The Distortion-Rate Function of Sampled Wiener Processes

Alon Kipnis, *Student Member, IEEE*, Andrea J. Goldsmith, *Fellow, IEEE*, and Yonina C. Eldar, *Fellow, IEEE*

***Abstract*— We consider the recovery of a continuous-time Wiener process from a quantized or a lossy compressed version of its uniform samples under limited bitrate and sampling rate. We derive a closed-form expression for the optimal tradeoff among sampling rate, bitrate, and quadratic distortion in this setting. This expression is given in terms of a reverse waterfilling formula over the asymptotic spectral distribution of a sequence of finite-rank operators associated with the optimal estimator of the Wiener process from its samples. We show that the ratio between this expression and the standard distortion rate function of the Wiener process, describing the optimal tradeoff between bitrate and distortion without a sampling constraint, is only a function of the number of bits per sample. We also consider a sub-optimal lossy compression scheme in which the continuous-time process is estimated from the output of an encoder that is optimal with respect to the discrete-time samples. We show that the latter is strictly greater than the distortion under optimal encoding but only by at most 3%. We, therefore, conclude that near optimal performance is attained even if the encoder is unaware of the continuous-time origin of the samples.**

***Index Terms*— Brownian motion, Wiener process, sampling, remote source coding, lossy compression, analog to digital conversion, compress and estimate, Brownian bridge.**

I. Introduction

THE Wiener process is a Gaussian stochastic process with stationary independent increments and continuous sample paths, with extensive applications in theoretical and applied science. In particular, the Wiener process models motion of diffusion particles, it is the driving process of risky financial assets in financial mathematics [2], it arises as the limiting law of sequential hypotesis testing procedures [3], it provides the basis for continuous-time martingale theory [4], and it is used to model phase noise in some communication channels [5]. In this work we are concerned with the problem of encoding the path of a Wiener process using a limited number of bits per unit time (bitrate). This is a source coding (lossy compression) problem that arises when a random signal whose probability law follows that of the Wiener process is stored or processed in digital memory, or transmitted over a link of limited capacity.

Manuscript received September 5, 2016; revised August 31, 2018; accepted October 1, 2018. Date of publication October 29, 2018; date of current version December 19, 2018. This work was supported in part by the NSF Center for Science of Information under Grant CCF-0939370 and in part by NSF–BSF under Grant 1609695. This paper was presented in part at the 2016 IEEE International Symposium on Information Theory [1].

A. Kipnis is with the Department of Statistics, Stanford University, Stanford, CA 94305 USA.

A. J. Goldsmith is with the Department of Electrical Engineering, Stanford University, Stanford, CA 94305 USA.

Y. C. Eldar is with the Department of Electrical Engineering, Technion–Israel Institute of Technology, Haifa 32000, Israel.

Communicated by M. Bloch, Associate Editor for Shannon Theory.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIT.2018.2878446

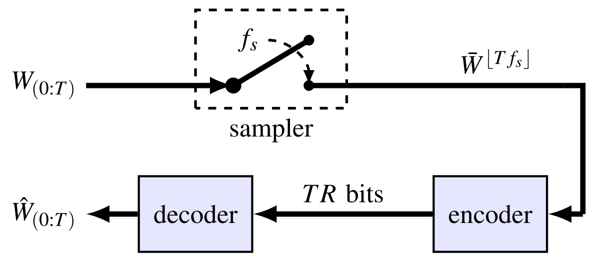

Fig. 1. Uniform sampling and source coding system model.

The optimal trade-off between bitrate and distortion in encoding and reconstructing the Wiener process is described by its *distortion-rate function* (DRF). This DRF was derived by Berger [6], and is based on an encoding of the coefficients of the Karhunen-Lòeve (KL) expansion of the Wiener process. These coefficients are obtained by integrating the Wiener path with respect to the KL basis elements. In practice, however, implementing such integration using purely analog components is extremely challenging since the Wiener process has equal energy in all its frequency components, whereas electronic devices tend to attenuate high frequencies. Consequently, algorithms based on the KL expansion typically operate in discrete-time or require some sort of time-discretization, namely, sampling [7]. In contrast to other processes that are bandlimited or have a finite rate of innovation [8], the *self-similarity* property of the Wiener path implies that its fluctuations scale with time resolution. It is therefore impossible to obtain an equivalent discrete-time representation of the Wiener process by sampling its path [9]. Consequently, Berger's achievability scheme, as well as any source coding approach that is based on transforming the Wiener path to discrete coefficients, is prone to sampling error in addition to the quantization distortion due to the bitrate constraint.

In order to account for the effect of sampling on the overall distortion in encoding the Wiener process, we consider in this work a combined setup of sampling and source coding as described in Fig. 1. In this setup the continuous-time Wiener process $W_{(\cdot)} = \{W_t\}_{t\geq 0}$ is first uniformly sampled at rate f_s over the time interval $[0, T]$, resulting in the finite dimensional random vector of samples $\bar{W}^{\lfloor T f_s \rfloor}$. This vector is then encoded using no more than TR bits, and ultimately the original

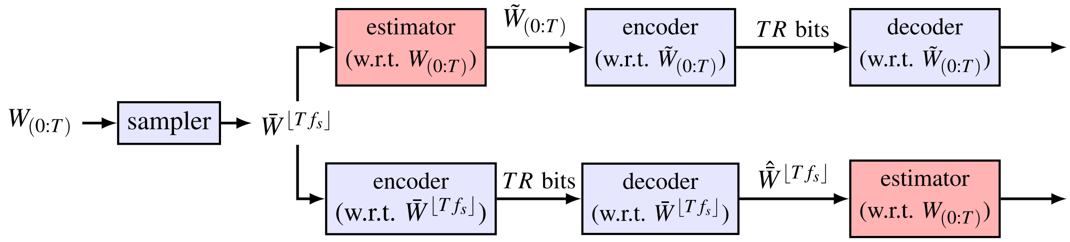

Fig. 2. Two source coding approaches. Upper Path: Estimate-and-Compress (EC). Lower Path: Compress-and-estimate (CE). EC achieves the minimal distortion in the combined sampling and source coding problem of Fig. 1. CE does not require parameters of the continuous-time model at the encoder.

Wiener path is reconstructed from this encoded version under a mean squared error (MSE) distortion criterion. We analyze the minimal distortion in the asymptotic regime of a large time horizon T as a function of the sampling rate f_s and the *bitrate* R. The optimal tradeoff among the three is described by the function $D(f_s, R)$, providing the minimal quadratic distortion in reconstructing the Wiener path when the sampling rate is f_s and the bitrate is R. Consequently, the ratio beteween $D(f_s, R)$ and the DRF of the Wiener process at bitrate R from [6] represents the excess distortion due to a sampling rate constraint f_s in encoding the Wiener process.

In the combined sampling and source coding setting illustrated in Fig. 1, the encoder has no direct access to the realizations of the source it is trying to describe. Therefore, this problem falls within the regime of *indirect* source coding (ISC), a.k.a. *remote* or *noisy* source coding [10, Ch. 3.5]. It is well-known that the optimal source code in such problems is attained by first estimating the original path from the samples, and then encoding this estimate using a source code that is optimal with respect to the estimated process [11]. As we explain in Section IV, this *estimate-and-compress* (EC) strategy allows us to characterize $D(f_s, R)$ by considering two separate problems: A minimal MSE (MMSE) estimation problem that does not involve coding, and a standard source coding problem with respect to the process resulting from this estimation.

Although the EC strategy leads to the minimal distortion under a sampling rate f_s and a bitrate R, it has an interesting caveat: It requires the availability of the sampling rate f_s at the encoder, or, equivalently, the time horizon T or the intensity of the process $\sigma^2 \triangleq \mathbb{E}(W_{t+1} - W_t)^2$ (any one of σ^2, T and f_s uniquely determines the other two). In other words, encoding in EC depends on the continuous-time model or the mechanism by which the samples were acquired. In some scenarios, the encoder may not be informed of any of these parameters or may be unaware of the continuous-time origin of the samples it is given. This situation arises, for example, when a model based on a continuous-time Wiener process is fitted to the measurements only after these were quantized or compressed to satisfy the bit constraint in the acquisition process. In this scenario of missing information, the encoder and decoder may operate according to a *compress-and-estimate* (CE) source coding strategy, as illustrated in Fig. 2: the encoder employs an optimal source code to compress the discrete-time samples subject to a quadratic distortion criterion. The only information required by the encoder in order to construct such an optimal code is the intensity of the Gaussian random walk resulting from uniformly sampling the Wiener process. This intensity equals σ^2/f_s, and can be estimated from the samples. The decoder, provided with the sampling rate (equivalently, T or σ^2), estimates the continuous-time path from the output of the encoder. The resulting distortion under a CE approach provides an upper bound on $D(f_s, R)$, and describes the excess distortion due to ignoring the continuous-time origin of the samples in the encoding stage.

In this work we analyze the distortion under a CE source coding approach when encoding is performed with respect to the random codebook that attains the DRF of the discrete-time vector of samples. We show that although encoding with respect to the samples as in CE is not equivalent to encoding with respect to the estimation of the Wiener process from its samples, the ratio between the distortion under this CE coding scheme is no more than 1.027 times higher than $D(f_s, R)$. That is, while information removed at the encoding stage differs between CE and EC, the performance difference between the two is relatively minor.

The main contribution of this paper is the characterization of the expected MSE distortion in the following cases:

(1) Minimal distortion under all possible bitrate R representations of uniform rate f_s samples of the Wiener process (the function $D(f_s, R)$).

(2) Minimal distortion in the CE scenario where the uniform rate f_s samples are encoded using a random codebook chosen to minimize the MSE distortion with respect to the sequence of samples. We denote this distortion by $D_{\mathrm{CE}}(f_s, R)$.

The characterization of $D(f_s, R)$ is achieved by first providing an information theoretic description of this function as the solution of a sequence of optimization problems involving only probability density functions of limited mutual information rate. This characterization leads to a similar information expression for the process obtained by estimating the Wiener process from its uniform samples. The KL transform of this signal defines a sequence of finite-rank operators, and the expression for $D(f_s, R)$ is given in terms of the limiting eigenvalue distribution of these operators. Finally, we analyze the ratio between $D(f_s, R)$ and the DRF of the Wiener process from [6], as well as the the ratio between $D(f_s, R)$ and the MMSE in estimating the Wiener process from its rate f_s uniform samples. These ratios describe the excess distortion due to sampling in the lossy compression of the Wiener process, and the excess distortion due to a bitrate constraint in the sampling of the Wiener process, respectively. We show that both excess distortions are only a function of the number of bits per sample R/f_s.

As opposed to $D(f_s, R)$ that describes the minimal distortion under any source code, in CE we consider a specific source code: the compression of the samples of the Wiener process using the achievable scheme for the DRF of the discrete-time Wiener process described in [6]. We show that when this source coding scheme is employed with coding bitrates converging to R from above as the time horizon goes to infinity, the resulting distortion converges to an expression we denote as $D_{\text{CE}}(f_s, R)$. This expression is defined in terms of the asymptotic eigenvalue distribution of the operators defining $D(f_s, R)$, as well as the asymptotic eigenvalues distribution of the covariance matrix of the samples of the Wiener process. Finally, we compare $D_{\text{CE}}(f_s, R)$ with $D(f_s, R)$, and conclude that the ratio between the two is bounded from above by 1.027. That is, the performance loss in using CE compared to the optimal source coding scheme is at most 2.7%. This loss can be seen as the penalty in ignoring the continuous-time origin of the samples at the encoder, not knowing the parameters of the continuous time model from which these samples were obtained, or possible suboptimality of the specific source code used to compress the discrete time samples.

We conclude the introduction by reviewing some related work. The DRF of the Wiener process was derived by Berger [6]. Gray [12] established similar results for the more general family of unstable auto-regressive processes. See also [13] for applications of Berger's and Gray's works in reliable communication of unstable processes [13]. The work of Neuhoff and Pradhan [14] derives conditions under which the DRF of a continuous-time stationary Gaussian process, possibly non-bandlimited, can be attained by encoding its samples obtained at asymptotically high rates. In contrast to [14], here we consider a non-stationary source signal and are interested in the optimal source coding performance under a fixed sampling rate, rather than the distortion in the limit of infinitesimally dense sampling grids. Closely related works are [15]–[17], which consider either the combined problem of lossy compression and sampling or other forms of dimensionality reduction. Our combined sampling and source coding setting also falls under the fixed sampling scenario with an informed encoder and decoder of [18]. The work [19] considers a problem of sampling a multi-dimensional Wiener process with limited resources, although without a constraint on the number of bits in representing the samples. A specific scenario of sampling the Wiener process under a communication constraint is considered in [20], to which our function $D(f_s, R)$ provides performance lower bounds. Unlike the results described above, the combined sampling and source coding setting of Fig. 1 allows us to derive the optimal tradeoff between distortion, bitrate, and sampling rate under any digital representation of these samples. We previously explored such tradeoffs for second-order Gaussian stationary processes [21]–[23] and for sparse signals [24].

It is well known that in indirect source coding settings such as in Fig. 1, the minimal distortion is attained via an EC strategy [10, Ch. 3.5], [11], [25]. The CE setting of [26] was proposed in order to study the performance in cases where estimation before compression is impossible due to lack of computation resources, missing information for performing this estimation such as the sampling rate in our setting, or simply an ad-hoc system design that is unaware of the indirect source. CE performance was recently explored in a compressed-sensing framework when the sampling matrix is unavailable at the encoder [27]. In contrast, in our setting, the source signal is not ergodic and the relation between the source signal and its observations is described by a deterministic pointwise sampling, hence the results of [26] and [27] do not apply. A discrete-time version of our setting is considered in [28].

The rest of this paper is organized as follows: In Section II we define a combined sampling and source coding problem for the Wiener process. In Section III we provide preliminary results that are based on known source coding results with respect to the Wiener process. We characterize the minimal distortion in the combined sampling and source coding problem of Fig. 1 in Section IV. The distortion under the CE approach is derived in Section V. Concluding remarks are provided in Section VI.

II. Problem Formulation

Let $W_{(\cdot)} = \{W_t,\, t \geq 0\}$ be a continuous-time Gaussian process with zero mean, autocovariance function

$$K_W(t,s) \triangleq \mathbb{E}[W_t W_s] = \sigma^2 \min\{t,s\}, \quad t,s \geq 0,$$

and $W_0 = 0$ almost surely. The standard definition of the Wiener process also requires that each realization of $W_{(\cdot)}$ has almost surely continuous paths [4]. In our setting, however, only the weaker assumption of almost surely Riemann integrability of the paths is required so that this path can be approximated in discrete-time in the L_2 sense.

We consider the system depicted in Fig. 1 to describe the random waveform $W_{(0:T)} \triangleq \{W_t,\, t \in [0,T]\}$ using a code of rate R bits per unit time. Unlike in the regular source coding problem for the Wiener process considered in [6], we assume that $W_{(0:T)}$ is first uniformly sampled at frequency f_s. Set $T_s \triangleq 1/f_s$ and let

$$N_T \triangleq \lfloor T/T_s \rfloor$$

to be the number of samples obtained by sampling the Wiener process over $[0,T]$ at rate f_s. Denote the vector of samples by

$$\bar{W}^{N_T} \triangleq \left\{W_{n/f_s},\, n \in \mathbb{N} \cap [0, f_s T]\right\}. \tag{1}$$

Above and throughout the paper, we use capital letters to denote random processes, random vectors and random vectors, and lower case letters to denote their realization. The superscript of a random vector denotes its dimension.

The *encoder* in Fig. 1 is a deterministic function

$$f : \mathbb{R}^{N_T} \to \left\{1, \ldots, 2^{\lfloor TR \rfloor}\right\}$$

that maps a real vector $\bar{w}^{N_T} \in \mathbb{R}^{N_T}$ to an index out of $2^{\lfloor TR \rfloor}$ possible indices. The *decoder*, upon receiving the index $f(\bar{w}^{N_T}) \in \{1, \ldots, 2^{\lfloor TR \rfloor}\}$, provides a *reconstructed* waveform

$\hat{w}_{(0:T)} = \left\{\hat{w}_t,\, t \in [0,T]\right\}$. Without loss of generality we assume that the decoder output is a member of the space of square Lebesgue integrable functions $L_2[0,T]$.

The optimal performance theoretically achievable (OPTA) in terms of the distortion in estimating $W_{(\cdot)}$ from its samples is defined as

$$D(f_s, R) = \liminf_{T\to\infty} D_T^{\mathrm{OPTA}}(R) \tag{2}$$

where

$$D_T^{\mathrm{OPTA}}(R) = \inf_{\mathsf{enc-dec}} \frac{1}{T}\int_0^T \mathbb{E}\left(W_t - \hat{W}_t\right)^2 dt,$$

and the infimum is taken over all encoders and decoders to and from a set of at most TR elements. We note that since $W_0 = 0$, replacing the limit infimum in (2) with the infimum over T leads to the trivial solution $D(f_s, R) = 0$. Our definition of OPTA avoids this degenerate case.

Without loss of generality, the OPTA can be written as

$$D(f_s, R) = \liminf_{T\to\infty} \inf_f \ \mathsf{mmse}\left(W_{(0:T)}|f(\bar{W}^{N_T})\right), \tag{3}$$

where

$$\mathsf{mmse}\left(W_{(0:T)}|f(\bar{W}^{N_T})\right) \triangleq \int_0^T \mathbb{E}\left(W_t - \mathbb{E}\left[W_t|f(\bar{W}_T^N)\right]\right)^2 dt$$

is the MMSE in estimating $W_{(0:T)}$ from $f(\bar{W}^{N_T})$. That is, compared to the definition of $D(f_s, R)$ in (3), we eliminate the dependency on the decoder by assuming that this provides the MMSE estimate of $W_{(0:T)}$ given the output of the encoder.

The main goal of this paper is to derive an expression for $D(f_s, R)$ in closed form, as well as to characterize $\mathsf{mmse}\left(W_{(0:T)}|f(\bar{W}^{N_T})\right)$ under an encoder f that follows the CE approach. Before doing so, we explore, in the next section, the connection between these two distortions to the DRF of the Wiener processes derived in [6] without sampling, and to the MMSE in sampling the Wiener process without a bitrate constraint.

III. Preliminaries

In this section we review known results on the optimal MSE attainable in encoding the continuous and discrete-time Wiener processes, and derive connections between these results and the combined sampling and source coding problem of Fig. 1. In particular, we show how these results lead to upper and lower bounds on $D(f_s, R)$. The notation and preliminary results provided in this section are used throughout the paper.

A. The Distortion-Rate Function of the Wiener Process

It is shown in [6] that the OPTA in encoding the Wiener process is given by Shannon's DRF of this process:

$$D_W(R) = \frac{2\sigma^2}{\pi^2 \ln 2} R^{-1} \approx 0.292\sigma^2 R^{-1}. \tag{4}$$

That is, the minimal expected distortion attainable in recovering a Wiener path from its encoded version is inversely proportional to the number of bits per unit time in this encoding. Compared to the combined sampling and source coding problem described in Fig. 1 where the source code is constrained to be a function of its uniform samples, (4) represents the OPTA when the source code is any functional of the Wiener path.

The achievability of (4) is based on the following procedure: divide the interval $[0,T]$ into L identical sub-intervals, each of length $T' = T/L$. For each $l = 0, \dots L-1$, expand the lth section of the path $W_{(0:T)}$ according to the KL expansion of the Wiener process over the interval $[0,T']$. The kth KL coefficient in this expansion is given by

$$\bar{X}_k^{(l)} \triangleq \frac{1}{T'}\int_0^{T'} \phi_k(t)\left(W_{t-lT'} - W_{lT'}\right) dt, \quad k \in \mathbb{N}, \tag{5}$$

where $\phi_k(t)$ is the kth KL eigenfunction of the Fredholm integral equation of the second kind [29] over the interval $[0,T']$ with Kernel $K_W(t,s)$. For each $k \in \mathbb{N}$, the sequence $(X_k^{(1)}, \dots, X_k^{(L-1)})$ consists of L i.i.d. Gaussian random variables with variance equal to the kth KL eigenvalue. We encode this sequence using a single code of $2^{R_l T}$ codewords that is optimal with respect to their scalar Gaussian distribution, where R_l is determined using Kolmogorov's waterfilling formula [30]. The reconstruction waveform is obtained by using the decoded KL coefficients and the KL eigenfunctions. Finally, in order to avoid unbounded distortion due to inaccurate block starting locations in reconstruction, Berger [6] suggested to encode the sequence of block starting locations $\{W_{lT'},\, l = 0, \dots, L-1\}$ using a separate bitstream obtained via a delta modulator [31].

It was shown in [6, Sec. IV] that by taking T and L to infinity such that L/T goes to zero, the bitrate required to encode $\{W_{lT'},\, l = 0, \dots, L-1\}$ is negligible, hence it can be provided to the decoder without increasing the overall bitrate. As a result, the scheme above attains distortion as close to $D_W(R)$ as desired while keeping the bitrate at most R.

As explained in the introduction, it is extremely challenging to realize the integration in (5) without first sampling the analog Wiener path. Instead, here we consider source coding schemes for $W_{(\cdot)}$ which assume that only the samples are available at the encoder, rather than the entire continuous-time path. The samples of $W_{(\cdot)}$ define a discrete-time Wiener process. In what follows we consider the optimal performance in encoding this process according to an MSE criterion subject to a bitrate constraint.

B. The Distortion-Rate Function of the Discrete-Time Wiener Process

The autocovariance function of the discrete-time process $\bar{W}_{[\cdot]} = \left\{W_{n/f_s},\, n = 0, 1, \dots\right\}$ obtained by sampling $W_{(\cdot)}$ at rate f_s, is given by

$$\mathbb{E}\left(\bar{W}_n \bar{W}_k\right) = \mathbb{E}\left(W_{nT_s} W_{kT_s}\right) = \frac{\sigma^2}{f_s}\min\{n,k\}.$$

The process $\bar{W}_{[\cdot]}$ is called a discrete-time Wiener process with intensity σ^2/f_s (a.k.a. a Gaussian random walk). A closed form expression for its DRF was also derived in [6], and can be written as follows:

$$D(\bar{R}_\theta) = \frac{\sigma^2}{f_s}\int_0^1 \min\left\{S_{\bar{W}}(\phi),\theta\right\}d\phi,$$
$$\bar{R}_\theta = \frac{1}{2}\int_0^1 \log^+\left[S_{\bar{W}}(\phi)/\theta\right]d\phi, \tag{6}$$

where $\bar{R}$ is the amount of bits per sample[1] of the code and

$$S_{\bar{W}}(\phi) \triangleq \frac{1}{4\sin^2(\pi\phi/2)} \tag{7}$$

is the asymptotic density of the eigenvalues of the matrix with entries $\min\{n,k\}$, $n,k = 0,\dots,N-1$, as N goes to infinity. Expression (7) gives the distortion as a function of the rate, or the rate as a function of the distortion, through a joint dependency on the parameter θ. Such a parametric representation is said to be of a *waterfilling* form, since only the part of $S_{\bar{W}}(\phi)$ below the water level parameter θ contributes to the distortion.

Keeping the bitrate $R = f_s\bar{R}$ fixed and increasing f_s, we see that the asymptotic behavior of the DRF of $\bar{W}_{[\cdot]}$ as f_s goes to infinity is given by (6) when $\bar{R}$ goes to zero or, equivalently, when θ goes to infinity. The latter can be obtained by expanding both expressions in (6) according to θ^{-1} which, after eliminating θ, leads to

$$D_{\bar{W}}(\bar{R}) \sim \frac{\sigma^2}{f_s}\left(\frac{2}{\pi^2\ln 2\bar{R}} + \frac{\bar{R}\ln 2}{12} + O(\bar{R}^{-2})\right)$$
$$= \frac{2\sigma^2}{\pi^2\ln 2}R^{-1} + \frac{\sigma^2\ln 2}{12}\frac{R}{f_s^2} + O(f_s^{-3}). \tag{8}$$

Note that the first term in (8) is the DRF of the continuous-time Wiener process (4). Thus, we have proven the following:

Proposition 1: Let $\bar{W}_{[\cdot]}$ be the process obtained by uniformly sampling the Wiener process $W_{(\cdot)}$ at sampling rate f_s. Then

$$\lim_{f_s\to\infty} D_{\bar{W}}(R/f_s) = D_W(R).$$

In fact, $D_{\bar{W}}(R/f_s)$ is monotonically increasing in f_s so that

$$\sup_{f_s>0} D_{\bar{W}}(R/f_s) = D_W(R). \tag{9}$$

Proposition 1 provides an intuitive explanation for a fact observed in [6]: the DRF of a discrete-time Wiener process at high distortion behaves as the DRF of a continuous-time Wiener process. Proposition 1 shows that this fact is simply the result of evaluating the DRF of the discrete-time Wiener process $\bar{W}_{[\cdot]}$ at high sampling rates, while holding the bitrate R fixed. Due to the high sampling rate, the number of bits per sample $\bar{R} = R/f_s$ goes to zero and the DRF of the discrete-time Wiener process is evaluated at the large distortion (low bit) limit. The fact that $D_{\bar{W}}(R/f_s)$ is monotonically increasing in f_s implies that the path of the sampled Wiener process becomes harder to describe as the frequency at which those samples are obtained increases.

[1]These units of measurements are consistent with our previous notations: the DRF of a source is evaluated as the number of bits per source symbol available for the code.

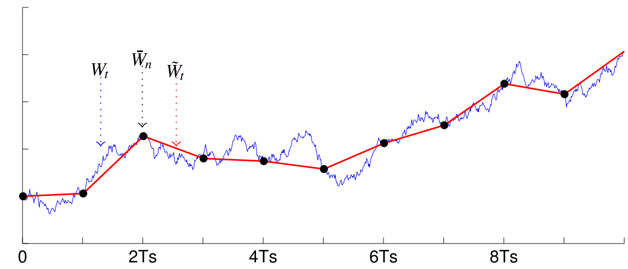

Fig. 3. Sample paths of a Wiener process $W_{(\cdot)}$, its discrete-time samples $\bar{W}_{[\cdot]}$, and it conditional expectation given these samples $\tilde{W}_{(\cdot)}$.

Since the paths of the Wiener process are Riemann integrable, the L_2 distance between any reasonable reconstruction technique (e.g., linear interpolation) of $W_{(0:T)}$ from $\bar{W}^{N_T}$ converges to zero as f_s goes to infinity. Therefore, in addition to convergence of their respective DRFs as expressed in Proposition 1, the path of the optimal reconstruction of $\bar{W}^{N_T}$ from its encoded version converges to the path of the reconstruction of $W_{(0:T)}$ from its encoded version in the L_2 sense. It follows that a distortion arbitrarily close to $D_W(R)$ can be obtained by the following procedure:

(i) Choose T large enough such that the distortion under Berger's achievability scheme for $W_{(0:T)}$ is close to $D_W(R)$.
(ii) Take f_s large enough such that $D_W(R)$ is close to $D_{\bar{W}}(R/f_s)$.
(iii) Encode $\bar{W}^{N_T}$ using a code that attains distortion close to $D_{\bar{W}}(R/f_s)$.
(iv) Estimate $W_{(0:T)}$ from the encoded version of $\bar{W}^{N_T}$.

Since this procedure falls under the system of Fig. 1, we necessarily have $\liminf_{f_s\to\infty} D(f_s,R) \le D_W(R)$, and hence

$$\lim_{f_s\to\infty} D(f_s,R) = D_W(R). \tag{10}$$

Following the characterization of $D(f_s,R)$ in Section IV below, we show that the convergence in (10) is inversely quadratic in f_s.

We now consider the other extreme in the combined sampling and source coding of Fig. 1: finite sampling rate and infinite bitrate.

C. Minimal MSE Under Sampling

A trivial lower bound on $D(f_s,R)$ is obtained by relaxing the bitrate constraint in Fig. 1 by letting R go to infinity. Under this relaxation, the function $D(f_s,R)$ reduces to the MMSE in estimating the Wiener process from its samples, denoted as $\mathsf{mmse}\left(W_{(\cdot)}|\bar{W}_{[\cdot]}\right)$. For $t > 0$, denote by t^+ and t^- the two points on the grid $\mathbb{Z}T_s$ closest to t, namely, $t^- = \lfloor tf_s\rfloor T_s$ and $t^+ = \lceil tf_s\rceil T_s$. Because of the Markov property of $W_{(\cdot)}$, the MMSE in estimating W_t from the process $\bar{W}_{[\cdot]}$ is given by linear interpolation between these two points:

$$\tilde{W}_t \triangleq \mathbb{E}\left[W_t|\bar{W}_{[\cdot]}\right] = \mathbb{E}\left[W_t|W_{t^+},W_{t^-}\right]$$
$$= \frac{t^+-t}{T_s}W_{t^-} + \frac{t-t^-}{T_s}W_{t^+}, \tag{11}$$

See Fig. 3 for an illustration of the path of the processes $W_{(\cdot)}$, $\bar{W}_{[\cdot]}$ and $\tilde{W}_{(\cdot)}$.

The instantaneous estimation error $B_t \triangleq W_t - \tilde{W}_t$ defines a *Brownian bridge* on any interval whose endpoints are on the grid $\mathbb{Z}T_s$. The autocovariance function of $B_{(\cdot)}$ is given by

$$K_B(t,s) = \mathbb{E}[B_t B_s] = \frac{\sigma^2}{T_s}\begin{cases}(t^+ - t\vee s)(t\wedge s - t^-) & nT_s \le t, s \le (n+1)T_s,\\ 0 & \text{otherwise,}\end{cases} \tag{12}$$

where $t \vee s$ and $t \wedge s$ denote the maximum and minimum of $\{t,s\}$, respectively. We conclude that

$$\mathsf{mmse}\left(W_t|\bar{W}_{[\cdot]}\right) = \mathsf{mmse}\left(W_t|\bar{W}_{N_T}, \bar{W}_{N_T+1}\right) = K_B(t,t),$$

and the average MMSE in estimating $W_{(\cdot)}$ from $\bar{W}_{[\cdot]}$ equals

$$\begin{aligned}\mathsf{mmse}\left(W_{(\cdot)}|\bar{W}_{[\cdot]}\right) &= \lim_{T\to\infty}\frac{1}{T}\int_0^T \mathsf{mmse}\left(W_t|\bar{W}_{[\cdot]}\right)dt\\ &= \frac{1}{T_s}\int_{nT_s}^{(n+1)T_s} K_B(t,t)dt = \frac{\sigma^2 T_s}{6} = \frac{\sigma^2}{6f_s}.\end{aligned} \tag{13}$$

For future use, we introduce the notation

$$\mathsf{mmse}(f_s) \triangleq \mathsf{mmse}\left(W_{(\cdot)}|\bar{W}_{[\cdot]}\right) = \frac{\sigma^2}{6f_s}.$$

From properties of the optimal MSE estimator, it follows that for any $T > 0$,

$$\mathsf{mmse}\left(W_{(0:T)}|f(\bar{W}^{N_T})\right) \tag{14}$$

$$= \mathsf{mmse}\left(W_{(0:T)}|\bar{W}\right) + \mathsf{mmse}\left(\tilde{W}_{(0:T)}|f(\bar{W}^{N_T})\right). \tag{15}$$

Since the optimization in (3) is only over the term $\mathsf{mmse}\left(\tilde{W}_{(0:T)}|f(\bar{W}^{N_T})\right)$, (15) implies that encoding $\bar{W}^{N_T}$ to best describe $W_{(0:T)}$ is equivalent to encoding $\bar{W}^{N_T}$ to best describe $\tilde{W}_{(0:T)}$. As a result, we conclude that the optimal encoding strategy for the system in Fig. 1 is *estimate-and-compress*: The encoder first estimates $W_{(0:T)}$ from the samples $\bar{W}^{N_T}$, and then applies an optimal source code to compress the estimate $\tilde{W}_{(0:T)}$ subject to the bitrate constraint. Furthermore, it follows that the OPTA can be written as

$$D(f_s, R) = \mathsf{mmse}(f_s) + D_{\tilde{W}}(R), \tag{16}$$

where $D_{\tilde{W}}(R)$ is the OPTA in encoding the continuous-time process $\tilde{W}_{(\cdot)}$ of (11) at rate R. In other words, (16) reduces the problem of deriving $D(f_s, R)$ to that of deriving the OPTA in encoding at rate R the MMSE estimation of $W_{(\cdot)}$ from its uniform rate f_s samples. This decomposition of the OPTA can be seen as a special case of a more general result dicussed in [11]. The relationships among the various processes and the distortion functions introduced thus far are illustrated in the diagram of Fig. 4.

Before considering the OPTA with respect to $\tilde{W}_{(\cdot)}$, which we defer to Section IV, we explore the relation between $D(f_s, R)$ to the distortion in estimating the samples $\bar{W}_{[\cdot]}$ from

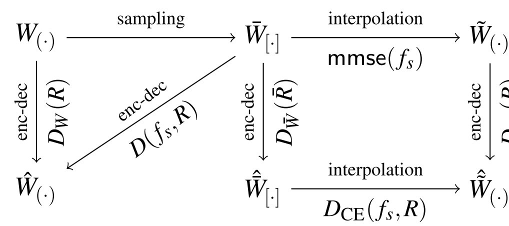

Fig. 4. Relations among the processes $W_{(\cdot)}$, $\bar{W}_{[\cdot]}$, $\tilde{W}_{(\cdot)}$ and their associated distortion functions. Each reconstruction operation is associated with a distortion quantity. In this paper we show that $D_{\tilde{W}}(R) \le D_{\bar{W}}(R/f_s) \le D_W(R) \le D(f_s, R) \le D_{\mathrm{CE}}(f_s, R) \le \mathsf{mmse}(f_s) + D_{\bar{W}}(R/f_s)$, where $D(f_s, R) = \mathsf{mmse}(f_s) + D_{\tilde{W}}(R)$.

an arbitrary finite bit representation of these samples. This relation provides a first upper estimate for $D(f_s, R)$, and will be used in Section V below to characterize the performance under the CE approach.

D. MSE in Discrete- and Continuous-Time

Consider an arbitrary finite bit representation $f(\bar{W}^{N_T}) \in \{1, \dots, 2^{\lfloor TR \rfloor}\}$ of the samples $\bar{W}^{N_T}$ in the system of Fig. 1. The two distortion functions associated with this representation are (1) the MMSE in estimating $W_{(0:T)}$, and (2) the MMSE in estimating $\bar{W}^{N_T}$. The following lemma connects these two distortions, and will be particularly useful in characterizing the distortion under CE in Section V below.

Lemma 2: Fix T, f_s, R, and an encoder $f : \mathbb{R}^{N_T} \to \{1, \dots, 2^{N_T R}\}$. Let $\Delta_n \triangleq \bar{W}_n - \mathbb{E}\left[\bar{W}_n|f(\bar{W}^{N_T})\right]$. The minimal MSE in estimating $W_{(0:T)}$ from $f(\bar{W}^{N_T})$ satisfies

$$\begin{aligned}&\mathsf{mmse}\left(W_{(0:T)}|f(\bar{W}^{N_T})\right)\\ &\quad\ge \mathsf{mmse}(W_{(0:T)}|\bar{W}^{N_T})\\ &\qquad + \frac{2}{3}\frac{1}{N_T}\sum_{n=1}^{N_T-1}\mathbb{E}\Delta_n^2 + \frac{1}{3}\frac{1}{N_T}\sum_{n=1}^{N_T}\mathbb{E}\Delta_n\Delta_{n+1},\end{aligned} \tag{17}$$

and

$$\begin{aligned}\mathsf{mmse}\left(W_{(0:T)}|f(\bar{W}^{N_T})\right) &\le \mathsf{mmse}(W_{(0:T)}|\bar{W}^{N_T})\\ &+\frac{2}{3}\frac{1}{N_T+1}\sum_{n=1}^{N_T+1}\mathbb{E}\Delta_n^2 + \frac{1}{3(N_T+1)}\sum_{n=1}^{N_T}\mathbb{E}\Delta_n\Delta_{n+1}.\end{aligned} \tag{18}$$

Proof: See Appendix B ■

Lemma 2 shows that for any finite bit representation of the samples, the expected MSE in recovering the samples

$$\mathsf{mmse}\left(\bar{W}^{N_T}|f\left(\bar{W}^{N_T}\right)\right) \triangleq \frac{1}{N_T}\sum_{n=1}^{N_T}\mathbb{E}\Delta_n^2$$

and the MSE in recovering the continuous-time Wiener process cannot be too far from each other. An interesting corollary of Lemma 2 arises if we consider a sequence of encoders $\{\bar{f}_N\}_{N\in\mathbb{N}}$ such that, together with the optimal MSE estimation of $\bar{W}^{N_T}$ from the output of these encoders, define a *good rate-distortion code* for $\bar{W}_{[\cdot]}$ [32], [33]. The term good rate-distortion code refers to the fact that the distortion attained by encoding and decoding approaches the DRF of $\bar{W}_{[\cdot]}$, namely

$$\lim_{N\to\infty} \mathsf{mmse}\left(\bar{W}_{[\cdot]}|\bar{f}_N(\bar{W}^N)\right) = D_{\bar{W}}(R/f_s). \tag{19}$$

The existence of such a sequence follows from the source coding theorem with respect to $\bar{W}_{[\cdot]}$ proved in [6], and leads to the following upper bound on $D(f_s, R)$:

Corollary 3: For any $R > 0$ and $f_s > 0$, let

$$\begin{aligned} D^U(f_s, R_\theta) &= \mathsf{mmse}(f_s) + D_{\bar{W}}(R_\theta/f_s) \\ &= \frac{\sigma^2}{6f_s} + \frac{\sigma^2}{f_s}\int_0^1 \min\left\{S_{\bar{W}}(\phi), \theta\right\} d\phi, \end{aligned} \tag{20}$$

where $S_{\bar{W}}(\phi)$ is given in (7) and

$$R_\theta = \frac{f_s}{2}\int_0^1 \log^+\left[S_{\bar{W}}(\phi)/\theta\right] d\phi.$$

The OPTA in estimating a path of the Wiener process from any rate-R encoding of its uniform samples at rate f_s satisfies

$$D(f_s, R) \le D^U(f_s, R).$$

Proof: See Appendix B ■

By considering $D^U(f_s, R)$ in the two extreme cases of the ratio of bitrate to sampling rate, as given below, we obtain estimates for the convergence rates of $D(f_s, R)$ to $D_W(R)$ and $\mathsf{mmse}(f_s)$ in the large sampling rate and bitrate asymptotic, respectively.

1) Low Sampling Rate: When $R \ge f_s$, (20) reduces to

$$D^U(f_s, R) = \mathsf{mmse}(f_s) + \frac{\sigma^2}{f_s}2^{-2R/f_s}. \tag{21}$$

In this regime we have

$$D(f_s,R) - \mathsf{mmse}(f_s) \le D^U(f_s,R) - \mathsf{mmse}(f_s) = \frac{\sigma^2}{f_s}2^{-2R/f_s}.$$

In particular, we conclude that for any $f_s > 0$,

$$\lim_{R\to\infty} D(f_s, R) = D_W(R). \tag{22}$$

2) High Sampling Rate: When f_s is high compared to R, (8) implies

$$D^U(f_s, R) = \frac{\sigma^2}{6f_s} + \frac{2\sigma^2}{\pi^2\ln 2}R^{-1} + O(f_s^{-2}), \tag{23}$$

and therefore

$$D(f_s, R) - D_W(R) \le D^U(f_s, R) - D_W(R) = O(f_s^{-1}).$$

In Section IV we will see that the upper bound $D^U(f_s, R)$ is loose except in trivial cases, and in particular that $D(f_s, R) - D_W(R)$ is $O(f_s^{-2})$. Furthermore, in Section V we derive a closed form expression for the distortion attained by a particular good sequence of encoders with respect to $\bar{W}_{[\cdot]}$. This distortion is shown to be strictly smaller than $D^U(f_s, R)$ and strictly larger than $D(f_s, R)$.

So far we considered elementary properties of $D(f_s, R)$. We concluded that $D(f_s, R)$ is bounded from below by $\mathsf{mmse}(f_s) = \sigma^2/(6f_s)$ and by $D_W(R)$ of (4), and from above by $D^U(f_s, R)$ of (20). We also showed that $D(f_s, R)$ converges to these expressions as f_s or R go to infinity, respectively. In the next section we provide an information theoretic characterization of $D(f_s, R)$, and use this characterization to derive it in closed form.

IV. The Fundamental Distortion-Sampling-Bitrate Limit

We now derive a closed form expression for the function $D(f_s, R)$ that describes the OPTA in recovering the Wiener process from an encoded version of its samples. This derivation is obtained by first proving a source coding theorem for the combined sampling and source coding problem of Fig. 1, and then evaluating the information expression resulting from this theorem.

A. A Combined Sampling and Source Coding Theorem

For $N \in \mathbb{N}$, denote by $\mathcal{P}(\bar{W}^N, R, T)$ the set of all probability distributions $\mathrm{P}_{\bar{W}^N, \hat{W}_{(0:T)}}$ over $\mathbb{R}^N \times \mathrm{L}_2[0, T]$ whose mutual information $I\left(\bar{W}^N; \hat{W}_{(0:T)}\right)$ exists and is limited to TR bits, and the marginal $\mathrm{P}_{\bar{W}^N}$ coincides with the distribution of the discrete-time Wiener process $\bar{W}_{[\cdot]}$ over $n = 0, \ldots, N$. The mutual information of a joint distribution $\mathrm{P}_{\bar{W}^N, \hat{W}_{(0:T)}}$ is defined in the standard manner as the supremum of expected information over all finite partitions of $\mathbb{R}^N$ and L_2 into a finite number of sets measurable with respect to $\mathrm{P}_{\bar{W}^N}$ and $\mathrm{P}_{\hat{W}_{(0:T)}}$, respectively [34].

For $T > 0$ and $R > 0$, define the function

$$D_T(R) = \inf \frac{1}{T}\int_0^T \mathbb{E}\left(W_t - \hat{W}_t\right)^2 dt,$$

where the infimum is over joint distributions $\mathrm{P}_{\bar{W}^{N_T}, \hat{W}_{(0:T)}} \in \mathcal{P}\left(\bar{W}^{N_T}, R, T\right)$.

The OPTA in the combined sampling and source coding setting of Fig. 1 is given as follows:

Theorem 4: For any $R > 0$ and $f_s > 0$,

$$D(f_s, R) = \limsup_{T\to\infty} D_T(R). \tag{24}$$

Proof: For any $T > 0$, define the following distortion measure on $\mathbb{R}^{N_T} \times \mathrm{L}_2[0, T]$:

$$\bar{d}\left(\bar{w}^{N_T}; \hat{w}_{(0:T)}\right) \triangleq \mathbb{E}\left[\frac{1}{T}\int_0^T (W_t - \hat{w}_t)^2 dt \,|\, \bar{W}^{N_T} = \bar{w}^{N_T}\right], \tag{25}$$

where $\hat{w}_{(0:T)}$ is an element of $\mathrm{L}_2[0, T]$, and the relation between $W_{(0:T)}$ and $\bar{W}^{N_T}$ is the same as in (1). In words, $\bar{d}$ is the averaged quadratic distortion between the reconstruction waveform $\hat{w}_{(0:T)}$ and all possible realizations of the random waveform $W_{(0:T)}$ whose values at the points $0, T_s, \ldots, N_T T_s$ are given by $\bar{w}^{N_T}$. By properties of conditional expectation we have

$$\mathbb{E}\left[\bar{d}\left(\bar{W}^{N_T}; \hat{W}_{(0:T)}\right)\right] = \frac{1}{T}\int_0^T \mathbb{E}\left(W_t - \hat{W}_t\right)^2 dt.$$

From the source coding theorem for i.i.d. random variables over arbitrary alphabets with a single-letter distortion measure [35], it follows that the OPTA in encoding $\bar{W}^{N_T}$ is obtained by minimizing over all joint probability distributions of $\bar{W}^{N_T}$ and $\hat{W}_{(0:T)}$ such that their mutual information is limited to TR bits. In the context of our problem, this source coding theorem implies an information representation for the OPTA under sampling at rate f_s of an information source consisting of multiple independent realizations of the waveform $W_{(0:T)}$. Since we are interested in describing a single realization of $W_{(0:T)}$ with T arbitrarily large, what is required is an argument that allows us to separate the path of $W_{(\cdot)}$ into multiple sections (blocks), and consider the joint encoding thereof as multiple realizations over a fixed-length finite interval.

When the continuous source is ergodic or, more generally, *asymptotic mean stationary*, such an argument is achieved by mixing properties of the probability space [36]. In our case, however, $W_{(\cdot)}$ is not asymptotic mean stationary since its variance diverges, so a different approach is required for separating the waveform into multiple i.i.d. sections in order to encode $W_{(\cdot)}$ over blocks. Such an approach was proposed by Berger [6]: use a separate bitstream to encode the endpoints of all length-T intervals. This task is equivalent to encoding a discrete-time Wiener process of variance $T\sigma^2$. It follows from [37, eq. (39)] that the distortion δ in an encoding of the latter using a delta modulator with $\bar{R} = RT$ bits is smaller than a constant times $\sigma^2/\bar{R}$. For any finite R, this number can be made arbitrarily small by taking the blocklength T large enough. That is, the endpoints $\bar{W}_{N_T}, \bar{W}_{2N_T}, \ldots,$ can be described with high accuracy using an arbitrarily small number of bits per unit time. Since the increments of $W_{(\cdot)}$ are independent, its statistics conditioned on the sequence of endpoints is the same as of multiple i.i.d. realizations of $W_{(0:T)}$, and Theorem 4 follows from the first part of the proof. ■

The representation (16) implies that $D(f_s, R)$ can be found by evaluating $D_{\tilde{W}}(R)$. An information theoretic expression for the latter follows from Theorem 4, as given in the following corollary.

Corollary 5: Fix $R > 0$ and $f_s > 0$. Then

$$D_{\tilde{W}}(R) = \limsup_{T\to\infty} \inf_{P_{\tilde{W}_{(0:T)},\hat{W}_{(0:T)}}} \frac{1}{T}\int_0^T \mathbb{E}\left(\tilde{W}_t - \hat{W}_t\right)^2, \quad (26)$$

where the infimum is taken over all joint distributions $\mathrm{P}_{\tilde{W}_{(0:T)},\hat{W}_{(0:T)}}$ over $\mathrm{L}_2[0,T]\times \mathrm{L}_2[0,T]$ whose mutual information exists and does not exceed TR bits, and whose marginal $\mathrm{P}_{\tilde{W}_{(0:T)}}$ coincides with the distribution of $\tilde{W}_{(\cdot)}$ of (11).

Proof: From the properties of optimal MSE estimation, $D_T(R)$ of Theorem 4 can be written as

$$D_T(R) = \mathsf{mmse}(W_{(0:T)}|\bar{W}^{N_T}) + \inf \frac{1}{T}\int_0^T \mathbb{E}\left(\tilde{W}_t - \hat{W}_t\right)^2 dt, \quad (27)$$

with optimization over the set $\mathcal{P}\left(\bar{W}^{N_T}, R, T\right)$. Since the mutual information is invariant to invertible transformations of the random vector $\bar{W}^{N_T}$ [38, Th. 1.4] and since $\tilde{W}_{(0:T)}$ is obtained from $\bar{W}^{N_T}$ by such a transformation, the optimization in (27) can be replaced by an optimization over joint distributions $\mathrm{P}_{\tilde{W}_{(0:T)},\hat{W}_{(0:T)}}$ over $\mathrm{L}_2[0,T]\times\mathrm{L}_2[0,T]$ with mutual information not exceeding TR bits, and whose marginal $\mathrm{P}_{\tilde{W}_{(0:T)}}$ coincides with the distribution of $\tilde{W}_{(\cdot)}$ reduced to the interval $[0,T]$. Corollary 5 now follows from (16) and since $D(f_s,R) = \limsup_{T\to\infty} D_T(R)$ by Theorem 4. ■

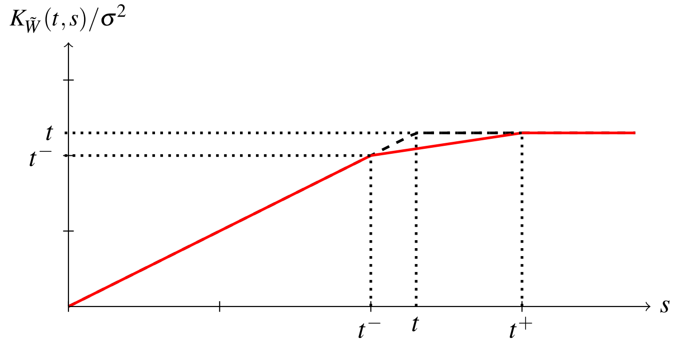

Fig. 5. The autocovariance function of the process $\tilde{W}_{(\cdot)}$ for a fixed $t \in (0, T)$.

Next, we derive $D_{\tilde{W}}(R)$ and $D(f_s, R)$ in closed form by solving the optimization problem in (26) and evaluating its limit as $T \to \infty$.

B. The DRF of $\tilde{W}_{(\cdot)}$

We use the KL expansion of $\tilde{W}_{(\cdot)}$ to evaluate $D_{\tilde{W}}(R)$. We have

$$W_t = \tilde{W}_t + B_t, \quad t \geq 0,$$

where $B_{(\cdot)}$ and $W_{(\cdot)}$ are independent processes. The covariance function of $\tilde{W}_{(\cdot)}$ is given by

$$K_{\tilde{W}}(t,s) = K_W(t,s) - K_B(t,s). \quad (28)$$

The function $K_{\tilde{W}}(t,s)$ is illustrated for a fixed $t \in (0,T)$ in Fig. 5. Corollary 5 implies that $D_{\tilde{W}}(R)$ is given as the limit in T of the second term in (27). Because $\tilde{W}_{(\cdot)}$ is a Gaussian process, this term is obtained by waterfilling over the eigenvalues in its KL transform [10]. These KL eigenvalues $\{\lambda_k,\ k = 1,2,\ldots\}$ and their corresponding eigenfunctions $\{\phi_k,\ k = 1,2,\ldots\}$ satisfy the Fredholm integral equation of the second kind [29]:

$$\lambda_k \phi_k(t) = \int_0^T K_{\tilde{W}}(t,s)\phi_k(s)ds. \quad (29)$$

Since $\tilde{W}_{(0:T)}$ is a linear combination of at most N_T elements, its kernel defined by its autocovariance function is of rank at most N_T. We show in Appendix A that N_T of the eigenvalues of $K_{\tilde{W}}(t,s)$ satisfying (29) are given by

$$\lambda_k = \frac{\sigma^2 T_s^2}{6}\frac{\left(2\cos(k\pi) - \sin\left(\frac{(2k-1)(N-1)\pi}{2N}\right)\right)}{\left(\cos(k\pi) + \sin\left(\frac{(2k-1)(N-1)\pi}{2N}\right)\right)}, \quad k = 1,\ldots,N_T, \quad (30)$$

and thus are the only eigenvalues of (29). We also show in Appendix A that as T goes to infinity with the ratio $f \triangleq k/T \approx kf_s/N_T$ kept constant for $0 < f < f_s$, the density of these eigenvalues converges to the function

$$T_s^2\left(S_{\bar{W}}\left(\frac{\pi f}{2 f_s}\right) - \frac{1}{6}\right), \quad 0 < f < f_s, \quad (31)$$

where $S_{\bar{W}}(\phi)$ is given in (7). Existence of this limiting eigenvalue density implies the following result:

Theorem 6: The DRF of the process $\tilde{W}_{(\cdot)}$, obtained by linearly interpolating the samples of a Wiener process at sampling rate f_s, is given by the following parametric expression:

$$D_{\tilde{W}}(R_\theta) = \frac{\sigma^2}{f_s}\int_0^1 \min\left\{\theta, S_{\tilde{W}}(\phi) - \frac{1}{6}\right\} d\phi, \tag{32a}$$

$$R_\theta = \frac{f_s}{2}\int_0^1 \log^+\left[\left(S_{\tilde{W}}(\phi) - \frac{1}{6}\right)/\theta\right] d\phi, \tag{32b}$$

where $S_{\tilde{W}}(\phi)$ is the limiting density of the eigenvalues in the KL expansion of $\bar{W}_{[\cdot]}$ given by (7).

Proof: The density function (31) satisfies the conditions of [10, Th. 4.5.4] (note that the stationarity property of the source is only needed in [10, Th. 4.5.4] to establish the existence of a density function, which in our case is given explicitly by (31)). This theorem implies that the waterfilling expression over the eigenvalues $\{\lambda_k\}_{k=1}^{N_T}$ converges, as T goes to infinity, to the waterfilling expression over the density $S_{\tilde{W}}(f)$. ■

We remark that as f_s goes to infinity, $D_{\tilde{W}}(R)$ converges to $D_W(R)$ as can be seen by eliminating θ from (32) in a similar way as in (8). In fact, this convergence already follows from the information representation of $D_{\tilde{W}}(R)$ and $D_W(R)$ in Corollary 5 and [6, Sec. IV], respectively, even without obtaining $D_{\tilde{W}}(R)$ in closed form. Indeed, the kernel $K_{\tilde{W}}(t,s)$ converges to the kernel $K_W(t,s)$ in the $L_2[0,T] \times L_2[0,T]$ sense. As a result, the corresponding sequence of compact integral operators defined by (29) converges, in the strong operator norm, to the operator defined by $K_W(t,s)$, showing that the eigenvalues (30) converge to the eigenvalues of the KL expansion for the Wiener process uniformly in T. This convergence of the eigenvalues implies convergence of $D_{\tilde{W}}(R)$ to $D_W(R)$, since both can be defined in terms of a uniformly bounded function of the eigenvalues of $K_{\tilde{W}}(t,s)$ and $K_W(t,s)$, respectively. Similar results were derived for cyclostationary Gaussian stationary processes in [39].

From a practical point of view, it is important to emphasize that although $\tilde{W}_{(\cdot)}$ is a continuous-time process, its KL coefficients can be obtained directly from the discrete-time samples $\bar{W}_{[\cdot]}$ and without performing any analog integration as opposed to the KL coefficients of $W_{(\cdot)}$ in (5). Indeed, assuming for simplicity that Tf_s is an integer, any integrable function $g(t)$ satisfies

$$\int_0^T g(u)\tilde{W}_u du = \sum_{n=0}^{f_sT-1}\left\{\bar{W}_n\bar{X}_n + \bar{W}_{n+1}\bar{Y}_n\right\}, \tag{33}$$

where

$$\bar{X}_n = \frac{1}{T_s}\int_{nT_s}^{(n+1)T_s} g(u)\left((n+1)T_s - u\right) du,$$
$$\bar{Y}_n = \frac{1}{T_s}\int_{nT_s}^{(n+1)T_s} g(u)\left(u - nT_s\right) du. \tag{34}$$

By taking $g(t)$ to be the kth eigenfunction in the KL decomposition of $\tilde{W}_{(\cdot)}$ as given in Appendix A, we see that the kth KL coefficient of $\tilde{W}_{(\cdot)}$ over $[0,T]$ can be expressed as a linear function of the samples $\bar{W}^{N_T}$. This last fact implies that, in contrast to the achivable scheme in [6], a source code which is based on the KL transform of $\tilde{W}_{(\cdot)}$ may be applied directly to a linear transformation of $\bar{W}^{N_T}$ and does not require analog integration as in (5).

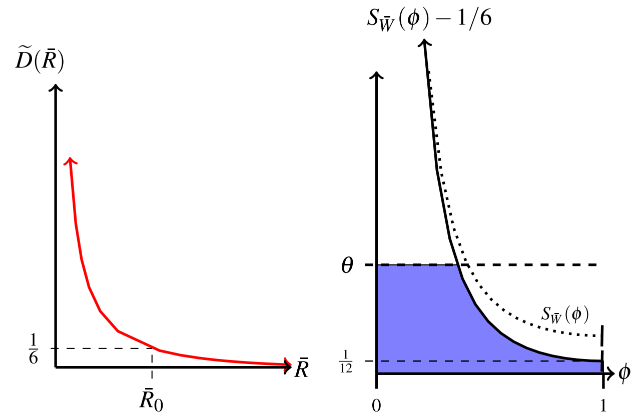

Fig. 6. Left: the function $\widetilde{D}(\bar{R})$ of (37). Right: waterfilling interpretation of the parametric equation (37) describing $\widetilde{D}(\bar{R})$.

We also note that the coefficients of this transformation, i.e., X_n and Y_n, $n = 1, \ldots, N_T$ of (34), depend on f_s. Therefore, the transformation the encoder applies to its input depends on the sampling rate or, equivalently, the duration T of the interval over which the N_T samples were obtained.

C. The DRF of the Wiener Process Given Its Samples

We are now ready to derive a closed-form expression for $D(f_s, R)$.

Theorem 7: The indirect DRF of the Wiener process $W_{(\cdot)}$ given its uniform samples at rate f_s and bitrate R is given by the following parametric form:

$$D(R_\theta) = \frac{\sigma^2}{6f_s} + \frac{\sigma^2}{f_s}\int_0^1 \min\left\{\theta, S_{\tilde{W}}(\phi) - \frac{1}{6}\right\} d\phi, \tag{35a}$$

$$R_\theta = \frac{f_s}{2}\int_0^1 \log^+\left[\left(S_{\tilde{W}}(\phi) - \frac{1}{6}\right)/\theta\right] d\phi. \tag{35b}$$

Proof: Expression (35) follows directly from (13), (16), and Theorem 6. ■

An alternative representation to (35a) is

$$D(f_s, R) = \frac{\sigma^2}{6f_s} + \frac{\sigma^2}{f_s}\widetilde{D}(\bar{R}), \tag{36}$$

where $\bar{R} = R/f_s$ and $\widetilde{D}(\bar{R})$ is given by

$$\widetilde{D}(R_\theta) = \int_0^1 \min\left\{\theta, S_{\tilde{W}}(\phi) - \frac{1}{6}\right\} d\phi, \tag{37a}$$

$$\bar{R}_\theta = \frac{1}{2}\int_0^1 \log^+\left[\left(S_{\tilde{W}}(\phi) - \frac{1}{6}\right)/\theta\right] d\phi. \tag{37b}$$

The function $\widetilde{D}(\bar{R})$ is dimensionless and only depends on the number of bits per sample $\bar{R}$. Figure 6 illustrates $\widetilde{D}(\bar{R})$ and a waterfilling interpretation of (37).

It follows from Theorem 7 that $D(f_s, R)$ is monotonically decreasing in f_s and converges to $D_W(R)$ as f_s goes to infinity. We remark that monotonicity of $D(f_s, R)$ in f_s is not evident in view of [21, Exm. VI.2], where it is shown that the DRF of an arbitrary Gaussian stationary process given its

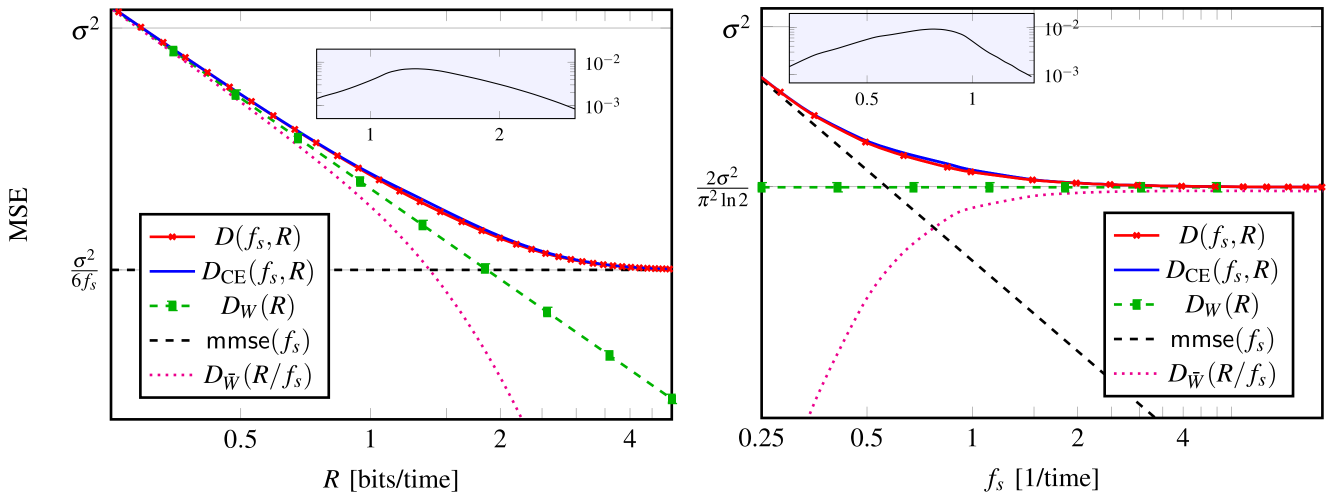

Fig. 7. The indirect DRF $D(f_s, R)$ of the Wiener process given its uniform samples and the compress-and-estimate upper bound $D_{\text{CE}}(f_s, R)$, both as a function of: (left) R, with $f_s = 1$ samples per unit time, and (right) f_s, with $R = 1$ bits per unit time. Also shown are the DRF of the Wiener process $D_W(R)$, the DRF of the discrete-time Wiener process $D_{\bar{W}}(R/f_s)$, and the minimal MSE in estimating the Wiener process from its samples $\mathsf{mmse}(f_s)$. In both figures all axes have logarithmic scales and the inset shows the difference $D_{\text{CE}}(f_s, R) - D(f_s, R)$.

samples may not be monotone in the sampling rate. Figure 7 illustrates expression (35) along with other distortion functions defined in this paper, and confirms the anticipated behavior of $D(f_s, R)$ as f_s or R goes to infinity that is predicted in (10) and (22). We next study expression (35) for $D(f_s, R)$ in the two regimes of low and high sampling rate f_s compared to the bitrate R, corresponding to high and low bits per sample $\bar{R}$, respectively.

1) Low Sampling Rates: As shown in Fig. 6, the minimal value of $S_{\bar{W}}(\phi) - 1/6$, the integrand in (35), is $1/12$. Whenever

$$\frac{R}{f_s} \geq \frac{1 + \log(\sqrt{3} + 2)}{2} \approx 1.45, \tag{38}$$

θ is smaller than $1/12$, in which case we can eliminate θ from (35) and obtain

$$D(f_s, R) = \frac{\sigma^2}{f_s}\left(\frac{1}{6} + \frac{2+\sqrt{3}}{6} 2^{-2R/f_s}\right). \tag{39}$$

2) High Sampling Rates: When $R \ll f_s$, θ is large compared to $S_{\bar{W}}(\phi) - 1/6$, and the integral in (32b) is non-zero only for small values of ϕ. Using the Taylor expansion of $\sin^{-2}(x)$, we obtain

$$D(f_s, R) = \frac{2\sigma^2}{\pi^2 \ln 2} R^{-1} + \frac{\sigma^2 \ln 2}{18} \frac{R}{f_s^2} + O\left(f_s^{-4}\right). \tag{40}$$

From (40) we have that, as anticipated in (10) and (20), $D(f_s, R)$ converges to $D_W(R)$ as f_s goes to infinity. However, this rate of convergence is inversely quadratic in f_s, rather than the inverse linear convergence rate implied by the upper bound $D^U(f_s, R)$ from Corollary 3.

The behavior of $D(f_s, R)$ in the two cases above quantifies the intuitive fact that the distortion is dominated by the minimal MSE distortion $\mathsf{mmse}(f_s)$ for high values of bits per sample $\bar{R}$, and by the lossy compression distortion $D_{\bar{W}}(R)$ for low values of $\bar{R}$. The transition between the two regimes occurs when the MMSE term in (36) equals the term $D_{\bar{W}}(R)$ associated with lossy compression distortion, i.e., at some $\bar{R}_0$ satisfying $\widetilde{D}(\bar{R}_0) = 1/6$, which can be found to be $\bar{R}_0 \approx 0.98$.

The excess distortion in encoding the Wiener process at bitrate R due to a rate f_s sampling constraint is described by the ratio

$$\rho_{\mathsf{smp}}(\bar{R}) \triangleq \frac{D(f_s, R)}{D_W(R)} = \frac{\pi^2 \ln 2}{2}\left(\frac{1}{6} + \widetilde{D}(\bar{R})\right)\bar{R}. \tag{41}$$

Similarly, the excess distortion in sampling the Wiener process at rate f_s due to a bitrate R quantization or lossy compression constraint on the samples is described by the ratio

$$\rho_{\mathsf{qnt}}(\bar{R}) \triangleq \frac{D(f_s, R)}{\mathsf{mmse}(f_s)} = 1 + 6\widetilde{D}(\bar{R}). \tag{42}$$

Both $\rho_{\mathsf{smp}}(\bar{R})$ and $\rho_{\mathsf{qnt}}(\bar{R})$ are only a function of $\bar{R}$, implying that the performance loss due either to sampling or lossy compression are fully characterized by the average number of bits per sample consumed by the digital representation. As an example, given any source code for the samples of the Wiener process allocating $\bar{R} = 1$ bits per sample on average, the distortion in recovering the process is at least

$$\rho_{\mathsf{smp}}(1) D_W(R) \approx 1.18\ D_W(R),$$

or

$$\rho_{\mathsf{qnt}}(1)\mathsf{mmse}(f_s) \approx 2.07\mathsf{mmse}(f_s).$$

These numbers reflect 18% loss compared to the optimal encoding without a sampling constraint, and 107% loss compared to recovering the process from its samples without quantizing them.

In the next section we study the distortion in recovering the Wiener process using a good sequence of encoders for the discrete-time process $\bar{W}_{[\cdot]}$, rather than an optimal sequence designed to attain $D_{\bar{W}}(R)$.

V. Compress-and-Estimate

In Subsection III-C we concluded that the OPTA in the combined sampling and source coding problem is obtained via an EC source coding strategy: First estimate $W_{(0:T)}$ from the samples $\bar{W}^{N_T}$ resulting in $\tilde{W}_{(0:T)}$, and then compress $\tilde{W}_{(0:T)}$ using an optimal source code adjusted to its distribution. In this section we consider an alternative coding strategy: First encode the vector of samples $\bar{W}^{N_T}$ using an optimal source code, such that the expected MSE in recovering $\bar{W}^{N_T}$ from its encoded version converges to its DRF $D_{\bar{W}}(\bar{R})$ of (6). Next, estimate $W_{(0:T)}$ from the the encoded representation of the samples using this code. We denote this scheme as compress-and-estimate (CE). See Fig. 2 for a block diagram of EC and CE.

In this section we provide a precise characterization of the distortion under CE in the case where the encoding of the samples is performed according to the scheme outlined in [6] for attaining the DRF of the discrete-time Wiener process $\bar{W}_{[\cdot]}$. Specifically, we derive a distortion expression we denote as $D_{\text{CE}}(f_s, R)$, and show that this distortion is achievable using a sequence of codes whose bitrate converges to R from above. We also show that when the bitrate is at most R, the distortion under this coding scheme is bounded from below by $D_{\text{CE}}(f_s, R)$. Finally, by comparing $D_{\text{CE}}(f_s, R)$ to $D(f_s, R)$, we conclude that the maximal ratio between the two is no greater than 1.027, indicating a maximal performance penalty of 2.7% in using CE over the optimal source coding scheme.

A. CE Encoding and Decoding

Let $\{\bar{f}_N\}_{N\in\mathbb{N}}$ be a sequence of encoders indexed by their blocklength $N \in \mathbb{N}$. Assume that the encoder $\bar{f}_N$ operates according to the random coding scheme outlined in [6] for achieving the DRF of the discrete-time Wiener process $\bar{W}_{[\cdot]}$. For completeness and further discussion, we now provide a detailed description of the encoder.

We describe the joint encoding of L blocks of samples obtained over the time lag TL, so that each block contains roughly $N_T = \lfloor Tf_s \rfloor$ samples. Denote by $\Sigma_{\bar{W}}$ the covariance matrix of the vector $\bar{W}^{N_T}$ and consider the unitary matrix $\mathbf{U}$ that satisfies

$$\Sigma_{\bar{W}} = \mathbf{U}^T \Lambda \mathbf{U},$$

where Λ is diagonal with the eigenvalues of $\Sigma_{\bar{W}}$ on its diagonal. These eigenvalues are given by [6, eq. (2)]

$$\lambda_n = \frac{\sigma^2/f_s}{4\sin^2\left(\frac{2n-1}{2N+1}\frac{\pi}{2}\right)}, \quad n = 1, \ldots, N_T,$$

where their respective eigenvectors $u_1, \ldots, u_{N_T}$ are the columns of $\mathbf{U}$. Given the $N_T L$ samples of $W_{(0:TL)}$, we consider the N_T length L sequences $B^{(1)}, \ldots, B^{(N_T)}$, defined by

$$B_l^{(n)} \triangleq u_n^T \bar{W}^{(l)} = \sum_{k=1}^{N_T} u_{k,n} \bar{W}_k^{(l)}, \quad l = 1, \ldots, L, \qquad n = 1, \ldots, N_T, \quad (43)$$

where $\bar{W}^{(l)} \triangleq \bar{W}_{(l-1)N_T}^{lN_T} - \bar{W}_{(l-1)N_T}$. In words, $B_n^{(l)}$ is the nth coefficient in the KL decomposition of the lth N_T-length block of $\bar{W}^{LN_T}$, where this block is initialized so that $\bar{W}_0^{(l)} = 0$. See Fig. 8 for an illustration of the relation between $B^{(n)}$ and $\bar{W}^{(l)}$.

block \ time	$B^{(1)}$	$B^{(2)}$	$\cdots$	$B^{(n_{\max})}$	$\cdots$	$B^{(N_T)}$
$l=1$	$u_1^T\bar{W}^{(1)}$	$u_2^T\bar{W}^{(1)}$	$\cdots$			$u_{N_T}^T\bar{W}^{(1)}$
$l=2$	$u_1^T\bar{W}^{(2)}$	$u_2^T\bar{W}^{(2)}$	$\cdots$			$u_{N_T}^T\bar{W}^{(2)}$
$\vdots$	$\vdots$	$\vdots$				
$l=L$	$u_1^T\bar{W}^{(L)}$	$u_2^T\bar{W}^{(L)}$	$\cdots$			$u_{N_T}^T\bar{W}^{(L)}$

Fig. 8. Description of encoding $\bar{W}^{LN_T}$ using the encoder $\bar{f}_{NL}$: divide the vector $\bar{W}^{LN}$ into the length N blocks $\bar{W}^{(1)}, \ldots, \bar{W}^{(L)}$ defined in (43). For $n = 1, \ldots, N$, form the vector $B^{(n)}$ consisting of the nth coefficient in the KL transform of each of the L blocks. Each such vector is encoded using a random Gaussian codebook of rate $L\bar{R}_n$ bits; $\bar{R}_n = 0$ for $N > n_{\max}$.

Given a bitrate budget R, a bitrate slackness parameter $\rho > 0$ and a blocklength N_T, we construct a codebook as follows: for each $n = 1, \ldots, N_T$, we draw $2^{\lfloor(\bar{R}_n+\rho)L\rfloor}$ codewords to describe $B^{(n)}$. Each codeword is obtained by L independent draws from the scalar normal distribution with zero mean and variance $[\lambda_n - \theta]^+$. Here $\bar{R}_n$ and θ are determined by

$$\bar{R} = \frac{1}{N_T}\sum_{n=1}^{N_T} \bar{R}_n,$$

where

$$\bar{R}_n = \frac{1}{2}\log^+ \frac{\lambda_n}{\theta},$$

and $\bar{R} \triangleq R/f_s$ is the number of bits per symbol. To each codeword $\hat{b}^{(l)}$ we associate a unique index $i_n \in \left\{1, \ldots, 2^{\lfloor R_n L \rfloor}\right\}$. We denote this codeword ensemble by $\mathcal{C}_n$, and reveal it to the encoder and decoder. Note that $\mathcal{C}_n$ is trivial whenever $\lambda_n \le \theta$, since then $R_n = 0$. Therefore, in practice, we only need to consider the encoding of $B^{(1)}, \ldots, B^{(n_{\max})}$ where $n_{\max}$ is the largest integer such that $\lambda_n > \theta$.

The encoding of a realization $\bar{w}^{LN_T}$ of $\bar{W}^{LN_T}$ is as follows: First divide $\bar{w}^{LN_T}$ into L blocks of length N_T each which we denote as $\bar{w}^{(1)}, \ldots, \bar{w}^{(L)}$. Then obtain the N_T length-L sequences $b^{(1)}, \ldots b^{(N_T)}$ from these blocks using (43). For each $n = 1, \ldots, N_T$, we associate the index i_n corresponding to the codeword $\hat{b}^{(n)}(i_n)$ of smallest Euclidean distance from $b^{(n)}$ in $\mathcal{C}_n$. The encoder outputs the indices $(i_1, \ldots, i_{N_T})$. In parallel to the representation of the block $\bar{W}^{LT}$ using $(i_1, \ldots, i_{N_T})$, in order to control the error due to uncertainty in block starting locations, the encoder sends a separate bitstream obtained using a delta modulator applied to the sequence of block starting points. As explained in [6, Sec. IV], the bitrate required for this representation goes to zero as T goes to infinity, and hence the total rate of the code we described is $R + o(1)$ where $o(1)$ goes to zero as both L and T/L go to infinity.

We note that encoding using $\{\bar{f}_N\}_{N\in\mathbb{N}}$ corresponds to the achievability scheme outlined in [6, Sec. IV] for attaining the DRF of the discrete-time Wiener process $\bar{W}_{[\cdot]}$. That is, for any

$\delta > 0$ and $\rho > 0$, there exists N large enough such that

$$\mathsf{mmse}\left(\bar{W}^N | \bar{f}_N\left(\bar{W}^N\right)\right) - \delta < D_{\bar{W}}(\bar{R}). \tag{44}$$

In our case, however, we are interested in recovering $W_{(\cdot)}$ rather than $\bar{W}_{[\cdot]}$, and hence the decoding in CE also involves the estimation of $W_{(0:T)}$ given the sequence consisting of decoded codewords. We now analyze the distortion with respect to $W_{(\cdot)}$ attained by using the sequence of encoders $\{\bar{f}_N\}_{N\in\mathbb{N}}$ defined above.

B. Distortion Analysis

In order to characterize the distortion attained by estimating the Wiener process from the encoding of its samples using $\bar{f}_N$, we define

$$D_{\mathrm{CE}}(f_s, R_\theta) = \frac{\sigma^2}{6f_s} + \frac{\sigma^2}{f_s}\int_0^1 \min\left\{\theta, S_{\bar{W}}(\phi)\right\}\frac{S_{\bar{W}}(\phi) - \frac{1}{6}}{S_{\bar{W}}(\phi)}d\phi \tag{45a}$$

$$R_\theta = \frac{f_s}{2}\int_0^1 \log^+\left[S_{\bar{W}}(\phi)/\theta\right]d\phi. \tag{45b}$$

Theorem 8: Fix sampling rate $f_s > 0$ and bitrate $R \geq 0$.

(1i) There exist sequences $\{R_n\}_{n\in\mathbb{N}}$ and $\{T_n\}_{n\in\mathbb{N}}$ with $R_n \to R$ and $T_n \to \infty$, such that, assuming $\bar{f}_{N_T}$ operates at rate R_n,

$$\lim_{n\to\infty} \mathsf{mmse}\left(W^{T_n} | \bar{f}_{\lfloor T_n f_s \rfloor}\left(\bar{W}^{\lfloor T_n f_s \rfloor}\right)\right) = D_{\mathrm{CE}}(f_s, R).$$

(ii) For any $\epsilon > 0$, there exists T_0 such that, for any $T > T_0$ and encoder $\bar{f}_{N_T}$,

$$\mathsf{mmse}\left(W_{(0:T)} | \bar{f}_{N_T}\left(\bar{W}^{N_T}\right)\right) \geq D_{\mathrm{CE}}(f_s, R).$$

Proof: See Appendix B. ■

Theorem 8 states that when the samples of the Wiener process are encoded using a minimum distance encoder with respect to a random codebook drawn from the distortion-rate achieving distribution, the resulting distortion is asymptotically given by $D_{\mathrm{CE}}(f_s, R)$ of (45). In particular, since $\{\bar{f}_N\}_{N\in\mathbb{N}}$ defines a good sequence of codes with respect to $\bar{W}_{[\cdot]}$, Theorem 8 tightens the upper bound of Corollary 3. Indeed, for any $R > 0$ and $f_s > 0$ we have

$$D_{\mathrm{CE}}(f_s, R) < D^U(f_s, R).$$

The expression $D_{\mathrm{CE}}(f_s, R)$ is illustrated in Fig. 7. We now analyze it in the two regimes of high and low sampling rate compared to the bitrate.

1) Low Sampling Rate: When $R \geq f_s$, (45) reduces to

$$D_{\mathrm{CE}}(f_s, R) = \frac{\sigma^2}{6f_s} + \frac{2}{3}D_{\bar{W}}(R/f_s) = \frac{\sigma^2}{6f_s} + \frac{2}{3f_s}2^{-2R/f_s}. \tag{46}$$

Comparing (46) with the optimal distortion in (40), we have

$$D_{\mathrm{CE}}(f_s, R) - D(f_s, R) = \frac{\sigma^2}{f_s}\frac{2-\sqrt{3}}{6}2^{-2\ R/f_s},$$

whenever $R/f_s \geq \left(1 + \log(\sqrt{3}+2)\right)/2$.

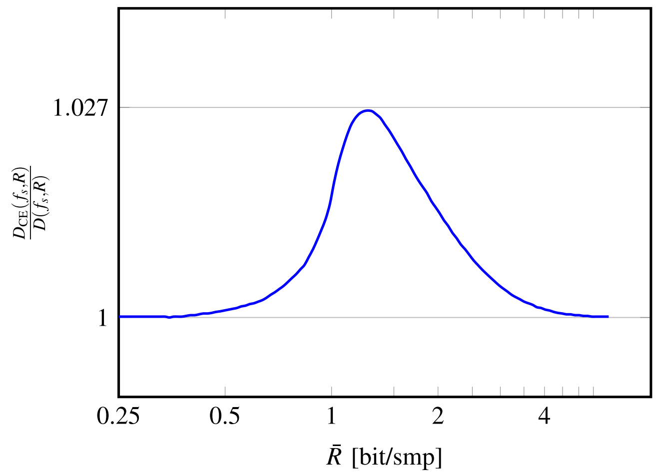

Fig. 9. The ratio $D_{\mathrm{CE}}(f_s, R)/D(f_s, R)$ versus $\bar{R} = R/f_s$ describing the performance loss in using CE compared to the optimal source coding scheme.

2) High Sampling Rate: Using the Taylor expansion of $\sin^{-2}(x)$, for $f_s \gg R$ we obtain

$$D_{\mathrm{CE}}(f_s, R) = D_W(R) + \frac{7}{36}\frac{R\ln 2}{f_s^2} + O(f_s^{-4}),$$

from which we conclude that, similarly to $D(f_s, R)$, $D_{\mathrm{CE}}(f_s, R)$ converges to $D_W(R)$ at a rate inversely quadratic in f_s.

As in the case of $D(f_s, R)$, the excess distortion ratios $D_{\mathrm{CE}}(f_s, R)/D_W(R)$ and $D_{\mathrm{CE}}(f_s, R)/\mathsf{mmse}(f_s)$ are both only functions of the number of bits per sample $\bar{R} = R/f_s$. Therefore, the ratio between $D_{\mathrm{CE}}(f_s, R)$ and $D(f_s, R)$, describing the performance loss in using CE compared to the optimal scheme, also depends only on $\bar{R}$. As illustrated in Fig. 9, this ratio is bounded from above by 1.027, indicating a maximal performance loss of only 2.7% in using CE compared to the optimal source coding scheme.

C. Discussion: Sub-Optimality of CE

In order to gain some insight into the difference between the performance of CE compared to the optimal source coding scheme, it is useful to focus on the term

$$\frac{1}{N_T}\sum_{n=0}^{N_T-1}\mathbb{E}\Delta_n\Delta_{n+1} \tag{47}$$

in the upper and lower bounds of Lemma 2. For simplicity, consider the regime $R \geq f_s$ in which we have

$$D_{\mathrm{CE}}(f_s, R) = \mathsf{mmse}(f_s) + \frac{2}{3}D_{\bar{W}}(f_s/R). \tag{48}$$

By evaluating (18) in the limit as T approaches infinity under the CE encoders $\{\bar{f}_N\}_{N\in\mathbb{N}}$ and comparing it with (48), it follows that (47) goes to zero under CE. We now argue that, as opposed to CE, under the optimal encoder the term (47) is negative. For this purpose, we illustrate in Fig. 10 two areas, each of which is associated with a different encoder: the EC encoder strives to minimize the difference between $\tilde{W}_{(0:T)}$ and its reconstruction $\hat{\tilde{W}}_{(0:T)}$ (integrated square of shaded red area).

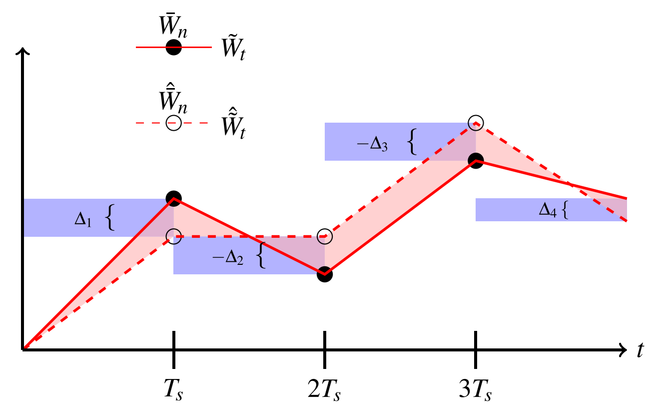

Fig. 10. The error in estimating $\bar{W}_{[\cdot]}$ and $\tilde{W}_{(\cdot)}$ corresponds to the blue and red areas, respectively. Negatively correlated Δ_n and Δ_{n+1} are preferred for minimizing the distortion with respect to $\tilde{W}_{(\cdot)}$, while the distortion with respect to $\bar{W}_{[\cdot]}$ is indifferent to this correlation.

The CE encoder strives to minimize the difference between $\bar{W}^{N_T}$ and its reconstruction $\hat{\bar{W}}^{N_T}$ (integrated square of shaded blue area). An inspection of these error terms reveals that the CE encoder is indifferent to the sign of Δ_n, whereas the distortion associated with the shaded red area is smaller whenever Δ_n and Δ_{n+1} alternate their signs (compare the red areas in the intervals $[T_s, 2T_s]$ and $[2T_s, 3T_s]$, corresponding to such sign alternation and no sign alternation, respectively). Therefore, an EC codebook favors a sign alternation from Δ_n to Δ_{n+1}, implying that (47) is negative under EC.

VI. Conclusions

We considered the estimation of the path of a continuous-time Wiener process from a bitrate-limited version of its uniform samples. We derived a closed form expression for the minimal distortion in this setting as a function of the sampling rate and the bitrate limitation on the encoded version of these samples. This expression allows us to determine the excess distortion in encoding the Wiener process under a uniform sampling constraint compared to its Shannon distortion-rate function, or, alternatively, the excess distortion in sampling the Wiener process under a quantization constraint compared to the MMSE from infinite precision samples.

In addition to the optimal encoding of the samples, we also considered a CE coding approach in which the Wiener process is estimated from an encoded version of its samples, whereas this encoding employs a code that minimizes the distortion with respect to the samples rather than the continuous-time process. We provided a closed form expression for the performance under this approach, and showed that the performance loss under this sub-optimal approach is smaller than 2.7% compared to the optimal source coding technique. Suboptimality of the CE approach can be seen as the price of ignoring the continuous-time origin of the samples at the encoder.

Appendix A

In this appendix we prove that the eigenvalues of the KL integral (29) are given by (30).

Equation (29) can be written as

$$\frac{\lambda}{\sigma^2}\phi(t) = \int_0^{t^-} s\phi(s)ds + \int_{t^-}^{t^+}\left(\frac{t-t^-}{T_s}s + t^-\frac{t^+-t}{T_s}\right)\phi(s)ds + t\int_{t^+}^{T}\phi_k(s)ds. \tag{49}$$

Differentiating the last expression leads to

$$\frac{\lambda}{\sigma^2}\phi'(t) = \int_{t^-}^{t^+}\frac{s-t^-}{T_s}\phi(s)ds + \int_{t^+}^{T}\phi_k(s)ds, \tag{50}$$

which implies

$$\frac{\lambda}{\sigma^2}\phi''(t) = 0. \tag{51}$$

We conclude that the solution to (49) is a piece-wise linear function on intervals of the form $[nTs, (n+1)T_s)$ for $n = 0, \ldots, N$, where $N = T/T_s$ (since the DRF is obtained by evaluating the solution as T goes to infinity, and since this limit exists, there is no loss in generality by assuming T/T_s is an integer), namely

$$\phi_k(t) = \frac{t^+-t}{T_s}a_k(t^-) + \frac{t-t^-}{T_s}b_k(t^-), \quad k = 1, 2, \ldots.$$

Equations (49) and (50) impose the following condition on the coefficients $a_k(t^-)$ and $b_k(t^-)$, for $t \in [0, T_sN]$:

$$\frac{\lambda}{T_s\sigma^2}\left(b_k(t^-) - a_k(t^-)\right) = \frac{1}{6T_s}\left(a_k(t^-) + b_k(t^-)\right) + \frac{1}{2}\int_{t^+}^{T}\left(a_k(s^-) + b_k(s^-)\right)ds.$$

By imposing the initial conditions in (49) and (50), it follows that the eigenfunctions in the KL transform are of the form

$$\phi_k(t) = \sqrt{A_k}\left(\frac{t^+-t}{T_s}\sin\left(\frac{2k-1}{2T}\pi t^-\right) + \frac{t-t^-}{T_s}\sin\left(\frac{2k-1}{2T}\pi t^+\right)\right), \quad k = 1, \ldots, N,$$

where A_k is a normalization constant. The corresponding eigenvalues can be found by evaluating (50), which leads to

$$\lambda_k = \frac{\sigma^2\ T_s^2}{6}\frac{\left(2\cos(k\pi) - \sin\left(\frac{(2\ k-1)(N-1)\pi}{2N}\right)\right)}{\left(\cos(k\pi) + \sin\left(\frac{(2\ k-1)(N-1)\pi}{2\ N}\right)\right)}, \quad k = 1, \ldots, N.$$

Appendix B

In this appendix we provide the proofs of Lemma 2, Corollary 3, and Theorem 8.

A. Proof of Lemma 2

Fix T and f_s. Set $\bar{R} = R/f_s$, $N = \lfloor T f_s \rfloor$ and $M = f(\bar{W}^N)$. We have

$$\mathsf{mmse}\big(W_{(0:T)}|M\big) = \mathsf{mmse}\Big(W_{(0:T)}|\bar{W}^N\Big) + \mathsf{mmse}\Big(\tilde{W}_{(0:T)}|M\Big), \tag{52}$$

hence we only focus on the term $\mathsf{mmse}\left(\tilde{W}_{(0:T)}|M\right)$. Denote $\hat{\bar{W}}_n = \mathbb{E}\left[\bar{W}_n|M\right]$. Then

$$\begin{aligned}\mathbb{E}\left[\tilde{W}_t|\hat{\bar{W}}^N\right] &= \frac{t^+ - t}{T_s}\mathbb{E}\left[W_{t^-}|\hat{\bar{W}}^N\right] + \frac{t - t^-}{T_s}\mathbb{E}\left[W_{t^+}|\hat{\bar{W}}^N\right] \\ &= \frac{t^+ - t}{T_s}\hat{\bar{W}}_{f_s t^-} + \frac{t - t^-}{T_s}\hat{\bar{W}}_{f_s t^+}.\end{aligned} \tag{53}$$

Now,

$$\begin{aligned}&\mathsf{mmse}\left(\tilde{W}^{NT_s}|\hat{\bar{W}}^N\right) \\ &= \frac{1}{NT_s}\int_0^{NT_s}\mathbb{E}\left(\tilde{W}_t - \mathbb{E}\left[\tilde{W}_t|\hat{\bar{W}}^N\right]\right)^2 dt \\ &= \frac{1}{NT_s}\sum_{n=0}^{N-1}\int_{nT_s}^{(n+1)T_s}\mathbb{E}\left(\tilde{W}_t - \mathbb{E}\left[\tilde{W}_t|\hat{\bar{W}}^N\right]\right)^2 dt \\ &\overset{(a)}{=} \frac{1}{NT_s}\sum_{n=0}^{N-1}\int_{nT_s}^{(n+1)T_s}\mathbb{E}\left(\tilde{W}_t - \frac{t-nT_s}{T_s}\hat{\bar{W}}_{n+1}\dots\right. \\ &\qquad\qquad \left. - \frac{T_s(n+1)-t}{T_s}\hat{\bar{W}}_n\right)^2 dt\end{aligned} \tag{54}$$

$$\overset{(b)}{=} \frac{1}{NT_s^3}\sum_{n=0}^{N-1}\int_{nT_s}^{(n+1)T_s}\mathbb{E}\big((t-nT_s)\Delta_{n+1} + ((n+1)T_s - t)\Delta_n\big)^2 dt, \tag{55}$$

where (a) follows from (53), and (b) follows since

$$\begin{aligned}\tilde{W}_t &= \frac{t-t^-}{T_s}W_{t^+} + \frac{t^+ - t}{T_s}W_{t^-} \\ &= \frac{t-t^-}{T_s}\bar{W}_{f_s t^+} + \frac{t^+ - t}{T_s}\bar{W}_{f_s t^-},\end{aligned}$$

and by introducing the notation

$$\Delta_n \triangleq \bar{W}_n - \hat{\bar{W}}_n = \bar{W}_n - \mathbb{E}\left[\bar{W}_n|M\right].$$

Evaluating the integral in (55) we obtain

$$\begin{aligned}&\mathsf{mmse}(\tilde{W}^{NT_s}|M) \\ &= \frac{1}{N}\sum_{n=0}^{N-1}\left(\frac{1}{3}\mathbb{E}\Delta_{n+1}^2 + \frac{1}{3}\mathbb{E}\Delta_n^2 + \frac{1}{3}\mathbb{E}\Delta_{n+1}\Delta_n\right) \\ &= \frac{2}{3}\frac{1}{N}\sum_{n=1}^{N-1}\mathbb{E}\Delta_n^2 + \frac{1}{3N}\sum_{n=0}^{N-1}\mathbb{E}\Delta_n\Delta_{n+1} + \frac{1}{3}\mathbb{E}\Delta_N^2 \\ &\geq \frac{2}{3}\frac{1}{N}\sum_{n=1}^{N-1}\mathbb{E}\Delta_n^2 + \frac{1}{3N}\sum_{n=0}^{N-1}\mathbb{E}\Delta_n\Delta_{n+1},\end{aligned}$$

where we used the fact that $\mathbb{E}\Delta_0 = 0$ because $W_0 = 0$ with probability one. Similarly, we have

$$\begin{aligned}&\mathsf{mmse}(\tilde{W}^{(N+1)T_s}|M) \\ &= \frac{2}{3}\frac{1}{N+1}\sum_{n=0}^{N}\mathbb{E}\Delta_n^2 + \frac{1}{3(N+1)}\sum_{n=0}^{N}\mathbb{E}\Delta_n\Delta_{n+1} + \frac{1}{3}\mathbb{E}\Delta_{N+1}^2 \\ &\leq \frac{2}{3}\frac{1}{N+1}\sum_{n=1}^{N+1}\mathbb{E}\Delta_n^2 + \frac{1}{3(N+1)}\sum_{n=1}^{N}\mathbb{E}\Delta_n\Delta_{n+1}.\end{aligned}$$

The bounds (17) and (18) follow from the last two inequalities and the fact that

$$\mathsf{mmse}(\tilde{W}^{NT_s}|M) \leq \mathsf{mmse}(\tilde{W}_{(0:T)}|M) \leq \mathsf{mmse}(\tilde{W}^{(N+1)T_s}|M).$$

B. Proof of Corollary 3

Set $N = \lfloor T f_s \rfloor$ and $\bar{R} = R/f_s$. By bounding $\mathbb{E}\Delta_{n+1}\Delta_n$ in (18) from above by $(\mathbb{E}\Delta_n^2 + \mathbb{E}\Delta_{n+1}^2)/2$, we obtain

$$\begin{aligned}&\frac{1}{N+1}\sum_{n=1}^{N}\mathbb{E}\Delta_n\Delta_{n+1} \\ &= \frac{1}{N+1}\sum_{n=0}^{N}\mathbb{E}\Delta_n\Delta_{n+1} \\ &\leq \frac{1}{N+1}\sum_{n=0}^{N}\left(\frac{\mathbb{E}\Delta_n^2}{2} + \frac{\mathbb{E}\Delta_{n+1}^2}{2}\right) \\ &= \frac{1}{N+1}\sum_{n=1}^{N}\mathbb{E}\Delta_n^2 + \frac{1}{2}\frac{1}{N+1}\mathbb{E}\Delta_{N+1}^2 \leq \frac{1}{N+1}\sum_{n=1}^{N+1}\mathbb{E}\Delta_n^2.\end{aligned}$$

It follows from (18) that

$$\begin{aligned}D(f_s, R) &\leq \mathsf{mmse}\left(W_{(0:T)}|\bar{f}(\bar{W}^N)\right) \leq \mathsf{mmse}(W_{(0:T)}|\bar{W}^N) \\ &\quad + \frac{2}{3}\frac{1}{N+1}\sum_{n=1}^{N+1}\mathbb{E}\Delta_n^2 + \frac{1}{3}\frac{1}{N+1}\sum_{n=1}^{N+1}\mathbb{E}\Delta_n^2 \\ &= \frac{1}{N+1}\sum_{n=1}^{N+1}\mathbb{E}\Delta_n^2 = \mathsf{mmse}(\bar{W}^{N+1}|\bar{f}\left(\bar{W}^N\right)).\end{aligned}$$

Under the sequence of encoders $\left\{\bar{f}_N,\ N \in \mathbb{N}\right\}$ in the limit $T \to \infty$, we have that $\mathsf{mmse}\left(\bar{W}^N|\bar{f}_N\left(\bar{W}^N\right)\right)$, and therefore $\mathsf{mmse}(\bar{W}^{N+1}|\bar{f}_N\left(\bar{W}^N\right)$, converges to $D_{\bar{W}}(\bar{R})$. In this limit we also have that $\mathsf{mmse}(W_{(0:T)}|\bar{W}^N)$ converges to $1/(6f_s)$, so that

$$D(f_s, R) \leq \frac{1}{6f_s} + D_{\bar{W}}(\bar{R}).$$

C. Proof of Theorem 8

For $L \in \mathbb{N}$ and $T > 0$ we consider the encoding of the vector of samples $\bar{W}^{LT}$ using the encoders $\{\bar{f}_N\}_{N\in\mathbb{N}}$ and the estimation of $W_{(0:LT)}$ from this encoding. Throughout the proof we make use of various simplifications for the notation in the paper, as per the following list:

- The distortion D is normalized by σ^2/f_s and, consequently, assume that any length N_T vector $\bar{W}^{(l)}$ and its reconstruction $\hat{\bar{W}}^{(l)}$ are normalized by $\sqrt{\sigma^2/f_s}$.

- $N \triangleq N_T \triangleq \lfloor T f_s \rfloor$.
- $\bar{f} \triangleq \bar{f}_N$ where the blocklength N is understood from the context.

We first consider properties of the joint distribution that attains the DRF of the vector $\bar{W}^N$. For a prescribed $\bar{R}$, let θ be such that

$$\bar{R} = \frac{1}{2}\sum_{k=1}^{N} \log^+ [\lambda_k/\theta].$$

Consider the eigenvalue decomposition of the matrix $\Sigma_{\bar{W}}$:

$$\Sigma_{\bar{W}} = \mathbf{U}^T \Lambda \mathbf{U},$$

where $\mathbf{U}$ is unitary and Λ is diagonal. The elements $\lambda_1, \ldots, \lambda_N$ on the diagonal of Λ are given by [6, eq. (2)]

$$\lambda_k = \frac{1}{4\sin^2\left(\frac{2k-1}{2N+1}\frac{\pi}{2}\right)}, \quad k = 1, 2, \ldots, N.$$

The columns of $\mathbf{U}$ are the eigenvectors in the KL transform of $\bar{W}^N$ corresponding to the eigenvalues $\lambda_1, \ldots, \lambda_N$, which are given by [6]

$$u_{k,n} = A_k \sin\left(\frac{2k-1}{2N+1}\pi n\right),$$

and where A_k is a normalization coefficient satisfying

$$A_k = \frac{1}{N}\sum_{n=0}^{N-1}\sin^2\left(\frac{2k-1}{2N+1}\pi n\right) = 1, \quad k = 1, 2, \ldots, N.$$

Given an encoder

$$g : \{1, \ldots, 2^{L\bar{R}_1}\} \times \cdots \times \{1, \ldots, 2^{L\bar{R}_{N_T}}\} \to \mathrm{L}_2[0, LT], \quad (56)$$

we denote

$$\hat{W}_{(0:LT)} = g\left(f\left(\bar{W}^{LN_T}\right)\right),$$

and

$$\Delta_n \triangleq W_n - \hat{\bar{W}}_n, \quad n = 1, \ldots, LN,$$

where $\hat{\bar{W}}_n = \hat{W}_{n/f_s}$.

In order to prove (i), it is enough to show that for any $\rho > 0$. $\varepsilon > 0$ and $\delta > 0$, there exists T and L large enough and a decoder g such that $L/T < \varepsilon$, and, if $\bar{W}^{LN}$ is encoded using f_{LN_T}, then

$$\frac{1}{T}\int_0^T \mathbb{E}\left(W_t - [g\left(f\left(\bar{W}^{LT}\right)\right)]_t\right)^2 \, dt < D_{\mathrm{CE}}(f_s, R) + \delta.$$

(The condition $L/T < \epsilon$ is required to guarantee that the bitrate consumed by the delta modulator is arbitrarily small). In order to prove (ii), we show that for any L, and a function g of the form (56), there exists T_0 such that

$$\frac{1}{T}\int_0^T \mathbb{E}\left(W_t - g\left(f\left(\bar{W}^{LT}\right)\right)\right)^2 \, dt \geq D_{\mathrm{CE}}(f_s, R),$$

whenever $\rho = 0$ and $T \geq T_0$.

We first prove the following claims:

I. Under the sequence of encoders $\{\bar{f}\}$,

$$\frac{1}{NL}\sum_{n=1}^{LN}\mathbb{E}\Delta_n\Delta_{n+1} = \frac{1}{N}\sum_{n=1}^{N}\sum_{k=1}^{N}u_{n,k}u_{n+1,k}\frac{1}{L}\sum_{l=1}^{L}\mathbb{E}\left[\left(B_l^{(k)} - \hat{B}_l^{(k)}\right)^2\right]. \quad (57)$$

II. For any $N \in \mathbb{N}$ and $\bar{R} > 0$,

$$\lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N-1}\sum_{k=1}^{N}u_{k,n}u_{k,n+1}\min\{\lambda_k, \theta\} = D_{\bar{W}}(\bar{R}) - \frac{1}{2}G(\bar{R}) \quad (58)$$

where

$$G(\bar{R}_\theta) = \int_0^1 \frac{\min\left\{S_{\bar{W}}(\phi), \theta\right\}}{S_{\bar{W}}(\phi)}d\phi.$$

III. For any f_s and R,

$$D_{\mathrm{CE}}(f_s, R) = \frac{1}{6\ f_s} + \frac{2}{3}D_{\bar{W}}(\bar{R}) + \frac{1}{3}\left(D_{\bar{W}}(\bar{R}) - \frac{1}{2}G(\bar{R})\right).$$

Proof of Claim I: For $l = 1, \ldots, L$ denote $Y^{(l)} = \mathbf{U}\bar{W}^{(l)}$ and $\hat{Y}^{(l)} = \mathbf{U}\hat{\bar{W}}^{(l)}$. Note that $B_l^{(n)} = Y_n^{(l)}$ and, since $B^{(n)}$ and $\hat{B}^{(n)}$ are independent from $B^{(k)}$ and $\hat{B}^{(k)}$ for $k \neq n$, we have that

$$\mathbb{E}\left(Y_n^{(l)} - \hat{Y}_n^{(l)}\right)\left(Y_k^{(l)} - \hat{Y}_k^{(l)}\right) = 0,$$

for $k \neq n$. Next,

$$\begin{aligned}
\frac{1}{NL}\sum_{n=1}^{LN}\mathbb{E}\Delta_n\Delta_{n+1}
&= \frac{1}{LN}\sum_{n=1}^{N-1}\sum_{l=1}^{L}\mathbb{E}\left(\bar{W}_n^{(l)} - \hat{\bar{W}}_n^{(l)}\right)\left(\bar{W}_{n+1}^{(l)} - \hat{\bar{W}}_{n+1}^{(l)}\right) \\
&= \frac{1}{N}\sum_{n=1}^{N-1}\frac{1}{L}\sum_{l=1}^{L}\mathbb{E}\left[\sum_{k=1}^{N}u_{n,k}\left(Y_k^{(l)} - \hat{Y}_k^{(l)}\right) \times \sum_{p=1}^{N}u_{n+1,p}\left(Y_p^{(l)} - \hat{Y}_p^{(l)}\right)\right] \\
&= \frac{1}{N}\sum_{n=1}^{N-1}\frac{1}{L}\sum_{l=1}^{L}\sum_{k=1}^{N}u_{n,k}u_{n+1,k}\mathbb{E}\left[\left(Y_k^{(l)} - \hat{Y}_k^{(l)}\right)^2\right] \\
&= \frac{1}{N}\sum_{n=1}^{N-1}\sum_{k=1}^{N}u_{n,k}u_{n+1,k}\frac{1}{L}\sum_{l=1}^{L}\mathbb{E}\left[\left(B_l^{(k)} - \hat{B}_l^{(k)}\right)^2\right]. \quad (59)
\end{aligned}$$

Proof of Claim II: We have

$$\begin{aligned}
&\frac{1}{N}\sum_{n=1}^{N-1} u_{k,n}u_{k,n+1} \\
&= \frac{A_k^2}{N}\sum_{n=1}^{N-1}\sin\left(\frac{2k-1}{2N+1}n\pi\right)\sin\left(\frac{2k-1}{2N+1}(n+1)\pi\right) \\
&= \frac{A_k^2}{2N}\sum_{n=1}^{N-1}\left(\cos\left(\frac{2k-1}{2N+1}\pi\right)-\cos\left(\frac{2k-1}{2N+1}\pi(2n+1)\right)\right) \\
&= \frac{A_k^2}{2}\cos\left(\frac{2k-1}{2N+1}\pi\right)+o(1) \qquad (60)
\end{aligned}$$

where the last transition is since

$$\sum_{n=1}^{N-1}\cos\left(\frac{2k-1}{2N+1}\pi(2n+1)\right)$$

is bounded in N. From (60) we obtain:

$$\begin{aligned}
&\frac{1}{N}\sum_{n=1}^{N-1}\mathbb{E}\Delta_n\Delta_{n+1} \\
&= \frac{1}{N}\sum_{n=1}^{N-1}\sum_{k=1}^{N} u_{k,n}u_{k,n}\min\{\theta,\lambda_k\} \\
&= \frac{1}{N}\sum_{k=1}^{N}(NA_k^2)\min\{\theta,\lambda_k\}\frac{1}{N}\sum_{n=1}^{N-1}u_{k,n}u_{k,n+1} \\
&= \frac{1}{2N}\sum_{k=1}^{N}\min\{\theta,\lambda_k\}(NA_k^2)\left(\cos\left(\frac{2k-1}{2N+1}\pi\right)+O(1)\right). \qquad (61)
\end{aligned}$$

We now take the limit $N\to\infty$ as $k/N\to\phi$, so the spectrum of $\Sigma_{\bar{W}}$ converges to $S_{\bar{W}}(\phi)$. Moreover, since

$$\begin{aligned}
A_k^{-2} &= \sum_{l=1}^{N}(u_{k,l})^2 = \sum_{l=1}^{N}\sin^2\left(\frac{2k-1}{2N+1}\pi l\right) \\
&= \sum_{l=1}^{N}\left(\frac{1}{2}-\frac{1}{2}\cos\left(2\frac{2k-1}{2N+1}\pi l\right)\right),
\end{aligned}$$

we have $NA_k^2\to 2$. Therefore, after multiplying by σ^2/f_s to obtain the un-normalized distortion, (61) converges to

$$\begin{aligned}
&\frac{\sigma^2}{f_s}\int_0^1\min\left\{S_{\bar{W}}(\phi),\theta\right\}\cos(\pi\phi)d\phi \\
&= D_{\bar{W}}(\bar{R})-2\frac{\sigma^2}{f_s}\int_0^1\min\left\{S_{\bar{W}}(\phi),\theta\right\}\sin^2(\pi\phi/2)d\phi \\
&= D_{\bar{W}}(\bar{R})-\frac{\sigma^2}{2f_s}\int_0^1\frac{\min\left\{S_{\bar{W}}(\phi),\theta\right\}}{S_{\bar{W}}(\phi)}d\phi \\
&= D_{\bar{W}}(\bar{R})-\frac{1}{2}G(\bar{R}).
\end{aligned}$$

Proof of Claim III: We have

$$D_{\bar{W}}(\bar{R})-\frac{1}{6}G(\bar{R})=\int_0^1\min\{S_{\bar{W}}(\phi),\theta\}\left(1-\frac{1}{6S_{\bar{W}}(\phi)}\right)d\phi,$$

so

$$D_{\mathrm{CE}}(f_s,R)=\frac{1}{6f_s}+D_{\bar{W}}(\bar{R})-\frac{1}{6}G(\bar{R}).$$

We now use Claims I-III to prove (i) and (ii) in Theorem 8. To show (ii), we fix $\bar{R}=R/f_s$ and consider the converse for the source coding theorem for encoding the L-dimensional vector source $B^{(k)}$, consisting of i.i.d. Gaussian random variables of variance λ_k, using $L\bar{R}_k$ bits. This converse implies that for any decoder g of the form (56),

$$\frac{1}{L}\sum_{l=1}^{L}\mathbb{E}\left[\left(B_l^{(k)}-\hat{B}_l^{(k)}\right)^2\right]\geq\lambda_k 2^{-2\bar{R}_k}=\min\{\theta,\lambda_k\}.$$

Therefore, using claim I,

$$\frac{1}{NL}\sum_{n=1}^{LN}\mathbb{E}\Delta_n\Delta_{n+1}\geq\frac{1}{N}\sum_{n=1}^{N}\sum_{k=1}^{N}u_{n,k}u_{n+1,k}\lambda_k\min\{\theta,\lambda_k\}.$$

In addition, we use the converse for the source coding theorem for the discrete-time Wiener process $\bar{W}_{[\cdot]}$ from [6] to obtain

$$\mathsf{mmse}\left(\bar{W}^{LN}|\bar{f}\left(\bar{W}^{LN}\right)\right)\geq D_{\bar{W}}(\bar{R}). \qquad (62)$$

It follows from II that for any $\epsilon>0$, there exists T_0 large enough such that, for any $T>T_0$,

$$\frac{1}{NL}\sum_{n=1}^{LN}\mathbb{E}\Delta_n\Delta_{n+1}+\epsilon/3 > D_{\bar{W}}(\bar{R})-\frac{1}{2}G(\bar{R}),$$

$$\mathsf{mmse}\left(\bar{W}^{LN}|\bar{f}\left(\bar{W}^{LN}\right)\right)+\epsilon/3 > D_{\bar{W}}(\bar{R}),$$

and

$$\mathsf{mmse}(W_{(0:LT)}|\bar{W}^{LT})+\varepsilon/3\geq\frac{1}{6f_s}.$$

Finally, from (17) we obtain

$$\begin{aligned}
&\mathsf{mmse}\left(W_{(0:LT)}|\bar{f}\left(\bar{W}^{LN}\right)\right) \\
&\geq \mathsf{mmse}(W_{(0:LT)}|\bar{W}^{LN}) \\
&\quad+\frac{2}{3}\mathsf{mmse}\left(\bar{W}^{LN}|\bar{f}\left(\bar{W}^{LN}\right)\right) \\
&> \frac{1}{6f_s}+\frac{2}{3}D_{\bar{W}}(\bar{R})+\frac{1}{3}\left(D_{\bar{W}}(\bar{R})-\frac{1}{2}G(\bar{R})\right)-\varepsilon.
\end{aligned}$$

Since L is arbitrary and using III, we conclude that for any $\varepsilon>0$, there exists T_0 such that

$$\mathsf{mmse}\left(W_{(0:T)}|\bar{f}\left(\bar{W}^{N}\right)\right)+\varepsilon\geq D_{\mathrm{CE}}(f_s,R).$$

In order to prove (i), fix $\rho,\varepsilon,\delta>0$, and consider a decoder g that, upon receiving $(i_1,\ldots,i_{n_{\max}})$, first computes the inverse transform $\mathbf{U}^T\hat{b}(i_n)$ for each index i_n and concatenates the resulting vectors to obtain $\hat{\bar{W}}^{LN_T}$. In order to estimate $W_{(0:TL)}$, the decoder uses an interpolation similar to (11):

$$\hat{W}_t\triangleq\frac{t^+-t}{T_s}\hat{\bar{W}}_{t^-}+\frac{t-t^-}{T_s}\hat{\bar{W}}_{t^+},\quad t\in[0,TL]. \qquad (63)$$

To analyze the distortion resulting from using this decoder, consider first the L dimensional vector $B^{(k)}$ using $\bar{R}_k + \rho$ codewords drawn i.i.d. from $\mathcal{N}(0, [\lambda_k - \theta]^+)$, where

$$\bar{R}_k = \begin{cases} \frac{1}{2}\log[\lambda_k/\theta], & k \leq k_{\max} \\ 0 & k > k_{\max}. \end{cases}$$

For any $T > 0$ there exists L_0 that is independent of T, such that for any $k = 1, \ldots, k_{\max}$ and $L > L_0$,

$$\frac{1}{L}\sum_{l=1}^{L} \mathbb{E}\left(B_l^{(k)} - \hat{B}_l^{(k)}\right)^2 - \varepsilon/3 \leq \lambda_k 2^{-2R_i} = \theta. \tag{64}$$

Substituting (64) in (59), we conclude that for $L \geq L_0$,

$$\frac{1}{NL}\sum_{n=1}^{LN} \mathbb{E}\Delta_n \Delta_{n+1} - \varepsilon/3 \leq \frac{1}{N}\sum_{n=1}^{N}\sum_{k=1}^{N} u_{n,k}u_{n+1,k}\min\{\lambda_k, \theta\}.$$

Next, let T_0 be such that for all $T > T_0$,

$$\left|\mathsf{mmse}\left(W^{L_0T}|\bar{W}^{L_0N}\right) - \mathsf{mmse}(f_s)\right| < \varepsilon/3.$$

Using the achievability side of the source coding theorem with respect to $\bar{W}_{[\cdot]}$ from [6], we may choose $T > T_0$ and $L = \sqrt{T} > L_0$ such that

$$\mathsf{mmse}\left(\bar{W}^{LN}|\bar{f}\left(\bar{W}^{LN}\right)\right) < D_{\bar{W}}(\bar{R}) + \varepsilon/3,$$

and therefore

$$\mathsf{mmse}\left(W_{(0:LT)}|\bar{f}\left(\bar{W}^{LN}\right)\right) \leq D_{\mathrm{CE}}(f_s, R) + \varepsilon.$$

References

[1] A. Kipnis, Y. C. Eldar, and A. J. Goldsmith, "Information rates of sampled Wiener processes," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jul. 2016, pp. 740–744.

[2] I. Karatzas and S. E. Shreve, *Methods of Mathematical Finance*, vol. 39. Springer, 1998.

[3] D. Siegmund, *Sequential Analysis: Tests and Confidence Intervals*. Springer, 2013.

[4] I. Karatzas and S. E. Shreve, *Brownian Motion and Stochastic Calculus* (Graduate Texts in Mathematics). New York, NY, USA: Springer, 1991.

[5] G. J. Foschini and G. Vannucci, "Characterizing filtered light waves corrupted by phase noise," *IEEE Trans. Inf. Theory*, vol. 34, no. 6, pp. 1437–1448, Nov. 1988.

[6] T. Berger, "Information rates of Wiener processes," *IEEE Trans. Inf. Theory*, vol. 16, no. 2, pp. 134–139, Mar. 1970.

[7] O. Rioul and P. Duhamel, "Fast algorithms for discrete and continuous wavelet transforms," *IEEE Trans. Inf. Theory*, vol. 38, no. 2, pp. 569–586, Mar. 1992.

[8] Y. C. Eldar, *Sampling Theory: Beyond Bandlimited Systems*. Cambridge, U.K.: Cambridge Univ. Press, 2015.

[9] W. Vervaat, "Sample path properties of self-similar processes with stationary increments," *Ann. Probab.*, vol. 13, no. 1, pp. 1–27, 1985.

[10] T. Berger, *Rate Distortion Theory: A Mathematical Basis for Data Compression*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1971.

[11] J. K. Wolf and J. Ziv, "Transmission of noisy information to a noisy receiver with minimum distortion," *IEEE Trans. Inf. Theory*, vol. IT-16, no. 4, pp. 406–411, Jul. 1970.

[12] R. Gray, "Information rates of autoregressive processes," *IEEE Trans. Inf. Theory*, vol. IT-16, no. 4, pp. 412–421, Jul. 1970.

[13] A. Sahai and S. Mitter, "The necessity and sufficiency of anytime capacity for stabilization of a linear system over a noisy communication link—Part I: Scalar systems," *IEEE Trans. Inf. Theory*, vol. 52, no. 8, pp. 3369–3395, Aug. 2006.

[14] D. L. Neuhoff and S. S. Pradhan, "Information rates of densely sampled data: Distributed vector quantization and scalar quantization with transforms for Gaussian sources," *IEEE Trans. Inf. Theory*, vol. 59, no. 9, pp. 5641–5664, Sep. 2013.

[15] S. S. Pradhan and K. Ramchandran, "Distributed source coding using syndromes (DISCUS): Design and construction," *IEEE Trans. Inf. Theory*, vol. 49, no. 3, pp. 626–643, Mar. 2003.

[16] Z. Xiong, A. D. Liveris, and S. Cheng, "Distributed source coding for sensor networks," *IEEE Signal Process. Mag.*, vol. 21, no. 5, pp. 80–94, Sep. 2004.

[17] E. Akyol, U. Mitra, E. Tuncel, and K. Rose, "Source coding in the presence of exploration-exploitation tradeoff," in *Proc. IEEE Int. Symp. Inf. Theory*, Jun./Jul. 2014, pp. 2057–2061.

[18] V. P. Boda and P. Narayan, "Sampling rate distortion," *IEEE Trans. Inf. Theory*, vol. 63, no. 1, pp. 563–574, Jan. 2017.

[19] K. Nar and T. Basar, "Sampling multidimensional Wiener processes," in *Proc. IEEE 53rd Annu. Conf. Decis. Control (CDC)*, Dec. 2014, pp. 3426–3431.

[20] Y. Sun, Y. Polyanskiy, and E. Uysal-Biyikoglu, "Remote estimation of the Wiener process over a channel with random delay," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2017, pp. 321–325, doi: 10.1109/ISIT.2017.8006542.

[21] A. Kipnis, A. Goldsmith, Y. C. Eldar, and T. Weissman, "Distortion rate function of sub-Nyquist sampled Gaussian sources," *IEEE Trans. Inf. Theory*, vol. 62, no. 1, pp. 401–429, Jan. 2016.

[22] A. Kipnis, Y. C. Eldar, and A. J. Goldsmith, "Fundamental distortion limits of analog-to-digital compression," *IEEE Trans. Inf. Theory*, vol. 64, no. 9, pp. 6013–6033, Sep. 2018.

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

[24] A. Kipnis, G. Reeves, Y. C. Eldar, and A. J. Goldsmith, "Compressed sensing under optimal quantization," in *Proc. IEEE Int. Symp. Inf. Theory*, Jun. 2017, pp. 2148–2152.

[25] H. S. Witsenhausen, "Indirect rate distortion problems," *IEEE Trans. Inf. Theory*, vol. IT-26, no. 5, pp. 518–521, Sep. 1980.

[26] A. Kipnis, S. Rini, and A. J. Goldsmith. (2017). "Compress and estimate in multiterminal source coding." [Online]. Available: https://arxiv.org/abs/1602.02201

[27] A. Kipnis, G. Reeves, and Y. C. Eldar, "Single letter formulas for quantized compressed sensing with Gaussian codebooks," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2018, pp. 71–75.

[28] G. Murray, A. Kipnis, and A. J. Goldsmith, "Lossy compression of decimated Gaussian random walks," in *Proc. 52nd Annu. Conf. Inf. Sci. Syst. (CISS)*, Mar. 2018, pp. 1–6.

[29] S. M. Zemyan, *The Classical Theory of Integral Equations: A Concise Treatment*. Springer, 2012.

[30] A. Kolmogorov, "On the Shannon theory of information transmission in the case of continuous signals," *IRE Trans. Inf. Theory*, vol. 2, no. 4, pp. 102–108, Dec. 1956.

[31] T. Fine, "Properties of an optimum digital system and applications," *IEEE Trans. Inf. Theory*, vol. 10, no. 4, pp. 287–296, Oct. 1964.

[32] A. Kanlis, S. Khudanpur, and P. Narayan, "Typicality of a good rate-distortion code," *Problems Inf. Transmiss.*, vol. 32, no. 1, pp. 96–103, 1996.

[33] T. Weissman and E. Ordentlich, "The empirical distribution of rate-constrained source codes," *IEEE Trans. Inf. Theory*, vol. 51, no. 11, pp. 3718–3733, Nov. 2005.

[34] M. S. Pinsker, *Information and Information Stability of Random Variables and Processes* (Holden-Day Series in Time Series Analysis). San Francisco, CA, USA: Holden-Day, 1964.

[35] T. Berger, "Rate distortion theory for sources with abstract alphabets and memory," *Inf. Control*, vol. 13, no. 3, pp. 254–273, 1968.

[36] R. M. Gray, *Entropy and Information Theory*, vol. 1. New York, NY, USA: Springer, 2011.

[37] T. Fine, "The response of a particular nonlinear system with feedback to each of two random processes," *IEEE Trans. Inf. Theory*, vol. IT-14, no. 2, pp. 255–264, Mar. 1968.

[38] I. Gelfand and A. Iaglom, *Calculation of the Amount of Information About a Random Function Contained in Another Such Function*. Providence, RI, USA: AMS, 1959.

[39] A. Kipnis, A. J. Goldsmith, and Y. C. Eldar, "The distortion rate function of cyclostationary Gaussian processes," *IEEE Trans. Inf. Theory*, vol. 64, no. 5, pp. 3810–3824, May 2018.

Alon Kipnis (S'14) received the B.Sc. degree in mathematics (summa cum laude) and the B.Sc. degree in electrical engineering (summa cum laude) in 2010, and the M.Sc in mathematics in 2012, all from Ben-Gurion University of the Negev. He recently received the Ph.D. degree in electrical engineering from Stanford University, California, where he is now a postdoctoral scholar

in the Department of Statistics. His research focuses on the intersection of data compression with signal processing, machine learning and statistics.

Andrea J. Goldsmith (S'90–M'93–SM'99–F'05) is the Stephen Harris professor in the School of Engineering and a professor of Electrical Engineering at Stanford University. She co-founded and served as Chief Technical Officer of Plume WiFi and of Quantenna (QTNA), and she currently serves on the Corporate or Technical Advisory Boards of Crown Castle Inc. (CCI), Interdigital Corp. (IDCC), Sequans (SQNS), Quantenna (QTNA) and Cohere. She has also held industry positions at Maxim Technologies, Memorylink Corporation, and AT&T Bell Laboratories. Dr. Goldsmith is a member of the National Academy of Engineering and the American Academy of Arts and Sciences and has received several awards for her work, including the IEEE Eric E. Sumner Technical Field Award, the ACM Athena Lecturer Award, the IEEE ComSoc Edwin H. Armstrong Achievement Award, the ComSoc Technical Achievement Awards in Communications Theory and in Wireless Communications, the ComSoc WICE mentoring award, and the Silicon Valley/San Jose Business Journal's Women of Influence Award. She is author of the book *Wireless Communications* and co-author of the books *MIMO Wireless Communications* and *Principles of Cognitive Radio*, all published by Cambridge University Press, as well as an inventor on 29 patents. Her research interests are in information theory and communication theory, and their application to wireless communications, distributed systems, and neuroscience. She received the B.S., M.S. and Ph.D. degrees in Electrical Engineering from U.C. Berkeley. Dr. Goldsmith participates actively in committees and conference organization for the IEEE Information Theory and Communications Societies and has served on the Board of Governors for both societies. She has been a Distinguished Lecturer for both societies, served as the President of the IEEE Information Theory Society in 2009, founded and chaired the student committee of the IEEE Information Theory society, and is the founding chair of the IEEE TAB Committee on Diversity and Inclusion. At Stanford she has served as Chair of Stanford's Faculty Senate and for multiple terms as a Senator, and on its Appointments and Promotions Advisory Board, Budget Group, Committee on Research, Planning and Policy Board, Commissions on Graduate and on Undergraduate Education, Faculty Women's Forum Steering Committee, and Task Force on Women and Leadership.

Yonina C. Eldar (S'98–M'02–SM'07–F'12) received the B.Sc. degree in Physics in 1995 and the B.Sc. degree in Electrical Engineering in 1996 both from Tel-Aviv University (TAU), Tel-Aviv, Israel, and the Ph.D. degree in Electrical Engineering and Computer Science in 2002 from the Massachusetts Institute of Technology (MIT), Cambridge. She is currently a Professor in the Department of Electrical Engineering at the Technion - Israel Institute of Technology, Haifa, Israel, where she holds the Edwards Chair in Engineering. She is also a Research Affiliate with the Research Laboratory of Electronics at MIT, an Adjunct Professor at Duke University, and was a Visiting Professor at Stanford University, Stanford, CA. She is a member of the Israel Academy of Sciences and Humanities (elected 2017), an IEEE Fellow and a EURASIP Fellow. Her research interests are in the broad areas of statistical signal processing, sampling theory and compressed sensing, optimization methods, and their applications to biology and optics. Dr. Eldar has received many awards for excellence in research and teaching, including the IEEE Signal Processing Society Technical Achievement Award (2013), the IEEE/AESS Fred Nathanson Memorial Radar Award (2014), and the IEEE Kiyo Tomiyasu Award (2016). She was a Horev Fellow of the Leaders in Science and Technology program at the Technion and an Alon Fellow. She received the Michael Bruno Memorial Award from the Rothschild Foundation, the Weizmann Prize for Exact Sciences, the Wolf Foundation Krill Prize for Excellence in Scientific Research, the Henry Taub Prize for Excellence in Research (twice), the Hershel Rich Innovation Award (three times), the Award for Women with Distinguished Contributions, the Andre and Bella Meyer Lectureship, the Career Development Chair at the Technion, the Muriel & David Jacknow Award for Excellence in Teaching, and the Technion's Award for Excellence in Teaching (two times). She received several best paper awards and best demo awards together with her research students and colleagues including the SIAM outstanding Paper Prize, the UFFC Outstanding Paper Award, the Signal Processing Society Best Paper Award and the IET Circuits, Devices and Systems Premium Award, and was selected as one of the 50 most influential women in Israel. She was a member of the Young Israel Academy of Science and Humanities and the Israel Committee for Higher Education. She is the Editor in Chief of *Foundations and Trends in Signal Processing*, a member of the IEEE Sensor Array and Multichannel Technical Committee and serves on several other IEEE committees. In the past, she was a Signal Processing Society Distinguished Lecturer, member of the IEEE Signal Processing Theory and Methods and Bio Imaging Signal Processing technical committees, and served as an associate editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING, the *EURASIP Journal of Signal Processing*, the *SIAM Journal on Matrix Analysis and Applications*, and the *SIAM Journal on Imaging Sciences*. She was Co-Chair and Technical Co-Chair of several international conferences and workshops. She is author of the book *Sampling Theory: Beyond Bandlimited Systems* and co-author of the books *Compressed Sensing* and *Convex Optimization Methods in Signal Processing and Communications*, all published by Cambridge University Press.