On Sampling Continuous-Time AWGN Channels

Guangyue Han[ID], *Senior Member, IEEE*, and Shlomo Shamai[ID], *Life Fellow, IEEE*

Abstract—For a continuous-time additive white Gaussian noise (AWGN) channel with possible feedback, it has been shown that as sampling gets infinitesimally fine, the mutual information of the associative discrete-time channels converges to that of the original continuous-time channel. We give in this paper more quantitative strengthenings of this result, which, among other implications, characterize how over-sampling approaches the true mutual information of a continuous-time Gaussian channel with bandwidth limit. The assumptions in our results are relatively mild. In particular, for the non-feedback case, compared to the Shannon-Nyquist sampling theorem, a widely used tool to connect continuous-time Gaussian channels to their discrete-time counterparts that requires the band-limitedness of the channel input, our results only require some integrability conditions on the power spectral density function of the input.

Index Terms—Continuous-time additive white Gaussian noise channel, the Shannon-Nyquist sampling theorem, mutual information, the I-MMSE relationship, stochastic differential equation.

I. INTRODUCTION

IN THIS paper, we are concerned with the following continuous-time AWGN channel:

$$Y(t) = \int_0^t X(s)ds + B(t), \quad t \geq 0, \tag{1}$$

where $\{B(t)\}$ denotes the standard Brownian motion. We will examine the channel (1) for both the non-feedback and feedback cases. More specifically, the non-feedback case refers to the scenario when the feedback is not allowed in the channel, and thereby the channel input $\{X(s)\}$ takes the following form:

$$X(s) = g(s, M), \tag{2}$$

where M is a random variable over a finite alphabet $\mathcal{M}$ which is independent of $\{B(t)\}$ and often interpreted as the *message* to be transmitted through the channel, and g is a real-valued deterministic function from $[0, \infty) \times \mathcal{M}$ to $\mathbb{R}$. By contrast,

for the feedback case, $X(s)$ may also depend on the previous output with the following form:

$$X(s) = g(s, M, Y_0^s), \tag{3}$$

where g is a *non-anticipative* functional from $[0, \infty) \times \mathcal{M} \times C[0, \infty)$ to $\mathbb{R}$ in the sense that, roughly speaking, with t interpreted as the present time, for any $m \in \mathcal{M}$ and $\varphi \in C[0, \infty)$, the value of $g(t, m, \varphi)$ only causally depends on the "history" of φ (see the rigorous definition in Section II), and $Y_0^s \triangleq \{Y(r) : 0 \leq r \leq s\}$ is the channel output up to time s that is fed back to the sender, which will be referred to as the *channel feedback* up to time s. Here, we remark that for the non-feedback case as in (2), it can be verified that for any $T > 0$,

$$I(X_0^T; Y_0^T) = I(M; Y_0^T),$$

which however does not hold true for the feedback case as in (3) when the channel is intricately characterized by the following stochastic differential equation:

$$Y(t) = \int_0^t g(s, M, Y_0^s)ds + B(t), \quad t \geq 0, \tag{4}$$

rather than a simple input-output equation.

For any $T > 0$, we say that $t_0, t_1, \ldots, t_n \in \mathbb{R}$[1] are *evenly spaced* over $[0, T]$ if $t_0 = 0$, $t_n = T$ and $t_i - t_{i-1} = T/n$ for all feasible i, and we will use $\delta_{T,n}$ to denote the stepsize, i.e., $\delta_{T,n} \triangleq t_1 - t_0$. Sampling the continuous-time channel (1) over the time window $[0, T]$ with respect to evenly spaced $t_0, t_1, \ldots, t_n$, we obtain the following discrete-time Gaussian channel[2]

$$Y(t_i) = \int_0^{t_i} X(s)ds + B(t_i), \quad i = 1, 2, \ldots, n. \tag{5}$$

It turns out that if the sampling is "fine" enough, the mutual information of the continuous-time Gaussian channel (1) over $[0, T]$ can be "well-approximated" by that of the discrete-time Gaussian channel (5). More precisely, it has been established in [17] that under some mild assumptions,

$$\lim_{n \to \infty} I(M; Y(\Delta_{T,n})) = I(M; Y_0^T), \tag{6}$$

where

$$\Delta_{T,n} \triangleq \{t_0, t_1, \ldots, t_n\},$$
$$Y(\Delta_{T,n}) \triangleq \{Y(t_0), Y(t_1), \ldots, Y(t_n)\}.$$

[1]Here and hereafter, all t_i depend on T and n, however we suppress this notational dependence for brevity.

[2]The sampler associated with (5) has been examined from a communication system design perspective and termed as the *integrate-and-dump* filter; see, e.g., [1], [3] and references therein.

This result connects a continuous-time Gaussian channels with its associative discrete-time versions, which has been used to recover a classical information-theoretic formula [16].

Strictly speaking, the above result is not new; as a matter of fact, it has been "known" for decades. Indeed, though not explicitly stated, a more general result, which implies (6) as a special corollary, follows from an appropriately modified argument in [7] (see more details in Appendix A). Moreover, the functional analysis approach as in [7] tackles more general channels and thus reveals the essence beneath the connection as in (6). By comparison, when it comes to merely establishing the result (6), our stochastic approach in [17] is indirect and cumbersome, only scratching the surface of the connection.

That being said, the stochastic calculus approach does allow us to capitalize on the peculiar characteristics of the continuous-time AWGN channel formulated as in (1), which is already evidenced by the approximation theorems established in [17] that do not seem to follow from the aforementioned functional analysis approach. The present paper will continue to employ the stochastic calculus approach to conduct a closer examination of (1) and quantitatively strengthen (6), particularly by zooming in on its convergence behavior. Our results encompass both the non-feedback case (Theorem III.1) and the feedback case (Theorem IV.4), which may be used for a finer analysis of continuous-time AWGN channels from an information-theoretic perspective. In particular, among other possible implications, our results characterize how over-sampling (see, e.g., [5] and references therein) approaches the true mutual information of a continuous-time band-limited AWGN channel over a finite time window (Corollary III.2). Though sampling has been extensively studied in the fields of signal processing and information theory (see, e.g., [14] and the references therein), we are not aware of any results in the literature that of comparable forms to our results.

We would like to add that the assumptions imposed in our results are rather mild. Indeed, to connect a continuous-time non-feedback AWGN channel to its discrete-time versions over an infinite time window, a conventional approach is to use the Shannon-Nyquist sampling theorem [22], [29], which requires the band-limitedness of the channel input. By comparison, with considerably stronger conclusions, Theorem III.1 and Theorem IV.4 do so over a finite time window, which is typically deemed more of practical interest; and moreover, Theorem III.1 only requires some integrability conditions on the power spectral density function of the channel input and Theorem IV.4 only requires some mild regularity conditions that are more or less standard in the theory of stochastic differential equations.

II. NOTATION AND PRELIMINARIES

We use $(\Omega, \mathcal{F}, \mathbb{P})$ to denote the underlying probability space, and $\mathbb{E}$ to denote the expectation with respect to the probability measure $\mathbb{P}$. As is typical in the theory of SDEs, we assume the probability space is equipped with a filtration $\{\mathcal{F}_t : 0 \leq t < \infty\}$, which satisfies the *usual conditions* [12] and is rich enough to accommodate the standard Brownian motion $\{B(t) : 0 \leq t < \infty\}$. Throughout the paper, we will use uppercase letters (e.g., $X, Y, Y^{(n)}$) to denote random variables or processes, and their lowercase counterparts (e.g., $x, y, y^{(n)}$) to denote their realizations.

Let $C[0, \infty)$ denote the space of all continuous functions over $[0, \infty)$, and for any $t > 0$, let $C[0, t]$ denote the space of all continuous functions over $[0, t]$. As usual, we will equip the space $C[0, \infty)$ with the filtration $\{\mathcal{B}_t\}_{0 \leq t < \infty}$, where $\mathcal{B}_\infty$ denotes the standard Borel σ-algebra on the space $C[0, \infty)$ and $\mathcal{B}_t = \pi_t^{-1}(\mathcal{B}_\infty)$, where $\pi_t : C[0, \infty) \to C[0, t]$ is defined as $(\pi_t x)(s) = x(t \wedge s)$. We say that a functional f from $[0, \infty) \times \mathcal{M} \times C[0, \infty)$ to $\mathbb{R}$ is *non-anticipative* if f is $\mathcal{B}([0, \infty) \otimes 2^{\mathcal{M}} \otimes \mathcal{B}_\infty$-measurable and for any $m \in \mathcal{M}$ and any $t \in [0, \infty)$, $f(t, m, \cdot)$ is $\mathcal{B}_t$-measurable, or equivalently, for $\varphi \in C[0, \infty)$,

$$f(t, m, \varphi) = f(t, m, \{\varphi(s); 0 \leq s \leq t\}).$$

For any $\varphi \in C[0, \infty)$, we use $\varphi(\{t_1, t_2, \ldots, t_n\})$ to denote $\{\varphi(t_1), \varphi(t_2), \ldots, \varphi(t_n)\}$ and φ_0^t to denote $\{\varphi(s) : 0 \leq s \leq t\}$. The sup-norm of φ_0^t, denoted by $\|\varphi_0^t\|$, is defined as $\|\varphi_0^t\| \triangleq \sup_{0 \leq s \leq t} |\varphi(s)|$; and similarly, we define $\|\varphi_0^t - \psi_0^t\| \triangleq \sup_{0 \leq s \leq t} |\varphi(s) - \psi(s)|$. For any $\varphi, \psi \in C[0, \infty)$, slightly abusing the notation, we define $\|\varphi_0^s - \psi_0^t\| \triangleq \|\hat{\varphi}_0^\infty - \hat{\psi}_0^\infty\|$, where $\hat{\varphi}, \hat{\psi} \in C[0, \infty)$ are "stopped" versions of φ, ψ at time s, t, respectively, with $\hat{\varphi}(r) \triangleq \varphi(r \wedge s)$ and $\hat{\psi}(r) \triangleq \psi(r \wedge t)$ for any $r \geq 0$.

Let X, Y, Z be random variables defined on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$, which will be used to illustrate most of the notions and facts in this section (note that the same notations may have different connotations in other sections). Note that in this paper, a random variable can be discrete-valued with a probability mass function, real-valued with a probability density function or path-valued (more precisely, $C[0, \infty)$- or $C[0, t]$-valued).

For any two probability measures μ and ν, we write $\mu \sim \nu$ to mean they are equivalent, namely, μ is absolutely continuous with respect to ν and vice versa. For any two path-valued random variables $X_0^t = \{X(s); 0 \leq s \leq t\}$ and $Y_0^t = \{Y(s); 0 \leq s \leq t\}$, we use $\mu_{X_0^t}$ and $\mu_{Y_0^t}$ to denote the probability distributions on $\mathcal{B}_t$ induced by X_0^t and Y_0^t, respectively; and if $\mu_{Y_0^t}$ is absolutely continuous with respect to $\mu_{X_0^t}$, we write the Radon-Nikodym derivative of $\mu_{Y_0^t}$ with respect to $\mu_{X_0^t}$ as $d\mu_{Y_0^t}/d\mu_{X_0^t}$. We use $\mu_{Y_0^t|Z=z}$ denote the probability distribution on $\mathcal{B}_t$ induced by Y_0^t given $Z = z$, and $d\mu_{Y_0^t|Z=z}/d\mu_{X_0^t|Z=z}$ to denote the Radon-Nikodym derivative of Y_0^t with respect to X_0^t given $Z = z$. Obviously, when Z is independent of X, $d\mu_{Y_0^t|Z=z}/d\mu_{X_0^t|Z=z} = d\mu_{Y_0^t|Z=z}/d\mu_{X_0^t}$.

By definition, for $\mathbb{E}[X|\sigma(Y, Z)]$, the conditional expectation of X with respect to the σ-algebra generated by Y and Z, there exists a $\sigma(Y) \otimes \sigma(Z)$-measurable function $\Psi(\cdot, \cdot)$ such that $\Psi(Y, Z) = \mathbb{E}[X|\sigma(Y, Z)]$. For notational convenience, we will in this paper simply write $\mathbb{E}[X|\sigma(Y, Z)]$ as $\mathbb{E}[X|Y, Z]$, and $\Psi(y, z)$ as $\mathbb{E}[X|y, z]$ and furthermore, $\Psi(Y, z)$ as $\mathbb{E}[X|Y, z]$.

A *partition* of the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is a disjoint collection of elements of $\mathcal{F}$ whose union is Ω. It is well known there is a one-to-one correspondence between finite partitions and finite sub-σ-algebras of $\mathcal{F}$. For a finite sub-σ-algebra $\mathcal{H} \subseteq \mathcal{F}$, let $\eta(\mathcal{H})$ denote the corresponding finite partition. The entropy of a finite partition $\xi = \{A_1, A_2, \cdots, A_m\}$, denoted

by $H(\xi)$, is defined as $H(\xi) \triangleq \sum_{i=1}^{m} -\mathbb{P}(A_i) \log \mathbb{P}(A_i)$, whereas the conditional entropy of ξ given another finite partition $\zeta = \{B_1, B_2, \ldots, B_n\}$, denoted by $H(\xi|\zeta)$, is defined as $H(\xi|\zeta) \triangleq \sum_{j=1}^{n} \sum_{i=1}^{m} -\mathbb{P}(A_i \cap B_j) \log \mathbb{P}(A_i|B_j)$. The mutual information between the above-mentioned two partitions ξ and ζ, denoted by $I(\xi; \zeta)$, is defined as $I(\xi; \zeta) \triangleq \sum_{j=1}^{n} \sum_{i=1}^{m} \mathbb{P}(A_i \cap B_j) \log \mathbb{P}(A_i \cap B_j)/\mathbb{P}(A_i)\mathbb{P}(B_j)$.

For the random variable X, we define

$$\eta(X) \triangleq \{\eta(\mathcal{H}) : \mathcal{H} \text{ is a finite sub-}\sigma\text{-algebra of } \sigma(X)\}.$$

The *entropy* of the random variable X, denoted by $H(X)$, is defined as

$$H(X) \triangleq \sup_{\xi \in \eta(X)} H(\xi).$$

The *conditional entropy* of Y given X, denoted by $H(Y|X)$, is defined as

$$H(Y|X) \triangleq \inf_{\xi \in \eta(X)} \sup_{\zeta \in \eta(Y)} H(\zeta|\xi).$$

Here, we note that if X and Y are independent, then obviously it holds that

$$H(Y|X) = H(Y). \tag{7}$$

The *mutual information* between X and Y, denoted by $I(X; Y)$, is defined as

$$I(X; Y) \triangleq \sup_{\xi \in \eta(X),\ \zeta \in \eta(Y)} I(\xi; \zeta).$$

A couple of properties of mutual information are in order. First, it can be shown, via a concavity argument, that the mutual information is always non-negative. Second, the mutual information is determined by the σ-algebras generated by the corresponding random variables. For example, for any random variables X', Y', X'', Y'',

$$I(X'; Y') = I(X''; Y'')$$
$$\text{if } \sigma(X') = \sigma(X'') \text{ and } \sigma(Y') = \sigma(Y'') \tag{8}$$

and

$$I(X'; Y') \leq I(X''; Y'')$$
$$\text{if } \sigma(X') \subseteq \sigma(X'') \text{ and } \sigma(Y') \subseteq \sigma(Y''). \tag{9}$$

For another example, we have

$$I(X; Y) = I(X, X; Y, Y + X), \quad I(X; Y) \leq I(X; Y, Z).$$

It turns out that for the case that X, Y, Z are all discrete random variables, all the above-mentioned notions are well-defined and can be computed rather explicitly: $H(X)$ can be computed as $H(X) = \mathbb{E}[-\log p_X(X)]$, where $p_X(\cdot)$ denotes the probability mass function of X; $H(Y|X)$ can be computed as $H(Y|X) = \mathbb{E}[-\log p_{Y|X}(Y|X)]$, where $p_{Y|X}(\cdot|\cdot)$ denotes the conditional probability mass function of Y given X; $I(X; Y)$ can be computed as

$$I(X; Y) = \mathbb{E}\left[\log \frac{p_{Y|X}(Y|X)}{p_Y(Y)}\right]. \tag{10}$$

The mutual information is intimately related to entropy. As an example, one verifies that

$$I(X; Y) = H(Y) - H(Y|X). \tag{11}$$

Note that the quality (11) may fail if non-discrete random variables are involved, since the corresponding entropies $H(Y)$ and $H(Y|X)$ can be infinity. For the case of real-valued random variables with density, this issue can be circumvented using the notion of differential entropy, as elaborated below.

Now, assume that Y is real-valued with probability density function $f_Y(\cdot)$. The *differential entropy* of Y, denoted by $h(Y)$, is defined as $h(Y) \triangleq \mathbb{E}[-\log f_Y(Y)]$. And the *differential conditional entropy* of Y given a finite partition $\xi = \{A_1, A_2, \ldots, A_n\}$, denoted by $h(Y|\zeta)$, is defined as $h(Y|\zeta) \triangleq \sum_{j=1}^{n} \mathbb{P}(A_i) \int f_{Y|A_i}(x) \log f_{Y|A_i}(x) dx$. The *differential conditional entropy* of Y given X (which can be discrete-, real- or path-valued), denoted by $h(Y|X)$, is defined as $h(Y|X) \triangleq \inf_{\xi \in \eta(X)} h(Y|\xi)$ (which, similarly as in (7), reduces to $h(Y)$ if Y is independent of X); in particular, if the conditional probability density function $f_{Y|X}(\cdot|\cdot)$ exists, then $h(Y|X)$ can be explicitly computed as $\mathbb{E}[-\log f_{Y|X}(Y|X)]$. As mentioned before, the aforementioned failure of (11) can be salvaged with the notion of differential entropy:

$$I(X; Y) = h(Y) - h(Y|X). \tag{12}$$

Here we emphasize that all the above-mentioned definitions naturally carry over to the setting when some/all of the invovled random variables are replaced by vectors of random variables. For a quick example, let $Y = \{Y_1, Y_2, \ldots, Y_n\}$, where each Y_i is a real-valued random variable with density. Then, the differential entropy $h(Y)$ of Y is defined as

$$h(Y) = h(Y_1, Y_2, \ldots, Y_n)$$
$$\triangleq \mathbb{E}[-\log f_{Y_1, Y_2, \ldots, Y_n}(Y_1, Y_2, \ldots, Y_n)],$$

where $f_{Y_1, Y_2, \ldots, Y_n}$ is the joint probability density function of $Y_1, Y_2, \ldots, Y_n$.

The notion of mutual information can be further extended to generalized random processes, which we will only briefly describe and we refer the reader to [7] for a more comprehensive exposition.

The mutual information between two generalized random processes $X = \{X(t)\}$ and $Y = \{Y(t)\}$ is defined as

$$I(X; Y) = \sup I(X(\phi_1), X(\phi_2), \ldots, X(\phi_m);$$
$$Y(\psi_1), Y(\psi_2), \ldots, Y(\psi_n)), \tag{13}$$

where the supremum is over all possible $n, m \in \mathbb{N}$ and all possible testing functions $\phi_1, \phi_2, \ldots, \phi_m$ and $\psi_1, \psi_2, \ldots, \psi_n$, and we have defined

$$X(\phi_i) = \int X(t)\phi_i(t) dt, \quad i = 1, 2, \ldots, m,$$
$$Y(\psi_j) = \int Y(t)\psi_j(t) dt, \quad j = 1, 2, \ldots, n.$$

It can be verified that the general definition of mutual information as in (13) includes all previous definitions as special cases; moreover, when one of X and Y, say, Y, is a random variable, the general definition boils down to

$$I(X; Y) = \sup I(X(\phi_1), X(\phi_2), \ldots, X(\phi_m); Y),$$

where the supremum is over all possible $n \in \mathbb{N}$ and all possible testing functions $\phi_1, \phi_2, \ldots, \phi_m$.

For the channel (1) with the input as in (2) or (3) , it is known that its mutual information over $[0, T]$ can be computed as (see, e.g., [11], [25]):

$$I(M; Y_0^T) = \begin{cases} \mathbb{E}\left[\log \dfrac{d\mu_{M,Y_0^T}}{d\mu_M \times \mu_{Y_0^T}}(M, Y_0^T)\right], & \text{if } \dfrac{d\mu_{M,Y_0^T}}{d\mu_M \times \mu_{Y_0^T}} \text{ exists}, \\ \infty, & \text{otherwise}, \end{cases} \tag{14}$$

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

III. THE NON-FEEDBACK CASE

In this section, adopting the notations in Section I, we examine the AWGN channel (1) for the non-feedback case and give quantitative strengthenings of (6), detailed in the following theorem.

Theorem III.1: For the continuous-time AWGN channel (1), suppose that the channel input $\{X(t)\}$ is a stationary stochastic process with power spectral density function $f(\cdot)$.[3] Then, the following two statements hold:

(a) Suppose $\int f(\lambda)|\lambda|d\lambda < \infty$. Then, for any T and n,

$$\sqrt{I(X_0^T; Y_0^T)} \leq \frac{1}{2}\left(\sqrt{2T\delta_{T,n}\int f(\lambda)|\lambda|d\lambda} + \sqrt{2T\delta_{T,n}\int f(\lambda)|\lambda|d\lambda + 4I(X_0^T; Y(\Delta_{T,n}))}\right).$$

(b) Suppose $\int f(\lambda)|\lambda|d\lambda < \infty$ and $\int f(\lambda)d\lambda < \infty$. Then, for any T and n,

$$I(X_0^T; Y_0^T) - I(X_0^T; Y(\Delta_{T,n})) \leq$$
$$T\sqrt{\delta_{T,n}}\left(\int f(\lambda)|\lambda|d\lambda\right)^{1/2}\left(\int f(\lambda)d\lambda\right)^{1/2}. \tag{15}$$

Consequently, for any T, choosing $n = n(T)$ such that $\lim_{T\to\infty} \delta_{T,n(T)} = 0$, we have

$$\frac{I(X_0^T; Y_0^T)}{T} - \frac{I(X_0^T; Y(\Delta_{T,n(T)}))}{T} = O\left(\sqrt{\delta_{T,n(T)}}\right), \tag{16}$$

as T tends to infinity.

Proof: Consider the following parameterized version of the channel (1):

$$Z(t) = \sqrt{snr}\int_0^t X(s)ds + B(t), \quad t \in [0, T], \tag{17}$$

where the parameter $snr > 0$ can be regarded as the signal-to-noise ratio of the channel. Obviously, when snr is fixed to be 1, $Z(t) = Y(t)$ for any $t \in [0, T]$, and moreover, the channel (17) is exactly the same as the channel (1).

Sampling the channel (17) with respect to sampling times $t_0, t_1, \ldots, t_n$ that are evenly spaced over $[0, T]$, we obtain the following discrete-time Gaussian channel:

$$Z(t_i) - Z(t_{i-1}) = \sqrt{snr}\int_{t_{i-1}}^{t_i} X(s)ds + B(t_i) - B(t_{i-1}),$$
$$i = 1, 2, \ldots, n, \tag{18}$$

which can be "normalized" as follows:

$$\frac{Z(t_i) - Z(t_{i-1})}{\sqrt{\delta_{T,n}}} = \sqrt{snr}\frac{\int_{t_{i-1}}^{t_i} X(s)ds}{\sqrt{\delta_{T,n}}} + \frac{B(t_i) - B(t_{i-1})}{\sqrt{\delta_{T,n}}},$$
$$i = 1, 2, \ldots, n, \tag{19}$$

where, at each time i, the channel noise $\frac{B(t_i)-B(t_{i-1})}{\sqrt{\delta_{T,n}}}$ is a standard Gaussian random variable and $\frac{\int_{t_{i-1}}^{t_i} X(s)ds}{\sqrt{\delta_{T,n}}}$ and $\frac{Z(t_i)-Z(t_{i-1})}{\sqrt{\delta_{T,n}}}$ should be regarded as the channel input and output, respectively.

Applying the continuous-time I-MMSE relationship (see Theorem 6 in [9]) to the channel (17), we have

$$\frac{d}{dsnr}I(X_0^T; Z_0^T) = \frac{1}{2}\int_0^T \mathbb{E}[(X(s) - \mathbb{E}[X(s)|Z_0^T])^2]ds.$$

Moreover, we deduce from (17) that

$$I(X_0^T; Z_0^T)|_{snr=1} = I(X_0^T; Y_0^T), \quad I(X_0^T; Z_0^T)|_{snr=0} = 0,$$

where we have in addition used the fact that $\{X(t)\}$ and $\{B(t)\}$ are independent in deriving the second equality. It then follows that the mutual information of the channel (1) can be computed as

$$I(X_0^T; Y_0^T) = I(X_0^T; Z_0^T)|_{snr=1} - I(X_0^T; Z_0^T)|_{snr=0}$$
$$= \int_0^1 \frac{d}{dsnr}I(X_0^T; Z_0^T)dsnr$$
$$= \frac{1}{2}\int_0^1 \int_0^T \mathbb{E}[(X(s) - \mathbb{E}[X(s)|Z_0^T])^2]dsdsnr.$$

Similarly, applying the discrete-time I-MMSE relationship (see Theorem 1[4] in [9]) to the channel (19), we have

$$I(X_0^T; Y(\Delta_{T,n})) = \frac{1}{2}\int_0^1 \sum_{i=1}^n \mathbb{E}\left[\left(\frac{\int_{t_{i-1}}^{t_i} X(s)ds}{\sqrt{\delta_{T,n}}}\right.\right.$$
$$\left.\left.- \mathbb{E}\left[\frac{\int_{t_{i-1}}^{t_i} X(s)ds}{\sqrt{\delta_{T,n}}}\bigg| Z(\Delta_{T,n})\right]\right)^2\right]dsnr.$$

Obviously, by (9), it holds true that $I(X_0^T; Y_0^T) \geq I(X_0^T; Y(\Delta_{T,n}))$. In the following, we will give an upper bound on their difference $I(X_0^T; Y_0^T) - I(X_0^T; Y(\Delta_{T,n}))$,

---

[3]More precisely, the channel input $\{X(t) : t \geq 0\}$ can be extended to a bi-infinite stationary stochastic process $\{X(t) : -\infty < t < \infty\}$ with power spectral density function $f(\cdot)$.

[4]More precisely, rather than the discrete-time I-MMSE relationship, we may have to resort to its extension to Gaussian vector channels (see Theorem 2 in [9]) or its further extension to Gaussian memory/feedback channels (see Theorem 3 in [10]) as the channel input to the channel (19) may not be an independent process.

thereby characterizing the closeness between the two quantities. Towards this goal, applying the well-known fact that for any random variables U, V and W with $\sigma(V) \subseteq \sigma(W)$,

$$\mathbb{E}[(U - \mathbb{E}[U|W])^2] \le \mathbb{E}[(U - \mathbb{E}[U|V])^2],$$

we first deduce that for each i,

$$\mathbb{E}\left[\left(\int_{t_{i-1}}^{t_i} X(s) - \mathbb{E}[X(s)|Z_0^T]\, ds\right)^2\right]$$
$$\le \mathbb{E}\left[\left(\int_{t_{i-1}}^{t_i} X(s) - \mathbb{E}[X(s)|Z(\Delta_{T,n})]\, ds\right)^2\right].$$

It then follows that

$$I(X_0^T; Y_0^T) - I(X_0^T; Y(\Delta_{T,n}))$$
$$= \frac{1}{2}\int_0^1\left(\int_0^T \mathbb{E}[(X(s) - \mathbb{E}[X(s)|Z_0^T])^2]ds\right.$$
$$\left. - \sum_{i=1}^n \mathbb{E}\left[\left(\int_{t_{i-1}}^{t_i} \frac{X(s) - \mathbb{E}[X(s)|Z(\Delta_{T,n})]}{\sqrt{\delta_{T,n}}}ds\right)^2\right]\right)dsnr$$
$$\le \frac{1}{2}\int_0^1\left(\int_0^T \mathbb{E}[(X(s) - \mathbb{E}[X(s)|Z_0^T])^2]ds\right.$$
$$\left. - \sum_{i=1}^n \mathbb{E}\left[\left(\int_{t_{i-1}}^{t_i} \frac{X(s) - \mathbb{E}[X(s)|Z_0^T]}{\sqrt{\delta_{T,n}}}ds\right)^2\right]\right)dsnr$$
$$= \frac{1}{2}\int_0^1\left(\int_0^T \mathbb{E}[R^2[X(s); Z_0^T]]ds\right.$$
$$\left. - \sum_{i=1}^n \mathbb{E}\left[\left(\int_{t_{i-1}}^{t_i} \frac{R[X(s); Z_0^T]}{\sqrt{\delta_{T,n}}}ds\right)^2\right]\right)dsnr$$
$$= \frac{1}{2}(S_1 + S_2),$$

where we have used the shorthand notation $R[X(s); Z_0^T]$ for $X(s) - \mathbb{E}[X(s)|Z_0^T]$ and

$$S_1 \triangleq \sum_{i=1}^n \int_0^1 \left(\int_{t_{i-1}}^{t_i} \mathbb{E}[R^2[X(s); Z_0^T]]ds\right.$$
$$\left. - \sum_{i=1}^n \int_{t_{i-1}}^{t_i} \mathbb{E}\left[R[X(s); Z_0^T]R[X(t_{i-1}); Z_0^T]\right]ds\right)dsnr,$$
$$S_2 \triangleq \sum_{i=1}^n \int_0^1 \mathbb{E}\left[\left(\int_{t_{i-1}}^{t_i} R[X(s); Z_0^T]ds\right)R[X(t_{i-1}); Z_0^T]\right]dsnr$$
$$- \sum_{i=1}^n \int_0^1 \mathbb{E}\left[\left(\int_{t_{i-1}}^{t_i} \frac{R[X(s); Z_0^T]}{\sqrt{\delta_{T,n}}}ds\right)^2\right]dsnr.$$

For the first term, we have that the following (20) holds, where we have used the Cauchy-Schwarz inequality for the last inequality. Now, applying the well-known fact that for any random variables U and V,

$$\mathbb{E}[(U - \mathbb{E}[U|V])^2] = \mathbb{E}[U^2] - \mathbb{E}[\mathbb{E}^2[U|V]] \le \mathbb{E}[U^2],$$

we deduce that

$$\mathbb{E}[R^2[X(s) - X(t_{i-1}); Z_0^T]] \le \mathbb{E}[(X(s) - X(t_{i-1}))^2],$$

and we continue from (20) to reach (21). Now, using the fact that, for any $u, v \in \mathbb{R}$,

$$\mathbb{E}[X(u)X(u + v)] = \int f(\lambda)e^{iv\lambda}d\lambda,$$

we have that (22) holds true.

For the second term, we have

$$S_2^2 = \left(\sum_{i=1}^n \int_{t_{i-1}}^{t_i} \int_0^1 \mathbb{E}\left[R[X(s); Z_0^T]\right.\right.$$
$$\left.\left.\times \int_{t_{i-1}}^{t_i} \frac{R[X(t_{i-1}) - X(u); Z_0^T]}{\delta_{T,n}}du\right]dsnrds\right)^2.$$

Starting from this and proceeding in a similar fashion as in (20)-(22), shown at the bottom of the next page, we derive

$$S_2^2 \le 2T\delta_{T,n}I(X_0^T; Y_0^T)\int f(\lambda)|\lambda|d\lambda.$$

Combining the bounds on S_1 and S_2, we have

$$I(X_0^T; Y_0^T) - I(X_0^T; Y(\Delta_{T,n}))$$
$$\le \sqrt{2T\delta_{T,n}}\left(I(X_0^T; Y_0^T)\int f(\lambda)|\lambda|d\lambda\right)^{1/2}. \quad (23)$$

It then immediately follows that

$$\sqrt{I(X_0^T; Y_0^T)} \le \frac{1}{2}\left(\sqrt{2T\delta_{T,n}\int f(\lambda)|\lambda|d\lambda}\right.$$
$$\left. + \sqrt{2T\delta_{T,n}\int f(\lambda)|\lambda|d\lambda + 4I(X_0^T; Y(\Delta_{T,n}))}\right),$$

establishing (a). Moreover, together with the fact that

$$I(X_0^T; Y_0^T) \le \frac{1}{2}\int_0^T \mathbb{E}[X^2(s)]ds = \frac{T}{2}\int f(\lambda)d\lambda,$$

the inequality (23) implies that

$$\left|I(X_0^T; Y_0^T) - I(X_0^T; Y(\Delta_{T,n}))\right|$$
$$\le T\sqrt{\delta_{T,n}}\left(\int f(\lambda)|\lambda|d\lambda\right)^{1/2}\left(\int f(\lambda)d\lambda\right)^{1/2},$$

establishing (b).   ∎

The following corollary, which immediately follows from Theorem III.1, characterizes, among others, how oversampling approaches the true mutual information of the channel (1) with bandwidth limit.

Corollary III.2: For the continuous-time AWGN channel (1), suppose that the channel input has bandwidth limit W and average power P, or more precisely, $f(\lambda) = 0$ for any $\lambda \in (-\infty, -W] \cup [W, \infty)$ and $\mathbb{E}[X^2(s)] \le P$ for any $s \ge 0$. Then, the following two statements hold:

(a) For any T and n,

$$\sqrt{I(X_0^T; Y_0^T)} \le \frac{1}{2}\left(\sqrt{2TPW\delta_{T,n}}\right.$$
$$\left. + \sqrt{2TPW\delta_{T,n} + 4I(X_0^T; Y(\Delta_{T,n}))}\right).$$

(b) For any T and n,

$$I(X_0^T; Y_0^T) - I(X_0^T; Y(\Delta_{T,n})) \le TP\sqrt{W\delta_{T,n}}.$$

Consequently, for each T, choosing $n = n(T)$ such that $\lim_{T \to \infty} \delta_{T, n(T)} = 0$, we have

$$\frac{I(X_0^T; Y_0^T)}{T} - \frac{I(X_0^T; Y(\Delta_{T,n}))}{T} = O\left(\sqrt{\delta_{T, n(T)}}\right),$$

as T tends to infinity.

Remark III.3: The I-MMSE relationship has played an important role in deriving our results for non-feedback AWGN channels. This powerful tool has been extended [10] to feedback AWGN channels in both discrete and continuous time. A natural question is whether the extended relationship can help us to derive counterpart results to Theorem III.1 for the feedback case. As tempting and promising as this idea may look, we failed to find a way to effectively apply the extended I-MMSE relationship, and our treatment for the feedback case, to be detailed in the next section, will use other tools and techniques from the theory of stochastic calculus. Here, we remark that the formulas derived in [10] are valid only with some extra assumption imposed and yet fail to hold true in general. For more detailed explanations and corrected formulas, see Arxiv:1401.3527.

IV. THE FEEDBACK CASE

In this section, adopting the notations in Section I, we give quantitative strengthenings of (6) for the AWGN channel (1) in the feedback case, which we remind the reader is characterized by the stochastic differential equation in (4).

The following regularity conditions may be imposed:

(I) **The uniform Lipschitz condition:** There exists a constant $L > 0$ such that for any $0 \le s_1, s_2, t_1, t_2 \le T$, any $M \in \mathcal{M}$ and any $Y_0^T, Z_0^T \in C[0, T]$,

$$|g(s_1, M, Y_0^{s_2}) - g(t_1, M, Z_0^{t_2})| \le L(|s_1 - t_1| + \|Y_0^{s_2} - Z_0^{t_2}\|).$$

(II) **The uniform linear growth condition:** There exists a constant $L > 0$ such that for any $M \in \mathcal{M}$ and any $Y_0^T \in C[0, T]$,

$$|g(t, M, Y_0^t)| \le L(1 + \|Y_0^t\|).$$

Remark IV.1: 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. These two conditions will be taken for granted in most practical

$$S_1^2 = \left(\sum_{i=1}^{n} \int_{t_{i-1}}^{t_i} \int_0^1 \mathbb{E}[R^2[X(s); Z_0^T]] ds nr ds - \sum_{i=1}^{n} \int_{t_{i-1}}^{t_i} \int_0^1 \mathbb{E}\left[R[X(s); Z_0^T] R[X(t_{i-1}); Z_0^T]\right] ds nr ds\right)^2$$

$$= \left(\sum_{i=1}^{n} \int_{t_{i-1}}^{t_i} \int_0^1 \mathbb{E}[R[X(s); Z_0^T] R[X(s) - X(t_{i-1}); Z_0^T]] ds nr ds\right)^2$$

$$\le n \sum_{i=1}^{n} \left(\int_{t_{i-1}}^{t_i} \int_0^1 \mathbb{E}[R[X(s); Z_0^T] R[X(s) - X(t_{i-1}); Z_0^T]] ds nr ds\right)^2$$

$$\le n \sum_{i=1}^{n} \int_{t_{i-1}}^{t_i} \int_0^1 \mathbb{E}[R^2[X(s); Z_0^T]] ds nr ds \int_{t_{i-1}}^{t_i} \int_0^1 \mathbb{E}[R^2[X(s) - X(t_{i-1}); Z_0^T]] ds nr ds \qquad (20)$$

$$S_1^2 \le n \sum_{i=1}^{n} \int_{t_{i-1}}^{t_i} \int_0^1 \mathbb{E}[R^2[X(s); Z_0^T]] ds nr ds \int_{t_{i-1}}^{t_i} \int_0^1 \mathbb{E}[(X(s) - X(t_{i-1}))^2] ds nr ds$$

$$= n \sum_{i=1}^{n} \int_{t_{i-1}}^{t_i} \int_0^1 \mathbb{E}[R^2[X(s); Z_0^T]] ds nr ds \int_{t_{i-1}}^{t_i} \int_0^1 \mathbb{E}[X^2(s) + X^2(t_{i-1}) - 2X(s)X(t_{i-1})] ds nr ds \qquad (21)$$

$$S_1^2 \le n \sum_{i=1}^{n} \int_{t_{i-1}}^{t_i} \int_0^1 \mathbb{E}[R^2[X(s); Z_0^T]] ds nr ds \int_{t_{i-1}}^{t_i} \int_0^1 \int 2f(\lambda)|1 - e^{i\lambda(s - t_{i-1})}| d\lambda ds nr ds$$

$$\le n \sum_{i=1}^{n} \int_{t_{i-1}}^{t_i} \int_0^1 \mathbb{E}[R^2[X(s); Z_0^T]] ds nr ds \int_{t_{i-1}}^{t_i} \int_0^1 \int 2f(\lambda)|\lambda|(s - t_{i-1}) d\lambda ds nr ds$$

$$\le n \delta_{T,n}^2 \int f(\lambda)|\lambda| d\lambda \int_0^1 \sum_{i=1}^{n} \int_{t_{i-1}}^{t_i} \mathbb{E}[R^2[X(s); Z_0^T]] ds ds nr$$

$$= T \delta_{T,n} \int f(\lambda)|\lambda| d\lambda \int_0^1 \int_0^T \mathbb{E}[(X(s) - \mathbb{E}[X(s)|Z_0^T])^2] ds ds nr$$

$$= 2T \delta_{T,n} I(X_0^T; Y_0^T) \int f(\lambda)|\lambda| d\lambda \qquad (22)$$

communication situations: as might be expected, the signals employed in practice will be much better-behaving. For a quick examples, it can be easily verified that the following linear feedback coding scheme

$$g(t, M, Y_0^t) = h(t, M) + \int_0^t k(s) dY(s)$$

satisfies the two conditions, provided some appropriate assumptions are imposed on the two functions $h(\cdot, \cdot)$ and $k(\cdot)$.

We will need the following lemma, which has already been established [17] in a slightly more general setting.

Lemma IV.2: Assume Conditions (I) and (II). Then, there exists a unique strong solution of (4) with initial value $Y(0) = 0$. Moreover, there exists $\varepsilon > 0$ such that

$$\mathbb{E}[e^{\varepsilon \|Y_0^T\|^2}] < \infty, \tag{24}$$

which immediately implies

$$\mathbb{P}\left(\int_0^T g^2(t, M, Y_0^t) dt < \infty\right) = \mathbb{P}\left(\int_0^T g^2(t, M, B_0^t) dt < \infty\right) = 1 \tag{25}$$

and

$$\int_0^T \mathbb{E}[|g(t, M, Y_0^t)|] dt < \infty. \tag{26}$$

We will also need the following lemma, whose proof is deferred to Appendix B.

Lemma IV.3: Let $Z_{max} = \max\{Z_1, Z_2, \ldots, Z_n\}$, where the i.i.d. random variables $Z_i \sim N(0, 1/n)$. Then, we have:

a) For any $0 < \varepsilon < 1$, $\mathbb{E}[Z_{max}^2] = O(n^{-(1-\varepsilon)})$.
b) For any $0 < \varepsilon < 1$, $\mathbb{E}[Z_{max}^4] = O(n^{-(2-\varepsilon)})$.
c) For any $0 < \varepsilon < 1$, $\mathbb{E}[e^{Z_{max}^2}] = 1 + O(n^{-(1-\varepsilon)})$.

As detailed below, the following theorem gives the aforementioned quantitative strengthening of (6). Here, we mention that with the channel input as in (4), the discrete-time channel (5) obtained by sampling the channel (1) over $[0, T]$ with respect to $\Delta_{T,n}$ takes the following form:

$$Y(t_i) = \int_0^{t_i} g(s, M, Y_0^s) ds + B(t_i), \quad i = 0, 1, \ldots, n. \tag{27}$$

Theorem IV.4: Fix $T > 0$ and assume Conditions (I) and (II). Then, for any $0 < \varepsilon < 1/2$, we have

$$I(M; Y_0^T) = I(M; Y(\Delta_{T,n})) + O(\delta_{T,n}^{1/2-\varepsilon}), \tag{28}$$

where we recall from Section II that $Y(\Delta_{T,n}) = \{Y(t_0), Y(t_1), \ldots, Y(t_n)\}$ and we remark the constant in the $O(\delta_{T,n}^{1/2-\varepsilon})$-term may depend on T.

Remark IV.5: Roughly speaking, (15) and (16) characterize the closeness between the mutual information of a continuous-time AWGN channel and its associative discrete time versions with respect to fixed T and scaled T, respectively, in the non-feedback setting. Noting that (28) is the counterpart result to (15) in the feedback setting, we remark that we are not able to establish a counterpart result to (16) in the feedback setting.

Proof: First of all, recall from Section II that for a stochastic process $\{X(s)\}$ and any $t \in \mathbb{R}_+$, we use $\mu_{X_0^t}$ to denote the distribution on $C[0, t]$ induced by X_0^t. Throughout

the proof, we only have to deal with the case $t = T$, and so we will simply write $\mu_{X_0^T}$ as μ_X. Moreover, we will rewrite $\Delta_{T,n}$ as Δ_n and $\delta_{T,n}$ as δ_n for notational simplicity.

Note that with (25) and (26), an application of Theorem 7.14 of [15] yields that

$$\mathbb{P}\left(\int_0^T \mathbb{E}^2[g(t, M, Y_0^t)|Y_0^t] dt < \infty\right) = 1. \tag{29}$$

Now for any $m \in \mathcal{M}$, fix $M = m$. One then verifies that the assumptions of Theorem C.1 are all satisfied (Theorem C.1 is a rephrased version of Lemma 7.7 of [15], whose general assumptions boil down to (25), (26) and (29) when restricted to our settings). It then follows that (Below "$\sim$" means "equivalent"; see Section II)

$$\mu_{Y|M=m} \sim \mu_B, \tag{30}$$

and moreover, with probability 1,

$$\frac{d\mu_{Y|M}}{d\mu_B}(Y_0^T|M) = \frac{1}{\mathbb{E}[e^{\rho_1(M, Y_0^T)}|Y_0^T, M]}, \tag{31}$$

where

$$\rho_1(m, Y_0^T) \triangleq -\int_0^T g(s, m, Y_0^s) dY(s) + \frac{1}{2} \int_0^T g(s, m, Y_0^s)^2 ds. \tag{32}$$

Since $\rho_1(M, Y_0^T)$ is a deterministic functional of M and Y_0^T, we conclude that, with probability 1,

$$\frac{d\mu_{Y|M}}{d\mu_B}(Y_0^T|M) = \frac{1}{e^{\rho_1(M, Y_0^T)}}. \tag{33}$$

Furthermore, using a parallel argument, we infer that

$$\mu_Y \sim \mu_B, \tag{34}$$

and moreover, with probability 1,

$$\frac{d\mu_Y}{d\mu_B}(Y_0^T) = \frac{1}{\mathbb{E}[e^{\rho_1(M, Y_0^T)}|Y_0^T]}. \tag{35}$$

Note that it follows from $\mathbb{E}[d\mu_B/d\mu_Y(Y_0^T)] = 1$ that $\mathbb{E}[e^{\rho_1(M, Y_0^T)}] = 1$, which is equivalent to

$$\mathbb{E}[e^{-\int_0^T g(s, M, Y_0^s) dB(s) - \frac{1}{2} \int_0^T g(s, M, Y_0^s)^2 ds}] = 1. \tag{36}$$

Then, a parallel argument as in the proof of Theorem 7.1 of [15] (which requires the condition (36)) further implies that, for any n,

$$\frac{d\mu_{Y(\Delta_n)|M}}{d\mu_{B(\Delta_n)}}(Y(\Delta_n)|M) = \frac{1}{\mathbb{E}[e^{\rho_1(M, Y_0^T)}|Y(\Delta_n), M]}, \quad a.s., \tag{37}$$

and

$$\frac{d\mu_{Y(\Delta_n)}}{d\mu_{B(\Delta_n)}}(Y(\Delta_n)) = \frac{1}{\mathbb{E}[e^{\rho_1(M, Y_0^T)}|Y(\Delta_n)]}, \quad a.s.. \tag{38}$$

Next, by the definition of mutual information, we have

$$I(M; Y_0^T) = \mathbb{E}\left[\log \frac{d\mu_{M,Y_0^T}}{d\mu_M \times \mu_{Y_0^T}}(M, Y_0^T)\right]$$

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

$$(39)$$

and

$$I(M; Y(\Delta_n)) \tag{40}$$

$$= \mathbb{E}\left[\log f_{Y(\Delta_n)|M}(Y(\Delta_n)|M)\right] - \mathbb{E}\left[\log f_{Y(\Delta_n)}(Y(\Delta_n))\right]$$

$$= \mathbb{E}\left[\log \frac{d\mu_{Y(\Delta_n)|M}}{d\mu_{B(\Delta_n)}}(Y(\Delta_n)|M)\right] - \mathbb{E}\left[\log \frac{d\mu_{Y(\Delta_n)}}{d\mu_{B(\Delta_n)}}(Y(\Delta_n))\right]. \tag{41}$$

Furthermore, by (39), (35), (31) and (33), we have

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

$$= -\mathbb{E}[\log \mathbb{E}[e^{\rho_1(M,Y_0^T)}|Y_0^T, M]] + \mathbb{E}[\log \mathbb{E}[e^{\rho_1(M,Y_0^T)}|Y_0^T]]$$

$$= -\mathbb{E}[\rho_1(M, Y_0^T)] + \mathbb{E}[\log \mathbb{E}[e^{\rho_1(M,Y_0^T)}|Y_0^T]],$$

and by (41), (38) and (37), we have

$$I(M; Y(\Delta_n))$$

$$= \mathbb{E}\left[\log \frac{d\mu_{Y(\Delta_n)|M}}{d\mu_{B(\Delta_n)}}(Y(\Delta_n)|M)\right] - \mathbb{E}\left[\log \frac{d\mu_{Y(\Delta_n)}}{d\mu_{B(\Delta_n)}}(Y(\Delta_n))\right]$$

$$= -\mathbb{E}[\log \mathbb{E}[e^{\rho_1(M,Y_0^T)}|Y(\Delta_n), M]] + \mathbb{E}[\log \mathbb{E}[e^{\rho_1(M,Y_0^T)}|Y(\Delta_n)]].$$

We next proceed in the following two steps.

Step 1. In this step, we show that for any $0 < \varepsilon < 1/2$,

$$\mathbb{E}\left[\log \frac{d\mu_{Y|M}}{d\mu_B}(Y_0^T|M)\right] - \mathbb{E}\left[\log \frac{d\mu_{Y(\Delta_n)|M}}{d\mu_{B(\Delta_n)}}(Y(\Delta_n)|M)\right] = O(\delta_n^{1/2-\varepsilon}).$$

Apparently, it suffices to show that for any $0 < \varepsilon < 1/2$,

$$\mathbb{E}[|\log \mathbb{E}[e^{\rho_1(M,Y_0^T)}|Y(\Delta_n), M] - \rho_1(M, Y_0^T)|] = O(\delta_n^{1/2-\varepsilon}). \tag{42}$$

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 $i = 0, 1, \ldots, n$, let $\bar{Y}_{\Delta_n}(t_i) = Y(t_i)$, and for any $t_{i-1} < s < t_i$ with $s = \lambda t_{i-1} + (1 - \lambda)t_i$ for some $0 < \lambda < 1$, let $\bar{Y}_{\Delta_n}(s) = \lambda Y(t_{i-1}) + (1 - \lambda)Y(t_i)$. Here we note that any sample path of $\bar{Y}_{\Delta_n,0}^T$ defined this way is continuous over $[0, T]$. Let $\bar{g}_{\Delta_n}(s, M, \bar{Y}_{\Delta_n,0}^s)$ denote the piecewise "flat" version of $g(s, M, \bar{Y}_{\Delta_n,0}^s)$ with respect to Δ_n; more precisely, for any $t_{i-1} \leq s < t_i$, $\bar{g}_{\Delta_n}(s, M, \bar{Y}_{\Delta_n,0}^s) = g(t_{i-1}, M, \bar{Y}_{\Delta_n,0}^{t_{i-1}})$.

Letting

$$\rho_2(\Delta_n, m, Y_0^T) \triangleq -\int_0^T \bar{g}_{\Delta_n}(s, m, \bar{Y}_{\Delta_n,0}^s)dY(s)$$

$$+ \frac{1}{2}\int_0^T \bar{g}_{\Delta_n}^2(s, m, \bar{Y}_{\Delta_n,0}^s)ds,$$

we have that (43), shown at the bottom of the page, holds true, where we have used the fact that

$$\mathbb{E}[e^{\rho_2(\Delta_n,M,Y_0^T)}|Y(\Delta_n), M] = e^{\rho_2(\Delta_n,M,Y_0^T)}, \tag{44}$$

since $\rho_2(\Delta_n, M, Y_0^T)$ only depends on M and $Y(\Delta_n)$.

We now prove the following rate of convergence:

$$\mathbb{E}\left[(\rho_1(M, Y_0^T) - \rho_2(\Delta_n, M, Y_0^T))^2\right] = O(\delta_n^{1-\varepsilon}), \tag{45}$$

where $0 < \varepsilon < 1$. To this end, we note that

$$\rho_1(M, Y_0^T) - \rho_2(\Delta_n, M, Y_0^T) = -\int_0^T (g(s) - \bar{g}_{\Delta_n}(s))dB(s)$$

$$- \frac{1}{2}\int_0^T (g(s) - \bar{g}_{\Delta_n}(s))^2 ds, \tag{46}$$

where we have rewritten $g(s, M, Y_0^s)$ as $g(s)$, and $\bar{g}_{\Delta_n}(s, M, \bar{Y}_{\Delta_n,0}^s)$ as $\bar{g}_{\Delta_n}(s)$. It then follows that (45) boils down to

$$\mathbb{E}\left[\left(-\int_0^T (g(s) - \bar{g}_{\Delta_n}(s))dB(s) - \frac{1}{2}\int_0^T (g(s) - \bar{g}_{\Delta_n}(s))^2 ds\right)^2\right] = O(\delta_n^{1-\varepsilon}). \tag{47}$$

To establish (47), notice that, by the Itô isometry [23], we have

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

We next prove that for any $0 < \varepsilon < 1$,

$$\mathbb{E}\left[\int_0^T (g(s) - \bar{g}_{\Delta_n}(s))^2 ds\right] = O(\delta_n^{1-\varepsilon}). \tag{48}$$

To see this, we note that, by Conditions (I) and (II), there exists $L_1 > 0$ such that for any $s \in [0, T]$ with $t_{i-1} \leq s < t_i$, (49), shown at the bottom of the next page, holds true. Similarly, there exists $L_2 > 0$ such that for any $s \in [0, T]$, (50), shown at the bottom of the next page, holds true, where we have used the non-negativity of the following expressions when deriving (a) therein:

$$\max_{s \in [t_{i-1},t_i]}(B(s) - B(t_{i-1})), -\min_{s \in [t_{i-1},t_i]}(B(s) - B(t_{i-1})),$$

$$\max_{s \in [t_{i-1},t_i]}(B(s) - B(t_i)), -\min_{s \in [t_{i-1},t_i]}(B(s) - B(t_i)).$$

It is well-known (see, e.g., Theorem 2.21 in [20]) that $\max_{s \in [t_{i-1},t_i]}(B(s) - B(t_{i-1}))$ is distributed as

$$\log \mathbb{E}[e^{\rho_1(M,Y_0^T)}|Y(\Delta_n), M] = \log \mathbb{E}[e^{\rho_2(\Delta_n,M,Y_0^T)+\rho_1(M,Y_0^T)-\rho_2(\Delta_n,M,Y_0^T)}|Y(\Delta_n), M]$$

$$= \log e^{\rho_2(\Delta_n,M,Y_0^T)}\mathbb{E}[e^{\rho_1(M,Y_0^T)-\rho_2(\Delta_n,M,Y_0^T)}|Y(\Delta_n), M]$$

$$= \rho_2(\Delta_n, M, Y_0^T) + \log \mathbb{E}[e^{\rho_1(M,Y_0^T)-\rho_2(\Delta_n,M,Y_0^T)}|Y(\Delta_n), M] \tag{43}$$

$|B(t_i) - B(t_{i-1})|$, which, together with the fact that $\{B(t)\}$ has independent increments, leads to

$$\mathbb{E}\left[\left(\max_i\left\{\max_{s\in[t_{i-1},t_i]}(B(s) - B(t_{i-1}))\right\}\right)^2\right]$$

$$= \mathbb{E}\left[\left(\max_i\{|B(t_i) - B(t_{i-1})|\}\right)^2\right]$$

$$\leq \mathbb{E}\left[\left(\max_i\{B(t_i) - B(t_{i-1})\}\right)^2\right]$$

$$+ \mathbb{E}\left[\left(\min_i\{B(t_i) - B(t_{i-1})\}\right)^2\right].$$

Now, applying Lemma IV.3 $a)$, we conclude that for any $0 < \varepsilon < 1$,

$$\mathbb{E}\left[\left(\max_i\left\{\max_{s\in[t_{i-1},t_i]}(B(s) - B(t_{i-1}))\right\}\right)^2\right] = O(\delta_n^{1-\varepsilon}).$$

And a completely parallel argument yields that for any $0 < \varepsilon < 1$,

$$\mathbb{E}\left[\left(\max_i\left\{-\min_{s\in[t_{i-1},t_i]}(B(s) - B(t_{i-1}))\right\}\right)^2\right] = O(\delta_n^{1-\varepsilon}),$$

and moreover,

$$\mathbb{E}\left[\left(\max_i\left\{\max_{s\in[t_{i-1},t_i]}(B(s) - B(t_i))\right\}\right)^2\right] = O(\delta_n^{1-\varepsilon}),$$

$$\mathbb{E}\left[\left(\max_i\left\{-\min_{s\in[t_{i-1},t_i]}(B(s) - B(t_i))\right\}\right)^2\right] = O(\delta_n^{1-\varepsilon}),$$

where for the latter two, we have used, in addition, the well-known fact that the time reversed Brownian motion is still a Brownian motion. Noticing that, by Lemma IV.2, $\|Y_0^T\|^2$ is integrable, we arrive at (48) for an arbitrarily small $\varepsilon > 0$.

A similar argument as above, coupled with Lemma IV.3 $b)$ (rather than Lemma IV.3 $a)$), will yield that for any $0 < \varepsilon < 1$,

$$\mathbb{E}\left[\left(\int_0^T (g(s) - \bar{g}_{\Delta_n}(s))^2 ds\right)^2\right] = O(\delta_n^{2-\varepsilon}), \qquad (51)$$

which, together with (47) and (48), implies (45), as desired.

We now prove the following rate of convergence:

$$\mathbb{E}[\log \mathbb{E}[e^{\rho_1(M,Y_0^T) - \rho_2(\Delta_n, M, Y_0^T)}|Y(\Delta_n), M]] = O(\delta_n^{1-\varepsilon}), \qquad (52)$$

where $0 < \varepsilon < 1$. To this end, we first note that for any n,

$$\mathbb{E}[\log \mathbb{E}[e^{\rho_1(M,Y_0^T) - \rho_2(\Delta_n, M, Y_0^T)}|Y(\Delta_n), M]]$$

$$\leq \log \mathbb{E}[\mathbb{E}[e^{\rho_1(M,Y_0^T) - \rho_2(\Delta_n, M, Y_0^T)}|Y(\Delta_n), M]]$$

$$= \log \mathbb{E}[e^{\rho_1(M,Y_0^T) - \rho_2(\Delta_n, M, Y_0^T)}]$$

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

where for the last equality, we have used (46). Moreover, it follows from (48) that for any n,

$$\int_0^T (g(s) - \bar{g}_{\Delta_n}(s))^2 ds < \infty, \quad a.s.,$$

and so the conditions of Lemma 2 of [8] are satisfied, which readily leads to

$$\mathbb{E}[e^{-\int_0^T (g(s) - \bar{g}_{\Delta_n}(s))dB(s) - \frac{1}{2}\int_0^T (g(s) - \bar{g}_{\Delta_n}(s))^2 ds}] \leq 1, \quad (53)$$

and furthermore,

$$\mathbb{E}[\log \mathbb{E}[e^{\rho_1(M,Y_0^T) - \rho_2(\Delta_n, M, Y_0^T)}|Y(\Delta_n), M]] \leq 0.$$

For another direction, we have

$$\mathbb{E}[\log \mathbb{E}[e^{\rho_1(M,Y_0^T) - \rho_2(\Delta_n, M, Y_0^T)}|Y(\Delta_n), M]]$$

$$\geq \mathbb{E}[\mathbb{E}[\log e^{\rho_1(M,Y_0^T) - \rho_2(\Delta_n, M, Y_0^T)}|Y(\Delta_n), M]]$$

$$|g(s, M, \bar{Y}_{\Delta_n,0}^s) - \bar{g}_{\Delta_n}(s, M, \bar{Y}_{\Delta_n,0}^s)| = |g(s, M, \bar{Y}_{\Delta_n,0}^s) - g(t_{i-1}, M, \bar{Y}_{\Delta_n,0}^{t_{i-1}})|$$

$$\leq L_1(|s - t_{i-1}| + \|\bar{Y}_{\Delta_n,0}^s - \bar{Y}_{\Delta_n,0}^{t_{i-1}}\|)$$

$$\leq L_1(|s - t_{i-1}| + |Y(t_i) - Y(t_{i-1})|)$$

$$\leq L_1\delta_n + L_1\delta_n + L_1\delta_n\|Y_0^T\| + |B(t_i) - B(t_{i-1})| \qquad (49)$$

$$|g(s, M, Y_0^s) - g(s, M, \bar{Y}_{\Delta_n,0}^s)|$$

$$\leq L_2\delta_n + L_2\delta_n\|Y_0^T\| + \max_i \max_{s\in[t_{i-1},t_i]} \max\{|B(s) - B(t_{i-1})|, |B(s) - B(t_i)|\}$$

$$\leq L_2\delta_n + L_2\delta_n\|Y_0^T\| + \max_i \left(\max\{\max_{s\in[t_{i-1},t_i]}(B(s) - B(t_{i-1})), -\min_{s\in[t_{i-1},t_i]}(B(s) - B(t_{i-1})),\right.$$

$$\left. \max_{s\in[t_{i-1},t_i]}(B(s) - B(t_i)), -\min_{s\in[t_{i-1},t_i]}(B(s) - B(t_i))\}\right)$$

$$\overset{(a)}{\leq} L_2\delta_n + L_2\delta_n\|Y_0^T\| + \max_i\left\{\max_{s\in[t_{i-1},t_i]}(B(s) - B(t_{i-1}))\right\} + \max_i\left\{-\min_{s\in[t_{i-1},t_i]}(B(s) - B(t_{i-1}))\right\}$$

$$+ \max_i\left\{\max_{s\in[t_{i-1},t_i]}(B(s) - B(t_i))\right\} + \max_i\left\{-\min_{s\in[t_{i-1},t_i]}(B(s) - B(t_i))\right\} \qquad (50)$$

$$
\begin{aligned}
&= \mathbb{E}[\mathbb{E}[\rho_1(M, Y_0^T) - \rho_2(\Delta_n, M, Y_0^T)|Y(\Delta_n), M]]\\
&= \mathbb{E}[\rho_1(M, Y_0^T) - \rho_2(\Delta_n, M, Y_0^T)]\\
&= -\frac{1}{2}\int_0^T \mathbb{E}[(g(s) - \bar{g}_{\Delta_n}(s))^2]ds,
\end{aligned}
$$

which, together with (48), leads to (52). It can be readily verified that (42) follows from (45) and (52), establishing Step 1.

Step 2. In this step, we will prove that for any $0 < \varepsilon < 1/2$, there exists a constant $C > 0$ such that for all n,

$$
-\mathbb{E}\left[\log\frac{d\mu_Y}{d\mu_B}(Y_0^T)\right] + \mathbb{E}\left[\log\frac{d\mu_{Y(\Delta_n)}}{d\mu_{B(\Delta_n)}}(Y(\Delta_n))\right] \leq C\delta_n^{1/2-\varepsilon}. \tag{54}
$$

First of all, note that by Theorem C.1, we have,

$$
\frac{d\mu_Y}{d\mu_B}(Y_0^T) = e^{\int_0^T \hat{g}(Y_0^s)dY(s) - \frac{1}{2}\int_0^T \hat{g}^2(Y_0^s)ds}, \tag{55}
$$

where $\hat{g}(Y_0^s) = \mathbb{E}[g(s, M, Y_0^s)|Y_0^s]$. And by (37), we have

$$
\frac{d\mu_{Y(\Delta_n)}}{d\mu_{B(\Delta_n)}}(Y(\Delta_n)) = \int \frac{d\mu_{Y(\Delta_n)|M}}{d\mu_{B(\Delta_n)}}(Y(\Delta_n)|m)d\mu_M(m)
$$

$$
= \int \frac{1}{\mathbb{E}[e^{\rho_1(M, Y_0^T)}|Y(\Delta_n), m]}d\mu_M(m).
$$

It then follows that (56), shown at the bottom of the page, holds true, where we have used the shorthand notation $\hat{\rho}_1(Y_0^T)$ for $-\int_0^T \hat{g}(Y_0^s)dY(s) + \frac{1}{2}\int_0^T \hat{g}^2(Y_0^s)ds$, and we have used (44) in deriving the last equality. It then follows that (57), shown at the bottom of the page, holds true, where we have applied Jensen's inequality with the convex functions $1/x$ and $-\log(x)$ when deriving (a) and (b), respectively, and we have used (55) and (33) when deriving (c). Now, applying the Cauchy-Schwarz inequality, we deduce that (58), shown at the bottom of the page, holds true, where we have applied Jensen's inequality with the convex function x^2 when deriving (a). Again, applying the Cauchy-Schwarz inequality, we deduce that (59), shown at the bottom of the page, holds true, where for the last inequality, we have used the fact that

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

which readily follows from a similar argument as in the derivation of (53). Now, using a largely parallel argument as in Step 1 coupled with Lemma IV.3 $c)$, we conclude that for

$$
\begin{aligned}
-\mathbb{E}\left[\log\frac{d\mu_Y}{d\mu_B}(Y_0^T)\right] + \mathbb{E}\left[\log\frac{d\mu_{Y(\Delta_n)}}{d\mu_{B(\Delta_n)}}(Y(\Delta_n))\right] &= -\mathbb{E}\left[\log e^{-\hat{\rho}_1(Y_0^T)}\right] + \mathbb{E}\left[\log\int \frac{1}{\mathbb{E}[e^{\rho_1(M, Y_0^T)}|Y(\Delta_n), m]}d\mu_M(m)\right]\\
&= \mathbb{E}\left[\log\int \frac{e^{\hat{\rho}_1(Y_0^T) - \rho_2(\Delta_n, m, Y_0^T)}}{\mathbb{E}[e^{\rho_1(M, Y_0^T) - \rho_2(\Delta_n, M, Y_0^T)}|Y(\Delta_n), m]}d\mu_M(m)\right]
\end{aligned} \tag{56}
$$

$$
\begin{aligned}
\mathbb{E}\left[\log\int \frac{e^{\hat{\rho}_1(Y_0^T) - \rho_2(\Delta_n, m, Y_0^T)}}{\mathbb{E}[e^{\rho_1(M, Y_0^T) - \rho_2(\Delta_n, M, Y_0^T)}|Y(\Delta_n), m]}d\mu_M(m)\right] &\\
\overset{(a)}{\leq} \mathbb{E}\left[\log\int e^{\hat{\rho}_1(Y_0^T) - \rho_2(\Delta_n, m, Y_0^T)}\mathbb{E}[e^{-\rho_1(M, Y_0^T) + \rho_2(\Delta_n, M, Y_0^T)}|Y(\Delta_n), m]d\mu_M(m)\right]&\\
\overset{(b)}{\leq} \log\mathbb{E}\left[e^{\hat{\rho}_1(Y_0^T) - \rho_2(\Delta_n, M, Y_0^T)}\mathbb{E}[e^{-\rho_1(M, Y_0^T) + \rho_2(\Delta_n, M, Y_0^T)}|Y(\Delta_n), M]\left(\frac{d\mu_Y}{d\mu_B}(Y_0^T)\right)\Big/\left(\frac{d\mu_{Y|M}}{d\mu_B}(Y_0^T|M)\right)\right]&\\
\overset{(c)}{=} \log\mathbb{E}\left[e^{\rho_1(M, Y_0^T) - \rho_2(\Delta_n, M, Y_0^T)}\mathbb{E}[e^{-\rho_1(M, Y_0^T) + \rho_2(\Delta_n, M, Y_0^T)}|Y(\Delta_n), M]\right]&
\end{aligned} \tag{57}
$$

$$
\begin{aligned}
\mathbb{E}^2\left[e^{\rho_1(M, Y_0^T) - \rho_2(\Delta_n, M, Y_0^T)}\mathbb{E}[e^{-\rho_1(M, Y_0^T) + \rho_2(\Delta_n, M, Y_0^T)}|Y(\Delta_n), M]\right]&\\
\leq \mathbb{E}\left[e^{2\rho_1(M, Y_0^T) - 2\rho_2(\Delta_n, M, Y_0^T)}\right]\mathbb{E}\left[\mathbb{E}^2[e^{-\rho_1(M, Y_0^T) + \rho_2(\Delta_n, M, Y_0^T)}|Y(\Delta_n), M]\right]&\\
\overset{(a)}{\leq} \mathbb{E}\left[e^{2\rho_1(M, Y_0^T) - 2\rho_2(\Delta_n, M, Y_0^T)}\right]\mathbb{E}\left[\mathbb{E}[e^{-2\rho_1(M, Y_0^T) + 2\rho_2(\Delta_n, M, Y_0^T)}|Y(\Delta_n), M]\right]&\\
= \mathbb{E}\left[e^{2\rho_1(M, Y_0^T) - 2\rho_2(\Delta_n, M, Y_0^T)}\right]\mathbb{E}\left[e^{-2\rho_1(M, Y_0^T) + 2\rho_2(\Delta_n, M, Y_0^T)}\right]&
\end{aligned} \tag{58}
$$

$$
\begin{aligned}
\mathbb{E}^2[e^{2\rho_1(M, Y_0^T) - 2\rho_2(\Delta_n, M, Y_0^T)}] &= \mathbb{E}^2[e^{-2\int_0^T (g(s) - \bar{g}_{\Delta_n}(s))dB(s) - \int_0^T (g(s) - \bar{g}_{\Delta_n}(s))^2 ds}]\\
&= \mathbb{E}^2[e^{-2\int_0^T (g(s) - \bar{g}_{\Delta_n}(s))dB(s) - 4\int_0^T (g(s) - \bar{g}_{\Delta_n}(s))^2 ds + 3\int_0^T (g(s) - \bar{g}_{\Delta_n}(s))^2 ds}]\\
&\leq \mathbb{E}[e^{-4\int_0^T (g(s) - \bar{g}_{\Delta_n}(s))dB(s) - 8\int_0^T (g(s) - \bar{g}_{\Delta_n}(s))^2 ds}]\mathbb{E}[e^{6\int_0^T (g(s) - \bar{g}_{\Delta_n}(s))^2 ds}]\\
&\leq \mathbb{E}[e^{6\int_0^T (g(s) - \bar{g}_{\Delta_n}(s))^2 ds}]
\end{aligned} \tag{59}
$$

IEEE TRANSACTIONS ON INFORMATION THEORY, VOL. 68, NO. 2, FEBRUARY 2022

any $0 < \varepsilon < 1$,

$$\mathbb{E}^2[e^{2\rho_1(M,Y_0^T)-2\rho_2(\Delta_n,M,Y_0^T)}] \leq \mathbb{E}[e^{6\int_0^T (g(s)-\bar{g}_{\Delta_n}(s))^2 ds}]$$
$$= 1 + O(\delta_n^{1-\varepsilon}),$$

which further leads to (54).

The theorem then immediately follows from Steps 1, 2 and the fact that

$$0 \leq I(M;Y_0^T) - I(M;Y(\Delta_{T,n})).$$

∎

V. Concluding Remarks

As opposed to the Brownian motion formulation in (1), a continuous-time AWGN channel can be alternatively characterized by the following white noise formulation:

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

where $\{Z(t)\}$ is a white Gaussian noise with flat spectral density 1, and slightly abusing the notation, we have still used X and Y, parameterized by $t \in \mathbb{R}$, to represent the channel input and output, respectively. While there is a comprehensive comparison between these two models in [17], we emphasize here that the Brownian motion formulation enables a more rigorous information-theoretic examination of AWGN channels and empowers the tools in the theory of stochastic calculus that seem to be essential for more quantitative results.

Indeed, we believe the framework and techniques developed in this work can be applied to a more quantitative investigation of sampling a wider range of Gaussian channels possibly with different input constraints and related issues, detailed below. First, it is possible that our approach can be applied to obtain a faster rate of convergence for sampling of a peak-power constrained AWGN channel (see, e.g., [24] and references therein) since, as observed in [21], the peak-power constraint can provide some much needed uniformity properties. Second, a result by Elias [4] has been used to re-derive [6] the somewhat surprising result that for feedback AWGN channels, the Schalkwijk-Kailath scheme yields the decoding error probability that decreases as a second-order exponent in block length. Noticing that Elias' argument in fact used discrete-time MMSE and our treatment essentially quantifies the difference between discrete-time and continuous-time MMSEs, it is worthwhile to investigate whether a rate of convergence result on the decoding error probability can be established for continuous-time feedback AWGN channels. Third, we can also consider sampling of continuous-time additive non-white Gaussian channels. To the best of our knowledge, results in this direction are scarce, but there are a great deal of efforts devoted to discrete-time feedback non-white Gaussian channels (see, e.g., [13] and references therein). Given the recently obtained counterpart results in discrete time [18], it is promising that our treatment can be adapted to such channels, in particular, additive channels with continuous-time auto-regressive and moving average Gaussian noises.

Appendix A
Proof of (61)

Consider the continuous-time AWGN channel (1) with the input as in (2) or (3). Assuming that the channel input $\{X(t)\}$ is integrable over $[0,T]$, $T > 0$, that is,

$$\int_0^T |X(t)| dt < \infty, \quad a.s.. \tag{60}$$

we will in this section prove that

$$\lim_{n \to \infty} I(M;Y(\Delta_{T,n})) = I(M;Y_0^T). \tag{61}$$

As previously mentioned, this result has been implicitly derived in [7], and so we only sketch its proof for brevity.

First of all, an appropriately modified version of the proof of Theorem 1.3 in [7] can be used to establish that

$$I(M;Y_0^T) = \sup I(M;Y(\{t_1, t_2, \ldots, t_n\})),$$

where the supremum is over all possible $n \in \mathbb{N}$ and $t_1, t_2, \ldots, t_n \in [0,T]$. It immediately follows that

$$\limsup_{n \to \infty} I(M;Y(\Delta_{T,n})) \leq I(M;Y_0^T).$$

So, to prove (61), it suffices to prove the other direction:

$$\liminf_{n \to \infty} I(M;Y(\Delta_{T,n})) \geq I(M;Y_0^T).$$

To this end, for each $m \in \mathbb{N}$, we choose a finite subset $\Pi_T^{(m)} \subseteq [0,T]$ such that

$$I(M;Y_0^T) = \lim_{m \to \infty} I(M;Y(\Pi_T^{(m)})).$$

Now, for fixed $\Pi_T^{(m)}$ and for each $n \in \mathbb{N}$, we choose $\hat{\Pi}_{T,n}^{(m)} \subseteq \Delta_{T,n}$ so that $Y(\hat{\Pi}_{T,n}^{(m)})$ is convergent to $Y(\Pi_T^{(m)})$ in distribution. Here we note that the existence of $\{\hat{\Pi}_{T,n}^{(m)}\}$ can be justified by the continuity of $\{Y(t)\}$, which is due to the assumed integrability of $\{X(t)\}$. Then, by Property II at Page 211 of [7], we have

$$\liminf_{n \to \infty} I(M;Y(\hat{\Pi}_{T,n}^{(m)})) \geq I(M;Y(\Pi_T^{(m)})).$$

Taking m to infinity and using the fact that $\hat{\Pi}_{T,n}^{(m)} \subseteq \Delta_{T,n}$ for all n, we have

$$\liminf_{n \to \infty} I(M;Y(\Delta_{T,n})) \geq I(M;Y_0^T),$$

as desired.

Appendix B
Proof of Lemma IV.3

We will only prove $a)$, the proof of $b)$ and $c)$ being largely parallel.

First of all, it can be verified that

$$\mathbb{E}[Z_{max}^2] = -t\mathbb{P}(Z_{max}^2 \geq t)\big|_0^\infty + \int_0^\infty \mathbb{P}(Z_{max}^2 \geq t) dt$$
$$= -\lim_{t \to \infty} t\mathbb{P}(Z_{max}^2 \geq t) + \int_0^\infty \mathbb{P}(Z_{max} \leq -\sqrt{t}) dt$$
$$+ \int_0^\infty \mathbb{P}(Z_{max} \geq \sqrt{t}) dt.$$

Authorized licensed use limited to: Technion Israel Institute of Technology. Downloaded on January 23,2022 at 08:09:04 UTC from IEEE Xplore. Restrictions apply.

Note that

$$\mathbb{P}(Z_{max} \leq -\sqrt{t}) = \mathbb{P}(Z_1 \leq -\sqrt{t}, Z_2 \leq -\sqrt{t}, \ldots, Z_n \leq -\sqrt{t})$$
$$= \mathbb{P}(\sqrt{n}Z_1 \geq \sqrt{nt}, \sqrt{n}Z_2 \geq \sqrt{nt}, \ldots, \sqrt{n}Z_n \geq \sqrt{nt})$$
$$= \left(\int_{\sqrt{nt}}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx \right)^n,$$

where we used the fact that each $\sqrt{n}Z_i$ is a standard normal random variable. Now, using the well-known fact that for any $x > 0$,

$$\frac{2}{\sqrt{\pi}} \int_x^{\infty} e^{-t^2} dt \leq e^{-x^2},$$

we derive

$$\mathbb{P}(Z_{max} \leq -\sqrt{t}) = \left(\int_{\sqrt{nt}}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx \right)^n$$
$$= \left(\frac{1}{2} \frac{2}{\sqrt{\pi}} \int_{\sqrt{\frac{nt}{2}}}^{\infty} e^{-s^2} ds \right)^n$$
$$\leq \left(\frac{1}{2} \right)^n e^{-n^2 t/2}. \tag{62}$$

It then follows that

$$\mathbb{P}(Z_{max} \geq \sqrt{t}) = 1 - \mathbb{P}(Z_{max} \leq \sqrt{t})$$
$$= 1 - \mathbb{P}(\sqrt{n}Z_1 \leq \sqrt{nt}, \sqrt{n}Z_2 \leq \sqrt{nt}, \ldots, \sqrt{n}Z_n \leq \sqrt{nt})$$
$$= 1 - \left(\int_{-\infty}^{\sqrt{nt}} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx \right)^n$$
$$= 1 - \left(1 - \int_{\sqrt{nt}}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx \right)^n.$$

Now, using the well-known fact that for any $x > -1$,

$$\frac{x}{1+x} \leq \log(1+x),$$

and for any $x \leq 0$,

$$1 + x \leq e^x,$$

we derive that for any x with $|x| < 1$,

$$(1-x)^n = e^{n \log(1-x)} \geq e^{-nx/(1-x)} \geq 1 - \frac{nx}{1-x}.$$

It then follows that

$$\mathbb{P}(Z_{max} \geq \sqrt{t}) = 1 - \left(1 - \int_{\sqrt{nt}}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx \right)^n$$

$$\leq 1 - \left(1 - \frac{1}{2} e^{-nt/2} \right)^n$$
$$\leq 1 - \left(1 - \frac{\frac{n}{2} e^{-nt/2}}{1 - \frac{1}{2} e^{-nt/2}} \right)$$
$$= \frac{\frac{n}{2} e^{-nt/2}}{1 - \frac{1}{2} e^{-nt/2}}. \tag{63}$$

Now, using (62) and (63), we have that for any $t > 0$,

$$t\mathbb{P}(Z_{max}^2 \geq t) = t\mathbb{P}(Z_{max} \leq -\sqrt{t}) + t\mathbb{P}(Z_{max} \geq \sqrt{t})$$
$$\leq t \left(\frac{1}{2} \right)^n e^{-n^2 t/2} + t \frac{\frac{n}{2} e^{-nt/2}}{1 - \frac{1}{2} e^{-nt/2}},$$

which immediately implies that

$$\lim_{t \to \infty} t\mathbb{P}(Z_{max}^2 \geq t) = 0. \tag{64}$$

Moreover, using (62), we derive

$$\int_0^{\infty} \mathbb{P}(Z_{max} \leq -\sqrt{t}) dt \leq \left(\frac{1}{2} \right)^n \int_0^{\infty} e^{-n^2 t/2} dt$$
$$= \frac{2}{n^2} \left(\frac{1}{2} \right)^n. \tag{65}$$

And, using (63), we have that for any $0 < a < 1$, (66), shown at the bottom of the page, holds true. Finally, combining (64), (65) and (66), we conclude that for any $0 < a < 1$,

$$\mathbb{E}[Z_{max}^2] = O(n^{-a}),$$

as desired.

Appendix C
Lemma 7.7 of [15] Restated

In this section, Lemma 7.7 of [15] and some conclusions in its proof will be restated using the notations that are more consistent with this paper.

Let $T > 0$ and consider a continuous random process $\xi = \{\xi(t) : 0 \leq t \leq T\}$ having the differential

$$d\xi(t) = C(t, \alpha, \xi) dt + D(t, \xi) dB(t), \quad \xi(0) = 0, \tag{67}$$

where $\alpha = \{\alpha(t) : 0 \leq t \leq T\}$ is a random process independent of the Brownian motion $\{B(t) : 0 \leq t \leq T\}$ whose trajectories belong to some measure space $(A_T, \mathcal{B}_{A_T})$. We shall assume the following conditions are satisfied.

$$\int_0^{\infty} \mathbb{P}(Z_{max} \geq \sqrt{t}) dt = \int_0^{\infty} 1 - \left(1 - \int_{\sqrt{nt}}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx \right)^n dt$$
$$\leq \int_0^{n^{-a}} 1 - \left(1 - \frac{1}{2} e^{-nt/2} \right)^n dt + \int_{n^{-a}}^{\infty} 1 - \left(1 - \frac{1}{2} e^{-nt/2} \right)^n dt$$
$$\leq \int_0^{n^{-a}} dt + \int_{n^{-a}}^{\infty} 1 - \left(1 - \frac{\frac{n}{2} e^{-nt/2}}{1 - \frac{1}{2} e^{-nt/2}} \right) dt$$
$$\leq \int_0^{n^{-a}} dt + \int_{n^{-a}}^{\infty} \frac{\frac{n}{2} e^{-nt/2}}{1 - \frac{1}{2} e^{-nt/2}} dt$$
$$= O(n^{-a}) + O(e^{-n^{1-a}})$$
$$= O(n^{-a}) \tag{66}$$

(A) The random process ξ is a strong solution to the equation (67).

(B) The functionals $C(t, a, x)$ and $D(t, x)$ are non-anticipative, and for each $a \in A_T$ and $x \in C[0, T]$,

$$\int_0^T |C(t, a, x)| dt < \infty, \quad \int_0^T D^2(t, x) dt < \infty.$$

(C) For any x and $\tilde{x}$ from $C[0, T]$,

$$|D(t, x) - D(t, \tilde{x})|^2 \leq L_1 \int_0^t |x(s) - \tilde{x}(s)|^2 dK(s)$$
$$+ L_2 |x(t) - \tilde{x}(t)|^2,$$
$$D^2(t, x) \leq L_1 \int_0^t (1 + x(s))^2 dK(s) + L_2 (1 + x^2(t)),$$
$$D^2(t, x) \geq L_0 > 0,$$

where $K(t)$ is a non-decreasing right continuous function with $0 \leq K(t) \leq 1$ and L_0, L_1, L_2 are positive constants.

(D)

$$\mathbb{P}\left(\int_0^T C^2(t, \alpha, \xi) dt < \infty\right) = \mathbb{P}\left(\int_0^T C^2(t, \alpha, \eta) dt < \infty\right) = 1,$$

where $\eta = \{\eta(t) : 0 \leq t \leq T\}$ is a strong solution to the equation

$$d\eta(t) = D(t, \eta) dB(t), \quad \eta(0) = 0.$$

(E)

$$\int_0^T \mathbb{E}[|C(t, \alpha, \xi)|] dt < \infty, \quad \mathbb{P}\left(\int_0^T \bar{C}^2(t, \xi)\right) = 1,$$

where $\bar{C}(t, \xi) = \mathbb{E}[C(t, \alpha, \xi)|\xi_0^t]$.

Then, we have the following theorem:

Theorem C.1: It holds true that $\mu_\xi \sim \mu_\eta$ and moreover, with probability 1,

$$\frac{d\mu_\eta}{d\mu_\xi}(\xi_0^T) = \exp\left\{-\int_0^T \frac{\bar{C}(t, \xi)}{D^2(t, \xi)} d\xi(t) + \frac{1}{2}\int_0^T \frac{\bar{C}^2(t, \xi)}{D^2(t, \xi)} dt\right\},$$

$$\frac{d\mu_\eta}{d\mu_\xi}(\xi_0^T) = \mathbb{E}\left[\frac{d\mu_\eta}{d\mu_{\xi|\alpha}}(\xi|\alpha) \,\middle|\, \xi_0^T\right].$$

REFERENCES

[1] A. Carlson and P. Crilly, *Communication Systems*, 5th ed. New York, NY, USA: McGraw-Hill, 2009.

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

[3] R. Deng, J. Zhou, and W. Zhang, "Bandlimited communication with one-bit quantization and oversampling: Transceiver design and performance evaluation," *IEEE Trans. Commun.*, vol. 69, no. 2, pp. 845–862, 2021.

[4] P. Elias, "Channel capacity without coding," MIT Res. Lab. Electron., Cambridge, MA, USA, Quart. Progr. Rep., Oct. 1956.

[5] J. Fan, S. Guo, X. Zhou, Y. Ren, G. Ye Li, and X. Chen, "Faster-than-Nyquist signaling: An overview," *IEEE Access*, vol. 5, pp. 1925–1940, 2017.

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

[7] I. Gel'fand and A. 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.

[8] I. V. Girsanov, "On transforming a certain class of stochastic processes by absolutely continuous substitution of measures," *Theory Probab. Appl.*, vol. 5, no. 3, pp. 285–301, Jan. 1960.

[9] D. Guo, S. S. Shitz, and S. Verdú, "Mutual information and minimum mean-square error in Gaussian channels," *IEEE Trans. Inf. Theory*, vol. 51, no. 4, pp. 1261–1282, Apr. 2005.

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

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

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

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

[14] A. Kipnis, Y. C. Eldar, and A. J. Goldsmith, "Analog-to-digital compression: A new paradigm for converting signals to bits," *IEEE Signal Process. Mag.*, vol. 35, no. 3, pp. 16–39, May 2018.

[15] R. Liptser and A. Shiryaev, *Statistics of Random Processes: I: General Theory*, 2nd ed. Berlin, Germany: Springer-Verlag, 2001.

[16] X. Liu, R. Busting, G. Han, and S. S. Shitz, "An elementary proof of a classical information-theoretic formula," in *Proc. ISIT*, Jul. 2019, pp. 1392–1396.

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

[18] T. Liu and G. Han, "Feedback capacity of stationary Gaussian channels further examined," *IEEE Trans. Inf. Theory*, vol. 65, no. 4, pp. 2492–2506, Apr. 2019.

[19] X. Mao, *Stochastic Differential Equations and Applications*. Bristol, U.K.: Horwood, 1997.

[20] P. Morters and Y. Peres, *Brownian Motion*, Cambridge, U.K.: Cambridge Univ. Press, 2010.

[21] T. H. Ng and G. Han, "Relative entropy convergence under Picard's iteration for stochastic differential equations," 2017, *arXiv:1710.05277*.

[22] H. Nyquist, "Certain factors affecting telegraph speed," *Bell Syst. Tech. J.*, vol. 3, no. 2, pp. 324–346, 1924.

[23] B. Oksendal, *Stochastic Differential Equations: An Introduction With Applications*. Berlin, Germany: Springer, 1995.

[24] L. H. Ozarow, A. D. Wyner, and J. Ziv, "Achievable rates for a constrained Gaussian channel," *IEEE Trans. Inf. Theory*, vol. IT-34, no. 3, pp. 365–370, May 1988.

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

[26] J. Schalkwijk, "A coding scheme for additive noise channels with feedback-II: Band-limited signals," *IEEE Trans. Inf. Theory*, vol. IT-12, no. 2, pp. 183–189, Apr. 1966.

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

[28] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, no. 3, pp. 379–423, 1948.

[29] C. E. Shannon, "Communication in the presence of noise," *Proc. Inst. Radio Eng.*, vol. 37, no. 1, pp. 10–21, 1949.

Guangyue Han (Senior Member, IEEE) received the B.S. and M.S. degrees in mathematics from Peking University, China, in 1997 and 2000, respectively, and the Ph.D. degree in mathematics from the University of Notre Dame, USA, in 2004. After three years with the Department of Mathematics, University of British Columbia, Canada, he joined the Department of Mathematics, The University of Hong Kong, China, in 2007. His main research areas are coding and information theory.

Shlomo Shamai (Life Fellow, IEEE) is with the Department of Electrical Engineering, Technion—Israel Institute of Technology, where he is currently a Technion Distinguished Professor, and holds the William Fondiller Chair of Telecommunications. He is an URSI Fellow, a member of the Israeli Academy of Sciences and Humanities, and a Foreign Member of the U.S. National Academy of Engineering. He was a recipient of the 2011 Claude E. Shannon Award, the 2014 Rothschild Prize in Mathematics/Computer Sciences and Engineering, and the 2017 IEEE Richard W. Hamming Medal. He is a co-recipient of the 2018 Third Bell Labs Prize for Shaping the Future of Information and Communications Technology. He is also a recipient of numerous technical and paper awards and recognitions of the IEEE (Donald G. Fink Prize Paper Award), Information Theory, Communications and Signal Processing Societies and EURASIP. He is listed as a Highly Cited Researcher in computer science for the years (2013, 2014, 2015, 2016, 2017, and 2018). He has served as an Associate Editor for the Shannon Theory of the IEEE TRANSACTIONS ON INFORMATION THEORY, and has also served twice on the Board of Governors of the Information Theory Society. He has also served on the Executive Editorial Board for the IEEE TRANSACTIONS ON INFORMATION THEORY, the IEEE Information Theory Society Nominations and Appointments Committee, and the IEEE Information Theory Society, and the Shannon Award Committee.