Two-level Decoupled Hamming Network for Associative Memory Under Noisy Environment

Liang Chen
Computer Science Department
Univerisy of Northern British Columbia
Prince George, BC, Canada V2N 4Z9
E-mail: lchen@ieee.org

Naoyuki Tokuda
SunFlare Company
Shinjuku Hirose Bldg, Yotsuya 4-7, Shinjuku-ku
Tokyo, Japan 160-0004
Email: tokuda_n@sunflare.co.jp

Akira Nagai
Advanced Media Network Center
Utsunomiya University
Utsunomiya, Tochigi, Japan 321-8585
Email:anagai@cc.utsunomiya-u.ac.jp

***Abstract*— Compared with a single level Hamming associative memory, a simple model based on uniform random noise analysis has proved a two-level decoupled Hamming network to be an efficient associative memory but with a lower capacity against uniform random noise. Introducing concentrated noise, our more detailed analysis shows that the deteriorated capacity of the two-level Hamming associative memory against uniform random noise is actually minor, but that it possesses a substantially increased capacity against concentrated noise which is more prevalent in practice than uniform random noise.**

I. Introduction

An associative memory is a memory that is addressed through its contents. When an input pattern, called a memory key, is presented, the associative memory should return a stored pattern coincident with the key. The behavior of an associative memory is very close to our human behavior in remembering something in the presence of a sensory cue. The coincidence between a memory key and a stored pattern need not and should not be perfect. The associate memory should be able to recall a stored pattern that is similar to the memory key such that noise polluted inputs can also be recognized.

Many neural models have been proposed for associative memories [1], [2], [3], [4]; but it seems that there is no practical package available up to now. Most of these models suffer from serious problems including complicated hardware designing and a huge number of interconnections [5].

Most recently, Ikeda and et.al[5] has presented a two-level decoupled Hamming memory, which is generalized from a single level Hamming associative memory. They showed that the two-level decoupled Hamming memory retains a high performance and allows for a simpler hardware implementation. But it has a lower capacity for tolerating uniform random noise in comparing with a single level Hamming associative memory[5]. [1]

This paper follows the basic assumptions of [5] on the uniform distribution of memory patterns and on the uniform distribution of random noise that corrupts a memory pattern into a memory key to be recognized. We further assume the presence of so-called concentrated noise, which corrupts parts of memory keys. We show that the deteriorated capacity of a two-level decoupled Hamming memory against uniform random noise of citeikeda is actually minor when errors on the independence of the random variables are corrected, and that this slightly deteriorated capacity is more than compensated by its substantially increased capacity against concentrated noise which is more prevalent in practice than uniform random noise.

[1]The derivation in [5] actually has minor errors. It claims that the random variables representing the numbers of votes received by different memories are independent. This is wrong as we can see this by a simple example where the memory set consists of only two memory patterns. When corrected, we show the deteriorated capacity against random noise turn out to be smaller.

II. Hamming Associative Memory, Two-level Hamming Memory and Basic Assumption on Memory Patterns

A. Hamming Associative Memory

Throughout this paper, we follow the definition of [5], discussing only the auto-associative memory for 2-D patterns. In this case, each memory pattern $\mathbf{X}^i_{M_1\times M_2}$ is an integer array of size $M_1 \times M_2$. Here $i \in \{1, 2, \cdots, m\}$ is the memory pattern index. A Hamming associative memory returns a memory pattern $\mathbf{X}^k_{M_1\times M_2}$ for an input memory key $\mathbf{Y}_{M_1\times M_2}$ if:

$$d(\mathbf{X}^k_{M_1\times M_2}, \mathbf{Y}_{M_1\times M_2}) < d(\mathbf{X}^i_{M_1\times M_2}, \mathbf{Y}_{M_1\times M_2})$$

for all $i \neq k$, where $d(\cdot,\cdot)$ is defined as the city block distance between two arrays.

To simplify the analysis, we follow [5]and only consider binary memory patterns, that is, each memory pattern is a 0-1 array. In this case, the city block distance actually reduces to the Hamming distance.

B. Two-Level Decoupled Hamming Memory

A two-level decoupled Hamming memory is set up by partitioning each of the memory patterns into a set of sub-memory patterns, and associating the corresponding sub-memory patterns of all the memory patterns with a local memory set. The decoupled Hamming memory is operated as follows: it firstly partitions the memory key into sub-memory keys in the same fashion as partitioning memory patterns, calculates the City-block distances locally between the sub-memory key and each sub-memory pattern in the corresponding local memory set to decide the closest sub-memory pattern to the sub-memory key in each local memory set, then returns a “closest” memory key by employing a voting mechanism on the results obtained from all local memory sets.

Formally, we always suppose M_1 and M_2 are divisible by R_1 and R_2 respectively, we partition each $\mathbf{X}^i_{M_1\times M_2}$ ($1 \le i \le m$) into a set of sub-memory patterns $\mathbf{X}^{i,1,1}_{R_1\times R_2}$, $\mathbf{X}^{i,1,2}_{R_1\times R_2}$, $\cdots$, $\mathbf{X}^{i,M_1/R_1,M_2/R_2}_{R_1\times R_2}$, and set up a local memory set $\{\mathbf{X}^{1,j,t}_{R_1\times R_2}, \mathbf{X}^{2,j,t}_{R_1\times R_2}, \cdots, \mathbf{X}^{m,j,t}_{R_1\times R_2}\}$ for each tuple (j,t) ($1 \le j \le M_1/R_1$ and $1 \le j \le M_2/R_2$). Then for any input memory key $\mathbf{Y}$, we partition it into $\{\mathbf{Y}^{1,1}_{R_1\times R_2}$, $\mathbf{Y}^{1,2}_{R_1\times R_2}$, $\cdots$, $\mathbf{Y}^{M_1/R_1,M_2/R_2}_{R_1\times R_2}\}$, and calculate the City-block distances between the sub-memory key $\mathbf{Y}^{j,t}_{M_1/R_1\times M_2/R_2}$ and the sub-memory patterns $\mathbf{X}^{i,j,t}_{M_1/R_1\times M_2/R_2}$, $1 \le i \le m$ to find the index of the closest sub-memory patterns to the sub-memory key. Finally we vote on the locally determined sub-memory pattern indexes using a simple majority scheme to return a memory pattern.

C. Basic Assumption on the Memory Patterns

We follow the basic assumptions of [5]; each memory pattern is generated by a uniform 0.5 probability of each bit being 1 or 0;with the memory key being a corrupted version of one of the memory patterns, named target memory, when subjected to certain noise.

We have many types of noise, of course. Basically, two types of noise have been treated in stability analysis literature [6], namely uniform random noise and concentrated noise. As in [5], we assume that a set of uniform random noise is capable of corrupting the target memory pattern into a memory key, by changing the value of each bit of a memory pattern by a probability of p. We further assume that, concentrated noise, on the other hand, appears in blocks and changes the values of all bits polluted into the opposite values (i.e., from 1 to 0, and 0 to 1) [2]. Formally, concentrated noise is here defined as a union of non-overlapped blocks of noise corrupting blocks of the two dimensional target memory pattern into the memory key, such that each bit on each corrupted block of the target memory pattern flips its value. To simplify the analysis, we assume each concentrated noise block is able to occupy as many sub-memory patterns as possible. We will pay particular attention to the sub-memory patterns which are "covered" by a noise block in its entirety which means in the analysis of the two-level decoupled Hamming memory all the bits in the sub-memory patterns flip their values. We call these sub-memory patterns as fully polluted sub-memory patterns by concentrated noise. For the parts of noise blocks that are not able to occupy an entire sub-memory pattern, we suppose that they distributed averagely over all the sub-memory patterns which are not fully polluted by concentrated noise.

The amount of uniform random noise could be measured by pM_1M_2, while the amount of concentrated noise by the total size S of noise blocks, each block is assumed to be of size $N_1 \times N_2$.

III. Capacity Analysis

To simplify the analysis, we only consider the case of $m = 2$. For $m > 2$, to determine the closest pattern to be returned with an input memory key, we use a "pair-wise best" scheme. That is, a memory pattern is returned if and only if it is independently "closer" to the memory key than any other memory pattern. For single level Hamming associative memories, the "pair-wise" best (closest) is always the best (closest) of all. However, this is not always true for a two-level decoupled Hamming associative memory. A simple example is given below: Suppose we partition each of the memory patterns into 7 sub-memory patterns, and associate the corresponding sub-memory patterns of all the memory patterns with a local memory set so that there are 7 local memory sets, namely L_1, L_2, $\cdots$, L_7. In L_1, L_2, the sub-memories of memory pattern *C* are "closer" to memory key than those of *A*, and the sub-memories of memory pattern *A* are "closer" to the memory key than those of *B*; in local memory sets L_3 and L_4, the sub-memories of memory pattern *A* are "closer" to the memory key than those of *B*, and the sub-memories of memory pattern *B* are "closer" to memory key than those of *C*; in the remaining 3 local memory sets, the sub-memories of memory pattern *B* are "closer" to memory key than those of *A*, and the sub-memories of memory pattern *A* are "closer" to memory key than those of *C*. In this example, the pair-wise best is memory pattern *A* while the best of all be memory pattern *B*.

Although a pair-wise best is not always the "best of all", there is no reason to believe "the best of all" is more reasonable. Actually, a pair-wise best satisfies a major criterion in voting theory [7], namely independence of irrelevant alternatives criterion, which means that societal preference of two alternatives should be independent of preferences for other options, or we could say that the winning candidate should remain to be the winner if one or more of the losing candidates drop out.

We would like to remind the reader here that we always use $\Phi(x)$ to denote $\dfrac{1}{2\pi}\displaystyle\int_{-\infty}^{x} e^{-\frac{y^2}{2}}dy$, which is the distribution function of a standard normal distribution. In the following discussion, we always assume M_1M_2 is a large number.

A. Capacity of One-Level Hamming Associative Memory

Theorem 1: The probability that a 1-level associative memory returns the target memory is:

$$\text{Prob}_{\text{1-level}} = \sum_{j=S}^{M_1M_2} \binom{M_1M_2 - S}{j - S} p^{j-S}(1-p)^{M_1M_2-j} \cdot \Phi\left(\frac{j - 0.5M_1M_2}{\sqrt{0.25M_1M_2}}\right). \quad (1)$$

For large $M_1M_2 - S$, this can be approximated to the following equation:

$$\text{Prob}_{\text{1-level}} = \Phi\left(\frac{(0.5-p)M_1M_2 - (1-p)S}{\sqrt{(M_1M_2 - S)p(1-p) + 0.25M_1M_2}}\right). \quad (2)$$

The proof is omitted because of space limitations.

B. Capacity of Two-Level Decoupled Hamming Associative Memory

Theorem 2: The probability that a two-level decoupled Hamming associative memory returns the target memory is:

$$\text{Prob}_{\text{2-level}} = \sum_{k+j<i\leq \frac{M_1M_2}{R_1R_2}-n_p}^{j+i\leq \frac{M_1M_2}{R_1R_2}-n_p,\, k\leq n_p} \left\{ \binom{n_p}{k} P_s^k (1-P_s)^{n_p-k} \binom{\frac{M_1M_2}{R_1R_2} - n_p}{i} P_t^i \cdot \binom{\frac{M_1M_2}{R_1R_2} - n_p - i}{j} P_n^j P_o^{\frac{M_1M_2}{R_1R_2}-n_p-i-j} \right\}. \quad (3)$$

where,

$$P_s = 1 - 0.5^{R_1R_2}, \quad (4)$$

$$P_t = \sum_{i=\lceil n_f\rceil}^{R_1R_2-1} \left[\binom{R_1R_2 - \lceil n_f\rceil}{i - \lceil n_f\rceil} p^{i-\lceil n_f\rceil}(1-p)^{R_1R_2-i} \cdot \sum_{j=i+1}^{R_1R_2} \binom{R_1R_2}{j} \left(\frac{1}{2}\right)^{R_1R_2} \right], \quad (5)$$

$$P_n = \sum_{i=\lceil n_f\rceil}^{R_1R_2} \left[\binom{R_1R_2 - \lceil n_f\rceil}{i - \lceil n_f\rceil} p^{i-\lceil n_f\rceil}(1-p)^{R_1R_2-i} \cdot \sum_{j=0}^{i-1} \binom{R_1R_2}{j} \left(\frac{1}{2}\right)^{R_1R_2} \right], \quad (6)$$

$$P_o = 1 - P_n - P_t, \quad (7)$$

[2]We do not consider "double" polluting, i.e., when a bit is polluted by both concentrated noise and uniformly random noise, we suppose its value changes only once.

$$n_f = \frac{\frac{S}{N_1 N_2} \times (N_1 N_2 - \lfloor \frac{N_1}{R_1} \rfloor \lfloor \frac{N_2}{R_2} \rfloor R_1 R_2}{\frac{M_1 M_2}{R_1 R_2} - \frac{S}{N_1 N_2} \lfloor \frac{N_1}{R_1} \rfloor \lfloor \frac{N_1}{R_1} \rfloor}, \quad (8)$$

$$n_p = \frac{S}{N_1 N_2} \cdot \lfloor \frac{N_1}{R_1} \rfloor \cdot \lfloor \frac{N_2}{R_2} \rfloor, \quad (9)$$

For large $R_1 R_2$, Equations 5 and 6 can be approximated by:

$$P_t = \sum_{i=\lceil n_f \rceil}^{R_1 R_2 - 1} \left[\binom{R_1 R_2 - \lceil n_f \rceil}{i - \lceil n_f \rceil} p^{i - \lceil n_f \rceil} (1-p)^{R_1 R_2 - i} \cdot \left(1 - \Phi \left(\frac{(i+1) - 0.5 R_1 R_2}{\sqrt{0.25 R_1 R_2}} \right) \right) \right], \quad (10)$$

$$P_n = \sum_{i=n_f}^{R_1 R_2} \left[\binom{R_1 R_2 - n_f}{i - n_f} p^{i - n_f} (1-p)^{R_1 R_2 - i} \cdot \Phi \left(\frac{(i-1) - 0.5 R_1 R_2}{\sqrt{0.25 R_1 R_2}} \right) \right], \quad (11)$$

When $R_1 R_2 - n_f$ is large, Equations 5 and 6 can further be simplified to:

$$P_t = \Phi \left(\frac{(0.5 - p) R_1 R_2 - (1-p) n_f}{\sqrt{(R_1 R_2 - n_f) p (1-p) + 0.25 R_1 R_2}} \right) \quad (12)$$

$$P_n = 1 - \Phi \left(\frac{(0.5 - p) R_1 R_2 - (1-p) n_f}{\sqrt{(R_1 R_2 - n_f) p (1-p) + 0.25 R_1 R_2}} \right) \quad (13)$$

Equation 3 can be simplified as follows:

When n_p is large, and P_s is not too close to 0 or 1:

$$\text{Prob}_{\text{2-level}} = \sum_{j < i \leq \frac{M_1 M_2}{R_1 R_2} - n_p}^{j + i \leq \frac{M_1 M_2}{R_1 R_2} - n_p} \left\{ \binom{\frac{M_1 M_2}{R_1 R_2} - n_p}{i} P_t^i \cdot \binom{\frac{M_1 M_2}{R_1 R_2} - n_p - i}{j} P_n^j P_o^{\frac{M_1 M_2}{R_1 R_2} - n_p - i - j} \cdot \Phi \left(\frac{i - j - 1 - n_p P_s}{\sqrt{n_p P_s (1 - P_s)}} \right) \right\}. \quad (14)$$

When $\frac{M_1 M_2}{R_1 R_2} - n_p$ is large, and $|P_n - P_t|$ is not too close to 1:

$$\text{Prob}_{\text{2-level}} = \sum_{k=0}^{n_p} \left[\binom{n_p}{k} P_s^k (1 - P_s)^{n_p - k} \cdot \Phi \left(\frac{-k - 1 - \left(\frac{M_1 M_2}{R_1 R_2} - n_p \right) (P_n - P_t)}{\sqrt{\left(\frac{M_1 M_2}{R_1 R_2} - n_p \right) (P_n + P_t - (P_n - P_t)^2)}} \right) \right]. \quad (15)$$

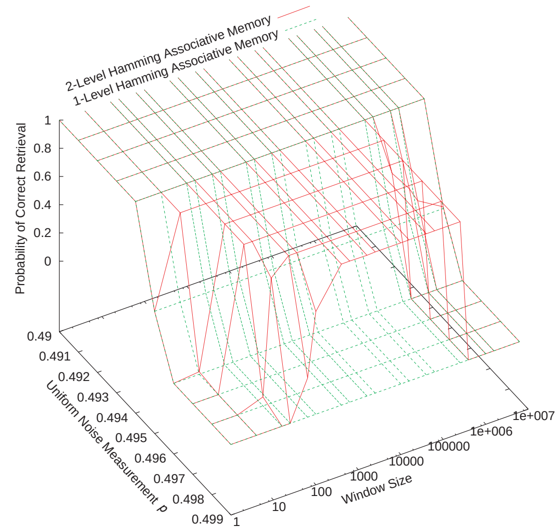
When both n_p and $\frac{M_1 M_2}{R_1 R_2} - n_p$ are large, and neither P_s nor $|P_n - P_t|$ is too close to 1: $\text{Prob}_{\text{2-level}} =$

$$\Phi \left(\frac{-\left(\frac{M_1 M_2}{R_1 R_2} - n_p \right) (P_n - P_t) - n_p P_s}{\sqrt{2 P_n P_t \left(\frac{M_1 M_2}{R_1 R_2} - n_p \right) + n_p P_s (1 - P_s)}} \right). \quad (16)$$

The proof is omitted because of space limitations.

C. Conclusion

Figure 1 shows the capacities with different amount of uniform noise and different region sizes, when some concentrated noise is present. The figure shows clearly that the two level Hamming associative memory always has larger capacity than one level Hamming associative memory does.

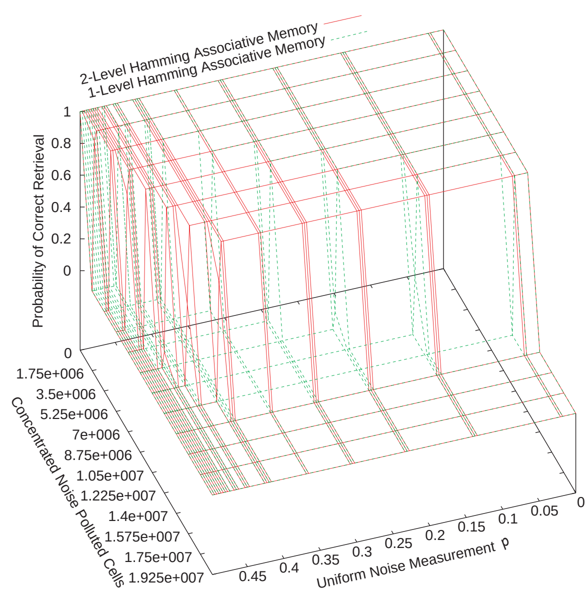


$S = 250000,\ M_1 = M_2 = 5000,\ N_1 = 60,\ N_2 = 30$

Fig. 1. Capacities of Hamming associative memories with different amount of uniform noise and different region sizes, when some concentrated noise is presented

Notice that, in figure 1, the amount of concentrated noise S is much smaller than that of uniform noise $pM_1 M_2$.

The improved capacity of two level Hamming associative memory can also be seen in figure 2, which illustrates the capacities of one-level and two-level Hamming associative memories with different uniform noise level p and different amount of concentrated noise S.



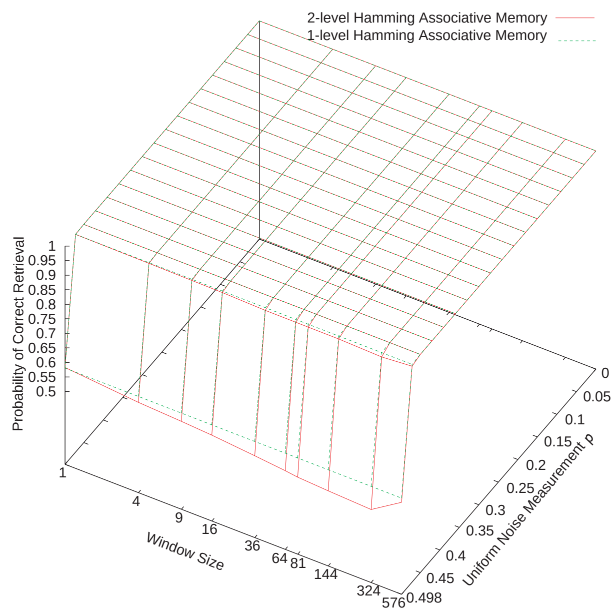
$R_1 = 40,\ R_2 = 10,\ M_1 = M_2 = 5000,\ N_1 = 60,\ N_2 = 30$

Fig. 2. Capacities of Hamming associative memories with different S and different amount of uniform noise

We should mention that, it is hard to see any difference between one-level and two level Hamming associative memories in these figures, for relatively small amount of concentrated noise S. However, there is some difference there, although it is extremely small when the memory pattern size M_1M_2 is large.

When the memory pattern size M_1M_2 is not too large, S is extremely small and p is large, we could notice by Figure 3, that 2-level the capacity of Hamming associative memory is a little smaller than that of 1-level Hamming associative memory. We must point out that, even if these deteriorated capacity against uniform noise is very small, it is important if there are *a lot* of memory patterns in the memory sets.

This is actually why the experimental analysis in [5] shows a substantially lower capacity of 2-level Hamming associative memory in comparing with 1level Hamming associative memory in case only uniform random noise is present. [3] However, as it can be seen in Figure 1, if the amount of concentrated noise is not extremely small like Figure 3, even if it is still far smaller than the amount of uniform noise pM_1M_2, the deteriorated capacity of 2-level Hamming associative memory against uniform noise is more than compensated by its substantially increased capacity against concentrated noise.



$S = 0,\ M_1 = M_2 = 72,\ N_1 = 6,\ N_2 = 3$

Fig. 3. Capacities of Hamming associative memories with different S and different region sizes, when concentrated noise is absent and the entire memory pattern size M_1M_2 is small

IV. Experiments

An experiment has been set up using 2 sets of frontal-view facial pictures of 100 persons; each set contains 100 pictures of different persons. We crop and scale each faces into a standard size of 72×72 pixels, using the position of the eyes, such that the line between two eyes is parallel to the horizontal axis, the inter-ocular distance is set to be 38 pixels.

We use one set as the memory set, another set as the test set.

[3]The training and test sets in the experiments in [5] is too artificially created according to the assumption of the model, they should not be able to be used to verify the correctness of the model itself.

In order to make the two level models more robust to small shift, we applied an elastic graph matching-type procedure. That is, when computing the local distance between the input and a memory image, we compute the distance in the standard position of the window as well as nearby positions (we perturb the position of the window a few pixels in each direction) and select the smallest among these candidate distances as the local distance.

Consequently, when computing the distance between the input and a memory image for the one level model, we list all the windows of size 66×66 in the input and in the memory image and compute the distances between windows in the input and windows in the memory image. Among these many different distances, we select the smallest as the distance between the input and the memory image.

We compare the recognition performance of 1-level and 2-level Hamming Associative Memory. Table I shows the recognition rate at different sizes of windows. This shows that the two level Hamming associative memory works much better than the one level Hamming associative memory.

TABLE I
Recognition Rates of 1-level and 2 level Hamming associative memory for facial recognition

Window Size for 2-Level Hamming Memory					1-Level Hamming Associative Memory
3×3	6×6	9×9	12×12	18×18	
84.5%	89%	92.5%	97%	87%	86%

V. Concluding Remarks

We have shown in this paper that the two-level Hamming associative memory has higher capacity than one-level Hamming associative memory against concentrated noise *at any intensity*. We also showed that, although two-level Hamming associative memory has slightly lower capacity against uniform random noise, its deteriorated capacity is more than compensated by its substantially increased capacity against concentrated noise, even if the amount of concentrated noise is not high. The presence of concentrated noise is very common in practice. We believe the presence of purely uniform noise is very rare in practice, these “pure” situations are unlikely to occur; because such a situation actually means that the noise is dispersed uniformly over the whole pattern *without* any exception, which makes the noise sound too regular and too predictable to be “random” noise. The superiority of the two-level Hamming associative level is evident..

The experiment on real dataset has clearly verified our theoretical conclusion.

References

[1] J. J. Hopfield, “Neural networks and physical systems with emergent collective computational abilities,” *Proceedings of the National Academy of Sciences*, vol. 79, pp. 2554–2558, 1982.
[2] T. Chiueh and R. Goodman, “Recurrent correlation associative memories,” *IEEE Transactions on Neural Networks*, vol. 2, no. 2, pp. 275–284, 1991.
[3] P. A. Chou, “The capacity of the kanerva associative memory,” *IEEE Transactions on Information Theory*, vol. 35, no. 2, pp. 281–298, 1989.
[4] M. H. Hassoun, Ed., *Associative Neural Memories: Theory and Implementation*. New York: Oxford University Press, 1993.
[5] N. Ikeda, P. Watta, M. Artiklar, and M. Hassoun, “A two-level hamming network for high performance associative memory,” *Neural Networks*, vol. 14, no. 9, pp. 1189–1200, 2001.
[6] L. Chen and N. Tokuda, “Stability analysis of regional and national voting schemes by a continuous model,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 15, no. 4, pp. 1037–1042, 2003.
[7] K. J. Arrow, *Social Choice and Individual Values*, 2nd ed. Yale University Press, 1963.