This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAC.2021.3071953, IEEE Transactions on Automatic Control

Optimal Causal Rate-Constrained Sampling of the Wiener Process

Nian Guo and Victoria Kostina

***Abstract*—We consider the following communication scenario. An encoder causally observes the Wiener process and decides when and what to transmit about it. A decoder estimates the process using causally received codewords in real time. We determine the causal encoding and decoding policies that jointly minimize the mean-square estimation error, under the long-term communication rate constraint of R bits per second. We show that an optimal encoding policy can be implemented as a causal sampling policy followed by a causal compressing policy. We prove that the optimal encoding policy samples the Wiener process once the innovation passes either $\sqrt{\frac{1}{R}}$ or $-\sqrt{\frac{1}{R}}$, and compresses the sign of the innovation (SOI) using a 1-bit codeword. The *SOI coding scheme* achieves the operational distortion-rate function, which is equal to $D^{\mathrm{op}}(R)=\frac{1}{6R}$. Surprisingly, this is significantly better than the distortion-rate tradeoff achieved in the limit of infinite delay by the best non-causal code. This is because the SOI coding scheme leverages the free timing information supplied by the zero-delay channel between the encoder and the decoder. The key to unlocking that gain is the event-triggered nature of the SOI sampling policy. In contrast, the distortion-rate tradeoffs achieved with deterministic sampling policies are much worse: we prove that the causal informational distortion-rate function in that scenario is as high as $D_{\mathrm{DET}}(R)=\frac{5}{6R}$. It is achieved by the uniform sampling policy with the sampling interval $\frac{1}{R}$. In either case, the optimal strategy is to sample the process as fast as possible and to transmit 1-bit codewords to the decoder without delay. We show that *the SOI coding scheme* also minimizes the mean-square cost of a continuous-time control system driven by the Wiener process and controlled via rate-constrained impulses.**

***Index Terms*—Causal lossy source coding, sequential estimation, sampling, rate-distortion theory, continuous-time tracking.**

I. Introduction

A. System Model

Consider the system in Figure 1. A source outputs a continuous-time standard Wiener process $\{W_t\}_{t=0}^T$ within the time horizon $[0,T]$. An encoder observes the process and decides to disclose information about it at a sequence of non-decreasing codeword-generating time stamps

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

These time stamps can be random and they can causally depend on the Wiener process. Consequently, the total number of time stamps N can also be random. At time τ_i, based on the past observed process $\{W_t\}_{t=0}^{\tau_i}$, the encoder chooses to generate a binary codeword U_i, with a length $\ell_i\in\mathbb{Z}^+$. Then, the codeword U_i is passed through a noiseless digital channel to the decoder without delay. Upon receiving the codeword U_i at time τ_i, based on all the received codewords U^i and the codeword-generating time stamps $\{\tau_1,\dots,\tau_i\}$, the decoder updates its running estimate of the Wiener process, yielding $\{\hat{W}_t\}_{t=\tau_i}^T$. The decoder updates its estimate $\{\hat{W}_t\}_{t=\tau_{i+1}}^T$ once the next codeword U_{i+1} is received at τ_{i+1}. The encoder and the decoder are synchronized in time.

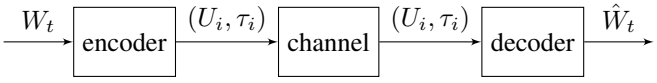


Fig. 1: System Model.

The communication between the encoder and the decoder is subject to the long-term average transmission rate constraint:

$$\frac{1}{T}\mathbb{E}\left[\sum_{i=1}^N\ell_i\right]\le R\ \text{(bits per sec)}. \tag{2}$$

The *distortion* is measured by the long-term mean-square error (MSE) between W_t and $\hat{W}_t$, $0\le t\le T$:

$$\frac{1}{T}\mathbb{E}\left[\int_0^T(W_t-\hat{W}_t)^2dt\right]\le d. \tag{3}$$

We aim to find the jointly optimal encoding and decoding policies that achieve the best tradeoffs between the rate in (2) and the MSE in (3).

B. Motivation and literature Review

With the rapid development of the Internet of Things, studying communication between nodes in networked tracking, robotic control, and navigation systems has become important. In these network systems, the communication between nodes is rate-constrained and delay-sensitive. This paper focuses on the following remote-tracking scenario. A transmitting node (encoder) sends the causally observed environment process, e.g. location, speed, temperature, to a receiving node. The receiving node (decoder) aims to recover the process in real time using the causally received information.

The commonly used communication constraint between the encoder and the decoder in prior literature [1]-[9] is the sampling frequency constraint. Finding sampling policies at the encoder and estimation policies at the decoder to jointly minimize the end-to-end distortion under transmission constraints falls into the realm of optimal scheduling and remote sequential estimation problems. Åström and Bernhardsson [1] compared uniform and symmetric threshold sampling policies[1] (referred to as Riemann and Lebesgue sampling, respectively)

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. This work was supported in part by the National Science Foundation (NSF) under grant CCF-1751356. A part of this work is presented at the 57th Annual Allerton Conference [36]; the conference version does not contain Section VI or any proofs.

[1]The symmetric threshold sampling policy samples the process if its current value exceeds or falls short of the previous sample by a certain threshold.

in continuous-time first-order stochastic systems with a Wiener process disturbance, and showed that the Lebesgue sampling gives a lower distortion than the Riemann sampling under the same average sampling frequency. Imer and Başar [2] considered causal estimation of i.i.d. processes under MSE and the constraint on the number of transmissions over a finite time horizon, and showed that the time-varying symmetric threshold sampling policy is optimal for i.i.d. Gaussian processes. Lipsa and Martins [3] proved that a time-varying symmetric threshold policy and a Kalman-like filter jointly minimize a discounted cost function consisting of MSE and a communication cost, for scalar discrete-time Gauss-Markov (GM) processes over a finite time horizon. For partially observed discrete-time GM processes, Wu et al. [4] fixed an event-triggered policy, where the encoder transmits only if the L-infinity norm of the measurement innovation exceeds a constant, and derived both the accurate and an approximate minimum MSE (MMSE) estimator to combine with that sampling policy. Rabi et al. [5] formulated the problem of causal estimation of the Wiener process under the constraint on the number of transmissions over a finite time horizon as an optimal stopping time problem. Rabi et al. [5] showed that the optimal deterministic sampling policy and the optimal event-triggered sampling policy for the Wiener process are a uniform policy and a symmetric threshold policy, respectively. Nar and Başar [6] extended the optimal stopping time problem in [5] to the multidimensional Wiener process, and proved that a symmetric threshold policy remains optimal over both finite and infinite time horizons. In particular, Nar and Başar [6] showed that the optimal threshold over the infinite horizon is a constant depending on the average sampling frequency. For autoregressive Markov processes driven by an i.i.d. process with unimodal and symmetric distribution, Charkravorty and Mahajan [7] proved that a symmetric threshold sampling policy together with a Kalman-like estimator achieves the best tradeoff between the estimation distortion and the number of transmissions. For the same scenario as in [7], Molin and Hirche [8] proposed an iterative algorithm to show that the optimal event-triggered policy converges to a symmetric threshold policy. Nayyar et al. [9] considered a scenario where the encoder relies on the energy harvested from the environment to transmit messages, and proved that the optimal sampling strategy is a symmetric threshold policy, provided that the finite-state Markov source and the distortion measure satisfy regularity conditions.

In contrast to the scenarios in [1]-[9], where the communication channel is assumed to be perfect, [10]-[13] consider imperfect communication channels. Sun et al. [10] proved that a symmetric threshold policy remains optimal even when the samples of the Wiener process experience an i.i.d. random transmission delay, but the threshold depends on the distribution of channel delay and is different from the one in [6]. Using dynamic programming, Gao et al. in [11] derived the optimal sampling, encoding, and decoding policies for the event-triggered sampling of an i.i.d. Laplacian source with subsequent transmission over a channel with a Gamma additive noise, under an average power constraint. For discrete-time first-order autoregressive Markov processes considered in [7]-[8], Ren et al. [12] introduced a fading channel between the encoder and the decoder, where a successful transmission depends on both the channel gains and the transmission power, and found the optimal encoding and decoding policies that minimize an infinite horizon cost function combining the MSE and the power usage. For first-order autoregressive sources considered in [7][8][12], Chakravorty and Mahajan [13] further proved that the optimal estimation policy is a Kalman-like filter and that the optimal sampling policy is symmetric threshold policy when the communication channel is a packet-drop channel with Markovian states.

In [1]-[9], the encoder transmits real-valued samples, while in practice, real numbers are quantized before they are transmitted in digital communication systems. For the non-causal lossy source coding of the uniformly sampled Wiener process, Kipnis et al. [14] derived the trade-offs between the sampling frequency, the communication bitrate, and the estimation MSE, achievable in the limit of infinite delay. However, the infinite delay introduced by classical rate-distortion theory in [14] is unsuitable in many delay-sensitive applications.

Quantized event-triggered control has attracted significant research attention in recent years [16]-[27]. Kofman and Braslavsky [16] designed a quantized event-triggered controller for noiseless partially observed continuous-time LTI systems to ensure asymptotic convergence of the system to the origin with zero average rate, seemingly violating the *data-rate theorem* [15]. Similar to [16], the fact that sampling time stamps of event-triggered policies carry information is also exploited in [17]-[20]. Pearson et al. [17] considered encoding the deterministic and possibly nonuniformly sampled states of noiseless continuous-time LTI systems into symbols in a finite alphabet with a *free* symbol representing the absence of transmission. For discrete-time linear systems with additive disturbances, Khina et al. [18] considered a setting where at each discrete-time instant, the encoder either transmits 1 bit or transmits the free symbol, and designed a quantizer with three bins using a Lloyd-Max algorithm with the quantization bin of the largest probability corresponding to the free symbol. Ling [19] designed a periodic event-triggered quantization policy to stabilize continuous-time LTI systems subject to i.i.d. feedback dropouts, bounded network delay, and bounded noise, which leads to a stabilizing rate that is lower than the one the data-rate theorem [15] requires for time-triggered policies. Khojasteh et al. [20] considered sampling noiseless continuous-time LTI systems where the state estimation error exceeds an exponentially decaying function. They found that for small enough delays, the information transmission rate required for stabilizing systems can be any positive value; it starts to increase once the delay exceeds a critical value. Quantized event-triggered control has also been studied for continuous-time LTI systems with bounded disturbances [21], for partially-observed continuous-time LTI systems without noise [22] and with bounded noise [23], for discrete-time noiseless linear systems [24], and for partially observed continuous-time LTI systems with time-varying network delay [25]. Event-triggered

control schemes to guarantee exponential stabilization were designed both for continuous-time LTI systems with bounded disturbances under a bounded rate constraint [26] and for noiseless continuous-time LTI systems under time-varying rates constraints and channel blackouts [27].

C. Contribution

In this paper, we adopt an information-theoretic approach to continuous-time causal estimation, by considering the optimal tradeoff between the achievable MSE and the average number of bits communicated. This is different from the models studied in [1]-[9], where communication cost is measured by the number of transmissions, and each infinite-precision transmission can carry an infinite amount of information. For communication over digital channels, a bitrate constraint, routinely considered in information theory, is more appropriate. Our setting is different from [14] in that we do not ignore delay: our distortion at time t is measured with respect to the actual value of the process at time t; whereas [14] permits an infinite delay, following a standard assumption in information theory. In contrast to the works [16]-[27] that do not claim or consider the optimality of the proposed event-triggered policies, we show the optimality of the SOI coding scheme for our problem setting in Section I-A.

We first show that an optimal encoding policy that achieves the operational distortion-rate function (ODRF) can be implemented as a causal sampling policy coupled with a compressing policy. Then, we prove that the optimal encoding policy is a symmetric threshold sampling policy with threshold $\pm\sqrt{\frac{1}{R}}$ and a 1-bit SOI compressor. The optimal decoding policy causally estimates the Wiener process by summing up the received innovations. This coding scheme, termed *the SOI coding scheme*, achieves the ODRF $D^{\mathrm{op}}(R) = \frac{1}{6R}$.

In the SOI coding scheme, the encoder continuously tracks the process and generates a bit once the process passes a pre-set threshold. To reconstruct the process, both those bits and their time stamps are required at the decoder. In the scenario where the sampler is process-agnostic, or the decoder has no access to timing information, one has to adopt a process-independent sampling policy. We prove that a uniform sampling policy with the sampling interval $\frac{1}{R}$ achieves the informational distortion-rate function (IDRF), which is equal to $D_{\mathrm{DET}}(R) = \frac{5}{6R}$. To define the IDRF for the deterministic sampling policies, we change the rate constraint (2) to a directed mutual information rate constraint, which serves as an information-theoretic lower bound to (2). To confirm that the IDRF is a meaningful gauge of what is achievable in the zero-delay causal compression, we implement the greedy Lloyd-Max compressor [18] to compress the innovations $W_{\tau_i} - \hat{W}_{\tau_{i-1}}$, and we verify that the performance of the resulting scheme is close to the IDRF.

To study the tradeoffs between the sampling frequency and the rate per sample under a rate per second constraint R, we define the operational and the informational distortion-frequency-rate functions (ODFRF and IDFRF). They are both achieved by the maximum frequency R and the minimum rate 1 bit/sample, implying that sampling the process as fast as possible under the rate constraint and transmitting 1-bit codewords to the decoder without delay is optimal.

Surprisingly, the distortion achieved by the SOI coding scheme is smaller than the distortion achieved by the best non-causal codes. This is because in the SOI coding scheme, the encoder and the decoder know the random sampling time stamps perfectly, whereas in the classical non-causal coding setting, the free timing information is not considered.

We show that the SOI coding scheme continues to be optimal when there is a fixed channel delay between the codeword-generating time and the codeword-delivery time. We also show that if the decoder is allowed to wait for the next codeword before decoding, the MSE can be further decreased.

Finally, we prove that the SOI coding scheme is also optimal in the rate-constrained event-triggered control scenario with a continuous-time stochastic plant driven by the Wiener process and controlled via impulse control. The SOI code minimizes the mean-square cost between the state of the stochastic plant and the desirable state 0.

D. Paper organization

In Section II, we define causal codes and distortion-rate functions. In Section III, we state the main results of this paper, including the optimal causal sampling and compressing policies and the tradeoffs between the sampling frequency and the rate per sample. In Section IV, we show the proof of the main results. In Section V, we discuss the distortion-rate tradeoffs when delays are allowed at both the encoder and the decoder, at the decoder only, and at the channel. In Section VI, we show the optimal causal sampling and compressing policies in a rate-constrained event-triggered control system.

E. Notations

We denote by $\{W_t\}_{t=\tau_i}^{\tau_{i+1}}$ and $\{W_t\}_{\tau_i<t<\tau_{i+1}}$ the parts of the Wiener process within the time intervals $[\tau_i, \tau_{i+1}]$ and (τ_i, τ_{i+1}), respectively. For $M \in \mathbb{Z}^+$, we put $[M] \triangleq \{1, \dots, M\}$. For a possibly infinite sequence $x = \{x_1, x_2, \dots\}$, we write $x^i = \{x_1, x_2, \dots, x_i\}$ to denote the vector of its first i elements.

II. Distortion-rate functions

A. Encoding and decoding policies

Definition 1 (standard Wiener process, e.g. [30])**.** *A standard Wiener process $\{W_t\}_{t\geq 0}$ is a stochastic process characterized by the following three properties:*

(i) time-homogeneity*: for all non-negative s and t, W_s and $W_{s+t} - W_t$ have the same distribution ($W_0 = 0$);*

(ii) independent increments*: $W_{t_i} - W_{s_i}$ $(i \geq 1)$ are independent whenever the intervals $(s_i, t_i]$ are disjoint;*

(iii) W_t follows the Gaussian distribution $\mathcal{N}(0, t)$.

Throughout, we assume that both the encoder and the decoder know the initial state $W_0 = 0$ at $\tau_0 = 0$. Next, we formally define the encoding and the decoding policies[2]. We

[2]We refer to them as encoding/decoding *policies*, not mappings, to emphasize their causal nature.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAC.2021.3071953, IEEE Transactions on Automatic Control

denote the set of continuous functions on the time interval $[0,t]$ by $\mathcal{C}_{[0,t]}$.

Definition 2 (An (R,d,T) causal code)**.** *An (R,d,T) causal code for the Wiener process $\{W_t\}_{t=0}^T$ is a pair of encoding and decoding policies defined as follows.*

The encoding *policy consists of*

(i) the causal sampling *policy $\pi_T = \{\tau_1, \tau_2, \dots\}$ that decides the codeword-generating time stamps in (1) that are stopping times of the filtration $\sigma(\{W_t\}_{t=0}^T)$, and*

(ii) the compressing *policy $f_T = \{\mathsf{f}_1, \mathsf{f}_2, \dots\}$*[3]*,*

$$\mathsf{f}_i \colon \mathcal{C}_{[0,T]} \to \left[2^{\ell i}\right]. \tag{4}$$

The codeword generated at time τ_i is $U_i = \mathsf{f}_i\left(\left\{W_{\min\{t,\tau_i\}}\right\}_{t=0}^T\right)$. The codewords' lengths must satisfy the long-term average rate constraint (2).

The decoding *policy causally maps the received codewords and the codeword-generating time stamps to a continuous-time process estimate $\{\hat{W}_t\}_{t=0}^T$ using*

$$\hat{W}_t \triangleq \mathbb{E}[W_t | U^i, \tau^i, t < \tau_{i+1}],\ t \in [\tau_i, \tau_{i+1}). \tag{5}$$

Together, the encoding and the decoding policies must satisfy the long-term MSE constraint in (3).

Allowing more freedom in the design of a decoding policy cannot yield a lower MSE because (5) is the MMSE estimator of W_t. This is a consequence of the zero-delay MSE constraint in (3). As we explain in Section V-B below, had we allowed delay at the decoder, we could have improved performance by e.g. using linear interpolation between recovered samples at the decoder. As we show in the proof of Theorem 1, given the optimal encoding policy, the optimal decoding policy in (5) reduces to

$$\hat{W}_t = \hat{W}_{\tau_i} \triangleq \mathbb{E}[W_{\tau_i} | U^i, \tau^i],\ t \in [\tau_i, \tau_{i+1}). \tag{6}$$

B. Operational distortion-rate function

Definition 3 (Operational distortion-rate function (ODRF))**.** *The ODRF is the minimum distortion compatible with rate R achievable by causal rate-R codes in the infinite time horizon:*

$$D^{\mathrm{op}}(R) \triangleq \limsup_{T\to\infty} \inf\{d : \exists\ (R,d,T)\ \textit{causal code}\}. \tag{7}$$

We define $\tau_{N+1} \triangleq T$ and define Π_T, F_T as the sets of all sampling and all compressing policies over the time horizon T respectively.

It turns out that the ODRF can be decomposed into the distortion due to sampling and the distortion due to quantization.

Proposition 1. *The ODRF for the Wiener process can be written as*

$$D^{\mathrm{op}}(R) = \limsup_{T\to\infty} \inf_{\pi_T \in \Pi_T} \frac{1}{T} \left\{ \mathbb{E}\left[\sum_{i=0}^{N} \int_{\tau_i}^{\tau_{i+1}} (W_t - W_{\tau_i})^2 dt\right] \right. \tag{8a}$$

$$\left. + \inf_{\substack{f_T \in F_T: \\ (2)}} \mathbb{E}\left[\sum_{i=1}^{N} (\tau_{i+1} - \tau_i)(W_{\tau_i} - \hat{W}_{\tau_i})^2\right] \right\}, \tag{8b}$$

where $\hat{W}_{\tau_i}$ is given in (6). Furthermore, if randomized compressing policies are allowed, there is no loss of optimality if at time τ_i, compressing policy f_T only takes into account the innovation $W_{\tau_i} - \hat{W}_{\tau_{i-1}}$, past codewords U^{i-1} and timing information τ^i, rather than the whole process up to time τ_i, as permitted by Definition 2.

Proof. Appendix A. □

In (8a), W_{τ_i} is the MMSE estimator of W_t at $t \in [\tau_i, \tau_{i+1})$, given the past samples $\{W_{\tau_j}\}_{j=1}^i$ and the codeword-generating times τ^i. The expectation in (8a) is the distortion due to causally estimating the Wiener process from its samples $\{W_{\tau_j}\}_{j=1}^i$ taken under the sampling policy π_T. The expectation in (8b) is the mean-square quantization error of the samples, accumulated over sampling intervals of length $\tau_{i+1} - \tau_i$, $i = 1, \dots, N$. According to the compressing policy in Proposition 1, the minimization problem in (8b) is the operational zero-delay causal distortion-rate function of the discrete-time stochastic process formed by the samples. Furthermore, the encoding policy can be implemented as a sampler followed by a compressor. See Figure 2. The sampler

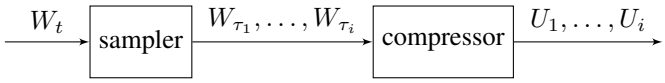


Fig. 2: Decomposition of the encoder.

takes measurements of the Wiener process under a sampling policy and outputs samples without delay to the compressor. Upon receiving a new sample, the compressor immediately generates a codeword under the compressing policy described in Proposition 1.

C. Informational distortion-rate function

The *directed information* $I(X^n \to Y^n)$ from a sequence X^n to a sequence Y^n is defined as [31]

$$I(X^n \to Y^n) = \sum_{i=1}^{n} I(X^i; Y_i | Y^{i-1}). \tag{9}$$

The directed information captures the information due to the causal dependence of Y^n on X^n.

A sampling policy $\pi_T = \{\tau_1, \tau_2, \dots\}$ is *deterministic* if its sampling time stamps (1) are deterministic. We denote by Π_T^{DET} the set of all deterministic sampling policies. Under a deterministic sampling policy, the total number of samples N within the time horizon $[0,T]$ is constant.

[3] In some scenarios, we allow randomness in the mapping f_i, replacing the deterministic mapping f_i in (5) by a transition probability kernel.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAC.2021.3071953, IEEE Transactions on Automatic Control

Definition 4 (Informational distortion-rate function (IDRF))**.** *The IDRF for the Wiener process under deterministic sampling policies is defined as*

$$D_{\mathrm{DET}}(R) \triangleq$$

$$\limsup_{T\to\infty} \inf_{\pi_T\in\Pi_T^{\mathrm{DET}}} \frac{1}{T}\left\{\mathbb{E}\left[\sum_{i=0}^{N}\int_{\tau_i}^{\tau_{i+1}}(W_t-W_{\tau_i})^2dt\right]+ \right. \tag{10a}$$

$$\left. \inf_{\substack{\otimes_{i=1}^{N} P_{\hat{W}_{\tau_i}|W^{\tau_i},\hat{W}^{\tau_{i-1}}}:\\ \frac{I(W^{\tau_N}\to\hat{W}^{\tau_N})}{T}\le R}} \mathbb{E}\left[\sum_{i=1}^{N}(\tau_{i+1}-\tau_i)(W_{\tau_i}-\hat{W}_{\tau_i})^2\right]\right\}, \tag{10b}$$

The minimization problem (10b) in $D_{\mathrm{DET}}(R)$ is the causal IDRF for the discrete-time stochastic process formed by the samples. Note that (10b) is minimized over the directed information rate, which gives an information-theoretic lower bound to the rate in (2). According to [34, Eq. (43)], we have

$$D_{\mathrm{DET}}^{\mathrm{op}}(R) \ge D_{\mathrm{DET}}(R), \tag{11}$$

where $D_{\mathrm{DET}}^{\mathrm{op}}(R)$ is the ODRF for deterministic sampling policies defined by (8) with the minimization constraint in (8a) replaced by $\pi_T \in \Pi_T^{\mathrm{DET}}$.

D. ODFRF and IDFRF

According to Proposition 1, an optimal encoder can be implemented as a sampler followed by a compressor. To gain insight into the tradeoffs between the sampling frequency f at the sampler and the rate per sample R_s at the compressor, we define an (f, R_s, d, T) causal code.

Definition 5 (An (f, R_s, d, T) causal code)**.** *An (f, R_s, d, T) causal code for the Wiener process $\{W_t\}_{t=0}^{T}$ is a triplet of causal sampling, compressing and decoding policies:*

(i) the causal sampling policy defined in Definition 2(i) $\pi_T = \{\tau_1, \tau_2, \dots\}$ satisfies the sampling frequency constraint

$$\frac{1}{T}\mathbb{E}[N] = f; \tag{12}$$

(ii) the compressing policy $f_T = \{\mathsf{f}_1, \mathsf{f}_2, \dots\}$[4] is

$$\mathsf{f}_i : \mathbb{R}\times\mathbb{R}^{i-1}\times\mathbb{R}^{i} \to \left[2^{\ell_i}\right]. \tag{13}$$

The codeword generated at time τ_i is $U_i = \mathsf{f}_i\left(W_{\tau_i}, U^{i-1}, \tau^i\right)$. The codewords' lengths must satisfy

$$\frac{1}{\mathbb{E}[N]}\mathbb{E}\left[\sum_{i=1}^{N}\ell_i\right] \le R_s \ (\textit{bits per sample}); \tag{14}$$

(iii) the decoding policy causally maps the received codewords and the codeword-generating time stamps to a continuous-time process estimate $\{\hat{W}_t\}_{t=0}^{T}$ using (6).

Together, the causal sampling, compressing, and decoding policies must satisfy the long-term MSE constraint in (3).

[4]Here we slightly abuse the notation: we have used f_T in Definition 2(ii), and have shown in Proposition 1 that the compressing policy f_T can be simplified to (13).

Definition 6 (Operational distortion-frequency-rate function(ODFRF))**.** *The ODFRF is the minimum distortion achievable by causal frequency-f and rate-R_s codes in the infinite time horizon:*

$$D^{\mathrm{op}}(f, R_s) \triangleq \limsup_{T\to\infty} \inf\{d : \exists\ (f, R_s, d, T)\ \textit{causal code}\}.$$

Using the method used to decompose $D^{\mathrm{op}}(R)$ in Proposition 1, we write $D^{\mathrm{op}}(f, R_s)$ as

$$\limsup_{T\to\infty} \inf_{\substack{\pi_T\in\Pi_T:\\ (12)}} \frac{1}{T}\left\{\mathbb{E}\left[\sum_{i=0}^{N}\int_{\tau_i}^{\tau_{i+1}}(W_t-W_{\tau_i})^2dt\right] \right. \tag{15a}$$

$$\left. + \inf_{\substack{f_T\in F_T:\\ (14)}} \mathbb{E}\left[\sum_{i=1}^{N}(\tau_{i+1}-\tau_i)(W_{\tau_i}-\hat{W}_{\tau_i})^2\right]\right\}, \tag{15b}$$

where the expectation in (15a) is the sampling distortion, and the expectation in (15b) is the quantization distortion.

The informational equivalent of $D^{\mathrm{op}}(f, R_s)$ replaces (14) by the constraint on the directed information, that is, for deterministic sampling policies,

$$\frac{1}{N}I(W^{\tau_N}\to\hat{W}^{\tau_N}) \le R_s. \tag{16}$$

Definition 7 (Informational distortion-frequency-rate function (IDFRF))**.** *The IDFRF for the Wiener process under deterministic sampling policies is defined as*

$$D_{\mathrm{DET}}(f, R_s) \triangleq$$

$$\limsup_{T\to\infty} \inf_{\substack{\pi_T\in\Pi_T^{\mathrm{DET}}:\\ (12)}} \frac{1}{T}\left\{\mathbb{E}\left[\sum_{i=0}^{N}\int_{\tau_i}^{\tau_{i+1}}(W_t-W_{\tau_i})^2dt\right] \right. \tag{17a}$$

$$\left. + \inf_{\substack{\otimes_{i=1}^{N} P_{\hat{W}_{\tau_i}|W^{\tau_i},\hat{W}^{\tau_{i-1}}}:\\ (16)}} \mathbb{E}\left[\sum_{i=1}^{N}(\tau_{i+1}-\tau_i)(W_{\tau_i}-\hat{W}_{\tau_i})^2\right]\right\} \tag{17b}$$

Similar to (10b), the optimization problem in (17b) is the IDRF for the Guass-Markov (GM) process formed by the samples, but the rate in (17b) is the rate per sample R_s rather than the rate per second R in (10b).

III. Main Results

The first theorem of this section shows the optimal causal sampling and compressing policies that achieve $D^{\mathrm{op}}(R)$.

Theorem 1. *In causal coding of the Wiener process, the optimal causal sampling policy is the following symmetric threshold sampling policy:*

$$\tau_{i+1} = \inf\left\{t \ge \tau_i : |W_t - W_{\tau_i}| \ge \sqrt{\frac{1}{R}}\right\},\ i = 0, 1, 2, \dots \tag{18}$$

The optimal compressing policy is the 1-bit sign-of-innovation (SOI) compressor:

$$U_i = \begin{cases} 1 & \textit{if}\quad W_{\tau_{i+1}} - W_{\tau_i} \ge 0 \\ 0 & \textit{if}\quad W_{\tau_{i+1}} - W_{\tau_i} < 0. \end{cases} \tag{19}$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAC.2021.3071953, IEEE Transactions on Automatic Control

The SOI coding scheme achieves the ODRF:

$$D^{\mathrm{op}}(R) = \frac{1}{6R}. \tag{20}$$

Proof. Section IV-A. □

Together with the optimal encoding policy in Theorem 1, the optimal decoding policy (6) accumulates the received noiseless innovations to estimate the current value of the process.

The next theorem shows the optimal deterministic sampling policy that achieves $D_{\mathrm{DET}}(R)$.

Theorem 2. *In causal coding of the Wiener process, the uniform sampling with the sampling interval equal to*

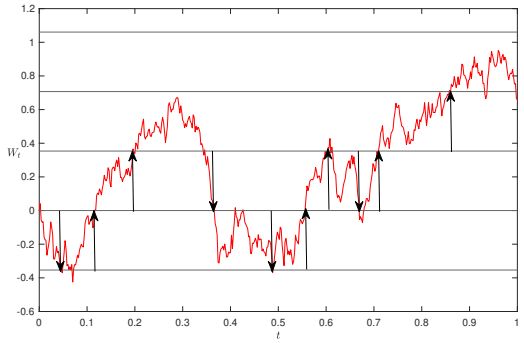
$$\tau_{i+1} - \tau_i = \frac{1}{R},\ i = 0, 1, 2, \ldots, \tag{21}$$

achieves

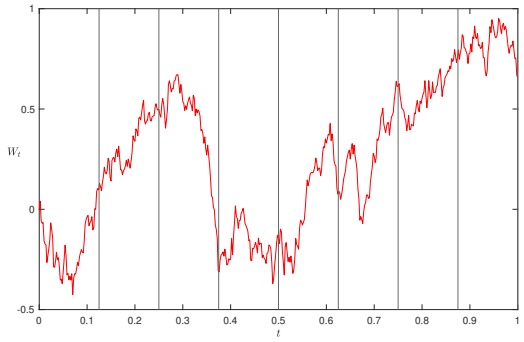
$$D_{\mathrm{DET}}(R) = \frac{5}{6R}. \tag{22}$$

Proof. Section IV-D. □

Figure 3a shows how the encoding policy of the SOI code in Theorem 1 works. The red curve represents the Wiener process. The gap between horizontal lines represents the sampling threshold $\sqrt{\frac{1}{R}}$. A down-arrow appears if the process innovation $W_t - W_{\tau_i}$ crosses the negative threshold, and codeword $U_i = 0$ is transmitted. An up-arrow appears if the process innovation $W_t - W_{\tau_i}$ crosses the positive threshold, and codeword $U_i = 1$ is transmitted. Figure 3b shows the uniform sampling policy in Theorem 2, where the gap between vertical lines represents the sampling interval $\frac{1}{R}$.



(a)



(b)

Fig. 3: (a) SOI code and (b) uniform sampling at $R = 8$.

Theorem 3. *In causal coding of the Wiener process, the ODRF satisfies*

$$D^{\mathrm{op}}(R) = \min_{\substack{f>0, R_s \geq 1: \\ f R_s \leq R}} D^{\mathrm{op}}(f, R_s), \tag{23a}$$

$$= D^{\mathrm{op}}(R, 1), \tag{23b}$$

and the IDRF under deterministic sampling policies satisfies

$$D_{\mathrm{DET}}(R) = \min_{\substack{f>0, R_s \geq 1: \\ f R_s \leq R}} D_{\mathrm{DET}}(f, R_s) \tag{24a}$$

$$= D_{\mathrm{DET}}(R, 1). \tag{24b}$$

Proof. See Section IV-B for the proof of (23). See Section IV-C for the proof of (24). □

Using Theorem 3, we conclude that the *working principle* of the optimal encoding policy is to transmit 1-bit codewords about the Wiener process as frequently as possible. The optimal encoding policies in both Theorems 1 and 2 follow this working principle.

In the setting of Theorem 2, although evaluating $D_{\mathrm{DET}}(R)$ does not give us an operational compressing policy, we know that the stochastic kernel that achieves the causal IDRF for discrete-time GM processes formed by the samples under uniform sampling policies has the form $\bigotimes_{i=1}^{\infty} P_{\hat{W}_{\tau_i}|W_{\tau_i}-\hat{W}_{\tau_{i-1}},\hat{W}_{\tau_{i-1}}}$ [33, Eq. (5.12)], suggesting that at the encoder, it is sufficient to compress the quantization innovation $W_{\tau_i} - \hat{W}_{\tau_{i-1}}$ only. The decoder computes the estimate $\hat{W}_{\tau_i}$ as $\hat{W}_{\tau_i} = \hat{W}_{\tau_{i-1}} + \mathsf{q}_i(W_{\tau_i} - \hat{W}_{\tau_{i-1}})$, where $\mathsf{q}_i = \mathsf{g}_i \circ \mathsf{f}_i$, $\mathsf{f}_i\left(W_{\tau_i} - \hat{W}_{\tau_{i-1}}\right) \in \left[2^{\ell_i}\right]$ is the i-th binary codeword, and $\mathsf{g}_i(\cdot) \in \mathbb{R}$ is the quantization representation point of its argument. In practice, one can use the *greedy Lloyd-Max quantizer* [18] that runs the Lloyd-Max algorithm for the quantization innovation in each step based on its prior pdf. Specifically, the prior pdf for the $(i+1)$-th step quantization innovation $W_{\tau_{i+1}} - \hat{W}_{\tau_i}$ can be computed by convolving the pdfs of the quantization error $W_{\tau_i} - \hat{W}_{\tau_i}$ and the process increment $W_{\tau_{i+1}} - W_{\tau_i}$. The globally optimal scheme has a negligible gain over the greedy Lloyd-Max algorithm even in the finite horizon [18].

Figure 4 displays distortion-rate tradeoffs obtained in Theorems 1 and 2, as well as a numerical simulation of the uniform sampler in Theorem 2 with the greedy Lloyd-Max quantizer. The symmetric threshold sampling policy followed by the 1-bit SOI compressor leads to a much lower MSE than uniform sampling. Indeed, according to Theorems 1 and 2, $\frac{D_{\mathrm{DET}}(R)}{D^{\mathrm{op}}(R)} = 5$, and $D^{\mathrm{op}}_{\mathrm{DET}}(R)$ for the uniform sampling is even higher than $D_{\mathrm{DET}}(R)$ by (11). Note that the greedy Lloyd-Max curve is rather close to the $D_{\mathrm{DET}}(R)$ curve, indicating that the IDRF is a meaningful gauge of what is attainable in zero-delay continuous-time causal compression.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAC.2021.3071953, IEEE Transactions on Automatic Control

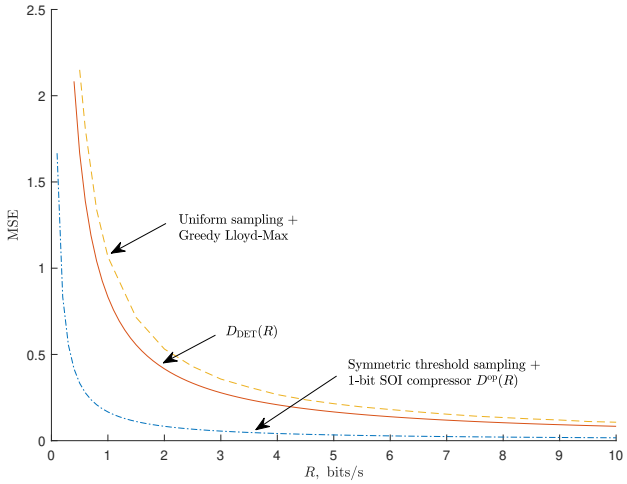


Fig. 4: MSE versus rate

IV. Proofs of the Main Results

A. Proof of Theorem 1

To prove that the SOI coding scheme in Theorem 1 achieves the ODRF, we derive a lower bound to the ODRF, and we show that the lower bound is achieved by the SOI coding scheme. The MSE achievable by causal rate-R codes is lower bounded as follows:

$$\inf_{\substack{\pi_T\in\Pi_T,\\ f_T\in F_T:\\ (2)}} \frac{1}{T}\mathbb{E}\left[\sum_{i=0}^{N}\int_{\tau_i}^{\tau_{i+1}}(W_t-\mathbb{E}[W_t|U^i,\tau^i,t<\tau_{i+1}])^2dt\right] \tag{25a}$$

$$\geq \inf_{\substack{\pi_T\in\Pi_T:\\ \frac{\mathbb{E}[N]}{T}\leq R}} \frac{1}{T}\mathbb{E}\left[\sum_{i=0}^{N}\int_{\tau_i}^{\tau_{i+1}}(W_t-\mathbb{E}[W_t|\{W_s\}_{s=0}^{\tau_i},\tau^i, t<\tau_{i+1}])^2dt\right] \tag{25b}$$

$$= \inf_{\substack{\pi_T\in\Pi_T:\\ \frac{1}{T}\mathbb{E}[N]\leq R}} \frac{1}{T}\mathbb{E}\left[\sum_{i=0}^{N}\int_{\tau_i}^{\tau_{i+1}}(W_t-\mathbb{E}[W_t|W_{\tau_i},\tau_i])^2dt\right] \tag{25c}$$

$$= \inf_{\substack{\pi_T\in\Pi_T:\\ \frac{1}{T}\mathbb{E}[N]\leq R}} \frac{1}{T}\mathbb{E}\left[\sum_{i=0}^{N}\int_{\tau_i}^{\tau_{i+1}}(W_t-W_{\tau_i})^2dt\right], \tag{25d}$$

where (25b) holds since $\sigma(U^i)\subset\sigma(\{W_t\}_{t=0}^{\tau_i})$ and

$$\mathbb{E}[N]\leq\mathbb{E}\left[\sum_{i=1}^{N}\ell_i\right]; \tag{26}$$

(25c) holds due to [38, Cor. 1.1 and Appendix H 2)], since the Wiener process satisfies the regularity conditions in [38, assumptions (i)-(iii) and Remark 2]; (25d) is due to the strong Markov property of the Wiener process.

It remains to show that the lower bound (25d) is achieved by the SOI coding scheme. First, the optimization problem in (25d) corresponds to determining the optimal sampling policy that minimizes the MSE subject to an average sampling frequency constraint, where N is the total number of samples taken within $[0,T]$. According to [6, Eq. (20)], the optimal sampling policy that achieves the $\limsup_{T\to\infty}$ of (25d) is given by (18). Second, each innovation $\Delta W_i\triangleq W_{\tau_{i+1}}-W_{\tau_i}$, $i=0,1,2,\cdots$ is equiprobabily distributed on the size-2 alphabet $\left\{\pm\sqrt{\frac{1}{R}}\right\}$. Thus ΔW_i can be noiselessly encoded using 1-bit codewords U^i, while satisfying the inequality in (26) with equality. It follows that the $\limsup_{T\to\infty}$ of (25d) is achieved by the SOI coding scheme.

From the equality in (25c) and the fact that the SOI coding scheme attains (25a), we conclude that the optimal decoding policy (5) can indeed be simplified to (6) given the optimal encoding policy in Theorem 1.

B. Proof of Theorem 3 (23)

$D^{\mathrm{op}}(f,R_s)$ is lower-bounded by the sampling distortion (15a), and (15a) is minimized by the symmetric threshold sampling policy with thresholds $\pm\sqrt{\frac{1}{f}}$ yielding the sampling distortion of $\frac{1}{6f}$ [6, Eq. (20)]. Since the SOI compressor results in zero quantization distortion (15b), it follows that

$$D^{\mathrm{op}}(f,R_s)=\frac{1}{6f},\ \forall R_s\geq 1. \tag{27}$$

We plug (27) into the minimization problem in (23a) to obtain:

$$\min_{\substack{f>0,R_s\geq 1:\\ fR_s\leq R}} D^{\mathrm{op}}(f,R_s)=D^{\mathrm{op}}(R,1)=\frac{1}{6R}. \tag{28}$$

Comparing (28) to (20), we conclude that (23) holds.

C. Proof of Theorem 3 (24)

Since the samples taken under a deterministic sampling policy form a GM process, we first compute $D_{\mathrm{DET}}(f,R_s)$ by building on existing results on the causal IDRF (17b) of discrete-time GM processes.

Lemma 1. *The IDFRF under deterministic sampling policies can be written as*

$$D_{\mathrm{DET}}(f,R_s)=\limsup_{N\to\infty}D_N(f,R_s), \tag{29a}$$

$$D_N(f,R_s)=\inf_{\substack{T^N\geq 0:\\ (30)}}\frac{f}{N}\left(\sum_{i=0}^{N}\frac{T_i^2}{2}+\min_{\substack{D^N\geq 0:\\ (31)}}\sum_{i=1}^{N}T_iD_i\right), \tag{29b}$$

where the minimization constraints in (29) are

$$\frac{1}{N}\sum_{i=0}^{N}T_i=\frac{1}{f}, \tag{30}$$

$$z\left(D^N\right)\triangleq\frac{1}{N}\left(\sum_{i=1}^{N-1}\log_2\left(1+\frac{T_i}{D_i}\right)+\log_2\left(\frac{T_0}{D_N}\right)\right)\leq 2R_s, \tag{31a}$$

$$D_{i-1}+T_{i-1}\geq D_i, i=1,\ldots,N,\ D_0=0. \tag{31b}$$

Proof. Appendix B. □

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAC.2021.3071953, IEEE Transactions on Automatic Control

The optimization variable T^N in (29) is the vector of sampling intervals $T^N = \{T_0, T_1, \ldots, T_N\}$ where

$$\begin{aligned} T_i &= \tau_{i+1} - \tau_i, \ i = 0, \ldots, N-1 \\ T_N &= T - \tau_N, \end{aligned} \tag{32}$$

that determines a deterministic sampling policy. The vector D^N in (29) contains sample distortions $D^N = \{D_1, \ldots, D_N\}$. $D_{\mathrm{DET}}(R)$ in (10) is related to $D_{\mathrm{DET}}(f, R_s)$ in (29) as follows,

$$D_{\mathrm{DET}}(R) = \limsup_{N\to\infty} \inf_{\substack{f>0, R_s\geq 1: \\ fR_s \leq R}} D_N(f, R_s). \tag{33}$$

Note that (24a) does not directly follow (33), since the right side of (24a) switches the order of lim sup and inf in (33).

We will use Lemmas 2 to 5 that follow to prove (24a).

Lemma 2. *$D_N(f, R_s)$ in (29b) is lower-bounded as*

$$D_N(f, R_s) \geq \underline{D}_N(f, R_s), \tag{34a}$$

$$\triangleq \inf_{\substack{T_0\geq 0, T_N\geq 0 \\ T_0+T_N\leq \frac{N}{f}}} \frac{f}{2}\left(\frac{T_0^2 + T_N^2 + 2\log_2 e\, \lambda^*(f, R_s, N)}{N} + \frac{N-1}{N} T^*(f,N) \sqrt{T^*(f,N)^2 + 4\log_2 e\, \lambda^*(f, R_s, N)} \right), \tag{34b}$$

where $T^*(f, N)$ is given by

$$T^*(f, N) \triangleq \frac{N}{f(N-1)} - \frac{T_0 + T_N}{N-1}, \tag{35}$$

and $\lambda^*(f, R_s, N) \geq 0$ is the unique solution to

$$z\left(D^{N*}\right) = 2R_s \tag{36}$$

with D^N in (31a) replaced by

$$D_i^* = \frac{-T_i + \sqrt{T_i^2 + 4\log_2 e\, \lambda^*(f, R_s, N)}}{2}, \ i = 1, \ldots, N-1, \tag{37a}$$

$$D_N^* = \frac{\lambda^*(f, R_s, N)\log_2 e}{T_N}, \tag{37b}$$

and T_i, $i = 1, \ldots, N-1$ in (31a) replaced by $T^*(f, N)$ (35).

Proof. Appendix C. □

Lemma 3. *$D_N(f, R_s)$ in (29b) is upper-bounded as*

$$D_N(f, R_s) \leq \bar{D}_N(f, R_s) \tag{38a}$$

$$\triangleq \frac{N}{f(N+1)^2} + \frac{\log_2 e\, \lambda^*(f, R_s, N) f}{N} + \frac{N-1}{2(N+1)} \sqrt{\left(\frac{N}{f(N+1)}\right)^2 + 4\log_2 e\, \lambda^*(f, R_s, N)}, \tag{38b}$$

where $\lambda^*(f, R_s, N) \geq 0$ is the unique solution to (36) with D^N in (31a) replaced by (37), and with T^N in (31a) replaced by

$$T_0 = T_1 = \cdots = T_N = \frac{N}{f(N+1)}. \tag{39}$$

Proof. Appendix D. □

Lemma 4. *$D_{\mathrm{DET}}(f, R_s)$ in (29a) is given by*

$$D_{\mathrm{DET}}(f, R_s) = \frac{1}{2f} + \frac{1}{f(2^{2R_s} - 1)}, \tag{40}$$

where (40) can be achieved by a uniform sampling policy with sampling intervals equal to $T_i = \frac{1}{f}$, $i = 0, 1, \ldots$

Proof. Appendix E. □

Lemma 5. *The equality in (24a) holds.*

Proof. Appendix F. □

To prove (24b), it remains to minimize $D_{\mathrm{DET}}(f, R_s)$ in (24a) over feasible f and R_s:

$$D_{\mathrm{DET}}(R) = \min_{R_s\geq 1} D_{\mathrm{DET}}\left(\frac{R}{R_s}, R_s\right) \tag{41a}$$

$$= D_{\mathrm{DET}}(R, 1) \tag{41b}$$

$$= \frac{1}{2R} + \frac{1}{3R} = \frac{5}{6R}, \tag{41c}$$

where (41a) holds because $D_{\mathrm{DET}}(f, R_s)$ in (40) decreases monotonically in f for any given $R_s \geq 1$, and (41b) holds because $D_{\mathrm{DET}}\left(\frac{R}{R_s}, R_s\right)$ increases monotonically as R_s increases in the range $R_s \geq 1$. Thus, $D_{\mathrm{DET}}(R)$ is achieved at $f = R$, $R_s = 1$. Note that $\frac{1}{2R}$ in (41c) comes from the sampling distortion and $\frac{1}{3R}$ comes from the causal IDRF for the discrete-time samples.

D. Proof of Theorem 2

From (41), we conclude that (22) holds. Using Lemma 4 and (24b), we conclude that the uniform sampling policy with sampling frequency R achieves $D_{\mathrm{DET}}(R)$.

V. Rate-constrained sampling with delays

In our communication scenario in Section I-A, the codewords are delivered from the encoder to the decoder without delay, and the distortion constraint (3) penalizes any delay at the encoder or the decoder. While those are realistic assumptions in some scenarios of remote tracking and control, in this section we consider how the achievable distortion-rate tradeoffs are affected if those assumptions are weakened.

A. Delay at the encoder and the decoder

In the scenario of encoding the entire process to preserve it for future, a large delay is permissible. In the extreme, the encoder and the decoder may wait until the end of the input process $\{W_t\}_{t=0}^T$ before coding. This corresponds to the classical scenario of non-causal (block) compression. The IDRF for this scenario is given by

$$D_{\mathrm{noncausal}}(R) = \lim_{T\to\infty} \inf_{\substack{P_{\{\hat{W}_t\}_{t=0}^T | \{W_t\}_{t=0}^T}: \\ \frac{1}{T} I(\{W_t\}_{t=0}^T; \{\hat{W}_t\}_{t=0}^T) \leq R}} \mathbb{E}\left[\frac{1}{T}\int_0^T (W_t - \hat{W}_t)^2 dt\right]. \tag{42}$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAC.2021.3071953, IEEE Transactions on Automatic Control

Using reverse water-filling over the power spectrum of the process, Berger [32] derived the IDRF for the Wiener process:

$$D_{\text{noncausal}}(R) = \frac{2\log_2 e}{\pi^2 R} \quad \text{bits/s.} \tag{43}$$

The IDRF (43) is a lower bound to its operational counterpart. As for the achievability, Berger showed that given a rate $R \geq 0$, and $\epsilon > 0$, there exists a code with rate $R+\epsilon$ that achieves the distortion $D_{\text{noncausal}}(R)+\epsilon$. Berger's coding scheme operates as follows [32]: the Wiener process is divided into successive time intervals of a large enough length of T seconds. For each interval, the Karhunen-Loève (KL) coefficients of the process are calculated, and at most $2^{T(R+\epsilon)}$ codewords are used to jointly encode these coefficients with the resulting MSE per second equal to $D_{\text{noncausal}}(R)+\epsilon$. In parallel with encoding the KL expansion coefficients, an integrating delta modulator is employed to encode each endpoint of the length-T intervals with MSE ϵ using ϵ bits on average.

Comparing $D_{\text{noncausal}}(R)$ in (43) with $D^{\text{op}}(R)$ in (20), we see that, surprisingly, the optimal zero-delay policy outperforms the best infinite delay one:

$$\frac{D^{\text{op}}(R)}{D_{\text{noncausal}}(R)} \approx 0.57. \tag{44}$$

This is because in zero-delay causal coding the timing information is free. Indeed, the decoder time-stamps the arrival of each codeword, and since the channel is delay-free, it knows the codeword-generating times. In classical noncausal (block) lossy compression, no encoder and decoder synchronization is assumed, and thus the encoder is tasked with encoding both the values of the Wiener process and the times corresponding to these values. In many operational scenarios of remote tracking and control, the encoder and decoder are naturally synchronized, providing free timing information. Since Berger's distortion-rate function in (43) does not take that into account, it cannot adequately characterize the fundamental information-theoretic limits in those scenarios.

B. Delay at the decoder

In the scenario of causal coding where a small delay is tolerable, e.g. speech communication, one can leverage both the free timing information and the coding delay to improve distortion-rate tradeoffs. A *one sample look-ahead decoder* waits for the next codeword $U_{\tau_{i+1}}$ before estimating W_t, $\tau_i \leq t < \tau_{i+1}$, thereby introducing a maximum average delay of $\mathbb{E}[\tau_{i+1}-\tau_i] = \frac{1}{R}$ at the decoder. As we are about to see, the one sample look-ahead decoder greatly reduces the MSE compared to the ODRF and IDRF under causal estimation.

With the encoding policy in Proposition 1, the decoder is permitted to estimate W_t at time t', $t \leq t' \leq T$ using not only the codewords received before time t, but also the extra codewords received during the time $[t,t']$. Using Wolf and Ziv's decomposition of the MSE in [28], the ODRF with decoder delay can be decomposed as

$$D^{\text{op}}_{\text{dec delay}}(R) = \limsup_{T\to\infty} \inf_{\substack{\pi_T\in\Pi_T \\ f_T\in F_T: \\ (2)}} \frac{1}{T}\mathbb{E}\Bigg[\sum_{i=0}^{N}\int_{\tau_i}^{\tau_{i+1}} (W_t-\bar{W}_t)^2 + \left(\bar{W}_t-\hat{W}_t\right)^2 dt\Bigg], \tag{45}$$

where $\bar{W}_t$ is the MMSE estimator of the process at the encoder using the samples and the sampling times: for $t \in [\tau_i, \tau_{i+1})$,

$$\bar{W}_t \triangleq \mathbb{E}[W_t|\{W_{\tau_j}\}_{j=1}^N, \tau^N] = \mathbb{E}[W_t|W_{\tau_i}, W_{\tau_{i+1}}, \tau_i, \tau_{i+1}], \tag{46}$$

where (46) holds because $W_t - (W_{\tau_i}, W_{\tau_{i+1}}\tau_i, \tau_{i+1}) - (\{W_{\tau_j}\}_{j=1}^{i-1}, \{W_{\tau_j}\}_{j=i+1}^{N}, \{\tau_j\}_{j=1}^{i-1}, \{\tau_j\}_{j=i+1}^{N})$ is a Markov chain. Given all the samples, $\bar{W}_t$ only depends on the previous sample and the next sample. In particular, when the samples are taken under a deterministic sampling policy, $(W_{\tau_i}, W_t, W_{\tau_{i+1}})$ is a Gaussian random vector, thus $\bar{W}_t$ in (46) is the linear interpolation between W_{τ_i} and $W_{\tau_{i+1}}$. The MMSE estimate $\hat{W}_t$ at the decoder of the process W_t using all the received information is given by

$$\hat{W}_t = \mathbb{E}[W_t|U^N, \tau^N] = \mathbb{E}[\bar{W}_t|U^N, \tau^N], \tag{47}$$

where (47) holds due to (46) and the Markov chain $W_t - (\{W_{\tau_j}\}_{j=1}^N, \tau^N) - U^N$. Since the one sample look-ahead decoder only waits until the next codeword U_{i+1} is received at τ_{i+1}, $\hat{W}_t$ is specified to $\mathbb{E}[\bar{W}_t|U^{i+1}, \tau^{i+1}]$ for $t \in [\tau_i, \tau_{i+1})$.

We append the one sample look-ahead decoder to the optimal encoding policy in Theorem 1 and calculate the resulting MSE. Under symmetric threshold sampling policies, the samples are not Gaussian, and the linear interpolation is suboptimal. Yet, if in (45) we substitute for $\bar{W}_t$ the suboptimal estimate $\frac{W_{\tau_{i+1}}+W_{\tau_i}}{2}$, then the resulting the MSE is equal to $\frac{1}{12R}$, a two-fold improvement over (20). We append the one sample look-ahead decoder to the uniform sampling policy in Theorem 2, and ignore the potential reduction in quantization distortion brought by the decoder's ability to look ahead. The resulting sampling distortion is $\frac{1}{T}\mathbb{E}\left[\sum_{i=0}^{N}\int_{\tau_i}^{\tau_{i+1}}(W_t-\bar{W}_t)^2\right] = \frac{1}{6R}$, a 3-fold improvement over the sampling distortion $\frac{1}{2R}$ (41c) causally attainable with a uniform sampling policy. Thus, the total MSE is at most $\frac{1}{2R}$, a 1.67-fold improvement over (22).

C. Delay at the channel

We consider the communication scenario in Figure 1 with a fixed channel delay between the codeword-generating time and the codeword-delivery time. We show that the optimal coding policy remains the SOI code in Theorem 1. We denote the channel delay by $\delta \geq 0$. If the sampling time is τ_i, the delivery time is $\tau_i+\delta$. The encoder and the decoder are clock-synchronized. The decoder knows the delivery time and the fixed delay, thus it knows the codeword-generating time. The distortion is however measured in real time as in Section I-A (3) rather than after a delay as in Sections V-A and V-B. The MSE that we aim to minimize under the rate constraint (2) is given by

$$D_{\text{ch}}(R) = \limsup_{T\to\infty} \inf_{\substack{\pi_T\in\Pi_T \\ f_T\in F_T: \\ (2)}} \frac{1}{T}\mathbb{E}\left[\sum_{i=0}^{N}\int_{\tau_i+\delta}^{\tau_{i+1}+\delta}(W_t-\hat{W}_t)^2 dt\right], \tag{48}$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAC.2021.3071953, IEEE Transactions on Automatic Control

where, similarly to [5] [10], we use the MMSE decoding policy given by

$$\hat{W}_t \triangleq \mathbb{E}[W_t|U^i, \tau^i],\ t \in [\tau_i + \delta, \tau_{i+1} + \delta). \tag{49}$$

Unlike Theorem 1 where we proved that conditioning on the event $t < \tau_{i+1}$ in the decoding policy (5) can be ignored to yield (6) without loss of optimality, here we do not delve into the issue of whether ignoring the known event $t < \tau_{i+1} + \delta$ in the conditional expectation (49) is optimal.

Proposition 2. *In causal coding of the Wiener process with a fixed channel delay δ and the decoding policy (49), the optimal causal code remains* the SOI coding scheme *in Theorem 1, and*

$$D_{\rm ch}(R) = \frac{1}{6R} + \delta. \tag{50}$$

Proof. Appendix G. □

The optimal sampling policy in the fixed-delay scenario coincides with the optimal sampling policy in the delay-free scenario. This differs from the result of [10], according to which the optimal causal sampling policy for the Wiener process through a channel with an i.i.d. delay Y_i is a symmetric threshold sampling policy:

$$\tau_{i+1} = \inf\{t + \tau_i + Y_i : |W_{t+\tau_i+Y_i} - W_{\tau_i}| \geq \beta\}, \tag{51}$$

where β is a threshold that depends on the distribution of Y_i and the sampling frequency constraint. The setting in [10] is different from ours in Section V-C, since the channel in [10] only serves one sample at a time. Because samples must wait in a queue before the previous sample is delivered, the optimal encoder in [10] takes a new sample after the previous sample is delivered, whereas in our setting, the encoder may take a new sample after or before the delivery of the previous sample. This results in the policy in [10] attaining a MSE in the constant-delay scenario no smaller than that indicated in (50). With the random delay, we cannot simply append an SOI compressor to the optimal symmetric threshold policy (51) to obtain the optimal rate-constrained code. Indeed, the innovation $W_{\tau_{i+1}} - W_{\tau_i}$ may not be a binary random variable for all $i = 0, 1, \ldots$ since waiting for the delivery of the previous sample may cause the thresholds not to be hit with equality at the time $\tau_i, i = 1, 2, \ldots$.

VI. Rate-constrained event-triggered control

The SOI coding scheme in Theorem 1 also applies to the rate-constrained event-triggered control scenario. The stochastic plant evolves according to

$$dX_t = Z_t dt + dW_t, \tag{52}$$

where W_t is the Wiener process and Z_t is the control signal generated by the controller. A sampler takes measurements of the plant X_t at a sequence of stopping times $\tau_1, \tau_2, \ldots$ adapted to the filtration generated by $\{X_t\}_{t=0}^T$. At time τ_i, the sampler outputs X_{τ_i}, and the compressor generates the codeword U_i based on causally received samples. At time τ_i, the controller uses the received codewords U^i to form an impulse control signal Z_{τ_i}. The communication between the compressor and the controller is subject to the bitrate constraint (2). We aim to find the optimal sampling and compressing policies that minimize the mean-square deviation of X_t from target state 0,

$$\limsup_{T\to\infty} \frac{1}{T}\mathbb{E}\left[\int_0^T X_t^2 dt\right]. \tag{53}$$

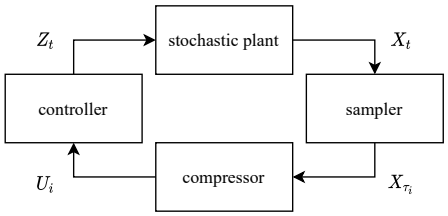


Fig. 5: Real-time communication over a channel with feedback

We restrict our control signal to be the impulse control as in [1],[39]. The impulse control only takes action at the stopping times (1) decided by the sampling policy, i.e., $Z_t \neq 0$ if and only if $t = \tau_1, \tau_2, \ldots$. The impulse control leads to

$$X_{\tau_i^+} = X_{\tau_i} + Z_{\tau_i},\ i = 0, 1, \ldots \tag{54}$$

where τ_i^+ represents the time just after τ_i [40].

Theorem 4. *In the rate-constrained event-triggered control system, the causal code that minimizes (53) is the SOI coding scheme in Theorem 1, and the optimal impulse control signal is*

$$Z_{\tau_i} = -(W_{\tau_i} - W_{\tau_{i-1}}), i = 0, 1, \ldots, \tag{55}$$

The minimum mean-square cost (53) is equal to $\frac{1}{6R}$.

Proof. Appendix H. □

In contrast, using a uniform sampler in this rate-constrained control system is suboptimal, since the resultant mean-square cost in (53) consists of sampling and compressing distortions and the distortion due to uniform sampling is equal to $\frac{1}{2R}$.

VII. Conclusion

The results in this paper contribute to the rich literature on optimal scheduling and remote estimation problems by introducing the transmission rate constraint beyond the popular sampling frequency constraint. The SOI coding scheme is optimal for causal estimation of the Wiener process under the rate constraint (Theorem 1). The performance of the SOI coding scheme is better than that achieved by the best non-causal code (Section V-A). This underscores the power of free information contained in the codeword arrival times that is not considered in the standard setting of non-causal compression. The SOI scheme remains optimal even if the channel introduces a fixed delay (Proposition 2). The key to transmitting information via timing is to use process-dependent rather than deterministic

sampling time stamps, because the latter contain zero information. The optimal deterministic sampling policy is uniform (Theorem 2). In either setting, the best strategy is to transmit 1-bit codewords as frequently as possible (Theorem 3). This is a consequence of the real-time distortion constraint (3). If a short delay is affordable, the MSE can be further reduced with a one sample look-ahead at the decoder (Section V-B). The SOI coding scheme also minimizes the mean-square cost of a stochastic plant driven by the Wiener process, and controlled via impulse control (Theorem 4). In our paper [37], we derive the optimal rate-constrained sampling policy for a class of continuous Markov processes satisfying symmetry conditions that the Wiener process meets, and we prove that the SOI code proposed in this paper remains optimal. Theorem 1 shows that the optimal threshold for the Wiener process is equal to $\sqrt{\frac{1}{R}}$. In [37], we show how to match the threshold to the statistics of the source more generally.

REFERENCES

[1] K. J. Åström and B. M. Bernhardsson, "Comparison of Riemann and Lebesgue sampling for first order stochastic systems," *Proceedings of the 41st IEEE Conference on Decision and Control*, Las Vegas, NV, USA, pp. 2011-2016 vol. 2, Dec. 2002.

[2] 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.

[3] 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, Sep. 2011.

[4] 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.

[5] 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.

[6] 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.

[7] 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.

[8] 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.

[9] 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, Sep. 2013.

[10] 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, June 2017.

[11] X. Gao, E. Akyol and T. Başar, "Optimal estimation with limited measurements and noisy communication," *2015 54th IEEE Conference on Decision and Control*, Osaka, Japan, pp. 1775-1780, Dec. 2015.

[12] X. Ren, J. Wu, K. H. Johansson, G. Shi, and L. Shi, "Infinite horizon optimal transmission power control for remote state estimation over fading channels," in *IEEE Transactions on Automatic Control*, vol. 63, no. 1, pp. 85-100, Jan. 2018.

[13] J. Chakravorty and A. Mahajan, "Remote estimation over a packet-drop channel with Markovian state", in *IEEE Transactions on Automatic Control*, June 2019.

[14] A. Kipnis, A. J. Goldsmith and Y. C. Eldar, "The distortion-rate function of sampled Wiener processes," in *IEEE Transactions on Information Theory*, vol. 65, no. 1, pp. 482-499, Jan. 2019.

[15] G. N. Nair, F. Fagnani, S. Zampieri and R. J. Evans, "Feedback Control Under Data Rate Constraints: An Overview," in *Proceedings of the IEEE*, vol. 95, no. 1, pp. 108-137, Jan. 2007.

[16] 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.

[17] 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.

[18] 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.

[19] 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.

[20] 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.

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

[22] 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.

[23] 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.

[24] 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.

[25] M. Sun, L. Huang, S. Wang, C. Mao and W. Xie, " Quantized control of event-triggered networked systems with time-varying delays," *Journal of the Franklin Institute*, May 2018.

[26] 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, June 2016.

[27] P. Tallapragada, M. Franceschetti and J. Cortés, "Event-triggered control under time-varying rates and channel blackouts,"in *IFAC Journal of Systems and Control*, Aug. 2019.

[28] J. K. Wolf and J. Ziv, "Transmission of noisy information to a noisy receiver with minimum distortion" in *IEEE Transactions on Information Theory*, vol. 16, no. 4, pp. 406-411, Jul. 1970.

[29] T. Tanaka, K. K. Kim, P. A. Parrilo and S. K. Mitter, "Semidefinite programming approach to Gaussian sequential rate-distortion trade-offs," in *IEEE Transactions on Automatic Control*, vol. 62, no. 4, pp. 1896-1910, Apr. 2017.

[30] G. R. Grimmett and D. R. Stirzaker, *Probability and Random Processes*. Oxford University Press, 2009.

[31] J. Massey, "Causality, feedback and directed information," *Proceedings International Symposium on Information Theory and its Applications*, pp. 303-305, Nov. 1990. quadratic rate-distortion function for Gaussian stationary sources," in *IEEE Transactions on Information Theory*, vol. 58, no. 5, pp. 3131-3152, May 2012.

[32] T. Berger, "Information rates of Wiener processes," in *IEEE Transactions on Information Theory*, vol. 16, no. 2, pp. 134-139, Mar. 1970.

[33] P. A. Stavrou, T. Charalambous, C. D. Charalambous, and S. Loyka, "Optimal estimation via nonanticipative rate distortion function and applications to time-varying Gauss-Markov processes," in *SIAM Journal on Control and Optimization*, pp. 3731-3765, Oct. 2018.

[34] V. Kostina and B. Hassibi, "Rate-cost tradeoffs in control," in *IEEE Transactions on Automatic Control*, vol. 64, no. 11, pp. 4525-4540, Nov. 2019.

[35] A. Braides, *-convergence for beginners*. Oxford University Press, 2002.

[36] N. Guo and V. Kostina, "Optimal causal rate-constrained sampling of the Wiener process", in *Proceedings 57th Annual Allerton Conference on Communication, Control and Computing*, pp. 1090-1097, Monticello, IL, Sep. 2019.

[37] N. Guo and V. Kostina, "Optimal causal rate-constrained sampling for a class of continuous Markov processes", in *IEEE International Symposium on Information Theory*, Los Angeles, CA, June 2020.
[38] N. Guo and V. Kostina, "Optimal causal rate-constrained sampling for a class of continuous Markov processes", *Arxiv preprint*, https://arxiv.org/abs/2002.01581, Apr. 2021.
[39] A. Bensoussan and J.-L. Lions, *Impulse control and quasi-variational inequalities*, Gauthier-Villars, Paris, 1984.
[40] Z. G. Li, C. Y. Wen, and Y. C. Soh, "Analysis and Design of Impulsive Control Systems", in *IEEE Transactions on Automatic Control*, pp. 894-897, vol. 46, No. 6, June 2001.
[41] D. L. McLeish, *Monte Carlo Simulation and Finance*, John Wiley and Sons, 2005.

APPENDIX

A. Proof of Proposition 1

The objective function in (7) decomposes as

$$\frac{1}{T}\mathbb{E}\left[\sum_{i=0}^{N}\int_{\tau_i}^{\tau_{i+1}}(W_t-\hat{W}_{\tau_i})^2dt\right] \tag{56a}$$

$$=\frac{1}{T}\mathbb{E}\left[\sum_{i=0}^{N}\int_{\tau_i}^{\tau_{i+1}}(W_t-W_{\tau_i})^2dt\right]+ \tag{56b}$$

$$\frac{1}{T}\mathbb{E}\left[\sum_{i=0}^{N}(\tau_{i+1}-\tau_i)(W_{\tau_i}-\hat{W}_{\tau_i})^2dt\right]+$$

$$\frac{1}{T}\mathbb{E}\left[\sum_{i=0}^{N}(W_{\tau_i}-\hat{W}_{\tau_i})\int_{\tau_i}^{\tau_{i+1}}(W_t-W_{\tau_i})dt\right]$$

$$=\frac{1}{T}\mathbb{E}\left[\sum_{i=0}^{N}\int_{\tau_i}^{\tau_{i+1}}(W_t-W_{\tau_i})^2dt\right]+ \tag{56c}$$

$$\frac{1}{T}\mathbb{E}\left[\sum_{i=0}^{N}(\tau_{i+1}-\tau_i)(W_{\tau_i}-\hat{W}_{\tau_i})^2dt\right],$$

where (56a) uses the simplified decoding policy (6) that is justified in the proof of Theorem 1; (56b) is obtained by substituting $W_t-W_{\tau_i}+W_{\tau_i}-\hat{W}_{\tau_i}$ for the term $W_t-\hat{W}_{\tau_i}$ in (56a), and (56c) holds due to the fact that $\int_{\tau_i}^{\tau_{i+1}}(W_t-W_{\tau_i})dt$ is orthogonal to $W_{\tau_i}-\hat{W}_{\tau_i}$ for all $i=0,1,2,\ldots,N$. Since the encoder only influences the second term in (56c), we move the minimization over the encoder f_T in (7) directly in front of the second term in (56c).

To show that f_i only encodes $W_{\tau_i}-\hat{W}_{\tau_{i-1}}$ given U^{i-1} and τ^i, we first recall a well-known fact. Consider the following source coding model in Figure 6, where $X\in\mathcal{X}$ and $Y\in\mathcal{Y}$ are available only at the encoder, C is the common information, $\hat{X}\in\hat{\mathcal{X}}$ is the reproduction. Encoder $P_{U|X,Y,C}$ and decoder $P_{\hat{X}|U,C}$ aim to achieve a given distortion $d=\mathbb{E}\left[\mathsf{d}(X,\hat{X})\right]$, where $\mathsf{d}\colon\mathcal{X}\times\hat{\mathcal{X}}\to\mathbb{R}^+$ is the distortion measure, subject to a constraint on the cardinality of the alphabet $\mathcal{U}$ of U. Since

$$\mathbb{E}\left[\mathsf{d}(X,\hat{X})\right]=\mathbb{E}[\mathbb{E}[\mathsf{d}(X,\hat{X})|U,C]], \tag{57}$$

and $(X,Y)-(U,C)-\hat{X}$, the knowledge of side information Y is useless at the encoder, i.e., for any encoder-decoder pair $\left(P_{U|X,Y,C},\ P_{\hat{X}|U,C}\right)$, the pair $\left(P_{U|X,C},P_{\hat{X}|U,C}\right)$, where $P_{U|X,C}$ is the marginal of $P_{U|X,Y,C}P_{Y|X,C}$, achieves the same expected distortion.

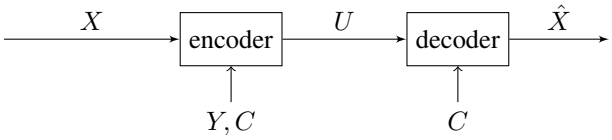


Fig. 6: Y only available at the encoder, C available at both the encoder and the deocder

In our problem, at time τ_1, we take $X=W_{\tau_1}$, $Y=\{W_t\}_{0<t<\tau_1}$, $C=\tau_1$, and $d(X,\hat{X})=(X-\hat{X})^2$. To achieve a given sample distortion $\mathbb{E}\left[(W_{\tau_1}-\hat{W}_{\tau_1})^2\right]$, the random compressing policy needs to only take into account W_{τ_1} and τ_1. Inductively, at time τ_i, the encoder knows $\{W_t\}_{t=0}^{\tau_i}$. Both the encoder and the decoder know U^{i-1} and τ^i. Since $\hat{W}_{\tau_{i-1}}$ is known once U^{i-1} and τ^{i-1} are given, we have

$$\mathbb{E}[(W_{\tau_i}-\hat{W}_{\tau_i})^2]= \\ \mathbb{E}\left[\left(W_{\tau_i}-\hat{W}_{\tau_{i-1}}-\mathbb{E}\left[W_{\tau_i}-\hat{W}_{\tau_{i-1}}|U^i,\tau^i\right]\right)^2\right]. \tag{58}$$

In (57), we take $X=W_{\tau_i}-\hat{W}_{\tau_{i-1}}$, $\hat{X}=\mathbb{E}[W_{\tau_i}-\hat{W}_{\tau_{i-1}}|U^i,\tau^i]$, $U=U_i$, $C=\{U^{i-1},\tau^i\}$, and Y is everything known at the encoder excluding X and C. It follows that for the purpose of achieving the sample distortion $\mathbb{E}[W_{\tau_i}-\hat{W}_{\tau_i}]^2$, at time τ_i, the randomized compressing policy needs to only take into account $W_{\tau_i}-\hat{W}_{\tau_{i-1}}$, U^{i-1}, and τ^i.

B. Proof of Lemma 1

We denote by $\tilde{D}_N(R_s)$ (17b) the IDRF for discrete-time samples of the Wiener process

$$W_{\tau_{i+1}}=W_{\tau_i}+V_{\tau_i},\ V_{\tau_i}\sim\mathcal{N}(0,T_i). \tag{59}$$

Using the representation of its dual in [29, Eq. (18)] derived using a semi-definite programming approach, we write

$$\tilde{D}_N(R_s)=\inf_{\substack{D_i\geq 0,\ i=1,\ldots,N:\\ D_{i-1}+T_{i-1}\geq D_i,\quad i=1,2,\cdots,N,\\ \frac{1}{N}\left(\sum_{i=1}^{N}\frac{1}{2}\log_2(D_{i-1}+T_{i-1})-\frac{1}{2}\log_2 D_i\right)\leq R_s.}}\ \sum_{i=1}^{N}T_iD_i. \tag{60}$$

Since the sampling intervals T^N are deterministic, we calculate the summand in (17a) as

$$\mathbb{E}\left[\int_{\tau_i}^{\tau_{i+1}}(W_t-W_{\tau_i})^2dt\right]=\mathbb{E}\left[\int_0^{T_i}W_t^2dt\right]=\frac{T_i^2}{2}. \tag{61}$$

Plugging (60) and (61) into (17), we write

$$D_{\mathrm{DET}}(f,R_s)=\limsup_{T\to\infty}\ \inf_{\pi_T\in\Pi_T^{\mathrm{DET}}\ (12)}\frac{1}{T}\left(\sum_{i=0}^{N}\frac{T_i^2}{2}+\tilde{D}_N(R_s)\right). \tag{62}$$

Note that as $T\to\infty$, the number of samples N must increase no slower than $\sqrt{T}$. Since the largest sampling interval satisfies $\max_{i=0,\ldots,N}T_i\geq\frac{T}{N+1}$, the summand in (62) $\frac{\max_i T_i^2}{2T}\geq\frac{T}{2(N+1)^2}$ will blow up to infinity if N increases slower than $\sqrt{T}$. Thus, we can replace the $\limsup_{T\to\infty}$ in

(62) by $\limsup_{N\to\infty}$, replace T in (62) by its equivalent $\frac{f}{N}$ in (12), and replace the minimization constraint (12) in (62) by its equivalent (30).

C. Proof of Lemma 2

We split $D_N(f,R_s)$ into layered optimization problems:

$$D_N(f,R_s) \triangleq \inf_{\substack{T_0\geq 0, T_N\geq 0:\\ T_0+T_N\leq \frac{N}{f}}} D_N(f,R_s,T_0,T_N), \tag{63a}$$

$$D_N(f,R_s,T_0,T_N) \triangleq \min_{\substack{T_1,\ldots,T_{N-1}\geq 0:\\ \frac{1}{N}\sum_{i=1}^{N-1}T_i=\frac{1}{f}-\frac{T_0+T_N}{N}}} \frac{f}{N}\left(\sum_{i=0}^{N}\frac{T_i^2}{2}+D_N\left(f,R_s,T^N\right)\right), \tag{63b}$$

$$D_N\left(f,R_s,T^N\right) \triangleq \min_{\substack{D^N\geq 0:\\ (31)}} \sum_{i=1}^{N} T_iD_i. \tag{63c}$$

We denote by $\underline{D}_N(f,R_s,T^N)$ the lower bound to $D_N(f,R_s,T^N)$ obtained by deleting the minimization constraint (31b) in (63c), i.e.,

$$\underline{D}_N(f,R_s,T^N) \triangleq \min_{\substack{D^N\geq 0:\\ (31a)}} \sum_{i=1}^{N} T_iD_i, \tag{64}$$

We denote by $\underline{D}_N(f,R_s,T_0,T_N)$ the corresponding lower bound to $D_N(f,R_s,T_0,T_N)$ in (63b):

$$\underline{D}_N(f,R_s,T_0,T_N) \triangleq \min_{\substack{T_1,\ldots,T_{N-1}\geq 0:\\ \frac{1}{N}\sum_{i=1}^{N-1}T_i=\frac{1}{f}-\frac{T_0+T_N}{N}}} \frac{f}{N}\left(\sum_{i=0}^{N}\frac{T_i^2}{2}+\underline{D}_N\left(f,R_s,T^N\right)\right). \tag{65}$$

We calculate the corresponding lower bound to $D_N(f,R_s)$:

$$\underline{D}_N(f,R_s) \triangleq \min_{\substack{T_0\geq 0, T_N\geq 0:\\ T_0+T_N\leq \frac{N}{f}}} \underline{D}_N(f,R_s,T_0,T_N). \tag{66}$$

We first show that the optimization problem in the right side of (64) is a convex optimization problem that satisfies Slater's condition, i.e., strong duality holds. Then, we solve its Lagrangian dual problem to get the optimal $D_1^*\ldots,D_N^*$ in (37) that achieve the minimum in the right side of (64), where $\lambda^*(f,R_s,N)\geq 0$ is the unique solution to (36).

The objective function $\sum_{i=1}^{N}T_iD_i$ (64) is an affine function in D^N. Furthermore, $z\left(D^N\right)$ is a convex function since

$$\frac{\partial^2 z\left(D^N\right)}{\partial D_i^2}=\frac{\log_2 eT_i(2D_i+T_i)}{N(D_i^2+D_iT_i)^2}\geq 0,\ \forall i=1,\ldots,N-1, \tag{67a}$$

$$\frac{\partial^2 z\left(D^N\right)}{\partial D_N^2}=\frac{\log_2 e}{ND_N^2}\geq 0, \tag{67b}$$

$$\frac{\partial^2 z\left(D^N\right)}{\partial D_i\partial D_j}=0,\ \forall i,j=1,\ldots,N. \tag{67c}$$

Therefore, the minimization problem in the right side of (64) is convex. Notice that $z(D,D,\ldots,D)$ decreases from $+\infty$ to $-\infty$ as D increases from 0 to ∞. Thus, there exists a $\tilde{D}\geq 0$ such that Slater's condition is satisfied, i.e.,

$$z\left(\tilde{D},\tilde{D},\ldots,\tilde{D}\right)<2R_s. \tag{68}$$

We conclude that 1) the strong duality holds, 2) $\underline{D}(f,R_s,T^N)$ can be obtained via its Lagrangian dual problem, and 3) there must exist an optimal Lagrangian multiplier $\lambda^*(f,R_s,N)\geq 0$ that satisfies the complementary slackness (36) in the KKT conditions. Indeed, (36) always has a non-negative solution $\lambda^*(f,R_s,N)$, since as a function of $\lambda^*(f,R_s,N)$, $z\left(D^{N*}\right)$ is continuous and monotonically decreasing from $+\infty$ to $-\infty$ as $\lambda^*(f,R_s,N)$ increases from 0 to $+\infty$.

D^{N*} in (37) is the solution to the Lagrangian function:

$$\min_{D^N}\sum_{i=1}^{N}T_iD_i+\lambda^*(f,R_s,N)\left(\sum_{i=1}^{N-1}\log_2\left(1+\frac{T_i}{D_i}\right)+\log_2\left(\frac{T_0}{D_N}\right)-2R_s\right), \tag{69}$$

where (37) is obtained by taking derivatives of the objective function of (69) with respect to each $D_i, i=1,2\ldots,N$.

Plugging D^{N*} (37) into (64), we obtain $\underline{D}_N\left(f,R_s,T^N\right)$ and proceed to evaluate $\underline{D}_N(f,R_s,T_0,T_N)$ in (65), which is

$$\underline{D}_N(f,R_s,T_0,T_N)=\min_{\substack{T_1,\ldots,T_{N-1}\geq 0:\\ \frac{1}{N}\sum_{i=1}^{N-1}T_i=\frac{1}{f}-\frac{T_0+T_N}{N}}} g(T_1,\ldots,T_{N-1}), \tag{70}$$

where

$$g(T_1,\ldots,T_{N-1})\triangleq\frac{f}{2N}\left[T_0^2+T_N^2+2\log_2 e\,\lambda^*(f,R_s,N)+\sum_{i=1}^{N-1}T_i\sqrt{T_i^2+4\log_2 e\,\lambda^*(f,R_s,N)}\right]. \tag{71}$$

We make use of the *Schur-convexity* of (71) to calculate $\underline{D}_N(f,R_s,T_0,T_N)$. Assume that a function $f(x^d)$ is symmetric, and its first partial derivative with respect to each x_i, $i=1,\ldots,d$ exists. Then, $f(x^d)$ is Schur-convex if and only if

$$(x_i-x_j)\left(\frac{\partial f(x^d)}{\partial x_i}-\frac{\partial f(x^d)}{\partial x_j}\right)\geq 0,\ \forall\, i,j=1,\ldots,d. \tag{72}$$

It is clear that $g(T_1,\ldots,T_{N-1})$ is symmetric since it is invariant to the permutations of $T_1,\ldots,T_{N-1}$. To calculate the partial derivatives of (71), we first compute the implicit differentiation $\frac{\partial\lambda^*(f,R_s,N)}{\partial T_i}$ by taking the derivative with respect to T_i on the both sides of (36), yielding

$$\frac{\partial\lambda^*(f,R_s,N)}{\partial T_i}=\frac{1}{\sqrt{T_i^2+4\log_2 e\,\lambda^*(f,R_s,N)}}\cdot\frac{2\lambda^*(f,R_s,N)}{1+\sum_{k=1}^{N-1}\frac{T_k}{\sqrt{T_k^2+4\log_2 e\,\lambda^*(f,R_s,N)}}}. \tag{73}$$

Using (73) to compute the first partial derivative, we obtain

$$\frac{\partial g(T_1,\ldots,T_{N-1})}{\partial T_i}=\frac{f}{N}\sqrt{T_i^2+4\log_2 e\,\lambda^*(f,R_s,N)}. \tag{74}$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAC.2021.3071953, IEEE Transactions on Automatic Control

Using (74), we can verify that $g(T_1,\dots,T_{N-1})$ satisfies (72). Therefore, $g(T_1,\dots,T_{N-1})$ is a Schur-convex function.

Let $x=(x_1,\dots,x_d)\in\mathbb{R}^d$, $y=(y_1,\dots,y_d)\in\mathbb{R}^d$ be two non-increasing sequences of real numbers. Recall that x is *majorized* by y if for each $k=1,\dots,d$, $\sum_{i=1}^k x_i\le\sum_{i=1}^k y_i$ with equality if $k=d$. For a Schur-convex function f, if x is majorized by y, then $f(x)\le f(y)$. In our case, the feasible T_i's must satisfy the constraint in (70). Any sequence $T_1,\dots,T_{N-1}$ that satisfies the constraint in (70) majorizes the sequence in (35). Thus, the infimum in (70) is achieved by the sequence $T_1^*,\dots,T_{N-1}^*$ in (35). Finally, plugging $T_1^*,\dots,T_{N-1}^*$ (35) into (70), and further plugging (70) into the right side of (66), we complete the proof.

D. Proof of Lemma 3

Plugging (39) into (37), we obtain:

$$D_1^*=\cdots=D_{N-1}^*=\frac{-\frac{N}{f(N+1)}+\sqrt{\left(\frac{N}{f(N+1)}\right)^2+4\log_2 e\,\lambda^*(f,R_s,N)}}{2}, \tag{75a}$$

$$D_N^*=\frac{f(N+1)}{N}\log_2 e\,\lambda^*(f,R_s,N), \tag{75b}$$

where $\lambda^*(f,R_s,N)$ is defined in Lemma 3. We first show that the T^N in (39) and the D^N in (75) satisfy the deleted constraint (31b). Then, we plug T^N (39) and D^N (75) as feasible solutions into the minimization problem associated with $D_N(f,R_s)$ in (29b) to obtain the upper bound in (38).

For $i=2,\dots,N-1$, the deleted constraint (31b) is satisfied trivially, since $D_{i-1}=D_i$ and $T_{i-1}\ge 0$. To prove that the deleted constraint (31b) also holds at $i=1$ and N, we upper bound $\lambda^*(f,R_s,N)$ for every $N>2$. If

$$T_1=\cdots=T_{N-1}, \tag{76}$$

we can rearrange terms in the complementary slackness condition (36) and conclude $x=\lambda^*(f,R_s,N)\log_2 e$ is the unique solution to the following equation,

$$h_N(T_0,T_N,T_1,R_s,x)-x=0, \tag{77}$$

where $h_N(T_0,T_N,T_1,R_s,x)$ is defined to be equal to

$$\frac{T_1^2}{2^{2R_s+\frac{2}{N-1}R_s-\frac{\log_2 T_0+\log_2 T_N}{N-1}+\frac{\log_2 x}{N-1}}-1}+\left(\frac{T_1}{2^{2R_s+\frac{2}{N-1}R_s-\frac{\log_2 T_0+\log_2 T_N}{N-1}+\frac{\log_2 x}{N-1}}-1}\right)^2. \tag{78}$$

The left side of (77) monotonically decreases as x increases.

Given R_s and plugging (39) into (77), we conclude that the $\lambda^*(f,R_s,N)$ in Lemma 3 is the unique solution to

$$h_N\left(\frac{N}{f(N+1)},\frac{N}{f(N+1)},\frac{N}{f(N+1)},R_s,x\right)-x=0, \tag{79}$$

Plugging

$$x=\frac{N^2}{2f^2(N+1)^2} \tag{80}$$

into (79), we observe that the left side of (79) is ≤ 0 for all $N>2$. Thus, we conclude

$$\lambda^*(f,R_s,N)\log_2 e\le\frac{N^2}{2f^2(N+1)^2},\ \forall\ N>2. \tag{81}$$

Plugging (81) into (75), we obtain

$$D_1^*\le\sqrt{\lambda^*(f,R_s,N)\log_2 e}\le\frac{N}{f(N+1)}, \tag{82a}$$

$$D_N^*\le\frac{N}{2f(N+1)}. \tag{82b}$$

Substituting (39) and (82) into (31b), we conclude that (31b) holds for $i=1$ and $i=N$. Now, we can plug (39) and (75) as feasible solutions into (29b) to obtain the right side of (38).

E. Proof of Lemma 4

From Lemmas 2 and 3, and (29a), we have

$$\liminf_{N\to\infty}\underline{D}_N(f,R_s)\le D_{\mathrm{DET}}(f,R_s)\le\limsup_{N\to\infty}\bar{D}_N(f,R_s). \tag{83}$$

We show that both bounds are equal to the right side of (40).

To compute the lower bound in (83), we need to understand the behavior of $T^*(f,N)$, $\lambda^*(f,R_s,N)$, and T_0^*, T_N^* as N goes to infinity, where T_0^*, T_N^* achieve the minimum of the left side of (83). T_0^* and T_N^* must increase as

$$T_0^*+T_N^*=O\left(\sqrt{N}\right), \tag{84}$$

or $\frac{T_0^{*2}+T_N^{*2}}{N}$ in (34b) will blow up to infinity as $N\to\infty$. Substituting (84) to (35), we obtain

$$T^*(f,N)=\frac{1}{f}+O\left(\frac{1}{\sqrt{N}}\right). \tag{85}$$

We proceed to compute

$$\lambda^*\triangleq\lim_{N\to\infty}\lambda^*(f,R_s,N). \tag{86}$$

For given T_0^*, T_N^*, and R_s, $x=\lambda^*(f,R_s,N)\log_2 e$ is the unique solution to (77) with T_0, T_N, and $T(N)$ replaced by T_0^*, T_N^*, and $T^*(f,N)$ in (35). We prove that

$$\lambda^*\log_2 e\ge\frac{1}{2^{2R_s}f^2}, \tag{87a}$$

$$\lambda^*\log_2 e\le\frac{1}{2f^2}. \tag{87b}$$

We substitute (84) and (85) into the left side of (77) and take $\lim_{N\to\infty}$ to conclude that

$$\lim_{N\to\infty}h_N\left(T_0^*,T_N^*,T^*(f,N),R_s,\frac{1}{2f^2}\right)-\frac{1}{2f^2}\le 0. \tag{88}$$

Since the left side of (77) is monotonically decreasing in x, we conclude (87a) holds. To prove (87b), we similarly compute

$$\lim_{N\to\infty}h_N\left(T_0^*,T_N^*,T^*(f,N),R_s,\frac{1}{2^{2R_s}f^2}\right)-\frac{1}{2^{2R_s}f^2}\ge 0. \tag{89}$$

Via the squeeze theorem, (87) implies

$$\lambda^*(f,R_s,N)=O(1). \tag{90}$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAC.2021.3071953, IEEE Transactions on Automatic Control

Plugging (84), (85), and (90) into (77), and taking $N \to \infty$ on both sides of (77), we obtain

$$\lambda^* \log_2 e = \frac{1}{f^2(2^{2R_s}-1)^2} + \frac{1}{f^2(2^{2R_s}-1)}. \tag{91}$$

Plugging (84), (85), and (91) into (34b) and taking $\lim_{N\to\infty}$, we compute that $\lim_{N\to\infty} \underline{D}_N(f, R_s)$ is equal to

$$\frac{1}{2f} + \frac{1}{f(2^{2R_s}-1)} + \lim_{N\to\infty} \inf_{\substack{T_0\geq 0, T_N \geq 0 \\ T_0+T_N \leq \frac{N}{f}}} \frac{f}{2}\left(\frac{T_0^2+T_N^2}{N}\right) \tag{92a}$$

$$= \frac{1}{2f} + \frac{1}{f(2^{2R_s}-1)}, \tag{92b}$$

where 0 is achieved in the last term of (92a) by choosing any pair of $T_0, T_N \geq 0$ that satisfies

$$T_0 + T_N = o\left(\sqrt{N}\right). \tag{93}$$

We choose T_0 and T_N in (39) that satisfy (93), such that together with $T_1, \ldots, T_{N-1}$ in (39), the lower bound of $D_{\mathrm{DET}}(f, R_s)$ in (83) is achieved.

Now, we compute the upper bound in the right side of (83). $\lambda^*(f, R_s, N)\log_2 e$ in (38b) is the unique solution to (77). Note that (91) holds for any T_0 and T_N that satisfy (84). Since T_0 and T_N in (39) satisfy (84), we conclude that the $\lim_{N\to\infty}$ of $\lambda^*(f, R_s, N)\log_2 e$ in (38b) is also equal to (91). Plugging (91) into (38b) and taking $\limsup_{N\to\infty}$, we calculate that the upper bound of $D_{\mathrm{DET}}(f, R_s)$ in (83) is equal to (92b).

Furthermore, we observe that the uniform sampling intervals (39) achieving both the upper and the lower bound of $D_{\mathrm{DET}}(f, R_s)$, converge to $\frac{1}{f}$ asymptotically. We conclude that the uniform sampling policy with the sampling interval $\frac{1}{f}$ achieves $D_{\mathrm{DET}}(f, R_s)$.

F. Proof of Lemma 5

The max-min inequality and (33) imply that

$$D_{\mathrm{DET}}(R) \leq \min_{\substack{f>0, R_s\geq 1: \\ fR_s \leq R}} \limsup_{N\to\infty} \bar{D}_N(f, R_s). \tag{94}$$

On the other hand,

$$D_{\mathrm{DET}}(R) \geq \lim_{N\to\infty} \inf_{\substack{f>0, R_s\geq 1: \\ fR_s \leq R}} \underline{D}_N(f, R_s) \tag{95a}$$

$$= \inf_{\substack{f>0, R_s\geq 1: \\ fR_s \leq R}} \lim_{N\to\infty} \underline{D}_N(f, R_s), \tag{95b}$$

where (95a) is by (33), and (95b) will be proved in the sequel. Using (83) with both bounds equal to each other, (94), and (95), we complete the proof of Lemma 5.

We proceed to prove (95b) via *the fundamental theorem of* Γ*-convergence*. Let $\mathcal{X}$ be a topological space and $G_N : \mathcal{X} \to [0, +\infty]$, $N = 1, 2, \ldots$, be a sequence of functions defined on $\mathcal{X}$. A sequence of functions G_N, $N = 1, 2, \ldots$ Γ*-converges* [35] to its Γ*-limit* $G\colon \mathcal{X} \to [0, +\infty]$ if:

(i) For every $x \in \mathcal{X}$, and for every sequence $x_N \in \mathcal{X}, N = 1, 2, \ldots$ converging to x,

$$G(x) \leq \liminf_{N\to\infty} G_N(x_N). \tag{96}$$

(ii) For every $x \in \mathcal{X}$, there exists a sequence $x_N \in \mathcal{X}, N = 1, 2, \ldots$ converging to x such that

$$G(x) \geq \limsup_{N\to\infty} G_N(x_N). \tag{97}$$

A sequence of functions G_N, $N = 1, 2, \ldots$ is *equicoercive* [35] if there exists a compact set $\mathcal{K}$ independent of N s.t.

$$\inf_{x\in\mathcal{X}} G_N(x) = \inf_{x\in\mathcal{K}} G_N(x). \tag{98}$$

The fundamental theorem of Γ-convergence [35] says that if G_N is equicoercive and Γ-converges to $G\colon \mathcal{X} \to [0, +\infty]$, then

$$\min_{x\in\mathcal{X}} G(x) = \lim_{N\to\infty} \inf_{x\in\mathcal{X}} G_N(x). \tag{99}$$

We will show that for any scalars $f > 0$, $R_s \geq 1$ and for any sequences $f_{(N)} \to f$, $R_{s(N)} \to R_s$, we have

$$\lim_{N\to\infty} \underline{D}_N(f_{(N)}, R_{s(N)}) = D_{\mathrm{DET}}(f, R_s), \tag{100}$$

which means in particular that $D_{\mathrm{DET}}(\cdot,\cdot)$ is the Γ-limit of $\underline{D}_N(\cdot,\cdot)$. We will also prove that $\underline{D}_N(f, R_s)$ is equicoercive, and (95b) will follow via the fundamental theorem of Γ-convergence. By verifying that the reasoning in (84)-(92) goes through replacing f and R_s by $f_{(N)}$ and $R_{s(N)}$ respectively, we conclude that (100) holds.

It remains to prove that $\underline{D}_N(f, R_s)$ is equicoercive. Ignoring the two non-negative $\lambda^*(f, R_s, N)$ terms in (34b), we observe that $\underline{D}_N(f, R_s)$ is lower bounded by

$$\inf_{\substack{T_0\geq 0, T_N \geq 0 \\ T_0+T_N \leq \frac{N}{f}}} \frac{f}{2}\left(\frac{T_0^2+T_N^2}{N} + \frac{N-1}{N} T^*(f, N)^2\right) \tag{101a}$$

$$= \inf_{\substack{T_0\geq 0, T_N \geq 0 \\ T_0+T_N \leq \frac{N}{f}}} \frac{1}{2}\Bigg(f\frac{T_0^2+T_N^2}{N} + \frac{N}{f(N-1)}\left(1 - \frac{f(T_0+T_N)}{N}\right)^2\Bigg), \tag{101b}$$

where (101b) is obtained by plugging (35) into (101a). We denote the objective function in (101b) by $q(T_0, T_N)$. We prove that $q(T_0, T_N)$ is a Schur-convex function: 1) $q(T_0, T_N)$ is symmetric, since it is invariant to the permutations of T_0 and T_N; 2) the first-order partial derivatives of $q(T_0, T_N)$ with respect to T_0 and T_N are

$$\frac{\partial q}{\partial T_0} = \frac{f}{N} T_0 + \frac{f}{N(N-1)}(T_0 + T_N) - \frac{1}{N-1}, \tag{102a}$$

$$\frac{\partial q}{\partial T_N} = \frac{f}{N} T_N + \frac{f}{N(N-1)}(T_0 + T_N) - \frac{1}{N-1}, \tag{102b}$$

where (102) satisfies (72). Using the property of Schur-convex functions stated in Lemma 2 after (74), we know that the minimum of $q(T_0, T_N)$ is achieved by

$$T_0 = T_N = a, \text{ for some } 0 \leq a \leq \frac{N}{2f}. \tag{103}$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAC.2021.3071953, IEEE Transactions on Automatic Control

Plugging (103) into $q(T_0, T_N)$, we find that the optimal a that minimizes $q(a, a)$ is given by

$$a = \frac{N}{(N+1)f}. \tag{104}$$

Plugging (103) and (104) into (101b), we obtain

$$\underline{D}_N(f, R_s) \geq \frac{N^2}{2f(N+1)^2}. \tag{105}$$

On the other hand, plugging (81) into the right side of (34), we obtain

$$\bar{D}_N(f, R_s) \leq \frac{3N}{2f(N+1)^2} + \frac{\sqrt{3}N(N-1)}{2f(N+1)^2}. \tag{106}$$

Choosing $f = R$ in (106), we conclude that

$$\inf_{\substack{f>0, R_s \geq 1 \\ fR_s \leq R}} \underline{D}_N(f, R_s) \leq \frac{3N}{2R(N+1)^2} + \frac{\sqrt{3}N(N-1)}{2R(N+1)^2}. \tag{107}$$

For any $f \in \left(0, \frac{R}{3+\sqrt{3}}\right)$, the right side of (105) is larger than the right side of (107). Thus, the infimum is attained in the following compact set:

$$f \in \left[\frac{R}{3+\sqrt{3}}, R\right], \tag{108}$$

where the upper bound of f is obtained by lower-bounding R_s by 1. Correspondingly, R_s lies in the following compact set:

$$R_s \in \left[1, 3+\sqrt{3}\right], \tag{109}$$

Using (108)–(109), we conclude that $\underline{D}_N(f, R_s)$ is equicoercive.

G. Proof of Proposition 2

We derive a lower bound to (48), and show that the lower bound is achieved by the SOI code. Note that (48) is lower bounded by

$$\limsup_{T\to\infty} \inf_{\substack{\pi_T \in \Pi_T: \\ \frac{\mathbb{E}[N]}{T} \leq R}} \frac{1}{T}\mathbb{E}\left[\sum_{i=0}^{N} \int_{\tau_i+\delta}^{\tau_{i+1}+\delta} (W_t - \tilde{W}_t)^2 dt\right], \tag{110}$$

where for $t \in [\tau_i + \delta, \tau_{i+1} + \delta)$,

$$\tilde{W}_t \triangleq \mathbb{E}[W_t | \{W_s\}_{s=0}^{\tau_i}, \tau^i] = W_{\tau_i}, \tag{111}$$

since $\sigma(U^i) \subset \sigma(\{W_s\}_{s=0}^{\tau_i})$, and (26) holds. Plugging (111) into the lower bound (110), we obtain the objective function:

$$\frac{1}{T}\mathbb{E}\left[\sum_{i=0}^{N} \int_{\tau_i}^{\tau_{i+1}} (W_t - W_{\tau_i})^2 dt\right] + \delta. \tag{112}$$

Note that the first part of (112) is equal to (25d) in the delay-free case, and δ is a fixed number. Following the arguments in the paragraph below (25d), we conclude that *the SOI coding scheme* achieves (110).

H. Proof of Theorem 4

We calculate the lower bound to the MSE in (53) and show that it can be achieved by *the SOI coding scheme*. From (52) and (54), we conclude that

$$X_t = X_{\tau_i^+} + W_t - W_{\tau_i}, \ t \in (\tau_i, \tau_{i+1}]. \tag{113}$$

The MSE in (53) is equal to

$$\limsup_{T\to\infty} \frac{1}{T}\mathbb{E}\left[\sum_{i=0}^{N} \int_{\tau_i}^{\tau_{i+1}} (X_{\tau_i^+} + W_t - W_{\tau_i})^2 dt\right] \tag{114a}$$

$$= \limsup_{T\to\infty} \frac{1}{T}\mathbb{E}\left[\sum_{i=0}^{N} \int_{\tau_i}^{\tau_{i+1}} X_{\tau_i^+}^2 dt\right] \tag{114b}$$

$$+ \limsup_{T\to\infty} \frac{1}{T}\mathbb{E}\left[\sum_{i=0}^{N} \int_{\tau_i}^{\tau_{i+1}} (W_t - W_{\tau_i})^2 dt\right] \tag{114c}$$

$$+2 \limsup_{T\to\infty} \mathbb{E}\left[\sum_{i=0}^{N} X_{\tau_i^+} \int_{\tau_i}^{\tau_{i+1}} (W_t - W_{\tau_i}) dt\right] \tag{114d}$$

$$\geq \limsup_{T\to\infty} \frac{1}{T}\mathbb{E}\left[\sum_{i=0}^{N} \int_{\tau_i}^{\tau_{i+1}} (W_t - W_{\tau_i})^2 dt\right], \tag{114e}$$

where (114a) is obtained by substituting (113) into (53); (114d) is equal to zero since $X_{\tau_i^+}$ is independent of $\int_{\tau_i}^{\tau_{i+1}} (W_t - W_{\tau_i}) dt$ and $\mathbb{E}\left[\int_{\tau_i}^{\tau_{i+1}} (W_t - W_{\tau_i}) dt\right] = 0$ for all $i = 0, 1, \ldots$ by the reflection principle of the Wiener process [41, Chap.3, Thm. A44]. To achieve the lower bound in (114e), we need $X_{\tau_i^+} = 0$ for all $i = 0, 1, \ldots$, thus, the optimal impulse control signal is $Z_{\tau_i} = -X_{\tau_i}$, $i = 0, 1, \ldots$, which is equal to (55). Using the arguments in the paragraph below (26) in the proof of Theorem 1, we can easily verify that the SOI coding scheme achieves the lower bound and therefore constitutes the optimal causal code.

Nian Guo is a Ph.D. candidate in Electrical Engineering at California Institute of Technology. She received a B. Eng. degree in Electrical Engineering from The University of Hong Kong in 2017, and a M.S. degree from California Institute of Technology in 2019. Her research interests include information theory, theory of random processes, control and wireless communication.

Victoria Kostina is a Professor of Electrical Engineering and Computing and Mathematical Science at Caltech. She received a Bachelor's degree from Moscow institute of Physics and Technology (2004), where she was affiliated with the Institute for Information Transmission Problems of the Russian Academy of Sciences, a Master's degree from University of Ottawa (2006), and a PhD from Princeton University (2013). She received the Natural Sciences and Engineering Research Council of Canada postgraduate scholarship (2009–2012), the Princeton Electrical Engineering Best Dissertation Award (2013), the Simons-Berkeley research fellowship (2015) and the NSF CAREER award (2017). Kostina's research spans information theory, coding, control, learning, and communications.