Capacity bounds for the CDMA system and a neural network: a moderate deviations approach

Matthias Löwe, Franck Vermet

CAPACITY BOUNDS FOR THE CDMA SYSTEM AND A NEURAL NETWORK: A MODERATE DEVIATIONS APPROACH

MATTHIAS LÖWE AND FRANCK VERMET

ABSTRACT. We study two systems that are based on sums of weakly dependent Bernoulli random variables that take values ± 1 with equal probabilities. We show that already one step of the so-called Soft-Decision Parallel Interference Cancellation, used in the third generation of mobile telecommunication CDMA, is able to considerably increase the number of users such a system can host. We also consider a variant of the well-known Hopfield model of neural networks. We show that this variant proposed by Amari and Yanai [2] has a larger storage capacity than the original model.
Both situations lead to the question of the moderate deviations behavior of a sum of weakly dependent Bernoulli random variables. We prove a moderate deviations principle for such a sum on the appropriate scale.

RÉSUMÉ. Nous étudions deux systèmes basés sur des sommes de variables aléatoires de Bernoulli valant ± 1 avec égale probabilité et faiblement dépendantes. Nous montrons qu'une seule étape de la méthode de suppression d'interférences SD-PIC, utilisée dans la troisième génération de télécommunication mobile CDMA, permet déjà d'augmenter considérablement le nombre d'utilisateurs supporté par le système. Nous considérons également une variante du modèle neuronal de Hopfield. Nous montrons que cette variante, proposée par Amari et Yanai [2], admet une capacité de stockage supérieure au modèle original.
Les deux situations conduisent à l'étude des déviations modérées d'une somme de variables aléatoires de Bernoulli faiblement corrélées. Nous montrons un principe de déviations modérées pour une telle somme convenablement normalisée.

1. INTRODUCTION

In this article we study two questions, which at first glance only have very little in common: How many users is a third generation mobile communication system able to host, if it works with a certain form of interference cancellation? and: How much information can be safely stored in an improved form of the Hopfield model of neural networks?
The reason that these questions can be treated in an integrative framework is that both models are formulated in terms of i.i.d. Bernoulli random variables that take the values $+1$ and -1 with equal probabilities. For both these models the questions addressed above then boil down to a problem of moderate deviations for a sum of Bernoulli random variables that are correlated in a rather peculiar way. Interestingly, this correlation is the same for the mobile communication system and the neural network and even more, one also is interested in the same order or magnitude for the deviations.
In recent years the analysis of moderate deviations has been introduced as one branch of the theory of limit theorems, next to well-established subjects such as central limit

Date: April 10, 2008.
1991 *Mathematics Subject Classification.* 82C32, 82B44, 60K35, 94A05, 94A15.
Key words and phrases. moderate deviations, large deviations, neural networks, storage capacity, Hopfield Model, Code-division multiple access (CDMA) systems, parallel interference cancellation.

theorems and large deviations. Here the term "moderate deviations" refers to the analysis of probabilities in a regime between the regime of the Central Limit Theorem and the regime of a Law of Large Numbers, which is referred to as the large deviations regime.
For example, it is known that for a sum of i.i.d. random variables $Y_1, Y_2, \ldots$, under appropriate conditions (the most general conditions can be found in [17]) the following "subexponential" estimates holds true in the moderate deviations regime, i.e. for a scaling between the $\frac{1}{\sqrt{n}}$ -scaling for a Central Limit Theorem and the $\frac{1}{n}$ -scaling for the Law of Large Numbers:

$$\lim_{n\to\infty} n^{1-2\alpha} \log \mathbb{P}\left(\sum_{i=1}^{n} Y_i \geq n^{\alpha} a\right) = -a^2/2$$

as well as

$$\lim_{n\to\infty} n^{1-2\alpha} \log \mathbb{P}\left(\sum_{i=1}^{n} Y_i \leq -n^{\alpha} a\right) = -a^2/2$$

for $a > 0$ and $1/2 < \alpha < 1$.
Often, such moderate deviations estimates are very useful (see e.g. [34]) as we will also see in the present paper. However, in our case the random variables will be dependent and this often makes life much more complicated.
The rest of this paper is organized as follows. In Section 2 we will address a problem in mobile telecommunication. This leads to a question about the moderate deviations behavior of a sum of weakly correlated Bernoulli variables. Once this is settled we will be able to analyze the number of users the telecommunications system can host. In Section 3, we will see that the analysis of the storage capacity of a variant of the well known Hopfield model of neural networks leads to a very similar question. Section 4 will treat the moderate deviations behavior of a sum of correlated Bernoulli variables in a theoretical way and will thus provide us with the main estimate for Sections 2 and 3.

Acknowledgement: We are very grateful to two anonymous referees for their valuable comments.

2. On the performance of third generation wireless communication systems

In this section we will study the performance of third generation mobile communication systems, which are based on a technique called code division multiple access (CDMA). In these systems, users receive their own coding sequence to distinguish them from other users. The resulting systems are much more flexible than systems used earlier and bandwith can be used much more efficiently.
In mathematical terms these CDMA systems can be described as follows. Suppose that k users want to transmit data across a channel simultaneously. To this end, each user multiplies his data by an individual coding sequence. The signals are decoded at the receiver by taking the inner product with the corresponding coding sequence. Ideally, all the coding sequences of k users are orthogonal. In this case, taking the inner product with the m'th coding sequence will yield solely the information sent by the m'th user ($1 \leq m \leq k$). However, in practice almost-orthogonal codes (pseudo-random codes) are used. The technique of coding signals in order to transmit various signals simultaneously is known as code division multiple access (CDMA), see for example [45]. So, let $b_m(t)$ be the data signal of the m'th user. We let $b_m(t) = b_{m[t/T]}$, for $1 \leq m \leq k$, where

$$b_m = (\ldots, b_{m,-1}, b_{m0}, b_{m1}, \ldots) \in \{-1, +1\}^{\mathbb{Z}}.$$

Here for some real number x, $[x]$ denotes the smallest integer larger than or equal to x. In order to encode this signal, for each m, $1 \leq m \leq k$, we have a sequence

$$a_m = (\ldots, a_{m,-1}, a_{m0}, a_{m1}, \ldots) \in \{-1, +1\}^{\mathbb{Z}}.$$

We put $a_m(t) = a_{m[t/T_c]}$, where $T_c = T/n$, for some integer n. The coded signal of the m'th user is then

$$s_m(t) = \sqrt{2P_m} b_m(t) a_m(t) cos(\omega_c t), \qquad 1 \leq m \leq k, \tag{1}$$

where P_m is the power of the m'th user and ω_c the carrier frequency. The factor $\cos(\omega_c t)$ is used to transmit the signal at frequency ω_c. CDMA turns out to reduce the signal-to-noise ratio of the transmitter and the jammer by a factor n. The code $a_m(t)$ is known to the mobile phone of the transmitting person and to the base station.
The total transmitted signal is given by

$$r(t) = \sum_{j=1}^{k} s_j(t).$$

To avoid technical complications we will assume that all users transmit using the same time grid. For simplicity, we also assume that we are working on a noiseless channel, such that the transmitted and received signals agree.
To retrieve the data bit b_{m1} from the received data, the signal $r(t)$ is multiplied by $a_m(t) cos(\omega_c t)$ and then averaged over $[0, T]$. For simplicity, we pick $\omega_c T_c = \pi f_c$, where $f_c \in \mathbb{N}$ to get

$$\frac{1}{T} \int_0^T r(t) a_m(t) cos(\omega_c t) dt = \frac{1}{2} \sqrt{2P_m}\, b_{m1} + \sum_{\substack{j=1 \\ \jmath \neq m}}^{k} \frac{1}{2} \sqrt{2P_j}\, b_{j1} \frac{1}{n} \sum_{i=1}^{n} a_{ji} a_{mi} \tag{2}$$

The above procedure is is often referred to as Matched Filter (MF). As can be seen from (2) the decoded signal consists of the desired bit (amplified by $\frac{1}{2}\sqrt{2P_m}$) and interference due to the other users. If the vectors $(a_{m1}, \ldots, a_{mn})$ and $(a_{j1}, \ldots, a_{jn})$, $j = 1 \ldots, k, j \neq m$, were orthogonal, then their inner product would disappear, such that

$$\sum_{i=1}^{n} a_{ji} a_{mi} = 0.$$

However in practice, the a-sequences are often generated by a random number generator. We model this by an array $(A_i^m), m = 1, 2, \ldots k, i = 1, 2, \ldots n$ of independent and identically distributed random variables taking values in $\{-1, +1\}$ and accepting both values with equal probability:

$$\mathbb{P}(A_i^m = +1) = \mathbb{P}(A_i^m = -1) = \frac{1}{2}.$$

In this case the signal (2) is turned to

$$Z_m^{(1)} := \frac{1}{2} \sqrt{2P_m}\, b_{m1} + \sum_{\substack{j=1 \\ \jmath \neq m}}^{k} \frac{1}{2} \sqrt{2P_j}\, b_{j1} \frac{1}{n} \sum_{i=1}^{n} A_i^j A_i^m.$$

Given this setup the signal b_{m1} can be estimated by

$$\hat{b}_{m1}^{(1)} := \operatorname{sgn}\left(Z_m^{(1)}\right),$$

where $\mathrm{sgn}(\cdot)$ is the sign-function while $\mathrm{sgn}(0) = U$ where U is a random variable with

$$\mathbb{P}(U = +1) = \mathbb{P}(U = -1) = \frac{1}{2}$$

and each time one needs to decide $\mathrm{sgn}(0)$ a fresh realization of U is taken.
The MF-system is used in the third generation telecommunication systems. The super-script $^{(1)}$ indicates that potentially $\hat{b}_{m1}^{(1)}$ is only a first step and further steps may be taken. We will show how such further steps can improve the performance.
We will be interested in the probability of a bit-error, i.e., in

$$\mathbb{P}(\hat{b}_{m1}^{(1)} \neq b_{m1}) \qquad \text{and} \qquad \mathbb{P}(\exists m : \hat{b}_{m1}^{(1)} \neq b_{m1})$$

and when this probability will go to zero, depending on k, when n goes to infinity. One can analyze this probability using Gaussian approximations as in [23], [27], [31], or other approximation techniques as in [32], [7], [37].
In [22] we were able to give sharp bounds on the number of users k (as a function of n) such that the probability of a bit error $\mathbb{P}(\exists m : \hat{b}_{m1}^{(1)} \neq b_{m1})$ converges to zero. More precisely, we were able to show that for $k = c\frac{n}{\log n}$ with $c < 1/2$ the above probability converges to zero, while for $c > 1/2$ this probability converges to one.
There are various ways to exploit that interference experienced in a CDMA system is different from completely random noise, and to thus improve the performance of matched filter systems. A straightforward technique is called *interference cancellation*. The idea is that we try to cancel the interference due to the other users (i.e., the users with subscript $j \neq m$). Interference cancellation comes in various flavours. The one we will focus on in this paper is called *soft decision parallel interference cancellation* (SD-PIC) (see e.g. [19]). It is best understood when first explaining the concept of *hard decision parallel interference cancellation* (HD-PIC) (see [11], [9], [25], [30], [1] and the references therein).
To this end, note that from our estimate $\hat{b}_{m1}^{(1)}$ for b_{m1}, we obtain an estimate for the signal $s_m(t)$ sent by the m'th user, where for $t \in [0, T]$, our estimator is given by

$$\hat{s}_m^{(1)}(t) = h_m(Z_m^{(1)})\ a_m(t)cos(\omega_c t), \qquad 1 \leq m \leq k,$$

where, for general powers,

$$h_m(x) = \begin{cases} \sqrt{2P_m}\,\mathrm{sgn}(x) & \text{for HD-PIC,} \\ 2x & \text{for SD-PIC.} \end{cases}$$

Obviously, for HD-PIC, we need to know the powers of the submitting users, whereas for SD-PIC this is not necessary. The total multiple access interference experienced by the m'th user (i.e., the interference of the m'th user from all other users) can be estimated by $\hat{r}_m^{(1)}(t) = \sum_{j \neq m} \hat{s}_j^{(1)}(t)$. In the absence of noise, the multiple access interference is the only possible cause for bit-errors.
The above estimate allows to correct our estimate of $b_{m,1}$ by subtracting the estimator of the multiple access interference from the signal. We thus obtain

$$\hat{b}_{m,1}^{(2)} = \mathrm{sgn}(Z_m^{(2)}),$$

where $Z_m^{(2)}$ is obtained by replacing $r(t)$ in (2) by $r(t) - \hat{r}_m^{(1)}(t)$, so that, if we set all powers equal to 2 to simplify things, we obtain

$$Z_m^{(2)} = b_{m1} + \sum_{\substack{j=1 \\ j \neq m}}^{k} \frac{1}{n} \sum_{i=1}^{n} A_i^j A_i^m (b_{j1} - h_j(Z_j^{(1)})).$$

If this procedure is successful, then we may iterate it. After $s-1$ (note that our first estimator does not involve any interference cancellation) steps of parallel interference cancellation, we obtain $\hat{b}_{m,1}^{(s)} = \operatorname{sgn}(Z_m^{(s)})$, where

$$Z_m^{(s)} = b_{m1} + \sum_{\substack{j=1\\ j\neq m}}^{k} \frac{1}{n} \sum_{i=1}^{n} A_i^j A_i^m \big(b_{j1} - h_j(Z_j^{(s-1)})\big).$$

The question is, of course, how reliable these estimators are. In other words, we need to investigate, how $\mathbb{P}(\exists m : \hat{b}_{m,1}^{(s)} \neq b_{m,1})$ behaves. In [22] we were able to show that a large number (order $\log n$) of iterations increases the capacity to $k = k_n = \delta n$ users (for some $\delta > 0$ small enough). However, simulations (cf. [22]) show that already one step of interference cancellation can improve the capacity and the performance of the system considerably. In this respect our theoretical results in [22] were rather poor. Our moderate deviations estimates in Section 4 below now give us the opportunity to give a theoretical analysis of the SD-PIC with just one step of parallel interference cancellation. We will see that this single step improves the capacity from $k_n = \frac{n}{2\log n}$ in the matched filter system to $k_n = \frac{n}{\sqrt{2\log n}}$ for one step of parallel interference cancellation in the SD-PIC.

Theorem 2.1 (Absence of bit-errors in SD-PIC systems with one iteration)**.** *Assume that* $k_n = \frac{n}{\sqrt{\gamma \log n}}$ *for some* $\gamma > 2$*. Then*

$$\lim_{n\to\infty} \mathbb{P}(\hat{b}_{m,1}^{(2)} = b_{m,1},\, \forall m = 1, \ldots, k_n) = 1. \tag{3}$$

Remarks 2.2. (1) In [22] we proved that for $k = k_n = \delta n$ with $\delta < (\sqrt{2}-1)^2$ there exists $M = M(\delta)$ such that after $s = M \log n$ many steps the probability of having no bit-error converges to one :

$$\lim_{n\to\infty} \mathbb{P}(\hat{b}_{m,1}^{(s)} = b_{m,1},\, \forall m = 1, \ldots, k_n) = 1.$$

However, for all practical purposes, n ranges between 32 and 512, in which case our bound on k_n in Theorem 2.1 (for just one iteration) is superior to the bound $k_n = \delta n$ quoted above for many iterations.

(2) In [22] we also showed that for the matched filter system $k_n = \frac{n}{2\log n}$ is the optimal bound, if we are interested in having no bit-errors. It would be interesting to see, whether the same holds true for the bound $k_n = \frac{n}{\sqrt{2\log n}}$ in Theorem 2.1. However, the corresponding techniques (basically tilting and negative association) do not easily carry over from the matched filter to the SD-PIC situation.

(3) It would be interesting to show a similar estimate for the HD-PIC system, which in simulations even performs better than the SD-PIC system. However, here the situation is mathematically much more involved.

Proof of Theorem 2.1. By definition, $\hat{b}_{m1}^{(2)} = \operatorname{sgn}(Z_m^{(2)})$, where

$$Z_m^{(2)} = b_{m1} + \sum_{\substack{j=1\\ j\neq m}}^{k} \frac{1}{n} \sum_{i=1}^{n} A_i^j A_i^m (b_{j1} - Z_j^{(1)})$$

and

$$Z_m^{(1)} = b_{m1} + \sum_{\substack{j=1\\ j\neq m}}^{k} b_{j1} \frac{1}{n} \sum_{i=1}^{n} A_i^j A_i^m.$$

Without loss of generality, we may assume that $b_{m1} = 1$, for all $m = 1, \dots, k$, and we have

$$\begin{aligned}\mathbb{P}[\exists\ m = 1, \dots, k,\ \hat{b}^{(2)}_{m1} \neq b_{m1}] &\leq k\ \mathbb{P}[\hat{b}^{(2)}_{11} \neq b_{11}] \\ &\leq k\ \mathbb{P}[Y_n \geq 1],\end{aligned}$$

where

$$Y_n = \frac{1}{n^2} \sum_{i_1=1}^{n} \sum_{i_2=1}^{n} \sum_{j_1=2}^{k} \sum_{j_2(\neq j_1)=1}^{k} A^1_{i_1} A^{j_1}_{i_1} A^{j_1}_{i_2} A^{j_2}_{i_2}.$$

We decompose Y_n into three parts

$$Y_n = \frac{1}{n^2} \sum_{i_1=1}^{n} \sum_{\substack{i_2=1 \\ i_2 \neq i_1}}^{n} \sum_{j_1=2}^{k} \sum_{\substack{j_2=2 \\ j_2 \neq j_1}}^{k} A^1_{i_1} A^{j_1}_{i_1} A^{j_1}_{i_2} A^{j_2}_{i_2} + \frac{1}{n^2} \sum_{j_1=2}^{k} (\sum_{i_1=1}^{n} A^1_{i_1} A^{j_1}_{i_1})^2 + \frac{k-2}{n^2} \sum_{j_1=2}^{k} \sum_{i_1=1}^{n} A^1_{i_1} A^{j_1}_{i_1}.$$

From here we obtain for any $\delta > 0$

$$\begin{aligned}\mathbb{P}[\hat{b}^{(2)}_{11} \neq b_{11}] \ \leq\ & \mathbb{P}[\frac{1}{n^2} \sum_{i_1=1}^{n} \sum_{\substack{i_2=1 \\ i_2 \neq i_1}}^{n} \sum_{j_1=2}^{k} \sum_{\substack{j_2=2 \\ j_2 \neq j_1}}^{k} A^1_{i_1} A^{j_1}_{i_1} A^{j_1}_{i_2} A^{j_2}_{i_2} \geq 1 - \delta] \\ & + \mathbb{P}[\frac{1}{n^2} \sum_{j_1=2}^{k} (\sum_{i_1=1}^{n} A^1_{i_1} A^{j_1}_{i_1})^2 \geq \delta_1] + \mathbb{P}[\frac{k-2}{n^2} \sum_{j_1=2}^{k} \sum_{i_1=1}^{n} A^1_{i_1} A^{j_1}_{i_1} \geq \delta_2],\end{aligned}$$

with $\delta_1, \delta_2 \in (0,1)$ to be chosen later and $\delta = \delta_1 + \delta_2$.
Now, we denote $A^1_{i1} A^{j1}_{i1}$ by $\bar{A}^{j1}_{i1}$ for all $i_1 = 1, \dots, n$ and $j_1 = 2, \dots, k$. These new variables $\bar{A}^{j1}_{i1}$ are i.i.d. symmetric Bernoulli random variables as the initial variables A^{j1}_{i1}. Then the second summand on the right will now be estimated with the help of an exponential Chebyshev inequality

$$\begin{aligned}\mathbb{P}[\frac{1}{n^2} \sum_{j_1=2}^{k} (\sum_{i_1=1}^{n} \bar{A}^{j_1}_{i_1})^2 \geq \delta_1] \ &\leq\ e^{-t\delta_1} \mathbb{E}[\exp(\frac{t}{n^2} \sum_{j_1=2}^{k} (\sum_{i_1=1}^{n} \bar{A}^{j_1}_{i_1})^2)] \\ &=\ e^{-t\delta_1} \prod_{j_1=2}^{k} \mathbb{E}_N \mathbb{E}[\exp(\sqrt{\frac{2t}{n^2}} (\sum_{i_1=1}^{n} \bar{A}^{j_1}_{i_1}) N)],\end{aligned}$$

where N is a standard normal random variable, and $\mathbb{E}_N$ is the expectation with respect to N. The last equality simply uses that for a Gaussian random variable G with zero mean and any $b \in \mathbb{R}$

$$\mathbb{E} e^{bG} = e^{\frac{1}{2} b^2 \mathbb{E} G^2}.$$

Now, using that

$$\mathbb{E}\left[e^{\sqrt{\frac{2t}{n^2}} \bar{A}^{j_1}_{i_1} N}\right] = \cosh\left(\sqrt{\frac{2t}{n^2}} N\right) \leq e^{\frac{t}{n^2} N^2}$$

we obtain for $t \in (0, n/2)$

$$\begin{aligned}\mathbb{P}\left[\frac{1}{n^2} \sum_{j_1=2}^{k} \left(\sum_{i_1=1}^{n} \bar{A}^{j_1}_{i_1}\right)^2 \geq \delta_1\right] \ &\leq\ e^{-t\delta_1} \left(\mathbb{E}_N\left[\exp(\frac{t}{n} N^2)\right]\right)^{k-1} \\ &\leq\ \exp\left(-t\delta_1 - \frac{k}{2} \log\left(1 - \frac{2t}{n}\right)\right).\end{aligned}$$

From the Taylor expansion of the natural logarithm we see that there exists $u_0 \in (0,1)$ such that for all $u \in (0,u_0)$, we have $-\log(1-u) \leq u+u^2$. We suppose now that $2t/n < u_0$. Then

$$\mathbb{P}\left[\frac{1}{n^2}\sum_{j_1=2}^{k}(\sum_{i_1=1}^{n}\bar{A}_{i_1}^{j_1})^2 \geq \delta_1\right] \leq \exp\left(-t(\delta_1-\frac{k}{n})+2k\frac{t^2}{n^2}\right)$$

If we choose $\delta_1 = \lambda\,\frac{k}{n},\ t=(\delta_1-\frac{k}{n})\frac{n^2}{4k}$ for some $\lambda>1$ such that $2t/n=(\lambda-1)/2<u_0$, then we finally get

$$\begin{aligned}\mathbb{P}\left[\frac{1}{n^2}\sum_{j_1=2}^{k}(\sum_{i_1=1}^{n}\bar{A}_{i_1}^{j_1})^2 \geq \delta_1\right] &\leq \exp\left(-\frac{(\delta_1-\frac{k}{n})^2}{8}\frac{n^2}{k}\right)\\ &\leq \exp\left(-\frac{(\lambda-1)^2}{8}k\right) \qquad (4)\end{aligned}$$

For the third term, we also obtain by the exponential Chebyshev inequality

$$\mathbb{P}\left[\frac{k-2}{n^2}\sum_{j_1=2}^{k}\sum_{i_1=1}^{n}\bar{A}_{i_1}^{j_1} \geq \delta_2\right] \leq \exp\left(-\frac{{\delta_2}^2 n^3}{2k^3}(1+r_n)\right), \qquad (5)$$

with $r_n \to 0$ as $n \to +\infty$.
For the first term, we can apply Theorem 4.1 in Section 4 below to obtain

$$\mathbb{P}\Big[\frac{1}{n^2}\sum_{i_1=1}^{n}\sum_{\substack{i_2=1\\ i_2\neq i_1}}^{n}\sum_{j_1=2}^{k}\sum_{\substack{j_2=2\\ j_2\neq j_1}}^{k}\bar{A}_{i_1}^{j_1}\bar{A}_{i_2}^{j_1}\bar{A}_{i_2}^{j_2} \geq 1-\delta\Big] \leq 2\exp\Big(-\frac{n^2}{2k^2}(1-\delta)^2\Big). \qquad (6)$$

Finally, putting the above estimates together, we see that the contribution stemming from the second summand is of order $e^{-\text{const.}k}$, while the one stemming from the first summand is of order $e^{-\text{const.}\frac{n^2}{k^2}}$ (possibly with different constants, of course), such that the contribution of the second summand becomes negligible as soon as

$$e^{-k} \ll e^{-n^2/k^2}$$

which is the case as soon as $k \geq n^\beta$ with $\beta > 2/3$. Choosing $k = c\frac{n}{\sqrt{\log n}}$, $\delta_2 = \sqrt{\frac{k}{n}}$, we obtain that

$$\begin{aligned}\mathbb{P}\big[\exists\ m=1,\dots,k,\ \hat{b}_{m1}^{(2)} \neq b_{m1}\big] &\leq k\exp\Big(-\frac{n^2}{2k^2}+\mathcal{O}((\log n)^{3/4})\Big)\\ &\leq k\exp\Big(-\frac{1}{2c^2}\log n+\mathcal{O}((\log n)^{3/4})\Big)\end{aligned}$$

which goes to 0 as $n \to +\infty$ if $c < \frac{1}{\sqrt{2}}$.

□

3. Capacity of the Hopfield model with a nonmonotonic dynamics

The Hopfield model is one of the best-studied models of a neural network. It was originally introduced by Pastur and Figotin [18] as a so-called frustrated spin system but it received most of its attention by its reinterpretation by Hopfield [24] as a very simple model for the associative memory in the brain (for reviews on mathematical work on this model, see [43], [44], [6], [5], [4] and all the references given there). The Hopfield model is based on a set of N formal neurons interconnected one to one. The state of the i-th unit is given

by a spin variable σ_i with values in $S = \{-1, +1\}$, and the state of the network by the configuration $\sigma = (\sigma_i)_{i=1,\dots,N} \in S^N$. This network is designed to memorize M patterns $\xi^\mu = (\xi_i^\mu)_{i=1,\dots,N}, \mu = 1, \dots, M$. Note that $M = M(N)$ may and actually will depend on N. First, we define a discrete time dynamical system on S^N: given a configuration $\sigma(t)$ at time t, the configuration of the network at time $t+1$ is

$$\sigma_i(t+1) = \operatorname{sgn}\left(\sum_{j=1}^N W_{ij}\sigma_j(t)\right),\ i = 1, \dots, N,$$

where the synaptic weights W_{ij} are defined according to the Hebb rule:

$$W_{ij} = \frac{1}{N}\sum_{\mu=1}^M \xi_i^\mu \xi_j^\mu, \qquad \text{for } i, j \in \{1, \dots, N\}. \tag{7}$$

The dynamics can be parallel, i.e., the updating is done simultaneously, or sequential if we update all σ_i one after the other. Under some conditions, basically on the number of stored patterns, starting from a configuration "close" to one of these patterns ξ^μ, the dynamics will converge to ξ^μ. Naturally, this is only possible if the patterns ξ^μ are stable points for the dynamics, and if we consider a starting point in a basin of attraction of an original pattern. So, the first and most basic question for this model is whether the patterns are fixed points of the dynamics: this question was first considered for i.i.d. symmetric Bernoulli patterns by McEliece *et al.* [36]. We will adopt their definition of storage capacity as the maximum number of patterns M that are stable under the dynamics described above, with a probability converging to one as $N \to +\infty$.
It appears that in the standard Hopfield model (where the patterns are i.i.d. and take values ± 1 with equal probabilities) for this notion of capacity, where no error is tolerated, and for this dynamics, the maximum number of stored patterns is $N/2\log N$. More precisely, if $M = cN/\log N$ for some $c < 1/2$ then any original pattern ξ^μ is stable with a probability converging to 1 as $N \to +\infty$ ([43], [47]), while if $M = cN/\log N$ for some $c > 1/2$, any original pattern is unstable ([3]). This notion of capacity has also been analyzed by Löwe and Vermet in more general situations: different types of dependent patterns ([33],[34]), patterns with $q \geq 2$ possible states for each pixel ([35]).
Various modifications of the original model have been proposed to increase the memory capacity. One interesting idea is to consider nonmonotonic neurons (see [40], [28], [42], [39], [41], [12]). Among these models, the so-called "partial reverse method" is worth to be mentioned. This method was proposed by Morita *et al* [40], [39]. There, the authors define a two-stage dynamics to realize the nonmonotonic behavior of a single neuron. This technique was generalized by Amari and Yanai [2]. For their model the authors obtain a memory capacity of order $N/\sqrt{\log N}$, larger than $N/\log N$ of the classical Hopfield model. However, their analysis relies on a non-rigorous approximation by Gaussian random variables, while a rigorous mathematical treatment of their improvement of the memory capacity is missing.
We will now define the dynamics proposed by Amari and Yanai, and we will see that the mathematical analysis of the stability of the original patterns is very close to the problem that we considered in the previous section for the SD-PIC model.

We consider the application $T = T_N \circ \dots \circ T_1 : \{-1, +1\}^N \to \{-1, +1\}^N$, where

$$T_i(\sigma) = \operatorname{sgn}\left(\sum_{j=1}^N (W - aW^2)_{ij}\ \sigma_j\right),\ i = 1, \dots, N.$$

Here $W = (W_{ij})_{\substack{i=1,\dots N,\\ j=1,\dots N}}$ is the matrix of the synaptic weights W_{ij} defined according to the Hebb rule (7), and $a \in (0,1)$ is a parameter. Amari and Yanai set $W_{ii} = 0$ for all $i = 1, \dots, N$, but this minor difference becomes negligible for N large and does not change the capacity. We will prove the following result, that was already announced in [2].

Theorem 3.1. *Under the hypotheses that the $(\xi_i^\mu)_{i,\mu}$ are i.i.d. random variables with*

$$\mathbb{P}(\xi_i^\mu = +1) = \mathbb{P}(\xi_i^\mu = -1) = \frac{1}{2}$$

the following holds true:

(i) If $a = \frac{1}{2}$ and M is such that $M = c\dfrac{N}{\sqrt{2\log N}}$, for some $c \in (0,1)$, then for any $\mu = 1, \dots, M$,

$$\lim_{N\to+\infty} \mathbb{P}[T(\xi^\mu) = \xi^\mu] = 1.$$

(ii) If $a \in (0,1)\ ) \ \setminus \{\frac{1}{2}\}$ and M is such that $M = c\dfrac{N}{\log N}$, for some $c \in (0, \frac{(1-a)^2}{2(1-2a)^2})$, then for any $\mu = 1, \dots, M$,

$$\lim_{N\to+\infty} \mathbb{P}[T(\xi^\mu) = \xi^\mu] = 1.$$

Remarks 3.2. (1) The dynamics proposed by Amari and Yanai is in fact related to the pseudo-inverse learning rule, also called projection learning rule ([14], [15], [26]). The idea of this learning rule is to search a matrix C which guarantees the stability of all original patterns through the strong conditions $C\xi^\mu = \xi^\mu$, for all $\mu = 1, \dots, M$. This can be written equivalently as $C\xi = \xi$, where ξ is the matrix $N \times M$ with columns $\xi^\mu = (\xi_i^\mu)_{i=1,\dots,N}, \mu = 1, \dots, M$. A non-trivial solution is the orthogonal projection matrix into the subspace spanned by the vectors $\xi^\mu, \mu = 1, \dots, M$, which can be written as $C = \xi\xi^+$, where ξ^+ is the Moore-Penrose pseudoinverse of ξ. Some papers, based on not totally rigourous techniques and simulations, indicate that this rule allows a higher capacity than the classical Hebbian learning rule (see [14], [15], [26]). But theoretically, this rule involves calculating the inverse of a $M \times M$ matrix to get the pseudoinverse matrix. To avoid matrix inversion, it is possible to consider iterative algorithms. For instance, a Neumann-type expansion for ξ^+ is a series representation of ξ^+ analogous to the Neumann expansion for the inverse of a non-singular matrix (see e.g. [8], [38]) :

$$\xi^+ = \alpha \sum_{k=0}^{\infty} (I - \alpha\xi^T\xi)^k \ \xi^T,$$

for $\alpha \in]0, \min_{d_{ii}\neq 0} 2/|d_{ii}|^2[$, where $D = (d_{ij})$ is the diagonal matrix in the singular value decomposition of ξ and ξ^T is the transpose of ξ. Truncating the series to the second term, we get

$$C = \xi\xi^+ \simeq C^{(2)} := 2\alpha(\tilde{W} - \frac{\alpha}{2}\tilde{W}^2),$$

where $\tilde{W} = N\ W$ and the dynamics

$$T_i(\sigma) = \mathrm{sgn}(\sum_{j=1}^{N} C_{ij}^{(2)}\sigma_j), i = 1, \dots, N$$

coincides with the dynamics proposed by Amari and Yanai.

(2) In the case $a \in (0,1)\ ) \setminus \{\frac{1}{2}\}$, the capacity is of the same order as for the classical Hopfield model ($a = 0$), with a larger constant for $a \in (0, 2/3)$.

(3) In the same paper, Amari and Yanai consider also the following dynamics :

$$T_i(\sigma) = \operatorname{sgn}\left(\sum_{j=1}^N W_{ij}\sigma_j + \sum_{j=1}^N W_{ij} f(\sum_{k=1}^N W_{jk}\sigma_k)\right), i = 1, \dots, N$$

where $f(u) = -au - (1-2a)\operatorname{sgn}(u)$, for $u \in \mathbb{R}, a > 0$. They assert that the capacity is $\frac{N}{\sqrt{2\log N}}$, independently of a. However, the situation is mathematically much more complicated, resembles in a way the HD-PIC systems from the previous sections and cannot be analyzed with the help of Theorem 4.1 in Section 4.

Proof. Without loss of generality, we may consider $\mu = 1$. We can also replace everywhere $\xi_i^1\xi_i^\mu$ by ξ_i^μ for all $i = 1, \dots, N$ and $\mu = 2, \dots, M$ for the same reasons as in the previous section. Then we have

$$\begin{aligned}\mathbb{P}[T(\xi^1) \neq \xi^1] &\leq N\ \mathbb{P}[T_1(\xi^1) \neq \xi_1^1] \\ &\leq N\ \mathbb{P}[\sum_{j=1}^N W_{1j} - a\sum_{j=1}^N W_{1j}\sum_{k=1}^N W_{jk} \leq 0].\end{aligned}$$

This last event can be rewritten as

$$A_N + B_N \sum_{k=2}^N \sum_{\mu=2}^M \xi_1^\mu \xi_k^\mu - \frac{a}{N^2}\sum_{\mu=2}^M (\sum_{j=1}^N \xi_j^\mu)^2 - \frac{a}{N^2}\sum_{j=2}^N \sum_{\mu_1=2}^M \sum_{\substack{k=2\\k\neq j}}^N \sum_{\substack{\mu_2=2\\\mu_2\neq\mu_1}}^M \xi_1^{\mu_1}\xi_j^{\mu_1}\xi_j^{\mu_2}\xi_k^{\mu_2} \leq 0,$$

with

$$A_N = (1-a)(1 + \frac{M-1}{N}) - a\frac{(M-1)^2}{N^2}$$

and

$$B_N = \frac{1-2a}{N} - 3a\frac{M-1}{N^2} + \frac{2a}{N^2}.$$

With this notation we obtain:

$$\begin{aligned}\mathbb{P}[T_1(\xi^1) \neq \xi_1^1] \leq\ & \mathbb{P}\Big[\frac{a}{N^2}\sum_{j=2}^N \sum_{\mu_1=2}^M \sum_{\substack{k=2\\k\neq j}}^N \sum_{\substack{\mu_2=2\\\mu_2\neq\mu_1}}^M \xi_1^{\mu_1}\xi_j^{\mu_1}\xi_j^{\mu_2}\xi_k^{\mu_2} \geq A_N - \delta\Big] \\ &+ \mathbb{P}\Big[\frac{a}{N^2}\sum_{\mu=2}^M (\sum_{j=1}^N \xi_j^\mu)^2 \geq \delta_1\Big] + \mathbb{P}\Big[-B_N \sum_{k=2}^N \sum_{\mu=2}^M \xi_1^\mu \xi_k^\mu \geq \delta_2\Big],\end{aligned}$$

with δ_1 and δ_2 strictly positive to be chosen later on and $\delta = \delta_1 + \delta_2$ such that $A_N - \delta > 0$. Note that, no matter whether $M = cN/\log N$ or $M = cN/\sqrt{\log N}$, M satisfies the conditions on k in Theorem 4.1, if we set $n = N$.
For the first term, we have

$$\mathbb{P}[\frac{a}{N^2}\sum_{j=2}^N \sum_{\mu_1=2}^M \sum_{\substack{k=2\\k\neq j}}^N \sum_{\substack{\mu_2=2\\\mu_2\neq\mu_1}}^M \xi_1^{\mu_1}\xi_j^{\mu_1}\xi_j^{\mu_2}\xi_k^{\mu_2} \geq A_N - \delta] = \mathbb{P}[\frac{a}{N^2}\sum_{j=2}^N \sum_{\mu_1=2}^M \sum_{\substack{k=2\\k\neq j}}^N \sum_{\substack{\mu_2=2\\\mu_2\neq\mu_1}}^M \xi_j^{\mu_1}\xi_j^{\mu_2}\xi_k^{\mu_2} \geq A_N - \delta],$$

and we apply Theorem 4.1 in the next Section to get

$$\mathbb{P}[\frac{a}{N^2}\sum_{j=2}^{N}\sum_{\mu_1=2}^{M}\sum_{\substack{k=2\\k\neq j}}^{N}\sum_{\substack{\mu_2=2\\\mu_2\neq\mu_1}}^{M}\xi_j^{\mu_1}\xi_j^{\mu_2}\xi_k^{\mu_2}\geq A_N-\delta]\leq\exp(-\frac{N^2}{2M^2}(\frac{A_N-\delta}{a})^2(1+\varepsilon_N)),$$

where $\varepsilon_N\to 0$ as $N\to+\infty$.
For the second term, we can prove as in the second part that we can find $\delta_1=\mathcal{O}(\frac{M}{N})$ and a constant $c_1>0$ such that

$$\mathbb{P}[\frac{a}{N^2}\sum_{\mu=2}^{M}(\sum_{j=1}^{N}\xi_j^{\mu})^2\geq\delta_1]\leq\exp(-c_1M).$$

For the third term, the exponential Markov inequality together with

$$\mathbb{E}\left[e^{t\xi_i^{\mu}}\right]=\cosh(t)\leq e^{t^2/2}$$

gives

$$\mathbb{P}[-B_N\sum_{k=2}^{N}\sum_{\mu=2}^{M}\xi_1^{\mu}\xi_k^{\mu}\geq\delta_2]\leq\exp(-\frac{{\delta_2}^2}{2{B_N}^2(M-1)(N-1)})$$

The important fact is that for $a=1/2$, B_N is of order M/N^2, while for $a\neq 1/2$, B_N is of order $1/N$. We will now treat the two cases separately.

• If $a=1/2$, we get from the preceding estimates

$$\mathbb{P}[T_1(\xi^1)\neq\xi_1^1] \;\leq\; e^{-\frac{N^2}{2M^2}((1-2\delta)^2+\mathcal{O}(\frac{M}{N}))(1+\varepsilon_N)}+e^{-c_1M}+e^{-\frac{2}{9}{\delta_2}^2\frac{N^3}{M^3}(1+\varepsilon'_N)}$$

If we choose $M=c\frac{N}{\sqrt{2\log N}},\delta_2=(\log N)^{-1/8}$, then we get

$$\mathbb{P}[T(\xi^1)\neq\xi^1] \;\leq\; Ne^{-\frac{1}{c}\log N(1+r_N)}+Ne^{-c_1M}+Ne^{-c_2(\log N)^{5/4}},$$

with $c_1,c_2>0$ and r_N such that $r_N\to 0$ as $N\to+\infty$. This implies that $P[T(\xi^1)\neq\xi^1]\to 0$ as $N\to+\infty$ if $c<1$.

• If $a\in(0,1\ )\ \setminus\{\frac{1}{2}\}$, we get from the preceding estimations

$$\mathbb{P}[T_1(\xi^1)\neq\xi_1^1] \;\leq\; e^{-\frac{N^2}{2M^2}(\frac{1-a-\delta_2}{a}+\mathcal{O}(\frac{M}{N}))^2(1+\varepsilon_N)}+e^{-c_1M}+e^{-\frac{1}{2(1-2a)^2}{\delta_2}^2\frac{N}{M}(1+\varepsilon'_N)} \qquad (8)$$

The third summand is the dominating term in (8). Therefore

$$\mathbb{P}[T(\xi^1)\neq\xi^1] \;\leq\; Ne^{-\frac{1}{2(1-2a)^2}{\delta_2}^2\frac{N}{M}(1+r'_N)},$$

with r'_N such that $r'_N\to 0$ as $N\to+\infty$. This implies that

$$\mathbb{P}[T(\xi^1)\neq\xi^1]\to 0 \quad \text{ as } N\to+\infty,$$

if $M=c_3N/\log N$ and $c_3<\frac{{\delta_2}^2}{2(1-2a)^2}$. Taking into account the constraint $\delta_2<1-a+\mathcal{O}(\frac{M}{N})$, we can choose any $c_3<\frac{(1-a)^2}{2(1-2a)^2}$ for N large enough. □

4. A useful moderate deviations result

In order to present a unifying framework for both applications in Sections 2 and 3, let us introduce an array of i.i.d. binary random variables $(A_i^j), i = 1, \dots, n, j = 1, \dots k$ with

$$\mathbb{P}(A_i^j = 1) = \mathbb{P}(A_i^j = -1) = \frac{1}{2} \qquad \text{for all } i, j.$$

The quantity of interest is

$$\Xi_n := \sum_{i_1=1}^{n} \sum_{\substack{i_2=1 \\ i_2 \neq i_1}}^{n} \sum_{j_1=1}^{k} \sum_{\substack{j_2=1 \\ j_2 \neq j_1}}^{k} A_{i_1}^{j_1} A_{i_2}^{j_1} A_{i_2}^{j_2}.$$

Note that

$$\mathbb{E}[\Xi_n] = 0 \qquad \text{for all } n.$$

We analyze the moderate deviations of Ξ_n on the scale $\frac{1}{n^2}$, i.e., we consider

$$X_n := \frac{1}{n^2} \Xi_n.$$

Note that the scale for a Law of Large Numbers and hence for a large deviation principle is $\frac{1}{n^2 k^2}$, while a Central Limit Theorem for Ξ_n holds on a scale $\frac{1}{nk}$ (cf. [13]). Hence we are genuinely in a moderate deviations regime as long as $k = o(n)$. Presumably, the techniques presented here will also be sufficient to analyze the moderate deviations of Ξ_n on other scales, but this is not needed for the applications presented above. In fact, we shall prove the following result.

Theorem 4.1. *For X_n defined as above and $k \geq n^\beta$ for some $\frac{14}{15} < \beta < 1$, but $k = o(n)$, and all $\gamma > 0$, the following moderate deviations principle holds true*

$$\lim_{n \to \infty} \frac{k^2}{n^2} \log \mathbb{P}[X_n \geq \gamma] = -\frac{\gamma^2}{2}$$

In particular, we obtain that for n large enough,

$$\mathbb{P}[X_n \geq \gamma] \leq \ 2 \ e^{-\frac{\gamma^2}{2} \frac{n^2}{k^2}} \tag{9}$$

Remarks 4.2.

- Note that the moderate deviations principle in Theorem 4.1 is the same as for a sum $n^2 k^2$ independent random variables scaled by a factor $1/n^2$. In this respect, the correlation of the summands in Ξ_n is relatively weak.
- For k too small the estimate in (9) in Theorem 4.1 is certainly wrong. For example, it was shown by van der Hofstad and Klok [20] that for $k = 2$ the speed of convergence is $\mathcal{O}(e^{-cn})$ for some constant $c > 0$ rather than $\mathcal{O}(e^{-cn^2})$. This could also be expected, since for $k = 2$, Ξ_n is close to a U-statistics and we are in a large deviations regime. Is is well known that for U-statistics the speed of convergence in the large deviations principle is the number of free variables, hence n rather than n^2 (see e.g. [16]). An interesting open question would therefore be to find an optimal bound on β that guarantees that the estimates of Theorem 4.1 and Lemma 4.3 below remain true. In the context of this paper we did not bother much about such an estimate, since in our applications k are always of the order $k = \frac{\kappa n}{\sqrt{\log n}}$ for some $\kappa > 0$, and then the estimates of Theorem 4.1 are certainly true.

The proof of Theorem 4.1 relies on the following lemma.

Lemma 4.3. *For $p = \mathcal{O}(\frac{n^2}{k^2})$ an even integer and $k \geq n^\beta$ for some $\beta > \frac{14}{15}$, but $k = o(n)$, it holds*

$$\mathbb{E}[(\frac{1}{n^2}\sum_{i_1=1}^{n}\sum_{\substack{i_2=1\\ i_2\neq i_1}}^{n}\sum_{j_1=1}^{k}\sum_{\substack{j_2=1\\ j_2\neq j_1}}^{k} A_{i_1}^{j_1}A_{i_2}^{j_1}A_{i_2}^{j_2})^p] = \frac{p!}{2^{p/2}(p/2)!}\frac{k^p}{n^p}(1+\varepsilon_{n,p}), \tag{10}$$

where $\lim_{n\to\infty}\varepsilon_{n,p} = 0$.

Proof. The sum

$$\Xi_n = \sum_{i_1=1}^{n}\sum_{\substack{i_2=1\\ i_2\neq i_1}}^{n}\sum_{j_1=1}^{k}\sum_{\substack{j_2=1\\ j_2\neq j_1}}^{k} A_{i_1}^{j_1}A_{i_2}^{j_1}A_{i_2}^{j_2}$$

can be written as

$$\Xi_n = \sum_{a\in\mathcal{A}} Y_a,$$

where

$$\mathcal{A} = \{a = (i_1, i_2, j_1, j_2) \in \{1,\dots,n\}\times\{1,\dots,n\}\times\{1,\dots,k\}\times\{1,\dots,k\}, i_1 \neq i_2, j_1 \neq j_2\}$$

and

$$Y_a = A_{i_1}^{j_1}A_{i_2}^{j_1}A_{i_2}^{j_2}$$

(these $Y's$ should, of course, not be confused with those occurring in the introduction). Then

$$(\Xi_n)^p = \sum_{a_1,\cdots,a_p\in\mathcal{A}} Y_{a_1}\dots Y_{a_p}.$$

Now suppose that p is even. Then

$$\mathbb{E}[Y_{a_1}\cdots Y_{a_p}] = 1,$$

if and only if all the A_i^j appear in the product with an even exponent, and

$$\mathbb{E}[Y_{a_1}\cdots Y_{a_p}] = 0$$

in all other cases.
We will show that the dominating term of $\mathbb{E}[(\Xi_n)^p]$ (in n and k) corresponds to the cases where the indices $a_1,\dots,a_p$ are themselves paired. As there are $\frac{p!}{2^{p/2}(p/2)!}$ many pairings of $1,\dots,p$ and each of the a_i's can be taken from $\mathcal{A}$, a set of size (slightly less than) n^2k^2, we have that the total contribution of terms stemming from paired indices $a_1,\dots,a_p$ is (less than)

$$\frac{p!}{2^{p/2}(p/2)!}k^p n^p.$$

However, also other terms add to this sum, which is to say: the A_i^j can occur in pairs, even though the Y_a are not paired. One simple example for $p = 4$ is the following collection of $a_i's$:

$$a_1 = (1,2,1,2), a_2 = (1,2,1,3), a_3 = (1,2,4,2), a_4 = (1,2,4,3).$$

However, for such cases there are additional constraints on the i and j, and it will turn out that their contribution is of lower order.
For $Y_a = A_{i_1}^{j_1}A_{i_2}^{j_1}A_{i_2}^{j_2}$ and $Y_{a'} = A_{i_1'}^{j_1'}A_{i_2'}^{j_1'}A_{i_2'}^{j_2'}$, we distinguish 3 types of partial or total pairings between these two terms :

(1) $i_1 = i_1', i_2 = i_2', j_1 = j_1'$ and $j_2 = j_2'$. This is the case of a total pairing, i.e. Y_a and $Y_{a'}$ coincide.

(2) The second type consists of partial pairings where two variables A_i^j of Y_a are paired respectively with two variables of $Y_{a'}$. This can occur under one of the following constraints :
- $i_1 = i'_1, i_2 = i'_2, j_1 = j'_1$,
- $i_1 = i'_2, i_2 = i'_1, j_1 = j'_1$,
- $i_1 = i'_2, i_2 = i'_1, j_1 = j'_2$ and $j_2 = j'_1$,
- $i_2 = i'_2, j_1 = j'_1$ and $j_2 = j'_2$,
- $i_2 = i'_2, j_1 = j'_2$ and $j_2 = j'_1$.

(3) The third type consists of partial pairings where a single variable A_i^j in the product constituting Y_a is paired with one variable of $Y_{a'}$. This can occur only under a constraint of the form: $i_e = i'_f$, and $j_g = j'_h$, where $e, f, g, h \in \{1, 2\}$. More precisely, the nine possible cases which produce such a pairing are :
- $i_1 = i'_1, j_1 = j'_1$,
- $i_1 = i'_2, j_1 = j'_1$,
- $i_1 = i'_2, j_1 = j'_2$,
- $i_2 = i'_1, j_1 = j'_1$,
- $i_2 = i'_2, j_1 = j'_1$,
- $i_2 = i'_2, j_1 = j'_2$,
- $i_2 = i'_1, j_2 = j'_1$,
- $i_2 = i'_2, j_2 = j'_1$,
- $i_2 = i'_2, j_2 = j'_2$.

Now consider $(\Xi_n)^p$. As Ξ_n consists of two sums with indices in $\{1, \dots, n\}$ and two sums with indices in $\{1, \dots, k\}$ the $p'th$ power $(\Xi_n)^p$ consists of $2p$ sums with indices varying from 1 to n and $2p$ sums with indices varying from 1 to k. In order to give a non-zero contribution to the expectation these sums have to be at least partially overlapping. But, if the variables Y_a are partially or totally paired, this reduces the number of such sums. More precisely, we have

$$\mathbb{E}[(\Xi_n)^p] \leq \sum_{\substack{p_1,p_2,p_3 \geq 0, \\ 3p_1+2p_2+p_3=\frac{3}{2}p}} N(p_1, p_2, p_3)\ n^{2p-2p_1-p_2-p_3}\ k^{2p-2p_1-2p_2-p_3}, \tag{11}$$

where p_i is the number of pairings of type i ($i = 1, 2, 3$) and $N(p_1, p_2, p_3)$ is the number of ways to pair the $3p$ variables A_i^j such that there are p_1 pairings of the Y_a's of type 1, p_2 pairings of the Y_a's of type 2 and p_3 pairings of the Y_a's of type 3. Once p_1, p_2, p_3 are fixed, the number of constraints implying equalities between different indices i_l and different indices j_l is known. Then, if we have bounds $u(p_1, p_2, p_3)$ and $v(p_1, p_2, p_3)$ for the number of free sums from 1 to n and the number of free sums from 1 to k respectively, $N(p_1, p_2, p_3)\ n^{u(p_1,p_2,p_3)}\ k^{v(p_1,p_2,p_3)}$ will bound the number of terms of type "(p_1, p_2, p_3)" appearing in $\mathbb{E}[(\Xi_n)^p]$.
In fact, we can prove that $n^{u(p_1,p_2,p_3)}\ k^{v(p_1,p_2,p_3)} \leq n^{2p-2p_1-p_2-p_3}\ k^{2p-2p_1-2p_2-p_3}$. Indeed, there are $3p$ random variables A_i^j in each summand of $(\Xi_n)^p$. These variables have to be organized in such a way that each A_i^j occurs an even number of times. These pairings of the A_i^j can be represented as $3p/2$ "edges" in a graph with vertices (i, j). By definition of the different types,

- For each pairing of type 1, we have 3 edges – we connect (i_1, j_1) to (i'_1, j'_1), (i_2, j_1) to (i'_2, j'_1), and (i_2, j_2) to (i'_2, j'_2).
- For a pairing of type 2, there are 2 edges (for instance, in the first case, (i_1, j_1) is connected to (i'_1, j'_1) and (i_2, j_1) to (i'_2, j'_1)).

- For a pairing of type 3, there is one edge.

Since in total there are $\frac{3}{2}p$ pairings, we obtain

$$3p_1 + 2p_2 + p_3 = 3/2p.$$

Moreover, each pairing eliminates one of the sums : for a pairing of type 1, the condition $i_1' = i_1$ eliminates the summation over i_1' (the index i_1 runs from 1 to n, but i_1' then is fixed). Similarly, in the same case, the sums over i_2', j_1' and j_2' can be eliminated.
For a pairing of type 2, at least three sums can be eliminated :
(a) one sum over an index "i" and two sums over indices "j",
or
(b) two sums over indices "i" and one sum over an index "j",
or
(c) two sums over indices "i" and two sums over indices "j".
Let $p_{2,l}$ for $l \in \{a, b, c\}$ the number of pairings of the three different cases (a), (b), (c), respectively. Trivially $p_2 = p_{2,a} + p_{2,b} + p_{2,c}$.
For a pairing of type 3, one sum over an index "i" and one sum over an index "j" disappear.
Following these lines, the number of summations over indices "i" (running from 1 to n), which is a priori $2p$ in $(\Xi_n)^p$ becomes $2p - 2p_1 - p_{2,a} - 2p_{2,b} - 2p_{2,c} - p_3$. In the same way, the number of summations over indices "j" (running from 1 to k) is reduced from $2p$ to $2p - 2p_1 - 2p_{2,a} - p_{2,b} - 2p_{2,c} - p_3$. Since $k \leq n$, we have that

$$n^{-2p_{2,b}} k^{-p_{2,b}} \leq n^{-p_{2,b}} k^{-2p_{2,b}}$$

which explains the bound $n^{2p-2p_1-p_2-p_3} k^{2p-2p_1-2p_2-p_3}$ in (11).
Now one checks that the dominating term (in n and k) in (11) is obtained by setting $p_1 = \frac{p}{2}$ and $p_2 = p_3 = 0$. One then obtains terms of order $\mathcal{O}(n^p k^p)$ (and this is the only way to obtain terms of that size). We can therefore estimate the expectation of $(\Xi_n)^p$ by

$$\mathbb{E}[(\Xi_n)^p] \leq \frac{p!}{2^{p/2}(p/2)!} k^p n^p (1 + \varepsilon_{n,p}),$$

with

$$\begin{aligned}
\varepsilon_{n,p} &= \sum_{\substack{p_1,p_2,p_3 \geq 0, p_1 \leq \frac{p}{2}-2, \\ 3p_1+2p_2+p_3=\frac{3}{2}p}} \frac{2^{p/2}(p/2)!}{p!} N(p_1, p_2, p_3)\ n^{p-2p_1-p_2-p_3}\ k^{p-2p_1-2p_2-p_3} \\
&\leq \sum_{p_1=0}^{p/2-2} \binom{p}{2p_1} \frac{(2p_1)!}{2^{p_1} p_1!} \frac{2^{p/2}(p/2)!}{p!} \sum_{\substack{p_2,p_3 \geq 0, \\ 3p_1+2p_2+p_3=\frac{3}{2}p}} N'(p_2, p_3)\ n^{-p/2+p_1+p_2}\ k^{-p/2+p_1}
\end{aligned}$$

In the latter inequality, $\binom{p}{2p_1}$ counts the number of possible choices of $2p_1$ indices amongst p and we use the fact that there are $\frac{(2p_1)!}{2^{p_1} p_1!}$ many pairings of $1, \ldots, 2p_1$. Moreover, we exploit

$$p_3 = \frac{3}{2}p - 3p_1 - 2p_2$$

to simplify the exponents of n and k. Now, if there are p_1 pairings of type 1, the remaining $q := p - 2p_1$ factors Y_a have to be such that all A_i^j composing their product are paired.

The factor $N'(p_2,p_3)$ is the number of ways to pair these $3q$ variables A_i^j with p_2 pairings of the Y_a's of type 2 and p_3 pairings of type 3. A trivial bound for $N'(p_2,p_3)$ is

$$N'(p_2,p_3) \leq (3q)^{3q}.$$

Moreover $p_2 \leq q/2$, which implies

$$n^{-p/2+p_1+p_2}\ k^{-p/2+p_1} \leq k^{-q/2}.$$

Putting things together and taking into account that there are at most p^3 possible choices for p_1,p_2,p_3, this means

$$\varepsilon_{n,p} \ \leq \ p^3 \max_{p_1=0,\dots,p/2-2} 2^{p/2-p_1} \frac{(p/2)!}{(p-2p_1)!p_1!} \left(\frac{(p-2p_1)^3}{\sqrt{k}} \right)^{p-2p_1}.$$

Estimating

$$\frac{(p/2)!}{(p-2p_1)!p_1!} \leq \frac{(p/2)!}{p_1!} \leq (\frac{p}{2})^{p/2-p1},$$

implies

$$\begin{aligned} \varepsilon_{n,p} \ &\leq \ p^3 \max_{p_1=0,\dots,p/2-2} (\sqrt{p}\frac{(p-2p_1)^3}{\sqrt{k}})^{p-2p_1} \\ &= \ p^3 \max_{u=4,6,8\dots,p} (u^3\sqrt{\frac{p}{k}})^u. \end{aligned}$$

Now, the function $g : u \mapsto \exp(u\log(u^3\sqrt{\frac{p}{k}}))$ is strictly decreasing on the interval $[4, \frac{1}{e}(\frac{k}{p})^{1/6}]$. If $k \geq n^\beta$ for some β in $(\frac{14}{15},1)$ and $p = \mathcal{O}(\frac{n^2}{k^2})$, then $\frac{1}{e}(\frac{k}{p})^{1/6} > p$ for n large enough. Therefore,

$$\max_{u=4,6,8\dots,p} (u^3\sqrt{\frac{p}{k}})^u = g(4) = 4^{12}\ \frac{p^2}{k^2},$$

and

$$\varepsilon_{n,p} \leq 4^{12}\ \frac{p^5}{k^2}.$$

Therefore, $\varepsilon_{n,p} \to 0$ as n goes to infinity, for $p = \mathcal{O}(\frac{n^2}{k^2})$ and $k \geq n^\beta$ for some β in $(\frac{14}{15},1)$. □

Now we are ready to prove Theorem 4.1.

Proof of Theorem 4.1. For the upper bound, we apply a Chebyshev–Markov inequality (with the increasing function $x \mapsto x^p$ for an even p) together with a technique introduced by Rio [46, page 44]. To this end recall that $p = \mathcal{O}(\frac{n^2}{k^2})$ and $k \geq n^\beta$ for some $\frac{14}{15} < \beta < 1$, but $k = o(n)$, such that we can apply the previous lemma. The announced inequality yields for each $\gamma > 0$

$$\begin{aligned} \mathbb{P}[X_n \geq \gamma] \ &\leq \ \gamma^{-2p}\mathbb{E}[(X_n)^{2p}] \\ &\leq \ \gamma^{-2p}\frac{(2p)!}{2^p p!}\left(\frac{k}{n}\right)^{2p}(1+\varepsilon_{n,2p}) \end{aligned} \tag{12}$$

where the first bound follows from a Chebyshev type inequality, while the second uses Lemma 4.3. Now dividing the Stirling approximation for $(2p)!$ by the Stirling approximation for $p!$ we see that for large enough p:

$$\frac{(2p)!}{2^p p!} \approx \frac{\left(\frac{2p}{e}\right)^{2p}\sqrt{2p}}{2^p\left(\frac{p}{e}\right)^p\sqrt{p}} = \sqrt{2}(2p/e)^p. \tag{13}$$

Inserting this into (12) we obtain

$$\mathbb{P}[X_n \geq \gamma] \leq \gamma^{-2p}\sqrt{2}(2p/e)^p\left(\frac{k}{n}\right)^{2p}(1+\varepsilon_{n,2p}).$$

Putting

$$s := \frac{\gamma^2}{2}\frac{n^2}{k^2},$$

and taking logarithms we arrive at

$$s + \log\mathbb{P}[X_n \geq \gamma] \leq s - p - p\log\left(\frac{s}{p}\right) + \log\sqrt{2} + \log((1+\varepsilon_{n,2p}))).$$

Now choose $p = \text{int}(s+\frac{1}{2})$, where $\text{int}(x)$ is the integer part of x. Note that this choice implies that indeed $p = \mathcal{O}(\frac{n^2}{k^2})$ and that

$$s \in [p - \frac{1}{2}, p + \frac{1}{2}].$$

Let

$$f_p(s) := s - p - p\log\left(\frac{s}{p}\right).$$

One easily sees that f_p is convex, f_p'' is decreasing on $[p-\frac{1}{2}, p+\frac{1}{2}]$ and therefore

$$f_p(s) \leq f_p(p-\frac{1}{2}) = -\frac{1}{2} + p\log(\frac{2p}{2p-1}) \leq -\frac{1}{2} + \log 2,$$

the latter since $x \mapsto x\log\frac{2x}{2x-1}$ is decreasing on $[1,+\infty)$.
Hence, we obtain

$$s + \log\mathbb{P}[X_n \geq \gamma] \leq \frac{3\log 2 - 1}{2} + \log((1+\varepsilon_{n,2p}))),$$

which implies

$$\mathbb{P}[X_n \geq \gamma] \leq 2\ e^{-s} \tag{14}$$

for n large enough. By the definition of s, this is exactly what we claimed for the upper bound.

The proof of the lower bound is based on the construction of a transformed measure, similarly as in the proof of Lemma 2.3 in [29], by König and Mörters. We define for any positive integer p the measure

$$d\widehat{\mathbb{P}}_{2p}(x) := \frac{x^{2p}}{Z_{2p}} d\mathbb{P}^{X_n}(x),$$

where $Z_{2p} = \int x^{2p} d\mathbb{P}^{X_n}$. Moreover, define the random variable $W_n = \log(|X_n|/\gamma)$. We first prove that for $p = \text{int}(\frac{\gamma^2}{2}\frac{n^2}{k^2})$ and all $\epsilon > 0$,

$$\lim_{n\to+\infty} \widehat{\mathbb{P}}_{2p}[|W_n| \leq \epsilon] = 1.$$

To prove this, we consider an arbitrary $\epsilon > 0$ and some small real $\alpha > 0$ such that $p\alpha$ is an integer. Then, by the Markov inequality, we get

$$\widehat{\mathbb{P}}_{2p}[W_n \geq \epsilon] = \widehat{\mathbb{P}}_{2p}[(X_n)^{2p\alpha} \geq \gamma^{2p\alpha} e^{2p\alpha\epsilon}] \leq \gamma^{-2p\alpha} e^{-2p\alpha\epsilon}\ \widehat{\mathbb{E}}_{2p}((X_n)^{2p\alpha}).$$

By definition of $\widehat{\mathbb{P}}_{2p}$, we have

$$\widehat{\mathbb{E}}_{2p}((X_n)^{2p\alpha}) = \frac{Z_{2p(1+\alpha)}}{Z_{2p}}.$$

Using our Lemma 4.3 and Stirling's formula, we see that

$$\widehat{\mathbb{E}}_{2p}((X_n)^{2p\alpha}) = (\frac{2p}{e})^{\alpha p}\ (1+\alpha)^{(1+\alpha)p} (\frac{k}{n})^{2\alpha p}\ (1+r_n),$$

where $r_n \to 0$ as $n \to +\infty$. We deduce that

$$\widehat{\mathbb{P}}_{2p}[W_n \geq \epsilon] \leq \exp(\alpha p(-1-2\epsilon + \frac{1+\alpha}{\alpha}\log(1+\alpha) + \log(\frac{2pk^2}{\gamma^2 n^2})))\ (1+r_n).$$

Choosing $p = \text{int}(\frac{\gamma^2}{2}\frac{n^2}{k^2})$, we get for n large enough

$$\widehat{\mathbb{P}}_{2p}[W_n \geq \epsilon] \leq C\exp(\alpha p(-1-2\epsilon + \frac{1+\alpha}{\alpha}\log(1+\alpha))),$$

for some constant $C > 0$. For a given $\epsilon > 0$, we can choose $\alpha > 0$ (depending on p) small enough, such that the term inside the exponential is strictly negative and αp is an integer going to infinity as n goes to infinity. For instance, we choose $\alpha = \text{int}(\sqrt{p})/p$ for p large enough. This implies that $\lim_{n\to\infty} \widehat{\mathbb{P}}_{2p}[W_n \geq \epsilon] = 0$.
Similarly, one shows that

$$\lim_{n\to\infty} \widehat{\mathbb{P}}_{2p}[W_n \leq -\epsilon] = 0.$$

Now, we can finish the proof of the lower bound. Let $\epsilon > 0$ arbitrarily fixed. We have then

$$\begin{aligned}
\mathbb{P}[X_n \geq \gamma e^{-\epsilon}] &= \frac{1}{2}\mathbb{P}[|X_n| \geq \gamma e^{-\epsilon}] \\
&\geq \frac{1}{2}\mathbb{P}[|W_n| \leq \epsilon] \\
&= \frac{1}{2}\ Z_{2p}\ \widehat{\mathbb{E}}_{2p}((X_n)^{-2p}\mathbf{1}_{\{|W_n|\leq\epsilon\}}) \\
&\geq \frac{1}{2}\ Z_{2p}\ e^{-2p\epsilon}\ \gamma^{-2p}\ \widehat{\mathbb{P}}_{2p}[|W_n| \leq \epsilon].
\end{aligned}$$

Choosing $p = \text{int}(\frac{\gamma^2}{2}\frac{n^2}{k^2})$ and using our Lemma 4.3 and Stirling's formula to estimate Z_{2p}, we get

$$\mathbb{P}[X_n \geq \gamma e^{-\epsilon}] \geq \frac{\sqrt{2}}{2} e^{-\frac{\gamma^2}{2}\frac{n^2}{k^2}(1+2\epsilon)}\ \widehat{\mathbb{P}}_{2p}[|W_n| \leq \epsilon]\ (1+r_n),$$

from which we deduce

$$\liminf_{n\to\infty} \frac{k^2}{n^2}\log\mathbb{P}[X_n \geq \gamma e^{-\epsilon}] \geq -\frac{\gamma^2}{2}(1+\epsilon).$$

Since we can take $\epsilon > 0$ arbitrarily small, we get the lower bound. □

References

[1] B. Aazhang, M.K. Varanasi; Multistage detection in asynchronous code-division multiple acces communications, IEEE Trans. Commun. 38, 509-519 (1990).

[2] S. Amari; H.-F. Yanai; auto-Associative Memory with Two-Stage Dynamics of Nonmonotonic Neurons; IEEE Trans. on Neural Networks 7, 803–815 (1996).

[3] A. Bovier; Sharp upper bounds for perfect retrieval in the Hopfield model; J. Appl. Prob. 36, 941–950 (1999).

[4] A. Bovier, Statistical mechanics of disordered system. A mathematical perspective, Cambridge Series in Statistical and Probabilistic Mathematics 18, Cambridge University Press (2006).

[5] A. Bovier, V. Gayrard; Hopfield models as a generalized mean field model; Preprint, in "Mathematics of spin glasses and neural networks", A. Bovier, P.Picco (eds.), "Progress in Probability", Birkhäuser (1998).

[6] Mathematical Aspects of Spin Glasses and Neural Networks, A. Bovier, P.Picco (eds.), "Progress in Probability", Birkhäuser, Boston (1998).

[7] R.K. Bahr, J.S. Sadowski; Direct-sequence spread-spectrum multiple-access communications with random signature sequences: a large deviations analysis, IEEE Trans. Inform. Th. 37, 514-527 (1991).

[8] A. Ben-Israel, A. Charnes ; Contribution to the theory of generalized inverses; J. SIAM 11, 667–699 (1963).

[9] R.M. Buehrer, A. Kaul, S. Striglis, B.D. Woerner; Analysis of DS-CDMA parallel interference cancellation with phase and timing errors, IEEE JSAC 14, 1522-1535 (1996).

[10] R.M. Buehrer, S.P. Nicoloso, S. Gollamudi; Linear versus non-linear interference cancellation, IEEE/IEICE Journal of Communications and Networks 1, 118-133 (1999).

[11] R.M. Buehrer, B.D. Woerner; Analysis of adaptive multistage interference cancellation for CDMA using an improved Gaussian approximation, IEEE Trans. Commun. 44, 1308-1329 (1996).

[12] B. Crespi; Storage capacity of non-monotonic neurons; Neural Networks 12, 1377–1389 (1999).

[13] P. de Jong; A Central Limit Theorem for Generalized Multilinear Forms; J. of Multivariate Anal. 34, 275-289 (1990).

[14] G. Dreyfus, I. Guyon, L. Personnaz; Information storage and retrieval in spin-glass like neural networks; J. Physique Lettres 46, L359–L365 (1985).

[15] G. Dreyfus, I. Guyon, L. Personnaz; Collective computational properties of neural networks : New learning mechanisms. Phys. Rev. A 34, 4217–4228 (1986).

[16] P. Eichelsbacher, M. Löwe; A large deviation principle for m-variate von Mises-statistics and U-statistics. J. Theoret. Probab. 8,807–824 (1995).

[17] P. Eichelsbacher, M. Löwe; Moderate deviations for i.i.d. random variables. ESAIM Probab. Stat. 7, 209–218 (2003).

[18] L.A. Pastur, A.L. Figotin; Exactly soluble model of a spin-glas; Sov. J. of Low Temprature Phys. 3(6), 378-383 (1977).

[19] R. van der Hofstad, G. Hooghiemstra, M.J. Klok; Large deviations for code division multiple access systems, SIAM Journ. Appl. Math. 62 1044–1065 (2002).

[20] R. van der Hofstad, M. J. Klok; Improving the performance of third-generation wireless communication systems. Adv. in Appl. Probab. 36, 1046–1084 (2004).

[21] R. van der Hofstad, M.J. Klok; Performance of DS-CDMA systems with optimal hard decision parallel interference cancellation, IEEE Trans. Inform. Theory 49, 2918–2940 (2003).

[22] R. van der Hofstad, M. Löwe, F. Vermet; The effect of system load on the existence of bit-errors in CDMA with and without parallel interference cancelation. IEEE Transactions on Information Theory 52, 4733–4741, (2006).

[23] J.M. Holtzman, ; A simple, accurate method to calculate spread spectrum multiple-access error probabilities, IEEE Trans. Commun., 40,461-464 (1992).

[24] J.J. Hopfield; Neural networks and physical systems with emergent collective computational abilities; Proc. nat. Acad. Sci. USA 79, 2554-2558 (1982).

[25] M. Juntti; Multiuser Demodulation for DS-CDMA Systems in Fading Channels, Ph.D. thesis, University of Oulu, Finland (1998).

[26] I. Kanter, H. Sampolinski; Associative recall of memory without errors; Phys. Rev. A 35, 380–392 (1987).

[27] M.J. Klok, G. Hooghiemstra, T. Ojanperä, R. Prasad; A novel technique for DS-CDMA system performance evaluation, VTC '99 spring, Houston, USA (1999).
[28] K. Kobayashi; On the capacity of a neuron with a nonmonotone output function; Network 2, 237–243 (1991).
[29] W. König, P. Mörters; Brownian intersection local times : upper tail asymptotics and thick points; Ann. Prob., 30, 1605–1656 (2002).
[30] M. Latva-aho; Advanced Receivers for Wideband CDMA Systems, Ph.D. Thesis, University of Oulu, Finland (1999).
[31] J.S. Lehnert, R.K. Morrow; Bit-to-bit-error dependence in slotted DS/SSMA packet systems with random signature sequences, IEEE Trans.Commun. COM-37, 1052-1061 (1989).
[32] J.S. Lehnert, M.B. Pursley; Error probabilities for binary direct sequence spread-spectrum communications with random signature sequences, IEEE Trans. Commun., 87-98 (1987).
[33] M. Löwe; On the storage capacity of Hopfield models with weakly correlated patterns; Annals of Appl. Probability 8, 1216–1250 (1999).
[34] M. Löwe, F. Vermet; The storage capacity of the Hopfield model and moderate deviations. Stat. & Prob. Lett. 75, 237–248 (2005).
[35] M. Löwe; F. Vermet; The Capacity of q-state Potts neural networks with Parallel Retrieval Dynamics; Stat. & Prob. Lett. 77, 1505–1514 (2007).
[36] R. McEliece, E. Posner, E. Rodemich. S. Venkatesh; The capacity of the Hopfield associative memory; IEEE Inf. Th. 33, 461-482 (1987).
[37] M.O. Sunay, P.J. Mclane; Calculating error probabilities for DS CDMA systems: When not to use the Gaussian approximation, IEEE Globecom 96, 3,1744-1749 (1996).
[38] S.K. Mitra, C. R. Rao; Generalized Inverse of Matrices and its Applications; Wiley, New York (1971).
[39] M. Morita; Associative memory with nonmonotone dynamics; Neural Networks 6, 115–126 (1993).
[40] M. Morita; S. Yoshizawa; K. Nakano; Analysis and improvement of the dynamics of autocorrelation associative memory; Trans. Inst. Elect., Inf., Comm. Eng. Japan J73-D-II, 232–242 (1990).
[41] N. Nishimori; I. Opris; Retrieval process of an associative memory with nonmonotonic input-output function; IEEE Int. Conf. on Neural Networks 1, 353–358 (1993).
[42] G. Palm; Memory capacities of local rules for synaptic modification; J. of Concepts in Neuroscience 2, 97–128 (1991).
[43] D. Petritis; Thermodynamic formalism of neural computing; Nonlinear Phenomena of Complex Systems 2, 86–146, Kluwer Acad. Publ., Dordrecht (1996).
[44] P. Picco; Artificial neural networks. A review from Physical and Mathematical point of view; Ann. I.H.P., section A, 64, 289-307 (1996).
[45] R. Prasad, R; CDMA for Wireless Personal Communications, Artech House (1996).
[46] E. Rio; Théorie asymptotique des processus aléatoires faiblement dépendants, Ed. Springer, Paris (2000).
[47] F. Vermet; Étude asymptotique d'un réseau neuronal : le modèle de mémoire associative de Hopfield; Ph.D. Thesis. University of Rennes 1, France (1994).

(Matthias Löwe) Fachbereich Mathematik und Informatik, Universität Münster, Einsteinstrasse 62, 48149 Münster, Germany
E-mail address, Matthias Löwe: maloewe@math.uni-muenster.de

(Franck Vermet) Laboratoire de Mathématiques, UMR CNRS 6205, Université de Bretagne Occidentale, 6, avenue Victor Le Gorgeu, CS 93837, F-29238 BREST Cedex 3, France
E-mail address, Franck Vermet: Franck.Vermet@univ-brest.fr