# Probabilistic Information Capacity of Hopfield Associative Memory 

Kesig Lee*<br>S. C. Kothari<br>Department of Computer Science, Iowa State University, Ames, Iowa 50011, USA<br>Dongwan Shin ${ }^{\dagger}$<br>Department of Statistics, Iowa State University, Ames, Iowa 50011, USA


#### Abstract

This paper defines a formal probabilistic notion for the information capacity of the Hopfield neural network model of associative memory. A mathematical expression is derived for the number of random binary patterns that can be stored as stable states in a Hopfield model of memory with $n$ neurons with a given probability. The derivation is based on a new approach using two powerful mathematical techniques: Brown's Martingale Central Limit Theorem and Gupta's transformation of the probability integral for a special case of the correlation matrix. The new approach provides a way for rigorously analyzing the complex dynamics of the Hopfield model. Our approach refines the current heuristic methods, which rely on simplifying assumptions about the dynamics of the model.


## 1. Introduction

The ability to recall memorized patterns is an important feature of human memory. In 1982, Hopfield [13] introduced a new model of associative memory based on a simple neural network model. Computational properties of his model emerge as collective properties of a system having a large number of simple neurons. Establishing empirically that such collective properties include default assignment, error correction, and spontaneous generalization, Hopfield demonstrated the attractiveness of the neural network model for many applications. Hopfield's work has rekindled the interest in neural networks.

[^0]An important step in studying the Hopfield network model is to mathematically quantify its performance as a memory. In the Hopfield model, a memory is defined as a stable state, that is, an activation state that remains unchanged with network iterations. The information capacity of the standard memory model is explicitly determined by the number of memory bits. In contrast, the information capacity of the Hopfield model is based on the dynamics of activation patterns in the neural network. Hence, the estimation of the information capacity in the Hopfield model is considerably more complex.

The dependence of the information capacity on the dynamics of the network has prompted researchers $[4,5,13,19,22,23]$ to consider probabilistic estimates of the information capacity of the Hopfield network based on simplifying assumptions. Amari [4, 5] and McEliece [19] have provided heuristic estimates of the capacity of the Hopfield network model based on statistical techniques. Weisbuch [22] also used statistical techniques with additional assumptions to provide the results of information capacity that were more consistent with simulation than were the other studies $[4,5,13,19]$.

We formalize the notion of the probabilistic information capacity in the Hopfield model. The formalization is useful for clarifying previous work $[4,5,6,13,19,22,23]$ and, more importantly, it paves the way for introducing powerful statistical techniques to rigorously analyze the information capacity. We provide a new approach for rigorously analyzing the complex dynamics of the Hopfield memory model. This approach is based on two powerful mathematical techniques: Brown's Martingale Central Limit Theorem [9] and Gupta's transformation [12].

In section 2 Hopfield's model of associative memory is reviewed and the notion of the probabilistic information capacity is formalized. Section 3 examines the current heuristic methods for the estimation of the information capacity. In section 4 our approach for analyzing the capacity of the Hopfield memory is discussed along with the main theorems. In section 5 the performance results based on our mathematical analysis are compared with the results from other theoretical $[4,5,19,23]$ and experimental [22] studies. Section 6 provides concluding remarks.

## 2. The Hopfield associative memory

This section provides a self-contained description of the Hopfield model of associative memory and introduces a formal notion of the probabilistic information capacity.

The Hopfield neural network model of associative memory consists of $n$ pairwise connected neurons. Any neuron $i$ can be in one of two states: $v_{i}=0$ (off) or $v_{i}=1$ (on).

Definition 2.1. A state vector $\mathbf{V}=\left[v_{1}, \ldots, v_{n}\right]$ is defined to be a binary vector whose $i$ th component corresponds to the state of the ith neuron.

Definition 2.2. A connection matrix is the $n \times n$ matrix $\mathbf{W}=\left(w_{i, j}\right)$, where the $(i, j)$ th entry of $\mathbf{W}$ is the strength of the synaptic connection from neuron $i$ to neuron $j$.

Each choice of $\mathbf{W}$ defines a specific neural network of $n$ neurons. In other words, the collective behavior of the neural network is entirely specified by $\mathbf{W}$. In fact, the matrix $\mathbf{W}$ acts as a decoding machine that can be recognized as a kind of information storage. The Hopfield model requires that $w_{i, j}=w_{j, i}$ and $w_{i, i}=0$.

According to the Hebbian learning rule [19], to memorize (store) $m$ patterns (state vectors) $\mathbf{V}^{1}, \mathbf{V}^{2}, \ldots, \mathbf{V}^{m}$ in the Hopfield neural network, each entry of the connection matrix $\mathbf{W}$ is computed by

$$
\begin{equation*}
w_{i, j}=\sum_{s=1}^{m}\left(2 v_{i}^{s}-1\right) \cdot\left(2 v_{j}^{s}-1\right) \tag{2.1}
\end{equation*}
$$

which can be rewritten as

$$
\begin{equation*}
w_{i, j}=\sum_{s=1}^{m} X_{i}^{s} \cdot X_{j}^{s} \tag{2.2}
\end{equation*}
$$

where $X_{i}^{s}=\left(2 v_{i}^{s}-1\right)$.
A randomly selected neuron receives inputs from connected neurons and changes its state in the following manner at each discrete time step $t$ :

$$
\begin{equation*}
v_{i}(t)=\operatorname{sgn}\left\{\sum_{j=1}^{n} w_{i, j} \cdot v_{j}(t-1)\right\}=\mathbf{T}(\mathbf{V}(t-1)) \tag{2.3}
\end{equation*}
$$

where

$$
\operatorname{sgn}(y)= \begin{cases}1 & \text { if } y \geq 0 \\ 0 & \text { otherwise }\end{cases}
$$

and $v_{i}(t)$ represents the state of the $i$ th neuron at time $t$. $\mathbf{T}$ is a nonlinear state transition operator.

Finally, the recalling process of the memorized (stored) patterns can be described as follows. Start with an initial state represented by a binary vector $\mathbf{V}(0)$. The state is changed iteratively according to equation (2.3). The iterative process is repeated until a state that remains unchanged with further network iterations is reached. The terminal state, call it $\mathbf{V}^{1}$, is said to be recalled from $\mathbf{V}(0)$.

Definition 2.3. A state vector (a pattern) $\mathbf{V}$ is called a stable state iff $\mathbf{V}$ is recalled from $\mathbf{V}$, namely, iff from equation (2.3)

$$
\begin{equation*}
\forall i \in\{1, \ldots, n\} \quad v_{i} \cdot\left(\sum_{j=1}^{n} w_{i, j} v_{j}\right) \geq 0 \tag{2.4}
\end{equation*}
$$

Let $\mathbf{S}_{m}$ denote a set of $m$ random binary patterns $\mathbf{V}^{1}, \ldots, \mathbf{V}^{m}$, each of size $n$. Consider a Hopfield network where the weights are as given by the Hebbian learning for storing all the patterns in $\mathbf{S}_{m}$. Let $\mathbf{P}\left(\mathbf{S}_{m}\right)$ denote the probability that all the patterns in $\mathbf{S}_{m}$ are in fact stable patterns.
Definition 2.4. Given an $\alpha \in[0,1]$, the information capacity $\mathbf{I}_{\alpha}$ is defined to be the maximum integer $m$ such that $\mathbf{P}\left(\mathbf{S}_{m}\right) \geq \alpha$.

## 3. Information capacity heuristics

In this section we first review the current heuristic methods for estimating the information capacity of the Hopfield model. After the review, the main ideas behind our approach are discussed. This section uses the framework developed in section 2.

### 3.1 Current quantitative heuristics and proposed extension

Several researchers $[4,5,13,19,22,23]$ have proposed statistical methods for exploring the capacity of the Hopfield memory. These methods depend upon two invariance conditions derived from equations (2.2), (2.3), and (2.4). Let the number of neurons be $n$ and the number of stored patterns be $m$. The invariance conditions (ICs) are:
IC 3.1. Given a stored pattern $\mathbf{V}^{s}$, an off-state neuron $i$ in the pattern remains off (0) if

$$
\begin{equation*}
0>\sum_{j=1, j \neq i}^{n} w_{i, j} \cdot v_{j}^{s}=-\sum_{j=1, j \neq i}^{n} v_{j}^{s}+\sum_{s^{\prime}=1, s^{\prime} \neq s}^{m} \sum_{j=1, j \neq i}^{n} X_{i}^{s^{\prime}} \cdot X_{j}^{s^{\prime}} \cdot v_{j}^{s} \tag{3.1}
\end{equation*}
$$

IC 3.2. Given a stored pattern $\mathbf{V}^{s}$, an on-state neuron $i$ in the pattern remains on (1) if

$$
\begin{equation*}
0 \leq \sum_{j=1, j \neq i}^{n} w_{i, j} \cdot v_{j}^{s}=\sum_{j=1, j \neq i}^{n} v_{j}^{s}+\sum_{s^{\prime}=1, s^{\prime} \neq s}^{m} \sum_{j=1, j \neq i}^{n} X_{i}^{s^{\prime}} \cdot X_{j}^{s^{\prime}} \cdot v_{j}^{s} \tag{3.2}
\end{equation*}
$$

To use the invariance conditions for statistical analysis of the information capacity, all approaches (those of Amari, McEliece, Weisbuch, and us) start from the following assumption.

Assumption 3.1. For $s=1, \ldots, m$ and $i=1, \ldots, n$, all $\mathrm{v}_{i}^{s}$ are identically, independently distributed (i.i.d.) random variables that take values either 1 or 0 with probability 0.5 . Note that this is a necessary assumption to handle random patterns.

For a given $s$ and $i$, define the signal term $\left(S_{i}^{s}\right)$ as $\sum_{j=1, j \neq i}^{n} v_{j}^{s}$ and the noise term $\left(N_{i}^{s}\right)$ as $\sum_{s^{\prime}=1, s^{\prime} \neq s}^{m} \sum_{j=1, j \neq i}^{n} X_{i}^{s^{\prime}} \cdot X_{j}^{s^{\prime}} \cdot v_{j}^{s}$. We want to evaluate the probability that the invariance condition for a neuron is satisfied. Because of symmetry, the probability is the same for both invariance conditions. The following discussion is based on IC 3.1.

The existing studies due to Amari, McEliece, and Weisbuch require an additional assumption.

Assumption 3.2. Each element in the noise term $X_{i}^{s^{\prime}} \cdot X_{j}^{s^{\prime}} \cdot v_{j}^{s}$ where $s^{\prime} \neq s$, $s^{\prime}=1, \ldots, m, j \neq i$, and $j=1, \ldots, n$ is an i.i.d. random variable.

Based on these assumptions (3.1 and 3.2) and the Central Limit Theorem, the current capacity heuristics lead to the result

$$
\begin{equation*}
-S_{i}^{s}+N_{i}^{s} \sim \mathbf{N}\left(-\frac{(n-1)}{2}, \frac{(m-1)(n-1)}{2}\right) \tag{3.3}
\end{equation*}
$$

where $\mathbf{N}$ means a normal distribution.
Assumption 3.2 is introduced to simplify the analysis, but it is not accurate because the elements of the noise terms are in fact not independent of each other. In Theorem 4.1 we prove, without using Assumption 3.2, the result

$$
\begin{equation*}
-S_{i}^{s}+N_{i}^{s} \sim \mathbf{N}\left(-\frac{(n-1)}{2}, \frac{(2 m-1)(n-1)}{4}\right) \tag{3.4}
\end{equation*}
$$

for all $s=1, \ldots, m$, and $i=1, \ldots, n$.
The existing studies derive the following result from equation (3.3). The probability that a neuron satisfies the invariance condition is given by

$$
\begin{equation*}
\mathbf{P}\left(S_{i}^{s}>N_{i}^{s}\right)=\Phi(z)=\int_{-\infty}^{z} \phi(t) d t \tag{3.5}
\end{equation*}
$$

where $\phi(t)=(1 / \sqrt{2 \pi}) e^{-t^{2} / 2}$ for $-\infty<t<\infty$, and $z=\sqrt{(n-1)} / \sqrt{2(m-1)}$.
Remark. In our case, by equation (3.4), $\mathbf{P}\left(S_{i}^{s}>N_{i}^{s}\right)=\Phi\left(z^{\prime}\right)$ where $z^{\prime}=$ $\sqrt{(n-1)} / \sqrt{2 m-1}$.

From equation (3.5), McEliece [19] approximated the expected number of failed neurons (not satisfying the invariance condition) in a stored pattern to $n(1-\Phi(z))$. Making the conjecture that the number of failed neurons approximately follows a Poisson distribution, he derived the probability that a stored pattern is indeed a stable state (fixed point) as follows. For a fixed probability $\beta$ (for example, $\beta=0.9999$ ),

$$
\begin{align*}
\beta & =\exp \{-n(1-\Phi(z))\} \\
& =\exp \left\{-n\left(1-\Phi\left(\frac{\sqrt{(n-1)}}{\sqrt{2(m-1)}}\right)\right)\right\} \tag{3.6}
\end{align*}
$$

Remark. If $X \sim \operatorname{Poisson}(\lambda)$ for a fixed $\lambda$, then $\mathrm{P}(X=x)=\left(e^{-\lambda} \cdot \lambda^{x}\right) / x$ !. In particular $\mathrm{P}(X=0)=\mathrm{P}($ no failure $)=e^{-\lambda}$.

From equation (3.6), McEliece derived the approximation that $m \sim$ $n / 2 \log n$ with the assumption that $n(1-\Phi(z))$ is a constant for all $n$, which
is not necessarily the case. Note that Amari [4] also derived similar result such that $m / n<1 /(2 \log n-\log \log n)$.

On the other hand, Amari $[4,5]$ and Weisbuch $[22,23]$ used the following additional assumption.

Assumption 3.3. $\forall s=1, \ldots, m, \forall i=1, \ldots, n$, each term $-S_{i}^{s}+N_{i}^{s}$ that is a random variable defining the invariance condition for a neuron is independent of each other.

Using equation (3.5) and Assumption 3.3, they concluded that the probability that all stored patterns are stable is

$$
\begin{equation*}
\mathbf{P}=\left(\mathbf{P}\left(S_{i}^{s}>N_{i}^{s}\right)\right)^{n m}=\{\Phi(z)\}^{n m} \tag{3.7}
\end{equation*}
$$

Amari [4] and Kamp [16] acknowledged that the dependency among the random variables $-S_{i}^{s}+N_{i}^{s}$ for $1 \leq i \leq n$ and $1 \leq s \leq m$ cannot be neglected through some simulations. In our proposed heuristic, Assumption 3.3 will not be used; instead we will account for the dependency among the random variables used in the assumption (Theorem 4.2 and Theorem 4.3).

In summary, with the help of powerful mathematical techniques, we will eliminate the use of incorrect assumptions (3.2 and 3.3), which so far have been used to simplify the analysis. In Theorem 4.4, we derive a mathematical expression for the information capacity by using the results from Theorem 4.2 , Theorem 4.3, and Gupta's transformation technique [12].

## 4. Capacity based on multivariate normal approximation

In this section we provide proofs for our claims in section 3.2. Our proofs are based on the invariance condition given by equation (3.1) and Assumption 3.1. Let

$$
\begin{equation*}
\mathbf{A}(i, s)=-S_{i}^{s}+N_{i}^{s}=-\sum_{j=1, j \neq i}^{n} v_{j}^{s}+\sum_{s^{\prime}=1, s^{\prime} \neq s}^{m} \sum_{j=1, j \neq i}^{n} X_{i}^{s^{\prime}} \cdot X_{j}^{s^{\prime}} \cdot v_{j}^{s} \tag{4.1}
\end{equation*}
$$

Proposition 4.1. Under Assumption 3.1, approximately,

$$
S_{i}^{s} \sim \mathbf{N}\left(\frac{n-1}{2}, \frac{n-1}{4}\right)
$$

Proof. By Assumption 3.1, $S_{i}^{s} \sim \mathbf{B}(n-1,1 / 2)$ where $\mathbf{B}(n, p)$ is a binomial distribution with parameters $n$ and $p$. Then, by the Central Limit Theorem, we can approximate $\mathbf{B}(n-1,1 / 2)$ by $\mathbf{N}((n-1) / 2,(n-1) / 4)$ for large $n$.

Proposition 4.2. Under Assumption 3.1, the variables $X_{i_{1}}^{s_{1}} \cdot X_{j_{1}}^{s_{1}} \cdot v_{j_{1}}^{s}$ and $X_{i_{2}}^{s_{2}} \cdot X_{j_{2}}^{s_{2}} \cdot v_{j_{2}}^{s}$ for $1 \leq i_{1}, i_{2}, j_{1}, j_{2} \leq n, 1 \leq s, s_{1}, s_{2} \leq m, s \neq s_{1}, s \neq s_{2}$ are mutually uncorrelated if $i_{1}, i_{2}$ are distinct from $j_{1}, j_{2}$ or $s_{1} \neq s_{2}$.

Proof. By Assumption 3.1 and equation (2.2),

$$
\begin{equation*}
\mathbf{E}\left(X_{i}^{s}\right)=0, \quad \forall s, i, 1 \leq s \leq m, 1 \leq i \leq n . \tag{4.2}
\end{equation*}
$$

By definition,

$$
\begin{aligned}
& \operatorname{Cov}\left(X_{i_{1}}^{s_{1}} \cdot X_{j_{1}}^{s_{1}} \cdot v_{j_{1}}^{s}, X_{i_{2}}^{s_{2}} \cdot X_{j_{2}}^{s_{2}} \cdot v_{j_{2}}^{s}\right) \\
& \quad=\mathbf{E}\left(X_{i_{1}}^{s_{1}} \cdot X_{j_{1}}^{s_{1}} \cdot v_{j_{1}}^{s} \cdot X_{i_{2}}^{s_{2}} \cdot X_{j_{2}}^{s_{2}} \cdot v_{j_{2}}^{s}\right)
\end{aligned}
$$

The proposition follows by considering different cases and applying the rule that $\mathbf{E}\left(t_{1} \cdot t_{2}\right)=\mathbf{E}\left(t_{1}\right) \cdot \mathbf{E}\left(t_{2}\right)$ if $t_{1}$ and $t_{2}$ are independent random variables.

Remark. While these random variables $\left(X_{i}^{s^{\prime}} \cdot X_{j}^{s^{\prime}} \cdot v_{j}^{s}\right.$ ) are identically distributed and are mutually uncorrelated, they are not independent of each other. For example, the variables $\left(X_{i}^{1} \cdot X_{j}^{1} \cdot v_{j}^{2}\right),\left(X_{i}^{1} \cdot X_{j}^{1} \cdot v_{j}^{3}\right)$, and $\left(X_{i}^{2} \cdot X_{j}^{2} \cdot v_{j}^{3}\right)$ are clearly not independent.

Proposition 4.3. Under Assumption 3.1, $\mathbf{E}\left(N_{i}^{s}\right)=0$ and $\operatorname{Var}\left(N_{i}^{s}\right)=$ $(m-1)(n-1) / 2$ for $1 \leq i \leq n$ and $1 \leq s \leq m$.

Proof. Fix $i$ and $s$, and select two different elements $X_{i}^{s_{1}^{\prime}} \cdot X_{j_{1}}^{s_{1}^{\prime}} \cdot v_{j_{1}}^{s}$ and $X_{i}^{s_{2}^{\prime}} \cdot X_{j_{2}}^{s_{2}^{\prime}} \cdot v_{j_{2}}^{s}$.

$$
\mathbf{E}\left(N_{i}^{s}\right)=\sum_{s^{\prime}=1, s^{\prime} \neq s}^{m} \sum_{j=1, j \neq i}^{n} \mathbf{E}\left(X_{i}^{s^{\prime}} \cdot X_{j}^{s^{\prime}} \cdot v_{j}^{s}\right)=0 .
$$

And, by Proposition 4.2,

$$
\begin{aligned}
\operatorname{Var} & \left(N_{i}^{s}\right) \\
= & \sum_{s^{\prime}=1, s^{\prime} \neq s}^{m} \sum_{j=1, j \neq i}^{m} \operatorname{Var}\left(X_{i}^{s^{\prime}} \cdot X_{j}^{s^{\prime}} \cdot v_{j}^{s}\right) \\
& +\sum_{\substack{s_{1}^{\prime}=1, s_{1}^{\prime} \neq s}}^{m} \sum_{\substack{s_{2}^{\prime}=1, s_{2}^{\prime} \neq s}}^{m} \sum_{j_{1}=1,1, j_{j}}^{n} \sum_{j_{2}=1,}^{n} \operatorname{Cov}\left(X_{i}^{s_{1}^{\prime}} \cdot X_{j_{1}}^{s_{1}^{\prime}} \cdot v_{j_{1}}^{s}, X_{i}^{s_{2}^{\prime}} \cdot X_{j_{2}}^{s_{2}^{\prime}} \cdot v_{j_{2}}^{s}\right) \\
= & \sum_{s^{\prime}=1, s^{\prime} \neq s}^{m} \sum_{j=1, j \neq i}^{n} \operatorname{Var}\left(X_{i}^{s^{\prime}} \cdot X_{j}^{s^{\prime}} \cdot v_{j}^{s}\right),
\end{aligned}
$$

because the set of random variables $\left\{X_{i}^{s^{\prime}} \cdot X_{j}^{s^{\prime}} \cdot v_{j}^{s}, 1 \leq i, j \leq n, 1, \leq s, s^{\prime} \leq m\right\}$ are identically distributed,

$$
\begin{aligned}
& =(m-1)(n-1) \operatorname{Var}\left(X_{i}^{1} \cdot X_{1}^{1} \cdot v_{1}^{s}\right) \\
& =(m-1)(n-1)\left\{\mathbf{E}\left(\left(X_{i}^{1} \cdot X_{1}^{1}\right)^{2} \cdot\left(v_{1}^{s}\right)^{2}\right)-\mathbf{E}\left(X_{i}^{1}\right) \cdot \mathbf{E}\left(X_{1}^{1} \cdot v_{1}^{s}\right)\right\} \\
& =(m-1)(n-1) \mathbf{E}\left(v_{1}^{s}\right)^{2}=(m-1)(n-1) / 2
\end{aligned}
$$

Proposition 4.4. Under Assumption 3.1, we have approximately $N_{i}^{s} \sim$ $\mathbf{N}(0,(m-1)(n-1) / 2)$

Proof. By Proposition 4.2, Proposition 4.3, and Brown's Martingale Central Limiting Theorem (refer to appendix A for a detailed proof).

Proposition 4.5. $S_{i}^{s}$ and $N_{i}^{s}$ are uncorrelated.

## Proof.

$$
\begin{aligned}
\operatorname{Cov}\left(S_{i}^{s}, N_{i}^{s}\right) & =\operatorname{Cov}\left(\sum_{j=1, j \neq i}^{n} v_{j}^{s}, \sum_{s^{\prime}=1, s^{\prime} \neq s}^{m} \sum_{j=1, j \neq i}^{n} X_{i}^{s^{\prime}} \cdot X_{j}^{s^{\prime}} \cdot v_{j}^{s}\right) \\
& =\sum_{j_{1}=1, j_{1} \neq i}^{n} \sum_{j_{2}=1, j_{2} \neq i}^{n} \sum_{s^{\prime}=1, s^{\prime} \neq s}^{m} \operatorname{Cov}\left(v_{j_{1}}^{s}, X_{i}^{s^{\prime}} \cdot X_{j_{2}}^{s^{\prime}} \cdot v_{j_{2}}^{s}\right) \\
& =0
\end{aligned}
$$

Because, by equation (4.2), $\forall j_{1}, j_{2}$, and $s^{\prime}$,

$$
\begin{aligned}
\operatorname{Cov}\left(v_{j_{1}}^{s}, X_{i}^{s^{\prime}} \cdot X_{j_{2}}^{s^{\prime}} \cdot v_{j_{2}}^{s}\right) & =\mathbf{E}\left(v_{j_{1}}^{s} \cdot X_{i}^{s^{\prime}} \cdot X_{j_{2}}^{s^{\prime}} \cdot v_{j_{2}}^{s}\right) \\
& =\mathbf{E}\left(v_{j_{1}}^{s}\right) \mathbf{E}\left(X_{i}^{s^{\prime}}\right) \mathbf{E}\left(X_{j_{2}}^{s^{\prime}} \cdot v_{j_{2}}^{s}\right) \\
& =0
\end{aligned}
$$

Proposition 4.6. If $S_{1} \xrightarrow{d} N_{1}$ and $S_{2} \xrightarrow{d} N_{2}$ where $N_{1}$ and $N_{2}$ are uncorrelated normal random variables, and d denotes convergence in distribution, then $S_{1}+S_{2} \xrightarrow{d} N_{1}+N_{2}$.

Proof. A well-known statistical theorem (refer to [7, 8]).
Theorem 4.1. Under Assumption 3.1, we have approximately $-S_{i}^{s}+N_{i}^{s} \sim$ $\mathbf{N}(-(n-1) / 2,(n-1)(2 m-1) / 4)$.

Proof. By Proposition $4.1\left(-S_{i}^{s} \sim \mathbf{N}(-(n-1) / 2,(n-1) / 4)\right)$, Proposition $4.4\left(N_{i}^{s} \sim \mathbf{N}(0,(n-1)(m-1) / 2)\right)$, Proposition 4.5, and Proposition 4.6.

Remark. With the assumption that $n$ is fixed and $m$ is large, McEliece claimed that $-S_{i}^{s}+N_{i}^{s} \sim \mathbf{N}(-(n-1) / 2,(n-1)(m-1) / 2)$. In other words, the effect of the variance of $S_{i}^{s}$ (i.e., $\left.(n-1) / 4 \ll(n-1)(m-1) / 2\right)$ is neglected.

Corollary 4.1. The probability that a neuron satisfies the invariance condition is

$$
\begin{aligned}
\mathbf{P}\left(-S_{i}^{s}+N_{i}^{s}<0\right) & =\Phi(\sqrt{(n-1)} / \sqrt{2 m-1}) \\
& =\int_{-\infty}^{\sqrt{(n-1)} / \sqrt{2 m-1}}(1 / \sqrt{2 \pi}) \cdot e^{-t^{2} / 2} d t
\end{aligned}
$$

for $-\infty<t<\infty$.

Now we need to answer the questions what is the probability that $n$ neurons in a stcred pattern satisfy the invariance condition and, more generally, what is the probability that $n m$ neurons of $m$ stored patterns satisfy the invariance condition. To answer these questions, first we check the dependency among the terms $-S_{i}^{s}+N_{i}^{s}$ for $n$ neurons in a stored pattern. Then we check the dependency among the terms $-S_{i}^{s}+N_{i}^{s}$ for the $i$ th $(1 \leq i \leq n)$ neurons in $m$ stored patterns.

Proposition 4.7. $\operatorname{Cov}(\mathbf{A}(1, s), \mathbf{A}(2, s))=(n+m-3) / 4, \forall s$ where $1 \leq s \leq$ $m$.

Proof. Note that

$$
\begin{aligned}
\operatorname{Cov}\left(S_{1}^{s}, S_{2}^{s}\right) & =\operatorname{Cov}\left(\sum_{j \neq 1, j=2}^{n} v_{j}^{s}, \sum_{j=1, j \neq 2}^{n} v_{j}^{s}\right) \\
& =\sum_{j_{1} \neq 1, j_{1}=2}^{n} \sum_{j_{2}=1, j_{2} \neq 2}^{n} \operatorname{Cov}\left(v_{j_{1}}^{s}, v_{j_{2}}^{s}\right) \\
& =(n-2) / 4
\end{aligned}
$$

and by the same arguments used in Proposition 4.5,

$$
\operatorname{Cov}\left(S_{1}^{s}, N_{2}^{s}\right)=\operatorname{Cov}\left(S_{2}^{s}, N_{1}^{s}\right)=0
$$

Also,

$$
\begin{aligned}
& \operatorname{Cov}\left(N_{1}^{s}, N_{2}^{s}\right) \\
& =\operatorname{Cov}\left(\sum_{\substack{s^{\prime}=1, s^{\prime} \neq s}}^{m} \sum_{\substack{j=2, j \neq 1}}^{n} X_{1}^{s^{\prime}} \cdot X_{j}^{s^{\prime}} \cdot v_{j}^{s}, \sum_{\substack{s^{\prime}=1, s^{\prime} \neq s}}^{m} \sum_{\substack{j=1, j \neq 2}}^{n} X_{2}^{s^{\prime}} \cdot X_{j}^{s^{\prime}} \cdot v_{j}^{s}\right) \\
& =\sum_{\substack{s_{1}^{\prime}=1, s_{1}^{\prime} \neq s}}^{m} \sum_{\substack{s_{2}^{\prime}=1, s_{2}^{\prime} \neq s}}^{m} \sum_{\substack{j_{1} \neq 1,1, j_{1}=2}}^{n} \sum_{j_{2}=1,}^{n} \operatorname{Cov}\left(X_{1}^{s_{1}^{\prime}} \cdot X_{j_{1} \neq 2}^{s_{1}^{\prime}} \cdot v_{j_{1}}^{s}, X_{2}^{s_{2}^{\prime}} \cdot X_{j_{2}}^{s_{2}^{\prime}} \cdot v_{j_{2}}^{s}\right) .
\end{aligned}
$$

By Proposition 4.2, the terms of Cov will be 0 except in the case $s_{1}^{\prime}=s_{2}^{\prime}$ and $j_{1}=2, j_{2}=1$.

$$
\begin{aligned}
& =\sum_{s^{\prime}=1, s^{\prime} \neq s}^{m} \operatorname{Cov}\left(X_{1}^{s^{\prime}} \cdot X_{2}^{s^{\prime}} \cdot v_{2}^{s}, X_{2}^{s^{\prime}} \cdot X_{1}^{s^{\prime}} \cdot v_{1}^{s}\right) \\
& =(m-1) \mathbf{E}\left(X_{1}^{s^{\prime}} \cdot X_{2}^{s^{\prime}} \cdot v_{2}^{s} \cdot X_{2}^{s^{\prime}} \cdot X_{1}^{s^{\prime}} \cdot v_{1}^{s}\right) \\
& =(m-1) \mathbf{E}\left(v_{1}^{s} \cdot v_{2}^{s}\right) \\
& =(m-1) \mathbf{E}\left(v_{1}^{s}\right) \cdot \mathbf{E}\left(v_{2}^{s}\right)=(m-1) / 4 .
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
\operatorname{Cov}(\mathbf{A}(1, s), \mathbf{A}(2, s)) & =(n-2) / 4+0+(m-1) / 4 \\
& =(n+m-3) / 4
\end{aligned}
$$

Theorem 4.2. The correlation coefficient $\rho_{A(1, s), A(2, s)}$ of $\mathbf{A}(1, s)$ and $\mathbf{A}(2, s)$ is $(n+m-3) /((n-1)(2 m-1))$ for $1 \leq s \leq m$.

Proof. By Theorem 4.1, $\sigma_{A(1, s)}^{2}=\sigma_{A(2, s)}^{2}=(n-1)(2 m-1) / 4$, and by Proposition 4.7, $\operatorname{Cov}(\mathbf{A}(1, s), \mathbf{A}(2, s))=(n+m-3) / 4$. Hence, $\rho_{A(1, s), A(2, s)}=$ $\operatorname{Cov}(\mathbf{A}(1, s), \mathbf{A}(2, s)) /\left(\sigma_{A(1, s)} \cdot \sigma_{A(2, s)}\right)=(n+m-3) /((n-1)(2 m-1))$.

Remark. $\lim _{n, m \rightarrow \infty} \rho_{A(1, s), A(2, s)}=0$. Hence, for large $n$ and $m$, we may be able to assume that there is no dependency among the terms $-S_{i}^{s}+N_{i}^{s}$ for $n$ neurons within a stored pattern.

Proposition 4.8. Under Assumption 3.1, we have $\operatorname{Cov}(\mathbf{A}(i, 1), \mathbf{A}(i, 2))=$ $(n-1)(m-2) / 4$ for $1 \leq i \leq n$.

Proof. Note that, by Assumption 3.1,

$$
\begin{aligned}
\operatorname{Cov}\left(S_{i}^{1}, S_{i}^{2}\right) & =\operatorname{Cov}\left(\sum_{j=1, j \neq i}^{n} v_{j}^{1}, \sum_{j=1, j \neq i}^{n} v_{j}^{2}\right) \\
& =\sum_{j_{1}=1, j_{1} \neq i}^{n} \sum_{j_{2}=1, j_{2} \neq i}^{n} \operatorname{Cov}\left(v_{j_{1}}^{1}, v_{j_{2}}^{2}\right)=0,
\end{aligned}
$$

and, by the same arguments used in Proposition 4.5,

$$
\operatorname{Cov}\left(S_{i}^{1}, N_{i}^{2}\right)=\operatorname{Cov}\left(S_{i}^{2}, N_{i}^{1}\right)=0
$$

Also,

$$
\begin{aligned}
& \operatorname{Cov}\left(N_{i}^{1}, N_{i}^{2}\right) \\
& =\operatorname{Cov}\left(\sum_{\substack{s^{\prime} \neq 1, s^{\prime}=2}}^{m} \sum_{\substack{j=1, j \neq i}}^{n} X_{i}^{s^{\prime}} \cdot X_{j}^{s^{\prime}} \cdot v_{j}^{1}, \sum_{\substack{s^{\prime}=1, s^{\prime} \neq 2}}^{m} \sum_{\substack{j=1, j \neq i}}^{n} X_{i}^{s^{\prime}} \cdot X_{j}^{s^{\prime}} \cdot v_{j}^{2}\right) \\
& =\sum_{\substack{s_{1}^{\prime} \neq 1, s_{1}^{\prime}=2}}^{m} \sum_{\substack{s_{2}^{\prime}=1, s_{2}^{\prime} \neq 2}}^{m} \sum_{\substack{j_{1}=1, j_{1} \neq i}}^{n} \sum_{j_{2}=1,} \operatorname{Cov}\left(X_{i}^{s_{1}^{\prime}} \cdot X_{j_{1} \neq i}^{s_{1}^{\prime}} \cdot v_{j_{1}}^{1}, X_{i}^{s_{2}^{\prime}} \cdot X_{j_{2}}^{s_{2}^{\prime}} \cdot v_{j_{2}}^{2}\right),
\end{aligned}
$$

and because the terms of Cov will be 0 except in the case $\left(s_{1}^{\prime}=s_{2}^{\prime}=s\right.$, $j_{1}=j_{2}=j$ ), by Proposition 4.2,

$$
\begin{aligned}
& =\sum_{s^{\prime} \neq 1,2, s^{\prime}=3}^{m} \sum_{j=1, j \neq i}^{n} \mathbf{E}\left(X_{i}^{s^{\prime}} \cdot X_{j}^{s^{\prime}} \cdot v_{j}^{1} \cdot X_{i}^{s^{\prime}} \cdot X_{j}^{s^{\prime}} \cdot v_{j}^{2}\right) \\
& =\sum_{s^{\prime}=3}^{m} \sum_{j=1, j \neq i}^{n} \mathbf{E}\left(v_{j}^{1} \cdot v_{j}^{2}\right)=(n-1)(m-2) / 4
\end{aligned}
$$

Therefore,

$$
\operatorname{Cov}(\mathbf{A}(i, 1), \mathbf{A}(i, 2))=(n-1)(m-2) / 4
$$

Theorem 4.3. The correlation coefficient $\rho_{A(i, 1), A(i, 2)}$ of $\mathbf{A}(i, 1)$ and $\mathbf{A}(i, 2)$ is $(m-2) /(2 m-1)$, for $1 \leq i \leq n$.

Proof. By the same arguments as in the proof of Theorem 4.2.

Remark. $\lim _{m \rightarrow \infty} \rho_{A(i, 1), A(i, 2)}=1 / 2$. Hence, the dependency among the terms $-S_{i}^{s}+N_{i}^{s}$ for the $i$ th neurons in $m$ stored patterns cannot be neglected (recall Assumption 3.3 and Amari's note).

Now, those dependencies both among the random variables (i.e., $-S_{i}^{s}+$ $N_{i}^{s}$ ) for the invariance condition of $n$ neurons in a stored pattern and among the random variables for the invariance condition of the $i$ th neurons in $m$ stored patterns should be considered to derive a statistical expression for the information capacity.

Because the former dependency is negligible by the remark following Theorem 4.2, we derive a statistical expression considering only the latter dependency. From Theorem 4.1, Proposition 4.8, and Theorem 4.3, we can derive the following:

$$
\left(\begin{array}{c}
U_{1} \\
\vdots \\
\vdots \\
U_{m}
\end{array}\right) \sim \mathbf{N}\left(\left(\begin{array}{c}
0 \\
\vdots \\
\vdots \\
0
\end{array}\right)\left(\begin{array}{ccccc}
1 & \rho & \rho & \ldots & \rho \\
\rho & 1 & \rho & \ldots & \rho \\
\vdots & & \ddots & & \vdots \\
\rho & \rho & \ldots & & 1
\end{array}\right)\right)
$$

where

$$
\begin{aligned}
\sigma_{A(i, 1)}^{2} & =\sigma_{A(i, 2)}^{2}=(n-1)(2 m-1) / 4 \\
U_{j} & =(A(i, j)+(n-1) / 2) / \sqrt{(n-1)(2 m-1) / 4} \\
\rho & =\operatorname{Cov}(A(i, 1), A(i, 2)) /\left(\sigma_{A(i, 1)} \cdot \sigma_{A(i, 2)}\right)=(m-2) /(2 m-1)
\end{aligned}
$$

Remark. For a fixed $m$ and $\forall\left(\lambda_{1}, \ldots, \lambda_{m}\right) \in \mathbf{R}^{m},\left(\lambda_{1}, \ldots, \lambda_{m}\right) \times \mathbf{Q}^{-1 / 2} \times$ $\left[U_{1}, \ldots, U_{m}\right]$ approximately follows $\mathbf{N}\left(0,\left[\lambda_{1}^{2}+\cdots+\lambda_{m}^{2}\right]\right)$ as $n \rightarrow \infty$ where () and [] mean a row vector and a column vector, respectively, and $\mathbf{Q}$ is the variance-covariance matrix of $U_{1}, \ldots, U_{m}$. Hence, by Cramer-Wold device ([7] page 397 ), $U_{j}$ 's are approximately jointly normal.

Theorem 4.4. The probability (i.e., $\left.\left\{\mathbf{P}\left(U_{1}<h, \ldots, U_{m}<h\right)\right\}^{n}\right)$ that all of the $m$ stored patterns are stable is

$$
\begin{equation*}
\left[\int_{-\infty}^{\infty} \Phi^{m}\left(\frac{h-\sqrt{\rho} \cdot z_{0}}{\sqrt{1-\rho}}\right) \cdot \frac{1}{\sqrt{2 \pi}} \cdot e^{-z_{0}^{2} / 2} d z_{0}\right]^{n} \tag{4.3}
\end{equation*}
$$

where $h=((n-1) / 2) / \sqrt{\sigma_{A(i, 1)}}$, and where $\rho$ and $U_{j}$ 's are defined as before. $\Phi$ denotes the standard normal CDF.

Proof. Let $\left(Z_{0}, Z_{1}, \ldots, Z_{m}\right)$ be i.i.d. $\mathbf{N}(0,1)$ variables. In the following transformation [12],

$$
\begin{gathered}
Y_{1}=\sqrt{\rho} Z_{0}+\sqrt{1-\rho} Z_{1}, \\
Y_{2}=\sqrt{\rho} Z_{0}+\sqrt{1-\rho} Z_{2}, \\
\vdots \\
Y_{m}=\sqrt{\rho} Z_{0}+\sqrt{1-\rho} Z_{m},
\end{gathered}
$$

$\mathbf{E}\left(Y_{i}\right)=0, \operatorname{Var}\left(Y_{i}\right)=\rho+(1-\rho)=1$, and $\operatorname{Cov}\left(Y_{1}, Y_{2}\right)=\rho$. Since the mean and the covariance of $\left(Y_{1}, \ldots, Y_{m}\right)$ are the same as the mean and the covariance of ( $U_{1}, \ldots, U_{m}$ ), and both have multivariate normal distribution, the distribution of $\left(U_{1}, \ldots, U_{m}\right)$ and $\left(Y_{1}, \ldots, Y_{m}\right)$ are identical. Therefore,

$$
\begin{aligned}
& \mathbf{P}\left(U_{1}<h, \ldots, U_{m}<h\right) \\
&=\mathbf{P}\left(Y_{1}<h, \ldots, Y_{m}<h\right) \\
& \quad=\mathbf{P}\left(\sqrt{\rho} Z_{0}+\sqrt{1-\rho} Z_{1}<h, \ldots, \sqrt{\rho} Z_{0}+\sqrt{1-\rho} Z_{m}<h\right) \\
& \quad=\mathbf{P}\left(Z_{1}<\left(h-\sqrt{\rho} Z_{0}\right) / \sqrt{1-\rho}, \ldots, Z_{m}<\left(h-\sqrt{\rho} Z_{0}\right) / \sqrt{1-\rho}\right) \\
& \quad=\int_{-\infty}^{\infty} \Phi^{m}\left(\left(h-\sqrt{\rho} \cdot z_{0}\right) / \sqrt{1-\rho}\right) \cdot 1 / \sqrt{2 \pi} \cdot e^{-z_{0}^{2} / 2} d z_{0} .
\end{aligned}
$$

## 5. Results

In this section we present the numerical results based on our theoretical work and compare them with the results of other theoretical studies (Amari, Hopfield, McEliece, and Weisbuch). The simulation results of Weisbuch [22] are used as criteria for the comparison.

Weisbuch's simulation results are based on testing $m n$ inequalities (invariance conditions, equation (3.1) or (3.2)) per network, and adjusting $m$ so that the probability is 0.5 for all inequalities to be verified. We and Weisbuch study the information capacity under the condition that the probability that all of the $m$ stored patterns are stable is 0.5 (i.e., $\mathbf{I}_{0.5}$ in Definition 2.4). On the other hand, Amari, Hopfield, and McEliece use the condition that most of $m$ stored patterns are stable, but the definition of most is left ambiguous in their studies. The comparison among Weisbuch's simulation study and the various theoretical studies is summarized in table 5.1.

The theoretical results of Hopfield [13], Amari [4], and McEliece [19] overestimate the information capacity in comparison to the simulation study of Weisbuch [22]. Our results and Weisbuch's theoretical results (using Assumptions 3.2 and 3.3 ) are close to the results of the simulation. The overestimation occurs because the studies $[4,5,13,19]$ are based on the condition that a stored pattern is stable, instead of on the stability of $m$ patterns.

| Methods | $m=$ number of patterns |  |  |  |  |  |
| :--- | ---: | ---: | ---: | ---: | ---: | ---: |
|  | 8 | 10 | 14 | 20 | 1,000 | 50,000 |
| Hopfield | 53 | 67 | 93 | 133 | 6,667 | 333,333 |
| Amari | 20 | 28 | 43 | 69 | 7,120 | 534,939 |
| McEliece | 21 | 30 | 47 | 76 | 7,783 | 576,046 |
| Weisbuch <br> (with assumptions) | 149 | 210 | 343 | 562 | 62,528 | $4,737,757$ |
| LKS (ours) | 146 | 206 | 337 | 552 | 61,842 | $4,698,517$ |
| Weisbuch's <br> simulation | 103 | 180 | 290 | 500 | N/A | N/A |

Table 1: The required number of neurons $n$ for storing $m$ patterns.

## 6. Conclusion

We have established a connection between the dynamics of the Hopfield Neural Network (HNN) and the theory of multivariate normal distribution, and use the connection to derive results about information capacity of the HNN. The current information capacity heuristics due to Amari, Hopfield, McEliece, and Weisbuch are carefully reviewed. These heuristics are inexact because they depend on simplifying assumptions that are not completely correct. We have been successful in eliminating the use of these assumptions through an exact analysis using powerful mathematical techniques [9, 12].

Based on our analysis and using the statistical software package IMSL [20], we derive numerical results for the information capacity of the Hopfield network. Our results are in close agreement with the results of Weisbuch's simulation [22]. We would have liked to do a more exhaustive simulation study including neural networks with a large number of neurons. However, such a study is currently infeasible because of prohibitively high computational requirement.

## Appendix A. Brown's Martingale Central Limit Theorem and Proposition 4.4

Definition. Let us define $S_{n}$ as $\sum_{k=1}^{n} Z_{k}$. Then $\left\{S_{n}\right\}_{n=1}^{\infty}$ is a Martingale with respect to $\left\{\vec{V}_{n}\right\}_{n=1}^{\infty}$ if $\mathbf{E}\left|S_{n}\right|<\infty$ for all $n$ and $\mathbf{E}\left(S_{n+1} \mid \vec{V}_{1}, \ldots, \vec{V}_{n}\right)=S_{n}$, where $\left\{Z_{n}\right\}$ is a sequence of random variables and $\left\{\vec{V}_{n}\right\}_{n=1}^{\infty}$ is a sequence of random vectors such that $\vec{V}_{n}=\left[X_{n}^{1}, \ldots, X_{n}^{m}\right]$ for a fixed $m$.

Brown's Martingale Central Limit Theorem. Assume that $\left\{S_{n}\right\}_{n=1}^{\infty}$ is a Martingale with respect to $\left\{\vec{V}_{n}\right\}_{n=1}^{\infty}$. Let $\delta_{n}^{2}=\mathbf{E}\left(Z_{n}^{2} \mid \vec{V}_{1}, \ldots, \vec{V}_{n-1}\right)$, $U_{n}^{2}=\sum_{i=1}^{n} \delta_{i}^{2}$, and $\&_{n}^{2}=\mathbf{E}\left(U_{n}^{2}\right)$. If (1) $U_{n}^{2} / \&_{n}^{2} \rightarrow 1$ in probability, and (2) $\lim _{n \rightarrow \infty}\left(\sum_{k=1}^{n} \mathbf{E}\left|Z_{k}\right|^{4}\right) / \&_{n}^{4}=0$, then $S_{n} / \&_{n} \Rightarrow \mathbf{N}(0,1)$.

Theorem. Under Assumption 3.1, approximately

$$
N_{i}^{s}(n, m) \sim \mathbf{N}\left(0, \frac{(m-1)(n-1)}{2}\right)
$$

This is the detailed proof of Proposition 4.4.
Proof. Fix $m$ and let (1) $Z_{n}=\sum_{s^{\prime}=2}^{m} X_{1}^{s^{\prime}} \cdot X_{n}^{s^{\prime}} \cdot v_{n}^{1}$, (2) $S_{n}=\sum_{k=2}^{n} Z_{k}$, (3) $\vec{V}_{n}=\left[X_{n}^{1}, \ldots, X_{n}^{m}\right]$, (4) $\delta_{n}^{2}=\mathrm{E}\left(Z_{n}^{2} \mid \vec{V}_{2}, \ldots, \vec{V}_{n-1}\right)$, (5) $U_{n}^{2}=\sum_{i=1}^{n} \delta_{i}^{2}$, and (6) $\&_{n}^{2}=\mathbf{E}\left(U_{n}^{2}\right)$ for $n=2,3, \ldots$ Because $\mathbf{E}\left(S_{n}\right)<\infty$, and

$$
\begin{aligned}
& \mathbf{E}\left(S_{n+1} \mid \vec{V}_{1}, \ldots, \vec{V}_{n}\right) \\
& =\sum_{j=2}^{n} \sum_{s^{\prime}=2}^{m} X_{1}^{s^{\prime}} \cdot X_{j}^{s^{\prime}} \cdot v_{j}^{1} \\
& \quad+X_{1}^{s^{\prime}} \mathbf{E}\left(\sum_{s^{\prime}=2}^{m} X_{n+1}^{s^{\prime}} \cdot v_{n+1}^{1} \mid\left[X_{j}^{1}, \ldots, X_{j}^{m}\right], j=1, \ldots, n\right) \\
& =\sum_{j=2}^{n} \sum_{s^{\prime}=2}^{m} X_{1}^{s^{\prime}} \cdot X_{j}^{s^{\prime}} \cdot v_{j}^{1}=S_{n} \quad \text { for } n=2,3, \ldots,
\end{aligned}
$$

$\left\{S_{n}\right\}_{n=2}^{\infty}$ is a Martingale with respect to $\left\{\vec{V}_{n}\right\}_{n=2}^{\infty}$. And,

$$
\begin{aligned}
\delta_{n}^{2}= & \mathbf{E}\left(Z_{n}^{2} \mid\left[X_{j}^{1}, \ldots, X_{j}^{m}\right], j=2, \ldots, n-1\right) \\
= & \sum_{s_{1}=2}^{m} \sum_{s_{2}=2}^{m} \mathbf{E}\left(X_{1}^{s_{1}} \cdot X_{1}^{s_{2}} \cdot X_{n}^{s_{1}} \cdot X_{n}^{s_{2}} \cdot\left(v_{n}^{1}\right)^{2} \mid\right. \\
& \left.\quad\left[X_{j}^{1}, \ldots, X_{j}^{m}\right], j=2, \ldots, n-1\right) \\
= & \sum_{s=2}^{m}\left(X_{1}^{s}\right)^{2} \cdot \frac{1}{2}=\frac{(m-1)}{2}
\end{aligned}
$$

Also,

$$
U_{n}^{2}=\sum_{k=2}^{n} \delta_{k}^{2}=\frac{(n-1)(m-1)}{2}
$$

Therefore,

$$
\&_{n}^{2}=\mathbf{E}\left(U_{n}^{2}\right)=\frac{(n-1)(m-1)}{2}
$$

and

$$
\frac{U_{n}^{2}}{\&_{n}^{2}}=1
$$

In addition to that,

$$
\lim _{n \rightarrow \infty} \frac{n \mathbf{E}\left|Z_{2}\right|^{4}}{(n-1)^{2}(m-1)^{2} / 4}=0
$$

Hence, by the Martingale Central Limit Theorem,

$$
\frac{S_{n}}{\&_{n}}=\frac{\sum_{j=2}^{n} \sum_{s=2}^{m} X_{1}^{s} \cdot X_{j}^{s} \cdot v_{j}^{1}}{\sqrt{(n-1)(m-1) / 2}}=\frac{N_{i}^{s}(n, m)}{\sqrt{(n-1)(m-1) / 2}} \Rightarrow \mathbf{N}(0,1)
$$

## References

[1] Y. S. Abu-Mostafa, "Neural Networks for Computing," AIP Conference Proceedings, Snowbird, UT, 151 (1986) 1-6.
[2] Y. S. Abu-Mostafa, "Connectivity versus Entropy," pages 1-8 in Neural Information Processing Systems (New York, American Institute of Physics, 1988).
[3] Y. S. Abu-Mostafa and J. St. Jacques, "Information Capacity of the Hopfield Model," IEEE Transactios on Information Theory, 31(4) (1985) 461-464.
[4] S. Amari, "On Mathematical Methods in the Theory of Neural Networks," IEEE Proceedings of the First International Conference on Neural Networks, 3 (1987) 3-10.
[5] S. Amari, K. Yoshida, and K. Kanatani, "A Mathematical Foundation for Statistical Neurodynamics," SIAM Journal on Applied Mathematics, 33(1) (1977) 95-126.
[6] D. J. Amit, Modeling Brain Function (New York, Cambridge University Press, 1989.)
[7] P. Billingsley, Probability and Measure (New York, John Wiley, 1986.)
[8] L. J. Brain, Introduction to Probability and Mathematical Statistics (Boston, MA, Duxbury Press, 1987.)
[9] B. B. Brown, "Martingale Central Limit Theorems," Annals of Mathematical Statistics, 42(1) (1971) 59-66.
[10] M. Cottell, "Stability and Attractivity in Associative Memory Networks," Biological Cybernetics, 58 (1988) 129-139.
[11] E. Gardner, "The Phase Space of Interactions in Neural Network Models," Journal of Physics A, 21 (1988) 257-270.
[12] S. S. Gupta, "Probability Integrals of Multivariate Normal and Multivariate $t$," Annals of Mathematical Statistics, 34(3) (1963) 792-828.
[13] J. J. Hopfield, "Neural Networks and Physical Systems with Emergent Collective Computational Abilities," Proceedings of the National Academy of Sciences, 79 (1982) 2554-2558.
[14] J. J. Hopfield, "Neurons with Graded Response Have Collective Computational Properties Like Those of Two-state Neurons," Proceedings of the National Academy of Sciences, 81 (1984) 3088-3092.
[15] J. J. Hopfield and D. W. Tank, "Neural Computation of Decision in Optimization Problems," Biological Cybernetics, 52 (1985) 141-152.
[16] Y. Kamp and M. Hasler, Recursive Networks for Associative Memory (West Sussex, England, John Wiley, 1990.)
[17] J. D. Keeler, "Basins of Attraction of Neural Network Model," AIP Conference Proceedings, Snowbird, UT, 151 (1986) 259-264.
[18] A. Krowitz, L. Rendell, and B. Hohensee, "The State Space of Memory Clusters in the Hopfield Network," IEEE Proceedings of the First International Conference on Neural Networks, 3 (1987) 339-346.
[19] R. J. McEliece, E. C. Posner, E. R. Rodemich, and S. S. Venkatesh, "The Capacity of the Hopfield Associative Memory," IEEE Transactions on Information Theory, 33(4) (1987) 461-482.
[20] Problem-Solving Software System, IMSL User's Manual Math/Library Fortran Subroutines for Mathematical Applications, Version 1.0 (1987).
[21] D. E. Rumelhart, J. L. McClelland, and PDP Research Group, Parallel Distributed Processing: Exploration in the Microstructure of Cognition: Vol. 1, Foundations (Cambridge, MA, MIT Press, 1986).
[22] G. Weisbuch, "Scaling Laws for the Attractors of Hopfield Networks," Journal de Physique Lettres, 46 (1985) L623-L630.
[23] G. Weisbuch and D. D'Humieres, "Determining the Dynamic Landscape of Hopfield Networks," NATO ASI Series, Disordered Systems and Biological Organization, F20 (1986) 187-191.


[^0]:    *Current address: R \& D Center, Consumer Electronics Business, Samsung Electronics, 416 Maetan-3 Dong, Kwonsun Gu, Suwon City, Kyungi-Do Korea
    ${ }^{\dagger}$ Current address: University of Suwon, Do Korea

