Optimal Causal Rate-Constrained Sampling for a Class of Continuous Markov Processes

Nian Guo and Victoria Kostina*

February 6, 2020

Abstract

Consider the following communication scenario. An encoder observes a stochastic process and causally decides when and what to transmit about it, under a constraint on bits transmitted per second. A decoder uses the received codewords to causally estimate the process in real time. We aim to find the optimal encoding and decoding policies that minimize the end-to-end estimation mean-square error under the rate constraint. For a class of continuous Markov processes satisfying regularity conditions, we show that the optimal encoding policy transmits a 1-bit codeword once the process innovation passes one of two thresholds. The optimal decoder noiselessly recovers the last sample from the 1-bit codewords and codeword-generating time stamps, and uses it as the running estimate of the current process, until the next codeword arrives. In particular, we show the optimal causal code for the Ornstein-Uhlenbeck process and calculate its distortion-rate function.

*N. Guo and V. Kostina are with the Department of Electrical Engineering, California Institute of Technology, Pasadena, CA, 91125 USA. E-mail: {nguo,vkostina}@caltech.edu

Index terms— Causal lossy source coding, sequential estimation, event-triggered sampling, zero-delay coding.

1 Introduction

1.1 System model and problem setup

Consider the system in Fig. 1. A source outputs a real-valued continuous-time stochastic process $\{X_t\}_{t=0}^T$, with state space $(\mathbb{R}, \mathcal{B}_\mathbb{R})$, where $\mathcal{B}_\mathbb{R}$ is the Borel σ-algebra on $\mathbb{R}$.

Figure 1: System Model. Sampling time τ_i and codeword U_i are chosen by the encoder's sampling and compressing policies, respectively.

An encoder tracks the input process $\{X_t\}_{t=0}^T$ and causally decides to transmit codewords about it at a sequence of stopping times

$$0 \leq \tau_1 \leq \tau_2 \leq \cdots \leq \tau_N \leq T \tag{1}$$

that are decided by a causal sampling policy. Thus, the total number of time stamps N can be random. The time horizon T can either be finite or infinite. At time τ_i, the encoder generates a codeword U_i according to a causal compressing policy, based on the process stopped at τ_i, $\{X_t\}_{t=0}^{\tau_i}$. Then, the codeword U_i is passed to the decoder without delay through a noiseless channel. At time t, $t \in [\tau_i, \tau_{i+1})$, the decoder estimates the input process X_t, yielding $\hat{X}_t$, based on all the received codewords and the codeword-generating time stamps, i.e. $(U_j, \tau_j),\ j = 1, 2, \ldots, i$.

The communication between the encoder and the decoder is subject to a constraint on the long-term average rate,

$$\frac{1}{T}\mathbb{E}\left[\sum_{i=1}^{N}\ell(U_i)\right] \leq R \text{ (bits per sec) } (T < \infty), \tag{2a}$$

$$\limsup_{T\to\infty}\frac{1}{T}\mathbb{E}\left[\sum_{i=1}^{N}\ell(U_i)\right] \leq R, \text{ (bits per sec) } (T = \infty), \tag{2b}$$

where $\ell\colon \mathbb{Z}_+ \to \mathbb{Z}_+$ denotes the length of its argument in bits, $\ell(x) = \lfloor \log_2(x) \rfloor + 1$ for $x > 0$, $\ell(0) = 1$. The *distortion* is measured by the long-term average mean-square error (MSE),

$$\frac{1}{T}\mathbb{E}\left[\int_0^T (X_t - \hat{X}_t)^2 dt\right] \leq d, (T < \infty), \tag{3a}$$

$$\limsup_{T\to\infty}\frac{1}{T}\mathbb{E}\left[\int_0^T (X_t - \hat{X}_t)^2 dt\right] \leq d. \ (T = \infty). \tag{3b}$$

We aim to find the encoding and decoding policies that achieve the optimal tradeoff between the communication rate (2) and the MSE (3).

1.2 The class of processes

Let $\{\mathcal{F}_t\}_{t=0}^T$ be the filtration generated by $\{X_t\}_{t=0}^T$. For τ an almost surely finite stopping time of $\{\mathcal{F}_t\}_{t=0}^T$, past until τ is defined as

$$\mathcal{F}_\tau = \{A \in \{\mathcal{F}_t\}_{t=0}^T : \{\tau \leq t\} \cap A \in \mathcal{F}_t, \forall t \in [0, T]\}. \tag{4}$$

Throughout, we assume that $\{X_t\}_{t=0}^T$ satisfies:

(i) *(Strong Markov property)* $\{X_t\}_{t=0}^T$ satisfies the strong Markov property: $X_{t+\tau}$ is independent of $\mathcal{F}_\tau$ given X_τ, for all $t \in [0, T-\tau]$ and all almost surely finite stopping times $\tau \in [0, T]$.

(ii) *(Continuous paths)* $\{X_t\}_{t=0}^T$ has continuous paths: X_t is almost surely continuous in t.

(iii) *(Mean-square residual error properties)* For all stopping times $\tau \in [0, T]$ and $t \in [\tau, T]$, the mean-square residual error of $\{X_t\}_{t=0}^{T}$, $Y_t = X_t - \mathbb{E}[X_t | X_\tau, \tau]$ satisfies:

(iii-a) Y_t is independent of $\mathcal{F}_\tau$; Y_t is independent of $\{Y_s\}_{s=\tau}^{r}$ given Y_r, for all $r \in [\tau, t]$.

(iii-b) Y_t can be expressed as

$$Y_t = q(t, s) Y_s + R(t, s, \tau), \tag{5}$$

where $s \in [\tau, t]$, $q(t, s)$ is a deterministic function of (t, s), and $R(t, s, \tau)$ is a random variable that may depend on (t, s, τ) and that has an even and quasi-concave pdf. Furthermore, $q(t, t) = 1$, $R(t, t, \tau) = 0$.

We assume that the initial state $X_0 = 0$ at time $\tau_0 = 0$, and that it is known both at the encoder and the decoder. The class of stochastic processes satisfying (i)-(iii) includes linear diffusion processes such as the Wiener process and the Ornstein-Uhlenbeck (OU) process, as well as the Lévy process with even and quasi-concave increments and continuous paths. These processes are widely used in financial mathematics and physics. The parameters $q(t, s)$ and $R(t, s, \tau)$ in (5) for the above three processes are specified in Table 1. Note

Processes	$q(t,s)$	$R(t,s,\tau)$
Wiener	1	W_{t-s}
OU	e^{t-s}	$\frac{\sigma}{\sqrt{2\theta}} e^{-\theta(t-s)} W_{e^{2\theta(t-s)}-1}$
Lévy	1	X_{t-s}

Table 1: $q(t, s)$ and $R(t, s, \tau)$ in (5) for the Wiener process, the OU process and the Lévy process with zero-mean increments. Here, $\{W_t\}_{t\geq 0}$ denotes the Wiener process.

that in all three cases in Table 1, the function $q(t, s)$ and the random variable

$R(t, s, \tau)$ only depend on the time difference $t - s$, but in general they may not be time-homogeneous.

1.3 Context

In wireless sensor networks and network control systems of the Internet of Things, nodes are spatially dispersed, communication between nodes is a limited resource, and delays are undesirable. We study the fundamental limits of the communication scenario in which the transmitting node (the encoder) observes a stochastic process, and wants to communicate it in real-time to the receiving node (the decoder).

Related work includes [1]-[10], where it is assumed that the encoder transmits real-valued samples of the input process and that the communication is subject to a sampling frequency constraint or a transmission cost. The causal sampling and estimation policies that achieve the optimal tradeoff between the sampling frequency and the distortion have been studied for the following *discrete-time processes*: the i.i.d process [1]; the Gauss-Markov process [2]; the partially observed Gauss-Markov process [3]; and, the first-order autoregressive Markov process $X_{t+1} = aX_t + V_t$ driven by an i.i.d. process $\{V_t\}$ with unimodal and even distribution [4][5]. The first-order autoregressive Markov process considered in [4][5] represents a discrete-time counterpart of the continuous-time process in (5) with $q(t, s) = a^{t-s}, R(t, s, \tau) = X_t - a^{t-s}X_s$. Chakravorty and Mahajan [4] showed that a threshold sampling policy with two constant thresholds and an innovation-based filter jointly minimize a discounted cost function consisting of the MSE and a transmission cost in the infinite time horizon. Molin and Hirche [5] proposed an iterative algorithm to find the sampling policy that achieves the minimum of a cost function consisting of a linear combination of the MSE and the transmission cost in the finite time horizon, and showed that the algorithm converges to a two-threshold policy.

The optimal sampling policies for some *continuous-time processes* have also been studied: the finite time-horizon Wiener and OU processes [7]; the infinite time-horizon multidimensional Wiener process [8]; the infinite-time horizon Wiener process [9]; and, the OU processes [10] with channel delay. The optimal causal sampling policies for the Wiener and the OU processes determined in [7]-[10] are threshold sampling policies, whose two thresholds are obtained by solving optimal stopping time problems via Snell's envelope. The proofs in [7]-[10] rely on a conjecture about the form of the MMSE decoding policy, implying that the causal sampling policies in [7]-[10] are optimal with respect to the conjectured decoding policy, rather than the optimal decoding policy. Namely, Rabi et al. [7] conjectured that the MMSE decoding policy under the optimal sampling policy is equal to the MMSE decoding policy under deterministic (process-independent) sampling policies without proof. Nar and Başar [8] arrived at the MMSE decoding policy for the Wiener process by referring to the results in [6], where the stochastic processes considered in [6] are in discrete-time and the increments of the discrete-time process are assumed to have finite support. Yet, the Wiener process is a continuous-time process with Gaussian increments having infinite support. Sun et al. [9] and Ornee and Sun [10] assumed that the decoding policy ignores the implied knowledge when no samples are received at the decoder, neglecting the possible influence of the sampling policy on the decoding policy.

Although the works [1]-[10] did not consider quantization effects, in digital communication systems, real-valued numbers are quantized into bits before transmission. Quantized event-triggered control schemes have been studied for the following systems: discrete-time linear systems with noise [11] and without noise [12]; continuous-time linear time-invariant (LTI) systems without noise

[13][14] and with bounded noise [15]-[17]; partially-observed continuous-time LTI systems without noise [18][19] and with bounded noise [20]. The quantized event-triggered control schemes in [11]-[20] are designed to stabilize the systems. The optimality of the proposed schemes was not considered in [11]-[20]. In our previous work [21], we introduced an information-theoretic framework for studying jointly optimal sampling and quantization policies by considering a long-term average bitrate constraint. We showed that the optimal event-triggered sampling policy for the Wiener process remains a two-threshold policy even under a bitrate constraint, while the optimal deterministic (process-independent) sampling policy is uniform.

1.4 Contribution

In the paper, we leverage the information-theoretic framework of our prior work [21], introduced in the context of the Wiener process, to study the jointly optimal sampling and quantization policies for the wider class of continuous-time processes introduced in Section 1.2. We prove that the optimal sampling policy is a two-threshold policy whether or not quantization is taken into account. We show that the optimal causal compressor is a sign-of-innovation compressor that generates 1-bit codewords representing the sign of the process innovation since the last sample. This surprisingly simple structure is a consequence of both the real-time distortion constraint (3), which penalizes coding delays, and the symmetry of the innovation distribution (iii), which ensures the optimality of the two-threshold sampling policy. Compared to the previous work on sampling of continuous-time processes [7]-[10], our results apply to a wider class of processes, namely, the processes satisfying (i)-(iii) in Section 1.2. Furthermore, we confirm the validity of the conjecture on the MMSE decoding policy in [7][9][10]. To do so, we use a set of tools that differs from that in [7]-[10]: where [7]-[10]

use Snell's envelope to find the optimal sampling policy under the conjecture on the form of the MMSE decoding policy, we apply majorization theory and real induciton to find the jointly optimal sampling and decoding policies. Finally, we show that the optimal causal code for the Ornstein-Uhlenbeck process generates a 1-bit codeword once the process innovation crosses one of the two thresholds, and calculate its distortion-rate function.

1.5 Notation

We denote by $\{X_t\}_{t=s}^{r}$ the portion of the stochastic process within the time interval $[s, r]$, and denote by $\{X_t\}_{t>s}^{r}$ the portion of the stochastic process within the time interval $(s, r]$. For a possibly infinite sequence $x = \{x_1, x_2, \ldots\}$, we write $x^i = \{x_1, x_2, \ldots, x_i\}$ to denote the vector of its first i elements. We use $X \leftarrow Y$ to represent replacing X by Y.

2 Causal frequency-constrained sampling

Before we show the optimal causal code in Section 3, we formulate the causal frequency-constrained sampling problem and find the optimal tradeoff between the sampling frequency and the MSE. In Theorem 1 in Section 2.2 below, we find the form of the optimal sampling policy. We will show in Theorem 3 in Section 3.2 that when coupled with an appropriate compressing policy, the optimal causal sampling policy in Theorem 1 attains the optimal tradeoff between the communication rate and the MSE.

2.1 Causal frequency-constrained code

Allowing the encoder to transmit real-valued samples $U_i = X_{\tau_i}$ instead of the $\mathbb{Z}_+$-valued codewords U_i, and replacing the bitrate constraint (2) by the average sampling frequency constraint

$$\frac{\mathbb{E}[N]}{T} \leq F \text{ } (\textit{samples per sec}), \text{ } (T < \infty), \tag{6a}$$

$$\limsup_{T\to\infty} \frac{\mathbb{E}[N]}{T} \leq F \text{ } (\textit{samples per sec}), \text{ } (T = \infty), \tag{6b}$$

we obtain the problem of *causal frequency-constrained sampling*. Next, we formally define the causal sampling and decoding policies.

Definition 1 ((F, d, T) causal frequency-constrained code)**.** *A time horizon-T causal frequency-constrained code for the stochastic process $\{X_t\}_{t=0}^T$ is a pair of causal sampling and decoding policies, characterized next.*

1. *The causal* sampling *policy, characterized by the $\mathcal{B}_{\mathbb{R}}$-valued process $\{\pi_t\}_{t=0}^T$ adapted to $\{\mathcal{F}_t\}_{t=0}^T$, decides the stopping times* (1)

$$\tau_{i+1} = \inf\{t \geq \tau_i, X_t \notin \pi_t\}, \tag{7}$$

   at which samples are generated.

2. *Given a causal sampling policy, the real-valued samples $\{X_{\tau_j}\}_{j=1}^i$ and sampling time stamps τ^i, the MMSE decoding policy is*

$$\bar{X}_t = \mathbb{E}[X_t | \{X_{\tau_j}\}_{j=1}^i, \tau^i, t < \tau_{i+1}], \text{ } t \in [\tau_i, \tau_{i+1}). \tag{8}$$

In an (F, d, T) code, the average sampling frequency must satisfy (6)*, while the MSE must satisfy*

$$\frac{1}{T}\mathbb{E}\left[\int_0^T (X_t - \bar{X}_t)^2\right] \leq d, (T < \infty) \tag{9a}$$

$$\limsup_{T\to\infty} \frac{1}{T}\mathbb{E}\left[\int_0^T (X_t - \bar{X}_t)^2\right] \leq d, (T = \infty). \tag{9b}$$

Allowing more freedom in designing the decoding policy will not lead to a lower MSE, since (8) is the MMSE estimator. Note that we cannot immediately simplify the expectation in (8) using the strong Markov property of $\{X_t\}_{t=0}^T$ ((i) in Section 1.2) at this point, since the expectation is also conditioned on $t < \tau_{i+1}$. We will show in Corollary 1.1 below that under the optimal causal sampling policy, (8) can indeed be simplified to (14).

In this work, we focus on the causal sampling policies satisfying the following natural assumptions.

(iv) The sampling interval between any two consecutive stopping times, $\tau_{i+1} - \tau_i$, satisfies

$$\mathbb{E}[\tau_{i+1} - \tau_i] < \infty,\ i = 0, 1, \ldots, \tag{10}$$

and the MSE within each interval satisfies

$$\mathbb{E}\left[\int_{\tau_i}^{\tau_{i+1}} (X_t - \bar{X}_t)^2 dt\right] < \infty,\ i = 0, 1, \ldots \tag{11}$$

(v) For all $i = 0, 1, \ldots$, the conditional pdf $f_{\tau_{i+1}|\tau_i}$ exists, and the process π_t is almost surely continuous in t on each of the intervals $[\tau_i, \tau_{i+1})$.

Note that (10) holds trivially if $T < \infty$. Sun et al. [9] and Ornee and Sun [10] also assumed (10) in their analyses of the infinite time horizon problems for the Wiener [9] and the OU [10] processes. We use (11) to obtain a simplified form of the distortion-frequency tradeoff for time-homogeneous processes (see (16) below). Furthermore, (11) allows us to prove (see (15) below) that the optimal sampling intervals $\tau_{i+1} - \tau_i$ form an i.i.d. process. In contrast, the sampling intervals of the causal sampling policy are assumed to form a regenerative process in [9][10]. We use (v) to show that the optimal sampling policy is a symmetric threshold sampling policy in the frequency-constrained setting, and this sampling policy remains optimal in the rate-constrained setting (see the discussion right before Theorem 3 below).

To quantify the tradeoffs between the sampling frequency (6) and the MSE (9), we introduce the distortion-frequency function.

Definition 2 (Distortion-frequency function (DFF))**.** *The DFF for causal frequency-constrained sampling of the process* $\{X_t\}_{t=0}^T$ *is the minimum MSE achievable by causal frequency-constrained codes,*

$$\begin{aligned}\underline{D}(F) \triangleq \inf\{d : \exists\ (F,d,T)\ \textit{causal} \\ \textit{frequency-constrained code satisfying (iv)(v)}\}.\end{aligned} \tag{12}$$

In the causal frequency-constrained sampling scenario, we say a causal sampling policy *optimal* if it achieves the DFF.

2.2 Optimal causal sampling policy

In Theorem 1 below, we show that the optimal sampling policy is a two-threshold policy that is symmetric with respect to the expected value of the process given the last sample and the last sampling time, henceforth referred to as a *symmetric threshold policy.* In Theorem 2, we show a simplified form of the policy for time-homogeneous processes.

Theorem 1. *The optimal causal sampling policy that achieves the DFF* (12) *in either finite or infinite time horizon for a class of continuous Markov processes satisfying assumptions (i)-(iii) in Section 1.2 is a symmetric threshold sampling policy of the form*

$$\begin{aligned}\tau_{i+1} = \inf\{t \geq \tau_i : & X_t - \mathbb{E}[X_t|X_{\tau_i}, \tau_i] \\ & \notin (-a(t,\tau_i,i), a(t,\tau_i,i))\},\end{aligned} \tag{13}$$

where the threshold a *is a non-negative deterministic function of* (t, τ_i, i)*.*

Proof. [26, Appendix A]. □

Theorem 1 shows that the optimal sampling policy is found within a much smaller set of sampling policies than that allowed in Definition 2: each set of π_t is an interval symmetric about $\mathbb{E}[X_t|X_{\tau_i}, \tau_i]$ that depends on $\{X_t\}_{t=0}^T$ only through the last sampling time and the number of samples taken until t. Using the form of the sampling policy (13), we show that the MMSE decoding policy (8) simplifies as follows.

Corollary 1.1. *In the setting of Theorem 1, the MMSE decoding policy that achieves the DFF is*

$$\bar{X}_t = \mathbb{E}[X_t|X_{\tau_i}, \tau_i],\ \ t \in [\tau_i, \tau_{i+1}). \tag{14}$$

Proof. [26, Appendix B]. □

Note that the expectation in (14) can be calculated at the decoder even without the knowledge of the sampling policy, whereas the expectation in (8) depends on the sampling policy at the encoder through the conditioning on the event that the next sample has not been taken yet, i.e. $t < \tau_{i+1}$. Corollary 1.1 confirms the conjecture in [7][9][10] on the form of the MMSE decoding policy.

Corollary 1.2. *In the setting of Theorem 1, the causal sampling policy that achieves the DFF satisfies (6) with equality.*

Proof. [26, Appendix C]. □

Corollary 1.2 indicates that the inequality in the sampling frequency constraint (6) can be simplified to an equality.

Definition 3 (time-homogeneous process)**.** *The process $\{X_t\}_{t=0}^T$ is called* time-homogeneous*, if for a stopping time $\tau \in [0, T]$ and a constant $s \in [0, T - \tau]$, $X_{s+\tau} - \mathbb{E}[X_{s+\tau}|X_\tau]$ follows a distribution that only depends on s.*

Theorem 2. *In the infinite time horizon, the optimal causal sampling policy that achieves the DFF for time-homogeneous continuous Markov processes satisfying assumptions (i)-(iii) in Section 1.2 is a symmetric threshold sampling policy of the form*

$$\begin{aligned}\tau_{i+1} = \inf\{t \geq \tau_i : & X_t - \mathbb{E}[X_t|X_{\tau_i}, \tau_i] \\ & \notin (-a'(t-\tau_i), a'(t-\tau_i))\},\end{aligned} \tag{15}$$

where the threshold a' is a non-negative deterministic function of $t - \tau_i$. The optimal thresholds of (15) are the solution to the following optimization problem,

$$\underline{D}(F) = \min_{\substack{\{a'(t)\}_{t\geq 0}: \\ \mathbb{E}[\tau_1]=\frac{1}{F}}} \frac{\mathbb{E}\left[\int_0^{\tau_1}(X_t - \mathbb{E}[X_t]^2)dt\right]}{\mathbb{E}[\tau_1]}. \tag{16}$$

Proof. [26, Appendix D]. ◻

Remark 1. *In the setting of Theorem 2, the sampling intervals $\tau_{i+1} - \tau_i$, $i = 0, 1, \ldots$ under a symmetric threshold sampling policy of the form (15) are i.i.d.*

Theorem 2 shows that the optimal sampling policy in Theorem 1 can be further simplified for time-homogeneous processes in the infinite time horizon. As a consequence of time homogeneity, thresholds in (15) only depend on the elapsed time from the last sampling time. In contrast, the thresholds in (13) depend on the last sampling time as well.

3 Causal rate-constrained sampling

In this section, we formally introduce the causal rate-constrained sampling problem, and leverage Theorem 1 in Section 2.2 to find the causal code that achieves the optimal tradeoff between the communication rate and the MSE.

3.1 Causal rate-constrained code

We formally define encoding and decoding policies, and define a distortion-rate function (DRF) to describe the tradeoffs between (2) and (3).

Definition 4 ((R, d, T) causal rate-constrained codes)**.** *A time horizon-T causal rate-constrained code for the stochastic process $\{X_t\}_{t=0}^{T}$ is a pair of encoding and decoding policies. The encoding policy consists of a causal sampling policy and a causal compressing policy.*

1. *The causal sampling policy, defined in Definition 1-1., decides the stopping times (1) at which codewords are generated.*

2. *The causal compressing policy, characterized by the $\mathbb{Z}_+$-valued process $\{f_t\}_{t=0}^{T}$ adapted to $\{\mathcal{F}_t\}_{t=0}^{T}$, decides the codeword to transmit at time τ_i,*

$$U_i = f_{\tau_i}. \tag{17}$$

Given an encoding policy, the MMSE decoding *policy uses the received codewords and codeword-generating time stamps to estimate the process,*

$$\hat{X}_t = \mathbb{E}[X_t | U^i, \tau^i, t < \tau_{i+1}], \ \ t \in [\tau_i, \tau_{i+1}). \tag{18}$$

In an (R, d, T) code, the lengths of the codewords must satisfy the average communication rate constraint R bits per sec in (2), while the MSE must satisfy (3).

Allowing more freedom in designing the decoding policy will not lead to a lower MSE, because (18) is the MMSE estimator.

Definition 5 (Distortion-rate function (DRF))**.** *The DRF for causal rate-constrained sampling of the process $\{X_t\}_{t=0}^{T}$ is the minimum MSE achievable* *by causal rate-R codes:*

$$D(R) \triangleq \inf\{d : \exists\ (R, d, T)\ \textit{causal}\ \textit{rate-constrained code satisfying (iv), (v)}\}. \tag{19}$$

We call a causal code *optimal* if it achieves the DRF.

3.2 Optimal causal codes

We proceed to show that the sampling policies in Theorem 1 remain optimal in the scenario of the rate-constrained sampling. Towards that end, we introduce a class of causal codes, namely, the sign-of-innovation (SOI) codes. We prove that an SOI code achieves the DRF in Definition 5 as long as the process satisfies the assumptions (i)-(iii) in Section 1.2.

Definition 6 (A Sign-of-innovation (SOI) code)**.** *The SOI code for a continuous-path process* $\{X_t\}_{t=0}^{T}$ *consists of an encoding and a decoding policy. Given a symmetric threshold sampling policy in (13) that satisfies (v), at each stopping time* τ_i*,* $i = 1, 2, \ldots,$ *the SOI encoding policy generates a* 1*-bit codeword*

$$U_i = \begin{cases} 1 & \textit{if} \quad X_{\tau_i} - \mathbb{E}[X_{\tau_i} | X_{\tau_{i-1}}, \tau_{i-1}] = a(\tau_i, \tau_{i-1}, i-1) \\ 0 & \textit{if} \quad X_{\tau_i} - \mathbb{E}[X_{\tau_i} | X_{\tau_{i-1}}, \tau_{i-1}] = -a(\tau_i, \tau_{i-1}, i-1). \end{cases} \tag{20}$$

At time τ_i*, the MMSE decoding policy noiselessly recovers* X_{τ_i}*,* $i = 1, 2, \ldots$ *via the received codewords* U^i*,*

$$X_{\tau_i} = (2U_i - 1)a(\tau_i, \tau_{i-1}, i-1) + \mathbb{E}[X_{\tau_i} | X_{\tau_{i-1}}, \tau_{i-1}], \tag{21}$$

and uses (14) as the estimate of X_t *until* U_{i+1} *arrives.*

Note that under (v), the continuous-path process is guaranteed to hit one of the boundaries of the symmetric set (13) with equality, implying that the 1-bit codeword in (20) together with the recovered samples $\{X_{\tau_j}\}_{j=1}^{i-1}$ suffices to recover $X_{\tau_i}, i = 1, 2, \ldots$ noiselessly at the decoder. We conjecture that the continuity of the optimal threshold $a(t, \tau_i, i)$ in (v) holds for the processes with continuous paths ((ii) in Section 1.2). Note that for the Wiener and the OU processes, $a(t, \tau_i, i)$ is a constant, and (v) is satisfied trivially.

Theorem 3. *For a process $\{X_t\}_{t=0}^{T}$ satisfying assumptions (i)-(iii) in Section 1.2, the SOI code, whose stopping times are decided by the optimal symmetric threshold sampling policy (13) of $\{X_t\}_{t=0}^{T}$ with average sampling frequency (6) $F = R$, is the optimal causal code that achieves the DRF (19).*

Proof. [26, Appendix E]. □

Theorem 3 illuminates the working principle of the optimal causal code for the stochastic processes considered in Section 1.2: The encoder transmits a 1-bit codeword representing the sign of the process innovation as soon as the innovation crosses one of the two symmetric thresholds. To achieve the DRF (19), the optimal causal code uses the minimum compression rate (1 bit per codeword) in exchange for the maximum average sampling frequency R.

Theorem 3 shows that the optimal codeword-generating times are the sampling times of the optimal causal sampling policy that satisfies piecewise continuity (v). Furthermore, the optimal decoding policy only depends on the thresholds of the sampling policy and the sampling time stamps. Thus, finding the optimal causal code is simplified to finding the optimal causal sampling policy.

3.3 Rate-constrained sampling of the OU process

Using Theorem 3 and (16), we can easily find the optimal causal code and its corresponding DRF for the OU process by finding the thresholds of the optimal causal sampling policy. The OU process is the solution to the following SDE:

$$dX_t = \theta(\mu - X_t)dt + \sigma dW_t, \tag{22}$$

where μ, θ, σ are positive constants, and W_t is the Wiener process. The OU process satisfies the conditions in Section 1.2. Under the assumption (iv) in Section 2.1 and the assumption that the sampling intervals form a regenerative process, Ornee and Sun [10] found the optimal sampling policy for the OU process in the infinite horizon by forming an optimal stopping problem. They solved the optimal stopping problem via the Snell's envelope which requires solving an SDE. We provide an easier method to find the optimal sampling policy for the OU process in [26, Appendix F]. We also show via Theorem 3 that the policy remains optimal when bitrate constraints are present.

Denote

$$R_1(v^2) \triangleq \frac{v^2}{\sigma^2} {}_2F_2\left(1, 1; \frac{3}{2}, 2; \frac{\theta}{\sigma^2}v^2\right), \tag{23}$$

$$R_2(v^2) \triangleq -\frac{v^2}{2\theta} + \frac{\sigma^2}{2\theta} R_1(v^2), \tag{24}$$

where ${}_2F_2$ is a generalized hypergeometric function.

Proposition 1. *For causal coding of the Ornstein-Uhlenbeck process, the optimal causal sampling policy is the symmetric threshold sampling policy given by*

$$\tau_{i+1} = \inf \left\{ t \geq \tau_i : |X_t - \mathbb{E}[X_t | X_{\tau_i}, \tau_i]| \geq \sqrt{R_1^{-1}\left(\frac{1}{R}\right)} \right\}, \tag{25}$$

The corresponding SOI code achieves the DRF, given by

$$D(R) = R \cdot R_2\left(R_1^{-1}\left(\frac{1}{R}\right)\right). \tag{26}$$

Proof. [26, Appendix F]. □

References

[1] O. C. Imer and T. Başar , "Optimal estimation with limited measurements," *International Journal of Systems Control and Communications*, vol. 2, no. 1-3, pp. 5-29, Jan. 2010.

[2] G. M. Lipsa and N. C. Martins, "Remote state estimation with communication costs for first-order LTI systems," in *IEEE Transactions on Automatic Control*, vol. 56, no. 9, pp. 2013-2025, Sept. 2011.

[3] J. Wu, Q. Jia, K. H. Johansson and L. Shi, "Event-based sensor data scheduling: trade-off between communication rate and estimation quality," in *IEEE Transactions on Automatic Control*, vol. 58, no. 4, pp. 1041-1046, Apr. 2013.

[4] J. Chakravorty and A. Mahajan, "Fundamental limits of remote estimation of autoregressive markov processes under communication constraints," in *IEEE Transactions on Automatic Control*, vol. 62, no. 3, pp. 1109-1124, Mar. 2017.

[5] A. Molin and S. Hirche, "Event-triggered state estimation: an iterative algorithm and optimality properties," in *IEEE Transactions on Automatic Control*, vol. 62, no. 11, pp. 5939-5946, Nov. 2017.

[6] A. Nayyar, T. Başar, D. Teneketzis, and V. V. Veeravalli, "Optimal strategies for communication and remote estimation with an energy harvesting sensor," *IEEE Transactions on Automatic Control*, vol. 58, no. 9, pp. 2246-2260, Sept. 2013.

[7] M. Rabi and G. V. Moustakides, and J. S. Baras, "Adaptive sampling for linear state estimation," in *SIAM Journal on Control and Optimization*, vol. 50, no. 2, pp. 672-702, Mar. 2012.

[8] K. Nar and T. Başar, "Sampling multidimensional Wiener processes," *53rd IEEE Conference on Decision and Control*, Los Angeles, CA, USA, pp. 3426-3431, Dec. 2014.

[9] Y. Sun, Y. Polyanskiy and E. Uysal-Biyikoglu, "Remote estimation of the Wiener process over a channel with random delay," *2017 IEEE International Symposium on Information Theory*, Aachen, Germany, pp. 321-325, Jun. 2017.

[10] T. Z. Ornee and Y. Sun, " Sampling for remote estimation through queues: age of information and beyond", *17-th International Symposium on Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks*, Avignon, France, Jun. 2019.

[11] A. Khina, Y. Nakahira, Y. Su and B. Hassibi, "Algorithms for optimal control with fixed-rate feedback," *2017 IEEE 56th Annual Conference on Decision and Control*, Melbourne, VIC, Australia, pp. 6015-6020, Dec. 2017.

[12] S. Yoshikawa, K. Kobayashi and Y. Yamashita, "Quantized event-triggered control of discrete-time linear systems with switching triggering conditions," *2017 56th Annual Conference of the Society of Instrument and Control Engineers of Japan*, Kanazawa, pp. 313-316, Sep. 2017.

[13] J. Pearson, J. P. Hespanha and D. Liberzon, "Control with minimal cost-per-symbol encoding and quasi-optimality of event-based encoders," in *IEEE Transactions on Automatic Control*, vol. 62, no. 5, pp. 2286-2301, May 2017.

[14] M. J. Khojasteh, P. Tallapragada, J. Cortés and M. Franceschetti, "The value of timing information in event-triggered control," in *IEEE Transactions on Automatic Control*, May 2019.

[15] D. Lehmann and J. Lunze, “Event-based control using quantized state information,” *IFAC Proceedings Volumes*, vol. 43, no. 19, pp.1-6, Sep. 2010.

[16] Q. Ling, “Periodic event-triggered quantization policy design for a scalar LTI system with i.i.d. feedback dropouts,” in *IEEE Transactions on Automatic Control*, vol. 64, no. 1, pp. 343-350, Jan. 2019.

[17] P. Tallapragada and J. Cortés, “Event-triggered stabilization of linear systems under bounded bit rates,” in *IEEE Transactions on Automatic Control*, vol. 61, no. 6, pp. 1575-1589, Jun. 2016.

[18] A. Tanwani, C. Prieurand M. Fiacchini, “Observer-based feedback stabilization of linear systems with event-triggered sampling and dynamic quantization,” in *Systems and Control Letters*, vol. 94, pp. 46-56, Aug. 2016.

[19] E. Kofman and J. H. Braslavsky, “Level crossing sampling in feedback stabilization under data-rate constraints,” in *Proceedings of the 45th IEEE Conference on Decision and Control*, San Diego, CA, USA, pp. 4423-4428, Dec. 2006.

[20] M. Abdelrahim, V. S. Dolk and W. P. M. H. Heemels, “Input-to-state stabilizing event-triggered control for linear systems with output quantization,” *2016 IEEE 55th Conference on Decision and Control*, Las Vegas, NV, USA, pp. 483-488, Dec. 2016.

[21] N. Guo and V. Kostina, ”Optimal causal rate-constrained sampling of the Wiener process,” 2019 57th Annual Allerton Conference on Communication, Control, and Computing (Allerton), Monticello, IL, USA, pp. 1090-1097, Sep. 2019.

[22] E. Çinlar, *Probability and Stochastics*. Springer-Verlag New York, 2014.

[23] B. Hajek, K. Mitzel, and S. Yang, "Paging and registration in cellular networks: Jointly optimal policies and an iterative algorithm", *IEEE Transactions on Information Theory*, vol. 54, no. 2, pp. 608-622, Feb. 2008.

[24] P. L. Clark, "The instructor's guide to real induction", *Mathematics Magazine*, pp. 136-150, Mar. 2019.

[25] M. Brown and S. M. Ross, "Renewal reward processes", *Issue 93 of Technical Report*, Cornell University, Ithaca, New York, 1969.

[26] N. Guo and V. Kostina, "Optimal causal rate-constrained sampling for a class of continuous Markov processes", *arxiv*, Jan. 2020.

A Proof of Theorem 1

First, we introduce Lemmas 1 to 4 below that supply majorization and real induction tools allowing us to prove by induction that there exists a symmetric threshold sampling policy of the form (13) that achieves a tradeoff between the average sampling frequency and the MSE no worse than an arbitrary sampling policy of the form (7). Then, in Proposition 2 below, we prove that the optimal tradeoff between the average sampling frequency and the MSE is achieved within the class of causal sampling policies only taking into account the last sampling time, the number of samples taken until the last sampling time, and the process starting from the last sampling time. Finally, given an arbitrary causal sampling policy in Proposition 2, we use Lemmas 1 to 4 to construct a symmetric threshold sampling policy with the same average sampling frequency as the given policy that leads to the same or lower MSE. To show this, we will use the real induction principles in Lemma 4 to prove that some majorization relation between two sampling policies holds on a continuous time interval, so that we can use properties of majorization in Lemmas 1 to 3 to compare the MSE under the two sampling policies.

Recall that f majorizes g, $f \succ g$, if and only if for any Borel measurable set $\mathcal{B} \in \mathcal{B}_{\mathbb{R}}$ with finite Lebesgue measure, there exits a Borel measurable set $\mathcal{A} \in \mathcal{B}_{\mathbb{R}}$ with the same Lebesgue measure, such that [2]

$$\int_{\mathcal{B}} g(x)dx \leq \int_{\mathcal{A}} f(x)dx. \tag{27}$$

Recall that we call a function $f : \mathbb{R} \to \mathbb{R}$ quasi-concave if for all $x, y \in \mathbb{R}$, $0 \leq \lambda \leq 1$,

$$f(\lambda x + (1-\lambda)y) \geq \min\{f(x), f(y)\}. \tag{28}$$

Lemma 1. *([2, Lemma 2]) Fix two pdfs f_X and g_X, such that f_X is even and quasi-concave and that f_X majorizes g_X, $f_X \succ g_X$. Fix a scalar $c > 0$, and a function $h\colon \mathbb{R} \to [0,1]$, such that*

$$\int_{\mathbb{R}} f_X(x)\mathbb{1}_{(-c,c)}(x)dx = \int_{\mathbb{R}} g_X(x)h(x)dx. \tag{29}$$

Then,

$$f_{X|X\in(-c,c)} \succ g'_X, \tag{30}$$

where the pdfs $f_{X|X\in(-a,a)}$ and g'_X are given by,

$$\begin{aligned} f_{X|X\in(-c,c)}(x) &= \frac{f_X(x)\mathbb{1}_{(-c,c)}(x)}{\int_{\mathbb{R}} f_X(x)\mathbb{1}_{(-c,c)}(x)dx} \\ g'_X(x) &= \frac{g_X(x)h(x)}{\int_{\mathbb{R}} g_X(x)h(x)dx}. \end{aligned} \tag{31}$$

Lemma 2. *([23, Lemma 6.7]) Fix two pdfs f_X and g_X, such that f_X is even and quasi-concave and that f_X majorizes g_X, $f_X \succ g_X$. Fix an even and quasi-concave pdf r_Y. Then, the convolution of f_X and r_Y majorizes the convolution* *of* g_X *and* r_Y*,*

$$f_X * r_Y \succ g_X * r_Y, \tag{32}$$

Furthermore, $f_X * r_Y$ *is even and quasi-concave.*

Lemma 3. *([2, Lemma 4]) Fix two pdfs* f_X *and* g_X *such that* f_X *is even and quasi-concave and that* f_X *majorizes* g_X*,* $f_X \succ g_X$*. Then,*

$$\int_{\mathbb{R}} x^2 f_X(x)dx \leq \int_{\mathbb{R}} (x-y)^2 g_X(x)dx, \ \forall y \in \mathbb{R}. \tag{33}$$

Lemma 4. *(Real induction [24, Thm. 2]) We call a subset* $S \subset [a,b]$*,* $a < b$ *inductive if*

(1) $a \in S$*;*

(2) If $a \leq x < b$*,* $x \in S$*, then there exists* $y > x$ *such that* $[x,y] \in S$*;*

(3) If $a \leq x < b$*,* $[a,x) \in S$*, then* $x \in S$*.*

If a subset $S \subset [a,b]$ *is inductive, then* $S = [a,b]$*.*

Before we state Proposition 2 below, we define the mean-square residual error (MSRE) process $\{Y_t\}_{t=0}^T$ and the residual error estimate (REE) process $\{\bar{Y}_t\}_{t=0}^T$ under an arbitrary causal sampling policy $\{\pi_t\}_{t=0}^T$ of the form (7) with stopping times $\tau_1, \tau_2, \ldots$:

$$Y_t = Y_t(\{\pi_s\}_{s=0}^T) \triangleq X_t - \mathbb{E}[X_t|X_{\tau_i}, \tau_i], t \in [\tau_i, \tau_{i+1}) \tag{34a}$$

$$\bar{Y}_t = \bar{Y}_t(\{\pi_s\}_{s=0}^T) \triangleq \bar{X}_t - \mathbb{E}[X_t|X_{\tau_i}, \tau_i], t \in [\tau_i, \tau_{i+1}), \tag{34b}$$

where $\bar{X}_t = \bar{X}_t(\{\pi_s\}_{s=0}^T)$ is the MMSE decoding policy defined in (8).

Proposition 2. *The optimal causal sampling policy that achieves DFF* (12) *for a continuous Markov process* $\{X_t\}_{t=0}^T$ *satisfying assumptions (i)-(iii) in Section 1.2 is characterized by a* $\mathcal{B}_{\mathbb{R}}$*-valued process* $\{\pi_t\}_{t=0}^T$*, such that at time* t*,*

$t \in [\tau_i, \tau_{i+1})$, $i = 0, 1, \ldots$, the value of π_t is determined by the last sampling time τ_i, the number of samples taken until the last sampling time i, and the process $\{Y_s\}_{s>\tau_i}^t$, i.e.

$$\pi_t = \pi_t(\{Y_s\}_{s>\tau_i}^t, \tau_i, i). \tag{35}$$

The policy $\{\pi_t\}_{t=0}^T$ decides the stopping times using the rule,

$$\tau_{i+1} = \inf\{t \geq \tau_i, Y_t \notin \pi_t\}. \tag{36}$$

Proof. Fix an arbitrary sampling policy $\{\pi_t\}_{t=0}^T$ of the form (7), whose stopping times are $\tau_1, \tau_2, \ldots$, and whose MMSE decoding policy is $\bar{X}_t$. Denote by N_s the number of stopping times of $\{\pi_t\}_{t=0}^T$ in $[0, s]$:

$$N_s = N_s(\{\pi_t\}_{t=0}^T) \triangleq \sum_{i=1}^{\infty} \mathbb{1}_{[0,s]}(\tau_i). \tag{37}$$

The MSE under $\{\pi_t\}_{t=0}^T$ is bounded below by

$$\mathbb{E}\left[\int_0^T (Y_s(\{\pi_t\}_{t=0}^T) - \bar{Y}_s(\{\pi_t\}_{t=0}^T))^2 dt\right] \tag{38a}$$

$$\geq \mathbb{E}\left[\int_0^{\tau_i} (Y_s(\{\pi_t\}_{t=0}^T) - \bar{Y}_s(\{\pi_t\}_{t=0}^T))^2 ds\right] \tag{38b}$$

$$+ \min_{\substack{\{\pi'_t\}_{t=0}^T \in \Pi_T: \\ \pi'_t = \pi_t, \forall t < \tau_i \\ \frac{\mathbb{E}[N'_T]}{T} \leq F}} \mathbb{E}\left[\int_{\tau_i}^T (Y_s(\{\pi'_t\}_{t=0}^T) - \bar{Y}_s(\{\pi'_t\}_{t=0}^T))^2 ds\right], \tag{38c}$$

where in the minimization constraint of (38c), Π_T represents the set of all causal sampling policies in $[0, T]$, $N'_T = N_T(\{\pi'_t\}_{t=0}^T)$. The lower bound in (38) corresponds to running $\{\pi_t\}_{t=0}^T$ until the i-th stopping time, and then switching to $\{\pi'_t\}_{t=\tau_i}^T$, where $\{\pi'_t\}_{t=\tau_i}^T$ is chosen to minimize (38c). Due to the Markov property of $\{Y_t\}_{t=0}^T$ implied by (iii-a), we observe that the knowledge of the stopping time τ_i, the number of samples taken i and $\{Y_s\}_{s=\tau_i}^T$ suffices to specify the minimization problem in (38c). We conclude that there exists a causal sampling policy of the form

$$\pi_t = \pi_t(\{Y_s\}_{s=\tau_i}^{t}, \tau_i, i) \tag{39}$$

that can achieve a tradeoff between distortion (9) and sampling frequency (6) no worse than an arbitrary policy in the more general class (7). Furthermore, since $Y_{\tau_i} = 0$ almost surely for all $i = 0, 1, \ldots$ by its definition in (34a), we can reduce the first entry of π_t in (39) to $\{Y_s\}_{s>\tau_i}^{t}$. □

Using Lemmas 1-4 and Proposition 2, we proceed to prove Theorem 1.

Fix a $\{\pi_t\}_{t=0}^{T}$ of the form (36). We construct a symmetric threshold sampling policy $\{\pi_t^{\text{sym}}\}_{t=0}^{T}$ of the form (13) with stopping times $\tau_1', \tau_2', \ldots$, such that

$$\mathbb{E}[N_T] = \mathbb{E}[N_T'], \tag{40}$$

$$\mathbb{E}\left[\sum_{i=0}^{N_T} \int_{\tau_i}^{\tau_{i+1}} (Y_t - \bar{Y}_t)^2 dt\right] \geq \mathbb{E}\left[\sum_{i=0}^{N_T'} \int_{\tau_i'}^{\tau_{i+1}'} (Y_t' - \bar{Y}_t')^2 dt\right], \tag{41}$$

where $\tau_0 = \tau_0' \triangleq 0$, $\tau_{N+1} = \tau_{N+1}' \triangleq T$, and

$$\begin{aligned} &N_T = N_T(\{\pi_s\}_{s=0}^{T}), N_T' = N_T(\{\pi_s^{\text{sym}}\}_{s=0}^{T}),\\ &Y_t = Y_t(\{\pi_s\}_{s=0}^{T}), Y_t' = Y_t(\{\pi_s^{\text{sym}}\}_{s=0}^{T}),\\ &\bar{Y}_t = Y_t(\{\pi_s\}_{s=0}^{T}), \bar{Y}_t' = Y_t(\{\pi_s^{\text{sym}}\}_{s=0}^{T}). \end{aligned} \tag{42}$$

We further simplify $\bar{Y}_t$ in (41) as

$$\bar{Y}_t = \mathbb{E}[Y_t | \{X_{\tau_j}\}_{j=1}^{i}, \tau^i, t < \tau_{i+1}],\ t \in [\tau_i, \tau_{i+1}) \tag{43a}$$

$$= \mathbb{E}[Y_t | \tau_i, t < \tau_{i+1}],\ t \in [\tau_i, \tau_{i+1}), \tag{43b}$$

where (43a) holds since $\mathbb{E}[X_t|X_{\tau_i},\tau_i] \in \sigma(\{X_{\tau_j}\}_{j=1}^i, \tau^i, t < \tau_{i+1})$; (43b) holds because Y_t is independent of $\{X_{\tau_j}\}_{j=1}^i, \tau^i$ due to (iii-a), and the event $\{t < \tau_{i+1}\}$ is independent of $\{X_{\tau_j}\}_{j=1}^i, \tau^{i-1}$ given τ_i due to (36). Since under $\{\pi_t^{\text{sym}}\}_{t=0}^T$, the event $\{t < \tau'_{i+1}\}$ is also independent of $\{X_{\tau'_j}\}_{j=1}^i, \tau'^{i-1}$ given τ'_i, we can write $\bar{Y}'_t$ as

$$\bar{Y}'_t = \mathbb{E}[Y'_t|\tau'_i, t < \tau'_{i+1}],\ t \in [\tau'_i, \tau'_{i+1}) \tag{44a}$$

$$= 0, \tag{44b}$$

where (44b) holds since Y'_t has an even and quasi-concave pdf due to the assumption (iii-b), and the pdf of Y_t conditioned on $\tau'_i, t < \tau'_{i+1}$ under a symmetric threshold sampling policy of the form (13) is still even and quasi-concave.

Denote by $\text{Supp}(\tau_i)$ the support of τ_i, i.e. for all $s \in \text{Supp}(\tau_i)$, we have $f_{\tau_i}(s) > 0$. Denote by ess $\sup(\tau_i)$ the essential supreme of τ_i, i.e. ess $\sup(\tau_i) \triangleq \sup \text{Supp}(\tau_i)$. We proceed to show how to construct a symmetric threshold sampling policy $\{\pi_t^{\text{sym}}\}_{t=0}^T$ such that for all $i = 0, 1, \ldots$ and for all $s \in \text{Supp}(\tau_i)$, $t \in \text{Supp}(\tau_{i+1})$, $0 \le s \le t \le T$, we have

$$\mathbb{P}[\tau'_{i+1} > t|\tau'_i = s] = \mathbb{P}[\tau_{i+1} > t|\tau_i = s], \tag{45}$$

and for all $t \in (s, \text{ess}\sup(\tau_{i+1}))$,

$$\mathbb{E}\left[(Y_t - \bar{Y}_t)^2|\tau_i = s, \tau_{i+1} > t\right] \ge \mathbb{E}\left[Y_t'^2|\tau'_i = s, \tau'_{i+1} > t\right], \tag{46}$$

Note that $\bar{Y}'_t = 0$ by (44). Observe that it suffices to show (45)–(46) to guarantee that (40)–(41) hold. We notice that the two sequences of stopping times $\{\tau_1, \tau_2, \ldots\}$ and $\{\tau'_1, \tau'_2, \ldots\}$ both satisfy Markov property due to the forms of the causal sampling policies (39) and (13). Indeed, (45) implies (40), because the Markov property of stopping times together with (45) implies that the joint distribution of $\tau_1, \tau_2, \dots$ is equal to the joint distribution of $\tau'_1, \tau'_2, \dots,$ while (45)–(46) imply (41) by the law of total expectation.

We can construct a sequence of thresholds $\{a(r, s, i)\}_{r=s}^{\text{ess sup}(\tau_{i+1})}$ such that for all $s \in \text{Supp}(\tau_i)$, $s \le t \le \text{ess sup}(\tau_{i+1})$,

$$\begin{aligned} &\mathbb{P}[Y'_r \in (-a(r,s,i), a(r,s,i)), \forall r \in [s,t] | \tau_i = s] \\ = &\mathbb{P}[Y_r \in \pi_r, \forall r \in [s,t] | \tau_i = s] \end{aligned} \tag{47}$$

thus (45) follows.

Next, we show (46) holds. We will prove using real induction that the conditions (a)–(c) stated next hold for all $s \in \text{Supp}(\tau_i)$, $t \in [s, \text{ess sup}(\tau_{i+1}))$:

(a) For any Borel measurable set $\mathcal{B} \in \mathcal{B}_{\mathbb{R}}$ with finite Lebesgue measure, there exists a Borel measurable set $\mathcal{A} \in \mathcal{B}_{\mathbb{R}}$ with the same Lebesgue measure, such that

$$\mathbb{P}[Y_t \in \mathcal{B} | \tau_i = s, \tau_{i+1} > t] \le \mathbb{P}[Y'_t \in \mathcal{A} | \tau'_i = s, \tau'_{i+1} > t]. \tag{48}$$

(b) The conditional cdf $\mathbb{P}[Y'_t \le y | \tau'_i = s, \tau'_{i+1} > t]$ is convex for $y < 0$ and is concave for $y > 0$.

(c) For any $y > 0$,

$$\mathbb{P}[Y'_t \in (0, y] | \tau'_i = s, \tau'_{i+1} > t] = \mathbb{P}[Y'_t \in [-y, 0) | \tau'_i = s, \tau'_{i+1} > t]. \tag{49}$$

Before we prove (a)–(c) using real induction, we will prove that if conditions (a)–(c) hold for $s \in \text{Supp}(\tau_i)$, $t \in [s, \text{ess sup}(\tau_{i+1}))$, then the following pdfs exist and satisfy

$$f_{Y'_t | \tau'_i = s, \tau'_{i+1} > t} \succ f_{Y_t | \tau_i = s, \tau_{i+1} > t}, \tag{50}$$

$$f_{Y'_t | \tau'_i = s, \tau'_{i+1} > t} \text{ is even and quasi-concave,} \tag{51}$$

for all $s \in \mathrm{Supp}(\tau_i)$, $t \in (s, \text{ess sup}(\tau_{i+1}))$. The validity of (46) will then follow via an application of Lemma 3 with $f_X \leftarrow f_{Y'_t|\tau'_i=s,\tau'_{i+1}>t}$ and $g_X \leftarrow f_{Y_t|\tau_i=s,\tau_{i+1}>t}$. Note that the reason that we do not directly prove that (50)–(51) hold for $s \in \mathrm{Supp}(\tau_i)$, $t \in [s, \text{ess sup}(\tau_{i+1}))$ is that $Y_s = 0$ almost surely given $\tau'_i = s, \tau'_{i+1} > s$, thus the densities in (50)–(51) do not exist at $t = s$. In contrast, conditions (a)–(c) are stated in terms of the probability measure allowing us to use real induction on a left-closed interval starting from $t = s$.

We proceed to show that the conditional pdfs in (50) exist for all $s \in \mathrm{Supp}(\tau_i), t \in (s, \text{ess sup}(\tau_{i+1}))$. We prove that $f_{Y_t|\tau_i=s,\tau_{i+1}>t}$ exists, and the proof that $f_{Y'_t|\tau'_i=s,\tau'_{i+1}>t}$ exists is similar. Since $R(t, s, s)$ is independent of the event $\{\tau_i = s, \tau_{i+1} > s\}$, implied by assumption (iii-a), we compute $f_{Y_t|\tau_i=s,\tau_{i+1}>s}$ using (5),

$$f_{Y_t|\tau_i=s,\tau_{i+1}>s} = f_{R(t,s,s)}. \tag{52}$$

Thus, $f_{Y_t|\tau_i=s,\tau_{i+1}>s}$ exists since $f_{R(t,s,s)}$ is a valid pdf by assumption (iii-b). To establish that $f_{Y_t|\tau_i=s,\tau_{i+1}>t}(y)$ exists, we compute

$$f_{Y_t|\tau_i=s,\tau_{i+1}>t}(y) = f_{Y_t|\tau_i=s,\tau_{i+1}>s,\tau_{i+1}>t}(y) \tag{53a}$$

$$= \frac{\mathbb{P}[\tau_{i+1} > t|\tau_{i+1} > s, \tau_i = s, Y_t = y] f_{Y_t|\tau_i=s,\tau_{i+1}>s}(y)}{\mathbb{P}[\tau_{i+1} > t|\tau_i = s, \tau_{i+1} > s]}, \tag{53b}$$

where (53a) holds since $\tau_{i+1} > t$ implies $\tau_{i+1} > s$. In (53b), we observe that for all $s \in \mathrm{Supp}(\tau_i)$, $t \in (s, \text{ess sup}(\tau_{i+1}))$, the pdf $f_{Y_t|\tau_{i+1}>s,\tau_i=s}$ exists by (52); the denominator of (53b) is larger than zero. We conclude that the pdf $f_{Y_t|\tau_i=s,\tau_i>t}$ exists for all $s \in \mathrm{Supp}(\tau_i)$, $t \in (s, \text{ess sup}(\tau_{i+1}))$.

We verify that (a)–(c) are equivalent to (50)–(51) for $t \in (s, \text{ess sup}(\tau_{i+1}))$, where (a) is equivalent to (50) due to the definition of majorization (27), and the proved fact that the pdfs $f_{Y'_t|\tau'_i=s,\tau'_{i+1}>t}$ and $f_{Y_t|\tau_i=s,\tau_{i+1}>t}$ exist; (b)–(c) are equivalent to (51) since for $t \in (s, \text{ess sup}(\tau_{i+1}))$, $f_{Y'_t|\tau'_i=s,\tau'_{i+1}>t}$ is quasi-concave if and only if (b) holds, and $f_{Y'_t|\tau'_i=s,\tau'_{i+1}>t}$ is even if and only if (c) holds. Thus, (a)–(c) hold for all $s \in \text{Supp}(\tau_i)$, $t \in [s, \text{ess sup}(\tau_{i+1}))$ implying that (50)–(51) hold for all $s \in \text{Supp}(\tau_i)$, $t \in (s, \text{ess sup}(\tau_{i+1}))$.

Now, we proceed to show that (a)–(c) hold for all $t \in [s, \text{ess sup}(\tau_{i+1}))$, given any $s \in \text{Supp}(\tau_i)$ using real induction. To verify that the condition (1) in Lemma 4 holds, we need to show that (a)–(c) hold for $t = s$. This is trivial since

$$\mathbb{P}[Y'_t = 0|\tau'_i = s, \tau'_{i+1} > t] = \mathbb{P}[Y_t = 0|\tau_i = s, \tau_{i+1} > t] = 1. \tag{54}$$

Next, we show that condition (3) in Lemma 4 holds, that is, assuming that (a)–(c) hold for all $t \in [s, r)$, $r \in (s, \text{ess sup}(\tau_{i+1})]$, we prove that (a)–(c) hold for $t = r$. Equivalently, we show that (50)–(51) hold for $t = r$. Let $\delta \in (0, r - s]$. At time $t = r$, we calculate the left side of (50),

$$f_{Y'_r|\tau'_i=s,\tau'_{i+1}>r}(y) = \lim_{\delta\to 0^+} f_{Y'_r|\tau'_i=s,\tau'_{i+1}>r-\delta,\tau'_{i+1}>r}(y) \tag{55a}$$

$$= \lim_{\delta\to 0^+} \frac{\mathbb{P}[\tau'_{i+1} > r|\tau'_{i+1} > r-\delta, \tau'_i = s, Y'_r = y] f_{Y'_r|\tau'_i=s,\tau'_{i+1}>r-\delta}(y)}{\int_{\mathbb{R}} \mathbb{P}[\tau'_{i+1} > r|\tau'_{i+1} > r-\delta, \tau'_i = s, Y'_r = y] f_{Y'_r|\tau'_i=s,\tau'_{i+1}>r-\delta}(y)dy} \tag{55b}$$

$$= \lim_{\delta\to 0^+} \frac{\mathbb{1}_{(-a(r,s,i),a(r,s,i))}(y) f_{Y'_r|\tau'_i=s,\tau'_{i+1}>r-\delta}(y)}{\int_{\mathbb{R}} \mathbb{1}_{(-a(r,s,i),a(r,s,i))}(y) f_{Y'_r|\tau'_i=s,\tau'_{i+1}>r-\delta}(y)dy}, \tag{55c}$$

where (55a) holds since the event $\tau'_{i+1} > r$ implies the event $\tau'_{i+1} > r - \delta$; the pdf $f_{Y'_r|\tau'_i=s,\tau'_{i+1}>r-\delta}$ in (55b) exists since (53) holds with Y_t, $\tau_i = s$, $\tau_{i+1} > s$, $\tau_{i+1} > t$ replaced by Y'_r, $\tau'_i = s$, $\tau'_{i+1} > s$, $\tau'_{i+1} > r - \delta$, respectively; (55c) holds since

$$\lim_{\delta\to 0^+} \mathbb{P}[\tau'_{i+1} > r|\tau'_{i+1} > r-\delta, \tau'_i = s, Y'_r = y] = \mathbb{1}_{(-a(r,s,i),a(r,s,i))}(y). \tag{56}$$

Similarly, the right side of (50) is equal to

$$\begin{aligned} & f_{Y_r|\tau_i=s,\tau_{i+1}>r}(y) \\ &= \lim_{\delta\to 0^+} \frac{\mathbb{P}[\tau_{i+1}>r|\tau_{i+1}>r-\delta,\tau_i=s,Y_r=y] f_{Y_r|\tau_i=s,\tau_{i+1}>r-\delta}(y)}{\int_{\mathbb{R}} \mathbb{P}[\tau_{i+1}>r|\tau_{i+1}>r-\delta,\tau_i=s,Y_r=y] f_{Y_r|\tau_i=s,\tau_{i+1}>r-\delta}(y)dy}, \end{aligned} \tag{57}$$

where the pdf $f_{Y_r|\tau_i=s,\tau_{i+1}>r-\delta}(y)$ exists since (53) holds with Y_t, $\tau_{i+1}>t$ replaced by Y_r, $\tau_{i+1}>r-\delta$ respectively.

To check that (50) holds at $t=r$, we first prove that $f_{Y'_r|\tau'_i=s,\tau'_{i+1}>r-\delta}$ majorizes $f_{Y_r|\tau_i=s,\tau_{i+1}>r-\delta}$. Note that $R(r,r-\delta,s)$ is independent of $\{Y_s\}_{s=0}^{r-\delta}$ due to assumption (iii-a), and thus is independent of the event $\{\tau'_{i+1}>r-\delta,\tau'_i=s\}$. We obtain Y'_r using (5),

$$f_{Y'_r|\tau'_i=s,\tau'_{i+1}>r-\delta} = f_{q(r,r-\delta)Y'_{r-\delta}|\tau'_i=s,\tau'_{i+1}>r-\delta} * f_{R(r,r-\delta,s)}. \tag{58}$$

By (58) and the inductive hypothesis that (a)–(c) holds for $t\in[s,r)$, the assumptions in Lemma 2 are satisfied with $f_X \leftarrow f_{q(r,r-\delta)Y'_{r-\delta}|\tau'_i=s,\tau'_{i+1}>r-\delta}$, $g_X \leftarrow f_{q(r,r-\delta)Y_{r-\delta}|\tau_i=s,\tau_{i+1}>r-\delta}$, $r_Y \leftarrow f_{R(r,r-\delta,s)}$. We conclude that

$$f_{Y'_r|\tau'_i=s,\tau'_{i+1}>r-\delta} \succ f_{Y_r|\tau_i=s,\tau_{i+1}>r-\delta}, \tag{59}$$

$$f_{Y'_r|\tau'_i=s,\tau'_{i+1}>r-\delta} \text{ is even and quasi-concave.} \tag{60}$$

Due to (60) and the fact that the indicator function in (55c) is over an interval symmetric about zero, we conclude (51) holds for $t=r$. By (45), (59) and (60), the assumptions in Lemma 1 are satisfied with $f_X \leftarrow f_{Y'_r|\tau'_i=s,\tau'_{i+1}>r-\delta}$, $g_X \leftarrow f_{Y_r|\tau_i=s,\tau_{i+1}>r-\delta}$, $f_{X|X\in(-c,c)} \leftarrow f_{Y'_r|\tau'_i=s,\tau'_{i+1}>r}$, and $g'_X \leftarrow f_{Y_r|\tau_i=s,\tau_{i+1}>r}$, $c \leftarrow a(r,s,i)$, $h \leftarrow \mathbb{P}[\tau_{i+1}>r|\tau_{i+1}>r-\delta,\tau_i=s,Y_t=y]$. Thus, we conclude that (50) holds for $t=r$. Therefore, (50)–(51) hold for $t=r$, i.e. (a)–(c) hold for $t=r$.

To prove that the condition (2) in Lemma 4 holds, we assume (a)–(c) hold for $t = r$, and prove that the following holds:

$$\lim_{\delta \to 0^+} f_{Y'_{r+\delta}|\tau'_i=s,\tau'_{i+1}>r+\delta} \succ \lim_{\delta \to 0^+} f_{Y_{r+\delta}|\tau_i=s,\tau_{i+1}>r+\delta}, \tag{61a}$$

$$\lim_{\delta \to 0^+} f_{Y'_{r+\delta}|\tau'_i=s,\tau'_{i+1}>r+\delta} \text{ is even and quasi-concave.} \tag{61b}$$

The right and the left sides of (61a) are equal to (55c) and (57) respectively with r replaced by $r + \delta$. It it easy to see that (58)–(60) and the assumptions in Lemma 1 hold with r replaced by $r + \delta$. Thus, we conclude that (61) holds.

Using the real induction in Lemma 4, we have shown that (a)–(c) hold for all $t \in [s, \text{ess sup}(\tau_{i+1}))$ given any $s \in \text{Supp}(\tau_i)$. Thus, (50)–(51) hold for all $s \in \text{Supp}(\tau_i)$, $t \in (s, \text{ess sup}(\tau_{i+1}))$.

B Proof of Corollary 1.1

Under the symmetric threshold sampling policy (13), the MMSE decoding policy in (8) can be expanded as, for $\tau_i \le t < \tau_{i+1}$,

$$\bar{X}_t = \mathbb{E}[X_t | \{X_{\tau_j}\}_{j=1}^i, \tau^i, t < \tau_{i+1}] \tag{62a}$$

$$= \bar{Y}_t + \mathbb{E}[X_t | X_{\tau_i}, \tau_i] \tag{62b}$$

$$= \mathbb{E}[X_t | X_{\tau_i}, \tau_i]. \tag{62c}$$

where $\bar{Y}_t$ is defined in (34b) and is equal to zero due to (44).

C Proof of Corollary 1.2

Given any causal sampling policy such that (6) is satisfied with strict inequality, we construct a causal sampling policy that satisfies (6) with equality and leads to the same or a lower MSE.

Given an arbitrary symmetric threshold sampling policy $\{\pi_t^{\text{sym}}\}_{t=0}^T$ of the form (13), we denote by $N_t = N_t(\{\pi_t^{\text{sym}}\}_{t=0}^T)$ the number of samples taken in $[0,t]$, and denote by $\tau_1, \tau_2, \ldots$ the stopping times. Let t', $t' \in (0,T)$ be a dummy deterministic time. We decompose the MSE under $\{\pi_t^{\text{sym}}\}_{t=0}^T$ as

$$\mathbb{E}\left[\sum_{i=0}^{N_{t'}-1}\int_{\tau_i}^{\tau_{i+1}}(X_t - \mathbb{E}[X_t|X_{\tau_i},\tau_i])^2 dt\right] \tag{63a}$$

$$+\mathbb{E}\left[\int_{\tau_{N_{t'}}}^{t'}(X_t - \mathbb{E}[X_t|X_{\tau_{N_{t'}}},\tau_{N_{t'}}])^2 dt\right] \tag{63b}$$

$$+\mathbb{E}\left[\int_{t'}^{\tau_{N_{t'}+1}}(X_t - \mathbb{E}[X_t|X_{\tau_{N_{t'}}},\tau_{N_{t'}}])^2 dt\right] \tag{63c}$$

$$+\mathbb{E}\left[\sum_{i=N_{t'}+1}^{N_T}\int_{\tau_i}^{\tau_{i+1}}(X_t - \mathbb{E}[X_t|X_{\tau_i},\tau_i])^2 dt\right], \tag{63d}$$

where $\tau_{N_T+1} \triangleq T$.

Given $\{\pi_t^{\text{sym}}\}_{t=0}^T$, we construct $\{\pi_t'\}_{t=0}^T$ by inserting an extra deterministic sampling time t'. The resultant MSE is the same as (63) with (63c) replaced by

$$\mathbb{E}\left[\int_{t'}^{\tau_{N_{t'}+1}}(X_t - \mathbb{E}[X_t|X_{t'}])^2 dt\right], \tag{64}$$

since a sample is taken at time t' under $\{\pi'\}_{t=0}^T$. Since the following holds,

$$\sigma(X_{\tau_{N_{t'}}},\tau_{N_{t'}}) \subseteq \sigma(\mathcal{F}_{t'}) \tag{65a}$$

$$\mathbb{E}[X_t|\mathcal{F}_{t'}] = \mathbb{E}[X_t|X_{t'}], \tag{65b}$$

where (65b) is due to the strong Markov process (i) in Section 1.2, we conclude that (63c) $\geq$ (64).

Thus, by introducing extra sampling times, we can achieve the same or a lower MSE. We can express the difference between the frequency constraint F

and the average sampling frequency under the given sampling policy as

$$F - \frac{\mathbb{E}[N_T]}{T} = I + D, \tag{66}$$

where $I \in \mathbb{N}$ represents the non-negative integer part, and $D \in (0, 1)$ represents the decimal part. By introducing I different deterministic sampling times, we can compensate the integer part I. By introducing a random sampling time stamp t with probability D to sample and probability $1 - D$ not to sample, we can compensate the decimal part. Therefore, for any sampling policy whose average sampling frequency is strictly less than F, we can always construct a sampling policy that achieves the maximum sampling frequency F and leads to the same or a lower MSE.

D Proof of Theorem 2

First, we introduce Lemma 5 stated next that will be helpful in proving (16). Second, we prove that the symmetric threshold sampling policy (13) in Theorem 1 can be reduced to (15) in the setting of Theorem 2, i.e. $\{X_t\}_{t\geq 0}$ has time-homogeneous property in Definition 3 and $T = \infty$. Then, we show that Remark 1 holds, and prove that (16) holds using Lemma 5.

Lemma 5. *(e.g. [25, Proposition 1(ii)]) Let* $N_t = \sum_{i=0}^{\infty} \mathbb{1}_{[0,t]}\left(\sum_{k=0}^{i} Z_k\right)$. *Suppose that* $Z_0, Z_1, \ldots$ *are i.i.d.,* $R_0, R_1, \ldots$ *are i.i.d rewards, and* $Y_t = \sum_{i=0}^{N_t} R_i$ *is the renewal reward process. If* $0 < \mathbb{E}[Z_i] < \infty$, $\mathbb{E}[|R_i|] < \infty$, *then*

$$\lim_{T\to\infty} \frac{\mathbb{E}[Y_T]}{T} = \frac{\mathbb{E}[R_0]}{\mathbb{E}[Z_0]}. \tag{67}$$

Since the stochastic process considered in Theorem 2 is infinitely long, we use the DFF in the infinite time horizon:

$$\underline{D}^{\infty}(F) = \inf_{\substack{\{\pi_t\}_{t\geq 0}\in\Pi: \\ (6b)}} \limsup_{T\to\infty} \frac{1}{T}\mathbb{E}\left[\int_0^T (Y_t - \bar{Y}_t)^2\right], \tag{68}$$

where Π is the set of all causal sampling policies of the form (13) satisfying (iv) and (v) in Section 1.2 over the infinite time horizon. We will prove that the causal sampling policy that achieves $\underline{D}^{\infty}(F)$ for time-homogeneous continuous Markov processes satisfying assumptions (i)–(iii) is of the form (15).

Note that for any stopping time τ, and any $t \geq \tau$, the following holds:

$$\{Y_t\}_{t\geq\tau} \text{ and } \{Y_{t-\tau}\}_{t-\tau\geq 0} \text{ are of the same distribution,} \tag{69}$$

$$\{Y_t\}_{t\geq\tau} \text{ is independent of } \{Y_t\}_{t=0}^{\tau}, \tag{70}$$

where (69) is due to the time-homogeneity of $\{X_t\}_{t\geq 0}$ in Definition 3, and (70) is due to (iii-a) in Section 1.2. Using (69)–(70) and assumption (iv), we will prove that the sampling policy that achieves the $\underline{D}^{\infty}(F)$ is of the form (15).

Given an arbitrary sampling policy $\{\pi_t\}_{t\geq 0}$ of the form (13), we define

$$\begin{aligned} Y_t &= Y_t(\{\pi_s\}_{s\geq 0}), \\ \bar{Y}_t &= \bar{Y}_t(\{\pi_s\}_{s\geq 0}). \end{aligned} \tag{71}$$

Denote by $\tau_1, \tau_2, \ldots$ the stopping times of $\{\pi_t\}_{t\geq 0}$. Assume that the sampling policy that achieves $\underline{D}^{\infty}(F)$ (68) is $\left\{\pi_t^{(a)}\right\}_{t\geq 0}$.

$$\underline{D}^{\infty}(F) \tag{72a}$$

$$= \inf_{\substack{\{\pi_t\}_{t\geq 0}\in\Pi: \\ \pi_t=\pi_t^{(a)}, t\leq\tau_i, \\ (6b)}} \limsup_{T\to\infty} \frac{1}{T}\mathbb{E}\left[\int_{\tau_i}^T (Y_t - \bar{Y}_t)^2 dt\right] \tag{72b}$$

$$= \inf_{\substack{\{\pi_t\}_{t\geq 0}\in\Pi: \\ (6b)}} \limsup_{T\to\infty} \frac{1}{T}\mathbb{E}\left[\int_0^{T-\tau_i} (Y_t - \bar{Y}_t)^2 dt\right] \tag{72c}$$

$$= \underline{D}^{\infty}(F), \tag{72d}$$

where (72b) is due to assumption (iv); (72c) is due to (69); the equality in (72d) is achieved since (72c) is upper-bounded by (72d) and is equal to (72a) simultaneously. Suppose that the sampling policies that achieve (72b)–(72c) are $\{\pi_t^{(b)}\}_{t\geq 0}$ and $\{\pi_t^{(c)}\}_{t\geq 0}$, respectively. From (72a) and (72b), we observe that

$$\left\{\pi_t^{(a)}\right\}_{t\geq\tau_i} = \left\{\pi_t^{(b)}\right\}_{t\geq\tau_i}, \ i = 0, 1, \ldots \tag{73}$$

We prove that under sampling policies satisfying assumption (iv),

$$\mathbb{E}\left[\int_{T-\tau_i}^{T} (Y_t - \bar{Y}_t)^2\right] < \infty, \tag{74}$$

so that we can conclude the following, using (72c), (72d) and (74),

$$\left\{\pi_t^{(c)}\right\}_{t\geq 0} = \left\{\pi_t^{(a)}\right\}_{t\geq 0}. \tag{75}$$

By assumption (iv) we know that there exist sampling policies such that $\mathbb{E}\left[\int_0^{\tau_i}(Y_t - \bar{Y}_t)^2 dt\right] < \infty$, thus there exist sampling policies such that (74) holds. Since the goal is to minimize the MSE, it sufficies to consider the sampling policies that leads to (74).

Due to (69), the probability distributions of $Y_t, t \in [0, T-\tau_i]$ in (72b) and $Y_t, t \in [\tau_i, T]$ (72c) are the same. Thus, the policy $\{\pi_t\}_{t\geq\tau_i} = \left\{\pi_{t-\tau_i}^{(a)}\right\}_{t-\tau_i\geq 0}$ achieves the infimum in (72b). We conclude

$$\left\{\pi_t^{(b)}\right\}_{t\geq\tau_i} = \left\{\pi_{t-\tau_i}^{(a)}\right\}_{t-\tau_i\geq 0}, \ i = 0, 1, \ldots \tag{76}$$

Using (73) and (76), we conclude that $\left\{\pi^{(a)}_{t-\tau_i}\right\}_{t-\tau_i\geq 0} = \left\{\pi^{(a)}_t\right\}_{t\geq\tau_i}$, $i = 0, 1, \ldots$, i.e.

$$a(s, 0, 0) = a(s + \tau_i, \tau_i, i). \tag{77}$$

Thus, (15) follows.

Next, we show Remark 1 using (15). We conclude that the sampling intervals T_i, $i = 0, 1, \ldots$ are independent due to (69) and the fact that the sampling policy (15) is independent to the process prior to the last stopping time; T_i, $i = 0, 1, \ldots$ are identically distributed due to (70) and the fact that the sampling policy (15) only takes into account the time elapsed from the last sampling time $t-\tau_i$, $t \in [\tau_i, \tau_{i+1})$, $i = 0, 1, \ldots$

We proceed to show that the optimization problem associated with $\underline{D}^{\infty}(F)$ can be reduced to (16) by Lemma 5. The assumptions in Lemma 5 are satisfied with $Z_i \leftarrow T_i$, $R_i \leftarrow \int_{\tau_i}^{\tau_{i+1}} (X_t - \mathbb{E}[X_t|X_{\tau_i}, \tau_i])^2 dt$. $T_0, T_1, \ldots$ are i.i.d. due to Remark 1. The expectation of T_i is finite by assumption (iv). The reward random variables R_i are i.i.d. due to (69)–(70) and Remark 1. Furthermore, the expectation of the reward is finite by assumption (iv). Therefore, using (67), we simplify the DFF in (12) to (16).

E Proof of Theorem 3

We derive a lower bound to the DRF in (19) and show that this lower bound can be achieved by the SOI coding scheme. We write the DRF in (19) as,

$$D(R) = \inf_{\substack{\{\pi_t\}_{t=0}^T \in \Pi_T, \\ \{f_t\}_{t=0}^T \in \mathcal{F}_T: \\ (2)}} \frac{1}{T}\mathbb{E}\left[\int_0^T (X_t - \hat{X}_t)^2 dt\right], \tag{78}$$

where Π_T and $\mathcal{F}_T$ are the set of all causal sampling policies and the set of all causal compressing policies on $[0, T]$ respectively, and $\hat{X}_t$ is given by (18).

Plugging $\hat{X}_t$ (18) into (78), we lower bound (78) as follows,

$$
\begin{aligned}
D(R) &= \inf_{\substack{\{\pi_t\}_{t=0}^T \in \Pi_T, \\ \{f_t\}_{t=0}^T \in \mathcal{F}_T: \\ (2)}} \frac{1}{T}\mathbb{E}\left[\sum_{i=0}^{N_T} \int_{\tau_i}^{\tau_{i+1}} (X_t - \mathbb{E}[X_t | U^i, \tau^i])^2 dt\right] && \text{(79a)} \\
&\geq \inf_{\substack{\{\pi_t\}_{t=0}^T \in \Pi_T: \\ \frac{\mathbb{E}[N_T]}{T} \leq R}} \frac{1}{T}\mathbb{E}\left[\sum_{i=0}^{N_T} \int_{\tau_i}^{\tau_{i+1}} (X_t - \mathbb{E}[X_t | \{X_s\}_{s=0}^{\tau_i}, \tau^i])^2 dt\right] && \text{(79b)} \\
&= \inf_{\substack{\{\pi_t\}_{t=0}^T \in \Pi_T: \\ \frac{\mathbb{E}[N_T]}{T} \leq R}} \frac{1}{T}\mathbb{E}\left[\sum_{i=0}^{N_T} \int_{\tau_i}^{\tau_{i+1}} (X_t - \bar{X}_t)^2 dt\right] && \text{(79c)} \\
&= \underline{D}(R), && \text{(79d)}
\end{aligned}
$$

where (79b) holds since $U^i \in \sigma(\{X_s\}_{s=0}^{\tau_i})$, and

$$
\frac{\mathbb{E}[N_T]}{T} \leq \frac{\mathbb{E}[\sum_{i=1}^{N_T} \ell(U_i)]}{T}. \tag{80}
$$

By examining the proof of Proposition 2 with $\{X_{\tau_j}\}_{j=0}^{i}$ in the conditional expectation of $\bar{Y}_t$ replaced by $\{X_s\}_{s=0}^{\tau_i}$, it is easy to see that Proposition 2 continues to hold, thus Theorem 1 and Corollary 1.1 continue to hold, and the conditional expectation $\mathbb{E}[X_t | \{X_s\}_{s=0}^{\tau_i}, \tau^i]$ in (79b) can be reduced to $\bar{X}_t$ in Corollary 1.1. Furthermore, from Theorem 1 we know that the optimal sampling policy that achieves $\underline{D}(R)$ in (79d) is of the form (13).

We proceed to show that the lower bound (79d) can be achieved with an SOI coding scheme. Since both the stochastic process and the thresholds are continuous, at each sampling time stamp, the innovation must achieve one of the two thresholds. Since the innovation has an alphabet of size 2, and the compressor knows the initial state X_0, we conclude that the 1-bit SOI compressor can noiselessly encode the innovation $X_{\tau_{i+1}} - \mathbb{E}[X_{\tau_{i+1}} | X_{\tau_i}, \tau_i]$, and that the samples can be recovered noiselessly at the decoder using (21). Moreover, since

$\ell(U_i) = 1$ under 1-bit SOI compressor, (80) holds with equality. We conclude (79a) is equal to (79d) under the SOI coding scheme.

F Proof of Proposition 1

Using (14), we calculate that for $t \in [\tau_i, \tau_{i+1})$,

$$X_t - \bar{X}_t \triangleq O_{t-\tau_i} = \frac{\sigma}{\sqrt{2\theta}} e^{-\theta(t-\tau_i)} W_{e^{2\theta(t-\tau_i)}-1}. \tag{81}$$

Let

$$T_i \triangleq \tau_{i+1} - \tau_i, i = 0, 1, 2, \dots, \tag{82}$$

$$R_1(v^2) = \frac{v^2}{\sigma^2} {}_2F_2\left(1, 1; \frac{3}{2}, 2; \frac{\theta}{\sigma^2} v^2\right), \tag{83}$$

$$R_2(v^2) = -\frac{v^2}{2\theta} + \frac{\sigma^2}{2\theta} R_1(v^2), \tag{84}$$

where ${}_2F_2$ is a generalized hypergeometric function[1]. Using (81)-(84) and Remark 1, we write the objective function of (16) as

$$\frac{\mathbb{E}\left[\int_0^{T_i} O_t^2 dt\right]}{\mathbb{E}[T_i]} \tag{85a}$$

$$= \frac{\mathbb{E}\left[R_2(O_{T_i}^2)\right]}{\mathbb{E}[R_1(O_{T_i}^2)]} \tag{85b}$$

$$\geq \frac{R_2(\mathbb{E}[O_{T_i}^2)])}{R_1(\mathbb{E}[O_{T_i}^2])}, \tag{85c}$$

[1]In contrast to the notations in [10], we use $R_1(v^2)$ and $R_2(v^2)$ instead of $R_1(v)$ and $R_2(v)$.

where (85b) is obtained by solving Dynkin's formula for $R_1(O_{T_i}^2)$ and $R_2(O_{T_i}^2)$ respectively, such that [10, Eq.(44)]

$$\mathbb{E}\left[\int_0^{T_i} O_t^2 dt\right] = \mathbb{E}\left[R_2(O_{T_i}^2)\right], \tag{86a}$$

$$\mathbb{E}[T_i] = \mathbb{E}[R_1(O_{T_i}^2)]; \tag{86b}$$

the lower bound (85c) is due to (83)–(84) and the fact that $\frac{A-C}{A} \geq \frac{B-C}{B}$ for $A \geq B \geq C \geq 0$. In particular,

$$A = \frac{\sigma^2}{2\theta}\mathbb{E}[R_1(O_{T_i}^2)],\ B = \frac{\sigma^2}{2\theta}R_1(\mathbb{E}[O_{T_i}^2]),\ C = \frac{1}{2\theta}\mathbb{E}[O_{T_i}^2], \tag{87a}$$

$$\mathbb{E}\left[R_2(O_{T_i}^2)\right] = A - C, \tag{87b}$$

$$R_2(\mathbb{E}[O_{T_i}^2)]) = B - C, \tag{87c}$$

where $A \geq B$ is obtained by applying Jensen's inequality to $R_1(v^2)$, where $R_1(v^2)$ is convex as a function of v^2.

By (86b) ([10, Eq.(43)]) and Jensen's inequality, we write the minimization constraint in (16) as,

$$R_1(\mathbb{E}[O_{T_i}^2]) \leq \mathbb{E}[R_1(O_{T_i}^2)] = \mathbb{E}[T_i] = \frac{1}{R}. \tag{88}$$

For any $R_1(\mathbb{E}[O_{T_i}^2])$ in the range (88), (85c) is a lower bound to (85a). Choosing $R_1(\mathbb{E}[O_{T_i}^2])$ satisfies (88) with equality leads to (85c) being equal to $D(R)$ in (26).

Plugging (25) into (85b), we verify that the lower bound in (85b) is achieved by the symmetric threshold sampling policy in (25).