COGNITIVE SYSTEMS RESEARCH PROGRAM

CORNELL UNIVERSITY ITHACA, N. Y.

REPORT NO. 4

COLLECTED TECHNICAL PAPERS
VOLUME 2

Edited By

F. ROSENBLATT

30 July, 1963

Prepared Under Contract No. NONR 401 (40)
and NSF GP-971.

CSRP

5 228 100

COGNITIVE SYSTEMS RESEARCH PROGRAM

Cornell University Ithaca, New York

14 Report No. 4

6 COLLECTED TECHNICAL PAPERS, VOLUME 2,

10 ed. ~~Edited~~ by

F. Rosenblatt

11 30 July, 1963

15 Prepared under Contract ~~No.~~ NONR 401(40) and NSF GP-971. Reproduction, translation, use and disposal in whole or in part by or for the United States Government is permitted.

TABLE OF CONTENTS

PREFACE

The Cognitive Systems Research Program, which currently employs a staff of about twenty people, is primarily concerned with the study of models of central nervous system functioning, and the testing and verification of such models by means of biological experiments. The twelve papers in this volume present the results of work performed under the Cognitive Systems Research Program since the publication of the first volume of Collected Technical Papers in January, 1961. The papers are divided into the following main groups, reflecting the principal areas of research conducted by the Program:

(1) Mathematical theory of neural networks (the first four papers).
(2) Biochemistry (one paper).
(3) Simulation and numerical experiments (four papers).
(4) Physical models and engineering studies (two papers).
(5) A bibliography of published work on perceptrons.

Considerable additional work in most of these categories has been accomplished during the period covered by this volume, but has either been reported elsewhere (see the bibliography) or will be described in separate reports to be issued shortly. In particular, detailed reports on the following topics are currently in preparation:

(1) Circuit designs for the Tobermory perceptron (by George Nagy).
(2) Manual for a general purpose perceptron simulation program (by Trevor Barker).
(3) Biochemical studies of neural transmitter substances (by Roger Marchbanks).
(4) Simulation studies of adaptive four-layer and cross-coupled perceptrons (by Frank Rosenblatt).
(5) Theory of biological memory mechanisms (by F. Rosenblatt).

The editor wishes to take this opportunity to express his appreciation for the work performed by all members of the CSRP staff, as well as the assistance of such collaborating organizations as the Courant Institute at New York

University, and the Stanford Research Institute, which have contributed substantially to the research reported here.

Frank Rosenblatt
Director, CSRP
July 30, 1963

A NOTE ON PROBABILISTIC CONVERGENCE IN ADAPTIVE FOUR-LAYER PERCEPTRONS*

By Harry Kesten

1. Introduction

Assume that there are n different stimuli. A sequence of stimuli is shown to a four-layer perceptron. We assume that the sequence of stimuli is a Markov chain, and that one stimulus is shown every Δt units of time. The procedure is as described in Principles of Neurodynamics, Section 16.2.1. The following small changes were made:

i) We have written $n_{i,j}$ for $N_a Q_{i,j}^{(1)}$

ii) The function ϕ was changed into a function ϕ satisfying $0 \leq \phi(x) \leq 1$, and

$$|\phi(x_1) - \phi(x_2)| \leq c_1 |x_1 - x_2| .$$

---

* EDITOR'S NOTE: When the analysis of adaptive four-layer perceptrons was first published (Block, Knight, and Rosenblatt, Reviews of Modern Physics, 34, 135-142, 1962) it was pointed out by a number of readers that convergence had not been demonstrated rigorously for the case of a random preconditioning sequence. Several critics, in fact, felt that the system would not converge at all under these conditions, since (if one waited long enough) a "disruptive" sequence of stimuli would be certain to occur, of sufficient duration to force the perceptron out of its steady-state condition. It was proposed by Rosenblatt that it should be possible to prove a theorem demonstrating that the perceptron would still converge in a weaker, probabilistic sense; namely, that by taking Δt, the time interval between successive stimuli, small enough, it should be possible to guarantee that the probability that the state of the perceptron differs from the previously predicted state at time t can be made arbitrarilly small. An appeal was made to Prof. Kesten for assistance in proving this conjecture, and the above paper is the result. (It is assumed that the reader is familiar with the earlier work by Block, Knight, and Rosenblatt. The version refered to by Kesten appears as Chapter 16 in Rosenblatt, Principles of Neurodynamics, Spartan Books, Washington, D.C., 1962.)

It is worth noting that the introduction of the Lipschitz condition for ϕ, in Kesten's analysis, is actually a step closer to a physically realistic system, since the assumption of an abrupt discontinuity of neural activity at the threshold is replaced by the assumption of a neuron which gradually "becomes active" as the input signal is increased. The frequency responses of biological neurons would, of course, satisfy this condition.

Under these circumstances it is shown that the solutions $\xi^{(i)}$ of the difference equations (3) (and boundary conditions (2)) are with large probability close to the solutions of the differential equation (10) (with boundary conditions (11)), when Δt is sufficiently small.

It is perhaps unpleasant that ϕ is required to satisfy a Lipschitz condition. The estimates show, however, that c_1 may be replaced by a function which depends on Δt, and that the function, $c_1(\Delta t)$, may tend to infinity as $\Delta t \to 0$, if $c_1(\Delta t).\Delta t$ does not increase too fast.

2. Analysis

Let ϕ be a function satisfying

(1) $$|\phi(x_1) - \phi(x_2)| \leq C_1|x_1 - x_2|, \quad 0 \leq \phi \leq 1$$

where C_1 is some positive constant.

For a given sequence of integers $j_{-1}, j_o, j_1, j_2, \ldots$ $(1 \leq j_k \leq n =$ total number of different stimuli) define $\xi^{(i)}(k)$ for $1 \leq i \leq n$ by

(2) $$\xi^{(i)}(0) = 0$$

(3) $$\xi^{(i)}(m+2) - \xi^{(i)}(m+1) = \eta.\Delta t.\phi[\beta^{(j_{m+1})} + \xi^{(j_{m+1})}(m+1)]\, n_{i,j_m} - \delta.\Delta t.\xi^{(i)}(m+1$$

where η, δ, Δt are given positive constants, $n_{i,j}$, $\beta^{(j)}$ are non-negative constants for $1 \leq i \leq n$, $1 \leq j \leq n$. (This clearly is the analogue of (16.6) Principles of Neurodynamics.)

LEMMA 1: If ξ is defined by (2) and (3), and if $\delta\, \Delta t \leq 1$, then

(4) $$0 \leq \xi^{(i)}(m+1) \leq (\frac{\eta}{\delta} + \eta.\Delta t) \max_j n_{i,j}.$$

(5) $$|\xi^{(i)}(m+2) - \xi^{(i)}(m+1)| \leq 3\eta.\Delta t \max_j n_{i,j}$$

Proof: For (4) we have

(6) $$\xi^{(i)}(m+2) = (1-\delta\, \Delta t)\, \xi^{(i)}(m+1) + \eta.\Delta t\, n_{i,j_m}\, \phi\, [\beta^{j_{m+1}} + \xi^{j_{m+1}}(m+1)].$$

It follows by induction (on m) that $\xi^{(i)}(m) \geq 0$ if $\delta\,\Delta t \leq 1$. As for the right-hand inequality of (4), assume that it is false, and that m_o is the first index for which it is violated, i.e.,

(7a) $\xi^{(i)}(m) \leq \left(\frac{\eta}{\delta} + \eta\,\Delta t\right) \max_j n_{i,j}$ for all $m \leq m_o$

(7b) $\xi^{(i)}(m_o + 1) > \left(\frac{\eta}{\delta} + \eta\,\Delta t\right) \max_j n_{i,j}$

Case 1: $\xi^{(i)}(m_o) \leq \frac{\eta}{\delta} \max_j n_{i,j}$

In this case, by (6) and (1)

$$\xi^{(i)}(m_o+1) \leq \xi^{(i)}(m_o) + \eta\cdot\Delta t \max_j n_{i,j} \leq \left(\frac{\eta}{\delta} + \eta\,\Delta t\right) \max_j n_{i,j}$$

which contradicts (7b).

Case 2: $\xi^{(i)}(m_o) > \frac{\eta}{\delta} \max_j n_{i,j}$

In this case, by (3) and (1),

$$\xi^{(i)}(m_o+1) - \xi^{(i)}(m_o) \leq \eta\,\Delta t \max_j n_{i,j} - \delta\,\Delta t\,\frac{\eta}{\delta} \max_j n_{i,j} \leq 0.$$

Hence $\xi^{(i)}(m_o+1) \leq \xi^{(i)}(m_o) \leq \left(\frac{\eta}{\delta} + \eta\,\Delta t\right) \max_j n_{i,j}$ which again contradicts (7b).

Thus (4) is always valid.

(5) follows immediately from (4) and (3), and $\delta\,\Delta t \leq 1$.

We now define a function differing slightly from $\xi^{(i)}$. For j=1,...,M we define $\zeta^{(i)}(kM+j)$ by

$$\zeta^{(i)}(kM) = \xi^{(i)}(kM) \text{ and}$$

(8) $$\zeta^{(i)}(kM+j) - \zeta^{(i)}(kM+j-1) = \eta\,\Delta t\,\phi\left[\beta^{j_{kM+j-1}} + \xi^{j_{kM+j-1}}(kM)\right] n_{i,j_{kM+j-2}} - \delta\,\Delta t\,\xi^{(i)}(kM)$$

$$j = 1,\ldots,M.$$

(Actually ζ also depends on k but it seems unnecessary to indicate this more explicitly).

<u>LEMMA 2</u>: Let $\tau = 3\eta \max_{i,j} n_{i,j}$

$$\rho = \frac{C_1 \tau^2}{3} + \delta\tau$$

(9) $\quad |\zeta^{(i)}((k+1)M) - \xi^{(i)}((k+1)M)| \leq \rho(M\Delta t)^2$

<u>Proof</u>: Comparing the definitions of ξ and ζ we see that the left hand side of (9) is bounded by

$$\eta\Delta t \max_{i,r} n_{i,r} \sum_{j=1}^{M} |\phi[\beta^{j_{kM+j-1}} + \xi^{j_{kM+j-1}}(kM)] - \phi[\beta^{j_{kM+j-1}} + \xi^{j_{kM+j-1}}(kM+j-1)|$$

$$+ \delta\, \Delta t \sum_{j=1}^{M} |\xi^{(i)}(kM+j-1) - \xi^{(i)}(kM)| \leq$$

$$\eta\Delta t \max_{i,r} n_{i,r} C_1 \sum_{j=1}^{M} \sup_{\nu} |\xi^{\nu}(kM) - \xi^{\nu}(kM+j-1)|$$

$$+ \delta\, \Delta t \sum_{j=1}^{M} \sup_{\nu} |\xi^{\nu}(kM) - \xi^{\nu}(kM+j-1)| \leq \quad \text{(by (5))}$$

$$\frac{C_1}{3} \tau\, \Delta t \sum_{j=1}^{M} (j-1)\tau\, \Delta t + \delta\, \Delta t \sum_{j=1}^{M} (j-1)\,\tau\, \Delta t$$

$$\leq \frac{C_1}{3} (M\tau\Delta t)^2 + \delta\tau(M\,\Delta t)^2$$

Let $\gamma^{(i)}$ be the solution of the differential equation

(10) $$\frac{d\gamma^{(i)}}{dt} = \sum_{j=1}^{n} \sum_{k=1}^{n} \eta f_{j,k}\, n_{i,j}\, \phi[\beta^{(k)} + \gamma^{(k)}(t)] - \delta\gamma^{(i)}(t)$$

satisfying

(11) $\gamma^{(i)}(0) = 0.$

Here $f_{j,k} \geq 0$, $\sum_{j,k} f_{j,k} = 1.$

We want to compare $\gamma^{(i)}(t)$ with $\xi^{(i)}([t/\Delta t])$ when there is a relation such as (14) between $f_{j,k}$ and the sequence $j_{-1}, j_0, \cdots$

First, we have the analogue of Lemma 1, i.e.,

LEMMA 3: For $t \geq 0$,

(12) $0 \leq \gamma^{(i)}(t) \leq \frac{\eta}{\delta} \max_j n_{i,j}$

(13) $\left| \frac{d\gamma^{(i)}}{dt} \right| \leq 2\eta \max_j n_{i,j} \leq \tau$.

Proof: $\frac{d\gamma^{(i)}}{dt} = \psi(t) - \delta\gamma^{(i)}(t)$

and $\gamma^{(i)}(0) = 0$ implies $\gamma^{(i)}(t) = e^{-\delta t} \int_0^t \psi(s) e^{\delta s} ds$. Any solution of (10) satisfies $\frac{d\gamma^{(i)}}{dt} = \psi(t) - \delta\gamma^{(i)}(t)$ for some ψ, $0 \leq \psi \leq \eta \max_j n_{i,j}$. This implies

$$0 \leq e^{-\delta t} \int_0^t \psi(s) e^{\delta s} ds \leq \frac{\eta}{\delta} \max_j n_{i,j}.$$

The second inequality in Lemma 3 follows immediately from the first one and (10).

Now let $\lambda_M^k(i,j) = \frac{1}{M} \cdot \{\text{number of } r,\ kM-1 \leq r \leq (k+1)M - 2, \text{ for which } j_r = i,\ j_{r+1} = j\}$ and assume that for some numbers $f_{i,j}$

(14) $|\lambda_m^k(i,j) - f_{i,j}| \leq M^{-1/2 + \alpha}$ for all $i,j \leq n$, $0 < \alpha < 1/2$.

LEMMA 4: Put $\varepsilon_k = \max_{i \le n} | \gamma^{(i)}(kM\Delta t) - \xi^{(i)}(kM) |$. Then

$$\varepsilon_{k+1} \le \left(1 + \frac{\rho M\Delta t}{\tau}\right) \varepsilon_k + 2\rho(M\Delta t)^2 + \frac{\tau}{3} n^2 M^{1/2+\alpha} \Delta t .$$

Proof: We want to estimate $\gamma^{(i)}((k+1)M\Delta t) - \xi^{(i)}((k+1)M)$.

Replacing $\xi^{(i)}((k+1)M)$ by $\zeta^{(i)}((k+1)M)$ introduces an error of at most $\rho(M\Delta t)^2$ by Lemma 2. However,

$$\zeta^{(i)}((k+1)M) - \zeta^{(i)}(kM) = \eta\Delta t \sum_{j=1}^{M} \phi[\beta^{j_{kM+j-1}} + \xi^{j_{kM+j-1}}(kM)] \cdot n_{i,j_{kM+j-2}}$$

$$-\delta\Delta t\, \xi^{(i)}(kM)$$

$$= \eta\, \Delta t\, M \sum_{r,s=1}^{n} \lambda_M^k (r,s)\, \phi[\beta^{(s)} + \xi^{(s)}(kM)]\, n_{i,r} - \delta\Delta t M \xi^{(i)}(kM) .$$

This last expression differs from

$$\eta\, \Delta t\, M \sum_{r,s=1}^{n} f_{r,s}\, \phi[\beta^{(s)} + \xi^{(s)}(mK)] n_{i,r} - \delta\Delta t M \xi^{(i)}(kM)$$

by at most

$$\eta\, \Delta t\, M\, n^2 M^{-1/2+\alpha} \max_r n_{i,r} \le \frac{\tau}{3} n^2 M^{1/2+\alpha} \Delta t.$$

Similarly,

$$\gamma^{(i)}((k+1)M\Delta t) - \gamma^{(i)}(kM\Delta t) = \int_{kM\Delta t}^{(k+1)M\Delta t} \left(\frac{d}{du}\gamma^{(i)}(u)\right) du$$

$$= M\Delta t \sum_{r=1}^{n} \sum_{s=1}^{n} \eta\, f_{r,s}\, n_{i,r}\, \phi[\beta^{(s)} + \gamma^{(s)}(kM\Delta t)] - \delta M\Delta t\, \gamma^{(i)}(kM\Delta t)$$

$$+ \eta \int_{kM\Delta t}^{(k+1)M\Delta t} \sum_r \sum_s f_{r,s}\, n_{i,r} \left\{\phi[\beta^{(s)} + \gamma^{(s)}(t)] - \phi[\beta^{(s)} + \gamma^{(s)}(kM\Delta t)]\right\} dt$$

$$- \delta \int_{kM\Delta t}^{(k+1)M\Delta t} \left\{\gamma^{(i)}(t) - \gamma^{(i)}(kM\Delta t)\right\} dt.$$

The absolute value of the last two integrals is at most

$$M\Delta t\eta \max_{i,r} n_{i,r} \sup_{s} \sup_{kM\Delta t\le t\le(k+1)M\Delta t} C_1|\gamma^{(s)}(t) - \gamma^{(s)}(kM\Delta t)|$$

$$+ \delta M \Delta t \sup_{kM\Delta t\le t\le(k+1)M\Delta t} |\gamma^{(i)}(t) - \gamma^{(i)}(kM\Delta t)| .$$

(13) shows that $\sup_{s} \sup_{kM\Delta t\le t\le(k+1)M\Delta t} |\gamma^{(s)}(t) - \gamma^{(s)}(kM\Delta t)| \le M\Delta t\tau$ so that

the integrals are bounded by

$$\frac{C_1}{3} (M\tau\Delta t)^2 + \delta\tau(M\Delta t)^2 \le \rho(M\Delta t)^2 .$$

Finally we have the estimate

$$\eta M\Delta t \sum_{r,s} f_{r,s}\, n_{i,r} |\phi[\beta^{(s)} + \xi^{(s)}(kM)] - \phi[\beta^{(s)} + \gamma^{(s)}(kM\Delta t)]|$$

$$+ \delta\Delta t M|\xi^{(i)}(kM) - \gamma^{(i)}(kM\Delta t)|$$

$$\le \eta M\Delta t \sup_{i,r} n_{i,r} C_1 \mathcal{E}_k + \delta\Delta t\, M\, \mathcal{E}_k = \frac{\rho}{\tau} M\Delta t\mathcal{E}_k .$$

Adding all error terms, we find

$$|\gamma^{(i)}((k+1)M\Delta t) - \xi^{(i)}((k+1)M)| \le 2\rho(M\Delta t)^2 + \frac{\tau}{3} n^2 M^{1/2+\alpha} \Delta t + \mathcal{E}_k(1+ \frac{\rho}{\tau} M\Delta t) .$$

<u>**THEOREM 1**</u>: Let $\xi^{(i)}(k)$ be determined by (2) and (3), and $\gamma^{(i)}(t)$ by (10) and (11). If (14) holds for all $k \le [\frac{t}{M\Delta t}]$ then

$$|\gamma^{(i)}(t) - \xi^{(i)}([t/\Delta t])| \le e^{\rho t/\tau} (4\tau M\Delta t + \frac{\tau^2 n^2}{3\rho} M^{-1/2+\alpha})$$

for $t/M\Delta t$ sufficiently large.

Proof: Let $t_1 = [t/M\Delta t]M\Delta t$ and $k_1 = [t/M\Delta t]$.

First we estimate

$$\max_i |\gamma^{(i)}(t_1) - \xi^{(i)}(k_1 M)| = \varepsilon_{k_1} .$$

By Lemma 4,

$$\varepsilon_{k_1} \leq (1+\frac{\rho M\Delta t}{\tau})\varepsilon_{k_1-1} + 2\rho(M\Delta t)^2 + \frac{\tau}{3} n^2 M^{1/2+\alpha}\Delta t)$$

$$\leq (1+\frac{\rho M\Delta t}{\tau})^2\varepsilon_{k_1-2} + (1+\frac{\rho M\Delta t}{\tau})[2\rho(M\Delta t)^2 + \frac{\tau}{3} n^2 M^{1/2+\alpha}\Delta t]$$

$$+ 2\rho(M\Delta t)^2 + \frac{\tau}{3} n^2 M^{1/2+\alpha}\Delta t \ldots .$$

$$\leq \frac{(1+\frac{\rho M\Delta t}{\tau})^{k_1} - 1}{\frac{\rho M\Delta t}{\tau}} (2\rho(M\Delta t)^2 + \frac{\tau}{3} n^2 M^{1/2+\alpha}\Delta t)$$

since $\varepsilon_0 = 0$.

Since $k_1 \leq t/M\Delta t$ we get

$$\varepsilon_{k_1} \leq (1+\frac{\rho M\Delta t}{\tau})^{t/M\Delta t} \cdot \frac{\tau}{\rho M\Delta t} [2\rho(M\Delta t)^2 + \frac{\tau}{3} n^2 M^{1/2+\alpha}\Delta t]$$

$$\leq e^{\rho t/\tau}(2\,\tau M\Delta t + \frac{\tau^2 n^2}{3\rho} M^{-1/2+\alpha}) .$$

By (5) and (13)

$$|\xi^{(i)}([t/\Delta t]) - \xi^{(i)}(k,M)| + |\gamma^{(i)}(t) - \gamma^{(i)}(t_1)| \leq 2M\Delta t \cdot \tau .$$

This proves the theorem.

LEMMA 5: If $j_{-1}, j_0, j_1, \ldots$ is a random sequence, forming a Markov chain with state space $\{1,2,\ldots,n\}$ containing exactly one ergodic class, and stationary probabilities

$$p_i = \lim_{m\to\infty} P\{j_m = i\}$$

and transition probabilities $p_{i,j} = P\{j_{m+1}=j \mid j_m = i\}$

then, putting $f_{i,j} = p_i\, p_{i,j}$, there exist constants B,C (depending on n, $p_{i,j}$)

such that for sufficiently small $\epsilon > 0$

$$P\{|\lambda_M^{(k)}(i,j) - f_{i,j}| \geq \epsilon \text{ for some } i,j\} \leq \frac{D}{\epsilon^2} e^{-B\epsilon^2 M} .$$

Proof: Melvin Katz and A. J. Thomasian, in "An exponential bound for functions of a Markov chain", Ann. Math. Stat. 31 (1960) 470-474, prove that for each fixed i,j

$$P\{|\lambda_M^{(k)}(i,j) - f_{i,j}| \geq \epsilon\} \leq \frac{C}{1-e^{-B\epsilon^2}} e^{-B\epsilon^2 M}$$

for constants B,C depending on p_i, $p_{i,j}$ and n only.

We obtain

$$P\{|\lambda_M^{(k)}(i,j) - f_{i,j}| \geq \epsilon \text{ for some } i,j\} \leq \frac{Cn^2}{1-e^{-B\epsilon^2}} e^{-B\epsilon^2 M} .$$

Since for small ϵ, $1 - e^{-B\epsilon^2} \geq \frac{B\epsilon^2}{2}$, the lemma follows.

N.B. In order to apply the Katz and Thomasian result we have to consider the Markov chain of pairs $(j_{-1}, j_0), (j_0, j_1), \ldots$.

COROLLARY: Under the assumptions of Lemma 5, for large M

$$P\{(14) \text{ holds for all } k \leq [t/M\Delta t]\}$$

$$\geq 1 - \frac{t}{M\Delta t} \frac{D}{M^{2\alpha-1}} e^{-BM^{2\alpha}} = 1 - \frac{Dt}{M^{2\alpha}\Delta t} e^{-BM^{2\alpha}} .$$

THEOREM 2: Under the assumptions of Lemma 5, for sufficiently large M and $0 < \alpha < 1/2$,

$$P\{|\gamma^{(i)}(t) - \xi^{(i)}([t/\Delta t])| \leq e^{\rho t/\tau} \left(4\tau M\Delta t + \frac{\tau^2 n^2}{3\rho} M^{-1/2+\alpha}\right) \text{ for } 1 \leq i \leq n\}$$

$$\geq 1 - \frac{Dt}{M^{2\alpha}\Delta t} e^{-BM^{2\alpha}} .$$

In particular, with $M = (\Delta t)^{-2/(3-2\alpha)}$,

$$P\{|\gamma^{(i)}(t) - \xi^{(i)}([t/\Delta t])| \leq e^{\rho t/\tau}\left(4\tau + \frac{\tau^2 n^2}{3\rho}\right)(\Delta t)^{(1-2\alpha)/(3-2\alpha)} \text{ for } 1 \leq i \leq n\}$$

$$\geq 1 - Dt\,(\Delta t)^{-(3-6\alpha)/(3-2\alpha)}\, e^{-B(\Delta t)^{-4\alpha/(3-2\alpha)}}$$

or, for any ϵ,

$$P\{|\gamma^{(i)}(t) - \xi^{(i)}([t/\Delta t])| \leq \epsilon \text{ for } 1 \leq i \leq n\} \geq 1-\epsilon$$

if Δt is sufficiently small.

Proof: Immediate from Theorem 1, and the corollary to Lemma 5.

This theorem justifies in a certain sense the approach in Sec. 16.2 of Principles of Neurodynamics. It shows that, even though the sequence of stimuli is a random sequence, with "large probability" the state of the perceptron is "close" to the solution of the differential equation (10) or equation (16.10) in the above reference.

Remark: It is worthwhile to look at the size of the constants in Theorem 2. D, B^{-1} depend on n, $p_{i,j}$ and will increase quite rapidly with n = total number of different stimuli.

$$\tau = 3\eta \max_{i,j} n_{i,j}$$

$$\rho = \frac{C_1\tau^2}{3} + \delta\tau .$$

Hence $\tau^2/\rho \leq 3/C_1$ (which tends to zero as $C_1 \to \infty$).

However, $\rho/\tau = (C_1\tau/3) + \delta$. If C_1 is replaced by a function $C_1(\Delta t)$, then $C_1(\Delta t)$ may tend to infinity as $\Delta t \to 0$, but we should have

$$e^{C_1(\Delta t)\cdot \tau t/3}(\Delta t)^{(1-2\alpha)/(3-2\alpha)} \to 0.$$

± 1 vs. 0,1

A NOTE ON THE CASE OF THE ADALINE VS. THE SIMPLE R-UNIT

By C. Kesler and G. Nagy

Periodic outbursts of speculation regarding the relative efficacy of the ±1 adaptation scheme advocated by Widrow[1] and the 0,1 scheme used in Rosenblatt's simple perceptron[2] have necessitated a more careful evaluation of the merits of the two systems. Two questions in particular must be answered:

1. Does either system show a marked superiority in the number of dichotomies achievable by a single threshold logic unit?
2. If a dichotomy may be achieved by both systems, is there a significant difference in convergence time?

This note proposes to show, first, that the adaline[1] and the simple R-unit[2] are virtually equivalent in regard to the problems they can solve, and second, that the relative speed of convergence depends on the average density of active input elements.

The answer to the first question is in the form of a theorem:

THEOREM: A threshold logic unit with variable threshold has solutions to the same set of dichotomies whether its input vector has components plus one and minus one, or zero and one, provided that minus ones in the one case are replaced by zeroes in the other.

<u>Corollary 1</u>: An adaline and a simple R-unit, both with an extra, permanently active input, have solutions to exactly the same set of dichotomies.

<u>Corollary 2</u>: If an adaline can do a problem, so can a simple R-unit with one extra, permanently active input.

<u>Corollary 3</u>: If a simple R-unit can do a problem, so can an adaline with one extra, permanently active input.

Proof

First we shall show that if a solution to a given dichotomy is known for ±1 input, then it is possible to specify a threshold which will yield a solution to the corresponding 0,1 input.

Consider an adaline with n inputs capable of dealing with some classification C where some of the stimuli $\bar{S}_i$ are in C^+ (the positive class), and some in C^- (the negative class). The $\bar{S}_i$ are n-dimensional vectors with components +1 or -1.

Then a solution may be described by the inequalities

$$\bar{S}_i \cdot \bar{W} > \Theta_a , \qquad \bar{S}_i \in C^+ \qquad (1a)$$

$$\bar{S}_i \cdot \bar{W} < \Theta_a , \qquad \bar{S}_i \in C^- \qquad (1b)$$

where $\bar{W}$ is a fixed solution weight vector and Θ_a is the threshold of the adaline.

Now add $\bar{U} \cdot \bar{W}$, where $\bar{U}$ is the n-dimensional vector whose components are all +1, to both sides of (1), and divide through by 2:

$$1/2\ (\bar{U}\cdot\bar{W} + \bar{S}_i\cdot\bar{W}) = 1/2\ (\bar{U} + \bar{S}_i)\cdot\bar{W} > 1/2\ (\Theta_a + \bar{U}\cdot\bar{W}),\ \bar{S}_i \in C^+ \qquad (2a)$$

$$1/2\ (\bar{U}\cdot\bar{W} + \bar{S}_i\cdot\bar{W}) = 1/2\ (\bar{U} + \bar{S}_i)\cdot\bar{W} < 1/2\ (\Theta_a + \bar{U}\cdot\bar{W}),\ \bar{S}_i \in C^- \qquad (2b)$$

Let $\bar{S}_i^* = 1/2\ (\bar{U} + \bar{S}_i)$, and $\Theta_r = 1/2\ (\Theta_a + \bar{U}\cdot\bar{W})$.

Then

$$\bar{S}_i^* \cdot \bar{W} > \Theta_r , \qquad \bar{S}_i^* \in C^+ \qquad (3a)$$

$$\bar{S}_i^* \cdot \bar{W} < \Theta_r , \qquad \bar{S}_i^* \in C^- \qquad (3b)$$

shows that $\bar{W}$ is also a solution to the corresponding classification with the minus ones in $\bar{S}_i$ replaced by zeroes in $\bar{S}_i^*$, and the threshold Θ_a of the adaline replaced by the threshold Θ_r of the R-unit.

To show that the adaline can solve a problem to which a solution exists in the R-unit with threshold Θ_r and solution vector $\bar{W}$, let

$$\bar{S}_i = 2\bar{S}_i^* - \bar{0}, \text{ and } \Theta_a = 2\Theta_r - \bar{0} \cdot \bar{W}.$$

Q.E.D.

The proof of the corollaries follows from the fact that for an R-unit or adaline with fixed threshold, an extra input which is permanently active is equivalent to an adaptive threshold.

It is worth noting that with the appropriate change in threshold, the same solution weight vector will serve in both cases.

To study relative speeds of convergence, simulation experiments were run on a threshold logic unit with 5 inputs. All possible dichotomies of 10 randomly chosen equal area stimuli were generated. For each dichotomy, it was either shown that no solutions existed, or the TLU was trained to a solution using quantized α-system error-correction reinforcement with zero starting weights, and cyclic presentation of stimuli.[2] 16 cases were computed (with different random stimulus sets) for each of the four possible nontrivial stimulus areas (4 plus ones, 3 plus ones, 2 plus ones, and 1 plus one). Since there are only 5 possible stimuli with 4 plus ones or with 1 plus one, the sets of 10 stimuli necessarily include repetitions of the same stimulus. No attempt was made to guarantee that these repetitions would be classified identically. Both the adaline and the simple R-unit learned exactly the same set of dichotomies (a demonstrable consequence of the equal size and odd number of components of the stimuli).

Dichotomies of 10 stimuli represent a fairly difficult task for a 5 input TLU; of the 65,536 dichotomies presented, only 3,186 were learned.

The relative speeds of convergence, denoting the average number of cycles through a ±1 stimulus sequence before perfect learning occurred divided by the corresponding figure for 0,1 inputs, were as follows:

Stimulus Size	Rel. Speed of Convergence
80%	.58
60%	.67
40%	.79
20%	1.20

A similar shift in favour of the 0,1 system for small stimuli was also observed in an SRI program[3]; nevertheless on the basis of those experiments it was concluded that the ±1 system is generally superior.

A heuristic argument to interpret our results might be constructed as follows. The degree of interference between two stimuli in opposite classes is proportional to their scalar product $\bar{S}_i \cdot \bar{S}_j$ or $\bar{S}^*_i \cdot \bar{S}^*_j$ ($\bar{S}_i$ still represents ±1 stimuli, and $\bar{S}^*_i$, 0,1 stimuli). With proper normalization such terms are analogous to the corresponding off-diagonal elements of the G-matrix[2], and for fast learning it is desirable to have such terms as small as possible.

As a guide to what we should expect in more complex cases, we calculate the expected value of $\bar{S}_i \cdot \bar{S}_j$ and of $\bar{S}^*_i \cdot \bar{S}^*_j$ for two n-bit stimuli whose elements are chosen independently with uniform probability $\frac{m}{n}$.

Then
$$E\left\{\bar{S}_i \cdot \bar{S}_j\right\} = n\left(2\frac{m}{n} - 1\right)^2 \qquad (4a)$$

$$E\left\{\bar{S}^*_i \cdot \bar{S}^*_j\right\} = n\left(\frac{m}{n}\right)^2 \qquad (4b)$$

After appropriate normalization (dividing (4a) by $E\left\{\bar{S}_i \cdot \bar{S}_i\right\}$ and (4b) by $E\left\{\bar{S}^*_i \cdot \bar{S}^*_i\right\}$, equations (4) yield the final interference terms:

$$Q\ (\text{adaline}) = \left(\frac{2m}{n} - 1\right)^2 \qquad (5a)$$

$$Q\ (\text{simple R-unit}) = \frac{m}{n} \qquad (5b)$$

It is seen that the interference term in the adaline reaches its minimum at $m = \frac{n}{2}$, while in the simple R-unit it is an increasing linear function of m. Traditionally, ±1 systems are run at the 50% activity level, while perceptrons aim for threshold settings insuring less than about 15% activity.

It is also interesting to note that, in the above instance, the interference terms Q (adaline) and Q (simple R-unit) are equal at $m = \frac{n}{4}$, corresponding to a stimulus size of 25%. This figure is in happy agreement with the simulation results.

The material presented in this note shows that the choice of activity scheme in a pattern recognition machine depends in relatively straightforward fashion on the expected input to the variable weight layer. For low density stimuli, the 0,1 scheme is preferable, while for larger patterns ±1 is faster.

REFERENCES

1. B. Widrow and G.M.E. Hoff, Jr., Adaptive Switching Circuits, TR No. 1553-1, Stanford Electronics Laboratories, Stanford, California, June, 1960.

2. F. Rosenblatt, Neurodynamics, Spartan Books, Washington D.C., 1962.

3. Brain, et al, An Experimental Comparison of Three Learning Machine Training Rules, SRI Quarterly Report No. 11, February, 1963.

FURTHER STUDIES OF REINFORCEMENT PROCEDURES AND RELATED PROBLEMS

by Carl Kesler

I. Introduction

This paper is concerned with the existence of solutions to classification problems in simple perceptrons (Ref. 4). It also contains some analysis and simulations of reinforcement procedures for simple perceptrons. Familiarity with CSRP Report No. 2 (Ref. 2) is assumed in several sections.

Section II contains corrections to CSRP Report No. 2.

In section III, several procedures are presented which attempt to solve a finite set, denoted by E, of constant coefficient linear inequalities. Three convergence theorems are proved about these reinforcement procedures. Then results are presented from four computer simulation experiments which compare different reinforcement procedures. From these results, it appears that if there is a solution to E, if solving E is relatively difficult, and if some parameters and conditions are properly specified, then the reinforcement procedures presented in section III will usually converge faster than those found in references 2 and 4.

In section IV, a method is outlined for quickly determining, with "high" probability or with certainty, whether a finite set of vectors, with integer components, is linearly dependent or linearly independent.

In section V, results are presented from two computer simulation experiments. From these results, it appears that a small minimal universal perceptron is not difficult to realize using "random" S-A connections.

In section VI, three lemmas, five theorems, and two false conjectures attack the problem of how many orthants are achievable by linear combinations of n m-dimensional vectors. A method is outlined which determines which (and hence how many) orthants are achievable by linear combinations of n m-dimensional vectors. Considering a finite set of vectors with integer components, the flow diagram in Figure 6 might either find that the vectors are linearly independent or find a linear relation among a linearly dependent subset of the vectors.

II. Corrections to CSRP Report No. 2

The following definitions will be required: Let C be any m by n matrix of known constants. Let c_i be the ith column vector of C^t, where C^t is the transpose of C and i=1,2,...,m. Let $x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$ be a column vector of unknown variables in N, where N is a real inner product vector space having n dimensions. Let $r = Cx = \begin{bmatrix} r_1 \\ \vdots \\ r_m \end{bmatrix}$. Let E be the following set of inequalities: $(c_i,x) > 0$, where (,) denotes the inner product of the contained vectors and i=1,2,...,m. A solution of E is any vector x such that $r_i > 0$ for i=1,2,...,m. Let Y denote the subset of E defined as follows. Any inequality, say the ith, is in Y if and only if there is a set of non-negative numbers $\{\alpha_j\}$ such that $\alpha_i > 0$ and $\sum_{j=1}^{m} \alpha_j c_j = 0$. Let U denote the set of numbers $\{j\}$ such that the jth inequality is in Y. Let V denote the set of numbers $\{i\}$ such that the ith inequality is in E-Y.

The definition of Y implies that there is a set of positive numbers $\{\delta_i\}$ such that $\sum_U \delta_i r_i = (\sum_U \delta_i c_i, x) = (0,x) = 0$. Thus Y can be thought of as the inequalities which are in contradictions; and E-Y can be thought of as the inequalities which are not in contradictions.

The following errors occur in reference 2: The equations at the bottom of page 15 might not hold because some k_i might be infinite ($1 \leq i \leq m$). The lemma on page 16 might not hold if V is empty. The sentence beginning on the third line of page 17 is incorrect. To correct these three errors, the material between the proof of Theorem 5 and the statement of Theorem 6 can be replaced by the following lemma.

Lemma: If E does not have a solution and if E-Y is not empty, there exists an infinite number of vectors x in N such that $r_i \begin{Bmatrix} =0 \text{ for } i \text{ in } U \\ >0 \text{ for } i \text{ in } V \end{Bmatrix}$.

Proof: Let Q be the subspace of N spanned by the set of vectors $\{c_j\}$ such that j is in U. Let P be the subspace of N which is orthogonal to Q and such that P+Q is the span of the column vectors of C^t. Let R be the orthogonal complement of P+Q. Then N is divided into three mutually orthogonal subspaces P,Q, and R, such that P+Q+R = N. Let the superscripts P,Q, or R on a vector denote that the vector is in P, Q, or R respectively. For j in U+V, define c_j^P and c_j^Q by $c_j^P + c_j^Q = c_j$.

Assume there is a set of non-negative numbers $\{\alpha_j\}$, not all zero, such that $\sum_V \alpha_j c_j^P = 0$. Then $\sum_V \alpha_j c_j = \sum_V \alpha_j c_j^Q = \sum_U \beta_j c_j$ for some set of numbers $\{\beta_j\}$. The definition of Y implies that there is a set of positive numbers $\{\delta_j\}$ such that $\sum_U \delta_j c_j = 0$. By subtracting this equation enough times from $\sum_U \beta_j c_j = \sum_V \alpha_j c_j$, one can find a set of negative numbers $\{\omega_j\}$ such that $\sum_U \omega_j c_j = \sum_V \alpha_j c_j$. Therefore some inequality is a member of both Y and E-Y, which is a contradiction. Consequently there does not exist a set of non-negative numbers $\{\alpha_j\}$, not all zero, such that $\sum_V \alpha_j c_j^P = 0$. (2.1)

Therefore by Theorem 3, a solution exists to the following set of inequalities: $(c_j^P, x) > 0$, where j is in V and (,) denotes the inner product of the contained vectors. Therefore by Theorem 5, if ABLE is applied to this set of inequalities with x initially equal to $\sum_V c_j^P$, then in a finite amount of time, one obtains a vector x such that x is a solution of this set of inequalities, $x = \sum_{V} \beta_j c_j^P$ for a set of positive numbers $\{\beta_j\}$, and x is in P. But such an x, as well as any positive number times x, satisfies $r_i \begin{Bmatrix} =0 \text{ for } i \text{ in } U \\ >0 \text{ for } i \text{ in } V \end{Bmatrix}$. Q.E.D.

III. Improved Simple Perceptron Error Correction Reinforcement Procedures

In this section, several procedures are presented which attempt to solve a finite set, denoted by E, of constant coefficient linear inequalities. Three convergence theorems are proved about these reinforcement procedures. Then results are presented from four computer simulation experiments which compare different reinforcement procedures. From these results, it appears that if there is a solution to E, if solving E is relatively difficult, and if some parameters and conditions are properly specified, then the reinforcement procedures presented in this section will usually converge faster than those found in references 2 and 4. The experimental parts of this section assume some knowledge of perceptron terminology, which may be obtained from reference 4.

Unless explicitly stated otherwise, it is assumed that all constants and variables mentioned below are real and finite.

The following definitions will be required: Let C be any m by n matrix of known constants. Let c_i be the ith column vector of C^t, where C^t is the transpose of C and i=1,2,...,m. Let $x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$ be an n-dimensional column vector of unknown variables. Let $r = Cx = \begin{bmatrix} r_1 \\ \vdots \\ r_m \end{bmatrix}$. Let δ be any known, non-negative constant. Let E be the following set of inequalities: $(c_i, x) > \delta$, where (,) denotes the inner product of the contained vectors and i=1,2,...,m. A solution of E is any vector x such that $r_i > \delta$ for i=1,2,...,m.

Let BELA denote the following error correction procedure applied to x. It is assumed that the initial value of x, call it x_{in}, is finite. The initial value of k is defined to be zero, where $k = \begin{bmatrix} k_1 \\ \vdots \\ k_m \end{bmatrix}$ is an m-dimensional column vector such that $\Delta x = C^t \Delta k$ during the application of BEIA to x. The inequalities in E are examined in any sequence, provided each is reexamined within a finite number of time units after its most recent examination, where the time unit is the time between successive examinations. Let time t begin at t = 0. It is assumed that there is a constant Γ such that each inequality is examined at least once before $t = \Gamma$. Let u be any known positive constant. Define the constant a by $\delta = u\alpha$. Let g be any known variable satisfying $-1 < -g_0 \leq g \leq g_0 < 1$, where g_0 is a known constant. Suppose any inequality, say the ith, is being examined. Define r_i^b and r_i^c to be the values of r_i just before and just after the examination respectively. Let Δ denote the change during the examination. If $r_i^b > \delta$, define $\Delta k = 0$. Then $r_i^c = r_i^b$. But if $r_i^b \leq \delta$, say that an error has occurred on the ith inequality and for j=1,2,...,m, define

$$\Delta k_j = \begin{cases} 0 & \text{if } j \neq i \text{ or } (c_j,c_j) = 0 \\ \max\left\{ \dfrac{(\delta - r_i^b)\,(1+g)}{(c_i,c_i)},\ u \right\} & \text{if } j = i \text{ and } (c_j,c_j) \neq 0 \end{cases}.$$

So if an error occurs on the ith inequality and if $(c_i,c_i) \neq 0$, then $r_i^c = r_i^b + (c_i,\Delta x) = \max\{\delta + g(\delta - r_i^b), r_i^b + u(c_i,c_i)\}$. On the basis of this equality, g could be called the "overshoot". The vector x may be doubled between examinations, but the following condition, called Condition 1, must be satisfied just before every doubling of x: The present value of (x,x) must be greater than the value of (x,x) just after the most recent doubling of x.

Let [ ] denote the integer part of the contained number. Let "integer BEIA" denote BEIA modified as follows. Let u be any known positive integer constant. If an error occurs on the ith inequality, then for j=1,2,...,m, define

$$\Delta k_j = \begin{cases} 0 & \text{if } j \neq i \text{ or } (c_j,c_j) = 0 \\ \max\left\{ \left[\dfrac{(\delta - r_i^b)\,(1+g)}{(c_i,c_i)}\right],\ u \right\} & \text{if } j = i \text{ and } (c_j,c_j) \neq 0 \end{cases}.$$

Let "cycle BEIA" denote integer BEIA restricted as follows. The sequence of examinations is restricted so that whenever any inequality is examined at any time t, it is also examined at time $t+\Gamma$. Let the first Γ examinations be called the first "cycle"; let the next Γ examinations be called the second "cycle"; etc. All conditions concerning the doubling of x must be of such a type that if x just after any cycle equals x just after a later cycle, then x will have become periodic.

An example will now be presented illustrating some of the motivation responsible for cycle BELA. The problem to be illustrated in Figure 1 has $\delta = 0$ and $C = \begin{bmatrix} +10 & +2 \\ -11 & -1 \end{bmatrix}$. Thus E consists of the following two inequalities:

$\left\{ \begin{array}{l} +10x_1 \ +2x_2 \ > 0 \\ -11x_1 \ -1x_2 \ > 0 \end{array} \right\}$. One error correction reinforcement procedure to be applied to x to find a solution of E is ABLE (Ref. 2) with the following conditions:

1. The two inequalities are examined in an alternating sequence, beginning with the first inequality.
2. x is initially zero.
3. At every error on any inequality, say the ith inequality, $b_i=1$.

The other error correction reinforcement procedure to be applied to x to find a solution of E is cycle BELA with the following conditions:

1. $\delta = 0$, $u = 1$, $g = 0.9$, $\Gamma = 2$
2. The first inequality is examined when $t = 0$ and the second inequality is examined when $t = 1$.
3. x is initially zero.
4. x is not to be doubled except between cycles.
5. x is to be doubled between cycles whenever Condition 1 is satisfied.

As can be seen from Figure 1, ABLE needed 25 examinations (and 25 errors) to attain a solution of E while cycle BELA needed 13 examinations (and 13 errors) to attain a solution of E.

Let BELA', integer BELA', and cycle BELA' respectively denote BELA, integer BELA, and cycle BELA modified as follows:

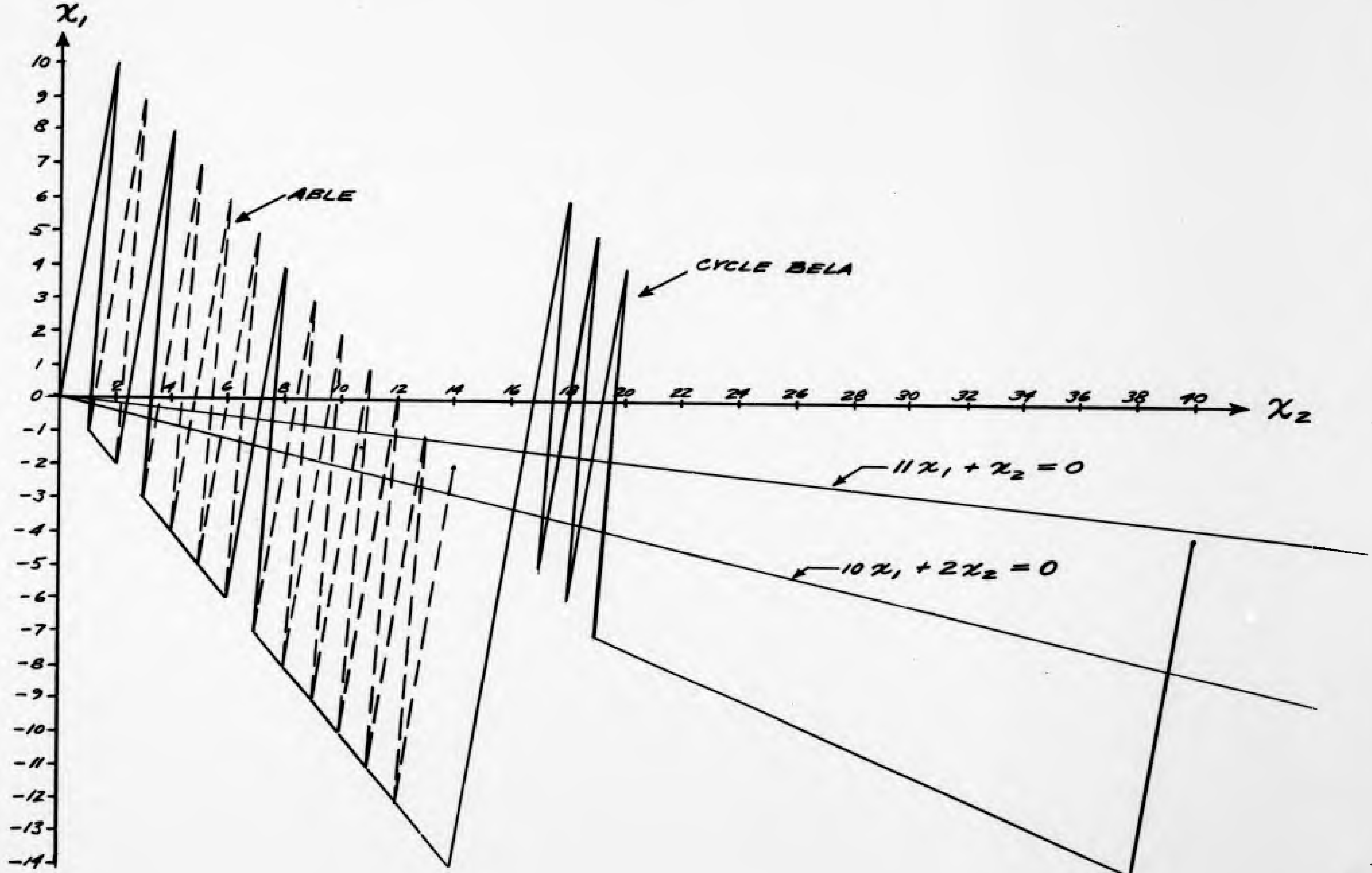


FIG. 1: COMPARISON OF ABLE AND CYCLE BELA ON A PROBLEM

1. Replace each doubling of x by a halving of u and δ.
2. Condition 1 now applies to the halving of u and δ instead of applying to the doubling of x.
3. Redefine [ ] by the following: For any number j, let [j] denote that integer multiple of $\frac{1}{2^h}$ defined by $j - \frac{1}{2^h} < [j] \le j$, where h is the number of times that u and δ have been halved.

Let x' = x during any particular application of BELA', integer BELA', or cycle BELA' to x. Let x" = x during the corresponding application of BELA, integer BELA, or cycle BELA to x. Then $2^h x' = x''$. (3.1)

<u>Theorem 1</u>: If a solution of E exists, BELA will yield a solution of E in a finite amount of time.

Proof: In this proof, assume a solution of E exists. By (3.1), if this theorem is true for BELA', it is also true for BELA. This theorem will therefore be proven for BELA'.

Suppose an error occurs on any inequality, say the ith. Assume

$\frac{(\delta - r_i^b)(1+g)}{(c_i, c_i)} \ge u$. Then $r_i^b + r_i^c \le -(1-g_o)(\delta - r_i^b) + 2\delta$. So $(\delta - r_i^b) \ge \frac{2\delta}{1-g_o}$

implies $r_i^b + r_i^c \le 0$. Therefore $\Delta(x,x) = (\Delta k_i)(r_i^b + r_i^c) \le$

$$\left\{ \begin{array}{ll} 0 & \text{when } \delta - r_i^b \ge \frac{2\delta}{1-g_o} \\ \frac{(\delta - r_i^b)(1+g_o)2\delta}{\underset{1\le i\le m}{\text{minimum}}\,(c_i,c_i)} \le \frac{4u^2a^2(1+g_o)}{(1-g_o)\underset{1\le i\le m}{\text{minimum}}\,(c_i,c_i)} & \text{when } \delta - r_i^b < \frac{2\delta}{1-g_o} \end{array} \right\}. \quad (3.2)$$

Now assume instead that $\frac{(\delta - r_i^b)(1+g)}{(c_i, c_i)} < u$. Then

$$\Delta(x,x) = (\Delta k_i)(r_i^b + r_i^c) \le u(2\delta + (u)\underset{1\le i\le m}{\text{maximum}}\,(c_i,c_i)) = u^2(2a + \underset{1\le i\le m}{\text{maximum}}(c_i,c_i)).$$

Combining this with (3.2) yields $\Delta(x,x) \le u^2 \max \left\{ \frac{4a^2(1+g_o)}{(1-g_o)\underset{1\le i\le m}{\text{minimum}}(c_i,c_i)},\ 2a + \underset{1\le i\le m}{\text{maximum}}(c_i,c_i) \right\} = u^2 y,$ (3.3) where this is the definition of the positive constant y.

Let s be any fixed solution of E. Let $\phi = \dfrac{(x,s)}{(s,s)}$. Let $\begin{bmatrix} z_1 \\ \vdots \\ z_n \end{bmatrix} = Cs$.

Suppose an error occurs on any inequality, say the ith. Then

$$\Delta\phi = \frac{z_i \Delta k_i}{(s,s)} \geq \frac{(u)\underset{1\leq i\leq m}{\text{minimum}}\, z_i}{(s,s)} \geq 0. \qquad (3.4)$$

Assume $\sum_{t=0}^{\infty} \sum_{i=1}^{m} \Delta k_i$ diverges, where Δ denotes the change during an examination. Then by (3.4), ϕ is unbounded as t approaches infinity. (3.5)

Therefore by (3.3) and (3.4), there is a time t_o such that for any error on any inequality, say the ith inequality, which occurs when $t > t_o$,

$$\Delta(x-\phi s, x-\phi s) = \Delta((x,x)-\phi^2(s,s)) \leq u^2 y-(\Delta k_i)z_i(2\phi+\Delta\phi) \leq u(uy-2\phi \underset{1\leq i\leq m}{\text{minimum}}\, z_i) \leq 0. \quad (3.6)$$

Let δ^{in} denote the initial value of δ. Then by (3.4) and (3.5), there is a time t_1 such that for $t > t_1$, $(c_i,\phi s) = \phi z_i > \dfrac{\delta^{in}}{2^h}$ for $i=1,2,\ldots,m$. (3.7)

Consider any i satisfying $1\leq i\leq m$. Let $\eta = \begin{bmatrix} \eta_1 \\ \vdots \\ \eta_n \end{bmatrix}$ be the shortest vector such that

$$(c_i,\phi s+\eta\ ) = \frac{\delta^{in}}{2^h}\ . \qquad (3.8)$$

Define the vector $\mu = \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_n \end{bmatrix}$ by $\eta = \left(\dfrac{\dfrac{\delta^{in}}{2^h} - (c_i,\phi s)}{(c_i,c_i)}\right) c_i + \mu.$ (3.9)

By (3.9), $(c_i,\phi s+\eta\ ) = (c_i,\mu) + \dfrac{\delta^{in}}{2^h}$. Therefore by (3.8), $(c_i,\mu) = 0$. Then by (3.8) and (3.9),

$$(\eta,\eta) = \left(\frac{\dfrac{\delta^{in}}{2^h} - (c_i,\phi s)}{\|c_i\|}\right)^2 \qquad (3.10)$$

Let $\gamma = \begin{bmatrix} \gamma_1 \\ \vdots \\ \gamma_n \end{bmatrix}$ be the shortest vector such that $(c_i, \phi s+\gamma) = \frac{\delta^{in}}{2^h}$ for some i satisfying $1 \leq i \leq m$, say $i=j$, and $(c_i, \phi s+\gamma) \geq \frac{\delta^{in}}{2^h}$ for all i satisfying $1 \leq i \leq m$. (3.11)

Then by (3.7) and (3.10), $$(\gamma,\gamma) = \underset{1 \leq i \leq m}{\text{minimum}} \left[\frac{\frac{\delta^{in}}{2^h} - (c_i, \phi s)}{\|c_i\|} \right]^2 \quad \text{for } t > t_1. \qquad (3.12)$$

Therefore by (3.5), (γ,γ) is unbounded as t approaches infinity. (3.13)

By (3.7) and (3.11), when $t > t_1$, $(x-\phi s, x-\phi s) < (\gamma,\gamma)$ implies that no error can occur on any inequality. (3.14)

By (3.6), (3.13), and (3.14), there is a finite time when x becomes a solution of E.

Now assume instead that $\sum_{t=0}^{\infty} \sum_{i=1}^{m} \Delta k_i$ converges and that BELA' does not yield a solution of E in a finite amount of time. Then $\lim_{t \to \infty} x$ exists. (3.15)

Therefore for $i=1,2,\ldots,m$, $\lim_{t \to \infty} r_i$ exists. (3.16)

If h is bounded as t approaches ∞, $\sum_{t=0}^{\infty} \sum_{i=1}^{m} \Delta k_i$ diverges.

Therefore h is unbounded as t approaches infinity. (3.17)

Assume $\lim_{t \to \infty} r_i < 0$ for some i, say $i=j$. Then by (3.17), there is a time t_2 such that for any error which occurs on the jth inequality when $t > t_2$,

$$\Delta k_j = \max \left\{ \frac{(\delta - r_j^b)(1+g)}{(c_j, c_j)}, u \right\} \geq \frac{(-r_j^b)(1-g_0)}{(c_j, c_j)} > d > 0,$$ where d is a suitable constant. Since this result implies that

$$\sum_{t=0}^{\infty} \sum_{i=1}^{m} \Delta k_i \text{ diverges, } \lim_{t \to \infty} r_i < 0 \text{ is not possible.} \qquad (3.18)$$

Let YE be the set of numbers $\{i\}$ such that $\lim_{t \to \infty} r_i = 0$. Let NE be the set of numbers $\{i\}$ such that $\lim_{t\to\infty} r_i > 0$. Then by (3.17), there is a time t_3 such that u and δ are halved when $t=t_3$, and such that for $t > t_3$ and i in NE, no error occurs on the ith inequality. Consequently by letting Δ denote the change during the time interval from t_3 to ∞, (3.15), (3.16), and (3.18), there is a set of numbers $\{\beta_i\}$ such that $\Delta(x,x) = 2(\Delta x, x+\Delta x)-(\Delta x,\Delta x) = 2(\sum_{YE} \beta_i c_i, x+\Delta x) - (\Delta x,\Delta x) \leq (2) \sum_{YE} \beta_i (c_i, x+\Delta x) = 0$. But there is a finite time when this result, (3.15), (3.17), and Condition 1 form a contradiction. Q.E.D.

Theorem 2: If a solution of E exists, integer BELA will yield a solution of E in a finite amount of time.

Proof: A proof will not be given because there is a proof of this theorem which is almost identical to the proof of Theorem 1. Q.E.D.

The following is an alternate form of Theorem 3 in reference 2. There is no solution of E if and only if there is a set of non-negative numbers $\{\alpha_i\}$, not all zero, such that

$$\sum_{i=1}^{m} \alpha_i c_i = 0. \qquad (3.19)$$

Statement (3.19) holds for the definitions given above, even though the definitions of E given above and in section II (or in reference 2) are slightly different.

Let Y denote the subset of E defined as follows. Any inequality, say the ith, is in Y if and only if there is a set of non-negative numbers $\{\alpha_j\}$ such that $\alpha_i > 0$ and $\sum_{j=1}^{m} \alpha_j c_j = 0$. Let U denote the set of numbers $\{i\}$ such that the ith inequality is in Y. Let V denote the set of numbers $\{i\}$ such that the ith inequality is in E-Y. Let W denote the set of numbers $\{i\}$ such that $(c_i, c_i) = 0$. Notice that W is a subset of U. The definition of Y implies that there is a set of positive numbers $\{\delta_i\}$ such that

$$\sum_U \delta_i r_i = (\sum_U \delta_i c_i, x) = (0,x) = 0. \qquad (3.20)$$

Thus Y can be thought of as the inequalities which are in contradictions; and E-Y can be thought of as the inequalities which are not in contradictions. By (3.19), a solution of E exists if and only if Y is empty.

Theorem 3: If a solution of E exists, cycle BELA will yield a solution of E in a finite amount of time. If no solution of E exists, if the elements of x_{in} and C are integers, and if cycle BELA is applied to x, then after a finite amount of time, no error will occur on any inequality in E-Y and x will become periodic.

Proof: Since the proof of Theorem 2 can be applied if U-W is empty, for the rest of this proof, assume U-W is not empty. Therefore $m \geq 2$.

Let N be the n-dimensional vector space which contains the column vectors of C^t. Let Q be the subspace of N spanned by the set of vectors $\{c_i\}$ such that i is in U. Let P be the subspace of N which is orthogonal to Q and such that P+Q is the span of the column vectors of C^t. Let R be the orthogonal complement of P+Q. Then N is divided into three mutually orthogonal subspaces P, Q, and R, such that P+Q+R = N. Let the superscripts P, Q, or R on a vector denote that the vector is in P, Q, or R respectively.

If during an application of cycle BELA' to x, x becomes periodic and a period contains one or more halvings of u and δ, then Condition 1 is contradicted. Thus by (3.1), if this theorem is true for cycle BELA', it is also true for cycle BELA. This theorem will therefore be proven for cycle BELA'. The portion of x which is in R does not affect r and is not altered by the application of cycle BELA' to x. Therefore, for the rest of this proof, assume that x is entirely in P+Q.

Define c_i^P, c_i^Q, x^P, x^Q, x_{in}^P, and x_{in}^Q by $c_i^P+c_i^Q=c_i$, $x^P+x^Q = x$, and $x_{in}^P+x_{in}^Q = x_{in}$. Then $(c_i^P,c_i^P)+(c_i^Q,c_i^Q) = (c_i,c_i)$, $(x^P, x^P)+(x^Q,x^Q) = (x,x)$, and $(x_{in}^P,x_{in}^P)+(x_{in}^Q,x_{in}^Q) = (x_{in},x_{in})$. Thus by the assumption that x_{in} is finite, x_{in}^P and x_{in}^Q are also finite. (3.21)

Suppose an error occurs on any inequality, say the ith. Then

$$\Delta(x,x) = (\Delta k_i)(r_i^b+r_i^c). \qquad (3.22)$$

Assume that i is in U+V-W and that $[\frac{(\delta-r_i^b)(1+g)}{(c_i,c_i)}] \geq u$. Then

$$r_i^b+r_i^c \leq -(1-g_o)\ (\delta-r_i^b)+2\delta. \tag{3.23}$$

Consequently $(\delta-r_i^b) \geq \frac{2\delta}{1-g_o}$ implies $r_i^b+r_i^c \leq 0$. Then by (3.22),

$$\Delta(x,x) \leq \left\{ \begin{array}{ll} 0 & \text{when } \delta-r_i^b \geq \frac{2\delta}{1-g_o} \\ \frac{(\delta-r_i^b)\ (1+g_o)2\delta}{\underset{U+V-W}{\text{minimum}}\ (c_i,c_i)} \leq \frac{4u^2a^2(1+g_o)}{1-g_o} & \text{when } \delta-r_i^b < \frac{2\delta}{1-g_o} \end{array} \right\} . \tag{3.24}$$

Now assume instead that i is in U+V-W and that $[\frac{(\delta-r_i^b)(1+g)}{(c_i,c_i)}] < u$. Then by (3.22), $\Delta(x,x) \leq u(2\delta+(u)\max_{U+V}(c_i,c_i)) = u^2(2a+\max_{U+V}(c_i,c_i))$. Therefore by (3.24)

and the fact that

$$\Delta(x,x)=0 \text{ if i is in W, } \Delta(x,x) \leq u^2 \max \left\{\frac{4a^2(1+g_o)}{1-g_o},\ 2a+\max_{U+V}(c_i,c_i)\right\} = u^2y, \tag{3.25}$$

where this is the definition of the positive constant y.

Next it will be shown that there is a constant Ω such that the following is true. If at any time t_o there begins an interval of length $\Delta t=\Omega$ which contains no error on any inequality in E-Y, then for $t > t_o$, no error will occur on any inequality in E-Y, and x will have become periodic before $t = t_o+\Omega$. (3.26)

Until (3.26) is proved, assume that we are restricting our attention to a finite time interval beginning at $t=t_o$, where t_o is any non-negative time.

$$\text{By (3.20), } 0 = \sum_U \delta_i r_i \geq (m-1)(\max_U \delta_i)(\min_U r_i) + (\min_U \delta_i)(\max_U r_i). \tag{3.27}$$

Let B denote a basis of Q consisting of c_i's such that i is in U. Let G denote a basis of Q obtained from B by the Gram-Schmidt orthogonalization process. Let B_i and G_i denote the ith element in B and G respectively. Then there is a set of numbers $\{\beta_{ij}\}$ such that each G_i satisfies $G_i=\sum_j \beta_{ij}B_j$. Thus $(x^Q,x^Q)=\sum_i (x^Q,G_i)^2= \sum_i(\sum_j \beta_{ij}(x^Q,B_j))^2=\sum_i(\sum_j \beta_{ij}(B_j,x))^2$. Therefore by (3.20), $\|x^Q\| = 1$ implies $\max_U r_i \geq \gamma$, where γ is a positive constant. Therefore $\|x^Q\| = v_1$ implies

$\max_U r_i \geq v_i \gamma$ for all non-negative numbers v_i. Then by (3.27),

$$(m-1)(\min_U r_i) \leq \frac{-(\min_U \delta_i)(\max_U r_i)}{\max_U \delta_i} \leq \frac{-(\min_U \delta_i)\gamma \, \|x^Q\|}{\max_U \delta_i} . \tag{3.28}$$

$$\text{Therefore } (x^Q, x^Q) \leq \left[\frac{(m-1)(\max_U \delta_i)(\min_U r_i)}{\gamma(\min_U \delta_i)} \right]^2 . \tag{3.29}$$

Let $\omega_\pi = \underset{\substack{\pi\text{th cycle} \\ U+V-W}}{\text{minimum}} r_i^b$, where the πth cycle is any cycle that occurs entirely after $t = t_o$. Let u_o be the value of u at $t = t_o$. Letting Δ denote change during an examination in the πth cycle,

$$\max_{U+V} \Delta r_i \leq (\max_{U+V} \Delta k_i) \max_{U+V}(c_i, c_j) \leq (\max_{U+V}(c_i, c_j)) \max \left\{ \frac{(\delta - \omega_\pi)(1+g_o)}{\underset{U+V-W}{\text{minimum}}(c_i, c_i)} , u_o \right\} .$$

Therefore by (3.20), just before the πth cycle,

$$0 \geq \min_U r_i \geq \min_{U+V} r_i > \omega_\pi - \delta - \Gamma(\max_{U+V}(c_i, c_j)) \max \left\{ \frac{(\delta - \omega_\pi)(1+g_o)}{\underset{U+V-W}{\text{minimum}}(c_i, c_i)} , u_o \right\} =$$

$$\min \left\{ (\omega_\pi - \delta) - \frac{\Gamma(\max_{U+V}(c_i, c_j))(\delta - \omega_\pi)(1+g_o)}{\underset{U+V-W}{\text{minimum}}(c_i, c_i)} , (\omega_\pi - \delta) - \Gamma u_o \max_{U+V}(c_i, c_j) \right\} \geq$$

$$(\omega_\pi - \delta)\left(2 + \frac{\Gamma(1+g_o)\max_{U+V}(c_i, c_j)}{\underset{U+V-W}{\text{minimum}}(c_i, c_i)} \right) - \Gamma u_o \max_{U+V}(c_i, c_j). \tag{3.30}$$

Therefore just before the πth cycle, $(\min_U r_i)^2 <$

$$\max \left\{ (\omega_\pi - \delta)^2 \left[1 + \frac{\Gamma(\max_{U+V}(c_i, c_j))(1+g_o)}{\underset{U+V-W}{\text{minimum}}(c_i, c_i)} \right]^2 , ((\omega_\pi - \delta) - \Gamma u_o \max_{U+V}(c_i, c_j))^2 \right\} . \tag{3.31}$$

Notice that (3.30) and (3.31) would still hold if the definitions of the first cycle, the second cycle, etc. were changed by beginning to count examinations after the first i examinations had passed, where $1 \leq i \leq \Gamma - 1$. (3.32)

Let the negative constant d be defined by $d = \frac{-1}{(1-g_o)} \left[a + \sqrt{a^2 + 2y\Gamma \max_{U+V}(c_i, c_i)} \right]$.

Let the πth cycle be called a "case 1" cycle when $\omega_\pi - \delta < 2u_o d$ and called a "case 2" cycle when $\omega_\pi - \delta \geq 2u_o d$. Assume the πth cycle is a case 1 cycle. Then

$$\omega_\pi - \delta < \frac{-4au_o}{1-g_o} \leq 0 \text{ and } \omega_\pi - \delta < \frac{-2u_o\sqrt{\dfrac{2y\Gamma \max(c_i,c_i)}{U+V}}}{1-g_o} \leq \frac{-2u_o\left(\dfrac{\max(c_i,c_i)}{U+V}\right)\sqrt{2\Gamma}}{1-g_o} < 0.$$

Thus $(-1+g_o)(\delta-\omega_\pi) + 2\delta \leq (\omega_\pi-\delta)(1-g_o) + 2au_o < -2au_o \leq 0$

$$\text{and } \left[\frac{(\delta-\omega_\pi)(1-g_o)}{\dfrac{\max(c_i,c_i)}{U+V}} \right] \geq [\, 2u_o\sqrt{2\Gamma}\,] > u_o. \tag{3.33}$$

Then by (3.22), (3.23), (3.25), and letting Δ denote the change during the πth cycle,

$$\begin{aligned}
\Delta(x,x) &\leq \left[\frac{(\delta-\omega_\pi)(1-g_o)}{\dfrac{\max(c_i,c_i)}{U+V}} \right] ((-1+g_o)(\delta-\omega_\pi) + 2\delta) + (\Gamma-1)u_o^2 y \\
&< - \frac{(\omega_\pi-\delta)(1-g_o)}{2\,\dfrac{\max(c_i,c_i)}{U+V}} ((1-g_o)(\omega_\pi-\delta) + 2au_o) + \Gamma u_o^2 y \\
&= f(\omega_\pi-\delta),
\end{aligned} \tag{3.34}$$

where this is the definition of the function $f(\omega_\pi-\delta)$. Therefore $f(\omega_\pi-\delta) \geq 0$

$$\begin{aligned}
\text{implies } \quad (\omega_\pi-\delta) &\geq \left(\frac{\dfrac{au_o(1-g_o)}{\frac{\max(c_i,c_i)}{U+V}} + \sqrt{\left[\dfrac{au_o(1-g_o)}{\frac{\max(c_i,c_i)}{U+V}}\right]^2 + \dfrac{2\Gamma u_o^2 y(1-g_o)^2}{\frac{\max(c_i,c_i)}{U+V}}}}{-\dfrac{(1-g_o)^2}{\frac{\max(c_i,c_i)}{U+V}}} \right) \\
&= \frac{-u_o}{(1-g_o)} \left[a + \sqrt{a^2 + 2y\Gamma \frac{\max(c_i,c_i)}{U+V}} \right] \\
&= u_o d.
\end{aligned}$$

Then by (3.34) and the assumption that the πth cycle is a case 1 cycle,

$$\Delta(x,x) < f(\omega_\pi-\delta) < f(2u_o d) < f\left(\tfrac{3}{2}u_o d\right) < 0. \tag{3.35}$$

Until (3.26) is proved, assume that the finite time interval beginning at $t=t_o$ contains no error on any inequality in E-Y. Then for any error when $t \geq t_o$, $\Delta(x,x) = \Delta(x^Q,x^Q)$. So by (3.35), (3.29), (3.31), (3.34), the second part of (3.33), and the assumption that the πth cycle is a case 1 cycle, somewhere within the

$$\left[\frac{\left(\dfrac{ \dfrac{(m-1)(\max \delta_i)}{U} \left[1 + \dfrac{ \dfrac{\Gamma(\max(c_i,c_j))(1+g_o)}{U+V} }{ \dfrac{\text{minimum}(c_i,c_i)}{U+V-W} } \right] }{ \dfrac{\gamma(\min \delta_i)}{U} } \right)^2 (\omega_\pi-\delta)^2 }{ -f(\omega_\pi-\delta) } \right] = \left[\frac{v_2(\omega_\pi-\delta)^2}{-f(\omega_\pi-\delta)} \right]$$

$$= \left[\frac{v_2}{\left(\dfrac{-f(\omega_\pi-\delta)}{(\omega_\pi-\delta)^2} \right)} \right] \leq \left[\frac{v_2}{\left(\dfrac{-f(2u_o d)}{(2u_o d)^2} \right)} \right] = \left[\frac{v_2(2u_o d)^2}{-f(2u_o d)} \right] = w$$

cycles just after the πth cycle, $\omega_j-\delta \geq 2u_o d$; (3.36)

the positive constants v_2 and w are defined by the equations (in this result) in which they first appear. In other words, (3.36) states that if the πth cycle is a case 1 cycle, then a case 2 cycle will occur within the w cycles immediately following the πth cycle, where w is a positive constant defined in (3.36). (3.37)

Now assume instead that the πth cycle is a case 2 cycle. Then by (3.37), a case 2 cycle will occur within the w+1 cycles just after the πth cycle. (3.38)
By the assumption that the πth cycle is a case 2 cycle, (3.29), and (3.30), just

before the πth cycle, $(x^Q,x^Q) <$

$$\left(\frac{ \dfrac{(m-1)(\max \delta_i)}{U} }{ \dfrac{\gamma(\min \delta_i)}{U} } \right)^2 \left(2u_o d \left[2 + \frac{ \dfrac{\Gamma(1+g_o)\max(c_i,c_j)}{U+V} }{ \dfrac{\text{minimum}\ (c_i,c_i)}{U+V-W} } \right] - \Gamma u_o \frac{\max(c_i,c_j)}{U+V} \right)^2 = u_o^2 z, \quad (3.39)$$

where this is the definition of the constant z.

By (3.25), (3.37), (3.38), and (3.39), $(x^Q,x^Q) < u_o^2 z + (w+1)\Gamma u_o^2 y$ for $t > t_o + \Gamma(1+w)$. (3.40)

Let t_1 be any particular time such that $t_1 \geq t_o$. Let u_1 be the value of u at $t=t_1$.

Then (3.40) can be generalized to the following: for $t > t_1+\Gamma(1+w)$,

$(x^Q,x^Q) < u_1^2 z + (w+1)\Gamma u_1^2 y.$ (3.41)

Thus for $t > t_1+\Gamma(1+w)$, the magnitude of each component of x^Q is less than $u_1\sqrt{z+(w+1)\Gamma y}$. Therefore because of the definition of [ ] and the assumption that the elements of C are integers, there is a constant v_3 such that the following is true: For any t_1 such that no halvings of u and δ occur when $t > t_1$, the number of different states x can be in when $t > t_1+\Gamma(1+w)$ is less than v_3. Thus for any t_1 such that no halvings of u and δ occur when $t > t_1$, x will be periodic before $t \geq t_1+\Gamma(1+w+v_3)+1$. Consequently if (3.26) is false, then for $t > t_o$, u and δ must be halved within the interval, of length $\Delta t = \Gamma(1+w+v_3)+1$, just after each halving of u and δ. (3.42)

Let u_{in} denote the initial value of u. Let τ be the smallest integer such that $\tau > 1$ and $\left(\frac{u_o}{2u_{in}}\right)^2 \geq \left(\frac{u_o}{2^\tau}\right)^2 (z+(w+1)\Gamma y)$. Assume that (3.26) is false. Then by (3.42), the assumption that the elements of x_{in} and C are integers, and Condition 1, $(x^Q,x^Q) \geq \left(\frac{u_o}{2u_{in}}\right)^2$ just after the second, and all subsequent, halvings of u and δ after $t=t_o$. (3.43)

By (3.42), the assumption that (3.26) is false, and by applying (3.41) to the time t_1 just after the τth halving of u and δ after $t=t_o$,

for $t > t_1+\Gamma(1+w)$, $(x^Q,x^Q) < \left(\frac{u_o}{2^\tau}\right)^2 (z+(w+1)\Gamma y) \leq \left(\frac{u_o}{2u_{in}}\right)^2$. Therefore by (3.43), u and δ cannot be halved when $t > t_1+\Gamma(1+w)$. (3.44)

By (3.42) and the assumption that (3.26) is false, the time t_1 just after the τth halving of u and δ after $t=t_o$ satisfies $t_1 \leq t_o + \tau(\Gamma(1+w+v_3)+1)$. Therefore by (3.44), u and δ cannot be halved when $t > t_o + \tau(\Gamma(1+w+v_3)+1)+\Gamma(1+w)$. (3.45)

By (3.42) and the assumption that (3.26) is false, u and δ must be halved in the interval $t_o+(\tau+1)(\Gamma(1+w+v_3)+1)+\Gamma(1+w)+2 > t > t_o + \tau(\Gamma(1+w+v_3)+1)+\Gamma(1+w)$. Therefore by (3.45), (3.26) is true with $\Omega = (\tau+1)(\Gamma(1+w+v_3)+1)+\Gamma(1+w)+2$. (3.46)

By (3.46), this proof is complete if E-Y is empty. Therefore for the rest of this proof, assume E-Y is not empty.

Let s be any fixed vector x obtained as outlined in the proof of the lemma in section II. Let $\phi = \frac{(x,s)}{(s,s)}$. Let $\begin{bmatrix} z_1 \\ \vdots \\ z_n \end{bmatrix} = Cs$. Suppose an error occurs on any inequality, say the ith. Then $\Delta\phi = 0$ unless i is in V, in which case

$$\Delta\phi = \frac{z_i \Delta k_i}{(s,s)} \geq \frac{(u)\min_V z_i}{(s,s)} \geq 0. \tag{3.47}$$

Also, by (3.25), $\Delta(x-\phi s, x-\phi s) = \Delta((x,x)-\phi^2(s,s)) \leq u^2 y - (\Delta k_i) z_i (2\phi + \Delta\phi)$. (3.48)

For the rest of this proof, assume that every interval of length $\Delta t = \Omega$ contains at least one error on an inequality in E-Y. (3.49)

Assume there is a constant h_o such that $h \leq h_o$ as t approaches infinity. Then by (3.47), at any error on an inequality in E-Y,

$$\Delta\phi \geq \frac{(u_{in})\min_V z_i}{2^{(h_o)}(s,s)} > 0. \tag{3.50}$$

By (3.47), (3.49), (3.50), and the assumption that x_{in} is finite, there is a time t_2 such that for $t > t_2$, $\phi > 0$. Therefore by (3.48), at any error when $t > t_2$ on any inequality, say the ith inequality, $\Delta(x-\phi s, x-\phi s) \leq u_{in}^2 y - \frac{u_{in} z_i (2\phi + \Delta\phi)}{2^{(h_o)}}$.

Thus by (3.49) and (3.50), there is a finite time when $(x-\phi s, x-\phi s) < 0$, which is a contradiction. (3.51)

Assume $\sum_{t=0}^{\infty} \sum_V \Delta k_i$ diverges, where Δ denotes the change during an examination.

Then by (3.47), ϕ monotonely approaches infinity as t approaches infinity. (3.52)

Therefore there is a time t_3 such that for $t > t_3$, $(c_i, \phi s) = \phi z_i > \frac{au_{in}}{2^h}$ for i in V. (3.53)

By (3.48), (3.49), (3.52), and letting Δ denote the change during an interval of length $\Delta t = \Omega$, there is a time t_4 such that for any interval of length $\Delta t = \Omega$ occurring after $t=t_4$, $\Delta(x-\phi s, x-\phi s) \leq \Omega u^2 y - 2\phi u(\min_V z_i) = u(u\Omega y - 2\phi(\min_V z_i)) \leq 0$.

Then by (3.48) and (3.52), $(x-\phi s, x-\phi s)$ is bounded as t approaches infinity. (3.54)

Let δ^{in} denote the initial value of δ. Consider any i in V. Let $\eta = \begin{bmatrix} \eta_1 \\ \vdots \\ \eta_n \end{bmatrix}$ be the shortest vector such that $(c_i, \phi s + \eta) = \frac{\delta^{in}}{2^h}$. (3.55)

Define the vector $\mu = \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_n \end{bmatrix}$ by $\eta = \left[\frac{\frac{\delta^{in}}{2^h} - (c_i, \phi s)}{(c_i, c_i)}\right] c_i + \mu$. (3.56)

By (3.56), $(c_i, \phi s + \eta) = (c_i, \mu) + \frac{\delta^{in}}{2^h}$. Then by (3.55), $(c_i, \mu) = 0$. Thus by (3.55) and (3.56), $(\eta, \eta) = \left(\frac{\frac{\delta^{in}}{2^h} - (c_i, \phi s)}{\|c_i\|}\right)^2$. (3.57)

Let $\psi = \begin{bmatrix} \psi_1 \\ \vdots \\ \psi_n \end{bmatrix}$ be the shortest vector such that $(c_i, \phi s + \psi) = \frac{\delta^{in}}{2^h}$ for some i in V, say i=j, and $(c_i, \phi s + \psi) \geq \frac{\delta^{in}}{2^h}$ for all i in V. (3.58)

Then by (3.53) and (3.57), $(\psi, \psi) = \min_V \left(\frac{\frac{\delta^{in}}{2^h} - (c_i, \phi s)}{\|c_i\|}\right)^2$ for $t > t_3$. (3.59)

Therefore by (3.52), (ψ, ψ) is unbounded as t approaches infinity. (3.60)

By (3.53) and (3.58), when $t > t_3$, $(x-\phi s, x-\phi s) < (\psi, \psi)$ implies that no error can occur on any inequality in E-Y. (3.61)

So there is a finite time when (3.49), (3.54), (3.60), and (3.61) form a contradiction. (3.62)

For the rest of this proof, assume that h is unbounded as t approaches infinity and that $\sum_{t=0}^{\infty} \sum_{V} \Delta k_i$ converges, where Δ denotes the change during an examination. (3.63)

Set $t_o = (j-1)\Gamma$, where j is any positive integer. Then by (3.28) and (3.30), just before the jth cycle,

$$(\delta-\omega_j) > \frac{\|x^Q\| - \dfrac{(m-1)(\max_U \delta_i)\Gamma u_o(\max_{U+V}(c_i,c_k))}{\gamma \min_U \delta_i}}{\dfrac{(m-1)(\max_U \delta_i)}{\gamma \min_U \delta_i}\left[2 + \dfrac{\Gamma(1+g_o)(\max_{U+V}(c_i,c_k))}{\operatorname{minimum}_{U+V-W}(c_i,c_i)}\right]} = \frac{\|x^Q\| - u_o v_4}{v_5}, \text{ where}$$

this is the definition of the positive constants v_4 and v_5. Therefore if $\|x^Q\| \geq u_o v_4 - 2u_o v_5 d$ just before the jth cycle, then $\omega_j - \delta < 2u_o d$. Therefore by (3.35) and letting Δ denote the change during the jth cycle, if $\|x^Q\| \geq u_o(v_4 - 2v_5 d)$ just before the jth cycle,

then $\Delta(x,x) < f(2u_o d) < f(\frac{3}{2} u_o d) < 0$. (3.64)

By (3.21), (3.63), and letting Δ denote the change during an examination,

$$\|x^P\| = \|x^P_{in} + \sum_{t=0}^{t} \sum_{V} \Delta k_i c_i^P\| \leq \|x^P_{in}\| + \sum_{t=0}^{t} \sum_{V} \|c_i^P \Delta k_i\|$$

$$\leq \|x^P_{in}\| + (\max_V \|c_i\|) \sum_{t=0}^{\infty} \sum_{V} \Delta k_i \leq v_6, \tag{3.65}$$

where v_6 is a positive constant.

Therefore $(x^Q,x^Q) \geq (x,x) - v_6^2$. Consequently by (3.25) and (3.64), there is a positive constant v_7 such that $(x,x) < v_7$. (3.66)

From the definition of s, there is a set of positive numbers $\{\beta_i\}$ such that $s = \sum_V \beta_i c_i^P$. So there is a set of numbers $\{\xi_i\}$ such that

$$\sum_V \beta_i c_i - s = \sum_V \beta_i c_i^Q = \sum_U \xi_i c_i. \tag{3.67}$$

The definition of Y implies that there is a set of positive numbers $\{\delta_i\}$ such that $\sum_{U} \delta_i c_i = 0$. By subtracting this equation enough times from (3.67), one can find a set of positive numbers $\{\alpha_i\}$ such that $\sum_{U+V} \alpha_i c_i = s$. Therefore $\phi \leqq -1$ implies $\min_{U+V} r_i \leqq -v_8 < 0$, where v_8 is a positive constant. Therefore if q is any non-negative number, $\phi \leqq -q$ implies $\min_{U+V} r_i \leqq -qv_8$. Therefore if q is any non-negative number, $\min_{U+V} r_i > -qv_8$ implies $\phi > -q$. (3.68)

By (3.21), (3.63), and the fact that $\lim_{t\to\infty} x^P = x^P_{in} + \sum_{t=0}^{\infty} \sum_{V} c_i^P \Delta k_i$, $\lim_{t\to\infty} x^P$ exists. Thus there is a set of numbers $\{\alpha_i\}$ such that $\lim_{t\to\infty} x^P = \sum_{V} \alpha_i c_i^P$. (3.69)

By (3.35), (3.66), and Condition 1, the first case 2 cycle will begin a finite amount of time after t=0; and if a case 2 cycle begins at any time t_5, then the next case 2 cycle will begin a finite amount of time after $t=t_5$. (3.70) By (3.30), if the πth cycle is a case 2 cycle, then just before the πth cycle,

$$\min_{U+V} r_i > (2u_o d)\left(2 + \frac{\Gamma(1+g_o)\max_{U+V}(c_i,c_j)}{\operatorname{minimum}_{U+V-W}(c_i,c_i)}\right) - \Gamma u_o \max_{U+V}(c_i,c_j).$$

So by (3.47), (3.63), (3.68), and (3.70), $\lim_{t\to\infty} \phi$ exists and $\lim_{t\to\infty} \phi = \lim_{t\to\infty} \frac{(x,s)}{(s,s)} \geqq 0$. (3.71)

From the definition of s, s satisfies the lemma in section II. There is a small positive constant β such that for i in V, $s + \beta c_i^P$ also satisfies the lemma in section II. Then by the arguments which led to (3.71), for i in V,

$$\lim_{t\to\infty} \frac{(x,s+\beta c_i^P)}{(s+\beta c_i^P, s+\beta c_i^P)} \text{ exists and } \lim_{t\to\infty} \frac{(x,s+\beta c_i^P)}{(s+\beta c_i^P, s+\beta c_i^P)} \geqq 0. \quad (3.72)$$

Assume $\lim_{t\to\infty} \frac{(x,s)}{(s,s)} = 0$ and $\lim_{t\to\infty} \frac{(x,s+\beta c_i^P)}{(s+\beta c_i^P, s+\beta c_i^P)} = 0$ for i in V. Then

$(\beta)\lim_{t\to\infty}(x,c_i^P) = \lim_{t\to\infty}(x,s+\beta c_i^P) - \lim_{t\to\infty}(x,s) = 0$ for i in V. Therefore $(\lim_{t\to\infty} x^P, c_i^P) =$

$\lim_{t\to\infty}(x^P, c_i^P) = \lim_{t\to\infty}(x,c_i^P) = 0$ for i in V. Then by (3.69), $(\sum_V \alpha_i c_i^P, c_i^P) = 0$ for i in V. Thus $(\sum_V \alpha_i c_i^P, \sum_V \alpha_i c_i^P) = 0$. Thus by (3.69), $\lim_{t\to\infty} x^P = 0$. Therefore for any particular positive constant ζ, there is a time t_6 such that for $t > t_6$, $(x,x) \leq \zeta + (x^Q, x^Q)$. Consequently by applying (3.64) to any cycle, say the jth, which occurs after $t = t_6$, and by letting Δ denote the change during the jth cycle, if $(x,x) \geq \zeta + u_o^2(v_4 - 2dv_5)^2$ just before the jth cycle, then $\Delta(x,x) < f(2u_o d) < f(\frac{3}{2}u_o d) < 0$. Then by (3.25), (3.66), and Condition 1, there is a time t_7 such that for $t > t_7$, $(x,x) < \zeta + u_o^2(v_4 - 2dv_5)^2 + 2\Gamma u_o^2 y$. Thus if ζ is chosen small enough, then there is a finite time when this result, (3.63), and Condition 1 form a contradiction. Thus by (3.71) and (3.72), either

$$\lim_{t\to\infty} \frac{(x,s)}{(s,s)} > 0 \text{ or } \lim_{t\to\infty} \frac{(x,s+\beta c_i^P)}{(s+\beta c_i^P, s+\beta c_i^P)} > 0 \text{ for some i in V.} \tag{3.73}$$

If $\lim_{t\to\infty} \frac{(x,s+\beta c_i^P)}{(s+\beta c_i^P, s+\beta c_i^P)} > 0$ for some i in V, say i=j, then replace s by $s+\beta c_j^P$.

$$\text{Then by (3.73), } \lim_{t\to\infty} \phi = \lim_{t\to\infty} \frac{(x,s)}{(s,s)} > 0. \tag{3.74}$$

By (3.47), (3.63), and (3.74), there is a time t_8 such that for $t > t_8$, $\phi > 0$ and $u(\Omega uy - (\frac{1}{2^\Omega})(\min_V z_i)2\phi) \leq 0$. Consider any interval of length $\Delta t = \Omega$ which occurs after $t = t_8$. Let u_2 and ϕ_2 denote the values of u and ϕ respectively at the beginning of the interval. Let Δ denote the change during the interval. Then by (3.47), (3.48), (3.49), and Condition 1,

$$\Delta(x-\phi s,x-\phi s) \leq u_2(\Omega u_2 y-(\frac{1}{2^\Omega})(\min_V z_i)2\phi_2) \leq 0. \tag{3.75}$$

By (3.48) and letting Δ denote the change during any examination which occurs after $t=t_8$, $\Delta(x-\phi s,x-\phi s) \leq u^2 y$. Therefore by (3.63) and (3.75), $\lim_{t\to\infty}(x-\phi s,x-\phi s)$ exists. Therefore $\lim_{t\to\infty}((x,x)-\phi^2(s,s))$ exists. Therefore by (3.74), $\lim_{t\to\infty}(x,x)$ exists. (3.76)

Let v_9 be any positive constant. Then by (3.35), (3.63), and (3.76), there is a time t_9 such that for $t > t_9$, $\min_U r_i^b \geq \min_{U+V} r_i^b > -v_9$. (3.77)

Therefore by (3.29), (3.31), (3.32), and (3.63), $\lim_{t\to\infty}(x^Q,x^Q) = 0$. Thus $\lim_{t\to\infty} x^Q = 0$. Therefore by (3.69), $\lim_{t\to\infty} x$ exists. (3.78)

Therefore for i in U+V, $\lim_{t\to\infty} r_i$ exists. Therefore by (3.30), (3.32), (3.63), and (3.77), $\lim_{t\to\infty} r_i \geq 0$ for i in U+V. (3.79)

Let YE be the set of numbers $\{i\}$ such that i is in U+V and $\lim_{t\to\infty} r_i = 0$. Let NE be the set of numbers $\{i\}$ such that i is in U+V and $\lim_{t\to\infty} r_i > 0$. Then by (3.63), there is a time t_{10} such that u and δ are halved when $t=t_{10}$, and such that for $t > t_{10}$ and i in NE, no error occurs on the ith inequality. Therefore by letting Δ denote the change during the time interval from t_{10} to ∞, (3.78), and (3.79), there is a set of numbers $\{\beta_i\}$ such that $\Delta(x,x) = 2(\Delta x,x+\Delta x)-(\Delta x,\Delta x) =$

$2(\sum_{YE}\beta_i c_i,x+\Delta x)-(\Delta x,\Delta x) \leq (2)\sum_{YE}\beta_i(c_i,x+\Delta x) = 0$. But there is a finite time when this result, (3.63), (3.76), and Condition 1 form a contradiction. Therefore by (3.49), (3.51), and (3.62), after a finite amount of time, an interval of length $\Delta t=\Omega$ occurs which contains no error on any inequality in E-Y. Thus by (3.26), the proof is complete. Q.E.D.

Some results from four computer simulation experiments will now be presented.

Experiment 1 compared the convergence time of two types of α-system (see reference 4) and one type of cycle BELA on a horizontal/vertical bar discrimination problem. Here is an outline of Experiment 1:

Perceptron type:	Binomial model elementary perceptron with a toroidally connected retina of simple S-units. Each A-unit had four excitatory input connections, no inhibitory input connections, and a threshold of 0.8. If the sum of the input signals to the R-unit was zero, the response was considered wrong.
Retinal dimensions:	20 by 20.
Number of A-units:	20.
Stimulus world:	The 40 possible 4 by 20 horizontal and vertical bars.
Training sequence:	As cycle BELA with: 1. $\Gamma = 40$ 2. alternation of the stimulus classes 3. the (i+2)th stimulus was the ith stimulus translated one retinal unit
Initial A-R weights:	Zero.
Criterion for saying that no solution exists:	When the type of cycle BELA shown in Table 1 was used, the first 100 cycles contained errors.
Number of perceptrons:	Out of the 40 generated, 25 had solutions.

Table 1 shows the total number of cycles containing errors for the last 20 perceptrons generated which had solutions (the first five perceptrons generated which had solutions were not run with any of the reinforcement procedures shown in Table 1).

Experiment 2 compared the convergence time of three types of cycle BELA on an E/F discrimination problem. The following is an outline of Experiment 2:

Perceptron type:	Simple perceptron with a toroidally connected retina of simple S-units and all possible S-A connections. Each S-A connection had a value randomly chosen from the set of integers $\{0,1,\ldots,9\}$. The output signal of an A-unit was equal to the algebraic sum, call if β, of its input signals if the one's digit of β was 5, 6, 7, 8, or 9; otherwise the output signal was zero. If the sum of the input signals to the simple R-unit was zero, the response was considered wrong.

Table 1. Experiment 1 convergence times.

Perceptron number.	Quantized α-system error correction reinforcement.	The following type of non-quantized α-system error correction reinforcement: at each error, quantized α-system reinforcement was repeatedly used until that stimulus was just correct.	Cycle BELA with δ=0, g=0.99, u=1, and doubling x between cycles whenever Condition 1 was satisfied.
1	120	128	26
2	49	50	19
3	254	229	35
4	55	54	19
5	34	43	14
6	321	333	55
7	52	53	9
8	65	24	14
9	463	382	44
10	94	109	22
11	2567	2355	87
12	492	416	43
13	102	98	25
14	203	140	27
15	37	40	14
16	760	720	51
17	22	26	16
18	276	246	48
19	34	29	13
20	59	69	13
Average:	302.95	277.20	29.70

Retinal dimensions: 3 by 5

Number of A-units: 18

Stimulus world: The 30 possible upright 2 by 5 E's and F's.

Training sequence: As cycle BELA with:

1. $\Gamma = 30$
2. alternation of the stimulus classes
3. for odd integers i, the (i+1)th stimulus was the F most similar to the E which was the ith stimulus
4. the (i+2)th stimulus was the ith stimulus slightly displaced

Initial A-R weights: Zero.

Criterion for saying that no solution exists: When the type of cycle BELA shown in the right-hand column of Table 2 was used, the first 100 cycles contained errors.

Number of perceptrons: Out of the 23 generated, 20 had solutions.

Table 2 shows the total number of cycles containing errors for the perceptrons which had solutions.

Table 2. Experiment 2 convergence times.

Perceptron number.	Cycle BELA with $\delta=0$, $g=0.99$, $u=1$, and doubling x between cycles whenever Condition 1 was satisfied.	Cycle BELA with $\delta=10^8$ for the first four cycles, $\delta=0$ after the first four cycles, $g=0.99$, $u=1$, and no doublings of x.	Cycle BELA with $\delta=10^8$ for the first four cycles, $\delta=0$ after the first four cycles, $g=0.75$, $u=1$, and no doublings of x.
1	49	37	28
2	10	7	8
3	22	16	15
4	23	26	21
5	5	6	5
6	8	8	8
7	47	31	31
8	15	17	18
9	30	22	21
10	11	8	8
11	11	10	9
12	23	17	16
13	7	7	7
14	70	70	68
15	24	24	21
16	86	67	52
17	7	9	9
18	106	80	63
19	126	68	54
20	5	7	6
Average:	34.25	26.85	23.40

Experiment 3 compared the convergence time of some perceptrons on two different bar discrimination problems. The following is an outline of Experiment 3:

Perceptron type:	As in Experiment 2.
Retinal dimensions:	10 by 10
Number of A-units:	10
Stimulus worlds:	The 20 possible 2 by 10 and 3 by 10 horizontal bars. The 20 possible 2 by 10 horizontal and vertical bars.
Training sequence:	As cycle BELA with: 1. $\Gamma = 20$ 2. alternation of the stimulus classes 3. the (i+2)th stimulus was the ith stimulus translated one retinal unit 4. in the horizontal bars problem, the (i+1)th stimulus and the ith stimulus had 20 S-units in common
Initial A-R weights:	Zero.
Criterion for saying that no solution exists:	The first 200 cycles contained errors.
Error correction procedure used:	Cycle BELA with $\delta=10^8$ for the first four cycles, $\delta=0$ after the first four cycles, g=0.75, u=1, and no doublings of x.
Number of perceptrons:	Out of the 40 generated, 16 had solutions on the horizontal bars problem, 18 had solutions on the horizontal and vertical bars problem, and 9 had solutions on both problems.

Table 3 shows the total number of cycles containing errors for the perceptrons which had solutions. A dash indicates that no solution exists.

Table 3. Experiment 3 convergence times.

Perceptron number.	Horizontal bars problem.	Horizontal and vertical bars problem.	Perceptron number.	Horizontal bars problem.	Horizontal and vertical bars problem.
39	6	-	19	-	-
38	7	18	18	-	-
37	17	-	17	-	-
36	-	43	16	-	18
35	8	26	15	13	-
34	5	9	14	-	-
33	-	7	13	-	60
32	-	-	12	-	6
31	18	-	11	26	7
30	-	-	10	8	-
29	-	-	9	7	10
28	-	-	8	-	-
27	14	-	7	-	-
26	4	11	6	-	47
25	17	19	5	11	28
24	6	-	4	-	-
23	-	5	3	-	-
22	-	-	2	-	8
21	5	8	1	-	-
20	-	50	0	-	-

Average for perceptrons with solutions on either problem:	10.75	21.11
Average for perceptrons with solutions on both problems:	10.00	15.11

Experiment 4 was like Experiment 3 except that the cases represented by a dash in Table 3 were not done, the stimulus sequence was different, and the error correction reinforcement procedures used were different. The following outline explains the stimulus sequence used in Experiment 4:

Training sequence: Each training stimulus was picked at random from the stimulus world.

Test sequence: After each 40 additional training stimuli were reinforced using an error correction procedure, the stimulus world was sequentially presented in what is called a "test". During each test, x was not altered and the number of errors which occurred was recorded.

Table 4 shows the total number of tests containing errors for each perceptron with a solution on the particular problem.

Table 4. Experiment 4 convergence times.

Perceptron number.	ABLE (see reference 2), where at any error on any inequality, say the ith inequality, $b_i=1$.		Integer BELA with $\delta=0$, $g=0.99$, $u=1$, and doubling x after tests whenever Condition 1 was satisfied.		Integer BELA with $\delta=10^8$ until after the second test, $\delta=0$ after the second test, $g=0.75$, $u=1$, and no doublings of x.	
	Horiz. bars problem	Horizontal and vertical bars problem	Horiz. bars problem	Horizontal and vertical bars problem	Horiz. bars problem	Horizontal and vertical bars problem
39	3		4		4	
38	11	137	4	21	4	16
37	95		13		11	
36		676		68		45
35	45	329	13	33	8	26
34	12	16	4	28	4	12
33		15		5		3
31	92		14		14	
27	123		13		7	
26	5	121	3	18	4	6
25	72	294	24	25	17	11
24	9		2		2	
23		9		5		5
21	1	24	1	12	2	9
20		9689		148		74
16		222		36		18
15	29		9		5	
13		>12750*		380		128
12		13		8		3
11	1772	21	69	6	28	7
10	29		7		7	
9	9	32	4	6	4	10
6		772		45		81
5	8	114	10	28	4	27
2		32		4		3
Average for the perceptrons with solutions on either problem:	144.68	>1403.66	12.12	48.66	7.81	26.88
Average for the perceptrons with solutions on both problems:	215.00	120.88	14.66	19.66	8.33	13.77

* Not completed because of the computing time involved.

Table 5. Horizontal bars problem percent-correct-in-test averages from Experiment 4.

Average of the 16 perceptrons which had solutions on the horizontal bars problem.			Test number.	Average of the 9 perceptrons which had solutions on both problems.		
ABLE.	Integer BELA with g=0.99.	Integer BELA with g=0.75.		ABLE.	Integer BELA with g=0.99.	Integer BELA with g=0.75.
72.5	71.0	60.7	1	75.0	70.6	58.9
79.4	82.2	66.6	2	78.9	81.2	66.7
82.2	87.5	86.3	3	82.3	87.8	89.5
80.0	88.2	87.9	4	83.9	91.7	91.2
84.7	89.1	92.9	5	85.0	92.8	95.0
83.8	91.0	95.7	6	85.6	93.9	96.2
84.4	91.0	96.0	7	87.8	93.4	96.7
85.0	94.7	97.9	8	87.3	96.2	97.8
88.8	93.2	96.6	9	92.8	94.5	96.7
90.0	94.4	96.9	10	91.2	96.2	96.2
91.3	99.1	99.7	20	95.0	98.4	99.5
93.8	99.7	100.0	30	95.6	99.5	100.0
95.0	99.1		40	95.6	98.4	
94.4	99.1		50	96.7	98.4	
95.7	99.4		60	97.8	98.9	
96.0	100.0		70	95.6	100.0	
96.3			80	97.8		
96.3			90	98.4		
98.2			100	97.8		
98.5			200	97.3		
98.5			300	97.3		
99.4			400	98.9		
99.4			500	98.9		
99.4			600	98.9		
98.8			700	97.8		
98.8			800	97.8		
99.1			900	98.4		
98.8			1000	97.8		
100.0			2000	100.0		

From Tables 5 and 6, one can plot some learning curves from Experiment 4 on 4 cycle semi-logarithmic graph paper. For a more complete description of the error correction reinforcement procedures mentioned in Tables 5 and 6, see Table 4 and the outline explaining the stimulus sequence used in Experiment 4.

Table 6. Horizontal and vertical bars problem percent-correct-in-test averages from Experiment 4.

Average of the 18 perceptrons which had solutions on the horizontal and vertical bars problem.			Test number.	Average of the 9 perceptrons which had solutions on both problems.		
ABLE.	Integer BELA with g=0.99.	Integer BELA with g=0.75.		ABLE.	Integer BELA with g=0.99.	Integer BELA with g=0.75.
68.9	69.5	55.3	1	67.3	69.5	53.9
74.2	76.2	59.8	2	75.6	75.0	60.0
75.0	79.2	83.1	3	75.6	81.2	85.0
77.5	79.8	84.2	4	76.7	77.8	81.2
77.0	85.0	86.2	5	76.2	83.4	84.5
81.7	85.3	88.7	6	84.5	82.8	85.0
81.4	87.8	91.4	7	83.9	87.8	90.6
78.7	86.2	89.8	8	78.4	87.3	91.2
77.3	87.0	90.9	9	76.7	86.7	91.2
83.4	86.7	92.0	10	84.5	85.0	93.4
88.4	92.3	95.0	20	88.9	94.5	96.7
89.8	95.3	97.3	30	90.0	97.8	100.0
90.6	97.0	97.5	40	89.5	100.0	
90.3	97.3	98.1	50	91.7		
90.0	98.1	98.4	60	90.6		
89.5	97.5	98.1	70	90.6		
89.2	98.4	98.7	80	89.5		
92.3	98.9	99.5	90	93.9		
92.5	97.8	99.5	100	95.0		
92.0	98.7	100.0	200	94.5		
95.6	98.9		300	98.4		
97.0	100.0		400	100.0		
97.0			500			
96.2			600			
97.8			700			
97.5			800			
97.5			900			
97.0			1000			
98.1			2000			
98.7			3000			
97.5			4000			
98.4			5000			
98.1			6000			
98.7			7000			
98.7			8000			
98.4			9000			
98.9			10000			

IV. The Problem of Finding Whether a Set of m n-dimensional Vectors, Over the Field of Integers, is Linearly Dependent or Linearly Independent

Let C be any m by n matrix of known constants. Let c_i be the ith column vector of C^t, where C^t is the transpose of C and $i=1,2,\ldots,m$. Choose one of the following definitions for the set of vectors $\{\delta_i\}$, where $i=1,2,\ldots,m$.

1. Let δ_i be the ith row vector of CC^t.
2. Let δ_i be the ith column vector of CC^t.

Theorem 4: For every set of m numbers $\{\beta_i\}$, $\sum_{i=1}^{m} \beta_i c_i = 0$ if and only if $\sum_{i=1}^{m} \beta_i \delta_i = 0$.

Proof: Assume $\sum_{i=1}^{m} \beta_i c_i = 0$. Then $\sum_{i=1}^{m} \beta_i (c_i, c_j) = 0$ for $j=1,2,\ldots,m$. Therefore $\sum_{i=1}^{m} \beta_i \delta_i = 0$. Now assume instead that $\sum_{i=1}^{m} \beta_i \delta_i = 0$. Then $\sum_{i=1}^{m} \beta_i (c_i, c_j) = 0$ for $j=1,2,\ldots,m$. Consequently $\left(\sum_{j=1}^{m} \beta_j c_j \, , \, \sum_{i=1}^{m} \beta_i c_i \right) = 0.$ Q.E.D.

Assume that the elements of C are integers and that one wishes to find whether the set of m n-dimensional vectors $\{c_i\}$ is linearly dependent or linearly independent. If $m > n$, the set $\{c_i\}$ is linearly dependent. By Theorem 4, if $m \leq n$, the determinant of CC^t is non-zero if and only if the set $\{c_i\}$ is linearly independent. The IBM 7090 computer flow diagram* in Figure 2 was therefore devised to deal with the problem of whether the determinant of D is zero or non-zero, where D is any m by m matrix of integers and where $m > 1$.

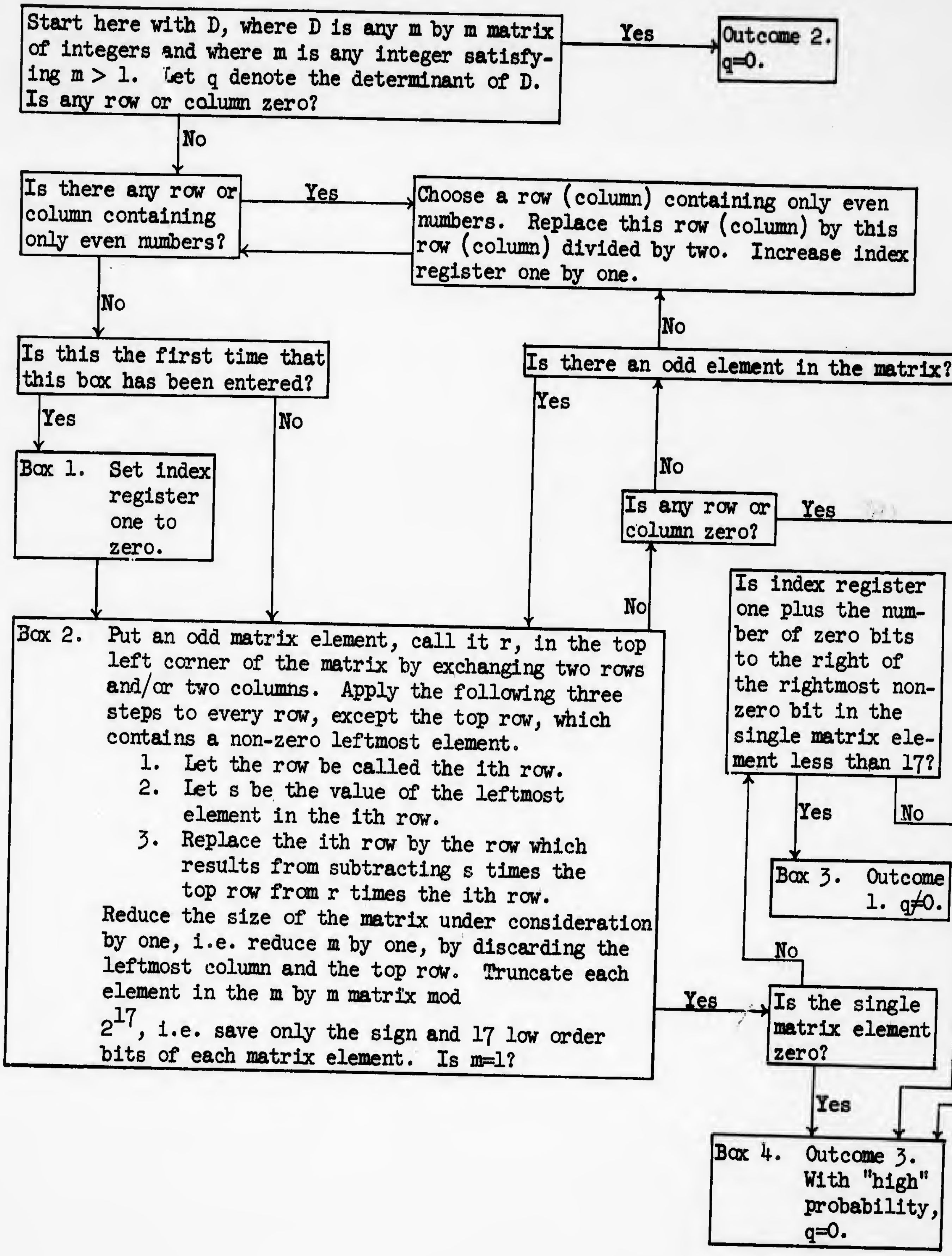
This flow diagram has the following three possible outcomes (exits):

Outcome 1. The determinant of D is non-zero.
Outcome 2. The determinant of D is zero.
Outcome 3. With "high" probability, the determinant of D is zero.

As will be shown later, "high" probability means that if the determinant of D is non-zero, this flow diagram will end in outcome 3 only if the value of the determinant of D is a non-zero integer multiple of 2^{17}. Outcome 3 was introduced so that this flow diagram could be made less time consuming.

* This flow diagram is easily adaptable to other computers or to a desk calculator.

Figure 2. Computer flow diagram.



Some explanation of the validity of the flow diagram in Figure 2 will now be presented.

If p is any integer, let [p] denote the result of truncating p mod 2^{17}, i.e. [p] is the number composed of the sign and 17 low order bits of p. Assume $a = 2^{17}b + c$, $d = 2^{17}e + f$, $g = 2^{17}h + c$, and $i = 2^{17}j + f$, where a,b,...,j are non-negative integers, $c < 2^{17}$, and $f < 2^{17}$. Then

$$|[a-d]| = \begin{cases} 0 & \text{if } c = f \\ 2^{17} + c - f & \text{if } c < f \text{ and } a > d \\ f - c & \text{if } c < f \text{ and } a < d \\ c - f & \text{if } c > f \text{ and } a > d \\ 2^{17} + f - c & \text{if } c > f \text{ and } a < d \end{cases} \text{ and}$$

$$|[g-i]| = \begin{cases} 0 & \text{if } c = f \\ 2^{17} + c - f & \text{if } c < f \text{ and } g > i \\ f - c & \text{if } c < f \text{ and } g < i \\ c - f & \text{if } c > f \text{ and } g > i \\ 2^{17} + f - c & \text{if } c > f \text{ and } g < i \end{cases} .$$

Therefore either $|[a-d]| = |[g-i]|$ or $|[a-d]| = 2^{17} - |[g-i]|$. (4.1)

Let d_i denote any element in the ith column of D. Then the determinant of D, call it q, can be written as

$q = \sum$ positive terms $(d_1 d_2 \ldots d_m) + \sum$ negative terms $(d_1 d_2 \ldots d_m)$. Then by (4.1), either

$|[\sum$ positive terms $([d_1][d_2]\ldots[d_m]) + \sum$ negative terms $([d_1][d_2]\ldots[d_m])]| = |[q]|$ or

$|[\sum$ positive terms $([d_1][d_2]\ldots[d_m]) + \sum$ negative terms $([d_1][d_2]\ldots[d_m])]| =$

$2^{17} - |[q]|$. (4.2)

Define a "bit-saving operation" to be any operation on an m by m matrix of integers which satisfies the two conditions given below. Let u and v denote the value of the 17 low order bits in the value of the determinant of the m by m matrix just before and just after the operation respectively.

1. If u=0, then v=0.
2. Assume that the less significant bits, i.e. the low order bits, are to the right. If u≠0, then the number of zero bits to the right of the rightmost non-zero bit in u is the same as the number of zero bits to the right of the rightmost non-zero bit in v.

Then by (4.2), the truncating operation in box 2 of the flow diagram in Figure 2 is a bit-saving operation. Therefore since all the other operations in box 2 are also bit-saving operations, box 2 as a whole is a bit-saving operation. Therefore box 3 is entered if and only if the value, of the determinant of the matrix under consideration when passing through box 1, has a non-zero bit in its 17 low order bits. Therefore the only way box 4 can be entered if q≠0 is for the 17 low order bits in the value, of the determinant of the matrix under consideration when passing through box 1, to be zero.

V. Minimal Universal Simple Perceptron Experiments

This section assumes some knowledge of perceptron terminology, which may be obtained from reference 4. A simple perceptron is defined in reference 4. A universal perceptron is any perceptron which has a solution for every classification of all the possible stimuli. A minimal perceptron is any perceptron which has only as many variable weights as the number of possible stimuli.

Consider any minimal simple perceptron which has w simple S-units, where w is any integer satisfying $w > 1$. Then since (2^w-1) is the number of possible stimuli (each stimulus must have at least one active S-unit), this perceptron has (2^w-1) A-units. Therefore one can form the (2^w-1) by (2^w-1) matrix D by having each row represent a different stimulus and by having each column represent the output signal of a different A-unit. Then by (3.19), this perceptron is a universal perceptron if and only if the determinant of D is non-zero. Therefore if the output signal of each A-unit in this perceptron is an integer, then the computer flow diagram in Figure 2 can be applied to D to see whether this perceptron is universal (Outcome 1), not universal (Outcome 2), or probably not universal (Outcome 3). That is what was done in the two experiments now to be described.

Experiment 1:

Perceptron type: Minimal simple perceptron with w simple S-units and all possible S-A connections. Each S-A connection had a value randomly chosen from the set of integers $\{0,1,\ldots,2^z-1\}$, where z is a parameter. Assume that the least significant bit is called the 1st bit, that the bit next to the 1st bit is called the 2nd bit, etc. The output signal of an A-unit was equal to the algebraic sum, call it β, of its input signals if the zth bit of β was a one; otherwise the output signal was zero. If the sum of the input signals to the simple R-unit was zero, the response was considered wrong.

Number of perceptrons: Ten perceptrons were generated for each line in Table 7.

Table 7. Results of Experiment 1.

w	z	Number which were not universal.	Number which were probably not universal.	Number which were universal.
3	1	7	3	0
3	3	3	0	7
3	6	0	0	10
7	1	6	4	0
7	3	0	2	8
7	6	0	0	10

Experiment 2:

Perceptron type: Minimal simple perceptron with w simple S-units and all possible S-A connections. Each S-A connection had a value randomly chosen from the set of integers $\{-(2^z-1),-(2^z-2),\ldots,-2,-1,0,0,1,2,\ldots,2^z-2,2^z-1\}$, where z is a parameter.

Consider any A-unit, say the ith. Consider any possible stimulus, say the jth. Let Θ denote the threshold of the ith A-unit. Let β denote the algebraic sum of the input signals arriving at the ith A-unit as a result of the jth stimulus. Then the output signal of the ith A-unit, as a result of the jth stimulus, was equal to

$$\begin{Bmatrix} \beta-\Theta \text{ if } \beta-\Theta > 0 \\ 0 \text{ if } \beta-\Theta \leq 0 \end{Bmatrix}.$$

If $\beta-\Theta > 0$, say that the ith A-unit was "active" to the jth stimulus; if $\beta-\Theta \leq 0$, say that the ith A-unit was "inactive" to the jth stimulus.

The threshold of each A-unit was an individually chosen constant such that each A-unit was active to about (100y) percent of the stimulus world, where y is a parameter. For example, consider any A-unit, say the ith. Let θ denote the threshold of the ith A-unit. The threshold of the ith A-unit was chosen such that:

1. θ was an integer.
2. the ith A-unit was active to equal or less than $[y(2^w-1)]$ of the (2^w-1) possible stimuli, where [ ] denotes the integer part of the contained number.
3. if θ had been decreased by one, the ith A-unit would have been active to more than $[y(2^w-1)]$ of the (2^w-1) possible stimuli, where [ ] denotes the integer part of the contained number.

If the sum of the input signals to the simple R-unit was zero, the response was considered wrong.

Number of perceptrons: Ten perceptrons were generated for each line in Table 8.

Table 8. Results of Experiment 2.

y	w	z	Number which were not universal.	Number which were probably not universal.	Number which were universal.
0.50	3	2	7	1	2
0.50	3	5	1	1	8
0.50	7	2	0	2	8
0.50	7	5	0	0	10
0.75	3	2	1	6	3
0.75	3	5	0	0	10
0.75	7	2	0	2	8
0.75	7	5	0	0	10

By comparing Tables 7 and 8, it is seen that for a given value of w, if the number of possible values for an S-A connection is about the same in Experiments 1 and 2, then about the same percentage of perceptrons will be universal in Experiments 1 and 2. From Experiments 1 and 2, it appears that a small minimal universal perceptron is not difficult to realize using "random" S-A connections.

VI. The Problem of How Many m-dimensional Orthants Are Achievable by Linear Combinations of n m-dimensional Vectors

Unless explicitly stated otherwise, it is assumed that all constants and variables mentioned below are real and finite.

An m-dimensional orthant can be thought of as a set consisting of all m-dimensional vectors whose components satisfy the following:

1. The components are all non-zero.
2. The corresponding components agree in sign.

Let M(m,n) denote the maximum number of m-dimensional orthants achievable by linear combinations of n m-dimensional vectors. Then using a result proven in Appendix B of Reference 1,

$$M(m,n) = \begin{cases} 2^m \text{ when } n \geq m \\ 2\left(\binom{m-1}{0}+\binom{m-1}{1}+\dots+\binom{m-1}{n-1}\right) \quad \text{when } m \geq n \end{cases}. \tag{6.1}$$

Let C be any m by n matrix of known constants. Let c_i be the ith column vector of C^t, where C^t is the transpose of C and $i=1,2,\dots,m$. From here on, let "number of achievable orthants" be an abbreviation of "number of m-dimensional orthants achievable by linear combinations of the n m-dimensional column vectors of C".

Since the number of achievable orthants is zero if any $c_i = 0$, from here on it will be assumed that each c_i is non-zero.

Let a "con" be any non-empty set of linearly dependent c_i's such that if any one member of the set is discarded, the remaining members of the set are linearly independent. Let a con's α be any non-zero m-dimensional column vector $\alpha = \begin{bmatrix}\alpha_1 \\ \vdots \\ \alpha_m\end{bmatrix}$ such that $C^t\alpha = 0$ and such that for $1 \leq i \leq m$, α_i is non-zero if and only if c_i is a member of this con.

Lemma 1: Any m-dimensional orthant cannot be achieved by a linear combination of the column vectors of C if and only if there is a con with an α such that every non-zero component of α has the same sign as the corresponding component of the orthant.

Proof: By (3.19), the "if" part of this lemma is true.

The flow diagram in Figure 3 proves the "only if" part of this lemma.

Figure 3. Flow diagram used in the proof of Lemma 1.

Start here. Set $j = 2$. Choose any m-dimensional orthant, call it Ω, which cannot be achieved by a linear combination of the column vectors of C.

↓

Has every set of j c_i's been chosen in box 1? — No → Box 1. — Yes → Increase j by one. (↑ back to the question)

Box 1. Choose a set of j c_i's, call it A, which has not yet been chosen in this box. Let B denote the set of j different numbers $\{i\}$ such that c_i is in A.

Is there a vector $\beta = \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_m \end{bmatrix}$ such that:

1. $\beta_i = 0$ for i not in B.
2. for i in B, either $\beta_i = 0$ or β_i has the same sign as the ith component of Ω.
3. for i in B, $(c_i, \sum_{k=1}^{m} \beta_k c_k)$ has the same sign as the ith component of Ω, where (,) denotes the inner product of the contained vectors.

(This condition cannot hold when $j=m$ because of the definition of Ω)?

Yes → back to "Has every set of j c_i's been chosen in box 1?"

No ↓

By (3.19), one can find a non-zero vector $\omega = \begin{bmatrix} \omega_1 \\ \vdots \\ \omega_m \end{bmatrix}$ such that:

1. $\omega_i = 0$ for i not in B.
2. for i in B, either $\omega_i = 0$ or ω_i has the same sign as the ith component of Ω.
3. $\sum_{i=1}^{m} \omega_i c_i = C^t \omega = 0.$ (6.2)

Choose such an ω. By (3.19) and the way in which B was obtained, $\omega_i \neq 0$ for i in B. (6.3)

Assume there is a linearly dependent set of c_i's, call it F, such that:

1. F is a subset of A.
2. F has less than j members and more than zero members.

Let G denote the set of numbers $\{i\}$ such that c_i is in F. Then by (6.3), one can find a non-zero vector $\delta = \begin{bmatrix} \delta_1 \\ \vdots \\ \delta_m \end{bmatrix}$ such that:

1. $\delta_i = 0$ for i not in G.
2. for $i = 1,2,\ldots,m$, $|\delta_i| \leq |\omega_i|$.
3. $\delta_i = -\omega_i$ for some i in G.
4. $\sum_{i=1}^{m} \delta_i c_i = C^t \delta = 0.$

Then $(\omega + \delta)$ satisfies (6.2) but not (6.3), which is a contradiction.

So by (6.2) and (6.3), A is a con with an α equal to ω such that every non-zero component of ω has the same sign as the corresponding component of Ω. Q.E.D.

Lemma 2: Each α of any con, say the jth con, is a non-zero multiple of every other α of the jth con.

Proof: Consider any con, say the jth con. Assume there are two α's of the jth con, say β and δ, such that β and δ are not non-zero multiples of each other. Then there is a non-zero number a such that:

1. $(a\beta+\delta)$ has more zero components than β.
2. $(a\beta+\delta)$ is non-zero.
3. $\sum_{i=1}^{m}(a\beta_i+\delta_i)c_i = C^t(a\beta+\delta) = 0.$

Therefore the jth con is not a con, which is a contradiction. Q.E.D.

Lemma 3: Let R be any set of u r-dimensional column (row) vectors, where r and u are any positive integers. For i=1,2,...,u, let π_i denote the ith member of R. Then there is a positive number w such that if ω denotes any r-dimensional column (row) vector whose length is at most w, and if ω is added to any π_i, say $\pi_i = \pi_j$, then every subset of R which was linearly independent is still linearly independent.

Proof: Let S_k denote the set of numbers $\{i\}$ such that the kth subset of R contains π_j and was linearly independent before ω was added to π_j. Let

$$w = 0.5\left[\underset{S_k}{\text{minimum}}\left(\underset{(\text{numbers } \beta_i)}{\text{minimum}} \left\| \pi_j - \sum_{\left\{\substack{i \text{ in } S_k \\ i \neq j}\right\}} \beta_i\pi_i \right\|\right)\right].$$ Q.E.D.

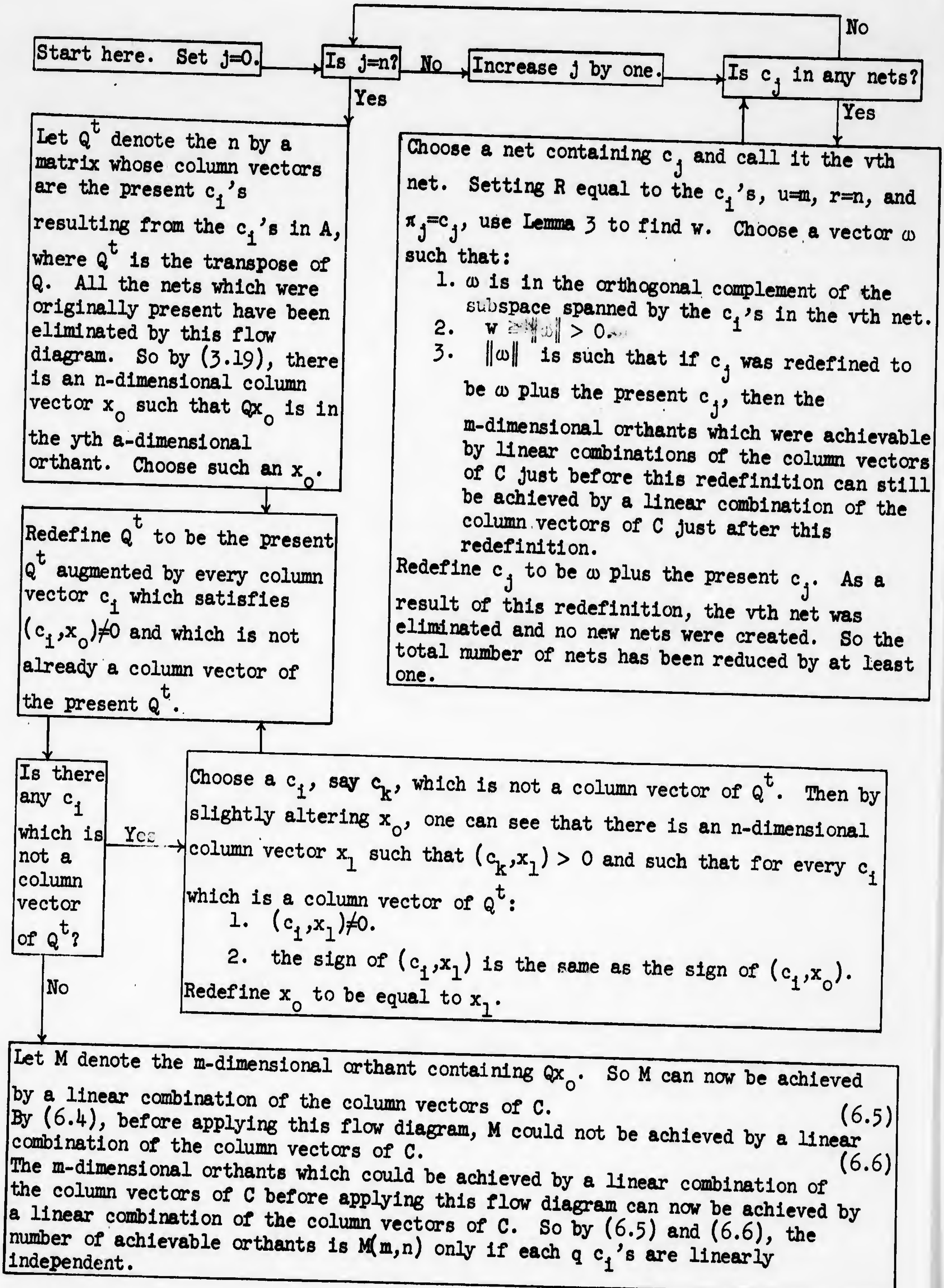
Theorem 5: Let $q = \min\{m,n\}$. Then the number of achievable orthants is $M(m,n)$ if and only if each q c_i's are linearly independent.

Proof: By (3.19) and (6.1), this theorem is true if $n \geq m$. So for the rest of this proof, assume $m > n$. Let a "net" be any con which has less than n+1 members.

The "only if" part of this theorem will now be proven. Assume there is a set of q c_i's which is linearly dependent. Then there is at least one net. Let A denote any particular net. Let a denote the number of members in A. Let P^t denote the n by a matrix whose column vectors are the c_i's in A, where P^t is the transpose of P. Then by (3.19), there is an a-dimensional orthant, say the yth, such that no matter what n-dimensional column vector x is chosen, Px cannot achieve the yth a-dimensional orthant. (6.4)

The flow diagram in Figure 4 completes the proof of the "only if" part of this theorem.

Figure 4. Flow diagram used in the proof of Theorem 5.



Now the "if" part of this theorem will be proved. For the rest of this proof, assume each q c_i's are linearly independent. Then each con has at least q+1 members and each q+1 c_i's are a con. (6.7)

Assume that there are m-n+1 α's of cons which are linearly independent. Then these linearly independent α's can be linearly combined to form a non-zero m-dimensional column vector δ such that:

1. δ has at least m-n zero components.
2. $C^t\delta = 0$.

Thus there are q c_i's which are linearly dependent, which is a contradiction. Therefore there are at most m-n linearly independent α's of cons. Therefore by (6.1), the maximum number of m-dimensional orthants achievable by linear combinations of the α's of cons is equal or less than M(m,m-n). (6.8)

Consider any m-dimensional orthant, say the zth, which cannot be achieved by a linear combination of the column vectors of C. Then by Lemma 1, there is a con, say the sth, with an α, say γ, such that every non-zero component of γ has the same sign as the corresponding component of the zth m-dimensional orthant. Let B be a set consisting of exactly one α of each con which has exactly q+1 members, exactly one of which is not also a member of the sth con. Then by (6.7), there is a linear combination of γ and the elements of B which is in the zth m-dimensional orthant.

Thus by (6.8), at most M(m,m-n) m-dimensional orthants cannot be achieved by a linear combination of the column vectors of C. Therefore since (6.1) implies that $M(m,m-n) + M(m,n) = 2^m$, the number of achievable orthants is M(m,n). Q.E.D.

<u>Theorem 6</u>: Let r be any integer satisfying $m > r > 0$. Then if each r+1 c_i's are linearly dependent, the number of achievable orthants is at most M(m,r).

Proof: Assume each r+1 c_i's are linearly dependent. Then there are at most r linearly independent c_i's. Thus there are at most r linearly independent column vectors of C. Therefore by (6.1), the number of achievable orthants is at most M(m,r). Q.E.D.

<u>Theorem 7</u>: Let r be any integer satisfying $m \geq r > 0$. Then if each r c_i's are linearly independent, the number of achievable orthants is at least M(m,r).

Proof: In this proof, assume each r c_i's are linearly independent. Let A denote any m by r matrix. Let B denote the set of r by r matrices whose row vectors are any r row vectors of A. Let B_i denote the ith element of B.

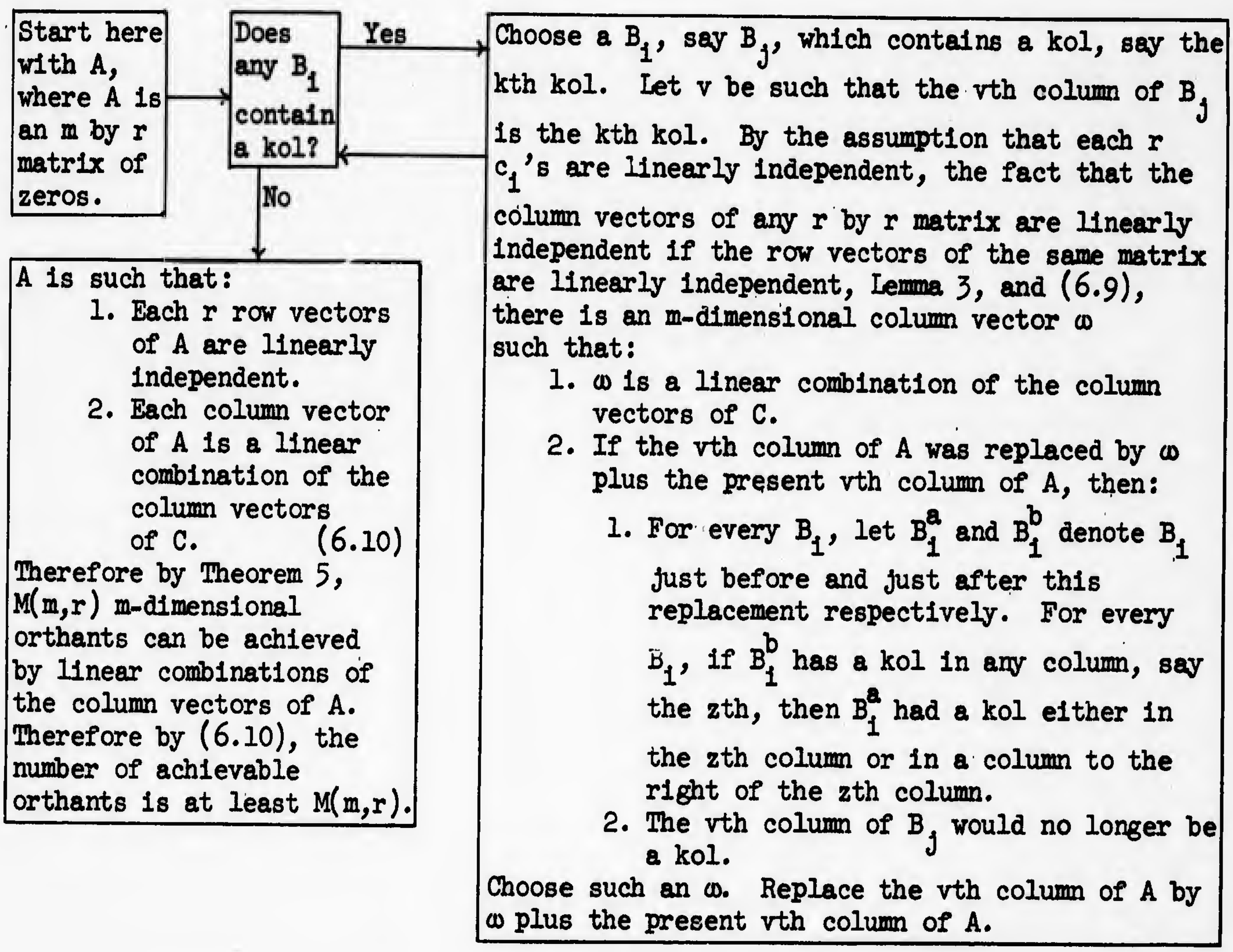
For each B_i, define a "kol" to be any column vector of B_i such that:

1. If there are any column vectors of B_i to the right of the kol, then they are linearly independent.
2. The kol together with the column vectors of B_i to the right of the kol are linearly dependent.

Then each B_i can have at most one kol. (6.9)

The flow diagram in Figure 5 completes this proof.

Figure 5. Flow diagram used in the proof of Theorem 7.



On the basis of Lemma 1, Lemma 2, Theorem 5, Theorem 6, and Theorem 7, it has been suggested that one or both of the following conjectures might be true. But it will be shown that both of the following conjectures are false.

<u>Conjecture 1</u>: If one knows only how many pairs, how many triplets,..., and how many m-tuplets of c_i's are linearly dependent, then one can find the number of achievable orthants.

Counterexample: If $C = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 2 \\ 1 & 3 \\ 1 & 1 \end{bmatrix}$, then three pairs, $\binom{6}{3}$ triplets, $\binom{6}{4}$ quadruplets, $\binom{6}{5}$ quintuplets, and one sextuplet of c_i's are linearly dependent; the number of achievable orthants is eight. If $C = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 1 & 1 \\ 1 & 1 \end{bmatrix}$, then three pairs, $\binom{6}{3}$ triplets, $\binom{6}{4}$ quadruplets, $\binom{6}{5}$ quintuplets, and one sextuplet of c_i's are linearly dependent; the number of achievable orthants is six.

Conjecture 2: If one knows only how many pairs, how many triplets,..., and how many m-tuplets of c_i's are cons, then one can find the number of achievable orthants.

Counterexample: If $C = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$, then three pairs, zero triplets, zero quadruplets, zero quintuplets, and zero sextuplets of c_i's are cons; the number of achievable orthants is sixteen. If $C = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix}$, then three pairs, zero triplets, zero quadruplets, zero quintuplets, and zero sextuplets of c_i's are cons; the number of achievable orthants is eight.

The following two definitions were motivated by (3.19). Say that any m-dimensional orthant is a "strictly non-achievable" orthant if and only if there is an m-dimensional column vector β such that $C^t\beta = 0$ and such that β is in the m-dimensional orthant. Say that any m-dimensional orthant is a "partially achievable" orthant if and only if the orthant is neither a strictly non-achievable orthant nor achievable by a linear combination of the column vectors of C.

<u>Theorem 8</u>: Let r be any integer satisfying $m \geq r > 0$. If the rank of C is r, then:

1. The number of strictly non-achievable orthants is at most $M(m,m-r)$.
2. The number of achievable orthants is at most $M(m,r)$.
3. The number of achievable orthants is $M(m,r)$ if and only if each r c_i's are linearly independent.

Proof: In this proof, assume the rank of C is r. Then there are at most r linearly independent column vectors of C. (6.11)

By (6.11) and (6.1), part 2 is true.

By (6.11) and Theorem 5, part 3 is true.

Let B denote the subspace containing every vector δ which satisfies $C^t\delta = 0$. Then the dimension of B is m-r. Thus the number of m-dimensional orthants achievable by linear combinations of the vectors in B is at most $M(m,m-r)$. Therefore part 1 is true. Q.E.D.

<u>Theorem 9</u>: Let r be any integer satisfying $m > r > 0$. If each r c_i's are linearly independent and each r+1 c_i's are linearly dependent, then:

1. The number of strictly non-achievable orthants is $M(m,m-r)$.
2. The number of partially achievable orthants is zero.
3. The number of achievable orthants is $M(m,r)$.

Proof: In this proof, assume each r c_i's are linearly independent and each r+1 c_i's are linearly dependent. Then:

1. By Theorems 6 and 7, part 3 is true.
2. Each con has at least r+1 members and each r+1 c_i's are a con. (6.12)

By (6.1), $M(m,m-r) + M(m,r) = 2^m$. (6.13)

Therefore by part 3, the number of m-dimensional orthants which cannot be achieved by a linear combination of the column vectors of C is $M(m,m-r)$. (6.14)

Consider any m-dimensional orthant, say the zth, which cannot be achieved by a linear combination of the column vectors of C. Then by Lemma 1, there is a con, say the sth, with an α, say γ, such that every non-zero component of γ has the same sign as the corresponding component of the zth m-dimensional orthant. Let B be a set consisting of one α of each con which has r+1 members, exactly one of which is not also a member of the sth con. Then by (6.12), there is a linear combination of γ and the elements of B which is in the zth m-dimensional orthant. Therefore there is a vector β in the zth m-dimensional orthant such that $C^t\beta = 0$. Therefore the zth m-dimensional orthant is a strictly non-achievable orthant.

Thus by (3.19) and (6.14), part 1 is true.

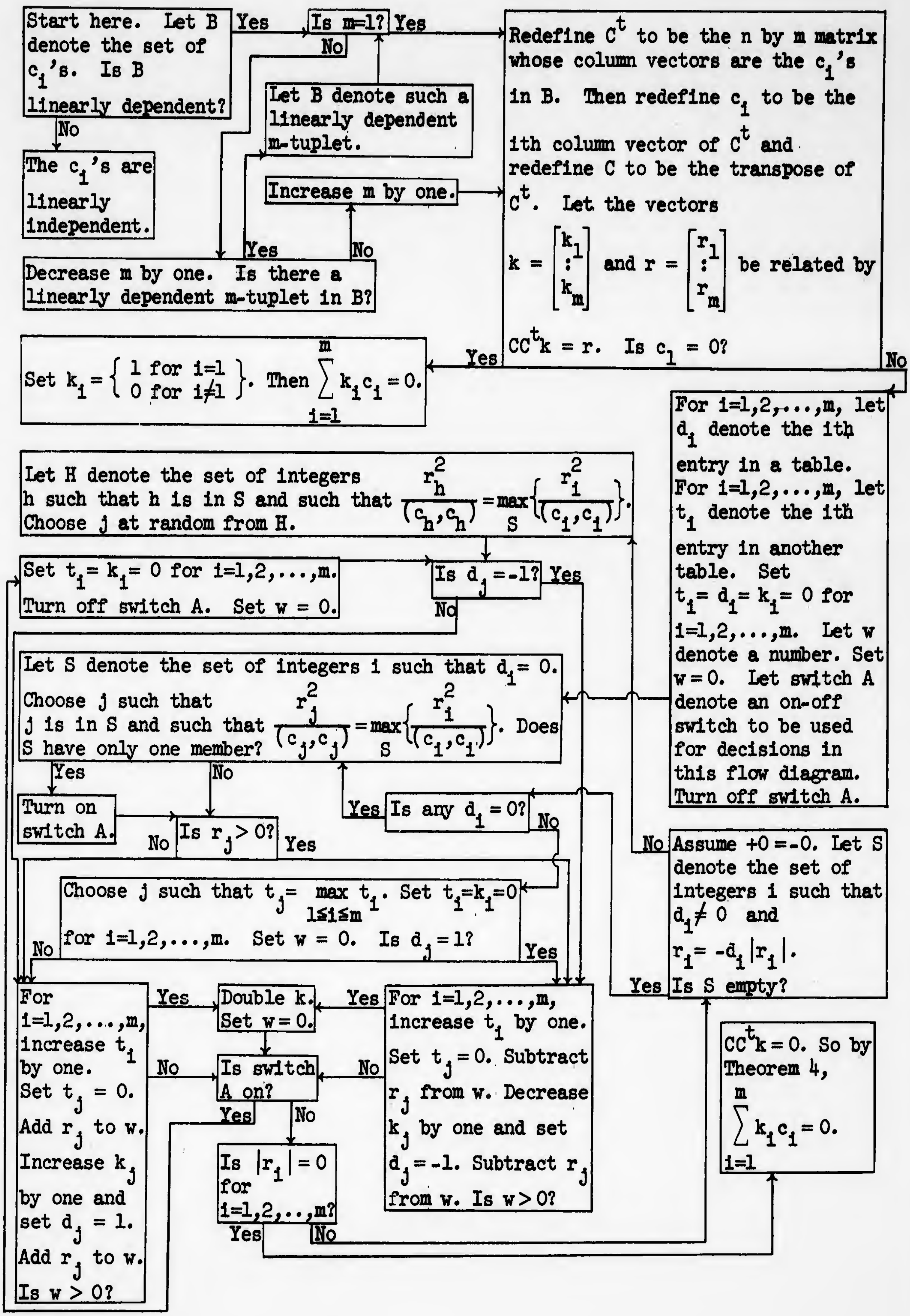
By (6.13), part 1, and part 3, part 2 is true. Q.E.D.

To find which m-dimensional orthants are achievable by a linear combination of the column vectors of C and/or to find the exact number of achievable orthants, one can form an "α-table" consisting of two α's, each a negative multiple of the other, of each con. Then by Lemmas 1 and 2, any m-dimensional orthant cannot be achieved by a linear combination of the column vectors of C if and only if there is an α in the α-table such that every non-zero component of this α has the same sign as the corresponding component of the orthant. Section IV, Lemma 2, and the following facts sometimes are useful in constructing an α-table:

1. Any con cannot be a subset of any other con.
2. If β is any linear combination of the α's in the α-table, then $C^t\beta = 0$.

Assume that the elements of C are integers. Then to see if the c_i's are linearly dependent, and if so, to find a linear relation among the c_i's, one can solve the Diophantine problem outlined on page 98 of reference 3. But in solving this Diophantine problem, the numbers involved often become too large to be easily handled. To avoid this difficulty, one can apply the flow diagram in Figure 6 to the c_i's. In a finite amount of time, the flow diagram in Figure 6 might either find that the c_i's are linearly independent or find a linear relation among a linearly dependent subset of the c_i's. Some theory and dozens of hand-computed examples went into the design of the flow diagram in Figure 6 in an effort to make it both simple and universal, but not very time consuming. The method outlined in Section IV can be used for the linear dependence tests in the flow diagram in Figure 6, although this might introduce a non-zero probability that the flow diagram in Figure 6 would never converge.

Figure 6. Flow diagram.



VII. References

1. Joseph, R.D., Contributions to Perceptron Theory, Cornell Aeronautical Laboratory Report No. VG-1196-G-7, Buffalo, 1960.
2. Kesler, C., Analysis and Simulations of a Nerve Cell Model, Cognitive Systems Research Program Report No. 2, Cornell University, Ithaca, 1961.
3. Niven, I., and Zuckerman, H.S., An Introduction to the Theory of Numbers, New York, John Wiley and Sons, 1960.
4. Rosenblatt, F., Principles of Neurodynamics; Perceptrons and the Theory of Brain Mechanisms, Washington, D.C., Spartan Books, 1962.

TWO THEOREMS ON MAJORITY DECISIONS

By F. Rosenblatt

The following two theorems were discovered in the course of recent investigations of perceptron memory mechanisms (Ref. 3). While they do not bear directly on any of the memory systems currently in use, the increasing use of majority decision schemes for such systems as Widrow's Madaline suggests that they may be of general interest. The first theorem, in particular, may be of interest in connection with optimum administrative structures, since it suggests that a committee with two equally reliable members can make decisions no better than a committee with one member (or, that a jury of twelve is no fairer than a jury of eleven).

THEOREM 1: Given any N interchangeable random variables, where N is odd, the probability that the majority of the variables are positive is equal to the probability that the majority of N+1 variables are positive (where ties are counted positive with probability .5).

Proof: Let P_m = probability that m or more events occur, out of a set of N possible events.

Let $\pi(k)$ = probability that any given set of k variables are positive. Since the given variables are interchangeable, $\pi(k)$ is identical for any choice of k variables.

Let $S_k = \sum p_k$ where p_k = probability that some particular set of k events occur, and the summation is over all possible sets of k events.

Then from Feller (Ref 2, Page 74) we have the equation:

$$P_m = S_m - \binom{m}{m-1} S_{m+1} + \binom{m+1}{m-1} S_{m+2} - \binom{m+2}{m-1} S_{m+3} + \dots \pm \binom{N-1}{m-1} S_N \quad (1)$$

Due to the interchangeability of the random variables, in the present case, it is clear that all terms p_k in the sum S_k are identical, and equal to $\pi(k)$. Consequently,

$$S_k = \binom{N}{k}\pi(k) \quad .$$

Now suppose N to be odd, and let $m = [N/2]+1 = \frac{N}{2} + \frac{1}{2}$. Then the probability that a majority of N variables is positive is given by:

$$P_{maj}(N) = P_m = \binom{N}{m}\pi(m) - \binom{m}{m-1}\binom{N}{m+1}\pi(m+1) + \binom{m+1}{m-1}\binom{N}{m+2}\pi(m+2)$$

$$- \binom{m+2}{m-1}\binom{N}{m+3}\pi(m+3) + \ldots \pm \binom{N-1}{m-1}\pi(N) \qquad (2)$$

For the majority of N+1 elements (where N is odd) we have the corresponding probability (giving half-weight to those cases where exactly m variables are positive),

$$P_{maj}(N+1) = \frac{1}{2}P_{[m]} + P_{m+1} \qquad (3)$$

where $P_{[m]}$ = probability that exactly m out of N+1 variables are positive, and P_{m+1} is the probability that m+1 or more variables are positive. After Feller, we have

$$P_{[m]} = S_m - \binom{m+1}{m}S_{m+1} + \binom{m+2}{m}S_{m+2} - \cdots \pm \binom{N+1}{m}S_{N+1} \qquad (4)$$

Thus, expanding (3) yields:

$$P_{maj}(N+1) = \frac{1}{2}\left[\binom{N+1}{m}\pi(m) - \binom{m+1}{m}\binom{N+1}{m+1}\pi(m+1) + \binom{m+2}{m}\binom{N+1}{m+2}\pi(m+2)\ldots \pm \binom{N+1}{m}\pi(N+1)\right]$$

$$+ \binom{N+1}{m+1}\pi(m+1) - \binom{m+2}{m}\binom{N+1}{m+2}\pi(m+2)\ldots \mp \binom{N}{m}\pi(N+1)$$

$$= \frac{1}{2}\binom{N+1}{m}\pi(m) - \left[\frac{1}{2}\binom{m+1}{m}\binom{N+1}{m+1} - \binom{N+1}{m+1}\right]\pi(m+1) + \left[\frac{1}{2}\binom{m+2}{m}\binom{N+1}{m+2} - \binom{m+1}{m}\binom{N+1}{m+2}\right]\pi(m+2)$$

$$\cdots \pm \left[\frac{1}{2}\binom{m+k}{m}\binom{N+1}{m+k} - \binom{m+k-1}{m}\binom{N+1}{m+k}\right]\pi(m+k)\ \ldots$$

$$\cdots \pm \left[\frac{1}{2}\binom{N+1}{m} - \binom{N}{m}\right]\pi(N+1) \qquad (5)$$

To prove the theorem, it is necessary to show that $P_{maj}(N) = P_{maj}(N+1)$.

Note that $P_{maj}(N)$ depends only on $\pi(m)$, $\pi(m+1),\ldots,\pi(N)$, while $P_{maj}(N+1)$ appears to depend on $\pi(N+1)$ as well. However, the coefficient of $\pi(N+1)$ in equation (5) can easily be shown to be equal to zero, so that we are left with terms in $\pi(m)\ldots\pi(N)$ in both equations.

For the remaining terms, the coefficients in Equation (2) take the form

$$\pm \binom{m+k-1}{m-1}\binom{N}{m+k} \qquad (6)$$

while in equation (5) the corresponding coefficient is

$$\pm \left[\frac{1}{2}\binom{m+k}{m}\binom{N+1}{m+k} - \binom{m+k-1}{m}\binom{N+1}{m+k}\right]. \qquad (7)$$

The signs of corresponding terms always agree in both equations. Therefore it remains to be shown that (6) and (7) are equal.

$$\binom{m+k-1}{m-1}\binom{N}{m+k} - \frac{1}{2}\binom{m+k}{m}\binom{N+1}{m+k} + \binom{m+k-1}{m}\binom{N+1}{m+k}$$

$$= \frac{(m+k-1)!\ N!}{k!\ (m-1)!\ (N-m-k)!\ (m+k)!} - \frac{(m+k)!\ (N+1)!}{2k!\ m!\ (N+1-m-k)!\ (m+k)!} + \frac{(m+k-1)!\ (N+1)!}{(k-1)!\ m!\ (N+1-m-k)!(m+k)}$$

$$= \frac{1}{2m!\ (m+k)!}\Big[2m[(m+k-1)(m+k-2)\ldots(k+1)][(N)(N-1)\ldots(N-m-k+1)]$$

$$-[(m+k)(m+k-1)\ldots(k+1)][(N+1)(N)\ldots(N-m-k+2)]$$

$$+2[(m+k-1)(m+k-2)\ldots(k)][(N+1)(N)\ldots(N-m-k+2)]\Big]$$

$$= \frac{1}{C}\ (2m(N-m-k+1)-(m+k)(N+1) + 2k(N+1)]$$

$$= \frac{1}{C}\left[2\left(\frac{N}{2}+\frac{1}{2}\right)\left[N-\left(\frac{N}{2}+\frac{1}{2}\right)-k+1\right] - \left[\left(\frac{N}{2}+\frac{1}{2}\right)+k\right](N+1)+2kN+2k\right] = 0.$$

This completes the proof of the theorem.

THEOREM 2: Given N uniformly correlated Gaussian variables, with mean h and unit variance, the limit (as N becomes infinite) of the probability that the majority are positive is $\phi\left(\frac{h}{\sqrt{\rho}}\right)$, where ρ = correlation between any pair of variables, and $\phi(Z)$ is the cumulative normal probability distribution, from $-\infty$ to Z.

Proof: The problem is equivalent to determining the probability that N/2 or more standardized variables (with mean zero and unit variance) are greater than -h, in an N-dimensional Gaussian distribution. This is given by

$$P_{maj} = \sum_{k=[N/2]+1}^{N} \binom{N}{k} P_k \tag{8}$$

where P_k = probability that exactly k out of N variables are $< h$, and N-k are $< -h$.

$$P_k = \int_{x_1=-\infty}^{h} \cdots \int_{x_k=-\infty}^{h} \int_{x_{k+1}=-\infty}^{-h} \cdots \int_{x_N=-\infty}^{-h} \phi' (x_1, x_2, \ldots x_N) dx_1 dx_2 \ldots dx_N \tag{9}$$

where ϕ' = the joint normal density function of the variables $x_1 \ldots x_N$.

After Curnow and Dunnett (Ref 1), for the case of uniformly correlated variables, this is equal to

$$P_k = \int_{-\infty}^{\infty} \prod_{i=1}^{N} \phi\left(\frac{h_i - Y\sqrt{\rho}}{\sqrt{1-\rho}}\right) \phi'(Y) dY \tag{10}$$

where $h_i = h$ for $i = 1, 2, \ldots k$, and $h_i = -h$ for $i = k+1, k+2, \ldots N$. This is equivalent to the expression

$$P_k = \int_{-\infty}^{\infty} \phi^k\left(\frac{h-Y\sqrt{\rho}}{\sqrt{1-\rho}}\right) \phi^{N-k}\left(\frac{-h - Y\sqrt{\rho}}{\sqrt{1-\rho}}\right) \phi'(Y) dY$$

$$= \int_{-\infty}^{\infty} F^k (1-F)^{N-k} \phi'(Y) dY \tag{11}$$

where $F = \phi\left(\frac{h-Y\sqrt{\rho}}{\sqrt{1-\rho}}\right)$.

Substituting (11) in equation (8), and moving the summation inside the integral, yields

$$P_{maj} = \int_{-\infty}^{\infty} \sum_{k=[N/2]+1}^{N} \binom{N}{k} F^k (1-F)^{N-k} \phi'(Y)dY \tag{12}$$

The integrand is seen to be the product of a cumulative binomial probability and a normal density function.

Now if N is large, the binomial distribution in (12) will have practically all of its mass concentrated in the neighborhood of the expected value, $k=FN$. Thus if $k = FN$ is included in the range of summation, the cumulative binomial distribution can be taken equal to 1, while if $k = FN$ is not included in the range of summation, the sum can be taken to be zero (for sufficiently large N). But it is clear that $k = FN$ will be included in the range of summation if $F > .5$, which will be true if $\frac{h-Y\sqrt{\rho}}{\sqrt{1-\rho}} > 0$, or $Y > -h/\sqrt{\rho}$. Thus, for $Y > -h/\sqrt{\rho}$, $\phi'(Y)$ in equation (12) is multiplied by 1, and for $Y < -h/\sqrt{\rho}$ it is multiplied by zero. Consequently, Equation (12) can be rewritten (for large N)

$$P_{maj} \xrightarrow[N\to\infty]{} \int_{\frac{-h}{\sqrt{\rho}}}^{\infty} \phi'(Y)dY$$

$$= \phi\left(\frac{h}{\sqrt{\rho}}\right) \tag{13}$$

which is what we set out to prove.

REFERENCES:

1. Curnow, R.N. and Dunnett, C.W. The Numerical Evaluation of certain multivariate Normal Integrals. Annals Math. Stat. 33 (1962), 571-579.

2. Feller, W. An Introduction to Probability Theory and its Applications John Wiley, New York, 1950.

3. Rosenblatt, F. A Model for Experiential Storage in Neural Networks, Proceedings of COINS Symposium (June, 1963), Spartan Books, Washington. (In press.)

A NUMERICAL METHOD FOR ESTIMATING MOLECULAR CONCENTRATIONS AND BINDING CONSTANTS IN A MIXTURE

By R. Marchbanks and F. Rosenblatt

In a number of experiments currently in progress, it is necessary to analyze a centrifugal fraction of brain tissue for the presence of large molecules which bind a known substrate, such as acetylcholine or serotonin. Since the fractions of interest generally contain a mixture of such binding molecules (the identities of which are unknown), it is of interest to determine the number of species of binding molecules which may be present, and the concentration and binding constant of each. The following technique is suggested for making such a determination from measurements of the total substrate bound as a function of the substrate concentration.

Let the mixture to be analyzed contain the proteins (or other large molecules) $P_1, P_2, \ldots, P_m$; let x_i = concentration of P_i , and let k_i = binding constant of P_i for the substrate S. Let [S] = concentration of S, and a_i = amount of S bound by P_i (measured as a concentration).

In an experiment in which the substrate S is added to the unknown protein mixture, it is possible to measure [S] and $\sum_i a_i$ (the total amount of substrate bound). It is desired to determine the values of x_i and k_i for all significant components of the mixture.

From the law of mass action, the following equation is known to apply: *

$$\sum_i a_i = [S] \left(\sum_j k_j x_j - \sum_j k_j a_j \right) \qquad (1)$$

From this we obtain:

$$a_i = \frac{k_i x_i [S]}{k_i [S] + 1} \qquad (2)$$

* See J.T. Edsall and J. Wyman, Biophysical Chemistry Vol. I, Chapt. 11, page 591. Academic Press, New York, 1958.

and

$$\frac{\sum_i a_i}{[S]} = \sum_{i=1}^{m} \frac{x_i}{[S] + \frac{1}{k_i}} \tag{3}$$

Now let $\sum a_i/[S] = F[S]$, and $1/k_i = c_i$. This yields the equation

$$F[S] = \sum_{i=1}^{m} \frac{x_i}{[S]+c_i} \tag{4}$$

where [S] and F[S] can be measured empirically.

Our objective is to estimate values of x_i and c_i (for an assumed value of m) which will give the best fit to the empirical data. A least squares method will be employed.

Let Y_j = measured value of F[S] for $[S] = S_j$. Suppose there are n such measurements (for n values of S_j). Then we must minimize the sum of squares,

$$\Sigma = \sum_j [Y_j - F[S_j]]^2$$

$$= \sum_{j=1}^{n} \left[Y_j - \sum_{i=1}^{m} \frac{x_i}{S_j+c_i} \right]^2 \tag{5}$$

Σ is a function of 2m independent variables $(x_1, x_2,\ldots,x_m;c_1, c_2,\ldots,c_m)$. For convenience of notation in the following equations, we will rename these variables as follows:

$$x_i \equiv z_i$$

$$c_i \equiv z_{m+i} \quad .$$

In terms of these variables, equation (5), for the sum of squares, becomes:

$$\Sigma = \sum_{j=1}^{n} \left[Y_j - \sum_{i=1}^{m} \frac{z_i}{S_j + z_{m+i}} \right]^2 \tag{6}$$

To minimize this expression, we obtain the partial derivatives of Σ with respect to each of the z_i, set all of these derivatives equal to zero, and then solve the resulting set of 2m simultaneous (non-linear) equations for the values of z_i.

The partial derivatives are given by the equations:

$$-1/2 \frac{\partial \Sigma}{\partial z_k} = \sum_{j=1}^{n} [Y_j - \sum_{i=1}^{m} \frac{z_i}{S_j + z_{m+i}}] \frac{1}{S_j + z_{m+k}} = 0 \qquad (k = 1, 2 \ldots, m)$$

$$\frac{1}{2z_k} \frac{\partial \Sigma}{\partial z_k} = \sum_{j=1}^{n} [Y_j - \sum_{i=1}^{m} \frac{z_i}{S_j + z_{m+1}}] \frac{1}{(S_j + z_{m+k})^2} \qquad (k = m+1, m+2, \ldots, 2m) \tag{7}$$

A method of solution by successive approximation is recommended for these equations (c.f. Milne, Numerical Calculus, Princeton University Press, 1949). Specifically, let z be the vector of components z_i, and let $F^k(z)$ be the k^{th} equation of the system (7). Then a procedure for solution is as follows:

(1) Assume approximate initial values for the variables $c_i = z_{m+i}$, and compute corresponding initial values of z_i for $i = 1, 2, \ldots, m$ from the first m equations of (7), which are linear in the unknown variables. Call these initial values $z_i(0)$. (Note that the initial values of c_i should all be different, to prevent a singular coefficient matrix in the simultaneous equations.)

(2) Given the initial values obtained as above, let $z - z(0) = \delta z$. Then, approximately,

$$0 = F^k + \sum_{i=1}^{2m} F_{ki} \, \delta z_i \qquad (k = 1, 2, \ldots, 2m) \tag{8}$$

where $F^k = F^k(z(0))$

$$F_{ki} = \frac{\partial F^k}{\partial z_i} \text{ at } z = z(0) .$$

This gives a system of 2m linear equations which can be solved for δz. An improved approximation is then given by $z(1) = z(0) + \delta z$.

(3) Step (2) is iterated until the desired accuracy is obtained.

The equations for F_{ki} which are required for the solution of (8) are:

$$F_{ki} = \begin{cases} -\sum_{j=1}^{n} \dfrac{1}{(S_j+z_{m+i})(S_j+z_{m+k})} & (k \leq m,\ i \leq m) \\[2ex] \sum_{j=1}^{n} \dfrac{z_{i-m}}{(S_j+z_i)^2 (S_j+z_k)} & (k \leq m,\ i > m,\ i \neq m+k) \\[2ex] \dfrac{1}{(S_j+z_i)^2} \sum_{j=1}^{n} \left[-Y_j + \sum_{q=1}^{m} \dfrac{z_q}{S_j+z_{m+q}} + \dfrac{z_i}{S_j+z_i} \right] & (k \leq m,\ i > m,\ i = m+k) \\[2ex] -\sum_{j=1}^{n} \dfrac{1}{(S_j+z_{m+i})(S_j+z_i)^2} & (k > m,\ i \leq m) \\[2ex] \sum_{j=1}^{n} \dfrac{z_{i-m}}{(S_j+z_i)^2 (S_j+z_k)^2} & (k > m,\ i > m,\ i \neq k) \\[2ex] \dfrac{1}{(S_j+z_i)^3} \sum_{j=1}^{n} \left[-2Y_j + 2\sum_{q=1}^{n} \dfrac{z_q}{S_j+z_{m+q}} + \dfrac{z_{i-m}}{S_j+z_i} \right] & (k > m,\ i > m,\ i = k) \end{cases} \tag{9}$$

Experimental results obtained by means of this technique will be reported in subsequent papers. An appropriate value for m (the number of proteins) can be found by solving the equations for increasingly large m, until components of negligible concentration begin to appear. The number of measurements, n, should always be greater than 2m, to avoid degenerate cases, and the accuracy of the fit can be improved by increasing n.

FURTHER SIMULATION EXPERIMENTS ON SERIES-COUPLED PERCEPTRONS

By C. Kesler and F. Rosenblatt

A number of simulation experiments on diverse topics are summarized in this paper. Preliminary data on some of these experiments have been presented previously, in Ref. 4. The experiments are grouped into four main categories: (1) Experiments on the modification of S-unit to A-unit connections; (2) A comparison of quantized and non-quantized reinforcement procedures; (3) Performance of four-layer similarity-constrained perceptrons; (4) Performance of a simulated cat's visual system in angle discrimination and alphabetic character discrimination.

1. Simulations of Three-Layer Perceptrons with Variable S-A Connections

Four discrimination experiments were run on the IBM 7090 computer to study the learning curves of four three-layer perceptrons which were designed to allow the S-A connection weights to change with time. Each experiment used only one perceptron with a 20 by 20 toroidally connected retina. In each experiment, the stimulus world consisted of the following E and F in an upright position everywhere on the 20 by 20 retina:

```
XXXXXXXXXX          XXXXXXXXXX
XXXXXXXXXX          XXXXXXXXXX
XX                  XX
XX                  XX
XX                  XX
XX                  XX
XXXXXXXX            XXXXXXXX
XXXXXXXX            XXXXXXXX
XX                  XX
XX                  XX
XX                  XX
XX                  XX
XXXXXXXXXX          XX
XXXXXXXXXX          XX
```

Thus there were 800 stimuli in the stimulus world, with the negative class consisting of 400 E's and the positive class consisting of 400 F's. The stimulus sequence was the same in all four experiments. This stimulus sequence consisted of 500 training stimuli, chosen at random from the stimulus world, followed by a test on 50 E's and 50 F's, followed by 500 additional training stimuli, then a repetition of the test, and so on. The set of test stimuli was selected to cover the retina in an approximately uniform fashion.

Three different procedures for modifying the weights of the S-A connections were studied in these experiments. Each of these procedures makes use of the concept of an "elastic perturbation procedure" (originally defined in Ref 3, Section 26.4) whereby a weight adjustment is tried on a tentative basis, and retained only if it leads to an improvement in the response of the system. Specifically, the three procedures are as follows:

Procedure 1: Consider any stimulus in the training sequence, say S_j. If the response to S_j is wrong, a quantized α-system error correction reinforcement is applied to the weights of the A-R connections. If this corrective reinforcement is sufficient to produce a correct response, the S-A weights are not changed, and the next stimulus in the training sequence is presented. If the response to S_j is still wrong after applying the error-correction reinforcement to the A-R weights, then the following procedure is used for modifying the S-A weights:

1. Select a set of n A-units at random, with uniform probability. Let P designate this selected set.
2. For every A-unit in P, negative quantized α-system (or γ-system) reinforcement is applied. *
3. A response improvement test is performed as follows: Measure the new input sum to the R-unit for stimulus S_j. If this sum has changed in an unfavorable direction (disagreeing in sign with the desired response) restore the previous S-A weights; otherwise go to the next stimulus in the training sequence.

In a modified form of this procedure, step 3 (the response improvement

* Negative reinforcement means that the sign is chosen so as to reverse the present activity state.

test) is omitted.

Procedure 2: Identical with Procedure 1, except that in step 2 of the S-A weight modification procedure, non-quantized reinforcement is substituted for quantized reinforcement. This has the effect of reversing the activity of every A-unit in the selected set, which has at least one active connection when S_j is presented. Here again, step 3 (the response improvement test) may be omitted in a modified form of the procedure.

Procedure 3: Identical with Procedure 2, except that the set P is chosen as follows: Consider any A-unit, say a_i. Let w_i denote the present weight of the connection from a_i to the R-unit. The probability that a_i is chosen to be in the set P is then $ce^{-k|w_i|}$, where c and k are parameters. (This tends to increase the stability of A-units which have acquired large weights, at the expense of A-units with small weights.) Thus the three procedures are characterized by the parameters n, c, and k, the choice of an α or γ system, and the inclusion or exclusion of the response improvement test, which makes reinforcement contingent upon its consequences. The four experiments which were performed to evaluate these procedures are as follows:

Experiment 1:

The perceptron employed in the first experiment was a binomially connected network with 50 A-units having 6 connections each, and a threshold of 3. In case of a zero signal to the R-unit, the response was considered wrong. Initial weights of S-A connections were 1, and initial weights of A-R connections were zero. Table 1 indicates the combinations of reinforcement parameters which were tested and the final level of performance in each case.

The results shown in this table seem to confirm the view that very little learning is taking place in this experiment. The mean performance for the 32 sets of parameters tested is only slightly better than 51 percent accuracy. The control case (where no S-A modification was performed) and the two alpha system cases with Procedure 3 (with the response improvement test incorporated in the program) were run out to 10,000 training stimuli, and the learning curves are shown in Figure 1. The parameters c and k were so chosen that the probability of selecting an A-unit for the set P is .001

TABLE 1: PARAMETERS AND TERMINAL PERFORMANCES IN EXPERIMENT 1.

S-A Reinfmt. Procedure	Response Improvement Test	Type of S-A Reinfmt.	m	c	k	Percent correct after 1500 stimuli
None (Control)	-	-	-	-	-	51
1	No	α	4	-	-	46
1	No	γ	4	-	-	49
1	No	α	8	-	-	46
1	No	γ	8	-	-	43
1	No	α	16	-	-	52
1	No	γ	16	-	-	51
1	Yes	α	4	-	-	46
1	Yes	γ	4	-	-	51
1	Yes	α	8	-	-	49
1	Yes	γ	8	-	-	51
1	Yes	α	16	-	-	47
1	Yes	γ	16	-	-	44
2	No	α	1	-	-	53
2	No	γ	1	-	-	44
2	No	α	2	-	-	50
2	No	γ	2	-	-	51
2	No	α	4	-	-	53
2	No	γ	4	-	-	48
2	Yes	α	1	-	-	54
2	Yes	γ	1	-	-	47
2	Yes	α	2	-	-	51
2	Yes	γ	2	-	-	48
2	Yes	α	4	-	-	54
2	Yes	γ	4	-	-	48
3	No	α	-	.05	.78	52
3	No	γ	-	.05	.78	49
3	No	α	-	.05	.98	49
3	No	γ	-	.05	.98	54
3	Yes	α	-	.05	.78	60
3	Yes	γ	-	.05	.78	51
3	Yes	α	-	.05	.98	47
3	Yes	γ	-	.05	.98	49

with weights of 4 (k = .98) and 5 (k = .78). Comparison with the control case shows that there is some slight improvement in performance, although it appears unlikely to rise much above 55 percent with these choices of parameters. The large fluctuations in the learning curves are due to the fact that each curve is obtained for a single perceptron, rather than the average of a set of perceptrons. This degree of instability is not unusual.

Experiment 2:

The perceptron was of the same type as in Experiment 1, but the number of connections was increased to twelve per A-unit, with a threshold of 4. The number of A-units was still equal to 50. Tests were limited to Procedure 3 with α-system reinforcement, since this seemed to fare best in the preceding experiment. The response improvement test was employed in each case. Learning curves for the three cases run (including a control case, with no S-A modification) are shown in Figure 2. Increasing the number of connections appears to have helped the performance considerably, although the instability with the small value of k is even more marked than before.

Experiment 3:

The same procedures were tested again with the number of A-units increased to 100, 6 connections to each A-unit, and a threshold of 3. Performance is shown in Fig. 3. Doubling the number of A-units appears to have helped more than doubling the number of connections (in the previous experiment), although the gain in performance relative to the control case is not so clear.

Experiment 4:

A Perceptron of the same type as before, but with 100 A-units and twelve connections to each A-unit (with threshold of 4), was tested for a number of different values of the parameters c and k. Results are shown in Figs. 4, 5 and 6. Note that as c increases and k decreases, the best performances tend to improve, although the fluctuations in the learning curve are more extreme. The great fluctuations may, in part, be due to the increased tendency towards instability of the active set of A-units; however, they are actually no greater than the fluctuations in the control case. It is worth noting that

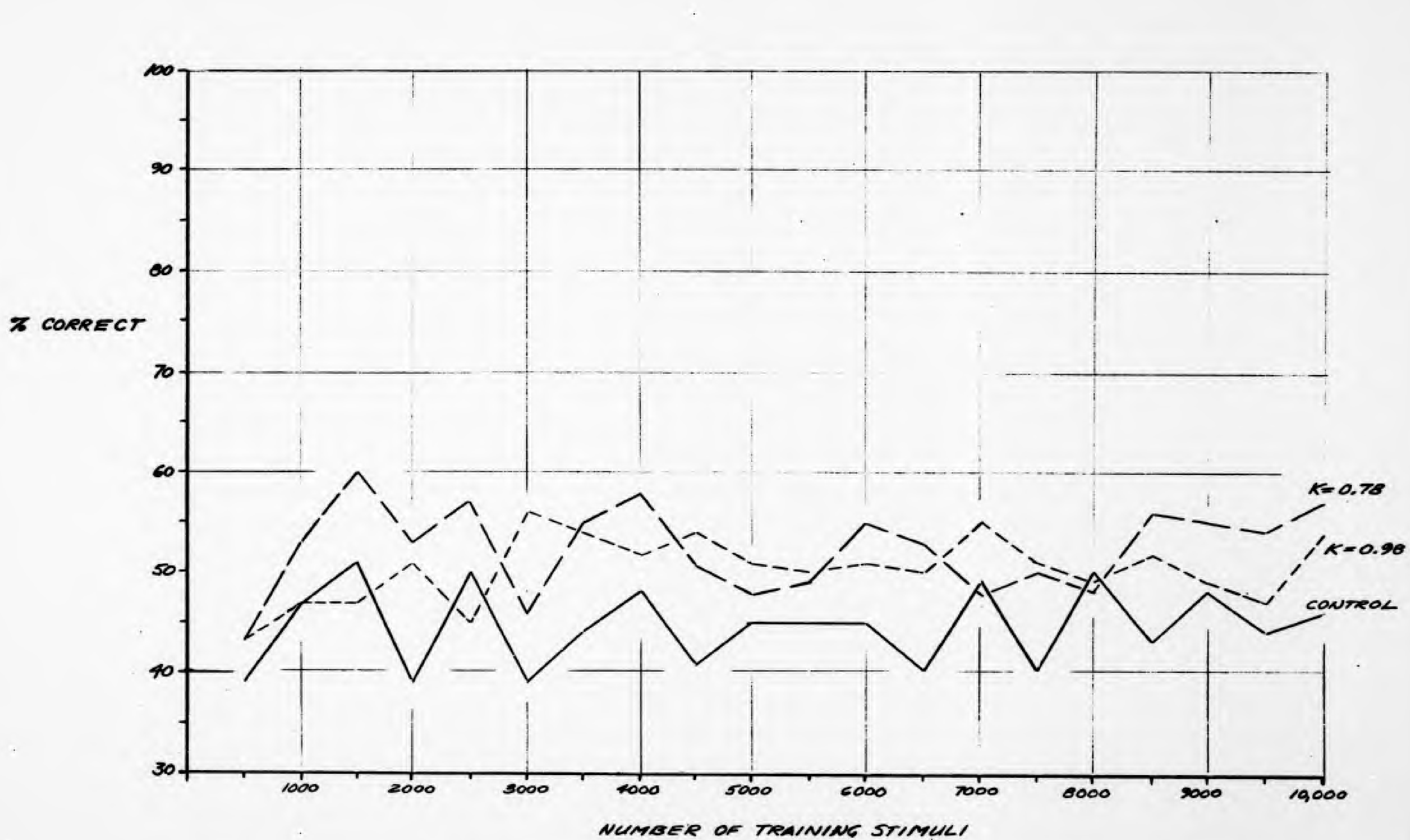


FIG. 1. EXPERIMENT I: E/F DISCRIMINATION. S-A MODIFICATION PROCEDURE 3, WITH α-SYSTEM REINFORCEMENT AND RESPONSE IMPROVEMENT TEST. 50 A-UNITS, 6 CONNECTIONS, θ=3, INITIAL WT. = 1, C = 0.05

FIG. 2. EXPERIMENT 2: E/F DISCRIMINATION. SAME PROCEDURE AS IN FIG. 1, WITH 50 A-UNITS, 12 CONNECTIONS, $\theta = 4$, INITIAL WT = 1, C = 0.05.

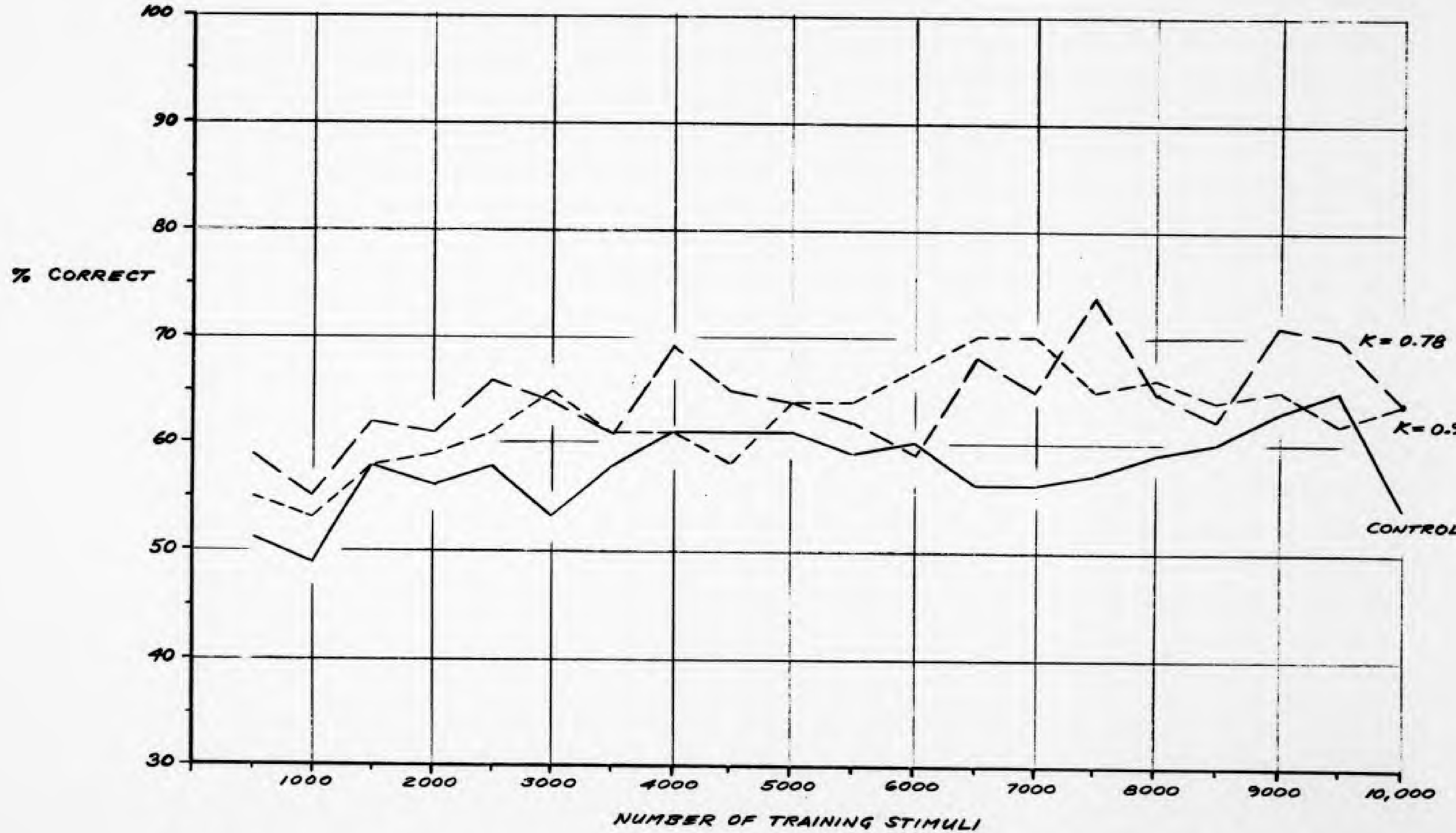


FIG. 3. EXPERIMENT 3: SAME PROCEDURES AS FIGS. 1 AND 2, WITH 100 A-UNITS, 6 CONNECTIONS, $\theta = 3$, INITIAL WT. = 1, C = 0.05.

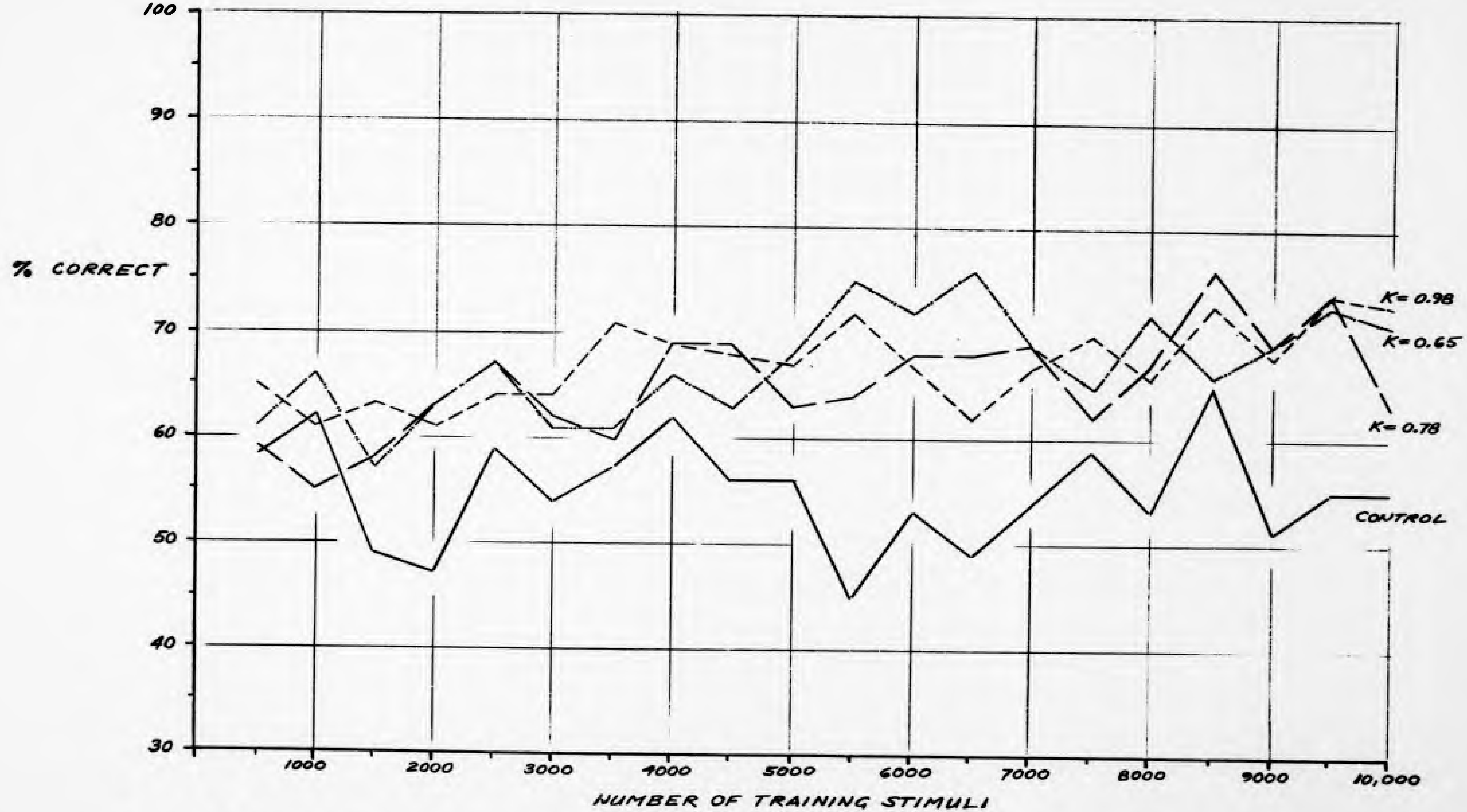


FIG. 4. EXPERIMENT 4: SAME AS FIGS. 1, 2, AND 3, BUT WITH 100 A-UNITS, 12 CONNECTIONS, $\theta = 4$, INITIAL WT. = 1, C = 0.05.

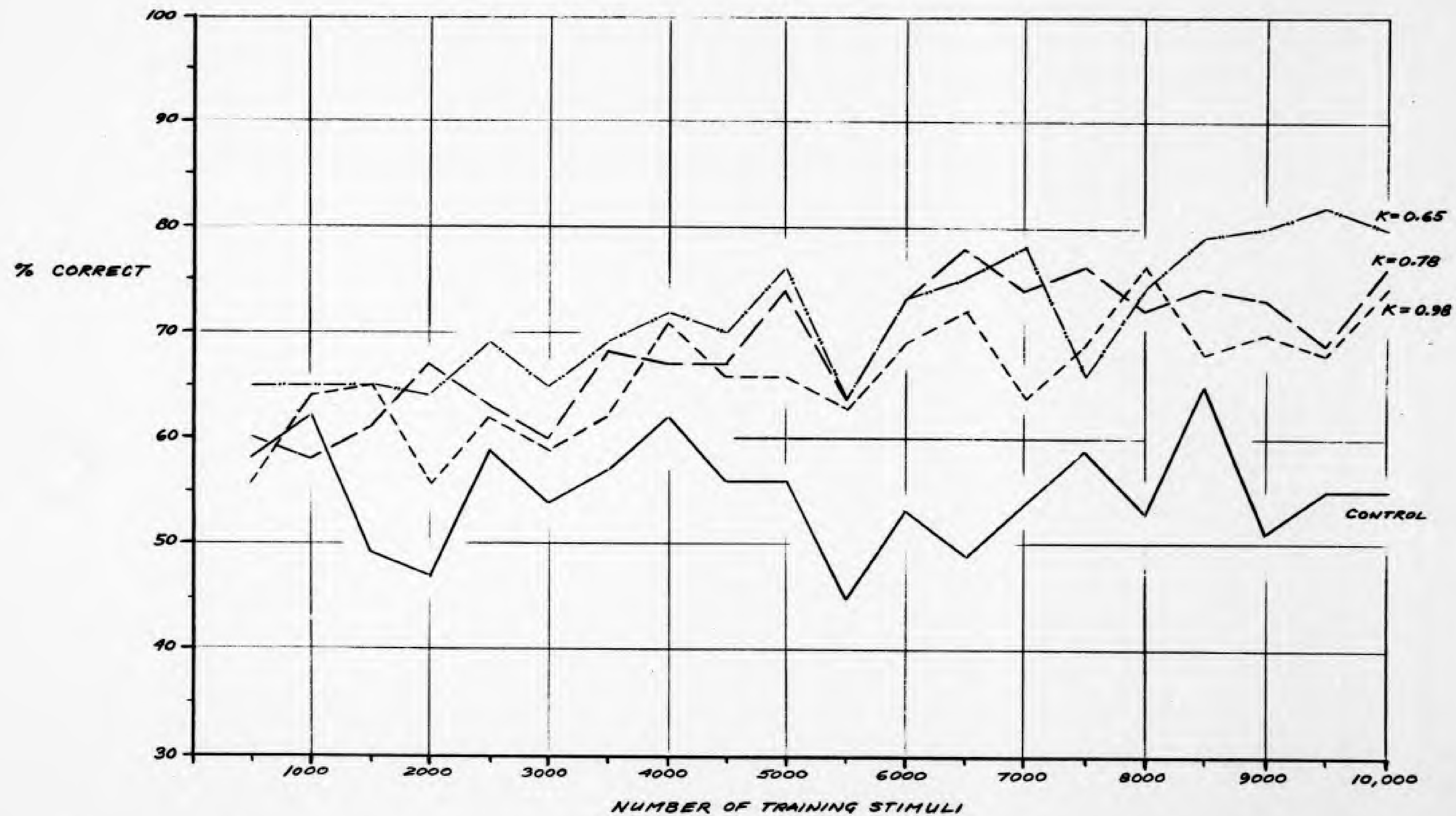


FIG. 5. EXPERIMENT 4: SAME CONDITIONS AS FIG. 4, WITH C INCREASED TO 0.10

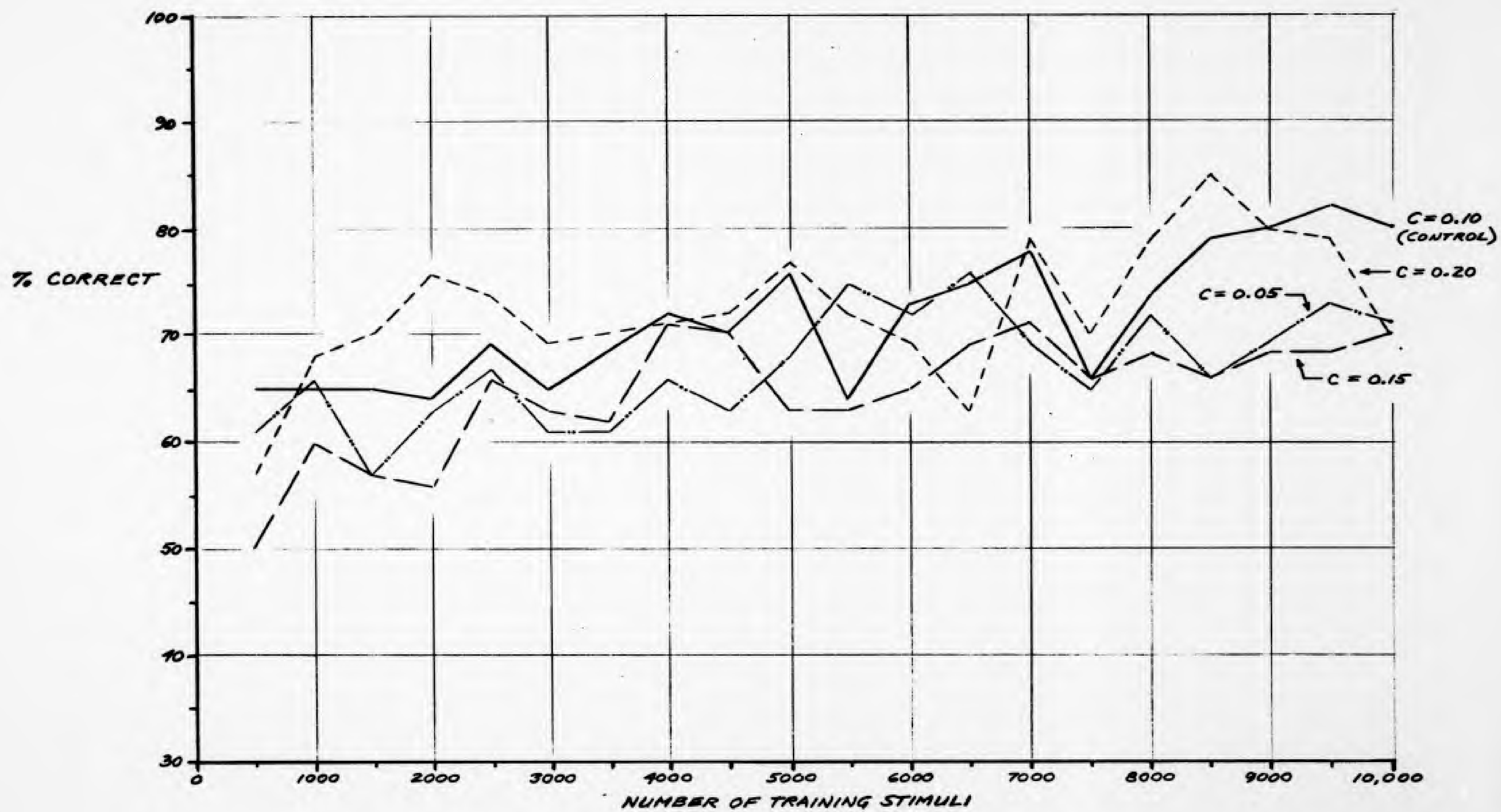


FIG. 6. EXPERIMENT 4: SAME CONDITIONS AS FIGS. 4 AND 5, SHOWING PERFORMANCE FOR 4 VALUES OF C, WITH K = 0.65

the best performance obtained (85 percent correct, with c = .20 and k = .65) is comparable to the performance of a four-layer similarity-constrained perceptron with 1500 $A^{(1)}$ units and 300 $A^{(2)}$ units, as shown in Fig. 11. This shows that the modifications of the S-A network are, indeed, producing a much more efficient set of A-units than were initially available.

It is also of interest to compare these results with the earlier experiments on S-A modification, illustrated in Figs. 38, 39, 69, and 70 of Ref. 3. The system employed here seems to be considerably more effective than any of those previously tried, although a direct comparison is hampered by the greater difficulty of the discrimination problem in the present case. As the discrimination problem becomes more difficult, it seems likely that the improvements due to S-A modification will be greater, since perceptrons which initially have no capability for finding a correct solution may acquire such a capability as a result of the modification process. This conclusion is borne out by the results of an experiment in which a perceptron which was known not to have a solution was compared with four perceptrons for which a solution was known to exist. The problem called for the discrimination of horizontal from vertical bars, and Procedure No. 2 was employed for S-A modification. Each of the five perceptrons tested had 20 A-units with four connections each. The S-A modification procedure was successful in obtaining a solution in the perceptron for which this was initially impossible.

2. Comparison of Quantized and Non-Quantized Reinforcement Procedures

The basic error-correction theorem for simple perceptrons (Ref. 3, Chapt. 5) guarantees that if a solution exists to a classification problem, then either a quantized or a non-quantized reinforcement procedure can be guaranteed to solve the problem. It was originally believed that a non-quantized procedure, which always supplies a sufficient amount of reinforcement to make the response correct, would learn faster than a quantized procedure, which gives a fixed increment to the connection weights regardless of the magnitude of the error. The following experiments were performed in order to compare performance with these two systems.

Five simple perceptrons with 100 A-units, 3 excitatory and 1 inhibitory connection to each, and a threshold of 2, and a 20 by 20 retina were tested on two problems. The first problem, which is quite easy, was the discrimination of horizontal and vertical bars (each bar 4 by 20 units). The second problem, which is quite difficult for simple perceptrons, was the discrimination of a captial E from a capital F. Training sequences consisted of a random sequence of stimuli, which might appear with equal probability anywhere in the retinal field. Performance on a fixed set of test stimuli was measured periodically. In order to measure variability due to training sequences, as well as variability due to perceptrons, each of the five perceptrons was trained with ten different random sequences for each of the two problems. In addition, tests on the E/F problem were carried out with a four-layer similarity constrained perceptron, having 100 A-units in the $A^{(2)}$ layer, and 500 $A^{(1)}$ units (See Ref. 3, Chapt. 15 for a detailed description of this model). This four-layer model was also tested with four different random sequences of stimuli. All runs were made once with quantized α-system error-correction, and once with a non-quantized procedure, in which sufficient reinforcement was given for each stimulus to make the response correct.

The main results of these experiments are shown in Figures 7, 8, and 9. The heavy line in each case represents the mean performance over all perceptrons and training sequences. The light lines represent the envelope of learning curves for the different perceptrons, each averaged over the 10 training sequences. The data suggest that the dispersion in performance on different training sequences (when the performance is averaged over perceptrons) is slightly less than the dispersion due to variations in the perceptrons themselves.

The principal conclusion is that there is no important advantage for the non-quantized system in these problems; in fact, on the more difficult problem (E vs. F) the quantized system seems to have a slight advantage. This advantage seems to be emphasized in the case of the four-layer similarity-constrained model, where all five perceptrons performed better with the quantized than the non-quantized procedure. This phenomenon does not have a rigorous theoretical explanation at this time, although it may be due to the fact that the larger

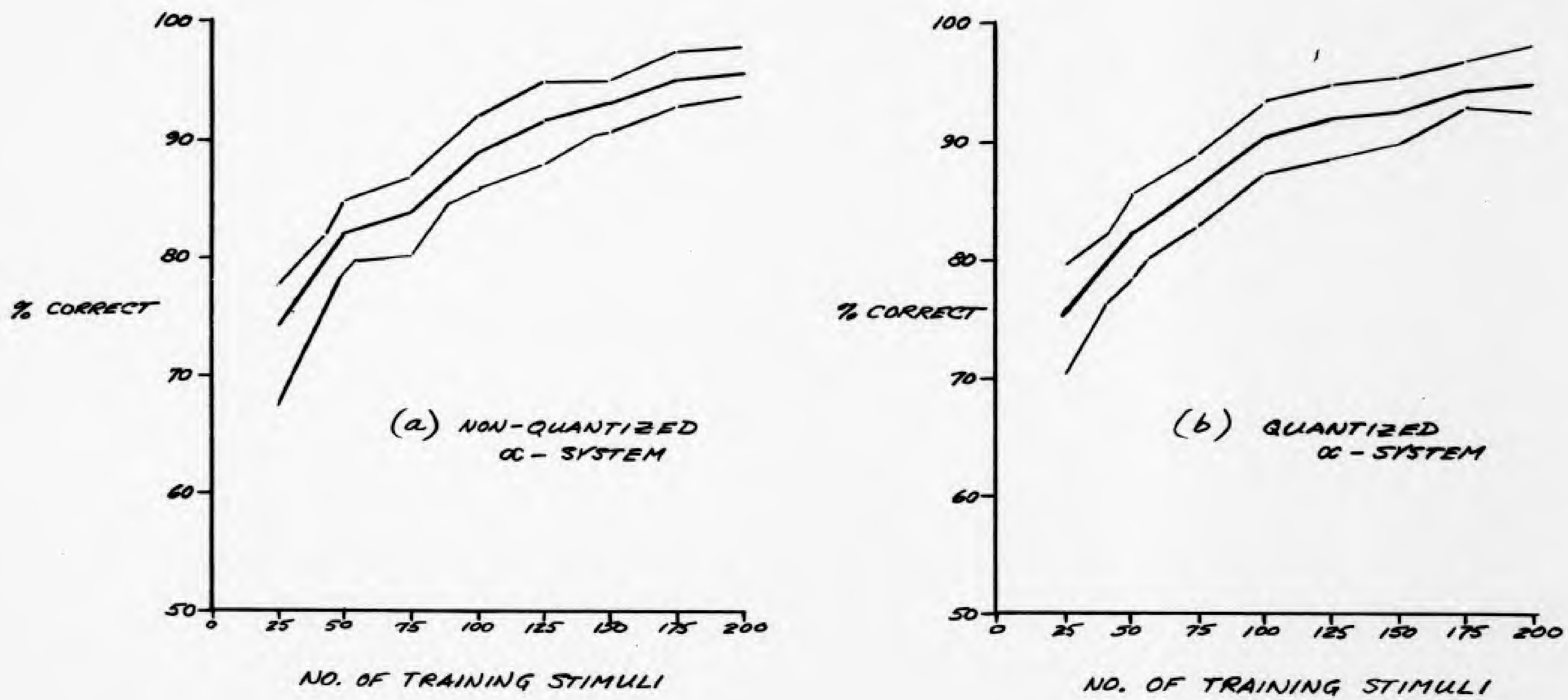


FIG. 7: HORIZONTAL/VERTICAL BAR DISCRIMINATION ON SIMPLE PERCEPTRON; ENVELOPE AND MEAN OF LEARNING CURVES OF 5 PERCEPTRONS AVERAGED OVER 10 SEQUENCES. $N_a = 100$, $x = 3$, $y = 1$, $\theta = 2$

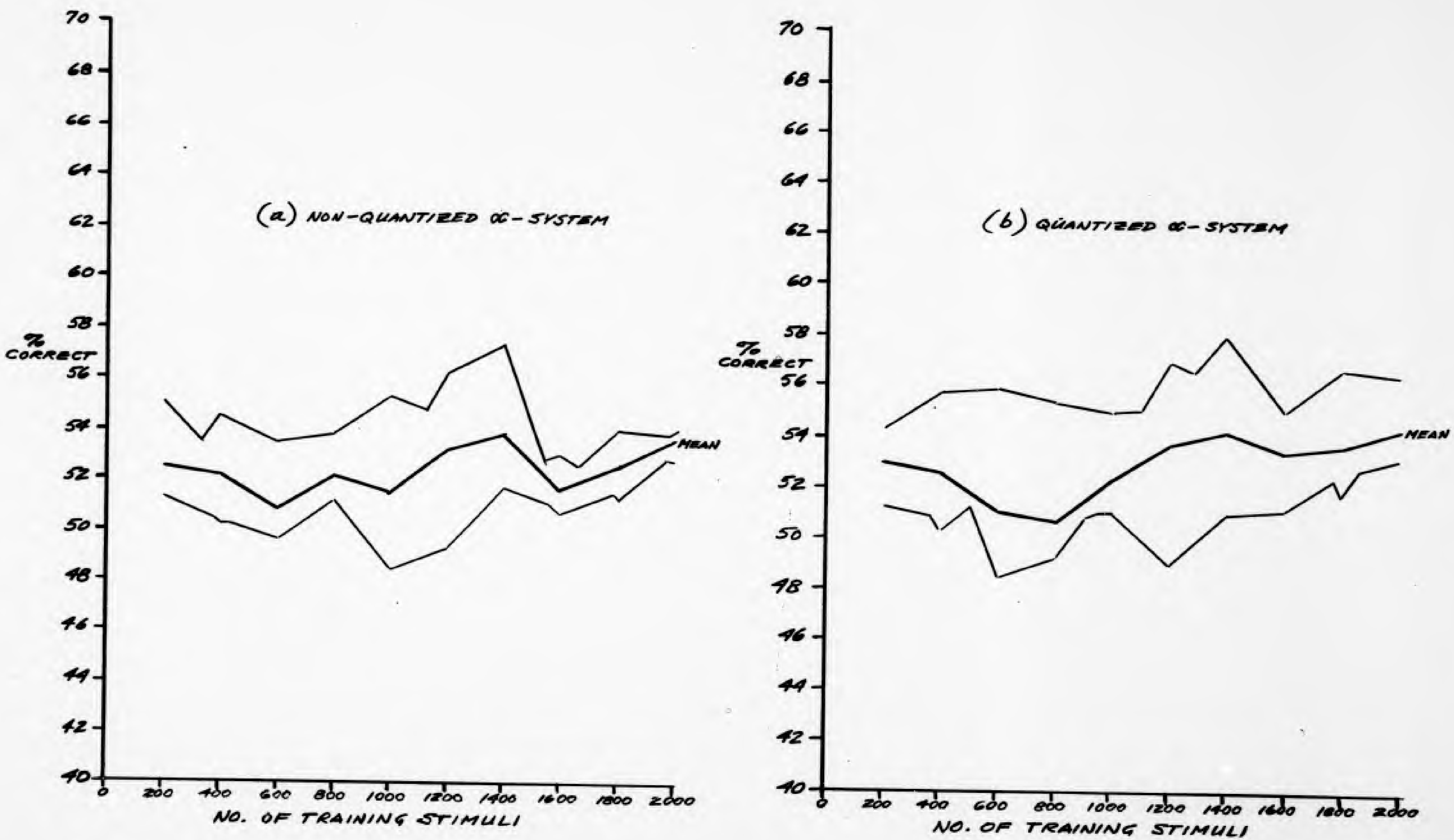


FIG. 8: SIMPLE PERCEPTRON. ENVELOPE AND MEAN OF LEARNING CURVES OF 5 PERCEPTRONS AVERAGED OVER 10 SEQUENCES. E vs. F. $N_a = 100$, $x = 3$, $y = 1$, $\theta = 2$

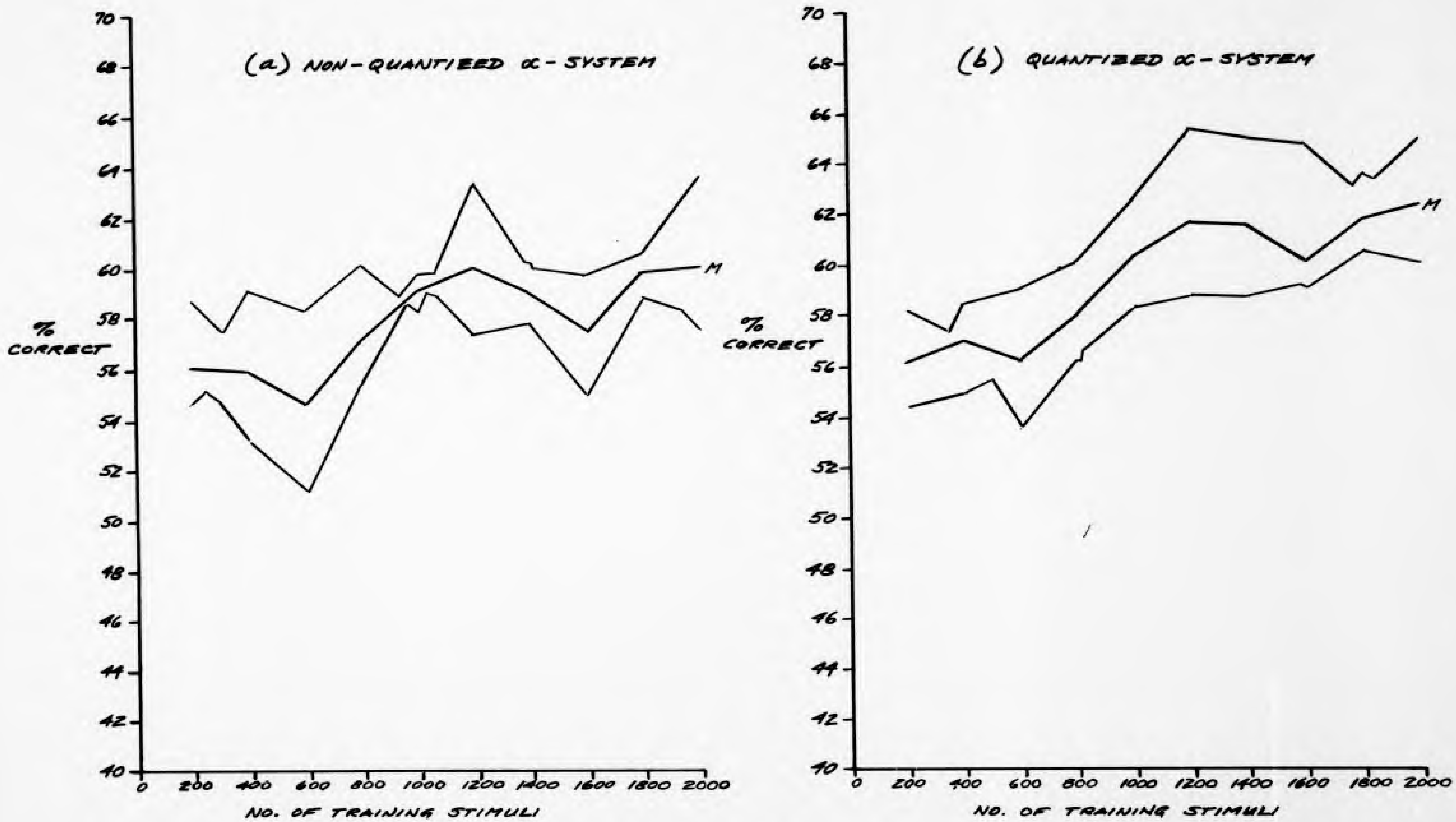


FIG. 9: 4-LAYER PERCEPTRON. ENVELOPE AND MEAN OF LEARNING CURVES OF 5 PERCEPTRONS AVERAGED OVER 10 SEQUENCES. E vs. F $N_a^{(2)} = 100$, $x = 3$, $y = 1$, $\theta = 2$, $m = 5$

perturbations of weights in the non-quantized procedure do more harm by disrupting previous learning than is the case in the quantized procedure.

A second effect worth noting is the greatly increased variability of performance on the more difficult problem, where the detailed organization of the perceptron network engenders striking differences in the learning curves. The particular perceptrons which performed well with the quantized procedure were the same ones which performed well with the non-quantized procedure, on all tests. There is also a slight suggestion of a correlation between performance on the horizontal/vertical bar test and performance on the E/F test, although the number of perceptrons run was too small to measure this correlation with confidence.

3. Performance of Four-Layer Similarity-Constrained Perceptrons

One set of learning curves for four-layer similarity-constrained perceptrons has already been presented in the previous section (Fig. 9). Previous results, reported in Ref. 4, have now been extended to obtain additional points in the early phase of learning curves for easy problems, and in an attempt to improve the parameters for the more difficult E/F discrimination problem. The results are shown in Figs. 10 and 11.

Figure 10 shows the mean performance curves of 25 perceptrons on a single random sequence of horizontal and vertical bars. Zero-signals to the R-unit were given half-credit. The number of first layer A-units (m) connected to each $A^{(2)}$ unit is shown for each curve. The curve for m = 1 is identical with the performance of a simple perceptron with N_a equal to the number of $A^{(2)}$ units. The perceptrons with m = 5 and m = 10 were tested after every five stimuli for the first fifty stimuli of the training sequence, in order to determine how fast they were learning. Even so, the mean performance of the largest perceptrons run ($N_a^{(2)}$ = 300, m = 5) was up to .9855 after the first five training stimuli, and 20 of the 25 perceptrons in the sample had already reached 100 percent accuracy.

Figure 11 shows the results of the E/F experiment, with the best set of parameters tested. The perceptrons were the same as those used for Figs. 4(b)

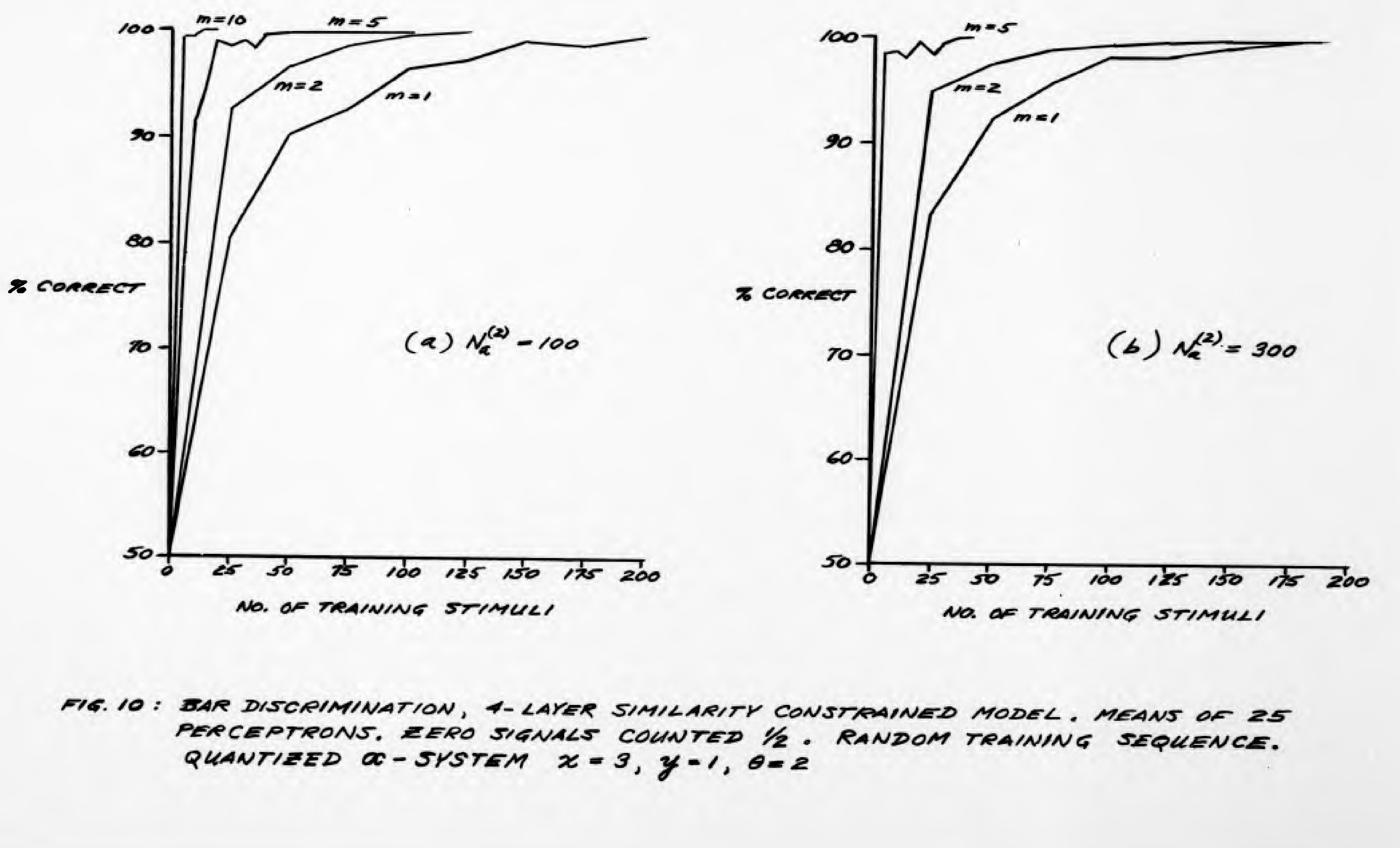


FIG. 10 : BAR DISCRIMINATION, 4-LAYER SIMILARITY CONSTRAINED MODEL. MEANS OF 25 PERCEPTRONS. ZERO SIGNALS COUNTED ½. RANDOM TRAINING SEQUENCE. QUANTIZED α-SYSTEM $x = 3$, $y = 1$, $\theta = 2$

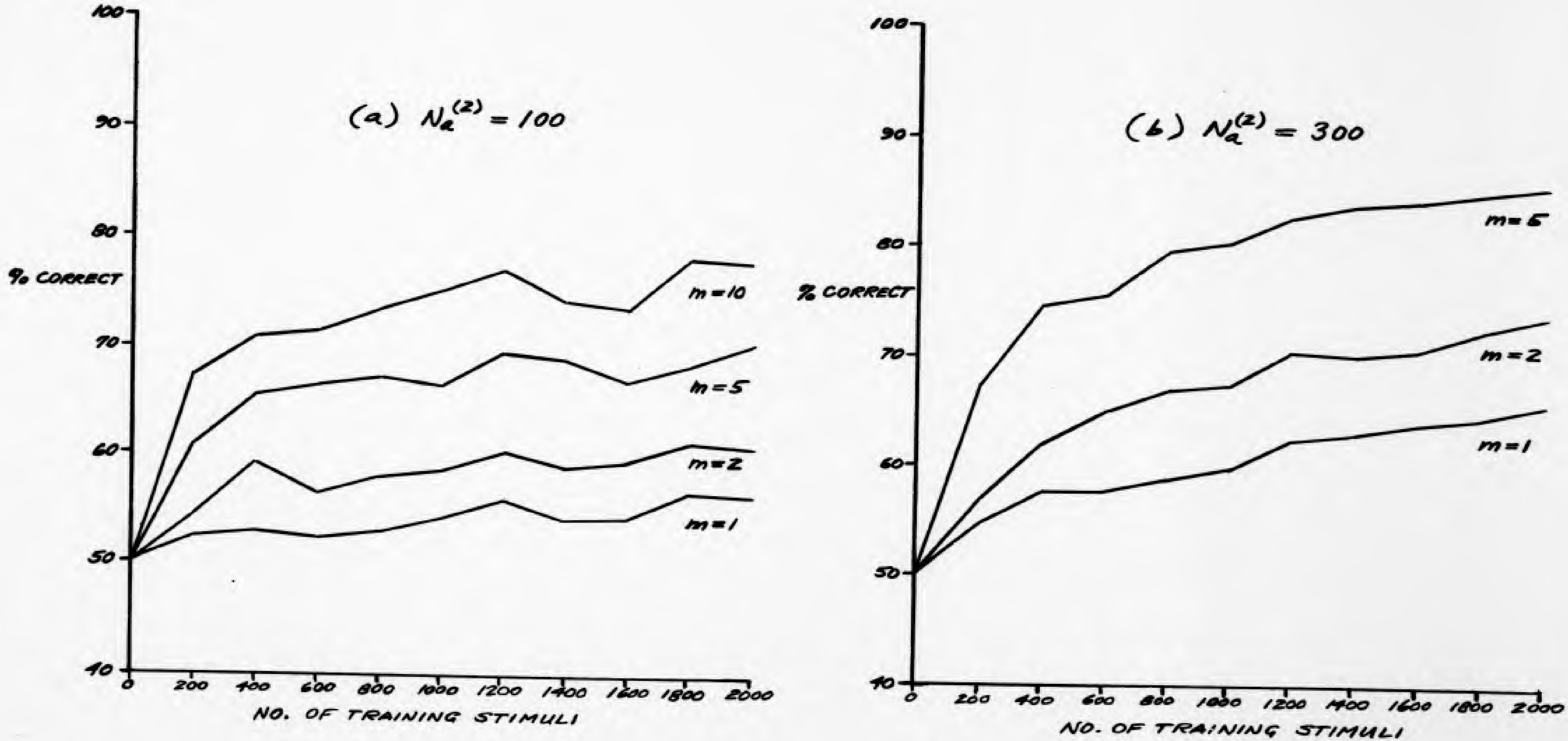


FIG. 11: E/F DISCRIMINATION, 4-LAYER SIMILARITY-CONSTRAINED MODELS
MEANS OF 25 PERCEPTRONS WITH SINGLE RANDOM TRAINING SEQUENCE.
TEST ON 100 STIMULI, ZERO SIGNALS COUNTED 1/2. $x = 5$ $y = 5$ $\theta = 3$

and 5(b) in Ref. 4, except that the threshold was raised from 2 to 3. This higher threshold has improved the performance appreciably, since it tends to increase the weight of A-units whose connection patterns form specific templates for discriminating E's from F's. One perceptron in the sample achieved a score of 95 percent after 1800 stimuli, which is the best performance obtained to date on this particular problem.

4. Performance of "Cat" Model on Line and Letter Discriminations

The five-layer "cat" model which is currently being tested was first described in Ref. 4. It makes use of recent data on the organization of receptive fields of single neurons in the visual cortex of the cat, as described by Hubel and Wiesel (Ref. 2). The retina consists of a 63 by 63 field. The first association layer contains 18,000 local property detectors, with receptive fields organized to act as line detectors and edge detectors, as described by Hubel and Wiesel. The second A-layer contains 2000 units, each of which recieves connections from nine similar line detectors or edge detectors, so that it responds to any one of a set of parallel lines and edges (analogous to Hubel's and Wiesel's "complex units" in the cat cortex). These units go to a randomly connected $A^{(3)}$ layer (with 1000 units), which is finally connected to the R-unit by means of adaptive connections. This organization is illustrated in Fig. 12. Connections from the $A^{(1)}$ layer directly to the $A^{(3)}$ layer are permitted in the model, but were not employed in the experiments reported here.

The first set of experiments was concerned with the discrimination of straight lines at different angles. Performance of the cat model is compared with performance of a simple perceptron with a 63 by 63 retina, on the same stimuli, in Figure 13. Figures 13 (a), (b), and (c) show tasks of increasing difficulty, as the angle between lines in the positive class (vertical) and lines in the negative class is reduced progressively from 90° to 10°. Note that for the simple perceptron (which must have retinal connections coinciding precisely with the stimulus line) these problems are all equally difficult. For the cat, the first problem seems to be somewhat easier than the other two, although the increase in difficulty is not marked. (These

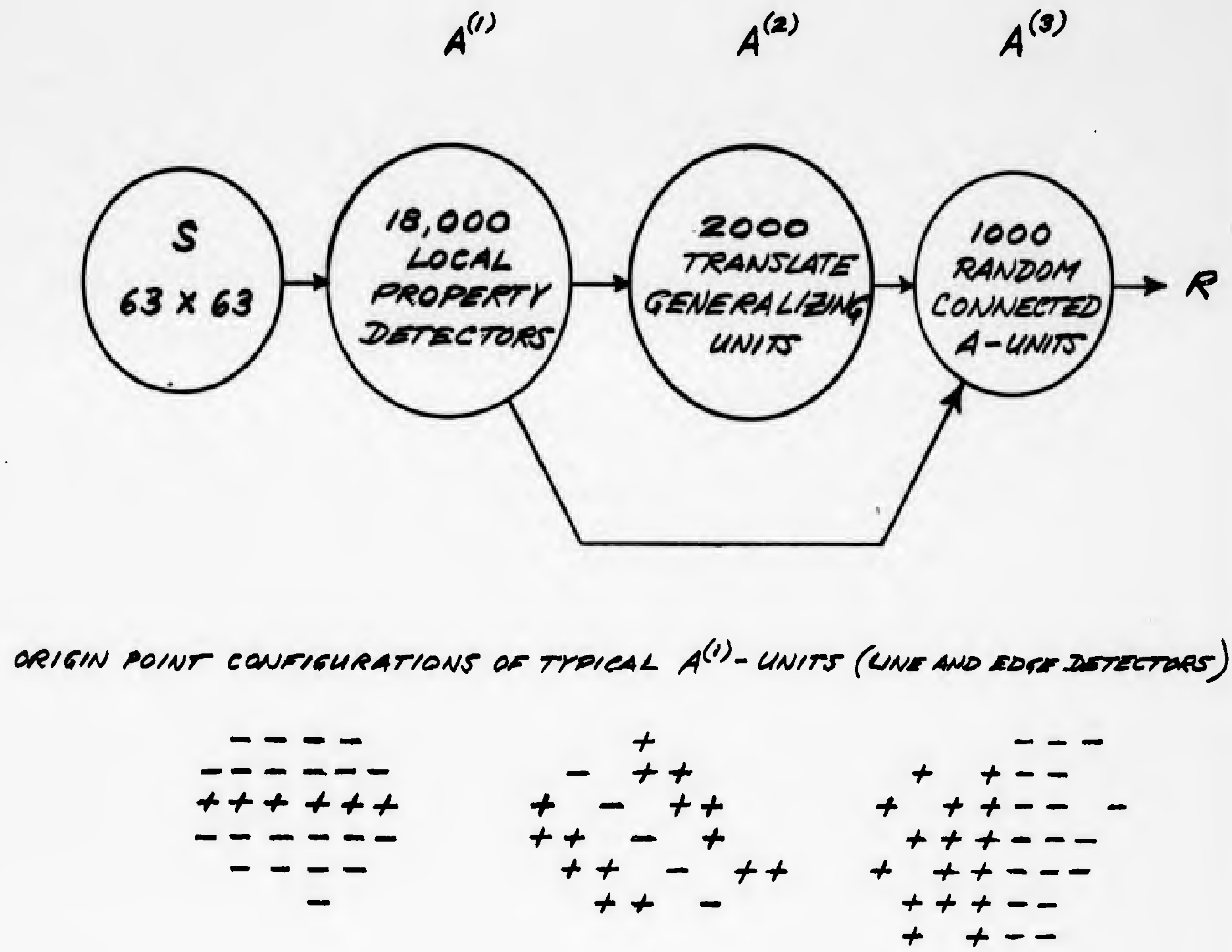
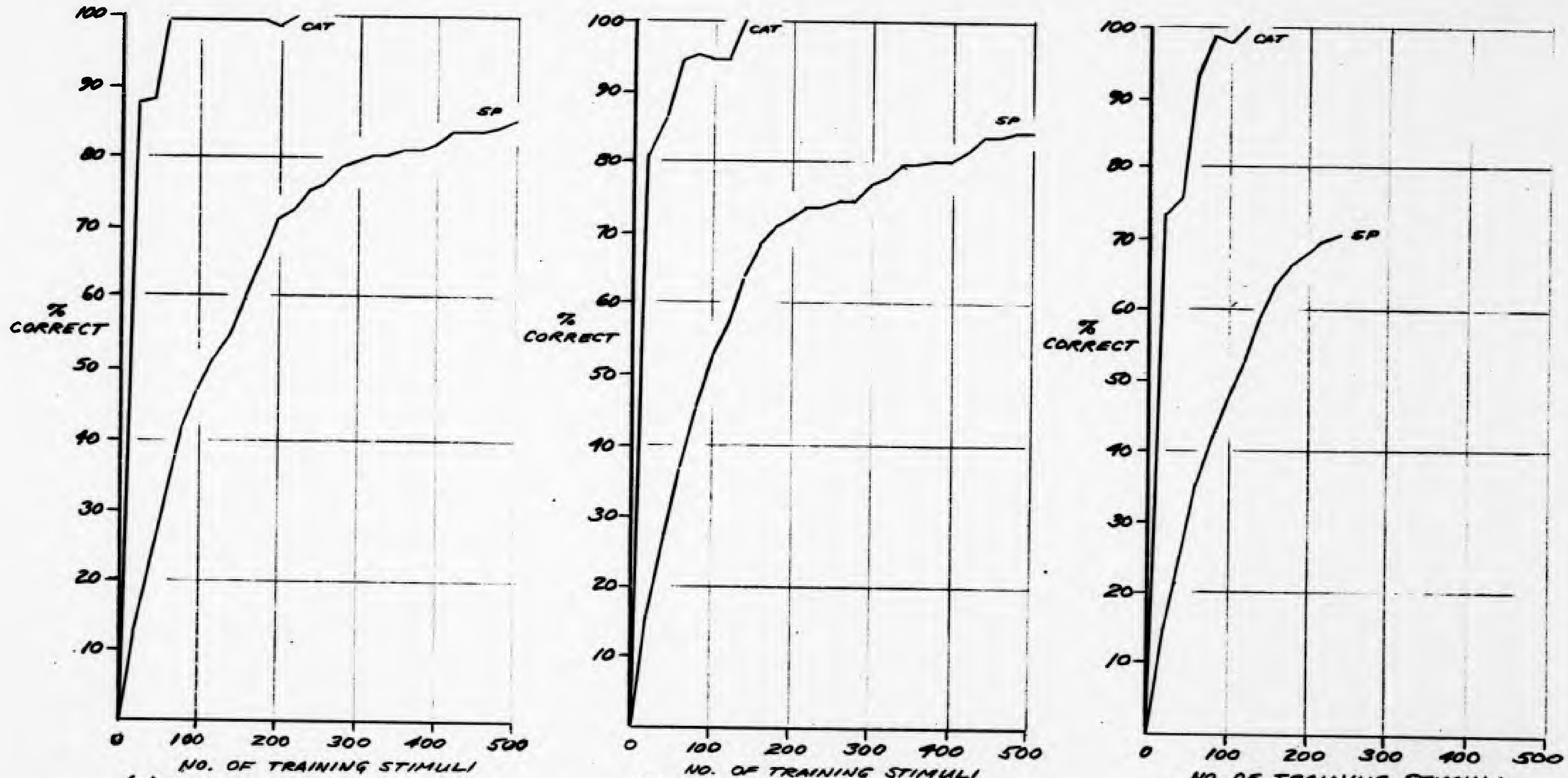


FIG. 12: PERCEPTRON EMPLOYING SIMULATED CAT CORTEX FOR SENSORY ANALYSIS.

learning curves start from 0 instead of .5, since zero signals to the R-unit were counted "wrong" in this program.)

The second set of experiments is shown in Fig. 14, and deals with discrimination of lines of different widths. All lines were vertical, and were either one retinal unit, two units, or three units in thickness. The markedly greater facility with which the cat was able to discriminate lines of thickness 1 from lines of thickness 2 (Fig. 14 (a)) as opposed to the discrimination of



(a) HORIZONTAL/VERTICAL LINE DISCRIMINATION

(b) VERTICAL/20° LINE DISCRIMINATION

(c) VERTICAL/10° LINE DISCRIMINATION

FIG. 13: LINE DISCRIMINATION AT THREE ANGLES
PARAMETERS FOR SIMPLE PERCEPTRON(SP): $N_a = 1000$, $x = y = 5$, $\theta = 2$
PARAMETERS FOR CAT MODEL $A^{(3)}$ UNITS: $x = y = 5$, $\theta = 2$

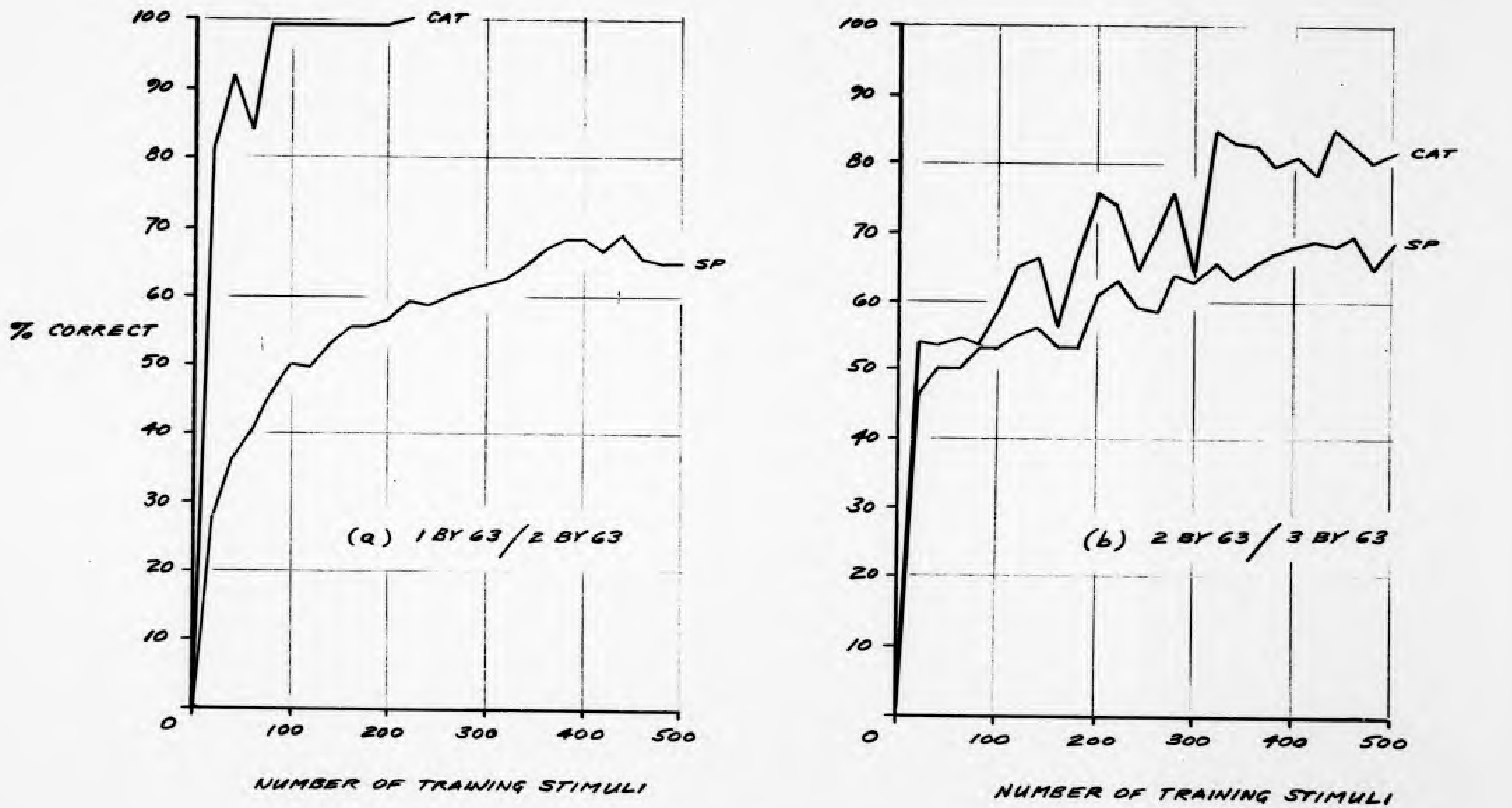


FIGURE 14: DISCRIMINATION OF THICK VERTICAL LINES FROM THIN VERTICAL LINES. SAME PARAMETERS AS FIGURE 13.

lines 2 units thick from lines 3 units thick (Fig. 14 (b)) is probably due in part to the construction of the templates for the $A^{(1)}$ connections. Lines which are only one unit thick would be expected to activate many of the line detectors, but few of the edge detectors, while the thicker lines would activate chiefly edge detectors. The distinction between the lines of medium thickness and lines of maximum thickness would not be so well marked, since both would tend to activate edge detectors, differing only in the magnitudes of the signals transmitted. (The $A^{(1)}$ and $A^{(2)}$ units in these experiments were linear units, transmitting a signal proportional to their input, provided the input signal exceeds the threshold of the unit.)

The last experiment is only partially complete at this time, but the data are of sufficient interest to warrant presentation of some preliminary results. In a recent series of perceptual experiments with pre-literate nursery school children, E.J. Gibson has obtained a confusion matrix for discrimination of alphabetic characters (Ref. 1). It was decided that a comparison of her confusion matrix with one obtained from our cat model, using the identical alphabetic characters (in digitalized form) would be worth while. So far, only three discriminations have been tested, and the learning curves for these are shown in Fig. 15. The relative performance in these three cases agrees perfectly with the results shown in Gibson's confusion matrix. It is particularly striking that the discrimination of M and W, which look virtually identical to the "cat" model, was also the most difficult for Gibson's children. While it is obviously premature to place much weight on these results, they suggest the possibility that the analyzing mechanisms present in the cat model may be quite close, in performance, to those which are operating in human infants, before more sophisticated perceptual tests have been learned. It will be particularly interesting to see whether discriminations of curved letters continue to show the close correlation with human performance which is being found for the straight-line letters.

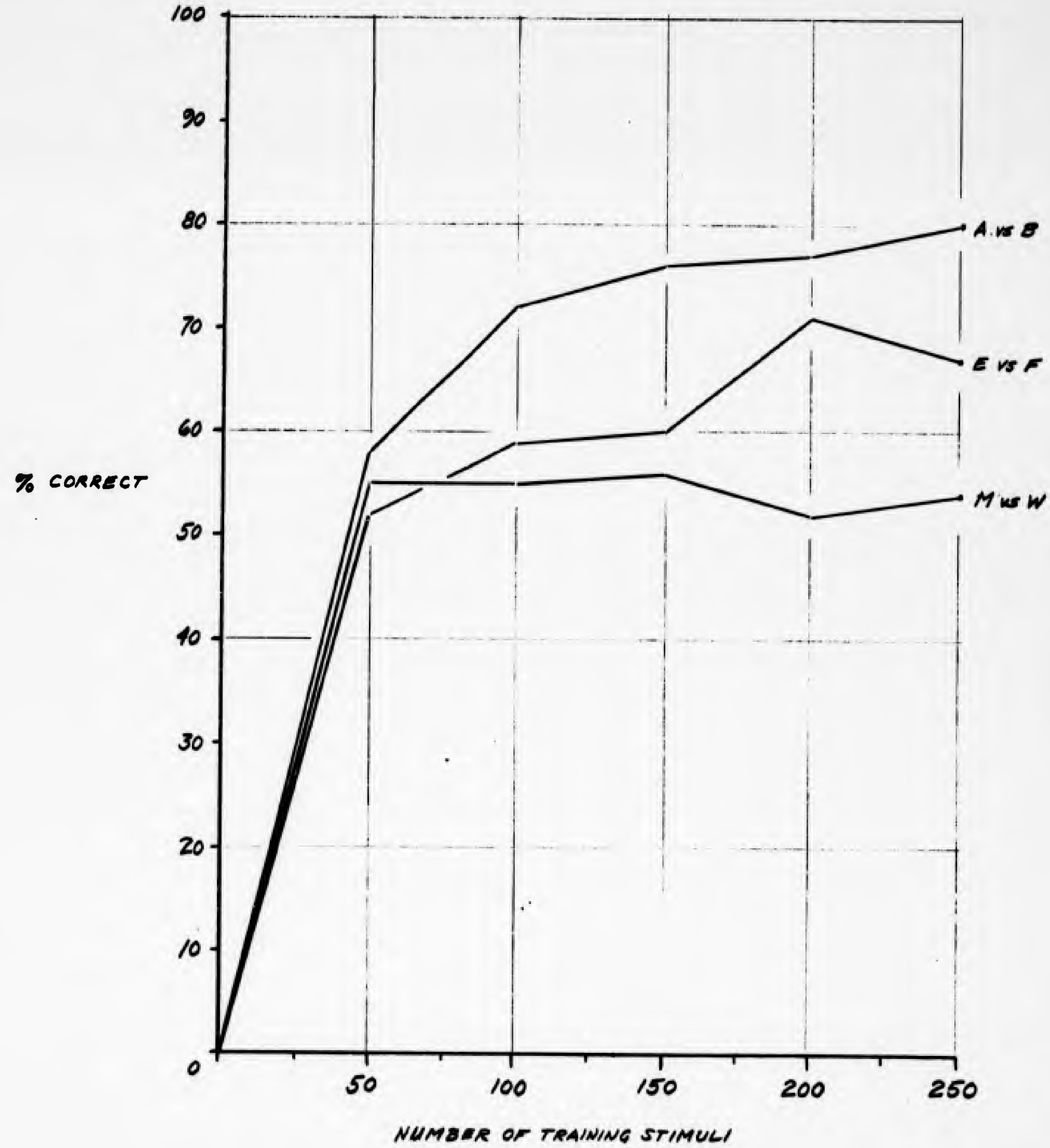


FIGURE 15: PERFORMANCE OF CAT MODEL ON GIBSON'S LETTERS

$A^{(1)}$: 6 TYPES OF MASKS, WITH EITHER $x=13$, $y=26$ OR $x=14$, $y=28$ AND $\theta=6$ OR 8

$A^{(2)}$: $\theta=1$, 9 CONNECTIONS TO EACH UNIT

$A^{(3)}$: $\theta=1$, $x=y=10$

REFERENCES

1. Gibson, E.J. Personal communication.

2. Hubel, D.H., and Wiesel, T.N., Receptive Fields, Binocular Interaction, and Functional Architecture in the Cat's Visual Cortex. J. Physiology, 160, 1962, 106-154.

3. Rosenblatt, F., Principles of Neurodynamics: Perceptrons and the Theory of Brain Mechanisms. Spartan Books, Washington, 1962.

4. Rosenblatt, F., A Comparision of Several Perceptron Models. In Yovits, Jacobi, and Goldstein (Eds.) Self-Organizing Systems, 1962, Spartan Books, Washington, 1962.

SOME STUDIES OF PARAMETER OPTIMIZATION FOR SIMPLE PERCEPTRONS

By F. Rosenblatt and C. Tappert

1. Introduction

Most theoretical and empirical studies of perceptrons (Ref. 2) have dealt with broad classes of networks, such as simple perceptrons, similarity-constrained models, or cross-coupled systems. These studies have either stated theorems applicable to the entire class of networks, or else they have drawn conclusions from conveniently chosen representatives of the class. Within these classes, however, wide variations in performance on any given test problem are likely to occur, depending on choice of parameters, and few attempts have been made to establish the optimum performance for any given class of network. Optimization studies are likely to prove important for at least two purposes:

(1) To assist in engineering design of practical systems.

(2) To assist in the systematic comparison of different types of network organization, to ascertain which gives the best performance.

Unfortunately, the task of finding optimum parameters for a given type of system is fraught with difficulties: The optimum organization depends on the nature of the environment, the task which the system is asked to perform, and the kind of training procedures employed. At the present time, we are unable to do more than scratch the surface of the problem, and the work presented here will be illustrative rather than difinitive. Several early conclusions on optimization of simple perceptrons (as in Section 7.1.3 of Ref. 2) now seem to be premature: it will become apparent from the following discussion that parametric combinations previously believed to be optimum may, in fact, be nothing more than a local "plateau" in the parameter space. It is hoped, however, that the observations reported here (which are confined to "simple perceptrons", consisting of three layers with a single R-unit) will illustrate several basic features of optimal performance, and suggest possible strategies for dealing with the problem in future work.

2. Some Theoretical Notes

Most studies of simple perceptrons published to date have dealt with models in which all A-units are constructed with identical parameters, i.e., the number of excitatory connections, the number of inhibitory connections, and the threshold are identical for every A-unit. An optimum organization may, however, require a mixture of A-units with different parameters, as shown by the following example.

Consider a perceptron with two sensory units, s_1 and s_2. The environment consists of three stimuli:

$$S_1 = (1,0)$$

$$S_2 = (0,1)$$

$$S_3 = (1,1)$$

Stimuli S_1 and S_2 are to be classified positively, and stimulus S_3 is to give a negative response. The perceptron is to have exactly two A-units. There exist eight possible A-unit activity vectors (each indicating the responses of a given A-unit to each of the three stimuli):

(1)	(0,0,0)	(5)	(1,0,0)
(2)	(0,0,1)	(6)	(1,0,1)
(3)	(0,1,0)	(7)	(1,1,1)
(4)	(0,1,1)	(8)	(1,1,0)

Of these eight vectors, only the first seven can be realized with simple A-units. The eighth, which requires that an A-unit respond to s_1 alone, and to s_2 alone, but not to the union of s_1 and s_2, is not possible. If we let x = number of excitatory connections to an A-unit, and y = number of inhibitory connections to an A-unit (with weights of +1 and -1), and θ = A-unit threshold, then the following activity vectors are possible, depending on choice of parameters:

With $\theta \leq (x-y)/2$ all vectors except (1) and (8) are possible.

With $\theta > (x-y)/2$ only vectors 1, 2, 4, and 6 are possible.

If we now consider the A-unit activity matrices (representing the activity of each of the two A-units for each of the three stimuli) we find that the seven admissible vectors yield 49 possible matrices (or 28 which are not equivalent by interchanging A-units). Of these, however, only the two matrices

$$\begin{pmatrix} 1 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix} \text{ and } \begin{pmatrix} 0 & 0 & 1 \\ 1 & 1 & 1 \end{pmatrix}$$

have solutions to the given classification problem. In other words, we must guarantee that one A-unit responds to all three stimuli (vector 7) while the other responds to S_3 alone (vector 2). Now with a random choice of A-units, this can be achieved by choosing $\theta \leq (x-y)/2$, provided we are exceptionally lucky in choosing precisely the right pair of A-units to yield the desired pair of activity vectors, out of the six vectors which might be realized with these parameters. Alternatively, we can pick one A-unit with parameters chosen so as to maximize the probability that it responds to all three stimuli ($y = 0$, and $x >> \theta$), and we can pick the second A-unit with parameters so chosen as to maximize the probability that it responds to S_3 alone (e.g., $y = 0$, and $\theta = x >> 0$). By picking large values of x in each case, it is clear that the probabilities of obtaining the desired response in each case will approach 1. The mixed parameter system is, therefore, clearly optimum for this problem.

In general, where the environment consists of classes which are similar in size of stimuli and distribution of stimuli over the retina, it seems likely that a single uniform choice of parameters may be optimum (although this is an intuitive, rather than a theoretically established conclusion). In heterogeneous environments, where the perceptron may be asked to perform arbitrary classifications, a mixed population of A-units with different parameters is more likely to give it the desired capability. In particular, the Poisson model perceptron (Ref. 2), in which the number of excitatory and inhibitory inputs to each A-unit is a random variable, seems worth considering in this connection.

For the remainder of this paper, however, we shall limit our attention to the performance of simple binomial perceptrons having a uniform population of A-units, and investigate the nature of optimum choices of parameters for such systems. In particular, we shall study the discrimination of horizontal from vertical lines, on a square retina, by such perceptrons.

Although the procedure generally employed to train a simple perceptron is an error-correction procedure, a quantitative analysis of performance as a function of system parameters is not available for this procedure. A suitable analysis has been completed by Joseph for the S-controlled reinforcement procedure, however (Ref. 2), and since optimum performance with the S-controlled procedure is closely related to ease of classification by means of an error correction procedure, the S-controlled system will be considered here. In particular, we shall try to elucidate the nature of the relationship between optimum performance and the parameters of the network, by means of the following example.

The performance of a perceptron, under the S-controlled procedure, is generally measured by the probability of obtaining a correct response to a test stimulus. This probability is a monotone function of the ratio $E^2(u_x)/\sigma^2(u_x)$, where

$E(u_x)$ = expected signal to the R-unit when the test stimulus S_x is shown, and
$\sigma^2(u_x)$ = variance of the R-unit input signal for stimulus S_x.

The equations for these variables are given in Ref. 2, Equations 7.3 and 7.6. Specifically, for the case of a fixed stimulus sequence where each stimulus in the environment appears exactly once, the equations are

$$E(u_x) = N_a \sum_j \rho_j Q_{jx} \qquad (1)$$

$$\sigma^2(u_x) = N_a \sum_j \sum_k \rho_j \rho_k (Q_{jkx} - Q_{jx} Q_{kx}) \qquad (2)$$

where $\rho_j = \begin{cases} +1 \text{ if stimulus } S_j \text{ is in the positive class} \\ -1 \text{ if stimulus } S_j \text{ is in the negative class} \end{cases}$

$Q_{jk\ldots x}$ = probability that an A-unit responds to all of the subscripted stimuli.

N_a = Number of A-units in the perceptron.

For comparing different systems, it is generally convenient to fix N_a at some convenient value. In the present case, performance measures will be computed for $N_a = 1$, so that this variable can be dropped from the above equations. (Changing N_a merely introduces a scale factor in the computed performance measures.)

For ease of analysis, we will consider the case of a square r x r retina, with an environment of r vertical lines (one unit in thickness and spanning the retina) and r horizontal lines, which form the positive and negative class, respectively. We also take the A-unit threshold $\theta > x/2$ (where x = number of excitatory connections to each A-unit). For these conditions, Q_{jk} for two different stimuli of the same class = 0, and Q_{jk} for stimuli of different classes $= Q_j^2$. It can therefore readily be shown from equations (1) and (2) that

$$E(u_x) = Q - rQ^2 \quad (3)$$

$$\sigma^2(u_x) = Q - (r+1)Q^2 + 2rQ^3 - r^2Q^4 \quad (4)$$

and

$$\frac{E^2(u_x)}{\sigma^2(u_x)} = \frac{Q - rQ^2}{1 - Q + rQ^2} \quad (5)$$

where $Q = Q_x$ = probability that an A-unit responds to any one of the line stimuli (see Chapter 6 of Ref. 2, and Ref. 1 for tables of Q functions). To optimize the perceptron, it is necessary to maximize the value of Equation (5).

It can be seen that this equation has the form $x/(1-x)$ which does not have a maximum, and is singluar at x = 1. But it can easily be demonstrated that for any choice of connection parameters, the numerator of (5) must be less than 1, and greater than zero. Subject to this constraint, it is clear that Equation (5) will be maximum when its numerator is maximum. Setting the derivative of the numerator equal to zero we have

$$1 - 2rQ = 0$$

and the optimum value of Q is therefore $1/2r$.

But this solution gives the optimum performance in terms of Q, the probability that an A-unit responds to a stimulus, rather than in terms of the physical parameters x, y, and θ, which characterize the network. In a particular connection scheme, such as the binomial model, it is possible that no values of x, y, and θ exist which yield the optimum value of Q exactly, although one may come very close by letting x, y, and θ range over a sufficiently large domain. As will be seen from the numerical examples in the following section, there may, in fact, be innumerable local "plateaus" in the x, y, θ space, without any true optimum, even though an optimum value of Q (or, more generally, an optimal set of Q functions) exists. Nonetheless, the optimum may be approached quite closely in practise, by considering a large variety of parameter combinations and selecting the best of them. In most practical cases, constraints are set by maximum permissible numbers of connections, and a network may be optimized within the given bounds. This approach is illustrated in the following section.

3. Performance Tables for Discrimination of Horizontal and Vertical Lines

The following numerical study was performed to determine optimum performance in horizontal/vertical line discrimination for perceptrons constrained to a maximum of 20 connections per A-unit, and to determine the effect of deviations from the optimum parameters upon performance. * The expectation and variance of the signal to the R-unit was computed from Equations (1) and (2), for $N_a = 1$, and the resulting ratios $E^2(u_x) / \sigma^2 (u_x)$ are shown in Tables 1 through 15 for a 5x5 retina, Tables 16 through 24 for a 20x20 retina, and Tables 25 through 30 for a 63 x 63 retina. Each table assumes a different value of θ. Values were computed for all combinations of x and y totalling 20 or less, with values of y taken through the optimum performance for each line of the table. (Further increases in y, keeping x and θ fixed, will always lead to a decrease in performance.) Each table begins with $x = 0$, as a minimum value for x. Since it seems clear from the nature of the problem that an optimum set of parameters will have a threshold no less than 2, each set of tables begins with $\theta = 2$, and goes up to the maximum threshold for which ratios greater than zero were computed.

The following conclusions appear from an examination of these tables:

(1) For a 5 x 5 retina, with $x+y \leq 20$, the optimum parameters for a simple binomial perceptron on this problem are $x = 3$, $y = 0$, $\theta = 2$. (This is close to the "optimum" parameters found for the horizontal/vertical bar discrimination problem, where the stimuli have similar proportions, in Table 3 of Ref. 2.) For a 20 x 20 retina, the optimum parameters are $x = 17$, $y = 3$, and $\theta = 3$, although nearly identical performance is obtained with $x = 9$, $y = 11$, and $\theta = 2$, as well as several other combinations of parameters. For the 63 x 63 retina, the optimum

---

* The writers are indebted to W. Blecher and T. Barker for their assistance in computing these tables.

TABLES OF $E^2(u_x)/\sigma^2(u_x)$ FOR HORIZONTAL/VERTICAL LINE DISCRIMINATION*

$N_a = 1$, r = Dimension of Retina

TABLE 1. $r = 5$, $\Theta = 2$.

x	y: 0	1	2	3	4	5	6	7	8	9	10	11	12
2	0331	0276											
3	0525	0514											
4	0328	0427	0486	0505	0494								
5	0225	0304	0382	0446	0488	0505	0500						
6	0185	0233	0292	0355	0415	0463	0494	0507	0502				
7	0171	0197	0235	0283	0337	0391	0440	0477	0501	0509	0502		
8	0166	0179	0202	0234	0276	0323	0372	0419	0459	0489	0506	0509	0500
9	0162	0169	0182	0203	0233	0269	0312	0357	0401	0442	0475	0498	
10	0158	0162	0170	0184	0204	0231	0264	0302	0344	0387	0427		
11	0153	0157	0162	0171	0185	0204	0229	0259	0295	0333			
12	0145	0151	0156	0162	0172	0185	0203	0227	0255				
13	0136	0144	0149	0155	0163	0172	0186	0203					
14	0126	0136	0143	0149	0155	0163	0173						
15	0115	0127	0136	0143	0149	0155							
16	0103	0118	0128	0136	0143								
17	0092	0108	0120	0129									
18	0081	0098	0111										
19	0071	0088											
20	0062												

TABLE 2. $r = 5$, $\Theta = 3$.

x	y: 0	1	2	3	4	5	6	7	8	9
3	0077	0062								
4	0241	0200								
5	0429	0377								
6	0508	0494								
7	0443	0488	0505	0497						
8	0344	0410	0461	0493	0505	0499				
9	0269	0327	0386	0437	0476	0500	0507	0498		
10	0223	0266	0315	0367	0416	0458	0488	0505	0507	0496
11	0196	0225	0262	0306	0353	0399	0441	0475	0497	0508
12	0181	0199	0225	0259	0298	0341	0384	0426	0461	
13	0172	0183	0201	0225	0255	0291	0331	0372		
14	0166	0173	0184	0201	0224	0252	0285			
15	0161	0166	0173	0185	0202	0223				
16	0155	0160	0166	0174	0186					
17	0150	0154	0159	0166						
18	0143	0148	0153							
19	0135	0142								
20	0127									

* Decimal points are understood at the left of each tabulated value.

TABLE 3. r = 5, θ = 4

x	y = 0	1	2	3	4	5	6	7
4	0016	0013						
5	0065	0053						
6	0158	0131						
7	0286	0244						
8	0416	0370						
9	0496	0469						
10	0494	0505	0496					
11	0436	0476	0500	0505	0495			
12	0363	0415	0458	0488	0504	0505	0493	
13	0299	0348	0397	0440	0475	0497	0507	0504
14	0252	0292	0337	0382	0425	0461	0488	
15	0219	0250	0287	0327	0370	0411		
16	0198	0220	0248	0282	0319			
17	0184	0199	0220	0246				
18	0174	0185	0200					
19	0167	0175						
20	0162							

TABLE 4. r = 5, θ = 5

x	y = 0	1	2	3	4
5	0003	0003			
6	0016	0013			
7	0046	0038			
8	0100	0083			
9	0180	0152			
10	0281	0242			
11	0386	0344			
12	0469	0434			
13	0504	0492			
14	0489	0504	0503		
15	0440	0475	0497	0506	0501
16	0380	0424	0461	0488	0503
17	0324	0368	0410	0448	
18	0278	0316	0357		
19	0242	0274			
20	0216				

TABLE 5. r = 5, θ = 6

x	y = 0	1	2
6	0001	0001	
7	0004	0003	
8	0012	0010	
9	0030	0025	
10	0062	0052	
11	0111	0093	
12	0178	0152	
13	0261	0226	
14	0350	0311	
15	0431	0394	
16	0486	0461	
17	0505	0499	
18	0489	0503	0505
19	0449	0478	
20	0397		

TABLE 6. $r = 5, \theta = 7$

x	y = 0	y = 1
7	0000	0000
8	0001	0001
9	0003	0003
10	0009	0007
11	0019	0016
12	0038	0032
13	0068	0057
14	0111	0094
15	0167	0143
16	0235	0205
17	0312	0276
18	0388	0351
19	0451	0420
20	0493	

TABLE 7. $r = 5, \theta = 8$

x	y = 0	y = 1
8	0000	0000
9	0000	0000
10	0001	0001
11	0002	0002
12	0006	0005
13	0012	0010
14	0024	0020
15	0042	0035
16	0068	0058
17	0104	0089
18	0152	0131
19	0209	0182
20	0275	

TABLE 8. $r = 5, \theta = 9.$

x	y = 0	y = 1
9	00000	00000
10	00000	00000
11	00002	00002
12	00006	00005
13	00017	00014
14	00038	00031
15	00078	00065
16	00147	00122
17	00255	00214
18	00418	00353
19	00648	00551
20	00957	

TABLE 9. $r = 5, \theta = 10.$

x	y = 0	y = 1
10	00000	00000
11	00000	00000
12	00000	00000
13	00002	00001
14	00005	00004
15	00011	00009
16	00025	00020
17	00049	00041
18	00091	00076
19	00157	00132
20	00257	

TABLE 10. r = 5, θ = 11.

x	y: 0	1
11	00000	00000
12	00000	00000
13	00000	00000
14	00000	00000
15	00001	00001
16	00003	00003
17	00008	00006
18	00016	00013
19	00031	00026
20	00056	

TABLE 11. r = 5, θ = 12.

x	y: 0	1
12	000000000	000000000
13	000000037	000000030
14	000000246	000000194
15	000001006	000000820
16	000003301	000002690
17	000009157	000007495
18	000022449	000018463
19	000049777	000041120
20	000101686	

TABLE 12. r = 5, θ = 13.

x	0	1
13	000000000	000000000
14	000000007	000000007
15	000000052	000000045
16	000000246	000000194
17	000000857	000000693
18	000002518	000002056
19	000006504	000005335
20	000015154	

TABLE 13. r = 5, θ = 14.

x	y: 0	1
14	000000000	000000000
15	000000000	000000000
16	000000007	000000007
17	000000060	000000045
18	000000216	000000171
19	000000678	000000551
20	000001840	

TABLE 14. r = 5, θ = 15.

x	y: 0	1
15	000000000	000000000
16	000000000	000000000
17	000000000	000000000
18	000000007	000000007
19	000000052	000000045
20	000000179	

TABLE 15. r = 5, θ = 16.

x	y: 0	1
16	000000000	000000000
17	000000000	000000000
18	000000000	000000000
19	000000000	000000000
20	000000007	

TABLE 16. r = 20, θ = 2.

x	y: 0	1	2	3	4	5	6	7	8	9	10	11
2	0024	0023										
3	0062	0060										
4	0098	0095										
5	0120	0118										
6	0128	0128										
7	0127	0129	0130	0131	0131	0131	0131					
8	0122	0124	0127	0129	0130	0132	0133	0133	0134	0134	0134	
9	0114	0118	0121	0124	0126	0128	0130	0132	0133	0134	0135	0136
10	0106	0110	0114	0117	0120	0123	0126	0128	0130	0132	0134	
11	0098	0103	0107	0110	0114	0117	0120	0123	0126	0128		
12	0091	0095	0100	0104	0107	0111	0114	0118	0121			
13	0085	0089	0093	0097	0101	0105	0108	0112				
14	0079	0083	0087	0091	0095	0099	0102					
15	0073	0078	0082	0085	0089	0093						
16	0069	0073	0077	0080	0084							
17	0065	0069	0072	0076								
18	0062	0065	0068									
19	0059	0062										
20	0056											

TABLE 17. r= 20, θ = 3.

x	y: 0	1	2	3
3	00012	00012		
4	00048	00045		
5	00113	00108		
6	00213	00203		
7	00345	00330		
8	00503	00483		
9	00674	00650		
10	00845	00818		
11	01000	00975		
12	01132	01109		
13	01233	01215		
14	01303	01291		
15	01344	01338		
16	01359	01361	01360	
17	01355	01362	01368	01371
18	01335	01348	01359	
19	01303	01321		
20	01263			

TABLE 18. r = 20, θ = 4.

x	y = 0	y = 1
4	00001	00001
5	00003	00003
6	00009	00008
7	00019	00002
8	00037	00035
9	00063	00060
10	00101	00096
11	00150	00143
12	00212	00203
13	00288	00276
14	00376	00361
15	00475	00457
16	00582	00561
17	00693	00670
18	00806	00781
19	00915	00890
20	01018	

TABLE 19. r = 20, θ = 5.

x	y = 0	y = 1
5	00000	00000
6	00000	00000
7	00001	00001
8	00002	00001
9	00003	00003
10	00006	00006
11	00011	00011
12	00018	00017
13	00028	00027
14	00042	00040
15	00061	00058
16	00084	00080
17	00113	00108
18	00149	00143
19	00192	00184
20	00241	

TABLE 20. r = 20, θ = 6.

x	y = 0	y = 1
6	00000	00000
7	00000	00000
8	00000	00000
9	00000	00000
10	00000	00000
11	00001	00001
12	00001	00001
13	00002	00002
14	00003	00003
15	00005	00005
16	00008	00008
17	00012	00011
18	00017	00016
19	00024	00023
20	00033	

TABLE 21. r = 20, θ = 7.

x	y = 0	y = 1
7	000000000	000000000
8	000000000	000000000
9	000000022	000000022
10	000000082	000000075
11	000000209	000000201
12	000000492	000000469
13	000001021	000000976
14	000001959	000001863
15	000003517	000003345
16	000005975	000005700
17	000009723	000009269
18	000015222	000014514
19	000023052	000021987
20	000033922	

TABLE 22. r= 20, θ = 8.

x	y: 0	y: 1
8	000000000	000000000
9	000000000	000000000
10	000000000	000000000
11	000000000	000000000
12	000000015	000000015
13	000000037	000000037
14	000000082	000000082
15	000000179	000000171
16	000000343	000000328
17	000000626	000000596
18	000001080	000001028
19	000001788	000001706
20	000002854	

TABLE 23. r= 20, θ = 9.

x	y: 0	y: 1
9	000000000	000000000
10	000000000	000000000
11	000000000	000000000
12	000000000	000000000
13	000000000	000000000
14	000000000	000000000
15	000000000	000000000
16	000000015	000000015
17	000000030	000000030
18	000000060	000000060
19	000000112	000000104
20	000000194	

TABLE 24 r = 20, θ = 10.

x	y: 0	y: 1
10	000000000	000000000
11	000000000	000000000
12	000000000	000000000
13	000000000	000000000
14	000000000	000000000
15	000000000	000000000
16	000000000	000000000
17	000000000	000000000
18	000000000	000000000
19	000000000	000000000
20	000000007	

TABLE 25. r = 63, θ = 2.

x	y = 0	1	2	3	4	5	6
2	00025	00024					
3	00071	00070					
4	00132	00131					
5	00199	00197					
6	00263	00261					
7	00320	00317					
8	00366	00364					
9	00402	00400					
10	00428	00426					
11	00445	00445					
12	00456	00456					
13	00462	00462	00463	00463	00463	00463	00463
14	00463	00464	00465	00466	00466	00467	00467
15	00461	00463	00464	00465	00466	00468	
16	00457	00459	00461	00462	00464		
17	00451	00453	00456	00458			
18	00444	00447	00449				
19	00436	00439					
20	00427						

TABLE 26. r = 63, θ = 3.

x	y = 0	1
3	00000	00000
4	00002	00002
5	00004	00004
6	00008	00008
7	00013	00013
8	00021	00020
9	00031	00030
10	00043	00042
11	00057	00056
12	00074	00073
13	00093	00091
14	00114	00112
15	00136	00135
16	00160	00158
17	00185	00183
18	00211	00209
19	00237	00234
20	00262	

TABLE 27. r = 63, θ = 4.

x	y = 0	1
4	000000060	000000060
5	000000313	000000305
6	000000924	000000909
7	000002131	000002101
8	000004217	000004150
9	000007495	000007376
10	000012331	000012144
11	000019133	000018835
12	000028312	000027873
13	000040345	000039719
14	000055693	000054844
15	000074863	000073723
16	000098363	000096872
17	000126675	000124767
18	000160307	000157915
19	000199743	000196777
20	000245430	

TABLE 28. $r = 63, \theta = 5$.

x	y: 0	y: 1
5	000000000	000000000
6	000000000	000000000
7	000000015	000000015
8	000000052	000000052
9	000000119	000000112
10	000000231	000000231
11	000000425	000000417
12	000000723	000000715
13	000001162	000001147
14	000001788	000001758
15	000002645	000002608
16	000003800	000003740
17	000005312	000005230
18	000007257	000007145
19	000009716	000009567
20	000012785	

TABLE 29. $r = 63, \theta = 6$.

x	y: 0	y: 1
6	000000000	000000000
7	000000000	000000000
8	000000000	000000000
9	000000000	000000000
10	000000000	000000000
11	000000000	000000000
12	000000007	000000007
13	000000022	000000022
14	000000037	000000037
15	000000067	000000067
16	000000104	000000104
17	000000164	000000164
18	000000246	000000246
19	000000358	000000358
20	000000507	

TABLE 30. $r = 63, \theta = 7$.

x	y: 0	y: 1
7	000000000	000000000
8	000000000	000000000
9	000000000	000000000
10	000000000	000000000
11	000000000	000000000
12	000000000	000000000
13	000000000	000000000
14	000000000	000000000
15	000000000	000000000
16	000000000	000000000
17	000000000	000000000
18	000000000	000000000
19	000000007	000000007
20	000000015	

occurs with x = 15, y = 5 and θ = 2. Note that in each case, the optimum θ is rather small.

(2) There is a considerable region in the phase space which gives performances close to the optimum, within the given constraints. For example, in Table 1, 12 out of 127 computed ratios are within 5 percent of the optimum value, and 21 ratios are within 10 percent of the optimum. In Table 2 (which does not include the optimum for this size retina) 9 out of 91 ratios are within 5 percent, and 19 are within 10 percent of the optimum performance. Even in Table 5 (where the threshold is three times its optimum value) 4 out of 30 computed values are within 5 percent and 6 are within 10 percent of the optimum. For the larger retinas, it is clear that there is a region in the neighborhood of the optimum for each table, where changes in parameter values make little difference in performance, i.e., the optima are not sharply defined, and small variations in parameters are unlikely to have a serious effect on performance. In Table 25, which includes the optimum for the large (63 x 63) retina, 35 out of 57 computed values are within 5 percent of optimum performance, and 42 are within 10 percent.

(3) For a fixed value of θ, the best performances seem to occur consistently in a single "connected" region of the x,y parameter space; there is no case of more than one plateau appearing within a single table. On the other hand, a slight change in θ may lead to a radical change in the location of the optimum. For example, increasing θ from 2 to 3, for the 5 x 5 retina, moves the optimum from x=3, y=0 to x=11, y=9. Increasing the threshold one more step (Table 3) changes it to x = 13, y=6. Similar phenomena are apparent with the 20 x 20 retina, where the optima change from x=9, y=11 to x=17, y=3 as θ goes from 2 to 3.

These numerical results, together with the theoretical conclusions of the preceding section, bear out the concept that a perceptron having a broad spectrum of A-unit types is likely to include sub-populations of A-units which are nearly optimum for any given problem, or set of problems. Thus it is generally unnecessary to undertake an extended search for optimum parameter values, if a perceptron is being designed as a "general purpose system", with a capability for a variety of different tasks. In such a case, it is more important to diversify the A-unit parameters than to optimize them for any particular problem which the perceptron may encounter. Moreover, the parameter space need not be densely covered by the sample of A-units included in the perceptron, in order to ensure that a subpopulation of nearly-optimum units exists for any given problem.

While these conclusions have been reached primarily on the basis of a particularly simple discrimination problem, it seems likely from our experience with simulated perceptrons on alphabetic character recognition and other, more difficult, experiments that they can be supported for most pattern-recognition problems of practical interest.

REFERENCES:

1. Rosenblatt, F. Tables of Q-Functions for Two Perceptron Models. Cornell Aeronautical Laboratory Report No. VG-1196-G-6, Buffalo, May, 1960. (Available from Office of Technical Services, Washington 25, D.C.)

2. Rosenblatt, F. Principles of Neurodynamics: Perceptrons and the Theory of Brain Mechanisms. Spartan Books, Washington, 1962.

SIMULATION EXPERIMENTS WITH A FIVE-LAYER PHONEME ANALYZING PERCEPTRON

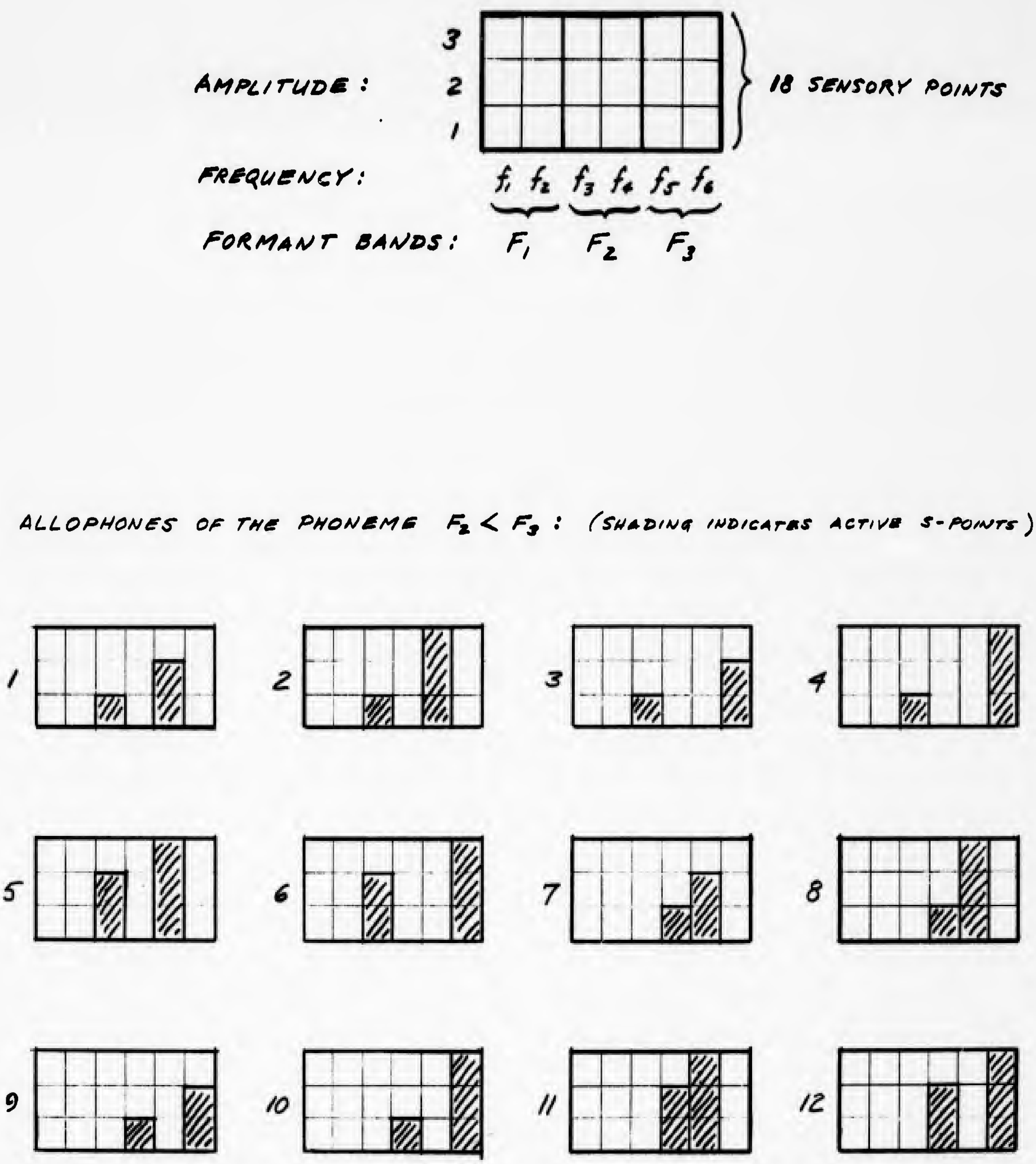
By W. Eisner, F. Rosenblatt, and R. Tuttle

A five-layer perceptron which appeared to have some capability for learning to distinguish the phonemes of a language, as a result of being trained to discriminate words (consisting of several phonemes each), was described by Rosenblatt in Ref. 1, Section 23.2. A number of digital simulation experiments designed to test this concept were initiated in June, 1961, using a Burroughs 220 computer. Four perceptron models were studied, all having five-layer organizations. Only limited success was obtained with these models, and further experiments have been discontinued at this time, although suggestions are made for possible new approaches.

The "environment" of these perceptrons is a greatly simplified model of a human language, consisting of words of two phonemes each. The phonemes are represented by patterns on a set of eighteen sensory units, divided among six frequency bands, with three different amplitude thresholds at each frequency (see Fig. 1). It was postulated that each phoneme would be identified by two formants, one occurring at a greater amplitude than the other. The frequency of the formants, and their relative amplitude, was to provide the information serving to identify the phoneme. The six available frequencies were classified into three "formant bands" of two frequencies each, and a phoneme consisted of some choice of frequency and amplitude in each of two formant bands. For example, a phoneme characterized by a combination of a low-amplitude formant in the middle band, and a high-amplitude formant in the high-frequency band might have 12 variations (or allophones), which are illustrated in Fig. 1. This phoneme is designated by the expression $F_2 < F_3$, indicating the non-zero formants and their relative amplitudes. Note that for the given phoneme, two allophones may activate disjoint sets of S-points. Despite the obvious simplifications, it was felt that this model would provide a degree and type of variability comparable to that of phonemes in a natural language. In all, the model permits 72 different allophone patterns, divided into six phoneme classes ($F_1 > F_2$; $F_1 < F_2$; $F_1 > F_3$; $F_1 < F_3$; $F_2 > F_3$; $F_2 < F_3$).

It was assumed that any pair of non-identical phonemes, occurring in sequence, could form a "word". Thus, with six phonemes, the language

FIG. 1. SENSORY ORGANIZATION AND TYPICAL STIMULUS PATTERNS



consisted of thirty words. Phonemes were only presented to the perceptron in word combinations, and were never individually identified for the perceptron, apart from the words in which they were used. The object of the training program was to teach the perceptron a five-bit response code for each word. The perceptrons were so designed that they could not learn this task unless they developed an internal "phoneme code" in the third layer of the network.

The organization of the five-layer perceptrons employed in these experiments is shown in Fig. 2. The sensory system consists of the eighteen S-units organized as in Fig. 1. The first layer of A-units, $A^{(1)}$, is organized as in a binomial-model perceptron, i.e., each A-unit receives a fixed number of excitatory and inhibitory connections, with fixed weights, originating from randomly chosen S-points. In the simulation experiments, there were 200 $A^{(1)}$ units. Each of these units is connected to each unit in the $A^{(2)}$ layer by means of variable-weighted connections. There are five $A^{(2)}$ units in most of the simulated models. Up to this point, all connections are assumed to have zero transmission time; the connections from $A^{(2)}$ to $A^{(3)}$, however, have delays of either one or two units, so that the $A^{(3)}$ units will tend to respond to specific sequences of duration 2, as described in Chapter 11 of Ref 1. The connections to the $A^{(3)}$ units again are organized according to a binomial scheme (origins chosen at random), and the $A^{(3)}$ outputs are fully coupled to the set of five R-units by means of variable connections. 100 $A^{(3)}$ units were used in the program.

When a word (a sequence of two phonemes) is presented to this perceptron, the sequence of events is as follows: After appearing in the S-units, each phoneme in turn sets up an activity state in the $A^{(1)}$ set, which is characteristic of that phoneme. From here, signals go to the small set of $A^{(2)}$ units, where it is hoped a unique state will occur for each phoneme. Since the number of $A^{(2)}$ units is so small that there are less $A^{(2)}$ states than the number of allophones in the language, the perceptron is obliged to encode a number of different allophones with a single $A^{(2)}$ state. It is hoped that parameters can be found which will encourage the system to generalize the $A^{(2)}$ responses over sets of related allophones (i.e., those representing a common phoneme). The $A^{(3)}$ units have thresholds and numbers of short-delay and long-delay connections chosen so that they will respond only after the word is complete. two successive states having appeared in the $A^{(2)}$ system. Thus the $A^{(3)}$ state represents the occurrence of a particular word. The appropriate response is then associated to the $A^{(3)}$ state by means of an α-system error

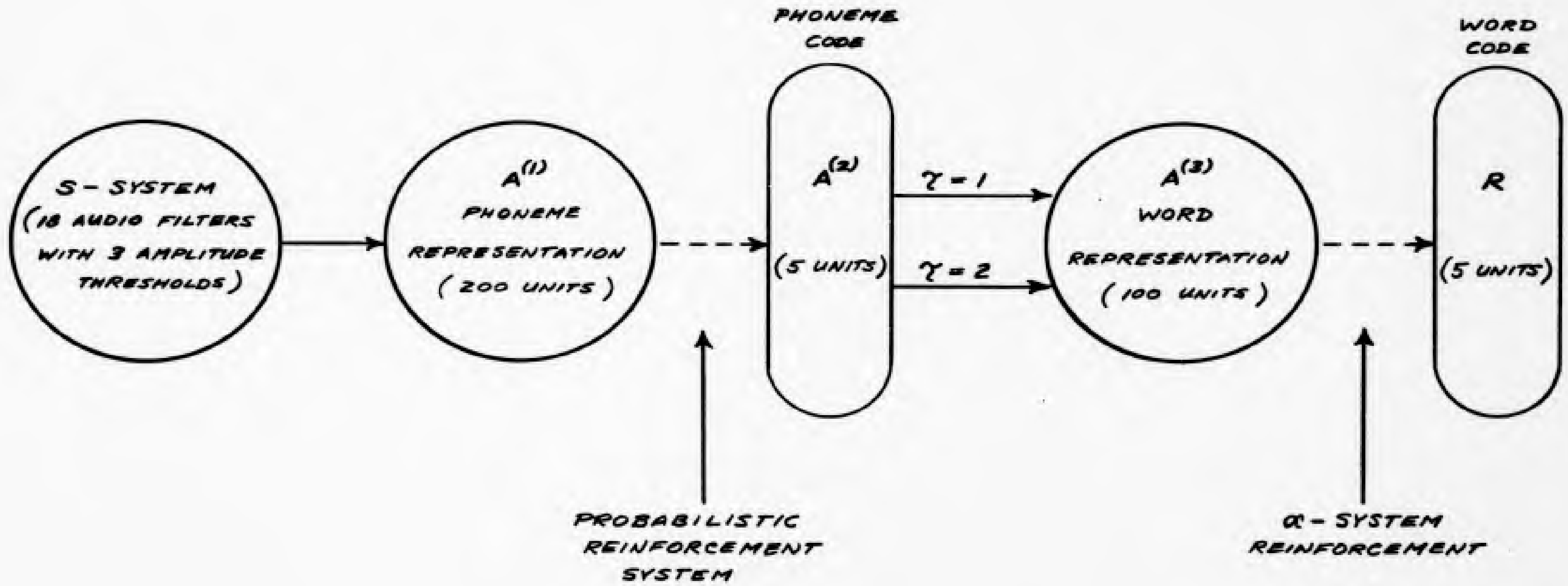


FIGURE 2: ORGANIZATION OF A 5-LAYER PHONEME-ANALYZING PERCEPTRON
(BROKEN LINES INDICATE VARIABLE-VALUED CONNECTIONS)

correction procedure (see Ref. 1, Chapter 5).

Clearly, the correct code can not be learned for every word in the language unless each phoneme can be discriminated from the others at the $A^{(2)}$ level. Thus, unless a phoneme code is found which assigns a unique code to each phoneme, there will be conflicts in the responses of the perceptron, and the system will never converge. The object of reinforcing the $A^{(1)}$ to $A^{(2)}$ connections is to try to encourage the formation of such a code, which not only distinguishes the phonemes from one another, but generalizes the allophones for a given phoneme to a single $A^{(2)}$ code. If such generalization occurs, then new variants of a previously learned word will automatically induce the correct response, even though the particular combination of allophones may never have occurred before. Several different reinforcement schemes for the preterminal connections (from $A^{(1)}$ to $A^{(2)}$) were tried in the following experiments.

Experiment 1:

The reinforcement procedure which was first studied for the $A^{(1)}$ to $A^{(2)}$ connections is described in Ref 1, page 542. Specifically, using the notation defined in the reference,

1. With each connection, c_{ij}, from an $A^{(1)}$ to an $A^{(2)}$ unit, is associated a time-dependent probability, $P_{ij}(t)$, called the instability coefficient of the connection.

2. Reinforcement at the preterminal level is applied only when a terminal reinforcement (of the $A^{(3)}$ to R connections) fails to correct an error. Otherwise, the values of these connections remain unchanged.

3. If preterminal reinforcement is applied at time t, all instability coefficients are changed by the amount $\Delta P_{ij} = a_i^*\varepsilon - \delta P_{ij}(t)$, $[0 < \varepsilon < 1]$. If no reinforcement is applied at time t, $\Delta P_{ij} = -\delta P_{ij}(t)$, i.e., the instability coefficients decay, making the connections more stable when correct responses are obtained.

4. If reinforcement is applied, assume that the current activity states of all $A^{(2)}$ units are "wrong", and apply the correction $\Delta v_{ij} = (-1)^{a_j^*} \cdot (a_i^*\eta)$ with probability $P_{ij}(t)$. (This is equivalent to an α-system error correction applied with an independent probability for each connection.)

Note that in this model, connections coming from $A^{(1)}$ units which are active at the time of an error are most likely to be reinforced, while $A^{(1)}$ units whose activity always leads to correct responses develop highly stable connections to the $A^{(2)}$ layer. It was hoped that this procedure would perturb the weights of the preterminal network until a satisfactory phoneme code was found, at which point all weights would become stable.

A flow diagram of the program used for simulating this perceptron (on a Burroughs 220 computer) is shown in Figure 3. In the course of the experiments, training periods were alternated with tests of performance. In this experiment, only one allophone was used for each phoneme (chosen at random from the twelve possibilities permitted by the model). The following parameters were employed to control the organization of the perceptron:

X_1 = number of excitatory connections from S-system to an $A^{(1)}$ unit.
Y_1 = number of inhibitory connections from S-system to an $A^{(1)}$ unit.
X_{21} = number of excitatory connections from $A^{(2)}$ to an $A^{(3)}$ unit with $\tau = 1$.
Y_{21} = number of inhibitory connections from $A^{(2)}$ to an $A^{(3)}$ unit with $\tau = 1$.
X_{22} = number of excitatory connections from $A^{(2)}$ to an $A^{(3)}$ unit with $\tau = 2$.
Y_{22} = number of inhibitory connections from $A^{(2)}$ to an $A^{(3)}$ unit with $\tau = 2$.
$\theta_1, \theta_2, \theta_3, \theta_R$ = thresholds of $A^{(1)}$, $A^{(2)}$, $A^{(3)}$, and R-units
$\mathcal{E}$ and δ = probability increment and decay rate, defined above.

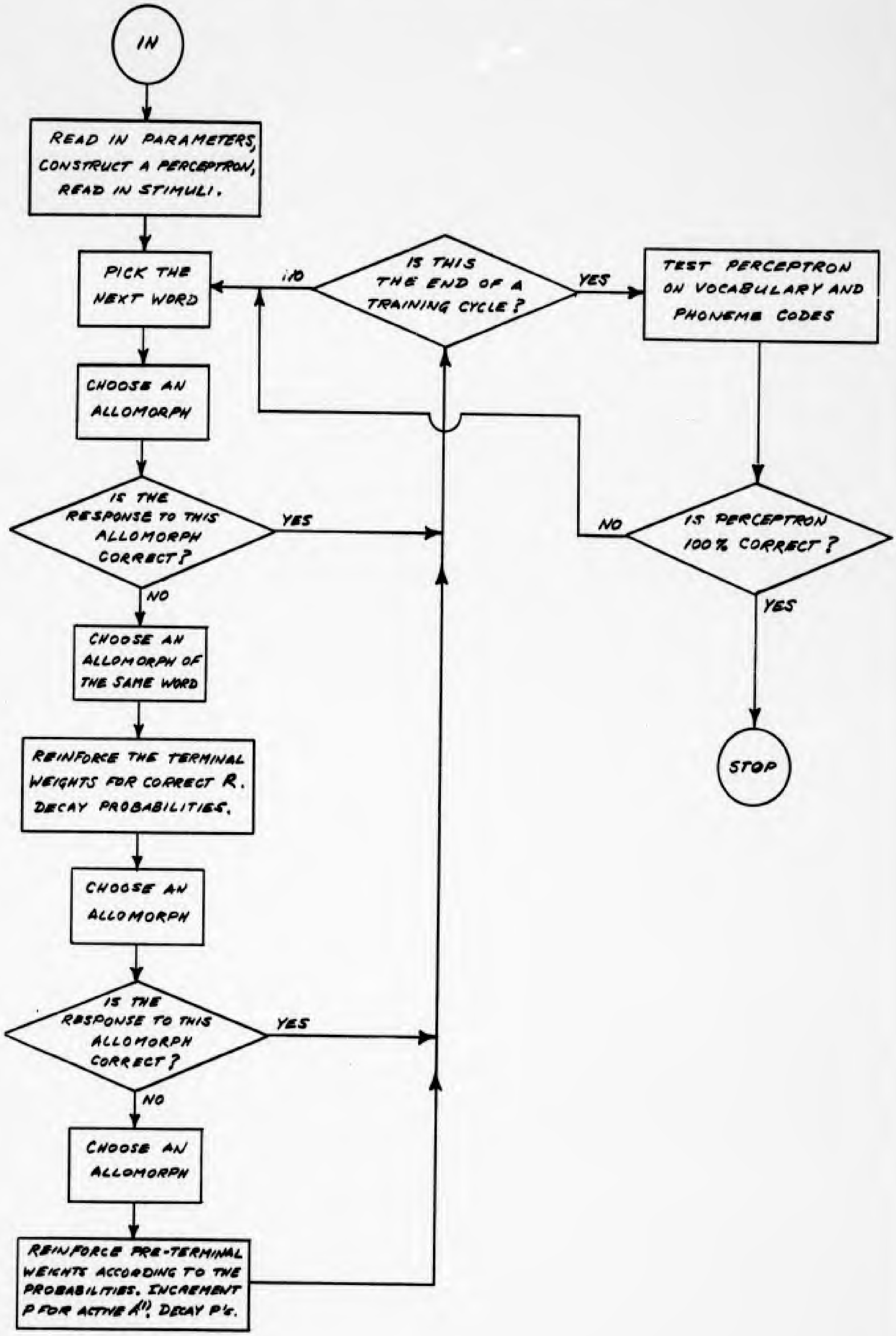
The best choice of parameters found for this experiment was:

$X_1 = 5,\ Y_1 = 2,\ X_{21} = Y_{21} = X_{22} = Y_{22} = 2,\ \theta_1 = \theta_3 = 3,\ \theta_2 = \theta_R = 1,\ \mathcal{E} = .9,\ \delta = .15.$

It was found that for more difficult problems, the probability increments and decay rate had to be made smaller than for this simplified problem.

Four perceptrons were generated using these parameters, and were tested on performance, after being trained on the complete vocabulary (restricted to one allophone per phoneme). Of the four perceptrons, all but one found a "good" code for the six phonemes after having been given each word only once. The one perceptron which failed to achieve a satisfactory code illustrates a basic problem with this type of organization, however: the possibility of stabilization of preterminal weights prior to achieving a complete solution. It became apparent that it was possible for a nearly correct solution to become stable due to the fact that a single correction applied to the terminal weights would always be sufficient to correct any error in the R-units;

FIG. 3. FLOW DIAGRAM OF FIVE LAYER AUDIO PERCEPTRONS



consequently, the $A^{(1)}$ to $A^{(2)}$ connections were no longer reinforced, and a perfect code was never found.

Experiment 2:

The procedure and choice of parameters was the same as for Experiment 1, except that two allophones of each phoneme were permitted, yielding a total of 12 allophones in the stimulus world. The program was constrained, however, to pick only one alomorph of a word (without varying the choice of allophones) in any one cycle through the flow diagram (Figure 3). With a random choice of allophones, this problem was found to be too difficult for the perceptron, and consequently the experimental design was further modified by choosing pairs of allophones which were expected to generalize readily to one another. (For example, one allophone of a phoneme might have an amplitude of three units in the lower frequency of a formant band, and the other would have an amplitude of two units in the same band.) In addition, the parameters were modified to provide for all-excitatory S to $A^{(1)}$ connections, which also tends to increase the generalization tendencies of the system. Finally, $\mathcal{E}$ and δ were modified to slow the rate of change of the probabilities. The final parameters used in Experiment 2 were:

$X_1 = 4$, $Y_1 = 0$, $X_{21} = Y_{21} = X_{22} = Y_{22} = 2$, $\theta_1 = 3$, $\theta_2 = \theta_R = 1$, $\theta_3 = 2$, $\mathcal{E} = .1$, and $\delta = .1$.

This perceptron was still unsuccessful in finding a satisfactory set of phoneme codes. After a few cycles, the new codes appeared to be changing in a "random" fashion, thus causing instability of previously learned responses at the terminal end of the network.

Experiment 3:

The parameters used in Experiment 2 were retained, but the perceptron was further modified in an attempt to increase the generalization tendency of the $A^{(2)}$ layer. Whereas the previous models reinforced connections from $A^{(1)}$ to $A^{(2)}$ units on the assumption that the $A^{(2)}$ response should always be reversed, the revised model assigned a higher probability for decrementing the weights than for incrementing them. A bias was thus introduced, tending to keep more $A^{(2)}$

units off than on, so that a typical phoneme code would consist of only one, or at most two $A^{(2)}$ units being active. Since only a small number of phoneme codes were now available, however, this resulted in a phenomenon of "over-generalization", with too many phonemes being assigned to the same code. The resulting conflicts among the codes for different phonemes led to a continuing oscillation of $A^{(3)}$ weights, and an inability to learn the problem.

<u>Experiment 4</u>:

In the previous models, the program kept track of the probability that an $A^{(1)}$ unit was "on" when errors occurred in the response, and changed the connections <u>from</u> such units to the $A^{(2)}$ units. The fourth model was changed to keep track of the probabilities that the states of the $A^{(2)}$ units were correlated with errors in the responses, and used these probabilities to govern the reinforcement of connections <u>to</u> erroneous $A^{(2)}$ units. With only this change introduced, it was found that the behavior was similar to the previous model, which tended to over-generalize. This resulted from the fact that successive words in the training sequence were apt to find almost the same probabilities associated with the $A^{(2)}$ units, and consequently tended to establish the same codes for their phonemes. An attempt was therefore made to achieve a compromise between the two systems, by using the products of the probabilities that the $A^{(1)}$ units were on when errors occured, and the probabilities that the $A^{(2)}$ units were in their present state when errors occurred. Thus a connection was likely to be modified only if it went from an active $A^{(1)}$ unit whose activity was correlated with response errors, to an $A^{(2)}$ unit whose current state was also correlated with response errors. The parameters employed were similar to those used above, but were chosen so that the probabilities would change very slowly. Specifically, the parameters were:

$X_1 = 4$, $Y_1 = 4$, $X_{21} = Y_{21} = X_{22} = Y_{22} = 2$, $\theta_1 = 3$, $\theta_2 = 1$, $\theta_3 = 2$, $\theta_R = 1$, $\varepsilon_1 = .01$, $\varepsilon_2 = .05$, $\delta_1 = .05$, $\delta_2 = .05$. (The subscripts on ε and δ refer to the $A^{(1)}$ probabilities and the $A^{(2)}$ probabilities, respectively).

This scheme provided the most promising results of the series of experiments (for the two allophone per phoneme environment). Although a perfect solution to the problem was not obtained, the learning curve was a reasonable one, and the perceptron seemed to be systematically searching for a suitable code. It seems likely that further study of this model, or a related one, might yield greater success than has been obtained to date.

The learning curve of this system exhibits several features of interest. The system will typically begin with a large number of conflicts (i.e., codes which are assigned to more than one phoneme) most of which are resolved by probabilistic reinforcement of the preterminal network. Since this means that some of the phoneme codes have been changed, much of the previous learning which has occurred in the terminal part of the network is no longer valid, and the percentage of correct responses to words drops accordingly. This, in turn, tends to increase the probabilities that preterminal corrections will be made, so that the codes are likely to be further modified while the perceptron is trying to adjust its terminal responses. The resulting learning curve tends to be "stepped", each plateau occurring when a better code has been evolved. Unfortunately, this stepwise improvement in performance tends to break down when the system comes close to a complete solution, due to the fact that the codes are then perturbed faster than the terminal network can be retrained. What is apparently needed is a mechanism for slowing down front-end changes while speeding up the terminal changes, or a better criterion for evaluating the coding at the $A^{(2)}$ level.

One approach to these problems would be the use of a two-pass training cycle. In the first pass (in which each word would be presented) only the $A^{(3)}$ to R-unit weights would be modified, so that the overall performance with the currently established code could be evaluated. The second pass would be the same as the present training procedure, modifying both terminal and preterminal weights, or else would be restricted to modify only the preterminal weights, but without changing the probabilities. Such a procedure seems analogous, in some ways, to some concept formation, in which a concept is radically changed on the basis of an estimate of how well it has "worked" over a period of time.

An alternative approach seems to be the use of an "elastic perturbation" reinforcement system, which was originally proposed in Ref 1, (Section 26.4), and has subsequently been used with considerable success in a number of simulation experiments. In this model, when an error occurs, one of the $A^{(2)}$ units would be chosen, at random, to be readapted. The exact rules for changing weights could be similar to those employed in the above systems. The resulting performance of the entire network would then be reevaluated. If the change in the $A^{(2)}$ code has led to an improvement in performance, the change would be stabilized, otherwise it would be "erased", the preterminal network returning to its previous condition. This would tend to retain those code features which proved useful, while allowing only those codes to change which actually interfere with the operation of the network. A preliminary program was written to study this procedure, but the experiments carried out were insufficient to permit any meaningful evaluation of performance.

In all of these five-layer perceptron experiments, the basic philosophy of operation was to allow the system to hunt, more or less "at random", for a phonemic organization of the environment which would make it possible to learn "words" consisting of several phonemes each. In retrospect, it seems questionable that the process by which a human learns to distinguish phonemes is likely to operate in this fashion. The active production of sounds, or imitation, seems to play a most important part in human speech perception, and it seems likely that the sounds which a person can reliably distinguish correspond closely to those which he can reliably produce. Thus, if our $A^{(2)}$ layer, rather than consisting of an arbitrary set of "neurons" with no inherent significance of their own, actually consisted of a set of motor neurons regulating speech production, there would, in fact, be only one correct code possible for each phoneme; the system could then distingusih incorrect codes by a mechanism permitting the registration of discrepancies between the "copied" sound (emitted under the control of its own $A^{(2)}$ units) and the sound heard as part of the spoken word. Future experiments (which have already been initiated by Carl Kesler) will be directed towards a model of this variety.

REFERENCE:

1. Rosenblatt, F. Principles of Neurodynamics: Perceptrons and the Theory of Brain Mechanisms. Spartan Books, Washington, D.C., 1962.

A PILOT STUDY ON THE APPLICABILITY OF PERCEPTRONS TO THE CLASSIFICATION OF BUBBLE CHAMBER PHOTOGRAPHS

By Sherman Chow

1. Statement of Problem

The purpose of this study is to show how a perceptron may be used to analyze data from photographs obtained in bubble chamber experiments.

When a particle enters a chamber, it leaves a trace on a light sensitive film. If this particle collides with another particle, the trace deviates from the horizontal direction and forms a "Y". Since the collision may occur at any point inside a chamber, the position of the "Y" cannot be predicted. The direction of deviation is a function of properties of the colliding particle, and, therefore, is also impossible to predict (Fig. 1 and Fig. 2).

It would be desirable to separate the photographs of experiments in which a collision occurs from those in which no collision occurs. Since large numbers of photographs are taken for each experiment, the task of sorting by human operators becomes tedious. This paper will evaluate the performance of a perceptron in classifying photographs in which no collision occurs from those in which a collision has occurred.

The perceptron and the tracks of the particles were simulated on a 7090 computer and experiments were carried out under various conditions.

1.1 A Comparison of Track Recognition Using a Digital Computer and the Perceptron.

Many attempts have been made to use a digital computer for pattern recognition (e.g., Refs. 1,4,8). Some of these schemes propose to store the shape of all patterns to be recognized in a digital computer memory. When a stimulus appears, the computer searches through the memory to find the stimulus which most closely resembles the one shown and bases its decision upon it. For problems where the shape of the stimuli are well-defined, and the size and orientation can be fixed, this method will give good results with limited

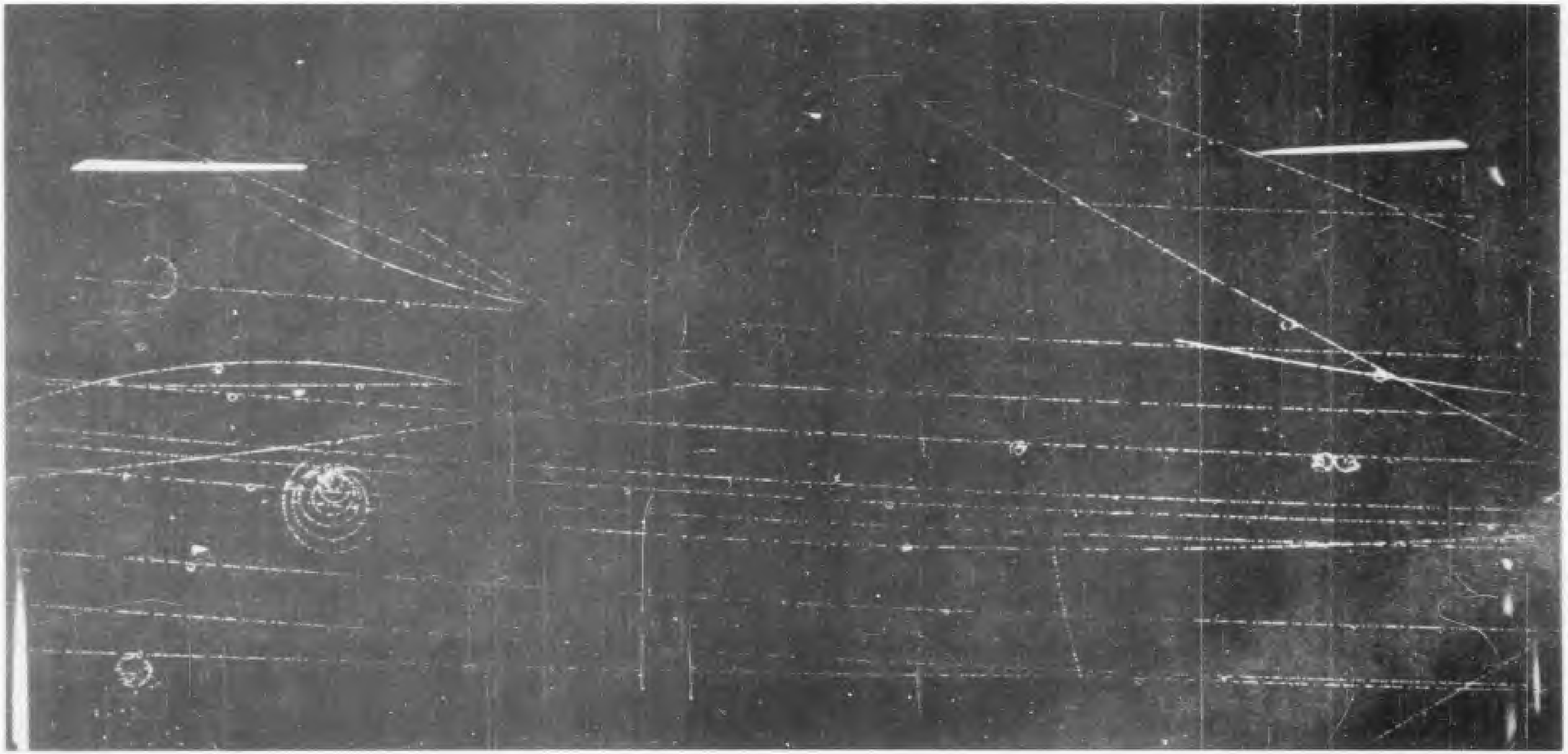


FIG. 1: PHOTOGRAPHS OF BUBBLE CHAMBER TRACKS
(Courtesy of Brookhaven National Laboratory)

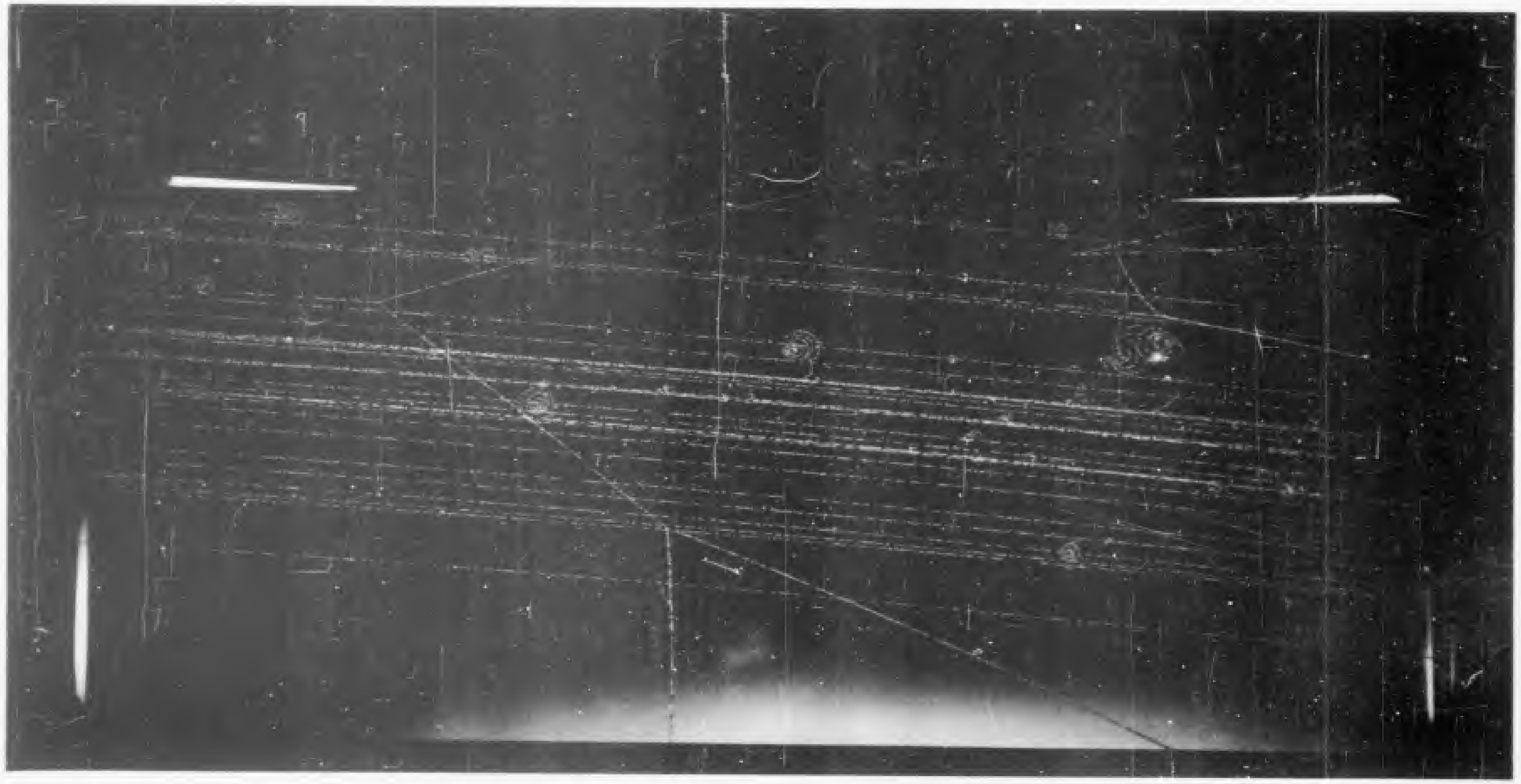


FIG. 2: PHOTOGRAPHS OF BUBBLE CHAMBER TRACKS
(Courtesy of Brookhaven National Laboratory)

equipment. In others, such as bubble tracks, the shape of the stimuli are not always the same; in fact, it is not possible to predict what the stimuli may do. If all possible patterns of an n by n square array are to be stored in memory, the total amount of storage required to do so would be $n^2 \times 2^{n^2}$ bits. For n of, for example, 64, this number will exceed the number of bits of storage available even in the largest computer.

Track recognition using digital computers has also been tried (c.f. Ref. 4). The most notable effort uses a digital computer and a scanning device which scans for lines in the photograph. The slopes of the lines are calculated and the lines are extended in the indicated direction. If any of the lines are shown to intersect, the computer recognizes this as an event. A problem encountered in this program is that there are a great many discontinuities in the lines, so that the computer is unable to determine the orientation of the pertinent lines.

The perceptron, which may be designed to generalize from one stimulus to a totally different one, does not require an exact correspondence between the training stimulus and the test stimulus (Ref. 6). Noise will not have any great effect upon identification of the stimuli provided that the noise is random in nature.

A preliminary study was undertaken by Kesler using a simple perceptron to detect angles and straight lines of bubble chamber tracks (Ref. 3). In these experiments, a 20 X 20 retina was used. Two classes of stimuli were generated: one having only horizontal lines, the other containing a mixture of horizontal lines and one angle. All the line segments are continuous.

The parameters of the perceptron were as follows:

Number of A-units	300
Number of excitatory connections per A-unit	6
Number of inhibitory connections per A-unit	0
Threshold of each A-unit	2

The connection of the A-units were arranged in such a way that an angle will cause some A-units to be active.

The results of these experiments show that a simple perceptron can learn all the classifications after about 500 trainings.

The experiments described in this paper are extensions of these experiments with a different type of perceptron, using larger stimuli and more variation in noise patterns to conform with actual bubble track photographs. It is believed that a perceptron containing a larger number of A-units will be necessary to perform classifications under these more difficult conditions.

2. Detailed Description of Track Recognition Perceptron

The perceptron is required to give information concerning the location of the collision as well as whether a collision has taken place. It was decided, therefore, to use a narrow retina and scan the photographs. The direction of the scan is perpendicular to the direction of motion of the particles. When a collision is detected, the perceptron can be made to indicate the position at which the collision was discovered.

The perceptron is, then, designed to recognize a collision only if the collision appears on the horizontal axis of the retina, midway between the horizontal edges.

The track recognition perceptron is a perceptron with one sensory layer followed by three A-layers. There is only one R-unit so that it can only dichotomize the stimuli into two classes. A diagram of this system is shown in Fig. 3. The variable connections are between the third A-layer and the R-unit.

2.1 Local Property Detectors

The retina consists of a 13 X 63 rectangular mosaic of light sensitive elements. The connections from the retina to the A-layer form "local property detectors" as suggested by the discovery of Hubel and Wiesel in the visual cortex of the cat (Ref. 2). They discovered that there are cells in the cat's visual cortex that will respond to lines and others that will respond to edges and gradients (Fig. 4). It is calculated that these "local property detectors" combined with proper connections to the deeper layers will result in more efficient use of the cells for processing the images (Ref. 6, pp. 512-521 and Ref. 7).

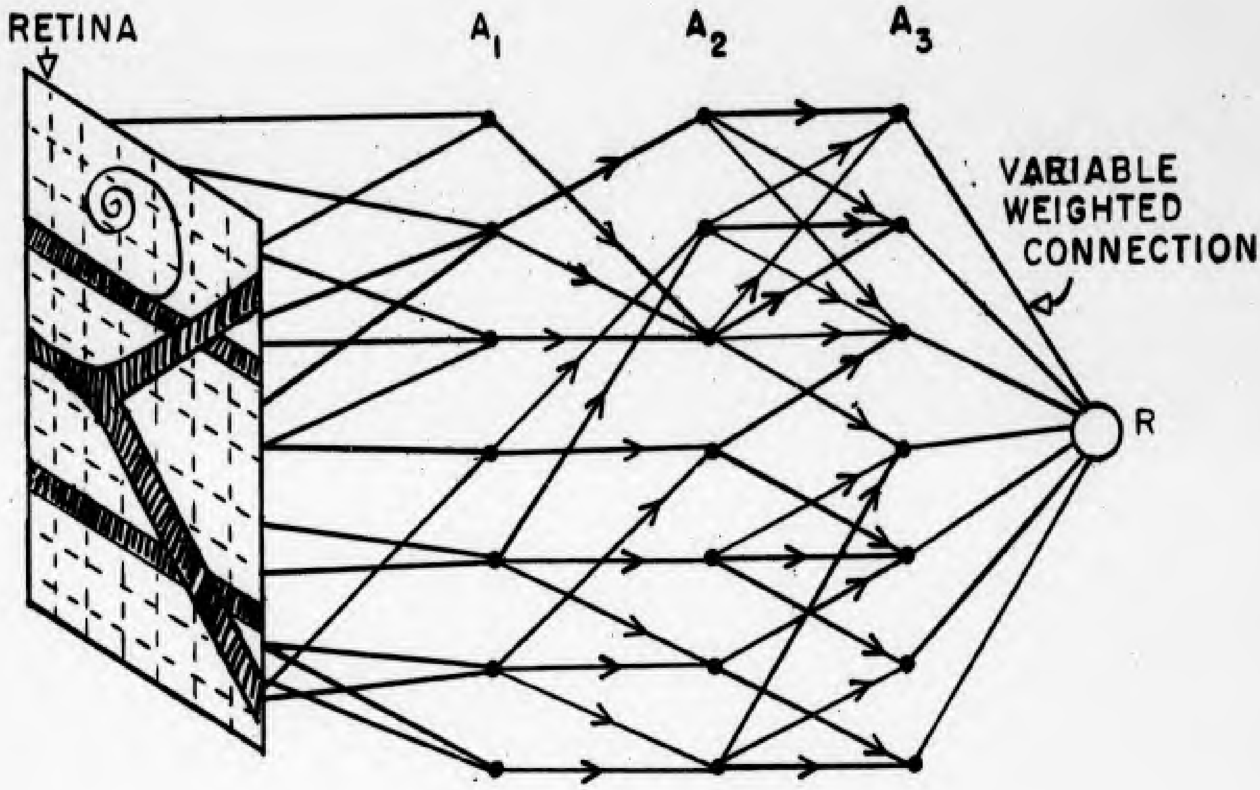


Fig. 3

Diagram of the Track Recognition Perceptron

2.2 The S to A_1 Connections

The S to A_1 connections are in the form of "local property detectors". For the particular problem of track analysis, the stimuli consist of lines and angles. There are 40 different types of local property detectors scattered on the retina. Some examples of these detectors are shown in Fig. 5. These "local property detectors" are constructed in a similar manner to those found in the cat, but due to the particular geometry of the stimuli, will be mostly "angle detectors", which actually do not exist in the cat.

The 40 local property detectors are made up of two groups of twenty different connection configurations. The difference between the groups is that the thresholds are set at 5 and 7 respectively.

2.3 The A_1 to A_2 Connections

The cells of the A_1 and A_2 layer are such that the value of the output will be equal to the input signal for inputs greater than the threshold, and 0 otherwise.

The cells of the A_1 layer will respond to tracks of a particular shape in a particular location. Since these tracks may appear anywhere along the horizontal axis of the retina, it is desirable that the perceptron respond to stimuli that are horizontal translates of each other in the same manner.

To accomplish this, the contour detectors of the same type along the horizontal axis will be connected to the same A_2 unit, and the threshold of the A_2 unit will be set at zero. Now the translates of the same stimuli will activate the same A_2 unit.

2.4 The A_2 to A_3 Connections

The connections between A_2 and A_3 are organized in a "binomial mode." This means that each A_3 unit receives a fixed number of connections from the A_2 units. The origins of these connections in the A_2 layer are picked at random. Since the signal to each A_3 unit is the difference of two binomially distributed random variables, this type of connection is called "binomial" (Ref. 6).

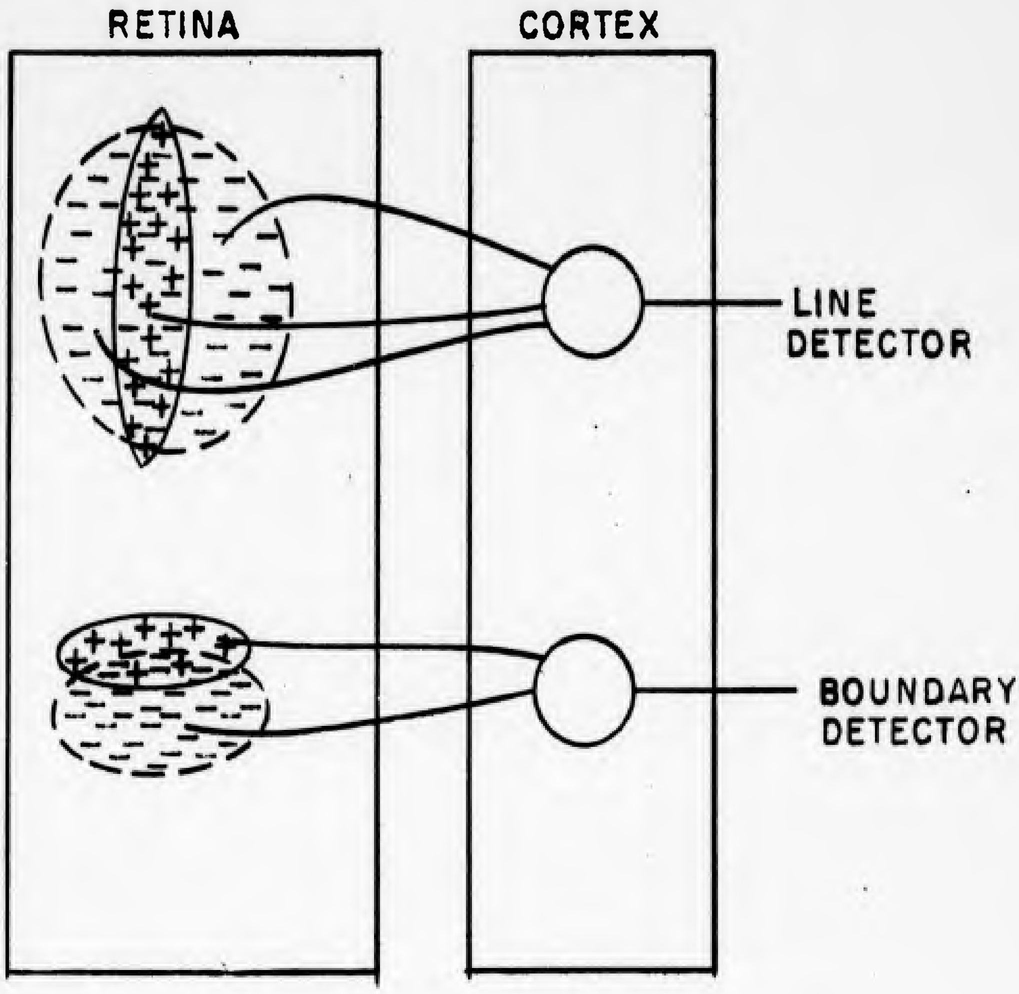
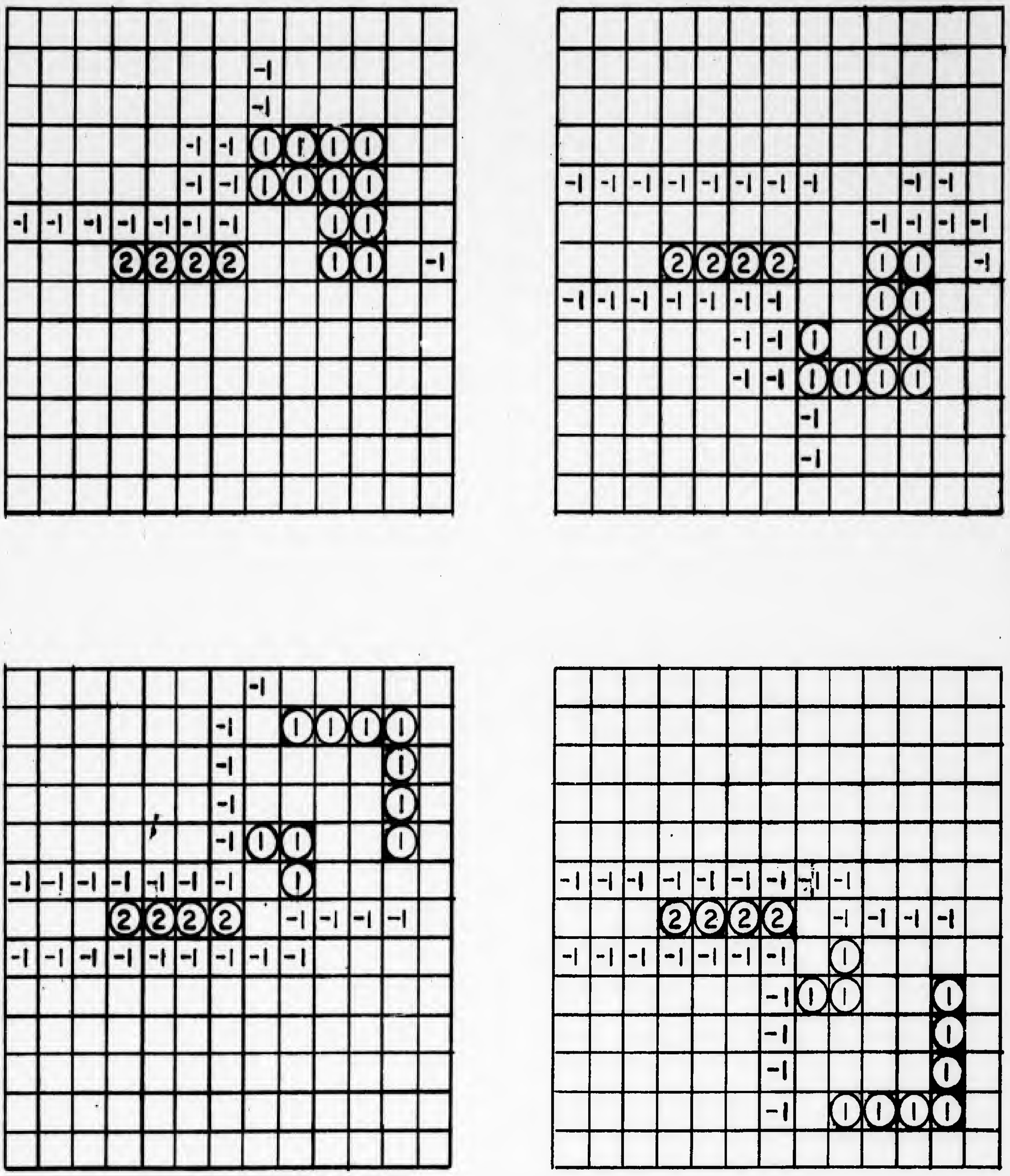


Fig. 4 Organization of Sensory Fields of Cells in the Visual Cortex of the Cat



(X) Denotes X "Excititory" Connections

_Y Denotes Y "Inhibitory" Connections

Fig. 5 Example of Local Property Detectors

2.5 The A_3 to R Connections

The A_3 to R connections are weighted, and the values of these weights may be changed during training.

3. The Perceptron Simulation Experiments

A perceptron is simulated on the 7090 digital computer (Fig. 6). The performance of this simulated perceptron is tested by generating a set of stimuli on the computer and computing the percentage of correct responses. The parameters of the perceptron can be easily changed (Ref. 5).

3.1 The Stimulus Generation Program

The stimuli generated by the program consist of points in a square field 13 X 63. In general, the stimuli are divided into two classes. For example, one class can consist of a set of pictures in which there are no collisions and the other of a set of pictures in which there are collisions. Alternate stimuli in the sequence always come from opposite classes.

Prototype stimuli are presented to the perceptron from punched cards. These may consist of a few samples of each class. The computer will automatically rotate, translate, expand or contract the stimuli given to create new stimuli. Examples of prototype stimuli are shown in Fig. 7A and 7B.

In some of the stimuli there will be curved tracks which represent electrons given off during a collision and curving under the influence of the magnetic field. These curved tracks are to be ignored by the perceptron. Other types of background noise such as spots due to photographic imperfections can also be simulated by the program.

These stimuli can be generated by the computer and stored on tape. These are available to the computer when they are needed.

The stimulus size is 63 X 63 while the size of the retina is 13 X 63 with the longer dimension in the direction of the particle track. A scanning procedure will be used by the perceptron to cover the whole stimulus.

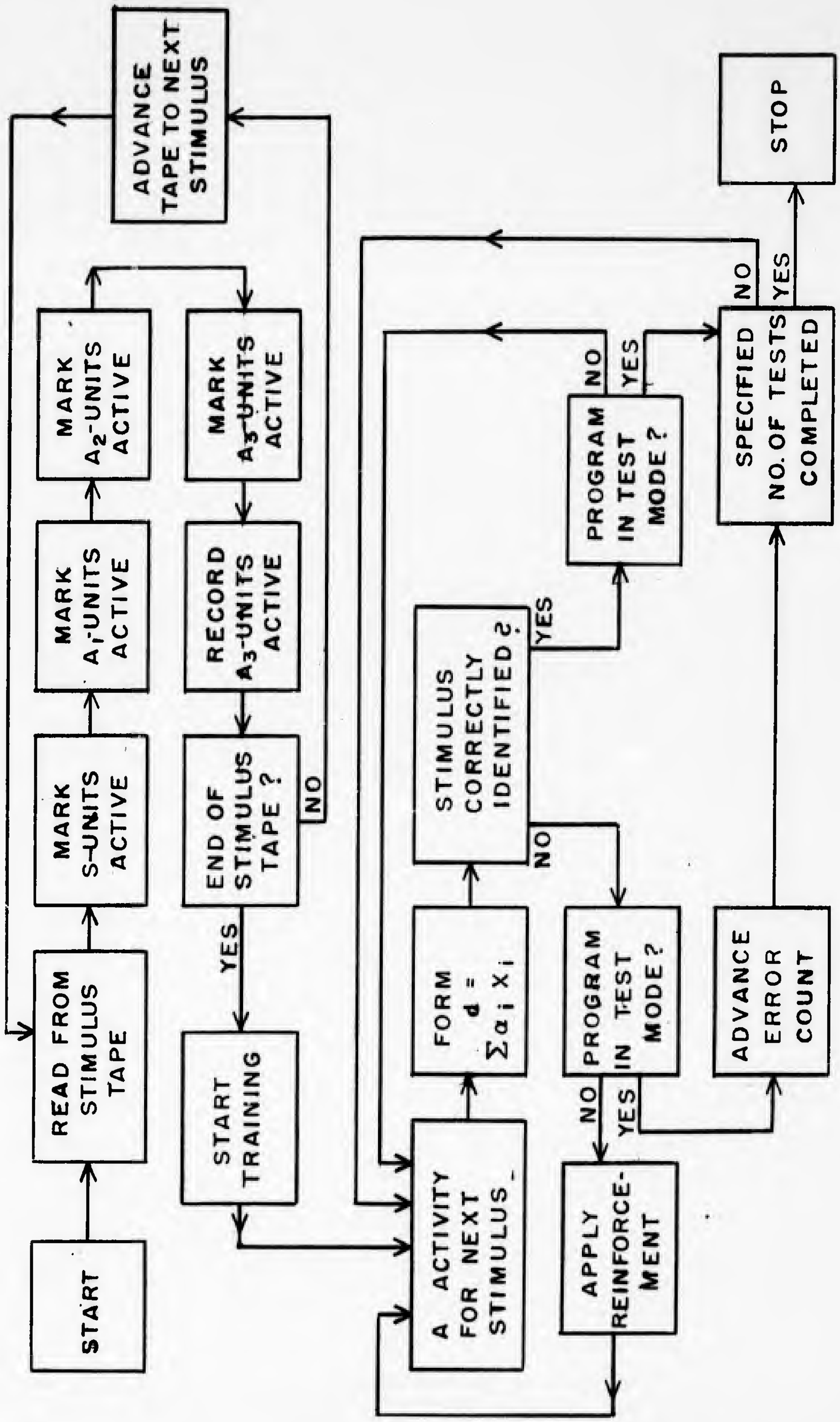


Fig. 6 Flow Diagram of Training and Test Cycles

3.2 The Perceptron Generation Program

The parameters of each perceptron are fed into the computer from punched cards, and each perceptron may be constructed from these parameters. The parameters fed into the computer consist of the following:

1. Number of cells in each layer
2. Number of excitatory and inhibitory connections per cell
3. Threshold of cells in each layer
4. The type of connections (see below)

If the cells of layer J and the cells of layer J + 1 are to be "binomially" connected, the program will generate K random numbers ranging from one to M where K is the total number of connections from the J to the J + 1 layer and M the number of cells in the Jth layer. This table will specify the location at which each connection will originate in the Jth layer. If the Jth and J + 1 layer are not randomly connected, the programmer may store the desired connection scheme into the memory of the computer from cards. This process may be repeated for each successive layer.

3.3 Training Sequence

When all the connections have been completed the training and testing sequence is ready to begin. The stimuli are read from tape and the active cells in the last layer for each stimulus is obtained; then these are stored. The activity of the i^{th} A-unit is designated x_i. The weights of A unit outputs will always start at $\alpha_1 = 0, \alpha_2 = 0 \ldots \alpha_n = 0$. Now

$$d = \sum_{i=1}^{n} \alpha_i x_i$$

will be calculated. Since alternate stimuli are from opposite classes then the inequalities

$$d_n > 0$$

$$d_{n+1} < 0$$

$$d_{n+2} > 0$$

must be satisfied where d_n is the signal to the R unit from the nth training stimulus on tape. If one of the inequalities is not satisfied, the α's corresponding to the active A-units are adjusted to decrease the error and the process is repeated until there are no more stimuli on tape or the process is ended by the program.

3.4 Test Sequence

Periodically, during the training sequence the process is stopped and a sequence of test stimuli is read from tape. The same procedure is followed in the test as in training except that α's are not adjusted. If an error has been made during this test sequence, it is recorded, thus giving a measure of the performance of the perceptron. If all the training stimuli are identified correctly, the whole program will halt, otherwise, it will continue until all the training stimuli have been used up.

3.5 Information Printed by the Program

The following information is printed out by the program.

1. All data fed into the computer from punched cards.
2. The number of cells in the A_2 layer excited by each stimulus.
3. The number of cells in the A_3 layer excited by each stimulus.
4. The value of all α's at the end of training procedure.
5. The number of training stimuli shown so far.
6. Input signal to the R-unit for each training and test stimulus.
7. Percentage of errors after each test sequence.

4 Experiment in Event Recognition

Perceptron characteristics:

A perceptron was simulated on the 7090 according to the rules described previously. The parameters of the perceptron are as follows:

1. Number of units in A_1 layer	2160
2. Number of units in A_2 layer	240
3. Number of units in A_3 layer	1000
4. Excitatory connections from A_2 to A_3	10
5. Inhibitory connections from A_2 to A_3	10
6. Threshold of A_1	5 and 7
7. Threshold of A_2	0

8. Threshold of A_3 2

9. Size of retina 13 X 63

The connections from the retina to A_1 units are in the form of "angle detectors". There are 40 different types, and examples of some configurations used are shown in Fig. 5.

4.1 The Stimuli

The stimuli are generated by the 7090 from information fed into the memory on punched cards. 770 stimuli are generated, consisting of two classes alternating, recorded on magnetic tape. The first 700 are used for training, and the last 70 for testing.

The stimuli have the following properties:

1. All collisions must leave at least one track that is more than $\pm 15^{\circ}$ and less than $\pm 75^{\circ}$ from the horizontal axis.
2. All other tracks must be less than $\pm 5^{\circ}$ from the horizontal axis.
3. There will be spiral-shaped tracks representing the motion of electrons splitting off and curving under the influence of the magnetic field.
4. All lines must be continuous.

The positive class will be comprised of all stimuli that have at least one collision on the horizontal axis in the center of the retinal field. The negative class will be comprised of all stimuli that have no collisions on the horizontal center line of the retina. Collisions away from the center line must be counted in this class because the stimuli are moved across the retina, and all collisions will eventually appear at the center.

The task of the perceptron will be to differentiate between the two classes. The simulation program is arranged so that training stimuli will be shown to the perceptron, and the perceptron will be trained using these 700 stimuli. The 70 test stimuli will be used to check the performance of the perceptron periodically during training. The test stimuli and the training stimuli are in no way similar except for the above rules. The perceptron must learn the classifications of the training stimuli and generalize to the test stimuli.

4.2 Result of Experiments Under Noiseless Conditions

The result of the above experiment shows that the perceptron is able to learn up to 98% of the stimuli correctly (Fig. 8). In the 70 test stimuli the only error made is misrecognition of the negative class. This means that all tracks containing collisions are correctly identified, while one error was made in which a track containing no collisions was mistaken for one containing a collision. This type of error is less serious than if a collision is missed by the perceptron. The performance of the perceptron may be biased by changing the R-unit threshold. Some quantitative results of R-unit threshold variation may be seen in Fig. 14.

4.3 Detection Experiments Under Noisy Conditions

In actual photographs of bubble chamber tracks the line segments will not be continuous. There also may be random noise in the background caused by imperfection in the photographic process.

Noise introduced in the simulation program is controlled by two parameters. An "illuminated" point can be turned off with probability P_1 and an "unilluminated" point can be turned on with probability P_2. Examples of noisy stimuli with various values of P_1 and P_2 are shown in Figs. 7A, 7B, 9, 10, 11 and 12. The performance of the perceptron under the various noise conditions is shown in Fig. 8.

The results of the experiments can be summarized as follows:

1. Noise of the type due to photographic imperfection represented by randomly placed spots in the stimulus did not cause serious deterioration in recognition. Random noise spots up to about 20 per cent of stimulus size have been tried and the decrease in performance due to this additive noise is small.
2. Noise due to discontinuity up to 20 percent of the total line length can be tolerated without seriously affecting the performance of the perceptron. When more than 20 percent of the tracks are missing from the photographs the performance is seriously affected.
3. When the threshold of the R-unit was raised from 0 to 18 all stimuli containing a collision were correctly identified. However, one more track was erroneously identified as a collision where there was no

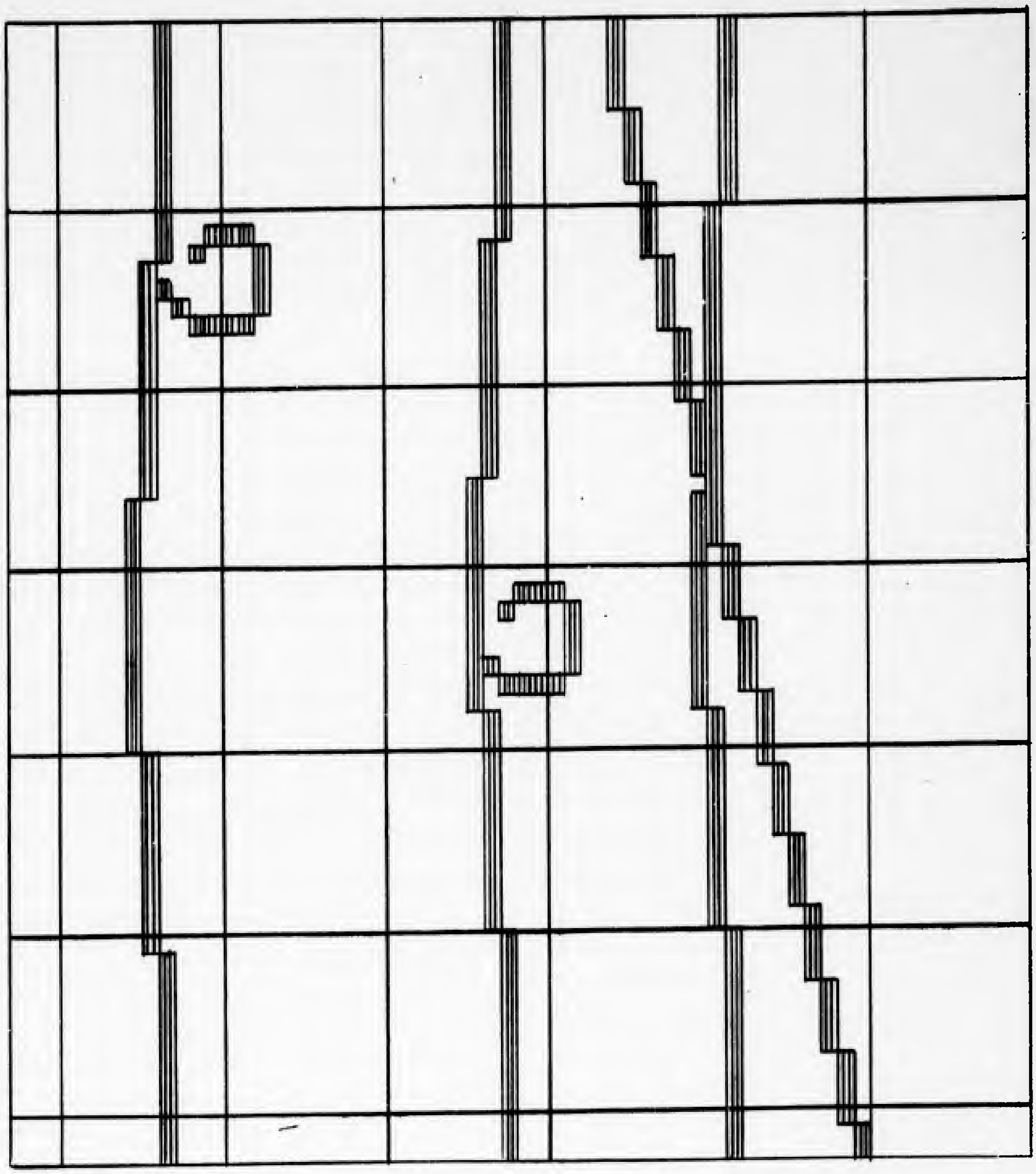


Fig. 7a Example of Stimulus Containing No Collision $P_1 = 0$ $P_2 = 0$

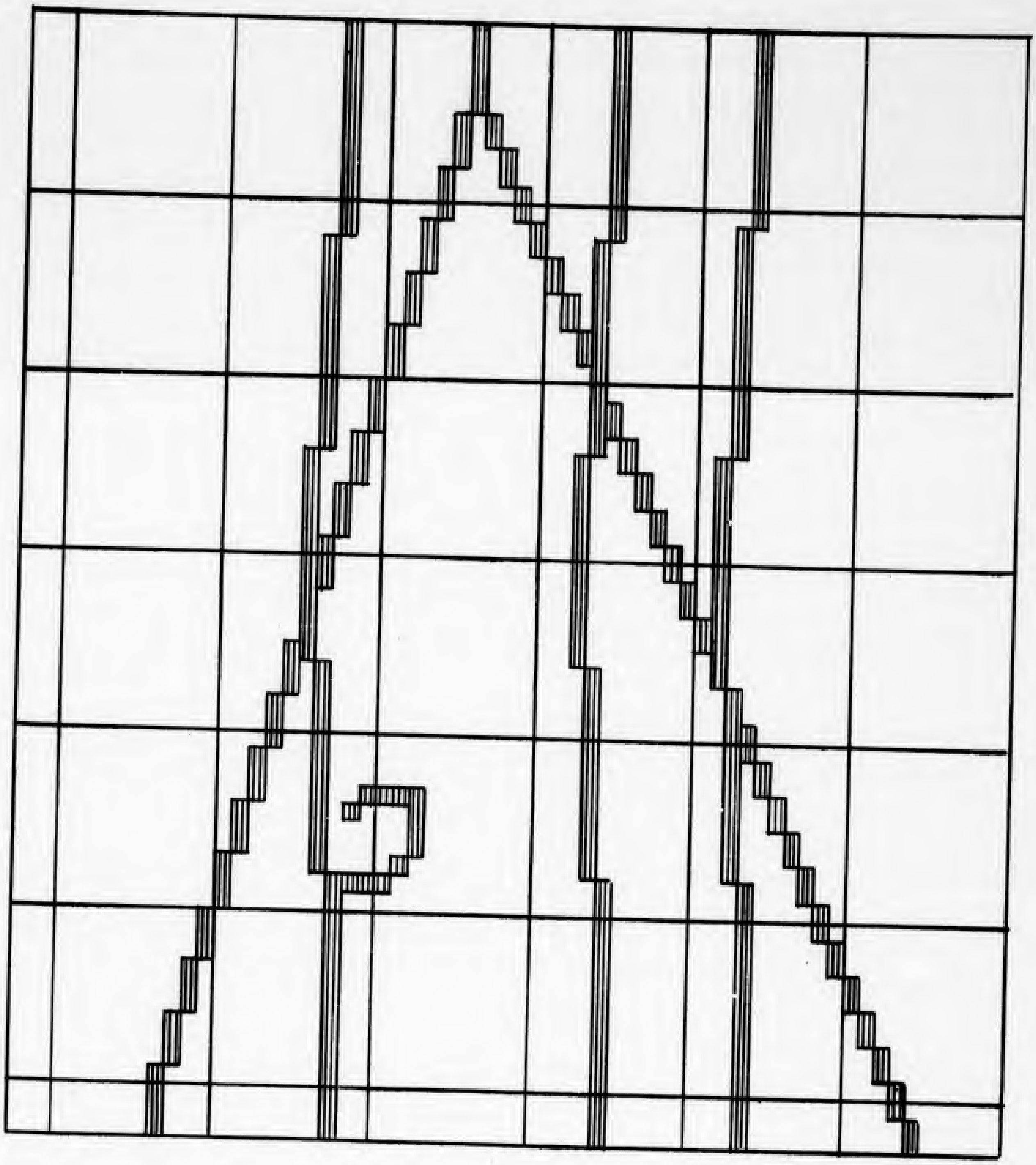


Fig. 7b Example of Stimulus Containing a Collision $P_1 = 0$ $P_2 = 0$

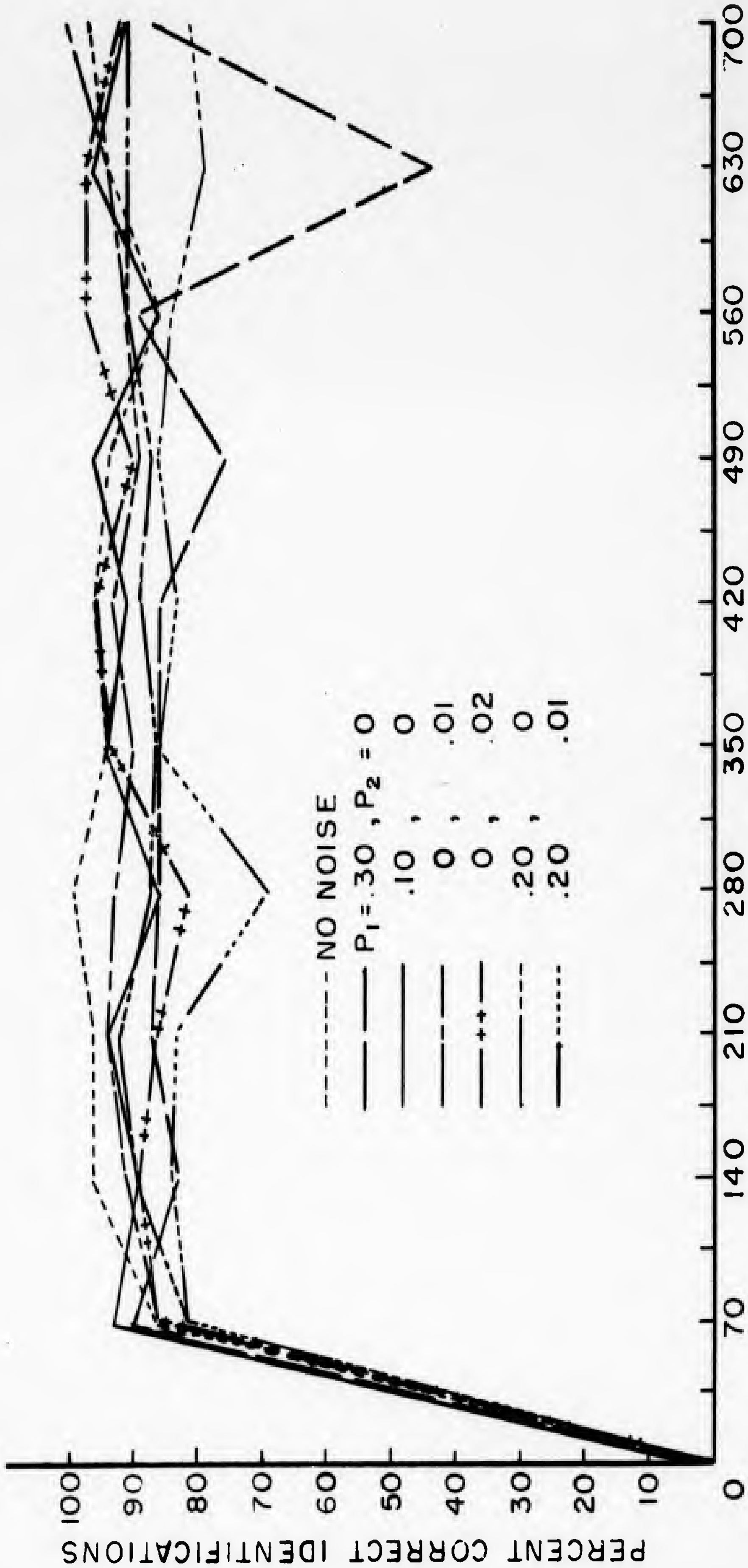


Fig. 8 Perceptron Performance Under Various Noise Conditions

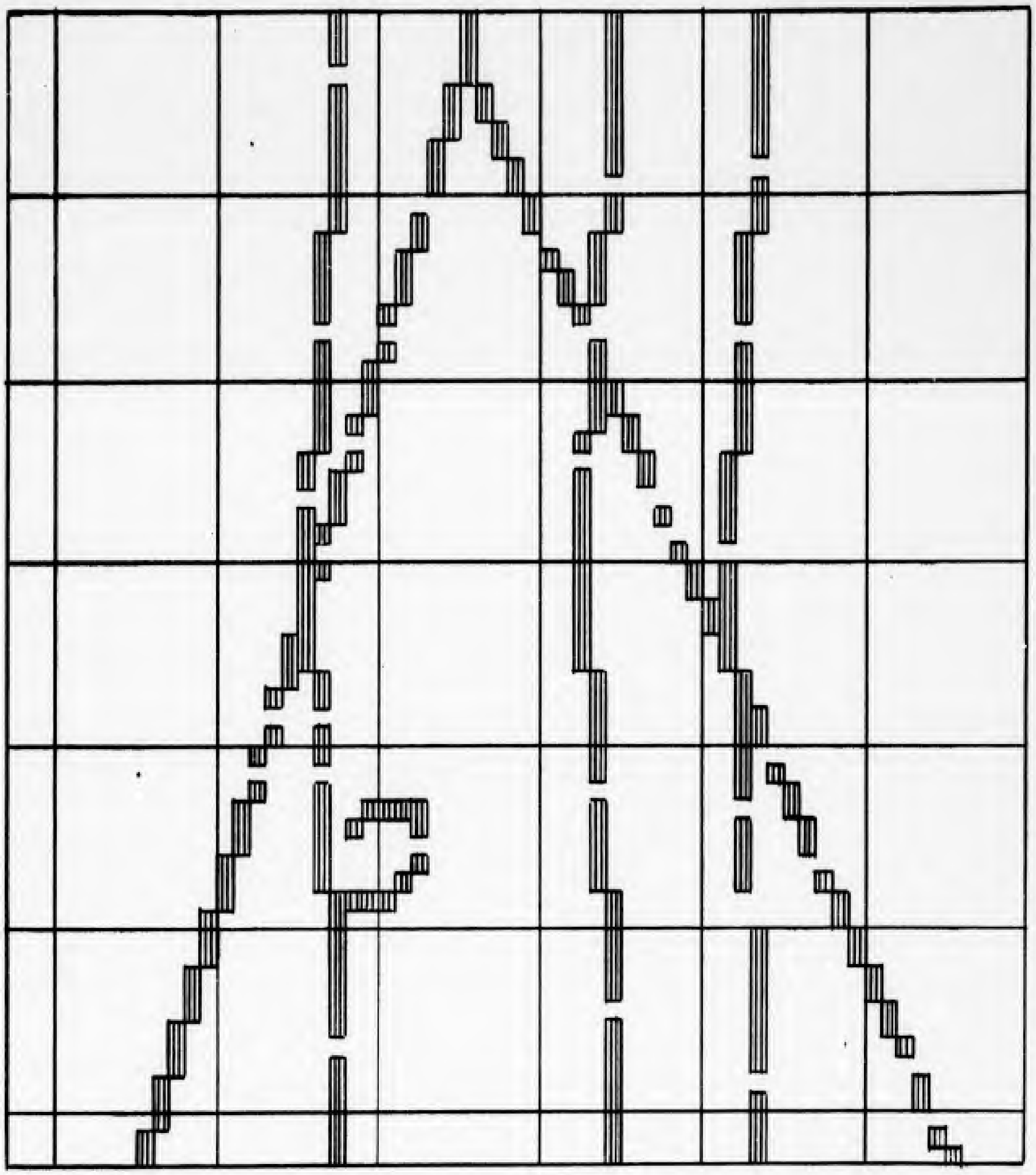


Fig. 9 Example of Stimulus $P_1 = .10$ $P_2 = 0$

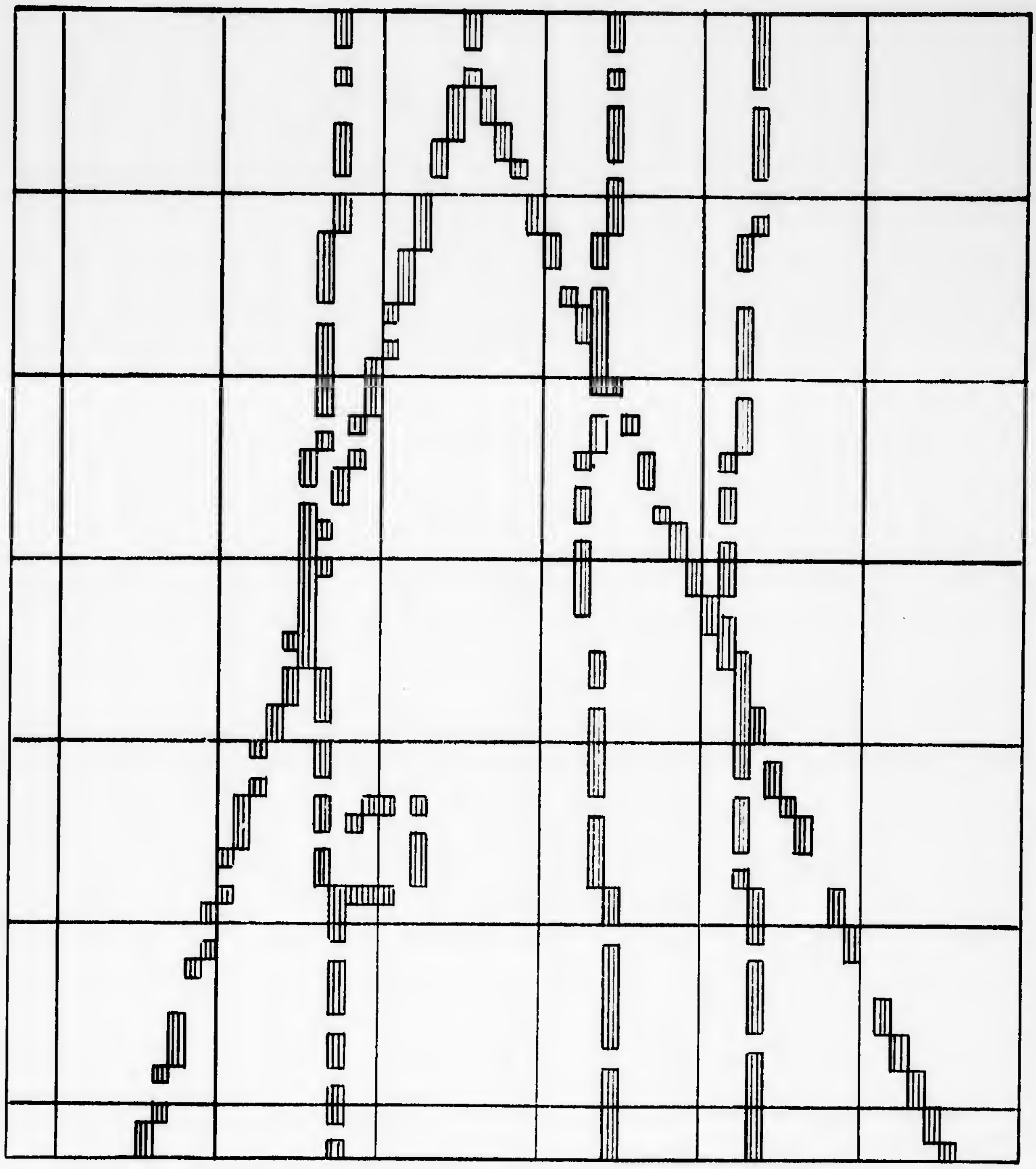


Fig. 10. Example of Stimulus; $P_1 = .20$ $P_2 = 0$

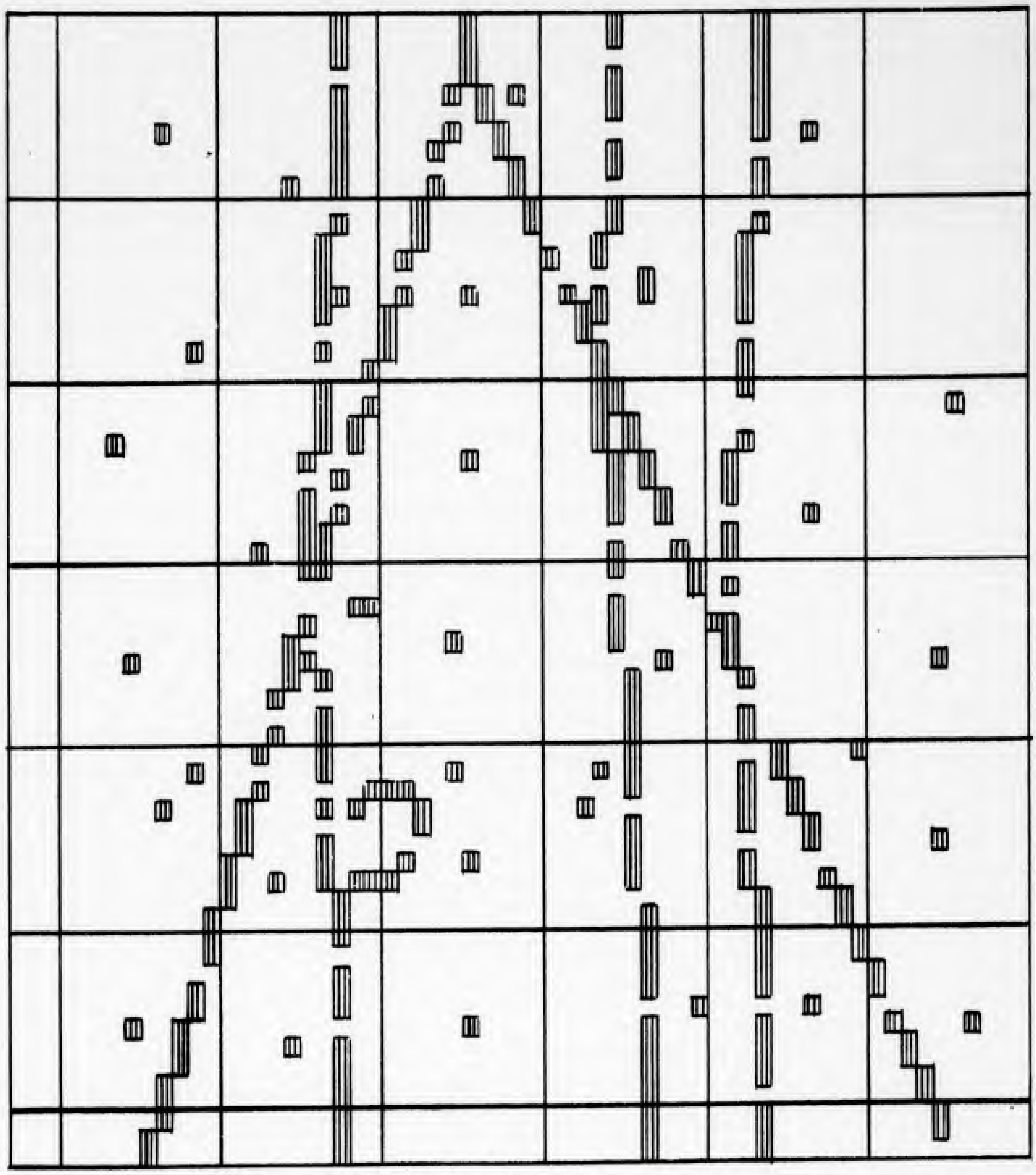


Fig. 11 Example of Stimulus; $P_1 = .20$ $P_2 = .01$

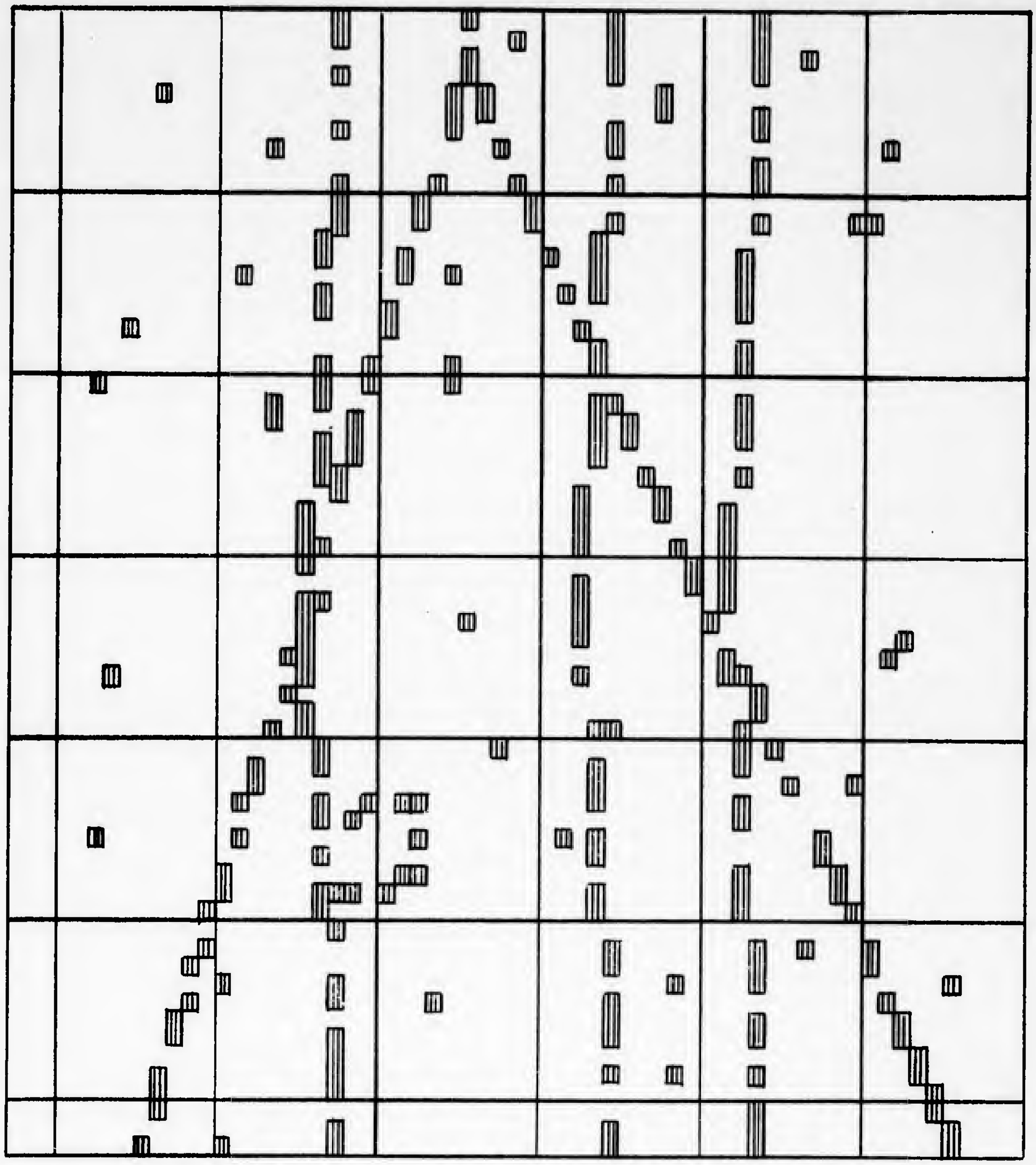


Fig. 12 Example of Stimulus; $P_1 = .30$ $P_2 = .01$

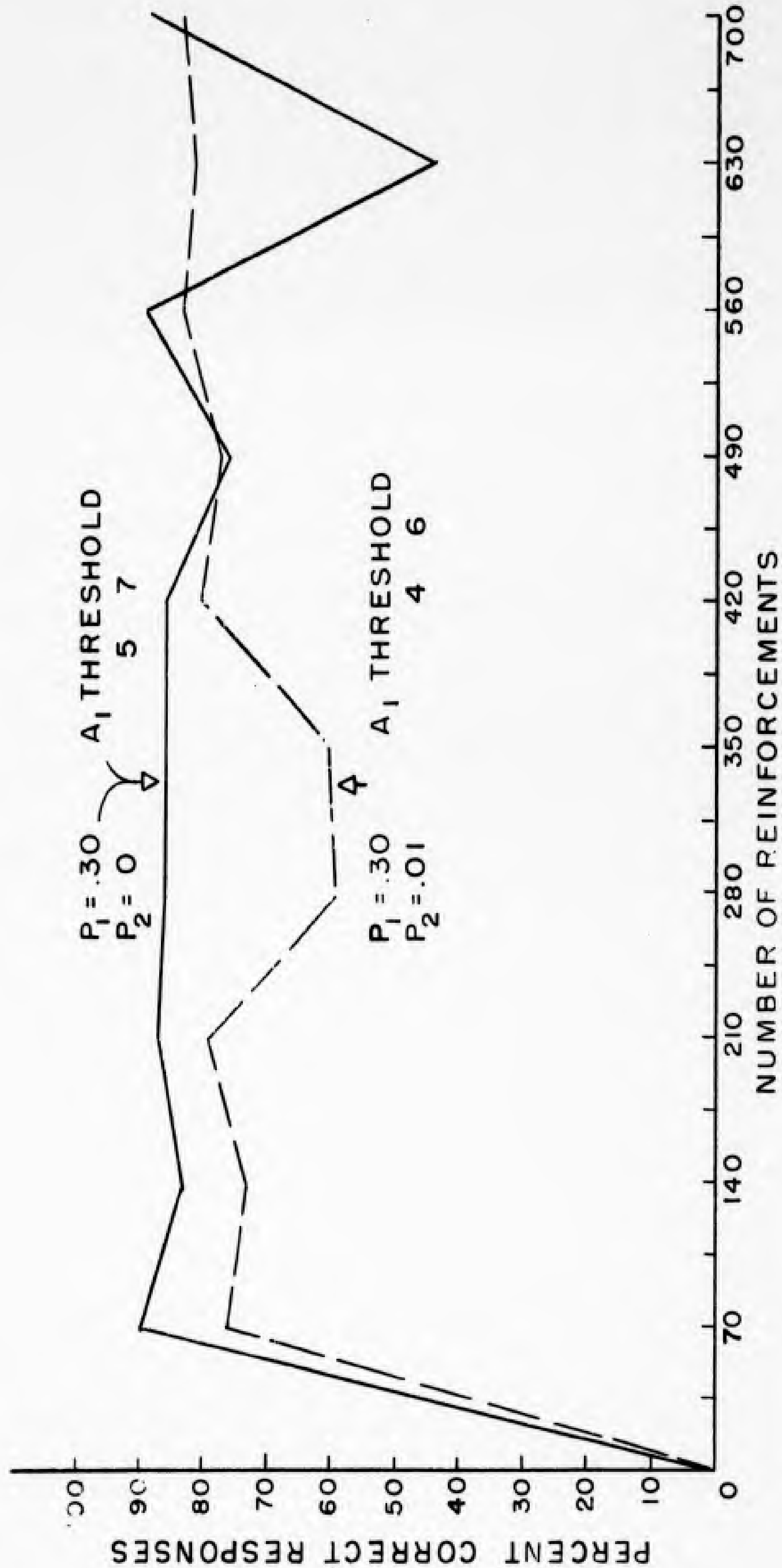


Fig. 13 Effect of Lowering A_1 Threshold Under Noisy Conditions

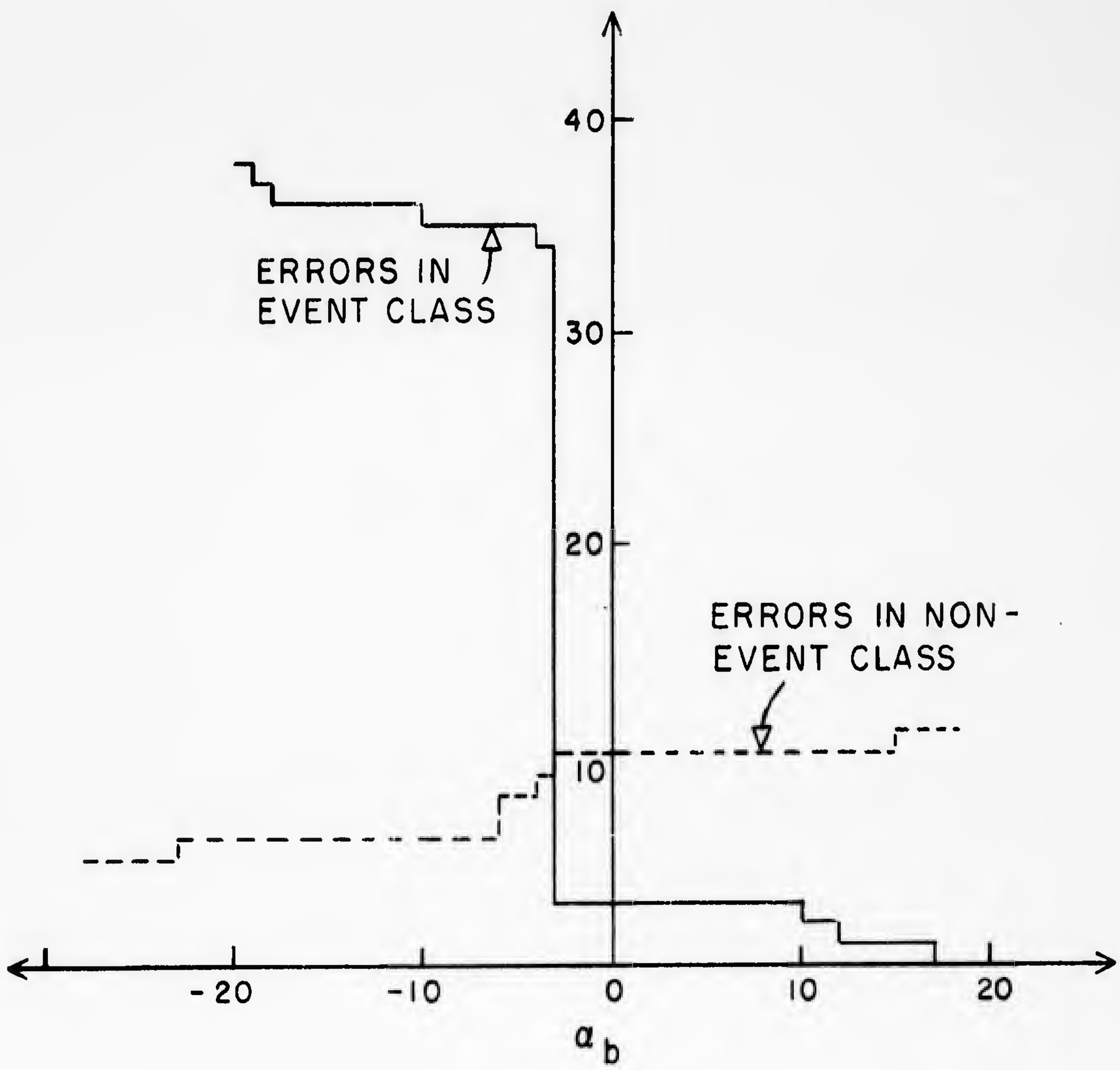


Fig. 14 R-Unit Threshold and Number of Errors in Each Class

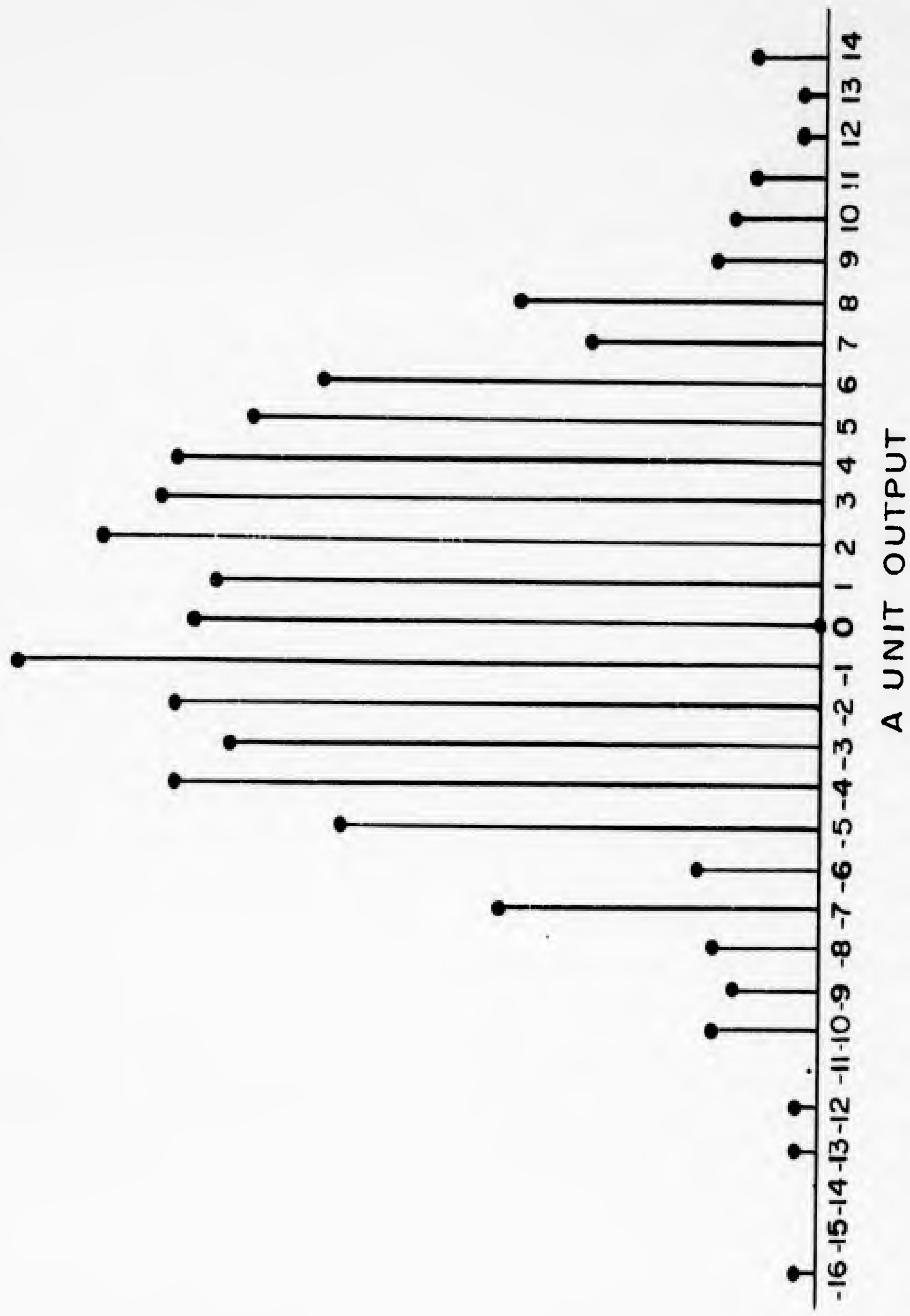


Fig. 15 Distribution of A-Unit Output

actual collision. Effects or errors from each of the two classes plotted against the R-unit threshold for values close to zero are shown in Fig. 14.

4. The values of the A-unit output ranged from +15 to -15 at the end of the last training cycle. For a distribution of weights see Fig. 15. These weights will give some idea of the range of A_3 unit output required in order to achieve a solution in a classification of this complexity.

4.4 The Effects of A_1 Unit Threshold

In the experiments described above the local property detectors are divided into two groups which are identical except for their thresholds. One group has a threshold of 5, and the other has a threshold of 7. It is found that with these thresholds performance in the case where $P_1 = .30$, $P_2 = 0$ is very poor. In fact, in some cases, the performance went below 50% (see Fig. 8). The large fluctuation in the learning curve would indicate that the perceptron had not found the solution at the end of the training.

From the configurations of the local property detectors used (see Fig. 5) it is hypothesized that with A_1 thresholds of 7 any stimulus with a large percentage of line segments missing would have a relatively small chance of activating an A_1 unit. To test this hypothesis, the A_1 units with a threshold of 7 were exchanged for a group with a threshold of 4. The experiment was repeated and the results are shown in Fig. 13. An example of a stimulus used is shown in Fig. 12. As seen from the graphs, the result of lowering the threshold seems to be that the performance of the perceptron improves when the threshold is decreased. This seems to indicate that the hypothesis is correct.

The second indication that the hypothesis is correct is that A_2 activity increased appreciably in every case with the lowered threshold. It can be concluded that with discontinuous line segments, performance may be improved by lowering the threshold of A_1 units.

Conclusions:

5. Design Criteria Derived from Simulation Experiments

The series of experiments was designed to find the parameters of a perceptron which will perform the track recognition problem satisfactorily. The following specifications were written with the limitation of available hardware in mind, and are based upon the particular perceptron which was simulated.

1. The A_1-units should have at least 50 inputs. This will include both the excitatory and inhibitory inputs. Experiments have been tried with smaller numbers of connections, but the perceptrons appear to be excessively sensitive to noise. There should also be an adjustable threshold θ for each unit that can be varied in value from 0 to 10.
2. The output of the A-unit should have at least thirty levels uniformly spaced. If quantized α-reinforcement is to be used, then the values of the output must be able to change by one integer step at a time.

5.1 Training Procedure

The suggested training procedure for this perceptron is as follows: A series of photographs can be selected from the experimental results. These can be classified by the experimenter, and used to train the perceptron. If necessary, the photographs classified by the perceptron as non-events can be reclassified by another perceptron trained from different photographs of the same experiment. In this manner the errors in that class can be reduced.

5.2 Additional Experiments

One additional series of experiments has been planned in this program. Some actual photographs will be digitalized and recorded on magnetic tape. These will be classified and arranged in the same sequence as the stimulus tape described in section 4.1. The experiments described in section 4.3 and 4.4 will be repeated.

This experiment will give more information on some parameters as yet undetermined.

1. Optimum magnification of the photograph for best performance.
2. Amount of overlap between successive scanning frames.
3. The distance between the cells in the S layer.

Another useful experiment would be to use the perceptron to differentiate the different interactions of the particle. For example, it would be useful to test whether a perceptron can be designed to differentiate "two prong" events from tracks with two or more prongs. For this experiment the "local property detectors" must be changed accordingly. With the information obtained in these experiments a hardware model of a track recognition perceptron might be built with a good chance of success.

BIBLIOGRAPHY

1. Highleyman, W.H., "Linear Decision Functions with Application to Pattern Recognition," Proc. IRE, Vol. 60, 1962, pp. 1501-1514.

2. Hubel, W. and Wiesel, T.N., "Receptive Fields, Binocular Interaction, and Functional Architecture in the Cat's Visual Cortex," Journal of Physiology, Vol. 160, 1962, pp. 106-154.

3. Kesler, C., "Preliminary Experiments on Perceptron Application to Bubble Chamber Event Recognition," Collected Technical Papers, Vol. 1, Report No. 1, Cognitive Systems Research Program, Cornell University, Ithaca, N.Y., January 1, 1961.

4. Mermelstein, P., Computer Detection of Vertices in Pictures of Nuclear Reaction, M.S. thesis (unpublished), M.I.T., September 1960.

5. Rosenblatt, F., "Perceptron Simulation Experiments", Proc. IRE, Vol. 48, pp. 301 - 309.

6. ________, Principles of Neurodynamics, Spartan Books, Washington, D.C., 1962.

7. ________, " A Comparison of Several Perceptron Models," in Yovits, Jacobi, and Goldstein (eds.) Self-Organizing Systems 1962, Spartan Books, Washington, D.C., 1962.

8. Sherman, H., "A Quasi-Topological Method for Recognition of Line Pattern", Proceedings of the International Conference on Information Processing, UNESCO, Paris, 1959.

A DESCRIPTION OF THE TOBERMORY PERCEPTRON

By F. Rosenblatt

1. Introduction

In the Spring of 1961, design work was initiated on a new physical model of a perceptron system, intended primarily for work in speech recognition. This system has since been named Tobermory, after H.H. Munro's talking cat. The original design called for a three-layer audio-perceptron, which has been described in Ref. 6. Subsequent modifications have converted this into a four-layer system, organized as shown in Fig. 1. Provisions are being made for the addition of a binocular visual input, having the organization of a four-layer similarity-constrained model (Refs. 7,8). The addition of further weighted connections is also contemplated as a possible future modification, to permit studies of selective attention controlled by audio inputs (Ref. 7, Chapt. 21) and to provide a new type of sequential memory (Ref. 9). A tentative design for this expanded system is shown in Fig. 2.

Since the number of adaptive weights in the connections of this system is considerably greater than in any physical model previously constructed, the highest priority task at the outset was to find a satisfactory economical integrator design. Electromechanical integrators, which were employed in the Mark I perceptron (Ref. 3), were rejected at the outset as too slow and too expensive for the Tobermory application. Studies of electrolytic and magnetic integrators were undertaken, and a magnetostrictive integrator (proposed by Charles Rosen at the Stanford Research Institute) was developed by George Nagy to a point where it would satisfy the cost and performance criteria which had been established (Ref. 4). More recently, however, a still simpler design employing tape-wound cores has been developed by Nagy in collaboration with the SRI group, and this will be the integrator actually employed in the system (See Refs. 1,2,5).

In the meantime, work has progressed on the design and construction of parts of the sensory system and R-units, which are independent of the integrator design. Parts of the system which are now under construction are shown in

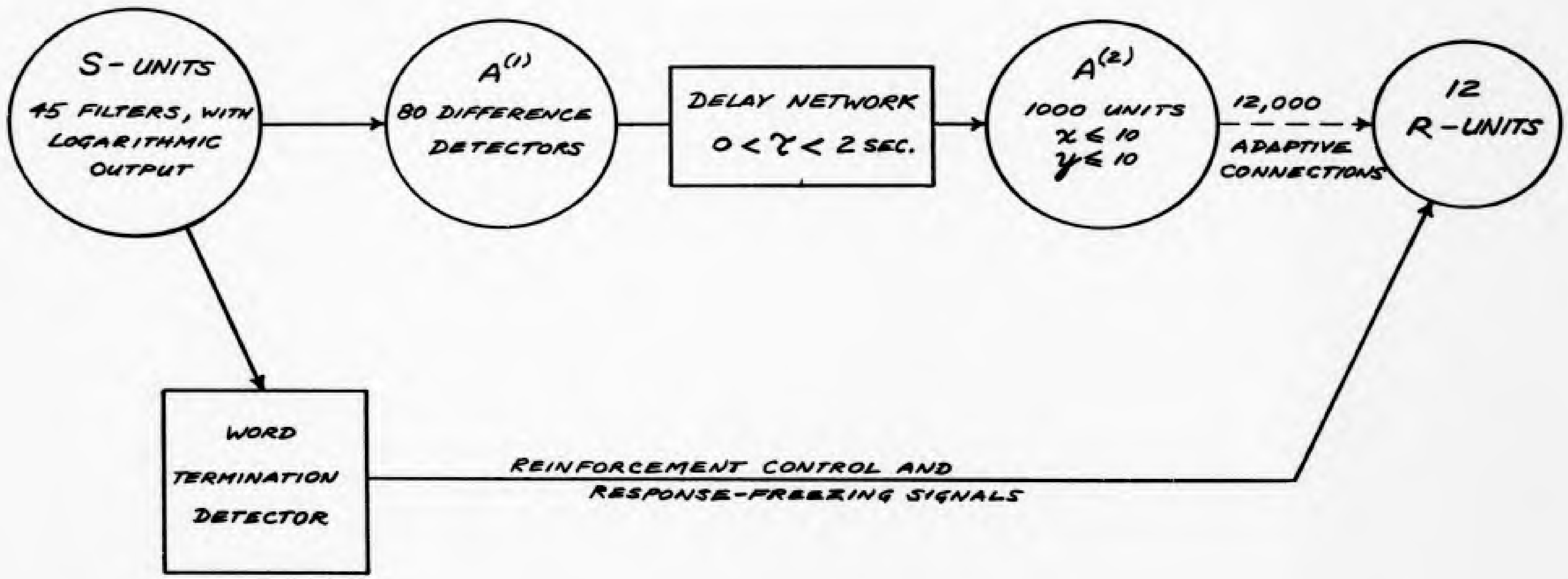
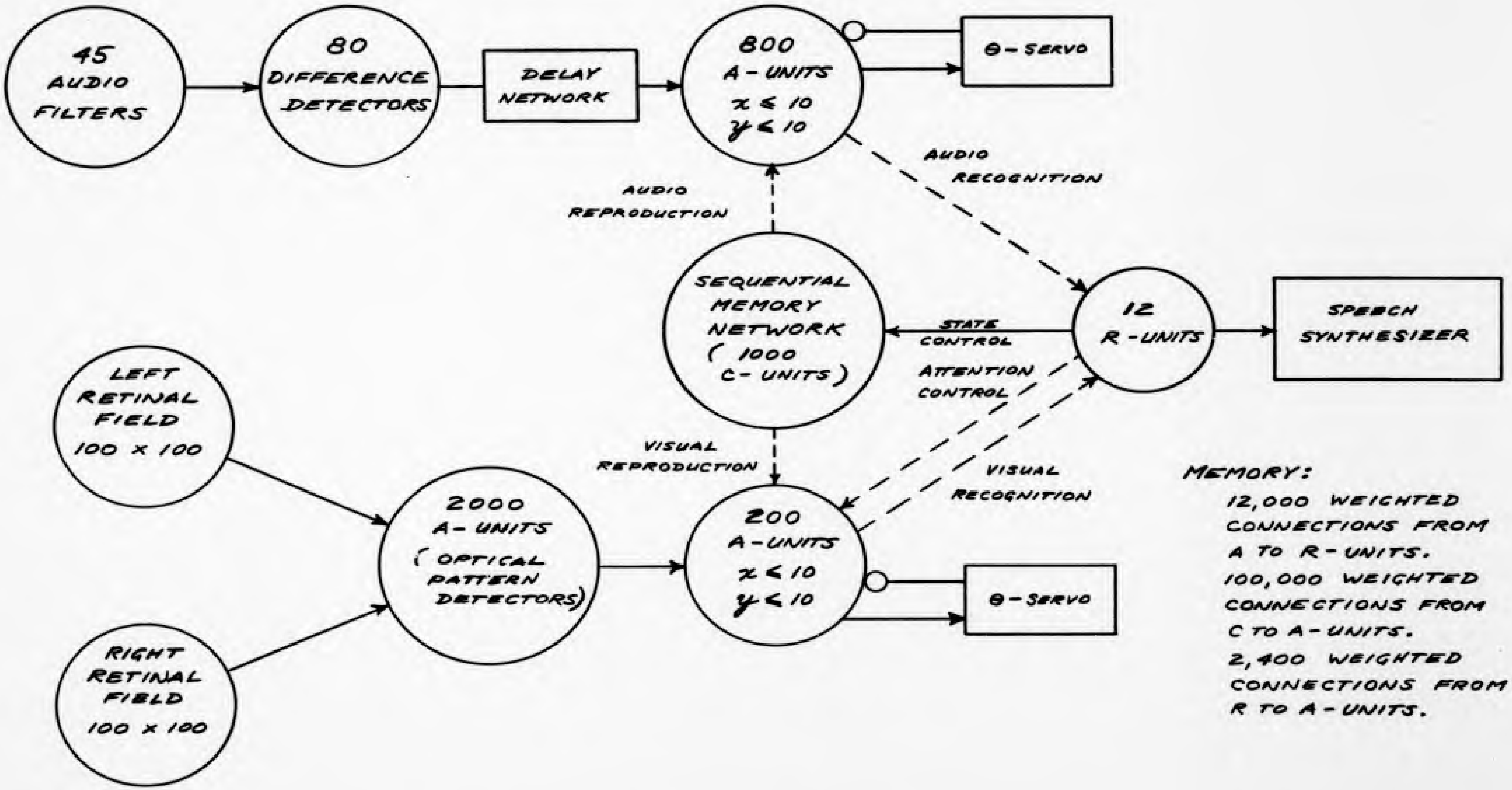


FIGURE 1: TOBERMORY ORGANIZATION: FOUR-LAYER PERCEPTRON WITH TIME-DELAY NETWORK BETWEEN SECOND AND THIRD LAYERS.



MEMORY:
- 12,000 WEIGHTED CONNECTIONS FROM A TO R-UNITS.
- 100,000 WEIGHTED CONNECTIONS FROM C TO A-UNITS.
- 2,400 WEIGHTED CONNECTIONS FROM R TO A-UNITS.

FIGURE 2: ENLARGED TOBERMORY SYSTEM

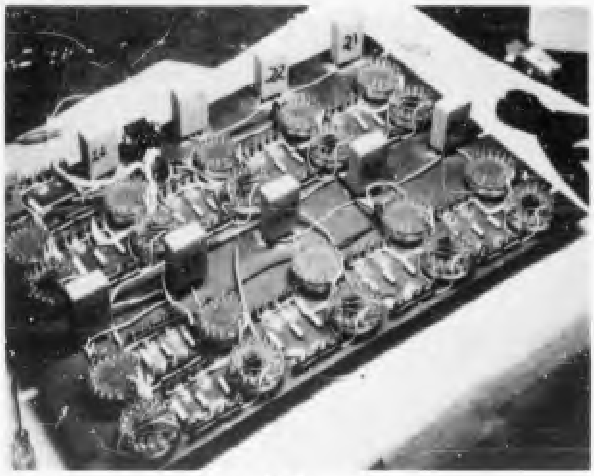
Fig. 3. It is hoped that most of the Tobermory system will be operable by the fall of 1963. The following description covers the detailed organization of the blocks shown in Figure 1.

2. The Sensory System

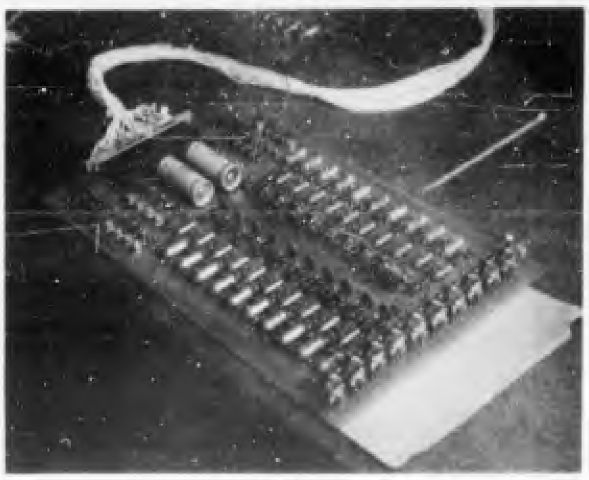
A detailed block diagram of the audio input system, $A^{(1)}$ units, and delay-connections to the $A^{(2)}$ system is shown in Figure 4.

The sensory inputs can be any mixture of signals from a tape recorder, microphone, a pair of audio oscillators, and a noise generator. An AGC amplifier with 25 db. compression can be switched into the system at the operator's discretion. From here the signals are amplified for input to the filter network. The signal coming from this main amplifier is available for display on an oscilloscope in the control room, as are the outputs of the filters. The audio signal also goes to monitoring speakers, a volume meter, and an amplitude measurement circuit which emits a voltage proportional to the logarithm of the average amplitude of the signal. This measurement is used to trigger the word termination detector (or pause detector) which is activated by a period of silence following an audio input signal. It is also averaged over a longer time period, to provide information to the perceptron on the amplitude profile of the input pattern, which would otherwise be lost in the frequency analyzing network. Both the "momentary amplitude" and the "average amplitude" are available, along with the logarithms of the 45 filter outputs, at Plug Board No. 1. The word termination detector initiates a timing sequence (with delays which can be set by the operator) which freezes the response state of the perceptron (to prevent random changes in response at the end of a word) and controls the period of reinforcement, so that the perceptron will be reinforced only for a short period while the word is "in register" in the delay network. This prevents reinforcement from being given, for example, during the first few phonemes of a long word, before the response of the system has been clearly established.

Each of the 45 filters can be set to a choice of three center frequencies and five bandwiths. If all filters are set to their first frequency position, they will cover the range from 30 to 4700 cycles per second, with a uniform

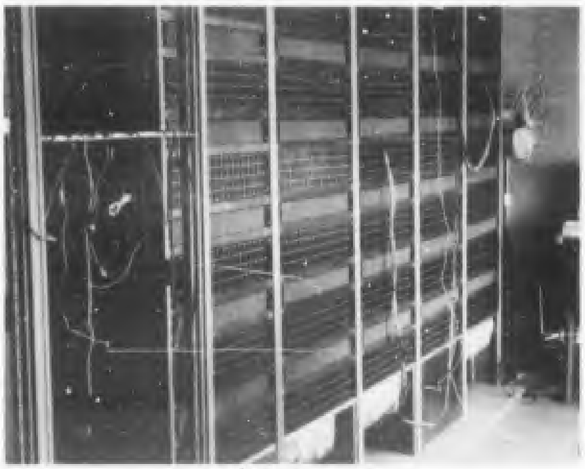


(a)



(b)

(c)



(d)

Fig. 3: PARTS OF THE TOBERMORY SYSTEM

(a) Panel With 10 Audio Filters

(b) Printed Circuit Board With 20 Delay Multivibrators

(c) Memory Module With 600 Integrators

(d) Plug Board No. 2 (Front View)

(e) Plug Board No. 2 (Rear View)

(e)

distribution in the Mel scale. In position No. 2, they cover the range from 47 to 7,000 cps, and in position No. 3 they cover the range from 60 to 9400 cps, likewise uniform in the Mel scale. The five bandwiths available for each filter give a choice of Q values of 1, 3, 5, 8, and 12 (corresponding to a choice of bandwiths equal to 100 percent, 33 percent, 20 percent, 12.5 percent, and 8.4 percent of the center frequencies).

Each of 40 differential amplifiers (representing the $A^{(1)}$ units of Fig. 1) can be connected to any pair of signals from filters or amplitude measuring devices, by means of the first plug board. Since all of these signals are represented in logarithmic form, the signal from the differential amplifier represents the ratio of the two amplitudes, rather than the absolute difference. This eliminates the need for a high quality AGC amplifier, and effectively normalizes the speech input for variability due to changes in volume, distance from microphone, etc. Each differential amplifier has two output channels, one of which carries a signal if the difference is positive, and the other if the difference is negative. Each of these differences is fed to a threshold gate, with an adjustable threshold. This system, then, effectively analyzes the profile of the frequency spectrum, or the ratio of the amplitudes at selected pairs of points throughout the spectrum. It is this set of ratios (now represented in digital form by the outputs of the 80 threshold gates) which characterizes the audio pattern for the subsequent parts of the system.

In order to represent the time dimension of the input pattern, the sets of eighty signals representing the momentary frequency spectrum are fed into a set of 1600 one-shot delay multivibrators, arranged in eighty channels of 20 multivibrators each. Whenever a threshold gate is activated (indicating that some ratio of frequency amplitudes has exceeded its threshold) it permits a pulse from the trigger multivibrator to touch off the first delay multivibrator in the corresponding chain. This signal travels down the line of multivibrators, triggering each one in turn. As each multivibrator is triggered, it also sends a signal to the main plug board (shown in Fig. 3) where it can be connected to any combination of $A^{(2)}$ units. Several signals may be travelling down a single multivibrator chain at any one time, representing the time-pattern of frequency components of the appropriate type. The delay for each multivibrator can be separately adjusted, over a range from 10 to

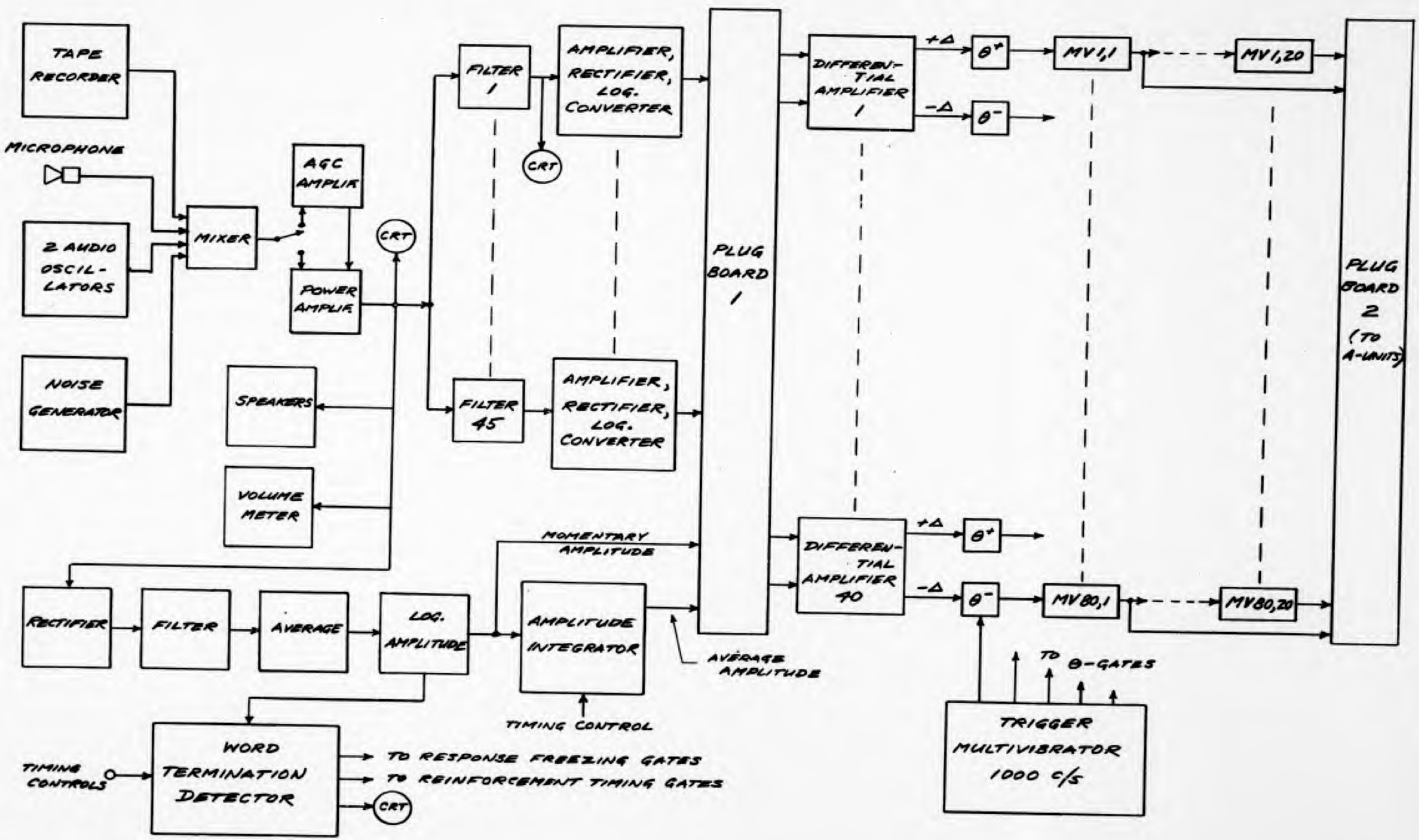


FIGURE 4: TOBERMORY SENSORY SYSTEM, DIFFERENCE UNITS, AND DELAY NETWORK

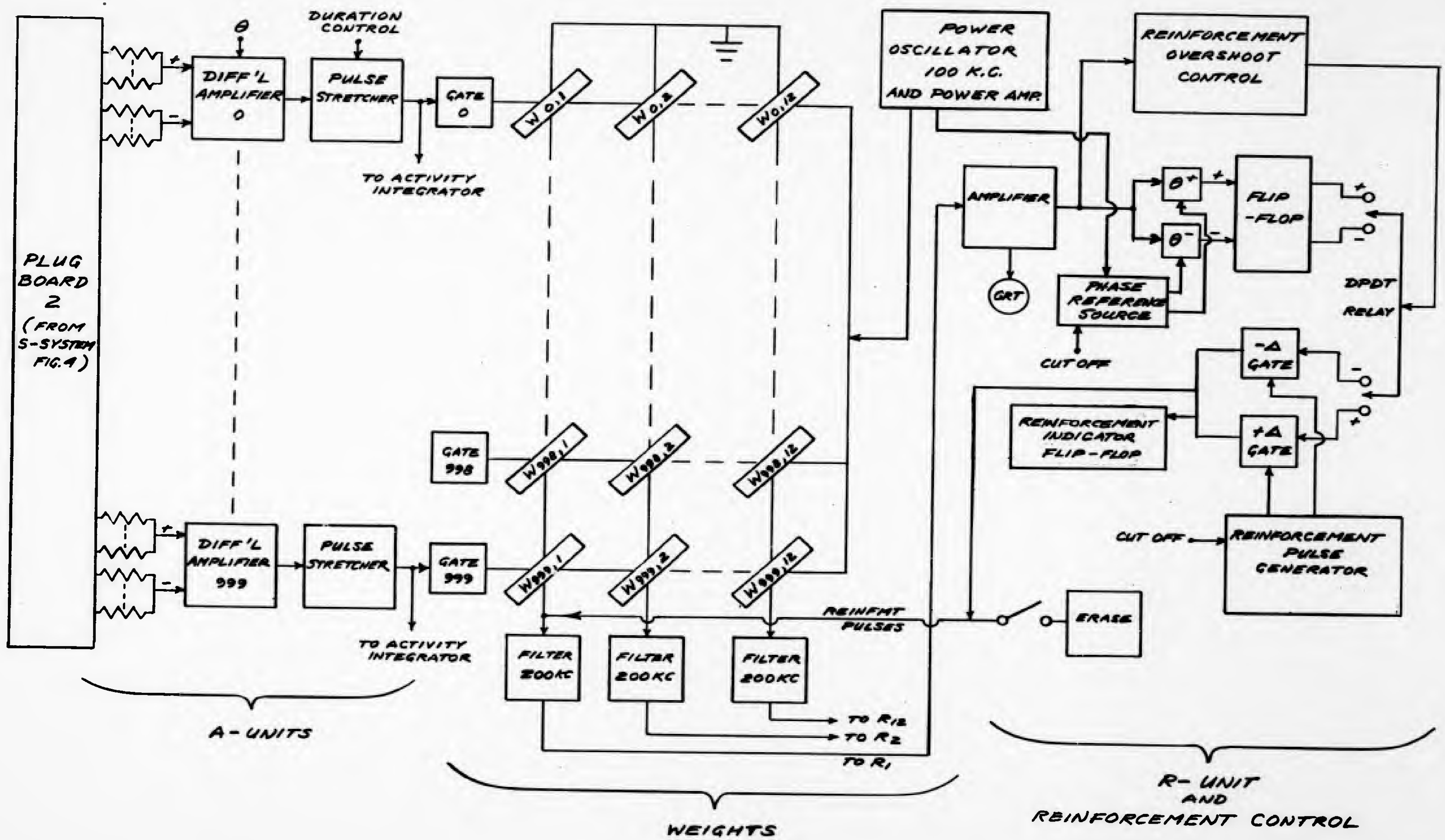


FIGURE 5: A-UNITS, WEIGHTED CONNECTIONS, AND R-UNITS.

100 milliseconds; thus the entire audio pattern represented at one time by the sensory system may cover any period from 0.2 to 2.0 seconds.

3. A-units, R-units, and Memory

The $A^{(1)}$ units have already been described above; these are simply the differential amplifiers which feed the delay system. The main Tobermory plug board contains ten excitatory input hubs and ten inhibitory input hubs for each $A^{(2)}$ unit. These inputs may be wired to any of the 1600 sensory points (the outputs of the delay multivibrators) each of which is represented by thirty hubs in the plug board. The organization of these A-units, their weighted connections to the R-units, and the R-units themselves, are shown in Fig. 5.

The A-units themselves consist of a differential amplifier, with an adjustable threshold, and a pulse stretcher, which produces an output pulse of a duration which can be adjusted between 3 and 50 milliseconds. The threshold and timing controls can be set independently for each A-unit. The A-unit outputs are accumulated by an activity integrator, to provide a display of the level of A-unit activity on the CRT in the control room. This integrator may also be used in the future to provide a servo control for the A-unit thresholds, to maintain a constant level of activity in the system.

The memory consists of 12,000 tape wound cores, arranged in a 12 by 1000 matrix, as shown in Fig. 5. When an A-unit is active, it opens a gate, permitting a 100 kc signal from the power oscillator to pass through the drive winding of the twelve cores to which the unit is connected. The signal from all active A-units in a column of the matrix is then summed by the output windings, and the second harmonic is extracted by the 200 kc bandpass filter for each R-unit. This second harmonic signal will agree in phase with the sign of the stored weights, and will agree in amplitude with the magnitude of the weights. Writing is done by gating 0.3 millisecond pulses to the same windings as are used for reading, and erasure (which can be performed independently for each R-unit) is done by switching an AC signal to these same windings. This matrix arrangement makes individual gates or switches for each integrator unnecessary, and provides a highly economical arrangement for a large matrix. A module of 600 of these integrators is shown in Fig. 3.

The sum of the signals from each column of the weight matrix goes to the input amplifier of the corresponding R-unit. The output of this amplifier can be displayed on the CRT, for monitoring of reinforcement effects. This amplifier can be cut off by the response freezing signal from the word termination detector (see Fig. 4), to prevent further changes in the state of the R-unit after an input message has been completed. The "positive phase" and "negative phase" are distinguished by comparison with a phase reference, and a separate positive and negative threshold (θ^+ and θ^-) is provided for each R-unit. If a signal of magnitude greater than either of these thresholds is received, the flip-flop, which records the state of the R-unit, is either set or reset. Thus, the R-unit may "change its mind" several times before the response is finally frozen. The output of the flip-flop is displayed in the control room, and may be printed by an automatic typewriter after each word.

Reinforcement is generally performed by an error correction procedure, which requires that the desired response must be set up before a word is presented. This desired response is registered by 12 DPDT relays such as the one in Fig. 5. As can be seen from the figure, if the desired response is positive, and the obtained response is negative, the positive pulse gate from the reinforcement pulse generator is opened. If the desired response is negative, and the obtained response is positive, the negative pulse gate is opened. The reinforcement pulse generator is turned on only during the period for which reinforcement is permitted, as determined by the word termination detector. This period will generally be set for a few milliseconds after the entire word is "in register", and overlapping the point at which the response state is frozen. The reinforcement overshoot control circuit is used to guarantee a brief continuation of reinforcement beyond the time at which a corrected response resets the flip-flop. The amount of this overshoot is adjustable for each R-unit. The circuit is designed to sense the input signal to the R-unit, and to hold the appropriate reinforcement gate open until its own threshold has been reached.

A sketch of the complete Tobermory system is shown in Fig. 6. The control room will be a soundproof room, with an automatic typewriter for construction

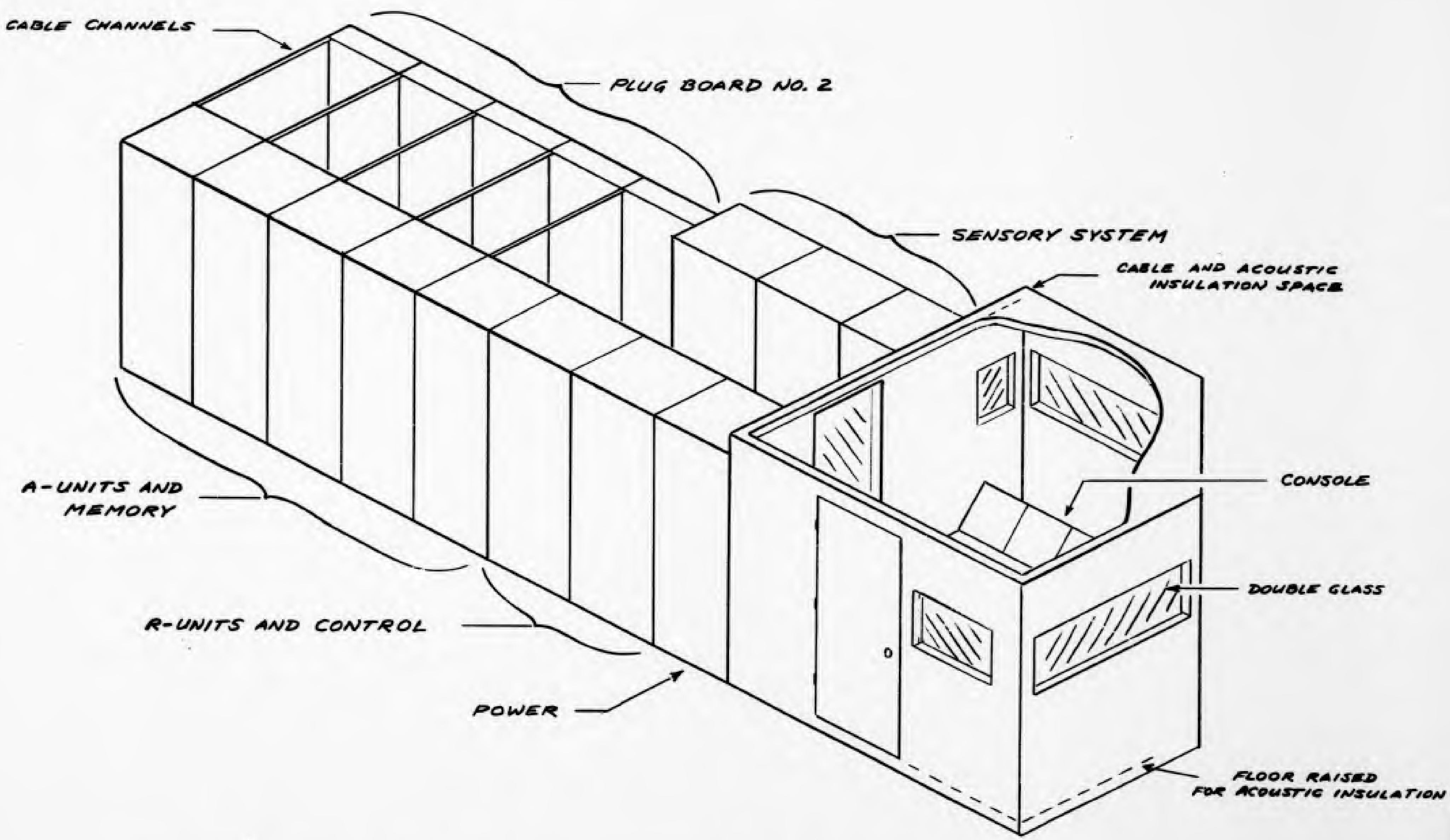


FIGURE 6: ISOMETRIC VIEW OF TOBERMORY SYSTEM

of program tapes and printing of output information. Manual control of all phases of operation is also possible, but it is expected that for most training routines, a tape controlled mode will be employed. The tape has two channels, one for the audio signals which the machine is to recognize, and the other for command codes which set up the desired response, and control the printing of output information between words.

It is hoped that the auditory system will be functional in the fall of 1963, at which time work will be initiated on the visual system, as shown in Fig. 2.

REFERENCES

1. Crafts, H.J. Components that learn and how to use them. Electronics, March 22, 1963, 49-53.
2. Crafts, H.J. A Magnetic Variable-Gain Component for Adaptive Networks Stanford Electronics Laboratories Tech. Report No. 1851-2, December, 1962.
3. Hay, J.C., Martin, F.C., and Wightman, C.W., The Mark I Perceptron, Design and Performance, Record of IRE National Convention, Part 2, New York, 1960.
4. Nagy, G. Analogue Memory Mechanisms for Neural Nets, CSRP Report No.3, Cornell University, 1962. Also, PhD thesis, Cornell University, 1962.
5. Nagy, G. Analogue Information Storage Elements, in Collected Technical Papers, Vol. 2, CSRP Report No. 4, Cornell University, 1963.
6. Rosenblatt, F. Applications of Perceptrons to Speech Recognition, in Proceedings of Stockholm Speech Communication Seminar, Royal Institute of Technology, Stockholm, 1963.
7. Rosenblatt, F. Principles of Neurodynamics; Perceptrons and the Theory of Brain Mechanisms, Spartan Books, Washington, D.C., 1962.
8. Rosenblatt, F. A Comparison of Several Perceptron Models, in Yovits, Jacobi, and Goldstein (Eds.) Self-Organizing Systems - 1962, Spartan Books, Washington, 1962.
9. Rosenblatt, F. A Model for Experiential Storage in Neural Networks, in Proceedings of COINS Symposium, Northwestern University, Spartan Books, 1963 (Awaiting Publication).

ANALOGUE INFORMATION STORAGE ELEMENTS

George Nagy

Cornell University, Ithaca, New York

Summary. Widespread and persistent interest in the implementation of multilinear logic, conditional probability computers, learning machines, and brain models has created a need for an inexpensive analogue or quasi-digital storage element. A number of possible approaches to this problem, ranging from the slow and reliable electromechanical systems to the many forms of charge and flux integration, are reviewed, and the suitability of each device for various fields of application is briefly discussed.

Introduction

During the course of the last decade several investigators have voiced the opinion that a workable solution to the problem of pattern recognition will have to be found before more sophisticated problems in the general field of "artificial intelligence" may be effectually tackled. The subfield of pattern recognition may be further subdivided, somewhat arbitrarily, into two schools of thought: the logic tree faction and the nerve net faction. The logic tree approach is based on the common binary operations, and is applicable chiefly to cases where the characteristics of all the patterns to be presented to the pattern recognition machine are specified ahead of time in full detail. This scheme exhibits complete structural staticity, and may be easily implemented with existing digital hardware. The nerve nets, on the other hand, often incorporate a number of variable, weighted connections, whose levels must be set during the course of a training routine to correspond to the probability density functions characterizing the input or the system structure. The object of this paper is to provide a comprehensive review of the kinds of memory elements which may be useful in implementing the information storage system of such nets.

The striking feature common to all these nets, be they part of a conditional probability computer, a pattern recognition device designed to classify bubble chamber photographs, or a model of the cat's visual cortex, is that the degree of interest and usefulness of the performance displayed increases with the number of variable strength links embodied in the system. Consequently, low unit cost is often the overriding consideration in choosing a storage device for a particular machine.

Fortunately, the logical design of most pattern recognition devices does not impose too stringent requirements on the performance of individual memory cells. As a rule, the outputs of a large number of weighted connections are added, and the correct classification of the input signal depends on whether the algebraic sum is greater or less than a given threshold (threshold logic), or greater or less than other similarly constituted sums (majority logic).

If only a coarse level setting arrangement is available, or if only a finite number of levels may be obtained, a well designed system will still converge to a solution, although a longer training sequence (adaptive period), or more adaptive links, may be necessary. Nor will malfunctioning of a few units crucially affect overall performance.

What is needed, then, is a low cost device with a range of about 30 to 40 discrete levels; a much higher resolution would be wasted unless the signal to noise ratio of the input patterns was extremely high. It should be possible to increment or decrement the value one level at a time, but it is not necessary to be able to reproduce a given level exactly, as it would be if the machine were to perform arithmetic operations. A way of resetting the value to some arbitrary zero level would also be desirable, in order to erase traces of previous learning before starting on a new problem.

The required speed of operation is more closely dependent on the particular application in view. For example, in a constantly updated radar monitor a faster reinforcement rate would be advantageous than in a character recognition device which would be retrained comparatively rarely. A speech processor operating in real time would likely have to be faster than the visual models built to date. Sophisticated configurations incorporating feedback, reverberating loops, and built-in decay,[16] would be able to take advantage of higher rates than a series-coupled system paced by the speed of the input equipment. Furthermore, one attempt at least has already been made to couple directly a parallel device of the nerve net type to a conventional sequential digital computer.[18]

The degree of permanence required may also vary from a few hours in a laboratory machine designed to check system performance to several months in an automatic page reader which would normally be left alone as long as radical changes in the type present in its input did not occur.

This completes the list of features which may be relevant in selecting a memory component. Let us now see where we may hope to find a device satisfying our rather modest demands.

State of the Art

1. Electromechanical Memory Elements. As had been the case with both digital and analogue computers, the first large scale parallel pattern recognition machine made extensive use of electromechanical elements. The

Mark I perceptron,[10] built at the Cornell Aeronautical Laboratory, includes 512 gear head direct current motors to actuate the potentiometers representing the values of its weighted connections. While in a larger system cost and space requirements could be cut down considerably by mounting a number of potentiometers -- possibly as many as forty or fifty -- on a common shaft, and clamping appropriate ratio arms to the rotating shaft by means of a magnetic clutch arrangement,[13] on the whole electro mechanical elements are obsolete for parallel storage.

2. Thermistors. The currents which flow through a thermistor raise its temperature through ohmic dissipation, and the temperature characteristics of the device are such that its conductance is thereby increased. Thus thermistors are ideally suited for the type of training required in certain four layer and cross-coupled systems:[4] only links originating at "active" signal generating units carry current, and are consequently reinforced. Unfortunately reinforcement is strictly monopolar. A further drawback is the short "half-memory" of thermistors: it is only of the order of three or four minutes.[1]

3. Photochromic Storage Devices. The characteristic curves of the photochromic or phototropic film on which these devices are based are displayed in Figure 1. The unique property of this film is that its transmittance near the center of the visible spectrum may be reversibly altered by exposure to high intensity radiation in the borderline regions. Curve A shows the transmittance of the film after it has been exposed to a flash of light of the spectral composition indicated by curve C (yellow filter), while curve B shows transmittance after an "erase" pulse through blue filter D. Curve E describes the "read" filter which has been found to

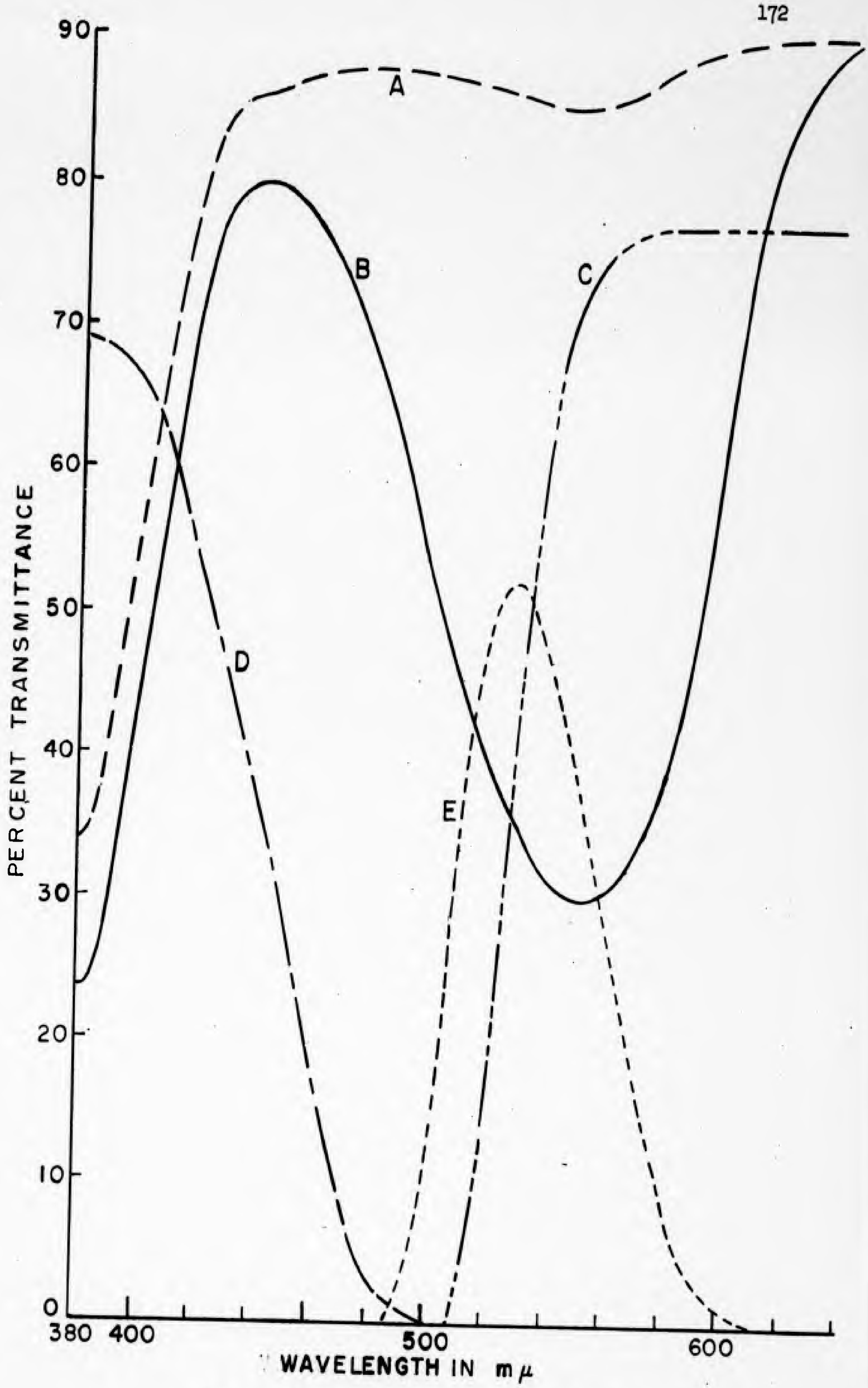


Figure 1. Photochromic Characteristics

interfere least with the condition of the film. Reading is, of course, performed at relatively much lower intensities than reinforcement; the overall transmissivity varies from about .3 to .8. The material is not too unstable: at room temperature it decays towards an equilibrium point with a time constant of several hours.

A rather elegant photochromic device, designed by Scott H. Cameron,[7] consists of a modified automatic slide projector. The input patterns, in the form of transparencies, are introduced into a collimator in front of the photochromic film. A photoresistor behind the film measures the total amount of light transmitted during a "read" cycle. If the intensity exceeds a preset threshold and the input pattern is to be classified in the "positive" class, a flash of "blue" light is triggered. If the input is to be classified "negative", and the intensity fails to reach threshold level, the film is flooded with "yellow" light. It may be shown that iteration of this procedure will enable the machine to form dichotomies among a broad class of patterns: in fact, the machine is designed to run through its magazine of slides until it stops making mistakes.

Photochromic devices are proving their usefulness in the two-layer processing of visual data, but there appears to be no simple way of adapting them to more complicated topologies.

4. Charge Integration. The engineer's concept of a circuit with a memory generally involves one or more charged capacitors, so it is reasonable to investigate whether these hold out any promise for present nerve net applications. Charge integration in capacitors presents the problem that for linear operation a constant current source, implying in practice a large series resistance, is required. This in turn renders incrementation

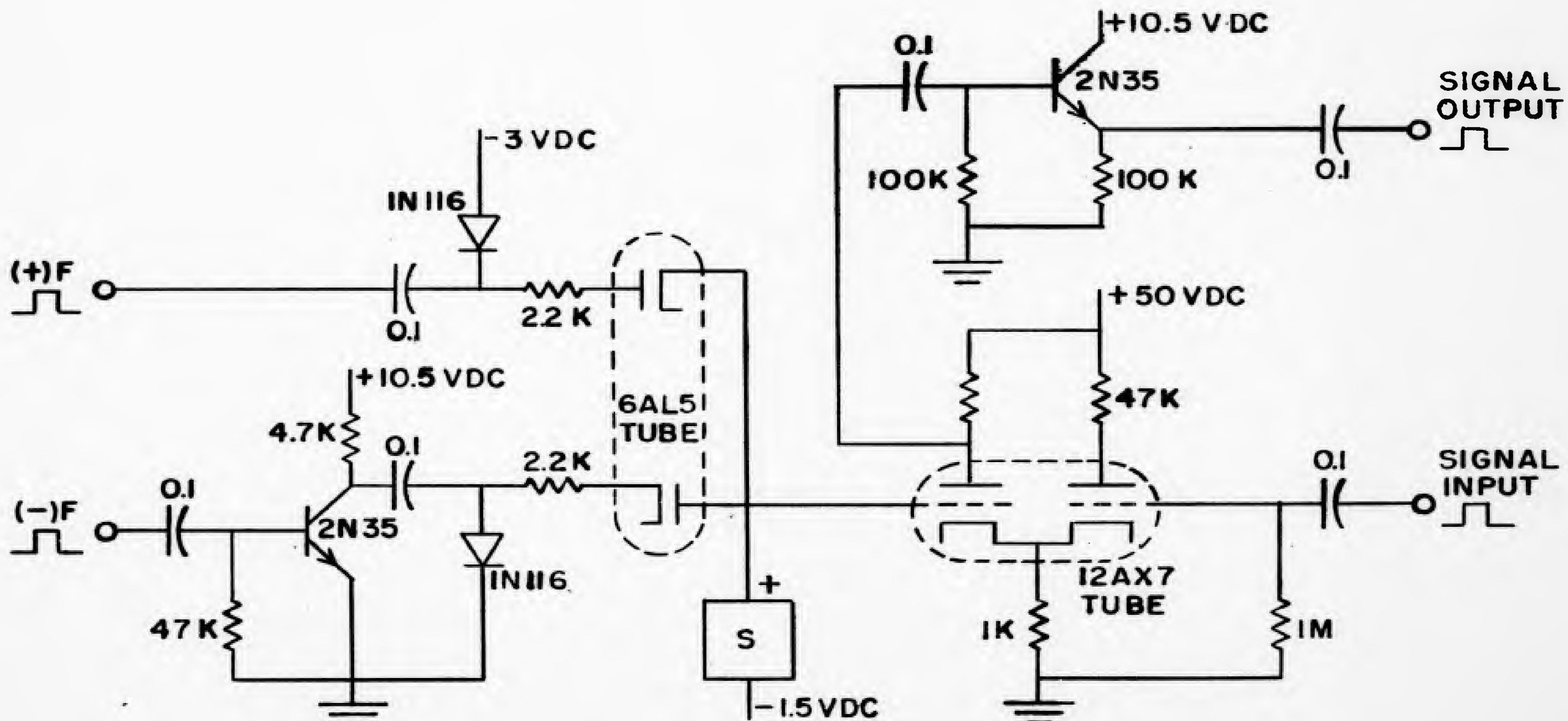


Figure 2. Babcock's "Refined Facilitator"

intolerably slow, since for reasonable storage times large capacitance values are necessary. The sensing of the charge offers a further problem.

These difficulties are largely overcome by Babcock's "Facilitator,"[2] shown in Figure 2. The facilitator is rather too expensive to use in conventional pattern recognition machines, and should be of interest chiefly to specialists in complex neuron interactions.

5. Solions. Solions[21] is the generic name of a family of amplifying devices which function by controlling and monitoring a reversible electrochemical reaction.

The reaction utilized in solions is a so-called "redox" reaction in which oxidation and reduction take place in turn. In the solion tetrode four inert electrodes are immersed in an electrolyte containing both the oxidized and the reduced species of an ion, and by controlling the charge transferred between the two input electrodes, a change in conductivity proportional to the input current may be obtained between the output electrodes.

Figure 3 is a simplified diagram of a solion tetrode connected as an integrator. The electrolyte used is an aqueous solution containing a small amount of iodine and a comparatively larger amount of potassium iodide. The amount of tri-iodide (resulting from the dissociation of the iodine in the presence of the potassium iodide) transferred from the Reservoir to the Integral Compartment by the input current is, by Faraday's Law, proportional to its integral with respect to time. The output current is proportional to the concentration of tri-iodide in the Integral Compartment, and hence to the integral of the input current. The polarized shield merely serves to reduce the tri-iodide concentration near it to the point where diffusion through the small perforations of the electrode is negligible.

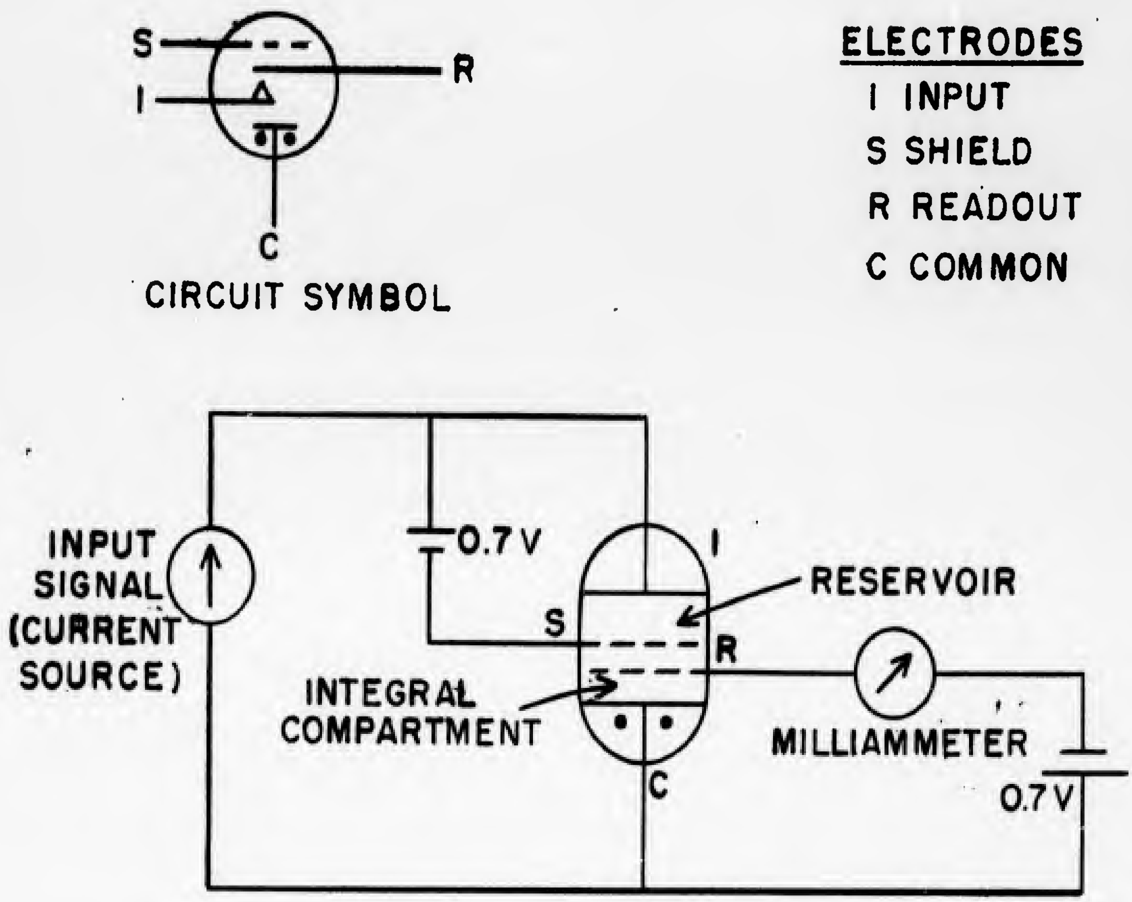


Figure 3. Solion Tetrode Connected as an Integrator

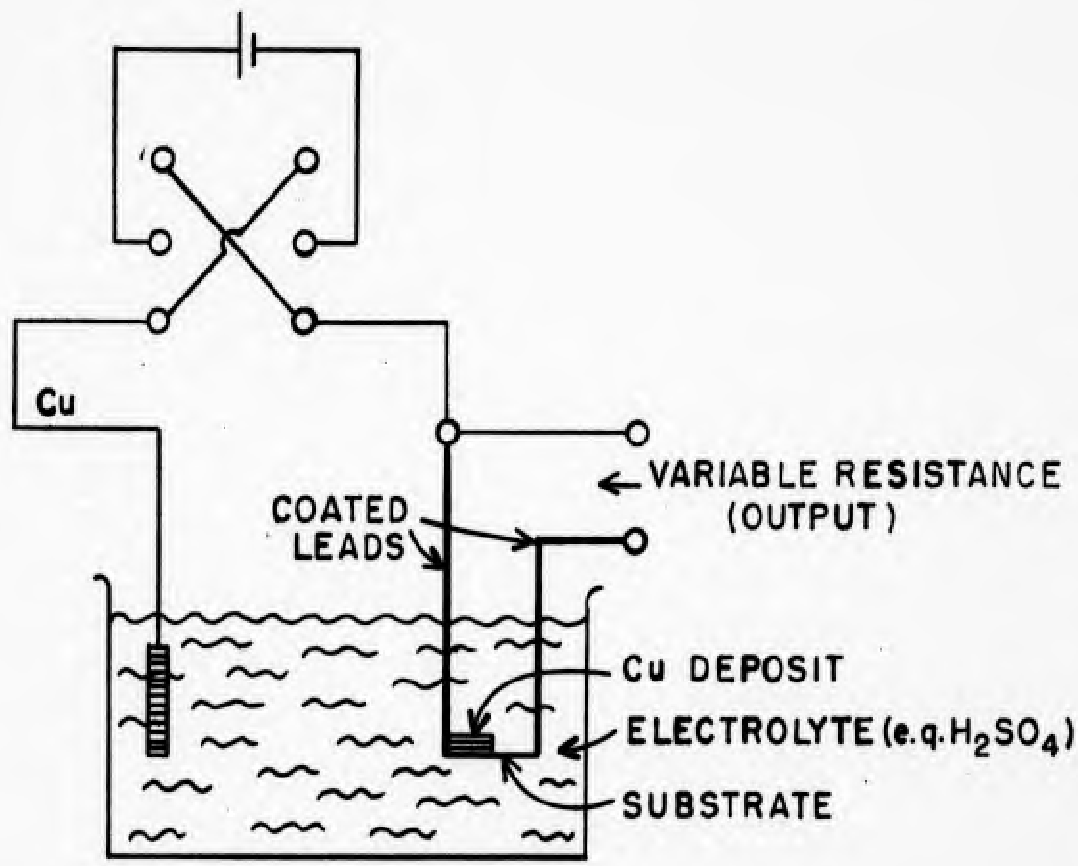


Figure 4. Schematic of Electrolytic Integrator

Because of the concentration potential resulting from the difference in ion concentration in the vicinity of the two input electrodes, the input impedance of the solion tetrode varies by a ratio of 5:1 in the operating range of the device. An even more serious drawback is the low output impedance of the device, which causes units connected together for the purpose of summing operations to discharge through one another. These difficulties are similar to those encountered with capacitors, though the time constant is greatly magnified in solions by the use of a liquid medium.

At constant temperature, the stability of isolated solions is reported to be excellent, with drifts of only a fraction of 1% over periods of several days. Reasonably high packing densities may already be achieved -- the volume of a tetrode now on the market is approximately .2" cu. in. -- but prices are still rather high. If solions are to be seriously considered for embodiment in large nerve nets, considerable redesigning is required, with the emphasis shifted from precision to ease of mass production.

6. Electrolytic Integrators. Yet another form of charge integration is exhibited in the "electrolytic integrator." The basic principle behind it is so simple that it had occurred to practically every investigator in need of a cheap and reversible memory device, but the first really workable device was developed by B. Widrow[19] using high precision electro-chemical techniques and, it is said, a number of incantations devised originally in connection with the touch-stone research program.

The electrolytic integrator, in its basic form, consists of two electrodes immersed in an electrolyte (see Figure 4) in such a way that it is possible to vary their resistance relative to one another by transferring metal in ionized form through the solution. In practice, one of the electrodes, the variable element, is a fairly high resistance conductor with two terminals accessible in order to detect resistance changes. The other electrode, the source, is simply a bar of metal.

The basic resistance of the variable element must lie in a relatively narrow range. If the basic resistance is too low compared to the resistance of the source metal, then in order to produce a detectable resistance change, very large amounts of the source metal must be deposited on the surface of the variable element. Since the maximum permissible plating current is limited by the need for even plating action, low basic resistance entails inadmissibly slow integrating action. If, on the other hand, the basic resistance is high compared to that of the solution, then the resistance change measured at the terminals of the variable element will again be small, due to the constant low resistance of the solution which is essentially in parallel with it.

Commonly used substrates include metallic oxide films deposited on glass, graphite, and thin resistance wires. The electrolyte is usually a solution of copper sulphate in water, with various chemical agents added to regulate the Ph factor and insure even plating characteristics. Currents of the order of a milliampere and time constants of a few seconds are typical of the small airtight integrator capsules now available.

A slight variation on the electrode resistance integrator, investigated by H. Y. Chiu[8] and others, deserves mention. Chiu advocates the use of cells where the resistance between the electrodes, rather than that of one of the electrodes, is changed as the result of copper transfer. For example, the cathode of such a cell may be a cylinder of copper foil, while the anode would consist of a thin gold wire concentric with the cathode. The reversibility of a process based on such a geometry is questionable, and the data obtained by Chiu are not too encouraging.

The Optimistor[22] is yet another version of the electrolytic integrator. Here a very thin layer of metal is deposited on a transparent substrate, and the thickness of the layer is sampled with a light beam. Because of the relatively large ratio of exposed surface to amount of material present, this arrangement is rather unstable.

Experiments have also been conducted on transferring silver ions through a thin film of silver bromide.[12] The nonlinearity of this process renders it suitable for coincidence mode selection of the points to be incremented in a memory matrix. The matrix consists simply of the points of intersection of thin silver wires electrolytically coated with a 10 micron tungsten bromide film. When current is passed through a point of intersection, silver ions are released from the wire acting as the anode, transported across the bromide film, and deposited on the cathode wire. Eventually a bridge of silver is built up between the wires, and the resistance between them changes from about 1 megohm to less than 10 ohms. Here again, reversibility is the chief problem. A program to investigate solid solutions with a view to electrolytic integrator applications is reported to be under way at the Cornell Aeronautical Laboratory.

7. <u>The Transpolarizer</u>. The transpolarizer,[15] an electrostatic analogue of the more widely known transfluxor, consists of two capacitors with a crystalline ferroelectric dielectric and a nearly rectangular hysteresis loop. The basic circuit is shown in Figure 5, and the mode of operation is as follows.

One of the capacitors, say C_1, is maintained in a polarized state by means of a d.c. bias. Then the transpolarizer is said to be in the unblocked state if C_2 is polarized in the same direction as C_1. In this case the two capacitors in series behave essentially as a single ferroelectric element,

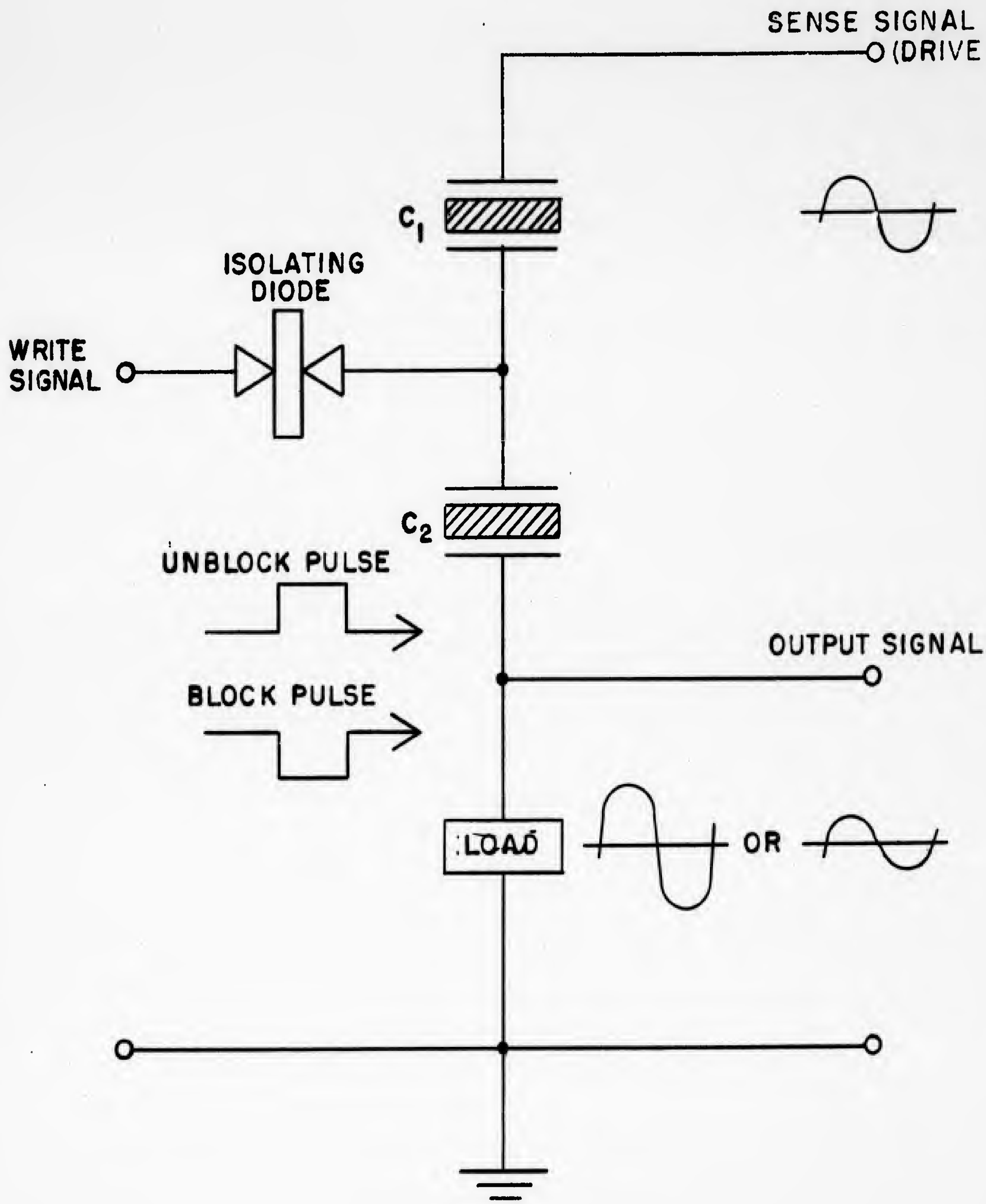


Figure 5. Basic Circuit of a Transpolarizer

and present a low impedance to a small a.c. sensing signal. If, however, C_2 is polarized oppositely to C_1, then any attempt to switch C_1 would result in C_2 being driven further into saturation. Hence no switching occurs, and the transpolarizer is said to be in the blocked state. The combination acts as a small linear capacitor, and therefore has a relatively high impedance at the driving frequency.

Making use of the partially blocked states of the transpolarizer, about thirty discrete and reproducible steps are attainable. The polarization is set to the desired level by 1 microsecond pulse of the appropriate polarity. With recently developed materials, such as tri-glycine sulfate (TGS) and tri-glycine fluoberyllate (TGFB), extremely stable operation may be expected, and sensing voltages several times as large as the coercive voltage may be safely applied.

8. Magnetic Flux Integration. Modelled on the core memories so widely used in digital computers, most flux integrators use the partial switching of the domains in a toroidal core under a current impulse as the basic increment. Differences between particular designs arise chiefly in the mode of non-destructive read-out employed.

One popular approach to the read-out problem makes use of quadrature fields. A weak "strobe" field is applied orthogonally to the "write" axis of magnetization; it causes the flux vector to rotate slightly, generating a voltage proportional to its rate of change (and hence its magnitude) in the read winding (which may be the same as the write winding). At the end of the strobe pulse, the flux vector springs back to its original preferred orientation by virtue of "domain elasticity." Figure 6 shows a core which was used in a small perceptron[11] utilizing this principle for information storage and modification.

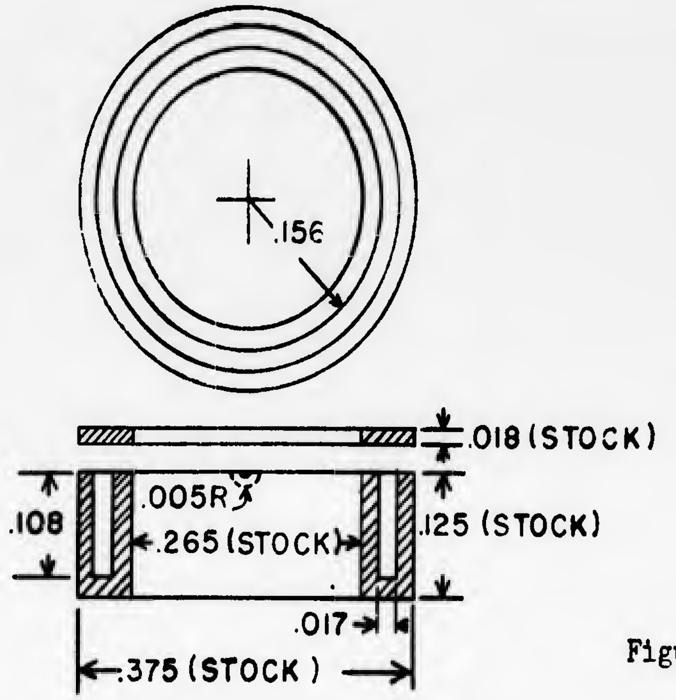


Figure 6. Aeronutronic Integrator Core and Steel Cover

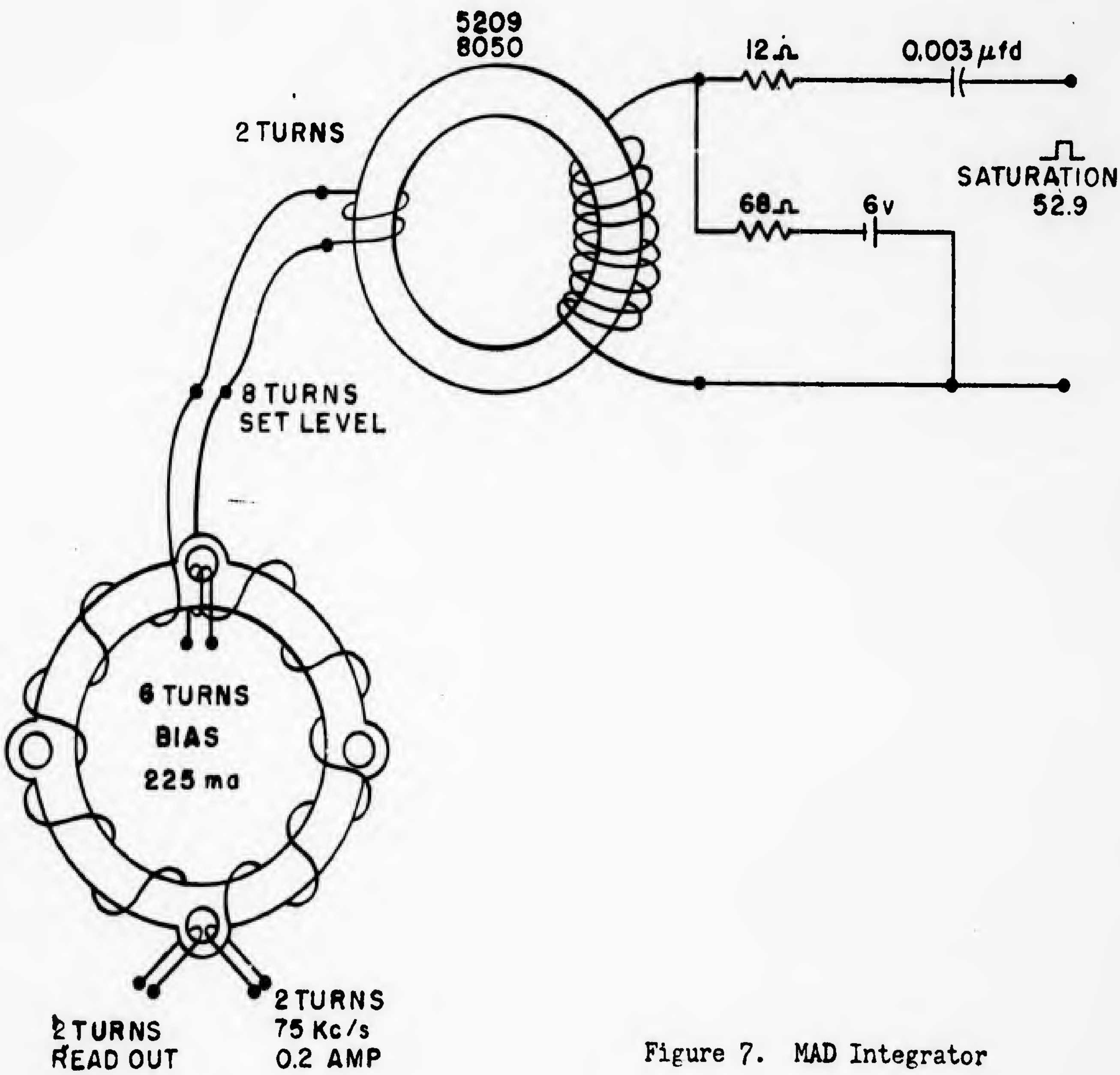


Figure 7. MAD Integrator

Another working analogue storage device has been developed at the Stanford Research Institute by A. Brain,[5] using the multiaperture device (MAD) derived earlier from the transfluxor configuration by Bennion and Crane.[3] A schematic of the MAD ferrite, as used for analogue storage, is shown in Figure 7. The current through the bias winding holds the core material near the inside perimeter of the core in a saturated condition, thus "trapping" any flux which may be present around the small apertures. Pulses through the set winding vary the total amount of flux in the core, with the amount of flux switched at each increment held constant by means of a "bucket" core.

A slightly different version of the toroidal flux integrator is now being patented by J. Divilbiss[2] of the University of Illinois. His device is shown in Figure 8. The novel feature here is the variable resistance short circuited loop, which controls the field available for flux switching.

A mode of readout first investigated by Widrow[18] has been recently perfected by H. J. Honerloh[12] at the Technische Hochschule in Karlsruhe, Germany, and is about to be incorporated into an 8,000 element "Lernmatrix." Here the readout signal is proportional to a difference frequency generated by core nonlinearity between two drive frequencies in the low broadcast range. The signal is very small, but the summing operation characteristic of parallel pattern recognition machines raises it above noise level. In the Lernmatrix the cores are simply threaded onto the read and write wires in the manner of conventional digital core memories, and incrementation takes place by coincidence mode switching.

Reversible flux switching in a tape-wound toroidal core, which takes place at harmonic frequencies of a drive current due to core non-linearity, has been successfully exploited for read-out purposes by H. S. Crafts.[22]

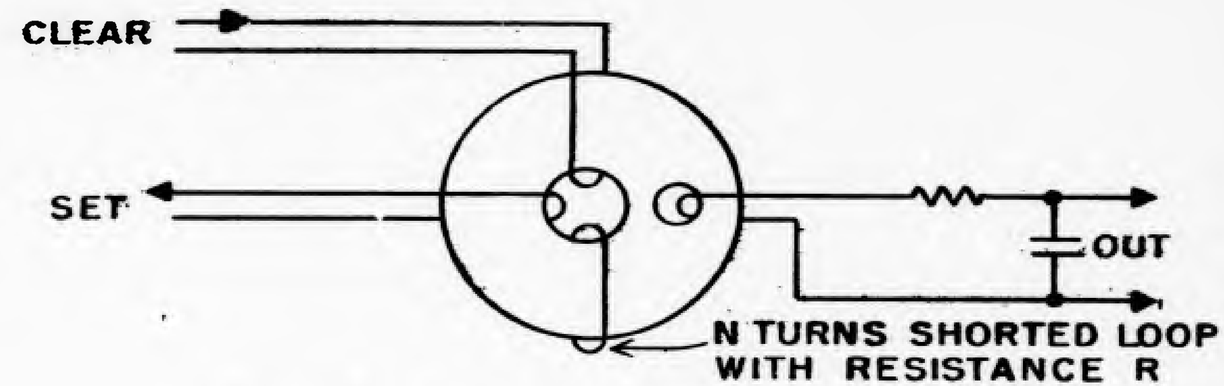


Figure 8. Divilbiss' Modified Transfluxor

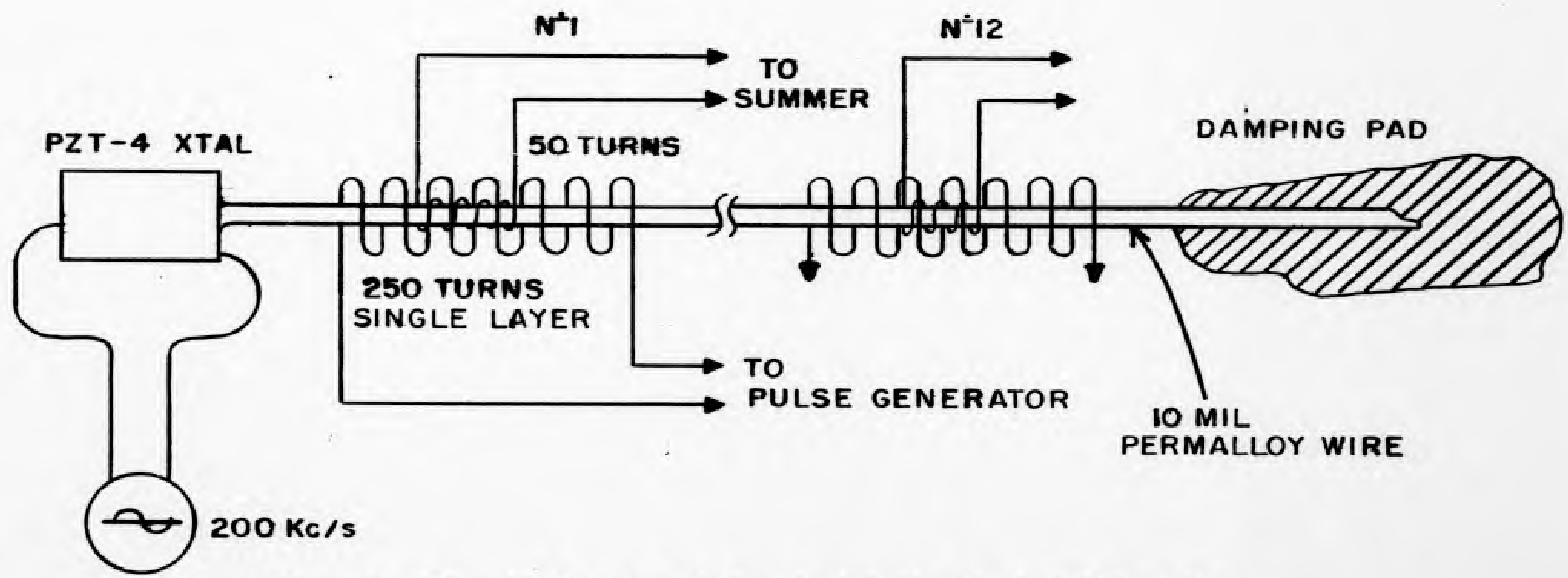


Figure 9. Schematic Diagram of Magnetostrictive Readout Integrator

An early version of the so-called second-harmonic integrator consists of two cores wound in such a manner that in the output winding the second-harmonic components, which are proportional to the flux level, add, while the fundamental cancels out. Because the high-level radio frequency drive effectively lowers the core coercivity, coincidence mode incrementation is readily practicable. Other features, such as relatively low cost, large number of levels available, and the simplicity of the required auxiliary circuitry, contribute to make this device a strong contender for a leading position among practical integrators.

We shall conclude this sampling of flux integration techniques with an idea originated by C. Rosen of the Stanford Research Institute. When a magnetostrictive element is acoustically excited, an alternating flux wave is generated whose magnitude is proportional to the initial magnetization of the element. A device based on this principle is illustrated on Figure 9: the flux carrying medium is the 10 mil permalloy wire, ultrasonically driven by a piezoelectric transducer. With this arrangement it is possible to obtain up to 80 discrete steps, using only a very modest amount of auxiliary circuitry. Further development work on this device is being carried on both at the Stanford Research Institute[6] and at Cornell University,[14] in conjunction with the construction of large scale visual and audio pattern recognition devices.

Conclusion

The list of analogue memory devices presented in the preceding paragraphs does not pretend to be exhaustive. One may gain some idea of the staggering variety of processes which may be potentially harnessed to fulfill nerve net memory functions by considering the number of physical phenomena characterized by a first order differential equation. Since the specifications for the simultaneous access storage in a pattern recognition computer are often rather loosely formulated, the choice between the various approaches available is not an easy one; a quantitative evaluation based on performance curves and cost would be a major research project in itself.

Among the alternatives presented, the various magnetic flux integrating devices offer perhaps the most flexibility for many applications. The popularity of the toroidal cores for analogue storage is bound to grow as improved fabricating methods and materials become available for thin film deposited cores, and for the application of printed circuit techniques to provide the necessary windings. Coincidence mode incrementation will no doubt be employed on all large scale machines.

It is also possible that improvements in electron-optical machining techniques, now being developed by K. Shoulders in Palo Alto,[17] may obviate the necessity for analogue storage. In principle, any pattern recognition machine using weighted connections may be simulated on a binary machine of sufficiently large capacity. With cryogenic storage elements, it may be possible to build large, fairly general purpose parallel computers, on which any specific connection scheme necessary for a given task may be established by external control.

Even the considerable improvements in components almost within reach are not likely to close the existing gap between proposed theoretical models and their hardware realization. As much as ever, it will remain up to the individual designer to maximize the yield of machines severely handicapped by the lack of a really cheap, reliable, and fast analogue memory component.

Bibliography

1. Arking, A., and Chiu, H.Y., Proposal for Construction of a Thermistor Perceptron, Laboratory of Nuclear Studies, Cornell University, Ithaca, 1959.

2. Babcock, M.L., Reorganization by Adaptive Automation, University of Illinois, Nonr 1834(21) Technical Report No. 1, Urbana, Illinois, January 1960.

3. Bennion, D.R. and Crane, H.D., "Design and Analysis of MAD Transfer Circuitry," Proc. W.J.C.C., 21-36, March 1959.

4. Block, H.D., Knight, B.W., Rosenblatt, F., Analysis of Four-Layer Series Coupled Perceptron, Review of Modern Physics, 1962, 34, 135-142.

5. Brain, A.E., Novikoff, A.B. and Bourne, C.P., Graphical Data Processing Research Study and Experimental Investigation, Report No. 4, Stanford Research Institute, Menlo Park, California, April 1961.

6. Brain, A.E. and Forsen, G.E., Graphical Data Processing Research, Study and Experimental Investigation, Report No. 7, March 1962, Stanford Research Institute, Menlo Park, California.

7. Cameron, S.H., Self Organizing Networks, Annual Report, Armour Research Foundation of Illinois Institute of Technology, Project No. E 154, Chicago, 1962.

8. Chiu, H.Y., An Investigation of the Possibility of Using Electrolytic Cells as A-Units in the Construction of a Perceptron, Laboratory for Nuclear Studies, Cornell University, Ithaca, 1959.

9. Divilbiss, J., A Magnetic Pulse Integrator, Disclosure of Invention and Letter of Transmittal, University of Illinois, Urbana, Ill., 1962.

10. Hay, J.C., Martin, F.C. and Wightman, C.W., "The Mark I Perceptron, Design and Performance," Record of the IRE National Convention, Part 2, New York, 1960.

11. Hawkins, J.K., Mansey, C.J. and Stafford, R.A., A Magnetic Integrator for the Perceptron Program, Summary Report, Aeronutronic Research Laboratory Publication No. U-1405, Newport Beach, California, 1961.

12. Hönerloh, H.J. and Kraft, H., "Technische Verwirklichung der Lernmatrix," in Lernende Automaten, H. Billing (Editor), R. Oldenbourg, Munich, 1962.

13. Minsky, M., Neural-Analog Networks and the Brain Model Problem (a thesis), Princeton University, 1954.

14. Nagy, G., Analogue Memory Mechanisms for Neural Nets, (a thesis) Cornell University, 1962.

15. Pulvari, P., "The Transpolarizer: An Electrostatically Controlled Circuit Impedance with Stored Setting," Proc. I.R.E. 47, 1117-1122, June 1959.

16. Rosenblatt, F., The Principles of Neurodynamics, Cornell Aeronautical Laboratory Report No. VG-1196-G-8, Buffalo, March 1961.

17. Shoulders, K.R., Research in Microelectronics Using Electron-Beam Activated Machining Techniques, Stanford Research Institute, Menlo Park, Cal. 1960.

18. Widrow, B., "A Radio-Frequency Non-Destructive Read-Out for Magnetic Core Memories," IRE Transactions on Electronic Computers, Vol. EC-3, pp. 12-15 (Dec. 1954).

19. Widrow, B., An Adaptive "Adaline" Neuron Using Chemical "Memistors," Solid State Electronics Laboratory Technical Report No. 1553-2, Stanford University, Stanford, October 1960.

20. Widrow, B. Mays, C.H., Project No. 1557-26 in Solid State Electronics Research, Consolidated Quarterly Report No. 14, Solid State Electronics Laboratory, Stanford Electronics Laboratories, Stanford University, Stanford, Cal.

21. An Introduction to Solions, Texas Research and Electronic Corporation, Dallas, June 1961.

22. Hoff, M.E., Crafts, H.S., and Angell, J.B., Components for Trainable Systems, Digest of Papers, 1962 International Solid State Circuits Conference.

A BIBLIOGRAPHY OF PERCEPTRON LITERATURE

By F. Rosenblatt

This bibliography represents a compilation of all technical writings which have come to our attention at the present time, which contribute to the theory, realization, or applications of perceptrons. Purely popular articles and news items have in general been omitted, except in a few cases where a serious attempt has been made to explain basic concepts. Likewise, a considerable amount of literature on closely related projects (such as Widrow's "Adaline" and "Madaline" systems, and Bledsoe and Browning's n-tuples) has been omitted, except where the connection with perceptron theory has been explicitly acknowledged by the authors. Reports and papers intended for limited circulation have been included only when the material presented is not published in more readily available sources. General reviews and other references which touch on perceptrons only incidentally are also omitted. Foreign language publications have been listed to the extent that they are known to us.

For work done prior to 1961, a comprehensive treatment can be found in Rosenblatt's Principles of Neurodynamics (Ref. 88). No comparable treatment of more recent work has yet been published.

1. Albert, A., Review of Rosenblatt, F., "Principles of Neurodynamics", Math. Reviews 24, 1962, No. B1682, p. 265.

2. Albert A., A Mathematical Theory of Pattern Recognition, Annals of Math. Stat. 34 (1963), 284-299.

3. Beckman Systems, Prop 1: A Program Optimizing Perceptron, Beckman Systems, California, Feb 17, 1961.

4. Block, H.D., Analysis of Perceptrons, Proc. Western Joint Computer Conf., 1961, 281-289.

5. Block, H.D., The Perceptron: A Model for Brain Function, Rev. Mod. Physics, 34, 1962, 123-135.

6. Block, H.D., Adaptive Neural Networks as Brain Models, Proc. of Symposium on Experimental Arithmetic, Amer. Math. Soc., Chicago, April 1962, (Awaiting publication).

7. Block, H.D., Knight, B.W., and Rosenblatt, F., Analysis of a Four-Layer, Series-Coupled Perceptron, Rev. Mod. Physics, 34, 1962, 135-142.

8. Block, H.D., Nilsson, N.J., and Duda, R.O., Feature Detection. In Proceedings of COINS Symposium, Spartan Books, Washington, 1963 (In press).

9. Brain, A.E., The Simulation of Neural Elements by Electrical Networks Based on Multi-Aperture Magnetic Cores, Proc. IRE, 49, 1961, 49-52.

10. Brain, A.E., et. al., Graphical Data Processing Research Study and Experimental Investigation, Quarterly Progress Reports 1-10, Covering period from 1 April, 1960 to 30 November, 1962. Stanford Research Institute, Menlo Park, Calif.

11. Brain, A.E., Forsen, G.E., Nilsson, N.J. and Rosen, C.A., Learning Machines, International Science and Technology, Nov. 1962, pp. 20-30.

12. Braverman, E.M., Some Design Problems of Machines which Classify Elements According to Signs Previously Unknown, Avtomatika i Telemekhanika, No. 10, 1960, pp. 1375-1386. [In Russian].

13. Brown, F.M., A Contribution to the Theory of the Rosenblatt Perceptron, Wright Air Development Division, July 11, 1960.

14. Bryan, J.S., Experiments in Adaptive Pattern Recognition, Bionics Symposium (1963), Air Force Systems Command, Wright-Patterson Air Force Base, Ohio.

15. Cameron, S., Self Organizing Networks, (Annual Report) Armour Research Foundation, Chicago, 1962.

16. Charnes, Abraham, On Some Fundamental Theorems of Perceptron Theory and their Geometry, Notices of the AMS, Vol. 10, No. 1, Part 1, Issue No. 65, Jan. 1963, p. 60. Also, awaiting publication in Proceedings of COINS Symposium, Spartan Books, Washington.

17. Cowan, J.D., Reviews of Block, H.D., "The Perceptron: A Model for Brain Functioning" (Rev. Mod. Phys.), Block, Knight, and Rosenblatt, "Analysis of a Four-Layer Series Coupled Perceptron" (Rev. Mod. Phys.), and Rosenblatt, F., "Strategic Approaches to the Study of Brain Models" (Illinois Symposium) Math Reviews, 25, 1963, Nos. 1057, 1058, and 1059, pp. 209-210.

18. Crafts, H.J., A Magnetic Variable-Gain Component for Adaptive Networks, Stanford Electronics Labs. Tech. Report No. 1851-2, December, 1962.

19. Daly, J., Joseph, R.D., and Kelly, P.M., Self-Organizing Logical Systems, Astropower, Inc., Costa Mesa, Calif., 1962.

20. Gamba, A., Optimum Performance of Learning Machines, Proc IRE, 49, 1961, 349-350.

21. Gamba, A., Supplement of Nuovo Cimento, Vol. 20, 1961, pp. 112-146, 221-232.

22. Geoffrion, A.M., The Origin and Significance of Perceptron Theory, J. of Industrial Engineering, 12, 1961, 339-405.

23. Glushkov, V.M., Theory of Instruction for One Class of Discrete Perceptrons, Zhurnal Vychisl. Matematiki i Matematich. Fiziki, No. 2, 1962, pp. 317-335. [In Russian].

24. Goldstein, G. (Editor), Perceptron Mark I, Digital Computer Newsletter, ONR, July, 1960, Vol. 12, No. 3, 1-4.

25. Good, I.J., Speculations on Perceptrons and Other Automata, IBM Res. Center, Guest Lecture, RC-115, 1959.

26. Harth, E.M., Random Network Approach to Mental Processes, Dept. of Physics, Syracuse University, 1963.

27. Hay, J.C., Lynch, B.E., Smith, D.R., and Murray, A.E., Mark I Perceptron Operators' Manual, Cornell Aero. Lab Report No. VG-1196-G-5, Buffalo, 1960.

28. Hay, J.C., Martin, F.C., and Wightman, C.W., The Mark I Perceptron, Design and Performance, Record of IRE National Convention, Part 2, New York, 1960.

29. Hay, J.C. and Wightman, C.W., The Mark I Perceptron, Research Trends, Cornell Aero. Lab., 1960, 8, No. 1, 1-4.

30. Hawkins, J.K., A Magnetic Integrator for the Perceptron Program, IRE National Convention Record, Part 2, 1960, 88-95.

31. Hawkins, J.K., Self-Organizing Systems - A Review and Commentary, Proc IRE, 49, 1961, 31-48.

32. Hawkins, J.K., and Munsey, C.J., A Magnetic Integrator for the Perceptron Program, Quarterly Progress Reports, Beginning July 1960, and Summary Report, 29 Sept., 1961.

33. Hawkins, J.K., Munsey, C.J., and Stafford, R.A., Research on Biax Type Elements and Associated Circuits (Biax Perceptron), Annual Summary Report, Aeronutronic Division, Ford Motor Co., Jan., 1963.

34. Hoffman, A., "The Whirling Dervish", A Simulation Study in Learning and Recognition Systems, IRE Convention Record, 1962, (Part 4), 153-160.

35. Holmes, W.S., Richmond, G.E., Leland H.R., and Spooner, M.G., Status and Planning Report on Perceptron Applicability to Photointerpretation, Cornell Aero. Lab. Report No. VE-1446-G-2, August, 1961.

36. Holmes, W.S., Leland, H.R., and Muerle, J.L., Recognition of Mixed-Font Imperfect Characters, in Fischer, Pollock, Radack, and Stevens (Eds.) Optical Character Recognition, Spartan Books, Washington DC, 1962.

37. Isaacs, S.S., The Perceptron Simulation Project, AML Report No. 126, Dec 1960, David Taylor Model Basin, Washington D.C.

38. Ivakhnenko, A.G., On the Application of the Theory of Transmission and Combined Control to the Synthesis and Analysis of Learning Systems, Avtomatika, (U.S.S.R.), 1961, No. 5, pp. 3-12. [In Ukrainian.]

39. Ivakhnenko, A.G., Self-Organizing Systems with Positive Feedbacks, Avtomatika, 1962, 33-50. [In Ukrainian.]

40. Ivakhnenko, A.G., The Advantage of One-Layer Cognitive Learning Systems, Avtomatika 7 (1962), No. 6, 10-19. [In Ukrainian.]

41. Ivakhnenko, A.G., Can a Cognitive Self-Learning System Discriminate the Moment of the Transformation of a Tadpole Into a Full-Fledged Frog? Avtomatika, 1963, No. 2, 31-40. [In Ukrainian.]

42. Ivakhnenko, A.G., Kleschkov, V.V., Otkhmezuri, G.L., and Shlesinger, M.I., Fundamental Monograph on the Theory of Perceptrons (Review of Rosenblatt's "Principles of Neurodynamics"), Avtomatika, 1963, No. 3, 84-90.

43. Joseph, R.D., On Predicting Perceptron Performance, Record of IRE National Convention, Part 2, New York, 1960.

44. Joseph, R.D., Contributions to Perceptron Theory, CAL Report VG-1196-G-7, Buffalo, 1960. Also Ph.D. Thesis, Cornell University.

45. Joseph, R.D., Two Theorems on Error Correction, Project PARA Tech. Memo No. 17, Cornell Aero. Lab., Buffalo, 1960.

46. Joseph, R.D., Kelly, P.M. and Viglione, S.S., An Optical Decision Filter, Astropower, Inc., Costa Mesa, Calif., July 1962, (Also Proceedings of the IEEE, 51, 1963, 1098-1118.)

47. Joseph, R.D. and Kelly, P.M., Predicting Machine Size for Self-Organizing Systems, Astropower, Inc., Costa Mesa, Calif., June 1962.

48. Joseph, R.D., Viglione, S.S., and Wolf, H.F., Application of Distributed Memory Systems to Space Navigation, Astropower, Inc., Report No. EL-6288, Oct. 1962.

49. Kac, M., A Note on Learning Signal Detection, IRE Trans. on Inf. Theory, IT-8 No. 1, 1962.

50. Keller, H., Finite Automata, Pattern Recognition, and Perceptrons, Report of AEC Computing and Applied Math Center, N.Y.U., New York, March 1, 1960. Published in J. of Assoc. of Computing Mach., Jan. 1961.

51. Kelly, P.M., Problems in Bio-Computer Design, Bionics Symposium, WADD Technical Report 60-600, 1960, pp. 215-237.

52. Kesler, C., Preliminary Experiments on Perceptron Application to Bubble Chamber Event Recognition, in Collected Technical Papers Vol. 1, CSRP Report No. 1, Cornell University, 1 Jan, 1961.

53. Kesler, C., Analysis and Simulations of a Nerve Cell Model, CSRP Rept. No. 2, Cornell University, 1 May 1961. Also MA thesis, Cornell, 1961.

54. Klass, P.J., Perceptron Shows its Ability to Learn, Aviation Week, 73, 1960, pp. 72-73, 75-77, 80.

55. Klass, P.J., Perceptron Tested for Photo Analysis, Aviation Week, 74, 1961, pp. 69-70, 73.

56. Konheim, A.G., A General Convergence Theorem for the Four Layer Series-Coupled Perceptron, IBM Research Report RC-461, March 17, 1961.

57. Konheim, A.G., A Geometric Convergence Theorem for the Perceptron, IBM Research Paper RC-621, Feb., 1962, To appear in J. of Soc. for Industrial and Applied Math.

58. Konheim, A.G., A New Class of Multi-Layer Series-Coupled Perceptrons, in Yovits, Jacobi, and Goldstein (Eds.) Self-Organizing Systems-1962, Spartan Books, Washington DC, 1962.

59. Kono, R., Review of Perceptron Studies at CAL, Mitsubishi Elec. Mfg. Co., Japan, 1962. [In Japanese.]

60. Kozybovsky, C.F., Analysis of Perceptrons, (Ukranian adaptation of H.D. Block, "Analysis of Perceptrons", W.J.C.C., 1961), Avtomatika, 1962, 91-97.

61. Minsky, M.L., Reviews of Rosenblatt "The Perceptron: A Theory of Statistical Separability in Cognitive Systems", and "Two Theorems of Statistical Separability" (CAL Reports VG-1196-G-1 and VG-1196-G-2), Math. Reviews, Vol. 22, No. 12B, Dec. 1961, Review No. 1330, A and B, Pages 2281-2282.

62. Murray, A.E., A Review of the Perceptron Program, Proc. of National Electronics Conf., Chicago, 1959, V. 15, 346-356.

63. Murray, A., Phase 1 Interim Report: Perceptron Applicability to Photo-Interpretation, CAL Report VE-1446-G-1, Nov. 1960.

64. Nagy, George, Analogue Memory Mechanisms for Neural Nets, CSRP Rept. No. 3, Cornell Univ., 1962. Also Ph.D. thesis, Cornell Univ. 1962.

65. Nagy, George, Analog Information Storage Elements, IRE Transactions on Electronic Computers, 1963 (In Press).

66. Nilsson, N., Learning Machines (Class Notes for EE 353, Stanford Univ., 1962.)

67. Novikoff, A., On Convergence Proofs for Perceptrons, Proceedings of Symposium on Mathematical Theory of Automata, Polytechnic Institute of Brooklyn, April, 1962. Published as Technical Report by SRI, Jan., 1963.

68. Otkhmezury, G.L., On Rosenblatt's Theorems and Perceptron-Type Systems, Avtomatika, 1963, No. 1, pp. 10-23. [In Ukrainian].

69. Otkhmezury, G.L., On the Properties of Signs and the Sixth Positive Feedback, Avtomatika, 1963, No. 2, 41-52. [In Ukrainian].

70. Palmieri, G., and Sanna, R., Automatic Probabilistic Programmer/Analyzer for Pattern Recognition, Instituto di Fisica, Univ. of Genoa, 1961.

71. Papert, S., Some Mathematical Models of Learning, Proc. of Fourth London Symposium on Information Theory, 1960.

72. Papert, S., Redundancy and Linear Logical Nets, Bionics Symposium, WADD Technical Report 60-600, 1960, pp. 181-195.

73. Pryor, C.N. Jr., The Perceptron as an Adaptive Classification Device, Naval Ordnance Lab., White Oak, Md., No. NOLTR 61-114, 1961.

74. Roberts, L.G., Pattern Recognition with an Adaptive Network, Record of IRE National Convention, Part 2, New York, 1960.

75. Rosenblatt, F., The Perceptron, A Perceiving and Recognizing Automaton (Project PARA), CAL Rept. No. 85-460-1, Jan. 1957.

76. Rosenblatt, F., A Probabilistic Model for Visual Perception, Proc. of the 15th International Congress of Psychology, North Holland Publishing Co., Amsterdam, 1957, pp. 296-297.

77. Rosenblatt, F., The Perceptron: A Theory of Statistical Separability in Cognitive Systems, CAL Rept. No. VG-1196-G-1, Jan. 1958.

78. Rosenblatt, F., The Design of an Intelligent Automaton, Research Reviews, ONR, Washington, Oct. 1958, 5-13 (Edited version also published in Research Trends, V. 6, No. 2, 1958, 1-7.)

79. Rosenblatt, F., The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain, Psych. Rev., 1958, 65, 386-408.

80. Rosenblatt, F., Two Theorems of Statistical Separability in the Perceptron, in Mechanization of Thought Processes (Vol I), H.M. Stationery Office, London, 1959. (Also published by Cornell Aero. Lab. as Report No. VG-1196-G-2, Sept., 1958.)

81. Rosenblatt, F., A Conjecture on the Biochemistry of Memory Mechanisms, Project PARA Tech. Memo. No. 10, Cornell Aero. Lab., Buffalo, 1959.

82. Rosenblatt, F., Perceptron Simulation Experiments, Proc. IRE, 1960, 48, 301-309.

83. Rosenblatt, F., On the Convergence of Reinforcement Procedures in Simple Perceptrons, CAL Rept. VG-1196-G-4, Buffalo, Feb. 1960.

84. Rosenblatt, F., Tables of Q-Functions for Two Perceptron Models, CAL Rept. VG-1196-G-6, Buffalo, May 1960.

85. Rosenblatt, F., Perceptual Generalization over Transformation Groups, in Yovits and Cameron (Eds) Self Organizing Systems, Pergamon Press, N.Y., 1960.

86. Rosenblatt, F., Perceptrons and Cognitive Systems, in Lernende Automaten, R. Oldenbourg, München, 1961.

87. Rosenblatt, F., Strategic Approaches to the Study of Brain Models, in Principles of Self-Organization, von Foersten and Zopf (Eds.), Pergamon Press, N.Y. 1962.

88. Rosenblatt, F., Principles of Neurodynamics: Perceptrons and the Theory of Brain Mechanisms, Spartan Books, Washington DC, 1962. (Also published by Cornell Aero. Lab. as Report No. VG-1196-G-8, 15 March, 1961.)

89. Rosenblatt, F., A Comparison of Several Perceptron Models, in Yovits, Jacobi, and Goldstein (Eds.) Self-Organizing Systems-1962, Spartan Books, Washington, 1962.

90. Rosenblatt, F., Analytic Techniques for the Study of Neural Nets, in Proc. of AIEE Joint Automatic Control Conference, 1962.

91. Rosenblatt, F., [Translation into Ukrainian of letter to Ivakhnenko], Avtomatika, 1962, 90-91.

92. Rosenblatt, F., Applications of Perceptrons to Speech Recognition, in Proc. of Speech Communication Seminar, Vol. 2, Royal Institute of Technology, Stockholm, 1963.

93. Rosenblatt, F., A Model for Experiential Storage in Neural Networks, Proceedings of COINS Symposium, Spartan Books, Washington, 1963. (In press).

94. Rosenblatt, F. (Ed.), Collected Technical Papers, Vol. 2, CSRP Report No. 4, Cornell University, July, 1963.

95. Shlesinger, M.I., Experiments on the Simulation of the "Alpha" System with Positive Feedback, Avtomatika, 1963, No. 2, 82-88. [In Ukrainian.]

96. Singleton, R.C., A Test for Linear Separability as Applied to Self-Organizing Machines, Stanford Research Institute Technical Report, May 1962.

97. Smith, Jack W., ADAP II - An Adaptive Routine for the LARC Computer, Navy Management Office, 24 Sept. 1962.

98. Warshaw, M., An Analysis of the Perceptron, Rand Corporation Report P-1884, Jan. 18, 1960.

DISTRIBUTION LIST

Addressee	No. of Copies
Assistant Sec. of Def. for Res. and Eng. Information Office Library Branch Pentagon Building Washington 25, D.C.	2
Armed Services Technical Information Agency Arlington Hall Station Arlington 12, Virginia	10
Chief of Naval Research Department of the Navy Washington 25, D.C. Attn Code 437, Information Systems Branch	2
Chief of Naval Operations OP-07T-12 Navy Department Washington 25, D.C.	1
Director, Naval Research Laboratory Technical Information Officer/Code 2000/ Washington 25, D.C.	6
Commanding Officer, Office of Naval Research Navy No. 100, Fleet Post Office New York, New York	10
Commanding Officer, O N R Branch Office 346 Broadway New York 13, New York	1
Commanding Officer, O N R Branch Office 495 Summer Street Boston 10, Massachusetts	1
Bureau of Ships Department of the Navy Washington 25, D.C. Attn: Code 607A NTDS	1
Bureau of Naval Weapons Department of the Navy Washington 25, D.C. Attn: RMWC Missile Weapons Control Div.	1

Addressee

(1 copy each)

Bureau of Naval Weapons
Department of the Navy
Washington 25, D.C.
Attn: RUDC ASW Detection and Control Div.

Bureau of Ships
Department of the Navy
Washington 25, D.C.
Attn: Communications Branch Code 686

Naval Ordnance Laboratory
White Oaks
Silver Spring 19, Maryland
Attn: Technical Library

David Taylor Model Basin
Washington 7, D.C.
Attn: Technical Library

Naval Electronics Laboratory
San Diego 52, California
Attn: Technical Library

University of Illinois
Control Systems Laboratory
Urbana, Illinois
Attn: D. Alpert

Air Force Cambridge Research Laboratories
Laurence C. Hanscom Field
Bedford, Massachusetts
Attn: Research Library, CRX2-R

Technical Information Officer
US Army Signal Research and Dev. Lab.
Fort Monmouth, New Jersey
Attn: Data Equipment Branch

National Security Agency
Fort George G. Meade, Maryland
Attn: R-4, Howard Campaigne

U.S. Naval Weapons Laboratory
Dahlgren, Virginia
Attn: Head, Compution Div., G.H. Gleissner

Addressee

(1 copy each)

National Bureau of Standards
Data Processing Systems Division
Room 239, Bldg. 10
Attn: A.K. Smilow
Washington 25, D.C.

Aberdeen Proving Ground, BRL
Aberdeen Proving Ground, Maryland
Attn: J.H. Giese, Chief Compution Lab.

Commanding Officer
O N R, Branch Office
John Crerar Library Bldg.
86 East Randolph Street
Chicago 1, Illinois

Commanding Officer
O N R Branch Office
1030 E. Green Street
Pasadena, California

Commanding Officer
O N R Branch Office
1000 Geary Street
San Francisco 9, California

National Bureau of Standards
Washington 25, D.C.
Attn: Mr. R.D. Elbourn

Syracuse University
Electrical Eng. Dpt.
Syracuse 10, New York
Attn: Dr. Stanford Goldman

Burroughs Corporation
Research Center
Paoli, Pennsylvania
Attn: R.A. Tracey

Office of Naval Research
Washington 25, D.C.
Attn: Code 455

Addressee

(1 copy each)

Lockheed Missiles and Space Company
3251 Hanover Street
Palo Alto, California
Attn: W.F. Main

Communications Sciences Lab
University of Michigan
180 Frieze Building
Ann Arbor, Michigan
Attn: Gordon E. Peterson

Univ. of Michigan
Ann Arbor, Michigan
Attn: Dept. of Psychology,
Prof. Arthur Melton

Univ. of Michigan
Ann Arbor, Michican
Attn: Dept. of Philosophy,
Prof. A.W. Burks

University of Pennsylvania
Philadelphia 4, Pennsylvania
Attn: Mr. R. Duncan Luce
Dept. of Psychology

Carnegie Institute of Technology
Dept. of Psychology
Pittsburgh 13, Pennsylvania
Attn: Prof. Bert F. Green, Jr.

Massachusetts Inst. of Technology
Research Lab. for Electronics
Cambridge, Mass.
Attn: Dr. Marvin Minsky

Stanford University
Stanford, California
Attn: Electronics Lab.,
Prof. John G. Linvill

University of Illinois
Urbana, Illinois
Attn: Electrical Engrg. Dept.,
Prof. H. Von Foerster

Addressee

(1 copy each)

Univ. of California
Institute of Eng. Research
Berkeley 4, California
Attn: Prof. A.J. Thomasian

National Science Foundation
Program Director for Documentation Research
Washington 25, D.C.
Attn: Helen L. Brownson

Univ. of California - LA
Los Angeles 24, California
Attn: Dept. of Engineering,
Prof. Gerald Estrin

Hebrew University
Jerusalem, Israel
Attn: Prof. Y. Bar-Hillel

Massachusetts Institute of Technology
Research Laboratory of Electronics
Attn: Prof. W. McCulloch

University of Illinois
Champaign Urbana, Illinois
Attn: John R. Pasta

Naval Research Laboratory
Washington 25, D.C.
Attn: Security Systems
Code 5266, Mr. G. Abraham

Zator Company
140 1/2 Mt. Auburn
Cambridge 38, Massachusetts
Attn: R.J. Solomonoff

Telecomputing Corporation
12838 Saticoy Street
North Hollywood, California
Data Instruments Division,
Field Engineering Dept.

NASA
Goddard Space Flight Center
Washington 25, D.C.
Attn: Arthur Shapiro

Addressee

(1 copy each)

Professor Marc Kac
Rockefeller Institute
New York, New York

Cornell University
Department of Psychology
Ithaca, New York
Attn: Prof. Robert B. MacLeod

Cornell University
Department of Psychology
Ithaca, New York
Attn: Prof. Julian Hochberg

Western Reserve University
Department of Anatomy-School of Medicine
Cleveland 6, Ohio
Attn: Professor Marcus Singer

Cornell University
Division of Modern Languages
Ithaca, New York
Attn: Prof. Charles Hockett

Cornell University
Department of Physics
Ithaca, New York
Attn: Prof. Phillip Morrison

Dr. W. Ross Ashby
University of Illinois
Dept. of Electrical Engineering
Urbana, Illinois

National Physical Laboratory
Teddington, Middlesex
England
Attn: Dr. A.M. Uttley, Superintendent, Autonomics Division

University College, London
Department of Elec. Eng.
Gower Street
London, W.C. 1, England
Attn: Dr. W.K. Taylor

Dr. George B. Yntema
United Aircraft Corporation
Research Laboratories
East Hartford 8, Connecticut

Addressee

(1 copy each)

Swarthmore College
Swarthmore, Pennsylvania
Attn: Department of Electrical Engrg., Prof. Carl Barus

Neurological Institute of McGill Univ.
3801 University Street
Montreal, Canada
Attn: Dr. Herbert Jasper

Dr. Jacob Beck
Harvard University
Memorial Hall
Cambridge 38, Massachusetts

University of Michigan
Ann Arbor, Michigan
Attn: Department of Psychology, Dr. James Olds

Diamond Ordnance Fuze Laboratory
Connecticut Ave. and Van Ness St.
Washington 25, D.C.
ORDTL-012, E.W. Channel

Harvard University
Cambridge, Massachusetts
Attn: School of Applied Science Dean Harvey Brook

Commanding Officer and Director
U.S. Naval Training Device Center
Port Washington
Long Island, New York
Attn: Technical Library

Office of Naval Research
Washington 25, D.C.
Attn: Code 450, Dr. R. Trumbull

The University of Chicago
Institute for Computer Research
Chicago 37, Illinois
Attn: Mr. Nicholas C. Metropolis, Director

Wright Air Development Division
Electronic Technology Laboratory
Wright-Patterson AFB, Ohio
Attn: Lt. Col. L.M. Butsch, Jr.
ASRUEB

Addressee

(1 copy each)

Laboratory for Electronics, Inc.
1079 Commonwealth Ave.
Boston 15, Massachusetts
Attn: Dr. H. Fuller

Stanford Research Institute
Computer Laboratory
Menlo Park, California
Attn: H.D. Crane

The Rand Corp.
1700 Main St.
Santa Monica, California
Attn: Numerical Analysis Dept.
Willis H. Ware

Univ. of Oregon
Portland, Oregon
Attn: Medical School, Dr. Tunturi

Zator Co.
140 1/2 Mt. Auburn Street
Cambridge 38, Mass.
Attn: Calvin N. Mooers

Massachusetts Institute of Technology
Cambridge 39, Mass.
Attn: Prof. John McCarthy, 26-007B

Office of Naval Research
Washington 25, D.C.
Attn: Code 430

Carnegie Institute of Technology
Pittsburgh, Pennsylvania
Attn: Director, Computation Cntr.,
Alan J. Perlis

Rome Air Development Center, RCOR
DCS/Operations, USAF
Griffiss Air Force Base, New York
Attn: Irving J. Gabelman

Air Force Office of Scientific Research
Directorate of Information Sciences
Washington 25, D.C.
Attn: Dr. Harold Wooster

Addressee

(1 copy each)

Hunter College
New York 21, N.Y.
Attn: Dean Mina Rees

Radio Corporation of America
306/2
Data Systems Division
8500 Balboa Blvd.
Van Nuys, California
Attn: Joseph E. Karroll

Mr. Sidney Kaplan
1814 Glen Park Avenue
Silver Spring, Maryland

Stanford Research Institute
Menlo Park, California
Attn: Dr. Charles Rosen
Applied Physics Laboratory

University of California
Department of Mathematics
Berkeley, California
Attn: Prof. H.J. Bremermann

National Bureau of Standards
Washington 25, D.C.
Attn: Miss Ida Rhodes, 220 Stucco Bldg.

University of Saskatchewan
College of Engineering
Saskatoon, Canada
Attn: H.C. Ratz

Northeastern University
360 Huntington Avenue
Boston, Massachusetts
Attn: Prof. L.O. Dolansky

Dr. Stanley Winkler
IBM Corporation
Federal Systems Division
326 E. Montgomery Avenue
Rockville, Maryland

L.G. Hanscom Field /AF-CRL-CRRB/
Bedford, Mass.
Attn: Dr. H.H. Zschirnt

Addressee

(1 copy each)

Rome Air Development Center
Griffiss Air Force Base
Rome, New York
Attn: Mr. Alan Barnun

Armour Research Foundation
10 West 35th Street
Chicago 16, Illinois
Attn: Mr. Scott Cameron
E.E. Research Dept.

Department of the Army
Office of the Chief of Research
and Development
Pentagon, Room 3D442
Washington 25, D.C.
Attn: Mr. L.H. Geiger

Carnegie Institute of Technology
Systems and Communication Sciences
Pittsburgh 13, Pennsylvania
Attn: Dr. Allen Newell

Cornell University
Cognitive Systems Research Program
Cornell Computing Center
Ithaca, New York
Attn: Prof. H.D. Block

Royal Aircraft Establishment,
Mathematics Dept.
Farnborough, Hampshire, England
Attn: Mr. R.A. Fairthorne
Minister of Aviation

Harvard Computation Lab.
Harvard University
Cambridge, Massachusetts
Attn: Dr. Anthony Oettinger

IBM Research
Yorktown Heights, New York
Attn: Hans Peter Luhn

Library of Congress
Washington 25, D.C.
Attn: John Sherrod

Addressee

(1 copy each)

National Biomedical Research Foundation Inc
8600 16th St., Suite 310
Silver Spring, Maryland
Attn: Dr. R.S. Ledley

National Bureau of Standards
Washington 25, D.C.
Attn: Mrs. Ethel Marden

Department of Commerce
U.S. Patent Office
Washington 25, D.C.
Attn: Mr. Herbert R. Koller

Thompson Ramo Woolridge Inc.
8433 Fallbrook Ave.
Canoga Park, California
Attn: D.R. Swanson

Rand Corporation
1700 Main Street
Santa Monica, California
Attn: Library

University of Chicago
Committee on Mathematical Biology
Chicago, Illinois
Attn: Prof. H.D. Landahl

University of Pennsylvania
Moore School of Electrical Engineering
200 South 33 rd Street
Philadelphia 4, Pennsylvania
Attn: Miss Anna Louise Campion

Department of the Army
Office of the Asst. COFD for Intelligence
Room 2B529, Pentagon
Washington, D.C.
Attn: John F. Kullgren

Mr. Robert F. Samson
Directorate of Intelligence and
Electronic Warfare
Griffiss Air Force Base
Rome, New York

Addressee

(1 copy each)

Army Research Office OCR+D
Department of Army
Washington 2, D.C.
Attn: Mr. Gregg McClurg

Division of Automatial Data Processing /AOP/
Department of State
Washington 25, D.C.
Attn: F.P. Diblasi,
19A16

Mr. Paul W. Howerton
Room 1053 M. Bldg.
Code 163
CIA
Washington, D.C.

Mr. Bernard M. Fry, Deputy Head
Office of Science Information Service
National Science Foundation
1951 Constitution Avenue, N.W.
Washington 25, D.C.

International Business Machines, Corp.
Advanced Systems Development Div.
San Jose 14, Calif.
Attn: I.A. Warheit

Harry Kesten
Cornell University
Dept. of Mathematics
Ithaca, New York

Applied Physics Laboratory
Johns Hopkins University
8621 Georgia Avenue
Silver Spring, Maryland
Attn: Document Library

Bureau of Supplies and Accounts, Chief
Navy Department
Washington, D.C.
Attn: Code W3

Bendix Products Division
Bendix Aviation Corporation
Southbend 20, Indiana
Attn: E.H. Crisler

Addressee

(1 copy each)

Officer in Charge
U.S. Naval Photographic Interpretation Center
4301 Suitland Road
Suitland, Maryland
Attn: Mr. J. Pickup

Dr. Noah S. Prywes
Moor School of Engineering
University of Pennsylvania
Philadelphia 4, Pennsylvania

Rabinow Engineering Co. Inc.
7712 New Hampshire Avenue
Washington 12, D.C.

National Security Agency
Fort George G. Meade, Maryland
Attn: R. -42, R. Wiggington

Peter H. Greene
Committee on Mathematical Biology
University of Chicago
Chicago, Illinois

Federal Aviation Agency
Bureau of Research and Development
Washington 25, D.C.
Attn/Rd-375/Mr. Harry Hayman

Veterans Administration
Department of Medicine and Surgery
Washington 25, D.C.
Attn/Research and Development Div.
H. Frieburger

Cornell Aeronautical Laboratory Inc.
PO Box 235
Buffalo 21, New York
Attn/Systems Requirements Dept/AE Murray/

Institute for Space Studies
475 Riverside Drive
New York 27, New York
Attn: Mr. Albert Arking

Mr. Donald F. Wilson
Code 5144
Naval Research Laboratory
Washington 25, D.C.

Addressee
(1 copy each)

University of California
Department of Engineering
Berkely 4, California
Attn: Prof. J.R. Singer

Navy Management Office
Data Processing Systems Division
Department of the Navy
Washington 25, D.C.
Attn: Mr. J. Smith

Lincoln Laboratory
Massachusettes Institute of Technology
Lexington 73, Massachusettes
Attn: Library

Prof. Silvio Ceccato
Universita Di Milano
Centro Di Cibernetica
Via Festa Del Perdono, 3
Milano, Italy

Brookhaven National Laboratry
Upton, Long Island
New York
Attn: Dr. Yoshio Shimamoto

Instituto di Fisica dell Universita
Genova, Italy
Attn: Prof. A. Gamba

Dr. Gunnar Fant
Royal Institute of Technology,
Speech Transmission Lab.
Stockholm 70, Sweden

Dr. C. R. Porter
Psychology Department
Howard University
Washington 1, D.C.

Electronics Research Laboratory
University of California
Berkeley 4, California
Attn: Director

Mr. Gordon Stanley
7685 South Sheridan Ct.
Littleton, Colorado
/Martin/Denver/

Addressee
(1 copy each)

Dr. Milton E. Rose
Lawrence Radiation Laboratory
University of California
Berkeley, California

A.J. Cote, Jr.
Applied Physics Laboratory, JHU
The Johns Hopkins University
8621 Georgia Avenue
Silver Spring, Maryland

McGill University
Department of Psychology
Montreal, Quebec-Canada
Attn: Professor Donald Hebb

Commanding Officer
RADC Griffiss Air Force Base
Rome, New York
Attn: Lt. Roger Geesey /RCWID/

Dr. Hong Yee Chiu
Institute for Space Studies
475 Riverside Drive
New York 27, New York

Dr. B. Mandelbrot
Littaver G-51
Harvard University
Cambridge 38, Massachusetts

Aeronutronics
Ford Road
Newport Beach, California
Attn: J.K. Hawkins

Cornell University Library
Ithaca, New York
Attn: Central Serial Record Dept.

University of Rochester
Center for Brain Research
Rochester 20, New York
Attn: E.R. John

Massachusetts Institute of Technology
Lincoln Laboratory
Lexington 73, Massachusetts
Attn: Dr. B. Farley

Addressee

(1 copy each)

Massachusetts Institute of Technology
Lincoln Laboratory
Lexington 73, Massachusetts
Attn: Dr. Oliver Selfridge

Dr. John Hay
62 Kensington Avenue
Northampton, Massachusetts

Mrs. Dorothea M. Myers, Librarian
Pennsylvania Research Associates, Inc.
Suite 508
133 South Thirty-Sixth Street
Philadelphia 4, Pennsylvania

National Bureau of Stds
Data Processing Systems Division
Wash 25, D.C.

Institute for Defense Analysis
Communications Research Division
Von Neumann Hall
Princeton, New Jersey

Juliana Van Solberghan 25
The Hague, Netherlands
Attn: Alexander Steiber

Institut fur Nachrichtenverarbeitung und
Nachrichtenubertragung
Technische Hochshule Karlsruhe
Kaiserstrasse 12, Karlsruhe, Germany
Attn: Prof. Dr. Ing. K. Steinbuch

Cornell Aeronautical Laboratory
Buffalo 21, New York
Attn: R. Stevens

Cornell University
Department of Mathematics
228 White Hall
Ithaca, New York
Attn: Prof. Wolfgang Fuchs

Cornell Aeronautical Laboratory
Buffalo 21, New York
Attn: W. Holmes

Addressee

(1 copy each)

Cornell University
Laboratory for Nuclear Studies
Ithaca, N.Y.
Attn: Prof. Salpeter

Cornell Aeronatuical Laboratory
Buffalo 21, N.Y.
Attn: Library

Mr. Earl B. Hunt
University of California
Western Management Science Inst.
Graduate School of Business Administration
Los Angeles 24, California

Dr. R.D. Joseph, Principal Mathematician
Astropower, Inc.
2968 Randolph Street
Costa Mesa, California

Dr. Peter M. Kelly
Philco Corporation
Philadelphia, Pa.

Ryuichi Kono
Mitsubishi Electric Manufacturing Co., LTD.
Electronics Works
80 Nakano, Minami Shimizu,
/Amagasaki,
Hyogo Prefecture, Japan

ONR Special Representative
%Hudson Laboratories
Columbia University
145 Palisade Street
Dobbs Ferry, New York

Mr. Saul Amarel
Radio Corporation of America
RCA Laboratories
David Sarnoff Research Center
Princeton, New Jersey

Mr. Seymour T. Levine
Research Division
Kollsman Instrument Corp.
80-08 45th Avenue
Elmhurst 73, New York

Addressee
(1 copy each)

Mr. Don Meyer
Ohio State University
1314 Kinnear Road
Columbus 12, Ohio

Dr. J.H. Andreae
Standard Telecommunication
Laboratories Limited
London Road, Harlow, Essex.
England

Allan Kiron
Office of Research and Development
U.S. Department of Commerce
Patent Office
Washington 25, D.C.

Texas Instruments Incorporated
Corporated Research and Engineering
Attn: Technical Reports Service
P.O. Box 5474
Dallas 22, Texas

Stanford Research Institute
Menlo Park, California
Attn: Dr. A.E. Brain

Harvard Medical School
Boston, Massachusetts
Dr. D.H. Hubel

National Bureau of Standards
Washington 25, D.C.
Attn: Dr. Russell Kirsch

Rockefeller Institute
New York 21, New York
Attn: Dr. Bruce W. Knight

University of Chicago
Committee on Mathematical Biology
Chicago, Illinois
Attn: Professor N. Rashevsky

California Institute of Technology
Department of Biology
Pasadena, California
Attn: Professor Ronald Sperry

Addressee
(1 copy each)

University of Illinois
Digital Computer Laboratory
Urbana, Illinois
Attn: Dr. Bruce McCormick

McGill University
Psychology Department
Montreal, Canada
Attn: Dr. Peter Milner

Massachusetts Institute of Technology
Research Laboratory of Electronics
Lexington 73, Mass.
Attn: Professor Walter Rosenblith

Air Force Office of Scientific Research
Information Research Div.
Washington 25, D.C.
Attn: R.W. Swanson

UNCLASSIFIED

UNCLASSIFIED