# The mutual information of a stochastic binary channel: validity of the Replica Symmetry Ansatz

ANTONIO TURIEL, ELKA KORUTCHEVA*and NÉSTOR PARGA
*Departamento de Física Teórica*
*Universidad Autónoma de Madrid*
*Canto Blanco, 28049 Madrid, Spain*

**Abstract**

We calculate the mutual information (MI) of a two-layered neural network with noiseless, continuous inputs and binary, stochastic outputs under several assumptions on the synaptic efficiencies. The interesting regime corresponds to the limit where the number of both input and output units is large but their ratio is kept fixed at a value $\alpha$. We first present a solution for the MI using the replica technique with a replica symmetric (RS) ansatz. Then we find an exact solution for this quantity valid in a neighborhood of $\alpha = 0$. An analysis of this solution shows that the system must have a phase transition at some finite value of $\alpha$. This transition shows a singularity in the third derivative of the MI. As the RS solution turns out to be infinitely differentiable, it could be regarded as a smooth approximation to the MI. This is checked numerically in the validity domain of the exact solution.

P.A.C.S. 05.20; 87.30
Short title: Mutual Information of a Stochastic Network
Submitted to Journal of Physics A

*Permanent address: G.Nadjakov Institute of Solid State Physics, Bulgarian Academy of Sciences, 1784 Sofia, Bulgaria.
Present address: Departamento de Física Fundamental, Universidad Nacional de Educación a Distancia, c/ Senda del Rey s/n, 28080 Madrid, Spain

# 1 Introduction

The aim of this work is to study the properties of a binary communication channel processing data from a Gaussian source, when the output state is stochastic. The architecture is a two-layered feedforward neural network with $N$ analogue input units and $P$ binary output units. The mutual information (MI) is evaluated in the large $N$ limit with $\alpha = \frac{P}{N}$ fixed. Research in this direction was previously done in [1], where the case of a noiseless binary channel was studied, and in [2] which dealt with the case of a Gaussian source corrupted with input noise.

The main motivation of this work is a technical one. In [1] and [2] the MI of binary channels was obtained by means of the replica technique and the replica symmetry ansatz [3]. However there have not been attempts to show the validity of this solution. In this paper we give an analytical solution of the MI of the channel without making use of the replica technique. In order to compare both methods, the replica symmetry ansatz (RSA) solution of a general stochastic binary channel is also evaluated. While the RSA yields an expression of the MI for all values of $\alpha$, the exact analytical solution turns out to be valid only up to some $\alpha = O(1)$. However, our conclusion is that the correct solution is the analytical one and that there is a (possibly large order) phase transition located at the value of $\alpha$ where the analytical solution ceases to be valid. The RSA solution has to be regarded just as a smooth approximation to the MI, interpolating between the correct small and large $\alpha$ regimes.

There are several other motivations for doing this investigation. Once the MI of the channel is known, the problem of extracting as much information as possible from the inputs can be addressed. This optimization problem leads to interesting data analysis. Optimizing the MI, a criterion known as the “infomax” principle [4], is a way of unsupervised learning (see, e.g., [5]). The parameters of the model (that is, of the channel) adapt according to this principle and in this way they learn the statistics of the environment (that is, of the source). One can also optimize the MI by adapting the transfer function itself [6, 7]. Another form of this type of unsupervised learning is the minimum redundancy criterion [8]. Both have been used to predict the receptive fields of the early visual system [9, 10, 11, 12]. The relation between them has been discussed in ref.[7]. Another motivation is that learning how to solve this particular non-linear channel could provide the techniques to deal with other type of non-linearities. Little is known on the properties of systems other than linear, except for threshold-linear networks [13, 14, 15] (treated with the replica technique), approximations for weak non-linear terms in the processing [16], some general properties of the low and large noise limits [7, 17] and an analytical treatment of binary commutication channels either noiseless [1] or with an input noise [2].

The paper is organized in the following way: The model is explained in the next Section, where the notation and the relevant quantities are also given. In Section 3, the exact calculation of one of the contributions to the MI (the “equivocation” term, see [18]) is presented. In Section 4 the evaluation of the other contribution to MI (the entropy term) is discussed. In its first subsection, this is done by the usual replica technique. The exact solution is obtained in the second subsection. In Section 5 the RS expression of the MI is analysed in several asymptotic regimes. A numerical analysis of the RSA solution is also presented at the end of this section. The comparison between the exact and the RSA solutions is done in Section 6. A discussion about the existence of a phase transition is given in Section 7. Section 8 is devoted to the analysis of the Replica Symmetry Breaking (RSB) solutions. The conclusions are contained in Section 9. Finally several technical questions are presented in the Appendices.

# 2 The model

We consider a two-layered neural network with N inputs $\vec{\xi} \in \Re^N$ and P binary outputs $\vec{v} \in (\mathcal{Z}_2)^P$. The input vector $\vec{\xi}$ is distributed as a Gaussian with zero mean and covariance matrix $\mathbf{C} \in M_{N\times N}(\Re)$:

$$\rho(\vec{\xi}\,) = \frac{e^{-\frac{1}{2}\vec{\xi}\,(\mathbf{C})^{-1}\vec{\xi}^{\,t}}}{\sqrt{\det{(2\pi\mathbf{C})}}}. \tag{1}$$

The feedforward connections are denoted by the matrix $\mathbf{J} \in M_{P\times N}(\Re)$ and its matrix elements by $\{J_{ij}\}(i = 1, ..., P; j = 1, ..., N)$. Instead of considering a fixed matrix we prefer to compute the average MI over an ensemble of stochastic binary channels. The $\{J_{ij}\}$ have also a Gaussian distribution, with zero mean value and two-point correlations $\mathbf{\Gamma}$:

$$\ll J_{ij} J_{i'j'} \gg = \delta_{ii'} \Gamma_{jj'}, \tag{2}$$

where the double angular brackets indicate the average over the channel ensemble and $\mathbf{\Gamma} \in M_{N\times N}(\Re)$ . Notice that those connections converging to different outputs are independently distributed. The coupling matrix $\mathbf{J}$ can be also regarded as $P$ random $N$-dimensional vectors $\vec{J}_i (i = 1, ..., P)$ given by the rows of $\mathbf{J}$. From eq.(2), we have that each of these rows is distributed independently as :

$$\rho(\vec{J}_i) = \frac{e^{-\frac{1}{2}\vec{J}_i(\mathbf{\Gamma})^{-1}\vec{J}_i^t}}{\sqrt{\det{(2\pi\mathbf{\Gamma})}}}\ . \tag{3}$$

Let us now define the local field as $\vec{h} = \mathbf{J}\,\vec{\xi}$. The output state is computed by means of the probability distribution $\wp(\vec{v}\,|\vec{h}\,)$ that the output vector is $\vec{v}$ for a given local field $\vec{h}$. Most interesting problems have a factorized output distribution, we will then assume that $\wp(\vec{v}\,|\vec{h}\,) = \prod_{i=1}^{P} \wp(v_i|h_i)$. We will also require the reasonable condition that $\wp(v_i|h_i) = f_\beta(h_i v_i)$ ($\beta$ denotes an output noise parameter), where $f_\beta(x)$ is an arbitrary function satisfying $0 \leq f_\beta(x) < 1$ and $f_\beta(x) + f_\beta(-x) = 1$.

Both the replica and the exact calculations will be done for any of those $f(x)$. However we will study with some detail the case (to be referred to as the Hyperbolic Tangent Transfer function, HTT):

$$f(x) = \frac{e^{\beta x}}{e^{\beta x} + e^{-\beta x}} = \frac{1}{2}(1 + \tanh(\beta x)) \ . \tag{4}$$

The deterministic channel [1] is obtained either when $f(x)$ is chosen as the Heaviside function $\theta$ or in the large $\beta$ limit of the HTT function.

Let us now define the mutual information $I(\vec{v}, \vec{\xi}\,|\mathbf{J})$ [18, 19, 20] between the input and output vectors, given the channel parameters $\mathbf{J}$: [1]

$$I(\vec{v}, \vec{\xi}\,|\mathbf{J}) = \sum_{\vec{v}, \vec{\xi}} \wp(\vec{v}, \vec{\xi}\,|\mathbf{J}) \log \frac{\wp(\vec{v}, \vec{\xi}\,|\mathbf{J})}{\wp(\vec{v}|\mathbf{J})\rho(\vec{\xi}\,)} \ , \tag{5}$$

where $\wp(\vec{v}|\mathbf{J})$ is the output vector distribution given $\mathbf{J}$. The joint probability $\wp(\vec{v}, \vec{\xi}\,|\mathbf{J})$ can be written as

$$\wp(\vec{v}, \vec{\xi}\,|\mathbf{J}) = \rho(\vec{\xi}\,)\wp(\vec{v}|\vec{\xi}\,, \mathbf{J}), \tag{6}$$

where $\wp(\vec{v}|\vec{\xi}\,, \mathbf{J})$ denotes the conditional distribution of the output vector $\vec{v}$ given the input $\vec{\xi}$ and the channel $\mathbf{J}$. Since the relation between the input and the local field $\vec{h}$ is deterministic, it can be substituted by $\wp(\vec{v}|\vec{h})$.

First we need to define the output entropy for fixed couplings $\mathbf{J}$

$$H(\vec{v}|\mathbf{J}) = -\sum_{\vec{v}} \wp(\vec{v}|\mathbf{J}) \log \wp(\vec{v}|\mathbf{J}), \tag{7}$$

and the entropy of the output conditioned by the input $\vec{\xi}$, again for fixed couplings (the equivocation term):

$$H(\vec{v}|\vec{\xi}\,, \mathbf{J}) = -\int d^N\vec{\xi} \quad \rho(\vec{\xi}\,) \sum_{\vec{v}} \wp(\vec{v}|\vec{\xi}\,, \mathbf{J}) \log \wp(\vec{v}|\vec{\xi}\,, \mathbf{J}) \ . \tag{8}$$

Then the MI can be expressed as

$$I(\vec{v}, \vec{\xi}\,|\mathbf{J}) = H(\vec{v}|\mathbf{J}) - H(\vec{v}|\vec{\xi}\,, \mathbf{J}). \tag{9}$$

We are interested in the mutual information $I = \ll I(\vec{v}, \vec{\xi}\,|\mathbf{J}) \gg$ averaged over the channel ensemble. Then, calling $I_1 = \ll H(\vec{v}|\mathbf{J}) \gg$ and $I_2 = \ll H(\vec{v}|\vec{\xi}\,, \mathbf{J}) \gg$, we have $I = I_1 - I_2$, or in terms of MI per input unit, $i = i_1 - i_2$, where $i_1 = \frac{I_1}{N}$ and $i_2 = \frac{I_2}{N}$.

Each term will be studied separately. In the next section we compute the equivocation term $I_2$. This can be obtained exactly by means of simple arguments. The output entropy term requires more care. It will be evaluated in section 4.1 using the replica technique and in section 4.2 using exact analytical methods.

[1] In eq. 5 and hereafter $\log(x) \equiv \frac{\ln(x)}{\ln 2}$. In the derivations, however, we make use of $\ln(x)$ because of the Taylor expansions.

## 3 The equivocation term $I_2$

Since $\wp(\vec{v}|\vec{\xi},\mathbf{J})$ factorizes, it is convenient to define the single output entropy, $\mathcal{H}(h_i)$, which is a *function* of $h_i$ :

$$\mathcal{H}(h_i) = -\sum_{v_i=\pm 1} \wp(v_i|h_i) \ln \wp(v_i|h_i) \ , \tag{10}$$

where one should keep in mind that $\vec{h} = \mathbf{J}\vec{\xi}$. Then :

$$H(\vec{v}|\vec{\xi},\mathbf{J}) = \sum_{i=1}^{P} \int d^N\vec{\xi} \ \ \rho(\vec{\xi})\mathcal{H}(h_i) \tag{11}$$

and

$$I_2 = \sum_{i=1}^{P} \int d\mathbf{J} \ \ \rho(\mathbf{J}) \int d^N\vec{\xi} \ \ \rho(\vec{\xi})\mathcal{H}(h_i). \tag{12}$$

$I_2$ can be easily evaluated in several simple examples. In the deterministic case it is zero. In the large noise limit ($\beta \to 0$) it reaches its upper bound $I_2 = P\ln 2$. For the HTT function, eq. (4), we have $\mathcal{H}(h) = \ln(e^{\beta h} + e^{-\beta h}) - \beta h \ \tanh \beta h$, and this single output entropy can be substituted in eq. (12) to obtain $I_2$. In Appendix A the details of such calculation are presented; here we recall the final formula eq. (58), valid for a matrix $\mathbf{M} = \mathbf{\Gamma C}$ having all its eigenvalues of the same order and any function $\mathcal{H}(h)$. The equivocation term *per* input unit is then:

$$i_2 = \alpha \int_{-\infty}^{\infty} dz \ \frac{e^{-z^2}}{\sqrt{\pi}} \ \mathcal{H}(\sqrt{2\bar{M}}\, z) \ . \tag{13}$$

with $\bar{M} = \mathrm{Tr}(\mathbf{M})$ (the trace of $\mathbf{M}$). For the sigmoidal HTT function, we obtain:

$$i_2 = \alpha \sum_{m=1}^{\infty} (-1)^{m+1} A_m, \tag{14}$$

where

$$A_m = \frac{2\beta_0}{\sqrt{\pi}} + (\frac{1}{m} - 2m\beta_0^2)(1 - \mathit{erf}(m\beta_0))e^{m^2\beta_0^2} \tag{15}$$

and $\beta_0 = \sqrt{2\bar{M}}\,\beta$, which we shall call the *reduced noise parameter*. Alternatively, we will also use the *reduced temperature* $T_0 = {\beta_0}^{-1}$. The symbol $\mathit{erf}(x)$ stands for the error function.

One can easily obtain several limits. For small $T_0$ :

$$i_2 \approx \alpha \frac{\pi^{3/2}}{6} T_0 \ , \tag{16}$$

where one observes that $i_2$ grows linearly with $T_0$ . For small $\beta_0$,

$$i_2 \approx \alpha\ \ln 2 - \frac{\alpha}{4}\,\beta_0^2\ , \tag{17}$$

in this case $i_2$ departs from $\alpha \ln 2$ quadratically with $\beta_0$ .

For other transfer functions one obtains the same qualitative result, although with different coefficients. This is simply because these coefficients depend on the derivatives of the transfer function in the neighborhood of $\beta = 0$ and $\beta = \infty$, respectively.

## 4 Calculation of $I_1$

We now compute the output entropy term defined in eq.(7). First we notice that the (discrete) probability density of the outputs $\vec{v}$ is given by:

$$\wp_{\vec{v}} \equiv \wp(\vec{v}|\mathbf{J}) = \int d^N\vec{\xi}\ \ \rho(\vec{\xi}\,)\wp(\vec{v}|\vec{\xi}\,,\mathbf{J})\ . \tag{18}$$

Since $\wp_{\vec{v}}$ is a probability density, $\sum_{\vec{v}} \wp_{\vec{v}} = 1$ and so

$$I_1 = -\lim_{n\to 0} \frac{\sum_{\vec{v}} \ll \wp_{\vec{v}}^{n+1} \gg\ -1}{n}\ . \tag{19}$$

$\ll \wp_{\vec{v}}^{n+1} \gg$ can be written in terms of replicated variables $\{\vec{\xi}^{\,a}\}$, $a = 0, 1, ..., n$ in the following way :

$$\ll \wp_{\vec{v}}^{n+1} \gg = \int d\mathbf{J}\ \ \rho(\mathbf{J}) \int \prod_{a=0}^{n} \left[ d^N\vec{\xi}^{\,a} \rho_{\vec{\xi}}(\vec{\xi}^{\,a}) \prod_{i=1}^{P} f(v_i h_i^a) \right]\ . \tag{20}$$

Here the local fields are:

$$h_i^a = \sum_{j=1}^{N} J_{ij}\ \xi_j^a\ . \tag{21}$$

We will only compute the integer order moments of $\wp_{\vec{v}}$. The continuous order moments will be obtained by naive extrapolation of them. Actually they can be obtained in a completely rigorous way (although in a rather complicated fashion) because the integer moments contain enough information to reconstruct the probability distribution of the variable $\wp_{\vec{v}}$.

It is obvious from the distribution of $\mathbf{J}$ that each element $J_{ij}$ is independent of the others with different output index $i$. Using this and the fact that each $\vec{J}_i$ has an even distribution (Gaussian) we have that $\ll \wp_{\vec{v}}^{n+1} \gg$ is independent of $\vec{v}$. Thus we can write :

$$I_1 = -\frac{2^P \ll \wp^{n+1} \gg\ -1}{n},\ n \to 0\ , \tag{22}$$

where $\wp$ stands for the conditional distribution $\wp_{\vec{v}}$ with a specific choice of $\vec{v}$ , e.g. $v_i = +1, i = 1, ..., P$.

In eq. (20) we can apply, for almost every $\mathbf{J}$, the Bessel-Plancherel identity in each of the integrals over the $\vec{\xi}^{a}$'s:

$$\int d^N\vec{\xi}^{a} \quad \frac{e^{-\frac{1}{2}\vec{\xi}^{at}(\mathbf{C})^{-1}\vec{\xi}^{a}}}{\sqrt{\det(2\pi\mathbf{C})}}\prod_{i=1}^{P} f(h_i^a) = \int d^P\vec{u}^a \quad e^{-2\pi^2\vec{u}^{at}\mathbf{\Delta}\vec{u}^a}\prod_{i=1}^{P}\hat{f}(u_i^a)\ . \tag{23}$$

The function $\hat{f}$ is the Fourier transform of $f$ and must be understood in the distributional sense, and $\mathbf{\Delta} = \mathbf{J}\,\mathbf{C}\,\mathbf{J}^t$. This expression holds for any value of $P$ and $N$ (see Appendix B for details). Then, the characteristic function $\hat{\rho}(\{\vec{u}^a\}_{a=0,\ldots,n})$ associated to the joint probability distribution of the replicated local fields, $\rho(\{\vec{h}^a\}_{a=0,\ldots,n})$ is:

$$\hat{\rho}(\{\vec{u}^a\}_{all\ a's}) = e^{-\frac{1}{2}\sum_{j=1}^{N}\ln\det(\mathbf{Id}_{P\times P}+4\pi^2 m_j\mathbf{U})}, \tag{24}$$

where the matrix $\mathbf{U}$ is the sum of all the projectors associated to each replica vector,

$$U_{ii'} = \sum_{a=0}^{n} u_i^a u_{i'}^a. \tag{25}$$

From eqs.(20) and (23) we obtain the following result:

$$\ll \wp^{n+1} \gg = \int \prod_{a=0}^{n} d^P\vec{u}^a\hat{\rho}(\{\vec{u}^a\})\prod_{a=0}^{n}\prod_{i=1}^{P}\hat{f}(u_i^a)\ . \tag{26}$$

This formula is exact for the moments of integer order of $\wp$. This will be the starting point of subsection 4.2, where we will calculate the moments of $\wp$ exactly. Before that we present the RSA solution in the next subsection.

## 4.1 The RSA approach

After a rather lengthy algebra we obtain the entropy term $I_1$ . Some details of the calculation are described in Appendix C; here we only give the final result. $I_1$ is a function of two order parameters, here called $x$ and $s$. Its value per input unit is given by:

$$i_1^{RSA} = \alpha R(x) + \frac{1}{2}\tau[\,\ln(\mathbf{Id}_{N\times N}+s\mathcal{G})\,] - \frac{s}{2(x+1)}\ , \tag{27}$$

where $\mathcal{G}$ is the normalized matrix, $\mathcal{G} = \frac{N\mathbf{M}}{M}$. We have made use of the symbol $\tau$ for a normalized trace operator, $\tau(\cdot) = \frac{1}{N}Tr(\cdot)$. The order parameters satisfy the self-consistent SP equations:

$$\begin{cases} x &= s\tau(\frac{\mathcal{G}^2}{\mathbf{Id}_{N\times N+s\mathcal{G}}})/\tau(\frac{\mathcal{G}}{\mathbf{Id}_{N\times N+s\mathcal{G}}}) \\ s &= -2\alpha(x+1)^2\frac{dR}{dx} \end{cases}, \tag{28}$$

The function $R(x)$ is the average entropy of an effective transfer function $g_x$ defined as:

$$g_x(y) = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} dw \; e^{-(w+y)^2} f(\sqrt{\frac{2\bar{M}}{1+x}} w) \; . \tag{29}$$

More precisely,

$$R(x) = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} dy \; e^{-y^2} S(g_x(\sqrt{x}\, y)) \; , \tag{30}$$

where $S(z)$ is the entropy of a binary probability, i.e. ,

$$S(z) = -z \ln z - (1-z) \ln (1-z) \; . \tag{31}$$

It is worth noting that for the deterministic case $g_x(y) = \frac{1}{2}(1 + erf(y))$, and for the fully random case $g_x(y) = \frac{1}{2}$. In the deterministic case, there is a simple relation between our parameters and those ($q$ and $\hat{q}$)used in [1]: $\hat{q} = s$ and $q = \frac{x}{1+x}$. We prefer to use $x$ instead of $q$ because it usually yields simpler expressions.

## 4.2 The Exact Solution

In this subsection we present an exact evaluation of $I_1$ valid for $\alpha \leq \alpha_c$, where $\alpha_c$ is of order one. It is necessary to assume that all the eigenvalues of $\mathbf{M} = \mathbf{\Gamma C}$ are of the same order. The details of the calculation are presented in Appendix D; we only give here the final result for the moments:

$$\begin{aligned} \ll \wp^{n+1} \gg = 2^{-P(n+1)} \; & e^{-\frac{1}{2} Tr[\, \ln (\mathbf{Id}_{N\times N} - \frac{2}{\pi} k^2\, n\, \alpha\, \mathcal{G})\, ]} \\ & e^{-\frac{n}{2} Tr[\, \ln (\mathbf{Id}_{N\times N} + \frac{2}{\pi} k^2\, \alpha\, \mathcal{G})\, ]} \; , \end{aligned} \tag{32}$$

where $k$ is defined as:

$$k = \int_{-\infty}^{\infty} dy \;\; y\, e^{-\frac{y^2}{2}} f(\sqrt{\bar{M}}\; y) \; . \tag{33}$$

Let us now extrapolate eq. (32) to non-integer $n$. This gives our analytical estimate of $i_1$ :

$$\begin{aligned} i_1^{an} &= -\frac{1}{N} \lim_{n\to 0} \frac{2^{-P(n+1)} \ll \wp^{n+1} \gg \; -1}{n} \\ &= \alpha \ln 2 - \frac{k^2}{\pi} \alpha + \frac{1}{2} \tau [\, \ln (\mathbf{Id}_{N\times N} + \frac{2k^2}{\pi} \alpha\, \mathcal{G})\, ] \; . \end{aligned} \tag{34}$$

The numerical constant $k$ is real-valued, and it is expected to be of order one. In fact, the maximum of $k$ is reached in the deterministic case ( $f(x) = \theta(x)$ ). This gives $k = 1$, independent of $\bar{M}$. Notice that the minimum of $k$ is reached by $f(x) = \theta(-x)$, giving $k = -1$. The minimum of $k^2$ is realized by $f(x) = \frac{1}{2}$, what gives $k = 0$.

The computation done in this section (eq. (32)) is equivalent to a Taylor expansion of the original equation for the moments, eq. (26). This can be checked by explicit evaluation of the derivatives of the two expressions.[2]

# 5 Analysis of the RSA solution

Using eq. (27), together with the SP equations (28), we obtain expansions of $i_1^{RSA}$ at small and large $\alpha$ and $\beta_0$ ($\beta_0 = \sqrt{2\bar{M}}\beta$). We will make explicit calculations for the deterministic and the HTT functions (the completely random case always gives $i_1 = \alpha \ln 2$).

## 5.1 Small $\alpha$ limit

Let us first investigate the deterministic case ($\beta_0 \to \infty$) in this regime. From eq. (28), we can see that $s \approx \frac{2}{\pi}\alpha$ and $x \approx s\,\tau(\mathcal{G}^2) \approx \frac{2\tau(\mathcal{G}^2)}{\pi}\alpha$. This gives the first two orders of the expansion of $i_1$ in powers of $\alpha$:

$$i_1^{RSA} \underset{\alpha\ll 1}{\approx} \alpha \ln 2 - \alpha^2 \frac{\tau(\mathcal{G}^2)}{\pi^2} \,, \tag{35}$$

where we see that, as expected, the second order is negative. The next order in $T_0$ gives, after solving the SP equations up to order $T_0$,

$$i_1^{RSA} \underset{\alpha\ll 1, T_0\ll 1}{\approx} \alpha \ln 2 - \alpha^2 \frac{\tau(\mathcal{G}^2)}{\pi^2} + \alpha^2 \frac{\tau(\mathcal{G}^2)}{3} T_0^2 \,. \tag{36}$$

This is a positive contribution. However, this does not mean that the MI increases with $T_0$; in fact, the term $i_2$ gives a larger contribution of order $\alpha T_0$, as can be seen in eq. (16). More precisely,

$$i^{RSA} \;=\; i_1^{RSA} - i_2 \;\approx \alpha \ln 2 - \frac{\pi^{3/2}}{6}\alpha T_0 - \alpha^2 \frac{\tau(\mathcal{G}^2)}{\pi^2} + \frac{\tau(\mathcal{G}^2)}{3}\alpha^2 T_0^2 \,. \tag{37}$$

We now calculate the first order correction in $\beta_0 \ll 1$ to $i_1^{RSA}$. The leading order is the fully stochastic case, and $i_1^{RSA} = i_2 = \alpha \ln 2$ ($x = s = 0$). Up to the next order, $i_1^{RSA}$ is:

$$i_1^{RSA} \underset{\alpha\ll 1, T_0\gg 1}{\approx} \alpha \ln 2 - \alpha^2 \frac{\tau(\mathcal{G}^2)}{16} \beta_0^4. \tag{38}$$

From eqs. (17) and (38), we obtain :

$$i^{RSA} \approx \frac{1}{4}\alpha\beta_0^2 - \frac{\tau(\mathcal{G}^2)}{16}\alpha^2\beta_0^4. \tag{39}$$

[2]To compute the derivatives of eq. (26) with respect to $\alpha$ one has first to make explicit its dependence on the parameter N by expressing $\mathbf{M}$ in terms of $\mathcal{G}$ ($\mathbf{M} = \frac{\bar{M}}{N}\mathcal{G}$). Then, after setting each derivative at $\alpha = 0$, the resulting integrals are easy to compute.

### 5.2 Large $\alpha$ limit

In this regime, and in the low temperature limit, we obtain from eq. (28) that $x \approx s$ and $s \approx A_0 \alpha \sqrt{x}$, where $A_0$ is the constant given in [1], [2] :
$A_0 = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} dz\; e^{-z^2} S(\frac{1+erf(z)}{2})$, $A_0 \approx 0.72$. From here we obtain $s \approx A_0^2 \alpha^2$, $x \approx A_0^2 \alpha^2$. Substituting these parameters in eq. (27) one obtains the known result [1], [2]:

$$i_1^{RSA} \underset{\alpha \gg 1}{\approx} \ln \alpha + \frac{1}{2} + \ln A_0 + \frac{1}{2} \tau[\, \ln \mathcal{G}\, ] \, . \tag{40}$$

Adding weak output noise, and assuming $\alpha\, T_0 \to 0$ we have:

$$i_1^{RSA} \underset{\alpha \gg 1, T_0 \ll 1}{\approx} \ln \alpha + \frac{1}{2} + \ln A_0 + \frac{1}{2} \tau[\, \ln \mathcal{G}\, ] + \frac{\pi^2 A_0^2}{12} \alpha^2 T_0^2 \, . \tag{41}$$

From here and eq. (16):

$$i^{RSA} \approx \ln \alpha + \ln A_0 + \frac{1}{2} + \frac{1}{2} \tau[\, \ln \mathcal{G}\, ] - \frac{\pi^{3/2}}{6} \alpha T_0 + \frac{\pi^2 A_0^2}{12} \alpha^2 T_0^2 \, . \tag{42}$$

In the opposite limit, $\beta_0 \ll 1$ (large temperatures), and also assuming $\alpha \beta_0^2$ small, it is straightforward to see that :

$$i_1^{RSA} \underset{\alpha \gg 1, T_0 \gg 1}{\approx} \alpha \ln 2 - \frac{\beta_0^2}{4} \alpha + \frac{1}{2} \tau[\, \ln (\mathbf{Id}_{N \times N} + \frac{1}{2} \beta_0^2 \alpha \mathcal{G})] \, , \tag{43}$$

which together with eq. (17) gives:

$$i^{RSA} \approx \frac{1}{2} \tau[\, \ln (\mathbf{Id}_{N \times N} + \frac{1}{2} \beta_0^2 \alpha \mathcal{G})\, ] \, . \tag{44}$$

which shows that the MI decays as $\beta_0^2$ when $\beta_0 \to 0$.

### 5.3 Numerical analysis

The plot of $\ln i^{RSA}$ (which is obtained combining eqs. (13) and (27)) versus $\ln \alpha$, using the HTT for several values of the reduced noise parameter $\beta_0 = \sqrt{2\bar{M}}\, \beta$, is shown in Fig. 1. The correlation matrix was taken proportional to the identity: $\mathbf{M} = \frac{\bar{M}}{N} \mathbf{Id}_{N \times N}$ . As expected, for each $\alpha$, the MI decreases as the temperature increases. It is also interesting that an increase in the temperature moves the saturation point (the change from the close-to-linear regime to the asymptotic one, in which the MI increases slower with $\alpha$) to greater values of $\alpha$.

## 6 Comparison between the exact and the RSA solutions

The analytical result presented in eq. (34) seems rather astonishing as it provides a very simple expression for $i_1$, compared with the cumbersome formulae of the RSA solution. Then

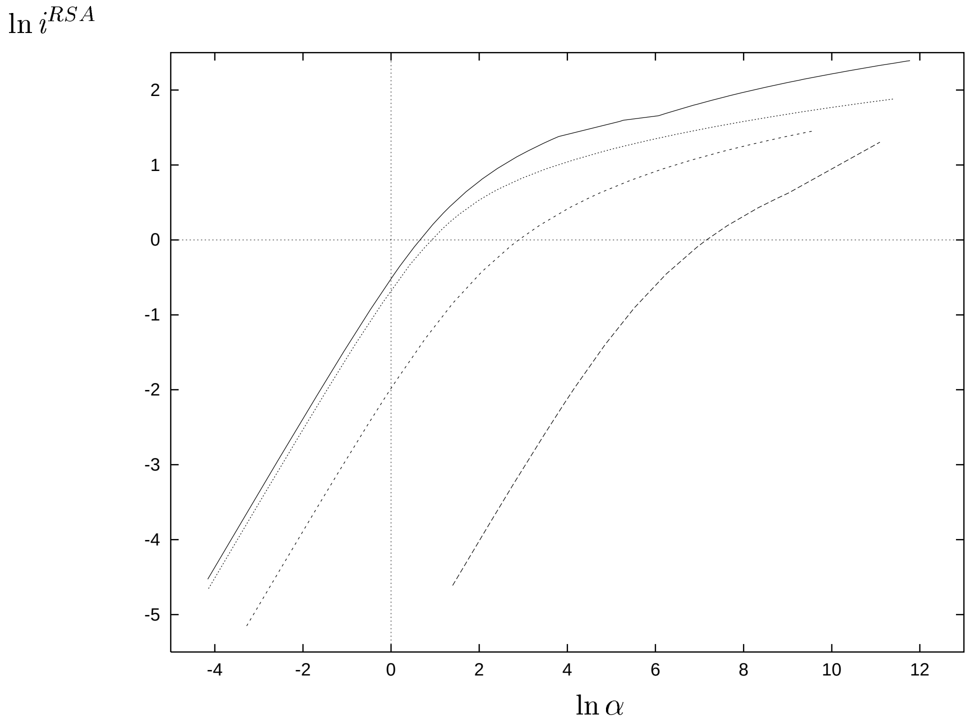


Figure 1: $\ln i^{RSA}$ vs. $\ln \alpha$ computed with the RSA for the deterministic transfer function and the HTT function, for several values of $\beta_0$ :
1. Solid line: $\beta_0 = \infty$ (Deterministic)
2. Dotted line: $\beta_0 = 10$ (near to deterministic)
3. Light dashed line: $\beta_0 = 1$
4. Dashed line: $\beta_0 = 0.1$ (not far from full stochasticity)

the following two questions arise: first, whether the two solutions do or do not coincide at least in the range of validity of the exact one. Secondly, if the exact MI can be analytically extended to greater values of $\alpha$. We will see that the answer to both questions is no, at least for the deterministic transfer function.

With respect to the first question, an expansion in powers of $\alpha$ can be easily evaluated for the deterministic case. It turns out that the corresponding Taylor coefficients coincide up to the second order, but the third is different. For instance, if the matrix $\mathbf{M}$ is proportional to the identity we observe that $i^{an} - i^{RSA} = i_1^{an} - i_1^{RSA} \approx \frac{4}{\pi^4}\alpha^3$ at the lowest order in $\alpha$. It should be noted that $i^{an}$ is always greater than $i^{RSA}$ (see figure 2). Both graphs are very close up to an undetermined value of $\alpha$ near $\alpha = 1$ ( $\ln \alpha \approx 0$ ), from which they split away fast. Detailed numerical studies for small $\alpha$ ( $\alpha \in [.0001\,,\ .005]$ ) confirmed a cubic divergence between the two MI's with coefficient $\approx \frac{4}{\pi^4}$ and a deviation form this value less

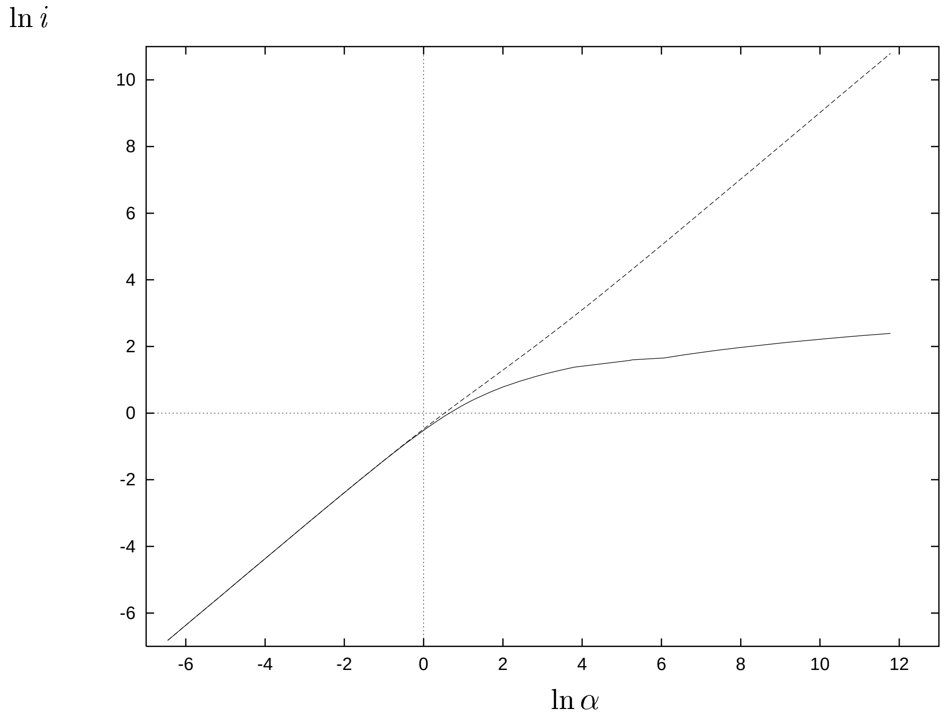


Figure 2: $\ln i$ vs. $\ln \alpha$ for the RSA solution (solid line) and for the analytical solution (dashed line), for the deterministic transfer function.

than 0.25 %.

As to the second question, the large $\alpha$ expansion of the RSA solution is (eq. (40)):

$$i_1^{RSA} \underset{\alpha \gg 1}{\approx} \ln \alpha + \frac{1}{2} + \ln A_0 + \frac{1}{2} \tau[\, \ln \mathcal{G}\, ] \; .$$

It is consistent with what is known about the continuous outputs, which should be reproduced when $\alpha$ goes to infinity. On the other hand, the analytical solution gives, in this limit:

$$i_1^{an} \underset{\alpha \gg 1}{\approx} \alpha \left( \ln 2 - \frac{1}{\pi} \right) + \frac{1}{2} \ln \alpha + \frac{1}{2} \ln \frac{2}{\pi} + \frac{1}{2} \tau[\, \ln \mathcal{G}\, ], \tag{45}$$

which is a qualitatively very different behaviour. Since the analytical solution is exact for small $\alpha$, with a convergence radius $O(1)$, the previous expansion suggests that the channel exhibits a phase transition.

# 7 Discussion of the phase transition

In this section we give a series of arguments to support the existence of a phase transition at $\alpha_c = O(1)$.

1. The first argument is provided by the behavior of the moments. The analytical computation of the integer moments, eq. (32), is exact in the thermodynamic limit. Yet, those moments cannot be correct for every value of $\alpha$. This is because they diverge at the values $\alpha_c^n = \pi/(2k^2 n)$. On the other hand since $\wp$ is a positive variable bounded by one (and then its moments should be less than one) one can conclude that eq. (26) presents critical points before those values. A natural guess would be that these singularities appear at values of $\alpha$ that follow the same behavior $1/(k^2 n)$. [3]

2. The critical point of $\ll \wp \ln \wp \gg$ is related to the critical points of the moments (26). We then expect that it has a phase transition at some $\alpha_c \sim 1/k^2$. As an example we consider the completely random channel ($k = 0$), where according to the previous argument the transition in pushed to infinity. In fact, the expansions of $i_1$ computed with the analytical and the RSA solutions coincide in the large T limit, in both regimes of $\alpha$ (eqs. (38) and (43)).

3. One could infer the existence of the critical point, observing the behaviour of the probability density $\rho(\vec{h})$. If one considers this distribution for only one replica, a dramatical change in the shape of the function takes place when $\alpha$ goes from 1 to 2. Considering its Fourier transform, eq. (53), it can be seen that $\hat{\rho}(\vec{u})$ behaves at $\infty$ like $u^{-N}$ ($u$ denotes the modulus of $\vec{u}$) , while the volume element behaves as $u^P$. This means that this function is integrable (i.e., $\hat{\rho}(\vec{u})$ is a $L^1$ function[4] ) up to $\alpha = \mathbf{1}$. It is also a square integrable function (i.e., it belongs to $L^2$) in this range. From $\alpha = 1$ to $\alpha = 2$ it is no longer a $L^1$ function, but it still belongs to $L^2$; and beyond $\alpha = 2$ it is no longer in $L^2$. What does this mean in terms of $\rho(\vec{h})$ ?

    - **Below** $\alpha = \mathbf{1}$, $\hat{\rho}(\vec{u}) \in L^1$. Consequently its Fourier transform $\rho(\vec{h})$ is bounded (that is, belongs to $L^\infty$). Besides, since $\rho(\vec{h})$ is a probability density it is also in $L^1$. The same argument holds for its derivatives in the thermodynamic limit. This is because derivation in $\vec{h}$-space is equivalent to multiplication by powers of $\vec{u}$ in $\vec{u}$-space. Since the order of the derivatives is finite, the leading behaviour in the thermodynamic limit is not changed.
    It follows that $\rho(\vec{h})$ and all its derivatives belong to $L^1 \cap L^\infty$. This means that $\rho(\vec{h})$ belongs to the Schwartz's class (see for instance [21]). Then, it is a very regular, fast decreasing function.

[3] **Remark:** Since the moments factorize as the product of contributions related to each eigenvalue of $\mathbf{M}$, we expect that there is a critical value of $\alpha$ for each of them. The functional behaviour at these transitions is the same, differing only in the critical value of $\alpha$ where they occur. This is not a serious complication, although one sould keep in mind that the distribution of eigenvalues of $\mathbf{M}$ is relevant.

[4] $L^q = \{f : \left[\int |f|^q\right]^{\frac{1}{q}} < +\infty\}$

- **Beyond $\alpha = 2$**, $\hat{\rho}$ is no longer in $L^2$, so $\rho$ cannot belong to $L^2$ either (since the Fourier transform is an isomorphism on $L^2$). Thus, $\rho$ cannot belong to $L^\infty$ (as $\rho$ belongs to $L^1$, then it would belong to $L^1 \cap L^\infty \subset L^2$, which is a contradiction ). Then the graph of $\rho$ is broken by one or more divergences to $\infty$.
- **Between $\alpha = 1$ and $\alpha = 2$**, the transition between the other two regimes has to occur.

For more than one replica, heuristic arguments permit to say that the main contributions to the characteristic function behaves like $u^{-N}$, independent of the number of replicas. The volume element behaves like $u^{P(n+1)}$. By the same arguments used in the case of a single replica, now $\rho(\{\vec{h}^a\}_{a=0}^n)$ exhibits a transition which takes place between $\alpha = 1/(n+1)$ and $\alpha = 2/(n+1)$. This is in agreement with the main conclusion obtained in the first comment.

Thus, we have proved that the joint probability distribution of the replicated fields, $\rho(\{\vec{h}^a\}_{a=0}^n)$, undergoes a phase transition at a some finite $\alpha$. Recalling that $\ll \wp^{n+1} \gg$ is calculated averaging $\prod_{a=0}^n \prod_{i=1}^P f(h_i^a)$ with this function, it is thus reasonable to think that the integer moments of $\wp$ and the MI could exhibit a phase transition caused by the transition in the own distribution.

4. Another argument in favor of the existence of a transition is given by the behavior of the information capacity. It has been proved [1] that this quantity has a third order transition for the deterministic channel. The high order of this transition makes the function rather smooth and the critical point hard to detect. The information capacity is only an upper bound of the MI, but it is plausible that the latter has a similar behavior.

These comments lead us to conclude that the MI undergoes a phase transition. What is then the meaning of the RSA solution? The expansion in powers of $\alpha$ of the RSA solution differs from that of the exact one at the third order, which is precisely the order of the transition for the information capacity. Besides, a detailed study of the RSA solution shows that the dependence of $i_1^{RSA}$ on $\alpha$ is infinitely smooth: this solution exhibitis no change in its behavior.

The conclusion is that the completely symmetric ansatz does not provide a wide enough family of solutions and the maximal MI is not attained by this ansatz. This explains why the exact solution is always above the RSA one. So, RSA *seems to be* a smooth regularization of the true MI. This would explain why it splits away from the true MI in a cubic way, suppossing that the latter possesses a third order transition. On the other hand, the behaviour at large $\alpha$ of the RSA solution is consistent with that of the information capacity and of the MI in a network with continuous output. Then, it is plausible that the RSA provides a smoothening for MI which asymptotically has the correct behavior, but which masks completely the critical point.

## 8 Further steps: Beyond the RSA

We have explored the possibility of breaking the replica symmetry by modifying the ansatz for $\tilde{\mathbf{U}}$ and $\tilde{\mathbf{V}}$ (See Appendix C). Our first attempt consisted in the usual RSB ansatz. After rather lenghty calculations this led us to exactly the same solution given by the RSA.

We also tried what can be called the Segregated Ansatz (SA), in which the first of the replicas is split from the other $n$. Then we assume:

$$\left\{ \begin{array}{ccll} \tilde{U}^0_{00} & = & U_0, & \\ \tilde{U}^0_{0b} = \tilde{U}^0_{b0} & = & U_1, & b = 1, ..., n \\ \tilde{U}^0_{aa} & = & U_2, & a = 1, ..., n \\ \tilde{U}^0_{ab} & = & U_3, & \forall a \neq b \ \in \{1, ..., n\} \end{array} \right. \tag{46}$$

and analogously for $\tilde{\mathbf{V}}_0$ .[5]

Under this ansatz the RS solution verifies again the SP equations. But in addition, an new infinite set of functions of $\alpha$ appeared that also verify the SP equations. The MI will then be given, at each $\alpha$, by the function providing the maximal MI. We observed that at large $\alpha$ this infinity of solutions contributes below the RSA. However, for arbitrary values of $\alpha$ the problem is too complicated to deal with.

## 9 Conclusions

In this paper we investigated the information processing by a noisy perceptron channel. Our network has $N$ real-valued input and $P$ binary output neurons, which state is determined by the joint probability distribution of the input and output states $P(\vec{v}, \vec{\xi})$. We performed the calculation for a general continuous, non decreasing and bounded transfer function, depending on a noise parameter. Our study generalizes previous results obtained for deterministic channels [1] using the replica technique. We also give the explicit expressions for the mutual information at different asymptotic regimes of the load parameter $\alpha = P/N$ and the noise $\beta$.

The mutual information per input unit can be decomposed in two pieces: $i = i_1 - i_2$. The second term, which extracts the wrong bits of information (the equivocation), can be calculated exactly because of the factorization of the probability. The entropic part $i_1$ is more difficult to compute. Here we computed it by means of the replica technique and analytical methods.

Our main result is that for values of $\alpha$ up to some value $O(1)$ there exists an exact solution for $i$, which we found explicitly (eq. (34)). This solution is *different* from the replica symmetry ansatz solution (eqs. (27) to (31)). A numerical computation of both solutions gives the remarkable result that they are *extremely close* to each other up to $\alpha \sim 1$ (Fig. 2).

[5]This ansatz is justified because it splits a typical $n \times n$ box from the matrices, which are $(n+1) \times (n+1)$. This splitting allows the segregated replica to behave independently from the others.

A small $\alpha$ expansion shows that the two solutions are equal up to the second order. Although the corresponding Taylor expansions differ above the third order, the numerical agreement up to $\alpha \sim 1$ is excellent (a relative difference of less than 0.9% up to $\alpha = 0.1$). This is due to intriguing cancellations between higher orders.

Our conclusion is that there exists a critical value $\alpha_c$ of order one, above which a drastic change of the mutual information occurs. This signals the appearance of a phase transition. Above $\alpha_c$ the analytical solution is not valid, one of the reasons is that it does not have the correct large $\alpha$ behaviour (it violates a bound given by the information capacity). On the other hand, even if the replica symmetric solution is wrong at small $\alpha$, it does have the correct asymptotic behaviour. Our interpretation of the RS solution is that it should be considered as a smooth regularization of the true mutual information, which is given by the analytical solution, eq. (34), for $\alpha < \alpha_c$. The precise value of $\alpha_c$ cannot be determined by our techniques. There is numerical evidence [22] supporting the validity of the analytical solution and the conjecture that the $RSA$ solution is an excellent interpolation between the small and large $\alpha$ behaviors. The analysis of the origin of the discrepancies between the RSA and the analytical approaches will be the subject of a future work.

We have also explored some other schemes beyond the completely symmetric ansatz. These are based on different types of replica symmetry breaking ansätze such as the usual breaking of the symmetry [3] and the separation of the first replica from the others. In the first case it was shown that the new solution coincides with the symmetric one. In the second, and because of the complexity of the problem, we have not been able to give an explicit final result.

## Note added in proof

Part of this work was presented at the "Interdisciplinary Workshop on Neural Networks", Würzburg, Germany,(October' 95) and at the "Fisica Estadistica'96" meeting, Zaragoza, Spain (May'96).

## Acknowledgements

We warmly thank J-P Nadal for fruitful discussions on this paper. This work was supported by a Spanish grant PB 96-47. Antonio Turiel is financially supported by a FPI grant from Comunidad Autónoma de Madrid. Elka Korutcheva was financially supported by the Spanish Ministry of Science and Education, and partially by Contract 608 from the Bulgarian Scientific Foundation. E.K. also thanks the Department of Fundamental Physics of UNED for useful discussions during the preparation of the present work.

# Appendices

## A Calculation of $I_2$

To compute $I_2$ in the general case, we use again the fact that $\vec{\xi}$ and $\vec{h}$ are deterministically related, which leads to

$$I_2 = \sum_{i=1}^{P} \int d\mathbf{J} \;\; \rho(\mathbf{J}) \int d^P\vec{h} \;\; \rho(\vec{h}|\mathbf{J})\mathcal{H}(h_i) \tag{47}$$

or, in terms of the Fourier transforms of the field distribution $\rho(\vec{h}|\mathbf{J})$ and of $\mathcal{H}(h_i)$ ($\hat{\rho}(\vec{u}|\mathbf{J})$ and $\hat{\mathcal{H}}(u_i)$, respectively)

$$I_2 = \sum_{i=1}^{P} \int d\mathbf{J} \;\; \rho(\mathbf{J}) \int d^P\vec{u} \;\; \hat{\rho}(\vec{u}|\mathbf{J})\hat{\mathcal{H}}(u_i). \tag{48}$$

The Fourier transform of the field distribution is computed in Appendix B. One has

$$\hat{\rho}(\vec{u}|\mathbf{J}) = e^{-2\pi^2\vec{u}\boldsymbol{\Delta}\vec{u}^t} \;, \tag{49}$$

where $\boldsymbol{\Delta} = \mathbf{J}\,\mathbf{C}\,\mathbf{J}^t \in M_{P\times P}(\Re)$. After integrating over the $\mathbf{J}$ in eq. (48), we obtain:

$$I_2 = \sum_{i=1}^{P} \int d^P\vec{u} \;\; \hat{\rho}(\vec{u})\hat{\mathcal{H}}(u_i) \;, \tag{50}$$

where $\hat{\rho}(\vec{u})$ is the characteristic function of $\rho(\vec{h})$. Although we cannot calculate the probability density of $\vec{h}$, we can have an explicit expression for its Fourier transform by comparing eqs. (48) and (50). Replacing eq.(49) in eq.(48) we obtain

$$\hat{\rho}(\vec{u}) = \frac{1}{\sqrt{\det\left(\mathbf{Id}_{NP\times NP} + 4\pi^2\mathbf{U}\otimes\mathbf{M}\right)}} \;. \tag{51}$$

Here,

- $\mathbf{Id}_{NP\times NP}$ is the identity matrix in NP dimensions .
- "$\bigotimes$" stands for the tensor product between $\mathbf{U}$ and $\mathbf{M}$.

- $\mathbf{M} = \mathbf{\Gamma}\,\mathbf{C}$ is a constant, $N$ dimensional matrix. Since $\vec{h}$ is of order one, $\mathrm{Tr}(\mathbf{M})$ is also of order one (Tr stands for the trace).
- $\mathbf{U}$ is a $P$ dimensional matrix defined as the projector on $\vec{u}$: $(\mathbf{U})_{ii'} = u_i u_{i'}$.

Since $\hat{\rho}(\vec{u})$ is invariant under similarity transformations of $\mathbf{M}$, it can be expressed as

$$\hat{\rho}(\vec{u}) = \frac{1}{\prod_{j=1}^{N} \sqrt{\det\left(\mathbf{Id}_{P\times P} + 4\pi^2 m_j \mathbf{U}\right)}}, \tag{52}$$

where $\{m_j\},\ j = 1, ..., N$, is the set of eigenvalues of $\mathbf{M}$. The matrix $\mathbf{U}$ has only one non-zero eigenvalue, which is $|\vec{u}|^2 = \vec{u} \cdot \vec{u}$ and thus:

$$\hat{\rho}(\vec{u}) = e^{-\frac{1}{2}\ln \prod_{j=1}^{N}(1+4\pi^2 m_j |\vec{u}|^2)} . \tag{53}$$

The computation of $I_2$ does not need the whole joint distribution $\rho(\vec{h})$ but only the marginals $\rho(h_i); i = 1, \ldots, P$. By permutation symmetry, it is obvious that all of them are given by the same function. Let us compute for example $\rho(h_1)$. Its Fourier transform is

$$\hat{\rho}(u_1) = \hat{\rho}(u_1, 0, \overset{P-1}{\ldots}, 0) \ = \ e^{-\frac{1}{2}\sum_{j=1}^{N} \ln\left(1+4\pi^2 m_j (u_1)^2\right)} \tag{54}$$

and since all the marginals are the same function, all the terms in $I_2$ are the same. Then $I_2$ reads

$$I_2 = P \int_{-\infty}^{\infty} dh\ \rho(h) \mathcal{H}(h) . \tag{55}$$

So far there is no hypothesis upon the matrix $\mathbf{M}$. Particularly interesting is the case in which all the $m_j$'s are of the same order, namely, of order $1/N$ (as we have already said, $\mathrm{Tr}(\mathbf{M})$ is $\mathrm{O}(1)$ ). In this particular case,

$$\hat{\rho}(u) \underset{N\gg 1}{\approx} e^{-\frac{1}{2}\sum_{j=1}^{N} 4\pi^2 m_j u^2} + O(e^{-N}) = e^{-2\pi^2 \bar{M} u^2} + O(e^{-N}), \tag{56}$$

where $\bar{M}$ $=\mathrm{Tr}(\mathbf{M})$. In the thermodynamic limit the term $O(e^{-N})$ becomes negligible and $\rho(h)$ is :

$$\lim_{N\to\infty} \rho(h) = \frac{e^{-h^2/(2\bar{M})}}{\sqrt{2\pi \bar{M}}}. \tag{57}$$

(Note that this expression makes explicit the reason why $\bar{M} = O(1)$). The conditional output entropy now is :

$$I_2 = P \int_{-\infty}^{\infty} dz\ \frac{e^{-z^2}}{\sqrt{\pi}}\ \mathcal{H}(\sqrt{2\bar{M}}\, z) . \tag{58}$$

## B Computation of $\hat{\rho}(\vec{u}|\mathbf{J})$ for $P > N$

We define the Fourier transform of a function $F(\vec{h})$ as the function $\hat{F}(\vec{u})$ given by :

$$\hat{F}(\vec{u}) = \int d^P\vec{h}\ \ F(\vec{h})e^{-2\pi i\vec{h}\cdot\vec{u}}\ . \tag{59}$$

The evaluation of $\hat{\rho}(\vec{u}|\mathbf{J})$ in the case $P \leq N$ is simple. This is because for almost every $\mathbf{J}$ the random vector $\vec{h}$ follows a Gaussian distribution with correlation matrix $\mathbf{\Delta} = \mathbf{J}\ \mathbf{C}\mathbf{J}^t$ and $\det(\mathbf{\Delta})\neq 0$. Then,

$$\hat{\rho}(\vec{u}|\mathbf{J}) = e^{-2\pi^2\vec{u}\mathbf{\Delta}\vec{u}^t}. \tag{60}$$

We now prove that this equation is still true when $P > N$. Let us first notice that in this case $\det(\mathbf{\Delta})$ is necessarily null, and consequently the random vector $\vec{h}$ is not Gaussian.

Let us compute $\wp(\vec{v}|\mathbf{J})$ for the particular vector $\vec{v} = (1, 1, ..., 1)$. Denoting this as $\wp$ we have:

$$\wp = \int d^N\vec{\xi}\ \ \frac{e^{-\frac{1}{2}\vec{\xi}\,(\mathbf{C})^{-1}\vec{\xi}^t}}{\sqrt{\det\,(2\pi\mathbf{C})}}\prod_{i=1}^{P} f(h_i) \quad = \quad \int d^P\vec{u}\ \ \hat{\rho}(\vec{u}|\mathbf{J})\prod_{i=1}^{P}\hat{f}(u_i) \tag{61}$$

with $h_i = \sum_{j=1}^N J_{ij}\ \xi_j$ and $u_i$ being its conjugate Fourier variable. For $P > N$, the first N components of $\vec{h}$ are independent random variables and the other $P - N$ depend upon the former (for almost every $\mathbf{J}$).

We split the matrix $\mathbf{J}$ into two matrices: $\mathbf{K} \in M_{N\times N}(\Re)$, $K_{jj'} = J_{jj'}$; and $\mathbf{L} \in M_{P-N\times N}(\Re)$, $L_{kj} = J_{N+k,j}, k = 1, ..., P - N;\ j = 1, ..., N$:

$$\mathbf{J}\ =\ \begin{pmatrix} \mathbf{K} \\ \hline \mathbf{L} \end{pmatrix}$$

and for almost every $\mathbf{J}$ , $\mathbf{K}$ is invertible. Then, we split $\vec{h} = (\vec{h}^{\,0}, \vec{h}^{\,1})$, $\vec{h}^{\,0} \in \Re^N$ and $\vec{h}^{\,1} \in \Re^{P-N}$. Moreover, $\vec{h}^{\,0}$ is gaussianly distributed with zero mean and covariance matrix $\mathbf{\Delta}_0 = \mathbf{K}\,\mathbf{C}\,\mathbf{K}^t$, and $\vec{h}^{\,1} = (\mathbf{L}\,\mathbf{K}^{-1})\ \vec{h}^{\,0}$.

In this way, we obtain $\prod_{i=1}^P f(h_i) = \prod_{j=1}^N f(h_j^0)\ \prod_{k=1}^{P-N} f([(\mathbf{L}\,\mathbf{K}^{-1})\ \vec{h}^{\,0}]_k)$, and hence $\wp$ can be written as:

$$\wp = \int d^N\vec{h}^{\,0}\ \ \frac{e^{-\frac{1}{2}\vec{h}^{\,0}(\mathbf{\Delta}_0)^{-1}\vec{h}^{\,0^t}}}{\sqrt{\det\,(2\pi\mathbf{\Delta}_0)}}\prod_{j=1}^{N} f(h_j^0)\ \prod_{k=1}^{P-N} f([(\mathbf{L}\,\mathbf{K}^{-1})\ \vec{h}^{\,0}]_k)\ . \tag{62}$$

If $g(\vec{h})$ is a function of vectorial argument, and $f(x)$ has real argument, we have that:

$$(g(\vec{h})\,f(\vec{a}\cdot\vec{h}))^{\wedge}(\vec{u}) = \int_{-\infty}^{\infty} dc\ \hat{g}(\vec{u} - c\,\vec{a})\hat{f}(c)\ , \tag{63}$$

where the hat symbol stands for the Fourier transform and $\vec{a}$ is an arbitrary constant vector. It should be noted that $\hat{g}$ is a multidimensional Fourier transform while $\hat{f}$ is the one-dimensional Fourier transform.

Let us denote by $\vec{d}_k$ the $(P-N)$ N-dimensional vectors defined by the rows of $\mathbf{L}\,\mathbf{K}^{-1}$. Applying the previous formula to the expression for $\wp$, and after using the Bessel-Plancherel identity, we obtain:

$$\wp = \int d^N\vec{u}' \;\; e^{-2\pi^2\vec{u}'\mathbf{\Delta}_0\vec{u}'^t} \int d^{P-N}\vec{c} \;\prod_{j=1}^{N} \hat{f}(u'_j - \sum_{k=1}^{P-N} c_k(\vec{d}_k)_j) \prod_{k=1}^{P-N} \hat{f}(c_k) \; . \tag{64}$$

Interchanging now the order of integrations and performing the change of variables $\vec{u}^{\,0}$, related via $\vec{u}' = \vec{u}^{\,0} + \sum_{k=1}^{P-N} c_k\vec{d}_k$, we obtain:

$$\wp = \int d^{P-N}\vec{c} \;\int d^N\vec{u}^{\,0} \;\prod_{j=1}^{N} \hat{f}(u_j^0) \prod_{k=1}^{P-N} \hat{f}(c_k)$$
$$e^{-2\pi^2(\vec{u}^{\,0}\mathbf{\Delta}_0\vec{u}^{\,0t} + \sum_{k,k'=1}^{P-N} c_k c_{k'}\vec{d}_k\mathbf{\Delta}_0\vec{d}^{\,t}_{k'} + 2\sum_{k=1}^{P-N} c_k\vec{u}^{\,0}\mathbf{\Delta}_0\vec{d}^{\,t}_k)} . \tag{65}$$

It is convenient to combine $\vec{u}^{\,0}$ and $\vec{c}$ in a single P dimensional vector $\vec{u} = (\vec{u}^{\,0}, \vec{c})$. Expressing eq. (65) in terms of this vector, we can use the vectors $\vec{d}_k$ to simplify the bilinear expression in the exponent as:

$$\wp = \int d^P\vec{u} \;\; e^{-2\pi^2\vec{u}\mathbf{\Delta}\vec{u}^t} \prod_{i=1}^{P} \hat{f}(u_i) \;\; , \;\; \mathbf{\Delta} = \mathbf{J}\,\mathbf{C}\mathbf{J}^t \; , \tag{66}$$

that depends only on $\mathbf{\Delta}$. From the right hand side of eq. (61) we have

$$\hat{\rho}(\vec{u}|\mathbf{J}) = e^{-2\pi^2\vec{u}\mathbf{\Delta}\vec{u}^t} \; . \tag{67}$$

# C RSA derivation for $I_1$

We now derive the Saddle Point (SP) equations. These are then simplified by using the Replica Symmetry Ansatz (RSA) [3]. The first order parameters are the overlap of two replicated Fourier transforms of the local field:

$$\tilde{U}_{ab} = \frac{1}{P}\vec{u}_a \cdot \vec{u}_b \; , \; a,b = 0,...,n \tag{68}$$

The pre-factor is taken in order to ensure that it is of order one. These are the elements of a matrix $\tilde{\mathbf{U}} \in M_{(n+1)\times(n+1)}(\Re)$ . Then, the Fourier transform of the joint distribution of the replicated local fields, eq. (24), can be expressed in terms of this matrix as:

$$\hat{\rho}(\{\vec{u}^a\}) = \prod_{a\le b} \int_{-\infty}^{\infty} dU_{ab}\delta(U_{ab} - \frac{1}{P}\vec{u}_a\vec{u}_b) e^{-\frac{1}{2}\sum_{j=1}^{N} \ln\det[\mathbf{Id}_{(n+1)\times(n+1)} + 4\pi^2 P m_j \tilde{\mathbf{U}}]} \; . \tag{69}$$

Now we introduce an order parameter $\tilde{\mathbf{V}}$, conjugated to $\tilde{\mathbf{U}}$. To linearize the quadratic form in the $\vec{u}^a$'s we will use $P$ new variables $\vec{w}_i$, which are $(n+1)$-dimensional vectors.

After substituting eq. (69) in eq. (26), we can perform the integrals over the $\vec{u}_\alpha$. Since these integrals are the anti-Fourier transforms of the $\hat{f}(u_i^a)$, the $\ll \wp^{n+1} \gg$ can be expressed in terms of the product of the transfer functions simply :

$$\ll \wp^{n+1} \gg = \int \prod_{a\le b} (-iPd\tilde{V}^{ab}\, d\tilde{U}_{ab})$$
$$e^{2\pi P \sum_{a\le b} \tilde{U}_{ab}\tilde{V}^{ab} - \frac{1}{2}\sum_{j=1}^{N} \ln\det[\mathbf{Id}_{(n+1)\times(n+1)} + 4\pi^2 m_j \tilde{\mathbf{U}}]}$$
$$\prod_{l=1}^{P} \int d^{n+1}\vec{w}_l \; \frac{e^{-\frac{1}{2}\vec{w}_l (\tilde{\mathbf{V}})^{-1}\vec{w}_l^t}}{\sqrt{\det(2\pi\tilde{\mathbf{V}})}} \prod_{a=0}^{n} f(\frac{w_l^a}{\sqrt{\pi}}) \; . \tag{70}$$

($i \equiv \sqrt{-1}$). This can be written as:

$$\ll \wp^{n+1} \gg = \int \prod_{a\le b} (-iPd\tilde{V}^{ab}\, d\tilde{U}_{ab})$$
$$e^{2\pi P \sum_{a\le b} \tilde{U}_{ab}\tilde{V}^{ab} - \frac{1}{2}\sum_{j=1}^{N} \ln\det[\mathbf{Id}_{(n+1)\times(n+1)} + 4\pi^2 P m_j \tilde{\mathbf{U}}] + P\ln Z(\tilde{\mathbf{V}})} \; , \tag{71}$$

where

$$Z(\tilde{\mathbf{V}}) = \int d^{n+1}\vec{w} \; \frac{e^{-\frac{1}{2}\vec{w}(\tilde{\mathbf{V}})^{-1}\vec{w}^t}}{\sqrt{\det(2\pi\tilde{\mathbf{V}})}} \prod_{a=0}^{n} f(\frac{w^a}{\sqrt{\pi}}). \tag{72}$$

In the large $N$ limit ($\alpha = \frac{P}{N}$ fixed), the integrals over $\tilde{\mathbf{U}}$ and $\tilde{\mathbf{V}}$ in eq. (71) can be solved by the SP method. This gives:

$$\ll \wp^{n+1} \gg \approx e^{G(\tilde{\mathbf{U}}_0, \tilde{\mathbf{V}}_0)} \tag{73}$$

where $\tilde{\mathbf{U}}_0$ and $\tilde{\mathbf{V}}_0$ are the SP values and

$$G = 2\pi P \sum_{a\le b} \tilde{U}_{ab}\tilde{V}^{ab} - \frac{1}{2}\sum_{j=1}^{N} \ln\det[\mathbf{Id}_{(n+1)\times(n+1)} + 4\pi^2 P m_j \tilde{\mathbf{U}}] + P\ln Z(\tilde{\mathbf{V}}). \tag{74}$$

The RSA is:

$$\left\{ \begin{array}{rcll} \tilde{U}^0_{aa} & = & U_0, & \forall a \\ \tilde{U}^0_{ab} & = & U_1, & \forall a \ne b \end{array} \right. \quad \text{and} \quad \left\{ \begin{array}{rcll} \tilde{V}^0_{aa} & = & V_0, & \forall a \\ \tilde{V}^0_{ab} & = & V_1, & \forall a \ne b \end{array} \right. \tag{75}$$

The starting point is eq. (73), where the function $G$ is evaluated with the RSA given in eq. (75). Defining the matrix $1\!\!1 \in M_{(n+1)\times(n+1)}(\Re)$ as $(1\!\!1\,)_{ab} = 1\ \forall\ a,b$, $\tilde{\mathbf{V}}$ can be expressed as:

$$\tilde{\mathbf{V}} = (v_0 - \frac{1}{2}v_1)\mathbf{Id}_{(n+1)\times(n+1)} + \frac{1}{2}v_1 1\!\!1 \tag{76}$$

and its inverse is:

$$(\tilde{\mathbf{V}})^{-1} = x_0\mathbf{Id}_{(n+1)\times(n+1)} + x_1 1\!\!1\ , \tag{77}$$

where $x_0 = \frac{1}{v_0-\frac{1}{2}v_1}$ and $x_1 = -\frac{\frac{1}{2}v_1}{(v_0-\frac{1}{2}v_1)(v_0+\frac{n}{2}v_1)}$. This form of $(\tilde{\mathbf{V}})^{-1}$ allows us to express $Z(\tilde{\mathbf{V}})$ (eq. (72)) in a more convenient way:

$$\begin{aligned} Z(\tilde{\mathbf{V}}) = (2\pi)^{-\frac{n+1}{2}}\, x_0^{\frac{n}{2}} \sqrt{x_0 + (n+1)x_1} \\ \int d^{n+1}\vec{w}\ \ e^{-\frac{1}{2}x_0\sum_{a=0}^{n} w_a^2 - \frac{1}{2}x_1(\sum_{a=0}^{n} w_a)^2} \prod_{a=0}^{n} f(w_i^a/\sqrt{\pi})\ . \end{aligned} \tag{78}$$

Notice that $Z(\tilde{\mathbf{V}}) \to \frac{1}{2}$ as $n \to 0$. We now expand $Z(\tilde{\mathbf{V}})$ up to order $n$, $Z \approx \frac{1}{2} + n\,h(x_0,x_1)$, $h = \partial Z/\partial n|_{n=0}$. Up to this order we obtain:

$$\begin{aligned} G \approx 2\pi P(n+1)v_0u_0 + \pi P\, n\, u_1v_1 - \frac{1}{2}\sum_{j=1}^{N}(1+4\pi^2Pm_ju_0) - \frac{n}{2}\sum_{j=1}^{N}\frac{4\pi^2Pm_ju_1}{1+4\pi^2Pm_ju_0} \\ -\frac{n}{2}\sum_{j=1} N\ln\left(1+4\pi^2Pm_j(u_0-u_1)\right) - P\ln 2 + 2Pnh(x_0,x_1)\ . \end{aligned} \tag{79}$$

The SP equations extremize G with respect to its variables. From the SP equation $\partial G/\partial v_0 = 0$ one obtains that $u_0$ is linear in $n$: $u_0 = n\tilde{u}$. Replacing eq. (79) in eq. (73), and this in eq. (22) we have:

$$I_1^{RSA} = -2\pi Pv_0\tilde{u} + \pi Pv_1u + 2\pi^2P\bar{M}\tilde{u} - 2\pi^2P\bar{M}u + \frac{1}{2}\sum_{j=1}^{N}\ln\left(1+4\pi^2Pm_ju\right) - 2Ph(x_0,x_1) \tag{80}$$

where $u = -u_1$. The SP equations are :

$$\left\{\begin{array}{rcl} v_0 & = & \pi\bar{M} \\ v_1 & = & 2\pi\bar{M} - 2\pi\sum_{j=1}^{N}\frac{m_j}{1+4\pi^2Pm_ju} \\ \tilde{u} & = & -\frac{1}{\pi}\partial h/\partial v_0 \\ u & = & \frac{2}{\pi}\partial h/\partial v_1 \end{array}\right. \tag{81}$$

$\tilde{u}$ is a Lagrange multiplier, that can be easily removed by substituting the value of $v_0$ in $I_1^{RSA}$.

The evaluation of $h(x_0, x_1)$ requires some care. We express $Z$ as:

$$Z(\tilde{\mathbf{V}}) = (2\pi)^{-\frac{n+1}{2}} x_0^{\frac{n}{2}} \sqrt{x_0 + (n+1)x_1} \int_{-\infty}^{\infty} dx \frac{e^{-\frac{1}{2}x^2}}{\sqrt{2\pi}} [\mathcal{L}(x)]^{n+1} , \tag{82}$$

where

$$\mathcal{L}(x) = \int_{-\infty}^{\infty} dw \; e^{-\frac{1}{2}x_0 w^2 + i\sqrt{x_1}\, xw} f(x/\sqrt{\pi}) . \tag{83}$$

Computing the term of order $n$ of $Z$ we obtain:

$$h(x_0, x_1) = -\frac{1}{2}\ln 2 - \frac{1}{2}\sqrt{\frac{x_0 + x_1}{2\pi x_0}} \tilde{M}(x_0, x_1) \tag{84}$$

with

$$\tilde{M}(x_0, x_1) = \int_{-\infty}^{\infty} dz \; e^{-\frac{x_0 + x_1}{2x_0} z^2} \tilde{F}(\tilde{g}(z)) , \tag{85}$$

where the function $\tilde{F}(y)$ is defined by:

$$\tilde{F}(y) = \frac{1}{2}\ln\frac{1-y}{1+y} - \frac{1}{2}\ln(1-y^2) \tag{86}$$

and its argument is :

$$\tilde{g}(z) = \sqrt{\frac{2x_0}{\pi}} \, e^{\frac{1}{2}\frac{x_1}{x_0}z^2} \int_{-\infty}^{\infty} dw \; e^{-\frac{1}{2}x_0 w^2} \sinh(\sqrt{-x_1}\, zw) f(w/\sqrt{\pi}) . \tag{87}$$

Substituting $h(x_0, x_1)$ given above in eq. (80), we obtain:

$$I_1^{RSA} = P\ln 2 + P\sqrt{\frac{x_0 + x_1}{2\pi x_0}} \, \tilde{M}(x_0, x_1) + \frac{1}{2}\sum_{j=1}^{N} \ln(1 + 4\pi^2 P m_j u) - \frac{1}{2}\sum_{j=1}^{N} \frac{4\pi^2 P m_j u}{1 + 4\pi^2 P m_j u} . \tag{88}$$

and the SP equations become:

$$\begin{cases} x_0 &= & 1/(\pi \sum_{j=1}^{N} \frac{m_j}{1+4\pi^2 P m_j u}) \\ x_1 &= & \frac{1}{\pi \bar{M}} - x_0 \\ u &= & -\frac{x_0^2}{(2\pi)^{3/2}} \left(\frac{\partial}{\partial x_0} - \frac{\partial}{\partial x_1}\right)\left[\sqrt{\frac{x_0+x_1}{x_0}} \tilde{M}(x_0, x_1)\right] \end{cases} \tag{89}$$

We can substitute in $I_1^{RSA}$ one of the parameters, for example $x_0$ . Defining $x = -\pi \bar{M} x_1$, $s = 4\pi^2 \bar{M} \, \alpha u$ and rearranging conveniently eqs. (88) and (89), we finally obtain eqs. (27) and (28).

## D Analytical derivation of $I_1$

This exact calculation starts from eq. (26). We assume that all the eigenvalues $\{m_j\}_{j=1,\dots,N}$ of the matrix $\mathbf{M}$ are at most of order $\frac{1}{N}$. We define now $\tilde{\mathbf{U}}$ in a slightly different way from eq. (68):

$$\tilde{U}_{ab} = \vec{u}^a \cdot \vec{u}^b. \tag{90}$$

These elements are of order $P$ and hence $m_j\tilde{\mathbf{U}}$ is of order $\alpha$. $\hat{\rho}(\{\vec{u}^a\})$ is computed as in subsection 4.1 . Then, the logarithm in the exponent of eq. (69) can be expanded around the identity matrix (what can be donde if $\alpha$ is less than $\frac{1}{4\pi^2 \bar{M}}$ times a geometrical factor of order one, that depends on $\mathcal{G}$)

$$\hat{\rho}(\{\vec{u}^a\}) = \prod_{m=1}^{\infty} e^{\frac{1}{2}(-1)^m \frac{(4\pi^2)^m}{m} Tr(\mathbf{M}^m) Tr(\tilde{\mathbf{U}}^m)} . \tag{91}$$

Let us remark that *this is not an approximation.* It is an *exact* derivation valid in a (undetermined) range of values $\alpha$ *of order one* . We can alternatively write this in the following form :

$$\hat{\rho}(\{\vec{u}^a\}) = e^{-2\pi^2 \bar{M} \sum_{a=0}^{n} |\vec{u}^a|^2} \prod_{m=2}^{\infty} e^{\frac{1}{2m}(-4\pi^2 \bar{M})^m N Tr(\mathcal{G}^m) Tr((\tilde{\mathbf{U}}/N)^m)} . \tag{92}$$

The second factor can now be expanded, leading to polynomials in traces of powers of $\tilde{\mathbf{U}}/N$.

Given a function $F(\{\vec{u}^a\})$, we now define its average with the transfer function (or shortly, its transfer-average) as:

$$\prec\!\!\prec F \succ\!\!\succ = \int \prod_{a=0}^{n} d^P\vec{u}^a \; e^{-2\pi^2 \bar{M} |\vec{u}^a|^2} F(\{\vec{u}^a\}) \prod_{a=0}^{n} \prod_{i=1}^{P} \hat{f}(u_i^a) . \tag{93}$$

Notice that

$$\prec\!\!\prec 1 \succ\!\!\succ = 2^{-P(n+1)}. \tag{94}$$

This property comes from the fact that $f(x) + f(-x) = 1$, and so $f(x) = 1/2 + a(x)$, where $a(x)$ is an odd function bounded by $-\frac{1}{2}$ and $\frac{1}{2}$. Notice that, although this will not be required here, in the physically reasonable cases $a(x)$ is an increasing and almost everywhere continuous function, $0 \leq a(x) \leq \frac{1}{2}$, $x > 0$ such that $a(x) \to \frac{1}{2}$ when $x \to \infty$.

In terms of these transfer-averages, the moments read:

$$\ll \wp^{n+1} \gg = 2^{-P(n+1)} + \sum_{\sigma=2}^{\infty} \sum_{\substack{2 \leq s_1 < \dots < s_\lambda \\ s_1 t_1 + \dots + s_\lambda t_\lambda = \sigma}} C_{s_1,\dots,s_\lambda}^{t_1,\dots,t_\lambda} \, \Lambda_{s_1,\dots,s_\lambda}^{t_1,\dots,t_\lambda} , \tag{95}$$

where

$$C^{t_1,\dots,t_\lambda}_{s_1,\dots,s_\lambda} = \frac{(-4\pi^2)^\sigma}{t_1!\dots t_\lambda!}(\frac{N\bar{M}}{2})^{t_1+\dots+t_\lambda}\ \frac{(Tr[(\mathcal{G})^{s_1}])^{t_1}}{s_1^{t_1}} \dots \frac{(Tr[(\mathcal{G})^{s_\lambda}])^{t_\lambda}}{s_\lambda^{t_\lambda}} \tag{96}$$

and

$$\Lambda^{t_1,\dots,t_\lambda}_{s_1,\dots,s_\lambda} = \prec\!\!\prec (Tr[(\tilde{\mathbf{U}}/N)^{s_1}])^{t_1} \dots (Tr[(\tilde{\mathbf{U}}/N)^{s_\lambda}])^{t_\lambda} \succ\!\!\succ . \tag{97}$$

These transfer-averages have a very simple expression in the thermodynamic limit, what allows us to rearrange the whole expression in a convenient way. First, we must notice that:

$$\int_{-\infty}^{\infty} du\, e^{-2\pi^2\bar{M}u^2}\ u^{2r}\hat{f}(u) = \frac{1}{2}\delta_{0r} \tag{98}$$

because $f(x) = \frac{1}{2} + a(x)$ and $a(x)$ is odd. We now prove a factorization property , eq. (103), of the transfer-averages of traces of $\tilde{\mathbf{U}}$ that will be useful to computate of $\ll \wp^{n+1} \gg$.

The trace of the r-th power of $\tilde{\mathbf{U}}$ can be written as:

$$Tr(\tilde{\mathbf{U}}^r) = \sum_{a_1,\dots,a_r=0}^{n}\ \sum_{i_1,\dots,i_r=1}^{P} u^{a_1}_{i_1}\, u^{a_1}_{i_2}\, u^{a_2}_{i_2}\, u^{a_2}_{i_3}\, u^{a_3}_{i_3}\, u^{a_3}_{i_4} \dots u^{a_{r-1}}_{i_{r-1}}\, u^{a_{r-1}}_{i_r}\, u^{a_r}_{i_r}\, u^{a_r}_{i_1} . \tag{99}$$

After taking the transfer-average on this expression, one should notice that the contribution of each term does not depend on the particular indeces $i$'s and $\alpha$'s present in that term. It only depends on the number of different variables in the term and the power of each variable. This is due to the independency and permutation symmetry of the $u$'s. It is possible to rearrange eq. (99), expressing it as the sum of each different contribution times a combinatorial factor. This factor is the number of terms giving that particular contribution. Since the contributions themselves are of order one, the thermodynamical limit is determined by the combinatorial factors. In this limit, $P \to \infty$ ($r$ kept finite) and the combinatorial factors scale as $P$ raised to the number of non-repeated indeces $i$. Then, no more than two $u$'s can be equal. Defining

$$\lambda(\bar{M}) = \int_{-\infty}^{\infty} du\, e^{-2\pi^2\bar{M}u^2}\ u\hat{f}(u) \tag{100}$$

and considering eq. (98) for $r = 0$ and $r = 1$, the transfer-average of eq. (99) can be expressed in terms of $\lambda$:

$$\prec\!\!\prec Tr[(\tilde{\mathbf{U}}/N)^r] \succ\!\!\succ \underset{N\to\infty}{=}\ 2^{-P(n+1)}\chi_r\ , \tag{101}$$

where

$$\chi_r = (n^r + (-1)^r n)(2\lambda)^{2r}\alpha^r\ . \tag{102}$$

By similar arguments, one can prove a useful factorization property. For the product of two traces we have:

$$\prec\!\!\prec Tr[(\tilde{\mathbf{U}}/N)^r]\, Tr[(\tilde{\mathbf{U}}/N)^s] \succ\!\!\succ = 2^{-P(n+1)}\, \chi_r\, \chi_s \tag{103}$$

and a similar factorization holds for the product of an arbitrary number of traces. Recalling eq. (95), this property allows us to write:

$$\ll \wp^{n+1} \gg = 2^{-P(n+1)}\; \exp\left[ N \sum_{m=1}^{\infty} \frac{1}{2m} (-4\pi^2 {\bar{M}})^m Tr[(\mathcal{G})^m]\, \chi_m \right] . \tag{104}$$

Substituting the explicit values of the $\chi$ 's, eq. (102), and preforming the sum, we have:

$$\ll \wp^{n+1} \gg = 2^{-P(n+1)}\; e^{-\frac{1}{2} Tr[\, \ln(\mathbf{Id}_{N\times N} + 16\pi^2\lambda^2 nP\mathbf{M})\,]}\; e^{-\frac{n}{2} Tr[\, \ln(\mathbf{Id}_{N\times N} - 16\pi^2\lambda^2 P\mathbf{M})\,]} . \tag{105}$$

These moments can be expressed in a more useful way. Defining $k$ by:

$$k = \int_{-\infty}^{\infty} dy\;\; y\, e^{-\frac{y^2}{2}} f(\sqrt{\bar{M}}\; y) \;, \tag{106}$$

we have $\lambda^2 = -\frac{k^2}{8\pi^3 \bar{M}}$. Using this relation in eq. (105), we finally obtain eq. (34).

## References


[1] Nadal J-P and Parga N 1993 *Network* **4** 295

[2] Korutcheva E, Nadal J-P and Parga N 1997 *Network* **8** 405

[3] Mezard M, Parisi G and Virasoro MA 1987 *Spin Glass Theory and Beyond* (World Scientific Publishing Co., Singapore)

[4] Linsker R. 1988 *Computer* **21** 105-117

[5] Hertz JA, Krogh A and Palmer RG 1991 *Introduction to the Theory of Neural Computation* (Adisson-Wesley, Redwood City CA.)

[6] Laughlin SB 1981 *Z. Naturf.* **C 36** 910

[7] Nadal JP and Parga N 1994 *Network* **4** 565

[8] Barlow HB 1961 *Sensory Communication* 217 (M.I.T. Press, Cambridge MA)

[9] Atick JJ 1992 *Network* **3** 213

[10] Atick JJ and Redlich A 1992 *Neural Computation* **4**, 196

[11] van Hateren JH 1992 *J. Comp. Physiology A* **171** 157

[12] van Hateren JH 1993 *Vision Research* **33** 257

[13] Treves A 1995 *J. Comput. Neurosci.* **2** 259

[14] Schultz S, Panzeri S, Rolls ET. and Treves A 1997 *Information Theory and the Brain* (edited by R. Baddeley, P. Földiak and P. Hancock; Cambridge University Press, Cambridge, U.K.)

[15] Schultz S and Treves A 1998 *Phys. Rev E* **57** 3302-3310

[16] Linsker R 1993 *Neural Inf. Proc. Systems* **5** 953

[17] Schuster HG 1992 *Phys. Rev. A* **46** 2131

[18] Shannon C.E. and Weaver W. 1949 *The Mathematical Theory of Communications* (Urbana, IL: University Illinois Press).

[19] Blahut RE 1988 *Principles and Practice of Information Theory.* (Adisson-Wesley, Cambridge MA)

[20] Cover TM and Thomas JA 1991 *Elements of Information Theory* (John Wiley & Sons)

[21] Rudin W 1974 *Real and Complex Analysis* (Mc Graw Hill, New York)

[22] Dominguez DRC, Maravall M, Turiel A, Ciria JC and Parga N "Numerical simulation of a binary communication channel: Comparison between a replica calculation and an exact solution" Submitted to *Europhysics Letters*

## List of Figures