IEICE TRANSACTIONS

on Fundamentals of Electronics, Communications and Computer Sciences

VOL. E100-A NO. 5
MAY 2017

A PUBLICATION OF THE ENGINEERING SCIENCES SOCIETY
The Institute of Electronics, Information and Communication Engineers
Kikai-Shinko-Kaikan Bldg., 5-8, Shibakoen 3 chome, Minato-ku, TOKYO, 105-0011 JAPAN

PAPER

Posterior Matching for Gaussian Broadcast Channels with Feedback

Lan V. TRUONG[†a)], *Nonmember and* **Hirosuke YAMAMOTO**[††b)], *Fellow*

SUMMARY In this paper, the posterior matching scheme proposed by Shayevits and Feder is extended to the Gaussian broadcast channel with feedback, and the error probabilities and achievable rate region are derived for this coding strategy by using the iterated random function theory. A variant of the Ozarow-Leung code for the general two-user broadcast channel with feedback can be realized as a special case of our coding scheme. Furthermore, for the symmetric Gaussian broadcast channel with feedback, our coding scheme achieves the linear-feedback sum-capacity like the LQG code and outperforms the Kramer code.

key words: *Gaussian broadcast channel with feedback, feedback, posterior matching, iterated function systems*

1. Introduction

The capacity region of the broadcast channel with M users (i.e. M receivers) is a well-known open problem. However, it is known that feedback can increase the capacity region for broadcast channels. Specially, Ozarow and Leung [1] proved for $M = 2$ that feedback can increase the capacity region of the additive white Gaussian broadcast channel (AWGN-BC) by cooperation between the users and the sender via feedback. Kramer [2] extended this coding scheme to the case of $M \geq 3$. Later, Elia [3] showed for $M = 2$ that the achievable rate region obtained by Ozarow and Leung [1] can be enlarged by using robust control theory. Ardestanizadeh et al. [4] proposed a coding scheme based on LQG (Linear Quadratic Gaussian) control approach for the symmetric AWGN-BC with feedback, and showed that their LQG code can attain the same achievable rate region as the Elia scheme [3] for $M = 2$ and outperforms the Kramer code [2] for the symmetric AWGN-BC with feedback for $M \geq 3$. The LQG code is derived based on a mapping from a feedback control problem to a linear code for the AWGN-BC with feedback. The achievable rate region is determined by the eigenvalues of the open-loop matrix of a linear system and the power constraint of channel input is related to the minimum power needed to stabilize the system using a feedback control signal.

Recently, Amor et al. [5], [6] showed that the rate regions achieved by linear feedback coding schemes over dual multi-antenna AWGN multi-access channels (MACs) and broadcast channels (BCs) with independent noises coincide, and the sum-rate achieved by the LQG code is optimal among all the linear-feedback coding schemes for the symmetric AWGN-BCs. This optimal sum-rate is called linear-feedback sum-capacity, and they showed for $M = 2$ that the linear-feedback sum-capacity of the scalar AWGN-BC with independent noises can be achieved by a simple rearrangement of Ozarow's MAC coding scheme [7]. (Refer to Remark 7 in Sect. 5 for more details.) However, it is not shown for $M \geq 3$ how to construct a coding scheme for AWGN-BCs with feedback by a rearrangement of a coding scheme for AWGN-MACs with feedback. Note that since Kramer's MAC coding scheme [2], which is a generalization of Ozarow's MAC coding scheme for $M \geq 3$, uses complex modulation coefficients, it is not easy to construct a BC coding scheme from Kramer's MAC coding scheme even if we try to use a rearrangement similar to the one used in [6].

In a more general setting, Gaspar et al. [8], [9] proposed a coding scheme for the AWGN-BC with correlated noises in the case of $M = 2$ with arbitrary noise covariance and in the case of $M \geq 3$ such that the noise of each user is a multiple of the same Gaussian noise. For example, they showed that for all noise correlations other than ± 1, the gap between the sum-rate of their scheme and the full-cooperation bound vanishes as the signal-to-noise ratio tends to infinity. Although their coding scheme works well in the asymptotic regime, it does not work well when the input power is not sufficiently large.

Shayevits and Feder [10] proposed the Posterior Matching (PM) Scheme for the point-to-point communication system with feedback, and they showed that the PM Scheme reduces to the Schalkwijk-Kailath scheme [11] when the channel is Gaussian. But, it is very hard to directly apply their scheme to AWGN-BCs because we need to assign multiple messages to a single vector and to refine the vector sequentially based on feedback to reduce the uncertainty of every user at the same time. To execute such a behavior, a higher order kernel is required for the reversed iterated function system (RIFS) used in the decoders, and all the decoders must know all the other decoders' messages. On the other hand, the indirect assignment methods used in the Ozarow-Leung code [1] or the Kramer code [2] can lead to only a suboptimal sum-rate compared with the Elia code [3] and the LQG code [4] as mentioned above. Specially, they assumed that the transmitted signal at each time n is a linear combination of different signal components, each of which is intended to decrease the uncertainty of each user, and they

Manuscript received August 22, 2016.
Manuscript revised December 29, 2016.

[†]The author is with National University of Singapore, Singapore.
[††]The author is with The University of Tokyo, Kashiwa-shi, 277-8561 Japan.
a) E-mail: lantruong@u.nus.edu
b) E-mail: hirosuke@ieee.org

DOI: 10.1587/transfun.E100.A.1165

also imposed a redundant restriction such that each signal component at time $n+1$ must be statistically independent of the signal feedbacked from the corresponding user at time n. This idea is originated from the Schalkwijk-Kailath scheme [11] and repeated in the Shayevitz-Feder scheme [10, Section A] to attain the capacity for point-to-point AWGN channels with feedback. But for the AWGN-BCs with feedback, this scheme cannot realize so good performance as the Elia code [3] and the LQG code [4].

In this paper, we extend the PM scheme [10] to AWGN-BCs with M users by devising a new encoding scheme for any M such that an $M \times M$ binary Hadamard matrix exists. Our encoding procedure can be considered as an optimization of the Kramer scheme [2] by using some mathematical tricks. The decoding scheme uses the same technique as the Shayevits-Feder scheme [10]. But our coding scheme is a general one for AWGN-BCs with feedback because it includes all the coding schemes treated in [1] and [2] as special cases, and we derive the achievable rate region of the proposed coding scheme. Then, we prove that a variant of the Ozarow-Leung scheme [1] obtained from our scheme can achieve the same achievable rate region as the original Ozarow-Leung scheme. Furthermore, we propose a coding scheme for physically non-degraded symmetric AWGN-BCs with feedback which can achieve the linear-feedback sum-capacity like the LQG code. Besides, since our coding scheme is a variant of the Kramer code, it has a potential to achieve not only the asymptotic capacity [8], [9] but also a good performance in non-asymptotic settings. More precisely, we can determine the code length (i.e. the repetition number of feedback) necessary to attain a given target of error probabilities and coding rates in our coding scheme in the same way as other PM schemes. This is an advantage over the Elia code [3] and the LQG code [4], in which we cannot determine the necessary code length because the decoding error exponent and achievable mean square error exponent are treated only in the asymptotic setting for these codes.

This paper is organized as follows. Section 2 presents the channel model and some mathematical preliminaries. A general time-varying coding scheme is proposed for AWGN-BCs with feedback in Sect. 3, and the achievable rate region and error probabilities for this general scheme are derived in Sect. 4. Section 5 shows that a variant of the Ozarow-Leung coding scheme can be obtained from our coding scheme. We show that the proposed coding scheme can achieve the linear-feedback sum-capacity for physically non-degraded symmetric AWGN-BCs with feedback in Sect. 6. Finally in Sect. 8, we compare the sum-rate for the AWGN-BC with the one for the AWGN-MAC.

2. Channel Model and Preliminaries

2.1 Mathematical Notations

Upper-case letters and lower-case letters denote random variables and their realizations, respectively. A real-valued random variable X is associated with a distribution $\mathbb{P}_X(\cdot)$ defined on the usual Borel σ-algebra over $\mathbb{R}$, and we write $X \sim \mathbb{P}_X$. The cumulative distribution function (c.d.f.) of X is given by $F_X(x) = \mathbb{P}_X((-\infty, x])$, and their inverse c.d.f is defined as $F_X^{-1}(t) \equiv \inf\{x : F_X(x) > t\}$. The uniform probability distribution over $(0, 1)$ is denoted by $\mathcal{U}$. In addition, we use the following notation. $(f \circ g)(x) \equiv f(g(x))$, $\mathbf{Y}_p^{q(m)} \equiv (Y_p^{(m)}, Y_{p+1}^{(m)}, ..., Y_q^{(m)})$ for $p \le q$, and $\mathrm{tr}(\mathbf{A})$ is the trace of matrix $\mathbf{A}$. In this paper, we use the following lemma:

Lemma 1 ([10, Lemma 1]): Let X be a continuous random variable with $X \sim \mathbb{P}_X$ and Θ be a uniform distribution random variable, i.e. $\Theta \sim \mathcal{U}$, and X be statistical independent of Θ. Then $F_X^{-1}(\Theta) \sim \mathbb{P}_X$ and $F_X(X) \sim \mathcal{U}$.

The binary Hadamard matrix [12] of order M is an $(M \times M)$ matrix of $+1$s and -1s such that $\mathbf{H}_M \mathbf{H}_M^T = M\mathbf{I}$ where $\mathbf{I}$ is the $(M \times M)$ identity matrix. It is not yet known for which values of M an $\mathbf{H}_M$ exists. However, we know that if the Hadamard matrix of order M exists then M is $1, 2, 4$, or a multiple of 4. Moreover, if M is of the form 2^m for a positive integer m we can construct $\mathbf{H}_M$ by using Sylvester's method. In addition, Paley's construction, which uses quadratic residues, can be used to construct Hadamard matrices of order M when M is equal to $p+1$ for a prime p and M is also a multiple of 4.

2.2 AWGN-BCs with Feedback

We extend the communication model treated in [1] to the case of AWGN-BCs. Consider the communication system shown in Fig. 1 such that one encoder and M decoders are connected via an AWGN-BC and all channel outputs are noiselessly feedbacked to the encoder. Let Θ_m be a random message point uniformly distributed over the unit interval that must be transmitted from the encoder to decoder $m \in \{1, 2, ..., M\}$. At each time n, the received signal of decoder m is

$$Y_n^{(m)} = X_n + Z_n + Z_n^{(m)}, \tag{1}$$

where $X_n \in \mathbb{R}$ is the symbol transmitted from the encoder at time n, and $Y_n^{(m)} \in \mathbb{R}$ is the signal received by decoder

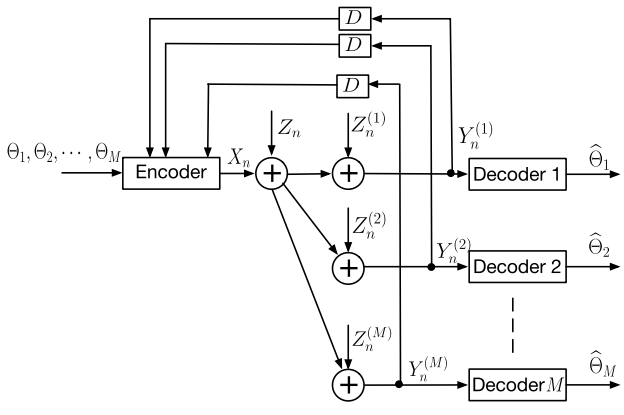

Fig. 1 M-user Gaussian broadcast channel with feedback.

m at time n. Z_n is a common white Gaussian noise with variance σ^2, and $Z_n^{(m)}$ are individual white Gaussian noises with variance σ_m^2 for $m = 1, 2, \cdots, M$. For physically non-degraded AWGN-BCs, we can set $\sigma^2 = 0$ and $\sigma_m^2 > 0$. We also assume that output symbols are casually feedbacked to the encoder and the transmitted symbol X_n at time n can depend on both messages $(\Theta_1, \Theta_2, ..., \Theta_M)$ and the previous channel output sequences $\left(\mathbf{Y}^{n-1(1)}, \mathbf{Y}^{n-1(2)}, \cdots, \mathbf{Y}^{n-1(M)}\right)$ where $\mathbf{Y}^{n-1(m)} \equiv (Y_1^{(m)}, Y_2^{(m)}, \cdots, Y_{n-1}^{(m)})$.

An *encoding scheme* for an AWGN-BC is a measurable transmission function $g_n : (0,1)^M \times \mathbb{R}^{(n-1)M} \to \mathbb{R}$, so that the channel input generated by the encoder is given by

$$X_n = g_n\left(\Theta_1, ..., \Theta_M, \mathbf{Y}^{n-1(1)}, \mathbf{Y}^{n-1(2)}, ..., \mathbf{Y}^{n-1(M)}\right). \tag{2}$$

A *decoding rule* for the AWGN-BC is the sequences of measurable mappings $\{\Delta_n^{(m)} : \mathbb{R}^n \to \mathcal{E}\}_{n=1}^{\infty}$, where $\mathcal{E}$ is the set of all open intervals in $(0, 1)$. We refer to $\Delta_n^{(m)}(\mathbf{y}^{n(m)})$ as the decoded interval of decoder m. The error probabilities at time n are defined as

$$p_{n,e}^{(m)} \equiv \mathbb{P}(\Theta_m \notin \Delta_n^{(m)}(\mathbf{Y}^{n(m)})) \tag{3}$$

for $m = 1, 2, \cdots, M$, and the corresponding coding rate at time n is defined by

$$R_n^{(m)} \equiv -\frac{1}{n} \log\left|\Delta_n^{(m)}\left(\mathbf{Y}^{n(m)}\right)\right|, \tag{4}$$

where $\left|\Delta_n^{(m)}(\mathbf{Y}^n)\right|$ is the length of the interval $\Delta_n^{(m)}(\mathbf{Y}^n)$.

We say that a coding scheme achieves a rate tuple $(R_1, R_2, \cdots, R_M)$ over an AWGN-BC if for all $m \in \{1, 2, ..., M\}$, it satisfies

$$\lim_{n\to\infty} \mathbb{P}\left(R_n^{(m)} < R_m\right) = 0, \tag{5}$$

$$\lim_{n\to\infty} p_{n,e}^{(m)} = 0. \tag{6}$$

The rate tuple is achieved within an input power constraint P if it also satisfies

$$\limsup_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \mathbb{E}[X_k^2] \le P. \tag{7}$$

An *optimal fixed rate* decoding rule for an AWGN-BC with feedback for rate tuple $(R_1, R_2, ..., R_M)$ is the one that decodes the tuple of fixed length intervals $(J_1, J_2, ..., J_M)$ satisfying $|J_m| = 2^{-nR_m}$ for each m, which maximizes each marginal posteriori probability, i.e.,

$$\triangle_n^{(m)}(\mathbf{y}^{n(m)}) = \underset{J_m \in \mathcal{E}:|J_m|=2^{-nR_m}}{\operatorname{argmax}} \mathbb{P}_{\Theta_m|\mathbf{Y}^{n(m)}}(J_m|\mathbf{y}^{n(m)}). \tag{8}$$

An *optimal variable rate* decoding rule with target error probabilities $p_{n,e}^{(m)} = \delta_n^{(m)}$ is the one that decodes the tuple of minimal-length intervals $(J_1, J_2, ..., J_M)$ such that each accumulated marginal posteriori probability exceeds corresponding target, i.e.,

$$\triangle_n^{(m)}(\mathbf{y}^{n(m)}) = \underset{J_m \in \mathcal{E}:\mathbb{P}_{\Theta_m|\mathbf{Y}^{n(m)}}(J_m|\mathbf{y}^{n(m)}) \ge 1-\delta_n^{(m)}}{\operatorname{argmin}} |J_m|. \tag{9}$$

Both decoding rules make good use of the marginal posterior distribution of the message point $\mathbb{P}_{\Theta_m|Y^n}$ which can be calculated online at the encoder and each decoder. Refer [10] for more details. Then, the following lemma holds.

Lemma 2 ([10, Lemma 3]): The achievability defined by (5)–(7) implies the achievability in the standard framework.

Remark 1: In the standard framework, a message i_m uniformly distributed over $\{1, 2, ..., 2^{n\tilde{R}_n^{(m)}}\}$ is sent to decoder m via a BC when the coding rate is $\tilde{R}_n^{(m)}$. It is shown in the proof of [10, Lemma 3] that if $\tilde{R}_n^{(m)}$ satisfies

$$\tilde{R}_n^{(m)} \le R_m + \frac{1}{n} \log\left(1 - \sqrt{p_{e,n}^{(m)}} - \tau_n\right) \tag{10}$$

for some $\tau_n > 0$ such that $\lim_{n\to\infty} \tau_n = 0$, then we can choose message points $\theta_{i_m,n}$ in $(0, 1)$ such that $\theta_{i_m+1,n} - \theta_{i_m,n} \ge 2^{-nR_m}$ for $1 \le i_m \le 2^{n\tilde{R}_n^{(m)}} - 1$ and the decoding error probability $\tilde{p}_{e,n}^{(m)}$ in the standard framework is upper bounded by

$$\tilde{p}_{e,n}^{(m)} < \sqrt{p_{e,n}^{(m)}}. \tag{11}$$

Note that the encoding in the standard framework can be realized by mapping each message i_m to $\theta_{i_m,n}$. Hence, if R_m is achievable in the sense of this section, then R_m is also achievable in the meaning of the standard framework. See [4] and [10] for the details of the proof of Lemma 2. Also note that since M independent message points $(\Theta_1, \Theta_2, \cdots, \Theta_M)$ are used in the encoding function g_n defined by (2), each message i_m can be mapped to the message point $\theta_{i_m,n}$ independently from other messages $i_{m'}$, $m' \ne m$. Therefore, Lemma 2 holds for the case of BCs in the same way as the case of point-to-point communication treated in [10].

3. A Time-Varying Coding Scheme for AWGN-BCs with Feedback

In this section, we propose *a time-varying coding scheme* for AWGN-BCs with feedback.

3.1 Encoding Scheme

Assume that the sender wants to send M messages $\{\Theta_m\}_{m=1}^M$ to M users, respectively, where Θ_m satisfying $\Theta_m \sim \mathcal{U}$ is the message for user m and Θ_m is independent of $\Theta_{m'}$ for $m' \ne m$.

Initialization at $n = 1$.[†]

For each m, $1 \le m \le M$:

[†]We use M time slots for the initialization. But for simplicity of notation, $n = 1$ is assigned for these M time slots.

- The encoder broadcasts a message $S_1^{(m)} = F_S^{-1}(\Theta_m)$, where $S \sim \mathcal{N}(0, P_0)$, and $P_0 > 0$ is determined based on the channel situation.
- User m receives $Y_1^{(m)} = S_1^{(m)} + Z_1 + Z_1^{(m)}$ and feedbacks $Y_1^{(m)}$ to the encoder.

Recursion for $n \geq 2$.

- The encoder creates a random variables $S_n^{(m)}$ defined by

$$S_n^{(m)} = \frac{1}{a_{n-1}^{(m)}} \left(S_{n-1}^{(m)} - b_{n-1}^{(m)} Y_{n-1}^{(m)} \right), \tag{12}$$

  where $a_{n-1}^{(m)} > 0$ and $b_{n-1}^{(m)}, m = 1, 2, ..., M$, are real numbers which are also chosen based on the channel situation.
- The encoder broadcasts the following signal to all the users:

$$X_n = \beta_n \sum_{m=1}^{M} \alpha_n^{(m)} S_n^{(m)}. \tag{13}$$

  Here, β_n is a real number, which is chosen to satisfy the input power constraint (7), and

$$\boldsymbol{\alpha}_n = [\ \alpha_n^{(1)} \quad \alpha_n^{(2)} \quad \cdots \quad \alpha_n^{(M)}\ ]^T \tag{14}$$

  is a modulated vector.

- User m receives the signal

$$Y_n^{(m)} = \beta_n \sum_{m=1}^{M} \alpha_n^{(m)} S_n^{(m)} + Z_n + Z_n^{(m)}, \tag{15}$$

  and it feedbacks $Y_n^{(m)}$ to the encoder.

3.2 Decoding Scheme

Recursion for $n \geq 2$:

- Each user m receives $Y_n^{(m)}$ given by (15).
- Each user m selects a fixed interval $J_1^{(m)} = (s_m, t_m) \subset \mathbb{R}$ with respect to $S_n^{(m)}$.
- Then, each user m estimates the interval $J_n^{(m)}$ for the $S_1^{(m)}$ as follows.

$$J_n^{(m)} = \left(T_n^{(m)}(s_m), T_n^{(m)}(t_m) \right) \tag{16}$$

  where

$$T_n^{(m)}(x) \equiv w_1^{(m)} \circ w_2^{(m)} \cdots \circ w_n^{(m)}(x) \tag{17}$$

  and

$$w_n^{(m)}(x) \equiv a_n^{(m)} x + b_n^{(m)} Y_n^{(m)}. \tag{18}$$

  Note that $a_n^{(m)} > 0$ ensures that $w_n^{(m)}(x)$ and $T_n^{(m)}(x)$ are monotonically increasing in x for any realization $\mathbf{y}^{n(m)}$ of $\mathbf{Y}^{n(m)}$.
- Finally, the decoded interval $\Delta_n^{(m)}(\mathbf{Y}^{n(m)})$ is determined for Θ_m as follows:

$$\Delta_n^{(m)}(\mathbf{Y}^{n(m)}) \equiv F_S \left(J_n^{(m)} \right), \tag{19}$$

  where $S \sim \mathcal{N}(0, P_0)$, and for the p.d.f. $f_S(t)$ of S,

$$F_S((a, b)) \equiv \left(\int_{-\infty}^{a} f_S(x)dx, \quad \int_{-\infty}^{b} f_S(x)dx \right). \tag{20}$$

4. Error Analysis for the Time-Varying Coding Scheme for AWGN-BCs with Feedback

In this section, we evaluate the performance of the time-varying posterior matching scheme proposed in Sect. 3.

Theorem 1: The time-varying coding scheme for the AWGN-BC given by Fig. 1 achieves any rate tuple $(R_1, R_2, ..., R_M)$ if it satisfies

$$R_m < R_m^* \equiv -\limsup_{n\to\infty} \log a_n^{(m)} \tag{21}$$

for $0 < \limsup_{n\to\infty} a_n^{(m)} < 1$ and $W_n^{(m)} \equiv \mathbb{E}[S_n^{(m)}]^2$ is upper bounded. Furthermore, the error probability $p_{n,e}^{(m)}$ satisfies that for every $m \in \{1, 2, ..., M\}$[†],

$$-\log p_{n,e}^{(m)} = o\left(2^{2n(R_m^* - R_m)}\right). \tag{22}$$

Remark 2: Eq. (22) means that $p_{n,e}^{(m)}$ can go to zero in the following way[††]:

$$\begin{aligned} p_{n,e}^{(m)} &\leq \kappa \exp_2\left(-\frac{2^{2n(R_m^* - R_m)}}{u(n)}\right) \\ &= \kappa \exp_2\left(-2^{2n(R_m^* - R_m - \frac{1}{2n}\log u(n))}\right) \end{aligned} \tag{23}$$

for some $\kappa > 0$ and any $u(n)$ satisfying that $\lim_{n\to\infty} u(n) = \infty$. Hence, if we use $u(n)$ satisfying $\lim_{n\to\infty}(1/2n)\log u(n) < \eta(R_m^* - R_m)$ for some η, $0 < \eta < 1$, then $p_{n,e}^{(m)}$ can go to zero with double exponential order. More precislely, κ can be determined from (30).

Proof Let $R_n^{(m)}$ be the instant rate to transmit message Θ_m to user m. For any fixed rate R_m, we have

$$\begin{aligned} \mathbb{P}\left(R_n^{(m)} < R_m\right) &\overset{(a)}{=} \mathbb{P}\left(-\frac{1}{n}\log\left|\Delta_n^{(m)}(\mathbf{Y}^{n(m)})\right| < R_m\right) \\ &= \mathbb{P}\left(|\Delta_n^{(m)}(\mathbf{Y}^{n(m)})| > 2^{-nR_m}\right) \\ &\overset{(b)}{\leq} \mathbb{P}\left(|J_n^{(m)}| > 2^{-nR_m}/K\right) \end{aligned} \tag{24}$$

where

[†] $f_1(n) = o(f_2(n))$ means that $\lim_{n\to\infty} f_1(n)/f_2(n) = 0$.
[††] $\exp_2(n) \equiv 2^n$.

$$K = \sup_{x\in\mathbb{R}}\{f_S(x)\}. \tag{25}$$

Here, (a) follows from (4), and (b) holds from (19), (20), and (25).

Note from (18) that for all $t, s \in \mathbb{R}$, we have

$$|w_n^{(m)}(t) - w_n^{(m)}(s)| = a_n^{(m)}|t-s|. \tag{26}$$

For $a_m \equiv \limsup_{n\to\infty} a_n^{(m)}$ we have $R_m^* \equiv \log a_m^{-1} > 0$ since $0 < a_m < 1$. Hence, for any rate $R_m < R_m^*$, we can find an $\epsilon > 0$ such that $R_m < \log(a_m+\epsilon)^{-1}$ and $a_m + \epsilon < 1$. Furthermore, there exists an $N_\epsilon \in \mathbb{N}$ such that $\sup_{n>N_\epsilon} a_n^{(m)} < a_m+\epsilon$. Define $v_m \equiv \sup_{1\le n\le N_\epsilon} a_n^{(m)}$. Then, from (24) and (26), we have

$$\begin{aligned}
&\mathbb{P}\left(R_n^{(m)} < R_m\right)\\
&\le \mathbb{P}\left(|J_n^{(m)}| > 2^{-nR_m}/K\right)\\
&\overset{(a)}{\le} K2^{nR_m}\mathbb{E}\left[\mathbb{E}\left(|w_1^{(m)}\circ w_2^{(m)}\cdots\circ w_n^{(m)}(t_m)\right.\right.\\
&\quad\left.\left.-w_1^{(m)}\circ w_2^{(m)}\cdots\circ w_n^{(m)}(s_m)|\ |\mathbf{Y}_2^{n(m)}\right)\right]\\
&\overset{(b)}{\le} K2^{nR_m}v_m\mathbb{E}\left[|w_2^{(m)}\circ w_3^{(m)}\cdots\circ w_n^{(m)}(t_m)\right.\\
&\quad\left.-w_2^{(m)}\circ w_3^{(m)}\cdots\circ w_n^{(m)}(s_m)|\right]\\
&\ \vdots\\
&\overset{(c)}{\le} K2^{nR_m}v_m^{N_\epsilon}\mathbb{E}\left[|w_{N_\epsilon+1}^{(m)}\circ w_{N_\epsilon+2}^{(m)}\cdots\circ w_n^{(m)}(t_m)\right.\\
&\quad\left.-w_{N_\epsilon+1}^{(m)}\circ w_{N_\epsilon+2}^{(m)}\cdots\circ w_n^{(m)}(s_m)|\right]\\
&\ \vdots\\
&\overset{(d)}{\le} K2^{nR_m}v_m^{N_\epsilon}(a_m+\epsilon)^{(n-N_\epsilon)}|J_1^{(m)}|,
\end{aligned}\tag{27}$$

where (a) follows from Markov's inequality and the law of iterated expectations, (b) follows from (26) and $v_m \equiv \sup_{1\le n\le N_\epsilon} a_n^{(m)}$, (c) is the recursive application of (b), and (d) follows from $\sup_{n>N_\epsilon} a_n^{(m)} < a_m+\epsilon$ and the recursive applications of (b).

From (27) and $a_m+\epsilon<1$, it is easy to see that $\mathbb{P}(R_n^{(m)} < R_m) \to 0$ holds if

$$|J_1^{(m)}| = o\left(2^{n(\log(a_m+\epsilon)^{-1}-R_m)}\right). \tag{28}$$

For $Q(x) \equiv \int_x^\infty \frac{1}{\sqrt{2\pi}}e^{-t^2/2}dt$ and $W_n^{(m)} = \mathbb{E}[S_n^{(m)}]^2$, we obtain[†]

$$\begin{aligned}
p_{n,e}^{(m)} &= \mathbb{P}\left(\Theta_m \notin \Delta_n^{(m)}\left(\mathbf{Y}^{n(m)}\right)\right)\\
&= \mathbb{P}\left(\Theta_m \notin F_S(J_n^{(m)})\right)\\
&\overset{(a)}{=} \mathbb{P}\left(S_1^{(m)} \notin J_n^{(m)}\right)\\
&\overset{(b)}{=} \mathbb{P}\left(S_n^{(m)} \notin J_1^{(m)}\right)\\
&\overset{(c)}{=} 2Q\left(\frac{|J_1^{(m)}|}{2\sqrt{W_n^{(m)}}}\right)\\
&\overset{(d)}{\sim} \frac{2\sqrt{W_n^{(m)}}}{\sqrt{2\pi}|J_1^{(m)}|}\exp\left(-\frac{|J_1^{(m)}|^2}{8W_n^{(m)}}\right).
\end{aligned}\tag{29}$$

[†] $f_1(n) \sim f_2(n)$ means that $\lim_{n\to\infty} f_1(n)/f_2(n) = 1$.

Here, (a) follows from the fact that Θ_m is uniformly distributed over $(0,1)$ and this equality holds for any realization $\mathbf{y}^{n(m)}$ of the random vector $\mathbf{Y}^{n(m)}$. (b) holds from $S_n^{(m)} = w_{n-1}^{(m)^{-1}}(S_{n-1})$, which originates from (12) and (18). (c) follows from the fact that $S_n^{(m)}$ is Gaussian with $\mathbb{E}[S_n^{(m)}] = 0$, which can be shown inductively from (12) and (15), and $J_1^{(m)}$ is symmetric if we set $s_m = -t_m$. (d) follows from that $Q(x)$ satisfies

$$\begin{aligned}
\frac{1}{\sqrt{2\pi}x}\left(1-\frac{1}{x^2}\right)\exp\left(-\frac{x^2}{2}\right) &< Q(x)\\
&< \frac{1}{\sqrt{2\pi}x}\exp\left(-\frac{x^2}{2}\right)
\end{aligned}\tag{31}$$

for any $x > 0$.

From $R_m < \log(a_m+\epsilon)^{-1} < R_m^*$, we can select $J_1^{(m)}$ satisfying (28) and $|J_1^{(m)}| \to \infty$ as $n\to\infty$. Furthermore, since $W_n^{(m)}$ is upper bounded by some W, we have

$$\frac{|J_1^{(m)}|^2}{8W_n^{(m)}} \ge \frac{|J_1^{(m)}|^2}{8W} \to \infty. \tag{32}$$

More precisely by substituting (28) into (30), $p_{n,e}^{(m)}$ satisfies

$$\begin{aligned}
-\log p_{n,e}^{(m)} &\sim \frac{|J_1^{(m)}|^2}{8W_n^{(m)}}\log \mathrm{e} - \log\frac{2\sqrt{W_n^{(m)}}}{\sqrt{2\pi}|J_1^{(m)}|}\\
&\sim \frac{|J_1^{(m)}|^2}{8W_n^{(m)}}\log \mathrm{e}\\
&= o\left(2^{2n(\log(a_m+\epsilon)^{-1}-R_m)}\right).
\end{aligned}\tag{33}$$

Since the above argument holds for any sufficiently small $\epsilon > 0$, we can attain

$$-\log p_{n,e}^{(m)} = o\left(2^{2n(R_m^*-R_m)}\right). \tag{34}$$

□

Remark 3: Since we can estimate R_m^* and know our desired rate R_m in advance, it is possible to choose ϵ appropriately as a target. This means that the decoding algorithm is technically realizable. However, there is a tradeoff between the transmission rate R_m (the possible values of ϵ) and the code length n. If R_m is very close to R_m^*, ϵ must be very small. As a result, the required N_ϵ becomes very large. Furthermore, since R_m is also very close to $\log(a_m+\epsilon)^{-1}$, the error probabilities $p_{n,e}^{(m)}$ decay slowly to zero. In the sequel, a very large code length n is required if R_m is close to R_m^*. On the contrary, if $R_m^* - R_m$ is large, we can choose quite large ϵ, which makes the required N_ϵ smaller and the decay of error probabilities faster.

Remark 4: In the case of finite n, $\mathbb{P}(R_n^{(m)} < R_m)$ is not

zero even if J_m satisfies (28). But this does not worsen the error probability $p_{n,e}^{(m)}$ if retransmission is allowed. Note that since the encoder obtains $\mathbf{y}^{n(m)}$ via the feedback channel, both the encoder and decoder m can know the value of $R_n^{(m)}$ for $\mathbf{y}^{n(m)}$. Hence, they can know whether event $\{R_n^{(m)} < R_m\}$ occurred or not when they received $\mathbf{y}^{n(m)}$. If event $\{R_n^{(m)} < R_m\}$ occurs, they discard this transmission and resend the same massage i_m in the standard framework. This retransmission decreases the coding rate of massage i_m from $\tilde{R}_n^{(m)}$ to $\tilde{R}_n^{(m)}(1 - \mathbb{P}(R_n^{(m)} < R_m))$. But, this degradation of coding rate is negligible if $\mathbb{P}(R_n^{(m)} < R_m)$ is sufficiently small.

Remark 5: If we cannot use the retransmission described in Remark 4, event $\{R_n^{(m)} < R_m\}$ makes a decoding error. In this case, we need to minimize the total decoding error probability given by $p_{n,e}^{(m)} + \mathbb{P}(R_n^{(m)} < R_m)$, and hence we cannot attain double exponential order. By setting $|J_1^{(m)}|^2(\log \mathrm{e})/8W = n(\log(a_m+\epsilon)^{-1} - R_m)$ in (27) and (30), the error exponent of the total error probability is given by

$$\begin{aligned}\lim_{\epsilon\to 0}\lim_{n\to\infty}&\left[-\frac{1}{n}\log\left(p_{n,e}^{(m)} + \mathbb{P}(R_n^{(m)} < R_m)\right)\right]\\ &\geq \lim_{\epsilon\to 0}\left[\log(a_m+\epsilon)^{-1} - R_m\right]\\ &= R_m^* - R_m. \end{aligned} \tag{35}$$

5. A Variant of the Ozarow-Leung Coding Scheme for Two-User AWGN-BCs with Feedback

Denote

$$\rho_n \equiv \frac{\mathbb{E}[S_n^{(1)}S_n^{(2)}]}{P/2}. \tag{36}$$

In this case, we set

$$P_0 = P/2, \tag{37}$$
$$\alpha_n^{(1)} = 1, \tag{38}$$
$$\alpha_n^{(2)} = g\ \mathrm{sgn}(\rho_n). \tag{39}$$

Here, $\mathrm{sgn}(x) \equiv 1$ if $x \geq 0$ and $\mathrm{sgn}(x) \equiv -1$ if $x < 0$. g is a nonnegative number which allows a trade-off between R_1^* and R_2^* [1]. We also define

$$\beta_n = \sqrt{\frac{2}{1+g^2+2g|\rho_n|}}, \tag{40}$$
$$a_n^{(1)} = \sqrt{\frac{\mathrm{var}(S_n^{(1)}|Y_n^{(1)})}{P/2}}, \tag{41}$$
$$a_n^{(2)} = \sqrt{\frac{\mathrm{var}(S_n^{(2)}|Y_n^{(2)})}{P/2}}, \tag{42}$$
$$b_n^{(1)} = \frac{\mathbb{E}[S_n^{(1)}Y_n^{(1)}]}{\mathrm{var}(Y_n^{(1)})}, \tag{43}$$
$$b_n^{(2)} = \frac{\mathbb{E}[S_n^{(2)}Y_n^{(2)}]}{\mathrm{var}(Y_n^{(2)})}. \tag{44}$$

By substituting (41)–(44) into (12), we can show for $m = 1$ and 2 that

$$S_{n+1}^{(m)} = F_S^{-1} \circ F_{S_n^{(m)}|Y_n^{(m)}}(S_n^{(m)}|Y_n^{(m)}). \tag{45}$$

(see [13], [14]).

From Lemma 1, each realization $y_n^{(m)}$ of $Y_n^{(m)}$ satisfies that

$$F_{S_n^{(m)}|Y_n^{(m)}}(S_n^{(m)}|y_n^{(m)}) \sim \mathcal{U}, \tag{46}$$

which means

$$F_{S_n^{(m)}|Y_n^{(m)}}(S_n^{(m)}|Y_n^{(m)}) \sim \mathcal{U}. \tag{47}$$

Since $S \sim \mathcal{N}(0, P_0) = \mathcal{N}(0, P/2)$, we have $S_1^{(m)} = F_S^{-1}(\Theta_m) \sim \mathcal{N}(0, P/2)$ from Lemma 1. Repeating this procedure we obtain

$$S_n^{(m)} \sim \mathcal{N}(0, P/2) \tag{48}$$

for any $n \geq 1$. In addition, we have from (13) and (15) that

$$\begin{aligned} X_n &= \beta_n[S_n^{(1)}\alpha_n^{(1)} + S_n^{(2)}\alpha_n^{(2)}]\\ &= \beta_n[S_n^{(1)} + g\ \mathrm{sgn}(\rho_n)S_n^{(2)}], \end{aligned} \tag{49}$$
$$Y_n^{(1)} = \beta_n[S_n^{(1)} + g\ \mathrm{sgn}(\rho_n)S_n^{(2)}] + Z_n + Z_n^{(1)}, \tag{50}$$
$$Y_n^{(2)} = \beta_n[S_n^{(1)} + g\ \mathrm{sgn}(\rho_n)S_n^{(2)}] + Z_n + Z_n^{(2)}. \tag{51}$$

Since $S_n^{(m)}$ satisfies $\mathbb{E}[S_n^{(m)}] = 0$ from (48), we have

$$\mathbb{E}[X_n] = \mathbb{E}[Y_n^{(1)}] = \mathbb{E}[Y_n^{(2)}] = 0. \tag{52}$$

Furthermore, from (48) we also have $\mathbb{E}[(S_n^{(m)})^2] = P/2$. Hence,

$$\mathbb{E}[X_n^2] = P, \tag{53}$$
$$\mathbb{E}[S_n^{(1)}Y_n^{(1)}] = (P/2)\beta_n(1 + g|\rho_n|), \tag{54}$$
$$\mathbb{E}[S_n^{(2)}Y_n^{(2)}] = (P/2)\beta_n\mathrm{sgn}(\rho_n)(g + |\rho_n|), \tag{55}$$
$$\mathrm{var}(Y_n^{(1)}) = P + \sigma^2 + \sigma_1^2, \tag{56}$$
$$\mathrm{var}(Y_n^{(2)}) = P + \sigma^2 + \sigma_2^2. \tag{57}$$

Note that the following relations hold. (Refer, e.g. [19, page 323].)

$$\mathrm{var}(S_n^{(1)}|Y_n^{(1)}) = \mathrm{var}(S_n^{(1)}) - \frac{(\mathbb{E}[S_n^{(1)}Y_n^{(1)}])^2}{\mathrm{var}(Y_n^{(1)})}, \tag{58}$$
$$\mathrm{var}(S_n^{(2)}|Y_n^{(2)}) = \mathrm{var}(S_n^{(2)}) - \frac{(\mathbb{E}[S_n^{(2)}Y_n^{(2)}])^2}{\mathrm{var}(Y_n^{(2)})}. \tag{59}$$

Substituting (53)–(59) into (41)–(44), we finally have

$$a_n^{(1)} = \sqrt{\frac{\sigma^2 + \sigma_1^2 + (Pg^2(1-\rho_n^2))/(1+g^2+2g|\rho_n|)}{P+\sigma^2+\sigma_1^2}}, \tag{60}$$

$$a_n^{(2)} = \sqrt{\frac{\sigma^2+\sigma_2^2+(P(1-\rho_n^2))/(1+g^2+2g|\rho_n|)}{P+\sigma^2+\sigma_2^2}}, \tag{61}$$

$$b_n^{(1)} = \frac{(P/2)\beta_n(1+g|\rho_n|)}{P+\sigma^2+\sigma_1^2}, \tag{62}$$

$$b_n^{(2)} = \frac{(P/2)\beta_n \mathrm{sgn}(\rho_n)(g+|\rho_n|)}{P+\sigma^2+\sigma_2^2}. \tag{63}$$

From (12) for $m = 1$ and 2, we have

$$\mathbb{E}[S_{n+1}^{(1)}S_{n+1}^{(2)}] = \frac{1}{a_n^{(1)}a_n^{(2)}}\left(\mathbb{E}[S_n^{(1)}S_n^{(2)}] - b_n^{(1)}\mathbb{E}[S_n^{(2)}Y_n^{(1)}] - b_n^{(2)}\mathbb{E}[S_n^{(1)}Y_n^{(2)}] + b_n^{(1)}b_n^{(2)}\mathbb{E}[Y_n^{(1)}Y_n^{(2)}]\right). \tag{64}$$

By substituting (36) and (60)–(63) into (64) and some calculations, ρ_n must satisfy

$$\rho_{n+1} = \frac{A\rho_n - \frac{PB}{D(|\rho_n|)}(g+|\rho_n|)(1+g|\rho_n|)\mathrm{sgn}(\rho_n)}{\sqrt{A}\sqrt{\left(\sigma^2+\sigma_1^2+\frac{Pg^2(1-\rho_n)^2}{D(|\rho_n|)}\right)\left(\sigma^2+\sigma_2^2+\frac{P(1-\rho_n^2)}{D(|\rho_n|)}\right)}}, \tag{65}$$

where

$$A = (P+\sigma^2+\sigma_1^2)(P+\sigma^2+\sigma_2^2), \tag{66}$$

$$B = P+\sigma^2+\sigma_1^2+\sigma_2^2, \tag{67}$$

$$D(x) = 1+g^2+2gx. \tag{68}$$

It is very difficult to affirm that the sequence $|\rho_n|$ is convergent. One strategy to overcome this difficulty is to keep $|\rho_n|$ unchanged (see [1]). Hence, we set $\rho_n = (-1)^{n+1}\rho$, where ρ is the biggest solution in $(0,1)$ of the following equation:

$$x + \frac{Ax - \frac{PB}{D(x)}(g+x)(1+gx)}{\sqrt{A}\sqrt{\left(\sigma^2+\sigma_1^2+\frac{Pg^2(1-x)^2}{D(x)}\right)\left(\sigma^2+\sigma_2^2+\frac{P(1-x^2)}{D(x)}\right)}} = 0. \tag{69}$$

Note that (69) has a solution in $(0,1)$ since the left hand side of (69) is negative at $x = 0$ and positive at $x = 1$.

Then, we have from (60) and (61) that

$$\limsup_{n\to\infty} a_n^{(1)} = \sqrt{\frac{\sigma^2+\sigma_1^2+(Pg^2(1-\rho^2))/(1+g^2+2g\rho)}{P+\sigma^2+\sigma_1^2}}, \tag{70}$$

$$\limsup_{n\to\infty} a_n^{(2)} = \sqrt{\frac{\sigma^2+\sigma_2^2+(P(1-\rho^2))/(1+g^2+2g\rho)}{P+\sigma^2+\sigma_2^2}}. \tag{71}$$

It is easy to verify that $0 < \limsup_{n\to\infty} a_n^{(m)} < 1$ for $m = 1$ and 2. Hence, from Theorem 1 the proposed scheme achieves any rate-pair (R_1, R_2) if

$$R_1 < R_1^* = -\limsup_{n\to\infty}\log a_n^{(1)} = \frac{1}{2}\log\left(\frac{P+\sigma^2+\sigma_1^2}{\sigma^2+\sigma_1^2+(Pg^2(1-\rho^2))/D(\rho)}\right), \tag{72}$$

$$R_2 < R_2^* = -\limsup_{n\to\infty}\log a_n^{(2)} = \frac{1}{2}\log\left(\frac{P+\sigma^2+\sigma_1^2}{\sigma^2+\sigma_2^2+(P(1-\rho^2))/D(\rho)}\right). \tag{73}$$

The error probabilities decay to zero as

$$-\log p_{n,e}^{(1)} = o\left(2^{2n(R_1^*-R_1)}\right), \tag{74}$$

$$-\log p_{n,e}^{(2)} = o\left(2^{2n(R_2^*-R_2)}\right). \tag{75}$$

Remark 6: The encoding scheme for $M = 2$ treated in this section is a variant of the Ozarow-Leung coding scheme [1] which is represented by a form of time-varying posterior matching [13], [14]. However, the performance of this code is worse than the one of the LQG code [4] and the Elia code [3]. Using the same approach, we can obtain a variant of the Kramer code [2] for $M > 2$. In the next section, we show that by choosing sequences $\alpha_n^{(m)}, b_n^{(m)}$ appropriately, we can achieve larger coding rate for $M \geq 2$. Specifically, we show that our proposed coding scheme for the symmetric AWGN-BCs with feedback attains the linear-feedback sum-capacity like the LQG code [4], which is larger than the achievable sum-rate of the Kramer code [2].

Remark 7: The Amor-Steinberg-Wigger (ASW) coding scheme [6] for 2-user asymmetric AWGN-BCs is constructed by a rearrangement of the Ozarow coding scheme for 2-user AWGN-MACs [7], where two messages are assigned to two vectors with different powers and the power of each message can vary at each time n. See [6, (189)]. But, the variant of the Ozarow-Leung coding scheme treated in this section uses a constant power at every time n as shown in (48). Therefore, for the 2-user asymmetric case, our coding scheme is generally inferior to the ASW coding scheme. However, in the 2-user *symmetric* case, our coding scheme can attain the linear-feedback sum-capacity, like the ASW coding scheme, as shown in Sect. 6. We conjecture that by choosing appropriately $a_n^{(m)}, b_n^{(m)}, \beta_n$ in general setting given by (12), (13), our coding scheme can also attain the same coding rates as the ASW coding scheme for the 2-user asymmetric case. Furthermore, it is expected that our coding scheme can be extended to the M-user asymmetric AWGN-BC channels with feedback easier than the ASW coding scheme because our scheme works for *real* AWGN-BC channels, but Kramer's MAC coding scheme [2], which is a generalization of the

Ozarow MAC coding scheme, uses a *complex* modulation. These extensions are interesting future works.

6. M-User Physically Non-Degraded Symmetric AWGN-BC with Feedback

In this section, we consider a physically non-degraded symmetric AWGN-BC with $\sigma_1^2 = \sigma_2^2 = \ldots = \sigma_M^2 = 1$ and $\sigma^2 = 0$. For this case, the following theorem holds.

Theorem 2: For the M-user physically non-degraded symmetric AWGN-BC with feedback satisfying $\sigma_1^2 = \sigma_2^2 = \cdots = \sigma_M^2 = 1$ and $\sigma^2 = 0$, the time-varying coding scheme proposed in Sect. 3 can achieve the linear-feedback sum-capacity, i.e. the sum-rate R_{sum} satisfying

$$R_{\text{sum}} = \sum_{m=1}^{M} R_m^* = \frac{1}{2}\log(1 + P\lambda), \tag{76}$$

where λ is the biggest solution in $[1, M]$ of the following equation:

$$(P\lambda + 1)^{M-1} = [(P/M)\lambda(M - \lambda) + 1]^M. \tag{77}$$

Theorem 2 will be proved in Sect. 7. The sum-rate given by (76) coincides with the sum-rate of the LGQ code [4, Theorem 2], which is the linear-feedback sum-capacity of the symmetric AWGN-BC treated in this section [6, Corollary 5].

From this theorem, like the MAC case, we can prove that for large M,

$$\sum_{m=1}^{M} R_m^* \approx \frac{1}{M}\log M + \frac{1}{2}\log\log M. \tag{78}$$

Refer [2, (72)] for details. This means that the difference of the sum-rate of AWGN-BC with between feedback and no feedback grows as $(\log\log M)/2$ similar to the case of MACs.

Next we derive the tight upper bounds of $p_{n,e}^{(m)}$ and $\mathbb{P}\left(R_n^{(m)} < R_m\right)$ for this symmetric case. Since $a_n^{(m)}$ can be fixed as $a_n^{(m)} = a$ for all m and n in this case as we will show in Sect. 7, it holds in (26) that

$$|w_n^{(m)}(t) - w_n^{(m)}(s)| = a|t - s|. \tag{79}$$

This means that we do not need to use ϵ in (27) in this case. Therefore, from (27) and (29), if we choose

$$|J_1^{(m)}| = \frac{2\sqrt{W}2^{n(-\log a - R_m)}}{u(n)} = o\left(2^{n(-\log a - R_m)}\right) \tag{80}$$

for any $u(n)$ and some constant W such that $\lim_{n\to\infty} u(n) = \infty$ and $W_n^{(m)} \le W$ for all n and m, we can construct the coding scheme satisfying

$$p_{n,e}^{(m)} \le 2Q\left(\frac{2^{n(-\log a - R_m)}}{u(n)}\right), \tag{81}$$

$$\mathbb{P}\left(R_n^{(m)} < R_m\right) \le K2^{nR_m}a^n|J_1^{(m)}| = \frac{2K\sqrt{W}}{u(n)}. \tag{82}$$

Remark 8: In the symmetric case treated in this section, it holds from (31) and (81) that for $R^* \equiv -\log a$,

$$-\log p_{n,e}^{(m)} = o\left(2^{2n(R^* - R_m)}\right). \tag{83}$$

Furthermore, it also holds from (35) that

$$\lim_{n\to\infty}\left[-\frac{1}{n}\log\left(p_{n,e}^{(m)} + \mathbb{P}(R_n^{(m)} < R_m)\right)\right] \ge R^* - R_m. \tag{84}$$

Since (81) gives the tight upper bound of $p_{e,n}^{(m)}$, we can know how many n is required to achieve the targets of $p_{n,e}^{(m)}$. On the other hand, the LQG code satisfies

$$p_{n,e}^{(m)} \le 4 \times 2^{-2n(-\log a - R_m - \epsilon_n)} \tag{85}$$

where $\epsilon_n \to 0$ as $n \to \infty$ [4, (48)]. However, we cannot know necessary n for the targets of $p_{n,e}^{(m)}$ in this code because the error exponent and achievable mean square error (MSE) exponents are only given in asymptotic settings. The same holds for the Elia code [3].

It is also worth noting that since our encoding scheme is a variant of the Kramer code, it has potential to achieve not only the symmetric capacity but also a good performance in asymmetric settings [9]. But it is very difficult for the LQG approach to treat the asymmetric setting.

7. Proof of Theorem 2

The following Lemmas 3 and 4 play important roles to prove Theorem 2.

Lemma 3: Let $\lambda^{(1)}, \lambda^{(2)}, \ldots, \lambda^{(M)}$ be a set of positive numbers satisfying:

$$\lambda^{(m+1)} = \frac{1 + (P/M)\lambda(M - \lambda)}{1 + P\lambda}\lambda^{(m)} \tag{86}$$

for $m = 1, 2, \ldots, M-1$, where $\lambda^{(1)} = \lambda$ is the biggest positive root of (77). Assuming that γ is a negative number satisfying

$$\gamma \ge -\frac{\lambda}{P\lambda + 1}, \tag{87}$$

then, we have $\lambda^{(m)} + \gamma > 0$ for all m.

Proof From (77) and (86), we have

$$\begin{aligned}\lambda^{(M)} &= \frac{[1 + (P/M)\lambda(M - \lambda)]^{M-1}}{(1 + P\lambda)^{M-1}}\lambda^{(1)} \\ &= \frac{1}{1 + (P/M)\lambda(M - \lambda)}\lambda^{(1)} \\ &= \frac{\lambda}{1 + (P/M)\lambda(M - \lambda)}\end{aligned} \tag{88}$$

Combining (88) with (87), we obtain

$$\begin{aligned}\lambda^{(M)}+\gamma &\geq \lambda^{(M)}-\frac{\lambda}{P\lambda+1}\\ &=\frac{\lambda}{1+(P/M)\lambda(M-\lambda)}-\frac{\lambda}{P\lambda+1}\\ &=\frac{(P/M)\lambda^3}{(1+P\lambda)(1+(P/M)\lambda(M-\lambda))}>0.\end{aligned} \tag{89}$$

Moreover, from (86) we have for all $m=1,2,...,M-1$ that

$$\begin{aligned}\lambda^{(m+1)}&=\frac{1+P\lambda-(P/M)\lambda^2}{1+P\lambda}\lambda^{(m)}\\ &\leq \lambda^{(m)}.\end{aligned} \tag{90}$$

This means that

$$\lambda^{(m)}+\gamma \geq \lambda^{(M)}+\gamma>0, \quad \forall m=1,2,...,M. \tag{91}$$

□

Lemma 4: For any positive number λ, the following simultaneous equations have a unique solution pair (b,γ) in $b>0$.

$$\gamma=\frac{P\lambda+1}{1+(P/M)\lambda(M-\lambda)}\left[\gamma+\frac{M}{P}b^2\right], \tag{92}$$

$$\gamma=\frac{1}{4b^2}\left[Mb^2+\frac{(P/M)\lambda^2}{1+P\lambda}\right]^2-\lambda. \tag{93}$$

Moreover, we have

$$0>\gamma\geq-\frac{\lambda}{1+P\lambda}, \tag{94}$$

$$Mb^2-2b\sqrt{\lambda+\gamma}+\frac{(P/M)\lambda^2}{1+P\lambda}=0. \tag{95}$$

Proof Eq. (93) is equivalent to (95), and (92) is equivalent to

$$\gamma=-\frac{(M/P)^2b^2(P\lambda+1)}{\lambda^2}. \tag{96}$$

Substituting (96) into (93), we have

$$-\frac{(M/P)^2b^2(P\lambda+1)}{\lambda^2}=\frac{1}{4b^2}\left[Mb^2+\frac{(P/M)\lambda^2}{1+P\lambda}\right]^2-\lambda, \tag{97}$$

which means

$$\begin{aligned}&\left[M^2+4\frac{(M/P)^2(P\lambda+1)}{\lambda^2}\right]b^4-2\left[\frac{P\lambda^2+2\lambda}{1+P\lambda}\right]b^2\\ &\quad+\frac{(P/M)^2\lambda^4}{(1+P\lambda)^2}=0.\end{aligned} \tag{98}$$

Since the discriminant of the above quadratic equation is equal to zero, this equation has a unique solution b^2 given by

$$b^2=\frac{(P\lambda^2+2\lambda)/(1+P\lambda)}{M^2+4[(M/P)^2(P\lambda+1)]/\lambda^2}. \tag{99}$$

Since we choose $b>0$ as the statement in Lemma 4, we get

$$b=\sqrt{\frac{(P\lambda^2+2\lambda)/(1+P\lambda)}{M^2+4[(M/P)^2(P\lambda+1)]/\lambda^2}}. \tag{100}$$

Furthermore, from (93) and $b>0$ we have $\gamma+\lambda>0$ and (95).

Since this equation has a real solution b, γ must satisfy

$$\lambda+\gamma\geq M\frac{(P/M)\lambda^2}{1+P\lambda}=\frac{P\lambda^2}{1+P\lambda}, \tag{101}$$

which means

$$\gamma\geq-\frac{\lambda}{1+P\lambda}. \tag{102}$$

On the other hand, we have $\gamma<0$ from (96). Hence (94) holds. □

Define a normalized covariance matrix by

$$\begin{aligned}\mathbf{R}_n&=\frac{1}{(P/M)}\mathbb{E}\left[\mathbf{S}_n\mathbf{S}_n^T\right]\\ &=\frac{1}{(P/M)}\begin{bmatrix}\mathbb{E}[S_n^{(1)}S_n^{(1)}] & \cdots & \mathbb{E}[S_n^{(1)}S_n^{(M)}]\\ \mathbb{E}[S_n^{(2)}S_n^{(1)}] & \cdots & \mathbb{E}[S_n^{(2)}S_n^{(M)}]\\ \vdots & \ddots & \vdots\\ \mathbb{E}[S_n^{(M)}S_n^{(1)}] & \cdots & \mathbb{E}[S_n^{(M)}S_n^{(M)}]\end{bmatrix}\\ &=\begin{bmatrix}r_n^{(1,1)} & \cdots & r_n^{(1,M)}\\ r_n^{(2,1)} & \cdots & r_n^{(2,M)}\\ \vdots & \ddots & \vdots\\ r_n^{(M,1)} & \cdots & r_n^{(M,M)}\end{bmatrix},\end{aligned} \tag{103}$$

where

$$r_n^{(m,k)}\equiv\frac{\mathbb{E}[S_n^{(m)}S_n^{(k)}]}{(P/M)}. \tag{104}$$

For $1\leq m\leq M$, let H_m be the m-th column vector of Hadamard matrix $\mathbf{H}$, and set vector $\boldsymbol{\alpha}_n\equiv[\ \alpha_n^{(1)},\alpha_n^{(2)},\cdots,\alpha_n^{(M)}\ ]^T=H_{(n-1 \bmod M)+1}$. In addition, we also set $b_n^{(m)}=b_n\,\alpha_n^{(m)}$ for each m where $\{b_n\}$ is a real sequence. We define a related matrix $\mathbf{G}_n$ by

$$\mathbf{G}_n=\mathbf{R}_n-\gamma_n\mathbf{I}_M, \tag{105}$$

where $\{\gamma_n\}$ is another real sequence.

Let $\lambda^{(1)},\lambda^{(2)},...,\lambda^{(M)}$ be the set of the positive numbers defined in Lemma 3. We first show by induction that if $\mathbf{G}_M$ is symmetric positive definite and all column vectors of $M\times M$ Hadamard matrix are eigenvectors of $\mathbf{G}_M$, then by suitably choosing sequences b_n,γ_n,β_n for all $n\geq M$, matrices $\mathbf{G}_n$ also satisfy the same properties. In addition, in this case, if $\lambda_M^{(1)}=\lambda^{(1)},\lambda_M^{(2)}=\lambda^{(2)},...,\lambda_M^{(M)}=\lambda^{(M)}$, we also have $\lambda_n^{(1)}=\lambda^{(1)},\lambda_n^{(2)}=\lambda^{(2)},...,\lambda_n^{(M)}=\lambda^{(M)}$ for all $n\geq M$. Here, $\lambda_n^{(m)}$ is the eigenvalue determined by the $[(n+m-2) \bmod M+1]$-th column vector of $M\times M$ Hadamard matrix for each $m=1,2,...,M$. For notation

simplicity, denote by $\lambda_n = \lambda_n^{(1)}$, and $\lambda = \lambda^{(1)}$, hereafter.

We first show that if $\mathbf{G}_n$ is symmetric definite and $\lambda_n^{(m)} = \lambda^{(m)}$ for $1 \le m \le M$, then $\mathbf{G}_{n+1}$ and $\lambda_{n+1}^{(m)}$ satisfying the same property. Denote

$$\mathbf{G}_n \equiv \begin{bmatrix} \rho_n^{(1,1)} & \cdots & \rho_n^{(1,M)} \\ \rho_n^{(2,1)} & \cdots & \rho_n^{(2,M)} \\ \vdots & \ddots & \vdots \\ \rho_n^{(M,1)} & \cdots & \rho_n^{(M,M)} \end{bmatrix}. \tag{106}$$

Then from (105), we obtain

$$\rho_n^{(m,k)} = r_n^{(m,k)} - \gamma_n \delta(m-k), \tag{107}$$

where $\delta(n) = 1$ if $n = 0$ and $\delta(n) = 0$ if $n \neq 0$. Since in our encoding scheme, X_n is given by (13), and $\mathbf{R}_n$ and $\mathbf{G}_n$ are defined by (103) and (105), respectively, the expected input power at time n, $\mathbb{E}[X_n^2]$, can be represented by

$$\begin{aligned} \mathbb{E}[X_n^2] &= \beta_n^2 \frac{P}{M} \boldsymbol{\alpha}_n^T \mathbf{R}_n \boldsymbol{\alpha}_n \\ &= \beta_n^2 \frac{P}{M} \left[\boldsymbol{\alpha}_n^T \mathbf{G}_n \boldsymbol{\alpha}_n + \gamma_n \boldsymbol{\alpha}_n^T \mathbf{I}_n \boldsymbol{\alpha}_n \right] \\ &= \beta_n^2 \frac{P}{M} \left[M\lambda + \gamma_n M \right] \\ &= P\beta_n^2(\lambda + \gamma_n), \end{aligned} \tag{108}$$

where the third equality holds from the fact that $\mathbf{G}_n \boldsymbol{\alpha}_n = \lambda \boldsymbol{\alpha}_n$ and $\boldsymbol{\alpha}_n^T \boldsymbol{\alpha}_n = ||\boldsymbol{\alpha}_n||_2^2 = M$.

On the other hand, since the relation between $S_n^{(m)}$ and $Y_n^{(m)}$ is given by (15) with $Z_n = 0$ and $\mathbb{E}[Z_n^{(m)}] = 0$, we obtain that

$$\begin{aligned} \mathbb{E}[S_n^{(m)} Y_n^{(k)}] &= \mathbb{E}\left[S_n^{(m)} \left(\beta_n \sum_{t=1}^{M} \alpha_n^{(t)} S_n^{(t)} + Z_n^{(m)} \right) \right] \\ &= \frac{P}{M} \beta_n \sum_{t=1}^{M} \alpha_n^{(t)} r_n^{(m,t)} \\ &= \frac{P}{M} \beta_n \sum_{t=1}^{M} \alpha_n^{(t)} [\rho_n^{(m,t)} + \gamma_n \delta(m-t)] \\ &= \frac{P}{M} \beta_n \boldsymbol{\alpha}_n^T \boldsymbol{\rho}_n^{(m)} + \frac{P}{M} \beta_n \gamma_n \alpha_n^{(m)}, \end{aligned} \tag{109}$$

where the third equality holds from (107), and

$$\boldsymbol{\rho}_n^{(m)} \equiv \left[\ \rho_n^{(m,1)}, \rho_n^{(m,2)}, \cdots, \rho_n^{(m,n)} \ \right]^T. \tag{110}$$

From the assumption that $\mathbf{G}_n$ is symmetric and λ_n is the eigenvalue associated with the eigenvector $\boldsymbol{\alpha}_n$ of this matrix, we have

$$\boldsymbol{\alpha}_n^T \mathbf{G}_n = \boldsymbol{\alpha}_n^T \mathbf{G}_n^T = \boldsymbol{\alpha}_n^T \lambda, \tag{111}$$

which means

$$\boldsymbol{\alpha}_n^T \boldsymbol{\rho}_n^{(m)} = \lambda \alpha_n^{(m)}. \tag{112}$$

Substituting (112) into (109), we obtain

$$\mathbb{E}[S_n^{(m)} Y_n^{(k)}] = \frac{P}{M} \beta_n (\lambda + \gamma_n) \alpha_n^{(m)}. \tag{113}$$

Furthermore, we also obtain

$$\begin{aligned} \mathbb{E}[Y_n^{(m)} Y_n^{(k)}] &= \mathbb{E}[(X_n + Z_n^{(m)})(X_n + Z_n^{(k)})] \\ &= \mathbb{E}[X_n^2] + \mathbb{E}[Z_n^{(m)} Z_n^{(k)}] \\ &= P\beta_n^2(\lambda + \gamma_n) + \delta(m-k). \end{aligned} \tag{114}$$

Note from (12) and $b_n^{(m)} = b_n \alpha_n^{(m)}$ that if we set $a_n^{(m)} = a_n$ for all m, our transmission scheme satisfies

$$S_{n+1}^{(m)} = \frac{1}{a_n} \left(S_n^{(m)} - b_n \alpha_n^{(m)} Y_n^{(m)} \right). \tag{115}$$

Therefore, we have

$$\begin{aligned} \mathbb{E}[S_{n+1}^{(m)} S_{n+1}^{(k)}] = \frac{1}{a_n^2} \Big(&\mathbb{E}[S_n^{(m)} S_n^{(k)}] - b_n \alpha_n^{(m)} \mathbb{E}[S_n^{(k)} Y_n^{(m)}] \\ &- b_n \alpha_n^{(k)} \mathbb{E}[S_n^{(m)} Y_n^{(k)}] + b_n^2 \alpha_n^{(m)} \alpha_n^{(k)} \mathbb{E}[Y_n^{(m)} Y_n^{(k)}] \Big). \end{aligned} \tag{116}$$

Then,

$$\begin{aligned} \frac{P}{M} r_{n+1}^{(m,k)} &= \frac{1}{a_n^2} \Bigg(\frac{P}{M} r_n^{(m,k)} - b_n \alpha_n^{(m)} \frac{P}{M} \beta_n (\lambda + \gamma_n) \alpha_n^{(k)} \\ &\quad - b_n \alpha_n^{(k)} \frac{P}{M} \beta_n (\lambda + \gamma_n) \alpha_n^{(m)} \\ &\quad + b_n^2 \alpha_n^{(m)} \alpha_n^{(k)} [P\beta_n^2(\lambda + \gamma_n) + \delta(m-k)] \Bigg) \\ &= \frac{1}{a_n^2} \Bigg(\frac{P}{M} r_n^{(m,k)} - 2 b_n \beta_n \frac{P}{M} (\lambda + \gamma_n) \alpha_n^{(m)} \alpha_n^{(k)} \\ &\quad + b_n^2 \alpha_n^{(m)} \alpha_n^{(k)} [P\beta_n^2(\lambda + \gamma_n) + \delta(m-k)] \Bigg). \end{aligned} \tag{117}$$

Hence, it holds from (107) and (117) that

$$\begin{aligned} \rho_{n+1}^{(m,k)} + \gamma_{n+1} \delta(m-k) = \frac{1}{a_n^2} \Big(&\rho_n^{(m,k)} + \gamma_n \delta(m-k) \\ &- 2 b_n \beta_n (\lambda + \gamma_n) \alpha_n^{(m)} \alpha_n^{(k)} + M b_n^2 \beta_n^2 (\lambda + \gamma_n) \alpha_n^{(m)} \alpha_n^{(k)} \\ &+ \frac{M}{P} b_n^2 \alpha_n^{(m)} \alpha_n^{(k)} \delta(m-k) \Big). \end{aligned} \tag{118}$$

Now, if we use

$$\gamma_{n+1} = \frac{1}{a_n^2} \left(\gamma_n + \frac{M}{P} b_n^2 \right), \tag{119}$$

then for all m, k we have

$$\begin{aligned} &\gamma_{n+1} \delta(m-k) \\ &\quad = \frac{1}{a_n^2} \left(\gamma_n \delta(m-k) + \frac{M}{P} b_n^2 \alpha_n^{(m)} \alpha_n^{(k)} \delta(m-k) \right). \end{aligned} \tag{120}$$

Combining (120) with (118), we obtain

$\rho_{n+1}^{(m,k)}$

$$= \frac{1}{a_n^2}\left[\rho_n^{(m,k)} - 2b_n\beta_n(\lambda+\gamma_n)\alpha_n^{(m)}\alpha_n^{(k)} + Mb_n^2\beta_n^2(\lambda+\gamma_n)\alpha_n^{(m)}\alpha_n^{(k)}\right]$$
$$= \frac{1}{a_n^2}\left(\rho_n^{(m,k)} - \left[2b_n\beta_n(\lambda+\gamma_n) - Mb_n^2\beta_n^2(\lambda+\gamma_n)\right]\alpha_n^{(m)}\alpha_n^{(k)}\right). \quad (121)$$

Now, for all $n \geq M$, we set

$$a_n = a = \sqrt{\frac{1+(P/M)\lambda(M-\lambda)}{1+P\lambda}}, \quad (122)$$

$b_n = b$ and $\gamma_n = \gamma < 0$ where (b, γ) is given in Lemma 4. In order to satisfy the input power constraint, we set β_n as follows.

$$\beta_n = \sqrt{\frac{1}{\lambda+\gamma_n}} = \sqrt{\frac{1}{\lambda+\gamma}}. \quad (123)$$

Then, it holds from (108) and (123) that $E[X_n^2] = P$ for all $n \geq M$. In addition, we also see from (95) and (123) that

$$2b_n\beta_n(\lambda+\gamma_n) - Mb_n^2\beta_n^2(\lambda+\gamma_n) = 2b\sqrt{\lambda+\gamma} - Mb^2 = \frac{(P/M)\lambda^2}{1+P\lambda}. \quad (124)$$

Substituting (124) into (121) we obtain the following recursion:

$$\rho_{n+1}^{(m,k)} = \frac{1+P\lambda}{1+(P/M)\lambda(M-\lambda)}\rho_n^{(m,k)} - \frac{(P/M)\lambda^2}{1+(P/M)\lambda(M-\lambda)}\alpha_n^{(m)}\alpha_n^{(k)}, \quad (125)$$

which means

$$\mathbf{G}_{n+1} = \frac{1+P\lambda}{1+(P/M)\lambda(M-\lambda)}\mathbf{G}_n - \frac{(P/M)\lambda^2}{1+(P/M)\lambda(M-\lambda)}\boldsymbol{\alpha}_n\boldsymbol{\alpha}_n^T. \quad (126)$$

We easily note from (126) that when $\mathbf{G}_n$ is symmetric, $\mathbf{G}_{n+1}$ is also symmetric. Denote $\mathbf{H}_n = [\boldsymbol{\alpha}_n \quad \boldsymbol{\alpha}_{n+1} \quad \cdots \quad \boldsymbol{\alpha}_{n+M-1}]$. By our induction assumption, the column vectors of $\mathbf{H}_n$ are M linearly independent eigenvectors of $\mathbf{G}_n$. Furthermore, it holds from (126) that

$$\mathbf{H}_{n+1}^T\mathbf{G}_{n+1}\mathbf{H}_{n+1} = \frac{1+P\lambda}{1+(P/M)\lambda(M-\lambda)}\mathbf{H}_{n+1}^T\mathbf{G}_n\mathbf{H}_{n+1} - \frac{(P/M)\lambda^2}{1+(P/M)\lambda(M-\lambda)}\mathbf{H}_{n+1}^T\boldsymbol{\alpha}_n\boldsymbol{\alpha}_n^T\mathbf{H}_{n+1}. \quad (127)$$

Note that since all column vectors of $\mathbf{H}_n$ are eigenvectors of $\mathbf{G}_n$, all the column vectors of the matrix $\mathbf{H}_{n+1}$ are also eigenvectors of $\mathbf{G}_n$. Hence we has the following eigenvalue decomposition

$$\boldsymbol{\Lambda}_n = \mathbf{H}_{n+1}^T\mathbf{G}_n\mathbf{H}_{n+1}, \quad (128)$$

where $\boldsymbol{\Lambda}_n = M\mathrm{diag}(\lambda^{(2)}, \lambda^{(3)}, ..., \lambda^{(M)}, \lambda^{(1)})$, which is a diagonal matrix. We also note that

$$\mathbf{H}_{n+1}^T\boldsymbol{\alpha}_n\boldsymbol{\alpha}_n^T\mathbf{H}_{n+1} = [\boldsymbol{\alpha}_n^T\mathbf{H}_{n+1}]^T\boldsymbol{\alpha}_n^T\mathbf{H}_{n+1} = M^2\mathrm{diag}(0, 0, ..., 0, 1) \quad (129)$$

because $\boldsymbol{\alpha}_{n+M} = \boldsymbol{\alpha}_n$, and hence

$$\boldsymbol{\alpha}_n^T\mathbf{H}_{n+1} = \boldsymbol{\alpha}_n^T\begin{bmatrix}\boldsymbol{\alpha}_{n+1} & \boldsymbol{\alpha}_{n+2} & \cdots & \boldsymbol{\alpha}_{n+M}\end{bmatrix} = \begin{bmatrix}0 & 0 & \cdots & M\end{bmatrix}. \quad (130)$$

From (127)–(130), $\mathbf{H}_{n+1}^T\mathbf{G}_{n+1}\mathbf{H}_{n+1}$ must be a diagonal matrix. Hence, all column vectors of $\mathbf{H}_{n+1}$ are eigenvectors of $\mathbf{G}_{n+1}$. Moreover, we obtain from (127) that for $1 \leq m \leq M-1$,

$$\lambda_{n+1}^{(m)} = \frac{1+P\lambda}{1+(P/M)\lambda(M-\lambda)}\lambda^{(m+1)} \overset{(a)}{=} \lambda^{(m)} \quad (131)$$

and

$$\lambda_{n+1}^{(M)} = \frac{1+P\lambda}{1+(P/M)\lambda(M-\lambda)}\lambda - \frac{(P/M)\lambda^2}{1+(P/M)\lambda(M-\lambda)}M = \frac{\lambda}{1+(P/M)\lambda(M-\lambda)} \overset{(a)}{=} \frac{1}{1+(P/M)\lambda(M-\lambda)}\left[\frac{1+P\lambda}{1+(P/M)\lambda(M-\lambda)}\right]^{M-1}\cdot\lambda^{(M)} \overset{(b)}{=} \lambda^{(M)} \quad (132)$$

where (a) and (b) holds from (77) and (86), respectively.

Therefore, from (126), (131) and (132), $\mathbf{G}_{n+1}$ is symmetric positive definite and $\lambda_{n+1}^{(m)} = \lambda^{(m)}$ if $\mathbf{G}_n$ is symmetric positive definite and $\lambda_n^{(m)} = \lambda^{(m)}$.

Next, we show that $\mathbf{G}_M$ can be derived from

$$\mathbf{G_1} = \lambda_0\mathbf{I}_M \quad (133)$$

by choosing of parameters $\gamma_n, b_n, \beta_n, a_n, b_n, \lambda_0$ appropriately for $1 \leq n \leq M$. In the same way as the case of $n \geq M$, we set $a_n = a, b_n = b, \gamma_n = \gamma$ where a and (b, γ) are given by (122) and Lemma 4, respectively. But, we allow that $\lambda_n^{(m)}$ depends on n for $1 \leq n \leq M$. Then, in the same way as (121), we obtain the following relation:

$$\rho_{n+1}^{(m,k)} = \frac{1}{a^2}\left[\rho_n^{(m,k)} - \left(2b\beta_n(\lambda_n+\gamma) - Mb^2\beta_n^2(\lambda_n+\gamma)\right)\alpha_n^{(m)}\alpha_n^{(k)}\right]. \quad (134)$$

Now, we consider $d_n, 1 \le n \le M-1$, that satisfies

$$2b\beta_n(\lambda_n+\gamma) - Mb^2\beta_n^2(\lambda_n+\gamma) = \left(\frac{1-d_n}{M}\right)\lambda_n. \tag{135}$$

Then, (134) becomes

$$\rho_{n+1}^{(m,k)} = \frac{1}{a^2}\left[\rho_n^{(m,k)} - \left(\frac{1-d_n}{M}\right)\lambda_n\alpha_n^{(m)}\alpha_n^{(k)}\right], \tag{136}$$

which means

$$\mathbf{G}_{n+1} = \frac{1}{a^2}\left[\mathbf{G}_n - \left(\frac{1-d_n}{M}\right)\lambda_n\boldsymbol{\alpha}_n\boldsymbol{\alpha}_n{}^T\right]. \tag{137}$$

Furthermore, in the same way as (131) and (132), we obtain

$$\lambda_{n+1}^{(m)} = \begin{cases}(1/a^2)\lambda_n^{(m+1)}, & m = 1, 2, ..., M-1,\\ (d_n/a^2)\lambda_n^{(1)}, & m = M.\end{cases} \tag{138}$$

From (137) and (138), we note that $\mathbf{G}_n$ is symmetric positive definite for $1 \le n \le M$ if d_n is positive.

We now derive d_n and λ_0 such that $\mathbf{G}_M$ has eigenvalues $\lambda^{(1)}, \lambda^{(2)}, ..., \lambda^{(M)}$, which are defined in Lemma 3. Note that $\lambda_1^{(m)} = \lambda_0$ for $1 \le m \le M$. Hence, applying (138) $M-1$ times, we obtain

$$\begin{aligned}\lambda_M^{(m)} &= \frac{d_{m-1}}{a^{2(M-1)}}\lambda_0\\ &= \left[\frac{1+P\lambda}{1+(P/M)\lambda(M-\lambda)}\right]^{M-1} d_{m-1}\lambda_0\\ &= \frac{[1+P\lambda]^{M-1}}{[1+(P/M)\lambda(M-\lambda)]^M}\\ &\qquad \times [1+(P/M)\lambda(M-\lambda)]d_{m-1}\lambda_0\end{aligned} \tag{139}$$

where $d_0 \equiv 1$. Since λ is the solution of (77), (139) means

$$\lambda_M^{(m)} = [1+(P/M)\lambda(M-\lambda)]d_{m-1}\lambda_0. \tag{140}$$

Hence, in order to satisfy $\lambda_M^{(m)} = \lambda^{(m)}$, λ_0 and d_{m-1} must satisfy

$$[1+(P/M)\lambda(M-\lambda)]d_{m-1}\lambda_0 = \lambda^{(m)}. \tag{141}$$

Since $\lambda_M^{(1)} = \lambda^{(1)} = \lambda$ and $d_0 = 1$, we obtain

$$\lambda_0 = \frac{\lambda}{1+(P/M)\lambda(M-\lambda)} \tag{142}$$

and

$$d_{m-1} = \frac{\lambda^{(m)}}{\lambda}. \tag{143}$$

On the other hand, it holds from (86) and (122) that

$$\lambda^{(m)} = a^{2(m-1)}\lambda. \tag{144}$$

Comparing (143) with (144), we have $d_{m-1} = a^{2(m-1)}$ for all $1 \le m \le M$. This means that

$$d_n = a^{2n}, \text{ for } 1 \le n \le M-1. \tag{145}$$

To complete the proof, we need to show that (135) has a positive solution β_n for $d_n = a^{2n}$. Note that (135) has a real solution $\beta_n b$ if

$$(\lambda_n+\gamma)^2 \ge (\lambda_n+\gamma)M\frac{1-d_n}{M}\lambda_n, \tag{146}$$

i.e.,

$$(\lambda_n+\gamma)(\gamma+d_n\lambda_n) \ge 0. \tag{147}$$

From (138), (142), and $\lambda_1^{(m)} = \lambda_0$ for $1 \le m \le M-1$, λ_n satisfies

$$\begin{aligned}\lambda_n = \lambda_n^{(1)} &= \frac{1}{a^{2(n-1)}}\lambda_1^{(n)}\\ &= \frac{1}{a^{2(n-1)}}\lambda_0\\ &= \frac{1}{a^{2(n-1)}}\left[\frac{\lambda}{1+(P/M)\lambda(M-\lambda)}\right].\end{aligned} \tag{148}$$

Therefore, from (122), (145), and(148), we obtain

$$\begin{aligned}\gamma + d_n\lambda_n &= \gamma + a^{2n}\frac{1}{a^{2(n-1)}}\left[\frac{\lambda}{1+(P/M)\lambda(M-\lambda)}\right]\\ &= \gamma + a^2\left[\frac{\lambda}{1+(P/M)\lambda(M-\lambda)}\right]\\ &= \gamma + \frac{1+(P/M)\lambda(M-\lambda)}{1+P\lambda}\\ &\qquad \times\left[\frac{\lambda}{1+(P/M)\lambda(M-\lambda)}\right]\\ &= \gamma + \frac{\lambda}{1+P\lambda} \ge 0,\end{aligned} \tag{149}$$

where the last inequality follows from (94). On the other hand, since it holds from (122) that $a^2 < 1$, we have $d_n = a^{2n} < 1$. Therefore, it holds from (149) that $\gamma + \lambda_n > \gamma + d_n\lambda_n \ge 0$, which means that (147) also holds. Hence, (135) has two positive solutions $b\beta_n$ by Vieta's theorem, but we choose smaller $b\beta_n$ to reduce the transmission power.

Finally, we check that $\mathbf{R}_1$ is realizable. From (105) and $\mathbf{G}_1 = \lambda_0\mathbf{I}_M$, we have

$$\begin{aligned}\mathbf{R}_1 &= \mathbf{G}_1 + \gamma\mathbf{I}_M = (\lambda_0+\gamma)\mathbf{I}_M\\ &= \left[\frac{\lambda}{1+(P/M)\lambda(M-\lambda)} + \gamma\right]\mathbf{I}_M.\end{aligned} \tag{150}$$

Since it holds for any positive λ that $\frac{\lambda}{1+P\lambda} < \frac{\lambda}{1+(P/M)\lambda(M-\lambda)}$, we have from (149) that

$$\frac{\lambda}{1+(P/M)\lambda(M-\lambda)} + \gamma > 0. \tag{151}$$

This means that the initialized random variable $S_1^{(m)} =$

$F_S^{-1}(\Theta_m)$ used in the encoding scheme given in Sect. 3 must satisfy

$$S \sim \mathcal{N}\left(0, (P/M)\left[\frac{\lambda}{1+(P/M)\lambda(M-\lambda)}+\gamma\right]\right). \tag{152}$$

Now, we evaluate the achievable rates and error probabilities. Our encoding scheme satisfies

$$\mathbb{E}[X_n^2] = P \quad \text{for } n \geq M, \tag{153}$$

and hence by the Cesàro Mean,

$$\limsup_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} \mathbb{E}[X_k^2] = P. \tag{154}$$

This means that the input power constraint is satisfied. Furthermore, for all $n \geq M$, we also have

$$\begin{aligned}\sum_{m=1}^{M} W_n^{(m)} &\equiv \sum_{m=1}^{M} \mathbb{E}[S_n^{(m)}]^2 \\ &= \frac{P}{M}\mathrm{tr}(\mathbf{R}_n) \\ &= \frac{P}{M}\left(\mathrm{tr}(\mathbf{G}_n) + M\gamma\right) \\ &= \frac{P}{M}\left(\sum_{m=1}^{M}\lambda_m + M\gamma\right) \\ &< \infty.\end{aligned} \tag{155}$$

Hence, we have $W_n \equiv \sup_m W_n^{(m)} < \infty$ since M is finite. Furthermore, since $1 \leq \lambda \leq M$, we have

$$0 < \limsup_{n\to\infty} a_n = a = \frac{1+(P/M)\lambda(M-\lambda)}{P\lambda+1} < 1. \tag{156}$$

Therefore, since the two conditions in Theorem 1 are satisfied, any rate less than the following R_m^* is achievable.

$$\begin{aligned}R_m^* &= -\limsup_{n\to\infty} \log a_n^{(m)} \\ &= -\log a \\ &= \frac{1}{2}\log\left(\frac{1+P\lambda}{1+(P/M)\lambda(M-\lambda)}\right) \equiv R^*.\end{aligned} \tag{157}$$

Hence, the following sum-rate is achievable:

$$\begin{aligned}\sum_{m=1}^{M} R_m^* &= \frac{M}{2}\log\left(\frac{1+P\lambda}{1+(P/M)\lambda(M-\lambda)}\right) \\ &= \frac{1}{2}\log\left(\frac{1+P\lambda}{1+(P/M)\lambda(M-\lambda)}\right)^M \\ &= \frac{1}{2}\log\left(1+P\lambda\right),\end{aligned} \tag{158}$$

where λ is the biggest solution in $[1, M]$ of (77).

8. Relation between AWGN-BCs and AWGN-MACs

The time-varying coding approach can be applied to the AWGN-MAC (multiple access channel) with feedback. It is shown in [13] that the time-varying coding scheme can achieve the linear-feedback sum-capacity for AWGN-MACs [18] as with the Kramer code [2] and the LQG code [4]. Let $R_{\mathrm{MAC}}(M, P)$ denote the achievable symmetric sum-rate by the time-varying code [13] for M-sender AWGN MACs with feedback where each encoder has power constraint P. Then, it is shown in [13, Theorem III] that

$$R_{\mathrm{MAC}}(M, P) = \frac{1}{2}\log(1+MP\lambda), \tag{159}$$

where λ is the biggest solution of

$$(1+MPx)^{M-1} = (1+Px(M-x))^M. \tag{160}$$

Comparing (159) with Theorem 2, we note that

$$R_{\mathrm{BC}}(M, P) = R_{\mathrm{MAC}}(M, P/M). \tag{161}$$

This shows that when we use the time-varying code under the same sum-power constraint P, the achievable sum-rate for MAC is equal to the one for BC. This relation between MAC and BC is already pointed out in [4] and [6]. From our results, we note that the posterior matching scheme can also attain this duality between MAC and BC.

9. Conclusion

We proposed a general coding scheme based on the posterior matching for AWGN-BCs with feedback, and we derived the achievable rate region and the decoding error probability of the proposed scheme. Then, we showed that a variant of the Ozarow-Leung coding scheme can be obtained as a special case of our scheme. Furthermore, we clarified how to realize the posterior matching for the physically non-degraded symmetric AWGN-BCs, and we showed the proposed coding scheme can attain the linear-feedback sum-capacity for these symmetric AWGN-BCs.

An interesting further research topic is to find a good sequences $a_n^{(m)}, b_n^{(m)}$ to attain good performance for more general settings treated in [8] and [9].

Acknowledgment

The authors thank the associate editor and reviewers for their helpful comments. This work is supported in part by JSPS Grant-in-Aid for Scientific Research, No. 25289111.

References

[1] L.H. Ozarow and S. Leung-Yan-Cheong, "An achievable region and outer bound for the Gaussian broadcast channel with feedback," IEEE Trans. Inf. Theory, vol.IT-30, no.4, pp.667–671, July 1984.
[2] G. Kramer, "Feedback strategies for white Gaussian interference networks," IEEE Trans. Inf. Theory, vol.48, pp.1423–1438, Jan. 2002.

[3] N. Elia, "When bode meets Shannon: Control oriented feedback communication schemes," IEEE Trans. Autom. Control, vol.49, no.9, pp.1477–1488, Sept. 2004.
[4] E. Ardestanizadeh, P. Minero, and M. Franceschetti, "LQG control approach to Gaussian broadcast channels with feedback," IEEE Trans. Inf. Theory, vol.58, pp.5267–5278, April 2012.
[5] S.B. Amor and M. Wigger, "Linear-feedback MAC-BC duality for correlated BC-noises, and iterative coding," Proc. 53rd Annual Allerton Conference on Communication, Control, and Computing, pp.1502–1509, Oct. 2015.
[6] S.B. Amor, Y. Steinberg, and M. Wigger, "MIMO MAC-BC duality with linear-feedback coding schemes," IEEE Trans. Inf. Theory, vol.61, no.11, pp.5976–5998, Nov. 2015
[7] L.H. Ozarow, "The capacity of the white Gaussian multiple access channel with feedback," IEEE Trans. Inf. Theory, vol.30, no.4, pp.623–629, July 1984.
[8] M. Wigger and M. Gastpar, "The pre-log of Gaussian broadcast with feedback can be two," Proc. Int. Symp. Information Theory, pp.545–546, June 2008.
[9] 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. Inf. Theory, vol.60, no.1, pp.54–57, Jan. 2014.
[10] O. Shayevitz and M. Feder, "Optimal feedback communication via posterior matching," IEEE Trans. Inf. Theory, vol.IT-57, no.3, pp.1186–1221, March 2011.
[11] J.P.M. Schalkwijk and T. Kailath, "A coding scheme for additive noise channels with feedback part I: No bandwidth constraint," IEEE Trans. Inf. Theory, vol.12, no.2, pp.172–182, April 1966.
[12] S.B. Wicker, Error Control Systems for Digital Communication and Storage, Prentice Hall, 1995.
[13] L.V. Truong, "Posterior matching scheme for Gaussian multiple access channel with feedback," Proc. IEEE Information Theory Workshop, pp.476–480, Nov. 2014.
[14] L.V. Truong and H. Yamamoto, "On the capacity of symmetric Gaussian interference channels with feedback," Proc. Int. Symp. Information Theory, pp.201–205, June 2015.
[15] T.M. Cover, "Broadcast channels," IEEE Trans. Inf. Theory, vol.18, no.1, pp.2–14, 1972.
[16] P. Bergmans, "A simple converge for broadcast channels with additive Gaussian noise," IEEE Trans. Inf. Theory, vol.20, no.2, pp.279–280, 1974.
[17] A.E. Gamal, "The feedback capacity of degraded broadcast channels," IEEE Trans. Inf. Theory, vol.24, no.3, pp.379–381, 1978.
[18] E. Ardestanizadeh, M. Wigger, Y.H. Kim, and T. Javidi, "Linear-feedback sum-capacity for Gaussian multiple access channels," IEEE Trans. Inf. Theory, vol.58, no.1, pp.224–236, 2012.
[19] S.M. Kay, Fundamentals of Statistical Signal Processing (Estimation Theory), 1st ed., Prentice-Hall, New Jersey, 1993.
[20] J.P.M. Schalkwijk, "A coding scheme for additive noise channels with feedback part II: Band-limited signal," IEEE Trans. Inf. Theory, vol.12, no.2, pp.183–189, April 1966.
[21] N.T. Gaarder and J.K. Wolf, "The capacity region of a multiple access discrete memoryless channel can increase with feedback," IRE Trans. Inf. Theory, vol.IT-21, no.1, pp.100–102, Jan. 1975.
[22] P. Diaconis and D. Freedman, "Iterated random functions," SIAM Rev., vol.41, no.1, pp.45–76, 1999.
[23] T.M. Cover and J.A. Thomas, Elements of Information Theory, 2nd ed., John Wiley & Sons, New Jersey, 2006.
[24] O. Shayevitz and M. Feder, "Communication with feedback via posterior matching," Proc. Int. Symp. Information Theory, pp.391–395, June 2007.

Lan V. Truong was born in Quang Binh province, Vietnam. He received the B.S.E. degree in Electronics and Telecommunications from Posts and Telecommunications Institute of Technology (PTIT), Hanoi, Vietnam in 2003. After many years of working as an operation and maintenance engineer (O&M) at MobiFone Telecommunications Corporation, Hanoi, Vietnam, he resumed his graduate studies at School of Electrical and Computer Engineering (ECE), Purdue University, West Lafayette, IN, United States and obtained the M.S.E. degree in 2011. From 2013 to June 2015, he was an academic lecturer at Department of Information Technology Specialization (ITS), FPT University, Hanoi, Vietnam. Presently, he is Ph.D. student at Department of Electrical and Computer Engineering (ECE), National University of Singapore (NUS). His research interests are information theory and its applications.

Hirosuke Yamamoto was born in Wakayama, Japan, in 1952. He received the B.E. degree from Shizuoka University, Shizuoka, Japan, in 1975 and the M.E. and Ph.D. degrees from the University of Tokyo, Tokyo, Japan, in 1977 and 1980, respectively, all in electrical engineering. In 1980, he joined Tokushima University. He was an Associate Professor at Tokushima University from 1983 to 1987, the University of Electro-Communications from 1987 to 1993, and the University of Tokyo from 1993 to 1999. Since 1999, he has been a Professor at the University of Tokyo and is currently with the Department of Complexity Science and Engineering at the university. In 1989–1990, he was a Visiting Scholar at the Information Systems Laboratory, Stanford University, Stanford, CA. His research interests are in Shannon theory, data compression algorithms, and information theoretic cryptology. Dr. Yamamoto served as the Chair of IEEE Information Theory Society Japan Chapter in 2002–2003, the TPC Co-Chair of the ISITA2004, the TPC Chair of the ISITA2008, the President of the SITA (Society of Information Theory and its Applications) in 2008–2009, the President of the ESS (Engineering Sciences Society) of IEICE in 2012–2013, an Auditor of IEICE in 2016–2017, an Associate Editor for Shannon Theory, the IEEE Transactions on Information Theory in 2007–2010, Editor-in-Chief for the IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences in 2009–2011. He is a Fellow of the IEICE and the IEEE.