

# Universal Decoding for Memoryless Gaussian Channels with a Deterministic Interference

Neri Merhav, *Senior Member, IEEE*

**Abstract**—A universal decoding procedure is proposed for memoryless Gaussian channels with deterministic interfering signals from a certain class. The universality of the proposed decoder is in the sense of being independent of the channel parameters and the unknown interfering signal, and at the same time attaining the same random coding error exponent as the optimal maximum likelihood (ML) decoder, which utilizes full knowledge of the channel parameters and the interfering signal. The proposed decoding rule can be regarded as a continuous-alphabet version of the universal maximum mutual information (MMI) decoder.

**Index Terms**—Universal decoding, maximum mutual information, random coding exponent, Gaussian channels.

## I. INTRODUCTION

UNIVERSAL DECODING for unknown finite-alphabet channels has been widely studied in the literature. For discrete memoryless channels (DMC's) Goppa [1] proposed a universal decoder, referred to as the *maximum mutual information* (MMI) decoder, which selects an input message that maximizes the empirical mutual information with the given output vector. Goppa has shown that if the receiver employs the MMI decoder, which is independent of the unknown channel statistics, the channel capacity is achievable. Csiszár and Körner [2] have sharpened this result and proved the existence of a deterministic universal fixed composition block code, which when decoded by the MMI decoder, yields the random coding error exponent for the given channel. Ziv [3] has investigated universal decoding for finite-alphabet, finite-state channels under a random coding regime and proposed a universal decoder based on the Lempel-Ziv algorithm [4] that attains the same random coding exponent as the optimal maximum likelihood (ML) decoder, which in turn, assumes full knowledge of the channel parameters. In the special case of a DMC and a fixed composition random code this decoder can be replaced by one that minimizes the empirical conditional entropy of the channel input given its output.

In this paper, we derive a parallel result for memoryless Gaussian channels with an unknown deterministic interfering signal from a fairly wide class. This interfering signal may represent a transmission from a competing source, a jammer,

or another noise source which cannot be modeled as a stationary Gaussian process. The empirical conditional entropy of the channel input given the channel output is induced by an auxiliary "backward channel" whose parameters are estimated from the given channel output vector and each one of the input vectors (codewords). The codewords are assumed to be chosen randomly near the surface of a Euclidean sphere whose radius corresponds to the power limitation. Similarly as in [3], it is shown that the proposed universal decoding rule attains the same error exponent as that of the optimal ML decoder which is fully informed of the channel parameters and the interfering signal. The main contribution of this work is in deriving an analogue to the MMI decoding principle for the continuous alphabet case. It also might serve as a step towards a derivation of a universal decoder for a Gaussian dispersive channel with unknown intersymbol interference (ISI) coefficients, which in turn has an important application in channel equalization (see, e.g., [5] and references therein). We present a conjecture as for the extension of the universal MMI decoder in the finite ISI case.

## II. STATEMENT OF MAIN RESULT

Consider a discrete-time, Gaussian memoryless channel characterized by

$$y_t = ax_t + z_t + w_t, \quad (1)$$

where  $x_t$  is the desired channel input,  $a \neq 0$  is an unknown fading parameter,  $\{w_t\}$  is zero-mean Gaussian white noise with an unknown variance  $\sigma^2 > 0$ ,  $z_t$  is an unknown deterministic interference, and  $y_t$  is the channel output. It is assumed that the noise  $\{w_t\}$  is statistically independent of the input  $\{x_t\}$  and that the interfering signal  $z_t$  can be represented by a series of given orthonormal bounded functions (e.g., sine and cosine functions) with an absolutely summable coefficient sequence, namely,

$$z_t = \sum_{i=1}^{\infty} b_i \phi_{i,t}, \quad t = 1, 2, \dots, \quad (2)$$

where  $\sum_{i=1}^{\infty} |b_i| < \infty$  and  $|\phi_{i,t}| \leq L$  for all  $i$  and  $t$ ,  $0 < L < \infty$ . The coefficients  $\{b_i\}_{i \geq 1}$  are assumed unknown.

Consider next, a codebook  $C = \{\mathbf{x}^1, \mathbf{x}^2, \dots, \mathbf{x}^M\}$  of  $M = 2^{nR}$  equiprobable messages  $\mathbf{x}^i = (x_1^i, x_2^i, \dots, x_n^i) \in \mathbb{R}^n$ ,  $i = 1, 2, \dots, M$  where  $R$  is the coding rate in bits per channel use. Clearly, if the parameter  $a$  and the interference signal  $z_t$  were known, the best decoder would have been the ML decoder, which in the Gaussian case considered here,

Manuscript received September 9, 1991; revised November 10, 1992. This work was presented in part at the IEEE International Symposium on Information Theory, San Antonio, TX, January 17–22, 1993. This work was done while the author was visiting AT&T Bell Laboratories, 600 Mountain Avenue, Murray Hill, NJ 07974.

The author is with the Department of Electrical Engineering, Technion—Israel Institute of Technology, Haifa 32000, Israel.  
IEEE Log Number 9208325.

selects the message  $\mathbf{x}^i$  that minimizes  $\sum_{t=1}^n (y_t - z_t - ax_t^i)^2$ . Similarly as in [3], the probability of error associated with the ML decoder will be denoted by  $P_{e,0}(C, R, n)$ .

Since the design of a codebook  $C$  that minimizes  $P_{e,0}(C, R, n)$  under an input power constraint is prohibitively complex for large  $n$ , we shall adopt the random coding approach, where each codeword is randomly chosen with respect to some probability density function (pdf)  $q(\mathbf{x})$ , and independently of all other codewords. For a given power constraint, a reasonable choice of  $q(\cdot)$  would be a Gaussian pdf restricted to an  $n$ -dimensional Euclidean sphere whose radius is about  $n \cdot \sigma_x^2$  (see, also [6, p. 326, (7.3.13)–(7.3.14)]). Specifically, let

$$q(\mathbf{x}) = \mu^{-1} \psi_{\Delta}(\mathbf{x}) \prod_{t=1}^n \exp\left\{-\frac{x_t^2}{2\sigma_x^2}\right\}, \quad (3)$$

where  $\mathbf{x} = (x_1, \dots, x_n)$ ,  $\psi_{\Delta}(\mathbf{x})$ , is the indicator function of the set  $D_{\Delta} \triangleq \{\mathbf{x} : |n^{-1} \sum_{t=1}^n x_t^2 - \sigma_x^2| \leq \Delta \sigma_x^2\}$ ,  $\Delta \ll 1$ , and  $\mu$  normalizes the measure (3) such that it would integrate to unity. It is well known [6, ch. 7] that  $\bar{P}_{e,0}(R, n) \triangleq E\{P_{e,0}(C, R, n)\}$ , where the expectation is taken over ensemble of randomly selected codebooks under  $q$ , decays exponentially for every rate  $R$  that is less than the channel capacity  $C = 0.5 \log(1 + a^2 \sigma_x^2 / \sigma^2)$ . The exponential rate of the error probability  $E(R) \triangleq -\lim_{n \rightarrow \infty} n^{-1} \log \bar{P}_{e,0}(R, n)$  is called the *random coding error exponent*. The previously defined input pdf  $q(\cdot)$  is known [6] to attain an exponent higher than that of the Gaussian pdf with the same variance, intuitively, because of the fact that it does not allow low energy codewords.

Since the fading parameter  $a$  and the interfering signal  $\{z_t\}$  are unknown, the ML decoder is obviously inapplicable. We next demonstrate a decoding procedure which is universal in the sense of being independent of  $a$  and  $\{z_t\}$ , and at the same time attaining  $E(R)$ . In other words, let  $P_{e,u}(C, R, n)$  denote the error probability associated with the universal rule for a given codebook  $C$ , and let  $\bar{P}_{e,u}(R, n) = E\{P_{e,u}(C, R, n)\}$ . Then,  $\bar{P}_{e,u}(R, n)$  decays exponentially at the same rate  $E(R)$  as that associated with the ML decoder. This is analogous to an earlier result by Ziv [1] for finite-alphabet, finite-state channels.

We now turn to present the proposed decoding rule. To this end, define an *auxiliary backward channel* of order  $k$  by the conditional pdf

$$\begin{aligned} V(\mathbf{x}|\mathbf{y}, \theta, k) &= \prod_{t=1}^n V(x_t|y_t, \theta, k) \\ &\triangleq \prod_{t=1}^n (2\pi\sigma_0^2)^{-1/2} \exp\left\{-\frac{1}{2\sigma_0^2} \left(x_t - \alpha y_t - \sum_{i=1}^k \beta_i \phi_{i,t}\right)^2\right\}, \end{aligned} \quad (4)$$

where  $\mathbf{y} = (y_1, \dots, y_n)$  and  $\theta \triangleq (\sigma_0^2, \alpha, \beta_1, \beta_2, \dots, \beta_k)$  is the parameter vector of the  $k$ th-order backward channel. Note that the definition of the auxiliary backward channel is completely *detached* from the underlying probabilistic model in the sense that it does not agree with the conditional pdf of  $\mathbf{x}$  given  $\mathbf{y}$  that is induced by the underlying joint pdf of

$\mathbf{x}$  and  $\mathbf{y}$ . In particular, it allows vectors  $\mathbf{x}$  outside  $D_{\Delta}$ . Let  $\{k_n\}_{n \geq 1}$  be any monotonically nondecreasing integer-valued sequence satisfying  $k_n \rightarrow \infty$  and  $k_n/n^{1/3} \rightarrow 0$  as  $n \rightarrow \infty$ . Our decoding rule will select a message  $\mathbf{x}^i$  that maximizes the function

$$u(\mathbf{x}^i, \mathbf{y}) \triangleq \frac{\max_{\theta} V(\mathbf{x}^i|\mathbf{y}, \theta, k_n)}{q(\mathbf{x}^i)} \quad (5)$$

among all  $M$  codebook messages.

**Theorem 1:** Assume that  $\{z_t\}$  can be expanded to a series of bounded orthonormal functions with an absolutely summable coefficient sequence  $\{b_i\}_{i \geq 1}$ . Let the codewords of  $C$  be chosen randomly with respect to the pdf  $q(\cdot)$  defined as in (3) and independently of each other. Then,

$$\limsup_{n \rightarrow \infty} \frac{1}{n} [\log \bar{P}_{e,u}(R, n) - \log \bar{P}_{e,0}(R, n)] \leq \lambda(\Delta), \quad (6)$$

where  $\Delta$  is as in (3) and  $\lambda(\Delta) \rightarrow 0$  as  $\Delta \rightarrow 0$ .

The proof appears in Section III.

The intuitive interpretation of (5) is that  $n^{-1} \log u(\mathbf{x}, \mathbf{y}) = n^{-1} \max_{\theta} \log V(\mathbf{x}|\mathbf{y}, \theta, k)/q(\mathbf{x})$  is an empirical version of the per-letter mutual information between  $\mathbf{x}$  and  $\mathbf{y}$ . Thus, we select the input  $\mathbf{x}^i$  that seems empirically “most dependent” upon the given output vector  $\mathbf{y}$ , which corresponds to the MMI principle [1], [2]. Note that on the support of  $q(\cdot)$ , the term  $n^{-1} \log q(\mathbf{x}^i)$  is nearly a constant independent of  $i$ . Thus, this decoding rule is essentially equivalent to one that maximizes  $\max_{\theta} V(\mathbf{x}|\mathbf{y}, \theta, k)$ , namely, maximum *a posteriori* (MAP) decoding. This is the continuous-alphabet counterpart of the minimum conditional entropy decoder of [3] for uniform input pdf's in the memoryless case.

It turns out that the extension of Theorem 1 to non-memoryless channels is not trivial. Consider, for example, a Gaussian channel with a linear intersymbol interference (ISI), characterized by  $y_t = \sum_{i=0}^k h_i x_{t-i} + w_t$ , where  $\{h_i\}_{i=0}^k$  is the channel impulse response and  $w_t$  is a Gaussian white noise. The difficulty appears to be in an appropriate definition of the auxiliary backward channel  $V(\cdot|\cdot)$ . A natural straightforward guess, in view of (4), could have been

$$\begin{aligned} V(\mathbf{x}|\mathbf{y}, \theta, k) &= (2\pi\sigma_0^2)^{-n/2} \prod_{t=1}^n \exp\left\{-\frac{1}{2\sigma_0^2} \left(x_t - \alpha y_t - \sum_{i=1}^k \beta_i x_{t-1}\right)^2\right\}. \end{aligned}$$

However, we were unable to prove that this backward channel, when plugged in (5), results in asymptotically optimal decoding. The problem seems to lie in the fact that the above conditional pdf depends on a  $2(k+1)$ -dimensional vector of sufficient statistics associated with  $\mathbf{x}$ , while there are only  $(k+2)$  degrees of freedom to adjust their conditional expectations. (For details, see Section III, Lemma 3). We conjecture that an appropriate definition of the backward channel in this case will be

$$V(\mathbf{x}|\mathbf{y}, \theta, k) = G_n(\theta, k, \mathbf{y}) \prod_{t=1}^n \exp\left\{-\frac{1}{2\sigma_0^2} \left(x_t - \sum_{i=1}^k \alpha_i x_{t-i}\right)^2 + \sum_{i=0}^k \beta_i x_{t-i} y_t\right\},$$

where  $\theta = (\sigma_0^2, \alpha, \dots, \alpha_k, \beta_0, \dots, \beta_k)$  and  $G_n(\theta, k, \mathbf{y})$  is a normalization factor chosen such that the pdf will integrate to unity.

### III. PROOF OF THEOREM 1

Let  $\mathbf{x}$  and  $\mathbf{y}$  be arbitrary fixed vectors and define

$$S_o(\mathbf{x}, \mathbf{y}) = \{\mathbf{x}' : W(\mathbf{y}|\mathbf{x}') > W(\mathbf{y}|\mathbf{x})\}, \quad (7)$$

where  $W(\mathbf{y}|\mathbf{x})$  is the conditional pdf associated with the channel, i.e.,

$$W(\mathbf{y}|\mathbf{x}) = (2\pi\sigma^2)^{-n/2} \prod_{t=1}^n \exp\left\{-\frac{1}{2\sigma^2}(y_t - z_t - ax_t)^2\right\}. \quad (8)$$

Similarly, let

$$S_u(\mathbf{x}, \mathbf{y}) = \{\mathbf{x}' : u(\mathbf{x}', \mathbf{y}) > u(\mathbf{x}, \mathbf{y})\}. \quad (9)$$

The average error probabilities associated with the ML decoder and the proposed decoder are given by

$$\bar{P}_{e,o}(R, n) = 1 - E\left\{\left[1 - \int_{S_o(\mathbf{x}, \mathbf{y})} q(\mathbf{x}') d\mathbf{x}'\right]^{2^{nR}-1}\right\} \quad (10)$$

and

$$\bar{P}_{e,u}(R, n) = 1 - E\left\{\left[1 - \int_{S_u(\mathbf{x}, \mathbf{y})} q(\mathbf{x}') d\mathbf{x}'\right]^{2^{nR}-1}\right\}, \quad (11)$$

respectively, where expectations are taken with respect to the joint pdf  $p(\mathbf{x}, \mathbf{y}) = q(\mathbf{x})W(\mathbf{y}|\mathbf{x})$ . Equation (10) (and similarly (11)) follows from a simple consideration explained in [3], which for the sake of completeness, will be repeated here. Fix a transmitted codeword  $\mathbf{x}$  and a received vector  $\mathbf{y}$ . Since other codewords  $\mathbf{x}'$  are selected randomly, the probability that  $\mathbf{x}'$  has a score  $W(\mathbf{y}|\mathbf{x}')$  smaller than  $W(\mathbf{y}|\mathbf{x})$  is  $(1 - \int_{S_o(\mathbf{x}, \mathbf{y})} q(\mathbf{x}') d\mathbf{x}')$ . Thus, in order for the decoded message to be correct, all  $(2^{nR} - 1)$  remaining codewords must provide a score lower than  $W(\mathbf{y}|\mathbf{x})$ . Since the codewords are chosen independently, this happens with probability  $[1 - \int_{S_o(\mathbf{x}, \mathbf{y})} q(\mathbf{x}') d\mathbf{x}]^{2^{nR}-1}$ . Finally, the overall correct decision probability is the expectation over  $\mathbf{x}$  and  $\mathbf{y}$  of the latter expression.

For  $\delta > 0$  we define the set

$$S_o^\delta(\mathbf{x}, \mathbf{y}) = \left\{\mathbf{x}' : \frac{1}{n} \log W(\mathbf{y}|\mathbf{x}') > \frac{1}{n} \log W(\mathbf{y}|\mathbf{x}) - \delta\right\} \quad (12)$$

and correspondingly,

$$\bar{P}_{e,o}^\delta(R, n) = 1 - E\left\{\left[1 - \int_{S_o^\delta(\mathbf{x}, \mathbf{y})} q(\mathbf{x}') d\mathbf{x}'\right]^{2^{nR}-1}\right\}, \quad (13)$$

In what follows, we shall compare the exponential behavior of  $\bar{P}_{e,u}(R, n)$  to that of  $\bar{P}_{e,o}^\delta(R, n)$  for a small  $\delta > 0$ . In the final step of the proof, this will be justified by showing that

$$\limsup_{n \rightarrow \infty} \frac{1}{n} [\log \bar{P}_{e,o}^\delta(R, n) - \log \bar{P}_{e,o}(R, n)] \leq \delta', \quad (14)$$

where  $\delta' \rightarrow 0$  as both  $\delta \rightarrow 0$  and  $\Delta \rightarrow 0$ . The use of  $\bar{P}_{e,o}^\delta(R, n)$  as an intermediate reference is done for technical reasons which will become clear in the sequel.

The following lemma is a modified version of [3, Corollary 1] for the continuous alphabet case considered here.

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

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log \int_{H_n^c} q(\mathbf{x})W(\mathbf{y}|\mathbf{x}) d\mathbf{x}d\mathbf{y} < -E(R), \quad (15)$$

where the superscript  $c$  denotes the complementary set. Then, for all large  $n$ ,

$$\bar{P}_{e,u}(R, n) \leq 2\bar{P}_{e,o}^\delta(R, n) \left[ \sup_{\mathbf{x}, \mathbf{y} \in H_n} \frac{\int_{S_u(\mathbf{x}, \mathbf{y})} q(\mathbf{x}') d\mathbf{x}'}{\int_{S_o^\delta(\mathbf{x}, \mathbf{y})} q(\mathbf{x}') d\mathbf{x}'} + 1.5 \right]. \quad (16)$$

The proof of Lemma 1 appears in the Appendix.

Lemma 1 will be useful if we can define such a sequence of sets  $H_n$  over which the ratio

$$\frac{\int_{S_u(\mathbf{x}, \mathbf{y})} q(\mathbf{x}') d\mathbf{x}'}{\int_{S_o^\delta(\mathbf{x}, \mathbf{y})} q(\mathbf{x}') d\mathbf{x}'} \quad (17)$$

is uniformly overbounded by a subexponential function of  $n$ , i.e.,  $e^{n\varepsilon_n}$ , where  $\varepsilon_n \rightarrow 0$  as  $n \rightarrow \infty$  uniformly for all  $(\mathbf{x}, \mathbf{y}) \in H_n$ . Once this is accomplished, the proof of the theorem will be complete.

The main difference between Lemma 1 and Corollary 1 of [3] is that in the latter the supremum over  $(\mathbf{x}, \mathbf{y})$  is not constrained. Observe, however, that in contrast to the finite alphabet case, here since the measurements are unbounded, then an unconstrained supremum in (16) might be arbitrarily large. Thus, the role of  $H_n$  is in dividing the space of pairs  $(\mathbf{x}, \mathbf{y})$  into two parts, where in one the supremum of (16) grows subexponentially and the other possesses a probability smaller than the desired exponential function  $e^{-nE(R)}$  (15) and hence negligible.

The set  $H_n$  will be defined as follows. For a given pair  $(\mathbf{x}, \mathbf{y})$  let

$$\hat{\theta} = (\hat{\sigma}_0^2, \hat{\alpha}, \hat{\beta}_1, \dots, \hat{\beta}_k) \triangleq \arg \max_{\theta} V(\mathbf{x}|\mathbf{y}, \theta, k). \quad (18)$$

Fix a positive constant  $B$  and define the set

$$H_n = H_n(B) \triangleq \left\{(\mathbf{x}, \mathbf{y}) : \left| \frac{1}{n} \sum_{t=1}^n x_t^2 - \sigma_x^2 \right| \leq \Delta \sigma_x^2, \frac{1}{n} \sum_{t=1}^n y_t^2 \leq B, \hat{\sigma}_0^2 \geq \frac{1}{B}, |\hat{\alpha}| \leq B, |\hat{\beta}_i| \leq B, i = 1, \dots, k\right\}. \quad (19)$$

The motivation behind this choice of  $H_n(B)$  is that it guarantees uniform continuity of  $n^{-1} \log V(\mathbf{x}|\mathbf{y}, \theta, k)$  with respect to

small perturbations of the sufficient statistics as will be seen in the sequel. We now have the following lemma which is proved in the Appendix.

**Lemma 2:** There exists a sufficiently large  $B$  such that  $\{H_n(B)\}_{n \geq 1}$  satisfies (15).

To overbound (17) within  $H_n(B)$ , we derive an upper bound on the numerator of (17) and a lower bound on its denominator, and show that these are exponentially equivalent. To this end, we first need to define a conditional type and establish some of its properties. For a given pair of vectors  $(\mathbf{x}, \mathbf{y})$  and  $\epsilon > 0$ , define the  $k$ th order conditional  $\epsilon$ -type of  $\mathbf{x}$  given  $\mathbf{y}$  as

$$T_\epsilon^k(\mathbf{x}|\mathbf{y}) = \left\{ \tilde{\mathbf{x}} \in \mathbf{R}^n : \left| \frac{1}{n} \sum_{t=1}^n \tilde{x}_t^2 - \frac{1}{n} \sum_{t=1}^n x_t^2 \right| \leq \epsilon, \tilde{\mathbf{x}} \in D_\Delta, \right. \\ \left. \left| \frac{1}{n} \sum_{t=1}^n \tilde{x}_t y_t - \frac{1}{n} \sum_{t=1}^n x_t y_t \right| \leq \epsilon, \right. \\ \left. \left| \frac{1}{n} \sum_{t=1}^n \tilde{x}_t \phi_{i,t} - \frac{1}{n} \sum_{t=1}^n x_t \phi_{i,t} \right| \leq \epsilon, i = 1, \dots, k \right\}, \quad (20)$$

where  $\tilde{\mathbf{x}} = (\tilde{x}_1, \dots, \tilde{x}_n)$ . The set  $T_\epsilon^k(\mathbf{x}|\mathbf{y})$  is regarded as a conditional type of  $\mathbf{x}$  given  $\mathbf{y}$  as it contains all vectors  $\tilde{\mathbf{x}} \in D_\Delta$  which, within  $\epsilon$ , have the same sufficient statistics as  $\mathbf{x}$ . This means that for every conditional type  $T_\epsilon^k(\mathbf{x}|\mathbf{y})$ ,  $\mathbf{x}, \mathbf{y} \in H_n(B)$ , and for any two vectors  $\mathbf{u} = (u_1, \dots, u_n)$  and  $\mathbf{v} = (v_1, \dots, v_n)$  in  $T_\epsilon^k(\mathbf{x}|\mathbf{y})$ , the conditional pdf's  $W(\mathbf{y}|\mathbf{u})$  and  $W(\mathbf{y}|\mathbf{v})$  are exponentially equivalent, provided that  $k$  is sufficiently large. Specifically, let  $\mathbf{u}, \mathbf{v} \in T_\epsilon^k(\mathbf{x}|\mathbf{y})$ . Then,

$$\left| \frac{1}{n} \log W(\mathbf{y}|\mathbf{u}) - \frac{1}{n} \log W(\mathbf{y}|\mathbf{v}) \right| \\ = \frac{1}{2\sigma^2} \left| \frac{1}{n} \sum_{t=1}^n (y_t - au_t - \sum_{i=1}^{\infty} b_i \phi_{i,t})^2 \right. \\ \left. - \frac{1}{n} \sum_{t=1}^n (y_t - av_t - \sum_{i=1}^{\infty} b_i \phi_{i,t})^2 \right| \\ \leq \frac{a^2}{2\sigma^2} \cdot \left| \frac{1}{n} \sum_{t=1}^n u_t^2 - \frac{1}{n} \sum_{t=1}^n v_t^2 \right| \\ + \frac{|a|}{\sigma^2} \cdot \left| \frac{1}{n} \sum_{t=1}^n u_t y_t - \frac{1}{n} \sum_{t=1}^n v_t y_t \right| \\ + \sum_{i=1}^{\infty} \frac{|b_i|}{\sigma^2} \cdot \left| \frac{1}{n} \sum_{t=1}^n u_t \phi_{i,t} - \frac{1}{n} \sum_{t=1}^n v_t \phi_{i,t} \right| \\ \leq \frac{a^2}{2\sigma^2} \cdot 2\epsilon + \frac{|a|}{\sigma^2} \cdot 2\epsilon + \sum_{i=1}^k \frac{|b_i|}{\sigma^2} \cdot 2\epsilon + \sum_{i=k+1}^{\infty} \frac{|b_i|}{\sigma^2} \cdot 2BL \\ \leq C_1 \cdot \epsilon + C_2 \cdot \zeta_k, \quad (21)$$

where  $C_1$  and  $C_2$  are some positive constants and  $\zeta_k \triangleq \sum_{i=k+1}^{\infty} |b_i|$  tends to zero as  $k \rightarrow \infty$  by the absolute summability of the sequence  $\{b_i\}_{i \geq 1}$ . Clearly, the right-most side of (21) can be made arbitrarily small by choosing  $\epsilon$  sufficiently small and  $k$  sufficiently large. Similarly,  $q(\mathbf{u})$  and  $q(\mathbf{v})$  are also exponentially equivalent, provided that they both belong

to the support  $S$  of  $q(\cdot)$ , namely,

$$\left| \frac{1}{n} \log q(\mathbf{u}) - \frac{1}{n} \log q(\mathbf{v}) \right| \leq C_3 \cdot \epsilon, \quad (22)$$

for some constant  $C_3$ .

The following lemma provides an upper and a lower bound on the volume of  $T_\epsilon^k(\mathbf{x}|\mathbf{y})$ , where the volume of a set  $A \subset \mathbf{R}^n$  is defined as  $\text{Vol}\{A\} \triangleq \int_A d\mathbf{x}$ .

**Lemma 3:** Let  $(\mathbf{x}, \mathbf{y} \in H_n(B))$  for some  $B > 0$ . Then, for every sufficiently small  $\epsilon > 0$  and every positive integer  $k$ , the volume of  $T_\epsilon^k(\mathbf{x}|\mathbf{y})$  is bounded as follows.

a)

$$\text{Vol}\{T_\epsilon^k(\mathbf{x}|\mathbf{y})\} \leq \frac{\exp[n\epsilon B^2(B+k+1)]}{\max_\theta V(\mathbf{x}|\mathbf{y}, \theta, k)}$$

b)

$$\text{Vol}\{T_\epsilon^k(\mathbf{x}|\mathbf{y})\} \geq \left[ \gamma_n(\Delta) - \frac{B(kL^2 + 25B)}{n\epsilon^2} \right] \cdot \frac{\exp[-n\epsilon B^2(B+k+1)]}{\max_\theta V(\mathbf{x}|\mathbf{y}, \theta, k)}$$

where  $L$  is the uniform bound on  $\phi_{i,t}$ , and  $\gamma_n(\Delta)$  is a sequence depending only on  $n$  and  $\Delta$  with the property  $\liminf_{n \rightarrow \infty} \gamma_n(\Delta) \geq 1/4$  for every  $\Delta > 0$ .

The proof of Lemma 3 appears in the Appendix.

We are now ready to derive a lower bound on the denominator of (17). Since  $\mathbf{x} \in S_\delta^o(\mathbf{x}, \mathbf{y})$ , then in view of (21), there exists a sufficiently small  $\epsilon > 0$  and a sufficiently large  $k$  (both depending on  $\delta$ ) such that  $T_\epsilon^k(\mathbf{x}|\mathbf{y}) \subset S_\delta^o(\mathbf{x}|\mathbf{y})$ . Thus, from (22) and part b) of Lemma 3 we get

$$\int_{S_\delta^o(\mathbf{x}, \mathbf{y})} q(\mathbf{x}') d\mathbf{x}' \geq \int_{T_\epsilon^k(\mathbf{x}|\mathbf{y})} q(\mathbf{x}') d\mathbf{x}' \\ \geq \text{Vol}\{T_\epsilon^k(\mathbf{x}|\mathbf{y})\} \cdot \inf_{\mathbf{x}' \in T_\epsilon^k(\mathbf{x}|\mathbf{y})} q(\mathbf{x}') \\ \geq \left[ \gamma_n(\Delta) - \frac{B(kL^2 + 25B)}{n\epsilon^2} \right] \cdot \frac{\exp[-n\epsilon B^2(B+k+1)]}{\max_\theta V(\mathbf{x}|\mathbf{y}, \theta, k)} \\ \cdot e^{-C_3 \cdot \epsilon} q(\mathbf{x}), \quad (23)$$

where we have used the fact that  $(\mathbf{x}, \mathbf{y}) \in H_n(B)$ .

We next overbound the numerator of (17). The underlying idea is to partition the set  $S_u(\mathbf{x}, \mathbf{y})$  into a subexponential number of conditional types, where for each conditional type,  $\int_{T_\epsilon^k(\mathbf{x}|\mathbf{y})} q(\mathbf{x}') d\mathbf{x}'$  is overestimated using part (b) of Lemma 3, in a manner similar to (23). However, this cannot be done directly, since not every  $\mathbf{x}' \in S_u(\mathbf{x}, \mathbf{y})$  is such that  $(\mathbf{x}', \mathbf{y}) \in H_n(B)$  and hence Lemma 3 cannot be applied to  $T_\epsilon^k(\mathbf{x}'|\mathbf{y})$ . To alleviate this difficulty, we first divide  $S_u(\mathbf{x}, \mathbf{y})$  into two subsets,  $S_u(\mathbf{x}, \mathbf{y}) \cap H_n(B_0|\mathbf{y})$  and  $S_u(\mathbf{x}, \mathbf{y}) \cap H_y^c(B_0|\mathbf{y})$ , where  $H_n(B_0|\mathbf{y}) \triangleq \{\mathbf{x}' : (\mathbf{x}', \mathbf{y}) \in H_n(B_0)\}$ ,  $B_0 \geq B$ , being a constant to be chosen later. Now, in the first subset we can apply Lemma 3 while the second has a very low probability provided that  $B_0$  is sufficiently large. Specifically, let  $B$  be so large that (15) holds and fix  $(\mathbf{x}, \mathbf{y}) \in H_n(B)$ . Similarly as

in Lemma 2, we can now choose  $B_0$  so large that for every  $\mathbf{y}' \in H_n(B|\mathbf{x}) \triangleq \{\mathbf{y}' : (\mathbf{x}, \mathbf{y}' \in H_n(B))\}$ , we have

$$\int_{H_n^c(B_0|\mathbf{y})} q(\mathbf{x}') d\mathbf{x}' \leq e^{-nQ(B_0)}, \quad (24)$$

for all large  $n$ , where  $Q(B_0) > 0$  can be made arbitrarily large by choosing  $B_0$  sufficiently large. Thus, we have

$$\begin{aligned} \int_{S_u(\mathbf{x}, \mathbf{y})} q(\mathbf{x}') d\mathbf{x}' &= \int_{S_u(\mathbf{x}, \mathbf{y}) \cap H_n(B_0|\mathbf{y})} q(\mathbf{x}') d\mathbf{x}' \\ &\quad + \int_{S_u(\mathbf{x}, \mathbf{y}) \cap H_n^c(B_0|\mathbf{y})} q(\mathbf{x}') d\mathbf{x}' \\ &\leq \int_{S_u(\mathbf{x}, \mathbf{y}) \cap H_n(B_0|\mathbf{y})} q(\mathbf{x}') d\mathbf{x}' + e^{-nQ(B_0)}. \end{aligned} \quad (25)$$

Now, as for the first term on the right-most side of (25), every  $\mathbf{x}'$  in the integration domain is such that  $(\mathbf{x}', \mathbf{y} \in H_n(B_0))$ . Hence, we can subdivide this domain into  $k$ th order conditional  $\varepsilon$ -types, whose volumes can be overestimated by Lemma 3 with  $B_0$  replacing  $B$ . Before doing this, we first have to overbound the number of conditional types needed to cover  $S_u(\mathbf{x}, \mathbf{y}) \cap H_n(B_0|\mathbf{y}) \subset H_n(B_0|\mathbf{y})$ . We note that in this subset  $n^{-1} \sum_{t=1}^n (x'_t)^2 \leq B_0$  and hence  $|n^{-1} \sum_{t=1}^n x'_t \phi_{i,t}| \leq B_0$ . Since  $\mathbf{y} \in H_n(B|\mathbf{x})$ , then  $n^{-1} \sum_{t=1}^n y_t^2 \leq B \leq B_0$ , and therefore  $|n^{-1} \sum_{t=1}^n x'_t y_t| \leq B_0$ . Thus, the number of conditional types  $\{T_\varepsilon^k(\mathbf{x}'|\mathbf{y})\}$  needed to cover  $H_n(B_0|\mathbf{y})$  is not larger than  $(2B_0/\varepsilon)^{k+2}$ . In view of these facts, the first term on the right-most side of (25) is bounded above as follows:

$$\begin{aligned} &\int_{S_u(\mathbf{x}, \mathbf{y}) \cap H_n(B_0|\mathbf{y})} q(\mathbf{x}') d\mathbf{x}' \\ &\leq \sum_{T_\varepsilon^k(\mathbf{x}'|\mathbf{y}) \subset S_u(\mathbf{x}, \mathbf{y}) \cap H_n(B_0|\mathbf{y})} \int_{T_\varepsilon^k(\mathbf{x}'|\mathbf{y})} q(\tilde{\mathbf{x}}) d\tilde{\mathbf{x}} \\ &\leq \left(\frac{2B_0}{\varepsilon}\right)^{k+2} \\ &\quad \cdot \sup_{\mathbf{x}' \in S_u(\mathbf{x}, \mathbf{y})} \left\{ \text{Vol}\{T_\varepsilon^k(\mathbf{x}'|\mathbf{y})\} \cdot \sup_{\tilde{\mathbf{x}} \in T_\varepsilon^k(\mathbf{x}'|\mathbf{y})} q(\tilde{\mathbf{x}}) \right\} \\ &\leq e^{C_3 \varepsilon n} \left(\frac{2B_0}{\varepsilon}\right)^{k+2} \exp[n\varepsilon B_0^2(B_0 + k + 1)] \\ &\quad \cdot \sup_{\mathbf{x}' \in S_u(\mathbf{x}, \mathbf{y})} \frac{q(\mathbf{x}')}{\max_\theta V(\mathbf{x}'|\mathbf{y}, \theta, k)} \\ &\leq e^{C_3 \varepsilon n} \left(\frac{2B_0}{\varepsilon}\right)^{k+2} \exp[n\varepsilon B_0^2(B_0 + k + 1)] \\ &\quad \cdot \frac{q(\mathbf{x})}{\max_\theta V(\mathbf{x}|\mathbf{y}, \theta, k)}, \end{aligned} \quad (26)$$

where the last step follows from the definition of  $S_u(\mathbf{x}, \mathbf{y})$ . Combining (23), (25) and (26), we get for all sufficiently large  $n$ ,

$$\sup_{(\mathbf{x}, \mathbf{y}) \in H_n(B)} \frac{\int_{S_u(\mathbf{x}, \mathbf{y})} q(\mathbf{x}') d\mathbf{x}'}{\int_{S_0^c(\mathbf{x}, \mathbf{y})} q(\mathbf{x}') d\mathbf{x}'}$$

$$\begin{aligned} &\leq \left[ \gamma_n(\Delta) - \frac{B(kL^2 + 25B)}{n\varepsilon^2} \right]^{-1} \\ &\quad \exp\left\{ 2n\varepsilon [C_3 + B_0^2(B_0 + k + 1)] \right\} \cdot \left(\frac{2B_0}{\varepsilon}\right)^{k+2} \\ &\quad \cdot \left[ 1 + e^{-nQ(B_0)} \sup_{(\mathbf{x}, \mathbf{y}) \in H_n(B)} \frac{\max_\theta V(\mathbf{x}|\mathbf{y}, \theta, k)}{q(\mathbf{x})} \right]. \end{aligned} \quad (27)$$

The supremum on the right-hand side of (27) is in turn bounded by

$$\begin{aligned} &\sup_{(\mathbf{x}, \mathbf{y}) \in H_n(B)} \frac{\max_\theta V(\mathbf{x}|\mathbf{y}, \theta, k)}{q(\mathbf{x})} \\ &\leq \frac{\sup_{(\mathbf{x}, \mathbf{y}) \in H_n(B)} \max_\theta V(\mathbf{x}|\mathbf{y}, \theta, k)}{\inf_{(\mathbf{x}, \mathbf{y}) \in H_n(B)} q(\mathbf{x})} \\ &\leq \frac{(2\pi e B^{-1})^{-n/2}}{\mu^{-1} e^{(1+\Delta)n/2}} \end{aligned} \quad (28)$$

Now,

$$\begin{aligned} \mu &= \int_{\mathbf{x} \in D_\Delta} d\mathbf{x} \exp\left\{ -\sum_{t=1}^n \frac{x_t^2}{2\sigma_x^2} \right\} \\ &\leq e^{-(1-\Delta)n/2} \cdot \text{Vol}\left\{ \mathbf{x} : \frac{1}{n} \sum_{t=1}^n x_t^2 \leq \sigma_x^2(1 + \Delta) \right\} \\ &\leq e^{-1(1-\Delta)n/2} \cdot [2\pi e \sigma_x^2(1 + \Delta)]^{n/2}, \end{aligned} \quad (29)$$

where the last inequality follows similarly to the derivation in (A.17)–(A.18) in the Appendix. Now choose  $B_0$  so large that

$$Q(B_0) > \frac{1}{2} [\log B + \log \sigma_x^2 + \log(1 + \Delta) + 2\Delta]. \quad (30)$$

Then, the last term in the product on the right-hand side of (27) tends to unity as  $n \rightarrow \infty$ . Thus, in order that (27) will be a subexponential function of  $n$ , we let  $\varepsilon = \varepsilon_n$  tend to zero and  $k = k_n$  tend to infinity such the following three conditions will hold simultaneously.

$$k_n \log \frac{1}{\varepsilon_n} = o(n), \quad (31a)$$

$$\lim_{n \rightarrow \infty} k_n \varepsilon_n = 0, \quad (31b)$$

and

$$\lim_{n \rightarrow \infty} \frac{n\varepsilon_n^2}{k_n} = C_4, \quad (31c)$$

where  $C_4$  is some sufficiently large constant. This happens if  $\varepsilon_n = 0(n^{-1/3})$  and hence  $k_n = 0(n^{1/3})$ . Thus, we have shown that

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \bar{P}_{\varepsilon, u}(R, n) \leq \lim_{n \rightarrow \infty} \frac{1}{n} \log \bar{P}_{\varepsilon, o}^\delta(R, n) \quad (32)$$

To complete the proof of the theorem, it remains to justify eq. (14). Note that both  $S_0(\mathbf{x}, \mathbf{y})$  and  $S_0^\delta(\mathbf{x}, \mathbf{y})$  correspond to a known channel and interference. The receiver subtracts  $z_t$  from  $y_t$  and treats the difference as an output of the channel  $W$  with no interference. Therefore, we now are dealing, in fact,

with a special case of  $W$  where  $b_i = 0$  for all  $i$ , namely, we redefine the channel as  $y_t = ax_t + w_t$ . Assume, without loss of generality, that  $a > 0$ . In this case,  $S_0(\mathbf{x}, \mathbf{y})$  and  $S_0^\delta(\mathbf{x}, \mathbf{y})$  satisfy

$$\begin{aligned} S_0(\mathbf{x}, \mathbf{y}) &= \left\{ \mathbf{x}' : \frac{1}{n} \sum_{i=1}^n x'_i y_t - \frac{1}{2n} \sum_{i=1}^n (x'_i)^2 \right. \\ &\quad \left. \geq \frac{1}{n} \sum_{i=1}^n x_t y_t - \frac{1}{2n} \sum_{i=1}^n x_t^2 \right\} \\ &\supset \left\{ \mathbf{x}' : \frac{1}{n} \sum_{i=1}^n x'_i y_t \geq \frac{1}{n} \sum_{i=1}^n x_t y_t + \Delta \sigma_x^2 \right\} \\ &\triangleq F(\mathbf{x}, \mathbf{y}, \Delta \sigma_x^2) \end{aligned} \quad (33)$$

and

$$\begin{aligned} S_0^\delta(\mathbf{x}, \mathbf{y}) &= \left\{ \mathbf{x}' : \frac{1}{n} \sum_{i=1}^n x'_i y_t - \frac{1}{2n} \sum_{i=1}^n (x'_i)^2 \right. \\ &\quad \left. \geq \frac{1}{n} \sum_{i=1}^n x_t y_t - \frac{1}{2n} \sum_{i=1}^n x_t^2 - \sigma^2 \delta \right\} \\ &\subset \left\{ \mathbf{x}' : \frac{1}{n} \sum_{i=1}^n x'_i y_t \geq \frac{1}{n} \sum_{i=1}^n x_t y_t - \Delta \sigma_x^2 - \delta \sigma^2 \right\} \\ &= F(\mathbf{x}, \mathbf{y}, -\Delta \sigma_x^2 - \delta \sigma^2) \end{aligned} \quad (34)$$

Now similarly as in Lemma 1, we can overbound the ratio  $\bar{P}_{o,e}^\delta(R, n) / \bar{P}_{e,o}(R, n)$  in terms of the worst case ratio

$$\frac{\int_{S_0^\delta(\mathbf{x}, \mathbf{y})} q(\mathbf{x}') d\mathbf{x}'}{\int_{S_0(\mathbf{x}, \mathbf{y})} q(\mathbf{x}') d\mathbf{x}'} \leq \frac{\int_{F(\mathbf{x}, \mathbf{y}, -\Delta \sigma_x^2 - \delta \sigma^2)} q(\mathbf{x}') d\mathbf{x}'}{\int_{F(\mathbf{x}, \mathbf{y}, \Delta \sigma_x^2)} q(\mathbf{x}') d\mathbf{x}'}, \quad (35)$$

over  $(\mathbf{x}, \mathbf{y}) \in H_n$ , where  $H_n$  is a set with high probability. A reasonable choice of  $H_n$  here is the set of pairs  $(\mathbf{x}, \mathbf{y})$  which satisfy

$$\left| \frac{1}{n} \sum_{i=1}^n x_t^2 - \sigma_x^2 \right| \leq \Delta \sigma_x^2, \quad (36)$$

$$\frac{1}{n} \sum_{i=1}^n y_t^2 \leq B, \quad (37)$$

and

$$\frac{1}{n} \sum_{i=1}^n x_t^2 - \frac{[\frac{1}{n} \sum_{i=1}^n x_t y_t]^2}{\frac{1}{n} \sum_{i=1}^n y_t^2} > \frac{1}{B}. \quad (38)$$

Using the same techniques as previously described, it is possible to overbound the numerator and underbound the denominator of the right-hand side of (35) in terms of the volumes of the conditional types

$$\begin{aligned} T_\varepsilon(\mathbf{x}|\mathbf{y}) &= \left\{ \tilde{\mathbf{x}} : \left| \frac{1}{n} \sum_{i=1}^n \tilde{x}_t^2 - \frac{1}{n} \sum_{i=1}^n x_t^2 \right| \leq \varepsilon, \quad \mathbf{x} \in D_\Delta, \right. \\ &\quad \left. \left| \frac{1}{n} \sum_{i=1}^n \tilde{x}_t y_t - \frac{1}{n} \sum_{i=1}^n x_t y_t \right| \leq \varepsilon \right\} \quad (\mathbf{x}, \mathbf{y}) \in H_n \end{aligned} \quad (39)$$

that are contained in  $F(\mathbf{x}, \mathbf{y}, \cdot)$ , and to establish (14).

## APPENDIX

*Proof of Lemma 1:* The proof involves merely a slight modification of the proof of [1, Corollary 1]. Let

$$P_{e,o}^\delta(\mathbf{x}, \mathbf{y}) = 1 - \left[ 1 - \int_{S_0^\delta(\mathbf{x}, \mathbf{y})} q(\mathbf{x}') d\mathbf{x}' \right]^{2^{nR}-1} \quad (A.1)$$

and

$$P_{e,u}(\mathbf{x}, \mathbf{y}) = 1 - \left[ 1 - \int_{S_u(\mathbf{x}, \mathbf{y})} q(\mathbf{x}') d\mathbf{x}' \right]^{2^{nR}-1} \quad (A.2)$$

Then, similarly to [1, (A.1)], we get

$$\begin{aligned} P_{e,u}(\mathbf{x}, \mathbf{y}) &\leq P_{e,o}^\delta(\mathbf{x}, \mathbf{y}) \cdot \frac{\int_{S_u(\mathbf{x}, \mathbf{y})} q(\mathbf{x}') d\mathbf{x}'}{\int_{S_0^\delta(\mathbf{x}, \mathbf{y})} q(\mathbf{x}') d\mathbf{x}'} \\ &\quad \cdot \left[ 1 - \frac{1}{2} (2^{nR} - 1) \int_{S_0^\delta(\mathbf{x}, \mathbf{y})} q(\mathbf{x}') d\mathbf{x}' \right]^{-1}. \end{aligned} \quad (A.3)$$

Next define the sets

$$S_1 = \left\{ (\mathbf{x}, \mathbf{y}) : \int_{S_0^\delta(\mathbf{x}, \mathbf{y})} q(\mathbf{x}') d\mathbf{x}' \leq (2^{nR} - 1)^{-1} \right\}, \quad (A.4)$$

and

$$S_2 = S_1 \cap H_n(B). \quad (A.5)$$

Now, similarly as in [1, (A.2)],

$$\begin{aligned} \bar{P}_{e,u}(R, n) &\leq 2 \sup_{(\mathbf{x}, \mathbf{y}) \in H_n(B)} \frac{\int_{S_u(\mathbf{x}, \mathbf{y})} q(\mathbf{x}') d\mathbf{x}'}{\int_{S_0^\delta(\mathbf{x}, \mathbf{y})} q(\mathbf{x}') d\mathbf{x}'} \\ &\quad \cdot \Pr\{S_2\} \cdot E\{P_{e,o}^\delta(\mathbf{x}, \mathbf{y}) | S_2\} + \Pr\{S_2^c\} \\ &\leq 2 \sup_{(\mathbf{x}, \mathbf{y}) \in H_n(B)} \frac{\int_{S_u(\mathbf{x}, \mathbf{y})} q(\mathbf{x}') d\mathbf{x}'}{\int_{S_0^\delta(\mathbf{x}, \mathbf{y})} q(\mathbf{x}') d\mathbf{x}'} \cdot \bar{P}_{e,o}^\delta(R, n) + 1 \\ &\quad - \Pr\{S_1\} + \Pr\{H_n^c(B)\}, \end{aligned} \quad (A.6)$$

where the last inequality follows from the union bound and the fact that the integral of a positive function over the entire sample space is larger than its integral over  $S_2$ . From [1, (A.4)], we have that

$$1 - \Pr\{S_1\} \leq 2\bar{P}_{e,o}^\delta(R, n) \quad (A.7)$$

From the assumption of the lemma, it follows that for all sufficiently large  $n$ , the probability that  $(\mathbf{x}, \mathbf{y})$  is outside  $H_n(B)$  will be less than  $\bar{P}_{e,o}^\delta(q, R, n)$ . This completes the proof of Lemma 1.  $\square$

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

$$\begin{aligned} \Pr\{H_n^c(B)\} &\leq \Pr\left\{\frac{1}{n}\sum_{t=1}^n y_t^2 > B\right\} + \Pr\{|\hat{\alpha}| > B\} \\ &\quad + \Pr\{\hat{\sigma}_0^2 < B^{-1}\} + \sum_{i=1}^k \Pr\{|\hat{\beta}_i| > B\}. \end{aligned} \quad (\text{A.8})$$

Thus, it should be shown that if  $B$  is chosen sufficiently large, each one of the probabilities on the right-hand side of (A.8) decays faster than  $e^{-nE(R)}$ . As for the first term, since

$$\frac{1}{n}\sum_{t=1}^n y_t^2 \leq \left[|a|\sqrt{\sigma_x^2(1+\Delta)} + \sqrt{\frac{1}{n}\sum_{t=1}^n w_t^2}\right]^2, \quad (\text{A.9})$$

then we have

$$\begin{aligned} \Pr\left\{\frac{1}{n}\sum_{t=1}^n y_t^2 > B\right\} \\ \leq \Pr\left\{\frac{1}{n}\sum_{t=1}^n w_t^2 > \left(\sqrt{B} - |a|\sqrt{\sigma_x^2(1+\Delta)}\right)^2\right\}, \end{aligned} \quad (\text{A.10})$$

which in turn 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 (see, e.g., [7, Lemma 12.9.1]). As for the remaining terms, we have the following expressions for the components of  $\hat{\theta}$ .

$$\hat{\alpha} = \frac{\frac{1}{n}\sum_{t=1}^n x_t(y_t - \sum_{i=1}^k \eta_i \phi_{i,t})}{\frac{1}{n}\sum_{t=1}^n (y_t - \sum_{i=1}^k \eta_i \phi_{i,t})^2}, \quad (\text{A.11})$$

where  $\eta_i = \sum_{t=1}^n y_t \phi_{i,t}$ ,

$$\hat{\beta}_i = \frac{1}{n}\sum_{t=1}^n x_t \phi_{i,t} - \hat{\alpha} \eta_i, \quad i = 1, \dots, k, \quad (\text{A.12})$$

and

$$\hat{\sigma}_0^2 = \frac{1}{n}\sum_{t=1}^n (x_t - \hat{\alpha} y_t - \sum_{i=1}^k \hat{\beta}_i \phi_{i,t})^2. \quad (\text{A.13})$$

Regarding the second term on the right-hand side of (A.8), we have

$$\begin{aligned} \Pr\{|\hat{\alpha}| > B\} &\leq \Pr\left\{\frac{\sigma_x^2(1+\Delta)}{\frac{1}{n}\sum_{t=1}^n (y_t - \sum_{i=1}^k \eta_i \phi_{i,t})^2} > B^2\right\} \\ &= \Pr\left\{\frac{1}{n}\sum_{t=1}^n (y_t - \sum_{i=1}^k \eta_i \phi_{i,t})^2 < \frac{\sigma_x^2(1+\Delta)}{B^2}\right\}. \end{aligned} \quad (\text{A.14})$$

Let  $B' \triangleq B^2/\sigma_x^2(1+\Delta)$ . Then, (A.14) is further bounded as follows.

$$\begin{aligned} \Pr\left\{\frac{1}{n}\sum_{t=1}^n (y_t - \sum_{i=1}^k \eta_i \phi_{i,t})^2 < \frac{1}{B'}\right\} \\ \leq \Pr\left\{\frac{1}{n}\sum_{t=1}^n (y_t - \sum_{i=1}^k \eta_i \phi_{i,t})^2 < \frac{1}{B'} \text{ and } \frac{1}{n}\sum_{t=1}^n y_t^2 \leq B\right\} \end{aligned}$$

$$+ \Pr\left\{\frac{1}{n}\sum_{t=1}^n y_t^2 > B\right\}. \quad (\text{A.15})$$

Again, the exponential decay rate of the second term on the right-hand side of (A.15) can be made arbitrarily large. As for the first term in (A.15), note that the coefficient vector  $\eta = (\eta_1, \dots, \eta_k)$  minimizes the quadratic form  $n^{-1}\sum_{t=1}^n (y_t - \sum_{i=1}^k \xi_i \phi_{i,t})^2$  over all vectors  $\xi = (\xi_1, \dots, \xi_k)$  in  $\mathbf{R}^k$ . Moreover, since the energy of  $\mathbf{y}$  is less than  $B$  within the set of interest, then the minimizing vector  $\xi = \eta$  must lie in the  $k$ -dimensional sphere  $\xi^T \xi \leq B$ , where  $T$  denotes transposition. Fix  $\delta > 0$ , define the grid  $G = \{\delta \cdot i : i = -\lceil B/\delta \rceil, \dots, -1, 0, 1, \dots, \lceil B/\delta \rceil\}$ , and let  $G^k$  denote the  $k$ th Cartesian power of  $G$ . From the uniform continuity of the above quadratic form within the set of all energy-limited vectors  $\mathbf{y}$ , one can find a sufficiently small value of  $\delta$  (depending on  $B$ ) such that there exists a vector  $\xi \in G^k$  satisfying

$$\frac{1}{n}\sum_{t=1}^n (y_t - \sum_{i=1}^k \xi_i \phi_{i,t})^2 \leq \frac{2}{B'}.$$

Thus,

$$\begin{aligned} \Pr\left\{\frac{1}{n}\sum_{t=1}^n (y_t - \sum_{i=1}^k \eta_i \phi_{i,t})^2 < \frac{1}{B'} \text{ and } \frac{1}{n}\sum_{t=1}^n y_t^2 \leq B\right\} \\ \leq \Pr\left\{\bigcup_{\xi \in G^k} \left\{\mathbf{y} : \frac{1}{n}\sum_{t=1}^n (y_t - \sum_{i=1}^k \xi_i \phi_{i,t})^2 < \frac{2}{B'}\right\}\right\} \\ \leq \sum_{\xi \in G^k} \Pr\left\{\frac{1}{n}\sum_{t=1}^n (y_t - \sum_{i=1}^k \xi_i \phi_{i,t})^2 < \frac{2}{B'}\right\}. \end{aligned} \quad (\text{A.16})$$

Thus, if we can show that each term on the right-most side of (A.16) can be made exponentially less than  $e^{-nE(R)}$ , the result for the first term in (A.15) will follow as  $G^k$  contains  $(\lceil 2B/\delta \rceil)^k$  points, which is a constant. (Even if  $k = k_n = o(n^{1/3})$ , then still this is a subexponential function of  $n$ .) To estimate each term on the right-hand side of (A.16), we first overbound the volume (Lebesgue measure) of the set

$$A_\xi = \left\{\mathbf{y} \in \mathbf{R}^n : \frac{1}{n}\sum_{t=1}^n (y_t - \sum_{i=1}^k \xi_i \phi_{i,t})^2 \leq \frac{2}{B'}\right\}.$$

Define an auxiliary pdf,

$$g(\mathbf{y}) = \left(\frac{4\pi}{B'}\right)^{-n/2} \prod_{t=1}^n \exp\left\{-\frac{B'}{4}(y_t - \sum_{i=1}^k \xi_i \phi_{i,t})^2\right\}. \quad (\text{A.17})$$

Let  $\text{Vol}\{A_\xi\} \triangleq \int_{A_\xi} d\mathbf{y}$  denote the volume of the set  $A_\xi$ . Then, we have

$$\begin{aligned} 1 &\geq \int_{A_\xi} g(\mathbf{y}) d\mathbf{y} \\ &\geq \text{Vol}\{A_\xi\} \cdot (4\pi/B')^{-n/2} \\ &\quad \inf_{\mathbf{y} \in A_\xi} \exp\left\{-\frac{B'}{4}\sum_{t=1}^n (y_t - \sum_{i=1}^k \xi_i \phi_{i,t})^2\right\} \end{aligned}$$

$$\geq \text{Vol}\{A_\xi\} \cdot \left(\frac{4\pi e}{B'}\right)^{-n/2}, \quad (\text{A.18})$$

and hence  $\text{Vol}\{A_\xi\} \leq (4\pi e/B')^{n/2}$ . Thus, each term in (A.16) is overbounded as follows:

$$\begin{aligned} \Pr\{A_\xi\} &= \int_{\mathbf{x} \in \mathbb{R}^n} d\mathbf{x} q(\mathbf{x}) \int_{\mathbf{y} \in A_\xi} d\mathbf{y} W(\mathbf{y}|\mathbf{x}) \\ &\leq \left[ \int_{\mathbf{x} \in \mathbb{R}^n} d\mathbf{x} q(\mathbf{x}) \right] \cdot \text{Vol}\{A_\xi\} \cdot (2\pi\sigma^2)^{-n/2} \\ &\leq \left(\frac{4\pi e}{B'}\right)^{n/2} \cdot (2\pi\sigma^2)^{-n/2} \\ &= \exp\left\{-n \cdot \frac{1}{2} \log\left(\frac{\sigma^2 B'}{2e}\right)\right\}, \quad (\text{A.19}) \end{aligned}$$

which can be made less than  $e^{-nE(R)}$  by choosing  $B$  sufficiently large. This completes the proof for the second term on the right-hand side of (A.8). The same technique exactly can be repeated for bounding the third term. Finally, each term  $\Pr\{|\hat{\beta}_i| > B\}$  of can be overbounded by the following consideration. From (A.12), we have

$$|\hat{\beta}_i| \leq \sqrt{\sigma_x^2(1+\Delta)} + |\hat{\alpha}| \cdot \sqrt{\frac{1}{n} \sum_{t=1}^n y_t^2}. \quad (\text{A.20})$$

Thus, by the union bound the probability that  $|\hat{\beta}_i|$  exceeds  $B$  is less than

$$\begin{aligned} \Pr\left\{|\hat{\alpha}| > \left[B - \sqrt{\sigma_x^2(1+\Delta)}\right]^{1/2}\right\} \\ + \Pr\left\{\frac{1}{n} \sum_{t=1}^n y_t^2 > B - \sqrt{\sigma_x^2(1+\Delta)}\right\}, \end{aligned}$$

which both have been shown already to decay faster than  $e^{-nE(R)}$  for a sufficiently large  $B$ . This completes the proof of Lemma 2.  $\square$

*Proof of Lemma 3:* Fix a pair  $(\mathbf{x}, \mathbf{y}) \in H_n(B)$  and let  $\rho_{xx} \triangleq n^{-1} \sum_{t=1}^n x_t^2$ ,  $\rho_{xy} \triangleq n^{-1} \sum_{t=1}^n x_t y_t$ , and  $\rho_x^i \triangleq n^{-1} \sum_{t=1}^n x_t \phi_{i,t}$ ,  $i = 1, 2, \dots, k$ . Consider the auxiliary backward channel of order  $k$  as defined in eq. (4), i.e., the backward channel

$$\tilde{x}_t = \alpha y_t + \sum_{i=1}^k \beta_i \phi_{i,t} + \nu_t, \quad (\text{A.21})$$

where  $\nu_t$  is a zero-mean Gaussian white noise with variance  $\sigma_0^2$ . Since  $(\mathbf{x}, \mathbf{y}) \in H_n(B)$  there exists a parameter vector  $\theta = (\sigma_0^2, \alpha, \beta_1, \dots, \beta_k)$  such that

$$E_\theta^k \left\{ \frac{1}{n} \sum_{t=1}^n \tilde{x}_t^2 \middle| \mathbf{y} \right\} = \rho_{xx} \quad (\text{A.22})$$

$$E_\theta^k \left\{ \frac{1}{n} \sum_{t=1}^n \tilde{x}_t y_t \middle| \mathbf{y} \right\} = \rho_{xy}, \quad (\text{A.23})$$

and

$$E_\theta^k \left\{ \frac{1}{n} \sum_{t=1}^n \tilde{x}_t \phi_{i,t} \middle| \mathbf{y} \right\} = \rho_x^i, \quad i = 1, 2, \dots, k, \quad (\text{A.24})$$

where  $E_\theta^k\{\cdot|\mathbf{y}\}$  denotes expectation with respect to  $V(\cdot|\mathbf{y}, \theta, k)$ . This parameter vector, that will be denoted by  $\hat{\theta}$ , can be found by solving  $(k+2)$  (A.22)–(A.24) with  $(k+2)$  unknowns. The fact that  $(\mathbf{x}, \mathbf{y}) \in H_n(B)$  guarantees that a solution exists (see (A.11)–(A.13)). The resulting  $\hat{\theta}$  attains the maximum of  $V(\mathbf{x}|\mathbf{y}, \theta, k)$  as can easily be seen.

To estimate the volume of  $T_\varepsilon^k(\mathbf{x}|\mathbf{y})$ , we derive upper and lower bounds on its probability under  $V(\cdot|\mathbf{y}, \hat{\theta}, k)$ . As for an upper bound, we have

$$\begin{aligned} 1 &\geq V(\{\tilde{\mathbf{x}} : \tilde{\mathbf{x}} \in T_\varepsilon^k(\mathbf{x}|\mathbf{y})\}|\mathbf{y}, \hat{\theta}, k) \\ &= \int_{T_\varepsilon^k(\mathbf{x}|\mathbf{y})} V(\tilde{\mathbf{x}}|\mathbf{y}, \hat{\theta}, k) d\tilde{\mathbf{x}} \\ &\geq \text{Vol}\{T_\varepsilon^k(\mathbf{x}|\mathbf{y})\} \cdot \inf_{\tilde{\mathbf{x}} \in T_\varepsilon^k(\mathbf{x}|\mathbf{y})} V(\tilde{\mathbf{x}}|\mathbf{y}, \hat{\theta}, k) \\ &\geq \text{Vol}\{T_\varepsilon^k(\mathbf{x}|\mathbf{y})\} \cdot V(\mathbf{x}|\mathbf{y}, \hat{\theta}, k) \\ &\quad \cdot \exp\left\{-n \left[ \frac{1}{2\hat{\sigma}_0^2} \left( \hat{\alpha}^2 + 2|\hat{\alpha}| + 2 \sum_{i=1}^k |\hat{\beta}_i| \right) \cdot \varepsilon \right]\right\} \\ &\geq \text{Vol}\{T_\varepsilon^k(\mathbf{x}|\mathbf{y})\} \cdot V(\mathbf{x}|\mathbf{y}, \hat{\theta}, k) \cdot \exp[-n\varepsilon B^2(B+k+1)], \quad (\text{A.25}) \end{aligned}$$

where the second last inequality follows from a derivation similar to (21) and the last inequality follows from the fact that  $(\mathbf{x}, \mathbf{y}) \in H_n(B)$ . Thus, we get

$$\text{Vol}\{T_\varepsilon^k(\mathbf{x}|\mathbf{y})\} \leq \frac{\exp[n\varepsilon B^2(B+k+1)]}{\max_\theta V(\mathbf{x}|\mathbf{y}, \theta, k)}, \quad (\text{A.26})$$

which completes the proof of part a) of the lemma.

As for part b), we use a similar technique by showing that the set  $T_\varepsilon^k(\mathbf{x}|\mathbf{y})$  has a high probability under  $V(\cdot|\mathbf{y}, \hat{\theta}, k)$ . By the union bound we have

$$\begin{aligned} V(T_\varepsilon^k(\mathbf{x}|\mathbf{y})|\mathbf{y}, \hat{\theta}, k) &\geq V(D_\Delta|\mathbf{y}, \hat{\theta}, k) \\ &\quad - V\left\{\left\{\tilde{\mathbf{x}} : \left| \frac{1}{n} \sum_{t=1}^n \tilde{x}_t^2 - \rho_{xx} \right| > \varepsilon\right\} \middle| \mathbf{y}, \hat{\theta}, k\right\} \\ &\quad - V\left\{\left\{\tilde{\mathbf{x}} : \left| \frac{1}{n} \sum_{t=1}^n \tilde{x}_t y_t - \rho_{xy} \right| > \varepsilon\right\} \middle| \mathbf{y}, \hat{\theta}, k\right\} \\ &\quad - \sum_{i=1}^k V\left\{\left\{\tilde{\mathbf{x}} : \left| \frac{1}{n} \sum_{t=1}^n \tilde{x}_t \phi_{i,t} - \rho_x^i \right| > \varepsilon\right\} \middle| \mathbf{y}, \hat{\theta}, k\right\}. \quad (\text{A.27}) \end{aligned}$$

We next underbound all the probabilities on the right-hand side of (A.27). Consider the backward channel defined in (A.21) and let

$$g(\boldsymbol{\nu}) = (2\pi\hat{\sigma}_0^2)^{-n/2} \prod_{t=1}^n \exp\left(-\frac{\nu_t^2}{2\hat{\sigma}_0^2}\right) \quad (\text{A.28})$$

denote the pdf of the noise, where  $\boldsymbol{\nu} = (\nu_1, \dots, \nu_n)$ , and let  $E_g(\cdot)$  denote expectation with respect to  $g(\cdot)$ . Thus,

$$V\left\{\left\{\tilde{\mathbf{x}} : \left| \frac{1}{n} \sum_{t=1}^n \tilde{x}_t y_t - \rho_{xy} \right| > \varepsilon\right\} \middle| \mathbf{y}, \hat{\theta}, k\right\}$$



$$\begin{aligned}
&= g \left\{ \nu : \left| \frac{1}{n} \sum_{t=1}^n \nu_t y_t \right| > \varepsilon \right\} \\
&\leq \frac{1}{\varepsilon^2} E_g \left( \frac{1}{n} \sum_{t=1}^n \nu_t y_t \right)^2 \\
&\leq \frac{1}{n\varepsilon^2} \left[ \sum_{t=1}^n y_t^2 \right] \cdot E_g \left\{ \frac{1}{n} \sum_{t=1}^n \nu_t^2 \right\} \\
&\leq \frac{B\hat{\sigma}_0^2}{n\varepsilon^2} \leq \frac{B^2}{n\varepsilon^2}, \tag{A.29}
\end{aligned}$$

where we have used the Chebychev inequality and the facts that  $(\mathbf{x}, \mathbf{y}) \in H_n(B)$  and  $\hat{\sigma}_0^2 \leq n^{-1} \sum_{t=1}^n x_t^2$ . In the same manner, it is easy to see that

$$V \left( \left\{ \tilde{\mathbf{x}} : \left| \frac{1}{n} \sum_{t=1}^n \tilde{x}_t \phi_{i,t} - \rho_{xx} \right| > \varepsilon \right\} \middle| \mathbf{y}, \hat{\theta}, k \right) \leq \frac{L^2 B}{n\varepsilon^2}, \tag{A.30}$$

where  $L$  is the uniform bound on  $\phi_{i,t}$ . Now,

$$\begin{aligned}
&V \left( \left\{ \tilde{\mathbf{x}} : \left| \frac{1}{n} \sum_{t=1}^n \tilde{x}_t^2 - \rho_{xx} \right| > \varepsilon \right\} \middle| \mathbf{y}, \hat{\theta}, k \right) \\
&= g \left\{ \nu : \left| \frac{1}{n} \sum_{t=1}^n \nu_t^2 - \hat{\sigma}_0^2 \right. \right. \\
&\quad \left. \left. + \frac{2}{n} \sum_{t=1}^n \left( \hat{\alpha} y_t + \sum_{i=1}^k \beta_i \phi_{i,t} \right) \nu_t \right| > \varepsilon \right\} \\
&\leq g \left\{ \nu : \left| \frac{1}{n} \sum_{t=1}^n \nu_t^2 - \hat{\sigma}_0^2 \right| > \frac{\varepsilon}{2} \right\} \\
&\quad + g \left\{ \nu : \left| \frac{2}{n} \sum_{t=1}^n \left( \hat{\alpha} y_t + \sum_{i=1}^k \beta_i \phi_{i,t} \right) \nu_t \right| > \frac{\varepsilon}{2} \right\}. \tag{A.31}
\end{aligned}$$

The second term on the right-most side of (A.31) can be overbounded similarly to (A.29), resulting in

$$g \left\{ \nu : \left| \frac{2}{n} \sum_{t=1}^n \left( \hat{\alpha} y_t + \sum_{i=1}^k \beta_i \phi_{i,t} \right) \nu_t \right| > \frac{\varepsilon}{2} \right\} \leq \frac{16B^2}{n\varepsilon^2}, \tag{A.32}$$

where we have used the fact that projecting a vector on a linear subspace shrinks its norm and hence,

$$\frac{1}{n} \sum_{t=1}^n \left( \hat{\alpha} y_t + \sum_{i=1}^k \hat{\beta}_i \phi_{i,t} \right)^2 \leq \frac{1}{n} \sum_{t=1}^n x_t^2 \leq \sigma_x^2 (1 + \Delta) \leq B. \tag{A.33}$$

The first term in (A.31) is again bounded by the Chebychev inequality, yielding

$$\begin{aligned}
g \left\{ \nu : \left| \frac{1}{n} \sum_{t=1}^n \nu_t^2 - \hat{\sigma}_0^2 \right| > \frac{\varepsilon}{2} \right\} &\leq \frac{4}{\varepsilon^2} E_g \left\{ \left( \frac{1}{n} \sum_{t=1}^n \nu_t^2 - \hat{\sigma}_0^2 \right)^2 \right\} \\
&= \frac{8\hat{\sigma}_0^4}{n\varepsilon^2} \leq \frac{8B^2}{n\varepsilon^2}. \tag{A.34}
\end{aligned}$$

The term  $V(D_\Delta | \mathbf{y}, \hat{\theta}, k)$ , depends only on  $n$  and  $\Delta$  but not on  $\varepsilon$ . Since  $\rho_{xx}$  must lie in the interval  $[\sigma_x^2(1 - \Delta), \sigma_x^2(1 + \Delta)]$ , then it is either an internal point in this interval or an endpoint. In the former case,  $V(D_\Delta | \mathbf{y}, \hat{\theta}, k) \rightarrow 1$  similarly as in (A.31)–(A.34). In the latter case, say,  $\rho_{xx} = \sigma_x^2(1 - \Delta)$ , the dominating event is

$$\begin{aligned}
0 &\leq \sum_{t=1}^n \tilde{x}_t^2 - \rho_{xx} \\
&= \left[ \frac{1}{n} \sum_{t=1}^n \nu_t^2 - \hat{\sigma}_0^2 \right] + \left[ \frac{2}{n} \sum_{t=1}^n \left( \hat{\alpha} y_t + \sum_{i=1}^k \beta_i \phi_{i,t} \right) \nu_t \right] \\
&\triangleq U_1 + U_2. \tag{A.35}
\end{aligned}$$

The probability of (A.35) is larger than the probability that both  $U_1 \geq 0$  and  $U_2 \geq 0$ . Now  $\Pr\{U_1 \geq 0\} \rightarrow 1/2$  by the central limit theorem. For every sequence  $\nu$  with a given energy  $\sum_t \nu_t^2$ , either  $\nu$  or  $-\nu$  (which both have the same probability density) is associated with a positive  $U_2$ . Thus, the probability that both  $U_1$  and  $U_2$  are nonnegative is asymptotically at least  $1/4$ . A similar consideration holds when  $\rho_{xx} = \sigma_x^2(1 + \Delta)$ . Therefore,  $V(D_\Delta | \mathbf{y}, \hat{\theta}, k) \triangleq \gamma_n(\Delta)$  is asymptotically never less than  $1/4$ . Combining (A.27), (A.29), (A.30), (A.32), (A.34) and (A.35), we get

$$\begin{aligned}
\gamma_n(\Delta) &\geq \frac{B(kL^2 + 25B)}{n\varepsilon^2} \\
&\leq V \left\{ T_\varepsilon^k(\mathbf{x} | \mathbf{y}) \middle| \mathbf{y}, \hat{\theta}, k \right\} \\
&\leq \text{Vol}\{T_\varepsilon^k(\mathbf{x} | \mathbf{y})\} \cdot \max_\theta V(\mathbf{x} | \mathbf{y}, \theta, k) \\
&\quad \cdot \exp[n\varepsilon B^2(B + k + 1)], \tag{A.36}
\end{aligned}$$

where the second inequality is derived similarly to (A.25). This completes the proof of part b) of Lemma 3.  $\square$

#### ACKNOWLEDGMENT

Interesting discussions with S. Shamai (Shitz), M. Feder, J. Ziv, H. Landau, and N. Seshadri are greatly appreciated. The author wishes to thank the anonymous referees for their useful comments and suggestions.

#### REFERENCES

- [1] V.D. Goppa, "Nonprobabilistic mutual information without memory," *Probl. Contr. Inform. Theory*, vol. 4, pp. 97–102, 1975.
- [2] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. New York: Academic, 1981.
- [3] J. Ziv, "Universal decoding for finite-state channels," *IEEE Trans. Inform. Theory*, vol. IT-31, no. 4, pp. 453–460, July 1985.
- [4] J. Ziv and A. Lempel, "Compression of individual sequences via variable-rate coding," *IEEE Trans. Inform. Theory*, vol. IT-24, no. 5, pp. 530–536, Sept. 1978.
- [5] S. U. H. Qureshi, "Adaptive equalization," *Proc. IEEE*, vol. 73, pp. 1349–1387, Sept. 1985.
- [6] R. G. Gallager, *Information Theory and Reliable Communication*. New York: Wiley, 1968.
- [7] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.