Fundamental Distortion Limits of Analog-to-Digital Compression

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

***Abstract*— Representing a continuous-time signal by a set of samples is a classical problem in signal processing. We study this problem under the additional constraint that the samples are quantized or compressed in a lossy manner under a limited bitrate budget. To this end, we consider a combined sampling and source coding problem in which an analog stationary Gaussian signal is reconstructed from its encoded samples. These samples are obtained by a set of bounded linear functionals of the continuous-time path, with a limitation on the average number of samples per unit time given in this setting. We provide a full characterization of the minimal distortion in terms of the sampling frequency, the bitrate, and the signal's spectrum. Assuming that the signal's energy is not uniformly distributed over its spectral support, we show that for each compression bitrate there exists a critical sampling frequency smaller than the Nyquist rate, such that the distortion in signal reconstruction when sampling at this frequency is minimal. Our results can be seen as an extension of the classical sampling theorem for bandlimited random processes in the sense that they describe the minimal amount of excess distortion in the reconstruction due to lossy compression of the samples and provide the minimal sampling frequency required in order to achieve this distortion. Finally, we compare the fundamental limits in the combined source coding and sampling problem to the performance of pulse code modulation, where each sample is quantized by a scalar quantizer using a fixed number of bits.**

***Index Terms*— Sampling, analog to digital, source coding, lossy compression, sub-Nyquist sampling, nonuniform sampling, Gaussian processes.**

I. Introduction

THE minimal sampling rate required for perfect reconstruction of a bandlimited continuous-time process from its samples is given by the celebrated works of Whittaker, Kotelnikov, Shannon and Landau [4]. These results, however, focus only on performance associated with sampling rates; they do not incorporate other sampling parameters, in particular the quantization precision of the samples. This work aims to develop a theory of sampling and associated fundamental performance bounds that incorporates both sampling rate as well as quantization precision.

The Shannon-Kotelnikov-Whittaker sampling theorem states that sampling a signal at its Nyquist rate is a sufficient condition for exact recreation of the signal from its samples. However, quoting Shannon [5]:

> …"we are not interested in exact transmission when we have a continuous [amplitude] source [signal], but only in transmission to within a certain [distortion] tolerance…".

It is in fact impossible to obtain an exact digital representation of any continuous amplitude signal due to the finite precision of the samples. Hence, any digital representation of an analog signal is prone to some error, regardless of the sampling rate. This raises the question as to whether the condition of Nyquist rate sampling can be relaxed when we are interested in converting an analog signal to bits at a given bitrate (bits per unit time), such that the associated point on the distortion-rate function (DRF) of the signal is achieved.

The DRF describes the minimal distortion for any digital representation of a given signal under a fixed number of bits per unit time. While this implies that the DRF provides a theoretical limit on the distortion as a result of analog to digital (A/D) conversion, in fact, A/D conversion involves both

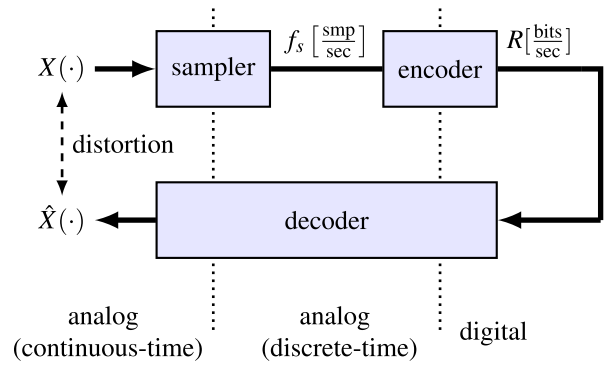

Fig. 1. Analog-to-digital compression (ADX) and reconstruction setting. Our goal is to derive the minimal distortion between the signal and its reconstruction from a lossy compressed version of its samples, where R is the compression bitrate and f_s is the sampling rate.

Manuscript received February 15, 2016; revised February 20, 2018 and April 30, 2018; accepted May 29, 2018. Date of publication July 20, 2018; date of current version August 16, 2018. This work was supported in part by the NSF Center for Science of Information under Grants CCF-0939370 and CCF-1320628, in part by NSF-BSF under Grant 2015711, and in part by the NSF/ENG/ECCS-BSF under Grant 1609695. Y. C. Eldar was supported in part by the European Union's Horizon 2020 Research and Innovation Program under Grant ERC-BNYQ and in part by the Israel Science Foundation under Grant 335/14. This paper was presented in part at the 2014 52nd Annual Allerton Conference on Communication, Control, and Computing [1], in part at the 2015 53rd Annual Allerton Conference on Communication, Control, and Computing [2], and in part at the 2015 Information Theory Workshop [3].

A. Kipnis was with the Department of Electrical Engineering, Stanford University, Stanford, CA 94305 USA. He is now with the Department of Statistics, Stanford University, Stanford, CA 94305 USA (e-mail: alonkipnis@gmail.com).

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

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

Communicated by E. Tuncel, Associate Editor for Source Coding.

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.2018.2857822

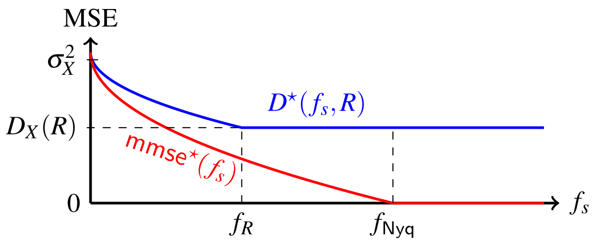

Fig. 2. Minimal distortion versus sampling rate. $D_X(R)$ is the information DRF describing the minimal distortion using lossy compression at bitrate R. $D^\star(R, f_s)$ is the minimal distortion using sampling at frequency f_s followed by lossy compression at bitrate R, and $\mathsf{mmse}(f_s)$ is the minimal distortion under sub-Nyquist sampling with infinite bit precision.

sampling a signal as well as converting those samples to bits, which entails some form of source coding, typically quantization. In some situations, it is possible to achieve the DRF of a continuous-time signal by mapping it into an equivalent discrete-time representation based on sampling at or above its Nyquist rate [6]. However, A/D technology limitations can preclude sampling signals at their Nyquist rate, particularly for wideband signals or under energy constraints [7], [8]. In such scenarios, the data available for source encoding is a sub-Nyquist sampled discrete-time representation of the signal [4], [8]. Our goal in this work is to consider the minimal distortion in recovering an analog signal from its samples with lossy compression of the samples at a prescribed bitrate, a setting which we call analog-to-digital compression (ADX) and is illustrated in Fig. 1. We are interested in particular in the optimal sampling rate to achieve this minimal distortion for a given lossy compression rate of the samples.

The distortion in ADX can be analyzed by considering the combined sampling and source coding setting studied in [9]. In this setting, the analog source signal is a Gaussian stationary process. This process, or a noisy version of it, is sampled at rate f_s, after which the samples are encoded using a code of rate R bits per unit time (R/f_s bits per sample on average). In the special case of scalar uniform sampling, zero noise, and assuming that f_s is above the Nyquist rate of the source signal, the encoder in Fig. 1 can estimate the signal with vanishing distortion prior to encoding it. As a result, in this case the distortion associated with sampling is zero, and the minimal ADX distortion is described by the DRF of the analog source signal. In this paper we ask the following question: given a source coding rate constraint R (for example, as a result of quantizing each sample using R/f_s bits), do we still need to sample at the Nyquist rate in order to achieve the DRF or is a lower sampling rate sufficient? By answering this question, we establish in this work a critical sampling rate f_R, which is in general lower than the Nyquist rate, such that sampling at this rate achieves the distortion-rate bound at bitrate R. This is illustrated in Fig. 2, where we see that sampling below the Nyquist rate is possible without additional distortion over that given by the DRF associated with Nyquist rate sampling.

Our results also imply that a picture similar to Fig. 2 holds even if we replace the uniform sampler by any bounded linear sampler. That is, each sample is obtained by a bounded linear functional applied to the continuous-time analog path, and we limit the average number of such samples obtained over a finite time interval to be at most f_s. In this case, the minimal f_s allowing zero sampling distortion under unlimited bitrate is the *spectral occupancy* or the *Landau rate* of the signal [10], [11], i.e., the Lebesgue measure of the support of it spectrum. We show that under a bitrate constraint R, the critical sampling rate f_R is always below the Landau rate for signals whose power is not uniformly distributed over their spectral support.

Our ADX setting also extends the expression for the fundamental distortion limit derived in [9] under uniform sampling to the class of all bounded linear samplers, and provides the optimal tradeoff between distortion, bitrate and sampling rate under a wide range of sampling models that are used in theory and practice. These include: filter-bank sampling, nonuniform sampling, multi-coset sampling [8], [11], and truncated wavelet transforms [4], [12]. In particular, the fundamental distortion limit we derive holds even if the process obtained by sampling the original signal does not have a known information theoretic distortion-rate characterization. For example, our results apply to sampling procedures resulting in non-ergodic processes.

When the signal is contaminated by noise before or during the sampling operation, there in no hope to achieve the DRF even with an unlimited sampling budget. Instead, the minimal distortion is described by the indirect DRF of the signal given its noisy version [13] [14, Sec. 4.5.4]. In this case, our results imply that the critical sampling rate f_R achieving the indirect DRF at bitrate R depends both on R and the noise, and can be attained in a similar manner as in the noise free setting.

Finally, we note that our ADX framework and characterization of the fundamental distortion limit hold even for signals that are not necessarily bandlimited, such as Gauss-Markov and autoregressive processes. For such a signal, the sampling distortion is non-zero for any finite sampling rate. Nevertheless, for each bitrate R, there exists a finite f_R such that the minimal ADX distortion in sampling at or above f_R equals the DRF of the signal. Consequently, f_R goes to infinity as R goes to infinity, and the asymptotic ratio R/f_R is the minimal number of bits per sample one must provide in order to make the ADX distortion vanish at the same rate as the sampling distortion.

In order to intuitively understand why optimal lossy compression performance can be attained by sampling below the Nyquist rate, one may consider the lossy compression of a signal represented by a sequence of independent Gaussian random variables. This representation is quite general since most signals of interest can be represented using their independent coefficients under some orthogonal basis transformation [15]. In order to compress such a sequence in an optimal manner subject to a minimum mean squared error (MSE) criterion in reconstruction, a random source code is obtained using the water-filling formula of Kolmogorov [16]. This formula implies that signal components with variances smaller than some threshold that depends on the bitrate are set to zero. As we explain in detail in Section II, the ratio between the number of coefficients exceeding this threshold and the original support of the distribution of the sequence can be seen as the optimal sampling rate required to attain the minimal distortion subject to the bit constraint.

For an analog stationary signal, its Fourier basis decomposition provides a canonical orthogonal representation. Hence, the main challenge in attaining the optimal lossy compression at bitrate R by sampling at rate f_R is in "aligning" the distribution of the sampled signal in the Fourier domain with the optimal lossy compression attaining distribution. When f_s is below f_R, the optimal alignment is described by a function $D^\star(f_s, R)$ defined by a water-filling formula over f_s spectral bands of maximal energy (or maximal SNR in the noisy version). As we show, this "alignmnet" is attainable by uniform multi-branch sampling using appropriate LTI pre-sampling operations. Together with a matching converse theorem with respect to $D^\star(f_s, R)$ under any bounded linear sampler, we conclude that $D^\star(f_s, R)$ fully characterizes the distortion in ADX. In particular, our results imply that the class of multi-branch LTI uniform sampling is optimal, in the sense that the distortion attained by any bounded linear sampler can be attained by a multi-branch uniform sampler with a sufficient number of sampling branches.

We also examine the distortion-rate performance of a very simple and sub-optimal A/D scheme known as pulse-code modulation (PCM). This scheme consists of a scalar quantizer with a fixed number of bits per sample as an encoder and a linear non-causal decoder. We analyze this A/D scheme under a fixed bitrate budget, and show that there exists a distortion minimizing sampling rate that optimally trades off distortion due to sampling and due to quantization precision. This optimal sampling rate is at or below the Nyquist rate, and experiences a similar dependency on the bitrate as the critical ADX rate f_R. Our results also imply that, as opposed to the behavior of the optimal ADX distortion $D^\star(f_s, R)$, oversampling a bandlimited signal in PCM has a detrimental effect on the distortion.

To put our work into context, we now briefly review some of the well-known sampling theories and their relation to our results. The celebrated Shannon-Kotelnikov-Wittaker sampling theorem asserts that a bandlimited deterministic signal $x(\cdot)$ with finite L_2 norm can be perfectly reconstructed from its uniform samples at frequency $f_s > f_{\mathsf{Nyq}}$, where f_{Nyq} is the bandwidth of the signal. This statement can be refined when the exact support $\operatorname{supp} S_x$ of the Fourier transform of $x(\cdot)$ is known: $x(\cdot)$ can be obtained as the limit in L_2 of linear combinations of the samples $x(\mathbb{Z}/f_s)$ iff for all $k \neq n \in \mathbb{Z}$, $(\operatorname{supp} S_x + f_s k) \cap (\operatorname{supp} S_x + f_s n) = \emptyset$, where a reconstruction formula is also available [17]. Lloyd [18] provided an equivalent result for stationary stochastic processes, where the Fourier transform is replaced by the power spectral density (PSD). When sampling at the Nyquist rate is not possible, the minimal MSE (MMSE) in estimating a Gaussian stationary process from its uniform samples can be expressed in terms of its PSD [19]–[21]. This MMSE in the case of multi-branch sampling was derived in [9, Sec. IV].

In general, the estimation of any regular Gaussian stationary process from its partial observations can be translated into the problem of projections into Hilbert spaces generated by complex exponentials [22], [23]. In particular, when the PSD is supported over a compact set $S \subset \mathbb{R}$, then the closed linear span (CLS) of exponentials with support over S is isomorphic, by the Fourier transform operator, to the *Paley-Wiener space* $\mathrm{Pw}(S)$ of functions with Fourier transform supported in S. In this space, optimal reconstruction of signals from their samples is possible when the samples define a *frame* [10], [24]. Beurling and Carleson [25] and Landau [26] showed that a sufficient and necessary condition for a discrete set of time samples to define a frame in $\mathrm{Pw}(S)$ is that its Beurling density (also called *uniform density*) exceeds the Lebesgue measure $\mu(S)$ of S. In our setting $\mu(S)$ is the spectral occupancy of the signal, which we also refer to as its *Landau* rate and denote it by f_{Lnd}. For the optimization of the sampling times see [27]–[29] and the references therein. We also refer to [4] and [30]–[32] [4], [30]–[32] for additional background on sampling theory and generalized sampling techniques.

On the other side of the ADX setting is the distortion in lossy compression at a limited bitrate R. The optimal trade-off between the average quadratic distortion and bitrate in the description of a Gaussian stationary process $X(\cdot)$ is given by its quadratic DRF, denoted here by $D_X(R)$. This DRF was initially derived by Pinsker [16], and then extended by Dubroshin and Tsybakov [13] to the case where the process is contaminated by Gaussian noise. Both the noisy case explored by Dubroshin and Tsybakov and the ADX characterized in this work fall within the *indirect* or *remote* source coding setting [14, Sec. 4.5.4], in which the encoder has no direct access to the signal it tries to describe. Indirect source coding problems were also considered in [33]–[35].

The interplay between bit resolution in source coding and sampling rates arise in numerous settings. For sampling rates above the Nyquist rate, non trivial trade-offs between the oversampling rate and bitrate, under different encoding scenarios, can be found in [6] and [37]–[40]. In order to explore the trade-off between lossy compression and sub-Nyquist sampling rates, a combined sampling and source coding problem was recently introduced in [9] assuming uniform sampling. The ADX can be seen as an extension of the setting in [9] to any bounded linear sampling technique, and the determination of the minimal sampling rate f_R attaining the optimal source coding performance. Finally, in the context of compressed sensing (CS) [40], the optimal trade-off between the sampling rate and bitrate is explored in the high bitrate asymptotic in [41] and for a finite bitrate in [42]. We note that our results are not directly relevant to CS since we focus on sampling continuous-time Gaussian signals that are not sparse in any basis. Nevertheless, the discrete-time counterpart of our results may be applied to CS to obtain a lower bound on the distortion when the signal's support is given as side information, or an upper bound on the distortion when the samples of the signal are encoded using a Gaussian codebook [43].

The rest of the paper is organized as follows: in Section II we provide intuition for the dependency between sampling and lossy compression in representing finite dimensional random vectors. In Section III we define the ADX problem and the class of bounded linear samplers we treat. Our main results are given in Section IV. In Section V we consider scalar quantization encoding and compare its performance to the minimal ADX distortion. Concluding remarks are provided in Section VI.

II. Lossy Compression of Finite Dimensional Signals

As an introduction to the ADX setup, it is instructive to consider a simpler setting involving the sampling and lossy compression of signals represented as finite dimensional random real vectors.

Let $X^n = (X_1, \ldots, X_n)$ be an n-dimensional Gaussian random vector with covariance matrix Σ_{X^n}, and let $Y^m = (Y_1, \ldots, Y_m)$ be a projected version of X^n defined by

$$Y^m = \boldsymbol{H} X^n, \tag{1}$$

where $\boldsymbol{H} \in \mathbb{R}^{m \times n}$ is a deterministic matrix and $m < n$. This projection of X^n into a lower dimensional space is the counterpart for the sampling operation in the ADX setting of Fig. 1. We consider the normalized MMSE estimate of X^n from a representation of Y^m using a limited number of bits.

Without constraining the number of bits, the distortion in this estimation is given by

$$\mathsf{mmse}(X^n|Y^m) \triangleq \frac{1}{n}\mathrm{trace}\left(\Sigma_{X^n} - \Sigma_{X^n|Y^m}\right), \tag{2}$$

where $\Sigma_{X^n|Y^m}$ is the conditional covariance matrix of X^n given Y^m. When Y^m is encoded using a code of no more than nR bits, the minimal distortion cannot be smaller than the indirect DRF of X^n given Y^m, denoted by $D_{X^n|Y^m}(R)$. This function is given by the following parametric expression [13]:

$$\begin{aligned} D_{X^n|Y^m}(R_\theta) &= \mathrm{trace}\left(\Sigma_{X^n}\right) - \sum_{i=1}^{m}\left[\lambda_i\left(\Sigma_{X^n|Y^m}\right) - \theta\right]^+, \\ R_\theta &= \frac{1}{2}\sum_{i=1}^{m}\log^+\left[\lambda_i\left(\Sigma_{X^n|Y^m}\right)/\theta\right] \end{aligned} \tag{3}$$

where $x^+ \triangleq \max\{x, 0\}$, $\log^+[x] \triangleq [\log(x)]^+$, and $\lambda_i\left(\Sigma_{X^n|Y^m}\right)$ is the ith eigenvalue of $\Sigma_{X^n|Y^m}$.

It follows from (2) that X^n can be recovered from Y^m with zero MMSE if and only if

$$\lambda_i\left(\Sigma_{X^n}\right) = \lambda_i\left(\Sigma_{X^n|Y^m}\right), \quad \forall i \in \{1, \ldots, n\}. \tag{4}$$

When this condition is satisfied, (3) takes on the form

$$\begin{aligned} D_{X^n}(R_\theta) &= \sum_{i=1}^{n}\min\left\{\lambda_i\left(\Sigma_{X^n}\right), \theta\right\}, \\ R_\theta &= \frac{1}{2}\sum_{i=1}^{n}\log^+\left[\lambda_i\left(\Sigma_{X^n}\right)/\theta\right] \end{aligned} \tag{5}$$

which is Kolmogorov's reverse water-filling expression for the DRF of the vector Gaussian source X^n [16], i.e., the minimal distortion in encoding X^n using codes of rate R bits per source realization. The key insight is that the requirements for equality between (3) and (5) are not as strict as (4): all that is needed is equality among those eigenvalues that affect the value of (5). In particular, assume that for a point (R, D) on $D_{X^n}(R)$, only $\lambda_n(\Sigma_{X^n}), \ldots \lambda_{n-m+1}(\Sigma_{X^n})$ are larger than θ, where the eigenvalues are organized in ascending order. Then we can choose the rows of $\boldsymbol{H}$ to be the m left eigenvectors corresponding to $\lambda_n(\Sigma_{X^n}), \ldots \lambda_{n-m+1}(\Sigma_{X^n})$. With this choice of $\boldsymbol{H}$, the m largest eigenvalues of $\Sigma_{X^n|Y^m}$ are identical to the m largest eigenvalues of Σ_{X^n}, and (5) is equal to (3).

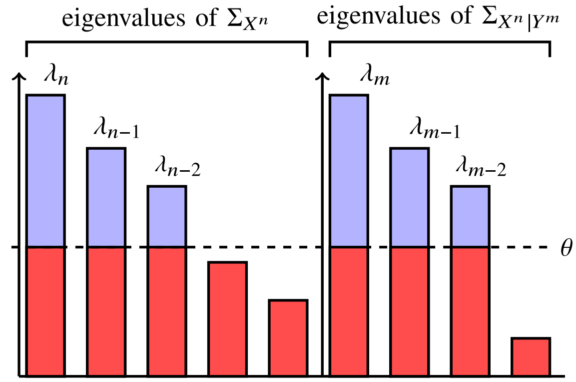

Fig. 3. Optimal sampling occurs whenever $D_{X^n}(R) = D_{X^n|Y^m}(R)$. This condition is satisfied even when $m < n$, as long as there is equality among the eigenvalues of Σ_{X^n} and $\Sigma_{X^n|Y^m}$ which are larger than the water-level parameter θ.

Since the rank of the sampling matrix is now $m < n$, we effectively performed sampling below the "Nyquist rate" of X^n without degrading the performance dictated by its DRF. One way to understand this phenomena is an alignment between the range of the sampling matrix $\boldsymbol{H}$ and the subspace over which X^n is represented, according to Kolmogorov's expression (5). When this expression implies that not all degrees of freedom are utilized by the optimal distortion-rate code, sub-sampling does not incur further performance loss provided the sampling matrix is aligned with the optimal code. This situation is illustrated in Fig. 3. Taking fewer rows than the actual rank of Σ_{X^n} is the finite-dimensional analog of sub-Nyquist sampling in the infinite-dimensional setting of continuous-time signals.

In the rest of this paper we explore the counterpart of the phenomena described above in the richer setting of continuous-time stationary processes that may or may not be bandlimited, and whose samples may be corrupted by additive noise. The precise problem description is given in the following section.

III. Problem Formulation and Preliminaries

A. ADX Setting

The ADX system is described in Fig. 4. We assume that $X(\cdot) = \{X(t),\ t \in \mathbb{R}\}$ is a zero-mean real Gaussian stationary process with a *known* PSD $S_X(f)$:

$$\mathbb{E}\left[X(t)X(s)\right] = \int_{-\infty}^{\infty} S_X(f) e^{2\pi j(t-s)f} df, \quad t, s \in \mathbb{R}. \tag{6}$$

In particular, $S_X(f)$ is in $\mathrm{L}_1(\mathbb{R})$ and the variance of $X(\cdot)$ is given by

$$\sigma_X^2 = \int_{-\infty}^{\infty} S_X(f) df.$$

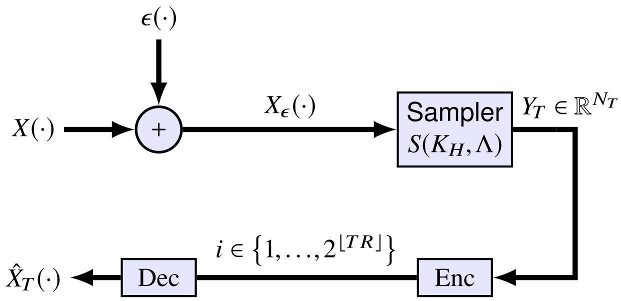

Fig. 4. ADX via a combined sampling and source coding setting with additive noise prior to sampling. We consider the distortion in recovering $X(\cdot)$ over $[-T/2, T/2]$ from a representation of its N_T samples using $\lfloor TR \rfloor$ bits, where N_T is the number of samples in $[-T/2, T/2]$ and N_T/T is bounded asymptotically by f_s.

The noise is another zero-mean real Gaussian stationary process $\epsilon(\cdot) = \{\epsilon(t), t \in \mathbb{R}\}$ independent of $X(\cdot)$ with PSD $S_\epsilon(f)$ and finite variance. We assume that the spectral measures of $X(\cdot)$ and $\epsilon(\cdot)$ are absolutely continuous with respect to the Lebesgue measure, so that their distribution is fully characterized by their PSDs.

The sampler S belongs to the class of bounded linear samplers to be defined in the sequel. This sampler receives the process

$$X_\epsilon(\cdot) \triangleq X(\cdot) + \epsilon(\cdot),$$

i.e., the noisy version of $X(\cdot)$, as its input. For a finite time horizon $T > 0$, the sampler S produces a finite number N_T of samples

$$Y_T \triangleq \left(Y_1, \ldots, Y_{N_T}\right) = S_T\left(X_\epsilon(\cdot)\right).$$

The assumption that the variance of the noise is finite excludes, for example, $\epsilon(\cdot)$ from being a white noise signal. This assumption is necessary to define sampling of $X_\epsilon(\cdot)$ in a meaningful way, as we explain below.

The encoder

$$f_{N_T} : \mathbb{R}^{N_T} \to \left\{1, \ldots, 2^{\lfloor TR \rfloor}\right\}, \tag{7}$$

receives the vector Y_T and outputs an index in $\left\{1, \ldots, 2^{\lfloor TR \rfloor}\right\}$. The decoder,

$$g_{N_T} : \left\{1, \ldots, 2^{\lfloor TR \rfloor}\right\} \to \mathbb{R}^{[-T/2, T/2]}, \tag{8}$$

upon receiving this index from the encoder, produces a reconstruction waveform $\hat{X}_T(\cdot)$. The goal of the joint operation of the encoder and the decoder is to minimize the average MSE

$$\frac{1}{T}\int_{-T/2}^{T/2} \mathbb{E}\left(X(t) - \widehat{X}_T(t)\right)^2 dt. \tag{9}$$

Given a particular bounded linear sampler S, and a bitrate R, we are interested in characterizing the function

$$D_T(S, R) \triangleq \inf \frac{1}{T}\int_{-T/2}^{T/2} \mathbb{E}\left(X(t) - \widehat{X}_T(t)\right)^2, \tag{10}$$

where the infimum is over all encoders and decoders of the form (7) and (8). We also consider the asymptotic version of (10):

$$D(S, R) \triangleq \liminf_{T\to\infty} D_T(S, R). \tag{11}$$

Before describing the class of bounded linear samplers, we remark on some of the properties of the ADX setting:

- Information loss in ADX is due to noise, sampling, and encoding. We do not consider limitations on the decoder that may exist in practice, such as memory or complexity.
- The additive noise $\epsilon(\cdot)$ may be seen as an external interference in transmitting $X(\cdot)$ or as noise associated with the sampling operation. With obvious adjustments, our setting can also handle a discrete-time noise vector with a stationary distribution added post sampling. For example, discrete-time white noise can be obtained from our setting by taking $\epsilon(\cdot)$ to be a flat spectrum noise with bandwidth equal to the sampling rate.
- For a finite time horizon T, the decoder is only required to recover $X(\cdot)$ over the interval $[-T/2, T/2]$. However, as follows from the description of the sampler below, each sample may depend on a realization of $X_\epsilon(\cdot)$ over the entire time-horizon (past and future). It is possible to restrict the sampler to be a function of $X(\cdot)$ only over $[-T/2, T/2]$ provided the conditional distribution of $X(\cdot)$ given its samples converges to an asymptotic distribution as T goes to infinity. Our asymptotic analysis below remains valid under this restriction due to the stationary distribution of $X(\cdot)$. Our setting also prohibits the sampler to depend on T; This restriction precludes adaptive sampling schemes such as in [44].
- Since $X(\cdot)$ is a stationary process, we can replace the interval $[-T/2, T/2]$ by any other interval of length T without affecting the main results.
- As opposed to common situations in source coding of stationary processes (e.g., [45, Lemma 10.6.2]), the liminf in (11) cannot be replaced by a simple infimum or a limit. One explanation for this difference is that, as we show below, the coding scheme that attains $D(f_s, R)$ essentially describes the estimator of $X(\cdot)$ from the samples Y_T, and the distribution of this estimator is not stationary in general.

B. Bounded Linear Sampling of Random Signals

We now describe the class of bounded linear samplers we use in the ADX setting of Fig. 4. Assume first that the input to the sampler is a deterministic signal $x(\cdot)$ in a class of signals $\mathscr{X}$. Each sample Y_n can then be seen as the result of applying a functional ϕ_n on $x(\cdot)$. In accordance with physical considerations of realizable systems, we require that the space of signals $\mathscr{X}$ is embedded in the Hilbert space of real functions of finite energy L_2, and that the functional defining the nth sample is linear and bounded. In other words, each sample is defined by an element of the dual space $\mathscr{X}^\star$ of $\mathscr{X}$. For this reason, we assume that $\mathscr{X}$ and $\mathscr{X}^\star$ are standard spaces of *test functions* and *distributions*, respectively [46], so that every distribution $\phi \in \mathscr{X}^\star$ has a Fourier transform in the Gelfand-Shilov sense [47]. Consequently, the bilinear operation $\langle \phi, x\rangle$ between $\phi \in \mathscr{X}^\star$ and $x \in \mathscr{X}$ satisfies the Plancherel identity

$$\langle \phi, x\rangle = \int_{-\infty}^{\infty} \mathscr{F}x(f)\left(\mathscr{F}\phi(f)\right)^* df, \tag{12}$$

where $\mathscr{F}$ is the Fourier transform and $*$ denotes complex conjugation. To summarize, for each $T > 0$ and assuming an appropriate class of input signals $\mathscr{X}$, the output of the sampler is defined by a set of N_T elements of $\mathscr{X}^\star$. We denote the samplers constructed in this manner as the class of *bounded linear samplers*.

Next, we consider bounded linear sampling of the random process $X_\epsilon(\cdot)$. Since the spectral measure of $X_\epsilon(\cdot)$ is absolutely continuous with respect to the Lebesgue measure μ, we have

$$\mathbb{E}\left[X_\epsilon(t)X_\epsilon(s)\right] = \int_{-\infty}^{\infty} w_t(f)w_s^*(f)S_{X_\epsilon}(f)df, \tag{13}$$

where we denoted $w_t(f) \triangleq e^{2\pi i f t}$. It follows from (13) that the mapping $X_\epsilon(t) \to w_t$ is an isometry, and as in [22], we extend this isometry to an isomorphism between the following Hilbert spaces: (1) the Hilbert space generated by the CLS of the process $X_\epsilon(\cdot)$ with norm $\|X_\epsilon(t)\|^2 = \mathbb{E}[X_\epsilon^2(t)]$, and (2) the space $\mathscr{W}(S_{X_\epsilon})$ which is the CLS of $\{w_t,\ t \in \mathbb{R}\}$ with an L_2 norm weighted by $S_{X_\epsilon}(f)$. This isomorphism allows us to define bounded linear sampling of $X_\epsilon(\cdot)$ by describing its operation on $\mathscr{W}(S_{X_\epsilon})$. Specifically, we identify $\mathscr{X}$ with the elements of $\mathscr{W}(S_{X_\epsilon})$ and set $\mathscr{X}^\star$ to be a space of distributions such that, for any $\phi \in \mathscr{X}^\star$, its Fourier transform $\hat{\phi}$ satisfies

$$\int_{-\infty}^{\infty} |\hat{\phi}(f)|^2 S_X(f)df < \infty. \tag{14}$$

For such $\phi_n \in \mathscr{X}^\star$, we define the sample

$$Y_n = \int_{-\infty}^{\infty} X_\epsilon(\tau)\phi_n^*(\tau)d\tau$$

to be the inverse image of $\langle \phi_n, w_t\rangle(f)$ under $X_\epsilon(t) \to w_t$. Although in most situations this inverse image cannot be found explicitly, we are only interested in the joint statistics of Y_n and $X(t)$, which is completely determined by

$$\mathbb{E}\left[Y_n X(t)\right] = \int_{-\infty}^{\infty} \langle \phi_n, w_t\rangle(f)e^{-2\pi i f t}S_{X_\epsilon}(f)dt. \tag{15}$$

In particular, condition (14) guarantees that the integral in (15) exists.

Example 1 (Pointwise Evaluation of Bandlimited Signals): Assume that ϕ_n is the Dirac distribution at t_n corresponding to pointwise evaluation at $t = t_n$, so that (14) holds and $\langle \phi_n, w_t\rangle = w_{t_n}$, whose inverse image is $X_\epsilon(t_n)$. If in addition $S_{X_\epsilon}(f)$ is supported on an open set $U \in \mathbb{R}$, then the element of $\mathscr{W}(S_{X_\epsilon})$ can be identified with the Paley-Wiener space of complex valued functions whose Fourier transform is supported on U. In this case, for most applications it is enough to take $\mathscr{X} = \mathscr{W}(S_{X_\epsilon})$ with its Hilbert space topology so that $\mathscr{X}^\star = \mathscr{X}$. For example, pointwise evaluation at $t = t_n$ for $x \in \mathscr{W}(S_{X_\epsilon}))$ is obtained by the inner product of x with $\mathscr{F}^{-1}(\mathbf{1}_U(f)e^{2\pi i f t_n})$, which is a member of $\mathscr{W}(S_{X_\epsilon})$.

In contrast to the scenario described in Example 1, we do not restrict ourselves to bandlimited signals at the sampler input. Thus, our setting supports any PSD $S_{X_\epsilon}(f)$ and corresponding set of functionals $\mathscr{X}^\star$ such that (14) holds.

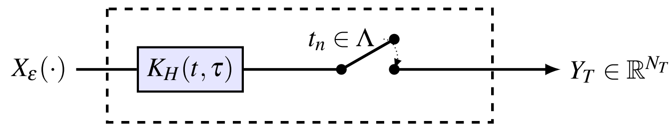

Fig. 5. Bounded linear sampler with a pre-sampling transformation with kernel K_H and a sampling set Λ.

Without loss of generality, it follows from the Schwartz kernel theorem [48] applied to $\mathscr{X} \times \mathscr{X}^\star$ that the sequence of functionals defining the samples can be described in terms of a bilinear kernel $K_H(t,s)$ on $\mathbb{R}\times\mathbb{R}$ and a discrete *sampling set* $\Lambda \subset \mathbb{R}$, as illustrated in Fig. 5. That is, the nth sample is given by

$$y_n \triangleq \int_{-\infty}^{\infty} X_\epsilon(s)K_H(t_n,s)ds.$$

In order to control the number of samples taken at every time horizon, we assume that Λ is uniformly discrete in the sense that there exists $\delta > 0$ such that $|t-s| > \delta$ for every non identical $t, s \in \Lambda$. For a time horizon T, we denote

$$\Lambda_T \triangleq \Lambda \cap [-T/2, T/2],$$

and define y_T to be the finite dimensional vector obtained by sampling $x_\epsilon(\cdot)$ at times $t_1,\ldots,t_n \in \Lambda_T$.

The assumption that Λ is uniformly discrete ensures that for any T, the *density* of Λ_T,

$$d_T(\Lambda) \triangleq \frac{\mathrm{card}\,(\Lambda_T)}{T},$$

if finite, and so is the limit

$$d^+(\Lambda) \triangleq \limsup_{T\to\infty} d_T(\Lambda).$$

We denote $d^+(\Lambda)$ as the *upper symmetric density* of Λ. Whenever it exists, we define the limit

$$d(\Lambda) = \lim_{T\to\infty} d(\Lambda_T) = \lim_{T\to\infty}\frac{\mathrm{card}\,(\Lambda\cap[-T/2,T/2])}{T},$$

as the *symmetric density* of Λ.

C. Multi-Branch LTI Uniform Sampling

An important special case of bounded linear sampling is described by the sampler in Fig. 6. This sampler has L sampling branches, where the lth branch consists of a linear time invariant (LTI) pre-sampling filter with transfer function $H_l(f)$ followed by a uniform sampler at rate f_s/L. Consequently, the nth sample produced by the lth branch is given by

$$Y_{l,n} = \int_{-\infty}^{\infty} h_l(nL/f_s - \tau)X_\epsilon(\tau).$$

We define

$$\mathbf{Y}_n = (Y_{1,n},\ldots,Y_{L,n})$$

as the nth output of all branches. For a finite time horizon T, the output of the sampler is

$$Y_T \triangleq \{\mathbf{Y}_n,\ |n| < \lfloor Tf_s/L\rfloor/2\},$$

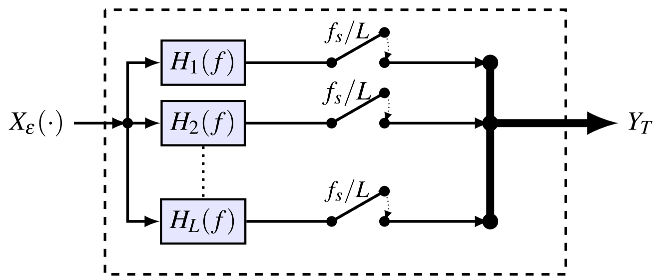

Fig. 6. Multi-branch linear time-invariant (MB-LTI) sampler.

so that Y_T incorporates at most $N_T = \lfloor T f_s \rfloor$ samples from the process at the input to the sampler.

The class of samplers obtained in this manner is called multi-branch LTI uniform samplers (MB-LTI), where we denote a single sampler from this class by $S_{f_s}(H_1, \ldots, H_L)$. In order to see that a MB-LTI is a bounded linear sampler, note that its nth sample can be defined by the functional $\phi_n = \phi_{kL+l}$, $k = 0, \ldots, N/L$, $l = 1, \ldots, L$,

$$\int_{-\infty}^{\infty} X_\epsilon(\tau)\phi_n(\tau)d\tau = \int_{-\infty}^{\infty} X_\epsilon(\tau)h_l(kL/f_s - \tau)d\tau.$$

A MB-LTI sampler belongs to the class of shift-invariant samplers [4], for which their output:

$$Y_\infty \triangleq \cup_{T>0} Y_{N_T}, \tag{16}$$

is invariant to time shifts by integer multiples of L/f_s in the input to the sampler $X_\epsilon(\cdot)$.

D. Properties of Optimal Encoding and Connection to Classical Results

We now explore basic properties of the functions $D_T(S, R)$ and $D(S, R)$ of (10) and (11) describing the minimal distortion in ADX. By doing so, we review previous results in sampling and source coding theory and explain their connection to our setting.

Denote by $\widetilde{X}_T(\cdot)$ the process that is obtained by MMSE estimation of $X(\cdot)$ from the vector of samples Y_T. Namely

$$\widetilde{X}_T(t) \triangleq \mathbb{E}[X(t)|Y_T], \quad t \in \mathbb{R}. \tag{17}$$

From properties of the conditional expectation, for any encoder f_{N_T} we have

$$\frac{1}{T}\int_{-T/2}^{T/2} \mathbb{E}\left(X(t) - \mathbb{E}[X(t)|Y_T]\right)^2 dt = \mathsf{mmse}_T(S) + \mathsf{mmse}\left(\widetilde{X}_T | f_{N_T}(Y_T)\right), \tag{18}$$

where $\widehat{X}_T(\cdot) = g_{N_T}\left(f_{N_T}(Y_T)\right)$,

$$\mathsf{mmse}_T(S) \triangleq \frac{1}{T}\int_{-T/2}^{T/2} \mathbb{E}\left(X(t) - \widetilde{X}_T(t)\right)^2 dt, \tag{19}$$

is the distortion associated only with sampling and noise, and

$$\mathsf{mmse}\left(\widetilde{X}_T | f_{N_T}(Y_T)\right) \triangleq \frac{1}{T}\int_{-T/2}^{T/2} \mathbb{E}\left(\widetilde{X}_T(t) - \mathbb{E}\left[\widetilde{X}_T(t) | f_{N_T}(Y_T)\right]\right)^2 dt$$

is the distortion associated with the lossy compression procedure, and depends on the sampler only through $\widetilde{X}_T(\cdot)$.

It follows from (18) that there is no loss in performance if the encoder tries to describe the process $\widetilde{X}_T(\cdot)$ subject to the bitrate constraint, rather than the process $X(\cdot)$. In addition, optimal decoding is obtained by outputting the conditional expectation of $\widetilde{X}_T(\cdot)$ given $f_{N_T}(Y_T)$. These observations, which hold in general in indirect source coding situations [14], were used in [13] to derive the indirect DRF of a pair of stationary Gaussian processes, and later in [34] to derive indirect DRF expressions in other settings. An extension of the principle presented in this decomposition to arbitrary distortion measures is discussed in [33].

The decomposition (18) is also related to the behavior of $D(S, R)$ under the two extreme cases illustrated in Fig. 2:

1) Unconstrained Bitrate: As the bitrate R goes to infinity, the MSE as a result of lossy compression goes to zero. Consequently, (18) implies that

$$\lim_{R\to\infty} D(S, R) = \inf_{R>0} D(S, R) = \mathsf{mmse}(S),$$

where $\mathsf{mmse}(S) = \liminf_{T\to\infty} \mathsf{mmse}_T(S)$.

As the sampling operation is linear and the signals are Gaussian, the optimal MSE estimator of $X(t)$ from Y_T is a linear function of Y_T. We therefore have

$$\mathbb{E}\left(X(t) - \mathbb{E}[X(t)|Y_T]\right)^2 = \inf_{\mathbf{a}\in\mathbb{R}^{N_T}} \mathbb{E}\left(X(t) - \sum_{t_n\in\Lambda_T} a_n Y_n\right)^2. \tag{20}$$

Under a MB-LTI sampler, an expression for (20) in the limit as T goes to infinity can be derived in closed form [20], [49], leading to a closed form expression for $\mathsf{mmse}(S)$. Although it is infeasible to obtain $\mathsf{mmse}_T(S)$ in a closed form for an arbitrary bounded linear sampler, it is sometimes possible to derive conditions on the density of λ such that $\mathsf{mmse}_T(S)$ converges to zero or to a MSE that is only due to noise. For example, assuming zero noise, $K_H(t, s) = \delta(t - s)$ the identity operator, and supp S_X is a finite union of bounded intervals, the condition on (20), and hence on $\mathsf{mmse}_T(S)$, to converge to zero is related to a classical problem in sampling theory studied by Beurling and Carleson [25] and Landau [26]. In order to see this relation, use (13) to translate the interpolation problem of (20) to the Hilbert space $\mathscr{W}(S_X)$. Interpolation in $\mathscr{W}(S_X)$ with vanishing MSE is equivalent to the same operation in the Paley-Wiener space of analytic functions whose Fourier transform vanishes outside supp S_X. Zero error in this interpolation is known to hold whenever the non-harmonic Fourier basis $\left\{e^{2\pi i t_n}, t_n \in \Lambda\right\}$ defines a frame in this Paley-Wiener space, i.e., there exists a universal constant $A > 0$ such that the L_2 norm of each function in this space is bounded by A times the energy of the samples of this function. Landau [26] showed that a necessary condition for this property is that the number of points in Λ that fall within any interval of length T is at least $f_{\mathsf{Lnd}}T$, perhaps minus a constant that is logarithmic in T. Since Landau's condition on the density of Λ implies $d^+(\Lambda) \geq f_{\mathsf{Lnd}}$, for $d^+(\Lambda) < f_{\mathsf{Lnd}}$ we necessarily have $\mathsf{mmse}(S) > 0$.

2) Unconstrained Sampling: The other lower bound in Fig. 2 describes the case when there is no loss in the sampling operation, so that the distortion is only due to lossy compression and noise. This situation occurs when the process $X|X_\epsilon(\cdot) \triangleq \{\mathbb{E}[X(t)|X_\epsilon(\cdot)], \ t \in \mathbb{R}\}$, whose spectral density is

$$S_{X|X_\epsilon}(f) = \frac{S_X^2(f)}{S_X(f) + S_\epsilon(f)}, \tag{21}$$

can be recovered from Y_T with zero MSE as $T \to \infty$. Note that $X|X_\epsilon(\cdot)$ is a Gaussian stationary process obtained by estimating $X(\cdot)$ using the non-causal Wiener filter. The resulting MSE in this estimation is

$$\mathsf{mmse}(X|X_\epsilon) \triangleq \sigma_X^2 - \int_{-\infty}^{\infty} S_{X|X_\epsilon}(f) df.$$

Since no limitation is imposed on the encoder in Fig. 4 except the bitrate, the encoder can estimate $X|X_\epsilon(\cdot)$ from Y_T and encode it at bitrate R as in standard source coding. When $T \to \infty$, the distortion in this procedure is given by [13]

$$D_{X|X_\epsilon}(R_\theta) = \mathsf{mmse}(X|X_\epsilon) + \int_{-\infty}^{\infty} \min\left\{S_{X|X_\epsilon}(f), \theta\right\} df, \tag{22a}$$

$$R_\theta = \frac{1}{2}\int_{-\infty}^{\infty} \log^+\left[S_{X|X_\epsilon}(f)/\theta\right] df. \tag{22b}$$

In the special case when $\epsilon(\cdot) \equiv 0$, (22) reduces to Pinsker's formula [16] for the DRF of $X(\cdot)$:

$$D_X(R_\theta) = \int_{-\infty}^{\infty} \min\{S_X(f), \theta\} df, \tag{23a}$$

$$R_\theta = \frac{1}{2}\int_{-\infty}^{\infty} \log^+[S_X(f)/\theta] df. \tag{23b}$$

Note that (23) is the continuous-time counterpart of (3).

When the minimal ADX distortion $D_T(S, R)$ approaches $D_{X|X_\epsilon}(R)$, or $D_X(R)$ in the non-noisy case, as $T \to \infty$, we say that the conditions for optimal sampling in ADX are met. Namely, optimal sampling occurs whenever

$$D(S, R) = D_{X|X_\epsilon}(R). \tag{24}$$

For example, (24) holds under MB-LTI sampling with a single sampling branch, provided $f_s \geq f_{\mathsf{Nyq}}$ and the passband of the pre-sampling filter $H(f)$ contains $\operatorname{supp} S_X$ (which equals to $\operatorname{supp} S_{X|X_\epsilon}$). More generally, it is possible to chose the pre-sampling filters of a MB-LTI sampler such that optimal sampling occurs for any $f_s \geq f_{\mathsf{Lnd}}$ [9, Sec. IV], [4], where f_{Lnd} is the Landau rate of $X(\cdot)$ (or its spectral occupancy). In these cases, we also have that $\mathsf{mmse}_T(S)$ of (19) converges to $\mathsf{mmse}(X|X_\epsilon)$, which is a sufficient condition for (24) to hold. As we shall see in the next section, this condition is not necessary, and optimal sampling can be attained by sampling below the Nyquist or Landau rates.

IV. The Fundamental Distortion Limit

We now provide the general definition of $D^\star(f_s, R)$, explore its basic properties, and use it to fully characterize the ADX distortion.

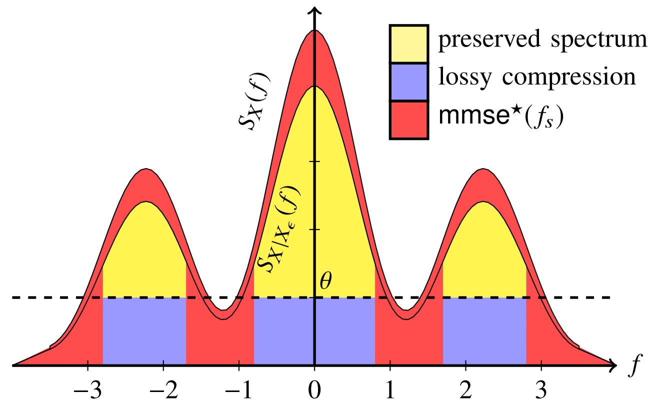

Fig. 7. Water-filling interpretation of the fundamental distortion limit $D^\star(f_s, R)$. The distortion is the sum of the sampling distortion ($\mathsf{mmse}^\star(f_s)$) and the lossy compression distortion. The set $F^\star_{f_s}$ defining $D^\star(f_s, R)$ is the support of the preserved spectrum.

A. Expression for ADX Distortion

We now define the function $D^\star(f_s, R)$, and later show that it describes the fundamental distortion in ADX.

Definition 1: For a sampling rate f_s and Gaussian signals $X(\cdot)$ and $\epsilon(\cdot)$, let $F^\star_{f_s}$ be a set of Lebesgue measure μ not exceeding f_s that maximizes

$$\int_F S_{X|X_\epsilon}(f) df = \int_F \frac{S_X^2(f)}{S_X(f) + S_\epsilon(f)} df, \tag{25}$$

over all sets F with $\mu(F) \leq f_s$. Define

$$D^\star(f_s, R_\theta) \triangleq \sigma_X^2 - \int_{F^\star_{f_s}} \left[S_{X|X_\epsilon}(f) - \theta\right]^+ (f) df, \tag{26}$$

where θ is determined by

$$R_\theta = \frac{1}{2}\int_{F^\star_{f_s}} \log^+\left[S_{X|X_\epsilon}(f)/\theta\right] df.$$

We also define the function $\mathsf{mmse}^\star(f_s)$ describing the sampling distortion in ADX as

$$\mathsf{mmse}^\star(f_s) = \sigma_X^2 - \int_{F^\star_{f_s}} S_{X|X_\epsilon}(f) df, \tag{27}$$

and note that

$$D^\star(f_s, R) = \mathsf{mmse}^\star(f_s) + \int_{F^\star_{f_s}} \min\left\{S_{X|X_\epsilon}(f), \theta\right\} df.$$

Graphical interpretations of $D^\star(f_s, R)$ and $\mathsf{mmse}^\star(f_s)$ are provided in Fig. 7. Compared to the classical water-filling formulas (23) and (22), the waterfilling in (26) is only over the part of the spectrum that is included in $F^\star_{f_s}$, whereas the complement of this part is associated with the sampling distortion $\mathsf{mmse}^\star(f_s)$.

The main results of this paper are a positive and negative coding statement in the ADX setting with respect to the function $D^\star(f_s, R)$, as per the following theorems:

Theorem 1 (Achievability): For any f_s and $\epsilon > 0$, there exists a MB-LTI sampler S with sampling rate f_s, such that, for any bitrate R, the distortion in ADX attained by sampling $X_\epsilon(\cdot)$ using S over a large enough time interval T, and encoding these samples using $\lfloor TR \rfloor$ bits, does not exceed $D^\star(f_s, R) + \epsilon$.

Theorem 2 (Converse): Let $S = (K_H, \Lambda)$ be a bounded linear sampler such that $\mathrm{card}(\Lambda_T) \le Tf_s$ for every $T > 0$. Then for any representation of the samples Y_T using at most TR bits, the MSE (9) in recovering $X(\cdot)$ is bounded from below by $D^\star(f_s, R)$. Furthermore, if $d^+(\Lambda) \le f_s$, then

$$D(S, R) \ge D^\star(f_s, R).$$

The proofs of Theorems 1 and 2 can be found in Appendices VI and VI, respectively. A sketch of these proofs is as follows. To prove Theorem 1, we use the expression for the ADX distortion under an MB-LTI sampler with L samplers derived in [9]. We then show that for any $\delta > 0$, there exists L large enough such that L filters $H_1, \ldots, H_L$ can be chosen to have disjoint supports whose union approximates $F^\star_{f_s}$ in the sense that the difference between $D(S, R)$ and $D^\star(f_s, R)$ is less than δ. The converse in Theorem 2 is first established for a MB-LTI sampler using an arbitrary number of sampling branches L. We then consider the distortion attained by a general linear bounded sampler $S = (\Lambda, K_H)$ over a finite time horizon T. We bound this distortion from below by the distortion in recovering $X(\cdot)$ over $[-T/2, T/2]$ using an encoding based on samples taken over a periodic extension of the sampling set. We then show that the samples obtained over this extension are equivalent to sampling the process using a specific MB-LTI sampler, so that the bound from the first part of the proof is valid for an arbitrary linear bounded sampler.

Before exploring additional properties of $D^\star(f_s, R)$, it is instructive to consider its behavior under various examples of the PSDs $S_X(f)$ and $S_\epsilon(f)$.

Example 2 (Rectangular PSD): Let $X_\Pi(\cdot)$ be the process with PSD

$$S_\Pi(f) = \sigma_X^2 \frac{\mathbf{1}_{|f|<W}(f)}{2W}. \tag{28}$$

Assume that $\epsilon(\cdot)$ is a flat spectrum noise within the band $[-W, W]$ such that $\gamma \triangleq S_\Pi(f)/S_\epsilon(f)$ is the SNR at the spectral component f. Under these conditions,

$$S_{X|X_\epsilon}(f) = \frac{\gamma}{1+\gamma} S_\Pi(f),$$

and the set $F^\star_{f_s}$ that maximizes (25) can be chosen as any subset of $[-W, W]$ with Lebesgue measure f_s. For simplicity we pick

$$F^\star_{f_s} = \{f : |f| < f_s/2\}, \tag{29}$$

and conclude that

$$D^\star(f_s, R) = \sigma_X^2 \begin{cases} 1 - f_s\left[\frac{\gamma}{2W(1+\gamma)}, \theta\right]^+ & f_s < 2W, \\ 1 - 2W \min\left[\frac{1}{2W(1+\gamma)}\right]^+ & f_s \ge 2W, \end{cases}$$

where θ is determined by

$$R = \frac{1}{2} \begin{cases} f_s\left(\log \frac{\sigma_X^2 \gamma}{2W(1+\gamma)} - \log\theta\right) & f_s < 2W, \\ 2W\left(\log \frac{\sigma_X^2 \gamma}{2W(1+\gamma)} - \log\theta\right) & f_s \ge 2W. \end{cases}$$

Since θ can be isolated from the last expression, we obtain

$$D^\star(f_s, R) = \sigma_X^2 \begin{cases} 1 - \frac{f_s}{2W}\frac{\gamma}{1+\gamma}\left(1 - 2^{-2R/f_s}\right) & f_s < 2W, \\ \frac{1}{1+\gamma} + \frac{\gamma}{1+\gamma} 2^{-R/W} & f_s \ge 2W. \end{cases} \tag{30}$$

We note that in the case $f_s \ge 2W$, (30) equals the DRF of $X_\Pi(\cdot)$ given $X_\Pi(\cdot) + X_\epsilon(\cdot)$ that is obtained from (22). Therefore only sampling at or above the Nyquist rate $f_{\mathsf{Nyq}} = 2W$ implies $D^\star(f_s, R) = D_{X|X_\epsilon}(R)$.

Example 3 (Triangular PSD): Let $X_\triangle(\cdot)$ be the process with PSD

$$S_\triangle(f) \triangleq \sigma_X^2 \frac{[1 - |f/W|]^+}{W}, \tag{31}$$

for some $W > 0$, and assume that $\epsilon(\cdot) \equiv 0$. Then

$$F^\star_{f_s} = \{f : |f| < f_s/2\},$$

and

$$D^\star(f_s, R) = \sigma_X^2 \begin{cases} \left(1 - \frac{f_s}{2W}\right)^2 + \theta f_s & f_s \le f_R, \\ \left(1 - \frac{f_R}{2W}\right)^2 + \theta f_R & f_s > f_R, \end{cases} \tag{32}$$

where $f_R \triangleq 2W(1 - \theta W)$, and θ is given by

$$R = \frac{1}{2} \begin{cases} \int_{-\frac{f_s}{2}}^{\frac{f_s}{2}} \left[\log S_\triangle(f) - \log\theta\right] & f_s \le f_R \\ \int_{-\frac{f_R}{2}}^{\frac{f_R}{2}} \log S_\triangle(f) - f_R \log\theta & f_s > f_R. \end{cases}$$

Note that in Example 3, the function $D^\star(f_s, R)$ in (32) is independent of f_s for the case $f_s \ge f_R$, and equals to the DRF of $X_\triangle(\cdot)$ given by Pinsker's expression (23). Consequently, for $X_\triangle(\cdot)$, the DRF is attained by sampling above f_R that is smaller than $2W$, which is the Nyquist rate of $X_\triangle(\cdot)$. Since the DRF is the minimal distortion subject only to the bitrate constraint regardless of the sampling mechanism, we conclude that the optimal distortion performance is attained by sampling below the Nyquist rate in this case. In the following subsection, we extend this observation to arbitrary PSDs.

B. Optimal Sampling Rate

We now consider the minimal sampling rate that leads to optimal sampling in ADX. We first note the following proposition, that follows from the definition of $D^\star(f_s, R)$.

Proposition 3 (Optimal Sampling Rate): For each point (R, D) on the graph of $D_{X|X_\epsilon}(R)$ associated with a water-level θ via (22), define

$$F_\theta \triangleq \left\{f : S_{X|X_\epsilon}(f) > \theta\right\},$$

and set $f_R = \mu(F_\theta)$. Then for all $f_s \ge f_R$,

$$D^\star(f_s, R) = D_{X|X_\epsilon}(R).$$

The proof of Proposition 3 is given in Appendix C. To gain some intuition into the results, consider the special case of zero noise and a unimodal $S_X(f)$ as illustrated in Fig. 8: fix a point (R, D) on the distortion rate curve of $X(\cdot)$ obtained from (23). The set $F_\theta = \{f \in \mathbb{R} : S_X(f) > \theta\}$ is the support of the

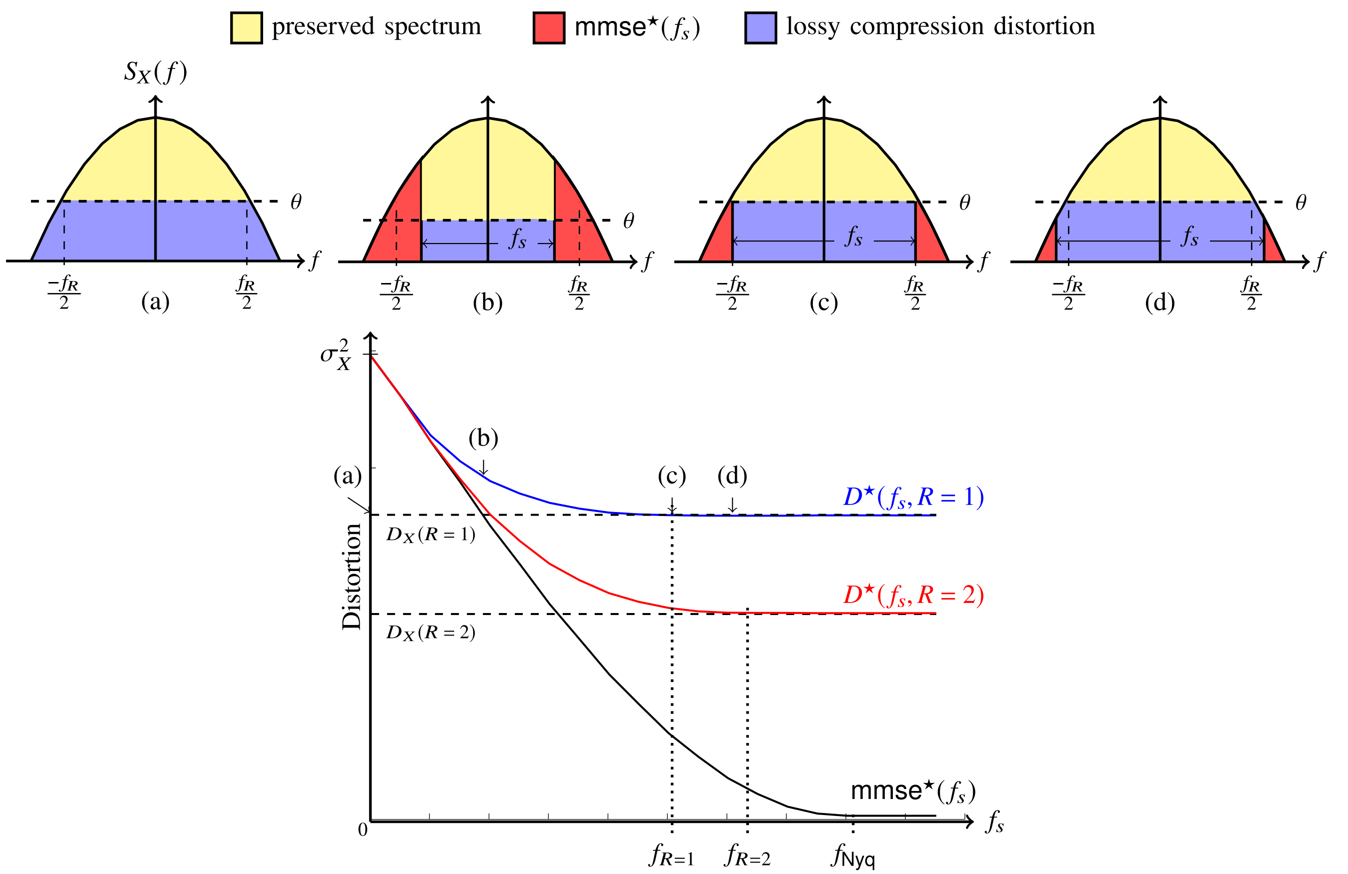

Fig. 8. An illustration of Proposition 3 for a unimodal PSD and zero noise: the distortion is the sum of the term $\mathsf{mmse}^\star(f_s)$ (red) and the lossy compression distortion (blue) in each water-filling scheme. The functions $D^\star(f_s, R)$ and $\mathsf{mmse}^\star(f_s)$ are illustrated versus f_s at the bottom for two fixed values of the bitrate R. Also shown is the DRF of $X(\cdot)$ at these values that is attained at the sub-Nyquist sampling rates marked by f_R. The marked points on the curve of $D^\star(f_s, R)$ with $R = 1$ correspond to the different water-filling scenarios (b)–(d).

non-shaded area in Fig. 8(a). We define the sampling rate f_R to be the Lebesgue measure of F_θ. Figure 8(b) shows the function $D^\star(f_s, R)$ for $f_s < f_R$, where the overall distortion is the sum of the term $\mathsf{mmse}^\star(f_s)$ given by the partially shaded area, and the water-filling term given by the blue area. Figures 8 (c) and (d) show the function $D^\star(f_s, R)$ for $f_s = f_R$ and $f_s > f_R$, respectively. The assertion of Proposition 3 is that the sum of the red area and the blue area stays the same for any $f_s \geq f_R$. It can also be seen from Fig. 8 that f_R increases with the source coding rate R and coincides with f_{Nyq} as $R \to \infty$. The bottom-right of Fig. 8 shows $D^\star(f_s, R)$ as a function of f_s for two fixed values of R.

We emphasize that the critical frequency f_R arising from Proposition 3 depends only on the PSD and on the operating point on the DRF curve of $X(\cdot)$ given $X_\epsilon(\cdot)$, which can be parametrized by either D, R or the water-level θ using (22). In fact, by inverting the function $D^\star(f_s, R)$ with respect to R, we obtain the following result.

Theorem 4 (Rate-Distortion Lower Bound): Given Gaussian stationary processes $X(\cdot)$ and $\epsilon(\cdot)$, sampling rate f_s and a target distortion $D > \mathsf{mmse}^\star(f_s)$, define

$$R^\star(f_s, D) \triangleq \begin{cases} \frac{1}{2}\int_{F^\star_{f_s}} \log^+\left(\frac{f_s S_{X|X_\epsilon}(f)}{D - \mathsf{mmse}^\star(f_s)}\right) df, & f_s < f_R, \\ R_{X|X_\epsilon}(D), & f_s \geq f_R, \end{cases} \tag{33}$$

where

$$R_{X|X_\epsilon}(D_\theta) = \frac{1}{2}\int_{-\infty}^{\infty} \log^+\left[S_{X|X_\epsilon}(f)/\theta\right] df,$$

is the indirect rate-distortion function of $X(\cdot)$ given $X_\epsilon(\cdot)$, $f_R = \mu\left(\left\{f : S_{X|X_\epsilon}(f) > \theta\right\}\right)$, and θ is determined by

$$D = \mathsf{mmse}(X|X_\epsilon) + \int_{-\infty}^{\infty} \min\left\{S_{X|X_\epsilon}(f), \theta\right\} df.$$

Then:

(i) The number of bits per unit time required to attain ADX distortion at most D is at least $R^\star(f_s, D)$.

(ii) For any $\epsilon > 0$ and $\rho > 0$, there exists T large enough and a MB-LTI sampler S at rate f_s such that

$$D_T(S, R^\star(f_s, D) + \rho) < D + \epsilon.$$

Proof: Theorem 4 is a restatement of Theorems 2 and 1 that is obtained using Proposition 3 and by inverting the role of the distortion and the bitrate. □

C. Discussion

Theorem 1 together with Proposition 3 extend the conditions for the equality (24), which, as argued in Subsection III-D, holds for $f_s \geq f_{\mathsf{Lnd}}$, to all sampling rates above f_R. This f_R depends on the bitrate R and is smaller than f_{Lnd} provided the signal power is not uniformly distributed over its spectral support (unlike $S_\Pi(f)$ of Example 2).

As R goes to infinity, the water-level θ goes to zero, the set F_θ coincides with the support of $S_{X|X_\epsilon}(f)$ and $S_X(f)$, and $D^\star(f_s, R)$ converges to $\mathsf{mmse}^\star(f_s)$. In particular, $f_R = \mu(F_\theta)$ converges to the spectral occupancy f_{Lnd} of $X(\cdot)$. In this limit, Proposition 3 then implies that $\mathsf{mmse}^\star(f_s) = \mathsf{mmse}(X|X_\epsilon)$ for all $f_s \geq f_{\mathsf{Lnd}}$. When the noise is zero, this last fact agrees with the Landau's necessary condition for stable sampling in the Paley-Wiener space [26].

The discussion in Section II on the finite-dimensional counterpart of ADX suggests the following intuition for our result assuming $\epsilon(\cdot) \equiv 0$: Pinsker's water-filling expression (23) implies that for a Gaussian stationary signal whose power is not uniformly distributed over its spectral occupancy, the distortion-rate function $D_X(R)$ is achieved by communicating only those bands with the highest energy. This means that fewer degrees of freedom are used in the signal's representation. Proposition 3 says that this reduction in degrees of freedom can be translated to a lower required sampling rate in order to achieve $D_X(R)$. The counterpart of this phenomena in the finite dimensional case is the condition for equality between (3) and (5) as discussed in Section II.

D. Examples

In the following examples the exact dependency of f_R on R and D is determined for various PSDs, and illustrated in Fig. 9.

Example 4 (Continuation of Example 2): Consider the PSDs $S_\Pi(f)$ and $S_\epsilon(f) = S_\Pi(f)/\gamma$ as in Example 2. In this case we have that $F_\theta = [-W, W]$ for all f_s, so that $f_R = 2W$. Therefore, in this example, $D^\star(f_s, R) = D_{X|X_\epsilon}(R)$ only for f_s larger than the Nyquist rate of $X_\Pi(f)$. This observation agrees with the expression for $D^\star(f_s, R)$ in (30).

Example 5 (Continuation of Example 3): Consider the situation of Example 3 with zero noise and PSD $S_\triangle(f)$. For a point $(R, D) \in [0, \infty) \times [0, \sigma_X^2]$ on the distortion-rate curve of $X_\triangle(\cdot)$, we have that $F_\theta = W[-1 + W\theta, 1 - W\theta]$ and hence $f_R = 2W(1 - W\theta)$. Indeed, this value for f_R agrees with (32), since for $f_s \geq f_R$ the function $D^\star(f_s, R)$ is independent of f_s and equals the DRF of $X_\triangle(\cdot)$.

The exact relation between R and f_R is given by

$$R = \frac{1}{2}\int_{-\frac{f_R}{2}}^{\frac{f_R}{2}} \log\left(\frac{1 - |f/W|}{1 - \frac{f_R}{2W}}\right) df = W \log \frac{1}{1 - \frac{f_R}{2\ln 2}} - \frac{f_R}{2W}. \tag{34}$$

Expressing f_R as a function of the distortion D leads to $f_R = 2W\sqrt{1 - D/\sigma_X^2}$.

Example 6 (Effect of Noise on f_R): Consider again $X_\triangle(\cdot)$ from Examples 3 and 5, but with $\epsilon(\cdot)$ a flat spectrum Gaussian noise with intensity σ_ϵ^2, i.e., $S_\epsilon(f) = \sigma_\epsilon^2 \mathbf{1}_{[-W,W]}$. The relation

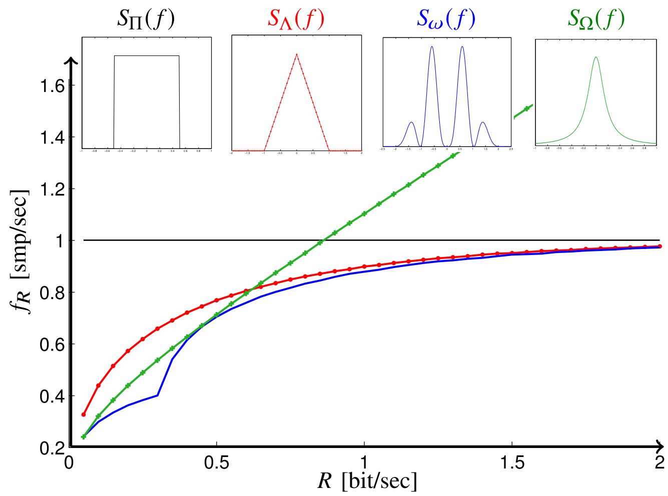

Fig. 9. The critical sampling rate f_R as a function of the bitrate R for the PSDs given in the small frames at the top of the figure and zero noise ($\epsilon(\cdot) \equiv 0$). Here $S_\Pi(f)$, $S_\triangle(f)$ and $S_\omega(f)$ have the same bandwidth while the support of $S_\Omega(f)$ is unbounded.

between R and f_R is given by:

$$\begin{aligned} R &= \int_{-\frac{f_R}{2}}^{\frac{f_R}{2}} \log\left[\frac{(1-\frac{f}{W})^2}{1 - \frac{f}{W} + W\sigma_\epsilon^2}\right] df \\ &\quad - f_R \log\left[\frac{(1-\frac{f_R}{2W})^2}{1 - \frac{f_R}{2\,W} + W\sigma_\epsilon^2}\right] \\ &= 2W \log \frac{1}{1 - \frac{f_R}{2W}} - W(1 + \sigma_\epsilon^2 W) \log \frac{1}{1 - \frac{f_R}{2W(1+\sigma_\epsilon^2 W)}} \\ &\quad - \frac{f_R}{2\ln 2}. \end{aligned}$$

The expression above decreases as the intensity of the noise σ_ϵ^2 increases. Since f_R increases with R, it follows that f_R decreases in σ_ϵ^2, as can be seen in Fig. 10 where f_R is plotted versus the SNR $\sigma_X^2/\sigma_\epsilon^2$ for two fixed values of R. The dependency between the critical sampling rate f_R and the SNR observed in Example 6 can be generalized to any signal PSD experiencing a uniform increase in the SNR: increase in SNR decreases $\mathsf{mmse}(X|X_\epsilon)$ and leads to the use of more spectral bands in the indirect source coding scheme that attains $D_{X|X_\epsilon}(R)$. As a result, more spectral bands of $S_{X|X_\epsilon}(f)$ must be utilized in order for $D^\star(f_s, R)$ to approach $D_{X|X_\epsilon}(R)$.

Next, we explore applications of the fundamental ADX distortion limit in the sampling and lossy compression of signals that are not bandlimited.

E. Sampling Non-Bandlimited Signals

Let $X(\cdot)$ be a stationary Gaussian signal that is not bandlimited in the sense that its Landau rate is infinite. For simplicity of discussion, we assume in this subsection that the noise $\epsilon(\cdot)$ is zero, hence the distortion is only due to sampling and lossy compression. Even under the zero noise assumption, it follows from [26] that it is impossible to recover such a signal with zero MSE from its samples obtained over any discrete set. Nevertheless, it follows from Proposition 3 that

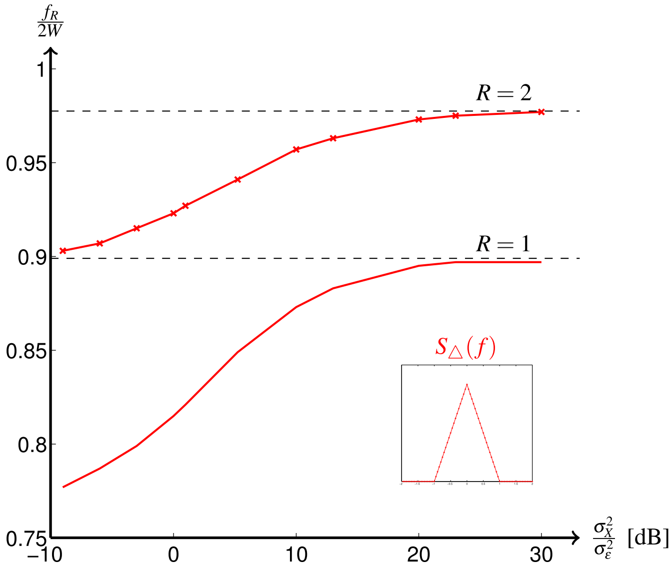

Fig. 10. The critical rate f_R as a function of the SNR $\sigma_X^2/\sigma_\epsilon^2$ for an input signal with PSD $S_\triangle(f)$ corrupted by a flat spectrum Gaussian noise for two fixed values of R. The dashed lines correspond to the value of f_R without noise.

for a point (R, D) on the DRF $D_X(R)$ of $X(\cdot)$, there exists a critical sampling rate f_R such that $D^\star(f_R, R) = D_X(R)$. In other words, the minimal distortion subject to the bitrate constrain R is attainable at a finite sampling rate f_R. Note, however, that unlike in the bandlimited case, f_R goes to infinity as R goes to infinity, hence in order to represent a signal that is not bandlimited in bits with vanishing distortion, both f_s and R must go to infinity.

For a bandlimited signal the ratio R/f_R is unbounded as R goes to infinity. This ratio is the maximal average number of bits per sample in an optimal digital representation of the signal from its samples with bitrate R. Consequently, the challenge in obtaining an optimal digital representation of a bandlimited signal with vanishing distortion is the design of a high-resolution quantizer to represent each sample obtained at the Landau rate. In contrast, for some non-bandlimited signals whose spectrum vanishes slow enough, the ratio R/f_R converges to a constant. As a result, the challenge in encoding such signals is incorporating information across a large number of samples and represent this information using a limited number of bits.

An example for a non-bandlimited signal, its critical sampling rate f_R, and the asymptotic behavior of R/f_R is as follows:

Example 7 (Gauss-Markov Process): Assume zero noise ($\epsilon(\cdot) \equiv 0$) and consider the Gauss-Markov process $X_\Omega(\cdot)$ whose PSD is

$$S_\Omega(f) = \frac{1/f_0}{(\pi f/f_0)^2 + 1}, \tag{35}$$

for some $f_0 > 0$. Note that the MMSE in recovering $X_\Omega(\cdot)$ from its uniform samples at rate f_s equals the area bounded by the tails of its PSD:

$$\mathsf{mmse}^\star(f_s) = 2\int_{f_s/2}^{\infty} S_\Omega(f)df = 1 - \frac{2}{\pi}\arctan\left(\frac{\pi f_s}{2f_0}\right). \tag{36}$$

For a point (R, D) on the distortion-rate curve of $X_\Omega(\cdot)$ and its corresponding θ, we have $f_R = \frac{2f_0}{\pi}\sqrt{\frac{1}{\theta f_0} - 1}$. Namely, sampling at this rate allows the encoding of $X_\Omega(\cdot)$ with minimal reconstruction distortion subject to the bitrate constraint R. Consequently, the distortion cannot be further reduced by sampling above this rate. The exact relation between R and f_R is given by

$$R = \frac{1}{\ln 2}\left(f_R - 2f_0\frac{\arctan\left(\frac{\pi f_R}{2f_0}\right)}{\pi}\right), \tag{37}$$

and is illustrated in Fig. 9. The high bitrate asymptotic of (37) implies $R/f_R \to 1/\ln 2$. Therefore, for R sufficiently large, $1/\ln 2 \approx 1.44$ is the maximal number of bits that can be used in encoding each sample of $X_\Omega(\cdot)$ in order to attain its DRF. If the number of bits per sample goes above this value, then the distortion is dominated by the sampling distortion as not enough samples are acquired for each bit in the digital representation. More generally, for any fixed number of bits per sample $\bar{R} = R/f_s$, the excess distortion due to sampling in the limit of high bitrate is given by

$$\lim_{R\to\infty}\frac{D^\star(R/\bar{R}, R)}{D_{X_\Omega}(R)} = \begin{cases} \frac{1+2^{-2(\bar{R}-1/\ln 2)}}{2}\bar{R}\ln 2 & \bar{R} > 1/\ln 2, \\ 1 & \bar{R} \le 1/\ln 2. \end{cases} \tag{38}$$

Note that a limit of the form (38) equals one for any bandlimited signal and for signals whose spectrum vanish fast enough, e.g., $S_X(f) = e^{-|f|}$.

V. Pulse-Code Modulation

So far we considered the conversion of analog signals to bits using bounded linear sampling and under optimal encoding of these samples to bits, subject only to the bitrate constraint R. In particular, we did not impose any limitations on the complexity or delay of the encoder and decoder aside from the bitrate at the encoder's output. Indeed, the achievablity of $D^\star(f_s, R)$ in Theorem 1 is obtained as the time horizon T grows to infinity, whereas the number of states assumed by the encoder and decoder grows exponentially in T.

In this section we are interested in imposing additional constraints on the restricted-bitrate representation of the samples and the recovery of $X(\cdot)$ beyond those associated with the achievable scheme of Theorem 1. Specifically, we now assume that the samples are obtained using a single sampling branch, the encoder maps each sample Y_n to its finite-bit representation $\hat{Y}_n$ at time n using a scalar quantizer with a fixed number of bits q, and the decoder recovers $X(\cdot)$ using a linear procedure. This form of encoding is known as *pulse-code modulation* (PCM) [50], [51]; we refer to [52, Sec I.A] for a historical overview. In order to focus on the effect of this sub-optimal encoding and decoding on the distortion-rate performance, we assume that no noise is added to $X(\cdot)$ prior to sampling. The extension of the distortion analysis below to the case in which such a noise is present is straightforward.

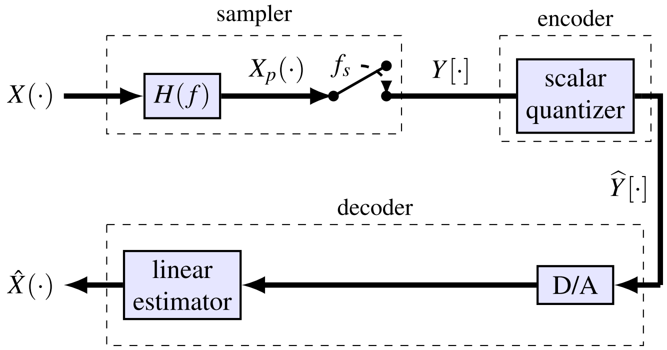

Fig. 11. Pulse-code modulation and reconstruction system.

A. PCM A/D Conversion and Reconstruction Setup

We consider the system described in Fig. 11, where the input $X(\cdot)$ is assumed to be a wide-sense stationary stochastic process with PSD $S_X(f)$, not necessarily Gaussian. This process is sampled using a pre-sampling filter $H(f)$ followed by a uniform sampler with sampling rate f_s. This is a special case of the multi-branch LTI uniform sampler of Fig. 6 with $L = 1$ and $H_1(f) = H(f)$. The sample $Y[n]$ at time n/f_s is mapped to a quantization level $\hat{Y}[n]$ using a procedure denoted as *fixed-length scalar quantization* [52]: we consider a set of M real numbers $\hat{y}_1, \ldots, \hat{y}_M$ called *reconstruction levels*. Each reconstruction level is assigned a digital number of length $q \triangleq \lceil \log M \rceil$, where q is the *bit resolution* of the quantizer. Upon receiving the input $Y[n]$, the quantizer outputs the nearest reconstruction level to $Y[n]$ among the set of reconstruction levels, what we denote $\hat{Y}[n]$. Using this notation, the *bitrate* of the digital representation, namely, the number of bits per unit time required to represent the process $\hat{Y}[\cdot]$, is $R = qf_s$.

Denote by $\eta[n]$ the quantization error, i.e.,

$$\hat{Y}[n] = Y[n] + \eta[n], \quad n \in \mathbb{Z}. \tag{39}$$

The variance of $\eta[n]$ depends on the square of the size of the quantization regions induced by the quantizer, i.e., the Voronoi sets associated with the reconstruction levels. The number of these sets increases exponentially in the bit resolution q and so does the radius of each set, provided all radii decrease uniformly [53]. As a result, the variance of $\eta[n]$ behaves as

$$\sigma_\eta^2 = c_0 2^{-2q}, \tag{40}$$

for some $c_0 > 0$. The constant c_0 depends on other statistical assumptions on the input to the quantizer. For example, if the amplitude of the input signal is bounded within the interval $(-A_m/2, A_m/2)$, then we may choose uniformly spaced quantization levels resulting in $c_0 = A_m/12$. If the input to the quantizer is Gaussian with variance σ_{in}^2 and the quantization rule is chosen according to the ideal point density allocation of the Lloyd algorithm [54], then [52, eq. (10)]

$$c_0 = \frac{\pi\sqrt{3}}{2}\sigma_{in}^2. \tag{41}$$

The non-linear relation between $Y[n]$ and $\hat{Y}[n]$ complicates the analysis. To simplify the problem, we adopt a common assumption in the signal processing literature (e.g. [55], [56]):

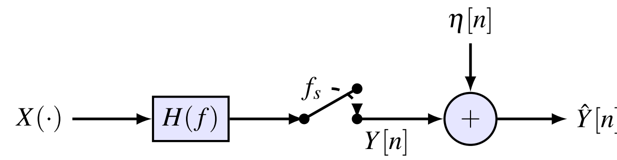

Fig. 12. Sampling and quantization system model.

(A) The process $\eta[\cdot]$ is zero mean, white (uncorrelated entries), and is uncorrelated with $Y[\cdot]$.

There exists a vast literature on conditions under which assumption (A) provides a good approximation to the system behavior. For example, in [53] it was shown that two consecutive samples $\eta[n]$ and $\eta[n+1]$ are approximately uncorrelated if the distribution of $Y[\cdot]$ is smooth enough, where this holds even if the sizes of the quantization regions are on the order of the variance of $Y[\cdot]$ [57]. This property justifies the assumption that the process $\eta[\cdot]$ is white. Bennett [58] showed that $\eta[\cdot]$ and $Y[\cdot]$ are approximately uncorrelated provided the PSD of $Y[\cdot]$ is smooth, the quantization regions are uniform and the quantizer resolution q is high. Since in our setting the quantizer resolution may also be relatively low when f_s approaches R, our analysis under (A) does not lead to an exact description of the performance limit under scalar quantization. Nevertheless, under (A) the distortion due to quantization decreases exponentially as a function of the quantizer bit precision and is proportional to the variance of the input signal. These two properties, which hold also under an exact analysis of the error due to scalar quantization with entropy coding [52], are the dominant factors in the MMSE analysis below.

B. Distortion Analysis

Under (A), the relation between the input and the output of the quantizer can be represented in the z domain by

$$\hat{Y}(z) = Y(z) + \eta(z). \tag{42}$$

This leads to the following relation between the corresponding PSDs:

$$\begin{aligned} S_{\hat{Y}}\left(e^{2\pi i\phi}\right) &= S_Y\left(e^{2\pi i\phi}\right) + S_\eta\left(e^{2\pi i\phi}\right) \\ &= f_s \sum_{k\in\mathbb{Z}} S_X(f - f_s k)\,|H(f - f_s k)|^2 + \sigma_\eta^2. \end{aligned} \tag{43}$$

The block diagram of a generic system that realizes the input-output relation (42) is given in Fig. 12, where, in accordance with (A), $\eta[\cdot]$ is a white noise independent of $X(\cdot)$. In what follows, we derive an expression for the linear MMSE in estimating $X(\cdot)$ from $\hat{Y}[\cdot]$ according to the relation (43) and an optimal choice of the pre-sampling filter $H(f)$, that minimizes this MSE.

The goal of the linear decoder is to provide a reconstruction signal $\hat{X}(\cdot)$ that minimizes

$$\lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} \mathbb{E}\left(X(t) - \hat{X}(t)\right)^2 \tag{44}$$

over all possible reconstruction signals of the form

$$\hat{X}(t) = \sum_{n\in\mathbb{Z}} w(t,n)\hat{Y}[n], \tag{45}$$

where $w(t,n)$ is square summable in n for every $t \in \mathbb{R}$. Note that this decoder is non-causal in the sense that the estimate of the source sample $X(t)$ is obtained from the entire history of the quantized signal $\hat{Y}[\cdot]$. Since all signals in Fig. 12 are assumed stationary, an expression for the minimal value of (44) subject to the constraint (45) can be found using standard linear estimation techniques, leading to the following proposition:

Proposition 5: Consider the system in Fig. 12. The minimal time-averaged MSE (44) in linear estimation of $X(\cdot)$ from $\hat{Y}[\cdot]$ is given by

$$D_{\mathsf{PCM}} \triangleq= \sigma_X^2 - \frac{1}{f_s}\int_{-\frac{f_s}{2}}^{\frac{f_s}{2}} \frac{\sum_{k\in\mathbb{Z}} S_X^2(f-f_sk)\left|H(f-f_sk)\right|^2}{\sum_{k\in\mathbb{Z}} S_X(f-f_sk)\left|H(f-f_sk)\right|^2 + \sigma_\eta^2/f_s} df. \tag{46}$$

Proof: See Appendix D. □

The effect of quantization noise is expressed in (46) by an additive noise with a constant PSD over the digital domain.

Using Hölder's inequality and monotonicity of the function $x \to \frac{x}{x+1}$, the integrand in (46) can be bounded for each f in the integration interval $(-f_s/2, f_s/2)$ by

$$\frac{(S^\star(f))^2}{S^\star(f)+\sigma_\eta^2/f_s}, \tag{47}$$

where

$$S^\star(f) = \sup_{k\in\mathbb{Z}} S_X(f-f_sk)\left|H(f-f_sk)\right|^2. \tag{48}$$

Since $S_X(f)$ is an L_1 function, the supremum in (48) is finite for all $f \in (-f_s/2, f_s/2)$ except for perhaps a set of Lebesgue measure zero. It follows that a lower bound on D_{PCM} is obtained by replacing the integrand in (46) with $S^\star(f)$.

Under the assumption that $S_X(f)$ is unimodal in the sense that it is symmetric and non-increasing for $f > 0$, for each $f \in [-f_s/2, f_s/2]$ the supremum in (48) is obtained for $k = 0$. This implies that (47) is achievable if the pre-sampling filter is a low-pass filter with cut-off frequency $f_s/2$, namely

$$H(f) = \begin{cases} 1, & |f| \le f_s/2, \\ 0, & \text{otherwise}. \end{cases} \tag{49}$$

This choice of $H(f)$ in (46) leads to

$$D_{\mathsf{PCM}} = \mathsf{mmse}^\star(f_s) + \int_{-\frac{f_s}{2}}^{\frac{f_s}{2}} \frac{S_X(f)}{1+\mathsf{snr}(f)} df, \tag{50}$$

where $\mathsf{mmse}^\star(f_s)$ is given by (27) and

$$\mathsf{snr}(f) \triangleq f_s S_X(f)/\sigma_\eta^2, \quad -\frac{f_s}{2} \le f \le \frac{f_s}{2}. \tag{51}$$

Henceforth, we will consider only processes with unimodal PSD, so that the MMSE under optimal pre-sampling filtering is given by (50).

C. PCM Distortion Under a Fixed Bitrate

From (51) we see that when the variance of the quantization noise is independent of f_s, than SNR in the system in Fig. 12 increases linearly in f_s. The MMSE of $X(\cdot)$ given $\hat{Y}[\cdot]$ then decreases by a factor of $1/f_s$ when f_s is large. However, when the bitrate $R = qf_s$ is fixed, the relation between σ_η^2 and f_s is given by

$$\sigma_\eta^2 = c_0 2^{-2q} = c_0 2^{-2R/f_s}. \tag{52}$$

Substituting (52) into (50) and (52), we obtain the following proposition:

Proposition 6: The MMSE in estimating $X(\cdot)$ from $\hat{Y}[\cdot]$ assuming (A) and $R = qf_s$ satisfies

$$D_{\mathsf{PCM}}(f_s, R) = \mathsf{mmse}^\star(f_s) + \int_{-\frac{f_s}{2}}^{\frac{f_s}{2}} \frac{S_X(f)}{1+\mathsf{snr}(f)} df \tag{53}$$

where

$$\mathsf{snr}(f) = \mathsf{snr}_{f_s,R}(f) = f_s\frac{2^{2R/f_s}}{c_0}S_X(f) \tag{54}$$

and $\mathsf{mmse}^\star(f_s)$ is given by (27).

We denote the two terms in the RHS of (53) as the *sampling distortion* and the *quantization distortion*, respectively. Note that when $R \to \infty$ the quantization error vanishes and the distortion in PCM is only due to sampling. Since we assumed unimodal PSD, the sampling distortion vanishes only for $f_s \ge f_{\mathsf{Nyq}}$.

Figure 14 shows $D_{\mathsf{PCM}}(f_s, R)$ as a function of f_s for a given R and various PSDs compared to their corresponding optimal ADX distortions $D^\star(f_s, R)$ of (26). In Fig. 14 and in other figures throughout, we take c_0 as in (41) which corresponds to an optimal point density of the Gaussian distribution whose variance is proportional to the signal at the input to the quantizer. The variance of the latter is given by

$$\sigma_{\mathsf{in}}^2 = \int_{-\infty}^{\infty} S_X(f)\left|H(f)\right|^2 df = \sigma_X^2 - \mathsf{mmse}^\star(f_s).$$

While σ_{in}^2 depends on the sampling rate f_s, it can be shown to have a negligible effect on $D_{\mathsf{PCM}}(f_s, R)$ for sampling rates close to f_{Nyq}, which is our main area of interest. We therefore ignore this dependency and continue our discussion assuming $\sigma_{\mathsf{in}}^2 = \sigma_X^2$.

D. The Optimal Sampling Rate

The quantization error in (53) is an increasing function of f_s (mainly due to the decrease in the exponent, but also due to the increase in σ_{in}^2), whereas the sampling error $\mathsf{mmse}_X^\star(f_s)$ decreases in f_s. This situation is illustrated in Fig. 13. The sampling rate $f_s^\star$ that minimizes $D_{\mathsf{PCM}}(f_s, R)$ is obtained at an equilibrium point where the derivatives of both terms are of equal magnitudes. Figure 14 shows that $f_s^\star$ depends on the particular shape of the input signal's PSD. If the signal is bandlimited, then we have the following result.

Corollary 7: For a bandlimited $X(\cdot)$, $f_s^\star$ that minimizes $D_{\mathsf{PCM}}(f_s, R)$ is at or below the Nyquist rate.

Proof: Since $\mathsf{snr}_{f_s,R}(f)$ is an increasing function of f_s in the interval $0 \le f_s \le R$, and since $\mathsf{mmse}^\star(f_s) = 0$ for

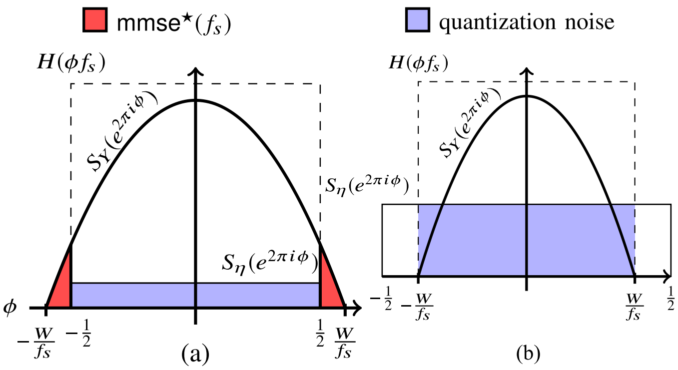

Fig. 13. Spectral representation of the distortion in PCM (53): (a) sampling below the Nyquist rate introduces sampling distortion $\mathsf{mmse}^\star(f_s)$. (b) As f_s increases, $\mathsf{mmse}^\star(f_s)$ decreases and vanishes when $f_s \geq f_{\mathsf{Nyq}}$, but the contribution of the in-band quantization noise increases due to lower bit-precision of each sample.

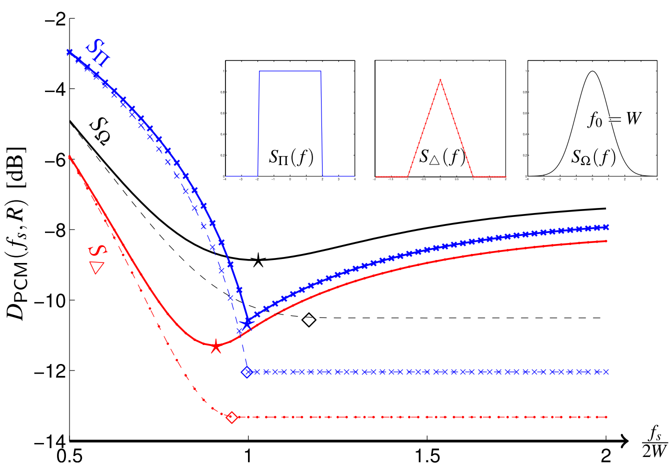

Fig. 14. PCM Distortion $D_{\mathsf{PCM}}(f_s, R)$ as a function of f_s for a fixed R and various PSDs, which are given in the small frames. The dashed curves are the corresponding minimal ADX distortions $D(f_s, R)$. The symbols $\star$ and $\diamond$ indicate the distortion at rates $f_s^\star$ and f_R, respectively.

$f_s \geq f_{\mathsf{Nyq}}$, for all $f_s > f_{\mathsf{Nyq}}$ we have that $D_{\mathsf{PCM}}(f_{\mathsf{Nyq}}, R) \leq D_{\mathsf{PCM}}(f_s, R)$. Therefore, the minimizing sampling rate cannot be greater than f_{Nyq}. □

How far $f_s^\star$ is below f_{Nyq} is determined by the derivative of $\mathsf{mmse}^\star(f_s)$, which equals $-2S_X(f_s/2)$. For example, in the case of $S_\Pi(f)$ of Examples 2 and 4, the derivative of $-2S_X(f_s/2)$ for $f_s < f_{\mathsf{Nyq}} = 2W$ is $-\sigma_X^2$. The derivative of the second term in (53) is smaller than σ_X^2 for most choices of system parameters.[1] It follows that 0 is in the sub-gradient of $D_{\mathsf{PCM}}(f_s, R)$ at $f_s = 2W$, and thus $f_s^\star = 2W$, i.e., Nyquist rate sampling is optimal when the energy of the signal is uniformly distributed over its bandwidth. We now consider the other PSDs illustrated in Fig. 14.

Example 8 (Triangular PSD): Let $S_\triangle(f)$ be the PSD of Examples 3 and 5. For any $f_s \leq f_{\mathsf{Nyq}} = 2W$, we have

$$\mathsf{mmse}^\star(f_s) = \sigma_X^2 \left(1 - \frac{f_s}{2W}\right)^2.$$

[1]This holds whenever $\left(2^{0.5R/W} - 1\right)^2 > \frac{c_0}{\sigma_X^2}$.

Since the derivative of $\mathsf{mmse}^\star(f_s)$, which is $-2S_\triangle(f_s/2)$, changes continuously from 0 to $-2\sigma_X^2/W$ as f_s varies from $2W$ to 0, we have $0 < f_s^\star < 2W$. The exact value of $f_s^\star$ depends on R and the ratio σ_X^2/c_0. It converges to $2W$ as the value of any of these two increases.

Example 9 (PSD of Unbounded Support): Consider the PSD $S_\Omega(f)$ of the Gauss-Markov process $X_\Omega(\cdot)$ in Example 7. Since $X_\Omega(f)$ is not bandlimited, Corollary 7 does not hold. Nevertheless, as can be seen in Fig. 14, there exists an optimal sampling rate $f_s^\star$ that balances the two trends as explained in Subsection V-D.

E. Discussion

Under a fixed bitrate constraint, oversampling no longer reduces the MMSE since increasing the sampling rate forces a reduction in the quantizer resolution and increases the magnitude of the quantization noise. As illustrated in Fig. 13, for any f_s below the Nyquist rate the bandwidths of both the signal and the quantization noise occupy the entire digital frequency domain, whereas the magnitude of the noise decreases as more bits are used in quantizing each sample.

It follows that $f_s^\star$ cannot be larger than the Nyquist rate (Corollary 7), and is strictly smaller than Nyquist when the energy of $X(\cdot)$ is not uniformly distributed over its bandwidth. In this case, some distortion due to sampling is preferred in order to increase the quantizer resolution. In other words, restricted to scalar quantization, the optimal rate R code is achieved by sub-Nyquist sampling. This behavior of $D_{PCM}(f_s, R)$ is similar to the behavior of the minimal ADX distortion $D^\star(f_s, R)$, as both provide an optimal sampling rate which balances sampling distortion and lossy compression distortion. On the other hand, oversampling introduces redundancy into the PCM representation, and yields a worse distortion-rate code than with $f_s = f_s^\star$. In this aspect the behavior of $D_{PCM}(f_s, R)$ is different than $D^\star(f_s, R)$ that represents the information theoretic bound, since the latter does not penalize oversampling as the optimal ADX encoder has the freedom to discard redundant samples when needed.

The similarity between $f_s^\star$ and f_R as a function of R is due to the fact that the optimal representation is obtained by discarding the same part of the signal under both the optimal lossy compression scheme or PCM. The observation that $f_s^\star \leq f_R$ in Examples 5 and 9 is explained by the diminishing effect of reducing the sampling rate on the overall error. That is, since $D_{\mathsf{PCM}}(f_s^\star, R) \geq D_X(R)$, the optimal lossy compression scheme is more sensitive to changes in the sampling rate than the sub-optimal implementation of A/D conversion via PCM.

VI. Conclusions

We considered an analog-to-digital compression (ADX) setting in which an analog signal is converted to bits by compressing its samples at a finite bitrate R, where these samples are obtained by any continuous linear sampling technique. We have shown that for any Gaussian stationary signal and any given bitrate R, there exists a critical sampling rate denoted f_R, such that the minimal distortion subject only to the bitrate constraint and the noise can be achieved by sampling at or above f_R. In particular, under nonuniform sampling, the minimal distortion subject to a bitrate constraint is attained only if the density of the sampling set exceeds f_R. The critical sampling rate f_R is strictly smaller than the Nyquist or Landau rates for processes whose power is not uniformly distributed over their spectral support. As the bitrate R increases, f_R increases as well and converges to the Nyquist or Landau rates as R goes to infinity. Furthermore, our results imply that such a finite critical sampling rate exists even for non-bandlimited signals, hence such signals can be converted to bits by sampling them at rate f_R without additional loss in information due to sampling.

The results in this paper also imply that with an optimized linear sampling technique, sampling below the Nyquist rate and above f_R does not degrade performance in the case where lossy compression of the samples is introduced. Since lossy compression due to quantization is an inherent part of any analog to digital conversion scheme, our work suggests that sampling below the Nyquist rate is optimal in terms of minimizing distortion in practice for most systems.

We also considered the case of a more restricted encoder and decoder which corresponds to pulse-code modulation (PCM) sampling and quantization. That is, instead of a vector quantizer whose block-length goes to infinity, PCM uses a zero-memory zero-delay quantizer. Under a fixed bitrate at the output of this quantizer, there exists a trade-off between bit-precision and sampling rate. We examined the behavior of this trade-off under an approximation on the scalar quantizer using additive white noise. We have shown through various examples that the optimal sampling rate in PCM experiences a similar behavior as the critical rate f_R, which is the minimal sampling rate under optimal source encoding-decoding of the samples.

There are a few important future research directions that arise from this work. While we restricted ourselves to bounded linear samplers, it is important to understand whether the distortion at a given sampling rate can be improved by considering non-linear sampling techniques. Indeed, such improvement is seen in the setting of [41], where a finite dimensional sampling system with a Gaussian input is considered. In addition, reduction of the optimal sampling rate under the bitrate constraint from the Nyquist rate to f_R can be understood as the result of a reduction in degrees of freedom in the compressed signal representation compared to the original source. It is interesting to understand whether a similar principle holds under non-Gaussian signal models (e.g., sparse signals), so that the sampling rate under a bitrate restriction can be reduced without incurring additional distortion. Finally, under sub-optimal encoding such as in PCM, it is important to characterize the conditions on the encoder under which oversampling has a detrimental effect on the distortion.

Appendix A
Proof of Theorem 1

For the MB-LTI sampler, the set Y_∞ of (16) is invariant under time shifts by an integer multiple of $1/f_s$ of the input $X_\epsilon(\cdot)$. Hence, for any $n \in \mathbb{Z}$, the distribution of $X(t)$ and $X(t+n/f_s)$ conditioned on the sigma algebra generated by Y_∞ are identical. It follows that the process $\widetilde{X}_T(\cdot)$ of (17) has an asymptotic distribution as $T \to \infty$ that is cyclostationary with period $1/f_s$ [59] (also known as $1/f_s$-ergodic [60]). Denote by

$$\widetilde{X}(t) = \mathbb{E}\left[X(t)|Y_\infty\right], \quad t \in \mathbb{R},$$

the process obeying the asymptotic distribution law of $\widetilde{X}_T(\cdot)$ when $T \to \infty$. It follows from (18) that with S a MB-LTI sampler, the asymptotic ADX distortion is given by

$$D(S, R) = \mathsf{mmse}(S) + D_{\widetilde{X}}(R),$$

where

$$\mathsf{mmse}(S) = \lim_{T\to\infty} \tfrac{1}{T} \int_{-T/2}^{T/2} \left(X(t) - \widetilde{X}(t)\right)^2 \, dt,$$

and $D_{\widetilde{X}}(R)$ is the DRF of the process $\widetilde{X}(\cdot)$.

We note that $\widetilde{X}(\cdot)$ can be derived in a closed form using a procedure that extends the Wiener filter, see [21], [49], [61]. Since cyclostationary processes are in particular asymptotic mean stationary processes [62], it follows from [45] that the DRF of $\widetilde{X}(\cdot)$ equals its information (a.k.a. Shannon's) DRF, i.e., the infimum over conditional probability distributions with mutual information rate not exceeding R. A closed form expression for this information DRF was derived in [9] in terms of the pre-sampling filters $H_1(f), \ldots, H_L(f)$ and the PSDs $S_X(f)$ and $S_\epsilon(f)$. Under the special case where the supports of $H_1(f), \ldots, H_L(f)$ are disjoint, this expression from [9] reduces to

$$D(S, R) = \mathsf{mmse}\,(S) + \sum_{l=1}^{L} \int_{-\frac{f_s}{2}}^{\frac{f_s}{2}} \min\left\{\widetilde{S}_l(f), \theta\right\} df \qquad \text{(55a)}$$

$$R_\theta = \frac{1}{2} \sum_{l=1}^{L} \int_{-\frac{f_s}{2}}^{\frac{f_s}{2}} \log^+\left[\widetilde{S}_l(f)/\theta\right] df, \qquad \text{(55b)}$$

where

$$\widetilde{S}_l(f) \triangleq \frac{\sum_{n\in\mathbb{Z}} S_X^2(f - f_s n)\mathbf{1}_{\mathrm{supp}\ H_l}(f - f_s n)}{\sum_{n\in\mathbb{Z}} \left[S_{X_\epsilon}(f - f_s n)\right]},$$

and

$$\mathsf{mmse}\,(S) = \sigma_X^2 - \sum_{l=1}^{L} \int_{-\frac{f_s}{2}}^{\frac{f_s}{2}} \widetilde{S}_l(f) df.$$

Next, let $F_{f_s}^\star$ be a set of Lebesgue measure at most f_s that maximizes (25). We now show that $F_{f_s}^\star$ can be approximated by L intervals of measure at most f_s/L. Let $\epsilon > 0$. Consider the measure μ_{S_X} defined by

$$\mu_S(A) = \int_A \frac{S_X^2(f)}{S_{X_\epsilon}(f)} df$$

for all Lebesgue measurable set A.

Since $S_X(f)$ is $\mathrm{L}_1(\mathbb{R})$, we can choose a set $G \subset F_{f_s}^\star$ such that $S_X(f)$ is bounded on G and such that $\mu_S(G) > \mu_S(F^\star) - \epsilon/3$. The measure μ_S is absolutely continuous with respect to the Lebesgue measure and hence is a regular measure [63]. Therefore, there exists M intervals $I_1, \ldots, I_M$ such that $\cup_{i=1}^M I_i \subset G$ and $\mu_S(\cup_{i=1}^M I_i) > \mu_S(G) - \epsilon/3 >$

$\mu_S(F^\star_{f_s}) - 2\epsilon/3$. We can assume that $I_1, \ldots, I_M$ are disjoint; otherwise we use $I_1' = I_1$, $I_2' = I_2 \setminus I_1'$, $I_3' = I_3 \setminus (I_1' \cup I_2')$, and so forth. Therefore, $\sum_{i=1}^M \mu(I_i) \le f_s$. For $\delta > 0$, let $L_i = \lfloor M\mu(I_i)/\delta \rfloor$ and $L = \sum_{i=1}^M L_i$. We now define L pre-sampling filters as follows: for each $i = 1, \ldots, M$, consider L_i disjoint intervals $I_{i,1}, \ldots, I_{i,j}$ of length $r = \delta/M$ that are sub-intervals of I_i. Since $\mu(I_i) \ge L_i r$, such L_i intervals exist and we set $A_i = I_i \setminus \cup_{j=1}^{L_i} I_{i,j}$. That is, A_i is the part of the interval I_i that is not covered by these L_i intervals. In particular, $\mu(A_i) \le r$. Finally, set the support of each filter $H_{i,j}$ to be $I_{i,j}$. Note that

$$\mu(\sum_{i,j} \text{supp } H_{i,j}) = \sum_{i=1}^M L_i r \le r \sum_{i=1}^M M\mu(I_i)/\delta \le f_s.$$

This way we have defined $L = L_1 + \ldots + L_M$ filters, each of passband of width $r \le f_s/L$. It remains to show that

$$\sum_{i,j} \mu_S(\text{supp } H_{i,j}) = \sum_{i,j} \int_{\text{supp } H_{i,j}} \frac{S_X^2(f)}{S_{X_\epsilon}(f)} df > \int_{F^\star_{f_s}} \frac{S_X^2(f)}{S_{X_\epsilon}(f)} df - \epsilon.$$

Denote by m_s the essential supremum of $S_X(f)$ on G and note that $S_{X|X_\epsilon}(f) \le m_s$ on G as well. We have

$$\mu_S(A_i) = \int_{A_i} S_{X|X_\epsilon}(f) df \le m_s \mu(A_i) \le m_s r.$$

It follows that

$$\begin{aligned}
&\mu_S(\sum_{i,j} \text{supp } H_{i,j}) \\
&= \sum_{i=1}^M \sum_{j=1}^{L_i} \mu_S(I_{i,j}) = \sum_{i=1}^M \mu_S(I_i) - \sum_{i=1}^M \mu_S(A_i) \\
&\ge \mu_S(G) - \epsilon/3 - M m_s r \ge \mu_S(F^\star_{f_s}) - 2\epsilon/3 - M m_s \delta.
\end{aligned}$$

Taking $\delta = \epsilon/(3\ M m_s)$ leads to the desired result.

To summarize, we constructed L interval $F_1, \ldots, F_L$, each of measure at most f_s/L, such that

$$\sum_{l=1}^L \int_{F_l} S_{X|X_\epsilon}(f) df + \delta > \int_{F^\star_{f_s}} S_{X|X_\epsilon}(f) df.$$

We now use use the following Proposition, proof of which can be found in [64, Proposition 3.4]:

Proposition 8: Fix $R > 0$ and set $A \subset \mathbb{R}$. For an integrable function f over A, define

$$\begin{aligned}
D(f) &= -\int_A [f(x) - \theta]^+ dx \\
R &= \frac{1}{2} \int_A \log^+ [f(x)/\theta] dx.
\end{aligned}$$

Let f and g be two integrable functions such that

$$\int_A f(x) dx \le \int_A g(x) dx.$$

Then $D(g) \le D(f)$.

We use Proposition 8 with $A = F^\star \cup F_1 \cup \ldots \cup F_L$,

$$f(x) = \mathbf{1}_{F^\star}(x) S_{X|X_\epsilon}(x),$$

and

$$g(x) = \mathbf{1}_{\cup_{l=1}^L F_l}(x) \left(S_{X|X_\epsilon}(f) + \delta \right).$$

Note that $D^\star(f_s, R)$ is a water-filling expression of the form (26) over $f(x)$ and A. Denote by D_δ the function defined by a water-filling expression over $g(x)$. Since $g(x) \ge f(x)$, it follows from Proposition 8 that

$$D_\delta \le D^\star(f_s, R).$$

Since D_δ is continuous in δ and since $\lim_{\delta \to 0} D_\delta = D(S, R)$, for $\epsilon > 0$ there exists L and δ such that $D(S, R) + \epsilon > D_\delta \ge D^\star(f_s, R)$.

Appendix B
Proof of Theorem 2

We first show that $D(S, R) \ge D^\star(f_s, R)$ for any bounded linear sampler $S = (K_H, \Lambda)$, with $d^+(\Lambda) \le f_s$. Consider the following cases of the sampler S in ADX:

(i) S is a MB-LTI uniform sampler of sampling rate f_s.
(ii) $S = (K_H, \Lambda)$ is a bounded linear sampler such that Λ is periodic with uniform density f_s.
(iii) $S = (K_H, \Lambda)$ is any bounded linear sampler such that $d^+(\Lambda) \le f_s$.

We show that case (iii) follows from (ii) which follows from (i).

A. Case (i)

Let S be an MB-LTI sampler with L sampling branches and fix a sampling rate f_s. For a given L, f_s, $S_X(f)$ and $S_\epsilon(f)$, consider a set of L filter $H_1, \ldots, H_L$ that minimizes $D(S, R)$. If follows from [9, Th. 21] that the support of each H_l in such a set is a bounded aliasing-free set for sampling at rate f_s/L. That is, for any $f_1, f_2 \in$ supp H_l, $f_1 \neq f_s$ modulo the grid $\mathbb{Z} f_s/L$. Since we are interested in bounding $D(S, R)$ and hence $D_T(S, R)$ from below, we can assume without loss of generality that the support of $H_1, \ldots, H_L$ satisfies the aliasing free condition. With this assumption, a closed form expression for $D(S, R)$ follows from [9, Th. 21]

$$D(S, R) = \sigma_X^2 - \sum_{l=1}^L \int_{\text{supp } H_l} \min \left[S_{X|X_\epsilon}(f) - \theta \right]^+ df \quad \text{(56a)}$$

$$R_\theta = \frac{1}{2} \sum_{l=1}^L \int_{\text{supp } H_l} \log^+ \left[S_{X|X_\epsilon}(f)/\theta \right] df. \quad \text{(56b)}$$

Furthermore, it follows from [9, Proposition 2] that the Lebesgue measure of $H_l^\star$ is at most f_s/L. Therefore, the Lebesgue measure of the union of supp $H_1, \ldots,$ supp H_L is at most f_s. Since Propositon 8 implies that a water-filling expression of the form (56a) is non-increasing in the function $S_{X|X_\epsilon}(f)$, it follows that (56a) is bounded from below by

$$\begin{aligned}
D^\star &= \sigma_X^2 - \int_{F^\star} \left[S_{X|X_\epsilon}(f) - \theta \right]^+ df \\
R &= \frac{1}{2} \int_{F^\star} \log^+ \left[S_{X|X_\epsilon}(f)/\theta \right] df,
\end{aligned}$$

which, by definition, equals $D^\star(f_s, R)$.

B. Case (ii)

Assume that the sampling set Λ is periodic with period T_0, i.e. it satisfies $\Lambda = \Lambda + T_0$. Assume moreover that $K_H(t + T_0 k, \tau) = K_H(t, \tau)$ for all $k \in \mathbb{Z}$, i.e. $K_H(t, \tau)$ is periodic in t with period T_0. Denote by L the number of points in Λ in the interval $[-T_0/2, T_0/2]$. Due to periodicity, $\lfloor L/T_0 \rfloor \leq d(\Lambda_T) \leq \lceil L/T_0 \rceil$ and so the symmetric density of Λ exists and equals L/T_0. Denote by $t_0, \ldots, t_{L-1}$ the L members of $\Lambda_{T_0} = \Lambda \cap [-T_0/2, T_0/2]$, where without loss of generality we can assume that $\pm T_0/2 \notin \Lambda$. Continue to enumerate the members of Λ that are larger than t_{L-1} in the positive direction by $t_L, t_{L+1}, \ldots$, and the elements of Λ smaller than t_0 in the negative direction by $t_{-1}, t_{-2}, \ldots$. By the periodicity of Λ, $t_{l+Lk} = t_l + T_0 k$ for all $l = 0, \ldots, L-1$ and $k \in \mathbb{Z}$. For $n = l + kL$, and $t_n < T/2$, each sample Y_n in the vector of samples Y_T satisfies

$$\begin{aligned} Y_n &= \int_{-\infty}^{\infty} K_H(t_{l+Lk}, s) X_\epsilon(s) ds \\ &= \int_{-\infty}^{\infty} K_H(t_l + T_0 k, s) X_\epsilon(s) ds \\ &= \int_{-\infty}^{\infty} K_H(t_l, s) X_\epsilon(s) ds = h_l(s - t_l) X_\epsilon(s) df, \end{aligned}$$

where, for $l = 0, \ldots, L-1$, we denoted $h_l(s) \triangleq K_H(t_l, s + t_l)$. We define the vector valued process $\mathbf{Y}[\cdot] = \{\mathbf{Y}[k],\ k \in \mathbb{Z}\}$ by

$$\mathbf{Y}[k] = (Y_{Lk}, Y_{Lk+1}, \ldots, Y_{Lk+L-1}), \quad k \in \mathbb{Z}.$$

That is, the kth sample of $\mathbf{Y}[\cdot]$ is a vector in $\mathbb{R}^L$ consists of L consecutive samples of $X_\epsilon(\cdot)$. Note that $Y[\cdot]$ is independent of the time horizon T. Since each $h_l(s)$ defines an LTI system, it follows that sampling with the periodic set Λ and the pre-processing system K_H is equivalent to sampling using L uniform sampling branches each of sampling rate $1/T_0$. From case (i) of the proof, it follows that $D(S, R) \geq D^\star(L/T_0, R) = D^\star(d(\Lambda), R)$.

C. Case (iii)

We now consider the general case of $S = (\Lambda, K_H)$ an arbitrary bounded linear sampler. For a sequence $\{T_n,\ n = 1, 2, \ldots\}$ such that $\lim_{n\to\infty} T_n = \infty$, denote

$$d_n \triangleq d_{T_n}(\Lambda) = \frac{\Lambda \cap [-T_n/2, T_n/2]}{T_n},$$

and let Y_{T_n} be the vector of d_n samples obtained by sampling $X_\epsilon(\cdot)$ using S over the interval $[-T_n/2, T_n/2]$. In addition, define the set $\tilde{\Lambda}_n$ to be the periodic extension of Λ_{T_n}, i.e,

$$\tilde{\Lambda}_n \triangleq \Lambda_{T_n} + T_n \mathbb{Z}.$$

Therefore, $\tilde{\Lambda}_n$ is a periodic sampling set with period T_n and, consequently, symmetric density d_n. We also extend $K_H(t, s)$ periodically as

$$\tilde{K}_n(t, s) \triangleq K_H([t], \tau)$$

where here and henceforth $[t]$ denotes t *modulo the grid* $T_n\mathbb{Z}$ (i.e. $t = [t] + kT_n$ where $k \in \mathbb{Z}$ and $0 \leq [t] < T_n$). Let $\tilde{S}_n \triangleq (\tilde{\Lambda}_n, \tilde{K}_n)$. We have

$$D_{T_n}(S, R) \overset{(a)}{\leq} D_{T_n}(\tilde{S}_n, R) \overset{(b)}{\geq} D(\tilde{S}_n, R) \overset{(c)}{\geq} D^\star(d_n, R), \quad (57)$$

where: (a) is becasue Y_{T_n} is a subset of the samples obtained by sampling with $\tilde{S}_n$, (b) follows since the distribution of the estimator of $X(\cdot)$ from the samples obtained by a MB-LTI sampler is cyclostationary, hence enlarging the time horizon T can only reduce distortion [45], and (c) is obtained from part (ii) of the proof.

Since $D^\star(f_s, R)$ is continuous and non-increasing in f_s, we have

$$\lim_{n\to\infty} D^\star(d_n, R) \geq D^\star(d^+(\Lambda), R).$$

Therefore, since any unbounded sequence of time horizons $\{T_n\}$ satisfies (57), we conclude that

$$\liminf_{T\to\infty} D_T(S, R) = D(S, R) \geq D^\star(d^+(\Lambda), R).$$

The finite time horizon statement of Theorem 2 is obtained using a similar procedure as steps (i)-(iii) above with the following changes: in (i), we consider $D_T(S, R)$ and note that since S is a MB-LTI sampler, the distribution of $X(\cdot)$ given Y_∞ is cyclostationary so $D_T(S, R) \geq D(S, R)$. As a result, the bound in step (ii) also bound $D_T(S, R)$ from below. Finally, in step (iii), we denote $d_T = d_T(\Lambda)$ and consider a periodic extended sampler $\tilde{S}_T \triangleq (\tilde{\Lambda}_T, \tilde{K}_T)$ where $\tilde{\Lambda}_T \triangleq \Lambda_T + T\mathbb{Z}$ and $\tilde{K}_T(t, s) \triangleq K_H([t], \tau)$. We have

$$\begin{aligned} D_T(S, R) &\overset{(a)}{\leq} D_T(\tilde{S}_T, R) \overset{(b)}{\geq} D(\tilde{S}_T, R) \\ &\overset{(c)}{\geq} D^\star(d_T, R) \overset{(d)}{\geq} D^\star(f_s, R). \end{aligned} \quad (58)$$

where (a) - (c) follow from the same arguments as in (57), and (d) is because $d_T \leq f_s$.

Appendix C
Proof of Proposition 3

Let (R, D) be a point on the curve $\left(R, D_{X|X_\epsilon}(R)\right)$. For θ such that

$$R = \frac{1}{2} \int_{-\infty}^{\infty} \log^+ \left[S_{X|X_\epsilon}(f)/\theta \right] df,$$

denote $F_\theta \triangleq \left\{ f \in \mathbb{R} \ :\ S_{X|X_\epsilon}(f) > \theta \right\}$, so that $f_R = \mu(F_\theta)$,

$$R = \frac{1}{2} \int_{F_\theta} \log \frac{S_{X|X_\epsilon}(f)}{\theta} df,$$

and

$$D = \sigma_X^2 - \int_{F_\theta} \left(S_{X|X_\epsilon}(f) - \theta \right) df. \quad (59)$$

Let $F^\star \subset \mathbb{R}$ be such that

$$D^\star(f_R, R) = \sigma_X^2 - \int_{F^\star} \left[S_{X|X_\epsilon}(f) df - \theta \right]^+, \quad (60)$$

and

$$R = \frac{1}{2} \int_{F^\star} \log^+ [S_{X|X_\epsilon}(f)/\theta] df.$$

From the definition of $D^\star(f_s, R)$, it follows that

$$\int_{F^\star} S_{X|X_\epsilon}(f) df \geq \int_{F_\theta} S_{X|X_\epsilon}(f) df. \quad (61)$$

Since the distortion expressions (59) and (60) are non-increasing in $\int_{F^\star} S_{X|X_\epsilon}(f)df$ and $\int_{F_\theta} S_{X|X_\epsilon}(f)df$, respectively, it follows from (61) that $D^\star(f_R, R) \leq D_{X|X_\epsilon}(R)$. In order to prove the reverse inequality, note that for any bounded linear sampler S and R we have

$$D(S, R) \geq D_{X|X_\epsilon}(R).$$

However, it follows from Theorem 1 that $D^\star(f_s, R)$ is achievable, and hence $D^\star(f_s, R) \geq D_{X|X_\epsilon}(R)$. Evidently, this same inequality can be derived directly from the definition of $D^\star(f_s, R)$ in (26) and the expression for $D_{X|X_\epsilon}(R)$ in (22) (that is, without using Theorem 1).

Appendix D
Proof of Proposition 5

For $0 \leq \Delta \leq 1$ define

$$X_\Delta[n] \triangleq X\left((n+\Delta)T_s\right), \quad n \in \mathbb{Z},$$

where $T_s \triangleq f_s^{-1}$. Also define $\hat{X}_\Delta[n]$ to be the optimal MSE estimator of $X_\Delta[n]$ from $\hat{Y}[\cdot]$, that is

$$\hat{X}_\Delta[n] = \mathbb{E}\left[X_\Delta[n]|\hat{Y}[\cdot]\right], \quad n \in \mathbb{Z}.$$

The MSE in (44) can be written as

$$\begin{aligned}
&\mathsf{mmse}_{X|\hat{Y}} \\
&= \lim_{N\to\infty} \frac{1}{2N+1} \int_{-N}^{N+1} \mathbb{E}\left(X(t) - \hat{X}(t)\right)^2 dt \\
&= \lim_{N\to\infty} \frac{1}{2N+1} \sum_{n=-N}^{N} \\
&\quad \times \int_0^1 \mathbb{E}\left(X\left((n+\Delta)T_s\right) - \hat{X}\left((n+\Delta)T_s\right)\right)^2 d\Delta \\
&= \lim_{N\to\infty} \frac{1}{2N+1} \sum_{n=-N}^{N} \int_0^1 \mathbb{E}\left(X_\Delta[n] - \hat{X}_\Delta[n]\right)^2 d\Delta \\
&= \int_0^1 \mathbb{E}\left(X_\Delta[n] - \hat{X}_\Delta[n]\right)^2 d\Delta. \qquad (62)
\end{aligned}$$

Note that $S_{X_\Delta}\left(e^{2\pi i\phi}\right) = S_Y\left(e^{2\pi i\phi}\right)$ and $X_\Delta[\cdot]$ and $\hat{Y}[\cdot]$ are jointly stationary with cross-PSD

$$\begin{aligned}
S_{X_\Delta \hat{Y}}\left(e^{2\pi i\phi}\right) &= S_{X_\Delta}\left(e^{2\pi i\phi}\right) \\
&= f_s \sum_{k\in\mathbb{Z}} S_X\left(f_s(k-\phi)\right) e^{2\pi i\Delta(k-\phi)}.
\end{aligned}$$

Denote by $S_{X_\Delta|\hat{Y}}\left(e^{2\pi i\phi}\right)$ the PSD of the estimator obtained by the discrete Wiener filter for estimating $X_\Delta[\cdot]$ from $\hat{Y}[\cdot]$. We have

$$\begin{aligned}
&S_{X_\Delta|\hat{Y}}\left(e^{2\pi i\phi}\right) \\
&= \frac{S_{X_\Delta\hat{Y}}\left(e^{2\pi i\phi}\right) S^*_{X_\Delta\hat{Y}}\left(e^{2\pi i\phi}\right)}{S_{\hat{Y}}\left(e^{2\pi i\phi}\right)} \\
&= \sum_{n,k} \frac{f_s^2 S_{X_a}\left(f_s(k-\phi)\right) S^*_{X_a}\left(f_s(n-\phi)\right) e^{2\pi i\Delta(k-n)}}{S_Y\left(e^{2\pi i\phi}\right) + S_\eta\left(e^{2\pi i\phi}\right)}, \qquad (63)
\end{aligned}$$

where $S_{X_a}(f) = S_X(f)H^*(f)$ is the cross-PSD of $X(\cdot)$ and the signal at the output of the filter $H(f)$. The estimation error in Wiener filtering is given by

$$\begin{aligned}
&\mathbb{E}\left(X_\Delta[n] - \hat{X}_\Delta[n]\right)^2 \\
&= \int_{-\frac{1}{2}}^{\frac{1}{2}} S_{X_\Delta}\left(e^{2\pi i\phi}\right) d\phi - \int_{-\frac{1}{2}}^{\frac{1}{2}} S_{X_\Delta|\hat{Y}}\left(e^{2\pi i\phi}\right) d\phi \\
&= \sigma_X^2 - \int_{-\frac{1}{2}}^{\frac{1}{2}} S_{X_\Delta|\hat{Y}}\left(e^{2\pi i\phi}\right) d\phi. \qquad (64)
\end{aligned}$$

Equations (62), (63) and (64) lead to

$$\begin{aligned}
&D_{PCM}(f_s, q, H) \\
&= \int_0^1 \mathbb{E}\left(X_\Delta[n] - \hat{X}_\Delta[n]\right)^2 d\Delta \\
&= \sigma_X^2 - \int_{-\frac{1}{2}}^{\frac{1}{2}} \int_0^1 S_{X_\Delta|\hat{Y}}\left(e^{2\pi i\phi}\right) d\phi \\
&\overset{a}{=} \sigma_X^2 - \int_{-\frac{1}{2}}^{\frac{1}{2}} \frac{f_s \sum_{k\in\mathbb{Z}} \left|S_{X_a}\right|^2 \left(f_s(k-\phi)\right)}{S_Y\left(e^{2\pi i\phi}\right) + S_\eta\left(e^{2\pi i\phi}\right)} d\phi, \qquad (65)
\end{aligned}$$

where (a) follows from (63) and the orthogonality of the functions $\left\{e^{2\pi xk}, k \in \mathbb{Z}\right\}$ over $0 \leq x \leq 1$. Equation (50) is obtained from (65) by changing the integration variable from ϕ to $f = \phi f_s$.

The optimal MMSE linear estimator of $X(t)$ from $\hat{Y}$ has the property that the estimation error is uncorrelated with any sample from $\hat{Y}[\cdot]$, namely,

$$\mathbb{E}\left[\left(X(t) - \sum_n w[n]\hat{Y}[n]\right)\hat{Y}[k]\right] = 0$$

for all $k \in \mathbb{Z}$. This implies that

$$\int_{-\infty}^{\infty} R_X(t - u - k/f_s)h(u)du = \sum_n w[n]R_{\hat{Y}}[n-k]. \qquad (66)$$

Taking the discrete time Fourier transform of both sides with respect to k in (66) leads to

$$\begin{aligned}
&f_s \sum_m S_X\left(f_s(\phi-k)\right) e^{-2\pi i t f_s(\phi-k)} H^*\left(f_s(\phi-k)\right) \\
&\qquad = W\left(e^{2\pi i\phi}\right) S_{\hat{Y}}\left(e^{2\pi i\phi}\right),
\end{aligned}$$

or

$$\begin{aligned}
&W\left(e^{2\pi i\phi}\right) \\
&\quad = \frac{f_s \sum_m S_X\left(f_s(\phi-k)\right) e^{-2\pi i t f_s(\phi-k)} H^*\left(f_s(\phi-k)\right)}{S_{\hat{Y}}\left(e^{2\pi i\phi}\right)}.
\end{aligned}$$

Note that the last expression equals the discrete-time Fourier transform of the function $\tilde{w}(t - n/f_s)$ (with respect to n), where the impulse response of $\tilde{w}(t)$ is given by

$$W(f) = \frac{H^*(f)S_X(f)}{\sum_{k\in\mathbb{Z}} |H(f)|^2 S_X(f - f_s k) + \sigma_\eta^2/f_s}, \qquad (67)$$

completing the proof. □

ACKNOWLEDGMENT

The authors thank Robert Gray for helpful discussions regarding the white noise approximation, and to the associate editor and an anonymous reviewer for constructive comments that greatly improved the paper.

REFERENCES

[1] A. Kipnis, A. J. Goldsmith, and Y. C. Eldar, "Gaussian distortion-rate function under sub-Nyquist nonuniform sampling," in *Proc. 52nd Annu. Allerton Conf. Commun., Control, Comput. (Allerton)*, 2014, pp. 874–880.

[2] A. Kipnis, Y. C. Eldar, and A. J. Goldsmith, "Optimal trade-off between sampling rate and quantization precision in Sigma-Delta A/D conversion," in *Proc. 53th Annu. Allerton Conf. Commun., Control, Comput. (Allerton)*, 2015, pp. 627–631.

[3] A. Kipnis, A. J. Goldsmith, and Y. C. Eldar, "Sub-Nyquist sampling achieves optimal rate-distortion," in *Proc. Inf. Theory Workshop (ITW)*, Apr. 2015, pp. 1–5.

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

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

[6] 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.

[7] R. H. Walden, "Analog-to-digital converter survey and analysis," *IEEE J. Sel. Areas Commun.*, vol. 17, no. 4, pp. 539–550, Apr. 1999.

[8] Y. C. Eldar and T. Michaeli, "Beyond bandlimited sampling," *IEEE Signal Process. Mag.*, vol. 26, no. 3, pp. 48–68, May 2009.

[9] 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.

[10] H. J. Landau, "Sampling, data transmission, and the Nyquist rate," *Proc. IEEE*, vol. 55, no. 10, pp. 1701–1706, Oct. 1967.

[11] R. Venkataramani and Y. Bresler, "Optimal sub-Nyquist nonuniform sampling and reconstruction for multiband signals," *IEEE Trans. Signal Process.*, vol. 49, no. 10, pp. 2301–2313, Oct. 2001.

[12] I. Daubechies and W. Sweldens, "Factoring wavelet transforms into lifting steps," *J. Fourier Anal. Appl.*, vol. 4, no. 3, pp. 247–269, 1998.

[13] R. Dobrushin and B. Tsybakov, "Information transmission with additional noise," *IRE Trans. Inf. Theory*, vol. 8, no. 5, pp. 293–304, Sep. 1962.

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

[15] D. L. Donoho, M. Vetterli, R. A. De Vore, and I. Daubechies, "Data compression and harmonic analysis," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2435–2476, Oct. 1998.

[16] 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.

[17] M. M. Dodson and A. M. Silva, "Fourier analysis and the sampling theorem," *Proc. Roy. Irish Acad., Sec. A, Math. Phys. Sci.*, vol. 24, no. 2, pp. 81–108, 1985.

[18] S. P. Lloyd, "A sampling theorem for stationary (wide sense) stochastic processes," *Trans. Amer. Math. Soc.*, vol. 92, no. 1, pp. 1–12, 1959.

[19] A. Balakrishnan, "A note on the sampling principle for continuous signals," *IRE Trans. Inf. Theory*, vol. 3, no. 2, pp. 143–146, Jun. 1957.

[20] D. Chan and R. Donaldson, "Optimum pre- and postfiltering of sampled signals with application to pulse modulation and data compression systems," *IEEE Trans. Commun. Technol.*, vol. CT-19, no. 2, pp. 141–157, Apr. 1971.

[21] M. B. Matthews, "On the linear minimum-mean-squared-error estimation of an undersampled wide-sense stationary random process," *IEEE Trans. Signal Process.*, vol. 48, no. 1, pp. 272–275, Jan. 2000.

[22] F. J. Beutler, "Sampling theorems and bases in a Hilbert space," *Inf. Control*, vol. 4, nos. 2–3, pp. 97–117, 1961.

[23] H. Dym and H. McKean, *Gaussian Processes, Function Theory, and the Inverse Spectral Problem*. Mineola, NY, USA: Dover, 2008.

[24] H. G. Feichtinger and K. Gröchenig, "Irregular sampling theorems and series expansions of band-limited functions," *J. Math. Anal. Appl.*, vol. 167, no. 2, pp. 530–556, 1992.

[25] A. Beurling and L. Carleson, *The Collected Works of Arne Beurling: Complex Analysis*, vol. 1. Cambridge, MA, USA: Birkhäuser, 1989.

[26] H. Landau, "Necessary density conditions for sampling and interpolation of certain entire functions," *Acta Mathematica*, vol. 117, no. 1, pp. 37–52, 1967.

[27] E. Mohammadi and F. Marvasti, "Sampling and distortion tradeoffs for bandlimited periodic signals," *IEEE Trans. Inf. Theory*, vol. 64, no. 3, pp. 1706–1724, Mar. 2018.

[28] J. Wu and N. Sun, "Optimum sensor density in distortion-tolerant wireless sensor networks," *IEEE Trans. Wireless Commun.*, vol. 11, no. 6, pp. 2056–2064, Jun. 2012.

[29] Y. Sung, H. V. Poor, and H. Yu, "Optimal node density for two-dimensional sensor arrays," in *Proc. 5th IEEE Sensor Array Multichannel Signal Process. Workshop*, Jul. 2008, pp. 271–274.

[30] J. R. Higgins, *Sampling Theory in Fourier and Signal Analysis: Foundations*. London, U.K.: Oxford Univ. Press, 1996.

[31] F. Marvasti, *Nonuniform Sampling: Theory and Practice*. New York, NY, USA: Springer, 2012.

[32] M. Unser and A. Aldroubi, "A general sampling theory for nonideal acquisition devices," *IEEE Trans. Signal Process.*, vol. 42, no. 11, pp. 2915–2925, Nov. 1994.

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

[34] 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.

[35] A. Kipnis, S. Rini, and A. J. Goldsmith, "The indirect rate-distortion function of a binary I.I.D source," in *Proc. IEEE Inf. Theory Workshop Fall (ITW)*, Oct. 2015, pp. 352–356.

[36] X. Liu, O. Simeone, and E. Erkip, "Lossy computing of correlated sources with fractional sampling," *IEEE Trans. Commun.*, vol. 61, no. 9, pp. 3685–3696, Sep. 2013.

[37] R. Zamir and M. Feder, "Rate-distortion performance in coding bandlimited sources by sampling and dithered quantization," *IEEE Trans. Inf. Theory*, vol. 41, no. 1, pp. 141–154, Jan. 1995.

[38] D. Marco and D. L. Neuhoff, "Entropy of highly correlated quantized data," *IEEE Trans. Inf. Theory*, vol. 56, no. 5, pp. 2455–2478, May 2010.

[39] A. Kumar, P. Ishwar, and K. Ramchandran, "High-resolution distributed sampling of bandlimited fields with low-precision sensors," *IEEE Trans. Inf. Theory*, vol. 57, no. 1, pp. 476–492, Jan. 2011.

[40] Y. C. Eldar and G. Kutyniok, *Compressed Sensing: Theory and Applications*. Cambridge, U.K.: Cambridge Univ. Press, 2012.

[41] Y. Wu and S. Verdú, "Optimal phase transitions in compressed sensing," *IEEE Trans. Inf. Theory*, vol. 58, no. 10, pp. 6241–6263, Oct. 2012.

[42] A. Kipnis, G. Reeves, Y. C. Eldar, and A. J. Goldsmith, "Fundamental limits of compressed sensing under optimal quantization," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, 2017, pp. 2148–2152.

[43] 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)*, 2018, pp. 71–75.

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

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

[46] A. H. Zemanian, *Distribution Theory and Transform Analysis: An Introduction to Generalized Functions, With Applications*. Chelmsford, MA, USA: Courier Corporation, 1965.

[47] I. Gelfand and G. Shilov, *Generalized Functions*, vol. 2. Boca Raton, FL, USA: Academic, 1968.

[48] I. M. Gelfand, N. Â. Vilenkin, and A. Feinstein, *Generalized Functions*, vol. 4. Boca Raton, FL, USA: Academic, 1964.

[49] Y. Chen, Y. C. Eldar, and A. J. Goldsmith, "Shannon meets Nyquist: Capacity of sampled Gaussian channels," *IEEE Trans. Inf. Theory*, vol. 59, no. 8, pp. 4889–4914, Aug. 2013.

[50] H. S. Black and J. O. Edson, "Pulse code modulation," *Trans. Amer. Inst. Electr. Eng.*, vol. 66, no. 1, pp. 895–899, Jan. 1947.

[51] B. M. Oliver, J. R. Pierce, and C. E. Shannon, "The Philosophy of PCM," *Proc. IRE*, vol. 36, no. 11, pp. 1324–1331, Nov. 1948.

[52] R. M. Gray and D. L. Neuhoff, "Quantization," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2325–2383, Oct. 1998.

[53] H. Viswanathan and R. Zamir, "On the whiteness of high-resolution quantization errors," *IEEE Trans. Inf. Theory*, vol. 47, no. 5, pp. 2029–2038, Jul. 2001.

[54] S. Lloyd, "Least squares quantization in PCM," *IEEE Trans. Inf. Theory*, vol. IT-28, no. 2, pp. 129–137, Mar. 1982.

[55] J. M. de la Rosa, "Sigma-Delta modulators: Tutorial overview, design guide, and state-of-the-art survey," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 58, no. 1, pp. 1–21, Jan. 2011.
[56] V. K. Goyal, A. K. Fletcher, and S. Rangan, "Compressive sampling and lossy compression," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 48–56, Mar. 2008.
[57] B. Widrow, "A study of rough amplitude quantization by means of Nyquist sampling theory," *IRE Trans. Circuit Theory*, vol. 3, no. 4, pp. 266–276, Dec. 1956.
[58] W. R. Bennett, "Spectra of quantized signals," *Bell Syst. Tech. J.*, vol. 27, no. 3, pp. 446–472, Jul. 1948.
[59] W. R. Bennett, "Statistics of regenerative digital transmission," *Bell Syst. Tech. J.*, vol. 37, no. 6, pp. 1501–1542, Nov. 1958.
[60] J. Nedoma, "On the ergodicity and r-ergodicity of stationary probability measures," *Z. Wahrsch. Verw. Gebiete*, vol. 4, no. 2, pp. 90–97, 1963.
[61] T. Michaeli and Y. C. Eldar, "High-rate interpolation of random signals from nonideal samples," *IEEE Trans. Signal Process.*, vol. 57, no. 3, pp. 977–992, Mar. 2009.
[62] R. Gray, *Probability, Random Processes, and Ergodic Properties*. New York, NY, USA: Springer, 2009.
[63] H. L. Royden and P. Fitzpatrick, *Real Analysis*, vol. 198, New York, NY, USA: Macmillan, 1988, no. 8.
[64] A. Kipnis, "Fundamental distortion limits of analog-to-digital compression," Ph.D. dissertation, Dept. Elect. Eng., Stanford Univ., Stanford, CA, USA, 2018.

Alon Kipnis (S'14) received his B.Sc. degree in mathematics (summa cum laude) and his B.Sc. degree in electrical engineering (summa cum laude), both in 2010, and his M.Sc. degree in mathematics in 2012, all from Ben-Gurion University of the Negev. He recently received his Ph.D. degree in electrical engineering from Stanford University, 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.

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 ofstatistical 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.

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 multiple public and private companies. 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, a Fellow of the IEEE and of Stanford, and has received several awards for her work, including the IEEE Sumner Technical Field Award, the ACM Athena Lecturer Award, the IEEE Comsoc Edwin H. Armstrong Achievement Award, the National Academy of Engineering Gilbreth Lecture Award, the Women in Communications Engineering 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, aswell as an inventor on 29 patents. She has also launched and led several multi-university researchprojects. Her research interests are in information theory and communication theory, and theirapplication to wireless communications and related fields. 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 on its 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.