Universal Decoding for Gaussian Intersymbol Interference Channels

Wasim Huleihel and Neri Merhav

Department of Electrical Engineering

Technion - Israel Institute of Technology

Haifa 32000, ISRAEL

E-mail: {wh@tx, merhav@ee}.technion.ac.il

Abstract

A universal decoding procedure is proposed for the intersymbol interference (ISI) Gaussian channels. The universality of the proposed decoder is in the sense of being independent of the various channel parameters, and at the same time, attaining the same random coding error exponent as the optimal maximum-likelihood (ML) decoder, which utilizes full knowledge of these unknown parameters. The proposed decoding rule can be regarded as a frequency domain version of the universal maximum mutual information (MMI) decoder. Contrary to previously suggested universal decoders for ISI channels, our proposed decoding metric can easily be evaluated.

Index Terms

Universal decoding, interference intersymbol (ISI), error exponents, maximum-likelihood (ML), random coding, maximum mutual information, Gaussian channels, deterministic interference.

I. INTRODUCTION

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 maximum likelihood (ML) decoder is precluded, and thus, universal decoders, independent of the

unknown channel, are sought. In designing such a decoder, there are two desirable properties that should be taken into account: The first is that the universal decoder performs asymptotically as well as the ML decoder had the channel law been known, and secondly, that the constructed decoding metric will be reasonably easy to calculate. This paper addresses the problem of universal decoding for intersymbol interference (ISI) Gaussian channels.

The topic of universal coding and decoding under channel uncertainty has received very much attention in the last four decades, see, for example, [1-15]. In the realm of memoryless channels, Goppa [2] explored the maximum mutual information (MMI) decoder, which chooses the codeword having the maximum empirical mutual information (MMI) with the channel output sequence. It was shown that this decoder achieves the capacity in the case of discrete memoryless channels (DMC). In [3], the problem of universal decoding for DMC's with finite input and output alphabets was studied. It was shown that the MMI decoder universally achieves the optimal random coding error exponent under the uniform random coding distribution over a certain type class. In [1], an analogous result was derived for a certain parametric class of memoryless Gaussian channels with an unknown deterministic interference signal. In the same paper, a conjecture was proposed concerning a universal decoder for ISI channels.

For channels with memory, there are several quite general results, each proposing a different universal decoder. In [5], the case of unknown finite-state channels with finite input and output alphabets for which the next channel state is a deterministic unknown function of the channel current state and current inputs and outputs, was considered. For uniform random codes over a given set, a universal decoder (that achieves the optimal random coding error exponent) which is based on the Lempel-Ziv algorithm was proposed. Later, in [6], it was shown that this decoder continues to be universally asymptotically optimum also for the class of finite-state channels with stochastic, rather than deterministic, next-state functions. In [7], sufficient conditions and a universal decoder (called the merging decoder) were proposed, for families of channels with memory. The idea was to employ many decoding lists in parallel, each one corresponding to one point in a dense grid (whose size grows with the input block length) in the index set. Accordingly, with regard to our work, it was shown that the proposed decoder universally achieves the optimal error exponent under the ISI channel. Unfortunately, as was mentioned before, this deocder is very hard to implement in practice due to its implicit structure and the fact that it requires to form a dense grid in the parameter space. In [8], a competitive minimax criterion was proposed. According to this approach, an optimum decoder is sought in the quest for minimizing (over all decision rules) the maximum (over all channels in the family) ratio between the error probability associated with a given channel and a given decision rule, and the error probability of the ML decoder for that channel, possibly

raised some power less than unity. This decoder is, again, very hard to implement for the ISI channel due its complicated decoding metric.

In this paper, we propose a universal decoder that asymptotically achieves the optimal error exponent, and contrary to previous proposed decoders, our proposed decoding metric can easily be calculated. The technique used in this paper is in line with the techniques which were established in [1, 16]. Specifically, similarly to [1], the main idea is to define an auxiliary "backward channel", which is a mathematical tool for assessing log-volumes of conditional typical sets of sequences with continuous-valued components. These log-volume terms play a pivotal role in the universal decoding metric. The backward channel is defined in a way that guarantees two properties: first, a measure concentration property, that is, assignment of high probability to a given conditional type by an appropriate choice of certain parameters, and secondly, the conditional density of the input given the output, associated with this backward channel should depend on the input and the output only via the sufficient statistics that define the conditional type class. Contrary to the problem considered in [1], the difficulty, in the ISI channel, stems from the fact that the choice of the backward channel is a non-trivial issue. It turns out that in this case, the passage to the frequency domain resolves this difficulty. The proposed decoding rule can be regarded as a frequency domain version of the universal maximum mutual information (MMI) decoder.

The remaining part of this paper is organized as follows. In Section II, we first present the model and formulate the problem. Then, the main results are provided and discussed. In Section III, we provide a proof outline where we discuss the techniques and methodologies that are utilized in order to prove the main result. Finally, in Section IV, the main results are proved.

II. MODEL FORMULATION AND MAIN RESULT

Consider a discrete time, Gaussian channel characterized by

$$y_t = \sum_{i=0}^{k} h_i x_{t-i} + w_t, \quad t = 0, 1, 2, \ldots, n \tag{1}$$

where $\{x_t\}$ are the channel inputs, $\{h_i\}_{i=0}^{k}$ is the unknown channel impulse response, $\{w_t\}$ is zero-mean Gaussian white noise with an unknown variance $\sigma^2 > 0$, and $\{y_t\}$ are the channel outputs. It will be assumed that the noise $\{w_t\}$ is statistically independent of the input $\{x_t\}$. We allow k to grow with n in the order of $k = o\left(n^{1/2}\right)$. In such a case, we further assume that the impulse response sequence $\{h_i\}_{i=0}^{\infty}$ is absolutely summable[1].

---

[1]This assumption can be relaxed to square summability of $\{h_i\}$.

The input is a codeword that is randomly and uniformly drawn over a codebook $\mathcal{C} = \left\{ \boldsymbol{x}^1, \ldots, \boldsymbol{x}^M \right\}$ of $M = 2^{nR}$ messages $\boldsymbol{x}^i = \left(x_1^i, \ldots, x_n^i \right) \in \mathbb{R}^n$, $i = 1, 2, \ldots, M$, where R is the coding rate in bits per channel use. In the following, the probability of error associated with the ML decoder, that knows the unknown parameters $\left(\sigma^2, h_0, \ldots, h_k \right)$, will be denoted by $P_{e,o}\left(\mathcal{C}, R, n \right)$. We shall adopt the random coding approach, where each codeword is randomly chosen with respect to a probability measure denoted by $\mu\left(\boldsymbol{x} \right)$. For a given power constraint, a reasonable choice of $\mu\left(\cdot \right)$ is the truncated Gaussian density restricted to the shell of an n-dimensional hypersphere whose radius is about $n\sigma_x^2$. To wit,

$$\mu\left(\boldsymbol{x} \right) = \nu^{-1}\psi_\Delta\left(\boldsymbol{x} \right) \prod_{t=0}^{n-1} \exp\left\{ -\frac{x_t^2}{2\sigma_x^2} \right\} \tag{2}$$

where $\psi_\Delta\left(\boldsymbol{x} \right)$ is the indicator function of the set

$$D_\Delta \overset{\triangle}{=} \left\{ \boldsymbol{x} : \left| \frac{1}{n}\sum_{t=0}^{n-1} x_t^2 - \sigma_x^2 \right| \leq \Delta\sigma_x^2 \right\} \tag{3}$$

where $\Delta \ll 1$, and ν normalizes the above measure such that it would integrate to unity. Note that $\mu\left(\boldsymbol{x} \right)$ is invariant to unitary transformations of $\boldsymbol{x}$. It is well-known [17, Chap. 7] that $\mu\left(\cdot \right)$ attains a higher error exponent than that of the respective Gaussian density with the same variance, at least for small rates, where the non-typical events (or, the large deviations events) are the dominant[2]. The analysis in this paper can also be carried for the case where the codewords are drawn independently and uniformly over a set $\mathcal{I}_n \subseteq \mathbb{R}^n$ that is endowed with a σ-algebra (e.g., an n-dimensional hypercube), and satisfy an average power constraint, as was considered in [7, Theorem 4]. Let $\bar{P}_{e,o}\left(R, n \right) \overset{\triangle}{=} \mathbb{E}\left\{ P_{e,0}\left(\mathcal{C}, R, n \right) \right\}$, where the expectation is taken over the ensemble of randomly selected codebooks under $\mu\left(\cdot \right)$. Finally, we define the random coding error exponent as $E\left(R \right) \overset{\triangle}{=} -\limsup_{n\to\infty} n^{-1}\log\bar{P}_{e,o}\left(R, n \right)$.

As was mentioned previously, we wish to find a decoding procedure which is universal in the sense of being independent of the unknown parameters, and at the same time attaining $E\left(R \right)$. Specifically, let $P_{e,u}\left(\mathcal{C}, R, n \right)$ designate the error probability associated with the universal rule for a given codebook $\mathcal{C}$, and let $\bar{P}_{e,u}\left(R, n \right) \overset{\triangle}{=} \mathbb{E}\left\{ P_{e,u}\left(\mathcal{C}, R, n \right) \right\}$. Then, we would like $\bar{P}_{e,u}\left(R, n \right)$ to decay exponentially with rate $E\left(R \right)$.

We now turn to present the proposed decoding rule. To this end, let $\tilde{\boldsymbol{x}}$ and $\tilde{\boldsymbol{y}}$ denote the discrete Fourier transforms (DFT) of the sequences $\{x_t\}$ and $\{y_t\}$, respectively, i.e., the m-th component of $\tilde{\boldsymbol{x}}$ is given by

$$\tilde{x}_m = \frac{1}{\sqrt{n}}\sum_{t=0}^{n-1} x_t e^{-j2\pi mt/n} \tag{4}$$

---

[2]Intuitively speaking, this is true because of the fact that it does not allow low energy codewords

where $j = \sqrt{-1}$ and similarly for $\tilde{\boldsymbol{y}}$. Then, define an auxiliary "backward channel" by the conditional measure

$$V\left(\tilde{\boldsymbol{x}}|\tilde{\boldsymbol{y}}, \boldsymbol{\theta}, k\right) = \prod_{m=0}^{n-1} \left(2\pi\sigma_0^2\right)^{-1/2} \exp\left\{-\frac{1}{2\sigma_0^2}\left|\tilde{x}_m - \tilde{y}_m \sum_{l=0}^{k} \alpha_l e^{\frac{2\pi jlm}{n}}\right|^2\right\} \tag{5}$$

where $\boldsymbol{\theta} \triangleq \left(\sigma_0^2, \alpha_0, \ldots, \alpha_k\right)$ is the parameters vector of the backward channel, in which $\{\alpha_l\}_{l=0}^{k}$ are complex-valued. It should be emphasized that the above definition of the auxiliary backward channel is completely unrelated to the underlying probabilistic model. In particular, it is not argued that $V\left(\tilde{\boldsymbol{x}}|\tilde{\boldsymbol{y}}, \boldsymbol{\theta}, k\right)$ is obtained from $\mu\left(\boldsymbol{x}\right)$ and the forward channel (1) by the Bayes rule, or any other relationship. For example, our backward channel allows vectors $\boldsymbol{x}$ that are outside the region D_Δ. Our decoding rule will select a message $\tilde{\boldsymbol{x}}^i$ that maximizes the metric

$$u\left(\tilde{\boldsymbol{x}}^i, \tilde{\boldsymbol{y}}\right) = \frac{\max_{\boldsymbol{\theta}} V\left(\tilde{\boldsymbol{x}}^i|\tilde{\boldsymbol{y}}, \boldsymbol{\theta}, k\right)}{\mu\left(\tilde{\boldsymbol{x}}^i\right)} \tag{6}$$

among all M codewords. The backward channel is a mathematical tool for assessing log-volumes of typical sets [1, 16, 18], and it should be defined in a way that guarantees two general properties: first, a measure concentration property, that is, assignment of high probability to a given conditional type by an appropriate choice of the parameters of this backward channel, and secondly, the conditional density of $\tilde{\boldsymbol{x}}$ given $\tilde{\boldsymbol{y}}$, associated with the backward channel should depend on $\tilde{\boldsymbol{x}}$ and $\tilde{\boldsymbol{y}}$ only via the sufficient statistics that define the conditional type class. Contrary to the problem considered in [1], the difficulty in the ISI channel stems from the fact that the choice of the backward channel is a non-trivial issue. Specifically, as will be seen in the sequel, an "appropriate" candidate backward channel must depend on a sufficient statistics vector (associated with $\boldsymbol{x}$) with dimension that equals to the number of degrees of freedom, which in turn adjust their conditional expectations. It turns out that in this case, the passage to the frequency domain is more "natural" and mathematically convenient due to the well-known asymptotic spectral properties of Toeplitz matrices (see, for example, [19]). To wit, it can be seen that the model in (1) can be written in a vector form $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{w}$ where $\boldsymbol{A} = \{a_{i,j}\} = \{h_{i-j}\}$ is a Toeplitz matrix. Now, by the spectral decomposition theorem [20], we know that there exists an orthonormal basis that diagonalizes the matrix $\boldsymbol{A}$. Projecting the observations onto this basis will simply decompose the original channel into a set of independent channels, which are simpler to analyze. While this is true for any matrix $\boldsymbol{A}$, for Toeplitz matrices we can asymptotically characterize their eigenvalues and eigenvectors in terms of the generating sequence $\{h_i\}_{i=0}^{k}$, which is a fundamental part in our analysis. We next give the main result of this paper.

Theorem 1 Let the codewords of $\mathcal{C}$ be chosen randomly and independently with respect to the density $\mu\left(\cdot\right)$ given in (2). Assume that the channel impulse response coefficients are absolutely summable $\{h_l\}_{l=1}^\infty \in \ell_1$, and that $k = o\left(n^{1/2}\right)$. Then,

$$\limsup_{n\to\infty} \frac{1}{n} \left[\log \bar{P}_{e,u}\left(R,n\right) - \bar{P}_{e,0}\left(R,n\right)\right] \leq \xi\left(\Delta\right) \tag{7}$$

where $\xi\left(\Delta\right) \to 0$ as $\Delta \to 0$, and $\bar{P}_{e,u}\left(R,n\right)$ is the average probability of error associated with the universal decoder given in (6).

The intuitive interpretation of (6) is that $n^{-1}\log u\left(\tilde{\boldsymbol{x}},\tilde{\boldsymbol{y}}\right) = n^{-1}\log\max_{\boldsymbol{\theta}} V\left(\tilde{\boldsymbol{x}}|\tilde{\boldsymbol{y}},\boldsymbol{\theta},k\right)/\mu\left(\tilde{\boldsymbol{x}}\right)$ is an empirical version of the per-letter mutual information between $\boldsymbol{x}$ and $\boldsymbol{y}$ in the frequency domain. Thus, we select the input $\tilde{\boldsymbol{x}}$ that seems empirically "most dependent" upon the given output vector $\tilde{\boldsymbol{y}}$ in the frequency domain, which corresponds to the MMI principle. The passage to the frequency domain asymptotically eliminates the strong interactions between the various components of the input vector, and transforms the original model into a set of n separable channels which are controlled by $\left(k+2\right)$ degrees of freedom. Note that on the support of $\mu\left(\cdot\right)$, the term $n^{-1}\log\mu\left(\tilde{\boldsymbol{x}}^i\right)$ is nearly a constant independent of i. Thus, the proposed decoding rule is essentially equivalent to one that maximizes $\max_{\boldsymbol{\theta}} V\left(\tilde{\boldsymbol{x}}|\tilde{\boldsymbol{y}},\boldsymbol{\theta},k\right)$, namely, maximum a posteriori (MAP) decoding.

Remark 1 In [1], a universal decoding procedure for memoryless Gaussian channels with a deterministic interference was proposed. Accordingly, we remark that Theorem 1 can be fairly easily extended to the channel model

$$y_t = \sum_{i=0}^k h_i x_{t-i} + z_t + w_t \tag{8}$$

where $\{z_t\}$ is an unknown deterministic interference that can be decomposed as a series expansion of orthonormal bounded functions with an absolutely summable coefficient sequence, namely,

$$z_t = \sum_{i=1}^\infty b_i \phi_{i,t}, \quad t = 1, 2, \ldots \tag{9}$$

where $\{b_i\} \in \ell_1$ and $|\phi_{i,t}| \leq L < \infty$ for all i and t. The coefficients $\{b_i\}$ are assumed deterministic and unknown. In this case, an appropriate definition of the auxiliary backward channel is

$$\tilde{V}\left(\tilde{\boldsymbol{x}}|\tilde{\boldsymbol{y}},\boldsymbol{\theta},k,q\right) = \prod_{m=1}^n \left(2\pi\sigma_0^2\right)^{-1/2} \exp\left\{-\frac{1}{2\sigma_0^2}\left|\tilde{x}_m - \tilde{y}_m \sum_{l=0}^k \alpha_l e^{\frac{2\pi jlm}{n}} - \sum_{i=1}^q \beta_i \tilde{\phi}_{i,m}\right|^2\right\} \tag{10}$$

where now $\boldsymbol{\theta} \stackrel{\triangle}{=} \left(\sigma_0^2, \alpha_0, \ldots, \alpha_k, \beta_1, \ldots, \beta_q\right)$ is the parameter vector of the backward channel, $\left\{\tilde{\phi}_{i,m}\right\}$ is the frequency transformed representation of $\{\phi_{i,t}\}$, and $q = q_n$ is assumed to be a monotonically

non-decreasing integer-valued sequence such that $q_n \to \infty$ and $q_n = o\left(n^{1/3}\right)$. Accordingly, the decoding rule will select a message $\boldsymbol{x}^i$ that maximizes the metric (6) (where V in (6) is replaced with $\tilde{V}$), among all M codewords. For simplicity of the exposition and to facilitate the reading of the proof of Theorem 1, we will assume the original model (1).

III. PROOF OUTLINE

In this section, before getting deep into the proof of Theorem 1, we discuss the techniques and the main steps which will be used in Section IV. In order to facilitate the explanations, we will need the following definitions: Let $\boldsymbol{x}$ and $\boldsymbol{y}$ be arbitrary vectors in $\mathbb{R}^n$ and define

$$\mathcal{S}_o\left(\boldsymbol{x}, \boldsymbol{y}\right) \triangleq \left\{\boldsymbol{x}' : W\left(\boldsymbol{y}|\boldsymbol{x}'\right) > W\left(\boldsymbol{y}|\boldsymbol{x}\right)\right\}, \tag{11}$$

$$\mathcal{S}_u\left(\boldsymbol{x}, \boldsymbol{y}\right) \triangleq \left\{\boldsymbol{x}' : u\left(\boldsymbol{x}', \boldsymbol{y}\right) > u\left(\boldsymbol{x}, \boldsymbol{y}\right)\right\}, \tag{12}$$

and

$$\mathcal{S}_0^\delta\left(\boldsymbol{x}, \boldsymbol{y}\right) \triangleq \left\{\boldsymbol{x}' : \frac{1}{n}\log W\left(\boldsymbol{y}|\boldsymbol{x}'\right) > \frac{1}{n}\log W\left(\boldsymbol{y}|\boldsymbol{x}\right) - \delta\right\}, \tag{13}$$

where $W\left(\boldsymbol{y}|\boldsymbol{x}\right)$ is the conditional pdf associated with the channel. In words, $\mathcal{S}_o\left(\boldsymbol{x}, \boldsymbol{y}\right)$ and $\mathcal{S}_u\left(\boldsymbol{x}, \boldsymbol{y}\right)$ are simply the sets of prospective incorrect codewords corresponding to the ML decoder, and the proposed universal decoder, respectively, assuming that $\boldsymbol{x}$ is the transmitted codewords and that $\boldsymbol{y}$ is the received vector. The set $\mathcal{S}_0^\delta\left(\boldsymbol{x}, \boldsymbol{y}\right)$ is just a δ-perturbed version of $\mathcal{S}_0\left(\boldsymbol{x}, \boldsymbol{y}\right)$ which will be used for technical reasons. Finally, we let $\bar{P}_{e,o}\left(R, n\right)$, $\bar{P}_{e,u}\left(R, n\right)$, and $\bar{P}_{e,o}^\delta\left(R, n\right)$ be the average error probabilities associated with the ML decoder, the proposed decoder, and the δ-perturbed decoder (see, (18)-(21)).

Generally speaking, the root of our analysis is Lemma 1, which was asserted and proved in [1, Lemma 1], and can be thought as a continuous extension of [5, Corollary 1]. This result relates between $\bar{P}_{e,o}^\delta\left(R, n\right)$ and $\bar{P}_{e,u}\left(R, n\right)$ as follows

$$\bar{P}_{e,u}\left(R, n\right) \leq 2\bar{P}_{e,o}^\delta\left(R, n\right) \left[\frac{3}{2} + \sup_{(\boldsymbol{x}, \boldsymbol{y}) \in H_n} \frac{\int_{\mathcal{S}_u(\boldsymbol{x}, \boldsymbol{y})} \mu\left(\boldsymbol{x}'\right) \mathrm{d}\boldsymbol{x}'}{\int_{\mathcal{S}_o^\delta(\boldsymbol{x}, \boldsymbol{y})} \mu\left(\boldsymbol{x}'\right) \mathrm{d}\boldsymbol{x}'}\right]. \tag{14}$$

where $\{H_n\}_{n \geq 1}$ is a sequence of sets of pairs $(\boldsymbol{x}, \boldsymbol{y})$ such that

$$\limsup_{n \to \infty} \frac{1}{n} \log \mathbb{P}\left\{H_n^c\right\} < -E\left(R\right). \tag{15}$$

Whence, we see that in order to show that $\bar{P}_{e,u}\left(R, n\right)$ and $\bar{P}_{e,o}\left(R, n\right)$ are exponentially the same, we just need to define a sequence $\{H_n\}_{n \geq 1}$ such that the ratio in (14)

$$\frac{\int_{\mathcal{S}_u(\boldsymbol{x}, \boldsymbol{y})} \mu\left(\boldsymbol{x}'\right) \mathrm{d}\boldsymbol{x}'}{\int_{\mathcal{S}_o^\delta(\boldsymbol{x}, \boldsymbol{y})} \mu\left(\boldsymbol{x}'\right) \mathrm{d}\boldsymbol{x}'} \tag{16}$$

is uniformly overbounded by a subexponential function of n, i.e., $e^{n\epsilon_n}$ where $\epsilon_n \to 0$ as $n \to \infty$ uniformly for all $(\boldsymbol{x}, \boldsymbol{y}) \in H_n$. Once this accomplished, the proof of the theorem will be complete. The main question is now how to define the sequence $\{H_n\}_{n \geq 1}$ properly? To answer this question, let us interpret its role. The set H_n simply divides the space of pairs $(\boldsymbol{x}, \boldsymbol{y})$ into two parts, where in the first part, the supremum in (14) is uniformly bounded by a subexponential function of n, and the second part possesses a probability smaller than the desired exponential function $e^{-nE(R)}$ and hence negligible (see, (15)). Obviously, given these requirements one can propose several candidates for H_n, namely, the choice is not unique. However, another important property that $\{H_n\}_{n \geq 1}$ should account for is that the function $n^{-1} \log V(\tilde{\boldsymbol{x}}|\tilde{\boldsymbol{y}}, \boldsymbol{\theta}, k)$ will be uniformly continuous w.r.t. small perturbations of the sufficient statistics (this idea will be emphasized in the analysis). To summarize, the first part in the forthcoming analysis is to define the sequence $\{H_n\}_{n \geq 1}$ such that (15) holds true, and that hopefully (14) will hold too. The proposed $\{H_n\}_{n \geq 1}$ is given in Lemma 2, and the main tool that is used in the proof is large deviations theory.

Following the first part, in the second part, we will eventually show that the chosen H_n fulfills the desired subexponential behavior of (16). Accordingly, we will overbound (16) within H_n as follows: we will derive an upper bound on the numerator of (16) and a lower bound on its denominator, and show that these are exponentially equivalent. To this end, we will need to define a conditional typical set of our continuous-valued input-output sequences, establish some of its properties, and particularly to calculate its volume (Lebesgue measure). This typical set of some sequence $\tilde{\boldsymbol{x}}$ given $\tilde{\boldsymbol{y}}$ will contain all the vectors which, within $\epsilon > 0$, have the same sufficient statistics as $\tilde{\boldsymbol{x}}$ induced by our backward channel (see (61) for a precise definition of this set). Then, we will provide upper and lower bounds (which are exponentially of the same order) on the volume of this typical set. To accomplish this, we will use methods that were previously used in [1, 16, 18], which are based on large deviations theory and methods that are customary to statistical physics. After that, we will show that for any two vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ that belong to this typical set, the conditional pdf's $W(\boldsymbol{y}|\boldsymbol{u})$ and $W(\boldsymbol{y}|\boldsymbol{v})$ are exponentially equivalent, that is, for sufficiently large n,

$$\left| \frac{1}{n} \log W(\boldsymbol{y}|\boldsymbol{u}) - \frac{1}{n} \log W(\boldsymbol{y}|\boldsymbol{v}) \right| < \zeta \tag{17}$$

for any $\zeta > 0$. Thus, given this property, we can easily provide a lower bound on the denominator of (16). Indeed, since $\boldsymbol{x} \in \mathcal{S}_o^\delta(\boldsymbol{x}, \boldsymbol{y})$, then in view of the last result, there exists a sufficiently small $\epsilon > 0$ such that the predefined typical set is essentially a subset of $\mathcal{S}_o^\delta(\boldsymbol{x}, \boldsymbol{y})$. Therefore, the integral over $\mathcal{S}_o^\delta(\boldsymbol{x}, \boldsymbol{y})$, in the denominator, can be underestimated as an integral over the typical set, and since we know its

volume (or, more precisely, a lower bound on it which is exponentially tight), it is not difficult to provide a lower bound on this integral (see (105) for more details). Providing an upper bound on the numerator is a little more involved. The underlying idea is to partition the set $\mathcal{S}_u\left(\boldsymbol{x}, \boldsymbol{y}\right)$ into a subexponential number of conditional types, where for each conditional type, the integral over the respective conditional type is overestimated using the upper bound on the volume. Finally, it will be shown that these two bounds are exponentially equivalent, which implies that (16) is subexponential function of n, as required.

IV. Proof of Theorem 1

For completeness, in this section, we will provide again some definitions that were already presented in short in the previous section. Let $\boldsymbol{x}$ and $\boldsymbol{y}$ be arbitrary vectors in $\mathbb{R}^n$ and define $\mathcal{S}_o\left(\boldsymbol{x}, \boldsymbol{y}\right)$ and $\mathcal{S}_u\left(\boldsymbol{x}, \boldsymbol{y}\right)$ as in eqs. (11) and (12), respectively. The average error probabilities associated with the ML decoder and the proposed decoder are given by (see, for example, [1])

$$\bar{P}_{e,o}\left(R, n\right) = 1 - \mathbb{E}\left\{\left[1 - \int_{\mathcal{S}_o(\boldsymbol{X}, \boldsymbol{Y})} \mu\left(\boldsymbol{x}'\right) \mathrm{d}\boldsymbol{x}'\right]^{2^{nR}-1}\right\} \tag{18}$$

and

$$\bar{P}_{e,u}\left(R, n\right) = 1 - \mathbb{E}\left\{\left[1 - \int_{\mathcal{S}_u(\boldsymbol{X}, \boldsymbol{Y})} \mu\left(\boldsymbol{x}'\right) \mathrm{d}\boldsymbol{x}'\right]^{2^{nR}-1}\right\}, \tag{19}$$

respectively, where the expectations are taken with respect to (w.r.t.) the joint distribution $\mu\left(\boldsymbol{x}\right) W\left(\boldsymbol{y}|\boldsymbol{x}\right)$, and we use the usual conventions where random vectors are denoted by capital letters in bold face font, and their sample values are denoted by the respective lower case letters. Similar convention will apply to scalar random variables (RVs), which will be denoted with same symbols without the bold face font. Finally, for $\delta > 0$ we define the set

$$\mathcal{S}_0^\delta\left(\boldsymbol{x}, \boldsymbol{y}\right) \triangleq \left\{\boldsymbol{x}' : \frac{1}{n}\log W\left(\boldsymbol{y}|\boldsymbol{x}'\right) > \frac{1}{n}\log W\left(\boldsymbol{y}|\boldsymbol{x}\right) - \delta\right\}, \tag{20}$$

and accordingly

$$\bar{P}_{e,o}^\delta\left(R, n\right) = 1 - \mathbb{E}\left\{\left[1 - \int_{\mathcal{S}_0^\delta(\boldsymbol{X}, \boldsymbol{Y})} \mu\left(\boldsymbol{x}'\right) \mathrm{d}\boldsymbol{x}'\right]^{2^{nR}-1}\right\}. \tag{21}$$

Finally, with a slight abuse of notation, we also use the notation $\mathcal{S}_o\left(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}}\right)$ which is defined as follows: Let $\tilde{\boldsymbol{x}}$ and $\tilde{\boldsymbol{y}}$ be the Fourier transforms of $\boldsymbol{x}$ and $\boldsymbol{y}$, respectively. Then, $\mathcal{S}_o\left(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}}\right) \triangleq \left\{\tilde{\boldsymbol{x}}' = \boldsymbol{F}^H \boldsymbol{x}' : \boldsymbol{x}' \in \mathcal{S}_0\left(\boldsymbol{x}, \boldsymbol{F}^H \tilde{\boldsymbol{y}}\right)\right\}$ where $\boldsymbol{F}$ is the DFT matrix, namely, $\boldsymbol{F} = \left\{e^{j2\pi ml/n}/\sqrt{n}\right\}_{m,l=0}^{n-1}$.

As was discussed earlier, our goal is to compare the exponential behavior of $\bar{P}_{e,u}(R,n)$ to that of $\bar{P}_{e,o}(R,n)$. To this end, we will instead compare the exponential behavior of $\bar{P}_{e,u}(R,n)$ to that of $\bar{P}_{e,o}^{\delta}(R,n)$ for small $\delta > 0$. In the final step of the proof, this will be justified by showing that

$$\limsup_{n \to \infty} \frac{1}{n} \left[\log \bar{P}_{e,o}^{\delta}(R,n) - \log \bar{P}_{e,o}(R,n) \right] \leq \delta' \tag{22}$$

where $\delta' \to 0$ as $\delta \to 0$ and $\Delta \to 0$. In the analysis, we will use the following lemma [1, Lemma 1 pp. 1263].

Lemma 1 Let $\{H_n\}_{n \geq 1}$ be a sequence of sets of pairs $(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}})$ of n-dimensional vectors such that

$$\limsup_{n \to \infty} \frac{1}{n} \log \mathbb{P}\{H_n^c\} < -E(R) \tag{23}$$

Then, for all large n,

$$\bar{P}_{e,u}(R,n) \leq 2\bar{P}_{e,o}^{\delta}(R,n) \left[\frac{3}{2} + \sup_{(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}}) \in H_n} \frac{\int_{\mathcal{S}_u(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}})} \mu(\boldsymbol{x}') \, \mathrm{d}\boldsymbol{x}'}{\int_{\mathcal{S}_o^{\delta}(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}})} \mu(\boldsymbol{x}') \, \mathrm{d}\boldsymbol{x}'} \right]. \tag{24}$$

Thus, by using Lemma 1, we see that in order to show that $\bar{P}_{e,u}(R,n)$ and $\bar{P}_{e,o}(R,n)$ are exponentially the same, we just need to find a sequence $\{H_n\}_{n \geq 1}$ such that the ratio

$$\frac{\int_{\mathcal{S}_u(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}})} \mu(\boldsymbol{x}') \, \mathrm{d}\boldsymbol{x}'}{\int_{\mathcal{S}_o^{\delta}(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}})} \mu(\boldsymbol{x}') \, \mathrm{d}\boldsymbol{x}'} \tag{25}$$

is uniformly overbounded by a subexponential function of n, i.e., $e^{n\epsilon_n}$ where $\epsilon_n \to 0$ as $n \to \infty$ uniformly for all $(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{x}}) \in H_n$. For a given pair $(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}})$, let us define $\hat{\boldsymbol{\theta}} = (\hat{\sigma}_0^2, \hat{\alpha}_0, \ldots, \hat{\alpha}_k)$ to be

$$\hat{\boldsymbol{\theta}} \stackrel{\triangle}{=} \arg\max_{\theta} V(\tilde{\boldsymbol{x}} | \tilde{\boldsymbol{y}}, \boldsymbol{\theta}, k). \tag{26}$$

The set H_n will be parametrized by a parameter $B > 0$ and defined as follows

$$H_n(B) \stackrel{\triangle}{=} \left\{ (\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}}) : \left| \frac{1}{n} \sum_{m=0}^{n-1} |\tilde{x}_m|^2 - \sigma_x^2 \right| \leq \Delta \sigma_x^2, \ \frac{1}{n} \sum_{m=0}^{n-1} |\tilde{y}_m|^2 \leq B, \ \hat{\sigma}_0^2 \geq \frac{1}{B} \right\}. \tag{27}$$

We have the following result.

Lemma 2 There exists a sufficiently large B such that $\{H_n(B)\}_{n \geq 1}$ satisfies (23).

Proof of Lemma 2: By the union bound we have that

$$\mathbb{P}\{H_n^c(B)\} \leq \mathbb{P}\left\{ \frac{1}{n} \sum_{t=0}^{n-1} Y_t^2 > B \right\} + \mathbb{P}\{\hat{\sigma}_0^2 < B^{-1}\}. \tag{28}$$

Thus, it should be shown that if B is sufficiently large, both probabilities on the right-hand side of (28) decays faster than $e^{-nE(R)}$. Regarding the first term, note that

$$\frac{1}{n}\sum_{t=0}^{n-1} y_t^2 \leq \left[\sqrt{\frac{1}{n}\|\boldsymbol{H}\boldsymbol{x}\|^2} + \sqrt{\frac{1}{n}\sum_{t=0}^{n-1} w_t^2}\right]^2 \tag{29}$$

$$\leq \left[\sqrt{\frac{1}{n}\|\boldsymbol{x}\|^2}\sqrt{\|\boldsymbol{H}^T\boldsymbol{H}\|_s} + \sqrt{\frac{1}{n}\sum_{t=0}^{n-1} w_t^2}\right]^2 \tag{30}$$

$$\leq \left[\sqrt{\sigma_x^2(1+\Delta)}\|\boldsymbol{H}\|_s + \sqrt{\frac{1}{n}\sum_{t=0}^{n-1} w_t^2}\right]^2 \tag{31}$$

where $\|\cdot\|_s$ denotes the spectral norm, and in the second inequality we have used the fact that $|\mathrm{tr}\,\boldsymbol{AB}| \leq \|\boldsymbol{B}\|_s\,\mathrm{tr}\,(\boldsymbol{A})$ for any $\boldsymbol{B}$ and nonnegative definite matrix $\boldsymbol{A}$. Due to the fact that $\{h_m\} \in \ell_1$ (essentially, $\{h_m\} \in \ell_2$ is suffice here) it can be shown that [19] the spectral norm $\|\boldsymbol{H}\|_s$ is uniformly bounded, that is for all matrix dimension n we have that $\|\boldsymbol{H}\|_s \leq M$ where $M > 0$. Therefore, we obtain that

$$\mathbb{P}\left\{\frac{1}{n}\sum_{t=0}^{n-1} Y_t^2 > B\right\} \leq \mathbb{P}\left\{\frac{1}{n}\sum_{t=0}^{n-1} W_t^2 > \left(\sqrt{B} - M\sqrt{\sigma_x^2(1+\Delta)}\right)^2\right\} \tag{32}$$

which can be made less than $e^{-nE(R)}$ by selecting a sufficiently large B, as can be shown by a simple application of the Chernoff bound. As for the remaining terms: by taking the gradient of $V(\tilde{\boldsymbol{x}}|\tilde{\boldsymbol{y}},\boldsymbol{\theta},k)$ w.r.t. $\boldsymbol{\theta}$, we obtain that the components of $\hat{\boldsymbol{\theta}}$ are given by the solutions of the following set of equations

$$\sum_{m=0}^{n-1} \tilde{x}_m\tilde{y}_m^* e^{-\frac{2\pi jmq}{n}} = \sum_{m=0}^{n-1} |\tilde{y}_m|^2 e^{-\frac{2\pi jmq}{n}} \sum_{l=0}^{k} \hat{\alpha}_l e^{\frac{2\pi jml}{n}}, \quad \text{for } q = 0,\ldots,k, \tag{33}$$

and

$$\hat{\sigma}_0^2 = \frac{1}{n}\sum_{m=0}^{n-1}\left|\tilde{x}_m - \tilde{y}_m\sum_{l=1}^{k}\hat{\alpha}_l e^{\frac{2\pi jml}{n}}\right|^2. \tag{34}$$

Note that

$$\mathbb{P}\{\hat{\sigma}_0^2 < B^{-1}\} \leq \mathbb{P}\left\{\hat{\sigma}_0^2 < B^{-1}, \frac{1}{n}\sum_{t=0}^{n-1}\left|\tilde{Y}_m\right|^2 \leq \sqrt{B}, \min_{0\leq m\leq n-1}\left|\tilde{Y}_m\right|^2 \geq \tau\right\}$$

$$+ \mathbb{P}\left\{\frac{1}{n}\sum_{t=0}^{n-1}\left|\tilde{Y}_m\right|^2 > \sqrt{B}\right\} + \mathbb{P}\left\{\max_{0\leq m\leq n-1}\left|\tilde{Y}_m\right|^2 \leq \tau\right\}$$

$$\leq \mathbb{P}\left\{\hat{\sigma}_0^2 < B^{-1}, \frac{1}{n}\sum_{t=0}^{n-1}\left|\tilde{Y}_m\right|^2 \leq \sqrt{B}, \min_{0\leq m\leq n-1}\left|\tilde{Y}_m\right|^2 \geq \tau\right\}$$

$$+ \mathbb{P}\left\{\frac{1}{n}\sum_{t=0}^{n-1}\left|\tilde{Y}_m\right|^2 > \sqrt{B}\right\} + \mathbb{P}\left\{\frac{1}{n}\sum_{t=0}^{n-1}\left|\tilde{Y}_m\right|^2 \leq \tau\right\} \tag{35}$$

where $\tau > 0$. As before, the exponential decay rate of the last two terms on the right-hand side of (35) can be made arbitrarily large by selecting a sufficiently large B and sufficiently small τ. As for the first term, we first note that by using (33), we have

$$\text{Re}\left\{\sum_{m=0}^{n-1}\sum_{q=0}^{k}\tilde{x}_m\tilde{y}_m^*\alpha_q^* e^{-\frac{2\pi jmq}{n}}\right\} = \text{Re}\left\{\sum_{m=0}^{n-1}\sum_{q=0}^{k}\alpha_q^*|\tilde{y}_m|^2 e^{-\frac{2\pi jmq}{n}}\sum_{l=0}^{k}\hat{\alpha}_l e^{\frac{2\pi jml}{n}}\right\} \tag{36}$$

$$= \sum_{m=0}^{n-1}|\tilde{y}_m|^2\left|\sum_{l=0}^{k}\hat{\alpha}_l e^{\frac{2\pi jml}{n}}\right|^2. \tag{37}$$

Thus, using the last result we obtain

$$\hat{\sigma}_0^2 = \frac{1}{n}\sum_{m=0}^{n-1}\left|\tilde{x}_m - \tilde{y}_m\sum_{l=0}^{k}\hat{\alpha}_l e^{\frac{2\pi jml}{n}}\right|^2 \tag{38}$$

$$= \frac{1}{n}\sum_{m=0}^{n-1}|\tilde{x}_m|^2 - 2\,\text{Re}\left\{\sum_{m=0}^{n-1}\sum_{l=0}^{k}\tilde{x}_m\tilde{y}_m^*\alpha_l^* e^{-\frac{2\pi jml}{n}}\right\} + \frac{1}{n}\sum_{m=0}^{n-1}|\tilde{y}_m|^2\left|\sum_{l=0}^{k}\hat{\alpha}_l e^{\frac{2\pi jml}{n}}\right|^2 \tag{39}$$

$$= \frac{1}{n}\sum_{m=0}^{n-1}|\tilde{x}_m|^2 - \frac{1}{n}\sum_{m=0}^{n-1}|\tilde{y}_m|^2\left|\sum_{l=0}^{k}\hat{\alpha}_l e^{\frac{2\pi jml}{n}}\right|^2, \tag{40}$$

which in turn must be nonnegative, and hence

$$\frac{1}{n}\sum_{m=0}^{n-1}|\tilde{y}_m|^2\left|\sum_{l=0}^{k}\hat{\alpha}_l e^{\frac{2\pi jml}{n}}\right|^2 \leq \frac{1}{n}\sum_{m=0}^{n-1}|\tilde{x}_m|^2 \leq \sigma_x^2\left(1+\Delta\right). \tag{41}$$

Thus, given that $\min_{0\leq m\leq n-1}|\tilde{y}_m|^2 \geq \tau$, by using (40) we obtain that

$$\frac{1}{n}\sum_{m=0}^{n-1}|\tilde{y}_m|^2\left|\sum_{l=0}^{k}\hat{\alpha}_l e^{\frac{2\pi jml}{n}}\right|^2 \geq \tau\frac{1}{n}\sum_{m=0}^{n-1}\left|\sum_{l=0}^{k}\hat{\alpha}_l e^{\frac{2\pi jml}{n}}\right|^2 \tag{42}$$

$$= \tau\sum_{l=0}^{k}\sum_{r=0}^{k}\hat{\alpha}_l\hat{\alpha}_r^*\frac{1}{n}\sum_{m=0}^{n-1}e^{\frac{2\pi jm(l-r)}{n}} \tag{43}$$

$$= \tau\sum_{l=0}^{k}|\hat{\alpha}_l|^2. \tag{44}$$

Therefore, invoking (41), we finally obtain that

$$\sum_{l=0}^{k}|\hat{\alpha}_l|^2 \leq \frac{\sigma_x^2\left(1+\Delta\right)}{\tau} \triangleq C\left(\tau,\Delta\right). \tag{45}$$

Now, recall that $\{\hat{\alpha}_l\}$ minimizes the quadratic norm

$$\frac{1}{n}\sum_{m=0}^{n-1}\left|\tilde{x}_m - \tilde{y}_m\sum_{l=0}^{k}\alpha_l e^{\frac{2\pi jml}{n}}\right|^2$$

over all vectors $\boldsymbol{\alpha} = (\alpha_0, \ldots, \alpha_k)$ in $\mathbb{C}^{k+1}$. Also, due to (45), the minimizing vector must lie in the $(k+1)$-dimensional hypersphere $\boldsymbol{\alpha}^H \boldsymbol{\alpha} \leq C(\tau, \Delta)$. Now, fix $\delta > 0$ and define the grid $\mathcal{G} \triangleq \{\delta \cdot i : i = -\lceil C(\tau, \Delta)/\delta \rceil, \ldots, -1, 0, 1, \ldots, \lceil C(\tau, \Delta)/\delta \rceil\}$, and let $\mathcal{G}^{k+1}$ designate the $(k+1)$th Cartesian power of $\mathcal{G}$. From the uniform continuity of the above quadratic form within the set of all energy limited vectors $\boldsymbol{y}$, one can find a sufficiently small value of δ (depending on C) such that there exists a vector $\boldsymbol{\alpha} = \boldsymbol{\alpha}_R + j\boldsymbol{\alpha}_I$ where $\boldsymbol{\alpha}_R, \boldsymbol{\alpha}_I \in \mathcal{G}^{k+1}$, i.e., the nearest neighbor of the minimizer, satisfying (given of course the event that $\hat{\sigma}_0^2 < B^{-1}$)

$$\frac{1}{n} \sum_{m=0}^{n-1} \left| \tilde{x}_m - \tilde{y}_m \sum_{l=0}^{k} \alpha_l e^{\frac{2\pi jml}{n}} \right|^2 \leq \frac{2}{B} + \delta' \tag{46}$$

where δ' is a sufficiently small value (depending on δ). For brevity, in the following, we will omit this negligible additive term. Whence

$$\mathbb{P}\left\{ \hat{\sigma}_0^2 < B^{-1}, \frac{1}{n} \sum_{t=0}^{n-1} \left| \tilde{Y}_m \right|^2 \leq \sqrt{B}, \min_{0 \leq m \leq n-1} \left| \tilde{Y}_m \right|^2 \geq \tau \right\}$$

$$= \mathbb{P}\left\{ \frac{1}{n} \sum_{m=0}^{n-1} \left| \tilde{X}_m - \tilde{Y}_m \sum_{l=0}^{k} \hat{\alpha}_l e^{\frac{2\pi jml}{n}} \right|^2 < \frac{1}{B}, \frac{1}{n} \sum_{t=0}^{n-1} \left| \tilde{Y}_m \right|^2 \leq \sqrt{B}, \min_{0 \leq m \leq n-1} \left| \tilde{Y}_m \right|^2 \geq \tau \right\} \tag{47}$$

$$\leq \mathbb{P}\left\{ \bigcup_{\boldsymbol{\alpha}_R, \boldsymbol{\alpha}_I \in \mathcal{G}^{k+1}} \left\{ \frac{1}{n} \sum_{m=0}^{n-1} \left| \tilde{X}_m - \tilde{Y}_m \sum_{l=0}^{k} \alpha_l e^{\frac{2\pi jml}{n}} \right|^2 < \frac{2}{B}, \frac{1}{n} \sum_{t=0}^{n-1} \left| \tilde{Y}_m \right|^2 \leq \sqrt{B} \right\} \right\} \tag{48}$$

$$\leq \sum_{\boldsymbol{\alpha}_R, \boldsymbol{\alpha}_I \in \mathcal{G}^{k+1}} \mathbb{P}\left\{ \frac{1}{n} \sum_{m=0}^{n-1} \left| \tilde{X}_m - \tilde{Y}_m \sum_{l=0}^{k} \alpha_l e^{\frac{2\pi jml}{n}} \right|^2 < \frac{2}{B}, \frac{1}{n} \sum_{t=0}^{n-1} \left| \tilde{Y}_m \right|^2 \leq \sqrt{B} \right\} \tag{49}$$

$$\leq \left(\left\lceil \frac{C(\tau, \Delta)}{\delta} \right\rceil \right)^{2k+2} \cdot \max_{\boldsymbol{\alpha}_R, \boldsymbol{\alpha}_I \in \mathcal{G}^{k+1}} \mathbb{P}\left\{ \frac{1}{n} \sum_{m=0}^{n-1} \left| \tilde{X}_m - \tilde{Y}_m \sum_{l=0}^{k} \alpha_l e^{\frac{2\pi jml}{n}} \right|^2 < \frac{2}{B}, \frac{1}{n} \sum_{t=0}^{n-1} \left| \tilde{Y}_m \right|^2 \leq \sqrt{B} \right\}. \tag{50}$$

Let us show that the term on the right-most side of can be made exponentially less than $e^{-nE(R)}$. Define the set

$$\mathcal{F}_\alpha \triangleq \left\{ (\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}}) : \frac{1}{n} \sum_{m=0}^{n-1} \left| \tilde{x}_m - \tilde{y}_m \sum_{l=0}^{k} \alpha_l e^{\frac{2\pi jml}{n}} \right|^2 < \frac{2}{B}, \frac{1}{n} \sum_{m=0}^{n-1} |\tilde{y}_m|^2 \leq \sqrt{B} \right\}. \tag{51}$$

Accordingly, define an auxiliary joint density

$$g(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}}) = \frac{1}{\left(2\pi^2/\sqrt{B} \right)^n} \prod_{m=0}^{n-1} \exp\left\{ -\frac{B}{2} \left| \tilde{x}_m - \tilde{y}_m \sum_{l=0}^{k} \alpha_l e^{\frac{2\pi jml}{n}} \right|^2 \right\} \exp\left\{ -\frac{1}{\sqrt{B}} |\tilde{y}_m|^2 \right\}. \tag{52}$$

Thus,

$$1 \geq \int_{\mathcal{F}_\alpha} g\left(\mathrm{d}\boldsymbol{x}, \mathrm{d}\boldsymbol{y}\right) \tag{53}$$

$$\geq \frac{\mathrm{Vol}\left\{\mathcal{F}_\alpha\right\}}{\left(2\pi^2/\sqrt{B}\right)^n} \inf_{(\boldsymbol{x},\boldsymbol{y})\in\mathcal{F}_\alpha} \left\{ \prod_{m=0}^{n-1} \exp\left\{ -\frac{B}{2}\left|\tilde{x}_m - \tilde{y}_m \sum_{l=0}^{k} \alpha_l e^{\frac{2\pi jml}{n}}\right|^2 \right\} \exp\left\{ -\frac{1}{\sqrt{B}}|\tilde{y}_m|^2 \right\} \right\} \tag{54}$$

$$\geq \frac{\mathrm{Vol}\left\{\mathcal{F}_\alpha\right\}}{\left(2\pi^2/\sqrt{B}\right)^n} \exp\left\{-2n\right\} \tag{55}$$

$$= \mathrm{Vol}\left\{\mathcal{F}_\alpha\right\} \left(\frac{2\pi^2 e^2}{\sqrt{B}}\right)^{-n}, \tag{56}$$

and therefore $\mathrm{Vol}\left\{\mathcal{F}_\alpha\right\} \leq \left(2\pi^2 e^2/\sqrt{B}\right)^n$. Thus, we now obtain that

$$\mathbb{P}\left\{\mathcal{F}_\alpha\right\} = \int_{(\boldsymbol{x},\boldsymbol{y})\in\mathcal{F}_\alpha} \mu\left(\boldsymbol{x}\right) W\left(\boldsymbol{y}|\boldsymbol{x}\right) \mathrm{d}\boldsymbol{x}\mathrm{d}\boldsymbol{y} \tag{57}$$

$$\leq \mathrm{Vol}\left\{\mathcal{F}_\xi\right\} \left(2\pi\sigma^2\right)^{-n/2} \nu^{-1} \tag{58}$$

$$\leq \left(2\pi\sigma^2\right)^{-n/2} \nu^{-1} \exp\left\{ -\frac{n}{2}\log\left(\frac{B}{2\pi^2 e^2}\right) \right\} \tag{59}$$

$$= \nu^{-1} \exp\left\{ -\frac{n}{2}\log\left(\frac{B\sigma^2}{\pi e^2}\right) \right\} \tag{60}$$

which, again, can be made less than $e^{-nE(R)}$ by selecting B sufficiently large. ∎

To overbound (25) within $H_n(B)$, we derive an upper bound on its numerator and a lower bound on its denominator, and show that these are exponentially equivalent. To this end, we first need to define a conditional typical set of our continuous-valued input-output sequences and establish some of its properties. For a given pair of vectors $(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}})$ and $\epsilon > 0$, define the kth order conditional ϵ-type of $\tilde{\boldsymbol{x}}$ given $\tilde{\boldsymbol{y}}$ as

$$\mathcal{T}_\epsilon^k\left(\tilde{\boldsymbol{x}}|\tilde{\boldsymbol{y}}\right) \triangleq \left\{ \tilde{\boldsymbol{x}}' \in \mathbb{C}^n : \left|\frac{1}{n}\sum_{m=0}^{n-1}|\tilde{x}_m|^2 - \frac{1}{n}\sum_{m=0}^{n-1}|\tilde{x}_m'|^2\right| \leq \epsilon \right.$$
$$\left|\frac{1}{n}\sum_{m=0}^{n-1}\mathrm{Re}\left\{\tilde{x}_m\tilde{y}_m^* e^{-\frac{2\pi jlm}{n}}\right\} - \frac{1}{n}\sum_{m=0}^{n-1}\mathrm{Re}\left\{\tilde{x}_m'\tilde{y}_m^* e^{-\frac{2\pi jlm}{n}}\right\}\right| \leq \epsilon,\ l = 0,\ldots,k,$$
$$\left.\left|\frac{1}{n}\sum_{m=0}^{n-1}\mathrm{Im}\left\{\tilde{x}_m\tilde{y}_m^* e^{-\frac{2\pi jlm}{n}}\right\} - \frac{1}{n}\sum_{m=0}^{n-1}\mathrm{Im}\left\{\tilde{x}_m'\tilde{y}_m^* e^{-\frac{2\pi jlm}{n}}\right\}\right| \leq \epsilon,\ l = 0,\ldots,k \right\}. \tag{61}$$

This set is regarded as a conditional type of $\tilde{\boldsymbol{x}}$ given $\tilde{\boldsymbol{y}}$ as it contains all vectors which, within ϵ, have the same sufficient statistics as $\tilde{\boldsymbol{x}}$ induced by our backward channel. In the following, we will show that for every conditional type $\mathcal{T}_\epsilon^k\left(\tilde{\boldsymbol{x}}|\tilde{\boldsymbol{y}}\right)$, and for any two vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ in $\mathcal{T}_\epsilon^k\left(\tilde{\boldsymbol{x}}|\tilde{\boldsymbol{y}}\right)$, the conditional pdf's

$W(\boldsymbol{y}|\boldsymbol{u})$ and $W(\boldsymbol{y}|\boldsymbol{v})$ are exponentially equivalent. This property will be used later on. To show that this is indeed the case, we will need the following lemma.

Lemma 3 Let L and n_b be natural numbers such that[3] $L = n/n_b$. Define the sets $\mathcal{G}_{1,\epsilon} \triangleq \{\epsilon \cdot i : \ i = 0, 1, \ldots, \lceil LP_x/\epsilon \rceil\}$. Also, let

$$\hat{\mathcal{T}}_\epsilon^k (\tilde{\boldsymbol{x}}|\tilde{\boldsymbol{y}}) \triangleq \left\{ \bigcup_{\boldsymbol{\mathcal{P}}^\epsilon} \overset{L}{\underset{l=1}{\times}} \mathscr{B}_l^\epsilon (P_l) \right\} \bigcap \tilde{\mathcal{T}}_\epsilon (\tilde{\boldsymbol{x}}|\tilde{\boldsymbol{y}}) \tag{62}$$

where $\times$ designates a Cartesian product, and

$$\tilde{\mathcal{T}}_\epsilon^k (\tilde{\boldsymbol{x}}|\tilde{\boldsymbol{y}}) \triangleq \left\{ \tilde{\boldsymbol{x}}' \in \mathbb{C}^n : \ \left| \frac{1}{n} \sum_{m=0}^{n-1} \mathrm{Re}\left\{ \tilde{x}_m \tilde{y}_m^* e^{-\frac{2\pi j l m}{n}} \right\} - \frac{1}{n} \sum_{m=0}^{n-1} \mathrm{Re}\left\{ \tilde{x}_m' \tilde{y}_m^* e^{-\frac{2\pi j l m}{n}} \right\} \right| \le \epsilon, \ l = 0, \ldots, k, \right.$$

$$\left. \left| \frac{1}{n} \sum_{m=0}^{n-1} \mathrm{Im}\left\{ \tilde{x}_m \tilde{y}_m^* e^{-\frac{2\pi j l m}{n}} \right\} - \frac{1}{n} \sum_{m=0}^{n-1} \mathrm{Im}\left\{ \tilde{x}_m' \tilde{y}_m^* e^{-\frac{2\pi j l m}{n}} \right\} \right| \le \epsilon, \ l = 0, \ldots, k \right\}, \tag{63}$$

and

$$\mathscr{B}_l^\epsilon (P_l) \triangleq \left\{ \tilde{\boldsymbol{x}}' \in \mathbb{C}^{n_b} : \ \left| \left\| \tilde{\boldsymbol{x}}' \right\|^2 - n_b P_l \right| \le \epsilon, \right\}$$

where

$$\boldsymbol{\mathcal{P}}^\epsilon \triangleq \left\{ \boldsymbol{P} \in \mathcal{G}_{1,\epsilon}^L : \ \left| \frac{1}{L} \sum_{i=1}^L P_i - \frac{1}{n} \sum_{m=0}^{n-1} |\tilde{x}_m|^2 \right| \le \epsilon \right\} \tag{64}$$

where $\mathcal{G}_{1,\epsilon}^L$ is the Lth Cartesian power of $\mathcal{G}_{1,\epsilon}$. Then,

$$\mathcal{T}_\epsilon^k (\tilde{\boldsymbol{x}}|\tilde{\boldsymbol{y}}) \subseteq \hat{\mathcal{T}}_\epsilon^k (\tilde{\boldsymbol{x}}|\tilde{\boldsymbol{y}}). \tag{65}$$

Proof: See Appendix A. ∎

Intuitively speaking, the difference between $\hat{\mathcal{T}}_\epsilon^k (\tilde{\boldsymbol{x}}|\tilde{\boldsymbol{y}})$ and $\mathcal{T}_\epsilon^k (\tilde{\boldsymbol{x}}|\tilde{\boldsymbol{y}})$ is that in the former we split each sequence $\tilde{\boldsymbol{x}}$ into L bins, where in each bin we fix the energy. Indeed, let $\tilde{\boldsymbol{u}}, \tilde{\boldsymbol{v}} \in \mathcal{T}_\epsilon^k (\tilde{\boldsymbol{x}}|\tilde{\boldsymbol{y}})$. Due to Lemma 3, we also have that $\tilde{\boldsymbol{u}}, \tilde{\boldsymbol{v}} \in \hat{\mathcal{T}}_\epsilon^k (\tilde{\boldsymbol{x}}|\tilde{\boldsymbol{y}})$. Then,

$$\left| \frac{1}{n} \log W(\boldsymbol{y}|\boldsymbol{u}) - \frac{1}{n} \log W(\boldsymbol{y}|\boldsymbol{v}) \right| = \frac{1}{2\sigma^2} \left| \frac{1}{n} \sum_{t=0}^{n-1} \left(y_t - \sum_{l=0}^k h_l u_{t-l} \right)^2 - \frac{1}{n} \sum_{t=0}^{n-1} \left(y_t - \sum_{l=0}^k h_l v_{t-l} \right)^2 \right|. \tag{66}$$

Recall that the model in (1) can be represented in the following vector form $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{w}$ where $\boldsymbol{A}$ is a Toeplitz matrix formed by the generating sequence $\{h_l\}$, that is $\boldsymbol{A} = \{a_{i,j}\}_{i,j} = \{h_{i-j}\}_{i,j}$. Now, by using the spectral decomposition theorem [20], we know that there exists a unitary matrix $\boldsymbol{R}$ that

---

[3]Without loss of generality, it is assumed that n_b (*bin* size) is a divisor of n, and that all L bins have the same size.

diagonalizes A. Accordingly, let $\{\lambda_l\}_{l=1}^n$ denote the singular values associated with this transformation. Thus, we obtain that

$$\left|\frac{1}{n}\log W\left(\boldsymbol{y}|\boldsymbol{u}\right) - \frac{1}{n}\log W\left(\boldsymbol{y}|\boldsymbol{v}\right)\right| = \frac{1}{2\sigma^2}\left|\frac{1}{n}\sum_{m=0}^{n-1}|\hat{y}_m - \lambda_m\hat{u}_m|^2 - \frac{1}{n}\sum_{t=0}^{n-1}|\hat{y}_m - \lambda_m\hat{v}_m|^2\right| \tag{67}$$

where $\hat{\boldsymbol{y}} = \boldsymbol{R}\boldsymbol{y}$ and similarly for $\{\hat{u}_m\}$ and $\{\hat{v}_m\}$. Continuing, we see that

$$\left|\frac{1}{n}\log W\left(\boldsymbol{y}|\boldsymbol{u}\right) - \frac{1}{n}\log W\left(\boldsymbol{y}|\boldsymbol{v}\right)\right| \leq \frac{1}{\sigma^2}\left|\frac{1}{n}\sum_{m=0}^{n-1}\mathrm{Re}\left\{\hat{y}_m^*\lambda_m\hat{u}_m\right\} - \frac{1}{n}\sum_{m=0}^{n-1}\mathrm{Re}\left\{\hat{y}_m^*\lambda_m\hat{v}_m\right\}\right|$$

$$+ \frac{1}{2\sigma^2}\left|\frac{1}{n}\sum_{m=0}^{n-1}|\lambda_m|^2\left(|\hat{u}_m|^2 - |\hat{v}_m|^2\right)\right|. \tag{68}$$

Now, we note that by Szegö's theorem [19-22], the Fourier basis asymptotically diagonalizes Toeplitz matrices. Accordingly, the asymptotic eigenvalues are given by the DFT of the generating sequence $\{h_l\}$, that is, for sufficiently large enough n and any $\varepsilon > 0$, we have that [20]

$$\left|\lambda_m - \sum_{l=0}^{k}h_l e^{-2\pi jml/n}\right| \leq \varepsilon, \quad m = 0, \ldots, n-1, \tag{69}$$

and by the same token[4], since the Fourier basis asymptotically diagonalizes A, the eigenvectors matrix R asymptotically equal to the Fourier basis F. Thus, using (61) and (69), we see that

$$\left|\frac{1}{n}\sum_{m=0}^{n-1}\mathrm{Re}\left\{\hat{y}_m^*\lambda_m\hat{u}_m\right\} - \frac{1}{n}\sum_{m=0}^{n-1}\mathrm{Re}\left\{\hat{y}_m^*\lambda_m\hat{v}_m\right\}\right|$$

$$\leq \sum_{l=0}^{k}|h_l|\left|\varepsilon + \frac{1}{n}\sum_{m=0}^{n-1}\mathrm{Re}\left\{\tilde{y}_m^*\tilde{u}_m e^{-2\pi jml/n}\right\} - \frac{1}{n}\sum_{m=0}^{n-1}\mathrm{Re}\left\{\tilde{y}_m^*\tilde{v}_m e^{-2\pi jml/n}\right\}\right| \tag{70}$$

$$\leq (\epsilon + \varepsilon)\sum_{l=0}^{k}|h_l| \leq (\epsilon + \varepsilon)\cdot C_1 \tag{71}$$

where in the last inequality we have used the fact that $\{h_k\}$ is absolutely summable. Now, regarding the second term on the right hand side (r.h.s.) of (68), we use the following approximation argument (which is asymptotically tight), that was used in [16, Sec. VI]. Recall that due to Szegö's theorem, we know that the Fourier basis asymptotically diagonalizes A, and that there exists a frequency response $H(\omega)$ that corresponds to the linear system induced by A, and is given by the Fourier transform of the sequence $\{h_i\}$. Then, we use the fact that every continuous function can be approximated arbitrarily well by a sequence of staircase functions with sufficiently small spacing between jumps. In other words, we

---

[4]Another approach is to first assume that A is a circulant matrix, and then the Fourier basis exactly diagonalizes A for any n, that is, the eigenvectors are given by the DFT matrix, and the eigenvalues are given by the DFT of $\{h_l\}$. Then, when taking the limit $n \to \infty$, using Szegö's theorem, this assumption can be dropped.

approximate the continuous frequency response $H(\omega)$ by a staircase function and then we take the width of each stair to zero. This approximation in turn corresponds to assuming that the eigenvalues, $\{\lambda_m\}$, are piecewise constant over the various L bins (see Lemma 3). At the final stage of the analysis (after taking the limit $n \to \infty$), we will take the limit $L \to \infty$ so that this approximation becomes superfluous. Thus, under this approximation, we obtain that

$$\left| \frac{1}{n} \sum_{m=0}^{n-1} |\lambda_m|^2 \left(|\tilde{u}_m|^2 - |\tilde{v}_m|^2 \right) \right| = \left| \frac{1}{L} \sum_{l=1}^{L} \frac{1}{n_b} \sum_{m \in \mathcal{I}_l} |\lambda_m|^2 \left(|\tilde{u}_m|^2 - |\tilde{v}_m|^2 \right) \right| \tag{72}$$

$$= \left| \frac{1}{L} \sum_{l=1}^{L} |\lambda_l|^2 \frac{1}{n_b} \sum_{m \in \mathcal{I}_l} \left(|\tilde{u}_m|^2 - |\tilde{v}_m|^2 \right) \right| \tag{73}$$

$$\leq \frac{1}{L} \sum_{l=1}^{L} |\lambda_l|^2 \left| \frac{1}{n_b} \sum_{m \in \mathcal{I}_l} \left(|\tilde{u}_m|^2 - |\tilde{v}_m|^2 \right) \right| \tag{74}$$

$$\leq \frac{\epsilon}{L} \sum_{l=1}^{L} |\lambda_l|^2 \leq \frac{\epsilon}{L} \sum_{l=1}^{L} \left[\sum_{v=0}^{k} |h_v| \right]^2 \leq \epsilon \cdot C_1^2. \tag{75}$$

Thus, we have shown that

$$\left| \frac{1}{n} \log W(\boldsymbol{y}|\boldsymbol{u}) - \frac{1}{n} \log W(\boldsymbol{y}|\boldsymbol{v}) \right| \leq \frac{(\epsilon + \varepsilon)}{\sigma^2} \cdot C_1 (1 + C_1). \tag{76}$$

Clearly, the right-most side of (76) can be made arbitrarily small by choosing ϵ sufficiently small and n, L sufficiently large. Similarly, $\mu(\boldsymbol{u})$ and $\mu(\boldsymbol{v})$ are also exponentially equivalent, provided that they both belong to the support of $\mu(\cdot)$, namely,

$$\left| \frac{1}{n} \log \mu(\boldsymbol{u}) - \frac{1}{n} \log \mu(\boldsymbol{v}) \right| \leq \epsilon \cdot C_2 \tag{77}$$

for some constant C_2. Next, we provide upper and lower bounds on the volume of $\mathcal{T}_\epsilon^k(\tilde{\boldsymbol{x}}|\tilde{\boldsymbol{y}})$, where the volume of a set $\mathcal{A} \subset \mathbb{R}^n$ is defined as $\mathrm{Vol}\{\mathcal{A}\} \triangleq \int_{\mathcal{A}} \mathrm{d}\boldsymbol{x}$.

Lemma 4 Let $(\boldsymbol{x}, \boldsymbol{y}) \in H_n(B)$ for some $B > 0$. Then, for every sufficiently small $\epsilon > 0$, the volume of $\mathcal{T}_\epsilon^k(\tilde{\boldsymbol{x}}|\tilde{\boldsymbol{y}})$ is bounded as follows

$$\frac{\exp\{-n\epsilon f(B, \Delta, k)\}}{\max_{\boldsymbol{\theta}} V(\tilde{\boldsymbol{x}}|\tilde{\boldsymbol{y}}, \boldsymbol{\theta}, k)} \left[1 - (2k + 12)\frac{B^2}{n\epsilon^2} \right] \leq \mathrm{Vol}\left\{ \mathcal{T}_\epsilon^k(\tilde{\boldsymbol{x}}|\tilde{\boldsymbol{y}}) \right\} \leq \frac{\exp\{n\epsilon f(B, \Delta, k)\}}{\max_{\boldsymbol{\theta}} V(\tilde{\boldsymbol{x}}|\tilde{\boldsymbol{y}}, \boldsymbol{\theta}, k)}, \tag{78}$$

in which $f(B, \Delta, k) \triangleq B\left[1 + \sqrt{k+1} \cdot C(B^{-1}, \Delta) \right]$ where $C(\cdot, \cdot)$ is defined in (45).

Proof of Lemma 4: Fix a pair $(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}}) \in H_n(B)$ and let

$$\rho_{xx} \triangleq n^{-1} \sum_{m=0}^{n-1} |\tilde{x}_m|^2, \tag{79}$$

$$\rho_R^l \overset{\triangle}{=} n^{-1} \sum_{m=0}^{n-1} \text{Re} \left\{ \tilde{x}_m \tilde{y}_m^* e^{-\frac{2\pi j l m}{n}} \right\}, \ l = 0, \ldots, k, \tag{80}$$

and

$$\rho_I^l \overset{\triangle}{=} n^{-1} \sum_{m=0}^{n-1} \text{Im} \left\{ \tilde{x}_m \tilde{y}_m^* e^{-\frac{2\pi j l m}{n}} \right\}, \ l = 0, \ldots, k. \tag{81}$$

Also, let $\hat{\boldsymbol{\theta}}$ designate the vector of parameters $\left(\sigma_0^2, \alpha_0, \ldots, \alpha_k \right)$ that corresponds to the solution of the following set of equations

$$\mathbb{E}_V \left\{ \sum_{m=0}^{n-1} \left| \tilde{X}_m \right|^2 \right\} = n\rho_{x,x}, \tag{82}$$

and

$$\mathbb{E}_V \left\{ \sum_{m=0}^{n-1} \text{Re} \left\{ \tilde{X}_m \tilde{y}_m^* e^{-\frac{2\pi j l m}{n}} \right\} \right\} = n\rho_R^l, \ l = 0, \ldots, k, \tag{83}$$

and

$$\mathbb{E}_V \left\{ \sum_{m=0}^{n-1} \text{Im} \left\{ \tilde{X}_m \tilde{y}_m^* e^{-\frac{2\pi j l m}{n}} \right\} \right\} = n\rho_I^l, \ l = 0, \ldots, k \tag{84}$$

where the expectation $\mathbb{E}_V$ is taken w.r.t. the backward channel $V \left(\cdot | \tilde{\boldsymbol{y}}, \boldsymbol{\theta}, k \right)$. This parameter vector can be found by solving the set of equations (33)-(34), namely, it attains the maximum of $V \left(\tilde{\boldsymbol{x}} | \tilde{\boldsymbol{y}}, \boldsymbol{\theta}, k \right)$ as can be easily seen. Then,

$$1 \geq V \left(\left\{ \mathcal{T}_\epsilon^k \left(\tilde{\boldsymbol{x}} | \tilde{\boldsymbol{y}} \right) \right\} | \tilde{\boldsymbol{y}}, \hat{\boldsymbol{\theta}}, k \right) \tag{85}$$

$$= \int_{\mathcal{T}_\epsilon^k (\tilde{\boldsymbol{x}} | \tilde{\boldsymbol{y}})} V(\bar{\boldsymbol{x}} | \tilde{\boldsymbol{y}}, \hat{\boldsymbol{\theta}}, k) \mathrm{d}\bar{\boldsymbol{x}} \tag{86}$$

$$\geq \text{Vol} \left\{ \mathcal{T}_\epsilon^k \left(\tilde{\boldsymbol{x}} | \tilde{\boldsymbol{y}} \right) \right\} \inf_{\bar{\boldsymbol{x}} \in \mathcal{T}_\epsilon^k (\tilde{\boldsymbol{x}} | \tilde{\boldsymbol{y}})} V(\bar{\boldsymbol{x}} | \tilde{\boldsymbol{y}}, \hat{\boldsymbol{\theta}}, k) \tag{87}$$

$$\geq \text{Vol} \left\{ \mathcal{T}_\epsilon^k \left(\tilde{\boldsymbol{x}} | \tilde{\boldsymbol{y}} \right) \right\} \exp \left\{ -n \left[\frac{1}{2\hat{\sigma}^2} \left(1 + 2 \sum_{l=0}^{k} |\hat{\alpha}_l| \right) \epsilon \right] \right\} V(\tilde{\boldsymbol{x}} | \tilde{\boldsymbol{y}}, \hat{\boldsymbol{\theta}}, k) \tag{88}$$

$$\geq \text{Vol} \left\{ \mathcal{T}_\epsilon^k \left(\tilde{\boldsymbol{x}} | \tilde{\boldsymbol{y}} \right) \right\} \exp \left\{ -n\epsilon B \left[1 + \sqrt{k+1} \cdot C \left(B^{-1}, \Delta \right) \right] \right\} V(\tilde{\boldsymbol{x}} | \tilde{\boldsymbol{y}}, \hat{\boldsymbol{\theta}}, k) \tag{89}$$

where the second last inequality readily follows from a derivation similar to (76) and the fact that $(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}}) \in H_n(B)$, and the last inequality follows from (45) along with the fact that for any sequence $\boldsymbol{z} = (z_1, \ldots, z_n)$, we have $\|\boldsymbol{z}\|_1 \leq \sqrt{n} \|\boldsymbol{z}\|_{\ell_2}$. Thus, we obtain

$$\text{Vol} \left\{ \mathcal{T}_\epsilon^k \left(\tilde{\boldsymbol{x}} | \tilde{\boldsymbol{y}} \right) \right\} \leq \frac{\exp \left\{ n\epsilon f \left(B, \Delta, k \right) \right\}}{\max_\theta V \left(\tilde{\boldsymbol{x}} | \tilde{\boldsymbol{y}}, \boldsymbol{\theta}, k \right)} = \exp \left\{ n\epsilon f \left(B, \Delta, k \right) \right\} \exp \left\{ n \log \left(\pi e \hat{\sigma}_0^2 \right) \right\} \tag{90}$$

For a lower bound on the volume, we first note that

$$1 = V \left(\left\{ \mathcal{T}_\epsilon^k \left(\tilde{\boldsymbol{x}} | \tilde{\boldsymbol{y}} \right) \cup \left\{ \mathcal{T}_\epsilon^k \left(\tilde{\boldsymbol{x}} | \tilde{\boldsymbol{y}} \right) \right\}^c \right\} | \tilde{\boldsymbol{y}}, \hat{\boldsymbol{\theta}}, k \right) \tag{91}$$

$$\leq \text{Vol}\left\{\mathcal{T}_\epsilon^k\left(\tilde{\boldsymbol{x}}|\tilde{\boldsymbol{y}}\right)\right\} \exp\left\{n\epsilon f\left(B,\Delta,k\right)\right\} \max_{\boldsymbol{\theta}} V\left(\tilde{\boldsymbol{x}}|\tilde{\boldsymbol{y}},\boldsymbol{\theta},k\right) + V\left(\left\{\mathcal{T}_\epsilon^k\left(\tilde{\boldsymbol{x}}|\tilde{\boldsymbol{y}}\right)\right\}^c|\tilde{\boldsymbol{y}},\hat{\boldsymbol{\theta}},k\right) \tag{92}$$

where the last inequality follows by the same considerations in (89). Using Boole's inequality

$$V\left(\left\{\mathcal{T}_\epsilon^k\left(\tilde{\boldsymbol{x}}|\tilde{\boldsymbol{y}}\right)\right\}^c|\tilde{\boldsymbol{y}},\hat{\boldsymbol{\theta}},k\right) \leq V\left(\left|\frac{1}{n}\sum_{m=0}^{n-1}\left|\tilde{X}_m\right|^2 - \rho_{xx}\right| > \epsilon\bigg|\tilde{\boldsymbol{y}},\hat{\boldsymbol{\theta}},k\right)$$
$$+ \sum_{l=0}^{k} V\left(\left|\frac{1}{n}\sum_{m=0}^{n-1}\text{Re}\left\{\tilde{X}_m\tilde{y}_m^* e^{-\frac{2\pi jlm}{n}}\right\} - \rho_R^l\right| > \epsilon\bigg|\tilde{\boldsymbol{y}},\hat{\boldsymbol{\theta}},k\right)$$
$$+ \sum_{l=0}^{k} V\left(\left|\frac{1}{n}\sum_{m=0}^{n-1}\text{Im}\left\{\tilde{X}_m\tilde{y}_m^* e^{-\frac{2\pi jlm}{n}}\right\} - \rho_I^l\right| > \epsilon\bigg|\tilde{\boldsymbol{y}},\hat{\boldsymbol{\theta}},k\right) \tag{93}$$

Now, due to (82)-(84), the events in (93) are large deviations events. For example, for the second term on the right hand side of (93), let us define the following Gaussian density

$$\delta_G\left(\boldsymbol{z}\right) = \frac{1}{\left(\pi\hat{\sigma}_0^2\right)^n}\exp\left\{-\frac{1}{\hat{\sigma}_0^2}\sum_{m=0}^{n-1}|z_m|^2\right\}. \tag{94}$$

Whence, by Chebychev's inequality we obtain, for any $0 \leq l \leq k$,

$$V\left(\left|\frac{1}{n}\sum_{m=0}^{n-1}\text{Re}\left\{\tilde{X}_m\tilde{y}_m^* e^{-\frac{2\pi jlm}{n}}\right\} - \rho_{xy}^l\right| > \epsilon\bigg|\tilde{\boldsymbol{y}},\hat{\boldsymbol{\theta}},k\right) = \delta_G\left\{\boldsymbol{Z}: \left|\frac{1}{n}\sum_{m=0}^{n-1}\text{Re}\left\{Z_m\tilde{y}_m^* e^{-\frac{2\pi jlm}{n}}\right\}\right| > \epsilon\right\}$$
$$\leq \frac{1}{\epsilon^2}\mathbb{E}_\delta\left(\frac{1}{n}\sum_{m=0}^{n-1}\text{Re}\left\{Z_m\tilde{y}_m^* e^{-\frac{2\pi jlm}{n}}\right\}\right)^2 \tag{95}$$
$$\leq \frac{1}{n\epsilon^2}\left[\sum_{m=0}^{n-1}|\tilde{y}_m|^2\right]\mathbb{E}_\delta\left\{\frac{1}{n}\sum_{m=0}^{n-1}|Z_m|^2\right\} \tag{96}$$
$$\leq \frac{B\hat{\sigma}_0^2}{n\epsilon^2} \leq \frac{B^2}{n\epsilon^2}. \tag{97}$$

For the third term on the right hand side of (93), we again have that

$$V\left(\left|\frac{1}{n}\sum_{m=0}^{n-1}\text{Im}\left\{\tilde{X}_m\tilde{y}_m^* e^{-\frac{2\pi jlm}{n}}\right\} - \rho_{xy}^l\right| > \epsilon\bigg|\tilde{\boldsymbol{y}},\hat{\boldsymbol{\theta}},k\right) \leq \frac{1}{\epsilon^2}\mathbb{E}_\delta\left(\frac{1}{n}\sum_{m=0}^{n-1}\text{Im}\left\{Z_m\tilde{y}_m^* e^{-\frac{2\pi jlm}{n}}\right\}\right)^2 \tag{98}$$
$$\leq \frac{B^2}{n\epsilon^2}. \tag{99}$$

Finally, exactly in the same way, one obtains that

$$V\left(\left|\frac{1}{n}\sum_{m=0}^{n-1}\text{Re}\left\{\left|\tilde{X}_m\right|^2 e^{\frac{2\pi jlm}{n}}\right\} - \rho_{xx}^l\right| > \epsilon\bigg|\tilde{\boldsymbol{y}},\hat{\boldsymbol{\theta}},k\right) \leq \frac{12B^2}{n\epsilon^2} \tag{100}$$

Therefore, using (92), (97), (99), and (100), we finally conclude that

$$\text{Vol}\left\{\mathcal{T}_\epsilon^k\left(\tilde{\boldsymbol{x}}|\tilde{\boldsymbol{y}}\right)\right\} \geq \frac{\exp\left\{-n\epsilon f\left(B,\Delta,k\right)\right\}}{\max_\theta V\left(\tilde{\boldsymbol{x}}|\tilde{\boldsymbol{y}},\boldsymbol{\theta},k\right)}\left[1 - \frac{12B^2}{n\epsilon^2} - 2k\frac{B^2}{n\epsilon^2}\right] \tag{101}$$

$$\geq \frac{\exp\left\{-n\epsilon f\left(B, \Delta, k\right)\right\}}{\max_{\boldsymbol{\theta}} V\left(\tilde{\boldsymbol{x}}|\tilde{\boldsymbol{y}}, \boldsymbol{\theta}, k\right)} \left[1 - \left(2k + 12\right)\frac{B^2}{n\epsilon^2}\right]. \tag{102}$$

∎

We are now ready to derive a lower bound on the denominator of (25). Since $\tilde{\boldsymbol{x}} \in \mathcal{S}_o^{\delta}\left(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}}\right)$, then, in view of (76), there exist a sufficiently small $\epsilon > 0$ and a sufficiently large n (both depending on δ) such that $\mathcal{T}_{\epsilon}^k\left(\tilde{\boldsymbol{x}}|\tilde{\boldsymbol{y}}\right) \subset \mathcal{S}_o^{\delta}\left(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}}\right)$. Thus, using Lemma 4, we get

$$\int_{\mathcal{S}_o^{\delta}(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}})} \mu\left(\boldsymbol{x}'\right) \mathrm{d}\boldsymbol{x}' \geq \int_{\mathcal{T}_{\epsilon}^k(\tilde{\boldsymbol{x}}|\tilde{\boldsymbol{y}})} \mu\left(\boldsymbol{x}'\right) \mathrm{d}\boldsymbol{x}' \tag{103}$$

$$\geq \mathrm{Vol}\left\{\mathcal{T}_{\epsilon}^k\left(\tilde{\boldsymbol{x}}|\tilde{\boldsymbol{y}}\right)\right\} \cdot \inf_{\boldsymbol{x}' \in \mathcal{T}_{\epsilon}^k(\tilde{\boldsymbol{x}}|\tilde{\boldsymbol{y}})} \mu\left(\boldsymbol{x}'\right) \tag{104}$$

$$\geq \frac{\exp\left\{-n\epsilon f\left(B, \Delta, k\right)\right\}}{\max_{\boldsymbol{\theta}} V\left(\tilde{\boldsymbol{x}}|\tilde{\boldsymbol{y}}, \boldsymbol{\theta}, k\right)} \left[1 - \left(2k + 12\right)\frac{B^2}{n\epsilon^2}\right] e^{-nC_2\epsilon} \mu\left(\tilde{\boldsymbol{x}}\right). \tag{105}$$

We next overbound the numerator of (25). The basic idea here is to decompose $\mathcal{S}_u\left(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}}\right)$ into subexponential number of conditional types, where for each conditional type, $\int_{\mathcal{T}_{\epsilon}^k(\tilde{\boldsymbol{x}}|\tilde{\boldsymbol{y}})} \mu\left(\boldsymbol{x}'\right) \mathrm{d}\boldsymbol{x}'$ is overestimated using Lemma 4. Yet, this cannot be done directly, simply because not every $\tilde{\boldsymbol{x}}' \in \mathcal{S}_u\left(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}}\right)$ is such that $(\tilde{\boldsymbol{x}}', \tilde{\boldsymbol{y}}) \in H_n\left(B\right)$ and hence we cannot apply Lemma 4 to $\mathcal{T}_{\epsilon}^k\left(\tilde{\boldsymbol{x}}'|\tilde{\boldsymbol{y}}\right)$. Thus, in order to alleviate this difficulty, let us divide $\mathcal{S}_u\left(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}}\right)$ into two subsets, $\mathcal{S}_u\left(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}}\right) \cap H_n\left(B_0|\tilde{\boldsymbol{y}}\right)$ and $\mathcal{S}_u\left(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}}\right) \cap H_n^c\left(B_0|\tilde{\boldsymbol{y}}\right)$, where $H_n\left(B_0|\tilde{\boldsymbol{y}}\right) \stackrel{\triangle}{=} \left\{\boldsymbol{x}' : (\boldsymbol{x}', \tilde{\boldsymbol{y}}) \in H_n\left(B_0\right)\right\}$, $B_0 \geq B$, being a constant to be chosen later. Now, in the first set we can apply Lemma 4 while the second has a very low probability provided that B_0 is sufficiently large. Let B be large enough so that (23) holds and fix $(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}}) \in H_n\left(B\right)$. Similarly to Lemma 2, one can choose B_0 so large such that for every $\boldsymbol{y}' \in H_n\left(B|\tilde{\boldsymbol{x}}\right)$, we have

$$\int_{H_n^c(B_0|\tilde{\boldsymbol{y}})} \mu\left(\boldsymbol{x}'\right) \mathrm{d}\boldsymbol{x}' \leq e^{-nQ(B_0)}, \tag{106}$$

for all large n, where $Q\left(B_0\right) > 0$ can be made arbitrarily large. Thus, we have

$$\int_{\mathcal{S}_u(\boldsymbol{x}, \boldsymbol{y})} \mu\left(\boldsymbol{x}'\right) \mathrm{d}\boldsymbol{x}' \leq \int_{\mathcal{S}_u(\boldsymbol{x}, \boldsymbol{y}) \cap H_n(B_0|\boldsymbol{y})} \mu\left(\boldsymbol{x}'\right) \mathrm{d}\boldsymbol{x}' + e^{-nQ(B_0)}. \tag{107}$$

Let us now subdivide the domain of the first term on the r.h.s. of the above inequality into conditional ϵ-types, whose volumes can be overestimated by Lemma 4. To this end, we will need the number of such sets required to cover the whole domain of integration, that is $\mathcal{S}_u\left(\boldsymbol{x}, \boldsymbol{y}\right) \cap H_n\left(B_0|\boldsymbol{y}\right) \subset H_n\left(B_0|\boldsymbol{y}\right)$. We note that within this set, $n^{-1}\sum_{m=0}^{n-1}|\tilde{x}'_m|^2 \leq B_0$, $n^{-1}\sum_{m=0}^{n-1}|\tilde{y}_m|^2 \leq B_0$, and hence also $n^{-1}\left|\sum_{m=0}^{n-1}\mathrm{Re}\left\{\tilde{x}'_m\tilde{y}_m^* e^{2\pi jlm/n}\right\}\right| \leq B_0$ and $n^{-1}\left|\sum_{m=0}^{n-1}\mathrm{Im}\left\{\tilde{x}'_m\tilde{y}_m^* e^{2\pi jlm/n}\right\}\right| \leq B_0$ for all $l = 0, \ldots, k$. Thus, the number of conditional types $\left\{\mathcal{T}_{\epsilon}^k\left(\boldsymbol{x}'|\boldsymbol{y}\right)\right\}$ needed to cover $H_n\left(B_0|\boldsymbol{y}\right)$ is not larger than $\left(2B_0/\epsilon\right)^{2k+3}$. Therefore,

$$\int_{\mathcal{S}_u(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}}) \cap H_n(B_0|\tilde{\boldsymbol{y}})} \mu\left(\boldsymbol{x}'\right) \mathrm{d}\boldsymbol{x}' \leq \sum_{\mathcal{T}_{\epsilon}^k(\tilde{\boldsymbol{x}}'|\tilde{\boldsymbol{y}}) \subset \mathcal{S}_u(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}}) \cap H_n(B_0|\tilde{\boldsymbol{y}})} \int_{\mathcal{T}_{\epsilon}^k(\tilde{\boldsymbol{x}}'|\tilde{\boldsymbol{y}})} \mu\left(\boldsymbol{x}''\right) \mathrm{d}\boldsymbol{x}'' \tag{108}$$

$$\leq \left(\frac{2B_0}{\epsilon}\right)^{2k+3} \sup_{\tilde{\boldsymbol{x}}' \in \mathcal{S}_u(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}})} \left\{ \text{Vol}\left\{\mathcal{T}_\epsilon^k\left(\tilde{\boldsymbol{x}}'|\tilde{\boldsymbol{y}}\right)\right\} \cdot \sup_{\boldsymbol{x}'' \in \mathcal{T}_\epsilon^k(\tilde{\boldsymbol{x}}'|\tilde{\boldsymbol{y}})} \mu\left(\boldsymbol{x}''\right) \right\}$$

$$\leq \left(\frac{2B_0}{\epsilon}\right)^{2k+3} \exp\left\{n\epsilon f\left(B, \Delta, k\right)\right\} e^{nC_2\epsilon} \sup_{\tilde{\boldsymbol{x}}' \in \mathcal{S}_u(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}})} \frac{\mu\left(\tilde{\boldsymbol{x}}'\right)}{\max_{\boldsymbol{\theta}} V\left(\tilde{\boldsymbol{x}}'|\tilde{\boldsymbol{y}}, \boldsymbol{\theta}, k\right)}$$

$$\leq \left(\frac{2B_0}{\epsilon}\right)^{2k+3} \exp\left\{n\epsilon f\left(B, \Delta, k\right)\right\} e^{nC_2\epsilon} \frac{\mu\left(\tilde{\boldsymbol{x}}\right)}{\max_{\boldsymbol{\theta}} V\left(\tilde{\boldsymbol{x}}|\tilde{\boldsymbol{y}}, \boldsymbol{\theta}, k\right)}. \tag{109}$$

Therefore, combining (105), (107), and (109), we get for all sufficiently large n,

$$\sup_{(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}}) \in H_n(B)} \frac{\int_{\mathcal{S}_u(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}})} \mu\left(\boldsymbol{x}'\right) \mathrm{d}\boldsymbol{x}'}{\int_{\mathcal{S}_0^\delta(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}})} \mu\left(\boldsymbol{x}'\right) \mathrm{d}\boldsymbol{x}'} \leq \left[1 - (2k+12)\frac{B^2}{n\epsilon^2}\right]^{-1} \left(\frac{2B_0}{\epsilon}\right)^{2k+3} e^{2n\epsilon[C_2 + f(B, \Delta, k)]}$$

$$\cdot \left[1 + e^{-nQ(B_0)} \sup_{(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}}) \in H_n(B)} \frac{\max_{\boldsymbol{\theta}} V\left(\tilde{\boldsymbol{x}}|\tilde{\boldsymbol{y}}, \boldsymbol{\theta}, k\right)}{\mu\left(\tilde{\boldsymbol{x}}\right)}\right]. \tag{110}$$

We next provide the conditions under which the last bound is indeed a subexponential function of n. To this end, let us first handle the squared brackets in (110), and show it tends to unity as $n \to \infty$ by choosing $Q(B_0)$ to be sufficiently large. Note that the supremum can be bounded by

$$\sup_{(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}}) \in H_n(B)} \frac{\max_{\boldsymbol{\theta}} V\left(\tilde{\boldsymbol{x}}|\tilde{\boldsymbol{y}}, \boldsymbol{\theta}, k\right)}{\mu\left(\tilde{\boldsymbol{x}}\right)} = \sup_{(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}}) \in H_n(B)} \frac{\left(\pi e \hat{\sigma}_0^2\right)^{-n/2}}{\mu\left(\tilde{\boldsymbol{x}}\right)} \tag{111}$$

$$\leq \frac{\left(\pi e B^{-1}\right)^{-n/2}}{\nu^{-1} e^{(1+\Delta)n/2}}, \tag{112}$$

and that the normalization constant ν can also be upper bounded as follows

$$\nu = \int_{\boldsymbol{x} \in \Psi_\Delta} \mathrm{d}\boldsymbol{x} \exp\left\{-\frac{1}{2\sigma_x^2} \sum_{t=0}^{n-1} x_t^2\right\} \leq e^{-1(1-\Delta)n/2} \left[2\pi e \sigma_x^2 (1 + \Delta)\right]^{n/2}. \tag{113}$$

Whence, using the last results and (110), we see that by choosing B_0 so large so that

$$Q(B_0) > \frac{1}{2} \left[\log B + \log \sigma_x^2 + \log(1 + \Delta) + 2\Delta\right], \tag{114}$$

the last term in the squared brackets in (110) tends to unity as $n \to \infty$, as required. Thus, in order that (110) will be a subexponential function of n, we let $\epsilon = \epsilon_n$ tend to zero and $k = k_n$, such that

$$\lim_{n \to \infty} \frac{1}{n} \log\left\{\left[1 - (2k_n + 12)\frac{B^2}{n\epsilon_n^2}\right]^{-1} \left(\frac{2B_0}{\epsilon_n}\right)^{2k_n+3} e^{2n\epsilon_n[C_2 + f(B, \Delta, k_n)]}\right\} = 0, \tag{115}$$

or, equivalently, that the following hold simultaneously

$$k_n \log \frac{1}{\epsilon_n} = o(n), \tag{116}$$

$$\lim_{n \to \infty} \sqrt{k_n} \epsilon_n = 0, \tag{117}$$

and

$$\lim_{n \to \infty} \frac{n \epsilon_n^2}{k_n} = C_2, \tag{118}$$

where C_2 is some sufficiently large constant, and (116), (117), and (118) follow from the midterm, right, and left terms on the left hand side of (115). This happens if $\epsilon_n = o\left(n^{-1/4}\right)$ and hence $k_n = o\left(n^{1/2}\right)$. Whence, we obtain that (110) is subexponential function of n, and thus

$$\lim_{n \to \infty} \frac{1}{n} \log \bar{P}_{e,u}(R, n) \leq \frac{1}{n} \log \bar{P}_{e,0}^{\delta}(R, n), \tag{119}$$

as required. Finally, to complete the proof of the theorem, it remains to show (22). Note that both $\mathcal{S}_0(\boldsymbol{x}, \boldsymbol{y})$ and $\mathcal{S}_0^{\delta}(\boldsymbol{x}, \boldsymbol{y})$ correspond to a known channel. This is, actually, a similar (and simpler) problem to that we considered above, and is very related to the problem considered in [1, Eqs. (33)-(39)], where (22) has been proven. In the sequel, we briefly describe how to obtain (22). Similarly to the above analysis, using Lemma 1, we would like to show that the ratio

$$\frac{\int_{\mathcal{S}_o^{\delta}(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}})} \mu(\boldsymbol{x}') \, \mathrm{d}\boldsymbol{x}'}{\int_{\mathcal{S}_o(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}})} \mu(\boldsymbol{x}') \, \mathrm{d}\boldsymbol{x}'} \tag{120}$$

is uniformly overbounded by a subexponential function of n, over $(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}}) \in H_n$ where H_n is defined exactly as in (27). For a given pair of vectors $(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}})$ and $\epsilon > 0$, define the kth order conditional ϵ-type $\mathcal{T}_\epsilon^k(\tilde{\boldsymbol{x}}|\tilde{\boldsymbol{y}})$ exactly as in (61). Accordingly, we know that for any $\tilde{\boldsymbol{u}}, \tilde{\boldsymbol{v}} \in \mathcal{T}_\epsilon^k(\tilde{\boldsymbol{x}}|\tilde{\boldsymbol{y}})$ the conditional pdf's $W(\boldsymbol{y}|\boldsymbol{u})$ and $W(\boldsymbol{y}|\boldsymbol{v})$ are exponentially equivalent, that is, (76) holds. Then, in view of the last fact, there exists a sufficiently small $\epsilon_1 > 0$ and a sufficiently large n such that $\mathcal{T}_\epsilon^k(\tilde{\boldsymbol{x}}|\tilde{\boldsymbol{y}}) \subset \mathcal{S}_o(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}})$, and another $\epsilon_2 > 0$ and a sufficiently large n (both depending on δ) such that $\mathcal{S}_o^{\delta}(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}}) \subset \mathcal{T}_\epsilon^k(\tilde{\boldsymbol{x}}|\tilde{\boldsymbol{y}})$. Then, using the same techniques as previously described, it is possible to overbound the numerator and underbound the denominator of the r.h.s. of (120) in terms of the volumes of the conditional types $\mathcal{T}_\epsilon^k(\tilde{\boldsymbol{x}}|\tilde{\boldsymbol{y}})$, and show that (120) is overbounded by a subexponential function of n.

APPENDIX A
PROOF OF LEMMA 3

We need to show the inclusion

$$\mathcal{T}_\epsilon^k(\tilde{\boldsymbol{x}}|\tilde{\boldsymbol{y}}) \subseteq \hat{\mathcal{T}}_\epsilon^k(\tilde{\boldsymbol{x}}|\tilde{\boldsymbol{y}}), \tag{A.1}$$

namely, for any $\bar{\boldsymbol{x}} \in \mathcal{T}_\epsilon^k(\tilde{\boldsymbol{x}}|\tilde{\boldsymbol{y}})$ also $\bar{\boldsymbol{x}} \in \hat{\mathcal{T}}_\epsilon^k(\tilde{\boldsymbol{x}}|\tilde{\boldsymbol{y}})$. Using the definitions of these sets we see that in order to show the above inclusion we only need to show that for every $\bar{\boldsymbol{x}} \in \mathcal{T}_\epsilon^k(\tilde{\boldsymbol{x}}|\tilde{\boldsymbol{y}})$, there exist a sequence

$\{P_m\}_{m=1}^L \in \mathcal{P}^\epsilon$ such that for any $1 \le l \le L$,

$$\left| \left\| \bar{\boldsymbol{x}}_{(l-1)n_b+1}^{ln_b} \right\|^2 - n_b P_l \right| \le \epsilon \tag{A.2}$$

where $\boldsymbol{x}_l^m \stackrel{\triangle}{=} (x_l, x_{l+1}, \ldots, x_m)$ for $m \ge l$. To this end, for each $1 \le l \le L$, P_l is chosen to be the nearest point to $\left\| \bar{\boldsymbol{x}}_{(l-1)n_b+1}^{ln_b} \right\|^2$ in the set $\mathcal{G}_{1,\epsilon}^L$, namely $P_l = \left\lfloor \left\| \bar{\boldsymbol{x}}_{(l-1)n_b+1}^{ln_b} \right\|^2 / (n_b \epsilon) \right\rfloor \cdot \epsilon$. Under this choice, obviously, (A.2) holds, and $\{P_l\}_{l=1}^L \in \mathcal{P}^\epsilon$, since

$$\left| \frac{1}{L} \sum_{l=1}^k P_l - P_x \right| = \left| \frac{1}{L} \sum_{l=1}^L \left\lfloor \frac{\left\| \bar{\boldsymbol{x}}_{(l-1)n_b+1}^{ln_b} \right\|^2}{n_b \epsilon} \right\rfloor \epsilon - P_x \right| \tag{A.3}$$

$$\le \left| \frac{1}{n} \sum_{l=1}^L \left\| \bar{\boldsymbol{x}}_{(l-1)n_b+1}^{ln_b} \right\|^2 \delta - P_x \right| \le \delta \tag{A.4}$$

where the last equality follows from the fact that $\bar{\boldsymbol{x}} \in \mathcal{T}_\epsilon^k (\tilde{\boldsymbol{x}}|\tilde{\boldsymbol{y}})$ and that $n = n_b L$.

REFERENCES

[1] 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.

[2] V. D. Goppa, "Nonprobabilistic mutual information without memory," *Probl. Cont. Information Theory*, vol. 4, pp. 97–102, 1975.

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

[4] 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.

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

[6] 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.

[7] 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.

[8] 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.

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

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

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

[12] 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.

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

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

[15] 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.

[16] W. Huleihel and N. Merhav, "Analysis of mismatched estimation errors using gradients of partition functions," *to appear in IEEE Trans. Inf. Theory*, April 2014.

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

[18] W. Huleihel and N. Merhav, "Asymptotic MMSE analysis under sparse representation modeling," *submitted to IEEE Trans. Inf. Theory*, Dec. 2013. [Online]. Available: http://arxiv.org/abs/1312.3417

[19] M. R. Gray, *Toeplitz and Circulant Matrices: A review*.   now, 2006.

[20] A. Widom, *Toeplitz Matrices*.   in Studies in Real and Complex Analysis, edited by I.I. Hirschmann, Jr., MAA Studies in Mathematics, Prentice-Hall, Englewood Cliffs, NJ, 1965.

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

[22] A. Bottcher and S. M. Grudsky, *Spectral Properties of Banded Toeplitz Matrices*.   SIAM, 2005.