} S, \tag{5}$$

and yields $\mathcal{C} = 1/2 \log(1+P)$ in (4), regardless of the distribution of S^N. The assignment in (5) is usually referred to as Dirty Paper Coding (DPC).

• **The "Carbon Copying onto Dirty Paper" (CCDP) channel.** The CCDP channel [3] is the M-user compound GP channel in which a channel output is obtained as the sum of the input, Gaussian noise and one of M Gaussian state sequences, all non-causally known at the transmitter, i.e.

$$Y_m^N = X^N + cS_m^N + Z_m^N, \quad m \in [1 \ldots M], \tag{6}$$

where $S_m^N \sim i.i.d\, \mathcal{N}(0, Q_m)$ and $\{S_m^N, \; m \in [1 \ldots M]\}$ have any jointly Gaussian distribution. In [3], the authors derive the first inner and outer bound for this model. The approximate capacity for the case of $M = 2$ and independent, unitary variance state sequences is derived as [14].

Theorem 6 (Outer Bound and Approximate Capacity for the 2-User CCDP Channel With Independent States [14]): The capacity of the 2-user CCDP channel with S_1^N, $S_2^N \sim$ i.i.d. $\mathcal{N}(0,1)$, $S_1^N \perp S_2^N$ is upper bounded as

$$\begin{aligned} \mathcal{C} &\leq R^{\rm OUT} \\ &= \begin{cases} \frac{1}{2}\log(1+P) + 1/2 & c^2 \leq 2 \\ \frac{1}{2}\log\left(\frac{P + c^2/2 + 1}{c^2}\right) & \\ \quad + \frac{1}{4}\log\left(\frac{c^2}{2}\right) + 1/2 & 2 \leq c^2 < 2(P+1) \\ \frac{1}{4}\log(P+1) & c^2 \geq 2(P+1), \end{cases} \end{aligned} \tag{7}$$

and the capacity is to within 1 bpcu *from the outer bound in* (7).

Capacity in Th. 6 is approached by sending the superposition of two codewords: the base codeword treats the states as additional noise while the top codeword is pre-coded against each of the state realizations for half of the time. Th. 6 shows that it substantially not possible to simultaneously pre-code the channel input against two independent channel states.

• **Writing onto Fast Fading Dirt (WFFD) channel.** For the WFFD channel with Gaussian fading, the authors of [15] optimize the achievable strategy in (4) over all jointly Gaussian distributions of S, U and X.

Theorem 7 (Achievability With Jointly Gaussian Signaling [9, Th. 1], [15, Sec. IV]): Consider the WFFD channel for $A^N \sim$ i.i.d. $\mathcal{N}(0,1)$ and let $\rho = (\rho_{XS}, \rho_{US}, \rho_{UX})$ and define $\mathbf{K} \subset [-1,1]^3$ as

$$\mathbf{K} = \left\{ \begin{array}{l} |\rho_t| < 1, \quad t \in \{\rm XS, US, UX\} \\ 1 + 2\rho_{XS}\rho_{US} - \rho_{XS}^2 - \rho_{US}^2 - \rho_{UX}^2 = 0 \end{array} \right\}, \tag{8}$$

then an inner bound to capacity is

$$\mathcal{C} \geq R^{\rm IN} = \max_{\rho \in \mathbf{K}} \; \mathbb{E}_\theta[R_\Gamma(\rho,\theta)| \; A = \theta], \tag{9}$$

for

$$\begin{aligned}R_\Gamma(\rho,\theta) &= \frac{1}{2}\log\Big((P+c^2+2\theta\rho_{XS}c\sqrt{P}+1)(1-\rho_{US}^2)\Big)\\ &\quad-\frac{1}{2}\log\big(P(1-\rho_{UX}^2)+c^2(1-\rho_{US}^2)\\ &\quad+2\theta c(\rho_{XS}-(\rho_{UX}\rho_{US}))\sqrt{P}+1\Big).\end{aligned} \tag{10}$$

Th. 7 attempts to generalize the result of [1] to the Gaussian fast fading case although, in all likelihood, one needs to consider a wider class of distributions than jointly Gaussian distributions to attain maximum in (4).

IV. WFFD Channel With a Discrete Fading Distribution

A. Uniform Antipodal Fading

We begin by providing the approximate capacity for the WFFD channel in which the fading is uniformly distributed over the set $\{-1,+1\}$. This is perhaps the simplest choice of fading distribution for the WFFD channel, yet this example well illustrates the main bounding techniques employed in the reminder of the paper.

Theorem 8 (Outer Bound and Approximate Capacity of the WFFD Channel With Antipodal Uniform Fading): Consider the WFFD channel in which A is uniformly distributed over the set $\{-1,+1\}$, then the capacity $\mathcal{C}$ is upper bounded as

$$\mathcal{C} \le R^{\rm OUT} = \begin{cases}\frac{1}{2}\log(P+1)+\frac{1}{2} & c^2\le 1\\ \frac{1}{2}\log(P+c^2+1) & \\ \quad -\frac{1}{4}\log(c^2)-\frac{1}{2} & 1<c^2<P+1\\ \frac{1}{4}\log(P+1)-\frac{1}{2} & c^2\ge P+1,\end{cases} \tag{11}$$

and the capacity is to within 1 bpcu *from the outer bound in* (11).

Proof: Achievability and converse proofs are as follows.

• **Achievability.** Consider the achievable strategy in which the channel input is obtained as the superposition of two codewords: (i) the codeword $X_{\rm SAN}^N$ (for *State As Noise*), at rate $R_{\rm SAN}$, which treats $cA^N\circ S^N$ as additional noise and (ii) the codeword $U_{\rm PAS}^N$ (for *Pre-coded Against the State*), at rate $R_{\rm PAS}$, which is pre-coded against S^N as in the WDP channel. This strategy attains the rate $R^{\rm IN}=R_{\rm SAN}+R_{\rm PAS}$ for

$$\begin{aligned}R_{\rm SAN} &\le I(Y;X_{\rm SAN}|A)\\ R_{\rm PAS} &\le I(Y;U_{\rm PAS}|X_{\rm SAN},A)-I(U_{\rm PAS};S).\end{aligned} \tag{12}$$

Through the assignment

$$\begin{aligned}X_{\rm SAN}\sim\mathcal{N}(0,\alpha P),\quad X_{\rm PAS}\sim\mathcal{N}(0,\overline{\alpha}P),\ X_{\rm SAN}\perp X_{\rm PAS}\\ X=X_{\rm SAN}+X_{\rm PAS},\quad U_{\rm PAS}=X_{\rm PAS}+c\frac{\overline{\alpha}P}{\overline{\alpha}P+1}S,\end{aligned} \tag{13}$$

we obtain the achievable rate

$$R^{\rm IN}(\alpha)\ge\frac{1}{2}\log\left(1+\frac{\alpha P}{1+\overline{\alpha}P+c^2}\right)+\frac{1}{4}\log(\overline{\alpha}P+1)-\frac{1}{2}. \tag{14}$$

Optimizing the expression in (14) over α yields in the inner bound

$$R^{\rm IN} = \begin{cases}\frac{1}{2}\log(1+P)-\frac{1}{2} & c^2\le 1\\ \frac{1}{2}\log(1+P+c^2) & \\ \quad -\frac{1}{4}\log(c^2)-1 & 1<c^2<P+1\\ \frac{1}{4}\log(1+P)-1 & c^2\ge P+1.\end{cases} \tag{15}$$

• **Converse.** From Fano's inequality we have

$$\begin{aligned}N(R-\epsilon_N) &\le I(Y^N;W|A^N)\\ &\le \sum_{j=1}^N H(Y_j|A_j)-H(Y^N|A^N,W)\\ &\le N\max_j H(Y_j|A_j)-H(Y^N|A^N,W)\\ &\le N\max_{P_{Y|A}}\ H(Y|A)-H(Y^N|A^N,W).\end{aligned} \tag{16}$$

The entropy term $\max_{P_{Y|A}}\ H(Y|A)$ in (16) is bounded as

$$\begin{aligned}&\max_{P_{Y|A}}\ H(Y|A)\\ &\le \max_{P_{Y|A}}\ \frac{1}{2}\left(H(X+cS+Z)+H(X-cS+Z_j)\right)\\ &\le \max_{|\rho_{XS}|\le 1}\ \frac{1}{2}\left(H(X_{Gj}+cS+Z)+H(X_{Gj}-cS+Z_j)\right),\end{aligned} \tag{17}$$

where (17) follows from the "Gaussian Maximizes Entropy (GME)" property by letting X_{Gj} be jointly Gaussian random variables with variance P and with correlation ρ_{XS} with S. Optimizing (17) over ρ_{XS} yields the upper bound

$$\max_{P_{Y|A}}\ H(Y|A)\le\frac{1}{2}\log(2\pi e)^2\left(P+c^2+1\right). \tag{18}$$

Define now $\overline{a}^N(a^N)=-a^N$ and notice that

$$\begin{aligned}&H(Y^N|W,A^N)\\ &\quad=\frac{1}{2}\sum_{a^N\in\{-1,+1\}^N}\frac{1}{2^N}\left(H(Y^N|W,A^N=-a^N)\right.\\ &\qquad\left.+H(Y^N|W,A^N=+a^N)\right),\end{aligned}$$

so that

$$\begin{aligned}&-H(Y^N|W,A^N)\\ &\le -\frac{1}{2^{N+1}}\sum_{a^N\in\{-1,+1\}^N}\\ &\qquad H(X^N-ca^N\circ S^N+Z^N,X^N+ca^N\circ S^N+Z^N|W)\end{aligned} \tag{19a}$$

$$\begin{aligned}&=-\frac{1}{2^{N+1}}\sum_{a^N\in\{-1,+1\}^N}\\ &\qquad H(2ca^N\circ S^N,X^N+ca^N\circ S^N+Z^N|W)\end{aligned} \tag{19b}$$

$$\begin{aligned}&=-\frac{1}{2^{N+1}}\sum_{a^N\in\{-1,+1\}^N}\left(H(2ca^N\circ S^N|W)\right.\\ &\qquad\left.+H(X^N+ca^N\circ S^N+Z^N|S^N,W)\right)\end{aligned} \tag{19c}$$

$$= -\frac{1}{2^{N+1}} \sum_{a^N \in \{-1,+1\}^N} \big(H(2ca^N \circ S^N) + H(X^N + ca^N \circ S^N + Z^N | S^N, W, X^N) \big) \quad (19d)$$

$$= -\frac{1}{2^{N+1}} \sum_{a^N \in \{-1,+1\}^N} H(2ca^N \circ S^N) + H(Z^N), \quad (19e)$$

where (19b) follows from the fact that transformation has unitary Jacobian. The equality in (19c) follows from the fact that $W \perp S^N$, (19d) from the fact that X^N is a function of W and S^N and from the Markov chain $W - [X^N, S^N] - Y^N$. Next, we observe that the terms in the summation in the RHS of (19e) are all identical and equal to $1/2 \log(2\pi e 4\ c^2) + 1/2 \log(2\pi e)$, so that

$$-H(Y^N|W, A^N) \le -\frac{N}{4}\log(2\pi e c^2) - \frac{N}{4}\log(2\pi e) - \frac{N}{2}. \quad (20)$$

Using (18) and (20) we rewrite the outer bound in (16) as

$$\begin{aligned} R^{\text{OUT}} &= \frac{1}{2}\log(2\pi e)^2 \left(P + c^2 + 1\right) \\ &\quad - \frac{1}{4}\log(2\pi e c^2) - \frac{1}{4}\log(2\pi e) - \frac{1}{2} \\ &= \frac{1}{2}\log(P + c^2 + 1) - \frac{1}{4}\log(c^2) - \frac{1}{2}. \end{aligned} \quad (21)$$

Note that, as a function of c^2, the expression in (21) has a minimum in $c^2 = P + 1$. From Lem. 4 we have the capacity is decreasing in c: for this reason, the channel in which c^2 is equal to $\min\{c^2, P + 1\}$ corresponds to a model with larger capacity. For this latter model, the outer bound in (21) still holds so that letting c^2 equal to $\min\{c^2, P+1\}$ in (21) provides an outer bound to the capacity of the original model. With this substitution and some further bounding for the case $c^2 < 1$, we obtain the outer bound in (11). By comparing the outer bound in (11) and the inner bound in (15), we verify that the they differ of at most 1 bpcu. ∎

The result in Th. 8 can be interpreted as follows: the parameter c controls the variance of the fading-time-state term and (i) for small values of c, treating the term $cA^N \circ S^N$ as additional noise results in a limited rate loss. (ii) when the variance of $cA^N \circ S^N$ is larger than the transmit power, then it is approximately optimal to pre-code against one fading realization, as this strategy grants correct decoding for half of the channel uses on average. Finally, (iii) in the intermediate regime capacity is approached by a linear combination of the previous two strategies.

Remark 9: The approximate capacity result for the two-user CCDP channel with independent, equal-variance states in Th. 6 has interesting similarities to the proof of Th. 8 and the approximate capacity expressions in (7) and (11) are also similar. From a high-level perspective, in the WFFD channel each fading realization can be thought of as a compound user in the CCDP channel so that the number of users grows with the transmission length, instead of being constant.

Remark 10: With respect to Rem. 5, we note that the result in Th. 6 can be generalized to the case of any i.i.d. state distributions by appropriately adapting the bounding step in (19) to yield

$$-H(Y^N|W, A^N) \le -\frac{N}{4}\log(c^2) - \frac{N}{2}H(S) - \frac{N}{2}, \quad (22)$$

instead of (20). This generalization of Th. 6 and the remaining results in the paper is not pursued for brevity.

B. WFFD Channel With a Discrete Fading Distribution With Mode Larger Than a Half

Theorem 11 (Outer Bound and Approximate Capacity for the WFFD Channel With a Fading Distribution of Mode Larger Than a Half): Consider the WFFD channel in which A is a discrete random variable such that

$$\exists\, m \in \mathcal{A}, \quad \text{s.t. } P_A(m) \ge \frac{1}{2}, \quad (23)$$

and let $Q_m = P_A(m)$ and $\overline{Q}_m = 1 - Q_m$, then the capacity $\mathcal{C}$ is upper bounded as

$$\mathcal{C} \le R^{\text{OUT}} = \begin{cases} \frac{1}{2}\log(1+P) + 1 & \overline{Q}_m \ge Q_m c^2(1+\mu_A^2) \\ \frac{1}{2}\log(1+P) - \frac{\overline{Q}_m}{2}\log\left(c^2(1+\mu_A^2)\right) + G_m & \overline{Q}_m < Q_m c^2(1+\mu_A^2) \le \overline{Q}_m(P+1) \\ \frac{Q_m}{2}\log(1+P) + G_m & Q_m c^2(1+\mu_A^2) > \overline{Q}_m(P+1), \end{cases} \quad (24)$$

for

$$G_m = \frac{1}{2}\mathbb{E}_A\left[\log\left(\frac{1+\mu_A^2}{(A-m)^2}\right) | A \ne m\right] + 3, \quad (25)$$

and the capacity is to within a gap of

$$G'_m = \frac{1}{2}\mathbb{E}_A\left[\log\left((1+\mu_A^2)\left(\frac{1}{A^2} + \frac{1}{(A-m)^2}\right)\right) | A \ne m\right] + 3, \quad (26)$$

from the outer bound in (24).

Proof: For the class of fading distributions in (23), state pre-cancellation can be attained for a portion Q_m of the channel uses on average: for this reason, the achievable strategy employed in the proof of Th. 8 is still effective. In the converse proof, we extend the idea of conjugate fading sequences in the proof of Th. 8 to the elements in the set of typical fading realizations. The full proof can be found in App. B. ∎

The next lemma provides a simplification of the result in Th. 11 under some conditions on the support of the fading distribution.

Lemma 12: If $|A| > \Delta$ and $|A - m| \ge \Delta$ for some $\Delta > 0$, then (25) *and* (27) *satisfy*

$$\begin{aligned} G_m &\le \frac{\overline{Q}_m}{2}\log\left(\frac{1+\mu_A^2}{\Delta^2}\right) + 3 \\ G'_m &\le G_m + \frac{1}{2}. \end{aligned}$$

The proof of Lem. 12 is omitted for brevity; this lemma shows that a tight characterization of capacity is possible when the mean of A is small and the points in the support are sufficiently far from the mode of the distribution.

1) WFFD Channel in the "Strong Fading" Regime:

Theorem 13 (Outer Bound and Approximate Capacity for the WFFD Channel in the "Strong Fading" Regime): Consider the WFFD channel with $c > 2$ and in which A is uniformly distributed over a discrete set $\mathcal{A} = \{\alpha_i\}_{i=1}^M$, $\alpha_1 < \alpha_2 < \ldots < \alpha_M$ such that

$$\alpha_1 \geq \frac{1}{c-1}, \tag{27a}$$

$$\alpha_{i+1} \geq c\alpha_i, \quad i \in [2 \ldots M-1], \tag{27b}$$

then, the capacity $\mathcal{C}$ is upper bounded as

$$\mathcal{C} \leq R^{\mathrm{OUT}} = \begin{cases} \frac{1}{2}\log(1+P) + G_s & \frac{1}{M} \geq \frac{(1+\mu_A^2)c^2}{M} \\ \frac{1}{2}\log(1+P+(1+\mu_A^2)c^2) - \frac{M-1}{2M}\log\left((1+\mu_A^2)c^2\right) + G_s & \frac{1}{M} < \frac{(1+\mu_A^2)c^2}{M} \leq \frac{M-1}{M}(P+1) \\ \frac{1}{2M}\log(1+P) + G_s & \frac{(1+\mu_A^2)c^2}{M} > \frac{M-1}{M}(P+1), \end{cases} \tag{28}$$

for

$$G_s = \frac{1}{2}\log(1+\mu_A^2) + \frac{1}{2}, \tag{29}$$

and the capacity is to within a gap of $G_s + 1$ from the outer bound in (28)*.*

Proof: As for Th. 8, the achievability proof relies on the simple combination of superposition coding and DPC. The converse bound involves defining $M-1$ conjugate sequences which are used to recursively bound the channel capacity by also providing a carefully-chosen genie-aided side information. The proof is provided in App. C. ■

Th. 13 effectively shows that, when the fading realizations are exponentially spaced apart while their mean is small, then it is not possible to simultaneously pre-code against multiple fading realizations. We refer to the conditions in (27) as the "strong fading" condition as the elements in support of the fading distribution grow exponentially large while their probability remains constant.

Remark 14: The result in Th. 13 can be generalized to the case in which $|a_{i+1}| \geq \kappa c|a_i|$ for some $\kappa \in \mathbb{R}^+$: in this case, (29) is expressed

$$G_s = \frac{1}{2}\log\kappa(1+\mu_A^2) + \frac{1}{2}. \tag{30}$$

As an example of the result in Th. 13, consider the case in which A is uniformly distributed over the set

$$\mathcal{A}(M) = \left\{\Delta, c\Delta, c^2\Delta, \ldots, c^{M-1}\Delta\right\},$$

for $M \geq 3$ where Δ is chosen so as to obtain unitary variance,[4] then $G_s = \frac{3}{2}$, $G_s' = \frac{5}{2}$. Note that the capacity goes to zero when both c and M grow to infinity.

C. Numerical Examples

• **Geometric distribution.** Consider the case in which A is distributed according to the following geometric distribution

$$P_A(k_a + n\Delta) = \overline{p}^n p, \quad n \in \mathbb{N}, \tag{31}$$

for some $p \in [0,1]$ with $\overline{p} = 1-p$ and $\Delta > 0$ and $\overline{p}\Delta^2 = \overline{p}^2$ (to obtain a unitary variance) and $k_a = -\Delta_A(1-p)/p$ (to obtain zero mean). For the fading distribution in (31), Th. 11 can be applied when $p \geq 1/2$: the outer bound in (24) and the gap from capacity depend on the value G_m and G_m' obtained as

$$\begin{aligned} G_m &= 2\sum_{n=1}^{\infty}\log\left(n\Delta+2\right)p(1-p)^n + 3 \\ &\geq -\overline{p}\log\Delta^2 + 3 \approx 3.15 \\ G_m' &= \sum_{n=1}^{\infty}\log\left(\frac{n^2\Delta^2}{(k_a+\Delta n)^2}+1\right)\overline{p}_A^n p + 3 \\ &\leq \frac{1}{2\ k_a^2}(1-p) + 3 \approx 3.65. \end{aligned}$$

• **Binomial Distribution.** Consider next the case in which A has the following Bernoulli distribution

$$P_A(k_a + n\Delta, N) = \binom{2N}{n}(1-p)^n p^{2N-n}, \tag{32}$$

for $n \in [-N \ldots +N]$ and $2Np(1-p) = \Delta^2$ to obtain variance unitary and $k_a = -N\Delta p$ to have a zero mean. Note that the case $N=2$ and $p = 1/2$ recovers the result in Th. 8. Using an approximation of the central Bernoulli coefficient we obtain that the Th. 11 is applicable to the fading distribution in (32) when $p \geq \frac{1}{2} + \frac{1}{2}\sqrt{1-2^{-2N+1}\sqrt{N}}$ in which case we have $G_m = \frac{1}{2}\log(2\pi e) + 3 + \mathcal{O}\left(\frac{1}{}N\right)$ and $G_m' = 3 + Q_m$.

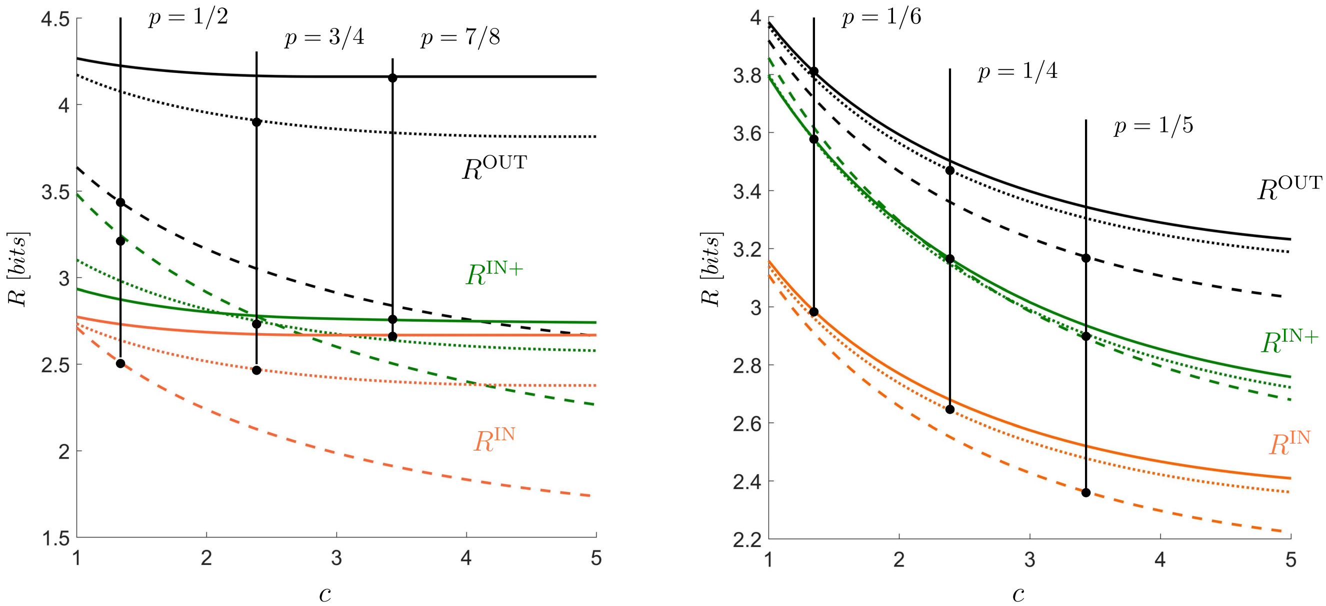
The capacity characterization in Th. 11 and Th. 13 holds for a generic P_A: for a given fading distribution, the evaluation of inner and outer bound can be refined to provide a tighter characterization of capacity. In Fig. 2a and Fig. 2b we provide such tighter characterization for the geometric and binomial distributions in (31) and (32), respectively.

Fig. 2a and Fig. 2b both contain three sets of curves for three different values of the parameter p in (31) and (32). For each value of the parameter p, we plot three curves: $R^{\mathrm{OUT}}/R^{\mathrm{IN}}$, a refined evaluation of the outer/inner bound in Th. 11 and $R^{\mathrm{IN}+}$, the inner bound in Th. 7.

V. WFFD Channel With a Continuous Fading Distribution

The results derived in Sec. IV are limited to the case of discrete fading distributions: although relevant from a theoretical standpoint, this scenario is not particularly meaningful in practical applications. In this section, we show results in Sec. IV can be extended to the case of a continuous fading distribution.

[4]More specifically, let $\Delta = \frac{1}{\sqrt{V}}$ with $V = \frac{1}{M}\frac{1-c^2\ M}{1-c^2} - \left(\frac{1}{M}\frac{(1-c^M)}{(1-c)}\right)^2$.



(a) The geometric fading distribution in (31).

(b) The binomial fading distribution in (32).

Fig. 2. Inner and outer bounds to the capacity of the WFFD channel for $P = 100$ and $c \in [1 \ldots 5]$.

We begin highlighting the sufficient conditions under which the presence of fading produces only a negligible rate loss from the AWGN capacity.

Lemma 15: Consider the WFFD channel and define κ as

$$\kappa = \frac{1}{2}\mathbb{E}_A\left[\log\left(c^2(A-\mu_A)^2+1\right)\right], \tag{33}$$

then the AWGN capacity can be attained to within κ bpcu.

The proof of Lem. 15 is omitted for brevity.

An upper bound on the capacity of the WFFD channel with continuous fading can be obtained by adapting the derivation in Th. 8.

Lemma 16 (Outer Bound for Positive-Defined Fading Distribution): Consider the WFFD channel in which $\mathbb{P}[A \geq 0] = 1$, then the capacity $\mathcal{C}$ is upper bounded as

$$\mathcal{C} \leq R^{\mathrm{OUT}} \tag{34}$$

$$= \begin{cases} \frac{1}{2}\log(P+1) + \mathcal{G}_c & c^2(1+\mu_A^2) \leq 1 \\ \frac{1}{2}\log(P+1) & \\ \quad -\frac{1}{4}\log(c^2(1+\mu_A^2)) + \mathcal{G}_c & 1 < c^2(1+\mu_A^2) < P+1 \\ \frac{1}{4}\log(P+1) + \mathcal{G}_c & c^2(1+\mu_A^2) \geq P+1, \end{cases}$$

for

$$\mathcal{G}_c = \frac{1}{2}, \mathbb{E}\left[\log\left(\frac{1+\mu_A^2}{A^2}\right)\right] + 1. \tag{35}$$

Proof: Only a sketch of the proof is provided. Define $B^N = \mathrm{sign}(S^N)$ and consider the channel in which $B^N \circ A^N$ is provided as an additional channel output, instead of A^N. The capacity of this channel is necessarily larger than the capacity of the original channel, as $A^N = |B^N \circ A^N|$, being A^N positive-defined. As $A'_i = A_i B_i \sim 1/2(P_A(a)+P_A(-a))$ is a symmetric distribution, we can adapt the converse proof in Th. 8 by similarly defining a conjugate sequence $\overline{a}'^N(a'^N) = -a'^N$ in (19). ■

Generally speaking, Lem. 16 only provides a loose upper bound to capacity, nonetheless it shows relatively simple conditions under which the capacity of the WFFD channel is at most half of the AWGN capacity. The next theorem extends the result in Th. 11 to the case of continuous fading distributions.

Theorem 17 (Outer Bound and Approximate Capacity for Narrow Fading): Consider the WFFD channel with $c \geq 1$ in which A is a continuous random variable with

$$\mathbb{P}\left[|A-\mu_A| \leq \frac{1}{c}\right] = Q_m \geq \frac{1}{2}, \tag{36}$$

then the expression in (24) for

$$G_m \leq \frac{\overline{Q}_m}{2}\log\left(1+\mu_A^2\right) + 4, \tag{37}$$

is an outer bound to capacity and the capacity is to within a gap of $G_m + 1/2$ bpcu from the outer bound in (24).

Proof: Only a sketch of the proof is provided. The achievability proof follows a similar derivation as the achievability proof of Th. 11. The converse proof is obtained in two steps: (i) first we show that the capacity of the channel with continuous fading distribution A is to within a constant gap from the capacity of the channel with discrete fading distribution A_Δ where A_Δ is obtained by uniformly quantizing A, then (ii) the result in Lem. 12 is applied to the model with fading distribution A^Δ to show the approximate capacity. In the following, we prove step (i) while only an outline of the proof of step (ii) is provided for brevity.

• **Gap from capacity.** Let the random variable A_Δ be defined as

$$\begin{aligned}\mathbb{P}[A_\Delta(A)=\gamma_k] &= \mathbb{P}[A\in I_k]\\ I_k &= \left[\mu_A+k\Delta-\frac{\Delta}{2},\mu_A+(k+1)\Delta+\frac{\Delta}{2}\right]\\ \gamma_k &= \mathbb{E}[A|A\in I_k],\end{aligned} \tag{38}$$

for $k\in\mathbb{Z}$ and some $\Delta\in\mathbb{R}^+$, that is, A_Δ is obtained by uniformly quantizing A with step size Δ and so that $\mathbb{E}[A]=\mathbb{E}[A_\Delta]$. Next, define

$$E^N = c(A^N-A_\Delta^N)\circ S^N+Z^N-Z_\Delta^N, \tag{39}$$

for $Z_\Delta^N \sim i.i.d.\ \mathcal{N}(0,1)$. An outer bound to capacity can be obtained by providing E^N in (39) to the receiver as a genie-aided side-information, that is

$$\begin{aligned}N(R-\epsilon_N) &\le I(Y^N,E^N;W|A^N)\\ &= I(Y^N-E^N,E^N;W|A^N) &(40a)\\ &= I(Y_\Delta^N;W|A^N)+I(E^N;W|A^N,Y_\Delta^N), &(40b)\end{aligned}$$

where (40a) follows from the fact that the transformation has unitary Jacobian while in (40b) follows by defining $Y_\Delta^N = X^N+cA_\Delta^N\circ S^N+Z_\Delta^N$. The term $I(E^N;W|A^N,Y_\Delta^N)$ is further bounded as

$$\begin{aligned}&I(E^N;W|A^N,Y_\Delta^N)\\ &= H(E^N|A^N,Y_\Delta^N)-H(E^N|A^N,Y_\Delta^N,W)\\ &= H(E^N|A^N,Y_\Delta^N)-H(Z^N|A^N,\overline{Z}^N,W,S^N,X^N)\\ &= H(E^N|A^N)-\frac{N}{2}\log 2\pi e\\ &\le N\max_{P_{E|A}} H(E|A)-\frac{N}{2}\log 2\pi e.\end{aligned}$$

Note that A_Δ is a deterministic function of A, so that the entropy term $H(E|A)$ can be bounded as

$$\begin{aligned}H(E|A) &\le \int_{\mathcal{A}}\frac{1}{2}\log 2\pi e\left(c^2(a-A_\Delta(a))^2+2\right)\mathrm{d}P_A\\ &\le \frac{1}{2}\log\left(c^2\Delta^2+2\right).\end{aligned}$$

From the above, we conclude that, by choosing $\Delta=1/c$ in (38), the capacity of the WFFD channel with fading distribution P_{A_Δ} is to within a gap of 1 bpcu from the capacity of the channel with fading distribution P_A. When the condition in (36) holds, the mode of A_Δ is $A_\Delta=\gamma_0\in[\mu_A-1/c,\mu_A+1/c]$ with $P_{A_\Delta}(\gamma_0)\ge 1/2$ and thus the result in Th. 11 can be applied. Note also that the distribution of A_Δ does not necessarily have unitary variance, so that Lem. 3 is invoked to normalize the fading variance. ∎

The result in Th. 17 is analogous to the result in Th. 11 as it identifies the condition under which it is approximately optimal for the transmitter to pre-code against one realization of the fading distribution while treating the remaining randomness in the fading sequence as additional noise.

Remark 18: Note that the condition in (36) can be generalized to

$$\mathbb{P}\left[|A-m|\le\frac{\kappa}{c}\right]=Q_m>\frac{1}{2}, \tag{41}$$

for some value $m\in\mathcal{A}$ to obtain a more general result than Th. 17. This yield an expression for G_m as in (27) and a gap from capacity as in (27).

The next theorem shows that there exists a class of fading distributions for which the capacity substantially reduces to the capacity of the channel without transmitter state knowledge. Let $1_{\{x\in\mathcal{X}\}}$ denote the indicator function for the set $x\in I$.

Theorem 19 (An Example With a Fat-Tailed Distribution): Consider a WFFD channel with $c>2$, then there exists a distribution of the form

$$P_A(a)=\frac{\alpha}{a}\cdot 1_{\{a\in I\}}, \tag{42}$$

such that capacity is upper bounded as

$$R^{\mathrm{OUT}}=\frac{1}{2}\log\left(1+\frac{P}{1+c^2}\right)+2, \tag{43}$$

and for which capacity is to within 3 bpcu *from the outer bound in* (43).

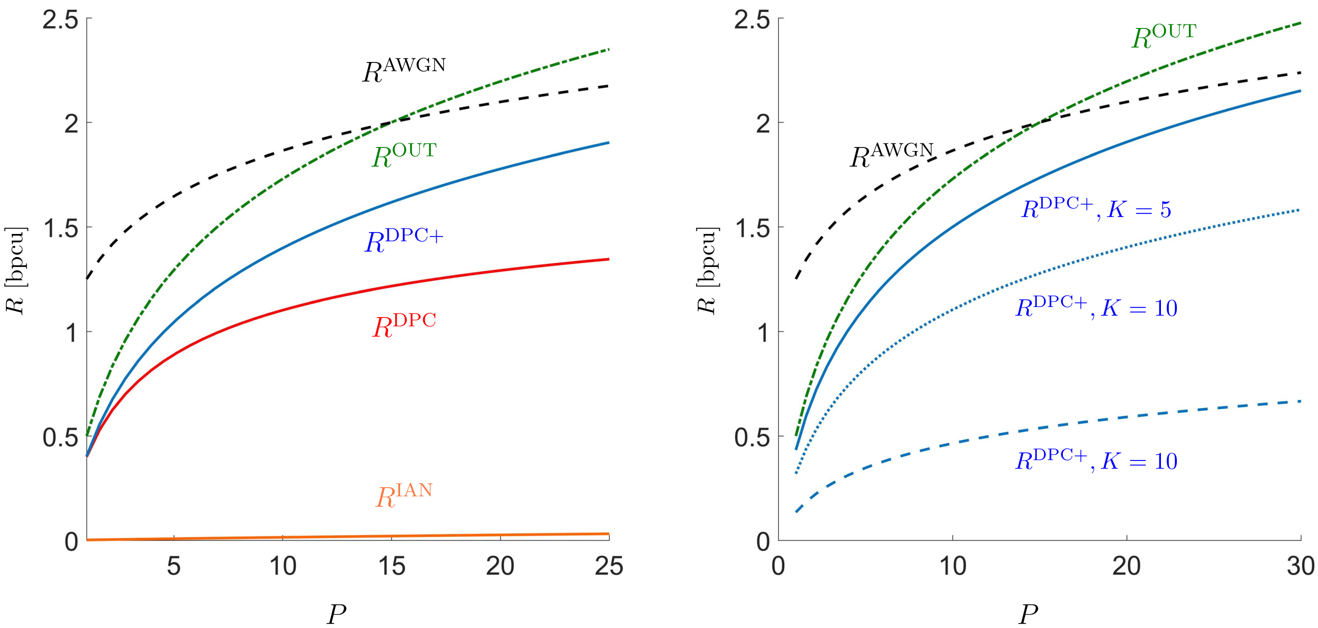
Proof: Quantizing the distribution in (42) as in Th. 17 in intervals of size $[c^{-k},c^{-(k-1)}]$ yields a random variable A_Δ which satisfies the conditions of Th. 13. The support I can be chosen as $[\kappa c^{-M-1},\kappa c^{-1}]$ for some sufficiently large M so that $(1+\mu_A^2)c^2/M\le(P+1)(M-1)/M$, thus yielding the outer bound in (43) while $\mu_A\le 1$. The achievability proof follows by treating the fading-times-state term as noise. The full proof is omitted for brevity. ∎

Remark 20: We are currently unable to determine the asymptotic behavior of capacity as c grows large: in this regime, one would expect state pre-coding to become ineffective as in Th. 19. For the case of zero mean fading this implies, in particular, that state knowledge at the transmitter does not provide any substantial rate advantage. In practical systems, state knowledge at the transmitter often come at the cost of an increase in complexity in the network architecture and transmitter design: as such, determining the fading regimes in which transmitter knowledge is rendered useless by the presence of fading is of great practical interest.

A. Numerical Examples

Although we are currently unable to characterize the capacity for many continuous fading distributions of practical relevance, such as Gaussian, Rayleigh or Rice distribution, Lem. 16 and Th. 17 do provide the first computable bounds for such fading scenarios. We present in this section numerical evaluations for the case of Rayleigh fading, in Fig. 3a, and Rice fading, in Fig. 3b. For the case of Rayleigh fading, the scale parameter is chosen so as to yield unitary variance while, for the case of Rice fading, the shape parameter is fixed and the scale parameter is chosen to yield unitary variance.[5] In Figs. 3a-3b, we plot the rate performance for a fixed c and varying P. The rate curves in these figures are labeled as following: (i) R^{AWGN} is the AWGN capacity, corresponding to the outer bound in Rem. 15, (ii) R^{OUT} is the outer bound

[5] Note that both the Rayleigh and the Rice distribution possess the scaling properties, so that the normalization in Lem. 3 preserves the fading distribution.



(a) Rayleigh fading.

(b) Rice fading for $K \in [5\ 10\ 15]$.

Fig. 3. Inner and outer bounds to the capacity of the WFFD for $c = 25$ and $P \in [1 \ldots 30]$.

in Lem. 16, (iii) R^{IAN} is the inner bound in which the fading-times-state sequence is treated as additional interference and corresponds to the achievable proof in Rem. 15. (iv) R^{DPC} corresponds to the achievable strategy in (5) which ignores the randomness in the fading realization and (v) $R^{\text{DPC+}}$ is the inner bound in Th. 7.

For Fig. 3a, we notice that the achievable scheme in Th. 7 performs surprisingly close to the outer bound in Lem. 16, especially in the low SNR regime. For Fig. 3b, observe that the achievable scheme in Th. 7 becomes less effective as the shape parameter increases. This coincides with the fact that, as the mean increases, the fading values are spread over a larger interval and thus DPC becomes less effective.

VI. Conclusions

This paper investigates the capacity of the Writing of Fast Fading Dirt (WFFD) channel, a variation of the classic "writing on dirty paper" channel in which the state sequence affected by fast fading process known only at the receiver. Accordingly, the output of the WFFD channel is obtained as the sum of the channel input, additive Gaussian noise and a fading-times-state term which is the element-wise product of the channel state, known only at the transmitter, and the fading process, known only at the receiver. We focus on the case in which the channel state is a white Gaussian process and the fading sequence is an i.i.d. sequence with either a discrete or a continuous distribution. The WFFD channel is a special case of the Gelf'and-Pinsker channel for which capacity is known: unfortunately, capacity is expressed as a solution of a maximization problem that cannot be easily determined, either in closed-form or through numerical evaluations. For this reason, we derive alternative inner and outer bounds and provide an approximate characterization of capacity for certain fading distributions.

For the WFFD channel with a discrete fading distribution, we determined capacity to within a small gap for two classes of distributions: distributions with mode larger than a half and uniform distributions in which the points in the support are incrementally spaced apart. For the WFFD channel with a continuous fading distribution, we derive the approximate capacity for the case in which more than half of the probability is concentrated in a small interval. In all these cases, capacity is approached by letting the channel input be the superposition of two codewords: a codeword treating the fading-times-state as additional noise and a codeword pre-coded against one realization of the fading times the state sequence. This relatively simple attainable strategy shows, from a high-level perspective, that robust state pre-cancellation is substantially unsuccessful for these fading distributions. This result is in sharp contrast with the performance attainable when the same fading process affects both the input and the state: in this case, fading does not essentially impede the effectiveness of interference pre-cancellation.

Appendix A
Proof of Lem. 4

Consider two sequences S_1^N and S_2^N such that $S_m^N \sim$ $i.i.d.\ \mathcal{N}(0, Q_m),\ m \in \{1, 2\}$, $S_1^N \perp S_2^N$ with $Q_1 + Q_2 = 1$ and let the channel state of the WFFD channel be obtained as $S^N = S_1^N + S_2^N$. Providing the sequence S_2^N to both the transmitter and receiver can only increase capacity, since they can disregard this extra information and operate as in the original channel. The channel in which S_2^N is provided to both encoder and decoder falls in the class of channels studied in

[16, Th. 1] for which capacity can be bounded as

$$\begin{aligned}\mathcal{C} &= \max_{X,U|S_1,S_2} I(X+cS_1A+Z,U|A,S_2) - I(U;S_1|S_2)\\ &\leq \max_{X,U|S_2,S_1} I(X+cS_1A+Z;U,S_2|A) - I(U,S_2;S_1) \quad (44a)\\ &= \max_{X,\widetilde{U}|S_1} I(X+cS_1A+Z;\widetilde{U}|A) - I(\widetilde{U};S_1), \quad (44b)\end{aligned}$$

where, (44a) follows from the independence of S_1 and S_2 by defining $\widetilde{U} = [U\ S_2]$ in (44b). Since S_2 no longer appears in (44b), it can be dropped from the maximization. From the result in Lem. 3, we have that (44b) equals the capacity of the channel in (1) for which $\widetilde{c} = c/\sqrt{Q_1}$ instead of c. From this equivalence, we conclude that the capacity of the WFFD channel is decreasing in the parameter c.

Appendix B
Proof of Th. 11

• **Achievability.** Consider the achievable strategy in Th. 8 and let the top codeword $U_{\rm PAS}$ in (13) be pre-coded against the sequence cmS^N as in the WDP channel. This assignment attains

$$\begin{aligned}R_{\rm PAS} &= \left[I(Y;U|X^{\rm SAN},A) - I(U;S)\right]^+\\ &\geq \frac{Q_m}{2}\log(1+\overline{\alpha}P)\\ &\quad + \sum_{a\in\mathcal{A}\setminus\{m\}} \frac{P_A(a)}{2}\\ &\quad \times \log\left(\frac{(1+c^2\ a^2+\overline{\alpha}P)(1+\overline{\alpha}P)}{\overline{\alpha}Pc^2(a-m)^2+\overline{\alpha}P+c^2a^2+1}\right)\\ &\geq \frac{Q_m}{2}\log(1+\overline{\alpha}P)\\ &\quad - \sum_{a\in\mathcal{A}\setminus\{m\}} \frac{P_A(a)}{2}\log\left(\frac{(a-m)^2}{a^2}+1\right),\end{aligned}$$

while the overall attainable rate $R^{\rm IN}(\alpha)$ in (14) becomes

$$\begin{aligned}R^{\rm IN}(\alpha) &= \frac{1}{2}\mathbb{E}_A\left[\log\left(1+\frac{\alpha P}{1+c^2A^2+\overline{\alpha}P}\right)\right]\\ &\quad + \frac{Q_m}{2}\log(1+\overline{\alpha}P)\\ &\quad - \sum_{a\in\mathcal{A}\setminus\{m\}} \frac{P_A(a)}{2}\log\left(\frac{(a-m)^2}{a^2}+1\right). \quad (45)\end{aligned}$$

The choice of $\overline{\alpha}P$ in (45) as

$$\overline{\alpha}^*P = \max\left\{\min\left\{\frac{Q_m}{\overline{Q}_m}c^2(1+\mu_A^2)-1, P\right\}, 0\right\}, \quad (46)$$

yields the inner bound

$$R^{\rm IN} = \begin{cases}\frac{1}{2}\log\left(1+\frac{P}{1+c^2(1+\mu_A^2)}\right) & \overline{Q}_m \geq Q_m c^2(1+\mu_A^2)\\ \frac{1}{2}\log(P+c^2(1+\mu_A^2)+1) - \frac{\overline{Q}_m}{2}\log\left(c^2(1+\mu_A^2)\right) - G_m & \overline{Q}_m \leq Q_m c^2(1+\mu_A^2) \leq \overline{Q}_m(P+1)\\ \frac{Q_m}{2}\log(1+P) - G_m & Q_m c^2(1+\mu_A^2) > \overline{Q}_m(P+1),\end{cases} \quad (47)$$

for

$$G_m = \frac{1}{2}\mathbb{E}_A\left[\log\left(\frac{(A-m)^2}{A^2}+1\right)|A\neq m\right]+1. \quad (48)$$

• **Converse.** Using Fano's, inequality we write

$$\begin{aligned}N(R-\epsilon_N) &\leq I(Y^N;W|A^N)\\ &\leq N\max_j H(Y_j|A_j)\\ &\quad - H(Y^N|W,A^N)\\ &\leq \frac{N}{2}\mathbb{E}_A\left[\log 2\pi e(P+A^2c^2+2|c||A|\sqrt{P}+1)\right]\\ &\quad - H(Y^N|W,A^N) \quad (49a)\\ &\leq \frac{N}{2}\log 2\pi e(P+c^2(1+\mu_A^2)+1)\\ &\quad - H(Y^N|W,A^N) + \frac{N}{2}, \quad (49b)\end{aligned}$$

where (49a) follows from the GME and (49b) follows from Jensen's inequality. Next, we derive a bound on the entropy term $H(Y^N|W,A^N)$ based on the properties of the set of typical fading realizations, $\mathcal{T}_\epsilon^N(P_A)$, defined as

$$\begin{aligned}&\mathcal{T}_\epsilon^N(P_A) \quad (50)\\ &= \left\{a^N,\ \left|\frac{1}{N}N(k|a^N)-P_A(k)\right| \leq \epsilon P_A(k),\ \forall\ k\ \in\mathcal{A}\right\}, \quad (51)\end{aligned}$$

where $N(k|a^N)$ is the number of symbols $k\in\mathcal{A}$ in the sequence a^N, that is

$$N(k|a^N) = \sum_{i=1}^N 1_{\{a_i=k\}}. \quad (52)$$

For the typical set in (51), we have

$$\begin{aligned}P(a^N) &\leq \frac{1}{2^{n(1+\epsilon)H(A)}}, \quad a^N\in\mathcal{T}_\epsilon^N \quad (53a)\\ \left|\mathcal{T}_\epsilon^N(P_A)\right| &\leq (1-\delta_\epsilon)2^{N(1-\epsilon)H(A)} \quad (53b)\\ N(k|a^N) &\leq NP_A(k)(a)(1-\epsilon), \quad (53c)\end{aligned}$$

for $\delta_\epsilon = 2|\mathcal{A}|e^{-N2\min_k P_A(k)}$. When the block-length N is sufficiently large, we have that $\epsilon \leq (Q_m-\frac{1}{2})/Q_m$ in (51) which implies $N(m|a^N) > 1/2$. For $N(m|a^N) > 1/2$, there exists a one-to-one mapping $\overline{a}^N(a^N):\ \mathcal{T}_\epsilon^N(P_A)\rightarrow\mathcal{T}_\epsilon^N(P_A)$ such that

$$\text{if } \overline{a}_i\neq m \text{ then } a_i=m, \quad \text{if } a_i\neq m \text{ then } \overline{a}_i=m, \quad (54)$$

that is, the sequence $\overline{a}^N(a^N)$ is obtained by permuting the $N-N(m|a^N)$ indexes for which $a_i\neq m$ with some $N-N(m|a^N)$ indexes for which $a_i=m$, while $2\,N(m|a^N)-N$ indexes are such that $a_i=\overline{a}_i=m$. Since the mapping $\overline{a}^N(a^N)$ in (54) is a one-to-one mapping of the typical set onto itself, we must have

$$\begin{aligned}&\sum_{a^N\in\mathcal{T}_\epsilon^N(P_A)} P(a^N)H(Y^N|W,A^N=a^N)\\ &= \sum_{a^N\in\mathcal{T}_\epsilon^N(P_A)} P\left(\overline{a}^N(a^N)\right)H\left(\overline{Y}^N|W,A^N=\overline{a}^N(a^N)\right), \quad (55)\end{aligned}$$

where $\overline{Y}^N$ in (55) is defined as

$$\overline{Y}^N = X^N + c\overline{a}^N S^N + \overline{Z}^N, \tag{56}$$

for $\overline{Z}^N \sim i.i.d.\ \mathcal{N}(0,1),\ \overline{Z}^N \perp Z^N$. Using the definitions above, we have that the entropy term $H(Y^N|W, A^N)$ can be bounded as

$$\begin{aligned}
&-H(Y^N|W,A^N) \quad \text{(57a)}\\
&= -\sum_{a^N\in\mathcal{A}^N} P(a^N)H(Y^N|W,A^N=a^N)\\
&\le -\sum_{a^N\in\mathcal{T}_\epsilon^N(P_A)} P(a^N)H(Y^N|W,A^N=a^N)\\
&= -\frac{1}{2}\sum_{a^N\in\mathcal{T}_\epsilon^N(P_A)} P(a^N)\left(H(Y^N|W,A^N=a^N)\right. \quad \text{(57b)}\\
&\quad \left.+H(Y^N|W,A^N=\overline{a}^N)\right)\\
&\le -\frac{1}{2}\sum_{a^N\in\mathcal{T}_\epsilon^N(P_A)} P(a^N)\\
&\quad \times\Big(H(X^N+ca^N\circ S^N+Z^N, X^N+c\overline{a}^N\circ S^N+\overline{Z}^N|W)\Big)\\
&= -\frac{1}{2}\sum_{a^N\in\mathcal{T}_\epsilon^N(P_A)} P(a^N)\\
&\quad \times H\Big(c(a^N-\overline{a}^N)\circ S^N+Z^N-\overline{Z}^N,\\
&\qquad X^N+c\overline{a}^N\circ S^N+\overline{Z}^N|W\Big)\\
&= -\frac{1}{2}\sum_{a^N\in\mathcal{T}_\epsilon^N(P_A)} P(a^N)\Big(H\Big(c(a^N-\overline{a}^N)\circ S^N+Z^N-\overline{Z}^N\Big)\\
&\quad \times +H(\overline{Y}^N|Y^N-\overline{Y}^N,W,S^N,X^N)\Big) \quad \text{(57c)}\\
&\le -\frac{1}{2}\sum_{a^N\in\mathcal{T}_\epsilon^N(P_A)} P(a^N)\\
&\quad \cdot\Big(H\Big(c(a^N-\overline{a}^N)\circ S^N+Z^N-\overline{Z}^N\Big)\\
&\quad +H(\overline{Z}^N|\overline{Z}^N-Z^N)\Big)\\
&= -\frac{1}{2}\sum_{a^N\in\mathcal{T}_\epsilon^N(P_A)} P(a^N)\\
&\quad \cdot H\Big(c(a^N-\overline{a}^N)\circ S^N+Z^N-\overline{Z}^N\Big)+\frac{N}{2}\log(\pi e), \quad \text{(57d)}
\end{aligned}$$

where (57c) follows from the fact that S^N and Z^N are independent from W. We continue the series of inequalities in (57) by noting that

$$\begin{aligned}
&\frac{1}{2}\sum_{a^N\in\mathcal{T}_\epsilon^N(P_A)} P(a^N)H\Big(c(a^N-\overline{a}^N)\circ S^N+Z^N-\overline{Z}^N\Big)\\
&\quad \le -\frac{1}{2}\frac{1}{2^{n(1+\epsilon)H(A)}} \quad \text{(58a)}\\
&\qquad \cdot\sum_{a^N\in\mathcal{T}_\epsilon^N(P_A)} H\Big(c(a^N-\overline{a}^N)\circ S^N+Z^N-\overline{Z}^N\Big)\\
&\quad \le -\frac{1}{2}\frac{1}{2^{n(1+\epsilon)H(A)}}\\
&\qquad \cdot\sum_{a^N\in\mathcal{T}_\epsilon^N(P_A)}\sum_{i=1}^N\Big(H\left(c(a_i-\overline{a}_i)S_i+Z_i-\overline{Z}_i\right)\Big), \quad \text{(58b)}
\end{aligned}$$

where (58a) follows from the bound in (53a) while (58b) follows from the fact that S^N and Z^N are i.i.d. sequences. From the definition of the mapping $\overline{a}^N(a^N)$, the sequence $a_i-\overline{a}_i$ can take three types of values: $m-k$, $k-m$ and 0 where k is any element of $\mathcal{A}\backslash\{m\}$. More specifically, $a_i-\overline{a}_i=m-k$ occurs $N(k|a^N)$ times, $a_i-\overline{a}_i=k-m$ occurs $N(k|a^N)$ times for all $k\in\mathcal{A}$ while $a_i-\overline{a}_i=0$ occurs $2(N-N(m|a^N))$ times. Using these observations, we write

$$\begin{aligned}
\text{(58b)} &= -\frac{1}{2}\frac{1}{2^{n(1+\epsilon)H(A)}}\sum_{a^N\in\mathcal{T}_\epsilon^N(P_A)}\\
&\quad \cdot\Bigg(\sum_{k\in\mathcal{A}\backslash\{m\}} 2N(k|a^N)H\left(c(m-k)S_i+Z_i-\overline{Z}_i\right)\\
&\quad +\frac{1}{2}(2N(m|a^N)-N)H\left(Z_i-\overline{Z}_i\right)\Bigg) \quad \text{(59a)}\\
&= -\frac{1}{2}\frac{1}{2^{n(1+\epsilon)H(A)}}\sum_{a^N\in\mathcal{T}_\epsilon^N(P_A)}\\
&\quad \times\Bigg(\sum_{k\in\mathcal{A}\backslash\{m\}} 2N(k|a^N)H\left(c(m-k)S_i+Z_i-\overline{Z}_i\right)\\
&\quad -\frac{1}{2}(2N(m|a^N)-N)\log(4\pi e)\Bigg) \quad \text{(59b)}\\
&\le -\frac{1}{2}\frac{1}{2^{n(1+\epsilon)H(A)}}\sum_{a^N\in\mathcal{T}_\epsilon^N(P_A)}\\
&\quad \times\sum_{k\in\mathcal{A}\backslash\{m\}} 2N(k|a^N)\frac{1}{2}\log 2\pi e(c^2(m-k)^2+2)\\
&= -\frac{1}{2^{n(1+\epsilon)H(A)}}(1-\delta_\epsilon)2^{n(1-\epsilon)H(A)} \quad \text{(59c)}\\
&\quad \cdot\sum_{k\in\mathcal{A}\backslash\{m\}} N(k|a^N)\frac{1}{2}\log 2\pi e(c^2(m-k)^2+2)\\
&= -\frac{1}{2^{n(1+\epsilon)H(A)}}(1-\delta_\epsilon)2^{n(1-\epsilon)H(A)}(1-\epsilon)N\\
&\quad \cdot\sum_{k\in\mathcal{A}\backslash\{m\}} P_A(k)(1-\epsilon)\frac{N}{2}\log 2\pi e(c^2(m-k)^2+2), \quad \text{(59d)}
\end{aligned}$$

where (59c) follows from the bound on the cardinality of the typical set in (53b) and (59d) from the definition in (51). For N is sufficiently large, we have

$$\begin{aligned}
&-H(Y^N|W,A^N)\\
&\quad \le -\sum_{k\in\mathcal{A}\backslash\{m\}} P_A(k)\frac{N}{2}\log 2\pi e\left(c^2(m-k)^2+2\right)\\
&\qquad -\frac{N\overline{Q}_m}{2}\log(2\pi e)-\epsilon_{\text{all}} \quad \text{(60a)}\\
&\quad \le -\frac{N\overline{Q}_m}{2}\log 2\pi e c^2-\frac{N}{2}\mathbb{E}_A[\log(A-m)^2|A\neq m]\\
&\qquad -\frac{N\overline{Q}_m}{2}\log(2\pi e)-\epsilon_{\text{all}}, \quad \text{(60b)}
\end{aligned}$$

for some ϵ_{all} that goes to zero as $N\to\infty$. Using the bound for (60) in (49b) and for some ϵ_{all} sufficiently small, we obtain the outer bound

$$\begin{aligned}R^{\rm OUT} &= \frac{1}{2}\log\left(2\pi e(P+c^2(1+\mu_A^2)+1)\right)\\&\quad-\frac{\overline{Q}_m}{2}\log(c^2(1+\mu_A^2))-\frac{1}{4}\log(\pi e)\\&\quad-\frac{1}{2}\mathbb{E}_A[\log\left(\frac{(A-m)^2}{1+\mu_A^2}\right)|A\neq m]+\frac{1}{2}. \qquad (61)\end{aligned}$$

Using Lem. 4, we can consider the assignment

$$\min\ \left\{\frac{\overline{Q}_m}{Q_m}(1+P), c^2(1+\mu_A)\right\}, \qquad (62)$$

for the term $c^2(1+\mu_A^2)$ in (61) which yields the expression in (24).

The gap between inner and outer bound of G'_m can be obtained by comparing the expressions in (24) and (47).

Appendix C
Proof of Th. 13

In the following, we provide the proof for the result in Th. 13 for the case of $M=3$. Only outline of the proof for the case of any $M>3$ is provided at the end of the section; the full derivation is omitted for brevity.

The achievability proof is a variation of the achievability proof of Th. 11 when letting the codeword $U_{\rm PAS}$ be pre-coded against the sequence $c\mu_A S^N$ as in the WDP channel.

• **Converse for** $M=3$**.** For $\mathcal{A}=\{\alpha_1,\alpha_2,\alpha_3\}$ and $\alpha_1\leq\alpha_2\leq\alpha_3$, we define two conjugate sequences of a^N, $\overline{a}^N_{(1)}(a^N)$ and $\overline{a}^N_{(2)}(a^N)$, as follows:

∘ the portion of a^N equal to α_1, is equal to α_2 in $\overline{a}^N_{(1)}$ and equal to α_3 in $\overline{a}^N_{(2)}$,

∘ the portion of a^N equal to α_2, is equal to α_3 in $\overline{a}^N_{(1)}$ and equal to α_1 in $\overline{a}^N_{(2)}$, and

∘ the portion of a^N equal to α_3, is equal to α_1 in $\overline{a}^N_{(1)}$ and equal to α_2 in $\overline{a}^N_{(2)}$.

From the definition of the mapping, $\overline{a}^N_{(1)}(a^N)$ and $\overline{a}^N_{(2)}(a^N)$, we have that

$$a^N\in\mathcal{T}_\epsilon^N(P_A) \iff \overline{a}^N_{(k)}(a^N)\in\mathcal{T}_\epsilon^N(P_A),\quad k\in[1,2], \qquad (63a)$$

moreover

$$\begin{aligned}&\sum_{a^N\in\mathcal{T}_\epsilon^N(P_A)} P_{A^N}(a^N)H(Y^N|W,A^N=a^N)\\&=\sum_{a^N\in\mathcal{T}_\epsilon^N(P_A)} P_{A^N}(\overline{a}^N_{(k)}(a^N))H(Y^N_{(k)}|W,A^N=\overline{a}^N_{(k)}(a^N)), \qquad (64)\end{aligned}$$

where $Y^N_{(k)}$ is defined similarly to (56) as

$$Y^N_{(k)}=X^N+c\overline{a}^N_{(k)}S^N+Z^N_{(k)}, \qquad (65)$$

for $Z^N_{(k)}\sim i.i.d.\ \mathcal{N}(0,1),\ k\in[1,2]$. Similarly to (50), Fano's inequality yields the bound

$$N(R-\epsilon_N)\leq\frac{N}{2}\log 2\pi e(P+c^2(1+\mu_A^2)+1) - H(Y^N|W,A^N)+\frac{N}{2}. \qquad (66)$$

Using the equivalence in (64), the term $H(Y^N|W,A^N)$ can be rewritten as

$$\begin{aligned}&-H(Y^N|W,A^N)\\&=-\sum_{a^N\in\mathcal{T}_\epsilon^N(P_A)}\frac{1}{3^N}H(Y^N|W,A^N=a^N)\\&=-\frac{1}{3^{N+1}}\sum_{a^N\in\mathcal{T}_\epsilon^N(P_A)}\big(H(Y^N|W,A^N=a^N)\\&\qquad+H(Y^N|W,A^N=a^N_{(1)})+H(Y^N|W,A^N=a^N_{(2)})\big)\end{aligned}$$

For $a^N\in\mathcal{T}_\epsilon^N(P_A)$, we have

$$\begin{aligned}&-H(Y^N|W,A^N=a^N)\\&\quad-H(Y^N|W,A^N=a^N_{(1)})-H(Y^N|W,A^N=a^N_{(2)})\\&\leq -H(Y^N,Y^N_{(1)},Y^N_{(2)}|W,A^N=a^N) \qquad (67a)\\&=-H\big(\{Y_i,Y_{(1),i},Y_{(2),i},\ \forall i\ a_i=\alpha_1\},\\&\quad\{Y_i,Y_{(1),i},Y_{(2),i},\ \forall i\ a_i=\alpha_2\},\\&\quad\{Y_i,Y_{(1),i},Y_{(2),i},\ \forall i\ a_i=\alpha_3\}|W,A^N=a^N\big) \qquad (67b)\\&=-H\big(\{Y_i,Y_i-Y_{(2),i},Y_{(2),i},\ \forall i\ a_i=\alpha_1\},\\&\quad\{Y_{(2),i}-Y_i,Y_{(1),i},Y_{(2),i},\ \forall i\ a_i=\alpha_2\},\\&\quad\{Y_i,Y_{(1),i}-Y_{(2),i},Y_{(2),i},\ \forall i\ a_i=\alpha_3\}|W,A^N=a^N\big), \qquad (67c)\end{aligned}$$

where (67b) follows by re-arranging the channel outputs according to the fading realization and (67c) follows from the fact this the transformation has unitary Jacobian. Consider the set

$$\begin{aligned}&\{Y_i-Y_{(1),i},\ \forall i\ a_i=\alpha_1\}\\&\quad\cup\{Y_{(2),i}-Y_i,\ \forall i\ a_i=\alpha_2\}\\&\quad\cup\{Y_{(1),i}-Y_{(2),i},\ \forall i\ a_i=\alpha_3\}, \qquad (68)\end{aligned}$$

from the definition of the conjugate sequences, we have that the set in (68) contains the elements of the vector

$$c(\alpha_2-\alpha_1)S^N+\widetilde{Z}^N_{21}, \qquad (69)$$

where $\widetilde{Z}^N_{21}$ is obtained as

$$\widetilde{Z}_{21,i}=\begin{cases}Z_i-Z_{(1),i} & a_i=\alpha_1\\ Z_{(2),i}-Z_i & a_i=\alpha_2\\ Z_{(1),i}-Z_{(2),i} & a_i=\alpha_3.\end{cases} \qquad (70)$$

Next, continuing the series of inequalities in (67), we have

$$\begin{aligned}&-3H(Y^N|W,A^N=a^N)\\&\leq -H\big(\{Y_i,Y_{(2),i},\ a_i=\alpha_1\}, \qquad (71a)\\&\quad\{Y_{(1),i},Y_{(2),i},\ a_i=\alpha_2\},\\&\quad\{Y_i,Y_{(1),i}\ a_i=\alpha_3\}\\&\quad|W,c(\alpha_2-\alpha_1)S^N+\widetilde{Z}^N_{21},A^N=a^N\big)\\&\quad-H(c(\alpha_2-\alpha_1)S^N+\widetilde{Z}^N_{21}|W)\\&\leq -H\big(\{Y_i-Y_{(2),i},Y_i,\ a_i=\alpha_1\}, \qquad (71b)\\&\quad\{Y_{(2),i}-Y_{(1),i},Y_{(2),i}\ a_i=\alpha_2\}, \qquad (71c)\\&\quad\{Y_{(1),i}-Y_i,Y_{(1),i}\ a_i=\alpha_3\}\\&\quad|W,c(\alpha_2-\alpha_1)S^N+\widetilde{Z}^N_{21},A^N=a^N\big)\\&\quad-H(c(\alpha_2-\alpha_1)S^N+\widetilde{Z}^N_{21}|W), \qquad (71d)\end{aligned}$$

where (71a) follows from the observation in (69) and (71d) follows again from the fact that this transformation has unitary Jacobian. Similarly to (68), we have that the set

$$\big\{\{Y_i - Y_{(2),i},\ a_i = \alpha_1\},\\ \{Y_{(2),i} - Y_{(1),i},\ a_i = \alpha_2\},\\ \{Y_{(1),i} - Y_i,\ a_i = \alpha_3\}\big\},$$

contains the elements of the vector

$$c(\alpha_3 - \alpha_1)S^N + \widetilde{Z}_{31}^N, \tag{72}$$

for $\widetilde{Z}_{31}^N$ defined similarly as in (70). With this observation, we write

$$\begin{aligned}
&-3H(Y^N|W, A^N = a^N)\\
&\le H\big(\{Y_i,\ a_i = \alpha_1\},\ \{Y_{(2),i}\ a_i = \alpha_2\},\ \{Y_{(1),i}\ a_i = \alpha_3\}\\
&\quad |W, c(\alpha_3-\alpha_1)S^N + \widetilde{Z}_{31}^N, c(\alpha_3-\alpha_1)S^N\\
&\quad + \widetilde{Z}_{21}^N, A^N = a^N\big)\\
&\quad - H\left(c(\alpha_3-\alpha_1)S^N + \widetilde{Z}_{31}^N|W, c(\alpha_2-\alpha_1)S^N + \widetilde{Z}_{21}^N\right)\\
&\quad - H\left(c(\alpha_2-\alpha_1)S^N + \widetilde{Z}_{21}^N|W\right) &(73a)\\
&\le H\big(\{Z_i,\ a_i = \alpha_1\}, \{Z_{(2),i}\ a_i = \alpha_2\}, \{Z_{(1),i}\ a_i = \alpha_3\} &(73b)\\
&\quad |W, S^N, \widetilde{Z}_{21}^N, \widetilde{Z}_{31}^N, A^N = a^N\big) &(73c)\\
&\quad - H\left(c(\alpha_3-\alpha_1)S^N + \widetilde{Z}_{31}^N|W, c(\alpha_2-\alpha_1)S^N + \widetilde{Z}_{21}^N\right) &(73d)\\
&\quad - H\left(c(\alpha_2-\alpha_1)S^N + \widetilde{Z}_{21}^N|W\right), &(73e)
\end{aligned}$$

We are now left with the task of evaluating the terms in (73c), (73d) and (73e) in closed-form. For the term in (73e), we have write

$$\begin{aligned}
&H(c(\alpha_2-\alpha_1)S^N + \widetilde{Z}_{21}^N)\\
&\quad = -\frac{N}{2}\log 2\pi e\left(c^2(\alpha_2-\alpha_1)^2 + 2\right)\\
&\quad \le -\frac{N}{2}\log 2\pi e\left(c^2(c^2-1)\alpha_1^2 + 2\right) &(74a)\\
&\quad \le -\frac{N}{2}\log 2\pi e\left(c^2+1\right) - \frac{1}{2}, &(74b)
\end{aligned}$$

where (74a) follows from the fact $\alpha_2 > c\alpha_1$ and (74b) follows from $\alpha_1 > 1/(c-1)$ as prescribed by (27). For the term in (73d), we have

$$\begin{aligned}
&-H(c(\alpha_3-\alpha_1)S^N + \widetilde{Z}_{31}^N|W, c(\alpha_2-\alpha_1)S^N + \widetilde{Z}_{21}^N)\\
&= -NH\left(c^2(\alpha_3-\alpha_1)(\alpha_2-\alpha_1)\left(1-\frac{c(\alpha_2-\alpha_1)}{c^2(\alpha_2-\alpha_1)^2+2}\right)S\right.\\
&\quad \left. + \widetilde{Z}_{13} - \frac{c^2(\alpha_3-\alpha_1)(\alpha_2-\alpha_1)}{c^2(\alpha_2-\alpha_1)^2+2}\widetilde{Z}_{12}\right) &(75a)\\
&\le -\frac{N}{2}\log 2\pi e\left(1 + c^2\frac{(\alpha_3-\alpha_1)^2}{2+c^2(\alpha_2-\alpha_1)^2}\right)\\
&\le -\frac{N}{2}\log 2\pi e\left(1 + c^2\frac{a^2(c-1)^2}{2+c^2(\alpha_2-(c-1)^{-1})^2}\right)\\
&\le -\frac{N}{2}\log 2\pi e\left(1 + \frac{1}{2}c^2\right), &(75b)
\end{aligned}$$

where, in (75b), we have used the fact that $\alpha_3 > c\alpha_2$, $\alpha_1 > 1/(1-c)$ and $c > 2$ by assumption. Finally, the term in (73c) only contains independent noise terms, so that

$$\begin{aligned}
&H\big(\{Z_i,\quad a_i = \alpha_1\},\\
&\qquad \{Z_{(2),i}\ a_i = \alpha_2\},\ \{Z_{(1),i}\ a_i = \alpha_3\}|\widetilde{Z}_{21}^N,\ \widetilde{Z}_{31}^N, A = a^N\big)\\
&= \frac{N}{2}\log\left(\frac{1}{3}\right). &(76)
\end{aligned}$$

Combining the bounds in (74), (75) and (76) we finally obtain the outer bound

$$\begin{aligned}
R^{\mathrm{OUT}} &\le \frac{1}{2}\log(P + c^2(1+\mu_A) + 1)\\
&\quad - \frac{3}{4}\log(c^2(1+\mu_A)) + \frac{3}{4}\log(1+\mu_A). &(77)
\end{aligned}$$

The final outer bound expression in (28) is obtained by using Lem. 4 to tighten the expression (77) with the appropriate choice of c. The gap between inner and outer bound is obtained similarly to Th. 11.

• **Sketch of the converse for** $M > 3$**.** The derivation for the case $M > 3$ is obtained by extending the derivation for $M = 3$ as follows. First, we define $M-1$ conjugate sequences as

$$\overline{a}_{(k)}^N(a^N) = \left\{a_i = \alpha_j \implies a_{(k),i} = \alpha_{\mathrm{mod}(k+j,M)}\right\}, \tag{78}$$

for $k \in [1\ldots M-1]$. Next, the bounding in (73) can be repeated recursively $M-1$ times: this yields $M-1$ terms of the form $H(\Delta_i S + \widetilde{Z}_i|\Delta_1\ S + \widetilde{Z}_1 \ldots \Delta_{i-1}S + \widetilde{Z}_{i-1})$ for $\Delta_i = \alpha_{i+1} - \alpha_1$, $\widetilde{Z}_{i1} = Z_i - Z_1$ as in (70), and

$$\begin{aligned}
&H(\Delta_i S + \widetilde{Z}_{(i+1)1}|\Delta_1 S + \widetilde{Z}_{21}\ldots\Delta_{i-1}S + \widetilde{Z}_{i1})\\
&\qquad = \frac{1}{2}\log\left(2\frac{c^2(\sum_{j=1}^i \Delta_j^2) + 2}{c^2\left(\sum_{j=1}^{i-1}\Delta_j^2\right) + 2}\right). &(79)
\end{aligned}$$

The conditions in (28) guarantee that

$$\begin{aligned}
&H(\Delta_i S + \widetilde{Z}_{(i+1)1}|\Delta_1 S + \widetilde{Z}_{21}\ldots\Delta_{i-1}S + \widetilde{Z}_{i1})\\
&\qquad \le -\frac{1}{2}\log(2\pi e c^2) + \frac{1}{2}, &(80)
\end{aligned}$$

for each $i \in [1\ldots M-1]$, yielding an outer bound in the spirit (77)

$$R^{\mathrm{OUT}} \le \frac{1}{2}\log(P + c^2(1+\mu_A) + 1) - \frac{M-1}{M}\log(c^2) \tag{81}$$

which can be tightened over the parameter c using Lem. 4. This tightening step finally yields the outer bound in (28).

ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers for their helpful and constructive comments that greatly contributed to improving the final version of the paper. They would also like to thank the editors for their generous comments and support during the review process.

REFERENCES

[1] M. Costa, "Writing on dirty paper," *IEEE Trans. Inf. Theory*, vol. IT-29, no. 3, pp. 439–441, May 1983.

[2] S. I. Gel'fand and M. S. Pinsker, "Coding for channel with random parameters," *Problems Control Inf. Theory*, vol. 9, no. 1, pp. 19–31, 1980.

[3] A. J. 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, May 2007.

[4] S. Rini and S. Shamai (Shitz), "On the capacity of the carbon copy onto dirty paper channel," *IEEE Trans. Inf. Theory*, vol. 63, no. 9, pp. 5907–5922, Sep. 2017.

[5] W. Zhang, S. Kotagiri, and J. N. Laneman, "Writing on dirty paper with resizing and its application to quasi-static fading broadcast channels," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2007, pp. 381–385.

[6] A. Hindy and A. Nosratinia, "Ergodic fading MIMO dirty paper and broadcast channels: Capacity bounds and lattice strategies," *IEEE Trans. Wireless Commun.*, vol. 16, no. 8, pp. 5525–5536, Aug. 2017.

[7] P. Grover and A. Sahai, "On the need for knowledge of the phase in exploiting known primary transmissions," in *Proc. IEEE Int. Symp. New Frontiers Dyn. Spectr. Access Netw. (DySPAN)*, Apr. 2007, pp. 462–471.

[8] S. Rini and S. Shamai (Shitz), "The impact of phase fading on the dirty paper coding channel," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2014, pp. 2287–2291.

[9] Y. Avner, B. M. Zaidel, S. Shamai (Shitz), and U. Erez, "On the dirty paper channel with fading dirt," in *Proc. IEEE Elect. Electron. Eng. Isr. (IEEEI)*, Nov. 2010, pp. 525–529.

[10] A. Bennatan, V. Aggarwal, Y. Wu, A. R. Calderbank, J. Hoydis, and A. Chindapol, "Bounds and lattice-based transmission strategies for the phase-faded dirty-paper channel," *IEEE Trans. Wireless Commun.*, vol. 8, no. 7, pp. 3620–3627, Jul. 2009.

[11] A. Khina and U. Erez, "On the robustness of dirty paper coding," *IEEE Trans. Commun.*, vol. 58, no. 5, pp. 1437–1446, May 2010.

[12] Z. Al-Qudah and W. A. Shehab, "Bounds on the achievable rates of faded dirty paper channel," *Int. J. Comput. Netw. Commun.*, vol. 9, no. 1, pp. 71–79, 2001.

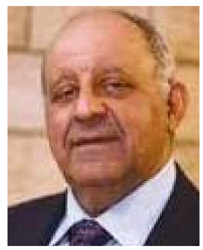
[13] I. Bergel, D. Yellin, and S. Shamai (Shitz), "A lower bound on the data rate of dirty paper coding in general noise and interference," *IEEE Wireless Commun. Lett.*, vol. 3, no. 4, pp. 417–420, Aug. 2014.

[14] S. Rini and S. Shamai (Shitz), "On capacity of the dirty paper channel with fading dirt in the strong fading regime," in *Proc. IEEE Inf. Theory Workshop (ITW)*, Nov. 2014, pp. 561–565.

[15] 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, Jan. 2008.

[16] 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, Jun. 2002.

Stefano Rini received the B.A. degree in computer science from the Politecnico di Milano, Italy, in 2005, and the M.S. degree in electrical and computer engineering and statistics and the Ph.D. degree in electrical and computer engineering, under the supervision of Prof. Tuninetti and co-advised by Prof. Devroye, with a thesis on Cognition and Cooperation in Wireless Networks: An Information Theoretic Perspective from The University of Illinois at Chicago, Chicago, IL, USA, in 2009 and 2011, respectively. In 2012, he was a Post-Doctoral Fellow with the Institute for Communications Engineering, Technical University of Munich, Germany, with Prof. Kramer. In 2013, he was also a Post-Doctoral Fellow with the Department of Electrical Engineering, Stanford University, USA, with Prof. Goldsmith. In 2014, he joined National Chiao Tung University, Taiwan, where he is currently an Assistant Professor with the Department of Electrical and Computer Engineering.



Shlomo Shamai (Shitz) (S'80–M'82–SM'88–F'94–LF'19) received the B.Sc., M.Sc., and Ph.D. degrees in electrical engineering from the Technion—Israel Institute of Technology, in 1975, 1981, and 1986, respectively.

From 1975 to 1985, he was with Communications Research Labs, as a Senior Research Engineer. Since 1986, he has been with the Department of Electrical Engineering, Technion—Israel Institute of Technology, where he is currently a Technion Distinguished Professor and the William Fondiller Chair of Telecommunications. His research interests encompass a wide spectrum of topics in information theory and statistical communications.

He is an URSI Fellow, a member of the Israeli Academy of Sciences and Humanities, and a Foreign Member of the U.S. National Academy of Engineering. He was a recipient of the 2011 Claude E. Shannon Award, the 2014 Rothschild Prize in Mathematics/Computer Sciences and Engineering, and the 2017 IEEE Richard W. Hamming Medal.

Dr. Shamai received the 1999 van der Pol Gold Medal from the Union Radio Scientifique Internationale. He was a co-recipient of the 2000 IEEE Donald G. Fink Prize Paper Award, the 2003 and 2004 joint IT/COM Societies Paper Award, the 2007 IEEE Information Theory Society Paper Award, the 2009 and 2015 European Commission FP7, Network of Excellence in Wireless COMmunications (NEWCOM++ and NEWCOM#) Best Paper Awards, the 2010 Thomson Reuters Award for International Excellence in Scientific Research, the 2014 EURASIP Best Paper Award (for the *EURASIP Journal on Wireless Communications and Networking*), and the 2015 IEEE Communications Society Best Tutorial Paper Award. He was also a recipient of the 1985 Alon Grant for distinguished young scientists and the 2000 Technion Henry Taub Prize for Excellence in Research. He has also served twice on the Board of Governors for the Information Theory Society. He has also served on the Executive Editorial Board for the IEEE TRANSACTIONS ON INFORMATION THEORY and the IEEE Information Theory Society Nominations and Appointments Committee. He has served as an Associate Editor of the Shannon Theory for the IEEE TRANSACTIONS ON INFORMATION THEORY.