Gaussian Intersymbol Interference Channels With Mismatch

Wasim Huleihel Salman Salamatian Neri Merhav Muriel Médard

Abstract

This paper considers the problem of channel coding over Gaussian intersymbol interference (ISI) channels with a given metric decoding rule. Specifically, it is assumed that the mismatched decoder has an incorrect assumption on the impulse response function. The mismatch capacity is the highest achievable rate for a given decoding rule. Existing lower bounds to the mismatch capacity for channels and decoding metrics with memory (as in our model) are presented only in the form of multi-letter expressions that have not been calculated in practice. Consequently, they provide little insight on the mismatch problem. In this paper, we derive computable single-letter lower bounds to the mismatch capacity, and discuss some implications of our results. Our achievable rates are based on two ensembles; the ensemble of codewords generated by an autoregressive process, and the ensemble of codewords drawn uniformly over a "type class" of real-valued sequences. Computation of our achievable rates demonstrates non-trivial behavior of the achievable rates as a function of the mismatched parameters. As a simple application of our technique, we derive also the random coding exponent associated with a mismatched decoder which assumes that there is no ISI at all. Finally, we compare our results with universal decoders which are designed *outside* the true class of channels that we consider in this paper.

I. Introduction

The mismatch capacity is the highest achievable rate for a given, possibly suboptimal, decoding rule. This scenario arises naturally when, due to imprecise channel measurement, the receiver performs maximum-likelihood decoding with respect to the wrong channel law, or when the receiver is intentionally designed to perform a suboptimal decoding rule due to implementation constraints. This problem has

W. Huleihel, S. Salamatian, and M. Médard are with the Research Laboratory of Electronics at the Massachusetts Institute of Technology, Cambridge, MA (e-mail: {wasimh,salmansa,medard}@mit.edu). The work of W. Huleihel was supported by the MIT-Technion Postdoctoral Fellowship.

N. Merhav is with the Andrew & Erna Viterbi Faculty of Electrical Engineering at the Technion-Israel Institute of Technology, Haifa 3200003, Israel (e-mail: merhav@ee.technion.ac.il).

been studied extensively, see e.g., [1-4] and many references therein. Finding a single-letter expression for the mismatch capacity is a long-standing open problem.

Most of the existing work on the mismatch capacity has focused on deriving achievable rates using random coding arguments for memoryless channels and decoding metrics. For a given block length, one typically selects a certain ensemble of rate–R codes and then studies the highest achievable rate for which the average probability of error still tends to zero as the block length tends to infinity. Different random coding ensembles yield different lower bounds to the mismatch capacity. For example, the ensemble of identically and independently distributed (i.i.d.) codewords leads to the generalized mutual information (GMI) rate [5-7]. Tighter lower bounds to the mismatch capacity can be derived using constant-composition ensembles [8, 9], and cost-constrained ensembles [3, 10]. Although the GMI is the weakest bound of this class, it has the advantage of being applicable also to channels over infinite alphabets, as its derivation relies on Gallager's bounding technique [11] rather than on the method of types. While superior to the GMI, the bound based on the constant-composition ensemble, relies heavily on the method of types[1] [12] and thus, at least at first glance, limited to channels over finite alphabets. See, however, [1] for some extensions to memoryless channels of an exponential type and to some channels with memory. In [3], this bound was also extended to general alphabets using an alternative derivation that does not require the method of types. In [13], the question of finding the best mismatched decoder (in the sense of maximizing the achievable rate) over a given family of linear decoders was considered, along with an efficient algorithm for computing this decoder. Finally, [10] considered a more comprehensive analysis of the random-coding error probability under various ensembles, including error exponents, second-order coding rates [14, 15], and refined asymptotic results based on the saddlepoint approximation [16]. In the discrete case, the results of [10] are tight in the error exponent sense, but for general alphabets there is no guarantee for ensemble tightness, as the analysis is based on Gallager's bounding technique.

For channels and decoding rules with memory, however, there are no known single-letter lower bounds, even in specific examples. The only existing lower bound, derived in [3], which holds for a general family of channels and decoding metrics with memory, appears in the form of a multi-letter expression. Unfortunately, this expression cannot be calculated in practice and it provides only little insight on the mismatch decoding problem.

Motivated by the last paragraph, in this work, we consider a specific class of channels with memory; Gaussian intersymbol interference (ISI) channels, with a mismatched decoding metric that is based upon wrong ISI coefficients (see Section III for a precise definition of our model). Considering this problem is important when, for example, the depth of the ISI is large (i.e, many taps), and thus the implementation of the optimal maximum-likelihood (ML) decoder is complicated. In such cases, one might want to

[1]More critically, the method of types is of limited applicability to channels with memory, rendering the bound inapplicable to such channels.

intentionally limit the depth of the assumed ISI in decoding metric so that to keep the decoding complexity within reasonable limits. Another possible motivation is that when the channel is slowly time-varying (e.g., fading), it might limit the block length one can work with (block length within which the channel is nearly fixed), and then estimation errors resulting from channel estimation can be significant. In such cases it is interesting to understand how this issue affects the achievable rates. As was demonstrated in [17], even for discrete memoryless channels (DMCs) and memoryless decoding metrics, the ensemble of i.i.d. input codewords is not optimal, and an improved bound on the mismatch capacity of the DMC can be obtained through a random coding argument applied to a superalphabet, or equivalently, inputs defined over product spaces (i.e., inputs with memory).

We consider two random coding ensembles. In the first, the random codewords are generated by an autoregressive (see eq. (11) for more details). For this ensemble, we derive a simple and computable single-letter lower bound to the mismatch capacity. The obtained rate is ensemble-tight, namely, it captures the exact maximum achievable rate for which the ensemble average error probability vanishes. Also, contrary to the above-mentioned multi-letter expressions, in the Gaussian ISI case, our achievable-rate formula is given in terms of frequency-domain integrals of certain spectral quantities, which are computable at least numerically. The main technical contribution in the derivation is a novel procedure to assess the exponential behavior of the error probability, using the saddle-point integration method (see, e.g., [18]). Specifically, as shall be seen in the proof of our main results, the probability of error associated with our mismatched decoder can be written as a function of the volumes (Lebesgue measure) of some "conditional typical set" of sequences with continuous-valued components. This typical set, of some input sequence $(x_1, x_2, \ldots, x_n)$, given an output sequence $(y_1, y_2, \ldots, y_n)$, will contain all sequences which, within $\epsilon > 0$, have the same sufficient statistics as $(x_1, x_2, \ldots, x_n)$ induced by our mismatched decoding rule. Accordingly, to analyze the probability of error we need to analyze the volume of this typical set. While this was also the main difficulty of [19, 20], and resolved using the "backward channel" technique, here, we use the saddle-point integration method which is, more direct, and simplifies the derivations significantly. Since we deal with a Gaussian channel, the above mentioned typical set depends on the input sequence only through certain simple statistics, such as, the correlation with the output sequence and auto-correlations (up to some order), and thus it is possible to get "single-letter" expressions, as opposed to general channels with (finite) memory, where only multi-letter formulas are available.

Then, using the same methods, we analyze also the ensemble of codewords which drawn according to the uniform distribution within a "type class" of real-valued sequences [19, 20] (see eq. (13) for more details). As before, for this ensemble we derive a computable single-letter lower bound to the mismatch capacity. The resulting formula is more complicated to compute compared to the previous ensemble. However, the fixed composition ensemble can be better than the autoregressive ensemble. As an illustrating example, consider the simple case where both the true channel and the decoding metric are memoryless. Specifically, the channel is given by $y_t = x_t + w_t$, for $t = 1, 2, \ldots, n$, where $\{w_t\}$

is a white Gaussian noise, while the mismatched decoder computes a ML estimate which correspond to the channel $\bar{y}_t = \alpha x_t + w_t$, for $t = 1, 2, \ldots, n$, and $\alpha > 0$ designates the mismatched parameter. It should be clear then that codewords that are drawn on the entire hypersurface of radius $\sqrt{nP_X}$ achieve the matched capacity irrespectively of the value of α.[2] This is no longer true if one generates codewords from the autoregressive codebook (in particular, i.i.d. codewords cannot achieve capacity if there is a mismatch, even though the true channel and the decoding metric are memoryless). This is illustrated in Fig. 1. In general scenarios (e.g., Fig. 2), we found that there is no special order between the autoregressive codebook and the above "fixed composition" ensemble, in terms of the achievable rates, namely, no ensemble is uniformly better than the other. Nevertheless, it seems that in most cases the fixed composition ensemble is more powerful as a function of the mismatched parameters.

It turns out that a byproduct of our analysis is an ensemble-tight characterization of the random coding error exponent. Exponentially tight analysis of the average probability of error was extensively studied before (see, e.g., [21-24]) mainly for discrete memoryless sources and channels. Here, on the other hand, as we deal with sources and channels with memory defined over infinite alphabets, the same methods cannot be applied. Specifically, to assess the exact exponential rate of the average error probability, we need to evaluate the log-volumes of some conditional typical sets of sequences with continuous-valued components. While this was also the main core of [19, 20], here, the saddle-point integration method simplifies the analysis considerably. Accordingly, to demonstrate the usefulness of our techniques, we consider the ensemble of codewords drawn on the entire hypersurface of radius $\sqrt{nP_X}$, and derive the *exact* random coding error exponent in the case of a memoryless decoding metric.

Finally, we consider also the problem of universal decoding which received very much attention in the last four decades [12, 19, 20, 25-36]. Indeed, as in the mismatch decoding problem, in many practical situations encountered in coded communication systems, the specific channel over which transmission is to be carried out is unknown to the receiver. The receiver only knows that the channel belongs to a given family of channels. In such a case, the implementation of the optimum ML decoder is of course precluded, and thus, universal decoders, independent of the unknown channel, and which perform asymptotically as well as the ML decoder had the channel law been known, are sought. In this paper, we look at the following scenario. Consider the Gaussian ISI channel, and assume that due to complexity issues concerning the implementation of the optimal ML decoder, the receiver intentionally uses a mismatched decoder which corresponds to a memoryless channel (namely, without ISI). Nonetheless, we allow the receiver to optimize his memoryless metric, namely, it can be a function of the true channel. Now, consider a different receiver which uses a universal decoder which is designed for a memoryless channel. In other words, the true family of channels is *outside* the class of channels for which the universal decoder is actually designed, i.e, mismatched universal decoder. Then, which approach yields higher rates? We

[2] This can be easily seen by expanding the mismatched decoding rule along with the fact that the different codewords have the same energy, and comparing to the maximum-liklihood decoder.

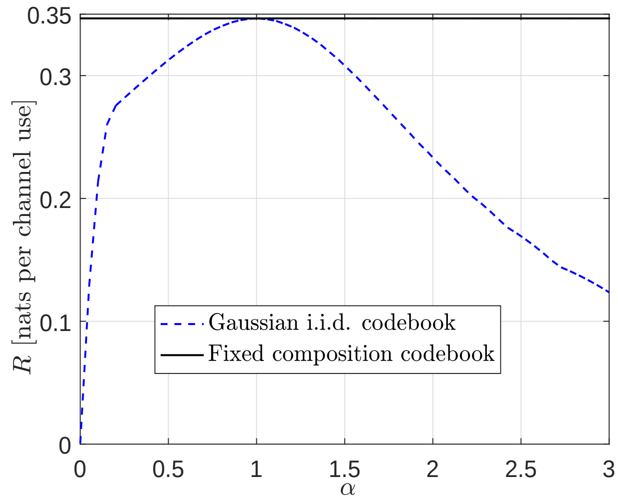

Fig. 1. Achievable rate as a function of the mismatched level α, for the additive white Gaussian noise channel (AWGN), using Gaussian i.i.d. codebook and fixed composition ensemble (i.e., codewords are drawn uniformly at random over the n-dimensional hypersphere of radius $\sqrt{nP_X}$), where $P_X = 1$ and $\sigma^2 = 1$. The fixed composition ensemble achieve the capacity of the AWGN, i.e., $1/2 \log 2$, regardless of the value of $\alpha > 0$.

show that both decoders achieve the same rates, and in fact achieve also the same error exponent. This means that, at least in the specific scenario described above, our results provide indications that universal decoders exhibit a robustness property with respect to (w.r.t.) the family of channels over which they are actually designed. In other words, this observation (potentially) suggests a certain expansion of the classic notion of universality to cases where the true underlying channel is outside the class.

The paper is organized as follows. In Section II we establish some notation. Then, in Section III, we present our system model and formalize the problem. In Section IV we assert our main results. Specifically, we first provide achievable rates under the autoregressive random coding ensemble and the fixed composition ensemble, respectively. Then, we consider the problem of mismatch universal decoding. Section V is devoted to the proofs of our main results. Finally, our conclusions appear in Section VI.

II. Notation Conventions

Throughout this paper, scalar random variables (RV's) will be denoted by capital letters, their sample values will be denoted by the respective lower case letters and their alphabets will be denoted by the respective calligraphic letters. A similar convention will apply to random vectors and matrices and their sample values, which will be denoted with same symbols in the bold face font. The expectation operator of a RV X will be denoted by $\mathbb{E}(X)$. When using vectors and matrices in a linear-algebraic format, n-dimensional vectors, like $\boldsymbol{x}$, will be understood as column vectors, the operators $(\cdot)^T$ and $(\cdot)^H$ will denote vector or matrix transposition and vector or matrix conjugate transposition, respectively, and so, $\boldsymbol{X}^T$

would be a row vector. The ℓ_2-norm of a vector $\boldsymbol{x}$ is denoted by $\|\boldsymbol{x}\|_2$. For two positive sequences $\{a_n\}$ and $\{b_n\}$, the notation $a_n \doteq b_n$ means equivalence in the exponential order, i.e., $\lim_{n\to\infty} \frac{1}{n}\log(a_n/b_n) = 0$, where in this paper, logarithms are defined w.r.t. the natural basis, that is, $\log(\cdot) = \ln(\cdot)$. Given two real numbers a and b, we denote by $[a:b]$ the set of integers $\{n \in \mathbb{N}:\ \lceil a \rceil \leq n \leq \lceil b \rceil\}$, and we let $a \wedge b \triangleq \min(a,b)$ and $a \vee b \triangleq \max(a,b)$. We define $\mathbb{R}_+ \triangleq \{x \in \mathbb{R}:\ x > 0\}$, and $\operatorname{sgn}(\cdot)$ is the sign function, i.e., $\operatorname{sgn}(x) = \frac{\mathrm{d}}{\mathrm{d}x}|x|$, for $x \neq 0$. The volume of a set $\mathcal{A} \subset \mathbb{R}^n$ is defined as $\operatorname{Vol}\{\mathcal{A}\} \triangleq \int_{\mathcal{A}} \mathrm{d}\boldsymbol{x}$. Finally, the indicator function on a set $\mathcal{A}$ will be denoted by $\mathbb{1}\{\mathcal{A}\}$.

III. Problem Setting

Consider a discrete time, K-tap Gaussian ISI channel, characterized by

$$y_t = \sum_{i=0}^{K} h_i x_{t-i} + w_t, \tag{1}$$

for $t = [1:n]$, where $\{x_t\}_{t=1-K}^n \in \mathbb{R}^n$ are the channel inputs, subjected to an average power constraint $\sum_{t=1}^n \mathbb{E}X_t^2 \leq nP_X$, $\{h_i\}_{i=0}^K \in \mathbb{R}^{K+1}$ are the ISI coefficients, $\{w_t\}_{t=1}^n$ is a zero-mean Gaussian white noise with variance σ^2, and $\{y_t\}_{t=1}^n \in \mathbb{R}^n$ are the channel outputs. It is assumed that $\{w_t\}_{t=1}^n$ is statistically independent of $\{x_t\}_{t=1}^n$. We denote by $W(\boldsymbol{y}|\boldsymbol{x})$, the conditional density of the channel output induced by (1), where $\boldsymbol{x} \triangleq (x_{1-K},\ldots,x_n)$ and $\boldsymbol{y} \triangleq (y_1,\ldots,y_n)$, and without loss of generality we assume that $x_{1-K} = \cdots = x_0 = 0$, namely, an overhead of zeroes at the beginning of each block. Alternatively, we may assume that $x_{-k} = x_{n-k}$, for $k = 0,\ldots,K-1$, that is, a circularity assumption on the input sequence [37]. As long as K is fixed and $n \to \infty$ these assumptions have no influence on either the achievable error exponents or the achievable rates. Accordingly, for any $\boldsymbol{x},\boldsymbol{y} \in \mathbb{R}^n$,

$$\log W(\boldsymbol{y}|\boldsymbol{x}) = -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{t=1}^n\left(y_t - \sum_{i=0}^K h_i x_{t-i}\right)^2. \tag{2}$$

A rate R block code of size n is a set of $M = e^{nR}$ equiprobable n-dimensional vectors (codewords), $\boldsymbol{x}_i = (x_{i,1},\ldots,x_{i,n}) \in \mathbb{R}^n$, for $1 \leq i \leq M$, to be transmitted over the channel (1). The decoder, upon receiving $\boldsymbol{y} \in \mathbb{R}^n$, estimates the message i of the transmitted codeword as the one that maximizes $\log V(\boldsymbol{y}|\boldsymbol{x}_i)$, henceforth referred as the *decoding metric*. If this decoding metric is not equivalent to that of the ML decoder (2), then we say that the decoder is *mismatched*.

We assume that the metric $V(\cdot|\cdot)$ is equivalent to that of the ML decoder of a Gaussian ISI channel with (possibly) different coefficients. Specifically, for any $\boldsymbol{x},\boldsymbol{y} \in \mathbb{R}^n$, $V(\boldsymbol{y}|\boldsymbol{x})$ is defined as,[3]

$$\log V(\boldsymbol{y}|\boldsymbol{x}) = -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{t=1}^n\left(y_t - \sum_{i=0}^K \alpha_i x_{t-i}\right)^2 \tag{3}$$

[3]Without loss of generality, we assume that the length of the mismatched filter is the same as the length of the true filter, with the understanding the one can always tap the shorter coefficients sequence with zeros.

where $\{\alpha_i\}_{i=0}^K \in \mathbb{R}^{K+1}$ are the mismatched ISI coefficients. In particular, when $\alpha_i = 0$ for $i = 1, \ldots, K$, then the decoder assumes that there is no ISI at all, i.e., the mismatched decoder is equivalent to the optimal ML decoder associated with the additive white Gaussian noise channel (AWGN), namely, $y_t = \alpha_0 x_t + w_t$, for $t = [1 : n]$.

An error is said to have occurred if the estimated index $\hat{i}$ differs from the correct one, i. A rate R is said to be achievable if, for every $\delta > 0$, there exists a sequence of codes $\{\mathfrak{C}_n\}_{n\geq 1}$ indexed by the block length n, with $M \geq e^{n(R-\delta)}$ and vanishing error probability $P_e(\mathfrak{C}_n)$ when decoding with the metric $V(\cdot|\cdot)$. The mismatch capacity, $C_{\text{ISI}}^{\text{Mis}}$ is the supremum of all achievable rates. The goal of this paper is to derive lower bounds to the mismatched capacity $C_{\text{ISI}}^{\text{Mis}}$.

IV. Main Results

In this section, we present and discuss our main results. Specifically, in Subsection IV-A we present achievable rates for the mismatch decoding problem. We start with the autoregressive ensemble, where codewords are drawn from an autoregressive process, and then we move forward to analyzing the fixed composition ensemble. Following these results, in Subsection IV-B, we consider the problem of mismatched universal decoding, as described in the Introduction.

A. Mismatched Achievable Rates

We establish first some notation. Let p be a non-negative integer, define $\mathcal{P}$ as the set of all vectors $\boldsymbol{\varphi} = (\varphi_1, \ldots, \varphi_p) \in \mathbb{R}^p$, such that all roots of the polynomial $z^p - \sum_{i=1}^p \varphi_i z^{p-i}$ lie strictly within the unit circle, and let $\varphi_0 = 0$. We let $A(\nu)$, for $\nu \in [0, 2\pi]$, be the Fourier transform of the sequence $\{\alpha_k\}_{k=0}^K$, i.e.,

$$A(\nu) \triangleq \sum_{k=0}^{K} \alpha_k e^{-jk\nu}, \quad \nu \in [0, 2\pi] \tag{4}$$

where $j \triangleq \sqrt{-1}$. Similarly, $\Phi(\cdot)$ and $H(\cdot)$ are the Fourier transforms of $\{\varphi_k\}_{k=0}^p$ and $\{h_k\}_{k=0}^K$, respectively. Next, define

$$\eta^2 \triangleq P_X \cdot \left[\frac{1}{2\pi}\int_0^{2\pi} \frac{\mathrm{d}\nu}{|1-\Phi(\nu)|^2}\right]^{-1}, \tag{5}$$

and note that η^2 depends on the choice of $\boldsymbol{\varphi}$. Let $\{\gamma_m\}_{m=0}^p$, with $\gamma_{-m} = \gamma_m$, for $m \in [1 : p]$, be defined as:

$$\gamma_m = \sum_{k=1}^{p} \varphi_k \gamma_{m-k} + \eta^2 \delta_m, \tag{6}$$

where δ_m is the Kronecker delta function. For $\nu \in [0, 2\pi]$, define:

$$S_X(\nu) \triangleq \frac{\eta^2}{|1-\Phi(\nu)|^2}, \tag{7}$$

$$S_Y(\nu) \triangleq |H(\nu)|^2 \cdot S_X(\nu) + \sigma^2, \tag{8}$$

namely, the input and output spectra, respectively, and for $\omega \in \mathbb{R}_+$,

$$f_\omega(\nu) \triangleq \frac{\omega}{2}|A(\nu)|^2 + \frac{1}{2\eta^2}\left[1 + |\Phi(\nu)|^2 - 2 \cdot \mathrm{Re}(\Phi(\nu))\right]. \tag{9}$$

Finally, define

$$\begin{aligned}\bar{I}_1(\boldsymbol{\alpha}, \boldsymbol{\varphi}) \triangleq& \frac{1}{2}\log(2\eta^2) - \min_{\omega \in \mathbb{R}_+}\left\{ -\frac{1}{4\pi}\int_0^{2\pi} \mathrm{d}\nu \log f_\omega(\nu) + \frac{\omega^2}{8\pi}\int_0^{2\pi} \mathrm{d}\nu \frac{S_Y(\nu)\left|A(\nu)\right|^2}{f_\omega(\nu)}\right.\\ &\left. -\omega\left[\sum_{l=0}^{K}\sum_{i=0}^{K}\alpha_l h_i \gamma_{l-i} - \left(\frac{1}{2}\left\|\boldsymbol{\alpha}\right\|_2^2\gamma_0 + \sum_{l=1}^{K}\sum_{k=0}^{K-l}\alpha_k\alpha_{k+l}\gamma_l\right)\right]\right\}\end{aligned} \tag{10}$$

where in case that $p < K$, γ_k for $k = p+1, \ldots, K$, are calculated using (6). It is a simple exercise to check that the minimization problem in (10) is convex. We are now ready to state our main result, a lower bound to the mismatch capacity associated with the system model described in Section III. The proof of the following result is given in Section V.

Theorem 1 Consider the Gaussian ISI channel model in (1), and the mismatched decoding metric in (3). Then, $C_{\mathrm{ISI}}^{\mathrm{Mis}} > \max_{\boldsymbol{\varphi} \in \mathcal{P}} \bar{I}_1(\boldsymbol{\alpha}, \boldsymbol{\varphi})$, where $\bar{I}_1(\boldsymbol{\alpha}, \boldsymbol{\varphi})$ is given in (10).

As was mentioned in the Introduction, previous works on the mismatch capacity focused mainly on standard random coding ensembles, where each codeword is independently and identically generated according to some given probability distribution. However, since the channel has memory, it is reasonable to consider ensembles over which there is a correlation between the symbols within each codeword. To achieve $\bar{I}_1(\boldsymbol{\alpha}, \boldsymbol{\varphi})$, for a given $\boldsymbol{\varphi}$, the codebook $\mathfrak{C}_n$ is generated as follows: For each message $i \in [1 : e^{nR}]$, we generate (independently) the sequence $\{X_t\}_{t=1}^n$ according to

$$X_t = \begin{cases} \sum_{i=1}^{p} \varphi_i \cdot X_{t-i} + \eta Z_t, & t \geq 1 \\ 0, & t < 1 \end{cases} \tag{11}$$

where $\{Z_t\}_t$ is white noise, η^2 is chosen such that $\mathbb{E}X_t^2 = P_X$, for $t = [1 : n]$, and thus it is given in (5). Since we assume that the roots of the polynomial $z^p - \sum_{i=1}^{p}\varphi_i z^{p-i}$ lie strictly within the unit circle, the above process is wide-sense stationary. For example, if $p = 1$, we get

$$X_t = \varphi_1 \cdot X_{t-1} + \sqrt{P_X\,(1 - \varphi_1^2)}Z_t, \tag{12}$$

for $t \geq 1$.

The role of $\boldsymbol{\varphi}$ is to shape the spectrum of the input process so as to mitigate the (undesired) effects of the mismatch decoding. Note, however, that any choice of $\{\varphi_m\}_{m=1}^p \in \mathcal{P}$ would result in a legitimate lower bound to the mismatch capacity. Generally speaking, our result can be interpreted as follows: the first term in (10) is associated with the differential entropy of the input process, and the second term corresponds to a certain conditional entropy of the input given the incorrectly processed output. Finally, we mention [7, Th. 2], where the Gaussian i.i.d. ensemble was studied using a different approach. It can be shown that, when specialized to the i.i.d. case (i.e., $\varphi_i = 0$ for all $i \in [1:p]$), our result in Theorem 1 coincides with [7, Th. 2].

We now study the fixed composition ensemble, where codewords are drawn uniformly at random within the a type class of real-valued sequences. Specifically, fix an arbitrary $\varepsilon > 0$, and pick $p \in \mathbb{N}$. Let Γ denote the set of all vectors $\boldsymbol{\gamma} = (\gamma_0, \gamma_1, \ldots, \gamma_p) \in \mathbb{R}^{p+1}$ with $\gamma_0 \triangleq P_X$, such that the matrix $\left\{\gamma_{|i-j|}\right\}_{i,j}$ is a positive-definite Toeplitz matrix. Also, let $\gamma_{-k} = \gamma_k$, for $k = [1:p]$. Define the sequence of sets $\mathcal{T}_\varepsilon^n(\boldsymbol{\gamma})$, for $n = 1, 2, \ldots$, as follows

$$\mathcal{T}_\varepsilon^n(\boldsymbol{\gamma}) \triangleq \left\{ \boldsymbol{x} \in \mathbb{R}^n : \left| \frac{1}{n} \sum_{t=1}^{n} x_t x_{t-k} - \gamma_k \right| < \varepsilon,\ k = [0:p] \right\}. \tag{13}$$

The codebook $\mathfrak{C}_n$ is generated by drawing M codewords independently and uniformly at random from $\mathcal{T}_\varepsilon^n(\boldsymbol{\gamma})$. The role of $\boldsymbol{\gamma}$ is to shape the spectrum of the input process, and accordingly, these parameters can be optimized. To state our main result we need some additional definitions. Let $\Pi_m(\boldsymbol{\alpha}) \triangleq \sum_{k=0}^{K-m} \alpha_k \alpha_{k+m}$, for $m \in [1:K]$, and

$$\mathcal{W} \triangleq \left\{ \{\omega_k\}_{k=0}^{p+1} \in \mathbb{R}_+^{p+1} :\ \omega_0 \geq \sum_{i=1}^{p} \omega_i + \omega_{p+1} \cdot \sum_{m=1+p\wedge K}^{K} |\Pi_m(\boldsymbol{\alpha})| \right\}. \tag{14}$$

For $\nu \in [0, 2\pi]$, and $\boldsymbol{\omega} \in \mathcal{W}$, define

$$g_{\boldsymbol{\omega}}(\nu) \triangleq \sum_{k=0}^{p} \omega_k \cos(k\nu) + \omega_{p+1} \cdot \sum_{k=1+p\wedge K}^{K} \Pi_k(\boldsymbol{\alpha}) \cos(k\nu), \tag{15}$$

and

$$\begin{aligned} \bar{I}_2(\boldsymbol{\alpha}, \boldsymbol{\gamma}) \triangleq \frac{1}{2} \log(2e\eta^2) - \min_{\boldsymbol{\omega} \in \mathcal{W}} \Bigg\{ & -\frac{1}{4\pi} \int_0^{2\pi} \log g_{\boldsymbol{\omega}}(\nu) d\nu + \frac{\omega_{p+1}^2}{8\pi} \int_{[0,2\pi]} \mathrm{d}\nu \frac{S_Y(\nu) \left|A(\nu)\right|^2}{g_{\boldsymbol{\omega}}(\nu)} \\ & -\omega_{p+1} \left[\sum_{k=0}^{K} \sum_{i=0}^{K} \alpha_k h_i \gamma_{k-i} - \sum_{l=1+p\wedge K}^{K} \Pi_l(\boldsymbol{\alpha}) \gamma_l \right] + \sum_{k=0}^{p} \omega_k \gamma_k \Bigg\} \end{aligned} \tag{16}$$

where η^2 above is calculated by first solving the Yule-Walker equations in (6) to find $\boldsymbol{\varphi}$, and then substituting in (5). Also, as before, when $p < K$, γ_k for $k = p+1, \ldots, K$, are calculated using (6). It can be shown that the minimization problem in (10) is convex. The following result is proved in Appendix B.

Theorem 2 Consider the Gaussian ISI channel model in (1), and the mismatched decoding metric in (3). Then, $C_{\text{ISI}}^{\text{Mis}} > \max_{\boldsymbol{\gamma} \in \Gamma} \bar{I}_2(\boldsymbol{\alpha}, \boldsymbol{\gamma})$, where $\bar{I}_2(\boldsymbol{\alpha}, \boldsymbol{\gamma})$ is given in (16).

Comparing the achievable rates in Theorems 1 and 2, we see that the achievable rate expression in (10) is expressed in terms of a one-dimensional optimization over ω, while the expression in (16) is expressed in terms of a minimization over a $(p+2)$-length vector $\boldsymbol{\omega}$. Indeed, as shall be seen in the proof of Theorem 2 the additional parameters $\{\omega_k\}_{k=0}^{p}$ correspond to the $p+1$ constraints imposed by the codebook itself.

As mentioned in the Introduction, a byproduct of our analysis is an ensemble-tight characterization of the random coding error exponent. Indeed, the proofs of Theorems 1 and 2 are based on an exponentially-tight analysis of the (ensemble) average probability of error. This implies also that the achievable rate in Theorem 1 is ensemble-tight, namely, one cannot achieve higher rates using the same random coding ensemble. We next characterize the random coding error exponent in the special case of $\alpha_k = 0$ for $1 \leq k \leq K$ (namely, a memoryless decoding metric), and we use the fixed composition ensemble in (13) with $p = 0$. We emphasize that using the same methods we can analyze the more general case, but for simplicity we opted to focus on the above configuration. Let $E(P_X, R)$ designate the random coding error exponent associated with the above setting, namely, $E(P_X, R) \triangleq \liminf_{n\to\infty} -\frac{1}{n}\log \bar{P}_e(n, R)$ where $\bar{P}_e(n, R)$ designates the ensemble average probability of error. For brevity, we omit the dependency of $E(P_X, R)$ on α_0, $\{h_k\}_{k=0}^K$, and σ^2. We start with some definitions. Let $\Pi_m(\boldsymbol{h}) \triangleq \sum_{k=0}^{K-m} h_k h_{k+m}$, for $m \in [1:K]$, and for $\nu \in [0, 2\pi]$, define

$$u_{\hat{\boldsymbol{\omega}}}(\nu) \triangleq \frac{1}{4}\left|\hat{\omega}_1 - h_0 + \sum_{k=1}^{K} h_k e^{-jk\nu}\right|^2 - \left(\frac{1}{2\sigma^2} + \hat{\omega}_2\right)\cdot\left[\hat{\omega}_0 + \frac{1}{2}\left\|\boldsymbol{h}\right\|_2^2 + \sum_{k=1}^{K}\Pi_k(\boldsymbol{h})\cos(k\nu)\right], \quad (17)$$

where $\hat{\boldsymbol{\omega}} \in \hat{\mathcal{W}}$, and $\hat{\mathcal{W}}$ is the set of all $\hat{\boldsymbol{\omega}}$ satisfying $u_{\hat{\boldsymbol{\omega}}}(\nu) > 0$, for any $\nu \in [0, 2\pi]$. Also, for $P_Y \in \mathbb{R}_+$, $|\rho_{XY}| \leq 1$, let

$$V(\hat{\boldsymbol{\omega}}, P_Y, \rho_{XY}) \triangleq \frac{1}{4\pi}\int_0^{2\pi} \log\left[4eP_X\sigma^2 u_{\hat{\boldsymbol{\omega}}}(\nu)\right]\mathrm{d}\nu - \hat{\omega}_0\cdot P_X - \hat{\omega}_1\cdot\rho_{XY}\sqrt{P_XP_Y} - \hat{\omega}_2\cdot P_Y. \quad (18)$$

Finally, for $|\beta| < 1$, define

$$I(\beta, h_0, \alpha_0) \triangleq -\frac{|\mathrm{sgn}(h_0) + \mathrm{sgn}(\alpha_0)|}{4}\log\left(1 - \beta^2\right), \quad (19)$$

and note that if h_0 and α_0 share the same sign then $I(\beta, h_0, \alpha_0) = -0.5\cdot\log\left(1-\beta^2\right)$. Otherwise, $I(\beta, h_0, \alpha_0) = 0$. The following result is proved in Appendix C.

Theorem 3 Consider the Gaussian ISI channel model in (1), and the mismatched decoding metric in (3), with $\alpha_k = 0$, for $1 \leq k \leq K$. Then,

$$E(P_X, R) = \min_{P_Y, \rho_{XY}} \max_{\hat{\boldsymbol{\omega}} \in \hat{\mathcal{W}}} \left\{V(\hat{\boldsymbol{\omega}}, P_Y, \rho_{XY}) + \left[I(\rho_{XY}, h_0, \alpha_0) - R\right]_+\right\}. \quad (20)$$

Based on Theorem 3 we can see that if h_0 and α_0 have different signs then $E(P_X, R) = 0$, for any $R \geq 0$ and any P_X. This is indeed reasonable due to the following reason: if, for example, $h_0 > 0$ but $\alpha_0 < 0$, then the mismatched decoder simply looks for the codeword which *minimizes* its empirical correlation with the output sequence $\boldsymbol{y}$. However, this is exactly the opposite operation of the optimal ML decoder which maximizes the empirical correlation with $\boldsymbol{y}$. Also, one can argue that 20 resembles the famous Csiszár-Körner-style error exponent function [12], namely, $\min_Q\{D_{\mathrm{KL}}(Q||P) + [I(Q) - R]_+\}$, where $D_{\mathrm{KL}}(Q||P)$ is the Kullback-Leibler (KL) divergence between two measures Q and P, and $I(Q)$ is the mutual information calculated w.r.t. Q. This indeed makes sense, and one can think of $\max_{\hat{\boldsymbol{\omega}}\in\hat{\mathcal{W}}} V(\hat{\boldsymbol{\omega}}, P_Y, \rho_{XY})$ in (20) as playing the role analogous to the KL-divergence. For example, if $K=0$, then it is a simple task to check that

$$\max_{\hat{\boldsymbol{\omega}}\in\hat{\mathcal{W}}} V(\hat{\boldsymbol{\omega}}, P_Y, \rho_{XY}) = -\frac{1}{2}\log\frac{(1-\rho_{XY}^2)P_Y}{\sigma^2} + \frac{1}{2\sigma^2}\left[P_Y - 2h_0\rho_{XY}\sqrt{P_XP_Y} + h_0^2P_X\right] - \frac{1}{2} \quad (21)$$

which is just the KL-divergence $D_{\mathrm{KL}}(Q_{XY}||P_{XY})$, with Q_{XY} and P_{XY} both being multivariate Gaussian distributions, with zero means, and covariances

$$\boldsymbol{\Sigma}_Q = \begin{bmatrix} P_X & \rho_{XY}\sqrt{P_XP_Y} \\ \rho_{XY}\sqrt{P_XP_Y} & P_Y \end{bmatrix} \quad \text{and} \quad \boldsymbol{\Sigma}_P = \begin{bmatrix} P_X & h_0\cdot P_X \\ h_0\cdot P_X & h_0^2\cdot P_X + \sigma^2 \end{bmatrix}, \quad (22)$$

respectively. In general, the term $\max_{\hat{\boldsymbol{\omega}}\in\hat{\mathcal{W}}} V(\hat{\boldsymbol{\omega}}, P_Y, \rho_{XY})$ can be thought as the asymptotic formula of the n-letter weighted KL-divergence $D_{\mathrm{KL}}(Q_{Y^n|X^n}||W_{Y^n|X^n}|\mu_{X^n})$, where $Q_{Y^n|X^n}$ is some test channel, and μ_{X^n} is the random coding distribution, i.e., a uniform measure over the n-dimensional hypersphere with radius $\sqrt{nP_X}$.

We next compare numerically the results obtained in Theorems 1 and 2. Fig. 2 presents a numerical comparison of the results obtained in Theorems 1 and 2, in the following scenario. We consider the two-tap Gaussian ISI channel with $h_0 = h_1 = 1/\sqrt{2}$, $\sigma^2 = 1$, and $P_X = 1$. The mismatched decoder has a fixed coefficient $\alpha_0 = 1/\sqrt{2}$, and we calculate the achievable rates as a function of α_1. The matched capacity can be calculated numerically and it is given by $C = 0.374$ nats per channel use. The achievable rates in Fig. 2 correspond to fixed composition ensemble with one correlation parameter (solid black curve), fixed composition ensemble without correlations (dashed-dotted blue curve), first-order autoregressive ensemble (dashed brown curve), and Gaussian i.i.d. codebook (solid v-marked red curve). It can be seen that the fixed composition ensemble with one correlation parameter is almost uniformly better than all the other ensembles. The Gaussian i.i.d. codebook is the worst ensemble. In this example, all ensembles achieve the maximum rate at the matched value of α_1. Interestingly, we see that all ensembles (and especially the fixed composition ensemble with correlation) behave differently in the regions $\alpha_1 < 1/\sqrt{2}$ and $\alpha_1 > 1/\sqrt{2}$. These regions, respectively, correspond to the "optimistic" and "pessimistic" assumptions of the receiver regarding the ISI channel. To wit, when $\alpha_1 < 1/\sqrt{2}$ ($\alpha_1 > 1/\sqrt{2}$) the receiver can be thought of as being optimistic (pessimistic) since he assumes that the ISI part is weaker (stronger) than what it really is. Accordingly, it seems that in terms of achievable rates, the price of optimism is higher than the price of pessimism.

Another comparison is shown in Fig. 3, where now $h_0 = 2/\sqrt{5}$, $h_1 = 1/\sqrt{5}$, $\sigma^2 = 1$, $P_X = 1$, and $\alpha_1 = 1$. The matched ISI capacity in this case is $C = 0.3625$ nats per channel use. In this case, since $\alpha_1 \neq h_1$, the mismatched channel is never the same as the true channel. Contrary to Fig. 2, it can be seen that here the maximum rate is not achieved at the matched value of α_0. This observation illustrates that, in general, there is no direct connection between achievable rates and the minimization of the ℓ_2 distance between the true channel and the mismatched channel. Instead, if one imposes a fixed ℓ_2 distance between the true response and the mismatched response, the achievable rates might very well depend of

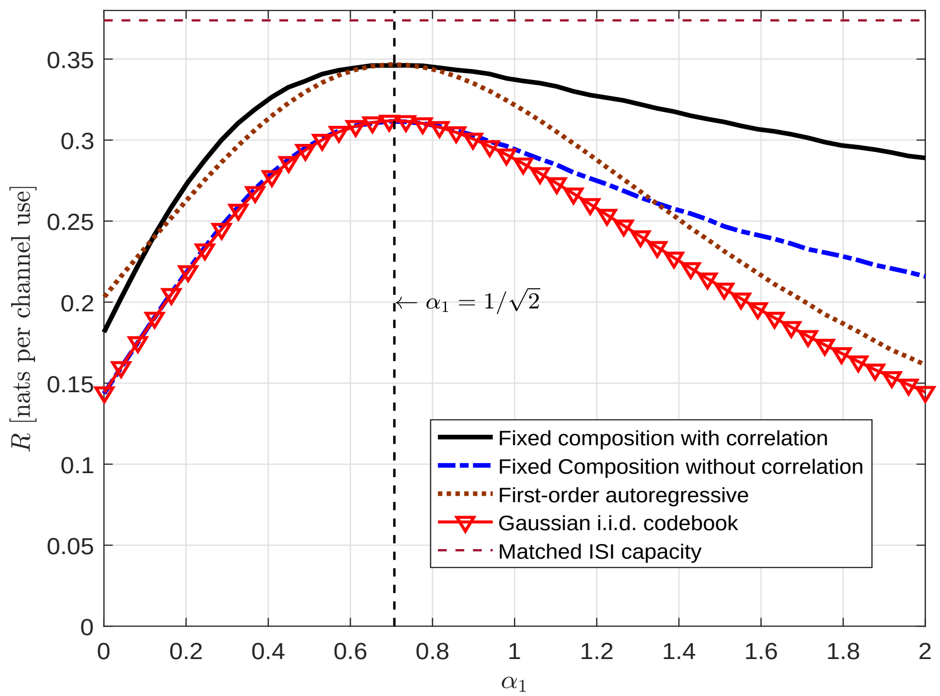

Fig. 2. Achievable rates as a function of the mismatched level α_1, for the Gaussian ISI channel with $h_0 = h_1 = 1/\sqrt{2}$, $\sigma^2 = 1$, $P_X = 1$, and $\alpha_0 = 1/\sqrt{2}$. The matched capacity is $C = 0.374$ nats per channel use. The achievable rates correspond to fixed composition ensemble with one correlation parameter (solid black curve), fixed composition ensemble without correlations (dashed-dotted blue curve), first-order autoregressive ensemble (dotted brown curve), and Gaussian i.i.d. codebook (solid v-mark red curve).

how well the mismatched response approximate the "good" part of the frequency domain representation of the true response. Similarly, in Fig. 4 we consider the case where $h_0 = h_1 = h_2 = 1/\sqrt{2}$, $\sigma^2 = 1$, and $P_X = 1$. The mismatched decoder has only one coefficient α_0 (namely, $K = 0$). Here, again, it can be seen that the maximum rate is not achieved at $\alpha_0 = 1/\sqrt{2}$. This illustrates that truncation (of the mismatched decoder) is not always optimal. Note that this result might be initially surprising since one can argue that, at least in the case of Gaussian i.i.d. codebook, if one decides to ignore the contribution of the ISI part, since input symbols are independent of each other, then this contribution plays the role of additional Gaussian additive noise. Accordingly, the best choice of the mismatch parameter should be the truncated one (namely $\alpha_0 = h_0$). This intuition is misleading due to the same reason mentioned above.

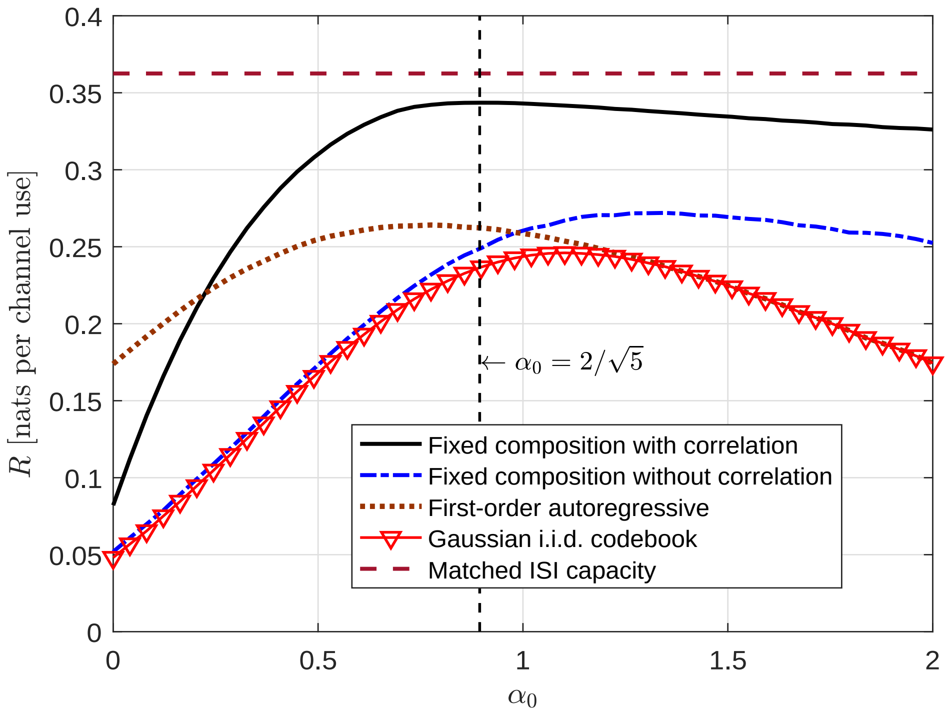

Fig. 3. Achievable rates as a function of the mismatched level α_0, for the Gaussian ISI channel with $h_0 = 2/\sqrt{5}$, $h_1 = 1/\sqrt{5}$, $\sigma^2 = 1$, $P_X = 1$, and $\alpha_1 = 1$. The matched capacity is $C = 0.3625$ nats per channel use. The achievable rates correspond to fixed composition ensemble with one correlation parameter (solid black curve), fixed composition ensemble without correlations (dashed-dotted blue curve), first-order autoregressive ensemble (dotted brown curve), and Gaussian i.i.d. codebook (solid v-mark red curve).

B. *Mismatched Universal Decoders*

As discussed in the Introduction in many situations in coded communication systems, channel uncertainty and variability preclude the implementation of the optimum ML decoder. In such cases, a good solution is provided by universal decoders which perform asymptotically as well as the ML decoder and yet do not require knowledge of the channel.

In this subsection, we analyze the following scenario: consider the same channel model presented in Section III. Then, assume that due to complexity issues concerning the implementation of the optimal ML decoder, the receiver uses the mismatched decoding metric in (3) with only α_0 being active, i.e., corresponding to a memoryless channel. Nonetheless, we allow our receiver to optimize this coefficient, namely, it can be a function of the true channel. Now, consider a different receiver which uses a universal decoder designed for a memoryless channel (namely without ISI). In other words, the true family of channels is *outside* the class of channels for which the universal decoder is actually designed, i.e, mismatched universal decoder. Then, which approach yields higher rates?

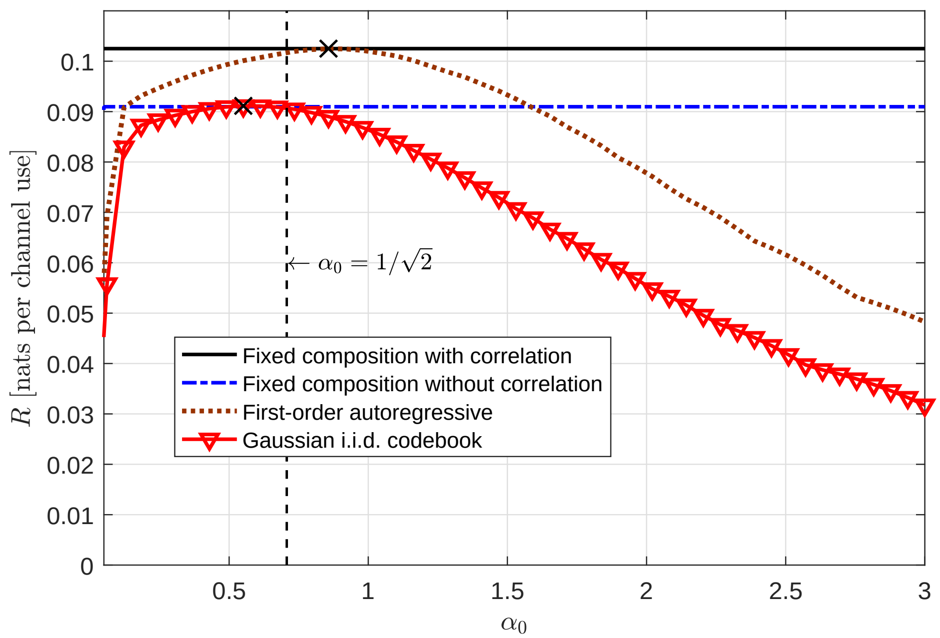

Fig. 4. Achievable rates as a function of the mismatched level α_0, for the Gaussian ISI channel with $h_0 = h_1 = h_2 = 1/\sqrt{2}$, $\sigma^2 = 1$, $P_X = 1$. The achievable rates correspond to fixed composition ensemble with one correlation parameter (solid black curve), fixed composition ensemble without correlations (dashed-dotted blue curve), first-order autoregressive ensemble (dotted brown curve), and Gaussian i.i.d. codebook (solid v-mark red curve). The cross symbols "X" refer to the maximum rate achieved by the corresponding ensembles.

In the sequel, for simplicity, we focus on the fixed composition ensemble without correlations. It is well-known (see, e.g., [1, 20, 28, 36, 38]) that, for the AWGN with codewords drawn uniformly and independently over the n-dimensional hypersphere with radius $\sqrt{nP_X}$, given an output sequence y^n, the decoder

$$\hat{i} = \arg\max_{i \in \mathfrak{C}_n} \left| \sum_{t=1}^{n} x_{t,i} y_t \right|, \tag{23}$$

is universal. Indeed, in this scenario, the generalized likelihood-ratio test (GLRT) is universal, and it is a simple exercise to check that the GLRT is equivalent to (23). Before we present our main result, we briefly comment that using the same techniques developed in this paper, one can consider more complicated ensembles, such as the one in (13), and universal decoders designed for the ISI channel [20] and not just for the AWGN, as described above. In principle, it makes sense that the random coding ensemble and the universal decoder would be consistent with one another in the sense of their assumptions on the memorylessness/memoryfulness of the channel. However, in the following discussion we demonstrate that we can consider the ensemble in (13) when the universal decoder is designed for a memoryless channel, with the understanding that we can actually consider also universal decoders designed for the

ISI channel [20]. Given $\varepsilon > 0$, and two sequences x^n and y^n, we define

$$\mathcal{T}_{\varepsilon}^{n}(\boldsymbol{x}|\boldsymbol{y}) \triangleq \left\{ \boldsymbol{x}' \in \mathbb{R}^n : \left| \frac{1}{n}\sum_{t=1}^{n} x_t' y_t - \frac{1}{n}\sum_{t=1}^{n} x_t y_t \right| < \epsilon \right\}. \tag{24}$$

Let $\mu(\cdot)$ designate the uniform measure over (13), and define

$$U(\boldsymbol{x},\boldsymbol{y}) \triangleq -\frac{1}{n}\log \mu\left(\mathcal{T}_{\varepsilon}^{n}(\boldsymbol{x}|\boldsymbol{y})\right) \tag{25}$$

$$= \frac{1}{n}\log \mathrm{Vol}\left(\mathcal{T}_{\varepsilon}^{n}(\boldsymbol{\gamma})\right) - \frac{1}{n}\log \mathrm{Vol}\left(\mathcal{T}_{\varepsilon}^{n}(\boldsymbol{\gamma}) \cap \mathcal{T}_{\varepsilon}^{n}(\boldsymbol{x}|\boldsymbol{y})\right). \tag{26}$$

Then, it can be shown that the decoder which maximizes the metric in (26) over the given codebook, is universal w.r.t. the ensemble in (13). Since the first term at the right hand side (r.h.s.) of (26) is in fact independent of $\boldsymbol{x}$ it can be omitted, and so to explicitly derive the above decoding metric, we just need to evaluate the log-volume of the set $\mathcal{T}_{\varepsilon}^{n}(\boldsymbol{\gamma}) \cap \mathcal{T}_{\varepsilon}^{n}(\boldsymbol{x}|\boldsymbol{y})$. This can be done using the same techniques used in the proof of Theorem 2. Returning back to our setting, we have the following result.

Theorem 4 Consider the Gaussian ISI channel model in (1), and the mismatched universal decoder in (23). Then, every

$$R < \frac{1}{2}\log\left(1 + \frac{h_0^2 \cdot P_X}{(\|\boldsymbol{h}\|_2^2 - h_0^2)\cdot P_X + \sigma^2}\right) \tag{27}$$

is achievable.

As before, we can obtain the random coding error exponent associated with the above universal decoder.

Theorem 5 Consider the Gaussian ISI channel model in (1), and the mismatched universal decoder in (23). Then,

$$E(P_X,R) = \min_{P_Y,\rho_{XY}} \max_{\hat{\boldsymbol{\omega}}\in\hat{\mathcal{W}}} \left\{ V(\hat{\boldsymbol{\omega}},P_Y,\rho_{XY}) + \left[-\frac{1}{2}\log(1-\rho_{XY}^2) - R\right]_+ \right\}. \tag{28}$$

We provide a proof sketch of Theorem 4 in Appendix D. The proof of Theorem 5 is essentially the same as the proof of Theorem 3, and thus omitted. From Theorem 4 we see that the achievable rate in (27) has the interpretation that the "mismatched input signal" (or the residue signal) is treated as additional Gaussian noise at the decoder. Accordingly, it is interesting to compare the above results with Theorems 2 and 3. Some technical calculations reveal that, for the specific scenario we consider above, the achievable rate $\max_{\alpha_0} \bar{I}_2(\alpha_0,P_X)$ in Theorem 2 is essentially *exactly* the same as (27). Moreover, the optimal α_0 which achieves (27) should be chosen as to match the sign of h_0, namely, if $h_0 > 0$ then α_0 can be any real but positive number, while if $h_0 < 0$ then α_0 can be any real but negative number. Accordingly, while the mismatched decoder needs some knowledge of the true channel, it achieves the same rates as the universal decoder which has no knowledge at all about the true channel. In particular, if the sign of α_0 was unknown to the mismatched decoder, and the mismatch decoder mistakenly assumes a different sign compared to the true channel, then it will achieve zero-rate. This basically demonstrates that, at least in the specific scenario described above, our results provide indications that universal decoders exhibit

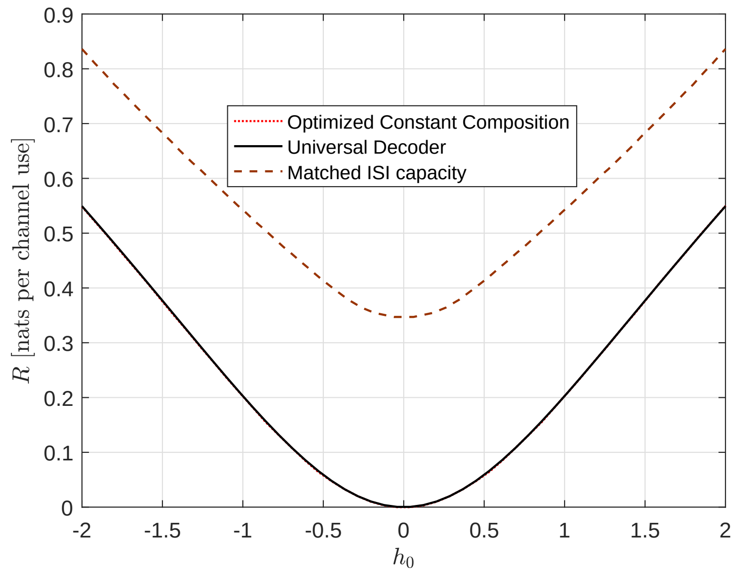

Fig. 5. Achievable rates as a function of the coefficient h_0, for the Gaussian ISI channel with two taps and $h_1 = 1$, $\sigma^2 = 1$, and $P_X = 1$. The achievable rates correspond to the optimized (over α_0) mismatched decoder (dotted red curve), and the mismatched universal decoder (solid black curve), both under fixed composition ensemble without correlations.

a robustness property w.r.t. the family of channels over which they are actually designed. To wit, even though our universal decoder is designed for a completely different class of channels, it still performs as well as the best mismatched ML decoder in the same family. This observation (potentially) suggests a certain expansion of the classic notion of universality to cases where the true underlying channel is outside the class. Similarly, comparing the error exponents in Theorem 3 and 5 we see that the only difference is the additional $|\text{sgn}(h_0) + \text{sgn}(\alpha_0)|/2$ term in (19). Accordingly, if h_0 and α_0 share the same sign, then both exponents coincide, otherwise, the exponent in Theorem 3 vanishes, while the exponent in Theorem 5 remains unaffected. Finally, we present in Fig. 5 a numerical calculation of the achievable rate in Theorem 4, for the two-tap Gaussian ISI channel with $h_1 = 1$, $\sigma^2 = 1$, and $P_X = 1$.

V. Proof of Theorem 1

The codebook $\mathfrak{C}_n = \{\boldsymbol{x}_i\}_{i=1}^M$ is generated as follows: For each message $i \in [1 : e^{nR}]$, we generate (independently) a sequence $\{X_t\}_{t=1}^n$ according to (11), where $\{Z_t\}_t$ is white Gaussian noise, η^2 is chosen such that $\mathbb{E}X_t^2 = P_X$, for any $t = [1 : n]$, and thus it is given in (5). Also, for technical reasons, we fix $X_t = 0$ for $t = n+1, \ldots, n+(p \vee K)$, and for each message $i \in [1 : e^{nR}]$. We will then think of all the forthcoming summation terms over $t = [1 : n]$ as summing up to $\bar{n} = n + p \vee K$ instead of n. This assumption is made only for convenience, but has no influence on either the achievable error exponents or the achievable rates, as long as $p \vee K$ is fixed and $n \to \infty$. Accordingly, we can write the following circularity relation

$$\sum_{t=1}^{\bar{n}} X_{t-i} X_{t-j} = \sum_{t=1}^{\bar{n}} X_t X_{t-(j-i)} \tag{29}$$

for $i \leq j \in [1 : p \vee K]$. For notational convenience, in the following we use n in place of $\bar{n}$. We let $\mu_p(\cdot)$ designate the probability density function of a sequence of RV's generated as described above. Finally, define

$$\gamma_m \triangleq \mathbb{E}\left\{X_t X_{t-m}\right\} \tag{30}$$

to be the auto-covariance of the autoregressive process, and recall that these coefficients can be found by evaluating (6).

Without loss of generality, we assume throughout, that the transmitted codeword is $\boldsymbol{x} \triangleq \boldsymbol{x}_1$. Accordingly, the average probability of error can be written as

$$\bar{P}_e(n,R) = \Pr\left[\bigcup_{i=2}^{M}\left\{\frac{1}{n}\log V\left(\boldsymbol{Y}|\boldsymbol{X}_i\right) \geq \frac{1}{n}\log V\left(\boldsymbol{Y}|\boldsymbol{X}_1\right)\right\}\right] \tag{31}$$

$$= \mathbb{E}\left\{\Pr\left[\bigcup_{i=2}^{M}\left\{\frac{1}{n}\log V\left(\boldsymbol{Y}|\boldsymbol{X}_i\right) \geq \frac{1}{n}\log V\left(\boldsymbol{Y}|\boldsymbol{X}_1\right)\right\}\middle|\,\mathcal{F}_0\right]\right\} \tag{32}$$

where $\mathcal{F}_0 \triangleq (\boldsymbol{X}_1, \boldsymbol{Y})$. Recall that for a pairwise independent events $\{\mathcal{A}_i\}_{i=1}^M$, we have [39, Sec. A2]

$$\frac{1}{2}\min\left\{1, \sum_{i=1}^{M}\Pr\left\{\mathcal{A}_i\right\}\right\} \leq \Pr\left\{\bigcup_{i=1}^{M}\mathcal{A}_i\right\} \leq \min\left\{1, \sum_{i=1}^{M}\Pr\left\{\mathcal{A}_i\right\}\right\},$$

and therefore,

$$\bar{P}_e(n,R) \doteq \mathbb{E}\left[\min\left\{1, M\cdot\Pr\left\{\frac{1}{n}\log V\left(\boldsymbol{Y}|\boldsymbol{X}_2\right) \geq \frac{1}{n}\log V\left(\boldsymbol{Y}|\boldsymbol{X}_1\right)\middle|\,\mathcal{F}_0\right\}\right\}\right]. \tag{33}$$

Thus, we need to assess the exponential behavior of the probability term in (33). Fix a positive constant $B > 0$, and define the sequence of sets

$$\mathcal{H}_n(B) \triangleq \left\{\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n :\ \frac{1}{n}\sum_{i=1}^{n} x_i^2 \leq B,\ \frac{1}{n}\sum_{i=1}^{n} y_i^2 \leq B\right\}, \tag{34}$$

for $n \geq 1$. Accordingly, (33) can be rewritten as follows

$$\begin{aligned}\bar{P}_e(n,R) \doteq\ & \mathbb{E}\left[\mathcal{I}\left\{\mathcal{H}_n(B)\right\}\cdot\min\left\{1, M\cdot\Pr\left\{\frac{1}{n}\log V\left(\boldsymbol{Y}|\boldsymbol{X}_2\right) \geq \frac{1}{n}\log V\left(\boldsymbol{Y}|\boldsymbol{X}_1\right)\middle|\,\mathcal{F}_0\right\}\right\}\right] \\ & + \mathbb{E}\left[\mathcal{I}\left\{\mathcal{H}_n^c(B)\right\}\cdot\min\left\{1, M\cdot\Pr\left\{\frac{1}{n}\log V\left(\boldsymbol{Y}|\boldsymbol{X}_2\right) \geq \frac{1}{n}\log V\left(\boldsymbol{Y}|\boldsymbol{X}_1\right)\middle|\,\mathcal{F}_0\right\}\right\}\right].\end{aligned} \tag{35}$$

Using the same techniques as in [19, Lemma 2] and [20, Lemma 2] it can be shown that there exists a sufficiently large B such that the first term at the r.h.s. of dominates in the exponential scale. This passage is mainly technical, and will be used to evaluate the volume of some typical sets as shall be clear in the sequel. To conclude, we have

$$\bar{P}_e(n,R) \doteq \mathbb{E}\left[\mathcal{I}\left\{\mathcal{H}_n(B)\right\}\cdot\min\left\{1, M\cdot\Pr\left\{\frac{1}{n}\log V\left(\boldsymbol{Y}|\boldsymbol{X}_2\right) \geq \frac{1}{n}\log V\left(\boldsymbol{Y}|\boldsymbol{X}_1\right)\middle|\,\mathcal{F}_0\right\}\right\}\right]. \tag{36}$$

For $(\boldsymbol{X}_1, \boldsymbol{Y}) = (\boldsymbol{x}, \boldsymbol{y}) \in \mathcal{H}_n(B)$, the inner probability term at the r.h.s. of (36) can be represented as

$$\Pr\left\{\frac{1}{n}\log V\left(\boldsymbol{y}|\boldsymbol{X}_2\right) \geq \frac{1}{n}\log V\left(\boldsymbol{y}|\boldsymbol{x}\right)\right\} = \int_{\boldsymbol{x}' \in \mathcal{V}(\boldsymbol{x},\boldsymbol{y})} \mathrm{d}\boldsymbol{x}' \mu_p(\boldsymbol{x}') \tag{37}$$

where

$$\mathcal{V}(\boldsymbol{x}, \boldsymbol{y}) \triangleq \left\{\boldsymbol{x}' \in \mathbb{R}^n : \ \frac{1}{n}\log V\left(\boldsymbol{y}|\boldsymbol{x}'\right) \geq \frac{1}{n}\log V\left(\boldsymbol{y}|\boldsymbol{x}\right)\right\}. \tag{38}$$

Using the saddle-point integration method we asses the exponential behavior of the r.h.s. of (37). To present the result, let $\omega \in \mathbb{R}_+$, and define for $1 \leq m \leq K$ and $1 \leq r \leq p$,

$$\Pi_m(\boldsymbol{\alpha}) \triangleq \sum_{k=0}^{K-m} \alpha_k \alpha_{k+m}, \tag{39}$$

$$\hat{\Pi}_r(\boldsymbol{\varphi}) \triangleq \sum_{k=1}^{p-r} \varphi_k \varphi_{k+r}, \tag{40}$$

and

$$\psi(\boldsymbol{x}, \boldsymbol{y}) \triangleq \sum_{l=0}^{K} \alpha_l \sum_{t=1}^{n} y_t x_{t-l} - \frac{1}{2}\left\|\boldsymbol{\alpha}\right\|_2^2 \sum_{t=1}^{n} x_t^2 - \sum_{l=1}^{K} \Pi_l(\boldsymbol{\alpha}) \sum_{t=1}^{n} x_t x_{t-l}. \tag{41}$$

Also, let $\boldsymbol{\Omega} \triangleq \boldsymbol{\Omega}_1 + \boldsymbol{\Omega}_2$ be the $n \times n$ symmetric Toeplitz matrix,

$$[\boldsymbol{\Omega}_1]_{i,j} \triangleq \frac{\omega}{2} \cdot \begin{cases} \left\|\boldsymbol{\alpha}\right\|_2^2, & \text{for } |i-j| = 0 \\ \Pi_{|i-j|}(\boldsymbol{\alpha}), & \text{for } 1 \leq |i-j| \leq K \\ 0, & \text{otherwise} \end{cases}, \tag{42}$$

and

$$[\boldsymbol{\Omega}_2]_{i,j} \triangleq \frac{1}{2\eta^2} \cdot \begin{cases} 1 + \left\|\boldsymbol{\varphi}\right\|_2^2, & \text{for } |i-j| = 0 \\ \hat{\Pi}_{|i-j|}(\boldsymbol{\varphi}) - \varphi_{|i-j|}, & \text{for } 1 \leq |i-j| \leq p \\ 0, & \text{otherwise} \end{cases}. \tag{43}$$

Let $\mathbf{A}$ be an $n \times n$ lower-triangular Toeplitz matrix where for $i \geq j$,

$$[\mathbf{A}]_{i,j} \triangleq \begin{cases} \alpha_{i-j}, & \text{for } 0 \leq i-j \leq K \\ 0, & \text{otherwise} \end{cases}. \tag{44}$$

Finally, define

$$q_n(\omega, \boldsymbol{x}, \boldsymbol{y}) \triangleq \frac{1}{2}\log\pi - \frac{1}{2n}\log\det\boldsymbol{\Omega} + \frac{\omega^2}{4n}\boldsymbol{y}^T\mathbf{A}\boldsymbol{\Omega}^{-1}\mathbf{A}^T\boldsymbol{y} - \frac{\omega}{n}\psi(\boldsymbol{x}, \boldsymbol{y}), \tag{45}$$

and

$$h_n(\boldsymbol{x}, \boldsymbol{y}) \triangleq \min_{\omega \in \mathbb{R}_+} q_n(\omega, \boldsymbol{x}, \boldsymbol{y}) - \frac{1}{2}\log(2\pi\eta^2). \tag{46}$$

The following lemma is proved in Appendix A.

Lemma 1 Fix $(\boldsymbol{x},\boldsymbol{y}) \in \mathcal{H}_n(B)$. Then, for any $\delta_1 > 0$, there exists $N_{\delta_1} \in \mathbb{N}$ large enough, such that for any $n > N_{\delta_1}$,

$$\left|\frac{1}{n}\log\int_{\mathcal{V}(\boldsymbol{x},\boldsymbol{y})} \mathrm{d}\boldsymbol{x}'\mu_p(\boldsymbol{x}') - h_n(\boldsymbol{x},\boldsymbol{y})\right| < \delta_1. \tag{47}$$

Using (36)-(37), and Lemma 1, we may conclude that

$$\bar{P}_e(n,R) \doteq \mathbb{E}\left[\mathcal{I}\left\{\mathcal{H}_n(B)\right\}\exp\left\{-n\left[-h_n(\boldsymbol{X},\boldsymbol{Y}) - \delta_1 - R\right]_+\right\}\right]. \tag{48}$$

The next step is taking the expectation w.r.t. $(\boldsymbol{X},\boldsymbol{Y})$, distributed according to $\mu_p \times W$. In the following, we calculate the limit of (46) as $n \to \infty$. Due to the fact that $q_n(\omega,\boldsymbol{x},\boldsymbol{y})$ converges uniformly to a limit almost surely, as shall be shown in a moment, we can interchange the order of the limit and the minimization in (46) [40]. Thus, we focus on the limit of $q_n(\omega,\boldsymbol{x},\boldsymbol{y})$, and we consider the asymptotics of each term in (45). Since $\boldsymbol{\Omega}$ is a Toepliz matrix, using Szegö theorem [41], we have

$$\lim_{n\to\infty}\frac{1}{n}\log\det\boldsymbol{\Omega} = \frac{1}{2\pi}\int_{[0,2\pi]}\mathrm{d}\nu\log f_\omega(\nu) \tag{49}$$

where

$$f_\omega(\nu) \triangleq \frac{\omega}{2}\left[\|\boldsymbol{\alpha}\|_2^2 + 2\sum_{k=1}^{K}\Pi_k(\boldsymbol{\alpha})\cos(k\nu)\right] + \frac{1}{2\eta^2}\left[1 + \|\boldsymbol{\varphi}\|_2^2 + 2\sum_{k=1}^{p}(\hat{\Pi}_k(\boldsymbol{\varphi}) - \varphi_k)\cos(k\nu)\right] \tag{50}$$

$$= \frac{\omega}{2}|A(\nu)|^2 + \frac{1}{2\eta^2}\left[1 + |\Phi(\nu)|^2 - 2\cdot\mathrm{Re}(\Phi(\nu))\right]. \tag{51}$$

Next, using the law of large numbers (LLN) and once again Szegö theorem, we get, with overwhelming probability as $n \to \infty$,

$$\lim_{n\to\infty}\frac{1}{n}\boldsymbol{y}^T\mathbf{A}\boldsymbol{\Omega}^{-1}\mathbf{A}^T\boldsymbol{y} = \lim_{n\to\infty}\frac{1}{n}\mathbb{E}\left(\boldsymbol{Y}^T\mathbf{A}\boldsymbol{\Omega}^{-1}\mathbf{A}^T\boldsymbol{Y}\right) \tag{52}$$

$$= \lim_{n\to\infty}\frac{1}{n}\,\mathrm{tr}\left\{\mathbf{A}\boldsymbol{\Omega}^{-1}\mathbf{A}^T\mathbf{R}_{\boldsymbol{y}}\right\} \tag{53}$$

$$= \frac{1}{2\pi}\int_{[0,2\pi]}\mathrm{d}\nu\frac{S_Y(\nu)\left|A(\nu)\right|^2}{f_\omega(\nu)} \tag{54}$$

where $S_X(\cdot)$, $S_Y(\cdot)$, and $f_\omega(\cdot)$, are defined in (7)-(9), respectively, and $S_Y(\cdot)$ is the spectral density function of $\boldsymbol{Y}$, namely, it is the Fourier transform of the auto-covariance

$$\mathrm{R}_Y(t,s) = \sum_{m,k=0}^{K} h_m h_k \mathbb{E}\left\{X_{t-m}X_{s-k}\right\} + \mathrm{R}_W(t-s) \tag{55}$$

$$= \sum_{m,k=0}^{K} h_{m_1}h_{m_2}\gamma_{t-s+k-m} + \mathrm{R}_W(t-s) \tag{56}$$

$$= \sum_{m,k=0}^{K} h_{m_1}h_{m_2}\gamma_{u+k-m} + \mathrm{R}_W(u) \tag{57}$$

$$= \mathrm{R}_Y(u), \tag{58}$$

where $\{\gamma_m\}$ are defined in (30), and thus

$$S_Y(\nu) = \sum_{m,k=0}^{K} h_{m_1} h_{m_2} S_X(\nu) e^{j(m-k)\nu} + \sigma^2 \tag{59}$$

$$= \left[\sum_{m=0}^{K} h_m^2 + 2 \cdot \sum_{m<k}^{K} h_m h_k \cos\left[(k-m)\nu\right]\right] S_X(\nu) + \sigma^2. \tag{60}$$

Finally, by the LLN and by using [42, Th. 2.4.2] we have, with overwhelming probability as $n \to \infty$,

$$\lim_{n\to\infty} \frac{1}{n}\psi(\boldsymbol{x},\boldsymbol{y}) = \sum_{l=0}^{K}\sum_{i=0}^{K} \alpha_l h_i \gamma_{l-i} - \left(\frac{1}{2}\left\|\boldsymbol{\alpha}\right\|_2^2 \gamma_0 + \sum_{l=1}^{K} \Pi_l(\boldsymbol{\alpha})\gamma_l\right). \tag{61}$$

Collecting the last results, we obtain that with probability approaching one as $n \to \infty$,

$$\begin{aligned}\lim_{n\to\infty} h_n(\boldsymbol{X},\boldsymbol{Y}) = \min_{\omega} \Bigg\{ & \frac{1}{2}\log\pi - \frac{1}{4\pi}\int_0^{2\pi} \mathrm{d}\nu \log f_\omega(\nu) + \frac{\omega^2}{8\pi}\int_{[0,2\pi]} \mathrm{d}\nu \frac{S_Y(\nu)\left|A(\nu)\right|^2}{f_\omega(\nu)} \\ & -\omega\left[\sum_{l=0}^{K}\sum_{i=0}^{K} \alpha_l h_i \gamma_{l-i} - \left(\frac{1}{2}\left\|\boldsymbol{\alpha}\right\|_2^2 \gamma_0 + \sum_{l=1}^{K}\Pi_l(\boldsymbol{\alpha})\gamma_l\right)\right]\Bigg\} - \frac{1}{2}\log(2\pi\eta^2) \\ \triangleq -\bar{I}_1(\boldsymbol{\alpha},\boldsymbol{\varphi}). \end{aligned} \tag{62}$$

Next, for any $\delta_2 > 0$, define the set

$$\mathcal{A}_n(\delta_2) \triangleq \left\{(\boldsymbol{x},\boldsymbol{y}):\ \left|h_n(\boldsymbol{x},\boldsymbol{y}) + \bar{I}_1(\boldsymbol{\alpha},\boldsymbol{\varphi})\right| < \delta_2\right\}. \tag{63}$$

Accordingly, from (48) we have

$$\bar{P}_e(n,R) \leq \mathbb{E}\left[\mathcal{I}\left\{\mathcal{A}_n(\delta_2)\right\}\exp\left\{-n\left[-h_n(\boldsymbol{X},\boldsymbol{Y}) - \delta_1 - R\right]_+\right\}\right] + \Pr\left\{\mathcal{A}_n^c(\delta_2)\right\}, \tag{64}$$

and in light of (62), taking $\delta_1 \to 0$ and $\delta_2 \to 0$, followed by $n \to \infty$, the last term at the r.h.s. of (64) is asymptotically negligible. Hence, in terms of achievable rate, we get that $\bar{P}_e(n,R)$ decays to zero as long as

$$R < \bar{I}_1(\boldsymbol{\alpha},\boldsymbol{\varphi}). \tag{65}$$

Since $\{\varphi_i\}_{i=1}^p \in \mathcal{P}$ were fixed parameters up to this point, we take the maximum of the r.h.s. of the above inequality over these parameters, which concludes the proof.

VI. CONCLUSIONS

In this paper, we considered the problem of channel coding over Gaussian ISI channels with a mismatched decoding rule. For this problem we provided two achievable rates using an autoregressive random coding ensemble, and fixed composition ensemble. We then presented a mismatched universal decoder, designed outside the true class of channels, and showed that it is robust. Finally, we discuss a certain generalization and some implications of our analysis. Using the same techniques developed in this paper, we can analyze more complicated scenarios of multi-user systems, such as, the ISI multiple-access channel [43], and the ISI broadcast channel. Specifically, to obtain ensemble-tight results, the only step in our proof that should be modified is the application of the truncated union-bound, which is not necessarily tight in the multi-user settings. To this end, one can use the tighter union bounds that were derived in [44, 45].

Appendix A
Proof of Lemma 1

Recall the fact that the step function $\mathbb{1}\left\{x \geq 0\right\}$ is the inverse Laplace transform of the function $1/s$, i.e.,

$$\mathbb{1}\left\{x \geq 0\right\} = \frac{1}{2\pi j}\int_{c-j\infty}^{c+j\infty} \mathrm{d}t \frac{\exp\left(tx\right)}{t}, \tag{A.1}$$

for any $c > 0$. Also, note that

$$\begin{aligned}
\mathcal{V}(\boldsymbol{x},\boldsymbol{y}) &= \left\{\boldsymbol{x}' \in \mathbb{R}^n :\ \log V(\boldsymbol{y}|\boldsymbol{x}') \geq \log V(\boldsymbol{y}|\boldsymbol{x})\right\} \\
&= \left\{\boldsymbol{x}' \in \mathbb{R}^n :\ \sum_{l=0}^{K}\alpha_l \sum_{t=1}^{n} y_t(x'_{t-l} - x_{t-l}) - \frac{1}{2}\sum_{l,k=0}^{K}\alpha_l\alpha_k\sum_{t=1}^{n}(x'_{t-l}x'_{t-k} - x_{t-l}x_{t-k}) \geq 0\right\} \\
&= \left\{\boldsymbol{x}' \in \mathbb{R}^n :\ \sum_{l=0}^{K}\alpha_l \sum_{t=1}^{n} y_t(x'_{t-l} - x_{t-l}) - \frac{1}{2}\left\|\boldsymbol{\alpha}\right\|_2^2\sum_{t=1}^{n}(x_t'^2 - x_t^2)\right. \\
&\qquad\qquad \left. - \sum_{l=1}^{K}\Pi_l(\boldsymbol{\alpha})\sum_{t=1}^{n}(x'_t x'_{t-l} - x_t x_{t-l}) \geq 0\right\} \qquad (\text{A.2}) \\
&= \left\{\boldsymbol{x}' \in \mathbb{R}^n :\ \psi(\boldsymbol{x}',\boldsymbol{y}) - \psi(\boldsymbol{x},\boldsymbol{y}) \geq 0\right\} \qquad (\text{A.3})
\end{aligned}$$

where $\psi(\boldsymbol{x},\boldsymbol{y})$ is given in (41), and (A.2) follows from the circularity assumption (29), indeed,

$$\begin{aligned}
\sum_{l=0}^{K}\sum_{k=0}^{K}\alpha_l\alpha_k\sum_{t=1}^{n}x_{t-l}x_{t-k} &= \left\|\boldsymbol{\alpha}\right\|_2^2\sum_{t=1}^{n}x_t^2 + 2\cdot\sum_{l=0}^{K}\sum_{k=l+1}^{K}\alpha_l\alpha_k\sum_{t=1}^{n}x_t x_{t-(k-l)} \qquad (\text{A.4}) \\
&= \left\|\boldsymbol{\alpha}\right\|_2^2\sum_{t=1}^{n}x_t^2 + 2\cdot\sum_{l=0}^{K}\sum_{m=1}^{K-l}\alpha_l\alpha_{m+l}\sum_{t=1}^{n}x_t x_{t-m} \qquad (\text{A.5}) \\
&= \left\|\boldsymbol{\alpha}\right\|_2^2\sum_{t=1}^{n}x_t^2 + 2\cdot\sum_{l=1}^{K}\Pi_l(\boldsymbol{\alpha})\sum_{t=1}^{n}x_t x_{t-l} \qquad (\text{A.6})
\end{aligned}$$

where $\Pi_l(\boldsymbol{\alpha})$ is defined in (39).

Then, we may write

$$\begin{aligned}
\int_{\mathcal{V}(\boldsymbol{x},\boldsymbol{y})}\mathrm{d}\boldsymbol{x}'\mu_p(\boldsymbol{x}') &= \int_{\mathbb{R}^n}\mathrm{d}\boldsymbol{x}'\mu_p(\boldsymbol{x}')\mathbb{1}\left\{\psi(\boldsymbol{x}',\boldsymbol{y}) - \psi(\boldsymbol{x},\boldsymbol{y}) \geq 0\right\} \qquad (\text{A.7}) \\
&= \frac{1}{2\pi j}\int_{\mathbb{R}^n}\mathrm{d}\boldsymbol{x}'\mu_p(\boldsymbol{x}')\int_{c-j\infty}^{c+j\infty}\mathrm{d}\omega\frac{1}{\omega}\exp\left\{\omega\left(\psi(\boldsymbol{x}',\boldsymbol{y}) - \psi(\boldsymbol{x},\boldsymbol{y})\right)\right\} \qquad (\text{A.8}) \\
&= \frac{1}{2\pi j}\int_{c-j\infty}^{c+j\infty}\mathrm{d}\omega\frac{1}{\omega}\exp\left\{-\omega\psi(\boldsymbol{x},\boldsymbol{y})\right\}\int_{\mathbb{R}^n}\mathrm{d}\boldsymbol{x}'\mu_p(\boldsymbol{x}')\exp\left\{\omega\psi(\boldsymbol{x}',\boldsymbol{y})\right\} \qquad (\text{A.9})
\end{aligned}$$

where

$$\mu_p(\boldsymbol{x}) = \frac{1}{(2\pi\eta^2)^{n/2}} \exp\left\{-\frac{1}{2\eta^2}\sum_{t=1}^{n}\left(x_t - \sum_{l=1}^{p}\varphi_l x_{t-l}\right)^2\right\}, \tag{A.10}$$

and due to the circularity assumption we may write

$$\mu_p(\boldsymbol{x}) = \frac{1}{(2\pi\eta^2)^{n/2}} \exp\left\{-\frac{1}{2\eta^2}\left[\left(1+\|\boldsymbol{\varphi}\|_2^2\right)\|\boldsymbol{x}\|_2^2 + 2\sum_{l=1}^{p}(\hat{\Pi}_l(\boldsymbol{\varphi})-\varphi_l)\sum_{t=1}^{n}x_t x_{t-l}\right]\right\} \tag{A.11}$$

$$\triangleq \frac{1}{(2\pi\eta^2)^{n/2}} \exp\left\{-\frac{1}{2\eta^2}L(\boldsymbol{x})\right\}. \tag{A.12}$$

Thus,

$$\int_{\mathcal{V}(\boldsymbol{x},\boldsymbol{y})} \mathrm{d}\boldsymbol{x}'\mu_p(\boldsymbol{x}') = \frac{1}{2\pi j}\frac{1}{(2\pi\eta^2)^{n/2}}\int_{c-j\infty}^{c+j\infty}\mathrm{d}\omega\frac{\exp\{-\omega\psi(\boldsymbol{x},\boldsymbol{y})\}}{\omega} \times \int_{\mathbb{R}^n}\mathrm{d}\boldsymbol{x}'\exp\left\{-\frac{1}{2\eta^2}L(\boldsymbol{x}')+\omega\psi(\boldsymbol{x}',\boldsymbol{y})\right\}. \tag{A.13}$$

Now, the last exponent at the r.h.s. of the above equality can be rewritten as

$$\frac{1}{2\eta^2}L(\boldsymbol{x}) - \omega\psi(\boldsymbol{x}',\boldsymbol{y}) = \boldsymbol{x}'^T\boldsymbol{\Omega}\boldsymbol{x}' - \omega\boldsymbol{y}^T\mathbf{A}\boldsymbol{x}' \tag{A.14}$$

$$= (\boldsymbol{x}' - \frac{\omega}{2}\boldsymbol{\Omega}^{-1}\mathbf{A}\boldsymbol{y})^T\boldsymbol{\Omega}(\boldsymbol{x}' - \frac{\omega}{2}\boldsymbol{\Omega}^{-1}\mathbf{A}\boldsymbol{y}) - \frac{\omega^2}{4}\boldsymbol{y}^T\mathbf{A}\boldsymbol{\Omega}^{-1}\mathbf{A}^T\boldsymbol{y} \tag{A.15}$$

where $\boldsymbol{\Omega}$ and $\mathbf{A}$ are defined in (42)-(43) and (44), respectively. Then,

$$\int_{\mathbb{R}^n}\mathrm{d}\boldsymbol{x}'\exp\left\{-\frac{1}{2\eta^2}L(\boldsymbol{x}')+\omega\psi(\boldsymbol{x}',\boldsymbol{y})\right\} = \left[\det\left(2\pi\cdot\frac{1}{2}\boldsymbol{\Omega}^{-1}\right)\right]^{1/2}\exp\left\{\frac{\omega^2}{4}\boldsymbol{y}^T\mathbf{A}\boldsymbol{\Omega}^{-1}\mathbf{A}^T\boldsymbol{y}\right\}$$

$$= \exp\left\{\frac{n}{2}\log\pi - \frac{1}{2}\log\det\boldsymbol{\Omega} + \frac{\omega^2}{4}\boldsymbol{y}^T\mathbf{A}\boldsymbol{\Omega}^{-1}\mathbf{A}^T\boldsymbol{y}\right\} \tag{A.16}$$

where in the first equality we have used the fact that $\boldsymbol{\Omega}$ is a symmetric positive-definite Toeplitz matrix.[4] Therefore, using the last result and (A.13), we get that

$$\int_{\mathcal{V}(\boldsymbol{x},\boldsymbol{y})} \mathrm{d}\boldsymbol{x}'\mu_p(\boldsymbol{x}') \doteq \frac{1}{(2\pi\eta^2)^{n/2}}\int_{c-j\infty}^{c+j\infty}\mathrm{d}\omega\frac{1}{\omega}\exp\left[nq_n(\omega,\boldsymbol{x},\boldsymbol{y})\right], \tag{A.17}$$

and $q_n(\omega,\boldsymbol{x},\boldsymbol{y})$ is defined in (45). The integral in (A.17) can now be assessed using the saddle-point method [18, 46]. The derivative of $q_n(\omega,\boldsymbol{x},\boldsymbol{y})$ vanishes at the value of ω that solves the equation $\partial q_n(\omega,\boldsymbol{x},\boldsymbol{y})/\partial\omega = 0$, where the gradient is taken w.r.t. ω. We will show that this saddle-point, denoted by ω^*, is in fact real-valued, i.e., $\omega^*\in\mathbb{R}_+$. Accordingly, we choose $c=\omega^*$, and thereby let the integration path pass through this saddle-point. Now, at ω^*, $q_n(\omega,\boldsymbol{x},\boldsymbol{y})$ has its maximum along the vertical direction $\omega = \omega^* + j\kappa,\ -\infty<\kappa<\infty$ (and hence it dominates the integral), but since it is a saddle-point it minimizes $q_n(\omega,\boldsymbol{x},\boldsymbol{y})$ in the horizontal direction (the real line), so we get

$$\int_{\mathcal{V}(\boldsymbol{x},\boldsymbol{y})} \mathrm{d}\boldsymbol{x}'\mu_p(\boldsymbol{x}') \doteq \frac{1}{(2\pi\eta^2)^{n/2}}\exp\left\{n\cdot\min_{\omega} q_n(\omega,\boldsymbol{x},\boldsymbol{y})\right\} \tag{A.18}$$

[4]Indeed, from (42)-(43) we see that $\boldsymbol{\Omega}$ is diagonally dominant matrix with positive diagonal elements, and thus positive-definite.

$$\triangleq \exp\left\{n \cdot h_n(\boldsymbol{x},\boldsymbol{y})\right\} \tag{A.19}$$

where $h_n(\boldsymbol{x},\boldsymbol{y})$ is defined in (46).

We next show that $\omega^* \in \mathbb{R}_+$, as claimed above. Indeed, the modulus of the integrand in (A.17) depends solely on the real part of the exponent of the integrand, namely, on $\mathrm{Re}\left\{q_n(\omega,\boldsymbol{x},\boldsymbol{y})\right\}$. Now, if we consider an arbitrary complex number $\omega = \omega_R + j\cdot\omega_I$, then we need to show that $\mathrm{Re}\left\{q_n(\omega,\boldsymbol{x},\boldsymbol{y})\right\}$ is maximized only at $\omega_I = 0$. Let $\bar{\boldsymbol{\Omega}}_1 \triangleq 2\omega^{-1}\cdot\boldsymbol{\Omega}_1$, and then by definition $\boldsymbol{\Omega} = \omega\cdot\bar{\boldsymbol{\Omega}}_1 + \boldsymbol{\Omega}_2$. Also, define $\mathbf{V} \triangleq (\omega_R\cdot\bar{\boldsymbol{\Omega}}_1+\boldsymbol{\Omega}_2)^2 + \omega_I^2\cdot\bar{\boldsymbol{\Omega}}_1^2$, where for a symmetric matrix $\mathbf{A}$, by $\mathbf{A}^2$ we mean $\mathbf{A}\mathbf{A}^T$. Recall that

$$q_n(\omega,\boldsymbol{x},\boldsymbol{y}) \triangleq \frac{1}{2}\log\pi - \frac{1}{2n}\log\det\boldsymbol{\Omega} + \frac{\omega^2}{4n}\boldsymbol{y}^T\mathbf{A}\boldsymbol{\Omega}^{-1}\mathbf{A}^T\boldsymbol{y} - \frac{\omega}{n}\psi(\boldsymbol{x},\boldsymbol{y}). \tag{A.20}$$

It is a simple exercise to check that $-\mathrm{Re}\left\{\log\det\boldsymbol{\Omega}\right\}$ is maximized at $\omega_I = 0$, and since the real part of the first and last terms in (A.20) are independent of ω_I, we focus on the real part of the third term, which after simple algebra boils down to

$$\mathrm{Re}\left\{\frac{\omega^2}{4}\boldsymbol{y}^T\mathbf{A}\boldsymbol{\Omega}^{-1}\mathbf{A}^T\boldsymbol{y}\right\} = \boldsymbol{y}^T\mathbf{A}\left[\mathbf{V}^{-1}\left(\omega_R^3\bar{\boldsymbol{\Omega}}_1 + \omega_R^2\boldsymbol{\Omega}_2 + \omega_R\omega_I^2\bar{\boldsymbol{\Omega}}_1 - \omega_I^2\boldsymbol{\Omega}_2\right)\right]\mathbf{A}^T\boldsymbol{y}. \tag{A.21}$$

To prove that (A.21) is maximized at $\omega_I = 0$ it suffices to show that

$$\mathbf{V}^{-1}\left(\omega_R^3\bar{\boldsymbol{\Omega}}_1 + \omega_R^2\boldsymbol{\Omega}_2 + \omega_R\omega_I^2\bar{\boldsymbol{\Omega}}_1 - \omega_I^2\boldsymbol{\Omega}_2\right) \preceq \left[\left(\omega_R\cdot\bar{\boldsymbol{\Omega}}_1+\boldsymbol{\Omega}_2\right)^2\right]^{-1}\left(\omega_R^3\bar{\boldsymbol{\Omega}}_1 + \omega_R^2\boldsymbol{\Omega}_2\right) \tag{A.22}$$

where for two matrices $\mathbf{X}$ and $\mathbf{Z}$, $\mathbf{X}\preceq\mathbf{Z}$ means that $\mathbf{Z}-\mathbf{X}$ is semi-positive definite matrix. Simple algebra reveals that to prove (A.22) it suffices to show that

$$\left(\omega_R^3\bar{\boldsymbol{\Omega}}_1 + \omega_R^2\boldsymbol{\Omega}_2\right)\bar{\boldsymbol{\Omega}}_1^2 \succeq \left(\omega_R\bar{\boldsymbol{\Omega}}_1 - \boldsymbol{\Omega}_2\right)\left(\omega_R\cdot\bar{\boldsymbol{\Omega}}_1+\boldsymbol{\Omega}_2\right)^2, \tag{A.23}$$

which follows by expanding the terms on the left and right hand side of (A.23), and using the fact that $\bar{\boldsymbol{\Omega}}_1$ and $\boldsymbol{\Omega}_2$ are semi-positive definite matrices.

Finally, we show that the convergence in (A.18) is uniform over $(\boldsymbol{x},\boldsymbol{y})\in\mathcal{H}_n(B)$, as claimed in the statement of Lemma 1. Precisely, we show that for any $\epsilon>0$, there exist an N_ϵ large enough and independent of $(\boldsymbol{x},\boldsymbol{y})$, such that for any $n > N_\epsilon$

$$\left|\frac{1}{n}\log\int_{\mathcal{V}(\boldsymbol{x},\boldsymbol{y})}\mathrm{d}\boldsymbol{x}'\mu_p(\boldsymbol{x}') - h_n(\boldsymbol{x},\boldsymbol{y})\right| < \epsilon. \tag{A.24}$$

Let $v(z)$ be an analytical function, and let $z_0\in\mathbb{R}$ be its unique saddle-point. By the analyticity of $v(\cdot)$, for any $\epsilon>0$, there exist a $\delta>0$ such that:

$$\left|v(z) - v(z_0) - \frac{1}{2}v''(z_0)(z-z_0)^2\right| \leq \epsilon(z-z_0)^2, \tag{A.25}$$

for any z such that $|z-z_0|<\delta$. Accordingly, it can be shown that [18]

$$\begin{aligned} e^{nv(z_0)}\cdot\left[\sqrt{\frac{2\pi}{|v''(z_0)+\epsilon|n}} - 2Q\left(\delta\sqrt{n\left(|v''(z_0)+\epsilon|\right)}\right)\right] &\leq \int_{z_0-j\infty}^{z_0+j\infty}\exp nv(z)\mathrm{d}z \\ &\leq e^{nv(z_0)}\sqrt{\frac{2\pi}{|v''(z_0)-\epsilon|n}} + C, \end{aligned} \tag{A.26}$$

for some constant C, and $Q(x) \triangleq \int_x^\infty (2\pi)^{-1/2} e^{-t^2/2} \mathrm{d}t$, for $x \in \mathbb{R}$. To prove (A.24), we show that the terms in the square brackets at the left and right hand sides of (A.26) can be lower and upper bounded by some universal constants. The proof consists of two steps. First, we will show that for any $\epsilon > 0$, there exists a universal $\delta > 0$ independent of $(\boldsymbol{x}, \boldsymbol{y})$, such that $q_n(\omega, \boldsymbol{x}, \boldsymbol{y})$ is close to its Taylor approximation around a ball of radius δ centered at the saddle-point. Then, we give a bound on $|q_n''(\omega, \boldsymbol{x}, \boldsymbol{y})|$, which will be independent of $(\boldsymbol{x}, \boldsymbol{y})$. The desired result will then follow by using the last fact, $|h''(z_0) \leq 0|$, and (A.26).

Recall that $q_n(\omega, \boldsymbol{x}, \boldsymbol{y})$ converges almost surely to a value $f(\omega)$, as $n \to \infty$, for any $(\boldsymbol{x}, \boldsymbol{y}) \in \mathcal{H}_n(B)$. Let $\omega_n^*(\boldsymbol{x}, \boldsymbol{y})$ be the sequence of saddle-points solving the minimization in (A.18). By the uniform continuity of $q_n(\omega, \boldsymbol{x}, \boldsymbol{y})$ w.r.t. $(\boldsymbol{x}, \boldsymbol{y})$, there is a radius $\Delta > 0$ such that for n sufficiently large, all saddle-points are contained in a ball centered at 0 and of radius Δ, i.e., $w_n^*(\boldsymbol{x}, \boldsymbol{y}) \in \mathcal{B}(0, \Delta)$ almost surely. Next, we define

$$M_\Delta \triangleq \max_{(\boldsymbol{x}, \boldsymbol{y}) \in \mathcal{H}_n(B)} \max_{|\omega - \omega_0| = \Delta} |q_n(\omega, \boldsymbol{x}, \boldsymbol{y})|. \tag{A.27}$$

By the Cauchy integration formula around the circle of radius Δ above, with parametrization γ, we can obtain the following expression of the Taylor approximation error $E(\omega, \boldsymbol{x}, \boldsymbol{y})$ [47, Section 3]

$$E(\omega, \boldsymbol{x}, \boldsymbol{y}) = q_n(\omega, \boldsymbol{x}, \boldsymbol{y}) - q_n(w^*, \boldsymbol{x}, \boldsymbol{y}) - \frac{1}{2} q_n''(w^*, \boldsymbol{x}, \boldsymbol{y})(\omega - w^*)^2 \tag{A.28}$$

$$= \sum_{k=3}^{\infty} \frac{(\omega - \omega^*)^k}{2\pi j} \oint_\gamma \frac{q_n(z, \boldsymbol{x}, \boldsymbol{y})}{(z - \omega^*)^{j+1}} \mathrm{d}z \tag{A.29}$$

where $\omega^* = \omega_n^*(\boldsymbol{x}, \boldsymbol{y})$. Now, let us consider a ball of radius $\delta < \Delta$ around the saddle-point, such that $\mathcal{B}(\omega^*, \delta) \subset \mathcal{B}(0, \Delta)$. Using (A.27) and (A.29), for any $\omega \in \mathcal{B}(\omega^*, \delta)$ we can bound $|R(\omega, \boldsymbol{x}, \boldsymbol{y})|$ as follows

$$|E(\omega, \boldsymbol{x}, \boldsymbol{y})| \leq \sum_{k=3}^{\infty} \frac{M_\Delta |\omega - \omega^*|^k}{\Delta^k} \tag{A.30}$$

$$\leq M_\Delta \frac{\delta^3}{\Delta^2(\Delta - \delta)} \triangleq \epsilon, \tag{A.31}$$

where the last step follows by evaluating the power series. Thus, for any $\epsilon > 0$, there exist a δ independent of $(\boldsymbol{x}, \boldsymbol{y})$ such that (A.25) is verified. Finally, by using Cauchy integration again, we have

$$|q_n''(\omega, \boldsymbol{x}, \boldsymbol{y})| = \frac{1}{\pi} \oint_\gamma \frac{q_n(z, \boldsymbol{x}, \boldsymbol{y})}{(z - \omega)^2} \tag{A.32}$$

$$\leq \frac{2M_\Delta}{\Delta^2}, \tag{A.33}$$

as required.

Appendix B
Proof of Theorem 2

In this section, we prove Theorem 2. Some technical details will be omitted, since they follow from similar steps used in the proof of Theorem 1. Fix an arbitrary $\epsilon > 0$, pick $p \in \mathbb{N}$, and a set of parameters

$\{\gamma_k\}_{k=0}^p \in \Gamma$. The codebook $\mathfrak{C}_n = \{\boldsymbol{x}_i\}_{i=1}^M$ is generated by drawing M codewords independently and uniformly at random over $\mathcal{T}_\varepsilon^n(\boldsymbol{\gamma}) \triangleq \mathcal{T}_\varepsilon^n$, defined in (13). Without loss of generality we assume that the codeword $\boldsymbol{x} \triangleq \boldsymbol{x}_1$ was sent. Then, similarly as in (36)-(37), we have

$$\bar{P}_e(n,R) \doteq \mathbb{E}\left[\mathcal{I}\left\{\mathcal{H}_n(B)\right\} \cdot \min\left\{1, M \cdot \Pr\left\{\frac{1}{n}\log V\left(\boldsymbol{Y}|\boldsymbol{X}_2\right) \geq \frac{1}{n}\log V\left(\boldsymbol{Y}|\boldsymbol{X}_1\right)\middle| \mathcal{F}_0\right\}\right\}\right]. \quad \text{(B.1)}$$

For $(\boldsymbol{X}_1, \boldsymbol{Y}) = (\boldsymbol{x}, \boldsymbol{y}) \in \mathcal{H}_n(B)$, the inner probability term at the r.h.s. of (B.1) can be represented as

$$\Pr\left\{\frac{1}{n}\log V\left(\boldsymbol{y}|\boldsymbol{X}_2\right) \geq \frac{1}{n}\log V\left(\boldsymbol{y}|\boldsymbol{x}\right)\right\} = \frac{1}{\mathrm{Vol}(\mathcal{T}_\varepsilon^n)}\int_{\boldsymbol{x}' \in \mathcal{V}(\boldsymbol{x},\boldsymbol{y}) \cap \mathcal{T}_\varepsilon^n} \mathrm{d}\boldsymbol{x}' \quad \text{(B.2)}$$

$$= \frac{\mathrm{Vol}\left(\mathcal{V}(\boldsymbol{x},\boldsymbol{y}) \cap \mathcal{T}_\varepsilon^n\right)}{\mathrm{Vol}(\mathcal{T}_\varepsilon^n)} \quad \text{(B.3)}$$

where

$$\mathcal{V}(\boldsymbol{x},\boldsymbol{y}) \cap \mathcal{T}_\varepsilon^n \triangleq \left\{\boldsymbol{x}' \in \mathcal{T}_\varepsilon^n : \ \frac{1}{n}\log V\left(\boldsymbol{y}|\boldsymbol{x}'\right) \geq \frac{1}{n}\log V\left(\boldsymbol{y}|\boldsymbol{x}\right)\right\} \quad \text{(B.4)}$$

$$= \left\{\boldsymbol{x}' \in \mathcal{T}_\varepsilon^n : \ \sum_{l=0}^K \alpha_l \sum_{t=1}^n y_t(x'_{t-l} - x_{t-l}) - \frac{1}{2}\left\|\boldsymbol{\alpha}\right\|_2^2 \sum_{t=1}^n (x_t'^2 - x_t^2) \right.$$

$$\left. - \sum_{l=1}^K \Pi_l(\boldsymbol{\alpha}) \sum_{t=1}^n (x'_t x'_{t-l} - x_t x_{t-l}) \geq 0\right\}. \quad \text{(B.5)}$$

Due to the fact that $\boldsymbol{x}', \boldsymbol{x} \in \mathcal{T}_\varepsilon^n$,

$$\left|\sum_{t=1}^n (x'_t x'_{t-l} - x_t x_{t-l})\right| \leq 2n\varepsilon, \ \ l \in [0:p], \quad \text{(B.6)}$$

and therefore,

$$\mathcal{V}^-(\boldsymbol{x},\boldsymbol{y}) \cap \mathcal{T}_\varepsilon^n \subseteq \mathcal{V}(\boldsymbol{x},\boldsymbol{y}) \cap \mathcal{T}_\varepsilon^n \subseteq \mathcal{V}^+(\boldsymbol{x},\boldsymbol{y}) \cap \mathcal{T}_\varepsilon^n, \quad \text{(B.7)}$$

in which

$$\mathcal{V}^+(\boldsymbol{x},\boldsymbol{y}) \triangleq \left\{\boldsymbol{x}' \in \mathbb{R}^n : \ \sum_{l=0}^K \alpha_l \sum_{t=1}^n y_t(x'_{t-l} - x_{t-l}) \right.$$

$$\left. - \sum_{l=1+p\wedge K}^K \Pi_l(\boldsymbol{\alpha}) \sum_{t=1}^n (x'_t x'_{t-l} - x_t x_{t-l}) + \left\|\boldsymbol{\alpha}\right\|_2^2 n\varepsilon + 2n\varepsilon \sum_{l=1}^K \Pi_l(\boldsymbol{\alpha}) \geq 0\right\}$$

$$\triangleq \left\{\boldsymbol{x}' \in \mathbb{R}^n : \ \psi'(\boldsymbol{x}',\boldsymbol{y}) - \psi'(\boldsymbol{x},\boldsymbol{y}) + nC^+(\boldsymbol{\alpha})\varepsilon \geq 0\right\} \quad \text{(B.8)}$$

where

$$\psi'(\boldsymbol{x},\boldsymbol{y}) \triangleq \sum_{l=0}^K \alpha_l \sum_{t=1}^n y_t x_{t-l} - \sum_{l=1+p\wedge K}^K \Pi_l(\boldsymbol{\alpha}) \sum_{t=1}^n x_t x_{t-l} \quad \text{(B.9)}$$

$$C^+(\boldsymbol{\alpha}) \triangleq \|\boldsymbol{\alpha}\|^2 + 2\sum_{l=1}^K \Pi_l(\boldsymbol{\alpha}), \quad \text{(B.10)}$$

and $\mathcal{V}^-(\boldsymbol{x},\boldsymbol{y})$ is defined similarly to (B.8) but with $C^+(\boldsymbol{\alpha})$ replaced by $C^-(\boldsymbol{\alpha}) \triangleq -C^+(\boldsymbol{\alpha})$.

The bulk of the argument resides in understanding the exponential behavior of (B.3). Using (B.7), we note that

$$\mathrm{Vol}\left(\mathcal{V}^{-}(\boldsymbol{x},\boldsymbol{y})\cap\mathcal{T}_{\varepsilon}^{n}\right)\leq\mathrm{Vol}\left(\mathcal{V}(\boldsymbol{x},\boldsymbol{y})\cap\mathcal{T}_{\varepsilon}^{n}\right)\leq\mathrm{Vol}\left(\mathcal{V}^{+}(\boldsymbol{x},\boldsymbol{y})\cap\mathcal{T}_{\varepsilon}^{n}\right). \tag{B.11}$$

In the sequel, we study the upper bound, with the understanding that the lower bound follows from similar steps. We write

$$\begin{aligned}\mathrm{Vol}\left(\mathcal{V}^{+}(\boldsymbol{x},\boldsymbol{y})\cap\mathcal{T}_{\varepsilon}^{n}\right)=\int_{\boldsymbol{x}'\in\mathbb{R}^n}&\mathbb{1}\left\{\psi'(\boldsymbol{x}',\boldsymbol{y})-\psi'(\boldsymbol{x},\boldsymbol{y})+n\varepsilon C^{+}(\boldsymbol{\alpha})\geq 0\right\}\\ &\cdot\prod_{k=0}^{p}\mathbb{1}\left\{\left|\frac{1}{n}\sum_{t=1}^{n}x_t'x_{t-k}'-\gamma_k\right|\leq\varepsilon\right\}\mathrm{d}\boldsymbol{x}'.\end{aligned} \tag{B.12}$$

Recalling that $\mathbb{1}\left\{|a|<\varepsilon\right\}=\left[\mathbb{1}\left\{a\leq\varepsilon\right\}-\mathbb{1}\left\{a\leq-\varepsilon\right\}\right]$, and using (A.1), we get

$$\begin{aligned}\mathrm{Vol}\left(\mathcal{V}^{+}(\boldsymbol{x},\boldsymbol{y})\cap\mathcal{T}_{\varepsilon}^{n}\right)\doteq\int_{\boldsymbol{x}'\in\mathbb{R}^n}\mathrm{d}\boldsymbol{x}'\int_{c_{p+1}-j\infty}^{c_{p+1}+j\infty}&\mathrm{d}\omega_{p+1}\frac{\exp\{\omega_{p+1}(\psi'(\boldsymbol{x}',\boldsymbol{y})-\psi'(\boldsymbol{x},\boldsymbol{y})+n\varepsilon C^{+}(\boldsymbol{\alpha}))\}}{\omega_{p+1}}\\ \cdot\prod_{k=0}^{p}\int_{c_k-j\infty}^{c_k+j\infty}&\mathrm{d}\omega_k\exp\left\{\omega_k\left(n\gamma_k-\sum_{t=1}^{n}x_t'x_{t-k}'\right)\right\}\left(\frac{e^{n\omega_k\varepsilon}-e^{-n\omega_k\varepsilon}}{w_k}\right).\end{aligned} \tag{B.13}$$

Exchanging the order of integrals and collecting terms, we obtain

$$\begin{aligned}\mathrm{Vol}\left(\mathcal{V}^{+}(\boldsymbol{x},\boldsymbol{y})\cap\mathcal{T}_{\varepsilon}^{n}\right)\doteq&\int_{c_{p+1}-j\infty}^{c_{p+1}+j\infty}\mathrm{d}\omega_{p+1}\frac{1}{\omega_{p+1}}\exp\{-\omega_{p+1}\psi'(\boldsymbol{x},\boldsymbol{y})+n\varepsilon C^{+}(\boldsymbol{\alpha})\}\\ &\cdot\int_{\bar{\mathcal{W}}}\mathrm{d}\boldsymbol{\omega}\left[\prod_{k=0}^{p}\left(\frac{e^{n\omega_k\varepsilon}-e^{-n\omega_k\varepsilon}}{w_k}\right)\exp\left\{n\omega_k\gamma_k\right\}\right]\\ &\cdot\int_{\boldsymbol{x}'\in\mathbb{R}^n}\mathrm{d}\boldsymbol{x}'\exp\{\omega_{p+1}\psi'(\boldsymbol{x}',\boldsymbol{y})\}\prod_{k=0}^{p}\exp\left\{-\omega_k\sum_{t=1}^{n}x_t'x_{t-k}'\right\},\end{aligned} \tag{B.14}$$

where $\bar{\mathcal{W}}\triangleq\{\boldsymbol{\omega}:c_k-j\infty\leq\omega_k\leq c_k+j\infty,\ k=0,\ldots,p\}$. Next, we note that

$$\begin{aligned}\omega_{p+1}\psi'(\boldsymbol{x}',\boldsymbol{y})-\sum_{l=0}^{p}\omega_l\sum_{t=1}^{n}x_t'x_{t-l}'&=\omega_{p+1}\sum_{l=0}^{K}\alpha_l\sum_{t=1}^{n}y_tx_{t-l}'-\omega_{p+1}\cdot\sum_{l=1+p\wedge K}^{K}\Pi_l(\boldsymbol{\alpha})\sum_{t=1}^{n}x_t'x_{t-l}'\\ &\quad-\sum_{l=0}^{p}\omega_l\sum_{t=1}^{n}x_t'x_{t-l}' &\text{(B.15)}\\ &=\omega_{p+1}\cdot\boldsymbol{y}^T\boldsymbol{A}\boldsymbol{x}'-\boldsymbol{x}'^T\boldsymbol{\Omega}_0\boldsymbol{x}' &\text{(B.16)}\\ &=-\left(\boldsymbol{x}'+\frac{\omega_{p+1}}{2}\boldsymbol{\Omega}_0^{-1}\boldsymbol{A}^T\boldsymbol{y}\right)^T\boldsymbol{\Omega}_0\left(\boldsymbol{x}'+\frac{\omega_{p+1}}{2}\boldsymbol{\Omega}_0^{-1}\boldsymbol{A}^T\boldsymbol{y}\right)\\ &\quad+\frac{\omega_{p+1}^2}{4}\boldsymbol{y}^T\boldsymbol{A}\boldsymbol{\Omega}_0^{-1}\boldsymbol{A}^T\boldsymbol{y}, &\text{(B.17)}\end{aligned}$$

where the last step follows by completing the square, and $\mathbf{\Omega}_0$ is the $n \times n$ symmetric Toeplitz matrix,

$$[\mathbf{\Omega}_0]_{i,j} \triangleq \begin{cases} \omega_0, & \text{for } |i-j| = 0 \\ 0.5 \cdot \omega_{|i-j|}, & \text{for } 1 \leq |i-j| \leq p \\ 0.5 \cdot \omega_{p+1} \cdot \Pi_{|i-j|}(\boldsymbol{\alpha}), & \text{for } 1 \leq p \wedge K \leq |i-j| \leq K \\ 0, & \text{otherwise} \end{cases}. \tag{B.18}$$

Therefore,

$$\int_{\boldsymbol{x}' \in \mathbb{R}^n} \mathrm{d}\boldsymbol{x}' \exp\{\omega_{p+1}\psi(\boldsymbol{x}', \boldsymbol{y})\} \prod_{l=0}^{p} \exp\left\{-\omega_l \sum_{t=1}^{n} x'_t x'_{t-l}\right\}$$
$$= \exp\left\{\frac{n}{2}\log\pi - \frac{1}{2}\log\det\mathbf{\Omega}_0 + \frac{\omega_{p+1}^2}{4}\boldsymbol{y}^T\boldsymbol{A}\mathbf{\Omega}_0^{-1}\boldsymbol{A}^T\boldsymbol{y}\right\}$$

where in the last equality we have used the fact that $\mathbf{\Omega}_0$ is a symmetric positive-definite Toeplitz matrix which follows because $\boldsymbol{\omega} \in \mathcal{W}$. Using the last result, we finally get

$$\mathrm{Vol}\left(\mathcal{V}^+(\boldsymbol{x}, \boldsymbol{y}) \cap \mathcal{T}_\varepsilon^n\right) \doteq \int_{\mathcal{W}} \mathrm{d}\boldsymbol{\omega} \exp\{n\bar{q}_n(\boldsymbol{\omega}, \boldsymbol{x}, \boldsymbol{y}, \varepsilon)\}, \tag{B.19}$$

where

$$\bar{q}_n(\boldsymbol{\omega}, \boldsymbol{x}, \boldsymbol{y}, \varepsilon) = \frac{\log\pi}{2} - \frac{1}{2n}\log\det\mathbf{\Omega}_0 + \frac{\omega_{p+1}^2}{4n}\boldsymbol{y}^T\boldsymbol{A}\mathbf{\Omega}_0^{-1}\boldsymbol{A}^T\boldsymbol{y} - \frac{\omega_{p+1}}{n}\psi'(\boldsymbol{x}, \boldsymbol{y}) + \sum_{l=0}^{p}\omega_l\gamma_l$$
$$+ \frac{1}{n}\sum_{l=0}^{p}\log\sinh(n\omega_l\varepsilon) + \varepsilon C^+(\boldsymbol{\alpha}). \tag{B.20}$$

Using similar steps as in Appendix A, the integral in (B.19) can be evaluated using the saddle-point method, resulting in

$$\mathrm{Vol}\left(\mathcal{V}^+(\boldsymbol{x}, \boldsymbol{y}) \cap \mathcal{T}_\varepsilon^n\right) \doteq \exp\left\{n\bar{h}_n(\boldsymbol{x}, \boldsymbol{y}, \varepsilon)\right\}, \tag{B.21}$$

where $\bar{h}_n(\boldsymbol{x}, \boldsymbol{y}, \varepsilon) = \min_{\boldsymbol{\omega} \in \mathcal{W}} q_n(\boldsymbol{\omega}, \varepsilon)$. Also, using the same steps as in (A.19)-(A.33), one can show that the saddle-point solution is in fact real-valued vector, and that the convergence (B.21) is uniform w.r.t. $(\boldsymbol{x}, \boldsymbol{y})$. Similarly, one can verify that

$$\lim_{\varepsilon\to 0}\lim_{n\to\infty}\frac{1}{n}\log\mathrm{Vol}(\mathcal{T}_\varepsilon^n) = \frac{1}{2}\log\left(2\pi e\eta^2\right). \tag{B.22}$$

Thus, we may conclude that

$$\bar{P}_e(n, R) \doteq \mathbb{E}\left[\mathcal{I}\left\{\mathcal{H}_n(B)\right\}\exp\left\{-n\left[\frac{1}{2}\log\left(2\pi e\eta^2\right) - \bar{h}_n(\boldsymbol{X}, \boldsymbol{Y}, \varepsilon) - R + o(\varepsilon)\right]_+\right\}\right]. \tag{B.23}$$

The next step is taking the expectation w.r.t. $(\boldsymbol{X}, \boldsymbol{Y})$, distributed according to $\mu_p \times W$. Similarly as in the proof of Theorem 1, we calculate the limit of each term in $\bar{q}_n(\boldsymbol{\omega}, \boldsymbol{x}, \boldsymbol{y}, \varepsilon)$. By Szegö theorem,

$$\frac{1}{n}\log\det\mathbf{\Omega}_0 = \frac{1}{2\pi}\int_0^{2\pi}\log g_{\boldsymbol{\omega}}(\nu)d\nu, \tag{B.24}$$

where $g_{\boldsymbol{\omega}}(\cdot)$ is defined in (15). Using the same steps as in (52)-(61), by the LLN and Szegö theorem, we get with overwhelming probability as $n \to \infty$,

$$\lim_{n\to\infty} \frac{1}{n}\boldsymbol{y}^T\mathbf{A}\boldsymbol{\Omega}_0^{-1}\mathbf{A}^T\boldsymbol{y} = \lim_{n\to\infty} \frac{1}{n}\mathbb{E}\left\{\boldsymbol{Y}^T\mathbf{A}\boldsymbol{\Omega}_0^{-1}\mathbf{A}^T\boldsymbol{Y}\right\} \tag{B.25}$$

$$= \lim_{n\to\infty} \frac{1}{n}\operatorname{tr}\left\{\mathbf{A}\boldsymbol{\Omega}_0^{-1}\mathbf{A}^T\mathbf{R}_{\boldsymbol{y}}\right\} \tag{B.26}$$

$$= \frac{1}{2\pi}\int_{[0,2\pi]} \mathrm{d}\nu \frac{S_Y(\nu)\left|A(\nu)\right|^2}{g_{\boldsymbol{\omega}}(\nu)}, \tag{B.27}$$

and

$$\lim_{n\to\infty} \frac{1}{n}\psi'(\boldsymbol{x},\boldsymbol{y}) = \sum_{l=0}^{K}\sum_{i=0}^{K}\alpha_l h_i \gamma_{l-i} - \sum_{l=1+p\wedge K}^{K}\Pi_l(\boldsymbol{\alpha})\gamma_l. \tag{B.28}$$

Also, recalling the fact that $\lim_{x\to\infty}(\log\sinh(x))/x = 1$,

$$\lim_{n\to\infty} \frac{1}{n}\sum_{l=0}^{p}\log\sinh(n\omega_l\varepsilon) = \varepsilon\sum_{l=0}^{p}\omega_l. \tag{B.29}$$

Collecting the last result, with probability approaching 1 as $n \to \infty$,

$$\lim_{\varepsilon\to 0}\lim_{n\to\infty}\bar{h}_n(\boldsymbol{X},\boldsymbol{Y},\varepsilon) = \min_{\boldsymbol{\omega}\in\Gamma}\left\{\frac{\log\pi}{2} - \frac{1}{4\pi}\int_0^{2\pi}\log g_{\boldsymbol{\omega}}(\nu)d\nu + \frac{\omega_{p+1}^2}{8\pi}\int_{[0,2\pi]}\mathrm{d}\nu\frac{S_Y(\nu)\left|A(\nu)\right|^2}{g_{\boldsymbol{\omega}}(\nu)}\right.$$
$$\left. - \omega_{p+1}\left[\sum_{l=0}^{K}\sum_{i=0}^{K}\alpha_l h_i\gamma_{l-i} - \sum_{l=1+p\wedge K}^{K}\Pi_l(\boldsymbol{\alpha})\gamma_l\right] + \sum_{k=0}^{p}\omega_k\gamma_k\right\} \tag{B.30}$$
$$\triangleq -\hat{I}_2(\boldsymbol{\alpha},\boldsymbol{\gamma}). \tag{B.31}$$

Next, for any $\delta > 0$, define the set

$$\mathcal{A}_n(\delta) \triangleq \left\{(\boldsymbol{x},\boldsymbol{y}):\ \left|\bar{h}_n(\boldsymbol{x},\boldsymbol{y},\varepsilon) + \hat{I}_2(\boldsymbol{\alpha},\boldsymbol{\gamma})\right| < \delta\right\}. \tag{B.32}$$

Accordingly, using (B.23) we have

$$\bar{P}_e(n,R) \le \mathbb{E}\left[\mathcal{I}\left\{\mathcal{A}_n(\delta)\right\}\exp\left\{-n\left[\frac{1}{2}\log\left(2\pi e\eta^2\right) - \bar{h}_n(\boldsymbol{X},\boldsymbol{Y},\varepsilon) - R + o(\varepsilon)\right]_+\right\}\right] + \Pr\left\{\mathcal{A}_n^c(\delta)\right\}, \tag{B.33}$$

and in light of (B.31), taking $\delta \to 0$ and $\varepsilon \to 0$ followed by $n \to \infty$, the last term at the r.h.s. of (B.33) is asymptotically negligible. Hence, in terms of achievable rate, we get that $\bar{P}_e(n,R)$ decays to zero as long as

$$R < \frac{1}{2}\log\left(2\pi e\eta^2\right) + \hat{I}_2(\boldsymbol{\alpha},\boldsymbol{\gamma}) = \bar{I}_2(\boldsymbol{\alpha},\boldsymbol{\gamma}). \tag{B.34}$$

Finally, since $\{\gamma_k\}_{k=1}^p \in \Gamma$ were fixed parameters up to this point, we take the maximum of the r.h.s. of the above inequality over these parameters, which concludes the proof.

APPENDIX C
PROOF OF THEOREM 3

We derive the error exponent in the case where $\alpha_k = 0$ for $1 \leq k \leq K$, using the ensemble in (13) with $p = 0$. Some technical details will be omitted, as they follow from similar steps used in the proof of Theorem 1. We use (B.23) in Appendix B, and we note that in the above special case (B.23) simplifies to

$$\bar{P}_e(n,R) \doteq \mathbb{E}\left[\mathcal{I}\left\{\mathcal{H}_n(B)\right\}\exp\left\{-n\left[I(\hat{\rho}_{XY},h_0,\alpha_0)-R+o(\varepsilon)\right]_+\right\}\right] \tag{C.1}$$

where $\hat{\rho}_{XY}$ is the empirical correlation coefficient defined as

$$\hat{\rho}_{XY} \triangleq \frac{\frac{1}{n}\sum_{t=1}^n X_tY_t}{\sqrt{P_X}\sqrt{\frac{1}{n}\sum_{t=1}^n Y_t^2}}, \tag{C.2}$$

and $I(\hat{\rho}_{XY},h_0,\alpha_0)$ is defined in (19). The next step is taking the expectation w.r.t. $(\boldsymbol{X},\boldsymbol{Y})$ distributed according to $\mu\times W$. To this end, let $\upsilon>0$, and define the set

$$\begin{aligned}\mathcal{L}_{\upsilon,\varepsilon}(P_X,P_Y,\rho_{XY}) \triangleq \Bigg\{\boldsymbol{x},\boldsymbol{y}\in\mathbb{R}^n:\ &\left|\frac{1}{n}\sum_{t=1}^n x_t^2-P_X\right|<\varepsilon,\ \left|\frac{1}{n}\sum_{t=1}^n y_t^2-P_Y\right|<\upsilon,\\ &\left|\frac{1}{n}\sum_{t=1}^n x_ty_t-\rho_{XY}\sqrt{P_XP_Y}\right|<\upsilon\Bigg\}.\end{aligned} \tag{C.3}$$

With this definition we see that the exponent term in (C.1) is almost constant over $\mathcal{L}_{\upsilon,\varepsilon}(P_X,P_Y,\rho_{XY})$, namely, $|I(\hat{\rho}_{XY})-I(\rho_{XY})|<\xi(\upsilon)$ with $\xi(\upsilon)\to 0$ as $\upsilon\to 0$. Then, we may write

$$\begin{aligned}\bar{P}_e(n,R) &\doteq \int_{\mathcal{H}_n(B)}\mathrm{d}\boldsymbol{x}\mathrm{d}\boldsymbol{y}\exp\left\{-n\left[I(\hat{\rho}_{XY})-R+o(\varepsilon)\right]_+\right\}\mu(\boldsymbol{x})W(\boldsymbol{y}|\boldsymbol{x})\\ &\doteq \int_{P_Y\in[0,B],|\rho_{XY}|\leq 1}\mathrm{d}P_Y\mathrm{d}\rho_{XY}\int_{\mathcal{H}_n(B)\cap\mathcal{L}_{\upsilon,\varepsilon}(P_X,P_Y,\rho_{XY})}\mathrm{d}\boldsymbol{x}\mathrm{d}\boldsymbol{y}\exp\left\{-n\left[I(\hat{\rho}_{XY})-R+o(\varepsilon)\right]_+\right\}\mu(\boldsymbol{x})W(\boldsymbol{y}|\boldsymbol{x})\\ &\doteq \max_{P_Y,\rho_{XY}}\int_{\mathcal{H}_n(B)\cap\mathcal{L}_{\upsilon,\varepsilon}(P_X,P_Y,\rho_{XY})}\mathrm{d}\boldsymbol{x}\mathrm{d}\boldsymbol{y}\exp\left\{-n\left[I(\hat{\rho}_{XY})-R+o(\varepsilon)\right]_+\right\}\mu(\boldsymbol{x})W(\boldsymbol{y}|\boldsymbol{x})\\ &\doteq \max_{P_Y,\rho_{XY}}\exp\left\{-n\left[I(\rho_{XY})+\xi(\upsilon)-R+o(\varepsilon)\right]_+-\frac{n}{2}\log(2\pi eP_X)\right\}\\ &\qquad\times\int_{\mathcal{H}_n(B)\cap\mathcal{L}_{\upsilon,\varepsilon}(P_X,P_Y,\rho_{XY})}\mathrm{d}\boldsymbol{x}\mathrm{d}\boldsymbol{y}W(\boldsymbol{y}|\boldsymbol{x})\end{aligned} \tag{C.4}$$

where the third (asymptotic) equality follows from the Laplace integration method. We next evaluate the integral term at the r.h.s. of (C.4) using the saddle-point method as in (B.12)-(B.14). We get

$$\begin{aligned}\int_{\mathcal{H}_n(B)\cap\mathcal{L}_{\upsilon,\varepsilon}}\mathrm{d}\boldsymbol{x}\mathrm{d}\boldsymbol{y}W(\boldsymbol{y}|\boldsymbol{x}) &\doteq \int_{\hat{\mathcal{W}}}\mathrm{d}\hat{\boldsymbol{\omega}}\exp\left\{n\left[\hat{\omega}_0P_X+\hat{\omega}_1\rho_{XY}\sqrt{P_XP_Y}+\hat{\omega}_2P_Y+o(\varepsilon)+o(\upsilon)\right]\right\}\\ &\quad\times\int_{\mathbb{R}^n\times\mathbb{R}^n}\mathrm{d}\boldsymbol{x}\mathrm{d}\boldsymbol{y}W(\boldsymbol{y}|\boldsymbol{x})\exp\left\{-\hat{\omega}_0\left\|\boldsymbol{x}\right\|_2^2-\hat{\omega}_1\sum_{t=1}^n x_ty_t-\hat{\omega}_2\left\|\boldsymbol{y}\right\|_2^2\right\}.\end{aligned} \tag{C.5}$$

Define

$$[\mathbf{H}_1]_{i,j} \triangleq \begin{cases} \hat{\omega}_0 + \frac{1}{2}\left\|\boldsymbol{h}\right\|_2^2, & \text{for } |i-j| = 0 \\ 0.5 \cdot \Pi_{|i-j|}(\boldsymbol{h}), & \text{for } 1 \leq |i-j| \leq K \cdot \\ 0, & \text{otherwise} \end{cases} \tag{C.6}$$

and let $\mathbf{H}_2$ be an $n \times n$ lower-triangular Toeplitz matrix where for $i \geq j$,

$$[\mathbf{H}_2]_{i,j} \triangleq \begin{cases} \hat{\omega}_1 - h_0, & \text{for } 0 \leq i-j = 0 \\ h_{i-j}, & \text{for } 1 \leq i-j \leq K \cdot \\ 0, & \text{otherwise} \end{cases} \tag{C.7}$$

Then, we have

$$\begin{aligned} \log W(\boldsymbol{y}|\boldsymbol{x}) - \hat{\omega}_0 \left\|\boldsymbol{x}\right\|_2^2 - \hat{\omega}_1 \sum_{t=1}^n x_t y_t - \hat{\omega}_2 \left\|\boldsymbol{y}\right\|_2^2 &= -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\left\|\boldsymbol{y}\right\|_2^2 + \sum_{k=0}^K h_k \sum_{t=1}^n y_t x_{t-k} \\ &\quad - \frac{1}{2}\left\|\boldsymbol{h}\right\|_2^2 \left\|\boldsymbol{x}\right\|_2^2 - \sum_{k=1}^K \Pi_k(\boldsymbol{h}) \sum_{t=1}^n x_t x_{t-k} - \hat{\omega}_0 \left\|\boldsymbol{x}\right\|_2^2 - \hat{\omega}_1 \sum_{t=1}^n x_t y_t - \hat{\omega}_2 \left\|\boldsymbol{y}\right\|^2 & \text{(C.8)} \\ &= -\frac{n}{2}\log(2\pi\sigma^2) - \left(\frac{1}{2\sigma^2} + \hat{\omega}_2\right)\boldsymbol{y}^T\boldsymbol{y} - \boldsymbol{x}^T\mathbf{H}_1\boldsymbol{x} - \boldsymbol{y}^T\mathbf{H}_2\boldsymbol{x} & \text{(C.9)} \\ &= -\frac{n}{2}\log(2\pi\sigma^2) - \left[\lambda\left\|\boldsymbol{y} + \frac{1}{2\lambda}\mathbf{H}_2\boldsymbol{x}\right\|_2^2 + \boldsymbol{x}^T\left(\mathbf{H}_1 - \frac{1}{4\lambda}\mathbf{H}_2^T\mathbf{H}_2\right)\boldsymbol{x}\right] & \text{(C.10)} \end{aligned}$$

where $\lambda \triangleq \frac{1}{2\sigma^2} + \hat{\omega}_2$. Therefore,

$$\begin{aligned} &\int_{\mathbb{R}^{\kappa} \times \mathbb{R}^n} \mathrm{d}\boldsymbol{x}\mathrm{d}\boldsymbol{y} W(\boldsymbol{y}|\boldsymbol{x}) \exp\left\{-\hat{\omega}_0\left\|\boldsymbol{x}\right\|_2^2 - \hat{\omega}_1 \sum_{t=1}^n x_t y_t - \hat{\omega}_2\left\|\boldsymbol{y}\right\|_2^2\right\} \\ &\qquad = \exp\left\{-\frac{n}{2}\log(2\pi\sigma^2) + n\log\pi - \frac{1}{2}\log\det\left(\frac{1}{4}\mathbf{H}_2^T\mathbf{H}_2 - \lambda\cdot\mathbf{H}_1\right)\right\} \end{aligned} \tag{C.11}$$

where the last equality follows because $\hat{\boldsymbol{\omega}} \in \hat{\mathcal{W}}$. Using the last result, we finally get

$$\int_{\mathcal{H}_n(B)\cap\mathcal{L}_{\upsilon,\varepsilon}} \mathrm{d}\boldsymbol{x}\mathrm{d}\boldsymbol{y} W(\boldsymbol{y}|\boldsymbol{x}) \doteq \int_{\mathcal{W}} \mathrm{d}\hat{\boldsymbol{\omega}} \exp\left\{n\left[\tilde{q}_n(\hat{\boldsymbol{\omega}},\upsilon) - \frac{1}{2}\log(2\pi\sigma^2) + \log\pi + o(\varepsilon) + o(\upsilon)\right]\right\}, \tag{C.12}$$

where

$$\tilde{q}_n(\hat{\boldsymbol{\omega}}, P_Y, \rho_{XY}, \upsilon) = \hat{\omega}_0 P_X + \hat{\omega}_1 \rho_{XY}\sqrt{P_X P_Y} + \hat{\omega}_2 P_Y - \frac{1}{2n}\log\det\left(\frac{1}{4}\mathbf{H}_2^T\mathbf{H}_2 - \lambda\cdot\mathbf{H}_1\right). \tag{C.13}$$

Using similar steps as in Appendix A, the integral in (B.19) can be evaluated using the saddle-point method, resulting in

$$\int_{\mathcal{H}_n(B)\cap\mathcal{L}_{\upsilon,\varepsilon}} \mathrm{d}\boldsymbol{x}\mathrm{d}\boldsymbol{y} W(\boldsymbol{y}|\boldsymbol{x}) \doteq \exp\left\{n\left[\tilde{h}_n(P_Y, \rho_{XY}, \upsilon) - \frac{1}{2}\log(2\pi\sigma^2) + \log\pi + o(\varepsilon) + o(\upsilon)\right]\right\}, \tag{C.14}$$

where $\tilde{h}_n(P_Y,\rho_{XY},\upsilon)=\min_{\hat{\boldsymbol{\omega}}\in\hat{\mathcal{W}}}\tilde{q}_n(\hat{\boldsymbol{\omega}},P_Y,\rho_{XY},\upsilon)$. Also, using the same steps as in (A.19)-(A.33), one can show that the saddle-point solution is in fact real-valued vector. Substituting the last result in (C.4), we get

$$\begin{aligned}\bar{P}_e(n,R) &\doteq \max_{P_Y,\rho_{XY}}\exp\left\{-n\left[I(\rho_{XY})+\xi(\upsilon)-R+o(\varepsilon)\right]_+-\frac{n}{2}\log(2\pi e P_X)\right\}\\ &\quad\times\exp\left\{n\left[\tilde{h}_n(P_Y,\rho_{XY},\upsilon)-\frac{1}{2}\log(2\pi\sigma^2)+\log\pi+o(\varepsilon)+o(\upsilon)\right]\right\} \qquad \text{(C.15)}\\ &=\max_{P_Y,\rho_{XY}}\exp\left\{-n\left[I(\rho_{XY})+\xi(\upsilon)-R+o(\varepsilon)\right]_+-\frac{n}{2}\log(4eP_X\sigma^2)\right.\\ &\qquad\qquad\left.+n\tilde{h}_n(P_Y,\rho_{XY},\upsilon)\right\}. \qquad \text{(C.16)}\end{aligned}$$

Now, using Szegö theorem,

$$\lim_{n\to\infty}\frac{1}{n}\log\det\left(\frac{1}{4}\mathbf{H}_2^T\mathbf{H}_2-\lambda\cdot\mathbf{H}_1\right)=\frac{1}{2\pi}\int_0^{2\pi}\log u_{\hat{\boldsymbol{\omega}}}(\nu)d\nu, \qquad \text{(C.17)}$$

where $u_{\hat{\boldsymbol{\omega}}}(\cdot)$ is defined in (17), and so,

$$\begin{aligned}\lim_{\upsilon\to0}\lim_{n\to\infty}\tilde{h}_n(P_Y,\rho_{XY},\upsilon) &= \min_{\hat{\boldsymbol{\omega}}\in\hat{\mathcal{W}}}\left\{\hat{\omega}_0P_X+\hat{\omega}_1\rho_{XY}\sqrt{P_XP_Y}+\hat{\omega}_2P_Y-\frac{1}{4\pi}\int_0^{2\pi}\log u_{\hat{\boldsymbol{\omega}}}(\nu)d\nu\right\}\\ &\triangleq -V(P_Y,\rho_{XY}). \qquad \text{(C.18)}\end{aligned}$$

Using (C.16) and (C.18), we finally get

$$\lim_{\upsilon,\varepsilon\to0}\lim_{n\to\infty}-\frac{1}{n}\log\bar{P}_e(n,R)=\min_{P_Y,\rho_{XY}}\left\{V(P_Y,\rho_{XY})+\frac{1}{2}\log(4eP_X\sigma^2)+\left[I(\rho_{XY})-R\right]_+\right\}. \qquad \text{(C.19)}$$

Appendix D

Proof of Theorem 4

Due to similarities to the proofs of Theorems 1 and 2, we provide only a proof sketch. Fix an arbitrary $\varepsilon>0$. Define the sequence of sets $\mathcal{T}_\varepsilon^n(P_X)$, for $n=1,2,\ldots$, as follows

$$\mathcal{T}_\varepsilon^n(P_X)\triangleq\left\{\boldsymbol{x}\in\mathbb{R}^n:\ \left|\frac{1}{n}\sum_{t=1}^n x_t^2-P_X\right|<\varepsilon\right\}. \qquad \text{(D.1)}$$

The codebook $\mathfrak{C}_n$ is generated by drawing M codewords independently and uniformly at random from $\mathcal{T}_\varepsilon^n(P_X)$. Then, the probability of error corresponding to the universal decoder in (23) is given by

$$\bar{P}_e(n,R)\doteq\mathbb{E}\left[\mathcal{I}\left\{\mathcal{H}_n(B)\right\}\cdot\min\left\{1,M\cdot\Pr\left\{\left|\boldsymbol{X}_2^T\boldsymbol{Y}\right|\geq\left|\boldsymbol{X}_1^T\boldsymbol{Y}\right|\middle|\mathcal{F}_0\right\}\right\}\right]. \qquad \text{(D.2)}$$

For $(\boldsymbol{X}_1,\boldsymbol{Y})=(\boldsymbol{x},\boldsymbol{y})\in\mathcal{H}_n(B)$, the inner probability term at the r.h.s. of (D.2) can be represented as

$$\begin{aligned}\Pr\left\{\left|\boldsymbol{X}_2^T\boldsymbol{y}\right|\geq\left|\boldsymbol{x}^T\boldsymbol{y}\right|\right\} &= 2\cdot\Pr\left\{\boldsymbol{X}_2^T\boldsymbol{y}\geq\left|\boldsymbol{x}^T\boldsymbol{y}\right|\right\} \qquad \text{(D.3)}\\ &\doteq\frac{\mathrm{Vol}\left(\mathcal{V}(\boldsymbol{x},\boldsymbol{y})\cap\mathcal{T}_\varepsilon^n(P_X)\right)}{\mathrm{Vol}(\mathcal{T}_\varepsilon^n(P_X))} \qquad \text{(D.4)}\end{aligned}$$

where

$$\mathcal{V}(\boldsymbol{x},\boldsymbol{y})\triangleq\left\{\boldsymbol{x}'\in\mathbb{R}^n:\ \sum_{i=1}^n x_i'y_i\geq\left|\sum_{i=1}^n x_iy_i\right|\right\}. \qquad \text{(D.5)}$$

Accordingly, using the same methods as in Appendix A, one can show that

$$\mathrm{Vol}\left(\mathcal{V}(\boldsymbol{x},\boldsymbol{y})\cap\mathcal{T}_{\varepsilon}^{n}(\,P_X)\right)\doteq\exp\left\{\frac{n}{2}\log\left[2\pi eP_X\left(1-\frac{\left|\sum_{i=1}^{n}x_iy_i\right|^2}{P_X\cdot\sum_{i=1}^{n}y_i^2}\right)\right]\right\},\tag{D.6}$$

and

$$\mathrm{Vol}(\mathcal{T}_{\varepsilon}^{n}(P_X))\doteq\exp\left\{\frac{n}{2}\log\left(2\pi eP_X\right)\right\}.\tag{D.7}$$

Thus, upon substitution in (D.4) and (D.2), and using the LLN, we get that the probability of error converges to zero as $n\to\infty$ as long as,

$$R<-\frac{1}{2}\log\left[1-\frac{h_0^2P_X}{\|\boldsymbol{h}\|_2^2\,P_X+\sigma^2}\right]\tag{D.8}$$

$$=\frac{1}{2}\log\left[1+\frac{h_0^2P_X}{(\|\boldsymbol{h}\|_2^2-h_0^2)\cdot P_X+\sigma^2}\right].\tag{D.9}$$

REFERENCES

[1] N. Merhav, G. Kaplan, A. Lapidoth, and S. Shamai (Shitz), “On information rates for mismatched decoders,” *IEEE Trans. on Inf. Theory*, vol. 40, no. 5, pp. 1953–1967, Nov. 1994.

[2] A. Lapidoth and P. Narayan, “Reliable communication under channel uncertainty,” *IEEE Trans. on Inf. Theory*, vol. 44, no. 5, pp. 2148–2177, Oct. 1998.

[3] A. Ganti, A. Lapidoth, and E. Telatar, “Mismatched decoding revisited: General alphabets, channels with memory, and the wide-band limit,” *IEEE Trans. Inf. Theory*, vol. 46, no. 7, p. 23152328, Nov. 2000.

[4] A. Somekh-Baruch, “A general formula for the mismatch capacity,” *IEEE Trans. on Inf. Theory*, vol. 61, no. 9, pp. 4554–4568, June 2015.

[5] I. G. Stiglitz, “Coding for a class of unknown channels,” *IEEE Trans. on Inf. Theory*, vol. IT-12, pp. 189–195, Apr. 1966.

[6] A. Kaplan and S. Shamai (Shitz), “Information rates of compound channels with application to antipodal signaling in a fading environment,” *AEU*, vol. 47, no. 4, pp. 228–239, Nov. 1993.

[7] I. C. Abou-Faycal, “An information theoretic study of reduced-complexity receivers for intersymbol interference channels,” Ph.D. dissertation, MIT, 2001.

[8] I. Csiszár and J. Körner, “Graph decomposition: A new key to coding theorems,” *IEEE Trans. on Inf. Theory*, vol. IT-27, pp. 5–12, Jan. 1981.

[9] J. Y. N. Hui, *Fundamental issues of multiple accessing*. Ph.D. dissertation, Mass. Inst. Technol., Cambridge, MA, 1983.

[10] J. Scarlett, A. Martinez, and A. Guillén i Fàbregas, “Mismatched decoding: Error exponents, second-order rates and saddlepoint approximations,” *IEEE Trans. on Inf. Theory*, vol. 60, no. 5, pp. 2647–2666, May 2014.

[11] R. G. Gallager, *Information Theory and Reliable Communication*. New York: Wiley, 1968.

[12] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. Academic Press, 1981.

[13] E. Abbe, M. Medard, S. Meyn, and L. Zheng, “Finding the best mismatched detector for channel coding and hypothesis testing,” in *2007 Information Theory and Applications Workshop*, Jan 2007, pp. 284–288.

[14] V. Polyanskiy, V. Poor, and S. Verdú, “Channel coding rate in the finite blocklength regime,” *IEEE Trans. on Inf. Theory*, vol. 56, no. 5, pp. 2307–2359, May 2010.

[15] M. Hayashi, “Information spectrum approach to second-order coding rate in channel coding,” *IEEE Trans. on Inf. Theory*, vol. 55, no. 11, pp. 4947–4966, Nov. 2009.

[16] J. L. Jensen, *Saddlepoint Approximations*. Oxford, U.K.: Oxford Univ. Press, 1995.

[17] I. Csiszár and P. Narayan, “Channel capacity for a given decoding metric,” *IEEE Trans. on Inf. Theory*, vol. 41, no. 1, pp. 35–43, Jan. 1995.

[18] N. G. De Bruijn, *Asymptotic Methods in Analysis*. Dover Publications, Inc. New York, 1981.

[19] N. Merhav, “Universal decoding for memoryless gaussian channels with a deterministic interference,” *IEEE Trans. on Inf. Theory*, vol. 39, no. 4, pp. 1261–1269, July 1993.

[20] W. Huleihel and N. Merhav, “Universal decoding for gaussian intersymbol interference channels,” *IEEE Trans. Inform. Theory*, vol. 61, no. 4, pp. 1606–161, Apr. 2015.

[21] N. Merhav, “Relations between random coding exponents and the statistical physics of random codes,” *IEEE Trans. on Inf. Theory*, vol. 55, no. 1, pp. 83–92, Jan. 2009.

[22] ——, “Erasure/list exponents for Slepian-Wolf decoding,” *IEEE Trans. on Inf. Theory*, vol. 60, no. 8, pp. 4463–4471, Aug. 2014.

[23] ——, “Exact random coding exponents of optimal bin index decoding,” *IEEE Trans. on Inf. Theory*, vol. 60, no. 10, pp. 6024–6031, Oct. 2014.

[24] W. Huleihel and N. Merhav, “Exact random coding error exponents for the two-user interference channel,” *submitted to IEEE Trans. Inform. Theory*, Mar. 2015.

[25] I. Csiszár, “Linear codes for sources and source networks: error exponents, universal coding,” *IEEE Trans. on Inf. Theory*, vol. IT-28, no. 4, pp. 585–592, July 1982.

[26] J. Ziv, “Universal decoding for finite-state channels,” *IEEE Trans. on Inf. Theory*, vol. IT-31, no. 4, pp. 453–460, July 1985.

[27] A. Lapidoth and J. Ziv, “On the universality of the lzbased noisy channels decoding algorithm,” *IEEE Trans. on Inf. Theory*, vol. 44, no. 5, pp. 1746–1755, Sep. 1998.

[28] M. Feder and A. Lapidoth, “Universal decoding for channels with memory,” *IEEE Trans. on Inf. Theory*, vol. 44, no. 5, pp. 1726–1745, Sep. 1998.

[29] M. Feder and N. Merhav, “Universal composite hypothesis testing: a competitive minimax approach,” *IEEE Trans. on Inf. Theory special issue in memory of Aaron D. Wyner*, vol. 48, no. 6, pp. 1504–1517, June 2002.

[30] Y. Lomnitz and M. Feder, “Communication over individual channels a general framework,” Mar. 2012. [Online]. Available: arXiv:1023.1406v1

[31] ——, “Universal communication over moduloadditive channels with an individual noise sequence,” May. 2012. [Online]. Available: arXiv:1012.2751v2

[32] V. Misra and T. Weissman, “The porosity of additive noise sequences,” May. 2012. [Online]. Available: arXiv:1025.6974v1

[33] O. Shayevitz and M. Feder, “Communicating using feedback over a binary channel with arbitrary noise sequence,” in *Proc. ISIT 2005*, Sep. 2005, pp. 1516–1520.

[34] ——, “Universal decoding for frequency-selective fading channels,” *IEEE Trans. on Inf. Theory*, vol. 51, no. 8, pp. 2770–2790, Aug. 2005.

[35] ——, “Universal decoding for frequency-selective fading channels,” *IEEE Trans. on Inf. Theory*, vol. 51, no. 8, pp. 2770–2790, Aug. 2005.

[36] N. Merhav, "Universal decoding for arbitrary channels relative to a given class of decoding metrics," *IEEE Trans. on Inf. Theory*, vol. 59, no. 9, pp. 5566–576, Sep. 2013.

[37] W. Hirt and J. L. Massey, "Capacity of the discrete-time gaussian channel with intersymbol interference," *IEEE Transactions on Information Theory*, vol. 34, no. 3, pp. 38–38, 1988.

[38] N. Merhav, "Universal decoding for arbitrary channels relative to a given class of decoding metrics," *IEEE Transactions on Information Theory*, vol. 59, no. 9, pp. 5566–5576, Sept 2013.

[39] N. Shulman, *Communication over an Unknown Channel via Common Broadcasting*. Ph.D. dissertation, Department of Electrical Engineering Systems, Tel Aviv University, July 2003.

[40] G. B. Folland, *Real Analysis*. John Wiley & Sons, 1984.

[41] U. Grenander and G. Szego, *Toeplitz Forms and Their Applications*. University of Calif. Press, Berkeley and Los Angeles, 1958.

[42] S. Ihara, *Information Theory for Continuous Systems*. Series on Probability & Statistics, World Scientific Pub Co Inc, 1993.

[43] R. S. Cheng and S. Verdú, "Gaussian multiaccess channels with isi: Capacity region and multiuser water-filling," *IEEE Transactions on Information Theory*, vol. 39, no. 3, pp. 773–785, 1993.

[44] J. Scarlett, A. Martinez, and A. Guillén i Fàbregas, "Multiuser coding techniques for mismatched decoding," *IEEE Trans. on Inf. Theory*, vol. 62, no. 7, pp. 3950–3970, July 2016.

[45] W. Huleihel and N. Merhav, "Random coding error exponents for the two-user interference channel," *IEEE Trans. on Inf. Theory*, vol. 63, no. 2, pp. 1019–1042, Feb. 2017.

[46] N. Merhav, "Statistical physics and information theory," *Foundations and Trends in Communications and Information Theory*, vol. 6, no. 1-2, pp. 1–212, Dec. 2010.

[47] R. Remmert, *Theory of complex functions*. Springer Science & Business Media, 2012, vol. 122.