

# A Coordinate System for Gaussian Networks

Emmanuel A. Abbe  
LCM, EPFL  
Lausanne, 1015 CH  
Email: emmanuel.abbe@epfl.ch

Lizhong Zheng  
RLE, MIT  
Cambridge, MA 02139  
Email: lizhong@mit.edu

**Abstract**—This paper studies network information theory problems where the external noise is Gaussian distributed. In particular, the Gaussian broadcast channel with coherent fading and the Gaussian interference channel are investigated. It is shown that in these problems, non-Gaussian code ensembles can achieve higher rates than the Gaussian ones. It is also shown that the strong Shamai-Laroia conjecture on the Gaussian ISI channel does not hold. In order to analyze non-Gaussian code ensembles over Gaussian networks, a geometrical tool using the Hermite polynomials is proposed. This tool provides a coordinate system to analyze a class of non-Gaussian input distributions that are invariant over Gaussian networks.

## I. INTRODUCTION

Let a memoryless additive white Gaussian noise (AWGN) channel be described by  $Y = X + Z$ , where  $Z \sim \mathcal{N}(0, v)$  is independent of  $X$ . If the input is imposed an average power constraint given by  $\mathbb{E}X^2 \leq p$ , the input distribution maximizing the mutual information is Gaussian. This is due to the fact that under second moment constraint, the Gaussian distribution maximizes the entropy, hence

$$\arg \max_{X: \mathbb{E}X^2=p} h(X + Z) \sim \mathcal{N}(0, p). \quad (1)$$

On the other hand, if we use a Gaussian input distribution, i.e.,  $X \sim \mathcal{N}(0, p)$ , the worst noise that can occur, i.e., the noise minimizing the mutual information, among noises with bounded second moment, is again Gaussian distributed. This can be shown using the *entropy power inequality* (EPI), cf. [17], which reduces in this setting to

$$\arg \min_{Z: h(Z)=\frac{1}{2} \log 2\pi ev} h(X + Z) \sim \mathcal{N}(0, v) \quad (2)$$

and implies

$$\arg \min_{Z: \mathbb{E}Z^2=v} h(X + Z) - h(Z) \sim \mathcal{N}(0, v). \quad (3)$$

Hence, in the single-user setting, when optimizing the mutual information as above, a Gaussian input is the best input for a Gaussian noise and a Gaussian noise is the worst noise for a Gaussian input. This provides a game equilibrium between user and nature, as defined in [11], p. 263. With these results, many problems in information theory dealing with Gaussian noise can be solved. However, in Gaussian networks, that is, in multi-user information theory problems where the external noise is Gaussian distributed, several new phenomena make the search for the optimal input ensemble more complex. Besides for some specific cases of Gaussian

networks, we still do not know how interference should be treated in general. Let us consider two users interfering on each other in addition to suffering from Gaussian external noise and say that the receivers treat interference as noise. Then, if the first user has drawn its code from a Gaussian ensemble, the second user faces a frustration phenomenon: using a Gaussian ensemble maximizes its mutual information but minimizes the mutual information of the first user. It is an open problem to find the optimal input distributions for this problem, as well as identifying the regime for which treating interference as noise is optimal. This is one illustration of the complications appearing in the network setting. Another example is regarding the treatment of the fading. Over a single-user AWGN channel, as long as the fading statistics are known at the receiver, whether the fading is deterministic or random does not affect the optimal input distribution. Indeed, from (1), it is clear that maximizing  $I(X; X + Z)$  or  $I(X; HX + Z|H)$  under an average power constraint is achieved by a Gaussian input. However, the situation is different if we consider a Gaussian broadcast channel (BC). When there is a deterministic fading, using (1) and (3), the optimal input distribution can be shown to be Gaussian. However, it has been an open problem to show whether Gaussian inputs are optimal or not for a Gaussian BC with a random fading known at the receiver, even if the fading is such that it is a degraded BC.

A reason for these open questions in the network information theoretic framework, is that Gaussian ensembles are roughly the only ensembles which we have known how to analyze over Gaussian networks, as non-Gaussian ensembles have left most problems in an intractable form. In this paper, a novel technique is developed to analyze a class of non-Gaussian input distributions over Gaussian noise channels. This technique is efficient to analyze the *competitive situations* occurring in the network problems described below. It allows in particular to find certain non-Gaussian ensembles that outperform Gaussian ones on a Gaussian BC with coherent fading channel, a two user interference channel, and it allows to disprove the strong Shamai-Laroia conjecture on the Gaussian intersymbol interference channel. This tool provides a new insight on Gaussian networks and confirms that non-Gaussian ensembles do have a role to play in these networks. We now introduce with more details different problems that are exhibiting competitive situations.

### A. Competitive Situations

1) *Fading Broadcast Channel*: Consider a degraded Gaussian BC with coherent memoryless fading, where the fading is indeed the same for both receivers, i.e.

$$\begin{aligned} Y_1 &= HX + Z_1, \\ Y_2 &= HX + Z_2 \end{aligned}$$

but  $Z_1 \sim \mathcal{N}(0, v_1)$  and  $Z_2 \sim \mathcal{N}(0, v_2)$ , with  $v_1 < v_2$ . The input  $X$  is imposed a power constraint denoted by  $p$ . Because the fading is coherent, each receiver also knows the realization of  $H$ , at each channel use. The fading and the noises are memoryless (iid) processes. Since this is a degraded broadcast channel, the capacity region is given by all rate pairs

$$(I(X; Y_1|U, H), I(U; Y_2|H))$$

with  $U - X - (Y_1, Y_2)$ . The optimal input distributions, i.e., the distributions of  $(U, X)$  achieving the capacity region boundary, are given by the following optimization problem, where  $\mu \in \mathbb{R}$ ,

$$\arg \max_{(U, X): \substack{U - X - (Y_1, Y_2) \\ \mathbb{E}X^2 \leq p}} I(X; Y_1|U, H) + \mu I(U; Y_2|H). \quad (4)$$

Note that the objective function in the above maximization is given by

$$h(Y_1|U, H) - h(Z_1) + \mu h(Y_2|H) - \mu h(Y_2|U, H).$$

Now, each term in this expression is individually maximized by a Gaussian distribution for  $U$  and  $X$ , but these terms are combined with different signs, and there is a competitive situation for which the maximizer is not identified using previously mentioned inequalities. When  $\mu \leq 1$ , one can show that Gaussian distributions are optimal. Also, if  $H$  is compactly supported, and if  $v$  is small enough as to make the support of  $H$  and  $1/vH$  non overlapping, the optimal distribution of  $(U, X)$  is jointly Gaussian (cf. [19]). However, in general the optimal distribution is unknown (theorems may be missing to conclude the optimality of Gaussian distributions or non-Gaussian codes may actually perform better than Gaussian ones).

2) *Interference Channel*: We consider the symmetric memoryless interference channel (IC) with two users and white Gaussian noise. The average power is denoted by  $p$ , the interference coefficients by  $a$ , and the respective noise by  $Z_1$  and  $Z_2$  (independent standard Gaussian). We define the following expression

$$\begin{aligned} mS_{a,p}(X_1^m, X_2^m) & \quad (5) \\ &= I(X_1^m; X_1^m + aX_2^m + Z_1^m) + I(X_2^m; X_2^m + aX_1^m + Z_2^m) \\ &= h(X_1^m + aX_2^m + Z_1^m) - h(aX_2^m + Z_1^m) \\ & \quad + h(X_2^m + aX_1^m + Z_2^m) - h(aX_1^m + Z_2^m), \end{aligned}$$

where  $X_1^m$  and  $X_2^m$  are independent random vectors of dimension  $m$  whose covariance's traces are bounded by  $mp$ , and  $Z_i^m$ ,  $i = 1, 2$ , are iid standard Gaussian. For any dimension  $m$  and any distributions of  $X_1^m$  and  $X_2^m$ ,  $S_{a,p}(X_1^m, X_2^m)$  is a lower bound to the sum-capacity. Moreover, it is tight by

taking  $m$  arbitrarily large and  $X_1^m$  and  $X_2^m$  maximizing (5). Now, a similar competitive situation as for the fading broadcast problem takes place: Gaussian distributions maximize each entropy term, but these terms are combined with different signs. Would we then prefer to take  $X_1$  and  $X_2$  Gaussian or not? This should depend on the value of  $a$ . If  $a = 0$ , we have two parallel AWGN channels with no interference, and Gaussian inputs are optimal. We can then expect that this might still hold for small values of  $a$ . It has been proved recently in [3], [13], [16], that the sum-capacity is achieved by treating interference as noise and with iid Gaussian inputs, as long as  $pa^3 + a - 1/2 \leq 0$ . Hence, in this regime, the iid Gaussian distribution maximizes (5) for any  $m$ . But if  $a$  is above that threshold and below 1, the problem is open.

Let us now review the notion of “treating interference as noise”. For each user, we say that the decoder is *treating interference as noise*, if it does not require the knowledge of the other user’s code book. However, we allow such decoders to have the knowledge of the distribution, under which the other user’s code book may be drawn. This is for example necessary to construct a sum-capacity achieving code<sup>1</sup> in [3], [13], [16], where the decoder of each user treats interference as noise but uses the fact that the other user’s code book is drawn from an iid Gaussian distribution. But, if we allow this distribution to be of arbitrarily large dimension  $m$  in our definition of treating interference as noise, we can get a misleading definition. Indeed, no matter what  $a$  is, if we take  $m$  large enough and a distribution of  $X_1^m, X_2^m$  maximizing (5), we can achieve rates arbitrarily close to the sum-capacity, yet, formally treating interference as noise. The problem is that the maximizing distributions in (5) may not be iid for an arbitrary  $a$ , and knowing it at the receiver can be as much information as knowing the other user’s code book (for example, if the distribution is the uniform distribution over a code book of small error probability). Hence, one has to be careful when taking  $m$  large. In this paper, we will only work with situations that are not ambiguous with respect to our definition of treating interference as noise. It is indeed an interesting problem to discuss what kind of  $m$ -dimensional distributions would capture the meaning of treating interference as noise that we want. This also points out that studying the maximizers of (5) relates to studying the concept of treating interference as noise or information. Since for any chosen distributions of the inputs we can achieve (5), the maximizers of (5) must have a different structure when  $a$  grows. For  $a$  small enough, iid Gaussian are maximizing distributions, but for  $a \geq 1$ , since we do not want to treat interference as noise, the maximizing distributions must have a “more complex structure”. This also underlines that an encoder can be drawn from a distribution which does not maximize (5) for any value of  $m$ , but yet, a decoder may exist in order to have a capacity achieving code. This happens if  $a \geq 1$ , iid Gaussian inputs will achieve the sum-capacity if the receiver decodes the message of both users (one can show that the problem is equivalent to having two

<sup>1</sup>in a low interference regime

MAC's). However, if  $a \geq 1$ , the iid Gaussian distribution does not maximize  $S_{a,p}(X_1, X_2)$  (for the dimension 1, hence for arbitrary dimensions).

In any cases, if the Gaussian distribution does not maximize (5) for the dimension 1, it means that iid Gaussian inputs and treating interference as noise is not capacity achieving, since a code which treats interference as noise and whose encoder is drawn from a distribution can be capacity achieving only if the encoder is drawn from a distribution maximizing (5). Hence, understanding better how to resolve the competitive situation of optimizing (5) is a consequent problem for the interference channel.

### B. ISI channel and strong Shamai-Laroia Conjecture

In the Gaussian intersymbol interference (ISI) channel, when a symbol  $x_i$  is transmitted at time  $i$ , the receiver observes a corrupted version of  $x_i$ , which contains the addition of Gaussian noise and the interference from other symbols at different times than  $i$ . We refer to [4], [12], [15], [14] and references therein for more details on the ISI channel. In view of analyzing this channel, the authors in [14] investigate the following inequality

$$I(X; X + \sum_{i=1}^{\infty} h_i X_i^G + Z) \stackrel{?}{\leq} I(X; X + \sum_{i=1}^{\infty} h_i X_i + Z),$$

where the  $X, X_i$ 's are i.i.d. with 0 mean and variance  $p$ , the  $X_i^G$ 's are i.i.d. Gaussian with 0 mean and variance  $p$  and  $Z$  is standard Gaussian. In [14], the inequality is conjectured when the  $h_i$ 's take a specific form, which corresponds to the coefficients arising when using an MMSE decision feedback equalizer on a Gaussian noise ISI channel. We also refer to [5], [6] for works related to this conjecture. Shlomo Shamai (Shitz), who referred to this conjecture as the weak Shamai-Laroia conjecture, brought to our attention that the inequality was conjectured even for a general choice of the interference coefficient  $h_i$ 's, as long as  $\sum_i |h_i|^2 < \infty$ , which he referred to as the strong Shamai-Laroia conjecture. In particular, by looking at the memory one case, it reduces to the following conjecture.

*Conjecture 1:* Let  $h, p, v \in \mathbb{R}_+$ ,  $X_1^G \sim \mathcal{N}(0, p)$  and  $Z \sim \mathcal{N}(0, v)$  (independent of  $X_1^G$ ). For all  $X, X_1$  i.i.d. with mean 0 and variance  $p$ , we have

$$I(X; X + hX_1^G + Z) \leq I(X; X + hX_1 + Z). \quad (6)$$

We show in this paper that this strong conjecture does not hold. In particular, we show how the technique developed in this paper can be used to find counter-examples to the strong conjecture and how it allows to find scalings between  $h$  and SNR for which the conjecture may still hold. This approach can then be used for tackling the weak Shamai-Laroia conjecture (work in progress).

There are many other examples in network information theory where such competitive situations occur. Our goal in this paper is to explore the degree of freedom provided by non-Gaussian input distributions. We show that the neighborhood

of Gaussian distributions can be parametrized in a specific way, as to simplify greatly the computations arising in competitive situations. We will be able to precisely quantify how much a certain amount of non-Gaussianness, which we will characterize by means of the Hermite polynomials, affects or helps us in maximizing the competitive entropic functional of previously mentioned problems.

## II. PROBLEM STATEMENT

### A. Fading BC

For the fading BC problem described in I-A1, we want to determine if/when the distribution of  $(U, X)$  maximizing (4) is Gaussian or not.

### B. IC

For the interference channel problem described in I-A2, we know from [3], [13], [16] that treating interference as noise and using iid Gaussian inputs is optimal when  $pa^3 + a - 1/2 > 0$ . We question when this coding scheme is no longer optimal. More generally, we want to analyze the maximizers of (5).

We distinguish the implication of such a threshold in both the synchronized and asynchronous users setting, as there will be an interesting distinction between these two cases. We recall how the synch and asynch settings are defined here. In the synch setting, each user of the IC sends their code words of a common block length  $n$  simultaneously, i.e., at time 1, they both send the first component of their code word, at time 2 the second component, etc. In the asynch setting, each user is still using code words of the same block length  $n$ , however, there might be a shift between the time at which the first and second users start sending their code words. We denote this shift by  $\tau$ , and assume w.l.o.g. that  $0 \leq \tau \leq n$ . In the totally asynch setting, we assume that  $\tau$  is drawn uniformly at random within  $\{0, \dots, n\}$ . We may also distinguish the cases where  $\tau$  is not known at the transmitter but at the receiver, and when  $\tau$  is not known at both. Note that if iid input distributions are used to draw the code books, and interference is treated as noise, whether the users are synch or asynch is not affecting the rate achievability<sup>2</sup>. However, if the users want to time-share over the channel uses, such as to fully avoid their interference, they will need synchronization.

*Definition 1:* Time sharing over a block length  $n$  (assumed to be even) with Gaussian inputs refers to using  $X_1$  Gaussian with covariance  $2P\hat{I}_{n/2}$  and  $X_2$  Gaussian with covariance  $2P\hat{I}_{n/2}^c$ , where  $\hat{I}_{n/2}$  is a diagonal matrix with  $n/2$  1's and 0's, and  $\hat{I}_{n/2}^c$  flips the 1's and 0's on the diagonal.

### C. ISI Channel and Strong Shamai-Laroia Conjecture

We want to determine whether conjecture 1 holds or not.

<sup>2</sup>hence, (5) with an iid distribution for  $X_1$  and  $X_2$  can still be defined for the totally asynch IC

#### D. General Problem

Our more general goal is to understand better the problem posed by any competitive situations. For this purpose, we formulate the following mathematical problem.

We start by changing the notation and rewrite (1) and (3) as

$$\arg \max_{f: m_2(f)=p} h(f \star g_v) = g_p \quad (7)$$

$$\arg \min_{f: m_2(f)=p} h(f \star g_v) - h(f) = g_p \quad (8)$$

where  $g_p$  denotes the Gaussian density with zero mean and variance  $p$ , and the functions  $f$  are density functions on  $\mathbb{R}$ , i.e., positive functions integrating to 1, and having a well-defined entropy and second moment  $m_2(f) = \int_{\mathbb{R}} x^2 f(x) dx$ .

We consider the local geometry by looking at densities of the form

$$f_\varepsilon(x) = g_p(x)(1 + \varepsilon L(x)), \quad x \in \mathbb{R}, \quad (9)$$

where  $L: \mathbb{R} \rightarrow \mathbb{R}$  satisfies

$$\inf_{x \in \mathbb{R}} L(x) > -\infty \quad (10)$$

$$\int_{\mathbb{R}} L(x) g_p(x) dx = 0. \quad (11)$$

With these two constraints on  $L$ ,  $f_\varepsilon$  is a valid density for  $\varepsilon$  sufficiently small. It is a perturbed Gaussian density, in a ‘‘direction’’  $L$ . Observe that,

$$m_1(f_\varepsilon) = 0 \quad \text{iff} \quad M_1(L) = \int_{\mathbb{R}} x L(x) g_p(x) dx = 0 \quad (12)$$

$$m_2(f_\varepsilon) = p \quad \text{iff} \quad M_2(L) = \int_{\mathbb{R}} x^2 L(x) g_p(x) dx = 0. \quad (13)$$

We are now interested in analyzing how these perturbations affect the output distributions through an AWGN channel. Note that, if the input distribution is a Gaussian  $g_p$  perturbed in the direction  $L$ , the output is a Gaussian  $g_{p+v}$  perturbed in the direction  $\frac{(g_p L) \star g_v}{g_{p+v}}$ , since

$$f_\varepsilon \star g_v = g_{p+v} \left( 1 + \varepsilon \frac{(g_p L) \star g_v}{g_{p+v}} \right).$$

*Convention:*  $g_p L \star g_v$  refers to  $(g_p L) \star g_v$ , i.e., the multiplicative operator precedes the convolution one.

For simplicity, let us assume in the following that the function  $L$  is a polynomial satisfying (10), (11).

*Lemma 1:* We have

$$D(f_\varepsilon \| g_p) = \frac{1}{2} \varepsilon^2 \|L\|_{g_p}^2 + o(\varepsilon^2)$$

$$D(f_\varepsilon \star g_v \| g_p \star g_v) = \frac{1}{2} \varepsilon^2 \left\| \frac{g_p L \star g_v}{g_{p+v}} \right\|_{g_{p+v}}^2 + o(\varepsilon^2).$$

where

$$\|L\|_{g_p}^2 = \int_{\mathbb{R}} L^2(x) g_p(x) dx.$$

Moreover, note that for any density  $f$ , if  $m_1(f) = a$  and  $m_2(f) = p + a^2$ , we have

$$h(f) = h(g_{a,p}) - D(f \| g_{a,p}). \quad (14)$$

Hence, the extremal entropic results of (7) and (8) are locally expressed as

$$\arg \min_{L: M_2(L)=0} \left\| \frac{g_p L \star g_v}{g_{p+v}} \right\|_{g_{p+v}}^2 = 0 \quad (15)$$

$$\arg \max_{L: M_2(L)=0} \left\| \frac{g_p L \star g_v}{g_{p+v}} \right\|_{g_{p+v}}^2 - \|L\|_{g_p}^2 = 0, \quad (16)$$

where 0 denotes here the zero function. If (15) is obvious, (16) requires a proof which will be done in section V. Let us define the following mapping,

$$\Gamma^{(+)} : L \in L_2(g_p) \mapsto \frac{g_p L \star g_v}{g_{p+v}} \in L_2(g_{p+v}; \mathbb{R}), \quad (17)$$

where  $L_2(g_p)$  denotes the space of real functions having a finite  $\|\cdot\|_{g_p}$  norm. This linear mapping gives, for a given perturbed direction  $L$  of a Gaussian input  $g_p$ , the resulting perturbed direction of the output through additive Gaussian noise  $g_v$ . The norm of each direction in their respective spaces, i.e., in  $L_2(g_p)$  and  $L_2(g_{p+v})$ , gives how far from the Gaussian distribution these perturbations are (up to a scaling factor). Note that if  $L$  satisfies (10)-(11), so does  $\Gamma^{(+)}L$  for the measure  $g_{p+v}$ . The result in (16) (worst noise case) tells us that this mapping is a contraction, but to tackle the competitive situations described in previous section, what would be helpful is a spectral analysis of this operator, to allow more quantitative results than the extreme-case results of (15) and (16).

In order to do so, one can express  $\Gamma^{(+)}$  as an operator defined and valued in the same space, namely  $L_2$  with the Lebesgue measure  $\lambda$ , which is done by inserting the Gaussian measure in the operator argument, and we can then proceed to a singular function/value analysis. Formally, let  $K = L\sqrt{g_p}$ , which gives  $\|K\|_\lambda = \|L\|_{g_p}$ , and let

$$\Lambda : K \in L_2(\lambda) \mapsto \frac{\sqrt{g_p} K \star g_v}{\sqrt{g_{p+v}}} \in L_2(\lambda) \quad (18)$$

which gives  $\|\Gamma^{(+)}L\|_{g_{p+v}} = \|\Lambda K\|_\lambda$ . Denoting by  $\Lambda^*$  the adjoint operator of  $\Lambda$ , we want to find the singular functions of  $\Lambda$ , i.e., the eigenfunctions  $K$  of  $\Lambda^*\Lambda$ :

$$\Lambda^*\Lambda K = \gamma K.$$

### III. RESULTS

#### A. General Result: Local Geometry and Hermite Coordinates

The following theorem gives the singular functions and values of the operator  $\Lambda$  defined in previous section.

*Theorem 1:*

$$\Lambda^*\Lambda K = \gamma K, \quad K \neq 0$$

holds for each pair

$$(K, \gamma) \in \left\{ \left( \sqrt{g_p} H_k^{[p]}, \left( \frac{p}{p+v} \right)^k \right) \right\}_{k \geq 0},$$

where

$$H_k^{[p]}(x) = \frac{1}{\sqrt{k!}} H_k(x/\sqrt{p})$$

$$H_k(x) = (-1)^k e^{x^2/2} \frac{d^k}{dx^k} e^{-x^2/2}, \quad k \geq 0, x \in \mathbb{R}.$$

The polynomials  $H_k^{[p]}$  are the normalized Hermite polynomials (for a Gaussian distribution having variance  $p$ ) and  $\sqrt{g_p} H_k^{[p]}$  are called the Hermite functions.

The following result contains the property of Hermite polynomials mostly used in our problems, and expresses Theorem 1 with Gaussian measures.

*Theorem 2:*

$$\Gamma^{(+)} H_k^{[p]} = \frac{g_p H_k^{[p]} \star g_v}{g_{p+v}} = \left( \frac{p}{p+v} \right)^{k/2} H_k^{[p+v]}, \quad (19)$$

$$\Gamma^{(-)} H_k^{[p+v]} = H_k^{[p+v]} \star g_v = \left( \frac{p}{p+v} \right)^{k/2} H_k^{[p]}. \quad (20)$$

Last Theorem implies Theorem 1, since

$$\Gamma^{(-)} \Gamma^{(+)} L = \gamma L \iff \Lambda^* \Lambda K = \gamma K$$

for

$$K = L \sqrt{g_p}.$$

Note: It is expected that these results are related to properties of the Ornstein-Uhlenbeck process.

**Valid Hermite directions:** In view of previous results, we would like to perturb Gaussians along Hermite polynomials. Then, over an additive Gaussian noise channel  $g_v$ , if we perturb the input  $g_p$  in the direction  $H_k^{[p]}$  by an amount  $\varepsilon$ , we perturb the output  $g_{p+v}$  in the direction  $H_k^{[p+v]}$  by an amount  $\left(\frac{p}{p+v}\right)^{k/2} \varepsilon$  and such a perturbation in  $H_k$  implies that the output entropy is reduced (compared to not perturbing) by  $\left(\frac{p}{p+v}\right)^k \frac{\varepsilon^2}{2}$ . However, we need to ensure that the perturbations are valid, namely, that (10) and (11) are satisfied, and this needs to be discussed carefully. For any  $p > 0$ ,  $\{H_k^{[p]}\}_{k \geq 0}$  is an orthonormal basis of  $L_2(g_p)$ , this can be found in [18]. Clearly,  $H_0^{[p]} = 1$  is not a valid direction; but the orthogonality property and  $H_0^{[p]} = 1$  imply that  $H_k^{[p]}$  satisfies (11) for any  $k \geq 1$ . On the other hand, it is only for even values of  $k$  that (10) is verified. Yet, any perturbation of the form  $L = \sum_{k=1}^K \alpha_k H_k^{[p]}$  is valid, as long as  $K$  is even and  $\alpha_K > 0$ . This will essentially allow us to consider  $H_k^{[p]}$  with  $k$  odd as directions. Indeed, for any  $\delta > 0$ , we have that  $H_k^{[p]} + \delta H_{4k}^{[p]}$  satisfies (10), whether  $k$  is even or not (we chose  $4k$  instead of  $2k$  for reasons that will become clear later). Now, if we consider the direction  $-H_k^{[p]}$ , (10) is not satisfied for both  $k$  even and odd. But again, for any  $\delta > 0$ , we have that  $-H_k^{[p]} + \delta H_{4k}^{[p]}$  satisfies (10). Hence, in order to ensure (10), we will often work in the proofs with  $\pm H_k^{[p]} + \delta H_{4k}^{[p]}$ , although it will essentially allow us to reach the performance achieved by any  $\pm H_k^{[p]}$  (odd or even), since we will then take  $\delta$  arbitrarily

small and use continuity arguments. Moreover, one can check that  $H_1$ , respectively  $H_2$  perturb a Gaussian distribution into another Gaussian distribution, with a different first moment, respectively second moment. It is only for  $k \geq 3$  that the  $H_k$  perturbations are not modifying the first two moments and are moving away from Gaussian distributions. Finally, note that, in order to make sense of the approximations in Lemma 1, the directions  $L$  must have a finite  $L_2$  norm with respect to the Gaussian measure, which is ensured and spanned by the functions expressed in the Hermite basis.

*Convention:* We drop the variance upper script in the Hermite terms whenever a Gaussian density with specified variance is perturbed, i.e., the density  $g_p(1 + \varepsilon H_k)$  always denotes  $g_p(1 + \varepsilon H_k^{[p]})$ , and  $g_p H_k$  always denotes  $g_p H_k^{[p]}$ , no matter what  $p$  is. Same treatment is done for  $\|\cdot\|_{g_p}$  and  $\|\cdot\|$ .

### B. Fading BC Result

The following result states that the capacity region of a degraded fading BC with Gaussian noise is not achieved by a Gaussian superposition code in general.

*Theorem 3:* Let

$$Y_1 = HX + Z_1,$$

$$Y_2 = HX + Z_2$$

with  $X$  such that  $\mathbb{E}X^2 \leq p$ ,  $Z_1 \sim \mathcal{N}(0, v)$ ,  $0 < v < 1$ ,  $Z_2 \sim \mathcal{N}(0, 1)$  and  $H, X, Z_1, Z_2$  mutually independent. There exist fading distributions and values of  $v$  for which the capacity achieving input distribution is non-Gaussian. More precisely, let  $U$  be any auxiliary random variable, with  $U - X - (Y_1, Y_2)$ . Then, there exists  $p, v$ , a distribution of  $H$  and  $\mu$  such that

$$(U, X) \mapsto I(X; Y_1 | U, H) + \mu I(U; Y_2 | H) \quad (21)$$

is maximized by a non jointly Gaussian distribution.

In the proof, we present a counter-example to Gaussian being optimal for  $H$  binary. In order to defeat Gaussian distributions, we construct input distributions using the Hermite coordinates. The proof also gives a condition on the fading distribution and the noise variance  $v$  for which a non-Gaussian distribution strictly improves on the Gaussian one.

### C. IC Result

Recall that

$$S_{a,p}(X_1, X_2)$$

$$= I(X_1; X_1 + aX_2 + Z_1) + I(X_2; X_2 + aX_1 + Z_2).$$

*Definition 2:* Let

$$F_k(a, p) = \lim_{\delta \searrow 0} \lim_{\varepsilon \searrow 0} \frac{2}{\varepsilon^2} [S_{a,p}(X_1, X_2) - S_{a,p}(X_1^G, X_2^G)],$$

where  $X_1^G, X_2^G \stackrel{\text{iid}}{\sim} g_p$ ,  $X_1 \sim g_p(1 + \varepsilon \tilde{H}_k)$ ,  $X_2 \sim g_p(1 - \varepsilon \tilde{H}_k)$ , and  $\tilde{H}_k$  is defined by

$$b \tilde{H}_k^{[p]} = \begin{cases} b(H_k^{[p]} + \delta H_{4k}^{[p]}), & \text{if } b \geq 0, \\ b(H_k^{[p]} - \delta H_{4k}^{[p]}), & \text{if } b < 0. \end{cases} \quad (22)$$

In other words,  $F_k(a, p)$  represents the gain (positive or negative) of using  $X_1$  perturbed along  $H_k$  and  $X_2$  perturbed along  $-H_k$  with respect to using Gaussian distributions. Note that the distributions we chose for  $X_1$  and  $X_2$  are not the most general ones, as we could have chosen arbitrary directions spanned by the Hermite basis to perturb the Gaussian densities. However, as explained in the proof of the theorem 4, this choice is sufficiently general for our purpose.

*Theorem 4:* We have for  $k \geq 2$

$$F_k(a, p) = \left[ \frac{a^2}{pa^2 + 1} \right]^k - \frac{(a^k - 1)^2}{(pa^2 + p + 1)^k}.$$

For any fixed  $p$ , the function  $F_k(\cdot, p)$  has a unique positive root, below which it is negative and above which it is positive.

*Theorem 5:* Treating interference as noise with iid Gaussian inputs does not achieve the sum-capacity of the symmetric IC (synch or asynch) and is outperformed by  $X_1 \sim g_p(1 + \varepsilon \tilde{H}_3)$  and  $X_2 \sim g_p(1 - \varepsilon \tilde{H}_3)$ , if  $F_3(a, p) > 0$ .

This Theorem is a direct consequence of Theorem 4.

*Proposition 1:* For the symmetric synch IC, time sharing improves on treating interference as noise with iid Gaussian distribution if  $F_2(a, p) > 0$ .

We now introduce the following definition.

*Definition 3:* Blind time sharing over a block length  $n$  (assumed to be even) between two users, refers to sending non-zero power symbols only at the instances marked with a 1 in  $(1, 0, 1, 0, 1, 0, \dots, 1, 0)$  for the first user, and zero power symbols only at the instances marked with a 1 in  $(1, 1, \dots, 1, 0, 0, \dots, 0)$  for the second user.

*Proposition 2:* For the symmetric totally asynch IC, if the receivers (but not transmitters) know the asynchronization delay, blind time sharing improves on treating interference as noise with iid Gaussian distributions if  $B_2(a, p) > 0$ , where  $B_2(a, p) = \frac{1}{4}(\log(1+2p) + \log(1 + \frac{2p}{1+2a^2p})) - \log(1 + \frac{p}{1+a^2p})$ . If the receivers do not know the asynchronization delay, blind time sharing cannot improve on treating interference as noise with iid Gaussian distributions if  $B_2(a, p) \leq 0$ .

*How to read these results:* We have four thresholds to keep track of:

- $T_1(p)$  is when  $pa^3 + a - \frac{1}{2} = 0$ . If  $a \leq T_1(p)$ , we know from [3], [13], [16] that iid Gaussian inputs and treating interference as noise is sum-capacity achieving.
- $T_2(p)$  is when  $F_2(a, p) = 0$ . If  $a > T_2(p)$ , we know from Prop. 1 that, if synchronization is permitted, time sharing improves on treating interference as noise with iid Gaussian inputs. This regime matches with the so-called moderate regime defined in [9].
- $T_3(p)$  is when  $F_3(a, p) = 0$ . If  $a > T_3(p)$ , we know from Prop. 5 that treating interference as noise with iid non-Gaussian distributions (opposites in  $H_3$ ) improves on the iid Gaussian ones.
- $T_4(p)$  is when  $B_2(a, p) = 0$ . If  $a > T_4(p)$ , we know from Prop. 2 that, even if the users are totally asynchronized, but if the receivers know the asynchronization delay, blind time sharing improves on treating interference as noise

with iid Gaussian inputs. If the receivers do not know the delay, the threshold can only appear for larger values of  $a$ .

The question is now, how are these thresholds ranked. It turns out that

$$0 < T_1(p) < T_2(p) < T_3(p) < T_4(p)$$

and if  $p = 1$ , the above inequality reads as

$$0.424 < 0.605 < 0.680 < 1.031.$$

This implies the following: For a decoder that treats interference as noise. Since  $T_2(p) < T_3(p)$ , it is first better to time share than using non-Gaussian distributions along  $H_3$ . But this is useful only if time-sharing is permitted, i.e., for the synch IC. However, for the asynch IC, since  $T_3(p) < T_4(p)$ , we are better off using the non-Gaussian distributions along  $H_3$  before a Gaussian input scheme, even with blind time-sharing, and even if the receiver could know the delay. We notice that there is still a gap between  $T_1(p)$  and  $T_2(p)$ , and we cannot say if, in this range, iid Gaussian inputs are still optimal, or if another class of non-Gaussian inputs (far away from Gaussians) can outperform them. In [8], another technique (which is related to ours but not equivalent) is used to find regimes where non-Gaussian inputs can improve on Gaussian ones on the same problem that we consider here. The threshold found in [8] is equal to 0.925 for  $p = 1$ , which is looser than the value of 0.680 found here (we pick  $p = 1$  as an arbitrary example).

Finally, the following interesting and curious fact has also been noticed. In theorem 4, we require  $k \geq 2$ . Nevertheless, if we plug  $k = 1$  in the right hand side of theorem 4 and ask for this expressions to be positive, we precisely get  $pa^3 + a - \frac{1}{2} > 0$ , i.e., the complement range delimited by  $T_1(p)$ . However, the right hand side of theorem 4 for  $k = 1$  is not equal to  $F_1(a, p)$  (this is explained in more details in the proof of theorem 4). Indeed, it would not make sense that moving along  $H_1$ , which changes the mean with a fixed second moment within Gaussians, would allow us to improve on the iid Gaussian scheme. Yet, getting to the exact same condition, when working on the problem of improving on the iid Gaussian scheme, seems to be a strange coincidence.

#### D. Strong Shamai-Laroia Conjecture

We show in Section V that conjecture 1 does not hold. We provide counter-examples to the conjecture, pointing out that the range of  $h$  for which the conjecture does not hold.

## IV. HERMITE CODING: FORMALITIES

In order to evaluate the entropy of a perturbation, i.e.,  $h(g_p(1 + \varepsilon L))$ , we can express it as the entropy of  $h(g_p)$  minus the divergence gap, as in (14), and then use Lemma 1 for the approximation. But this is correct if  $g_p(1 + \varepsilon L)$  has the same first two moments as  $g_p$ . Hence, if  $L$  contains only  $H_k$ 's with  $k \geq 3$ , the previous argument can be used. But if

$L$  contains  $H_1$  and/or  $H_2$  terms, the situation can be different. Next Lemma describes this. Recall that for  $\delta, p > 0$ ,

$$b\tilde{H}_k^{[p]} = \begin{cases} b(H_k^{[p]} + \delta H_{4k}^{[p]}), & \text{if } b \geq 0, \\ b(H_k^{[p]} - \delta H_{4k}^{[p]}), & \text{if } b < 0. \end{cases} \quad (23)$$

*Lemma 2:* We have for any  $\alpha_k \in \mathbb{R}, k \geq 1, \varepsilon > 0$

$$\begin{aligned} & h(g_p(1 + \varepsilon \sum_{k \geq 1} \alpha_k \tilde{H}_k)) = \\ & h(g_p) - D(g_p(1 + \varepsilon \sum_{k \geq 1} \alpha_k \tilde{H}_k) \| g_p) + \frac{\varepsilon \alpha_2}{\sqrt{2}}. \end{aligned}$$

Finally, when we convolve two perturbed Gaussian distributions, we get

$$\begin{aligned} & g_a(1 + \varepsilon H_j) \star g_b(1 + \varepsilon H_k) = \\ & g_{a+b} + \varepsilon [g_a H_j \star g_b + g_a \star g_b H_k] + \varepsilon^2 g_a H_j \star g_b H_k. \end{aligned}$$

We already know from Theorem 2 how to describe the terms in  $\varepsilon$ , what we still need is to describe the term in  $\varepsilon^2$ . We have the following.

*Lemma 3:* We have

$$g_a H_k^{[a]} \star g_b H_l^{[b]} = C g_{a+b} H_{k+l}^{[a+b]},$$

where  $C$  is a constant depending only on  $a, b, k$  and  $l$ . In particular if  $k = l = 1$ , we have  $C = \frac{\sqrt{2ab}}{a+b}$ .

Note: In general

$$C = \binom{k+l}{k}^{1/2} \left(\frac{a}{a+b}\right)^{k/2} \left(\frac{b}{a+b}\right)^{l/2},$$

this result is however proved in a separate note [1], since it is not used in this paper.

## V. PROOFS

We start by reviewing the proof of (16), as it brings interesting facts. We then prove the main result.

*Proof of (16):*

We first assume that  $f_\varepsilon$  has zero mean and variance  $p$ . Using the Hermite basis, we then take  $L = \sum_{k \geq 3} \alpha_k H_k^{[p]}$ . Using (19), we can express (16) as

$$\sum_{k \geq 3} \alpha_k^2 \left(\frac{p}{p+v}\right)^k - \sum_{k \geq 3} \alpha_k^2 \quad (24)$$

which is clearly negative. Hence, we have proved that

$$\| \frac{g_p L \star g_v}{g_{p+v}} \|_{g_{p+v}}^2 \leq \|L\|_{g_p}^2 \quad (25)$$

and (16) is maximized by taking  $L = 0$ . Note that we can get tighter bounds than the one in previous inequality, indeed the tightest, holding for  $H_3$ , is given by

$$\| \frac{g_p L \star g_v}{g_{p+v}} \|_{g_{p+v}}^2 \leq \left(\frac{p}{p+v}\right)^3 \|L\|_{g_p}^2 \quad (26)$$

(this clearly holds if written as a series like in (24)). Hence, locally the contraction property can be tightened, and locally,

we have stronger EPI's, or worst noise case. Namely, if  $\nu \geq \left(\frac{p}{p+v}\right)^3$ , we have

$$\arg \min_{f: m_1(f)=0, m_2(f)=p} h(f \star g_v) - \nu h(f) = g_p \quad (27)$$

and if  $\nu < \left(\frac{p}{p+v}\right)^3$ ,  $g_p$  is outperformed by non-Gaussian distributions. Now, if we consider the constraint  $m_2(f) \leq p$ , we get that if  $\nu \geq \frac{p}{p+v}$ ,

$$\arg \min_{f: m_2(f) \leq p} h(f \star g_v) - \nu h(f) = g_p \quad (28)$$

and if  $\nu < \frac{p}{p+v}$ ,  $g_p$  is outperformed by  $g_{p-\delta}$  for some  $\delta > 0$ . It would then be interesting to study if these tighter results hold in a greater generality than for the local setting.

*Proof of Theorem 2:*

We want to show

$$\begin{aligned} H_k^{[p+v]} \star g_v &= \left(\frac{p}{p+v}\right)^{k/2} H_k^{[p]}, \\ g_p H_k^{[p]} \star g_v &= \left(\frac{p}{p+v}\right)^{k/2} g_{p+v} H_k^{[p+v]}, \end{aligned}$$

which is proved by an induction on  $k$ , using the following properties (Appell sequence and recurrence relation) of Hermite polynomials:

$$\frac{\partial}{\partial x} H_{k+1}^{[p]}(x) = \sqrt{\frac{k+1}{p}} H_k^{[p]}(x)$$

$$\frac{\partial}{\partial x} \left( g_p(x) H_k^{[p]}(x) \right) = -\sqrt{\frac{k+1}{p}} g_p(x) H_{k+1}^{[p]}(x).$$

*Proof of Theorem 3:*

We refer to (21) as the mu-rate. Let us first consider Gaussian codes, i.e., when  $(U, X)$  is jointly Gaussian, and see what mu-rate they can achieve. Without loss of generality, we can assume that  $X = U + V$ , with  $U$  and  $V$  independent and Gaussian, with respective variance  $Q$  and  $R$  satisfying  $P = Q + R$ . Then, (21) becomes

$$\frac{1}{2} \mathbb{E} \log \left( 1 + \frac{RH^2}{v} \right) + \mu \frac{1}{2} \mathbb{E} \log \frac{1 + PH^2}{1 + RH^2}. \quad (29)$$

Now, we pick a  $\mu$  and look for the optimal power  $R$  that must be allocated to  $V$  in order to maximize the above expression. We are interested in cases for which the optimal  $R$  is not at the boundary but at an extremum of (29), and if the maxima is unique, the optimal  $R$  is found by the first derivative check, which gives  $\mathbb{E} \frac{H^2}{v+RH^2} = \mu \mathbb{E} \frac{H^2}{1+RH^2}$ . Since we will look for  $\mu, v$ , with  $R > 0$ , previous condition can be written as

$$\mathbb{E} \frac{RH^2}{v+RH^2} = \mu \mathbb{E} \frac{RH^2}{1+RH^2}. \quad (30)$$

We now check if we can improve on (29) by moving away from the optimal jointly Gaussian  $(U, X)$ . There are several ways to perturb  $(U, X)$ , we consider first the following case. We keep  $U$  and  $V$  independent, but perturb them away from Gaussian's in the following way:

$$p_{U_\varepsilon}(u) = g_Q(u)(1 + \varepsilon(H_3^{[Q]}(u) + \delta H_4)) \quad (31)$$

$$p_{V_\varepsilon}(v) = g_R(v)(1 - \varepsilon(H_3^{[R]}(v) - \delta H_4)) \quad (32)$$

with  $\varepsilon, \delta > 0$  small enough. Note that these are valid density functions and that they preserve the first two moments of  $U$  and  $V$ . The reason why we add  $\delta H_4$ , is to ensure that (13) is satisfied, but we will see that for our purpose, this can essentially be neglected. Then, using Lemma 2, the new distribution of  $X$  is given by

$$p_X(x) = g_P(x) \left(1 + \varepsilon \left(\frac{Q}{P}\right)^{\frac{3}{2}} H_3^{[P]} - \varepsilon \left(\frac{R}{P}\right)^{\frac{3}{2}} H_3^{[P]}\right) + f(\delta)$$

where  $f(\delta) = \delta g_P(x) \varepsilon \left(\left(\frac{Q}{P}\right)^{\frac{4}{2}} H_4^{[P]} + \left(\frac{R}{P}\right)^{\frac{4}{2}} H_4^{[P]}\right)$ , which tends to zero when  $\delta$  tends to zero. Now, by picking  $P = 2R$ , we have

$$p_X(x) = g_P(x) + f(\delta). \quad (33)$$

Hence, by taking  $\delta$  arbitrarily small, the distribution of  $X$  is arbitrarily close to the Gaussian distribution with variance  $P$ . We now want to evaluate how these Hermite perturbations perform, given that we want to maximize (21), i.e.,

$$h(Y_1|U, H) - h(Z_1) + \mu h(Y_2|H) - \mu h(Y_2|U, H). \quad (34)$$

We wonder if, by moving away from Gaussian distributions, the gain achieved for the term  $-h(Y_2|U, H)$  is higher than the loss suffered from the other terms. Using Theorem 2, Lemma 1 and Lemma 2, we are able to precisely measure this and we get

$$\begin{aligned} h(Y_1|U = u, H = h) &= h(g_{hu, v+Rh^2} (1 - \varepsilon \left(\frac{Rh^2}{v+Rh^2}\right)^{\frac{3}{2}} H_3^{[hu, v+Rh^2]})) \\ &= \frac{1}{2} \log 2\pi e(v+Rh^2) - \frac{\varepsilon^2}{2} \left(\frac{Rh^2}{v+Rh^2}\right)^3 + o(\varepsilon^2) + o(\delta) \end{aligned}$$

$$\begin{aligned} h(Y_2|U = u, H = h) &= \frac{1}{2} \log 2\pi e(1+Rh^2) - \frac{\varepsilon^2}{2} \left(\frac{Rh^2}{1+Rh^2}\right)^3 + o(\varepsilon^2) + o(\delta) \end{aligned}$$

and because of (33)

$$h(Y_2|H = h) = \frac{1}{2} \log 2\pi e(1+Ph^2) + o(\varepsilon^2) + o(\delta).$$

Therefore, collecting all terms, we find that for  $U_\varepsilon$  and  $V_\varepsilon$  defined in (31) and (32), expression (42) reduces to

$$\begin{aligned} I_G - \frac{\varepsilon^2}{2} \mathbb{E} \left(\frac{RH^2}{v+RH^2}\right)^3 + \mu \frac{\varepsilon^2}{2} \mathbb{E} \left(\frac{RH^2}{1+RH^2}\right)^3 \\ + o(\varepsilon^2) + o(\delta) \end{aligned} \quad (35)$$

where  $I_G$  is equal to (29) (and is the mu-rate obtained with Gaussian inputs). Hence, if for some distribution of  $H$  and some  $v$ , we have that

$$\mu \mathbb{E} \left(\frac{RH^2}{1+RH^2}\right)^k - \mathbb{E} \left(\frac{RH^2}{v+RH^2}\right)^k > 0, \quad (36)$$

when  $k = 3$  and  $R$  is optimal for  $\mu$ , we can take  $\varepsilon$  and  $\delta$  small enough in order to make (35) strictly larger than

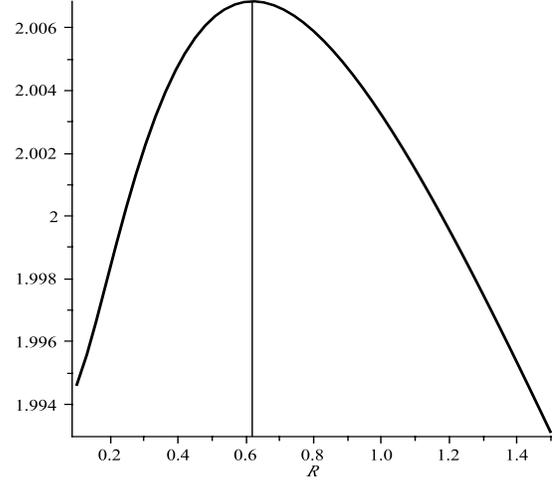


Fig. 1. Gaussian mu-rate, i.e., expression (29), plotted as a function of  $R$  for  $\mu = 5/4$ ,  $v = 1/4$ ,  $P = 1.24086308$  and  $H$  binary  $\{1; 10\}$ . Maxima at  $R = 0.62043154$ .

$I_G$ . We have shown how, if verified, inequality (36) leads to counter-examples of the Gaussian optimality, but with similar expansions, we would also get counter-examples if the following inequality holds for any power  $k$  instead of 3, as long as  $k \geq 3$ . Let us summarize what we obtained: Let  $R$  be optimal for  $\mu$ , which means that (30) holds if there is only one maxima (not at the boarder). Then, non-Gaussian codes along Hermite's strictly outperforms Gaussian codes, if, for some  $k \geq 3$ , (36) holds. If the maxima is unique, this becomes

$$\frac{\mathbb{E}T(v)^k}{\mathbb{E}T(1)^k} < \frac{\mathbb{E}T(v)}{\mathbb{E}T(1)}$$

where

$$T(v) = \frac{RH^2}{v+RH^2}.$$

So we want the Jensen gap of  $T(v)$  for the power  $k$  to be small enough compared to the Jensen gap of  $T(1)$ .

We now give an example of a fading distribution for which the above conditions can be verified. Let  $H$  be binary, taking values 1 and 10 with probability half and let  $v = 1/4$ . Let  $\mu = 5/4$ , then for any values of  $P$ , the maximizer of (29) is at  $R = 0.62043154$ , cf. Figure 1, which corresponds in this case to the unique value of  $R$  for which (30) is satisfied. Hence if  $P$  is larger than this value of  $R$ , there is a corresponding fading BC for which the best Gaussian code splits the power on  $U$  and  $V$  with  $R = 0.62043154$  to achieve the best mu-rate with  $\mu = 5/4$ . To fit the counter-examples with the choice of Hermite perturbations made previously, we pick  $P = 2R$ . Finally, for these values of  $\mu$  and  $R$ , (36) can be verified for  $k = 8$ , cf. Figure 2, and the corresponding Hermite code (along  $H_8$ ) strictly outperforms any Gaussian codes.

Note that we can consider other non-Gaussian encoders, such as when  $U$  and  $V$  are independent with  $U$  Gaussian and  $V$  non-Gaussian along Hermite's. Then, we get the following

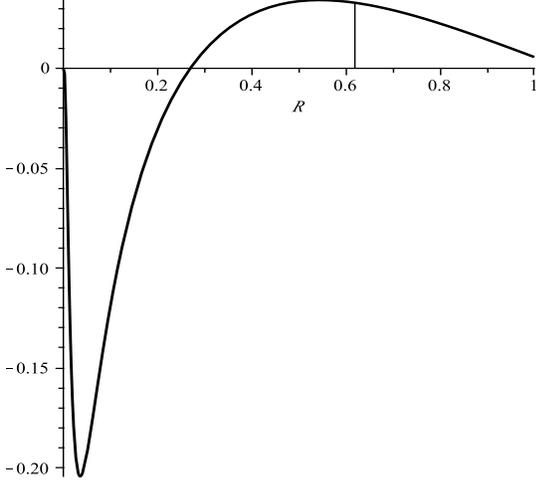


Fig. 2. LHS of (36) as a function of  $R$ , for  $\mu = 5/4$ ,  $v = 1/4$ ,  $k = 8$  and  $H$  binary  $\{1; 10\}$ , positive at  $R = 0.62043154$ .

condition. If for  $k \geq 3$  and  $R$  optimal for  $\mu$ , we have

$$\mathbb{E} \left( \frac{RH^2}{v + RH^2} \right)^k \quad (37)$$

$$< \mu \left[ \mathbb{E} \left( \frac{RH^2}{1 + RH^2} \right)^k - \left( \frac{R}{P} \right)^k \mathbb{E} \left( \frac{PH^2}{1 + PH^2} \right)^k \right], \quad (38)$$

then Gaussian encoders are not optimal. Notice that previous inequality is stronger than the one in (36) for fixed values of the parameters. Yet, it can still be verified for valid values of the parameters and there are also codes with  $U$  Gaussian and  $V$  non-Gaussian that outperform Gaussian codes for some degraded fading BCs.

*Proof of Theorem 4:*

Let  $\varepsilon, \delta > 0$  and let  $X_1$  and  $X_2$  be respectively distributed as  $g_p(1 + \varepsilon[H_k + \delta H_{4k}])$  and  $g_p(1 - \varepsilon[H_k - \delta H_{4k}])$ , where  $k \neq 1, 2$ . We have

$$I(X_1, X_1 + aX_2 + Z_1) = h(X_1 + aX_2 + Z_1) - h(aX_2 + Z_1)$$

where  $X_k^G$  are independent Gaussian 0-mean and  $p$ -variance random variables. Hence, we need to evaluate the contribution of each divergence appearing in previous expression, in order to know if the perturbations are improving on the Gaussian distributions. Let us first analyze  $h(X_1 + aX_2 + Z_1)$ . The density of  $X_1 + aX_2 + Z_1$  is given by

$$g_p(1 + \varepsilon[H_k + \delta H_{4k}]) \star g_{a^2p}(1 - \varepsilon[H_k - \delta H_{4k}]) \star g_1, \quad (39)$$

which, from Theorem 2, is equal to

$$g_{p+a^2p+1} \left( 1 + \varepsilon \left\{ \left[ \left( \frac{p}{p+a^2p+1} \right)^{\frac{k}{2}} H_k + \delta \left( \frac{p}{p+a^2p+1} \right)^{2k} H_{4k} \right] - \left[ \left( \frac{a^2p}{p+a^2p+1} \right)^{\frac{k}{2}} H_k - \delta \left( \frac{a^2p}{p+a^2p+1} \right)^{2k} H_{4k} \right] - \varepsilon L \right\} \right)$$

where

$$L = \frac{g_p[H_k + \delta H_{4k}] \star g_{a^2p}[H_k - \delta H_{4k}] \star g_1}{g_{p+a^2p+1}}.$$

Note that each direction in each line of the bracket  $\{\cdot\}$  above, including  $L$ , satisfy (10) and (11). Using Lemma 3, we have

$$L = \frac{g_p[H_k + \delta H_{4k}] \star g_{a^2p+1} \left[ \left( \frac{a^2p}{a^2p+1} \right)^{\frac{k}{2}} H_k - \left( \frac{a^2p}{a^2p+1} \right)^{2k} \delta H_{4k} \right]}{g_{p+a^2p+1}} \\ = C_1 H_{2k}^{[p+a^2p+1]} + C_2 H_{5k}^{[p+a^2p+1]} + C_3 H_{8k}^{[p+a^2p+1]}, \quad (40)$$

where  $C_1, C_2, C_3$  are constants. Therefore, the density of  $X_1 + aX_2 + Z_1$  is a Gaussian  $g_{p+a^2p+1}$  perturbed along the direction  $H_k$  in the order  $\varepsilon$  and several  $H_l$  with  $l \geq 2k$  in the order  $\varepsilon^2$  (and other directions but that have a  $\delta$  order). So we can use Lemma 2 and write

$$h(X_1 + aX_2 + Z_1) = h(X_1^G + aX_2^G + Z_1) - D(X_1 + aX_2 + Z_1 || X_1^G + aX_2^G + Z_1)$$

Using Lemma 1, we have

$$D(X_1 + aX_2 + Z_1 || X_1^G + aX_2^G + Z_1) \\ \doteq \frac{\varepsilon^2}{2} \left\| \left[ \left( \frac{p}{p+a^2p+1} \right)^{\frac{k}{2}} H_k + \delta \left( \frac{p}{p+a^2p+1} \right)^{2k} H_{4k} \right] - \left[ \left( \frac{a^2p}{p+a^2p+1} \right)^{\frac{k}{2}} H_k - \delta \left( \frac{a^2p}{p+a^2p+1} \right)^{2k} H_{4k} \right] \right\|^2 \\ = \frac{\varepsilon^2}{2} (1 - a^k)^2 \left( \frac{p}{p+a^2p+1} \right)^k + \frac{\varepsilon^2}{2} o(\delta).$$

Hence

$$h(X_1 + aX_2 + Z_1) = h(X_1^G + aX_2^G + Z_1) - \frac{\varepsilon^2}{2} (1 - a^k)^2 \left( \frac{p}{p+a^2p+1} \right)^k + \frac{\varepsilon^2}{2} o(\delta).$$

Similarly, we get

$$D(aX_2 + Z_1 || aX_2^G + Z_1) \doteq \frac{\varepsilon^2}{2} \left( \frac{a^2p}{a^2p+1} \right)^k + \frac{\varepsilon^2}{2} o(\delta)$$

and

$$I(X_1, X_1 + aX_2 + Z_1) \doteq I(X_1^G, X_1^G + aX_2^G + Z_1) + \frac{\varepsilon^2}{2} \left[ \left( \frac{a^2p}{a^2p+1} \right)^k - (1 - a^k)^2 \left( \frac{p}{p+a^2p+1} \right)^k \right] + \frac{\varepsilon^2}{2} o(\delta).$$

Finally, we have

$$I(X_2, X_2 + aX_1 + Z_2) = I(X_1, X_1 + aX_2 + Z_1)$$

and

$$\begin{aligned} & I(X_1, X_1 + aX_2 + Z_1) + I(X_2, X_2 + aX_1 + Z_2) \\ & \doteq I(X_1^G, X_1^G + aX_2^G + Z_1) + I(X_2^G, X_2^G + aX_1^G + Z_2) \\ & + \varepsilon^2 \left[ \left( \frac{a^2 p}{a^2 p + 1} \right)^k - (1 - a^k)^2 \left( \frac{p}{p + a^2 p + 1} \right)^k \right] + \varepsilon^2 o(\delta). \end{aligned}$$

Hence, if for some  $k \neq 3$  we have

$$\left( \frac{a^2}{a^2 p + 1} \right)^k - \frac{(a^k - 1)^2}{(p + a^2 p + 1)^k} > 0 \quad (41)$$

we can improve on the iid Gaussian distributions  $g_p$  by using the respective Hermite perturbations.

Now, we could have started with  $X_1$  and  $X_2$  distributed as  $g_p(1 + \varepsilon b_k \tilde{H}_k)$  and  $g_p(1 + \varepsilon c_k \tilde{H}_k)$ , where  $\tilde{H}_k$  is defined in (23). With similar expansions, we would then get that we can improve on the Gaussian distributions if for some some  $b_k, c_k$  and  $k \neq 1, 2$  we have

$$\begin{aligned} & \left[ \left( \frac{a^2 p}{a^2 p + 1} \right)^k - \frac{(a^2 p)^k + p^k}{(p + a^2 p + 1)^k} \right] (b_k^2 + c_k^2) \\ & - 4 \left( \frac{ap}{p + a^2 p + 1} \right)^k b_k c_k > 0. \end{aligned}$$

But the quadratic function

$$(b, c) \in \mathbb{R}^2 \mapsto \gamma(b^2 + c^2) - 2\delta bc,$$

with  $\delta > 0$ , can be made positive if and only if  $\gamma + \delta > 0$ , and is made so by taking  $b_k = -c_k$ . Hence, the initial choice we made about  $X_1$  and  $X_2$  is optimal. Moreover, note that for this distribution of  $X_1$  and  $X_2$ , we could have actually chosen  $k = 2$  as well. Because, even if Lemma 2 tells us that we must use correction terms, these correction terms will cancel out when we consider the sum-rate, since  $b_k = -c_k$  and since the correction is in  $\varepsilon$ . There is however another problem when using  $k = 2$ , which is that  $g_p(1 + \varepsilon H_2)$  has a larger second moment than  $p$ . However, if we use a scheme of block length 2, we can compensate this excess on the first channel use with the second channel use, and because of the symmetry, we can achieve the desired rate. But this is allowed only with synchronization. We could also have used perturbations that are mixtures of Hermite's, such as  $g_p(1 + \varepsilon \sum_k b_k H_k)$ . We would then get mixtures of previous equations as our condition. But in the current problem this will not be helpful. Finally, perturbing iid Gaussian inputs in a independent but non i.d. way, i.e., to perturb different components in different Hermite directions, cannot improve on our scheme, from previous arguments. The only option which is not investigated here (but in a work in progress), is to perturb iid Gaussian inputs in a non independent manner. Finally, if we work with  $k = 1$ , the proof sees the following modification. In (40), we now have a term in  $H_2$ . However, even if this term is in the order  $\varepsilon^2$ , we can no longer neglect it, since from Lemma

2, a  $\varepsilon^2 H_2$  term in the direction comes out as a  $\frac{\varepsilon^2}{\sqrt{2}}$  term in the entropy. Hence, we do not get the above condition for  $k = 1$ , but the one obtained by replacing  $(a^k - 1)^2$  with  $(a^2 + 1)$ , and the condition for positivity can never be fulfilled.

*Proof of Proposition 1:*

From Theorem 4, we know that when treating interference as noise and when  $F_2(a, p) > 0$ , it is better to use encoders drawn from the 2 dimensional distributions  $X_1^2$  and  $X_2^2$ , where  $(X_1)_1 \sim g_p(1 + \varepsilon \tilde{H}_2)$ ,  $(X_2)_1 \sim g_p(1 - \varepsilon \tilde{H}_2)$ ,  $(X_1)_2 \sim g_p(1 - \varepsilon \tilde{H}_2)$  and  $(X_2)_2 \sim g_p(1 + \varepsilon \tilde{H}_2)$ , as opposed to using Gaussian distributions. But perturbations in  $H_2$  are changing the second moment of the input distribution. Hence, this scheme is mimicing a time-sharing in our local setting. Moreover, a direct computation also allows to show that, constraining each user to use Gaussian inputs of arbitrarily block length  $n$ , with arbitrary covariances having a trace bounded by  $nP$ , the optimal covariances are  $pI_n$  if  $F_2(a, p) \leq 0$ , and otherwise, are given by a time-sharing scheme (cf. Definition 1 for the definition of a Gaussian time-sharing scheme and covariance matrices).

*Proof of Proposition 2:*

Note that when using blind time-sharing, no matter what the delay in the asynchronization of each user is, the users are interfering in  $n/4$  channel uses and have each a non-intefering channel in  $n/4$  channel uses (the rest of the  $n/4$  channel uses are not used by any users). Hence, if the receiver have the knowledge of the asynchronization delay, the following sum-rate can be achieved:  $\frac{1}{4}(\log(1 + 2p) + \log(1 + \frac{2p}{1 + 2a^2 p}))$ . And if the delay is unknown to the receivers, the previous sum-rate can surely not be improved on.

*Disproof of Conjecture 1:*

This proof uses similar steps as previous proofs. Using Lemma (14), we express

$$I(X; X + hX_1^G + Z) \leq I(X; X + hX_1 + Z).$$

as

$$\begin{aligned} & -D(X + hX_1^G + Z || X^G + hX_1^G + Z) \\ & \leq -D(X + hX_1 + Z || X^G + hX_1^G + Z) \\ & + D(hX_1 + Z || hX_1^G + Z). \end{aligned} \quad (42)$$

We then pick  $X, X_1 \sim g_p(1 + \varepsilon \tilde{H}_k)$  and assume that  $k$  is even for now. We then have

$$\begin{aligned} & X + hX_1^G + Z \\ & \sim g_p(1 + \varepsilon H_k) \star g_{ph^2+v} \\ & = g_{p+ph^2+v} \left( 1 + \varepsilon \left( \frac{p}{p + ph^2 + v} \right)^{k/2} H_k \right) \end{aligned}$$

and

$$\begin{aligned} & D(X + hX_1^G + Z || X^G + hX_1^G + Z) \\ & \doteq \frac{\varepsilon^2}{2} \left( \frac{p}{p + ph^2 + v} \right)^k. \end{aligned}$$

Similarly,

$$\begin{aligned}
& X + hX_1 + Z \\
& \sim g_p(1 + \varepsilon H_k) \star g_{ph^2}(1 + H_k) \star g_v \\
& = g_p(1 + \varepsilon H_k) \star g_{ph^2+v}(1 + (\frac{ph^2}{ph^2+v})^{k/2} H_k) \\
& g_{p+ph^2+v}(1 + \varepsilon(\frac{p}{p+ph^2+v})^{k/2} H_k) \\
& + \varepsilon(\frac{ph^2}{p+ph^2+v})^{k/2} H_k
\end{aligned}$$

and

$$\begin{aligned}
& D(X + hX_1 + Z || X^G + hX_1^G + Z) \\
& \doteq \frac{\varepsilon^2}{2} (\frac{p}{p+ph^2+v})^k (1 + h^k)^2.
\end{aligned}$$

Finally,

$$\begin{aligned}
& hX_1 + Z \sim g_{ph^2}(1 + \varepsilon H_k) \star g_v \\
& = g_{h^2p+v}(1 + \varepsilon(\frac{ph^2}{ph^2+v})^{k/2} H_k)
\end{aligned}$$

and

$$D(hX_1 + Z || hX_1^G + Z) \doteq \frac{\varepsilon^2}{2} (\frac{ph^2}{ph^2+v})^k,$$

Therefore, (42) is given by

$$\begin{aligned}
& - \left( \frac{p}{p+ph^2+v} \right)^k \\
& \leq - \left( \frac{p}{p+ph^2+v} \right)^k (1 + h^k)^2 + \left( \frac{ph^2}{ph^2+v} \right)^k + o(1)
\end{aligned}$$

and if

$$\begin{aligned}
& \left( \frac{p}{p+ph^2+v} \right)^k (1 + h^k)^2 - \left( \frac{p}{p+ph^2+v} \right)^k \\
& - \left( \frac{ph^2}{ph^2+v} \right)^k > 0
\end{aligned} \tag{43}$$

for some  $k$  even and greater than 4, we have a counter example to the strong conjecture. Note that, using the same trick as in previous proofs, that is, perturbing along  $\tilde{H}_k$  instead of  $H_k$ , we get that if (43) holds for any  $k = 3$ , we also have a counter example to the strong conjecture. Defining  $s := p/v = \text{SNR}$ , (43) is equivalent to

$$G(h, s, k) := \frac{h^{2k} + 2h^k}{(1 + h^2 + 1/s)^k} - \left( \frac{h^2}{h^2 + 1/s} \right)^k > 0. \tag{44}$$

As shown in Figure 3, this can indeed happen. Note that the range where (44) holds increases when  $s$  decreases, i.e., when SNR decreases. Indeed, when  $s = +\infty$ , which corresponds to dropping the additive noise  $Z$ , we do not get a counter-example to the conjecture. But in the presence of Gaussian noise, the conjecture does not hold for some distributions of  $X, X_1$  as illustrated above. Moreover, for  $k = 3$ , we get that  $G(h, s, 3) > 0$  is equivalent to

$$p(h, s) := (h^3 + 2)(sh^2 + 1)^3 - h^3(sh^2 + s + 1) > 0. \tag{45}$$

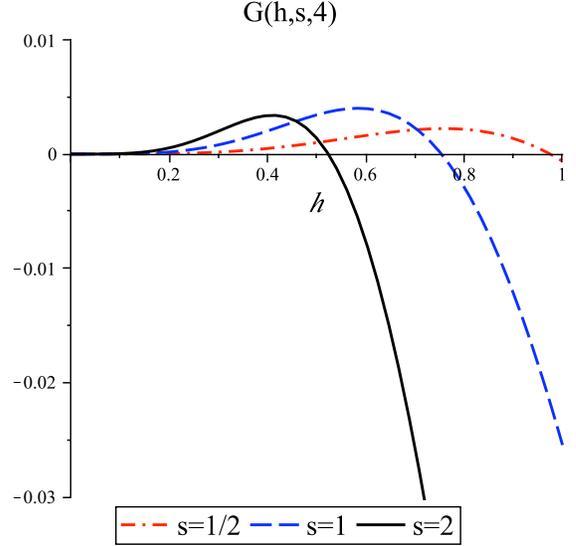


Fig. 3. Function  $h \mapsto G(h, s, 4)$  for different values of  $s$

For any fixed  $s$ ,  $p(h, s)$  is a polynomial of degree 7 in  $h$ , having a single non-zero root  $r(s)$ . Hence, if  $h \geq r(s)$ , the conjecture may still hold. Numerical evidence suggest that  $r(s)$  behaves like  $1/s$  (more precisely  $1.26/s$  for large  $s$ ). This means that if the interference coefficient  $h$  is at least of the order of  $1/\text{SNR}$ , the conjecture may still hold.

## VI. DISCUSSION

We have developed a technique to analyze codes drawn from non-Gaussian ensembles using the Hermite polynomials. If the performance of non-Gaussian inputs is usually hard to analyze, we showed how with this tool, it reduces to the analysis of analytic power series. This allowed us to show that Gaussian inputs are in general not optimal for degraded fading Gaussian BC, although they might still be optimal for many fading distributions. For the IC problem, we found that in the asynchronous setting and when treating interference as noise, using non-Gaussian code ensembles ( $H_3$  perturbations) can strictly improve on using Gaussian ones, when the interference coefficient is above a given threshold, which significantly improves on the existing threshold (cf. [8]). We have also recovered the threshold of the moderate regime by using  $H_2$  perturbations in the synch setting, showing that this global threshold is reflected in our local setting. We also met “mysteriously” in our local setting the other global threshold found in [3], [13], [16], below which treating interference as noise with iid Gaussian inputs is optimal. It is worth noting that this two global thresholds (moderate regime and noisy interference) are recovered with our tool from a common analytic function. We hope to understand this better with a work in progress.

The Hermite technique provides not only counter-examples to the optimality of Gaussian inputs but it also gives insight on the competitive situations in Gaussian network problems. For

example, in the fading BC problem, the Hermite technique gives a condition on what kind of fading distributions and degradedness (values of  $v$ ) non-Gaussian inputs must be used. It also points out that the perturbation in  $H_3$  are most effective when carried in an opposite manner for the two users, so as to make the distribution of  $X$  close to Gaussian.

Finally, in a different context, local results could be “lifted” to corresponding global results in [2]. There, the localization is made with respect to the channels and not the input distribution, yet, it would be interesting to compare the local with the global behavior for the current problem too. The fact that we have observed some global results locally, as mentioned previously, gives hope for possible local to global extensions. A work in progress aims to use our tool beyond the local setting, in particular, by analyzing all sub-Gaussian distributions. Moreover, there are interesting connections between the results developed in this paper and the properties of the Ornstein-Uhlenbeck process. Indeed, some of these properties have already been used in [7] to solve the long standing entropy monotonicity conjecture, and we are currently investigating these relations from closer.

#### ACKNOWLEDGMENT

The authors would like to thank Tie Liu and Shlomo Shamai for pointing out problems relevant to the application of the proposed tool, Daniel Stroock for discussions on Hermite polynomials, and Emre Telatar for stimulating discussions and helpful comments.

#### REFERENCES

- [1] E. Abbe, “Global properties of the Hermite Coordinate System”, manuscript, 2009.
- [2] E. Abbe and L. Zheng, “Linear universal decoding: a local to global geometric approach”, Submitted IEEE Trans. Inform. Theory, 2008.
- [3] V. S. Annapureddy and V. V. Veeravalli, “Gaussian Interference Network: sum Capacity in the Low Interference Regime and New Outer Bounds on the Capacity Region”, Submitted IEEE Trans. Inform. Theory, 2008.
- [4] I. N. Andersen, “Sample-whitened matched filters”, IEEE Trans. Inform. Theory, vol. IT-19, no. 5, pp. 653-660, Sept. 1973.
- [5] D. Arnold and H.-A. Loeliger, “On the information rate of binary-input channels with memory”, Proc. 2001 IEEE Int. Conf. on Communications, Helsinki, Finland, pp. 2692-2695, June 11-14, 2001.
- [6] D. Arnold, H.-A. Loeliger, P. O. Vontobel, A. Kavcic, and W. Zeng, “Simulation-based computation of information rates for channels with memory”, IEEE Trans. on Inform. Theory, vol. 52, no. 8, pp. 3498-3508, Aug. 2006.
- [7] S. Artstein, K. Ball, F. Barthe and A. Naor, “Solution of Shannon’s Problem on the Monotonicity of Entropy”, Journal of the American Mathematical Society 17, 975-982 (2004).
- [8] E. Calvo, J. Fonollosa, and J. Vidal, “The Totally Asynchronous Interference Channel with Single-User Receivers”, ISIT 2009, Seoul.
- [9] M. Costa, “On the Gaussian interference channel,” IEEE Trans. Inform. Theory, vol. IT-31, pp. 607-615, Sept. 1985.
- [10] T. M. Cover, “Comments on broadcast channel”, IEEE Trans. on Inform. Theory vol. 44, n. 6, pp. 2524-2530, Oct. 1998.
- [11] T. M. Cover and J. A. Thomas, “Elements of Information Theory”, John Wiley & Sons, New York, NY, 1991.
- [12] G. D. Fomey, “Maximum-likelihood sequence estimation of digital sequences in the presence of intersymbol interference”, IEEE Trans. Inform. Theory, vol. IT-18, no. 3, pp. 363-378, May 1972.
- [13] A. S. Motahari, A. K. Khandani, “Capacity Bounds for the Gaussian Interference Channel”, CoRR abs/0801.1306; (2008).
- [14] S. Shamai (Shitz) and R. Laroia, “The intersymbol interference channel: Lower bounds on capacity and channel precoding loss”, IEEE Trans. Inform. Theory, vol. 42, pp. 1388-1404, Sept. 1996.
- [15] S. Shamai (Shitz), L.H. Ozarow, and A.D. Wyner, “Information rates for a discrete-time Gaussian channel with intersymbol interference and stationary inputs”, IEEE Trans. Inform. Theory, vol. 37, no. 6, pp. 1527-1539, Nov. 1991.
- [16] X. Shang, G. Kramer, B. Chen, “Outer bound and noisy-interference sum-rate capacity for symmetric Gaussian interference channels”, CISS 2008: 385-389.
- [17] A.J. Stam, “Some inequalities satisfied by the quantities of information of Fisher and Shannon”, Information and Control, 2: 101-112, 1959.
- [18] D. W. Stroock, “A Concise Introduction to the Theory of Integration”, Third Edition, Birkhäuser, 1999.
- [19] D. Tuninetti and S. Shamai, “On two-user fading Gaussian broadcast channels with perfect channel state information at the receivers”, Int. Symp. on Inform. Theory (ISIT), 2003.