Differential Entropy of the Conditional Expectation Under Additive Gaussian Noise

Arda Atalik, Alper Köse, and Michael Gastpar, *Fellow, IEEE*

***Abstract*—The conditional mean is a fundamental and important quantity whose applications include the theories of estimation and rate-distortion. It is also notoriously difficult to work with. This paper establishes novel bounds on the differential entropy of the conditional mean in the case of finite-variance input signals and additive Gaussian noise. The main result is a new lower bound in terms of the differential entropies of the input signal and the noisy observation. The main results are also extended to the vector Gaussian channel and to the natural exponential family. Various other properties such as upper bounds, asymptotics, Taylor series expansion, and connection to Fisher Information are obtained. Two applications of the lower bound in the remote-source coding and CEO problem are discussed.**

***Index Terms*—Differential entropy, conditional mean estimator, Gaussian noise, exponential family, remote source coding problem, CEO problem.**

I. Introduction and Motivation

THE conditional expectation is a quantity of fundamental interest with myriad applications. It is an intuitively pleasing estimator of an underlying signal, given a noisy observation. For example, it is well known that subject to a mean-squared error (MSE) criterion, the conditional expectation is the optimal estimator. More generally, the conditional expectation is a sufficient statistic for a large class of problems. Applications of conditional expectation thus include detection and estimation [1], [2], [3], multiterminal hypothesis testing [4], [5], information bottleneck [6], [7], privacy funnel [8], [9], equalization in communication systems, and statistical physics.

While the conditional mean has a straightforward formula, it is usually impossible to express it in closed form. This hampers the exploration of its key properties, such as its moments or its differential entropy. The goal of the present investigation is to shed light on the *differential entropy* of the conditional mean $h(\mathsf{E}[X|Y])$. This quantity has a multitude of applications. We present and develop one of these applications, which concerns problems of noisy (or *remote*) source coding, where the encoding device does not get to observe the source X of interest, but rather only a noisy version Y of the source.

Manuscript received 2 December 2021; revised 18 July 2022 and 16 September 2022; accepted 18 September 2022. Date of publication 3 October 2022; date of current version 17 October 2022. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Caroline Chaux. This work was supported by the Swiss National Science Foundation under Grant 200364. This work was presented in part at the 2021 IEEE Information Theory Workshop, Japan [DOI: 10.1109/ITW48936.2021.9611440].

Arda Atalik is with the Department of Electrical and Electronics Engineering, Bilkent University, 06800 Ankara, Turkey (e-mail: arda.atalik@bilkent.edu.tr).

Alper Köse is with the Electrical and Electronics Engineering Department, Bogazici University, 34470 Istanbul, Turkey (e-mail: alper.kose@boun.edu.tr).

Michael Gastpar is with the School of Computer and Communication Sciences, École Polytechnique Fédérale de Lausanne, 1015 Lausanne, Switzerland (e-mail: michael.gastpar@epfl.ch).

Digital Object Identifier 10.1109/TSP.2022.3211403

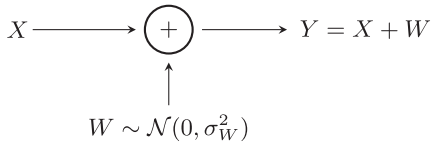

Fig. 1. The additive Gaussian noise observation model.

In this paper, our primary focus concerns the additive Gaussian noise model

$$Y = X + W \tag{1}$$

where W is a zero-mean Gaussian random variable of variance σ_W^2, independent of the signal X, and where X has an arbitrary distribution. This is illustrated pictorially in Fig. 1. Estimation in Gaussian noise has been a central topic at the intersection of estimation and information theory for the last sixty years. Consequently, in this special case, the conditional expectation $\mathsf{E}[X|Y]$ has been studied in a wealth of works and many key tools are known. A first such tool that will be of importance to our derivations is *Tweedie's Formula* [10], which connects the conditional mean to the score function. Following up on this important formula, the *Hatsell-Nolte Identity* was discovered [11], which relates the conditional mean and variance. Yet another key formula is *Brown's Identity* [12], which shows a connection between the minimum mean-square error (MMSE) and Fisher Information.

It is also interesting to compare and contrast $h(\mathsf{E}[X|Y])$ to a more common quantity, namely, the conditional entropy of X given Y, usually denoted as $h(X|Y)$. While a thorough study is outside of the scope of the present paper, we may observe that the two quantities behave quite differently. Starting from the Gaussian noise model of (1), one may consider the limiting cases as σ_W^2 tends to either of its extremal values. First, as σ_W^2 tends to zero, we can observe that $h(\mathsf{E}[X|Y])$ simply tends to $h(X)$ while $h(X|Y)$ diverges to $-\infty$. On the other end of the scale, as σ_W^2 tends to ∞, we observe that $h(\mathsf{E}[X|Y])$ diverges to $-\infty$ while $h(X|Y)$ tends to $h(X)$. In the case where X is Gaussian, too, one can derive closed-form expressions for both

quantities, as follows:

$$\begin{aligned} h(X|Y) &= h(X) + h(Y|X) - h(Y) \\ &= h(X) + h(W) - h(Y) \\ &= \frac{1}{2}\log\left(2\pi e \frac{\sigma_X^2\,\sigma_W^2}{\sigma_X^2+\sigma_W^2}\right) \qquad (2) \\ h(\mathsf{E}\left[X|Y\right]) &= h\left(\frac{\sigma_X^2}{\sigma_X^2+\sigma_W^2}Y + \frac{\sigma_W^2}{\sigma_X^2+\sigma_W^2}\mu_X\right) \\ &= h(Y) + \log\frac{\sigma_X^2}{\sigma_X^2+\sigma_W^2} \\ &= \frac{1}{2}\log\left(2\pi e \frac{\sigma_X^4}{\sigma_X^2+\sigma_W^2}\right). \qquad (3) \end{aligned}$$

Finally, we would like to emphasize the following *dual* entropy power inequalities, which are valid for arbitrary inputs [13, Sec. 8.6]:

$$N(X|Y) \le \mathsf{E}\left[\mathsf{Var}\left(X|Y\right)\right] \qquad (4a)$$

$$N(\mathsf{E}\left[X|Y\right]) \le \mathsf{Var}\left(\mathsf{E}\left[X|Y\right]\right), \qquad (4b)$$

and satisfied with equalities for Gaussian X.

A. Contributions

- We provide a new lower bound which relates the differential entropy of the conditional mean to that of input and output in Fig. 1:

$$h(\mathsf{E}\left[X|Y\right]) \ge 2h(X) - h(Y). \qquad (5)$$

- We also derive novel upper bounds on $h(\mathsf{E}[X|Y])$ and show several applications of the new bounds, most notably in the context of so-called remote source coding.
- We extend our bound to the case where the noisy observation process is not characterized by adding Gaussian noise, but by a general exponential family distribution.
- We also extend our bounds to the case of vector signals.

B. Related Work

Differential entropy plays a key role in estimation theory as it measures the uncertainty in a given probabilistic scenario and it has been gaining importance in many different fields such as data science and machine learning, biology and neuroscience, economics, and other experimental sciences as highlighted in [14].

Mean-squared error (MSE) is one of the most commonly used error metrics in estimation theory and plays a significant role in many real-life applications ranging from signal detection in communications to regression problems in machine learning. Typically, in an estimation problem, the objective aims at minimizing a cost function, thus it is highly significant to find the estimator which leads to the *minimum mean-square error* (MMSE). As widely known, the conditional mean is optimal in the MSE sense, i.e., the MMSE estimate of X observing $Y = y$ is found by the conditional mean of X given $Y = y$. As a result, conditional mean is widely used in signal processing applications, e.g., signal detection [15], noise cancellation [16], frequency estimation [17], and target tracking [18].

In [19], a fundamental derivative identity connecting the mutual information and MMSE is discovered and in [20], it has been further explored. Note that by contrast to the model in Fig. 1, the model in [19] fixes the noise variance to unity and instead (but equivalently) lets the noisy observation be $Y = \sqrt{\gamma}X + W$. Our results are more naturally expressed in terms of the model in Fig. 1. Properties of MMSE such as monotonicity, convexity, and infinite differentiability as a function of snr have been shown in [21], while its functional properties as a function of input-output distribution have been analyzed in [22], [23]. Recently, in [24], [25], the authors have focused on the derivatives of the conditional mean with respect to the observation, and many previously known identities in the literature have been recovered.

C. Outline

- In Section II, we provide a new lower bound which relates the differential entropy of the conditional mean to that of input and output in Fig. 1.
- In Section III, we study further properties of the differential entropy of the conditional mean such as upper bounds, Taylor series expansion, low-and-high input variance asymptotics, and connection to Fisher Information. For different input distributions, the bounds on $h(\mathsf{E}[X|Y])$ are illustrated.
- In Section IV, two applications of the lower bound in the remote source coding [26] are investigated.
- In Section V, the main result is extended to the natural exponential families.
- In Section VI, the lower and upper bounds are extended to the vector additive Gaussian noise model.

D. Notation

We use uppercase letters $X,\ Y$ to denote random variables and lowercase letters $x,\ y$ to denote their realizations. In slight abuse of notation, we use boldface uppercase letters $\mathbf{X},\ \mathbf{Y}$ to denote both random vectors as well as (deterministic) matrices. The distinction will be clear from context. Given a square-integrable, absolutely continuous random variable X with a probability density function (PDF) $p_X(x)$, its mean $\mathsf{E}[X]$ is denoted as μ_X, its variance $\mathsf{Var}(X)$ is denoted as σ_X^2, and its differential entropy is

$$h(X) = -\int p_X(x)\,\log p_X(x)\,dx \qquad (6)$$

where $\log$ denotes the natural logarithm. The entropy power of X is

$$N(X) = e^{2h(X)}/(2\pi e), \qquad (7)$$

the conditional entropy power of X given Y is $N(X|Y) = e^{2h(X|Y)}/(2\pi e)$, and the Fisher information of X is $J(X) = \int p_X(x)\,(\frac{d}{dx}\,\log p_X(x))^2 dx$. To denote X is a Gaussian random variable with mean μ_X and variance σ_X^2, we use the notation $X \sim \mathcal{N}(\mu_X, \sigma_X^2)$.

For two random variables (X, Y) with a density function $p_{X,Y}(x, y)$, the mutual information is $I(X;Y) = h(Y) - h(Y|X) = h(X) - h(X|Y)$. Denote the conditional expectation and variance of X given Y as $\mathsf{E}[X|Y]$ and $\mathsf{Var}(X|Y)$, respectively, and the corresponding mean-square error as

$$\mathsf{mmse}\left(X|Y\right) = \mathsf{E}\left[\mathsf{Var}\left(X|Y\right)\right] = \mathsf{E}\left[\left(X - \mathsf{E}\left[X|Y\right]\right)^2\right]. \quad (8)$$

It is conceptually straightforward to give an expression for the probability density function (PDF) of the conditional mean $\mathsf{E}[X|Y]$, and thus, for the differential entropy of $\mathsf{E}[X|Y]$. Some simplifications can be applied in the special case of additive Gaussian noise considered in the present paper, see e.g. [24], but in general, the resulting expressions are intractable. Therefore, in our work, we will not leverage the probability density function of $\mathsf{E}[X|Y]$ directly.

E. The Case of Gaussian Inputs

In this section, we briefly review the well-known formulas for the case where the underlying source X is Gaussian. That is, $X \sim \mathcal{N}(\mu_X, \sigma_X^2)$. For this case, all the quantities defined in Section I-D can be calculated analytically. The probability density function of $X|Y = y$ can be calculated by Bayes' rule to find

$$\begin{aligned} f_{X|Y}(x|y) &= \frac{p_{Y|X}(y|x)\, p_X(x)}{p_Y(y)} \\ &= \frac{p_W(y-x)\, p_X(x)}{p_Y(y)} \\ &= \frac{\exp\left(-\frac{\left(x-\left(\frac{\sigma_X^2}{\sigma_X^2+\sigma_W^2}y+\frac{\sigma_W^2}{\sigma_X^2+\sigma_W^2}\mu_X\right)\right)^2}{2\left(\frac{\sigma_X\sigma_W}{\sqrt{\sigma_X^2+\sigma_W^2}}\right)^2}\right)}{\frac{\sigma_X\sigma_W}{\sqrt{\sigma_X^2+\sigma_W^2}}\sqrt{2\pi}}. \end{aligned} \quad (9)$$

That is, conditioned on $Y = y$, X has Gaussian distribution with mean $\frac{\sigma_X^2}{\sigma_X^2+\sigma_W^2}y + \frac{\sigma_W^2}{\sigma_X^2+\sigma_W^2}\mu_X$, and variance $\frac{\sigma_X^2\sigma_W^2}{\sigma_X^2+\sigma_W^2}$. Thus,

$$\mathsf{E}\left[X|Y\right] = \frac{\sigma_X^2}{\sigma_X^2+\sigma_W^2}Y + \frac{\sigma_W^2}{\sigma_X^2+\sigma_W^2}\mu_X, \quad (10)$$

$$\mathsf{Var}\left(X|Y\right) = \frac{\sigma_X^2\sigma_W^2}{\sigma_X^2+\sigma_W^2} = \mathsf{mmse}\left(X|Y\right). \quad (11)$$

From (10) and (11), it is observed that $\mathsf{E}[X|Y]$ is Gaussian with mean μ_X, and variance $\frac{\sigma_X^4}{\sigma_X^2+\sigma_W^2}$.

$$\mathsf{E}\left[\mathsf{E}\left[X|Y\right]\right] = \mathsf{E}\left[X\right] = \mu_X, \quad (12)$$

$$\begin{aligned} \mathsf{Var}\left(\mathsf{E}\left[X|Y\right]\right) &= \mathsf{Var}\left(\frac{\sigma_X^2}{\sigma_X^2+\sigma_W^2}Y + \frac{\sigma_W^2}{\sigma_X^2+\sigma_W^2}\mu_X\right) \\ &= \mathsf{Var}\left(\frac{\sigma_X^2}{\sigma_X^2+\sigma_W^2}Y\right) \\ &= \left(\frac{\sigma_X^2}{\sigma_X^2+\sigma_W^2}\right)^2\left(\mathsf{Var}(X)+\mathsf{Var}(W)\right) \\ &= \frac{\sigma_X^4}{\sigma_X^2+\sigma_W^2}. \end{aligned} \quad (13)$$

The Fisher information of $\mathsf{E}[X|Y]$ is the reciprocal of its variance since it is Gaussian, i.e.,

$$J\left(\mathsf{E}\left[X|Y\right]\right) = \frac{\sigma_X^2+\sigma_W^2}{\sigma_X^4}, \quad (14)$$

and the main inequality (5) is satisfied with equality:

$$h(\mathsf{E}\left[X|Y\right]) = \frac{1}{2}\log\left(2\pi e\frac{\sigma_X^4}{\sigma_X^2+\sigma_W^2}\right) \quad (15)$$

$$= 2h(X) - h(Y). \quad (16)$$

II. Main Results

The differential entropy of the conditional expectation shows up as a lower/upper bound in certain multi-terminal information theory problems as shown in [26]. Thus, it is useful to derive tight bounds on $h(\mathsf{E}[X|Y])$ to obtain further insights. In general, deriving upper bounds to the differential entropy is not difficult using the maximum entropy argument, while lower bounds are less trivial and they might require extra assumptions on the input distribution such as log-concavity. Using the main result of [27], one obtains $h(\mathsf{E}[X|Y]) \geq \frac{1}{2}\log(4\mathsf{Var}(\mathsf{E}[X|Y])) \geq \frac{1}{2}\log(4\frac{\sigma_X^4}{\sigma_X^2+\sigma_W^2})$ as long as the density of $\mathsf{E}[X|Y]$ is log-concave, which might be tedious to check.

For the case of additive Gaussian noise, we present a new lower bound which combines the maximum entropy argument in the conditional setting with the identity (17), and is applicable regardless of the input distribution. Our theorem is based on the following lemma, which relates the differential entropy of the conditional expectation to that of the output.

Lemma 1: (Calculation of $h(\mathsf{E}[X|Y])$): For the model given in (1) with $\sigma_W^2 > 0$, the differential entropy of the conditional mean can be written as

$$h(\mathsf{E}\left[X|Y\right]) = h(Y) + \mathsf{E}\left[\log\left(\frac{1}{\sigma_W^2}\mathsf{Var}\left(X|Y\right)\right)\right]. \quad (17)$$

Proof: This lemma follows by careful application of several known tools, including Tweedie's formula [10] and the Hatsell-Nolte identity [11]. A full proof is provided in Appendix B. ■

We can now leverage this lemma to the lower bound on the differential entropy of the conditional mean.

Theorem 1: Let X be an arbitrary continuous random variable with finite variance and $Y = X + W$, where W is a zero mean Gaussian with variance σ_W^2 and independent of X. Then,

$$h(\mathsf{E}\left[X|Y\right]) \geq 2h(X) - h(Y). \quad (18)$$

Furthermore, equality is achieved if and only if X is Gaussian.

Proof: The proof of this theorem starts from Lemma 1 and lower bounds the second summand on the right hand side of (17) by a maximum entropy argument. Specifically, observe that $h(X|Y=y) \leq \frac{1}{2}\log(2\pi e\,\mathsf{Var}(X|Y=y))$ for every y. Taking the expectation of both sides, one obtains

$$\mathsf{E}\left[\log\left(\mathsf{Var}\left(X|Y\right)\right)\right] \geq 2\left(h(X|Y) - \frac{1}{2}\log\left(2\pi e\right)\right) \quad (19)$$

$$= 2\left(h(X|Y) - h(W) + \log \sigma_W\right)$$
$$= 2\left(h(X) - h(Y) + \log \sigma_W\right) \quad (20)$$

where (20) follows by the definition of mutual information

$$I(X;Y) = h(Y) - h(W) = h(X) - h(X|Y). \quad (21)$$

Combining (17) and (20), the $\log \sigma_W$ term cancels out and the desired result follows. The uniqueness of the equality condition follows from the uniqueness of the Gaussian distribution as a maximum entropy distribution under a variance constraint. ■

To the best of our knowledge, (18) is not known in the literature outside of the special equality case when the input is Gaussian [26]. As immediate applications, (18) can be used to compare the tightness of rate-distortion lower bounds in remote source coding under Gaussian noise, which is explained in Section IV-A; and novel rate loss bounds can be derived as explained in Section IV-B.

Remark 1: Our main result, Theorem 1, can be expressed in terms of entropy powers as

$$N(\mathsf{E}[X|Y])\, N(Y) \geq (N(X))^2. \quad (22a)$$

There is a pleasing *duality* with a known result about variances,

$$\mathsf{Var}(\mathsf{E}[X|Y])\ \sigma_Y^2 \geq (\sigma_X^2)^2. \quad (22b)$$

This last inequality follows directly from the fact that $\mathsf{mmse}(X|Y)$ cannot be larger than the mean-squared error of the best *linear* estimator,

$$\mathsf{mmse}(X|Y) \leq \frac{\sigma_X^2 \sigma_W^2}{\sigma_Y^2}, \quad (23)$$

combined with the fact that

$$\mathsf{mmse}(X|Y) + \mathsf{Var}(\mathsf{E}[X|Y]) = \sigma_X^2, \quad (24)$$

which is the law of total variance.

III. Further Properties of the Differential Entropy and Entropy Power of Conditional Mean

A. Upper Bounds

By the concavity of the logarithm, one can use Jensen's Inequality to derive an upper bound.

Lemma 2: (Upper Bounds of Differential Entropy): For the model given in 1, the differential entropy of the conditional mean can be upper bounded as

$$h(\mathsf{E}[X|Y]) = h(Y) + \mathsf{E}[\log \mathsf{Var}(X|Y)] - \log \sigma_W^2 \quad (25)$$
$$\leq h(Y) + \log \mathsf{E}[\mathsf{Var}(X|Y)] - \log \sigma_W^2 \quad (26)$$
$$= h(Y) + \log \mathsf{mmse}(X|Y) - \log \sigma_W^2 \quad (27)$$
$$\leq h(Y) + \log\left(\frac{\sigma_X^2}{\sigma_X^2 + \sigma_W^2}\right) \quad (28)$$
$$\leq \frac{1}{2}\log\left(2\pi e \frac{\sigma_X^4}{\sigma_X^2 + \sigma_W^2}\right). \quad (29)$$

Proof: To prove this lemma, we note that (26) follows from the Jensen's Inequality, (27) is by definition of mmse, and (28) and (29) follow from the maximization of mmse and differential entropy, respectively. ■

Each inequality in Lemma 2 is satisfied with equality if and only if the input X is Gaussian. That is, for Gaussian inputs, $\mathsf{Var}(X|Y)$ is constant almost surely, and mmse and differential entropy are maximized [21]. From (29), it is evident that the maximum differential entropy of the conditional mean is achieved when the input X is Gaussian even though it minimizes the variance of $\mathsf{E}[X|Y]$, which we record in the following corollary:

Corollary 1: (Maximum Entropy): For the model of Fig. 1, we have

$$\max_{p_X} h(\mathsf{E}[X|Y]) = \frac{1}{2}\log\left(2\pi e \frac{\sigma_X^4}{\sigma_X^2 + \sigma_W^2}\right), \quad (30)$$

and the maximum is achieved when $X \sim \mathcal{N}(\mu_X, \sigma_X^2)$.

In the sequel, we will sometimes find it convenient to rewrite our upper and lower bounds on the differential entropy of the conditional mean in terms of its entropy power. Namely, (18), (27), and (29) can be equivalently expressed as

$$\frac{N^2(X)}{N(Y)} \leq N(\mathsf{E}[X|Y]) \leq N(Y)\frac{(\mathsf{mmse}(X|Y))^2}{\sigma_W^4} \leq \frac{\sigma_X^4}{\sigma_X^2 + \sigma_W^2}. \quad (31)$$

Remark 2: The upper bound (31) can also be written as

$$N(\mathsf{E}[X|Y]) \leq \sigma_X^2 - \sigma_W^2 \frac{\sigma_X^2}{\sigma_Y^2}. \quad (32)$$

There is also a *dual* result about variances, which follows from [26, Eqn. 16] and the law of total variance.

$$\mathsf{Var}(\mathsf{E}[X|Y]) \leq \sigma_X^2 - \sigma_W^2 \frac{N(X)}{N(Y)}. \quad (33)$$

B. Comparison of the Lower and Upper Bounds

When the input X is Gaussian, the upper and lower bounds are satisfied with equality. On the other extreme, as the input distribution approaches to a discrete random variable, the gap between $2h(X) - h(Y)$ and $h(\mathsf{E}[X|Y])$ increases since the input differential entropy $h(X)$ approaches to $-\infty$. To illustrate, suppose the input is a zero mean Gaussian mixture random variable with two components centered around -1, 1. That is,

$$X = 2B - 1 + \widetilde{X} \quad (34)$$

where B and $\widetilde{X}$ are Bernoulli($\frac{1}{2}$) and Gaussian random variables with mean 0 and variance $\sigma_X^2 - 1$, respectively; and B, $\widetilde{X}$, and W[1] are independent. We refer to this example as Gaussian mixture input. In Fig. 2, the bounds are illustrated for the case when the input X follows Gaussian mixture, exponential and uniform distribution. Note that $h(Y)$, $h(\mathsf{E}[X|Y])$, and mmse are calculated numerically for exponential and uniform inputs as there are no closed-form expressions. For the Gaussian mixture input, all quantities are calculated numerically. We also calculated the bounds for other input distributions, e.g., Laplace and Triangular, and observed that there are no visible gaps. Without loss of generality, the noise variance is set to unity in both simulations.

[1] W is the zero-mean Gaussian noise in (1).

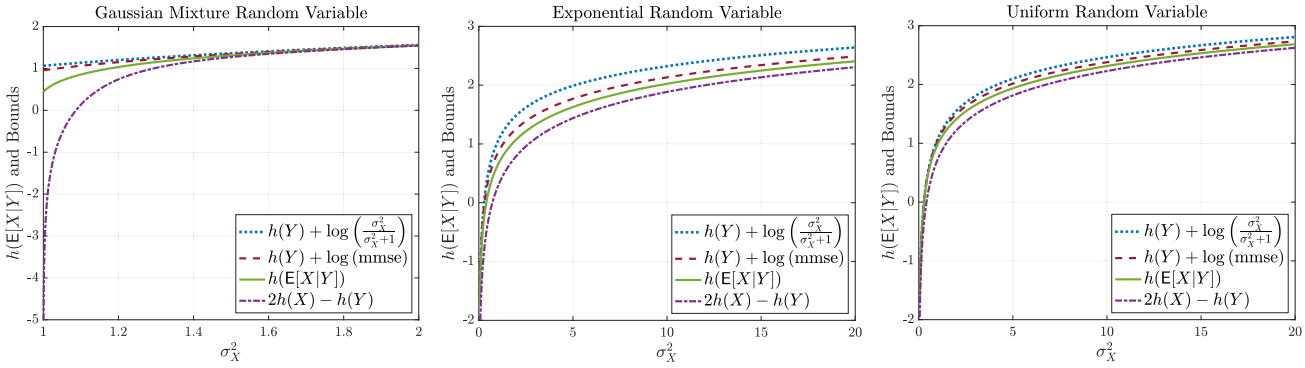

Fig. 2. $h(\mathsf{E}[X|Y])$ and the bounds (18), (27), (28) for Gaussian mixture, exponential and uniform inputs.

Observe from (2) that the bounds are tighter for input distributions close to Gaussian in terms of Kullback-Leibler (KL) divergence, and both the upper and lower bounds become tighter as the input variance increases.

C. Input Variance Asymptotics

We fix the noise variance σ_W^2, and study low and high σ_X^2 asymptotics of $h(\mathsf{E}[X|Y])$ and $N(\mathsf{E}[X|Y])$.

Lemma 3 (Low σ_X^2 Asymptotics): As $\sigma_X^2 \to 0^+$

$$h(\mathsf{E}[X|Y]) \sim \log \sigma_X^2 \tag{35a}$$

$$N(\mathsf{E}[X|Y]) \sim \sigma_X^4 \tag{35b}$$

i.e., $\lim_{\sigma_X^2 \to 0^+} \frac{h(\mathsf{E}[X|Y])}{\log \sigma_X^2} = 1$ and $\lim_{\sigma_X^2 \to 0^+} \frac{N(\mathsf{E}[X|Y])}{\sigma_X^4} = 1$.

Proof: Observe that as σ_X^2 approaches 0^+, since Y converges to $\mu_X + W$ almost surely, the first term in the right hand side of (17) approaches to the differential entropy of the noise, i.e., $h(Y) = -\int p_Y(y) \log p_Y(y) dy \to -\int p_W(y) \log p_W(y) dy = h(W)$. For the second term, it is easy to check that $\mathsf{Var}(X|Y)$ converges to $\mathsf{Var}(X) = \sigma_X^2$ almost surely, which implies that $\mathsf{E}[\log \mathsf{Var}(X|Y)] \to \log \sigma_X^2$. Combining these, we obtain (35a), which is equivalent to (35b). ■

Remark 3: (High σ_X^2 Asymptotics): As $\sigma_X^2 \to \infty$

$$h(\mathsf{E}[X|Y]) = \mathcal{O}(\log \sigma_X) \tag{36a}$$

$$N(\mathsf{E}[X|Y]) = \mathcal{O}\left(\sigma_X^2\right). \tag{36b}$$

Proof: Proof directly follows from (29), i.e., $h(\mathsf{E}[X|Y]) \le \frac{1}{2}\log(2\pi e) + \frac{1}{2}\log(\frac{\sigma_X^2}{\sigma_X^2+\sigma_W^2}) + \frac{1}{2}\log \sigma_X^2$ and $N(\mathsf{E}[X|Y]) \le \frac{\sigma_X^4}{\sigma_X^2+\sigma_W^2}$. ■

D. Taylor Series Expansion

In this subsection, we discuss Taylor series approximation to $\mathsf{E}[\log(\mathsf{Var}(X|Y))]$. Assume that all moments of $\mathsf{Var}(X|Y)$ are finite. Denote the k^{th} central moment of $\mathsf{Var}(X|Y)$ by c_k, i.e., $\mathsf{E}[(\mathsf{Var}(X|Y) - \mathsf{E}[\mathsf{Var}(X|Y)])^k] = c_k$ and $\mathsf{mmse}(X|Y)$ by mmse. Since $\log(\cdot)$ is sufficiently differentiable, we have

$$\mathsf{E}[\log \mathsf{Var}(X|Y)] = \log \mathsf{mmse} - \frac{c_2}{2\mathsf{mmse}^2} + \frac{c_3}{3\,\xi_S^3} \tag{37}$$

where ξ satisfies

$$|\xi - \mathsf{E}[\mathsf{Var}(X|Y)]| < |\mathsf{Var}(X|Y) - \mathsf{E}[\mathsf{Var}(X|Y)]|. \tag{38}$$

Using (17), Taylor expansion of $h(\mathsf{E}[X|Y])$ follows.

$$h(\mathsf{E}[X|Y]) = h(Y) + \log\left(\frac{\mathsf{mmse}}{\sigma_W^2}\right) - \frac{c_2}{2\mathsf{mmse}^2} + \frac{c_3}{3\,\xi_S^3}. \tag{39}$$

Observe that neglecting the last two terms in (39), we recover the upper bound in (27), and neglecting only the last term, we obtain an approximation to $h(\mathsf{E}[X|Y])$.

$$h(\mathsf{E}[X|Y]) \approx h(Y) + \log\left(\frac{\mathsf{mmse}}{\sigma_W^2}\right) - \frac{\mathsf{Var}(\mathsf{Var}(X|Y))}{2\mathsf{mmse}^2}. \tag{40}$$

E. Connection to Fisher Information

A lower bound for the Fisher Information of V directly follows from Stam's Inequality [28].

$$J(\mathsf{E}[X|Y]) \ge \frac{1}{N(\mathsf{E}[X|Y])}. \tag{41}$$

We can further lower bound this by applying (31), obtaining

$$J(\mathsf{E}[X|Y]) \ge \frac{1}{N(Y)} \frac{\sigma_W^4}{(\mathsf{mmse}(X|Y))^2} \tag{42}$$

$$\ge \frac{\sigma_X^2 + \sigma_W^2}{\sigma_X^4}, \tag{43}$$

and noting that equality is achieved when the input is Gaussian. This lower bound on $J(\mathsf{E}[X|Y])$ is of interest since it does not involve the differential entropy of $\mathsf{E}[X|Y]$.

F. Comparison With Costa's Entropy Power Inequality

Define $Y_\alpha \triangleq X + \alpha W$, where $\alpha \in [0,1]$, and X, W satisfy the same assumptions as the main model (1). Observe from (20) that the lower bound (18), applied to the modified model Y_α is equivalent to

$$N(Y_\alpha) \ge \alpha^2 \sigma_W^2 N(X) \exp\left(-\mathsf{E}[\log \mathsf{Var}(X|Y_\alpha)]\right). \tag{44}$$

Costa's entropy power inequality (EPI) [29] gives a lower bound on $N(Y_\alpha)$ for $\alpha \in [0,1]$ in terms of $N(X)$ and $N(Y_1)$.

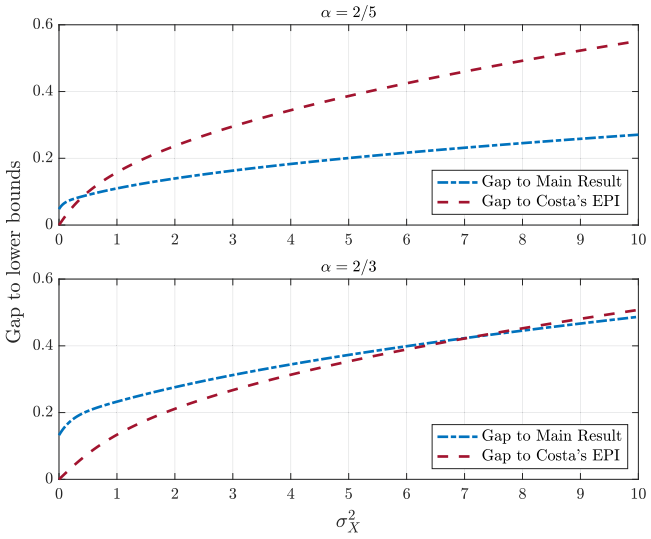

Fig. 3. Comparison of the tightness of lower bounds for $N(Y_\alpha)$: Gap to Main Result and Gap to Costa's EPI refer to (46a) and (46b), respectively.

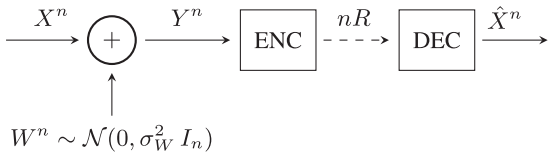

Fig. 4. The AWGN remote source coding problem.

That is,

$$N(Y_\alpha) \geq (1-\alpha^2)N(X) + \alpha^2 N(Y_1). \tag{45}$$

It is tempting to compare (44) with the Costa's EPI as they both provide lower bounds on the entropy power of the output Y_α. In Fig. 3, we compare these bounds when the input X is a uniform random variable with zero mean and variance σ_X^2. Moreover, we set $\sigma_W = 1$ and consider $\alpha \in \{\frac{2}{5}, \frac{2}{3}\}$. The figure illustrates the gap to the main result, which is

$$N(Y_\alpha) - \alpha^2\, N(X)\, \exp\left(-\mathsf{E}\left[\log \mathsf{Var}\left(X|Y_\alpha\right)\right]\right), \tag{46a}$$

and the gap to Costa's EPI, which is

$$N(Y_\alpha) - \left((1-\alpha^2)N(X) + \alpha^2 N(Y_1)\right). \tag{46b}$$

As expected, (44) performs better for small values of α.

IV. Applications

A. Lower Bounds of the Rate Distortion Function in the Remote Source Coding Problem

An important application of Inequality (18) can be found in the remote source coding problem. Specifically, consider the source coding problem illustrated in Fig. 4: An encoder observes the underlying source X subject to additive white Gaussian noise W. The noisy observation is Y and can be encoded using R bits per sample. The decoder produces a reconstruction $\hat{X}$ to within the smallest possible mean-squared error. For a formal problem statement, we refer to [26]. The smallest possible rate to attain a target distortion D is referred to as the *remote rate-distortion function,* denoted as $R_X^R(D)$. For the case where the underlying source X, not necessarily Gaussian, has finite differential entropy, [26] discusses two different lower bounds for the remote rate-distortion function, namely

$$R_X^R(D) \geq \frac{1}{2}\log^+ \frac{N(\mathsf{E}\left[X|Y\right])}{D} + \frac{1}{2}\log^+ \frac{N(Y)}{N(Y) - \frac{N(X)}{D}N(W)}, \tag{47}$$

and

$$R_X^R(D) \geq \frac{1}{2}\log^+ \frac{N(X)}{D} + \frac{1}{2}\log^+ \frac{N(X)}{N(Y) - \frac{N(X)}{D}\sigma_W^2} \tag{48}$$

where $D > \mathsf{E}[(X - \mathsf{E}[X|Y])^2]$ and $\log^+ x = \max\{0, \log x\}$. At the time of their writing, a comparison between the right-hand sides (RHS) of (47) and (48) was not available. We prove that the RHS of (47) is always greater than or equal to the RHS of (48).

Proposition 1: (A comparison of lower bounds for the remote rate-distortion function):

$$\frac{1}{2}\log^+ \frac{N(\mathsf{E}\left[X|Y\right])}{D} + \frac{1}{2}\log^+ \frac{N(Y)}{N(Y) - \frac{N(X)}{D}N(W)} \geq \frac{1}{2}\log^+ \frac{N(X)}{D} + \frac{1}{2}\log^+ \frac{N(X)}{N(Y) - \frac{N(X)}{D}\sigma_W^2}. \tag{49}$$

Furthermore, equality is achieved if and only if X is Gaussian.

Proof: First, note that since W is Gaussian, we have that $N(W) = \sigma_W^2$. Moreover, by the entropy power inequality,

$$N(Y) \geq N(X) + N(W), \tag{50}$$

and clearly, $N(Y) > N(Y) - \frac{N(X)}{D}N(W)$. Combining these observations, we conclude that the second term in the RHS of (47) is always greater than or equal to the second term in (48)

To compare the first terms in the inequalities, observe that if $D \geq N(X) \geq N(\mathsf{E}[X|Y])$, both $\log^+ \frac{N(\mathsf{E}[X|Y])}{D}$ and $\log^+ \frac{N(X)}{D}$ are equal to 0, so the lower bound in (47) is greater. If $N(X) \geq N(\mathsf{E}[X|Y]) > D$, since $N(\mathsf{E}[X|Y])N(Y) \geq (N(X))^2$ from Proposition (18), it is again guaranteed that the lower bound in (47) is greater. Finally, the last possible and a little more complicated case is $N(X) > D \geq N(\mathsf{E}[X|Y])$. In this case, the comparison of the lower bounds seen in (47) and (48) is equivalent to the comparison of $DN(Y)$ and $(N(X))^2$. Taking the logarithm of both sides, we thus have to compare

$$2h(Y) + \log 2\pi e D \underset{(47)}{\overset{(48)}{\lessgtr}} 4h(X). \tag{51}$$

Note that we have assumed $D \geq N(\mathsf{E}[X|Y])$ for this case at the beginning. This is equivalent to the following inequality.

$$\log 2\pi e D \geq 2h(\mathsf{E}\left[X|Y\right]). \tag{52}$$

Combining Theorem 1 (that is, (18)) with (51) and (52), one obtains

$$2h(Y) + \log 2\pi eD \geq 2h(Y) + 2h(\mathsf{E}[X|Y]) \geq 4h(X). \quad (53)$$

Therefore, we conclude that the lower bound in (47) is always greater than or equal to the lower bound in (48). ■

B. Novel Rate-Loss Bounds for the CEO Problem

In this section, we apply Theorem 1 to the so-called CEO problem. In this problem, a single underlying source X is observed by M encoders. Each encoder receives a noisy version of the source X, denoted as $Y_i = X + W_i$, for $i = 1, 2, \ldots, M$. In our consideration, the noises W_i are assumed to be zero-mean Gaussian, independent of each other, and of variance σ_W^2. Each encoder compresses its observation using R_i bits. All M compressed representations are given to a single central decoder whose goal is to produce a reconstruction of the underlying source X to with mean-squared error D. The smallest possible sum-rate required to attain a distortion D is denoted by $R_X^{\mathrm{CEO}}(D)$. We precisely follow the exact problem statement and notation used in [26].

The *rate loss* in the CEO problem denotes the difference between $R_X^{\mathrm{CEO}}(D)$ and the much smaller rate that would be required if all encoders were to cooperate fully, *i.e.,* the rate required by a single encoder having access to all M noisy source observations. Evidently, if the encoders are allowed to cooperate fully, then the problem is exactly the remote rate-distortion problem discussed in Section IV-A above, but with reduced variance σ_W^2/M. We denote the corresponding rate by $R_X^{\mathrm{R}}(D)$, and the rate loss by

$$L(D) \triangleq R_X^{\mathrm{CEO}}(D) - R_X^{\mathrm{R}}(D). \quad (54)$$

In this section, we establish a novel bound on this rate loss.

To develop our results we will use the auxiliary notation

$$Y(M) = \frac{1}{M}\sum_{i=1}^{M} Y_i. \quad (55)$$

A lower bound for $L(D)$ is presented in [26]. For

$$\sigma_X^2 \frac{\sigma_W^2/M}{\sigma_X^2 + \sigma_W^2/M} < D < N(X)\frac{\sigma_W^2/M}{N(Y(M)) - N(X)}, \quad (56)$$

the lower bound on the rate loss establishes that

$$L(D) \geq \frac{M}{2}\log\left(\frac{1}{\frac{N(Y(M))}{N(X)} - \frac{\sigma_W^2}{M}\frac{1}{D}}\right) - \frac{1}{2}\log\left(\frac{\sigma_X^2}{N(X)}\frac{1}{1 - \frac{\sigma_W^2}{M}\left(\frac{1}{D} - \frac{1}{\sigma_X^2}\right)}\right). \quad (57)$$

For $M \uparrow \infty$ and under some regularity conditions further discussed in Appendix D, the bound becomes, for $0 < D < \frac{1}{J(X)}$

$$L(D) \geq \frac{\sigma_W^2}{2}\left(\frac{1}{D} - J(X)\right) - \frac{1}{2}\log\frac{\sigma_X^2}{N(X)}. \quad (58)$$

The novel bound presented here is an *upper* bound on the rate loss, developed in the following subsections.

1) Cooperation Bound: The first ingredient of the novel upper bound on the rate loss is an improved lower bound on $R_X^{\mathrm{R}}(D)$. To this end, we will utilize both $N(\mathsf{E}[X|Y(M)])$ and $\mathsf{mmse}(X|Y(M))$, i.e., for all $D > \mathsf{mmse}(X|Y(M))$

$$R_X^{\mathrm{R}}(D) \geq \frac{1}{2}\log^+\frac{N(\mathsf{E}[X|Y(M)])}{D - \mathsf{mmse}(X|Y(M))}. \quad (59)$$

One can weaken (59) to omit the calculation of $\mathsf{mmse}(X|Y(M))$. In that case, one obtains for all $D > N(X)\,\sigma_W^2/(M\,N(Y(M)))$,

$$R_X^{\mathrm{R}}(D) \geq \frac{1}{2}\log^+\frac{N(\mathsf{E}[X|Y(M)])}{D} + \frac{1}{2}\log^+\frac{M\,N(Y(M))}{M\,N(Y(M)) - \frac{N(X)}{D}N(W)} \quad (60)$$

$$= \begin{cases} \frac{1}{2}\log\frac{N(\mathsf{E}[X|Y(M)])}{D - \frac{\sigma_W^2}{M}\frac{N(X)}{N(Y(M))}} & \text{if } D < N(\mathsf{E}[X|Y(M)]) \\ \frac{1}{2}\log\frac{1}{1 - \frac{\sigma_W^2}{M}\frac{N(X)}{D\,N(Y(M))}} & \text{otherwise} \end{cases} \quad (61)$$

where $Y(M) = X + \sum_{i=1}^{M} W_i/M$. As we have shown in Section IV-A, this bound is tighter than the other lower bound in [26] for any finite[2] M.

2) Novel Rate Loss Upper Bound: In order to upper bound the rate loss $L(D)$, we utilize the upper bound on the CEO sum-rate distortion by Eswaran and Gastpar [26], which states that for $D > \sigma_X^2\sigma_W^2/(M\,\sigma_{Y(M)}^2)$,

$$R_X^{\mathrm{CEO}}(D) \leq \frac{1}{2}\log^+\frac{\sigma_X^2}{D} + \frac{M}{2}\log^+\frac{M\,\sigma_X^2}{M\sigma_{Y(M)}^2 - \frac{\sigma_X^2}{D}\sigma_W^2} \quad (62)$$

$$= \frac{1}{2}\log^+\frac{\sigma_X^2}{D} + \frac{M}{2}\log^+\frac{1}{1 + \frac{\sigma_W^2}{M}\left(\frac{1}{\sigma_X^2} - \frac{1}{D}\right)} \quad (63)$$

$$= \begin{cases} \frac{1}{2}\log\frac{\sigma_X^2}{D} + \frac{M}{2}\log\frac{1}{1 + \frac{\sigma_W^2}{M}\left(\frac{1}{\sigma_X^2} - \frac{1}{D}\right)} & \text{if } D < \sigma_X^2 \\ 0 & \text{otherwise} \end{cases}. \quad (64)$$

Since $N(\mathsf{E}[X|Y(M)]) \leq \mathsf{Var}(\mathsf{E}[X|Y(M)]) \leq \sigma_X^2$ and $\sigma_X^2\sigma_W^2/(M\,\sigma_{Y(M)}^2) \geq N(X)\,N(W)/(M\,N(Y(M)))$, we have two region of interests as we subtract (61) from (64) to obtain the new upper bound.

[2]Observe that as $M \uparrow \infty$, the second term vanishes, and the bound becomes $R_X^{\mathrm{R}}(D) \geq \frac{1}{2}\log^+\frac{N(X)}{D}$ as expected. This is also true for the other lower bound.

Theorem 2: For $D > \sigma_X^2\sigma_W^2/(M\,\sigma_{Y(M)}^2)$, we have the inequality seen in (65).

$$L(D) \leq \begin{cases} \frac{1}{2}\log\left(\frac{\sigma_X^2}{N(\mathsf{E}[X|Y(M)])}\left(1-\frac{\sigma_W^2}{M\,D}\frac{N(X)}{N(Y(M))}\right)\right) \\ \quad +\frac{M}{2}\log\frac{1}{1+\frac{\sigma_W^2}{M}\left(\frac{1}{\sigma_X^2}-\frac{1}{D}\right)} & \text{if } C_1 \\ \frac{1}{2}\log\left(\frac{\sigma_X^2}{D}\left(1-\frac{\sigma_W^2}{M\,D}\frac{N(X)}{N(Y(M))}\right)\right) \\ \quad +\frac{M}{2}\log\frac{1}{1+\frac{\sigma_W^2}{M}\left(\frac{1}{\sigma_X^2}-\frac{1}{D}\right)} & \text{if } C_2 \end{cases} \tag{65}$$

where C_1 is the condition $D < N(\mathsf{E}[X|Y(M)])$ and C_2 is the condition $N(\mathsf{E}[X|Y(M)]) \leq D < \sigma_X^2$.

As the number of agents approaches infinity, (65) simplifies to the following.

Corollary 2: As $M \uparrow \infty$, the upper bound on the loss becomes

$$L(D) \leq \begin{cases} \frac{1}{2}\log\frac{\sigma_X^2}{N(X)} + \frac{1}{2}\sigma_W^2\left(\frac{1}{D}-\frac{1}{\sigma_X^2}\right) & \text{if } D < N(X) \\ \frac{1}{2}\log\frac{\sigma_X^2}{D} + \frac{1}{2}\sigma_W^2\left(\frac{1}{D}-\frac{1}{\sigma_X^2}\right) & \text{if } N(X) \leq D < \sigma_X^2 \end{cases}. \tag{66}$$

It is also possible to use (59) to obtain a tighter upper bound on the rate loss for arbitrary M.

Theorem 3: For $\mathsf{mmse}(X|Y(M)) < D < \mathsf{mmse}(X|Y(M)) + N(\mathsf{E}[X|Y(M)])$, the rate loss is upper bounded as

$$\begin{aligned} L(D) &\leq \frac{1}{2}\log 2\pi e\sigma_X^2 - h(\mathsf{E}\,[X|Y(M)]) \\ &\quad + \frac{1}{2}\log\left(1-\frac{\mathsf{mmse}\,(X|Y(M))}{D}\right) \\ &\quad + \frac{M}{2}\log\frac{1}{1-\frac{\sigma_W^2}{M}\left(\frac{1}{D}-\frac{1}{\sigma_X^2}\right)} \qquad (67) \\ &= \frac{1}{2}\log\frac{\sigma_X^2}{N(\mathsf{E}\,[X|Y(M)])} \\ &\quad + \frac{1}{2}\log\left(1-\frac{\mathsf{mmse}\,(X|Y(M))}{D}\right) \\ &\quad + \frac{M}{2}\log\frac{1}{1-\frac{\sigma_W^2}{M}\left(\frac{1}{D}-\frac{1}{\sigma_X^2}\right)}, \qquad (68) \end{aligned}$$

and for $\mathsf{mmse}(X|Y(M)) + N(\mathsf{E}[X|Y(M)]) < D < \sigma_X^2$, we have

$$L(D) \leq \frac{1}{2}\log\frac{\sigma_X^2}{D} + \frac{M}{2}\log\frac{1}{1-\frac{\sigma_W^2}{M}\left(\frac{1}{D}-\frac{1}{\sigma_X^2}\right)}. \tag{69}$$

Remark 4: Note that (68) is minimized for Gaussian inputs since both $N(\mathsf{E}[X|Y(M)])$ and $\mathsf{mmse}(X|Y(M))$ is maximized in that case. Furthermore, the theorem simplifies to

$$\begin{aligned} L(D) &\leq \frac{1}{2}\log\left(1-\frac{\sigma_W^2}{M}\left(\frac{1}{D}-\frac{1}{\sigma_X^2}\right)\right) \\ &\quad + \frac{M}{2}\log\frac{1}{1-\frac{\sigma_W^2}{M}\left(\frac{1}{D}-\frac{1}{\sigma_X^2}\right)} \qquad (70) \\ &= \frac{M-1}{2}\log\frac{1}{1-\frac{\sigma_W^2}{M}\left(\frac{1}{D}-\frac{1}{\sigma_X^2}\right)} \qquad (71) \end{aligned}$$

for any $\frac{\sigma_W^2\sigma_X^2}{M\sigma_X^2+\sigma_W^2} < D < \sigma_X^2$. The exact loss for the Gaussian input is well-known [26].

$$\begin{aligned} L_{\mathcal{N}}(D) &= \frac{M-1}{2}\log\left(\frac{D}{\frac{\sigma_X^2+\sigma_W^2/M}{\sigma_X^2}D-\frac{\sigma_W^2}{M}}\right) \qquad (72) \\ &= \frac{M-1}{2}\log\frac{1}{1-\frac{\sigma_W^2}{M}\left(\frac{1}{D}-\frac{1}{\sigma_X^2}\right)}. \qquad (73) \end{aligned}$$

Hence, the new upper bound is tight for Gaussian inputs, irrespective of M.

Remark 5:

- As $M \uparrow \infty$, Theorem 2 and 3 yield the same asymptotics, i.e., Corollary 2. Yet, Theorem 3 is guaranteed to be tighter in the finite régime.
- It is important to note that the Gaussian input maximizes the lower bound (57), whereas it minimizes the upper bound (68). Hence, the bounds are tight for the inputs that are close to Gaussian distribution.

3) Previous Rate Loss Upper Bound: We also note that the following upper bound on the rate loss $L(D)$ appears in [30].

$$\begin{aligned} L(D) &\leq \frac{M-1}{2}\log\frac{1}{1-\frac{\sigma_W^2}{M}\left(\frac{1}{D}-\frac{1}{\sigma_X^2}\right)} \qquad (74) \\ &\quad + \frac{1}{2}\log\left(1+\frac{\left(\sigma_X^2-D\right)\left(M\cdot\left(D+2\sqrt{D\cdot\sigma_W^2}\right)+\sigma_W^2\right)}{D\cdot\left(M\sigma_X^2+\sigma_W^2\right)-\sigma_W^2\sigma_X^2}\right) \\ &= \frac{M-1}{2}\log\frac{1}{1-\frac{\sigma_W^2}{M}\left(\frac{1}{D}-\frac{1}{\sigma_X^2}\right)} \qquad (75) \\ &\quad + \frac{1}{2}\log\left(1+\left(\frac{1}{D}-\frac{1}{\sigma_X^2}\right)\left(\frac{D+2\sqrt{D}\sigma_W+\frac{\sigma_W^2}{M}}{1-\frac{\sigma_W^2}{M}\left(\frac{1}{D}-\frac{1}{\sigma_X^2}\right)}\right)\right), \qquad (76) \end{aligned}$$

and as $M \uparrow \infty$, we have

$$\begin{aligned} L(D) &\leq \frac{\sigma_W^2}{2}\left(\frac{1}{D}-\frac{1}{\sigma_X^2}\right) \\ &\quad + \frac{1}{2}\log\left(1+\left(\frac{1}{D}-\frac{1}{\sigma_X^2}\right)\left(D+2\sqrt{D\sigma_W^2}\right)\right). \end{aligned} \tag{77}$$

4) Comparison of the Rate Loss Bounds: Comparing (68) and (76) for any input distributions is tedious. For Gaussian inputs, we proved that the former is tighter than the latter. For other distributions such as Laplace, Exponential, Uniform we present numerical results in Fig. 5, and Fig. 6. Note that the new upper bound is tighter than the general upper bound and its dependence on the number of users M is more prominent at low distortions. For a detailed analysis on the rate loss in the AWGN CEO problem, we refer to [31].

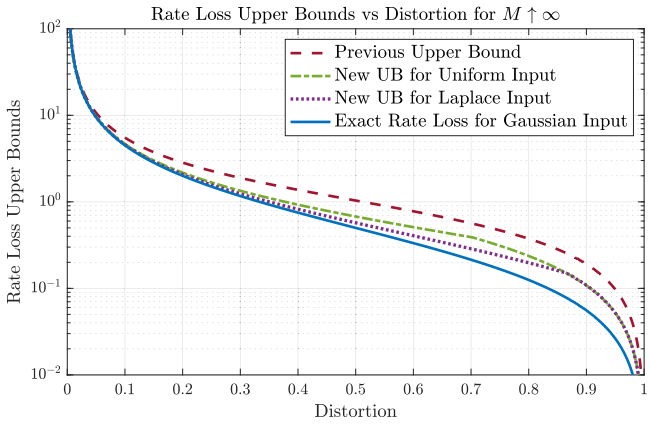

Fig. 5. Comparison of Upper Bounds for $\sigma_X^2 = \sigma_W^2 = 1$.

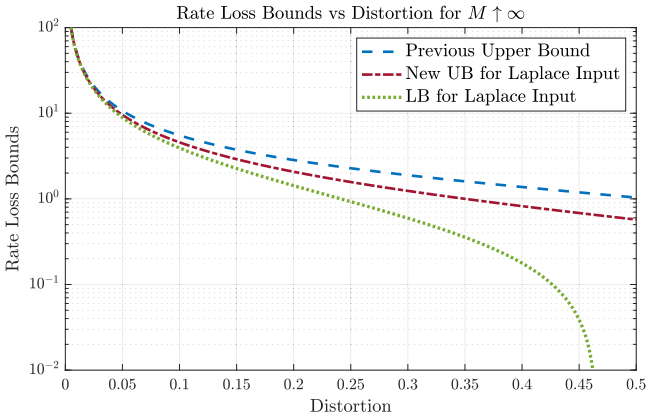

Fig. 6. Comparison of Bounds for $\sigma_X^2 = \sigma_W^2 = 1$.

V. Extension to Exponential Families

We investigate the exponential family generalization of the differential entropy of the conditional expectation. We allow the input to be of any distribution with finite variance and differential entropy, and the output given the input is restricted to the exponential family with the canonical parameter equal to the input.

A. Natural Exponential Family Setup

In this section, we adopt the notation used in [32], [33] and replace the model of Fig. 1 by the following more general model:

$$X \sim q(\cdot) \text{ and } Y \mid X = x \sim p_x(y) = e^{xy - A(x)} p_b(y) \quad (78)$$

where x is the canonical parameter of the family, $A(x)$ is the cumulant generating function (CGF),[3] $p_b(y)$ is the absolutely continuous base measure, and $q(\cdot)$ is the PDF of the absolutely continuous random variable X with known mean and variance. For this model, the following lower bound on the differential entropy of the conditional expectation holds.

[3]Derivatives of CGF yield the cumulants. See [32] for further explanations.

Theorem 4: For the model of (78), we have

$$h(\mathsf{E}[X|Y]) \geq 2h(X) - h(Y) + 2\left(h(Y|X) - \frac{1}{2}\log(2\pi e)\right). \quad (79)$$

Proof: The proof follows by analogy to Theorem 1. By Bayes' Rule, the posterior density of X given $Y = y$ is

$$p(x|y) = p_x(y)q(x)/p(y) \quad (80)$$

where $p(y)$ is the marginal density which is calculated as

$$p(y) = \int p_x(y)q(x)\,dx. \quad (81)$$

Plugging (78) in (80), we obtain

$$p(x|y) = e^{xy - \log\frac{p(y)}{p_b(y)}}\left(q(x)\,e^{-A(x)}\right), \quad (82)$$

which is a natural exponential family with canonical parameter y and CGF $\log\frac{p(y)}{p_b(y)}$. Taking derivatives of the CGF with respect to y, we obtain the natural exponential family generalization of *Tweedie's formula*, provided the differentiation under the integral sign is justified. We discuss these details in Appendix C.

$$\mathsf{E}[X|Y=y] = \frac{d}{dy}\log\left(\frac{p(y)}{p_b(y)}\right) \quad (83)$$

$$\mathsf{Var}(X|Y=y) = \frac{d}{dy}\mathsf{E}[X|Y=y] = \frac{d^2}{dy^2}\log\left(\frac{p(y)}{p_b(y)}\right). \quad (84)$$

Hence, we have the same formula for calculating $h(\mathsf{E}[X|Y])$. Specifically, we can observe the following steps:

$$h(\mathsf{E}[X|Y]) = h(Y) + \mathsf{E}\left[\log\left|\frac{d}{dY}\mathsf{E}[X|Y]\right|\right] \quad (85)$$

$$= h(Y) + \mathsf{E}[\log(\mathsf{Var}(X|Y))] \quad (86)$$

$$\geq h(Y) + 2h(X|Y) - \log(2\pi e) \quad (87)$$

$$= 2h(X) - h(Y) + 2h(Y|X) - \log(2\pi e) \quad (88)$$

where (85) follows immediately from the change of variable formula, (86) follows from (83), and (84). In order to justify (87), we can use the following argument. Since

$$h(X|Y=y) \leq \frac{1}{2}\log(2\pi e\,\mathsf{Var}(X|Y=y)) \quad (89)$$

for every y, we take expectations of both sides to obtain

$$\mathsf{E}[\log(\mathsf{Var}(X|Y))] \geq 2\left(h(X|Y) - \frac{1}{2}\log(2\pi e)\right). \quad (90)$$

Finally, (88) follows from the definition. That is,

$$I(X;Y) = h(X) - h(X|Y) = h(Y) - h(Y|X). \quad (91)$$

■

Remark 6:

- Note that (89) inherently assumes X is supported on $\mathbb{R}$, and therefore, it can be tightened in case a support constraint

is added to the model (78). We refer to [34] for a detailed discussion on this matter.

- From (89), it is evident that the equality is achieved in (79) if and only if $X|Y=y$ is Gaussian. Hence, the bound is tighter for posterior distributions close to Gaussian in terms of KL divergence.

We note that if we specialize the model of (78) to the AWGN model of Fig. 1, then the corrective term in the second line of (79) vanishes and we obtain Theorem 1. In that case, $p_b(y)$ is the Gaussian density with zero mean and σ_W^2 variance and

$$x \triangleq \frac{\mu}{\sigma_W^2},\ A(x) \triangleq \frac{1}{2}\sigma_W^2\, x^2 = \frac{1}{2}\frac{\mu^2}{\sigma_W^2}. \tag{92}$$

Hence, $Y|X=x$ is a Gaussian random variable with mean $\sigma_W^2\, x$ and variance σ_W^2, and the model (78) corresponds to a modified AWGN channel $Y=\sigma_W^2 X+W$. Introducing a change of variable $\widetilde{X} \triangleq \sigma_W^2 X$ gives the desired result, i.e., $h(\mathsf{E}[\widetilde{X}|Y]) \geq 2h(\widetilde{X}) - h(Y)$. In general, however, the corrective term can be positive or negative, which is illustrated by the example given below in Section V-B.

Remark 7: It is important to note that the presented setup is for the natural exponential family, which includes the Normal distribution with known variance, the Poisson distribution, the Gamma distribution with known shape parameter α, the Binomial distribution with known number of trials, and the Negative Binomial distribution with known r. As we work with differential entropy, an important example in this setup is the Gamma distribution.

B. Example: Gamma Distribution

As a concrete example of the model of (78), we now consider the Gamma distribution, *i.e.,* we let X be any positive, absolutely continuous random variable with finite variance σ_X^2 and differential entropy $h(X)$, and

$$X \sim p_X(\cdot) \ \text{ and } \ Y \mid X=x \sim \frac{x^\alpha}{\Gamma(\alpha)} y^{\alpha-1} e^{-xy}. \tag{93}$$

That is, the conditional distribution of Y given $X=x$ is Gamma with known shape parameter α and rate parameter $x>0$.[4] In this case, $h(Y|X)$ is calculated as $-\mathsf{E}[\log X]+\alpha+\log\Gamma(\alpha)+(1-\alpha)\psi(\alpha)$. Define the corrective term

$$\Delta(p_X,\alpha) \triangleq \alpha+\log\Gamma(\alpha)+(1-\alpha)\psi(\alpha) - \frac{1}{2}\log(2\pi e) - \mathsf{E}[\log X]. \tag{94}$$

Depending on p_X and α, $\Delta(p_X,\alpha)$ can be positive or negative, and Theorem 4 evaluates to

$$h(\mathsf{E}[X|Y]) \geq 2h(X) - h(Y) + 2\Delta(p_X,\alpha). \tag{95}$$

Observe that this model corresponds to a multiplicative model rather than the additive model we considered in Fig. 1, i.e.,

$$Y = X^{-1}\, G \tag{96}$$

[4]For a fixed shape parameter, the canonical parameter of Gamma distribution is the additive inverse of the rate parameter. However, this does not change the Theorem 4

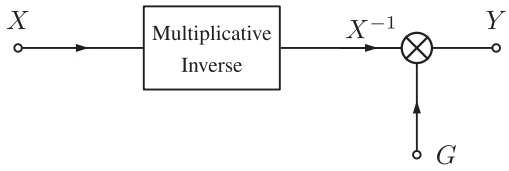

Fig. 7. The Gamma Model.

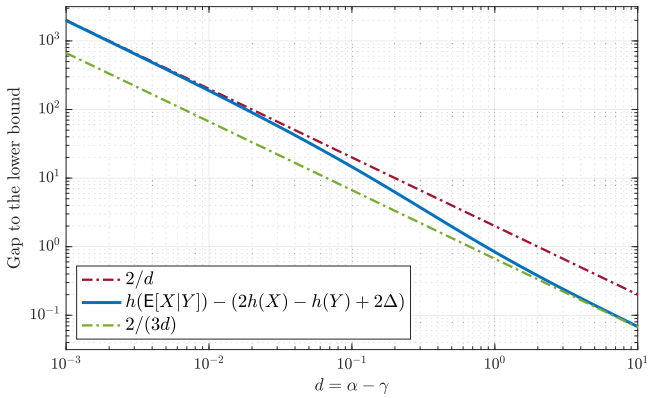

Fig. 8. Gap to the lower bound vs d.

where G is Gamma with shape parameter α and rate parameter 1, independent of X. This is illustrated in Fig. 7.

To illustrate the Theorem 4 for a specific input distribution, let X be a Beta-prime random variable, i.e., its PDF is

$$p_X(x) = \begin{cases} \frac{\Gamma(\alpha)}{\Gamma(\alpha-\gamma)\Gamma(\gamma)}(x-1)^{\alpha-\gamma-1}x^{-\alpha} & \text{if } x \geq 1 \\ 0 & \text{otherwise} \end{cases} \tag{97}$$

where $\alpha > \gamma > 2$ so that its variance is finite. It is easy to check that for this choice of $p_X(\cdot)$, Y is Gamma with shape parameter γ and rate parameter 1, i.e., $p_Y(y) = \frac{e^{-y}\, y^{\gamma-1}}{\Gamma(\gamma)}$ for $y>0$. Conditioned on $Y=y$,

$$X = 1+T \tag{98}$$

where T is Gamma with shape parameter $\alpha-\gamma$ and rate parameter y. Thus,

$$\mathsf{E}[X|Y] = 1+\frac{\alpha-\gamma}{Y} \tag{99}$$

$$\mathsf{Var}(X|Y) = \frac{\alpha-\gamma}{Y^2} \tag{100}$$

and every term in (79) can be calculated analytically as a function of α and γ. Furthermore, the gap to the lower bound in (95) depends only on the difference $d \triangleq \alpha-\gamma$:

$$h(\mathsf{E}[X|Y]) - (2h(X) - h(Y) + 2\Delta) = \log\left(\frac{2\pi e d}{(\Gamma(d))^2}\right) + 2(d-1)\psi(d) - 2\,d. \tag{101}$$

Using the series expansion of the gamma and digamma functions, one can show that the gap is asymptotically equivalent to $\frac{2}{d}$ and $\frac{2}{3d}$ for $d \downarrow 0$ and $d \uparrow \infty$, respectively. We illustrate the tightness of the lower bound as a function of d in Fig. 8.

VI. Extension to the Vector Case

In this section, we consider the extension of the main result (18) and the upper bound (26) under the vector Gaussian noise model, i.e., the input-output relationship is governed by

$$\mathbf{Y} = \mathbf{A}\mathbf{X} + \mathbf{W} \tag{102}$$

where $\mathbf{W} \in \mathbb{R}^n$ is a zero mean Gaussian random vector with positive-definite covariance matrix $\mathbf{K_W}$, $\mathbf{X} \in \mathbb{R}^n$ is and $\mathbf{A}$ is a full-rank, $n \times n$ matrix. It is assumed that $\mathbf{X}$ and $\mathbf{W}$ are independent, and the only assumption on $\mathbf{X}$ is that its covariance matrix $\mathbf{K_X}$ is full-rank, i.e., $\mathbf{X}$ is non-degenerate.

Denote the conditional variance matrix by

$$\mathsf{Var}\left(\mathbf{X}|\mathbf{Y}\right) \triangleq \mathsf{E}\left[\mathbf{X}\mathbf{X}^T|\mathbf{Y}\right] - \mathsf{E}\left[\mathbf{X}|\mathbf{Y}\right]\mathsf{E}\left[\mathbf{X}^T|\mathbf{Y}\right], \tag{103}$$

and the MMSE matrix by

$$\mathsf{MMSE}(\mathbf{X}|\mathbf{Y}) \triangleq \mathsf{E}\left[\mathsf{Var}\left(\mathbf{X}|\mathbf{Y}\right)\right], \tag{104}$$

and the Jacobian matrix of a transformation $\boldsymbol{\phi} : \mathbf{R}^n \mapsto \mathbf{R}^m$ by $\mathbf{J_y}\boldsymbol{\phi}(\mathbf{y})$ with i^{th} row, j^{th} column element being $\frac{\partial \phi_j}{\partial y_i}$. The result of this section is that a similar lower bound is valid under the model in (102).

Proposition 2: (A Lower Bound of Differential Entropy in Vector Case): For the model given in (102), we have

$$h\left(\mathsf{E}\left[\mathbf{X}|\mathbf{Y}\right]\right) \geq 2h\left(\mathbf{X}\right) - h\left(\mathbf{Y}\right) + \log\det\mathbf{A}. \tag{105}$$

Furthermore, equality is achieved when X is Gaussian.

Proof: Since $I(\mathbf{X};\mathbf{Y}) = h(\mathbf{Y}) - h(\mathbf{W}) = h(\mathbf{X}) - h(\mathbf{X}|\mathbf{Y})$, (105) is equivalent to

$$\begin{aligned} h(\mathbf{X}|\mathbf{Y}) &\leq h(\mathbf{W}) - \frac{1}{2}h(\mathbf{Y}) \\ &\quad + \frac{1}{2}h(\mathsf{E}\left[\mathbf{X}|\mathbf{Y}\right]) - \frac{1}{2}\log\det\mathbf{A} \qquad (106) \\ &= \frac{1}{2}\log\left((2\pi e)^n\ \det\mathbf{K_W}\right) - \frac{1}{2}h(\mathbf{Y}) \\ &\quad + \frac{1}{2}h(\mathsf{E}\left[\mathbf{X}|\mathbf{Y}\right]) - \frac{1}{2}\log\det\mathbf{A}. \qquad (107) \end{aligned}$$

By the maximum entropy argument,

$$h(\mathbf{X}|\mathbf{Y}) \leq \frac{1}{2}\mathsf{E}\left[\log\left((2\pi e)^n\ \det\mathsf{Var}\left(\mathbf{X}|\mathbf{Y}\right)\right)\right]. \tag{108}$$

Hence, it is sufficient to show that

$$\begin{aligned} &\frac{1}{2}\mathsf{E}\left[\log\left((2\pi e)^n\ \det\mathsf{Var}\left(\mathbf{X}|\mathbf{Y}\right)\right)\right] \\ &\quad \leq \frac{1}{2}\log\left((2\pi e)^n\ \det(\mathbf{A}^{-1}\,\mathbf{K_W})\right) - \frac{1}{2}h(\mathbf{Y}) \\ &\qquad + \frac{1}{2}h(\mathsf{E}\left[\mathbf{X}|\mathbf{Y}\right]) \end{aligned} \tag{109}$$

which is equivalent to

$$\mathsf{E}\left[\log\det\left(\mathbf{A}\,\mathbf{K_W^{-1}}\,\mathsf{Var}\left(\mathbf{X}|\mathbf{Y}\right)\right)\right] \leq h(\mathsf{E}\left[\mathbf{X}|\mathbf{Y}\right]) - h(\mathbf{Y}). \tag{110}$$

By the change of variables formula,

$$h(\mathsf{E}\left[\mathbf{X}|\mathbf{Y}\right]) = h(\mathbf{Y}) + \mathsf{E}\left[\log|\det\mathbf{J_Y}(\mathsf{E}\left[\mathbf{X}|\mathbf{Y}\right])|\right] \tag{111}$$

provided that the transformation $\mathbf{y} \mapsto \mathsf{E}[\mathbf{X}|\mathbf{Y}=\mathbf{y}]$ is diffeomorphic. Thus, it remains to justify

$$\mathsf{E}\left[\log|\det\mathbf{J_Y}(\mathsf{E}\left[\mathbf{X}|\mathbf{Y}\right])|\right] \geq \mathsf{E}\left[\log\det\left(\mathbf{A}\,\mathbf{K_W^{-1}}\,\mathsf{Var}\left(\mathbf{X}|\mathbf{Y}\right)\right)\right] \tag{112}$$

and $\mathbf{y} \mapsto \mathsf{E}[\mathbf{X}|\mathbf{Y}=\mathbf{y}]$ is diffeomorphic. Under Gaussian noise, the variance identity of Hatsell and Nolte [24] gives

$$\mathbf{J_y}\mathsf{E}\left[\mathbf{X}|\mathbf{Y}=\mathbf{y}\right] = \mathbf{A}^{-1}\,\mathbf{K_W^{-1}}\,\mathbf{A}\,\mathsf{Var}\left(\mathbf{X}|\mathbf{Y}=\mathbf{y}\right)\mathbf{A}^T \tag{113}$$

for every realization $\mathbf{y}$. Hence, (112) is satisfied with equality and $\mathbf{y} \mapsto \mathsf{E}[\mathbf{X}|\mathbf{Y}=\mathbf{y}]$ is diffeomorphic provided that $\mathbf{K_X}$ is full-rank. When the input is Gaussian, (108) and therefore the main lower bound (105) are satisfied with equality:

$$\begin{aligned} h(\mathsf{E}\left[\mathbf{X}|\mathbf{Y}\right]) &= 2h\left(\mathbf{X}\right) - h\left(\mathbf{Y}\right) + \log\det\mathbf{A} \\ &= \tfrac{1}{2}\log\det(2\pi e(\mathbf{A}\mathbf{K_X})^2(\mathbf{A}\mathbf{K_X}\mathbf{A}^T + \mathbf{K_W})^{-1}). \end{aligned}$$

■

By the concavity of the log determinant, Jensen's Inequality gives an upper bound to $h(\mathsf{E}[\mathbf{X}|\mathbf{Y}])$.

Lemma 4: (Upper Bounds of Differential Entropy in Vector Case): For the model given in (102), we have

$$\begin{aligned} h(\mathsf{E}\left[\mathbf{X}|\mathbf{Y}\right]) &= h(\mathbf{Y}) \\ &\quad + \mathsf{E}\left[\log\det\left(\mathbf{A}\,\mathbf{K_W^{-1}}\,\mathsf{Var}\left(\mathbf{X}|\mathbf{Y}\right)\right)\right] \qquad (114) \\ &\leq h(\mathbf{Y}) + \log\det\mathsf{E}\left[\mathsf{Var}\left(\mathbf{X}|\mathbf{Y}\right)\right] \\ &\quad + \log\det\left(\mathbf{A}\,\mathbf{K_W^{-1}}\right) \qquad (115) \\ &= h(\mathbf{Y}) + \log\det\mathsf{MMSE}(\mathbf{X}|\mathbf{Y}) \\ &\quad + \log\det\left(\mathbf{A}\,\mathbf{K_W^{-1}}\right) \qquad (116) \\ &\leq \frac{1}{2}\log\left((2\pi e)^n\ \det(\mathbf{K_X} + \mathbf{K_W})\right) \\ &\quad + \log\det\mathsf{MMSE}(\mathbf{X}|\mathbf{Y}) \\ &\quad + \log\det\left(\mathbf{A}\,\mathbf{K_W^{-1}}\right) \qquad (117) \end{aligned}$$

Proof: (114) follows from Hatsell and Nolte Identity [24] combined with the change of variables, (115) follows from Jensen's Inequality, (116) is by the definition of MMSE matrix and (117) is by the maximum entropy argument. ■

Remark 8: (A specific case): One could set $\mathbf{A} = \mathbf{I}_n$, i.e., the simple extension of the scalar model in the vector setting:

$$\mathbf{Y} = \mathbf{X} + \mathbf{W} \tag{118}$$

for a Gaussian random vector $\mathbf{W}$ with covariance matrix $\mathbf{K_W}$, and an arbitrary random vector $\mathbf{X}$ with finite covariance matrix $\mathbf{K_X}$. In this setup, the main inequality becomes

$$h\left(\mathsf{E}\left[\mathbf{X}|\mathbf{Y}\right]\right) \geq 2h\left(\mathbf{X}\right) - h\left(\mathbf{Y}\right), \tag{119}$$

which is the same bound as in the scalar case, cf. (18). Similarly, the upper bound (116) becomes

$$\begin{aligned} h\left(\mathsf{E}\left[\mathbf{X}|\mathbf{Y}\right]\right) &\leq h(\mathbf{Y}) + \log\det\mathsf{MMSE}(\mathbf{X}|\mathbf{Y}) \\ &\quad + \log\det\left(\mathbf{K_W^{-1}}\right), \end{aligned} \tag{120}$$

which is an extension of (27).

Remark 9: Similar to the scalar case, (108) and therefore the main lower bound (105) are tight for input distributions close to Gaussian in terms of KL divergence. For a fixed input distribution, if the input or the additive noise is scaled so that $\mathbf{K_X}$ and $\mathbf{K_W}$ vary, (105) becomes tighter as $\det(\mathbf{K_X}\,\mathbf{K_W^{-1}})$ increases.

APPENDIX A
DERIVATION OF HATSELL-NOLTE IDENTITY

As the original proofs are not explicit, we include derivations for Tweedie's Formula and for the Hatsell-Nolte identity relying on multiple uses of differentiation under the integral sign. Hence, the tool we need for this proof is the well-known *Leibniz Rule* which we state here for completeness and future reference:

Theorem 5: (Leibniz Integral Rule):[35] Let $(S, \mathcal{S}, \mu)$ be a measure space. Let f be a complex valued function defined on $\mathbb{R} \times S$. Let $\delta > 0$, and suppose that for $y \in (x-\delta, x+\delta)$ we have

i) $p_Y(y) = \int_S \xi(y,s)\,\mu(ds)$ with $\int_S |\xi(y,s)|\,\mu(ds) < \infty$.
ii) For fixed s, $\frac{\partial \xi}{\partial y}\,(y,s)$ exists and is a continuous function of y.
iii) $\int_S \sup_{\theta\in[-\delta,\delta]} |\frac{\partial \xi}{\partial y}(x+\theta,s)|\mu(ds) < \infty$.

Then, $p'_Y(x) = \int_S \frac{\partial \xi}{\partial x}(x,s)\,\mu(ds)$.

As we use Tweedie's Formula in the proof of Hatsell-Nolte Identity, we begin by proving the former.

Lemma 5: (Tweedie's Formula): For the model given in (1),

$$\mathsf{E}\left[X \mid Y=y\right] = y + \sigma_W^2 \frac{d}{dy}\,\log p_Y(y). \tag{121}$$

Proof: Let ϕ_W denote the zero-mean Gaussian PDF with variance σ_W^2. By independence, the density p_Y is the convolution of p_X and ϕ_W:

$$p_Y(y) = \int_{\mathbb{R}} p_X(s)\,\phi_W(y-s)\,ds. \tag{122}$$

Multiplying both sides by σ_W^2 and taking derivative of (122) w.r.t. y, we obtain

$$\sigma_W^2 p'_Y(y) = \sigma_W^2 \frac{d}{dy}\int_{\mathbb{R}} p_X(s)\,\phi_W(y-s)\,ds \tag{123}$$

$$\overset{(a)}{=} \int_{\mathbb{R}} p_X(s)\,(s-y)\,\phi_W(y-s)\,ds \tag{124}$$

$$= \int_{\mathbb{R}} s\,p_X(s)\,\phi_W(y-s)\,ds - y\,p_Y(y) \tag{125}$$

where we used the fact that $\sigma_W^2 \frac{d}{ds}\,\phi_W(s) = -s\,\phi_W(s)$, and Step (a) follows from the Leibniz Integral Rule, as we argue carefully below. Let us divide both sides of (125) by $p_Y(y)$:

$$\sigma_W^2 \frac{p'_Y(y)}{p_Y(y)} = \sigma_W^2 \frac{d}{dy}\,\log p_Y(y) \tag{126}$$

$$= \int_{\mathbb{R}} s\,\frac{p_X(s)\,\phi_W(y-s)}{p_Y(y)}\,ds - y. \tag{127}$$

Observe that $\mathsf{E}[X \mid Y=y] = \int_{\mathbb{R}} s\,\frac{p_X(s)\,\phi_W(y-s)}{p_Y(y)}\,ds$ since the joint density of X, Y is simply $p_{X,Y}(x,y) = p_X(x)\,\phi_W(y-x)$. Hence, we obtain the desired result

$$\mathsf{E}\left[X \mid Y=y\right] = y + \sigma_W^2 \frac{d}{dy}\,\log p_Y(y). \tag{128}$$

For the justification of Step (a), one can use the Leibniz Integral Rule stated above in Theorem 5. To use this theorem, we now verify that Conditions *(i), (ii),* and *(iii)* are satisfied. First, observe that $p_Y(y) = \int_{\mathbb{R}} \phi_W(y-s)\,p_X(s)\,ds \Rightarrow \xi(y,s) = \phi_W(y-s)$ and $\mu(ds) = p_X(s)\,ds$. With this, we observe the following:

i) Since $\xi(y,s) = |\xi(y,s)|$ for all $y, s \in \mathbb{R}$; we have $\int_{\mathbb{R}} |\xi(y,s)|\,\mu(ds) = p_Y(y) < \infty$.
ii) For any fixed s, $\frac{\partial \xi}{\partial y}\,(y,s) = \sigma_W^{-2}(s-y)\,\phi_W(y-s)$ obviously exists and is a continuous function of y for every $y \in \mathbb{R}$.
iii)

$$\left|\frac{\partial \xi}{\partial y}\,(x+\theta,s)\right| = \sigma_W^{-2}\,|(s-x-\theta)\,\phi_W(x+\theta-s)| \tag{129}$$

$$\leq \frac{(|s|+|x|+|\theta|)}{\sigma_W^3\sqrt{2\,\pi}} \tag{130}$$

for all $s,\ x,\ \theta \in \mathbb{R}$. Hence,

$$\int_S \sup_{\theta\in[-\delta,\delta]} \left|\frac{\partial \xi}{\partial y}(x+\theta,s)\right| \mu(ds) \tag{131}$$

$$\leq \int_S \frac{|s|+|x|+\delta}{\sigma_W^3\sqrt{2\,\pi}}\,\mu(ds) \tag{132}$$

$$< \infty \tag{133}$$

for every $\delta > 0$ and $x \in \mathbb{R}$ since we assume that X is square-integrable, thus integrable. Observe that (iii) is satisfied for every $\delta > 0$, hence we have

$$p'_Y(y) = \frac{d}{dy}\int_{\mathbb{R}} \phi_W(y-s)\,p_X(s)\,ds \tag{134}$$

$$= \int_{\mathbb{R}} \frac{\partial}{\partial y}\phi_W(y-s)\,p_X(s)\,ds \quad \forall y \in \mathbb{R}. \tag{135}$$

This concludes the proof that the Leibniz Integral rule applies, and thus, concludes the proof of Tweedie's formula. ■

Theorem 6 (Hatsell-Nolte Identity): For the model given in (1),

$$\sigma_W^2 \frac{d}{dy}\mathsf{E}\left[X \mid Y=y\right] = \mathsf{Var}\left(X \mid Y=y\right). \tag{136}$$

Proof: By definition, $\mathrm{Var}(X \mid Y=y) = \mathsf{E}[X^2 \mid Y=y] - \mathsf{E}[X \mid Y=y]^2$. Hence, it is sufficient to derive a formula for $\mathsf{E}[X^2 \mid Y=y]$ by multiplying (125) by σ_W^2 and taking derivative w.r.t y:

$$\sigma_W^4 p''_Y(y) \overset{(b)}{=} \int_{\mathbb{R}} s\,(s-y)\,p_X(s)\,\phi_W(y-s)\,ds - \sigma_W^2 p_Y(y) - \sigma_W^2 y\,p'_Y(y)$$

$$\sigma_W^4 p'_Y(y) = \sigma_W^2 \int_{\mathbb{R}} s\,p_X(s)\,\phi_W(y-s)\,ds - \sigma_W^2 y\,p_Y(y) \tag{137}$$

$$= \int_{\mathbb{R}} s^2\, p_X(s)\, \phi_W(y-s)\, ds - y \int_{\mathbb{R}} s\, p_X(s)\, \phi_W(y-s)\, ds - \sigma_W^2 p_Y(y) - \sigma_W^2 y\, p_Y'(y). \quad (138)$$

Dividing both sides of (138) by $p_Y(y)$ and recalling $\mathsf{E}[X^p \,|\, Y=y] = \int_{\mathbb{R}} s^p \, \frac{p_X(s)\,\phi_W(y-s)}{p_Y(y)}\, ds$, we obtain

$$\sigma_W^4 \frac{p_Y''(y)}{p_Y(y)} = \mathsf{E}\left[X^2 \,|\, Y=y\right] - y\,\mathsf{E}\left[X \,|\, Y=y\right] - \sigma_W^2 \left(1 + y\,\frac{p_Y'(y)}{p_Y(y)}\right) \quad (139)$$

$$= \mathsf{E}\left[X^2 \,|\, Y=y\right] - y\left(y + \sigma_W^2 \frac{p_Y'(y)}{p_Y(y)}\right) - \sigma_W^2 \left(1 + y\,\frac{p_Y'(y)}{p_Y(y)}\right), \quad (140)$$

which implies

$$\mathsf{E}\left[X^2 \,|\, Y=y\right] = \sigma_W^4 \frac{p_Y''(y)}{p_Y(y)} + 2\sigma_W^2 y\,\frac{p_Y'(y)}{p_Y(y)} + y^2 + \sigma_W^2. \quad (141)$$

Combining (128) and (141), we obtain

$$\mathrm{Var}\left(X \,|\, Y=y\right) = \sigma_W^4 \frac{p_Y''(y)}{p_Y(y)} + 2\sigma_W^2 y\,\frac{p_Y'(y)}{p_Y(y)} + y^2 + \sigma_W^2 - \left(y + \sigma_W^2 \frac{p_Y'(y)}{p_Y(y)}\right)^2 \quad (142)$$

$$= \sigma_W^4 \frac{p_Y''(y)\,p_Y(y) - p_Y'(y)^2}{p_Y(y)^2} + \sigma_W^2 \quad (143)$$

$$= \sigma_W^2 \frac{d}{dy}\left(y + \sigma_W^2 \frac{p_Y'(y)}{p_Y(y)}\right) \quad (144)$$

$$= \sigma_W^2 \frac{d}{dy}\mathsf{E}\left[X \,|\, Y=y\right]. \quad (145)$$

What remains is to justify the step (b), which follows from the Leibniz Integral Rule stated above in Theorem 5. To use this theorem, we verify that Conditions *(i), (ii),* and *(iii)* are satisfied. In this setting,

$$p_Y'(y) = \sigma_W^{-2} \int_{\mathbb{R}} (s-y)\,\phi_W(y-s)\,p_X(s)\,ds$$

$\Rightarrow \xi(y,s) = \sigma_W^{-2}(s-y)\,\phi_W(y-s)$ and $\mu(ds) = p_X(s)\,ds$. With this, we observe the following:

i) $\int_{\mathbb{R}} |\xi(y,s)|\,\mu(ds) = \int_{\mathbb{R}} \sigma_W^{-2}|s-y|\,\phi_W(y-s)\,\mu(ds) \le \int_{\mathbb{R}} (|s|+|y|)\,\frac{1}{\sigma_W^3\sqrt{2\,\pi}}\,\mu(ds) < \infty$ for all $y \in \mathbb{R}$.

ii) For any fixed s,

$$\frac{\partial \xi}{\partial y}(y,s) = \sigma_W^{-4}\left((s-y)^2 - \sigma_W^2\right)\,\phi_W(y-s)$$

obviously exists and is a continuous function of y for every $y \in \mathbb{R}$.

iii)

$$\left|\frac{\partial \xi}{\partial y}(x+\theta,s)\right| = \sigma_W^{-4}\left|(s-x-\theta)^2 - \sigma_W^2\right|\,|\phi_W(x+\theta-s)| \le \frac{\left|s^2 - 2\,(x+\theta)\,s + (x+\theta)^2 - \sigma_W^2\right|}{\sigma_W^5\sqrt{2\,\pi}} \le \frac{s^2 + 2\,(|x|+|\theta|)\,|s|}{\sigma_W^5\sqrt{2\,\pi}} + \frac{x^2+\theta^2+2\,|x|\,|\theta| + \sigma_W^2}{\sigma_W^5\sqrt{2\,\pi}}$$

for all $s,\ x,\ \theta \in \mathbb{R}$. Hence,

$$\int_S \sup_{\theta\in[-\delta,\delta]} \left|\frac{\partial \xi}{\partial y}(x+\theta,s)\right| \mu(ds) \le \int_S \left(\frac{s^2 + 2\,(|x|+\delta)\,|s|}{\sigma_W^5\sqrt{2\,\pi}} + \frac{x^2+\delta^2+2\,|x|\,\delta + \sigma_W^2}{\sigma_W^5\sqrt{2\,\pi}}\right)\mu(ds) < \infty$$

for every $\delta > 0$ and $x \in \mathbb{R}$ since we assume that X is square-integrable. Since (iii) is satisfied for all $\delta > 0$,

$$p_Y''(y) = \frac{d}{dy}\,p_Y'(y) = \int_{\mathbb{R}} \frac{\partial}{\partial y}\left(\sigma_W^{-2}(s-y)\,\phi_W(y-s)\right)\,p_X(s)\,ds.$$

This concludes the proof that the Leibniz Integral rule applies, and thus, concludes the proof of the Hatsell-Nolte Identity. ■

APPENDIX B
PROOF OF LEMMA 1

To establish Lemma 1, the starting point is the well-known formula for the differential entropy of a transformed random variable, which we state here for completeness.

Lemma 6: (Differential Entropy of Diffeomorphic Transformations of a Random Variable): Let $y \mapsto \varphi(y)$ be a C^1-diffeomorphic[5] transformation on $\mathbb{R}$, and Y be an absolutely continuous random variable with PDF $p_Y(\cdot)$ and finite differential entropy $h(Y)$. Then, the differential entropy of $\varphi(Y)$ satisfies the following equation.

$$h(\varphi(Y)) = h(Y) + \mathsf{E}\left[\log\left|\frac{d\varphi(Y)}{dY}\right|\right]. \quad (146)$$

Proof: Since $\varphi(\cdot)$ is C^1-diffeomorphic, the probability density function of $\varphi(Y)$, denoted as $p_{\varphi(Y)}(\cdot)$, can be written as

$$p_{\varphi(Y)}(\phi) = \frac{p_Y(\varphi^{-1}(\phi))}{|\varphi'(\varphi^{-1}(\phi))|}. \quad (147)$$

[5] $\varphi(\cdot)$ and $\varphi^{-1}(\cdot)$ are continuously differentiable.

Using (147) and a change of variable $y \triangleq \varphi^{-1}(\phi)$, $h(\varphi(Y))$ is simplified to

$$\begin{aligned} h(\varphi(Y)) &= -\int_{-\infty}^{\infty} p_{\varphi(Y)}(\phi)\log p_{\varphi(Y)}(\phi)\,d\phi \\ &= -\int_{-\infty}^{\infty} \frac{p_Y(\varphi^{-1}(\phi))}{|\varphi'(\varphi^{-1}(\phi))|}\log\frac{p_Y(\varphi^{-1}(\phi))}{|\varphi'(\varphi^{-1}(\phi))|}\,d\phi \\ &= -\int_{-\infty}^{\infty} \frac{p_Y(y)}{|\varphi'(y)|}\log\frac{p_Y(y)}{|\varphi'(y)|}\,\varphi'(y)dy \\ &= -\int_{-\infty}^{\infty} p_Y(y)\log\frac{p_Y(y)}{|\varphi'(y)|}\,dy \qquad (148) \\ &= h(Y) + \mathsf{E}\left[\log\left|\frac{d\varphi(Y)}{dY}\right|\right] \qquad (149) \end{aligned}$$

where (148) follows from the C^1-diffeomorphic assumption. That is, $\varphi(\cdot)$ must be a strictly monotonic function, and analyzing two cases [6] separately both yield (148). ■

We apply this lemma to the function $y \mapsto \mathsf{E}[X|Y=y]$ in the additive white Gaussian noise model. It is straightforward to show that in this case, the function $y \mapsto \varphi(y)$ and its inverse are real-analytic provided that X is a non-degenerate random variable (see for example Lemma 2 and Lemma 3 of [24]). Therefore, $y \mapsto \varphi(y)$ is $C^\infty-$diffeomorphic and (146) can be applied whenever $\sigma_X^2 > 0$. Using (146), we immediately obtain

$$h(\mathsf{E}[X|Y]) = h(Y) + \mathsf{E}\left[\log\left|\frac{d\mathsf{E}[X|Y]}{dY}\right|\right] \qquad (150)$$

$$= h(Y) + \mathsf{E}\left[\log\left(\frac{1}{\sigma_W^2}\mathsf{Var}(X|Y)\right)\right] \qquad (151)$$

where the last line follows from the Hatsell-Nolte identity.

Appendix C
Exponential Family Generalization of Tweedie's Formula

Our objective in this section is to state and prove the Exponential Family generalization of *Tweedie's Formula*, which is also based on multiple uses of differentiation under the integral sign.

Lemma 7 (Tweedie's Formula for Exponential Family): For the model in (78), define $\nu(y) \triangleq \frac{p(y)}{p_b(y)}$. Suppose the following conditions hold for every y in the support of $\nu(y)$.

$$\nu(y) < \infty, \qquad (152a)$$

$$\mathsf{E}[|X|\,|Y=y] < \infty, \qquad (152b)$$

$$\mathsf{E}\left[X^2|Y=y\right] < \infty. \qquad (152c)$$

Then, we have

$$\mathsf{E}[X|Y=y] = \frac{d}{dy}\log\nu(y), \qquad (153)$$

$$\mathsf{Var}(X|Y=y) = \frac{d^2}{dy^2}\log\nu(y). \qquad (154)$$

[6] increasing and decreasing $\varphi(\cdot)$

Proof: By (82), we have

$$\begin{aligned} \mathsf{E}[X|Y=y] &= \int x\,p(x|y)\,dx \\ &= \int x\,e^{xy-\log\nu(y)}\left(q(x)\,e^{-A(x)}\right)dx. \qquad (155) \end{aligned}$$

Taking the logarithmic derivative of $\nu(y)$, we obtain

$$\begin{aligned} \frac{d\log\nu(y)}{dy} &= \frac{\frac{d\nu(y)}{dy}}{\nu(y)} \\ &= \frac{\frac{d}{dy}\int e^{xy-A(x)}q(x)\,dx}{\nu(y)} \\ &\overset{(c)}{=} \frac{\int \frac{d}{dy}\left(e^{xy-A(x)}q(x)\right)dx}{\nu(y)} \\ &= \int x\,e^{xy-\log\nu(y)}\left(q(x)\,e^{-A(x)}\right)dx, \qquad (156) \end{aligned}$$

which is the same expression as (155). To justify (c), we can use Theorem 5. By (152a),

$$\begin{aligned} \int e^{xy-A(x)}q(x)\,dx &= \int \nu(y)\,p(x|y)\,dx \qquad (157) \\ &= \nu(y)\int p(x|y)\,dx \qquad (158) \\ &< \infty, \qquad (159) \end{aligned}$$

so (i) is satisfied. Checking the condition (ii) is trivial, i.e., for every x in the support of $q(\cdot)$, $y \mapsto e^{xy}$ is continuous for every y in the support of $\nu(y)$. Finally, the condition (iii) holds due to (152a) and (152b). Differentiating (156) once more, we obtain

$$\begin{aligned} \frac{d^2\log\nu(y)}{dy^2} &= \frac{d}{dy}\int x\,e^{xy-\log\nu(y)}\left(q(x)\,e^{-A(x)}\right)dx \\ &\overset{(d)}{=} \int \frac{d}{dy}\left(x\,e^{xy-\log\nu(y)}\left(q(x)\,e^{-A(x)}\right)\right)dx \\ &= \int x^2\,e^{xy-\log\nu(y)}\left(q(x)\,e^{-A(x)}\right)dx \\ &\quad - \int x\frac{\nu'(y)}{\nu(y)}\,e^{xy-\log\nu(y)}\left(q(x)\,e^{-A(x)}\right)dx \\ &= \mathsf{E}\left[X^2|Y=y\right] - \left(\mathsf{E}[X|Y=y]\right)^2. \qquad (160) \end{aligned}$$

Justification of (d) is similar to (c): (i) holds due to (152a) and (152b), (ii) follows by the same reasoning as (c), and (iii) holds due to (152a) and (152c). ■

Remark 10: The Gamma distribution example given in Sec. V-B satisfies (152a), (152b), (152c). That is,

$$\begin{aligned} \nu(y) &= \int_0^\infty x^\alpha e^{-xy}\,q(x)\,dx \qquad (161) \\ &\le \int_0^\infty \left(\frac{\alpha}{e\,y}\right)^\alpha q(x)\,dx \qquad (162) \\ &< \infty. \qquad (163) \end{aligned}$$

Similarly,

$$\mathsf{E}\left[X|Y=y\right] \leq \frac{1}{\nu(y)}\left(\frac{\alpha+1}{e\,y}\right)^{\alpha+1} < \infty$$

$$\mathsf{E}\left[X^2|Y=y\right] \leq \frac{1}{\nu(y)}\left(\frac{\alpha+2}{e\,y}\right)^{\alpha+2} < \infty.$$

Appendix D
Calculation of κ_X

In [26], the lower bound is expressed as, for $0 < D < \frac{N(X)}{\kappa_X}$

$$L(D) \geq \frac{\sigma_W^2}{2}\left(\frac{1}{D} - \frac{\kappa_X}{N(X)}\right) - \frac{1}{2}\log\frac{\sigma_X^2}{N(X)} \tag{164}$$

where

$$\kappa_X = \lim_{s\to 0^+}\frac{d}{ds}N(X+\sqrt{s}G). \tag{165}$$

Under regularity conditions, κ_X can be simplified to

$$\kappa_X = \lim_{s\to 0^+}\frac{d}{ds}N(X+\sqrt{s}G) \tag{166}$$

$$= \lim_{s\to 0^+} 2N(X+\sqrt{s}G)\,\frac{d}{ds}h(X+\sqrt{s}G) \tag{167}$$

$$= 2N(X)\lim_{s\to 0^+}\frac{d}{ds}h(X+\sqrt{s}G) \tag{168}$$

$$= 2N(X)\lim_{s\to 0^+}\frac{1}{2}\,J(X+\sqrt{s}G) \tag{169}$$

$$= N(X)\,J(X), \tag{170}$$

which thus yields (58).

Acknowledgment

The authors are grateful to the Associate Editor and anonymous reviewers for a meticulous reading of the draft and insightful feedback which helped improve the manuscript.

References

[1] A. Banerjee, X. Guo, and H. Wang, "On the optimality of conditional expectation as a Bregman predictor," *IEEE Trans. Inf. Theory*, vol. 51, no. 7, pp. 2664–2669, Jul. 2005.
[2] D. Blackwell, "Conditional expectation and unbiased sequential estimation," *Ann. Math. Statist.*, vol. 18, no. 1, pp. 105–110, 1947.
[3] A. Dytso, M. Fauß, and H. V. Poor, "Bayesian risk with Bregman loss: A Cramér–Rao type bound and linear estimation," *IEEE Trans. Inf. Theory*, vol. 68, no. 3, pp. 1985–2000, Mar. 2022.
[4] T. Han, "Hypothesis testing with multiterminal data compression," *IEEE Trans. Inf. Theory*, vol. IT-33, no. 6, pp. 759–772, Nov. 1987.
[5] A. Zaidi, "Hypothesis testing against independence under Gaussian noise," in *Proc. IEEE Int. Symp. Inf. Theory*, Los Angeles, CA, USA, 2020, pp. 1289–1294.
[6] N. Tishby, F. C. Pereira, and W. Bialek, "The information bottleneck method," in *Proc. 37th Annu. Allerton Conf. Commun., Control, Comput.*, 1999, pp. 368–377.
[7] A. Zaidi, I. Estella-Aguerri, and S. Shamai, "On the information bottleneck problems: Models, connections, applications and information theoretic views," *Entropy*, vol. 22, no. 2, 2020, Art. no. 151.
[8] A. Makhdoumi, S. Salamatian, N. Fawaz, and M. Médard, "From the information bottleneck to the privacy funnel," in *Proc. IEEE Inf. Theory Workshop*, 2014, pp. 501–505.
[9] F. P. Calmon, A. Makhdoumi, and M. Médard, "Fundamental limits of perfect privacy," in *Proc. IEEE Int. Symp. Inf. Theory*, 2015, pp. 1796–1800.
[10] H. E. Robbins, "An empirical Bayes approach to statistics," in *Proc. 3rd Berkeley Symp. Math. Statist. Probability*, 1956, vol. 1, pp. 157–163.
[11] C. Hatsell and L. Nolte, "Some geometric properties of the likelihood ratio (Corresp.)," *IEEE Trans. Inf. Theory*, vol. IT-17, no. 5, pp. 616–618, Sep. 1971.
[12] L. D. Brown, "Admissible estimators, recurrent diffusions, and insoluble boundary value problems," *Ann. Math. Statist.*, vol. 42, no. 3, pp. 855–903, Jun. 1971.
[13] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. Hoboken, NJ, USA: Wiley, 2006.
[14] S. Verdú, "Empirical estimation of information measures: A literature guide," *Entropy*, vol. 21, no. 8, 2019, Art. no. 720.
[15] M. Tanahashi and H. Ochiai, "A new reduced-complexity conditional-mean based MIMO signal detection using symbol distribution approximation technique," *IEEE Trans. Signal Process.*, vol. 59, no. 11, pp. 5644–5651, Nov. 2011.
[16] U. Spagnolini, "Cancellation of polarized impulsive noise using an azimuth-dependent conditional mean estimator," *IEEE Trans. Signal Process.*, vol. 46, no. 12, pp. 3333–3344, Dec. 1998.
[17] B. James, B. D. O. Anderson, and R. C. Williamson, "Conditional mean and maximum likelihood approaches to multiharmonic frequency estimation," *IEEE Trans. Signal Process.*, vol. 42, no. 6, pp. 1366–1375, Jun. 1994.
[18] M. I. Miller, A. Srivastava, and U. Grenander, "Conditional-mean estimation via jump-diffusion processes in multiple target tracking/recognition," *IEEE Trans. Signal Process.*, vol. 43, no. 11, pp. 2678–2690, Nov. 1995.
[19] D. Guo, S. Shamai, and S. Verdu, "Mutual information and minimum mean-square error in Gaussian channels," *IEEE Trans. Inf. Theory*, vol. 51, no. 4, pp. 1261–1282, Apr. 2005.
[20] D. P. Palomar and S. Verdu, "Gradient of mutual information in linear vector Gaussian channels," *IEEE Trans. Inf. Theory*, vol. 52, no. 1, pp. 141–154, Jan. 2006.
[21] D. Guo, Y. Wu, S. S. Shitz, and S. Verdú, "Estimation in Gaussian noise: Properties of the minimum mean-square error," *IEEE Trans. Inf. Theory*, vol. 57, no. 4, pp. 2371–2385, Apr. 2011.
[22] Y. Wu and S. Verdu, "Functional properties of minimum mean-square error and mutual information," *IEEE Trans. Inf. Theory*, vol. 58, no. 3, pp. 1289–1301, Mar. 2012.
[23] A. Dytso, H. V. Poor, R. Bustin, and S. Shamai, "On the structure of the least favorable prior distributions," in *Proc. IEEE Int. Symp. Inf. Theory*, 2018, pp. 1081–1085.
[24] A. Dytso, H. V. Poor, and S. Shamai Shitz, "A general derivative identity for the conditional mean estimator in Gaussian noise and some applications," in *Proc. IEEE Int. Symp. Inf. Theory*, Los Angeles, CA, USA, 2020, pp. 1183–1188.
[25] A. Dytso and M. Cardone, "A general derivative identity for the conditional expectation with focus on the exponential family," in *Proc. IEEE Inf. Theory Workshop*, 2021, pp. 1–6.
[26] K. Eswaran and M. Gastpar, "Remote source coding under Gaussian noise: Dueling roles of power and entropy power," *IEEE Trans. Inf. Theory*, vol. 65, no. 7, pp. 4486–4498, Jul. 2019.
[27] A. Marsiglietti and V. Kostina, "A lower bound on the differential entropy of log-concave random vectors with applications," *Entropy*, vol. 20, no. 3, Mar. 2018, Art. no. 185.
[28] A. J. Stam, "Some inequalities satisfied by the quantities of information of fisher and shannon," *Inf. Control*, vol. 2, no. 2, pp. 101–112, 1959.
[29] M. Costa, "A new entropy power inequality," *IEEE Trans. Inf. Theory*, vol. IT-31, no. 6, pp. 751–760, Nov. 1985.
[30] P. L. Dragotti and M. Gastpar, *Distributed Source Coding: Theory, Algorithms and Applications*. Cambridge, MA, USA: Academic Press, 2009.
[31] A. Atalik, A. Köse, and M. Gastpar, "The price of distributed: Rate loss in the CEO problem," in *Proc. 56th Annu. Conf. Inf. Sci. Syst.*, 2022, pp. 125–130.
[32] C. N. Morris, "Natural exponential families with quadratic variance functions," *Ann. Statist.*, vol. 10, no. 1, pp. 65–80, 1982.
[33] B. Efron, "Tweedie's formula and selection bias," *J. Amer. Stat. Assoc.*, vol. 106, no. 496, pp. 1602–1614, 2011.
[34] D. Dowson and A. Wragg, "Maximum-entropy distributions having prescribed first and second moments (Corresp.)," *IEEE Trans. Inf. Theory*, vol. IT-19, no. 5, pp. 689–693, Sep. 1973.
[35] R. Durrett, *Probability: Theory and Examples*. Cambridge, U.K.: Cambridge Univ. Press, 2020.

Arda Atalik received the B.S. degree (*summa cum laude*) and the M.S. degree in electrical engineering from Bilkent University, Ankara, Turkey. He is currently working toward the Ph.D. degree in data science with New York University (NYU), New York, NY, USA, under the supervision of Sumit Chopra, Daniel Sodickson, and Kyunghyun Cho. After his B.S. degree, he completed a summer internship with Erdal Arikan. He spent one year at EPFL's Laboratory for Information in Networked Systems, headed by Michael Gastpar. Before joining NYU, he was a Research Engineer with the telecom industry for three years. His research interests include statistical machine learning, signal processing, and information theory. He was the recipient of a full scholarship and Academic Excellence Award at Bilkent University, and Center for Data Science Fellowship at NYU.

Alper Köse received the B.Sc. degree (Hons.) in electrical and electronics engineering from Bogazici University, Istanbul, Turkey, in 2015, and the M.Sc. degree in electrical and electronics engineering along with a minor in computer science from the Swiss Federal Institute of Technology Lausanne, Lausanne, Switzerland. He completed his M.Sc. thesis with the Research Laboratory for Electronics, Massachusetts Institute of Technology, Cambridge, MA, USA, in 2017. He is currently working toward the Ph.D. degree with Bogazici University, focusing on communications and information theory. His research interests include machine learning, communications, and networking.

Michael Gastpar (Fellow, IEEE) received the Dipl. El.-Ing. degree from the Eidgenössische Technische Hochschule, Zürich, Switzerland, in 1997, the M.S. degree in electrical engineering from the University of Illinois at Urbana-Champaign, Champaign, IL, USA, in 1999, and the Doctorat ès Science degree from the Ecole Polytechnique Fédérale Lausanne (EPFL), Lausanne, Switzerland, in 2002. He was also a student in engineering and philosophy with the Universities of Edinburgh, Edinburgh, U.K. and Universities of Lausanne, Lausanne, Switzerland. During 2003–2011, he was an Assistant and tenured Associate Professor with the Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, Berkeley, CA, USA. Since 2011, he has been a Professor with the School of Computer and Communication Sciences, EPFL. He was also a Professor with the Delft University of Technology, Delft, The Netherlands, and Researcher with the Mathematics of Communications Department, Bell Labs, Lucent Technologies, Murray Hill, NJ, USA. His research interests include network information theory and related coding and signal processing techniques, with applications to sensor networks and neuroscience. He was the recipient of the IEEE Communications Society and Information Theory Society Joint Paper Award in 2013 and the EPFL Best Thesis Award in 2002. He was an Information Theory Society Distinguished Lecturer during 2009–2011, Associate Editor for Shannon Theory for the IEEE TRANSACTIONS ON INFORMATION THEORY during 2008–2011, and was the Technical Program Committee Co-chair for the 2010 and 2021 International Symposia on Information Theory (Austin, TX, USA, and Melbourne, Australia).