Zero-Delay Source-Channel Coding with a One-Bit ADC Front End and Correlated Receiver Side Information

Morteza Varasteh†, Borzoo Rassouli†, Osvaldo Simeone* and Deniz Gündüz†
† Department of Electrical and Electronic Engineering, Imperial College London, London, UK.
* Department of Informatics, King's College London, UK.
{m.varasteh12; b.rassouli12; d.gunduz}@imperial.ac.uk, osvaldo.simeone@kcl.ac.uk.

***Abstract*—Zero-delay transmission of a Gaussian source over an additive white Gaussian noise (AWGN) channel is considered with a one-bit analog-to-digital converter (ADC) front end and a correlated side information at the receiver. The design of the optimal encoder and decoder is studied for two different performance criteria, namely the mean squared error (MSE) distortion and the distortion outage probability (DOP), under an average power constraint on the channel input. For both criteria, necessary optimality conditions for the encoder and the decoder are derived, which are then used to numerically obtain encoder and decoder mappings that satisfy these conditions. Using these conditions, it is observed that the numerically optimized encoder (NOE) under the MSE distortion criterion is periodic, and its period increases with the correlation between the source and the receiver side information. For the DOP, it is instead seen that the NOE mappings periodically acquire positive and negative values, which decay to zero with increasing source magnitude, and the interval over which the mapping takes non-zero values becomes wider with the correlation between the source and the side information. Finally, inspired by the mentioned properties of the NOE mappings, parameterized encoder mappings with a small number of degrees of freedom are proposed for both distortion criteria, and their performance is compared with that of the NOE mappings.**

***Index Terms*- Joint source channel coding, zero-delay transmission, mean squared error distortion, distortion outage probability, one-bit ADC, correlated side information.**

I. Introduction

Current wireless communication systems enable reliable transmission of specific high-rate content types, such as JPEG and MPEG, by exploiting near capacity-achieving channel codes and highly optimized compression algorithms. However, many emerging applications, such as the Internet-of-Things (IoT) or machine-to-machine (M2M) communications, impose further constraints on the cost and complexity of communication devices, or on the available energy and the end-to-end latency, which render many of the known codes and modulation techniques inapplicable. For example, in time-sensitive control applications, such as the monitoring of power lines for attacks or failures in a smart grid or the detection and prevention of epileptic seizures through embedded sensors, the underlying signals should be measured and transmitted to the receiving-end under extreme latency constraints. In such scenarios, neither measuring multiple signals to improve the compression efficiency nor using the channel many times to approach the channel capacity is possible. Here, we model such a communication scenario with the extreme *zero-delay* constraint, imposing the transmission of a single sample of the underlying signal over a single use of the channel.

A key component of the front end of any digital receiver is the analog-to-digital converter (ADC) that is typically connected to each receiving antenna. The energy consumption of an ADC (in Joules/sample) increases exponentially with its resolution (in bits/sample) [2]. This is causing a growing concern regarding the energy consumption of digital receivers, either due to the increasing number of receiving antennas, e.g., for massive multiple-input multiple-output (MIMO) transceivers [3], or due to the limited availability of energy, e.g., in energy harvesting terminals [4]. Energy-efficient operation of digital receivers may hence impose constraints on the resolution of the ADCs that can be employed for each receiving antenna.

Motivated by communication among energy- and complexity-limited sensor nodes under extreme latency constraints, we study the zero-delay transmission of analog sensor measurements to a receiver equipped with a 1-bit ADC front end. Considering that the transmitter and the receiver should be physically close to each other when communicating at low power, we further assume that the receiving node has its own correlated measurement of the transmitted source sample (see Figure 1). We consider two standard performance criteria, namely the mean squared error (MSE) distortion and the distortion outage probability (DOP). Our goal here is to gain insights into the structure and the performance of the optimal encoder and decoder functions when the source sample and the side information are jointly Gaussian.

This work contributes to a line of research that endeavors to

M. Varasteh and B. Rassouli have been supported by the British Council Institutional Links Program under grant number 173605884.

D. Gunduz has received funding from the European Research Council (ERC) through Starting Grant BEACON (agreement No. 677854).

The work of O. Simeone has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 725731), and was partially supported by the U.S. NSF through grant CCF-1525629.

This work was presented in part at the 2016 IEEE Information Theory Workshop [1].

understand the impact of front-end ADC limitations on the fundamental performance limits of communication systems. The capacity analysis of a real discrete-time AWGN channel with a K-level ADC front end is studied in [5], proving the sufficiency of $K + 1$ constellation points at the encoder. Furthermore, it is shown in [5] that BPSK modulation achieves the capacity when the receiver front end is limited to a 1-bit ADC. In [6], the authors prove that, in the low signal-to-noise ratio (SNR) regime, the symmetric threshold 1-bit ADC is suboptimal, while asymmetric threshold quantizers and asymmetric signalling constellations are needed to obtain the optimal performance. The generalization of the analysis to multiple-input multiple-output (MIMO) fading systems is put forth in [7], and, more recently, to massive MIMO systems in [3] and [8]. In [9] the authors of this work considered the zero-delay transmission set-up analysed here, but in the absence of correlated side information at the receiver. It is noted that the zero-delay constraint prevents the application of the channel capacity results in [5], [8], and as it will be seen, the presence of correlated side information at the receiver significantly modifies the optimal design problem with respect to the set-up studied in [9].

The main contributions of this work are as follows. We derive necessary optimality conditions for encoder and decoder mappings for both of the performance criteria under consideration, namely the MSE and the DOP. Then, we numerically obtain an encoder mapping that satisfies the corresponding necessary condition through the gradient descent algorithm. For the MSE criterion, we observe that, in a manner similar to the case with an infinite resolution front end studied in [10]–[12], the numerically optimized encoder (NOE) mapping is periodic. Furthermore, the period of this function depends on the correlation coefficient between the source and the side information, and is independent of the input power constraint, or equivalently the channel SNR. Motivated by the structure of the NOE mappings, we also propose simple parameterized mappings, which, although being suboptimal, approach the performance of NOE mappings, while requiring significantly less computations for optimization. For the DOP criterion, we observe that the NOE mappings periodically acquire positive and negative values with decaying magnitude with the absolute value of the source. It is also observed that, as the correlation between the source and the side information increases, the number of changes between positive and negative values in the encoder mapping, as well as the size of the intervals of the source output values for which the encoder mapping is non-zero, increase. A parameterized encoder mapping is also proposed for the DOP criterion and its performance is compared with that of the NOE mapping.

The rest of the paper is organized as follows. In Section II, the system model is explained. Section III focuses on the MSE criterion. We study the optimal design of the encoder and the decoder in Section III-A, while Sections III-B, III-C and III-D present different parameterized encoding schemes. In Section III-E, we consider the scenario in which the side information is also available at the encoder, and, by leveraging the results in [9], we obtain a lower bound on the performance of the original problem with decoder-only side information. As another reference result, in Section III-F, we present the Shannon lower bound for the decoder-only side information problem. Focusing on the DOP criterion in Section IV, we first consider the optimal design of the encoder and decoder in Section IV-A. Next, in Section IV-B, we propose a suboptimal parameterized encoder mapping under the DOP criterion. In Section IV-C, as for the MSE counterpart, we consider the case in which the side information is also available at the encoder. In Section V, numerical results are provided, and Section VI concludes the paper.

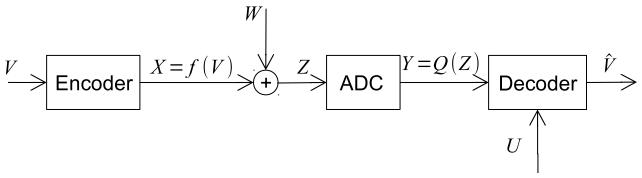

Figure 1. System model for the zero-delay transmission of a Gaussian source sample over an AWGN channel with a one-bit ADC front end and correlated side information at the receiver.

Notations: Throughout the paper upper case and lowercase letters denote random variables and their realizations, respectively. The standard normal distribution is denoted by $\mathcal{N}(0,1)$, and its probability density function (pdf) by $\Phi(\cdot)$. $\mathbb{E}[\cdot]$ and $\Pr(\cdot)$ stand for the expectation and probability, respectively. $Q(\cdot)$ denotes the complementary cumulative distribution function (CCDF) of the standard normal distribution, defined as

$$Q(z) \triangleq \frac{1}{\sqrt{2\pi}} \int_{z}^{\infty} e^{-\frac{x^2}{2}} dx. \tag{1}$$

The boundaries of integrals are from $-\infty$ to ∞ unless stated otherwise. We denote the pdf of a standard bivariate normal distribution with correlation r as

$$\Phi(v,u) = \frac{1}{2\pi\sqrt{1-r^2}} e^{-\frac{1}{2(1-r^2)}\left(v^2+u^2-2rvu\right)}, \tag{2}$$

and the conditional pdf for these variables as

$$\Phi(v|u) = \frac{1}{\sqrt{2\pi(1-r^2)}} e^{-\frac{(v-ru)^2}{2(1-r^2)}}. \tag{3}$$

II. System Model

We consider the system model in Figure 1, in which a single Gaussian source sample $V \sim \mathcal{N}(0,\sigma_v^2)$ is transmitted over a single use of a channel characterized by AWGN followed by a one-bit ADC front end. Unlike the model studied in [9], the receiver has access to side information $U \sim \mathcal{N}(0,\sigma_u^2)$, which is correlated with the source V. The correlation matrix of the source and the side information is given by

$$\Lambda = \begin{bmatrix} \sigma_v^2 & r\sigma_v\sigma_u \\ r\sigma_v\sigma_u & \sigma_u^2 \end{bmatrix}, \tag{4}$$

where $r \in [-1, 1]$ denotes the correlation coefficient.

The encoded signal is obtained as $X = f(V)$, where $f : \mathbb{R} \to \mathbb{R}$ is a mapping that is constrained to satisfy an average power constraint $\mathbb{E}[f(V)^2] \leq P$. At the receiver, the received noisy signal is modelled as

$$Z = f(V) + W, \tag{5}$$

where $W \sim \mathcal{N}(0, \sigma_w^2)$ is independent of the source and side information. The noisy signal Z is quantized with a one-bit ADC producing the received signal as

$$Y = Q(Z) = \begin{cases} 0 & Z \geq 0, \\ 1 & Z < 0. \end{cases} \tag{6}$$

We define the signal-to-noise ratio (SNR) as $\gamma = P/\sigma_w^2$. Based on Y and U, the decoder produces an estimate $\hat{V}$ of V using a decoding function $g : \{0, 1\} \times \mathbb{R} \to \mathbb{R}$, i.e., $\hat{V} = g(Y, U)$.

Two performance criteria are considered in this paper, namely, the MSE distortion

$$\bar{D} = \mathbb{E}\left[(V - \hat{V})^2\right], \tag{7}$$

and the DOP

$$\epsilon(D) = \Pr\left((V - \hat{V})^2 \geq D\right). \tag{8}$$

In both cases, we aim at studying the optimal encoder mapping f, along with the corresponding optimal estimator g at the receiver, such that $\bar{D}$ and $\epsilon(D)$ are minimized subject to the average power constraint. More specifically, as it is common in related works (see, e.g., [10]), we consider the unconstrained minimization

$$\underset{f,g}{\text{minimize}} \quad L(f, g, \lambda), \tag{9}$$

where

$$L(f, g, \lambda) = \begin{cases} \bar{D} + \lambda E[f(V)^2] & \text{for the MSE criterion,} \\ \epsilon(D) + \lambda E[f(V)^2] & \text{for the DOP criterion,} \end{cases} \tag{10}$$

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. MSE distortion criterion

In this section, we study the performance of the system model in Figure 1 under the MSE distortion criterion. In the following, we first consider the optimal design of the encoder and the decoder, and obtain a necessary condition for the optimality of an encoder mapping. Then we numerically obtain an encoder mapping that aims at satisfying these necessary conditions via gradient descent. As lower complexity alternatives, and inspired by the shape of the obtained NOE mappings, we propose three different parameterized encoding schemes, namely periodic linear transmission (PLT), periodic BPSK transmission (PBT) and multi-sine transmission (MST). Then, as lower bounds, we consider the MSE in the presence of side information at both the encoder and the decoder, and the Shannon lower bound.

A. Optimal Encoder and Decoder Design

The design goal is to minimize the Lagrangian in (9) for the MSE distortion criterion. The following proposition provides a necessary condition for the optimal encoder mapping.

Proposition III.1. *The optimal encoder mapping f for problem (9) under the MSE distortion criterion must satisfy the implicit equation*

$$2\sqrt{2\pi}\sigma_w\sigma_u\lambda f(v)e^{\frac{f(v)^2}{2\sigma_w^2}} = 2vA(v) - B(v), \tag{11}$$

where $\lambda \geq 0$ and is given. The functions $A(v)$ and $B(v)$ are defined as

$$A(v) \triangleq \int \Phi\left(\frac{u}{\sigma_u}\middle|\frac{v}{\sigma_v}\right)(g(0, u) - g(1, u))\, du, \tag{12a}$$

$$B(v) \triangleq \int \Phi\left(\frac{u}{\sigma_u}\middle|\frac{v}{\sigma_v}\right)\left(g(0, u)^2 - g(1, u)^2\right) du, \tag{12b}$$

where $g(y, u)$, for $y = 0, 1$, is the optimal MMSE estimator given by

$$g(y, u) = \frac{\int v\Phi\left(\frac{v}{\sigma_v}\middle|\frac{u}{\sigma_u}\right) Q\left(\frac{(-1)^{y+1}f(v)}{\sigma_w}\right) dv}{\int \Phi\left(\frac{v}{\sigma_v}\middle|\frac{u}{\sigma_u}\right) Q\left(\frac{(-1)^{y+1}f(v)}{\sigma_w}\right) dv}. \tag{13}$$

Furthermore, the gradient of the Lagrangian function $L(f, g, \lambda)$ over f, for g given as in (13), is given by

$$\nabla L = 2\lambda f(v) - \frac{e^{-\frac{f(v)^2}{2\sigma_w^2}}}{\sqrt{2\pi}\sigma_w\sigma_u}(2vA(v) - B(v)). \tag{14}$$

Proof: See Appendix B.

Remark III.1. *The parameter λ in (11) is the Lagrange multiplier, which is unique for a given power constraint [13, Theorem 3.3.1][1]. We also remark that, by following steps similar to [10, Theorem 3], the same optimality condition in (11) can also be derived by considering the decoder $g(\cdot, \cdot)$ as a fixed functional of the encoder $f(\cdot)$. Hence, condition (11) justifies an alternating optimization scheme in which f and g are optimized iteratively while keeping one of them fixed at each iteration.*

To elaborate on the necessary condition obtained in (11), we consider two extreme values of the correlation coefficient r. If we have an independent side information, that is, when $r = 0$, it can be easily verified that the condition (11) coincides with the result obtained in [9, Proposition III.1], where there is no side information at the receiver. The optimal mapping in this case is an odd function. Plot of the optimal encoder mapping for different SNR values is shown in Figure 7. It can be seen that, for high SNR (large γ), the mapping tends to binary antipodal signalling, whereas for low SNR (small γ), it tends to a linear mapping. In contrast, with perfect side information, i.e., $r = \pm 1$, we have $\Phi(u/\sigma_u|v/\sigma_v) = \pm\frac{\sigma_u}{\sigma_v}\delta(u - v)$, where $\delta(\cdot)$ is the Dirac delta function, and, from (13), it is seen that $g(y, u)$ is

[1]Uniqueness of the Lagrange multiplier λ, requires that the solution is regular [13, Section 3.3.1]. The regularity of a solution is verified, since there is only one constraint.

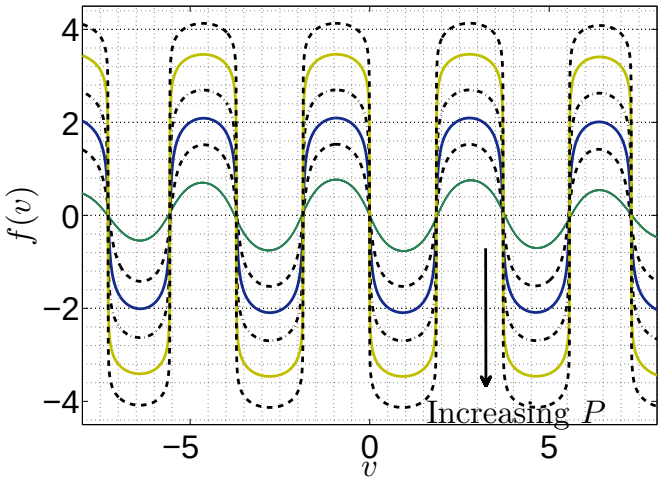

Figure 2. NOE mappings under the MSE criterion with different average power values and $r = 0.85$ ($\sigma_v = \sigma_w = 1$). Increasing the power constraint P has no impact on the period of the NOE mapping, which instead depends on r (see Figure 3).

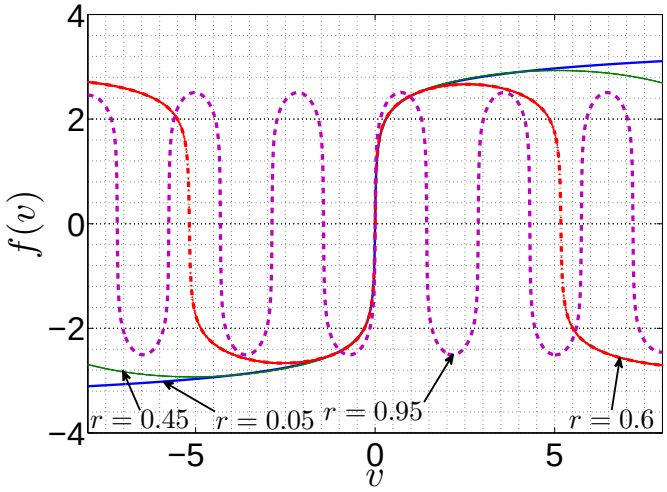

Figure 3. NOE mappings under the MSE criterion for different correlation coefficients r and an average power constraint $P = 5$ ($\sigma_v^2 = \sigma_w^2 = 1$).

the MMSE estimate of V given U, namely $g(y,u) = \frac{\sigma_v}{\sigma_u}u$. Therefore, from (12a) and (12b) along with (11) it is verified that $f(\cdot) = 0$.

In the following, we present NOE mappings obtained via gradient descent by using the gradient (14). It is observed that, due to the correlated receiver side information, the resulting encoder mappings are periodic, with a period that depends on the correlation coefficient r. Similar periodic mappings have been found to be optimal in [10] for the case with an infinite-resolution front end.

NOE mappings under the MSE distortion criterion: In order to derive the NOE mappings we apply a gradient descent-based iterative algorithm. The algorithm performs a gradient descent search in the opposite direction of the derivative of the Lagrangian (10) with respect to the encoder mapping $f(\cdot)$. The update is obtained by

$$f_{i+1}(v) = f_i(v) - \mu \nabla_f L, \tag{15}$$

where i is the iteration index, $\nabla_f L$ is defined in (14), and $\mu > 0$ is the step size. The algorithm can be initialized with an arbitrary mapping. Here, we use a linear mapping with slope close to zero for initialization. It is noted that the algorithm is not guaranteed to converge to a global optimal solution.

In Figure 2, NOE mappings for the MSE distortion criterion obtained using the aforementioned gradient descent algorithm are plotted for different average power constraints, for a correlation coefficient of $r = 0.85$. We note the periodic structure of the mappings, which is in line with the results in [10] for an infinite resolution front end. In contrast, the optimal mapping obtained in [9] when $r = 0$ is a monotonically increasing function. We also observe that the average power constraint P does not affect the period of the mapping. In Figure 3, NOE mappings for an average power of $P = 5$ are plotted for different correlation coefficients. We see that the period of the mapping instead depends on r: the higher the correlation coefficient r is, the smaller the period of the mapping is.

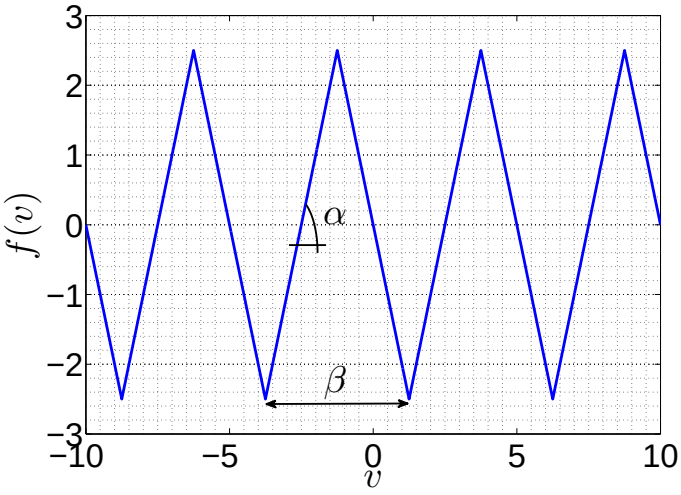

Figure 4. Illustration of th.e PLT encoder mapping for $\alpha = 2$ and $\beta = 2.5$.

Furthermore, Figure 2 shows that the SNR, or P, affects the shape of the encoder mapping in each period, where for low SNR the optimal mapping resembles a periodic linear function, while for high SNR, the optimal mapping resembles a periodic BPSK encoder. Motivated by these observations and by the results in [9] for the case of no side information, we now propose three simple parameterized encoder mappings, whose performance will be compared with the NOE mapping in Section V.

B. Periodic Linear Transmission (PLT)

Targeting the low-SNR regime, the first proposed encoder mapping is a periodic linear function with period β and slope α within each period. The encoder function is given by

$$f_{\text{PLT}}(v) = \alpha(-1)^{\left\lfloor \frac{v}{\beta}+\frac{1}{2} \right\rfloor} \left(\beta \left\lfloor \frac{v}{\beta} + \frac{1}{2} \right\rfloor - v \right), \tag{16}$$

where $\lfloor x \rfloor$ is the largest integer less than or equal to x. In Figure 4, an illustration of this mapping for $\alpha = 2$ and $\beta = 2.5$ is shown. To satisfy an average power constraint of P, the following condition must be satisfied by (α, β)

$$\begin{aligned}&\mathbb{E}[f(V)^2] = \\ &\alpha^2\left(\sigma_v^2 + \beta^2 \sum_{i=-\infty}^{\infty} i^2 \left(Q\left(\frac{-\frac{\beta}{2}+i\beta}{\sigma_v}\right) - Q\left(\frac{\frac{\beta}{2}+i\beta}{\sigma_v}\right)\right)\right. \\ &\left. -\frac{2\beta\sigma_v}{\sqrt{2\pi}} \sum_{i=-\infty}^{\infty} i\left(e^{-\frac{\left(-\frac{\beta}{2}+i\beta\right)^2}{2\sigma_v^2}} - e^{-\frac{\left(\frac{\beta}{2}+i\beta\right)^2}{2\sigma_v^2}}\right)\right) \le P. \end{aligned} \tag{17}$$

Remark III.2. *The parameters α and β are optimized under a given average power constraint P in order to minimize the MSE distortion $\bar{D}$; that is, we solve the following optimization problem*

$$\begin{aligned} &\underset{\alpha,\beta}{\text{minimize}} \quad \bar{D} \\ &\text{s.t.} \quad \mathbb{E}\left[f(V)^2\right] \le P, \end{aligned} \tag{18}$$

where $\bar{D}$ is given by

$$\begin{aligned}\bar{D} =& \sigma_v^2 - \frac{1}{\sigma_u \sigma_v} \iint v\Phi\left(\frac{v}{\sigma_v}, \frac{u}{\sigma_u}\right) \\ &\cdot \left(g(1,u)Q\left(\frac{f(v)}{\sigma_w}\right) + g(0,u)Q\left(\frac{-f(v)}{\sigma_w}\right)\right) dudv. \end{aligned} \tag{19}$$

We solve the above optimization problem numerically using the interior-point algorithm implemented by the fmincon *function in MATLAB software, where a large value for β, and an α such that the power constraint P is satisfied with equality, are chosen as initialization.*

C. Periodic BPSK Transmission (PBT)

The second proposed encoder mapping, unlike PLT, targets the high-SNR regime and adopts digital modulation with two levels, namely, γ and $-\gamma$, with a period of δ. The mapping is defined as

$$f_{\text{PBT}}(v) = \gamma\left(1 + 2(1 - 2Q(v)) \cdot \text{mod}\left(\left\lfloor \frac{2v}{\delta} \right\rfloor\right)_2\right), \tag{20}$$

where $\text{mod}(\cdot)_2$ return its argument modulo 2. In Figure 5, an illustration of this mapping for $\gamma = 0.2$ and $\delta = 2.5$ is shown. The numerical optimization over γ and δ is also performed using the fmincon function.

D. Multi-Sine Transmission (MST)

Due to the periodic nature of the NOE mapping under the MSE distortion criterion, as seen in Figures 2 and 3, the NOE can be approximated by a sum of sinusoidal functions with odd frequencies. Accordingly, we consider the following parametrization of the encoder mapping $f(\cdot)$ given as

$$\begin{aligned} f_{\text{MST}}(v) =& m_0 \sin(wv) + m_1 \sin(3wv) \\ &+ m_2 \sin(5wv) + m_3 \sin(7wv), \end{aligned} \tag{21}$$

where m_0, m_1, m_2, m_3 and w are the optimization parameters. In Figure 6, an illustration of this mapping is shown for $m_0 = 1$, $m_1 = 0.2$, $m_2 = 0.06$, $m_3 = 0.015$ and $w = 1$. Optimization over the parameters m_0, m_1, m_2, m_3 and w is performed by using the interior-point algorithm as explained in Remark III.2.

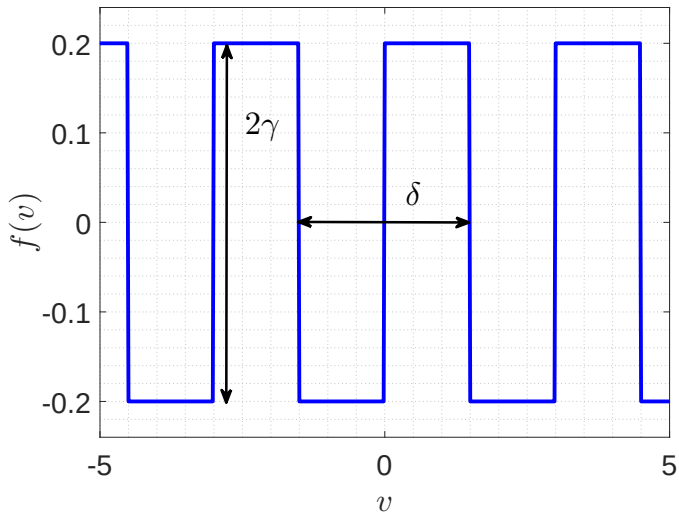

Figure 5. Illustration of the PBT encoder mapping for $\gamma = 0.2$ and $\delta = 3$.

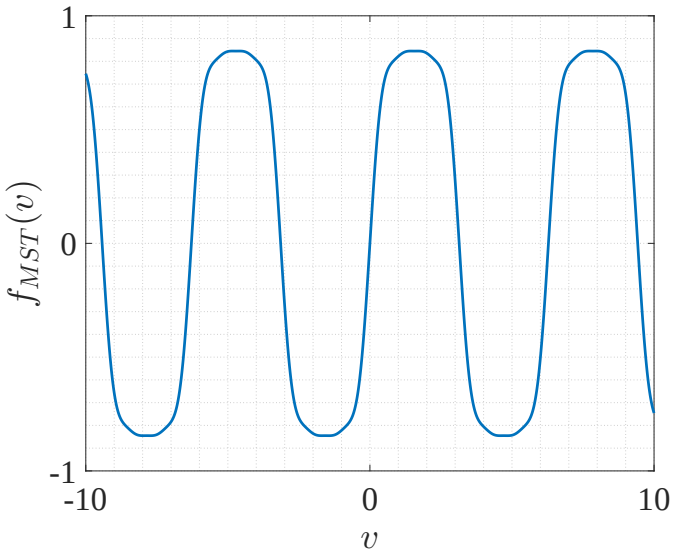

Figure 6. Illustration of the MST encoder mapping for $m_0 = 1$, $m_1 = 0.2$, $m_2 = 0.06$, $m_3 = 0.015$ and $w = 1$.

E. Side Information Available at Both the Encoder and Decoder

Here, we consider the scenario in which both the encoder and the decoder have access to the side information U. Source V can be written as a linear combination of the side information U and an independent noise term. Since the first term is known perfectly both at the encoder and the decoder, all transmission power should be used for the transmission of the independent component. Therefore, without loss of optimality, the encoder can encode the error

$$T = V - \frac{\sigma_v}{\sigma_u} rU, \tag{22}$$

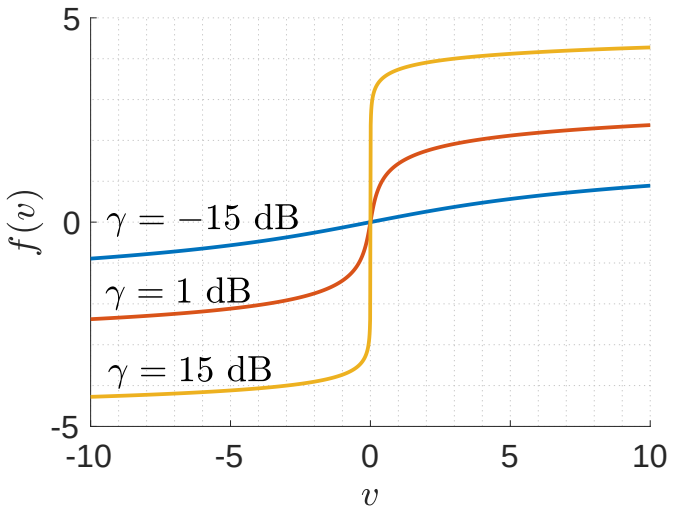

Figure 7. Illustration of the optimal encoder mapping when there is no side information at the receiver, i.e., $r = 0$ ($\sigma_v^2 = \sigma_w^2 = 1$).

where the random variable $\sigma_v r U/\sigma_u$ is the MMSE estimate of V given U, which can be computed at both the encoder and the decoder. Since the random variable T, which is distributed as $\mathcal{N}(0, \sigma_t^2)$, with $\sigma_t^2 = \sigma_v^2(1-r^2)$, is independent of the side information U, the encoder can directly encode the error T via a mapping function $\tilde{f}(t)$ ignoring the presence of the side information U at the receiver. Therefore, the problem reduces to the one studied in [9] and discussed in Section III-A (see Figure 7). As a result, a mapping $f(v) = \tilde{f}(v - \sigma_v r u/\sigma_u)$ is optimal, where $\tilde{f}(\cdot)$ is the optimal mapping with no side information shown in Figure 7. Therefore, the optimal mapping is centred on the MMSE estimate $\sigma_v r u/\sigma_u$. We will see in Section V that, when the side information is not available at the encoder, the NOE consists of periodic replicas of a mapping similar to $\tilde{f}(\cdot)$ in Figure 7.

F. Shannon Lower Bound (SLB)

A lower bound on the MSE distortion can be obtained by relaxing the zero-delay constraint, and using the Shannon's source-channel separation theorem. In [5], it is shown that the capacity of the AWGN channel with a 1-bit ADC in (6) is given by

$$C = 1 - h\left(Q\left(\sqrt{\text{SNR}}\right)\right), \tag{23}$$

where $h(\cdot)$ is the binary entropy function defined as $h(p) \triangleq -p\log_2 p - (1-p)\log_2(1-p)$. Furthermore, the rate-distortion function of a Gaussian source with correlated Gaussian side information at the receiver is given by the Wyner-Ziv rate-distortion function [14]

$$R(\bar{D}) = \frac{1}{2}\left[\log_2 \frac{\sigma_v^2(1-r^2)}{\bar{D}}\right]^+, \tag{24}$$

where $[x]^+ = \max(0, x)$. Combining (23) and (24) a lower bound on the MSE distortion $\bar{D}$ is obtained as

$$\bar{D}_{\text{lower}} = (1-r^2)\sigma_v^2 2^{-2\left(1-h\left(Q\left(\sqrt{\text{SNR}}\right)\right)\right)}. \tag{25}$$

IV. DOP Criterion

In this section, we consider the optimization of the system in Figure 1 under the DOP criterion. We first derive necessary conditions for an optimal encoder and decoder pair. Then we obtain a lower bound by considering the availability of the side information also at the transmitter.

A. Optimal Encoder and Decoder Design

We first obtain the necessary optimality condition of an encoder mapping f for a given decoder g. Then, we obtain the optimal decoder g for a given encoder mapping f.

Optimal encoder: For a fixed decoder function $g(y,u)$, we define the intervals

$$I_y(u) = \left\{v : (v - g(y,u))^2 < D\right\}, \ y = 0, 1. \tag{26}$$

Each interval $I_0(u)$ and $I_1(u)$ in (26) corresponds to the set of source values that are within the allowed distortion target D of the reconstruction points $g(0,u)$ and $g(1,u)$, respectively, when the side information is $U = u$. Hence, the following claims hold: (*i*) For all source realizations v in the set $(I_0(u) \cup I_1(u))^C = \{v : \min_{y=0,1}(v - g(y,u))^2 \geq D\}$, outage occurs since no reconstruction point $g(y,u)$ satisfies the distortion constraint (superscript C denotes the complement set). We refer to this event as *source outage*. (*ii*) For all source realizations in the interval $I_0(u) \cap I_1(u)$, either of the reconstruction points yields a distortion not larger than the target value D. Therefore, regardless of which output $(g(0,u), g(1,u))$ is selected by the receiver, no outage occurs.

With these observations in mind, the next proposition characterizes the optimal encoder mapping f for a given decoder g.

Proposition IV.1. *Given a target distortion D, and a decoder with reconstruction function $g(\cdot,\cdot)$, the optimal mapping $f(\cdot)$ for the problem (10) under the DOP criterion satisfies*

$$f(v) = \frac{e^{-\frac{f(v)^2}{2\sigma_w^2}}}{2\lambda\sqrt{2\pi}}\Big(Pr\left(U \in S_{0\setminus 1}(v)\right) - Pr\left(U \in S_{1\setminus 0}(v)\right)\Big). \tag{27}$$

where $S_{0\setminus 1}(v)$ and $S_{1\setminus 0}(v)$ are defined as

$$\begin{aligned} S_{0\setminus 1}(v) &\triangleq \{u : \ b_{0l}(u) \leq v \leq b_{0r}(u)\}, \\ S_{1\setminus 0}(v) &\triangleq \{u : \ b_{1l}(u) \leq v \leq b_{1r}(u)\}, \end{aligned} \tag{28}$$

and $b_{1r}(u)$, $b_{1l}(u)$, $b_{0r}(u)$ and $b_{0l}(u)$ are defined as below

$$b_{0r}(u) \triangleq \begin{cases} g(0,u) + \sqrt{D} & g(0,u) \geq g(1,u), \\ \min\left\{g(1,u) - \sqrt{D}, g(0,u) + \sqrt{D}\right\} & g(0,u) < g(1,u), \end{cases} \tag{29a}$$

$$b_{0l}(u) \triangleq \begin{cases} \max\left\{g(1,u) + \sqrt{D}, g(0,u) - \sqrt{D}\right\} & g(0,u) \geq g(1,u), \\ g(0,u) - \sqrt{D} & g(0,u) < g(1,u), \end{cases} \tag{29b}$$

$$b_{1r}(u) \triangleq$$

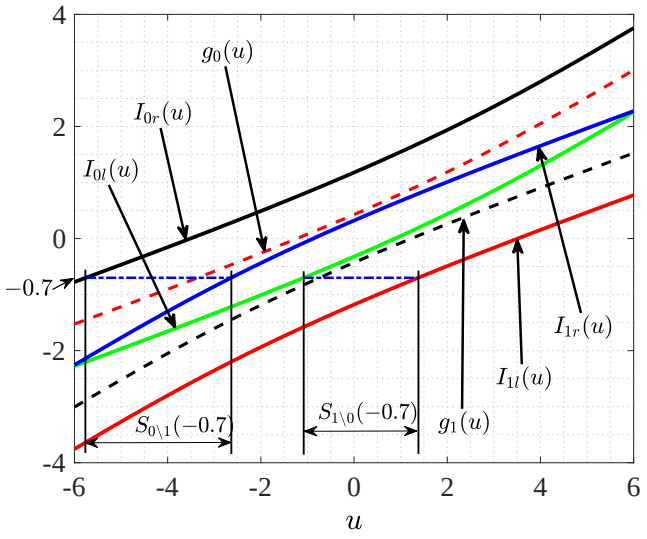

Figure 8. Illustration of the intervals $S_{0\setminus1}(v)$ and $S_{1\setminus0}(v)$ for $v = -0.7$, and functions $I_{0r}(u), I_{0l}(u), I_{1r}(u), I_{1l}(u)$ for a given decoder mapping $g(Y, U)$.

$$\begin{cases} \min\left\{g(1,u)+\sqrt{D}, g(0,u)-\sqrt{D}\right\} & g(0,u) \geq g(1,u), \\ g(1,u)+\sqrt{D} & g(0,u) < g(1,u), \end{cases} \tag{29c}$$

$$b_{1l}(u) \triangleq \begin{cases} g(1,u)-\sqrt{D} & g(0,u) \geq g(1,u), \\ \max\left\{g(1,u)-\sqrt{D}, g(0,u)+\sqrt{D}\right\} & g(0,u) < g(1,u). \end{cases} \tag{29d}$$

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

$$\nabla L = 2\lambda f(v) - \frac{e^{-\frac{f(v)^2}{2\sigma_w^2}}}{\sqrt{2\pi}} \cdot \Big(Pr\left(U \in S_{0\setminus1}(v)\right) - Pr\left(U \in S_{1\setminus0}(v)\right)\Big). \tag{30}$$

Proof: See Appendix C.

The sets $S_{0\setminus1}(v)$ and $S_{1\setminus0}(v)$ and functions $I_{0r}(u), I_{0l}(u), I_{1r}(u), I_{1l}(u)$ are illustrated in Figure 8 for an arbitrary decoding function $g(Y, U)$.

Optimal decoder: Assuming that the encoder mapping f is given, we now aim to minimize the Lagrangian function in (10) for the DOP criterion over the decoding function g. The next proposition characterizes the optimal decoder mapping for a given encoder f.

Proposition IV.2. *Given a target distortion D and an encoder mapping $f(\cdot)$, the optimal decoder $g(\cdot,\cdot)$ for the problem (10) under the DOP criterion is obtained as*

$$g(y,u) \in \arg\max_{\hat{v}} \int_{\hat{v}-\sqrt{D}}^{\hat{v}+\sqrt{D}} \Phi\left(\frac{v}{\sigma_v}\Big|\frac{u}{\sigma_u}\right) Q\left(\frac{(-1)^{y+1} f(v)}{\sigma_w}\right) dv. \tag{31}$$

Proof: See Appendix D.

To provide further insights into Propositions IV.1 and IV.2, it is worth considering the two extreme cases of side information correlation. When $r = 0$ the decoder outputs $g(0,u) = \hat{v}_0$ and $g(1,u) = \hat{v}_1$ are independent of the side information U, and the conditions derived here coincide with those obtained in [9, Proposition IV.1] for the optimal mapping when there is no side information at the receiver. Instead, with $r = 1$, we can choose $g(y,u) = g(u) = \pm\frac{\sigma_v}{\sigma_u}u$; and hence, we have $f(v) = 0$ for all v.

Remark IV.1. *In the low SNR regime, from (31), we have $g(y,u) \simeq \frac{r\sigma_v}{\sigma_u}u$, $y = 0, 1$. Therefore, in the asymptotic low SNR regime, the DOP at the receiver is found as (see Appendix E)*

$$\lim_{SNR\to 0} \epsilon(D) = 2Q\left(\frac{\sqrt{D}}{\sigma_v\sqrt{1-r^2}}\right). \tag{32}$$

In Section V, we validate (32) in the asymptotic low SNR regime.

Similarly to the MSE distortion, in the following, we present NOE mappings obtained via gradient descent by using (30). It will be observed that, due to the correlated receiver side information, the resulting encoder mappings periodically acquire positive and negative values, which decay to zero with increasing source magnitude, and the interval over which the mapping is nonzero becomes wider as the correlation coefficient increases.

NOE mappings under the DOP criterion: Similarly to the MSE distortion criterion, we apply a gradient descent-based iterative algorithm. The update is obtained by (15), where $\nabla_f L$ is defined in (30). The algorithm is initialized with a linear mapping with slope close to zero.

Remark IV.2. *While the optimality condition for the MSE criterion is shown to be the same as for an alternating optimization of the encoder and decoder mappings (see Remark III.1), in the case of DOP, we have a priori focused our attention to the use of alternating optimization. Therefore, under both criteria the numerical algorithm employed to obtain the NOE mappings is equivalent to the one described in [10]. We also remark that different power constraints are imposed by means of a linear search over the Lagrange multiplier λ.*

In Figure 9, NOE mappings for different power constraints and correlation coefficients are shown under the DOP criterion. For small values of the correlation coefficient, such as $r = 0.1$, the NOE mappings resemble the optimal mappings in the absence of receiver side information obtained in [9], which correspond to a binary transmitter as seen in Figure 9. We observe that the domain of the mapping is bounded, unlike for the MSE criterion, since values of the source that differ by more than $\sqrt{D}$ from the reconstruction points yield an outage irrespective of the mapping. As the correlation between the source and the side information increases, the domain of the mapping expands. Motivated by this observation, we now propose a simple parameterized encoder mapping, which will be compared with the NOE mapping in Section V.

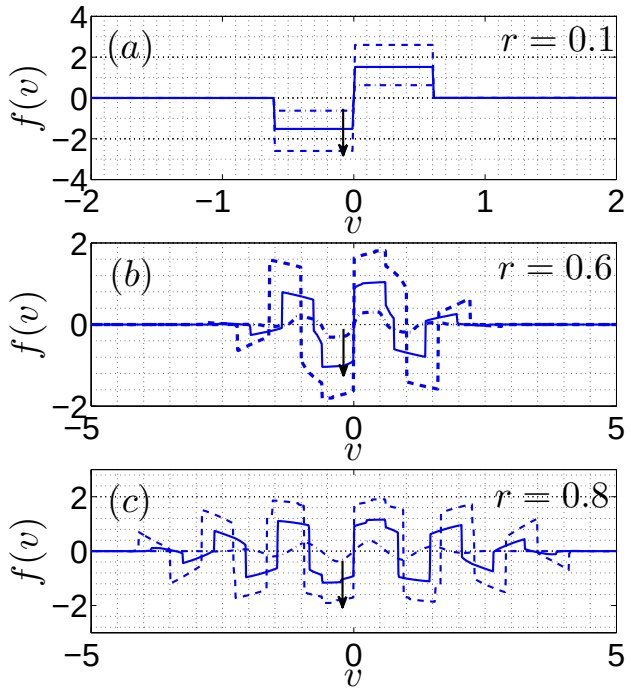

Figure 9. NOE mappings under the DOP criterion for different correlation coefficients, ($\sigma_v^2 = \sigma_w^2 = 1$). The power constraint P of the mappings increases in the direction of the arrow.

B. Periodically Decaying Transmission (PDT)

The proposed encoder mapping is a periodically decaying function parameterized with the variables δ, γ, ζ, and is given by

$$f_{\text{PDT}}(v) = \zeta^{\left\lfloor \frac{2|v|}{\delta} \right\rfloor} f_{\text{PBT}}(v), \tag{33}$$

where $f_{\text{PBT}}(v)$ is as defined in (20) and $0 \leq \zeta \leq 1$. In Figure 10, an illustration of this mapping for $\zeta = 0.5$, $\delta = 3$ and $\gamma = 0.2$ is shown. The optimization over ζ, δ and γ in order to minimize the DOP is performed by the interior-point algorithm as explained in Remark III.2.

C. Side Information Available at Both Encoder and Decoder

When the side information U is also available at the encoder, using the optimal decoder under the DOP criterion, the encoder can reconstruct the source as

$$\arg\min_{\hat{v}} \Pr(|V - \hat{v}|^2 \geq D | U = u) \tag{34a}$$

$$= \arg\max_{\hat{v}} \int_{\hat{v}-\sqrt{D}}^{\hat{v}+\sqrt{D}} \Phi\left(\frac{v}{\sigma_v} \middle| \frac{u}{\sigma_u}\right) dv \tag{34b}$$

$$= \frac{\sigma_v}{\sigma_u} r u. \tag{34c}$$

From (34), we conclude that the best estimate of a Gaussian source from Gaussian side information under the DOP criterion equals the MMSE estimate. Given that (34c) can be reconstructed at both encoder and decoder, as in Section III-E, the optimal encoder uses the optimal mapping for the scenario without side information [9] as applied to the error signal in (22). In the following proposition, we show that the optimal decoder under the DOP criterion is obtained by summing the estimates computed on the basis of the side information u and the channel output y, separately.

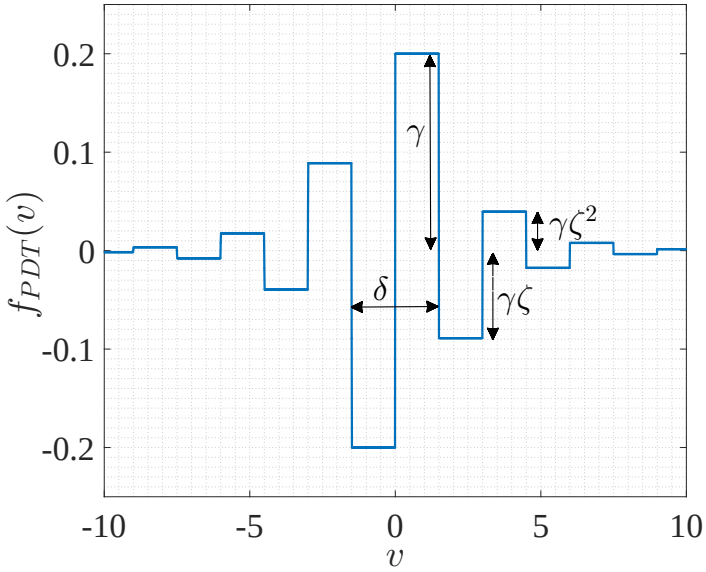

Figure 10. Illustration of the PDT encoder mapping for $\zeta = 0.5$, $\gamma = 0.2$ and $\delta = 3$.

Proposition IV.3. *Given a target distortion D and side information U being available at both encoder and decoder, the optimal decoder $g(\cdot,\cdot)$ for the problem (10) under the DOP criterion is obtained as*

$$g(y,u) = \frac{r\sigma_v}{\sigma_u} u + \hat{t}_y, \quad y = 0, 1, \tag{35}$$

where $\hat{t}_y$ represents the optimal decoder for a Gaussian source with variance $(1-r^2)\sigma_v^2$ as a function of Y, which is given in [9, Proposition IV.2]. The resulting minimum DOP is obtained as

$$\begin{aligned}\epsilon(D) =& 2Q\left(\frac{2\sqrt{D}-a}{\sqrt{1-r^2}\sigma_v}\right) + 2Q\left(\frac{t}{\sigma_w}\right) \\ &\cdot\left(Q\left(\frac{a}{\sqrt{1-r^2}\sigma_v}\right) - Q\left(\frac{2\sqrt{D}-a}{\sqrt{1-r^2}\sigma_v}\right)\right),\end{aligned} \tag{36}$$

where t is the solution of the equation $te^{\frac{t^2}{2\sigma_w^2}} = \frac{1}{2\sqrt{2\pi}\sigma_w\lambda}$.

Proof: See Appendix F.

V. Numerical Results

In this section, we present numerical results with the aim of assessing the performance of the encoder/ decoder pairs obtained in the previous sections. In the following, we first discuss the numerical results for the MSE distortion criterion, followed by the DOP criterion.

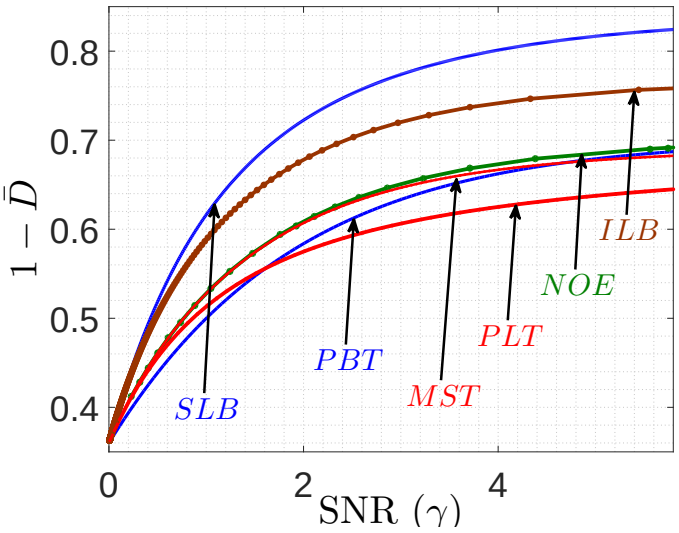

Figure 11. Complementary MSE distortion vs. SNR for $r = 0.6$ ($\sigma_v^2 = \sigma_w^2 = 1$).

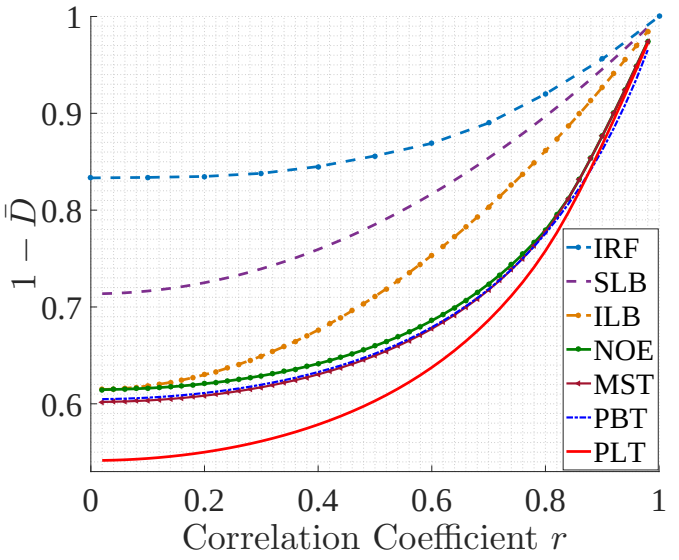

Figure 12. Complementary MSE distortion versus correlation coefficient r under the average power constraint $P = 5$ ($\sigma_v^2 = \sigma_w^2 = 1$).

MSE criterion: In Figure 11, we plot the complementary MSE distortion $(1 - \bar{D})$ versus SNR for the NOE, as well as for the PLT, MST and PBT schemes, for correlation coefficient $r = 0.6$. The SLB and the MSE distortion achieved when both the encoder and the decoder have access to the side information U, which is referred to as the informed-encoder lower bound (ILB), are also included for comparison. Following the discussion above, we observe that the performance of PBT is close to that of NOE at high SNR values. On the other hand, for low SNRs, PLT outperforms PBT and approaches the NOE performance. In the low to medium SNR values, the performance of MST is very close to the one obtained by the NOE mappings, surpassing the performance of both PBT and PLT. Its performance for the high SNR values can be improved by introducing additional harmonics, at the expense of increased complexity.

In Figure 12, we plot the complementary MSE distortion $(1 - \bar{D})$ versus the correlation coefficient r for a fixed average power constraint of $P = 5$. We note from Figure 12 that the ILB is tight in the low and high correlation regime with respect to SLB, and in general there is a loss in the MSE distortion by not having the side information at the encoder. We recall that this is not the case with infinite resolution and infinite block-length. We also observe that PLT is tighter when the correlation is higher, while it performs quite poorly when the side information quality is poor. On the other hand, for this P values PBT performs relatively close to NOE for the whole range of side information correlation values. As for MST, it is observed that in the low correlation coefficient r values, it performs similarly to PBT, while for the high correlation coefficient r values, it almost closes that gap with NOE. For an infinite resolution front end and side information at the receiver (studied in [11]), the gradient of the Lagrangian can be computed as

$$\nabla L = 2\lambda f(v) - v \iint g^{'}(u, f(v) + w)\Phi(u|v)\Phi_W(w)dwdu, \tag{37}$$

where $g^{'}(u, f(v) + w)$ is the derivative of the MMSE estimator $g(\cdot, \cdot)$ with respect to the second argument, and is given by

$$g^{'}(x, y) = \frac{d}{dy}\frac{\int v\Phi(v|x)\Phi_W(y - f(v))dv}{\int \Phi(v|x)\Phi_W(y - f(v))dv}. \tag{38}$$

Using the resulting gradient descent algorithm, the performance of the NOE mappings with infinite resolution front end (IRF) can be evaluated, and it is included in Figure 12 for comparison. As we observe from Figure 12, an infinite resolution front end yields significant gains for low values of the correlation coefficient r, with diminishing performance gains as r grows larger due to the increasing reliance of the decoder on the side information.

DOP criterion: In Figure 13, we plot the complementary DOP, $1 - \epsilon(D)$, versus SNR for NOE mappings as well as the ILB and PDT, for correlation coefficients $r = 0,\ 0.6,\ 0.8$. We observe that in the low SNR regime the DOP achieved by NOE is close to the ILB. This is because, in the low SNR regime the source estimate can be obtained based mainly on the side information. We also observe that, as the SNR increases, the DOP saturates to the source outage probability, which is independent of the SNR. As for the PDT, it is observed that the performance gap is not negligible, however, from the complexity point of view, PDT is much simpler than the NOE mappings.

VI. Conclusions

We considered the problem of transmission of a Gaussian source over an AWGN channel to a receiver equipped with a one-bit ADC front end. We also considered the availability of a correlated side information at the receiver. We studied this problem under two distinct performance criteria, namely the

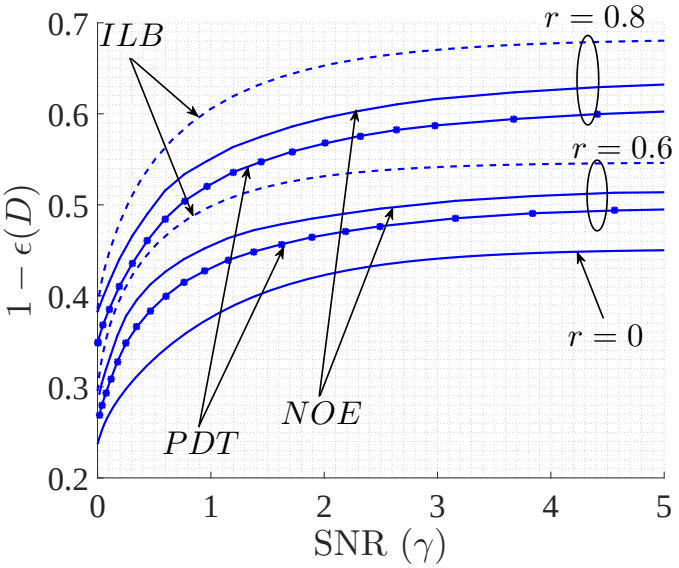

Figure 13. Complementary DOP vs. SNR for $r = 0,\ 0.6,\ 0.8$ ($\sigma_v^2 = \sigma_w^2 = 1$) and $D = 0.09$.

MSE distortion and the DOP, while imposing an average power constraint at the transmitter. Assuming that the transmission is zero-delay, in the sense that it maps every single source output to a single channel input, we obtained necessary conditions for the optimal encoder and decoder mappings under both performance criteria. In the comparison to the previous work in [9], we observed that the availability of correlated side information at the receiver has a significant impact on the shape of the optimal encoder mapping. For instance, as in the case of infinite-resolution front end [10], the optimal mapping becomes periodic under the MSE distortion criterion. We observed that the period of the optimal mapping depends on the correlation coefficient between the source and the side information, but it is not affected by the transmitter power condition. For the DOP criterion, the availability of the side information enlarges the domain of the mapping, i.e., a larger set of source sample values are mapped to a non-zero channel input. Interesting future research direction include investigating the effect of higher level ADCs on the performance of the system, obtaining optimized structures when there is fading in the channel or when there are multi observations at the receiver.

VII. Acknowledgment

The authors with to thank the anonymous reviewers for their numerous helpful comments.

VIII. Appendices

A. Preliminaries: Calculus of Variations

In the proof of the Proposition III.1, we use variational calculus [15, Section 7] to obtain necessary optimality conditions. The next theorem summarizes the key result that will be needed.

Theorem A.1. *Let F, G_0 and G_i, $i = 1, \ldots, n$, be continuous functionals of (f, H, t), $(f, r_1, \ldots, r_n, u, t)$ and (f, t, u), respectively, where H and r_i, $i = 1, \ldots, n$, are given by*

$$H(t) = \int_{t_1}^{t_2} G_0(f(t), r_1(u), \ldots, r_n(u), u, t)du, \tag{39}$$

$$r_i(u) = \int_{t_1}^{t_2} G_i(f(v), v, u)dv, \quad i = 1, \ldots, n. \tag{40}$$

Also, let F, G_0 and G_i, $i = 1, \ldots, n$, have continuous partial derivatives with respect to (f, H), $(f, r_1, \ldots, r_n)$ and f, respectively. Consider the following minimization problem

$$\underset{f}{\text{minimize}} \quad L(f) \triangleq \int_{t_1}^{t_2} F(f(t), H(t), t)dt. \tag{41}$$

Define ∇L as

$$\begin{aligned}\nabla L \triangleq\ & F^f(f(t), H(t), t) \\ & + F^H(f(t), H(t), t) \int_{t_1}^{t_2} G_0^f(f(t), r_1(u), \ldots, r_n(u), u, t)du \\ & + \int_{t_1}^{t_2}\int_{t_1}^{t_2} F^H(f(v), H(v), v) \sum_{i=1}^{n} \Big(G_i^f(f(t), t, u) \\ & \qquad \cdot G_0^{r_i}(f(v), r_1(u), \ldots, r_n(u), u, v) \Big) dvdu,\end{aligned} \tag{42}$$

where F^f and F^H denote the partial derivatives of the functional F with respect to f and H, respectively; and G_0^f and $G_0^{r_i}$ denote the partial derivatives of the functional G_0 with respect to f and r_i, respectively. Similarly, G_i^f denotes the partial derivative of the functional G_i with respect to f. A necessary condition for a function f to be a solution to the minimization problem in (41) is

$$\nabla L = 0. \tag{43}$$

Proof: Following the conventional approach in the calculus of variations, we perturb the function $f(t)$ by an arbitrary function $\eta(t)$ that vanishes on the boundary points t_1 and t_2 [15]. Let $\delta_f L \triangleq \frac{dL(f+\alpha\eta)}{d\alpha}\Big|_{\alpha=0}$ be the resulting Gateaux derivative of the functional L with respect to the parameter α. We have

$$\delta_f L = \frac{d}{d\alpha} \int_{t_1}^{t_2} F(f(t) + \alpha\eta(t), H^\alpha(t), t)dt \Big|_{\alpha=0}, \tag{44}$$

where $H^\alpha(t)$ is defined as

$$H^\alpha(t) \triangleq \int_{t_1}^{t_2} G_0(f(t) + \alpha\eta(t), r_1^\alpha(u), \ldots, r_n^\alpha(u), u, t)du, \tag{45}$$

and $r_i^\alpha(u),\ i = 1, \ldots, n$, are defined as

$$r_i^\alpha(u) \triangleq \int_{t_1}^{t_2} G_i(f(v) + \alpha\eta(v), v, u)dv, \quad i = 1, \ldots, n. \tag{46}$$

Evaluating the derivative in (44), we have

$$\delta_f L = \int_{t_1}^{t_2} \Bigg[\eta(t)F^f(f(t), H(t), t) + \frac{dH^\alpha(t)}{d\alpha}F^H(f(t), H(t), t)\Bigg]dt, \tag{47}$$

where $\frac{dH^\alpha(t)}{d\alpha}$ is the Gateaux derivative of the functional $H(t)$, which is computed as

$$\frac{dH^\alpha(t)}{d\alpha} = \frac{d}{d\alpha}\int_{t_1}^{t_2} G_0(f(t) + \alpha\eta(t), r_1^\alpha(u), \ldots, r_n^\alpha(u), u, t)du\Big|_{\alpha=0} \tag{48}$$

$$= \int_{t_1}^{t_2} \Bigg(\eta(t)G_0^f(f(t), r_1(u), \ldots, r_n(u), u, t) + \sum_{i=1}^{n} \frac{dr_i^\alpha(u)}{d\alpha}G_0^{r_i}(f(t), r_1(u), \ldots, r_n(u), u, t)\Bigg) du, \tag{49}$$

where

$$\frac{dr_i^\alpha(u)}{d\alpha} \triangleq \frac{d}{d\alpha}\int_{t_1}^{t_2} G_i(f(v) + \alpha\eta(v), v, u)dv\Big|_{\alpha=0} \tag{50}$$

$$= \int_{t_1}^{t_2} \eta(v)G_i^f(f(v), v, u)dv. \tag{51}$$

By plugging (51) into (49), we can write

$$\frac{dH^\alpha(t)}{d\alpha} = \int_{t_1}^{t_2} \eta(t)G_0^f(f(t), r_1(u), \ldots, r_n(u), u, t)du, + \int_{t_1}^{t_2}\int_{t_1}^{t_2}\sum_{i=1}^{n}\Bigg(\eta(v)G_i^f(f(v), v, u) \cdot G_0^{r_i}(f(t), r_1(u), \ldots, r_n(u), u, t)\Bigg)dvdu. \tag{52}$$

By substituting (52) into (47) we have

$$\delta_f L = \int_{t_1}^{t_2} \eta(t)F^f(f(t), H(t), t)dt + \int_{t_1}^{t_2} F^H(f(t), H(t), t) \cdot \Bigg(\int_{t_1}^{t_2} \eta(t)G_0^f(f(t), r_1(u), \ldots, r_n(u), u, t)du + \int_{t_1}^{t_2}\int_{t_1}^{t_2}\sum_{i=1}^{n}\eta(v)G_i^f(f(v), v, u) \cdot G_0^{r_i}(f(t), r_1(u), \ldots, r_n(u), u, t)dvdu\Bigg)dt \tag{53}$$

$$= \int_{t_1}^{t_2} \eta(t)\Bigg(F^f(f(t), H(t), t) + F^H(f(t), H(t), t) \cdot \int_{t_1}^{t_2} G_0^f(f(t), r_1(u), \ldots, r_n(u), u, t)du\Bigg)dt + \int_{t_1}^{t_2}\int_{t_1}^{t_2}\int_{t_1}^{t_2}\eta(v)F^H(f(t), H(t), t)\sum_{i=1}^{n}\Bigg(G_i^f(f(v), v, u) \cdot G_0^{r_i}(f(t), r_1(u), \ldots, r_n(u), u, t)\Bigg)dvdudt \tag{54}$$

$$= \int_{t_1}^{t_2} \eta(t)\Bigg(F^f(f(t), H(t), t) + F^H(f(t), H(t), t) \cdot \int_{t_1}^{t_2} G_0^f(f(t), r_1(u), \ldots, r_n(u), u, t)du\Bigg)dt + \int_{t_1}^{t_2}\int_{t_1}^{t_2}\int_{t_1}^{t_2}\eta(t)F^H(f(v), H(v), v)\sum_{i=1}^{n}\Bigg(G_i^f(f(t), t, u) \cdot G_0^{r_i}(f(v), r_1(u), \ldots, r_n(u), u, v)\Bigg)dvdudt \tag{55}$$

$$= \int_{t_1}^{t_2} \eta(t)\Bigg(F^f(f(t), H(t), t) + F^H(f(t), H(t), t) \cdot \int_{t_1}^{t_2} G_0^f(f(t), r_1(u), \ldots, r_n(u), u, t)du + \int_{t_1}^{t_2}\int_{t_1}^{t_2} F^H(f(v), H(v), v)\sum_{i=1}^{n}\Bigg(G_i^f(f(t), t, u) \cdot G_0^{r_i}(f(v), r_1(u), \ldots, r_n(u), u, v)\Bigg)dvdu\Bigg)dt. \tag{56}$$

Since $\eta(t)$ in (56) is an arbitrary function, the necessary condition for f to be a solution is that the term inside the round brackets in (56) is zero. This concludes the proof.

□

Remark A.1. *In Proposition IV.1, we will consider the minimization of the functional*

$$L(f) = \frac{1}{\sigma_v}\int_{t_1}^{t_2} \Phi\left(\frac{t}{\sigma_v}\right)\tilde{F}(t, f(t))\,dt. \tag{57}$$

By writing the Euler-Lagrange equation from [15, Section 7.5, Equation (2)], it can be easily verified that the solution of (57) needs to satisfy

$$\nabla L \triangleq \tilde{F}^f\left(u, f(u)\right) = 0,\ u \in [t_1, t_2], \tag{58}$$

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

B. Proof of Proposition III.1

Note that under the MSE distortion criterion the optimal decoder is MMSE, which for this system model is obtained as follows

$$\begin{aligned}
\hat{v} &= g(y,u) && (59)\\
&= \mathbb{E}[V|Y=y, U=u] && (60)\\
&= \frac{\int v\Phi\left(\frac{v}{\sigma_v}, \frac{u}{\sigma_u}\right)\Pr(Y=y|U=u,V=v)dv}{\Pr(Y=y,U=u)} && (61)\\
&= \frac{\int v\Phi\left(\frac{v}{\sigma_v}\middle|\frac{u}{\sigma_u}\right)Q\left(\frac{(-1)^{y+1}f(v)}{\sigma_w}\right)dv}{\int \Phi\left(\frac{v}{\sigma_v}\middle|\frac{u}{\sigma_u}\right)Q\left(\frac{(-1)^{y+1}f(v)}{\sigma_w}\right)dv} && (62)\\
&\triangleq \frac{r_{2y+1}}{r_{2y+2}}. && (63)
\end{aligned}$$

Due to the orthogonality principle of the MMSE estimation, it can be easily verified that $\bar{D} = \sigma_v^2 - \mathbb{E}[V\hat{V}]$. Rewriting the Lagrangian $L(f,g,\lambda)$ for the MSE distortion criterion and dropping constants that are independent of f, we have

$$\underset{f}{\text{minimize}}\ -\mathbb{E}[V\hat{V}] + \lambda\mathbb{E}[f(V)^2]. \tag{64}$$

By expanding the objective function in (64), it can be written as

$$\begin{aligned}
&\frac{-1}{\sigma_w\sigma_v\sigma_u}\int\int\int vg(y,u)\Phi\left(\frac{v}{\sigma_v},\frac{u}{\sigma_u}\right)\Phi\left(\frac{w}{\sigma_w}\right)dwdudv\\
&\quad+\frac{\lambda}{\sigma_v}\int\Phi\left(\frac{v}{\sigma_v}\right)f(v)^2dv && (65)\\
&=\frac{-1}{\sigma_v\sigma_u}\int\int v\left(g(1,u)Q\left(\frac{f(v)}{\sigma_w}\right)+g(0,u)Q\left(\frac{-f(v)}{\sigma_w}\right)\right)\\
&\quad\cdot\Phi\left(\frac{v}{\sigma_v},\frac{u}{\sigma_u}\right)dudv+\frac{\lambda}{\sigma_v}\int\Phi\left(\frac{v}{\sigma_v}\right)f(v)^2dv && (66)\\
&=\frac{-1}{\sigma_v}\int\left(v\int\frac{1}{\sigma_u}\Phi\left(\frac{v}{\sigma_v},\frac{u}{\sigma_u}\right)\left(\frac{r_3(u)}{r_4(u)}Q\left(\frac{f(v)}{\sigma_w}\right)\right.\right.\\
&\quad\left.\left.+\frac{r_1(u)}{r_2(u)}Q\left(\frac{-f(v)}{\sigma_w}\right)\right)du+\lambda\Phi\left(\frac{v}{\sigma_v}\right)f(v)^2\right)dv, && (67)
\end{aligned}$$

Note that (67) is in the form of (41) with $F(f,H(v),v)$ and $H(v)$ defined as

$$F(f,H(v),v)=\frac{1}{\sigma_v}\left(-vH(v)+\lambda\Phi\left(\frac{v}{\sigma_v}\right)f(v)^2\right), \tag{68}$$

and

$$H(v)=\int G_0\left(f(v),r_1(u),\ldots,r_4(u),u,v\right)du, \tag{69}$$

where $G_0\left(f(v),r_1(u),\ldots,r_4(u),u,v\right)$ is given by

$$\begin{aligned}
G_0\left(f(v),r_1(u),\ldots,r_4(u),u,v\right)&=\frac{1}{\sigma_u}\Phi\left(\frac{v}{\sigma_v},\frac{u}{\sigma_u}\right)\\
&\cdot\left(\frac{r_3(u)}{r_4(u)}Q\left(\frac{f(v)}{\sigma_w}\right)+\frac{r_1(u)}{r_2(u)}Q\left(\frac{-f(v)}{\sigma_w}\right)\right), && (70)
\end{aligned}$$

and $G_i, i=1,...,4$, are the arguments of the integrals r_{2y+1} and r_{2y+2} for $y=0,1$ defined in (63). Now we can apply the necessary condition in (42). To this end, we compute

$$\begin{aligned}
&F^f(f(v),H(v),v)=\frac{2\lambda}{\sigma_v}\Phi\left(\frac{v}{\sigma_v}\right)f(v), && (71a)\\
&F^H(f(v),H(v),v)=\frac{-v}{\sigma_v}, && (71b)\\
&G_0^f(f(v),r_1(u),\ldots,r_4(u),u,v)=\\
&\quad\frac{e^{-\frac{f(v)^2}{2\sigma_w^2}}}{\sigma_w\sigma_u\sqrt{2\pi}}\Phi\left(\frac{v}{\sigma_v},\frac{u}{\sigma_u}\right)\left(\frac{r_1(u)}{r_2(u)}-\frac{r_3(u)}{r_4(u)}\right), && (71c)\\
&G_0^{r_{2y+1}}(f(v),r_1(u),\ldots,r_4(u),u,v)=\\
&\quad\frac{1}{\sigma_u r_{2y+2}(u)}\Phi\left(\frac{v}{\sigma_v},\frac{u}{\sigma_u}\right)Q\left(\frac{(-1)^{y+1}f(v)}{\sigma_w}\right), && (71d)\\
&G_0^{r_{2y+2}}(f(v),r_1(u),\ldots,r_4(u),u,v)=\\
&\quad-r_{2y+1}(u)G_0^{r_{2y+1}}(f(v),r_1(u),\ldots,r_4(u),u,v), && (71e)\\
&G_{2y+2}^f(f(v),v,u)=(-1)^y\Phi\left(\frac{v}{\sigma_v},\frac{u}{\sigma_u}\right)\frac{e^{-\frac{f(v)^2}{2\sigma_w^2}}}{\sqrt{2\pi}\sigma_w}, && (71f)\\
&G_{2y+1}^f(f(v),v,u)=v(-1)^yG_{2y+2}^f(f(v),v,u). && (71g)
\end{aligned}$$

Substituting (71) in (42), the necessary condition in (41) is obtained as

$$\begin{aligned}
\nabla L=&\frac{2\lambda}{\sigma_v}\Phi\left(\frac{v}{\sigma_v}\right)f(v)-\frac{ve^{-\frac{f(v)^2}{2\sigma_w^2}}}{\sigma_v\sigma_w\sigma_u\sqrt{2\pi}}\\
&\cdot\int\Phi\left(\frac{v}{\sigma_v},\frac{u}{\sigma_u}\right)\left(\frac{r_1(u)}{r_2(u)}-\frac{r_3(u)}{r_4(u)}\right)du\\
&-\int\int\frac{t}{\sigma_v}\left(v\Phi\left(\frac{v}{\sigma_v},\frac{u}{\sigma_u}\right)\frac{-e^{-\frac{f(v)^2}{2\sigma_w^2}}}{\sqrt{2\pi}\sigma_w}\right.\\
&\cdot\frac{1}{\sigma_u r_4(u)}\Phi\left(\frac{t}{\sigma_v},\frac{u}{\sigma_u}\right)Q\left(\frac{f(t)}{\sigma_w}\right)\\
&+\Phi\left(\frac{v}{\sigma_v},\frac{u}{\sigma_u}\right)\frac{-e^{-\frac{f(v)^2}{2\sigma_w^2}}}{\sqrt{2\pi}\sigma_w}\cdot\frac{-r_3(u)}{\sigma_u r_4(u)^2}\Phi\left(\frac{t}{\sigma_v},\frac{u}{\sigma_u}\right)Q\left(\frac{f(t)}{\sigma_w}\right)\\
&+v\Phi\left(\frac{v}{\sigma_v},\frac{u}{\sigma_u}\right)\frac{e^{-\frac{f(v)^2}{2\sigma_w^2}}}{\sqrt{2\pi}\sigma_w}\cdot\frac{1}{\sigma_u r_2}\Phi\left(\frac{t}{\sigma_v},\frac{u}{\sigma_u}\right)Q\left(\frac{-f(t)}{\sigma_w}\right)\\
&+\Phi\left(\frac{v}{\sigma_v},\frac{u}{\sigma_u}\right)\frac{e^{-\frac{f(v)^2}{2\sigma_w^2}}}{\sqrt{2\pi}\sigma_w}\cdot\frac{-r_1(u)}{\sigma_u r_2(u)^2}\\
&\left.\cdot\Phi\left(\frac{t}{\sigma_v},\frac{u}{\sigma_u}\right)Q\left(\frac{-f(t)}{\sigma_w}\right)\right)dtdu=0. && (72)
\end{aligned}$$

Rewriting (72), we have

$$
\begin{aligned}
&2\sqrt{2\pi}\sigma_w\sigma_u\lambda\Phi\left(\frac{v}{\sigma_v}\right)f(v)e^{\frac{f(v)^2}{2\sigma_w^2}} = \\
&v\int\Phi\left(\frac{v}{\sigma_v},\frac{u}{\sigma_u}\right)\left(\frac{r_1(u)}{r_2(u)}-\frac{r_3(u)}{r_4(u)}\right)du \\
&-\int\int\Phi\left(\frac{v}{\sigma_v},\frac{u}{\sigma_u}\right)\cdot\frac{vt}{r_4(u)}\Phi\left(\frac{t}{\sigma_v},\frac{u}{\sigma_u}\right)Q\left(\frac{f(t)}{\sigma_w}\right)dtdu \\
&+\int\int\Phi\left(\frac{v}{\sigma_v},\frac{u}{\sigma_u}\right)\cdot\frac{tr_3(u)}{r_4(u)^2}\Phi\left(\frac{t}{\sigma_v},\frac{u}{\sigma_u}\right)Q\left(\frac{f(t)}{\sigma_w}\right)dtdu \\
&+\int\int\Phi\left(\frac{v}{\sigma_v},\frac{u}{\sigma_u}\right)\cdot\frac{vt}{r_2(u)}\Phi\left(\frac{t}{\sigma_v},\frac{u}{\sigma_u}\right)Q\left(\frac{-f(t)}{\sigma_w}\right)dtdu \\
&-\int\int\Phi\left(\frac{v}{\sigma_v},\frac{u}{\sigma_u}\right)\cdot\frac{tr_1(u)}{r_2(u)^2}\Phi\left(\frac{t}{\sigma_v},\frac{u}{\sigma_u}\right)Q\left(\frac{-f(t)}{\sigma_w}\right)dtdu \quad (73)\\
&=v\int\Phi\left(\frac{v}{\sigma_v},\frac{u}{\sigma_u}\right)\left(\frac{r_1(u)}{r_2(u)}-\frac{r_3(u)}{r_4(u)}\right)du \\
&-v\int\Phi\left(\frac{v}{\sigma_v},\frac{u}{\sigma_u}\right)\cdot\frac{r_3(u)}{r_4(u)}du \\
&+\int\Phi\left(\frac{v}{\sigma_v},\frac{u}{\sigma_u}\right)\cdot\frac{r_3(u)^2}{r_4(u)^2}du \\
&+v\int\Phi\left(\frac{v}{\sigma_v},\frac{u}{\sigma_u}\right)\cdot\frac{r_1(u)}{r_2(u)}du \\
&-\int\Phi\left(\frac{v}{\sigma_v},\frac{u}{\sigma_u}\right)\cdot\frac{r_1(u)^2}{r_2(u)^2}du. \quad (74)
\end{aligned}
$$

Finally, by some elementary manipulations the result in (11) is obtained.

□

C. Proof of Proposition IV.1

Assume that a decoder function $g(Y,U)$ is given. By expanding the Lagrangian function $L(f,g,\lambda)$ for the DOP we have

$$
\begin{aligned}
L(f,g,\lambda) &= \epsilon(D)+\lambda\mathbb{E}[f(V)^2] \quad (75)\\
&= \frac{1}{\sigma_u}\int\epsilon(D|U=u)\Phi\left(\frac{u}{\sigma_u}\right)du \\
&\quad+\frac{\lambda}{\sigma_v}\int\Phi\left(\frac{v}{\sigma_v}\right)f(v)^2dv. \quad (76)
\end{aligned}
$$

Expanding $\epsilon(D|U=u)=\Pr\left((V-\hat{V})^2\geq D|U=u\right)$ we have

$$
\begin{aligned}
\epsilon(D|U=u) &= \Pr(V\in I_0(u)\setminus I_1(u), Y=1|U=u) \\
&\quad+\Pr(V\in I_1(u)\setminus I_0(u), Y=0|U=u) \\
&\quad+\Pr(V\in(I_0(u)\cup I_1(u))^C, |\hat{V}-V|^2\geq D|U=u) \\
&\quad+\Pr(V\in(I_0(u)\cap I_1(u)), |\hat{V}-V|^2\geq D|U=u) \quad (77a)\\
&= \frac{1}{\sigma_v}\int\limits_{v\in I_0(u)\setminus I_1(u)}\Phi\left(\frac{v}{\sigma_v}\Big|\frac{u}{\sigma_u}\right)Q\left(\frac{f(v)}{\sigma_w}\right)dv \\
&\quad+\frac{1}{\sigma_v}\int\limits_{v\in I_1(u)\setminus I_0(u)}\Phi\left(\frac{v}{\sigma_v}\Big|\frac{u}{\sigma_u}\right)Q\left(\frac{-f(v)}{\sigma_w}\right)dv \\
&\quad+\frac{1}{\sigma_v}\int\limits_{v\in(I_0(u)\cup I_1(u))^C}\Phi\left(\frac{v}{\sigma_v}\Big|\frac{u}{\sigma_u}\right)dv, \quad (77b)
\end{aligned}
$$

where we used the fact that no outage occurs when $V\in I_0(U)\cap I_1(U)$. Substituting (77b) in (76), we can write the Lagrangian $L(f,g,\lambda)$ as

$$
\begin{aligned}
&L(f,g,\lambda) = \\
&\frac{1}{\sigma_v\sigma_u}\int\Phi\left(\frac{u}{\sigma_u}\right)\int\Phi\left(\frac{v}{\sigma_v}\Big|\frac{u}{\sigma_u}\right)G(u,v,f(v))\,dvdu \\
&\quad+\frac{\lambda}{\sigma_v}\int\Phi\left(\frac{v}{\sigma_v}\right)\lambda f^2(v)dv \quad (78)\\
&= \frac{1}{\sigma_v}\int\Phi\left(\frac{v}{\sigma_v}\right) \\
&\quad\left[\int\frac{1}{\sigma_u}\Phi\left(\frac{u}{\sigma_u}\Big|\frac{v}{\sigma_v}\right)G(u,v,f(v))\,du+\lambda f^2(v)\right]dv, \quad (79)
\end{aligned}
$$

with $G(u,v,f(v))$ defined as

$$
G(u,v,f(v))\triangleq\begin{cases}Q\left(\frac{f(v)}{\sigma_w}\right) & v\in(I_0(u)\setminus I_1(u)),\\ Q\left(\frac{-f(v)}{\sigma_w}\right) & v\in(I_1(u)\setminus I_0(u)),\\ 1 & v\in(I_0(u)\cup I_1(u))^C,\\ 0 & v\in(I_0(u)\cap I_1(u)).\end{cases} \quad (80)
$$

Applying the Euler-Lagrange equation (57), we have

$$
\nabla L = 2\lambda f(v)+\frac{1}{\sigma_u}\int\Phi\left(\frac{u}{\sigma_u}\Big|\frac{v}{\sigma_v}\right)G^f(u,v,f(v))\,du=0, \quad (81)
$$

where $G^f(u,v,f(v))$ is the derivative of $G(u,v,f(v))$ with respect to the function f and is obtained as

$$
G^f(u,v,f(v)) = \begin{cases}\frac{-1}{\sqrt{2\pi}}e^{-\frac{f(v)^2}{2\sigma_w^2}} & v\in(I_0(u)\setminus I_1(u)),\\ \frac{1}{\sqrt{2\pi}}e^{-\frac{f(v)^2}{2\sigma_w^2}} & v\in(I_1(u)\setminus I_0(u)),\\ 0 & v\in(I_0(u)\cap I_1(u)) \text{ or } v\in(I_0(u)\cup I_1(u))^C.\end{cases} \quad (82)
$$

Note that the integration in (81) is over the side information u. In the following, we aim at identifying the boundaries of the intervals of u, such that, for a given source output v we have $G^f(u,v,f(v))\neq 0$. To do so, we characterize the intervals as

$$
\begin{aligned}
I_0(u)\setminus I_1(u) &= (b_{0l}(u), b_{0r}(u)),\\
I_1(u)\setminus I_0(u) &= (b_{1l}(u), b_{1r}(u)). \quad (83)
\end{aligned}
$$

Note that if $v\in I_0(u)\cap I_1(u)$ and $v\in(I_0(u)\cup I_1(u))^C$, we have $G^f(u,v,f(v))=0$. For a given side information realization u, $g(0,u)$ and $g(1,u)$ are two points. Hence, depending on the condition that $g(0,u)$ is equal to, less than, or greater than $g(1,u)$, we have different situations for $I_0(u)$ and $I_1(u)$ in (82).

Case 1) $g(0,u)=g(1,u)$: In this case the two intervals $I_0(u)$ and $I_1(u)$ overlap completely, and therefore, $I_0(u)\setminus I_1(u)$ and $I_1(u)\setminus I_0(u)$ are both empty sets.

Case 2) $g(0,u)>g(1,u)$: In this case $b_{0l}(u)$, $b_{0r}(u)$, $b_{1r}(u)$ and $b_{1l}(u)$ are obtained as

$$
\begin{aligned}
b_{0r}(u) &= g(0,u)+\sqrt{D},\\
b_{0l}(u) &= \max\left\{g(1,u)+\sqrt{D}, g(0,u)-\sqrt{D}\right\},
\end{aligned}
$$

$$b_{1r}(u) = \min\left\{g(1,u)+\sqrt{D}, g(0,u)-\sqrt{D}\right\},$$
$$b_{1l}(u) = g(1,u)-\sqrt{D}. \tag{84}$$

Case 3) $g(0,u) < g(1,u)$: In this case $b_{0l}(u)$, $b_{0r}(u)$, $b_{1r}(u)$ and $b_{1l}(u)$ are obtained as

$$b_{0r}(u) = \min\left\{g(0,u)+\sqrt{D}, g(1,u)-\sqrt{D}\right\},$$
$$b_{0l}(u) = g(0,u)-\sqrt{D},$$
$$b_{1r}(u) = g(1,u)+\sqrt{D},$$
$$b_{1l}(u) = \max\left\{g(1,u)-\sqrt{D}, g(0,u)+\sqrt{D}\right\}. \tag{85}$$

It can be easily verified that for a given source output v, the side information range corresponding to $G^f(u,v,f(v)) \neq 0$ can be obtained as $S_{0\setminus 1}(v) \cup S_{1\setminus 0}(v)$, where $S_{0\setminus 1}(v)$ and $S_{1\setminus 0}(v)$ are defined as

$$S_{0\setminus 1}(v) \triangleq \{u:\ b_{0r}(u) \geq v \geq b_{0l}(u)\},$$
$$S_{1\setminus 0}(v) \triangleq \{u:\ b_{1r}(u) \geq v \geq b_{1l}(u)\}. \tag{86}$$

Finally, we can simplify (81) as

$$\nabla L = \frac{-1}{\sqrt{2\pi}\sigma_u} e^{-\frac{f(v)^2}{2\sigma_w^2}} \int\limits_{u\in S_{0\setminus 1}(v)} \Phi\left(\frac{u}{\sigma_u}\middle|\frac{v}{\sigma_v}\right) du + \frac{1}{\sqrt{2\pi}\sigma_u} e^{-\frac{f(v)^2}{2\sigma_w^2}} \int\limits_{u\in S_{1\setminus 0}(v)} \Phi\left(\frac{u}{\sigma_u}\middle|\frac{v}{\sigma_v}\right) du + 2\lambda f(v). \tag{87}$$

Imposing (87) to be zero we have

$$f(v) = \frac{e^{-\frac{f(v)^2}{2\sigma_w^2}}}{2\lambda\sqrt{2\pi}} \left(\Pr\left(U \in S_{0\setminus 1}(v)\right) - \Pr\left(U \in S_{1\setminus 0}(v)\right)\right). \tag{88}$$

□

D. Proof of Proposition IV.2

The optimal decoder functions, i.e., $g(0,u)$ and $g(1,u)$ can be obtained as

$$g(0,u) = \arg\min_{\hat{v}} \ \Pr\left(|V-\hat{v}|^2 \geq D | U=u, Y=0\right) \tag{89}$$
$$= \arg\max_{\hat{v}} \ \Pr\left(|V-\hat{v}|^2 < D | U=u, Y=0\right) \tag{90}$$
$$= \arg\max_{\hat{v}} \ \frac{1}{\sigma_v\sigma_u} \int\limits_{\hat{v}-\sqrt{D}}^{\hat{v}+\sqrt{D}} p_{V|U,Y}(t|u,Y=0)\, dt \tag{91}$$
$$= \arg\max_{\hat{v}} \int\limits_{\hat{v}-\sqrt{D}}^{\hat{v}+\sqrt{D}} \Phi\left(\frac{t}{\sigma_v}\middle|\frac{u}{\sigma_u}\right) Q\left(\frac{-f(t)}{\sigma_w}\right) dt. \tag{92}$$

We note that, since the mapping $f(v)$ is given, it could be possible that for some encoder mapping f and side information realization u, more than one output is obtained in (92). From the DOP point of view, there is no difference in choosing either of these points. Therefore, we have

$$g^*(0,u) \in \arg\max_{\hat{v}} \int\limits_{\hat{v}-\sqrt{D}}^{\hat{v}+\sqrt{D}} \Phi\left(\frac{t}{\sigma_v}, \frac{u}{\sigma_u}\right) Q\left(\frac{-f(t)}{\sigma_w}\right) dt. \tag{93}$$

$g(1,u)$ is derived similarly to $g(0,u)$.

□

E. Proof of (32) in Remark IV.1

In the low SNR regime (large values of σ_w^2), it can be verified from (27) that the encoder mapping tends to an all-zero function. Hence, the DOP can be evaluated as

$$\epsilon(D) = 1 - \Pr(|V-\hat{V}|^2 < D) \tag{94a}$$
$$= 1 - \frac{1}{\sigma_u}\int \Pr(|V-\hat{V}|^2 < D | U=u)\Phi\left(\frac{u}{\sigma_u}\right) du \tag{94b}$$
$$= 1 - \frac{1}{\sigma_u}\int \Phi\left(\frac{u}{\sigma_u}\right) \Pr\left(\left|V - \frac{r\sigma_v}{\sigma_u}u\right|^2 < D | U=u\right) du \tag{94c}$$
$$= 1 - \frac{1}{\sigma_u\sigma_v}\int \Phi\left(\frac{u}{\sigma_u}\right) \int\limits_{\frac{r\sigma_v}{\sigma_u}u-\sqrt{D}}^{\frac{r\sigma_v}{\sigma_u}u+\sqrt{D}} \Phi\left(\frac{v}{\sigma_v}\middle|\frac{u}{\sigma_u}\right) du \tag{94d}$$
$$= 1 - \frac{1}{\sigma_u}\int \Phi\left(\frac{u}{\sigma_u}\right) \cdot \left(Q\left(\frac{-\sqrt{D}}{\sigma_v\sqrt{1-r^2}}\right) - Q\left(\frac{\sqrt{D}}{\sigma_v\sqrt{1-r^2}}\right)\right) du \tag{94e}$$
$$= 2Q\left(\frac{\sqrt{D}}{\sigma_v\sqrt{1-r^2}}\right). \tag{94f}$$

□

F. Proof of Proposition IV.3

We first define as $f(\cdot)$ the encoder mapping applied to the error in (22). Assuming that the side information is available at both the encoder and the decoder, the optimal decoder can be obtained as

$$g(y,u) = \arg\min_{\hat{v}} \ \Pr\left(|V-\hat{v}|^2 \geq D | Y=y, U=u\right) + \lambda\mathbb{E}[\tilde{f}(T)^2] \tag{95a}$$
$$= \arg\min_{\hat{v}} \int\limits_{\hat{v}-\sqrt{D}}^{\hat{v}+\sqrt{D}} \Phi\left(\frac{v}{\sigma_v}\middle|\frac{u}{\sigma_u}\right) \cdot Q\left(\frac{(-1)^{y+1}\tilde{f}\left(v - \frac{r\sigma_v u}{\sigma_u}\right)}{\sigma_w}\right) dv - \lambda\mathbb{E}[\tilde{f}(T)^2] \tag{95b}$$
$$= \arg\min_{\hat{v}} \int\limits_{\hat{v}-\sqrt{D}}^{\hat{v}+\sqrt{D}} e^{-\frac{\left(v-\frac{r\sigma_v u}{\sigma_u}\right)^2}{2\sigma_v^2(1-r^2)}} \cdot Q\left(\frac{(-1)^{y+1}\tilde{f}\left(v - \frac{r\sigma_v u}{\sigma_u}\right)}{\sigma_w}\right) dv - \lambda\mathbb{E}[\tilde{f}(T)^2]$$

$$= \arg\min_{\hat{v}} \int_{\hat{v}-\sqrt{D}-\frac{r\sigma_v u}{\sigma_u}}^{\hat{v}+\sqrt{D}-\frac{r\sigma_v u}{\sigma_u}} e^{-\frac{t^2}{2\sigma_{\tilde{V}}^2(1-r^2)}} \cdot Q\left(\frac{(-1)^{y+1}\tilde{f}(t)}{\sigma_w}\right) dt - \lambda\mathbb{E}[\tilde{f}(T)^2] \tag{95c}$$

$$= \frac{r\sigma_v u}{\sigma_u} + \arg\min_{\hat{t}} \int_{\hat{t}-\sqrt{D}}^{\hat{t}+\sqrt{D}} e^{-\frac{t^2}{2\sigma_{\tilde{V}}^2(1-r^2)}} \cdot Q\left(\frac{(-1)^{y+1}\tilde{f}(t)}{\sigma_w}\right) dt - \lambda\mathbb{E}[\tilde{f}(T)^2] \tag{95d}$$

$$= \frac{r\sigma_v u}{\sigma_u} + \hat{t}_Y, \tag{95e}$$

where in (95c) we used the transformation $v - \frac{r\sigma_v u}{\sigma_u} = t$; in (95d) we replaced $\hat{v} - r\sigma_v u/\sigma_u$ with $\hat{t}$ by adding $r\sigma_v u/\sigma_u$ to the resultant argument. Finally, the second term in (95e) represent the optimal decoder when there is no side information as derived in [9, Proposition IV.2].

The DOP in (8) can be evaluated as

$$\epsilon(D) = \frac{1}{\sigma_v}\int \Phi\left(\frac{u}{\sigma_u}\right) \Pr\left(|V-\hat{V}|^2 \geq D | U=u\right) du \tag{96a}$$

$$= \frac{1}{\sigma_v}\int \Phi\left(\frac{u}{\sigma_u}\right) \Pr\left(\left|V - \frac{r\sigma_v u}{\sigma_u} - \hat{t}_Y\right|^2 \geq D \middle| U=u\right) du \tag{96b}$$

$$= \frac{1}{\sigma_v}\int \Phi\left(\frac{u}{\sigma_u}\right) \Bigg[\Pr\left(V \in (I_0(u)\cup I_1(u))^C\right) + \Pr\left(V \in I_0(u)\setminus I_1(u), \hat{t}_Y = t_1\right) + \Pr\left(V \in I_1(u)\setminus I_0(u), \hat{t}_Y = t_0\right)\Bigg] du \tag{96c}$$

$$= \frac{1}{\sigma_u\sigma_v}\int \Phi\left(\frac{u}{\sigma_u}\right) \Bigg[\int_{(I_0(u)\cup I_1(u))^C} \Phi\left(\frac{v}{\sigma_v}\middle|\frac{u}{\sigma_u}\right) dv + Q\left(\frac{t}{\sigma_w}\right)\Bigg(\int_{I_1(u)\setminus I_0(u)} \Phi\left(\frac{v}{\sigma_v}\middle|\frac{u}{\sigma_u}\right) dv + \int_{I_0(u)\setminus I_1(u)} \Phi\left(\frac{v}{\sigma_v}\middle|\frac{u}{\sigma_u}\right) dv\Bigg)\Bigg], \tag{96d}$$

where we have defined

$$I_y(u) \triangleq \left\{v : \left(v - \frac{r\sigma_v u}{\sigma_u} - \hat{t}_y\right)^2 \leq D\right\}, \; y = 0, 1. \tag{97}$$

□

References

[1] 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, 2016, pp. 449–453.

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

[3] C. Risi, D. Persson, and E. G. Larsson, "Massive MIMO with 1-bit ADC," *CoRR*, vol. abs/1404.7736, 2014. [Online]. Available: http://arxiv.org/abs/1404.7736

[4] D. Gunduz, K. Stamatiou, N. Michelusi, and M. Zorzi, "Designing intelligent energy harvesting communication systems," *IEEE communications magazine*, vol. 52, pp. 210–216, 2014.

[5] 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.

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

[7] 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.

[8] S. Jacobsson, G. Durisi, M. Coldrey, U. Gustavsson, and C. Studer, "One-bit massive MIMO: Channel estimation and high-order modulations," *CoRR*, vol. abs/1504.04540, 2015. [Online]. Available: http://arxiv.org/abs/1504.04540

[9] M. Varasteh, B. Rassouli, O. Simeone, and D. Gündüz, "Joint source-channel coding with one-bit ADC front end," *CoRR*, vol. abs/1604.06578, 2016. [Online]. Available: http://arxiv.org/abs/1604.06578

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

[11] M. S. Mehmetoglu, E. Akyol, and K. Rose, "Deterministic annealing-based optimization for zero-delay source-channel coding in networks," *IEEE Trans. Commun.*, vol. 63, no. 12, pp. 5089–5100, Dec. 2015.

[12] X. Chen and E. Tuncel, "Zero-delay joint source-channel coding using hybrid digital-analog schemes in the wyner-ziv setting," *IEEE Trans. Commun.*, vol. 62, no. 2, pp. 726–735, Feb. 2014.

[13] D. P. Bertsekas, *Nonlinear Programming*. Athena Scientific, 1999.

[14] A. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE Trans. Information Theory*, vol. 22, no. 1, pp. 1–10, January 1976.

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

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, and leading the Information Processing and Communications Lab. His research interests lie in the areas of communications and information theory, machine learning, and security 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 is the General Co-chair of the 2018 Workshop on Smart Antennas, and previously 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.