On Continuous-Time Gaussian Channels *

Xianming Liu
Huazhong University of Science and Technology
email: xmliu@hust.edu.cn

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

April 9, 2017

Abstract

We establish natural connections between continuous-time Gaussian feedback/memory channels and their associated discrete-time versions in the forms of sampling and approximating theorems. It turns out that these connections, together with relevant tools from stochastic calculus, can enhance our understanding of continuous-time Gaussian channels in terms of giving alternative interpretations to some long-held "folklores", recovering known results from new perspectives, and obtaining new results inspired by the insights and ideas that come along with the connections. In particular, we derive the capacity regions of a continuous-time white Gaussian multiple access channel, a continuous-time white Gaussian interference channel, and a continuous-time white Gaussian broadcast channel; furthermore, applying the the sampling and approximating theorems and the ideas and techniques in their proofs, we analyze how feedback affects the capacity regions of families of continuous-time multi-user one-hop Gaussian channels: feedback will increase the capacity regions of some continuous-time white Gaussian broadcast and interference channels, while it will not increase capacity regions of continuous-time white Gaussian multiple access channels.

Index Terms: *continuous-time channel, Gaussian channel, feedback, memory, sampling theorem, approximating theorem, network information theory, capacity, capacity region, mutual information*

1 Introduction

Continuous-time Gaussian channels were considered at the very inception of information theory. In his celebrated paper [74] birthing information theory, Shannon studied the following point-to-point continuous-time white Gaussian channels:

$$Y(t) = X(t) + Z(t), \quad t \in \mathbb{R}, \tag{1}$$

where $X(t)$ is the channel input with average power limit P, $Z(t)$ is the white Gaussian noise with flat power spectral density 1 and $Y(t)$ are the channel output. Shannon actually

*Results in this paper have been partially presented in the 2014 IEEE ISIT [56].

only considered the case that the channel has bandwidth limit ω, namely, the channel input X and the noise Z, and therefore the output Y all have bandwidth limit ω (alternatively, as in (9.54) of [15], this can be interpreted as the original channel (1) concatenated with an ideal bandpass filter with bandwidth limit ω). Using the celebrated Shannon-Nyquist sampling theorem [62, 75], the continuous-time channel (1) can be equivalently represented by a parallel Gaussian channel:

$$Y_n^{(\omega)} = X_n^{(\omega)} + Z_n^{(\omega)}, \quad n \in \mathbb{Z}, \tag{2}$$

where the noise process $\{Z_n^{(\omega)}\}$ is i.i.d. with variance 1. Regarding the "space" index n as time, the above parallel channel can be interpreted as a discrete-time Gaussian channel associated with the continuous-time channel (1). It is well known from the theory of discrete-time Gaussian channels that the capacity of the channel (2) can be computed as

$$C^{(\omega)} = \omega \log \left(1 + \frac{P}{2\omega}\right). \tag{3}$$

Then, the capacity C of the channel (1) can be computed by taking the limit of the above expression as ω tends to infinity:

$$C = \lim_{\omega \to \infty} C^{(\omega)} = P/2. \tag{4}$$

The bandwidth approach consisting of (1)-(4) typifies how one explains explains the long-held folklore "a continuous-time Gaussian channel is the limit of associated discrete-time Gaussian channels as bandwidth limit tends to infinity, or equivalently, as the signal-to-noise tends to zero". Moments of reflection, however, reveals that the bandwidth approach as above for the channel capacity (with bandwidth limit or not) is heuristic in nature: For one thing, a bandwidth-limited signal cannot be time-limited, which renders it infeasible to define the data transmission rate if assuming a channel has bandwidth limit. In this regard, rigorous treatments coping with this issue and other technicalities can be found in [86, 22]; see also [76] for a relevant in-depth discussion. Another issue is that, even disregarding the above nuisance arising from the bandwidth limit assumption, the bandwidth approach only gives a lower bound for the capacity of (1): it shows that $P/2$ is achievable via a class of special coding schemes, but it is not clear that why transmission rate higher than $P/2$ cannot be achieved by other coding schemes. The capacity of (1) was rigorously studied in [21, 9], and a complete proof establishing $P/2$ as its de facto capacity can be found in [3, 4].

In this paper, instead of using a white Gaussian noise, we would rather follow [39] to use a Brownian motion to formulate a white Gaussian channel. Under this formulation, the white Gaussian noise channel (1) becomes:

$$Y(t) = \int_0^t X(s)ds + B(t), \tag{5}$$

where, slightly abusing the notation, we still use $Y(t)$ to denote the output corresponding to the input $X(s)$, and $B(t)$ denotes the standard Brownian motion ($Z(t)$ can be viewed as a generalized derivative of $B(t)$); equivalently, the channel (5) can be seen as the original channel (1) concatenated with an integrator circuit. As opposed to white Gaussian noises, which only exist as generalized functions [69], Brownian motions are well-defined stochastic processes and have been extensively studied in probability theory. Here we remark that, via a routine orthonormal decomposition argument, both of the two channels are equivalent to a parallel channel consisting of infinitely many Gaussian sub-channels [5].

An immediate and convenient consequence of such a formulation is that many notions in discrete time, including mutual information and typical sets, carry over to the continuous-time setting, which will rid us of the nuisances arising from the bandwidth limit assumption. Indeed, such a framework yields a fundamental formula for the mutual information of the channel (5) [16, 43] and a clean and direct proof [43] that the capacity of (5) is $P/2$; moreover, as evidenced by numerous results collected in [39] on point-to-point Gaussian channels, the use of Brownian motions elevate the level of rigorousness of our treatment, and equip us with a wide range of established techniques and tools from stochastic calculus. Here we remark that Girsanov's theorem, one of the most important theorems in stochastic calculus, lays the foundation of our rigourous treatment; for those who are interested in the technical details in our proofs, we refer to Chapters 6 and 7 of [55], where Girsanov's theorem and its numerous variants are discussed in great details.

Furthermore, as elaborated in Remark 3.9, the Brownian motion formulation is also versatile enough to accommodate feedback and memory in such a way that memory and feedback can be naturally translated to the discrete-time setting: the pathwise continuity of sample paths of a Brownian motion allow the inheritance of temporal causality when the channel is sampled (see Section 2) or approximated (see Section 3). On the other hand, the white Gaussian noise formulation is facing inherent difficulty as far as inheriting temporal causality is concerned: in converting (1) to (2), while $X_n^{(w)}$ are obtained as "time" samples of $X(t)$, $Z_n^{(w)}$ are in fact "space" samples of $Z(t)$, as they are merely the coefficients of the (extended) Karhunen-Loeve decomposition of $Z(t)$ [23, 33, 34]; see also [45] for an in-depth discussion on this.

In this paper, we are concerned with continuous-time Gaussian memory/feedback channels under the Brownian motion formulation in the point-to-point or multi-user setting. Below, we summarize the contributions of this paper.

The point-to-point continuous-time white Gaussian memory/feedback channel can be characterized by the following stochastic differential equation:

$$Y(t) = \int_0^t g(s, W_0^s, Y_0^s) ds + B(t), \quad t \in [0, T], \tag{6}$$

where g is a function from $[0, T] \times C[0, T] \times C[0, T]$ to $\mathbb{R}$. Note that (6) can be interpreted

1) either as a feedback channel, where W_0^s can be rewritten as M, interpreted as the message to be transmitted through the channel, and $g(s)$ can be rewritten as $X(s)$, interpreted as the channel input, which depends on M and Y_0^s, the channel output up to time s that is fed back to the sender,

2) or as a memory channel, where W_0^s can rewritten as X_0^s, interpreted as the channel input, g is "part" of the channel, and $Y(t)$, the channel output at time t, depends on X_0^t and Y_0^t, the channel input and output up to time t that are present in the channel as memory, respectively.

Note that, strictly speaking, the third parameter of g in (6) should be Y_0^{s-}, which, however, can be equivalently replaced by Y_0^s due to the continuity of sample paths of $\{Y(t)\}$.

In Section 2, we prove Theorems 2.1 and 2.4, sampling theorems for a continuous-time Gaussian feedback/memory channel, which naturally connect such a channel with their sampled discrete-time versions. And in Section 3, we prove Theorems 3.3 and 3.5, the so-called approximating theorems, which connect a continuous-time Gaussian feedback/memory channel with its approximated discrete-time versions (in the sense of the Euler-Maruyama approximation [35]). Roughly speaking, a sampling theorem says that a time-sampled channel is "close" to the original channel if the sampling is fine enough, and an approximating theorem says that an approximated channel is "close" to the original channel if the approximation is fine enough, both in an information-theoretic sense. Note that, as elaborated in Remark 3.8, certain version of the approximating theorem boils down to the sampling theorem when there is no memory and feedback in the channel.

Apparently a sampling theorem is of practical value due to the fact it deals with the "real" values of the channel output; and as will be shown in this paper, approximating theorems seem to be surprisingly useful in a number of respects despite the fact it only deals with the "approximated" values of the channel output: it can certainly provide alternative rigorous tools in translating results from discrete time to continuous time; more importantly, as elaborated in Section 4, it also gives us the insights and intuitive ideas in the point-to-point continuous-time setting, which will further inspire us to deliver rigorous treatments of multi-user continuous-time Gaussian channels in Section 5.

More specifically, in Section 5, we derive the capacity regions of a continuous-time white Gaussian multiple access channel (Theorem 5.1), a continuous-time white Gaussian interference channel (Theorem 5.8), and a continuous-time white Gaussian broadcast channel (Theorem 5.11); and applying the the approximating theorems and the ideas and techniques in their proofs, we use relevant results and proofs in discrete time to analyze how feedback affects the capacity regions of families of continuous-time multi-user one-hop Gaussian channels: feedback will increase the capacity regions of some continuous-time Gaussian broadcast (Theorem 5.16) and interference channels (Theorem 5.10), while it will not increase capacity regions of a continuous-time physically degraded Gaussian broadcast channel (Theorem 5.15) and a continuous-time Gaussian multiple access channels (Theorem 5.1).

2 Sampling Theorems

In this section, we will establish sampling theorems for the channel (6), which naturally connect such channels with their sampled discrete-time versions.

Consider the following regularity conditions:

(a) The solution $\{Y(t)\}$ to the stochastic differential equation (6) uniquely exists;

(b)
$$\mathbb{P}\left(\int_0^T g^2(t, W_0^t, Y_0^t)dt < \infty\right) = \mathbb{P}\left(\int_0^T g^2(t, W_0^t, B_0^t)dt\right) = 1;$$

(c)

$$\int_0^T \mathbb{E}[|g(t, W_0^t, Y_0^t)|]dt < \infty.$$

Now, for any $n \in \mathbb{N}$, choose $t_{n,0}, t_{n,1}, \ldots, t_{n,n} \in \mathbb{R}$ such that

$$0 = t_{n,0} < t_{n,1} < \cdots < t_{n,n-1} < t_{n,n} = T,$$

and let $\Delta_n \triangleq \{t_{n,0}, t_{n,1}, \ldots, t_{n,n}\}$. Sampling the channel (6) over the time interval $[0, T]$ with respect to Δ_n, we obtain its sampled discrete-time version as follows:

$$Y(t_{n,i}) = \int_0^{t_{n,i}} g(s, W_0^s, Y_0^s)ds + B(t_{n,i}), \quad i = 0, 1, \ldots, n. \tag{7}$$

The sequence $\{\Delta_n\}$ is said to be *increasingly refined* if $\Delta_n \subset \Delta_{n+1}$ for any $n \in \mathbb{N}$ and

$$\delta_{\Delta_n} \triangleq \max_{i=1,3,\ldots,n} (t_{n,i} - t_{n,i-1}) \to 0. \tag{8}$$

The sequence $\{\Delta_n\}$ is said to be *equidistant* if $t_{n,i} - t_{n,i-1} = T/n$ for all feasible i, and we will use δ_n to denote the stepsize of an equidistant Δ_n, i.e., $\delta_n \triangleq t_{n,1} - t_{n,0} = T/n$. Here we note that an equidistant sequence $\{\Delta_n\}$ is not increasingly refined, despite the fact that (8) holds true for such a sequence.

Roughly speaking, the following sampling theorem states that for any sequence of increasingly refined samplings, the mutual information of the sampled discrete-time channel (7) will converge to that of the original channel (6).

Theorem 2.1. *Assume Conditions (a)-(c). For any increasingly refined* $\{\Delta_n\}$*, we have*

$$\lim_{n\to\infty} I(W_0^T; Y(\Delta_n)) = I(W_0^T; Y_0^T),$$

where $Y(\Delta_n) \triangleq \{Y_{t_{n,0}}, Y_{t_{n,1}}, \ldots, Y_{t_{n,n}}\}$.

Proof. First of all, an application of Theorem 7.14 of [55] with Conditions (b) and (c) yields that

$$P\left(\int_0^T E^2[g(t, W_0^t, Y_0^t)|Y_0^t]dt < \infty\right) = 1.$$

Then one verifies that the assumptions of Lemma 7.7 of [55] are all satisfied, which implies that for any w,

$$\mu_Y \sim \mu_{Y|W=w} \sim \mu_B,$$

where "$\sim$" means "equivalent", and moreover,

$$\frac{d\mu_{Y|W}}{d\mu_B}(Y_0^T) = \frac{1}{\mathbb{E}[e^{-\int_0^T g(s)dY(s) + \frac{1}{2}\int_0^T g(s)^2 ds}|Y_0^T, W_0^T]}, \quad \frac{d\mu_Y}{d\mu_B}(Y_0^T) = \frac{1}{\mathbb{E}[e^{-\int_0^T g(s)dY(s) + \frac{1}{2}\int_0^T g(s)^2 ds}|Y_0^T]}, \tag{9}$$

where we have rewritten $g(s, W_0^s, Y_0^s)$ as $g(s)$ for notational simplicity. Here we remark that $\mathbb{E}[e^{-\int_0^T g(s)dY(s) + \frac{1}{2}\int_0^T g(s)^2 ds}|Y_0^T, W_0^T]$ is in fact equal to $e^{-\int_0^T g(s)dY(s) + \frac{1}{2}\int_0^T g(s)^2 ds}$, but we keep it the way it is as above for an easy comparison.

Note that it follows from $\mathbb{E}[d\mu_B/d\mu_Y(Y_0^T)] = 1$ that

$$\mathbb{E}[e^{-\int_0^T g(s)dY(s)+\frac{1}{2}\int_0^T g(s)^2 ds}] = 1,$$

which is equivalent to

$$\mathbb{E}[e^{-\int_0^T g(s)dB(s)-\frac{1}{2}\int_0^T g(s)^2 ds}] = 1.$$

Then, a parallel argument as in the proof of Theorem 7.1 of [55] further implies that for any Δ_n,

$$\frac{d\mu_{Y|W}}{d\mu_B}(Y(\Delta_n)) = \frac{1}{\mathbb{E}[e^{-\int_0^T g(s)dY(s)+\frac{1}{2}\int_0^T g(s)^2 ds}|Y(\Delta_n), W_0^T]}, \quad \frac{d\mu_Y}{d\mu_B}(Y(\Delta_n)) = \frac{1}{\mathbb{E}[e^{-\int_0^T g(s)dY(s)+\frac{1}{2}\int_0^T g(s)^2 ds}|Y(\Delta_n)]}, \tag{10}$$

where, similarly as before, we have defined

$$B(\Delta_n) \triangleq \{B_{t_{n,0}}, \ldots, B_{t_{n,n}}\},$$

and moreover,

$$\frac{d\mu_{Y|W}}{d\mu_B}(Y(\Delta_n)) \triangleq \frac{d\mu_{Y(\Delta_n)|W}}{d\mu_{B(\Delta_n)}}(Y(\Delta_n)), \quad \frac{d\mu_Y}{d\mu_B}(Y(\Delta_n)) \triangleq \frac{d\mu_{Y(\Delta_n)}}{d\mu_{B(\Delta_n)}}(Y(\Delta_n)).$$

Then, by definition, we have

$$I(W_0^T; Y(\Delta_n)) = \mathbb{E}\left[\log \frac{d\mu_{Y|W}}{d\mu_B}(Y(\Delta_n))\right] - \mathbb{E}\left[\log \frac{d\mu_Y}{d\mu_B}(Y(\Delta_n))\right].$$

Notice that it can be easily checked that $e^{-\int_0^T g(s)dY(s)+\frac{1}{2}\int_0^T g(s)^2 ds}$ integrable, which further implies that $\left\{\frac{d\mu_{Y|W}}{d\mu_B}(Y(\Delta_n))\right\}$ and $\left\{\frac{d\mu_Y}{d\mu_B}(Y(\Delta_n))\right\}$ are both martingales, and therefore,

$$\frac{d\mu_{Y|W}}{d\mu_B}(Y(\Delta_n)) \to \frac{d\mu_{Y|W}}{d\mu_B}(Y_0^T), \quad \frac{d\mu_Y}{d\mu_B}(Y(\Delta_n)) \to \frac{d\mu_Y}{d\mu_B}(Y_0^T), \text{ a.s.}$$

Now, by Jensen's inequality, we have

$$\mathbb{E}\left[-\int_0^T g(s)dY(s) + \frac{1}{2}\int_0^T g(s)^2 ds \middle| Y(\Delta_n), W_0^T\right] \leq \log \mathbb{E}[e^{-\int_0^T g(s)dY(s)+\frac{1}{2}\int_0^T g(s)^2 ds}|Y(\Delta_n), W_0^T],$$

and, by the easy fact that $\log x \leq x$ for any $x > 0$, we have

$$\log \mathbb{E}[e^{-\int_0^T g(s)dY(s)+\frac{1}{2}\int_0^T g(s)^2 ds}|Y(\Delta_n), W_0^T] \leq \mathbb{E}[e^{-\int_0^T g(s)dY(s)+\frac{1}{2}\int_0^T g(s)^2 ds}|Y(\Delta_n), W_0^T].$$

It then follows that

$$\left|\log \mathbb{E}[e^{-\int_0^T g(s)dY(s)+\frac{1}{2}\int_0^T g(s)^2 ds}|Y(\Delta_n), W_0^T]\right| \leq \left|\mathbb{E}\left[-\int_0^T g(s)dY(s) + \frac{1}{2}\int_0^T g(s)^2 ds \middle| Y(\Delta_n), W_0^T\right]\right|$$
$$+ \mathbb{E}[e^{-\int_0^T g(s)dY(s)+\frac{1}{2}\int_0^T g(s)^2 ds}|Y(\Delta_n), W_0^T].$$

Applying the general Lebesgue dominated convergence theorem (see, e.g., Theorem 19 on Page 89 of [70]), we then have

$$\lim_{n\to\infty} \mathbb{E}\left[\log \frac{d\mu_{Y|W}}{d\mu_B}(Y(\Delta_n))\right] = \mathbb{E}[\log \mathbb{E}[e^{-\int_0^T g(s)dY(s)+\frac{1}{2}\int_0^T g(s)^2 ds}|Y_0^T, W_0^T]] = \mathbb{E}\left[\log \frac{d\mu_{Y|W}}{d\mu_B}(Y_0^T)\right].$$

A completely parallel argument yields that

$$\lim_{n\to\infty} \mathbb{E}\left[\log \frac{d\mu_Y}{d\mu_B}(Y(\Delta_n))\right] = \mathbb{E}[\log \mathbb{E}[e^{-\int_0^T g(s)dY(s)+\frac{1}{2}\int_0^T g(s)^2 ds}|Y_0^T]] = \mathbb{E}\left[\log \frac{d\mu_Y}{d\mu_B}(Y_0^T)\right].$$

So, with the definition

$$I(W_0^T; Y_0^T) = \mathbb{E}\left[\log \frac{d\mu_{Y|W}}{d\mu_B}(Y_0^T)\right] - \mathbb{E}\left[\log \frac{d\mu_Y}{d\mu_B}(Y_0^T)\right],$$

we conclude that

$$\lim_{n\to\infty} I(W_0^T; Y(\Delta_n)) = \mathbb{E}\left[\log \frac{d\mu_{Y|W}}{d\mu_B}(Y_0^T)\right] - \mathbb{E}\left[\log \frac{d\mu_Y}{d\mu_B}(Y_0^T)\right] = I(W_0^T; Y_0^T).$$

□

We next show that the conclusion of Theorem 2.1 can be strengthened with some extra regularity conditions: The following theorem shows that continuous-time white Gaussian channels can be sampled into discrete-time Gaussian channels without losing much information as long as the stepsize of the sampling is small enough. Below and hereafter, defining the "distance" $\|U_0^s - V_0^t\|$ between U_0^s and V_0^t with $0 \leq s \leq t$ as

$$\|U_0^s - V_0^t\| \triangleq \sup_{r\in[0,s]} |U(r) - V(r)| + \sup_{r\in[s,t]} |U(s) - V(r)|, \tag{11}$$

we may assume the following three regularity conditions for the channel (6):

(d) **Uniform Lipschitz condition:** There exists a constant $L > 0$ such that for any $0 \leq s_1, s_2, s_3, t_1, t_2, t_3 \leq T$, any U_0^T, V_0^T, Y_0^T and Z_0^T,

$$|g(s_1, U_0^{s_2}, Y_0^{s_3}) - g(t_1, V_0^{t_2}, Z_0^{t_3})| \leq L(|s_1 - t_1| + \|U_0^{s_2} - V_0^{t_2}\| + \|Y_0^{s_3} - Z_0^{t_3}\|);$$

(e) **Uniform linear growth condition:** There exists a constant $L > 0$ such that for any W_0^T and any Y_0^T,

$$|g(t, W_0^t, Y_0^t)| \leq L(1 + \|W_0^t\| + \|Y_0^t\|),$$

where

$$\|W_0^t\| = \sup_{r\in[0,t]} |W(r)|, \quad \|Y_0^t\| = \sup_{r\in[0,t]} |Y(r)|;$$

(f) **Regularity conditions on W:** There exists $\varepsilon > 0$ such that

$$\mathbb{E}[e^{\varepsilon\|W_0^T\|^2}] < \infty,$$

and for any $K > 0$, there exists $\varepsilon' > 0$ such that

$$\mathbb{E}[e^{K\sup_{|s-t|\leq\varepsilon'}(W(s)-W(t))^2}] < \infty,$$

and there exists a constant $L > 0$ such that for any $\varepsilon'' > 0$,

$$\mathbb{E}[\sup_{|s-t|\leq\varepsilon''}(W(s) - W(t))^2] \leq L\varepsilon''.$$

Remark 2.2. The uniform Lipschitz condition, uniform linear growth condition and their numerous variants are typical assumptions that can guarantee the existence and uniqueness of the solution to a given stochastic differential equation; see, e.g., [57].

We will need the following lemma, which says that Conditions (d)-(f) imply Conditions (a)-(c).

Lemma 2.3. *Assume Conditions (d)-(f). Then, there exists a unique strong solution of (6) with initial value $Y(0) = 0$. Moreover, there exists $\varepsilon > 0$ such that*

$$\mathbb{E}[e^{\varepsilon \|Y_0^T\|^2}] < \infty, \tag{12}$$

which immediately implies Conditions (b) and (c).

Proof. With Conditions (d)-(f), the proof of the existence and uniqueness of the solution to (6) is somewhat standard; see, e.g., Section 5.4 in [57]. So, in the following, we will only prove (12).

For the stochastic differential equation (6), applying Condition (e), we deduce that there exists $L_1 > 0$ such that

$$\begin{aligned} \|Y_0^T\| &\leq \int_0^T L_1(1 + \|W_0^t\| + \|Y_0^t\|)dt + \|B_0^T\| \\ &\leq L_1 T + L_1 T \|W_0^T\| + \|B_0^T\| + \int_0^T L_1 \|Y_0^t\| dt. \end{aligned}$$

Then, applying the Gronwall inequality followed by a straightforward bounding analysis, we deduce that there exists $L_2 > 0$ such that

$$\begin{aligned} \|Y_0^T\| &\leq (L_1 T + L_1 T \|W_0^T\| + \|B_0^T\|) e^{\int_0^T L_1 dt} \\ &= e^{L_1 T}(L_1 T + L_1 T \|W_0^T\| + \|B_0^T\|) \\ &= L_2 + L_2 \|W_0^T\| + L_2 \|B_0^T\|. \end{aligned}$$

Now, for any $\varepsilon > 0$, applying Doob's submartingale inequality, we have

$$\begin{aligned} \mathbb{E}[e^{\varepsilon \|Y_0^T\|^2}] &\leq \mathbb{E}[e^{\varepsilon (L_2 + L_2 \|W_0^T\| + L_2 \|B_0^T\|)^2}] \\ &\leq \mathbb{E}[e^{3\varepsilon (L_2^2 + L_2^2 \|W_0^T\|^2 + L_2^2 \|B_0^T\|^2)}] \\ &= e^{3\varepsilon L_2^2} \mathbb{E}[e^{3\varepsilon L_2^2 \|W_0^T\|^2}] \mathbb{E}[e^{3\varepsilon L_2^2 \|B_0^T\|^2}] \\ &= e^{3\varepsilon L_2^2} \mathbb{E}[e^{3\varepsilon L_2^2 \|W_0^T\|^2}] \mathbb{E}[\sup_{0 \leq t \leq T} e^{3\varepsilon L_2^2 B(t)^2}] \\ &\leq 4 e^{3\varepsilon L_2^2} \mathbb{E}[e^{3\varepsilon L_2^2 \|W_0^T\|^2}] \mathbb{E}[e^{3\varepsilon L_2^2 B(T)^2}], \end{aligned}$$

which, by Condition (f), is finite provided that ε is small enough. □

Roughly speaking, the following theorem states that if the stepsizes of the samplings tend to 0, the mutual information of the channel (7) will converge to that of the channel (6). Note that in this theorem, we do not need the assumption that $\Delta_n \subset \Delta_{n+1}$ for all n, which is required in Theorem 2.4.

Theorem 2.4. *Assume Conditions (d)-(f). For any sequence $\{\Delta_n\}$ with $\delta_{\Delta_n} \to 0$ as n tends to infinity, we have*

$$\lim_{n\to\infty} I(W_0^T; Y(\Delta_n)) = I(W_0^T; Y_0^T).$$

Proof. We proceed in the following steps.

Step 1. In this step, we establish the theorem assuming that there exists $C > 0$ such that for all W_0^T and all Y_0^T,

$$\int_0^T g^2(s, W_0^s, Y_0^s) ds < C. \tag{13}$$

By the definition of mutual information, (9) and (10), we have

$$\begin{aligned} I(W_0^T; Y(\Delta_n)) &= \mathbb{E}\left[\log \frac{d\mu_{Y|W}}{d\mu_B}(Y(\Delta_n))\right] - \mathbb{E}\left[\log \frac{d\mu_Y}{d\mu_B}(Y(\Delta_n))\right] \\ &= -\mathbb{E}[\log \mathbb{E}[e^{-\int_0^T g(s)dY(s) + \frac{1}{2}\int_0^T g(s)^2 ds} | Y(\Delta_n), W_0^T]] + \mathbb{E}[\log \mathbb{E}[e^{-\int_0^T g(s)dY(s) + \frac{1}{2}\int_0^T g(s)^2 ds} | Y(\Delta_n)]] \\ &= -\mathbb{E}[F_n] + \mathbb{E}[G_n], \end{aligned}$$

where, for notational simplicity, we have rewritten $g(s, W_0^s, Y_0^s)$ as $g(s)$.

Step 1.1. In this step, we prove that as n tends to infinity,

$$F_n \to -\int_0^T g(s)dY(s) + \frac{1}{2}\int_0^T g(s)^2 ds, \tag{14}$$

in probability.

Let $\bar{Y}_{\Delta_n,0}^T$ denote the piecewise linear version of Y_0^T with respect to Δ_n; more precisely, for any $i = 0, 1, \ldots, n$, $\bar{Y}_{\Delta_n}(t_{n,i}) = Y(t_{n,i})$, and for any $t_{n,i-1} < s < t_{n,i}$ with $s = \lambda t_{n,i-1} + (1-\lambda)t_{n,i}$ for some $0 < \lambda < 1$, $\bar{Y}_{\Delta_n}(s) = \lambda Y(t_{n,i-1}) + (1-\lambda)Y(t_{n,i})$. Let $\bar{g}_{\Delta_n}(s, W, \bar{Y}_{\Delta_n,0}^s)$ denote the piecewise "flat" version of $g(s, W, \bar{Y}_{\Delta_n,0}^s)$ with respect to Δ_n; more precisely, for any $t_{n,i-1} \le s < t_{n,i}$, $\bar{g}_{\Delta_n}(s, W_0^s, \bar{Y}_{\Delta_n,0}^s) = g(t_{n,i-1}, W_0^{t_{n,i-1}}, \bar{Y}_{\Delta_n,0}^{t_{n,i-1}})$.

Rewriting $\bar{g}_{\Delta_n}(s, W_0^s, \bar{Y}_{\Delta_n,0}^s)$ as $\bar{g}_{\Delta_n}(s)$, we have

$$\begin{aligned} F_n &= -\log \mathbb{E}[e^{-\int_0^T g(s)dY(s) + \frac{1}{2}\int_0^T g^2(s)ds} | Y(\Delta_n), W_0^T] \\ &= -\log \mathbb{E}[e^{-\int_0^T \bar{g}_{\Delta_n}(s)dY(s) + \frac{1}{2}\int_0^T \bar{g}_{\Delta_n}^2(s)ds - \int_0^T (g(s)-\bar{g}_{\Delta_n}(s))dY(s) + \frac{1}{2}\int_0^T (g^2(s)-\bar{g}_{\Delta_n}^2(s))ds} | Y(\Delta_n), W_0^T] \\ &= -\log e^{-\int_0^T \bar{g}_{\Delta_n}(s)dY(s) + \frac{1}{2}\int_0^T \bar{g}_{\Delta_n}^2(s)ds} \mathbb{E}[e^{-\int_0^T (g(s)-\bar{g}_{\Delta_n}(s))dB(s) - \frac{1}{2}\int_0^T (g(s)-\bar{g}_{\Delta_n}(s))^2(s)ds} | Y(\Delta_n), W_0^T] \\ &= -\int_0^T \bar{g}_{\Delta_n}(s)dY(s) - \frac{1}{2}\int_0^T \bar{g}_{\Delta_n}^2(s)ds - \log \mathbb{E}[e^{-\int_0^T (g(s)-\bar{g}_{\Delta_n}(s))dB(s) - \frac{1}{2}\int_0^T (g(s)-\bar{g}_{\Delta_n}(s))^2 ds} | Y(\Delta_n), W_0^T], \end{aligned}$$

where we have used the fact that

$$\mathbb{E}[e^{-\int_0^T \bar{g}_{\Delta_n}(s)dY(s) + \frac{1}{2}\int_0^T \bar{g}_{\Delta_n}^2(s)ds} | Y(\Delta_n), W_0^T] = e^{-\int_0^T \bar{g}_{\Delta_n}(s)dY(s) + \frac{1}{2}\int_0^T \bar{g}_{\Delta_n}^2(s)ds},$$

since $\bar{g}_{\Delta_n}(s)$ is a function depending only on W_0^T and $Y(\Delta_n)$.

We now prove the following convergence:

$$\mathbb{E}\left[\left(\left(-\int_0^T \bar{g}_{\Delta_n}(s)dY(s) - \frac{1}{2}\int_0^T \bar{g}_{\Delta_n}^2(s)ds\right) - \left(-\int_0^T g(s)dY(s) - \frac{1}{2}\int_0^T g^2(s)ds\right)\right)^2\right] \to 0, \tag{15}$$

which will imply that

$$-\int_0^T \bar{g}_{\Delta_n}(s)dY(s) - \frac{1}{2}\int_0^T \bar{g}^2_{\Delta_n}(s)ds \to -\int_0^T g(s)dY(s) - \frac{1}{2}\int_0^T g^2(s)ds$$

in probability. Apparently, to prove (15), we only need to prove that

$$\mathbb{E}\left[\left(-\int_0^T (g(s)-\bar{g}_{\Delta_n}(s))dB(s) - \frac{1}{2}\int_0^T (g(s)-\bar{g}_{\Delta_n}(s))^2 ds\right)^2\right] \to 0. \tag{16}$$

To establish (16), notice that, by the Itô isometry [63], we have

$$\mathbb{E}\left[\left(\int_0^T (g(s)-\bar{g}_{\Delta_n}(s))dB(s)\right)^2\right] = \mathbb{E}\left[\int_0^T (g(s)-\bar{g}_{\Delta_n}(s))^2 ds\right],$$

which means we only need to prove that as $n \to \infty$,

$$\mathbb{E}\left[\int_0^T (g(s)-\bar{g}_{\Delta_n}(s))^2 ds\right] \to 0. \tag{17}$$

To see this, we note that, by Conditions (d) and (e), there exists $L_1 > 0$ such that for any $s \in [0, T]$ with $t_{n,i-1} \leq s < t_{n,i}$,

$$\begin{aligned}
&|g(s, W_0^s, \bar{Y}^s_{\Delta_n,0}) - \bar{g}_{\Delta_n}(s, W_0^s, \bar{Y}^s_{\Delta_n,0})| \\
&\quad = |g(s, W_0^s, \bar{Y}^s_{\Delta_n,0}) - g(t_{n,i-1}, W_0^{t_{n,i-1}}, \bar{Y}^{t_{n,i-1}}_{\Delta_n,0})| \\
&\quad \leq L_1(|s-t_{n,i-1}| + \|W_0^s - W_0^{t_{n,i-1}}\| + \|\bar{Y}^s_{\Delta_n,0} - \bar{Y}^{t_{n,i-1}}_{\Delta_n,0}\|) \\
&\quad \leq L_1(|s-t_{n,i-1}| + \|W_0^s - W_0^{t_{n,i-1}}\| + |Y(t_{n,i}) - Y(t_{n,i-1})|) \\
&\quad \leq L_1\delta_{\Delta_n} + L_1 \sup\nolimits_{r\in[t_{n,i-1},t_{n,i}]} |W(r) - W(t_{n,i-1})| \\
&\quad + L_1\delta_{\Delta_n} + L_1\delta_{\Delta_n}\|W_0^T\| + L_1\delta_{\Delta_n}\|Y_0^T\| + |B(t_{n,i}) - B(t_{n,i-1})|.
\end{aligned} \tag{18}$$

Moreover, by Lemma 2.3 and Condition (f), both $\|Y_0^T\|$ and $\|W_0^T\|$ are quadratically integrable. And furthermore, by Condition (f), we deduce that for any $t_{n,i-1} \leq s < t_{n,i}$,

$$\mathbb{E}[\sup\nolimits_{r\in[t_{n,i-1},t_{n,i}]}(W(r) - W(t_{n,i-1}))^2] \leq L_2\delta_{\Delta_n}, \tag{19}$$

for some $L_2 > 0$, and by the Itô isometry, we deduce that

$$\mathbb{E}[(B(t_{n,i}) - B(t_{n,i-1}))^2] = (t_{n,i} - t_{n,i-1}) \leq \delta_{\Delta_n}. \tag{20}$$

It can be readily checked that (18), (19) and (20) imply (17), which in turn implies (15), as desired.

We now prove that as n tends to infinity,

$$\mathbb{E}[|\mathbb{E}[e^{-\int_0^T (g(s)-\bar{g}_{\Delta_n}(s))dB(s) - \frac{1}{2}\int_0^T (g(s)-\bar{g}_{\Delta_n}(s))^2 ds}|Y(\Delta_n), W_0^T] - 1|] \to 0, \tag{21}$$

which will imply that

$$\log \mathbb{E}[e^{-\int_0^T (g(s)-\bar{g}(s))dB(s) - \frac{1}{2}\int_0^T (g(s)-\bar{g}(s))^2 ds}|Y(\Delta_n), W_0^T] \to 0$$

in probability and furthermore (14). To establish (21), we first note that

$$
\begin{aligned}
&\mathbb{E}[|\mathbb{E}[e^{-\int_0^T(g(s)-\bar{g}_{\Delta_n}(s))dB(s)-\frac{1}{2}\int_0^T(g(s)-\bar{g}_{\Delta_n}(s))^2ds}|Y(\Delta_n),W_0^T]-1|] \\
&\leq \mathbb{E}[\mathbb{E}[|e^{-\int_0^T(g(s)-\bar{g}_{\Delta_n}(s))dB(s)-\frac{1}{2}\int_0^T(g(s)-\bar{g}_{\Delta_n}(s))^2ds}-1||Y(\Delta_n),W_0^T]] \\
&= \mathbb{E}[|e^{-\int_0^T(g(s)-\bar{g}_{\Delta_n}(s))dB(s)-\frac{1}{2}\int_0^T(g(s)-\bar{g}_{\Delta_n}(s))^2ds}-1|] \\
&\leq \mathbb{E}\left[\left|-\int_0^T(g(s)-\bar{g}_{\Delta_n}(s))dB(s)-\frac{1}{2}\int_0^T(g(s)-\bar{g}_{\Delta_n}(s))^2ds\right|e^{|-\int_0^T(g(s)-\bar{g}_{\Delta_n}(s))dB(s)-\frac{1}{2}\int_0^T(g(s)-\bar{g}_{\Delta_n}(s))^2ds|}\right] \\
&\leq \mathbb{E}\left[\left|-\int_0^T(g(s)-\bar{g}_{\Delta_n}(s))dB(s)-\frac{1}{2}\int_0^T(g(s)-\bar{g}_{\Delta_n}(s))^2ds\right|^2\right]\mathbb{E}\left[e^{2|-\int_0^T(g(s)-\bar{g}_{\Delta_n}(s))dB(s)-\frac{1}{2}\int_0^T(g(s)-\bar{g}_{\Delta_n}(s))^2ds|}\right].
\end{aligned}
$$

By (15), we have that as n tends to infinity,

$$
\mathbb{E}\left[\left|-\int_0^T(g(s)-\bar{g}_{\Delta_n}(s))dB(s)-\frac{1}{2}\int_0^T(g(s)-\bar{g}_{\Delta_n}(s))^2ds\right|^2\right]\to 0.
$$

It then follows that, to prove (21), we only need to prove that if δ_{Δ_n} is small enough,

$$
\mathbb{E}\left[e^{2|-\int_0^T(g(s)-\bar{g}_{\Delta_n}(s))dB(s)-\frac{1}{2}\int_0^T(g(s)-\bar{g}_{\Delta_n}(s))^2ds|}\right]<\infty. \tag{22}
$$

Since

$$
\begin{aligned}
&\mathbb{E}\left[e^{2|-\int_0^T(g(s)-\bar{g}_{\Delta_n}(s))dB(s)-\frac{1}{2}\int_0^T(g(s)-\bar{g}_{\Delta_n}(s))^2ds|}\right] \\
&\leq \mathbb{E}\left[e^{2(-\int_0^T(g(s)-\bar{g}_{\Delta_n}(s))dB(s)-\frac{1}{2}\int_0^T(g(s)-\bar{g}_{\Delta_n}(s))^2ds)}\right]+\mathbb{E}\left[e^{2(\int_0^T(g(s)-\bar{g}_{\Delta_n}(s))dB(s)+\frac{1}{2}\int_0^T(g(s)-\bar{g}_{\Delta_n}(s))^2ds)}\right],
\end{aligned} \tag{23}
$$

we only have to prove that the two terms in the above upper bound are both finite provided that δ_{Δ_n} is small enough. Note that for the first term, applying the Cauchy-Schwarz inequality, we have

$$
\begin{aligned}
\mathbb{E}[e^{2(-\int_0^T(g(s)-\bar{g}_{\Delta_n}(s))dB(s)-\frac{1}{2}\int_0^T(g(s)-\bar{g}_{\Delta_n}(s))^2ds)}] &= \mathbb{E}[e^{\int_0^T 2(g(s)-\bar{g}_{\Delta_n}(s))dB(s)-\int_0^T 4(g(s)-\bar{g}_{\Delta_n}(s))^2ds+3\int_0^T(g(s)-\bar{g}_{\Delta_n}(s))^2ds}] \\
&\leq \mathbb{E}[e^{\int_0^T 4(g(s)-\bar{g}_{\Delta_n}(s))dB(s)-\int_0^T 8(g(s)-\bar{g}_{\Delta_n}(s))^2ds}]\mathbb{E}[e^{6\int_0^T(g(s)-\bar{g}_{\Delta_n}(s))^2ds}].
\end{aligned}
$$

It is well known that an application of Fatou's lemma yields that

$$
\mathbb{E}[e^{\int_0^T 4(g(s)-\bar{g}_{\Delta_n}(s))dB(s)-\int_0^T 8(g(s)-\bar{g}_{\Delta_n}(s))^2ds}]\leq 1, \tag{24}
$$

and by (18), we deduce that there exists $L_3>0$ such that

$$
\mathbb{E}[e^{6\int_0^T(g(s)-\bar{g}_{\Delta_n}(s))^2ds}]\leq e^{L_3\delta_{\Delta_n}^2}\mathbb{E}[e^{L_3\|B_0^{\delta_{\Delta_n}}\|^2}]\mathbb{E}[e^{L_3\delta_{\Delta_n}^2\|Y_0^T\|^2}]\mathbb{E}[e^{L_3\delta_{\Delta_n}^2\|W_0^T\|^2}]\mathbb{E}[e^{L_3\sup_{|s-t|\leq\delta_{\Delta_n}}|W(s)-W(t)|^2}].
$$

Note that it follows from Doob's submartingale inequality that if δ_{Δ_n} is small enough,

$$
\mathbb{E}[e^{L_3\|B_0^{\delta_{\Delta_n}}\|^2}]<\infty,
$$

and by Lemma 2.3, we also deduce that if δ_{Δ_n} is small enough,

$$
\mathbb{E}[e^{L_3\delta_{\Delta_n}^2\|Y_0^T\|^2}]<\infty,
$$

which, together with Condition (f), yields that for the first term in (23)

$$\mathbb{E}[e^{2(-\int_0^T (g(s)-\bar{g}_{\Delta_n}(s))dB(s)-\frac{1}{2}\int_0^T (g(s)-\bar{g}_{\Delta_n}(s))^2 ds)}] < \infty. \tag{25}$$

A completely parallel argument will yield that for the second term in (23)

$$\mathbb{E}\left[e^{(2\int_0^T (g(s)-\bar{g}_{\Delta_n}(s))dB(s)+\frac{1}{2}\int_0^T (g(s)-\bar{g}_{\Delta_n}(s))^2 ds)}\right] < \infty,$$

which, together with (25), immediately implies (22), which in turn implies (21), as desired.

Step 1.2. In this step, we prove that as n tends to infinity,

$$G_n \to \log \mathbb{E}[e^{-\int_0^T g(s)dY(s)+\frac{1}{2}\int_0^T g(s)^2 ds}|Y_0^T], \tag{26}$$

in probability.

First, we note that G_n can be rewritten as

$$\begin{aligned} G_n &= -\log \mathbb{E}[e^{-\int_0^T g(s)dY(s)+\frac{1}{2}\int_0^T g^2(s)ds}|Y(\Delta_n)] \\ &= -\log \mathbb{E}[\mathbb{E}[e^{-\int_0^T g(s)dY+\frac{1}{2}\int_0^T g^2(s)ds}|Y_0^T]|Y(\Delta_n)]. \end{aligned}$$

And, by Theorem 7.23 of [55],

$$\mathbb{E}[e^{-\int_0^T g(s)dY(s)+\frac{1}{2}\int_0^T g^2(s)ds}|Y_0^T] = \int \frac{d\mu_{Y|W=w}}{d\mu_B}(Y_0^T)d\mu_W(w),$$

where

$$\frac{d\mu_{Y|W=w}}{d\mu_B}(Y_0^T) = e^{\int_0^T g(w_0^s)dY(s)-\frac{1}{2}\int_0^T g^2(w_0^s)ds},$$

where we have rewritten $g(s, w_0^s, Y_0^s)$ as $g(w_0^s)$ for notational simplicity. It then follows that

$$\begin{aligned} G_n &= -\log \mathbb{E}\left[\int e^{\int_0^T g(w_0^s)dY(s)-\frac{1}{2}\int_0^T g^2(w_0^s)ds} d\mu_W(w) \middle| Y(\Delta_n)\right] \\ &= -\log \int \mathbb{E}\left[e^{\int_0^T g(w_0^s)dY(s)-\frac{1}{2}\int_0^T g^2(w_0^s)ds} \middle| Y(\Delta_n)\right] d\mu_W(w). \end{aligned}$$

Now, we consider the following difference:

$$\begin{aligned} &\mathbb{E}[e^{-\int_0^T g(s)dY(s)+\frac{1}{2}\int_0^T g^2(s)ds}|Y_0^T] - \mathbb{E}[e^{-\int_0^T g(s)dY(s)+\frac{1}{2}\int_0^T g^2(s)ds}|Y(\Delta_n)] \\ &= \int e^{\int_0^T g(w_0^s)dY(s)-\frac{1}{2}\int_0^T g^2(w_0^s)ds} d\mu_W(w) - \mathbb{E}\left[\int e^{\int_0^T g(w_0^s)dY(s)-\frac{1}{2}\int_0^T g^2(w_0^s)ds} d\mu_W(w) \middle| Y(\Delta_n)\right] \\ &= \int e^{\int_0^T g(w_0^s)dY(s)-\frac{1}{2}\int_0^T g^2(w_0^s)ds} d\mu_W(w) - \int \mathbb{E}\left[e^{\int_0^T g(w_0^s)dY(s)-\frac{1}{2}\int_0^T g^2(w_0^s)ds} \middle| Y(\Delta_n)\right] d\mu_W(w) \\ &= \int e^{\int_0^T g(w_0^s)dY(s)-\frac{1}{2}\int_0^T g^2(w_0^s)ds} - e^{\int_0^T \bar{g}_{\Delta_n}(w_0^s)dY(s)-\frac{1}{2}\int_0^T \bar{g}_{\Delta_n}(w_0^s)^2 ds} d\mu_W(w) \\ &\qquad - \int \mathbb{E}[e^{\int_0^T g(w_0^s)dY(s)-\frac{1}{2}\int_0^T g^2(w_0^s)ds} - e^{\int_0^T \bar{g}_{\Delta_n}(w_0^s)dY(s)-\frac{1}{2}\int_0^T \bar{g}^2_{\Delta_n}(w_0^s)ds}|Y(\Delta_n)]d\mu_W(w) \\ &= I_n - J_n. \end{aligned}$$

Note that

$$\begin{aligned}\mathbb{E}[|J_n|] &= \mathbb{E}\left[\left|\int \mathbb{E}[e^{\int_0^T g(w_0^s)dY(s)-\frac{1}{2}\int_0^T g^2(w_0^s)ds} - e^{\int_0^T \bar{g}_{\Delta_n}(w_0^s)dY(s)-\frac{1}{2}\int_0^T \bar{g}^2_{\Delta_n}(w_0^s)ds}|Y(\Delta_n)]d\mu_W(w)\right|\right] \\ &\leq \mathbb{E}\left[\int \mathbb{E}\left[\left|e^{\int_0^T g(w_0^s)dY(s)-\frac{1}{2}\int_0^T g^2(w_0^s)ds} - e^{\int_0^T \bar{g}_{\Delta_n}(w_0^s)dY(s)-\frac{1}{2}\int_0^T \bar{g}^2_{\Delta_n}(w_0^s)ds}\right| |Y(\Delta_n)\right] d\mu_W(w)\right] \\ &= \int \mathbb{E}\left[\left|e^{\int_0^T g(w_0^s)dY(s)-\frac{1}{2}\int_0^T g^2(w_0^s)ds} - e^{\int_0^T \bar{g}_{\Delta_n}(w_0^s)dY(s)-\frac{1}{2}\int_0^T \bar{g}^2_{\Delta_n}(w_0^s)ds}\right|\right] d\mu_W(w).\end{aligned}$$

Applying the inequality that for any $x, y \in \mathbb{R}$,

$$|e^x - e^y| = |e^y(e^{x-y}-1)| \leq e^y(|x-y|e^{x-y} + |x-y|e^{y-x}) = |x-y|(e^x + e^{2y-x}), \tag{27}$$

we continue

$$\begin{aligned}\mathbb{E}[|J_n|] &\leq \int \mathbb{E}\left[\left|e^{\int_0^T g(w_0^s)dY(s)-\frac{1}{2}\int_0^T g^2(w_0^s)ds} - e^{\int_0^T \bar{g}_{\Delta_n}(w_0^s)dY(s)-\frac{1}{2}\int_0^T \bar{g}^2_{\Delta_n}(w_0^s)ds}\right|\right] d\mu_W(w) \\ &\leq \int \mathbb{E}\left[\left|\int_0^T g(w_0^s) - \bar{g}_{\Delta_n}(w_0^s)dY(s) - \frac{1}{2}\int_0^T g^2(w_0^s) - \bar{g}^2_{\Delta_n}(w_0^s)ds\right|\right. \\ &\quad \times \left.\left(e^{\int_0^T g(w_0^s)dY(s)-\frac{1}{2}\int_0^T g^2(w_0^s)ds} + e^{(2\int_0^T \bar{g}_{\Delta_n}(w_0^s)dY(s)-\int_0^T \bar{g}^2_{\Delta_n}(w_0^s)ds)-(\int_0^T g(w_0^s)dY(s)-\frac{1}{2}\int_0^T g^2(w_0^s)ds)}\right)\right] d\mu_W(w) \\ &\leq \int \mathbb{E}\left[\left|\int_0^T (g(w_0^s) - \bar{g}_{\Delta_n}(w_0^s))dB(s)\right| + \left|\int_0^T (g(w_0^s) - \bar{g}_{\Delta_n}(w_0^s))(g(s) - \frac{1}{2}g(w_0^s) - \frac{1}{2}\bar{g}_{\Delta_n}(w_0^s))ds\right|\right. \\ &\quad \times \left.\left(e^{\int_0^T g(w_0^s)dY(s)-\frac{1}{2}\int_0^T g^2(w_0^s)ds} + e^{(2\int_0^T \bar{g}_{\Delta_n}(w_0^s)dY(s)-\int_0^T \bar{g}^2_{\Delta_n}(w_0^s)ds)-(\int_0^T g(w_0^s)dY(s)-\frac{1}{2}\int_0^T g^2(w_0^s)ds)}\right)\right] d\mu_W(w) \\ &\leq \int \mathbb{E}\left[\left|\int_0^T (g(w_0^s) - \bar{g}_{\Delta_n}(w_0^s))dB(s)\right| + (L\delta_{\Delta_n} + L\sup_{|s-t|\leq\delta_{\Delta_n}}|w(s)-w(t)| + L\delta_{\Delta_n}\right. \\ &\quad + L\delta_{\Delta_n}\|w_0^T\| + L\delta_{\Delta_n}\|Y_0^T\| + \sup_{|s-t|\leq\delta_{\Delta_n}}|B(s)-B(t)|)\left(\int_0^T \left|g(s) - \frac{1}{2}g(w_0^s) - \frac{1}{2}\bar{g}_{\Delta_n}(w_0^s)\right| ds\right) \\ &\quad \times \left.\left(e^{\int_0^T g(w_0^s)dY(s)-\frac{1}{2}\int_0^T g^2(w_0^s)ds} + e^{(2\int_0^T \bar{g}_{\Delta_n}(w_0^s)dY(s)-\int_0^T \bar{g}^2_{\Delta_n}(w_0^s)ds)-(\int_0^T g(w_0^s)dY(s)-\frac{1}{2}\int_0^T g^2(w_0^s)ds)}\right)\right] d\mu_W(w).\end{aligned}$$

Now, using (18), Condition (f) and the Itô isometry, we deduce that as $n \to \infty$,

$$\int \mathbb{E}\left[\left|\int_0^T g(w_0^s) - \bar{g}_{\Delta_n}(w_0^s)dB(s)\right|^2\right] d\mu_W(w) \to 0,$$

and as n tends to infinity,

$$\begin{aligned}&\int \mathbb{E}[(L\delta_{\Delta_n} + L\sup_{|s-t|\leq\delta_{\Delta_n}}|w(s)-w(t)| + L\delta_{\Delta_n} \\ &+L\delta_{\Delta_n}\|w_0^T\| + L\delta_{\Delta_n}\|Y_0^T\| + \sup_{|s-t|\leq\delta_{\Delta_n}}|B(s)-B(t)|)^2]d\mu_W(w) \to 0.\end{aligned}$$

Now, using a similar argument as above with (13) and Lemma 2.3, we can show that for any constant K,

$$\mathbb{E}[e^{\int_0^T K\bar{g}^2_{\Delta_n}(s)ds}] = \mathbb{E}[e^{\int_0^T K(\bar{g}_{\Delta_n}(s)-g(s)+g(s))^2ds}] = \mathbb{E}[e^{\int_0^T K(2(\bar{g}_{\Delta_n}(s)-g(s))^2+2g^2(s))ds}] < \infty,$$

provided that n is large enough, which, coupled with a similar argument as in the derivation of (25), proves that for n large enough,

$$\int \mathbb{E}\left[\left(e^{\int_0^T g(w_0^s)dY(s)-\frac{1}{2}\int_0^T g^2(w_0^s)ds} + e^{(2\int_0^T \bar{g}(w_0^s)dY(s)-\int_0^T \bar{g}^2(w_0^s)ds)-(\int_0^T g(w_0^s)dY(s)-\frac{1}{2}\int_0^T g^2(w_0^s)ds)}\right)^2\right] d\mu_W(w) < \infty,$$

and furthermore

$$\int \left(\int_0^T \left|g(s)-\frac{1}{2}g(w_0^s)-\frac{1}{2}\bar{g}_{\Delta_n}(w_0^s)\right| ds\right)^2$$
$$\times\left(e^{\int_0^T g(w_0^s)dY(s)-\frac{1}{2}\int_0^T g^2(w_0^s)ds} + e^{(2\int_0^T \bar{g}_{\Delta_n}(w_0^s)dY(s)-\int_0^T \bar{g}^2_{\Delta_n}(w_0^s)ds)-(\int_0^T g(w_0^s)dY(s)-\frac{1}{2}\int_0^T g^2(w_0^s)ds)}\right)^2 d\mu_W(w) < \infty,$$

which further implies that as n tends to infinity,

$$\mathbb{E}[|J_n|] \to 0. \tag{28}$$

In a similar fashion, we can also show that

$$\mathbb{E}[|I_n|] \to 0.$$

Now, we are ready to conclude that

$$\mathbb{E}|\mathbb{E}[e^{-\int_0^T g(s)dY(s)+\frac{1}{2}\int_0^T g^2(s)ds}|Y_0^T] - \mathbb{E}[e^{-\int_0^T g(s)dY(s)+\frac{1}{2}\int_0^T g^2(s)ds}|Y(\Delta_n)]| \to 0,$$

which further implies that as n tends to infinity,

$$\mathbb{E}[e^{-\int_0^T g(s)dY(s)+\frac{1}{2}\int_0^T g^2(s)ds}|Y(\Delta_n)] \to \mathbb{E}[e^{-\int_0^T g(s)dY+\frac{1}{2}\int_0^T g^2(s)ds}|Y_0^T]$$

in probability, which in turn implies (26), as desired.

Step 1.3. In this step, we show the convergence of $\{\mathbb{E}[F_n]\}$ and $\{\mathbb{E}[G_n]\}$ and further establish the theorem under the condition (13).

Now, using the concavity of the log function and the fact that $\log x \leq x$, we can obtain the upper bounds and lower bounds of F_n and G_n as follows:

$$F_n \leq \left|\mathbb{E}\left[-\int_0^T g(s)dY(s)+\frac{1}{2}\int_0^T g^2(s)ds\middle| Y_{\Delta_n}, W_0^T\right]\right| + \mathbb{E}\left[e^{-\int_0^T g(s)dY(s)+\frac{1}{2}\int_0^T g^2(s)ds}\middle| Y_{\Delta_n}, W_0^T\right],$$

$$F_n \geq -\left|\mathbb{E}\left[-\int_0^T g(s)dY(s)+\frac{1}{2}\int_0^T g^2(s)ds\middle| Y_{\Delta_n}, W_0^T\right]\right| - \mathbb{E}\left[e^{-\int_0^T g(s)dY(s)+\frac{1}{2}\int_0^T g^2(s)ds}\middle| Y_{\Delta_n}, W_0^T\right],$$

and

$$G_n \leq \left|\mathbb{E}\left[-\int_0^T g(s)dY(s)+\frac{1}{2}\int_0^T g^2(s)ds\middle| Y_{\Delta_n}\right]\right| + \mathbb{E}\left[e^{-\int_0^T g(s)dY(s)+\frac{1}{2}\int_0^T g^2(s)ds}\middle| Y_{\Delta_n}\right],$$

$$G_n \geq -\left|\mathbb{E}\left[-\int_0^T g(s)dY(s)+\frac{1}{2}\int_0^T g^2(s)ds\middle| Y_{\Delta_n}\right]\right| - \mathbb{E}\left[e^{-\int_0^T g(s)dY(s)+\frac{1}{2}\int_0^T g^2(s)ds}\middle| Y_{\Delta_n}\right].$$

And furthermore, using a similar argument as in **Step 1.1**, we can show that as n tends to infinity,

$$\mathbb{E}\left[-\int_0^T g(s)dY(s)+\frac{1}{2}\int_0^T g^2(s)ds\middle| Y(\Delta_n), W_0^T\right] = \left(-\int_0^T \bar{g}_{\Delta_n}(s)dY(s)+\frac{1}{2}\int_0^T \bar{g}^2_{\Delta_n}(s)ds\right)$$
$$\times\, \mathbb{E}\left[-\int_0^T (g(s)-\bar{g}_{\Delta_n})(s)dY(s)+\frac{1}{2}\int_0^T (g^2(s)-\bar{g}^2_{\Delta_n}(s))ds\middle| Y(\Delta_n), W_0^T\right]$$
$$\to \left(-\int_0^T g(s)dY(s)+\frac{1}{2}\int_0^T g^2(s)ds\right)$$

and

$$\mathbb{E}[e^{-\int_0^T g(s)dY(s)+\frac{1}{2}\int_0^T g^2(s)ds}|Y(\Delta_n), W_0^T]$$
$$= \mathbb{E}[e^{-\int_0^T \bar{g}_{\Delta_n}(s)dY(s)+\frac{1}{2}\int_0^T \bar{g}^2_{\Delta_n}(s)ds-\int_0^T (g(s)-\bar{g}_{\Delta_n}(s))dY(s)+\frac{1}{2}\int_0^T (g^2(s)-\bar{g}^2_{\Delta_n}(s))ds}|Y(\Delta_n), W_0^T]$$
$$= e^{-\int_0^T \bar{g}_{\Delta_n}(s)dY(s)+\frac{1}{2}\int_0^T \bar{g}^2_{\Delta_n}(s)ds}\mathbb{E}[e^{-\int_0^T (g(s)-\bar{g}_{\Delta_n}(s))dB(s)-\frac{1}{2}\int_0^T (g(s)-\bar{g}_{\Delta_n}(s))^2 ds}|Y(\Delta_n), W_0^T]$$
$$\to e^{-\int_0^T g(s)dY(s)+\frac{1}{2}\int_0^T g^2(s)ds}.$$

It then follows from the general Lebesgue dominated convergence theorem that

$$\lim_{n\to\infty}\mathbb{E}[F_n]\to\mathbb{E}\left[-\int_0^T g(s)dY(s)+\frac{1}{2}\int_0^T g(s)^2 ds\right].$$

A parallel argument can be used to show that

$$\lim_{n\to\infty}\mathbb{E}[G_n]=\mathbb{E}[\log\mathbb{E}[e^{-\int_0^T g(s)dY(s)+\frac{1}{2}\int_0^T g(s)^2 ds}|Y_0^T]].$$

So, under the condition (13), we have shown that

$$\lim_{n\to\infty} I(W_0^T; Y(\Delta_n)) = I(W_0^T; Y_0^T).$$

Step 2. In this step, we will use the convergence in **Step 1** and establish the theorem without the condition (13).

Following Page 264 of [55], we define, for any k,

$$\tau_k = \begin{cases}\inf\{t\le T: \int_0^t g^2(s, W_0^s, Y_0^s)ds \ge k\}, \text{ if } \int_0^T g^2(s, W_0^s, Y_0^s)ds\ge k\\ \infty, \text{ if } \int_0^T g^2(s, W_0^s, Y_0^s)ds < k.\end{cases} \tag{29}$$

Then, we again follow [55] and define a truncated version of g as follows:

$$g_{(k)}(t, \gamma_0^t, \phi_0^t) = g(t, \gamma_0^t, \phi_0^t)\mathbf{1}_{\int_0^t g^2(s,\gamma_0^t,\phi_0^s)ds<k}.$$

Now, define a truncated version of Y as follows:

$$Y_{(k)}(t) = \rho\int_0^t g_{(k)}(s, W_0^s, Y_0^s)ds + B(t), \quad t\in[0,T],$$

which, as elaborated on Page 265 in [55], can be rewritten as

$$Y_{(k)}(t) = \rho \int_0^t g_{(k)}(s, W_0^s, Y_{(k),0}^s) ds + B(t), \quad t \in [0, \tau_k \wedge T]. \tag{30}$$

Note that for fixed k, the system in (30) satisfies the condition (13), and so the theorem holds true. To be more precise, note that

$$I(W_0^T; Y_0^{\tau_k}) = \mathbb{E}\left[\log \frac{d\mu_{\tau_k, Y|W}}{d\mu_{\tau_k, B}}(Y_0^{\tau_k})\right] - \mathbb{E}\left[\log \frac{d\mu_{\tau_k, Y}}{d\mu_{\tau_k, B}}(Y_0^{\tau_k})\right],$$

where $\mu_{\tau_k, Y}$ and $\mu_{\tau_k, B}$ respectively denote the truncated versions of μ_Y and μ_B (from time 0 to time τ_n). Applying Theorem 7.10 in [55], we obtain

$$\frac{d\mu_{\tau_k, Y|W}}{d\mu_{\tau_k, B}}(Y_0^{\tau_n}) = e^{\int_0^{\tau_k} g(s) dY(s) - \frac{1}{2}\int_0^{\tau_k} g^2(s) ds},$$

and

$$\frac{d\mu_{\tau_k, Y}}{d\mu_{\tau_k, B}}(Y_0^{\tau_k}) = e^{\int_0^{\tau_k} \hat{g}(s) dY(s) - \frac{1}{2}\int_0^{\tau_k} \hat{g}^2(s) ds},$$

where

$$\hat{g}(s) = \mathbb{E}[g(s, W_0^s, Y_0^s) | Y_0^s].$$

It then follows that

$$I(W_0^T; Y_0^{\tau_k}) = \frac{1}{2}\mathbb{E}\left[\int_0^{\tau_k} (g(s) - \hat{g}(s))^2 ds\right].$$

Notice that it can be easily verified that $\tau_k \to T$ as k tends to infinity, which, together with the monotone convergence theorem, further yields that monotone increasingly,

$$I(W_0^T; Y_0^{\tau_k}) = \frac{1}{2}\mathbb{E}\left[\int_0^{\tau_k} (g(s) - \hat{g}(s))^2 ds\right] \to I(W_0^T; Y_0^T) = \frac{1}{2}\mathbb{E}\left[\int_0^{T} (g(s) - \hat{g}(s))^2 ds\right],$$

as k tends to infinity. By **Step 1**, for any fixed k_i,

$$\lim_{n \to \infty} I(W_0^T; Y(\Delta_n \cap [0, \tau_{k_i}])) = I(W_0^T; Y_0^{\tau_{k_i}}),$$

which means that there exists a sequence $\{n_i\}$ such that, as i tends to infinity, we have, monotone increasingly,

$$I(W_0^T; Y(\Delta_{n_i} \cap [0, \tau_{k_i}])) \to I(W_0^T; Y_0^T).$$

Since, by the fact that $Y_0^{\tau_k}$ coincides with Y_0^T on the interval $[0, \tau_k \wedge T]$, we have

$$I(W_0^T; Y(\Delta_{n_i})) \geq I(W_0^T; Y(\Delta_{n_i} \cap [0, \tau_{k_i}])).$$

Now, using the fact that

$$I(W_0^T; Y(\Delta_{n_i})) \leq I(W_0^T; Y_0^T),$$

we conclude that as i tends to infinity,

$$\lim_{i\to\infty} I(W_0^T; Y(\Delta_{n_i})) = I(W_0^T; Y_0^T).$$

A similar argument can be readily applied to any subsequence of $\{I(W_0^T; Y(\Delta_n))\}$, which will establish the existence of its further subsubsequence that converges to $I(W_0^T; Y_0^T)$, which implies that

$$\lim_{n\to\infty} I(W_0^T; Y(\Delta_n)) = I(W_0^T; Y_0^T).$$

The proof of the theorem is then complete. □

Remark 2.5. Consider the following continuous-time Gaussian feedback channel:

$$Y(t) = \int_0^t X(s, M, Y_0^s)ds + B(t), \quad t \in [0, T].$$

The arguments in the proof of Theorem 2.4 can yield a sampling theorem for minimum mean square error (MMSE), a quantity of central importance in estimation theory. More precisely, under the assumptions of Theorem 2.4, the MMSE

$$\int_0^T \mathbb{E}[(X(s) - \mathbb{E}[X(s)|Y_0^T])^2]ds = \int_0^T \mathbb{E}[X^2(s)] - \mathbb{E}[\mathbb{E}^2[X(s)|Y_0^T]]ds$$

is the limit of the MMSE based on the samples with respect to Δ_n,

$$\int_0^T \mathbb{E}[(X(s) - \mathbb{E}[X(s)|Y(\Delta_n))^2]ds = \int_0^T \mathbb{E}[X^2(s)] - \mathbb{E}[\mathbb{E}^2[X(s)|Y(\Delta_n)]]ds,$$

as n tends to infinity.

To see this, note that the above-mentioned convergence follows from the fact that

$$\mathbb{E}[\mathbb{E}^2[X(s)|Y_0^T]] = \mathbb{E}\left[\left(\frac{\int X(s, m_0^s, Y_0^s)d\mu_{Y|M=m}(Y_0^T)/d\mu_B d\mu_M(m)}{d\mu_Y(Y_0^T)/d\mu_B}\right)^2\right],$$

and

$$\mathbb{E}[\mathbb{E}^2[X(s)|Y(\Delta_n)]] = \mathbb{E}\left[\left(\frac{\int X(s, m_0^s, Y_0^s)d\mu_{Y|M=m}(Y(\Delta_n))/d\mu_B d\mu_M(m)}{d\mu_Y(Y(\Delta_n))/d\mu_B}\right)^2\right],$$

and the proven facts that

$$d\mu_Y(Y(\Delta_n))/d\mu_B = \mathbb{E}[e^{-\int_0^T XdY + \frac{1}{2}\int_0^T X^2 ds}|Y(\Delta_n)]$$

converges to

$$d\mu_Y(Y_0^T)/d\mu_B = \mathbb{E}[e^{-\int_0^T XdY + \frac{1}{2}\int_0^T X^2 ds}|Y_0^T],$$

and that

$$d\mu_{Y|M}(Y(\Delta_n))/d\mu_B = \mathbb{E}[e^{-\int_0^T XdY + \frac{1}{2}\int_0^T X^2 ds}|M, Y(\Delta_n)]$$

converges to

$$d\mu_{Y|M}(Y_0^T)/d\mu_B = \mathbb{E}[e^{-\int_0^T X dY + \frac{1}{2}\int_0^T X^2 ds}|M, Y_0^T],$$

and parallel argument as in establishing the convergence of $\{\mathbb{E}[F_n]\}$ and $\{\mathbb{E}[G_n]\}$ as in the proof of Theorem 2.4.

Similarly, we can also conclude that under the assumptions of Theorem 2.4, the causal MMSE

$$\int_0^T \mathbb{E}[(X(s) - \mathbb{E}[X(s)|Y_0^s])^2]ds = \int_0^T \mathbb{E}[X^2(s)] - \mathbb{E}[\mathbb{E}^2[X(s)|Y_0^s]]ds$$

is the limit of the sampled causal MMSE

$$\int_0^T \mathbb{E}[(X(s) - \mathbb{E}[X(s)|Y(\Delta_n \cap [0,s])])^2]ds = \int_0^T \mathbb{E}[X^2(s)] - \mathbb{E}[\mathbb{E}^2[X(s)|Y(\Delta_n \cap [0,s])]]ds,$$

as n tends to infinity.

3 Approximating Theorems

In this section, we will establish approximating theorems for the channel (6), which naturally connect such channels with their approximated discrete-time versions.

An application of the Euler-Maruyama approximation [35] with respect to Δ_n to (6) will yield a discrete-time sequence $\{Y^{(n)}(t_{n,i}) : i = 0, 1, \dots, n\}$ and a continuous-time process $\{Y^{(n)}(t) : t \in [0,T]\}$, a linear interpolation of $\{Y(t_{n,i})\}$, as follows: Initializing with $Y^{(n)}(0) = 0$, for each $i = 0, 1, \dots, n-1$, we recursively define

$$Y^{(n)}(t_{n,i+1}) = Y^{(n)}(t_{n,i}) + \int_{t_{n,i}}^{t_{n,i+1}} g(s, W_0^{t_{n,i}}, Y_0^{(n),t_{n,i}})ds + B(t_{n,i+1}) - B(t_{n,i}), \tag{31}$$

which is immediately followed by the linear interpolation below:

$$Y^{(n)}(t) = Y^{(n)}(t_{n,i}) + \frac{t - t_{n,i}}{t_{n,i+1} - t_{n,i}}(Y^{(n)}(t_{n,i+1}) - Y^{(n)}(t_{n,i})), \quad t_{n,i} \le t \le t_{n,i+1}. \tag{32}$$

Parallel to Lemma 2.3, we have the following lemma.

Lemma 3.1. *Assume Conditions (d)-(f). Then, there exists $\varepsilon > 0$ and a constant $C > 0$ such that for all n,*

$$\mathbb{E}[e^{\varepsilon \sup_{0 \le t \le T}(Y^{(n)}(t))^2}] < C. \tag{33}$$

Proof. A discrete-time version of the proof of Theorem 2.3 implies that there exists $\varepsilon > 0$ and a constant $C > 0$ such that for all n

$$\mathbb{E}[e^{\varepsilon \sup_{i \in \{0,1,\dots,n\}}(Y^{(n)}(t_{n,i}))^2}] < C,$$

which, together with (32), immediately implies (33). □

The following lemma is parallel to Theorem 10.2.2 in [35].

Lemma 3.2. *Assume Conditions (d)-(f). Then, there exists a constant $C > 0$ such that for all n,*

$$\mathbb{E}[\sup_{0\le t\le T} |Y^{(n)}(t) - Y(t)|^2] \le C\delta_{\Delta_n}.$$

Proof. Note that for any n, we have

$$Y(t_{n,i+1}) = Y(t_{n,i}) + \int_{t_{n,i}}^{t_{n,i+1}} g(s, W_0^s, Y_0^s)ds + B(t_{n,i+1}) - B(t_{n,i}),$$

and

$$Y^{(n)}(t_{n,i+1}) = Y^{(n)}(t_{n,i}) + \int_{t_{n,i}}^{t_{n,i+1}} g(s, W_0^{t_{n,i}}, Y_0^{(n),t_{n,i}})ds + B(t_{n,i+1}) - B(t_{n,i}).$$

It then follows that

$$Y(t_{n,i+1}) - Y^{(n)}(t_{n,i+1}) = Y(t_{n,i}) - Y^{(n)}(t_{n,i}) + \int_{t_{n,i}}^{t_{n,i+1}} (g(s, W_0^s, Y_0^s) - g(s, W_0^{t_{n,i}}, Y_0^{(n),t_{n,i}}))ds. \tag{34}$$

Now, for any t, choose n_0 such that $t_{n,n_0} \le t < t_{n,n_0+1}$. Now, a recursive application of (34), coupled with Conditions (d) and (e), yields that for some $L > 0$,

$$\begin{aligned}
Y(t) - Y^{(n)}(t) &= \sum_{i=0}^{n_0} \int_{t_{n,i}}^{t_{n,i+1}} (g(s, W_0^s, Y_0^s) - g(t_{n,i}, W_0^{t_{n,i}}, Y_0^{(n),t_{n,i}}))ds + \int_{t_{n,n_0+1}}^{t} (g(s, W_0^s, Y_0^s) - g(t_{n,i}, W_0^{t_{n,n_0+1}}, Y_0^{(n),t_{n,n_0+1}}))ds \\
&\le \sum_{i=0}^{n_0} \int_{t_{n,i}}^{t_{n,i+1}} L|s - t_{n,i}| + L\|W_0^s - W_0^{t_{n,i}}\| + L\|Y_0^s - Y_0^{(n),s}\| + L\|Y_0^{(n),s} - Y_0^{(n),t_{n,i}}\|ds \\
&+ \int_{t_{n,n_0+1}}^{t} L|s - t_{n,n_0+1}| + L\|W_0^s - W_0^{t_{n,n_0+1}}\| + L\|Y_0^s - Y_0^{(n),s}\| + L\|Y_0^{(n),s} - Y_0^{(n),t_{n,n_0+1}}\|ds.
\end{aligned}$$

Noticing that for any s with $t_{n,i} \le s < t_{n,i+1}$, we have

$$\|Y_0^{(n),s} - Y_0^{(n),t_{n,i}}\|^2 \le \|Y_0^{(n),t_{n,i+1}} - Y_0^{(n),t_{n,i}}\|^2 \le 2\left|\int_{t_{n,i}}^{t_{n,i+1}} g(s, W_0^{t_{n,i}}, Y_0^{(n),t_{n,i}})ds\right|^2 + 2|B(t_{n,i+1}) - B(t_{n,i})|^2,$$

which, together with Condition (e) and the fact that for all n and i,

$$\mathbb{E}[|B(t_{n,i+1}) - B(t_{n,i})|^2] = O(\delta_{\Delta_n}), \tag{35}$$

implies that

$$\mathbb{E}[\|Y_0^{(n),s} - Y_0^{(n),t_{n,i}}\|^2] = O(\delta_{\Delta_n}). \tag{36}$$

Noting that the constants in the two terms $O(\delta_{\Delta_n})$ in (35) and (36) can be chosen uniform over all n, a usual argument with the Gronwall inequality and Condition (f) applied to $\mathbb{E}[\|Y_0^t - Y_0^{(n),t}\|^2]$ completes the proof of the theorem. □

We are now ready to prove the following theorem:

Theorem 3.3. *Assume Conditions (d)-(f). Then, we have*

$$\lim_{n\to\infty} I(W_0^T; Y^{(n)}(\Delta_n)) = I(W_0^T; Y_0^T).$$

Proof. We proceed in two steps.

Step 1. In this step, we establish the theorem assuming that there exists a constant $C > 0$ such that for all W_0^T and all Y_0^T,

$$\int_0^T g^2(s, W_0^s, Y_0^s)ds < C. \tag{37}$$

We first note that straightforward computations yield

$$\begin{aligned} f_{Y^{(n)}(\Delta_n)|W}(y^{(n)}(\Delta_n)|w_0^T) &= \prod_{i=1}^n f(y_{t_{n,i}}^{(n)}|y_{t_{n,0}}^{(n),t_{n,i-1}}, w_0^{t_{n,i-1}}) \\ &= \prod_{i=1}^n \frac{1}{\sqrt{2\pi(t_{n,i}-t_{n,i-1})}} \exp\left(-\frac{(y_{t_{n,i}}^{(n)} - y_{t_{n,i-1}}^{(n)} - \int_{t_{n,i-1}}^{t_{n,i}} g(s, w_0^{t_{n,i-1}}, y_0^{(n),t_{n,i-1}})ds)^2}{2(t_{n,i}-t_{n,i-1})}\right), \end{aligned}$$

(here we have used the shorter notations $y_{t_{n,i}}^{(n)}, y_{t_{n,i-1}}^{(n)}$ for $y^{(n)}(t_{n,i}), y^{(n)}(t_{n,i-1})$, respectively) and

$$f_{Y^{(n)}(\Delta_n)}(y^{(n)}(\Delta_n)) = \int \prod_{i=1}^n \frac{1}{\sqrt{2\pi(t_{n,i}-t_{n,i-1})}} \exp\left(-\frac{(y_{t_{n,i}}^{(n)} - y_{t_{n,i-1}}^{(n)} - \int_{t_{n,i-1}}^{t_{n,i}} g(s, w_0^{t_{n,i-1}}, y_0^{(n),t_{n,i-1}})ds)^2}{2(t_{n,i}-t_{n,i-1})}\right) d\mu_W(w),$$

which further lead to

$$f_{Y^{(n)}(\Delta_n)|W}(Y^{(n)}(\Delta_n)|W_0^T) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi(t_{n,i}-t_{n,i-1})}} \exp\left(-\frac{(Y_{t_{n,i}}^{(n)} - Y_{t_{n,i-1}}^{(n)} - \int_{t_{n,i-1}}^{t_{n,i}} g(s, W_0^{t_{n,i-1}}, Y_0^{(n),t_{n,i-1}})ds)^2}{2(t_{n,i}-t_{n,i-1})}\right), \tag{38}$$

and

$$f_{Y^{(n)}(\Delta_n)}(Y^{(n)}(\Delta_n)) = \int \prod_{i=1}^n \frac{1}{\sqrt{2\pi(t_{n,i}-t_{n,i-1})}} \exp\left(-\frac{(Y_{t_{n,i}}^{(n)} - Y_{t_{n,i-1}}^{(n)} - \int_{t_{n,i-1}}^{t_{n,i}} g(t_{n,i-1}, w_0^{t_{n,i-1}}, Y_0^{(n),t_{n,i-1}})ds)^2}{2(t_{n,i}-t_{n,i-1})}\right) d\mu_W(w). \tag{39}$$

With (38) and (39), we have

$$I(W_0^T; Y^{(n)}(\Delta_n)) = \mathbb{E}[\log f_{Y^{(n)}(\Delta_n)|W}(Y^{(n)}(\Delta_n)|W_0^T)] - \mathbb{E}[\log f_{Y^{(n)}(\Delta_n)}(Y^{(n)}(\Delta_n))]$$

$$= \mathbb{E}\left[\log \prod_{i=1}^n \exp\left(-\frac{-2\int_{t_{n,i-1}}^{t_{n,i}} g(s, W_0^{t_{n,i-1}}, Y_0^{(n),t_{n,i-1}})ds\ (Y_{t_{n,i}}^{(n)} - Y_{t_{n,i-1}}^{(n)}) + (\int_{t_{n,i-1}}^{t_{n,i}} g(s, W_0^{t_{n,i-1}}, Y_0^{(n),t_{n,i-1}})ds)^2}{2(t_{n,i} - t_{n,i-1})}\right)\right]$$

$$- \mathbb{E}\left[\log \int \prod_{i=1}^n \exp\left(-\frac{-2\int_{t_{n,i-1}}^{t_{n,i}} g(s, w_0^{t_{n,i-1}}, Y_0^{(n),t_{n,i-1}})ds\ (Y_{t_{n,i}}^{(n)} - Y_{t_{n,i-1}}^{(n)}) + (\int_{t_{n,i-1}}^{t_{n,i}} g(s, w_0^{t_{n,i-1}}, Y_0^{(n),t_{n,i-1}})ds)^2}{2(t_{n,i} - t_{n,i-1})}\right) d\mu_W(w)\right]$$

$$= \mathbb{E}\left[\sum_{i=1}^n \left(-\frac{-2\int_{t_{n,i-1}}^{t_{n,i}} g(s, W_0^{t_{n,i-1}}, Y_0^{(n),t_{n,i-1}})ds\ (Y_{t_{n,i}}^{(n)} - Y_{t_{n,i-1}}^{(n)}) + (\int_{t_{n,i-1}}^{t_{n,i}} g(s, W_0^{t_{n,i-1}}, Y_0^{(n),t_{n,i-1}})ds)^2}{2(t_{n,i} - t_{n,i-1})}\right)\right]$$

$$- \mathbb{E}\left[\log \int \exp \sum_{i=1}^n \left(-\frac{-2\int_{t_{n,i-1}}^{t_{n,i}} g(s, w_0^{t_{n,i-1}}, Y_0^{(n),t_{n,i-1}})ds\ (Y_{t_{n,i}}^{(n)} - Y_{t_{n,i-1}}^{(n)}) + (\int_{t_{n,i-1}}^{t_{n,i}} g(s, w_0^{t_{n,i-1}}, Y_0^{(n),t_{n,i-1}})ds)^2}{2(t_{n,i} - t_{n,i-1})}\right) d\mu_W(w)\right].$$

On the other hand, it is well known (see, e.g., [39]) that

$$\begin{aligned}
I(W; Y_0^T) &= \mathbb{E}\left[\log \frac{d\mu_{Y|W}}{d\mu_B}(Y_0^T)\right] - \mathbb{E}\left[\log \frac{d\mu_Y}{d\mu_B}(Y_0^T)\right] \\
&= \mathbb{E}\left[\log \exp\left[\int_0^T g(s, W_0^s, Y_0^s)dY(s) - \frac{1}{2}\int_0^T g^2(s, W_0^s, Y_0^s)ds\right]\right] \\
&- \mathbb{E}\left[\log \int \exp\left[\int_0^T g(s, w_0^s, Y_0^s)dY(s) - \frac{1}{2}\int_0^T g^2(s, w_0^s, Y_0^s)ds d\mu_W(w)\right]\right] \\
&= \mathbb{E}\left[\int_0^T g(s, W_0^s, Y_0^s)dY(s) - \frac{1}{2}\int_0^T g^2(s, W_0^s, Y_0^s)ds\right] \\
&- \mathbb{E}\left[\log \int \exp\left[\int_0^T g(s, w_0^s, Y_0^s)dY(s) - \frac{1}{2}\int_0^T g^2(s, w_0^s, Y_0^s)ds d\mu_W(w)\right]\right].
\end{aligned}$$

Now, we compute

$$\int_0^T g(s, W_0^s, Y_0^s)dY(s) - \sum_{i=1}^n \frac{\int_{t_{n,i-1}}^{t_{n,i}} g(s, W_0^{t_{n,i-1}}, Y_0^{(n),t_{n,i-1}})ds(Y_{t_{n,i}}^{(n)} - Y_{t_{n,i-1}}^{(n)})}{t_{n,i} - t_{n,i-1}}$$

$$= \int_0^T g(s, W_0^s, Y_0^s)dY(s) - \sum_{i=1}^n g(t_{n,i-1}, W_0^{t_{n,i-1}}, Y_0^{(n),t_{n,i-1}})(Y_{t_{n,i}}^{(n)} - Y_{t_{n,i-1}}^{(n)})$$

$$- \sum_{i=1}^n \frac{\int_{t_{n,i-1}}^{t_{n,i}} (g(s, W_0^{t_{n,i-1}}, Y_0^{(n),t_{n,i-1}}) - g(t_{n,i-1}, W_0^{t_{n,i-1}}, Y_0^{(n),t_{n,i-1}}))ds(Y_{t_{n,i}}^{(n)} - Y_{t_{n,i-1}}^{(n)})}{t_{n,i} - t_{n,i-1}}.$$

It can be easily checked that the second term of the right hand side of the above equality converges to 0 in mean. For the first term, we have

$$\int_0^T g(s, W_0^s, Y_0^s) dY(s) - \sum_{i=1}^n g(t_{n,i-1}, W_0^{t_{n,i-1}}, Y_0^{(n),t_{n,i-1}})(Y_{t_{n,i}}^{(n)} - Y_{t_{n,i-1}}^{(n)})$$
$$= \sum_{i=1}^n \int_{t_{n,i-1}}^{t_{n,i}} g(s, W_0^s, Y_0^s) dY(s) - \sum_{i=1}^n g(t_{n,i-1}, W_0^{t_{n,i-1}}, Y_0^{(n),t_{n,i-1}})(Y_{t_{n,i}} - Y_{t_{n,i-1}})$$
$$+ \sum_{i=1}^n g(t_{n,i-1}, W_0^{t_{n,i-1}}, Y_0^{(n),t_{n,i-1}})((Y_{t_{n,i}} - Y_{t_{n,i-1}}) - (Y_{t_{n,i}}^{(n)} - Y_{t_{n,i-1}}^{(n)}))$$
$$= \sum_{i=1}^n \int_{t_{n,i-1}}^{t_{n,i}} g(s, W_0^s, Y_0^s) dY(s) - \sum_{i=1}^n \int_{i=1}^n g(t_{n,i-1}, W_0^{t_{n,i-1}}, Y_0^{(n),t_{n,i-1}}) dY(s)$$
$$+ \sum_{i=1}^n g(t_{n,i-1}, W_0^{t_{n,i-1}}, Y_0^{(n),t_{n,i-1}})((Y_{t_{n,i}} - Y_{t_{n,i-1}}) - (Y_{t_{n,i}}^{(n)} - Y_{t_{n,i-1}}^{(n)}))$$
$$= \sum_{i=1}^n \int_{t_{n,i-1}}^{t_{n,i}} (g(s, W_0^s, Y_0^s) - g(t_{n,i-1}, W_0^{t_{n,i-1}}, Y_0^{(n),t_{n,i-1}})) dY(s)$$
$$+ \sum_{i=1}^n g(t_{n,i-1}, W_0^{t_{n,i-1}}, Y_0^{(n),t_{n,i-1}})((Y_{t_{n,i}} - Y_{t_{n,i-1}}) - (Y_{t_{n,i}}^{(n)} - Y_{t_{n,i-1}}^{(n)})).$$

It then follows from Conditions (d) and (e), Lemmas 2.3, 3.1 and 3.2 that

$$\mathbb{E}\left[\left|\sum_{i=1}^n \frac{\int_{t_{n,i-1}}^{t_{n,i}} g(s, W_0^{t_{n,i-1}}, Y_0^{(n),t_{n,i-1}}) ds (Y_{t_{n,i}}^{(n)} - Y_{t_{n,i-1}}^{(n)})}{t_{n,i} - t_{n,i-1}} - \int_0^T g(s, W_0^s, Y_0^s) dY(s)\right|\right] = O(\delta_{\Delta_n}^{\frac{1}{2}}). \tag{40}$$

And using a similar argument as above, we deduce that

$$\mathbb{E}\left[\left|\frac{1}{2}\sum_{i=1}^n \frac{(\int_{t_{n,i-1}}^{t_{n,i}} g(s, W_0^{t_{n,i-1}}, Y_0^{(n),t_{n,i-1}}) ds)^2}{t_{n,i} - t_{n,i-1}} - \frac{1}{2}\int_0^T g(s, W_0^s, Y_0^s)^2 ds\right|\right] = O(\delta_{\Delta_n}^{\frac{1}{2}}). \tag{41}$$

It then follows from (40) and (41) that as n tends to infinity,

$$\mathbb{E}\left[\left|\sum_{i=1}^n \left(-\frac{-2\int_{t_{n,i-1}}^{t_{n,i}} g(s, W_0^{t_{n,i-1}}, Y_0^{(n),t_{n,i-1}}) ds (Y_{t_{n,i}}^{(n)} - Y_{t_{n,i-1}}^{(n)}) + (\int_{t_{n,i-1}}^{t_{n,i}} g(s, W_0^{t_{n,i-1}}, Y_0^{(n),t_{n,i-1}}) ds)^2}{2(t_{n,i} - t_{n,i-1})}\right)\right.\right.$$
$$\left.\left. - \int_0^T g(s, W_0^s, Y_0^s) dY(s) + \frac{1}{2}\int_0^T g(s, W_0^s, Y_0^s)^2 ds\right|\right] = O(\delta_{\Delta_n}^{\frac{1}{2}}). \tag{42}$$

We now establish the following convergence:

$$\mathbb{E}\left[\log \int \exp A^{(n)}(w) d\mu_W(w)\right] \to \mathbb{E}\left[\log \int \exp A(w) d\mu_W(w)\right]. \tag{43}$$

where

$$A^{(n)}(w) = \sum_{i=1}^{n} \left(-\frac{-2\int_{t_{n,i-1}}^{t_{n,i}} g(s, w_0^{t_{n,i-1}}, Y_0^{(n),t_{n,i-1}})ds(Y_{t_{n,i}}^{(n)} - Y_{t_{n,i-1}}^{(n)}) + (\int_{t_{n,i-1}}^{t_{n,i}} g(s, w_0^{t_{n,i-1}}, Y_0^{(n),t_{n,i-1}})ds)^2}{2(t_{n,i} - t_{n,i-1})} \right).$$

and let

$$A(w) = \int_0^T g(s, w_0^s, Y_0^s)dY(s) - \frac{1}{2}\int_0^T g^2(s, w_0^s, Y_0^s)ds.$$

Note that using a parallel argument as the derivation of (42), we can establish

$$\mathbb{E}\int \left|A^{(n)}(w) - A(w)\right| d\mu_W(w) \to 0, \tag{44}$$

as n tends to infinity; and similarly as in the derivation of (28), from Conditions (d), (e) and (f), Lemmas 2.3, 3.1 and 3.2, we deduce that

$$\mathbb{E}\left[\int \left|\exp A^{(n)}(w) - \exp A(w)\right| d\mu_W(w)\right] \to 0 \tag{45}$$

as n tends to infinity. And note that we always have

$$\left|\log\int \exp A^{(n)}(w)d\mu_W(w)\right| \leq \int \exp A^{(n)}(w)d\mu_W(w) + \left|\int A^{(n)}(w)d\mu_W(w)\right|. \tag{46}$$

So, by the general Lebesgue dominated convergence theorem with (44), (45) and (46), we have

$$\mathbb{E}\left[\log\int \exp A^{(n)}(w)d\mu_W(w)\right] \to \mathbb{E}\left[\log\int \exp A(w)d\mu_W(w)\right].$$

So, under the condition (37), we have established the theorem.

Step 2. In this step, we will use the convergence in **Step 1** and establish the theorem without the condition (13).

Defining the stopping τ_k, $g_{(k)}$ and $Y_{(k)}$ as in the proof of Theorem 2.4, we again have:

$$Y_{(k)}(t) = \rho\int_0^t g_{(k)}(s, W_0^s, Y_{(k),0}^s)ds + B(t), \quad t \in [0, \tau_k \wedge T].$$

For any fixed k, applying the Euler-Maruyama approximation as in (31) and (32) to the above channel with respect to Δ_n, we obtain the process $Y_{(k)}^{(n)}(\cdot)$.

Now, by the fact that

$$\begin{aligned} I(W_0^T; Y^{(n)}(\Delta_n)) &= \mathbb{E}[\log f_{Y^{(n)}(\Delta_n)|W}(Y^{(n)}(\Delta_n)|W_0^T)] - \mathbb{E}[\log f_{Y^{(n)}(\Delta_n)}(Y^{(n)}(\Delta_n))] \\ &= \mathbb{E}[A^{(n)}(W)] - \mathbb{E}\left[\log\int \exp A^{(n)}(w)d\mu_W(w)\right] \\ &\geq 0, \end{aligned}$$

we deduce that

$$
\begin{aligned}
&\mathbb{E}[\log f_{Y^{(n)}(\Delta_n)}(Y^{(n)}(\Delta_n))]\\
&\leq \mathbb{E}[\log f_{Y^{(n)}(\Delta_n)|W}(Y^{(n)}(\Delta_n)|W_0^T)]\\
&= \mathbb{E}\left[\sum_{i=1}^n \left(-\frac{-2\int_{t_{n,i-1}}^{t_{n,i}} g(s, W_0^{t_{n,i-1}}, Y_0^{(n),t_{n,i-1}})ds\ (Y_{t_{n,i}}^{(n)} - Y_{t_{n,i-1}}^{(n)}) + (\int_{t_{n,i-1}}^{t_{n,i}} g(s, W_0^{t_{n,i-1}}, Y_0^{(n),t_{n,i-1}})ds)^2}{2(t_{n,i} - t_{n,i-1})}\right)\right]\\
&= \mathbb{E}\left[\sum_{i=1}^n \frac{(\int_{t_{n,i-1}}^{t_{n,i}} g(s, W_0^{t_{n,i-1}}, Y_0^{(n),t_{n,i-1}})ds)^2}{2(t_{n,i} - t_{n,i-1})}\right]\\
&\leq \mathbb{E}\left[\sum_{i=1}^n \frac{(\int_{t_{n,i-1}}^{t_{n,i}} g(s, W_0^{t_{n,i-1}}, Y_{(k),0}^{(n),t_{n,i-1}})^2 ds) \int_{t_{n,i-1}}^{t_{n,i}} ds}{2(t_{n,i} - t_{n,i-1})}\right]\\
&= \mathbb{E}\left[\sum_{i=1}^n \int_{t_{n,i-1}}^{t_{n,i}} g(s, W_0^{t_{n,i-1}}, Y_{(k),0}^{(n),t_{n,i-1}})^2 ds\right]\\
&= \mathbb{E}\left[\int_0^T g(s, W_0^{\lfloor s\rfloor_{\Delta_n}}, Y_0^{(n),\lfloor s\rfloor_{\Delta_n}})^2 ds\right],
\end{aligned}
$$

where $\lfloor s\rfloor_{\Delta_n}$ denote the unique number n_0 such that $t_{n,n_0} \leq s < t_{n,n_0+1}$. Now, using the easily verifiable fact that

$$
\frac{1}{\int \exp A^{(n)}(w) d\mu_W(w)} = \mathbb{E}[\exp(-A^{(n)}(W))|Y_0^T],
$$

and Jensen's inequality, we deduce that

$$
\mathbb{E}\left[\log \frac{1}{\int \exp A^{(n)}(w) d\mu_W(w)}\right] = \mathbb{E}\left[\log \mathbb{E}[\exp(-A^{(n)}(W))|Y_0^T]\right] \leq \log \mathbb{E}[\exp(-A^{(n)}(W))] \leq 0,
$$

where for the last inequality, we have applied Fatou's lemma as in deriving (24). It then follows that

$$
0 \leq \mathbb{E}[\log f_{Y^{(n)}(\Delta_n)}(Y^{(n)}(\Delta_n))] \leq \mathbb{E}\left[\sum_{i=1}^n \int_{t_{n,i-1}}^{t_{n,i}} g(s, W_0^{t_{n,i-1}}, Y_0^{(n),t_{n,i-1}})^2 ds\right],
$$

which further implies that

$$
I(W_0^T; Y^{(n)}(\Delta_n)) \leq \mathbb{E}\left[\sum_{i=1}^n \int_{t_{n,i-1}}^{t_{n,i}} g(s, W_0^{t_{n,i-1}}, Y_0^{(n),t_{n,i-1}})^2 ds\right]. \tag{47}
$$

Now, using the fact that $Y^{(n)}$ and $Y_{(k)}^{(n)}$ coincide over $[0, \tau_k \wedge T]$, one verifies that for any

$\varepsilon > 0$,

$$\begin{aligned}
I(W_0^T; Y^{(n)}(\Delta_n)) - I(W_0^T; Y^{(n)}_{(k),\Delta_n}) &\leq \mathbb{E}\left[\int_{\tau_k}^T g(s, W_0^{\lfloor s \rfloor_{\Delta_n}}, Y_0^{(n),\lfloor s \rfloor_{\Delta_n}})^2 ds\right] \\
&\leq \mathbb{E}\left[\int_{\tau_k}^T g(s, W_0^{\lfloor s \rfloor_{\Delta_n}}, Y_0^{(n),\lfloor s \rfloor_{\Delta_n}})^2 ds; T - \tau_k \leq \varepsilon\right] + \mathbb{E}\left[\int_{\tau_k}^T g(s, W_0^{\lfloor s \rfloor_{\Delta_n}}, Y_0^{(n),\lfloor s \rfloor_{\Delta_n}})^2 ds; T - \tau_k > \varepsilon\right] \\
&\leq \int_{T-\varepsilon}^T \mathbb{E}\left[g(s, W_0^{\lfloor s \rfloor_{\Delta_n}}, Y_0^{(n),\lfloor s \rfloor_{\Delta_n}})^2\right] ds + \mathbb{E}\left[\int_{\tau_k}^T g(s, W_0^{\lfloor s \rfloor_{\Delta_n}}, Y_0^{(n),\lfloor s \rfloor_{\Delta_n}})^2 ds; T - \tau_k > \varepsilon\right].
\end{aligned}$$

Using the easily verifiable fact that $\{\tau_k\}$ converges to T in probability uniformly over all n and the fact that ε can be arbitrarily small, we conclude that as k tends to infinity, uniformly over all n,

$$I(W_0^T; Y^{(n)}_{(k)}(\Delta_n)) \to I(W_0^T; Y^{(n)}(\Delta_n)). \tag{48}$$

Next, an application of the monotone convergence theorem, together with the fact that $\tau_k \to T$ as k tends to infinity, yields that monotone increasingly

$$I(W_0^T; Y_0^{\tau_n}) = \frac{1}{2}\mathbb{E}\left[\int_0^{\tau_n} (g(s) - \hat{g}(s))^2 ds\right] \to I(W_0^T; Y_0^T) = \frac{1}{2}\mathbb{E}\left[\int_0^T (g(s) - \hat{g}(s))^2 ds\right]$$

as n tends to infinity. By **Step 1**, for any fixed k_i,

$$\lim_{n\to\infty} I(W_0^T; Y^{(n)}_{(k_i)}(\Delta_n)) = I(W_0^T; Y_0^{\tau_{k_i}}),$$

which means that there exists a sequence $\{n_i\}$ such that, as i tends to infinity,

$$I(W; Y^{(n_i)}_{(k_i)}(\Delta_n)) \to I(W; Y_0^T).$$

Moreover, by (48),

$$\lim_{i\to\infty} I(W_0^T; Y^{(n_i)}_{(k_i)}(\Delta_{n_i})) = \lim_{i\to\infty} I(W_0^T; Y^{(n_i)}(\Delta_n)),$$

which further implies that

$$\lim_{i\to\infty} I(W_0^T; Y^{(n_i)}(\Delta_n)) = I(W_0^T; Y_0^T).$$

The theorem then follows from a usual subsequence argument as in the proof of Theorem 2.4. □

Remark 3.4. Consider the following continuous-time Gaussian feedback channel:

$$Y(t) = \int_0^t X(s, M, Y_0^s) ds + B(t), \quad t \in [0, T]. \tag{49}$$

For any Δ_n, we define $\tilde{X}^{(n)}(\cdot)$ as follows: for any t with $t_{n,i} \leq t < t_{n,i+1}$,

$$\tilde{X}^{(n)}(t) = \sum_{j=0}^{i-1} \int_{t_{n,j}}^{t_{n,j+1}} X(s, M, Y_0^{(n),t_{n,j}}) ds + \int_{t_{n,i}}^t X(s, M, Y_0^{(n),t_{n,i}}) ds,$$

Writing $\tilde{X}(t, M, Y_0^{(n),t})$ as $\tilde{X}^{(n)}(t)$ for simplicity, (31) can be rewritten as

$$Y^{(n)}(t_{n,i+1}) = Y^{(n)}(t_{n,i}) + \tilde{X}^{(n)}(t_{n,i+1}) - \tilde{X}^{(n)}(t_{n,i}) + B(t_{n,i+1}) - B(t_{n,i}),$$

for which it can be readily checked that

$$I(\tilde{X}^{(n)}(\Delta_n) \to Y^{(n)}(\Delta_n)) = I(M; Y^{(n)}(\Delta_n)). \tag{50}$$

Theorem 3.3 and the above observation can be used to define continuous-time directed mutual information. To be more precise, the continuous-time directed information from X_0^T to Y_0^T of the channel (49) can be defined as

$$I(X_0^T \to Y_0^T) \triangleq \lim_{n\to\infty} I(\tilde{X}_{\Delta_n}^{(n)} \to Y_{\Delta_n}^{(n)}). \tag{51}$$

Consider the following continuous-time Gaussian channel with possibly delayed feedback:

$$Y(t) = \int_0^t X(s, M, Y_0^{s-D})ds + B(t), \quad t \in [0, T], \tag{52}$$

where $D \geq 0$ denotes the possible delay of the feedback. In [83], the notion of continuous-time directed information from X_0^T to Y_0^T is defined as follows:

$$I_D(X_0^T \to Y_0^T) = \inf_{\Delta_n} \sum_{i=1}^{n} I(Y_{t_{n,i-1}}^{t_{n,i}}; X_{t_{n,0}}^{t_{n,i}} | Y_{t_{n,0}}^{t_{n,i-1}}). \tag{53}$$

It is proven that for the case $D > 0$, using this notion, a connection between information theory and estimation theory can be established as follows:

$$I_D(X_0^T \to Y_0^T) = \frac{1}{2}\int_0^T \mathbb{E}[(X(t) - \mathbb{E}[X(t)|Y_0^t])^2]dt. \tag{54}$$

On the other hand though, it is easy to see that for the case $D = 0$, i.e., there is no delay in the feedback as in (49), the definition in (53) and the equality as in (54) may run into some problems: Consider the extreme scenario and choose $X_t = -Y_t$, then clearly the right hand side of (54) should be equal to 0. On the other hand though, for the left hand side, each small interval in (53) will yield

$$I_D(X_0^{t_{i+1}}; Y_{t_i}^{t_{i+1}} | Y_0^{t_i}) = H(Y_{t_i}^{t_{i+1}} | Y_0^{t_i}),$$

which should be infinite (to see this, note that under the assumption that $X_t = -Y_t$, Y_t has to be an OrnsteinUhlenbeck process, which is a Gaussian Markov process). This further implies that the left-hand side of (54) is infinite, a contradiction. On the other hand, be it the case $D > 0$ or $D = 0$, with the definition in (51), Theorem 3.3 however promises:

$$I(X_0^T \to Y_0^T) = I(M_0^T; Y_0^T) = \frac{1}{2}\int_0^T \mathbb{E}[(X(t) - \mathbb{E}[X(t)|Y_0^t])^2]dt.$$

For any Δ_n, let $W^{(n)}(t)$ denote the piecewise linear version of W_0^T with respect to Δ_n; more precisely, for any $i = 0, 1, \ldots, n$, $W^{(n)}(t_{n,i}) = W(t_{n,i})$, and for any $t_{n,i-1} < s < t_{n,i}$ with $s = \lambda t_{n,i-1} + (1-\lambda)t_{n,i}$ where $0 < \lambda < 1$, $W^{(n)}(s) = \lambda W(t_{n,i-1}) + (1-\lambda)W(t_{n,i})$. The following modified Euler-Maruyama approximation with respect to Δ_n applied to the channel (6) yields a discrete-time sequences $\{Y^{(n)}(t_{n,i}) : i = 0, 1, \ldots, n\}$ and a continuous-time processes $\{Y^{(n)}(t) : t \in [0, T]\}$ as follows: Initializing with $Y^{(n)}(0) = 0$, for each $i = 0, 1, \ldots, n-1$, we recursively define

$$\hat{Y}^{(n)}(t_{n,i+1}) = \hat{Y}^{(n)}(t_{n,i}) + \int_{t_{n,i}}^{t_{n,i+1}} g(s, W_0^{(n),t_{n,i}}, \hat{Y}_0^{(n),t_{n,i}})ds + B(t_{n,i+1}) - B(t_{n,i}), \qquad (55)$$

which is immediately followed by a linear interpolation below:

$$\hat{Y}^{(n)}(t) = \hat{Y}^{(n)}(t_{n,i}) + \frac{t - t_{n,i}}{t_{n,i+1} - t_{n,i}}(\hat{Y}^{(n)}(t_{n,i+1}) - \hat{Y}^{(n)}(t_{n,i})), \quad t_{n,i} \leq t \leq t_{n,i+1}. \qquad (56)$$

Now, using a parallel argument in the proof of Theorem 3.3, we have the following approximating theorem.

Theorem 3.5. *Assume Conditions (d)-(f). Then, we have*

$$\lim_{n\to\infty} I(W_{\Delta_n}^{(n)}; \hat{Y}_{\Delta_n}^{(n)}) = I(W_0^T; Y_0^T).$$

Remark 3.6. When the channel (6) is interpreted as a feedback channel, both $W^{(n)}$ and W are precisely M. When the channel (6) is interpreted as a memory channel, Theorem 3.5 states that the mutual information between its input and output is the limit of that of its approximated input and output (in the sense of the above-mentioned modified Euler-Maruyama approximation).

Remark 3.7. Other variants of the Euler-Maruyama approximation can also be applied to the channel. For instance, under Conditions (d)-(f), for the following variant,

$$Y^{(n)}(t_{n,i+1}) = Y^{(n)}(t_{n,i}) + \int_{t_{n,i}}^{t_{n,i+1}} X(t_{n,i}, W_0^{t_{n,i}}, Y_0^{(n),t_{n,i}})ds + B(t_{n,i+1}) - B(t_{n,i}), \qquad (57)$$

a parallel argument as in the proof of Theorem 3.3 will show that

$$\lim_{n\to\infty} I(W_0^T; Y^{(n)}(\Delta_n)) = I(W_0^T; Y_0^T);$$

and for the variant

$$\hat{Y}^{(n)}(t_{n,i+1}) = \hat{Y}^{(n)}(t_{n,i}) + \int_{t_{n,i}}^{t_{n,i+1}} X(t_{n,i}, W_0^{(n),t_{n,i}}, \hat{Y}_0^{(n),t_{n,i}})ds + B(t_{n,i+1}) - B(t_{n,i}), \qquad (58)$$

we also have

$$\lim_{n\to\infty} I(W_{\Delta_n}^{(n)}; Y^{(n)}(\Delta_n)) = I(W_0^T; Y_0^T).$$

Remark 3.8. When there is no feedback or memeory, Theorem 3.3 boils down to Theorem 2.4: obviously we will have

$$Y^{(n)}(t) = Y(t),$$

which means that Theorem 3.3 actually states

$$\lim_{n\to\infty} I(W_0^T; Y(\Delta_n)) = I(W_0^T; Y_0^T),$$

which is precisely the conclusion of Theorem 2.4. And moreover, by Remark 3.4, we also have

$$\lim_{n\to\infty} I(\tilde{X}^{(n)}(\Delta_n); Y(\Delta_n)) = \lim_{n\to\infty} I(\tilde{X}^{(n)}(\Delta_n) \to Y(\Delta_n)) = I(W_0^T; Y_0^T).$$

Remark 3.9. Apparently, in their full generality, sampling and approximating theorems have subtle differences despite their similar conclusions. Taking advantage of the continuity of sample paths of a Brownian motion, Theorems 2.1 and 2.4 naturally connect continuous-time Gaussian memory/feedback channels with their discrete-time counterparts, whose outputs are precisely sampled outputs of the original continuous-time Gaussian channel. As a result, a sampling theorem possess practical values compared with an approximating theorem, which is somewhat "artificial" in the sense that the outputs of the associated discrete-time channels are only approximated outputs of the original continuous-time channels. Nonetheless, as elaborated later, Theorem 3.3 and 3.5 play a key role in connecting continuous-time and discrete-time channels, as the Euler-Maruyama approximation of a continuous-time channel yields the form of a discrete-time channel typically takes, which allows translation from the results and ideas from the discrete-time setting to the continuous-time setting.

Remark 3.10. Consider the following continuous-time Gaussian feedback channel:

$$Y(t) = \int_0^t X(s, M, Y_0^s) ds + B(t), \quad t \in [0, T].$$

Then, parallel to Remark 2.5, the arguments in the proof of Theorem 3.3 can yield an approximating theorem for the MMSE term. More precisely, with the assumptions in Theorem 3.3, the MMSE

$$\int_0^T \mathbb{E}[(X(s) - \mathbb{E}[X(s)|Y_0^T])^2] ds = \int_0^T \mathbb{E}[X^2(s)] - \mathbb{E}[\mathbb{E}^2[X(s)|Y_0^T]] ds$$

is the limit of the approximated MMSE

$$\int_0^T \mathbb{E}[(X^{(n)}(s) - \mathbb{E}[X^{(n)}(s)|Y^{(n)}(\Delta_n)])^2] ds = \int_0^T \mathbb{E}[(X^{(n)}(s))^2] - \mathbb{E}[\mathbb{E}^2[X^{(n)}(s)|Y^{(n)}(\Delta_n)]] ds,$$

as n tends to infinity. In more detail, the above-mentioned convergence follows from the fact that

$$\int_0^T \mathbb{E}[(X^{(n)}(s))^2] ds \to \int_0^T \mathbb{E}[(X(s))^2] ds$$

and the fact that

$$\mathbb{E}[\mathbb{E}^2[X(s)|Y_0^T]] = \mathbb{E}\left[\left(\frac{\int X(s, m_0^s, Y_0^s) \exp(A(m)) d\mu_M(m)}{\int \exp(A(m)) d\mu_M(m)}\right)^2\right],$$

and the fact that

$$\mathbb{E}[\mathbb{E}^2[X^{(n)}(s)|Y^{(n)}(\Delta_n)]] = \mathbb{E}\left[\left(\frac{\int X^{(n)}(s, m_0^s, Y_0^s)\exp(A^{(n)}(m))d\mu_M(m)}{\int \exp(A^{(n)}(m))d\mu_M(m)}\right)^2\right].$$

Then, using a similar argument as in the proof of Theorem 3.3, we can show

$$\lim_{n\to\infty} \mathbb{E}[\mathbb{E}^2[X^{(n)}(s)|Y^{(n)}(\Delta_n)]] = \mathbb{E}[\mathbb{E}^2[X(s)|Y_0^T]],$$

which implies the claimed convergence.

Similarly, we can also conclude that with the assumptions in Theorem 3.3, the causal MMSE

$$\int_0^T \mathbb{E}[(X(s) - \mathbb{E}[X(s)|Y_0^s])^2]ds = \int_0^T \mathbb{E}[X^2(s)] - \mathbb{E}[\mathbb{E}^2[X(s)|Y_0^s]]ds$$

is the limit of the approximated causal MMSE

$$\int_0^T \mathbb{E}[(X^{(n)}(s) - \mathbb{E}[X^{(n)}(s)|Y^{(n)}(\Delta_n \cap [0, s])^2]ds = \int_0^T \mathbb{E}[(X^{(n)}(s))^2] - \mathbb{E}[\mathbb{E}^2[X^{(n)}(s)|Y^{(n)}(\Delta_n \cap [0, s])]ds,$$

as n tends to infinity.

4 Another Heuristic Approach

Consider the following continuous-time white Gaussian channel with feedback

$$Y(t) = \int_0^t X(s, M, Y_0^s)ds + B(t), \quad t \geq 0, \tag{59}$$

satisfying the power constraint: there exists $P > 0$ such that for any T

$$\int_0^T X^2(s, M, Y_0^s)ds \leq PT. \tag{60}$$

As mentioned in Section 1, it is well-known that the capacity of the above channel is $P/2$ (The same result can be established under alternative power constraints; see, e.g., [39]).

When there is no feedback in the channel, i.e., the channel (59) is actually the same as (1), using the bandwidth approach as in (1)-(4), one can derive the non-feedback capacity heuristically.

In this section, we use our approximation theorems to give an alternative way to "derive" the capacity of (59), which will be referred to as *the approximation approach* in the remainder of the paper. Compared to the bandwidth approach, our approach can handle the feedback due to the fact the Euler-Maruyama approximation preserve the temporal causality. Below we roughly explain this new approach, which will be further developed and used in Section 5, where multiple users may be involved in a communication system.

For fixed $T > 0$, consider the equidistant sequence Δ_n with stepsize $\delta_n = T/n$. Applying the Euler-Maruyama approximation (57) to the channel (59) over the time window $[0, T]$, we obtain

$$Y^{(n)}(t_{n,i+1}) = Y^{(n)}(t_{n,i}) + \int_{t_{n,i}}^{t_{n,i+1}} X(t_{n,i}, M_0^{t_{n,i}}, Y_0^{(n),t_{n,i}})ds + B(t_{n,i+1}) - B(t_{n,i}). \tag{61}$$

By Theorem 3.3 and Remark 3.7, we have

$$I(M;Y_0^T) = \lim_{n\to\infty} I(M;Y^{(n)}(\Delta_n)). \tag{62}$$

Our strategy is to "establish" the capacity for the discrete-time channel (61) first, and then the capacity for the continuous-time channel (59) using the "closeness" between the two channels, as claimed by approximating theorems.

For the converse part, we first note that

$$\begin{aligned} I(M;Y^{(n)}(\Delta_n)) &= \sum_{i=1}^{n} H(Y^{(n)}(t_{n,i}) - Y^{(n)}(t_{n,i-1}) | Y_{t_{n,0}}^{(n),t_{n,i-2}}) - \sum_{i=1}^{n} H(B(t_{n,i}) - B(t_{n,i-1})) \\ &\leq \sum_{i=1}^{n} H(Y^{(n)}(t_{n,i}) - Y^{(n)}(t_{n,i-1})) - \sum_{i=1}^{n} H(B(t_{n,i}) - B(t_{n,i-1})). \end{aligned}$$

It then follows from the fact

$$\begin{aligned} Var(Y^{(n)}(t_{n,i}) - Y^{(n)}(t_{n,i-1})) &= \mathbb{E}[(Y^{(n)}(t_{n,i}) - Y^{(n)}(t_{n,i-1}))^2] \\ &= \mathbb{E}[\delta_n^2 (X^{(n)}(t_{n,i-1}))^2] + \mathbb{E}[(B(t_{n,i}) - B(t_{n,i-1}))^2] \\ &= \mathbb{E}[\delta_n^2 (X^{(n)}(t_{n,i-1}))^2] + \delta_n, \end{aligned}$$

that

$$I(M;Y_{\Delta_n}^{(n)}) \leq \frac{1}{2} \sum_{i=0}^{n} \log(1 + \delta_n (X^{(n)}(t_{n,i}))^2) \tag{63}$$

$$\leq \frac{1}{2} \sum_{i=0}^{n} \delta_n (X^{(n)}(t_{n,i}))^2, \tag{64}$$

which, by (62), immediately yields

$$I(M;Y_0^T) \leq \frac{1}{2} \int_0^T X^2(s) ds \leq \frac{PT}{2}, \tag{65}$$

which establishes the converse part.

For the availability part, note that if we assume all $X^{(n)}(t_{n,i-1})$ are independent of the Brownian motion B with $\mathbb{E}[X^{(n)}(t_{n,i-1})] = P$, then the inequalities in (63) and (64) will become equalities. The part then follows from a usual random coding argument with codes generated by the distribution of $X^{(n)}$ (or more precisely, a linear interpolation of $X^{(n)}$). It is clear that as n tends to infinity, the process $X^{(n)}$ behaves more and more like a white Gaussian process. This observation echoes Theorem 6.4.1 in [39], whose proof rigorously shows that an Ornstein-Uhlenbeck process that oscillates "extremely" fast will achieve the capacity of (59).

Roughly speaking, as opposed to the bandwidth approach, the above approximation approach establishes a continuous-time Gaussian feedback channel as the limit of the associated discrete-time channels (see (61)) as the time lapse between channel uses and the signal-to-noise ratio for each channel use shrink to zero proportionately (note that in the above arguments, both of them are equal to δ_n). Here, we remark that the above approach is heuristic in nature: Theorem 3.3 does require Conditions (d)-(f), which are much stronger than the power constraint (60). Nevertheless, this approach is of fundamental importance to our treatment of continuous-time Gaussian channels: as elaborated in Section 5, not only can it provide an alternative way to establish new results, more importantly, it can also give insights to or even inspirations for our rigorous treatment employing relevant tools from stochastic calculus.

5 Continuous-Time Multi-User Gaussian Channels

Extending Shannon's fundamental theorems on point-to-point communication channels to general networks with multiple sources and destinations, network information theory aims to establish the fundamental limits on information flows in networks and the optimal coding schemes that achieve these limits. The vast majority of researches on network information theory to date have been focusing on networks in discrete time. In a way, this phenomenon can find its source from Shannon's original treatment of continuous-time point-to-point channels, where such channels were examined through their associated discrete-time versions. This insightful viewpoint has exerted major influences on the bulk of the related literature on continuous-time Gaussian channels, oftentimes prompting a model shift from the continuous-time setting to the discrete-time one right from the beginning of a research attempt.

The primary focus of this section is the capacity regions of families of continuous-time multi-user one-hop white Gaussian channels, including continuous-time multi-user white Gaussian multiple access channels (MACs), interference channels (ICs) and broadcast channels (BCs). We also examine how feedback affects the capacity regions of the above-mentioned channels. To deliver the rigourous proofs of our results, we will directly work within the continuous-time setting, employing tools from stochastic calculus; on the other hand though, we also explain the results using the connections between discrete-time and continuous-time channels, which has been established in our sampling and approximating theorems. Indeed, many results can be translated from the discrete-time setting to the continuous-time setting, such as that feedback increases the capacity region of Gaussian ICs and BCs, and that feedback does not increase the capacity region of physically degraded BCs. Nevertheless, there is a seeming "exception": as opposed to discrete-time Gaussian MACs, feedback does not increase the capacity region some of continuous-time Gaussian MACs, which, somewhat surprisingly, can also be explained by the above-mentioned connections.

Below, we summarize the results in this section. To put our results into a relevant context, we will first list some related results in discrete time.

Gaussian MACs. When there is no feedback, the capacity region of a discrete-time memoryless MAC is relatively better understood: a single-letter characterization has been established by Ahlswede [1] and the capacity region of a Gaussian MAC was explicitly derived in Wyner [87] and Cover [13]. On the other hand, the capacity region of MACs with feedback still demands more complete understanding, despite several decades of great effort by many authors: Cover and Leung [14] derived an achievable region for a memoryless MAC with feedback. In [84], Willems showed that Cover and Leung's region is optimal for a class of memoryless MACs with feedback where one of the inputs is a deterministic function of the output and the other input. More recently, Bross and Lapidoth [10] improved Cover and Leung's region, and Wu *et al.* [85] extended Cover and Leung's region for the case where non-causal state information is available at both senders. An interesting result has been obtained by Ozarow [64], who derived the capacity region of a memoryless Gaussian MAC with two users via a modification of the Schalkwijk-Kailath scheme [72]; moreover, Ozarow's result showed that in general, the capacity region for a discrete memoryless MAC is increased by feedback. The capacity region of more general MACs has also been considered; see, e.g., [53, 58, 60, 78, 11, 48, 49, 25, 26, 66] and references therein. Unfortunately, none of the above-mentioned work gives an explicit characterization of the capacity region of a generic multiple access channel with feedback, which is widely believed to be highly intractable.

In Section 5.1, inspired by our sampling theorems, we employ a stochastic calculus approach to derive the capacity region of a continuous-time white Gaussian MAC with m senders and with/without feedback. It turns out that for such a channel, the feedback does not increase the capacity region, which, at first sight, may see at odds with the aforementioned Ozarow's result and the conclusion of our sampling/approximating theorems. This, however, can be roughly explained by the well-known fact that "$>$" may become "$=$" when taking the limit (indeed, $a_n > b_n$ does not necessarily imply $\lim_{n\to\infty} a_n > \lim_{n\to\infty} b_n$). ■

Gaussian ICs. The capacity regions of discrete-time Gaussian ICs are largely unknown except for certain special scenarios: The capacity region of Gaussian ICs with strong interference has been established in Sato [71], Han and Kobayashi [27]. The sum-capacity of Gaussian ICs with weak interference has been simultaneously derived in [73, 2, 61]. The half-bit theorem on the tightness of the Han-Kobayashi bound [27] was proven in [20]. The approximation of the Gaussian IC by the q-ary expansion deterministic channel was first proposed by Avestimehr, Diggavi, and Tse [6]. Outer and inner bounds on the feedback capacity region of Gaussian interference channels are established by Suh and Tse [77]. Note that all the above-mentioned work deal with ICs with two pairs of senders and receivers. For more than two user pairs, special classes of Gaussian ICs have been examined using the scheme of interference alignment; see an extensive list of references in [19].

In Section 5.2, using a similar approach that we developed for continuous-time Gaussian MACs, we derive the capacity region of a continuous-time white Gaussian IC with m pairs of senders and receivers and without feedback. And we also use a translated version of the argument in [77] and our approximating theorem to show that feedback does increase the capacity region of certain continuous-time white Gaussian IC. ■

Gaussian BCs. The capacity regions of discrete-time Gaussian BCs without feedback are well known [12, 8]. And it has been shown by El Gamal [18] that feedback cannot increase the capacity region of a physically degraded Gaussian BC. On the other hand, it was shown by Ozarow and Leung [65] that feedback can increase the capacity of stochastically degraded Gaussian BCs, whose capacity regions are far less understood.

In Section 5.3, using a continuous-time version of entropy power inequality, we derive the capacity region of a continuous-time BC with m receivers and without feedback. And inspired our sampling theorems, we use a modified argument in [18] to show that feedback does not increase the capacity region of a physically degraded continuous-time Gaussian BC, and on the other hand, a translated version of the argument in [65] to show that feedback does increase the capacity region of certain continuous-time Gaussian BC. ■

5.1 Gaussian MACs

Consider a continuous-time white Gaussian MAC with m users, which can be characterized by

$$Y(t) = \int_0^t X_1(s, M_1, Y_0^s)ds + \int_0^t X_2(s, M_2, Y_0^s)ds + \cdots + \int_0^t X_m(s, M_m, Y_0^s)ds + B(t), \quad t \geq 0, \tag{66}$$

where X_i is the channel input from sender i, which depends on M_i, the message sent from sender i, which is independent of all messages from other senders, and possibly on the feedback Y_0^s, the channel output up to time s. Note that, with the presence of feedback, the existence and uniqueness of Y is in fact a tricky mathematical problem, however, we will simply assume that all the inputs X_i are appropriately chosen such that Y uniquely exists.

For $T, R_1, \dots, R_m, P_1, \dots, P_m > 0$, a $(T, (e^{TR_1}, \dots, e^{TR_m}), (P_1, \dots, P_m))$-code for the MAC (66) consists of m sets of integers $\mathcal{M}_i = \{1, 2, \dots, e^{TR_i}\}$, the *message alphabet* for user i, $i = 1, 2, \dots, m$, and m *encoding functions*, $X_i : \mathcal{M}_i \to C[0, T]$, which satisfy the following power constraint: for any $i = 1, 2, \dots, m$,

$$\frac{1}{T} \int_0^T X_i^2(s, M_i, Y_0^s)ds \leq P_i, \tag{67}$$

and a *decoding function*,

$$g : C[0, T] \to \mathcal{M}_1 \times \mathcal{M}_2 \times \cdots \times \mathcal{M}_m.$$

The average probability of error for the above code is defined as

$$P_e^{(T)} = \frac{1}{e^{T(\sum_{i=1}^m R_i)}} \sum_{(M_1, M_2, \dots, M_m) \in \mathcal{M}_1 \times \mathcal{M}_2 \times \cdots \times \mathcal{M}_m} P\{g(Y_0^T) \neq (M_1, M_2, \dots, M_m) \mid (M_1, M_2, \dots, M_m) \text{ sent}\}.$$

A rate tuple $(R_1, R_2, \dots, R_m)$ is said to be **achievable** for the MAC if there exists a sequence of $(T, (e^{TR_1}, \dots, e^{TR_m}), (P_1, \dots, P_m))$-codes with $P_e^{(T)} \to 0$ as $T \to \infty$. The **capacity region** of the MAC is the closure of the set of all the achievable $(R_1, R_2, \dots, R_m)$ rate tuples.

The following theorem gives an explicit characterization of the capacity region.

Theorem 5.1. *Whether there is feedback or not, the capacity region of the continuous-time white Gaussian MAC (66) is*

$$\{(R_1, R_2, \dots, R_m) \in \mathbb{R}_+^m : R_i \leq P_i/2, \quad i = 1, 2, \dots, m\}.$$

In the following, we will give the proof of Theorem 5.1. For notational convenience only, we will assume $m = 2$, the case with a generic m being completely parallel. We will first need the following lemma.

Lemma 5.2. *For any $\epsilon > 0$, there exist two independent Ornstein-Uhlenbeck processes $\{X_i(s) : s \geq 0\}$, $i = 1, 2$, satisfying the following power constraint:*

$$\text{for } i = 1, 2, \text{ there exists } P_i > 0 \text{ such that for all } t > 0, \ \frac{1}{t} \int_0^t E^2[X_i(s)]ds = P_i, \tag{68}$$

such that for all T,

$$|I_T(X_1, X_2; Y)/T - (P_1 + P_2)/2| \leq \epsilon, \tag{69}$$

and

$$|I_T(X_1; Y|X_2)/T - P_1/2| \leq \epsilon, \quad |I_T(X_2; Y|X_1)/T - P_2/2| \leq \epsilon, \tag{70}$$

moreover,

$$|I_T(X_1; Y)/T - P_1/2| \leq \epsilon, \quad |I_T(X_2; Y)/T - P_2/2| \leq \epsilon, \tag{71}$$

where

$$Y(t) = \int_0^t X_1(s)ds + \int_0^t X_2(s)ds + B(t), \quad t \geq 0. \tag{72}$$

Here (and often in the remainder of the paper) the subscript T means that the (conditional) mutual information is computed over the time period $[0, T]$.

Proof. For $a > 0$, consider the following two independent Ornstein-Uhlenbeck processes $X_i(t)$, $i = 1, 2$, given by

$$X_i(t) = \sqrt{2aP_i} \int_{-\infty}^t e^{-a(t-s)} dB_i(s),$$

where B_i, $i = 1, 2$, are independent standard Brownian motions. Obviously, for X_i defined as above, (68) is satisfied. A parallel version of the proof of Theorem 6.2.1 of [39] yields that

$$I_T(X_1, X_2; Y) = I_T(X_1 + X_2; Y) = \frac{1}{2} \int_0^T E[(X_1(t) + X_2(t) - E[X_1(t) + X_2(t)|Y_0^t])^2]dt.$$

It then follows from Theorem 6.4.1 in [39] (applied to the Ornstein-Uhlenbeck process $X_1(t) + X_2(t)$) that as $a \to \infty$,

$$I_T(X_1, X_2; Y)/T = I_T(X_1 + X_2; Y)/T \to (P_1 + P_2)/2,$$

uniformly in T, which establishes (69).

For $i = 1, 2$, define

$$\tilde{Y}_i(t) = \int_0^t X_i(s)ds + B(t), \quad t > 0.$$

As in the proof of Theorem 6.4.1 in [39], we deduce that for $i = 1, 2$, $I_T(X_i; \tilde{Y}_i)/T$ tend to $P_i/2$ uniformly in T. Now, since X_1 and X_2 are independent, we have for any fixed T,

$$I_T(X_1; Y|X_2) = I_T(X_1; \tilde{Y}_1|X_2) = I_T(X_1; \tilde{Y}_1),$$

and

$$I_T(X_2; Y|X_1) = I_T(X_2; \tilde{Y}_2|X_1) = I_T(X_2; \tilde{Y}_2),$$

which immediately implies (70).

Now, by the chain rule of mutual information,

$$I_T(X_1, X_2; Y) = I_T(X_1; Y) + I_T(X_2; Y|X_1) = I_T(X_2; Y) + I_T(X_1; Y|X_2),$$

which, together with (69) and (70), implies (71).

□

Remark 5.3. With X_i, $i = 1, 2$, regarded as channel inputs, (72) can be reinterpreted as a white Gaussian MAC. For $i \neq j$, $I(X_i; Y)$, the reliable transmission rate of X_i when X_j is not known can be arbitrarily close to $I(X_i; Y|X_j)$, the reliable transmission rate of X_i when X_j is known. In other words, for white Gaussian MACs, knowledge about other user's inputs will not help to achieve faster transmission rate, and therefore, they can be simply treated as noises. An more intuitive explanation of this result is as follows: for the Ornstein-Uhlenbeck process X_i as specified in the proof, its power spectral density can be computed as

$$f_i(\lambda) = \frac{2aP_i}{2\pi(\lambda^2 + a^2)},$$

which is "negligible" compared to that of the white Gaussian noise (which is the constant 1) as a tends to infinity. Lemma 5.2 is a key ingredient for deriving the capacity regions of white Gaussian MACs, and, as elaborated later in the paper, those of white Gaussian ICs and BCs as well.

We also need some result on the information stability of continuous-time Gaussian processes. Let $(U, V) = \{(U(t), V(t)), t \geq 0\}$ be a continuous Gaussian system (which means $U(t), V(t)$ are pairwise Gaussian stochastic processes). Define

$$\varphi^{(T)}(u, v) = \frac{d\mu_{UV}^{(T)}}{d\mu_U^{(T)} \times \mu_V^{(T)}}(u, v), \qquad (u, v) \in C[0, T] \times C[0, T],$$

where $\mu_U^{(T)}$, $\mu_V^{(T)}$ and $\mu_{UV}^{(T)}$ denote the probability distributions of U_0^T, V_0^T and their joint distribution, respectively. For any $\varepsilon > 0$, we denote by $\mathcal{T}_\varepsilon^{(T)}$ the ε-typical set:

$$\mathcal{T}_\varepsilon^{(T)} = \left\{(u, v) \in C[0, T] \times C[0, T]; \frac{1}{T}|\log \varphi^{(T)}(u, v) - I_T(U, V)| \leq \varepsilon\right\}.$$

The pair (U, V) is said to be *information stable* [67] if for any $\varepsilon > 0$,

$$\lim_{T \to \infty} \mu_{UV}^{(T)}(\mathcal{T}_\varepsilon) = 1.$$

The following theorem is a rephrased version of Theorem 6.6.2. in [39].

Lemma 5.4. *The Gaussian system (U, V) is information stable provided that*

$$\lim_{T \to \infty} \frac{I_T(U; V)}{T^2} = 0.$$

Lemma 5.4 will be used in the proof of Theorem 5.1 to establish, roughly speaking, that almost all sequences are jointly typical.

We are now ready for the proof of Theorem 5.1

Proof of Theorem 5.1. **The converse part.** In this part, we will show that for any sequence of $(T, (e^{TR_1}, e^{TR_2}), (P_1, P_2))$-codes with $P_e^{(T)} \to 0$ as $T \to \infty$, the rate pair (R_1, R_2) will have to satisfy

$$R_1 \leq P_1/2, \qquad R_2 \leq P_2/2.$$

Fix T and consider the above-mentioned $(T, (e^{TR_1}, e^{TR_2}), (P_1, P_2))$-code. By the code construction, it is possible to estimate the messages (M_1, M_2) from the channel output Y_0^T with a low probability of error. Hence, the conditional entropy of (M_1, M_2) given Y_0^T must be small; more precisely, by Fano's inequality,

$$H(M_1, M_2|Y_0^T) \leq T(R_1 + R_2)P_e^{(T)} + H(P_e^{(T)}) = T\varepsilon_T,$$

where $\varepsilon_T \to 0$ as $T \to \infty$. Then, we have

$$H(M_1|Y^T) \leq H(M_1, M_2|Y^T) \leq T\varepsilon_T, \quad H(M_2|Y^T) \leq H(M_1, M_2|Y^T) \leq T\varepsilon_T.$$

Now, we can bound the rate R_1 as follows:

$$\begin{aligned} TR_1 &= H(M_1) \\ &= I(M_1; Y_0^T) + H(M_1|Y_0^T) \\ &\leq I(M_1; Y_0^T) + T\varepsilon_T \\ &\leq H(M_1) - H(M_1|Y_0^T) + T\varepsilon_T \\ &\leq H(M_1|M_2) - H(M_1|Y_0^T, M_2) + T\varepsilon_T \\ &= I(M_1; Y_0^T|M_2) + T\varepsilon_T. \end{aligned}$$

Applying Theorem 6.2.1 in [39], we have

$$I(M_1; Y_0^T|M_2) = \frac{1}{2}E\left[\int_0^T E[(X_1 + X_2 - \hat{X}_1 - \hat{X}_2)^2|M_2]dt\right] = \frac{1}{2}\int_0^T E[(X_1+X_2-\hat{X}_1-\hat{X}_2)^2]dt,$$

where $\hat{X}_i(t) = E[X_i(t)|Y_0^T, M_2]$, $i = 1, 2$. Noticing that $X_2 = \hat{X}_2$, we then have

$$I(M_1; Y_0^T|M_2) = \frac{1}{2}\int_0^T E[(X_1 - \hat{X}_1)^2]dt],$$

which, together with (67), implies that $R_1 \leq P_1/2$. A completely parallel argument will yield that $R_2 \leq P_2/2$.

The achievability part. In this part, we will show that as long as (R_1, R_2) satisfying

$$0 \leq R_1 < P_1/2, \quad 0 \leq R_2 < P_2/2, \tag{73}$$

we can find a sequence of $(T, (e^{TR_1}, e^{TR_2}), (P_1, P_2))$-codes with $P_e^{(T)} \to 0$ as $T \to \infty$. The argument consists of several steps as follows.

Codebook generation: For a fixed $T > 0$ and $\varepsilon > 0$, assume that X_1 and X_2 are independent Ornstein-Uhlenbeck processes over $[0, T]$ with respective variances $P_1 - \varepsilon$ and $P_2 - \varepsilon$, and that (R_1, R_2) satisfying (73). Generate e^{TR_1} independent codewords $X_{1,i}$, $i \in \{1, 2, \ldots, e^{TR_1}\}$, of length T, according to the distribution of X_1. Similarly, generate e^{TR_2} independent codewords $X_{2,j}$, $j \in \{1, 2, \ldots, e^{TR_2}\}$, of length T, according to the distribution of X_2. These codewords (which may not satisfy the power constraint in (67)) form the codebook, which is revealed to the senders and the receiver.

Encoding: To send message $i \in \mathcal{M}_1$, sender 1 sends the codeword $X_{1,i}$. Similarly, to send $j \in \mathcal{M}_2$, sender 2 sends $X_{2,j}$.

Decoding: For any fixed $\varepsilon > 0$, let $\mathcal{T}_\varepsilon^{(T)}$ denote the set of *jointly typical* (x_1, x_2, y) sequences, which is defined as follows:

$$\mathcal{T}_\varepsilon^{(T)} = \{(x_1, x_2, y) \in C[0,T] \times C[0,T] \times C[0,T] : |\log \varphi_1(x_1, x_2, y) - I_T(X_1, X_2; Y)| \leq T\varepsilon,$$

$$|\log \varphi_2(x_1, x_2, y) - I_T(X_1; X_2, Y)| \leq T\varepsilon, |\log \varphi_3(x_1, x_2, y) - I_T(X_2; X_1, Y)| \leq T\varepsilon\},$$

where

$$\varphi_1(x_1, x_2, y) = \frac{d\mu_{X_1 X_2 Y}}{d\mu_{X_1 X_2} \times \mu_Y}(x_1, x_2, y),$$

$$\varphi_2(x_1, x_2, y) = \frac{d\mu_{X_1 X_2 Y}}{d\mu_{X_1} \times \mu_{X_2 Y}}(x_1, x_2, y),$$

$$\varphi_3(x_1, x_2, y) = \frac{d\mu_{X_1 X_2 Y}}{d\mu_{X_2} \times \mu_{X_1 Y}}(x_1, x_2, y).$$

Here we remark that it is easy to check that the above Randon-Nykodym derivatives are all well-defined; see, e.g., Theorem 7.7 of [55] for sufficient conditions for their existence. Based on the received output $y \in C[0,T]$, the receiver chooses the pair (i,j) such that

$$(x_{1,i}, x_{2,j}, y) \in \mathcal{T}_\varepsilon^{(T)},$$

if such a pair (i,j) exists and is unique; otherwise, an error is declared. Moreover, an error will be declared if the chosen codeword does not satisfy the power constraint in (67).

Analysis of the probability of error: Now, for fixed $T, \varepsilon > 0$, define

$$E_{ij} = \{(X_{1,i}, X_{2,j}, Y) \in \mathcal{T}_\varepsilon^{(T)}\}.$$

By symmetry, we assume, without loss of generality, that (1,1) was sent. Define $\pi^{(T)}$ to be the event that

$$\int_0^T (X_{1,1}(t))^2 dt > P_1 T, \quad \int_0^T (X_{2,1}(t))^2 dt > P_2 T.$$

Then, $\hat{P}_e^{(T)}$, the error probability for the above coding scheme (where codewords violating the power constraint are allowed), can be upper bounded as follows:

$$\hat{P}_e^{(T)} = P(\pi^{(T)} \cup E_{11}^c \bigcup \cup_{(i,j) \neq (1,1)} E_{ij})$$

$$\leq P(\pi^{(T)}) + P(E_{11}^c) + \sum_{i \neq 1, j=1} P(E_{i1}) + \sum_{i=1, j \neq 1} P(E_{1j}) + \sum_{i \neq 1, j \neq 1} P(E_{ij}).$$

So, for any $i, j \neq 1$, we have

$$\hat{P}_e^{(T)} \leq P(\pi^{(T)}) + P(E_{11}^c) + e^{TR_1} P(E_{i1}) + e^{TR_2} P(E_{1j}) + e^{TR_1 + TR_2} P(E_{ij})$$

Using the well-known fact that an Ornstein-Uhlenbeck process is ergodic [54, 50], we deduce that $P(\pi^{(T)}) \to 0$ as $T \to \infty$. And by Lemma 5.4 and Theorem 6.2.1 in [39], we have

$$\lim_{T \to \infty} P((X_{1,1}, X_{2,1}, Y) \in \mathcal{T}_\varepsilon^{(T)}) = 1 \text{ and thus } \lim_{T \to \infty} P(E_{11}^c) = 0.$$

Now, we have for any $i \neq 1$,

$$\begin{aligned} P(E_{i1}) &= P((X_{1,i}, X_{2,1}, Y) \in \mathcal{T}_\varepsilon^{(T)}) \\ &= \int_{(x_1,x_2,y)\in\mathcal{T}_\varepsilon^{(T)}} d\mu_{X_1}(x_1) d\mu_{X_2Y}(x_2,y) \\ &= \int_{\mathcal{T}_\varepsilon^{(T)}} \frac{1}{\varphi_1(x_1,x_2,y)} d\mu_{X_1X_2Y}(x_1,x_2,y) \\ &\leq \int_{\mathcal{T}_\varepsilon^{(T)}} e^{-I_T(X_1;X_2,Y)+\varepsilon T} d\mu_{X_1X_2Y}(x_1,x_2,y) \\ &= e^{-I_T(X_1;Y|X_2)+\varepsilon T}, \end{aligned}$$

where we have used the independence of X_1 and X_2, and the consequent fact that

$$I_T(X_1; X_2, Y) = I_T(X_1; X_2) + I_T(X_1; Y|X_2) = I_T(X_1; Y|X_2).$$

Similarly, we have, for $j \neq 1$,

$$P(E_{1j}) \leq e^{-I_T(X_2;Y|X_1)+\varepsilon T},$$

and for $i, j \neq 1$,

$$P(E_{ij}) \leq e^{-I_T(X_1,X_2;Y)+\varepsilon T}.$$

It then follows that

$$\hat{P}_e^{(T)} \leq P(\pi^{(T)}) + P(E_{11}^c) + e^{TR_1+\varepsilon T-I_T(X_1;Y|X_2)} + e^{TR_2+\varepsilon T-I_T(X_2;Y|X_1)} + e^{TR_1+TR_2+\varepsilon T-I_T(X_1,X_2;Y)}.$$

By Lemma 5.2, one can choose independent OU processes X_1, X_2 such that $I_T(X_1; Y|X_2)/T \to (P_1-\epsilon)/2$, $I_T(X_2; Y|X_1)/T \to (P_2-\epsilon)/2$ and $I_T(X_1, X_2; Y)/T \to (P_1+P_2-2\epsilon)$ uniformly in T. This implies that with ϵ chosen sufficiently small, we have $\hat{P}_e^{(T)} \to 0$, as $T \to \infty$. In other words, there exists a sequence of good codes (which may not satisfy the power constraint) with low average error probability. Now, from each of the above codes, we delete the worse half of the codewords (any codeword violating the power constraint will be deleted since it must have error probability 1). Then, with only slightly decreased transmission rate, the remaining codewords will satisfy the power constraint and will have small maximum error probability (and thus small average error probability $P_e^{(T)}$), which implies that the rate pair (R_1, R_2) is achievable. □

Remark 5.5. The achievability part can be proven alternatively, which will be roughly described as follows: for arbitrarily small $\epsilon > 0$, by Lemma 5.2, one can choose independent Ornstein-Uhlenbeck processes X_i with respective variances $P_i - \epsilon$, $i = 1, 2$, such that $I_T(X_i; Y)/T$ approaches $(P_i-\epsilon)/2$. Then, a parallel random coding argument with X_j, $j \neq i$, being treated as noise at receiver i shows that the rate pair $((P_1 - \epsilon)/2, (P_2 - \epsilon)/2)$ can be approached, which yields the achievability part.

Remark 5.6. When there is no feedback, Theorem 5.1 can be heuristically explained using the bandwidth approach as in (2)-(4) (this heuristical approach in this example should be well-known; see, e.g., Exercise 15.26 in [15]): Consider the following continuous-time white Gaussian multiple access channel with two senders:

$$Y(t) = X_1(t) + X_2(t) + Z(t), \quad t \in \mathbb{R}, \tag{74}$$

where X_i, $i = 1, 2$, is the input from the i-th user with average power limit P_i. Similarly as before, consider its associated discrete-time version corresponding to bandwidth limit ω:

$$Y_n = X_{1,n}^{(\omega)} + X_{2,n}^{(\omega)} + Z_n^{(\omega)}, \quad n \in \mathbb{Z}.$$

Then, it is well known [19] that the outer bound on the capacity region can be computed as

$$\left\{(R_1, R_2) \in \mathbb{R}_+^2 : R_1 \leq W \log\left(1 + \frac{P_1}{2\omega}\right), R_2 \leq W \log\left(1 + \frac{P_2}{2\omega}\right)\right\},$$

and the inner bound as

$$\left\{(R_1, R_2) \in \mathbb{R}_+^2 : R_1 \leq \omega \log\left(1 + \frac{P_1}{2\omega}\right), R_2 \leq \omega \log\left(1 + \frac{P_2}{2\omega}\right), R_1 + R_2 \leq \omega \log\left(1 + \frac{P_1 + P_2}{2\omega}\right)\right\}.$$

(Here, it is known [87, 13] that the outer bound can be tightened to coincide with the inner bound, which, however, is not needed for this example.) It is easy to verify that the two bounds also collapse into the same region as ω tends to infinity:

$$\left\{(R_1, R_2) \in \mathbb{R}_+^2 : R_1 \leq P_1/2, R_2 \leq P_2/2\right\},$$

which is "expected" to be the capacity region of (74). And a similar argument holds for more than two senders as well through a parallel extension.

Remark 5.7. Following the spirits of our approximating theorems, we will use the approximation approach as in Section 4 to give an alternative way to explain Theorem 5.1 assuming the presence of the feedback. For simplicity only, we consider the following continuous-time Gaussian MAC with two senders:

$$Y(t) = \int_0^t X_1(s, M_1, Y_0^s)ds + \int_0^t X_2(s, M_2, Y_0^s)ds + B(t), \quad t \geq 0, \tag{75}$$

with the power constraints: there exist $P_1, P_2 > 0$ such that for all T,

$$\int_0^T X_1^2(s, M_1, Y_0^s)ds \leq P_1 T, \quad \int_0^T X_2^2(s, M_2, Y_0^s)ds \leq P_2 T. \tag{76}$$

Apply the Euler-Maruyama approximation to the above channel over the time window $[0, T]$ with respect to the equidistant Δ_n with $\delta_n = T/n$, we obtain

$$Y^{(n)}(t_{n,i+1}) = Y^{(n)}(t_{n,i}) + \int_{t_{n,i}}^{t_{n,i+1}} X_1(s, M_1, Y_0^{(n),t_{n,i}})ds + \int_{t_{n,i}}^{t_{n,i+1}} X_2(s, M_2, Y_0^{(n),t_{n,i}})ds + B(t_{n,i+1}) - B(t_{n,i}).$$

Now, straightforward computations and a usual concavity argument then yields that for large n,

$$\begin{aligned}
I(M_1; Y^{(n)}(\Delta_n)|M_2) &= H(Y^{(n)}(\Delta_n)|M_2) - H(Y^{(n)}(\Delta_n)|M_1, M_2) \\
&= \sum_{i=1}^{n} H(Y^{(n)}(t_{n,i})|Y_{t_{n,0}}^{(n),t_{n,i-1}}, M_2) - \sum_{i=1}^{n} H(Y^{(n)}(t_{n,i})|Y_{t_{n,0}}^{(n),t_{n,i-1}}, M_1, M_2) \\
&\le \sum_{i=1}^{n} \frac{1}{2} \log \left(\mathbb{E} \left(\int_{t_{n,i-1}}^{t_{n,i}} X_1(s, M_1, Y_0^{(n),t_{n,i-1}}) ds \right)^2 + \delta_n \right) - \frac{1}{2} \log(\delta_n) \\
&\le \sum_{i=1}^{n} \frac{1}{2} \log \left(\left(\int_{t_{n,i-1}}^{t_{n,i}} \mathbb{E} X_1(s, M_1, Y_0^{(n),t_{n,i-1}})^2 ds \right) \delta_n + \delta_n \right) - \frac{1}{2} \log(\delta_n) \\
&= \sum_{i=1}^{n} \frac{1}{2} \log \left(\int_{t_{n,i-1}}^{t_{n,i}} \mathbb{E} X_1(s, M_1, Y_0^{(n),t_{n,i-1}})^2 ds + 1 \right) \\
&\le \sum_{i=1}^{n} \sum_{i=1}^{n} \frac{1}{2} \int_{t_{n,i-1}}^{t_{n,i}} \mathbb{E} X_1(s, M_1, Y_0^{(n),t_{n,i-1}})^2 ds \\
&\approx \sum_{i=1}^{n} \sum_{i=1}^{n} \frac{1}{2} \int_{t_{n,i-1}}^{t_{n,i}} \mathbb{E} X_1(s, M_1, Y_0^{s})^2 ds \\
&\le \frac{P_1 T}{2}.
\end{aligned}$$

A completely parallel argument will yield that

$$I(M_2; Y^{(n)}(\Delta_n)|M_1) \le \frac{P_2 T}{2}.$$

It then follows from Theorem 3.3 the region below give an outer bound of the capacity region:

$$\{(R_1, R_2) : 0 \le R_1 \le P_1/2, \quad 0 \le R_2 \le P_2/2\}. \tag{77}$$

To see that this outer bound can be achieved, set $X_1(s), X_2(s)$, $t_{n,i} \le s \le t_{n,i+1}$, in (75) to be independent Gaussian random variables with variances P_1, P_2, respectively. Then, one verifies that for large n,

$$\begin{aligned}
I(M_1; Y(\Delta_n)) &= H(Y(\Delta_n)) - H(Y(\Delta_n)|M_1) \\
&= \sum_{i=1}^{n} H(Y(t_{n,i})) - \sum_{i=1}^{n} H(Y(t_{n,i})|M_1) \\
&= \sum_{i=1}^{n} \frac{1}{2} \log(P_1 \delta_n^2 + P_2 \delta_n^2 + \delta_n) - \frac{1}{2} \log(P_2 \delta_n^2 + \delta_n) \\
&= \sum_{i=1}^{n} \frac{1}{2} \log \left(1 + \frac{P_1 \delta_n^2}{P_2 \delta_n^2 + \delta_n} \right) \\
&\approx P_1 T/2.
\end{aligned}$$

Similarly one can prove that

$$I(M_2; Y_{\Delta_n}) \approx P_2 T/2.$$

It then follows that the outer bound in (77) can be achieved.

Here we remark that similarly as in Section 4, the constructed processes X_1 and X_2 behave like "fast-oscillating" Ornstein-Uhlenbeck processes, which echoes Remark 5.3 and gives another explanation to Lemma 5.2, a key lemma in our rigourous treatment.

5.2 Gaussian ICs

Consider the following continuous-time white Gaussian interference channel having no feedback and with m pairs of senders and receivers: for $i = 1, 2, \ldots, m$,

$$Y_i(t) = a_{i1} \int_0^t X_1(s, M_1) ds + a_{i2} \int_0^t X_2(s, M_2) ds + \cdots + a_{im} \int_0^t X_m(s, M_m) ds + B_i(t), \quad t \geq 0, \tag{78}$$

where X_i is the channel input from sender i, which depends on M_i, the message sent from sender i, which is independent of all messages from other senders, and $a_{ij} \in \mathbb{R}$, $i, j = 1, 2, \ldots, m$, is the channel gain from sender j to receiver i, all $B_i(t)$ are (possibly correlated) standard Brownian motions.

For $T, R_1, \ldots, R_m, P_1, \ldots, P_m > 0$, a $(T, (e^{TR_1}, \ldots, e^{TR_m}), (P_1, \ldots, P_m))$-code for the IC (78) consists of m sets of integers $\mathcal{M}_i = \{1, 2, \ldots, e^{TR_i}\}$, the *message alphabet* for user i, $i = 1, 2, \ldots, m$, and m *encoding functions*, $X_i : \mathcal{M}_i \to C[0, T]$ satisfying the following power constraint: for any $i = 1, 2, \ldots, m$,

$$\frac{1}{T} \int_0^T X_i^2(s, M_i) ds \leq P_i, \tag{79}$$

and m *decoding functions*, $g_i : C[0, T] \to \mathcal{M}_i$, $i = 1, 2, \ldots, m$.

The average probability of error for the $(T, (e^{TR_1}, \ldots, e^{TR_m}), (P_1, \ldots, P_m))$-code is defined as

$$P_e^{(T)} = \frac{1}{e^{T(\sum_{i=1}^m R_i)}} \sum_{(M_1, M_2, \ldots, M_m) \in \mathcal{M}_1 \times \mathcal{M}_2 \times \cdots \times \mathcal{M}_m} P\{g_i(Y_{i,0}^T) \neq M_i, i = 1, 2, \ldots, m \mid (M_1, M_2, \ldots, M_m) \text{ sent}\}.$$

A rate tuple $(R_1, R_2, \ldots, R_m)$ is said to be **achievable** for the IC if there exists a sequence of $(T, (e^{TR_1}, \ldots, e^{TR_m}), (P_1, \ldots, P_m))$-codes with $P_e^{(T)} \to 0$ as $T \to \infty$. The **capacity region** of the IC is the closure of the set of all the achievable $(R_1, R_2, \ldots, R_m)$ rate tuples.

The following theorem explicitly characterizes the capacity region of the above IC:

Theorem 5.8. *The capacity region of the continuous-time white Gaussian IC (78) is*

$$\{(R_1, R_2, \ldots, R_m) \in \mathbb{R}_+^m : R_i \leq a_{ii}^2 P_i/2, \quad i = 1, 2, \ldots, m\}.$$

Proof. For notational convenience only, we only prove the case when $n = 2$; the case when n is generic is similar.

The converse part. In this part, we will show that for any sequence of $(T, (e^{TR_1}, e^{TR_2}), (P_1, P_2))$ codes with $P_e^{(T)} \to 0$, the rate pair (R_1, R_2) will have to satisfy

$$R_1 \leq a_{11}^2 P_1/2, \quad R_2 \leq a_{22}^2 P_2/2. \tag{80}$$

Fix T and consider the above-mentioned $(T, (e^{TR_1}, e^{TR_2}), (P_1, P_2))$ code. By the code construction, for $i = 1, 2$, it is possible to estimate the messages M_i from the channel output $Y_{i,0}^T$ with an arbitrarily low probability of error. Hence, by Fano's inequality, for $i = 1, 2$,

$$H(M_i | Y_{i,0}^T) = T\varepsilon_{i,T},$$

where $\varepsilon_{i,T} \to 0$ as $T \to \infty$. We then have

$$TR_1 = H(M_1) = H(M_1|M_2) = I(M_1; Y_1|M_2) + H(M_1|M_2, Y_1) \leq I(M_1; Y_1|M_2) + T\varepsilon_{1,T},$$

As in the proof of Theorem 5.1, we have

$$I(M_1; Y_{1,0}^T | M_2) = \frac{a_{11}^2}{2} \int_0^T E[(X_1(s) - E[X_1(s)|M_2, Y_{1,0}^s])^2] ds.$$

It then follows that

$$TR_1 \leq \frac{a_{11}^2}{2} \int_0^T E[(X_1(s) - E[X_1(s)|M_2, Y_{1,0}^s])^2] ds + T\varepsilon_{1,T},$$

which implies that $R_1 \leq a_{11}^2 P_1/2$. With a parallel argument, one can derive that $R_2 \leq a_{22}^2 P_2/2$. The proof for the converse part is then complete.

The achievability part. We only sketch the proof of this part. For arbitrarily small $\epsilon > 0$, by Lemma 5.2, one can choose independent Ornstein-Uhlenbeck processes X_i with respective variances $P_i - \epsilon$, $i = 1, 2$, such that $I_T(X_i; Y)/T$ approaches $a_{ii}^2 (P_i - \epsilon)/2$. Then, a parallel random coding argument as in the proof of Theorem 5.1 with X_j, $j \neq i$, being treated as noise at receiver i shows that the rate pair $(a_{11}^2(P_1 - \epsilon)/2, a_{22}^2(P_2 - \epsilon)/2)$ can be approached, which yields the achievability part. □

Remark 5.9. Theorem 5.8 can be heuristically derived using a similar argument employing Theorem 3.3 as in Remark 5.7.

The proof of the following theorem is essentially a translated version of the argument in [77] coupled with the approximation approach as in Section 4, and so we only provide a sketch of the proof.

Theorem 5.10. *Feedback strictly increases the capacity region of certain continuous-time Gaussian interference channel.*

Proof. Consider the following symmetric continuous-time Gaussian interference channel with two pairs of senders and receivers:

$$Y_1(t) = \sqrt{snr} \int_0^t X_1(s) ds + \sqrt{inr} \int_0^t X_2(s) ds + B_1(t),$$

$$Y_2(t) = \sqrt{inr} \int_0^t X_1(s)ds + \sqrt{snr} \int_0^t X_2(s)ds + B_2(t),$$

where snr, inr denote the signal-to-noise, interference-to-noise ratios, respectively, $B_1(t), B_2(t)$ are independent standard Brownian motions, and the average power of X_1, X_2 are assumed to be 1.

Following [77], we consider the following coding scheme over two stages, each of length T_0. In the first stage, transmitters 1 and 2 send codewords $X_{1,0}^{T_0}$ and $X_{2,0}^{T_0}$ with rates R_1 and R_2, respectively. In the second stage, using feedback, transmitters 1 and 2 decode $X_{2,0}^{T_0}$ and $X_{1,0}^{T_0}$, respectively. This can be decoded if

$$R_1, R_2 \leq \frac{inr}{2}.$$

Then, transmitters 1 and 2 send $X_{1,T_0}^{2T_0}$ and $X_{2,T_0}^{2T_0}$, respectively such that for any $0 \leq t \leq T_0$,

$$X_1(T_0 + t) = X_2(t), \quad X_2(T_0 + t) = -X_1(t).$$

Then during the two stages, receiver 1 receives

$$Y_1(t) = \sqrt{snr} \int_0^t X_1(s)ds + \sqrt{inr} \int_0^t X_2(s)ds + B_1(t), \quad 0 \leq t \leq T_0,$$

and

$$Y_1(T_0 + t) = \sqrt{snr} \int_0^{T_0+t} X_1(s)ds + \sqrt{inr} \int_0^{T_0+t} X_2(s)ds + B_1(T_0 + t), \quad 0 \leq t \leq T_0,$$

which immediately gives rise to

$$\begin{aligned} Y_1(T_0 + t) - Y_1(T_0) &= \sqrt{snr} \int_{T_0}^{T_0+t} X_1(s)ds + \sqrt{inr} \int_{T_0}^{T_0+t} X_2(s)ds + B_1(T_0 + t) - B_1(T_0) \\ &= \sqrt{snr} \int_0^t X_2(s)ds - \sqrt{inr} \int_0^t X_1(s)ds + B_1(T_0 + t) - B_1(T_0). \end{aligned}$$

We then have that for any $0 \leq t \leq T_0$,

$$\sqrt{snr}Y_1(t) - \sqrt{inr}(Y_1(T_0+t) - Y_1(T_0)) = (snr+inr) \int_0^t X_1(s)ds + \sqrt{snr}B_1(t) - \sqrt{inr}(B_1(T_0+t) - B(T_0)),$$

which means the codeword $X_{1,0}^{T_0}$ can be decoded at the second stage if

$$R_1 \leq \frac{snr + inr}{2}.$$

A completely parallel argument yields that the codeword $X_{2,0}^{T_0}$ can be decoded at the second stages if

$$R_2 \leq \frac{snr + inr}{2}.$$

All in all, after the two stages, the two codewords $X_{1,0}^{T_0}$ and $X_{2,0}^{T_0}$ can be decoded as long as

$$R_1, R_2 \leq \frac{inr}{2};$$

in other words, coding rate $(\frac{inr}{2}, \frac{inr}{2})$ is achievable, which, if assuming $inr > snr$, will imply that feedback strictly increases the capacity region. □

5.3 Gaussian BCs

In this section, we consider a continuous-time white Gaussian BC with m receivers, which is characterized by: for $i = 1, 2, \ldots, m$,

$$Y_i(t) = \sqrt{snr_i} \int_0^t X(s, M_1, M_2, \ldots, M_m) ds + B_i(t), \quad t \geq 0, \tag{81}$$

where X is the channel input, which depends on M_i, the message sent from sender i, which is uniformly distributed over a finite alphabet $\mathcal{M}_i$ and independent of all messages from other senders, snr_i is the signal-to-noise ratio in the channel for user i, $B_i(t)$ are (possibly correlated) standard Brownian motions.

For $T, R_1, R_2, \ldots, R_m, P > 0$, a $(T, (e^{TR_1}, \ldots, e^{TR_m}), P)$-code for the BC (81) consists of m set of integers $\mathcal{M}_i = \{1, 2, \ldots, e^{TR_i}\}$, the *message set* for receiver i, $i = 1, 2, \ldots, m$, and an *encoding function*, $X : \mathcal{M}_1 \times \mathcal{M}_2 \times \cdots \times \mathcal{M}_m \to C[0, T]$, which satisfies the following power constraint:

$$\frac{1}{T} \int_0^T X^2(s, M_1, M_2, \ldots, M_m) ds \leq P, \tag{82}$$

and m *decoding functions*, $g_i : C[0, T] \to \mathcal{M}_i$, $i = 1, 2, \ldots, m$.

The average probability of error for the $(T, (e^{TR_1}, e^{TR_2}, \ldots, e^{TR_m}), P)$-code is defined as

$$P_e^{(T)} = \frac{1}{e^{T(\sum_{i=1}^m R_i)}} \sum_{(M_1, M_2, \ldots, M_m) \in \mathcal{M}_1 \times \mathcal{M}_2 \times \cdots \times \mathcal{M}_m} P\{g_i(Y_0^T) \neq M_i, i = 1, 2, \ldots, m \mid (M_1, M_2, \ldots, M_m) \text{ sent}\}.$$

A rate tuple $(R_1, R_2, \ldots, R_m)$ is said to be **achievable** for the BC if there exists a sequence of $(T, (e^{TR_1}, e^{TR_2}, \ldots, e^{TR_m}), P)$-codes with $P_e^{(T)} \to 0$ as $T \to \infty$. The **capacity region** of the BC is the closure of the set of all the achievable $(R_1, R_2, \ldots, R_m)$ rate tuples.

The following theorem explicitly characterizes the capacity region of the above BC:

Theorem 5.11. *The capacity region of the continuous-time white Gaussian BC (81) is*

$$\left\{ (R_1, R_2, \ldots, R_m) \in \mathbb{R}_+^m : \frac{R_1}{snr_1} + \frac{R_2}{snr_2} + \cdots + \frac{R_m}{snr_m} \leq \frac{P}{2} \right\}.$$

One of the important tools that plays a key role in discrete-time network information theory is the entropy power inequality [15, 19], which can be applied to compare information-theoretic quantities involving different users. The following lemma, which, despite its strikingly different form, serves the typical function of a discrete-time entropy power inequality.

Lemma 5.12. *Consider a continuous-time white Gaussian channel characterized by the following equation*

$$Y(t) = \sqrt{snr} \int_0^t X(s) ds + B(t), \quad t \geq 0,$$

where $snr \geq 0$ denotes the signal-to-noise ratio in the channel and M is the message to be transmitted through the channel. Then, for any fixed T, $I_T(M; Y)/snr$ is a monotone decreasing function of snr.

Proof. For notational convenience, in this proof, we write $I_T(M; Y)$ as $I_T(snr)$. By Theorem 6.2.1 in [39], we have

$$I_T(snr) = \frac{snr}{2} \int_0^T E[(X(s) - E[X(s)|Y_0^s])^2]ds,$$

and Theorem 6 in [24], we have (the derivative is with respect to snr)

$$I_T'(snr) = \frac{1}{2} \int_0^T E[(X(s) - E[X(s)|Y_0^T])^2]ds.$$

It then follows that

$$\begin{aligned}\left(\frac{I_T(snr)}{snr}\right)' &= \frac{1}{snr}\left(I_T'(snr) - \frac{I_T(snr)}{snr}\right) \\ &= \frac{1}{2snr}\left(\int_0^T E[(X(s) - E[X(s)|Y_0^T])^2]ds - \int_0^T E[(X(s) - E[X(s)|Y_0^s])^2]ds\right) \leq 0,\end{aligned}$$

which immediately implies the lemma. □

Proof of Theorem 5.11. For notational convenience only, we prove the case when $n = 2$, the case when n is generic being parallel.

The converse part. Without loss of generality, we assume that

$$snr_1 \geq snr_2.$$

We will show that for any sequence of $(T, (e^{TR_1}, e^{TR_2}), P)$ codes with $P_e^{(T)} \to 0$ as $T \to \infty$, the rate pair (R_1, R_2) will have to satisfy

$$\frac{R_1}{snr_1} + \frac{R_2}{snr_2} \leq \frac{P}{2}. \tag{83}$$

Fix T and consider the above-mentioned $(T, (e^{TR_1}, e^{TR_2}), P)$-code. By the code construction, for $i = 1, 2$, it is possible to estimate the messages M_i from the channel output $Y_{i,0}^T$ with an arbitrarily low probability of error. Hence, by Fano's inequality, for $i = 1, 2$,

$$H(M_i|Y_{i,0}^T) \leq TR_i P_e^{(T)} + H(P_e^{(T)}) = T\varepsilon_{i,T},$$

where $\varepsilon_{i,T} \to 0$ as $T \to \infty$. It then follows that

$$TR_1 = H(M_1) = H(M_1|M_2) \leq I(M_1; Y_{1,0}^T|M_2) + T\varepsilon_{1,T}, \tag{84}$$

$$TR_2 = H(M_2) \leq I(M_2; Y_{2,0}^T) + T\varepsilon_{2,T}. \tag{85}$$

By the chain rule of mutual information, we have

$$I(M_1, M_2; Y_{2,0}^T) = I(M_2; Y_{2,0}^T) + I(M_1; Y_{2,0}^T|M_2) \geq I(M_2; Y_{2,0}^T) + \frac{snr_2}{snr_1} I(M_1; Y_{1,0}^T|M_2), \tag{86}$$

where, for the inequality above, we have applied Lemma 5.12. Now, by Theorem 6.2.1 in [39], we have

$$I(M_1, M_2; Y_{2,0}^T) = \frac{snr_2}{2} \int_0^T E[(X(s) - E[X(s)|Y_{2,0}^s])^2] ds \leq \frac{snr_2}{2} \int_0^T E[X^2(s)] ds,$$

which, together with (84), (85), (86) and (82), immediately implies the converse part.

The achievability part. We only sketch the proof of this part. For an arbitrarily small $\epsilon > 0$, by Theorem 6.4.1 in [39], one can choose an Ornstein-Uhlenbeck processes $\tilde{X}$ with variance $P - \epsilon$, such that $I_T(\tilde{X}; Y_i)/T$ approaches $snr_i(P - \epsilon)/2$. For any $0 \leq \lambda \leq 1$, let

$$X(t) = \sqrt{\lambda} X_1(t) + \sqrt{1-\lambda} X_2(t), \quad t \geq 0,$$

where X_1 and X_2 are independent copies of $\tilde{X}$. Then, by a similar argument as in the proof of Lemm 5.2, we deduce that $I_T(X_1; Y_1)/T, I_T(X_2; Y_2)/T$ approach $snr_1 \lambda (P - \epsilon)/2$, $snr_2(1-\lambda)(P-\epsilon)/2$, respectively. Then, a parallel random coding argument as in the proof of Theorem 5.1 such that

- when encoding, X_i only carries the message meant for receiver i;
- when decoding, receiver i treats X_j, $j \neq i$, as noise,

shows that the rate pair $(snr_1 \lambda (P-\epsilon)/2, snr_2(1-\lambda)(P-\epsilon)/2)$ can be approached, which immediately establishes the achievability part. □

Remark 5.13. For the achievability part, instead of using the power sharing scheme as in the proof, one can also employ the following time sharing scheme: set X to be X_1 for λ fraction of the time, and X_2 for $1 - \lambda$ fraction of the time. Then, it is straightforward to check this scheme also achieves the rate pair $(snr_1 \lambda (P-\epsilon)/2, snr_2(1-\lambda)(P-\epsilon)/2)$. This, from a different perspective, echoes the observation in [52] that time sharing achieves the capacity region of a white Gaussian BC as the bandwidth limit tends to infinity.

Remark 5.14. Theorem 5.11 can be heuristically derived using a similar argument employing the approximation approach as in Remark 5.7.

We have the following theorem, whose proof is inspired by the ideas in Section 4 and parallels that the argument in [18].

Theorem 5.15. *Consider the following continuous-time physically degraded Gaussian broadcast channel with one sender and two receivers:*

$$Y_1(t) = \int_0^t X(s, M_1, Y_{1,0}^s, Y_{2,0}^s) ds + \sqrt{N_1} B_1(t),$$

$$Y_2(t) = \int_0^t X(s, M_2, Y_{1,0}^s, Y_{2,0}^s) ds + \sqrt{N_1} B_1(t) + \sqrt{N_2} B_2(t),$$

where $N_1, N_2 > 0$, and B_1, B_2 are independent standard Brownian motions, and the channel input $X(s)$ is assumed to satisfy Conditions (d)-(f). Then, feedback does not increase the capacity region of the above channel.

Proof. Let X be a $(T, (e^{TR_1}, e^{TR_2}), P)$-code. By the code construction, for $i = 1, 2$, it is possible to estimate the messages M_i from the channel output $Y_{i,0}^T$ with an arbitrarily low probability of error. Hence, by Fano's inequality, for $i = 1, 2$,

$$H(M_i|Y_{i,0}^T) \leq TR_iP_e^{(T)} + H(P_e^{(T)}) = T\varepsilon_{i,T},$$

where $\varepsilon_{i,T} \to 0$ as $T \to \infty$. It then follows that

$$TR_1 = H(M_1) = H(M_1|M_2) \leq I(M_1; Y_{1,0}^T|M_2) + T\varepsilon_{1,T},$$

$$TR_2 = H(M_2) \leq I(M_2; Y_{2,0}^T) + T\varepsilon_{2,T}.$$

Now the Euler-Maruyama approximation with respect to the equidistant Δ_n of stepsize $\delta_n = T/n$ applied to the continuous-time physically degraded Gaussian BC yields:

$$Y_1^{(n)}(t_{n,i}) - Y_1^{(n)}(t_{n,i-1}) = \int_{t_{n,i-1}}^{t_{n,i}} X(s, M, Y_{1,t_{n,0}}^{(n),t_{n,i-1}}, Y_{2,t_{n,0}}^{(n),t_{n,i-1}})ds + \sqrt{N_1}B_1(t_{n,i}) - \sqrt{N_1}B_1(t_{n,i-1}),$$

$$Y_2^{(n)}(t_{n,i}) - Y_2^{(n)}(t_{n,i-1}) = \int_{t_{n,i-1}}^{t_{n,i}} X(s, M, Y_{1,t_{n,0}}^{(n),t_{n,i-1}}, Y_{2,t_{n,0}}^{(n),t_{n,i-1}})ds + \sqrt{N_1}B_1(t_{n,i}) - \sqrt{N_1}B_1(t_{n,i-1}) + \sqrt{N_2}B_2(t_{n,i}) - \sqrt{N_2}B_2(t_{n,i-1}).$$

Then, by Theorem 3.3, we have

$$\begin{aligned} I(M_2; Y_{2,0}^T) &= \lim_{n\to\infty} I(M_2; Y_2^{(n)}(\Delta_n)) \\ &= \lim_{n\to\infty} I(M_2; \Delta Y_2^{(n)}(\Delta_n)) \\ &= \lim_{n\to\infty} H(\Delta Y_2^{(n)}(\Delta_n)) - H(\Delta Y_2^{(n)}(\Delta_n)|M_2), \end{aligned}$$

where $\Delta Y_2^{(n)}(\Delta_n) \triangleq \{Y_2^{(n)}(t_{n,i}) - Y_2^{(n)}(t_{n,i-1}) : i = 1, 2, \cdots, n\}$. Note that

$$H(\Delta Y_2^{(n)}(\Delta_n)) \leq \sum_{i=1}^{n} \log(2\pi e(P\delta_n^2 + N_2\delta_n)),$$

and

$$\begin{aligned} H(\Delta Y_2^{(n)}(\Delta_n)|M_2) &= \sum_{i=1}^{n} H(Y_2^{(n)}(t_{n,i}) - Y_2^{(n)}(t_{n,i-1})|Y_{2,t_{n,0}}^{(n),t_{n,i-1}}, M_2) \\ &\geq \sum_{i=1}^{n} H(\sqrt{N_2}B_2(t_{n,i}) - \sqrt{N_2}B_2(t_{n,i-1})) \\ &= \sum_{i=1}^{n} \log(2\pi e N_2\delta_n), \end{aligned}$$

which implies that there exists an $\alpha \in [0, 1]$ such that

$$H(\Delta Y_2^{(n)}(\Delta_n)|M) = \sum_{i=1}^{n} \frac{n}{2} \log(2\pi e(\alpha P\delta_n^2 + N_2\delta_n)).$$

It then follows from Theorem 3.3 that

$$I(M_2; Y_{2,0}^T) \le \frac{1}{2} \lim_{n \to \infty} \sum_{i=1}^n \log \frac{P\delta_n^2 + N_2 \delta_n}{\alpha P \delta_n^2 + N_2 \delta_n} = \frac{(1-\alpha)PT}{2N_2}.$$

Next we consider

$$\begin{aligned} I(M_1; Y_1^{(n)}(\Delta_n)|M_2) &= H(Y_1^{(n)}(\Delta_n)|M_2) - H(Y_1^{(n)}(\Delta_n)|M_1, M_2) \\ &= H(Y_1^{(n)}(\Delta_n)|M_2) - \sum_{i=1}^n H(Y_1^{(n)}(t_{n,i})|M_1, M_2, Y_{1,t_{n,0}}^{(n),t_{n,i-1}}) \\ &\le H(Y_1^{(n)}(\Delta_n)|M_2) - \sum_{i=1}^n H(Y_1^{(n)}(t_{n,i})|M_1, M_2, Y_{1,t_{n,0}}^{(n),t_{n,i-1}}, Y_{1,t_{n,0}}^{(n),t_{n,i-1}}) \\ &= H(Y_1^{(n)}(\Delta_n)|M_2) - \frac{1}{2} \sum_{i=1}^n \log(2\pi e N_1 \delta_n). \end{aligned}$$

Now, using Lemma 1 in [18] (an extension of the entropy power inequality), we obtain

$$H(Y^{(n)}(\Delta_n)|M_2) \ge \frac{n}{2} \log(2^{2H(Y_1^{(n)}(\Delta_n))|M_2)/n} + 2\pi e(N_2 - N_1)\delta_n),$$

which immediately implies that

$$H(Y_{1,t_{n,0}}^{(n),t_{n,i}}|M_2) \le \frac{1}{2} \sum_{i=1}^n \log(2\pi e(\alpha P \delta_n^2 + N_1 \delta_n))$$

and furthermore, by Theorem 3.3,

$$\begin{aligned} I(M_1; Y_{1,0}^T|M_2) &\le \lim_{n \to \infty} \frac{1}{2} \sum_{i=1}^n \log(2\pi e(\alpha P \delta_n^2 + N_1 \delta_n)) - \frac{1}{2} \sum_{i=1}^n \log(2\pi e N_1 \delta_n) \\ &= \lim_{n \to \infty} \frac{1}{2} \sum_{i=1}^n \log\left(1 + \frac{\alpha P \delta_n}{N_1}\right) \\ &= \frac{\alpha P T}{2N_1}. \end{aligned}$$

Now, by Theorem 5.11, we conclude that feedback capacity region is exactly the same non-feedback capacity region; in other words, feedback does not increase the capacity region of a physically degraded continuous-time Gaussian BC. □

The following theorem follows from a translated version of the argument in [65] coupled with the approximation approach as in Section 4, and so we only provide the sketch of the proof.

Theorem 5.16. *Feedback increases the capacity region of certain continuous-time stochastically degraded Gaussian broadcast channel.*

Proof. Consider the following symmetric continuous-time Gaussian broadcast channel:

$$Y_1(t) = \int_0^t X(s)ds + B_1(t),$$

$$Y_2(t) = \int_0^t X(s)ds + B_2(t),$$

where B_1, B_2 are independent standard Brownian motions, and X satisfies the average power constraint P. By Theorem 5.11, without feedback, the capacity region is the set of rate pairs (R_1, R_2) such that

$$R_1 + R_2 \leq \frac{P}{2}. \tag{87}$$

With feedback, one can use the following variation [65] of the Schalkwijk-Kailath coding scheme [72] over $[0, T]$ at discrete time points $\{t_{n,i}\}$ that form an equidistant Δ_n of stepsize δ_n: after some proper initialization, we send, at time $t_{n,i}$, $X_{t_{n,i}} = X_{1,t_{n,i}} + X_{2,t_{n,i}}$, where

$$X_{1,t_{n,i}} = \gamma_i(X_{1,t_{n,i-1}} - \mathbb{E}[X_{1,t_{n,i-1}}|Y_{1,t_{n,i-1}}]),$$

$$X_{2,t_{n,i}} = -\gamma_i(X_{2,t_{n,i-1}} - \mathbb{E}[X_{2,t_{n,i-1}}|Y_{2,t_{n,i-1}}]),$$

where γ_i is chosen so that $\mathbb{E}[X_i^2] \leq P$ for each i. Going through a completely parallel argument as in [65] and taking the limit as n tends to infinity afterwards, we can then obtain that

$$R_1 = R_2 = \frac{1}{2} \sum_{n \to \infty} \sum_{i=1}^{n} \frac{\log\left(1 + \frac{P\delta_n(1+\rho^*)/2}{(1+P\delta_n(1-\rho^*))/2}\right)}{T} = \frac{P(1+\rho^*)}{4} \tag{88}$$

are achievable, where $\rho^* > 0$ satisfies the condition

$$\rho^*(1 + (P+1)(1 + P(1-\rho^*)/2)) = \frac{P(P+2)(1-\rho^*)}{2}.$$

The claim that feedback strictly increases the capacity region then follows from (87), and (88) and the fact that $\rho^* > 0$. □

6 Conclusions and Future Work

In this paper, we have established natural connections between continuous-time Gaussian memory/feedback channels with their associated discrete-time versions in the forms of sampling and approximating theorems that preserve temporal causality. It turns out that these connections can help us to achieve deeper understanding of continuous-time Gaussian channels from several perspectives: not only can these theorems help us to reinterpret and recover some known facts and results, they can also provide us the rigorous tools, intuitive ideas and inspirations to obtain new results for continuous-time Gaussian channels either in the point-to-point or multi-user setting.

On the other hand though, there are many questions that remain unanswered and a number of directions that need to be further explored.

The first direction is to strengthen our sampling and approximating theorems. Note that both Theorem 2.4 and Theorem 3.3 require Conditions (d)-(f), which are stronger than the average power constraint typically imposed in either practical or theoretical considerations. Obviously, this will render the two theorems fall short of what is required to be applicable in some practical and theoretical situations. For instance, despite the fact that our approximating theorem gives intuitive explanations to or even inspirations for the rigorous treatment of continuous-time multi-user Gaussian channels in Section 5, it fails to rigorously establish Theorems 5.1, 5.8 and 5.11. The stochastic calculus approach employed in Section 5 requires only the power constraints, which can be loosely explained by the fact that Girsanov's theorems (or, more precisely, its several variants) only requires as weak conditions. It is certainly worthwhile to explore whether the assumptions in our sampling and approximating theorems can be relaxed either in general or for some special settings. Another topic in this direction is the rate of convergence in the sampling and approximating theorems. Note that, in an information-theoretic sense, our sampling and approximating theorems have established a continuous-time Gaussian channel as the limit of a sequence of discrete-time ones as the time lapse in between the channel uses and the signal-to-noise ratio in the corresponding discrete-time channel proportionately shrink to zero. On the other hand though, a more quantitative analysis on how fast these discrete-time channels will "approach" the continuous-time one would be of great value in both practice and theory.

The second direction is to further develop relevant information-theoretic tools in stochastic calculus. As mentioned in Section 5, in our proof of Theorem 5.11, Lemma 5.12, which may take a strikingly different look compared with a discrete-time entropy power inequality, serves the same function that an entropy power inequality would in the discrete-time setting; in other words, it is a continuous-time form of the entropy power inequality. On the other hand though, we note that Theorem 5.15 does require Conditions (d)-(f), stronger conditions than the average power constraint, and we somehow didn't find a version of Theorem 5.15 that only assumes the power constraint, as we have done in establishing Theorems 5.1, 5.8 and 5.11. Among many other reasons, the lack of such a theorem can be an indication that some corresponding information-theoretic tools in stochastic calculus are simply missing, and thereby more information-theoretical tools from stochastic calculus need to be developed.

The third direction is to broaden the applications of the approach employed in this work. Although we have shown that for continuous-time Gaussian ICs and BCs, feedback may increase the capacity region of continuous-time Gaussian ICs and BCs, the feedback capacity regions of these two types of channels remain unknown, and so an immediate problem would be to find a simple expression for them using the approach employed in this work, as we have done for continuous-time MACs. One can also consider generalizing the results in this work to the situation with general Gaussian noises [36, 30]. For this topic, note that there exist in-depth studies [33, 34, 31, 42, 32, 37, 41, 38, 7, 40, 68] on continuous-time point-to-point general Gaussian channels with possible feedback. A first step in this direction can be sampling or approximating theorems for stationary Gaussian processes: as elaborated in [28], such theorems for stationary Gaussian processes can help to connect continuous-time and discrete-time stationary Gaussian channels, for which Kim's variational formulation of discrete-time stationary Gaussian feedback capacity [47] proves to rather effective. Of course, further topics also include exploring whether the ideas and techniques in this paper can be applied to multi-hop channels with more general Gaussian noises, or even continuous-time stochastic networks with more general noises and topology.

The fourth direction is to explore the application of our methodology to ultra-wideband communications, since the channels we have considered are "essentially" infinite-bandwidth communication systems. The literature on multi-user Gaussian networks operating at the ultra-wideband regime is vast, and much work has been focused on the interplay between bandwidth, power and transmission rate. Of greater relevance to this work are [59, 79, 52, 80, 81, 82], where, among many other results, the asymptotic behavior of multi-user Gaussian broadcast channels, as the bandwidth tends to infinity (or equivalently, as the power tends to zero), is discussed in great depth. Most of the above-mentioned work used the bandwidth approach to deal with the continuous-time Gaussian channel. Exemplified by the results obtained in this work, it is worthwhile to explore whether our approach employing the approximating theorems and stochastic calculus can provide another way to examine ultra-wideband communication systems with possible memory, feedback and interference, which the bandwidth approach may have difficulty to handle.

Acknowledgement. We would like to thank Jun Chen, Young-Han Kim, Tsachy Weissman and Wenyi Zhang for insightful suggestions and comments, and for pointing out relevant references.

References

[1] R. Ahlswede. Multi-way communication channels. *IEEE ISIT*, 1973.

[2] V. Annapureddy and V. Veeravalli. Gaussian interference networks: Sum capacity in the low interference regime and new outer bounds on the capacity region. *IEEE Trans. Info. Theory*, vol. 55, no. 7, pp. 3032-3050, 2009.

[3] R. Ash. Capacity and error bounds for a time-continuous Gaussian channel. *Information and Control*, vol. 6, pp. 14-27, 1963.

[4] R. Ash. Further discussion of a time-continuous Gaussian channel. *Information and Control*, vol. 7, pp. 78-83, 1964.

[5] R. Ash. *Information Theory*, Wiley Interscience, New York, 1965.

[6] A. Avestimehr, S. Diggavi and D. Tse. Wireless network information flow: A deterministic approach. *IEEE Trans. Info. Theory*, vol. 57, no. 4, pp. 1872-1905, 2011.

[7] R. Baker and S. Ihara. Information capacity of the stationary Gaussian channel. IEEE Trans. Info. Theory, vol. 37, no. 5, pp. 1314-1326, 1991.

[8] P. Bergmans. Random coding theorem for broadcast channels with degraded components. *IEEE Trans. Info. Theory*, vol. 19, pp. 197-207, 1973.

[9] P. Bethoux. Test et estimations concernant certaines functions aleatoires en particulier Laplaciennes. *Ann. Inst. Henri Poincare*, vol. 27, pp. 255-322, 1962.

[10] S. Bross and A. Lapidoth. An improved achievable region for the discrete memoryless two-user multiple-access channel with noiseless feedback. *IEEE Trans. Info. Theory*, vol. 51, no. 3, pp. 811-833, 2005.

[11] R. Cheng and S. Verdu. Gaussian multiaccess channels with ISI: Capacity region and multiuser water-filling. *IEEE Trans. Info. Theory*, vol. 39, no. 3, pp. 773-785, 1993.

[12] T. Cover. Broadcast channels. *IEEE Trans. Info. Theory*, vol. 18, pp. 2-14, 1972.

[13] T. Cover. Some advances in broadcast channels. *Advances in Communication Systems*, Academic Press, San Francisco, vol. 4, pp. 229-260, 1975.

[14] T. Cover and C. Leung. An achievable rate region for the multiple-access channel with feedback. *IEEE Trans. Info. Theory*, vol. 27, no. 3, pp. 292-298, 1981.

[15] T. Cover and J. Thomas. *Elements of Information Theory*, 2nd ed., Wiley Interscience, New York, 2006.

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

[17] R. Durrett. *Probability: Theory and Examples*, 4th ed., Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press, Cambridge, 2010.

[18] A. El Gamal. The capacity of the physically degraded Gaussian broadcast channel with feedback. *IEEE Trans. Info. Theory*, vol. 27, no. 4, pp. 508-511, 1981.

[19] A. El Gamal and Y. Kim. *Network Information Theory*, Cambridge University Press, New York, 2011.

[20] R. Etkin, D. Tse and H. Wang. Gaussian interference channel capacity to within one bit. *IEEE Trans. Info. Theory*, vol. 54, no. 12, pp. 5534-5562, 2008.

[21] R. Fortet. Hypothesis testing and Estimation for Laplacian Functions. *Fourth Berkeley Symposium on Mathematical Statistics and Probability*, vol. 1, pp. 289-305, 1961.

[22] R. Gallager. *Information Theory and Reliable Communication*, Wiley, New York, 1968.

[23] A. Gelfand and I. Yaglom. Calculation of the amount of information about a random function contained in another such function. *Uspekhi Mat. Nauk*, vol. 12, pp. 3-52, 1957. Trans. in *Am. Math. Sot. Trans.*, Ser. 2, vol. 12, pp. 199-247, 1959.

[24] D. Guo, S. Shamai and S. Verdu. Mutual information and minimum mean-square error in Gaussian channels. *IEEE Trans. Info. Theory*, vol. 51, no. 4, pp. 1261-1282, 2005.

[25] T. Han. *Information-Spectrum Method in Information Theory*, Springer, New York, 2003.

[26] T. Han. An information-spectrum approach to capacity theorems for the general multiple-access channel. *IEEE Trans. Info. Theory*, vol. 44, no. 7, pp. 2773-2795, 1998.

[27] T. Han and K. Kobayashi. A new achievable rate region for the interference channel. *IEEE Trans. Info. Theory*, vol. 27, no. 1, pp. 49-60, 1981.

[28] G. Han and J. Song. Extensions of the I-MMSE relationship to Gaussian channels with feedback and memory. *IEEE Trans. Info. Theory*, vol. 62, no. 10, pp. 5422-5445, 2016.

[29] T. Hida. *Canonical representations of Gaussian processes and their applications*, Memoirs of the College of Science, University of Kyoto, Series A: Mathematics, vol. 33, no. 1, pp. 109-155, 1960.

[30] T. Hida and M. Hitsuda. *Gaussian processes*, vol. 120, American Mathematical Soc., 1993.

[31] M. Hitsuda. Mutual information in Gaussian channels. *J. Multivariate Anal.*, vol. 4, pp. 66-73, 1974.

[32] M. Hitsuda and S. Ihara. Gaussian channels and the optimal coding. *J. Multivariate Anal.*, vol. 5, pp. 106-118, 1975.

[33] R. Huang and R. Johnson. Information capacity of time-continuous channels. *IEEE Trans. Info. Theory*, IT-8, pp. l91-l98, 1962.

[34] R. Huang and R. Johnson. Information transmission with time-continuous random processes. *IEEE Trans. Info. Theory*, vol. 9, no. 2, pp. 84-94, 1963.

[35] P. Kloeden and E. Platen. *Numerical Solution of Stochastic Differential Equations*, Stochastic Modelling and Applied Probability, vol. 23, Springer-Verlag, Berlin Heidelberg, 1992.

[36] I. Ibragimov and Y. Rozanov. *Gaussian Random Processes*, Springer-Verlag, New York, 1978.

[37] S. Ihara, S. On the capacity of the continuous time Gaussian channel with feedback. *J. Multivariate Anal.*, vol. 10, pp. 319-331, 1980.

[38] S. Ihara. Capacity of mismatched Gaussian channels with and without feedback. *Probability Theory Rel. Fields*, vol. 84, pp. 453-471, 1990.

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

[40] S. Ihara. Coding theorems for a continuous-time Gaussian channel with feedback. *IEEE Trans. Info. Theory*, vol. 40, no. 6, pp. 2014-2045, 1994.

[41] S. Ihara. Mutual information in stationary channels with additive noise. *IEEE Trans. Info. Theory*, vol. 31, no. 5, pp. 602-606, 1985.

[42] B. Jacob, M. Zakai and J. Ziv. On the ε-entropy and the rate-distortion function of certain non-Gaussian processes. *IEEE Trans. Info. Theory*, vol. 20, no. 4, pp. 517-524, 1974.

[43] T. Kadota, M. Zakai and J. Ziv. Mutual information of the white Gaussian channel with and without feedback. *IEEE Trans. Info. Theory*, vol. 17, pp. 368-371, 1971.

[44] I. Karatzas and S. Shreve. *Brownian Motion and Stochastic Calculus*, New York: Springer-Verlag, 1991.

[45] Y. Kim. *Gaussian Feedback Capacity*, Ph.D. Thesis, Stanford University, 2006.

[46] Y. Kim. Feedback capacity of the first-order moving average Gaussian channel. *IEEE Trans. Info. Theory*, vol. 52, no. 7, pp. 3063-3079, 2006.

[47] Y. Kim. Feedback capacity of stationary Gaussian channels. *IEEE Trans. Info. Theory*, vol. 56, no. 1, pp. 57-85, 2010.

[48] G. Kramer. *Directed Information for Channels with Feedback*, Ph.D. Dissertation, Swiss Federal Institute of Technology (ETH), Zurich, Switzerland, 1998.

[49] G. Kramer. Capacity results for the discrete memoryless network. *IEEE Trans. Info. Theory*, vol. 49, no. 1, pp. 4-21, 2003.

[50] Y. Kutoyants. *Statistical inference for ergodic diffusion processes*, Springer, London, 2004.

[51] A. Lapidoth. On the asymptotic capacity of stationary Gaussian fading channels. *IEEE Trans. Info. Theory*, vol. 51, no. 2, pp. 437-446, 2005.

[52] A. Lapidoth, E. Telatar and R. Urbanke. On wide-band broadcast channels. *IEEE Trans. Info. Theory.* vol. 49, no. 12, pp. 3250-3258, 2003.

[53] A. Lapidoth and M. Wigger. On the Gaussian MAC with imperfect feedback. In *Proc. 24th IEEE Conv. Electrical and Electronics Engineers in Israel (IEEEI06)*, Eilat, Israel, pp. 203-207, 2006.

[54] A. Leon-Garcia. *Probability, Statistics, and Random Processes for Electrical Engineering*, Pearson/Prentice Hall, N.J., 2008

[55] R. Liptser and A. Shiryaev. *Statistics of random processes (I): General theory*, 2nd edition, Springer-Verlag, Berlin, 2001.

[56] X. Liu and G. Han. Recent results in continuous-time network information theory. *IEEE ISIT*, pp. 2421-2425, 2014.

[57] X. Mao. *Stochastic Differential Equations and Applications*, Horwood, 1997.

[58] N. Martins and T. Weissman. Coding schemes for additive white noise channels with feedback corrupted by quantization or bounded noise. *IEEE Trans. Info. Theory*, submitted for publication.

[59] R. McEliece and L. Swanson. A note on the wide-band Gaussian broadcast channel. *IEEE Trans. Commun.*, vol. 35, pp. 452-453, 1987.

[60] N. Merhav and T. Weissman. Coding for the feedback Gelfand-Pinsker channel and the feedforward Wyner-Ziv source. *IEEE Trans. Info. Theory*, vol. 52, no. 9, pp. 4207-4211, 2006.

[61] A. Motahari and A. Khandani. Capacity bounds for the Gaussian interference channel. *IEEE Trans. Info. Theory*, vol. 55, no. 2, pp. 620-643, 2009.

[62] H. Nyquist. Certain factors affecting telegraph speed. *The Bell System Technical Journal*, pp. 324, 1924.

[63] B. Oksendal. *Stochastic Differential Equations: An Introduction with Applications*, Springer, Berlin, 1995.

[64] L. Ozarow. The capacity of the white Gaussian multiple access channel with feedback. *IEEE Trans. Info. Theory*, vol. 30, no. 4, pp. 623-628, 1984.

[65] L. Ozarow and S. Leung. An achievable region and outer bound for the Gaussian broadcast channel with feedback. *IEEE Trans. Info. Theory*, vol. 30, no. 4, pp. 667-671, 1984.

[66] H. Permuter, T. Weissman and J. Chen. Capacity region of the finite-state multiple-access channel with and without feedback. *IEEE Trans. Info. Theory*, vol. 55, no. 6, pp. 2455-2477, 2009.

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

[68] M. Pinsker, V. Prelov and S. Verdu. Sensitivity of the ε-entropy of stationary continuous-time Gaussian processes. *Probl. Peredachi Inf.*, vol. 33, no. 2, pp. 3-25, 1997.

[69] H. Poor. *An introduction to signal detection and estimation*, Springer-Verlag, New York, 1994.

[70] H. Royden. *Real analysis*, 4th edition, Prentice Hall, Boston, 2010.

[71] H. Sato. On the capacity region of a discrete two-user channel for strong interference. *IEEE Trans. Info. Theory*, vol. 24, no. 3, pp. 377-379, 1978.

[72] J. Schalkwijk and T. Kailath. Coding scheme for additive noise channels with feedback I: No bandwidth constraint. *IEEE Trans. Info. Theory*, vol. 12, no. 2, pp. 172-182, 1966.

[73] X. Shang, G. Kramer and B. Chen. A new outer bound and the noisy-interference sum-rate capacity for Gaussian interference channels. *IEEE Trans. Info. Theory*, vol. 55, no. 2, pp. 689-699, 2009.

[74] C. Shannon. A mathematical theory of communication. *The Bell System Technical Journal*, vol. 27, pp. 379-423, 1948.

[75] C. Shannon. Communication in the presence of noise. *Proc. IRE*, vol. 37, pp. 10-21, 1949.

[76] D. Slepian. On Bandwidth. *Proc. IEEE*, vol. 64, no. 3, 1976.

[77] C. Suh and D. Tse. Feedback capacity of the Gaussian interference channel to within 2 Bits. *IEEE Trans. Info. Theory*, vol. 57, no. 5, pp. 2667-2685, 2011.

[78] S. Verdu. Multiple-access channels with memory with and without frame synchronism. *IEEE Trans. Info. Theory*, vol. 35, no. 3, pp. 605-619, 1989.

[79] S. Verdu. On Channel Capacity per Unit Cost. *IEEE Trans. Info. Theory*, vol. 36, no. 5, pp. 1019-1030, 1990.

[80] S. Verdu. Spectral efficiency in the wideband regime. *IEEE Trans. Info. Theory*, vol. 48, no. 6, pp. 1319-1343, 2002.

[81] S. Verdu. Recent results on the capacity of wideband channels in the low-power regime. *IEEE Wireless Commun.*, vol. 1, pp. 40-47, 2002.

[82] S. Verdu, G. Caire and D. Tuninetti. Is TDMA optimal in the low power regime? *IEEE ISIT*, p. 193, 2002.

[83] T. Weissman, Y. Kim and H. Permuter. Directed information, causal estimation, and communication in continuous time. *IEEE Trans. Info. Theory*, vol. 59, no. 3, pp. 1271-1287, 2013.

[84] F. Willems. The feedback capacity region of a class of discrete memoryless multiple access channels. *IEEE Trans. Info. Theory*, vol. 28, no. 1, pp. 93-95, 1982.

[85] W. Wu, S. Vishwanath and A. Arapostathis. On the capacity of multiple access channels with state information and feedback. *IEEE Trans. Info. Theory*, submitted for publication.

[86] A. Wyner. The capacity of the band-limited Gaussian channel, *The Bell System Technical Journal*, pp. 359-395, 1966. Also reprinted in *Key Papers in the Development of Information Theory*, edited by D. Slepian, IEEE Press, New York, pp. 190-193, 1974.

[87] A. Wyner. Recent results in the Shannon theory. *IEEE Trans. Info. Theory*, vol. 20, no. 1, pp. 2-10, 1974.