On the Dirty Paper Channel with Fast Fading Dirt

Stefano Rini* and Shlomo Shamai (Shitz)†

* National Chiao-Tung University, Hsinchu, Taiwan, E-mail: stefano@nctu.edu.tw
† Technion-Israel Institute of Technology, Haifa, Israel, E-mail: sshlomo@ee.technion.ac.il

Abstract

Costa’s “writing on dirty paper” result establishes that full state pre-cancellation can be attained in the Gel’fand-Pinsker problem with additive state and additive white Gaussian noise. This result holds under the assumptions that full channel knowledge is available at both the transmitter and the receiver. In this work we consider the scenario in which the state is multiplied by an ergodic fading process which is not known at the encoder. We study both the case in which the receiver has knowledge of the fading and the case in which it does not: for both models we derive inner and outer bounds to capacity and determine the distance between the two bounds when possible. For the channel without fading knowledge at either the transmitter or the receiver, the gap between inner and outer bounds is finite for a class of fading distributions which includes a number of canonical fading models. In the capacity approaching strategy for this class, the transmitter performs Costa’s pre-coding against the mean value of the fading times the state while the receiver treats the remaining signal as noise. For the case in which only the receiver has knowledge of the fading, we determine a finite gap between inner and outer bounds for two classes of discrete fading distribution. The first class of distributions is the one in which there exists a probability mass larger than one half while the second class is the one in which the fading is uniformly distributed over values that are exponentially spaced apart. Unfortunately, the capacity in the case of a continuous fading distribution remains very hard to characterize.

Index Terms

Gel’fand-Pinsker Problem; Writing on Fading Dirt; Ergodic Fading; Imperfect Channel Side Information;

I. INTRODUCTION

In the Gel’fand-Pinsker (GP) model [1] the output of a point-to-point memoryless channel is obtained as a function of the channel input, a noise term and a state variable which is non-causally provided to the transmitter but is unknown at the receiver. In this channel the state may represent the interference caused by another user in a wireless network which is also communicated to the transmitter by the network infrastructure. In the original set up, both transmitter and receiver are assumed to have perfect channel knowledge: while it is reasonable to assume that a transmitter knows the channel toward its intended receiver and vice-versa, it is not always realistic to suppose that a transmitter knows the channel between an interfering user and the receiver. This is especially true in wireless network where the channel conditions vary continuously over time and reliable channel estimates are hard to obtain. For this reason, we wish to investigate ultimate rate advantages that can be gained when only partial channel knowledge is available at each user.

The “writing on dirty paper” result from Costa [2] establishes a closed-form characterization of the capacity of the GP problem in the additive state and additive white Gaussian noise setting. Perhaps surprisingly, the presence of the state does not reduce the capacity of this model, regardless of the distribution or power of this sequence. In this work we are interested in characterizing the effect of fading on the capacity of this model and determine the optimal transmission strategies in this scenario. In the literature, different variations of Costa’s set up which also include fading have already been considered. The “writing on fading dirt” channel in [3] is a variation of the channel of [2] in which both the channel input and the state sequence are multiplied by a fading value known at the receiver but not at the transmitter. The authors of [3] evaluate the achievable region with Costa’s assignment and show that the rate loss from full state pre-cancellation is vanishing in both the ergodic and quasi-static fading case. In the “compound dirty-paper” channel of [4] only the state is multiplied by a quasi-static fading coefficient know at the receiver but unknown at the transmitter. For this model, an inner bound based on lattice strategies is derived to compensate for the channel uncertainty at the transmitter. Achievable rates under Gaussian signaling and lattice strategies for this channel are derived in [5] while an outer bound for the vector case is derived in [6]. Outer and inner bounds to the capacity of the writing on fading dirt channel with phase fading are derived in [7]. The approximate capacity of this channel is obtained in [8] for the case of binomial and uniform phase fading case.

In this paper we study the “writing on fading dirt” model in which the state sequence is multiplied by an ergodic fading coefficient which is not known at the transmitter. We derive inner and outer bounds to capacity for both the case in which the fading is known at the receiver and for the case in which it is not. When neither the transmitter nor receiver have fading knowledge, we show that the outer bound can be attained to within a finite gap for a class of fading distribution which includes the Gaussian, the uniform and the Rayleigh distribution but does not include the log-normal distribution. For the case in which only the receiver has fading knowledge, we show a finite gap between inner and outer bound for two classes of discrete
Fig. 1. The Dirty Paper Channel with Fast Fading Dirty (DPC-FFD). The dotted line represent the state information provided at the transmitter.

Fig. 2. The Dirty Paper Channel with Fast Fading Dirty and Receiver Channel Side Information (DPC-FFD-RCSI).

distributions: when the fading distribution has a mass function greater than a half and when it is uniformly distributed over a set of points that are exponentially spaced apart.

The remainder of the paper is organized as follows: Sec. II introduces the channel model and the some related results. Sec. III investigates the capacity when neither the transmitter nor the receiver have fading knowledge while Sec. IV the case when only the receiver has fading knowledge. Finally, Sec. V concludes the paper.

II. DIRTY PAPER CHANNEL WITH FAST FADING DIRTY

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

\[ Y_i = X_i + c A_i S_i + Z_i, \quad i \in \{1 \ldots N\}, \]

for \( c \in \mathbb{R} \) and where \( X_i \) is the channel input, \( S_i \) the state, \( A_i \) the fading realization and \( Z_i \) the additive noise. The channel input \( X_i \) is subject to a second moment constraint \( \mathbb{E} [ |X_i|^2 ] \leq P \) while the state \( S_i \) and the noise term \( Z_i \) are distributed as

\[ S_i \sim \mathcal{N}(\mu_S, 1), \quad Z_i \sim \mathcal{N}(0, 1), \quad i.i.d. \]

where \( \mathcal{N}(\mu, \sigma^2) \) indicates the Gaussian Random Variable (RV) with mean \( \mu \) and variance \( \sigma^2 \). The fading RV \( A_i \) is drawn from a distribution \( p_A \) which has variance one and mean \( \mu_A \). The state sequence \( S^N \) is assumed to be non-causally available at the transmitter while fading sequence \( A^N \) is unknown at both the transmitter or the receiver.

A related model to the DPC-FFD in Fig. 1 is the model in which the fading sequence is provided to the receiver. We refer to this model as the Dirty Paper Channel with Fast Fading Dirty and Receiver Channel Side Information (DPC-FFD-RCSI), also depicted in Fig. 2. For the DPC-FFD-RCSI the receiver side information can be seen as an additional channel output, that is, the channel output is the vector \([Y_i, A_i]\) for \( Y_i \) in (1).

Remark II.1. Mean of the state and the fading. The channel output in (1) can be rewritten as

\[
Y = X + c (A_0 - \mu_A) (S_0 - \mu_S) + Z = X + c (A_0 S_0 - \mu_A S_0 - \mu_A \mu_S) + Z, \tag{3}
\]

where \( A_0 = A - \mu_A \) and \( S_0 = S - \mu_S \). That Each of the term in (3) can be seen as follows

- \( c \mu_A A_0 \) can be cancelled at the receiver when it posses fading knowledge. Without receiver fading knowledge, this term is unknown at both the receiver and the transmitter and is equivalent to additive noise.
- \( c \mu_A S_0 \) can be pre-cancelled with Costa coding by the transmitter as in (2) (Costa pre-coding in the following).
- \( c A_0 S_0 \) requires the cooperation of both transmitter and receiver, since they each have a knowledge of one of the terms of the multiplication.
A. Generalized channel

Consider a more general channel than that of (1) in which
\[ \tilde{S}_i \sim \mathcal{N}(\mu, \sigma_S^2), \quad \tilde{Z}_i \sim \mathcal{N}(0, \sigma_Z^2), \]
\[ \tilde{A} \quad \text{s.t.} \quad \mathbb{E}[\tilde{A}] = \mu_A, \quad \forall \var{\tilde{A}} = \sigma_A^2, \]
and \( \tilde{Y}_i = \tilde{X}_i + c\tilde{A}_i\tilde{S}_i + \tilde{Z}_i \), for some \( c \in \mathbb{R} \) and where \( \tilde{X} \) is subject to the average power constraint \( \mathbb{E}[[\tilde{X}]^2] \leq \tilde{P} \). The model in (1) is shown to be identical to the model in (8) by considering the equivalent channel output
\[ Y = \frac{\tilde{Y}_i}{\sigma_Z} = \frac{\tilde{X}_i}{\sigma_Z} + c\frac{\sigma_A\sigma_S}{\sigma_Z} \left( \frac{\tilde{A}_i}{\sigma_A} \tilde{S}_i \right) + \frac{\tilde{Z}_i}{\sigma_Z}, \]
which has the same distribution as the channel output in (1) for
\[ c = \frac{c}{\sigma_A\sigma_S/\sigma_Z}, \quad S = \frac{\tilde{S}_i}{\sigma_S}, \quad A = \frac{\tilde{A}_i}{\sigma_A}, \]
\[ X = \tilde{X}_i/\sigma_Z, \quad P = \tilde{P}/\sigma_Z. \]
The equivalence in (5) implies that we can re-scale the fading or the state without affecting the capacity of the channel, as long as the equivalence in (6) is preserved. This means that one cannot properly distinguish the magnitude of the fading with respect to the magnitude of the state. This fact, intuitively, has important implication on the capacity of the channel since even a small amount of fading can be amplified by a large state variance.

B. Related Results

Gelfand-Pinsker channel. The DPC-FFD and the DPC-FFD-RCSI are a special case of the GP problem for which capacity is obtained in (1).

Theorem II.2. Capacity of the DPC-FFD(-RCSI) (1). The capacity \( \mathcal{C} \) of the DPC-FFD in (8) is
\[ \mathcal{C} = \max_{P_U, X_S} I(Y; U) - I(U; S), \]
while the capacity of the DPC-FFD-RCSI is obtained from (7) by considering the channel output \( [Y, A] \).

The expression in (7) contains an auxiliary RV \( U \) and is expressed as the maximization over the distribution \( P_U, X_S \). For thin reason a closed-form expression cannot be evaluated easily, either analytically or numerically.

Dirty paper channel with receiver side information and phase fading. In [8], we have derived the approximate capacity of the DPC-FFD-RCSI for the case in which \( p_A \) is a circularly binomial distribution.

Theorem II.3. Capacity of the DPC-FFD-RCSI with circularly binomial fading [8, Th. IV.5]. Consider the DPC-FFD-RCSI
\[ Y_i = X_i + e^{j\theta_i}S_{R,i} + Z_i, \quad i \in [1...N], \]
where the state \( S_{R,i} \) is a Gaussian RV with zero mean and variance \( Q \) and while the fading is \( A = \exp\{\theta\} \) for
\[ P_0(t) = \frac{1}{2} \left( 1 + \frac{\Delta}{|t|} \right) + \frac{1}{2} \left( 1 - \frac{\Delta}{|t|} \right), \quad \Delta \in [0, \pi/2], \]
then, if \( \pi/4 \leq \Delta \leq \pi/2 \), the capacity lies to within constant gap of 3 bits per channel use from the outer bound
\[ R_{\text{OUT}} = \left\{ \begin{array}{ll} \log(P + 1) + 2 & c^2 \leq 1 \\ \frac{3}{2} \log(P + 1) + 2 & c^2 \geq 1 \\ \frac{1}{2} \log(P + 1) + \frac{1}{2} \log(1 + (\sqrt{P} + c)^2) & c^2 \leq P + 1 \\ -\frac{1}{4} \log(2c^2) + 2 & 1 < c^2 < P + 1 \end{array} \right. \]
where \( c = \sin(\Delta)\sqrt{Q} \).

Carbon copying onto dirty paper. A model related to the DPC-FFD is the “carbon copying onto dirty paper” of [9]. In this channel model there are \( M \) possible state sequences \( S_j \) that can possibly affect in the channel output. The transmitter has knowledge of each sequence but does not know which one will appear. Correct decoding must be granted regardless of the state realization and for each of the possible channel outputs,
\[ Y_j^N = X^N + cS_j^N + Z_j^N, \quad j \in [1...M], \]
where \( S_j^N \) is an i.i.d. Gaussian sequences. In [9] inner and outer bound to the capacity region are derived but capacity has yet to been determined.
III. THE DIRTY PAPER CHANNEL WITH FAST FADING DIRT

We begin by investigating the capacity of DPC-FFD in Fig 1 since no closed-form expression for the optimization in (7) is available, we derive a novel outer bound that is expressed solely as a function of the channel parameters. This outer bound can be approached by a simple achievable strategy in which the transmitter performs Costa pre-coding against the term c\(\mu_A S\), the average realization of the fading times state.

For the DPC-FFD the term c\(\mu_A S\) acts as additional noise, since it is unknown at both the transmitter and the receiver: this reason in the following we assume that \(\mu_S = 0\).

**Theorem III.1. Outer bound and partial approximate capacity for DPC-FFD.**

Consider the DPC-FFD in Fig. 1 and let \(h(A) = \frac{1}{2} \log(2\pi e^2)\) for some \(\alpha \in [0, 1]\), then the capacity \(C\) is upper bounded as

\[
C \leq R^{\text{OUT}} = \frac{1}{2} \log \left( \frac{P + \frac{1}{c^2\alpha}}{e^2\alpha} + \frac{1}{\alpha} \right) + \frac{1}{2},
\]

and the capacity is to within a gap \(G\) bits/channel-use from \(R^{\text{OUT}}\) where

\[
G = -\frac{\log(\alpha)}{2} + \frac{1}{2}.
\]

**Proof:** The proof can be found in App. A.

The gap from capacity in Th. III.1 can be easily evaluated for some canonical fading distributions.

**Lemma III.2. Gap from for some fading distributions.**

- When \(A\) is Gaussian distributed with mean \(\mu_A\) and unitary variance, the capacity is known to within a gap \(G_{\gamma}\)

\[
G_{\gamma} = \frac{1}{2}.
\]

- When \(A\) is uniformly distributed between \([\mu_A - \frac{3}{2}, \mu_A + \frac{3}{2}]\), the capacity can be attained to within a gap \(G_{y}\)

\[
G_{y} = -\frac{1}{2} \log \left( \frac{2\pi e}{12} \right) + \frac{1}{2} \leq 1.
\]

- When \(A\) is Rayleigh distributed, i.e. \(A = \sqrt{U^2 + V^2}\) for \(U, V \sim \mathcal{N}(0, 2/(4 - \pi))\) and independent, capacity can be attained to within a gap \(G_{R}\) defined as

\[
G_{R} = -\frac{1}{2} \log(1) + \gamma + 1 + \frac{1}{2} \leq 2.08,
\]

where \(\gamma\) is the Euler-Mascheroni constant.

- When \(A\) is log-normal distributed, i.e. \(A = \sigma^2 e^{-\sigma^2} (e^{\sigma^2} - 1)^{-1}\) for \(Z_A \sim (\mu, \sigma^2)\), capacity can be attained to within a gap \(G_{\log}\) defined as

\[
G_{\log} = \log \left( e^{\sigma^2} - 1 \right) + \mu + \sigma^2 + \frac{1}{2} \leq \mu + 2\sigma^2 - \frac{1}{2},
\]

which is not a finite value for all values of \(\mu\) and \(\sigma^2\).

The result in Th. III.1 is substantially a negative result since it establishes that, for a number of fading distributions for which \(\alpha\) is close to one, the best strategy is to Costa pre-code against the mean value of the fading times the state and treat the term \(A_0S_0\) as additional noise. This strategy performs very poorly when compared to the full state pre-cancellation and indeed, for any choice of the power \(P\), capacity tends to a small constant as the term \(c^2\) increases.

Note that the gap \(G\) in (12) for the log-normal distribution is not bounded: the variance of this distribution grows exponentially with \(\sigma^2\) while the entropy grows logarithmically with \(\sigma^2\), therefore \(\alpha\) can be made arbitrarily small and \(G\) arbitrarily large.

In actuality, we expect the outer bound in (11) to be close to capacity for a larger set of distributions than that for which \(\alpha\) is close to one. The difficulty in developing a more general result lies in the lack of tighter outer bound.

Note also that this result does not hold for discrete fading distributions and thus does not include extensions of the result in Th. III.3 for the case with no RCSI.

IV. DIRTY PAPER CHANNEL WITH FAST FADING DIRT AND RECEIVER SIDE INFORMATION

We now turn our attention to the DCP-FFD-RCSI: also for this channel capacity can be obtained from Th. III.2 but the optimization is extremely hard to express in closed-form. This case is significantly harder to study than the case with no receiver fading information because of the distributed way in which transmitter and receiver can cooperate in dealing with the term \(cAS\). As an illustrative example, consider the DPC-FFD with no additive noise and in which the state and the input are...
restricted to take value $\pm 1$, that is

$$Y = X + AS, \quad X, S \in \{-1, 1\},$$

while $A$ has any distribution. Given the cardinality of the input, the capacity of this channel is at most 1 bit/channel-use. This rate can be attained by setting $X(-1) = X(+1) = 1/2$, independent from $S$ and by setting $U = XS$ and independent from $S$ in (7). With this assignment, $U$ can be recovered from the channel output by considering the squared channel output, in fact:

$$(Y^2 | A = a) = X^2 + a^2 S^2 + 2a XS = 1 + a^2 + 2a U,$$

so that $U = (Y^2 - 1 - A^2)/2A$, regardless of the distribution of $A$. This simple example shows that the maximization in (7) might yield some unexpected results.

Given the difficulty of the problem at hand, we are able to make only partial progress in characterizing the capacity of the DPC-FFD-RCSI. In the following we provide two approximate capacity results for two classes of discrete distributions of $A$: (i) for the class of discrete distributions in which one of the probability masses is larger or equal to one half and (ii) for the class of uniform distributions over the discrete set in which points are incrementally spaced apart. Both results are a generalization of our previous result in Th. IV.1 and employ a similar inner bound in which the transmitter simply performs Costa pre-coding against one realization of the fading times the state. Our contributions is, therefore, to identify a set of channels in which Costa pre-coding is optimal, although it is clear that this coding strategy might not be capacity achieving in general.

Note that, for the DPC-FFD-RCSI, we again consider the case in which $\mu_S$ is equal to zero; since the receiver has knowledge of $A$, it can subtract $c\mu_S A$ from the channel output.

Let’s consider first the class of distribution in which there exists an outcome $A = a'$ with $P_A(a') \geq 1/2$: this class of distributions generalizes the distribution considered in our result in Th. IV.1. For this fading model the transmitter can Costa pre-code against the realization $a'S$ and obtain full state cancellation for approximatively a portion $P_A(a')$ of the time. The performance of this strategy can be improved upon letting the channel input be composed of two codewords: one treating the state times fading as noise and one that Costa pre-codes against $a'S$. By optimizing over the power allocated to each codeword, one obtains a larger inner bound.

**Theorem IV.1. Approximate capacity for a discrete distribution with a mass larger than half.**

Consider a DPC-FFD-RCSI in Fig. 2 and let $A$ have a discrete distribution $P_A(a')$ with support $\mathcal{A}$ where there exists $A = a'$ such that $P_A(a') \geq 1/2$. Define moreover

$$P'_A = P_A(a'), \quad P'_A = 1 - P_A(a')$$

$$G = P'_A \mathbb{E}[\log(\frac{a-a'}{|a|})^2] = a \neq a'$$

$$G' = P'_A \mathbb{E} \left[ \log \left( \frac{(a-a')^2}{a^2} + 1 \right) | a \neq a' \right],$$

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

$$\mathcal{C} \leq R^\text{OUT} = \begin{cases} \frac{1}{2} \log(1 + P) + 1 & P'_A \leq P'_A c^2 \\ \frac{P'}{2} \log(1 + P) & P'_A c^2 \leq P'_A (P + 1) \\ \frac{P'}{2} \log(P c^2) + 1 - G & P'_A > P'_A (P + 1) \end{cases}$$

an the capacity lies to within $G' - G + 3$ bits per channel use from $R^\text{OUT}$.

**Proof:** The proof can be found in App. B.

The result of Th. IV.1 can be evaluated for some discrete fading distributions.

**Lemma IV.2. Gap from for some discrete distributions.**

- When $A$ is distributed according to a geometric distribution, i.e.

$$P_A(k_a + n\Delta) = (1 - p)^n p, \quad n \in \mathbb{N},$$

for some $p \in [0, 1], \Delta > 0$ and $p^2 \Delta^2 = \overline{p}$ (to obtain a unitary variance), Th. IV.1 can be applied for $p \leq 1/2$. For this choice of $p$, $A = k_a$ has probability larger than a half and the best strategy for the transmitter is to Costa pre-code against the sequence $ck_aS$ or otherwise treat the fading times state as noise. The value of the outer bound in (15) depends on the value $G$, while...
the gap from capacity on $G'$ which are obtained as
\[
G = 2 \sum_{n=1}^{\infty} \log (n\Delta) p(1-p)^n \geq -(1-p) \log \Delta^2 \\
G' = \sum_{n=1}^{\infty} \log \left( \frac{n^2\Delta^2}{(k_a+\Delta n)^2} + 1 \right) (1-p)^n \leq \frac{1}{2k_a} (1-p),
\]
for which
\[
G' - G \leq (1-p)(k_a^{-2} + \log \Delta^2).
\]
The gap between inner and outer bound goes to infinite as $\Delta$ goes to zero: in this regime the channel reduces to the classic DPC with no fading for which the bounding techniques in Th. II.2 are no longer tight. Note that (17) goes to infinity as $k_a$ goes to zero, but this is only a consequence of the bounding in (16).

- **Binomial Distribution.** Consider now the case in which $A$ has a binomial distribution of the form
\[
p_A(k_a+n\Delta, N) = \binom{2N}{n} (1-p)^n p^{2N-n}, \quad n \in [-N \ldots +N],
\]
and $2Np(1-p) = \Delta^2$ to maintain the variance unitary. By simple enumeration we see that for $N > 1$ no assignment of $p$ gives a probability mass larger than a half. For $N = 1$ we have only one $p$ which makes the theorem applicable: $p = 1/2$ which corresponds to the probability vector $\{1/4,1/2,1/4\}$. This result extends the case where the probability vector is $\{1/2,1/2\}$ which corresponds to the case in Th. II.2.

Another possible extension of the result in Th. II.2 is the case in which $A$ is uniformly distributed over a set with more than two elements. In the following we indeed show such a generalization: the caveat is that the points in the support of the distribution must be increasingly spaced apart. This result is similar in spirit to our result in [10] for the DPC with slow fading, that is for the channel in which a fading coefficient is randomly drawn from a set of possible values before transmission and is constant for the whole block-length. The intuitive interpretation of this result is as follows: when two fading value are sufficiently spaced apart, the transmitter cannot exploit the correlation between the two different channel outputs corresponding to the two different fading realizations. For this reason the best choice for the transmitter is to Costa pre-code against one realization of the fading times state. This argument can be repeated for any two pair of fading values to obtain a condition that holds for any set of points.

**Theorem IV.3.** Approximate capacity in the “strong fading” regime.
Consider the case in which $A$ is uniformly distributed over the set
\[
\mathcal{A}(M) = \{a_1 \leq a_2 \leq \ldots \leq a_M, \quad a_i \in \mathbb{R} \},
\]
with $\text{Var}(A) = 1$ and let $\Delta_i$ be the distance between two consecutive points in $\mathcal{A}$, that is
\[
\Delta_i = a_{i+1} - a_i,
\]
then, if
\[
\Delta_{i+1}^2 \geq (\alpha c^2 - 1)^2 \sum_{n=0}^{i} \Delta_n^2,
\]
for some $\alpha \geq 0$ and for $i \in [1 \ldots M]$, then capacity is to within a 1 bit per channel use from the outer bound
\[
K_{\text{OUT}} = \left\{ \begin{array}{ll}
\frac{1}{M} \log(1+P) + 1 & M^{-1} \leq \frac{c}{M} \\
\frac{1}{2M} \log(1+P) + \frac{M-1}{M} \log(c^2) + 1 + \log \alpha & \frac{c^2}{M} > \frac{M-1}{M} (P+1) \\
\frac{1}{M} \log(1+P) + 1 + \log \alpha & \frac{c^2}{M} \leq \frac{M-1}{M} (P+1)
\end{array} \right.
\]

**Proof:** The proof is provided in App. C.

As an example of Th. IV.3 consider the case in which $\alpha = c^2/(c^2+1)$ and for $a_0 = 0$: in this case the condition in (20) translates to the set $\mathcal{A}(M)$ defined as
\[
\mathcal{A}(M) = \{0, \Delta_0, c\Delta_0, c^2\Delta_0 \ldots M^{-2} c\Delta_0 \},
\]
where $\Delta_0$ is determined so that the variance is equal to one, that is
\[
\frac{\Delta_0^2}{M} \left( 1 - c^{2M-2} \right) = \left( \frac{\Delta_0}{M} \right)^2 \left( 1 - c^{M-1} \right) = 1,
\]
which follows from the properties of the geometric series.

Note that Th. [IV.3] implies that, when $c^2$ is much larger than $P$, then the capacity of the DPC-FFD-RCSI as $1/M$ times the capacity of the channel without state.

We conclude by providing an outer bound for the case of a continuous fading distribution. Unfortunately this bound is not tight in general: this reflect the fact that the outer bounding techniques employed so far are too crude to address this general case.

**Theorem IV.4. Outer Bound for continuous fading distributions.**

Consider the case in which $A$ has a continuous distribution such that there exists an interval $I = [a, b] \subset \mathbb{R}$ with $P_A(I) \geq 1/2$, let moreover

$$a' \in [a, b] \quad \text{s.t.} \quad P(a')(b-a) = P(I)$$

$$G = \int_{\mathbb{R}} \log((a-a')^2) \, dP_a,$$

then the capacity $C$ is upper bounded as

$$C \leq R_{\text{OUT}} = \begin{cases} \frac{4}{3} \log(1+P) + 1 & P_A(I) \leq P_A(I)^2 \\ \frac{P_A(I)}{2} \log(1+P) + \frac{P_A(I)^2}{2} \log(Pc^2) + 1 - G & P_A(I)^2 \leq P_A(I)(P + 1) \\ \frac{P_A(I)}{2} \log(1+P) + 1 - G & P_A(I)(P + 1) > P_A(I)^2 \end{cases}$$

**Proof:** The proof is provided in App. [10].

It is straightforward to verify that the above bound cannot be attained by simply performing Costa pre-coding against a value of $ca'S$: in fact this strategy achieves

$$R_{\text{IN}} = \frac{1}{2} \log(1+P) - \frac{1}{2} E_A \log \left( \frac{Pc^2}{P + c^2a^2} + 1 - (a-a')^2 + 1 \right)$$

$$\approx \frac{1}{2} \log(1+P) - \frac{1}{2} E_A \left[ \log \left( \min \{a^2c^2, P\} \frac{(a-a')^2}{a^2} + 1 \right) \right],$$

which goes to zero as $P$ or $c^2$ grow unless $A$ is mostly concentrated around $a' \pm 1/c^2$.

V. Conclusion

In this paper we studied a variation of the classic dirty paper channel in which the channel state is multiplied by a fast fading process which is unknown at the transmitter. We consider both the case in which the decoder has knowledge of the fading and the case in which it does not. For this model we derive inner and outer bounds to capacity and bound the difference between the two when possible. When fading knowledge is not available at the receiver, the gap between inner and outer bounds is small for a number of classic fading distributions but it is not bounded for others. When fading knowledge is available at the receiver we can characterize capacity for some specific discrete distributions of the fading. The closed-form characterization of the capacity for the case with receiver fading knowledge remains unknown in general.

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] M. Costa, “Writing on dirty paper,” IEEE Trans. Inf. Theory, vol. 29, no. 3, pp. 439–441, 1983.
[3] 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 International Symposium on Information Theory (ISIT), Nice, France, 2010, pp. 381–385.
[4] A. Khina and U. Erez, “On the robustness of dirty paper coding,” Communications, IEEE Transactions on, vol. 58, no. 5, pp. 1437–1446, 2010.
[5] Y. Avner, B. M. Zaidel, S. Shamai, and U. Erez, “On the dirty paper channel with fading dirt,” in Electrical and Electronics Engineers in Israel (IEEEI), 2010 IEEE 26th Convention of. IEEE, 2010, pp. 000 525–000 529.
[6] D. T. Kao and A. Sabharwal, “An upper bound on the capacity of vector dirty paper with unknown spin and stretch,” arXiv preprint arXiv:1305.3934, 2013.
[7] P. Grover and A. Sahai, “On the need for knowledge of the phase in exploiting known primary transmissions,” in New Frontiers in Dynamic Spectrum Access Networks, 2007. DySPAN 2007. 2nd IEEE International Symposium on. IEEE, 2007, pp. 462–471.
[8] S. Rini and S. Shamai, “The impact of phase fading on the dirty paper channel,” arXiv preprint arXiv:1401.4236, 2014.
[9] A. Khisti, U. Erez, A. Lapidoth, and G. Wornell, “Carbon copying onto dirty paper,” IEEE Trans. Inf. Theory, vol. 53, no. 5, pp. 1814–1827, May 2007.
[10] S. Rini and S. Shamai, “On capacity of the dirty paper channel with fading dirt in the strong fading regime,” arXiv preprint arXiv:1405.2221, 2014.
APPENDIX

A. Proof of Th. III.1

Consider the following series of inequalities developed from Fano’s inequality

\[
N(R - \epsilon_N) \leq I(Y^N; W)
\]

\[
\leq I(Y^N; W|S^N)
\]

\[
= h(Y^N|S^N) - h(Y^N|W,S^N,X^N)
\]

\[
\leq N \max_i h(Y_i|S_i) - h(Y^N|W,S^N,X^N).
\]

For the term \(\max_i h(Y_i|S_i)\) we can use the Gaussian Maximizes Entropy (GEM) property and conclude that

\[
\max_i h(Y_i|S_i) \leq \frac{1}{2} \log \left( \frac{P + c^2 + 1}{\alpha} \right).
\] (23)

For the term \(-h(Y^N|W,S^N,X^N)\) we have:

\[
-h(Y^N|W,S^N,X^N)
\]

\[
= -h(cS^N A^N + Z^N|W,S^N,X^N)
\]

\[
= -Nh(cSA + Z|S)
\] (24a)

\[
= N(-h(cSA + Z|S) + h(Z|cSA + Z,S))
\] (24c)

\[
= N(-h(cSA|S) - h(Z) + h(Z|cSA + Z,S))
\] (24d)

\[
\leq -Nh(SA|S) - N\log |c| + 1.
\] (24e)

The term \(-h(S,A|S)\) can be rewritten as

\[
-h(S,A|S) = -h(A) + \mathbb{E}_S \left[ \frac{1}{2} \log(s^2) \right] = -h(A) + \gamma,
\]

where \(\gamma\) is the Euler’s constant \(\gamma \approx 0.577\). Combining the bounds in (23) and (24) we obtain the expression in (11).

For the inner bound, we consider Costa’s dirty paper coding strategy to pre-cancel \(\mu_A S\) while disregarding the remaining randomness in the fading. This strategy attains

\[
R^{\text{IN}} = I(Y;U|A) - I(U;S) = h(U|S) - h(U|Y,A).
\]

Considering now the assignment in which \(X\) and \(U\)

\[
X \sim \mathcal{N}(0,P), \quad U = X + \frac{P}{P + 1 + c^2} \mu_A S,
\]

which achieves

\[
R^{\text{IN}} \geq \frac{1}{2} \log \left( 1 + \frac{P}{c^2 + 1} \right).
\] (25)

The difference between outer and inner bound is then

\[
G = \frac{1}{2} \log \left( \frac{P + 1 + c^2}{\alpha(P + 1 + c^2 \alpha)} \right) + \frac{1}{2} \leq \frac{1}{2} \log \left( \frac{1}{\alpha} \right) + \frac{1}{2},
\]

which concludes the proof.
B. Proof of Th. 4.1

Using Fano’s inequality we write

\[ N(R - \varepsilon) \leq I(Y^N; W|A^N) \]  
\[ \leq \frac{N}{2} \mathbb{E} \left[ \log 2\pi e(P + c^2 + 1) \right] - h(Y^N|A^N, W) \]  
\[ \leq \frac{N}{2} \mathbb{E} \left[ \log 2\pi e(P + c^2 + 1) \right] - \sum_{i=1}^{N} h(y_i|Y^{i-1}, A^N, W) \]  
\[ \leq \frac{N}{2} \mathbb{E} \left[ \log 2\pi e(P + c^2 + 1) \right] - \sum_{i=1}^{N} h(y_i|A_i, Y^{i-1}, A^{i-1}, W) \]  
\[ \leq \frac{N}{2} \mathbb{E} \left[ \log 2\pi e(P + c^2 + 1) \right] - \sum_{i=1}^{N} h(y_i|A_i, U_i), \]  

where \( U_i = [Y^{i-1}, A^{i-1}, W] \) has been defined only to reduce the notation. For the negative entropy term let’s drop the \( i \) index and write

\[ -h(Y_i|A_i, U_i) = h(Y_i|A, U) \]  
\[ = -\sum_{a'} P_{a'} h(Y|U, A = a) \]  
\[ = -\sum_{a'} P_{a'} h(X + caS + Z|U) \]  
\[ = -P_{a'} h(X + caS + Z|U) - \sum_{\substack{d \neq a'}} P_{a'} h(X + caS + Z_d|U) \]  
\[ = -(P_{a'} - (1 - P_{a'})) h(X + caS + Z|U) - \sum_{\substack{d \neq a'}} P_{a'} (h(X + caS + Z_d|U) + h(X + caS + Z_d'|U)) \]  
\[ \leq -(2P_{a'} - 1) h(X + caS + Z|U, X, S) - \sum_{\substack{d \neq a'}} P_{a'} (h(X + caS + Z_d + X + ca'd + Z_d'|U)) \]  
\[ = -(2P_{a'} - 1) h(Z') - \sum_{\substack{d \neq a'}} P_{a'} (h(c(a - a')S + Z_d - Z', X + ca'S + Z_d'|U)) \]  
\[ = -(2P_{a'} - 1) \log(2\pi e) \]  
\[ = -\sum_{\substack{d \neq a'}} P_{a'} (h(c(a - a')S + Z_d - Z'|U) + h(X + ca'S + Z'|c(a - a')S + Z_d - Z', U)) \]  
\[ \leq -(2P_{a'} - 1) \log(2\pi e) \]  
\[ = -\sum_{\substack{d \neq a'}} P_{a'} (h(c(a - a')S + Z_d - Z'|U) - h(X + ca'S + Z'|c(a - a')S + Z_d - Z', U, X)) \]  
\[ = -(2P_{a'} - 1) \log(2\pi e) \]  
\[ = \sum_{\substack{d \neq a'}} P_{a'} (h(c(a - a')S + Z_d - Z') + h(Z'|Z_d - Z')) \]  
\[ = -(2P_{a'} - 1) \log(2\pi e) \]  
\[ = -\sum_{\substack{d \neq a'}} P_{a'} \left( \frac{1}{2} \log(2\pi e(c^2(a - a')^2 + 2)) + \frac{1}{2} \log(2\pi e) + \frac{1}{4} \right), \]  

Now we have

\[ - \sum_{\substack{d \neq a'}} P_{a'} \frac{1}{2} \log(2\pi e) = -\frac{P_{a'}}{2} \log(2\pi e), \]
and similarly we have

\[- \sum_{\forall \alpha \neq \alpha'} P_{\alpha} \frac{1}{2} \log(2\pi e(c^2(a - a')^2 + 2)) \]

\[\leq - \sum_{\forall \alpha \neq \alpha'} P_{\alpha} \frac{1}{2} \log(2\pi e(c^2(a - a')^2)) \]

\[= - \frac{P_{\alpha}}{2} \log c^2 - \sum_{\forall \alpha \neq \alpha'} P_{\alpha} \frac{1}{2} \log(2\pi e((a - a')^2)) \]

Using the previous two bounds we obtain the outer bound expression

\[R^{\text{OUT}} = \frac{1}{2} \log \left(2\pi e(P + c^2 + 1)\right) - \frac{P_{\alpha}}{2} \log c^2 - \frac{P_{\alpha}}{2} \log(2\pi e) - G + \frac{1}{4} \]

We can optimize the above expression over the parameter $c^2$: this minimization corresponds to the outer bound obtained by providing some of the side information $S$ also to the decoder. This is obtained by considering the state to be $cS$ and provide a fraction $\alpha cS$ so that the remaining level of interference is $c^2 S$.

The optimal value of $c^2$ in (30) is

\[(c^2)^* = \min \left\{ \frac{P_{\alpha}}{P_{\alpha}'(1 + P)}, c^2 \right\} \]

When $P_{\alpha}'c^2 \geq P_{\alpha}'(1 + P)$ this optimization yield the tighter outer bound than that of (30)

\[R^{\text{OUT}} = \frac{1}{2} \log \left(1 + P \left(1 + \frac{P_{\alpha}}{P}\right)\right) - \frac{P_{\alpha}}{2} \log \left(\frac{P_{\alpha} 1 + P}{P_a a^2}\right) - G \]

so that the outer bound can be further simplified as

\[R^{\text{OUT}} = \left\{ \begin{array}{ll}
\frac{P_{\alpha}}{2} \log(1 + P) + \frac{P_{\alpha}}{2} \log(Pc^2) + 1 - G & P_{\alpha}'c^2 \leq P_{\alpha}'(P + 1) \\
\frac{P_{\alpha}}{2} \log(1 + P) + 1 - G & P_{\alpha}'c^2 > P_{\alpha}'(P + 1)
\end{array} \right. \]

For the inner bound consider the simple scenario in which the transmitter pre-codes against the realization $a'$, which occurs more than half of the time, as in the Costa problem. That is, consider the assignment

\[X \sim \mathcal{N}(0, P) \]

\[U = X + \frac{P}{P + 1} d'eS \]

this assignment attains full interference pre-cancellation for a portion $P_{\alpha}'$ of the time since

\[R^{\text{IN}} = \mathbb{E}_{A} [I(Y; U|A) - I(U; S)]^+ \]

\[= \frac{P_{\alpha}'}{2} \log(1 + P) + \sum_{\forall \alpha \neq \alpha'} P_{\alpha} \frac{1}{2} \log \left(\frac{(1 + c^2 a^2 + P)(1 + P)}{P_{\alpha}'c^2(a - a')^2 + P + c^2 a^2 + 1}\right) \]
the latter term is bounded as

\[
\sum_{\alpha, \alpha' \neq \alpha} \frac{P_a}{2} \log \left( \frac{(1 + c^2a^2 + P)(1 + P)}{Pc^2(a - \alpha)^2 + P + c^2a^2 + 1} \right)
\]

(37a)

\[
= \sum_{\alpha, \alpha' \neq \alpha} \frac{P_a}{2} \log (1 + P) - \frac{P_a}{2} \log \left( \frac{Pc^2}{P + c^2a^2 + 1} (a - \alpha')^2 + 1 \right)
\]

(37b)

\[
= \sum_{\alpha, \alpha' \neq \alpha} \frac{P_a}{2} \log (1 + P) - \frac{P_a}{2} \log \left( \frac{Pac^2}{P + c^2a^2 + 1} (a - \alpha')^2 + 1 \right)
\]

(37c)

\[
\geq \sum_{\alpha, \alpha' \neq \alpha} \frac{P_a}{2} \log (P) - \frac{P_a}{2} \log \left( \min \left\{ \frac{Pac^2}{2}, \frac{(a - \alpha')^2}{a^2} + 1 \right\} \right)
\]

(37d)

\[
\geq \sum_{\alpha, \alpha' \neq \alpha} \frac{P_a}{2} \log \left( \min \left\{ 1, \frac{a^2c^2}{P} \right\} \right) - \frac{1}{2} \log \left( \frac{a - \alpha'}{a} + 1 \right)
\]

(37e)

\[
\geq \sum_{\alpha, \alpha' \neq \alpha} \frac{P_a}{2} \log \left( \min \left\{ 1, \frac{a^2c^2}{P} \right\} \right) - \frac{1}{2} \log \left( \frac{a - \alpha'}{a} + 1 \right)
\]

(37f)

\[
\geq \sum_{\alpha, \alpha' \neq \alpha} \frac{P_a}{2} \log \left( \frac{(a - \alpha')^2}{a^2} + 1 \right) = G'
\]  

(37g)

This attainable rate can be improved upon by using two codewords: one that treats the interference as noise. We can assign power \( \alpha \) to one codeword and power \( \bar{\alpha} = 1 - \alpha \) to the other and successively optimize over the power assigned to each codeword. This yield the achievable rate

\[
R^N = \max_{\alpha \in [0,1]} E_A \left[ \frac{1}{2} \log \left( 1 + \frac{\alpha P}{1 + c^2a^2 + \alpha P} \right) + \frac{P}{2} \log (1 + \alpha P) \right] - \frac{G'}{2}
\]

(38)

\[
\geq \max_{\alpha \in [0,1]} \frac{1}{2} \log \left( 1 + \frac{\alpha P}{1 + c^2a^2 + \alpha P} \right) + \frac{P}{2} \log (1 + \alpha P) - \frac{G'}{2}
\]

(39)

the optimal value of \( \bar{\alpha}P \) is then

\[
\bar{\alpha}P = \max \left\{ \min \left\{ \frac{A}{P^A}, 2 \right\}, 0 \right\}
\]

(40)

so that the optimized outer bound is

\[
\begin{cases}
\frac{1}{2} \log \left( 1 + \frac{P}{1 + c} \right) & \bar{P}^A \leq P_A c^2 \\
\frac{P}{2} \log (1 + P) + \frac{P}{2} \log (Pc^2) - \frac{P}{2} - 2 & \bar{P}^A < P_A c^2 \leq \bar{P}^A (P + 1) \\
\frac{P}{2} \log (1 + P) - \frac{P}{2} - 2 & P_A c^2 > \bar{P}^A (P + 1)
\end{cases}
\]

(41)

A gap between inner and outer bound of 3 bits in the interval \( P_A c^2 > \bar{P}^A \) can be obtained by simply comparing the two expressions. For the case in which \( P_A c^2 \leq \bar{P}^A \) we have that \( c^2 \leq 1 \) so that the capacity can be approached to within 1 bit by treating the interference as noise with a variance partially known at the receiver.

C. Proof of Th. [IV.3]

The proof proceeds by induction: we first consider the case where \( M = 2 \) then \( M = 3 \) and then proceed to the general case by induction from these first two terms.

Consider the bounding of \( 28 \) for the first the case \( M = 2 \) so that only two terms appear in expectation: \( A = a_1 \) and \( A = a_2 \)

\[
\frac{1}{M} (-h(X + ca_1S + Z|W, U) - h(X + ca_2S + Z|W, U))
\]

(42a)

\[
\leq -\frac{1}{M} h(X + ca_1S + Z, X + ca_2S + Z|W, U)
\]

(42b)

\[
\leq -\frac{1}{M} h(c(a_2 - a_1)S + Z_1 - Z_2, X + ca_1S + Z|W, U)
\]

(42c)

\[
\leq -\frac{1}{M} (h(c(a_2 - a_1)S + Z_1 - Z_2) + h(X + ca_2S + Z|W, U, c(a_2 - a_1)S + Z))
\]

(42d)

This bounding is sufficient to obtain the approximate capacity for the case of \( M = 2 \) since it gives rise to the term \( 1/2 \cdot 1/2 \log(c^2) \). We \( M = 3 \) we wish to obtain the term \( 2/3 \cdot 1/2 \log(c^2) \), to do so we want to combine the term \( h(X + ca_2S + Z|W, U, c(a_2 - a_1)S + Z) \) and \( h(X + ca_2S + Z|W, U, c(a_2 - a_1)S + Z) \) and
with the term \( h(X + ca_3S + Z | W, U) \). We do so as

\[
- \frac{1}{M} \left[ h(X + ca_3S + Z | W, U) + h(X + ca_1S + Z_1 | W, U, c(a_2 - a_1)S + Z_2 - Z_1) \right] 
\]

\[
\leq \frac{1}{M} \left[ h(X + ca_3S + Z_3, X + ca_2S + Z_2 | W, U, c(a_2 - a_1)S + Z_2 - Z_1) \right] 
\]

\[
= \frac{1}{M} \left[ h(c(a_3 - a_2)S + Z_3 - Z_2, X + ca_2S + Z_2 | W, U, c(a_2 - a_1)S + Z_2 - Z_1) \right] 
\]

\[
= \frac{1}{M} h(c(a_3 - a_2)S + Z_3 - Z_2 | c(a_2 - a_1)S + Z_2 - Z_1) 
\]

\[
+ \frac{1}{M} h(X + ca_2S + Z_2 | W, U, c(a_2 - a_1)S + Z_2 - Z_1, c(a_3 - a_2)S + Z_3 - Z_2) 
\]

\[
\text{When } M = 3 \text{ we wish for the term } h(c(a_3 - a_2)S + Z_3, X + ca_2S + Z_2 | c(a_2 - a_1)S + Z_2 - Z_1) \text{ to scale as } 1/2 \log(c^2), \text{ for this to happen we need to determine the appropriate value of } a_i.
\]

We choose \( Z_3 = -Z_1 \) so that the noises become independent terms with variance 2, for convenience we also define

\[
\tilde{Z}_i = Z_{i+1} - Z_i 
\]

\[
\tilde{Z}_i \sim \mathcal{N}(0, 2), \text{ i.i.d.} 
\]

With this notation we write

\[
h(c(a_3 - a_2)S + Z_3 - Z_2 | c(a_2 - a_1)S + Z_2 - Z_1) 
\]

\[
= h \left( c^2(a_3 - a_2)(a_2 - a_1) \left( 1 - \frac{c(a_2 - a_1)}{c^2(a_2 - a_1)^2 + 1} \right) + (Z_3 - Z_2) - \frac{c^2(a_3 - a_2)(a_2 - a_1)}{c^2(a_2 - a_1)^2 + 1} (Z_2 - Z_1) \right) 
\]

Using the definition of \( \Delta_i \) above we can express the above entropy as

\[
h(c(a_3 - a_2)S + Z_3 - Z_2 | c(a_2 - a_1)S + Z_2 - Z_1) 
\]

\[
h(\tilde{Z}_2 | c\tilde{Z}_2S + \tilde{Z}_1) 
\]

\[
= \frac{1}{2} \log \left( c^2\Delta_2^2 + 2 - \frac{c^4\Delta_2^2 \Delta_1^2}{c^2\Delta_1^2} \right) 
\]

\[
= \frac{1}{2} \log \left( c^2\Delta_2^2 + 2 - \frac{c^4\Delta_2^2 \Delta_1^2}{c^2\Delta_1^2} \right) 
\]

\[
= \frac{1}{2} \log \left( \frac{2c^2(\Delta_2^2 + \Delta_1^2) + 1}{c^2\Delta_2^2 + 2} \right) 
\]

\[
\leq \frac{1}{2} \log \left( \frac{2c^2(\Delta_2^2 + \Delta_1^2) + 1}{c^2\Delta_2^2 + 2} \right) 
\]

so the term scales as \( 1/2 \log(c^2) \) when

\[
\Delta_2^2 + \Delta_1^2 \geq \alpha c^2\Delta_1^2 
\]

\[
\Delta_2^2 \geq (\alpha c^2 - 1)\Delta_1^2 
\]

for some \( \alpha > 0 \). In order to repeat this bounding for all the elements in \( \mathcal{A} \) we can simply consider the case for \( M = 4 \).

In this case we rewrite the conditional entropy term containing the channel input \( X \) as

\[
h(X + ca_2S + Z_2 | W, U, c(a_2 - a_1)S + Z_2 - Z_1, c(a_3 - a_2)S + Z_3 - Z_2) 
\]

\[
= h(X + ca_2S + Z_2 | W, U, c\Delta_1S + \tilde{Z}_1, c\Delta_2S + \tilde{Z}_2) 
\]

This term can be bounded together with the entropy term \( H(Y | A = a_1) \) as

\[
h(\Delta_3S + \tilde{Z}_3 | W, U, c\Delta_1S + \tilde{Z}_1, c\Delta_2S + \tilde{Z}_2) 
\]

\[
= \frac{1}{2} \log \left( \frac{2c^2(\Delta_3^2 + \Delta_1^2 + \Delta_2^2) + 2}{c^2(\Delta_3^2 + \Delta_1^2 + \Delta_2^2) + 2} \right) 
\]

which also follows from the maximal ratio combining principle. In this case the condition on \( \Delta_3 \) is now

\[
\Delta_3 \geq (\alpha c^2 - 1)(\Delta_1 + \Delta_2) 
\]
More in general we have that a scaling of $1/2\log(c^2)$ of the negative entropy term is possible when

$$\Delta_{i+1} \geq (a c^2 - 1) \sum_{n=1}^{i} \Delta_{i+1}$$

(50)

With the above recursion we come to the outer bound

$$R^{\text{OUT}} = \frac{1}{2} \log(1 + c^2 + P) - \frac{M - 1}{2M} \log(c^2)$$

(51)

This expression correspond to the expression in (30) in the proof of Th. IV.1 consequently it can be optimized over $c$ as such said expression.

The inner bound to match this outer bound is also obtained in a similar fashion as the inner bound in (36) for $P'_A = 1/M$. Note that this optimization does not depend on the value of $P'_A$ and thus carries over naturally to the case when $P'_A < 1/2$.

D. Proof of Th. IV.4

We proceed in the bounding from Fano’s inequality up to (28) in App. A. For the negative entropy term let’s drop the $i$ index and write

$$h(Y_i|A_i, U_i) = h(Y|A, U)$$

(52a)

$$= \int_R p_a h(X + caS + Z|U) \, da$$

(52b)

$$= \int_R p_a h(X + caS + Z|U) \, da + \int_{\mathbb{R} \setminus I} p_a h(X + caS + Z|U) \, da$$

(52c)

Since

$$\frac{1}{2} \log(2\pi e) \leq h(X + caS + Z|U) \leq \frac{1}{2} \log(P + c^2 a^2 + 2Pca + 1)$$

(53)

and $I$ is a closed interval, we can apply the mean value theorem and conclude that

$$\int_I p_a h(X + caS + Z|U, A = a) \, da = P_A(I) h(X + caS + Z|U) \text{ \quad \quad (54)}$$

Note that this holds even if the distribution $P_{X,S}$ has some discrete points because of the convolution with the distribution of $Z$. Now we can write

$$\int_R p_a h(X + caS + Z|U) \, da + \int_{\mathbb{R} \setminus I} p_a h(X + caS + Z|U) \, da$$

$$= P_A(I) - (1 - P_A(I)) h(X + ca'S + Z|U) + \int_{\mathbb{R} \setminus I} p_a \left( h(X + caS + Z|U) + h(X + ca'S + Z|U) \right) \, da$$

$$\geq P_A(I) - (1 - P_A(I)) h(X + ca'S + Z|U, X, S) + \int_{\mathbb{R} \setminus I} p_a \left( h(X + caS + Z, X, ca'S + Z|U) \right) \, da$$

$$\geq \frac{P_A(I) - (1 - P_A(I))}{2} \log(2\pi e) + \int_{\mathbb{R} \setminus I} p_a \left( h(c(a - a')S + Z, X, ca'S + Z|U) \right) \, da$$

$$\geq \frac{P_A(I) - (1 - P_A(I))}{2} \log(2\pi e) + \int_{\mathbb{R} \setminus I} p_a \left( \frac{1}{2} \log 2\pi e \left( c^2 (a - a')^2 + 2 \right) + h(Z|U, c(a - a')S + Z, S, X) \right)$$

$$\geq \frac{P_A(I)}{2} \log(2\pi e) + \frac{P_A(\mathbb{R} \setminus I)}{2} \log 2\pi ec^2 + \int_{\mathbb{R} \setminus I} \frac{P_a}{2} \log \left( (a - a')^2 \right)$$

$$\geq \frac{P_A(I)}{2} \log(2\pi e) + \frac{P_A(\mathbb{R} \setminus I)}{2} \log 2\pi ec^2 + G$$

(55h)

This yields the same outer bound as (30) but with an updated expression for $G$. As for Thm. IV.1 we can optimize the expression in $c$ and obtain the same outer bound.