2019 57th Annual Allerton Conference on Communication, Control, and Computing (Allerton)
Allerton Park and Retreat Center
Monticello, IL, USA, September 24-27, 2019

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 makes real-time estimation of the process using causally received codewords. 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 unlock 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.**

I. Introduction

A. System Model

Consider the system in Fig. 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

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

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, the encoder chooses to generate a binary codeword U_i, with a length $\ell_i \in \mathbb{Z}^+$, based on the past observed process $\{W_t\}_{t=0}^{\tau_i}$. 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 $\{\tau_1, \ldots, \tau_i\}$, 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}.

Fig. 1: System Model.

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

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

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

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

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. Literature Review

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 area of optimal scheduling and sequential estimation problems. Åström and Bernhardsson [1] compared uniform and symmetric threshold sampling policies (referred to as Riemann and Lebesgue sampling, respectively) 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 the problem of causally estimating i.i.d and Gauss-Markov discrete-time processes under the constraint on the total number of transmissions over a finite time horizon, and showed via dynamic programming that the optimal sampling policy is event-triggered. For causal estimation of multidimensional discrete-time Gauss-Markov processes, Cogill et al. [3] proved that the sampling policy that samples when the squared error exceeds some constant, leads to a cost that is within a factor of 6 of the optimal achievable cost. Using dynamic programming and majorization theory, Lipsa and Martins [4] 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,

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.

for scalar discrete-time Gauss-Markov processes over a finite time horizon. For partially observed discrete-time Gauss-Markov processes, Wu et al. [5] derived the accurate and approximate minimum MSE (MMSE) estimator for an event-triggered sampler, and showed the relation between the transmission frequency and the threshold of the event-triggered policy. Rabi et al. [6] formulated the problem of causal estimation of the Wiener process under the constraint on the total number of transmissions over a finite time horizon as an optimal stopping time problem, and solved it iteratively to show that a time-varying symmetric threshold policy is optimal. Nar and Başar [7] extended the optimal stopping time problem in [6] 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 [7] showed that the optimal threshold in the infinite time horizon is a constant depending on the average sampling frequency. Sun et al. in [8] further 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 for this symmetric threshold policy depends on the channel delay, and is different from the one in [7]. For autogressive Markov processes driven by an i.i.d. process with a unimodal and symmetric distribution, Charkravorty and Mahajan [9] showed that a symmetric threshold sampling policy and a Kalman-like estimator jointly achieve the optimal tradeoff between the estimation distortion and transmission cost in the infinite time horizon. Kipnis et al. [10] considered the non-causal lossy source coding of a uniformly sampled Wiener process, and derived the tradeoffs between the frequency, bitrate and the estimation MSE. Kofman and Braslavsky [11] designed a quantized event-triggered controller for noiseless partially observed continuous-time LTI systems with an unknown initial state to ensure asymptotic convergence of the system to the origin with zero average rate, seemingly violating the data-rate theorem. Event-triggered control schemes to guarantee stabilization were designed for continuous-time LTI systems in [11]-[13].

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 also different from [10] 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 [10] permits an infinite delay, following a standard assumption in information theory.

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 $D^{\text{op}}(R) = \frac{1}{6R}$.

In the SOI coding scheme, the encoder continuously tracks the process, generating a bit once the process passes the threshold. To reconstruct the process, both those bits and their time stamps are required at the decoder. In the scenario where, due to implementation constraint, the sampler is process-agnostic, or the decoder has no access to timing information, one has to adopt a deterministic 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 $D_{\text{DET}}(R) = \frac{5}{6R}$. 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). This is a consequence of our real-time distortion constraint. Had we allowed delay, coding gains would have been possible by, for example, jointly compressing blocks of those bits. 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 [20] to compress the innovations $W_{\tau_i} - \hat{W}_{\tau_{i-1}}$, and 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 operational and informational distortion-frequency-rate function (ODFRF and IDFRF). The ODFRF and the IDFRF are both minimized by the maximum sampling 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. The reason is that in the SOI coding scheme, the encoder and the decoder know the random sampling time stamps perfectly, whereas in classical non-causal coding, this free timing information is not considered.

We also show that the SOI coding scheme with a different sampling threshold continues to be optimal when there is a random i.i.d. channel delay between the codeword-generating time and the codeword-delivery time. Finally, we show that if the decoder is allowed to wait for only the next codeword before decoding, the MSE can be further decreased.

In this paper, we provide the proof sketches, with detailed proofs relegated to the long version [25].

D. Paper organization

In Section II, we define causal codes, distortion-rate and distortion-frequency-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 discuss the distortion-rate tradeoffs when delays are allowed at both the encoder and the decoder, at the decoder only, and at the communication channel.

E. Notations

We denote by $\{W_t\}_{t=\tau_i}^{\tau_{i+1}}$ the parts of the Wiener process within time intervals $[\tau_i, \tau_{i+1}]$. 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. For $M \in \mathbb{Z}^+$, $[M] \triangleq \{1, \dots, M\}$.

II. Distortion-rate functions

In this section, we define the operational and the informational causal distortion-rate functions, and we show that an optimal encoder can be separated into a sampler followed by a compressor.

A. Encoding and decoding policies

The standard Wiener process is defined as follows.

Definition 1. *(standard Wiener process, e.g. [17]) 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 encoder and decoder know the initial state $W_0 = 0$ at $\tau_0 = 0$.

Next, we formally define the *encoding* and *decoding* policies[1]. Denote the set of continuous functions on the time interval $[0, t]$ by $\mathcal{C}_{[0,t]}$. Define the Wiener process stopped at a stopping time τ (e.g. [21, Eq. 3.9]) as:

$$W_t(\tau) = \begin{cases} W_t & \text{if } t \leq \tau \\ W_\tau & \text{if } t > \tau. \end{cases} \tag{4}$$

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 $f_T = \{\mathsf{f}_1, \mathsf{f}_2, \dots\}^2$,*

$$\mathsf{f}_i \colon \mathcal{C}_{[0,T]} \to \left[2^{\ell i}\right]. \tag{5}$$

The codeword generated at time τ_i is $U_i = \mathsf{f}_i\left(\{W_t(\tau_i)\}_{t=0}^T\right)$. 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*

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

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

The decoding policy in (6) forces the estimate $\hat{W}_t$ to be equal to the conditional expectation of W_t given all the received information, which is constant between two consecutive codeword-generating time stamps. Allowing more freedom in the design of a decoding policy cannot yield a lower MSE because (6) is the MMSE estimator of W_t during $t \in [\tau_i, \tau_{i+1})$. This is a consequence of the zero-delay MSE constraint (3) at the decoder. As we explain in Section IV-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.

B. Operational distortion-rate function

We now define the 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 limit of infinite time horizon:*

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

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*

$$D^{\text{op}}(R) =$$

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

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

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

Furthermore, if randomized compressing policies are allowed, there is no loss of optimality if at time τ_i, a compressing policy 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 sketch: The detailed proof is in [25, Appen. A]. ■

In (8a), W_{τ_i} is the MMSE estimator of W_t at $t \in [\tau_i, \tau_{i+1})$, given the past lossless samples $\{W_{\tau_j}\}_{j=1}^i$ and the codeword-generating time stamps τ^i. The expectation in (8a) is the

[1]We refer to encoding and decoding *policies* to emphasize their causal nature.

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

sampling distortion due to causally estimating the Wiener process from its lossless 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 $\tau_{i+1}-\tau_i,\ i=1,\ldots,N$. According to the compressing policy described 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 Fig. 2. The sampler takes measurements of the Wiener process under

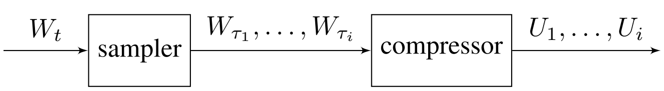

Fig. 2: Decomposition of the encoder.

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 [18]

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

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

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

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

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

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

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

The minimization problem (10b) in $D_{\mathrm{DET}}(R)$ is the causal IDRF for the discrete-time Gauss-Markov 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 considered in (2). Thus, the following relation holds according to [24, Eq. (43)].

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

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. Operational and informational distortion-frequency-rate function

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[3] $\pi_T = \{\tau_1, \tau_2, \ldots\}$ satisfies the average sampling frequency constraint

$$\frac{1}{T}\mathbb{E}(N) = f; \tag{12}$$

(ii) the compressing policy $f_T = \{\mathsf{f}_1, \mathsf{f}_2, \ldots\}$[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

$$\mathbb{E}\left(\sum_{i=1}^{N} \ell_i\right) \le \mathbb{E}(N) R_s; \tag{14}$$

(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).

We define the operational distortion-frequency-rate function.

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 limit of infinite time horizon:*

$$D^{\mathrm{op}}(f, R_s) \triangleq \limsup_{T\to\infty} \inf\{d : \exists\ (f, R_s, d, T) \text{ causal code}\}. \tag{15}$$

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

$$D^{\mathrm{op}}(f, R_s) =$$

$$\limsup_{T\to\infty} \inf_{\substack{\pi_T \in \Pi_T:\\ (12)}} \frac{1}{T}\Bigg\{\mathbb{E}\left(\sum_{i=0}^{N}\int_{\tau_i}^{\tau_{i+1}} (W_t - W_{\tau_i})^2 dt\right) \tag{16a}$$

$$+ \inf_{\substack{f_T \in F_T:\\ (14)}} \mathbb{E}\left(\sum_{i=1}^{N} (\tau_{i+1}-\tau_i)(W_{\tau_i} - \hat{W}_{\tau_i})^2\right)\Bigg\}, \tag{16b}$$

[3]The causal sampling policy is defined in Definition 2(i)

[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).

where the expectation in (16a) is the sampling distortion, and the expectation in (16b) is the mean-square quantization error of the samples weighted by the lengths of sampling intervals $\tau_{i+1} - \tau_i$, $i = 1, \ldots, N$.

We define the informational distortion-frequency-rate function for deterministic sampling policies. The informational equivalent of $D^{\rm op}(f, R_s)$ replaces (14) by the constraint on the directed information, that is, for deterministic sampling policies,

$$\frac{1}{N} I(W^{\tau_N} \to \hat{W}^{\tau_N}) \leq R_s. \tag{17}$$

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

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

$$\limsup_{T\to\infty} \inf_{\substack{\pi_T \in \Pi_T^{\rm DET}:\\ (12)}} \frac{1}{T}\Bigg\{\mathbb{E}\Bigg(\sum_{i=0}^{N}\int_{\tau_i}^{\tau_{i+1}} (W_t - W_{\tau_i})^2 dt\Bigg) \tag{18a}$$

$$+ \inf_{\substack{\otimes_{i=1}^N P_{\hat{W}_{\tau_i}|W^{\tau_i}, \hat{W}^{\tau_{i-1}}}:\\ (17)}} \mathbb{E}\Bigg(\sum_{i=1}^{N} (\tau_{i+1} - \tau_i)(W_{\tau_i} - \hat{W}_{\tau_i})^2\Bigg)\Bigg\} \tag{18b}$$

Similar to $D_{\rm DET}(R)$ in Definition 3, (18b) is the IDRF for the Gauss-Markov process formed by the samples, but it is worth noticing that the rate considered in (18b) is the rate per sample R_s rather than the rate per second R considered in (10b).

III. Main Results

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

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

$$\tau_{i+1} = \inf\left\{t \geq \tau_i : |W_t - W_{\tau_i}| \geq \sqrt{\frac{1}{R}}\right\},\ i = 0, 1, 2, \ldots \tag{19}$$

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

$$U_i = \begin{cases} 1 & \text{if} \quad W_{\tau_{i+1}} - W_{\tau_i} \geq 0 \\ 0 & \text{if} \quad W_{\tau_{i+1}} - W_{\tau_i} < 0. \end{cases} \tag{20}$$

The SOI coding scheme achieves the ODRF:

$$D^{\rm op}(R) = \frac{1}{6R}. \tag{21}$$

Proof sketch: The detailed proof is in [25, Sec. IV-A]. We first prove that the SOI coding scheme in Theorem 1 achieves (21) and satisfies (2) (achievability). We lower bound $D^{\rm op}(R)$ in (8) by the sampling distortion in (8a) under the maximum sampling frequency constraint R, and leverage the results of [7] to calculate that this lower bound is equal to (21) (converse). ■

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 $D_{\rm DET}(R)$.

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

$$\tau_{i+1} - \tau_i = \frac{1}{R},\ i = 0, 1, 2, \ldots, \tag{22}$$

achieves

$$D_{\rm DET}(R) = \frac{5}{6R}. \tag{23}$$

Proof sketch: The detailed proof is in [25, Sec. IV-D]. The value of $D_{\rm DET}(R)$ is found by solving the minimization problem in (25a). The optimality of the uniform sampling policy follows by evaluating (18) with that policy. ■

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

$$D^{\rm op}(R) = \min_{\substack{f>0, R_s \geq 1:\\ f R_s \leq R}} D^{\rm op}(f, R_s), \tag{24a}$$

$$= D^{\rm op}(R, 1) \tag{24b}$$

and the IDRF under deterministic sampling policies satisfies

$$D_{\rm DET}(R) = \min_{\substack{f>0, R_s \geq 1:\\ f R_s \leq R}} D_{\rm DET}(f, R_s) \tag{25a}$$

$$= D_{\rm DET}(R, 1) \tag{25b}$$

Proof sketch: The detailed proofs of (24) and (25) are in [25, Sec. IV-B] and [25, Sec. IV-C] respectively. To prove (24), using a reasoning similar to the proof of Theorem 1, we find that $D^{\rm op}(f, R_s)$ is equal to the sampling distortion $\frac{1}{6f}$ for any $R_s \geq 1$. To prove (25), we leverage the semidefinite representation [15] of the causal IDRF for the Gauss-Markov process formed by N samples, and write $D_{\rm DET}(R)$ as a limit of optimization problems parameterized by N. We show upper and lower bounds to each problem that coincide in the limit of N, yielding

$$D_{\rm DET}(R) = \frac{1}{2f} + \frac{1}{f(2^{2R_s} - 1)}. \tag{26}$$

To justify (25a), we use the fundamental theorem of Γ-convergence. Solving the minimization problem in (25a), we obtain (25b). ■

Using Theorem 3, we can formulate the *working principle* of an optimal encoding policy as follows. A sampler takes measurements of the Wiener process as fast as possible subject to a rate constraint, and the most recent sample is used to generate a 1-bit codeword, which is transmitted to the decoder without delay. In the setting of Theorem 1, the 1-bit SOI compressor associated with the symmetric threshold sampling policy uses the most recent sample to calculate the innovation and to produce a 1-bit codeword. In the setting of Theorem 2, although evaluating $D_{\rm DET}(R)$

does not give us an operational compressing policy, we know that the stochastic kernel that achieves the causal IDRF for discrete-time Gauss-Markov processes formed by the samples under uniform sampling policies has the form $\bigotimes_{i=1}^{\infty} P_{\hat{W}_{\tau_i}|W_{\tau_i}-\hat{W}_{\tau_{i-1}},\hat{W}_{\tau_{i-1}}}$ [23, 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 $\mathsf{g}_i(c) \in \mathbb{R}$ is the quantization representation point corresponding to $c \in \left[2^{\ell_i}\right]$. In practice, one can use the *greedy Lloyd-Max compressor* [20] that runs the Lloyd-Max algorithm for the quantization innovation in each step based on the prior probability of the quantization innovation. Specifically, the prior for $(i+1)$-th step is the pdf of the quantization innovation $W_{\tau_{i+1}} - \hat{W}_{\tau_i}$, which can be computed as the convolution of 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 time horizon [20].

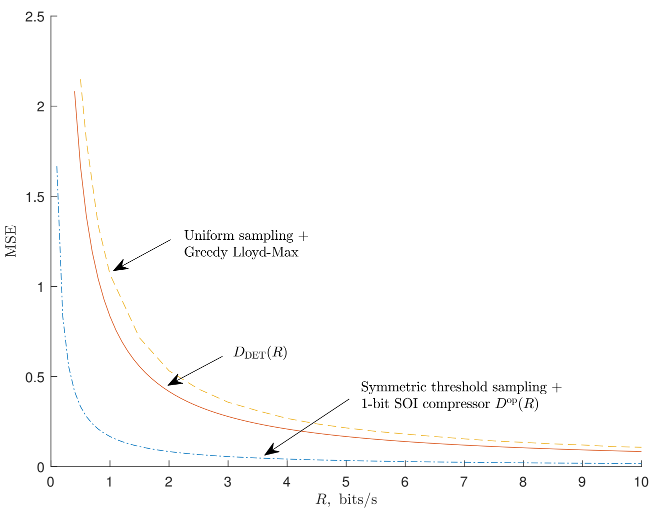

Fig. 3: MSE versus rate

Fig. 3 displays distortion-rate tradeoffs obtained in Theorems 1 and 2, as well as a numerical simulation of the uniform sampling in Theorem 2 with the greedy Lloyd-Max quantization of innovations. The symmetric threshold sampling policy followed by a 1-bit SOI compressor leads to a much lower MSE than uniform sampling. Indeed, according to Theorems 1 and 2, $\frac{D_{\text{DET}}(R)}{D^{\text{op}}(R)} = 5$, and $D^{\text{op}}_{\text{DET}}(R)$ for the uniform sampling is even higher than $D_{\text{DET}}(R)$ by (11). Note that the greedy Lloyd-Max curve is rather close to the $D_{\text{DET}}(R)$ curve, indicating that the IDRF is a meaningful gauge of what is attainable in zero-delay continuous-time causal compression.

The optimal sampling policies of Theorems 1 and 2, i.e. the symmetric threshold and the uniform sampling policies, are the same as the corresponding optimal sampling policies that achieve the minimum sampling distortion [6, Sec. 3.1] [7] subject to an average sampling frequency constraint (12) with $f = R$. The value of $D^{\text{op}}(R)$ (21) achieved by the symmetric threshold sampling policy is the same as the sampling distortion, since the 1-bit SOI compressor is able to compress each innovation noiselessly due to the size-2 alphabet of the innovations, resulting in zero quantization distortion (8b). In contrast, for deterministic sampling policies, quantization distortion is unavoidable, since the samples are Gaussian. If we only consider the constraint on the sampling frequency, the optimal deterministic sampling policy for the Wiener process is uniform sampling [6, Sec. 3.1]. Nevertheless, the result in Theorem 2 implies that uniform sampling is still optimal in the IDRF sense, whether or not the quantization distortion is taken into account.

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

$$D_{\text{noncausal}}(R) = \lim_{T\to\infty} \inf_{\substack{P_{\{\hat{W}_t\}_{t=0}^T|\{W_t\}_{t=0}^T}: \\ \frac{1}{T} I(\{W_t\}_{t=0}^T;\{\hat{W}_t\}_{t=0}^T)\leq R}} \mathbb{E}\left(\frac{1}{T}\int_0^T (W_t - \hat{W}_t)^2 dt\right) \tag{27}$$

Berger [22] derived the distortion-rate function for the Wiener process using reverse water-filling over the power spectrum of the process,

$$D_{\text{noncausal}}(R) = \frac{2\log_2 e}{\pi^2 R} \quad \text{bits/s.} \tag{28}$$

The ODRF continues to be lower-bounded by the IDRF in this non-causal scenario, $D^{\text{op}}_{\text{noncausal}}(R) \geq D_{\text{noncausal}}(R)$ (cf. (11)). As for the achievability, Berger showed that (28) can be achieved in the following sense: 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 [22]: the Wiener process is divided into successive time intervals of a large enough length 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 a resulting MSE per second equal to $D_{\text{noncausal}}(R) + \epsilon$. In parallel with the KL expansion coefficients encoding scheme, an integrating delta modulator is employed to encode each endpoint of the length-T intervals with MSE per second ϵ using ϵ bits per second.

Comparing $D_{\text{noncausal}}(R)$ in (28) with $D^{\text{op}}(R)$ in (21), we see that, surprisingly, the optimal zero-delay policy outperforms the best infinite delay one:

$$\frac{D^{\text{op}}(R)}{D_{\text{noncausal}}(R)} \approx 0.57. \tag{29}$$

This is because in zero-delay causal coding, the timing information is free. Indeed, the decoder knows the codeword-generating time stamps that are stopping times of the filtration generated by the Wiener process. 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 time stamps 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 (28) 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 some small delay is tolerated but the data is not recorded for storage, 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, $\tau_i \leq t < \tau_{i+1}$, 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, this one sample look-ahead decoder greatly reduces the MSE compared to the ODRF obtained in (21) 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 $[t, t']$. In the extreme, $t' = T$, the decoder can jointly use all the codewords and codeword-generating time stamps in time horizon $[0, T]$ to recover the Wiener process. Using Wolf and Ziv's decomposition of MSE in [14], the ODRF with decoder delay can be decomposed as

$$D^{\text{op}}_{\text{dec delay}}(R) = \limsup_{T\to\infty} \inf_{\substack{\pi_T\in\Pi_T\\ f_T\in F_T:\\ (2)}} \frac{1}{T}\mathbb{E}\Bigg(\sum_{i=0}^{N}\int_{\tau_i}^{\tau_{i+1}} (W_t - \bar{W}_t)^2 + \left(\bar{W}_t - \hat{W}_t\right)^2 dt\Bigg), \tag{30}$$

where $\hat{W}_t = \mathbb{E}(W_t|U^N, \tau^N)$ is the MMSE estimator of the process at the decoder using all the received information, and $\bar{W}_t$ is the MMSE estimator of the process at the encoder using the samples and the times that they were taken: for $t \in [\tau_i, \tau_{i+1})$,

$$\bar{W}_t \triangleq \mathbb{E}(W_t|\{W_{\tau_j}\}_{j=1}^N, \tau^N) = \mathbb{E}(W_t|W_{\tau_i}, W_{\tau_{i+1}}, \tau_i, \tau_{i+1}), \tag{31}$$

where (31) holds because $W_t - (W_{\tau_i}, W_{\tau_{i+1}}\tau_i, \tau_{i+1}) - (\{W_{\tau_j}\}_{j=1}^{i-1}, \{W_{\tau_j}\}_{j=i+1}^{N}, \{\tau_j\}_{j=1}^{i-1}, \{\tau_j\}_{j=i+1}^{N})$ form a Markov chain in that order. Therefore, given all the noiseless 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 (31) is the linear interpolation between W_{τ_i} and $W_{\tau_{i+1}}$.

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 necessarily Gaussian, and the linear interpolation can be suboptimal. Yet, if in (30) we substitute for $\bar{W}_t$ a 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 (21). 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 by one sample. 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 $\frac{1}{2R}$ causally attainable with a uniform sampling policy. Thus, the total MSE is at most $\frac{1}{2R}$, a 1.67-fold improvement over (23).

C. Channel delay

Consider the communication scenario in Fig. 1 with a random channel delay between the codeword-generating time stamp and the codeword-delivery time stamp. The decoder sends an acknowledgement to the encoder once it receives a codeword, and a new codeword is generated only after the previous codeword is delivered. A random delay in the communication channel disrupts the synchronization of timing information, worsening the achievable distortion-rate tradeoffs.

Let Y_i be the channel delay. Assume that the initial channel delay is $Y_0 = 0$, and that $0 \leq Y_i \leq \tau_{i+1} - \tau_i$, $\forall i = 1, \ldots, N$, $\tau_{N+1} = T$ and $Y_{N+1} = 0$. The ODRF under the channel delay can be written as

$$D^{\text{op}}_{\text{channel delay}}(R) = \limsup_{T\to\infty} \inf_{\substack{\pi_T\in\Pi_T\\ f_T\in F_T:\\ (2)}} \frac{1}{T}\mathbb{E}\Bigg(\sum_{i=0}^{N}\int_{\tau_i+Y_i}^{\tau_{i+1}+Y_{i+1}} \left(W_t - \hat{W}_t\right)^2 dt\Bigg), \tag{32}$$

where if $t \in [\tau_i + Y_i, \tau_{i+1} + Y_{i+1})$, the optimal decoding policy $\hat{W}_t$ is equal to the following MMSE estimator

$$\hat{W}_{\tau_i+Y_i} \triangleq \mathbb{E}(W_t|U^i, \tau^i + Y^i) = \mathbb{E}(W_{\tau_i+Y_i}|U^i, \tau^i + Y^i), \tag{33}$$

and the codeword U_i is generated based on the past process $\{W_t\}_{t=0}^{\tau_i}$ (Definition 2, (ii)), as in the scenario without the channel delay.

Proposition 2. *i) If the delay Y_i is independent of the Wiener* *process, then*

$$D^{\mathrm{op}}_{\text{channel delay}}(R) - D^{\mathrm{op}}(R) \le \lim_{T\to\infty} \frac{1}{TR}\mathbb{E}\left(\sum_{i=1}^{N} Y_i\right), \tag{34}$$

where N *in* (34) *is the total number of samples taken under the symmetric threshold sampling policy* (19) *within the time duration* T.

ii) If the channel delay Y_i *is independent of the Wiener process and is i.i.d. distributed as a random variable* Y*, the optimal encoding policy is a symmetric threshold policy* (6) *with the new threshold* $\sqrt{\beta}$ *calculated by solving,*

$$\mathbb{E}\left[\max(\beta, W_Y^2)\right] = \max\left(\frac{1}{f}, \frac{\mathbb{E}\left[\max(\beta^2, W_Y^4)\right]}{2\beta}\right). \tag{35}$$

followed by a 1-bit SOI compressor (20)*. The optimal decoding policy still recovers the samples noiselessly by summing up the received innovations.*

Proof sketch: Proposition 2 ii) is proven in the same way as Theorem 1, leveraging the result of Sun et al. [8], who proved that the optimal sampling policy for the Wiener process under a sampling frequency constraint and an i.i.d. channel delay is the symmetric threshold policy with threshold $\sqrt{\beta}$. For the proof of Proposition 2 i), we upper bound $D^{\mathrm{op}}_{\text{channel delay}}(R)$ by the distortion achieved by the SOI coding scheme in Theorem 1. See Appendix G in [25]. ■

V. Conclusion

The results in this paper contribute to the rich literature on optimal scheduling and causal sequential estimation problems by introducing a transmission rate constraint beyond the popular sampling frequency constraint. The SOI coding scheme is optimal for causal estimation of the Wiener process under an expected rate constraint (Theorem 1). The performance of the SOI coding scheme is much better than that of the best non-causal code (Section IV-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 (block) compression. The SOI scheme with a different threshold remains optimal even if the channel introduces an i.i.d. random delay (Proposition 2). The key to transmit 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 lowest possible rate 1-bit codewords as frequently as possible (Theorem 3). This is a consequence of the real-time distortion constraint (3). If a delay is affordable, the MSE can be further reduced with only one sample look-ahead at the decoder (Section IV-B).

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, 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, 2010.

[3] R. Cogill, S. Lall and J. P. Hespanha, "A Constant Factor Approximation Algorithm for Event-Based Sampling," *2007 American Control Conference*, New York, NY, USA, pp. 305-311, 2007.

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

[5] 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.

[6] 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.

[7] 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.

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

[9] 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.

[10] 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.

[11] 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.

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

[13] P. Tallapragada and J. Cortés, "Event-Triggered Stabilization of Linear Systems Under Bounded Bit Rates," in *IEEE Transactions on Automatic Control*, vol. 61, no. 6, pp. 1575-1589, Jun. 2016.

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

[15] 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.

[16] R. Durrett, *Probability: Theory and Examples*, 4th ed. Cambridge University Press, 2010.

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

[18] J. Massey, "Causality, feedback and directed information," *Proceedings International Symposium on Information Theory and its Applications (ISITA-90)*, pp. 303-305, Nov. 1990.

[19] M. S. Derpich and J. Østergaard, "Improved upper bounds to the causal quadratic rate-distortion function for Gaussian stationary sources," in *IEEE Transactions on Information Theory*, vol. 58, no. 5, pp. 3131-3152, 2012.

[20] 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 (CDC)*, Melbourne, VIC, Australia, pp. 6015-6020, Dec. 2017.

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

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

[23] 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," *SIAM Journal on Control and Optimization*, pp. 3731-3765, Oct. 2018.

[24] V. Kostina and B. Hassibi, "Rate-cost tradeoffs in control," in *IEEE Transactions on Automatic Control*, Apr. 2019.

[25] N. Guo and V. Kostina, "Optimal Causal Rate-Constrained Sampling of the Wiener Process", Sep. 2019, http://arxiv.org/abs/1909.01317.