Finite-size correction and variance of the mutual information of random linear estimation with non-Gaussian priors: A replica calculation

Theodoros G. Tsironis and Aris L. Moustakas

Department of Physics, National Kapodistrian University of Athens, Athens, Greece
and Archimedes/Athena RC, Athens, Greece

(Received 2 November 2023; revised 3 April 2024; accepted 15 April 2024; published 6 June 2024)

Random linear vector channels have been known to increase the transmission of information in several communications systems. For Gaussian priors, the statistics of a key metric, namely, the mutual information, which is related to the free energy of the system, have been analyzed in great detail for various types of channel randomness. However, for the realistic case of non-Gaussian priors, only the average mutual information has been obtained in the asymptotic limit of large channel matrices. In this paper, we employ methods from statistical physics, namely, the replica approach, to calculate the finite-size correction and the variance of the mutual information with non-Gaussian priors, both for the case of correlated Gaussian and uncorrelated non-Gaussian channel matrices in the same asymptotic limit. Furthermore, using the same methodology, we show that higher order cumulants of the mutual information should vanish in the large-system-size limit. In addition, we obtain closed-form expressions for the minimum mean-square error finite-size corrections and variance for both Gaussian and non-Gaussian channels. Finally, we provide numerical verification of the results using numerical methods on finite-sized systems.

DOI: 10.1103/PhysRevE.109.064114

I. INTRODUCTION

Random linear vector channels have long been seen as a standard paradigm in analyzing inference. For example, their study has revolutionized the wireless communications industry in the last 30 years, initially with the introduction of CDMA (code-division multiple access), in which symbols from different transmitters are spread in time (or frequency) through pseudorandom code vectors, followed by multi-antenna array processing using multiple input, multiple output [1,2]. Other applications include random linear regression and compressed sensing [3–6]. More recently, the ideas stemming from random linear vector channels have been generalized to machine learning and neural networks [7–10].

A random linear vector channel model can be compactly defined in the following way:

$$\bar{\mathbf{y}} = \frac{1}{\sqrt{M}}\mathbf{H}\mathbf{x} + \bar{\mathbf{z}}. \tag{1}$$

Given an M-dimensional input vector $\mathbf{x}$, the above equation returns an N-dimensional output vector $\bar{\mathbf{y}}$, linearly related to $\mathbf{x}$ through the random matrix $\mathbf{H}$ in the presence of a (typically Gaussian) noise $\bar{\mathbf{z}}$. In (1), $\mathbf{H}$ can be seen either as a channel matrix with $\mathbf{x}$ being the input signal vector in the context of communications or, equivalently, as a matrix of N covariate data vectors of size M multiplying the unknown vector $\mathbf{x}$ to be estimated in the case of linear regression. For concreteness, in the remainder of the paper, we will refer to $\mathbf{H}$ as channel matrix.

The Bayesian inference problem then can be expressed using the conditional probability

$$P(\mathbf{x}|\bar{\mathbf{y}},\mathbf{H}) = \frac{P(\mathbf{x})P(\bar{\mathbf{y}}|\mathbf{x},\mathbf{H})}{Z(\bar{\mathbf{y}}|\mathbf{H})}, \tag{2}$$

where $P(\mathbf{x})$ is the distribution of priors and $P(\bar{\mathbf{y}}|\mathbf{x},\mathbf{H})$ is the noise distribution. The normalization $Z(\bar{\mathbf{y}}|\mathbf{H}) = \mathbb{E}_{\mathbf{x}}[P(\bar{\mathbf{y}}|\mathbf{x},\mathbf{H})]$ plays the role of the partition function of the effective spins $\mathbf{x}$, in the presence of quenched interactions and external field given, respectively, by $\mathbf{H}$ and $\bar{\mathbf{y}}$. The obvious analogy to a spin glass has allowed for fruitful applications of ideas from statistical physics.

A key metric is the mutual information, defined as

$$I(X,\bar{Y}|\mathbf{H}) = -\mathbb{E}_{\bar{\mathbf{y}}}[\ln(Z(\bar{\mathbf{y}}|\mathbf{H})] + \mathbb{E}_{\mathbf{x},\bar{\mathbf{y}}}[\ln P(\bar{\mathbf{y}}|\mathbf{x},\mathbf{H})], \tag{3}$$

where the first term corresponds to $h(\bar{Y}|\mathbf{H})$, the entropy of the output $\bar{\mathbf{y}}$ and the second to (minus) the entropy of the noise, $h(\bar{Y}|X,\mathbf{H})$. The mutual information represents the maximum information rate achievable for a given prior distribution $P(\mathbf{x})$. The so-called ergodic mutual information, i.e., the expectation of the mutual information with respect to $\mathbf{H}$, $\mathbb{E}_{\mathbf{H}}[I(X,\bar{Y}|\mathbf{H})]$ corresponds to the average free energy of the system, up to the second term above, which is a constant.

The case of the Gaussian distribution of priors has attracted significant attention in the information theoretic community, since, in addition to the fact that it maximizes the mutual information, in this case $I(X,\bar{Y}|\mathbf{H})$ takes a closed-form expression [1,2,11]. This allowed the application of analytic approaches based on random matrix theory (RMT) for the evaluation of the ergodic mutual information in the large system limit, where the size of the input and output vectors grow indefinitely, but at a fixed ratio [11–13].

More realistically (and perhaps more interestingly), the prior distribution of the elements of $\mathbf{x}$ is discrete, taken from a constellation of points on the complex plane, the simplest one being proportional to ± 1. In this latter case, the model is very similar to a spin-glass system, where RMT does not suffice to tackle the problem. As a result, the analysis for such systems was based on the replica method [14], which has been applied with great success to calculate the asymptotic ergodic mutual information and other relevant quantities in communications systems [15–19], compressed sensing [3–5], and, more recently, in the context of machine learning [7–10], where the channel matrix and output vector correspond to data and their labels, respectively, and the signal corresponds to the weights of a nonlinear neuron function. The results were based on the assumption that the solution exhibits replica symmetry. Nevertheless, more recently the above replica-symmetric expression for the average mutual information was rigorously proven to be correct in a quite general setting [20–22], including rotationally invariant random channel matrices [23,24].

It is important to point out that in the context of communications, the randomness in the external channel matrix $\mathbf{H}$ has a different timescale of variation compared to $\bar{\mathbf{z}}$. Specifically, $\mathbf{H}$ may be assumed to be constant over the period of a data packet transmission (typically $\gtrsim$100 msec), while $\bar{\mathbf{z}}$ (and therefore $\bar{\mathbf{y}}$) changes many times over that period and hence can be averaged out. Therefore, the quantity of interest is $I(X, \bar{Y}|\mathbf{H})$. In contrast, the fluctuations of $I(X, \bar{Y}|\mathbf{H})$ with respect to $\mathbf{H}$ signify its variation between different data packet transmissions and hence are measurable and have a physical significance. Similarly, in the case of linear regression, the added noise is unknown and hence have to be averaged out first, for the fixed covariates given. Thus, in this case higher moments of the mutual information with respect to $\mathbf{H}$ are necessary to characterize its fluctuations in a fading environment, where the channel elements may vary due to multiple scattering. In particular, the variance is especially important, since higher order cumulants are known to vanish for Gaussian channels and priors in the large system limit [25,26], thus making the distribution of the mutual information asymptotically Gaussian. The calculation of the second moment of the mutual information for Gaussian inputs by Refs. [25–27] allowed for the evaluation of the outage probability, an important metric for fading channels, which corresponds to the probability that a data-packet encoded with a given rate cannot be decoded at the output due to channel variations.

In this paper, we apply the replica approach to obtain a closed-form expression for the variance of the mutual information for a subclass of correlated Gaussian vector channels with arbitrary non-Gaussian priors. We also calculate the variance and the $O(1)$ correction to the mean of the mutual information when the channel is non-Gaussian with uncorrelated entries. The analysis is valid in the large system-size limit, for which we also argue that all higher moments vanish, thus resulting to an asymptotically Gaussian distribution for the mutual information.

The practical relevance of these results is evident given that, as mentioned above, prior distributions are typically non-Gaussian in actual communications systems, for which it is important to evaluate the fluctuations of the mutual information due to channel variations. In addition, the analysis of non-Gaussian vector channels is particularly relevant for lattice-valued vector channels as in CDMA transmission, as well as in the presence of erasures, in which a number of channel coefficients becomes negligible due to obstacles (shadowing). In the related setting of in high-dimensional regression, the case of non-Gaussian, heavy-tailed vector channels was recently analyzed in Ref. [28]. Furthermore, explicit calculations of the variance of the free energy in spin-glass-type problems in the replica-symmetric phase is relatively scarce [29,30] and hence our methodology may provide additional intuition in this context. Finally, this approach is also applicable to generalized linear vector channels, which have been used in the context of machine learning and neural networks [8–10].

Outline

In the next section, we define the system model and summarize the analytical results in the paper. In Sec. III, we provide the mathematical framework and describe the main steps that lead us to the results for Gaussian-distributed matrix elements, while Sec. IV generalizes the results for non-Gaussian matrix elements. In Sec. V, we analyze the resulting equations providing representative solutions as well as some numerical validations for small matrix sizes, and in Sec. VI we conclude.

II. MODEL DEFINITION

We consider an M-transmitter (input) and N-receiver (output) random linear vector channel system, for which the channel equation is given by (1) with independent and identically distributed noise $\bar{\mathbf{z}}$, with elements having complex unit variance. We assume that the channel matrix $\mathbf{H}$ can be written as

$$\mathbf{H} = \mathbf{C}^{1/2}\mathbf{G}, \tag{4}$$

where $\mathbf{C}$ is the correlation matrix at the output side, with the normalization condition that $\mathrm{tr}\mathbf{C} = N$, and where the matrix $\mathbf{G}$ will occasionally be expressed in terms of its rows, i.e., $\mathbf{G} = [\mathbf{g}_1, \mathbf{g}_2, \dots, \mathbf{g}_N]^T$. $\mathbf{C}$ can be diagonalized through a unitary matrix $\mathbf{U}$, such that $\mathbf{C} = \mathbf{U}^\dagger\mathbf{\Lambda}\mathbf{U}$, where $\mathbf{\Lambda}$ is a diagonal matrix with the eigenvalues of $\mathbf{C}$, $\lambda_i \geqslant 0$, for $i = 1, \dots, N$, on the diagonal. In this paper, we will treat two distinct cases of random $\mathbf{H}$. First, we will assume that $\mathbf{H}$ is complex Gaussian for general $\mathbf{C}$. Subsequently, we will treat the case of uncorrelated entries of $\mathbf{H}$ and hence assume diagonal $\mathbf{C}$, i.e., $\mathbf{U} = \mathbf{I}$, but for general distributions of the entries of each row, characterized only by their fourth cumulant κ_{4j}. We also assume that the entries have a finite $(4 + \epsilon)$-moment, so the necessary convergence is guaranteed for the Edgeworth expansion, which we employ. Note that this assumption is optimal in terms of moment conditions but, in general, it may be further tightened. For exposition purposes, we will discuss here the general case, specializing in the two cases when necessary.

Assuming that the output side knows the channel matrix $\mathbf{H}$, it is equivalent to analyze the vector $\mathbf{y} = \mathbf{U}\bar{\mathbf{y}}$, with elements given by

$$\mathbf{y} = \frac{1}{\sqrt{M}}\mathbf{\Lambda}^{1/2}\mathbf{U}\mathbf{G}\mathbf{x} + \mathbf{z}, \tag{5}$$

where $\mathbf{z} = \mathbf{U}\bar{\mathbf{z}}$. Furthermore, we assume that the elements of $\mathbf{x}$, namely, x_j, for $j = 1, \dots, M$, follow the distribution $p_j(x_j)$, with zero-mean, variance ρ_j, zero covariance between its real and imaginary parts, and finite higher moments. Here, ρ_j corresponds to the signal-to-noise (SNR) ratio of input j.

The mutual information of the system $I(X, Y|\mathbf{H})$ can then be expressed as (3). Given independence, the input distribution can be written as a product, i.e., $P(\mathbf{x}) = \prod_j p_j(x_j)$, while the Gaussian noise distribution is given by

$$P(\mathbf{y}|\mathbf{x}, \mathbf{H}) = \frac{e^{-|\mathbf{y}-\frac{1}{\sqrt{M}}\mathbf{\Lambda}^{1/2}\mathbf{U}\mathbf{G}\mathbf{x}|^2}}{\pi^N}. \tag{6}$$

It can be easily shown that $h(Y|X, \mathbf{H}) = N\ln(\pi e)$, independently of $\mathbf{H}$. We will express the output entropy in the form

$$h(Y|\mathbf{H}) = -\int d\mathbf{y} Z(\mathbf{y}|\mathbf{H}) \ln Z(\mathbf{y}|\mathbf{H}), \tag{7}$$

where the output distribution is expressed as

$$Z(\mathbf{y}|\mathbf{H}) = \sum_{\mathbf{x}} P(\mathbf{x}) P(\mathbf{y}|\mathbf{x}, \mathbf{H}). \tag{8}$$

A closely related metric to the mutual information [31] is the minimum mean square error (MMSE), also known as the Bayesian error

$$\mathrm{MMSE}(\mathbf{H}) = \mathbb{E}_{\mathbf{y}}[|\mathbf{x} - \hat{\mathbf{x}}(\mathbf{y}, \mathbf{H})|^2], \tag{9}$$

where $\mathbf{x}_0$ is the input signal, while $\hat{\mathbf{x}}(\mathbf{y}, \mathbf{H})$ is the posterior mean estimator:

$$\hat{\mathbf{x}}(\mathbf{y}, \mathbf{H}) = \sum_{\mathbf{x}} \mathbf{x} P(\mathbf{x}|\mathbf{y}, \mathbf{H}). \tag{10}$$

A. Summary of results

1. Definitions

As emphasized in Ref. [16], underlying much of the analysis here is the observation that in the large-system limit the output corresponding to each input element can be described through an *auxiliary* Gaussian effective noise channel $p(y|x; \bar{q})$ given by

$$p(y|x; \bar{q}) = \frac{\bar{q}}{\pi} e^{-\bar{q}|y-x|^2}, \tag{11}$$

with variance $1/\bar{q}$. Thus, the marginal distribution of y for input j is then expressed as

$$p_j(y; \bar{q}) = \sum_x p_j(x) p(y|x; \bar{q}). \tag{12}$$

Averages over this distribution will appear as $\mathbb{E}_{y,j}[\cdot]$. Similarly, the conditional probability for input j given y takes the form

$$p_j(x|y; \bar{q}) = \frac{p_j(x) e^{-\bar{q}|y-x|^2}}{\sum_{x'} p_j(x') e^{-\bar{q}|y-x'|^2}}, \tag{13}$$

with an expectation over this distribution denoted as $\mathbb{E}_j[\cdot|y; \bar{q}]$. As a result, the posterior mean or Bayesian estimator $\hat{x}(y; \bar{q})$ can be expressed as

$$\hat{x}_j(y; \bar{q}) = \sum_x x p_j(x|y; \bar{q}) = \mathbb{E}_j[x|y; \bar{q}]. \tag{14}$$

The mutual information of the channel takes the form

$$I_{y,x_j}(\bar{q}) = -\mathbb{E}_j[\ln p_j(y; \bar{q})|y; \bar{q}] - \ln\left(\frac{e\pi}{\bar{q}}\right). \tag{15}$$

The corresponding posterior variances and covariances of the input can be expressed as

$$v_{1j}(y; \bar{q}) = \mathbb{E}_j[\mathrm{Re}(x - \hat{x}_j(y; \bar{q}))^2|y; \bar{q}], \tag{16a}$$

$$v_{2j}(y; \bar{q}) = \mathbb{E}_j[\mathrm{Im}(x - \hat{x}_j(y; \bar{q}))^2|y; \bar{q}], \tag{16b}$$

$$\begin{aligned} v_{3j}(y; \bar{q}) = \mathbb{E}_j[(&\mathrm{Re}(x)\mathrm{Im}(x) - \mathrm{Re}(\hat{x}_j(y; \bar{q})) \\ &\times \mathrm{Im}(\hat{x}_j(y; \bar{q})))|y; \bar{q}]. \end{aligned} \tag{16c}$$

Thus, the mean square error of the posterior mean estimator of input j takes the form

$$\begin{aligned} \varepsilon_j &= \mathbb{E}_{y,j}[\mathbb{E}_j[|x - \hat{x}_j(y; \bar{q})|^2|y; \bar{q}]] \\ &= \mathbb{E}_{y,j}[v_{1j}(y; \bar{q}) + v_{2j}(y; \bar{q})]. \end{aligned} \tag{17}$$

As described in Refs. [16,18], we are interested in this above channels in the specific case where $\bar{q}$ is related to the mean square error averaged over inputs

$$\varepsilon = \frac{1}{M} \sum_{j=1}^{M} \varepsilon_j \tag{18}$$

through

$$\bar{q} = \frac{1}{M} \sum_{i=1}^{N} \frac{\lambda_i}{1 + \varepsilon\lambda_i}. \tag{19}$$

In addition, under mild conditions, the expectation of the normalized mutual information (or average free energy) $\mathbb{E}_{\mathbf{H}}[I(X, Y|\mathbf{H})]/M$ has been proven rigorously to converge in the large size limit as follows:

Theorem (asymptotic normalized mutual information [20–22]). Consider the function Φ_{RS} defined through

$$\Phi_{RS}(\varepsilon, \bar{q}) = \left(\frac{1}{M} \sum_{i=1}^{N} \ln(1 + \lambda_i\varepsilon) - \varepsilon\bar{q} + \frac{1}{M} \sum_{j=1}^{M} I_{y,x_j}(\bar{q})\right). \tag{20}$$

As long as Φ has at most three stationary points, then in the limit $M \to \infty$, with fixed $\beta = \frac{M}{N}$, the expectation of the mutual information takes the form

$$\lim_{M\to\infty} \mathbb{E}_{\mathbf{H}}\left[\frac{I(X, Y|\mathbf{H})}{M}\right] = C, \tag{21}$$

where C is given by

$$C = \inf_{\varepsilon \leqslant \rho} \sup_{\bar{q}} \Phi_{\mathrm{RS}}(\varepsilon, \bar{q}). \tag{22}$$

The above theorem provides strong evidence for the validity of replica symmetry, and thus we expect our results to be valid under the same conditions as (21). The expression for C together with (18) and (19) was first obtained in Ref. [18], which generalized results in Refs. [15,16] to correlated channels. Given that the fixed point equations may (and indeed sometimes do) have multiple solutions, the $\inf_\varepsilon \sup_{\bar{q}}$ condition in (22) guarantees that we keep the correct solution, which corresponds to the one with the lowest value of Φ_{RS}. The existence of multiple solutions is tied to the existence of metastable states, which leads to first-order transitions. Nevertheless, for all relevant input distributions, from the point of view of communications, the conditions of the above theorem hold.

We may also define the following related quantities:

$$m_1 = \frac{1}{M}\sum_{i=1}^{N}\frac{\lambda_i^2}{(1+\lambda_i\varepsilon)^2}, \tag{23}$$

$$m_2 = \frac{1}{M}\sum_{j=1}^{M}\varepsilon_j^2, \tag{24}$$

and

$$\Delta_j = 2\mathbb{E}_{y,j}[v_{1j}(y;\bar{q})^2 + v_{2j}(y;\bar{q})^2 + 2v_{3j}(y;\bar{q})^2]. \tag{25}$$

Finally, letting

$$\Gamma_j(y;\bar{q}) = 2\mathbb{E}_j\left[\sum_{\substack{m,n=0\\ m>n}}^{3}|x_m - x_n|^2(x_0-x_2)^*(x_1-x_3)\mathrm{Re}(x_0^*x_1)\middle| y;\bar{q}\right],$$

where $\{x_0,\ldots,x_3\}$ are random variables drawn independently from the distribution $p_j(x|y;\bar{q})$ in (13), and subsequently defining

$$\Gamma_j = \mathbb{E}_{y,j}\Gamma_j(y;\bar{q}), \tag{26}$$

we are ready to state our results.

2. Main results

Our results come in the form of results based on the replica method. As discussed in the beginning of Sec. II, we will present results for two cases, namely, when $\mathbf{G}$ is Gaussian with a left-correlation matrix $\mathbf{C}$ and when $\mathbf{G}$ is independent and identically distributed, but generally non-Gaussian entries. In both cases, we assume that the first and third moments of each entry vanish, that their second moment is unity, and their fourth cumulant is given by κ_{4j} (which vanishes in the Gaussian case).

Result 1 (bias of the mutual information). In the limit $M\to\infty$, with fixed $\beta=\frac{M}{N}$, the expectation of the mutual information takes the form

$$\lim_{M\to\infty}[\mathbb{E}_{\mathbf{H}}[I(X,Y|\mathbf{H})] - MC] = -\lim_{M\to\infty}\frac{m_1}{2M}\sum_{j=1}^{M}\kappa_{4j}\Delta_j, \tag{27}$$

where C is given by (22).

It can be seen that for Gaussian channel matrices, there is no finite-size correction to the mutual information, a result that has been known to hold for the case of Gaussian inputs [25]. The next result relates to the variance of the mutual information.

Result 2 (variance of mutual information). In the limit $M\to\infty$, with converging $\beta=\frac{M}{N}$, the variance of the mutual information is given by

$$\lim_{M\to\infty}\mathrm{var}_{\mathbf{H}}[I(X,Y|\mathbf{H})] = \lim_{M\to\infty}\left[-\ln(1-m_1m_2) + \frac{m_1}{M}\sum_{j=1}^{M}\kappa_{4j}\varepsilon_j^2\right]. \tag{28}$$

Moving to higher cumulants, the following result establishes that all higher cumulants of the mutual information vanish in the large-system-size limit, therefore making the distribution of the mutual information asymptotically Gaussian:

Result 3 (higher order cumulants of mutual information). In the limit $M\to\infty$, with converging $\beta=\frac{M}{N}$, all cumulants of the mutual information beyond the variance vanish.

In Appendix B, we use the methodology applied by Ref. [25] for the case of Gaussian inputs, to obtain a controlled, asymptotic expansion of higher order terms around the saddle-point solution, in particular, showing that the skewness of the mutual information vanishes in the large-system limit. This approach can be readily generalized for all higher cumulants. As a result, the distribution of the mutual information becomes asymptotically Gaussian in this limit. In addition, it should be noted that the above methodology can be applied to find the higher order corrections to the mutual information and the variance, both of which also vanish in the same limit (see Ref. [25]).

Regarding the minimum mean-square error, its asymptotic variance can be obtained from the following:

Result 4 (variance of MMSE). In the limit $M\to\infty$, with converging $\beta=\frac{M}{N}$, the variance of the normalized Bayes optimum error is given by

$$\lim_{M\to\infty}[\mathrm{var}_{\mathbf{H}}[\mathrm{MMSE}(\mathbf{H})]] = \lim_{M\to\infty}\left[V_G + \frac{m_1}{M}\sum_{j=1}^{M}\kappa_{4j}\Delta_j^2\right], \tag{29}$$

where V_G is the MMSE variance for Gaussian matrices, given by

$$V_G = \frac{m_1}{1-m_1m_2}\frac{1}{M}\sum_{j=1}^{M}\Delta_j^2 + \left(\frac{m_1}{1-m_1m_2}\frac{1}{M}\sum_{j=1}^{M}\Delta_j\varepsilon_j\right)^2.$$

It is worth pointing out that when all x_j are identically distributed, the expression for V_G reduces to

$$V_G = \frac{m_1}{(1-m_1m_2)^2}\Delta^2. \tag{30}$$

Finally, as in the case of the mutual information, the MMSE has a nonvanishing finite-size correction in the case of non-Gaussian channels, as seen by the following:

Result 5 (bias of the MMSE). In the limit $M\to\infty$, with fixed $\beta=\frac{M}{N}$, the expectation of the MMSE takes the form

$$\lim_{M\to\infty}[\mathbb{E}_{\mathbf{H}}[\mathrm{MMSE}(\mathbf{H})] - \mathrm{M}\varepsilon] = -\lim_{M\to\infty}\frac{m_1}{2M}\sum_{j=1}^{M}\kappa_{4j}\Gamma_j. \tag{31}$$

The derivation for all above results in the case of Gaussian matrices ($\kappa_4 = 0$) appears in Sec. III. The corrections due to non-Gaussian channel matrices are analyzed in Sec. IV.

3. Relation to previous results

It is worth noting that the above expressions reduce to known results for Gaussian input distribution, when the channel matrix is non-Gaussian with kurtosis κ_4. In this case, the average mean square error for input j in (17) takes the simple form

$$\varepsilon_{jg} = \frac{\rho_j}{1+\rho_j\bar{q}} \tag{32}$$

for $j = 1, \dots, M$, and with $\bar{q}$ given by (19), with the subscript g indicating the Gaussian prior case, while the corresponding values of the parameters ε_g, m_{1g}, m_{2g} are obtained, applying the above to (18) and (23). As a result, the correction to the expectation of the mutual information takes exactly the same form as before, namely,

$$\lim_{M\to\infty}[\mathbb{E}[I(X,Y|\mathbf{H})] - MC_g] = -\kappa_4 m_{1g} m_{2g}, \tag{33}$$

where the expression for the average mutual information per input element,

$$C_g = \frac{1}{M}\sum_{i=1}^{N}\ln(1+\lambda_i\varepsilon) + \frac{1}{M}\sum_{j=1}^{M}\ln(1+\rho_j\bar{q}) - \bar{q}\varepsilon_g, \tag{34}$$

has been rederived in the past using various methods [25–27]. In addition, the variance of the mutual information becomes

$$\mathrm{var}[I(X,Y|\mathbf{H})] = -\ln(1-m_{1g}m_{2g}) + \kappa_4 m_{1g} m_{2g}. \tag{35}$$

The above expressions coincide with corresponding results in Ref. [27] when the channel statistics are the same, namely, $\mathbb{E}[H_{ia}H^*_{jb}] = \delta_{ij}\delta_{ab}\lambda_i$.

III. THE REPLICA TRICK: GAUSSIAN CHANNELS

In this section, we will introduce the main mathematical tools of the paper. We will initially assume that the channel matrix $\mathbf{G}$ is Gaussian. In the next section, we will generalize the analysis to non-Gaussian matrices and calculate the corresponding corrections. Before we proceed, we present a bird's eye view of our main calculation.

The central role in what follows is played by a function of two real variables $\mathcal{F}(\mu_1, \mu_2)$, which is constructed so $\mathrm{var}[I] = -\partial_{\mu_1}\partial_{\mu_2}\mathcal{F}(1, 1)$, as described in (39). Since $\mathcal{F}$ is hard to calculate at arbitrary (μ_1, μ_2), we assume that it is equal to the "obvious" analytical continuation of its restriction on $\mathbb{Z}^2$, which we will calculate. This is the content of one of the replica method assumptions, which will be detailed below.

To obtain an expression for $\mathcal{F}(\mu_1, \mu_2)$ with both arguments being positive integers, in Sec. III A we take advantage of the Gaussianity of the channel and the introduction of delta functions to get an integral expression of the form of (55), namely, $e^{-\mathcal{F}} = \int \mathcal{D}\mathcal{X} e^{-M\Phi(\mathcal{X})}$, where Φ is a function of two matrices $\mathbf{S}$, $\bar{\mathbf{S}}$, referred to jointly for compactness as $\mathcal{X}$, of dimension equal to $\mu_1 + \mu_2$. We evaluate this integral asymptotically, using the saddle-point method for matrices $\mathbf{S}$ and $\bar{\mathbf{S}}$. The validity of the replica symmetry, as proven in Refs. [20–22], physically means that there is a single relevant pure state governing the large-system behavior of the system. As a result, this necessitates that saddle-point values $\mathbf{S}^*$ and $\bar{\mathbf{S}}^*$ are replica-metric with respect to its indices. The identification of the dominating saddle-point and matrices $\mathbf{S}^*$ and $\bar{\mathbf{S}}^*$ is performed in Sec. III B and corresponds to the application of Varadhan's lemma and Cramer's theorem as in Refs. [15,16]. We shall then see that the derivative of $\mathcal{F}$ that interests us is of subleading order in the exponential integral and so we need to keep corrections to its asymptotic calculation. Obtaining these corrections requires analyzing the fluctuations of Φ around the replica symmetric saddle point, which is performed in Sec. III C.

A. The generating function

To evaluate the moments of the mutual information (3), one needs to perform the integrals over $\mathbf{y}$ and the average over $\mathbf{H}$ in (7). This is generally difficult due to the existence of $\ln Z(\mathbf{y}|\mathbf{H})$ in the integrand. To tackle this problem, the *replica* trick is devised, in which the log-normal generating function $\mathcal{F}(\mu)$ is introduced,

$$\mathcal{F}(\mu) = -\ln\left\{\int d\mathbf{y}\,\mathbb{E}_{\mathbf{H}}\left[\frac{Z(\mathbf{y}|\mathbf{H})^{\mu}}{(\pi e)^{N(1-\mu)}}\right]\right\}, \tag{36}$$

where the denominator has been introduced to include the noise entropy. The derivative of the above quantity with respect to μ at $\mu = 1$ is the quantity of interest, here the mutual information:

$$\left.\frac{\partial\mathcal{F}}{\partial\mu}\right|_{\mu=1} = \mathbb{E}_{\mathbf{H}}[I(X,Y|\mathbf{H})]. \tag{37}$$

The basic prescription of this approach is to evaluate $\mathcal{F}(\mu)$ at integer values of μ and then analytically continue the result in the vicinity of $\mu = 1$, so as to be able to take the derivative in (36). The proof that such an analytic continuation is valid is usually not possible. However, in a number of similar problems it has recently been shown that the results obtained through the replica method are exact [32,33]. To obtain the variance of the mutual information, we need to generalize the above formula to

$$\mathcal{F}(\mu_1,\mu_2) = -\ln\left\{\iint d\mathbf{y}_1 d\mathbf{y}_2\,\mathbb{E}_{\mathbf{H}}\left[\frac{Z(\mathbf{y}_1|\mathbf{H})^{\mu_1}Z(\mathbf{y}_2|\mathbf{H})^{\mu_2}}{(\pi e)^{N(2-\mu_1-\mu_2)}}\right]\right\}. \tag{38}$$

Then

$$\left.\frac{\partial^2\mathcal{F}(\mu_1,\mu_2)}{\partial\mu_1\partial\mu_2}\right|_{\mu_1=\mu_2=1} = -\mathrm{var}_{\mathbf{H}}[I(X,Y|\mathbf{H})] \tag{39}$$

and, of course,

$$\left.\frac{\partial\mathcal{F}(\mu_1,1)}{\partial\mu_1}\right|_{\mu_1=1} = \mathbb{E}_{\mathbf{H}}[I(X,Y|\mathbf{H})]. \tag{40}$$

As a result, in what follows we will make use of the following ansatz:

Assumption 1 (replica assumption). $\mathcal{F}(\mu_1, \mu_2)$ evaluated at positive integer values of μ_1, μ_2 can be analytically continued for positive real values, specifically in the neighborhood of $\mu_1 = \mu_2 = 1$.

In the above expressions, when μ_1 is a positive integer, the term $Z(\mathbf{y}_1|\mathbf{H})^{\mu_1}$ can be expressed in an explicit way as follows:

$$Z(\mathbf{y}_1|\mathbf{H})^{\mu_1} = \frac{1}{\pi^{N\mu_1}} \sum_{\mathbf{X}_1} P(\mathbf{X}_1) e^{-\sum_{\alpha=1}^{\mu_1} |\mathbf{y}_1 - \frac{\mathbf{\Lambda}^{1/2}}{\sqrt{M}} \mathbf{G}\mathbf{x}_1^\alpha|^2}, \tag{41}$$

where we have introduced the matrix $\mathbf{X}_1$, with elements x_{1j}^α for $j = 1, \ldots, M$ and $\alpha = 1, \ldots, \mu_1$, with columns denoted, respectively, by $\mathbf{x}_1^\alpha$ for $\alpha = 1, \ldots, \mu_1$ and $\mathbf{x}_{1j}$ for $j = 1, \ldots, M$, and we have overloaded the definition of $P(\cdot)$ to denote also $P(\mathbf{X}_1) = \prod_\alpha \prod_j p_j(x_{1j}^\alpha)$. At a later stage, we will generalize the expression of $\mathcal{F}$ by using a slightly different distribution of the $\mathbf{x}_1^\alpha$, namely,

$$P(\mathbf{X}_1) \to P_{h_1}(\mathbf{X}_1) = P(\mathbf{X}_1) e^{-\frac{h_1}{2} \sum_{\alpha \neq \beta} |\mathbf{x}_1^\alpha - \mathbf{x}_1^\beta|^2}. \tag{42}$$

$P(\mathbf{X}_2)$ will similarly correspond to the distribution of the $M \times \mu_2$ matrix $\mathbf{X}_2$. For compactness, sometimes we will use the notation $\mathbf{X} = [\mathbf{X}_1, \mathbf{X}_2]$. The resulting function will be denoted as $\mathcal{F}(\mu_1, \mu_2; h_1, h_2)$ for concreteness. In this case, one can also see that

$$\frac{\partial}{\partial \mu_1} \frac{\partial \mathcal{F}(\mu_1, 1; h_1, 0)}{\partial h_1}\bigg|_{\mu_1 = 1; h_1 = 0} = \mathbb{E}_{\mathbf{H}}[\mathrm{MMSE}(\mathbf{H})] \tag{43}$$

while

$$\frac{\partial^2}{\partial \mu_1 \partial \mu_2} \frac{\partial^2 \mathcal{F}(\mu_1, \mu_2; h_1, h_2)}{\partial h_1 \partial h_2}\bigg|_{\substack{\mu_1 = \mu_2 = 1 \\ h_1 = h_2 = 0}} = -\mathrm{var}_{\mathbf{H}}[\mathrm{MMSE}(\mathbf{H})]. \tag{44}$$

For simplicity, we start with the case $h_1 = h_2 = 0$. We may now express $\mathcal{F}$ as

$$\begin{aligned} \mathcal{F} = &-\ln \Bigg\{ \sum_{\mathbf{X}_1 \mathbf{X}_2} P(\mathbf{X}_1) P(\mathbf{X}_2) \\ &\times \mathbb{E}_{\mathbf{G}} \Bigg[\int \frac{d\mathbf{y}_1}{\pi^N} e^{-\sum_{\alpha_1=1}^{\mu_1} |\mathbf{y}_1 - \frac{1}{\sqrt{M}} \mathbf{\Lambda}^{1/2} \mathbf{U} \mathbf{G} \mathbf{x}_1^{\alpha_1}|^2 + N(\mu_1 - 1)} \\ &\times \int \frac{d\mathbf{y}_2}{\pi^N} e^{-\sum_{\alpha_2=1}^{\mu_2} |\mathbf{y}_2 - \frac{1}{\sqrt{M}} \mathbf{\Lambda}^{1/2} \mathbf{U} \mathbf{G} \mathbf{x}_2^{\alpha_2}|^2 + N(\mu_2 - 1)} \Bigg] \Bigg\}. \end{aligned} \tag{45}$$

Integrating over $\mathbf{y}_1$ and $\mathbf{y}_2$ and ignoring terms of order $(\mu_1 - 1)^2$ or $(\mu_2 - 1)^2$, we get

$$\mathcal{F} = -\ln \Bigg\{ \sum_{\mathbf{X}_1 \mathbf{X}_2} P(\mathbf{X}_1) P(\mathbf{X}_2) \mathbb{E}_{\mathbf{G}}[e^{-\mathrm{tr}(\mathbf{\Lambda U V \Pi V^\dagger U^\dagger})}] \Bigg\}. \tag{46}$$

In the above expression, $\mathbf{\Pi}$ is a $(\mu_1 + \mu_2)$-dimensional, block-diagonal matrix given by

$$\mathbf{\Pi} = \begin{pmatrix} \mathbf{\Pi}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{\Pi}_2 \end{pmatrix}, \tag{47}$$

where $\mathbf{\Pi}_1 = \mathbf{I}_{\mu_1} - \mathbf{u}_1 \mathbf{u}_1^\dagger$ is a $\mu_1 \times \mu_1$ projection matrix orthogonal to the unit vector $\mathbf{u}_1 = [1, \ldots, 1]^T / \sqrt{\mu_1}$ with unit elements, and with $\mathbf{\Pi}_2$ having a similar definition. In addition, $\mathbf{V} = [\mathbf{V}_1, \mathbf{V}_2]$ is a $N \times (\mu_1 + \mu_2)$-dimensional matrix given by the concatenation of the matrices $\mathbf{V}_1$ and $\mathbf{V}_2$ along the second index

$$\begin{aligned} \mathbf{V}_1 &= \frac{1}{\sqrt{M}} \mathbf{G} \mathbf{X}_1, \\ \mathbf{V}_2 &= \frac{1}{\sqrt{M}} \mathbf{G} \mathbf{X}_2. \end{aligned} \tag{48}$$

The rows of the matrices are uncorrelated, i.e., $\mathbb{E}[v_{1i\alpha}^* v_{1j\beta}] = \mathbb{E}[v_{2i\alpha}^* v_{2j\beta}] = \mathbb{E}[v_{1i\alpha}^* v_{2j\beta}] = 0$ for $i \neq j$.

For equal values of i, their covariance matrix for fixed $\mathbf{X}_1$, $\mathbf{X}_2$ is given by

$$\begin{aligned} \mathbf{Q}_1 &= \mathbb{E}[\mathbf{V}_1^\dagger \mathbf{V}_1] = \frac{1}{M} \mathbf{X}_1^\dagger \mathbf{X}_1, \\ \mathbf{Q}_2 &= \mathbb{E}[\mathbf{V}_2^\dagger \mathbf{V}_2] = \frac{1}{M} \mathbf{X}_2^\dagger \mathbf{X}_2, \\ \mathbf{R} &= \mathbb{E}[\mathbf{V}_1^\dagger \mathbf{V}_2] = \frac{1}{M} \mathbf{X}_1^\dagger \mathbf{X}_2. \end{aligned} \tag{49}$$

We will now make the assumption that the matrix $\mathbf{G}$ is Gaussian, deferring the general non-Gaussian case to the next section. As a result, we will take the distribution of $\mathbf{V}$ to be Gaussian, namely,

$$f_G(\mathbf{V}; \mathbf{S}) = \frac{\exp[-\mathrm{Tr}[\mathbf{V} \mathbf{S}^{-1} \mathbf{V}^\dagger]]}{(\pi^{\mu_1 + \mu_2} \det(\mathbf{S}))^N}, \tag{50}$$

where

$$\mathbf{S} = \begin{pmatrix} \mathbf{Q}_1 & \mathbf{R} \\ \mathbf{R}^\dagger & \mathbf{Q}_2 \end{pmatrix}. \tag{51}$$

We note that $\mathbf{Q}_1$ and $\mathbf{Q}_2$ are μ_1- and μ_2-dimensional Hermitian matrices, hence $\mathbf{Q}_{1,\alpha\beta} = \mathbf{Q}_{1,\beta\alpha}^*$, etc. In addition, $\mathbf{R}$ is a $\mu_1 \times \mu_2$-dimensional matrix. As a result, after integrating over $\mathbf{y}_1$, $\mathbf{y}_2$, and $\mathbf{G}$ we obtain

$$\mathcal{F} = -\ln \Bigg\{ \sum_{\mathbf{X}_1 \mathbf{X}_2} P(\mathbf{X}_1) P(\mathbf{X}_2) \prod_{i=1}^N \frac{1}{\det(\mathbf{I} + \lambda_i \mathbf{\Pi} \mathbf{S})} \Bigg\}. \tag{52}$$

We now introduce δ functions which allow us to sum $\mathbf{X}_1$, $\mathbf{X}_2$ independently over different shells of values of the elements of the matrices $\mathbf{Q}_1$, $\mathbf{Q}_2$, $\mathbf{R}$. We also need to introduce a dual matrix $\bar{\mathbf{S}}$, defined in a similar way as $\mathbf{S}$ in (51) through $\bar{\mathbf{Q}}_1$, $\bar{\mathbf{Q}}_2$, $\bar{\mathbf{R}}$ to express the expectations with respect to $\mathbf{X}_1$, $\mathbf{X}_2$ of the δ functions below in terms of their corresponding moments generating functions using Fourier expansions. Let us clarify here that the delta function with complex input can be defined by extending the definition for real input as $\delta(z) = \delta(\mathrm{Re}(z)) \delta(\mathrm{Im}(z))$.

$$\begin{aligned} \mathcal{F} &= -\ln \Bigg\{ \int d\mathbf{S} \prod_{i=1}^N \frac{1}{\det(\mathbf{I} + \lambda_i \mathbf{\Pi} \mathbf{S})} \cdot \sum_{\mathbf{X}_1 \mathbf{X}_2} P(\mathbf{X}_1) P(\mathbf{X}_2) \prod_{\alpha \geqslant \beta} \delta(M S_{\alpha\beta} - [\mathbf{X}^\dagger \mathbf{X}]_{\alpha\beta}) \Bigg\} \\ &= -\ln \Bigg\{ \int d\mathbf{S} d\bar{\mathbf{S}} \sum_{\mathbf{X}_1 \mathbf{X}_2} P(\mathbf{X}_1) P(\mathbf{X}_2) \prod_{i=1}^N \frac{1}{\det(\mathbf{I} + \lambda_i \mathbf{\Pi} \mathbf{S})} \cdot \exp\Bigg[\sum_{\alpha \geqslant \beta} \bar{S}_{\alpha\beta}^* (M S_{\alpha\beta} - [\mathbf{X}^\dagger \mathbf{X}]_{\alpha\beta}) \Bigg] \Bigg\}, \end{aligned} \tag{53}$$

where in the integrations above the differential elements are

$$d\mathbf{S} = M^{(\mu_1+\mu_2)^2}\prod_{\alpha} dS_{\alpha\alpha}\prod_{\beta<\alpha} dS_{\alpha\beta r}dS_{\alpha\beta i},$$

$$d\bar{\mathbf{S}} = \prod_{\alpha}\frac{1}{2\pi i}d\bar{S}_{\alpha\alpha}\prod_{\alpha>\beta}\frac{d\bar{S}_{\alpha\beta r}}{\pi i}\frac{d\bar{S}_{\alpha\beta i}}{\pi i}. \tag{54}$$

The subscripts r, i correspond to the real and imaginary parts of the matrix elements and the lower triangular parts of $\mathbf{S}$ and $\bar{\mathbf{S}}$ are defined by Hermitianity. The integrals of the unbarred variables are over the real line, while the barred variables are integrated over the imaginary line (and hence the i in the denominator).

We can then express $\mathcal{F}$ as follows:

$$e^{-\mathcal{F}} = \int d\mathbf{S}d\bar{\mathbf{S}}e^{-M\Phi}, \tag{55}$$

with

$$\Phi = \frac{1}{M}\sum_{i=1}^{N}\ln\det(\mathbf{I}+\lambda_i\mathbf{\Pi}\mathbf{S}) + \mathrm{tr}[\bar{\mathbf{Q}}_1\mathbf{Q}_1] + \mathrm{tr}[\bar{\mathbf{Q}}_2\mathbf{Q}_2] + \mathrm{tr}[\bar{\mathbf{R}}^\dagger\mathbf{R}+\bar{\mathbf{R}}\mathbf{R}^\dagger]$$
$$-\frac{1}{M}\sum_{j=1}^{M}\ln\left(\sum_{\mathbf{x}_{1j},\mathbf{x}_{2j}} P_j(\mathbf{x}_{1j})P_j(\mathbf{x}_{2j})\exp[[\mathbf{x}_{1j};\mathbf{x}_{2j}]\bar{\mathbf{S}}[\mathbf{x}_{1j}^\dagger;\mathbf{x}_{2j}^\dagger]^T]\right). \tag{56}$$

It should be emphasized that since the channel matrix $\mathbf{H}$ is Gaussian, (55) together with (56), make for an exact representation of $\mathcal{F}$, which is well-defined for integer values of μ_1 and μ_2. In the next section, we will use the framework of the replica method to analyze this representation. Specifically, we will need to devise natural generalizations to the common replica method assumptions to make them applicable in our program [7,14].

B. Saddle-point analysis

To move on, we will need to analyze the asymptotic behavior of the integral in (55) by taking the large M limit. In principle, $\mathcal{F}(\mu_1,\mu_2)$ should first be analytically continued at noninteger μ_1,μ_2, before the large system-size limit can be taken. However, as we shall see, the integrals over $\mathbf{Q}_1$, $\bar{\mathbf{Q}}_1$, etc., cannot be evaluated for general M even for integer μ_1,μ_2. Therefore, we will make the following assumptions.

Assumption 2 (interchanging limits). The limits $M\to\infty$ and $\mu_1,\mu_2\to 1^+$ in evaluating $\mathcal{F}(\mu_1,\mu_2)$ in (52) can be interchanged by first taking the former and then the latter without affecting the final answer.

Having this in mind, we exploit the fact that in the large M limit, $\mathcal{F}$ is well approximated by the saddle-point solution of the above functional. Specifically, from Varadhan's lemma [34], the integral over $\mathbf{S}$ is dominated by the infimum of the integrand in (55) or, equivalently, the infimum of Φ in (56) over $\mathbf{Q}_1$, $\mathbf{Q}_2$, and $\mathbf{R}$. Furthermore, from Cramér's theorem, the supremum of Φ with respect to $\bar{\mathbf{Q}}_1$, $\bar{\mathbf{Q}}_2$, $\bar{\mathbf{R}}$ characterizes the integral over $\bar{\mathbf{S}}$ for large M, thus resulting in a saddle point of Φ. Once we have obtained the relevant saddle point, we will also need to analyze the fluctuations around the solution, so we may evaluate $\mathcal{F}$ to $O(1)$ accuracy in M.

To narrow down the search of saddle points of Φ, we exploit the fact that (56) is symmetric under the exchange of replica indices. Hence, we seek a replica-symmetric saddle point. Therefore, we make the following additional assumption:

Assumption 3 (replica symmetry). At the relevant saddle-point solution, the matrices $\mathbf{Q}_k$, $\bar{\mathbf{Q}}_k$, for $k=1,2$ and $\mathbf{R}$, $\bar{\mathbf{R}}$ are symmetric under interchange of any two matrix indices.

This is a key assumption to this calculation. Usually, the validity of this assumption is tested by performing stability analysis around the saddle point [14]. Fortunately, in our case, it has been shown [15,35] that the replica symmetric saddle-point is stable and, additionally, that the replica symmetric solution is exact for a class of problems including the current one [20–22]. The validity of replica symmetry in this scenario is known to be tightly related with the Bayesian-optimality setting of the problem, i.e., the perfect knowledge of all underlying priors [8]. Specifically, it has been studied in great detail how Bayesian optimality implies the Nishimori condition [36], which in turn places the model firmly in the replica-symmetric regime [37]. Therefore, we seek matrices $\mathbf{Q}_1$, $\mathbf{Q}_2$, which have identical diagonal elements and identical off-diagonal elements, while the matrix $\mathbf{R}$, due to the same symmetry, needs to have all elements equal. Similar requirements are imposed on the matrices $\bar{\mathbf{Q}}_1$, $\bar{\mathbf{Q}}_2$, and $\bar{\mathbf{R}}$. As a result, a compact form of the desired saddle point solution is the following:

$$\begin{aligned}\mathbf{Q}_k &= (\rho-q)\mathbf{I}_{\mu_k} + \mu_k q\mathbf{u}_k\mathbf{u}_k^\dagger,\\ \bar{\mathbf{Q}}_k &= (\bar{\rho}-\bar{q})\mathbf{I}_{\mu_k} + \mu_k\bar{q}\mathbf{u}_k\mathbf{u}_k^\dagger,\\ \mathbf{R} &= p\sqrt{\mu_1\mu_2}\mathbf{u}_1\mathbf{u}_2^\dagger,\\ \bar{\mathbf{R}} &= \bar{p}\sqrt{\mu_1\mu_2}\mathbf{u}_1\mathbf{u}_2^\dagger,\end{aligned} \tag{57}$$

where $k=1,2$. The values of the above parameters are found by demanding that Φ is a saddle point, i.e., by setting the corresponding partial derivatives to zero.

Based on the above, we find that at the saddle point, $\mathbf{\Pi}_1$, $\mathbf{\Pi}_2$ are orthogonal to $\mathbf{R}$, in the sense that $\mathbf{\Pi}_1\mathbf{R}$ and $\mathbf{R}\mathbf{\Pi}_2$ have zero elements. As a result, the first term in (55) does not have any p dependence resulting in vanishing saddle-point values of both p and $\bar{p}$, i.e., $p=\bar{p}=0$.

As a result, the saddle-point equations for ρ, q and $\bar{\rho}, \bar{q}$ are given below, to leading order in $\mu_k - 1$,

$$\bar{q} = \frac{1}{M}\sum_{i=1}^{N}\frac{\lambda_i}{1+(\rho-q)\lambda_i}, \tag{58}$$

$$\bar{r} = -(\mu_k - 1)\bar{q}, \tag{59}$$

$$\rho = \frac{1}{M}\sum_{j=1}^{M}\sum_{x} p_j(x)|x|^2 = \frac{1}{M}\sum_{j=1}^{M}\rho_j, \tag{60}$$

$$q_j = \mathbb{E}_{y,j}\left[\left|\frac{\sum_x x p_j(x)e^{-\bar{q}|y-x|^2}}{\sum_x p_j(x)e^{-\bar{q}|y-x|^2}}\right|^2\right], \tag{61}$$

$$q = \frac{1}{M}\sum_{j=1}^{M} q_j, \tag{62}$$

for $k = 1, 2$. It is worth pointing out that the quantity $\rho_j - q_j$ is the minimum mean square error for the signal of the jth input element. Indeed, taking the difference $\rho_j - q_j$, we see that

$$\varepsilon_j = \rho_j - q_j = \mathbb{E}_{y,j}[|x_{0j} - \hat{x}_j(y;\bar{q})|^2], \tag{63}$$

where x_{0j} is the symbol from the jth input element and, correspondingly, $\rho - q$ is the average minimum mean square error over all inputs,

$$\varepsilon = \rho - q = \frac{1}{M}\sum_{j=1}^{M}\mathbb{E}_{y,j}[|x_{0j} - \hat{x}_j(y;\bar{q})|^2], \tag{64}$$

and this leads to (17) and (18).

Inserting the replica symmetric values of the matrices $\mathbf{Q}_k$, $\bar{\mathbf{Q}}_k$, $\mathbf{R}$, and $\bar{\mathbf{R}}$ into (56) we see that, to leading order in M, $\mathcal{F}(\mu_1, \mu_2) \approx -M\Phi^*$, where Φ^* is evaluated at the saddle point,

$$\Phi^* = (\mu_1 - 1)C + (\mu_2 - 1)C, \tag{65}$$

where C is given in (22), which specifies that it is an infimum with respect to ε and hence the elements of $\mathbf{Q}_1$ and a supremum with respect to $\bar{q}$, i.e., the elements of $\bar{\mathbf{Q}}_1$.

Remark. The fact that this is the only possible solution for the saddle-point values of the matrices $\mathbf{R}^*$ and $\bar{\mathbf{R}}^*$ can be seen in another way. If, for some reason, this wasn't the case, then we would have had a finite saddle-point correction to Φ^*, which would be proportional to $(\mu_1 - 1)(\mu_2 - 1)$. This would then correspond to an $O(M)$ correction to the variance (or the mean) of the mutual information, which would have been clearly visible in the numerics of Sec. V B.

As a result, the ergodic mutual information per input dimension is given by [18]

$$\frac{1}{M}\mathbb{E}_{\mathbf{H}}[I(X,Y|\mathbf{H})] = \left.\frac{\partial\Phi}{\partial\mu_1}\right|_{\mu_1=\mu_2=1} = C. \tag{66}$$

In addition, if we include the h_1 dependence in (42), it is easy to see using (43) that the minimum mean square error is given by

$$\mathbb{E}_{\mathbf{H}}[\mathrm{MMSE}(\mathbf{H})] = M\varepsilon. \tag{67}$$

C. Second-order analysis

In anticipation that the variance and the finite-size correction to the mutual information are going to be of order unity in M, we will need to evaluate the second-order statistics of the functional Φ. To do this, we expand Φ to second order around the saddle point to calculate the fluctuations around it, obtaining the following expression:

$$\begin{aligned}\delta^2\Phi = \sum_{k=1}^{2}\Bigg\{&\mathrm{tr}\delta\bar{\mathbf{Q}}_k^\dagger\delta\mathbf{Q}_k - \frac{1}{2M}\sum_{i=1}^{N}\left(\frac{\lambda_i}{1+\lambda_i\varepsilon}\right)^2\mathrm{tr}[\delta\mathbf{Q}_k\mathbf{\Pi}_1\delta\mathbf{Q}_k\mathbf{\Pi}_1]\\ &-\frac{1}{2M}\sum_{j=1}^{M}\left[\sum_{\mathbf{x}_k}\mathcal{P}_j(\mathbf{x}_k)(\mathbf{x}_k^\dagger\delta\bar{\mathbf{Q}}_k\mathbf{x}_k)^2 - \left(\sum_{\mathbf{x}_k}\mathcal{P}_j(\mathbf{x}_k)\mathbf{x}_k^\dagger\delta\bar{\mathbf{Q}}_k\mathbf{x}_k\right)^2\right]\Bigg\} + \mathrm{tr}\delta\bar{\mathbf{R}}^\dagger\delta\mathbf{R} + \mathrm{tr}\delta\bar{\mathbf{R}}\delta\mathbf{R}^\dagger\\ &-\frac{1}{M}\sum_{i=1}^{N}\left(\frac{\lambda_i}{1+\lambda_i\varepsilon}\right)^2\mathrm{tr}[\delta\mathbf{R}^\dagger\mathbf{\Pi}_1\delta\mathbf{R}\mathbf{\Pi}_2] - \frac{1}{M}\sum_{j=1}^{M}\sum_{\mathbf{x}_1,\mathbf{x}_2}\mathcal{P}_j(\mathbf{x}_1)\mathcal{P}_j(\mathbf{x}_2)\mathbf{x}_1^\dagger\delta\bar{\mathbf{R}}\mathbf{x}_2\mathbf{x}_2^\dagger\delta\bar{\mathbf{R}}^\dagger\mathbf{x}_1.\end{aligned} \tag{68}$$

In the above, $\delta\bar{\mathbf{Q}}_k$, $\delta\mathbf{Q}_k$, $\delta\bar{\mathbf{R}}$, $\delta\mathbf{R}$ are variations of the corresponding matrices about their values on the saddle point and $\mathcal{P}_j(\mathbf{x}_k)$ are the saddle-point probabilities of the corresponding signal vectors $\mathbf{x}_k = [x_1, x_2, \ldots, x_{\mu_k}]$ for the input j, which is given by

$$\mathcal{P}_j(\mathbf{x}_k) = \frac{P(\mathbf{x}_k)e^{\sum_{\alpha,\beta=1}^{\mu_k}x_\alpha^*\bar{\mathbf{Q}}_{\alpha\beta}x_\beta}}{\sum_{\mathbf{x}_k}P(\mathbf{x}_k)e^{\sum_{\alpha,\beta=1}^{\mu_k}x_\alpha^*\bar{\mathbf{Q}}_{\alpha\beta}x_\beta}} \tag{69}$$

for $k = 1, 2$. Now, we note that the variations are only coupled in pairs, i.e., $\delta\mathbf{Q}_k$ with $\delta\bar{\mathbf{Q}}_k$, and $\delta\mathbf{R}$ with $\delta\bar{\mathbf{R}}$, etc. Simple number counting of the degrees of freedom in each sector shows that integration over $\delta\mathbf{Q}_k$, $\delta\bar{\mathbf{Q}}_k$ will provide corrections proportional to $(\mu_1 - 1)^2 + (\mu_2 - 1)^2$, and they provide the variance $\mathbb{E}_{\mathbf{H},\mathbf{y}}[\ln Z(\mathbf{y}|\mathbf{H})^2] - \mathbb{E}_{\mathbf{H},\mathbf{y}}[\ln Z(\mathbf{y}|\mathbf{H})]^2$ and hence do not contribute to the variance of the mutual information. (Nevertheless, they contribute to higher order $O(M^{-1})$ corrections of the variance.) Therefore, we will not further analyze these terms and focus only on the terms including $\delta\mathbf{R}$ and $\delta\bar{\mathbf{R}}$. We first observe that the quadratic term in $\delta\mathbf{R}$ does not involve fluctuations proportional to $\mathbf{u}_1\mathbf{u}_2^\dagger$, due to the projection operators $\mathbf{\Pi}_1$, $\mathbf{\Pi}_2$ appearing in the corresponding term in (68). Thus, it is convenient to express the fluctuations of $\mathbf{R}$, $\bar{\mathbf{R}}$ in a basis where $\mathbf{u}_1$ and $\mathbf{u}_2$ are denoted as the left-side and right-side basis vectors with index 0. Thus, $\delta\mathbf{R}_{00} = \mathbf{u}_1^\dagger \delta\mathbf{R}\mathbf{u}_2$, $\delta\mathbf{R}_{0,\alpha} = \mathbf{u}_1^\dagger \delta\mathbf{R}(\mathbf{I} - \mathbf{u}_2\mathbf{u}_2^\dagger)_\alpha$, etc. Furthermore, we express the fluctuations in terms of their real and imaginary parts as follows:

$$\delta\mathbf{R}_{\alpha\beta} = \delta\mathbf{R}_{\alpha\beta 1} + i\delta\mathbf{R}_{\alpha\beta 2}. \tag{70}$$

We may now express the relevant part of $\delta^2\Phi$ as

$$\begin{aligned}\delta^2\Phi = &\sum_{\alpha_1,\alpha_2,\ell} \begin{bmatrix}\delta\mathbf{R}_{\alpha_1\alpha_2\ell}\\ \delta\bar{\mathbf{R}}_{\alpha_1\alpha_2\ell}\end{bmatrix}^T \mathbf{V}_1 \begin{bmatrix}\delta\mathbf{R}_{\alpha_1\alpha_2\ell}\\ \delta\bar{\mathbf{R}}_{\alpha_1\alpha_2\ell}\end{bmatrix}\\ &+ \sum_{\alpha_1,\ell} \begin{bmatrix}\delta\mathbf{R}_{\alpha_1 0\ell}\\ \delta\bar{\mathbf{R}}_{\alpha_1 0\ell}\end{bmatrix}^T \mathbf{V}_2 \begin{bmatrix}\delta\mathbf{R}_{\alpha_1 0\ell}\\ \delta\bar{\mathbf{R}}_{\alpha_1 0\ell}\end{bmatrix}\\ &+ \sum_{\alpha_2,\ell} \begin{bmatrix}\delta\mathbf{R}_{0\alpha_2\ell}\\ \delta\bar{\mathbf{R}}_{0\alpha_2\ell}\end{bmatrix}^T \mathbf{V}_2 \begin{bmatrix}\delta\mathbf{R}_{0\alpha_2\ell}\\ \delta\bar{\mathbf{R}}_{0\alpha_2\ell}\end{bmatrix}\\ &+ \sum_{\ell} \begin{bmatrix}\delta\mathbf{R}_{00\ell}\\ \delta\bar{\mathbf{R}}_{00\ell}\end{bmatrix}^T \mathbf{V}_3 \begin{bmatrix}\delta\mathbf{R}_{00\ell}\\ \delta\bar{\mathbf{R}}_{00\ell}\end{bmatrix},\end{aligned} \tag{71}$$

where the value ranges of the indices are $\alpha_1 = 1, \ldots, \mu_1 - 1$, $\alpha_2 = 1, \ldots, \mu_2 - 1$ and $\ell = 1, 2$, and the block-Hessian matrices $\mathbf{V}_i$ for $i = 1, 2, 3$ can be expressed as

$$\mathbf{V}_1 = \begin{bmatrix}-m_1 & 1\\ 1 & -m_2\end{bmatrix}, \tag{72}$$

$$\mathbf{V}_2 = \begin{bmatrix}0 & 1\\ 1 & -m_3\end{bmatrix}, \tag{73}$$

$$\mathbf{V}_3 = \begin{bmatrix}0 & 1\\ 1 & -m_4\end{bmatrix}, \tag{74}$$

where the quantities m_i for $i = 1, \ldots, 4$ are

$$m_1 = \frac{1}{M}\sum_{i=1}^{N} \frac{\lambda_i^2}{(1+\lambda_i\varepsilon)^2}, \tag{75}$$

$$m_2 = \frac{1}{M}\sum_{j=1}^{M} \varepsilon_j^2, \tag{76}$$

$$m_3 = \frac{1}{M}\sum_{j=1}^{M} \rho_j\varepsilon_j, \tag{77}$$

$$m_4 = \frac{1}{M}\sum_{j=1}^{M} \rho_j^2. \tag{78}$$

Since the upper left element of $\mathbf{V}_2$ and $\mathbf{V}_3$ is zero, their determinant is equal to -1. We may now integrate over the fluctuations $\delta\mathbf{R}$, $\delta\bar{\mathbf{R}}$ using the volume element appearing in (55). In Appendix A, we prove the following result:

$$\det(\mathbf{V}_1) = (m_1 m_2 - 1) < 0. \tag{79}$$

As a result, the two eigenvalues of $\mathbf{V}_1$ have opposite signs. This behavior is typical for a saddle-point calculation. To perform the integration, one needs to deform the contour of the eigenamplitude of the negative eigenvalue to move along the imaginary axis. The same should happen for the cases of $\mathbf{V}_2$ and $\mathbf{V}_3$. The integration of half the variables over the imaginary axis also takes care of the imaginary numbers appearing in (54).

In summary, to leading order in $\mu_1 - 1$, $\mu_2 - 1$, $\mathcal{F}$ can be expressed as

$$\begin{aligned}\mathcal{F} = &(\mu_1 - 1)MC + (\mu_2 - 1)MC\\ &+ (\mu_1 - 1)(\mu_2 - 1)\ln(1 - m_1 m_2)\\ &+ \mathcal{O}((\mu_1 - 1)^2, (\mu_2 - 1)^2)\end{aligned} \tag{80}$$

from which taking the derivative over μ_1 and μ_2 at $\mu_1 = \mu_2 = 1$ gives us

$$\mathrm{var}_{\mathbf{H}}[I(X, Y|\mathbf{H})] = -\ln(1 - m_1 m_2). \tag{81}$$

This is the final expression of Result 2 for Gaussian channel matrices ($\kappa_4 = 0$). In Appendix B, we will sketch how higher order cumulants vanish in the large-N limit, resulting to the mutual information being asymptotically Gaussian, with mean MC and variance $-\ln(1 - m_1 m_2)$, as above. Nevertheless, that methodology can also be used to obtain higher order corrections to the mean and variance as well.

1. Statistics of MMSE

To get the variance of the mean-square error, we proceed in the same way after including the h dependence in (42) and using the corresponding expression of $\mathcal{F}(\mu_1, \mu_2; h_1, h_2)$. In this case, $\delta^2\Phi$ in (68) is changed only in the fact that $\mathcal{P}_j(\mathbf{x}_k^j)$ in (69) is generalized to

$$\mathcal{P}_j\left(\mathbf{x}_k^j\right)e^{-\frac{h_k}{2}\sum_{\alpha\neq\beta}|x_{k\alpha}-x_{k\beta}|^2} \approx \mathcal{P}_j(\mathbf{x}_k)\left(1 - \frac{h_k}{2}\sum_{\alpha\neq\beta}|x_{k\alpha} - x_{k\beta}|^2\right) \tag{82}$$

for $k = 1, 2$. After going through the same steps as before, we see that only m_2 and m_3 are changed to

$$m_{2h} = \frac{1}{M}\sum_{j=1}^{M}(\rho_j - q_j - h_1\Delta_j)(\rho_j - q_j - h_2\Delta_j), \tag{83}$$

$$m_{3h} = \frac{1}{M}\sum_{j=1}^{M}\rho_j(\rho_j - q_j - h_1\Delta_j), \tag{84}$$

where

$$\Delta_j = -\mathbb{E}_{y,j}[\mathbb{E}_j[|x_0 - x_1|^2(x_0 - x_2)^*(x_1 - x_3)|y; \bar{q}]]. \tag{85}$$

After some algebra, the above result can be expressed as

$$\Delta_j = 2\mathbb{E}_{y,j}[v_1(y; \bar{q})^2 + v_2(y; \bar{q})^2 + 2v_3(y; \bar{q})^2], \tag{86}$$

where the $v_1(y; \bar{q})$, $v_2(y; \bar{q})$ and $v_3(y; \bar{q})$ have been defined in (16a), (16b), and (16c), respectively. As a result, the variance of the normalized mean square error can be obtained from (44) as follows:

$$\begin{aligned}\mathrm{var}[\mathrm{MMSE}(\mathbf{H})] = &\frac{m_1}{1 - m_1 m_2}\frac{1}{M}\sum_{j=1}^{M}\Delta_j^2\\ &+ \left(\frac{m_1}{1 - m_1 m_2}\frac{1}{M}\sum_{j=1}^{M}\Delta_j\varepsilon_j\right)^2.\end{aligned} \tag{87}$$

IV. BEYOND GAUSSIAN MATRIX ELEMENTS

We will now discuss the case when the elements of the channel matrix have non-Gaussian statistics. As mentioned in Sec. II, we will only treat the case where matrix $\mathbf{H}$ has independently distributed elements, hence matrix $\mathbf{C}$ in (4) is diagonal and $\mathbf{U} = \mathbf{I}$. It is important to point out that the rows of $\mathbf{G}$, namely, the vectors $\mathbf{g}_i$, for $i = \dots, N$, enter the calculation independently, and only through the finite-sized $\mu_1 + \mu_2$ dimensional rows $\mathbf{v}_i$ of the matrix $\mathbf{V}$, defined in (48). These rows are now independent given $\mathbf{X}$ and, in the large M limit, their elements become jointly Gaussian, with density given by the marginals of $f_G(\mathbf{V}, \mathbf{S})$ defined in (50). In this section, we will study the deviations from normality in a perturbative fashion. In particular, as first suggested in Ref. [15], we will make use of an Edgeworth expansion about the Gaussian density, which is essentially an asymptotic expansion of the distribution of $\mathbf{v}_i = \frac{1}{\sqrt{M}}\mathbf{g}_i\mathbf{X}$.

Starting from (46) and setting $\mathbf{U} = \mathbf{I}$, the expectation can be decomposed into a product of N expectations over $\mathbf{v}_i$ in the form

$$\prod_{i=1}^{N} \mathbb{E}_{\mathbf{v}_i}[e^{-\lambda_i \mathbf{v}_i^\dagger \mathbf{\Pi} \mathbf{v}_i}]. \tag{88}$$

The following proposition describes how we might evaluate these expectations up to an $o(1/M)$ error.

Proposition (Edgeworth expansion for complex variables). Let $\{g_j\}$ be a sequence of independent proper random variables in $\mathbb{C}$ of zero mean, unit variance, vanishing third moments, and with $\mathbb{E}|g_j|^{4+\epsilon}$ bounded independently of j for some $\epsilon > 0$. Denote $\kappa_{4j} = \mathbb{E}|g_j|^4 - 2$ the fourth cumulants of $\{g_j\}_{j=1}^{M}$. Then we have

$$\left|\left(\mathbb{E}_{\mathbf{v}_i} - \int d\mathbf{v}_i \psi(\mathbf{v}_i)\right) e^{-\lambda_i \mathbf{v}_i^\dagger \mathbf{\Pi} \mathbf{v}_i}\right| = o\left(\frac{1}{M}\right)\sqrt{\lambda_i}\|\mathbf{S}\|^{1/2}, \tag{89}$$

with

$$\psi(\mathbf{v}) = \left(1 + \frac{1}{4M}\sum_{\alpha\beta\gamma\delta=1}^{\mu_1+\mu_2} \kappa_{\alpha\beta\gamma\delta}\partial_\alpha^*\partial_\beta\partial_\gamma^*\partial_\delta\right) f_G(\mathbf{v}; \mathbf{S}), \tag{90}$$

in which ∂_α and ∂_α^* denote Wirtinger derivatives with respect to v_α and v_α^*, respectively, and f_G is as defined in (50).

In addition,

$$\kappa_{\alpha\beta\gamma\delta} = \frac{1}{M}\sum_{j=1}^{M} \kappa_{4j} x_{j\alpha}^* x_{j\beta} x_{j\gamma}^* x_{j\delta}. \tag{91}$$

Proof. If we treat the complex random variables $g_j\mathbf{x}_j$ as real vectors in $\mathbb{R}^{2(\mu_1+\mu_2)}$, then the above proposition is a straightforward application of Theorem 3.6 in Ref. [38]. The compact expression of (90), with partial derivatives with respect to v_α and v_α^*, can be translated directly to corresponding derivatives over the real and imaginary parts of $\mathbf{v}$, after noting that the $\partial_\alpha^*\partial_\beta\partial_\gamma\partial_\delta$ and $\partial_\alpha\partial_\beta\partial_\gamma\partial_\delta$ terms vanish in the integration of $e^{-\mathbf{v}^\dagger \mathbf{\Pi} \mathbf{v}}$. ■

Remark. Note that from Theorem 3.10 of Ref. [39], a uniform bound of the $2 + \epsilon$ moments of the entries of $\mathbf{H}$ is sufficient for the universality of the Marchenko-Pastur Law of the matrix $\frac{1}{M}\mathbf{H}^\dagger\mathbf{H}$. Note also that the eigenvalue statistics are determined from the mutual information of the Gaussian input channel, since in that case the MMSE is the Stieljes transform. Here we see that a uniform bound of the $4 + \epsilon$ moments of the entries of $\mathbf{H}$ is sufficient for the universality of the first-order corrections of the mutual information of the system. Additionally, from Theorem 3.6 of Ref. [38] we can see that a uniform bound of the $5 + \epsilon$ moments of the entries of $\mathbf{H}$ is sufficient to capture in the above result. Specifically, under that condition we can replace $o(1/M)$ with $\mathcal{O}(1/M^{3/2})$. However, since typically the odd-order cumulants of g_j vanish due to symmetry, the next-leading term is of order $O(M^{-2})$, which would result in a correction of $O(M^{-1})$ to the mean and variance of the mutual information, provided that the sixth-order cumulant $\kappa_{6j} < \infty$.

As a result of the above proposition, we obtain

$$\mathcal{F} = -\ln\left\{\sum_{\mathbf{X}_1\mathbf{X}_2} P(\mathbf{X}_1)P(\mathbf{X}_2)\prod_{i=1}^{N}\frac{1}{\det(\mathbf{I} + \lambda_i\mathbf{\Pi}\mathbf{S})}\left[1 + \frac{\lambda_i^2}{2M^2}\sum_{j=1}^{M}\kappa_{4j}(\mathbf{x}_j^\dagger\mathbf{\Pi}(\mathbf{I} + \lambda_i\mathbf{S}\mathbf{\Pi})^{-1}\mathbf{x}_j)^2 + o\left(\frac{1}{M}\right)\sqrt{\lambda}\|\mathbf{S}\|\right]\right\}. \tag{92}$$

Here, we need to express the correction in the second line as a product over the inputs j (for $j = 1, \cdots, M$) of correction terms that have the form of averages over the output i (for $i = 1 \dots, N$). To achieve this exchange in the above expression, we can exponentiate the terms in the second line, and then make use of the following inequalities for the logarithm:

$$\begin{aligned}
&\frac{1}{M}\sum_{j=1}^{M}\ln\left(1 + \frac{\lambda_i^2}{2M}\kappa_{4j}(\mathbf{x}_j^\dagger\mathbf{\Pi}(\mathbf{I} + \lambda_i\mathbf{S}\mathbf{\Pi})^{-1}\mathbf{x}_j)^2 + o\left(\frac{1}{M}\right)\right)\\
&\leqslant \ln\left(1 + \frac{\lambda_i^2}{2M^2}\sum_{j=1}^{M}\kappa_{4j}(\mathbf{x}_j^\dagger\mathbf{\Pi}(\mathbf{I} + \lambda_i\mathbf{S}\mathbf{\Pi})^{-1}\mathbf{x}_j)^2 + o\left(\frac{1}{M}\right)\right)\\
&\leqslant \frac{\lambda_i^2}{2M^2}\sum_{j=1}^{M}\kappa_{4j}(\mathbf{x}_j^\dagger\mathbf{\Pi}(\mathbf{I} + \lambda_i\mathbf{S}\mathbf{\Pi})^{-1}\mathbf{x}_j)^2 + o\left(\frac{1}{M}\right),
\end{aligned} \tag{93}$$

where we hid the dependence of the correction on $\mathbf{x}$ and λ for compactness.

After such a manipulation, going over the same steps as in (52) changes the functional Φ to

$$\Phi = \frac{1}{M}\sum_{i=1}^{N} \ln\det(\mathbf{I}+\lambda_i \mathbf{\Pi}\mathbf{S}) + \mathrm{tr}[\bar{\mathbf{Q}}_1\mathbf{Q}_1] + \mathrm{tr}[\bar{\mathbf{Q}}_2\mathbf{Q}_2] + \mathrm{tr}[\bar{\mathbf{R}}^\dagger\mathbf{R} + \bar{\mathbf{R}}\mathbf{R}^\dagger]$$
$$- \frac{1}{M}\sum_{j=1}^{M}\ln\left[\sum_{\mathbf{x}} p_j(\mathbf{x})e^{\mathbf{x}^\dagger\bar{\mathbf{S}}\mathbf{x}}\cdot\left(1+\frac{\kappa_{4j}}{2M^2}\sum_{i=1}^{N}\lambda_i^2(\mathbf{x}^\dagger\mathbf{\Pi}(\mathbf{I}+\lambda_i\mathbf{S}\mathbf{\Pi})^{-1}\mathbf{x})^2 + o(1)\right)\right]. \tag{94}$$

As a result, evaluated at the replica-symmetric saddle point, the correction to $\mathcal{F}$ due to the non-Gaussianity of the channel takes the form

$$\delta\mathcal{F} = -\frac{1}{2M^2}\sum_{i=1}^{N}\left(\frac{\lambda_i}{1+\lambda_i\varepsilon}\right)^2\cdot\sum_{j=1}^{M}\kappa_{4j}\sum_{\mathbf{x}_1\mathbf{x}_2}\mathcal{P}_j(\mathbf{x}_1)\mathcal{P}_j(\mathbf{x}_2)(\mathbf{x}_1^\dagger\mathbf{\Pi}_1\mathbf{x}_1 + \mathbf{x}_2^\dagger\mathbf{\Pi}_2\mathbf{x}_2)^2 + o(1). \tag{95}$$

In the above expression, the term proportional to $\mathbf{x}_1^\dagger\mathbf{\Pi}_1\mathbf{x}_1\mathbf{x}_2^\dagger\mathbf{\Pi}_2\mathbf{x}_2$ contributes to the variance of the mutual information. Indeed, the analysis is similar to the derivation of the saddle-point equation (18):

$$\sum_{\mathbf{x}_1}\mathcal{P}_j(\mathbf{x}_1)\mathbf{x}_1^\dagger\mathbf{\Pi}_1\mathbf{x}_1$$
$$= \sum_{\mathbf{x}_1}\mathcal{P}_j(\mathbf{x}_1)\sum_{\alpha,\beta}\left(|x_{1\alpha}|^2\delta_{\alpha\beta} - \frac{1}{\mu_1}x_{1\alpha}^* x_{1\beta}\right)$$
$$= (\mu_1-1)(\rho_j - q_j) = (\mu_1-1)\varepsilon_j + \mathcal{O}((\mu_1-1)^2). \tag{96}$$

An identical expression is obtained for the term proportional to $\mathbf{x}_2^\dagger\mathbf{\Pi}_2\mathbf{x}_2$, which is proportional to (μ_2-1). Putting these together, we get

$$\delta\mathrm{var}[I(X,Y|\mathbf{H})] = \frac{m_1}{M}\sum_{j=1}^{M}\kappa_{4j}\varepsilon_j^2. \tag{97}$$

To investigate the variance of the MMSE, a similar analysis culminates in the following evaluation:

$$\frac{1}{2}\sum_{\mathbf{x}}\mathcal{P}_j(\mathbf{x})|x_1-x_2|^2\mathbf{x}^\dagger\mathbf{\Pi}_1\mathbf{x} = (\mu-1)\Delta_j + \mathcal{O}((\mu-1)^2), \tag{98}$$

and the conclusion is

$$\delta\mathrm{var}[\mathrm{MMSE}(\mathbf{H})] = \frac{m_1}{M}\sum_{j=1}^{M}\kappa_{4j}\Delta_j^2. \tag{99}$$

In a similar manner, the term proportional to $(\mathbf{x}_1^\dagger\mathbf{\Pi}_1\mathbf{x}_1)^2$ contributes to the correction of the mean mutual information. In this case, up to $\mathcal{O}((\mu_1-1)^2)$ terms, we have

$$\sum_{\mathbf{x}}\mathcal{P}_j(\mathbf{x})(\mathbf{x}^\dagger\mathbf{\Pi}_1\mathbf{x})^2$$
$$= \sum_{\mathbf{x}}\mathcal{P}_j(\mathbf{x})\sum_{\alpha,\beta,\gamma,\delta}x_\alpha^*\left(\delta_{\alpha\beta}-\frac{1}{\mu_1}\right)x_\beta x_\gamma^*\left(\delta_{\gamma\delta}-\frac{1}{\mu_1}\right)x_\delta$$
$$= (\mu_1-1)\sum_{\mathbf{x}}\mathcal{P}_j(\mathbf{x})x_1^*(x_1-x_2)\sum_{\alpha,\beta}x_\alpha^*\left(\delta_{\alpha\beta}-\frac{1}{\mu_1}\right)x_\beta, \tag{100}$$

from which we get

$$\sum_{\mathbf{x}}\mathcal{P}_j(\mathbf{x})(\mathbf{x}^\dagger\mathbf{\Pi}_1\mathbf{x})^2$$
$$= (\mu_1-1)\sum_{\mathbf{x}}\mathcal{P}_j(\mathbf{x})x_1^*x_2(x_1^*(x_2-x_4)$$
$$+ (x_2-x_3)^*x_1 - x_2^*x_4 - x_3^*x_2 + 2x_3^*x_4),$$

which finally becomes equal to Δ_j.

In conclusion, the correction to the mean mutual information is

$$\delta\mathbb{E}[I(X,Y|\mathbf{H})] = -\frac{m_1}{2M}\sum_{j=1}^{M}\kappa_{4j}\Delta_j. \tag{101}$$

Along this approach, one can also discuss the bias of the MMSE, which results in the calculation

$$\lim_{\mu\to1}\frac{1}{2}\sum_{\mathbf{x}}\mathcal{P}_j(\mathbf{x})|x_1-x_2|^2(\mathbf{x}^\dagger\mathbf{\Pi}_1\mathbf{x})^2$$
$$= (\mu-1)\Gamma_j + \mathcal{O}((\mu-1)^2). \tag{102}$$

The resulting bias is

$$\delta\mathbb{E}[\mathrm{MMSE}(\mathbf{H})] = -\frac{m_1}{2M}\sum_{j=1}^{M}\kappa_{4j}\Gamma_j. \tag{103}$$

V. ANALYSIS OF RESULTS

We will now illustrate the above asymptotic results with specific examples. We will evaluate the *binary* mutual information, normalizing all analytic results of the means and standard deviations in the plots by ln(2). For simplicity, we will only discuss the case of equal variance input elements and treat a specific class of discrete prior distributions common in communications theory. Specifically, we will mainly use the binary distribution with equiprobable values $x = \pm\sqrt{\rho}$ and, in addition, the complex quaternary and hexadecimal distributions, which have equiprobable points on a square lattice of 4 and 16 values on the complex plane, respectively, such that the mean of the distribution is zero and its variance equal to ρ.

Finally, to assess the effect of channel non-Gaussianity, we will use one specific case of i.i.d. sparse channel matrix

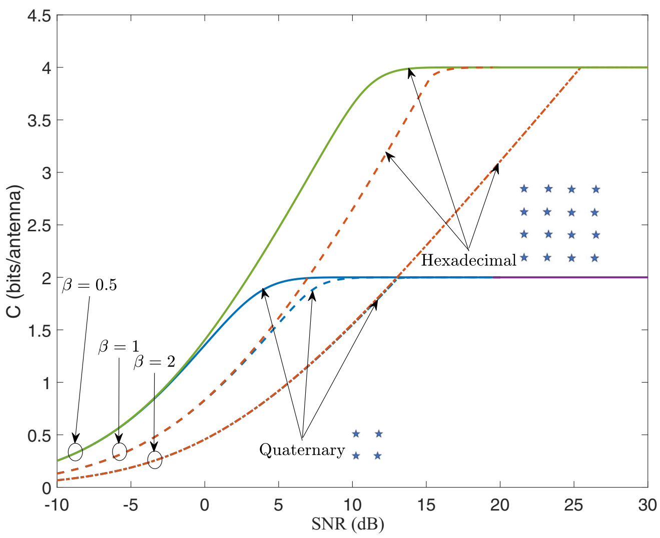

FIG. 1. Analytic expression of the mean mutual information as a function of SNR for various values of $\beta = M/N = 1/2, 1, 2$ and two different prior distributions, namely, quaternary and hexadecimal (the prior lattices are depicted next to their name). The discontinuities of the curves of hexadecimal lattice prior distribution for $\beta = 1, 2$ and quaternary lattice priors for $\beta = 2$ are due to the thermodynamic transition between the two branches of the solution, as discussed in Sec. V.

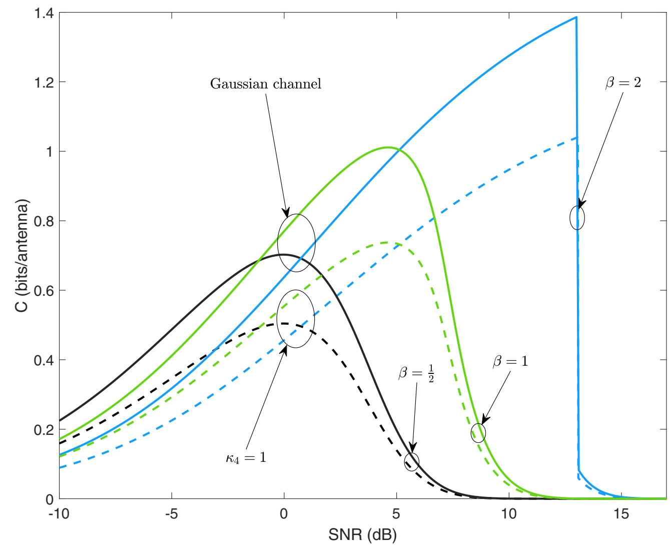

FIG. 2. Standard deviation of the mutual information as a function of SNR for the quaternary prior distribution and $\beta = 0.5, 1, 2$. Two cases of channel statistics have been employed, namely, the case of complex independent Gaussian elements with vanishing curtosis $\kappa_4 = 0$ and the case of independent non-Gaussian elements with $\kappa_4 = 1$. The discontinuities are, as in Fig. 1, due to the thermodynamic transition between the two branches of the solution.

elements with $\kappa_4 = 1$, with distribution $f(G) = p\delta^2(G) + (1-p)f_g(G/\sigma)$, where $p = 1/3$ is the probability of $G = 0$ and $f_g(G/\sigma)$ the circular complex Gaussian density with variance chosen equal to $\sigma^2 = 1/(1-p)$, so that $\mathbb{E}[|G|^2] = 1$.

A. Asymptotic results

Unlike the logarithmic increase with SNR for Gaussian prior distributions [1,2], the average mutual information per input dimension for finite priors peaks at the maximum entropy of the prior distribution, irrespective of β. This can be seen directly in Fig. 1, where the quaternary lattice curves converge to 2 bits/input, while the hexadecimal lattices curves have limiting values at 4 bits/input. It is worth pointing out the discontinuities at high SNR. These are due to a first-order transition between a physical and a metastable nonphysical solution discussed in detail in the past [15] and correspond to the point where the two solutions cross. It should be emphasized here that, due to the stability of the replica-symmetry solution in these systems [15,16], the mutual information of the lower branch of the solution will be always the *true* mutual information in the thermodynamic limit, which, as discussed before, corresponds to the inf sup over $\bar{q}$ and ε, respectively, appearing in (22), as was shown rigorously in Refs. [23,24,8].

Figure 2 depicts the dependence of the standard deviation of the mutual information for various values of β, channel statistics, and prior distributions. The general trend of the dependence is not surprising. Initially, the standard deviation increases with SNR, until eventually it starts decreasing to zero. This decrease is tied to the fact that for increasing SNR the mutual information becomes asymptotically close to its maximum value and therefore its value becomes increasingly independent of the channel realization. Furthermore, it can be seen that for $\kappa_4 = 1$ the standard deviation is larger compared to the Gaussian channel case with $\kappa_4 = 0$. This behavior is directly related with the form of the non-Gaussian corrections to the variance, as seen in (28). Finally, the discontinuities appearing in the average mutual information plots in Fig. 1 are present in the standard deviation as well, and for the same reasons. When the system undergoes a first-order transition, switching from one branch of the coexistence to another, its variance will also transition in a discontinuous way. Of course, this behavior can only be expected in the thermodynamic limit. In addition, Fig. 3 depicts the dependence of the finite order correction to the mutual information due to the non-Gaussian character of the channel. As seen in (27), it depends only on and is proportional to the curtosis of the channel coefficients.

B. Numerical results

We will now provide some comparisons between the analytic results provided in this paper, valid in the asymptotic limit, and numerically generated values of the statistics of the mutual information for finite sized systems. We will solely treat the case of binary input priors, since they are the least computationally demanding. For every fixed value of the channel matrix $\mathbf{G}$, we generate $\mathcal{N} = 10^5$ instantiations of the output vector $\mathbf{y} = \frac{\sqrt{\rho}}{\sqrt{M}}\mathbf{G}\mathbf{x}_0 + \mathbf{z}$, drawing random samples of the vectors $\mathbf{z} \in \mathcal{CN}(0, \mathbf{I}_N)$ and $\mathbf{x}_0 \in (-1, 1)^M$. We then evaluate the empirical value of the mutual information

$$\hat{I}(\mathbf{G}) = -\frac{1}{\mathcal{N}}\sum_{i=1}^{\mathcal{N}} \ln_2\left(\sum_{\mathbf{x}} \frac{1}{2^M}\frac{P(\mathbf{y}_i|\mathbf{G}, \mathbf{x})}{P(\mathbf{y}_i|\mathbf{G}, \mathbf{x}_{0,i})}\right), \tag{104}$$

where the sum inside the log is over all 2^M possible input vectors $\mathbf{x}$. The statistics over the channel variations are then

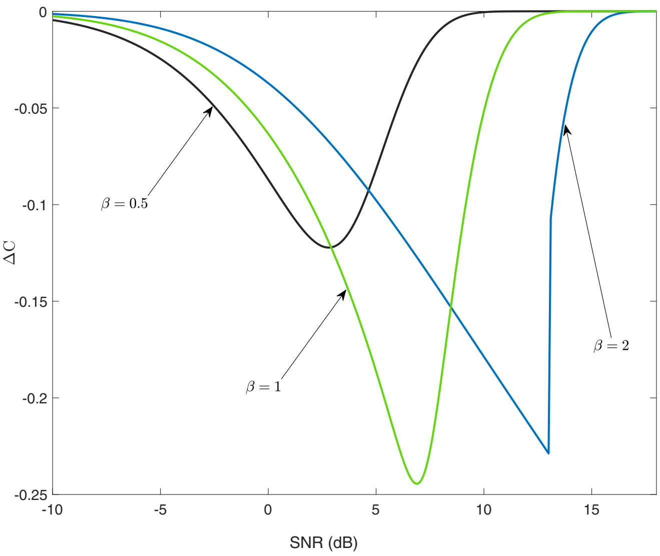

FIG. 3. Finite-size correction to the mutual information for quaternary prior distributions and $\beta = 0.5$, 1, 2, due to the non-Gaussian character of the channel elements. The correction is proportional to κ_4 and here it is depicted for $\kappa_4 = 1$.

obtained by generating 10^3 instances of the channel matrix with the desired distribution.

Starting with the average mutual information, in Fig. 4 we plot the asymptotic average mutual information per input dimension as a function of SNR, together with the corresponding numerically evaluated expectation of the mutual information for Gaussian uncorrelated channels, for two matrix size values M, N with fixed ratio β. We see that the larger matrix size curve is closer to the theoretical one, indicating convergence. In the inset, we plot the finite size correction to the average mutual information due to the non-Gaussian character of the channel (for $\kappa_4 = 1$) and compare the theoretical curve with numerically generated ones. Again, we see that as the system size increases, with fixed β-value, the curve approaches the theoretical one. In Fig. 5, we plot the standard deviation of the mutual information for Gaussian uncorrelated channels, while in the inset we plot the standard deviation of the mutual information for non-Gaussian channels with $\kappa_4 = 1$. Once again, the convergence to the analytic result for larger system sizes is evident, although for non-Gaussian channels the deviation is larger, which is to be expected due to the additional finite-size corrections appearing due to the non-Gaussian character of the channel as seen in Sec. IV. As can be seen, in Figs. 4 and 5, to meaningfully converge to the asymptotic analytic result, especially in the large SNR regime, we need to increase the system size by a factor of 2, e.g., to $N = 20$, $M = 40$, which, however, is beyond our numerical capabilities due to the necessity of summing over all instantiations of the input inside the logarithm in (104), which increases exponentially with size. Other approaches, which are numerically more efficient, such as the AMP algorithm [6], are not of any use here because, while they have been proved to converge to the correct asymptotic limit of the MMSE (and the mutual information per dimension), their finite-size corrections and variance are, in fact, different from the ones discussed here. It should also be mentioned that the difference in the speed of convergence of small and

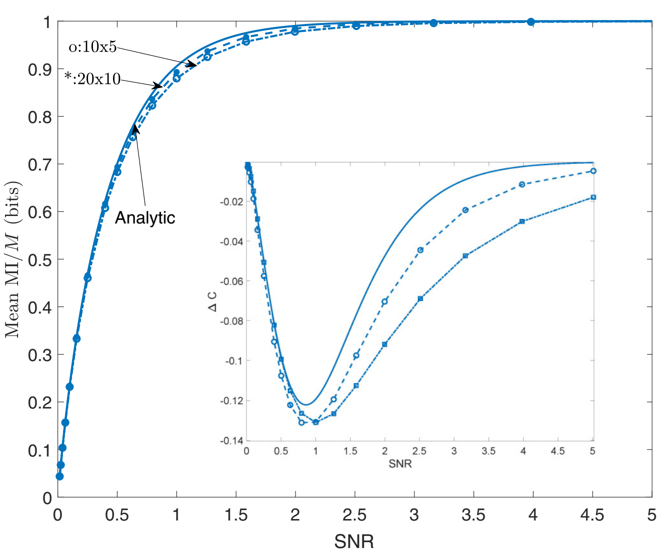

FIG. 4. Plot of the expectation of the mutual information as a function of SNR for independent complex Gaussian channels and binary priors. The analytic curve is compared to two sizes of input and output vectors, namely, $M = 5$, $N = 10$ and $M = 10$, $N = 20$, indicating convergence for increasing system sizes. In the inset, the finite-size correction of the mutual information for independent complex non-Gaussian channel elements. The analytic curve results from (27) for $\kappa_4 = 1$, while the numerical curves are obtained from the difference between the numerically obtained means of the mutual information with Gaussian channel elements and non-Gaussian elements generated from the distribution discussed in the beginning of Sec. V.

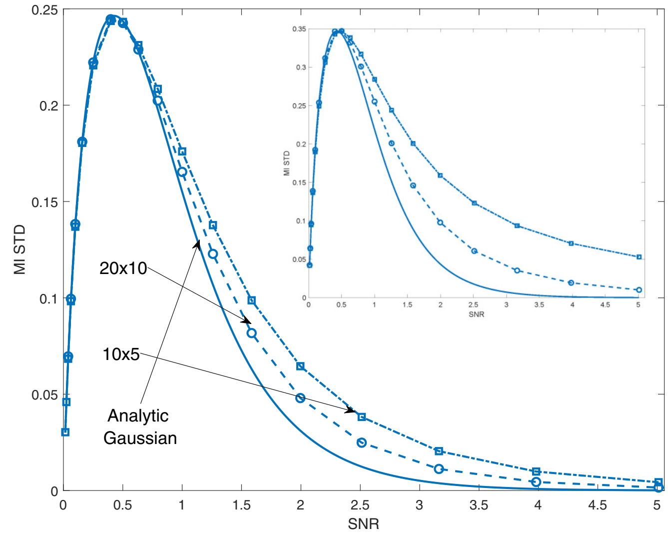

FIG. 5. Standard deviation of the mutual information with binary prior and $\beta = 1/2$ as a function of SNR. The thick solid curves correspond to the asymptotic values using Gaussian channel elements in (28), which are compared to numerically generated curves for two system sizes as in Fig. 4. The inset depicts the standard deviation, both analytical and numerical for the same system sizes as above, for non-Gaussian matrix elements generated from the distribution discussed in the beginning of Sec. V.

larger SNR is not surprising and is well-known in the case of Gaussian inputs, where the mutual information statistics have been studied in more detail. For example, in that case, it has been shown that the large-N and large-SNR limits do not commute, giving different expressions of the variance for the mutual information [40].

VI. CONCLUSION

In conclusion, we have obtained closed-form expressions for the variance and finite-size corrections of the mutual information with non-Gaussian input signals for the case of Gaussian correlated vector channels, as well as the case of uncorrelated non-Gaussian channels. Also, we have shown how in the large M-limit all higher order cumulants of the mutual information vanish, thus making its distribution converge to a Gaussian. Our approach is asymptotic in nature, does not take into account any finite size corrections to the mean and the variance, and hence does not give any information about the speed of convergence. While the approach discussed in Appendix B and in Ref. [25] can be generalized to obtain finite-size corrections to the mean and variance of the mutual information, and thus provide information about the convergence speed, this analysis is beyond the scope of the current paper. Furthermore, we have calculated the variance of the minimum mean-squared error for the case of Gaussian channels, which can be straightforwardly generalized to non-Gaussian channels. These quantities are important when considering the nonasymptotic regime with finite-sized systems and one is interested in the error probability of vector communications channels. Our analysis is formally valid in the asymptotic regime and we have relied on the replica approach. Specifically, we took advantage of the fact that the replica-symmetric solution for the mutual information per input dimension has been proven to be exact in the thermodynamic limit. The methodology we introduce here can be readily applied for the case of generalized linear models used in machine learning [22], where the replica-symmetric solution has been shown to be exact.

ACKNOWLEDGMENTS

This work was partially supported by Project MIS No. 5154714 of the National Recovery and Resilience Plan Greece 2.0 funded by the European Union under the NextGenerationEU Program.

APPENDIX A: PROOF OF EQ. (79)

To prove the above equation, we start from (75):

$$m_1 = \frac{1}{M}\sum_{i=1}^{N}\frac{\lambda_i^2}{(1+\lambda_i\varepsilon)^2} = \frac{1}{M\varepsilon}\sum_{i=1}^{N}\frac{\lambda_i^2\varepsilon}{(1+\lambda_i\varepsilon)^2} \leqslant \frac{1}{M\varepsilon}\sum_{i=1}^{N}\frac{\lambda_i}{1+\lambda_i\varepsilon} = \frac{\bar{q}}{\varepsilon}. \tag{A1}$$

Furthermore, from (63) we see that ε_j is the minimum mean square error. Hence

$$\varepsilon_j = \mathbb{E}_{y,j}[|x_0 - \hat{x}_j(y,\bar{q})|^2] \leqslant \mathbb{E}_{y,j}[|x_0 - y|^2] = \frac{1}{\bar{q}}, \tag{A2}$$

since $\hat{x}_j(y,\bar{q})$, as a function of y minimizes the error. Hence,

$$m_2 = \frac{1}{M}\sum_{j=1}^{M}\varepsilon_j^2 \leqslant \frac{1}{\bar{q}M}\sum_{i=1}^{M}\varepsilon_j = \frac{\varepsilon}{\bar{q}}, \tag{A3}$$

which, combined with (A1), proves (79).

APPENDIX B: HIGHER ORDER CUMULANTS

In this Appendix, we will sketch how to prove that higher order moments vanish in the large M limit. Starting with the skewness of MI, we need to analyze the quantity

$$\mathcal{F}(\mu_1,\mu_2,\mu_3) = -\ln\left\{\iiint\int d\mathbf{y}_1 d\mathbf{y}_2 d\mathbf{y}_3 \mathbb{E}_{\mathbf{H}}\left[\frac{Z(\mathbf{y}_1|\mathbf{H})^{\mu_1}Z(\mathbf{y}_2|\mathbf{H})^{\mu_2}Z(\mathbf{y}_3|\mathbf{H})^{\mu_3}}{(\pi e)^{N(3-\mu_1-\mu_2-\mu_3)}}\right]\right\}, \tag{B1}$$

and then evaluate the derivative $\partial^3_{\mu_1,\mu_2,\mu_3}\mathcal{F}$ at $\mu_1=\mu_2=\mu_3=1$. The analysis in this case follows through in a similar way as in the previous sections. For higher order cumulants, we can generalize to similarly defined quantities $\mathcal{F}(\mu_1,\dots,\mu_k)$ for $k>3$, which can then be differentiated with respect to $\mu_1,\dots,\mu_k$. The evaluation of (B1) for integer values of μ_1,μ_2,μ_3 follows along the same lines as in Sec. III. In this context, the matrix $\mathbf{S}$ in (51) takes the form

$$\mathbf{S} = \begin{pmatrix} \mathbf{Q}_1 & \mathbf{R}_{12} & \mathbf{R}_{13} \\ \mathbf{R}_{21} & \mathbf{Q}_2 & \mathbf{R}_{23} \\ \mathbf{R}_{31} & \mathbf{R}_{32} & \mathbf{Q}_3 \end{pmatrix}, \tag{B2}$$

with $\mathbf{R}_{k\ell} = \frac{1}{M}\mathbf{X}_k^\dagger\mathbf{X}_\ell = \mathbf{R}_{\ell k}^\dagger$, for $k,\ell=1,2,3$, the $\mu_k\times\mu_\ell$ covariance matrices defined as in (49). Similarly, one can define the matrix $\bar{\mathbf{S}}$. The fixed point values of the matrices $\mathbf{R}_{k\ell}$, $\bar{\mathbf{R}}_{k\ell}$ at the replica symmetric point can be shown to vanish. Expanding the corresponding functional Φ around the saddle point, we have

$$e^{-\mathcal{F}} \approx e^{-M\Phi^*}\int d\mathbf{S}d\bar{\mathbf{S}}e^{-M(\delta^2\Phi+\delta^3\Phi+\dots)}, \tag{B3}$$

where $\delta^2\Phi$ are the corresponding second-order fluctuations as in (68) and (71), while $\delta^3\Phi$ are higher order terms in the expansion over $\delta\mathbf{R}_{k\ell}$, $\delta\mathbf{Q}_k$, etc. As in (68), the quadratic terms in $\delta\mathbf{R}_{k\ell}$, $\delta\bar{\mathbf{R}}_{k\ell}$ are decoupled from the corresponding quadratic terms in $\delta\mathbf{Q}_k$, $\delta\bar{\mathbf{Q}}_k$ and, hence, the latter will be disregarded, as they will play a role in this context as well, at least in leading order. After integrating the quadratic terms, the cubic and higher order-terms around the saddle point can be treated in a perturbative fashion, resulting in

$$e^{-\mathcal{F}} \approx e^{-M\Phi^* - \mathcal{F}_2} \left\langle\!\left\langle 1 - M\delta^3\Phi + \frac{(M\delta^3\Phi)^2}{2} \dots \right\rangle\!\right\rangle, \tag{B4}$$

where $\mathcal{F}_2$ is the result over the Gaussian integration of the fluctuating fields $\mathbf{R}_{k\ell}$ etc. and the brackets $\langle\langle\cdot\rangle\rangle$ corresponds to the expectation with respect to the Gaussian fluctuations, which have the following correlations:

$$\begin{aligned}
\left\langle\!\left\langle \delta\mathbf{R}_{k\ell}^{a_k b_\ell} \delta\mathbf{R}_{k'\ell'}^{*a_{k'} b_{\ell'}} \right\rangle\!\right\rangle &= -\frac{m_1}{M \det \mathbf{V}_1} \delta_{k,k'} \delta_{\ell,\ell'} \delta_{a_k,a_{k'}} \delta_{b_\ell,b_{\ell'}}, \\
\left\langle\!\left\langle \delta\bar{\mathbf{R}}_{k\ell}^{a_k b_\ell} \delta\bar{\mathbf{R}}_{k'\ell'}^{*a_{k'} b_{\ell'}} \right\rangle\!\right\rangle &= -\frac{m_2}{M \det \mathbf{V}_1} \delta_{k,k'} \delta_{\ell,\ell'} \delta_{a_k,a_{k'}} \delta_{b_\ell,b_{\ell'}}, \\
\left\langle\!\left\langle \delta\bar{\mathbf{R}}_{k\ell}^{a_k b_\ell} \delta\mathbf{R}_{k'\ell'}^{*a_{k'} b_{\ell'}} \right\rangle\!\right\rangle &= -\frac{1}{M \det \mathbf{V}_1} \delta_{k,k'} \delta_{\ell,\ell'} \delta_{a_k,a_{k'}} \delta_{b_\ell,b_{\ell'}},
\end{aligned} \tag{B5}$$

etc., for $a_\ell, b_\ell = 1, \dots, (\mu_\ell - 1)$. (In the case where either a_ℓ, b_ℓ are zero, then the matrix $\mathbf{V}_1$ is exchanged with $\mathbf{V}_2$ or $\mathbf{V}_3$, as seen in (71). The leading relevant terms of $\delta^3\Phi$ that are related to $\mathbf{R}_{k\ell}$, $\bar{\mathbf{R}}_{k\ell}$ can be expressed as

$$\begin{aligned}
M\delta^3\Phi = &\sum_{i=1}^{N} \left(\frac{\lambda_i}{1+\varepsilon\lambda_i} \right)^3 \\
&\times [\mathrm{Tr}\{\delta\mathbf{R}_{12}\Pi_2\delta\mathbf{R}_{23}\Pi_3\delta\mathbf{R}_{13}^\dagger\Pi_1 + \mathrm{H.c.}\}] \\
&- \sum_{j=1}^{M} \sum_{\mathbf{x}_1,\mathbf{x}_2,\mathbf{x}_3} \mathcal{P}_j(\mathbf{x}_1)\mathcal{P}_j(\mathbf{x}_2)\mathcal{P}_j(\mathbf{x}_3) \\
&\times \mathbf{x}_1^\dagger \delta\bar{\mathbf{R}}_{12}\mathbf{x}_2\mathbf{x}_2^\dagger\delta\bar{\mathbf{R}}_{23}\mathbf{x}_3\mathbf{x}_3^\dagger\delta\bar{\mathbf{R}}_{13}^\dagger\mathbf{x}_1.
\end{aligned} \tag{B6}$$

Since clearly $\langle\langle\delta^3\Phi\rangle\rangle = 0$, the first nonvanishing contribution to the skewness of the mutual information results from $\langle\langle(\delta^3\Phi)^2\rangle\rangle$, which is $O(M^{-1})$ and hence vanishes in the large-system size limit. A similar analysis to higher cumulants shows that they vanish in the same limit.

[1] G. J. Foschini and M. J. Gans, On limits of wireless communications in a fading environment when using multiple antennas, Wireless Pers. Commun. **6**, 311 (1998).

[2] I. E. Telatar, Capacity of multi-antenna Gaussian channels, Eur. Trans. Telecommun. Relat. Technol. **10**, 585 (1999).

[3] Y. Kabashima, T. Wadayama, and T. Tanaka, A typical reconstruction limit for compressed sensing based on lp-norm minimization, J. Stat. Mech.: Theory Exp. (2009) L09003.

[4] S. Rangan, A. K. Fletcher, and V. K. Goyal, Asymptotic analysis of MAP estimation via the replica method and applications to compressed sensing, IEEE Trans. Inf. Theory **58**, 1902 (2012).

[5] S. Ganguli and H. Sompolinsky, Statistical mechanics of compressed sensing, Phys. Rev. Lett. **104**, 188701 (2010).

[6] D. L. Donoho, A. Maleki, and A. Montanari, Message-passing algorithms for compressed sensing, Proc. Natl. Acad. Sci. USA **106**, 18914 (2009).

[7] Y. Kabashima, Inference from correlated patterns: A unified theory for perceptron learning and linear vector channels, in *Journal of Physics: Conference Series* (IOP Publishing, Kyoto, Japan, 2008), Vol. 95, p. 012001.

[8] J. Barbier, F. Krzakala, N. Macris, L. Miolane, and L. Zdeborová, Optimal errors and phase transitions in high-dimensional generalized linear models, Proc. Natl. Acad. Sci. USA **116**, 5451 (2019).

[9] F. Gerace, F. Krzakala, B. Loureiro, L. Stephan, and L. Zdeborová, Gaussian universality of perceptrons with random labels, Phys. Rev. E **109**, 034305 (2024).

[10] A. Maillard, G. B. Arous, and G. Biroli, Landscape complexity for the empirical risk of generalized linear models, in *Proceedings of The First Mathematical and Scientific Machine Learning Conference* (PMLR, Princeton University, Princeton, NJ, USA, 2020), pp. 287–327.

[11] S. Verdú and S. Shamai, Spectral efficiency of CDMA with random spreading, IEEE Trans. Inf. Theory **45**, 622 (1999).

[12] P. B. Rapajic and D. Popescu, Information capacity of a random signature multiple-input multiple-output chanel, IEEE Trans. Commun. **48**, 1245 (2000).

[13] A. L. Moustakas, H. U. Baranger, L. Balents, A. M. Sengupta, and S. H. Simon, Communication through a diffusive medium: Coherence and capacity, Science **287**, 287 (2000).

[14] M. Mézard, G. Parisi, and M. A. Virasoro, *Spin Glass Theory and Beyond* (World Scientific, Singapore, 1987).

[15] T. Tanaka, A statistical-mechanics approach to large-system analysis of CDMA multiuser detectors, IEEE Trans. Inf. Theory **48**, 2888 (2002).

[16] D. Guo and S. Verdu, Randomly spread CDMA: Asymptotics via statistical physics, IEEE Trans. Inf. Theory **51**, 1983 (2005).

[17] R. R. Müller, Channel capacity and minimum probability of error in large dual antenna array systems with binary modulation, IEEE Trans. Signal Process. **51**, 2821 (2003).

[18] C.-K. Wen, P. Ting, and J.-T. Chen, Asymptotic analysis of MIMO wireless systems with spatial correlation at the receiver, IEEE Trans. Commun. **54**, 349 (2006).

[19] R. R. Müller, D. Guo, and A. L. Moustakas, Vector precoding for wireless MIMO systems and its replica analysis, IEEE J. Sel. Areas Commun. **26**, 530 (2008).

[20] G. Reeves and H. D. Pfister, The replica-symmetric prediction for compressed sensing with Gaussian matrices is exact, in *2016 IEEE International Symposium on Information Theory (ISIT)* (IEEE, Barcelona, Spain, 2016), pp. 665–669.

[21] J. Barbier, N. Macris, M. Dia, and F. Krzakala, Mutual information and optimality of approximate message-passing in random linear estimation, IEEE Trans. Inf. Theory **66**, 4270 (2020).

[22] J. Barbier and N. Macris, The adaptive interpolation method: A simple scheme to prove replica formulas in bayesian inference, Probab. Theory Relat. Fields **174**, 1133 (2019).

[23] J. Barbier, N. Macris, A. Maillard, and F. Krzakala, The mutual information in random linear estimation beyond i.i.d. matrices, in *2018 IEEE International Symposium on Information Theory (ISIT)* (IEEE, Vail, CO, USA, 2018), pp. 1390–1394.

[24] A. Maillard, B. Loureiro, F. Krzakala, and L. Zdeborová, Phase retrieval in high dimensions: Statistical and computational phase transitions, Adv. Neural Inf. Process. Syst. **33**, 11071 (2020).

[25] A. L. Moustakas, S. H. Simon, and A. M. Sengupta, MIMO capacity through correlated channels in the presence of correlated interferers and noise: A (not so) large *N* analysis, IEEE Trans. Inf. Theory **49**, 2545 (2003).

[26] W. Hachem, O. Khorunzhiy, P. Loubaton, J. Najim, and L. Pastur, A new approach for capacity analysis of large dimensional multi-antenna channels, IEEE Trans. Inf. Theory **54**, 3987 (2008).

[27] W. Hachem, P. Loubaton, and J. Najim, A CLT for information-theoretic statistics of Gram random matrices with a given variance profile, Ann. Appl. Probab. **18**, 2071 (2008).

[28] U. Adomaityte, L. Defilippis, B. Loureiro, and G. Sicuro, High-dimensional robust regression under heavy-tailed data: Asymptotics and universality, arXiv:2309.16476.

[29] M. Aizenman, J. L. Lebowitz, and D. Ruelle, Some rigorous results on the Sherrington-Kirkpatrick spin glass model, Commun. Math. Phys. **112**, 3 (1987).

[30] A. E. Alaoui, F. Krzakala, and M. I. Jordan, Finite size corrections and likelihood ratio fluctuations in the spiked Wigner model, arXiv:1710.02903.

[31] D. Guo, S. S. (Shitz), and S. Verdú, Mutual information and minimum mean-square error in Gaussian channels, IEEE Trans. Inf. Theory **51**, 1261 (2005).

[32] F. Guerra and F. L. Toninelli, The thermodynamic limit in mean field spin glass models, Commun. Math. Phys. **230**, 71 (2002).

[33] M. Talagrand, The Parisi formula, Ann. Math. **163**, 221 (2006).

[34] A. Dembo and O. Zeitouni, *Large Deviations Techniques and Applications* (Springer-Verlag Inc., New York, 1998).

[35] H. Nishimori, *Statistical Physics of Spin Glasses and Information Processing: An Introduction* (Oxford University Press, Oxford, 2001).

[36] L. Zdeborová and F. Krzakala, Statistical physics of inference: Thresholds and algorithms, Adv. Phys. **65**, 453 (2016).

[37] H. Nishimori, Exact results and critical properties of the Ising model with competing interactions, J. Phys. C **13**, 4071 (1980).

[38] F. Götze and C. Hipp, Asymptotic expansions in the central limit theorem under moment conditions, Z. Wahrscheinlichkeitstheorie Verw. Geb. **42**, 67 (1978).

[39] Z. Bai and J. W. Silverstein, *Spectral Analysis of Large Dimensional Random Matrices* (Springer, New York, NY, USA, 2010), Vol. 20.

[40] P. Kazakopoulos, P. Mertikopoulos, A. L. Moustakas, and G. Caire, Living at the edge: A large deviations approach to the outage MIMO capacity, IEEE Trans. Inf. Theory **57**, 1984 (2011).