ORIGINAL PAGE
BLACK AND WHITE PHOTOGRAPH

5/8-61

153/8

N91-71368

Willard R. Taber, Ph.D.

General Dynamics
San Diego, California

Dr. Taber majored in physics and astronomy at the University of Iowa, receiving his B.S. in 1973. Following graduation, he worked in the field of soil physics and chemistry, writing computer programs to identify ecologically important variables for nuclear reactor sites. He also began work in computer science at Texas A&M on medical imaging at the College of Medicine. From 1978 through 1984, Dr. Taber designed and built a small computer, performed hardware maintenance on Interdata minicomputers at the chip level, and integrated computers with laboratory equipment, such as spectrophotometers. A major part of his work involved building image and signal processing software. He graduated from Texas A&M in 1984, joining General Dynamics shortly thereafter.

FUZZY LOGIC OPERATORS AND NEURON ACTIVATION FIELDS

Abstract

A neural structure in light of fuzzy sets and operators is examined. During a study of underwater acoustic signatures, it was discovered that a simple version of the avalanche could be improved for classification purposes by adding two simulated hardware memories (or latches) to each neuron. The performance of the new structure, called a neuron ring, approximates the performance of cross correlation. Only simple operators, such as the sigma-count and the triangular norms MAX and MIN are necessary. In brief, the pattern exciting the neuron is viewed as simply a means to induce a possibility distribution in the neuron. The height of the distribution is partial support for the hypothesis in question. The summation of support after a time sequence of excitation is support for the hypothesis.

Fuzzy Sets and Neural Networks

W.R. Taber and R.O. Deich

General Dynamics, Electronics Division, Box 85310
San Diego, CA 92138, MZ 7202-K

INTRODUCTION

Fuzzy sets [15,17] and their operators have interesting applications in neural networks. The performance of fuzzy decision rules for pattern classification provides a compelling reason to use graded set membership functions. As we will show, fuzzy operators enable a simple structure, the neuron ring, to classify non-stationary patterns in the presence of severe noise. Without fuzzy processing, the ring is exceptionally sensitive to noise. Its operation as a shift invariant filter is not appropriate.

The underlying problem with non-fuzzy rings is that decision rules examine terminal activation instead of the historical activation record.

In this paper, we report the performance of neural structures trained with undersea ship signatures from San Diego Bay and elsewhere. We compare performance to cross correlation - a conventional processing algorithm that seldom fails. Backpropagation [9,10,13] performance is also compared both with and without stationarity assumptions.

Reference to ship signatures should not sidetrack the reader from recognizing the contributions of fuzzy processing to pattern recognition. Fuzzy processing may be the best model for non-stationary patterns - those patterns that change their descriptive statistics over time[1]. That ship acoustic energy survives to propagate over geographic distances is amazing. Yet the digital computer, simulating fuzzy rings, correctly identifies ships in the presence of pseudo-white, colored, chirp, and other types of noise. This being the case, the algorithm outlined in this paper has intrinsic value apart from its underwater application.

FUZZY SET EXAMPLE

The indicator function of a standard set is either a 1 or a 0. Either the set element is present or it is not. In distinction, the indicator function of a fuzzy set admits graded membership. An element can be present or absent, or it may be present to a degree. A simple example will illustrate this concept.

Suppose a man with a full head of hair is the subject of an experiment to quantify baldness. The experiment consists of plucking a single hair then recording the answer to the question, "Is this man bald?" The first time, the answer is definitely "no". However, continued experiments with the same subject and other observers will lead to positive answers. The outcome of the experiment is justification for statements of the form "the probability that this man is bald is .50."

Are there other ways to estimate baldness? The answer is yes, and fuzzy set theory provides an approach.

A fuzzy practitioner views bald individuals as a set and tries to estimate an individual's membership from data. For example, he estimates the number of hairs on the average head, then estimates the number of hairs on the subject's head. Without solicitation, he is able to use an estimate of the

[1] See reference 2 for definitions.

number of hairs as the numerator of the ratio of the subject's hair to the average. This is an estimate of the approximate baldness degree. His membership in the set of bald individuals is about the ratio.

Often, there is no information or even requirement to justify probability estimation either from a frequentist's or from a Bayesian's viewpoint. It is in these situations that fuzzy sets offer complementary value.

THE NEURON RING

There are a number of connection intensive networks for pattern classification [7]. The Grossberg avalanche [5] cascades neural elements to learn and recognize spatiotemporal patterns. In signal processing terminology, the avalanche recognizes nonstationary patterns. Hecht-Nielsen [6] further reduced the connectivity of this structure in the commercial SPR (spatiotemporal pattern recognizer) feedforward network.

The neuron ring resembles the SPR but has new architectural features. Its closest approximation is the torus of Goles [4]. The ring's processing element is the DPNL- the dot product neuron with latches that hold time (T) and activation (M) values. These are visible in figure 1.

The DPNL operates in the following manner. During training, the reference pattern is hardwired (fast-learn mode) into the neuron as vector Z. During recognition, the test pattern U is dotted with Z yielding a scalar. This quantity initializes an accumulator in DPNL *i*. The next operation multiplies the unit's old activation, X_{old}, by -a. Another term is computed with *decouple* gating the activation of the previous neuron. The activation sum is multiplied by a gain factor A, yielding the activation increment δX_i^{new}.

Decouple couples a small fraction of activation or *encouragement* forward. When zero, the signal enables a special mode for spectral classification using permutations of firing order as similarity metrics. When decoupled, the DPNLs activate independently, leaving an audit trail of firing order.

Equation (1) relates these quantities for DPNL *i* with a variation in the terminology established by Hecht-Nielsen for the SPR.

Figure 2 illustrates the tertiary structure of a ring assembly. The lateral feedforward and the single return are its gross features. The input layer is not shown.

A pattern sequence excites the ring to a graded activation level. Rather than invoke the all or none firing principle, the DPNL exports its activation untouched except for hard limiting the value to the unit interval [0,1]. The basis for excitation is partial correlation by dot product. That is, correlation at zero time lag. For pattern vectors of unit length, the dot product is in the unit interval. The higher the number, the greater the similarity between V_1 and V_2 on the unit hypersphere in $\Re^n$. Each pattern excites every DPNL. Initially, both latches are reset to 0. As the patterns arrive at the DPNL, the registers latch new values. The max latch, M, changes only when the current activation exceeds M.

After complete pattern presentation, the activation field or max latch array is ana-

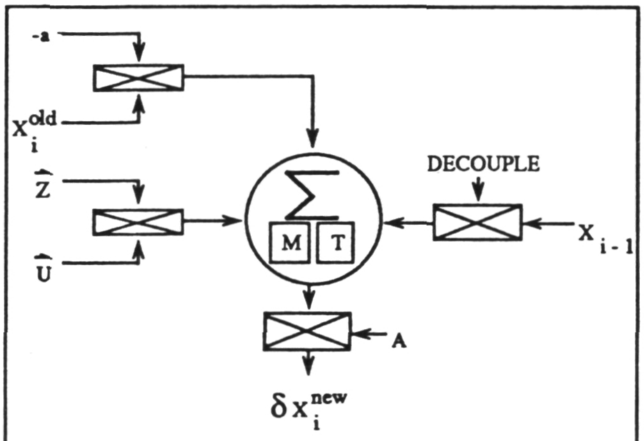


Figure 1. DPNL

lyzed with fuzzy primitives. It is this historical record and processing that is absent in most network paradigms.

$$X_i^{new} = X_i^{old} + A[-a\, X_i^{old} + b\, I_1 + c\, I_2]^+$$

Equation 1

X_i^{new}	= Activation for neuron i
X_i^{old}	= Old activation
A	= Attack factor
a	= Decay constant for old activation
b	= Gain for activation from previous neuron (decouple)
I_1	= Activation from previous neuron
c	= Gain for dot product
I_2	= Dot product of pattern $\vec{Z}_i$ with unknown $\vec{U}_i$

Before discussing fuzzy activation field processing, we first examine two point-sensitive decision rules in common use before this study. They are:

D_1: The ring with the highest acceptable activation in its last neuron wins.

D_2: The ring with the first activation of 1 wins the competition.

We eliminate D_1 immediately. Imagine a ring with 100 neurons. Suppose the test input is identical to the training pattern except that time slice vector 99 is an attenuated version of the true vector. DPNL 100 will not be fully excited. Another ring, conditioned on a different signal, may win at the end of time slice 100 by a simple twist of fate at time 99.

The other rule has competitive merit but it makes little sense when applied in a noisy environment with rampant phase errors. Transients may induce random neuron firing.

Point failure is serious because a system which allows it to occur discounts historical evidence in favor of the current state as does a markov process. Other decision rules are possible. Before addressing the rule(s) of choice, we will discuss some aspects of sampling and phase error.

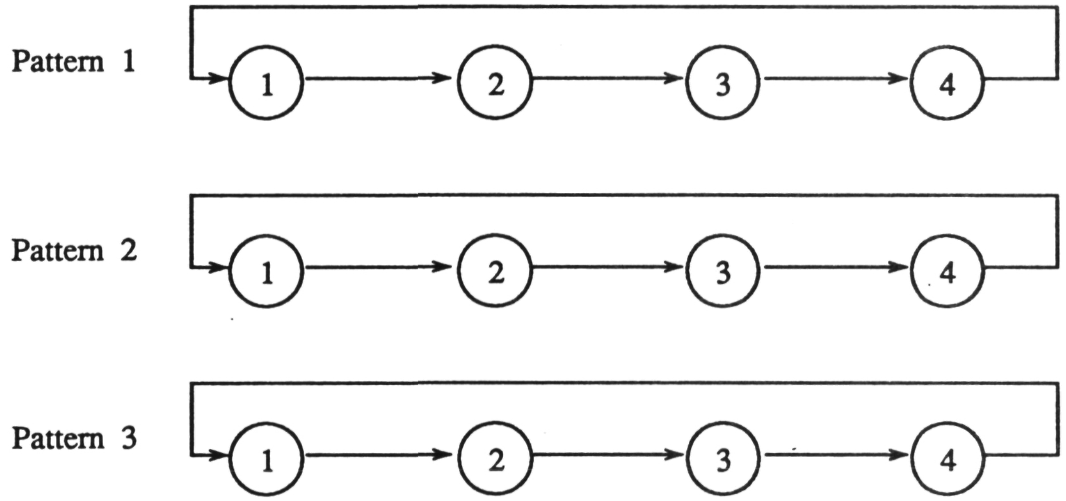


Figure 2: A three pattern ring assembly. Pattern i conditions ring i. Each DPNL supports a microhypothesis (MH) for time j.

Phase error can be illustrated with an example from the undersea application. Assume the acoustic signature is a periodic and deterministic function of propeller angle as it spins. Suppose the training pattern was sampled when the propeller was vertical. The remaining samples follow at equal intervals. During a sea trial, the probability that sampling started with a vertical propeller is small. This being the case, the test and the training signal are similar except for a phase difference.

The re-entrant ring compensates for phase, although the effect is large only for small rings. It makes no difference which DPNL is first stimulated; DPNLs activate around the ring by virtue of the syndetic lines. Phase displacement lies latent in the T latch chain.

A better decision rule or heuristic for pattern classification will now be sought.

Each DPNL continually provides a statistic for testing the hypothesis the pattern is as expected as a correlation by-product. The first DPNL in a ring holds the pattern vector for time slice 1. Therefore, it estimates the grade of membership or suitability of the test pattern's first vector. The second DPNL estimates the suitability of the second pattern vector. Anthropomorphically, operation is a question and answer sequence: "How well, on a unit scale, do you like what you see at this time?" The problem is to decide, that of all the activations generated by a DPNL during presentation, which best indicates support for the hypothesis?

The answer is stated without proof; it is the height of the time series of activation for the DPNL, otherwise called the fuzzy possibility measure. It is just the contents of M.

We estimate the compatibility of the test pattern with the ring as a whole. The appropriate operator is the sigma-count [16] of the fuzzy set M. Kosko [8] proved the sigma-count, ΣC, is a positive measure of set cardinality. It represents support for the ring's hypothesis.

Up to the present time, the discussion has been limited to a single ring. More signals require more rings. The supervisory system, if it exists, recruits empty rings and conditions them as necessary based on mean squared error criteria.

Assuming ΣC_i is support for signal *i*, what rule robustly adjudicates the race for classification?

This discussion argues that no point estimate from the ring during excitation will suffice as a fuzzy statistic, unless its value is unity. We propose the following calculations as a foundation for a better decision rule.

$$\text{Support}_i = \Sigma M_{ij}$$

$$\text{Non-support}_i = \text{Card} - \text{support}_i$$

where support is belief in the hypothesis the pattern is *i*, and *j* is the DPNL index. Card is the cardinality of the ring's non-fuzzy superset, i.e., the number of DPNLs per ring. Non-support is the degree to which the hypothesis is not warranted.

A walk through figure 3 data will clarify the procedure for classification. Morphologically, the assembly that generated the data had 10 rings. Each had 20 DPNLs, one per time slot.

The numbers in the top matrix are the contents of the max latches at the end of the excitation. Rows index the pattern while columns index time.

The contents of the max latch for DPNL for pattern 1, time 18, is 9 (the lack of a decimal is a concession to display technology). The 114 under ΣC is the support for pattern 1 while its non-support is 200 - 114 = 86. With the implied decimal, these values are .9, 11.4, 20, and 8.6.

The certainty ratio (CR) is calculated:

$$CR = Support_i / \Sigma\,(Support_i)$$

Next, each CR is divided by the maximum in the CR column. Finally, "FUZZY MEMBERSHIP" (FM) is reported as a percentage.

The quantities under "FUZZY MEMBERSHIP" indicate the degree to which the training pattern is supported by the data if a decision must be made. Its associated degree of non-support is 100 - itself. For pattern 1, the values are 100 and 0.

If a delayed decision is permitted, the ΣC column is more appropriate than FM. If the support for any ship is lower than a threshold, classification can be deferred until more samples are available.

The formal definition of the possibility computations is as follows. Let the time slot set for a single neuron be A = {1, 2, 3, ..., n}. Let x_i be the activation of a neuron at time *i*:

$$x = \{x_1 \quad x_2 \quad x_3 \ldots \quad x_n\}$$

STATISTICS FOR TEST FILE: SHIPS INPUT SIGNAL: Boat 2

WINDOW =	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	ΣC
Boat 2	0	0	2	4	6	5	5	5	4	6	6	7	7	7	6	8	7	$\boxed{9}$	10	10	$\boxed{114}$
Boat 3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Elizabeth	0	0	2	4	3	3	3	3	3	2	2	2	2	2	2	2	2	4	5	5	51
SEINER	0	0	2	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	14
FF1041A	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
FF1041B	0	0	2	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	12
FFG41B	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
FFG41C	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
DREDGE	0	0	0	2	1	1	1	1	1	1	1	1	1	1	1	1	1	2	2	1	20
ZODIAC	0	0	2	4	5	5	5	4	4	4	4	4	3	3	3	3	3	3	4	6	69

CERTAINTY MEASURES FOR INPUT SIGNAL: Boat 2

TRAINING SIGNAL	CERTAINTY RATIO	FUZZY MEMBERSHIP
Boat 2	$\boxed{0.407}$	$\boxed{100}$
Boat 3	0.000	0
Elizabeth	0.182	45
SEINER	0.050	12
FF1041A	0.000	0
FF1041B	0.043	11
FFG41B	0.000	0
FFG41C	0.000	0
DREDGE	0.071	18
ZODIAC	0.246	61

CLOSEST MATCH FOR INPUT SIGNAL: Boat 2
CURRENT PERTURBATION PERCENT = 40
TEST NUMBER = 25

FIGURE 3. Computer screen depicting the excitation signal Boat 2 with 40 % noise in its power spectrum. Numbers in the top matrix are the contents of the max latch at termination. The last column is the measure of support for the hypothesis implied by the row heading.

Then the possibility distribution:

$$\Pi_X = x_1/1 + x_2/2 + x_n/n$$

is induced on the neuron. The term $x_1/1$ means: the possibility that <u>the signal is appropriate</u> given the signal at time 1 is x_1. Then the possibility measure

$$\pi(A) = \max(x_1, x_2, x_3, \ldots x_n)$$

is held in the max latch, M.

Either the ΣC or the FM statistic is now considered a better decision statistic than those used by either D_1 or D_2. We adopt a decision rule:

$$D_3 = \mathrm{arc}(\bigcup_i \Sigma C_i)$$

or

$$D_4 = \mathrm{arc}(\bigcup_i FM_i)$$

where the functor *arc* is a pointer back to the pattern name. Thus $D_3(114)$ is boat 2.

One objective of pattern recognition is to generalize [1] , to go from a specific signal to the class to which it belongs. An admissible algorithm is driven by the joint similarity between the training pattern(s) and the set of all patterns produced by the same signal source. It infers that while the test signal is different from any in the training set, it has the gross properties of, say, a frigate. There are at least two variations on generalization. The first is mentioned only for the sake of completeness.

Variation 1 examines the output string of the supervisory system if present. For example, the ring's postprocessor declares the signal to be the <u>frigate</u> Mir on the basis of its emissions. Variation 1 parses the text string for the underlined word and subsequently declares the class to be <u>frigate</u>.

Variation 2 scrutinizes the output of the ring - the ΣC or FUZZY MEMBERSHIP tal data alone. Figure 3 illustrates that two of the fishing boats excite the ring but that the Zodiac raft does also. These boats have much in common in the frequency domain. On a broader note, the ring permits the testing of any fuzzy hypothesis that can be constructed from the universe of discourse.

Analysis of the activation history or field is facilitated with an element of set theory called the power set - the set of all subsets from the universe of discourse. Its cardinality is 2^n.

Two constructs derivable from the power set are the frame of discernment or disjunctive frame (Strat [11]) and the frame of concernment or conjunctive frame. In all, they contain $2^{(n+1)}-(n+1)$ unique hypotheses and from them, any hypothesis with conjunction/disjunction is constructible. For example, support for the hypothesis:

H(*Elizabeth)*

is 51 in the ΣC column. We can also test:

H(the *Elizabeth* or boat 3)

with MAX, the fuzzy set union operator.

The arithmetic is a straightforward application of the sigma-count, and the Frank [3] triangular norms and co-norms MAX and MIN:

H(*Elizabeth*) = 51

H(*Elizabeth* or boat 3) =
MAX(0, 51) = 51

H(*Elizabeth* or boat 3)

AND
H(FF1041A or FF1041B) =
MIN(12,51)= 12

Yager [14] and Zadeh [15,17] discuss other operators for reasoning with uncertain information. Similar exercises apply to FM.

EXPERIMENTS AND NOISE

Signatures from ten vessels were collected from San Diego Bay with an omnidirectional hydrophone. They joined an extensive library of marine mammal vocalizations, munition launches, seismic explosions, and other acoustic events.

We soon developed a comprehensive procedure for simulating ship encounters and testing performance. Classification merit is equated to the probability of correct classification, PCC. Graphs of this function have PCC on the vertical axis and percent noise perturbation on the horizontal. The graph indicates the sensitivity level of the procedure under test to levels of increasing noise. Noise is added as percentage of signal power from 0 to 100 percent. A noise level of 25% implies there is 25% uncertainty in the true value of any frequency bin.

All signatures had ambient bay noise. They also contained aperiodic impulse spikes from nearby power rails. These were attenuated with a median filter (Taber [12]) before Fourier transformation.

Uniformly distributed pseudo-white noise is not the only kind of noise in the ocean. As an aid to more comprehensive situations, we used five noise types.

- uniform white
- ramp up with frequency
- ramp down with frequency
- time shift
- convex combinations of signals

Uniform noise occurs in narrow band samples. Our passband was 0-3.5 KHz. In some cases, the uniform noise assumption is justifiable. However, for example, the sudden appearance of a second boat in the water injects frequency and range dependent noise.

Ramp up and ramp down are analogous to linear chirp in radar. The amount of noise is frequency dependent; the amplitude of the zero mean noise either increases or decreases with frequency, simulating ocean anomalies and opening and closing ranges.

Time shifts are expected. Seldom will the training signal be an exact replica of the test. An ocean buoy for monitoring harbor traffic must contend with tracking vessels over a range of perhaps hundreds of miles. In the laboratory, we simulate the buoy by taking the test signal from a different tape or tape segment than the training signal.

Finally, convex combinations of existing signals test the resolving power of the network to identify fleets of ships. Can it generate activation consistent with known signal mixes? For example, if we mix 75% destroyer with 25% frigate, can the network confirm the 75/25 split? Convexity means that for any combination C, of signals, S_k,

$$C = \alpha S_1 + \beta S_2 + \gamma S_3 \ldots,$$

the coefficients sum to 1.

RESULTS

Our experiments tested the ring's ability to recognize signals in the presence of severe noise. Figures 4 and 5 illustrate the varied behavior of the algorithms in uniform zero mean noise. Cross correlation almost always recognizes the signal. Backpropagation (BP) trained with the average Fourier power spectrum rather than the entire non-stationary spectrum does not perform adequately. Its failure justifies the use of the non-stationary signals for training in spite of the expensive training sweeps. The BP network for figure 4 has 16 input neurons, 9 middle layer neurons, and 5 in the output layer (16:9:5). The BP networks for figures 5 and 7 are 320:20:5. BP training time on a Sun 3/140 for ten non-stationary signals is approximately 30 minutes. Training time for the ring is less than a second if training time is measured from the time a pattern is interned to the time the network is prepared to recognize signals.

The non-fuzzy rules make the ring into a very narrow bandwidth matched filter for uniform noise. It does not matter whether D_1 or D_2 is selected; the results are similar to the trace in figure 4b.

The ramp up and ramp down tests caused a single fault in the PCC graph. All plots for the fuzzy ring were constant at 100% PCC for noise levels that started at 25% in the 0-60 Hz band and escalated to 75% in the 3.5 KHz band, and constant for

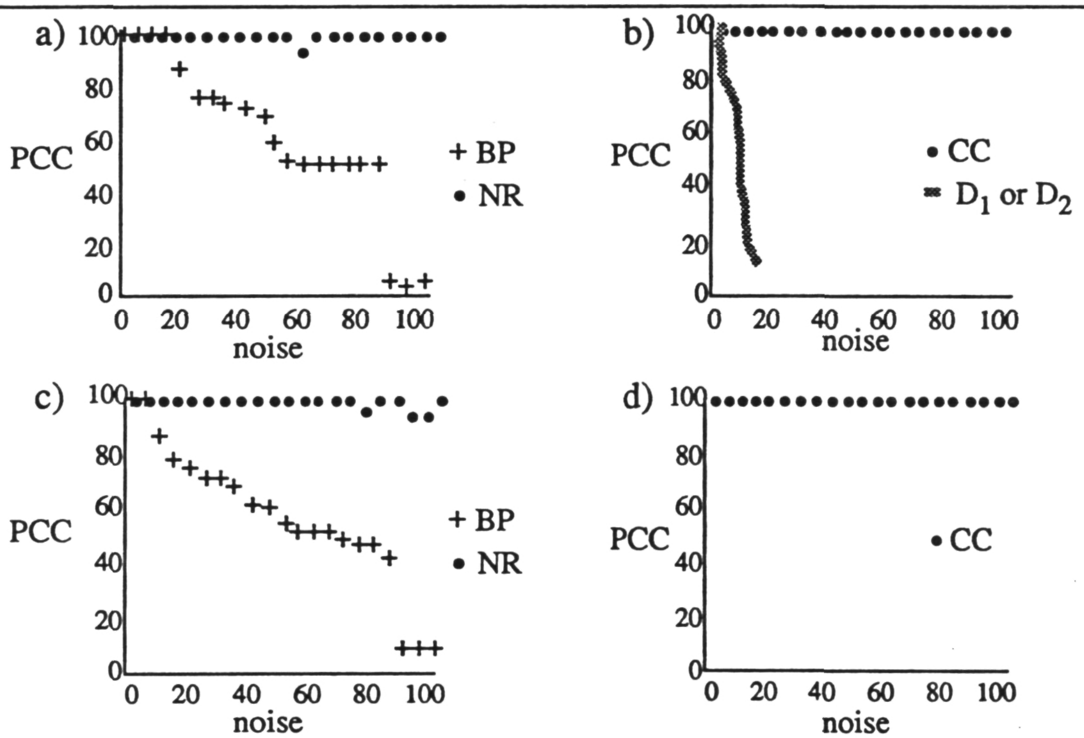


Figure 4. Probability of correct classification (PCC) as a function of additive noise percentage for back-propagation (BP), the neuron ring (NR), cross correlation (CC), and the non-fuzzy ring structure using non-fuzzy rules D_1 or D_2. Non-fuzzy performance is approximate. BP trained on spatial data ; spatio-temporal patterns produced by averaging Fourier data records. Top and bottom graphs pertain to the frigate FF 1041A and to the *Elizabeth*, respectively. Each trace is based on 5000 simulated ship encounters or trials.

the reverse situation for noise ramping down from 75% to 25% with increasing frequency. Thus, the effective correct classification percentage is 99.99% for these tests.

The format of the convexity test was simply to mix the signals then excite the network with the mixture. Work is still in progress to detect if the mix percentage

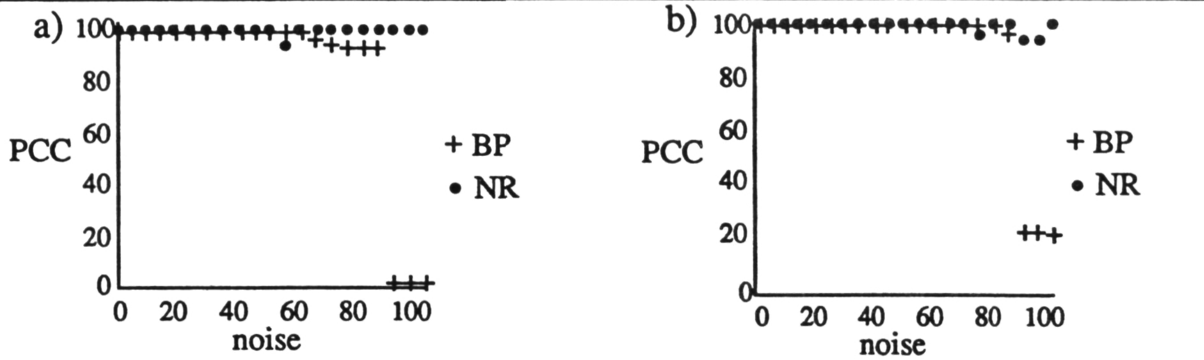


Figure 5. Performance of backpropagation and the fuzzy ring in uniform additive zero mean percentage noise for the frigate and boat 2. Training is with spatio-temporal patterns.

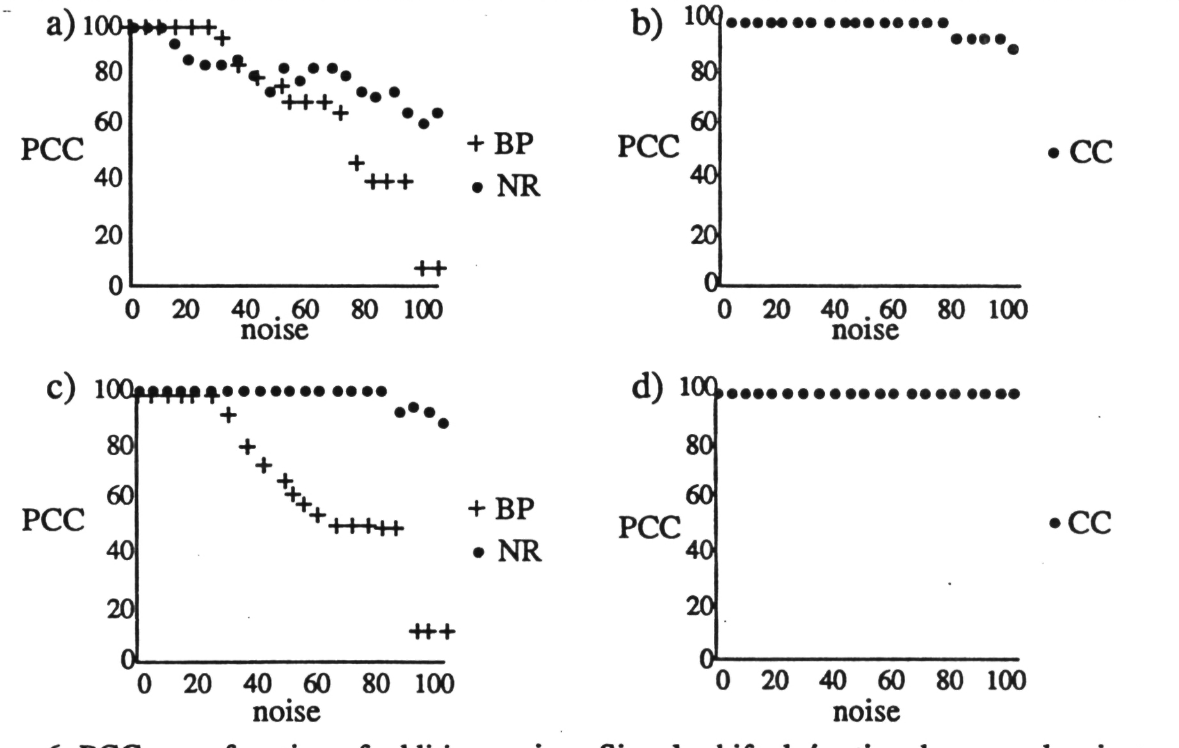


Figure 6. PCC as a function of additive noise. Signals shifted in time by several minutes. Top diagrams are for the frigate FF1041A and the bottom diagrams are for boat 2. BP training is spatial only.

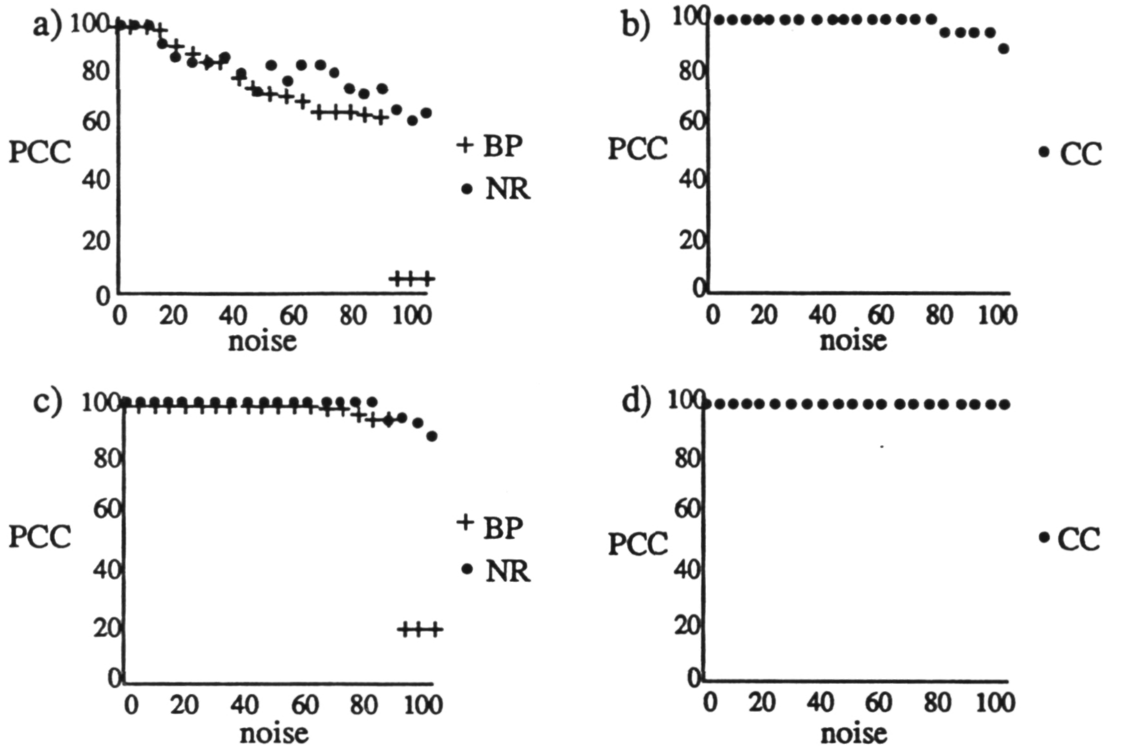


Figure 7

Figure 7. Probability of correct classification as a function of uniform additive noise. Time shift ~ 2 minutes. a) BP and NR for boat 2; b) cross correlation for boat 2. c) BP and NR for the frigate; and d) cross correlation for the frigate. Training is spatio-temporal.

propagates to the decision metrics discussed earlier.

SUMMARY

This study implies that the analysis of the historical activation record or activation field is more effective than using point estimates, at least for simple structures such as the neuron ring. We simulated over a quarter of a million ship encounters in the study; the graphs indicate typical performance. The breadth of the noise and signal characteristics lend credence to the thesis of this paper; namely, that excitation induces a possibility distribution on the neuron's activation. The total support for the ring's hypothesis is the sum of each neuron's possibility measure. This support is a better indicator of pattern class than those used before this study.

REFERENCES

1. Bellman, R., Kalaba, R., & Zadeh, L., Abstraction and pattern classification, Journal of Mathematical Analysis and Applications, **13**, 1-7, 1986.

2. Bendat, J.S. and A.G. Piersol. *Random data: analysis and measurement procedures*, second edition, John Wiley & Sons, New York, 1986.

3. Frank, M.J. On the simultaneous associativity of F(x,y) and x+y - F(x,y). Aequat. Math., **19**, 194-226, 1979

4. Goles, E. Lyapunov functions associated to automata networks. *in* Automata Networks in Computer Science, Theory and Applications, Princeton University Press, F.F. Soulie, Ed., 1987. pp 58-81.

5. Grossberg, S., A prediction theory for some nonlinear functional-differential equations: I, learning of lists, Journal of Mathematical Analysis & Application, **21**, 643-694, 1968.

6. Hecht-Nielsen, R., Nearest matched filter classification of spatiotemporal patterns, Applied Optics, **26**, 1892-1899, 1987.

7. Keller, J.M. and D.J. Hunt. Incorporating fuzzy membership functions into the perceptron algorithm, IEEE Transactions on Pattern Analysis and Machine Intelligence, **PAMI-7**, no.6, 693-699, 1979.

8. Kosko, B., Counting with Fuzzy Sets, IEEE Transactions on Pattern Analysis and Machine Intelligence, **PAMI-8**, no. 4, 556-557, 1986.

9. Parker, D., Learning logic, invention report, SBI-64, File 1, Office of technology licensing, Stanford University, 1982.

10. Rumelhart, D. E., McClelland, J. L., Eds., *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, Vols 1 and 2, MIT Press, Cambridge, 1986.

11. Strat, T. Continuous Belief Functions, Proceedings, AAAI-84, 308 - 313, 1984.

12. Taber, W.R. An Imaging System for the Study of Fluorescent Protein Movement in the Microcirculation. Ph.D. Dissertation, Texas A&M University, 1984.

13. Werbos, P., Beyond regression: New tools for prediction and analysis in the behavioral sciences, Ph.D. Dissertation, Harvard University, 1974.

14. Yager, R.R. Toward a General Theory of Reasoning with Uncertainty. I: Nonspecificity and Fuzziness, International Journal of Intelligent Systems, **1**, no. 1, Spring, 1986.

15. Zadeh, L. A., Fuzzy Sets, Information and Control, **8** 338-353, 1965.

16. Zadeh, L. A., Test-score semantics for natural languages and meaning-representation via PRUF, Technical note 247, SRI International, 1981.

17. Zadeh, L. A., Making Computers Think Like People, IEEE Spectrum, Aug. 1984.

Fuzzy Logic and Neural Networks

by

W. R. Taber

General Dynamics
Electronics Division

San Diego, California

WHAT CAN FUZZY LOGIC

DO

FOR NEURAL NETWORKS ?

This paper describes the results of a study to find minimal neural structures that are able to recognize ships from underwater recordings.

We propose some architectural change to the neuron.

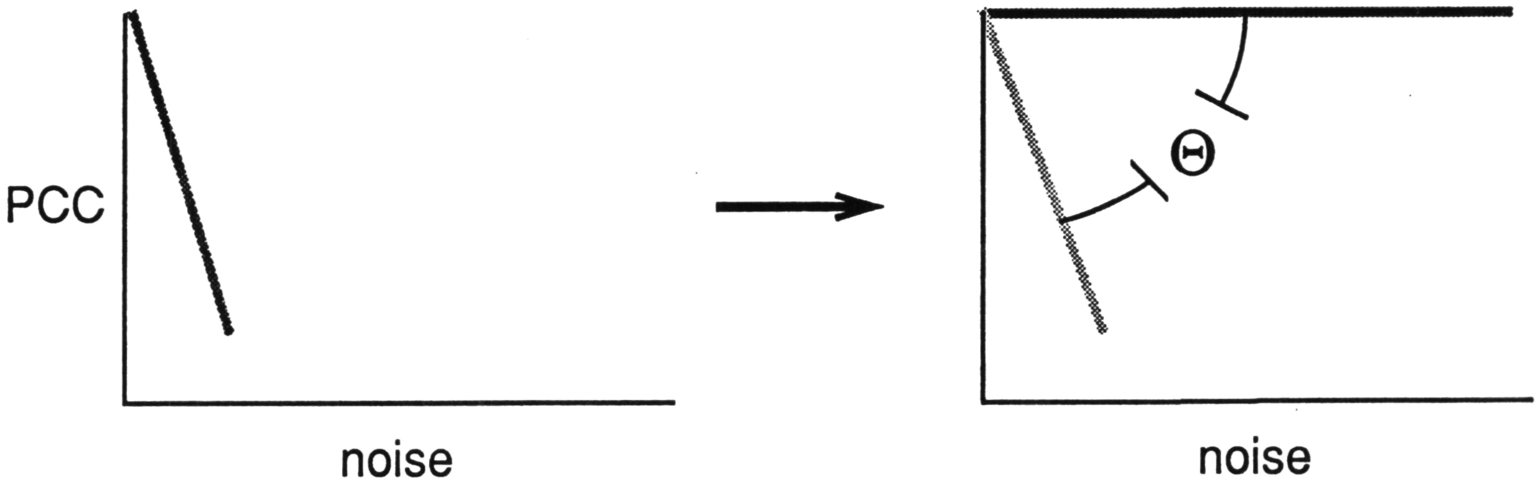
Experiments were designed to test the ability of neural networks to correctly classify ships. During the tests, it became clear that the avalanche worked as an extremely narrow band matched filter.

We went back to the basics and asked:

"What is the neuron doing?"

"Can we boost performance?"

$\uparrow$ Performance = $\downarrow \Theta$



GENERAL DYNAMICS
ELECTRONICS DIVISION

AT THE HEART OF MOST NEURAL NETWORKS

DOT PRODUCT NEURON
PERFORMS

$$u \bullet v = x$$

u = prestored reference vector

v = unknown or test vector

x = nascent excitation scalar

$$F(x) \rightarrow a \; \varepsilon \; (0,1)$$

WHY THIS WORKS

Suppose u, v are positive unit vectors on Hypersphere in R^n

$$\frac{u \bullet v}{||u|| \bullet ||v||} = \cos \phi$$

But $||u|| \bullet ||v|| = 1$ (unit vectors)

So $\cos \phi = u \bullet v \; \varepsilon \; (0,1)$

CORRELATION

$$R_{uv}(r \Delta t) = \frac{1}{N - r} \sum_{i=1}^{N-r} U_i \cdot V_{i+r}$$

for time delay = 0:

$$R_{uv} = \frac{1}{N} \sum_{i=1}^{N} U_i \cdot V_i$$

if N is constant over all patterns.

$$\therefore U \cdot V = R_{uv}$$

Activation by correlation

BASIC OPERATION

$$U \bullet V = x$$

$$\text{Activation} = F(x)$$

$$\text{Example: } F = \frac{1}{1 + e^{-x}}$$

What happens to information if $F(x) \neq x$?

a. We make-up spurious information

or

b. We discount good information

a. If we add spurious information (spurium?)

are we degrading performance ?

b. If we discount good information,

are we degrading performance ?

The problem domain is underwater acoustics

We would like a system to tell us what ship or kind of ship is in the water

We designed a series of experiments to find a system model

DATA CAPTURE

- Record ships with a hydrophone
- Filter the analog waveform at 8 KHz
- Digitize .6 seconds at 20 KHz
- 1024 pt Fourier transform on 20 time slots per ship
- Compute power spectra
- Train neuron rings for the pattern set

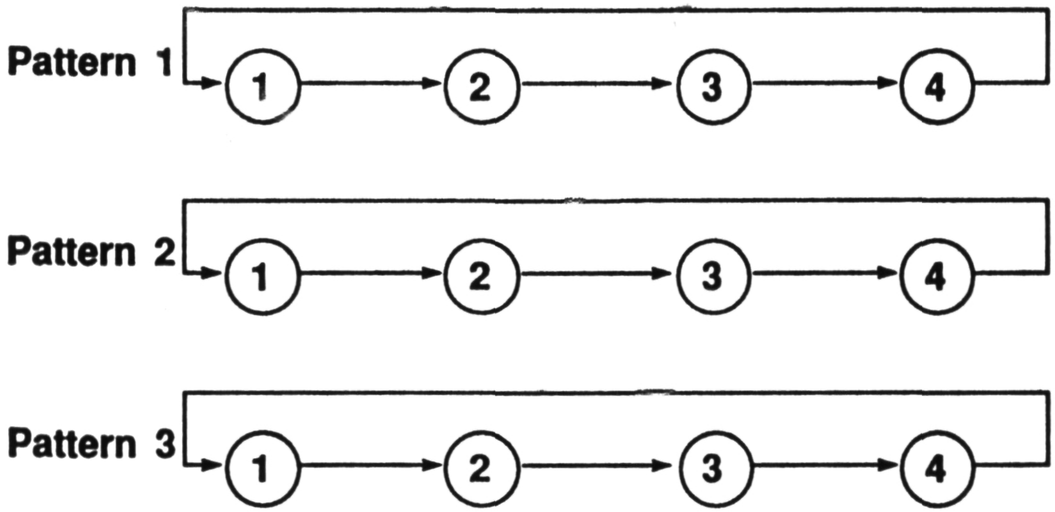
TEST

- Pick test signal
- Add noise to test
- Classify

NOISE TYPES

- Pseudo-white
- Ramp up with frequency
- Ramp down with frequency
- Time shift
- Convex combinations of existing signals

NEURON RING



$$X_i^{new} = X_i^{old} + A[-a\, X_i^{old} + b\, I_1 + c\, I_2]^+$$

X_i^{new} = Activation for neuron i

X_i^{old} = Old activation

A = Attack factor

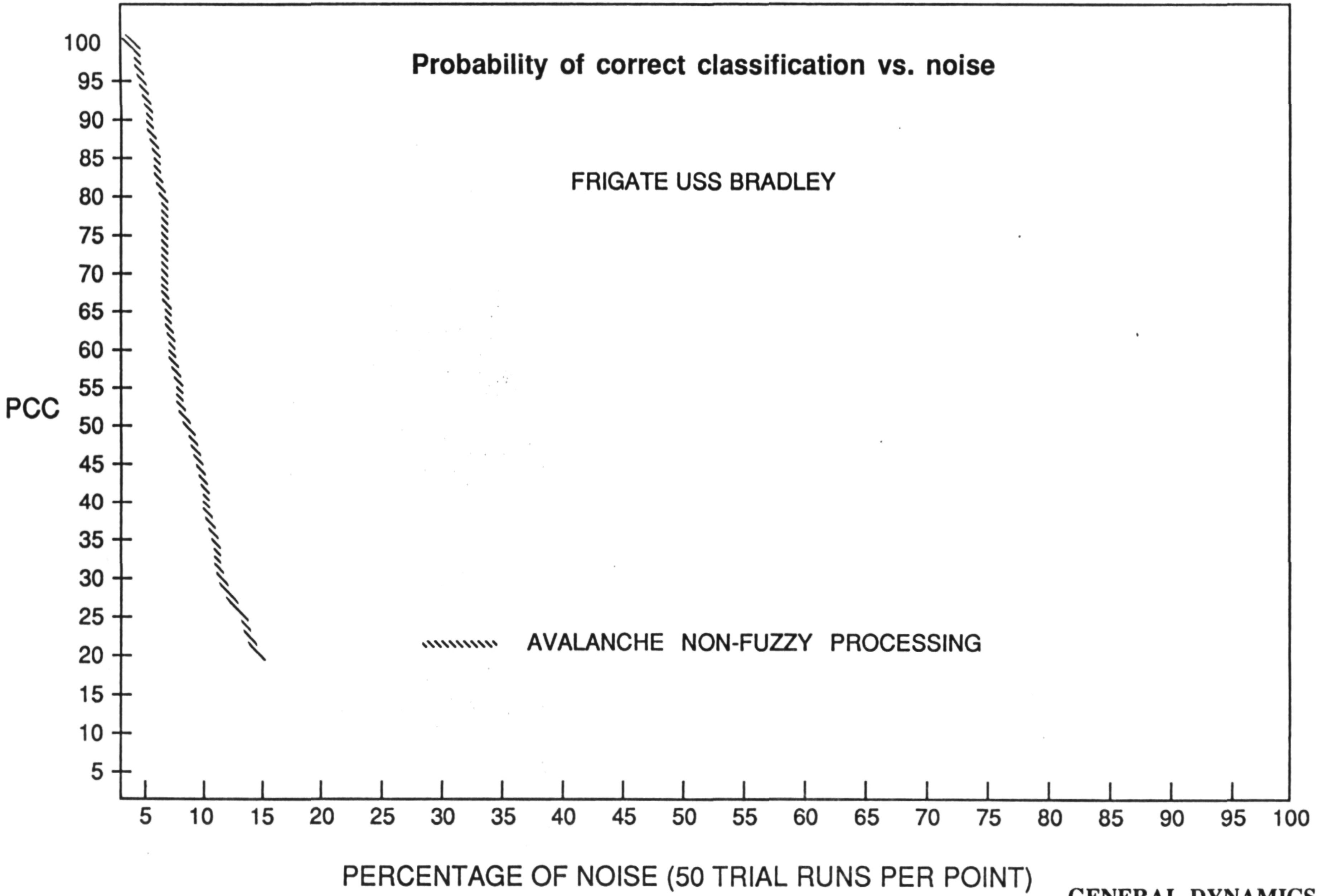
a = Decay constant for old activation

b = Gain for activation from previous neuron (decouple)

I_1 = Activation from previous neuron

c = Gain for dot product

I_2 = Dot product of pattern $\vec{Z}_i$ with unknown $\vec{U}_i$



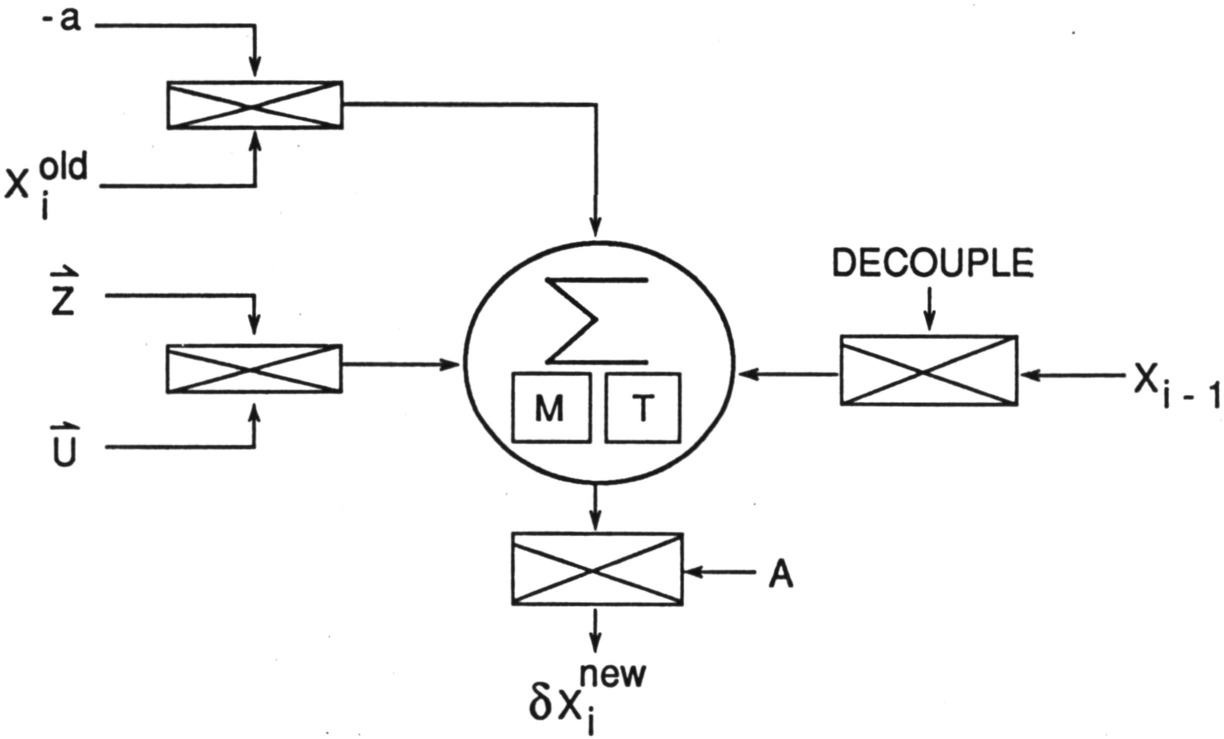
GENERAL DYNAMICS
ELECTRONICS DIVISION

Consider fuzzy sets and multi-valued logic.

- Distribution free
- Graded membership is an attractive idea

Probability	Possibility
P (e or ~e) = 1	Po (e or ~e) = max (e or ~e)
P (.8 or .2) = 1	Po (.8 or .2) = .8

This enables fuzzy processing



The Dot Product Neuron with Latches

LET'S ALLOW THE FOLLOWING F:

$$F(x) = \begin{cases} x & \text{if } 0 \le x \le 1 \\ 1 & \text{if } x > 1 \end{cases}$$

THEN WE MINIMIZE THE MODIFICATION OF INFORMATION

PROCEDURE

- Gate each pattern to every neuron
- Present all 20 vectors
- Decide which pattern it is $D_1 \ldots D_4$

D_1: The ring with the highest acceptable activation in its last neuron wins.

D_2: The ring with the first activation of 1 wins the competition.

$$D_3 = \mathrm{arc}(\bigcup_i \Sigma C_i)$$

$$D_4 = \mathrm{arc}(\bigcup_i FM_i)$$

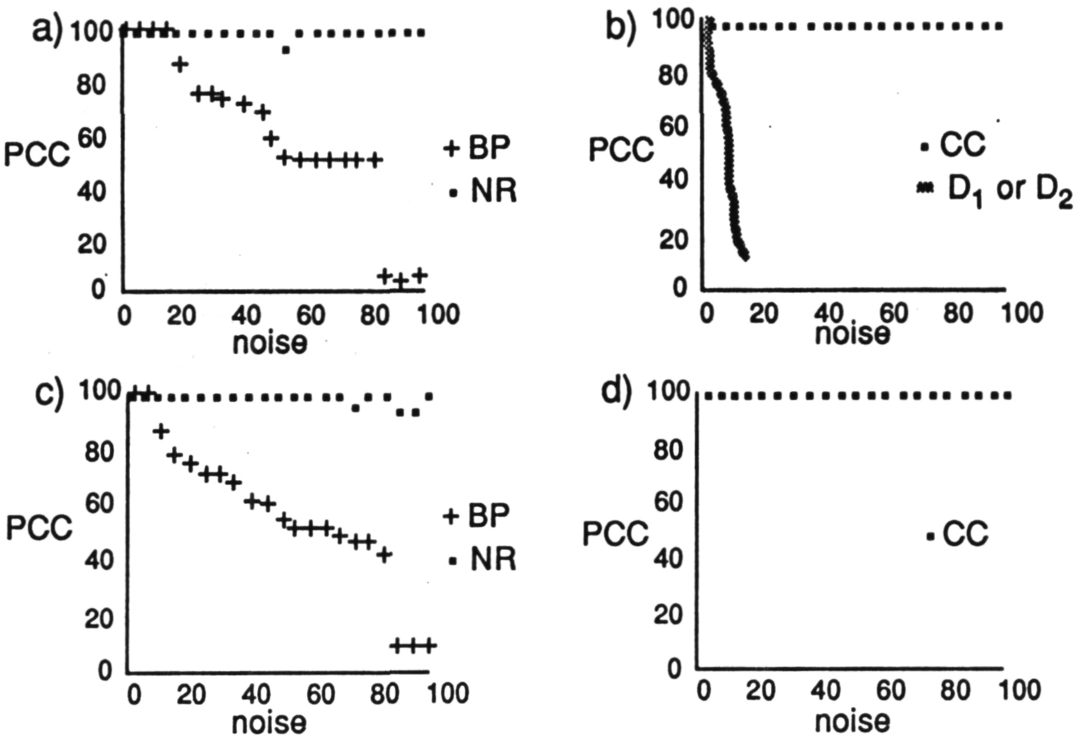


Figure 4. Probability of correct classification (PCC) as a function of additive noise percentage for back-propagation (BP), the neuron ring (NR), cross correlation (CC), and the non-fuzzy ring structure using non-fuzzy rules D_1 or D_2. Non-fuzzy performance is approximate. BP trained on spatial data only; spatio-temporal patterns reduced by averaging Fourier data records. Top frigate; bottom boat 2.

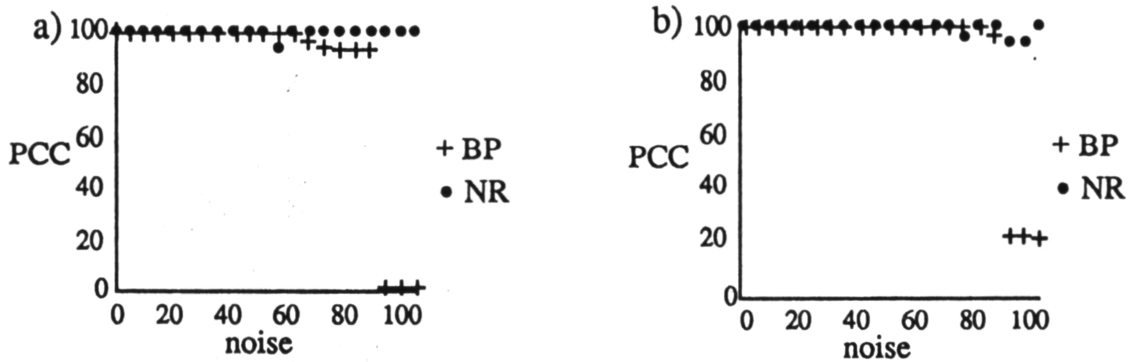


Figure 5. Performance of backpropagation and the fuzzy ring in uniform additive zero mean noise

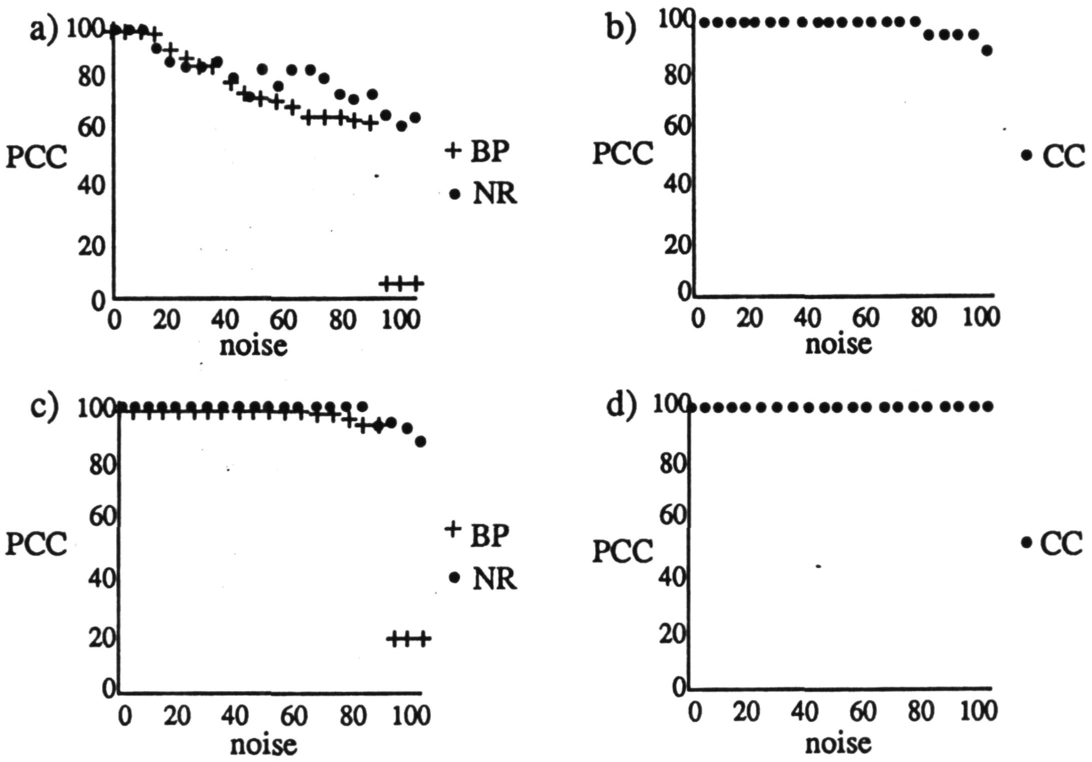


Figure 7

Figure 7. Probability of correct classification as a function of uniform additive noise. Time shift ~ 2 minutes. a) BP and NR for boat 2; b) cross correlation for boat 2. c) BP and NR for the frigate; and d) cross correlation for the frigate. Training is spatio-temporal.

FUZZY PROCESSING

1. Patterns induce a Possibility distribution on the neuron
2. Height of the distribution equates to possibility measure
3. ΣC of the heights $\longrightarrow$ classification

GENERAL DYNAMICS
ELECTRONICS DIVISION

DECISION RULES

NON-FUZZY

D_1: The ring with the highest acceptable activation in its last neuron wins.

D_2: The ring with the first activation of 1 wins the competition.

FUZZY

$$D_3 = \mathrm{arc}\left(\bigcup_i \Sigma C_i\right)$$

$$D_4 = \mathrm{arc}\left(\bigcup_i FM_i\right)$$

ΣC = Sigma-count
arc = pre-image of the argument

STATISTICS FOR TEST FILE: SHIPS   INPUT SIGNAL: Boat 2

WINDOW =	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	ΣC
Boat 2	0	0	2	4	6	5	5	5	4	6	6	7	7	7	6	8	7	$\boxed{9}$	10	10	$\boxed{114}$
Boat 3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Elizabeth	0	0	2	4	3	3	3	3	3	2	2	2	2	2	2	2	2	4	5	5	51
SEINER	0	0	2	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	14
FF1041A	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
FF1041B	0	0	2	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	12
FFG41B	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
FFG41C	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
DREDGE	0	0	0	2	1	1	1	1	1	1	1	1	1	1	1	1	1	2	2	1	20
ZODIAC	0	0	2	4	5	5	5	4	4	4	4	4	3	3	3	3	3	3	4	6	69

CERTAINTY MEASURES FOR INPUT SIGNAL: Boat 2

TRAINING SIGNAL	CERTAINTY RATIO	FUZZY MEMBERSHIP
Boat 2	$\boxed{0.407}$	$\boxed{100}$
Boat 3	0.000	0
Elizabeth	0.182	45
SEINER	0.050	12
FF1041A	0.000	0
FF1041B	0.043	11
FFG41B	0.000	0
FFG41C	0.000	0
DREDGE	0.071	18
ZODIAC	0.246	61

CLOSEST MATCH FOR INPUT SIGNAL: Boat 2
CURRENT PERTURBATION PERCENT = 40
TEST NUMBER = 25

GENERAL DYNAMICS
ELECTRONICS DIVISION

SUMMARY

We fuzzified the neuron

F(x) = x if $0 \le x \le 1$ **1 if $x > 1$** **0 if $x < 0$**	**graded membership**
M latches maximum F(x) **T latches time**	**possibility**
ΣC of M yields hypothesis support	**graded membership**
Θ approaches 0	**near optimal performance**