Joint Communication and Binary State Detection

Hamdi Joudeh and Frans M. J. Willems

Abstract

In this paper we study basic joint communication and sensing settings from an information theoretic perspective. In the considered settings, a transmitter sends an encoded message through a pair of noisy channels connected to a receiver and a sensor. The receiver is interested in decoding the message from its noisy observation. The sensor has access to the message as side information, and instead is interested in estimating an unknown yet fixed binary state of the channel. This basic setup models scenarios where a wireless transceiver broadcasts an information bearing waveform, to be decoded by possibly more than one receiver, and then observes a modulated echo of the waveform from which it wishes to detect an element or object in the environment. We consider binary symmetric and Gaussian settings with multiplicative binary states, and establish the fundamental trade-off between reliable message communication and efficient state detection in these settings. This trade-off is captured by the message communication rate against the state detection error exponent. The results give insights into the benefits of carrying out communication and sensing jointly.

The authors are with the Department of Electrical Engineering, Eindhoven University of Technology, 5600 MB Eindhoven, The Netherlands (e-mail: h.joudeh@tue.nl; f.m.j.willems@tue.nl). This work was supported in part by the Netherlands Organisation for Scientific Research (NWO) under the project Integrated Cooperative Automated Vehicles (i-CAVE).

1 Introduction

We consider a setup in which a transmitter with a random message is connected to a receiver and a sensor through a pair of noisy channels. The transmitter encodes its message and sends the resulting codeword over the two channels. The receiver is interested in decoding the message from a noisy observation of the codeword. The sensor has access to the message (and hence the codeword) as side information, and instead wishes to estimate an unknown yet fixed state (or parameter) of its channel from a state-modulated noisy observation of the codeword. An illustration of this setup, with a multiplicative state and additive noise, is shown in Fig. 1. Our goal is to understand the fundamental performance trade-off between reliable message communication and efficient state estimation or detection.

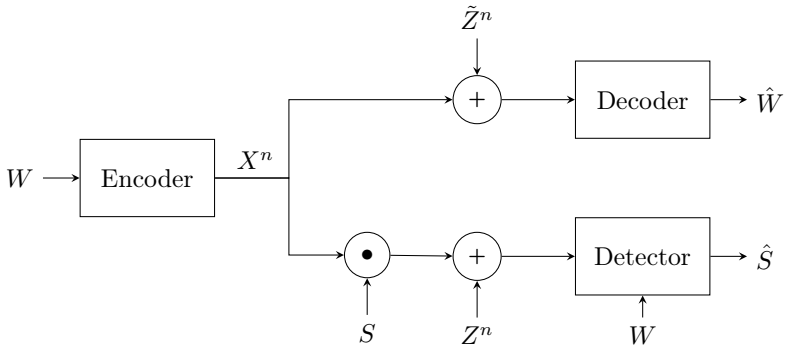

Figure 1: A joint communication and state detection setup with a transmitter (encoder), a receiver (decoder), and a sensor (detector).

First, let us demonstrate that there is indeed a performance trade-off between the two objectives. Consider an instance of the setup in Fig. 1, where the state S and all sequences X^n, $\tilde{Z}^n$ and Z^n are binary, and the sum is modulo-2. Each of the noise sequences $\tilde{Z}^n$ and Z^n is assumed to have independent and identically distributed (i.i.d.) entries. Given that a codeword $X^n = x^n$ is transmitted, the problem of state detection at the sensor is essentially a binary hypothesis test of $0 \cdot x^n$ against $1 \cdot x^n$, see, e.g., [1]. The worst-case error (i.e. over $S \in \{0, 1\}$) of such test is minimized by selecting x^n to be a codeword of all ones, as zero entries are useless for distinguishing between $0 \cdot x^n$ and $1 \cdot x^n$. Such transmission of all ones carries no information to the receiver. Communication at positive rates requires injecting zeros into the transmission, which in turn leads to degradation in state detection. The fundamental trade-off in such setting, as n grows large, is established in Theorem 1 of this paper (see Section 3).

Motivation and related work: The setup in Fig. 1 is motivated by problems of joint (or integrated) communication and sensing, which have received increased interest of late, see, e.g., [2–5] and references therein. For instance, the setup in Fig. 1 with an on-off multiplicative state is a basic model for scenarios where a vehicle sends a single waveform with the dual purpose of communicating information to other vehicles and detecting the presence of an obstacle from the waveform echo, i.e. joint communication and radar sensing. In this case, the transmitter (encoder) and the sensor (detector) are physically co-located, which justifies the assumption that the detector has access to the message (and transmitted waveform) as side information. One may also think of a related scenario where a satellite uses a single waveform to perform active remote sensing while communicating with a ground station.

An information theoretic formulation for the joint communication and sensing problem has been proposed in [6], where the authors consider a setting in which a transmitter communicates a message to a receiver over a state-dependent stationary memoryless channel, and at the same time estimates the state of the channel through means of generalized feedback (i.e. a strictly causal channel output observed by the transmitter). The trade-off between message communication and state estimation is characterized

in terms of a capacity-distortion function. The formulation and results of [6] have been extended to multi-user scenarios, i.e. multiple access and broadcast channels, in [7–9].

While the problem considered in [6] (then extended in [7–9]) and the one we formulate here share the same motivation, they differ in their underlying assumptions, hence leading to distinct approaches. A primary difference is that the channel state in [6–9] is assumed to be an i.i.d. process, varying from one symbol (or channel use) to another. In our problem formulation, the channel state is assumed to remain fixed throughout the transmission block. Our model is hence better suited for scenarios where the period over which the state remains unchanged is considerably larger than the timescale of a single communication block (or packet). One can envisage several practical scenarios where this is the case, e.g. as in the case of detecting the presence or absence of a road obstacle. The other key difference is the availability of generalized feedback in [6–9], which is not incorporated in our model. Instead we assume that the sensor has access to the message as side information, hence effectively co-locating it with the transmitter, yet channel inputs cannot be adapted based on the sensor's past observations.

Contribution: We study two instances of the joint communication and state detection problem illustrated in Fig. 1, where in both cases the state is assumed to be multiplicative and binary (on-off), hence reducing the sensing problem to a binary hypothesis test of signal versus noise.

First, we consider a binary symmetric setting with binary input and outputs. We establish the fundamental trade-off between reliable message communication and efficient state detection, captured by the message communication rate against the state detection error exponent (Theorem 1, Section 3). The trade-off in this setting is governed by a parameter that measures the minimum fraction of ones any codeword is allowed to have. We show that the optimal rate-exponent trade-off is achieved using constant-composition codes, which guarantee a fixed fraction of ones across all codewords.

Next, we consider a Gaussian setting with real input and outputs and additive Gaussian noise. Somewhat surprisingly, it turns out that there is no fundamental trade-off between the message communication rate and the state detection error exponent in this case, as the maximum rate and exponent can be simultaneously achieved (Theorem 2, Section 3). The key ingredient here is the use of almost constant-power codes, which achieve optimal performance for both message communication and state detection.

Notation: We use common notation which is explained along the way. For a positive integer integer a, the set $\{1, 2, \ldots, a\}$ is denoted by $[a]$. An indicator function $\mathbb{1}[\cdot]$ is equal to one when the condition in the argument is satisfied, and equal to zero otherwise. A Bernoulli distribution with parameter (i.e. mean) p is denoted by $\text{Bern}(p)$. A Gaussian distribution with mean μ and variance σ^2 is denoted by $\mathcal{N}(\mu, \sigma^2)$. Logarithms denoted by $\log(\cdot)$ are of base 2.

2 Problem Setting

Consider a setting with a transmitter, a receiver and a sensor. The transmitter has access to a message W drawn uniformly at random from a finite set $\mathcal{W}$. The message is mapped (i.e. encoded) into an input sequence $x^n(W) \triangleq x_1(W), x_2(W), \ldots, x_n(W)$ drawn from $\mathcal{X}^n$, where $\mathcal{X}$ is the input alphabet and $n \in \mathbb{N}$ is the length of the input sequence (or number of channel uses). The input sequence is transmitted over n uses of a noisy broadcast channel to the receiver and the sensor.

The receiver obtains a noisy output sequence $\tilde{Y}^n \triangleq \tilde{Y}_1, \tilde{Y}_2, \ldots, \tilde{Y}_n$ drawn from $\tilde{\mathcal{Y}}^n$, where $\tilde{\mathcal{Y}}$ is the corresponding output alphabet, through a stationary memoryless channel with a transition law of

$$P_{\tilde{Y}^n|X^n}(\tilde{y}^n|x^n) = \prod_{i=1}^{n} P_{\tilde{Y}|X}(\tilde{y}_i|x_i). \tag{1}$$

The receiver is interested in decoding the message W, and hence maps the observed sequence $\tilde{Y}^n$ into a message estimate $\hat{w}(\tilde{Y}^n)$. A message *decoding error* occurs whenever $\hat{w}(\tilde{Y}^n) \neq W$.

The sensor obtains a noisy output sequence $Y^n \triangleq Y_1, Y_2, \ldots, Y_n$ drawn from $\mathcal{Y}^n$, where $\mathcal{Y}$ is the corresponding output alphabet. The channel to the sensor is memoryless, yet it depends on a state S which takes values in a finite set $\mathcal{S}$ and remains fixed throughout the transmission. Given $S = s$, the transition law of the channel is given by

$$P^s_{Y^n|X^n}(y^n|x^n) = \prod_{i=1}^{n} P^s_{Y|X}(y_i|x_i). \tag{2}$$

The sensor is interested in detecting the state S, and has access to the message W (and hence the input sequence $x^n(W)$) as side information. The observed sequence Y^n and the side information W are mapped into a state estimate $\hat{s}(Y^n, W)$. A state *detection error* occurs whenever $\hat{s}(Y^n, W) \neq S$.

In this work, we restrict our attention to a *binary* state S, and hence we assume that $\mathcal{S} = \{0, 1\}$. We also focus on two basic settings: a binary symmetric setting and a Gaussian setting.

2.1 Binary symmetric setting

In the binary symmetric setting, we assume that $\mathcal{X} = \tilde{\mathcal{Y}} = \mathcal{Y} = \{0, 1\}$. In the i-th channel use, where $i \in [n]$, the outputs at the receiver and the sensor are respectively given by

$$\tilde{Y}_i = X_i \oplus \tilde{Z}_i \tag{3}$$

$$Y_i = S \cdot X_i \oplus Z_i \tag{4}$$

where $\oplus$ denotes the modulo-2 sum. $\tilde{Z}_i \sim \text{Bern}(p)$ and $Z_i \sim \text{Bern}(q)$ are independent additive noise terms with crossover parameters $p, q \in (0, 0.5)$, each assumed to be i.i.d. over channel uses. It is readily seen that the channel to the receiver has transition probabilities of

$$P_{\tilde{Y}|X}(\tilde{y}|x) = \begin{cases} 1-p, & x = \tilde{y} \\ p, & x \neq \tilde{y} \end{cases} \tag{5}$$

while the transition probabilities of the state-dependent channel to the sensor are given by

$$P^0_{Y|X}(y|x) = \begin{cases} 1-q, & y = 0 \\ q, & y = 1 \end{cases} \quad \text{and} \quad P^1_{Y|X}(y|x) = \begin{cases} 1-q, & y = x \\ q, & y \neq x. \end{cases} \tag{6}$$

2.2 Gaussian setting

Here we have $\mathcal{X} = \tilde{\mathcal{Y}} = \mathcal{Y} = \mathbb{R}$, and the input-output relationships for each $i \in [n]$ are given by

$$\tilde{Y}_i = X_i + \tilde{Z}_i \tag{7}$$

$$Y_i = S \cdot X_i + Z_i. \tag{8}$$

$\tilde{Z}_i \sim \mathcal{N}(0, \tilde{\sigma}^2)$ and $Z_i \sim \mathcal{N}(0, \sigma^2)$ are independent zero-mean Gaussian noise terms with variances $\tilde{\sigma}^2$ and σ^2 respectively, each assumed to be i.i.d. over channel uses. Admissible input sequences in $\mathbb{R}^n$ are those that satisfy the following average transmission power constraint:

$$\frac{1}{n} \sum_{i=1}^{n} |x_i|^2 \leq \mathsf{P} \tag{9}$$

where $\mathsf{P} > 0$ is the average transmission power budget. The transition probability laws of the above channels are given by the following Gaussian density functions

$$P_{\tilde{Y}|X}(\tilde{y}|x) = \frac{1}{\sqrt{2\pi\tilde{\sigma}^2}} \exp\left(-\frac{1}{2\tilde{\sigma}^2}(\tilde{y} - x)^2\right) \tag{10}$$

$$P^s_{Y|X}(y|x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(y - sx)^2\right). \tag{11}$$

We define the SNRs at the receiver and the sensor as $\mathsf{SNR}_1 \triangleq \frac{\mathsf{P}}{\tilde{\sigma}^2}$ and $\mathsf{SNR}_2 \triangleq \frac{\mathsf{P}}{\sigma^2}$, respectively.

2.3 Codes and performance measures

An (M_n, n) code for the above setting consists of a message set $\mathcal{W} = [M_n]$ and the following mappings:

- *Encoding function* $x^n : [M_n] \to \mathcal{X}^n$. This is employed by the transmitter, and maps each message $w \in [M_n]$ into a codeword (i.e. input sequence) $x^n(w) \in \mathcal{X}^n$.
- *Message decoding function* $\hat{w} : \tilde{\mathcal{Y}}^n \to [M_n] \cup \{\mathrm{e}\}$. This is employed by the receiver, and maps each output sequence $\tilde{y}^n \in \tilde{\mathcal{Y}}^n$ into a decoded message $\hat{w}(\tilde{y}^n) \in [M_n] \cup \{\mathrm{e}\}$, where e indicates an error.
- *State detection function* $\hat{s} : \mathcal{Y}^n \times [M_n] \to \mathcal{S} \cup \{\mathrm{e}\}$. This is employed by the sensor, and maps each pair of output sequence and message $(y^n, w) \in \mathcal{Y}^n \times [M_n]$ into a detected state $\hat{s}(y^n, w) \in \{0, 1\} \cup \{\mathrm{e}\}$.

For any (M_n, n) code, the corresponding *codebook* is given by $\mathcal{C}_n \triangleq \{x^n(w) : w \in [M_n]\}$, which is a subset of $\mathcal{X}^n$. The communication *rate* of an (M_n, n) code is given by $\frac{1}{n} \log M_n$ bits/channel use. The maximum message *decoding error probability* at the receiver is defined as

$$\lambda_n \triangleq \max_{w \in [M_n]} \mathbb{P}\left[\hat{w}(\tilde{Y}^n) \neq W | W = w\right]. \tag{12}$$

The maximum state *detection error probability* at the sensor is defined as

$$\varepsilon_n \triangleq \max_{w \in [M_n]} \max_{s \in \mathcal{S}} \mathbb{P}\Big[\hat{s}(Y^n, W) \neq S | W = w, S = s\Big]. \tag{13}$$

We are interested in the fundamental trade-off between the rate of reliable communication and the state detection error in the regime of asymptotically large block-length. This trade-off is formalized as follows.

Definition 1. The rate-exponent tuple (R, E) is said to be achievable if there exists a sequence of (M_n, n) codes, where $n \in \mathbb{N}$, for which $\lim_{n\to\infty} \lambda_n = 0$ (i.e. vanishing decoding error probability) and

$$\lim_{n\to\infty} \frac{1}{n} \log M_n = R \tag{14}$$

$$\lim_{n\to\infty} \frac{1}{n} \log \frac{1}{\varepsilon_n} = E. \tag{15}$$

The rate-exponent region $\mathcal{R}$ is the closure of the the set of all achievable pairs (R, E), that is

$$\mathcal{R} \triangleq \mathrm{cl}\big\{(R, E) : (R, E) \text{ is achievable}\big\}. \tag{16}$$

The error exponent E measures the efficiency of state detection at the sensor as the block-length n grows large. For example, given an achievable exponent E, the detection error probability is approximately equal to $\varepsilon_n \approx 2^{-nE}$ for large n, and hence achieving a detection error probability target of $\varepsilon^\star$ requires roughly $n^\star \approx \frac{1}{E} \log \frac{1}{\varepsilon^\star}$ channel uses. Simultaneously, given that a rate R is also achievable, then we can reliably communicate about $n^\star R$ bits of information to the receiver with a small decoding error, as R measures the efficiency of message communication in bits per channel use for large n. It is worthwhile highlighting that a missing figure of merit here is the message decoding error exponent, which captures the exponential decay rate of the message decoding error probability λ_n, see, e.g., [10]. Analysing this decoding error exponent, however, is outside the scope of the current work.

3 Main Results

In this section, we present the main result of this paper alongside some observations and insights.

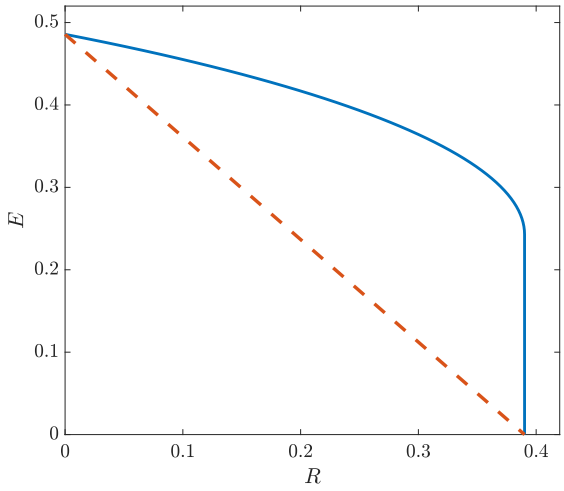

Figure 2: Rate-exponent trade-off for the binary symmetric setting with $p, q = 0.15$. The boundary of the optimal region is illustrated by the solid blue line. The red dashed line corresponds to time-sharing, where message communication and state detection are carried out separately in non-overlapping blocks of channel uses.

3.1 Binary symmetric setting

We define the binary entropy function as $H(u) = u \log \frac{1}{u} + (1-u) \log \frac{1}{1-u}$ and the binary relative entropy function as $D(u\|q) \triangleq u \log \frac{u}{q} + (1-u) \log \frac{1-u}{1-q}$, where $u, q \in [0,1]$.

Theorem 1. *For the binary symmetric setting, $\mathcal{R}$ consists of all tuples $(R, E) \in \mathbb{R}_+^2$ that satisfy*

$$\begin{aligned} R &\leq H(\alpha * p) - H(p) \\ E &\leq \alpha D(0.5\|q) \end{aligned} \tag{17}$$

for some $\alpha \in [0.5, 1]$, where $\alpha * p \triangleq \alpha(1-p) + (1-\alpha)p$.

The achievability of Theorem 1 is presented in Section 4, while the converse is given in Section 6. An illustration of the rate-exponent region in (17) is shown in Fig. 2.

The parameter α in the statement of Theorem 1 can be understood as the minimum fraction of ones a codeword is allowed to have. For instance, $\alpha = 1$ corresponds to a transmission of all ones, which is best for the purpose of state detection at the sensor, as zero inputs are useless for distinguishing between $S = 0$ and $S = 1$. In this case, a state detection error exponent of $D(0.5\|q)$ is achieved, which can be understood as follows. The typical outcome of the noise process Z^n at the sensor is a sequence of roughly qn ones and $(1-q)n$ zeros, hence flipping around qn bits of the state-multiplied codeword $S \cdot x^n$. Successful detection of S is possible as long as less than half the bits in $S \cdot x^n$ are flipped. From large deviation theory, the probability that Z^n produces a sequence with at least $0.5n$ ones, hence causing a detection error, is proportional to $2^{-nD(0.5\|q)}$, hence yielding an error exponent of $D(0.5\|q)$.

Achieving the maximum exponent of $D(0.5\|q)$ comes at the expense of the message communication rate, as an all ones transmission bears no information at all (top-left corner point of the region in Fig. 2). Achieving non-zero rates requires selecting $\alpha < 1$, which in turn decreases the exponent to $\alpha D(0.5\|q)$, as only a fraction α of channel uses are useful for state detection in this case (for the worst-case codeword). The maximum message communication rate is achieved when $\alpha = 0.5$, where the empirical distribution of ones and zeros in any codeword can be made to match the capacity-achieving input distribution for the binary symmetric channel (BSC). In this case, the exponent is given by $0.5D(0.5\|q)$, as at least half of the entries in any codeword are still useful for state detection (right-endpoint of the curved segment in Fig. 2). This is still much more efficient than carrying out message communication and state detection separately, as seen in Fig. 2.

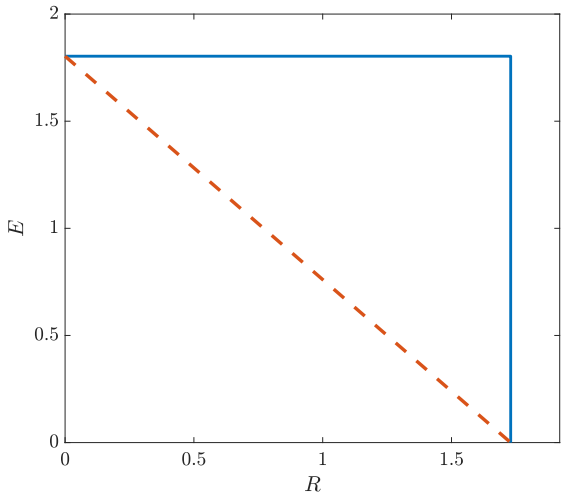

Figure 3: Rate-exponent trade-off for the Gaussian setting with $\mathsf{SNR}_1, \mathsf{SNR}_2 = 10$ dB. The boundary of the optimal region is illustrated by the solid blue line. The red dashed line corresponds to time-sharing.

3.2 Gaussian setting

Theorem 2. *For the Gaussian setting, $\mathcal{R}$ consists of all tuples $(R, E) \in \mathbb{R}_+^2$ that satisfy*

$$\begin{aligned} R &\leq \frac{1}{2}\log\left(1+\mathsf{SNR}_1\right) \\ E &\leq \frac{1}{8}\mathsf{SNR}_2 \log e. \end{aligned} \tag{18}$$

The achievability of Theorem 2 is presented in Section 4, and the converse is given in Section 6. Perhaps somewhat surprisingly, in the Gaussian setting there is in fact no fundamental trade-off between message communication and state detection: the maximum rate of $\frac{1}{2}\log\left(1+\mathsf{SNR}_1\right)$ and the maximum exponent of $\frac{1}{8}\mathsf{SNR}_2 \log e$ can be simultaneously achieved as illustrated in Fig. 3.

Given that codeword x^n is sent by the transmitter, state detection at the sensor is essentially to decide whether an observed sequence y^n is a noisy version of 0^n or x^n, where 0^n is an all zeros n-sequence. Maximum likelihood detection in this case boils down to a minimum-distance decision rule: $S = \hat{s}$ if y^n is closer to $\hat{s} \cdot x^n$ than to $(1-\hat{s}) \cdot x^n$ in the Euclidean sense. Detection errors are hence minimized by maximising the Euclidean distance between 0^n and x^n, which is accomplished by selecting x^n to be on the surface of an n-ball of radius $\sqrt{n\mathsf{P}}$. As for message communication to the receiver, it is well known that the capacity-achieving input distribution in this case is Gaussian with mean zero and variance P. By the sphere hardening property, a typical codeword x^n generated by the capacity-achieving input distribution will, with high probability, lie in a thin spherical shell that includes the surface of an n-ball with radius $\sqrt{n\mathsf{P}}$. Therefore, restricting codewords to have Euclidean norms close to $\sqrt{n\mathsf{P}}$ achieves the best possible detection error exponent, and is sufficient to achieve rates close to capacity as n grows large.

Remark 1. It is worthwhile noting that the maximum exponent in (18) is equal to

$$\frac{1}{8}\mathsf{SNR}_2 \log e = D\Big(\mathcal{N}\big(0.5\sqrt{\mathsf{P}}, \sigma^2\big) \parallel \mathcal{N}\big(0, \sigma^2\big)\Big) \tag{19}$$

which is the relative entropy between two Gaussian distributions, $\mathcal{N}\big(0.5\sqrt{\mathsf{P}}, \sigma^2\big)$ and $\mathcal{N}\big(0, \sigma^2\big)$. Therefore, the region in Theorem 2 can be expressed as the set of non-negative tuples (R, E) that satisfy

$$\begin{aligned} R &\leq \frac{1}{2}\log\left(1+\mathsf{SNR}_1\right) \\ E &\leq D\Big(\mathcal{N}\big(0.5\sqrt{\mathsf{P}}, \sigma^2\big) \parallel \mathcal{N}\big(0, \sigma^2\big)\Big). \end{aligned} \tag{20}$$

This bears close resemblance to the binary symmetric setting result in Theorem 1. Intuitively, the relative entropy expression is explained by the fact that the detection error probability is dominated by the probability of the event that the channel produces an output which is close to the maximum likelihood decision boundary, e.g. an outcome of $\mathcal{N}(0,\sigma^2)$ would look as if it was generated by $\mathcal{N}(0.5\sqrt{\mathsf{P}},\sigma^2)$. The latter distribution is also known as a tilted distribution, and often appears in characterisations of hypothesis testing error exponents, see, e.g. [1] and [11, Ch. 3.5].

4 Achievability: Binary Symmetric Setting

To show that the region $\mathcal{R}$ in Theorem 1 is achievable, it is sufficient to show that for any parameter $\alpha \in [0.5, 1]$, the rate-exponent tuple (R, E) with

$$R = \ H(\alpha * p) - H(p) - \epsilon \ \text{ and } \ E = \alpha D(0.5\|q) - \epsilon \tag{21}$$

is achievable, where $\epsilon > 0$ can be made as small as desired. All remaining points in the region can be achieved using a time-sharing argument, and the closure of all such points gives (17).

Now let us fix $\alpha \in [0.5, 1]$ and define a binary input random variable $X \sim \text{Bern}(\alpha)$. For such input, the mutual information at the output of the receiver is given by $I(X;\tilde{Y}) = H(\alpha * p) - H(p)$. It is known from the channel coding theorem that for any small $\epsilon > 0$, the rate $R = I(X;\tilde{Y}) - \epsilon$ is achievable, i.e. there exists a sequence of (M_n, n) codes for which $\lambda_n \to 0$ and $\frac{1}{n}\log M_n \to R$ as $n \to \infty$. We wish to show that at least one such sequence of codes achieves an exponent $E = \alpha D(0.5\|q) - \epsilon$. To this end, we start by analysing the state detection error probability ε_n for an arbitrary codebook $\mathcal{C}_n$.

4.1 Achievable error exponent

We start with the following definition which will be useful further on.

Definition 2. For a finite alphabet $\mathcal{X}$, the *composition* of any sequence $x^n \in \mathcal{X}^n$ is defined as

$$\pi(x|x^n) \triangleq \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}\left[x_i = x\right], \ x \in \mathcal{X} \tag{22}$$

which is also known as the *empirical distribution* or *type*. In the binary case $\mathcal{X} = \{0,1\}$, the composition is fully determined by $\pi(1|x^n)$. Hence we will refer to $\pi(1|x^n)$ as the composition of x^n.

Now suppose that an arbitrary message $W = w$ is selected, where $w \in [M_n]$, and the transmitter sends the binary codeword $x^n(w) \in \mathcal{C}_n$. For ease of notation, we drop the message index and simply denote $x^n(w)$ by x^n. Given a received sequence $Y^n = y^n$, and with knowledge of w (and hence x^n), the sensor estimates the state using the maximum likelihood (ML) detection rule described as follows:

$$\hat{s}(y^n, w) = \begin{cases} 0, & P^0_{Y^n|X^n}(y^n|x^n) > P^1_{Y^n|X^n}(y^n|x^n) \\ 1, & P^0_{Y^n|X^n}(y^n|x^n) < P^1_{Y^n|X^n}(y^n|x^n) \\ \text{e}, & P^0_{Y^n|X^n}(y^n|x^n) = P^1_{Y^n|X^n}(y^n|x^n). \end{cases} \tag{23}$$

When $S = s$, a detection error occurs whenever the sensor observes a sequence y^n that leads to $\hat{s}(y^n, w) \neq s$ according to the rule in (23). The probability of such event is bounded above as:

$$\mathbb{P}\left[\hat{s}(Y^n, W) \neq S | W = w, S = s\right] = \sum_{y^n \in \mathcal{Y}^n} P^s_{Y^n|X^n}(y^n|x^n)\mathbb{1}\left[P^{1-s}_{Y^n|X^n}(y^n|x^n) \geq P^s_{Y^n|X^n}(y^n|x^n)\right] \tag{24}$$

$$\leq \sum_{y^n \in \mathcal{Y}^n} P^s_{Y^n|X^n}(y^n|x^n)\sqrt{\frac{P^{1-s}_{Y^n|X^n}(y^n|x^n)}{P^s_{Y^n|X^n}(y^n|x^n)}} \tag{25}$$

$$= \sum_{y^n \in \mathcal{Y}^n} \sqrt{P^1_{Y^n|X^n}(y^n|x^n) P^0_{Y^n|X^n}(y^n|x^n)} \tag{26}$$

$$= \sum_{y_1, y_2, \ldots, y_n} \prod_{i=1}^{n} \sqrt{P^1_{Y|X}(y_i|x_i) P^0_{Y|X}(y_i|x_i)} \tag{27}$$

$$= \prod_{i=1}^{n} \sum_{y \in \{0,1\}} \sqrt{P^1_{Y|X}(y|x_i) P^0_{Y|X}(y|x_i)} \tag{28}$$

$$= \left(2\sqrt{q(1-q)}\right)^{\sum_{i=1}^{n} \mathbb{1}[x_i = 1]} \tag{29}$$

$$= \left(2\sqrt{q(1-q)}\right)^{n\pi(1|x^n)} \tag{30}$$

$$= 2^{-n\pi(1|x^n) D(0.5\|q)}. \tag{31}$$

The equality in (29) follows from the observation that

$$\sum_{y \in \{0,1\}} \sqrt{P^1_{Y|X}(y|x) P^0_{Y|X}(y|x)} = \begin{cases} 1, & x = 0 \\ 2\sqrt{q(1-q)}, & x = 1. \end{cases} \tag{32}$$

The equality in (31) holds by definition of the relative entropy. The above upper bound is known as the Bhattacharyya bound, and it is a special case of the Chernoff bound, see, e.g. [11, Ch. 2.3].

From (24)–(31), we obtain an upper bound for the achievable detection error probability ε_n by taking the maximum over both s and w. Note that maximising over $w \in [M_n]$ translates to maximising over $x^n \in \mathcal{C}_n$ in the right-hand-side of (31), and hence we have

$$\varepsilon_n \leq \max_{x^n \in \mathcal{C}_n} 2^{-n\pi(1|x^n) D(0.5\|q)}. \tag{33}$$

Now suppose that we have a sequence of (M_n, n) codes with vanishing decoding error probability. For such sequence of codes, the upper bound in (33) implies that the following exponent is achievable:

$$E = \lim_{n \to \infty} \min_{x^n \in \mathcal{C}_n} \pi(1|x^n) D(0.5\|q). \tag{34}$$

Therefore, our problem reduces to showing that for any small $\epsilon, \delta > 0$, there exists a sequence of (M_n, n) codes with vanishing decoding error probability such that

$$\lim_{n \to \infty} \frac{1}{n} \log M_n \geq H(\alpha * p) - H(p) - \epsilon \quad \text{and} \quad \lim_{n \to \infty} \min_{x^n \in \mathcal{C}_n} \pi(1|x^n) \geq \alpha - \delta. \tag{35}$$

Next, we show the existence of such sequence of codes.

4.2 Constant-composition codes

Definition 3. An (M_n, n, α_n) *constant-composition* binary code has M_n codewords in $\{0,1\}^n$, all of which have the same composition α_n, where $\alpha_n \in [0,1]$ such that $n\alpha_n$ is integer. That is

$$\pi(1|x^n) = \alpha_n, \text{ for all } x^n \in \mathcal{C}_n. \tag{36}$$

It is well known that the capacity of the BSC, or any other discrete memoryless channel, is achievable using constant-composition codes [12]. This result extends to all rates below capacity as seen next.

Theorem 3. *For the BSC in* (3)*, and for any* $\alpha \in [0.5, 1]$ *and* $\epsilon, \delta > 0$*, there exists a sequence of* (M_n, n, α_n) *constant-composition codes with vanishing decoding error probability, and for which*

$$\lim_{n \to \infty} \frac{1}{n} \log M_n \geq H(\alpha * p) - H(p) - \epsilon \quad \textit{and} \quad \lim_{n \to \infty} \alpha_n \geq \alpha - \delta. \tag{37}$$

A proof for Theorem 3 is presented in Appendix A. By setting $\delta = \epsilon/D(0.5\|q)$ in Theorem 3, and combining the result with (34), we conclude that the rate-exponent tuple (R, E) in (21) is achievable. This completes the proof of achievability for Theorem 1.

Remark 2. Theorem 3 can be shown to hold from [12, Problem 6.19]. The proof presented in Appendix A, however, avoids the use of strong typicality and analysis based on the method of types, as done in [12]. While our proof may be of interest in its own right, one of its advantages is that its main steps can be adapted to deal with the Gaussian setting, as seen in the proof of Theorem 4.

5 Achievability: Gaussian setting

To show that $\mathcal{R}$ in Theorem 2 is achievable, we only need to show that the tuple (R, E) with

$$R = \frac{1}{2}\log\left(1 + \frac{\mathsf{P}}{\tilde{\sigma}^2}\right) - \epsilon \quad \text{and} \quad E = \frac{\mathsf{P}}{8\sigma^2}\log e - \epsilon \tag{38}$$

is achievable, where $\epsilon > 0$ can be made as small as desired.

5.1 Achievable error exponent

We start with the following definition of average power.

Definition 4. The average power of a real sequence $x^n \in \mathbb{R}^n$ is defined as

$$\mathcal{P}(x^n) \triangleq \frac{1}{n}\sum_{i=1}^{n} |x_i|^2. \tag{39}$$

Upon receiving a noisy sequence $Y^n = y^n$, and with knowledge of the selected message $W = w$ and the corresponding input sequence $X^n = x^n$, the sensor estimates the state using the ML detection rule in (23). An upper bound for the detection error probability, given that $W = w$ and $S = s$, is obtained by adapting the Bhattacharyya bound in Section 4.1 to the Gaussian setting as follows:

$$\mathbb{P}\left[\hat{s}(Y^n, W) \neq S | W = w, S = s\right] \leq \int_{y^n \in \mathcal{Y}^n} \sqrt{P^1_{Y^n|X^n}(y^n|x^n) P^0_{Y^n|X^n}(y^n|x^n)}\mathrm{d}y^n \tag{40}$$

$$= \int_{y_1,\ldots,y_n} \prod_{i=1}^{n} \sqrt{P^1_{Y|X}(y_i|x_i) P^0_{Y|X}(y_i|x_i)}\mathrm{d}y_1 \ldots \mathrm{d}y_n \tag{41}$$

$$= \prod_{i=1}^{n} \int_{y\in\mathbb{R}} \sqrt{P^1_{Y|X}(y|x_i) P^0_{Y|X}(y|x_i)}\mathrm{d}y \tag{42}$$

$$= \prod_{i=1}^{n} \int_{y\in\mathbb{R}} \frac{1}{\sqrt{2\pi\sigma^2}} \sqrt{\exp\left(-\frac{\left(y - x_i\right)^2}{2\sigma^2}\right)\exp\left(-\frac{y^2}{2\sigma^2}\right)}\mathrm{d}y \tag{43}$$

$$= \prod_{i=1}^{n} \int_{y\in\mathbb{R}} \frac{1}{\sqrt{2\pi\sigma^2}} \sqrt{\exp\left(-\frac{1}{\sigma^2}\left(y - \frac{1}{2}x_i\right)^2 - \frac{x_i^2}{4\sigma^2}\right)}\mathrm{d}y \tag{44}$$

$$= \prod_{i=1}^{n} \exp\left(-\frac{x_i^2}{8\sigma^2}\right) \int_{y\in\mathbb{R}} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}\left(y - \frac{1}{2}x_i\right)^2\right)\mathrm{d}y \tag{45}$$

$$= \exp\left(-\frac{\sum_{i=1}^{n}|x_i|^2}{8\sigma^2}\right) \tag{46}$$

$$= \exp\left(-n\frac{\mathcal{P}(x^n)}{8\sigma^2}\right). \tag{47}$$

From the above, we obtain an upper bound for the achievable detection error probability ε_n by taking the maximum of the bound in (47) over both $s \in \{0, 1\}$ and $x^n \in \mathcal{C}_n$, leading to

$$\varepsilon_n \leq \max_{x^n \in \mathcal{C}_n} 2^{-n \frac{\mathcal{P}(x^n)}{8\sigma^2} \log e}. \tag{48}$$

Now suppose that we have a sequence of (M_n, n) codes with vanishing decoding error probability. For such sequence of codes, the upper bound in (48) implies that the following exponent is achievable:

$$E = \lim_{n \to \infty} \min_{x^n \in \mathcal{C}_n} \frac{\mathcal{P}(x^n)}{8\sigma^2} \log e. \tag{49}$$

Hence the achievability problem reduces to showing that there exists a sequence of (M_n, n) codes with vanishing decoding error probability, which satisfy the average power constraint in (9), and for which

$$\lim_{n \to \infty} \frac{1}{n} \log M_n \geq \frac{1}{2} \log \left(1 + \frac{\mathsf{P}}{\tilde{\sigma}^2}\right) - \epsilon \quad \text{and} \quad \lim_{n \to \infty} \min_{x^n \in \mathcal{C}_n} \mathcal{P}(x^n) \geq \mathsf{P} - \delta \tag{50}$$

for any $\epsilon, \delta > 0$. Setting $\delta = (8\sigma^2\epsilon)/(\log e)$, the desired (R, E) tuple is achieved. Next, we show the existence of the desired sequence of codes.

5.2 Almost constant-power codes

Next, we introduce a class of codes for the Gaussian channel.

Definition 5. An $(M_n, n, \mathsf{P}_n, \delta)$ *almost constant-power* code has M_n codewords in $\mathbb{R}^n$, all of almost the same power P_n. That is, for some $0 < \delta \ll \mathsf{P}_n$, we have

$$\mathsf{P}_n - \delta \leq \mathcal{P}(x^n) \leq \mathsf{P}_n, \text{ for all } x^n \in \mathcal{C}_n. \tag{51}$$

As it turns out, almost constant-power codes achieve the capacity of the Gaussian channel.

Theorem 4. *For the Gaussian channel in* (7)*, and for any $\epsilon, \delta > 0$, there exists a sequence of $(M_n, n, \mathsf{P}_n, \delta)$ almost constant-power codes with $\mathsf{P}_n \leq \mathsf{P}$ and vanishing decoding error probability, and for which*

$$\lim_{n \to \infty} \frac{1}{n} \log M_n \geq \frac{1}{2} \log \left(1 + \frac{\mathsf{P}}{\tilde{\sigma}^2}\right) - \epsilon \quad \textit{and} \quad \lim_{n \to \infty} \mathsf{P}_n = \mathsf{P}. \tag{52}$$

A proof for Theorem 4 is presented in Appendix B. Setting $\delta = (8\sigma^2\epsilon)/(\log e)$ in Theorem 4, and combining with (49), it follows that the rate-exponent tuple (R, E) with

$$R = \frac{1}{2} \log \left(1 + \frac{\mathsf{P}}{\tilde{\sigma}^2}\right) - \epsilon \quad \text{and} \quad E = \frac{\mathsf{P}}{8\sigma^2} \log e - \epsilon \tag{53}$$

is achievable, which in turn completes the proof of achievability for Theorem 2.

Remark 3. Constant-power codes, in which all codewords have the same power of $n\mathsf{P}$, were analysed by Shannon in a paper that dates back to 1959 [13]. There, it was shown that constant-power codes achieve the capacity of the Gaussian channel under an average power constraint. This classical result by Shannon directly implies Theorem 4. The proof by Shannon, however, is involved and relies on a geometric analysis of the reliability function (i.e. the channel coding exponent). The proof we present in Appendix 4 avoids such intricate analysis and instead relies on a multi-letter achievability argument through, e.g., weak typicality, as well as a simple application of the entropy power inequality.

6 Converse

We start by deriving a lower bound for the state detection error probability ε_n which holds for any detection rule, given a fixed codebook $\mathcal{C}_n$. To this end, we make use of a general result by Shannon, Gallager and Berlekamp [14] on the probability of decoding error for a block code with two codewords. Note that in our setting, given that a codeword x^n is sent by the transmitter, state detection at the sensor is equivalent to distinguishing between two possible state-multiplied codewords: $0 \cdot x^n$ and $1 \cdot x^n$.

Given that a codeword $x^n \in \mathcal{C}_n$ is selected, the function $\mu : [0,1] \to (-\infty, 0]$ is defined as

$$\mu(\nu|x^n) \triangleq \log \sum_{y^n \in \mathcal{Y}^n} P^0_{Y^n|X^n}(y^n|x^n)^{1-\nu} P^1_{Y^n|X^n}(y^n|x^n)^{\nu}. \tag{54}$$

In the Gaussian case, with a continuous alphabet and well-defined densities, we instead write

$$\mu(\nu|x^n) \triangleq \log \int_{y^n \in \mathbb{R}^n} P^0_{Y^n|X^n}(y^n|x^n)^{1-\nu} P^1_{Y^n|X^n}(y^n|x^n)^{\nu} \mathrm{d}y^n. \tag{55}$$

For the settings of interest to us here, $P^s_{Y^n|X^n}(y^n|x^n) > 0$ holds for all $s \in \{0,1\}$, $y^n \in \mathcal{Y}^n$ and $x^n \in \mathcal{X}^n$. Therefore, the quantities in (54) and (55) are well defined for any $\nu \in [0,1]$. It is also worthwhile noting that $\mu(0.5|x^n)$ is equal to the logarithm of the Bhattacharyya coefficient in (26) and (40).

The next result is a straightforward adaptation of the corollary that follows [14, Theorem 5]. We denote by $\varepsilon_n(s|x^n)$ the detection error for state $s \in \{0,1\}$, given codeword $x^n \in \mathcal{C}_n$

Shannon-Gallager-Berlekamp lower bound. *Let $\nu^\star$ minimizes $\mu(\nu|x^n)$ over $\nu \in [0,1]$, and $\mu''(\nu|x^n)$ be the second derivative of $\mu(\nu|x^n)$. For any state detection function, the following lower bound holds:*

$$\log \varepsilon_n(s|x^n) \geq \mu(\nu^\star|x^n) - \sqrt{2\mu''(\nu^\star|x^n)} - 2, \ \ s \in \{0,1\}. \tag{56}$$

Moreover, the term $\sqrt{2\mu''(\nu^\star|x^n)}$ is proportional to $\sqrt{n}$.

From the above result, $\log \varepsilon_n$ is bounded below as follows:

$$\begin{aligned}
\log \varepsilon_n &= \max_{x^n \in \mathcal{C}_n} \max_{s \in \{0,1\}} \log \varepsilon_n(s|x^n) && (57)\\
&\geq \max_{x^n \in \mathcal{C}_n} \left\{ \mu(\nu^\star|x^n) - \sqrt{2\mu''(\nu^\star|x^n)} \right\} - 2 && (58)\\
&\geq \mu(\nu^\star|x^{\star n}) - \sqrt{2\mu''(\nu^\star|x^{\star n})} - 2 && (59)
\end{aligned}$$

where $x^{\star n}$ maximizes $\mu(\nu^\star|x^n)$ over $x^n \in \mathcal{C}_n$. Now suppose that we have a sequence of codebooks $(\mathcal{C}_n)_{n \in \mathbb{N}}$. The lower bound in (59) implies that the state detection error exponents is bounded above as

$$E \leq \lim_{n \to \infty} \left\{ -\frac{1}{n} \max_{x^n \in \mathcal{C}_n} \mu(\nu^\star|x^n) \right\} \tag{60}$$

provided that the limit exists. (60) follows from (59) by noting that $\frac{1}{n}\sqrt{2\mu''(\nu^\star|x^{\star n})} + \frac{2}{n} \to 0$ as $n \to \infty$.

6.1 Binary symmetric setting

We now specialize (60) to the binary symmetric setting. For any $x^n \in \{0,1\}^n$, we have

$$\begin{aligned}
\mu(\nu|x^n) &= \sum_{i=1}^{n} \log \sum_{y \in \{0,1\}} P^0_{Y|X}(y_i|x_i)^{1-\nu} P^1_{Y|X}(y_i|x_i)^{\nu} && (61)\\
&= n\pi(1|x^n) \log \left(q^{1-\nu}(1-q)^{\nu} + q^{\nu}(1-q)^{1-\nu} \right) && (62)
\end{aligned}$$

which can be obtained by extending the steps in (26)–(30). The function $q^{1-\nu}(1-q)^{\nu}+q^{\nu}(1-q)^{1-\nu}$ is convex in $\nu \in [0,1]$, and it is minimized by $\nu^{\star}=0.5$ (see [10, Ch. 5.3]). Therefore, we have

$$\mu(\nu^{\star}|x^n) = n\pi(1|x^n)\log\left(2\sqrt{q(1-q)}\right) = -n\pi(1|x^n)D(0.5\|q). \tag{63}$$

It follows from (60) that E is bounded as

$$E \leq \lim_{n\to\infty} \min_{x^n\in\mathcal{C}_n} \pi(1|x^n)D(0.5\|q). \tag{64}$$

Since the above upper bound is also achievable (see Section 4.1), it follows that it is the optimal detection error exponent given a sequence of codebooks. It is also evident that $0 \leq E \leq D(0.5\|q)$.

From the above, it follows that any tuple $(R,E) \in \mathcal{R}$ can be expressed as $\big(R, \alpha D(0.5\|q)\big)$ for some parameter $\alpha \in [0,1]$. The boundary points of $\mathcal{R}$ are characterized by $\big(R(\alpha), \alpha D(0.5\|q)\big)$, where

$$R(\alpha) \triangleq \sup\left\{R : \big(R, \alpha D(0.5\|q)\big) \in \mathcal{R}\right\}. \tag{65}$$

Therefore, any upper bound for $R(\alpha)$ will result in an outer bound for the rate-exponent region $\mathcal{R}$. For the range of parameters $\alpha \in [0, 0.5)$, we use the simple upper bound given by

$$R(\alpha) \leq 1 - H(p),\ \alpha \in [0, 0.5) \tag{66}$$

which follows from the capacity of the BSC. We focus on $\alpha \in [0.5, 1]$ henceforth.

Consider a sequence of (M_n, n) binary codes with vanishing decoding error at the receiver and detection error exponent of $E = \alpha D(0.5\|q)$ at the sensor, for some $\alpha \in [0.5, 1]$. From (64), we have

$$E = \alpha D(0.5\|q) \implies \lim_{n\to\infty} \min_{x^n\in\mathcal{C}_n} \pi(1|x^n) \geq \alpha. \tag{67}$$

Using α_n to denote $\min_{x^n\in\mathcal{C}_n} \pi(1|x^n)$, then for any $n \in \mathbb{N}$ we have

$$\pi\big(1|x^n(w)\big) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}\left[x_i(w)=1\right] \geq \alpha_n,\ \forall w \in [M_n]. \tag{68}$$

The set of inequalities in (68) imply

$$\frac{1}{n}\sum_{i=1}^{n} \mathbb{E}\left[x_i(W)\right] = \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}\left[X_i\right] \geq \alpha_n \tag{69}$$

where the expectations are with respect to W, which is uniformly distributed on $[M_n]$, and $X_i = x_i(W)$ is a random variable corresponding to the i-th entry of a randomly selected codeword from $\mathcal{C}_n$.

Note that (68) can be seen as a per-codeword *minimum cost* constraint. Unlike the standard setting where a maximum cost must not be exceeded (see, e.g., [15, Ch. 3.3]), here no codeword should fall short of a prescribed minimum cost. We wish to find an upper bound for $R(\alpha)$ under such constraint. By a standard application of Fano's inequality and the data processing inequality, we proceed as

$$\begin{aligned}
\log M_n &\leq I\big(X^n; \tilde{Y}^n\big) + n\epsilon_n && (70)\\
&\leq \sum_{i=1}^{n} H(\tilde{Y}_i) - nH(p) + n\epsilon_n && (71)\\
&= \sum_{i=1}^{n} H\left(\mathbb{E}[X_i] * p\right) - nH(p) + n\epsilon_n && (72)\\
&\leq nH\left(\frac{1}{n}\sum_{i=1}^{n} \mathbb{E}[X_i] * p\right) - nH(p) + n\epsilon_n. && (73)
\end{aligned}$$

In the above, ϵ_n is a term that goes to zero as n goes to infinity, while (72) holds since $\tilde{Y}_i = X_i \oplus \tilde{Z}_i$, $X_i \sim \text{Bern}\big(\mathbb{E}[X_i]\big)$ and $\tilde{Z}_i \sim \text{Bern}(p)$. The inequality in (73) follows from Jensen's inequality and the concavity of $H(\beta * p)$ in $\beta \in [0,1]$ (see Appendix C). From the above, we proceed as follows

$$R(\alpha) \leq \lim_{n\to\infty} H\left(\frac{1}{n}\sum_{i=1}^{n} \mathbb{E}[X_i] * p\right) - H(p) \tag{74}$$

$$= H\left(\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}[X_i] * p\right) - H(p) \tag{75}$$

$$\leq H\left(\alpha * p\right) - H(p) \tag{76}$$

where (75) follows by continuity. The function $\beta * p = \beta(1-2p) + p$ is increasing in β for any $p \in (0, 0.5)$, and $H(u)$ is decreasing in $u \in [0.5, 1]$. Since $\beta * p$ is in $[0.5, 1]$ whenever $\beta \in [0.5, 1]$ and $p \in (0, 0.5)$, then $H(\beta * p)$ is decreasing in $\beta \in [0.5, 1]$. The inequality in (76) follows by combining this with

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}[X_i] \geq \lim_{n\to\infty} \alpha_n \geq \alpha \geq 0.5. \tag{77}$$

From the upper bounds in (66) and (76), it follows that the rate-exponent region $\mathcal{R}$ is included in the outer bound given by the set of all non-negative rate-exponent tuples (R, E) that satisfy

$$\begin{aligned} R &\leq H(\max\{\alpha, 0.5\} * p) - H(p) \\ E &\leq \alpha D(0.5 \| q) \end{aligned} \tag{78}$$

for some $\alpha \in [0, 1]$. The region described in (78) is equivalent to the one given in (17) in Theorem 1, which was shown to be achievable in Section 4. This proves the converse of Theorem 1.

6.2 Gaussian setting

In the Gaussian setting, and for any $x^n \in \mathbb{R}^n$, we have

$$\mu(\nu|x^n) = \sum_{i=1}^{n} \log \int_{y\in\mathbb{R}} P^0_{Y|X}(y|x_i)^{1-\nu} P^1_{Y|X}(y|x_i)^{\nu} \mathrm{d}y \tag{79}$$

$$= n(\nu^2 - \nu)\frac{\mathcal{P}(x^n)}{2\sigma^2}\log e \tag{80}$$

obtained by extending the calculation in (40)–(47) as follows:

$$\int_{y\in\mathbb{R}} P^0_{Y|X}(y|x_i)^{1-\nu} P^1_{Y|X}(y|x_i)^{\nu} \mathrm{d}y = \int_{y\in\mathbb{R}} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{\nu\big(y - x_i\big)^2}{2\sigma^2}\right) \exp\left(-\frac{(1-\nu)y^2}{2\sigma^2}\right) \mathrm{d}y \tag{81}$$

$$= \int_{y\in\mathbb{R}} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{\big(y - \nu x_i\big)^2}{2\sigma^2} - \frac{(\nu - \nu^2)x_i^2}{2\sigma^2}\right) \mathrm{d}y \tag{82}$$

$$= \exp\left(-\frac{(\nu-\nu^2)x_i^2}{2\sigma^2}\right) \int_{y\in\mathbb{R}} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}\left(y - \nu x_i\right)^2\right) \mathrm{d}y \tag{83}$$

$$= \exp\left(-\frac{\nu(1-\nu)x_i^2}{2\sigma^2}\right). \tag{84}$$

The function $(\nu^2 - \nu)$ in (80) is convex in [0,1], and it is minimized by $\nu^\star = 0.5$. Hence we have

$$\mu(\nu^\star|x^n) = -n\frac{\mathcal{P}(x^n)}{8\sigma^2}\log e. \tag{85}$$

Now consider a sequence of (M_n, n) codes for the Gaussian setting, which satisfy the power constraint in (9), and which have a vanishing probability of decoding error at the receiver. From the capacity of the Gaussian channel with an average power constraint, it follows that the rate is bounded above as

$$R \leq \frac{1}{2}\log\left(1+\mathsf{SNR}_1\right). \tag{86}$$

On the other hand, from (85) and (60), the state detection error exponent E is bounded above as

$$E \leq \lim_{n\to\infty} \min_{x^n\in\mathcal{C}_n} \frac{\mathcal{P}(x^n)}{8\sigma^2}\log e \tag{87}$$

$$\leq \frac{1}{8}\mathsf{SNR}_2 \log e. \tag{88}$$

This proves the converse of Theorem 2.

7 Conclusion

In this paper, we brought forth what is perhaps a new perspective on the information-theoretic formulation of joint communication and sensing problems. Contrary to the i.i.d. state process assumed in [6], in our proposed formulation we assume that the unknown state to be sensed remains fixed throughout the transmission period. This assumption renders the rate-distortion approach in [6] unsuitable for the case considered here. Alternatively, we proposed to characterize the performance trade-off in terms of a rate-exponent region. We established the optimal rate-exponent regions for a binary symmetric setting and a Gaussian setting, both with multiplicative binary states. The methods we employed to derive the optimal rate-exponent regions include tools that are commonly used to study binary hypothesis testing problems.

The problem formulation proposed in this paper can be extended and generalized in several directions. For instance, one may consider more general discrete memoryless channels where the binary state is not necessarily multiplicative, but is used to identify two distinct channel laws $P^0_{Y|X}$ and $P^1_{Y|X}$ for the sensor. Another interesting direction is to consider M-ary states instead of binary states. For the Gaussian setting, it will be interesting to consider the problem of joint communication and parameter estimation, where the state to be estimated is a real-valued parameter. It is also of interest to extend the model to capture parameters of practical interest as, e.g, time delay and angle-of-arrival which are relevant in radar sensing applications. Finally, in detection problems, it is occasionally the case that different types of errors are not weighted equally. For example, in obstacle detection to avoid road collisions, a missed detection can be much worse than a false alarm. This motivates investigating the problem within a Neyman-Pearson binary hypothesis testing framework, where the objective is to minimize one type of error probability (e.g. missed detection) while keeping the other type (e.g. false alarm) below a certain threshold. Characterising the rate-exponent trade-off under such formulation is of interest.

A Proof of Theorem 3

Let k be a positive integer and $\beta \in [0,1]$ be such that $k\beta$ is also an integer. β is a valid composition for binary sequences of length k, and the set of all such sequences is given by

$$\mathcal{T}^k(\beta) \triangleq \left\{x^k \in \{0,1\}^k : \pi(1|x^k) = \beta\right\}. \tag{89}$$

The set $\mathcal{T}^k(\beta)$ is known as a composition class (or a type class), and it has a cardinality of $|\mathcal{T}^k(\beta)| = \binom{k}{k\beta}$. It is well known that $|\mathcal{T}^k(\beta)|$ is bounded in terms of the binary entropy $H(\beta)$ as [12, Lemma 2.3]:

$$\frac{1}{(k+1)}2^{kH(\beta)} \leq |\mathcal{T}^k(\beta)| \leq 2^{kH(\beta)}. \tag{90}$$

We now define a distribution P_{X^k} on $\{0,1\}^k$ as follows:

$$P_{X^k}(x^k) = \begin{cases} \frac{1}{|\mathcal{T}^k(\beta)|}, & x^k \in \mathcal{T}^k(\beta) \\ 0, & x^k \notin \mathcal{T}^k(\beta). \end{cases} \tag{91}$$

For a k-letter input X^k drawn from (91), let $I(X^k;\tilde{Y}^k)$ be the corresponding mutual information at the output of the receiver (i.e. the output of the BSC in (3)). The following achievability result holds.

Lemma 1. *For the BSC in* (3)*, and for any $\epsilon' > 0$, there exists a sequence of (M_n, n, α_n) constant-composition codes with vanishing decoding error probability, and for which*

$$\lim_{n\to\infty} \frac{1}{n} \log M_n \geq \frac{1}{k} I(X^k;\tilde{Y}^k) - \epsilon' \quad \textit{and} \quad \lim_{n\to\infty} \alpha_n = \beta. \tag{92}$$

Proof. The proof relies on defining an extended channel that takes k-letter inputs and produces k-letter outputs (see, e.g., [15, Ch. 4.3]). Consider an extended channel which takes X^k as input and produces $\tilde{Y}^k$ as output. This is clearly a discrete memoryless channel. The random coding argument in the achievability of the the channel coding theorem implies that for any R such that $kR < I(X^k;\tilde{Y}^k)$ bits/extended channel use, there exists a sequence of (L_j, j) codes with $L_j = \lceil 2^{jkR} \rceil$, where $j \in \mathbb{N}$, with a vanishing decoding error probability as $j \to \infty$ (see, e.g., [16]). Moreover, from the structure of the input distribution in (91), it is evident that all binary codewords in this sequence of codes have the same composition β.

An (M_n, n) code for the original channel, where $n \in \mathbb{N}$, can be constructed from a corresponding (L_j, j) code for the extended channel by setting $j = \lfloor \frac{n}{k} \rfloor$. The remaining $n - k\lfloor \frac{n}{k} \rfloor$ symbols of each codeword in the new code are set to 0. All codewords in this code will have a composition of

$$\alpha_n = \frac{1}{n} \left\lfloor \frac{n}{k} \right\rfloor k\beta. \tag{93}$$

It evident that a sequence of (M_n, n, α_n) constant-composition codes constructed in the above fashion for the BSC in (3) will have vanishing decoding error probability as $n \to \infty$. Moreover, we have

$$\lim_{n\to\infty} \frac{1}{n} \log M_n = \lim_{n\to\infty} \frac{1}{n} \log L_{\lfloor \frac{n}{k} \rfloor} = R. \tag{94}$$

This holds since $\frac{1}{n} \log L_{\lfloor \frac{n}{k} \rfloor}$, where $L_{\lfloor \frac{n}{k} \rfloor} = \left\lceil 2^{\lfloor \frac{n}{k} \rfloor kR} \right\rceil$, is bounded as follows:

$$\left(\frac{n-k}{n} \right) R \leq \frac{1}{n} \log L_{\lfloor \frac{n}{k} \rfloor} \leq R + \frac{1}{n} \tag{95}$$

with both bounds converging to R as $n \to \infty$ for any finite k. Finally, it is clear from (93) that $\alpha_n \to \beta$ as $n \to \infty$. This completes the proof. □

Note that have introduced the multi-letter characterisation in Lemma 1 in order to obtain an achievability result for constant-composition codes using a standard random coding argument. We now proceed to characterize the limiting behaviour of $\frac{1}{k} I(X^k;\tilde{Y}^k)$ as k grow large. To this end, we define a sequence of compositions $(\beta_k)_{k\in\mathbb{N}}$, where $\beta_k \in [0,1]$ such that $k\beta_k$ is an integer for all $k \in \mathbb{N}$, and $\lim_{k\to\infty} \beta_k = \alpha$. This sequence of compositions yields a sequence of multi-letter input distributions $(P_{X^k})_{k\in\mathbb{N}}$, where each distribution $P_{X^k}(x^k)$ is defined by setting $\beta = \beta_k$ in (91).

Lemma 2. *The limit of $\frac{1}{k} I(X^k;\tilde{Y}^k)$ as $k \to \infty$ exists and is given by*

$$\lim_{k\to\infty} \frac{1}{k} I(X^k;\tilde{Y}^k) = H(\alpha * p) - H(p). \tag{96}$$

Proof. We start by deriving an upper bound for $\frac{1}{k}I(X^k;\tilde{Y}^k)$. This is given as follows:

$$\frac{1}{k}I(X^k;\tilde{Y}^k) = \frac{1}{k}H(\tilde{Y}^k) - \frac{1}{k}H(\tilde{Y}^k|X^k) \tag{97}$$

$$\leq \frac{1}{k}\sum_{i=1}^{k} H(\tilde{Y}_i) - H(p) \tag{98}$$

$$= \underbrace{H(\beta_k * p) - H(p)}_{\triangleq a_k}. \tag{99}$$

In the above, the inequality follows from the independence bound, while the last equality holds since $\tilde{Y}_i = X_i \oplus \tilde{Z}_i$, with $X_i \sim \text{Bern}(\beta_k)$ and $\tilde{Z}_i \sim \text{Bern}(p)$.[1] Taking the limit of the upper bound, we obtain

$$\lim_{k\to\infty} a_k = \lim_{k\to\infty} H(\beta_k * p) - H(p) \tag{100}$$

$$= H(\alpha * p) - H(p). \tag{101}$$

which holds since binary entropy function $H(\beta * p)$ is continuous in β (see Appendix C). We now move on to deriving a lower bound. To this end, we invoke Mrs. Gerber's lemma [17] (see also [15, Ch. 2.1]).

Mrs. Gerber's Lemma. *Let X^k be a binary random sequence with entropy $H(X^k)$ and $\tilde{Z}^k$ be a sequence of i.i.d.* Bern(p) *random variables. Then the following inequality holds*

$$\frac{1}{k}H\big(X^k \oplus \tilde{Z}^k\big) \geq H\left(H^{-1}\left(\frac{H(X^k)}{k}\right) * p\right) \tag{102}$$

where $H^{-1} : [0,1] \to [0, 0.5]$ is the inverse of the binary entropy function, satisfying $H(H^{-1}(v)) = v$ and $H^{-1}(H(u)) = \min\{u, 1-u\}$, for all $v, u \in [0,1]$.

Equipped with Mrs. Gerber's Lemma, we bound $\frac{1}{k}I(X^k;\tilde{Y}^k)$ below as follows:

$$\frac{1}{k}I(X^k;\tilde{Y}^k) = \frac{1}{k}H(\tilde{Y}^k) - H(p) \tag{103}$$

$$\geq H\left(H^{-1}\left(\frac{H(X^k)}{k}\right) * p\right) - H(p) \tag{104}$$

$$\geq \underbrace{H\left(H^{-1}\left(H(\beta_k) - \frac{\log(k+1)}{k}\right) * p\right) - H(p)}_{\triangleq b_k}. \tag{105}$$

To see that (105) holds, we first note from the lower bound in (90) that $H(X^k)$ is bounded below as

$$H(X^k) = \log|\mathcal{T}^k(\beta_k)| \geq kH(\beta_k) - \log(k+1). \tag{106}$$

The inequality in (105) follows by combining (106) with the fact that $H\left(H^{-1}(v) * p\right)$ is increasing in v for any fixed $p \in (0, 0.5)$ (see Appendix C). Taking the limit of the lower bound in (105), we get

$$\lim_{k\to\infty} b_k = \lim_{k\to\infty} H\left(H^{-1}\left(H(\beta_k) - \frac{\log(k+1)}{k}\right) * p\right) - H(p) \tag{107}$$

$$= H\left(H^{-1}\left(H(\alpha)\right) * p\right) - H(p) \tag{108}$$

$$= H\big(\min\{\alpha, 1-\alpha\} * p\big) - H(p) \tag{109}$$

$$= H\left(\alpha * p\right) - H(p). \tag{110}$$

In the above, (108) follows from the continuity of $H\big(H^{-1}(v) * p\big)$ in v (see Appendix C). Since the limits of the upper bound in (99) and the lower bound in (105) coincide, the proof of Lemma 2 is complete. □

[1] Note that $X_i \sim \text{Bern}(\beta_k)$ as for any $i \in [k]$, we have $X_i = 1$ in a total of $\binom{k-1}{k\beta_k - 1}$ of the $\binom{k}{k\beta_k}$ sequences in $\mathcal{T}^k(\beta_k)$. Since sequences are uniformly distributed in $\mathcal{T}^k(\beta_k)$, we have $\mathbb{P}[X_i = 1] = \binom{k-1}{k\beta_k}/\binom{k}{k\beta_k} = \beta_k$.

The result of Lemma 2 implies that for any $\epsilon'' > 0$ there exists some finite k_1 for which

$$\frac{1}{k} I(X^k; \tilde{Y}^k) \geq H(\alpha * p) - H(p) - \epsilon'', \text{ for all } k \geq k_1. \tag{111}$$

Moreover, since $\beta_k \to \alpha$ as $k \to \infty$, then for any $\delta > 0$ there exists a finite k_2 for which

$$\beta_k \geq \alpha - \delta, \text{ for all } k \geq k_2. \tag{112}$$

Taking $k = \max\{k_1, k_2\}$ and using Lemma 1, it follows that there exists a sequence of (M_n, n, α_n) codes with vanishing decoding error probability and for which

$$\lim_{n\to\infty} \frac{1}{n} \log M_n \geq H(\alpha * p) - H(p) - (\epsilon' + \epsilon'') \quad \text{and} \quad \lim_{n\to\infty} \alpha_n \geq \alpha - \delta. \tag{113}$$

This completes the proof of Theorem 3.

B Proof of Theorem 4

We start with a few useful definitions. For any positive integer k, a k-ball of radius $r \geq 0$ is defined as

$$\mathcal{B}_k(r) \triangleq \left\{ x^k \in \mathbb{R}^k : \sqrt{k\mathcal{P}(x^k)} \leq r \right\}. \tag{114}$$

The volume of $\mathcal{B}_k(r)$ is given by

$$V\big(\mathcal{B}_k(r)\big) \triangleq \int_{x^k \in \mathcal{B}_k(r)} \mathrm{d}x_1 \cdots \mathrm{d}x_k = \frac{\big(\pi r^2\big)^{\frac{k}{2}}}{\Gamma\left(\frac{k}{2}+1\right)} \tag{115}$$

where $\Gamma\left(\frac{k}{2}+1\right)$ is the Euler gamma function, which is equal to

$$\Gamma\left(\frac{k}{2}+1\right) = (1+\gamma_k)\sqrt{\pi k}\left(\frac{k}{2e}\right)^{\frac{k}{2}} \tag{116}$$

from some real constant γ_k which depends on k, and which satisfies $\gamma_k \to 0$ as $k \to \infty$. A k-spherical-shell of outer radius $r \geq 0$ and inner radius ηr, for some $\eta \in [0,1]$, is defined as

$$\mathcal{S}_n(r,\eta) \triangleq \left\{ x^n \in \mathbb{R}^n : \eta r \leq \sqrt{k\mathcal{P}(x^k)} \leq r \right\}. \tag{117}$$

The volume of $\mathcal{S}_n(r,\eta)$ is equal to the volume of $\mathcal{B}_k(r) \setminus \mathcal{B}_k(\eta r)$, and it is hence given by

$$V\big(\mathcal{S}_n(r,\eta)\big) = V\big(\mathcal{B}_n(r)\big) - V\big(\mathcal{B}_n(\eta r)\big) \tag{118}$$

$$= \frac{\big(\pi r^2\big)^{\frac{k}{2}}}{\Gamma\left(\frac{k}{2}+1\right)}\big(1-\eta^k\big) \tag{119}$$

$$= \left(2\pi e \frac{r^2}{k}\right)^{\frac{k}{2}} \cdot \left(\frac{1-\eta^k}{1+\gamma_k}\right) \cdot \frac{1}{\sqrt{\pi k}}. \tag{120}$$

We now define the following distribution (or density) on $\mathbb{R}^k$

$$P_{X^k}(x^k) = \begin{cases} \frac{1}{V(\mathcal{S}_k(\sqrt{k\mathsf{P}},\eta))}, & x^n \in \mathcal{S}_k(\sqrt{k\mathsf{P}},\eta) \\ 0, & x^n \notin \mathcal{S}_k(\sqrt{k\mathsf{P}},\eta) \end{cases} \tag{121}$$

where we set $\eta = \sqrt{1-(\delta/\mathsf{P})}$. For a k-letter input X^k drawn from (121), let $I(X^k;\tilde{Y}^k)$ be the corresponding mutual information at the output of the Gaussian channel in (7). The following result holds.

Lemma 3. *For the Gaussian channel in* (7)*, and for any $\epsilon' > 0$, there exists a sequence of $(M_n, n, \mathsf{P}_n, \delta)$ almost constant-power codes with $\mathsf{P}_n \leq \mathsf{P}$ and vanishing decoding error probability, and for which*

$$\lim_{n\to\infty} \frac{1}{n} \log M_n \geq \frac{1}{k} I(X^k; \tilde{Y}^k) - \epsilon' \quad \textit{and} \quad \lim_{n\to\infty} \mathsf{P}_n = \mathsf{P}. \tag{122}$$

Proof. As in the proof of Lemma 1, we consider a k-letter extension of the Gaussian channel in (7). From the achievability proof of the channel coding theorem for Gaussian channels (see, e.g., [16]), it follows that for any rate $kR < I(X^k; \tilde{Y}^k)$ bits/extended channel use, there exists a sequence of (L_j, j) codes with $L_j = \left\lceil 2^{jkR} \right\rceil$ for all $j \in \mathbb{N}$ and vanishing decoding error probability. Moreover, the structure of the k-letter input distribution in (121) guarantees that all extended codewords in this sequence of codes have an almost fixed average power, which is at least $\eta^2\mathsf{P} = \mathsf{P} - \delta$ and at most P.

We construct an (M_n, n) code for the original channel from an (L_j, j) code for the extended channel by choosing $j = \left\lfloor \frac{n}{k} \right\rfloor$, and setting the remaining $n - k\left\lfloor \frac{n}{k} \right\rfloor$ symbols of each codeword in the new code to 0. Any codeword x^n in this new code will have an average power satisfying

$$\frac{1}{n}\left\lfloor \frac{n}{k} \right\rfloor k(\mathsf{P} - \delta) \leq \mathcal{P}(x^n) \leq \frac{1}{n}\left\lfloor \frac{n}{k} \right\rfloor k\mathsf{P}. \tag{123}$$

Defining $\mathsf{P}_n \triangleq \frac{1}{n}\left\lfloor \frac{n}{k} \right\rfloor k\mathsf{P}$, it is evident that $\mathsf{P}_n \leq \mathsf{P}$, and hence the average power constraint in (9) is not violated. (123) implies that $\mathsf{P}_n - \delta \leq \mathcal{P}(x^n) \leq \mathsf{P}_n$ for every codeword x^n in the new code, and hence we have an $(M_n, n, \mathsf{P}_n, \delta)$ almost constant-power code. A sequence of $(M_n, n, \mathsf{P}_n, \delta)$ codes constructed in the above fashion for the Gaussian channel in (7) will have vanishing decoding error probability and $\mathsf{P}_n \to \mathsf{P}$ as $n \to \infty$. Moreover, $\frac{1}{n}\log M_n \to R$ as $n \to \infty$, which holds by the same argument used in (94). □

We now characterize the limiting behaviour of $\frac{1}{k} I(X^k; \tilde{Y}^k)$ as $k \to \infty$. To this end, we consider a sequence $(P_{X^k})_{k\in\mathbb{N}}$ of multi-letter densities, where each density $P_{X^k}(x^k)$ is defined as in (121). We have the following result.

Lemma 4. *The limit of $\frac{1}{k} I(X^k; \tilde{Y}^k)$ as $k \to \infty$ exists as is given by*

$$\lim_{k\to\infty} \frac{1}{k} I(X^k; \tilde{Y}^k) = \frac{1}{2} \log\left(1 + \frac{\mathsf{P}}{\tilde{\sigma}^2}\right). \tag{124}$$

Proof. An upper bound for $\frac{1}{k} I(X^k; \tilde{Y}^k)$ is derived in a standard manner as:

$$\begin{aligned}
\frac{1}{k} I(X^k; \tilde{Y}^k) &= \frac{1}{k} h(\tilde{Y}^k) - \frac{1}{k} h(\tilde{Y}^k | X^k) && (125)\\
&\leq \frac{1}{k} \sum_{i=1}^{k} h(\tilde{Y}_i) - \frac{1}{2} \log\left(2\pi e \tilde{\sigma}^2\right) && (126)\\
&\leq \frac{1}{k} \sum_{i=1}^{k} \frac{1}{2} \log\left(2\pi e(\mathbb{E}[X_i^2] + \tilde{\sigma}^2)\right) - \frac{1}{2} \log\left(2\pi e \tilde{\sigma}^2\right) && (127)\\
&\leq \frac{1}{2} \log\left(1 + \frac{1}{k} \sum_{i=1}^{k} \frac{\mathbb{E}[X_i^2]}{\tilde{\sigma}^2}\right) && (128)\\
&\leq \frac{1}{2} \log\left(1 + \frac{\mathsf{P}}{\tilde{\sigma}^2}\right). && (129)
\end{aligned}$$

In the above, (127) holds since $\tilde{Y}_i$ has an expected power of $E[X_i^2] + \tilde{\sigma}^2$, and the fact that a Gaussian distribution maximizes the differential entropy under an expected power constraint. (128) follows from Jensen's inequality. Now we turn to deriving a lower bound $\frac{1}{k} I(X^k; \tilde{Y}^k)$. To this end, we will require the entropy power inequality (see, e.g., [16, Theorem 17.7.3]).

Entropy Power Inequality. *Let X^k and $\tilde{Z}^k$ be independent random vectors with probability density functions on $\mathbb{R}^n$. Then the differential entropy of $X^k + \tilde{Z}^k$ is bounded below as*

$$\frac{1}{k}h(X^k + \tilde{Z}^k) \geq \frac{1}{2}\log\left(2^{2h(X^k)/k} + 2^{2h(\tilde{Z}^k)/k}\right). \tag{130}$$

From using the entropy power inequality, we obtain the following lower bound

$$\begin{aligned}
\frac{1}{k}I(X^k;\tilde{Y}^k) &= \frac{1}{k}h(\tilde{Y}^k) - \frac{1}{2}\log\left(2\pi e\tilde{\sigma}^2\right) && (131)\\
&\geq \frac{1}{2}\log\left(2^{2h(X^k)/k} + 2\pi e\tilde{\sigma}^2\right) - \frac{1}{2}\log\left(2\pi e\tilde{\sigma}^2\right) && (132)\\
&= \frac{1}{2}\log\left(2\pi e\mathsf{P}2^{\frac{2}{k}\log\left(\frac{1-\eta^k}{1+\gamma_k}\right) - \frac{2}{k}\log\sqrt{\pi k}} + 2\pi e\tilde{\sigma}^2\right) - \frac{1}{2}\log\left(2\pi e\tilde{\sigma}^2\right) && (133)\\
&= \frac{1}{2}\log\left(1 + \frac{\mathsf{P}}{\tilde{\sigma}^2}2^{-\frac{2}{k}\log\left(\frac{1+\gamma_k}{1-\eta^k}\cdot\sqrt{\pi k}\right)}\right). && (134)
\end{aligned}$$

In the above, (133) follows from

$$\begin{aligned}
h(X^k) &= \log V(\mathcal{S}_k(\sqrt{k\mathsf{P}},\eta)) && (135)\\
&= \frac{k}{2}\log\left(2\pi e\mathsf{P}\right) + \log\left(\frac{1-\eta^k}{1+\gamma_k}\right) - \log\sqrt{\pi k} && (136)
\end{aligned}$$

Since $\frac{2}{k}\log\left(\frac{1+\gamma_k}{1-\eta^k}\cdot\sqrt{\pi k}\right) \to 0$ as $k\to\infty$, the lower bound in (134) converges to $\frac{1}{2}\log\left(1+\frac{\mathsf{P}}{\tilde{\sigma}^2}\right)$ as $k\to\infty$. Since this matches the upper bound in (129), the proof is complete. □

The result of Lemma 4 implies that for any $\epsilon'' > 0$ there is some finite k for which

$$\frac{1}{k}I(X^k;\tilde{Y}^k) \geq \frac{1}{2}\log\left(1+\frac{\mathsf{P}}{\tilde{\sigma}^2}\right) - \epsilon''. \tag{137}$$

Selecting such k and using Lemma 3, it follows that there exists a sequence of $(M_n, n, \mathsf{P}_n, \delta)$ almost constant-power codes with $\mathsf{P}_n \leq \mathsf{P}$ and vanishing decoding error probability, and for which

$$\lim_{n\to\infty}\frac{1}{n}\log M_n \geq \frac{1}{2}\log\left(1+\frac{\mathsf{P}}{\tilde{\sigma}^2}\right) - (\epsilon'+\epsilon'') \quad\text{and}\quad \lim_{n\to\infty}\mathsf{P}_n = \mathsf{P}. \tag{138}$$

This completes the proof of Theorem 4.

C Binary Entropy Function

Here we state some properties of the binary entropy function and its inverse, which are used in the proofs related to Theorem 1. The binary entropy function $H:[0,1]\to[0,1]$ is defined as

$$H(u) \triangleq -u\log u - (1-u)\log(1-u). \tag{139}$$

The inverse of the binary entropy function $H^{-1}:[0,1]\to[0,0.5]$ satisfies

$$H(H^{-1}(v)) = v \text{ and } H^{-1}(H(u)) = \min\{u, 1-u\}. \tag{140}$$

Monotonicity: $H(u)$ is increasing in $u\in[0,0.5]$, and $H^{-1}(v)$ is increasing in $v\in[0,1]$. For any $p\in(0,0.5)$, the function $\alpha * p \triangleq \alpha(1-p)+p(1-\alpha) = \alpha(1-2p)+p$ is increasing in $\alpha\in[0,1]$. Moreover, by further restricting α to be in $[0,0.5]$, we will have $\alpha * p$ in $[0,0.5]$ as well. It follows that the function $H\left(H^{-1}(v) * p\right)$ is increasing in $v\in[0,1]$, given that $p\in(0,0.5)$.

Continuity: $H(u)$ is continues in $u \in [0,1]$, which can be seen from its differentiability on $(0,1)$. The function $\alpha * p$ is continuous in $\alpha \in [0,1]$, and hence the composition $H(\alpha * p)$ is continuous in α.

$H^{-1}(v)$ can be equivalently defined as the inverse of $H(u)$ with a domain restricted to $[0, 0.5]$. The fact that $H(u)$ is continuous and strictly monotonic on $[0, 0.5]$ implies that its inverse $H^{-1}(v)$ is continuous in its argument $v \in [0,1]$. The function $H\big(H^{-1}(v) * p\big)$ is continuous in $v \in [0,1]$, as it is a composition of $H(\alpha * p)$ and $H^{-1}(v)$, which are both continuous in their respective arguments.

Concavity: It is well known that $H(u)$ is concave in $u \in [0,1]$. Moreover, $\alpha * p$ is affine in α. It follows that $H(\alpha * p)$ is concave in $\alpha \in [0,1]$, as it is the composition of a concave function with an affine function.

References

[1] R. Blahut, "Hypothesis testing and information theory," *IEEE Trans. Inf. Theory*, vol. 20, no. 4, pp. 405–417, 1974.

[2] L. Zheng, M. Lops, Y. C. Eldar, and X. Wang, "Radar and communication coexistence: An overview," *IEEE Signal Process. Magazine*, vol. 36, no. 5, pp. 85–99, 2019.

[3] F. Liu, C. Masouros, A. P. Petropulu, H. Griffiths, and L. Hanzo, "Joint radar and communication design: Applications, state-of-the-art, and the road ahead," *IEEE Trans. Commun.*, vol. 68, no. 6, pp. 3834–3862, 2020.

[4] J. A. Zhang, F. Liu, C. Masouros, R. W. Heath Jr, Z. Feng, L. Zheng, and A. Petropulu, "An overview of signal processing techniques for joint communication and radar sensing," *arXiv:2102.12780*, 2021.

[5] F. Liu, Y. Cui, C. Masouros, J. Xu, T. X. Han, Y. C. Eldar, and S. Buzzi, "Integrated sensing and communications: Towards dual-functional wireless networks for 6G and beyond," *arXiv:2108.07165*, 2021.

[6] M. Kobayashi, G. Caire, and G. Kramer, "Joint state sensing and communication: Optimal tradeoff for a memoryless case," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, 2018, pp. 111–115.

[7] M. Kobayashi, H. Hamad, G. Kramer, and G. Caire, "Joint state sensing and communication over memoryless multiple access channels," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, 2019, pp. 270–274.

[8] M. Ahmadipour, M. Wigger, and M. Kobayashi, "Joint sensing and communication over memoryless broadcast channels," in *Proc. IEEE Inf. Theory Workshop (ITW)*, 2021, pp. 1–5.

[9] M. Ahmadipour, M. Kobayashi, M. Wigger, and G. Caire, "An information-theoretic approach to joint sensing and communication," *arXiv:2107.14264*, 2021.

[10] R. G. Gallager, *Information theory and reliable communication.* New York: John Wiley & Sons, Inc., 1968.

[11] A. J. Viterbi and J. K. Omura, *Principles of digital communication and coding.* New York: McGraw-Hill, Inc., 1979.

[12] I. Csiszár and J. Körner, *Information theory: coding theorems for discrete memoryless systems*, 2nd ed. Cambridge University Press, 2011.

[13] C. E. Shannon, "Probability of error for optimal codes in a Gaussian channel," *Bell Syst. Tech. J.*, vol. 38, no. 3, pp. 611–656, 1959.

[14] C. E. Shannon, R. G. Gallager, and E. R. Berlekamp, "Lower bounds to error probability for coding on discrete memoryless channels. I," *Information and Control*, vol. 10, no. 1, pp. 65–103, 1967.

[15] A. El Gamal and Y.-H. Kim, *Network information theory*. Cambridge University Press, 2011.

[16] T. M. Cover and J. A. Thomas, *Elements of information theory*, 2nd ed. New York: John Wiley & Sons, Inc., 2006.

[17] A. Wyner and J. Ziv, "A theorem on the entropy of certain binary sequences and applications–I," *IEEE Trans. Inf. Theory*, vol. 19, no. 6, pp. 769–772, 1973.