Information Loss and Anti-Aliasing Filters in Multirate Systems

Bernhard C. Geiger and Gernot Kubin
Graz University of Technology, Austria
{geiger,g.kubin}@ieee.org

***Abstract*—This work investigates the information loss in a decimation system, i.e., in a downsampler preceded by an anti-aliasing filter. It is shown that, without a specific signal model in mind, the anti-aliasing filter cannot reduce information loss, while for a simple signal-plus-noise model it can. For the Gaussian case, the optimal anti-aliasing filter is shown to coincide with the one obtained from energetic considerations. For a non-Gaussian signal corrupted by Gaussian noise, the Gaussian assumption yields an upper bound on the information loss, suggesting filter design principles based on second-order statistics.**

I. Introduction

Multi-rate systems are ubiquitously used in digital systems to increase (upsample) or decrease (downsample) the rate at which a signal is processed. Especially downsampling is a critical operation since it can introduce aliasing, like sampling, and thus can cause information loss. Standard textbooks on signal processing deal with this issue by recommending an anti-aliasing filter prior to downsampling – resulting in a cascade which is commonly known as a decimator [1, Ch. 4.6]. In these books, this anti-aliasing filter is usually an ideal low-pass filter with a cut-off frequency of π/M, for an M-fold decimation system (cf. Fig. 1). Unser showed that this choice is optimal in terms of the mean-squared reconstruction error (MSE) only if the input process is such that the passband portion of its power spectral density (PSD) exceeds all aliased components [2]. Similarly, it was shown by Tsatsanis and Giannakis [3], that the filter minimizing the MSE is piecewise constant, M-aliasing-free (i.e., the aliased components of the M-fold downsampled frequency response do not overlap), and has a passband depending on the PSD of the input process. Specifically, the filter which permits most of the energy to pass aliasing-free is optimal in the MSE sense.

In this paper we consider a design objective vastly different from the MSE: information. The fact that information, compared to energy, can yield more successful system designs has long been recognized, e.g., for (non-linear) adaptive filters [4] or for state estimation using linear filters [5]. In information theory, transceiver filter design based on mutual information is covered in, e.g., [6], [7]. That information-theoretic design seems to become a trend recently is understandable: After all, it is information one wants to transmit, not energy. Finally, quantifying information relieves us from having to specify a reconstruction procedure: The information lost in the decimation system is independent from signal reconstruction, therefore a separate design of these two system components should be possible.

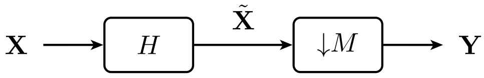

Fig. 1. Simple decimation system consisting of a linear filter H and an M-fold downsampler.

Our first result is surprising: Given mild assumptions on the input process of the decimation system, the information loss can be bounded *independently* of the anti-aliasing filter (see Section III). The reason is that under these assumptions every bit of the input process is treated equivalently, regardless of the amount of energy by which it is represented. In order to remedy this counter-intuitivity, Section IV considers Gaussian processes with a specific signal model in mind: The input to the decimation system is a relevant data signal corrupted by noise. As a corollary to a more general result, we show that for white noise the anti-aliasing filter minimizing the information loss coincides with the optimal filter of [3]. Since in most cases the Gaussian assumption is too restrictive, we let the signal process have arbitrary distribution in Section V, but keep the noise Gaussian. Following the approach of Plumbley in [8], we prove that the Gaussian assumption for the signal process yields an upper bound on the information loss in the general case. In other words, designing a filter based on the PSDs of the signal and noise processes guarantees a minimum information transfer over the decimation system. This justifies a filter design based on second-order statistics, i.e., on energetic considerations, also from an information-theoretic perspective. In Section VI we illustrate our results in a simple toy example.

Due to the lack of space, we only give an outline of our proofs. An extended version of this manuscript is currently in preparation.

II. Preliminaries and Notation

Throughout this work we adopt the following notation: $\mathbf{Z}$ is a real-valued random process, whose n-th sample is the random variable (RV) Z_n. We abbreviate $Z_i^j := \{Z_i, Z_{i+1}, \dots, Z_j\}$. The differential entropy [9, Ch. 8] and the Rényi information dimension [10] of Z_i^j are $h(Z_i^j)$ and $d(Z_i^j)$, respectively, provided these quantities exist and are finite. Finally, we define the M-fold blocking $\mathbf{Z}^{(M)}$ of $\mathbf{Z}$

as the sequence of M-dimensional RVs $Z_1^{(M)} := Z_1^M$, $Z_2^{(M)} := Z_{M+1}^{2M}$, and so on.

In this work, we often consider a process $\mathbf{Z}$ satisfying

Assumption 1. $\mathbf{Z}$ is stationary, has finite marginal differential entropy $h(Z_n)$, finite Shannon entropy of the quantized RV $\lfloor Z_n \rfloor$, and finite differential entropy rate

$$\bar{h}(\mathbf{Z}) := \lim_{n\to\infty} \frac{1}{n} h(Z_1^n) = \lim_{n\to\infty} h(Z_n|Z_1^{n-1}). \quad (1)$$

As a direct consequence of Assumption 1, the information dimension satisfies $d(Z_1^n) = n$ for all n, and the mutual information rate with a process $\mathbf{W}$ jointly stationary with $\mathbf{Z}$ exists and equals [11, Thm. 8.3]

$$\bar{I}(\mathbf{Z};\mathbf{W}) := \lim_{n\to\infty} \frac{1}{n} I(Z_1^n; W_1^n). \quad (2)$$

We introduce two measures of information loss for stationary stochastic processes: The first is an extension of the *relative information loss*[1] $l(Z \to g(Z))$ to stochastic processes (cf. [12], [13]):

$$l(\mathbf{Z} \to g(\mathbf{Z})) := \lim_{n\to\infty} l((Z_1^n \to g(Z_1^n)) \quad (3)$$

where we abused notation by applying g coordinate-wise. The second notion is an extension of [14], where we introduced the *relevant information loss*. Let $\mathbf{W}$ be a process statistically related to and jointly stationary with $\mathbf{Z}$, representing the relevant information content of $\mathbf{Z}$; for example, $\mathbf{W}$ might be the sign of $\mathbf{Z}$, or $\mathbf{Z}$ might be a noisy observation of $\mathbf{W}$. Then, the information loss rate relevant w.r.t. $\mathbf{W}$ is

$$\bar{L}_{\mathbf{W}}(\mathbf{Z} \to g(\mathbf{Z})) := \bar{I}(\mathbf{W};\mathbf{Z}) - \bar{I}(\mathbf{W}; g(\mathbf{Z})) \quad (4)$$

provided the quantities exists.

III. Relative Information Loss in a Downsampler

Consider the scenario depicted in Fig. 1, where $\mathbf{X}$ satisfies Assumption 1. It can be shown that if the linear filter H is stable and causal, also the output process $\tilde{\mathbf{X}}$ satisfies Assumption 1. Moreover, such a filter has no effect on the information content of the stochastic process in the sense that, for $\mathbf{S}$ jointly stationary with $\mathbf{X}$, $\bar{I}(\mathbf{X};\mathbf{S}) = \bar{I}(\tilde{\mathbf{X}};\mathbf{S})$.

To analyze the information loss rate in the downsampling device, we employ the relative information loss rate,

$$l(\tilde{\mathbf{X}}^{(M)} \to \mathbf{Y}) := \lim_{n\to\infty} l((\tilde{X}^{(M)})_1^n \to Y_1^n) \quad (5)$$

where we applied M-fold blocking to ensure that the mapping between $(\tilde{X}^{(M)})_1^n$ and Y_1^n is static. Downsampling, $Y_n := \tilde{X}_{nM}$, is now a projection to a single coordinate, hence [12]

$$l((\tilde{X}^{(M)})_1^n \to Y_1^n) = \frac{d((\tilde{X}^{(M)})_1^n|Y_1^n)}{d((\tilde{X}^{(M)})_1^n)} = \frac{n(M-1)}{nM}. \quad (6)$$

[1]Roughly speaking, $l(Z \to g(Z))$ captures the *percentage* of information lost by applying the function g to the RV Z.

Since the filter H is stable and causal and, thus, has no influence on the information content of the stochastic process, we abuse notation in (a) below and combine (3) with (6) to

$$l(\mathbf{X}^{(M)} \to \mathbf{Y}) \stackrel{(a)}{=} l(\tilde{\mathbf{X}}^{(M)} \to \mathbf{Y}) = \frac{M-1}{M}. \quad (7)$$

The amount of information lost in the decimation system in Fig. 1 is the same for all stable, causal filters H.

The question remains whether an *ideal* anti-aliasing filter can prevent information loss, since it guarantees that the downsampling operation is invertible. To show that the answer to this question is negative, take, for example,

$$H(\mathrm{e}^{\jmath\theta}) = \begin{cases} 1, & \text{if } |\theta| < \frac{\pi}{M} \\ 0, & \text{else} \end{cases}. \quad (8)$$

We decompose $\mathbf{X}$ in an M-channel filterbank: The k-th channel is characterized by analysis and synthesis filters being constant in the frequency band $(k-1)/M \le |\theta| < k/M$ and zero elsewhere. Let $\mathbf{Y}_{(k)}$ be the (M-fold downsampled) process in the k-th channel – clearly, $\mathbf{Y} \equiv \mathbf{Y}_{(1)}$. It can be shown that every $\mathbf{Y}_{(k)}$ satisfies Assumption 1 if $\mathbf{X}$ is Gaussian. Thus we obtain

$$l(\mathbf{X}^{(M)} \to \mathbf{Y}) \stackrel{(a)}{=} l(\mathbf{Y}_{(1)},\dots,\mathbf{Y}_{(M)} \to \mathbf{Y}_{(1)}) = \frac{M-1}{M} \quad (9)$$

where the information is again lost in a projection and where we abused notation in (a) since the filterbank decomposition is invertible. The ideal anti-aliasing low-pass filter prevents information from being lost in the downsampler by *destroying information itself.*

If the filter H is a cascade of a causal, stable filter and of one with a piecewise-constant transfer function (with less trivial intervals as pass-bands), the analysis still holds; Information is either lost in the filter or in the downsampler:

Theorem 1. *For a Gaussian process $\mathbf{X}$ satisfying Assumption 1, the relative information loss rate in the decimation system depicted in Fig. 1 satisfies*

$$l(\mathbf{X}^{(M)} \to \mathbf{Y}) \ge \frac{M-1}{M} \quad (10)$$

for every anti-aliasing filter H with finitely many pass-band intervals.

Sketch of the proof: The inequality is trivial, since H can destroy an arbitrarily large amount of information. We sketch the proof for H being piecewise constant with passband intervals having rational endpoints, i.e., being integer multiples of $1/L$ with L sufficiently large. As before, we apply a filterbank decomposition of $\mathbf{X}$, this time with LM channels. The cascade of H and the filters of the filterbank is either identical to zero or to one. Using polyphase decomposition, it can be shown that M-fold downsampling amounts to adding M bands, some of which are set to zero by the filter. Summation is a memoryless operation, hence the information loss depends on the information dimension of the random sums. Since the information dimension of a scalar cannot exceed one, the bound is obtained. ■

The reason for this seemingly counter-intuitive result is that, without a specific signal model, the amount of information is not necessarily proportional to the amount of energy by which it is represented: There is no reason to prefer a specific frequency band over another. This in some sense parallels our result on the relative information loss in principal components analysis (PCA), where we showed that PCA cannot reduce the amount of information being lost in reducing the dimensionality of the data [12].

IV. Relevant Information Loss: Gaussian Case

To remove the counter-intuitivity of the previous section, we adapt the signal model: Let $\mathbf{X}$ be a noisy observation of a signal process $\mathbf{S}$, i.e., $X_n = S_n + N_n$, where $\mathbf{S}$ and $\mathbf{N}$ are independent, jointly stationary Gaussian processes with smooth PSDs $S_S(\mathrm{e}^{\jmath\theta})$ and $S_N(\mathrm{e}^{\jmath\theta})$, respectively, and which satisfy Assumption 1. The information loss rate relevant w.r.t. $\mathbf{S}$ is given by

$$\bar{L}_{\mathbf{S}^{(M)}}(\mathbf{X}^{(M)} \to \mathbf{Y}) := \bar{I}(\mathbf{S}^{(M)}; \mathbf{X}^{(M)}) - \bar{I}(\mathbf{S}^{(M)}; \mathbf{Y}) \quad (11)$$

and measures how much of the information $\mathbf{X}$ conveys about $\mathbf{S}$ is lost in each time step due to downsampling.

While in the general case the filter which minimizes $\bar{L}_{\mathbf{S}^{(M)}}(\mathbf{X}^{(M)} \to \mathbf{Y})$ is hard to find, for this particular signal model the solution is surprisingly intuitive:

Theorem 2. *Let $\mathbf{S}$ and $\mathbf{N}$ be jointly stationary Gaussian processes with smooth PSDs $S_S(\mathrm{e}^{\jmath\theta})$ and $S_N(\mathrm{e}^{\jmath\theta})$ and which satisfy Assumption 1. Let $X_n = S_n + N_n$. Then, the M-aliasing-free energy compaction filter for $S_S(\mathrm{e}^{\jmath\theta})/S_N(\mathrm{e}^{\jmath\theta})$ minimizes the relevant information loss rate in the decimation system depicted in Fig. 1.*

The energy compaction filter for a given PSD can be constructed easily: The M-fold downsampled PSD consists of M aliased components; for each frequency point $\theta \in [-\pi/M, \pi/M]$, at least one of them is maximal. The passbands of the energy compaction filter correspond to exactly these maximal components [2], [3]. See also (12) below.

In particular, since for white Gaussian noise $\mathbf{N}$ the energy compaction filter for $S_S(\mathrm{e}^{\jmath\theta})/S_N(\mathrm{e}^{\jmath\theta})$ coincides with the energy compaction filter for $S_S(\mathrm{e}^{\jmath\theta})$, the filter that lets most of the signal's energy pass aliasing-free is also optimal in terms of information.

Sketch of the proof: Instead of minimizing $\bar{L}_{\mathbf{S}^{(M)}}(\mathbf{X}^{(M)} \to \mathbf{Y})$ we maximize

$$\bar{I}(\tilde{\mathbf{S}}^{(M)}; \mathbf{Y}) = \lim_{n\to\infty} \frac{1}{n}\left(h(Y_1^n) - h(Y_1^n|\tilde{S}_1^{nM})\right).$$

with $\tilde{\mathbf{S}}$ being $\mathbf{S}$ filtered by H. But $h(Y_1^n) = h(\tilde{X}_M, \dots, \tilde{X}_{nM})$ and $h(Y_1^n|\tilde{S}_1^{nM}) = h(\tilde{N}_M, \dots, \tilde{N}_{nM})$, where $\tilde{\mathbf{N}}$ is the noise process filtered by H. By Gaussianity, the mutual information rate reads

$$\bar{I}(\tilde{\mathbf{S}}^{(M)}; \mathbf{Y}) = \frac{1}{4\pi}\int_{-\pi}^{\pi} \ln\left(1 + \frac{\sum_{k=0}^{M-1} \frac{S_S(\mathrm{e}^{\jmath\theta_k})}{S_N(\mathrm{e}^{\jmath\theta_k})} H_k(\theta)}{\sum_{k=0}^{M-1} H_k(\theta)}\right) d\theta$$

where $\theta_k := \frac{\theta - 2k\pi}{M}$ and $H_k(\theta) := S_N(\mathrm{e}^{\jmath\theta_k})|H(\mathrm{e}^{\jmath\theta_k})|^2$. Maximizing the integral is done by maximizing the fraction inside the logarithm, which is, essentially, a weighted average of the ratios $S_S(\mathrm{e}^{\jmath\theta_k})/S_N(\mathrm{e}^{\jmath\theta_k})$. The maximum is obtained if

$$H_l(\theta) = \begin{cases} 1, & \text{for smallest } l \text{ s.t. } \forall k: \frac{S_S(\mathrm{e}^{\jmath\theta_l})}{S_N(\mathrm{e}^{\jmath\theta_l})} \geq \frac{S_S(\mathrm{e}^{\jmath\theta_k})}{S_N(\mathrm{e}^{\jmath\theta_k})} \\ 0, & \text{else} \end{cases} \quad (12)$$

i.e., if H is related to the energy compaction filter via $H_k(\theta) = S_N(\mathrm{e}^{\jmath\theta_k})|H(\mathrm{e}^{\jmath\theta_k})|^2$. Since $|\bar{h}(\mathbf{N})| < \infty$, a filter H' with $|H'(\mathrm{e}^{\jmath\theta})|^2 = 1/S_N(\mathrm{e}^{\jmath\theta})$ does not change the information content; thus, H can be chosen as the energy compaction filter. ∎

V. Relevant Information Loss: General Case

Although the result for Gaussian processes is interesting due to its closed form, it is of little practical relevance. In many cases, at least the relevant part of $\mathbf{X}$, the signal process $\mathbf{S}$, will be non-Gaussian. We thus drop the restriction that $\mathbf{S}$ is Gaussian, but we still assume Gaussianity of $\mathbf{N}$.

One can expect that in this more general case a closed-form solution for H will not be available. However, *assuming* that $\mathbf{S}$ is Gaussian, yields an upper bound on the information rate $\bar{I}(\mathbf{S}^{(M)}; \mathbf{Y})$. It can also be shown that the Gaussian assumption provides an upper bound on the relevant information loss rate. To this end, we employ the approach of Plumbley [8], who showed that, with a specific signal model, PCA can be justified from an information-theoretic perspective (cf. also [14]).

Theorem 3. *Let H be stable and causal, let $\mathbf{S}$ and $\mathbf{N}$ be jointly stationary and satisfy Assumption 1, and let $X_n = S_n + N_n$. $\mathbf{N}$ is Gaussian, and $\mathbf{S}_G$ is Gaussian with the same PSD as $\mathbf{S}$. Let $X_{G,n} = S_{G,n} + N_n$, and let $\mathbf{Y}_G$ be the corresponding output processes of the decimation system, respectively. Then,*

$$\bar{L}_{\mathbf{S}^{(M)}}(\mathbf{X}^{(M)} \to \mathbf{Y}) \leq \bar{L}_{\mathbf{S}_G^{(M)}}(\mathbf{X}_G^{(M)} \to \mathbf{Y}_G). \quad (13)$$

Sketch of the proof: We start from

$$\bar{L}_{\mathbf{S}^{(M)}}(\mathbf{X}^{(M)} \to \mathbf{Y}) = \lim_{n\to\infty} \frac{1}{n}\left(h(\tilde{X}_1^{nM}) - h(\tilde{N}_1^{nM}) - h(Y_1^n) + h(Y_1^n|\tilde{S}_1^{nM})\right)$$

where we exploited the fact that the filter does not change the information content of the process. We then apply $h(Y_1^n|\tilde{S}_1^{nM}) = h(\tilde{N}_M, \dots, \tilde{N}_{nM})$ and $h(Y_1^n) = h(\tilde{X}_M, \dots, \tilde{X}_{nM})$. Then, the only term depending on non-Gaussian RVs is the conditional differential entropy

$$h(\tilde{X}_1^{M-1}, \dots, \tilde{X}_{(n-1)M+1}^{nM-1}|\tilde{X}_M, \dots, \tilde{X}_{nM})$$

which is positive in above equation for the relevant information loss rate. This differential entropy is upper bounded by the one of Gaussian RVs $(X_G)_1^{nM}$ with the same first and second moments. The bound is achieved by replacing $\mathbf{S}$ by $\mathbf{S}_G$. ∎

A consequence of this theorem is that filter design by energetic considerations, i.e., by considering the PSDs of the signal only, has performance guarantees also in information-theoretic terms. One has to consider, though, that the filter H

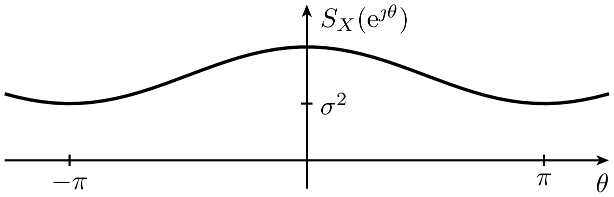

Fig. 2. Power spectral density of $\mathbf{X}$

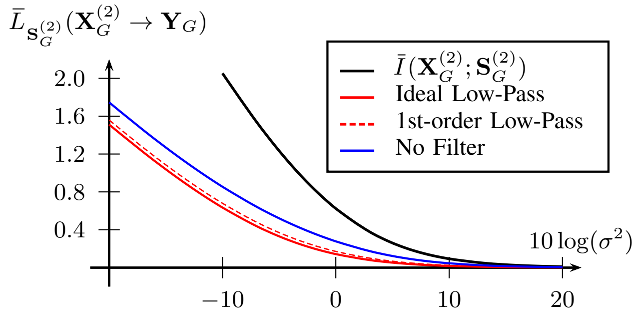

Fig. 3. Upper bounds on the relevant information loss rate in nats as a function of the noise variance σ^2 for various filter options.

optimal in the sense of the upper bound might not coincide with the filter optimal w.r.t. $\bar{L}_{\mathbf{S}^{(M)}}(\mathbf{X}^{(M)} \to \mathbf{Y})$.

VI. Example

We now illustrate our results with an example: Let the PSD of $\mathbf{S}$ be given by $S_S(\mathrm{e}^{\jmath\theta}) = 1 + \cos\theta$ and let $\mathbf{N}$ be independent white Gaussian noise with variance σ^2, i.e., $S_N(\mathrm{e}^{\jmath\theta}) = \sigma^2$. The PSD of $\mathbf{X}$ is depicted in Fig. 2. We consider downsampling by a factor of $M = 2$. Were $\mathbf{S}$ Gaussian too, the optimal filter would be an ideal low-pass filter with cut-off frequency $\pi/2$ (cf. Theorem 2).

If we assume that $\mathbf{S}$ is non-Gaussian, Theorem 3 allows us to design a finite-order filter which minimizes an upper bound on the relevant information loss rate. In particular, it can be shown that among all first-order FIR filters with impulse response $h[n] = \delta[n] + c\delta[n-1]$, the filter with $c = 1$ minimizes the Gaussian bound.

Fig. 3 shows the upper bound on the relevant information loss rate as a function of the noise variance σ^2 for the ideal low-pass filter and the optimal first-order FIR filter compared to the case where no filter is used. In addition, the available information $\bar{I}(\mathbf{X}_G^{(2)};\mathbf{S}_G^{(2)}) = 2\bar{I}(\mathbf{X}_G;\mathbf{S}_G)$ is plotted, which decreases with increasing noise variance. Indeed, filtering can reduce the relevant information loss rate compared to omitting the filter. This is in stark contrast with the results of Section III, in which we showed that the relative information loss rate equals $1/2$ regardless of the filter. The reason is that in Section III we did not have a signal model in mind, treating every bit of information equally. As soon as one knows which aspect of a stochastic process is relevant, one can successfully apply signal processing methods to retrieve as much information as possible (or to remove as much of the irrelevant information as possible, cf. [14]).

Interestingly, as Fig. 3 shows, the improvement of a first-order FIR filter over direct downsampling is significant. Using low-order filters is beneficial also from a computational perspective: To the best of our knowledge, the optimization problem does not permit a closed-form solution for the filter coefficients in general. Thus, numerical procedures will benefit from the fact that the number of coefficients can be kept small. Moreover, while the optimal first-order FIR filter is independent of the noise variance σ^2, numerical calculations suggest that the optimal second-order FIR with impulse response $h[n] = \delta[n] + c\delta[n-1] + \delta[n-2]$ has a coefficient c depending on σ^2.

VII. Conclusion

In this work we analyzed the information loss in a decimation system as a function of its constituting anti-aliasing filter. In particular, we showed that without a signal model in mind, anti-aliasing filtering is futile since it cannot reduce the information loss even if ideal filters are permitted. The situation changes for a simple signal-plus-Gaussian-noise model, where the information loss w.r.t. the signal process can be reduced by properly choosing the filter. As a direct consequence, we concluded that filter design based on second-order statistics of the process can be justified from an information-theoretic perspective.

References

[1] A. V. Oppenheim and R. W. Schafer, *Discrete-Time Signal Processing*, 3rd ed. Upper Saddle River, NJ: Pearson Higher Ed., 2010.

[2] M. Unser, "On the optimality of ideal filters for pyramid and wavelet signal approximation," *IEEE Trans. Signal Process.*, vol. 41, no. 12, pp. 3591–3596, Dec. 1993.

[3] M. Tsatsanis and G. Giannakis, "Principal component filter banks for optimal multiresolution analysis," *IEEE Trans. Signal Process.*, vol. 43, no. 8, pp. 1766–1777, Aug. 1995.

[4] J. C. Principe, *Information Theoretic Learning: Renyi's Entropy and Kernel Perspectives*, ser. Information Science and Statistics. New York, NY: Springer, 2010.

[5] J. I. Galdos and D. E. Gustafson, "Information and distortion in reduced-order filter design," *IEEE Trans. Inf. Theory*, vol. 23, no. 2, pp. 183–194, Mar. 1977.

[6] N. Al-Dhahir and J. M. Cioffi, "Block transmission over dispersive channels: transmit filter optimization and realization, and MMSE-DFE receiver performance," *IEEE Trans. Inf. Theory*, vol. 42, no. 1, pp. 137–160, Jan. 1996.

[7] A. Scaglione, S. Barbarossa, and G. B. Giannakis, "Filterbank transceivers optimizing information rate in block transmissions over dispersive channels," *IEEE Trans. Inf. Theory*, vol. 45, no. 3, pp. 1019–1032, Apr. 1999.

[8] M. Plumbley, "Information theory and unsupervised neural networks," Cambridge University Engineering Department, Tech. Rep. CUED/F-INFENG/TR. 78, 1991.

[9] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. Hoboken, NJ: Wiley Interscience, 2006.

[10] A. Rényi, "On the dimension and entropy of probability distributions," *Acta Mathematica Hungarica*, vol. 10, no. 1-2, pp. 193–215, Mar. 1959.

[11] R. M. Gray, *Entropy and Information Theory*. New York, NY: Springer, 1990.

[12] B. C. Geiger and G. Kubin, "Relative information loss in the PCA," in *Proc. IEEE Information Theory Workshop (ITW)*, Lausanne, Sep. 2012, pp. 562–566, extended version available: `arXiv:1204.0429 [cs.IT]`.

[13] ——, "On the rate of information loss in memoryless systems," Apr. 2013, `arXiv:1304.5057 [cs.IT]`.

[14] ——, "Signal enhancement as minimization of relevant information loss," in *Proc. ITG Conf. on Systems, Communication and Coding*, Munich, Jan. 2013, pp. 1–6, extended version available: `arXiv:1205.6935 [cs.IT]`.