



Nils J. Nilsson
Computer Science Department
Stanford University
Stanford, CA. 94305-4110

9 March 1965

Proposal for Research
SRI No. ESU 65-12
Extension to Contract AF 30(602)-3448

CALCULUS OF NETWORKS OF ADAPTIVE ELEMENTS

Prepared for:

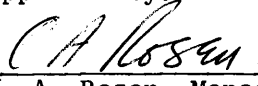
Rome Air Development Center
Griffiss Air Force Base
New York

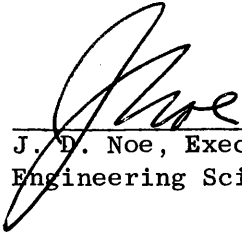
Prepared by:

Richard O. Duda
Research Engineer
Applied Physics Laboratory

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

Approved by:


C. A. Rosen, Manager
Applied Physics Laboratory


J. D. Noe, Executive Director
Engineering Science and Industrial Research

Copy No. 19

I INTRODUCTION AND BACKGROUND

Intensive effort during the last four years has led to a variety of achievements in the field of trainable pattern-classifying systems. Particular progress has been made in the construction, application, and theory of learning machines--trainable systems that often consist of networks of adaptive threshold logic units (TLUs). The following developments are illustrative of some of the accomplishments of the learning machine research program at the Stanford Research Institute:

- (1) Development of low-cost, high-speed, electronically adjustable weighting elements^{1,2*}
- (2) Design and construction of a large-scale learning machine³
- (3) Application of learning machine techniques to certain weather prediction problems⁴
- (4) Development of techniques for determining structural features of patterns⁵
- (5) Development of techniques for clustering similar patterns⁶
- (6) Contributions to the theory of the trainability⁷⁻¹¹ and capacity¹² of a threshold logic unit
- (7) Investigations into the mathematical theory of networks of threshold logic units.^{13,14}

The mathematical results obtained so far represent the beginnings of a theoretical base for understanding the many proposed trainable systems. The ultimate goal of the current RADC contract** is to provide such a base by developing a calculus of networks of adaptive elements. The major steps that have been taken toward this goal are reported in detail in Ref. 15, and the present state of this research can be summarized as follows:

- (1) A framework for a general theory of trainable pattern-classifying machines has been developed. In this framework, the central questions involve (a) the characterization, (b) the training, and (c) the capacity of various classes of decision surfaces--linear surfaces, piecewise linear surfaces, quadric surfaces, Φ -surfaces,*** etc.

* References are listed at the end of the proposal.

** Contract No. AF 30(602)-3448.

*** Φ -surfaces include quadric surfaces, polynomial surfaces of any order, and other surfaces whose discriminant functions are linear in their adaptive parameters.

- (2) These central questions have been well answered for two-category classification with a linear surface-- i.e., classification by a TLU.
- (3) All of the results for a TLU have been extended to two-category classification by a Φ -surface.
- (4) Partial results have been obtained for R-category classification by linear surfaces (and Φ -surfaces). A training theorem and special-case capacity results are available, and the promise of additional results is very good.
- (5) The mathematical theory of networks of TLUs remains incomplete, despite considerable effort and some interesting and hard-won results.^{13,14}

One of the problems encountered in a systematic study of networks of TLUs is the great number of classes of such systems available for study. Some of the more important structures, described in detail in the attached Appendix, are as follows:

Two-category classifiers:

- (1) Threshold logic unit
- (2) Linear machine
- (3) Piecewise linear machine
- (4) Committee machine

R-category classifiers:

- (1) Parallel-connected dichotomizers
 - (a) Tokoloshe
- (2) Linear machine
- (3) Piecewise linear machine.

In addition to the variety of interesting structures, there is a variety of plausible training algorithms for each structure, such as the Ridgway algorithm¹⁶ for the committee machine, and the Nilsson algorithm¹⁶ for the piecewise linear machine. Good heuristic arguments can usually be marshalled for using each of these structures and each of these training algorithms, but mathematical proofs are usually difficult to obtain.

In view of the difficulties encountered in a direct mathematical attack on these problems, it is proposed that the analytical program be augmented by an experimental program of carefully designed simulation studies. This program would have as its primary goal a critical examination and comparison of proposed machine organizations and training algorithms. Theoretical investigation could then be concentrated on the most successful structures and training rules. In addition to this very systematic study, a certain amount of exploratory research should be performed to allow the creation and investigation of significantly different organizations and training rules. Such a combined analytical and experimental attack, we

believe, holds the greatest promise of broadening and extending the theoretical foundations of trainable systems. SRI proposes that the present contract, AF 30(602)-3448, be extended for a one-year period to allow this work to be performed.

II OBJECTIVES

The ultimate objective of this research project is to develop a Calculus of Networks of Adaptive Elements. Such a calculus should include:

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

The phrase "various kinds of decision surfaces" shall include linear surfaces, piecewise linear surfaces, and Φ -surfaces.

The specific objectives of the proposed extension of the project shall include the following:

- (1) To extend the theoretical results on training and capacity for a TLU to a linear machine, and whenever possible, to piecewise linear machines
- (2) To prepare for further theoretical advances by conducting a well planned program of exploratory research on new organizations and training procedures using computer simulations
- (3) To conduct a critical examination and comparison of various proposed organizations including linear machines, piecewise linear machines, committee machines, and parallel-connected dichotomizers.

III WORK TO BE PERFORMED

The development of a general theory of trainable pattern-classifying machines requires an understanding of R-category problems, $R > 2$, and an understanding of networks of adaptive devices. The greatest potential for concrete analytical results for the R-category problem undoubtedly lies in the linear machine. A convergence proof for a fixed-increment training procedure is known,¹⁶ and special-case capacity results have been obtained. Since in many ways the linear machine is a natural generalization of a TLU, it is proposed that primary theoretical effort be devoted to extending the training and capacity results for a TLU to a linear machine.

Some of the results for a linear machine may be directly extendable to a piecewise linear machine, and whenever possible this will be done. However, the training of a piecewise linear machine raises qualitatively

different problems, since the very structure of the machine can be changed during training. Considerable theoretical effort, augmented by simulation experiments whenever appropriate, will be devoted to the analysis of this important trainable system.

Probably the most fruitful way to compare and contrast the numerous machine organizations described in the attached Appendix is through a carefully planned and executed program of experiments. The performance of the different systems depends upon the characteristics of the patterns that they are required to classify. To provide a comprehensive test of the different structures, more than one class of patterns would probably be needed. Moreover, it is highly desirable that the patterns used be suitable as standards for comparing the performance of adaptive systems with that of other pattern-classifying techniques. Certain familiar classes of patterns, such as unimodal and multimodal gaussian patterns, or binary patterns with statistically independent components, have proven to be suitable in the past, and are possible candidates for this study. However, further consultation with RADC is needed before a final selection can be made.

Standard criteria are available for objectively comparing the classification performance of the different systems on these patterns. These include learning curves, convergence time, final error rate on training data, and error rate on independent testing data. Special features of certain systems that may be advantageous for special problems will be noted, but will not be considered in the final overall evaluation.

IV METHOD OF APPROACH

The mathematical problems outlined in the preceding section shall be attacked by a combination of those techniques that have been most successful in the past. These include techniques from the fields of probability theory, matrix algebra, n-dimensional geometry, switching theory, coding theory, and modern algebra. The simulation experiments will be performed on either an SDS 910 computer or a Burroughs B5000 computer, the former being more suitable for exploratory research, and the latter being more suitable for runs requiring extensive computation.

V REPORTS

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

VI PERSONNEL

This work would be performed by staff members of the Applied Physics Laboratory and the Mathematical Sciences Department. In addition, it is proposed that provision be made for continuing the services of a consultant whose past contributions have been of definite benefit to this project.

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

Ball, Geoffrey H. - Senior Research Engineer, Radio Systems Laboratory

Dr. Ball received the degree of A.B. in Applied Physics and Engineering Science at Harvard University in 1955, and his M.S. and Ph.D. degrees in Electrical Engineering from Stanford University in 1960 and 1962, respectively.

Initially, his graduate study was concerned with statistical communication theory and information theory. Summer work at Stanford Research Institute in pattern recognition and on the perceptron type of learning machine led to further work in artificial intelligence. His dissertation, entitled "An Invariant Input for a Pattern Recognition Machine," dealt with the preprocessing of optical patterns to obtain position, orientation, and scale invariant parameters of these patterns.

In March 1962, upon completion of his doctoral studies, Dr. Ball joined ITT Federal Laboratories, where he was Manager of the Artificial Intelligence Group. This group was engaged in work on pattern recognition and learning machines for application to electromagnetic and audio signals.

Dr. Ball rejoined Stanford Research Institute in October 1963. During the past nine months he has developed, in conjunction with other SRI staff members, a new data analysis process called ISODATA. ISODATA (short for Iterative Self-Organizing Data Analysis Technique (A)) automatically, without categorization data, groups together patterns that are similar. The technique appears to have utility for analysis of economic and sociological data as well as physical sciences data and, as importantly, for the design of preprocessing in pattern recognition.

Dr. Ball is a member of the Institute of Electrical and Electronics Engineers, the IEEE Professional Technical Groups on Information Theory and on Electronic Computers, the American Association for the Advancement of Science, Sigma Xi, The Society for Industrial and Applied Mathematics, and the Institute of Mathematical Statistics.

Cover, Dr. Thomas M. - Consultant

Dr. Cover received a B.S. degree in Physics from the Massachusetts Institute of Technology in June 1960. He received an M.S. degree in Electrical Engineering in June 1961 and a Ph.D. in Electrical Engineering in June 1964, both from Stanford University.

During the summers of 1959 and 1960 Dr. Cover was employed in the Ballistics Department of the Grand Central Rocket Company. He was a consultant to the RAND Corporation in the fields of radar and communication theory from June 1961 until April 1964. In the summer of 1963 he was an Engineering Research Assistant at Stanford Research Institute, where he was engaged in the mathematical study of classification capabilities of perceptron-like, adaptive networks. At present, he is a consultant for SRI and an Assistant Professor of Electrical Engineering at Stanford University, where he is engaged in research in statistical data processing and pattern recognition.

Dr. Cover is a member of Sigma Xi.

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 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.

Fossum, Haakon - Research Engineer, Applied Physics Laboratory

Mr. Fossum graduated from the Norwegian Institute of Technology (NIT), Division for Electronic Engineering in June 1960 and received his Diploma of Eng. degree from the same institute in 1961. After his graduation, he held a temporary position as Research Assistant in the Control System Laboratory at NIT, working in the field of electronic instrumentation, until he joined the Norwegian Army in September 1960. During his military service, he was attached to the Norwegian Defense Research Establishment (NDRE), Division for Electronics, Kjeller, Norway, where his field of work included tunnel-diode high-frequency amplifiers, high-frequency pulsed transmitters, analog-to-digital converters, and digital circuit design. After having finished his duty, he continued his work at NDRE as a research engineer in charge of analog simulations and guidance-and-control circuit engineering on an infrared homing missile project.

In 1964, Mr. Fossum received a postgraduate fellowship from the Royal Norwegian Council for Industrial and Scientific Research, in order to pursue advanced studies in the U.S.A. in the field of adaptive control and learning-machine techniques. He joined the Learning Machine Group at SRI in October 1964 and has since been engaged in work on digital simulations of different pattern-recognizing-machine organizations, hardware implementations and analog weight research.

Hall, David J. - Research Engineer, Applied Physics Laboratory

Mr. Hall joined the staff of Stanford Research Institute in April 1962. He received a B.S. degree in Electrical Engineering from the University of Witwatersrand, Johannesburg, South Africa in 1954, and an M.S. degree in Electrical Engineering from London University in 1957.

From 1957 to March 1962 Mr. Hall was Assistant Chief Engineer for F. G. Slack & Company in Johannesburg, where he worked with quotations,

design, and the development and production of a specialized range of industrial electronic products. He is particularly interested in solid-state switching, relay logic, psychometric methods, shaft signalling, mining communication systems, and safety devices with fail-safe characteristics; also in industrial applications of radio isotopes, particularly for density control.

Mr. Hall was in the South African military service from 1950 to 1954, serving as a full lieutenant, with duties in radar, sound ranging, surveying of gun positions, etc. He was awarded the Swan Memorial Scholarship in London (1956) and the British Admiralty Research Grant (1956).

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. While there she was engaged in the formulation of mathematical problems concerning the structure and training of adaptive machines. In June 1963 she transferred to the Mathematical Sciences Department, where 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. She received a B.S. degree in 1960, an M.S. degree in 1962, both in Mathematics, and an M.S. degree in 1964 in Electrical Engineering, all from Stanford University.

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, the American Mathematical Society, and the Mathematical Association of America.

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 on learning machines at Stanford University and at the University of Calif., at Berkeley. He expects to publish in April 1965 a monograph describing 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 communication problems.

Before coming to SRI, Dr. Nilsson completed a three-year term of active duty in the U. S. Air Force. 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, and has 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.

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 directing a program including major projects in microelectronics, learning machines, and artificial intelligence.

In 1940-1943 he served with the British Air Commission dealing with inspection, and technical investigations of aircraft radio systems, components, and instrumentation. From 1943 to 1946, he was successively in charge of the Radio Department, Spot-Weld Engineering, 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, engaged in the development of intercommunication and electronic control systems. From 1950 to 1957 he was employed at the Electronics Laboratory, General Electric Co., Syracuse, New York, and 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 was shortly given responsibility for developing the Applied Physics Laboratory.

His fields of specialization 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).

Dr. Rosen is a Senior Member of the Institute of Electrical and Electronics Engineers, a member of the American Physical Society, and the Scientific Research Society of America.

VII 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 price is detailed in the attached Cost Breakdown.

VIII CONTRACT FORM

It is requested that any contract resulting from this proposal be written as an extension to Contract No. AF 30(602)-3448 between the United States Air Force and Stanford Research Institute.

IX ACCEPTANCE PERIOD

This proposal will remain in effect until 30 June 1965. If consideration of the proposal requires a longer period, the Institute will be glad to consider a request for an extension in time.

X SECURITY CLASSIFICATION

Stanford Research Institute holds a Top Secret facility clearance which may be verified through the cognizant military security agency, Western Contract Management Region (RWIP), AFSC, U. S. Air Force Unit PO, Los Angeles, California 90045. Staff assignments will be in accordance with the level of security assigned to the work.

APPENDIX

A SURVEY OF SEVERAL TRAINABLE PATTERN CLASSIFYING SYSTEMS

by

Nils J. Nilsson

1. Introduction

One result of the last few years of research in artificial intelligence has been the invention of a variety of trainable systems for pattern classification. Many of these systems contain networks of simple trainable units arranged in various ways. The systems are called trainable because they possess adjustable "weights" that can be iteratively modified while the system is being exposed to typical patterns.

There are two important aspects to these systems: structure and dynamics. Structure has to do with the configuration of the basic elements in the system; dynamics concerns the way in which the adjustable weights are modified during training. The purpose of this appendix is to describe the structure of some of the more important systems. For most of the systems to be described here there are several different adjustment or training procedures; we shall refer the reader to the literature for a description of these.

In general, we shall be interested in discussing systems for classifying patterns into R categories, where R is any integer greater than unity. A case of special interest is $R = 2$. In fact we shall treat pattern dichotomizers first and then generalize our descriptions to classifiers for $R > 2$.

2. Pattern Dichotomizers

a. The Dot-Product Unit

A basic element of trainable pattern classifiers is a device, called a dot-product unit, which computes a weighted sum, S , of its inputs.

Suppose the inputs to a dot-product unit are the components of a pattern. Let these components, say, d in number, be denoted by the symbols x_1, x_2, \dots, x_d . Then, the dot-product unit computes the following weighted sum:

$$S = w_1 x_1 + w_2 x_2 + \dots + w_d x_d + w_{d+1} \quad (1)$$

The coefficients $w_1, w_2, \dots, w_d, w_{d+1}$, known as weights, are adjustable. The $(d+1)$ th weight, w_{d+1} , does not multiply one of the pattern components. It can be interpreted as an adjustable bias component of the sum. Often it is included as the coefficient of a $(d+1)$ th input x_{d+1} whose value is always equal to one.

If the pattern inputs are represented by a d -dimensional point or vector $\vec{X} = (x_1, x_2, \dots, x_d)$ and the weights are represented by a d -dimensional point or vector, $\vec{W} = (w_1, w_2, \dots, w_d)^*$, then the sum, S , includes the dot-product of these two vectors:

$$S = \vec{X} \cdot \vec{W} + w_{d+1} \quad (2)$$

hence the name dot-product unit.

A dot-product unit is illustrated in Fig. 1(a). Since it is such a common building block in pattern classifiers, we shall use the special symbol in Fig. 1(b) to represent it in future diagrams.

b. The Threshold Logic Unit (TLU) and Linear Machines

Perhaps the simplest pattern dichotomizer is the threshold logic unit (TLU). A trainable TLU, illustrated in Fig. 2, consists of a dot-product unit with adjustable weights and a threshold element. For patterns of d components (x_1, x_2, \dots, x_d) , the TLU can have up to $(d+1)$ inputs,

* By a point \vec{X} (or \vec{W}) we mean that point in d -dimensional Euclidean space whose coordinates are equal to the input components (or weights). By a vector \vec{X} (or \vec{W}) we mean that vector directed from the origin of d -dimensional Euclidean space to the point \vec{X} (or \vec{W}).

each of which is weighted before summing. The first d inputs are the components of the pattern, and the $(d+1)$ th input, x_{d+1} , always has the value one. Thus the value of the threshold can be fixed at zero, and the effective threshold is determined by the weight, w_{d+1} . Then if the sum, S , of the $(d+1)$ weighted inputs exceeds zero, the TLU responds with a $+1$ output, and if the sum is less than zero, the TLU responds with a -1 output. These two possible outputs are then associated with the two pattern categories.

Various methods^{11,18,19} exist for training the TLU, by iterative weight adjustment, so that it will appropriately dichotomize a set of training patterns. Some information about the effectiveness of a TLU as a pattern dichotomizer is also available. Specifically, the number of random patterns that a TLU can be trained to dichotomize (this number is called the "capacity" of a TLU) is known to be twice the number of adjustable weights.²⁰

Many of the properties of the TLU pattern dichotomizer can best be described in geometric terms. For any given set of weights, the equation

$$\vec{W} \cdot \vec{X} + w_{d+1} = 0 \quad (3)$$

defines the locus of points (representing patterns) producing an undefined or border-line TLU output. This locus is a surface called a hyperplane and divides the space of pattern points into two regions: points causing $+1$ TLU outputs are on the "positive" side of the hyperplane, and those causing -1 TLU outputs are on the other (or negative) side.

The TLU has an interesting generalization for $R > 2$, but to recognize this generalization we must first modify the TLU into an apparently different, but actually equivalent, structure, which we shall now do. Suppose we consider a dichotomizer with two sets of adjustable weights instead of one. We shall represent these sets by the two vectors $\vec{W}_1 = (w_{11}, \dots, w_{d1})$ and $\vec{W}_2 = (w_{12}, \dots, w_{d2})$ along with the $(d+1)$ th weights $w_{d+1,1}$ and $w_{d+1,2}$. This dichotomizer forms two weighted sums

$$S_1 = \vec{W}_1 \cdot \vec{X} + w_{d+1,1}$$

and

$$S_2 = \vec{W}_2 \cdot \vec{X} + w_{d+1,2}$$

(4)

The two dot-product units are followed by a maximum selector, and classification of a pattern \vec{X} is made by noting the larger of these two sums. If $S_1 > S_2$, the dichotomizer responds with an output of +1; if $S_1 < S_2$, the dichotomizer responds with an output of -1. These two possible outputs are then associated with the two pattern categories. This dichotomizer is illustrated in Fig. 3. It is called a linear machine because the pattern components occur linearly (rather than raised to some power, say) in the sums S_1 and S_2 .

One can easily see that the TLU and the linear machine are identical if

$$\vec{W} = \vec{W}_1 - \vec{W}_2$$

and

$$w_{d+1} = w_{d+1,1} - w_{d+1,2}$$

(5)

The TLU is the more economical to construct and train since there is only one set of weights; the linear machine (for $R = 2$) has pedagogic interest because it motivates our later discussions of R-category classifiers.

c. Piecewise Linear (PWL) Machines

An important generalization of the linear machine is the piecewise linear (PWL) machine. For $R = 2$, it consists of two banks of dot-product units. Suppose there are L_1 dot-product units in the first bank and L_2 in the second bank. Then there will be a total of $L_1 + L_2$ sets of adjustable weights. Each set can be represented by a weight vector $\vec{W}_i = (w_{1i}, \dots, w_{di})$, $i = 1, \dots, L_1 + L_2$, and a $(d+1)$ th weight, $w_{d+1,i}$, $i = 1, \dots, L_1 + L_2$. For any given pattern \vec{X} , the $(L_1 + L_2)$ sums are then given by the expressions

$$S_i = \vec{W}_i \cdot \vec{X} + w_{d+1,i} \quad \text{for } i = 1, \dots, L_1 + L_2. \quad (6)$$

These dot-product units are followed by a maximum selector. If, for a pattern \vec{X} , the largest sum belongs to the first bank, the dichotomizer responds with an output of +1; if the largest sum belongs to the second bank, the dichotomizer responds with a -1. These two possible outputs are then associated with the two pattern categories.

The PWL machine is illustrated in Fig. 4. Training methods²¹ exist for adjusting the weights of a PWL machine, although the properties of these methods are not yet fully understood. The PWL machine bases its classification of any pattern on a comparison (a correlation) of this pattern against a multiplicity of stored prototype patterns (the weight vectors) for each category. On the other hand, the simpler linear machine or TLU compares the input pattern against only one stored pattern for each category.

d. Committee Machines

A committee machine is a dichotomizer that bases its classification on the majority vote of an odd number of trainable TLUs. Each TLU responds with a +1 or -1 to any pattern, and a "vote-taker" sums the outputs of the TLUs. This sum must be either positive or negative since there are an odd number of TLUs. If the sum is positive, the committee machine responds with an output of +1; if the sum is negative the committee machine responds with an output of -1. These two outputs are then associated with the two pattern categories.

The committee machine is illustrated in Fig. 5. This organization has been used in at least two pattern-recognition devices.^{22,23} Methods have been developed and tested that permit training²⁴ of the committee machine, and some theoretical knowledge of the properties of the machine has been gained.¹⁴

3. Pattern Classifiers for $R > 2$

a. Parallel Organizations of Dichotomizers

One approach to the design of pattern-classifying systems for $R > 2$ uses a parallel organization of dichotomizers. The dichotomizers may be of any type; for example they may be any of those described in

Sec. 2 of this appendix. A classifier employing some number, say q , of independent dichotomizers can classify patterns into as many as 2^q different categories since there are 2^q different possible output combinations. This type of dichotomizer uses a code-converter to implement a coding scheme whereby the R pattern categories are associated with the 2^q possible dichotomizer output combinations. Such a classifier is illustrated in Fig. 6.

It has been found useful to use a value of q such that 2^q is much larger than the number of categories R . Such a redundant use of dichotomizers permits the use of error-correcting codes. For example, suppose $q = 9$ and $R = 32$. There exists a single-error-correcting 9-bit code with 5 information bits.²⁵ For each of the 32 code words belonging to this code there corresponds one of the pattern categories. This correspondence defines the desired outputs for each of the 9 dichotomizers which can then be trained simultaneously. When a pattern is presented to this machine the response of 5 of the dichotomizers can be used to specify the 32 categories after any single errors have been corrected in the complete 9-bit output.

Another type of code employs as many dichotomizers as there are categories ($q = R$). Each dichotomizer is trained to be a specialist. The i th dichotomizer then would respond with a +1 to all patterns in category i and with a -1 to all other patterns. Here we must interpret illegal response combinations (all R dichotomizers responding with a -1, or more than one dichotomizer responding with a +1) as a reject or confused decision.

One of the outstanding research problems connected with the parallel dichotomizer organization is the question: which coding scheme to use? Very little is yet known, for example, about the trade-off between redundancy in the number of dichotomizers and accuracy of classification.

b. Parallel Organizations of TLUs

In implementations²³ of the parallel dichotomizer systems just discussed, each constituent dichotomizer is usually a structure more complicated than a single TLU; for example, each might be a committee

machine or a PWL machine. The tendency to concentrate on these more complex dichotomizers has obscured an important special case in which each dichotomizer consists of a single TLU. In this case it would seem necessary to compensate for the simplicity of the constituent dichotomizers by a more redundant use of them. Thus, 2^q is usually very much larger than R .

Some interesting coding schemes have been proposed to associate the different combinations of TLU responses with the pattern categories. One such scheme, employed by a machine organization called Tokoloshe¹⁷ used $q = 63$ and $R = 64$. There exists a code consisting of 64 code words, each of 63 bits and each code word differing from the other code words in precisely 32 bits. These 64 code words were used to define the desired response of the 63 TLUs for each of the 64 pattern categories. In using such a pattern classifier, the properties of the code permit up to 15 errors in the individual TLUs without incurring a classification error.

Similar codes exist for any q such that $q = 2^p - 1$ where p is any integer. They are called maximal-length shift-register sequences²⁶ and have the following properties:

- (1) The code length is $2^p - 1$ bits.
- (2) There are 2^p distinct code words.
- (3) Each code word differs from all the others in precisely 2^{p-1} bits allowing up to $[2^{p-2} - 1]$ errors.

Additional redundancy in the number of TLU dichotomizers can be obtained by using codes employing a higher ratio of word length to number of words. For example, a maximal-length shift-register sequence code with $2^p > R$ could be used. Suppose $p = 6$ and $R = 16$. In this case the 64 different code words might be assigned to the 16 categories by associating 4 code words with each category. This extra redundancy permits a larger number of permissible combinations of TLU responses for each category.

Other types of codes might also be useful for this type of organization. It has been suggested that it would be appropriate for the code words corresponding to the pattern categories to differ from each other in a manner that reflects the relative importance of each category and the relative costs of errors. Thus, if it is important to avoid confusing categories i and j , the corresponding code words should differ in a large number of bits.

Except for some of these intuitive suggestions, which experience has not yet raised above the level of doubt, very little is yet known about how best to provide the requisite redundancy of TLUs in such a classifier.

It is interesting to note that a pattern classifier consisting of parallel committee machine dichotomizers could also be interpreted as a parallel organization of single TLU dichotomizers. This is so because each committee dichotomizer consists of a layer of parallel TLUs. The TLUs from the different dichotomizers can be lumped together and a code specified that would yield the same category decisions as does the combination of committee machines.

c. Linear and Piecewise Linear Machines

Both the linear and piecewise linear pattern dichotomizers have obvious generalizations for $R > 2$. In the case of the linear machine there are R dot-product units each with adjustable weight sets. The weight sets can be represented by weight vectors $\vec{W}_i = (w_{1i}, w_{2i}, \dots, w_{di})$, $i = 1, \dots, R$, and $(d+1)$ th weights $w_{d+1,i}$, $i = 1, \dots, R$. For any patterns, \vec{X} , these dot-product units compute R sums given by the expressions:

$$S_i = \vec{W}_i \cdot \vec{X} + w_{d+1,i} \quad i = 1, \dots, R \quad (7)$$

The dot-product units are followed by a maximum selector that indicates which sum is the largest. If the i_0 -th sum is largest, the input pattern is assigned to category i_0 .

The R-category linear machine is illustrated in Fig. 7. Its pattern-classifying properties are fairly well understood,²⁰ and methods do exist for training it.¹⁸ This structure has been suggested often in the pattern-recognition literature. It is called a learning matrix by Steinbuch²⁷ and has also been investigated by Griffin, et al.²⁸

The piecewise linear machine for $R > 2$ consists of R banks of dot-product units, each with adjustable weight sets. Suppose that the j -th bank contains L_j dot-product units so there are $\sum_{j=1}^R L_j$ of them in all. Let us represent each weight set by a weight vector $\vec{W}_i = (w_{1i}, \dots, w_{di})$, $i = 1, \dots, R$ and a $(d+1)$ th weight $w_{d+1,i}$, $i = 1, \dots, R$. For any pattern, S , the dot-product units compute $\sum_{j=1}^R L_j$ sums given by the expressions:

$$S_i = \vec{W}_i \cdot \vec{X} + w_{d+1,i} \quad i = 1, \dots, \sum_{j=1}^R L_j \quad (8)$$

The dot-product units are followed by a maximum selector that indicates which bank contains the largest sum. If the i_0 -th bank contains the largest sum, the input pattern is assigned to category i_0 .

The R-category PWL machine is illustrated in Fig. 8. Some methods have been proposed and tested for training it,²¹ while research continues on more efficient training methods.

Of the various methods suggested for organizing R-category pattern-classifying machines, it is not yet known which, if any, are generally preferable. A wide choice confronts the designer of trainable pattern recognition systems, and more experimenting and research with these structures are necessary before wise selections can be made in any given situation.

4. Φ -Preprocessors

Each of the pattern classifiers that we have considered so far used the same basic element: a dot-product unit.

If the weight set is represented by the vector \vec{W} and a $(d+1)$ th weight w_{d+1} , then for any pattern S , the dot-product unit computes

$$S = \vec{X} \cdot \vec{W} + w_{d+1} \quad (9)$$

The adjustable parameters of this sum (i.e., the components of \vec{W} and w_{d+1}) occur linearly in S , and many of the theoretical results concerned with training and capacity depend only on this fact. These results are easily generalized²⁰ to machines containing elements that compute weighted sums of functions of the components of \vec{X} .

If the pattern \vec{X} is transformed by some non-adjustable vector-valued function $\vec{\Phi}(\vec{X})$ into a new set of components, these new components can be used as the input to the trainable pattern classifier often with enhanced results. We speak of any device making such a transformation as a Φ -processor.

An example of a Φ -processor is a device which for any pattern \vec{X} with d components (x_1, x_2, \dots, x_d) produces a set of $\frac{d(d+3)}{2}$ new components which are all the products of the x_i taken up to two at a time (i.e., $x_1^2, x_2^2, \dots, x_d^2, x_1x_2, x_1x_3, \dots, x_{d-1}x_d, x_1, x_2, \dots, x_d$). These $\frac{d(d+3)}{2}$ components can then be applied to the dot-product units of a pattern classifier. If the pattern classifier is a TLU, the combination of a Φ -processor and a trainable TLU can implement hyper-surfaces much more complex than the hyperplane implemented by a TLU alone. For example, when products of up to two of the x_i are computed by a Φ -processor, the tandem combination can implement general hyperquadric surfaces that have significantly more power to dichotomize pattern sets than does the simpler hyperplane.

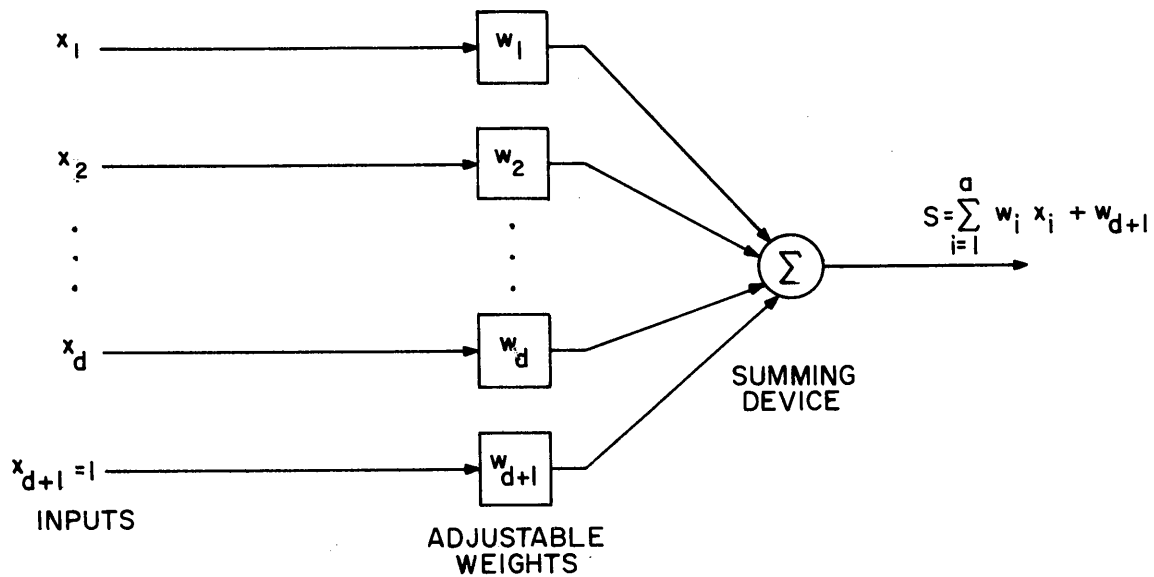
Thus, the Φ -processor is a technique to enhance the classification power of relatively simple trainable organizations. Additional research is needed to determine whether or not it competes favorably with the techniques involving the more complex trainable organizations that we have already described.

REFERENCES

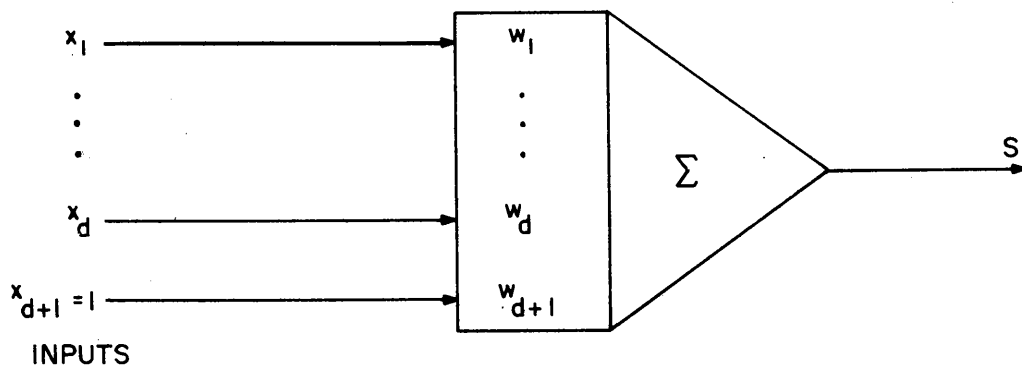
1. Brain, A.E., "The Simulation of Neural Elements by Electrical Networks based on Multi-Aperture Cores," Proc. IRE, 49, pp. 49-52 (January 1961).
2. Crafts, H. S., "Design of a Magnetic Variable-Gain Component for Adaptive Networks," 1963 WESCON, San Francisco, California.
3. Rosen, C. A., et al., "A Large, Self-Contained Learning Machine," 1963 WESCON, San Francisco, California.
4. Duda, R. O., and J. W. Machanik, "An Adaptive Prediction Technique and its Application to Weather Forecasting," 1963 WESCON, San Francisco, California.
5. Block, H. D., N. J. Nilsson, and R. O. Duda, "Determination and Detection of Features in Patterns," Technical Documentary Report No. RADC-TDR-63-467, Information Processing Branch, Rome Air Development Center, Research and Technology Division, Air Force Systems Command, Griffiss Air Force Base, New York (December 1963).
6. Ball, G. H. and D. J. Hall, "Some Fundamental Concepts and Synthesis Procedures for Pattern Recognition Preprocessors," International Conference on Microwaves, Circuit Theory, and Information Theory, Tokyo, Japan (September 1964).
7. 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).
8. Singleton, R. C., "A Test for Linear Separability as Applied to Self-Organizing Machines," Technical Report, SRI Project 3605, Contract 3438(00), Stanford Research Institute, Menlo Park, California (May 1963).
9. Nilsson, N. J. and R. C. Singleton, "An Experimental Comparison of Three Learning Machine Training Rules," Technical Documentary Report No. RADC-TDR-64-33, Information Processing Branch, Rome Air Development Center, Research and Technology Division, Air Force Systems Command, Griffiss Air Force Base, New York (February 1964).
10. Efron, B., "The Perceptron Correction Procedure in Non-Separable Situations," Technical Documentary Report No. RADC-TDR-63-533, Information Processing Branch, Rome Air Development Center, Research and Technology Division, Air Force Systems Command, Griffiss Air Force Base, New York (February 1964).
11. Duda, R. O. and R. C. Singleton, "Training a Threshold Logic Unit with Imperfectly Classified Patterns," 1964 WESCON, Los Angeles, California.

12. Cover, T. M., "Classification and Generalization Capabilities of Linear Threshold Units," Technical Documentary Report No. RADC-TDR-64-32, Information Processing Branch, Rome Air Development Center, Research and Technology Division, Air Force Systems Command, Griffiss Air Force Base, New York (February 1964).
13. Kaylor, D. J., "A Mathematical Model of a Two-Layer Network of Threshold Elements," Technical Documentary Report No. RADC-TDR-63-534, Information Processing Branch, Rome Air Development Center, Research and Technology Division, Air Force Systems Command, Griffiss Air Force Base, New York (March 1964).
14. Ablow, C. M. and D. J. Kaylor, "A Committee Solution of the Pattern Recognition Problem," Rome Air Development Center, Technical Documentary Report (to be published in 1965).
15. Nilsson, N. J., "Linear Separability of Signal Space as a Basis for Adaptive Mechanisms," Technical Documentary Report No. RADC-TDR-64-145, Information Processing Branch, Rome Air Development Center, Research and Technology Division, Air Force Systems Command, Griffiss Air Force Base, New York (May 1964).
16. Nilsson, N. J., Learning Machines: Foundations of Trainable Pattern Classifying Systems (McGraw-Hill Book Company, New York; to be published in March, 1965).
17. Brain, A. E., et al., "Graphical Data Processing Research Study and Experimental Investigation," Report No. 8 (pp. 9-13) and No. 9 (pp. 3-10), Contract DA 26-039 SC-78343, Stanford Research Institute, Menlo Park, California (June 1962 and September 1962).
18. Nilsson, N. J., Learning Machines: Foundations of Trainable Pattern Systems, Chapters 4 and 5.
19. Widrow, B., and M. E. Hoff, "Adaptive Switching Circuits," Stanford Electronics Laboratories Technical Report No. 1553-1, Stanford University, Stanford, California (June 30, 1960).
20. Nilsson, N. J., Learning Machines: Foundations of Trainable Pattern Classifying Systems, Chapter 2.
21. Nilsson, N. J., Learning Machines: Foundations of Trainable Pattern Classifying Systems, Chapter 7.
22. Widrow, B., "Generalization and Information Storage in Networks of Adaline 'Neurons'," Self-Organizing Systems--1962, Yovits, Jacobi and Goldstein (Eds.), pp. 435-461 (Spartan Press, Washington, D.C., 1962).
23. Brain, A. E., et al., "A Large, Self-Contained Learning Machine," 1963 WESCON paper No. 6.1, August, 1963.

24. Nilsson, N. J., Learning Machines: Foundations of Training Pattern Classifying Systems, Chapter 6.
25. Brain, A. E., et al., "Graphical Data Processing Research Study and Experimental Investigation," Report No. 17, pp. 9-10, Contract DA 36-039 AMC-03247(E), Stanford Research Institute, Menlo Park, California (October 1964).
26. Peterson, W. W., Error-Correcting Codes, pp. 147 ff (M.I.T.Press, Cambridge, Massachusetts, and John Wiley & Sons, Inc., New York, 1961).
27. Steinbuch, K. and U. A. W. Piske, "Learning Matrices and Their Applications," IEEE Trans. EC-12, 5, pp. 846-862 (December 1963).
28. Griffin, J. S., et al., "Pattern Identification Using Adaptive Linear Decision Function," paper No. 6.4 presented at 1963 WESCON, August 20-23, 1963.



(a)
ELEMENTS OF A DOT-PRODUCT UNIT



(b)
SCHEMATIC FOR A DOT-PRODUCT UNIT

FIG. 1. A DOT-PRODUCT UNIT

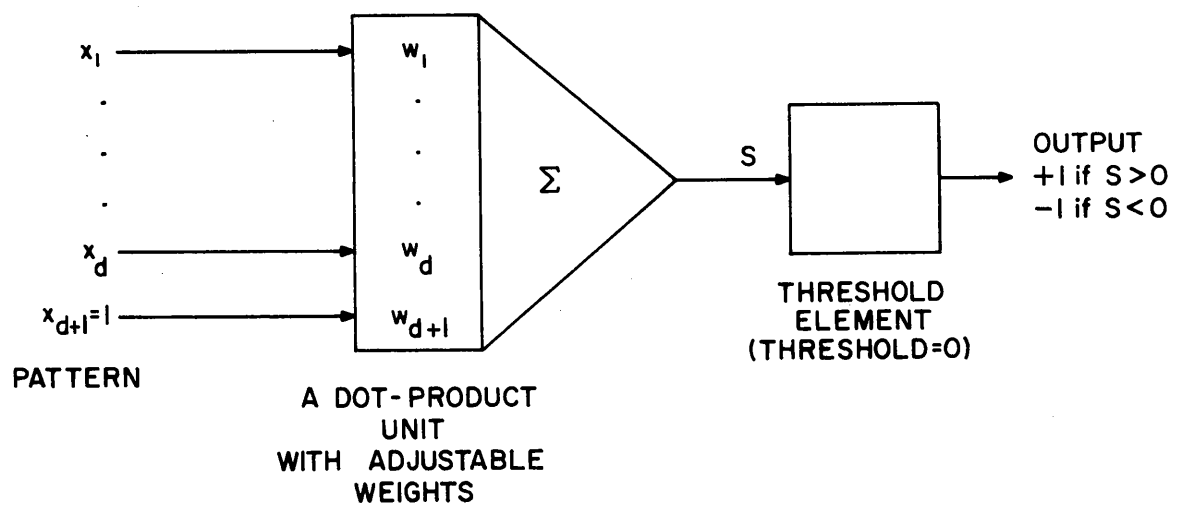


FIG. 2 A THRESHOLD LOGIC UNIT (TLU)

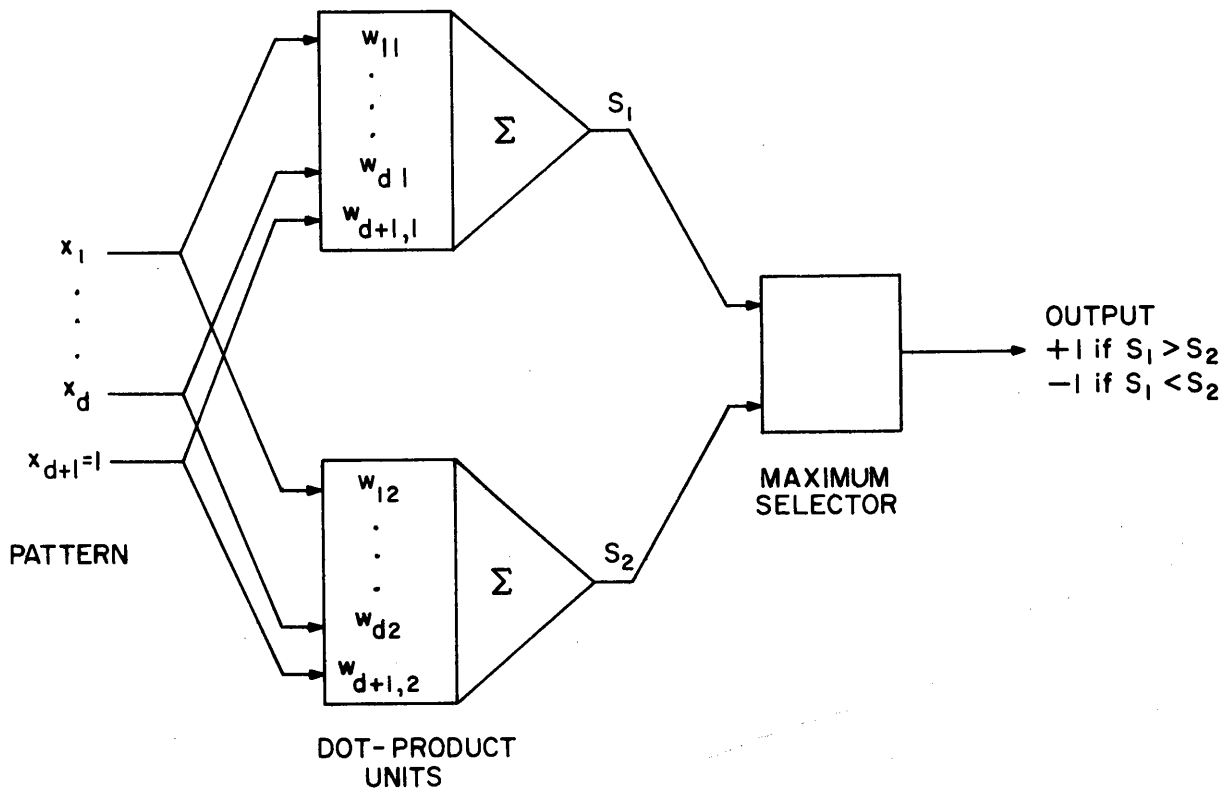


FIG. 3 A LINEAR MACHINE ($R = 2$)

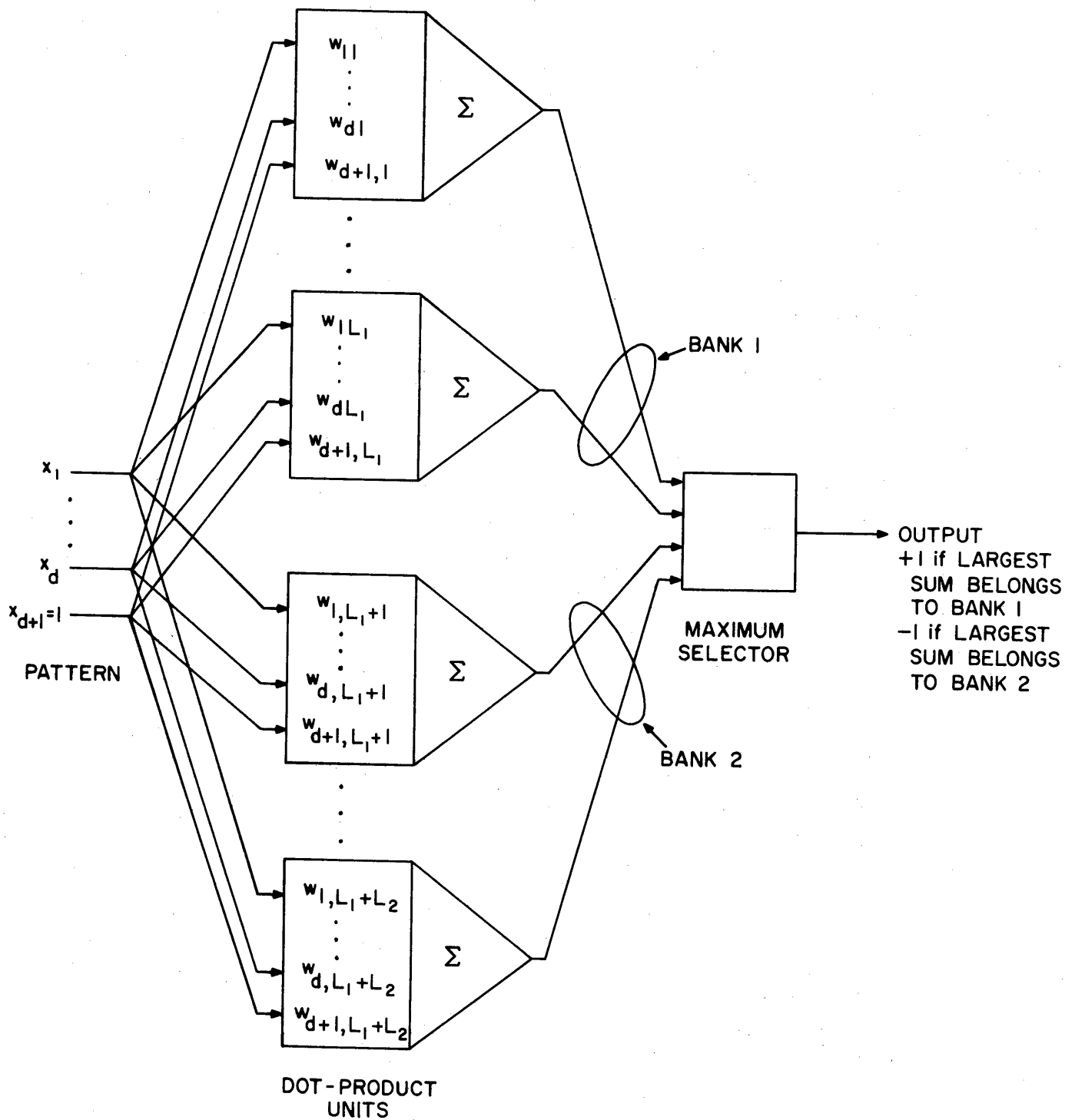


FIG. 4 A PIECEWISE LINEAR (PWL) MACHINE

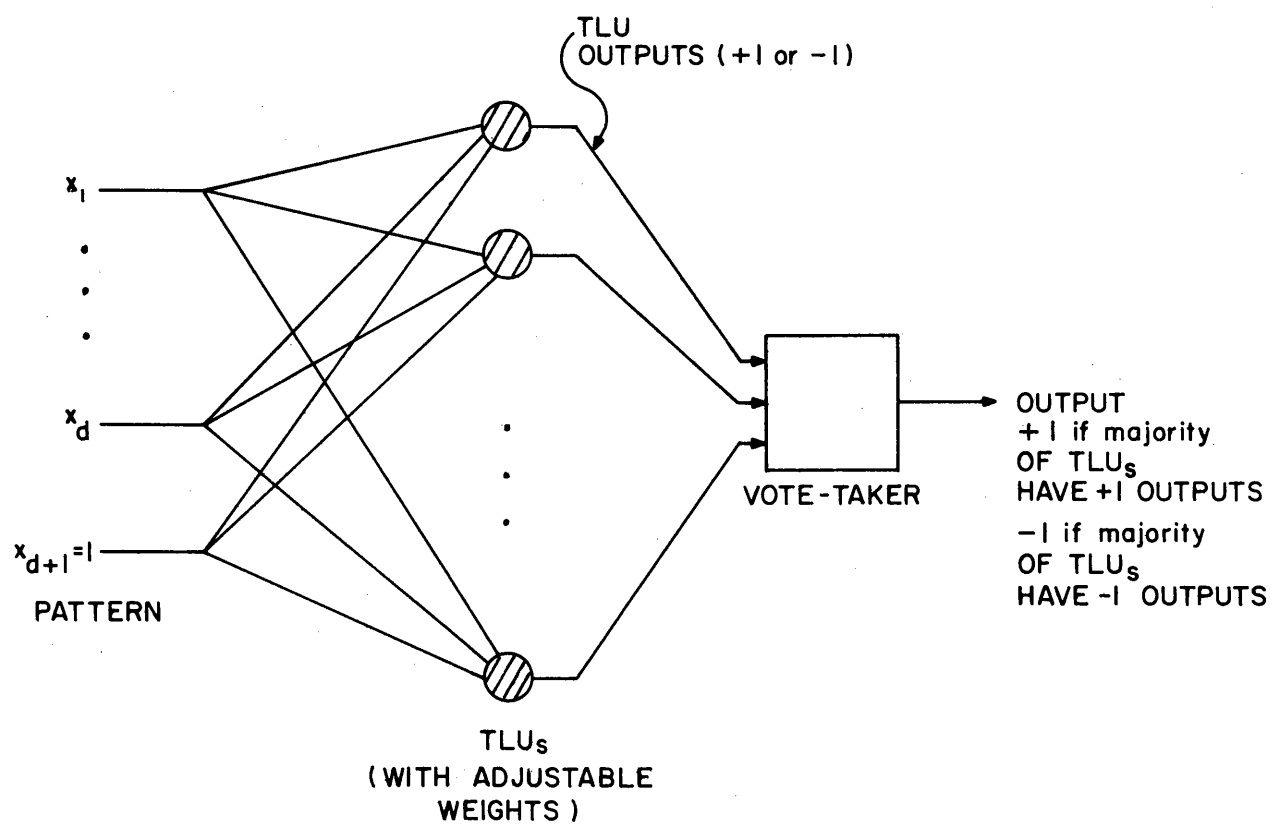
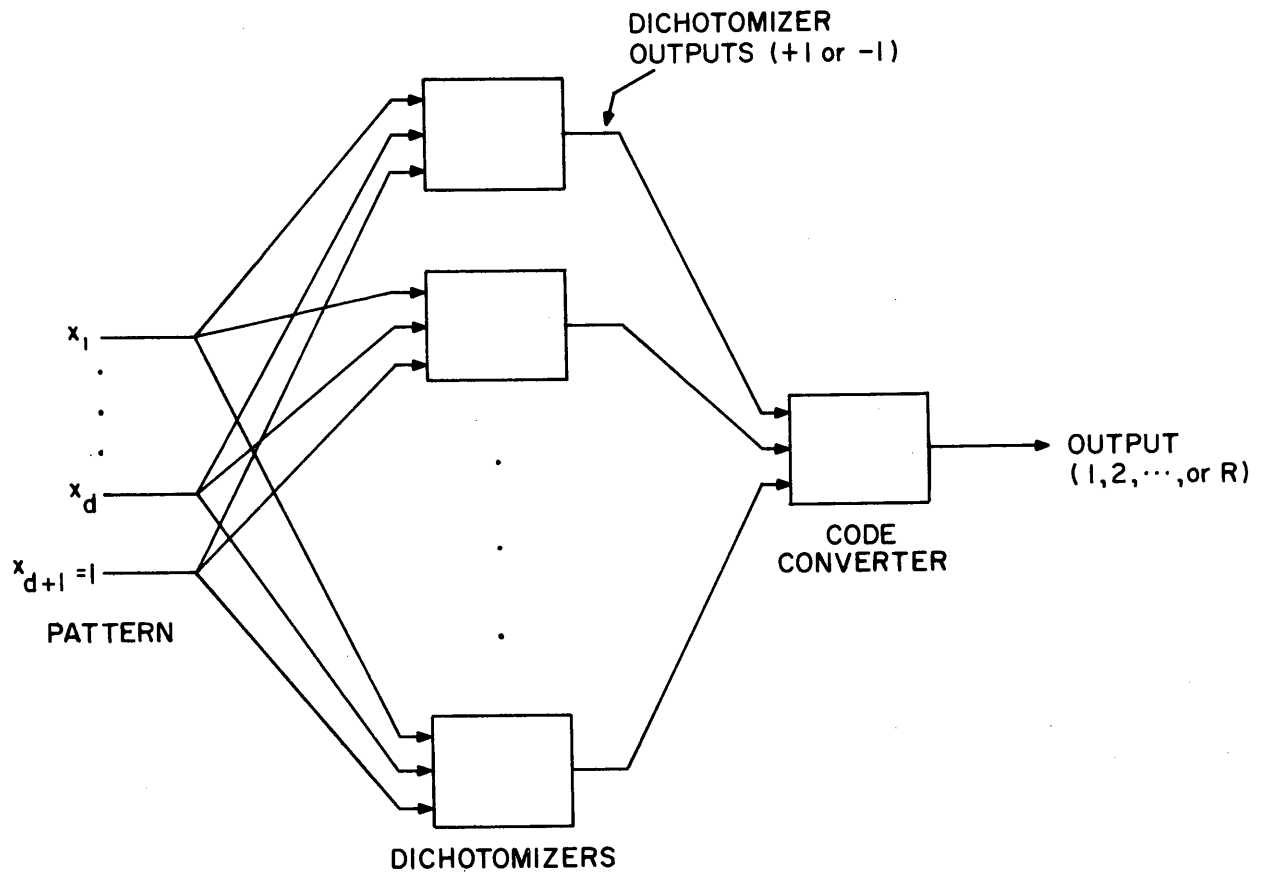


FIG. 5 A COMMITTEE MACHINE



- NOTE: 1. THE DICHOTOMIZERS MAY BE OF ANY TYPE. A SPECIAL CASE OF INTEREST OCCURS WHEN THEY ARE SINGLE TLU_s .
2. THE CODE CONVERTER MAY IMPLEMENT VARIOUS CODING SCHEMES INCLUDING ERROR-CORRECTING AND "SPECIALIST" CODES.

FIG. 6 AN R-CATEGORY CLASSIFIER USING
PARALLEL DICHOTOMIZERS

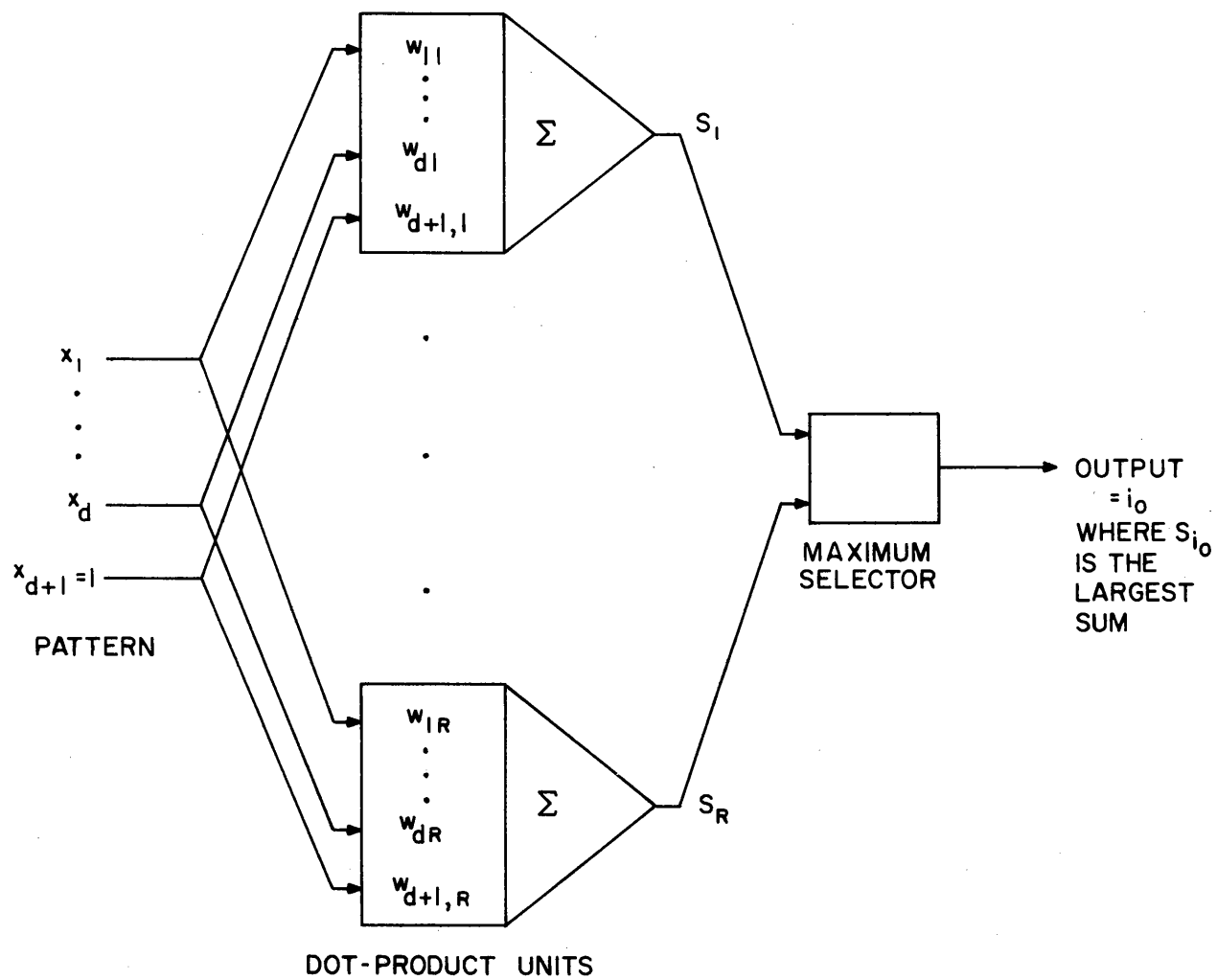


FIG. 7 AN R-CATEGORY LINEAR MACHINE

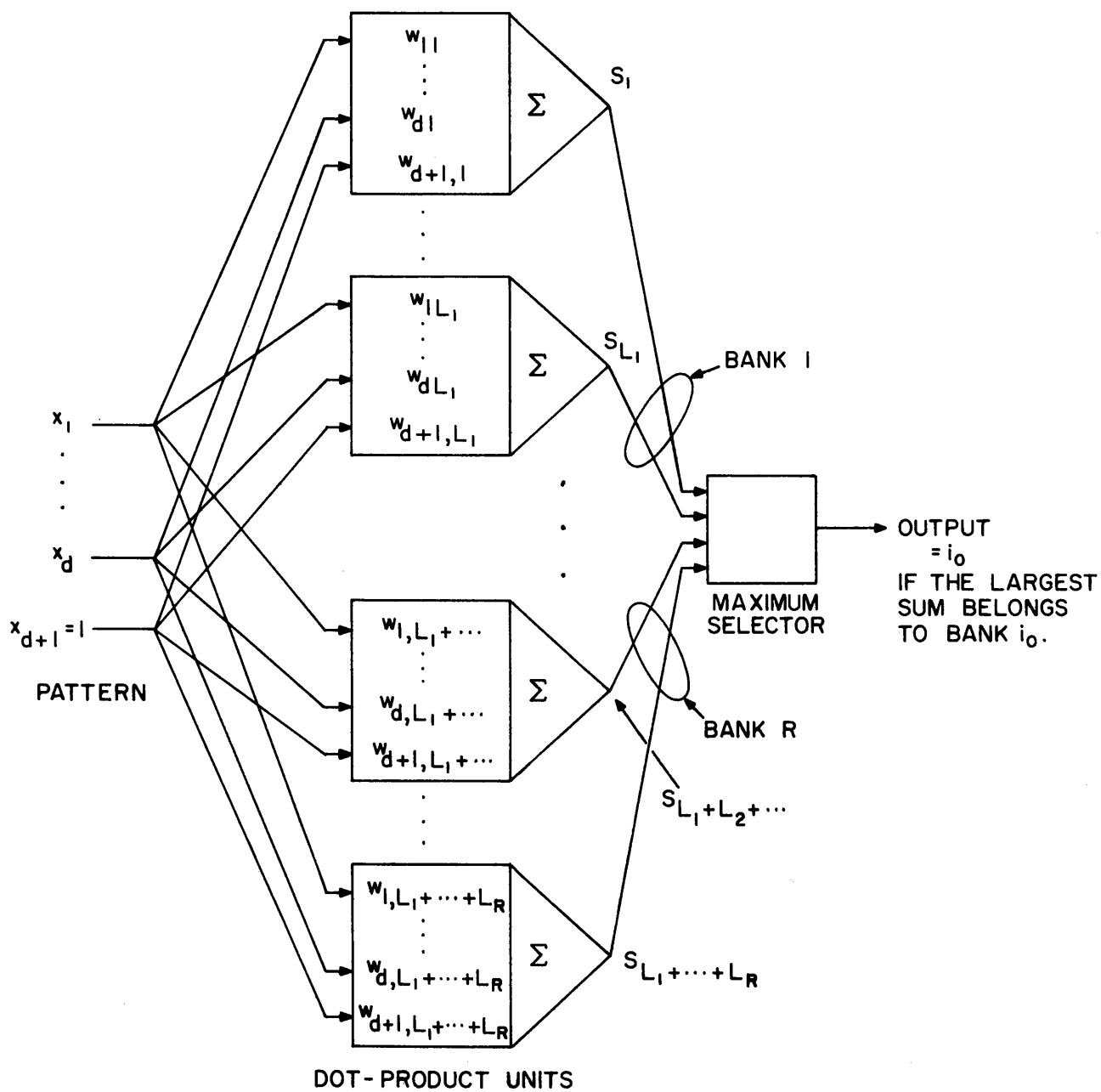


FIG. 8 AN R-CATEGORY PNL MACHINE

Proposal for Research
SRI No. ESU 65-12
Extension to Contract AF 30(602)-3448

9 March 1965

COST BREAKDOWN

Personnel Costs

Supervisory, 84 hours @ \$[REDACTED]/hour	\$ [REDACTED]
Research Engineer, 2020 hours @ \$[REDACTED]/hour	[REDACTED]
Research Mathematician, 673 hours @ \$[REDACTED]/hour	[REDACTED]
Editorial, 84 hours @ \$[REDACTED]/hour	[REDACTED]
Secretarial, 336 hours @ \$[REDACTED]/hour	[REDACTED]
Programmers, 504 hours @ \$[REDACTED]/hour	[REDACTED]

Total Direct Labor

Payroll Burden at 16%*

Total Salaries and Wages

Overhead at 86% of Direct Labor*

Total Personnel Costs

Direct Costs

Travel and Subsistence:

3 cross-continental trips @ [REDACTED]/trip

6 days subsistence @ [REDACTED]

Shipping and Communications

Computer time, 15 hours on 7090

(or equivalent) @ [REDACTED]/hour

Report Production Costs

Consultant, 30 days @ [REDACTED]/day

Total Direct Costs

Total Estimated Costs

Fee

Total Price

*The rates quoted are those currently approved for billing and estimating purposes. It is requested that contracts provide for provisional reimbursement at rates acceptable to the Contracting Officer 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.