Upper Bounds on Error Probabilities for Continuous-Time White Gaussian Channels with Feedback

Shunsuke Ihara
Graduate School of Information Science
Nagoya University
Nagoya, 464-8601 Japan
Email: ihara@zm.commufa.jp

***Abstract*—In information transmission over additive white Gaussian channels with feedback, the use of feedback link to improve the performance of communication systems has been studied by a number of authors. It is well known that the error probability in information transmission can be substantially reduced by using feedback, namely, under the average power constraint, the error probability decreases more rapidly than the exponential of any order. Recently, for discrete-time additive white Gaussian channels, Gallager and Nakiboğlu proposed a feedback coding scheme such that the resulting error probability $P_e(N)$ decreases with an exponential order αN which is linearly increasing with block length N, where α is a positive constant. In this paper, we consider continuous-time additive white Gaussian channels with feedback. The aim is to prove a stronger result on the multiple-exponential decay of the error probability. More precisely, for *any* positive constant α, there exists a feedback coding scheme such that the resulting error probability $P_e(T)$ at time T decreases more rapidly than the exponential of order αT as $T \to \infty$.**

Index Terms - White Gaussian channel with feedback, feedback coding scheme, multiple-exponential decrease of error probability, average power constraint

I. Introduction

In information transmission over additive white Gaussian noise channels (AWGC) with feedback, it is known that the minimum error probability, under the average power constraint, converges to zero faster than the exponential of any order (cf. [1]–[7]). In 1966, Shalkwijk and Kailath [1] proposed a coding scheme and demonstrated a rather surprising result that the resulting error probability converges to zero double exponentially fast. For notational convenience we introduce the notation

$$\exp_n(x) \stackrel{\triangle}{=} \exp\{\exp_{n-1}(x)\}, \quad n = 1, 2, ...,$$

to denote the exponential function of order n, where $\exp_0(x) = x$. It has been known that, for any positive integer K, there exists a coding scheme under which the error probability $P_e(T)$ at time T decreases more rapidly than the exponential of order K, i.e.

$$P_e(T) = o\left(\frac{1}{\exp_K(T)}\right), \quad T \to \infty, \tag{1}$$

(cf. [2]–[7]). Hereafter, this kind of decrease of the error probability will be referred to as the multiple-exponential decay. Recently Gallager and Nakiboğlu [7] proposed a coding scheme for the discrete-time AWGC and successfully demonstrated the multiple-exponential decay (1) of the resulting error probability at all rates below capacity. In addition, they showed that the error probability decreases with an exponential order which is linearly increasing with block length, i.e. for *some* positive constant α, there exists a coding scheme under which the error probability $P_e(T)$ decreases as

$$P_e(T) = o\left(\frac{1}{\exp_{\lfloor \alpha T \rfloor}(T)}\right), \quad T \to \infty, \tag{2}$$

where $\lfloor x \rfloor$ denotes the maximum integer not greater than x. The results (1) and (2) due to Gallager and Nakiboğlu [7] are generalized to general discrete-time Gaussian channels where the additive noises are not necessarily white but stationary Gaussian processes ([8]).

In this paper we treat the continuous-time AWGC with feedback. The aim of this paper is to prove a stronger result on the multiple-exponential decay of the error probability. More precisely, we shall show that, for *any* positive constant α, there exists a coding scheme under which the error probability $P_e(T)$ decreases as (2). It should be emphasized that the order $\lfloor \alpha T \rfloor$ of the exponent in (2) is linear in T and the coefficient α may be taken arbitrarily large. Needless to say, the known result (1) is an easy consequence of (2).

The continuous-time AWGC is presented by

$$Y(t) = \int_0^t X(u)du + B(t), \quad t > 0, \tag{3}$$

where $X = \{X(t)\}$ is an input signal, $Y = \{Y(t)\}$ the corresponding output, and the noise $B = \{B(t)\}$ a Brownian motion. We assume that an average power constraint

$$\lim_{T\to\infty} \frac{1}{T}\int_0^T E[X(t)^2]dt \le P \tag{4}$$

is imposed on the input signal, where $P > 0$ is a constant. It is well known that the capacity C of the AWGC subject to

(4) is not increased by feedback and is equal to $C = P/2$. In the following, the terminal time T of the channel uses and the rate $R > 0$ are fixed, unless otherwise mentioned. The message $U_0 \equiv U_0(T)$ with entropy R is a random variable, independent of the noise $\{B(t)\}$, such that

$$\Pr(U_0 = m) = \frac{1}{M_T}, \quad m \in \mathcal{M}_T \equiv \{1, 2, ..., M_T\}, \tag{5}$$

where $M_T = e^{RT}$. For notational convenience, we denote $U_0(T)$ and M_T simply as U_0 and M, respectively, if no confusions occur. In this paper we assume that the feedback link is noiseless and without time-lag. The input signal $X(t)$ and the decoding message $\widehat{U}(t)$ at t are given in the forms

$$X(t) = \varphi_T(t, U_0, Y_0^{t-}) \quad \text{and} \quad \widehat{U}(t) = \psi_T(t, Y_0^t),$$

respectively, where φ_T and ψ_T are measurable functions, $Y_0^{t-} \equiv \{Y(u); u < t\}$, and $Y_0^t \equiv \{Y(u); u \le t\}$. We denote by

$$P_e(t) \stackrel{\triangle}{=} \Pr(\widehat{U}(t) \ne U_0)$$

the resulting error probability.

Let us briefly explain our coding scheme. The precise definition of the scheme will be given in Section III. To define the encoding scheme and decoding scheme, the schemes investigated in [1], [3], [6], [7] are helpful and useful. We divide the time interval $(0, T]$ into sub-intervals $\mathbf{T}_k = (T_{k-1}, T_k]$, $k = 1, 2, ...$ Although, the length $|\mathbf{T}_1| = T_1$ of the first sub-interval $\mathbf{T}_1$ should be long enough, the lengths of other intervals $\mathbf{T}_k$, $k \ge 2$, may be chosen arbitrary. For example, we may define as $T_k = T_{k-1} + \Delta$, $k \ge 2$, where $\Delta > 0$ is an arbitrary constant. The receiver decodes at time T_k and denotes by $\widehat{U}(T_k)$ the decoding message. On the interval $\mathbf{T}_{k+1}$, the transmitter sends linearly scaled versions of the decoding error $W_{k+1} = U_0 - \widehat{U}(T_k)$ with help of linear feedback. The encoding scheme on $\mathbf{T}_{k+1}$ is based on the linear filtering theory. If the message U_0 is correctly decoded at T_k, i.e. $\widehat{U}(T_k) = U_0$, then $W_{k+1} = 0$, meaning that no signals are input on the interval $\mathbf{T}_{k+1}$. This is one of basic ideas to define an optimal coding scheme under the average power constraint (cf. [3], [6], [7]).

A brief review of linear filtering theory is given in Section II. In Section III we define our coding scheme and derive a formula to calculate the resulting error probability (Theorem 1). The asymptotic behavior of the error probability is evaluated in Section IV. The main result (2) is proved in Theorem 2.

II. AWGC WITH LINEAR FEEDBACK

In this section we briefly review some known results on coding schemes based on filtering theory. Let us consider an AWGC

$$\widetilde{Y}(t) = \int_0^t \xi(u)\, du + B(t), \quad t > 0, \tag{6}$$

without feedback and a corresponding AWGC

$$Y(t) = \int_0^t \{\xi(u) - \widehat{\xi}(u)\}\, du + B(t), \quad t > 0, \tag{7}$$

with linear feedback, where $\xi = \{\xi(t)\}$ is a message, independent of the noise $B = \{B(t)\}$, with expectation $E[\xi(t)] = 0$ and covariance function $R(t, u) = E[\xi(t)\xi(u)] \in L^2([0, T]^2)$, and $\widehat{\xi}(t)$ is given in the form

$$\widehat{\xi}(t) = \int_0^t h(t, u)\, dY(u), \tag{8}$$

where $h(t, u) \in L^2([0, T]^2)$. A Volterra kernel is a function $k(t, u) \in L^2([0, T]^2)$ such that $k(t, u) = 0$, $t < u$. For the covariance function $R(t, u) \in L^2([0, T]^2)$ of $\{\xi(t)\}$, there exists a Volterra kernel $k(t, u)$ which satisfies the Wiener-Hopf equation

$$k(t, u) + \int_0^t k(t, s) R(s, u)\, ds = R(t, u), \quad 0 \le u \le t. \tag{9}$$

The kernel $k(t, u)$ is called the resolvent of $R(t, u)$. For the Volterra kernel $k(t, u)$ there exists a unique Volterra kernel $h(t, u)$, called the resolvent of $k(t, u)$, such that

$$h(t, u) - \int_u^t h(t, s) k(s, u)\, ds = k(t, u), \quad 0 \le u \le t. \tag{10}$$

The kernel $h(t, u)$ also satisfies the equation

$$h(t, u) - \int_u^t k(t, s) h(s, u) ds = k(t, u), \quad 0 \le u \le t. \tag{11}$$

We summarize fundamental properties of the schemes (6) and (7). See [9], [10] and [11, Theorem 6.3.2, 6.3.6] for the proof.

Proposition 1: (i) Let $h(t, u)$ be the solution of (10). Then the process $Y = \{Y(t)\}$ given by

$$Y(t) = \widetilde{Y}(t) - \int_0^t \int_0^s k(s, u)\, d\widetilde{Y}(u)\, ds \tag{12}$$

is the unique solution of the stochastic equation (7). Moreover, we have

$$\widehat{\xi}(t) = \int_0^t h(t, u) dY(u) = \int_0^t k(t, u) d\widetilde{Y}(u) \tag{13}$$

and

$$\mathcal{L}_t(Y) = \mathcal{L}_t(\widetilde{Y}), \tag{14}$$

where $\mathcal{L}_t(Y)$ and $\mathcal{L}_t(\widetilde{Y})$ denote the linear spaces spanned by $\{Y(s); s \le t\}$ and $\{\widetilde{Y}(s); s \le t\}$, respectively.

(ii) If $k(t, u)$ is the resolvent of $R(t, u)$, then $\widehat{\xi}(t)$ of (13) is the orthogonal projection of $\xi(t)$ on $\mathcal{L}_t(Y)$. Moreover, if $\{\xi(t)\}$ is a Gaussian process, then $\widehat{\xi}(t) = E[\xi(t)|Y_0^t]$ is the conditional expectation of $\xi(t)$ given Y_0^t.

(iii) If the covariance function $R(t, u)$ is continuous, then the resolvent $k(t, u)$ is continuous on $\{(t, u); 0 \le u \le t\}$, and the mean squared error is given by

$$E[|\xi(t) - \widehat{\xi}(t)|^2] = k(t, t), \quad t > 0. \tag{15}$$

III. Coding Scheme

In this section we propose our coding scheme. We divide the time axis $(0,T]$ into subintervals $\mathbf{T}_k = (T_{k-1}, T_k]$, $k = 1,2,...$, $(0 = T_0 < T_1 < T_2 < \cdots)$ and denote by $|\mathbf{T}_k| = T_k - T_{k-1}$ the length of the interval $\mathbf{T}_k$. On each interval $\mathbf{T}_k$, the AWGC (3) can be rewritten in the form

$$Y_k(t) = \int_0^t X_k(u)du + B_k(t), \quad 0 < t \le |\mathbf{T}_k|, \tag{16}$$

where $X_k(t) = X(t+T_{k-1})$, $Y_k(t) = Y(t+T_{k-1}) - Y(T_{k-1})$ and $B_k(t) = B(t+T_{k-1}) - B(T_{k-1})$. Note that each $B_k \equiv \{B_k(t); 0 < t \le |\mathbf{T}_k|\}$ is a Brownian motion and $B_1, B_2, ...,$ are mutually independent. We also note that, for each k, (16) presents an AWGC. On the interval $\mathbf{T}_k$, the transmitter transmits scaled versions of the decoding error

$$W_k \triangleq U_0 - \widehat{U}(T_{k-1}) \tag{17}$$

with help of linear feedback, where U_0 is the original message (see (5)) and $\widehat{U}(T_{k-1})$ is the decoded message at time T_{k-1}. The definition of $\widehat{U}(T_{k-1})$ will be given by (22). Precisely, the input signal $X_k(t)$ on $\mathbf{T}_k$ is defined by

$$X_k(t) = \sqrt{P} a_k \exp(Pt/2)(W_k - \widehat{W}_k(t)), \quad 0 < t \le |\mathbf{T}_k|, \tag{18}$$

where $\widehat{W}_k(t)$ is given by

$$a_k \widehat{W}_k(t) \triangleq \sqrt{P} \int_0^t \exp(-Pu/2) dY_k(u), \tag{19}$$

and $a_k > 0$ is a constant satisfying

$$a_k^2 E[W_k^2] = a_k^2 E[|U_0 - \widehat{U}(T_{k-1})|^2] = 1. \tag{20}$$

Then, receiving the output signal

$$Y_k(t) = \sqrt{P} a_k \int_0^t \exp(Pu/2)(W_k - \widehat{W}_k(u))du + B_k(t), \quad 0 < t \le |\mathbf{T}_k|, \tag{21}$$

the receiver reproduces at time T_k a decoded message $\widehat{U}(T_k)$ defined by

$$\widehat{U}(T_k) \triangleq \widehat{U}(T_{k-1}) + \widetilde{W}_k, \quad k = 1,2,..., \tag{22}$$

where

$$\widetilde{W}_k \triangleq \left\lceil \frac{\widehat{W}_k(|\mathbf{T}_k|)}{1 - \exp(-P|\mathbf{T}_k|)} - \frac{1}{2} \right\rceil \tag{23}$$

($\lceil x \rceil$ denotes the minimum integer not less than x) and

$$\widehat{U}(T_0) \triangleq E[U_0] = (M+1)/2. \tag{24}$$

It is clear from (17) and (22) that $\{W_k\}$ satisfies

$$W_{k+1} = W_k - \widetilde{W}_k. \tag{25}$$

So far, we have defined $\widehat{U}(t)$ only for $t = T_k$. For simplicity and completeness we define $\widehat{U}(t)$ by

$$\widehat{U}(t) \triangleq \widehat{U}(T_k), \quad T_k \le t < T_{k+1}.$$

Let us define $\xi_k(t)$ by

$$\xi_k(t) \triangleq \sqrt{P} a_k \exp(Pt/2) W_k, \quad 0 < t \le |\mathbf{T}_k|. \tag{26}$$

Then, corresponding to the feedback channel (21), the channel without feedback is presented by

$$\begin{aligned} \widetilde{Y}_k(t) &= \sqrt{P} a_k W_k \int_0^t \exp(Pu/2)du + B_k(t) \\ &= \int_0^t \xi_k(u)du + B_k(t), \quad 0 < t \le |\mathbf{T}_k|. \end{aligned} \tag{27}$$

A random variable

$$\widehat{B}_k(t) \triangleq \frac{\sqrt{P}\exp(-Pt)}{1 - \exp(-Pt)} \int_0^t \exp(Pu/2)\, dB_k(u) \tag{28}$$

will play important roles to evaluate the error probability. It is clear that $\widehat{B}_k(t)$ is a Gaussian random variable with expectation 0 and variance

$$E[\widehat{B}_k(t)^2] = \frac{1}{\exp(Pt) - 1}. \tag{29}$$

Using the known results summarized in Proposition 1, we can prove the following proposition concerning the properties of the coding scheme proposed above.

Proposition 2: (i) The message W_k is independent of the noise $B_k = \{B_k(t); t \le |\mathbf{T}_k|\}$, and $E[W_k] = 0$.
(ii) $\widehat{W}_k(t)$ is the orthogonal projection of W_k on $\mathcal{L}_t(Y_k) = \mathcal{L}_t(\widetilde{Y}_k)$ and satisfies

$$\begin{aligned} a_k \widehat{W}_k(t) &= \sqrt{P}\exp(-Pt) \int_0^t \exp(Pu/2) d\widetilde{Y}_k(u) && (30) \\ &= \{1 - \exp(-Pt)\} \left\{ a_k W_k + \widehat{B}_k(t) \right\}. && (31) \end{aligned}$$

(iii) The decoding error $W_{k+1} = U_0 - \widehat{U}(T_k)$ is presented by

$$W_{k+1} = -\left\lceil \frac{\widehat{B}_k(|\mathbf{T}_k|)}{a_k} - \frac{1}{2} \right\rceil. \tag{32}$$

(iv) We have $E[X_k(t)^2] = P$. The constraint (4) is satisfied with equality.

Proof: (i) By definition $\widehat{W}_{k-1}(|\mathbf{T}_{k-1}|)$, $\widetilde{W}_{k-1}$, $\widehat{U}(T_{k-1})$ and W_k are functions of $(U_0, B_1, ..., B_{k-1})$ which is independent of B_k. Therefore, W_k is independent of B_k. We omit the proof of $E[W_k] = 0$.
(ii) Since $E[W_k] = 0$, $E[\xi_k(t)] = 0$ and the covariance function of $\{\xi_k(t)\}$ is given by

$$R(t,u) = E[\xi_k(t)\xi_k(u)] = P\exp\{P(t+u)/2\}. \tag{33}$$

Note that the covariance function does not depend on k. For the covariance function $R(t,u)$ of (33), it is easy to show that the kernel

$$k(t,u) = P\exp\{-P(t-u)/2\}, \quad 0 < u \le t, \tag{34}$$

is the resolvent of $R(t,u)$, i.e., the solution of (9), and that the kernel

$$h(t,u) = P\exp\{P(t-u)/2\}, \quad 0 < u \le t, \tag{35}$$

is the resolvent of $k(t,u)$, i.e., the solution of (10). Therefore, we know from (13) that the orthogonal projection $\widehat{\xi}_k(t)$ of $\xi_k(t)$ on $\mathcal{L}_t(Y_k)=\mathcal{L}_t(\widetilde{Y}_k)$ is given by

$$\begin{aligned}\widehat{\xi}_k(t) &= P\exp(-Pt/2)\int_0^t \exp(Pu/2)d\widetilde{Y}_k(u)\\ &= P\exp(Pt/2)\int_0^t \exp(-Pu/2)dY_k(u). \quad (36)\end{aligned}$$

This means that $\widehat{W}_k(t)$ defined by (19) is nothing but the orthogonal projection of W_k on $\mathcal{L}_t(Y_k)=\mathcal{L}_t(\widetilde{Y}_k)$. We also know from (36) that $\widehat{W}_k(t)$ is written as (30). Eq. (31) is clear from (27), (28) and (30):

$$\begin{aligned}a_k\widehat{W}_k(t) &= P\exp(-Pt)a_kW_k\int_0^t \exp(Pu)du\\ &\quad +\sqrt{P}\exp(-Pt)\int_0^t \exp\left(Pu/2\right)dB_k(u)\\ &= \{1-\exp(-Pt)\}\left\{a_kW_k+\widehat{B}_k(t)\right\}.\end{aligned}$$

(iii) Since $W_k=U_0-\widehat{U}(T_{k-1})$ is integer-valued, (32) follows from (23), (25) and (31):

$$\begin{aligned}W_{k+1} &= W_k-\widetilde{W}_k\\ &= W_k-\left\lceil W_k+\frac{\widehat{B}_k(|\mathbf{T}_k|)}{a_k}-\frac{1}{2}\right\rceil\\ &= -\left\lceil \frac{\widehat{B}_k(|\mathbf{T}_k|)}{a_k}-\frac{1}{2}\right\rceil.\end{aligned}$$

(iv) Since $R(t,u)$ of (33) is continuous, it follows from (15) and (34) that

$$E[X_k(t)^2]=E[|\xi_k(t)-\widehat{\xi}_k(t)|^2]=k(t,t)=P.$$

We are now in a position to give a formula to calculate the error probability $Pr(\widehat{U}(T_k)\neq U_0)$.

Theorem 1: Under the coding scheme proposed in this section, the error probability $P_e(T_k)=Pr(\widehat{U}(T_k)\neq U_0)$ is given by

$$P_e(T_k)=2Q\left(\beta_k a_k\right), \quad k\geq 1, \quad (37)$$

where

$$Q(x)=\frac{1}{\sqrt{2\pi}}\int_x^\infty \exp\left(-y^2/2\right)dy \quad (38)$$

is the complementary distribution function of $N(0,1)$ and

$$\beta_k \triangleq \frac{\sqrt{\exp(P|\mathbf{T}_k|)-1}}{2}. \quad (39)$$

Proof: It is clear from (32) that

$$\begin{aligned}P_e(T_k) &= Pr\left(\widehat{U}(T_k)\neq U_0\right)=Pr\left(W_{k+1}\neq 0\right)\\ &= Pr\left(\widehat{B}_k(|\mathbf{T}_k|)\notin\left(-\frac{a_k}{2},\frac{a_k}{2}\right]\right). \quad (40)\end{aligned}$$

Since $\widehat{B}_k(|\mathbf{T}_k|)\sim N(0,(2\beta_k)^{-2})$ (see (29)), (37) follows from (40).

IV. Asymptotic Behavior of Error Probability

In this section we evaluate the asymptotic behavior of the error probability and prove our main result (2) in the following theorem.

Theorem 2: Assume that the AWGC (3) with feedback is subject to the average power constraint (4) and that the rate R is less the capacity C. Then, for any constant $\alpha>0$, there exists a coding scheme such that the resulting error probability $P_e(T)=Pr(\widehat{U}(T)\neq U_0)$ satisfies

$$\lim_{T\to\infty} P_e(T)\,\exp_{\lfloor \alpha T\rfloor}(T)=0. \quad (41)$$

For notational convenience, we denote as $f(x)\sim g(x)$ if $\lim_{x\to\infty} f(x)/g(x)=1$. The asymptotic behavior of the function $Q(x)$ defined by (38) is

$$Q(x)\sim\frac{1}{\sqrt{2\pi}x}\exp\left(-x^2/2\right), \quad x\to\infty.$$

In particular, there exists a constant $x_0>0$ such that

$$2Q(x)<\exp\left(-x^2/2\right), \quad \forall x\geq x_0. \quad (42)$$

The following lemma due to Gallager and Nakiboğlu [7] is useful to evaluate the asymptotic behavior of the error probability (see [7] for the proof).

Lemma 1: Let η be a Gaussian random variable with distribution $N(0,\sigma^2)$ and define $\widetilde{\eta}$ by $\widetilde{\eta}=\left\lceil\frac{\eta}{c}-\frac{1}{2}\right\rceil$. Then $E[\widetilde{\eta}^2]$ is upper-bounded by

$$E[\widetilde{\eta}^2]\leq\frac{1.6\sigma}{c}\exp\left(-\frac{c^2}{8\sigma^2}\right), \quad (43)$$

where $c\geq 4\sigma$ is a constant.

We are now in a position to prove Theorem 2.

Proof of Theorem 2: Let δ, D_0 and Δ be positive constants such that

$$R<(1-\delta)C, \quad D_0<(1-\delta)C-R, \quad \Delta<\delta/\alpha. \quad (44)$$

Let T be the terminal time of the information transmission. We use the coding scheme proposed in Section III, where T_k is determined by

$$T_k=(1-\delta)T+(k-1)\Delta, \quad k=1,2,..., \quad (45)$$

and denote by $P_e(T_k)=Pr(\widehat{U}(T_k)\neq U_0)$ the error probability. Note that $|\mathbf{T}_k|=\Delta$ and $\beta_k=2^{-1}\sqrt{\exp(P\Delta)-1}$ $(k\geq 2)$ do not depend on k. Therefore, to evaluate the asymptotic behavior of the error probability $P_e(T_k)=2Q(\beta_k a_k)$, it is sufficient to examine the asymptotic behavior of a_k, $k\to\infty$. Using (37) and (42), we know that $P_e(T_k)$ is upper bounded by

$$P_e(T_k)=2Q(\beta_k a_k)<\exp\left(-\frac{\beta_k^2a_k^2}{2}\right), \quad k\geq 1. \quad (46)$$

Since $\widehat{B}_k(|\mathbf{T}_k|) \sim N(0, (2\beta_k)^{-2})$, noting (17) and applying Lemma 1 to (32), we obtain

$$E[|U_0 - \widehat{U}(T_k)|^2] \le \frac{1.6}{2\beta_k a_k} \exp\left(-\frac{\beta_k^2 a_k^2}{2}\right) < \frac{1}{\beta_k a_k} \exp\left(-\frac{\beta_k^2 a_k^2}{2}\right). \quad (47)$$

Then, using (20), we have the key inequality

$$a_{k+1}^2 > \beta_k a_k \exp\left(\frac{\beta_k^2 a_k^2}{2}\right), \quad k \ge 1. \quad (48)$$

We shall prove the inequality

$$P_e(T_k) \exp_k \left\{\frac{\exp(2D_0 T)}{2}\right\} < 1, \quad k \ge 1. \quad (49)$$

Since $M = \exp(RT)$ and $E[|U_0 - E[U_0]|^2] = (M^2 - 1)/12$, it is clear from (20) that

$$a_1^2 = \frac{1}{E[(U_0 - E[U_0])^2]} = \frac{12}{M^2 - 1} \sim 12 \exp(-2RT).$$

Since $|\mathbf{T}_1| = T_1 = (1-\delta)T$, $4\beta_1^2 = \exp(PT_1) - 1 \sim \exp\{(1-\delta)PT\} = \exp\{2(1-\delta)CT\}$ and

$$\beta_1^2 a_1^2 \sim 3 \exp\{2((1-\delta)C - R)T\}.$$

Therefore, noting (44), we have

$$\beta_1^2 a_1^2 > \exp(2D_0 T), \quad (50)$$

if T is large enough. Then the inequality (49) for $k = 1$ follows from (46) and (50). To prove (49) for $k \ge 2$, we shall show the inequality

$$a_k^2 > \beta_{k-1} a_{k-1} \exp_{k-1} \left\{\frac{\exp(2D_0 T)}{2}\right\}, \quad k \ge 2, \quad (51)$$

by induction. It is clear from (48) and (50) that

$$a_2^2 > \beta_1 a_1 \exp\left(\frac{\beta_1^2 a_1^2}{2}\right) > \beta_1 a_1 \exp\left\{\frac{\exp(2D_0 T)}{2}\right\},$$

meaning that (51) is true for $k = 2$. We now suppose that (51) is true for $k = j$, namely

$$a_j^2 > \beta_{j-1} a_{j-1} \exp_{j-1} \left\{\frac{\exp(2D_0 T)}{2}\right\}. \quad (52)$$

Since a_{j-1} is large enough, we may assume that

$$\beta_j^2 \beta_{j-1} a_{j-1} > 2. \quad (53)$$

Then it follows from (48), (52) and (53) that

$$\begin{aligned} a_{j+1}^2 &> \beta_j a_j \exp\left(\frac{\beta_j^2 a_j^2}{2}\right) \\ &> \beta_j a_j \exp\left\{\frac{\beta_j^2 \beta_{j-1} a_{j-1}}{2} \exp_{j-1}\left(\frac{\exp(2D_0 T)}{2}\right)\right\} \\ &> \beta_j a_j \exp_j \left\{\frac{\exp(2D_0 T)}{2}\right\}, \end{aligned}$$

meaning that (51) is true for $k = j + 1$. Thus, we have shown that (51) is true for all $k \ge 2$. Combining (46), (51) and (53), we have

$$\begin{aligned} P_e(T_k) &< \exp\left(-\frac{\beta_k^2 a_k^2}{2}\right) \\ &< \exp\left\{-\frac{\beta_k^2 \beta_{k-1} a_{k-1}}{2} \exp_{k-1}\{\exp(2D_0 T)/2\}\right\} \\ &< \exp\{-\exp_{k-1}\{\exp(2D_0 T)/2\}\}, \quad k \ge 2. \end{aligned}$$

Thus we have obtained (49).

Note that $\exp_k(T) = o(\exp_k\{\exp(2D_0 T)/2\})$. Then it is clear from (49) that

$$P_e(T_k) \exp_k(T) = o(1). \quad (54)$$

Since $\Delta < \delta/\alpha$,

$$T_{\lfloor \alpha T \rfloor} < (1-\delta)T + \lfloor \alpha T \rfloor \Delta < (1-\delta)T + \delta T = T$$

and

$$P_e(T) \le P_e(T_{\lfloor \alpha T \rfloor}).$$

Then, putting $k = \lfloor \alpha T \rfloor$ in (54), we obtain (41).

We have shown that, for any large coefficient α, one can construct a coding scheme such that the resulting error probability $P_e(T)$ decreases with T more rapidly than the exponential of order αT. We have also seen that, to realize a large α, we need to take the length $|\mathbf{T}_k| = \Delta(< \delta/\alpha)$ of subintervals small enough. This is possible for the continuous-time AWGC, and the situation is different in the discrete-time AWGC.

REFERENCES

[1] J. P. M. Schalkwijk and T. Kailath, "A coding scheme for additive noise channels with feedback, Part I: No bandwidth constraints," *IEEE Trans. Inform. Theory*, vol. IT-12, no. 2, pp. 172–182, Apr. 1966.

[2] M. S. Pinsker, "The probability of error in block transmission in a memoryless Gaussian channel with feedback," *Problems Inform. Transmission*, vol. 4, pp. 1–14, 1968.

[3] A. J. Kramer, "Improving communication reliability by use of an intermittent feedback channel," *IEEE Trans. Inform. Theory*, vol. IT-15, no. 1, pp. 52–60, Jan. 1969.

[4] K. Sh. Zigangirov, "Upper bound for the error probability for Gaussian channels with feedback," *Problems Inform. Transmission*, vol. 6, pp. 159–163, 1970.

[5] S. Ihara and L. G. Xin, "Error exponents for discrete-time white Gaussian channels with feedback," *Proc. Meijo Winter School 2003, Quantum Information and Complexity*, World Scientific, Singapore, pp. 227–243, 2004.

[6] S. Ihara, "Probability of error in information transmission over continuous-time Gaussian channel with feedback," *Proc. Int. Symp. Inf. Theory*, Toronto, Canada, pp. 895–899, 2008.

[7] R. G. Gallager and B. Nakiboğlu, "Variations on a theme by Schalkwijk and Kailath," *IEEE Trans. Inform. Theory*, vol. IT-56, no. 1, pp. 6–18, Jan. 2010.

[8] S. Ihara, "Upper bounds of error probabilities for stationary Gaussian channels with feedback," *IEEE Trans. Inform. Theory*, vol. IT-58, no. 12, pp. 7068–7072, Dec. 2012.

[9] T. Kailath, "An innovations approach to least-squares estimation, Part I: Linear filtering in additive white noise," *IEEE Trans. Automat. Control*, vol. AC-13, pp. 646–655, 1968.

[10] R. S. Liptser and A. N. Shiryayev, *Statistics of Random Processes I, II*, Springer Verlag, New York, 1977.

[11] S. Ihara, *Information Theory for Continuous Systems*, World Scientific, Singapore, 1993.

[12] M. S. Pinsker, *Information and Information Stability of Random Variables and Processes*, Holden-Day, San Francisco, 1964.