Zero-Delay Source-Channel Coding with a Low-Resolution ADC Front End

Morteza Varasteh[†], Borzoo Rassouli[†], Osvaldo Simeone[*] and Deniz Gündüz[†]

[†] Department of Electrical and Electronic Engineering, Imperial College London, London, U.K.

[*] Department of Informatics, King's College London, UK.

{m.varasteh12; b.rassouli12; d.gunduz}@imperial.ac.uk, osvaldo.simeone@kcl.ac.uk.

Abstract—**Motivated by the practical constraints arising in emerging sensor network and Internet-of-Things (IoT) applications, the zero-delay transmission of a Gaussian measurement over a real single-input multiple-output (SIMO) additive white Gaussian noise (AWGN) channel is studied with a low-resolution analog-to-digital converter (ADC) front end. Joint optimization of the encoder and the decoder mapping is tackled under both the mean squared error (MSE) distortion and the distortion outage probability (DOP) criteria, with an average power constraint on the channel input. Optimal encoder and decoder mappings are identified for a one-bit ADC front end under both criteria. For the MSE distortion, the optimal encoder mapping is shown to be non-linear in general, while it tends to a linear encoder in the low signal-to-noise ratio (SNR) regime, and to an antipodal digital encoder in the high SNR regime. This is in contrast to the optimality of linear encoding at all SNR values in the presence of a full-precision front end. For the DOP criterion, it is shown that the optimal encoder mapping is piecewise constant and can take only two opposite values when it is non-zero. For both the MSE distortion and the DOP criteria, necessary optimality conditions are then derived for K-level ADC front ends as well as front ends with multiple one-bit ADCs. These conditions are used to obtain numerically optimized solutions. Extensive numerical results are also provided in order to gain insights into the structure of the optimal encoding and decoding mappings.**

Index Terms- **Analog-to-digital converter, distortion outage probability, joint source channel coding, mean squared error distortion, zero-delay transmission.**

I. INTRODUCTION

The power consumed by analog-to-digital converters (ADCs) grows exponentially with the number of bits and linearly with the sampling rate [2], [3]. This technological limitation constrains the resolution of the ADCs used in devices that need to operate under stringent power budgets, such as sensor nodes and mobile devices. As an extreme case, one-bit ADCs are of particular interest, since they can be realized using a simple threshold comparator and without the need for automatic gain control [4], [5].

Motivated by these considerations, the impact of a one-bit ADC front end on the performance of a communication system has been studied in [6]–[12] for various models, as briefly reviewed below. While these works focus on reliable transmission of digital information over long blocks, in many applications, such as the Internet-of-Things (IoT), cyber-physical systems and wireless sensor networks, the low-delay transfer of analog measurements is crucial. Specific examples include uncoded video transmission techniques [13], [14], memoryless amplify-and-forward relays, wearable sensors that detect neurological impulses [15], and real-time control applications for the IoT [16], [17]. In light of these emerging applications, this work considers the zero-delay transmission of an analog Gaussian source over a real single-input multiple-output (SIMO) additive white Gaussian noise (AWGN) channel followed by a low-resolution ADC front end. As illustrated in Figure 1, this problem refers to the transmission of a single sample of a Gaussian source over an individual use of the SIMO AWGN channel. With an infinite resolution front end and equal bandwidth of the Gaussian source and the AWGN channel [18], it is well known that, linear transmission and minimum mean squared error (MMSE) estimation are, respectively, the optimal encoder and decoder under an average power constraint and a mean squared error (MSE) distortion measure. Nonetheless, in the finite resolution scenario studied here, the optimal encoder and decoder mappings have been unknown.

A. Related Works

Within the spectrum of information theoretic analysis of joint source-channel coding, while the classical Shannon theoretic infinite block-length regime occupies one end of the spectrum, the zero-delay transmission of a single source sample over single channel use lies at the opposite end. Intermediate, but still asymptotic, analyses have also been put forth recently by investigating second-order approximations [19], [20]. For the problem of zero-delay transmission, there is no explicit method to obtain the optimal encoding and decoding mappings, except for the special case of statistically matched source-channel pairs [21], [18]. This special case includes the setting studied in this paper when the receiver front end has infinite resolution. Various solutions have been proposed in the literature for specific unmatched source-channel pairs. Notable examples

are the space-filling curves proposed by Shannon [22] and Kotelnikov [23], and later extended in [24]–[28], for the delay-limited transmission of a Gaussian source over an AWGN channel with bandwidth mismatch between the source and the channel. Other solutions include [24], [29]–[31].

Communication with a finite resolution receiver front end received considerable recent attention, focusing mostly on the channel coding aspects. In [6], it is shown that antipodal signalling, or BPSK, is capacity achieving for a real-valued AWGN channel with a one-bit ADC front end, whereas, for the complex counterpart, QPSK is optimal. While these results hold under the assumption that the one-bit ADC is symmetric (that is, it is a zero-threshold comparator), in [7] it is shown that, in the low signal-to-noise ratio (SNR) regime, a symmetric quantizer is not optimal, and the optimal performance is achieved by flash signalling [8, Def. 2] together with an optimized asymmetric quantizer. In [9], it is shown that, for a point-to-point Rayleigh fading channel with a one-bit ADC front end, under the assumption of perfect channel state information (CSI) at the receiver, QPSK is capacity achieving. Instead, when CSI is not available at the receiver, reference [10] proves that QPSK is optimal above an SNR threshold that depends on the coherence time of the channel, while, for lower SNRs, on-off QPSK achieves the capacity. For the point-to-point multiple-input multiple-output (MIMO) scenarios with a one-bit ADC front end at each receive antenna, the capacity is unknown. In [11], it is argued that, with perfect CSI at the receiver, QPSK is optimal at very low SNRs, while, in [12], upper and lower bounds on the capacity with perfect transmitter CSI are presented. To the best of our knowledge, joint source-channel coding with a finite resolution ADC is first studied in [1], which is extended to the scenario with a correlated receiver side information in [32].

B. Main Contributions and Organization of the paper

We study the optimization of encoding and decoding mappings for the transmission of a Gaussian source sample over the single use of a real SIMO AWGN channel with a low-resolution ADC front end (see Figure 1). We consider two different criteria, namely the MSE distortion and the distortion outage probability (DOP), with an average power constraint on the channel input. Optimal solutions are derived for the case of a one-bit ADC front end for both criteria. For the MSE distortion, we show that the optimal encoder mapping tends to a linear encoder, which is optimal with a full-precision front end, only in the low-SNR regime. For the DOP criterion, we derive the optimal encoder mapping, showing that it is piecewise constant and that it can take only two opposite values when it is non-zero. For both the MSE distortion and the DOP criteria, we study necessary optimality conditions for K-level ADC front ends as well as for front ends with multiple one-bit ADCs. Extensive numerical results are also provided in order to gain insights into the structure of the optimal encoding and decoding mappings.

The rest of the paper is organized as follows. In Section II, we introduce the system model. In Section III, we consider
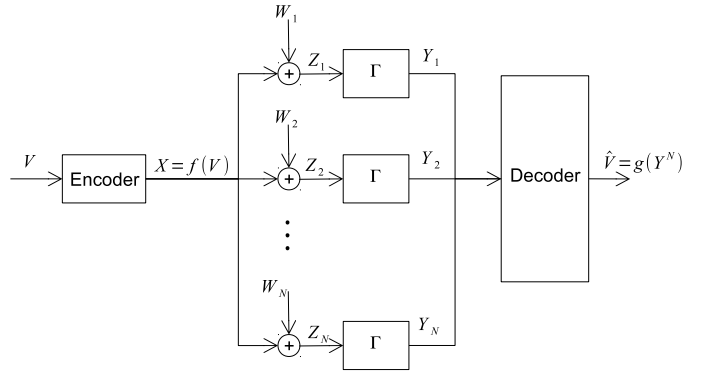

Figure 1: System model for the transmission of a single Gaussian source sample over a quantized SIMO AWGN channel with N ADC front ends.

the design of the optimal transceiver under the MSE distortion criterion when the receiver has a single observation of the source. In Section IV, we study the same design problem under the DOP criterion. In Section V, we study the more general case in which the receiver makes multiple one-bit observations under both the MSE distortion and the DOP criteria. In Section VI, numerical results are provided, followed by the conclusions in Section VII.

Notation: Throughout the paper, $\mathbb{R}$ denotes the set of real numbers; uppercase and lowercase letters denote random variables and realizations, respectively. We use b_j^N to denote the bit-wise representation of the number $2^N - j$, $j = 1, \ldots, 2^N$ with length N. $\mathbb{E}[\cdot]$ and $\Pr(\cdot)$ denote the expectation and probability operators, respectively. Let $f'(x) = \frac{df(x)}{dx}$, $f''(x) = \frac{d^2 f(x)}{dx^2}$ denote the first and second order derivatives of the continuously differentiable function f with respect to its argument. The standard normal distribution is denoted by $\mathcal{N}(0, 1)$, with cumulative distribution function $\Phi(\cdot)$ and the complementary cumulative distribution function (CCDF) by $Q(\cdot)$, which is given by

$$Q(z) = \frac{1}{\sqrt{2\pi}} \int_z^\infty e^{-\frac{t^2}{2}} dt.$$

Unless stated otherwise, boundaries of integrals are from $-\infty$ to ∞.

II. System Model

We consider the system model in Figure 1, in which a single sample of a Gaussian source $V \sim \mathcal{N}(0, \sigma_v^2)$ is transmitted over a single use of a quantized SIMO AWGN channel. The encoded signal is given as $X = f(V)$, where $f : \mathbb{R} \to \mathbb{R}$ is a mapping from the source sample to the channel input, with average transmission power $P = \mathbb{E}[f(V)^2]$. The receiver makes N noisy measurements of the encoded signal, which are digitized by means of a low-resolution ADC front end. Mathematically, each noisy received signal is modelled as

$$Z_i = f(V) + W_i, \ i = 1, ..., N, \tag{1}$$

where the noise $W_i \sim \mathcal{N}(0, \sigma_{w_i}^2)$, $i = 1, ..., N$, is independent over index i. Each received signal Z_i is quantized with a scalar K-level ADC producing a quantized signal

$$Y_i = \Gamma(Z_i), \ i = 1, \ldots, N. \tag{2}$$

The scalar K-level ADC is characterized by fixed quantization intervals and corresponding quantized levels, namely

$$\Gamma(z) = y_{(j)}, \quad \text{for} \quad z \in \left[z_{(j-1)}, z_{(j)}\right), \ j = 1, \ldots, K, \tag{3}$$

where $z_{(j-1)}$ and $z_{(j)}$ are the lower and upper bounds of the interval corresponding to the quantized signal $y_{(j)}$, respectively, for $j = 1, \ldots, K$, and we have $z_{(0)} = -\infty$ and $z_{(K)} = \infty$. Note that the ADCs employed to quantize different channel outputs all have the same quantization intervals and reconstruction levels.

For most of the paper, we will consider a symmetric one-bit ADC with threshold $z_{(1)} = 0$ and reconstruction levels $y_{(1)} = 1$ and $y_{(2)} = 0$:

$$\Gamma(z) = \begin{cases} 0 & z \geq 0, \\ 1 & z < 0. \end{cases} \tag{4}$$

We define the SNR as

$$\gamma = \frac{NP}{\sum\limits_{i=1}^{N} \sigma_{w_i}^2}. \tag{5}$$

Based on the quantized signals $(Y_1, ..., Y_N) \triangleq Y^N$, the decoder produces an estimate $\hat{V}$ of V using a decoding function $g : \{y_{(1)}, \ldots, y_{(K)}\}^N \to \mathbb{R}$, i.e., $\hat{V} = g(Y^N)$.

Two performance criteria are considered, namely the MSE distortion, which is defined as

$$\bar{D} = \mathbb{E}\left[(V - \hat{V})^2\right], \tag{6}$$

and the DOP, which is instead defined as

$$\epsilon(D) = \Pr\left((V - \hat{V})^2 \geq D\right). \tag{7}$$

In both cases, we aim at studying the optimal encoder function f, along with the corresponding optimal estimator g at the decoder, such that $\bar{D}$ and $\epsilon(D)$ are minimized, subject to an average power constraint. More specifically, as it is common in related works (see, e.g., [31]), we consider the unconstrained minimization

$$\underset{f,g}{\text{minimize}} \quad L(f, g, \lambda), \tag{8}$$

where

$$L(f, g, \lambda) = \begin{cases} \bar{D} + \lambda \mathbb{E}[f(V)^2] & \text{for the MSE criterion,} \\ \epsilon(D) + \lambda \mathbb{E}[f(V)^2] & \text{for the DOP criterion,} \end{cases} \tag{9}$$

with $\lambda \geq 0$ being a Lagrange multiplier that defines the relative weight given to the average transmission power $\mathbb{E}[f(V)^2]$ as compared to the distortion criterion.

III. SINGLE OBSERVATION: MSE DISTORTION

In this section, we study the design of the encoder and the decoder under the MSE criterion by focusing on the case of a single observation ($N = 1$). For the one-bit ADC in (4), we obtain the optimal encoder and decoder in Section III-A. Furthermore, we consider the conventional linear transmission and digital modulation schemes for reference in Section III-B and Section III-C, respectively. Finally, we consider the extensions to a K-level front end, and obtain a necessary condition on the optimal mapping in Section III-D. For brevity, throughout this section, we drop the subscript $i = 1$ identifying the observation index.

A. Optimal Encoder and Decoder for a One-Bit Front End

To elaborate on the optimal encoder and decoder for the one-bit ADC in (4), without loss of generality, we write the receiver mapping as

$$g(Y) = \hat{V} = \begin{cases} \hat{v}_{(1)} & Y = 0, \\ \hat{v}_{(2)} & Y = 1, \end{cases} \tag{10}$$

which is defined by the pair of parameters $(\hat{v}_{(1)}, \hat{v}_{(2)})$. In (10), since, for any encoder mapping f, the MMSE estimator is optimal under the MSE criterion, and hence also for problem (8), we have $\hat{v}_{(1)} = \mathbb{E}[V|Y = 0]$ and $\hat{v}_{(2)} = \mathbb{E}[V|Y = 1]$. The next proposition provides the optimal encoder mapping.

Proposition III.1. *The optimal mapping f for problem (8) under the MSE criterion is unique up to a sign, is an odd function of v, and is defined by the implicit equation*

$$f(v) e^{\frac{f(v)^2}{2\sigma_w^2}} = \frac{v}{\sqrt{2\pi}\sigma_w \lambda}. \tag{11}$$

Proof: See Appendix B.

Illustrations of the optimal mappings satisfying (11) will be given in Section VI. Here, we observe that, by expanding the Taylor series of the exponential function in (11), it can be easily verified that, in the low SNR regime, that is, as $\sigma_w^2 \to \infty$, the optimal mapping satisfies the condition $f(v) \propto v$, that is, it approaches a linear mapping. Furthermore, given that the optimal mapping $f(v)$ is odd, we can write

$$\hat{v}_{(1)} = \mathbb{E}[V|Y = 0] \tag{12a}$$

$$= \frac{\frac{1}{\sigma_v} \int v \Phi\left(\frac{v}{\sigma_v}\right) \Pr(Y = 0|V = v) dv}{\Pr(Y = 0)} \tag{12b}$$

$$= \frac{-2}{\sigma_v} \int v \Phi\left(\frac{v}{\sigma_v}\right) Q\left(\frac{f(v)}{\sigma_w}\right) dv \tag{12c}$$

$$= \frac{2\sigma_v}{\sqrt{2\pi}} - \frac{4}{\sigma_v} \int_0^{\infty} v \Phi\left(\frac{v}{\sigma_v}\right) Q\left(\frac{f(v)}{\sigma_w}\right) dv, \tag{12d}$$

and hence the average distortion can be simplified as

$$\bar{D} = \sigma_v^2 - \mathbb{E}[V\hat{V}] \tag{13a}$$

$$= \sigma_v^2 - \frac{1}{2}\left(\hat{v}_{(1)}\mathbb{E}[V|\hat{V} = \hat{v}_{(1)}] \right.$$

$$\left. + \hat{v}_{(2)}\mathbb{E}[V|\hat{V} = \hat{v}_{(2)}]\right) \tag{13b}$$

$$= \sigma_v^2 - \hat{v}_{(1)}^2, \tag{13c}$$

where (13a) is due to the orthogonality property of MMSE estimation; (13b) follows from the fact that the optimal encoder is odd; and (13c) is due to the chain of equalities $\mathbb{E}[V|\hat{V} = \hat{v}_{(1)}] = \mathbb{E}[V|Y = 0] = \hat{v}_{(1)} = -\hat{v}_{(2)} = -\mathbb{E}[V|\hat{V} = \hat{v}_{(2)}]$.

B. Linear Transmission for One-Bit Front End

Here we consider the performance of linear transmission in the presence of a one-bit ADC front end. The encoder mapping for linear transmission is given by

$$f(v) = \sqrt{\frac{P}{\sigma_v^2}} v. \tag{14}$$

As seen in Section III-A, linear transmission is asymptotically optimal in the low-SNR asymptotic regime. In the following, we elaborate on its performance for any given channel SNR γ. The MSE distortion $\bar{D}_l$ achieved by linear transmission, can be found by calculating the integral in (12d) with $f(v)$ given in (14) and substituting the result in (13c). As a result, the MSE distortion for the linear mapping is obtained as

$$\bar{D}_l = \sigma_v^2 \left(1 - \frac{2}{\pi} \left(\frac{\gamma}{\gamma + 1} \right) \right). \tag{15}$$

Note that for $\gamma = 0$, we have $\bar{D}_l = \sigma_v^2$. On the other hand, in the high SNR regime, i.e., as $\gamma \to \infty$, we obtain $\bar{D}_l = \sigma_v^2(1 - 2/\pi)$. Both distortions can be argued to be asymptotically optimal. In fact, for zero SNR, even with an infinite-resolution front end, the MMSE estimate is given by $\hat{V} = 0$, which yields $\bar{D} = \sigma_v^2$. Instead, for infinite SNR, the best mapping is given by the optimal binary quantizer, which yields $\bar{D} = \sigma_v^2(1 - 2/\pi)$ (see, e.g., [33, Section 10.1]).

C. Digital Transmission for One-bit Front End

Here we consider a conventional digital transmission scheme, which is based on quantizing and mapping the source to a discrete constellation for transmission over the channel. Accordingly, the source is quantized to one of the M levels, each characterized by the interval $[v_{(l-1)}, v_{(l)})$, $l = 1, ..., M$, where $v_{(M)} = \infty$, $v_{(0)} = -\infty$, and $v_{(l)} \geq v_{(l-1)}$ for all $l = 1, ..., M$. Each interval $[v_{(l-1)}, v_{(l)})$ is mapped to the corresponding channel input $X = x_{(l)}$. We take the constellation of possible transmission points to be $\{X = A(2l - 1 - M), l = 1, ..., M\}$, for some parameter $A \geq 0$, such that the average power constraint is satisfied. Note that, when M is even, this corresponds to the M-PAM modulation, while if M is odd, the constellation includes the zero-power signal, i.e., $x_{\left(\frac{M+1}{2}\right)} = 0$. The average transmission power can be written as $\mathbb{E}[X^2] = \sum_{l=1}^{M} x_{(l)}^2 \cdot \Pr(X = x_{(l)})$, where $\Pr(X = x_{(l)}) = \frac{1}{\sigma_v} \int_{v_{(l-1)}}^{v_{(l)}} \Phi(v/\sigma_v) \, dv$.

The average achievable distortion for M levels of symmetric digital transmission, i.e., $v_{(l)} = -v_{(M-l)}$, can be easily obtained as

$$\bar{D}_{d,M} = \sigma_v^2 - \left(\frac{2}{\sigma_v} \sum_{l=1}^{M} Q(S_l) \int_{v_{(l-1)}}^{v_{(l)}} v \Phi \left(\frac{v}{\sigma_v} \right) dv \right)^2, \tag{16}$$

where $S_l^2 = \frac{(2l-1-M)^2 \gamma}{\sum_{l=1}^{M} (2l-1-M)^2 \Pr(x_{(l)})}$.

As a special case, when $M = 2$, setting the quantization threshold as $v_{(1)} = 0$, we obtain BPSK transmission. The resulting achievable distortion can be computed from (16) as

$$\bar{D}_{d,2} = \sigma_v^2 \left(1 - \frac{2}{\pi} \left(1 - 2Q\left(\sqrt{\gamma}\right) \right)^2 \right). \tag{17}$$

We observe that, as for linear transmission, when $\gamma \to \infty$, we have $\bar{D}_{d,2} = \sigma_v^2(1 - 2/\pi)$, and when $\gamma \to 0$, we have $\bar{D}_{d,2} = \sigma_v^2$. From (17), one can check that the slope of the average distortion for BPSK transmission as $\gamma \to 0$ is $-4\sigma_v^2/\pi^2$, whereas the slope for linear transmission, obtained from (13c), is $-2\sigma_v^2/\pi$. This means that linear transmission has a decline in the distortion in the low SNR regime; therefore, it outperforms BPSK transmission, in this regime.

As another example, for $M = 3$, we set the quantization thresholds as $v_{(1)} = -c$ and $v_{(2)} = c$, so that $[-c, c]$ is the interval of source values for which the transmission symbol is $x_{(2)} = 0$. The MSE distortion can be computed by solving the following optimization problem with line search:

$$\bar{D}_{d,3} = \min_{c \geq 0} \; \sigma_v^2 \left(1 - \frac{2e^{-\frac{c^2}{\sigma_v^2}}}{\pi} \cdot \kappa_\gamma(c)^2 \right), \tag{18}$$

where $\kappa_\gamma(c) \triangleq 1 - 2Q\left(\sqrt{\frac{\gamma}{2Q\left(\frac{c}{\sigma_v}\right)}} \right)$.

D. K-Level Front End

In this section, we consider the system model in Figure 1 with a single observation, i.e., $N = 1$, but with a K-level ADC front end as in (3). As in the case of single one-bit observation, without loss of generality, we write the receiver mapping as

$$g(Y) = \hat{V} = \hat{v}_{(j)}, \; \text{if } Y = y_{(j)}, \tag{19}$$

for $j = 1, \ldots, K$. In the next proposition, we obtain a necessary optimality condition for the encoding and decoding functions f and g.

Proposition III.2. *The optimal encoder and decoder mappings f and g for the K-level ADC front end in (3) satisfy the necessary conditions*

$$f(v) = \frac{1}{2\sqrt{2\pi}\sigma_w \lambda} \sum_{j=1}^{K} \left(\hat{v}_{(j)} \left(2v - \hat{v}_{(j)} \right) \right.$$
$$\left. \cdot \left(e^{-\frac{\left(z_{(j-1)} - f(v)\right)^2}{2\sigma_w^2}} - e^{-\frac{\left(z_{(j)} - f(v)\right)^2}{2\sigma_w^2}} \right) \right), \tag{20}$$

and (19) with

$$\hat{v}_{(j)} = \frac{\int v \Phi\left(\frac{v}{\sigma_v}\right) \left(Q\left(\frac{z_{(j-1)} - f(v)}{\sigma_w}\right) - Q\left(\frac{z_{(j)} - f(v)}{\sigma_w}\right) \right) dv}{\int \Phi\left(\frac{v}{\sigma_v}\right) \left(Q\left(\frac{z_{(j-1)} - f(v)}{\sigma_w}\right) - Q\left(\frac{z_{(j)} - f(v)}{\sigma_w}\right) \right) dv}, \tag{21}$$

for $j = 1, ..., K$. Furthermore, the gradient of the Lagrangian function $L(f, g, \lambda)$ over f for g in (19) is given as

$$\nabla L = 2\lambda f(v) - \frac{1}{\sqrt{2\pi}\sigma_w} \sum_{j=1}^{K} \left(\hat{v}_{(j)} \left(2v - \hat{v}_{(j)} \right) \right.$$
$$\left. \cdot \left(e^{-\frac{\left(z_{(j-1)} - f(v) \right)^2}{2\sigma_w^2}} - e^{-\frac{\left(z_{(j)} - f(v) \right)^2}{2\sigma_w^2}} \right) \right). \quad (22)$$

Proof: See Appendix C.

As detailed later, the gradient in (22), along with (21), will be used in Section VI to obtain numerically optimized encoders and decoders.

IV. SINGLE OBSERVATION: DISTORTION OUTAGE PROBABILITY

In this section, we study the optimal encoder and decoder under the DOP criterion defined in (7) for the case of a single observation ($N = 1$). We first study the case of a one-bit front end in Section IV-A, and then we extend the results to a K-level front end in Section IV-B.

A. Optimal Encoder and Decoder for a One-Bit Front End

With no loss of generality, the decoder is given as in (10) for some reconstruction points $(\hat{v}_{(1)}, \hat{v}_{(2)})$. To proceed, we first focus on the optimization of the encoder mapping f for a given decoder in (10). We then tackle the problem of minimizing the DOP over the reconstruction points $(\hat{v}_{(1)}, \hat{v}_{(2)})$.

To elaborate, we define the intervals

$$I_j \triangleq \{v : (v - \hat{v}_{(j)})^2 < D\}, \quad (23)$$

for $j = 1, 2$, which are depicted in Figure 2. Each interval I_j, corresponds to the set of source values that are within the allowed distortion D of the reconstruction point $\hat{v}_{(j)}$. The following claims hold: (*i*) For all source outputs v in the set $(I_1 \cup I_2)^C = \{v : \min_{j=1,2}(v - \hat{v}_{(j)})^2 > D\}$, outage occurs (superscript C denotes the complement set). We refer to this event as *source outage*. (*ii*) For all source values in the interval $I_1 \cap I_2$, either of the reconstruction points yield a distortion no more than the target value D. Therefore, regardless of which of the two reconstruction levels, $\hat{v}_{(1)}$ and $\hat{v}_{(2)}$), is selected by the receiver, no outage occurs. From observations (*i*) and (*ii*), it easily follows that, for all source values v inside the intervals $(I_1 \cup I_2)^C$ and $(I_1 \cap I_2)$, the optimal mapping is $f(v) = 0$, since, for both intervals, the occurrence of an outage event is independent of the transmitted signal.

From the discussion above, we only need to specify the optimal mapping for the intervals $I_1 \setminus I_2$ and $I_2 \setminus I_1$. This should be done by accounting not only for the *source outage* event mentioned above, but also for the *channel outage* events. In particular, the distortion outage probability $\epsilon(D)$ can be written as

$$\epsilon(D) = \Pr\left(V \in (I_1 \cup I_2)^C \right)$$
$$+ \Pr\left(V \in (I_1 \setminus I_2), \hat{V} = \hat{v}_{(2)} \right)$$
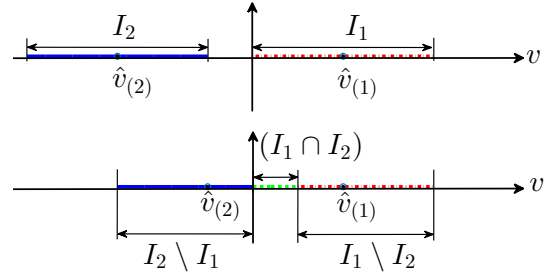

Figure 2: Illustration of the intervals I_1, I_2 and $(I_1 \cap I_2)$ that characterize the optimal encoder for the DOP criterion, for two different cases depending on the $(\hat{v}_{(1)}, \hat{v}_{(2)})$ values.

$$+ \Pr\left(V \in (I_2 \setminus I_1), \hat{V} = \hat{v}_{(1)} \right), \quad (24)$$

where the first term accounts for the source outage event, while the second and third terms are the probabilities of outage due to channel transmission errors. For instance, the second term is the probability that the decoder selects $\hat{V} = \hat{v}_{(2)}$ while V is in the interval $I_1 \setminus I_2$ (see Figure 2). The next proposition characterizes the optimal encoder mapping.

Proposition IV.1. *Given a target distortion D, and arbitrary reconstruction points $\hat{v}_{(1)}$ and $\hat{v}_{(2)}$, the optimal mapping f for the problem (8) is given by*

$$f(v) = \begin{cases} 0 & v \in (I_1 \cup I_2)^C \cup (I_1 \cap I_2), \\ -u & v \in (I_2 \setminus I_1), \\ u & v \in (I_1 \setminus I_2), \end{cases} \quad (25)$$

where u is the unique solution of

$$u e^{\frac{u^2}{2\sigma_w^2}} = \frac{1}{2\sqrt{2\pi}\sigma_w\lambda}. \quad (26)$$

Proof: See Appendix D.

We note here that, for given $\lambda \geq 0$, the optimal u is independent of the values of $\hat{v}_{(1)}$ and $\hat{v}_{(2)}$. Examples of optimal encoders will be provided in Section VI. In the next proposition, we turn to the optimization of the reconstruction levels $(\hat{v}_{(1)}, \hat{v}_{(2)})$.

Proposition IV.2. *The optimal reconstruction points $(\hat{v}_{(1)}, \hat{v}_{(2)})$, are given by*

$$\hat{v}_{(1)} = \sqrt{D} - a^*, \quad (27a)$$
$$\hat{v}_{(2)} = -\hat{v}_{(1)}, \quad (27b)$$

where a^* is obtained from

$$a^* = \underset{a \in [0, \sqrt{D}]}{\arg\min} \ 2Q\left(\frac{2\sqrt{D} - a}{\sigma_v} \right) + 2\left(Q\left(\frac{u}{\sigma_w} \right) + \lambda u^2 \right)$$
$$\cdot \left(Q\left(\frac{a}{\sigma_v} \right) - Q\left(\frac{2\sqrt{D} - a}{\sigma_v} \right) \right), \quad (28)$$

where u is obtained by solving (26).

Proof: See Appendix E.

To summarize, the optimal encoder and decoder are obtained as follows. First, given the Lagrange multiplier $\lambda \geq 0$, the

value of u is obtained by solving (26). Then the decoder's reconstruction points $(\hat{v}_{(1)}, \hat{v}_{(2)})$ are computed from (27)-(28). Finally, the optimal encoder mapping is given by (25). The next remark elaborates on the optimal encoder and decoder in two asymptotic SNR regimes.

Remark IV.1. *If λ is large, i.e., in the low-SNR regime, from (26) we have $u \approx 0$. Also from (26) it can be verified that $\lambda u^2 = \frac{e^{-\frac{u^2}{\sigma_w^2}}}{8\pi\sigma_w^2 \lambda} \approx 0$. Hence, from (28) we obtain*

$$a^* \approx \underset{a \in [0, \sqrt{D}]}{\arg\min} \ Q\left(\frac{a}{\sigma_v}\right) + Q\left(\frac{2\sqrt{D} - a}{\sigma_v}\right), \qquad (29)$$

yielding $\hat{v}_{(1)} = \hat{v}_{(2)} = 0$, that is, $I_1 = I_2$ and $\epsilon(D) = 2Q\left(\frac{\sqrt{D}}{\sigma_v}\right)$. On the other hand, for small values of λ, corresponding to the high-SNR regime, the variable u becomes large. Also, from (26) we have the equality $\lambda = 1/\left(2\sqrt{2\pi}\sigma_w u e^{\frac{u^2}{2\sigma_w^2}}\right)$ and Hence the approximations $\lambda u^2 = u/\left(2\sqrt{2\pi}\sigma_w e^{\frac{u^2}{2\sigma_w^2}}\right) \approx 0$, and

$$a^* \approx \underset{a \in [0, \sqrt{D}]}{\arg\min} \ 2Q\left(\frac{2\sqrt{D} - a}{\sigma_v}\right), \qquad (30)$$

which yield distinct intervals I_1 and I_2 with $\hat{v}_{(1)} = -\hat{v}_{(2)} = \sqrt{D}$, and $\epsilon(D) = 2Q\left(\frac{2\sqrt{D}}{\sigma_v}\right)$.

B. K-Level Front End

Here we turn our attention to the case of K-level front end under the DOP criterion. With no loss of optimality, the decoder is given as in (19) for some reconstruction levels $\{\hat{v}_{(1)}, \ldots, \hat{v}_{(K)}\}$ to be optimized. For a subset $\mathcal{V} \subset \{\hat{v}_{(1)}, ..., \hat{v}_{(K)}\}$, let $I_{\mathcal{V}}$ be the set of source outputs v for which the quadratic distance between v and the reconstruction points in set $\mathcal{V}$ is less than D, while the quadratic distance with respect to the reconstruction points in $\mathcal{V}^C$ is larger than D. This set is defined as

$$I_{\mathcal{V}} = \bigcap_{\hat{v}_{(j)} \in \mathcal{V}} I_j \setminus \bigcup_{\hat{v}_{(j)} \in \mathcal{V}^C} I_j, \quad |\mathcal{V}| = 1, ..., K, \qquad (31)$$

where I_j is given as

$$I_j \triangleq \left\{ v : |v - \hat{v}_{(j)}|^2 < D \right\}. \qquad (32)$$

We also define the set $I_\emptyset$ corresponding to $\mathcal{V} = \emptyset$ as

$$I_\emptyset \triangleq \left(\bigcup_{\mathcal{V}:|\mathcal{V}|\neq 0} I_{\mathcal{V}} \right)^C, \qquad (33)$$

that is, the set of values $v \in \mathbb{R}$ that do not have a reconstruction value $\hat{v}_{(j)}$ within a distance $\sqrt{D}$. Note that the sets $\left\{ I_{\mathcal{V}} : \mathcal{V} \subset \{\hat{v}_{(1)}, ..., \hat{v}_{(K)}\} \right\}$ form a partition of the whole real line.

As for the single one-bit front end (24), the DOP depends on both source and channel outage events. Note that the source

outage occurs if $V \in I_\emptyset$, whereas the channel outage occurs when $V \notin I_\emptyset$ and $V \notin I_{\{\hat{v}_{(1)}, \ldots, \hat{v}_{(K)}\}}$, with no outage occurring when $V \in I_{\{\hat{v}_{(1)}, \ldots, \hat{v}_{(K)}\}}$. In the following proposition, we present necessary optimality conditions for the encoder and decoder mappings.

Proposition IV.3. *For a K-level ADC front end, the optimal encoder and decoder mappings f and g satisfy the necessary conditions*

$$f(v) = \frac{1}{2\lambda} G(f, v), \qquad (34)$$

where $G(f, v)$ is defined as

$$G(f, v) \triangleq$$
$$\begin{cases} 0 & v \in I_{\mathcal{V}} : |\mathcal{V}| = 0, K, \\ \sum_{j: \ \hat{v}_{(j)} \in \mathcal{V}} \frac{\xi(j, f(v))}{\sqrt{2\pi}\sigma_w} & v \in I_{\mathcal{V}} : \ |\mathcal{V}| = 1, \ldots, K-1, \end{cases} \qquad (35)$$

where $\xi(j, f(v)) \triangleq e^{-\frac{\left(z_{(j-1)} - f(v)\right)^2}{2\sigma_w^2}} - e^{-\frac{\left(z_{(j)} - f(v)\right)^2}{2\sigma_w^2}}$, and (19) with

$$\hat{v}_{(j)} = \arg\max_t \int_{t-\sqrt{D}}^{t+\sqrt{D}} \Phi\left(\frac{v}{\sigma_v}\right)$$
$$\cdot \left(Q\left(\frac{z_{(j-1)} - f(v)}{\sigma_w}\right) - Q\left(\frac{z_{(j)} - f(v)}{\sigma_w}\right) \right) dv. \qquad (36)$$

Furthermore, the gradient of the Lagrangian function $L(f, g, \lambda)$ over f for g in (19) is given as

$$\nabla L = 2\lambda f(v) - \frac{1}{2} G(f, v). \qquad (37)$$

Proof: See Appendix F.

As for Proposition III.2, in Section VI, we will use (36) and (37) to obtain numerically optimized encoders and decoders, that satisfy the necessary optimality conditions.

V. MULTIPLE OBSERVATIONS

In this section, we study the more general case in which the receiver has $N > 1$ noisy one-bit quantized observations, Y^N, of the transmitted source sample. For both the MSE distortion and the DOP criteria, without loss of generality, we write the receiver mapping as

$$g(Y^N) = \hat{V} = \hat{v}_{(j)}, \ \text{if} \ Y^N = b_j^N, \qquad (38)$$

where b_j^N is the N-length binary representation of the number $2^N - j$ for $j = 1, \ldots, 2^N$. Note that there are 2^N reconstruction levels $\hat{v}_{(j)}$, each of which corresponds to a different configuration of the received signal $Y^N \in \{0, 1\}^N$. We will denote the k-th element of the vector b_j^N by $b_j^N(k)$.

A. MSE Criterion

Recalling that the optimal decoding function under the MSE criterion is the MMSE estimator, the optimal decoder satisfies $\hat{v}_{(j)} = \mathbb{E}[V|Y^N = b_j^N]$. In the next proposition, we provide necessary conditions for the optimal encoder and decoder

mappings along with an expression for the gradient of the Lagrangian in (9) over the encoder mapping f.

Proposition V.1. *Given a front end with N one-bit ADCs, the optimal encoder and decoder mappings f and g, respectively, satisfy the necessary conditions*

$$f(v) = \frac{1}{2\lambda} \sum_{j=1}^{2^N} \Theta\left(N, f(v), j\right) \hat{v}_{(j)} \left(2v - \hat{v}_{(j)}\right), \qquad (39)$$

and (38) with

$$\hat{v}_{(j)} \triangleq \frac{\int v \Phi\left(\frac{v}{\sigma_v}\right) \prod_{i=1}^{N} Q\left(\frac{(-1)^{b_j^N(i)+1} f(v)}{\sigma_{w_i}}\right) dv}{\int \Phi\left(\frac{v}{\sigma_v}\right) \prod_{i=1}^{N} Q\left(\frac{(-1)^{b_j^N(i)+1} f(v)}{\sigma_{w_i}}\right) dv}, \qquad (40)$$

where $\Theta\left(N, f(v), j\right)$ is defined as

$$\Theta\left(N, f(v), j\right) \triangleq \sum_{k=1}^{N} \left(\frac{(-1)^{b_j^N(k)} e^{-\frac{f(v)^2}{2\sigma_{w_k}^2}}}{\sqrt{2\pi}\sigma_{w_k}} \right.$$
$$\left. \cdot \prod_{l=1,l\neq k}^{N} Q\left(\frac{(-1)^{b_j^N(l)+1} f(v)}{\sigma_{w_l}}\right) \right). \qquad (41)$$

Furthermore, the gradient of the Lagrangian function $L(f, g, \lambda)$ over f for g in (38) is given as

$$\nabla L = 2\lambda f(v) - \sum_{j=1}^{2^N} \hat{v}_{(j)} \Theta\left(N, f(v), j\right) \left(2v - \hat{v}_{(j)}\right). \qquad (42)$$

Proof: See Appendix G.

Remark V.1. *In order to obtain numerically optimized encoder (NOE) mappings in Section VI, we will adopt a gradient-descent algorithm. This method updates the current iterate $f^{(i)}(v)$ as*

$$f^{(i+1)}(v) = f^{(i)}(v) - \mu \nabla L, \qquad (43)$$

where i is the iteration index and $\mu > 0$ is the step size. ∇L is the derivative of the Lagrangian (8) with respect to the mapping f at $f^{(i)}$, which can be found in (22) and (42) for K-level ADC and multiple one-bit ADCs, respectively. The algorithm is initialized with the linear mapping specified in (14). It is noted that the algorithm is not guaranteed to converge to a global optimal solution.

B. DOP Criterion

In this subsection, we consider the DOP criterion. The analysis follows the same steps as in Section IV-B. In particular, we define the different intervals $I_\mathcal{V}$, where $\mathcal{V}$ is a subset of $\{\hat{v}_{(1)}, \ldots, \hat{v}_{(2^N)}\}$ as in (31). Based on this definition, in the next proposition, we derive necessary optimality conditions for the encoder and decoder mappings.

Proposition V.2. *Given a front end with N one-bit ADCs, the optimal encoder and decoder mappings f and g satisfy the necessary conditions*

$$f(v) = \frac{1}{2\lambda} \tilde{G}(f, v), \qquad (44)$$

where $\tilde{G}(f, v)$ is defined as

$$\tilde{G}(f, v) \triangleq$$
$$\begin{cases} 0 & v \in I_\mathcal{V} : |\mathcal{V}| = 0, 2^N, \\ \displaystyle\sum_{k: \hat{v}_{(k)} \in \mathcal{V}} \left(\sum_{i=1}^{N} \frac{(-1)^{b_k^N(i)} e^{-\frac{f(v)^2}{2\sigma_{w_i}^2}}}{\sqrt{2\pi}\sigma_{w_i}} \right. & v \in I_\mathcal{V} : |\mathcal{V}| = 1, \ldots, 2^N - 1, \\ \qquad \left. \cdot \prod_{\substack{l=1 \\ l\neq i}}^{N} Q\left(\frac{(-1)^{b_k^N(l)+1} f(v)}{\sigma_{w_l}}\right) \right) \end{cases}$$
$$\qquad (45)$$

and (38) with

$$\hat{v}_{(j)} = \arg\max_{t} \int_{t-\sqrt{D}}^{t+\sqrt{D}} \Phi\left(\frac{v}{\sigma_v}\right) \prod_{i=1}^{N} Q\left(\frac{(-1)^{b_j^N(i)+1} f(v)}{\sigma_{w_i}}\right) dv, \qquad (46)$$

for $j = 1, \ldots, 2^N$. Furthermore, the gradient of the Lagrangian function $L(f, g, \lambda)$ over f for g in (38) is given as

$$\nabla L = 2\lambda f(v) - \frac{1}{2} \tilde{G}(f, v). \qquad (47)$$

Proof: See Appendix H.

Remark V.2. *From the optimal decoders in (40) and (46), under the MSE and DOP criteria, respectively, it can be easily verified that, with equal noise variances for the N observations, i.e., $\sigma_{w_i}^2 = \sigma_w^2$ for $i = 1, ..., N$, the reconstruction points $\hat{v}_{(j)}$ corresponding to b_j^N vectors with the same number of ones are equal. As a consequence, there are only $N + 1$ effective reconstruction points instead of 2^N.*

Remark V.3. *In Section VI, we obtain NOE mappings using the gradient-descent method as follows:*

1) *Initialize the set of reconstruction points $\{\hat{v}_{(j)}\}$, $j = 1, \ldots, 2^B$;*
2) *Optimize the encoder mapping corresponding to the decoder (19) and (38) for K-level ADC and multiple one-bit ADCs, respectively, for the given $\{\hat{v}_{(j)}\}$, $j = 1, \ldots, 2^B$, using the gradient-descent algorithm (43), with ∇L defined as in (37) for $K > 2$, and as in (47) for $N > 1$;*
3) *Find the optimal reconstruction points corresponding to the obtained encoder mapping using (36) for $K > 2$, and (46) for $N > 1$;*
4) *If convergence is not obtained, go back to step 2.*

VI. NUMERICAL RESULTS

In this section, we provide some illustrations of the results derived above by means of numerical examples. We consider
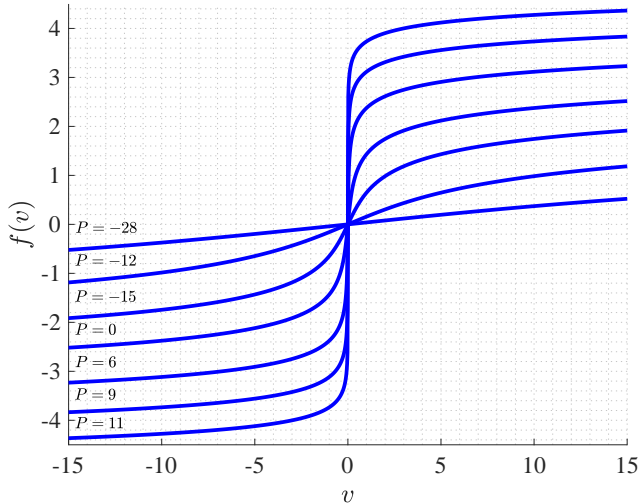
Figure 3: Optimal encoder mappings for the MSE distortion criterion under different power constraints P (dB) ($\sigma_w^2 = \sigma_v^2 = 1$).

Figure 4: Complement of the MSE $(1 - \bar{D})$ (dB) versus the SNR (γ) (dB) for the optimal encoder obtained in Proposition III.1 compared with the linear and digital schemes studied in Sections III-A, III-B and III-C ($\sigma_w^2 = \sigma_v^2 = 1$).

the MSE distortion and the DOP in separate subsections. Throughout this section, we set $\sigma_v^2 = \sigma_w^2 = 1$, so that the SNR in (5) is proportional to the power constraint P.

A. MSE Criterion

We start by considering the MSE distortion in a single measurement system ($N = 1$) with a one-bit ADC front end ($K = 2$). We then investigate alternative front ends with more levels ($K > 2$) or more observations ($N > 1$) and present a performance comparison of different front end architectures.

For the case $N = 1$ and $K = 2$, Figure 3 shows the optimal mappings obtained from Proposition III.1 for different values of P. The value of the Lagrange multiplier λ in (11) is obtained by means of bisection so as to satisfy the power constraint $\mathbb{E}[f(V)^2] = P$. Figure 3 shows that the optimal mapping has a linear behaviour around the origin, and that it saturates as the absolute value of the source sample increases. As discussed in Section III-A, for low SNR, the saturation occurs only for very large absolute values of the source; and therefore, the numerically obtained solution for (11) tends to a linear function, while, for high SNR, the numerically obtained solution tends to a step function, resembling digital transmission.

The conclusions above are corroborated by Figure 4, in which the MSE distortion of the optimal, linear and digital transmission schemes are plotted versus the SNR γ. For clarity of illustration, we plot the complementary MSE distortion $1 - \bar{D}$, where we note that $\bar{D} = \sigma_v^2 = 1$ is achievable by setting $\hat{V} = 0$ irrespective of the received signal. The figure confirms that linear transmission approaches optimality at low SNR, whereas, for high SNR, digital schemes outperform linear transmission. It is also seen that, for digital transmission, increasing the number of constellation points generally
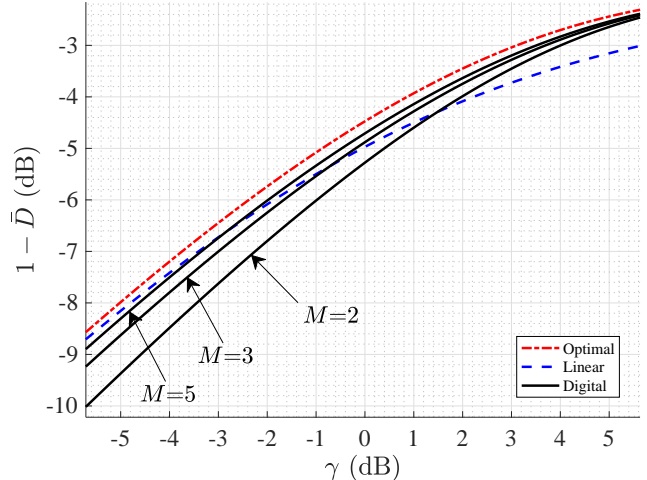
improves the performance, although, in the high SNR regime, binary transmission is sufficient to achieve the minimum MSE.

We now investigate the MSE performance with the K-level ADC ($N = 1$, $K > 2$) front end studied in Section III-D and with the multiple one-bit ADCs front end ($N > 1$, $K = 2$) considered in Section V-A. For these architectures, for which Proposition III.2 and Proposition V.1 provide respective necessary optimality conditions, we resort to a gradient-descent approach, as described in Remark V.1.

We now first illustrate the NOE mappings for the two types of front end architectures and then provide a performance comparison. In Figure 5, the obtained NOE mappings are illustrated for a $K = 4$ and a $K = 8$ level ADC front end, respectively, under the MSE distortion criterion. The decision thresholds for $K = 4$ and $K = 8$ are chosen as $[\infty, -d, 0, d, \infty]$ and as $[-\infty, -3d, -2d, -d, 0, d, 2d, 3d, \infty]$, respectively, and the parameter d is optimized by means of a line search. From the results obtained in Figure 5, it is noticed that, in the low SNR regime, the NOE mappings for both $K = 4$ and $K = 8$ approach linear transmission as discussed in Section III-A. Instead, as the SNR γ increases, they resemble digital mappings with K constellation points.

Considering now the alternative architecture with N ADCs each with $K = 2$ quantization levels, Figure 6 illustrates the NOE mappings for $N = 2$ and $N = 7$ under different power constraints. It is again observed that, when the SNR is low, the NOE mappings approach linear transmission. With increasing SNR the NOE mapping exhibits a saturation behaviour, although, as opposed to the $N = 1$ case in Figure 3, the linear behaviour around the origin does not disappear completely.

We now compare the two front end architectures, under the constraint that they both provide B output bits per received sample. The first architecture uses all bits to quantize a single
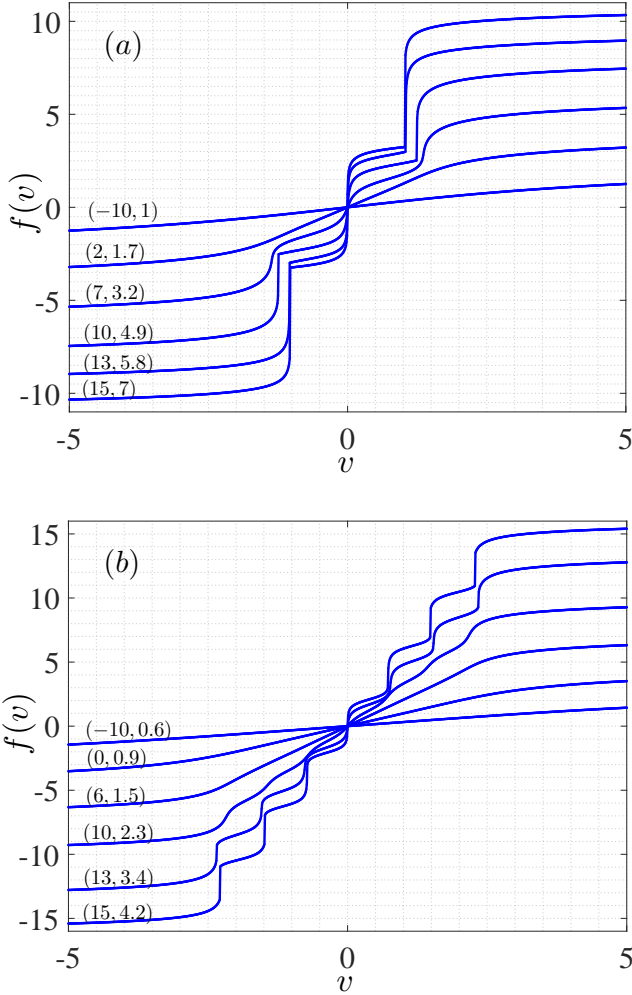
Figure 5: NOE mappings for different power constraints (dB) and under the MSE distortion criterion ($\sigma_w^2 = \sigma_v^2 = 1$): (a) $K = 4$; (b) $K = 8$. The curves are labelled by the pair (P, d), where P is the average power constraint (dB) and d is the quantization step size of the ADC front end.

Figure 6: NOE mappings for different power constraints (dB) and under the MSE distortion criterion $\sigma_w^2 = \sigma_v^2 = 1$: (a) $N = 2$, $K = 2$; (b), $N = 7$, $K = 2$. The curves are labelled by their corresponding average power constraint (dB).

observation, i.e., $N = 1$ and $K = 2^B$, while the second one outputs B one-bit measurements, i.e., $N = B$, $K = 2$. In Figure 7, the achievable MSE distortion is shown for the two architectures for $B = 1$, 2, 3, along with the Shannon bound [18, Equation 21], which corresponds to the theoretical performance limit for $N = 1$ and $K = \infty$. It is seen that, for the same number of bits per sample, B, as the SNR (γ) increases, the 2^B-level ADC architecture outperforms the receiver with B one-bit ADCs, whereas, the opposite is true for low SNR. This shows that, for high SNR, it is more beneficial to invest additional output bits in improving the ADC resolution. In contrast, for low SNR, it is preferable to increase the number of observations in order to improve the effective SNR by collecting independent measurements of the transmitted signal. We also note that, we found this conclusion to hold when the threshold values of the N one-bit ADCs
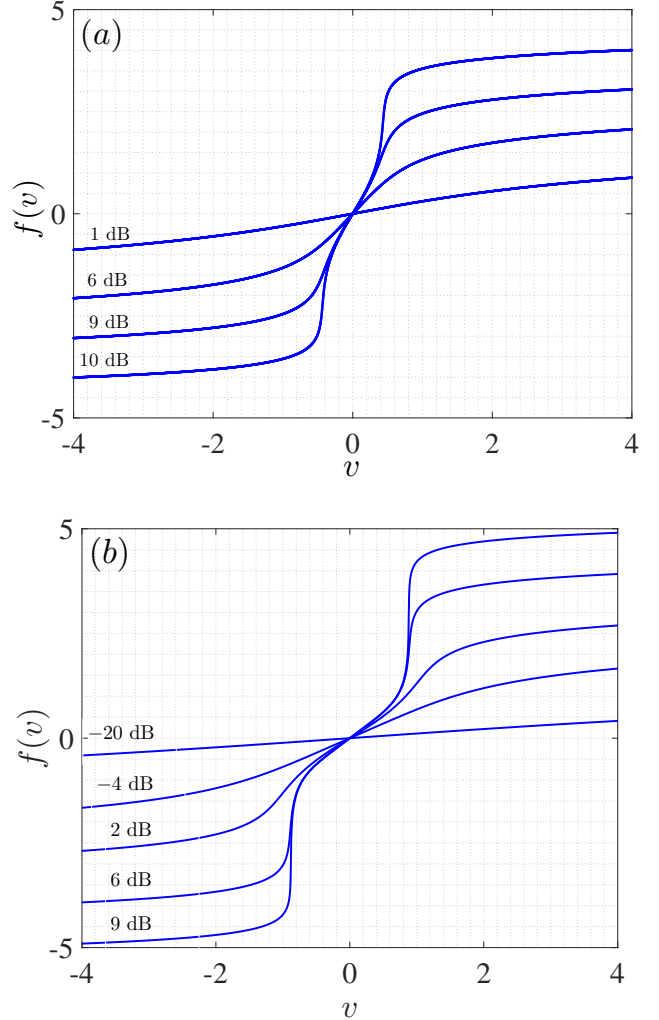
are allowed to be distinct, and optimized. Similar observations have also been made when the mappings are restricted to be linear (not shown here).

B. DOP Criterion

Here, we study the optimal mapping and performance under the DOP criterion. We start by considering the case of a one-bit ADC front end with a single observation, i.e., $N = 1$, $K = 2$. The optimal mapping along with the corresponding optimal reconstruction points $(\hat{v}_{(1)}, \hat{v}_{(2)})$ for three different values of the power constraint P are shown in Figure 8, as obtained in Propositions IV.1 and IV.2. It is seen that, as the SNR decreases, the optimal reconstruction points, $\hat{v}_{(1)}$ and $\hat{v}_{(2)}$, tend to zero, while, for high SNR, they tend to $\hat{v}_{(1)} = -\hat{v}_{(2)} = \sqrt{D}$, as per Remark IV.1. This observation can be explained as follows. For low SNR, the DOP is dominated by the probability of channel outage, i.e., by the last two terms in (24), which
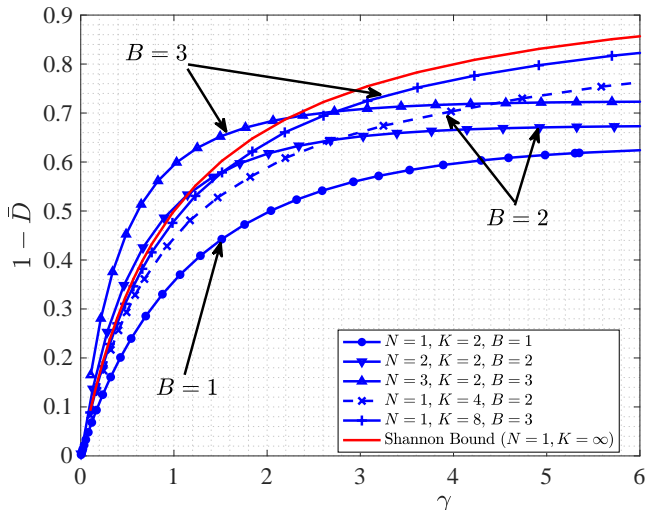
Figure 7: Complement of the MSE $(1-\bar{D})$ vs. SNR (γ) (linear scale) for different ADC architectures with B-bit outputs ($\sigma_v^2 = 1$). The variance of the AWGN in all the scenarios and for all the observations is one ($\sigma_w^2 = 1$).

Figure 8: Optimal encoder mappings for $N = 1$, $K = 2$ under the DOP criterion, for different values of the SNR (γ). The dots are the reconstruction points $\hat{v}_{(1)}$ and $\hat{v}_{(2)}$ ($\sigma_w^2 = \sigma_v^2 = 1$ and $D = 0.3$).

are zero for $\hat{v}_{(1)} = \hat{v}_{(2)} = 0$. In contrast, for high SNR, the optimal solution aims at minimizing the probability of source outage events, i.e., the first term in (24), which requires $\hat{v}_{(1)} = -\hat{v}_{(2)} = \sqrt{D}$.

Continuing the analysis of the $N = 1$, $K = 2$ case, in Figure 9, the complement of the DOP, $1-\epsilon(D)$, is plotted with respect to the SNR for different values of D. As SNR decreases, based on the discussion above and Remark IV.1, DOP tends to the first term in (24) when $\hat{v}_{(1)} = \hat{v}_{(2)} = 0$, which can be computed as $Q(\sqrt{D}/\sigma_v)$. Furthermore, as the SNR increases, DOP tends to the first term in (24) but with $\hat{v}_{(1)} = -\hat{v}_{(2)} = \sqrt{D}$, resulting in $\epsilon(D) = Q(2\sqrt{D}/\sigma_v)$, indicated by the dashed lines in Figure 9.

In order to derive NOE mapping under the DOP criterion for $K > 2$ or $N > 1$, we apply iterative gradient-descent based algorithm based on the results obtained in Propositions IV.3 and V.2 as described in Remark V.3.

For low SNR, based on the results in Section IV-A (see also Figure 8) the reconstruction points are close to zero; and hence, we can initialize the algorithm with all-zero values when λ is very large. Therefore, we first set a large value for λ and consider all-zero vector as the initial mapping. Then, we consider successively smaller values of λ, i.e., increase the SNR. We use the reconstruction points obtained for the previous value of the Lagrange multiplier λ to initialize the algorithm for the current value of λ. This approach is known as noise channel relaxation (NCR) [34].

In Figure 10, the complement of the DOP, $1 - \epsilon(D)$, is shown for two architectures with $B = 2$ output bits, namely one observation with 4-level ADC ($N = 1$, $K = 4$), and two observations with one-bit ADCs ($N = 2$, $K = 2$), as well as for the architecture with a one-bit ADC ($N = 1$, $K = 2$), for a target distortion of $D = 0.09$. In accordance with the
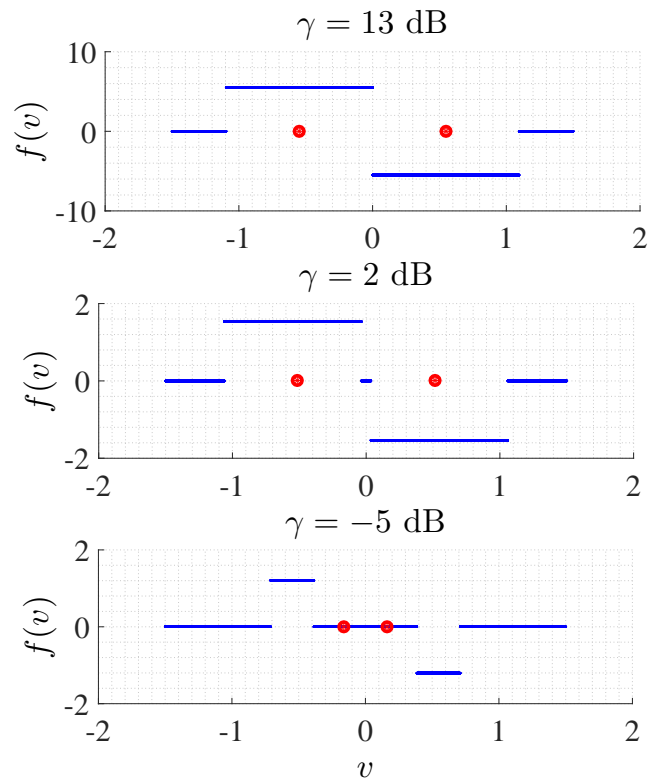
discussion above, all architectures have the same performance for low SNR, namely $\epsilon(D) = 2Q(\sqrt{D}/\sigma_v) = 2Q(\sqrt{0.09})$. Furthermore, in a manner similar to the discussion on Figure 7 for the MSE criterion, in the low SNR regime, it is beneficial to increase the number of observations, whereas at high SNR, it is preferable to increase the ADC resolution. In this regard, we note that, in the high SNR the optimal $K = 4$ levels are selected to minimize the probability of source outage, that is, the probability $\Pr(V \in I_\emptyset)$, yielding the reconstruction points $\hat{V} = \{-3\sqrt{D}, -\sqrt{D}, \sqrt{D}, 3\sqrt{D}\}$ and a DOP of $2Q(4\sqrt{D}/\sigma_v) = 2Q(1.2) = 0.2301$. Instead, for $N = 2$ and $K = 2$, the three effective reconstruction points (see Remark V.2) at high SNR tend to $\hat{V} = \{-2\sqrt{D}, 0, 2\sqrt{D}\}$ and the minimum DOP is lower bounded by $Q(3\sqrt{D}/\sigma_v) = 2Q(.9) = 0.3681$.

VII. Conclusions and Future Work

We have considered the zero-delay transmission of a single sample of a Gaussian source over a quantized SIMO AWGN channel. We first studied a system with a one-bit ADC front end at the receiver under two distinct performance criteria, namely the mean square error (MSE) distortion and the distortion outage probability (DOP). For the MSE distortion, we have shown that the optimal encoder mapping is odd, and that, in the low SNR regime, linear transmission approaches the optimal performance, whereas digital transmission becomes
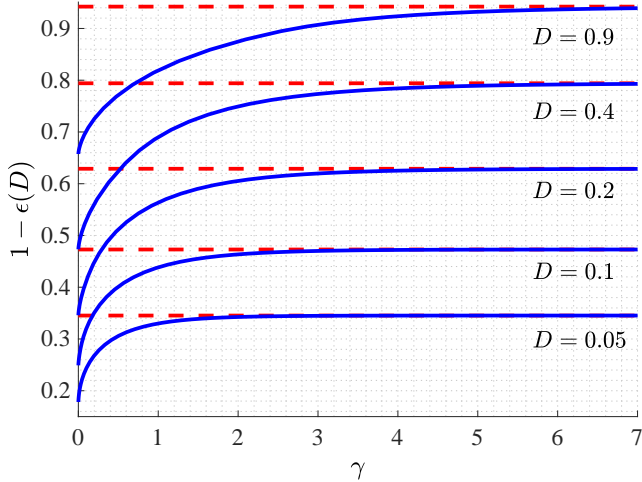
Figure 9: Complement of the DOP $(1-\epsilon(D))$ versus SNR (γ) in linear scale. Dashed lines represent the DOP in the high SNR regime for the corresponding distortion target $(\sigma_w^2 = \sigma_v^2 = 1)$.
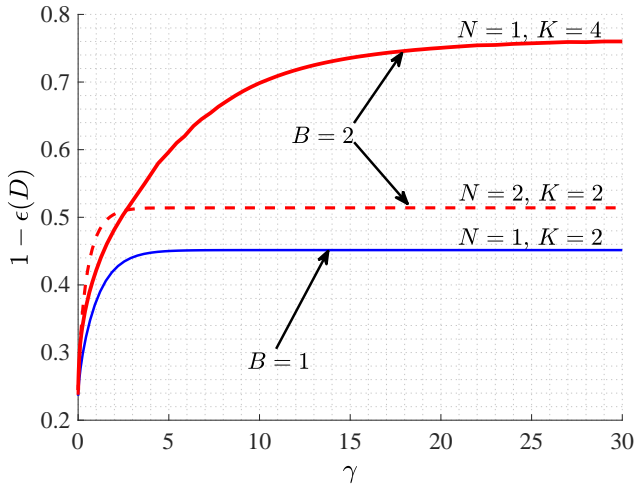

Figure 10: Complement of the DOP $(1 - \epsilon(D))$ versus SNR (γ) in linear scale for different ADC architectures with B-bit outputs $(\sigma_w^2 = \sigma_v^2 = 1)$.

optimal in the high SNR regime. For the DOP criterion, we have obtained the optimal structure of the encoder and the decoder, demonstrating that the optimal encoder function is symmetric and piecewise constant. For both the MSE distortion and the DOP criteria, we also derived necessary optimality conditions for the encoder and decoder mappings for a K-level ADC front end and for multiple one-bit ADC observations.

Among open problems that are left for future research, we mention here the joint optimization of the quantization intervals of the ADC front end with the encoding and decoding functions. Another interesting problem is the zero-delay joint source-channel coding over fading channels with finite-

resolution ADCs.

VIII. Appendices

A. Preliminaries: Calculus of Variations

In the proofs of the propositions reported above, we leverage the standard method in variational calculus to obtain necessary optimality conditions [35, Section 7]. The following lemma presents the key result that will be used throughout the following appendices. For the sake of brevity, we drop the arguments of functions and functionals where no confusion can arise. We also use the notation F^f and $F^{f'}$ to denote the partial derivatives of a functional F with respect to functions f and f', respectively.

Lemma A.1. *Let G_i, $i = 1, \ldots, n$, be continuous functionals of (f, f', u) and have continuous partial derivatives with respect to (f, f'). Also, let F be a continuous functional of $(f, f', r_1, \ldots, r_n, u)$, where r_i is a functional of G_i given as*

$$r_i = \int_{t_1}^{t_2} G_i(f(t), f'(t), t)dt, \quad i = 1, \ldots, n. \tag{48}$$

Let F has continuous partial derivatives with respect to $(f, f', r_1, \ldots, r_n)$. Consider the following minimization problem

$$\underset{f}{minimize} \quad L(f) \triangleq \int_{t_1}^{t_2} F(f(t), f'(t), r_1, \ldots, r_n, t)dt. \tag{49}$$

Define the functional derivative ∇L as

$$
\begin{aligned}
\nabla L \triangleq & F^f(f(u), f'(u), r_1, \ldots, r_n, u) \\
& - \frac{d}{du} F^{f'}(f(u), f'(u), r_1, \ldots, r_n, u) \\
& + \sum_{i=1}^{n} \Bigg(\bigg(G_i^f(f(u), f'(u), u) - \frac{d}{du} G_i^{f'}(f(u), f'(u), u) \bigg) \\
& \qquad \cdot \int F^{r_i}(f(t), f'(t), r_1, \ldots, r_n, t)dt \Bigg),
\end{aligned}
\tag{50}
$$

where F^{r_i} is the derivative of the functional F with respect to r_i. Similarly, $G_{i,}^f, G_i^{f'}$ are derivatives of the functional G_i with respect to f, f', respectively. A necessary condition for a function f to be a solution of the problem (49) is

$$\nabla L = 0, \ \forall u \in [t_1, t_2]. \tag{51}$$

Proof: Following the conventional approach in the calculus of variations, we perturb the function $f(t)$ by an arbitrary function $\eta(t)$, which vanishes on the boundary points t_1 and t_2 [35]. Let $\delta_f L \triangleq \frac{dL(f+\alpha\eta)}{d\alpha}\Big|_{\alpha=0}$ be the Gateaux derivative of the functional L with respect to the parameter α. We have

$$\delta_f L =$$

$$\frac{d}{d\alpha} \int_{t_1}^{t_2} F(f(t) + \alpha\eta(t), f'(t) + \alpha\eta'(t), r_1^{\alpha}, \ldots, r_n^{\alpha}, t)dt \Bigg|_{\alpha=0},$$

$$\tag{52}$$

where r_i^α is defined as

$$r_i^\alpha = \int_{t_1}^{t_2} G_i(f(u) + \alpha\eta(u), f'(u) + \alpha\eta'(u), u)du, \quad (53)$$

for $i = 1, \ldots, n$. Note that in (52), the order of the derivative with respect to α and the integral can be exchanged due to the Lebesgue dominated convergence theorem [36]. Therefore, the Gateaux derivative $\delta_f L$ can be written as

$$\delta_f L = \int_{t_1}^{t_2} \Big[\eta(t) F^f(f(t), f'(t), r_1, \ldots, r_n, t)$$
$$+ \eta'(t) F^{f'}(f(t), f'(t), r_1, \ldots, r_n, t) \Big] dt$$
$$+ \int_{t_1}^{t_2} \sum_{i=1}^{n} \frac{dr_i^\alpha}{d\alpha}\Big|_{\alpha=0} \cdot F^{r_i}(f(t), f'(t), r_1, \ldots, r_n, t) dt \quad (54)$$
$$= \int_{t_1}^{t_2} \eta(t) \Big[F^f(f(t), f'(t), r_1, \ldots, r_n, t)$$
$$- \frac{d}{dt} F^{f'}(f(t), f'(t), r_1, \ldots, r_n, t) \Big] dt$$
$$+ \underbrace{F^{f'}(f(t), f'(t), r_1, \ldots, r_n, t)\eta(t)\Big|_{t_1}^{t_2}}_{=0}$$
$$+ \int_{t_1}^{t_2} \sum_{i=1}^{n} \frac{dr_i^\alpha}{d\alpha}\Big|_{\alpha=0} \cdot F^{r_i}(f(t), f'(t), r_1, \ldots, r_n, t) dt,$$
$$(55)$$

where (55) is due to integration by parts, and the fact that $\eta(t)$ vanishes at t_1 and t_2 by construction. We compute

$$\frac{dr_i^\alpha}{d\alpha}\Big|_{\alpha=0} = \frac{d}{d\alpha} \int_{t_1}^{t_2} G_i(f(u) + \alpha\eta(u), f'(u) + \alpha\eta'(u), u)du \Big|_{\alpha=0}$$
$$(56)$$
$$= \int_{t_1}^{t_2} \Big[\eta(u) G_i^f(f(u), f'(u), u)$$
$$+ \eta'(u) G_i^{f'}(f(u), f'(u), u) \Big] du \quad (57)$$
$$= \int_{t_1}^{t_2} \eta(u) \Big[G_i^f(f(u), f'(u), u)$$
$$- \frac{d}{du} G_i^{f'}(f(u), f'(u), u) \Big] du$$
$$+ G_i^{f'}(f(u), f'(u), u)\eta(u)\Big|_{t_1}^{t_2} \quad (58)$$
$$= \int_{t_1}^{t_2} \eta(u) \Big[G_i^f(f(u), f'(u), u)$$

$$- \frac{d}{du} G_i^{f'}(f(u), f'(u), u) \Big] du, \quad i = 1, \ldots, n. \quad (59)$$

By plugging (59) into (55) we finally have

$$\delta_f L = \int_{t_1}^{t_2} \eta(u) \Big[F^f(f(u), f'(u), r_1, \ldots, r_n, u)$$
$$- \frac{d}{du} F^{f'}(f(u), f'(u), r_1, \ldots, r_n, u) \Big] du$$
$$+ \sum_{i=1}^{n} \Bigg[\int_{t_1}^{t_2} F^{r_i}(f(t), f'(t), r_1, \ldots, r_n, t) dt$$
$$\cdot \int_{t_1}^{t_2} \eta(u) \Big(G_i^f(f(u), f'(u), u)$$
$$- \frac{d}{du} G_i^{f'}(f(u), f'(u), u) \Big) du \Bigg] \quad (60)$$
$$= \int_{t_1}^{t_2} \eta(u) \Bigg[F^f(f(u), f'(u), r_1, \ldots, r_n, u)$$
$$- \frac{d}{du} F^{f'}(f(u), f'(u), r_1, \ldots, r_n, u)$$
$$+ \sum_{i=1}^{n} \Bigg(\Big(G_i^f(f(u), f'(u), u) - \frac{d}{du} G_i^{f'}(f(u), f'(u), u) \Big)$$
$$\cdot \int_{t_1}^{t_2} F^{r_i}(f(t), f'(t), r_1, \ldots, r_n, t) dt \Bigg) \Bigg] du. \quad (61)$$

Since $\eta(u)$ is an arbitrary function and the term multiplying $\eta(u)$ is continuous, it must be zero everywhere on the interval $[t_1, t_2]$. Thus, the optimal solution must satisfy the following equality

$$F^f(f(u), f'(u), r_1, \ldots, r_n, u)$$
$$- \frac{d}{du} F^{f'}(f(u), f'(u), r_1, \ldots, r_n, u)$$
$$+ \sum_{i=1}^{n} \Bigg(\Big(G_i^f(f(u), f'(u), u) - \frac{d}{du} G_i^{f'}(f(u), f'(u), u) \Big)$$
$$\int_{t_1}^{t_2} F^{r_i}(f(t), f'(t), r_1, \ldots, r_n, t) dt \Bigg) = 0, \quad \forall u \in [t_1, t_2], \quad (62)$$

which yields (51).

$\blacksquare$

Remark A.1. *When the functional F is independent of r_i, $i = 1, \ldots, n$, then $F^{r_i}(f, f', u) = 0$, and therefore, we obtain the well known Euler-Lagrange equation*

$$F^f(f, f', u) - \frac{d}{du} F^{f'}(f, f', u). \quad (63)$$

Remark A.2. *As a special case of Lemma (A.1), it will be useful in Appendices B, C and G to consider the minimization of the functional*

$$L(f) = \frac{1}{\sigma_v} \int_{t_1}^{t_2} \Phi\left(\frac{t}{\sigma_v}\right) \tilde{F}(f(t), r_1, \dots, r_n, t) dt, \qquad (64)$$

where we recall that $\Phi(\cdot)$ is the Gaussian probability density function with mean zero and variance one, and r_i, $i = 1, \dots, n$, are of the form given by

$$r_i = \frac{1}{\sigma_v} \int_{t_1}^{t_2} \Phi\left(\frac{t}{\sigma_v}\right) \tilde{G}_i(f(t), t) dt, \quad i = 1, \dots, n. \qquad (65)$$

From Lemma A.1, setting

$$F = \frac{1}{\sigma_v} \Phi\left(\frac{t}{\sigma_v}\right) \tilde{F}(f(t), r_1, \dots, r_n, t), \qquad (66)$$

$$G_i = \frac{1}{\sigma_v} \Phi\left(\frac{t}{\sigma_v}\right) \tilde{G}_i(f(t), t), \ i = 1, \dots, n, \qquad (67)$$

the solution for this problem needs to satisfy $\nabla L = 0$, where the functional derivative ∇L is found as

$$\nabla L \triangleq \tilde{F}^f(f(u), r_1, \dots, r_n, u) + \sum_{i=1}^{n} \left(\tilde{G}_i^f(f(u), u) \right.$$
$$\left. \cdot \int \frac{1}{\sigma_v} \Phi\left(\frac{t}{\sigma_v}\right) \tilde{F}^{r_i}(f(t), r_1, \dots, r_n, t) dt \right), u \in [t_1, t_2], \qquad (68)$$

with $\tilde{G}_i$, $i = 1, \dots, n$, and $\tilde{F}$ being continuous functionals of (f, t) and $(f, r_1, \dots, r_n, t)$, respectively, and having continuous partial derivatives with respect to f, and $(f, r_1, \dots, r_n)$, respectively.

Remark A.3. *In Appendices B, D, F and H we will consider the minimization of the functional*

$$L(f) = \frac{1}{\sigma_v} \int_{t_1}^{t_2} \Phi\left(\frac{t}{\sigma_v}\right) \tilde{F}(t, f(t)) dt. \qquad (69)$$

From Remark A.1, setting

$$F(f, t) = \frac{1}{\sigma_v} \Phi\left(\frac{t}{\sigma_v}\right) \tilde{F}(f(t), t), \qquad (70)$$

the solution for the minimization of the problem (69) needs to satisfy

$$\nabla L \triangleq \tilde{F}^f(u, f(u)) = 0, \ u \in [t_1, t_2], \qquad (71)$$

where $\tilde{F}$ is continuous with respect to f and has continuous partial derivatives with respect to f.

B. PROOF OF PROPOSITION III.1

As discussed in Section III, the MMSE estimator $g(\cdot) = \mathbb{E}[V|\cdot]$ is optimal for any encoder mapping f. Due to the orthogonality principle of the MMSE estimation, we can write $\bar{D} = \sigma_v^2 - \mathbb{E}[V\hat{V}]$. Rewriting the Lagrangian in (9) for the MSE distortion criterion, and dropping the constants that are independent of f, we have

$$L_s(f, g, \lambda) \triangleq L(f, g, \lambda) - \sigma_v^2 \qquad (72)$$
$$= -\mathbb{E}[V\hat{V}] + \lambda \mathbb{E}[f(V)^2]. \qquad (73)$$

In the following, we prove the proposition by means of three key lemmas. In the first, using (68), we obtain a necessary condition for the optimal encoder mapping. Then, using this necessary condition, we show that the Lagrangian function in (73) takes its minimum value when f is odd. Finally, tackling the optimization problem in (8) over odd functions, and using (71), we obtain the result of Proposition III.1.

Lemma B.1. *The optimal encoder function f for the problem (8), has to satisfy*

$$f(v) e^{\frac{f(v)^2}{2\sigma_w^2}} = a_f(v + b_f), \qquad (74)$$

where a_f and b_f are defined as

$$a_f \triangleq \frac{-r_1}{\sqrt{2\pi}\sigma_w \lambda r_2(1 - r_2)}, \qquad (75a)$$

$$b_f \triangleq -\frac{(1 - 2r_2)r_1}{2r_2(1 - r_2)}, \qquad (75b)$$

with r_1 and r_2 defined as

$$r_1 \triangleq \frac{1}{\sigma_v} \int v\Phi\left(\frac{v}{\sigma_v}\right) Q\left(\frac{f(v)}{\sigma_w}\right) dv, \qquad (76a)$$

$$r_2 \triangleq \frac{1}{\sigma_v} \int \Phi\left(\frac{v}{\sigma_v}\right) Q\left(\frac{f(v)}{\sigma_w}\right) dv. \qquad (76b)$$

Proof: Expanding the objective function $L_s(f, g, \lambda)$ in (73), we have

$$L_s(f, g, \lambda) = -\mathbb{E}_V \left[V\mathbb{E}_{\hat{V}|V}[\hat{V}|V] - \lambda f(V)^2 \right] \qquad (77a)$$

$$= \frac{1}{\sigma_v} \int \Phi\left(\frac{v}{\sigma_v}\right) \left(vQ\left(\frac{f(v)}{\sigma_w}\right) (\hat{v}_{(1)} - \hat{v}_{(2)}) + \lambda f(v)^2 \right) dv \qquad (77b)$$

$$= \frac{1}{\sigma_v} \int \Phi\left(\frac{v}{\sigma_v}\right) \left(vQ\left(\frac{f(v)}{\sigma_w}\right) \left(\frac{r_1}{r_2(r_2 - 1)}\right) + \lambda f(v)^2 \right) dv, \qquad (77c)$$

where $\hat{v}_{(1)} = \frac{r_1}{r_2 - 1}$ and $\hat{v}_{(2)} = \frac{r_1}{r_2}$. We observe that (77c) can be stated in the form in (64) by setting

$$\tilde{F} = vQ\left(\frac{f(v)}{\sigma_w}\right) \cdot \left(\frac{r_1}{r_2(r_2 - 1)}\right) + \lambda f(v)^2, \qquad (78a)$$

$$\tilde{G}_1 = vQ\left(\frac{f(v)}{\sigma_w}\right), \qquad (78b)$$

$$\tilde{G}_2 = Q\left(\frac{f(v)}{\sigma_w}\right). \qquad (78c)$$

Note that $\nabla L = \nabla L_s$. Therefore, from (68) we have the necessary condition

$$\nabla L = \frac{vr_1 e^{-\frac{f(v)^2}{2\sigma_w^2}}}{\sqrt{2\pi}\sigma_w r_2(1 - r_2)} + 2\lambda f(v)$$

$$+\frac{vr_1e^{-\frac{f(v)^2}{2\sigma_w^2}}}{\sqrt{2\pi}\sigma_w r_2(1-r_2)}-\frac{(1-2r_2)r_1^2}{(r_2(1-r_2))^2}\frac{e^{-\frac{f(v)^2}{2\sigma_w^2}}}{\sqrt{2\pi}\sigma_w}$$

$$=\frac{2r_1e^{-\frac{f(v)^2}{2\sigma_w^2}}}{\sqrt{2\pi}\sigma_w r_2(1-r_2)}\left(v-\frac{(1-2r_2)r_1}{2r_2(1-r_2)}\right)+2\lambda f(v)=0. \tag{79}$$

By solving (79) with respect to $f(v)$, we obtain (74), which concludes the proof.

$\square$

Based on (74), we restrict the minimization of the objective function over encoder mappings $f_{a,b}(v)$ that satisfy

$$f_{a,b}(v)e^{\frac{f_{a,b}(v)^2}{2\sigma_w^2}}=\frac{1}{a}(v-b), \tag{80}$$

for some parameters a and b. Note that taking the derivative of (80) with respect to b we have

$$f'_{a,b}(v)=\frac{1}{ae^{\frac{f_{a,b}(v)^2}{2\sigma_w^2}}\left(1+2f_{a,b}(v)^2\right)}, \tag{81}$$

which implies that for $a>0$, the function $f_{a,b}(v)$ is monotonically increasing.

Defining the function $f_o(\cdot)$ as the unique solution of

$$f_o(v)e^{\frac{f_o(v)^2}{2}}=v, \tag{82}$$

it will be convenient to write

$$f_{a,b}(v)=\sigma_w f_o\left(\frac{v-b}{a\sigma_w}\right). \tag{83}$$

This shows that any function $f_{a,b}(v)$ can be seen as a scaled and shifted version of the function that satisfies (82). The function $f_o(v)$ is plotted in Figure B1, along with an example of the function $f_{a,b}(v)$ for $a=4$, $b=10$. It can be easily verified that $f_o(v)$ is odd, that is, $f_o(-v)=-f_o(v)$. Furthermore, functions of the form $f_{a,0}(v)$, that is, with $b=0$, are odd as well.

In the following lemma, we show that the Lagrangian functional in (73) takes its minimum value only over functions of the form $f_{a,b}(v)$ where $b=0$ for any a, which shows that the optimal function $f_{a,b}(v)$ is odd as summarized in Corollary B.3.

Lemma B.2. *The optimal solution of the problem*

$$\underset{b}{minimize}\quad L(f_{a,b},g,\lambda), \tag{84}$$

is achieved for $b=0$ for any fixed a.

Proof: We prove the lemma assuming that $a,b\geq0$. This is without loss of generality given the relationship $L(f_{a,b},g,\lambda)=L(f_{a,-b},g,\lambda)=L(f_{-a,b},g,\lambda)=L(f_{-a,-b},g,\lambda)$. This can be verified by substituting $f_{a,-b}$, $f_{-a,b}$ and $f_{-a,-b}$ in (73) and performing the changes of variables $v=t-2b$, $v=2b-t$ and $v=-t$, respectively. In the following, we show that decreasing the value of $b\geq0$, reduces the average power of the mapping, i.e., $\mathbb{E}[f_{a,b}(V)^2]$, and increases $\mathbb{E}[V\hat{V}]$. Therefore, the function $L(f_{a,b},g,\lambda)$ becomes smaller by decreasing the value of $b\geq0$.

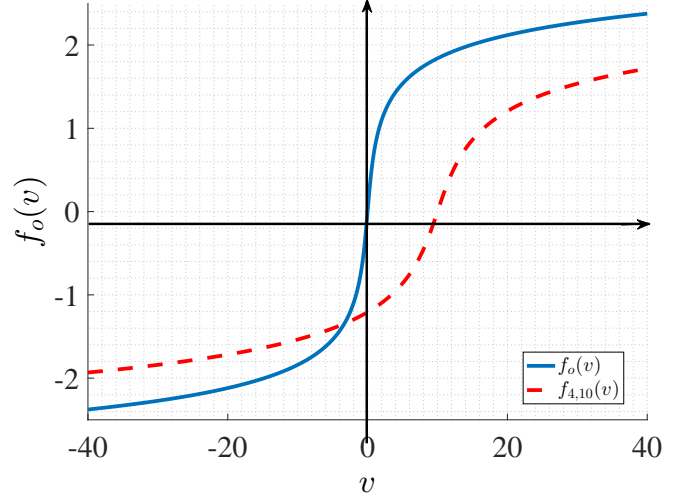

Figure B1: The function $f_o(v)$ in (82) and the function $f_{a,b}(v)$ in (80) with $a=4$, $b=10$.

1) $\mathbb{E}[f_{a,b}(V)^2]$ is a strictly increasing function of b: By writing the average power of the function $f_{a,b}(v)$ we have

$$\mathbb{E}[f_{a,b}(V)^2]=\frac{1}{\sigma_v}\int\Phi\left(\frac{v}{\sigma_v}\right)f_{a,b}(v)^2dv \tag{85}$$

$$=\frac{\sigma_w}{\sigma_v}\int\Phi\left(\frac{v}{\sigma_v}\right)f_o\left(\frac{v-b}{a\sigma_w}\right)^2dv \tag{86}$$

$$=\frac{a\sigma_w^2}{\sigma_v}\int\Phi\left(\frac{a\sigma_w v+b}{\sigma_v}\right)f_o\left(v\right)^2dv. \tag{87}$$

Differentiating $\mathbb{E}[f_{a,b}(V)^2]$ with respect to b, we have

$$-\frac{d\mathbb{E}[f_{a,b}(V)^2]}{db}$$

$$=\frac{a\sigma_w^2}{\sigma_v^3}\int(a\sigma_w v+b)\Phi\left(\frac{a\sigma_w v+b}{\sigma_v}\right)f_o\left(v\right)^2dv \tag{88}$$

$$\leq\frac{a\sigma_w^2}{\sigma_v^3}\int(a\sigma_w v+b)\Phi\left(\frac{a\sigma_w v+b}{\sigma_v}\right)\Psi(v)dv, \tag{89}$$

where $\Psi(v)$ is defined as

$$\Psi(v)\triangleq\begin{cases}f_o(v)^2 & v\leq-\frac{b}{a\sigma_w},\\ f_o\left(v+\frac{2b}{a\sigma_w}\right)^2 & v>-\frac{b}{a\sigma_w}.\end{cases} \tag{90}$$

It is noted that the inequality in (89) holds due to the fact that the integrand is positive and that we have the inequalities $\Psi(v)\geq f_o(v)$ for $v\geq-b/(a\sigma_w)$. Also, $\Psi(v)$ is an even function shifted to the left by $\frac{b}{a\sigma_w}$. Hence, due to the fact that $(a\sigma_w v+b)\phi(\frac{a\sigma_w v+b}{\sigma_v})$ is odd with respect to $v=-b/(a\sigma_w)$, the integrand on the right hand side of the inequality in (89) is an odd function shifted to the left by $b/(a\sigma_w)$. Therefore, the integral is zero, completing the proof. Note that (89) holds with equality only when $b=0$.

2) $\mathbb{E}[V\hat{V}]$ is a decreasing function of b: We first consider the noiseless scenario, i.e., $W = 0$. In the noiseless scenario, for the reconstruction points we have

$$\hat{V} = \begin{cases} \mathbb{E}[V|v \geq b] & Y = 0, \\ \mathbb{E}[V|v < b] & Y = 1 \end{cases} \tag{91}$$

$$= \begin{cases} \dfrac{\sigma_v e^{-\frac{b^2}{2}}}{\sqrt{2\pi}Q\left(\frac{b}{\sigma_v}\right)} & Y = 0, \\ \dfrac{-\sigma_v e^{-\frac{b^2}{2}}}{\sqrt{2\pi}\left(1-Q\left(\frac{b}{\sigma_v}\right)\right)} & Y = 1. \end{cases} \tag{92}$$

Therefore, $\mathbb{E}[V\hat{V}]$ can be written as

$$\mathbb{E}[V\hat{V}] = \Pr(V \geq b)\hat{v}_{(1)}\mathbb{E}[V|\hat{V} = \hat{v}_{(1)}]$$
$$+ \Pr(V < b)\hat{v}_{(2)}\mathbb{E}[V|\hat{V} = \hat{v}_{(2)}] \tag{93}$$

$$= \frac{\sigma_v^2 e^{-b^2}}{2\pi Q\left(\frac{b}{\sigma_v}\right)\left(1 - Q\left(\frac{b}{\sigma_v}\right)\right)} \triangleq \tilde{g}(b). \tag{94}$$

It can be verified that the function $\tilde{g}(b)$ is even, and that it takes its maximum at $b = 0$ as seen in Figure B2 (for the proof see Appendix I).

Now, considering the noisy received signal, we expand $\mathbb{E}[V\hat{V}]$ as

$$\mathbb{E}[V\hat{V}] = \frac{1}{\sigma_w}\int \Phi\left(\frac{w}{\sigma_w}\right)\mathbb{E}[V\hat{V}|W = -w]dw, \tag{95}$$

where

$$\mathbb{E}[V\hat{V}|W = -w]$$
$$= \mathbb{E}[V|Y = 0, W = -w]^2 \Pr(Y = 0|W = -w)$$
$$+ \mathbb{E}[V|Y = 1, W = -w]^2 \Pr(Y = 1|W = -w) \tag{96}$$

$$= \frac{\sigma_v e^{-\left(f_{a,b}^{-1}(w)\right)^2}}{2\pi Q\left(\frac{f_{a,b}^{-1}(w)}{\sigma_v}\right)} + \frac{\sigma_v e^{-\left(f_{a,b}^{-1}(w)\right)^2}}{2\pi\left(1 - Q\left(\frac{f_{a,b}^{-1}(w)}{\sigma_v}\right)\right)} \tag{97}$$

$$= \tilde{g}\left(f_{a,b}^{-1}(w)\right). \tag{98}$$

Notice that the inverse of the function $f_{a,b}(v)$ exists because it is one-to-one and is defined on the whole real line. Therefore,

$$\mathbb{E}[V\hat{V}] = \frac{1}{\sigma_w}\int \Phi\left(\frac{w}{\sigma_w}\right)\tilde{g}\left(f_{a,b}^{-1}(w)\right)dw. \tag{99}$$

We want to show that

$$\int e^{-\frac{w^2}{2}}\tilde{g}\left(f_{a,b}^{-1}(w)\right)dw \leq \int e^{-\frac{w^2}{2}}\tilde{g}\left(f_{a,0}^{-1}(w)\right)dw. \tag{100}$$

Using the change of variables $w = f_{a,0}(v)$ and the equality $f_{a,b}^{-1}(w) = b + f_{a,0}^{-1}(w)$, inequality (100) can be written as

$$\int e^{-\frac{f_{a,0}^2(v)}{2}}\tilde{g}(b+v)f_{a,0}'(v)dv \leq \int e^{-\frac{f_{a,0}^2(v)}{2}}\tilde{g}(v)f_{a,0}'(v)dv. \tag{101}$$

Inequality (101) is equivalent to

$$\int_{-\infty}^{-\frac{b}{2}} e^{-\frac{f_{a,0}^2(v)}{2}}\left(\tilde{g}(b+v) - \tilde{g}(v)\right)f_{a,0}'(v)dv \leq$$
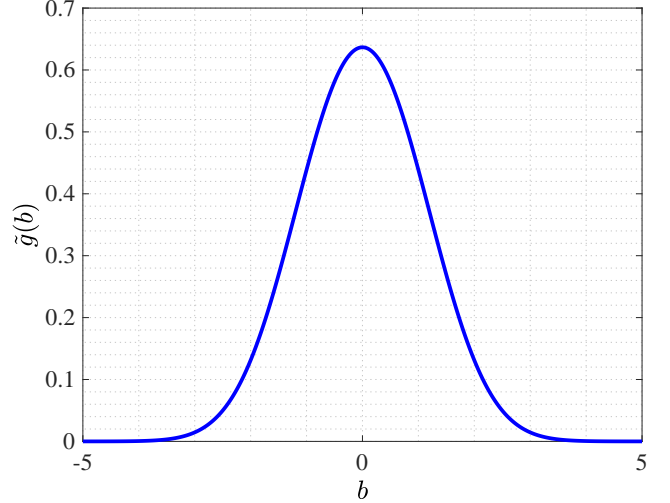

Figure B2: Plot of the function $\tilde{g}(b)$ in (94).

$$\int_{-\frac{b}{2}}^{\infty} e^{-\frac{f_{a,0}^2(v)}{2}}\left(\tilde{g}(v) - \tilde{g}(b+v)\right)f_{a,0}'(v)dv. \tag{102}$$

Using the transformation $-b - v = t$ for the left hand side of (102), we have

$$\int_{-\frac{b}{2}}^{\infty} e^{-\frac{f_{a,0}^2(-b-t)}{2}}\left(\tilde{g}(-t) - \tilde{g}(-b-t)\right)f_{a,0}'(-t-b)dt$$

$$= \int_{-\frac{b}{2}}^{\infty} e^{-\frac{f_{a,0}^2(b+t)}{2}}\left(\tilde{g}(t) - \tilde{g}(b+t)\right)f_{a,0}'(t+b)dt, \tag{103}$$

where the second equality is due to the fact that functions $\tilde{g}$, $f_{a,0}^2$, $f_{a,0}'$ are even. Therefore, (102) is equivalent to

$$\int_{-\frac{b}{2}}^{\infty}\left[e^{-\frac{f_{a,0}^2(b+t)}{2}}f_{a,0}'(t+b) - e^{-\frac{f_{a,0}^2(t)}{2}}f_{a,0}'(t)\right]$$

$$\cdot\left[\tilde{g}(t) - \tilde{g}(b+t)\right]dt \leq 0. \tag{104}$$

The inequality in (104) concludes the proof and follows from the following facts:

- Since $g(t)$ is even and decreasing for $t > 0$, we have $\tilde{g}(t) \geq \tilde{g}(t+b) > 0$ for $t \geq -\frac{b}{2}$, and hence the second bracket is nonnegative;
- Since $f_{a,0}(\cdot)$ is nondecreasing, we have the inequality $e^{-\frac{f_{a,0}^2(t)}{2}} \geq e^{-\frac{f_{a,0}^2(b+t)}{2}}$;
- From (81) given that the function $f_{a,0}'(t)$ is even and decreasing for $t > -\frac{b}{2}$, we have the inequality $f_{a,0}'(t) \geq f_{a,0}'(t+b) > 0$ for $t > -\frac{b}{2}$.

Therefore, the inequality in (100) is concluded.

$\square$

Corollary B.3. *The optimal mapping for problem (9) under the MSE criterion is odd.*

Proof: By Lemma B.1, the optimal mapping can be written in the form $f_{a,b}(v)$ without loss of optimality. By Lemma B.2, we conclude that the optimal mapping is in the form $f_{a,0}(v)$; and hence, it is odd.

Lemma B.4. *A necessary condition for an optimal solution to problem (8) is*

$$\nabla L = \frac{-2v}{\sqrt{2\pi}\sigma_w}e^{-\frac{f(v)^2}{2\sigma_w^2}} + 2\lambda f(v) = 0. \qquad (105)$$

Proof: Since the optimal mapping is odd by Corollary B.3, it can be easily verified that $\hat{v}_{(1)} = -\hat{v}_{(2)}$. Furthermore, without loss of optimality we can assume $\hat{v}_{(1)} \geq 0$. Therefore, using (13c) problem (8) can be restated as

$$\underset{f}{\text{minimize}} \quad -\hat{v}_{(1)} + \lambda E[f(V)^2], \qquad (106)$$

where $\hat{v}_{(1)}$ is obtained as in (12d). Expanding (106), we can write

$$\underset{f}{\text{minimize}} \quad \frac{-1}{\sigma_v}\int_{-\infty}^{\infty}\Phi\left(\frac{v}{\sigma_v}\right)\left(2vQ\left(\frac{f(v)}{\sigma_w}\right)-\lambda f(v)^2\right)dv. \quad (107)$$

Since (107) is of the form (69) with

$$\tilde{F} = -2vQ\left(\frac{f(v)}{\sigma_w}\right) + \lambda f(v)^2, \qquad (108)$$

we have the necessary condition in (105) using (71). This concludes the proof. ■

C. PROOF OF PROPOSITION III.2

Expanding the objective function $L(f, g, \lambda)$ as in (77a) we have

$$L_s(f, g, \lambda)$$
$$= \frac{-1}{\sigma_w}\int\Phi\left(\frac{v}{\sigma_v}\right)\left(v\int g\left(Q\left(f(v)+w\right)\right)\Phi\left(\frac{w}{\sigma_w}\right)dw - \lambda f(v)^2\right)dv$$
(109a)

$$= \frac{-1}{\sigma_v}\int\Phi\left(\frac{v}{\sigma_v}\right)\left(v\sum_{j=1}^{K}\hat{v}_{(j)}\left(Q\left(\frac{z_{(j-1)}-f(v)}{\sigma_w}\right)\right.\right.$$
$$\left.\left. -Q\left(\frac{z_{(j)}-f(v)}{\sigma_w}\right)\right)-\lambda f(v)^2\right)dv, \qquad (109b)$$

where we recall that $\hat{v}_{(j)}$ is the MMSE estimation of the source when the received signal is $Y = y_{(j)}$, i.e., $Y = f(V) + W \in [z_{(j-1)}, z_{(j)})$. We have $\hat{v}_{(j)} = \frac{r_{1j}}{r_{2j}}$ with

$$r_{1j} = \frac{1}{\sigma_v}\int v\Phi\left(\frac{v}{\sigma_v}\right)\left(Q\left(\frac{z_{(j-1)}-f(v)}{\sigma_w}\right)-Q\left(\frac{z_{(j)}-f(v)}{\sigma_w}\right)\right)dv,$$
(110a)

$$r_{2j} = \frac{1}{\sigma_v}\int \Phi\left(\frac{v}{\sigma_v}\right)\left(Q\left(\frac{z_{(j-1)}-f(v)}{\sigma_w}\right)-Q\left(\frac{z_{(j)}-f(v)}{\sigma_w}\right)\right)dv,$$
(110b)

for $j = 1, ..., K$. Since (109b) is of the form (64) with

$$\tilde{F} = -v\sum_{j=1}^{K}\frac{r_{1j}}{r_{2j}}\left(Q\left(\frac{z_{(j-1)}-f(v)}{\sigma_w}\right)-Q\left(\frac{z_{(j)}-f(v)}{\sigma_w}\right)\right)$$
$$+\lambda f(v)^2, \qquad (111)$$

and

$$\tilde{G}_{1j} = v\left(Q\left(\frac{z_{(j-1)}-f(v)}{\sigma_w}\right)-Q\left(\frac{z_{(j)}-f(v)}{\sigma_w}\right)\right),$$
$$\tilde{G}_{2j} = \left(Q\left(\frac{z_{(j-1)}-f(v)}{\sigma_w}\right)-Q\left(\frac{z_{(j)}-f(v)}{\sigma_w}\right)\right), \qquad (112)$$

for $j = 1, ..., K$. By writing the necessary condition in (68), we have

$$\nabla L = -\frac{v}{\sqrt{2\pi}\sigma_w}\sum_{j=1}^{K}\frac{r_{1j}}{r_{2j}}\left(e^{-\frac{\left(z_{(j-1)}-f(v)\right)^2}{2\sigma_w^2}}-e^{-\frac{\left(z_{(j)}-f(v)\right)^2}{2\sigma_w^2}}\right)$$
$$-\frac{v}{\sqrt{2\pi}\sigma_w}\sum_{j=1}^{K}\frac{r_{1j}}{r_{2j}}\left(e^{-\frac{\left(z_{(j-1)}-f(v)\right)^2}{2\sigma_w^2}}-e^{-\frac{\left(z_{(j)}-f(v)\right)^2}{2\sigma_w^2}}\right)$$
$$+\frac{1}{\sqrt{2\pi}\sigma_w}\sum_{j=1}^{K}\frac{r_{1j}^2}{r_{2j}^2}\left(e^{-\frac{\left(z_{(j-1)}-f(v)\right)^2}{2\sigma_w^2}}-e^{-\frac{\left(z_{(j)}-f(v)\right)^2}{2\sigma_w^2}}\right)$$
$$+2\lambda f(v) = 0. \qquad (113)$$

Solving for f, we have

$$f(v) = \frac{1}{2\sqrt{2\pi}\sigma_w\lambda}\sum_{j=1}^{K}\left(\hat{v}_{(j)}\left(2v-\hat{v}_{(j)}\right)\right.$$
$$\left. \cdot\left(e^{-\frac{\left(z_{(j-1)}-f(v)\right)^2}{2\sigma_w^2}}-e^{-\frac{\left(z_{(j)}-f(v)\right)^2}{2\sigma_w^2}}\right)\right). \quad (114)$$

■

D. PROOF OF PROPOSITION IV.1

We start by expanding the Lagrangian function (9) for the DOP criterion as

$$L(f, g, \lambda) = \epsilon(D) + \lambda\mathbb{E}[f(V)^2] \qquad (115a)$$
$$= \Pr\left(V \in (I_1 \cup I_2)^C\right)$$
$$+ \Pr\left(V \in (I_1 \setminus I_2), \hat{V} = \hat{v}_{(2)}\right)$$
$$+ \Pr\left(V \in (I_2 \setminus I_1), \hat{V} = \hat{v}_{(1)}\right)$$
$$+ \frac{\lambda}{\sigma_v}\int\Phi\left(\frac{v}{\sigma_v}\right)f(v)^2dv, \qquad (115b)$$

where we have used the decomposition in (24). The probabilities in (115b) can be written as

$$\Pr\left(V \in (I_1 \cup I_2)^C\right) = \frac{1}{\sigma_v}\int_{v\in(I_1\cup I_2)^C}\Phi\left(\frac{v}{\sigma_v}\right)dv, \quad (116a)$$

$$\Pr\left(V \in (I_1 \setminus I_2), \hat{V} = \hat{v}_{(2)}\right)$$

$$= \frac{1}{\sigma_v} \int_{v \in I_1 \setminus I_2} \Phi\left(\frac{v}{\sigma_v}\right) Q\left(\frac{f(v)}{\sigma_w}\right) dv, \quad (116b)$$

$$\Pr\left(V \in (I_2 \setminus I_1), \hat{V} = \hat{v}_{(1)}\right)$$

$$= \frac{1}{\sigma_v} \int_{v \in I_2 \setminus I_1} \Phi\left(\frac{v}{\sigma_v}\right) Q\left(\frac{-f(v)}{\sigma_w}\right) dv. \quad (116c)$$

Since the intervals on which the integrals in (116) are taken do not overlap and span the real line, the Lagrangian in (115) can be written in the form of (69) with $\tilde{F}(v,f)$ defined as

$$\tilde{F}(v,f) = \begin{cases} 1 + \lambda f(v)^2 & v \in (I_1 \cup I_2)^C, \\ Q\left(\frac{f(v)}{\sigma_w}\right) + \lambda f(v)^2 & v \in (I_1 \setminus I_2), \\ Q\left(\frac{-f(v)}{\sigma_w}\right) + \lambda f(v)^2 & v \in (I_2 \setminus I_1), \\ \lambda f(v)^2 & v \in (I_1 \cap I_2). \end{cases}$$
(117)

Using the optimality condition in (71) and differentiating $\tilde{F}(v,f)$ in (117) with respect to f, we have the necessary condition $\tilde{F}^f = 0$ with

$$\tilde{F}^f(v,f) =$$
$$\begin{cases} 2\lambda f(v) & v \in (I_1 \cup I_2)^C \cup (I_1 \cap I_2), \\ \frac{-1}{\sqrt{2\pi}\sigma_w} e^{\frac{-f(v)^2}{2\sigma_w^2}} + 2\lambda f(v) & v \in (I_1 \setminus I_2), \\ \frac{1}{\sqrt{2\pi}\sigma_w} e^{\frac{-f(v)^2}{2\sigma_w^2}} + 2\lambda f(v) & v \in (I_2 \setminus I_1). \end{cases}$$
(118)

Therefore, we have

$$f(v) = \begin{cases} 0 & v \in (I_1 \cup I_2)^C \cup (I_1 \cap I_2), \\ \frac{1}{2\lambda\sqrt{2\pi}\sigma_w} e^{\frac{-f(v)^2}{2\sigma_w^2}} & v \in (I_1 \setminus I_2), \\ \frac{-1}{2\lambda\sqrt{2\pi}\sigma_w} e^{\frac{-f(v)^2}{2\sigma_w^2}} & v \in (I_2 \setminus I_1). \end{cases}$$
(119)

∎

E. Proof of Proposition IV.2

Given λ, and obtaining the value of u from (26), we aim to minimize the objective function in (9) for the DOP with respect to the reconstruction points $\hat{v}_{(1)}$ and $\hat{v}_{(2)}$. Since by Proposition IV.1 the non-zero optimal values of $f(v)$ are opposite of one another over $I_2 \setminus I_1$ and $I_1 \setminus I_2$, we can rewrite (9) for the DOP as follows

$$L(f,g,\lambda) = \frac{1}{\sigma_v}\left(\int_{(I_1 \cup I_2)^C} \Phi\left(\frac{v}{\sigma_v}\right) dv \right.$$

$$+ Q\left(\frac{u}{\sigma_w}\right)\left(\int_{I_2 \setminus I_1} \Phi\left(\frac{v}{\sigma_v}\right) dv + \int_{I_1 \setminus I_2} \Phi\left(\frac{v}{\sigma_v}\right) dv \right)$$

$$+ \lambda\left(\int_{I_2 \setminus I_1} \Phi\left(\frac{v}{\sigma_v}\right) u^2 dv + \int_{I_1 \setminus I_2} \Phi\left(\frac{v}{\sigma_v}\right) u^2 dv \right)\right) \quad (120)$$

$$= \frac{1}{\sigma_v}\left(\int_{(I_1 \cup I_2)^C} \Phi\left(\frac{v}{\sigma_v}\right) dv \right.$$

$$\left. + \left(Q\left(\frac{u}{\sigma_w}\right) + \lambda u^2\right) \cdot \int_{(I_2 \setminus I_1) \cup (I_1 \setminus I_2)} \Phi\left(\frac{v}{\sigma_v}\right) dv \right), \quad (121)$$

where u is the solution of $ue^{\frac{u^2}{2\sigma_w^2}} = \frac{1}{2\sigma_w\sqrt{2\pi}\lambda}$. From (121), it can be shown that (9) for the DOP is minimized if I_1 and I_2 lie around the origin, i.e., $0 \in I_1 \cap I_2$, and they are symmetric with respect to the origin; that is, if $I_1 = [l,h]$ for $l \leq 0 \leq h$, then $I_2 = [-h, -l]$. From (26), we have $\lambda u^2 = \frac{ue^{\frac{-u^2}{2\sigma_w^2}}}{2\sigma_w\sqrt{2\pi}}$. By substituting this into (121) we have

$$L(f,g,\lambda) = \frac{1}{\sigma_v}\left(\int_{(I_1 \cup I_2)^C} \Phi\left(\frac{v}{\sigma_v}\right) dv \right.$$

$$\left. + C(u) \cdot \int_{(I_2 \setminus I_1) \cup (I_1 \setminus I_2)} \Phi\left(\frac{v}{\sigma_v}\right) dv \right), \quad (122)$$

where $C(u) \triangleq Q\left(\frac{u}{\sigma_w}\right) + \frac{ue^{\frac{-u^2}{2\sigma_w^2}}}{2\sigma_w\sqrt{2\pi}}$. By differentiating $C(u)$ with respect to u it can be verified that $C(u)$ is decreasing with u and is less than or equal to $1/2$ for all $u \geq 0$. We now prove the claim (the intervals are around the origin and symmetric) for the case in which the intervals I_1 and I_2 do not overlap. The proof is similar for the case in which the intervals overlap.

Consider two symmetric non-overlapping intervals I_1^s and I_2^s, such that $I_1^s = [-2\sqrt{D}, 0]$ and $I_2^s = [0, 2\sqrt{D}]$. Note that we have

$$\frac{1}{\sigma_v} \int_{I_1 \cup I_2} \Phi\left(\frac{v}{\sigma_v}\right) dv$$

$$\leq \frac{1}{\sigma_v} \int_{I_1^s \cup I_2^s} \Phi\left(\frac{v}{\sigma_v}\right) dv \quad (123)$$

$$= \frac{1}{\sigma_v} \int_{I_1 \cup I_2} \Phi\left(\frac{v}{\sigma_v}\right) dv + A_{\text{diff}}, \quad (124)$$

where $A_{\text{diff}} \geq 0$. Therefore, we can write

$$L(f,g,\lambda) = \frac{1}{\sigma_v}\left(\int_{(I_1 \cup I_2)^C} \Phi\left(\frac{v}{\sigma_v}\right) dv \right.$$

$$\left. + C(u) \cdot \int_{I_1 \cup I_2} \Phi\left(\frac{v}{\sigma_v}\right) dv \right) \quad (125)$$

$$= \frac{1}{\sigma_v}\left(\int_{(I_1^s \cup I_2^s)^C} \Phi\left(\frac{v}{\sigma_v}\right) dv + A_{\text{diff}} \right.$$

$$+C(u)\cdot\left(\int_{I_1^s\cup I_2^s}\Phi\left(\frac{v}{\sigma_v}\right)dv-A_{\text{diff}}\right)\right) \tag{126}$$

$$=\frac{1}{\sigma_v}\left(\int_{(I_1^s\cup I_2^s)^C}\Phi\left(\frac{v}{\sigma_v}\right)dv\right.$$

$$+C(u)\cdot\int_{I_1^s\cup I_2^s}\Phi\left(\frac{v}{\sigma_v}\right)dv\right)+A_{\text{diff}}(1-C(u)) \tag{127}$$

$$=L^s(f,g,\lambda)+A_{\text{diff}}(1-C(u)), \tag{128}$$

where $L^s(f,g,\lambda)$ represents the Lagrangian for symmetric non-overlapping intervals. Since $A_{\text{diff}}(1-C(u))\geq 0$, it follows that

$$L(f,g,\lambda)\geq L^s(f,g,\lambda),$$

proving the claim.

Focusing on symmetric intervals, we define the positive parameter $0\leq a\leq\sqrt{D}$ such that $I_2=[-2\sqrt{D}+a,a]$ and $I_1=[-a,2\sqrt{D}-a]$, and $I_1\cap I_2=[-a,a]$. Therefore, (121) can be rewritten as

$$L(f,g,\lambda)=2Q\left(\frac{2\sqrt{D}-a}{\sigma_v}\right)+2\left(Q\left(\frac{u}{\sigma_w}\right)+\lambda u^2\right)$$

$$\cdot\left(Q\left(\frac{a}{\sigma_v}\right)-Q\left(\frac{2\sqrt{D}-a}{\sigma_v}\right)\right). \tag{129}$$

Optimizing (129) over a completes the proof of Proposition IV.2.   ∎

F. Proof of Proposition IV.3

Necessary optimality condition of an encoder mapping f for a given set of reconstruction points $\{\hat{v}_{(1)},\ldots,\hat{v}_{(K)}\}$: We start by expanding $\epsilon(D)$ with respect to the different intervals defined in (31). We have

$$\epsilon(D)=1-\Pr\left((V-\hat{V})^2<D\right) \tag{130a}$$

$$=1-\sum_{\substack{\mathcal{V}:\\ \mathcal{V}\subset\{\hat{v}_{(1)},\ldots,\hat{v}_{(K)}\}}}\Pr\left((V-\hat{V})^2<D|V\in I_\mathcal{V}\right)\Pr\left(V\in I_\mathcal{V}\right) \tag{130b}$$

$$=1-\Pr\left((V-\hat{V})^2<D|V\in I_\emptyset\right)\Pr\left(V\in I_\emptyset\right)$$

$$-\Pr\left((V-\hat{V})^2<D|V\in I_{\{\hat{v}_{(1)},\ldots,\hat{v}_{(K)}\}}\right)$$

$$\cdot\Pr\left(V\in I_{\{\hat{v}_{(1)},\ldots,\hat{v}_{(K)}\}}\right)$$

$$-\sum_{\substack{\mathcal{V}:\mathcal{V}\subset\{\hat{v}_{(1)},\ldots,\hat{v}_{(K)}\}\\|\mathcal{V}|=1,\ldots,K-1}}\Pr\left((V-\hat{V})^2<D|V\in I_\mathcal{V}\right)\Pr\left(V\in I_\mathcal{V}\right) \tag{130c}$$

$$=1-\Pr\left(V\in I_{\{\hat{v}_{(1)},\ldots,\hat{v}_{(K)}\}}\right)$$

$$-\sum_{\substack{\mathcal{V}:\mathcal{V}\subset\{\hat{v}_{(1)},\ldots,\hat{v}_{(K)}\}\\|\mathcal{V}|=1,\ldots,K-1}}\Pr\left((V-\hat{V})^2<D|V\in I_\mathcal{V}\right)\Pr\left(V\in I_\mathcal{V}\right) \tag{130d}$$

$$=1-\Pr\left(V\in I_{\{\hat{v}_{(1)},\ldots,\hat{v}_{(K)}\}}\right)$$

$$-\sum_{\substack{\mathcal{V}:\mathcal{V}\subset\{\hat{v}_{(1)},\ldots,\hat{v}_{(K)}\}\\|\mathcal{V}|=1,\ldots,K-1}}\frac{1}{\sigma_v}\int_{v\in I_\mathcal{V}}\Phi\left(\frac{v}{\sigma_v}\right)\Pr\left(W+f(v)\in\zeta_\mathcal{V}\right)dv \tag{130e}$$

$$=1-\Pr\left(V\in I_{\{\hat{v}_{(1)},\ldots,\hat{v}_{(K)}\}}\right)$$

$$-\frac{1}{\sigma_v}\sum_{\substack{\mathcal{V}:\mathcal{V}\subset\{\hat{v}_{(1)},\ldots,\hat{v}_{(K)}\}\\|\mathcal{V}|=1,\ldots,K-1}}\int_{v\in I_\mathcal{V}}\Phi\left(\frac{v}{\sigma_v}\right) \tag{130f}$$

$$\cdot\sum_{j:\hat{v}_{(j)}\in\mathcal{V}}\left(Q\left(\frac{z_{(j-1)}-f(v)}{\sigma_w}\right)-Q\left(\frac{z_{(j)}-f(v)}{\sigma_w}\right)\right)dv, \tag{130g}$$

where $\zeta_\mathcal{V}\triangleq\bigcup_{j:\hat{v}_{(j)}\in\mathcal{V}}\left[z_{(j-1)},z_{(j)}\right)$. Note that the different sets in (130f) partition the real line, that is, they are disjoint and their union is the real line. Substituting (130f) in (14), the Lagrangian $L(f,g,\lambda)$ can be written in the form of (69) with $\tilde{F}(v,f)$ defined as

$$\tilde{F}(v,f)\triangleq$$
$$\begin{cases}1+\lambda f(v)^2 & v\in I_\emptyset,\\ \lambda f(v)^2 & v\in I_{\{\hat{v}_{(1)},\ldots,\hat{v}_{(K)}\}},\\ 1-\sum_{j:\hat{v}_{(j)}\in\mathcal{V}}\left(Q\left(\frac{z_{(j-1)}-f(v)}{\sigma_w}\right)\right. & v\in I_\mathcal{V},\ |\mathcal{V}|=1,\ldots,K-1.\\ \left.-Q\left(\frac{z_{(j)}-f(v)}{\sigma_w}\right)\right)+\lambda f(v)^2\end{cases} \tag{131}$$

Writing the necessary condition in (71) and setting to zero leads to (35).

Optimal decoder function g for a given encoder mapping f: For a given encoder mapping f, the optimal decoder can be obtained as

$$\hat{v}_{(j)}=\arg\min_t\ \Pr((V-t)^2\geq D|Y=y_{(j)}) \tag{132a}$$

$$\equiv\arg\max_t\ \Pr\left((V-t)^2<D,Y=y_{(j)}\right) \tag{132b}$$

$$=\arg\max_t\int_{t-\sqrt{D}}^{t+\sqrt{D}}\int_{z_{(j-1)}-f(v)}^{z_{(j)}-f(v)}\Phi\left(\frac{v}{\sigma_v}\right)\Phi\left(\frac{w}{\sigma_w}\right)dwdv \tag{132c}$$

$$=\arg\max_t\int_{t-\sqrt{D}}^{t+\sqrt{D}}\Phi\left(\frac{v}{\sigma_v}\right)\left(Q\left(\frac{z_{(j-1)}-f(v)}{\sigma_w}\right)\right.$$

$$\left.-Q\left(\frac{z_{(j)}-f(v)}{\sigma_w}\right)\right)dv. \tag{132d}$$

∎

G. Proof of Proposition V.1

Define $y_i \triangleq \Gamma(f(v) + w_i)$, $w^N \triangleq [w_1, ..., w_N]^T$, $y^N \triangleq [y_1, ..., y_N]^T$. Expanding $L(f, g, \lambda)$ in (9) for the MSE we have

$$L_s(f, g, \lambda)$$

$$= \frac{-1}{\sigma \sigma_v} \int_v \int_{w^N} v g\left(y^N\right) \Phi\left(\frac{v}{\sigma_v}\right) \prod_{i=1}^N \Phi\left(\frac{w_i}{\sigma_{w_i}}\right) dw^N dv$$

$$+ \lambda \frac{1}{\sigma_v} \int \Phi\left(\frac{v}{\sigma_v}\right) f(v)^2 dv \tag{133a}$$

$$= \frac{-1}{\sigma \sigma_v} \int \Phi\left(\frac{v}{\sigma_v}\right) \left(v \int_{w^N} g\left(y^N\right) \prod_{i=1}^N \Phi\left(\frac{w_i}{\sigma_{w_i}}\right) dw^N - \lambda f(v)^2 \right) dv \tag{133b}$$

$$= \frac{-1}{\sigma \sigma_v} \int \Phi\left(\frac{v}{\sigma_v}\right) \left(v \sum_{j=1}^{2^N} \hat{v}_{(j)} \Pr\left(\hat{V} = \hat{v}_{(j)} | V = v\right) - \lambda f(v)^2 \right) dv, \tag{133c}$$

where $\hat{v}_{(j)} = \mathbb{E}[V | Y^N = b_j^N] = \mathbb{E}[V | Y_1 = b_j^N(1), \ldots, Y_N = b_j^N(N)]$ and $\sigma = \prod_{i=1}^N \frac{1}{\sigma_{w_i}}$. We can compute

$$\Pr\left(\hat{V} = \hat{v}_{(j)} | V = v\right) = \prod_{i=1}^N \Pr\left(\Gamma(f(v) + W_i) = b_j^N(i)\right) \tag{134a}$$

$$= \prod_{i=1}^N Q\left(\frac{(-1)^{b_j^N(i)+1} f(v)}{\sigma_{w_i}}\right), \tag{134b}$$

for $j = 1, ..., 2^N$. Substituting (134b) in (133c), we have

$$L_s(f, g, \lambda) = \frac{-1}{\sigma_v} \int \Phi\left(\frac{v}{\sigma_v}\right) \left(v \sum_{j=1}^{2^N} \hat{v}_{(j)} \right.$$

$$\left. \cdot \prod_{i=1}^N Q\left(\frac{(-1)^{b_j^N(i)+1} f(v)}{\sigma_{w_i}}\right) - \lambda f(v)^2 \right) dv, \tag{135}$$

where for $\hat{v}_{(j)}$ we have

$$\hat{v}_{(j)} = \frac{\frac{1}{\sigma_v} \int v \Phi\left(\frac{v}{\sigma_v}\right) \prod_{i=1}^N Q\left(\frac{(-1)^{b_j^N(i)+1} f(v)}{\sigma_{w_i}}\right) dv}{\frac{1}{\sigma_v} \int \Phi\left(\frac{v}{\sigma_v}\right) \prod_{i=1}^N Q\left(\frac{(-1)^{b_j^N(i)+1} f(v)}{\sigma_{w_i}}\right) dv} \triangleq \frac{r_{1_j}}{r_{2_j}}, \tag{136}$$

for $j = 1, \ldots, 2^N$. Since (135) is of the form (64) with

$$\tilde{F} = -v \sum_{j=1}^{2^N} \hat{v}_{(j)} \prod_{i=1}^N Q\left(\frac{(-1)^{b_j^N(i)+1} f(v)}{\sigma_{w_i}}\right) + \lambda f(v)^2, \tag{137}$$

$$\tilde{G}_{1_j} = v \prod_{i=1}^N Q\left(\frac{(-1)^{b_j^N(i)+1} f(v)}{\sigma_{w_i}}\right), \; j = 1, \ldots, 2^N, \tag{138}$$

$$\tilde{G}_{2_j} = \prod_{i=1}^N Q\left(\frac{(-1)^{b_j^N(i)+1} f(v)}{\sigma_{w_i}}\right), \; j = 1, \ldots, 2^N. \tag{139}$$

Writing down the optimality condition in (68), we have

$$\nabla L = - v \sum_{j=1}^{2^N} \frac{r_{1_j}}{r_{2_j}} \Theta\left(N, f(v), j\right) + 2\lambda f(v)$$

$$- \sum_{j=1}^{2^N} \frac{v \cdot \Theta\left(N, f(v), j\right)}{r_{2_j}} \cdot r_{1_j}$$

$$+ \sum_{j=1}^{2^N} \frac{r_{1_j} \cdot \Theta\left(N, f(v), j\right)}{r_{2_j}^2} \cdot r_{1_j} \tag{140a}$$

$$= - \sum_{j=1}^{2^N} \frac{r_{1_j}}{r_{2_j}} \Theta\left(N, f(v), j\right) \left(2v - \frac{r_{1_j}}{r_{2_j}} \right) + 2\lambda f(v) \tag{140b}$$

$$= - \sum_{j=1}^{2^N} \Theta\left(N, f(v), j\right) \hat{v}_{(j)} \left(2v - \hat{v}_{(j)} \right) + 2\lambda f(v) = 0, \tag{140c}$$

where

$$\Theta\left(N, f(v), j\right) = \sum_{k=1}^N \left(\frac{(-1)^{b_j^N(k)} e^{-\frac{f(v)^2}{2\sigma_{w_k}^2}}}{\sqrt{2\pi} \sigma_{w_k}} \right.$$

$$\left. \prod_{l=1, l \neq k}^N Q\left(\frac{(-1)^{b_j^N(l)+1} f(v)}{\sigma_{w_l}}\right) \right). \tag{141}$$

Therefore, the optimal mapping must be in the form given by (42). ∎

H. Proof of Proposition V.2

Necessary optimality condition of an encoder mapping f for a given set of reconstruction points $\{\hat{v}_{(1)}, \ldots, \hat{v}_{(K)}\}$: We expand $\epsilon(D)$

$$\epsilon(D) = 1 - \Pr\left((V - \hat{V})^2 < D\right) \tag{142a}$$

$$= 1 - \frac{1}{\sigma \sigma_v} \iint_{v, w^N : |v - g(y^N)|^2 < D} \Phi\left(\frac{v}{\sigma_v}\right) \prod_{i=1}^N \Phi\left(\frac{w_i}{\sigma_{w_i}}\right) dw^N dv \tag{142b}$$

$$= 1 - \frac{1}{\sigma \sigma_v} \sum_{\mathcal{V}:\mathcal{V} \subset \left\{\hat{v}_{(1)}, \ldots, \hat{v}_{(2^N)}\right\}} \int_{v \in I_\mathcal{V}} \int_{w^N : g(y^N) \in \mathcal{V}} \Phi\left(\frac{v}{\sigma_v}\right)$$

$$\cdot \prod_{i=1}^N \Phi\left(\frac{w_i}{\sigma_{w_i}}\right) dw^N dv \tag{142c}$$

$$= 1 - \frac{1}{\sigma_v} \sum_{\mathcal{V}:\mathcal{V} \subset \left\{\hat{v}_{(1)}, \ldots, \hat{v}_{(2^N)}\right\}} \int_{v \in I_\mathcal{V}} \Phi\left(\frac{v}{\sigma_v}\right)$$

$$\cdot \sum_{j : \hat{v}_{(j)} \in \mathcal{V}} \left(\prod_{i=1}^N Q\left(\frac{(-1)^{b_j^N(i)+1} f(v)}{\sigma_{w_i}}\right) \right) dv \tag{142d}$$

$$= 1 - \frac{1}{\sigma_v} \int\limits_{v \in I_{\left\{ \hat{v}_{(1)}, \dots, \hat{v}_{(2^N)} \right\}}} \Phi\left(\frac{v}{\sigma_v}\right) dv$$

$$- \frac{1}{\sigma_v} \sum_{\substack{\mathcal{V}: \mathcal{V} \subset \left\{ \hat{v}_{(1)}, \dots, \hat{v}_{(2^N)} \right\} \\ |\mathcal{V}| = 1, \dots, 2^N - 1}} \int\limits_{v \in I_{\mathcal{V}}} \Phi\left(\frac{v}{\sigma_v}\right)$$

$$\cdot \sum_{j: \hat{v}_{(j)} \in \mathcal{V}} \left(\prod_{i=1}^{N} Q\left(\frac{(-1)^{b_j^N(i)+1} f(v)}{\sigma_{w_i}} \right) \right) dv, \quad (142e)$$

where $\sigma = \prod_{i=1}^{N} \sigma_{w_i}$. Substituting (142e) in (9), the Lagrangian $L(f, g, \lambda)$ can be written in the form of (69) with $\tilde{F}(v, f)$ defined as

$$\tilde{F}(v, f) \triangleq \quad (143)$$

$$\begin{cases} 1 + \lambda f(v)^2 & v \in I_\emptyset, \\ \lambda f(v)^2 & v \in I_{\left\{ \hat{v}_{(1)}, \dots, \hat{v}_{(2^N)} \right\}}, \\ 1 + \lambda f(v)^2 & v \in I_{\mathcal{V}} : |\mathcal{V}| = 1, \dots, 2^N - 1. \\ - \sum_{j: \hat{v}_{(j)} \in \mathcal{V}} \prod_{i=1}^{N} Q\left(\frac{(-1)^{b_j^N(i)+1} f(v)}{\sigma_{w_i}} \right) \end{cases}$$

$$(144)$$

Writing the necessary condition in (71) and setting to zero yields the result in (45).

Optimal decoder function g for a given encoder mapping f: Assume that the encoder mapping is given. Following a similar approach to the derivation of the optimal decoder for a K-level ADC front end, the optimal decoder at the receiver is obtained as

$$\hat{v}_{(j)} = \arg\max_{t} \int\limits_{t-\sqrt{D}}^{t+\sqrt{D}} \Phi\left(\frac{v}{\sigma_v}\right) \Pr(Y_1 = b_j^N(1), \dots, Y_N = b_j^N(N)) dv$$

$$(145a)$$

$$= \arg\max_{t} \int\limits_{t-\sqrt{D}}^{t+\sqrt{D}} \Phi\left(\frac{v}{\sigma_v}\right) \prod_{i=1}^{N} Q\left(\frac{(-1)^{b_j^N(i)+1} f(v)}{\sigma_{w_i}} \right) dv.$$

$$(145b)$$

∎

I. Proof of the Decreasing Characteristic of (94) with Respect to $|b|$

It can be easily verified from (94) that $\tilde{g}(b)$ is even. In the following, using the bounding techniques for the Q function, we show that $\tilde{g}(b)$ is decreasing with respect to the absolute value of b. We assume $\sigma_v^2 = 1$ for brevity of the presentation, but the results are valid for any value of σ_v^2. We consider the cases $b \geq 1$ and $0 \leq b \leq 1$ separately.

- $b \geq 1$: Taking the derivative of (94) with respect to b, we have

$$\tilde{g}'(b) = \frac{e^{-b^2} g_1(b)}{2\pi \left(Q(b) \left(1 - Q(b)\right) \right)^2}. \quad (146)$$

where $g_1(b) \triangleq \Phi(b)(1 - 2Q(b)) - 2bQ(b)(1 - Q(b))$. It can be easily verified that $\tilde{g}'(0) = 0$. It is enough to show that $g_1(b) < 0$ for $b > 1$. Using the lower bound $Q(b) \geq \frac{b}{1+b^2}\Phi(b)$, we have

$$g_1(b) \leq \Phi(b)(1 - 2Q(b)) - \frac{2b^2}{1 + b^2}\Phi(b)(1 - Q(b)) \quad (147)$$

$$= \frac{\Phi(b)}{1 + b^2} \left(1 - b^2 - 2Q(b)\right), \quad (148)$$

which is obviously negative for $b > 1$.

- $0 \leq b \leq 1$: Taking the second derivative of $g_1(b)$, we have

$$\tilde{g}_1''(b) = -\Phi(b) \left((b^2 - 3)(1 - 2Q(b)) + 2b\Phi(b)\right). \quad (149)$$

Using the first two terms of the Taylor series expansion of the Q-function as a lower bound, that is, $Q(b) \geq 1/2 - \Phi(0)b$, we can write

$$\tilde{g}_1''(b) \geq -\Phi(b) \left((b^2 - 3)2\Phi(0)b + 2b\Phi(b)\right) \quad (150)$$

$$\geq -2b\Phi(0)\Phi(b)(b^2 - 2) \geq 0. \quad (151)$$

Therefore, $g_1(b)$ is convex for $0 \leq b \leq 1$. Noticing that $g_1(0) = 0$, and $g_1(1) < 0$, concludes that $g_1(b) < 0$ for all $b > 0$.

∎

References

[1] M. Varasteh, O. Simeone, and D. Gündüz, "Joint source-channel coding with One-Bit ADC front end," *IEEE Int. Sym. Inf. Theory*, Jul. 2016.

[2] R. Walden, "Analog-to-digital converter survey and analysis," *IEEE Journal on Selected Areas in Comm.*, vol. 17, no. 4, pp. 539–550, Apr. 1999.

[3] B. Murmann, "ADC performance survey," *CoRR*, vol. abs/1404.7736, 1997-2014. [Online]. Available: http://web.stanford.edu/~murmann/adcsurvey.html

[4] S. H. Park, O. Simeone, O. Sahin, and S. S. Shitz, "Fronthaul compression for cloud radio access networks: Signal processing advances inspired by network information theory," *IEEE Signal Processing Magazine*, vol. 31, no. 6, pp. 69–79, Nov 2014.

[5] I. D. O'Donnell and R. W. Brodersen, "An ultra-wideband transceiver architecture for low power, low rate, wireless systems," *IEEE Trans. Vehicular Technology*, vol. 54, no. 5, pp. 1623–1631, Sep. 2005.

[6] J. Singh, O. Dabeer, and U. Madhow, "On the limits of communication with low-precision analog-to-digital conversion at the receiver," *IEEE Trans. Commun.*, vol. 57, no. 12, pp. 3629–3639, Dec. 2009.

[7] T. Koch and A. Lapidoth, "At low snr, asymmetric quantizers are better," *IEEE Trans. Inf. Theory*, vol. 59, no. 9, pp. 5421–5445, Sep. 2013.

[8] S. Verdu, "Spectral efficiency in the wideband regime," *IEEE Trans. Inf. Theory*, vol. 48, no. 6, pp. 1319–1343, Jun. 2002.

[9] S. Krone and G. Fettweis, "Fading channels with 1-bit output quantization: Optimal modulation, ergodic capacity and outage probability," in *Information Theory Workshop (ITW), 2010 IEEE*, Aug 2010, pp. 1–5.

[10] A. Mezghani and J. A. Nossek, "Analysis of rayleigh-fading channels with 1-bit quantized output," in *IEEE Int. Sym. Inf. Theory*, Toronto, Jul. 2008, pp. 260–264.

[11] A. Mezghani and J. Nossek, "On ultra-wideband MIMO systems with 1-bit quantized outputs: Performance analysis and input optimization," *IEEE Int. Sym. Inf. Theory*, pp. 1286–1289, Jun. 2007.

[12] J. Mo and R. Heath, "Capacity analysis of one-bit quantized MIMO systems with transmitter channel state information," *IEEE Trans. on Signal Processing*, vol. 63, no. 20, pp. 5498–5512, Oct. 2015.

[13] H. Cui, C. Luo, C. W. Chen, and F. Wu, "Robust uncoded video transmission over wireless fast fading channel," in *IEEE Conference on Computer Commun.*, Apr. 2014.

[14] D. Liu and J. Wu, "Uncoded video transmission in multi-users networks," in *Proc. Int'l Conf. Commun. and Networking*, Aug. 2015.

[15] F. Massé, M. V. Bussel, A. Serteyn, J. Arends, and J. Penders, "Miniaturized wireless ecg monitor for real-time detection of epileptic seizures," *ACM Trans. Embed. Comput. Syst.*, vol. 12, no. 4, pp. 102:1–102:21, 2013.

[16] M. Condoluci, G. Araniti, T. Mahmoodi, and M. Dohler, "Enabling the iot machine age with 5g: Machine-type multicast services for innovative real-time applications," *IEEE Access*, vol. 4, pp. 5555–5569, 2016.

[17] W. K. Hsieh, W. H. Hsieh, J. L. Chen, and C. Y. Lin, "Self-configuration and smart binding control on iot applications," in *Int'l Conf. Advanced Commun. Technology (ICACT)*, Jan. 2016.

[18] T. J. Goblick, "Theoretical limitations on the transmission of data from analog sources," *IEEE Trans. Inf. Theory*, vol. 11, no. 11, pp. 558–567, Nov. 1965.

[19] Y. Polyanskiy, H. V. Poor, and S. Verdu, "Channel coding rate in the finite blocklength regime," *IEEE Trans. Inf. Theory*, vol. 56, no. 5, pp. 2307–2359, May 2010.

[20] V. Kostina and S. Verdu, "Fixed-length lossy compression in the finite blocklength regime," *IEEE Trans. Inf. Theory*, vol. 58, no. 6, pp. 3309–3338, Jun. 2012.

[21] M. Gastpar, B. Rimoldi, and M. Vetterli, "To code, or not to code: Lossy source-channel communication revisited," *IEEE Trans. Inf. Theory*, vol. 49, no. 5, pp. 1147–1158, May 2003.

[22] C. Shannon, "Communication in the presence of noise," *Proceedings of the IEEE*, vol. 86, no. 2, pp. 447–457, Feb 1998.

[23] V. A. Kotel'nikov, *The Theory of Optimum Noise Immunity*. New York: McGraw-Hill Book Company, Inc., 1959.

[24] A. Fuldseth and T. A. Ramstad, "Bandwidth compression for continuous amplitude channels based on vector approximation to a continuous subset of the source signal space," in *IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, vol. 4, Apr 1997, pp. 3093–3096 vol.4.

[25] S. Y. Chung, "On the construction of some capacity approaching coding schemes," Ph.D. dissertation, Massachusetts Institute of Technology, 2000.

[26] V. A. Vaishampayan and S. I. R. Costa, "Curves on a sphere, shift-map dynamics, and error control for continuous alphabet sources," *IEEE Trans. Inf. Theory*, vol. 49, no. 7, pp. 1658–1672, Jul. 2003.

[27] T. A. Ramstad, "Shannon mappings for robust communication," *Telektronikk*, vol. 98, no. 1, pp. 114–128, 2002.

[28] F. Hekland, P. Floor, and T. Ramstad, "Shannon-Kotel'nikov mappings in joint source-channel coding," *IEEE Trans. Commun.*, vol. 57, no. 1, pp. 94–105, Jan. 2009.

[29] P. A. Floor, T. A. Ramstad, and N. Wernersson, "Power constrained channel optimized vector quantizers used for bandwidth expansion," in *Int. Sym. Wireless Commun. Systems*, Oct. 2007, pp. 667–671.

[30] J. Karlsson and M. Skoglund, "Optimized low-delay source-channel-relay mappings," *IEEE Trans. Commun.*, vol. 58, no. 5, pp. 1397–1404, May 2010.

[31] E. Akyol, K. B. Viswanatha, K. Rose, and T. A. Ramstad, "On zero-delay source-channel coding," *IEEE Trans. Inf. Theory*, vol. 60, no. 12, pp. 7473–7489, Dec. 2014.

[32] M. Varasteh, B. Rassouli, O. Simeone, and D. Gündüz, "Zero-delay joint source-channel coding with a 1-bit ADC front end and receiver side information," in *IEEE Inf. Theory Workshop (ITW)*, Cambridge, UK, Sep. 2016.

[33] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. Wiley-Interscience, 2006.

[34] S. Gadkari and K. Rose, "Robust vector quantizer design by noisy channel relaxation," *IEEE Trans. Commun.*, vol. 47, no. 8, pp. 1113–1116, Aug. 1999.

[35] D. Luenberger, *Optimization by Vector Space Methods*. New York: John Wiley & Sons, Inc, 1969.

[36] H. Royden and P. Fitzpatrick, *Real Analysis*. Pearson, 2010.

Morteza Varasteh [S'13] received the B.S. degree in electrical and electronics engineering from Tabriz University, Tabriz, Iran, in 2009, and the M.S. degree in communication systems engineering from Sharif University of Technology, Tehran, Iran, in 2011, and the Ph.D. degree in communications from Imperial College London, London, UK, in 2016. He is currently a postdoctoral research associate in the Communications and Signal Processing group at Imperial College London. His research interests are in the general areas of information theory, wireless communications, and optimization theory.

Borzoo Rassouli [S'13] received his M.Sc. degree in communication systems engineering from University of Tehran, Iran, in 2012, and Ph.D. degree from Imperial College London in 2016. He is currently a postdoctoral researcher in the Intelligent Systems and Networks Group at Imperial College. His research interests are in the general areas of information theory, wireless communications, and detection and estimation theory.

Osvaldo Simeone [SM'14] is a Professor of Information Engineering with the Department of Informatics at King's College London. He received an M.Sc. degree (with honors) and a Ph.D. degree in information engineering from Politecnico di Milano, Milan, Italy, in 2001 and 2005, respectively. He was previously a Professor affiliated with the Center for Wireless Information Processing (CWiP) at the New Jersey Institute of Technology (NJIT). His research interests include wireless communications, information theory, optimization and machine learning. Dr Simeone is a co-recipient of the 2017 JCN Best Paper Award, the 2015 IEEE Communication Society Best Tutorial Paper Award and of the Best Paper Awards of IEEE SPAWC 2007 and IEEE WRECOM 2007. He was awarded a Consolidator grant by the European Research Council (ERC) in 2016. His research has been supported by the U.S. NSF, the ERC, the Vienna Science and Technology Fund, as well by a number of industrial collaborations. Dr Simeone is a co-author of a monograph, an edited book published by Cambridge University Press and more than one hundred research journal papers. He is currently a Distinguished Lecturer of the Information Theory Society and a member of the Signal Processing for Communications and Networking Technical Committee. He is a Fellow of the IEEE.

Deniz Gündüz [S'03-M'08-SM'13] received the B.S. degree in electrical and electronics engineering from METU, Turkey in 2002, and the M.S. and Ph.D. degrees in electrical engineering from NYU Polytechnic School of Engineering in 2004 and 2007, respectively. After his PhD, he served as a postdoctoral research associate at Princeton University, and as a consulting assistant professor at Stanford University. He was a research associate at CTTC in Barcelona, Spain until September 2012, when he joined the Electrical and Electronic Engineering Department of Imperial College London, UK, where he is currently a Reader in information theory and communications.

His research interests lie in the areas of communications and information theory with special emphasis on wireless networks, joint source-channel coding, and privacy in cyber-physical systems. Dr. Gündüz is an Editor of the IEEE TRANSACTIONS ON COMMUNICATIONS, and the IEEE TRANSACTIONS ON GREEN COMMUNICATIONS AND NETWORKING. He is the recipient of a Starting Grant of the European Research Council (ERC) in 2016, IEEE Communications Society Best Young Researcher Award for the Europe, Middle East, and Africa Region in 2014, Best Paper Award at the 2016 IEEE Wireless Communications and Networking Conference (WCNC), and the Best Student Paper Award at the 2007 IEEE International Symposium on Information Theory (ISIT). He served as the General Co-chair of the 2016 IEEE Information Theory Workshop, and a Co-chair of the PHY and Fundamentals Track of the 2017 IEEE Wireless Communications and Networking Conference.