Analysis of Mismatched Estimation Errors Using Gradients of Partition Functions*

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

We consider the problem of signal estimation (denoising) from a statistical-mechanical perspective, in continuation to a recent work on the analysis of mean-square error (MSE) estimation using a direct relationship between optimum estimation and certain partition functions. The paper consists of essentially two parts. In the first part, using the aforementioned relationship, we derive single-letter expressions of the mismatched MSE of a codeword (from a randomly selected code), corrupted by a Gaussian vector channel. In the second part, we provide several examples to demonstrate phase transitions in the behavior of the MSE. These examples enable us to understand more deeply and to gather intuition regarding the roles of the real and the mismatched probability measures in creating these phase transitions.

Index Terms

Minimum mean-square error (MMSE), mismatched MSE, partition function, statistical-mechanics, conditional mean estimation, phase transitions, threshold effect.

I. INTRODUCTION

The connections and the interplay between information theory, statistical physics and signal estimation have been known for several decades [1-4], and they are still being studied from a variety of aspects, see, for example [5-17] and many references therein.

Recently, in [6], the well known I-MMSE relation [8], which relates the mutual information and the derivative of the minimum mean-square error (MMSE), was further explored using a statistical physics perspective. Specifically, in their analysis, the authors of [6] exploit the natural "mapping" between information theory problems and certain models of many-particle systems in statistical mechanics (see, e.g., [18, 19]). One of the main contributions in [6] is the demonstration of the usefulness of statistical-mechanical tools (in particular, utilizing the fact that the mutual information can be viewed as the partition function of a certain physical system) in assessing MMSE via the I-MMSE relation of [8]. More recently, Merhav [5] proposed a more flexible method, whose main idea is that, for the purpose of evaluating the covariance matrix of the MMSE estimator, one may use other information measures, which have the form of a partition function and hence can be analyzed using methods of statistical physics (see, e.g., [18-26] and many references therein). The main advantage of the proposed approach over the I-MMSE relations, is its full generality: Any joint probability function $P(x, y)$, where x and y designate the channel input to be estimated and the channel output, respectively, can be handled (for example, the channel does not have to be additive or Gaussian). Moreover, using this approach, any mismatch, both in the source and the channel, can be considered.

This paper is a further development of [5] in the above described direction. Particularly, in [5, Section IV. A], the problem of mismatched estimation of a codeword, transmitted over an additive white Gaussian (AWGN) channel, was considered. It was shown that the mismatched MSE exhibits phase transitions at some rate thresholds, which depend upon the real and the mismatched parameters of the problem, and the behavior of the receiver. To wit, the mismatched MSE acts inherently differently for a *pessimistic* and *optimistic* receivers, where in the example considered in [5, Section IV. A] pessimism literally means that the estimator assumes that the channel is worse than it really is (in terms of signal-to-noise ratio (SNR)), and the vice versa for optimism. In this paper, we extend the above described model to a much more general one; the Gaussian vector channel, which has a plenty of applications in communications and signal processing. It is important to emphasize that compared to [5, 6], it will be seen that: (1) the mathematical analysis is much more complicated (consisting of some new concepts), and (2) the notions of pessimism and optimism described above, also play a significant role in this model, although their physical meanings in general are not obvious. Moreover, in contrast to previous work on mismatched estimation, in this paper, the interesting case of channel mismatch is explored, namely, the receiver has a wrong assumption on the channel. In order to demonstrate the usefulness of the theoretical results derived for the general model, we also provide a few examples associated with some specific channel transfer functions, and draw conclusions and insights regarding the threshold effects in the behavior of

the partition function and the MSE.

As was mentioned earlier, we consider the Gaussian vector channel model

$$\boldsymbol{Y} = \boldsymbol{A}\boldsymbol{X} + \boldsymbol{N}, \tag{1}$$

where $\boldsymbol{N} \in \mathbb{R}^n$ is a Gaussian white noise vector and $\boldsymbol{A}$ is a deterministic $n \times n$ matrix representing a linear transformation induced by a given linear system. The vector $\boldsymbol{X} \in \mathbb{R}^n$ is chosen uniformly at random from a codebook (which is itself selected at random as well). There are several motivations for codeword estimation. One example is that of a user that, in addition to its desired signal, receives also a relatively strong interference signal, which carries digital information intended to other users, and which comes from a codebook whose rate exceeds the capacity of this crosstalk channel between the interferer and our user, so that the user cannot fully decode this interference. Nevertheless, our user would like to estimate the interference as accurately as possible for the purpose of cancellation. Furthermore, we believe that the tools/concepts developed in this paper for handling matched and mismatched problems, can be used in other applications in signal processing and communication. Such examples are denoising (see for example, [27-29]), mismatched decoding (for example, [30]), blind deconvolution (for example, [31, 32]), and many other applications. Note that although the aforementioned examples are radically different (in terms of their basic models and systematization), they will all suffer from mismatch when estimating the input signals.

In the special case of matched estimation, it will be shown that the MMSE is asymptotically given by

$$\lim_{n \to \infty} \frac{\text{mmse}\left(\boldsymbol{X} \mid \boldsymbol{Y}\right)}{n} = \begin{cases} \frac{1}{2\pi} \int_0^{2\pi} \frac{P_x}{1 + |\mathbf{H}(\omega)|^2 P_x \beta} \mathrm{d}\omega, & \text{if } R > R_c \\ 0, & \text{if } R \leq R_c \end{cases} \tag{2}$$

where

$$R_c \triangleq \frac{1}{4\pi} \int_0^{2\pi} \ln\left(1 + |\mathbf{H}(\omega)|^2 P_x \beta\right) \mathrm{d}\omega, \tag{3}$$

in which mmse $\left(\boldsymbol{X} \mid \boldsymbol{Y}\right)$ is the estimation error results from estimating $\boldsymbol{X}$ based on $\boldsymbol{Y}$, using the MMSE estimator, $1/\beta$ and P_x denote the noise variance and the transmitted power, respectively, and $\mathbf{H}(\omega)$ is the frequency response of the linear system $\boldsymbol{A}$. As can be seen from the above formula, for $R < R_c$ the MMSE essentially vanishes since the correct codeword can be reliably decoded, whereas for $R > R_c$, the MMSE is simply the estimation error which results by the Wiener filter that would have been applied had the input been a zero-mean, i.i.d. Gaussian process, with variance $1/\beta$. Accordingly, it will be seen that for $R > R_c$ the MMSE estimator is simply the Wiener filter. It is important to emphasize that while

the above result may seem to be a natural generalization of the results in [5, 6] (where $\boldsymbol{A}$ is taken to be identity matrix), the analysis (and results) of the mismatched case is by far more complicated and non-trivial. Indeed, it will be seen that in the mismatched case, the MSE is essentially separated into two cases, each exhibiting a completely different behavior. Further physical insights regarding the above result and other results will be presented later on.

The remaining part of this paper is organized as follows. In Section II, we first establish notation conventions. Then, the model considered is presented and the problem is formulated. In Section III, the main results are stated and discussed. In Section IV, we provide a few examples which illustrate the theoretical results. In Section V, we discuss the techniques and methodologies that are utilized in order to prove the main results, along with a brief background and summary on the basic relations between the conditional mean estimator, as well as its error covariance matrix and the aforementioned partition function, which were derived in [5]. In Section VI, the main results are proved. Finally, our conclusions appear in Section VII.

II. NOTATION CONVENTIONS AND PROBLEM FORMULATION

A. Notation Conventions

Throughout this paper, scalar random variables (RV's) will be denoted by capital letters, their sample values will be denoted by the respective lower case letters and their alphabets will be denoted by the respective calligraphic letters. A similar convention will apply to random vectors and their sample values, which will be denoted with same symbols in the bold face font. Thus, for example, $\boldsymbol{X}$ will denote a random vector $(X_1, \ldots, X_n)$ and $\boldsymbol{x} = (x_1, \ldots, x_n)$ is a specific vector value in $\mathcal{X}^n$, the n-th Cartesian power of $\mathcal{X}$. The notations $\boldsymbol{x}_i^j$ and $\boldsymbol{X}_i^j$, where i and j are integers and $i \leq j$, will designate segments $(x_i, \ldots, x_j)$ and $(X_i, \ldots, X_j)$, respectively. Probability functions will be denoted generically by the letter P or P'. In particular, $P(\boldsymbol{x}, \boldsymbol{y})$ is the joint probability mass function (in the discrete case) or the joint density (in the continuous case) of the desired channel input vector $\boldsymbol{x}$ and the observed channel output vector $\boldsymbol{y}$. Accordingly, $P(\boldsymbol{x})$ will denote the marginal of $\boldsymbol{x}$, $P(\boldsymbol{y} \mid \boldsymbol{x})$ will denote the conditional probability or density of $\boldsymbol{y}$ given $\boldsymbol{x}$, induced by the channel, and so on.

The expectation operator of a generic function $f(\boldsymbol{x}, \boldsymbol{y})$ with respect to (w.r.t.) the joint distribution of $\boldsymbol{X}$ and $\boldsymbol{Y}$, $P(\boldsymbol{x}, \boldsymbol{y})$, will be denoted by $\boldsymbol{E}\{f(\boldsymbol{X}, \boldsymbol{Y})\}$. Accordingly, $\boldsymbol{E}'\{f(\boldsymbol{X}, \boldsymbol{Y})\}$ means that the expectation is performed w.r.t. $P'(\boldsymbol{x}, \boldsymbol{y})$. The conditional expectation of the same function given that $\boldsymbol{Y} = \boldsymbol{y}$, denoted $\boldsymbol{E}\{f(\boldsymbol{X}, \boldsymbol{Y}) \mid \boldsymbol{Y} = \boldsymbol{y}\}$ and which is obviously identical to $\boldsymbol{E}\{f(\boldsymbol{X}, \boldsymbol{y}) \mid \boldsymbol{Y} = \boldsymbol{y}\}$, is, of course, a function of $\boldsymbol{y}$. On substituting $\boldsymbol{Y}$ in this function, this becomes a random variable which

will be denoted by $\boldsymbol{E}\{f(\boldsymbol{X}, \boldsymbol{Y}) \mid \boldsymbol{Y}\}$. When using vectors and matrices in a linear-algebraic format, n-dimensional vectors, like $\boldsymbol{x}$ (and $\boldsymbol{X}$), will be understood as column vectors, the operators $(\cdot)^T$ and $(\cdot)^H$ will denote vector or matrix transposition and vector or matrix conjugate transposition, respectively, and so, $\boldsymbol{x}^T$ would be a row vector. For two positive sequences $\{a_n\}$ and $\{b_n\}$, the notation $a_n \doteq b_n$ means equivalence in the exponential order, i.e., $\lim_{n\to\infty} \frac{1}{n} \log(a_n/b_n) = 0$. For two sequences $\{a_n\}$ and $\{b_n\}$, the notations $a_n \sim b_n$ and $a_n \lesssim b_n$ mean $\lim_{n\to\infty}(a_n/b_n) = 1$ and $\lim_{n\to\infty}(a_n/b_n) \leq 1$, respectively. Finally, the indicator function of an event $\mathcal{A}$ will be denoted by $\mathbb{1}\{\mathcal{A}\}$.

B. Model and Problem Formulation

Let $\mathcal{C} = \{\boldsymbol{x}_0, \ldots, \boldsymbol{x}_{M-1}\}$ denote a codebook of size $M = e^{nR}$, which is selected at random (and then revealed to the estimator) in the following manner: Each $\boldsymbol{x}_i$ is drawn independently under the uniform distribution over the surface of the n-dimensional hypersphere, which is centered at the origin, and whose radius is $\sqrt{nP_x}$. Finally, let $\boldsymbol{X}$ assume a uniform distribution over $\mathcal{C}$. We consider the Gaussian vector channel model

$$\boldsymbol{Y} = \boldsymbol{A}\boldsymbol{X} + \boldsymbol{N}, \tag{4}$$

where $\boldsymbol{Y}$, $\boldsymbol{X}$ and $\boldsymbol{N}$ are random vectors in $\mathbb{R}^n$, designating the channel output vector, the transmitted codeword and the noise vector, respectively. It is assumed that the components of the noise vector, $\boldsymbol{N}$, are i.i.d., zero-mean, Gaussian random variables with variance $1/\beta$, where β is a given positive constant designating the signal-to-noise ratio (SNR) (for $P_x = 1$), or the inverse temperature in the statistical-mechanical jargon. We further assume that $\boldsymbol{X}$ and $\boldsymbol{N}$ are statistically independent. Finally, the channel matrix, $\boldsymbol{A} \in \mathbb{R}^{n \times n}$, is assumed to be a given deterministic Toeplitz matrix, whose entries are given by the coefficients of the impulse response of a given linear system. Specifically, let $\{h_k\}$ denote the generating sequence (or impulse response) of $\boldsymbol{A}$, so that $\boldsymbol{A} = \{a_{i,j}\}_{i,j} = \{h_{i-j}\}_{i,j}$, and let $\mathbf{H}(\omega)$ designate the frequency response (Fourier transform) of $\{h_k\}$.

As was mentioned previously, we analyze the problem of mismatched codeword estimation which is formulated as follows: Consider a mismatched estimator which is the conditional mean of $\boldsymbol{X}$ given $\boldsymbol{Y}$, based on an incorrect joint distribution $P'(\boldsymbol{x}, \boldsymbol{y})$, whereas the true joint distribution continues to be $P(\boldsymbol{x}, \boldsymbol{y})$. Accordingly, the *mismatched MSE* is defined as

$$\mathrm{mse}(\boldsymbol{X} \mid \boldsymbol{Y}) \overset{\triangle}{=} \boldsymbol{E}\left\|\boldsymbol{X} - \boldsymbol{E}'\{\boldsymbol{X} \mid \boldsymbol{Y}\}\right\|^2 \tag{5}$$

where $\boldsymbol{E}'\{\boldsymbol{X} \mid \boldsymbol{Y}\}$ is the conditional expectation w.r.t. the mismatched measure P'. In this paper, the following mismatch mechanism is assumed: The input measure is matched, i.e., $P(\boldsymbol{x}) = P'(\boldsymbol{x})$ (namely,

the mismatched estimator knows the true code), both conditional measures ("channels") $P(\cdot \mid \boldsymbol{x})$ and $P'(\cdot \mid \boldsymbol{x})$ are Gaussian, but are associated with different channel matrices. More precisely, while the true channel matrix (under P) is $\boldsymbol{A}$, the assumed channel matrix (under P') is $\boldsymbol{A}'$, another Toeplitz matrix, generated by the impulse response $\{h'_k\}$, whose frequency response is $\mathbf{H}'(\omega)$. It should be pointed out, however, that the analysis in this paper can be easily carried out also for the case of mismatch in the input distribution, or mismatch in the noise distribution, which has been already considered in [5]. Using the theoretical tools derived in [5], the mismatched MSE (and the MMSE as a special case) will be derived for the model described above.

A very important function, which will be pivotal to our derivation of both the mismatched estimator and the MSE, is the *partition function*, which is defined as follows.

Definition 1 (Partition Function) Let $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_n)^T$ be a column vector of n real-valued parameters. The partition function w.r.t. the joint distribution $P(\boldsymbol{x}, \boldsymbol{y})$, denoted by $Z(\boldsymbol{y}, \boldsymbol{\lambda})$, is defined as

$$Z(\boldsymbol{y}, \boldsymbol{\lambda}) \triangleq \sum_{\boldsymbol{x} \in \mathcal{X}^n} \exp\left\{\boldsymbol{\lambda}^T \boldsymbol{x}\right\} P(\boldsymbol{x}, \boldsymbol{y}). \tag{6}$$

In the above definition, it is assumed that the sum (or integral, in the continuous case) converges uniformly at least in some neighborhood of $\boldsymbol{\lambda} = \mathbf{0}$ [1]. Accordingly, under the above described model, the mismatched partition function is given by

$$Z'(\boldsymbol{y}, \boldsymbol{\lambda}) \triangleq \sum_{\boldsymbol{x} \in \mathcal{C}} \exp\left\{\boldsymbol{\lambda}^T \boldsymbol{x}\right\} P'(\boldsymbol{x}, \boldsymbol{y}) \tag{7}$$

$$= (2\pi/\beta)^{-n/2} \sum_{\boldsymbol{x} \in \mathcal{C}} e^{-nR} \exp\left[-\beta \left\|\boldsymbol{y} - \boldsymbol{A}'\boldsymbol{x}\right\|^2/2 + \boldsymbol{\lambda}^T \boldsymbol{x}\right]. \tag{8}$$

Remark 1 In the above definition, the role of $\boldsymbol{\lambda}$ will be understood later on. In a nutshell, the idea [5] is that the gradient of $\ln Z'(\boldsymbol{y}, \boldsymbol{\lambda})$ w.r.t. $\boldsymbol{\lambda}$, computed at $\boldsymbol{\lambda} = \mathbf{0}$, simply gives the mismatched MSE estimator, $\boldsymbol{E}'\{\boldsymbol{X} \mid \boldsymbol{y}\}$, and the expectation of the Hessian of $\ln Z'(\boldsymbol{y}, \boldsymbol{\lambda})$ w.r.t. $\boldsymbol{\lambda}$, computed at $\boldsymbol{\lambda} = \mathbf{0}$, gives the MSE. Nevertheless, in the next section, where we present the main results, the dependency of the different quantities in $\boldsymbol{\lambda}$ will not be apparent, as they will already be computed at $\boldsymbol{\lambda} = \mathbf{0}$.

III. Main Results and Discussion

In this section, our main results are presented and discussed. The proofs of these results are provided in Section VI. The asymptotic MMSE, which is obtained as a special case of the mismatched case ($P = P'$),

---

[1] In case that this assumption does not hold, one can instead, parametrize each component λ_i of $\boldsymbol{\lambda}$ as a purely imaginary number $\lambda_i = j\omega_i$ where $i = \sqrt{-1}$, similarly to the definition of the characteristics function.

is given in the following theorem.

Theorem 1 (Asymptotic MMSE) Consider the model defined in Subsection II-B, and assume that the sequence $\{h_k\}_k$ is square summable. Then, the asymptotic MMSE is given by

$$\lim_{n\to\infty} \frac{\mathrm{mmse}\,(\boldsymbol{X}\mid\boldsymbol{Y})}{n} = \begin{cases} \frac{1}{2\pi}\int_0^{2\pi} \frac{P_x}{1+|\mathbf{H}(\omega)|^2 P_x\beta}\mathrm{d}\omega, & R > R_c \\ 0, & R \le R_c \end{cases} \tag{9}$$

where

$$R_c \triangleq \frac{1}{4\pi}\int_0^{2\pi} \ln\left(1 + |\mathbf{H}(\omega)|^2\, P_x\beta\right)\mathrm{d}\omega. \tag{10}$$

From the above result, it can be seen that for $R > R_c$ the MMSE is simply the estimation error which results by the Wiener filter that would have been applied had the input been a zero-mean, i.i.d. Gaussian process, with variance $1/\beta$. Accordingly, it is also shown in Section VI that the MMSE estimator is exactly the Wiener filter.

In the next theorem, we present the mismatched MSE. In contrast to the MMSE, unfortunately, the MSE does not lend itself to a simple closed-form expression. As will be seen in Section VI, this complexity stems from the complicated dependence of the partition function on $\boldsymbol{\lambda}$. Nevertheless, despite of the following non-trivial expressions, it should be emphasized that the obtained MSE expression has a single-letter formula, and thus, practically, it can be easily calculated at least numerically. Let us define the following auxiliary variables

$$P_a(\omega) \triangleq \frac{|\mathbf{H}'(\omega)|^2\,\beta\left(2 + P_x\beta\,|\mathbf{H}(\omega)|^2\right) + \gamma_0}{\left(|\mathbf{H}'(\omega)|^2\,\beta + \gamma_0\right)^2} \tag{11}$$

where γ_0 is chosen such that $\int_0^{2\pi} P_a(\omega)\,\mathrm{d}\omega = 2\pi P_x$. Next define

$$\mathbf{B}(\omega) \triangleq \frac{\left(|\mathbf{H}'(\omega)|^2 + \gamma_0\right) - 2\left(|\mathbf{H}'(\omega)|^2\,\beta\left(2 + P_x\beta\,|\mathbf{H}(\omega)|^2\right) + \gamma_0\right)}{\left(|\mathbf{H}'(\omega)|^2 + \gamma_0\right)^3} \tag{12}$$

$$\mathbf{C}(\omega) \triangleq \frac{2\beta^2\,|\mathbf{H}'(\omega)|^2\left(|\mathbf{H}(\omega)|^2 + \frac{1}{\beta}\right)\mathbf{B}(\omega)}{\sqrt{1 + 4\beta^2\,|\mathbf{H}'(\omega)|^2\,P_a(\omega)\left(|\mathbf{H}(\omega)|^2 + \frac{1}{\beta}\right)}} \tag{13}$$

$$\vartheta \triangleq 2 + \frac{\int_0^{2\pi}\left[\frac{P_x}{P_x\gamma_0 + P_x|\mathbf{H}'(\omega)|^2\beta} - \mathbf{C}(\omega)\right]\mathrm{d}\omega}{\int_0^{2\pi} \mathbf{B}(\omega)\,\mathrm{d}\omega}, \tag{14}$$

and

$$\mathbf{\Xi}_1\left(\omega\right) \triangleq -\frac{\beta \mathbf{H}'^*\left(\omega\right)}{\left(\left|\mathbf{H}'\left(\omega\right)\right|^2 + \gamma_0\right)^2}\left[\vartheta - \frac{2\left(\left|\mathbf{H}'\left(\omega\right)\right|^2 + \gamma_0\right)^2 P_a\left(\omega\right) + 2\beta^2\left|\mathbf{H}'\left(\omega\right)\right|^2\left(\left|\mathbf{H}\left(\omega\right)\right|^2 + \frac{1}{\beta}\right)}{\sqrt{1 + 4\beta^2\left|\mathbf{H}'\left(\omega\right)\right|^2 P_a\left(\omega\right)\left(\left|\mathbf{H}\left(\omega\right)\right|^2 + \frac{1}{\beta}\right)}}\right]. \tag{15}$$

Let $\epsilon_{s,0}$, $\alpha_{1,0}$ and $\alpha_{2,0}$ be the solution of the following set of three simultaneous equations:

$$R + \frac{1}{4\pi}\int_0^{2\pi}\ln\left(\frac{2\epsilon_{s,0}}{P_x\left|\mathbf{H}'\left(\omega\right)\right|^2\alpha_{2,0} + 2P_x\alpha_{1,0}\epsilon_{s,0}}\right)\mathrm{d}\omega = 0 \tag{16}$$

$$\frac{1}{2\pi}\int_0^{2\pi}\frac{4\alpha_{1,0}\epsilon_{0,s}^2 + \left|\mathbf{H}'\left(\omega\right)\right|^2\alpha_{2,0}\left[\left(\left|\mathbf{H}\left(\omega\right)\right|^2 P_x + \frac{1}{\beta}\right)\alpha_{2,0} + 2\epsilon_{s,0}\right]}{\left(\left|\mathbf{H}'\left(\omega\right)\right|^2\alpha_{2,0} + 2\alpha_{1,0}\epsilon_{s,0}\right)^2}\mathrm{d}\omega = P_x \tag{17}$$

$$\frac{1}{2\pi}\int_0^{2\pi}\frac{4\alpha_{1,0}^2\epsilon_{s,0}^2\left(1 + P_x\beta\left|\mathbf{H}\left(\omega\right)\right|^2\right) + 4\left|\mathbf{H}'\left(\omega\right)\right|^2\alpha_{1,0}\epsilon_{s,0}^2\beta + 2\left|\mathbf{H}'\left(\omega\right)\right|^4\alpha_{2,0}\epsilon_{s,0}\beta}{2\beta\epsilon_{s,0}\left(\left|\mathbf{H}'\left(\omega\right)\right|^2\alpha_{2,0} + 2\alpha_{1,0}\epsilon_{s,0}\right)^2}\mathrm{d}\omega = 1. \tag{18}$$

Then, we define

$$\mathbf{K}\left(\omega\right) \triangleq 2\beta\epsilon_{s,0}\left(\left|\mathbf{H}'\left(\omega\right)\right|^2\alpha_{2,0} + 2\alpha_{1,0}\epsilon_{s,0}\right)^2 \tag{19}$$

$$\mathbf{T}\left(\omega\right) \triangleq 4\alpha_{1,0}^2\epsilon_{s,0}^2\left(1 + P_x\beta\left|\mathbf{H}\left(\omega\right)\right|^2\right) + 4\beta\left|\mathbf{H}'\left(\omega\right)\right|^2\alpha_{1,0}\epsilon_{s,0}^2 + 2\left|\mathbf{H}'\left(\omega\right)\right|^4\alpha_{2,0}\beta\epsilon_{s,0} \tag{20}$$

$$\mathbf{D}\left(\omega\right) \triangleq \left|\mathbf{H}'\left(\omega\right)\right|^2\alpha_{2,0} + 2\alpha_{1,0}\epsilon_{s,0} \tag{21}$$

$$\mathbf{R}\left(\omega\right) \triangleq 4\alpha_{1,0}\epsilon_{s,0}^2 + \left|\mathbf{H}'\left(\omega\right)\right|^2\alpha_{2,0}\left[\alpha_{2,0}\left(\left|\mathbf{H}\left(\omega\right)\right|^2 P_x + 2\epsilon_{s,0}\right)\right] \tag{22}$$

$$\mathbf{Q}\left(\omega\right) \triangleq \epsilon_{s,0}\left(P_x\left|\mathbf{H}'\left(\omega\right)\right|^2\alpha_{2,0} + 2P_x\alpha_{1,0}\epsilon_{s,0}\right) \tag{23}$$

$$V \triangleq \frac{1}{2\pi}\int_0^{2\pi}\frac{P_x\left|\mathbf{H}'\left(\omega\right)\right|^2\alpha_{2,0}}{\epsilon_{s,0}\left(P_x\left|\mathbf{H}'\left(\omega\right)\right|^2\alpha_{2,0} + 2P_x\alpha_{1,0}\epsilon_{s,0}\right)}\mathrm{d}\omega \tag{24}$$

$$F \triangleq \frac{1}{V}\frac{1}{2\pi}\int_0^{2\pi}\frac{P_x\left|\mathbf{H}'\left(\omega\right)\right|^2 r_2 + 2P_x\epsilon_{s,0}r_1}{P_x\left|\mathbf{H}'\left(\omega\right)\right|^2\alpha_{2,0} + 2P_x\alpha_{1,0}\epsilon_{s,0}}\mathrm{d}\omega \tag{25}$$

$$\gamma_1 \triangleq \frac{1}{2\pi}\int_0^{2\pi}\left[\frac{8\alpha_{1,0}\epsilon_{s,0}^2\left(1 + P_x\beta\left|\mathbf{H}\left(\omega\right)\right|^2\right) + 4\beta\left|\mathbf{H}'\left(\omega\right)\right|^2\epsilon_{s,0}^2}{\mathbf{K}\left(\omega\right)}\right.$$
$$\left. - \frac{8\mathbf{T}\left(\omega\right)\beta\epsilon_{s,0}^2\left(\left|\mathbf{H}'\left(\omega\right)\right|^2\alpha_{2,0} + 2\alpha_{1,0}\epsilon_{s,0}\right)}{\mathbf{K}^2\left(\omega\right)}\right]\mathrm{d}\omega \tag{26}$$

$$\gamma_2 \triangleq \frac{1}{2\pi}\int_0^{2\pi}\frac{2\mathbf{K}\left(\omega\right)\beta\epsilon_{s,0}\left|\mathbf{H}'\left(\omega\right)\right|^4 - 4\mathbf{T}\left(\omega\right)\beta\epsilon_{s,0}\left(\left|\mathbf{H}'\left(\omega\right)\right|^2\alpha_{2,0} + 2\alpha_{1,0}\epsilon_{s,0}\right)\left|\mathbf{H}'\left(\omega\right)\right|^2}{\mathbf{K}^2\left(\omega\right)}\mathrm{d}\omega \tag{27}$$

$$\gamma_3 \triangleq \frac{1}{2\pi} \int_0^{2\pi} \left[\frac{8\alpha_{1,0}^2 \epsilon_{s,0} \left(1 + P_x \beta |\mathbf{H}(\omega)|^2\right) + 8\beta \epsilon_{s,0} |\mathbf{H}'(\omega)|^2 \alpha_{1,0} + 2\beta \alpha_{2,0} |\mathbf{H}'(\omega)|^4}{\mathbf{K}(\omega)} \right.$$
$$\left. - \frac{\mathbf{T}(\omega) \left[2\beta \left(|\mathbf{H}'(\omega)|^2 \alpha_{2,0} + 2\alpha_{1,0} \epsilon_{s,0}\right)^2 + 8\beta \epsilon_{s,0} \alpha_{1,0} \left(|\mathbf{H}'(\omega)|^2 \alpha_{2,0} + 2\alpha_{1,0} \epsilon_{s,0}\right) \right]}{\mathbf{K}^2(\omega)} \right] \mathrm{d}\omega \quad (28)$$

$$\Upsilon(\omega) \triangleq \frac{-4\beta \alpha_{1,0} \epsilon_{s,0} \alpha_{2,0} - \beta \alpha_{2,0}^2 |\mathbf{H}'(\omega)|^2}{\mathbf{K}^2(\omega)} \quad (29)$$

$$\eta_1 \triangleq \frac{1}{2\pi} \int_0^{2\pi} \frac{4\mathbf{D}(\omega) \epsilon_{s,0}^2 - 4\mathbf{R}(\omega) \epsilon_{s,0}}{\mathbf{D}^3(\omega)} \mathrm{d}\omega \quad (30)$$

$$\eta_2 \triangleq \frac{1}{2\pi} \int_0^{2\pi} \left[\frac{|\mathbf{H}'(\omega)|^2 \left[\left(|\mathbf{H}(\omega)|^2 P_x + \frac{1}{\beta}\right) \alpha_{2,0} + 2\epsilon_{s,0} \right]}{\mathbf{D}^2(\omega)} \right.$$
$$\left. + \frac{|\mathbf{H}'(\omega)|^2 \alpha_{2,0} \left(|\mathbf{H}(\omega)|^2 P_x + \frac{1}{\beta}\right) \mathbf{D}(\omega) - 2\mathbf{R}(\omega) |\mathbf{H}'(\omega)|^2}{\mathbf{D}^3(\omega)} \mathrm{d}\omega \right] \quad (31)$$

$$\eta_3 \triangleq \frac{1}{2\pi} \int_0^{2\pi} \frac{8\mathbf{D}(\omega) \alpha_{1,0} \epsilon_{s,0} + 2\mathbf{D}(\omega) |\mathbf{H}'(\omega)|^2 \alpha_{2,0} - 4\mathbf{R}(\omega) \alpha_{1,0}}{\mathbf{D}^3(\omega)} \mathrm{d}\omega \quad (32)$$

$$\Lambda(\omega) \triangleq \frac{\alpha_{2,0}^2}{\mathbf{D}^2(\omega)} \quad (33)$$

$$r_1 = \frac{\eta_2 \gamma_3 - \gamma_2 \eta_3}{\gamma_2 \eta_1 - \eta_2 \gamma_1} \quad (34)$$

$$r_2 = \frac{\eta_1 \gamma_3 - \gamma_1 \eta_3}{\gamma_1 \eta_2 - \eta_1 \gamma_2} \quad (35)$$

$$\mathbf{J}_1(\omega) = \frac{\eta_2 \Upsilon(\omega) - \gamma_2 \Lambda(\omega)}{\gamma_2 \eta_1 - \eta_2 \gamma_1} \quad (36)$$

$$\mathbf{J}_2(\omega) = \frac{\eta_1 \Upsilon(\omega) - \gamma_1 \Lambda(\omega)}{\gamma_1 \eta_2 - \eta_1 \gamma_2} \quad (37)$$

$$\mathbf{J}(\omega) \triangleq \frac{1}{2\pi} \frac{\mathbf{J}_1(\omega) \int_0^{2\pi} \frac{2\epsilon_{s,0}^2 P_x}{\mathbf{Q}(\omega)} \mathrm{d}\omega + \mathbf{J}_2(\omega) \int_0^{2\pi} \frac{2\epsilon_{s,0} P_x |\mathbf{H}'(\omega)|^2}{\mathbf{Q}(\omega)} \mathrm{d}\omega}{V(1-F)} \quad (38)$$

$$\boldsymbol{\Xi}_2(\omega) \triangleq -2\mathbf{J}(\omega) \mathbf{H}'^*(\omega). \quad (39)$$

Finally, let

$$E_g \triangleq P_x - \mathrm{Re}\left(\frac{1}{\pi} \int_0^{2\pi} \boldsymbol{\Xi}_2(\omega)^* \mathbf{H}^*(\omega) P_x \mathrm{d}\omega \right) + \frac{1}{2\pi} \int_0^{2\pi} |\boldsymbol{\Xi}_2(\omega)|^2 \left(|\mathbf{H}(\omega)|^2 P_x + \frac{1}{\beta}\right) \mathrm{d}\omega, \quad (40)$$

and

$$E_p \triangleq P_x - \mathrm{Re}\left(\frac{1}{\pi} \int_0^{2\pi} \boldsymbol{\Xi}_1(\omega)^* \mathbf{H}^*(\omega) P_x \mathrm{d}\omega \right) + \frac{1}{2\pi} \int_0^{2\pi} |\boldsymbol{\Xi}_1(\omega)|^2 \left(|\mathbf{H}(\omega)|^2 P_x + \frac{1}{\beta}\right) \mathrm{d}\omega, \quad (41)$$

and we define the following critical rates

$$R_e \triangleq \frac{1}{4\pi} \int_0^{2\pi} \ln\left(P_x \gamma_0 + P_x \beta \left|\mathbf{H}'(\omega)\right|^2\right) d\omega \tag{42}$$

$$R_d \triangleq \frac{1}{2} + \beta P_x \frac{1}{2\pi} \int_0^{2\pi} \mathrm{Re}\left(\mathbf{H}'^*(\omega)\mathbf{H}(\omega)\right) d\omega$$

$$+ \frac{1}{4\pi} \int_0^{2\pi} \left|\mathbf{H}'(\omega)\right|^2 \beta \left(\frac{\left|\mathbf{H}'(\omega)\right|^2 \beta \left(2 + \beta \left|\mathbf{H}(\omega)\right|^2 P_x\right) + \gamma_0}{\left(\left|\mathbf{H}'(\omega)\right|^2 \beta + \gamma_0\right)^2} - P_x \right)$$

$$- \frac{1}{4\pi} \int_0^{2\pi} \frac{\left|\mathbf{H}'(\omega)\right|^2 \beta \left(3 + 2P_x\beta\left|\mathbf{H}(\omega)\right|^2\right) + \gamma_0}{\left(\left|\mathbf{H}'(\omega)\right|^2 \beta + \gamma_0\right)} \tag{43}$$

$$R_c \triangleq R_e + R_d \tag{44}$$

$$R_g \triangleq -\frac{1}{4\pi} \int_0^{2\pi} \ln\left(\frac{2\tilde{\epsilon}}{P_x \left|\mathbf{H}'(\omega)\right|^2 \tilde{\alpha}_2 + 2P_x\tilde{\alpha}_1\tilde{\epsilon}}\right) d\omega \tag{45}$$

where $\tilde{\alpha}_1$ and $\tilde{\alpha}_2$ solve the set of two simultaneous equations

$$\frac{1}{2\pi} \int_0^{2\pi} \frac{4\tilde{\alpha}_1\tilde{\epsilon}^2 + \left|\mathbf{H}'(\omega)\right|^2 \tilde{\alpha}_2 \left[\left(\left|\mathbf{H}(\omega)\right|^2 P_x + \frac{1}{\beta}\right)\tilde{\alpha}_2 + 2\tilde{\epsilon}\right]}{\left(\left|\mathbf{H}'(\omega)\right|^2 \tilde{\alpha}_2 + 2\tilde{\alpha}_1\tilde{\epsilon}\right)^2} d\omega = P_x \tag{46}$$

$$\frac{1}{2\pi} \int_0^{2\pi} \frac{4\tilde{\alpha}_1^2\tilde{\epsilon}^2\left(1 + P_x\beta\left|\mathbf{H}(\omega)\right|^2\right) + 4\left|\mathbf{H}'(\omega)\right|^2 \tilde{\alpha}_1\tilde{\epsilon}^2\beta + 2\left|\mathbf{H}'(\omega)\right|^4 \tilde{\alpha}_2\tilde{\epsilon}\beta}{2\beta\tilde{\epsilon}\left(\left|\mathbf{H}'(\omega)\right|^2 \tilde{\alpha}_2 + 2\tilde{\alpha}_1\tilde{\epsilon}\right)^2} d\omega = 1. \tag{47}$$

and

$$\tilde{\epsilon} = \frac{1}{2\beta} + \frac{P_x}{4\pi} \int_0^{2\pi} \left|\mathbf{H}'(\omega) - \mathbf{H}(\omega)\right|^2 d\omega. \tag{48}$$

We are now in a position to state our main theorem.

Theorem 2 (Mismatched MSE) Consider the model defined in Subsection II-B, and assume that the sequence $\{h_k\}_k$ is square summable. The (asymptotic) mismatched MSE is given as follows:

a) For $R_d \geq 0$

$$\lim_{n\to\infty} \frac{\mathrm{mse}\left(\boldsymbol{X} \mid \boldsymbol{Y}\right)}{n} = \begin{cases} 0, & R \leq R_c \\ E_p, & R > R_c \end{cases}. \tag{49}$$

b) For $R_d < 0$

$$\lim_{n \to \infty} \frac{\text{mse} \left(\boldsymbol{X} \mid \boldsymbol{Y} \right)}{n} = \begin{cases} 0, & R \leq R_g \\ E_g, & R_g < R \leq R_e \\ E_p, & R > R_e \end{cases} \tag{50}$$

In the jargon of statistical mechanics of spin arrays (see for example [33, Ch. 6]), the ranges of rates $R \leq R_c$ for $R_d \geq 0$, $R \leq R_g$ for $R_d < 0$, and $R \leq R_c$ in the matched case, correspond to the ordered phase (or ferromagnetic phase) in which the partition function is dominated by the correct codeword (and hence so is the posterior). Accordingly, in this range the MSE asymptotically vanishes, which literally means reliable communication. The intermediate range, $R_g < R \leq R_e$, which appears only in the mismatched case and only for $R_d < 0$, is analogous to the glassy phase (or "frozen" phase), in which the partition function is dominated by a sub-exponential number of wrong codewords. Intuitively, in this range, we may have the illusion that there is relatively little uncertainty about the transmitted codeword, but this is wrong due to the mismatch (as the main support of the mismatched posterior belongs to incorrect codewords). The remaining range corresponds to the paramagnetic phase, in which the partition function is dominated by an exponential number of wrong codewords. In Section IV, we will link between each one of the two cases $R_d \geq 0$ and $R_d < 0$, to "pessimistic" and "optimistic" behaviors of the receiver, which were already mentioned in the Introduction.

It is tempting to think that there should not be a range of rates for which the MSE (MMSE) vanishes, as we deal with an estimation problem rather than a decoding problem. Nonetheless, since codewords are being estimated, and there are a finite number of them, for low enough rates (up to some critical rate) the posterior is dominated by the correct codeword, and thus asymptotically, the estimation can be regarded as a maximum a posteriori probability (MAP) estimation, and so the MSE vanishes. In the same breath, note that this is not the case if mismatch in the input distribution is considered. For example, if the receiver's assumption on the transmitted energy is wrong, then no matter how low the rate is, there will always be an inherent error which stems from the fallacious averaging over a hypersphere with wrong radius (wrong codebook). Precisely, in this case, the estimated codeword will differ from the real one by an inevitable scaling of $\sqrt{P_x'/P_x}$, where P_x' is the mismatched power.

Finally, it is important to emphasize that the mismatched MSE estimator and the MMSE estimator can also be obtained as a byproduct of the analysis. However, since they will add only little further insights into the problem, we do not present them here. The interested reader can find their explicit expressions

in Section VI.

Remark 2 Although we have assumed that the transmitted codeword has a flat spectrum, the analysis can readily be extended to any input spectral density $S_x(\omega)$. In Section VI, we discuss the technical issues that should be considered in order to modify the analysis to hold for this generalization. As a concrete simple example, in the case of MMSE estimation, one obtains

$$\lim_{n \to \infty} \frac{\text{mmse}\left(\boldsymbol{X} \mid \boldsymbol{Y}\right)}{n} = \begin{cases} \frac{1}{2\pi} \int_0^{2\pi} \frac{S_x(\omega)}{1 + |\mathbf{H}(\omega)|^2 S_x(\omega)\beta} \mathrm{d}\omega, & R > R_c \\ 0, & R \le R_c \end{cases} \tag{51}$$

where

$$R_c \triangleq \frac{1}{4\pi} \int_0^{2\pi} \ln\left(1 + |\mathbf{H}(\omega)|^2 S_x(\omega)\beta\right) \mathrm{d}\omega. \tag{52}$$

Nevertheless, our assumption on flat input spectrum is reasonable when there is uncertainty at the encoder concerning the frequency response of the channel, as there are no "preferred" frequencies. Finally, note that as an application of the above issue, one may wish to consider the minimization of the MMSE w.r.t. the input spectral density.

IV. EXAMPLES

In this section, we provide a few examples in order to illustrate the theoretical results presented in the previous section. In particular, we present and explore the phase diagrams and the MSE's as functions of the rate and some parameters of the mismatched channel. The main goal in these examples is further understanding of the role of the true and the mismatched probability measures in creating phase transitions.

Example 1 We start with a simple example where both $\mathbf{H}(\omega)$ and $\mathbf{H}'(\omega)$ are low-pass filters (LPFs) that differ in their cutoff frequencies and gains

$$\mathbf{H}(\omega) = \begin{cases} 1, & |\omega| \le \frac{\pi}{2} \\ 0, & \text{else} \end{cases}, \tag{53}$$

and

$$\mathbf{H}'(\omega) = \begin{cases} \chi, & |\omega| \le \omega_c \\ 0, & \text{else} \end{cases} \tag{54}$$

for some $\chi > 0$ and $0 \le \omega_c \le \pi$. In the numerical calculations, we chose $\beta = P_x = 1$. Figures 1 and 2 show, respectively, the phase diagrams and the MSE's as functions of R and ω_c, for various values of

the gain χ. The first obvious observation is that the maximum range of rates for which the ferromagnetic phase dominates the partition function occurs at $\omega_c = \pi/2$ for each gain, as expected. Next, consider the case of $\chi = 1$, which means that the gain is matched. In this case, it is observed that for $\omega_c \leq \pi/2$, there are two phases: the ferromagnetic phase and the paramagnetic phase, and hence, based on Theorem 2, $R_d \geq 0$. On the other hand, for $\omega_c > \pi/2$, the glassy phase begins to play a role, and thus $R_d < 0$. Intuitively speaking, the case of $\omega_c \leq \pi/2$ corresponds to a pessimistic assumption of the receiver - lower bandwidth which translates to lower effective SNR, while $\omega_c > \pi/2$ corresponds to an optimistic assumption - higher effective SNR. These behaviors are consistent with the results obtained in [5], where the case of mismatch in the noise variance was considered (while assuming that $\boldsymbol{A} = \boldsymbol{A}'$ is the identity matrix).

In [5], $R_d > 0$ simply translates to $\beta > \beta'$ (the mismatched noise variance is larger than the actual one), namely, the estimator is pessimistic, while in the case of the reversed inequality it is overly optimistic. Accordingly, in the pessimistic case, the partition function exhibits a single phase transition, but at the price of a lower critical rate (compared to the matched case), which means that the range of rates for which reliable communication is possible is smaller. In the optimistic case, however, there is no loss in the critical rate, but there is a price of an additional phase transition. Now, for $\chi \neq 1$, the notions of pessimism and optimism are not a priori obvious. For example, it can be seen that for $\chi < 1$, and for a large enough cutoff frequency ω_c, the mismatched estimator can be regarded as an optimistic one. Also, for $\chi > 1$, apparently, the "price" of being too optimistic in the gain results in a dominant range of the glassy phase. Finally, note that the fact that the range of rates for which the ferromagnetic region dominates the partition function (namely, vanishing MSE) is decreasing with the excess of the optimism (e.g., for $\chi = 1$ and increasing of the cutoff frequency) is reasonable[2]. Indeed, the uncertainty in the frequency domain, causes the receiver to assume that the codewords are distributed in some subspace of the n-dimensional hypersphere. The size of this subspace is, of course, increasing as the receiver's assumption is more optimistic. Accordingly, the probability of error also increases, and thus the threshold rate for reliable communication decreases.

---

[2]In [5], in contrast to our case, for $\beta < \beta'$ ($R_d < 0$), the critical rate R_e is fixed for any mismatched noise variance value, namely, it is independent of the optimistic behavior of the receiver.
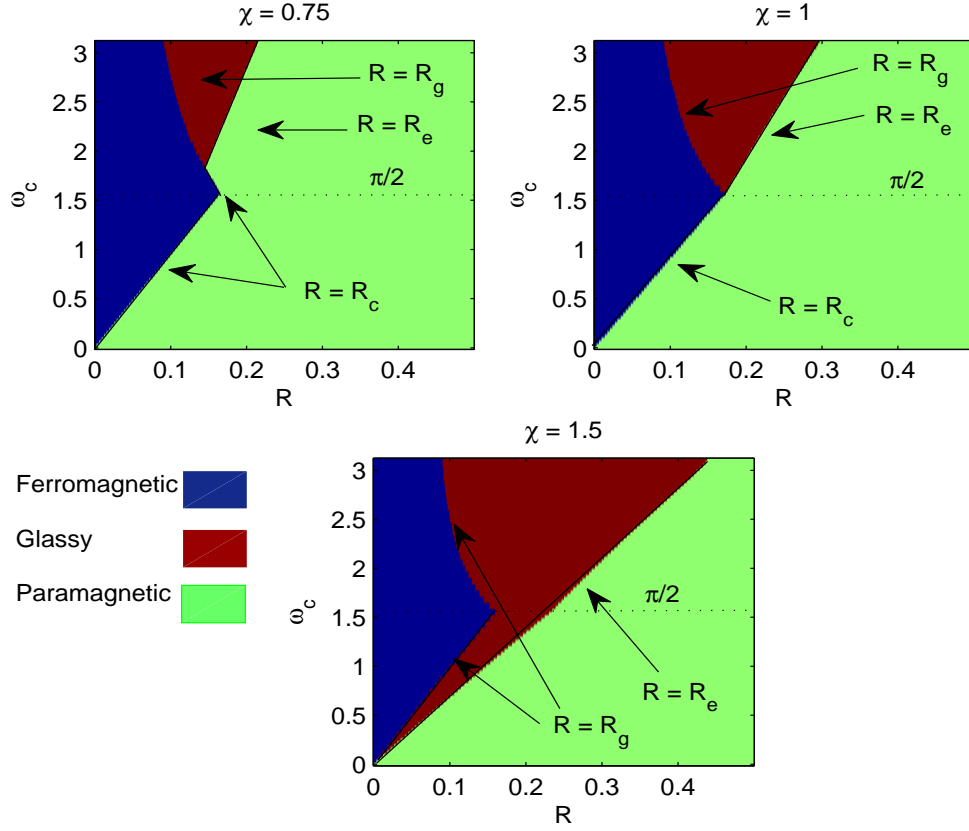
Fig. 1. Example 1: Phase diagram in the plane of R vs. ω_c with various gain values. The arrows are directed towards the boundaries of the various phase transitions.

Example 2 Let $\mathbf{H}(\omega)$ be a multiband filter given by

$$\mathbf{H}(\omega) = \begin{cases} 1, & \left|\omega \pm \frac{3\pi}{8}\right| \leq \frac{\pi}{8} \text{ or } \left|\omega \pm \frac{7\pi}{8}\right| \leq \frac{\pi}{8} \\ 0, & \text{else} \end{cases}, \tag{55}$$

and let the mismatched filter be given by a band-pass filter

$$\mathbf{H}'(\omega) = \begin{cases} 1, & \omega_L \leq |\omega| \leq \omega_R \\ 0, & \text{else} \end{cases}, \tag{56}$$

with constant bandwidth, $\omega_R - \omega_L = \pi/8$, i.e., smaller than the real one. In the numerical calculations, we again chose $\beta = P_x = 1$. Figures 3 and 4 show, respectively, the phase diagram and the MSE as functions of R and ω_L. First, observe that for $\omega_R < \pi/4$, which means that $\mathbf{H}'(\omega)$ and $\mathbf{H}(\omega)$
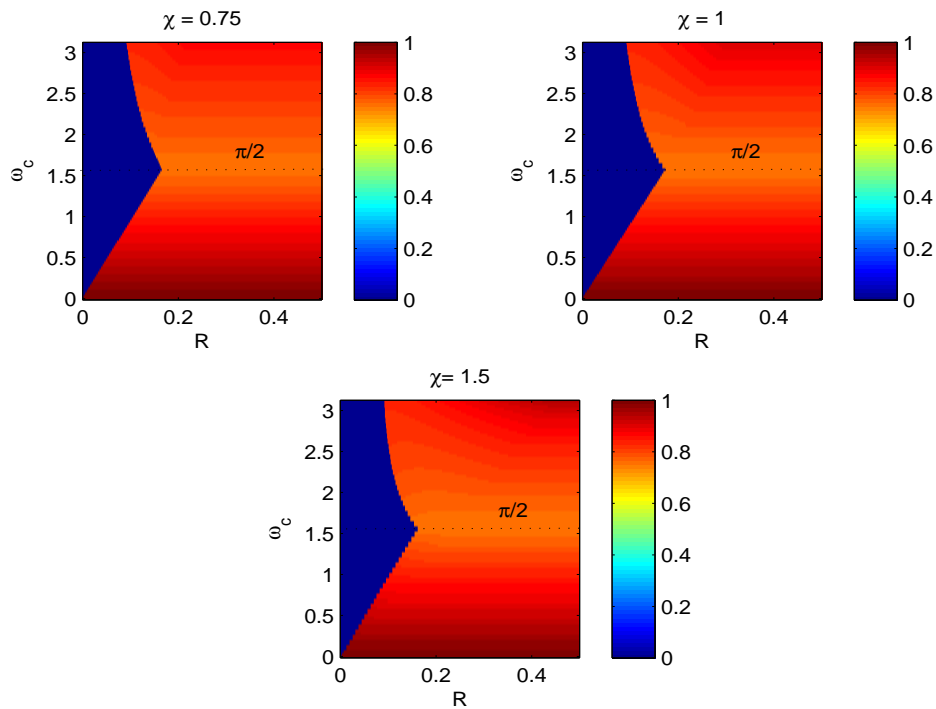
Fig. 2. Example 1: Mismatched MSE as a function of R and ω_c with various gain values.

are equal to one over non intersecting frequency ranges, there is no ferromagnetic phase, as expected. Accordingly, for $\omega_R > \pi/4$, the ferromagnetic phase begins to play a role, and it can be seen that for $\pi/4 + \pi/8 < \omega_R < \pi/2$, which means maximal intersection between the two filters, the range of rates for which the ferromagnetic phase dominates the partition function is maximal. Since the matched filter has two bands, obviously, the same behavior appears also in the second band. Thus, in this example, we actually obtain two disjoint glassy (and ferromagnetic) regions, which correspond to the two bands of the matched filter. Also, as shown in Fig. 4, in the ranges where no ferromagnetic phase exists, the MSE within the paramagnetic phase is larger than the MSE within the regions where ferromagnetic phase does exists, as one would expect.

Remark 3 Example 2 actually demonstrates that there can be arbitrarily many phase transitions. Generally speaking, for a matched multiband filter with N disjoint bands, and a mismatched bandpass filter (with small enough bandwidth), there are N disjoint glassy and ferromagnetic phases.
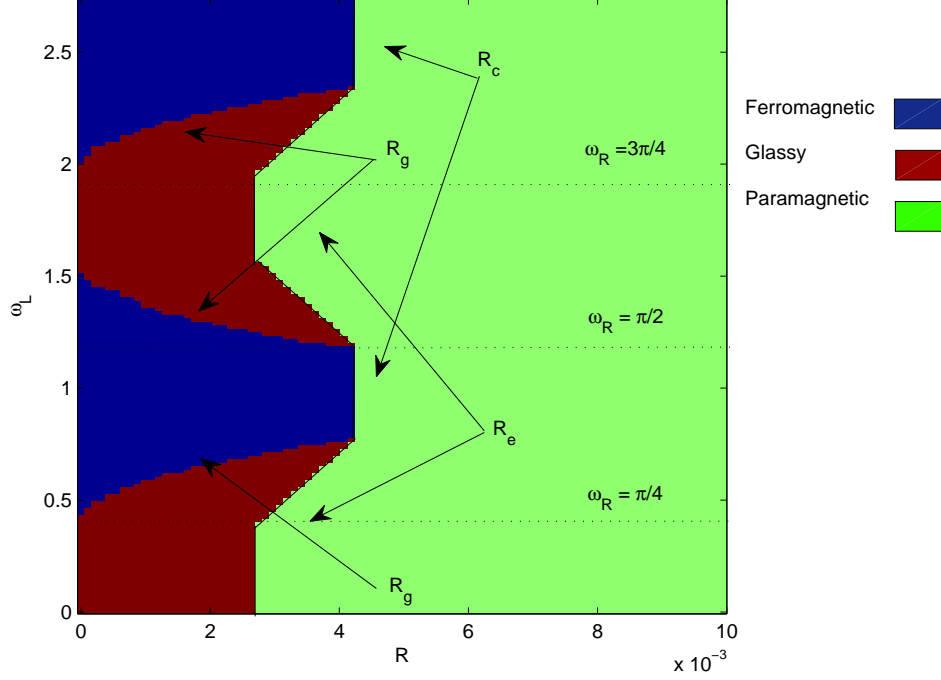
Fig. 3.   Example 2: Phase diagram in the plane of R vs. ω_L.

Example 3 In this example, we consider more realistic filters. Let $\mathbf{H}(z)$ denote a Type-II FIR filter given by (in the $\mathcal{Z}$ domain)

$$\mathbf{H}(z) = \left(1 - e^{j0.8\pi}z^{-1}\right)^2 \left(1 - e^{-j0.8\pi}z^{-1}\right)^2, \tag{57}$$

and let the mismatched filter be given is

$$\mathbf{H}'(z) = \left(1 - z_0 z^{-1}\right) \left(1 - z_0^* z^{-1}\right) \left(1 - e^{j0.8\pi}z^{-1}\right) \left(1 - e^{-j0.8\pi}z^{-1}\right) \tag{58}$$

where z_0 is a mismatched zero. In the numerical calculations, we chose again $\beta = P_x = 1$. Fig. 5 shows the amplitude response of the real and the mismatched filters for various angular frequencies defined as $\phi \overset{\triangle}{=} \arg(z_0)$. Figures 6 and 7 show, respectively, the phase diagram and the MSE as functions of R and ϕ. In this example, the roles of the differences between the true and mismatched filters, are emphasized. Starting with the obvious, observe that the maximal range of rates for which the ferromagnetic region dominates the partition function occurs at $\phi = 0.8\pi$, as expected. Less trivially, for angular frequencies within the range $[0.2\pi, 0.25\pi]$, the ferromagnetic region is negligible. Looking at Fig. 5, it can be seen that within this range of angular frequencies, the true and the mismatched filters are "almost orthogonal"
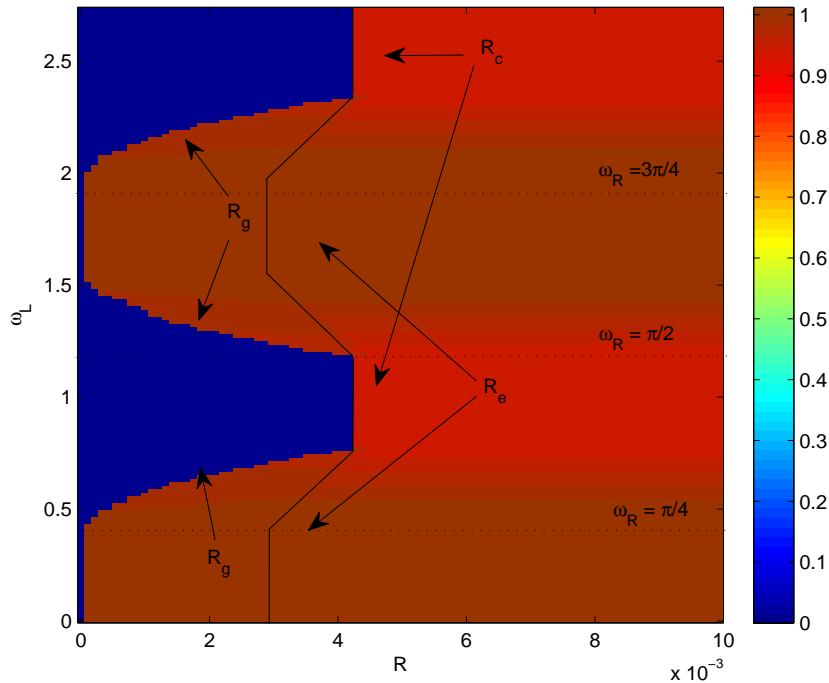
Fig. 4.   Example 2: Mismatched MSE as a function of R and ω_L.

in the L_2 sense, namely, their inner product is almost zero. Accordingly, using the methods in Section VI, it can be easily shown that for orthogonal filters we have that $R_g = 0$, namely, no ferromagnetic region exists (note that in this example, R_g is never equal to zero since the filters are never orthogonal). Finally, for angular frequencies within the range $[0, 0.2\pi]$, the ferromagnetic region returns to play a role. Indeed, Fig. 5 shows that, within this range, the matched and the mismatched filters "share" more similarities (in the sense of larger inner product).

Example 4 Let $\mathbf{H}(z)$ be given by

$$\mathbf{H}(z) = z - 2\cos(0.8\pi) + z^{-1}$$
$$= z \cdot \left(1 - e^{j0.8\pi} z^{-1}\right)\left(1 - e^{-j0.8\pi} z^{-1}\right) \tag{59}$$

and let the mismatched filter be given as

$$\mathbf{H}'(z) = \mathbf{H}(z) z^{-d} \tag{60}$$

where $d \in \mathbb{Z}$ is a mismatched delay. As before, in the numerical calculations, we chose $\beta = P_x = 1$.
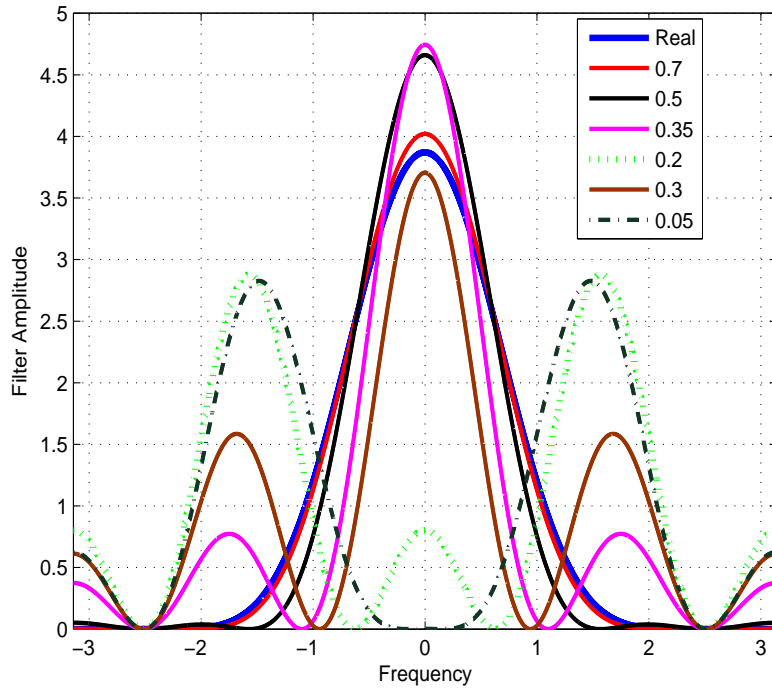
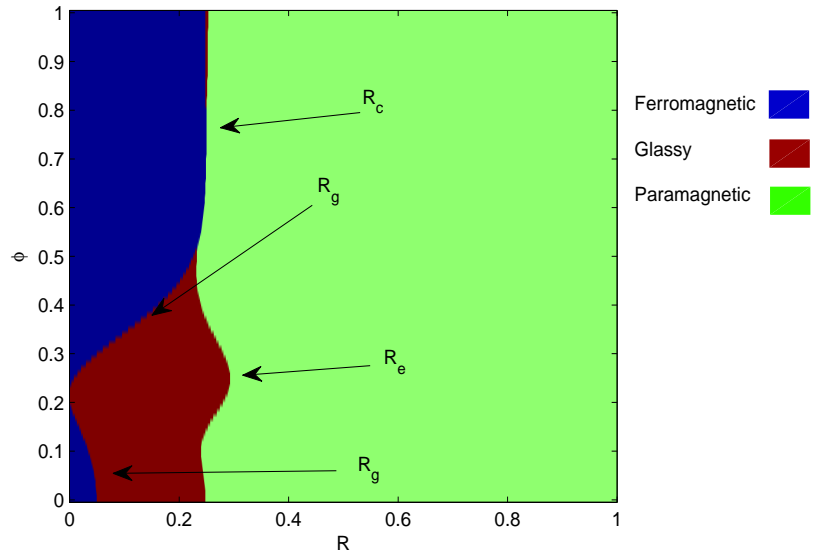Fig. 5.   Example 3: Amplitude of the real filter and mismatch filters for several phases.

Fig. 6.   Example 3: Phase diagram in the plane of R vs. ϕ.
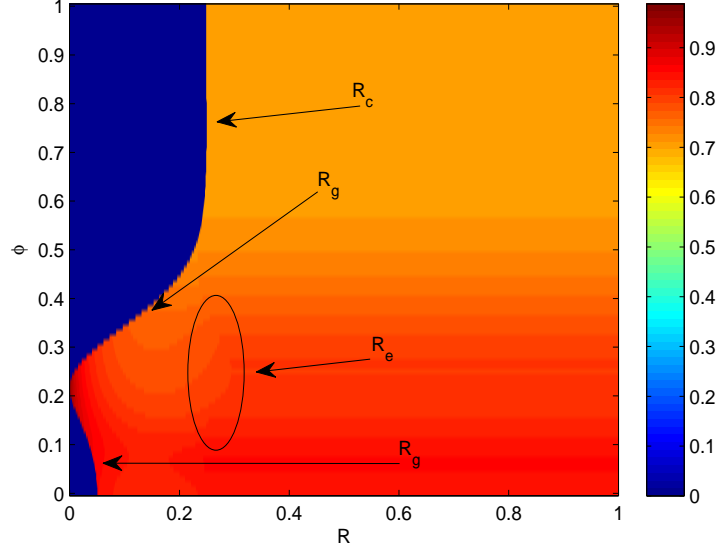
Fig. 7.   Example 3: Mismatched MSE as a function of R and ϕ.

Figures 8 and 9 show, respectively, the phase diagram and the MSE as functions of R and d. First, we see that R_e is constant, approximately equal to 0.29, which makes sense since R_e is given by

$$R_e = \frac{1}{4\pi} \int_0^{2\pi} \ln\left[P_x \left(\gamma_0 + \left| \mathbf{H}'(\omega) \right|^2 \beta \right) \right] \mathrm{d}\omega, \tag{61}$$

and thus independent of the delay (note that according to (11) γ_0 is also independent of the delay). Next, let us take a look at R_d given in (45)

$$\begin{aligned}
R_d =& \frac{1}{2} + \beta P_x \frac{1}{2\pi} \int_0^{2\pi} \mathrm{Re}\left(\mathbf{H}'^*(\omega) \mathbf{H}(\omega) \right) \mathrm{d}\omega \\
&+ \frac{1}{4\pi} \int_0^{2\pi} \left| \mathbf{H}'(\omega) \right|^2 \beta \left(\frac{\left| \mathbf{H}'(\omega) \right|^2 \beta \left(2 + \beta \left| \mathbf{H}(\omega) \right|^2 P_x \right) + \gamma_0}{\left(\left| \mathbf{H}'(\omega) \right|^2 \beta + \gamma_0 \right)^2} - P_x \right) \\
&- \frac{1}{4\pi} \int_0^{2\pi} \frac{\left| \mathbf{H}'(\omega) \right|^2 \beta \left(3 + 2 P_x \beta \left| \mathbf{H}(\omega) \right|^2 \right) + \gamma_0}{\left(\left| \mathbf{H}'(\omega) \right|^2 \beta + \gamma_0 \right)}.
\end{aligned} \tag{62}$$

In contrast to R_e, R_d does depend on the delay via the second term, which in the case considered takes the form $\mathrm{Re}\left(\mathbf{H}'^*(\omega) \mathbf{H}(\omega) \right) = \left| \mathbf{H}(\omega) \right|^2 \cos(\omega d)$. Actually, in the settings considered, it is easy to show that $\gamma_0 = 1/P_x = 1$, thus obtaining

$$R_d = \frac{1}{2\pi} \int_0^{2\pi} \mathrm{Re}\left(\mathbf{H}'^*(\omega) \mathbf{H}(\omega) \right) \mathrm{d}\omega - \frac{1}{2\pi} \int_0^{2\pi} \left| \mathbf{H}'(\omega) \right|^2 \mathrm{d}\omega \tag{63}$$

$$= \frac{1}{2\pi} \int_0^{2\pi} \left| \mathbf{H}' \left(\omega \right) \right|^2 \cos \left(\omega d \right) \mathrm{d}\omega - \frac{1}{2\pi} \int_0^{2\pi} \left| \mathbf{H}' \left(\omega \right) \right|^2 \mathrm{d}\omega \tag{64}$$

$$= \frac{1}{2\pi} \int_0^{2\pi} \left| \mathbf{H}' \left(\omega \right) \right|^2 \left[\cos \left(\omega d \right) - 1 \right] \mathrm{d}\omega \leq 0. \tag{65}$$

Therefore, we obtain that R_d is non-positive, and hence for all ϕ (except the trivial case of $\phi = 0$) there is a glassy phase. This result is consistent with Figures 8 and 9. More importantly, it can be observed that the MSE vanishes (or equivalently, the ferromagnetic phase dominates the partition function) only in case $d = 0$, namely, zero delay. This is a reasonable result, as a delay of one sample (linear phase) is enough to cause a serious degradation in the MSE. Actually, for any fixed rate the error is constant, independently of the delay, as one would expect. Finally, note that the MSE is larger in the glassy region than in the paramagnetic region[3]. This is also a reasonable result: As the rate increases, and hence more codewords are possible, since the MSE estimator is actually a weighted average (w.r.t. the posterior) over the codewords, the MSE can only decrease (each codeword in the codebook contributes approximately the same estimation error). Accordingly, for small codebooks (low rates) the MSE is larger, since the averaging is performed over "fewer" codewords.

V. PROOF OUTLINE AND TOOLS

A. Proof Outline

In this section, before getting deep into the proof of Theorem 2, we discuss the techniques and the main steps which will be used in Section VI. Generally speaking, the evaluation of the mismatched partition function, $Z' \left(\boldsymbol{y}, \boldsymbol{\lambda} \right)$, for a typical $\boldsymbol{y}$, essentially boils down to the evaluation of the exponential order of

$$\Pr \left\{ \frac{1}{2} \left\| \boldsymbol{y} - \boldsymbol{A}' \boldsymbol{X}_1 \right\|^2 - \frac{\boldsymbol{\lambda}^T \boldsymbol{X}_1}{\beta} \approx n\epsilon \right\} \tag{66}$$

for every value of ϵ in some range. In case that $\boldsymbol{A}' = \boldsymbol{I}$ [5, 6], this probability can be calculated fairly easily. Indeed, in this case, the above probability is equivalent to calculating the probability that a randomly chosen vector $\boldsymbol{X}$ on the n-dimensional hypersphere shell would have an empirical correlation coefficient ρ (induced by the constraint $\left\| \boldsymbol{y} - \boldsymbol{x} \right\|^2 / 2 - \boldsymbol{\lambda}^T \boldsymbol{x} / \beta \approx n\epsilon$) with a given vector $\boldsymbol{y}' = \boldsymbol{y} + \boldsymbol{\lambda}/\beta$. Geometrically, this probability is actually the probability that $\boldsymbol{X}$ falls within a cone of half angle $\arccos \left(\rho \right)$ around $\boldsymbol{y}'$ (for more details, see [34, 35]). However, in our case, because of the "interactions"[4] between

---

[3]Note that the MSE, in contrast to the MMSE, must not be monotonically increasing as a function of the rate.

[4]In the considered settings, the posterior, is proportional to $\exp \left\{ -\beta \left\| \boldsymbol{y} - \boldsymbol{A}' \boldsymbol{x} \right\|^2 / 2 \right\}$, and after expansion of the norm, the exponent includes an "external-field term," proportional to $\boldsymbol{y}^H \boldsymbol{A}' \boldsymbol{x}$, and a "pairwise spin-spin interaction term," proportional to $\left\| \boldsymbol{A}' \boldsymbol{x} \right\|^2$.
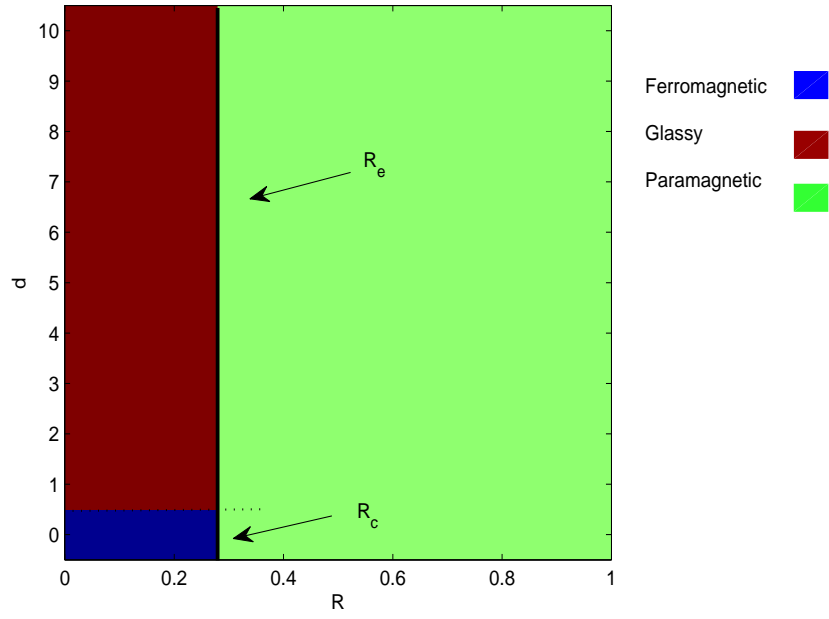
Fig. 8. Example 4: Phase diagram in the plane of R vs. d.
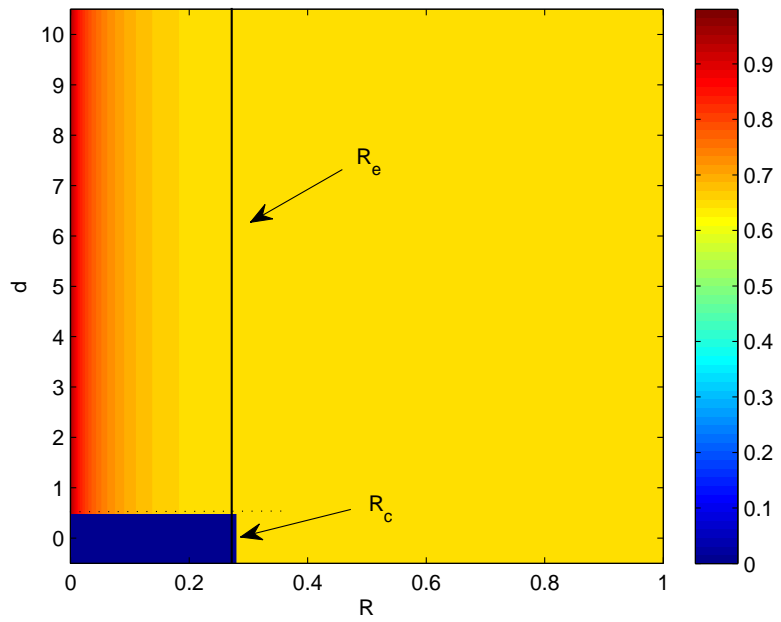
Fig. 9. Example 4: Mismatched MSE as a function of R and d.

different components of $\boldsymbol{X}$, which are induced by $\boldsymbol{A}'$, the methods in the aforementioned papers are not directly applicable. In our case, the purpose is to estimate the probability that a randomly chosen vector $\boldsymbol{X}$ on the n-dimensional hypersphere shell would fall within the intersection of this hypersphere and the n-dimensional hyperellipsoid (which is induced by the event in (66)). All our attempts to approach this calculation using the "geometric" route have failed. Thus, we will use a different route.

The main idea in our approach is, to "eliminate" the interactions between the different components of $\boldsymbol{X}$, by passing to the frequency domain. Since $\boldsymbol{A}'$ is a Toeplitz matrix, according to Szegö's theorem [36-39], it is asymptotically diagonalized by the discrete Fourier transform (DFT) matrix (if $\boldsymbol{A}'$ is a circulant matrix then the DFT matrix exactly diagonalizes it). Thus, multiplying both sides of (4) by the DFT matrix, $\boldsymbol{F}^H = \left\{ e^{-j2\pi ml/n}/\sqrt{n} \right\}_{m,l=0}^{n-1}$, we "asymptotically"[5] have that

$$\tilde{\boldsymbol{Y}} = \boldsymbol{\Sigma}\tilde{\boldsymbol{X}} + \tilde{\boldsymbol{N}} \tag{67}$$

where $\boldsymbol{\Sigma} \triangleq \mathrm{diag}\,(\sigma_1, \ldots, \sigma_n)$, $\tilde{\boldsymbol{X}} \triangleq \boldsymbol{F}^H \boldsymbol{X}$, $\tilde{\boldsymbol{Y}} \triangleq \boldsymbol{F}^H \boldsymbol{Y}$ and $\tilde{\boldsymbol{N}} \triangleq \boldsymbol{F}^H \boldsymbol{N}$. Accordingly, we evaluate (66), using

$$\mathrm{Pr}\left\{ \frac{1}{2} \left\| \tilde{\boldsymbol{y}} - \boldsymbol{\Sigma}\tilde{\boldsymbol{X}}_1 \right\|^2 - \frac{\tilde{\boldsymbol{\lambda}}^T \tilde{\boldsymbol{X}}_1}{\beta} \approx n\epsilon \right\} \tag{68}$$

where $\tilde{\boldsymbol{\lambda}} = \boldsymbol{F}^T \boldsymbol{\lambda}$. Now, in order to evaluate (68), it is desirable to estimate the volume[6] of the following set: For a given pair of vectors $(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}})$ and $\delta > 0$, we define the conditional δ-type of $\tilde{\boldsymbol{x}}$ given $\tilde{\boldsymbol{y}}$ as

$$\mathcal{T}_\delta\,(\tilde{\boldsymbol{x}} \mid \tilde{\boldsymbol{y}}) \triangleq \left\{ \tilde{\boldsymbol{x}} \in \mathbb{R}^n : \; \left| \|\tilde{\boldsymbol{x}}\|^2 - nP_x \right| \leq \delta, \; \left| \frac{\|\tilde{\boldsymbol{y}} - \boldsymbol{\Sigma}\tilde{\boldsymbol{x}}\|^2}{2} - \frac{\tilde{\boldsymbol{\lambda}}^T \tilde{\boldsymbol{x}}}{\beta} - n\epsilon \right| \leq \delta \right\}. \tag{69}$$

This set is regarded as a conditional type of (wrong) codewords $\tilde{\boldsymbol{x}}$ given $\tilde{\boldsymbol{y}}$ as it contains all vectors which, within δ, have the same energy related to the partition function (8). After calculating the volume of (69), the probability in (68) can then be easily estimated. However, as was previously mentioned, calculating the volume of such a set is a tedious task when approaching it directly. We will use instead the following relaxation. We start with partitioning the components of $\tilde{\boldsymbol{x}}$ into k *bins*, each of dimension n_b, such that $k = n/n_b$, and we approximate the eigenvalues, which are the diagonal elements of $\boldsymbol{\Sigma}$, to be piecewise constant over these bins. This partition literally means that we transform the original model

---

[5]Rigorously, in the proof, we first assume that $\boldsymbol{A}'$ is a circulant matrix, and thus (67) is exact for any n. Then, when taking the limit $n \to \infty$, using Szegö's theorem, this assumption will be dropped. Finally, note that the assumption of the square summability of the generating sequence $\{h_k\}$ in the theorems presented earlier, is made in order to use Szegö's theorem.

[6]Recall that the volume of a set $\mathcal{A} \subset \mathbb{R}^n$ is defined as $\mathrm{Vol}\,\{\mathcal{A}\} \triangleq \int_{\mathcal{A}} \mathrm{d}\boldsymbol{x}$.

in (67) into k subchannels, each having the form

$$Y_{i,r} = \sigma_r X_i + N_i, \quad i = (r-1) n_b + 1, \ldots, r \cdot n_b, \tag{70}$$

for $r = 1, \ldots, k$. With this partitioning in mind, at the final stage of the analysis (after taking the limit $n \to \infty$), we take the limit $k \to \infty$. This partitioning will enable to calculate the desired volume. Then, using large deviations considerations, the mismatched partition function will be obtained. Finally, in order to derive the MSE, we will use the tools of [5], which are briefly presented in the following subsection.

B. Optimum Estimation Relations - Background and Summary

1) Matched Case: Let $\boldsymbol{X} = (X_1, \ldots, X_n)$ and $\boldsymbol{Y} = (Y_1, \ldots, Y_m)$ be two random vectors, jointly distributed according to a given probability function $P(\boldsymbol{x}, \boldsymbol{y})$. The conditional mean estimator of $\boldsymbol{X}$ based on $\boldsymbol{Y}$, i.e., $\hat{\boldsymbol{X}} = \boldsymbol{E}\{\boldsymbol{X} \mid \boldsymbol{Y}\}$ is well known to minimize the MSE $\boldsymbol{E}\left(X_i - \hat{X}_i\right)^2$ for all $i = 1, \ldots, n$. Accordingly, the MMSE in estimating X_i equals to $\boldsymbol{E}\left\{(X_i - \boldsymbol{E}\{X_i \mid \boldsymbol{Y}\})^2\right\}$, i.e., the expected conditional variance of X_i given $\boldsymbol{Y}$. More generally, the MMSE error covariance matrix is an $n \times n$ matrix whose (i,j)-th element is given by $\boldsymbol{E}\{(X_i - \boldsymbol{E}\{X_i \mid \boldsymbol{Y}\})(X_j - \boldsymbol{E}\{X_j \mid \boldsymbol{Y}\})\}$. This matrix can be represented as the expectation (w.r.t. $\boldsymbol{Y}$) of the conditional covariance matrix of $\boldsymbol{X}$ given $\boldsymbol{Y}$, henceforth denoted by $\mathrm{cov}(\boldsymbol{X} \mid \boldsymbol{Y})$. In particular, using the orthogonality principle, the MMSE error covariance matrix is given by

$$\boldsymbol{E}\{\mathrm{cov}(\boldsymbol{X} \mid \boldsymbol{Y})\} = \boldsymbol{E}\{\boldsymbol{X}\boldsymbol{X}^T\} - \boldsymbol{E}\{\boldsymbol{E}\{\boldsymbol{X} \mid \boldsymbol{Y}\}\boldsymbol{E}\{\boldsymbol{X}^T \mid \boldsymbol{Y}\}\}. \tag{71}$$

Based on Definition 1, the following relations readily follow

$$\boldsymbol{E}\{\boldsymbol{X} \mid \boldsymbol{Y} = \boldsymbol{y}\} = \nabla_0 g(\boldsymbol{\lambda}) \ln Z(\boldsymbol{y}, \boldsymbol{\lambda}) \tag{72}$$

$$\boldsymbol{E}\{\mathrm{cov}(\boldsymbol{X} \mid \boldsymbol{Y})\} = \boldsymbol{E}\{\nabla_0^2 \ln Z(\boldsymbol{Y}, \boldsymbol{\lambda})\} \tag{73}$$

where for a generic function g, we use $\nabla_0 g(\boldsymbol{\lambda})$ and $\nabla_0^2 g(\boldsymbol{\lambda})$ to designate $\nabla_{\boldsymbol{\lambda}} g(\boldsymbol{\lambda})|_{\boldsymbol{\lambda}=0}$ and $\nabla_{\boldsymbol{\lambda}}^2 g(\boldsymbol{\lambda})|_{\boldsymbol{\lambda}=0}$, respectively, and $\nabla_{\boldsymbol{\lambda}}$ and $\nabla_{\boldsymbol{\lambda}}^2$ denote the gradient and Hessian operators w.r.t. $\boldsymbol{\lambda}$, respectively. Finally, it is easy to verify that the following relation holds

$$\boldsymbol{E}\{\mathrm{cov}(\boldsymbol{X} \mid \boldsymbol{Y})\} = \boldsymbol{E}\{\boldsymbol{X}\boldsymbol{X}^T\} - \boldsymbol{E}\left\{[\nabla_0 \ln Z(\boldsymbol{Y}, \boldsymbol{\lambda})][\nabla_0 \ln Z(\boldsymbol{Y}, \boldsymbol{\lambda})]^T\right\}, \tag{74}$$

and upon taking the trace of the above equation one obtains

$$\mathrm{mmse}(\boldsymbol{X} \mid \boldsymbol{Y}) \stackrel{\triangle}{=} \sum_{i=1}^{n} \boldsymbol{E}\left\{(X_i - \boldsymbol{E}\{X_i \mid \boldsymbol{Y}\})^2\right\}$$

$$= \sum_{i=1}^{n} \left[\boldsymbol{E}\left\{X_i^2\right\} - \boldsymbol{E}\left\{ \left. \left[\frac{\partial \ln Z\left(\boldsymbol{Y}, \boldsymbol{\lambda}\right)}{\partial \lambda_i} \right]^2 \right|_{\boldsymbol{\lambda}=0} \right\} \right]. \tag{75}$$

Further relations between information measures and estimation quantities can be found in [5, 6].

2) Mismatched Case: Consider a mismatched estimator which is the conditional mean of $\boldsymbol{X}$ given $\boldsymbol{Y}$, based on an incorrect joint distribution $P'\left(\boldsymbol{x}, \boldsymbol{y}\right)$, whereas the true joint distribution continues to be $P\left(\boldsymbol{x}, \boldsymbol{y}\right)$. Then, the following relation holds

$$\boldsymbol{E}\left\{\text{cov}'\left(\boldsymbol{X} \mid \boldsymbol{Y}\right)\right\} \triangleq \boldsymbol{E}\left\{ \left(\boldsymbol{X} - \boldsymbol{E}'\left\{\boldsymbol{X} \mid \boldsymbol{Y}\right\}\right)\left(\boldsymbol{X} - \boldsymbol{E}'\left\{\boldsymbol{X} \mid \boldsymbol{Y}\right\}\right)^T \right\}$$

$$= \boldsymbol{E}\left\{\boldsymbol{X}\boldsymbol{X}^T\right\} - \boldsymbol{E}_P\left\{\boldsymbol{E}\left\{\boldsymbol{X} \mid \boldsymbol{Y}\right\} \boldsymbol{E}'\left\{\boldsymbol{X}^T \mid \boldsymbol{Y}\right\}\right\}$$

$$- \boldsymbol{E}\left\{\boldsymbol{E}'\left\{\boldsymbol{X} \mid \boldsymbol{Y}\right\} \boldsymbol{E}\left\{\boldsymbol{X}^T \mid \boldsymbol{Y}\right\}\right\}$$

$$+ \boldsymbol{E}\left\{\boldsymbol{E}'\left\{\boldsymbol{X} \mid \boldsymbol{Y}\right\} \boldsymbol{E}'\left\{\boldsymbol{X}^T \mid \boldsymbol{Y}\right\}\right\}, \tag{76}$$

where $\text{cov}'\left(\boldsymbol{X} \mid \boldsymbol{Y}\right) \triangleq \left(\boldsymbol{X} - \boldsymbol{E}'\left\{\boldsymbol{X} \mid \boldsymbol{Y}\right\}\right)\left(\boldsymbol{X} - \boldsymbol{E}'\left\{\boldsymbol{X} \mid \boldsymbol{Y}\right\}\right)^T$. Upon taking the trace of (76), one obtains

$$\text{mse}\left(\boldsymbol{X} \mid \boldsymbol{Y}\right) \triangleq \sum_{i=1}^{n} \boldsymbol{E}\left\{ \left(X_i - \boldsymbol{E}'\left\{X_i \mid \boldsymbol{Y}\right\}\right)^2 \right\}$$

$$= \sum_{i=1}^{n} \left[\boldsymbol{E}\left\{X_i^2\right\} - 2\boldsymbol{E}\left\{ \left. \frac{\partial \ln Z\left(\boldsymbol{Y}, \boldsymbol{\lambda}\right)}{\partial \lambda_i} \right|_{\boldsymbol{\lambda}=0} \cdot \left. \frac{\partial \ln Z'\left(\boldsymbol{Y}, \boldsymbol{\lambda}\right)}{\partial \lambda_i} \right|_{\boldsymbol{\lambda}=0} \right\} \right.$$

$$\left. + \boldsymbol{E}\left\{ \left. \left[\frac{\partial \ln Z'\left(\boldsymbol{Y}, \boldsymbol{\lambda}\right)}{\partial \lambda_i} \right]^2 \right|_{\boldsymbol{\lambda}=0} \right\} \right]. \tag{77}$$

VI. PROOF OF THEOREM 2

For a given $\boldsymbol{y}$, the mismatched partition function is given by[7]

$$Z'\left(\boldsymbol{y}, \boldsymbol{\lambda}\right) = \sum_{\boldsymbol{x} \in \mathcal{C}} e^{-nR} \exp\left[-\beta \left\|\boldsymbol{y} - \boldsymbol{A}'\boldsymbol{x}\right\|^2 / 2 + \boldsymbol{\lambda}^T\boldsymbol{x} \right] \tag{78}$$

$$= e^{-nR} \exp\left[-\beta \left\|\boldsymbol{y} - \boldsymbol{A}'\boldsymbol{x}_0\right\|^2 / 2 + \boldsymbol{\lambda}^T\boldsymbol{x}_0 \right] \tag{79}$$

$$+ \sum_{\boldsymbol{x} \in \mathcal{C}\backslash\{\boldsymbol{x}_0\}} e^{-nR} \exp\left[-\beta \left\|\boldsymbol{y} - \boldsymbol{A}'\boldsymbol{x}\right\|^2 / 2 + \boldsymbol{\lambda}^T\boldsymbol{x} \right] \tag{80}$$

$$\triangleq Z_c'\left(\boldsymbol{y}, \boldsymbol{\lambda}\right) + Z_e'\left(\boldsymbol{y}, \boldsymbol{\lambda}\right) \tag{81}$$

---

[7]Note that there should be a normalization factor of $(2\pi/\beta)^{-n/2}$ in (78). Nonetheless, since this constant is independent of $\boldsymbol{\lambda}$, it has no effect on the MSE (which is obtained by the gradient of $\ln Z'\left(\boldsymbol{y}, \boldsymbol{\lambda}\right)$ w.r.t. $\boldsymbol{\lambda}$). Hence, for simplicity of notation, it is omitted.

where without loss of generality, the transmitted codeword is assumed to be $\boldsymbol{x}_0$, and $Z'_c(\boldsymbol{y}, \boldsymbol{\lambda})$ and $Z'_e(\boldsymbol{y}, \boldsymbol{\lambda})$ are the partial partition functions induced by the correct codeword and the wrong codewords, respectively. By the law of large numbers (LLN), $\|\boldsymbol{y} - \boldsymbol{A}'\boldsymbol{x}_0\|^2 \approx \|(\boldsymbol{A} - \boldsymbol{A}')\boldsymbol{x}_0\|^2 + n/\beta$, and therefore, with high probability

$$Z'_c(\boldsymbol{y}, \boldsymbol{\lambda}) \doteq e^{-nR} \exp\left\{-\frac{\beta}{2}\left[\|(\boldsymbol{A} - \boldsymbol{A}')\boldsymbol{x}_0\|^2 + \frac{n}{\beta}\right] + \boldsymbol{\lambda}^T\boldsymbol{x}_0\right\} \tag{82}$$

$$= \exp\left\{-n\left(R + \frac{1}{2} + \frac{\beta\|(\boldsymbol{A} - \boldsymbol{A}')\boldsymbol{x}_0\|^2}{2n}\right) + \boldsymbol{\lambda}^T\boldsymbol{x}_0\right\}. \tag{83}$$

More precisely, for any $\epsilon > 0$,

$$\exp\left\{-n\left(R + \frac{1}{2} + \frac{\beta\|(\boldsymbol{A} - \boldsymbol{A}')\boldsymbol{x}_0\|^2}{2n} + \epsilon\right) + \boldsymbol{\lambda}^T\boldsymbol{x}_0\right\} \le Z'_c(\boldsymbol{y}, \boldsymbol{\lambda})$$

$$\le \exp\left\{-n\left(R + \frac{1}{2} + \frac{\beta\|(\boldsymbol{A} - \boldsymbol{A}')\boldsymbol{x}_0\|^2}{2n} - \epsilon\right) + \boldsymbol{\lambda}^T\boldsymbol{x}_0\right\} \tag{84}$$

with probability tending to one as $n \to \infty$. As for $Z'_e(\boldsymbol{y}, \boldsymbol{\lambda})$, we have

$$Z'_e(\boldsymbol{y}, \boldsymbol{\lambda}) = e^{-nR}\int_{\mathbb{R}} \mathcal{N}(\epsilon)\, e^{-n\beta\epsilon}\mathrm{d}\epsilon \tag{85}$$

where

$$\mathcal{N}(\epsilon) \triangleq \sum_{i=1}^{M-1} \mathbb{1}\left\{\boldsymbol{x}_i : \frac{\|\boldsymbol{y} - \boldsymbol{A}'\boldsymbol{x}_i\|^2}{2} - \frac{\boldsymbol{\lambda}^T\boldsymbol{x}_i}{\beta} \approx n\epsilon\right\}, \tag{86}$$

to wit, $\mathcal{N}(\epsilon)$ is the number of codewords $\{\boldsymbol{x}_i\}$ in $\mathcal{C} \setminus \{\boldsymbol{x}_0\}$ for which $\|\boldsymbol{y} - \boldsymbol{A}'\boldsymbol{x}_i\|^2/2 - \boldsymbol{\lambda}^T\boldsymbol{x}_i/\beta \approx n\epsilon$, namely, between $n\epsilon$ and $n(\epsilon + \mathrm{d}\epsilon)$. We proceed in two steps: First, the typical exponential order of $\mathcal{N}(\epsilon)$ is computed, and then (85) is calculated.

Step 1: Given $\boldsymbol{y}$, $\mathcal{N}(\epsilon)$ is a sum of $(M-1)$ i.i.d. Bernoulli random variables and therefore, its expected value is given by

$$\boldsymbol{E}\{\mathcal{N}(\epsilon)\} = \sum_{i=1}^{M-1} \Pr\left\{\frac{\|\boldsymbol{y} - \boldsymbol{A}'\boldsymbol{X}_i\|^2}{2} - \frac{\boldsymbol{\lambda}^T\boldsymbol{X}_i}{\beta} \approx n\epsilon\right\} \tag{87}$$

$$= (e^{nR} - 1) \cdot \Pr\left\{\frac{\|\boldsymbol{y} - \boldsymbol{A}'\boldsymbol{X}_1\|^2}{2} - \frac{\boldsymbol{\lambda}^T\boldsymbol{X}_1}{\beta} \approx n\epsilon\right\}. \tag{88}$$

Assuming that $\boldsymbol{A}'$ is a circulant matrix[8], it is known that the discrete Fourier transform (DFT) matrix diagonalizes it [36-39], and thus multiplying both sides of equation (4) by the DFT matrix, $\boldsymbol{F}^H$, one

---

[8]Recall that this assumption is only an intermediate step in the analysis, and will be dropped later on. Alternatively, instead of this assumption, one could use the spectral decomposition theorem, to find an orthonormal basis which diagonalizes the matrix $\boldsymbol{A}'$, and project (4) on this basis, to obtain the form of (89).

obtains

$$\tilde{Y} = \Sigma' \tilde{X} + \tilde{N} \tag{89}$$

where $\Sigma' \triangleq \mathrm{diag}\,(\sigma'_1, \ldots, \sigma'_n)$, $\tilde{X} \triangleq F^H X$, $\tilde{Y} \triangleq F^H Y$ and $\tilde{N} \triangleq F^H N$. Since a unitary operator is applied on X, then $\tilde{X}$ is still uniformly drawn on the n-hyperesphere with radius $\sqrt{nP_x}$ (as in the original setting). Similarly, $\tilde{N}$ has the same statistics as before, namely, its components are i.i.d. complex Gaussian random variables with zero mean and variance $1/\beta$. For simplicity of notation, in the following, the "tilde" sign over the various variables will be omitted, keeping the original notation. Therefore, instead of evaluating (88), the exponential order of

$$\Pr\left\{\frac{\|y - \Sigma' X_1\|^2}{2} - \frac{\lambda^T X_1}{\beta} \approx n\epsilon\right\}, \tag{90}$$

will be evaluated, where $F^T \lambda \mapsto \lambda^9$. For a given pair of vectors (x, y) and $\delta > 0$, define the conditional δ-type of x given y as

$$\mathcal{T}_\delta\,(x \mid y) \triangleq \left\{x \in \mathbb{R}^n : \ \left|\|x\|^2 - nP_x\right| \le \delta, \ \left|\frac{\|y - \Sigma' x\|^2}{2} - \frac{\lambda^T x}{\beta} - n\epsilon\right| \le \delta\right\}. \tag{91}$$

The following lemma is proved in Appendix A.

Lemma 1 Let k and n_b be natural numbers such that $k = n/n_b{}^{10}$. Define the sets $\mathcal{G}_{1,\delta} \triangleq \{\delta \cdot i : \ i = 0, 1, \ldots, \lceil kP_x/\delta \rceil\}$ and $\mathcal{G}_{2,\delta} \triangleq \{\delta \cdot i : \ i = -\lceil k/\delta \rceil, \ldots, -1, 0, 1, \ldots, \lceil k/\delta \rceil\}$. Also, let

$$\hat{\mathcal{T}}_\delta\,(x \mid y) \triangleq \bigcup_{\mathcal{P}^\delta \cap \mathcal{R}_{\mathcal{P}}^\delta} \overset{k}{\underset{m=1}{\bigtimes}} \mathscr{B}_m^\delta\,(P_m, \rho_m) \tag{92}$$

where $\times$ designates a Cartesian product, and

$$\mathscr{B}_m^\delta\,(P_m, \rho_m) \triangleq \left\{x \in \mathbb{R}^{n_b} : \ \left|\left\|x_{(m-1)n_b+1}^{mn_b}\right\|^2 - n_b P_m\right| \le \delta, \right.$$

$$\left. \left|\mathrm{Re}\left\{\sum_{i \in \mathcal{I}_m} \sigma'_i \bar{y}_i x_i\right\} - n_b \rho_m \sqrt{\tilde{P}_{y,m}\tilde{P}_{\sigma,m}}\right| \le \delta\right\} \tag{93}$$

---

[9] Note that $F^T \lambda$ may be a complex quantity (in contrast to λ). This fact will be taken into account later on.

[10] Without loss of generality, it is assumed that n_b (*bin* length) is a divisor of n, and that the k various bins have equal sizes.

where $\mathcal{I}_m \triangleq [(m-1)\,n_b + 1, mn_b]$, $\bar{y}_i \triangleq y_i^* + \frac{\lambda_i}{\beta \sigma_i'}$, $\tilde{P}_{y,m} \triangleq \frac{1}{n_b} \sum_{i \in \mathcal{I}_m} |\bar{y}_i|^2$, $\tilde{P}_{\sigma,m} \triangleq \frac{1}{n_b} \sum_{i \in \mathcal{I}_m} |\sigma_i' x_i|^2$, and[11]

$$\boldsymbol{\mathcal{P}}^\delta \triangleq \left\{ \boldsymbol{P} \in \mathcal{G}_{1,\delta}^k : \left| \frac{1}{k} \sum_{i=1}^k P_i - P_x \right| \leq \delta \right\} \tag{94}$$

$$\boldsymbol{\mathcal{R}}_{\boldsymbol{\mathcal{P}}}^\delta \triangleq \left\{ \boldsymbol{\rho} \in \mathcal{G}_{2,\delta}^k : \left| \frac{1}{k} \sum_{i=1}^k \rho_i \sqrt{\tilde{P}_{y,i} \tilde{P}_{\sigma,i}} - \tilde{\rho} \right| \leq \delta \right\} \tag{95}$$

where $\mathcal{G}_{1,\delta}^k$ and $\mathcal{G}_{2,\delta}^k$ are the kth Cartesian power of $\mathcal{G}_{1,\delta}$ and $\mathcal{G}_{2,\delta}$, respectively, and

$$\tilde{\rho} \triangleq \frac{\frac{1}{n} \sum_{i=1}^n |\sigma_i' x_i|^2 + P_y - 2\epsilon}{2} \tag{96}$$

where $P_y \triangleq \frac{1}{n} \sum_{i=1}^n |y_i|^2$. Then,

$$\hat{\mathcal{T}}_{\delta/k} (\boldsymbol{x} \mid \boldsymbol{y}) \subseteq \mathcal{T}_\delta (\boldsymbol{x} \mid \boldsymbol{y}) \subseteq \hat{\mathcal{T}}_\delta (\boldsymbol{x} \mid \boldsymbol{y}). \tag{97}$$

Next, the eigenvalues, $\{\sigma_i'\}_i$, are approximated to be piecewise constant over the various bins. At the final stage of the analysis (after taking the limit $n \to \infty$), we will take the limit $k \to \infty$ so that this approximation becomes superfluous. Accordingly, under this approximation, $\tilde{P}_{\sigma,m} = |\sigma_m'|^2 P_m$, and (with abuse of notation)

$$\hat{\mathcal{T}}_\delta^k (\boldsymbol{x} \mid \boldsymbol{y}) = \bigcup_{\boldsymbol{\mathcal{P}}^\delta \cap \boldsymbol{\mathcal{R}}_{\boldsymbol{\mathcal{P}}}^\delta} \overset{k}{\underset{m=1}{\times}} \mathscr{B}_m^\delta (P_m, \rho_m) \tag{98}$$

where now

$$\mathscr{B}_m^\delta (P_m, \rho_m) \triangleq \left\{ \boldsymbol{x} \in \mathbb{R}^{n_b} : \left| \left\| \boldsymbol{x}_{(m-1)n_b+1}^{mn_b} \right\|^2 - n_b P_m \right| \leq \delta, \right.$$
$$\left. \left| \mathrm{Re} \left\{ \sigma_m' \sum_{i \in \mathcal{I}_m} \bar{y}_i x_i \right\} - n_b \rho_m \sqrt{\tilde{P}_{y,m} |\sigma_m'|^2 P_m} \right| \leq \delta \right\} \tag{99}$$

and

$$\boldsymbol{\mathcal{R}}_{\boldsymbol{\mathcal{P}}}^\delta \triangleq \left\{ \boldsymbol{\rho} \in \mathcal{G}_{2,\delta}^k : \left| \frac{1}{k} \sum_{i=1}^k |\sigma_i'| \rho_i \sqrt{\tilde{P}_{y,i} P_i} - \tilde{\rho} \right| \leq \delta \right\}, \tag{100}$$

where

$$\tilde{\rho} \triangleq \frac{\frac{1}{k} \sum_{i=1}^k |\sigma_i'|^2 P_i + P_y - 2\epsilon}{2}. \tag{101}$$

---

[11]The purpose of the subscript symbol in $\boldsymbol{\mathcal{R}}_{\boldsymbol{\mathcal{P}}}^\delta$ is to emphasize the dependence of it on $\boldsymbol{\mathcal{P}}$. More precisely, these sets should be understood as joint-power-correlation allocations, which are "living" in the intersection $\boldsymbol{\mathcal{P}}^\delta \cap \boldsymbol{\mathcal{R}}_{\boldsymbol{\mathcal{P}}}^\delta$. Accordingly, P_m and ρ_m are the power and correlation constraints within the mth bin, respectively.

In the following, the volume of $\mathcal{T}_\delta(\boldsymbol{x} \mid \boldsymbol{y})$ is evaluated. On the one hand, using Lemma 1, one obtains that

$$\text{Vol}\left\{\mathcal{T}_\delta(\boldsymbol{x} \mid \boldsymbol{y})\right\} \leq \text{Vol}\left\{\hat{\mathcal{T}}_\delta^k(\boldsymbol{x} \mid \boldsymbol{y})\right\} \tag{102}$$

$$\leq \sum_{\boldsymbol{\mathcal{P}}^\delta \cap \boldsymbol{\mathcal{R}}_{\boldsymbol{\mathcal{P}}}^\delta} \text{Vol}\left\{\underset{m=1}{\overset{k}{\times}} \mathscr{B}_m^\delta(P_m, \rho_m)\right\} \tag{103}$$

$$\leq N_{\delta,k} \cdot \max_{\boldsymbol{\mathcal{P}}^\delta \cap \boldsymbol{\mathcal{R}}_{\boldsymbol{\mathcal{P}}}^\delta} \text{Vol}\left\{\underset{m=1}{\overset{k}{\times}} \mathscr{B}_m^\delta(P_m, \rho_m)\right\} \tag{104}$$

where the second inequality follows for the union bound, and $N_{k,\delta}$ is a constant depending on k and δ (but not on n). This constant can be roughly bounded by

$$N_{\delta,k} \leq \left|\boldsymbol{\mathcal{P}}^\delta\right| \left|\boldsymbol{\mathcal{R}}_{\boldsymbol{\mathcal{P}}}^\delta\right| \leq \left(\frac{kP_x + 2\delta}{\delta}\right)^k \left(2 \cdot \frac{k+\delta}{\delta}\right)^k. \tag{105}$$

On the other hand,

$$\text{Vol}\left\{\mathcal{T}_\delta(\boldsymbol{x} \mid \boldsymbol{y})\right\} \geq \text{Vol}\left\{\hat{\mathcal{T}}_{\delta/k}^k(\boldsymbol{x} \mid \boldsymbol{y})\right\} \tag{106}$$

$$\geq \max_{\boldsymbol{\mathcal{P}}^{\delta/k} \cap \boldsymbol{\mathcal{R}}_{\boldsymbol{\mathcal{P}}}^{\delta/k}} \text{Vol}\left\{\underset{m=1}{\overset{k}{\times}} \mathscr{B}_m^{\delta/k}(P_m, \rho_m)\right\}. \tag{107}$$

The following lemma is proved in Appendix B.

Lemma 2 For every $m = 1, \ldots, k$ and $\nu > 0$,

$$(1 - \nu) \exp\left\{\frac{n_b}{2} \ln\left(\pi e \vartheta_{o,u}^2\right)\right\} \leq \text{Vol}\left\{\mathscr{B}_m^\delta(P_m, \rho_m)\right\} \leq \exp\left\{\frac{n_b}{2} \ln\left(\pi e \vartheta_o^2\right)\right\}. \tag{108}$$

where

$$\vartheta_{\delta,+}^2 = P_m + \delta - P_m(\rho_m - \delta)^2 \tag{109}$$

and

$$\vartheta_{\delta,-}^2 = P_m - \delta - P_m(\rho_m + \delta)^2. \tag{110}$$

In particular,

$$\lim_{\delta \to 0} \lim_{n_b \to \infty} \frac{1}{n_b} \ln \text{Vol}\left\{\mathscr{B}_m^\delta(P_m, \rho_m)\right\} = \frac{1}{2} \ln\left(\pi e P_m\left(1 - \rho_m^2\right)\right). \tag{111}$$

Now,

$$\text{Vol}\left\{\underset{m=1}{\overset{k}{\times}} \mathscr{B}_m^\delta(P_m, \rho_m)\right\} = \prod_{m=1}^k \text{Vol}\left\{\mathscr{B}_m^\delta(P_m, \rho_m)\right\}. \tag{112}$$

Whence, using Lemma 2, (112), (104) and (107), one obtains that

$$\text{Vol}\left\{\mathcal{T}_\delta\left(\boldsymbol{x}\mid\boldsymbol{y}\right)\right\} \geq \max_{\boldsymbol{\mathcal{P}}^{\delta/k}\cap\boldsymbol{\mathcal{R}}_{\boldsymbol{\mathcal{P}}}^{\delta/k}}\left(\pi e\right)^{n/2}\exp\left\{\frac{n_b}{2}\sum_{m=1}^{k}\ln\left(\vartheta_{\delta/k,-}^2\right)\right\} \tag{113}$$

and

$$\text{Vol}\left\{\mathcal{T}_\delta\left(\boldsymbol{x}\mid\boldsymbol{y}\right)\right\} \leq N_{\delta,k}\cdot\max_{\boldsymbol{\mathcal{P}}^{\delta}\cap\boldsymbol{\mathcal{R}}_{\boldsymbol{\mathcal{P}}}^{\delta}}\left(\pi e\right)^{n/2}\exp\left\{\frac{n_b}{2}\sum_{m=1}^{k}\ln\left(\vartheta_{\delta,+}^2\right)\right\}. \tag{114}$$

Thus,

$$\lim_{\delta\to0}\lim_{n\to\infty}\frac{1}{n}\ln\text{Vol}\left\{\mathcal{T}_\delta\left(\boldsymbol{x}\mid\boldsymbol{y}\right)\right\} = \frac{1}{2}\ln\left(\pi e\right) + \max_{\boldsymbol{\mathcal{P}}\cap\boldsymbol{\mathcal{R}}_{\boldsymbol{\mathcal{P}}}}\left\{\frac{h}{2}\sum_{m=1}^{k}\ln\left(P_m\left(1-\rho_m^2\right)\right)\right\} \tag{115}$$

where

$$\boldsymbol{\mathcal{P}} \triangleq \left\{\boldsymbol{P}\in\mathbb{R}^k:\ \frac{1}{k}\sum_{i=1}^{k}P_i = P_x\right\} \tag{116}$$

$$\boldsymbol{\mathcal{R}}_{\boldsymbol{\mathcal{P}}} \triangleq \left\{\boldsymbol{\rho}\in\mathbb{R}^k:\ \frac{1}{k}\sum_{i=1}^{k}\left|\sigma_i'\right|\rho_i\sqrt{\tilde{P}_{y,i}P_i} = \tilde{\rho}\right\}. \tag{117}$$

Finally, the probability in (90), is given by

$$\lim_{n\to\infty}\frac{1}{n}\ln\Pr\left\{\frac{\left\|\boldsymbol{y}-\boldsymbol{\Sigma}'\boldsymbol{X}_1\right\|^2}{2}-\frac{\boldsymbol{\lambda}^T\boldsymbol{X}_1}{\beta}\approx n\epsilon\right\} = \lim_{h\to0}\lim_{\delta\to0}\lim_{n\to\infty}\frac{1}{n}\ln\left(\frac{\text{Vol}\left\{\mathcal{T}_\delta^k\left(\boldsymbol{x}\mid\boldsymbol{y}\right)\right\}}{\text{Vol}\left\{\mathcal{T}_{x,\delta}^n\right\}}\right), \tag{118}$$

in which $\mathcal{T}_{x,\delta}^n$ is the set of n-dimensional $\boldsymbol{x}$-complex vectors with norm $\sqrt{nP_x}$.

Lemma 3 The volume of $\mathcal{T}_{x,\delta}^n$ is given by

$$\lim_{\delta\to0}\lim_{n\to\infty}\frac{1}{n}\ln\text{Vol}\left\{\mathcal{T}_{x,\delta}^n\right\} = \frac{1}{2}\ln\left(\pi e P_x\right). \tag{119}$$

Proof: Readily follows by using almost the same proof of Lemma 2 (see Appendix B). ∎

Thus, applying Lemma 3 on (118), one obtains[12]

$$\Pr\left\{\frac{\left\|\boldsymbol{y}-\boldsymbol{\Sigma}'\boldsymbol{X}_1\right\|^2}{2}-\frac{\boldsymbol{\lambda}^T\boldsymbol{X}_1}{\beta}\approx n\epsilon\right\} \doteq \exp\left\{n\tilde{\Gamma}\left(\epsilon\right)\right\} \tag{120}$$

with probability tending to one as $n\to\infty$, and

$$\tilde{\Gamma}\left(\epsilon\right) \triangleq \lim_{h\to0}\max_{\boldsymbol{\mathcal{P}},\boldsymbol{\mathcal{R}}_{\boldsymbol{\mathcal{P}}}}\left\{\frac{h}{2}\sum_{m=1}^{k}\ln\left(\frac{P_m}{P_x}\left(1-\rho_m^2\right)\right)\right\}. \tag{121}$$

---

[12]Note that at this stage, using once again the dominated convergence theorem (DCT) [40] and Szegö's theorem [36-39], we can refine the bin sizes by taking the limit $h \triangleq n_b/n = 1/k \to 0$, and then to solve a variational problem. However, it turns out that it is better to refine the bin sizes only at the last stage of the analysis.

Therefore, using (88)

$$\boldsymbol{E}\left\{\mathcal{N}\left(\epsilon\right)\right\} \doteq \exp\left\{n\left(R+\tilde{\Gamma}\left(\epsilon\right)\right)\right\}. \tag{122}$$

To finish step 1, the following lemma is proposed and proved in Appendix C[13].

Lemma 4 Let

$$\mathscr{E} \stackrel{\triangle}{=} \left\{\epsilon \in \mathbb{R}: \ R+\tilde{\Gamma}\left(\epsilon\right)>0\right\}. \tag{123}$$

Then,

$$\lim_{n\to\infty} \frac{1}{n}\ln\mathcal{N}\left(\epsilon\right) = \begin{cases} R+\tilde{\Gamma}\left(\epsilon\right), & \epsilon \in \mathscr{E} \\ -\infty, & \text{else} \end{cases} \tag{124}$$

with probability (w.p.) 1.

Step 2: Using Lemma 4, (85), and Varadhan's theorem [41], one obtains that [42, 33, Ch. 2],

$$Z'_e\left(\boldsymbol{Y},\boldsymbol{\lambda}\right) \doteq e^{-nR} \max_{\epsilon\in\mathscr{E}} \exp\left\{n\left(R+\tilde{\Gamma}\left(\epsilon\right)-\beta\epsilon\right)\right\} \tag{125}$$

$$= \exp\left\{n\left[\max_{\epsilon\in\mathscr{E}}\left\{\tilde{\Gamma}\left(\epsilon\right)-\beta\epsilon\right\}\right]\right\}, \tag{126}$$

namely, w.p. 1,

$$\lim_{n\to\infty} \frac{\ln Z'_e\left(\boldsymbol{Y},\boldsymbol{\lambda}\right)}{n} = \max_{\epsilon\in\mathscr{E}}\left\{\tilde{\Gamma}\left(\epsilon\right)-\beta\epsilon\right\}. \tag{127}$$

Let $\Gamma\left(\epsilon\right)$ be defined as in (121), but without the limit over h. It is verified in Appendix D that the maximization and the limit over h can be interchanged, namely, (127) can be rewritten as follows[14]

$$\frac{\ln Z'_e\left(\boldsymbol{Y},\boldsymbol{\lambda}\right)}{n} \sim \lim_{h\to 0}\max_{\epsilon\in\mathscr{E}}\left\{\Gamma\left(\epsilon\right)-\beta\epsilon\right\} \tag{128}$$

with probability tending to one. For simplicity of notation, in the following, the notion of *typical* sequences is used to describe an event that is happening with high probability. For example, we say that for a *typical*

---

[13]Lemma 4 simply states that, if we chose ϵ such that, $R+\tilde{\Gamma}\left(\epsilon\right)>0$, then the energy level ϵ will be "typically" populated with an exponential number of codewords, concentrated very strongly around its mean $\boldsymbol{E}\left\{\mathcal{N}\left(\epsilon\right)\right\}$. Otherwise (which means that $\boldsymbol{E}\left\{\mathcal{N}\left(\epsilon\right)\right\}$ is exponentially small), the energy level ϵ will not be populated by any codewords "typically".

[14]Another approach to "handle" the limit over h is, to first prove the theorem for a linear system whose frequency response is a staircase function (namely, "ignoring evaluate" the limit over h in (118)). Then, using the fact that every frequency response can be approximated arbitrarily well by a sequence of staircase functions with sufficiently small spacing between jumps (Szegö's theorem), the main theorem is proved. Note that (128) literally means that the partition function for any transfer function is obtained via a limit (w.r.t. h) of a sequence of partition functions corresponding to staircase functions with spacings h.

realization of $\boldsymbol{y}$, $Z'_e(\boldsymbol{y}, \boldsymbol{\lambda})$ is given by the right hand side of (127), with the meaning that it happens with probability tending to one as $n \to \infty$. Also, in the following, in order not to drag the limit over h, it will be omitted and then reverted when it has a role.

Next, an explicit expression for $Z'_e(\boldsymbol{y}, \boldsymbol{\lambda})$ is derived. Based on (116), (117), and (121), $\Gamma(\epsilon)$ can be rewritten as

$$\max_{\{P_i\}_{i=1}^k, \{\rho_i\}_{i=1}^k} \quad \frac{h}{2} \sum_{m=1}^k \ln\left(\frac{P_m}{P_x}\left(1 - \rho_m^2\right)\right)$$

$$\text{s.t.} \quad \frac{1}{k}\left(\sum_{m=1}^k |\sigma'_m| \rho_m \sqrt{P_m \tilde{P}_{y,m}} - \frac{1}{2}|\sigma_m|^2 P_x - \frac{1}{2}|\sigma'_m|^2 P_m - \frac{1}{2\beta}\right) = -\epsilon$$

$$\frac{1}{k}\sum_{m=1}^k P_m = P_x. \tag{129}$$

Proposition 1 Let $\{\mu_i\}_{i=1}^k$ be a vector of real scalars such that $\sum_i \mu_i = k$. Then, (129) can be transformed into

$$\max_{\{P_i\}_{i=1}^k, \{\rho_i\}_{i=1}^k, \{\mu_i\}_{i=1}^k} \quad \frac{h}{2} \sum_{m=1}^k \ln\left(\frac{P_m}{P_x}\left(1 - \rho_m^2\right)\right)$$

$$\text{s.t.} \quad |\sigma'_i| \rho_i \sqrt{P_i \tilde{P}_{y,i}} - \frac{1}{2}|\sigma_i|^2 P_x - \frac{1}{2}|\sigma'_i|^2 P_i - \frac{1}{2\beta} = -\mu_i \epsilon, \quad i = 1, \ldots, k$$

$$\frac{1}{k}\sum_{m=1}^k P_m = P_x; \quad \frac{1}{k}\sum_{m=1}^k \mu_m = 1. \tag{130}$$

Proof of Proposition 1: Given a solution of (130), it is to verify that it is feasible for the optimization problem given by (129). Conversely, given a solution, $\{P_m^*, \rho_m^*\}$, of (129), by taking

$$\mu_m^* = -\frac{|\sigma'_m| \rho_m^* \sqrt{P_m^* \tilde{P}_{y,m}} - \frac{1}{2}|\sigma_m|^2 P_x - \frac{1}{2}|\sigma'_m|^2 P_m^* - \frac{1}{2\beta}}{\epsilon}, \tag{131}$$

it can be seen that $\{P_m^*, \rho_m^*, \mu_m^*\}$ is feasible for (130). Thus, the two problems are equivalent. ∎

Using the first constraint in (130), the optimization problem in (130) can be transformed into

$$\max_{\{P_i\}_{i=1}^k, \{\mu_i\}_{i=1}^k} \quad \frac{h}{2} \sum_{m=1}^k \ln\left\{\frac{P_m}{P_x}\left[1 - \left(\frac{\frac{1}{2}|\sigma_m|^2 P_x + \frac{1}{2}|\sigma'_m|^2 P_m + \frac{1}{2\beta} - \mu_m \epsilon}{|\sigma'_m| \sqrt{P_m \tilde{P}_{y,m}}}\right)^2\right]\right\}$$

$$\text{s.t.} \quad \frac{1}{k}\sum_{i=1}^k P_i = P_x; \quad \frac{1}{k}\sum_{m=1}^k \mu_m = 1. \tag{132}$$

Therefore, for a typical realization of the vector $\boldsymbol{y}$, $Z'_e(\boldsymbol{y}, \boldsymbol{\lambda})$ is given by

$$\frac{\ln Z'_e(\boldsymbol{y}, \boldsymbol{\lambda})}{n} \sim \max_{\boldsymbol{\mathcal{P}}} \max_{\mathscr{E}, \{\mu_i\} \in \mathscr{M}_k}$$

$$\frac{h}{2} \sum_{m=1}^{k} \left(\ln \left\{ \frac{P_m}{P_x} \left[1 - \left(\frac{\frac{1}{2} |\sigma_m|^2 P_x + \frac{1}{2} |\sigma_m'|^2 P_m + \frac{1}{2\beta} - \mu_m \epsilon}{|\sigma_m'| \sqrt{P_m \tilde{P}_{y,m}}} \right)^2 \right] \right\} - 2\beta \mu_m \epsilon \right), \tag{133}$$

in which

$$\mathscr{M}_k \triangleq \left\{ (\mu_1, \ldots, \mu_k) \in \mathbb{R}^k : \frac{1}{k} \sum_{i=1}^{k} \mu_i = 1 \right\}. \tag{134}$$

Using the subadditivity property of the maximum norm one obtains (for typical $\boldsymbol{y}$)

$$\frac{\ln Z_e'(\boldsymbol{y}, \boldsymbol{\lambda})}{n} \lesssim \max_{\boldsymbol{\mathcal{P}}, \{\mu_i\}} \frac{h}{2} \sum_{m=1}^{k}$$

$$\max_{\mathscr{E}} \ln \left\{ \frac{P_m}{P_x} \left[1 - \left(\frac{\frac{1}{2} |\sigma_m|^2 P_x + \frac{1}{2} |\sigma_m'|^2 P_m + \frac{1}{2\beta} - \mu_m \epsilon}{|\sigma_m'| \sqrt{P_m \tilde{P}_{y,m}}} \right)^2 \right] \right\} - 2\beta \mu_m \epsilon. \tag{135}$$

Note that except the subadditivity, in the above optimization the maximization is carried over $\{\mu_i\} \in \mathbb{R}^k$ rather than $\{\mu_i\} \in \mathscr{M}_k$ (as it should be), hence increasing further the bound. Changing the variables, $\mu_m \epsilon \mapsto \epsilon_m$, the values of ϵ_m for which the derivative vanishes are the solutions of the following equation

$$\frac{2 \left(\frac{1}{2\beta} - \epsilon_m + \frac{1}{2} |\sigma_m|^2 P_x + \frac{1}{2} |\sigma_m'|^2 P_m \right)}{|\sigma_m'|^2 \tilde{P}_{y,m} P_m \left(1 - \left(\frac{\frac{1}{2} |\sigma_m|^2 P_x + \frac{1}{2} |\sigma_m'|^2 P_m + \frac{1}{2\beta} - \mu_m \epsilon}{|\sigma_m'| \sqrt{P_m \tilde{P}_{y,m}}} \right)^2 \right)} - 2\beta = 0, \tag{136}$$

which after simple algebra, boils down to a quadratic equation whose solutions are

$$\epsilon_{m,1}^* = \frac{2 + |\sigma_m'|^2 \beta P_m + |\sigma_m|^2 \beta P_x + \sqrt{1 + 4\beta^2 |\sigma_m'|^2 \tilde{P}_{y,m} P_m}}{2\beta} \tag{137}$$

$$\epsilon_{m,2}^* = \frac{2 + |\sigma_m'|^2 \beta P_m + |\sigma_m|^2 \beta P_x - \sqrt{1 + 4\beta^2 |\sigma_m'|^2 \tilde{P}_{y,m} P_m}}{2\beta}. \tag{138}$$

Substitution of $\epsilon_{m,1}^*$ in the objective function of (135) reveals that $\epsilon_{m,1}^*$ is not in the objective function domain, and thus only $\epsilon_{m,2}^*$ is considered. In the following, the case $\epsilon_{m,2}^* \in \mathscr{E}$ is first analyzed. Substituting $\epsilon_{m,2}^*$ in (135), one obtains (for typical $\boldsymbol{y}$)

$$\frac{\ln Z_e'(\boldsymbol{y}, \boldsymbol{\lambda})}{n} \lesssim \max_{\boldsymbol{\mathcal{P}}, \{\mu_i\}} \frac{h}{2} \sum_{m=1}^{k} \ln \left\{ \frac{P_m}{P_x} \left[1 - \left(\frac{\frac{1}{2} |\sigma_m|^2 P_x + \frac{1}{2} |\sigma_m'|^2 P_m + \frac{1}{2\beta} - \epsilon_{m,2}^*}{|\sigma_m'| \sqrt{P_m P_{y,m}}} \right)^2 \right] \right\} - 2\beta \epsilon_{m,2}^*. \tag{139}$$

Let γ be the Lagrange multiplier associated with the power constraint. Then, the derivative of the objective function in (139) w.r.t. P_m is given by

$$-|\sigma_m'|^2 \beta + \frac{1 - \sqrt{1 + 4 |\sigma_m'|^2 \beta^2 P_m^* \tilde{P}_{y,m}}}{2 P_m^*} - \gamma = 0, \tag{140}$$

which vanishes at

$$P_m^* = \frac{|\sigma_m'|^2 \beta \left(1 + \beta \tilde{P}_{y,m}\right) + \gamma}{\left(|\sigma_m'|^2 \beta + \gamma\right)^2}, \tag{141}$$

independently of $\epsilon_{m,2}^*$, and γ is chosen such that $\sum_i P_i = kP_x$. Therefore (for typical $\boldsymbol{y}$),

$$\frac{\ln Z_e'(\boldsymbol{y}, \boldsymbol{\lambda})}{n} \lesssim \frac{h}{2} \sum_{m=1}^{k} \ln \left\{ \frac{P_m^*}{P_x} \left[1 - \left(\frac{\frac{1}{2}|\sigma_m|^2 P_x + \frac{1}{2}|\sigma_m'|^2 P_m^* + \frac{1}{2\beta} - \epsilon_m^*}{|\sigma_m'| \sqrt{P_m^* \tilde{P}_{y,m}}} \right)^2 \right] \right\} - 2\beta\epsilon_m^* \tag{142}$$

$$\triangleq F_{\text{par}},$$

where $\epsilon_m^* \triangleq \epsilon_{m,2}^* (P_m^*)$. Hence, an upper bound, F_{par}, on $\ln Z_e'(\boldsymbol{y}, \boldsymbol{\lambda})/n$ is obtained. On the other hand, by taking

$$\epsilon^* = \frac{\sum_{i=1}^{k} \epsilon_i^*}{k} \tag{143}$$

$$\mu_m^* = \frac{k\epsilon_m^*}{\sum_{i=1}^{k} \epsilon_i^*}, \tag{144}$$

and (141), this bound is achieved. Summarizing the above results, $Z_e'(\boldsymbol{y}, \boldsymbol{\lambda})$ is given by (for typical $\boldsymbol{y}$)

$$\frac{\ln Z_e'(\boldsymbol{y}, \boldsymbol{\lambda})}{n} \sim \begin{cases} F_{\text{par}}, & \Gamma(\epsilon^*) + R > 0 \\ \\ \Gamma(\epsilon_s) - \beta\epsilon_s, & \Gamma(\epsilon^*) + R \leq 0 \end{cases}. \tag{145}$$

Since at the final step of the calculation, the partition function (or its derivative w.r.t. $\boldsymbol{\lambda}$) is evaluated at $\boldsymbol{\lambda} = \boldsymbol{0}$, the range $\Gamma(\epsilon^*) + R > 0$ should be computed at the vicinity of $\boldsymbol{\lambda} = \boldsymbol{0}$. First, note that $\tilde{P}_{y,i}$, given in Lemma 1, can be written as

$$\tilde{P}_{y,i} = |\sigma_i|^2 P_x + \frac{1}{\beta} + \frac{2}{\beta} \frac{1}{n_b} \text{Re} \left(\frac{1}{\sigma_i'} \sum_{r \in i\text{-th bin}} y_r \lambda_r \right) + \frac{1}{\beta^2 |\sigma_i'|^2} \frac{1}{n_b} \sum_{r \in i\text{-th bin}} |\lambda_r|^2. \tag{146}$$

Hence, substituting $\boldsymbol{\lambda} = \boldsymbol{0}$ in (141), one obtains

$$P_m^*|_{\boldsymbol{\lambda}=\boldsymbol{0}} = \frac{|\sigma_m'|^2 \beta \left(1 + \beta \left.\tilde{P}_{y,m}\right|_{\boldsymbol{\lambda}=\boldsymbol{0}}\right) + \gamma_0}{\left(|\sigma_m'|^2 \beta + \gamma_0\right)^2} \tag{147}$$

$$= \frac{|\sigma_m'|^2 \beta \left(2 + \beta |\sigma_m|^2 P_x\right) + \gamma_0}{\left(|\sigma_m'|^2 \beta + \gamma_0\right)^2} \tag{148}$$

where γ_0 is chosen such that

$$kP_x = \sum_m P_m^*|_{\boldsymbol{\lambda}=\boldsymbol{0}}. \tag{149}$$

Substitution of $P_m^*|_{\boldsymbol{\lambda}=0}$ and $\epsilon_m^*\left(P_m^*|_{\boldsymbol{\lambda}=0}\right)|_{\boldsymbol{\lambda}=0}$ in $\Gamma\left(\epsilon\right)$, reveals that

$$\Gamma\left(\epsilon^*\right)|_{\boldsymbol{\lambda}=0} = \frac{h}{2}\sum_{m=1}^{k}\ln\left\{\frac{P_m^*|_{\boldsymbol{\lambda}=0}}{P_x}\left[1 - \left(\frac{\frac{1}{2}\left|\sigma_m\right|^2 P_x + \frac{1}{2}\left|\sigma_m'\right|^2 P_m^*|_{\boldsymbol{\lambda}=0} + \frac{1}{2\beta} - \epsilon_m^*|_{\boldsymbol{\lambda}=0}}{\left|\sigma_m'\right|\sqrt{P_m^*|_{\boldsymbol{\lambda}=0}\left.\tilde{P}_{y,m}\right|_{\boldsymbol{\lambda}=0}}}\right)^2\right]\right\},$$
(150)

and that

$$\epsilon_m^*|_{\boldsymbol{\lambda}=0} = \frac{2 + \left|\sigma_m'\right|^2\beta\,P_m^*|_{\boldsymbol{\lambda}=0} + \left|\sigma_m\right|^2\beta P_x - \sqrt{1 + 4\beta^2\left|\sigma_m'\right|^2\tilde{P}_{y,m}\,P_m^*|_{\boldsymbol{\lambda}=0}}}{2\beta}.$$
(151)

Then, substituting (151) in the mth term of the sum in (150), it becomes

$$\ln\left(\frac{\sqrt{1 + 4\left|\sigma_m'\right|^2\beta^2\,P_m^*|_{\boldsymbol{\lambda}=0}\left.\tilde{P}_{y,m}\right|_{\boldsymbol{\lambda}=0}} - 1}{2P_x\left|\sigma_m'\right|^2\beta^2\left.\tilde{P}_{y,m}\right|_{\boldsymbol{\lambda}=0}}\right),$$
(152)

which after substitution of (147), boils down to

$$\frac{1}{P_x\gamma_0 + \left|\sigma_m'\right|^2 P_x\beta}.$$
(153)

Hence, substituting (153) in (150), one obtains

$$\Gamma\left(\epsilon^*\right)|_{\boldsymbol{\lambda}=0} = -\frac{h}{2}\sum_{m=1}^{k}\ln\left(P_x\gamma_0 + \left|\sigma_m'\right|^2 P_x\beta\right).$$
(154)

Accordingly, the region $\Gamma\left(\epsilon^*\right) + R \le 0$ is equivalent to

$$R \le \frac{h}{2}\sum_{m=1}^{k}\ln\left(P_x\gamma_0 + \left|\sigma_m'\right|^2 P_x\beta\right) \triangleq R_e,$$
(155)

and hence

$$\frac{\ln Z_e'\left(\boldsymbol{y},\boldsymbol{\lambda}\right)}{n} \sim \begin{cases} F_{\text{par}}, & R > R_e \\ \Gamma\left(\epsilon_s\right) - \beta\epsilon_s, & R \le R_e \end{cases}.$$
(156)

The next step in the evaluation of $Z'\left(\boldsymbol{y},\boldsymbol{\lambda}\right)$, is taking into account $Z_c'\left(\boldsymbol{y},\boldsymbol{\lambda}\right)$. To this end, the following relation is used

$$\lim_{n\to\infty}\frac{\ln\left(e^{-na} + e^{-nb}\right)}{n} = -\min\left(a,b\right).$$
(157)

Accordingly, within the range $R > R_e$, for a typical code and realizations of the vector $\boldsymbol{y}$, we search rates for which $Z_c'\left(\boldsymbol{y},\boldsymbol{0}\right) > Z_e'\left(\boldsymbol{y},\boldsymbol{0}\right)$, namely,

$$\frac{\ln Z_c'\left(\boldsymbol{y},\boldsymbol{0}\right)}{n} > F_{\text{par}}|_{\boldsymbol{\lambda}=0}.$$
(158)

Recall that $F_{\text{par}}|_{\boldsymbol{\lambda}=0}$ is given by

$$F_{\text{par}}|_{\boldsymbol{\lambda}=0} = \Gamma\left(\epsilon^*\right) - \beta\epsilon^* \tag{159}$$

$$= -\frac{h}{2} \sum_{m=1}^{k} \left\{ \ln\left(P_x\gamma_0 + P_x\left|\sigma'_m\right|^2\beta\right) + 2\beta\epsilon^*_m|_{\boldsymbol{\lambda}=0} \right\}, \tag{160}$$

and that

$$\frac{\ln Z'_c\left(\boldsymbol{y}, \boldsymbol{0}\right)}{n} = -\left(R + \frac{1}{2} + \frac{\beta}{2n}\left\|\left(\boldsymbol{A}' - \boldsymbol{A}\right)\boldsymbol{x}_0\right\|^2\right) \tag{161}$$

$$= -R - \frac{1}{2} - \frac{h\beta}{2}\sum_{m=1}^{k}\left|\sigma'_m - \sigma_m\right|^2 P_x. \tag{162}$$

Hence the inequality in (158) becomes

$$R < \frac{h}{2}\sum_{m=1}^{k}\ln\left(P_x\gamma_0 + P_x\left|\sigma'_m\right|^2\beta\right) - \frac{1}{2} - \frac{h\beta}{2}\sum_{m=1}^{k}\left|\sigma'_m - \sigma_m\right|^2 P_x$$

$$+ \frac{h}{2}\sum_{m=1}^{k}\left\{2 + \left|\sigma'_m\right|^2\beta\, P^*_m|_{\boldsymbol{\lambda}=0} + \left|\sigma_m\right|^2\beta P_x - \sqrt{1 + 4\beta^2\left|\sigma'_m\right|^2\tilde{P}_{y,m}\,P^*_m|_{\boldsymbol{\lambda}=0}}\right\} \tag{163}$$

$$= \frac{h}{2}\sum_{m=1}^{k}\ln\left(P_x\gamma_0 + P_x\left|\sigma'_m\right|^2\beta\right) + \frac{1}{2} + h\beta\sum_{m=1}^{k}\text{Re}\left(\sigma'^*_m\sigma_m\right)P_x$$

$$+ \frac{h\beta}{2}\sum_{m=1}^{k}\left|\sigma'_m\right|^2\left(P^*_m|_{\boldsymbol{\lambda}=0} - P_x\right) - \frac{h}{2}\sum_{m=1}^{k}\sqrt{1 + 4\beta^2\left|\sigma'_m\right|^2\tilde{P}_{y,m}\,P^*_m|_{\boldsymbol{\lambda}=0}}. \tag{164}$$

Substituting P^*_m, given in (148), in the last two terms of (164), one obtains

$$R < \frac{h}{2}\sum_{m=1}^{k}\ln\left(P_x\gamma_0 + P_x\left|\sigma'_m\right|^2\beta\right) + \frac{1}{2} + h\beta\sum_{m=1}^{k}\text{Re}\left(\sigma'^*_m\sigma_m\right)P_x$$

$$+ \frac{h\beta}{2}\sum_{m=1}^{k}\left|\sigma'_m\right|^2\beta\left(\frac{\left|\sigma'_m\right|^2\beta\left(2 + \beta\left|\sigma_m\right|^2 P_x\right) + \gamma_0}{\left(\left|\sigma'_m\right|^2\beta + \gamma_0\right)^2} - P_x\right)$$

$$- \frac{h}{2}\sum_{m=1}^{k}\frac{\left|\sigma'_m\right|^2\beta\left(3 + 2P_x\beta\left|\sigma_m\right|^2\right) + \gamma_0}{\left(\left|\sigma'_m\right|^2\beta + \gamma_0\right)}. \tag{165}$$

Refining the bin sizes by taking the limit $h \to 0$, while using Szegö's theorem, it is shown in Appendix E that (165) becomes

$$R < R_e + R_d \stackrel{\triangle}{=} R_c \tag{166}$$

where

$$R_e \stackrel{\triangle}{=} \frac{1}{4\pi}\int_0^{2\pi}\ln\left(P_x\gamma_0 + \left|\mathbf{H}'\left(\omega\right)\right|^2 P_x\beta\right)\mathrm{d}\omega, \tag{167}$$

and

$$R_d \triangleq \frac{1}{2} + \beta P_x \frac{1}{2\pi} \int_0^{2\pi} \mathrm{Re}\left(\mathbf{H}'^*\left(\omega\right)\mathbf{H}\left(\omega\right)\right)\mathrm{d}\omega$$

$$+ \frac{1}{4\pi}\int_0^{2\pi}\left|\mathbf{H}'\left(\omega\right)\right|^2 \beta\left(\frac{\left|\mathbf{H}'\left(\omega\right)\right|^2\beta\left(2 + \beta\left|\mathbf{H}\left(\omega\right)\right|^2 P_x\right) + \gamma_0}{\left(\left|\mathbf{H}'\left(\omega\right)\right|^2\beta + \gamma_0\right)^2} - P_x\right)$$

$$- \frac{1}{4\pi}\int_0^{2\pi}\frac{\left|\mathbf{H}'\left(\omega\right)\right|^2\beta\left(3 + 2P_x\beta\left|\mathbf{H}\left(\omega\right)\right|^2\right) + \gamma_0}{\left(\left|\mathbf{H}'\left(\omega\right)\right|^2\beta + \gamma_0\right)}. \tag{168}$$

Hence, within the range $R > R_e$, $Z'_c\left(\boldsymbol{y}, \mathbf{0}\right) > Z'_e\left(\boldsymbol{y}, \mathbf{0}\right)$ (again, typical code and realization vector $\boldsymbol{y}$) for

$$\{R < R_c\} \cap \{R > R_e\} = \{R_e < R < R_e + R_d = R_c\}, \tag{169}$$

which is a non-empty set if R_d is positive. Next, within the range $R \leq R_e$, $Z'_c\left(\boldsymbol{y}, \mathbf{0}\right) > Z'_e\left(\boldsymbol{y}, \mathbf{0}\right)$ for rates which satisfy (for typical code and realization of $\boldsymbol{y}$)

$$\frac{\ln Z'_c\left(\boldsymbol{y}, \mathbf{0}\right)}{n} > \Gamma\left(\epsilon_s\right) - \beta\epsilon_s|_{\boldsymbol{\lambda}=0}. \tag{170}$$

First, recall that ϵ_s satisfies $R + \Gamma\left(\epsilon_s\right) = 0$, and hence $\Gamma\left(\epsilon_s\right) = -R$. Thus, (170) can be rewritten as

$$-R - \frac{1}{2} - \frac{h\beta}{2}\sum_{m=1}^{k}\left|\sigma'_m - \sigma_m\right|^2 P_x > -R - \beta\epsilon_s|_{\boldsymbol{\lambda}=0}, \tag{171}$$

which is equivalent to

$$\epsilon_s|_{\boldsymbol{\lambda}=0} > \frac{1}{2\beta} + \frac{h}{2}\sum_{m=1}^{k}\left|\sigma'_m - \sigma_m\right|^2 P_x. \tag{172}$$

Applying $\Gamma\left(\cdot\right)$ to (172), one obtains

$$\Gamma\left(\epsilon_s|_{\boldsymbol{\lambda}=0}\right) > \Gamma\left(\frac{1}{2\beta} + \frac{h}{2}\sum_{m=1}^{k}\left|\sigma'_m - \sigma_m\right|^2 P_x\right), \tag{173}$$

and hence

$$R < -\Gamma\left(\frac{1}{2\beta} + \frac{h}{2}\sum_{m=1}^{k}\left|\sigma'_m - \sigma_m\right|^2 P_x\right)\bigg|_{\boldsymbol{\lambda}=0} \triangleq R_g, \tag{174}$$

where

$$\Gamma\left(\frac{1}{2\beta} + \frac{h}{2}\sum_{m=1}^{k}\left|\sigma'_m - \sigma_m\right|^2 P_x\right) =$$

$$\max_{\{P_i\}_{i=1}^{k}, \{\rho_i\}_{i=1}^{k}} \frac{h}{2}\sum_{m=1}^{k}\ln\left(\frac{P_m}{P_x}\left(1 - \rho_m^2\right)\right)$$

$$\text{s.t.}\quad \frac{1}{k}\left(\sum_{m=1}^{k}\left|\sigma'_m\right|\rho_m\sqrt{P_m\tilde{P}_{y,m}} - \frac{1}{2}\left|\sigma'_m\right|^2\left(P_m - P_x\right) - \mathrm{Re}\left(\sigma'^*_m\sigma_m\right)P_x\right) = 0$$

$$\frac{1}{k} \sum_{m=1}^{k} P_m = P_x. \tag{175}$$

To conclude, $Z'(\boldsymbol{y}, \boldsymbol{\lambda})$ is given by (for a typical code and $\boldsymbol{y}$)

$$\frac{\ln Z'(\boldsymbol{y}, \boldsymbol{\lambda})}{n} \sim \begin{cases} F_{\text{par}}, & R > R_e \vee R_c \\ \Gamma(\epsilon_s) - \beta\epsilon_s, & R_g < R \leq R_e \\ \ln Z'_c(\boldsymbol{y}, \boldsymbol{\lambda})/n, & \{R_e \leq R \leq R_c\} \cup \{R \leq R_g \wedge R_e\} \end{cases} \tag{176}$$

where $a \vee b \triangleq \max(a, b)$ and $a \wedge b \triangleq \min(a, b)$. In the following, the relation

$$R_d > 0 \implies R_e < R_g, \tag{177}$$

is verified. Recall that R_d follows from the requirement that

$$-\left(R + \frac{1}{2} + \frac{\beta}{2n}\|(\boldsymbol{A} - \boldsymbol{A}')\boldsymbol{x}_0\|^2\right) = \frac{\ln Z'_c(\boldsymbol{y}, \boldsymbol{0})}{n} \tag{178}$$

$$\geq F_{\text{par}}|_{\boldsymbol{\lambda}=0} \tag{179}$$

$$= \Gamma(\epsilon^*) - \beta\epsilon^*|_{\boldsymbol{\lambda}=0} = -R_e - \beta\epsilon^*|_{\boldsymbol{\lambda}=0}, \tag{180}$$

which can be rewritten as

$$R \leq R_e + \beta\epsilon^*|_{\boldsymbol{\lambda}=0} - \left(\frac{1}{2} + \frac{\beta}{2n}\|(\boldsymbol{A}' - \boldsymbol{A})\boldsymbol{x}_0\|^2\right), \tag{181}$$

and thus R_d is given by

$$R_d = \beta\epsilon^*|_{\boldsymbol{\lambda}=0} - \left(\frac{1}{2} + \frac{\beta}{2n}\|(\boldsymbol{A}' - \boldsymbol{A})\boldsymbol{x}_0\|^2\right). \tag{182}$$

Accordingly, $R_d > 0$ is equivalent to

$$\beta\epsilon^*|_{\boldsymbol{\lambda}=0} > \frac{1}{2} + \frac{\beta}{2n}\|(\boldsymbol{A} - \boldsymbol{A}')\boldsymbol{x}_0\|^2. \tag{183}$$

Now, within the range $R \leq R_e$, $Z'_c(\boldsymbol{y}, \boldsymbol{0}) \geq Z'_e(\boldsymbol{y}, \boldsymbol{0})$ if (172)

$$\beta\epsilon_s|_{\boldsymbol{\lambda}=0} > \frac{1}{2} + \frac{\beta}{2n}\|(\boldsymbol{A} - \boldsymbol{A}')\boldsymbol{x}_0\|^2. \tag{184}$$

However, $R \leq R_e$ is equivalent to $\epsilon^* \notin \mathscr{E}$, and thus $\epsilon_s|_{\boldsymbol{\lambda}=0} \geq \epsilon^*|_{\boldsymbol{\lambda}=0}$. Therefore, if $R_d > 0$, the following holds

$$\beta\epsilon_s|_{\boldsymbol{\lambda}=0} \geq \beta\epsilon^*|_{\boldsymbol{\lambda}=0} > \frac{1}{2} + \frac{\beta}{2n}\|(\boldsymbol{A} - \boldsymbol{A}')\boldsymbol{x}_0\|^2. \tag{185}$$

Whence, (184) holds true within the whole region $R \leq R_e$, and therefore $R_e < R_g$. Thus, for $R_d > 0$, $Z'(\boldsymbol{y}, \boldsymbol{\lambda})$ becomes (for a typical code realization $\boldsymbol{y}$)

$$\frac{\ln Z'(\boldsymbol{y}, \boldsymbol{\lambda})}{n} \sim \begin{cases} F_{\text{par}}, & R > R_c \\ \ln Z'_c(\boldsymbol{y}, \boldsymbol{\lambda})/n, & R \leq R_c \end{cases}. \tag{186}$$

If however, $R_d < 0$, then $R_g \leq R_e$, and hence (for a typical code realization $\boldsymbol{y}$)

$$\frac{\ln Z'(\boldsymbol{y}, \boldsymbol{\lambda})}{n} \sim \begin{cases} F_{\text{par}}, & R > R_e \\ -R - \beta \epsilon_s, & R_g < R \leq R_e \\ \ln Z'_c(\boldsymbol{y}, \boldsymbol{\lambda})/n, & R \leq R_g \end{cases}. \tag{187}$$

Recall that ϵ_s is the solution of the equation

$$\Gamma(\epsilon_s) + R = 0, \tag{188}$$

where $\Gamma(\epsilon_s)$ is given by

$$\max_{\{P_i\}_{i=1}^k, \{\rho_i\}_{i=1}^k} \frac{h}{2} \sum_{m=1}^k \ln\left(\frac{P_m}{P_x}\left(1 - \rho_m^2\right)\right)$$

$$\text{s.t.} \quad \frac{1}{k}\left(\sum_{m=1}^k |\sigma'_m| \rho_m \sqrt{P_m \tilde{P}_{y,m}} - \frac{1}{2}|\sigma_m|^2 P_x - \frac{1}{2}|\sigma'_m|^2 P_m - \frac{1}{2\beta}\right) = -\epsilon_s$$

$$\frac{1}{k}\sum_{m=1}^k P_m = P_x. \tag{189}$$

Similarly to the optimization problem in (132), the above maximization problem can be rewritten as

$$\max_{\{P_i\}_{i=1}^k, \{\rho_i\}_{i=1}^k} \frac{h}{2} \sum_{m=1}^k \ln\left\{\frac{P_m}{P_x}\left[1 - \left(\frac{\frac{1}{2}|\sigma_m|^2 P_x + \frac{1}{2}|\sigma'_m|^2 P_m + \frac{1}{2\beta} - \mu_m \epsilon_s}{\sigma'_m \sqrt{P_m \tilde{P}_{y,m}}}\right)^2\right]\right\}$$

$$\text{s.t.} \quad \frac{1}{k}\sum_{m=1}^k P_m = P_x, \quad \frac{1}{k}\sum_{m=1}^k \mu_m = 1. \tag{190}$$

Accordingly, the derivative of the objective function w.r.t. P_m vanishes at

$$P_m^* = \frac{4\alpha_1 \epsilon_s^2 + |\sigma'_m|^2 \alpha_2 \left(\tilde{P}_{y,m}\alpha_2 + 2\epsilon_s\right)}{\left(|\sigma'_m|^2 \alpha_2 + 2\alpha_1 \epsilon_s\right)^2} \tag{191}$$

and the derivative w.r.t. μ_m it vanishes at

$$\mu_m^* = \frac{4\alpha_1^2 \epsilon_s^2 \left(1 + P_x \beta |\sigma_m|^2\right) + 4|\sigma'_m|^2 \alpha_1 \epsilon_s \left(\alpha_2 - \beta \tilde{P}_{y,m}\alpha_2 + \beta \epsilon_s + P_x \beta \alpha_2 |\sigma_m|^2\right)}{2\beta \epsilon_s \left(|\sigma'_m|^2 \alpha_2 + 2\alpha_1 \epsilon_s\right)^2}$$

$$+ \frac{|\sigma'_m|^4 \alpha_2 \left(\alpha_2 - \beta \tilde{P}_{y,m}\alpha_2 + 2\beta\epsilon_s + P_x\beta\alpha_2 |\sigma_m|^2\right)}{2\beta\epsilon_s \left(|\sigma'_m|^2 \alpha_2 + 2\alpha_1\epsilon_s\right)^2} \tag{192}$$

where α_1 is chosen such that $kP_x = \sum_m P_m^*$, and α_2 is chosen such that $k = \sum_m \mu_m^*$. Substituting the above maximizers in the objective function one obtains

$$\Gamma(\epsilon_s) = \frac{h}{2} \sum_{m=1}^{k} \ln\left(\frac{2\epsilon_s}{P_x |\sigma'_m|^2 \alpha_2 + 2P_x\alpha_1\epsilon_s}\right). \tag{193}$$

For completeness, a closed-form expression for R_g is derived. Based on (174)

$$R_g = -\left.\Gamma\left(\frac{1}{2\beta} + \frac{h}{2} \sum_{m=1}^{k} |\sigma'_m - \sigma_m|^2 P_x\right)\right|_{\boldsymbol{\lambda}=0}. \tag{194}$$

Using (193), and upon taking the limit $h \to 0$ (while using Szegö's theorem, as was done in (168))

$$R_g = -\frac{1}{4\pi} \int_0^{2\pi} \ln\left(\frac{2\tilde{\epsilon}}{P_x |\mathbf{H}'(\omega)|^2 \tilde{\alpha}_2 + 2P_x\tilde{\alpha}_1\tilde{\epsilon}}\right) d\omega \tag{195}$$

where $\tilde{\alpha}_1$ and $\tilde{\alpha}_2$ solve the simultaneous equations

$$\frac{1}{2\pi} \int_0^{2\pi} \frac{4\tilde{\alpha}_1\tilde{\epsilon}^2 + |\mathbf{H}'(\omega)|^2 \tilde{\alpha}_2 \left[\left(|\mathbf{H}(\omega)|^2 P_x + \frac{1}{\beta}\right) \tilde{\alpha}_2 + 2\tilde{\epsilon}\right]}{\left(|\mathbf{H}'(\omega)|^2 \tilde{\alpha}_2 + 2\tilde{\alpha}_1\tilde{\epsilon}\right)^2} d\omega = P_x \tag{196}$$

$$\frac{1}{2\pi} \int_0^{2\pi} \frac{4\tilde{\alpha}_1^2\tilde{\epsilon}^2 \left(1 + P_x\beta |\mathbf{H}(\omega)|^2\right) + 4 |\mathbf{H}'(\omega)|^2 \tilde{\alpha}_1\tilde{\epsilon}^2\beta + 2 |\mathbf{H}'(\omega)|^4 \tilde{\alpha}_2\tilde{\epsilon}\beta}{2\beta\tilde{\epsilon} \left(|\mathbf{H}'(\omega)|^2 \tilde{\alpha}_2 + 2\tilde{\alpha}_1\tilde{\epsilon}\right)^2} d\omega = 1, \tag{197}$$

and

$$\tilde{\epsilon} = \frac{1}{2\beta} + \frac{P_x}{4\pi} \int_0^{2\pi} |\mathbf{H}'(\omega) - \mathbf{H}(\omega)|^2 d\omega. \tag{198}$$

Obtaining $Z'(\boldsymbol{y}, \boldsymbol{\lambda})$, using the tools presented in Subsection V-B, the MSE is now derived. The MSE estimator of the ith component (*chip*) of $\boldsymbol{x}'$, within the qth bin, is given by the derivative of $Z'(\boldsymbol{y}, \boldsymbol{\lambda})$ w.r.t. λ_{q_i} evaluated at $\boldsymbol{\lambda} = \boldsymbol{0}$[15]. The derivative of F_{par} is given by

$$\left.\frac{\partial F_{\text{par}}}{\partial \lambda_{q_i}}\right|_{\boldsymbol{\lambda}=0} = -\frac{h}{2} \sum_{l=1}^{k} \left\{\frac{P_x}{P_x\gamma_0 + P_x |\sigma'_l|^2 \beta} \left.\frac{\partial\gamma}{\partial\lambda_{q_i}}\right|_{\boldsymbol{\lambda}=0} + \left.\frac{\partial\psi_l}{\partial\lambda_{q_i}}\right|_{\boldsymbol{\lambda}=0}\right\}. \tag{199}$$

---

[15] A very similar analysis applies also to the derivative $\frac{\partial}{\partial\lambda_i} \ln Z(\boldsymbol{y}, \boldsymbol{\lambda})$, which is essentially a weighted average over x_i with weights proportional to $\boldsymbol{E}\mathcal{N}(\epsilon)e^{-\beta\epsilon}$ for $\epsilon \in \mathscr{E}$. Thus, the exponentially dominant weight is due to the term that maximizes the exponent [5, 6]. Hence, in this case, the commutativity between the derivative w.r.t. $\boldsymbol{\lambda}$ and the limit $n \to \infty$ is legitimate. Another approach to justify the interchange of the order of these operations is to use well-known results (for example, [43, Ch. 16],[44, 45]) on functional properties of a limit function, which are applicable in our case due to the uniform convergence of the various relevant terms (see Appendix D).

Let $x_q \triangleq \partial\gamma/\partial\lambda_{q_i}|_{\boldsymbol{\lambda}=0}$. Using (141), one obtains

$$
\left.\frac{\partial P_q^*}{\partial\lambda_{q_i}}\right|_{\boldsymbol{\lambda}=0} = \frac{\left(\left|\sigma_q'\right|^2 \beta^2 \left.\frac{\partial\tilde{P}_{y,q}}{\partial\lambda_{q_i}}\right|_{\boldsymbol{\lambda}=0} + x_q\right)\left(\left|\sigma_q'\right|^2 + \gamma_0\right)^2}{\left(\left|\sigma_q'\right|^2 + \gamma_0\right)^4}
$$

$$
- \frac{2\left(\left|\sigma_q'\right|^2 + \gamma_0\right)\left(\left|\sigma_q'\right|^2 \beta\left(1 + \beta\tilde{P}_{y,q}\big|_{\lambda=0}\right) + \gamma_0\right)x_q}{\left(\left|\sigma_q'\right|^2 + \gamma_0\right)^4}, \tag{200}
$$

and for $l \neq q$

$$
\left.\frac{\partial P_l^*}{\partial\lambda_{q_i}}\right|_{\boldsymbol{\lambda}=0} = \frac{\left(\left|\sigma_l'\right|^2 + \gamma_0\right)^2 x_q - 2\left(\left|\sigma_l'\right|^2 + \gamma_0\right)\left(\left|\sigma_l'\right|^2 \beta\left(1 + \beta\tilde{P}_{y,l}\big|_{\lambda=0}\right) + \gamma_0\right)x_q}{\left(\left|\sigma_l'\right|^2 + \gamma_0\right)^4} \tag{201}
$$

where by using (146)

$$
\left.\frac{\partial\tilde{P}_{y,q}}{\partial\lambda_{q_i}}\right|_{\boldsymbol{\lambda}=0} = \frac{2}{\beta\sigma_q'}\frac{y_{q_i}}{n_b} \tag{202}
$$

$$
\beta\tilde{P}_{y,q}\Big|_{\lambda=0} = 1 + |\sigma_q|^2 P_x\beta. \tag{203}
$$

Since γ_0 is chosen to satisfy $\sum_r P_r^*|_{\boldsymbol{\lambda}=0} = kP_x$, it follows that

$$
0 = \left.\frac{\partial}{\partial\lambda_{q_i}}\sum_{r=1}^k P_r^*\right|_{\boldsymbol{\lambda}=0} = \sum_{r=1}^k \left.\frac{\partial P_r^*}{\partial\lambda_{q_i}}\right|_{\boldsymbol{\lambda}=0} \tag{204}
$$

$$
= \frac{\left|\sigma_q'\right|^2 \beta^2 \left.\frac{\partial\tilde{P}_{y,q}}{\partial\lambda_{q_i}}\right|_{\boldsymbol{\lambda}=0}\left(\left|\sigma_q'\right|^2 + \gamma_0\right)^2}{\left(\left|\sigma_q'\right|^2 + \gamma_0\right)^4}
$$

$$
+ x_q\sum_{r=1}^k \frac{\left(\left|\sigma_r'\right|^2 + \gamma_0\right)^2 - 2\left(\left|\sigma_r'\right|^2 + \gamma_0\right)\left(\left|\sigma_r'\right|^2 \beta\left(1 + \beta\tilde{P}_{y,r}\big|_{\lambda=0}\right) + \gamma_0\right)}{\left(\left|\sigma_r'\right|^2 + \gamma_0\right)^4}, \tag{205}
$$

and thus

$$
x_q = \frac{\left|\sigma_q'\right|^2 \beta^2 \left.\frac{\partial\tilde{P}_{y,q}}{\partial\lambda_{q_i}}\right|_{\boldsymbol{\lambda}=0}}{\left(\left|\sigma_q'\right|^2 + \gamma_0\right)^2 C} = \frac{2\sigma_q'^*\beta}{n_b\left(\left|\sigma_q'\right|^2 + \gamma_0\right)^2 C}y_{q_i}, \tag{206}
$$

where

$$
C \triangleq \sum_{r=1}^k \frac{\left(\left|\sigma_r'\right|^2 + \gamma_0\right) - 2\left(\left|\sigma_r'\right|^2 \beta\left(1 + \beta\tilde{P}_{y,r}\big|_{\lambda=0}\right) + \gamma_0\right)}{\left(\left|\sigma_r'\right|^2 + \gamma_0\right)^3}. \tag{207}
$$

Next, $\partial\psi_l/\partial\lambda_{q_i}|_{\boldsymbol{\lambda}=0}$, is calculated. Using the definition of ψ_m in (11) one obtains

$$
\left.\frac{\partial\psi_q}{\partial\lambda_{q_i}}\right|_{\boldsymbol{\lambda}=0} = \left|\sigma_q'\right|^2 \beta \left.\frac{\partial P_q^*}{\partial\lambda_{q_i}}\right|_{\boldsymbol{\lambda}=0} - \frac{2\beta^2\left|\sigma_q'\right|^2\left(\left.\frac{\partial P_{y,q}^*}{\partial\lambda_{q_i}}\right|_{\boldsymbol{\lambda}=0} P_q^*\big|_{\boldsymbol{\lambda}=0} + \tilde{P}_{y,q}\big|_{\boldsymbol{\lambda}=0}\left.\frac{\partial P_q^*}{\partial\lambda_{q_i}}\right|_{\boldsymbol{\lambda}=0}\right)}{\sqrt{1 + 4\beta^2\left|\sigma_q'\right|^2 \tilde{P}_{y,q}\big|_{\boldsymbol{\lambda}=0} P_q^*\big|_{\boldsymbol{\lambda}=0}}}, \tag{208}
$$

and for $l \neq q$

$$\left. \frac{\partial \psi_l}{\partial \lambda_{q_i}} \right|_{\boldsymbol{\lambda}=0} = |\sigma_l'|^2 \beta \left. \frac{\partial P_l^*}{\partial \lambda_{q_i}} \right|_{\boldsymbol{\lambda}=0} - \frac{2\beta^2 |\sigma_l'|^2 \left. \tilde{P}_{y,l} \right|_{\boldsymbol{\lambda}=0} \left. \frac{\partial P_l^*}{\partial \lambda_{q_i}} \right|_{\boldsymbol{\lambda}=0}}{\sqrt{1 + 4\beta^2 |\sigma_l'|^2 \left. \tilde{P}_{y,l} \right|_{\boldsymbol{\lambda}=0} P_l^* \big|_{\boldsymbol{\lambda}=0}}}. \tag{209}$$

Substituting (206), (208) and (209) in (199), the MSE estimator in the range $R > R_c$ and $R > R_e$, for $R_d > 0$ and $R_d < 0$, respectively, (note that all the terms are dependent on y_{q_i} linearly via x_q) is given by[16]

$$\boldsymbol{E}' \{X_{q_i} \mid \boldsymbol{Y}\} \sim \left. \frac{\partial n F_{\text{par}}}{\partial \lambda_{q_i}} \right|_{\boldsymbol{\lambda}=0} = \xi_{1,q} Y_{q_i} \tag{210}$$

where

$$\xi_{1,q} = -\frac{1}{2} \frac{2\sigma_q'^* \beta}{\left(|\sigma_q'|^2 + \gamma_0\right)^2 C} \sum_{l=1}^{k} \left\{ \frac{P_x}{P_x \gamma_0 + P_x |\sigma_l'|^2 \beta} + B_l - C_l \right\}$$

$$- \frac{1}{2} \left(1 - \frac{2 \left(|\sigma_q'|^2 + \gamma_0\right)^2 P_q^* \big|_{\boldsymbol{\lambda}=0}}{\sqrt{1 + 4\beta^2 |\sigma_q'|^2 \left. \tilde{P}_{y,q} \right|_{\boldsymbol{\lambda}=0} P_q^* \big|_{\boldsymbol{\lambda}=0}}} \right.$$

$$\left. - \frac{2\beta^2 |\sigma_q'|^2 \left. \tilde{P}_{y,q} \right|_{\boldsymbol{\lambda}=0}}{\sqrt{1 + 4\beta^2 |\sigma_q'|^2 \left. \tilde{P}_{y,q} \right|_{\boldsymbol{\lambda}=0} P_q^* \big|_{\boldsymbol{\lambda}=0}}} \right) \frac{2\sigma_q'^* \beta}{\left(|\sigma_q'|^2 + \gamma_0\right)^2}, \tag{211}$$

with

$$B_l = \frac{\left(|\sigma_l'|^2 + \gamma_0\right)^2 - 2 \left(|\sigma_l'|^2 + \gamma_0\right) \left(|\sigma_l'|^2 \beta \left(1 + \beta \tilde{P}_{y,l} \big|_{\boldsymbol{\lambda}=0}\right) + \gamma_0\right)}{\left(|\sigma_l'|^2 + \gamma_0\right)^4} \tag{212}$$

$$C_l = \frac{2\beta^2 |\sigma_l'|^2 \left. \tilde{P}_{y,l} \right|_{\boldsymbol{\lambda}=0} B_l}{\sqrt{1 + 4\beta^2 |\sigma_l'|^2 \left. \tilde{P}_{y,l} \right|_{\boldsymbol{\lambda}=0} P_l^* \big|_{\boldsymbol{\lambda}=0}}}. \tag{213}$$

Next, the MSE estimator in the region $R_g < R \leq R_e$ for $R_d < 0$ is derived. The derivative of the partition function w.r.t. λ_{q_i} is given by

$$\left. \frac{\partial F_{\text{glas}}}{\partial \lambda_{q_i}} \right|_{\boldsymbol{\lambda}=0} = -\beta \left. \frac{\partial \epsilon_s}{\partial \lambda_{q_i}} \right|_{\boldsymbol{\lambda}=0}. \tag{214}$$

Recall that ϵ_s is the solution of the equation

$$\Gamma(\epsilon_s) + R = 0, \tag{215}$$

---

[16]The relation between the right and the left hand sides of (210) is an asymptotic equality between two random variables, in the sense that the difference between them converges to zero w.p. 1.

where $\Gamma\left(\epsilon_s\right)$ is given as

$$\max_{\{P_i\}_{i=1}^k,\{\rho_i\}_{i=1}^k} \frac{h}{2}\sum_{m=1}^k \ln\left(\frac{P_m}{P_x}\left(1-\rho_m^2\right)\right)$$

$$\text{s.t.} \quad \frac{1}{k}\left(\sum_{m=1}^k |\sigma_m'|\,\rho_m\sqrt{P_m\tilde{P}_{y,m}} - \frac{1}{2}|\sigma_m|^2 P_x - \frac{1}{2}|\sigma_m'|^2 P_m - \frac{1}{2\beta}\right) = -\epsilon_s$$

$$\frac{1}{k}\sum_{m=1}^k P_m = P_x. \tag{216}$$

Similarly to the optimization problem in (132), the maximization problem in (216) can be rewritten as

$$\max_{\{P_i\}_{i=1}^k,\{\rho_i\}_{i=1}^k} \frac{h}{2}\sum_{m=1}^k \ln\left\{\frac{P_m}{P_x}\left[1-\left(\frac{\frac{1}{2}|\sigma_m|^2 P_x + \frac{1}{2}|\sigma_m'|^2 P_m + \frac{1}{2\beta} - \mu_m\epsilon_s}{\sigma_m'\sqrt{P_m\tilde{P}_{y,m}}}\right)^2\right]\right\}$$

$$\text{s.t.} \quad \frac{1}{k}\sum_{m=1}^k P_m = P_x,\; \frac{1}{k}\sum_{m=1}^k \mu_m = 1. \tag{217}$$

The derivative of the objective function w.r.t. P_m vanishes at

$$P_m^* = \frac{4\alpha_1\epsilon_s^2 + |\sigma_m'|^2\,\alpha_2\left(\tilde{P}_{y,m}\alpha_2 + 2\epsilon_s\right)}{\left(|\sigma_m'|^2\,\alpha_2 + 2\alpha_1\epsilon_s\right)^2} \tag{218}$$

and the derivative w.r.t. μ_m, vanishes at

$$\mu_m^* = \frac{4\alpha_1^2\epsilon_s^2\left(1 + P_x\beta\,|\sigma_m|^2\right) + 4|\sigma_m'|^2\,\alpha_1\epsilon_s\left(\alpha_2 - \beta\tilde{P}_{y,m}\alpha_2 + \beta\epsilon_s + P_x\beta\alpha_2\,|\sigma_m|^2\right)}{2\beta\epsilon_s\left(|\sigma_m'|^2\,\alpha_2 + 2\alpha_1\epsilon_s\right)^2}$$

$$+ \frac{|\sigma_m'|^4\,\alpha_2\left(\alpha_2 - \beta\tilde{P}_{y,m}\alpha_2 + 2\beta\epsilon_s + P_x\beta\alpha_2\,|\sigma_m|^2\right)}{2\beta\epsilon_s\left(|\sigma_m'|^2\,\alpha_2 + 2\alpha_1\epsilon_s\right)^2} \tag{219}$$

where α_1 is chosen to such that $kP_x = \sum_m P_m^*$, and α_2 is chosen such that $k = \sum_m \mu_m^*$. Substituting the above maximizers in the objective function of (217) one obtains

$$\Gamma\left(\epsilon_s\right) = \frac{h}{2}\sum_{m=1}^k \ln\left(\frac{2\epsilon_s}{P_x\,|\sigma_m'|^2\,\alpha_2 + 2P_x\alpha_1\epsilon_s}\right). \tag{220}$$

Thus, (215) becomes

$$\frac{h}{2}\sum_{m=1}^k \ln\left(\frac{2\epsilon_s}{P_x\,|\sigma_m'|^2\,\alpha_2 + 2P_x\alpha_1\epsilon_s}\right) + R = 0. \tag{221}$$

Let $x_q \triangleq \partial\epsilon_s/\partial\lambda_{q_i}|_{\boldsymbol{\lambda}=0}$, $\dot{\alpha}_{1,q} \triangleq \partial\alpha_1/\partial\lambda_{q_i}|_{\boldsymbol{\lambda}=0}$, $\dot{\alpha}_{2,q} \triangleq \partial\alpha_2/\partial\lambda_{q_i}|_{\boldsymbol{\lambda}=0}$, $\alpha_{1,0} \triangleq \alpha_1|_{\boldsymbol{\lambda}=0}$, $\alpha_{2,0} \triangleq \alpha_2|_{\boldsymbol{\lambda}=0}$ and $\epsilon_{s,0} \triangleq \epsilon_s|_{\boldsymbol{\lambda}=0}$. Differentiating (221) w.r.t. λ_{q_i} one obtains

$$0 = \sum_{m=1}^k \frac{P_x\,|\sigma_m'|^2\,\alpha_{2,0} + 2P_x\alpha_{1,0}\epsilon_{s,0}}{2\epsilon_{s,0}}$$

$$\times \frac{2\left(P_x \left|\sigma'_m\right|^2 \alpha_{2,0} + 2P_x\alpha_{1,0}\epsilon_{s,0}\right) x_q - 2\epsilon_{s,0}\left(P_x \left|\sigma'_m\right|^2 \dot{\alpha}_{2,q} + 2P_x\dot{\alpha}_{1,q}\epsilon_{s,0} + 2P_x\alpha_{1,0}x_q\right)}{\left(P_x \left|\sigma'_m\right|^2 \alpha_{2,0} + 2P_x\alpha_{1,0}\epsilon_{s,0}\right)^2} \tag{222}$$

$$= \sum_{m=1}^{k} \frac{P_x \left|\sigma'_m\right|^2 \alpha_{2,0}x_q - \epsilon_{s,0}\left(P_x \left|\sigma'_m\right|^2 \dot{\alpha}_{2,q} + 2P_x\dot{\alpha}_{1,q}\epsilon_{s,0}\right)}{\epsilon_{s,0}\left(P_x \left|\sigma'_m\right|^2 \alpha_{2,0} + 2P_x\alpha_{1,0}\epsilon_{s,0}\right)}, \tag{223}$$

and thus

$$x_q = \frac{U_1}{U_2} \tag{224}$$

where

$$U_1 \triangleq \sum_{m=1}^{k} \frac{P_x \left|\sigma'_m\right|^2 \dot{\alpha}_{2,q} + 2P_x\dot{\alpha}_{1,q}\epsilon_{s,0}}{P_x \left|\sigma'_m\right|^2 \alpha_{2,0} + 2P_x\alpha_{1,0}\epsilon_{s,0}} \tag{225}$$

and

$$U_2 \triangleq \sum_{m=1}^{k} \frac{P_x \left|\sigma'_m\right|^2 \alpha_{2,0}}{\epsilon_{s,0}\left(P_x \left|\sigma'_m\right|^2 \alpha_{2,0} + 2P_x\alpha_{1,0}\epsilon_{s,0}\right)}. \tag{226}$$

Hence, in order to calculate x_q one needs to find $\epsilon_{s,0}, \alpha_{1,0}, \alpha_{2,0}, \dot{\alpha}_{1,q}, \dot{\alpha}_{2,q}$. The terms $\epsilon_{s,0}, \alpha_{1,0}, \alpha_{2,0}$ are calculated using the set of simultaneous equations

$$\Gamma\left(\epsilon_s\right)|_{\boldsymbol{\lambda}=0} + R = 0 \tag{227a}$$

$$\frac{1}{k} \sum_{m=1}^{k} P_m^*|_{\boldsymbol{\lambda}=0} = P_x \tag{227b}$$

$$\frac{1}{k} \sum_{m=1}^{k} \mu_m^*|_{\boldsymbol{\lambda}=0} = 1, \tag{227c}$$

and accordingly, the terms $\dot{\alpha}_{1,q}, \dot{\alpha}_{2,q}$ are calculated using the set of equations

$$\sum_{m=1}^{k} \left.\frac{\partial P_m^*}{\partial \lambda_{q_i}}\right|_{\boldsymbol{\lambda}=0} = 0 \tag{228a}$$

$$\sum_{m=1}^{k} \left.\frac{\partial \mu_m^*}{\partial \lambda_{q_i}}\right|_{\boldsymbol{\lambda}=0} = 0. \tag{228b}$$

Given $\epsilon_{s,0}, \alpha_{1,0}, \alpha_{2,0}$, closed-form expressions for $\dot{\alpha}_{1,q}, \dot{\alpha}_{2,q}$ are now derived. Using (218), (228a) can be written as

$$\eta_1\dot{\alpha}_{1,q} + \eta_2\dot{\alpha}_{2,q} + \eta_3 x_q + \eta_q \left|\sigma'_q\right|^2 \dot{\tilde{P}}_{y,q} = 0 \tag{229}$$

where $\dot{\tilde{P}}_{y,q} \triangleq \left.\partial \tilde{P}_{y,q}/\partial \lambda_{q_i}\right|_{\boldsymbol{\lambda}=0}$, and

$$\eta_1 \triangleq \sum_{m=1}^{k} \frac{4D_m\epsilon_{s,0}^2 - 4R_m\epsilon_{s,0}}{D_m^3} \tag{230}$$

$$\eta_2 \triangleq \sum_{m=1}^{k} \frac{D_m \left|\sigma'_m\right|^2 \left[\left(\left|\sigma_m\right|^2 P_x + \frac{1}{\beta}\right)\alpha_{2,0} + 2\epsilon_{s,0}\right]}{D_m^3}$$

$$+ \frac{\left|\sigma'_m\right|^2 \alpha_{2,0}\left(\left|\sigma_m\right|^2 P_x + \frac{1}{\beta}\right)D_m - 2R_m \left|\sigma'_m\right|^2}{D_m^3} \tag{231}$$

$$\eta_3 \triangleq \sum_{m=1}^{k} \frac{8D_m \alpha_{1,0}\epsilon_{s,0} + 2D_m \left|\sigma'_m\right|^2 \alpha_{2,0} - 4R_m \alpha_{1,0}}{D_m^3} \tag{232}$$

$$\eta_q \triangleq \frac{\alpha_{2,0}^2}{D_q^2}, \tag{233}$$

in which

$$D_m \triangleq \left|\sigma'_m\right|^2 \alpha_{2,0} + 2\alpha_{1,0}\epsilon_{s,0} \tag{234}$$

$$R_m \triangleq 4\alpha_{1,0}\epsilon_{s,0}^2 + \left|\sigma'_m\right|^2 \alpha_{2,0}\left[\alpha_{2,0}\left(\left|\sigma_m\right|^2 P_x + 2\epsilon_{s,0}\right)\right]. \tag{235}$$

Similarly, using (219), (228b) can be written as

$$\gamma_1 \dot{\alpha}_{1,q} + \gamma_2 \dot{\alpha}_{2,q} + \gamma_3 x_q + \gamma_q \left|\sigma'_q\right|^2 \dot{\tilde{P}}_{y,q} = 0 \tag{236}$$

where

$$\gamma_1 \triangleq \sum_{m=1}^{k} \frac{8\alpha_{1,0}\epsilon_{s,0}^2 \left(1 + P_x \beta \left|\sigma_m\right|^2\right) + 4\beta \left|\sigma'_m\right|^2 \epsilon_{s,0}^2}{K_m}$$

$$- \frac{8T_m \beta \epsilon_{s,0}^2 \left(\left|\sigma'_m\right|^2 \alpha_{2,0} + 2\alpha_{1,0}\epsilon_{s,0}\right)}{K_m^2} \tag{237}$$

$$\gamma_2 \triangleq \sum_{m=1}^{k} \frac{2K_m \beta \epsilon_{s,0} \left|\sigma'_m\right|^4 - 4T_m \beta \epsilon_{s,0} \left(\left|\sigma'_m\right|^2 \alpha_{2,0} + 2\alpha_{1,0}\epsilon_{s,0}\right)\left|\sigma'_m\right|^2}{K_m^2} \tag{238}$$

$$\gamma_3 \triangleq \sum_{m=1}^{k} \frac{8\alpha_{1,0}^2 \epsilon_{s,0}\left(1 + P_x \beta \left|\sigma_m^2\right|\right) + 8\beta\epsilon_{s,0} \left|\sigma'_m\right|^2 \alpha_{1,0} + 2\beta\alpha_{2,0} \left|\sigma'_m\right|^4}{K_m}$$

$$- \frac{T_m \left[2\beta\left(\left|\sigma'_m\right|^2 \alpha_{2,0} + 2\alpha_{1,0}\epsilon_{s,0}\right)^2 + 8\beta\epsilon_{s,0}\alpha_{1,0}\left(\left|\sigma'_m\right|^2 \alpha_{2,0} + 2\alpha_{1,0}\epsilon_{s,0}\right)\right]}{K_m^2} \tag{239}$$

$$\gamma_q \triangleq \frac{-4\beta\alpha_{1,0}\epsilon_{s,0}\alpha_{2,0} - \beta\alpha_{2,0}^2 \left|\sigma'_q\right|^2}{K_q^2}, \tag{240}$$

in which

$$K_m \triangleq 2\beta\epsilon_{s,0}\left(\left|\sigma'_m\right|^2 \alpha_{2,0} + 2\alpha_{1,0}\epsilon_{s,0}\right)^2 \tag{241}$$

$$T_m \triangleq 4\alpha_{1,0}^2 \epsilon_{s,0}^2 \left(1 + P_x \beta \left|\sigma_m\right|^2\right) + 4\beta \left|\sigma'_m\right|^2 \alpha_{1,0}\epsilon_{s,0}^2 + 2 \left|\sigma'_m\right|^4 \alpha_{2,0}\beta\epsilon_{s,0}. \tag{242}$$

Thus, solving the pair of equations, (229) and (236), one obtains

$$\dot{\alpha}_{1,q} = \frac{\gamma_3/\gamma_2 - \eta_3/\eta_2}{\eta_1/\eta_2 - \gamma_1/\gamma_2} x_q + \frac{\gamma_q/\gamma_2 - \eta_q/\eta_2}{\eta_1/\eta_2 - \gamma_1/\gamma_2} \left|\sigma'_q\right|^2 \dot{P}_{y,q} \triangleq r_1 x_q + J_{1q} \left|\sigma'_q\right|^2 \dot{P}_{y,q} \tag{243}$$

$$\dot{\alpha}_{2,q} = \frac{\gamma_3/\gamma_1 - \eta_3/\eta_1}{\eta_2/\eta_1 - \gamma_2/\gamma_1} x_q + \frac{\gamma_q/\gamma_1 - \eta_q/\eta_1}{\eta_2/\eta_1 - \gamma_2/\gamma_1} \left|\sigma'_q\right|^2 \dot{P}_{y,q} \triangleq r_2 x_q + J_{2q} \left|\sigma'_q\right|^2 \dot{P}_{y,q}. \tag{244}$$

Substituting $\dot{\alpha}_{1,q}$ and $\dot{\alpha}_{2,q}$ in (224), simple rearrangement of terms reveals that

$$x_q = \frac{J_{1,q} \sum_{m=1}^{k} 2\epsilon_{s,0}^2 P_x/Q_m + J_{2,q} \sum_{m=1}^{k} 2\epsilon_{s,0} P_x \left|\sigma'_m\right|^2 /Q_m}{V(1-F)} \left|\sigma'_q\right|^2 \dot{P}_{y,q} \tag{245}$$

where

$$V \triangleq \sum_{m=1}^{k} \frac{P_x \left|\sigma'_m\right|^2 \alpha_{2,0}}{\epsilon_{s,0} \left(P_x \left|\sigma'_m\right|^2 \alpha_{2,0} + 2P_x \alpha_{1,0} \epsilon_{s,0} \right)} \tag{246}$$

$$F \triangleq \frac{1}{V} \sum_{m=1}^{k} \frac{\epsilon_{s,0} \left(P_x \left|\sigma'_m\right|^2 r_2 + 2P_x \epsilon_{s,0} r_1 \right)}{\epsilon_{s,0} \left(P_x \left|\sigma'_m\right|^2 \alpha_{2,0} + 2P_x \alpha_{1,0} \epsilon_{s,0} \right)} \tag{247}$$

$$Q_m \triangleq \epsilon_{s,0} \left(P_x \left|\sigma'_m\right|^2 \alpha_{2,0} + 2P_x \alpha_{1,0} \epsilon_{s,0} \right). \tag{248}$$

Let

$$J_q \triangleq \frac{J_{1,q} \sum_{m=1}^{k} 2\epsilon_{s,0}^2 P_x/Q_m + J_{2,q} \sum_{m=1}^{k} 2\epsilon_{s,0} P_x \left|\sigma'_m\right|^2 /Q_m}{V(1-F)}, \tag{249}$$

and so

$$x_q = J_q \left|\sigma'_q\right|^2 \dot{P}_{y,q} = J_q \frac{2\sigma'^*_q}{n_b \beta} y_{q_i}. \tag{250}$$

Therefore,

$$\boldsymbol{E}' \{X_{q_i} \mid \boldsymbol{Y}\} \sim \left. \frac{\partial n F_{\text{glas}}}{\partial \lambda_{q_i}} \right|_{\boldsymbol{\lambda}=0} \tag{251}$$

$$= -n\beta \left. \frac{\partial \epsilon_s}{\partial \lambda_{q_i}} \right|_{\boldsymbol{\lambda}=0} = \xi_{q,2} \cdot y_{q_i}. \tag{252}$$

where

$$\xi_{q,2} \triangleq -J_q \frac{2\sigma'^*_q}{h}. \tag{253}$$

Finally, the mismatched MSE estimator in the region $R \leq R_g$ and $R \leq R_c$ for $R_d < 0$ and $R_d > 0$, respectively, is derived. Based on (83), it readily follows that

$$\boldsymbol{E}' \{X_{q_i} \mid \boldsymbol{Y}\} \sim \left. \frac{\partial \ln Z_{Q,c}}{\partial \lambda_{q_i}} \right|_{\boldsymbol{\lambda}=0} = X_{q_i}. \tag{254}$$

To conclude, the mismatched MSE estimator is given as follows.

For $R_d \geq 0$

$$\boldsymbol{E}' \{X_{q_i} \mid \boldsymbol{Y}\} \sim \begin{cases} X_{q_i}, & R \leq R_c \\ \xi_{q,1} Y_{q_i}, & R > R_c \end{cases}.$$ (255)

For $R_d < 0$

$$\boldsymbol{E}' \{X_{q_i} \mid \boldsymbol{Y}\} \sim \begin{cases} X_{q_i}, & R \leq R_g \\ \xi_{q,2} Y_{q_i}, & R_g < R \leq R_e \\ \xi_{q,1} Y_{q_i}, & R > R_e \end{cases}$$ (256)

where the above equalities are asymptotic equalities between two random variables, in the sense that the difference between them converges to zero in probability.

The mismatched MSE is given by

$$\mathrm{mse}\,(\boldsymbol{X} \mid \boldsymbol{Y}) = \sum_{i=1}^{n} \boldsymbol{E}\left\{X_i^2\right\} - 2\,\mathrm{Re}\left(\boldsymbol{E}\left\{\boldsymbol{E}\left(X_i \mid \boldsymbol{Y}\right) \boldsymbol{E}'^*\left(X_i \mid \boldsymbol{Y}\right)\right\}\right)$$
$$+ \boldsymbol{E}\left\{\left|\boldsymbol{E}'\left(X_i \mid \boldsymbol{Y}\right)\right|^2\right\}.$$ (257)

Therefore, based on (257), in order to calculate the MSE, the MMSE estimator should be obtained first. Substituting $\boldsymbol{A} = \boldsymbol{A}'$ in R_d, given in (168), one can see that $R_d = 0$. Thus, the MMSE estimator is given by

$$\boldsymbol{E}\{X_{q_i} \mid \boldsymbol{Y}\} \sim \begin{cases} \xi_{1,q} Y_{q_i}, & R > R_e \\ X_{q_i}, & R \leq R_e \end{cases}.$$ (258)

In order to find R_e, according to (155), γ_0 is needed. However, in this case it can readily be verified that $\gamma_0 = 1/P_x$, and thus

$$R_{c,M} \triangleq R_e = \frac{1}{4\pi} \int_0^{2\pi} \ln\left(1 + |\mathbf{H}(\omega)|^2 \beta P_x\right) d\omega.$$ (259)

Finally, substitution of $\sigma_m = \sigma_m'$ in (211), reveals that

$$\xi_{1,q} = \frac{\beta \sigma_q^* P_x}{1 + |\sigma_q|^2 P_x \beta},$$ (260)

and thus

$$\boldsymbol{E}\{X_{q_i} \mid \boldsymbol{Y}\} \sim \begin{cases} \frac{\beta \sigma_q^* P_x}{1 + |\sigma_q|^2 P_x \beta} Y_{q_i}, & R > R_{c,M} \\ X_{q_i}, & R \leq R_{c,M} \end{cases}.$$ (261)

Based on the second term of the sum in (257), several cases should be considered. For $R_d > 0$, since $R_c < R_{c,M}$, there are three regions: $R < R_c$, $R_c < R < R_{c,M}$ and $R > R_{c,M}$. For $R < R_c$, both the matched and the mismatched estimators are asymptotically equal to X_{q_i} with high probability, and thus

$$\text{mse}\left(\boldsymbol{X} \mid \boldsymbol{Y}\right) = 0. \tag{262}$$

For $R_c < R < R_{c,M}$ one readily obtains

$$\frac{\text{mse}\left(\boldsymbol{X} \mid \boldsymbol{Y}\right)}{n} = P_x - 2h \operatorname{Re}\left(\sum_{m=1}^{k} \xi_{m,1}^* \sigma_m^* P_x\right) + h \sum_{m=1}^{k} |\xi_{m,1}|^2 \left(|\sigma_m|^2 P_x + \frac{1}{\beta}\right), \tag{263}$$

and similarly, for $R_{c,M} < R$,

$$\frac{\text{mse}\left(\boldsymbol{X} \mid \boldsymbol{Y}\right)}{n} = P_x - 2h \operatorname{Re}\left(\sum_{m=1}^{k} \xi_{m,1}^* \frac{\beta \sigma_m^* P_x}{1 + |\sigma_m|^2 P_x \beta}\left(|\sigma_m|^2 P_x + \frac{1}{\beta}\right)\right)$$

$$+ h \sum_{m=1}^{k} |\xi_{m,1}|^2 \left(|\sigma_m|^2 P_x + \frac{1}{\beta}\right) \tag{264}$$

$$= P_x - 2h \operatorname{Re}\left(\sum_{m=1}^{k} \xi_{m,1}^* \sigma_m^* P_x\right) + h \sum_{m=1}^{k} |\xi_{m,1}|^2 \left(|\sigma_m|^2 P_x + \frac{1}{\beta}\right). \tag{265}$$

Thus, the MSE's in the last two ranges are the same. In the same way, the MSE for $R_d < 0$ is calculated. For $R \leq R_g$

$$\text{mse}\left(\boldsymbol{X} \mid \boldsymbol{Y}\right) = 0. \tag{266}$$

For $R_g < R \leq R_e$

$$\frac{\text{mse}\left(\boldsymbol{X} \mid \boldsymbol{Y}\right)}{n} = P_x - 2h \operatorname{Re}\left(\sum_{m=1}^{k} \xi_{m,2}^* \sigma_m^* P_x\right) + h \sum_{m=1}^{k} |\xi_{m,2}|^2 \left(|\sigma_m|^2 P_x + \frac{1}{\beta}\right) \triangleq \text{mse}_1, \tag{267}$$

and for $R > R_e$

$$\frac{\text{mse}\left(\boldsymbol{X} \mid \boldsymbol{Y}\right)}{n} = P_x - 2h \operatorname{Re}\left(\sum_{m=1}^{k} \xi_{m,1}^* \sigma_m^* P_x\right) + h \sum_{m=1}^{k} |\xi_{m,1}|^2 \left(|\sigma_m|^2 P_x + \frac{1}{\beta}\right) \triangleq \text{mse}_2. \tag{268}$$

Finally, take the limit $h \to 0$ (after $n \to \infty$). Using Szegö's theorem (as was done in (168)), one obtains ($i = 1, 2$)

$$\lim_{n \to \infty} \text{mse}_i = P_x - \frac{P_x}{\pi} \int_0^{2\pi} \operatorname{Re}\left(\boldsymbol{\Xi}_i^*\left(\omega\right) \mathbf{H}^*\left(\omega\right)\right) \mathrm{d}\omega + \frac{1}{2\pi} \int_0^{2\pi} |\boldsymbol{\Xi}_i\left(\omega\right)|^2 \left(|\mathbf{H}\left(\omega\right)|^2 + \frac{1}{\beta}\right) \mathrm{d}\omega \tag{269}$$

where $\boldsymbol{\Xi}_i\left(\omega\right)$, for $i = 1, 2$, are given in (15) and (39).

In the matched case, for $R > R_{c,M}$

$$\text{mmse}\left(\boldsymbol{X} \mid \boldsymbol{Y}\right) = \sum_{i=1}^{n} \boldsymbol{E}\left\{X_i^2\right\} - \boldsymbol{E}\left\{|\boldsymbol{E}\left\{X_i \mid \boldsymbol{Y}\right\}|^2\right\} \tag{270}$$

$$= nP_x - \sum_{m=1}^{k} \sum_{i_m=1}^{n_b} \boldsymbol{E}\left\{|\boldsymbol{E}\left\{X_{i_m} \mid \boldsymbol{Y}\right\}|^2\right\} \tag{271}$$

$$= nP_x - \sum_{m=1}^{k} \left|\frac{\beta \sigma_m^* P_x}{1 + |\sigma_m|^2 P_x \beta}\right|^2 n_b \left(|\sigma_m|^2 P_x + \frac{1}{\beta}\right) \tag{272}$$

$$= nP_x - n \sum_{m=1}^{k} h \frac{|\sigma_m|^2 P_x^2}{\frac{1}{\beta} + |\sigma_m|^2 P_x} = n \sum_{m=1}^{k} h \frac{P_x}{1 + |\sigma_m|^2 P_x \beta}, \tag{273}$$

which upon taking the limit $h \to 0$, becomes

$$\lim_{n\to\infty} \frac{\text{mmse}\left(\boldsymbol{X} \mid \boldsymbol{Y}\right)}{n} = \frac{1}{2\pi} \int_0^{2\pi} \frac{P_x}{1 + |\mathbf{H}\left(\omega\right)|^2 P_x \beta} \mathrm{d}\omega. \tag{274}$$

Remark 4 (Generalization to Any Input Spectral Distribution) As was mentioned in Section III, the above analysis can be modified to hold for any input spectral density $S_x\left(\omega\right)$. Technically speaking, the following modification should be considered: Let $P_{x,m}$ be the (real) transmitted power over the mth bin. Then, because of the separable form of the partition function over the bins, we will essentially obtain exactly the same results with the exception of $P_{x,m}$ instead of P_x. Precisely, instead of P_x which appears in the numerator of the logarithm function in (133), one should simply replace it to $P_{x,m}$. Following the same lines of derivation, at the final stage of the refinement of the bin sizes, we will finally obtain the spectral density $S_x\left(\omega\right)$ as a limit function of $\{P_{x,m}\}_m$.

VII. CONCLUSION

In this paper, we considered the problem of mismatched estimation of codewords corrupted by a Gaussian vector channel. The derivation was build upon a simple relation between the MSE and a certain function, which can be viewed as a partition function, and hence be analyzed using methods of statistical mechanics. As a special case, the MMSE estimator and its respective estimation error was derived. In particular, it was shown that the MSE essentially separated into two cases each exhibiting a different behavior: In one case, the MSE exhibits single phase transition, which divides the MSE into ferromagnetic and paramagnetic phases. In the other case, the MSE exhibits two phase transitions, which divide the MSE into three phases consisting of the two previous phases and a third glassy phase. Then, using the theoretical results obtained, a few numerical examples were analyzed, by exploring the phase diagrams and the MSE's as functions of the mismatched parameters in each problem. This leads to physical intuitions regarding the threshold effects and the role of the mismatched measure in creating them. Indeed, it was shown that the aforementioned separation of the MSE is linked to pessimism and optimism behaviors of the receiver, according to its mismatched assumption on the channel. Note that in

contrast to previous related papers [5, 6], in which the explored examples did not completely emphasize the necessity of the use of the analysis techniques of statistical physics for deriving the MSE, we believe that the considered problem in this paper does, as standard information theoretic approaches do not lend themselves to rigorous analysis. Finally, we believe that the tools developed in this paper for handling optimum estimation problems, can be used in other applications. One such application, which has been already considered for a simple model is estimation of signals of partial support [6, Section V. D] which has motivation in compressed sensing applications. It would be natural to generalize the model considered in [6, Section V. D] to a much more rich and applicable one (in the spirit of the considered model in this paper), and perhaps assessing the MSE using the concepts developed in this paper.

Appendix A
Proof of Lemma 1

Proof: We first show the inclusion

$$\mathcal{T}_\delta \left(\boldsymbol{x} \mid \boldsymbol{y} \right) \subseteq \hat{\mathcal{T}}_\delta \left(\boldsymbol{x} \mid \boldsymbol{y} \right), \tag{A.1}$$

namely, for any $\boldsymbol{x} \in \mathcal{T}_\delta \left(\boldsymbol{x} \mid \boldsymbol{y} \right)$ also $\boldsymbol{x} \in \hat{\mathcal{T}}_\delta \left(\boldsymbol{x} \mid \boldsymbol{y} \right)$. Recall that

$$\mathscr{B}_m^\delta \left(P_m, \rho_m \right) \triangleq \left\{ \boldsymbol{x} \in \mathbb{R}^{n_b} : \left| \left\| \boldsymbol{x}_{(m-1)n_b+1}^{mn_b} \right\|^2 - n_b P_m \right| \leq \delta, \right.$$

$$\left. \left| \mathrm{Re} \left\{ \sum_{i \in \mathcal{I}_m} \sigma_i' \bar{y}_i x_i \right\} - n_b \rho_m \sqrt{\tilde{P}_{y,m} \tilde{P}_{\sigma,m}} \right| \leq \delta \right\}, \tag{A.2}$$

and that

$$\mathcal{T}_\delta \left(\boldsymbol{x} \mid \boldsymbol{y} \right) \triangleq \left\{ \boldsymbol{x} \in \mathbb{R}^n : \left| \| \boldsymbol{x} \|^2 - n P_x \right| \leq \delta, \left| \frac{\| \boldsymbol{y} - \boldsymbol{\Sigma}' \boldsymbol{x} \|^2}{2} - \frac{\boldsymbol{\lambda}^T \boldsymbol{x}}{\beta} - n \epsilon \right| \leq \delta \right\}. \tag{A.3}$$

First, note that the second constraint in (A.3) can be rewritten as

$$\left| \mathrm{Re} \left\{ \frac{1}{n} \sum_{i=1}^n \sigma_i' \bar{y}_i x_i \right\} - \rho \right| \leq \delta \tag{A.4}$$

where

$$\tilde{\rho} = \frac{\frac{1}{n} \sum_{i=1}^n |\sigma_i' x_i|^2 + P_y - 2\epsilon}{2}. \tag{A.5}$$

Then, for any $\boldsymbol{x} \in \mathcal{T}_\delta \left(\boldsymbol{x} \mid \boldsymbol{y} \right)$, we first show that there exist a sequence $\{P_m\}_{m=1}^k \in \mathcal{P}^\delta$ such that for any $1 \leq m \leq k$,

$$\left| \left\| \boldsymbol{x}_{(m-1)n_b+1}^{mn_b} \right\|^2 - n_b P_m \right| \leq \delta. \tag{A.6}$$

To this end, for each $1 \leq m \leq k$, P_m is chosen to be the nearest point to $\left\|\boldsymbol{x}^{mn_b}_{(m-1)n_b+1}\right\|^2$ in the set $\mathcal{G}^k_{1,\delta}$, namely $P_m = \left\lfloor \left\|\boldsymbol{x}^{mn_b}_{(m-1)n_b+1}\right\|^2 / (n_b\delta) \right\rfloor \cdot \delta$. Under this choice, obviously, (A.6) holds, and $\{P_m\}^k_{m=1} \in \boldsymbol{\mathcal{P}}^\delta$, since

$$\left| \frac{1}{k}\sum_{m=1}^k P_m - P_x \right| = \left| \frac{1}{k}\sum_{m=1}^k \left\lfloor \frac{\left\|\boldsymbol{x}^{mn_b}_{(m-1)n_b+1}\right\|^2}{n_b\delta} \right\rfloor \delta - P_x \right| \tag{A.7}$$

$$\leq \left| \frac{1}{n}\sum_{m=1}^k \left\|\boldsymbol{x}^{mn_b}_{(m-1)n_b+1}\right\|^2 \delta - P_x \right| \leq \delta \tag{A.8}$$

where the last equality follows from the fact that $\boldsymbol{x} \in \mathcal{T}_\delta(\boldsymbol{x} \mid \boldsymbol{y})$. Next, we show that there exist a sequence $\{\rho_m\}^k_{m=1} \in \mathcal{R}^\delta_{\mathcal{P}}$ such that for any $1 \leq m \leq k$,

$$\left| \mathrm{Re}\left\{ \sum_{i \in \mathcal{I}_m} \sigma'_i \bar{y}_i x_i \right\} - n_b \rho_m \sqrt{\tilde{P}_{y,m}\tilde{P}_{\sigma,m}} \right| \leq \delta. \tag{A.9}$$

Similarly, by taking

$$\rho_m = \left\lfloor \frac{\mathrm{Re}\left\{ \sum_{i \in \mathcal{I}_m} \sigma'_i \bar{y}_i x_i \right\}}{n_b\delta\sqrt{\tilde{P}_{y,m}\tilde{P}_{\sigma,m}}} \right\rfloor \cdot \delta \in \mathcal{G}^k_{2,\delta}, \tag{A.10}$$

obviously, (A.9) holds, and also $\{P_m, \rho_m\} \in \boldsymbol{\mathcal{P}}^\delta \cap \mathcal{R}^\delta_{\mathcal{P}}$, since

$$\left| \frac{1}{k}\sum_{m=1}^k \rho_m \sqrt{\tilde{P}_{y,m}\tilde{P}_{\sigma,m}} - \tilde{\rho} \right| = \left| \frac{1}{k}\sum_{m=1}^k \left\lfloor \frac{\mathrm{Re}\left\{ \sum_{i \in \mathcal{I}_m} \sigma'_i \bar{y}_i x_i \right\}}{n_b\delta\sqrt{\tilde{P}_{y,m}\tilde{P}_{\sigma,m}}} \right\rfloor \cdot \delta\sqrt{\tilde{P}_{y,m}\tilde{P}_{\sigma,m}} - \tilde{\rho} \right| \tag{A.11}$$

$$\leq \left| \frac{1}{n}\sum_{m=1}^k \mathrm{Re}\left\{ \sum_{i \in \mathcal{I}_m} \sigma'_i \bar{y}_i x_i \right\} - \tilde{\rho} \right| \tag{A.12}$$

$$= \left| \mathrm{Re}\left\{ \frac{1}{n}\sum_{i=1}^n \sigma'_i \bar{y}_i x_i \right\} - \rho \right| \leq \delta \tag{A.13}$$

where the last equality follows from the fact that $\boldsymbol{x} \in \mathcal{T}_\delta(\boldsymbol{x} \mid \boldsymbol{y})$. For the second inclusion, we need to show that $\hat{\mathcal{T}}_{\delta/k}(\boldsymbol{x} \mid \boldsymbol{y}) \subseteq \mathcal{T}_\delta(\boldsymbol{x} \mid \boldsymbol{y})$. For any $\boldsymbol{x} \in \hat{\mathcal{T}}_{\delta/k}(\boldsymbol{x} \mid \boldsymbol{y})$

$$\left| \|\boldsymbol{x}\|^2 - nP_x \right| = \left| \sum_{m=1}^k \left\|\boldsymbol{x}^{mn_b}_{(m-1)n_b+1}\right\|^2 - nP_x \right|$$

$$= \left| \sum_{m=1}^k \left\|\boldsymbol{x}^{mn_b}_{(m-1)n_b+1}\right\|^2 - \sum_{m=1}^k n_b P_m \right|$$

$$\leq \sum_{m=1}^k \left| \left\|\boldsymbol{x}^{mn_b}_{(m-1)n_b+1}\right\|^2 - n_b P_m \right| \leq k\frac{\delta}{k} = \delta \tag{A.14}$$

where the second equality follows from the definition of $\mathcal{P}^\delta$, the third inequality follows from the triangle inequality, and the forth inequality follows from the definition of $\mathscr{B}_m^\delta(P_m, \rho_m)$. In the same way, for any $\boldsymbol{x} \in \hat{\mathcal{T}}_{\delta/k}(\boldsymbol{x} \mid \boldsymbol{y})$

$$\left| \text{Re} \left\{ \sum_{i=1}^n \sigma_i' \bar{y}_i x_i \right\} - n\rho \sqrt{P_{\bar{y}} \left(\frac{1}{n} \sum_{i=1}^n |\sigma_i' x_i|^2 \right)} \right| = \left| \sum_{m=1}^k \text{Re} \left\{ \sum_{i \in \mathcal{I}_m} \sigma_i' \bar{y}_i x_i \right\} - n_b \sum_{m=1}^k \rho_m \sqrt{\tilde{P}_{y,m} \tilde{P}_{\sigma,m}} \right|$$

$$\leq k \frac{\delta}{k} = \delta \tag{A.15}$$

where the first equality follows from the definition of $\mathcal{R}_{\mathcal{P}}^\delta$, and the second inequality follows from the triangle inequality and the definition of $\mathscr{B}_m^\delta(P_m, \rho_m)$. Thus $\hat{\mathcal{T}}_{\delta/k}(\boldsymbol{x} \mid \boldsymbol{y}) \subseteq \mathcal{T}_\delta(\boldsymbol{x} \mid \boldsymbol{y}) \subseteq \hat{\mathcal{T}}_\delta(\boldsymbol{x} \mid \boldsymbol{y})$. ∎

APPENDIX B
PROOF OF LEMMA 2

Proof: For simplicity of notation, the following conventions are used. Calculating the volume of $\mathscr{B}_m^\delta(P_m, \rho_m)$ is equivalent to calculating the volume of the set

$$\mathcal{F}_\delta(P_x, \rho) \triangleq \left\{ \boldsymbol{x} \in \mathbb{R}^n : \left| \|\boldsymbol{x}\|^2 - nP_x \right| \leq \delta, \left| \text{Re} \left\{ \sum_{i=1}^n x_i y_i^* \right\} - n\rho \sqrt{P_x P_y} \right| \leq \delta \right\} \tag{B.1}$$

where $P_x \triangleq \|\boldsymbol{x}\|^2 / n$ and $P_y \triangleq \|\boldsymbol{y}\|^2 / n$, for a given vector $\boldsymbol{y} \in \mathbb{C}^n$. Due to the symmetry of the vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ in the DFT domain (recall that in the time domain the considered vectors are real), i.e., $x_i = x_{n-i}^*$ for $i = 1, \ldots, n$ (and similarly for $\boldsymbol{y}$), for the volume calculation of (B.1), only vectors with dimension $n/2$ should be considered, while the other half is fixed. Accordingly, the constraints in (B.1) take the form

$$\left| \sum_{i=1}^{n/2} |x_i|^2 - \frac{n}{2} P_x \right| \leq \delta, \tag{B.2}$$

and

$$\left| \text{Re} \left\{ \sum_{i=1}^{n/2} x_i y_i^* \right\} - \frac{n}{2} \rho \sqrt{P_x P_y} \right| \leq \delta. \tag{B.3}$$

Let $m \triangleq n/2$. Consider the following Gaussian measure

$$\text{d}\gamma_G^m(\boldsymbol{x}) \triangleq \frac{1}{(\pi \vartheta^2)^m} \exp \left\{ -\frac{1}{\vartheta^2} \sum_{i=1}^m |x_i - ay_i|^2 \right\} \text{d}\boldsymbol{x} \tag{B.4}$$

where $a, \vartheta^2 \in \mathbb{R}$. Then,

$$1 = \gamma_G^m \{ \mathbb{R}^m \} \geq \gamma_G^m \{ \mathcal{F}_\delta(P_x, \rho) \} \tag{B.5}$$

$$= \int_{\mathcal{F}_\delta} \frac{1}{(\pi\vartheta^2)^m} \exp\left\{ -\frac{1}{\vartheta^2} \sum_{i=1}^m |x_i - ay_i|^2 \right\} \mathrm{d}\boldsymbol{x} \tag{B.6}$$

$$\geq \int_{\mathcal{F}_\delta} \frac{1}{(\pi\vartheta^2)^m} \exp\left\{ -\frac{m}{\vartheta^2} \left[P_x + \delta - 2a\sqrt{P_x P_y}\,(\rho - \delta) + a^2 P_y \right] \right\} \mathrm{d}\boldsymbol{x} \tag{B.7}$$

$$= \mathrm{Vol}\left\{ \mathcal{F}_\delta\,(P_x, \rho) \right\} \frac{1}{(\pi\vartheta^2)^m} \exp\left\{ -\frac{m}{\vartheta^2} \left[P_x + \delta - 2a\sqrt{P_x P_y}\,(\rho - \delta) + a^2 P_y \right] \right\}. \tag{B.8}$$

It is easy to verify that

$$a_o \triangleq \sqrt{\frac{P_x}{P_y}}\,(\rho - \delta), \tag{B.9}$$

and

$$\vartheta_o^2 \triangleq P_x + \delta - 2a\sqrt{P_x P_y}\,(\rho - \delta) + a^2 P_y \tag{B.10}$$

$$= P_x + \delta - P_x\,(\rho - \delta)^2 \tag{B.11}$$

maximize the right hand side of (B.8) (w.r.t. a and ϑ^2). Thus, on the one hand,

$$\mathrm{Vol}\left\{ \mathcal{F}_\delta\,(P_x, \rho) \right\} \leq \exp\left\{ m \ln\left(\pi e \vartheta_o^2 \right) \right\}. \tag{B.12}$$

On the other hand,

$$1 = \gamma_G^m \left\{ \mathcal{F}_\delta\,(P_x, \rho) \cup \mathcal{F}_\delta^c\,(P_x, \rho) \right\} \tag{B.13}$$

$$= \int_{\mathcal{F}_\delta} \frac{1}{(\pi\vartheta^2)^m} \exp\left\{ -\frac{1}{\vartheta^2} \sum_{i=1}^m |x_i - ay_i|^2 \right\} \mathrm{d}\boldsymbol{x} + \gamma_G^m \left\{ \mathcal{F}_\delta^c\,(P_x, \rho) \right\} \tag{B.14}$$

$$\leq \mathrm{Vol}\left\{ \mathcal{F}_\delta\,(P_x, \rho) \right\} \exp\left\{ -m \ln\left(\pi e \vartheta_{o,u}^2 \right) \right\} + \gamma_G^m \left\{ \mathcal{F}_\delta^c\,(P_x, \rho) \right\} \tag{B.15}$$

where the last inequality follows by the same considerations as before, and

$$\vartheta_{o,u}^2 = P_x - \delta - P_x\,(\rho + \delta)^2. \tag{B.16}$$

Using Boole's inequality

$$\gamma_G^m \left\{ \mathcal{F}_\delta^c\,(P_x, \rho) \right\} \leq \gamma_G^m \left\{ \boldsymbol{x} : \left| \|\boldsymbol{x}\|^2 - mP_x \right| > \delta \right\} + \gamma_G^m \left\{ \boldsymbol{x} : \left| \mathrm{Re}\left\{ \sum_{i=1}^n x_i y_i^* \right\} - m\rho\sqrt{P_x P_y} \right| > \delta \right\}. \tag{B.17}$$

It is easy to verify that the parameters a and ϑ that are maximizing the Gaussian measure are given by

$$a_M = \frac{\mathrm{Re}\left(\sum_{i=1}^m x_i y_i^* \right)}{P_y} \triangleq \frac{\tilde{\rho}}{P_y} \tag{B.18}$$

$$\vartheta_M^2 = \frac{1}{m} \sum_{i=1}^m |x_i|^2 - \frac{\mathrm{Re}\left(\sum_{i=1}^m x_i y_i^* \right)^2}{P_y} \triangleq \tilde{P}_x - \frac{\tilde{\rho}^2}{P_y} \tag{B.19}$$

where $\tilde{\rho}$ and $\tilde{P}_x$ are the empirical correlation and the input variance, respectively. Let $\gamma_{G,M}$ denote the Gaussian measure associated with the parameters a_M, ϑ_M, namely, $\gamma_{G,M}$ is given by (B.4) with $a = a_M$ and $\vartheta^2 = \vartheta_M^2$. Accordingly, it is easy to verify that under $\gamma_{G,M}$, the following hold

$$\boldsymbol{E}_{\gamma_{G,M}} \left\{ \|\boldsymbol{X}\|^2 \right\} = mP_x, \tag{B.20}$$

and

$$\boldsymbol{E}_{\gamma_{G,M}} \left\{ \mathrm{Re}\left[\sum_{i=1}^n X_i y_i^* \right] \right\} = m\sqrt{P_x P_y} \rho. \tag{B.21}$$

Thus, using the LLN, the two terms on the right hand side of (B.17) are negligible as $m \to \infty$, namely,

$$\gamma_G^m \left\{ \mathcal{F}_\delta^c \left(P_x, \rho \right) \right\} \le \epsilon \tag{B.22}$$

for any $\epsilon > 0$. Thus,

$$\mathrm{Vol}\left\{ \mathcal{F}_\delta \left(P_x, \rho \right) \right\} \ge (1 - \epsilon) \exp\left\{ m \ln\left(\pi e \vartheta_{o,u}^2 \right) \right\}. \tag{B.23}$$

Finally, combining (B.12), (B.23), and taking the limit $\delta \to 0$, the lemma follows. ∎

APPENDIX C
PROOF OF LEMMA 4

Proof: Recall that

$$\boldsymbol{E}\left\{ \mathcal{N}\left(\epsilon \right) \right\} \doteq \exp\left\{ n\left(R + \Gamma\left(\epsilon \right) \right) \right\}, \tag{C.1}$$

and that [17]

$$\mathrm{var}\left\{ \mathcal{N}\left(\epsilon \right) \right\} \doteq \exp\left\{ n\left(R + \Gamma\left(\epsilon \right) \right) \right\} \left(1 - \exp\left\{ n\Gamma\left(\epsilon \right) \right\} \right). \tag{C.2}$$

Thus,

$$\frac{\mathrm{var}\left\{ \mathcal{N}\left(\epsilon \right) \right\}}{\left(\boldsymbol{E}\left\{ \mathcal{N}\left(\epsilon \right) \right\} \right)^2} \doteq \exp\left\{ -n\left(R + \Gamma\left(\epsilon \right) \right) \right\}. \tag{C.3}$$

For any $\epsilon \notin \mathscr{E}$, the expectation of $\mathcal{N}\left(\epsilon \right)$ can be written as $\boldsymbol{E}\left\{ \mathcal{N}\left(\epsilon \right) \right\} \doteq e^{-nC_1}$ where $C_1 = R + \Gamma\left(\epsilon \right) > 0$. Thus, by Markov inequality (since $\mathcal{N}\left(\epsilon \right) \in \mathbb{N} \cup \{0\}$)

$$\mathbb{P}\left\{ \mathcal{N}\left(\epsilon \right) > 0 \right\} \le \boldsymbol{E}\left\{ \mathcal{N}\left(\epsilon \right) \right\} \doteq e^{-nC_1}. \tag{C.4}$$

[17]Given $\boldsymbol{y}$, $\mathcal{N}\left(\epsilon \right)$ is a sum of $M - 1$ i.i.d. Bernoulli random variables and therefore its variance is $(M - 1)p(1 - p)$, where p is the success probability, which in our case, was shown to be given by $p = \exp\left\{ n\Gamma\left(\epsilon \right) \right\}$.

On the other hand, for any $\epsilon \in \mathscr{E}$ and $\delta > 0$, using Chebyshev's inequality

$$\mathbb{P}\left\{\left|\frac{\mathcal{N}(\epsilon)}{\boldsymbol{E}\{\mathcal{N}(\epsilon)\}} - 1\right| > \delta\right\} \leq \frac{\text{var}\{\mathcal{N}(\epsilon)\}}{\gamma(\boldsymbol{E}\{\mathcal{N}(\epsilon)\})^2} \doteq e^{-nC_2} \tag{C.5}$$

where $C_2 = R + \gamma(\epsilon) > 0$. Thus, in this case, $\mathcal{N}(\epsilon)$ is concentrated very strongly around $\boldsymbol{E}\{\mathcal{N}(\epsilon)\}$. Finally, let $\mathcal{A}_n \overset{\triangle}{=} \{|\mathcal{N}(\epsilon) - \boldsymbol{E}\{\mathcal{N}(\epsilon)\}\mathbb{1}\{\mathscr{E}\}| > \delta\}$. Then, using (C.4) and (C.5), it is easy to verify that

$$\sum_{i=1}^{\infty} \mathbb{P}(\mathcal{A}_n) < \infty. \tag{C.6}$$

Thus, using Borel-Cantelli Lemma, one obtains that

$$\mathbb{P}\left\{\limsup_{n \to \infty} \mathcal{A}_n\right\} = 0, \tag{C.7}$$

and hence (124) follows. ∎

APPENDIX D
PROOF OF (128)

Equation (128) follows by the following lemma.

Lemma 5 Let $f : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ be a smooth function such that

$$g(x) = \lim_{h \to a} f(x, h), \tag{D.1}$$

uniformly for every $x \in \mathbb{R}$. Assume that $\lim_{h \to a} \max_x f(x, h)$ exist. Then,

$$\lim_{h \to a} \max_x f(x, h) = \max_x \lim_{h \to a} f(x, h). \tag{D.2}$$

Proof of Lemma 5: Let

$$\lambda \overset{\triangle}{=} \lim_{h \to a} \max_x f(x, h), \tag{D.3}$$

and

$$g(x_0) \overset{\triangle}{=} \max_x g(x) = \max_x \lim_{h \to a} f(x, h). \tag{D.4}$$

Based on (D.3), $\forall \epsilon_1 > 0$ there exist $\delta_1 > 0$ such that

$$\left|\max_x f(x, h) - \lambda\right| < \epsilon_1 \tag{D.5}$$

whenever $0 < h - a < \delta_1$. Accordingly, by (D.1), $\forall \epsilon_2 > 0$ there exist $\delta_2 > 0$ such that

$$|f(x, h) - g(x)| < \epsilon_2 \tag{D.6}$$

whenever $0 < h - a < \delta_2$. Let us assume by contradiction that (without loss of generality)

$$\Delta \stackrel{\triangle}{=} |g(x_0) - \lambda| > 0. \tag{D.7}$$

However, by using the triangle inequality, one obtains that

$$0 < \Delta = |g(x_0) - \lambda| \leq \left| g(x_0) - \max_x f(x, h) \right| + \left| \max_x f(x, h) - \lambda \right|, \tag{D.8}$$

and hence

$$0 < \Delta - \epsilon_1 \leq \left| g(x_0) - \max_x f(x, h) \right| \leq |g(x_0) - f(x_0, h)|, \tag{D.9}$$

for $0 < h - a < \min(\delta_1, \delta_2)$, which contradicts the assumption in (D.1) (or (D.6)). Thus, $\delta = 0$. ∎

Remark 5 As the proof shows, Lemma 5 remains valid for functions $f : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$.

In our case, the assumptions of Lemma 5 hold true: the uniform convergence is due to the absolutely (square) summability of the sequence $\{h_k\}$ and Szegö's theorem, and the existence the limit over the maximization problem indeed exists as was obtained. Thus, the order of limit over h and the maximization over ϵ in (128) can be interchanged.

APPENDIX E

DERIVATION OF (168)

Szegö's theorem [36-39] basically states that, for a sequence of Toeplitz matrices $\boldsymbol{T}_n = \{t_{i-j}\}_{i,j}$ with dimension $n \times n$, for which $\{t_k\}$ is absolutely (square) summable, the following holds

$$\lim_{n \to \infty} \frac{1}{n} \sum_{k=0}^{n-1} F(\tau_{n,k}) = \frac{1}{2\pi} \int_0^{2\pi} F(\mathbf{T}(\omega)) \, \mathrm{d}\omega \tag{E.1}$$

where $\{\tau_{n,k}\}_k$ are the eigenvalues of $\boldsymbol{T}_n$, $\mathbf{T}(\omega)$ is the Fourier transform of $\{t_k\}$, and $F(\cdot)$ is some polynomial function. Furthermore, if $\boldsymbol{T}_n$ are Hermitian, then (E.1) holds true for any continuous function $F(\cdot)$.

In our case, however, the matrices $\boldsymbol{A}$ and $\boldsymbol{A}'$ are not necessarily Hermitian. Nevertheless, based on (165), it can be seen that the dependency of the various non-linear terms (except the third term) on the eigenvalues is only via $|\sigma'_m|^2$, which can be regarded as eigenvalues of the Hermitian matrix $\boldsymbol{A}^H \boldsymbol{A}$, and so Szegö's theorem can be applied. Regarding the third term in the right hand side of (165), it can be shown [38] that a product of Toeplitz matrices also satisfies Szegö's theorem, namely,

$$\lim_{n \to \infty} \frac{1}{n} \sum_{k=0}^{n-1} F(\rho_{n,k}) = \frac{1}{2\pi} \int_0^{2\pi} F(\mathbf{T}(\omega) \mathbf{S}(\omega)) \, \mathrm{d}\omega \tag{E.2}$$

where $\{\rho_{n,k}\}_k$ are the eigenvalues of product of the Toeplitz matrices, $\boldsymbol{T}_n\boldsymbol{S}_n$, and $\mathbf{T}\left(\omega\right)$ and $\mathbf{S}\left(\omega\right)$ are the respective Fourier transforms. Accordingly, since the third term in (165) is originated from a product of Toeplitz matrices (162), (E.2) can be used. Therefore, a direct application of (E.1) and (E.2) on (165), we finally obtain (168). Finally, note that these considerations are utilized to justify the other places in the paper (for example, (195) and (269)) in which Szegö's theorem is applied.

REFERENCES

[1] R. S. Bucy, "Information and filtering," *Inf. Sci.*, vol. 18, pp. 179–187, 1979.

[2] T. E. Duncan, "On the calculation of mutual information," *SIAM J. Appl. Math.*, vol. 19, no. 1, pp. 215–220, 1970.

[3] T. Kailath, "The innovations approach to detection ans estimation theory," *Proc. IEEE*, vol. 58, no. 5, pp. 680–695, May 1970.

[4] J. Seidler, "Bounds on the mean-square error and the quality of domain decisions based on mutual information," *IEEE Trans. Inf. Theory*, vol. IT-17, no. 6, pp. 655–665, Nov. 1971.

[5] N. Merhav, "Optimum estimation via gradients of partition functions and information measures: A statistical-mechanical perspective," *IEEE Trans. Inf. Theory*, vol. 57, no. 6, pp. 3887–3898, June 2011.

[6] N. Merhav, D. Guo, and S. Shamai, "Statistical physics of signal estimation in Gaussian noise: theory and examples of phase transitions," *IEEE Trans. Inf. Theory*, vol. 56, no. 3, pp. 1400–1416, Mar. 2010.

[7] D. Guo, "Relative entropy and score function: New information-estimation relationships through arbitrary additive perturbations," *presented at the Int. Symp. Information Theory, Seoul, South Korea*, Jun./Jul. 2009.

[8] D. Guo, S. Shamai, and S. Verdú, "Mutual information and minimum mean-square error in Gaussian channels," *IEEE Trans. Inf. Theory*, vol. 51, no. 4, pp. 1261–1282, Apr. 2005.

[9] ——, "Additive non-Gaussian noise channels: Mutual information and conditional mean estimation," in *Proc. IEEE Int. Symp. Inf. Theory*. Adelaide, Australia, Sep. 2005, pp. 719–723.

[10] ——, "Mutual information and conditional mean estimation in Poisson channels," *IEEE Trans. Inf. Theory*, vol. 54, no. 5, pp. 1837–1849, May 2008.

[11] D. P. Palomar and S. Verdú, "Gradient of mutual information in linear vector gaussian channels," *IEEE Trans. Inf. Theory*, vol. 52, no. 1, pp. 141–154, Jan. 2006.

[12] ——, "Representation of mutual information via input estimates," *IEEE Trans. Inf. Theory*, vol. 53, no. 2, pp. 453–470, Feb. 2007.

[13] M. Raginsky and T. P. Coleman, "Mutual information and posterior estimates in channels of exponential family type," in *Proc. IEEE Workshop Inf. Theory*. Taormina, Italy, Oct. 2009, pp. 399–403.

[14] S. Verdú, "Mismatched estimation and relative entropy," *IEEE Trans. Inf. Theory*, vol. 56, no. 8, pp. 3712–3720, Aug. 2010.

[15] T. Weissman, "The relationship between causal and non-causal mismatched estimation in continuous-time AWGN channels," *IEEE Trans. Inf. Theory*, vol. 56, no. 9, pp. 4256–4273, Sep. 2006.

[16] A. Atar and T. Weissman, "Mutual information, relative entropy, and estimation in the Poisson channel," *IEEE Trans. Inf. Theory*, vol. 58, no. 3, pp. 1302–1318, Mar. 2012.

[17] R. Bustin and S. Shamai, "MMSE of "Bad" coeds," *IEEE Trans. Inf. Theory*, vol. 59, no. 2, pp. 733–743, Feb. 2013.

[18] N. Sourlas, "Sping-glass models as error-correcting codes," *Nature*, vol. 339, pp. 693–695, Jan. 1989.

[19] ——, "Spin glasses, error-correcting codes and finite-temperatures," *Europhys. Lett.*, vol. 25, pp. 159–164, 1994.

[20] F. Cousseau, K. Mimura, T. Omori, and M. Okada, "Statistical mechanics of lossy compression for non-monotonic multilayer perceptrons," *Phys. Rev. E*, vol. 78, p. 021124, Jul. 2008.

[21] T. Hosaka and Y. Kabashima, "Statistical mechanical approach to error exponents of lossy data compression," *J. Phys. Soc. Japan*, vol. 74, no. 1, pp. 488–497, Jan. 2005.

[22] Y. Iba, "The Nishimori line and Bayesian statistics," *J. Phys. A: Math. Gen*, vol. 32, pp. 3875–3888, 1999.

[23] Y. Kabashima and T. Hosaka, "Statistical mechanics for source coding with a fidelity criterion," *Progr. Theoret. Phys.*, pp. 197–204, 2005.

[24] Y. Kabashima, K. Nakamura, and J. vanMourik, "Statistical mechanics of typical set decoding," *Phys. Rev. E*, vol. 66, pp. 197–204, 2002.

[25] Y. Kitagawa and T. Tanaka, "Optimal spreading sequences in large CDMA systems: A statistical mechanics approach," in *Proc. Int. Symp. Information Theory*, vol. 13731377. Toronto, ON, Canada, Jul. 20, 2008.

[26] T. Tanaka, "A statistical-mechanics approach to large-system analysis of CDMA multiuser detectors," *IEEE Trans. Inf. Theory*, vol. 48, no. 11, pp. 2888–2910, Nov. 2002.

[27] T. Weissman, A. Ordentlich, G. Seroussi, S. Verdú, and J. M. Weinberger, "Universal discrete denoising: Known channel," *IEEE Trans. Inf. Theory*, vol. 51, no. 1, pp. 5–28, Jan. 2005.

[28] G. Gemelos, S. Sigurjonsson, and T. Weissman, "Universal minimax discrete denoising under channel uncertainty," *IEEE Trans. Inf. Theory*, vol. 52, no. 8, pp. 3476–3497, Aug. 2006.

[29] S. Jalali and T. Weissman, "Denoising via MCMC-based lossy compression," *IEEE Trans. Sig. Process.*, vol. 60, no. 6, pp. 3092–3100, Jun. 2012.

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

[31] D. Donoho, "On minimum entropy deconvolution," *Applied time series analysis*, pp. 565–608, 1981.

[32] M. Fozunbal, "On regret of parametric mismatch in minimum mean square error estimation," in *Proc. IEEE Int. Symp. Inf. Theory*. Austin, Texas, U.S.A., June 2010, pp. 1408–1412.

[33] A. Mézard, M. Montanari, *Information, Physics and Computation*. Oxford, U.K.: Oxford Univ. Press., 2009.

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

[35] A. D. Wyner, "A bound on the number of distinguishable functions which are time-limited and," *SIAM J. Appl. Math.*, vol. 24, no. 3, pp. 289–297, May 1973.

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

[37] 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.

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

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

[40] W. Rudin, *Principles of Mathematical Analysis*. 3rd ed. New York: McGraw-Hill, 1976.

[41] A. Dembo and O. Zeitouni, *Large Deviations Techniques and Applications*. Springer, 1998.

[42] G. Galavotti, *Statistical Mechanics: A Short Treatise*. Springer Verlag, New York, 1999.

[43] V. A. Zorich, *Mathematical Analysis II*.   Springer, 2009.

[44] O. J. Frink, "Differentiation of sequences," *Amer. Math. Soc.*, vol. 41, pp. 553–560, Dec. 1934.

[45] R. C. Steinlage, "Nearly uniform convergence and interchange of limits," *Publications de l'Institut Mathématique*, vol. 26, no. 12, pp. 115–129, June 1971.