# Feedback Capacity of OU-Colored AWGN Channels

Jun Su
The University of Hong Kong
junsu@smail.nju.edu.cn

Guangyue Han
The University of Hong Kong
ghan@hku.hk

Shlomo Shamai (Shitz)
Technion-Israel Institute of Technology
sshlomo@ee.technion.ac.il

2023 IEEE International Symposium on Information Theory (ISIT) | 978-1-6654-7554-9/23/$31.00 ©2023 IEEE | DOI: 10.1109/ISIT54713.2023.10206821

***Abstract*—We derive an explicit formula of the feedback capacity for a continuous-time OU-colored AWGN channel. Among many others, this result shows that at least in some cases, the continuous-time Schalkwijk-Kailath coding scheme achieves the feedback capacity for such a channel, and feedback may not increase the capacity of a continuous-time ACGN channel even if the noise process is colored.**

## I. Introduction

We start with the following continuous-time additive white Gaussian noise (AWGN) channel

$$\boldsymbol{y}(t)=\boldsymbol{x}(t)+\boldsymbol{w}(t),\ \ -\infty<t<+\infty, \tag{1}$$

where the channel noise $\{\boldsymbol{w}(t)\}$ is a white Gaussian process with unit double-sided spectral density, $\{\boldsymbol{x}(t)\}$ is the channel input and $\{\boldsymbol{y}(t)\}$ is the channel output. Since $\{\boldsymbol{w}(t)\}$ can be regarded as the derivative $\{\dot{B}(t)\}$ of the standard Brownian motion $\{B(t)\}$ in the generalized sense [1], [2], or equivalently, $\{B(t)\}$ is the integral of $\{\boldsymbol{w}(t)\}$, the AWGN channel as in (1) can be alternatively characterized by

$$Y(t)=\int_0^t X(u)du+B(t),\ \ t\geq 0, \tag{2}$$

where $X=\{X(t)\}$ is the channel input and $Y=\{Y(t)\}$ is the channel output. Unlike white Gaussian noise, which is a generalized stochastic process in the sense of Schwartz's distribution [3], Brownian motion is an ordinary stochastic process that has been extensively studied in stochastic calculus. Evidently, the two formulations as in (1) and (2) allow us to examine an AWGN channel from different perspectives; in particular, the use of Brownian motion equips us with a wide range of established tools and techniques in stochastic calculus (see, e.g., [4]–[7] and references therein).

This paper is concerned with the following continuous-time additive colored Gaussian noise (ACGN) channel

$$\boldsymbol{y}(t)=\boldsymbol{x}(t)+\boldsymbol{z}(t),\ \ -\infty<t<+\infty, \tag{3}$$

where the channel noise $\boldsymbol{z}=\{\boldsymbol{z}(t)\}$ is a (possibly colored and generalized) stationary Gaussian process. Evidently, AWGN channels are a degenerated case of ACGN channels. Similarly as above, the ACGN channel as in (3) can be alternatively characterized by

$$Y(t)=\int_0^t X(u)du+Z(t),\ \ \ t\geq 0, \tag{4}$$

where $Z=\{Z(t)\}$ is the (generalized) integral of $\boldsymbol{z}$. Following [4], the treatment of ACGN channels in this work is mainly based on the formulation in (4).

For any $M\in\mathbb{N}$ and $T>0$, an $(M,T)$ *code* for the ACGN channel (4) consists of the following:

$(a)$ A message index $W$ independent of $\{Z(t); t\in[0,T]\}$ and uniformly distributed over $\{1,2,...,M\}$.

$(b)$ For the non-feedback case, an encoding function $g_u:\{1,2,...,M\}\to\mathbb{R},\ u\in[0,T]$, yielding codewords $X(u)=g_u(W)$; for the feedback case, an encoding function $g_u:\{1,2,...,M\}\times C[0,u]\to\mathbb{R},\ u\in[0,T]$, yielding codewords $X(u)=g_u(W,Y_0^{u-})$. For both cases, the classical average power constraint is satisfied:

$$\frac{1}{T}\int_0^T \mathbb{E}[|X(u)|^2]du\leq P.$$

$(c)$ A decoding functional $\hat{g}:C[0,T]\to\{1,2,...,M\}$.

Here we remark that for the feedback case, it follows from the pathwise continuity of $\{Y(t)\}$ that $X(t)=g_t(W,Y_0^t)$, and therefore the channel output $\{Y(t)\}$ is in fact the unique solution to the following stochastic functional differential equation:

$$dY(t)=g_t(W,Y_0^t)dt+dZ(t). \tag{5}$$

The *error probability* $\pi^{(T)}$ for the $(M,T)$ code as above is defined as

$$\pi^{(T)}=\mathrm{P}(\hat{g}(Y_0^T)\neq W).$$

A rate $R$ is *achievable* if there exists a sequence of $([e^{TR}],T)$ codes with $\lim_{T\to\infty}\pi^{(T)}=0$. The *channel capacity* is defined as the supremum of all achievable rates, denoted by $C_0(P)$ for the non-feedback case and $C_f(P)$ for the feedback case.

The literature on continuous-time ACGN channels is vast, and so below we only survey those results that are most relevant to this work. It has been shown by Huang and Johnson [8], [9] that $C_0(P)$ can be achieved by a Gaussian input. For a special family of ACGN channels, Hitsuda [10] has applied a canonical representation method to derive a fundamental formula for the channel mutual information (see Lemma III.2); based on this result, Ihara [11] showed that $C_f(P)$ can be achieved by a Gaussian input with an additive feedback term. Employing a Hilbert space approach, Baker [12], [13] has derived a theoretical formula for $C_0(P)$, which however is somewhat difficult to evaluate. When it comes to effective computation of $C_0(P)$ or $C_f(P)$, to the best of our knowledge, there are only a few results featuring an "explicit" and "computable" formula, detailed below. Here, we remark that Baker, Ihara and Hitsuda have studied the capacity of some families of ACGN channels, yet under different types of power constraints (see [10], [12]–[14]).


978-1-6654-7554-9/23/$31.00 ©2023 IEEE

1. For the ACGN channel formulated as in (3), when $\boldsymbol{z}$ is a stationary Gaussian process with rational spectrum, $C_0(P)$ can be determined by the water-filling method (see, e.g., [14]–[17]). More specifically,
$$C_0(P) = \frac{1}{4\pi}\int_{-\infty}^{\infty} \log\left[\max\left(\frac{A}{S_{\boldsymbol{z}}(x)}, 1\right)\right] dx,$$
where $S_{\boldsymbol{z}}(x)$ is the spectral density function of the noise process $\boldsymbol{z}$ and the water level $A$ is a constant determined by
$$P = \int_{[S_{\boldsymbol{z}}(x)\leq A]} (A - S_{\boldsymbol{z}}(x))dx.$$
2. For the AWGN channel as in (1) or (2), it is a classical result that $C_0(P) = P/2$ and feedback does not increase the channel capacity, that is to say, $C_f(P) = P/2$ (see, e.g., [4], [5], [18]). Moreover, $C_f(P)$ can be achieved by a linear feedback coding scheme [19]–[21].

In this paper, we will focus our attention on a special family of ACGN channels, which is characterized as
$$Y(t) = \int_0^t X(u)du + B(t) + \lambda \int_0^t \int_{-\infty}^s e^{-\kappa(s-u)} dB(u)ds \tag{6}$$
for $t \geq 0$, where $\lambda \in \mathbb{R}, \kappa > 0$. Note that the channel above can be alternatively characterized by
$$\boldsymbol{y}(t) = \boldsymbol{x}(t) + \boldsymbol{w}(t) + \lambda \boldsymbol{u}(t), \quad -\infty < t < \infty, \tag{7}$$
where, as before, $\boldsymbol{w}(t) = \dot{B}(t)$ is a white Gaussian process, and $\boldsymbol{u}(t) = \int_{-\infty}^t e^{-\kappa(t-u)}dB(u)$ is a stationary Ornstein-Uhlenbeck (OU) process, arguably the simplest nontrivial continuous-time stationary Gaussian process. Evidently, when $\lambda = 0$, (7) boils down to (1), and when $\lambda \neq 0$, the channel input, after going through an AWGN channel, will be further corrupted by an OU noise. For this reason, we may henceforth refer to the channel (6) as an OU-colored AWGN channel.

Note that when $\lambda = -\kappa$, the sum of the last two terms in (6) can be regarded as a continuous-time autoregressive (CAR) process with the first order (see, e.g., [22]), which has been of great interest to physicists and engineers (see, e.g., [23]). Here we remark that discrete-time ACGN channels with many types of noise (e.g., autoregressive (AR) processes, moving average (MA) processes, or more generally autoregressive moving average (ARMA) processes) have been extensively studied (see, e.g., [24]–[34]).

The main contribution in this work is an explicit characterization of the feedback capacity of an OU-colored AWGN channel. Before this work, no "explicit" and "computable" feedback capacity formula is known for any nontrivial stationary ACGN channel. Throughout the remainder of this paper, the notations $C_0(P)$ and $C_f(P)$ will be reserved for such a channel.

We will first derive a lower bound on $C_f(P)$, which turns out to be tight for some cases. To achieve this, we will examine the following ACGN channel
$$Y(t) = \int_0^t X(u)du + B(t) + \int_0^t \int_0^s h(s,u)dB(u)ds \tag{8}$$
for $t \geq 0$, where $h(s,u)$ is a Volterra kernel on $L^2([0,T]^2)$ for any $T > 0$. Here we emphasize that the channel (8) may not correspond to a stationary ACGN channel as in (3). However, it can be shown that $\{B(t)+\int_0^t\int_0^s h(s,u)dB(u)ds\}$ is equivalent to the Brownian motion $\{B(t)\}$ (see [35]), which renders the channel (8) more amenable to in-depth mathematical analysis, as evidenced by relevant results in the literature (see, e.g., [10], [11], [36]).

More specifically, let $\{\Theta(t)\}$ be the message process, and let $\bar{I}_{\text{SK}}(\Theta; Y)$ denote the mutual information rate between $\{\Theta(t)\}$ and $\{Y(t)\}$ under the so-called continuous-time Schalkwijk-Kailath (SK) coding scheme. We will show (Theorem IV.2) that $\bar{I}_{\text{SK}}(\Theta; Y) = Pr_P^2$, where $r_P$ is the limit of the unique solution to an ordinary differential equation, and moreover, one of the real roots of a third-order polynomial. It turns out that an OU-colored AWGN channel can be regarded as a special case of (8), and therefore $\bar{I}_{\text{SK}}(\Theta; Y)$ can help provide a lower bound on $C_f(P)$.

With the aforementioned lower bound, we are ready to derive an explicit expression of $C_f(P)$. More specifically, by examining a discrete-time approximation of the channel (6), we prove (Theorem V.1) that for the case $-2\kappa < \lambda < 0$, $C_f(P)$ is upper bounded by $\bar{I}_{\text{SK}}(\Theta; Y)$, which means $C_f(P) = \bar{I}_{\text{SK}}(\Theta; Y)$; for the other cases, we show $C_f(P) = C_0(P) = P/2$. As a byproduct, this result shows that feedback may not increase the capacity of a continuous-time ACGN channel even if the noise process is colored. By contrast, for a discrete-time ACGN channel, feedback does not increase the capacity if and only if the noise spectrum is white (see [28, Cor. 4.3]).

The remainder of the paper is organized as follows. In Section II, we introduce necessary notation and terminlogies. In Section III, we review the coding theorem for the feedback capacity and introduce the continuous-time SK coding scheme. Section IV provides an asymptotic characterization of $\bar{I}_{\text{SK}}(\Theta; Y)$ for a subclass of ACGN channels, which helps provide a lower bound on $C_f(P)$. In Section V, we derive an explicit formula for $C_f(P)$.

## II. Notation and Terminlogies

We use $(\Omega, \mathcal{F}, \mathbb{P})$ to denote the underlying probability space, and $\mathbb{E}$ to denote the expectation with respect to the probability measure $\mathbb{P}$. As is typical in the theory of stochastic calculus, we assume the probability space is equipped with a filtration $\{\mathcal{F}_t : t \geq 0\}$, which satisfies the *usual conditions* [37] and is rich enough to accommodate a standard Brownian motion.

Let $C[0,\infty)$ denote the space of all continuous functions over $[0,\infty)$, and let $C^1[0,\infty)$ be the space of all functions in $C[0,\infty)$ that have continuous derivatives on $[0,\infty)$. For any $T > 0$, let $C[0,T]$ denote the space of all continuous functions over $[0,T]$. Let $X, Y$ be random variables defined on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$, which will be used to illustrate most of the notions and facts in this section (the same notations may have different connotations in other sections). Note that in this paper, a random variable can be real-valued with a probability density function, or path-valued (more precisely, $C[0,\infty)$- or $C[0,T]$-valued).

For any two path-valued random variables $X_0^T = \{X(t); 0 \le t \le T\}$ and $Y_0^T = \{Y(t); 0 \le t \le T\}$, we use $\mu_{X_0^T}$ and $\mu_{Y_0^T}$ to denote the probability distributions on $C[0,T]$ induced by $X_0^T$ and $Y_0^T$, respectively, and $\mu_{X_0^T} \times \mu_{Y_0^T}$ the product distribution of $\mu_{X_0^T}$ and $\mu_{X_0^T}$; moreover, we will use $\mu_{X_0^T, Y_0^T}$ to denote their joint probability distribution on $C[0,T] \times C[0,T]$. Besides, we use $\mathcal{F}_t(Y)$ to denote the $\sigma$-field generated by $Y_0^t$.

For any two probability measures $\mu$ and $\nu$, we write $\mu \sim \nu$ to mean they are equivalent, namely, $\mu$ is absolutely continuous with respect to $\nu$ and vice versa. By Hitsuda [35], if a Gaussian process $\{Z(t)\}$ is equivalent to a given Brownian motion, then there exists a (possibly different) Brownian motion $\{B(t)\}$ such that $\{Z(t)\}$ can be uniquely represented by

$$Z(t) = B(t) + \int_0^t \int_0^s h(s,u) dB(u) ds, \tag{9}$$

where $h(s,u)$ is a Volterra kernel in $L^2([0,T]^2)$ for any $T > 0$, i.e., $h(s,u) = 0$ if $s < u$ and $\int_0^T \int_0^T h(s,u)^2 ds du < \infty$ for any $T > 0$. Conversely, for a given Brownian motion $\{B(t)\}$, if $\{Z(t)\}$ has a representation in the form (9), then $\{Z(t)\}$ is equivalent to $\{B(t)\}$. Note that, for any $T > 0$, there exists a Volterra kernel $l(s,u) \in L^2([0,T]^2)$, referred to as the *resolvent kernel* of $h(s,u)$, such that

$$\begin{aligned} -h(s,u) &= l(s,u) + \int_u^s h(s,v) l(v,u) dv \\ &= l(s,u) + \int_u^s l(s,v) h(v,u) dv \end{aligned} \tag{10}$$

for any $s, u \in [0,T]$ (see [38, Chapter 2]).

The *mutual information* $I(X;Y)$ between two real-valued random variables $X, Y$ is defined as

$$I(X;Y) = \mathbb{E}\left[\log \frac{f_{X,Y}(X,Y)}{f_X(X) f_Y(Y)}\right],$$

where $f_X, f_Y$ denote the probability density functions of $X, Y$, respectively, and $f_{X,Y}$ denotes their joint probability density function. For two $C[0,T]$-valued random variables $X_0^T, Y_0^T$, we define

$$I(X_0^T; Y_0^T) = \begin{cases} \mathbb{E}\left[\log \frac{d\mu_{X_0^T, Y_0^T}}{d\mu_{X_0^T} \times \mu_{Y_0^T}}(X_0^T, Y_0^T)\right], & \text{if } \frac{d\mu_{X_0^T, Y_0^T}}{d\mu_{X_0^T} \times \mu_{Y_0^T}} \text{ exists,} \\ \infty, & \text{otherwise,} \end{cases}$$

where $d\mu_{X_0^T, Y_0^T} / d\mu_{X_0^T} \times \mu_{Y_0^T}$ denotes the Radon-Nikodym derivative of $\mu_{X_0^T, Y_0^T}$ with respect to $\mu_{X_0^T} \times \mu_{Y_0^T}$.

The notion of mutual information can be further extended to generalized random processes. Due to the space limit, we refer the reader to [39] for a comprehensive exposition.

## III. CONTINUOUS-TIME SK CODING

In this section, we shall examine the continuous-time ACGN channel (8). Throughout this section, let $Z(t) = B(t) + \int_0^t \int_0^s h(s,u) dB(u) ds$.

For the purpose of presenting a coding theorem for the feedback capacity, instead of transmitting a message index $W$, a random variable taking values from a finite alphabet, we will transmit a message process $\Theta = \{\Theta(t)\}$, a real-valued random process. Then, compared to (5), the associated stochastic functional differential equation will take the following form:

$$dY(t) = g_t(\Theta(t), Y_0^t) dt + dZ(t),$$

where we have set $X(t) = g_t(\Theta(t), Y_0^t)$. Following [4], we consider the so-called *$T$-block feedback capacity* $C_{f,T}(P) = \sup_{(\Theta,X)} \frac{1}{T} I(\Theta_0^T; Y_0^T)$, where the supremum is taken over all pairs $(\Theta, X)$ satisfying the following constraint

$$\frac{1}{T} \int_0^T \mathbb{E}[X^2(t)] dt \le P. \tag{11}$$

Now, we define $\bar{I}(\Theta; Y) = \limsup_{T\to\infty} \frac{1}{T} I(\Theta_0^T; Y_0^T)$, provided the limit exists, and furthermore define $C_{f,\infty}(P) = \sup_{(\Theta,X)} \bar{I}(\Theta; Y)$, where the supremum is taken for all pairs $(\Theta, X)$ satisfying

$$\overline{\lim_{T\to\infty}} \frac{1}{T} \int_0^T \mathbb{E}[X^2(t)] dt \le P. \tag{12}$$

Then, the aforementioned coding theorem for the feedback capacity is stated below.

**Theorem III.1** ( [40, Th. 1]). *Assume that*

$$\lim_{T\to\infty} \frac{1}{T} C_{f,T}(P) = 0.$$

*If $R < C_{f,\infty}(P)$ and $P$ is continuous point of $C_{f,\infty}(P)$, then the rate $R$ is achievable. Conversely, if a rate $R$ is achievable, then $R \le C_{f,\infty}(P)$.*

The following lemma generalizes the classical I-CMMSE relationship in [5], [41].

**Lemma III.2** (Hitsuda and Ihara [10]). *Suppose $\int_0^T \mathbb{E}[X^2(t)] dt < \infty$. Then, we have*

$$I(\Theta_0^T; Y_0^T) = \frac{1}{2} \int_0^T \mathbb{E}\left[\left|X_l(t) - \mathbb{E}[X_l(t)|\mathcal{F}_t(Y)]\right|^2\right] dt,$$

*where $\{X_l(t); t \in [0,T]\}$ is a random process defined by*

$$X_l(t) = X(t) + \int_0^t l(t,u) X(u) du,$$

*and $l(s,u)$ is the resolvent kernel of $h(s,u)$ in $L^2([0,T]^2)$.*

Consider the additive feedback coding scheme $(\Theta, X) = (\{\Theta(t)\}, \{X(t)\})$ with $X(t) = \Theta(t) - \zeta(t)$, where $\zeta = \{\zeta(t)\}$ represents the feedback term, causally dependent on the output $Y = \{Y(t)\}$, and is appropriately chosen such that the stochastic functional differential equation

$$Y(t) = \int_0^t \Theta(s) - \zeta(s) ds + Z(t) \tag{13}$$

admits a unique solution. The following lemma characterizes the optimal coding scheme for $C_{f,T}(P)$ more explicitly.

**Theorem III.3** ( [11, Th. 3] Reformulated). *For the continuous-time ACGN channel (8) under the constraint (11),*

*$C_{f,T}(P)$ can be achieved by a Gaussian pair $(\Theta, X)$ of the following form*

$$X(t) = \Theta(t) - \mathbb{E}[\Theta(t)|\mathcal{F}_t(Y^*)], \quad t \in [0,T], \tag{14}$$

*where*

$$Y^*(t) = \int_0^t \Theta(s)ds + Z(t).$$

*Moreover, $\mathcal{F}_t(Y^*) = \mathcal{F}_t(Y)$ for any $t \in [0,T]$, and so the pair $(\Theta, X)$ characterizes an additive feedback coding scheme of the form (13) where $\zeta(t) = \mathbb{E}[\Theta(t)|\mathcal{F}_t(Y)]$.*

The essence of the above theorem is that we can restrict our attention to the coding schemes of the form as in (14). Following the spirits of the classical Schalkwijk-Kailath (SK) coding scheme, we formulate in our notation the continuous-time version of the celebrated SK coding scheme $(\Theta, X)$ in the form of

$$X(t) = \Theta(t) - \zeta(t) = A(t)\Theta_0 - A(t)\mathbb{E}[\Theta_0|\mathcal{F}_t(Y^*)] \tag{15}$$

satisfying $\mathbb{E}[X^2(t)] = P$ for any $t \geq 0$, where $\Theta_0$ is a standard Gaussian random variable and $A(t)$ is some deterministic function.

In general, the above continuous-time SK coding scheme can be invalid in the sense that $A(t)$ may not exist. However, in Sections IV and V, we will show that the continuous-time SK coding scheme is valid for a subclass of ACGN channels (8) and is also optimal for some special families of ACGN channels.

## IV. Mutual Information Rate

In this section, we narrow our attention to the special family of ACGN channels (8) in which the resolvent kernel $l(t,s)$ of $h(t,s)$ can be written as

$$l(t,s) = \frac{l_u(s)}{l_d(t)}, \quad \text{for } t \geq s, \tag{16}$$

where $l_u(t) \in C[0, +\infty)$ and $l_d(t) \in C^1[0, +\infty)$.

We first prove a lemma characterizing the asymptotics of the solution $g$ to the following ordinary differential equation (ODE)

$$\begin{cases} g'(t) = -Pg^3(t) + \dfrac{P}{\sqrt{2}}g^2(t) + p(t)g(t) + \dfrac{1}{\sqrt{2}}q(t), \\ g(0) = \dfrac{1}{\sqrt{2}}, \end{cases} \tag{17}$$

where $p(t), q(t) \in C[0,\infty)$ satisfying $\lim_{t\to\infty} p(t) = p$ and $\lim_{t\to\infty} q(t) = q$.

**Lemma IV.1.** *For every $P > 0$, the ODE (17) admits a unique solution $g(t) \in C^1[0,\infty)$. Moreover, $\lim_{t\to\infty} g(t)$ exists, which is one of the real roots of the following cubic equation:*

$$-Py^3 + \frac{P}{\sqrt{2}}y^2 + py + \frac{q}{\sqrt{2}} = 0.$$

Equipped with Lemma IV.1, we are ready to prove our first main result, described below.

**Theorem IV.2.** *Assume the resolvent kernel $l(t,s)$ of $h(t,s)$ in (8) can be written in the form (16) with*

$$\lim_{t\to\infty} \frac{l_u(t)}{l_d(t)} = \alpha, \quad \lim_{t\to\infty} \frac{l'_d(t)}{l_d(t)} = \beta. \tag{18}$$

*Then, we have*

$$\bar{I}_{SK}(\Theta; Y) = Pr_P^2, \tag{19}$$

*where $r_P = \lim_{t\to\infty} g(t)$ and $g$ is the solution of the ODE (17) with $p(t) = -l'_d(t)/l_d(t)$ and $q(t) = (l_u(t) + l'_d(t))/l_d(t)$. Moreover, $r_P$ is one of the real roots of the following cubic equation:*

$$-Py^3 + \frac{P}{\sqrt{2}}y^2 - \beta y + \frac{\beta + \alpha}{\sqrt{2}} = 0. \tag{20}$$

**Remark IV.3.** *It turns out that from the proof of Theorem IV.2, $r_P$ is uniquely determined by $l(t,s)$, rather than the choice of $l_u, l_d$.*

To illustrate possible applications of the above theorem, we give the following two examples.

**Example IV.4.** When $l(t,s) \equiv 0$, the channel (8) boils down to the AWGN channel (2). Apparently, one can choose $l_u \equiv 0$ and $l_d \equiv 1$, yielding $\bar{I}_{SK}(\Theta; Y) = P/2$, which is widely known as the capacity of the channel (2).

**Example IV.5.** When $l(t,s) = 1$, it turns out that the channel (8) boils down to

$$Y(t) = \int_0^t X(s)ds + B(t) - \int_0^t \int_0^s e^{u-s} dB(u)ds.$$

Apparently, it can be verified that $l_u \equiv l_d \equiv c$, where $c$ is a non-zero constant. Thus, we have $\alpha = 1, \beta = 0$, yielding that $\bar{I}_{SK}(\Theta; Y)$ is the unique positive root of the cubic equation $P(x+1)^2 = 2x^3$. This recovers Proposition 1 in [42].

## V. Feedback Capacity

In this section, we focus on the OU-colored AWGN channel (6). From now on, let

$$Z(t) = B(t) + \lambda \int_0^t \int_{-\infty}^s e^{-\kappa(s-u)} dB(u)ds, \ \lambda \in \mathbb{R}, \kappa > 0,$$

and let $\zeta_0$ denote $\int_{-\infty}^0 e^{\kappa s} dB(s)$.

The following theorem is our main result in which we derive an explicit formula for $C_f(P)$.

**Theorem V.1.** *$C_f(P)$ is determined in the following two cases:*

1. *if $\lambda \leq -2\kappa$ or $\lambda \geq 0$, then $C_f(P) = P/2$;*
2. *if $-2\kappa < \lambda < 0$, then $C_f(P)$ is the unique positive root of the third-order polynomial*

$$P(x+\kappa)^2 = 2x(x + |\kappa + \lambda|)^2. \tag{21}$$

### A. Proof of the Converse Part

**Lemma V.2.** *For any $T > 0$, the $T$-block feedback capacity $C_{f,T}(P)$ of the OU-colored AWGN channel (6) is upper bounded by*

$$C_{f,T}(P) \leq \begin{cases} \dfrac{P}{2}, & \text{if } \lambda \leq -2\kappa \text{ or } \lambda \geq 0; \\ x_0(P;\lambda,\kappa), & \text{if } -2\kappa < \lambda < 0, \end{cases} \tag{22}$$

*where $x_0(P;\lambda,\kappa)$ is the positive root of the polynomial (21). Moreover, (22) also holds when $T$ is replaced by $\infty$.*

*Proof of Lemma V.2.* We only prove the case that $T > 0$, which together with the definition of $C_{f,\infty}(P)$, immediately implies the "moreover" part.

By Theorem III.3, we can prove (22) by considering any Gaussian pair $(\Theta, X)$ of the form (14) in which $X$ satisfies the constraint (11). Then, it is known [11] that there exists a Volterra kernel $K(t,s)$ on $L^2([0,T]^2)$ such that $\mathbb{E}[\Theta(t)|\mathcal{F}_t(Y^*)] = \int_0^t K(t,s)dY^*(s)$. It turns out that it suffices to prove the lemma under the assumption that $K(t,s)$ is continuous on the set $\{(t,s) \in [0,T]^2; t \geq s\}$. The remainder of the proof is divided into two steps.

**Step 1.** In this step, we shall introduce a sequence of ARMA(1,1) Gaussian channels constructed from the OU-colored AWGN channel (6).

For any $n \in \mathbb{N}$, we consider a partition $\{t_k^{(n)}; k = 0,1,...,n\}$ of $[0,T]$ satisfying $t_{k+1}^{(n)} - t_k^{(n)} = \delta_n$ for all $k$, where $\delta_n = T/n$. For any $k = 0,1,...,n-1$, define $B_k^{(n)} = B(t_{k+1}^{(n)}) - B(t_k^{(n)})$ and

$$\widetilde{Z}_k^{(n)} = B_k^{(n)} + \lambda d_k^{(n)}\left(m(\delta_n)\zeta_0 + \sum_{i=0}^{k-1} e^{\kappa t_{i+1}} B_i^{(n)}\right), \tag{23}$$

where $d_k^{(n)} \triangleq \int_{t_k^{(n)}}^{t_{k+1}^{(n)}} e^{-\kappa s} ds$ and $m(x) \triangleq \sqrt{2\kappa x/(1-e^{-2\kappa x})}$. It can be readily verified that $\{\widetilde{Z}_k^{(n)}/\sqrt{\delta_n}\}$ is a stationary ARMA(1,1) process of the following form

$$\frac{\widetilde{Z}_{k+1}^{(n)}}{\sqrt{\delta_n}} = e^{-\kappa\delta_n}\frac{\widetilde{Z}_k^{(n)}}{\sqrt{\delta_n}} + \frac{B_{k+1}^{(n)}}{\sqrt{\delta_n}} + \left(\frac{\lambda}{\kappa} - \left(\frac{\lambda}{\kappa}+1\right)e^{-\kappa\delta_n}\right)\frac{B_k^{(n)}}{\sqrt{\delta_n}}. \tag{24}$$

Furthermore, for any $k = 0,1,...,n-1$, we define $Y_k^{*,(n)} = \Theta_k^{(n)} + \widetilde{Z}_k^{(n)}$ and $Y_k^{(n)} = \Theta_k^{(n)} - \zeta_k^{(n)} + \widetilde{Z}_k^{(n)}$, where $\Theta_k^{(n)} = \int_{t_k^{(n)}}^{t_{k+1}^{(n)}} \Theta(s)ds$ and $\zeta_k^{(n)} = \delta_n \sum_{i=0}^{k-1} K(t_k^{(n)}, t_i^{(n)}) Y_i^{*,(n)}$. Then, we have that, for $k = 0,1,...,n-1$

$$\frac{Y_k^{(n)}}{\sqrt{\delta_n}} = \frac{\Theta_k^{(n)} - \zeta_k^{(n)}}{\sqrt{\delta_n}} + \frac{\widetilde{Z}_k^{(n)}}{\sqrt{\delta_n}}. \tag{25}$$

In particular, it follows from (24) that (25) corresponds to $n$-block ARMA(1,1) Gaussian channels with feedback.

**Step 2.** This step will be devoted to approximating $P/2$ and $x_0(P;\lambda,\kappa)$ by feedback capacities of the sequence of ARMA(1,1) Gaussian channels (25).

To this end, we have the following chain of inequalities:

$$\begin{aligned} \frac{1}{T}I(\Theta_0^T; Y_0^T) &\overset{(a)}{=} \frac{1}{T}I(\Theta_0^T; Y_0^{*,T}) \\ &\overset{(b)}{\leq} \varliminf_{n\to\infty} \frac{1}{T}I(\{\Theta_k^{(n)}\}; \{Y_k^{*,(n)}\}) \\ &\overset{(c)}{=} \varliminf_{n\to\infty} \frac{1}{T}I(\{\Theta_k^{(n)}\}; \{Y_k^{(n)}\}) \\ &\overset{(d)}{\leq} \varliminf_{n\to\infty} \frac{1}{\delta_n}C_{FB,n}(P\delta_n + \frac{e(\delta_n)}{n}) \\ &\overset{(e)}{\leq} \varliminf_{n\to\infty} \frac{1}{\delta_n}C_{FB}(P\delta_n + \frac{e(\delta_n)}{n}) \\ &\overset{(f)}{\leq} \begin{cases} \dfrac{P}{2}, & \text{if } \lambda \leq -2\kappa \text{ or } \lambda \geq 0; \\ x_0(P;\lambda,\kappa), & \text{if } -2\kappa < \lambda < 0, \end{cases} \end{aligned} \tag{26}$$

where $e(\delta_n)$ is some function dependent on $\{\Theta(t)\}$ with the property $\lim_{n\to\infty} e(\delta_n) = 0$, where $C_{FB,n}(P)$ denotes the $n$-block feedback capacity [43] of the channel (25) under the constraint that the average power of the channel input is bounded by $P$ (see [28]) and $C_{FB}(P)$ denotes feedback capacity. Now, with (a)-(f) validified, (22) immediately follows from (26) and Theorem III.3. □

*Proof of the Converse Part.* The converse part immediately follows from Theorem III.1 and Lemma V.2. □

### B. Proof of the Achievability Part

Since Case (1) follows immediately from the easily verifiable fact that $C_0(P) = P/2$ and $C_0(P) \leq C_f(P)$, we focus on Case (2) in what follows.

By [44, Th. 7.15], we have $\mu_Z \sim \mu_B$. Moreover, it follows from (9) (10) that there exists a standard Brownian motion $\{V(t)\}$ on $(\Omega, \{\mathcal{F}_t(Z)\}, \mathbb{P})$ such that $V(t) = Z(t) + \int_0^t \int_0^s l_{\mathrm{OU}}(s,u)dZ(u)ds$, where the Volterra kernel $l_{\mathrm{OU}}(s,u)$ is equal to

$$\begin{cases} \dfrac{\lambda(2\kappa+\lambda)^2 e^{(\kappa+\lambda)u} + \lambda^2(2\kappa+\lambda)e^{-(\kappa+\lambda)u}}{\lambda^2 e^{-(\kappa+\lambda)s} - (\lambda+2\kappa)^2 e^{(\kappa+\lambda)s}}, & \text{if } \lambda+\kappa \neq 0; \\ \dfrac{\kappa(\kappa u+1)}{\kappa s+2}, & \text{if } \lambda+\kappa = 0. \end{cases}$$

It is easy to see that $l_{\mathrm{OU}}$ satisfies all the conditions in Theorem IV.2. Consequently, we have $\overline{I}_{\mathrm{SK}}(\Theta;Y) = Pr_{\mathrm{OU}}^2$, where $r_{\mathrm{OU}}$ is one of the real roots of the following equation

$$-Py^3 + \frac{P}{\sqrt{2}}y^2 - |\lambda+\kappa|y + \frac{1}{\sqrt{2}}\kappa = 0. \tag{27}$$

It can be easily verified that the equation (27) has the unique positive root for all $-2\kappa \leq \lambda \leq 0$. Then, substituting $y = \sqrt{x/P}$ into (27), we have that $\overline{I}_{\mathrm{SK}}(\Theta;Y)$ is the unique positive root of (21), which implies $C_{f,\infty}(P) \geq x_0(P;\lambda,\kappa)$. This, together with Lemma V.2 and Theorem III.1, immediately yields $C_f(P) \geq x_0(P;\lambda,\kappa)$, as desired.

**Acknowledgment**: The work of S. Shamai has been supported by the European Union's Horizon 2020 Research and Innovation Programme, grant agreement no. 694630.

## REFERENCES


[1] L. Koralov and Y. G. Sinai, *Theory of Probability and Random Processes*. Springer Science & Business Media, 2007.

[2] I. M. Gel'fand and N. Y. Vilenkin, *Generalized Functions: Applications of Harmonic Analysis*. Academic Press, 2014, vol. 4.

[3] N. Obata, *White Noise Calculus and Fock Space*. Springer, 2006.

[4] S. Ihara, *Information Theory for Continuous Systems*. World Scientific, 1993, vol. 2.

[5] T. Kadota, M. Zakai, and J. Ziv, "Mutual information of the white Gaussian channel with and without feedback," *IEEE Transactions on Information Theory*, vol. 17, no. 4, pp. 368–371, 1971.

[6] X. Liu and G. Han, "On continuous-time Gaussian channels," *Entropy*, vol. 21, no. 1, p. 67, 2019.

[7] G. Han and S. Shamai, "On sampling continuous-time AWGN channels," *IEEE Transactions on Information Theory*, vol. 68, no. 2, pp. 782–794, 2021.

[8] R. Huang and R. Johnson, "Information capacity of time-continuous channels," *IRE Transactions on Information Theory*, vol. 8, no. 5, pp. 191–198, 1962.

[9] ——, "Information transmission with time-continuous random processes," *IEEE Transactions on Information Theory*, vol. 9, no. 2, pp. 84–94, 1963.

[10] M. Hitsuda and S. Ihara, "Gaussian channels and the optimal coding," *Journal of Multivariate Analysis*, vol. 5, no. 1, pp. 106–118, 1975.

[11] S. Ihara, "On the capacity of the continuous time Gaussian channel with feedback," *Journal of Multivariate Analysis*, vol. 10, no. 3, pp. 319–331, 1980.

[12] C. Baker, "Capacity of the mismatched Gaussian channel," *IEEE Transactions on Information Theory*, vol. 33, no. 6, pp. 802–812, 1987.

[13] C. R. Baker, "Channel models and their capacity," *Sen, P.K. (ed.) Contributions to Statistics; Essays in Honor of Norman L. Johnson,*, pp. 1–16, 1983.

[14] C. R. Baker and S. Ihara, "Information capacity of the stationary Gaussian channel," *IEEE Transactions on Information Theory*, vol. 37, no. 5, pp. 1314–1326, 1991.

[15] R. M. Fano, "Transmission of information: A statistical theory of communications," *American Journal of Physics*, vol. 29, no. 11, pp. 793–794, 1961.

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

[17] R. G. Gallager, *Information Theory and Reliable Communication*. Springer, 1968, vol. 588.

[18] C. E. Shannon, "A mathematical theory of communication," *The Bell System Technical Journal*, vol. 27, no. 3, pp. 379–423, 1948.

[19] S. Ihara, "Optimal coding in white Gaussian channel with feedback," *Lecture Notes Math*, vol. 330, pp. 120–123, 1973.

[20] R. S. Liptser, "Optimal encoding and decoding for transmission of a Gaussian Markov signal in a noiseless-feedback channel," *Problemy Peredachi Informatsii*, vol. 10, no. 4, pp. 3–15, 1974.

[21] S. Ihara, "Coding theory in white Gaussian channel with feedback," *Journal of Multivariate Analysis*, vol. 4, no. 1, pp. 74–87, 1974.

[22] P. J. Brockwell, "Continuous-time ARMA processes," *Handbook of Statistics*, vol. 19, pp. 249–276, 2001.

[23] Fowler, *Statistical Mechanics*. CUP Archive, 1967.

[24] T. Liu and G. Han, "Feedback capacity of stationary Gaussian channels further examined," *IEEE Transactions on Information Theory*, vol. 65, no. 4, pp. 2492–2506, 2018.

[25] ——, "The ARMA(k) Gaussian feedback capacity," in *2017 IEEE International Symposium on Information Theory (ISIT)*. IEEE, 2017, pp. 211–215.

[26] G. Han and T. Liu, "ARMA(1) Gaussian feedback capacity revisited," in *2016 International Symposium on Information Theory and Its Applications (ISITA)*. IEEE, 2016, pp. 111–115.

[27] S. Yang, A. Kavcic, and S. Tatikonda, "On the feedback capacity of power-constrained Gaussian noise channels with memory," *IEEE Transactions on Information Theory*, vol. 53, no. 3, pp. 929–954, 2007.

[28] Y.-H. Kim, "Feedback capacity of stationary Gaussian channels," *IEEE Transactions on Information Theory*, vol. 56, no. 1, pp. 57–85, 2009.

[29] ——, "Feedback capacity of the first-order moving average Gaussian channel," *IEEE Transactions on Information Theory*, vol. 52, no. 7, pp. 3063–3079, 2006.

[30] S. Ihara, "On the feedback capacity of the first-order moving average Gaussian channel," *Japanese Journal of Statistics and Data Science*, vol. 2, pp. 491–506, 2019.

[31] O. Sabag, V. Kostina, and B. Hassibi, "Feedback capacity of MIMO Gaussian channels," in *2021 IEEE International Symposium on Information Theory (ISIT)*. IEEE, 2021, pp. 7–12.

[32] A. Gattami, "Feedback capacity of Gaussian channels revisited," *IEEE Transactions on Information Theory*, vol. 65, no. 3, pp. 1948–1960, 2018.

[33] S. Fang and Q. Zhu, "Feedback capacity of parallel ACGN channels and Kalman filter: Power allocation with feedback," *ArXiv Preprint ArXiv:2102.02730*, 2021.

[34] M. S. Derpich and J. Østergaard, "Comments on feedback capacity of stationary Gaussian channels," *IEEE Transactions on Information Theory*, 2022, doi:10.1109/TIT.2022.3182270.

[35] M. Hitsuda, "Representation of Gaussian processes equivalent to Wiener process," *Osaka Journal of Mathematics*, vol. 5, no. 2, pp. 299–312, 1968.

[36] ——, "Mutual information in Gaussian channels," *Journal of Multivariate Analysis*, vol. 4, no. 1, pp. 66–73, 1974.

[37] I. Karatzas and S. Shreve, *Brownian Motion and Stochastic Calculus*. Springer Science & Business Media, 1991, vol. 113.

[38] F. Smithies, *Integral Equations*. CUP Archive, 1958, no. 49.

[39] I. M. Gelfand and A. M. Yaglom, *Calculation of the amount of information about a random function contained in another such function*. American Mathematical Society Providence, 1959.

[40] S. Ihara, "Coding theorem for the continuous time Gaussian channel with feedback II," *Probability Theory and Mathematical Statistics: Proceedings of the Sixth USSR-Japan Symposium, Kiev, USSR*, pp. 132–142, 1991.

[41] T. E. Duncan, "On the calculation of mutual information," *SIAM Journal on Applied Mathematics*, vol. 19, no. 1, pp. 215–220, 1970.

[42] S. Ihara, "Capacity of mismatched Gaussian channels with and without feedback," *Probability Theory and Related Fields*, vol. 84, no. 4, pp. 453–471, 1990.

[43] T. M. Cover and S. Pombra, "Gaussian feedback capacity," *IEEE Transactions on Information Theory*, vol. 35, no. 1, pp. 37–43, 1989.

[44] R. S. Lipster and A. N. Shiryaev, *Statistics of Random Processes I: General Theory*, 1977.