

STANFORD RESEARCH INSTITUTE

MENLO PARK, CALIFORNIA



February 24, 1964

Nilsson

Proposal for Research
SRI No. ESU 64-15

CALCULUS OF NETWORKS OF ADAPTIVE ELEMENTS

Prepared for:

Rome Air Development Center
Griffiss Air Force Base
New York

In Response to PR 64-707, dated February 3, 1964

Prepared by:

Nils J. Nilsson
Head, Learning Machine Group
Applied Physics Laboratory

Approved:

C. A. Rosen
C. A. Rosen, Manager
Applied Physics Laboratory

R. W. Noe
J. D. Noe, Director
Engineering Sciences Division

Copy No. 7

CALCULUS OF NETWORKS OF ADAPTIVE ELEMENTS

I INTRODUCTION AND BACKGROUND

In response to Rome Air Development Center Purchase Request No. 64-707, dated February 3, 1964, this proposal outlines a program of research aimed at developing a Calculus of Networks of Adaptive Elements.

During the past three years, important advances have been made in the field of trainable pattern-classifying machines. Particularly productive has been the work on those machines based on networks of adaptive elements. These machines, here called learning machines, often consist of networks of adaptive threshold logic units (TLUs). The following developments are illustrative of some of the achievements of the learning machine research program at the Stanford Research Institute: (a) development of low-cost, high-speed, electronically adjustable weighting elements,^{1,2*} (b) design and construction of a large-scale learning machine,³ (c) successful application of learning machine techniques to certain weather prediction problems,⁴ (d) contributions to the theory of the trainability^{5,6,7,8} and capacity⁹ of a threshold logic unit, (e) investigations into the mathematical theory of networks of TLUs,¹⁰ and (f) development of techniques for determining basic structural features in patterns.¹¹

Progress has been rapid and substantial in capitalizing on the relatively few sound theoretical concepts that have emerged in the past five years. There is, however, no well developed body of unifying theoretical principles, although a current RADC contract** has permitted us to initiate appropriate studies. The results, reported in Refs 7,8,9,10, and 11, have, we believe, laid the foundation for developing such a body of mathematical knowledge that can best be described as a Calculus of Networks of Adaptive Elements. SRI now proposes to carry out research to enlarge and contribute to this body in an orderly manner.

In order to appraise the present status of theoretical development, we have prepared the attached Appendix, which outlines in detail the theory that is known and well understood, and indicates problem areas where more research is needed. The following points summarize the conclusions of the Appendix.

- (1) The mathematical theory underling the basic building block, the adaptive TLU, is fairly complete. The TLU implements a linear decision surface (LDS) about which much is known. This knowledge can be subdivided into three parts:

*References are listed at the end of the proposal.

**Contract No. AF 30(602)-2943.

- (a) Mathematical descriptions of the separability achievable by an LDS
 - (b) Training theorems for an LDS, and
 - (c) capacity theorems for an LDS.
- (2) All of the mathematical results that apply to an LDS can be simply extended to a more general class of surfaces called Φ -surfaces. Φ -surfaces include polynomial and other surfaces and have a simple implementation. The fact that these surfaces are trainable is a significant result. It is anticipated that the presently known theorems concerning the training and capacity of Φ -surfaces will be useful in the study of the more complicated decision surfaces implemented by networks of adaptive elements.
- (3) The mathematical theory underlying networks of TLUs is still sketchy and incomplete. Some initial results have been obtained for layered networks. In particular, they implement decision surfaces which are piecewise-linear.

Furthermore, layered networks are efficient in the sense that with only small numbers of component elements they can implement quite complicated surfaces. It is proposed that research on layered networks be continued with emphasis on existence, training, and capacity theorems for piecewise linear decision surfaces.

II OBJECTIVES AND WORK TO BE PERFORMED

The ultimate objective of the proposed research project is to develop a Calculus of Networks of Adaptive Elements. The mathematical knowledge developed by this research will have direct applications in many areas of data processing including automatic pattern recognition.

The specific objectives of this program shall include, but not necessarily be limited to the following items of work:

- (1) A mathematical study of the conditions for the existence of solutions to pattern classifying problems with various kinds of decision surfaces.
- (2) A mathematical study of the adaptive control and manipulation (training) of various kinds of decision surfaces
- (3) A mathematical study of the statistical capacity of various kinds of decision surfaces

In all the above items the phrase "various kinds of decision surfaces" shall include linear surfaces, piecewise linear surfaces, and Φ -surfaces. The results of the above studies will be organized into a unified theoretical structure that shall include existence theorems, training theorems and capacity theorems.

III METHOD OF APPROACH

The problems outlined in the preceding section shall be attacked by a combination of mathematical techniques that have so far proven eminently successful. These include techniques from the fields of probability theory, matrix algebra, n-dimensional geometry, switching theory, and modern algebra. In addition, digital computer simulations will be used whenever it is felt that such experimenting will either add insight into the solution of problems or will be helpful in checking theoretical results.

IV REPORTS

A final technical report will be submitted within one month after the termination of the proposed work. In addition, monthly progress letters shall be submitted and occasional technical notes will be written as required to record important milestones.

V PERSONNEL

This work would be performed by staff members of the Applied Physics Laboratory, the Mathematical Sciences Department, and the Computer Techniques Laboratory of the Engineering Sciences Division.

Biographies of key personnel that would be associated with the project follow.

Nilsson, Nils J. - Head, Learning Machine Group

Applied Physics Laboratory

In August 1961 Dr. Nilsson joined the staff of Stanford Research Institute where he has participated in and led research in pattern recognition and self-organizing machines. He has taught courses in Learning Machines at Stanford University and the University of California, Berkeley. He soon expects to publish a Monograph covering recent theoretical work in Learning Machines.

Dr. Nilsson received an M.S. degree in Electrical Engineering in 1956 and a Ph.D. degree in 1958, both from Stanford University. While a graduate student at Stanford he held a National Science Foundation Fellowship. His field of graduate study was the application of statistical techniques to radar and communications problems.

Before coming to SRI, Dr. Nilsson completed a three-year term of active duty in the United States Air Force. He was stationed at the Rome Air Development Center, Griffiss Air Force Base, New York. His duties entailed research in advanced radar techniques, signal analysis, and the application of statistical techniques to radar problems. He has written several papers on various aspects of radar signal processing. While stationed at the Rome Air Development Center, Dr. Nilsson held an appointment as Lecturer in the Electrical Engineering Department of Syracuse University.

Dr. Nilsson is a member of Sigma Xi, Tau Beta Pi, and the Institute of Electrical and Electronics Engineers.

Duda, Richard O. - Research Engineer, Applied Physics Laboratory

Dr. Duda received a B.S. degree in 1958 and an M.S. degree in 1959, both in Electrical Engineering, from the University of California at Los Angeles. In 1962 he received a Ph.D. degree from the Massachusetts Institute of Technology, where he specialized in network theory and communication theory.

Between 1955 and 1958 he was engaged in electronic component and equipment testing and design at Lockheed and ITT Laboratories. From 1959 to 1961 he concentrated on control system analysis and analog simulation, including adaptive control studies for the Titan II and Saturn C-1 boosters, at Space Technology Laboratories.

In September 1962 Dr. Duda joined the staff of Stanford Research Institute, where he has been working on problems of preprocessing for pattern recognition and on the theory and applications of learning machines.

Dr. Duda is a member of Phi Beta Kappa, Tau Beta Pi, Sigma Xi, and the Institute of Electrical and Electronics Engineers.

Elspas, Bernard - Senior Research Engineer, Computer Techniques Laboratory

Dr. Elspas received a B.E.E. degree from the City College of New York in 1946, and M.E.E. degree from New York University in 1948, and a Ph.D. degree in Electrical Engineering in 1955 from Stanford University. From 1949 to 1951 he was a Research Assistance and a Research Associate in the Electron Tube Group at New York University. From 1951 to 1954 he was a Research Assistant at the Electronic Research Laboratory at Stanford. He held a National Science Foundation Pre-Doctoral Fellowship from 1952 to 1954. Upon completing his studies, Dr. Elspas did research in the Stanford Applied Electronics Laboratory on the application of statistical communication theory in radar systems. He has taught courses in communication theory for the University of California Engineering Extension program, and in coding theory at Stanford University.

In 1955 Dr. Elspas joined the staff of Stanford Research Institute, where he participated in the development and testing program of the ERMA computer and has been engaged in the study of some basic problems in sequential switching theory. The statistical analysis and synthesis of communications systems is another area in which he has specialized. He carried out an analysis of the vulnerability of FSK teletype systems to various kinds of jamming, and he has also been engaged in research on signal-analysis techniques. Dr. Elspas' recent work has been concerned principally with the design and instrumentation of efficient error-correcting codes, and with the development of advanced techniques for the logical design of sequential digital networks.

Dr. Elspas is a member of Sigma Xi, the Scientific Research Society of America, the Institute of Electrical and Electronics Engineers, and the IEEE Professional Technical Groups on Information Theory and on Electronic Computers.

Munson, John H. - Research Physicist, Applied Physics Laboratory

Since joining SRI in 1963 Dr. Munson has been engaged in learning machine research and applications. His activities have included the exploration of combined digital computer-learning machine systems and their potential application for advanced automata.

Dr. Munson received a B.Sc. degree with honors from the California Institute of Technology in 1960. He received an M.A. degree in 1962 and a Ph.D. degree in 1964 (to be formally conferred in June 1964), both from the University of California at Berkeley, in the field of Physics. He held a National Merit Scholarship award as an undergraduate, and a National Science Foundation Fellowship as a graduate student.

In his doctoral research in nuclear physics, Dr. Munson participated in the design and use of a computer-connected system for measurements on bubble-chamber film. He was primarily engaged in machine-language, FORTRAN, and hybrid computer programming, real-time man-machine systems, and graphical pattern recognition. This past experience has also included work in reactor physics, data analysis, and analog computers.

Dr. Munson is a member of Tau Beta Pi.

Singleton, Richard C. - Research Mathematical Statistician
Mathematical Sciences Department

Dr. Singleton received both B.S. and M.S. degrees in Electrical Engineering in 1950 from the Massachusetts Institute of Technology. In 1952 he received the M.B.A. degree from Stanford University Graduate School of Business. He holds also the degree of Ph.D. in Mathematical Statistics from Stanford University, conferred in 1960. His Ph.D. research was in the field of stochastic models of inventory processes, applying the general theory of Markov processes.

Dr. Singleton has been a member of the staff of Stanford Research Institute since January 1952. During this period, he has engaged in operations research studies, in the application of electronic computers to business data processing, and in general consulting in the area of mathematical statistics. His work the past several years has been mainly on the mathematical theory of self-organizing machines, magnetic-core switches, and error-correcting codes. He has written several articles for professional journals.

Before joining the Institute staff in 1952, Dr. Singleton's industrial experience included work in the product engineering and industrial engineering departments at Philco Corporation in Philadelphia, and employment as the chief engineer for a radio broadcasting station. He was an Instructor while doing graduate work at M.I.T.

Dr. Singleton is a member of a number of professional societies, including the Institute of Mathematical Statistics, the Institute of

Electrical and Electronics Engineers, the Operations Research Society of America, the Research Society of America, Eta Kappa Nu, and Sigma Xi.

Kaylor, Donna J. - Mathematician, Mathematical Sciences Department

Mrs. Kaylor joined the staff of Stanford Research Institute in April 1962 as a Mathematician with the Applied Physics Laboratory, where she was engaged in the formulation of mathematical problems concerning the structure and training of adaptive machines. Since her transfer in June 1963 to the Mathematical Sciences Department, she has continued to study the structure of networks of threshold elements.

Mrs. Kaylor attended the University of California at Davis from 1956 to 1958 and Stanford University from 1958 to 1960, receiving her B.S. degree in Mathematics. She received her M.S. degree in Mathematics (including background in Electrical Engineering) from Stanford in June 1962.

During the summers of 1959, 1960, and 1961, Mrs. Kaylor was employed as a mathematician for the U.S. Naval Radiological Defense Laboratory, Military Evaluations Division, in San Francisco. Her work there included analysis of performance of radiological countermeasures systems and analysis of ocean currents and their effect on the detection of radioactive ocean masses.

Mrs. Kaylor is a member of Phi Beta Kappa, American Mathematical Society, and the Mathematical Association of America.

Ablow, Clarence M. - Head Applied Mathematics Group,
Manager, Mathematical Sciences Department

Dr. Ablow received, in 1951, a Ph.D. degree in Applied Mathematics from Brown University. He then became a Research Specialist at the Boeing Airplane Company, engaged in Applied Mathematics, and remained there until he joined the staff of Stanford Research Institute in 1955. At the Institute he has been concerned with problems in continuum mechanics, heat transfer, and chemical kinetics. This work has lead to a number of publications in technical journals as listed below.

He is a member of Phi Beta Kappa, Sigma Xi, the American Mathematical Society, the Mathematical Association of America, and the Society for Industrial and Applied Mathematics.

Publications:

C. M. Ablow, "The Strength of Seismic Shock in an Elastic Earth Under Blast Loading," Proceedings, Fourth Natl. Congress Appl. Mechanics, Berkeley, June 1962.

C. M. Ablow and Henry Wise, "Diffusion and Heterogeneous Reaction, IV. Effects of Gas-Phase Reaction and Convective Flow," J. Chem. Phys., Vol. 35, No. 1 (July 1961).

C. M. Ablow and M. W. Evans, "Theories of Detonation," Chem. Revs., Vol. 61, No. 2 (April 1961).

C. M. Ablow, "Wave Refraction at an Interface," Quart. Appl. Math. Vol. XVIII, No. 1 (April 1960).

C. M. Ablow and C. L. Perry, "Iterative Solutions of the Dirichlet Problem for $\Delta u = u^2$," J. Soc. Indust. Appl. Math., Vol. 7, No. 4 (December 1959).

C. M. Ablow and Henry Wise, "Diffusion and Heterogeneous Reaction. I. The Dynamics of Radical Reactions," J. Chem. Phys., Vol. 29, No. 3 (September 1958).

C. M. Ablow and S. E. Rea, "Transient Air Temperatures in a Duct," Trans. Amer. Soc. Mech. Engrs., Vol. 79, No. 7 (October 1957).

C. M. Ablow and Henry Wise, "Burning of a Liquid Droplet. III. Conductive Heat Transfer Within the Condensed Phase During Combustion," J. Chem. Phys., Vol. 27, No. 2 (August 1957).

C. M. Ablow and Georges Brigham, "An Analog Solution of Programming Problems," J. Operations Res. Soc., Vol. 3, No. 4 (November 1955).

Rosen, Charles A. - Manager, Applied Physics Laboratory

Dr. Rosen received a B.E.E. degree from the Cooper Union Institute of Technology in 1940. He received an M.Eng. in Communications from McGill University in 1950, and a Ph.D. degree in Electrical Engineering (minor, Solid-State Physics) from Syracuse University in 1956.

Since December 1959 Dr. Rosen, as Manager of the Applied Physics Laboratory, has been engaged in the technical planning and build-up of facilities and personnel to carry out major projects in microelectronics and learning machines.

In 1940-1943 he served with the British Air Commission as a Senior Examiner dealing with inspection, and technical investigations of aircraft radio systems, components, and instrumentation. During the period 1943 to 1946, he was successively in charge of the Radio Department, Spot-Weld Engineering Group, and Aircraft Electrical and Radio Design at Fairchild Aircraft, Ltd., Longueuil, Quebec, Canada. From 1946 to 1950 he was a co-partner in Electrolabs Reg'd., Montreal, in charge of development of inter-communication and electronic control systems. In 1950 he was employed at the Electronics Laboratory, General Electric Co., Syracuse, New York, where he was successively Assistant Head of the Transistor Circuit Group, Head of the Dielectric Devices Group, and Consulting Engineer, Dielectric and Magnetic Devices Subsection. In August 1957 Dr. Rosen joined the staff of Stanford Research Institute where he helped to develop the Applied Physics Laboratory.

His fields of specialty include learning machines, dielectric and piezoelectric devices, electro-mechanical filters, and a general acquaintance with the solid-state device field. He has contributed substantially as co-author to two books, Principles of Transistor Circuits, R. F. Shea, editor (John Wiley and Sons, Inc., 1953) and Solid State Dielectric and Magnetic Devices, H. Katz, editor (John Wiley and Sons, Inc., 1959).

VI ESTIMATED TIME AND CHARGES

The estimated time required to complete this project and report its results is 13 months. The Institute could begin work upon acceptance of this proposal. The estimated costs are detailed in the attached Cost Breakdown.

VII CONTRACT FORM

It is requested that any contract resulting from this proposal be written on a cost-plus-fixed-fee basis under Basic Agreement No. AF 33(657)-5112 between the United States Air Force and Stanford Research Institute.

VIII ACCEPTANCE PERIOD

This proposal will remain in effect until 30 June 1964. If consideration of the proposal requires a longer period, the Institute will be glad to consider a request for an extension in time.

IX SECURITY CLASSIFICATION

Stanford Research Institute holds a Top Secret facility clearance which may be validated through the cognizant military security agency Western Contract Management Region (RWIP), United States Air Force, Mira Loma Air Force Station, Mira Loma, California. Staff assignments will be in accordance with the level of security assigned to the work.

APPENDIX

A SURVEY OF THE STATUS OF LEARNING MACHINE THEORY

APPENDIX

A SURVEY OF THE STATUS OF LEARNING MACHINE THEORY

The purpose of this appendix is to outline the status of theoretical knowledge about networks of adaptive threshold logic units (TLUs) and the decision surfaces they implement. We shall organize this knowledge under three main headings: (1) the adaptive TLU, (2) augmented TLU devices that implement Φ -surfaces, and (3) networks of adaptive TLUs. Theoretical knowledge under the first two headings is now beginning to grow rapidly, whereas extensive theory has yet to be developed under the third heading.

1. The Adaptive TLU

a. The Mathematical Significance of Separability

To implement a particular dichotomy of a set of pattern vectors by an adaptive TLU it is necessary and sufficient that these vectors be separable by a hyperplane. This geometric statement can be made in an alternative way if each of the pattern vectors in one of the two classes is replaced by its negative; the derived set of patterns must lie in a proper cone if an adaptive TLU is to dichotomize the original set.*

Several authors,^{6,12} have devised tests on a dichotomized set of pattern vectors to determine whether or not the dichotomy is linearly separable.

If N pattern vectors, each of D dimensions, are chosen in such a way that they are in general position (no D of them lying on a $D-1$ dimensional subspace), then there are precisely $C_{N,D}$ linearly separable dichotomies of these N -pattern vectors.** Various authors⁹ have shown that

$$\begin{aligned} C_{N,D} &= 2 \sum_{i=0}^{D-1} \binom{N-1}{i} \quad \text{for } N \geq D \\ &= 2^N \quad \text{for } N < D . \end{aligned}$$

* We assume here that the TLU has a threshold value equal to zero. The effect of non-zero thresholds can be realized by increasing the dimension of the pattern space by one.

** These results are of suggestive value in the case where the pattern vectors are the vertices of the unit cube and thus are not in general position.

b. Training Procedures

If a dichotomy of a set of pattern vectors is linearly separable, various procedures exist for specifying the set of weight values (called a solution weight vector) of the TLU which will implement the dichotomy. It is well known that linear and quadratic programming techniques^{13,14} can be employed to find these weight values, but we are interested here in other methods, which we shall call training procedures. A training procedure has the following characteristics:

- (1) The pattern vectors are presented to an adaptive TLU one at a time (in any sequence in which each of them occurs infinitely often) to determine the TLU response.
- (2) If the TLU response to a pattern is incorrect, an adjustment (adaptation) is immediately made in the weight values. Otherwise, the weight values are left unchanged.

There are several training procedures that have been proven to be effective:

(1) Motzkin-Schoenberg Procedure¹⁵

If the weight values are to be adjusted, they are adjusted according to the following rule:

$$\vec{W}' = \vec{W} - \lambda \frac{\vec{W} \cdot \vec{P}}{\vec{P} \cdot \vec{P}} \vec{P}$$

where

$\vec{W}'$ = New weight vector

$\vec{W}$ = Old weight vector

$\vec{P}$ = Pattern vector inaccurately categorized by TLU with weights given by $\vec{W}$.

For $0 < \lambda < 2$, this rule produces a sequence of weight vectors that converge to a point on the boundary of the region of solution weight vectors unless one member of the sequence is itself a solution weight vector, in which case the sequence then terminates. For $\lambda = 2$, the sequence of weight vectors terminates at a solution.

(2) Rosenblatt-Widrow Procedure^{16,17}

The following weight vector adjustment is made

$$\vec{W}' = \vec{W} - K \frac{\vec{W} \cdot \vec{P}}{|\vec{W} \cdot \vec{P}|} \vec{P}$$

where K is any constant. This rule is guaranteed to produce a solution weight vector (when one exists) in at most a finite number of steps. If a solution weight vector does not exist, it has been shown⁸ that the length of the weight vector remains bounded.

c. Some Properties of Training Procedures

There exists an upper bound⁶ on the number of steps required by the Rosenblatt-Widrow procedure to achieve a solution weight vector. Unfortunately, this upper bound is of little use in estimating the number of steps required for a given classification problem. Additional research might well provide a reasonable answer to this need.

Even though training time cannot be accurately estimated beforehand, it has been established, both experimentally and theoretically, that for binary patterns a (+1, -1) representation leads to more rapid convergence than does a (1, 0) representation.

d. The Statistical Capacity of Adaptive TLUs

Cover⁹ has shown that if N pattern vectors, each of D dimensions, are chosen according to any of a wide class of probability distributions, and if these pattern vectors are given independent, random, binary categorizations, then they are linearly separable with probability $P_{N,D}$ where

$$P_{N,D} = \frac{1}{2^N} \sum_{i=0}^{D-1} \binom{N-1}{i} .$$

Because $P_{2D,D} = \frac{1}{2}$, and because of the pronounced threshold effect of $P_{kD,D}$ near $k = 2$ for large D , it is reasonable to define the capacity, C , of an adaptive TLU as twice the dimension of the pattern. That is,

$$C = 2D .$$

Any attempt to train a TLU to classify correctly more than C randomly chosen, D -dimensional patterns is almost bound to fail if D is large. On the other hand any attempt to train a TLU on fewer than C randomly chosen patterns is almost bound to succeed.

2. Augmented TLU Devices Which Implement Φ -Surfaces

a. Implementation of Φ -surfaces

Suppose we augment the inputs to a TLU by including, in addition to the individual components of the pattern, functions of these components. For example, an augmented TLU might have additional inputs equal to all of the cross products and squares of the individual components. Such a TLU would be capable of implementing a general quadric or second-degree (instead of linear or first-degree) surface.. Other examples of decision surfaces that

can be implemented by similarly augmented TLUs are polynomial surfaces of any degree. Indeed any family of surfaces whose defining equation can be written as a linear function of the augmented TLU weights can thus be implemented. We call such surfaces Φ -surfaces. They are a large and useful class.

b. Properties of Φ -surfaces

It has been shown that all the results on linear surfaces (implemented by ordinary TLUs) can be extended to Φ -surfaces (implemented by augmented TLUs). In particular the following results should be mentioned.

- (1) All Φ -surfaces are trainable--The theorems on error-correction procedures imply that if it is known that a pattern can be correctly dichotomized by members of a particular Φ -surface family, then convergence to a separating surface in that family in a finite time is guaranteed. As an example, suppose we know that some dichotomy of a set of patterns can be achieved by a quadric surface. Then a movable quadric surface can be trained to perform the desired dichotomy.
- (2) The capacity of a Φ -surface depends only on the number of degrees of freedom of the surface--All the capacity results that apply to ordinary TLUs can be extended to augmented TLUs. Let us define the number of degrees of freedom, F , of a Φ -surface as the number of variable weights in the augmented TLU. We then have the result

$$C = 2F$$

where C is the capacity of the augmented TLU implementing a Φ -surface. Note that the capacity does not depend on the type of Φ -surface, only on the number of degrees of freedom. For example, the capacity of an augmented TLU implementing general quadric (second degree) surfaces in D dimensions is $(D+1)(D+2)$. This number is to be compared with $2D$ which is the capacity of an ordinary TLU.

3. Networks of Adaptive Threshold Elements

a. The Committee Problem

When a single adaptive TLU is incapable of implementing a given dichotomy, we are led to inquire into the conditions under which a network of TLUs can together implement the dichotomy. Suppose there are

K TLUs, each of which has as its input the pattern to be classified. If the response of each of these K TLUs is polled, the consensus can be taken to be the network response. We are interested in the mathematical conditions on pattern sets for the existence of a committee of K TLUs whose consensus correctly dichotomizes the set of patterns.

By consensus we mean voting, and we distinguish two possible voting procedures:

- (1) Each TLU has an equal vote
- (2) The various TLUs have adaptable votes.

In either of the above cases the pattern can be categorized on the basis of a simple majority vote or variations such as a larger-than-majority (say, 2/3) vote.

A committee of TLUs is a special case of layered TLU networks. It is known that such networks implement piecewise linear decision surfaces. Such surfaces can be made quite complex even with a small number of elements in the network. Because of the resultant efficiency of such networks it thus becomes important to study the properties of generalized piecewise linear decision surfaces. Very little is known about these properties at the present time. For example, it is not yet known how to adapt such surfaces or what their statistical capacity is (except in special cases to be discussed below). Work to date has centered on the committee networks which implement only a subclass of the generalized piecewise linear surfaces.

A general theory of committees of the TLUs has yet to be developed, but some information is now known for certain special cases which we shall now discuss. We first restrict ourselves to the case of requiring only simple majorities for pattern classification. Fairly complete statements^{10,18} can be made about this restricted case if the dimension of the patterns, D , is equal to 2. Dichotomizing N two-dimensional patterns in the most disadvantageous way can require, at most, a committee of size $N-1$ if N is even and N if N is odd. Furthermore, committees of these sizes are necessary and sufficient to implement all dichotomies of N two-dimensional patterns.

Precise statements about the choice of weight vectors for each TLU in the committee can also be made. An important consequence derived from these results states that for $D = 2$, it is never necessary to adapt the vote strengths of the committee members. It is expected that continued research will begin to provide answers to more general questions about committee networks.

b. The Training of Committee Networks

There are no theorems yet known that are parallel to the Motzkin-Schoenberg and Rosenblatt-Widrow theorems for training committees. Two problems exist. One is to train the committee members themselves, and the other is to adapt appropriately the voting strengths of the committee

members. These problems are usually referred to as the problem of adapting two layers of weights. Several heuristic procedures have been suggested for training the committee members¹⁹ which have been experimentally tested and found quite efficient. Efforts are continuing to abstract principles from these heuristics which can form the basis of solid mathematical theorems.

c. The Statistical Capacity of Committee Networks

Many unsolved problems need further research before capacity formulas can be given for committees. Again, the special case of $D = 2$ has been solved. The probability that N randomly categorized, two-dimensional patterns can be dichotomized by a K -member committee is given by the expression

$$P_{N,2}^{(K)} = \frac{1}{2^{N-1}} \sum_{i=1}^K \binom{N-1}{i} \text{ for } K < N-1 \text{ and } K \text{ odd.}$$

There remain many unanswered questions concerned with the theory of networks of adaptive TLUs. Of particular interest are theorems dealing with the trainability and capacity of these networks. The solutions to these questions are needed to provide a sound basis for a theory of learning machines.

4. Tabular Summary of Solved and Unsolved Problems

The present status of the calculus of Networks of Adaptive Elements is conveniently summarized in Table I. The column headings pertain to the categories of the knowledge we seek about such networks. The row headings describe the various kinds of decision surfaces implemented by the networks studied to date. Each row heading is subdivided into two parts: $R = 2$ and $R > 2$. The symbol R stands for the number of categories into which the input patterns are to be separated. (For certain of the networks the two-category problem is qualitatively different than the poly-category problem.)

Subject Area Type of Decision Surface (Network)	Existence Theorems	Training Theorems	Statistical Capacity Theorems
<u>Linear</u>			
$R = 2$ (TLU)	*	*	*
$R > 2$ (Bank of Summers)	†	*	†
<u>Φ-Surfaces</u>			
$R = 2$ (Augmented TLU)	*	*	*
$R > 2$ (Bank of Augmented Summers)	†	*	†
<u>Piecewise Linear</u>			
$R = 2$ (Network of TLUs)	†	†	†
$R > 2$ (Banks of Summers)			
$R = 2$ (Networks of Summers)	†	†	†
$R > 2$			

TABLE I

PRESENT STATUS OF THE CALCULUS OF
NETWORKS OF ADAPTIVE ELEMENTS

† Indicates area in which further extensive research is needed.

* Indicates area that is fairly well understood.

REFERENCES

1. Brain, A. E., "The Simulation of Neural Elements by Electrical Networks based on Multi-Aperture Cores," Proc. IRE, 49, pp. 49-52 (January 1961).
2. Crafts, H. S., "Design of a Magnetic Variable-Gain Component for Adaptive Networks," 1963 WESCON
3. Rosen, C. A., et al, "A Large, Self-Contained Learning Machine," 1963 WESCON
4. Duda, R. O., and J. W. Machanik, "An Adaptive Prediction Technique and its Application to Weather Forecasting," 1963 WESCON
5. Novikoff, A.B.J., "On Convergence Proofs for Perceptrons," Technical Report, SRI Project 3605, Contract Nonr 3438(00), Stanford Research Institute, Menlo Park, California (January 1963).
6. Singleton, R. C., "A Test for Linear Separability as Applied to Self-Organizing Machines," Technical Report, SRI Project 3605 Contract Nonr 3438(00), Stanford Research Institute, Menlo Park, California (May 1963).
7. Nilsson, N. J., and R. C. Singleton, "An Experimental Comparison of Three Learning Machine Training Rules," Technical Note submitted to RADC under contract No. AF 30(602)-2943, Stanford Research Institute, Menlo Park, California (November 1963).
8. Efron, B. "The Perceptron Correction Procedure in Non-Separable Situations," Technical Note submitted to RADC under contract No. AF 30(602)-2943, Stanford Research Institute, Menlo Park, California (August 1963).
9. Cover, T. M., "Classification and Generalization Capabilities of Linear Threshold Units," Technical Note submitted to RADC under Contract No. AF 30(602)-2943, Stanford Research Institute, Menlo Park, California (November 1963).
10. Kaylor, D. J., "A Mathematical Model of a Two-Layer Network of Threshold Elements," Technical Note submitted to RADC under Contract No. AF 30(602)-2943, Stanford Research Institute, Menlo Park, California (October 1963).
11. Block, H. D., N. J. Nilsson, and R. O. Duda, "Determination and Detection of Features in Patterns," Technical Note submitted to RADC under contract no. AF 30(602)-2943, Stanford Research Institute, Menlo Park, California (June 1963).

12. Gableman, Irving J., "A Note on the Realization of Boolean Functions Using a Single Threshold Element," Proc. IRE 50, pp. 225-226 (February 1962).
13. Minnick, R. C., "Linear-Input Logic," IRE Trans., PGEC-10, pp. 6-16 (March 1961).
14. Rosen, B. R. "Pattern Separation by Convex Programming," Report No. 30, Applied Mathematics and Statistics Laboratories Report, Stanford University, Stanford, California (June 28, 1963).
15. Motzkin, T. S., and I. J. Schoenberg, "The Relaxation Method for Linear Inequalities," Canadian J.Math. 6, 3, pp. 393-404 (1954).
16. Rosenblatt, F., Principles of Neurodynamics: Perceptrons and the Theory of Brain Mechanisms (Spartan Books, Washington, D.C. 1961).
17. Widrow, B., "Generalization and Information Storage in Networks of Adaline Neurons," pp. 435-461 Self-Organizing Systems, Yovits, Jacobi, Goldstein, (Eds.) (Spartan Books, N.Y. 1962).
18. Ablow, L., and D. Kaylor, Unpublished Research Memorandum, Stanford Research Institute, (1963).
19. Ridgeway, W., "An Adaptive Logic System with Generalizing Properties," Technical Report No. 1556-1, Stanford Electronics Laboratories, Stanford University (April 1962).

Proposal for Research
Stanford Research Institute No. ESU 64-15

COST BREAKDOWN

Personnel Costs

Supervisory, $\frac{1}{2}$ man-month at	██████████	\$	██████████
Senior Professional, 8 man-months at	██████████		██████████
Professional, $14\frac{1}{2}$ man-months at	██████████		██████████
Editorial, $\frac{1}{2}$ man-month at	██████████		██████████
Secretarial, Report Clerical, 2 man-months at	██████████		██████████
TOTAL DIRECT LABOR			██████████
PAYROLL BURDEN AT 16%			██████████
TOTAL SALARIES & WAGES			██████████
OVERHEAD AT 95% OF SALARIES AND WAGES			██████████
TOTAL PERSONNEL COSTS			██████████

Direct Costs

Travel and Subsistence (2 cross-continental trips at ██████████ each; 4 days subsistence at ██████████ per day.)	\$	██████████
Shipping and Communication		██████████
Computer time, 10 hours, B5000 (or equiva- lent) at ██████████ per hour		██████████
Report Production Costs		██████████
TOTAL DIRECT COSTS		██████████
TOTAL ESTIMATED COSTS		\$ ██████████
FIXED FEE		██████████
TOTAL CONTRACT COST		██████████

* The rates quoted above represent our current cost experience. It is requested that the contract provide for reimbursement at these rates on a provisional basis, subject to retroactive adjustment to fixed rates negotiated on the basis of historical cost data. Included in payroll burden are such costs as vacation and sick leave pay, social security taxes, and contributions to employee benefit plans.