# Insufficiency of Linear-Feedback Schemes in Gaussian Broadcast Channels with Common Message

Youlong Wu, Paolo Minero, and Michèle Wigger

*Abstract*—We consider the $K \geq 2$-user memoryless Gaussian broadcast channel (BC) with feedback and common message only. We show that linear-feedback schemes with a message point, in the spirit of Schalkwijk&Kailath's scheme for point-to-point channels or Ozarow&Leung's scheme for BCs with private messages, are strictly suboptimal for this setup. Even with perfect feedback, the largest rate achieved by these schemes is strictly smaller than capacity $C$ (which is the same with and without feedback). In the extreme case where the number of receivers $K \to \infty$, the largest rate achieved by linear-feedback schemes with a message point tends to 0.

To contrast this negative result, we describe a scheme for *rate-limited* feedback that uses the feedback in an intermittent way, i.e., the receivers send feedback signals only in few channel uses. This scheme achieves all rates $R$ up to capacity $C$ with an $L$-th order exponential decay of the probability of error if the feedback rate $R_{\mathrm{fb}}$ is at least $(L-1)R$ for some positive integer $L$.

*Index Terms*—Broadcast channel, channel capacity, feedback, reliability.

## I. INTRODUCTION

We consider the $K \geq 2$-user Gaussian broadcast channel (BC) where the transmitter sends a single common message to all receivers. This setup arises in, for instance, in wireless networks where a base station broadcasts control information to a set of users or in multicast applications. For this setup, even perfect feedback cannot increase capacity. Feedback can however reduce the minimum probability of error for a given blocklength.

It has been shown that perfect feedback allows to have a double-exponential decay of the probability of error in the blocklength in Gaussian point-to-point channels [1], [2] or other memoryless Gaussian networks such as the multiple-access channel (MAC) [3] and the BC with private messages [4]. These super-exponential decays of the probability of error are achieved by Schalkwijk&Kailath type schemes that first map the message(s) into real message point(s) and then send as their channel inputs linear combinations of

the message point(s) and the past feedback signals. We call such schemes *linear-feedback schemes with message points* or *linear-feedback schemes* for short. Such schemes are known to achieve the capacity of Gaussian point-to-point channels (memoryless or with memory) [1], [2] and the sum-capacity of the two-user memoryless Gaussian MAC [3]. For $K \geq 3$-user Gaussian MACs they are optimal among a large class of schemes [5], [6], and for Gaussian BCs with private messages, they achieve the largest sum-rates known to date [7], [8], [9], [10].

In this paper we show that linear-feedback schemes with a message point are strictly suboptimal for the $K$-user memoryless Gaussian BC with common message and fail to achieve capacity. As a consequence, for this setup, linear-feedback schemes also fail to achieve double-exponential decay of the probability of error for rates close to capacity. To our knowledge, this is the first example of a memoryless Gaussian network with perfect feedback, where linear-feedback schemes with message points are shown to be strictly suboptimal. In all previously studied networks with perfect feedback, they attained the optimal performance or the best so far performance. (In case of noisy feedback, they are known to perform badly even in the memoryless Gaussian point-to-point channel [11].)

In the asymptotic scenario of infinitely many receivers $K \to \infty$, the performance of linear-feedback schemes with a message point completely collapses: the largest rate that is achievable with these schemes tends to 0 as $K \to \infty$. This latter result holds under some mild assumptions regarding the variances of the noises experienced at the receivers, which are for example met when all the noise variances are equal. Notice that, in contrast, the capacity of the $K$-user Gaussian BC with common message does not tend to 0 as $K \to \infty$ when e.g., all the noise variances are equal. In this case, the capacity does not depend on $K$, because it is simply given by the point-to-point capacity to the receiver with the largest noise variance.

That the performance of linear-feedback schemes with a common message point degenerates with increasing number of users $K$ is intuitively explained as follows. At each time instant, the transmitter sends a linear combination of the message point and past noise symbols. Resending the noise symbols previously experienced at some Receiver $k$ can be beneficial for this Receiver $k$ because it allows it to mitigate the noise corrupting previous outputs. However, resending these noise symbols is of no benefit for all other Receivers $k' \neq k$ and

only harms them. Therefore, the more receivers there are, the more noise symbols the transmitter sends in each channel use that are useless for a given Receiver $k$. Our result hinges upon the independence of the noises at difference receivers. In the case of correlated noises a noise symbol can be beneficial to multiple receivers. In the extreme case where all noises are identical, for instance, the BC degenerates to a point-to-point channel and Schalkwijk&Kailath's scheme is capacity achieving.

For the memoryless Gaussian point-to-point channel [1] and MAC [4], the (sum-)capacity achieving linear-feedback schemes with message points transmit in each channel use a scaled version of the linear minimum mean square estimation (LMMSE) errors of the message points given the previous channel outputs. The same strategy is however strictly suboptimal—even among the class of linear-feedback schemes with message points—when sending private messages over a Gaussian BC [7]. It is unknown whether LMMSE estimates are optimal among linear-feedback schemes when sending a common message over the Gaussian BC.

In our proof that any linear-feedback scheme with a message point cannot achieve the capacity of the Gaussian BC with common message, the following proposition is key: For any sequence of linear-feedback schemes with a *common* message point that achieves rate $R > 0$, one can construct a sequence of linear-feedback schemes that achieves the rate tuple $R_1 = \ldots = R_K = R$ when sending $K$ *private* message points with a linear-feedback scheme. This proposition shows that the class of linear-feedback schemes with message points cannot take advantage of the fact that all the $K \geq 2$ receivers are interested in the same message.

To contrast the bad performance of linear-feedback schemes, we present a coding scheme that uses the feedback in a intermittent way (only in few time slots the receivers send feedback signals) and that achieves double-exponential decay of the probability of error for all rates up to capacity. In our scheme it suffices to have rate-limited feedback with feedback rate $R_{\rm fb}$ no smaller than the forward rate $R$. If the feedback rate $R_{\rm fb} < R$ then, even for the setup with only one receiver, the probability of error can decay only exponentially in the blocklength [12]. This implies immediately that also for the $K \geq 2$ receivers BC with common message no double-exponential decay in the probability of error is achievable when $R_{\rm fb} < R$. When the feedback rate $R_{\rm fb} > (L-1)R$, for some positive integer $L$, then our intermittent-feedback scheme can achieve an $L$-th order exponential decay in the probability of error. That means, it achieves a probability of error of the form $P_{\rm e}^{(n)} = \exp(-\exp(\exp(\ldots \exp(\Omega(n)))))$, where there are $L$ exponential terms and where $\Omega(n)$ denotes a function that satisfies $\underline{\lim}_{n\to\infty} \frac{\Omega(n)}{n} > 0$.

Our intermittent-feedback scheme is inspired by the scheme in [12] for the memoryless Gaussian point-to-point channel with rate-limited feedback. Also the schemes in [13] and [14] for the memoryless Gaussian point-to-point channel with perfect feedback are related. In fact, in our scheme communication takes place in $L$ phases. In the first phase, the transmitter uses a Gaussian code of power $P$ to send the common message to the $K$ Receivers. The transmission in

phase $l \in \{2, \ldots, L\}$ depends on the feedback signals. After each phases $l \in \{1, \ldots, L-1\}$ each Receiver $k$ feeds back a temporary guess of the message. Now, if one receiver's temporary guesses after phase $(l-1)$ is wrong, then in phase $l$ the transmitter resends the common message using a new code. If all receivers' temporary guesses after phase $(l-1)$ were correct, in phase $l$ the transmitter sends the all-zero sequence. In this latter case, no power is consumed in phase $l$. The receivers' final guess is their temporary guess after phase $L$.

The fact that the described scheme can achieve an $L$-th order decay of the probability of error, roughly follows from the following inductive argument. Assume that the probability of the event "one of the receivers' guesses is wrong after phase $l$", for $l \in \{1, \ldots, L-1\}$, has an $l$-th order exponential decay in the blocklength. Then, when sending the common message in phase $l+1$, the transmitter can use power that is $l$-th order exponentially large in the blocklength without violating the *expected* average blockpower constraint. With such a code, in turn, the probability that after phase $l+1$ one of the receivers has a wrong guess can have an $(l+1)$-th order exponential decay in the blocklength.

The rest of the paper is organized as follows. This section is concluded with some remarks on notation. Section II describes the Gaussian BC with common message and defines the class of linear-feedback schemes with a message point. Section III introduces the Gaussian BC with private messages and defines the class of linear-feedback schemes with private message points. Section IV presents our main results. Finally, Sections V and VI contain the proofs of our Theorems 1 and 2.

*Notation:* Let $\mathbb{R}$ denote the set of reals and $\mathbb{Z}^+$ the set of positive integers. Also, let $\mathcal{K}$ denote the discrete set $\mathcal{K} := \{1, \ldots, K\}$, for some $K \in \mathbb{Z}^+$. For a finite set $\mathcal{A}$, we denote by $|\mathcal{A}|$ its cardinality and by $\mathcal{A}^j$, for $j \in \mathbb{Z}^+$, its $j$-fold Cartesian product, $\mathcal{A}^j := \mathcal{A}_1 \times \ldots \times \mathcal{A}_j$.

We use capital letters to denote random variables and small letters for their realizations, e.g. $X$ and $x$. For $j \in \mathbb{Z}^+$, we use the short hand notations $X^j$ and $x^j$ for the tuples $X^j := (X_1, \ldots, X_j)$ and $x^j := (x_1, \ldots, x_j)$. Vectors are displayed in boldface, e.g., $\mathbf{X}$ and $\mathbf{x}$ for a random and deterministic vector. Further, $|\cdot|$ denotes the modulus operation for scalars and $\|\cdot\|$ the norm operation for vectors. For matrices we use the font $\mathsf{A}$, and we use $\|\mathsf{A}\|_F$ to denote its Frobenius norm.

The abbreviation i.i.d. stands for *independent and identically distributed*. All logarithms are taken with base $e$, i.e., $\log(\cdot)$ denotes the natural logarithm. We denote by $\mathcal{Q}(\cdot)$ the tail probability of the standard normal distribution. The operator $\circ$ is used to denote function composition.

We use the Landau symbols: $o(1)$ denotes any function that tends to 0 as $n \to \infty$.

## II. SETUP

### A. System Model and Capacity

We consider the $K \geq 2$-receiver Gaussian BC with common message and feedback depicted in Figure 1. Specifically, if $X_i$ denotes the transmitter's channel input at time-$i$, the time-$i$
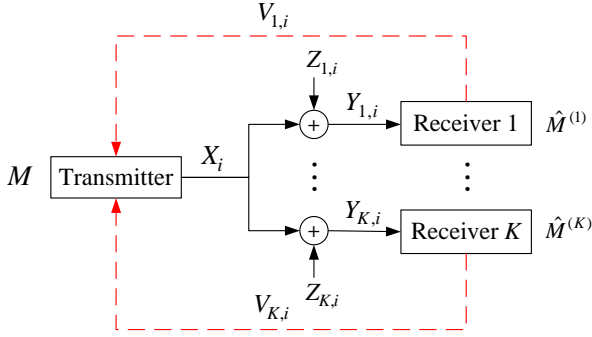
Fig. 1. $K$-receiver Gaussian Broadcast channel with feedback and common message only.

channel output at Receiver $k \in \mathcal{K}$ is

$$Y_{k,i} = X_i + Z_{k,i} \tag{1}$$

where $\{Z_{k,i}\}_{i=1}^n$ models the additive noise at Receiver $k$. The sequence of noises $\{(Z_{1,i}, \ldots, Z_{K,i})\}_{i=1}^n$ is a sequence of i.i.d. centered Gaussian vectors, each of diagonal covariance matrix

$$\mathsf{K}_z = \begin{pmatrix} \sigma_1^2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_K^2 \end{pmatrix}. \tag{2}$$

Without loss of generality, we assume that

$$\sigma_1^2 \geq \sigma_2^2 \geq \ldots \geq \sigma_K^2. \tag{3}$$

The transmitter wishes to convey a common message $M$ to all receivers, where $M$ is uniformly distributed over the message set $\mathcal{M} := \{1, ..., \lfloor e^{nR} \rfloor\}$ independent of the noise sequences $\{Z_{1,i}\}_{i=1}^n, \ldots, \{Z_{K,i}\}_{i=1}^n$. Here, $n$ denotes the blocklength and $R > 0$ the rate of transmission. It is assumed that the transmitter has either *rate-limited* or *perfect* feedback from all receivers. That means, after each channel use $i \in \{1, \ldots, n\}$, each Receiver $k \in \mathcal{K}$ feeds back a signal $V_{k,i} \in \mathcal{V}_{k,i}$ to the transmitter. The feedback alphabet $\mathcal{V}_{k,i}$ is a design parameter of the scheme. In the case of rate-limited feedback, the signals from Receiver $k$ have to satisfy:

$$\sum_{i=1}^n H(V_{k,i}) \leq nR_{\mathrm{fb}}, \quad k \in \mathcal{K} \tag{4}$$

where $R_{\mathrm{fb}}$ denotes the symmetric feedback rate. In the case of perfect feedback, we have no constraint on the feedback signals $\{V_{k,i}\}_{i=1}^n$, and it is thus optimal to choose $\mathcal{V}_{k,i} = \mathbb{R}$ and

$$V_{k,i} = Y_{k,i}, \tag{5}$$

because in this way any processing that can be done at the receivers can also be done at the transmitter.

An encoding strategy is comprised of a sequence of encoding functions $\{f_i^{(n)}\}_{i=1}^n$ of the form

$$f_i^{(n)} \colon \mathcal{M} \times \mathcal{V}_1^{i-1} \times \ldots \times \mathcal{V}_K^{i-1} \to \mathbb{R} \tag{6}$$

that is used to produce the channel inputs as

$$X_i = f_i^{(n)}(M, V_1^{i-1}, \ldots, V_K^{i-1}), \quad i \in \{1, \ldots, n\}. \tag{7}$$

We impose an expected average block-power constraint $P$ on the channel input sequence:

$$\frac{1}{n} \mathbb{E} \left[ \sum_{i=1}^n X_i^2 \right] \leq P. \tag{8}$$

Each Receiver $k \in \mathcal{K}$ decodes the message $M$ by means of a decoding function $g_k^{(n)}$ of the form

$$g_k^{(n)} \colon \mathbb{R}^n \to \mathcal{M}. \tag{9}$$

That means, Receiver $k$ produces as its guess

$$\hat{M}^{(k)} = g_k^{(n)}(Y_k^n). \tag{10}$$

An error occurs in the communication if

$$(\hat{M}^{(k)} \neq M), \tag{11}$$

for some $k \in \mathcal{K}$. Thus, the average probability of error is

$$P_e^{(n)} := \Pr \left[ \bigcup_{k \in \mathcal{K}} \left( \hat{M}^{(k)} \neq M \right) \right]. \tag{12}$$

We say that *a rate $R > 0$ is achievable* for the described setup if for every $\epsilon > 0$ there exists a sequence of encoding and decoding functions $\left\{ \{f_i^{(n)}\}_{i=1}^n, \{g_k^{(n)}\}_{k=1}^K \right\}_{n=1}^\infty$ as in (6) and (9) and satisfying the power constraint (8) such that for sufficiently large blocklengths $n$ the probability of error $P_e^{(n)} < \epsilon$. The supremum of all achievable rates is called the *capacity*. The capacity is the same in the case of perfect feedback, of rate-limited feedback (irrespective of the feedback rate $R_{\mathrm{fb}}$), and without feedback. We denote it by $C$ and by assumption (3) it is given by

$$C = \frac{1}{2} \log \left( 1 + \frac{P}{\sigma_1^2} \right). \tag{13}$$

Our main interest in this paper is in the speed of decay of the probability of error at rates $R < C$.

**Definition 1.** *Given a positive integer $L$, we say that the $L$-th order exponential decay in the probability of error is achievable at a given rate $R < C$, if there exists a sequence of schemes of rate $R$ such that their probabilities of error $\{P_e^{(n)}\}_{n=1}^\infty$ satisfy*

$$\varliminf_{n \to \infty} \frac{1}{n} \log \log \ldots \log(-\log P_e^{(n)}) > 0, \tag{14}$$

*where the number of logarithms in (14) is $L$.*

### B. Linear-Feedback Schemes with a Message Point

When considering perfect feedback, we will be interested in the class of coding schemes where the feedback is only used in a linear fashion. Specifically, we say that a scheme is a linear-feedback scheme with a message point, if the sequence of encoding functions $\{f_i^{(n)}\}_{i=1}^n$ is of the form

$$f_i^{(n)} = \Phi^{(n)} \circ L_i^{(n)} \tag{15}$$

with

$$\Phi^{(n)} \colon M \mapsto \Theta^{(n)} \in \mathbb{R} \tag{16a}$$

$$L_i^{(n)} \colon (\Theta^{(n)}, Y_1^{i-1}, \ldots, Y_K^{i-1}) \mapsto X_i \tag{16b}$$

where $\Phi^{(n)}$ is an arbitrary function on the respective domains and $L_i^{(n)}$ is a *linear mapping* on the respective domains. There is no constraint on the decoding functions $g_1^{(n)}, \ldots, g_K^{(n)}$.

By the definition of a linear-feedback coding scheme in (16), for each blocklength $n$, if we define $\mathbf{X} = (X_1, \ldots, X_n)^{\mathsf{T}}$, $\mathbf{Y}_k = (Y_{k,1}, \ldots, Y_{k,n})^{\mathsf{T}}$, and $\mathbf{Z}_k = (Z_{k,1}, \ldots, Z_{k,n})^{\mathsf{T}}$, for $k \in \mathcal{K}$, the channel inputs can be written as:

$$\mathbf{X} = \Theta^{(n)} \cdot \mathbf{d}^{(n)} + \sum_{k=1}^{K} \mathsf{A}_k^{(n)} \mathbf{Z}_k, \tag{17}$$

for some $n$-dimensional vector $\mathbf{d}^{(n)}$ and $n$-by-$n$ strictly lower-triangular matrices $\mathsf{A}_1^{(n)}, \ldots, \mathsf{A}_K^{(n)}$. (The lower-triangularity of $\mathsf{A}_1^{(n)}, \ldots, \mathsf{A}_K^{(n)}$ ensures that the feedback is used in a strictly causal fashion.) Thus, for a given blocklength $n$, a linear-feedback scheme is described by the tuple

$$\Phi^{(n)}, \mathbf{d}^{(n)}, \mathsf{A}_1^{(n)}, \ldots, \mathsf{A}_K^{(n)}, g_1^{(n)}, \ldots, g_K^{(n)}. \tag{18}$$

It satisfies the average block-power constraint (8) whenever

$$\sum_{k=1}^{K} \|\mathsf{A}_k^{(n)}\|_F^2 \sigma_k^2 + \|\mathbf{d}^{(n)}\|^2 \mathsf{E}\left[|\Theta^{(n)}|^2\right] \leq nP. \tag{19}$$

The supremum of all rates that are achievable with a sequence of linear-feedback schemes with a message point is denoted by $C^{(\mathrm{Lin})}$.

## III. FOR COMPARISON: SETUP WITH PRIVATE MESSAGES AND PERFECT FEEDBACK

### A. System Model and Capacity Region

For comparison, we also discuss the scenario where the transmitter wishes to communicate a private message $M_k$ to each Receiver $k \in \mathcal{K}$ over the Gaussian BC in Figure 1. The messages $M_1, \ldots, M_K$ are assumed independent of each other and of the noise sequences $\{Z_{1,i}\}_{i=1}^n, \ldots, \{Z_{K,i}\}_{i=1}^n$ and each $M_k$ is uniformly distributed over the set $\mathcal{M}_k := \{1, \ldots, \lfloor e^{nR_k} \rfloor\}$. For this setup we restrict attention to perfect feedback. Thus, here the channel inputs are produced as

$$X_i = f_{\mathrm{priv},i}^{(n)}(M_1, \ldots, M_K, Y_1^{i-1}, \ldots, Y_K^{i-1}), \ i \in \{1, \ldots, n\}. \tag{20}$$

Receiver $k$ produces the guess

$$\hat{M}_k = g_{\mathrm{priv},k}^{(n)}(Y_k^n) \tag{21}$$

where the sequence of decoding function $\{g_{\mathrm{priv},k}^{(n)}\}_{k=1}^K$ is of the form

$$g_{\mathrm{priv},k}^{(n)} \colon \mathbb{R}^n \to \{1, \ldots, \lfloor e^{nR_k} \rfloor\}, \tag{22}$$

A rate tuple $(R_1, \ldots, R_K)$ is said to be achievable if for every blocklength $n$ there exists a set of $n$ encoding functions as in (20) satisfying the power constraint (8) and a set of $K$ decoding functions as in (22) such that the probability of decoding error tends to 0 as the blocklength $n$ tends to infinity, i.e.,

$$\lim_{n \to \infty} \Pr\left[(M_1, \ldots, M_K) \neq (\hat{M}_1, \ldots, \hat{M}_K)\right] = 0.$$

The closure of the set of all achievable rate tuples $(R_1, \ldots, R_K)$ is called the *capacity region*. We denote it

$\mathcal{C}_{\mathrm{private}}$. This capacity region is unknown to date. (The sum-capacity in the high-SNR asymptotic regime is derived in [8].) Achievable regions were presented in [7], [8], [9]; the tightest known outer bound on capacity for $K = 2$ users was presented in [4] based on the idea of revealing one of the output sequences to the other receiver. This idea generalizes to $K \geq 2$ users, and leads to the following outer bound [5], [15]:

**Lemma 1.** *If the rate tuple* $(R_1, \ldots, R_K)$ *lies in* $\mathcal{C}_{\mathrm{private}}$, *then there exist coefficients* $\alpha_1, \ldots, \alpha_K$ *in the closed interval* $[0, 1]$ *such that for each* $k \in \mathcal{K}$,

$$R_k \leq \frac{1}{2} \log\left(1 + \frac{\alpha_k P}{(1 - \alpha_1 - \ldots - \alpha_k)P + N_k}\right) \tag{23}$$

*where*

$$N_k = \left(\sum_{k'=1}^{k} \frac{1}{\sigma_{k'}^2}\right)^{-1}, \qquad k \in \mathcal{K}. \tag{24}$$

*Proof:* Let a genie reveal each output sequence $Y_k^n$ to Receivers $k+1, \ldots, K$. The resulting BC is physically degraded, and thus its capacity is the same as without feedback [16] and known. Evaluating this capacity region readily gives the outer bound in the lemma. ∎

### B. Linear-Feedback Schemes with Message Points

A *linear-feedback scheme with message points* for this setup with independent messages consists of a sequence of $K$ decoding functions as in (22) and of a sequence of encoding functions $\{f_{\mathrm{priv},i}^{(n)}\}_{i=1}^n$ of the form

$$f_{\mathrm{priv},i}^{(n)} = \Phi_{\mathrm{priv}}^{(n)} \circ L_{\mathrm{priv},i}^{(n)} \tag{25}$$

with

$$\Phi_{\mathrm{priv}}^{(n)} \colon \begin{pmatrix} M_1 \\ \vdots \\ M_K \end{pmatrix} \mapsto \boldsymbol{\Theta} := \begin{pmatrix} \Theta_1 \\ \vdots \\ \Theta_K \end{pmatrix} \in \mathbb{R}^K \tag{26a}$$

$$L_{\mathrm{priv},i}^{(n)} \colon (\boldsymbol{\Theta}, Y_1^{i-1}, \ldots, Y_K^{i-1}) \mapsto X_i \tag{26b}$$

where $\Phi_{\mathrm{priv}}^{(n)}$ is an arbitrary function on the respective domains and $L_{\mathrm{priv},i}^{(n)}$ is a *linear mapping* on the respective domains.

We denote the closure of the set of rate tuples $(R_1, \ldots, R_K)$ that are achievable with a linear-feedback scheme with message points by $\mathcal{C}_{\mathrm{private}}^{(\mathrm{Lin})}$. This region is unknown to date.

## IV. MAIN RESULTS

The main question we wish to answer is whether for the Gaussian BC with common message a super-exponential decay in the probability of error is achievable for all rates $R < C$. We first show that the class of linear-feedback schemes with message point fails in achieving this goal even with perfect feedback, because it does not achieve capacity (Theorem 1 and Corollary 1). As the number of receivers $K$ increases, the largest rate that is achievable with linear-feedback schemes with a message point vanishes (Proposition 2). However, as we show then, a super-exponential decay in the probability of error is still possible by means of an intermittent feedback scheme similar to [12] (Theorem 2).

**Proposition 1.** *If a sequence of linear-feedback schemes with a message point achieves a* common *rate $R > 0$, then there exists a sequence of linear-feedback schemes with message points that achieves the* private *rates $(R, \ldots, R) \in \mathbb{R}^K$:*

$$0 < R \le C^{(\text{Lin})} \quad \Longrightarrow \quad (R, \ldots, R) \in \mathcal{C}^{(\text{Lin})}_{\text{private}}. \tag{27}$$

*Proof:* See Section V. ∎

Proposition 1 and the upper bound in Lemma 1 yield the following result:

**Theorem 1.** *We have:*

$$C^{(\text{Lin})} \le \frac{1}{2} \log \left( 1 + \frac{\alpha_1^\star P}{(1 - \alpha_1^\star)P + \sigma_1^2} \right) \tag{28}$$

*where $\alpha_1^\star$ lies in the open interval $(0, 1)$ and is such that there exist $\alpha_2^\star, \ldots, \alpha_K^\star \in (0, 1)$ that satisfy*

$$\alpha_1^\star + \alpha_2^\star + \ldots + \alpha_K^\star = 1 \tag{29a}$$

*and for $k \in \{2, \ldots, K\}$:*

$$\begin{aligned}
&\frac{1}{2} \log \left( 1 + \frac{\alpha_k^\star P}{(1 - \alpha_1^\star - \alpha_2^\star - \ldots - \alpha_k^\star)P + N_k} \right) \\
&= \frac{1}{2} \log \left( 1 + \frac{\alpha_1^\star P}{(1 - \alpha_1^\star)P + \sigma_1^2} \right)
\end{aligned} \tag{29b}$$

*where the noise variances $\{N_k\}_{k=1}^K$ are defined in (24).*

Since $\alpha_1^\star$ is strictly smaller than 1, irrespective of $K$ and the noise variances $\sigma_1^2, \ldots, \sigma_K^2$, we obtain the following corollary.

**Corollary 1.** *Linear-feedback schemes with a message point cannot achieve the capacity of the Gaussian BC with common message:*

$$C^{(\text{Lin})} < C \tag{30}$$

*where the inequality is strict.*

**Proposition 2.** *If the noise variances $\{\sigma_k^2\}_{k=1}^K$ are such that*

$$\sum_{k=1}^{\infty} N_k = \infty, \tag{31}$$

*then*

$$\lim_{K \to \infty} C^{(\text{Lin})} = 0. \tag{32}$$

*Proof:* See Appendix A. ∎

In Figure 2 we plot the upper bond on $C^{(\text{Lin})}$ shown in (28), Theorem 1, as a function of the number of receivers $K$, which have all the same noise variance $\sigma_1^2 = \ldots = \sigma_K^2 = 1$. As we observe, this upper bound, and thus also $C^{(\text{Lin})}$, tends to 0 as $K$ tends to infinity

**Theorem 2.** *For any positive rate $R < C$, if the feedback rate*

$$R_{\text{fb}} \ge (L - 1)R, \tag{33}$$

*for some positive integer $L$, then it is possible to achieve an $L$-th order exponential decay of the probability of error in the blocklength.*
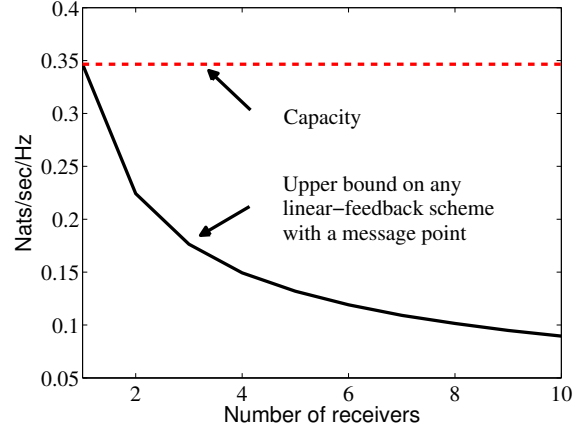
*Proof:* See Section VI. ∎



Fig. 2. Upper bound (28) on the rates achievable with linear-feedback schemes with a message point in function of the number of receivers $K$.

## V. Proof of Proposition 1

Let $\delta > 0$ be a small real number. Fix a sequence of rate-$R > 0$, power-$(P - \delta)$ linear-feedback schemes that sends a common message point over the Gaussian BC with probability of error $P_e^{(n)}$ tending to 0 as $n \to \infty$. For each $n \in \mathbb{Z}^+$, let

$$\Phi^{(n)}, \mathbf{d}^{(n)}, \mathsf{A}_1^{(n)}, \ldots, \mathsf{A}_K^{(n)}, g_1^{(n)}, \ldots, g_K^{(n)} \tag{34}$$

denote the parameters of the blocklength-$n$ scheme, which satisfy the power constraint

$$\mathsf{E}\left[|\Theta^{(n)}|^2\right] \cdot \|\mathbf{d}^{(n)}\|^2 + \sum_{k=1}^K \left\|\mathsf{A}_k^{(n)}\right\|_F^2 \le n(P - \delta) \tag{35}$$

where $\Theta^{(n)} = \Phi^{(n)}(M)$.

We have the following lemma.

**Lemma 2.** *For each blocklength $n$, there exist $n$-dimensional row-vectors $\mathbf{v}_1^{(n)}, \ldots, \mathbf{v}_K^{(n)}$ of unit norms,*

$$\|\mathbf{v}_1^{(n)}\|^2 = \cdots = \|\mathbf{v}_K^{(n)}\|^2 = 1, \tag{36}$$

*and $K$ indices $j_1^{(n)}, \ldots, j_K^{(n)} \in \{1, \ldots, n\}$ such that for each $k \in \mathcal{K}$ the following three limits holds:*

1)

$$R \le \varliminf_{n \to \infty} -\frac{1}{2n} \log c_k^{(n)} \tag{37}$$

*where*

$$c_k^{(n)} := \sigma_k^2 \left\|\mathbf{v}_k^{(n)}\big(\mathsf{I} + \mathsf{A}_k^{(n)}\big)\right\|^2 + \sum_{k' \in \mathcal{K} \setminus \{k\}} \sigma_{k'}^2 \left\|\mathbf{v}_{k'}^{(n)} \mathsf{A}_{k'}^{(n)}\right\|^2; \tag{38}$$

2)

$$\lim_{n \to \infty} \frac{1}{n} \mathsf{E}\left[\left(X_{j_k^{(n)}}^{(n)}\right)^2\right] = 0 \tag{39}$$

*where for $i \in \{1, \ldots, n\}$, $X_i^{(n)}$ denotes the $i$-th channel input of the blocklength-$n$ scheme; and*

3)

$$\lim_{n \to \infty} \frac{1}{2n} \log \left(|v_{k, j_k^{(n)}}^{(n)}|\right) = 0 \tag{40}$$

*where for $i \in \{1, \ldots, n\}$, $v_{k,i}^{(n)}$ denotes the i-th component of the vector $\mathbf{v}_k^{(n)}$.*

*Proof:* See Appendix B.  ∎

**Remark 1.** *In the statement of the above lemma, the vector $\mathbf{v}_k^{(n)}$ is a scaled version of the LMMSE filter of the the input given observations $Y_{k,1}, \ldots, Y_{k,n}$, and $c_k^{(n)}$ represents the volume of uncertainty about the message point at receiver k (hence R is bounded by its rate of decay). The last two claims of Lemma 2 hinge upon the fact that the channel input is power limited and therefore there must exists channel inputs that use less or equal than average power.*

In the following, let for each $n \in \mathbb{Z}^+$, $\mathbf{v}_1^{(n)}, \ldots, \mathbf{v}_K^{(n)}$ be $n$-dimensional unit-norm row-vectors and $j_1^{(n)}, \ldots, j_K^{(n)}$ be positive integers satisfying the limits (37), (39), and (40).

We now construct a sequence of linear-feedback schemes with message points that can send $K$ independent messages $M_1, \ldots, M_K$ to Receivers $1, \ldots, K$ at rates

$$R_k \geq \left( \varlimsup_{n \to \infty} -\frac{1}{2n} \log c_k^{(n)} \right) - \epsilon, \qquad k \in \mathcal{K}, \quad (41)$$

for an arbitrary small $\epsilon > 0$ with: 1) a probability of error that tends to 0 as the blocklength tends to infinity and 2) with an average blockpower that is no larger than $P$ when the blocklength is sufficiently large. By (37), since $\delta, \epsilon > 0$ can be chosen arbitrary small, and since $C^{(\text{Lin})}$ is continuous in the power $P$ (Remark 2 ahead) and is defined as a supremum, the result in Proposition 1 will follow.

We describe our scheme for blocklength-$(n + 2K)$, for some fixed $n \in \mathbb{Z}^+$. Our scheme is based on the parameters $\mathsf{A}_1^{(n)}, \ldots, \mathsf{A}_K^{(n)}$ in (34), on the vectors $\mathbf{v}_1^{(n)}, \ldots, \mathbf{v}_K^{(n)}$, and on the indices $j_1^{(n)}, \ldots, j_K^{(n)}$ where $\mathbf{v}_k^{(n)}$ and $j_k^{(n)}$, $k \in \mathcal{K}$ are defined in Lemma 2. For ease of notation, when describing our scheme in the following, we drop the superscript $(n)$, i.e., we write

$$\mathsf{A}_1, \ldots, \mathsf{A}_K, \mathbf{v}_1, \ldots, \mathbf{v}_K, \quad \text{and} \quad j_1, \ldots, j_K.$$

We also assume that

$$j_1 \leq j_2 \leq \ldots \leq j_K. \quad (42)$$

(If this is not the case, we simply relabel the receivers.) Also, to further simplify the description of the linear-feedback coding and the decoding, we rename the $n + 2K$ transmission slots as depicted in Figure 3. Transmission starts at slot $1 - K$ and ends at slot $n$; also, after each slot $j_k$, for $k \in \mathcal{K}$, we introduce an additional slot $\tilde{j}_k$. We call the slots $1 - K, \ldots, 0$ the *initialization slots*, the slots $\tilde{j}_1, \ldots, \tilde{j}_K$ the *extra slots*, and the remaining slots $1, 2, 3, \ldots, n$ the *regular slots*.

In our scheme, the message points $\{\Theta_k\}_{k=1}^K$ are constructed as in the Ozarow-Leung scheme [4]:

$$\Theta_k := 1/2 - \frac{M_k - 1}{\lfloor e^{(n+2K)R_k} \rfloor}, \qquad k \in \mathcal{K}. \quad (43)$$

These messages are sent during the initialization phase. Specifically, in the initialization slots $i = 1 - K, \ldots, 0$, the trans-

mitter sends the $K$ message points $\Theta_1, \ldots, \Theta_K$:

$$X_{1-k} = \sqrt{\frac{P}{\mathsf{Var}(\Theta_k)}} \Theta_k, \qquad k \in \mathcal{K}. \quad (44)$$

In the regular slots $i = 1, \ldots, n$, the transmitter sends the same inputs as in the scheme with common message described by the parameters in (34), but without the component from the message point and where for each $k \in \mathcal{K}$ the noise sample $Z_{k,j_k}$ is replaced by $Z_{k,\tilde{j}_k}$. Thus, defining the $n$-length vector of regular inputs $\mathbf{X} \triangleq (X_1, X_2, X_3, \ldots, X_n)^\mathsf{T}$, we have

$$\mathbf{X} = \sum_{k=1}^K \mathsf{A}_k \tilde{\mathbf{Z}}_k \quad (45)$$

where for $k \in \mathcal{K}$,

$$\tilde{\mathbf{Z}}_k := (Z_{k,1}, Z_{k,2}, \ldots, Z_{k,j_k-1}, Z_{k,\tilde{j}_k}, Z_{k,j_k+1}, \ldots, Z_{k,n})^\mathsf{T} \quad (46)$$

denotes the $n$-length noise vector experienced at Receiver $k$ during the regular slots $1, \ldots, j_k - 1$, the extra slot $\tilde{j}_k$, and the regular slots $j_k + 1, \ldots, n$.

Since for each $k \in \mathcal{K}$, the extra slot $\tilde{j}_k$ precedes all regular slots $j_k + 1, \ldots, n$, the strict lower-triangularity of the matrices $\mathsf{A}_1, \ldots, \mathsf{A}_K$ ensures that in (45) the feedback is used in a strictly causal way.

In each extra slot $\tilde{j}_k$, for $k \in \mathcal{K}$, the transmitter sends the regular input $X_{j_k}$, but now with the noise sample $Z_{k,1-k}$,

$$X_{\tilde{j}_k} = X_{j_k} + Z_{k,1-k}. \quad (47)$$

The noise sample $Z_{k,1-k}$ is of interest to Receiver $k$ (and only to Receiver $k$) because from this noise sample and $Y_{k,1-k}$ one can recover $\Theta_k$, see (44). Therefore—as described shortly—in the decoding, Receiver $k$ considers the extra output $Y_{k,\tilde{j}_k}$ which contains $Z_{k,1-k}$ whereas all other receivers $k' \neq k$ instead consider the regular outputs $Y_{k',j_k}$ which do not have the $Z_{k,1-k}$-component.

The decoding is similar as in the Ozarow-Leung scheme. However, here, each Receiver $k \in \mathcal{K}$ only considers the initialization output $Y_{k,1-k}$, the regular outputs $Y_{k,1}, \ldots, Y_{k,j_k-1}, Y_{k,j_k+1}, \ldots, Y_{k,K}$ and the extra output $Y_{k,\tilde{j}_k}$, see also Figure 4. Specifically, Receiver $k$ forms the $n$-length vector

$$\tilde{\mathbf{Y}}_k := (Y_{k,1}, \ldots, Y_{k,j_k-1}, Y_{k,\tilde{j}_k}, Y_{k,j_k+1}, \ldots, Y_{k,n})^\mathsf{T}, \quad (48)$$

and produces the LMMSE estimate $\hat{Z}_{k,1-k}$ of the noise $Z_{k,1-k}$ based on the vector $\tilde{\mathbf{Y}}_k$. It then forms

$$\hat{\Theta}_k = \sqrt{\frac{\mathsf{Var}(\Theta_k)}{P}} \left( Y_{k,1-k} - \hat{Z}_{k,1-k} \right). \quad (49)$$

and performs nearest neighbor decoding to decode its desired Message $M_k$ based on $\hat{\Theta}_k$.

We now analyze the described scheme. The expected block-power of our scheme is:

$$\sum_{i=1-K}^0 \mathsf{E}\left[|X_i|^2\right] + \sum_{i=1}^n \mathsf{E}\left[|X_i|^2\right] + \sum_{k=1}^K \mathsf{E}\left[|X_{\tilde{j}_k}|\right]$$

$$\leq KP + n(P - \delta) + \sum_{k=1}^K \mathsf{E}\left[|X_{j_k}|^2\right] + \sum_{k=1}^K \sigma_k^2 \quad (50)$$
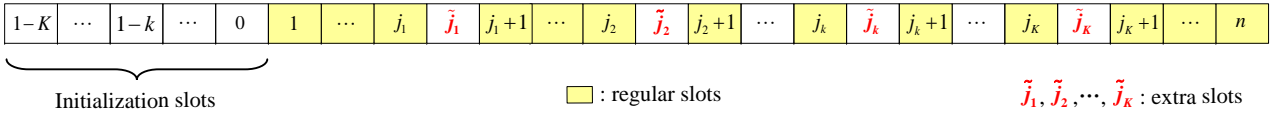
Fig. 3. Labeling of the transmission slots for our blocklength-$(n + 2K)$ scheme.
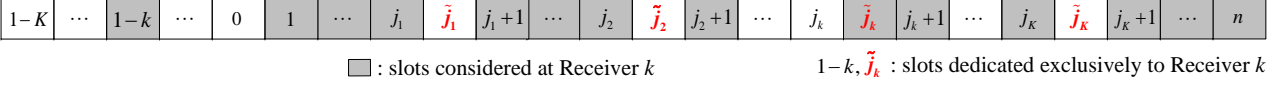


Fig. 4. Transmissions considered at Receiver $k$ and transmissions dedicated exclusively to Receiver $k$.

where the inequality follows from (44), (45), and (47), and from (35), which assures that the regular inputs $X_1, \ldots, X_n$ are block-power constrained to $n(P - \delta)$. Further, since the indices $j_1, \ldots, j_K$ satisfy Assumption (39),

$$\lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^{K} \mathsf{E}\left[|X_{j_k}|^2\right] = 0, \qquad (51)$$

and thus for sufficiently large $n$ the proposed scheme for independent messages is average blockpower constrained to $P$.

We analyze the probability of error. Notice that

$$\hat{\Theta}_k = \Theta_k + E_k \qquad (52)$$

where

$$E_k := \sqrt{\frac{\mathrm{Var}(\Theta_k)}{P}} \left(Z_{k,1-k} - \hat{Z}_{k,1-k}\right) \qquad (53)$$

is zero-mean Gaussian of variance

$$\mathsf{Var}(E_k) = \frac{\mathsf{Var}(\Theta_k)}{P} \sigma_k^2 e^{-2I\left(Z_{k,1-k}; \tilde{\mathbf{Y}}_k\right)}. \qquad (54)$$

Equation (54) is justified by

$$
\begin{aligned}
I\left(Z_{k,1-k} \; ; \; \tilde{\mathbf{Y}}_k\right) \\
= h(Z_{k,1-k}) - h\left(Z_{k,1-k} \big| \tilde{\mathbf{Y}}_k\right) \\
= \frac{1}{2} \log \left(\frac{\sigma_k^2}{\mathsf{Var}\left(Z_{k,1-k} - \hat{Z}_{k,1-k}\right)}\right) \qquad (55)
\end{aligned}
$$

where the last equality follows because $Z_{k,1-k}$ and $\tilde{\mathbf{Y}}_k$ are jointly Gaussian, and thus the LMMSE estimation error $Z_{k,1-k} - \hat{Z}_{k,1-k}$ is independent of the observations $\tilde{\mathbf{Y}}_k$.

The nearest neighbor decoding rule is successful if $|E_k|$ is smaller than half the distance between any two message points. Since $E_k$ is Gaussian and independent of the message point, the probability of this happening is

$$
\begin{aligned}
\Pr\left[\hat{M}_k \neq M_k\right] &\leq \Pr\left[|E_k| \geq \frac{1}{2 \cdot \lfloor e^{(n+2K)R_k}\rfloor}\right] \\
&= 2\mathcal{Q}\left(\frac{e^{I(Z_{k,1-k}; \tilde{\mathbf{Y}}_k)}}{2 \cdot \lfloor e^{(n+2K)R_k}\rfloor} \cdot \frac{P}{\mathsf{Var}(\Theta_k)\sigma_k^2}\right).
\end{aligned}
$$

We conclude that the probability of error tends to 0, double-exponentially, whenever

$$R_k < \lim_{n \to \infty} \frac{1}{n} I\left(Z_{k,1-k}; \tilde{\mathbf{Y}}_k\right). \qquad (56)$$

Notice that the vector $\tilde{\mathbf{Y}}_k$ as defined in (48), satisfies

$$\tilde{\mathbf{Y}}_k = \sum_{k' \in \mathcal{K} \setminus \{k\}} \mathsf{A}_{k'} \tilde{\mathbf{Z}}_{k'} + (\mathsf{I} + \mathsf{A}_k)\tilde{\mathbf{Z}}_k + \mathbf{e}_{j_k} Z_{k,1-k} \qquad (57)$$

where for each $i \in \{1, \ldots, n\}$ the vector $\mathbf{e}_i$ is the $n$-length unit-norm vector with all zero entries except at position $i$ where the entry is 1. Thus, by the data processing inequality,

$$
\begin{aligned}
&I(Z_{k,1-k}; \tilde{\mathbf{Y}}_k) \\
&\geq I(Z_{k,1-k}; \mathbf{v}_k^\mathsf{T} \tilde{\mathbf{Y}}_k) \\
&= \frac{1}{2} \log \left(1 + \frac{|v_{k,j_k}|^2}{\sigma_k^2 \|\mathbf{v}_k(\mathsf{I} + \mathsf{A}_k)\|^2 + \sum_{k' \in \mathcal{K} \setminus \{k\}} \sigma_{k'}^2 \|\mathbf{v}_{k'} \mathsf{A}_{k'}\|^2}\right) \\
&= \frac{1}{2} \log \left(1 + \frac{|v_{k,j_k}|^2}{c_k}\right) \qquad (58)
\end{aligned}
$$

where the first equality follows by (57) and the joint Gaussianity of all involved random variables and the second equality follows by the definition of $c_k$ in (38).

Combining (56) and (58), we obtain that the probability $\Pr\left[\hat{M}_k \neq M_k\right]$ tends to 0 as $n \to \infty$ whenever

$$R_k < \lim_{n \to \infty} \frac{1}{2n} \log \left(1 + \frac{|v_{k,j_k}|^2}{c_k}\right). \qquad (59)$$

(Recall that the quantities $j_k$, $c_k$, and $v_{k,j_k}$ depend on $n$, but here we do not show this dependence for readability.)

Further, by the converse in (37),

$$
\begin{aligned}
0 < R &\leq \lim_{n \to \infty} \frac{-1}{2n} \log c_k \\
&= \lim_{n \to \infty} \frac{1}{2n} \log \frac{|v_{k,j_k}|^2}{c_k} \qquad (60) \\
&= \lim_{n \to \infty} \frac{1}{2n} \log \left(1 + \frac{|v_{k,j_k}|^2}{c_k}\right) \qquad (61)
\end{aligned}
$$

where the first equality holds by Condition (40) and the second equality holds because (60) implies that the ratio $\frac{|v_{k,j_k}|^2}{c_k}$ tends to infinity with $n$.

Combining (59) with (61) establishes that for arbitrary $\epsilon > 0$ there exists a rate tuple $(R_1, \ldots, R_K)$ satisfying (41) such that the described scheme with independent messages achieves probability of error that tends to 0 as the blocklength tends to infinity.

**Remark 2.** *In the spirit of the scheme for private messages described above, one can construct a linear-feedback scheme*

*with a common message point that has arbitrary small probability of error whenever*

$$R < \varlimsup_{n \to \infty} -\frac{1}{2n} \log c_k, \qquad k \in \mathcal{K}.$$

*Combined with the converse in (37), this gives a (multiletter) characterization of $C^{(\mathrm{Lin})}$. Based on this multi-letter characterization one can show the continuity of $C^{(\mathrm{Lin})}$ in the transmit-power constraint $P$.*

## VI. PROOF OF THEOREM 2: CODING SCHEME ACHIEVING $L$-TH ORDER EXPONENTIAL DECAY

The scheme is based on the scheme in [12], see also [13], [14]. Fix a positive rate $R < C$ and a positive integer $L$. Assume that

$$R_{\mathrm{fb}} \geq R(L-1). \tag{62}$$

Also, fix a large blocklength $n$ and small numbers $\epsilon, \delta > 0$ such that

$$R < C(1-\delta) \tag{63}$$

and

$$(1-\epsilon)^{-1} < 1 + \delta. \tag{64}$$

Define

$$n_1 := (1-\epsilon)n \tag{65}$$

and for $l \in \{2, \dots, L\}$

$$n_l := n_1 + \frac{\epsilon n}{L-1}(l-1). \tag{66}$$

Notice that by (64) and (65),

$$\frac{n}{n_1} < 1 + \delta. \tag{67}$$

The coding scheme takes place in $L$ phases. After each phase $l \in \{1, \dots, L\}$, each Receiver $k \in \mathcal{K}$ makes a temporary guess $\hat{M}_l^{(k)}$ of message $M$. The final guess is the guess after phase $L$:

$$\hat{M}^{(k)} = \hat{M}_L^{(k)}, \tag{68}$$

Define the probability of error after phase $l \in \{1, \dots, L\}$:

$$P_{e,l}^{(n)} := \Pr\left[\bigcup_{k \in \mathcal{K}} \hat{M}_l^{(k)} \neq M\right] \tag{69}$$

and thus

$$P_e^{(n)} = P_{e,L}^{(n)}. \tag{70}$$

### A. Code Construction

We construct a codebook $\mathcal{C}_1$ that

- is of blocklength $n_1$,
- is of rate $R_{\mathrm{phase},1} = \frac{n}{n_1}R$,
- satisfies an expected average block-power constraint $P$, and
- when used to send a common message over the Gaussian BC in (1) and combined with an optimal decoding rule, it achieves probability of error $\rho_1$ not exceeding

$$\rho_1 \leq e^{-n(\zeta - o(1))} \tag{71}$$

for some $\zeta > 0$.

Notice that such a code exists because, by (63) and (67), the rate of the code $\frac{n}{n_1}R < C(1-\delta^2)$, and because the error exponent of the BC with common message without feedback is positive for all rates below capacity.[1]

Let

$$\gamma_1 := \rho_1. \tag{72}$$

For $l$ from 2 to $L$, do the following.

Construct a codebook $\mathcal{C}_l$ that:

- is of blocklength $\frac{\epsilon n}{L-1} - 1$,
- is of rate $R_{\mathrm{phase},l} := \frac{R(L-1)}{\epsilon - (L-1)/n}$,
- satisfies an expected average block-power constraint $P/\gamma_{l-1}$,
- when used to send a common message over the Gaussian BC in (1) and combined with an optimal decoding rule, it achieves probability of error $\rho_l$ not exceeding

$$\rho_l \leq \exp(-\underbrace{\exp \circ \dots \circ \exp}_{l-1 \text{ times}}(\Omega(n))). \tag{73}$$

Define

$$\gamma_l := \rho_l + 2 \sum_{k \in \mathcal{K}} \mathcal{Q}\left(\frac{\sqrt{P/\gamma_{l-1}}}{2\sigma_k}\right). \tag{74}$$

(As shown in Section VI-C ahead, $\gamma_l$ upper bounds $P_{e,l}^{(n)}$ defined in (69).) By (73) and (74), inductively one can show that

$$\gamma_l \leq \exp(-\underbrace{\exp \circ \dots \circ \exp}_{l-1 \text{ times}}(\Omega(n))). \tag{75}$$

In Appendix C, we prove that such codes $\mathcal{C}_2, \dots, \mathcal{C}_L$ exist.

### B. Transmission

Transmission takes place in $L$ phases.

*1) First phase with channel uses $i = 1, \dots, n_1$:* During the first $n_1$ channel uses, the transmitter sends the codeword in $\mathcal{C}_1$ corresponding to message $M$.

After observing the channel outputs $Y_k^{n_1}$, Receiver $k \in \mathcal{K}$ makes a temporary decision $\hat{M}_1^{(k)}$ about $M$. It then sends this temporary decision $\hat{M}_1^{(k)}$ to the transmitter over the feedback channel:

$$V_{k,n_1} = \hat{M}_1^{(k)}. \tag{76}$$

All previous feedback signals from Receiver $k$ are deterministically 0.

*2) Phase $l \in \{2, \dots, L\}$ with channel uses $i \in \{n_{l-1}+1, \dots, n_l\}$:* The communication in phase $l$ depends on the receivers' temporary decisions $\hat{M}_{l-1}^{(1)}, \dots, \hat{M}_{l-1}^{(K)}$ after the previous phase $(l-1)$. These decisions have been communicated to the transmitter over the respective feedback links.

If in phase $(l-1)$ at least one of the receivers made an incorrect decision,

$$(\hat{M}_{l-1}^{(k)} \neq M), \qquad \text{for some } k \in \mathcal{K}, \tag{77}$$

[1]The positiveness of the error exponent for the Gaussian BC with common message and without feedback follows from the fact that without feedback the probability of error for the Gaussian BC with common messages is at most $K$ times the probability of error to the weakest receiver.

then in channel use $n_{l-1} + 1$ the transmitter sends an error signal to indicate an error:

$$X_{n_l+1} = \sqrt{P/\gamma_{l-1}}. \tag{78}$$

During the remaining channel uses $i = n_{l-1}+2, \ldots, n_l$ it then retransmits the message $M$ by sending the codeword from $\mathcal{C}_l$ that corresponds to $M$.

On the other hand, if all receivers' temporary decisions to the phase $(l-1)$ were correct,

$$\hat{M}_{l-1}^{(1)} = \hat{M}_{l-1}^{(2)} = \ldots = \hat{M}_{l-1}^{(K)} = M, \tag{79}$$

then the transmitter sends 0 during the entire phase $l$:

$$X_i = 0, \qquad i = n_{l-1}+1, \ldots, n_l. \tag{80}$$

In this case, no power is consumed in phase $l$.

The receivers first detect whether the transmitter sent an error signal in channel use $n_{l-1} + 1$. Depending on the output of this detection, they either stick to their temporary decision in phase $(l-1)$ or make a new decision based on the transmissions in phase $l$. Specifically, if

$$Y_{k,n_{l-1}+1} < T_{l-1} \tag{81}$$

where

$$T_{l-1} := \frac{\sqrt{P/\gamma_{l-1}}}{2}, \tag{82}$$

then Receiver $k \in \mathcal{K}$ decides that its decision $\hat{M}_{l-1}^{(k)}$ in phase $(l-1)$ was correct and keeps it as its temporary guess of the message $M$:

$$\hat{M}_l^{(k)} = \hat{M}_{l-1}^{(k)}. \tag{83}$$

If instead,

$$Y_{k,n_{l-1}+1} \geq T_{l-1}, \tag{84}$$

Receiver $k$ decides that its temporary decision $\hat{M}_{l-1}^{(k)}$ was wrong and discards it. It then produces a new guess $\hat{M}_l^{(k)}$ by decoding the code $\mathcal{C}_l$ based on the outputs $Y_{k,n_{l-1}+2}, \ldots, Y_{k,n_l}$.

After each phase $l \in \{2, \ldots, L-1\}$, each Receiver $k \in \mathcal{K}$ feeds back to the transmitter its temporary guess $\hat{M}_l^{(k)}$:

$$V_{k,n_l} = \hat{M}_l^{(k)}. \tag{85}$$

All other feedback signals $V_{k,n_{l-1}+1}, \ldots, V_{k,n_l-1}$ in phase $l$ are deterministically 0.

After $L$ transmission phases, Receiver $k$'s final guess is

$$\hat{M}^{(k)} = \hat{M}_L^{(k)}. \tag{86}$$

Thus, an error occurs in the communication if

$$(\hat{M}_L^{(k)} \neq M), \quad \text{for some } k \in \mathcal{K}. \tag{87}$$

## C. Analysis

In view of (62), by (76) and (85), and because all other feedback signals are deterministically 0, our scheme satisfies the feedback rate constraint in (4).

We next analyze the probability of error and we bound the consumed power. These analysis rely on the following events. For each $k \in \mathcal{K}$ and $l \in \{1, \ldots, L\}$ define the events:

- $\mathcal{E}_l^{(k)}$: Receiver $k$'s decision in phase $l$ is wrong:

$$\hat{M}_l^{(k)} \neq M; \tag{88}$$

- $\mathcal{E}_{T,l}^{(k)}$: Receiver $k$ observes

$$Y_{k,n_l+1} < T_l; \tag{89}$$

- $\mathcal{E}_{\rho,l}^{(k)}$: Decoding Message $M$ based on Receiver $k$'s phase-$l$ outputs $Y_{k,n_{l-1}+2}, \ldots, Y_{k,n_l}$ using codebook $\mathcal{C}_l$ results in an error.

Define also the events:

$\mathcal{E}_{1,l}$: All receivers' decisions in phase $(l-1)$ are correct, and at least one Receiver $k \in \mathcal{K}$ obtains an error signal in channel use $n_{l-1} + 1$ :

$$\left( \bigcap_{k \in \mathcal{K}} (\mathcal{E}_{l-1}^{(k)})^c \right) \cap \left( \bigcup_{k \in \mathcal{K}} (\mathcal{E}_{T,l-1}^{(k)})^c \right). \tag{90}$$

$\mathcal{E}_{2,l}$: At least one Receiver $k \in \mathcal{K}$ makes an incorrect decision in phase $(l-1)$ but obtains no error signal in channel use $n_{l-1} + 1$:

$$\bigcup_{k \in \mathcal{K}} \left( \mathcal{E}_{l-1}^{(k)} \cap \mathcal{E}_{T,l-1}^{(k)} \right). \tag{91}$$

$\mathcal{E}_{3,l}$: At least one Receiver $k \in \mathcal{K}$ makes an incorrect temporary decision in phase $(l-1)$, and at least one Receiver $k' \in \mathcal{K}$ observes $Y_{k',n_{l-1}+1} \geq T_{l-1}$ and errs when decoding $M$ based on its phase-$l$ outputs $Y_{k',n_{l-1}+2}, \ldots, Y_{k',n_l}$:

$$\left( \bigcup_{k \in \mathcal{K}} \mathcal{E}_{l-1}^{(k)} \right) \cap \left( \bigcup_{k' \in \mathcal{K}} \left( (\mathcal{E}_{T,l}^{(k')})^c \cap \mathcal{E}_{\rho,l}^{(k')} \right) \right). \tag{92}$$

For each $l \in \{1, \ldots, L\}$, the probability $P_{e,l}^{(n)}$ is included in the union of the events $(\mathcal{E}_{1,l} \cup \mathcal{E}_{2,l} \cup \mathcal{E}_{3,l})$, and thus, by the union bound,

$$P_{e,l}^{(n)} \leq \Pr[\mathcal{E}_{1,l}] + \Pr[\mathcal{E}_{2,l}] + \Pr[\mathcal{E}_{3,l}]. \tag{93}$$

In particular, by (70) and (93), the probability of error of our scheme

$$P_e^{(n)} \leq \Pr[\mathcal{E}_{1,L}] + \Pr[\mathcal{E}_{2,L}] + \Pr[\mathcal{E}_{3,L}]. \tag{94}$$

We bound each summand in (94) individually, starting with

$\Pr[\mathcal{E}_{1,L}]$. By (90), we have

$$\begin{aligned}
\Pr[\mathcal{E}_{1,L}] &= \Pr\left[\left(\bigcap_{k\in\mathcal{K}}(\mathcal{E}_{L-1}^{(k)})^c\right)\cap\left(\bigcup_{k\in\mathcal{K}}(\mathcal{E}_{T,L-1}^{(k)})^c\right)\right] \\
&\leq \Pr\left[\bigcup_{k\in\mathcal{K}}(\mathcal{E}_{T,L-1}^{(k)})^c\Big|\bigcap_{k\in\mathcal{K}}(\mathcal{E}_{L-1}^{(k)})^c\right] \\
&\leq \sum_{k=1}^{K}\Pr\left[(\mathcal{E}_{T,L-1}^{(k)})^c\Big|\bigcap_{k\in\mathcal{K}}(\mathcal{E}_{L-1}^{(k)})^c\right] \\
&= \sum_{k=1}^{K}\mathcal{Q}\left(\frac{T_{L-1}}{\sigma_k}\right)
\end{aligned}$$

(95)

where the first inequality follows by Bayes' rule and because a probability cannot exceed 1; the second inequality by the union bound; and the last equality because in the event $\left(\bigcap_{k\in\mathcal{K}}(\mathcal{E}_{L-1}^{(k)})^c\right)$, we have $X_{n_{L-1}+1} = 0$ and thus $Y_{k,n_{L-1}+1} \sim \mathcal{N}(0,\sigma_k^2)$.

Next, by (91) and similar arguments as before, we obtain,

$$\begin{aligned}
\Pr[\mathcal{E}_{2,L}] &= \Pr\left[\bigcup_{k=\in\mathcal{K}}\left(\mathcal{E}_{L-1}^{(k)}\cap\mathcal{E}_{T,L-1}^{(k)}\right)\right] \\
&\leq \sum_{k=1}^{K}\Pr\left[\mathcal{E}_{L-1}^{(k)}\cap\mathcal{E}_{T,L-1}^{(k)}\right] \\
&\leq \sum_{k=1}^{K}\Pr\left[\mathcal{E}_{T,L-1}^{(k)}|\mathcal{E}_{L-1}^{(k)}\right] \\
&= \sum_{k=1}^{K}\mathcal{Q}\left(\frac{T_{L-1}}{\sigma_k}\right).
\end{aligned}$$

(96)

Finally, by (92) and similar arguments as before,

$$\begin{aligned}
\Pr[\mathcal{E}_{3,L}] &= \Pr\left[\left(\bigcup_{k\in\mathcal{K}}\mathcal{E}_{L-1}^{(k)}\right)\cap\left(\bigcup_{k'\in\mathcal{K}}\left((\mathcal{E}_{T,L}^{(k')})^c\cap\mathcal{E}_{\rho,L}^{(k')}\right)\right)\right] \\
&\leq \Pr\left[\bigcup_{k'\in\mathcal{K}}\left((\mathcal{E}_{T,L}^{(k')})^c\cap\mathcal{E}_{\rho,L}^{(k')}\right)\Big|\bigcup_{k\in\mathcal{K}}\mathcal{E}_{L-1}^{(k)}\right] \\
&\leq \Pr\left[\bigcup_{k'\in\mathcal{K}}\mathcal{E}_{\rho,L}^{(k')}\Big|\bigcup_{k\in\mathcal{K}}\mathcal{E}_{L-1}^{(k)}\right] \\
&\leq \rho_L
\end{aligned}$$

(97)

where the last inequality follows by the definition of $\rho_L$.

In view of (82) and (94)–(97),

$$\begin{aligned}
P_e^{(n)} &\leq \Pr[\mathcal{E}_{1,L}] + \Pr[\mathcal{E}_{2,L}] + \Pr[\mathcal{E}_{3,L}] \\
&\leq \rho_L + 2\sum_{k\in\mathcal{K}}\mathcal{Q}\left(\frac{\sqrt{P/\gamma_{L-1}}}{2\sigma_k}\right) \\
&= \gamma_L
\end{aligned}$$

(98)

where the equality follows by the definition of $\gamma_L$ in (74). Combining this with the $L$-th order exponential decay of $\gamma_L$, see (75), we obtain

$$\lim_{n\to\infty} -\frac{1}{n}\underbrace{\log\log\ldots\log}_{L-1\text{ times}}(-\log P_e^{(n)}) > 0,$$

(99)

Now consider the consumed expected average block-power. Similarly to (98), we can show that for $l\in\{1,\ldots,L-1\}$,

$$P_{e,l}^{(n)} \leq \gamma_l.$$

(100)

Since in each phase $l\in\{2,\ldots,L\}$ we consume power $P/\gamma_{l-1}$ in the event (77) and power 0 in the event (79), by the definition in (69),

$$\frac{1}{n}\mathsf{E}\left[\sum_{i=1}^{n}X_i^2\right] \leq \frac{1}{n}\Big(P(1-\epsilon)n+\sum_{l=2}^{L}P_{e,l-1}^{(n)}\frac{P}{\gamma_{l-1}}\frac{\epsilon n}{L-1}\Big) \leq P$$

(101)

where the second inequality follows from (100).

This completes the proof of Theorem 2.

## APPENDIX A
### PROOF OF PROPOSITION 2

We show that under assumption (31),

$$\lim_{K\to\infty} \alpha_1^\star = 0,$$

(102)

which implies (32).

Notice that (29b) implies for $k\in\{1,\ldots,K-1\}$:

$$\frac{\alpha_K^* P}{N_K} = \frac{\alpha_k^* P}{(1-\alpha_1^\star-\alpha_2^*\ldots-\alpha_k^*)P+N_k}.$$

(103)

Since for each $k$, the term $(1-\alpha_1^\star-\alpha_2^*-\ldots-\alpha_k^*)$ is nonnegative,

$$\alpha_k^* \geq \frac{N_k}{N_K}\alpha_K^*, \qquad k\in\{1,\ldots,K-1\}.$$

(104)

Thus, by (29a),

$$1 = \sum_{k=1}^{K}\alpha_k^* \geq \sum_{k=1}^{K}\frac{N_k}{N_K}\alpha_K^*$$

and

$$\alpha_K^* \leq \frac{N_K}{\sum_{k=1}^{K}N_k}.$$

We conclude that, for every finite positive integer $K$,

$$R_K \leq \frac{1}{2}\log\left(1+\frac{P}{\sum_{k=1}^{K}N_k}\right),$$

and under Assumption (31), in the limit as $K\to\infty$,

$$\lim_{K\to\infty} R_K = 0.$$

## APPENDIX B
### PROOF OF LEMMA 2

We first prove the converse (37). Fix a blocklength $n$. By Fano's inequality, for each $k\in\mathcal{K}$,

$$\begin{aligned}
nR &= H(M^{(n)}) \\
&\leq I\left(M^{(n)};Y_{k,1}^{(n)},\ldots,Y_{k,n}^{(n)}\right) + \epsilon(n) \\
&\leq I\left(\Theta^{(n)};Y_{k,1}^{(n)},\ldots,Y_{k,n}^{(n)}\right) + \epsilon(n) \\
&\stackrel{(a)}{\leq} I\left(\bar{\Theta}^{(n)};\bar{Y}_{k,1}^{(n)},\ldots,\bar{Y}_{k,n}^{(n)}\right) + \epsilon(n)
\end{aligned}$$

(105)

where $\frac{\epsilon(n)}{n} \to 0$ as $n \to \infty$ and where we defined the tuple $(\bar{\Theta}^{(n)}, \bar{Y}_{k,1}^{(n)}, \ldots, \bar{Y}_{k,n}^{(n)})$ to be jointly Gaussian with the same covariance matrix as the tuple $(\Theta^{(n)}; Y_{k,1}^{(n)}, \ldots, Y_{k,n}^{(n)})$. Inequality (a) holds because the Gaussian distribution maximizes differential entropy under a covariance constraint.

Now, since $\bar{\Theta}^{(n)}, \bar{Y}_{k,1}^{(n)}, \ldots, \bar{Y}_{k,n}^{(n)}$ are jointly Gaussian, there exists a linear combination $\sum_{i=1}^n v_{k,i}^{(n)} \bar{Y}_{k,i}^{(n)}$ such that

$$I\left(\bar{\Theta}^{(n)}; \bar{Y}_{k,1}^{(n)}, \ldots, \bar{Y}_{k,n}^{(n)}\right) = I\left(\bar{\Theta}^{(n)}; \sum_{i=1}^n v_{k,i}^{(n)} \bar{Y}_{k,i}^{(n)}\right). \quad (106)$$

(In fact, the linear combination is simply the LMMSE-estimate of $\bar{\Theta}^{(n)}$ based on $\bar{Y}_{k,1}^{(n)}, \ldots, \bar{Y}_{k,n}^{(n)}$.) Defining the $n$-dimensional row-vector $\mathbf{v}_k^{(n)} = (v_{k,1}^{(n)}, \ldots, v_{k,n}^{(n)})$, in view of (106), we have

$$I\left(\bar{\Theta}^{(n)}; \bar{Y}_{k,1}^{(n)}, \ldots, \bar{Y}_{k,n}^{(n)}\right)$$
$$= \frac{1}{2} \log\left(1 + \frac{(\mathbf{v}_k^{(n)} \mathbf{d}^{(n)})^2 \mathsf{Var}(\Theta^{(n)})}{c_k^{(n)}}\right) \quad (107)$$

where $c_k^{(n)}$ is as defined in (38).

Notice that the right-hand side of (107) does not depend on the norm of $\mathbf{v}_k^{(n)}$ (as long as it is non-zero) but only on the direction. Therefore, without loss of generality, we can assume that

$$\|\mathbf{v}_k^{(n)}\|^2 = 1. \quad (108)$$

By (105) and (107), we conclude that for each $k \in \mathcal{K}$, there exists a unit-norm vector $\mathbf{v}_k^{(n)}$ such that

$$R \le \varliminf_{n\to\infty} \frac{1}{2n} \log\left(1 + \frac{(\mathbf{v}_k^{(n)} \mathbf{d}^{(n)})^2 \mathsf{Var}(\Theta^{(n)})}{c_k^{(n)}}\right). \quad (109)$$

Since by assumption $R > 0$, (109) implies that the ratio $(\mathbf{v}_k^{(n)} \mathbf{d}^{(n)})^2 \mathsf{Var}(\Theta^{(n)})/c_k^{(n)}$ tends to infinity and thus

$$R \le \varliminf_{n\to\infty} \frac{1}{2n} \log\left(\frac{(\mathbf{v}_k^{(n)} \mathbf{d}^{(n)})^2 \mathsf{Var}(\Theta^{(n)})}{c_k^{(n)}}\right). \quad (110)$$

Now, consider the average block-power constraint (35). Since the trace of a positive semidefinite matrix is non-negative and $\mathsf{Var}(\Theta^{(n)}) \le \mathsf{E}[|\Theta^{(n)}|^2]$, by (35), for each $n \in \mathbb{Z}^+$:

$$\|\mathbf{d}^{(n)}\|^2 \mathsf{E}[|\Theta^{(n)}|^2] \le n(P-\delta). \quad (111)$$

Since $\|\mathbf{v}_k^{(n)}\| = 1$, (108), by the Cauchy-Schwarz Inequality,

$$(\mathbf{v}_k^{(n)} \mathbf{d}^{(n)})^2 \mathsf{Var}(\Theta^{(n)}) \le n(P-\delta) \quad (112)$$

and as a consequence

$$\varliminf_{n\to\infty} \frac{1}{2n} \log\left((\mathbf{v}_k^{(n)} \mathbf{d}^{(n)})^2 \mathsf{Var}(\Theta^{(n)})\right) \le 0. \quad (113)$$

Combining this with (110), proves the desired inequality (37).

The proof of Inequalities (39) and (40) relies on Lemmas 3 and 4 at the end of this appendix. Notice that the monotonicity of the log-function and the nonnegativity of the norm combined with (37) imply that for each $k \in \mathcal{K}$,

$$R \le \varliminf_{n\to\infty} -\frac{1}{2n} \log\left\|\mathbf{v}_k^{(n)}(\mathsf{I} + \mathsf{A}_k^{(n)})\right\|^2, \quad (114)$$

where recall that we assumed $R > 0$.

Define for each $k \in \mathcal{K}$ and positive integer $n$ the set

$$\mathcal{S}_k^{(n)} := \left\{i \in \{1, \ldots, n\}: v_{k,i}^{(n)} > n^{-2\log n}\right\}. \quad (115)$$

By Lemma 3 and Inequality (114), the cardinality of each set $\mathcal{S}_k^{(n)}$ is unbounded,

$$|\mathcal{S}_k^{(n)}| \to \infty \quad \text{as} \quad n \to \infty, \quad k \in \mathcal{K}. \quad (116)$$

Applying now Lemma 4 to $p = P - \delta$, to

$$\pi_i^{(n)} = \mathsf{E}\left[\left(X_i^{(n)}\right)^2\right], \quad (117)$$

and to $\mathcal{T}^{(n)} = \mathcal{S}_k^{(n)}$ implies that for each $k \in \mathcal{K}$ there exists a sequences of indices $\{j_k^{(n)} \in \mathcal{S}_k^{(n)}\}_{n=1}^\infty$ that satisfies (39). Since every sequence of indices $\{i^{(n)} \in \mathcal{S}_k^{(n)}\}_{n=1}^\infty$ also satisfies (40), this concludes the proof of the lemma.

**Lemma 3.** *For each $n \in \mathbb{Z}^+$, let $\mathsf{A}^{(n)}$ be a strictly lower-triangular $n$-by-$n$ matrix and $\mathbf{v}^{(n)}$ an $n$-dimensional row-vector. Let $a_{i,j}^{(n)}$ denote the row-$i$, column-$j$ entry of $\mathsf{A}^{(n)}$ and $v_i^{(n)}$ denote the $i$-th entry of $\mathbf{v}^{(n)}$. Assume that the elements $a_{i,j}^{(n)}$ are bounded as*

$$|a_{i,j}^{(n)}|^2 \le np \quad (118)$$

*for some real number $p > 0$, and that the inequality*

$$\varliminf_{n\to\infty} -\frac{1}{2n} \log\|\mathbf{v}^{(n)}(\mathsf{I} + \mathsf{A}^{(n)})\|^2 \ge \Gamma \quad (119)$$

*holds for some real number $\Gamma > 0$. Then, for each $\epsilon \in (0, \Gamma)$ and for all sufficiently large $n$ the following implication holds: If*

$$|v_j^{(n)}| > e^{-n(\Gamma-\epsilon)} \quad (120a)$$

*for some index $j \in \{1, \ldots, n\}$, then there must exist an index $i \in \{j+1, \ldots, n\}$ such that*

$$|v_i^{(n)}| \ge \frac{|v_j^{(n)}| - e^{-n(\Gamma-\epsilon)}}{n^{\frac{3}{2}} \sqrt{p}}. \quad (120b)$$

*If moreover, the vectors $\{\mathbf{v}^{(n)}\}_{n=1}^\infty$ are of unit norm, then the cardinality of the set*

$$\mathcal{S}^{(n)} := \left\{j \in \{1, \ldots, n\} : |v_j^{(n)}| > n^{-2\log(n)}\right\} \quad (121)$$

*is unbounded in $n$.*

*Proof:* Fix $\epsilon \in (0, \Gamma)$ and let $n$ be sufficiently large so that

$$-\frac{1}{2n} \log\left\|\mathbf{v}^{(n)}(\mathsf{I} + \mathsf{A}^{(n)})\right\|^2 \ge \Gamma - \epsilon. \quad (122)$$

This is possible by (119).

Since $\mathsf{A}^{(n)}$ is strictly lower-triangular,

$$\|\mathbf{v}^{(n)}(\mathsf{I} + \mathsf{A}^{(n)})\|^2 = \sum_{j=1}^n \left(v_j^{(n)} + \sum_{i=j+1}^n v_i^{(n)} a_{i,j}^{(n)}\right)^2$$
$$\ge \sum_{j=1}^n \left(|v_j^{(n)}| - \left|\sum_{i=j+1}^n v_i^{(n)} a_{i,j}^{(n)}\right|\right)^2$$
$$\ge \left(|v_j^{(n)}| - \left|\sum_{i=j+1}^n v_i^{(n)} a_{i,j}^{(n)}\right|\right)^2$$

and by (122) and the monotonicity of the log-function, for all $j \in \{1, \ldots, n\}$:

$$-\frac{1}{2n} \log \left( |v_j^{(n)}| - | \sum_{i=j+1}^{n} v_i^{(n)} a_{i,j}^{(n)} | \right)^2 \geq \Gamma - \epsilon.$$

Thus,

$$|v_j^{(n)}| \leq | \sum_{i=j+1}^{n} v_i^{(n)} a_{i,j}^{(n)}| + e^{-n(\Gamma-\epsilon)}$$

and by (118):

$$|v_j^{(n)}| - e^{-n(\Gamma-\epsilon)} \leq | \sum_{i=j+1}^{n} v_i^{(n)} a_{i,j}^{(n)}|$$
$$\leq \sum_{i=j+1}^{n} |v_i^{(n)}| \sqrt{np}. \tag{123}$$

If $|v_j^{(n)}| \leq e^{-n(\Gamma-\epsilon)}$, then the sum on the right-hand side of (123) can be empty, i.e., $j = n$. However, if

$$|v_j^{(n)}| > e^{-n(\Gamma-\epsilon)}, \tag{124}$$

then the sum needs to have at least one term. Indeed, if (124) holds and $i < n$, there must exist an index $i \in \{j+1, ..., n\}$ such that

$$\frac{1}{n} \left( |v_j^{(n)}| - e^{-n(\Gamma-\epsilon)} \right) \leq |v_i^{(n)}| \sqrt{np}, \tag{125}$$

which is equivalent to the desired bound (120b).

We now prove the second part of the lemma, i.e., the unboundedness of the cardinalities of the sets $\mathcal{S}^{(n)}$, where we assume that the vectors $\{\mathbf{v}^{(n)}\}$ are of unit norm. In the following, let $n$ be sufficiently large so that the first part of the lemma, Implication (120), holds and so that

$$\frac{1}{\sqrt{n}} > \frac{1}{n^{2\log(n)}} > e^{-n(\Gamma-\epsilon)} \tag{126}$$

and for every $\ell \in \{1, \ldots, \log(n)\}$

$$\frac{1}{n^{(3\ell+1)/2} p^{\ell/2}} - e^{-n(\Gamma-\epsilon)} n^{-3/2} p^{-1/2} \frac{1-n^{-3\ell/2} p^{-\ell/2}}{1-n^{-3/2} p^{-1/2}}$$
$$> \frac{1}{n^{2\log(n)}} \tag{127}$$

Since $\|\mathbf{v}^{(n)}\|^2 = 1$, for each $n$, there must exist an index $i_0^{(n)} \in \{1, \ldots, n\}$ such that

$$|v_{i_0^{(n)}}^{(n)}| \geq \frac{1}{\sqrt{n}}, \tag{128}$$

and by (126)

$$|v_{i_0^{(n)}}^{(n)}| > n^{2\log(n)} > e^{-n(\Gamma-\epsilon)}. \tag{129}$$

We conclude by (120) that there exists an index $i_1^{(n)} \in \{i_0^{(n)} + 1, \ldots, n\}$ satisfying

$$|v_{i_1^{(n)}}^{(n)}| \geq \frac{|v_{i_0^{(n)}}^{(n)}| - e^{-n(\Gamma-\epsilon)}}{n^{\frac{3}{2}} \sqrt{p}}$$
$$\geq \frac{1}{n^2 \sqrt{p}} - \frac{e^{-n(\Gamma-\epsilon)}}{n^{\frac{3}{2}} \sqrt{p}} \tag{130}$$

where the inequality follows from (128). By (126) and (127), (applied for $\ell = 1$), Inequality (130) implies that

$$|v_{i_1^{(n)}}^{(n)}| > e^{-n(\Gamma-\epsilon)}, \tag{131}$$

and consequently, by (120), there exists an index $i_2^{(n)} \in \{i_1^{(n)} + 1, \ldots, n\}$ satisfying

$$|v_{i_2^{(n)}}^{(n)}| \geq \frac{|v_{i_1^{(n)}}^{(n)}| - e^{-n(\Gamma-\epsilon)}}{n^{\frac{3}{2}} \sqrt{p}} \tag{132}$$
$$\geq \frac{1}{n^{7/2} p} - \frac{e^{-n(\Gamma-\epsilon)}}{n^3 p} - \frac{e^{-n(\Gamma-\epsilon)}}{n^{\frac{3}{2}} \sqrt{p}} \tag{133}$$
$$> e^{-n(\Gamma-\epsilon)}, \tag{134}$$

where the last inequality follows by (126) and (127) (applied for $\ell = 2$).

Repeating these arguments iteratively, we conclude that it is possible to find indices $1 \leq i_0^{(n)} < i_1^{(n)} < \ldots < i_{\log(n)}^{(n)} < n$ such that for each $\ell \in \{1, \ldots, \log(n)\}$:

$$|v_{i_\ell^{(n)}}^{(n)}| \geq \frac{1}{n^{3(\ell+1)/2} p^{\ell/2}} - e^{-n(\Gamma-\epsilon)} \sum_{j=1}^{\ell} \left( n^{-3/2} p^{-1/2} \right)^j$$
$$= \frac{1}{n^{3(\ell+1)/2} p^{\ell/2}}$$
$$- e^{-n(\Gamma-\epsilon)} n^{-3/2} p^{-1/2} \frac{1-n^{-3\ell/2} p^{-\ell/2}}{1-n^{-3/2} p^{-1/2}} \tag{135}$$
$$> \frac{1}{n^{2\log(n)}} \tag{136}$$
$$> e^{-(\Gamma-\epsilon)} \tag{137}$$

where the last two inequalities follow from (126) and (127). This proves that for sufficiently large $n$ the cardinality of the set $\mathcal{S}^{(n)}$ as defined in (121) is at least $\log(n)$ and thus unbounded in $n$. ∎

**Lemma 4.** *For each positive integer $n$, let $(\pi_1^{(n)}, \ldots, \pi_n^{(n)})$ be a tuple of nonnegative real numbers that satisfy*

$$\frac{1}{n} \sum_{i=1}^{n} \pi_i^{(n)} \leq p \tag{138}$$

*for some real number $p > 0$, and let $\mathcal{T}^{(n)}$ be a subset of the indices from 1 to $n$,*

$$\mathcal{T}^{(n)} \subseteq \{1, \ldots, n\}, \tag{139}$$

*that satisfies*

$$|\mathcal{T}^{(n)}| \to \infty \quad \text{as} \quad n \to \infty. \tag{140}$$

*Then, there exists a sequence of indices $\left\{ i^{(n)} \in \mathcal{T}^{(n)} \right\}_{n=1}^{\infty}$ such that*

$$\varlimsup_{n\to\infty} \frac{1}{n} \pi_{i^{(n)}}^{(n)} = 0. \tag{141}$$

*Proof:* Since all numbers $\pi_i^{(n)}$ are nonnegative, for every sequence of indices $\{i^{(n)} \in \mathcal{T}^{(n)}\}_{n=1}^{\infty}$,

$$\varlimsup_{n\to\infty} \frac{1}{n} \pi_{i^{(n)}}^{(n)} \geq 0. \tag{142}$$

We thus have to prove that there exists at least one sequence of indices $\{i^{(n)} \in \mathcal{T}^{(n)}\}_{n=1}^{\infty}$ that satisfies

$$\varlimsup_{n \to \infty} \frac{1}{n} \pi_{i^{(n)}}^{(n)} \leq 0. \tag{143}$$

We prove this by contradiction. Assume that for each sequence of indices $\{i^{(n)} \in \mathcal{T}^{(n)}\}_{n=1}^{\infty}$

$$\varlimsup_{n \to \infty} \frac{1}{n} \pi_{i^{(n)}}^{(n)} > 0. \tag{144}$$

Define for each $n \in \mathbb{Z}^{+}$

$$\pi_{\min}^{(n)} := \min_{i \in \mathcal{T}^{(n)}} \pi_i^{(n)}, \tag{145}$$

and define the limit

$$\delta_{\min} := \varlimsup_{n \to \infty} \frac{1}{n} \pi_{\min}^{(n)}, \tag{146}$$

which by Assumption (144) is strictly positive,

$$\delta_{\min} > 0. \tag{147}$$

Now, since all the terms $\pi_i^{(n)}$ are nonnegative:

$$\frac{1}{n} \sum_{i=1}^{n} \pi_i^{(n)} \geq \frac{1}{n} \sum_{i \in \mathcal{T}^{(n)}} \pi_i^{(n)} \geq \frac{1}{n} \pi_{\min}^{(n)} |\mathcal{T}^{(n)}|, \tag{148}$$

where the second inequality follows by the definition in (145). By (146) and (147) and by the undboundedness of the cardinality of the sets $\mathcal{T}^{(n)}$, we conclude that the sum in (148) is unbounded in $n$, which contradicts Assumption (138) and thus concludes our proof. ∎

## APPENDIX C
### EXISTENCE OF CODE $\mathcal{C}_2, \ldots, \mathcal{C}_L$ WITH THE DESIRED PROPERTIES

The proof is by induction: for each $\ell \in \{2, \ldots, L\}$, when proving the existence of the desired $\mathcal{C}_\ell$, we assume that

$$\gamma_{l-1} \leq \exp(-\underbrace{\exp \circ \ldots \circ \exp}_{l-2 \text{ times}}(\Omega(n))). \tag{149}$$

For $l = 2$, Inequality (149) follows from (71).

By [17], for all rates

$$\tilde{R} < \frac{1}{2} \log \frac{2 + \sqrt{\tilde{P}^2/\sigma^4 + 4}}{4},$$

and for sufficiently large $n$ there exists a blocklength-$\tilde{n}$, rate-$\tilde{R}$ non-feedback coding scheme for the memoryless Gaussian point-to-point channel with noise variance $\sigma^2$, with expected average block-power no larger than $\tilde{P}$ and with probability of error $P_e$ satisfying

$$P_e \leq e^{-\tilde{n}(E(\tilde{R}, \tilde{P}/\sigma^2) - \epsilon')} \tag{150}$$

for some fixed $\epsilon' > 0$ and

$$E(\tilde{R}, \tilde{P}) = \frac{\tilde{P}}{4\sigma^2} \left(1 - \sqrt{1 - e^{-2\tilde{R}}}\right). \tag{151}$$

Since the probability of error of a non-feedback code over the Gaussian BC with common message is at most $K$ times the probability of error to the weakest receiver, we conclude that for all $\tilde{P} > 0$ and

$$0 < \tilde{R} < \frac{1}{2} \log \frac{2 + \sqrt{\tilde{P}^2/\sigma_1^4 + 4}}{4}, \tag{152a}$$

there exists a rate-$\tilde{R}$ code with power $\tilde{P}$ and blocklength $\tilde{n}$ that for the Gaussian BC with common message achieves probability of error

$$P_e^{(\text{BC})} \leq K e^{-\tilde{n}\left(\frac{P}{4\sigma_1^2}\left(1 - \sqrt{1 - e^{-2\tilde{R}}}\right) - \epsilon'\right)}. \tag{152b}$$

Now apply this statement to $\tilde{R} = R_{\text{phase},l}$, $\tilde{P} = P/\gamma_{l-1}$ and $\tilde{n} = \frac{\epsilon n}{L-1} - 1$. Since for sufficiently large $n$, by (149),

$$R_{\text{phase},l} < \frac{1}{2} \log \frac{2 + \sqrt{\frac{P^2}{\gamma_{l-1}^2 \sigma_1^4} + 4}}{4}, \tag{153}$$

we conclude by (152) that there exists a code $\mathcal{C}_l$ of rate-$R_{\text{phase},l}$, block-power $P/\gamma_{l-1}$, blocklength $\frac{\epsilon n}{L-1} - 1$ and probability of error $\rho_l$ satisfying

$$\begin{aligned}
\rho_l &\leq K e^{-\left(\frac{\epsilon n}{L-1} - 1\right)\left(\frac{P}{4\gamma_{l-1}\sigma_1^2}\left(1 - \sqrt{1 - e^{-2\frac{R(L-1)}{\epsilon(L-1)/n}}}\right) - \epsilon'\right)} \\
&\leq \exp(-\underbrace{\exp \circ \ldots \circ \exp}_{l-1 \text{ times}}(\Omega(n)))
\end{aligned} \tag{154}$$

where the inequality follows again by (149).

By the definition of $\gamma_l$ in (74), Inequalities (154) and (149) also yield:

$$\gamma_l \leq \exp(-\underbrace{\exp \circ \ldots \circ \exp}_{l-1 \text{ times}}(\Omega(n))). \tag{155}$$

## REFERENCES

[1] J. P. M. Schalkwijk and T. Kailath, "A coding scheme for additive noise channels with feedback-I: no bandwidth constraint," *IEEE Trans. on Inf. Theory*, vol. 12, no. 2, pp. 183–189, Apr. 1966.

[2] Y.-H. Kim, "Feedback capacity of stationary Gaussian channels," *IEEE Trans. on Inf. Theory*, vol. 56, no. 1, pp. 57–85, Jan. 2010.

[3] L. H. Ozarow, "The capacity of the white Gaussian multiple access channel with feedback," *IEEE Trans. on Inf. Theory*, vol. 30, no. 4, pp. 623–629, July 1984.

[4] L. H. Ozarow, and S. K. Leung-Yan-Cheong, "An achievable region and outer bound for the Gaussian broadcast channel with feedback," *IEEE Trans. on Inf. Theory*, vol. 30, no. 4, pp. 667–671, July 1984.

[5] G. Kramer, "Feedback strategies for white Gaussian interference networks," *IEEE Trans. on Inf. Theory*, vol. 48, no. 6, pp. 1423–1438, June 2002.

[6] E. Ardestanizadeh, M. Wigger, Y.-H. Kim, and T. Javidi, "Linear sum-capacity for Gaussian multiple access channels with feedback," *IEEE Trans. on Inf. Theory*, vol. 58, no. 1, pp. 224–236, Jan. 2012.

[7] E. Ardestanizadeh, P. Minero and M. Franceschetti, "LQG control approach to Gaussian broadcast channels with feedback," *IEEE Trans. on Inf. Theory*, vol. 58, no. 8, pp. 5267–5278, Aug. 2012.

[8] M. Gastpar, A. Lapidoth, Y. Steinberg, and M. Wigger, "Coding schemes and asymptotic capacity for the Gaussian broadcast and interference channels with feedback," *IEEE Trans. on Inf. Theory*, vol. 60, no. 1, pp. 54–71, Jan. 2014.

[9] M. Gastpar, A. Lapidoth, Y. Steinberg, and M. Wigger, "New achievable rates for the Gaussian broadcast channel with feedback," in *Proc. of 8th ISWCS*, 2011, Aachen, Germany, Nov. 6-9, 2011, pp. 579–583.

[10] S. Belhadj Amor, Y. Steinberg, and M. Wigger, "Duality with linear-feedback schemes for the scalar Gaussian MAC and BC", in *Proc. of 2014 IZS*, Zurich, Switzerland, Feb. 26–28, 2014, pp. 25–28.

[11] Y.-H. Kim, A. Lapidoth, and T. Weissman, "On the reliability of Gaussian channels with noisy feedback," in *Proc. of the 44th Allerton*, Monticello, Illinois, Sep. 2006, pp. 364–371.

[12] R. Mirghaderi, A. Goldsmith, T. Weissman, "Achievable error exponents in the Gaussian channel with rate-limited feedback", *IEEE Trans. on Inf. Theory*, vol. 59, no. 12 pp. 8144–8156, Dec. 2013.

[13] R. G. Gallager and B. Nakiboglu, "Variations on a theme by Schalkwijk and Kailath," *IEEE Trans. on Inf. Theory*, vol. 56, no. 1, pp. 6–17, Jan. 2010.

[14] A. Sahai, S. C. Draper, and M. Gastpar, "Boosting reliability over AWGN networks with average power constraints and noiseless feedback," in *Proc. of ISIT 2005*, Adelaide, Australia, Sep. 2005, pp. 402–406.

[15] P. P. Bergmans, "A simple converse for broadcast channels with additive white Gaussian noise," *IEEE Trans. on Inf. Theory*, vol. 20, no. 2, pp. 279–280, Mar. 1974.

[16] A. El Gamal, "The feedback capacity of degraded broadcast channels," *IEEE Trans. on Inf. Theory*, vol. 24, no. 3, pp. 379–381, May 1978.

[17] C. E. Shannon, "Probability of error for optimal codes in a Gaussian channel," *Bell Systems Tech. Journal*, vol. 38, no. 3, pp. 611–656, May 1959.

**Youlong WU** (S'13) received the B.S. degree in electrical engineering from Wuhan University, Wuhan, China, in 2007. He received the M.S. degree in electrical engineering from Shanghai Jiaotong University, Shanghai, China, in 2011. He is currently pursuing the Ph.D. Degree at Telecom ParisTech, in Paris, France. His research interests are in information theory and wireless communication.

**Paolo Minero** (M'11) received the Laurea degree (with highest honors) in electrical engineering from the Politecnico di Torino, Torino, Italy, in 2003, the M.S. degree in electrical engineering from the University of California at Berkeley in 2006, and the Ph.D. degree in electrical engineering from the University of California at San Diego in 2010. He is an Assistant Professor in the Department of Electrical Engineering at the University of Notre Dame. Before joining the University of Notre Dame, he was a postdoctoral scholar at the University of California at San Diego for six month. His research interests are in communication systems theory and include information theory, wireless communication, and control over networks. Dr. Minero received the U.S. Vodafone Fellowship in 2004 and 2005, and the Shannon Memorial Fellowship in 2008.

**Michèle Wigger** (S'05-M'09) received the M.Sc. degree in electrical engineering (with distinction) and the Ph.D. degree in electrical engineering both from ETH Zurich in 2003 and 2008, respectively. In 2009 she was a postdoctoral researcher at the ITA center at the University of California, San Diego. Since December 2009 she is an Assistant Professor at Telecom ParisTech, in Paris, France. Her research interests are in information and communications theory.