

# On the information loss in memoryless systems: the multivariate case

**Conference Paper**

**Author(s):**

Geiger, Bernhard C.; Kubin, Gernot

**Publication date:**

2012

**Permanent link:**

<https://doi.org/10.3929/ethz-a-007050390>

**Rights / license:**

[In Copyright - Non-Commercial Use Permitted](#)

# On the Information Loss in Memoryless Systems: The Multivariate Case

Bernhard C. Geiger\*, Gernot Kubin\*

\*Signal Processing and Speech Communication Laboratory, Graz University of Technology, Austria  
{geiger,gernot.kubin}@tugraz.at

**Abstract**—In this work we give a concise definition of information loss from a system-theoretic point of view. Based on this definition, we analyze the information loss in memoryless input-output systems subject to a continuous-valued input. For a certain class of multiple-input, multiple-output systems the information loss is quantified. An interpretation of this loss is accompanied by upper bounds which are simple to evaluate.

Finally, a class of systems is identified for which the information loss is necessarily infinite. Quantizers and limiters are shown to belong to this class.

## I. INTRODUCTION

In the XXXI. Shannon lecture Han argued that information theory links information-theoretic quantities, such as entropy and mutual information, to operational quantities such as source, channel, capacity, and error probability [1]. In this work we try to make a new link to an operational quantity not mentioned by Han: information loss. Information can be lost, on the one hand, in erasures or due to superposition of noise as it is known from communication theory. Dating back to Shannon [2] this loss is linked to the conditional entropy of the input given the output, at least in discrete-amplitude, memoryless settings. On the other hand, as stated by the data processing inequality (DPI, [3]), information can be lost in deterministic, noiseless systems. It is this kind of loss that we will treat in this work, and we will show that it makes sense to link it to the same information-theoretic quantity.

The information loss in input-output systems is very sparsely covered in the literature. Aside from the DPI for discrete random variables (RV) and static systems, some results are available for jointly stationary stochastic processes [4]. Yet, all these results just state that *information is lost*, without quantifying this loss. Only in [5] the information lost by collapsing states of a discrete-valued stochastic process is quantified as the difference between the entropy rates at the input and the output of the memoryless system.

Conversely, energy loss in input-output systems has been deeply analyzed, leading to meaningful definitions of transfer functions and notions of passivity, stability, and losslessness. Essentially, it is our aim to develop a system theory not from an energetic, but from an information-theoretic point of view. So far we analyzed the information loss of discrete-valued stationary stochastic processes in finite-dimensional dynamical input-output systems [6], where we proposed an upper bound on the information loss and identified a class of information-preserving systems (the information-theoretic

counterpart to lossless systems). In [7] the information loss of continuous RVs in memoryless systems was quantified and bounded in a preliminary way. In this work, extending [7], we analyze the information loss for static multiple-input, multiple-output systems which are subject to a continuous input RV. Unlike in our previous work, we permit functions which lose an infinite amount of information and present the according conditions. Aside from that we provide a link between information loss and differential entropy, a quantity which is not invariant under changes of variables. The next steps towards an information-centered system theory are the analysis of discrete-time dynamical systems with continuous-valued stationary input processes and a treatment of information loss in multirate systems.

In the remainder of this paper we give a mathematically concise definition of information loss (Section II). After restricting the class of systems in Section III, in Section IV we provide exact results for information loss together with simple bounds, and establish a link to differential entropies. Finally, in Section V we show under which conditions the information loss becomes infinite. An extended version of this paper, accompanied by several examples illustrating the theoretical results, is available in [8].

## II. A DEFINITION OF INFORMATION LOSS

When talking about the information loss induced by processing of signals, it is of prime importance to accompany this discussion by a well-based definition of information loss going beyond, but without lacking, intuition. We try to meet this objective with the following

**Definition 1.** Let  $X$  be an RV<sup>1</sup> on the sample space  $\mathcal{X}$ , and let  $Y$  be obtained by transforming  $X$ . We define the information loss induced by this transform as

$$L(X \rightarrow Y) = \sup_{\mathcal{P}} \left( I(\hat{X}; X) - I(\hat{X}; Y) \right) \quad (1)$$

where the supremum is over all partitions  $\mathcal{P}$  of  $\mathcal{X}$ , and where  $\hat{X}$  is obtained by quantizing  $X$  according to the partition  $\mathcal{P}$  (see Fig. 1).

This Definition is motivated by the data processing inequality (cf. [3]), which states that the expression under the supremum is always non-negative: Information loss is the worst-case reduction of information about  $\hat{X}$  induced by

<sup>1</sup>Note that  $X$  and all other involved RVs need not be scalar-valued.

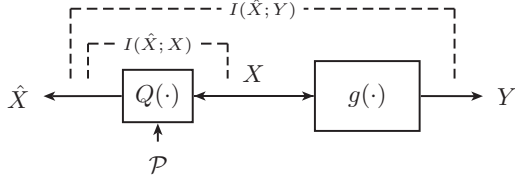


Fig. 1. Model for computing the information loss of a memoryless input-output system  $g$ .  $Q$  is a quantizer with partition  $\mathcal{P}$ .

transforming  $X$ . We now try to shed a little more light on Definition 1 in the following

**Theorem 1.** *The information loss of Definition 1 is given by the conditional entropy of the input given the output, i.e.,*

$$L(X \rightarrow Y) = H(X|Y). \quad (2)$$

*Proof:* We start by noticing that

$$I(\hat{X}; X) - I(\hat{X}; Y) = H(\hat{X}|Y) \leq H(X|Y) \quad (3)$$

by the definition of mutual information and since both  $\hat{X}$  and  $Y$  are functions of  $X$ . The inequality in (3) is due to data processing ( $\hat{X}$  is a function of  $X$ ). We now show that in the supremum over all partitions equality can be achieved.

To this end, observe that among all partitions of the sample space of  $X$  there is a sequence  $\{\mathcal{P}_n\}$  of increasingly fine partitions<sup>2</sup> such that

$$\lim_{n \rightarrow \infty} \hat{X}_n = X \quad (4)$$

where  $\hat{X}_n$  is the quantization of  $X$  induced by partition  $\mathcal{P}_n$ . By the axioms of entropy (e.g., [9, Ch. 14]),  $H(\hat{X}_n|Y)$  is an increasing sequence in  $n$  with limit  $H(X|Y)$ . Thus, equality in (3) is achieved, which completes the proof. ■

This Theorem shows that the supremum in Definition 1 is achieved for  $\hat{X} \equiv X$ , i.e., when we compute the difference between the *self-information* of the input and the information the output of the system contains about its input. In addition to that, the Theorem suggests a natural way to measure the information loss via measuring mutual informations, as it is depicted in Fig. 1. As we will see later (cf. Theorem 3), the considered partition does not even have to be infinitely fine.

### III. PROBLEM STATEMENT

Let  $\mathbf{X} = [X_1, X_2, \dots, X_N]$  be an  $N$ -dimensional RV with a probability measure  $P_{\mathbf{X}}$  absolutely continuous w.r.t. the Lebesgue measure  $\mu$  ( $P_{\mathbf{X}} \ll \mu$ ). We require  $P_{\mathbf{X}}$  to be concentrated on  $\mathcal{X} \subseteq \mathbb{R}^N$ . This RV, which possesses a unique probability density function (PDF)  $f_{\mathbf{X}}$ , is the input to the following multivariate, vector-valued function:

**Definition 2.** Let  $\mathbf{g}: \mathcal{X} \rightarrow \mathcal{Y}$ ,  $\mathcal{X}, \mathcal{Y} \subseteq \mathbb{R}^N$ , be a surjective, Borel-measurable function defined in a piecewise manner:

$$\mathbf{g}(\mathbf{x}) = \begin{cases} \mathbf{g}_1(\mathbf{x}), & \text{if } \mathbf{x} \in \mathcal{X}_1 \\ \mathbf{g}_2(\mathbf{x}), & \text{if } \mathbf{x} \in \mathcal{X}_2 \\ \vdots & \end{cases} \quad (5)$$

<sup>2</sup>i.e.,  $\mathcal{P}_{n+1}$  is a refinement of  $\mathcal{P}_n$ .

where  $\mathbf{x} = [x_1, x_2, \dots, x_N]$  and  $\mathbf{g}_i: \mathcal{X}_i \rightarrow \mathcal{Y}_i$  bijectively. Furthermore, let the Jacobian matrix  $\mathcal{J}_{\mathbf{g}}(\cdot)$  exist on the closures of  $\mathcal{X}_i$ . In addition to that, we require the Jacobian determinant,  $|\det \mathcal{J}_{\mathbf{g}}(\cdot)|$ , to be non-zero  $P_{\mathbf{X}}$ -almost everywhere.

In accordance with previous work [7] the  $\mathcal{X}_i$  are disjoint sets of positive  $P_{\mathbf{X}}$ -measure which unite to  $\mathcal{X}$ , i.e.,  $\bigcup_i \mathcal{X}_i = \mathcal{X}$  and  $\mathcal{X}_i \cap \mathcal{X}_j = \emptyset$  if  $i \neq j$ . Clearly, also the  $\mathcal{Y}_i$  unite to  $\mathcal{Y}$ , but need not be disjoint. This definition ensures that the preimage  $\mathbf{g}^{-1}[\mathbf{y}]$  of each element  $\mathbf{y} \in \mathcal{Y}$  is a countable set.

Using the method of transformation [9, pp. 244] one obtains the PDF of the  $N$ -dimensional output RV  $\mathbf{Y} = [Y_1, Y_2, \dots, Y_N]$  as

$$f_{\mathbf{Y}}(\mathbf{y}) = \sum_{\mathbf{x}_i \in \mathbf{g}^{-1}[\mathbf{y}]} \frac{f_{\mathbf{X}}(\mathbf{x}_i)}{|\det \mathcal{J}_{\mathbf{g}}(\mathbf{x}_i)|} \quad (6)$$

where the sum is over all elements of the preimage. Note that since  $\mathbf{Y}$  possesses a density, the corresponding probability measure  $P_{\mathbf{Y}}$  is also absolutely continuous w.r.t. the Lebesgue measure.

### IV. MAIN RESULTS

We now state our main results:

**Theorem 2.** *The information loss induced by a function  $\mathbf{g}$  satisfying Definition 2 is given as*

$$H(\mathbf{X}|\mathbf{Y}) = \int_{\mathcal{X}} f_{\mathbf{X}}(\mathbf{x}) \log \left( \frac{\sum_{\mathbf{x}_i \in \mathbf{g}^{-1}[\mathbf{g}(\mathbf{x})]} \frac{f_{\mathbf{X}}(\mathbf{x}_i)}{|\det \mathcal{J}_{\mathbf{g}}(\mathbf{x}_i)|}}{\frac{f_{\mathbf{X}}(\mathbf{x})}{|\det \mathcal{J}_{\mathbf{g}}(\mathbf{x})|}} \right) d\mathbf{x}. \quad (7)$$

The proof of this Theorem can be found in the Appendix of [8] and, in a modified version for univariate functions, in [7]. Note that for univariate functions the Jacobian determinant is replaced by the derivative of the function.

**Corollary 1.** *The information loss induced by a function  $\mathbf{g}$  satisfying Definition 2 is given as*

$$H(\mathbf{X}|\mathbf{Y}) = h(\mathbf{X}) - h(\mathbf{Y}) + \mathbb{E} \{ \log |\det \mathcal{J}_{\mathbf{g}}(\mathbf{X})| \} \quad (8)$$

*Proof:* The proof is obtained by recognizing the PDF of  $\mathbf{Y}$  inside the logarithm in (7) and by splitting the logarithm. ■

This result is particularly interesting because it provides a link between information loss and differential entropies already anticipated in [9, pp. 660]. There, it was claimed that

$$h(\mathbf{Y}) \leq h(\mathbf{X}) + \mathbb{E} \{ \log |\det \mathcal{J}_{\mathbf{g}}(\mathbf{X})| \} \quad (9)$$

where equality holds iff  $\mathbf{g}$  is bijective. While (9) is actually another version of the DPI, Corollary 1 quantifies how much information is lost by processing. In addition to that, a very similar expression denoted as *folding entropy* has been presented in [10], although in a completely different setting analyzing the entropy production of *autonomous* dynamical systems.

We now introduce a discrete RV  $W$  which depends on the set  $\mathcal{X}_i$  from which  $\mathbf{X}$  was taken. In other words, we have  $W =$

$i$  iff  $\mathbf{x} \in \mathcal{X}_i$ . One can interpret this RV as being generated by a vector quantization of  $\mathbf{X}$  with a partition  $\mathcal{P} = \{\mathcal{X}_i\}$ . With this new RV we can state

**Theorem 3.** *The information loss is identical to the uncertainty about the set  $\mathcal{X}_i$  from which the input was taken, i.e.,*

$$H(\mathbf{X}|\mathbf{Y}) = H(W|\mathbf{Y}). \quad (10)$$

The proof closely follows the proof provided in [7] and thus is omitted. However, this equivalence suggests a way of measuring information loss by means of proper quantization: Since  $H(W|\mathbf{Y}) = I(W;\mathbf{X}) - I(W;\mathbf{Y})$  the loss can be determined by measuring mutual informations, which in this case are always finite (or, at least, bounded by  $H(W)$ ).

The interpretation derived from Theorem 3 allows us now to provide upper bounds on the information loss:

**Theorem 4.** *The information loss is upper bounded by*

$$H(\mathbf{X}|\mathbf{Y}) \leq \int_{\mathcal{Y}} f_{\mathbf{Y}}(\mathbf{y}) \log \text{card}(\mathbf{g}^{-1}[\mathbf{y}]) d\mathbf{y} \quad (11)$$

$$\leq \log \left( \sum_i \int_{\mathcal{Y}_i} f_{\mathbf{Y}}(\mathbf{y}) d\mathbf{y} \right) \quad (12)$$

$$\leq \max_{\mathbf{y}} \log \text{card}(\mathbf{g}^{-1}[\mathbf{y}]) \quad (13)$$

where  $\text{card}(B)$  is the cardinality of the set  $B$ .

*Proof:* We give here only a sketch of the proof: The first inequality results from bounding  $H(W|\mathbf{Y} = \mathbf{y})$  by the entropy of a uniform distribution on the preimage of  $\mathbf{y}$ . Jensen's inequality yields the second line of the Theorem. The coarsest bound is obtained by replacing the cardinality of the preimage by its maximal value. ■

In this Theorem, we bounded the information loss given a certain output by the cardinality of the preimage. While the first bound considers the fact that the cardinality may actually depend on the output itself, the last bound incorporates the maximum cardinality only. In cases where the function from Definition 2 is defined not on a countable but on a finite number of subdomains, this finite number can act as an upper bound (cf. [7]). Another straightforward upper bound, which is independent from the bounds in Theorem 4 is obtained from Theorem 3 by removing conditioning:

$$H(\mathbf{X}|\mathbf{Y}) \leq H(W) = - \sum_i p_i \log p_i \quad (14)$$

where  $p_i = P_{\mathbf{X}}(\mathcal{X}_i) = \int_{\mathcal{X}_i} f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x}$ .

A further implication of introducing this discrete RV  $W$  is that it allows us to perform investigations about reconstructing the input from the output. Currently, a Fano-type inequality bounding the reconstruction error by the information loss is under development. In addition to that, new upper bounds on the information loss related to the reconstruction error of optimal (in the *maximum a posteriori* sense) and of simpler, sub-optimal estimators are analyzed.

## V. FUNCTIONS WITH INFINITE INFORMATION LOSS

We now drop the requirement in Definition 2 that the  $\mathbf{g}$  possesses partial inverses to analyze a wider class of Borel-measurable functions  $\mathbf{g}: \mathcal{X} \rightarrow \mathcal{Y}$ . We keep the requirement that  $P_{\mathbf{X}} \ll \mu$  and thus  $\mathbf{X}$  possesses a density  $f_{\mathbf{X}}$  (positive on  $\mathcal{X}$  and zero elsewhere). We maintain

**Theorem 5.** *Let  $\mathbf{g}: \mathcal{X} \rightarrow \mathcal{Y}$  be a Borel-measurable function and let the continuous RV  $\mathbf{X}$  be the input to this function. If there exists a set  $B \subseteq \mathcal{Y}$  of positive  $P_{\mathbf{Y}}$ -measure such that the preimage  $\mathbf{g}^{-1}[\mathbf{y}]$  is uncountable for every  $\mathbf{y} \in B$ , then the information loss is infinite.*

*Proof:* With  $B \subseteq \mathcal{Y}$  we notice that

$$H(\mathbf{X}|\mathbf{Y}) \geq \int_B H(\mathbf{X}|\mathbf{Y} = \mathbf{y}) dP_{\mathbf{Y}}(\mathbf{y}) \quad (15)$$

where the integral is now written as Lebesgue integral, since  $P_{\mathbf{Y}}$  now not necessarily possesses a density.

Since on  $B$  the preimage of every element is uncountable, we obtain with [4] and the references therein  $H(\mathbf{X}|\mathbf{Y} = \mathbf{y}) = \infty$  for all  $\mathbf{y} \in B$ , and, thus,  $H(\mathbf{X}|\mathbf{Y}) = \infty$ . ■

Note that the requirement of  $B$  being a set of positive  $P_{\mathbf{Y}}$ -measure cannot be dropped, as one of the examples provided in [8] illustrates. We immediately obtain the following

**Corollary 2.** *Let  $\mathbf{g}: \mathcal{X} \rightarrow \mathcal{Y}$  be a Borel-measurable function and let the continuous RV  $\mathbf{X}$  be the input to this function. If the probability measure of the output,  $P_{\mathbf{Y}}$ , possesses a non-vanishing discrete component, the information loss is infinite.*

*Proof:* According to the Lebesgue-Radon-Nikodym theorem [11, pp. 121] every measure can be decomposed in a component absolutely continuous w.r.t.  $\mu$  and a component singular to  $\mu$ . The latter part can further be decomposed into a singular continuous and a discrete part, where the latter places positive  $P_{\mathbf{Y}}$ -mass on points. Let  $\mathbf{y}^*$  be such a point, i.e.,  $P_{\mathbf{Y}}(\mathbf{y}^*) > 0$ . As an immediate consequence,  $P_{\mathbf{X}}(\mathbf{g}^{-1}[\mathbf{y}^*]) > 0$ , which is only possible if  $\mathbf{g}^{-1}[\mathbf{y}^*]$  is uncountable ( $P_{\mathbf{X}} \ll \mu$ ). ■

This result is also in accordance with intuition, as the analysis of a simple quantizer shows: While the entropy of the input RV is infinite ( $I(\hat{\mathbf{X}};\mathbf{X}) \rightarrow \infty$  for  $\hat{\mathbf{X}} \rightarrow \mathbf{X}$ ; cf. [9, pp. 654]), the quantized output can contain only a finite amount of information ( $I(\hat{\mathbf{X}};\mathbf{Y}) \rightarrow H(\mathbf{Y}) < \infty$ ). In addition to that, the preimage of each possible output value  $\mathbf{y}$  is a set of positive  $P_{\mathbf{X}}$ -measure. The loss, as a consequence, is infinite.

While for the quantizer the preimage of each possible output value is a set of positive measure, there certainly are functions for which some outputs have a countable preimage and some whose preimage is a non-null set. An example of such a system is the limiter [9, Ex. 5-4]. For such systems it can be shown that both the information loss  $L(\mathbf{X} \rightarrow \mathbf{Y}) = H(\mathbf{X}|\mathbf{Y})$  and the information transfer  $I(\mathbf{X}; \mathbf{Y})$  are infinite.

Finally, there exist functions  $\mathbf{g}$  for which the preimages of all output values  $\mathbf{y}$  are null sets, but which still fulfill the conditions of Theorem 5. Functions which project  $\mathcal{X}$  on a lower-dimensional subspace of  $\mathbb{R}^N$  fall into that category.

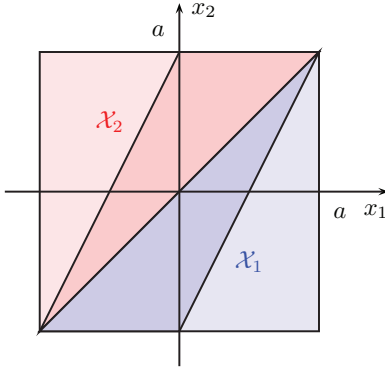


Fig. 2. The partitioned functions  $\mathbf{g}_i$  restricted to the domain of either color are bijective. Furthermore, the overall function  $\mathbf{g}$  is bijective if restricted to areas with light shading.

### VI. EXAMPLE: A TWO-DIMENSIONAL TRANSFORM WITH FINITE INFORMATION LOSS

In this Section we illustrate our theoretical results with the help of an example. The logarithm is taken to base 2. Let  $\mathbf{X}$  be uniformly distributed on the square  $\mathcal{X} = [-a, a] \times [-a, a]$ , or equivalently, the two constituting RVs  $X_1$  and  $X_2$  are independent and uniformly distributed on  $[-a, a]$ . In other words, while  $f_{\mathbf{X}}(\mathbf{x}) = 1/4a^2$  for all  $\mathbf{x} \in \mathcal{X}$ , we have  $f_X(x_i) = 1/2a$  for  $x_i \in [-a, a]$  and  $i = 1, 2$ .

We consider a function  $\mathbf{g}$  performing the mapping:

$$Y_1 = X_1 \quad (16)$$

$$Y_2 = |X_1 - X_2| \quad (17)$$

The corresponding Jacobian matrix is a triangular matrix

$$\mathcal{J}_{\mathbf{g}}(\mathbf{x}) = \begin{bmatrix} 1 & 0 \\ \text{sgn}(x_1 - x_2) & \text{sgn}(x_2 - x_1) \end{bmatrix} \quad (18)$$

where  $\text{sgn}(\cdot)$  is the sign-function. From this immediately follows that the magnitude of the determinant of the Jacobian matrix is unity for all possible values of  $\mathbf{X}$ , i.e.,  $|\det \mathcal{J}_{\mathbf{g}}(\mathbf{x})| = 1$  for all  $\mathbf{x} \in \mathcal{X}$ . The subsets of  $\mathcal{X}$  on which the partitioned functions  $\mathbf{g}_i$  are bijective are no intervals in this case; they are the triangular halves of the square induced by  $x_1 = x_2$  (see Fig. 2)

$$\mathcal{X}_1 = \{[x_1, x_2] \in \mathcal{X} : x_1 > x_2\} \quad (19)$$

$$\mathcal{X}_2 = \{[x_1, x_2] \in \mathcal{X} : x_1 \leq x_2\}. \quad (20)$$

The preimage of  $\mathbf{g}(\mathbf{x})$  is, in any case,

$$\{[x_1, x_2], [x_1, 2x_1 - x_2]\} \cap \mathcal{X}. \quad (21)$$

The transform  $\mathbf{g}$  is bijective whenever  $[x_1, 2x_1 - x_2] \notin \mathcal{X}$ , i.e., if  $|2x_1 - x_2| > a$ .

With the PDF of  $\mathbf{X}$  and of its components we obtain for the information loss

$$H(\mathbf{X}|\mathbf{Y}) = \int_{-a}^a \int_{-a}^a \frac{1}{4a^2} \log \left( \frac{\frac{1}{2a} + f_X(2x_1 - x_2)}{\frac{1}{2a}} \right) dx_1 dx_2 \quad (22)$$

which is non-zero only if  $-a \leq 2x_1 - x_2 \leq a$  (numerator and denominator cancel otherwise; no loss occurs in the bijective domain of the function). As a consequence,

$$H(\mathbf{X}|\mathbf{Y}) = \int_{-a}^a \int_{\frac{x_2-a}{2}}^{\frac{x_2+a}{2}} \frac{\log 2}{4a^2} dx_1 dx_2 = \int_{-a}^a \frac{1}{4a} dx_2 = \frac{1}{2}. \quad (23)$$

The information loss is identical to a half bit. This is intuitive by looking at Fig. 2, where it can be seen that any information loss occurs only on one half of the domain  $\mathcal{X}$  (shaded in stronger colors). By destroying the sign information, in this area the information loss is equal to one bit.

### VII. CONCLUSION

In this work, we proposed a mathematically concise definition of information loss for the purpose of establishing a system theory from an information-theoretic point of view. For a certain class of multivariate, vector-valued functions and continuous input variables this information loss was quantified, and the result is accompanied by convenient upper bounds. We further showed a connection between information loss and the differential entropies of the input and output variables.

Finally, a class of systems has been identified for which the information loss is infinite. Vector-quantizers and limiters belong to that class, but also functions which project the input space onto a space of lower dimensionality.

### REFERENCES

- [1] T. S. Han, "Musing upon information theory," XXXI Shannon Lecture, 2010, presented at IEEE Int. Sym. on Information Theory (ISIT).
- [2] C. E. Shannon, "A mathematical theory of communication," *Bell Systems Technical Journal*, vol. 27, pp. 379–423, 623–656, Oct. 1948.
- [3] R. M. Gray, *Entropy and Information Theory*. New York, NY: Springer, 1990.
- [4] M. S. Pinsker, *Information and Information Stability of Random Variables and Processes*. San Francisco: Holden Day, 1964.
- [5] S. Watanabe and C. T. Abraham, "Loss and recovery of information by coarse observation of stochastic chain," *Information and Control*, vol. 3, no. 3, pp. 248–278, Sep. 1960.
- [6] B. C. Geiger and G. Kubin, "Some results on the information loss in dynamical systems," in *Proc. IEEE Int. Sym. Wireless Communication Systems (ISWSC)*, Aachen, Nov. 2011, pp. 794–798, extended version available: [arXiv:1106.2404](https://arxiv.org/abs/1106.2404) [cs.IT].
- [7] B. C. Geiger, C. Feldbauer, and G. Kubin, "Information loss in static nonlinearities," in *Proc. IEEE Int. Sym. Wireless Communication Systems (ISWSC)*, Aachen, Nov. 2011, pp. 799–803, extended version available: [arXiv:1102.4794](https://arxiv.org/abs/1102.4794) [cs.IT].
- [8] B. C. Geiger and G. Kubin, "On the information loss in memoryless systems: The multivariate case," Sep. 2011, [arXiv:1109.4856](https://arxiv.org/abs/1109.4856) [cs.IT].
- [9] A. Papoulis and U. S. Pillai, *Probability, Random Variables and Stochastic Processes*, 4th ed. New York, NY: McGraw Hill, 2002.
- [10] D. Ruelle, "Positivity of entropy production in nonequilibrium statistical mechanics," *J.Stat.Phys.*, vol. 85, pp. 1–23, 1996.
- [11] W. Rudin, *Real and Complex Analysis*, 3rd ed. New York, NY: McGraw-Hill, 1987.