ARODR 62-2

U. S. ARMY RESEARCH OFFICE (DURHAM)

PROCEEDINGS OF THE SEVENTH CONFERENCE ON THE DESIGN OF EXPERIMENTS IN ARMY RESEARCH DEVELOPMENT AND TESTING

U. S. ARMY RESEARCH OFFICE (DURHAM)
BOX CM, DUKE STATION
DURHAM, NORTH CAROLINA

U. S. ARMY RESEARCH OFFICE (DURHAM)

Report No. 62-2
August 1962

PROCEEDINGS OF THE SEVENTH CONFERENCE ON THE DESIGN OF EXPERIMENTS IN ARMY RESEARCH DEVELOPMENT AND TESTING

Sponsored by the Army Mathematics Steering Committee

conducted at

U. S. Army Signal Research & Development Laboratory
Fort Monmouth, New Jersey
18-20 October 1961

U. S. Army Research Office (Durham)
Box CM, Duke Station
Durham, North Carolina

Engineering
Diff
3-14-66

TABLE OF CONTENTS

* This paper was presented at the Conference. It does not appear in these proceedings.

TABLE OF CONTENTS (cont.) Page

* This paper was presented at the Conference. It does not appear in these proceedings.

TABLE OF CONTENTS (cont.) Page

TABLE OF CONTENTS (cont.) Page

* This paper was presented at the Conference. It does not appear in these proceedings.

FOREWORD

The U. S. Army Signal Research and Development Laboratory of Fort Monmouth, New Jersey served as host to the Seventh Conference on the Design of Experiments in Army Research, Development and Testing. This Laboratory is the Signal Corps' major scientific arm. It has the responsibility to conceive and develop vital detection and communications equipment. Starting during World War I as the Radio Laboratories with a small group of officers, enlisted men and civilians, the Laboratory has grown to an ultra-modern facility. Surviving through wartime expansions, economy waves and name changes, USASRDL has attained a hard-earned reputation as a leader in most phases of electronics research and development. Included among the many accomplishments of the Laboratory are such things as the development of the essential vehicular and fixed radios used in World War I, the walkie-talkie, radar, the world's first radar contact with the moon, radar storm detection, the first feasible mass-production technique using printed circuits, the first solar batteries for satellites, the world's first communications satellite, and the world's first weather satellite.

The research and development atmosphere of USASRDL provided an ideal locale for a conference on the Design of Experiments. The sponsoring group--the Army Mathematics Steering Committee--was pleased to receive an invitation from Colonel H. McD. Brown to use the facilities under his Command for the 1961 conference. Colonel Brown named Messrs. J. A. McClung and Joseph Weinstein as cochairmen for this meeting. At this time the AMSC would like to thank these gentlemen for the excellent local arrangements of the Conference and for the effectiveness with which they provided for the needs of all who participated and attended this conference.

At the Seventh Conference on the Design of Experiments Drs. R. L. Anderson, John Hammersley, G. S. Watson, and G. A. Watterson delivered the invited addresses. Estimation of variance components, Monte Carlo methods, hazard analysis, and time series and spectral analysis were, respectively, the topics treated by these specialists. Professor R. M. Thrall served as Chairman of the Panel Discussion on Simulation. He arranged for Colonel A. W. DeQuoy, Mr. J. H. Moss, and Dr. Gustave Rabson to discuss various aspects of simulation, with Dr. Hammersley serving as a commentator on the papers presented. In addition to these parts of the program, 10 papers were given in Clinical Sessions, and 19 papers in the Technical Sessions.

This volume of the Proceedings contains 37 of the papers which were presented at Conference. In order to contribute to a wider dissemination of knowledge and use of modern statistical principles in the design of experiments, particularly for Army research, development and testing scientists and engineers, the AMSC is making these articles available in this form.

The Seventh Conference was attended by 152 registrants and participants from over 70 different organizations. Speakers and panelists came from the Armour Research Foundation; Bethesda-Chevy Chase High School; Booz-Allen Applied Research, Inc.; Massachusetts Institute of Technology; Mathematics Research Center, University of Wisconsin; Montgomery Blair High School; North Carolina State College; Operations Research Inc.; Phillips Andover Academy; Research Analysis Corporation; Research Triangle Institute; U. S. Bureau of Mines; Cornell University; University of Delaware; University of Georgia; Harvard University; Oxford University; Princeton University; University of Toronto; Woodrow Wilson High School and 11 Army facilities.

The members of the Army Mathematics Steering Committee take this opportunity to express their thanks to the many speakers and other research workers who participated in the Conference; to Colonel H. McD. Brown for making available the excellent facilities of USASRDL for the conference; and to J. A. McClung and Joseph Weinstein for organizing a most interesting and informative tour of the facilities of the U. S. Army Signal Research and Development Laboratory as well as presenting a movie documenting some of the research work being conducted at the Laboratory.

Finally, the Chairman wishes to express his appreciation to his Advisory Committee: F. G. Dressel (Secretary), Fred Frishman, Boyd Harshbarger, Frank E. Grubbs, H. L. Lucas Jr., Clifford J. Maloney, and Joseph Weinstein for their help in selecting the invited speakers and formalizing the plans for this conference.

S. S. Wilks
Professor of Mathematics
Princeton University

SEVENTH CONFERENCE ON THE DESIGN OF EXPERIMENTS
IN ARMY RESEARCH, DEVELOPMENT AND TESTING
18-20 October 1961

U. S. Army Signal Research and Development Laboratory

Tuesday, 17 October

REGISTRATION: 1830-2230 (Eastern Daylight Saving Time)
Hotel Berkeley - Carteret - Asbury Park, New Jersey

ORIENTATION FILM: 2130-2215 - Hotel Berkeley - Carteret - Crystal Terrace
A Film, USASRDL Report 1960, outlining the areas of operations at the Laboratories will be shown.

Wednesday, 18 October

BREAKFAST: 0700-0800 Coffee Shop - Berkeley-Carteret Hotel
Busses will take conferees to U. S. Army Signal R & D Laboratory

REGISTRATION: 0830-0900 Hexagon Building

GENERAL SESSION 1: 0900-1145 Hexagon Building - Room (Main Auditorium)

Calling of Conference to Order:
Mr. Joseph Weinstein, Local Chairman

Welcome:
Colonel Raymond H. Bates, Deputy Commander
U. S. Signal Research & Development Laboratory

Chairman:
Colonel George W. Taylor, Commanding Officer
Army Research Office (Durham)

Time Series and Spectral Analysis:
Dr. G. A. Watterson, Virginia Polytechnic Institute

Monte Carlo Methods:
Dr. John Hammersley, Oxford University and Princeton University

LUNCH: 1200-1330 Hotel Berkeley-Carteret Coffee Shop and Grill

TOURS: 1400

DINNER: 1800-1900 Hotel Berkeley-Carteret Coffee Shop & Grill
Two technical sessions are scheduled for Wednesday night.

TECHNICAL SESSION I: 1930-2145 Hotel Berkeley-Carteret Hunt Suite 'A'

Chairman: F. J. Anscombe, Princeton University

The Construction and Analysis of Non-Orthogonal Plans for the 2^n Factorial Experiments
Sidney Addelman, Statistics Research Division, Research Triangle Institute

Use of the Up-and-Down Method with Factorial Designs
R. L. Grant and R. W. VanDolah, Explosive Research Laboratory, U. S. Bureau of Mines

A General Formula and Positional Index Algorithm for Orthogonal Contrasts in Factorial Designs
Erwin Biser, Systems Division, U. S. Army Signal Research and Development Laboratory

TECHNICAL SESSION II: 1930-2145 Hotel Berkeley-Carteret Skyline Room

Chairman: James F. O'Neal, Springfield Armory

A Semi-Automatic Gaming System
John L. Donaldson, Research Analysis Corporation
Thomas R. Shaw, Operations Research Inc., Silver Spring, Md.

Transient Nuclear Radiation Effects on Electron Tubes and Transistors
Richard G. Saelens, U. S. Army Signal Research and Development Laboratory

The Development of Subminiature Tube Handbook Information
J. A. Zoellner, Analysis and Programming, Electromagnetic Compatability Analysis Center, Armour Research Foundation

Thursday, 19 October

BREAKFAST: 0700-0800 Coffee Shop

Technical Session III and Clinical Sessions A and B will run from 0820 to 1030. Technical Sessions IV, V, and VI scheduled from 1100 to 1230 complete the morning phase of the program. General Session 2 is a panel discussion and is timed from 1400-1600. The Subcommittee on Probability and Statistics of the Army Mathematics Steering Committee will meet at 1630; all members of the conference are invited to attend this committee meeting.

PROGRAM (cont.)

Starting at 1930 there will be a discussion of education programs now being conducted in certain installations. This will be followed by a game in which members of the audience are invited to participate.

TECHNICAL SESSION: 0820-1030 Hotel Berkeley-Carteret Crystal Terrace

Chairman: John P. Purtell, Research Branch, Watervliet Arsenal

Reliability Testing and Estimation for Single and Multiple Environments Using Increased Severity Methods
S. K. Einbinder and Ingram Olkin, Picatinny Arsenal

Reliability, Probability and Bionomial Inference
W. A. Thompson, Jr., University of Delaware

An Experiment on Aircraft Vulnerability
Garth McCormick and Bruce Taylor, Research Analysis Corporation

CLINICAL SESSION A: 0820-1030 Hotel Berkeley-Carteret Hunt Suite 'A'

Chairman: A. Hammer, Springfield Armory

Panel Members: O. P. Bruno, Ballistic Research Laboratories
F. E. Grubbs, Ballistic Research Laboratories
Boyd Harshbarger, Virginia Polytechnic Institute
C. J. Maloney, U. S. Army Biological Warfare Labs.
S. S. Wilks, Princeton University
Marvin Zelen, U. S. Army Mathematics Research Center, The University of Michigan

A Method of Weapon System Analysis
Harry Smith, Picatinny Arsenal

Variation of Artillery Ammunition Expenditure with Intelligence
(Members of the ORO 1961 Summer Research Program for Young People)
Robert H. Hobbs (Group Leader), Mass. Institute of Technology
Simon H. Kahan, Harvard University
Thomas H. Brylawski, Woodrow Wilson High School, Wash., D. C.
Andrew H. Levy, Phillips Andover Academy, Andover, Mass.
Peter E. Lobban, Bethesda-Chevy Chase High School, Bethesda, Md.
Michael M. Weisfield, Montgomery Blair High School, Silver Spring, Maryland
Sponsoring Agency: Research Analysis Corporation

An Approach to Sensitivity Analysis of CARMONETTE (A Small Unit Combat Monte Carlo Simulation)
Richard J. Matteis and William C. Suhler, Research Analysis Corp.

CLINICAL SESSION B: 0820-1030 Hotel Berkeley-Carteret Skyline Room

Chairman: Fred Frishman, Army Research Office, Washington; Office, Chief of Research and Development

Panel Members: R. L. Anderson, North Carolina State College
F. J. Anscombe, Princeton University
R. E. Bechhofer, Cornell University
A. C. Cohen, Jr., The University of Georgia
H. L. Lucas, North Carolina State College
G. S. Watson, University of Toronto

The Complex Nature of Reliability
A. Bulfinch, Picatinny Arsenal

Surveillance Inspection of Textile Materials
William S. Cowie, Laboratory and Control Section, Textile Engineering Branch, TC & F Division, QMR & E Command, Natick, Massachusetts

Problems Involved in Developing and Analyzing Durability Data from Field Tests of Textile Footwear Items
Harold R. Rush, Quartermaster Research and Engineering Field Agency, Fort Lee, Virginia

BREAK: 1030-1100 Oval Lounge - Hotel Berkeley-Carteret

TECHNICAL SESSION IV: 1100-1230 Skyline Room

Chairman: J. F. McAreavy, Headquarters, U. S. Army Ordnance Weapons Command

A Predictor Model for Stability Estimates in the Rotating Drum
Cecil Orain Eckard, U. S. Army Chemical Corps, Fort Detrick, Frederick, Maryland

Disease Severity Quantitation, III
Clifford Joseph Maloney, U. S. Army Chemical Corps, Fort Detrick, Frederick, Maryland

Statistical Studies of Plaque Results in Virus Assay
Francis Marion Wadley and Walter Dean Foster, U. S. Army Chemical Corps, Fort Detrick, Frederick, Maryland

TECHNICAL SESSION V: 1100-1230 Hotel Berkeley-Carteret Hunt Suite 'A'

Chairman: Erwin Biser, U. S. Army Signal Research and Development Laboratory

A Confidence Interval for the Reliability of Multi-Component Systems
John K. Abraham, Surveillance Branch, Weapon Systems Laboratory Ballistic Research Laboratories, U. S. Army Ordnance, Aberdeen Proving Ground

Reliability of Compliance with One-Side Specification Limits when Data is Normally Distributed
E. L. Bombara, Army Rocket & Guided Missile Agency, Redstone Arsenal

A General Approach to Engineering Tolerance Specification
Sheldon G. Levin, Diamond Ordnance Fuze Laboratories

TECHNICAL SESSION VI: 1100-1230 Hotel Berkeley-Carteret Crystal Terrace

Chairman: J. J. Gergen, Army Research Office (Durham)

A Further Analysis of Missile Range Tracking Systems
Oliver Lee Kingsley, Range Instrumentation Development Division, White Sands Missile Range

Efficiency of Average Radius as an Estimate of Circular Probable Error when True Center of Impact is Unknown
Robert I. McKeague, Jr., U. S. Army Ordnance Ammunition Command

LUNCH: 1230-1330 Hotel Berkeley-Carteret Coffee Shop & Grill

GENERAL SESSION 2: 1400-1600 Hotel Berkeley-Carteret Crystal Terrace

Panel Discussion on Simulation
Chairman: Dr. Robert M. Thrall, The University of Michigan
Panel Members: Colonel Alfred W. DeQuoy, Chief, Strategy and Tactics Analysis Group, Department of Army
Dr. John Hammersley, Oxford University and Princeton University
Mr. John H. Moss, Research Analysis Corporation
Dr. Gustave Rabson, The University of Michigan

BREAK: 1600-1630 Oval Lounge - Berkeley-Carteret

GENERAL SESSION 3: 1630 Hotel Berkeley-Carteret Crystal Terrace Room

Subcommittee on Probability and Statistics
Chairman: Dr. Clifford Joseph Maloney, U. S. Army Chemical Corps
Fort Detrick, Frederick, Maryland

The Subcommittee on Probability and Statistics of the Army Mathematics Steering Committee, as one of its assigned duties, is required to survey the needs of the Army for development of new techniques in its field "with a major impact on Army research development and testing", and call these to the attention of the Chief of R & D at least annually.

This open meeting is scheduled during the conference in order to provide an opportunity for Army scientists generally to call specific attention to any such requirements that they may be aware of at this time. Full attendance and participation in this Subcommittee meeting is encouraged.

DINNER: 1800-1900 Hotel Berkeley-Carteret Coffee Shop & Grill

EDUCATIONAL AND GAME SESSION: 1930-2200 Crystal Terrace Room
Hotel Berkeley-Carteret

Chairman: Ralph D. Doner, Army Rocket and Guided Missile Agency,
U. S. Army Ordnance Missile Command

Training Programs in Statistics
A. Bulfinch, Picatinny Arsenal

A Review of a Statistical Workshop
Walter D. Foster, U. S. Army Chemical Corps, Fort Detrick, Md.
Theodore W. Horner, Booz-Allen Applied Research, Inc., Bethesda, Maryland

Ye Olde Confidence Game
(In this interesting game the audience takes an active part. Please read the abstract of this paper.)
Theodore W. Horner, Booz-Allen Research, Inc.
Walter D. Foster, U. S. Army Chemical Corps

Friday*, 20 October

BREAKFAST: 0700-0800 Coffee Shop & Grill - Hotel Berkeley-Carteret

Clinical Session C carries a security classification of CONFIDENTIAL. Technical Session VII and Clinical Sessions C & D run from 0830-0945. General Session 4 is called from 1015-1230.

TECHNICAL SESSION VII: 0830-0945 Hotel Berkeley-Carteret Crystal Terrace

Chairman; M. J. Pascual, Research Branch, Watervliet Arsenal

Some Aspects of Linear Regression Systems
William S. Mallios, Institute of Statistics, North Carolina State College

Some Results Concerning the Reduction of Product Variability through the Use of Variance Component Analysis
Richard R. Prairie, Institute of Statistics North Carolina State College

CLINICAL SESSION C: 0830-0945 Hotel Berkeley-Carteret, Skyline Room

Security Classification - CONFIDENTIAL

Chairman: Ira A. DeArmon, Jr., Operations Research Group, U. S. Army Chemical Corps

Panel Members: O. P. Bruno, Ballistic Research Laboratories
F. E. Grubbs, Ballistic Research Laboratories
Boyd Harshbarger, Virginia Polytechnic Institute
S. S. Wilks, Princeton University
Marvin Zelen, U. S. Army Mathematics Research Center, The University of Wisconsin

Use of Statistical Designs in Laboratory Environmental Testing of Adaption Kits
Daniel J. Taravella, Picatinny Arsenal

---

* Due to the length of the program and the agreed checkout time of 1330, conferees are requested to pack their bags and complete their financial transactions during the period from 0800-0830 or during the morning coffee break.

A Series of Two-Phase Experiments
Emil H. Jebe, Institute of Science and Technology, The University of Michigan

CLINICAL SESSION D: 0830-0945 Hotel Berkeley-Carteret Hunt Suite 'A'

Chairman: Dorothy M. Gilford, Logistics and Mathematical Statistics Branch, Office of Naval Research

Panel Members: R. E. Bechhofer, Cornell University
A. C. Cohen, Jr., The University of Georgia
H. L. Lucas, North Carolina State College
C. J. Maloney, U. S. Army Biological Warfare Laboratories
G. A. Watterson, Virginia Polytechnic Institute

Fitting the "Modified Exponential" Function by a Multiple Regression Method
Willis LeRoy Hasty, U. S. Army Chemical Corps, Fort Detrick, Frederick, Maryland

Problems Related to a Bio-Assay for Spore-Germination Inhibitors Associated with Uredospores
K. R. Bromfield, Crops Division, U. S. Army Chemical Corps, Chemical Corps Biological Laboratories, Fort Detrick, Md.

BREAK: 0945-1015 Hotel Berkeley-Carteret Oval Lounge

GENERAL SESSION 4: 1015-1230 Hotel Berkeley-Carteret Crystal Terrace Room

Chairman: Dr. S. S. Wilks, Princeton University

Designs for Estimating Variance Components
Dr. R. L. Anderson, Institute for Statistics, North Carolina State College

Hazard Analysis
Dr. G. S. Watson, University of Toronto

LUNCH: 1230-1330 Hotel Berkeley-Carteret Coffee Shop & Grill

TIME SERIES AND SPECTRAL ANALYSIS

G. A. Watterson
Virginia Polytechnic Institute

INTRODUCTION. One of the classical problems in statistics is the following. An experiment produces a result which is a random variable having a density function $f(x)$ say. In order to investigate the nature of this density, several independent experiments are performed, yielding observations

$$x_1, x_2, \ldots, x_n .$$

The joint density (the "likelihood") of these observations is then

$$f(x_1, x_2, \ldots, x_n) = f(x_1) f(x_2) \ldots f(x_n) , \tag{1}$$

and assuming the form of $f(x)$ is known, the parameters may be estimated by maximum likelihood.

A more general problem is the following. Suppose the observations

$$x_1, x_2, \ldots, x_n$$

are taken on an experiment, where the subscript refers to the time at which the observation is made. For such time-series, one can seldom assume independence between the observations. There are two points of view which can be taken about the analysis of such observations. The lack of independence is often a nuisance -- we wish to analyze the observations by the use of standard techniques such as t and F tests, confidence limits, regression analysis, but cannot because of the lack of independence. On the other hand in some situations the lack of independence is just that aspect in which we are most interested. If we wish to predict something about the future on the basis of past experience, it would, indeed, be a nuisance if the future was independent of the past. So, depending on one's point of view, correlation can either be a hinderance or a help.

One of the most fruitful assumptions that can be made, at least as far as mathematical theory is concerned, is that of (weak) stationarity. By this, we mean that the expected values, and the variances, of our observations are constant over time, and that the covariance between two observations depends on the time between them, but not on just when we take the observations. Whether this assumption is true or not of course depends on the practical situation. We give three examples.

(i) A radar installation detects echos of radio signals bounced back from a physical object. Ideally, no statistics is required and the observed echos are explained deterministically; if the object is in motion, the signals will not be stationary. But suppose no object exists, and the only "signal" received is due to the "noise" generated by the tubes in the receiver. This noise may very well satisfy the requirements for a stationary, random, time series.

(ii) Ocean waves tend to make a ship rock. If the waves have most of their energy concentrated in a frequency range corresponding to the natural frequency of oscillation of the ship, then a dangerous rocking motion may result. Thus we wish to study the frequency structure of waves, which over a short time may be approximately stationary in the above sense, although over longer periods of time will exhibit non-stationary features due to tides, etc.

(iii) The word "time" used above need not really mean "time". An axle for a truck may be specified as having a diameter of one inch; if the production machinery is satisfactory then one can expect that the average diameter of a batch of axles will be close to one inch wherever the measurement is taken, but certainly, observations made on the same axle will have a correlation depending on the distance apart that they are taken. If the production machinery is not working as it should, then one end of the axles may be larger than the other and the expected values would not be stationary along the axle.

In this paper, we intend to review some of the techniques that may be used for handling correlated data taken at equally spaced time intervals, under the assumption of stationarity. Two approaches can be made. One I call the "statistician's" approach, in which one tries to estimate the parameters in the joint distribution of the observations. The other, here called the "electrical engineer's" approach, concerns the estimation of how important, or otherwise, are certain frequencies apparent in observed time series. In choosing my labels, I am doing an injustice to statisticians -

the frequency interpretation is probably the most interesting in practice, and certainly statisticians are also interested in it. In any case, the two approaches are intimately connected although not obviously so. Some of the more recent books discussing time series analysis are [2], [4], [5].

STATISTICIAN'S APPROACH. Let us take as our time scale unit, the interval between successive observations. The assumptions of stationarity imply

$$E(x_1) = E(x_2) = \dots = E(x_t) = \dots = E(x_n) = \mu \text{ say},$$

(2) $\operatorname{Cov}(x_{t_1}, x_{t_2}) = c(t_1 - t_2) = c(t_2 - t_1)$ say, for all $t_1, t_2 = 1, 2, \dots, n$

and in particular, $\operatorname{Var}(x_1) = \operatorname{Var}(x_2) = \dots = \operatorname{Var}(x_n) = c(0)$.

If, in addition, we may assume that the joint distribution of the observations is normal, with density function

(3) $$f(x_1, x_2, \dots, x_n) = (1/2\pi)^{n/2} \left|\Sigma^{-1}\right|^{1/2} e^{-\frac{1}{2}(\underline{x}-\underline{\mu})' \Sigma^{-1}(\underline{x}-\underline{\mu})}$$

where

$$\Sigma = \begin{bmatrix} c(0) & c(1) & \cdots & c(n-1) \\ c(1) & c(0) & c(1) & c(n-2) \\ & \ddots & \ddots & \ddots \\ & & & c(1) \\ c(n-1) & \cdots & c(1) & c(0) \end{bmatrix}$$

is the matrix of the variances and covariances, and

$$\underline{x} = \begin{bmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ \cdot \\ x_n \end{bmatrix}, \qquad \underline{\mu} = \begin{bmatrix} \mu \\ \mu \\ \cdot \\ \cdot \\ \cdot \\ \mu \end{bmatrix},$$

then we see that there are $n + 1$ parameters μ, $c(0)$, $c(1)$, ..., $c(n - 1)$ to be estimated. Of course, from one sample of size n one can hardly hope to estimate all $n + 1$ parameters reliably. We will see shortly how these can sometimes be reduced in number by making additional assumptions. A time-series with (3) as density, is called "Gaussian".

For samples obtained independently, the most generally useful estimate of μ is

$$\bar{x} = \frac{\sum_{t=1}^{n} x_t}{n}$$

since it is unbiased, $E(\bar{x}) = \mu$, and has a variance σ^2/n (or in our notation $c(0)/n$) which can be made arbitrarily small by taking a sufficiently large sample. However, in time-series analysis, $\bar{x}$ is not the maximum likelihood estimate of μ even when (3) is taken to be the likelihood. There are estimates of smaller variance than $\bar{x}$, see [9]. Nevertheless $\bar{x}$ is almost invariably used since it is unbiased, easy to compute, and its competitors involve a knowledge of the covariances or at least their estimation. Whether $\bar{x}$ has a variance which decreases as n increases is, of course, a very important question, and forms the basis of the Ergodic Theorem. Roughly speaking, provided the covariance between observations decreases fairly rapidly as the time between them increases, then the observed average of a time series approaches the population mean μ as the number of observations increase. For, we have

$$\mathrm{Var}(\bar{x}) = \mathrm{Var}\left(\frac{\sum_{1}^{n} x_t}{n}\right) = 1/n^2 \sum_{t_2=1}^{n} \sum_{t_1=1}^{n} \mathrm{Cov}(x_{t_1}, x_{t_2}) ,$$

$$= 1/n^2 \sum_{t_2=1}^{n} \sum_{t_1=1}^{n} c(t_1 - t_2) ,$$

which can be seen to equal

$$= \frac{c(0)}{n} + \frac{2}{n} \sum_{t=1}^{n-1} (1 - t/n)\, c(t) .$$

If we write $\rho(t)$ = correlation between observations spaced t apart

$$= \frac{\mathrm{Cov}(x_1, x_{1+t})}{\sqrt{\mathrm{Var}(x_1) \cdot \mathrm{Var}(x_{1+t})}} = \frac{c(t)}{c(0)} .$$

we have

$$\text{(4)} \qquad \mathrm{Var}(\bar{x}) = \frac{c(0)}{n} \left\{ 1 + 2 \sum_{t=1}^{n-1} (1 - t/n)\, \rho(t) \right\} .$$

Thus, provided $\dfrac{\sum_{1}^{n-1} \rho(t)}{n} \rightarrow 0$ as $n \rightarrow \infty$, $\mathrm{Var}(\bar{x}) \longrightarrow 0$.

When the conditions of the theorem are fulfilled, we do not need to independently replicate the realizations to get a good estimate of μ ; all we need is a single, long, realization. But what if the conditions do not hold? For example, suppose we try to estimate the average rainfall over U.S.A. by choosing only one recording station at random, say New York, to provide us our data. Clearly, for a long series of observations from New York we can get a good estimate of the average rainfall at New York, but this tells us very little about the entire country. In the wider view, the New York observations are correlated no matter how far they are separated in time; they are all subject to the same particular environmental factors. Then, replication using other recording stations would be necessary.

Turning now to the estimation of the covariances, our assumptions of stationarity included the condition that the $n - t$ pairs

$$x_1, \quad x_{1+t}; \quad x_2, \quad x_{t+2}; \quad \ldots; \quad x_{n-t}, \quad x_n$$

each had the covariance $c(t)$. It is natural to estimate $c(t)$ by

$$\hat{c}(t) = \frac{(x_1 - \bar{x})(x_{1+t} - \bar{x}) + (x_2 - \bar{x})(x_{2+t} - \bar{x}) + \ldots + (x_{n-t} - \bar{x})(x_n - \bar{x})}{n - t}$$

(5) $$= \frac{\sum_{j=1}^{n-t} (x_j - \bar{x})(x_{j+t} - \bar{x})}{n - t}, \qquad t = 0, 1, \ldots, n-1 .$$

In partiular, the variance $c(0)$ would be estimated by

(6) $$\hat{c}(0) = \frac{\sum_{j=1}^{n} (x_j - \bar{x})^2}{n}$$

which is, of course, very familiar from elementary statistics. Clearly some of these estimates are better than others. In (6) we are averaging n terms, but when we come to estimating $c(n-1)$ we have only one pair of observations to use. In any case, the estimates will be biased in general, and their exact distribution is difficult to find.

The correlations $\rho(t) = \frac{c(t)}{c(0)}$ can most naturally be estimated by

$$\hat{\rho}(t) = \frac{\hat{c}(t)}{\hat{c}(0)} \qquad t = 1, 2, \ldots, n-1 \tag{7}$$

using (5) and (6). Again, and especially for small n, these estimates are somewhat biased, and for those with t close to n their variances are large. A plot of these estimated correlations, called the "correlogram", will give a general idea of how the dependence of an observation on the previous ones behaves.

Under fairly general conditions - roughly that the true correlations $\rho(t)$ tend to zero fairly rapidly as $t \to \infty$ - it has been shown that our estimates $\bar{x}$, $\hat{c}(t)$, $\hat{\rho}(t)$ are asymptotically normal as $n \to \infty$. The exact, small sample, distributions depend on the assumptions made about the joint density $f(x_1, x_2, \ldots, x_n)$, and even when this is multivariate normal only $\bar{x}$ has a simple (normal) distribution. Various versions of estimates $\hat{\rho}(t)$ have been investigated; for their moments see [2], and their distributions see [1], [3], [12].

One device which is of frequent use in explaining an observed time series, using fewer parameters than those introduced above, is the autoregressive model. As an example, one might, to a first approximation, expect that the stock market index today, x_t, is dependent on the index yesterday, x_{t-1}, on the direction of the previous trend (whether upward or downward), $x_{t-1} - x_{t-2}$, and also on additional factors peculiar to today, e_t. We might have as a model

$$x_t = \alpha_1 x_{t-1} + \alpha_2 (x_{t-1} - x_{t-2}) + e_t \ ,$$

or, written slightly differently,

$$x_t = \beta_1 x_{t-1} + \beta_2 x_{t-2} + e_t \quad . \tag{8}$$

A model of this type, where the dependence of x_t on previous values is linear, is called a (linear) autoregression. The parameters of interest are the regression coefficients β_1, β_2, ... and the variance of the e_t. Least squares estimation of course yields the usual expressions encountered in ordinary regression analysis, although here, x plays the dual role of dependent and independent variable. In fairly general situations, the sampling properties of the estimates are closely approximated by the usual distributions used in regression theory. For significance tests in autoregressive models see [10], [6], [7].

<u>ELECTRICAL ENGINEER'S APPROACH</u>. When frequency of oscillation is more interesting than the serial correlations, a somewhat different approach is taken. Even a simple model like

$$x_t = -\beta x_{t-1} + e_t \quad , \tag{9}$$

with $\beta > 0$, can produce observations which tend to be alternatively positive, and negative; the pattern can of course be disrupted by the error term e_t. The most important single frequency here would be 1 cycle per 2 time units. A more general model such as (8) can exhibit oscillations of any frequency. To show the connection between the two approaches, let us consider again the serial covariance function. Any linear combination of x_1, x_2, ..., x_n, say $a_1x_1 + a_2x_2 + \ldots + a_nx_n$ must have a nonnegative variance. That is

$$\mathrm{Var}(a_1 x_1 + a_2 x_2 + \ldots + a_n x_n) = \sum_{t_2=1}^{n} \sum_{t_1=1}^{n} a_{t_1} a_{t_2} \mathrm{Cov}(x_{t_1}, x_{t_2})$$

$$= \sum_{t_2=1}^{n} \sum_{t_1=1}^{n} a_{t_1} a_{t_2} c(t_1 - t_2)$$

$$\geq 0 .$$

By choosing, firstly, $a_t = \cos t\Theta$, and secondly, $a_t = \sin t\Theta$, and combining the two results, one can show that the covariance function $c(t)$ cannot be completely arbitrary, but must be expressible as

(10) $$c(t) = c(0) \int_0^{\pi} \cos(t\Theta)\, dF(\Theta) \quad ,$$

where $F(\Theta)$ is some cumulative distribution function on $(0, \pi)$, see [2]. Of course $F(\Theta)$ is <u>not</u> the distribution of x_t; for the moment, it is just some function needed to explain the second order moment of the x_t series. To get an idea of what $F(\Theta)$ represents, we consider an example:

Let

(11) $$x_t = a \cos \Theta_1 t + b \sin \Theta_1 t \quad , \qquad 0 \leq \Theta_1 \leq \pi \quad ,$$

where a, b are independent random variables with $E(a) = E(b) = 0$, $\mathrm{Var}(a) = \mathrm{Var}(b) = \sigma^2$. Then x_t is itself a random variable, although clearly all possible realizations are trigonometric functions of frequency $\theta_1/2\pi$ with amplitude and phase determined at random. The realization will appear to be completely deterministic, but in the population of possible realizations we have

$$E(x_t) = 0, \quad \mathrm{Var}(x_t) = \cos^2\theta_1 t\,\mathrm{Var}(a) + \sin^2\theta_2 t\,\mathrm{Var}(b)$$

$$= (\cos^2\theta_1 t + \sin^2\theta_2 t)\sigma^2 = \sigma^2 ,$$

and

$$\mathrm{Cov}(x_{t_1}, x_{t_2}) = \mathrm{Var}(a)\cos\theta_1 t_1 \cos\theta_1 t_2 + \mathrm{Var}(b)\sin\theta_1 t_1 \sin\theta_1 t_2$$

$$= \sigma^2 \cos\left[(t_1 - t_2)\right]\theta_1 .$$

Thus we have

$$c(t) = \sigma^2 \cos t\,\theta_1 ,$$

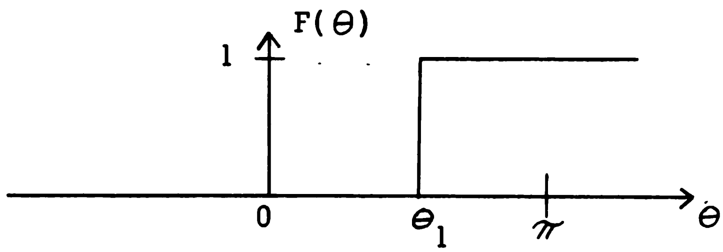
and the model (11) is a stationary (random) time series. Now comparing

$$c(t) = \sigma^2 \cos t\,\theta_1$$

with (10),

$$c(t) = c(0)\int_0^{\pi} \cos(t\,\theta)\,dF(\theta)$$

we see that $F(\theta)$ must have the form



Interpreting this result, $F(\theta)$ shows that the only frequency of interest corresponds to $\theta = \theta_1$, which is obvious from (11). By importance, we mean that the total variance $c(0) = \sigma^2$ can be explained by trigonometric terms, with random coefficients, at the frequency in question. But of course most time series are not of this type. In general, we have that

$$F(\theta_2) - F(\theta_1)$$

represents the proportion of the variance ("energy") of the series which is due to frequencies $\theta/2\pi$ in the range $(\theta_1/2\pi, \quad \theta_2/2\pi)$. When $F(\theta)$ is differentiable, we write

$$f(\theta) = \frac{dF(\theta)}{d\theta}$$

and the above proportion is $\int_{\theta_1}^{\theta_2} f(\theta)\, d\theta$.

$F(\theta)$ is called the spectral distribution, while $f(\theta)$ is called the spectral density. To the engineer, a knowledge of $F(\theta)$ or $f(\theta)$ tells him all he needs to know about the importance of discrete frequencies, or a continuous range of frequencies.

Consider the special case where $f(\theta) = 1/\pi$, $0 \leq \theta \leq \pi$. This uniform density tells us that all frequencies are equally important, and in musical terms this would be "noise" rather than an acceptable musical sound. Technically, the series is called "white noise". Let us consider the consequences. In (10) we have

$$c(t) = c(0)\int_0^{\pi} \cos t\theta \, dF(\theta) = c(0)\int_0^{\pi} \cos t\theta \, f(\theta)\, d\theta$$

$$= c(0)/\pi \int_0^{\pi} \cos t\theta \, d\theta$$

$$= \begin{cases} c(0) & \text{if } t = 0 \\ 0 & \text{if } t = \pm 1, \pm 2, \pm 3, \ldots \end{cases}$$

Thus the covariance of variables at different times is zero. Such would be the case if the $x_1, x_2, \ldots, x_n$ were independent. Hence the case with a uniform spectral density is an extreme one, and is often used as a null hypothesis in significance testing.

But now, how does one estimate the spectral density? If we have a model in mind, e.g.

$$x_t = -\beta x_{t-1} + e_t$$

where the e_t are themselves a white noise process, it can be shown that the spectral density is

$$f(\theta) = 1/\pi \frac{1-\beta^2}{1+2\beta\cos\theta+\beta^2}, \quad 0 \leq \theta \leq \pi,$$

and the obvious way to estimate this is to estimate β by regression analysis and plug in to the formula. The significance test $\beta = 0$ is the same as testing $f(\theta) = 1/\pi$, that is, whether the series is only "white noise" or not.

If however, no model has been established, then one can proceed as follows. Assuming a spectral density exists, by inverting the formula (10) we find

$$(12) \quad f(\theta) = 1/\pi + 2/\pi \sum_{t=1}^{\infty} \frac{c(t)}{c(0)} \cos t\theta = 1/\pi + 2/\pi \sum_{t=1}^{\infty} \rho(t) \cos t\theta .$$

Now if we have n observations, we do not have any observations spaced (in time) by more than $n - 1$ units apart, and so $\rho(n)$, $\rho(n + 1)$, ... cannot be estimated at all. We may be willing to assume these are all zero in view of the long time lags involved. Then we might estimate $f(\theta)$ by

$$(13) \qquad \hat{f}(\theta) = 1/\pi + 2/\pi \sum_{t=1}^{n-1} \hat{\rho}(t) \cos t\theta$$

where the estimate $\hat{\rho}(t)$ is given in (5), (7). Certainly the estimate may be biased because terms have been left out. It suffers from the other disability that it includes estimates $\hat{\rho}(t)$ for t close to $n - 1$, which are unreliable in the sense that large variances may be expected in view of the few pairs of observations that they are based on. A slightly different estimate, but not much better than the one above, is the "periodogram". This can be defined as follows:

$$I_n(\theta) = [2/n]\left\{\left[\sum_1^n (x_t - \bar{x}) \cos\theta t\right]^2 + \left[\sum_1^n (x_t - \bar{x}) \sin\theta t\right]^2\right\} ,$$

(although various authors use different multiplying constants) and can be shown to equal

$$(14) \qquad 2\hat{c}(o) + 4\sum_{t=1}^{n-1} (1 - t/n)\,\hat{c}(t) \cos t\theta \quad .$$

By comparing (14) with (12) we can see that the periodogram actually estimates $2\pi c(0) f(\theta)$; the multiplier $(1 - t/n)$ in (14) reduces the importance of the terms $\hat{c}(t)$ for t large, which is advantageous in view of their bad sampling properties, but may increase the bias of the estimate. One way that the bad effects of the variance of the $\hat{c}(t)$ or the $\hat{\rho}(t)$, can be eliminated is to disregard entirely those based on few observations. Thus instead of (13) one might use

$$\hat{f}(\theta) = 1/\pi + (2/\pi)\sum_{t=1}^{m} \hat{\rho}(t) \cos t\theta$$

where m is much smaller than n. This results in an estimate which has a comparatively low variance, but if the neglected terms are important, will produce a badly biased estimate. One has to make a compromise decision as to the practical importance of the bias and variance. This subject has been studied, and several estimates proposed, by various people, see e.g. [11]. Significance tests against the null hypothesis of white noise, or any other particular density $f(\theta)$, can be made using the periodogram. Some of these tests assume normality of the underlying distribution, a consequence being that the periodogram ordinates at $\theta_j = 2\pi j/n$, $j = 1, 2, \ldots, [\frac{1}{2}(n - 1)]$ have essentially independent χ^2 distributions with 2 degrees of freedom when the series is actually white noise. See [5], [8] for the distribution theory and significance tests for estimates of the spectral density.

<u>SUMMARY</u>. We have made a quick survey of how stationary time series can be analysed in terms of covariances, correlations, or by autoregressive

models; also we have considered how frequency effects can be investigated. We have not discussed the problems associated with observations taken at unequal time intervals, or the vast field of problems with continuous time recording, or the concept of multivariate time series, or the analysis of non-stationary processes. Many of these aspects are partly solved, but remain a fruitful area for research.

REFERENCES

[1] Anderson, R. L. Distribution of the serial correlation coefficient. Ann. Math. Statist. 13, (1942), 1-13.

[2] Bartlet, M. S. Stochastic Processes. Cambridge U.P. (1955).

[3] Daniels, H. E. The approximate distribution of serial correlation coefficients. Biometrika, 43, (1956) 169-185.

[4] Grenander, V. and Rosenblatt, M. Statistical Analysis of Stationary Time Series. Wiley (1957).

[5] Hannan, E. J. Time Series Analysis. Methuen Monograph (1960).

[6] Jenkins, G. H. Tests of hypotheses in the linear autoregressive model. I. Null hypothesis distribution in the Yale scheme. Biometrika, 41 (1954) 405-419.

[7] Jenkins, G. H. Tests of hypotheses in the linear autoregressive model. I. Null distributions for higher order schemes. Non null distributions. Biometrika, 43 (1956) 186-199.

[8] Jones, R. H. Spectral estimates and their distributions. Tech. Report, Dept. of Army Signal Corps DA 36-039 SC-78130/10 (1960).

[9] Jowett, G. H. The comparison of means of sets of observations from sections of independent stochastic series. J. R. Statist. Soc. B 17, (1955) 208-227.

[10] Mann, H. B. and Wald, A. On the statistical treatment of linear stochastic difference equations. Econometrica 11 (1943) 173-220.

[11] Parzen, E. On asymptotically efficient consistent estimates of the spectral density function of a stationary time series. J. R. Statist. Soc. B 20 (1958) 303-322.

[12] Watson, G. S. On the joint distribution of the circular serial correlation coefficients. Biometrika 43 (1956) 101-168.

MONTE CARLO METHODS

J. M. Hammersley
Oxford University and Princeton University

The research worker too often feels that he must say something new and original whenever he says anything, whereas the politician knows quite well that he should hammer away at some old cliché (preferably filched from elsewhere so that it can be disavowed in emergency) until the public believes it. I have filched this remark from a Nobel prizewinner (of whose capacity for original research there can consequently be no doubt), and I propose to follow his advice and to subject you to a political speech this morning.

"The Monte Carlo method," said Dr. Curtiss in a foreword to the proceedings of a symposium on the subject [1], "may briefly be described as the device of studying an artificial stochastic model of a physical or mathematical process. The device is certainly not new. Moreover, the theory of stochastic processes has been a subject of study for quite some time, and the novelty of the Monte Carlo method does not lie here. The novelty lies rather in the suggestion that where an equation arising in a non-probabilistic context demands a numerical solution not easily obtainable by standard numerical methods, there may exist a stochastic process with distributions or parameters which satisfy the equation, and it may actually be more efficient to construct such a process and compute the statistics than to attempt to use those standard methods. Simple and natural as this suggestion seems, once it is made, someone had to make it first in a voice loud enough to attract notice. The voices seem to have been chiefly those of Ulam and von Neumann, though Enrico Fermi, not elsewhere mentioned in these Proceedings, also contributed."

It seems to me entirely right and proper that the credit should go to Ulam, von Neumann, and Fermi in this way, for indeed it was they who showed the world at large what Monte Carlo methods might do. By way of contrast Lord Kelvin, who employed Monte Carlo methods sixty years ago to study the Boltzmann equation [8] and other topics still under active examination at Los Alamos, failed to make himself heard: I would not have known of his work had it not been brought to my attention by Dr. Stephen Brush (of the Radiation Laboratory at Livermore), who has a particular interest in the history of mathematics.

We are accustomed to dividing physicists, say, into theoretical and experimental physicists; but it is a less familiar practice to classify mathematicians in the same way. Nevertheless, experimental mathematics is a genuine subject of study, and Monte Carlo methods comprise one of its principal branches. The essential difference between the theoretician and the experimentalist is that the former postulates and deduces whereas the latter observes and infers. As with other subjects, this is not a dichotomy into exclusive parts; and the work of the theoretical mathematician is an adjunct and a complement to that of the experimental mathematician, and vice versa. In all experimental work, the better the experimental technique the more reliable are the answers.

In a Monte Carlo experiment, the basic observational material consists of random numbers. The experimental technique is to combine these random numbers by some arithmetical procedure to form an estimator, which produces an estimate that is a solution of the problem under study. As a rule, it is not hard to concoct unbiased estimators. The difficult part of the art is find an estimator with a respectably small variance. Of course, variances can be made small by taking large samples; but in general this is not a rewarding course of action, for the size of the final standard error is inversely proportional to the square root of the sample size. To this extent, the relative efficiency of a Monte Carlo procedure may be defined as inversely proportional to the product of the variance of the estimator and the amount of computation expended on obtaining it.

Since this is a political talk, I shall be much concerned with principles. The overriding principle of Monte Carlo work is to cheat: indeed it is this which distinguishes it from straightforward simulation. The game theorists tell us that you cannot guarantee to win a fair game without cheating. Now there is an important difference between the statistician's task in Monte Carlo work and in the handling of data obtained from the average physical or biological experiment. In the latter, there are experimental errors dictated by Nature; and, though every good statistician should utilize efficient experimental designs to mitigate and balance out these errors, there often comes a point beyond which further elimination of the errors is either not possible or would destroy the very purpose of the investigation. But a Monte Carlo experiment is artificial, the creation of the experimenter himself; and, if the errors are immoderately large, that is simply the fault of the experimenter. In Monte Carlo work we can take heed of Lord Rutherford's dictum: "If your experiment needs statistics, you ought to have done a better experiment." In a sense, all good Monte Carlo work is self-liquidating: although we start out with random numbers in order to solve a problem, which may seem to be intractable by conventional numerical analysis, nevertheless we should strive to reduce their influence on the final result, and one should always seize any opportunity to replace a part or even the whole of the sampling experiment by exact analysis.

By way of illustration I shall consider the problem of evaluating integrals. This is not so special a case as it may seem: for most Monte Carlo work treats of the expected values of estimators, and these are simply integrals over sample space. Usually the sample space has a large number of dimensions, if not infinitely many; and this brings its attendant troubles. However, for simplicity I shall work with one dimension only, which will be enough to exhibit those basic principles that I want to discuss this morning.

Suppose then that we have a function $f(x)$ which satisfies

$$0 < f(x) < 1 \,, \tag{1}$$

and that we wish to evaluate

$$\theta = \int_0^1 f(x)\, dx \,. \tag{2}$$

If we draw the curve of $y = f(x)$ in the (x, y) Euclidean plane and construct the unit square S with corners at $(0, 0)$, $(0, 1)$, $(1, 0)$ and $(1, 1)$, then θ is the fraction of the area of S which lies below the curve. There can scarcely be a cruder Monte Carlo procedure than to take a sample of n points, each uniformly and independently distributed over S, to observe the number r of these points which fall below the curve, and to use r/n as an estimator of θ. It is unbiased, distributed binomially, and has sampling variance $\theta(1-\theta)/n$. For purposes of comparison between one method and other, the denominator is superfluous and we shall suppress it and work instead with

$$\theta(1-\theta), \tag{3}$$

which represents the reciprocal of the efficiency of the method inasmuch as it is proportional to the product of the sampling variance and the amount of work expended in obtaining the estimate.

An alternative procedure would be to take a random number ξ, where here and hereafter ξ with or without a suffix denotes a number uniformly distributed between 0 and 1, and to use

$$t = f(\xi) \tag{4}$$

as an estimator of θ. More generally, in practice one would use the mean of n such quantities (4); but, as already explained, the case $n = 1$ provides us with all the comparative information we need. If we replace x by ξ in (2), we see at once that (4) is unbiased. Its sampling variance is

$$\int_0^1 f^2 \, dx - \theta^2. \tag{5}$$

By (1), we have $f^2 < f$; and hence in comparison with (3)

$$\int_0^1 f^2 \, dx - \theta^2 < \int_0^1 f \, dx - \theta^2 = \theta - \theta^2 = \theta(1 - \theta). \tag{6}$$

Consequently (4) provides a better estimator than we got from choosing a point uniformly at random in S. The improvement comes from using the exact value of f corresponding to the random ξ, instead of the less exact information on the relative position (below or above) of the point to the curve. This is an example of partial replacement of experiment by a piece of calculation or exact analysis. The relative efficiency of the two methods comes from the ratio of (3) to (5), adjusted by some factor representing the relative amounts of computing time: in one method, we need to choose two random numbers to fix a point in the square and we have to decide whether the point is above or below the curve; in the other method we have only one random number to choose, but we have to calculate f exactly for this random argument. The adjustment will depend upon the form of the function f; and I leave it to you to make the assessment in various particular cases according to your taste.

For reference purposes, I shall call (4) crude Monte Carlo. The other method of choosing a point in the square is so abysmally bad that it does not merit a name.

Evidently the sampling errors in the crude Monte Carlo method arise from fluctuations in the value of $f(\xi)$ as ξ ranges over its possible values from 0 to 1. If f is a reasonably smooth function, it will undergo less fluctuation in a shorter interval. This suggests breaking the range of integration up into several pieces by points $0 = a_0 < a_1 < \ldots < a_m = 1$, estimating the several integrals over the respective subintervals by crude Monte Carlo methods, and finally adding the results together to obtain an estimate of θ. Simple linear transformations make allowance for the changed lengths of the subintervals, and from these we obtain the estimator

$$t = \sum_{j=1}^{m} (a_j - a_{j-1}) f[a_{j-1} + (a_j - a_{j-1}) \xi_j]. \tag{7}$$

This method is known as stratified Monte Carlo. It is reminiscent of the sundry linear formulae (trapezoidal rule, Simpson's rule, etc.) which occur in classical numerical analysis for evaluating integrals, and indeed it shares much in common with them. The question naturally arises of how we should choose the numbers a_j to make the method as efficient as possible. A full discussion of this would take me too far afield, so I shall content myself with saying that (i) the problem is substantially the same as that encountered in sampling surveys and can be found under the heading of stratified sampling in the standard textbooks on sampling methods, and (ii) broadly speaking, a pretty good procedure is to choose the subintervals to equalize the variation of f in each subinterval.

Next consider rewriting (2) in the form

$$\theta = \int_0^1 f(x)\, dx = \int_0^1 \frac{f(x)}{g(x)}\, g(x)\, dx, \tag{8}$$

where $g(x)$ is a frequency function over the interval $(0, 1)$: that is to say

(9) $$g(x) > 0 \quad \text{and} \quad \int_0^1 g(x)\, dx = 1.$$

It follows from (8) that, if η is a random variable from the distribution whose frequency function is g, then

(10) $$t = f(\eta)/g(\eta)$$

is an unbiased estimator of θ. This holds for any g, and we would like to know how to choose g to minimize the sampling variance. According to the principles already explained, we shall get a small variance if t in (10) is practically constant. Indeed we might try to make t exactly a constant by choosing $g = cf$, where c is a constant. We can determine the constant c from (2) and (9):

(11) $$1 = \int_0^1 g(x)\, dx = \int_0^1 cf(x)\, dx = c\,\theta.$$

Thus we want to sample from the distribution whose frequency function is $f(x)/\theta$; and we could do this if only we knew θ, the answer to the problem we are engaged on solving! This is asking too much: and in practice we compromise by selecting g to meet as best as we can the two conflicting requirements:-

(a) $g(x)$ must be a simple enough function for us to be able to integrate it analytically and thereby ensure that is satisfies (9), or normalize if it does not already satisfy (9); and

(b) $g(x)$ must be a reasonably good mimic of the function $f(x)$, whose integral we do not know, so that t in (10) is substantially constant at least over a good part of the range of integration.

If we succeed in satisfying (b) to a reasonable extent, then g will be large when f is, and accordingly the majority of values η will cluster around the points where f has its largest or most important values. For this reason (10) is known as importance sampling.

The expected value of the sum of several random variables is the sum of their expectations, whether or not these random variables are independent. Hence no bias will be introduced if in (7) we make the various ξ_j dependent. At the same time dependence can reduce the variance. For instance, consider the case $m = 2$, $a_1 = a$, $\xi_1 = 1 - \xi_2 = \xi$, for which (7) becomes

$$(12) \qquad t = af(a\xi) + (1-a)\, f[1-(1-a)\xi] = T_a f(\xi),$$

where T_a is the functional operator defined by (12). If f is a monotone function, the two terms in the central member of (12) will tend to balance one another out, one being high when the other is low; and the variance of t will be correspondingly reduced. Because the two terms of (12) reduce the variance by acting against each other, this method is known as the antithetic variate method. A different kind of dependence, appropriate for integrands with a hump or a trough in the middle is $\xi_1 = \xi_2 = \xi$; and (7) would instead become

$$(13) \qquad t = af(a\xi) + (1-a)\, f[a + (1-a)\xi] = S_a f(\xi).$$

In general, this second transformation is not so powerful a means of reducing variance as the antithetic variate transformation. However, very striking gains of efficiency may result from a combination of both methods. In each case we may ask how a should be chosen. Now (12) shows that antithetic variate estimation applied to the integrand f is equivalent to crude Monte Carlo applied to the transformed function $T_a f$; and the latter will have a small variance if we can choose a to make $T_a f$ as constant as possible. As a rough and ready means of achieving this, we might equate the values of $T_a f$ at the ends of the range of integration. This gives $a = \alpha$ where α is the root of

$$(14) \qquad f(\alpha) = (1-\alpha)\, f(1) + \alpha f(0).$$

The choice of $\underline{a}$ in (13) is a more difficult matter, discussed in the original paper on the subject [6]. However, if (12) and (13) are to be used in combination with one another, the appropriate thing is to apply T_α first, where α is given by (14), followed by as many applications of $S_{1/2}$ as may be desired.

As an example of the foregoing remarks, consider the case

(14) $$f(x) = \sin(\pi x/2).$$

The sampling variances (per random number used) and the ratios of these to the variance for the crude Monte Carlo method are:-

Method	Sampling Variance	Ratio	Rough estimate of computing time	Efficiency
Crude Monte Carlo	$9{\cdot}5 \times 10^{-2}$	1.0	1	1.0
Importance sampling	$6{\cdot}8 \times 10^{-3}$	1.4×10^{1}	3	4.7
Antithetic T_α	$3{\cdot}6 \times 10^{-4}$	2.7×10^{2}	2	1.3×10^{2}
Antithetic $S_{1/2}T_\alpha$	$2{\cdot}2 \times 10^{-5}$	4.4×10^{3}	4	1.1×10^{3}
Antithetic $S^3_{1/2}T_\alpha$	$8{\cdot}4 \times 10^{-8}$	1.1×10^{6}	16	6.9×10^{4}

The importance function used for the second line of the above table was $g(x) = 2x$, which bears about the right relative degree of simplicity to f encountered in real applications. In the final line (T_α followed by three applications of $S_{1/2}$) there will be 16 observations per random number used: thus the variance is reduced by a factor of a million at the expense of a sixteenfold increase in computing, which means an overall efficiency gain of about 70,000. Further examples, including a detailed discussion of a genuine application to a six dimensional integral, appear in [6]. The theory of the subject ([5], [7]) shows that we need only consider dependencies of the forms stated above, namely $\xi_1 = \xi_2$ or $\xi_1 = 1 - \xi_2$. Further efficiency gains result from using values of the integrand outside the range of integration [3].

In the methods discussed above, there is usually some choice (e.g., choice of importance function, choice of stratification point a in the antithetic variate methods, etc.). If the choice is not made in an optimum fashion, there will be some loss of efficiency (though seldom much loss because of the flatness of minima); but there will be no introduction of bias however bad the choice. To this extent, the methods are robust.

There is of course no reason why one should not apply several methods simultaneously; and there are a variety of useful methods which I have omitted from this talk. It is also worth using numbers which are not random on certain occasions. This and other questions are discussed in [4].

True to political tenets, I have said nothing profound or new; but I hope I have exposed a few of the tricks of the trade.

REFERENCES.

[1] J. H. CURTISS et al. "Monte Carlo method" Nat. Bur. Stand. Appl. Math. Ser. 12 (1951).

[2] J. H. HALTON "On the efficiency of certain quasi-random sequences of points in evaluating multidimensional integrals" Numerisch Mathematik 2 (1960) 84-90.

[3] J. H. HALTON and D. C. HANDSCOMB "A method for increasing the efficiency of Monte Carlo integration" J. Assoc. Comp. Mach. 4 (1957) 329-340.

[4] J. M. HAMMERSLEY "Monte Carlo methods for solving multivariable problems" Ann. New York Acad. Sci. 86 (1960) 844-874.

[5] J. M. HAMMERSLEY and J. G. MAULDON "General principles of antithetic variates" Proc. Camb. Phil. Soc. 52 (1956) 476-481.

[6] J. M. HAMMERSLEY and K. W. MORTON "A new Monte Carlo technique: antithetic variates" Proc. Camb. Phil. Soc. 52 (1956) 449-475.

[7] D. C. HANDSCOMB "Proof of the antithetic variates theorem for $n > 2$" Proc. Camb. Phil. Soc. 54 (1958) 300-301.

[8] Lord KELVIN "Nineteenth century clouds over the dynamical theory of heat and light" Phil. Mag. (6) 2 (1901) 1-40.

[9] K. W. MORTON "A generalization of the antithetic variate technique for evaluating integrals" J. Math. and Phys. 36 (1957) 289-293.

[10] J. W. TUKEY "Anthithesis or regression?" Proc. Camb. Phil. Soc. 53 (1957) 923-924.

THE CONSTRUCTION AND ANALYSIS OF NON-ORTHOGONAL PLANS FOR THE 2^n FACTORIAL EXPERIMENTS

Sidney Addelman
Research Triangle Institute

INTRODUCTION. There are many situations in which an experimenter must estimate the important effects of a symmetrical factorial experiment with as few trials as possible. There are instances where, say, all two-factor interactions are important and the orthogonal plan necessary to estimate these parameters requires more trials than one can afford. If the experimenter is restricted to orthogonal fractional replicate plans he must then either abandon the investigation or lose information on some of the interactions that may be important.

Consider, for example, a situation where it is desirable to estimate the main effects and two-factor interaction effects of the 2^7 experiment and no more than 50 trials can be made. It is well known that a 1/2 replicate of the 2^7 experiment allows orthogonal estimates of all main effects and two-factor interactions and that a 1/4 replicate does not. The 1/2 replicate plan requires 64 trials and the 1/4 replicate plan requires 32 trials. It then seems reasonable to inquire whether a plan with 48 trials can be constructed that yields information on all main effects and two-factor interaction effects. The consideration of all possible subsets of 48 treatment combinations from the totality of 128 possible combinations in the 2^7 experiment is a tedious task. Therefore, an investigation has been made of the use of subsets which consist of a number, k, of the possible 2^m distinct $1/2^m$ fractional replicates defined by a particular identity relationship.

DEVELOPMENT OF NON-ORTHOGONAL PLANS. The treatment combinations of the 2^n experiment may be represented as the points of a n-dimensional lattice with axes $x_1, x_2, \ldots, x_n$. Each axis of the lattice will have two points, 0 and 1. Thus, for example, the treatment combinations of the 2^2 factorial experiment, with factors A and B, which are usually represented by (1), a, b and ab can also be represented by the points (0, 0), (1, 0), (0, 1) and (1, 1), respectively. The points (0, 0) and (1, 1) both satisfy the equation $x_1 + x_2 = 0$ (modulo 2) and comprise the 1/2 replicate of the 2^2 experiment defined by the identity relationship $I = AB$. The symbol AB_0 can be used to denote the set of treatment combinations for which $x_1 + x_2 = 0$ (modulo 2). Similarly AB_1 denotes the set of treatment combinations for which $x_1 + x_2 = 1$ (modulo 2).

Consider n factors, A, B, C, ..., L, each with two levels. Then $ABC \cdots L_i$ denotes the set of treatment combinations for which $x_1 + x_2 + \ldots + x_n = i$, where $i = 0$ or 1. A 0 or 1 can be associated with every effect or interaction of the identity relationship. For example, the 4 possible 1/4 replicates of the 2^5 experiment defined by the identity relationship

$$I = ABC = ADE = BCDE$$

can be represented by the four relationships

$$I = ABC_0 = ADE_0 = BCDE_0$$

$$I = ABC_0 = ADE_1 = BCDE_1$$

$$I = ABC_1 = ADE_0 = BCDE_1$$

$$I = ABC_1 = ADE_1 = BCDE_0$$

or can be displayed in tabular form as in Table 1.

TABLE 1

STRUCTURE OF THE 1/4 REPLICATES OF A 2^5 EXPERIMENT

Identity relationship	Fractional replicate			
	1	2	3	4
I				
ABC	0	0	1	1
ADE	0	1	0	1
BCDE	0	1	1	0

Since the interaction BCDE is the generalized interaction of ABC and ADE the subscripts associated with BCDE in each of the four fractional replicates may be obtained as the sum (modulo 2) of the corresponding subscripts of ABC and ADE.

The treatment combinations which constitute a 3/4 replicate plan of the 2^5 experiment may be obtained by selecting the treatment combinations that occur in any 3 of the four 1/4 replicates given in Table 1.

If each of the k subscripts associated with a member of the identity relationship defining a $k/2^m$ replicate of the 2^n experiment are identical, then that effect is completely confounded with the mean, μ. If the k subscripts are not identical then that member of the identity relationship is partially confounded with the mean. If, for any interaction, k is an even number and half of the subscripts are 0 and the other half are 1, then that interaction is unconfounded with the mean.

The following theorem, which can be easily verified, is helpful in the construction of non-orthogonal plans.

<u>Theorem 1</u>. In a $k/2^m$ replicate plan for the 2^n factorial experiment no main effect or interaction need be completely confounded with the mean if $k \geq (m + 1)$.

<u>Corollary</u>. In a $k/2^m$ replicate plan for the 2^n experiment, if $k = (m - u)$, where $u = 0, 1, 2, \ldots$, then $(u + 1)$ interactions and their

generalized interactions will be completely confounded with the mean.

If $k = (m - u)$, it is often possible to construct a non-orthogonal plan so that the interactions which are completely confounded with the mean will contain at least five factors. This can always be arranged if there exists a $1/2^{u+1}$ replicate of the 2^n experiment which permits uncorrelated estimates of the main effects and all two-factor interaction effects, when higher order interactions are negligible. These plans are sometimes called plans of Resolution V. If it is not possible to have only five-factor and higher order interactions completely confounded with the mean, then some two-factor interactions will not be estimable.

Of the k subscripts associated with a member of the identity relationship defining a $k/2^m$ replicate of the 2^n experiment let t be 0 and $(k - t)$ be 1. The following theorem can easily be verified:

Theorem 2. If an interaction of the identity relationship defining a $k/2^m$ replicate of the 2^n experiment has an odd number of factors, the off-diagonal elements of the information matrix, corresponding to the partially confounded effects that are determined by the interaction, are equal to $(k - 2t)\,2^{n-m}$. If an interaction has an even number of factors, the off-diagonal elements of the information matrix associated with that interaction are equal to $-(k - 2t)\,2^{n-m}$.

We shall adopt the rule that when k is odd, an odd-factor interaction will have an odd number of the k subscripts associated with it equal to 1 and an even-factor interaction will have an even number of the k subscripts associated with it equal to 1. With this rule the off-diagonal elements of the information matrix associated with each member of the identity relationship, for which the absolute value of $(k - 2t)$ is constant, will have the same value when k is odd. When k is even it may often be desirable to have $(1/2)k$ subscripts equal to 0 and $(1/2)k$ subscripts equal to 1 for some interactions so that they are unconfounded with the mean.

If we also adopt the procedure of grouping the effects and interactions of interest in such a way that those which are partially confounded with each other are contiguous, then the information matrix will consist of various sized blocks about the main diagonal with off-diagonal blocks of zero elements. The variance-covariance matrix can then be obtained by inverting each of the blocks about the diagonal separately.

If the above procedures are followed, the blocks which lie about the diagonal of the information matrix of a $3/2^m$ replicate of the 2^n experiment will be of the form

$$2^{n-m+2} I - 2^{n-m} J$$

where I is the identity matrix of rank p and J is a $p \times p$ matrix of l's. It is easily verified that the inverse of this matrix, the variance-covariance matrix, is

$$\frac{1}{2^{n-m+2}} \left[I + \frac{1}{4-p} J \right]$$

When $p = 2$ the block on the diagonal of the variance-covariance matrix is

$$\frac{1}{2^{n-m+3}} \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix}$$

and when $p = 3$ the block is given by

$$\frac{1}{2^{n-m+2}} \begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{bmatrix}$$

It is evident from the form of the variance-covariance matrix that if for any block in the information matrix $p = 4$, then that block must be singular so that an inverse of that block does not exist. In such a situation one must assume that one of the interactions in that block is negligible and thus reduce the size of the block to a 3×3, making it non-singular.

The yield of a treatment combination in a $k/2^m$ replicate plan for the 2^n experiment can be written in terms of the main effects and interactions:

$$y_{ijk\ldots} = \mu \pm \frac{1}{2}A \pm \frac{1}{2}B \pm \frac{1}{2}AB \pm \frac{1}{2}C \pm AC \pm \text{etc.} + \text{error}$$

where the sign

on A is - if $i = 0$ and + if $i = 1$

on B is - if $j = 0$ and + if $j = 1$

on C is - if $k = 0$ and + if $k = 1$

and so on,

and the sign on a term involving several letters is the product of the signs on the individual letters.

The estimates of the $3/2^m$ replicate plan for the 2^n experiment can be shown to have the following forms:

(i) If X, Y, and Z denote effects and/or interactions that are partially confounded with each other, their estimates are given by

$$\begin{bmatrix} \hat{X} \\ \hat{Y} \\ \hat{Z} \end{bmatrix} = \frac{1}{2^{n-m+1}} \begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{bmatrix} \begin{bmatrix} [X] \\ [Y] \\ [Z] \end{bmatrix}$$

where [X] denotes the sum of the treatment combinations whose expectations contain X positively, minus the sum of the combinations whose expectations contain X negatively, [Y] and [Z] being similarly defined. The variances and covariances are obtained from the variance-covariance matrix and can be shown to be

$$\operatorname{var}(\hat{X}) = \operatorname{var}(\hat{Y}) = \operatorname{var}(\hat{Z}) = \sigma^2/2^{n-m-1}$$

$$\operatorname{cov}(\hat{X},\hat{Y}) = \operatorname{cov}(\hat{X},\hat{Z}) = \operatorname{cov}(\hat{Y},\hat{Z}) = \sigma^2/2^{n-m}$$

(ii) If μ, X and Y are partially confounded with one another, their estimates are given by

$$\begin{bmatrix} 2\hat{\mu} \\ \hat{X} \\ \hat{Y} \end{bmatrix} = \frac{1}{2^{n-m+1}} \begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{bmatrix} \begin{bmatrix} T \\ [X] \\ [Y] \end{bmatrix}$$

where T denotes the sum of all treatment combinations.

$$\operatorname{var}(\hat{X}) = \operatorname{var}(\hat{Y}) = \sigma^2/2^{n-m-1}$$

$$\operatorname{cov}(2\hat{\mu},\hat{X}) = \operatorname{cov}(2\hat{\mu},\hat{Y}) = \operatorname{cov}(\hat{X},\hat{Y}) = \sigma^2/2^{n-m}$$

(iii) If X and Y are partially confounded with each other and with no others, their estimates are given by

$$\begin{bmatrix} \hat{X} \\ \hat{Y} \end{bmatrix} = \frac{1}{2^{n-m+2}} \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix} \cdot \begin{bmatrix} [X] \\ [Y] \end{bmatrix}$$

$$\text{var}\ (\hat{X}) = \text{var}\ (\hat{Y}) = 3\sigma^2/2^{n-m+1}, \quad \text{cov}\ (\hat{X},\hat{Y}) = \sigma^2/2^{n-m+1}$$

(iv) If μ and X are partially confounded with each other and no others, their estimates are given by

$$\begin{bmatrix} 2\hat{\mu} \\ \hat{X} \end{bmatrix} = \frac{1}{2^{n-m+2}} \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix} \cdot \begin{bmatrix} T \\ [X] \end{bmatrix}$$

$$\text{var}\ (\hat{X}) = 3\sigma^2/2^{n-m+1}, \quad \text{cov}\ (2\hat{\mu},\hat{X}) = \sigma^2/2^{n-m+1}.$$

The correlation of the estimates in a $3/2^m$ replicate of the 2^n experiment which are partially confounded with each other can be shown to equal $1/(5 - p)$, where $1 < p < 5$. Hence, if $p = 2$, the correlation is equal to 1/3 and if $p = 3$ the correlation is equal to 1/2.

<u>EXAMPLES OF NON-ORTHOGONAL PLANS</u>. The structure of some useful non-orthogonal plans are presented in the following tables.

TABLE 2

A 3/4 REPLICATE OF THE 2^4 EXPERIMENT

Identity relationship	Fractional replicate		
	1	2	3
I			
ABC	0	0	1
ABD	0	1	0
CD	0	1	1

TABLE 3

A 3/8 REPLICATE OF THE 2^7 EXPERIMENT

Identity relationship	Fractional replicate		
	1	2	3
I			
ABCDE	1	1	1
ABF	0	0	1
CDEF	1	1	0
AEG	0	1	0
BCDG	1	0	1
BEFG	0	1	1
ACDFG	1	0	0

TABLE 4

A 3/16 REPLICATE OF THE 2^8 EXPERIMENT

Identity relationship	Fractional replicate 1	2	3
I			
ABCDE	1	1	1
ABFGH	1	1	1
CDEFGH	0	0	0
ACF	0	0	1
BDEF	1	1	0
BCGH	1	1	0
ADEGH	0	0	1
BEG	0	1	0
ACDG	1	0	1
AEFH	1	0	1
BCDFH	0	1	0
ABCEFG	0	1	1
DFG	1	0	0
CEH	1	0	0
ABDH	0	1	1

A USEFUL APPLICATION OF THE NON-ORTHOGONAL PLANS. In the application of factorial patterns there sometimes arise situations where the experimenter feels that one or more additional factors should have been included in an experiment which has just been performed. Using the non-orthogonal plans described in this paper, the original plan can be augmented to yield information on the additional factors, without losing the information obtained with the original design.

If the original experimental plan was a 2^3 factorial arrangement and an additional factor must be added, it can be assumed that the additional factor was held at its 0 level in the original plan and that the original 8 trials were a 1/2 replicate of the 2^4 experiment. By adding 4 treatment combinations which constitute another 1/4 replicate of the 2^4 experiment, the additional factor, as well as the interaction of that factor with the original three factors, can be estimated. The augmented plan is a 3/4 replicate of the 2^4 experiment. A fifth factor can be introduced by adding 4 more treatment combinations and the resulting plan is a 4/8 replicate of the 2^5 experiment. If a sixth factor is introduced, the resulting plan will be a 5/16 replicate of the 2^6 experiment. When more than one factor is added to the original orthogonal plan the resulting non-orthogonal plan does not permit the estimation of the interactions among the additional factors.

Table 5 gives the structure of the 2^3 plan augmented to a 5/16 replicate of the 2^6 experiment. Imbedded in this plan are a 4/8 replicate of the 2^5 and a 3/4 replicate of the 2^4 experiments. The letters A, B and C denote the original 3 factors and D, E and F denote the additional factors.

TABLE 5

STRUCTURE OF A 2^3 PLAN AUGMENTED TO A 5/16 REPLICATE OF THE 2^6 EXPERIMENT

Identity relationship	Fractional replicate 1	2	3	4	5
I					
D	0	0	1	0	0
ABC	0	1	0	1	1
ABCD	0	1	1	1	1
E	0	0	0	1	0
DE	0	0	1	1	0
ABCE	0	1	0	0	1
ABCDE	0	1	1	0	1
F	0	0	0	0	1
DF	0	0	1	0	1
ABCF	0	1	0	1	0
ABCDF	0	1	1	1	0
EF	0	0	0	1	1
DEF	0	0	1	1	1
ABCEF	0	1	0	0	0
ABCDEF	0	1	1	0	0

USE OF THE UP-AND-DOWN METHOD WITH FACTORIAL DESIGNS

R. L. Grant and R. W. Van Dolah
Explosive Research Laboratory, Bureau of Mines
Pittsburgh, Pennsylvania

The adaptation of the up-and-down method to experimental designs for which it might be appropriate has been hampered by the fact that at least 50 trials per sequence are usually recommended (1).* Where the individual trials are expensive, as for explosives and ordnance experimentation, the cost of such long sequences becomes prohibitive. During the past several years, the Explosives Research Laboratory of the Bureau of Mines has sought to exploit the advantages of the up-and-down method and of factorial experiments by combining the two into a single design and to overcome the cost objection of the up-and-down method by the economical use of shorter sequences. The principle of the model based on the combination design is to assume a factorial design and accept as particular response values the means as calculated from the corresponding randomized up-and-down sequences. Since the up-and-down method is founded on an efficient design for estimating a mean and since factorial experiments are based on efficient designs for evaluating factor effects, a combination of the two should also be relatively efficient. Such a design should be applicable to situations of wide occurrence in explosives and ordnance experimentation where experiments are conducted of the go-no-go, or success-failure, type and where it is desired to study the effects of one or more factors. However, unless the sequences are relatively short the number of trials and the cost of the experiment will be large for this design. In the course of this work, attention was given to meeting the requirements of the up-and-down method in order to increase its efficiency and thus permit shorter sequences. These requirements are that the basic distribution must be normal, its standard deviation be known approximately, and the up-and-down interval be approximately equal to the standard deviation.

This paper reviews the work conducted at the Bureau with this combination model for experiments with coal mine explosives. The statistical principles are described and the application of the design is illustrated with five factorial experiments for evaluating certain effects which influence the safety of these explosives.

---

*Underlined numbers in parentheses refer to items in the list of references at the end of this report.

THEORY. A conventional two-factor experiment is assumed; the design for this is illustrated in Figure 1a.* Here there are p levels of factor A and r levels of factor C. The response, or experimental observation, for a particular combination of A and C is designated by Y, where Y_{11} denotes the combination of factors A and C each at the first level and Y_{ij} denotes the general response. The model is

$$\text{(1a)} \qquad Y_{ij} = \mu + \alpha_i + \gamma_j + (\alpha\gamma)_{ij} + \epsilon_{ij}$$

$$\text{(1b)} \qquad = m + a_i + c_j + (ac)_{ij} + e_{ij}$$

where equation (1a) represents the universe and (1b) the sample estimate. (The notation of Anderson and Bancroft (3) has been followed.)

To obtain a response Y an up-and-down sequence is performed according to the design illustrated in Figure 1b. Sequence 1 is that for combination A_1C_1 and sequence n for A_pC_r. Each up-and-down sequence yields a mean and is generally designated as an $\bar{X}$. The mean of any given sequence is assumed as a Y response for use in the design of Figure 1a. Or,

$$\text{(2)} \qquad \bar{X}_{11} \text{ of Figure 1b} = Y_{11} \text{ of Figure 1a,}$$

$$\bar{X}_{ij} \text{ of Figure 1b} = Y_{ij} \text{ of Figure 1a.}$$

Thus, the mean as determined by the up-and-down sequence is accepted as a measurement of sensitivity of the explosive or item of ordnance under study and this measurement serves as the corresponding response value for the factorial experiment.

It will be noted that there are pr combinations in the factorial design of Figure 1a and therefore there must be the same number of up-and-down sequences in Figure 1b. That is,

*Figures appear at the end of the article.

$$pr = n \tag{3}$$

is a property of the combination design represented by equations (1) and (2). The experiments described in this report were conducted according to this combination design. It is clear that although the two-factor design has been cited, the combination design may be composed of the up-and-down design and any factorial design.

Randomization is achieved as follows. Essentially, the n up-and-down sequences are conducted simultaneously and in random order. To illustrate, if p is 3 and r is 3, then n will be 9. Numbers from 1 to 9 are assigned to the 9 different combinations randomly and the individual trials, or shots for explosives, made in order of the assigned number.

Orthogonality, or rectangular symmetry, of the factorial design of Figure 1a is assured by the requirement that each up-and-down sequence of Figure 1b be of the same length. This means that there should be the same number of yes-no pairs in each sequence. For example, 10 yes-no pairs will require a minimum of 20 trials, or shots, with explosives.

The method of calculation for the mean of each up-and-down sequence is given by Dixon and Massey (2). After entering the appropriate response values in the factorial design, this is analyzed in the usual way (3, 8).

VALIDITY OF THE UP-AND-DOWN METHOD. The following three conditions must be satisfied for valid application of the up-and-down method (2, 4):

(1) The underlying, or stimulus, variable of the experiment must be normally distributed.

(2) The standard deviation of this variable must be known, at least approximately.

(3) The predetermined intervals of this variable employed in the up-and-down steps should be within the range of one-half to two standard deviations.

Although the normality requirement may be examined best by conducting relatively long up-and-down preliminary sequences, these cannot always be justified. Short sequences or sequences which are an integral part of the investigation will then be advisable. In this investigation a test for normality was made with a 60-shot sequence with explosive A, a selected permissible explosive (Figure 2). The size of the interval between the steps was based on the logarithm of the charge weight of the explosive expressed in grams. At the time this experiment was started, the best estimate of the standard deviation was 0.06 log units and accordingly this was chosen for the steps of the sequence. A plot of the distribution of the ignitions, or Y's, is shown in Figure 3 as a histogram. The test for normality using the chi-square method described in (1, 5) permitted the conclusion that this distribution was acceptably fitted by a normal curve. The best-fitting normal curve is drawn in Figure 3.

This sequence gave an estimate of the standard deviation of 0.081 log units, thus fulfilling requirement (2). Therefore, for requirement (3) an interval of magnitude between half and twice this estimate, or 0.04 and 0.16, would be reasonably satisfactory for the up-and-down steps. The actual interval of 0.06 used for this sequence was evidently well chosen.

The sequence of Figure 2 also provided information concerning the reliability of short sequences. To obtain this, the 60-shot sequence is first subdivided into three equal sequences each with 10 yes-no pairs. For the 30-pair sequence and for each of the 10-pair sequences, 95 percent confidence limits of the respective means were calculated. The results of these calculations are plotted as intervals in Figure 4a. The assumption is made that the best estimate available from the data of the 95 percent confidence interval for a 10-pair, or 20-shot, sequence is the confidence interval based on the 30-pair, or 60-shot, sequence adjusted to 10 pairs. This adjustment to 10 pairs is made by multiplying the 95 percent confidence interval of 30 pairs by $\sqrt{30}/\sqrt{10}$, or 1.732. These adjusted confidence limits are shown as dashed lines in Figure 4a. The mean of any 20-shot up-and-down sequence should fall within this adjusted interval and, if such is the case, one may conclude that satisfactory reproducibility of the mean has been achieved. Figure 4a shows that each of the three means based on 10 pairs, or 20 shots, were within the adjusted interval.

Three additional 60-shot, or 30-pair, sequences were made with the same explosive and test procedure (Figures 4b, c, and d). Each mean based on 20 shots, or 10 pairs, fell within the corresponding adjusted interval. Thus, these 240 shots indicated that satisfactory reproducibility of a mean based on a minimum of 20 shots was achieved 12 out of 12 times with the up-and-down method. This, of course, reflects an important property of this method, namely, that the mean is measured with increased efficiency.

Therefore, the conclusions were made that the requirements of the up-and-down method were being met and that sequences of a minimum length of 20 shots would be reasonably satisfactory for our experiments.

FACTORIAL EXPERIMENTS WITH EXPLOSIVES. Five factorial experiments have been conducted using the combination design described above as the model. Each experiment dealt with some aspect of the safety of coal mine explosives in the presence of a flammable atmosphere. The explosives were either special formulations prepared by explosives manufacturers for the Bureau of Mines or commercial (permissible) explosives. The charges of explosives were fired from steel cannons into an explosive mixture of natural gas in air. Figure 5 shows the operator loading the explosive charge into the borehole of the steel cannon. The weight of the charge varied from 150 to 1,300 grams. On the left of the photo is shown the end of the steel gallery which contains the natural gas in air at the time of the shot. Figure 6 presents an overall view of the gallery taken at the instant of ignition of the gas in the gallery by an explosive charge fired into it. The volume of flammable gas which is ignited is 625 cubic feet. The philosophy of this testing procedure is based on the hypothesis that safe explosives will require relatively large charge weights to ignite the gas and dangerous explosives will require smaller charge weights. Accordingly, this is a sensitivity experiment because there will be a critical charge weight above which more or less consistent ignitions will be obtained and below which nonignitions will result. For such a test, the up-and-down method is suitable. The experiments presented below were run over a period of several years and can be described only briefly.

Experiment 1. The object was to study the effects of the particle size of the ammonium nitrate and the types of carbonaceous material of the explosive on their incendivity to the natural gas atmosphere. There were three grades of particle size and five types of carbonaceous material. Accordingly, this was designed as a 3 x 5 factorial with two replications in order to estimate interaction between the two factors. The results of the experiment are given in Table 1a in which each number in the body of the table is a W_{50} value. This is defined as the mean weight of explosive as determined from an up-and-down sequence consisting of a nominal 20-shot sequence. This weight will be expected to produce ignitions 50 percent of the time. A total of 612 shots were fired in the course of this experiment. Table 1b presents the analysis of variance. The conclusions were that: (a) The particle size of the ammonium nitrate in permissible-type explosives has a highly significant effect on the incendivity of the explosive to natural gas (methane plus ethane) in air, with coarse ammonium nitrate producing less incendive explosives than the fine. (b) The type of carbonaceous material and the interaction between the two main factors have no significant effect on the incendivity (6).

Experiment 2. Several new types of stemming for holding and increasing the confinement of the explosive in the coal borehole in mines have been proposed. A study was made of the safety characteristics of these by determining their effectiveness in reducing the likelihood of ignition of the gas. Table 2a shows the combined results of three separate factorial experiments with a total of seven types of stemming materials and nine samples of explosives. However, the question of interest was the behavior of a given stemming material with the standard, or control, stemming and, accordingly, these results were analyzed by comparing a particular stemming with the standard stemming, which was one pound of dry fireclay. The analyses of variance for the principal comparisons are given in Table 2b. The conclusions were as follows: The stemming materials which were significantly better than the standard fireclay were dry sodium chloride in an asbestos container, ordinary water, gelled water, and a saturated sodium chloride solution in water, each contained in a plastic bag. The wet fireclay did not differ significantly from the standard dry fireclay but the special asbestos stemming device was distinctly inferior, and therefore relatively hazardous, as compared with the standard fireclay (7).

Experiment 3. A two-factor 3 x 7 experiment was performed to establish the percentage of natural gas which represented the mixture of maximum ignitibility when the explosives were fired into the gas-air mixtures. The combustibles in the natural gas were analyzed and expressed as methane plus ethane. In this factorial experiment one factor was the gas concentration at 7 levels, 7.0, 7.5, to 10.0 percent, and the second factor was three typical permissible explosives. The principal object was to determine whether the gas concentration had a significant effect on the result expressed as a W_{50} value of the explosive and, if so, to ascertain gas concentration representing maximum ignitibility. Table 3a shows the results and Table 3b the analysis of variance. The conclusion was that

TABLE 1a. - Results of Experiment 1, a 3 x 5 factorial

The upper figure is the result of the first replicate and the lower figure is the result of the second replicate. Each result is a W_{50} value in grams of explosive.

	Ammonium nitrate			
	Fine	Medium	Coarse	Total
Carbonaceous material:				
Wood meal	436	424	514	
	374	487	480	2,715
Fine bagasse	467	493	514	
	401	473	507	2,855
Coarse bagasse	412	480	507	
	401	480	493	2,773
Starch	436	473	500	
	418	487	487	2,801
Walnut meal	473	487	473	
	467	500	529	2,929
Total, first replicate	2,224	2,357	2,508	7,089
Total, second replicate	2,061	2,427	2,496	6,984
Total, both replicates	4,285	4,784	5,004	14,073
Means	428.5	478.4	500.4	

TABLE 1b. - Results of analysis of variance

Source of variance	Sum of squares	Degrees of freedom	Mean squares	F	$F_{.05}$	$F_{.01}$
Particle size of ammonium nitrate	27,146	2	13,573.0	21.86**	3.74	6.51
Carbonaceous material	4,426	4	1,106.5	1.78	3.11	-
Interaction	3,036	8	379.5	.61	2.70	-
Replicates	368	1	368.0	.59	4.60	-
Error	8,693	14	620.9	-	-	-
Total	43,693	29	-	-	-	-

TABLE 2a. - Results of Experiment 2, a series of r x 2 factorials. Each result represents a W_{50} value based on a nominal 20-shot series

Explosive	1 lb. dry fireclay	½ lb. water in plastic bag	Special asbestos stemming device	1 lb. wet fireclay	¼ lb. dry salt	½ lb. water-salt solution in plastic bag	½ lb. gelled water in plastic bag
1	582	632	-	-	-	-	-
2	507	591	168	461	258	-	-
3	473	487	198	522	599	-	-
4	687	660	265	599	599	-	-
5	574	727	331	536	430	-	-
6	522	607	288	-	-	717	727
7	424	536	218	-	-	591	624
8	544	436	210	-	-	514	551
9	551	789	544	-	-	906	747

TABLE 2b. - Analyses of variance for the data of Table 2a

Source of variance	Sum of squares	Degrees of freedom	Mean squares	F	$F_{.05}$	$F_{.01}$
A. Comparing ½ lb. water in plastic bag with 1 lb. dry fireclay:						
Stemming methods	20,067	1	20,067	3.92	5.32	-
Explosives	103,001	8	12,875			
Error	40,917	8	5,115			
B. Comparing 1 special asbestos stemming device with 1 lb. dry fireclay:						
Stemming methods	265,225	1	265,225	35.02**	5.59	12.25
Explosives	89,624	7	12,803			
Error	53,013	7	7,573			
C. Comparing 1 lb. wet fireclay with 1 lb. dry fireclay:						
Stemming methods	1,891	1	1,891	1.14	10.13	-
Explosives	31,362	3	10,454			
Error	4,962	3	1,654			
D. Comparing ¼ lb. dry salt with 1 lb. dry fireclay:						
Stemming methods	15,753	1	15,753	1.26	10.13	-
Explosives	69,094	3	23,031			
Error	37,426	3	12,475			
E. Comparing ½ lb. water-salt solution with 1 lb. dry fireclay:						
Stemming methods	58,996	1	58,996	4.73	10.13	-
Explosives	60,860	3	20,287			
Error	37,423	3	12,474			
F. Comparing ½ lb. gelled water with 1 lb. dry fireclay:						
Stemming methods	46,208	1	46,208	9.88	10.13	-
Explosives	21,555	3	7,185			
Error	14,037	3	4,679			

TABLE 3a. - Results of Experiment 3

Explosive	Gas, percent 7.0	7.5	8.0	8.5	9.0	9.5	10.0
1	633	500	500	651	659	789	866
2	660	717	599	616	589	599	607
3	894	858	651	632	688	800	1,026
Totals	2,187	2,075	1,750	1,899	1,936	2,188	2,499
Means	729.0	691.7	583.3	633.0	645.3	729.3	833.0

TABLE 3b. - Analysis of variance of results in Table 3a

Source of variation	Sum of squares	Degrees of freedom	Mean squares	F	$F_{.05}$	$F_{.01}$
Gas, linear	21,600	1	21,600	2.01	4.75	-
Gas, quadratic	90,517	1	90,517	8.40*	4.75	9.33
Gas, higher	8,133	4	2,033	.90	3.26	-
Explosives	109,484	2	54,742	5.08*	3.88	6.93
Error	129,281	12	10,773	-	-	-
Total	359,098	20	-	-	-	-

the gas concentration, methane plus ethane, had a significant effect on the ignitibility of the gas mixture and the concentration of maximum ignitibility was approximately 8 percent. As orthogonality was maintained in this experiment, the calculation methods described by Anderson and Bancroft (3) and Cochran and Cox (8) were applicable. Single degree of freedom analysis showed the quadratic regression for the gas to be significant. Therefore, the relationship between the gas concentration and ignitibility, assumed as the inverse of the W_{50} values, is expressed as the parabolic curve of Figure 7.

Experiment 4. One of the early experiments in this series was a one-factor experiment to determine the effect of the quantity of sodium chloride in the explosive on its incendivity to the 8 percent gas mixture. A feature of this experiment was background randomization of all other factors which conceivably could affect the result. Accordingly, Experiment 4 was a series of smaller randomized experiments in which the sodium chloride was varied deliberately and the other factors varied randomly. This was accomplished by determining the W_{50} values on a relatively large number, in this case, 87, explosives of varying compositions. Although this experiment represents considerable work, much of the data was a by-product of the regular testing schedules. Table 4a shows the results placed in the form of a one factor experiment with unequal numbers of subsamples. From the 87 W_{50} values suitable calculations gave ratios, called improvement indexes, for the table of data for analysis. Table 4b presents the results of the analysis of variance. The large F value permitted the strong conclusion that added sodium chloride, up to 20 percent, has a highly significant effect in reducing the incendivity of formulations of permissible explosives (4).

Experiment 5. In a manner similar to that of Experiment 4, a study was made of the effect of the particle size of the sodium chloride constituent (Tables 5a and 5b). The conclusion was that the fine sodium chloride reduced the incendivity of the explosives significantly more than the coarse salt at the 90 percent probability level (or the 10 percent confidence level) (4).

CONCLUSIONS. With respect to the experimental design it was concluded that: (1) The up-and-down method may be combined with a factorial design to provide a useful combination design that gives valid conclusions for relatively complex experiments with explosives. (2) Satisfaction of the requirements of the up-and-down method permits the use of sequences of a minimum of 20, rather than 50 trials, or shots, with this design and leads to appreciable reduction in the cost of the experiments.

REFERENCES

1. Office of Scientific Research and Development. Statistical Analysis for a New Procedure in Sensitivity Experiments. Statistical Research Group, Princeton Univ., Applied Mathematics Panel, Rept. 4040 (AMP Rept. 101.1R), July 1944, 58 pp.

2. Dixon, W. J., and Massey, F. J., Jr. Introduction to Statistical Analysis. McGraw-Hill Book Co., Inc., New York, N. Y., 2nd Edition, 1957, 488 pp.

3. Anderson, R. L., and Bancroft, T. A. Statistical Theory in Research, McGraw-Hill Book Co., Inc., New York, N.Y., 1952, 399 pp.

4. Hanna, N. E., Damon, G. H., and Van Dolah, R. W. Reducing the Incendivity of Permissible Explosives by Sodium Chloride. Bureau of Mines Rept. of Investigations 5683, 1960, 19 pp.

5. Grant, R. L., Mason, C. M., Damon, G. H., and Van Dolah, R. W. Influence of Freespace on Ignition of Firedamp by Explosives. Bureau of Mines Rept. of Investigations 5486, 1959, 21 pp.

6. Hanna, N. E., Grant, R. L., and Van Dolah, R. W. Factors Influencing the Incendivity of Permissible Explosives: Ammonium Nitrate and Carbonaceous Material. Bureau of Mines Rept. of Investigations 5867, 1961, 11 pp.

7. Van Dolah, R. W., Hanna, N. E., and Grant, R. L. Relative Efficacy of Stemming Materials in Reducing Incendivity of Permissible Explosives. Bureau of Mines unpublished report, June 20, 1961, presented at the Restricted International Conference of Directors of Safety in Mines Research, Warsaw, Poland, October 1961.

8. Cochran, W. G., and Cox, G. M. Experimental Designs. John Wiley & Sons, Inc., New York, N. Y., 1957, 611 pp.

TABLE 4a. - Effect of quantity of sodium chloride: 55 improvement indexes derived from 87 W_{50} values

Explosive No.	Improvement index 1/			
	$\frac{(W_{50})_3}{(W_{50})_0}$	$\frac{(W_{50})_{10}}{(W_{50})_0}$ and $\frac{(W_{50})_{10}}{(W_{50})_3}$ 2/	$\frac{(W_{50})_{15}}{(W_{50})_0}$	$\frac{(W_{50})_{20}}{(W_{50})_0}$
1..............	1.21	-	-	-
2..............	1.08	-	-	-
3..............	1.07	-	-	-
4..............	1.17	-	-	-
5..............	.96	-	-	-
6..............	1.20	-	-	-
7..............	1.30	-	-	-
8..............	1.10	-	-	-
9..............	-	1.28	-	-
10..............	-	1.09	-	-
11..............	-	1.07	-	-
12..............	-	1.07	-	-
13..............	-	1.44	-	-
14..............	-	1.51	-	-
15..............	-	1.05	-	-
16..............	-	1.19	-	-
17..............	-	1.09	-	-
18..............	-	2.29	2.91	2.91
19..............	-	1.50	-	-
20..............	-	1.39	-	-
21..............	-	1.37	-	-
22..............	-	1.30	-	-
23..............	-	1.31	-	-
24..............	-	1.25	-	-
25..............	-	1.60	-	-
26..............	-	1.45	-	-
27..............	-	1.58	-	-
28..............	-	1.58	-	-
29..............	-	1.79	2.21	2.59
30..............	-	2.05	1.91	3.16
31..............	-	1.23	1.40	1.91
32..............	-	1.53	1.70	6.51
33..............	-	1.00	1.24	1.28
34..............	-	1.23	-	2.74
35..............	-	1.77	-	5.23
36..............	-	2.29	4.68	7.66
37..............	-	2.32	3.03	4.05

1/ Subscript denotes percentage of sodium chloride in formulations.

2/ $\frac{(W_{50})_{10}}{(W_{50})_0}$ assumed same as $\frac{(W_{50})_{10}}{(W_{50})_3}$

TABLE 4b. - Analysis of variance

Source of variance	Sum of squares	Degrees of freedom	Mean squares	F	$F_{0.05}$	$F_{0.01}$
Improvement index (columns)..............	47.84	3	15.95	15.92**	2.79	4.19
Error.......................	51.10	51	1.002	-	-	-
Total....................	98.94	54	-	-	-	-

TABLE 5a. - Improvement indexes showing the effect of particle size of sodium chloride

Formulation No.	Stemming	Improvement index: $\frac{W_{50}\ (\text{coarse})}{W_{50}\ (\text{No NaCl})}$	Improvement index: $\frac{W_{50}\ (\text{fine})}{W_{50}\ (\text{No NaCl})}$
1...........	Fireclay	-	-
2...........	do.	1.45	-
3...........	do.	-	1.60
4...........	do.	1.51	-
5...........	do.	-	1.51
6...........	2 plugs	-	-
7...........	do.	1.79	-
8...........	do.	-	2.05
9...........	do.	2.21	-
10...........	do.	-	1.91
11...........	do.	2.59	-
12...........	do.	-	3.16
13...........	None	-	-
14...........	do.	1.23	-
15...........	do.	-	1.53
16...........	do.	1.23	-
17...........	do.	-	1.77
18...........	do.	1.40	-
19...........	do.	-	1.70
20...........	do.	1.91	-
21...........	do.	-	6.51
22...........	do.	2.74	-
23...........	do.	-	5.23
24...........	do.	-	-
25...........	do.	2.32	-
26...........	do.	-	2.29
27...........	do.	3.03	-
28...........	do.	-	4.68
29...........	do.	4.05	-
30...........	do.	-	7.66

TABLE 5b. - Analysis of variance

Source of variance	Sum of squares	Degrees of freedom	Mean squares	F	$F_{.10}$	$F_{.05}$
Particle size (columns)	7.69	1	7.69	2.99	2.93	4.26
Error	61.76	24	2.573	-	-	-
Total	69.45	25	-	-	-	-

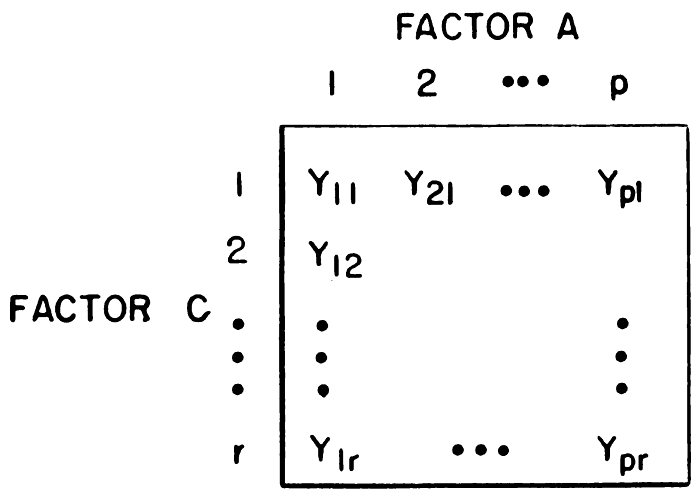


FIGURE 1a

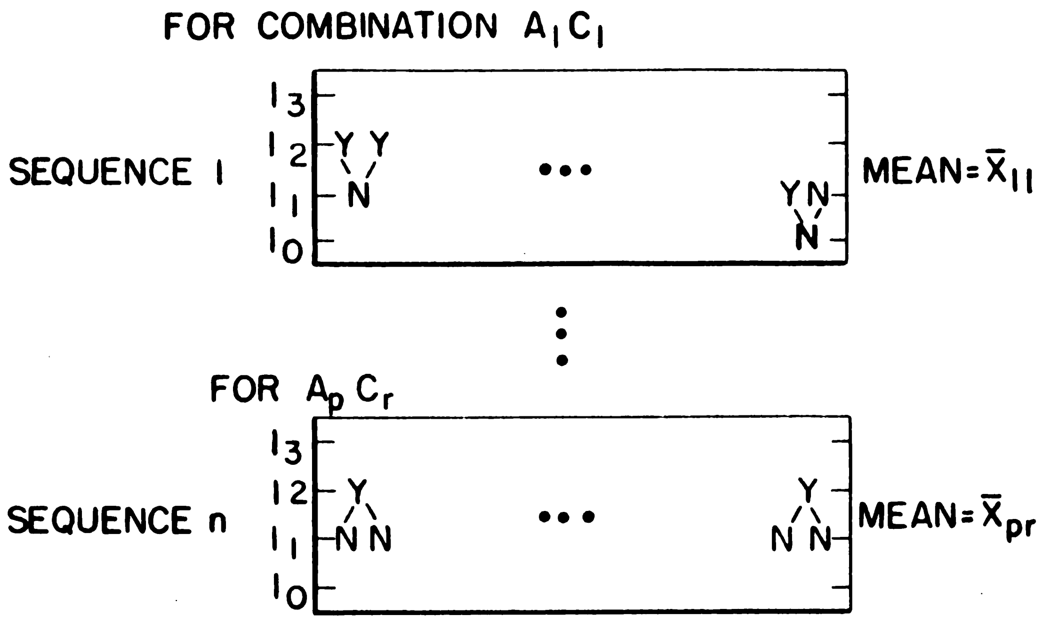


FIGURE 1b

FIGURE 1a.-DESIGN FOR A FACTORIAL EXPERIMENT WITH TWO FACTORS A AND C.

FIGURE 1b.-DESIGN FOR n RANDOMIZED UP-AND-DOWN SEQUENCES, WHERE n EQUALS pr.

8-29-61 L-7017

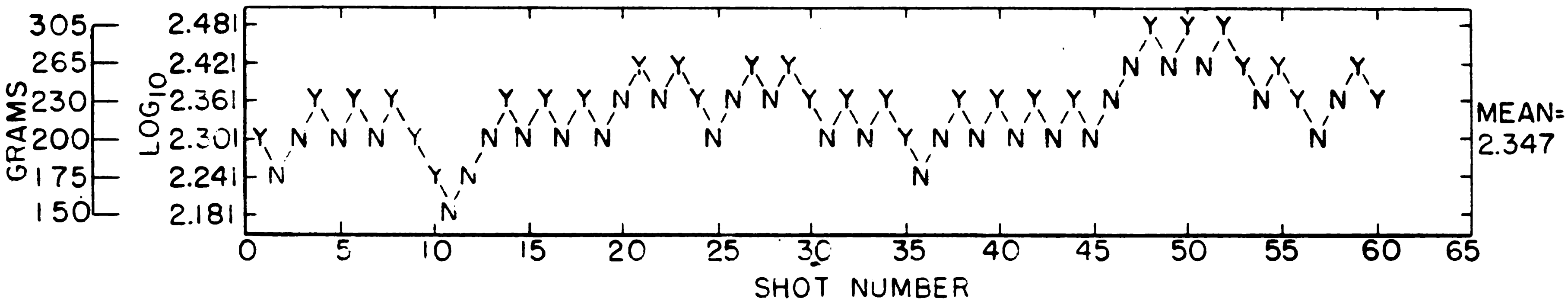


FIGURE 2.- SEQUENCE FOR TEST FOR NORMALITY WITH EXPLOSIVE A.

L-6966 8-16-61

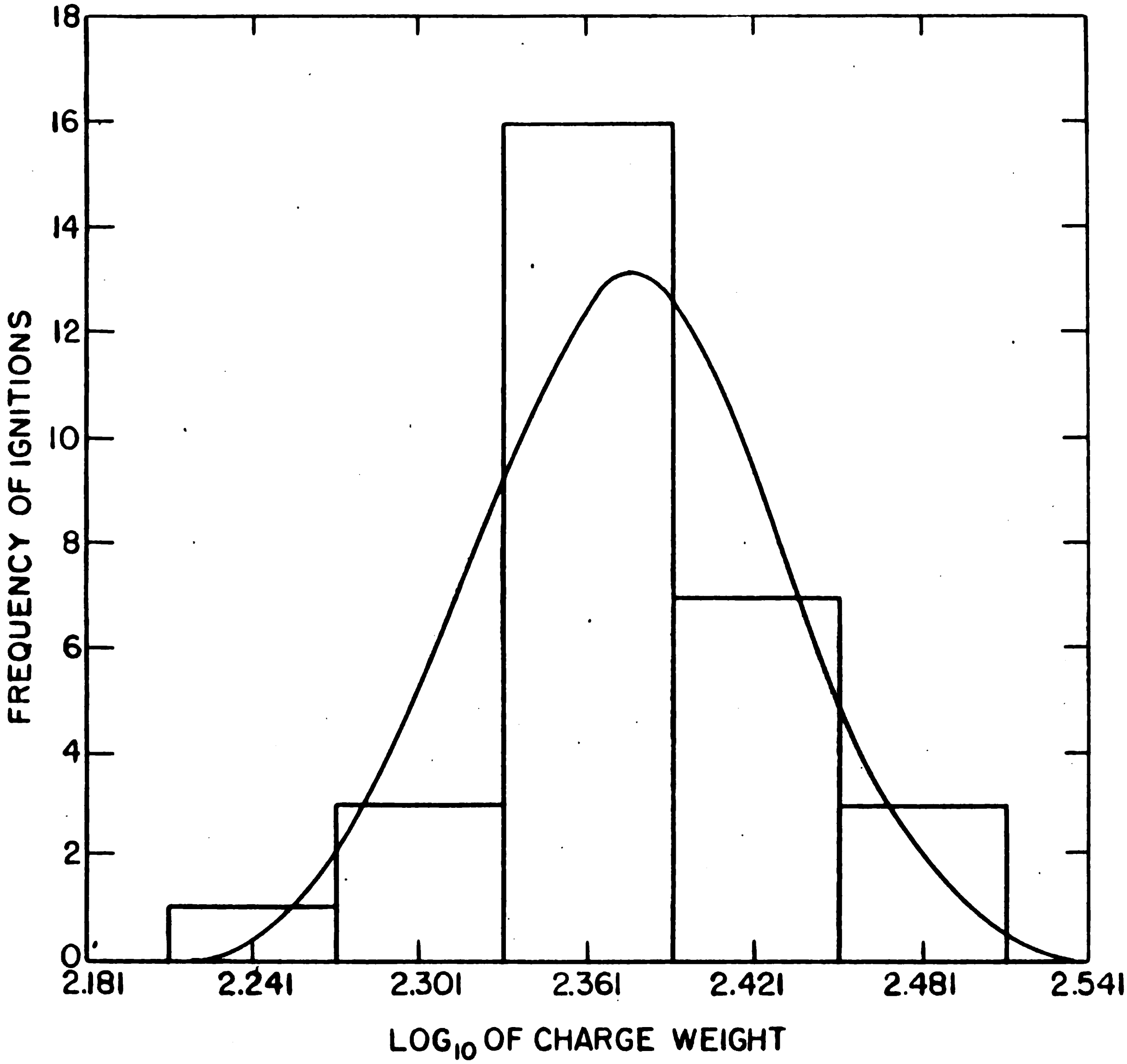


FIGURE 3.-HISTOGRAM AND FITTED NORMAL CURVE FOR FREQUENCY OF IGNITIONS IN UP-AND-DOWN SEQUENCE IN FIGURE 2.

L-6967 8-16-61

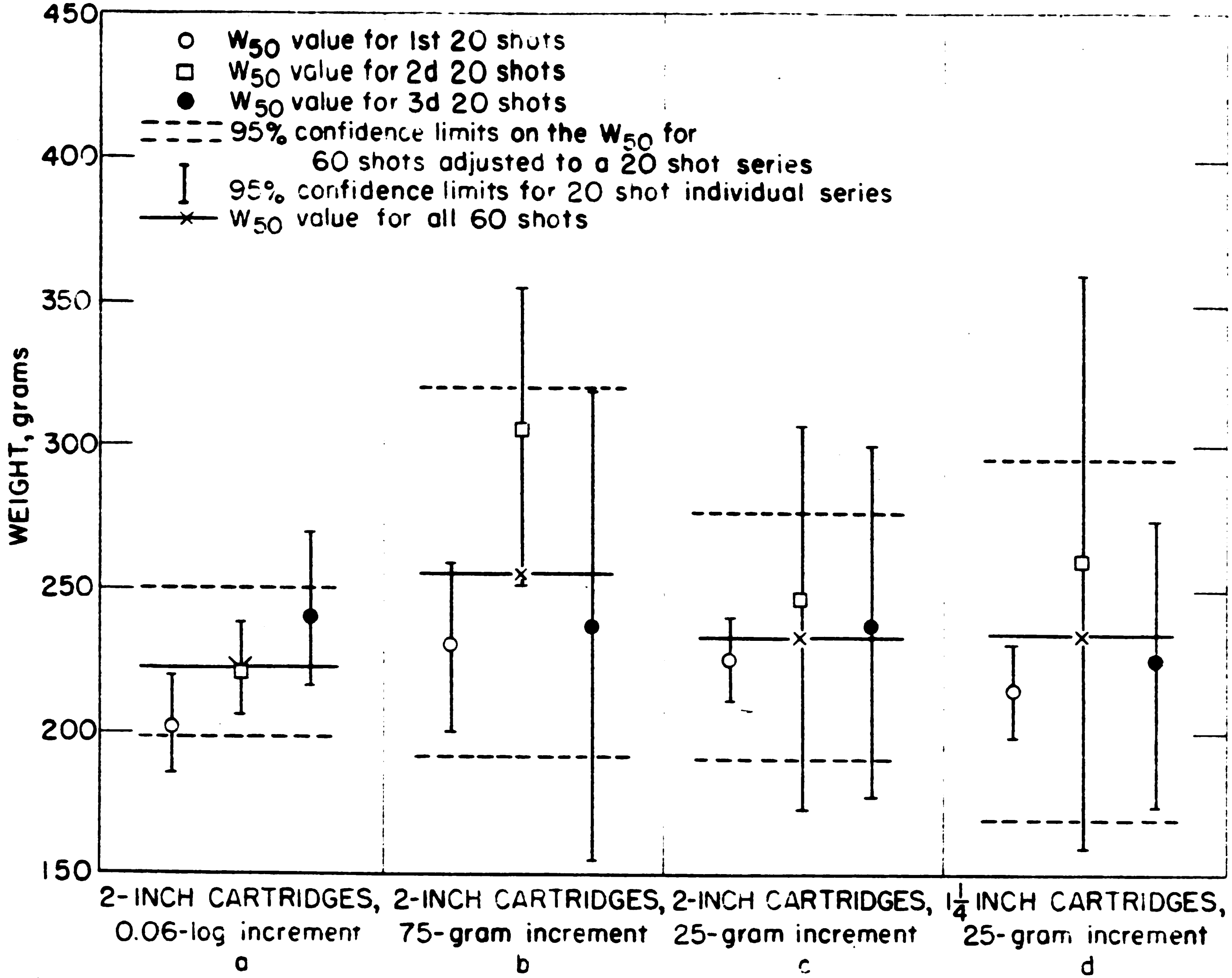


Figure 4.- Reproducibility of means of short sequences of the up-and-down method. The W_{50} values obtained with groups of 20 shots are compared with the 95-percent confidence limits for all 60 shots adjusted to 20 shots. All trials with explosive A.

L-7012 8-25-61

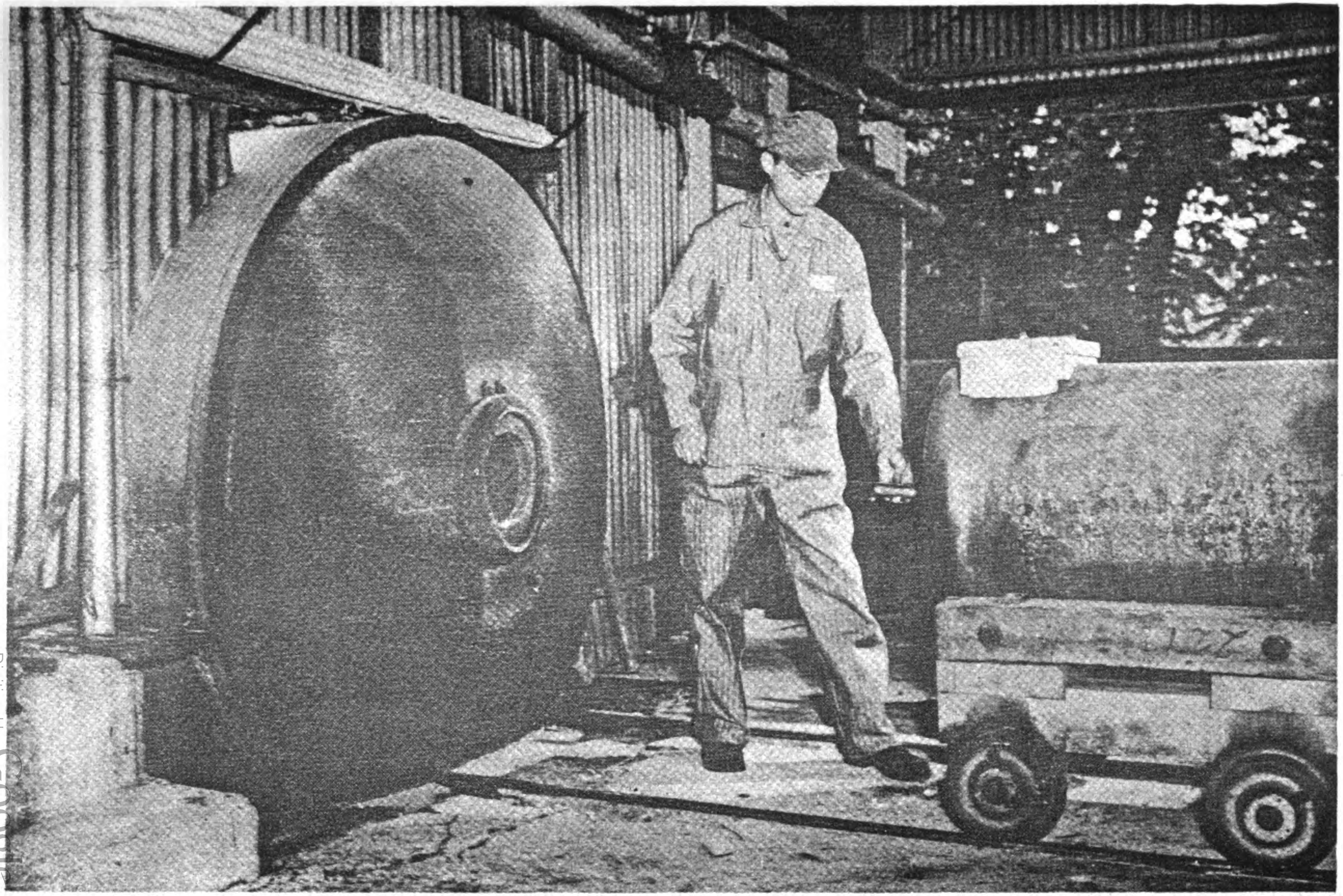


Figure 5.--Operator loading steel canon with explosive charge.

Figure 6.--A typical ignition of the natural gas-air mixture in the gallery by a shot with explosive.

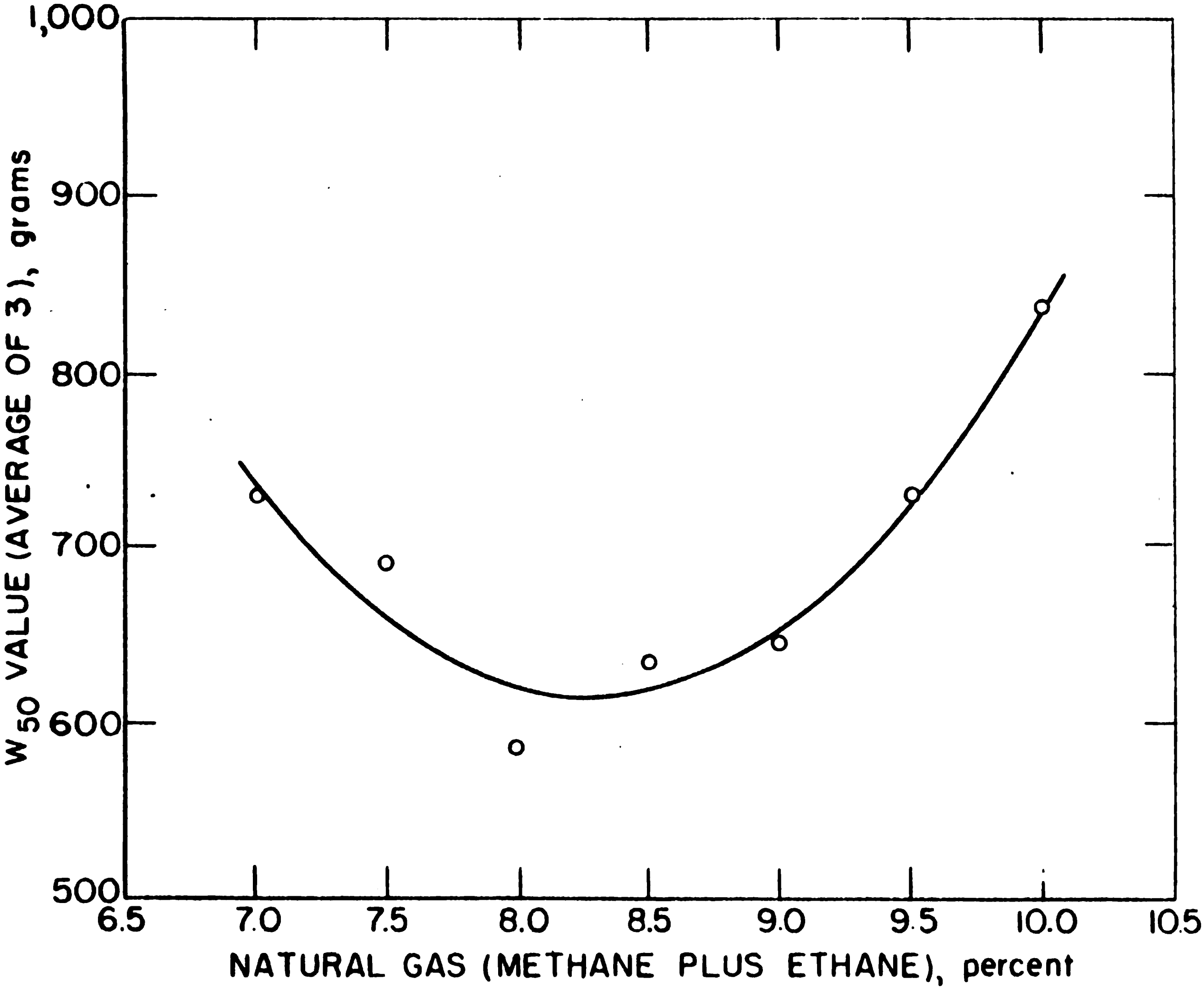


FIGURE 7.-EFFECT OF NATURAL GAS CONCENTRATION ON IGNITIBILITY OF GAS-AIR MIXTURES BY PERMISSIBLE EXPLOSIVES. CURVE PLOTTED THROUGH MEAN VALUES OF THE RESULTS FOR THE THREE EXPLOSIVES. MAXIMUM IGNITIBILIY IS APPROXIMATELY 8.3 PERCENT GAS.

8-3-61 L-7018

GENERAL FORMULAS AND A POSITIONAL INDEX-ALGORITHM FOR GENERATING ORTHOGONAL CONTRASTS IN MULTI-VARIABLE STATISTICAL DESIGNS

Erwin Biser
Systems Division, Surveillance Department
U. S. Army Signal Research and Development Laboratory,
Fort Monmouth, New Jersey

ABSTRACT. This report deals with the development and application of general formulas and a number-positional algorithm to generate the effect functions (effects and interactions) of the factors in orthogonal multi-variable statistical designs. The positional indices serve to establish a biunique correspondence between the elements of a data-matrix and their associated coefficient-multipliers.

The formulas, symmetric functions of the contrast indices, level indices, and factor indices, facilitate the unique identification and computation of the effects and interactions of a desired set or subset of factors in orthogonal designs.

The algorithm and general formulas presented in this report are ideally suited for a computer. The factors, their levels, the associated λ-matrices of polynomial values, and the elements of the data-matrix are uniquely represented by sets of positionally ordered numerical indices as subscripts and superscripts. This situation is amply conducive to machine computations that involve sums of products.

The application of the general formula to a $5 \times 4 \times 3 \times 2$ orthogonal design is elucidated by charts and tables. The report contains a comprehensive summary of formulas for generating the elements of an orthogonal contrast matrix, as well as of the symbolic notation for the positional representation of the index algorithm.

The algorithm developed leads to a significant simplification of the usual techniques of analysis of variance.

GLOSSARY OF TERMS AND SYMBOLS

Algorithm. A symbolic technique and/or method used in mathematical disciplines.

Analysis of Variance. A statistical technique for estimating how much of the total variation in a set of data can be attributed to one or more assignable causes of variation.

Contrast. A comparison or difference between two means or groups of means of a set of data. A contrast can be represented as a Linear combination of these means with known coefficients. When the sum of these coefficients is zero the contrast is said to be orthogonal.

Design of Experiment. An experiment which chooses the important factors, the selection of levels, and the order in which the treatments are taken.

Effect. The effect of a factor is the change in response produced by a change in the level of this factor. The differences between the means of the higher and lower levels of one factor, averaged over all levels of the other factors, constitute the effect, or more specifically, its main effect.

Experiment. A planned set of operations (trials) which lead to a corresponding set of observations, these being the results of the individual trials constituting the experiment.

Factor. Denotes any feature of the experimental conditions which may be deliberately varied from trial to trial. It may represent, for example, temperature, pressure, velocity of a chemical reaction, or azimuth, elevation, slant range for obtaining target positions. Factors may be qualitative (when the levels cannot be arranged in any order of magnitude) or quantitative (when the levels can be arranged in some order of magnitude).

Factorial Experiment. One which studies the effects of a number of different factors on some observable quantity by varying two or more of these factors simultaneously.

Interaction. If the effect of one factor is dependent upon the level chosen for another factor, the two factors are said to interact, or, that an interaction is present.

Level of a Factor. The various fixed values of a factor examined in a factorial experiment are known as levels. The term applies to qualitative as well as quantitative factors. For example in a three factor experiment there may be methods, batches and temperatures involved. There may be two methods (two levels) M_1 and M_2, four batches (four levels) B_1, B_2, B_3, B_4, and three temperatures (three levels) T_1, T_2, T_3. We designate this experiment as a $2 \times 3 \times 4$ factorial experiment.

Replication. Repetition of the whole or part of an experiment a number of times in order to establish the effect of a given treatment more accurately and to provide an estimate of the variation between experimental units receiving the same treatment.

Treatment. The set of levels of all factors used in a given trial is called the treatment or treatment combination. The term treatment is also used to denote the different procedures whose effects are to be measured.

$$\left.\begin{array}{l} i_1, i_2, \ldots i_m \\ j_1, j_2, \ldots j_m \\ p, q, \ldots v \end{array}\right\} \quad \text{indices indicating summation}$$

$i_m j_m$ Element in the ith row and jth column of the orthogonal polynomial coefficient matrix for the mth factor F_m.

$i_m = 1, 2, 3, \ldots, N_m - 1$ yields the linear, quadratic, cubic, ..., $(N_m - 1)$ contrasts respectively.

$j_m = 1, 2, 3, \ldots, N_m$ refers to the first, second, third, ..., N_{mth} levels for the factors F_1, F_2, F_3, ..., F_m respectively.

$\lambda^{(m)}_{i_m j_m}$ represent row vectors of the transpose matrix of the Fisher orthogonal polynomial coefficients of factor F_m with the restriction that $i_m \neq 0$.

$\Lambda^{(m)}_{i_m j_m}$ represent the normalized form of $\lambda^{(m)}_{i_m j_m}$.

$C_{i_1 i_2 \cdots i_m}$ represents the general element of the orthogonal contrast (interaction) matrix of the factors F_1, F_2, ..., F_m.

G. A. Grand average.

r Number of replications per cell (treatment).

$\prod_{k=1}^{m} N_k$ $N_1 \cdot N_2 \cdot \ldots \cdot N_m$.

$S_{j_1 j_2 \ldots j_m}$ General element of data-matrix for factors $F_1, F_2, \ldots, F_m$.

TABLE OF CONTENTS

GENERAL FORMULAS AND A POSITIONAL INDEX-ALGORITHM FOR GENERATING ORTHOGONAL CONTRASTS IN MULTI-VARIABLE STATISTICAL DESIGNS

1. INTRODUCTION . Orthogonal experimental designs are being used extensively in statistical work for many reasons. The principal advantages that accrue from using orthogonal designs are the following:

a. The main effects and interactions can be estimated independently of each other, i.e. the estimate of any effect is unaltered by changes in one or more of the other effects.

b. The work of computing the effects and interactions and of interpreting the results is very much simplified.

c. Fully orthogonal designs are more efficient in that they make possible, for a given number of trials, a more precise estimation of the effects.

The purpose of this paper is to present and elucidate the derivation of a compact general formula for obtaining the orthogonal contrasts or comparisons - effects and interactions - of one or more factors in a multi-factor statistical design. The levels of these factors are equally spaced; this, to be sure, constitutes a constraint on the design, but it has the advantage of facilitating the regression analysis of the treatment sum of squares. The analytical procedure makes use of matrices of orthogonal polynomial values tabulated in tables of orthogonal polynomials (given in Table XXIII, Statistical Tables for Biological Agricultural and Medical Research by Fisher and Yates; Oliver and Boyd, London, 1949, Hafner Publishing Company, New York.)

These polynomial values, namely the values of the orthogonal polynomials at the equally spaced levels, serve as coefficients of the elements of a data-matrix, i.e. the set of data arranged according to a factorial structure (treatment combinations). The tables of values of these orthogonal coefficients are of invaluable aid in simplifying the computation of the linear, quadratic, cubic, etc., components of the factor interactions.

Thus results of the theory of orthogonal polynomials are fruitfully brought to bear on the problem of computing orthogonal contrasts (effects and interactions) for designs characterized by equal spacings of the factor levels.

2. SUMMARY.

a. A general formula and an efficient Positional Index Algorithm are developed for generating all the elements of an orthogonal contrast (interaction) matrix. This formula and algorithm yield all the effect functions (effects and interactions) of the factors in orthogonal multifactor statistical experiments (the levels of each factor are equally spaced). The spacings of the levels need not be the same for all factors.

b. The algorithm is particularly suited for a computer since the indices, both the subscripts and the superscripts have positional significance and the formulas consist of sums of products.

c. The positional indices serve to establish a biunique correspondence between the elements of a data-matrix and their associated coefficient-multipliers.

d. The index algorithm and general formula facilitate the unique identification and computation of the effects and interactions (contrasts) of a desired set or subset of factors, as well as the total sum of observations of orthogonal multifactor experiments.

e. The formula and algorithm are general in that they are not restricted to a specific number of factors.

f. The formulas, symmetric functions of the contrast indices, level indices, and factor indices, facilitate an expeditious selection of the factors whose interaction is desired.

3. DISCUSSION

a. A Heuristic Approach. It is deemed advisable to introduce the application of the theory of orthogonal contrasts by way of a simple example.

Consider a 2 x 2 factorial experiment given below:

$i \downarrow$ \ $j \rightarrow$	1	2
1	S_{11}	S_{12}
2	S_{21}	S_{22}

DATA-MATRIX. Table 1.

The matrix in Table 1 represents a data-matrix of a two factor experiment symbolized by S_{ij}; where the subscripts i and j refer to rows and columns respectively. Furthermore, i refers to the levels of F_1, the first factor and j to those of F_2, the second factor. Thus: $i = 1, 2$; $j = 1, 2$. The juxtapostion i and j in S_{ij} has positional significance, in that the first subscript (from left to right) refers to factor 1 (F_1) and the second refers to factor 2 (F_2). The concept of positional notation is introduced here for the purpose of stressing its significance and use in the subsequent development and derivation of general formulas and algorithms for orthogonal contrasts. S_{21} stands for the treatment in which F_1 (factor 1) is at the second level and F_2 (factor 2) is at the first level.

Let the data or observations in the data-matrix be arranged in the following way, i.e. as a vector, or a row (or column) matrix:

$$S: \left[S_{11}, S_{12}, S_{21}, S_{22} \right]$$

Now we know that the main effects and interactions can be expressed by the following scheme:

	TREATMENT COMBINATION				
	(1)	(a)	(b)	(ab)	DIVISOR
[A]	-1	+1	-1	+1	2
[B]	-1	-1	+1	+1	2
[AB]	+1	-1	-1	+1	2

TABLE 2.

Here A corresponds to F_1

B corresponds to F_2.

If the treatment symbols are interpreted as:

$(1) \equiv a_1 b_1 \equiv S_{11}$ (both factors at their lowest levels)

$(a) \equiv a_2 b_1 \equiv S_{21}$

$(b) \equiv a_1 b_2 \equiv S_{12}$

$(ab) \equiv a_2 b_2 \equiv S_{22}$ (both factors at their highest levels),

then:

$$[A] \quad \text{The main effect of } A = (1/2)\left[S_{22} - S_{12} + S_{21} - S_{11}\right]$$

$$\text{(3a-1)} \quad [B] \quad \text{The main effect of } B = (1/2)\left[S_{22} + S_{12} - S_{21} - S_{11}\right]$$

$$[AB] \quad \text{The AB interaction} = (1/2)\left[S_{22} - S_{12} - S_{21} + S_{11}\right].$$

$[A]$ can be represented as the difference or comparison of <u>two</u> <u>means</u>:

$$\text{(3a-2)} \quad [A] = \left[\frac{S_{22} + S_{21}}{2}\right] - \left[\frac{S_{12} + S_{11}}{2}\right];$$

likewise for $[B]$: $[B] = \left[\dfrac{S_{22} + S_{12}}{2}\right] - \left[\dfrac{S_{21} + S_{11}}{2}\right]$;

whereas: $[AB] = \left[\dfrac{S_{22} + S_{11}}{2}\right] - \left[\dfrac{S_{21} + S_{12}}{2}\right]$.

Let us rearrange the elements in Table 2 to correspond to the elements of S', the transpose of S. We shall simply interchange only the 2nd and 3rd columns of the matrix in Table 2. We obtain the lambda matrix:

$$\lambda \equiv \left[\lambda_{ij}\right] \equiv \begin{bmatrix} -1 & -1 & +1 & +1 \\ -1 & +1 & -1 & +1 \\ +1 & -1 & -1 & +1 \end{bmatrix} \tag{3a-3}$$

λ_{ij} = the element in the i-th row and j-th column; i = 1, 2, 3; j = 1, 2, 3, 4.

If we pre-multiply the column-martix

$$S' = \begin{bmatrix} S_{11} \\ S_{12} \\ S_{21} \\ S_{22} \end{bmatrix}$$ by the matrix λ,

we obtain:

$$\lambda S' = C, \tag{3a-4}$$

where C is a 3xl column-matrix given by:

$$[C] = \begin{bmatrix} C_1 = (-S_{11} - S_{12} + S_{21} + S_{22}) \\ C_2 = (-S_{11} + S_{12} - S_{21} + S_{22}) \\ C_3 = (S_{11} - S_{12} - S_{21} + S_{22}) \end{bmatrix} \equiv \begin{bmatrix} C_1 \\ C_2 \\ C_3 \end{bmatrix}. \tag{3a-5}$$

Each element of the C-matrix is termed a contrast or a comparison, i.e., a difference of means (if we take into account the appropriate divisor). The first element, C_1, is the main A effect; the second element, C_2, the main B effect; and the third, C_3, is the AB effect.

As has been pointed out, these contrasts can be put on a mean basis by introducing proper divisors.

b. Orthogonal Contrasts:

From an examination of the contrast matrix in (3a-5) it can be seen that each element of this matrix, the C-matrix, is a linear form of the means of treatment combinations. C_1, for instance, is given by

(3b-1)

$$-S_{11} - S_{12} + S_{21} + S_{22} \equiv (-1)\, S_{11} + (-1)\, S_{12} + (+1)\, S_{21} + (+1)\, S_{22}\, . \; . \; .$$

The coefficients of S_{ij} (3b-1) are elements λ_{ij} (j = 1, to 4) of the λ-matrix in (3a-3). These coefficients have the following property

(3b-1)

$$\begin{aligned} \lambda_{11} + \lambda_{12} + \lambda_{13} + \lambda_{14} &= 0 \\ \lambda_{21} + \lambda_{22} + \lambda_{23} + \lambda_{24} &= 0 \\ \lambda_{31} + \lambda_{32} + \lambda_{33} + \lambda_{34} &= 0, \end{aligned}$$

This can be written as:

(3b-2)

$$\sum_{j=1}^{4} \lambda_{ij} = 0; \quad (i = 1, 2, 3).$$

Thus from (3a-3) it is clear that:

(3b-3)

$$\left.\begin{aligned} -1 \; -1 \; +1 \; +1 &= 0 \\ -1 \; +1 \; -1 \; +1 &= 0 \\ +1 \; -1 \; -1 \; +1 &= 0 \end{aligned}\right\}.$$

The coefficients of the orthogonal contrasts possess another property, namely, that expressed by the following equation:

(3b-4) $$\sum_j \lambda_{ij} \lambda_{kj} = 0 , \quad i \neq k ,$$

(j refers to columns),

where <u>i and k refer to different contrasts</u> of the set; i.e., the inner product of any pair of rows of the coefficients in the λ-matrix of an orthogonal matrix equals to zero.

Thus from (3a-3) it is seen that: (taking the first and third row)

(3b-5) $$(-1)(+1) + (-1)(-1) + (1)(-1) + (1)(1) = 0.$$

It is this property that enables one to estimate the effect of, say, factor A, independently of factor B, and of the effect of the AB interaction.

The notion of contrast can be put on a more formal basis. It is pertinent to present some of the salient structural characteristics of the concept of contrasts.

<u>Contrasts</u> are <u>comparisons</u> or <u>differences</u> between two means or groups of means. A contrast among parameters $S_1, S_2, \ldots, S_n$ is a linear function of the S_j with known constant coefficients subject to the condition that the sum of the coefficients is zero:

(3b-6)
$$C_1 = \lambda_{11} S_1 + \lambda_{12} S_2 + \ldots + \lambda_{1n} S_n .$$
$$C_1 = \sum_{j=1}^{n} \lambda_{1j} S_j .$$

C_1 is a contrast if $\sum_{j=1}^{n} \lambda_{1j} = 0.$

Two contrasts

$$C_1 = \lambda_{11} s_1 + \lambda_{12} s_2 + \cdots + \lambda_{1n} s_n$$

$$C_2 = \lambda_{21} s_1 + \lambda_{22} s_2 + \cdots + \lambda_{2n} s_n$$

are said to be orthogonal if

(3b-7) $$\lambda_{11} \lambda_{21} + \lambda_{12} \lambda_{22} + \cdots + \lambda_{1n} \lambda_{2n} = 0,$$

or more compactly written as:

(3b-8) $$\sum_{j=1}^{n} \lambda_{1j} \lambda_{2j} = 0.$$

The sums of squares (SS) associated with any contrast C_i is given by:

(3b-9) $$SS(C_i) = (C_i)^2 \Big/ \sum_{j=1}^{n} \left(\lambda_{ij} \right)^2 .$$

EXAMPLE: The following simple example is given with the aim in mind of concretizing some of the abstract notions on contrasts presented thus far:

A \ B	b_1	b_2	TOTALS
a_1	S_{11} 13	S_{12} 18	$S_{11} + S_{12}$ 31
a_2	S_{21} 15	S_{22} 22	$S_{21} + S_{22}$ 37
TOTALS	$S_{11} + S_{21}$ 28	$S_{12} + S_{22}$ 40	$S_{..} = \sum_{i=1}^{2} \sum_{j=1}^{2} S_{ij} = 68$; $\bar{S}_{..} = 68/4 = 17$

TWO-BY-TWO TABLE. Table 3

$S_{..}$ is the total sum of treatment measurements: grand total.

$\bar{S}_{..}$ is the grand mean.

The main effect of A = (37-31)/2 = 3.

The main effect of B = (40-28)/2 = 6.

The interaction of A and B = (13 + 22 - 18 - 15)/2 = 1.

The effect of A in the presence of b_1 is equal to 15-13 = 2.

The effect of A in the presence of b_2 is equal to 22-18 = 4.

Similarly, measures can be obtained for the effects of B in the presence of a_1 and in the presence of a_2 respectively. Note that $(3)^2 + (6)^2 + (1)^2 = 46$. This the total sum of squares (TSS) given by the usual formula:

$$\mathrm{TSS}(S_{ij}) = \sum_{i,j} (S_{ij} - \bar{S}..)^2$$

(3b-10)
$$= (13 - 17)^2 + (18 - 17)^2 + (15 - 17)^2 + (22 - 17)^2$$

$$= 46.$$

Let us apply the matrix of orthogonal coefficients given in (3a-3) and make use of equation (3a-4):

(3b-11)
$$\lambda S' = \begin{bmatrix} -1 & -1 & +1 & +1 \\ -1 & +1 & -1 & +1 \\ & & & \\ +1 & -1 & -1 & +1 \end{bmatrix} \begin{bmatrix} 13 \\ 18 \\ 15 \\ 22 \end{bmatrix} = C,$$

(contrast matrix)
$$[C] = \begin{bmatrix} 6 \\ 12 \\ 2 \end{bmatrix} = \begin{bmatrix} C_1 \\ C_2 \\ C_3 \end{bmatrix},$$

C_1 = the A-effect total
C_2 = the B-effect total
C_3 = the AB interaction total.
The sum of squares of the contrast elements is given by (3b-9)

$$SS(C_i) = (C_i)^2 \Big/ \sum_j (\lambda_{ij})^2$$

$$SS(C_1) = 36/\,(1^2 + 1^2 + 1^2 + 1^2) = 9$$

$$SS(C_2) = 144/(1^2 + 1^2 + 1^2 + 1^2) = 36$$

$$SS(C_3) = 4/(1^2 + 1^2 + 1^2 + 1^2) = 1$$

$$TSS = \sum_{i=1}^{3}\left[C_i^2 \Big/ \sum_{j=1}^{4} (\lambda_{ij})^2\right] = 9 + 36 + 1 = 46.$$

c. A Replicated 5 x 4 Design. Let us turn our attention to a two factor experiment with F_1 (factor A) and F_2 (factor B) at four and five levels respectively. The levels of F_1 and F_2 are equally spaced. The spacings of A and B need not be the same. The experiment is conducted with three replications for each treatment combination:

$$\begin{aligned} &N_1 \text{ (Number of levels of } F_1) = 4 \\ (3c\text{-}1) \qquad &N_2 \text{ (Number of levels of } F_2) = 5 \\ &r \text{ (Number of replications)} = 3 \;. \end{aligned}$$

The following tableau gives the data matrix of the experiment:

$[S_{ij}]$: DATA MATRIX OF 5 x 4 EXPERIMENT

A \ B	b_1	b_2	b_3	b_4	b_5	TOTALS
a_1	S_{11}	S_{12}	S_{13}	S_{14}	S_{15}	$\sum S_{1j} = \sum a_1$
a_2	S_{21}	S_{22}	S_{23}	S_{24}	S_{25}	$\sum S_{2j} = \sum a_2$
a_3	S_{31}	S_{32}	S_{33}	S_{34}	S_{35}	$\sum S_{3j} = \sum a_3$
a_4	S_{41}	S_{42}	S_{43}	S_{44}	S_{45}	$\sum S_{4j} = \sum a_4$
TOTALS	$\sum S_{i1} = \sum b_1$	$\sum S_{i2} = \sum b_2$	$\sum S_{i3} = \sum b_3$	$\sum S_{i4} = \sum b_4$	$\sum S_{i5} = \sum b_5$	$\sum\sum S_{ij} = \sum a_i = \sum b_j$

TABLE 4A.

S_{ij} is the data matrix of this experiment; $i = 1$ to 4; $j = 1$ to 5.
Table XXIII of The Fisher and Yates Tables give the following orthogonal coefficient matrix (the matrix of polynomial values):

For factor A ($N_1 = 4$) this matrix is

$$(3c\text{-}2) \qquad [P(A)] = \begin{bmatrix} -3 & +1 & -1 \\ -1 & -1 & +3 \\ +1 & -1 & -3 \\ +3 & +1 & +1 \end{bmatrix}.$$

The first column is associated with the linear, the second column with the quadratic and the third column with the cubic effects respectively.

Let us form the matrix $\left[\lambda_{ij}^{(1)}\right]$, the transpose of P(A) in (3c-2).

$$\text{(3c-3)} \quad \left[\lambda_{ij}^{(1)}\right] = \begin{bmatrix} \lambda_{11}^{(1)} & \lambda_{12}^{(1)} & \lambda_{13}^{(1)} & \lambda_{14}^{(1)} \\ \lambda_{21}^{(1)} & \lambda_{22}^{(1)} & \lambda_{23}^{(1)} & \lambda_{24}^{(1)} \\ \lambda_{31}^{(1)} & \lambda_{32}^{(1)} & \lambda_{33}^{(1)} & \lambda_{34}^{(1)} \end{bmatrix}$$

$i = 1, 2, 3$
$j = 1, 2, 3, 4$

$$= \begin{bmatrix} -3 & -1 & +1 & +3 \\ +1 & -1 & -1 & +1 \\ -1 & +3 & -3 & +1 \end{bmatrix}$$

$$= \begin{bmatrix} L_1^{(1)} & L_2^{(1)} & L_3^{(1)} & L_4^{(1)} \\ Q_1^{(1)} & Q_2^{(1)} & Q_3^{(1)} & Q_4^{(1)} \\ C_1^{(1)} & C_2^{(1)} & C_3^{(1)} & C_4^{(1)} \end{bmatrix}.$$

$\left[\lambda_{1j}^{(1)}\right]$, the first row is the linear row;

$\left[\lambda_{2j}^{(1)}\right]$, the second row is the quadratic row;

$\left[\lambda_{3j}^{(1)}\right]$, the third row is the cubic row.

The superscript (1) in $\left[\lambda_{ij}^{(1)}\right]$ refers to the F_1, factor A.

One can obtain the <u>contrast matrix</u> for <u>the A-effects</u> by <u>post-multiplying</u> the $\left[\lambda_{ij}^{(1)}\right]$ matrix for factor A, given in (3c-3) by the column-matrix

$$\begin{bmatrix} \sum a_1 \\ \sum a_2 \\ \sum a_3 \\ \sum a_4 \end{bmatrix}:$$

$$\text{(3c-4)} \quad \underset{(k = 1,\ 2,\ 3)}{\left[C_k(A)\right]} = \begin{bmatrix} L_1^{(1)} & L_2^{(1)} & L_3^{(1)} & L_4^{(1)} \\ Q_1^{(1)} & Q_2^{(1)} & Q_3^{(1)} & Q_4^{(1)} \\ C_1^{(1)} & C_2^{(1)} & C_3^{(1)} & C_4^{(1)} \end{bmatrix} \begin{bmatrix} \sum a_1 \\ \sum a_2 \\ \sum a_3 \\ \sum a_4 \end{bmatrix},$$

where

$$\text{(3c-5)} \qquad \sum a_i = \sum\sum s_{ij}, \ i = 1, 2, 3, 4,$$

(see Table 4A).

$$
(3c\text{-}6)\quad \begin{cases}
C_1 = L_1^{(1)}\sum a_1 + L_2^{(1)}\sum a_2 + L_3^{(1)}\sum a_3 + L_4^{(1)}\sum a_4 \\
\quad = \text{Linear effect total of A} \\
C_2 = Q_1^{(1)}\sum a_1 + Q_2^{(1)}\sum a_2 + Q_3^{(1)}\sum a_3 + Q_4^{(1)}\sum a_4 \\
\quad = \text{Quadratic effect total of A} \\
C_3 = C_1^{(1)}\sum a_1 + C_2^{(1)}\sum a_2 + C_3^{(1)}\sum a_3 + C_4^{(1)}\sum a_4 \\
\quad = \text{Cubic effect total of A}
\end{cases}
$$

$$
(3c\text{-}7)\quad \underset{(k = 1,\ 2,\ 3)}{\left[C_k\ (A)\right]} = \begin{bmatrix} C_1\ (A) \\ C_2\ (A) \\ C_3\ (A) \end{bmatrix}.
$$

Note that there are three A contrasts; this number is one less than the number of levels of A, which is four ($N_1 = 4$).

The sum of squares (SS) of the element $C_k(A)$, the k-th element of contrast matrix of A is given by the following expression:

$$
(3c\text{-}8)\quad \underset{(k = 1,\ 2,\ 3)}{SS\left[C_k\ (A)\right]} = \frac{(C_k)^2}{rN_2\sum_{j=1}^{4}\left[\lambda_{ij}^{(1)}\right]^2},
$$

where $\left[\lambda_{ij}^{(1)}\right]$ is given in (3c-3); and where

N_2 = Number of levels of F_2 (factor B),

r = Number of replications of each treatment.

In this example $N_2 = 5$; $r = 3$. See (3c-1).

The total sum of squares (TSS) for factor A is given by the following expression:

(3c-9)
$$\text{TSS}(A) = \sum_{k=1}^{3} \left\{ \left[C_k\right]^2 \Big/ r N_2 \sum_{j=1}^{4} \left[\lambda_{ij}^{(1)}\right]^2 \right\}$$

(3c-10)
$$= \sum_{i=1}^{3} \left\{ \left[C_i\right]^2 \Big/ 15 \sum_{j=1}^{4} \left[\lambda_{ij}^{(1)}\right]^2 \right\}$$

Analogously it can be shown that the contrast matrix for factor B ($\equiv F_2$) is given by the following expression:

(3c-11)
$$\left[C_k(B)\right] = \begin{bmatrix} L_1^{(2)} & L_2^{(2)} & L_3^{(2)} & L_4^{(2)} & L_5^{(2)} \\ Q_1^{(2)} & Q_2^{(2)} & Q_3^{(2)} & Q_4^{(2)} & Q_5^{(2)} \\ C_1^{(2)} & C_2^{(2)} & C_3^{(2)} & C_4^{(2)} & C_5^{(2)} \\ q_1^{(2)} & q_2^{(2)} & q_3^{(2)} & q_4^{(2)} & q_5^{(2)} \end{bmatrix} \begin{bmatrix} \sum b_1 \\ \sum b_2 \\ \sum b_3 \\ \sum b_4 \\ \sum b_5 \end{bmatrix}.$$
$(k = 1, 2, 3, 4)$

Where:

$$(3c\text{-}12)\quad \begin{bmatrix} L_1^{(2)} & L_2^{(2)} & L_3^{(2)} & L_4^{(2)} & L_5^{(2)} \\ Q_1^{(2)} & Q_2^{(2)} & Q_3^{(2)} & Q_4^{(2)} & Q_5^{(2)} \\ C_1^{(2)} & C_2^{(2)} & C_3^{(2)} & C_4^{(2)} & C_5^{(2)} \\ q_1^{(2)} & q_2^{(2)} & q_3^{(2)} & q_4^{(2)} & q_5^{(2)} \end{bmatrix} = \begin{bmatrix} \lambda_{11}^{(2)} & \lambda_{12}^{(2)} & \lambda_{13}^{(2)} & \lambda_{14}^{(2)} & \lambda_{15}^{(2)} \\ \lambda_{21}^{(2)} & \lambda_{22}^{(2)} & \lambda_{23}^{(2)} & \lambda_{24}^{(2)} & \lambda_{25}^{(2)} \\ \lambda_{31}^{(2)} & \lambda_{32}^{(2)} & \lambda_{33}^{(2)} & \lambda_{34}^{(2)} & \lambda_{35}^{(2)} \\ \lambda_{41}^{(2)} & \lambda_{42}^{(2)} & \lambda_{43}^{(2)} & \lambda_{44}^{(2)} & \lambda_{45}^{(2)} \end{bmatrix}$$

$$= \begin{bmatrix} \lambda_{ij}^{(2)} \end{bmatrix} = \begin{bmatrix} -2 & -1 & 0 & +1 & +2 \\ +2 & -1 & -2 & -1 & +2 \\ -1 & +2 & 0 & -2 & +1 \\ +1 & -4 & +6 & -4 & +1 \end{bmatrix}.$$

$i = 1, 2, 3, 4$
$j = 1, 2, 3, 4, 5$

Note there are four B contrasts; i.e., one less than the number of levels of $B (N_2 = 5)$.

$$(3c\text{-}13)\quad \begin{bmatrix} C_k(B) \end{bmatrix} = \begin{bmatrix} C_1(B) \\ C_2(B) \\ C_3(B) \\ C_4(B) \end{bmatrix}$$

$(k=1,2,3,4)$

The sum of squares of the k-th contrast of B is given by:

$$SS\left[C_k(A)\right] = \frac{(C_k)^2}{rN_1\sum_{j=1}^{5}\left[\lambda_{ij}^{(2)}\right]^2}. \tag{3c-14}$$

Where N_1 = the number of levels of $A\,(=F_1)$

d. Derivation of Expression for Elements of Contrast Matrix (Two Factor Experiment).

Let us consider the data matrix $\left[S_{ij}\right]$ given in Table 4A ($i = 1$ to 4; $j = 1$ to 5).

The last column in this table, consisting of $\sum a_1$, $\sum a_2$, $\sum a_3$, and of $\sum a_4$ can be obtained by postmultiplying the data matrix $\left[S_{ij}\right]$ by the column-matrix consisting of 1's in the following manner:

$$\begin{bmatrix} S_{11} & S_{12} & S_{13} & S_{14} & S_{15} \\ S_{21} & S_{22} & S_{23} & S_{24} & S_{25} \\ S_{31} & S_{32} & S_{33} & S_{34} & S_{35} \\ S_{41} & S_{42} & S_{43} & S_{44} & S_{45} \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} \sum S_{1j} \\ \sum S_{2j} \\ \sum S_{3j} \\ \sum S_{4j} \end{bmatrix}. \tag{3d-1}$$

In view of equations (3c-3) and (3c-4), the contrast-matrix of A can be expressed as follows:

(3d-2)

$$\left[C_k(A)\right] = \begin{bmatrix} \lambda_{11}^{(1)} & \lambda_{12}^{(1)} & \lambda_{13}^{(1)} & \lambda_{14}^{(1)} \\ \lambda_{21}^{(1)} & \lambda_{22}^{(1)} & \lambda_{23}^{(1)} & \lambda_{24}^{(1)} \\ \lambda_{31}^{(1)} & \lambda_{32}^{(1)} & \lambda_{33}^{(1)} & \lambda_{34}^{(1)} \end{bmatrix} \begin{bmatrix} S_{11} & S_{12} & S_{13} & S_{14} & S_{15} \\ S_{21} & S_{22} & S_{23} & S_{24} & S_{25} \\ S_{31} & S_{32} & S_{33} & S_{34} & S_{35} \\ S_{41} & S_{42} & S_{43} & S_{44} & S_{45} \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}.$$

$(k = 1, 2, 3)$

Thus it is clear that <u>postmultiplying</u> the data matrix by the column-matrix of 1's has the <u>effect of eliminating the B factor</u>, since it merely adds the data row-wise yielding the vector in the last column of Table 4; this is the column vector on the right hand of equation (3d-1). Equation (3d-2) can also be put in the following form:

(3d-3)

$$\left[C_k(A)\right] = \begin{bmatrix} L_1^{(1)} & L_2^{(1)} & L_3^{(1)} & L_4^{(1)} \\ Q_1^{(1)} & Q_2^{(1)} & Q_3^{(1)} & Q_4^{(1)} \\ C_1^{(1)} & C_2^{(1)} & C_3^{(1)} & C_4^{(1)} \end{bmatrix} \begin{bmatrix} \Sigma S_{1j} & 0 & 0 & 0 \\ 0 & \Sigma S_{2j} & 0 & 0 \\ 0 & 0 & \Sigma S_{3j} & 0 \\ 0 & 0 & 0 & \Sigma S_{4j} \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}.$$

$(k = 1, 2, 3)$

In order to obtain the contrast matrix for factor B ($=F_2$), the data matrix, $\left[S_{ij}\right]$ is <u>premultiplied</u> by the row-matrix of 1's; the resulting matrix is premultiplied by the matrix of orthogonal coefficients corresponding to the number of levels of factor $B(=F_2)$.* Thus analogously to (3d-2) the contrast matrix for factor B is given by the following expression ($N_2 = 5$):

(3d-4)

$$[C_k(B)] = \begin{bmatrix} L_1^{(2)} & L_2^{(2)} & L_3^{(2)} & L_4^{(2)} & L_5^{(2)} \\ Q_1^{(2)} & Q_2^{(2)} & Q_3^{(2)} & Q_4^{(2)} & Q_5^{(2)} \\ C_1^{(2)} & C_2^{(2)} & C_3^{(2)} & C_4^{(2)} & C_5^{(2)} \\ q_1^{(2)} & q_2^{(2)} & q_3^{(2)} & q_4^{(2)} & q_5^{(2)} \end{bmatrix} \cdot \left(\begin{bmatrix} 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} S_{11} & S_{12} & S_{13} & S_{14} & S_{15} \\ S_{21} & S_{22} & S_{23} & S_{24} & S_{25} \\ S_{31} & S_{32} & S_{33} & S_{34} & S_{35} \\ S_{41} & S_{42} & S_{43} & S_{44} & S_{45} \end{bmatrix} \right)^T$$

($k = 1, 2, 3, 4$)

where

(3d-5)

$$\underset{k = 1 \text{ to } 4\ (=N_2 - 1)}{[C_k(B)]} = \begin{bmatrix} C_1(B) \\ C_2(B) \\ C_3(B) \\ C_4(B) \end{bmatrix}.$$

The contrast matrix for factor B can be written in a form analogous to (3d-3):

---

* Actually, the transpose of the resulting matrix product is premultiplied by the coefficient matrix; analogously for the product of the two right extreme matrices in equation (3d-6).

(3d-6)

$$[C_k(B)] = \begin{bmatrix} L_1^{(2)} & L_2^{(2)} & L_3^{(2)} & L_4^{(2)} & L_5^{(2)} \\ Q_1^{(2)} & Q_2^{(2)} & Q_3^{(2)} & Q_4^{(2)} & Q_5^{(2)} \\ C_1^{(2)} & C_2^{(2)} & C_3^{(2)} & C_4^{(2)} & C_5^{(2)} \\ q_1^{(2)} & q_2^{(2)} & q_3^{(2)} & q_4^{(2)} & q_5^{(2)} \end{bmatrix} \left\{ \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} \sum S_{i1} & 0 & 0 & 0 & 0 \\ 0 & \sum S_{i2} & 0 & 0 & 0 \\ 0 & 0 & \sum S_{i3} & 0 & 0 \\ 0 & 0 & 0 & \sum S_{i4} & 0 \\ 0 & 0 & 0 & 0 & \sum S_{i5} \end{bmatrix} \right\}^T .$$

$(k = 1, 2, 3, 4)$

Thus far the presentation dealt with the contrast matrices for the A and B factors singly, i.e., no interaction terms were involved. What is needed is to develop an expression for obtaining the general element of the contrast matrix for the <u>AB interaction</u>, or the <u>interaction contrast matrix</u>.

Let the data matrix of the experiment, denoted by $[S_{ij}]$, be <u>premultiplied</u> by the matrix of orthogonal coefficients for factor A, namely, the $[\lambda_{ij}^{(1)}]$ matrix given in (3c-3); and postmultiplied by the <u>transpose</u> of the matrix of orthogonal coefficients for factor B, namely, the $[\lambda_{ij}^{(2)}]$ matrix given in (3c-12):

(3d-7)

$$\begin{bmatrix} L_1^{(1)} & L_2^{(1)} & L_3^{(1)} & L_4^{(1)} \\ Q_1^{(1)} & Q_2^{(1)} & Q_3^{(1)} & Q_4^{(1)} \\ C_1^{(1)} & C_2^{(1)} & C_3^{(1)} & C_4^{(1)} \end{bmatrix} \begin{bmatrix} S_{11} & S_{12} & S_{13} & S_{14} & S_{15} \\ S_{21} & S_{22} & S_{23} & S_{24} & S_{25} \\ S_{31} & S_{32} & S_{33} & S_{34} & S_{35} \\ S_{41} & S_{42} & S_{43} & S_{44} & S_{45} \end{bmatrix} \begin{bmatrix} L_1^{(2)} & Q_1^{(2)} & C_1^{(2)} & q_1^{(2)} \\ L_2^{(2)} & Q_2^{(2)} & C_2^{(2)} & q_2^{(2)} \\ L_3^{(2)} & Q_3^{(2)} & C_3^{(2)} & q_3^{(2)} \\ L_4^{(2)} & Q_4^{(2)} & C_4^{(2)} & q_4^{(2)} \\ L_5^{(2)} & Q_5^{(2)} & C_5^{(2)} & q_5^{(2)} \end{bmatrix} .$$

This can be seen to equal to:

$$\begin{bmatrix} \sum L_i^{(1)} s_{i1} & \sum L_i^{(1)} s_{i2} & \sum L_i^{(1)} s_{i3} & \sum L_i^{(1)} s_{i4} & \sum L_i^{(1)} s_{i5} \\ \sum Q_i^{(1)} s_{i1} & \sum Q_i^{(1)} s_{i2} & \sum Q_i^{(1)} s_{i3} & \sum Q_i^{(1)} s_{i4} & \sum Q_i^{(1)} s_{i5} \\ \sum C_i^{(1)} s_{i1} & \sum C_i^{(1)} s_{i2} & \sum C_i^{(1)} s_{i3} & \sum C_i^{(1)} s_{i4} & \sum C_i^{(1)} s_{i5} \end{bmatrix} \begin{bmatrix} L_1^{(2)} & Q_1^{(2)} & C_1^{(2)} & q_1^{(2)} \\ L_2^{(2)} & Q_2^{(2)} & C_2^{(2)} & q_2^{(2)} \\ L_3^{(2)} & Q_3^{(2)} & C_3^{(2)} & q_3^{(2)} \\ L_4^{(2)} & Q_4^{(2)} & C_4^{(2)} & q_4^{(2)} \\ L_5^{(2)} & Q_5^{(2)} & C_5^{(2)} & q_5^{(2)} \end{bmatrix}$$

where $\sum \equiv \sum_{i=1}^{N_1 = 4}$

(3d-8)

$$= C_k(AB) \qquad (k = 1, 2, \cdots, 12)$$

. This is the AB <u>interaction contrast</u> <u>matrix</u>.

Equation (3d-8) can be seen to be equivalent to the following expression:

(3d-9)

$$\begin{bmatrix}
\sum\limits_{i,j=1}^{N_1,N_2} L_i^{(1)} L_j^{(2)} S_{ij} & \sum\limits_{i,j=1}^{N_1,N_2} L_i^{(1)} Q_j^{(2)} S_{ij} & \sum\limits_{i,j=1}^{N_1,N_2} L_i^{(1)} C_j^{(2)} S_{ij} & \sum\limits_{i,j=1}^{N_1,N_2} L_i^{(1)} q_j^{(2)} S_{ij} \\
\sum\limits_{i,j=1}^{N_1,N_2} Q_i^{(1)} L_j^{(2)} S_{ij} & \sum\limits_{i,j=1}^{N_1,N_2} Q_i^{(1)} Q_j^{(2)} S_{ij} & \sum\limits_{i,j=1}^{N_1,N_2} Q_i^{(1)} C_j^{(2)} S_{ij} & \sum\limits_{i,j=1}^{N_1,N_2} Q_i^{(1)} q_i^{(2)} S_{ij} \\
\sum\limits_{i,j=1}^{N_1,N_2} C_i^{(1)} L_j^{(2)} S_{ij} & \sum\limits_{i,j=1}^{N_1,N_2} C_i^{(1)} Q_j^{(2)} S_{ij} & \sum\limits_{i,j=1}^{N_1,N_2} C_i^{(1)} C_j^{(2)} S_{ij} & \sum\limits_{i,j=1}^{N_1,N_2} C_i^{(1)} q_j^{(2)} S_{ij}
\end{bmatrix}$$

where

$$\sum_{i,j=1}^{N_1,N_2} \equiv \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} .$$

In view of (3c-3), equation (3d-9) can be represented by the following expression:

(3d-10)

$$[C_k(AB)] = \quad (k = 1, 2, \cdots 12)$$

$$\begin{bmatrix} \sum_{i,j=1}^{N_1,N_2} \lambda_{1i}^{(1)}\lambda_{1j}^{(2)}s_{ij} & \sum_{i,j=1}^{N_1,N_2} \lambda_{1i}^{(1)}\lambda_{2j}^{(2)}s_{ij} & \sum_{i,j=1}^{N_1,N_2} \lambda_{1i}^{(1)}\lambda_{3j}^{(2)}s_{ij} & \sum_{i,j=1}^{N_1,N_2} \lambda_{1i}^{(1)}\lambda_{4j}^{(2)}s_{ij} \\ \sum_{i,j=1}^{N_1,N_2} \lambda_{2i}^{(1)}\lambda_{1j}^{(2)}s_{ij} & \sum_{i,j=1}^{N_1,N_2} \lambda_{2i}^{(1)}\lambda_{2j}^{(2)}s_{ij} & \sum_{i,j=1}^{N_1,N_2} \lambda_{2i}^{(1)}\lambda_{3j}^{(2)}s_{ij} & \sum_{i,j=1}^{N_1,N_2} \lambda_{2i}^{(1)}\lambda_{4j}^{(2)}s_{ij} \\ \sum_{i,j=1}^{N_1,N_2} \lambda_{3i}^{(1)}\lambda_{1j}^{(2)}s_{ij} & \sum_{i,j=1}^{N_1,N_2} \lambda_{3i}^{(1)}\lambda_{2j}^{(2)}s_{ij} & \sum_{i,j=1}^{N_1,N_2} \lambda_{3i}^{(1)}\lambda_{3j}^{(2)}s_{ij} & \sum_{i,j=1}^{N_1,N_2} \lambda_{3i}^{(1)}\lambda_{4j}^{(2)}s_{ij} \end{bmatrix}$$

where

$$\sum_{i,j=1}^{N_1,N_2} \equiv \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} .$$

$i = 1$ to N_1 (=4 in this experiment)

$j = 1$ to N_2 (=5 in this experiment),

where $\left[\lambda^{(1)}_{1i}\right]$, $\left[\lambda^{(1)}_{2i}\right]$, $\left[\lambda^{(1)}_{3i}\right]$ are the linear, quadratic, and cubic row-matrices associated with F_1, which is factor A. Note that the superscript "(1)" refers to factor designated by 1, namely F_1 (≡A). Also note that the subscripts "1", "2" and "3" refer to the linear, quadratic, and cubic row-vectors of the matrix of orthogonal coefficients. Pari passu, similar descriptions hold for the row-vectors associated with factor designated by 2, namely F_2:

$$F_2 \ (\equiv B): \quad \left[\lambda^{(2)}_{1j}\right], \left[\lambda^{(2)}_{2j}\right], \left[\lambda^{(2)}_{3j}\right], \left[\lambda^{(2)}_{4j}\right] .$$

The equation for the general element of the interaction matrix, C_{pq}, namely the element in the p-th row and q-th column of the interaction matrix, is as follows:

$$C_{pq} = \sum_{k=1}^{N_1} \sum_{j=1}^{N_2} \lambda^{(1)}_{pk} \lambda^{(2)}_{qj} S_{kj} \tag{3d-11}$$

$p = 1$ to $N_1 - 1$ (=3 in this experiment)

$q = 1$ to $N_2 - 1$ (=4 in this experiment)

where N_1 = the number of levels of F_1 (≡ A),

N_2 = the number of levels of F_2 (≡ B).

The AB interaction orthogonal contrast matrix is a 3 x 4 matrix given as follows:

$$[C_{pq}] = \begin{bmatrix} C_{11} & C_{12} & C_{13} & C_{14} \\ C_{21} & C_{22} & C_{23} & C_{24} \\ C_{31} & C_{32} & C_{33} & C_{34} \end{bmatrix} . \tag{3d-12}$$

This notational matrix can be represented in a more familiar symbolism:

$$[C_{pq}] = \begin{bmatrix} A_{Lin}B_{Lin} & A_{Lin}B_{Quad} & A_{Lin}B_{Cubic} & A_{Lin}B_{Quart} \\ A_{Quad}B_{Lin} & A_{Quad}B_{Quad} & A_{Quad}B_{Cubic} & A_{Quad}B_{Quart} \\ A_{Cubic}B_{Lin} & A_{Cubic}B_{Quad} & A_{Cubic}B_{Cubic} & A_{Cubic}B_{Quart} \end{bmatrix} . \tag{3d-13}$$

It is noteworthy to point to the fruitful significance of the positional notation used in equation (3d-11):

(1) There are only two indices (subscripts) in the general symbol C_{pq}. This indicates there are only two factors involved, namely F_1 and F_2. Thus the number of subscripts in the general symbol designating the general element of the contrast matrix equals to the number of factors involved in the experiment.

(2) The first subscript (from left to right) of the general symbol C_{pq} is also the first subscript of the orthogonal coefficients associated with the first factor, F_1.

(3) The second subscript of the general element is the first subscript of the orthogonal coefficients associated with the second factor ($\equiv F_2$).

(4) The second subscripts of $\lambda^{(1)}_{pk}$ and $\lambda^{(2)}_{qj}$ orthogonal coefficients are also the subscripts of the elements of the data matrix (S_{kj} in equation (3d-11)).

(5) For any set of values of the doublet (p, q) each element of the data-matrix is uniquely associated with one and only one element of the $\lambda^{(1)}_{pk}$ and one and only one element of $\lambda^{(2)}_{qj}$ coefficients. Thus in computing C_{23}, the element S_{41} is associated uniquely with $\lambda^{(1)}_{24}$ and with $\lambda^{(2)}_{31}$. This establishes the biunique mapping of the indices of the orthogonal coefficients and those of the elements of the data matrix, for any set of values of the indices of the elements of the contrast matrix. The subscripts of the general term of the contrast matrix identify the sources (factors in their proper positions) of information.

Let us now turn to the problem of utilizing this compact symbolism to obtain the A and B effects. This entails a slight innovation in equation (3d-11). We introduce the symbols $\lambda^{(1)}_{0k}$ and $\lambda^{(2)}_{0j}$ where

(3d-14)
$$\lambda^{(1)}_{0k} = 1 \text{ for } k = 1, 2, \cdots, N_1$$
$$\lambda^{(2)}_{0j} = 1 \text{ for } j = 1, 2, \cdots, N_2 \; .$$

Thus the A and B effects (totals) are given by the two following expressions respectively:

(3d-15)
$$(\text{A-effect}) = C_{p0} = \sum_{k=1}^{N_1} \sum_{j=1}^{N_2} \lambda^{(1)}_{pk} \; \lambda^{(2)}_{0j} S_{kj}$$
$$= \sum_{k=1}^{N_1} \sum_{j=1}^{N_2} \lambda^{(1)}_{pk} S_{kj} \quad ,$$

(3d-16)

$$(\text{B-effect}) = C_{0q} = \sum_{k=1}^{N_1} \sum_{j=1}^{N_2} \lambda_{0k}^{(1)} \lambda_{qj}^{(2)} S_{kj}$$

$$= \sum_{k=1}^{N_1} \sum_{j=1}^{N_2} \lambda_{qj}^{(2)} S_{kj}$$

The total of all observations in the experiment is given by:

(3d-17)

$$C_{00} = \sum_{k=1}^{N_1} \sum_{j=1}^{N_2} \lambda_{0k}^{(1)} \lambda_{0j}^{(2)} S_{kj}$$

$$= \sum_{k=1}^{N_1} \sum_{j=1}^{N_2} S_{kj}$$

Numerical Example

Consider the following experiment, exhibited in the following table:

$\left[S_{kj}\right]$: Data Matrix of 4 x 3 Experiment

k = 1 to 4
j = 1 to 3

	b_1	b_2	b_3	TOTALS
a_1	S_{11} 50	S_{12} 90	S_{13} 130	$\sum a_1 = 270$
a_2	S_{21} 30	S_{22} 80	S_{23} 110	$\sum a_2 = 220$
a_3	S_{31} 70	S_{32} 90	S_{33} 150	$\sum a_3 = 310$
a_4	S_{41} 40	S_{42} 70	S_{43} 90	$\sum a_4 = 200$
TOTALS	$\sum b_1 = 190$	$\sum b_2 = 330$	$\sum b_3 = 480$	$\sum a_k = \sum b_j = \sum \sum S_{kj} = 1{,}000$

TABLE 4B.

(3d-18)
$$\left[\lambda^{(1)}_{pk}\right] = \begin{bmatrix} -3 & -1 & +1 & +3 \\ +1 & -1 & -1 & +1 \\ -1 & +3 & -3 & +1 \end{bmatrix}$$

$p = 1$ to 3
$k = 1$ to 4

(3d-19)
$$\left[\lambda^{(2)}_{qj}\right] = \begin{bmatrix} -1 & 0 & +1 \\ +1 & -2 & +1 \end{bmatrix}$$

$q = 1, 2;\ j = 1, 2, 3$

(3d-20)
$$N_1 = 4$$
$$N_2 = 3 \quad .$$

The effects (total) of A are given by:

(3d-21)
$$C_{p0} = \sum_{k=1}^{4} \sum_{j=1}^{3} \lambda^{(1)}_{pk} \ \lambda^{(2)}_{0j} \, S_{kj}$$

$p = 1, 2, 3\ (=N_1 - 1).$

The <u>Linear effect (total)</u> of A is given by:

(3d-22)
$$C_{10} = \sum_{k=1}^{4} \sum_{j=1}^{3} \lambda^{(1)}_{1k} \, S_{kj}$$

(3d-23)
$$C_{10} = \lambda_{11}^{(1)} S_{11} + \lambda_{11}^{(1)} S_{12} + \lambda_{11}^{(1)} S_{13} + \lambda_{12}^{(1)} S_{21} + \lambda_{12}^{(1)} S_{22} + \lambda_{12}^{(1)} S_{23} + \lambda_{13}^{(1)} S_{31} + \lambda_{13}^{(1)} S_{32} + \lambda_{13}^{(1)} S_{33} + \lambda_{14}^{(1)} S_{41} + \lambda_{14}^{(1)} S_{42} + \lambda_{14}^{(1)} S_{43} .$$

The expression (3d-23) is equivalent to:

(3d-24)
$$\lambda_{11}^{(1)}(S_{11} + S_{12} + S_{13}) + \lambda_{12}^{(1)}(S_{21} + S_{22} + S_{23}) + \lambda_{13}^{(1)}(S_{31} + S_{32} + S_{33}) + \lambda_{14}^{(1)}(S_{41} + S_{42} + S_{43}) .$$

The latter in turn is equivalent to the matrix product:

(3d-25)
$$\begin{bmatrix} \lambda_{11}^{(1)} & \lambda_{12}^{(1)} & \lambda_{13}^{(1)} & \lambda_{14}^{(1)} \end{bmatrix} \begin{bmatrix} \sum a_1 \\ \sum a_2 \\ \sum a_3 \\ \sum a_4 \end{bmatrix}$$

(For $\sum a_1$, $\sum a_2$, etc. see TABLE 5),

$$C_{10} = (-3)(270) + (-1)(220) + 1(310) + 3(200) = \underline{-120} .$$

The Linear effect of $A = -120/12 = -10$ (per treatment).

Likewise, C_{20}, the quadratic effect (total) of A equals to the matrix product:

$$C_{20} = \left[\lambda_{21}^{(1)}\ \lambda_{22}^{(1)}\ \lambda_{23}^{(1)}\ \lambda_{24}^{(1)}\right] \begin{bmatrix} \Sigma a_1 \\ \Sigma a_2 \\ \Sigma a_3 \\ \Sigma a_4 \end{bmatrix}. \tag{3d-26}$$

$$= (+1)(270) + (-1)(220) + (-1)(310) + (+1)(200) = -60.$$

The quadratic effect of $A = -60/12 = \underline{-5}$ (per treatment)

(The number of observations is twelve).

The cubic effect (total) of A: C_{30} $(= \sum_{k=1}^{4} \sum_{j=1}^{3} \lambda_{3k}^{(1)} S_{kj})$ can be represented by:

$$C_{30} = \left[\lambda_{31}^{(1)}\ \lambda_{32}^{(1)}\ \lambda_{33}^{(1)}\ \lambda_{34}^{(1)}\right] \begin{bmatrix} \Sigma a_1 \\ \Sigma a_2 \\ \Sigma a_3 \\ \Sigma a_4 \end{bmatrix}. \tag{3d-27}$$

Thus, C_{30}, the cubic effect total of factor A, equals $(-1)(270) + 3(220) - 3(310) + 1(200) = -340$.

The cubic effect of $A = -340/12$

$\approx \underline{-28.34}$ (per treatment)

$$(3d\text{-}28)\qquad \begin{aligned}\left[\lambda^{(1)}_{1k}\right] &= \left[-3\ \ -1\ \ +1\ \ +3\right]\\ \left[\lambda^{(1)}_{2k}\right] &= \left[+1\ \ -1\ \ -1\ \ +1\right]\\ \left[\lambda^{(1)}_{3k}\right] &= \left[-1\ \ +3\ \ -3\ \ +1\right].\end{aligned}$$

Similar procedures can be employed to obtain the linear and quadratic components of the B-effect. Now $\sum b_i$ replaces the $\sum a_i$ in expressions (3d-25) to (3d-28); and the $\lambda^{(2)}$-matrix of (3d-19) replaces the $\lambda^{(1)}$-matrix in these expressions. Expression (3d-16) constitutes a more general formula for obtaining the B-effect.

Let us now compute some elements of the AB interaction contrast matrix:

The expression for C_{22} is given below

$$(3d\text{-}29)\qquad C_{22}\,(\equiv A_Q B_Q) = \sum_{k=1}^{N_1=4}\ \sum_{j=1}^{N_2=3} \lambda^{(1)}_{2k}\ \lambda^{(2)}_{2j}\ S_{kj}\ ,$$

where $\left[\lambda^{(1)}_{2k}\right] = \left[+1\ \ -1\ \ -1\ \ +1\right]$; $k = 1$ to 4

and $\left[\lambda^{(2)}_{2j}\right] = \left[+1\ \ -2\ \ +1\right]$; $j = 1$ to 3

$$(3d\text{-}30)\qquad \begin{aligned} C_{22} = {} & \lambda^{(1)}_{21}\lambda^{(2)}_{21}S_{11} + \lambda^{(1)}_{21}\lambda^{(2)}_{22}S_{12} + \lambda^{(1)}_{21}\lambda^{(2)}_{23}S_{13}\\ & + \lambda^{(1)}_{22}\lambda^{(2)}_{21}S_{21} + \lambda^{(1)}_{22}\lambda^{(2)}_{22}S_{22} + \lambda^{(1)}_{22}\lambda^{(2)}_{23}S_{23}\\ & + \lambda^{(1)}_{23}\lambda^{(2)}_{21}S_{31} + \lambda^{(1)}_{23}\lambda^{(2)}_{22}S_{32} + \lambda^{(1)}_{23}\lambda^{(2)}_{23}S_{33}\\ & + \lambda^{(1)}_{24}\lambda^{(2)}_{21}S_{41} + \lambda^{(1)}_{24}\lambda^{(2)}_{22}S_{42} + \lambda^{(1)}_{24}\lambda^{(2)}_{23}S_{43}\end{aligned}$$

(3d-31)

$$C_{22} = \left\{ \begin{array}{l} (+1)\ (+1)\ \ 50 + (+1)\ (-2)\ \ 90 + (+1)\ (+1)\ \ 130 = 0 \\ + (-1)\ (+1)\ \ 30 + (-1)\ (-2)\ \ 80 + (-1)\ (+1)\ \ 110 = +20 \\ + (-1)\ (+1)\ \ 70 + (-1)\ (-2)\ \ 90 + (-1)\ (+1)\ \ 150 = -40 \\ + (+1)\ (+1)\ \ 40 + (+1)\ (-2)\ \ 70 + (\ 1)\ (+1)\ \ 90 = -10 \end{array} \right\} = -30.$$

What is done in (3d-30) and (3d-31) is to superimpose on the data matrix, S_{kj}, the matrix formed from $\sum_{k=1}^{N_1=4} \sum_{j=1}^{N_2=3} \lambda^{(1)}_{2k} \lambda^{(2)}_{2j}$ namely, the matrix:

$$\begin{bmatrix} \lambda^{(1)}_{21} \lambda^{(2)}_{21} & \cdots\cdots & \lambda^{(1)}_{21} \lambda^{(2)}_{23} \\ \vdots & & \vdots \\ \lambda^{(1)}_{24} \lambda^{(2)}_{21} & \cdots\cdots & \lambda^{(1)}_{24} \lambda^{(2)}_{23} \end{bmatrix}.$$

The superimposition is unique in that every element S_{kj} is multiplied by the product $\lambda^{(1)}_{2k} \lambda^{(2)}_{2j}$ with due regard to the positional sigificance of the indices. This is clearly seen in (3d-30).

(3d-32)

$$C_{10}\,(\equiv A_{Lin}) = \sum_{k=1}^{N_1=4} \sum_{j=1}^{N_2=3} \lambda^{(1)}_{1k} \lambda^{(2)}_{0j} S_{kj}$$

where $\left[\lambda^{(1)}_{1k}\right] = \left[-3 \quad -1 \quad +1 \quad +3\right]$; $\lambda^{(2)}_{0j} = 1$, for all j

(3d-33)

$$C_{10} = \begin{cases} (-3)\ 50 + (-3)\ 90 + (-3)\ 130 = -810 \\ (-1)\ 30 + (-1)\ 80 + (-1)\ 110 = -220 \\ (+1)\ 70 + (1)\ 90 + (1)\ 150 = +310 \\ (+3)\ 40 + (3)\ 70 + (3)\ 90 = 600 \end{cases}$$

$$C_{10} = \underline{-120}.$$

We can include the <u>total sum of observations</u> (C_{00}), the A and B effects totals by having p, and q, in C_{pq}, assume the values 0; <u>p = 0, 1, 2, 3</u>; <u>q= 0, 1, 2</u>.

(3d-34)
$$\left[C_{pq}\right] = \begin{bmatrix} C_{00} & C_{01} & C_{02} \\ C_{10} & C_{11} & C_{12} \\ C_{20} & C_{21} & C_{22} \\ C_{30} & C_{31} & C_{32} \end{bmatrix}$$

(3d-35)
$$\left[C_{pq}\right] = \begin{bmatrix} 1{,}000 & +290 & +10 \\ -120 & -90 & +30 \\ -60 & -30 & -30 \\ -340 & -30 & -190 \end{bmatrix}.$$

It is to be noted that C_{00}, <u>the total sum</u>, <u>does not represent a contrast</u>.

e. The Extension of Formula and Algorithm to Three-, and Four-Factor Orthogonal Designs.

The Case of Three Factors: In the case of three factors A ($\equiv F_1$), B ($\equiv F_2$), and C ($\equiv F_3$), the elements of the interaction contrast matrix can be generated by the following expression (It is immaterial which factors are first, second, third, etc.):

$$C_{pqr} = \sum_{k=1}^{N_1} \sum_{j=1}^{N_2} \sum_{\ell=1}^{N_3} \lambda_{pk}^{(1)} \lambda_{qj}^{(2)} \lambda_{r\ell}^{(3)} S_{kj\ell} \quad . \tag{3e-1}$$

From the positional notation, it is evident that p, q, r refer to F_1, F_2, F_3 respectively. Note the symmetrical arrangement of the indices p, q, r (in the general element of the contrast matrix) and that of the indices k, j, ℓ (the indices of the elements of the data matrix).

If

$$N_1 = 4, \; N_2 = 3, \; N_3 = 2 \tag{3e-2}$$

p, q, r take one of the following values:

$$\begin{aligned} p &= \{1, 2, 3\} \\ q &= \{1, 2\} \\ r &= \{1\} \end{aligned} \tag{3e-3}$$

The interaction matrix (in groups of three factor effects) consists of six elements, and can be represented as follows:

$$(3e\text{-}4) \qquad \left[C_{pqr}\right] = \begin{bmatrix} C_{111} & C_{121} \\ C_{211} & C_{221} \\ C_{311} & C_{321} \end{bmatrix} .$$

A particular element of this matrix, for instance C_{121} $(=A_L B_Q C_L)$ can be computed by the following expression (using (3e-1)):

$$(3e\text{-}5) \qquad C_{121} = \sum_{k=1}^{4} \sum_{j=1}^{3} \sum_{\ell=1}^{2} \lambda^{(1)}_{1k} \lambda^{(2)}_{2j} \lambda^{(3)}_{1\ell} S_{kj\ell} .$$

It is evident from (3e-4) that each element of the data matrix, $S_{kj\ell}$, is multiplied by its uniquely associated multiplier consisting of the product of three unique elements from the orthogonal polynomial matrices of the factors A, B, and C respectively. Thus for each element of the contrast matrix there is a biunique one-to-one correspondence between a set of multipliers and the elements of the data-matrix.

In computing, for example, C_{121}, the element of the data-matrix, say S_{432}, is multiplied by the mutiplier, $\lambda^{(1)}_{14} \lambda^{(2)}_{23} \lambda^{(3)}_{12}$; the element S_{321} by $\lambda^{(1)}_{13} \lambda^{(2)}_{22} \lambda^{(3)}_{11}$, etc.

Note that the indices 1, 2, 1 of the contrast element C_{121} appear sequentially (in accordance with the positional notation) as the first indices of $\lambda^{(1)}$, $\lambda^{(2)}$, $\lambda^{(3)}$ respectively (as subscripts).

The computational scheme for obtaining the contrast C_{121} under the condition (3e-2) can be displayed as follows:

(3e-6)

$$C_{121} = \begin{bmatrix} \lambda^{(1)}_{11}\lambda^{(2)}_{21}\lambda^{(3)}_{11}S_{111} + \lambda^{(1)}_{11}\lambda^{(2)}_{21}\lambda^{(3)}_{12}S_{112} + \cdots + \lambda^{(1)}_{11}\lambda^{(2)}_{23}\lambda^{(3)}_{12}S_{132} \\ + \lambda^{(1)}_{12}\lambda^{(2)}_{21}\lambda^{(3)}_{11}S_{211} + \lambda^{(1)}_{12}\lambda^{(2)}_{21}\lambda^{(3)}_{12}S_{212} + \cdots + \lambda^{(1)}_{12}\lambda^{(2)}_{23}\lambda^{(3)}_{12}S_{232} \\ + \lambda^{(1)}_{13}\lambda^{(2)}_{21}\lambda^{(3)}_{11}S_{311} + \lambda^{(1)}_{13}\lambda^{(2)}_{21}\lambda^{(3)}_{12}S_{312} + \cdots + \lambda^{(1)}_{13}\lambda^{(2)}_{23}\lambda^{(3)}_{12}S_{332} \\ + \lambda^{(1)}_{14}\lambda^{(2)}_{21}\lambda^{(3)}_{11}S_{411} + \lambda^{(1)}_{14}\lambda^{(2)}_{21}\lambda^{(3)}_{12}S_{412} + \cdots + \lambda^{(1)}_{14}\lambda^{(2)}_{23}\lambda^{(3)}_{12}S_{432} \end{bmatrix} .$$

Let $S_{kj\ell}$ consist of the following data:

(3e-7)

$$\left.\begin{matrix} k=1,\ldots,4 \\ j=1,\ldots,3 \\ \ell=1,2 \end{matrix}\right\} \qquad \left[S_{kj\ell}\right] = \begin{bmatrix} 2 & 3 & 4 & 5 & 6 & 7 \\ 1 & 2 & 3 & 5 & 5 & 6 \\ 3 & 4 & 4 & 5 & 7 & 8 \\ 1 & 3 & 3 & 4 & 4 & 5 \end{bmatrix}$$

(3e-8)

$$N_A = N_1 = 4$$
$$N_B = N_2 = 3$$
$$N_C = N_3 = 2 \quad .$$

The orthogonal polynomial matrices for the factors A , B , and C are given as follows:

(3e-9)

$$[\lambda(A)] = \begin{bmatrix} \lambda^{(1)}_{11} & \lambda^{(1)}_{12} & \lambda^{(1)}_{13} & \lambda^{(1)}_{14} \\ \lambda^{(1)}_{21} & \lambda^{(1)}_{22} & \lambda^{(1)}_{23} & \lambda^{(1)}_{24} \\ \lambda^{(1)}_{31} & \lambda^{(1)}_{32} & \lambda^{(1)}_{33} & \lambda^{(1)}_{34} \end{bmatrix}$$

$(N_1 = 4)$

(3e-10)
$$\left[\lambda^{(1)}(A)\right] = \begin{bmatrix} -3 & -1 & +1 & +3 \\ +1 & -1 & -1 & +1 \\ -1 & +3 & -3 & +1 \end{bmatrix}$$

(3e-11)
$$\left[\lambda^{(1)}_{1k}\right] = \begin{bmatrix} -3 & -1 & +1 & +3 \end{bmatrix}$$
$$\left[\lambda^{(1)}_{2k}\right] = \begin{bmatrix} +1 & -1 & -1 & +1 \end{bmatrix}$$
$$\left[\lambda^{(1)}_{3k}\right] = \begin{bmatrix} -1 & +3 & -3 & +1 \end{bmatrix}$$

$\underline{k = 1, 2, 3, 4}$

(3e-12)
$$\left[\lambda^{(2)}(B)\right] = \begin{bmatrix} \lambda^{(2)}_{11} & \lambda^{(2)}_{12} & \lambda^{(2)}_{13} \\ \lambda^{(2)}_{21} & \lambda^{(2)}_{22} & \lambda^{(2)}_{23} \end{bmatrix} = \begin{bmatrix} -1 & 0 & +1 \\ +1 & -2 & +1 \end{bmatrix},$$

where

(3e-13)
$$\left[\lambda^{(2)}_{1j}\right] = \left[-1 \quad 0 \quad +1\right]$$
$$\left[\lambda^{(2)}_{2j}\right] = \left[1 \quad -2 \quad +1\right]$$
$$\underline{j = 1,\ 2,\ 3}$$

(3e-14)
$$\left[\lambda^{(3)}(C)\right] = \left[\lambda^{(3)}_{11} \quad \lambda^{(3)}_{12}\right] = \left[-1 \quad +1\right].$$

(3e-15)
$$\left[\lambda^{(3)}_{1\ell}\right] = \left[-1 \quad +1\right]$$
$$\underline{\ell = 1,\ 2.}$$

To compute the contrast C_{121} $(\equiv A_L B_Q C_L)$ the array of the product of multipliers $\left\{\lambda^{(1)}_{1k}, \quad \lambda^{(2)}_{2j}, \quad \lambda^{(3)}_{1\ell}\right\}$ is given as follows:

(3e-16)
$$\begin{bmatrix} 3 & -3 & -6 & 6 & 3 & -3 \\ 1 & -1 & -2 & 2 & 1 & -1 \\ -1 & 1 & 2 & -2 & -1 & 1 \\ -3 & 3 & 6 & -6 & -3 & 3 \end{bmatrix}.$$

They represent the values of the corresponding elements of the array:

$$
(3e\text{-}17)\quad \begin{bmatrix}
\lambda_{11}^{(1)}\lambda_{21}^{(2)}\lambda_{11}^{(3)} & \lambda_{11}^{(1)}\lambda_{21}^{(2)}\lambda_{12}^{(3)} & \cdots & \lambda_{11}^{(1)}\lambda_{23}^{(2)}\lambda_{12}^{(3)} \\
\lambda_{12}^{(1)}\lambda_{21}^{(2)}\lambda_{11}^{(3)} & \lambda_{12}^{(1)}\lambda_{21}^{(2)}\lambda_{12}^{(3)} & \cdots & \lambda_{12}^{(1)}\lambda_{23}^{(2)}\lambda_{12}^{(3)} \\
\lambda_{13}^{(1)}\lambda_{21}^{(2)}\lambda_{11}^{(3)} & \lambda_{13}^{(1)}\lambda_{21}^{(2)}\lambda_{12}^{(3)} & \cdots & \lambda_{13}^{(1)}\lambda_{23}^{(2)}\lambda_{12}^{(3)} \\
\lambda_{14}^{(1)}\lambda_{21}^{(2)}\lambda_{11}^{(3)} & \lambda_{14}^{(1)}\lambda_{21}^{(2)}\lambda_{12}^{(3)} & \cdots & \lambda_{14}^{(1)}\lambda_{23}^{(2)}\lambda_{12}^{(3)}
\end{bmatrix}.
$$

Upon multiplying the values in (3e-16) by their uniquely corresponding elements of the data matrix of (3e-7) and adding we obtain:

$$
(3e\text{-}18)\quad \left\{\begin{array}{lcr}
 & 6\ -9\ -24\ +30\ +18\ -21 = & 0 \\
+ & & \\
 & 1\ -2\ -6\ +10\ +5\ -6 = & 2 \\
+ & & \\
 & -3\ +4\ +8\ -10\ -7\ +8 = & 0 \\
+ & & \\
 & -3\ +9\ +18\ -24\ -12 + 15 = & \underline{3} \\
 & & 5
\end{array}\right.
$$

$$
(3e\text{-}19)\qquad C_{121} = (\equiv A_L B_Q C_L) = 5.
$$

This method enables one to superimpose uniquely the matrix of products of orthogonal polynomial coefficients, as given in (3e-17) in this example, on the data-matrix of (3e-7) with the object of computing the elements of the contrast matrix.

Likewise all the elements of the contrast matrix given in (3e-4) can be computed in this manner.

If the values of p, q, r in (3e-3) are extended to include zero,

(3e-20) $$p = \{0, 1, 2, 3\}, \quad q = \{0, 1, 2\}, \quad r = \{0, 1\}$$

one obtains, in addition to three factor effects, two factor effects, linear effects (main effects totals), and the sum of the total number of observations. The latter is not a contrast. The matrix is now given by the following expression:

(3e-21) $$C_{pqr} = \sum_{k=1}^{4} \sum_{j=1}^{3} \sum_{\ell=1}^{2} \lambda_{pk}^{(1)} \lambda_{qj}^{(2)} \lambda_{r\ell}^{(3)} s_{kj\ell}$$

$$(N_1 = 4, \quad N_2 = 3, \quad N_3 = 2)$$

where

(3e-22) $$\lambda_{0k}^{(1)} = 1, \quad \lambda_{0j}^{(2)} = 1, \quad \lambda_{0\ell}^{(3)} = 1$$

for all values of k, j, ℓ.

The contrast matrix of the elements given in (3e-21), subject to the conditions of (3e-20), is given as follows:

(3e-23)

$$[C_{pqr}] \equiv \begin{bmatrix} C_{000} & C_{001} & C_{010} & C_{011} & C_{020} & C_{021} \\ C_{100} & C_{101} & C_{110} & C_{111} & C_{120} & C_{121} \\ C_{200} & C_{201} & C_{210} & C_{211} & C_{220} & C_{221} \\ C_{300} & C_{301} & C_{310} & C_{311} & C_{320} & C_{321} \end{bmatrix}.$$

There are twenty four (4 x 3 x 2) elements of which C_{000}, the total sum of observations, is not a contrast.

The terms, two indices of which are equal to zero, are linear effects (totals); the terms one index of which is zero constitute the two-factor effects (totals); the terms all three indices of which are not zero are the three-factor effects (totals). There is only one element (not a contrast) all the indices of which are equal to zero; this is the total sum of all the observations (C_{000}).

The Case of Four Factors: For a design of four factors, with the levels of each factor equally spaced (the spacings of one factor need not be the same as those of the levels of another) the expression of the general element is given by:

(3e-24)

$$C_{pqrs} = \sum_{k=1}^{N_1} \sum_{j=1}^{N_2} \sum_{\ell=1}^{N_3} \sum_{m=1}^{N_4} \lambda^{(1)}_{pk} \lambda^{(2)}_{qj} \lambda^{(3)}_{r\ell} \lambda^{(4)}_{sm} S_{kj\ell m}$$

(3e-25)

$$\begin{aligned} p &= 0, 1, \cdots, N_1 - 1 \\ q &= 0, 1, \cdots, N_2 - 1 \\ r &= 0, 1, \cdots, N_3 - 1 \\ s &= 0, 1, \cdots, N_4 - 1 \end{aligned}$$

where:

(3e-26)

$$N_1 = \text{number of levels of } F_1 \ (\equiv A)$$

$$N_2 = \text{number of levels of } F_2 \ (\equiv B)$$

$$N_3 = \text{number of levels of } F_3 \ (\equiv C)$$

$$N_4 = \text{number of levels of } F_4 \ (\equiv D) \ .$$

The contrast for any two-factor interaction, say, the AC interaction can be obtained by means of the following expression:

(3e-27)

$$C_{p0r0} = \sum_{k=1}^{N_1} \sum_{j=1}^{N_2} \sum_{\ell=1}^{N_3} \sum_{m=1}^{N_4} \lambda_{pk}^{(1)} \lambda_{0j}^{(2)} \lambda_{r\ell}^{(3)} \lambda_{0m}^{(4)} S_{kj\ell m} \ .$$

Since $\lambda_{0j}^{(2)} = \lambda_{0m}^{(4)} = 1$, C_{p0r0} becomes

(3e-28)

$$C_{p0r0} = \sum_{k=1}^{N_1} \sum_{j=1}^{N_2} \sum_{\ell=1}^{N_3} \sum_{m=1}^{N_4} \lambda_{pk}^{(1)} (1) \lambda_{r\ell}^{(3)} (1) S_{kj\ell m} \ .$$

Because of the positional notation of the indices, the zero's in C_{p0r0} indicate the <u>elimination as regards interaction effects</u> of the factors with corresponding <u>positional places</u>, namely, in this case F_2 ($\equiv$ B) and F_4 ($\equiv$ D). This is indicated in the righthand member of (3e-28) by equating to one (1), the polynomial coefficients corresponding to these factors (F_2 and F_4), namely $\lambda_{0j}^{(2)}$ and $\lambda_{0m}^{(4)}$. The appearance of a zero in a given position of the indices of the general term for the contrast constitutes a rule for filling in the number 1 for the value of the positionally corresponding polynomial coefficient.

As an example, in computing C_{1020} ($\equiv A_L C_Q$), the element of the data-matrix, say, S_{3245} ($N_1 \geqslant 3$; $N_2 \geqslant 2$; $N_3 \geqslant 4$; $N_4 \geqslant 5$) is multiplied by the product $\lambda_{13}^{(1)}(1)\,\lambda_{24}^{(3)}(1)$; and the element S_{2134} is multiplied by $\lambda_{12}^{(1)}(1)\,\lambda_{23}^{(3)}(1)$. A similar procedure is used with regards to the remaining elements of the data-matrix. Note that the indices 1 and 2, occurring in the first and third places (from left to right) in C_{1020} correspond to their occurrence in the first position indices of $\lambda^{(1)}$ and $\lambda^{(3)}$ namely, the orthogonal polynomial coefficients corresponding to F_1 ($\equiv$ A) and F_3 ($\equiv$C) respectively. Likewise, C_{0qr0} is given by:

(3e-29)

$$C_{0qr0} = \sum_{k=1}^{N_1} \sum_{j=1}^{N_2} \sum_{\ell=1}^{N_3} \sum_{m=1}^{N_4} \lambda_{0k}^{(1)}\,\lambda_{qj}^{(2)}\,\lambda_{r\ell}^{(3)}\,\lambda_{0m}^{(4)}\,S_{kj\ell m}$$

$$C_{0qr0} = \sum_{k=1}^{N_1} \sum_{j=1}^{N_2} \sum_{\ell=1}^{N_3} \sum_{m=1}^{N_4} (1)\,\lambda_{qj}^{(2)}\,\lambda_{r\ell}^{(3)}(1)\,S_{kj\ell m}\,.$$

This scheme can readily be extended to any finite number of factors, as is shown in the next section.

f. A Generalized Formula and Algorithm for Orthogonal Contrasts

The generalization of the formulas given in the preceding section can be carried out in the following manner:

Instead of the symbols p, q, r, s, t $\cdots$, as subscripts in the expression for the general term for the orthogonal contrasts involving m factors, we shall use the symbols: i_1, i_2, i_3, $\cdots i_m$. i_1 refers to F_1; i_2 to factor F_2; and i_m to F_m (the m-th factor). The contrast is written as follows:

$$C_{i_1 i_2 i_3 \cdots i_m} .$$

The symbols j_1, j_2, j_3 $\cdots$ j_m will be used as subscripts for identifying the elements of the data-matrix. The elements of the data-matrix will be symbolized by: $S_{j_1 j_2 j_3 \ldots j_m}$, where j_1 refers to the number of levels of F_1, j_2 to the number of levels of F_2, $\ldots$, and j_m refers to the number of levels of F_m, the m-th factor. (All levels are equally spaced.)

The doublets $i_1 j_1$; $i_2 j_2$; $i_3 j_3$; $\cdots$, $i_m j_m$ will be used as subscript indices of the orthogonal polynomial matrices (the λ's) corresponding to F_1; F_2; F_3; $\cdots$; F_m. To avoid confusion we shall adopt the superscript symbols (1); (2); (3); $\cdots$; (m) for the λ's:

$$\lambda^{(1)}_{i_1 j_1};\ \lambda^{(2)}_{i_2 j_2};\ \lambda^{(3)}_{i_3 j_3};\ \ldots;\ \lambda^{(m)}_{i_m j_m} .$$

Note that the first subscript for the λ's is taken from the set of subscripts $\{i_k\}$ $(1 \leq k \leq m)$ of contrast, $C_{i_1 i_2 \cdots i_m}$.

The second subscript for the λ's is taken from the set of subscripts $\{j_k\}$ $(1 \leq k \leq m)$ of the data matrix, $S_{1}\,j_2 j_3 \cdots j_m$.

The following represents the formulation, explanation, and definition of the expressions and algorithm for generating orthogonal contrasts (note that the total sum of observations, $C_{00 \ldots 0}$, is not a contrast).

LOGIC, SYMBOLIC NOTATION, EXPRESSIONS AND FORMULAS FOR ORTHOGONAL CONTRASTS

ELEMENT OF DATA-MATRIX FOR FACTORS $F_1, F_2, \ldots, F_m$.

$$S_{j_1 j_2 \ldots j_m}$$

SUBSCRIPTS: $j_1, j_2, \ldots, j_m$

$j_1 = 1, 2, \ldots, N_1$

$j_2 = 1, 2, \ldots, N_2$

. . . .

$j_m = 1, 2, \ldots, N_m$

FACTORS: $F_1, F_2, \ldots, F_m$

LEVELS OF FACTORS:

N_1 = NUMBER OF LEVELS OF F_1

N_2 = " " " " F_2

. . . .

N_m = " " " " F_m

VALUES OF SUBSCRIPTS:

$i_1 = 0, 1, \ldots, N_1 - 1$	$j_1 = 1, 2, \ldots, N_1$
$i_2 = 0, 1, \ldots, N_2 - 1$	$j_2 = 1, 2, \ldots, N_2$
. . . .	. . . .
$i_m = 0, 1, \ldots, N_m - 1$	$j_m = 1, 2, \ldots, N_m$

GENERAL ELEMENT OF ORTHOGONAL CONTRAST MATRIX FOR FACTORS $F_1, F_2, \ldots, F_m$.

$$C_{i_1 i_2 \ldots i_m} = \sum_{j_1=1}^{N_1} \sum_{j_2=1}^{N_2} \cdots \sum_{j_m=1}^{N_m} \lambda^{(1)}_{i_1 j_1} \lambda^{(2)}_{i_2 j_2} \cdots \lambda^{(m)}_{i_m j_m} S_{j_1 j_2 \ldots j_m}$$

$$= \sum_{\overline{j_1 j_2 \ldots j_m}}^{\overline{N_1, N_2, \ldots, N_m}} \lambda^{(1)}_{i_1 j_1} \lambda^{(2)}_{i_2 j_2} \cdots \lambda^{(m)}_{i_m j_m} S_{j_1 j_2 \ldots j_m}$$

WHERE

$$\sum_{j_1=1}^{N_1} \sum_{j_2=1}^{N_2} \cdots \sum_{j_m=1}^{N_m} \equiv \sum_{\overline{j_1 j_2 \ldots j_m}}^{\overline{N_1, N_2, \ldots, N_m}}$$

NOTE: $\lambda_{i_m j_m}$ IS THE ELEMENT IN THE i^{th} ROW AND j^{th} COLUMN OF THE ORTHOGONAL COEFFICIENT MATRIX FOR THE m^{th} FACTOR F_m.

MATRIX OF ORTHOGONAL POLYNOMIAL COEFFICIENTS OF m th FACTOR F_m.

$$\left[\lambda^{(m)}_{i_m j_m}\right] = \begin{bmatrix} \lambda^{(m)}_{11} & \lambda^{(m)}_{12} & \cdots & \lambda^{(m)}_{1,N_m} \\ \lambda^{(m)}_{31} & \lambda^{(m)}_{32} & \cdots & \lambda^{(m)}_{3,N_m} \\ \lambda^{(m)}_{31} & \lambda^{(m)}_{32} & \cdots & \lambda^{(m)}_{3,N_m} \\ \cdots & \cdots & \cdots & \cdots \\ \lambda^{(m)}_{N_m-1,1} & \lambda^{(m)}_{N_m-1,2} & \cdots & \lambda^{(m)}_{N_m-1,N_m} \end{bmatrix} = \begin{bmatrix} \text{LINEAR CONTRAST COEFFICIENTS} \\ \text{QUADRATIC} \quad \text{"} \quad\quad \text{"} \\ \text{CUBIC} \quad \text{"} \quad\quad \text{"} \\ \cdots\cdots\cdots\cdots \\ (N_m-1)\text{th} \quad \text{"} \quad\quad \text{"} \end{bmatrix}$$

MATRIX METHOD FOR OBTAINING A-EFFECT IN A TWO FACTOR DESIGN

$N_A = 4$ (FOUR LEVELS OF A)

$N_B = 5$ (FIVE LEVELS OF B)

$$\underset{(K=1,2,3)}{\left[C_K(A)\right]} = \begin{bmatrix} \lambda_{11}^{(1)} & \lambda_{12}^{(1)} & \lambda_{13}^{(1)} & \lambda_{14}^{(1)} \\ \lambda_{21}^{(1)} & \lambda_{22}^{(1)} & \lambda_{23}^{(1)} & \lambda_{24}^{(1)} \\ \lambda_{31}^{(1)} & \lambda_{32}^{(1)} & \lambda_{33}^{(1)} & \lambda_{34}^{(1)} \end{bmatrix} \begin{bmatrix} S_{11} & S_{12} & S_{13} & S_{14} & S_{15} \\ S_{21} & S_{22} & S_{23} & S_{24} & S_{25} \\ S_{31} & S_{32} & S_{33} & S_{34} & S_{35} \\ S_{41} & S_{42} & S_{43} & S_{44} & S_{45} \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} \lambda_{11}^{(1)} & \lambda_{12}^{(1)} & \lambda_{13}^{(1)} & \lambda_{14}^{(1)} \\ \lambda_{21}^{(1)} & \lambda_{22}^{(1)} & \lambda_{23}^{(1)} & \lambda_{24}^{(1)} \\ \lambda_{31}^{(1)} & \lambda_{32}^{(1)} & \lambda_{33}^{(1)} & \lambda_{34}^{(1)} \end{bmatrix} \begin{bmatrix} \Sigma S_{1j} \\ \Sigma S_{2j} \\ \Sigma S_{3j} \\ \Sigma S_{4j} \end{bmatrix}$$

MATRIX METHOD FOR OBTAINING B-EFFECT IN A TWO-FACTOR DESIGN

$N_A = 4$ (FOUR LEVELS OF A)
$N_B = 5$ (FIVE LEVELS OF B)

$$
\left[C_k(B)\right] = \begin{bmatrix} \lambda_{11}^{(2)} & \lambda_{12}^{(2)} & \lambda_{13}^{(2)} & \lambda_{14}^{(2)} & \lambda_{15}^{(2)} \\ \lambda_{21}^{(2)} & \lambda_{22}^{(2)} & \lambda_{23}^{(2)} & \lambda_{24}^{(2)} & \lambda_{25}^{(2)} \\ \lambda_{31}^{(2)} & \lambda_{32}^{(2)} & \lambda_{33}^{(2)} & \lambda_{34}^{(2)} & \lambda_{35}^{(2)} \\ \lambda_{41}^{(2)} & \lambda_{42}^{(2)} & \lambda_{43}^{(2)} & \lambda_{44}^{(2)} & \lambda_{45}^{(2)} \end{bmatrix} \left\{ \begin{bmatrix} 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} S_{11} & S_{12} & S_{13} & S_{14} & S_{15} \\ S_{21} & S_{22} & S_{23} & S_{24} & S_{25} \\ S_{31} & S_{32} & S_{33} & S_{34} & S_{35} \\ S_{41} & S_{42} & S_{43} & S_{44} & S_{45} \end{bmatrix} \right\}^{T}
$$

$(k = 1, 2, 3, 4)$

$$
= \begin{bmatrix} \lambda_{11}^{(2)} & \lambda_{12}^{(2)} & \lambda_{13}^{(2)} & \lambda_{14}^{(2)} & \lambda_{15}^{(2)} \\ \lambda_{21}^{(2)} & \lambda_{22}^{(2)} & \lambda_{23}^{(2)} & \lambda_{24}^{(2)} & \lambda_{25}^{(2)} \\ \lambda_{31}^{(2)} & \lambda_{32}^{(2)} & \lambda_{33}^{(2)} & \lambda_{34}^{(2)} & \lambda_{35}^{(2)} \\ \lambda_{41}^{(2)} & \lambda_{42}^{(2)} & \lambda_{43}^{(2)} & \lambda_{44}^{(2)} & \lambda_{45}^{(2)} \end{bmatrix} \begin{bmatrix} \Sigma S_{i1} \\ \Sigma S_{i2} \\ \Sigma S_{i3} \\ \Sigma S_{i4} \\ \Sigma S_{i5} \end{bmatrix}
$$

THE INTERACTION CONTRAST MATRIX FOR A TWO FACTOR EXPERIMENT

$N_1 = 4, \quad N_2 = 5$

$p = 1, 2, 3, = N_1 - 1$

$q = 1, 2, 3, 4 = N_2 - 1$

$$[C_{pq}] = \begin{bmatrix} \Sigma\Sigma \lambda^{(1)}_{1i} \lambda^{(2)}_{1j} S_{ij} & \Sigma\Sigma \lambda^{(1)}_{1i} \lambda^{(2)}_{2j} S_{ij} & \Sigma\Sigma \lambda^{(1)}_{1i} \lambda^{(2)}_{3j} S_{ij} & \Sigma\Sigma \lambda^{(1)}_{1i} \lambda^{(2)}_{4j} S_{ij} \\ \Sigma\Sigma \lambda^{(1)}_{2i} \lambda^{(2)}_{1j} S_{ij} & \Sigma\Sigma \lambda^{(1)}_{2i} \lambda^{(2)}_{2j} S_{ij} & \Sigma\Sigma \lambda^{(1)}_{2i} \lambda^{(2)}_{3i} S_{ij} & \Sigma\Sigma \lambda^{(1)}_{2i} \lambda^{(2)}_{4j} S_{ij} \\ \Sigma\Sigma \lambda^{(1)}_{3i} \lambda^{(2)}_{1j} S_{ij} & \Sigma\Sigma \lambda^{(1)}_{3i} \lambda^{(2)}_{2j} S_{ij} & \Sigma\Sigma \lambda^{(1)}_{3i} \lambda^{(2)}_{3j} S_{ij} & \Sigma\Sigma \lambda^{(1)}_{3i} \lambda^{(2)}_{4j} S_{ij} \end{bmatrix}$$

WHERE $\Sigma\Sigma \equiv \sum_{i=1}^{N_1=4} \sum_{j=1}^{N_2=5}$

MATRIX METHOD FOR OBTAINING INTERACTION CONTRASTS FOR A TWO FACTOR EXPERIMENT

$N_1 = 4 \qquad N_2 = 5$

$$\left[C_K(AB)\right] = \begin{bmatrix} \Sigma L_i^{(1)} S_{i1} & \Sigma L_i^{(1)} S_{i2} & \Sigma L_i^{(1)} S_{i3} & \Sigma L_i^{(1)} S_{i4} & \Sigma L_i^{(1)} S_{i5} \\ \Sigma Q_i^{(1)} S_{i1} & \Sigma Q_i^{(1)} S_{i2} & \Sigma Q_i^{(1)} S_{i3} & \Sigma Q_i^{(1)} S_{i4} & \Sigma Q_i^{(1)} S_{i5} \\ \Sigma C_i^{(1)} S_{i1} & \Sigma C_i^{(1)} S_{i2} & \Sigma C_i^{(1)} S_{i3} & \Sigma C_i^{(1)} S_{i4} & \Sigma C_i^{(1)} S_{i5} \end{bmatrix} \begin{bmatrix} L_1^{(2)} & Q_1^{(2)} & C_1^{(2)} & q_1^{(2)} \\ L_2^{(2)} & Q_2^{(2)} & C_2^{(2)} & q_2^{(2)} \\ L_3^{(2)} & Q_3^{(2)} & C_3^{(2)} & q_3^{(2)} \\ L_4^{(2)} & Q_4^{(2)} & C_4^{(2)} & q_4^{(2)} \\ L_5^{(2)} & Q_5^{(2)} & C_5^{(2)} & q_5^{(2)} \end{bmatrix}$$

$(K = 1, 2, \ldots 12)$

WHERE $\Sigma \equiv \sum_{i=1}^{N_1 = 4}$

INTERACTION CONTRAST MATRIX FOR A TWO-FACTOR EXPERIMENT

$$N_1 = 4$$
$$N_2 = 5$$
$$p = 1,2,3, \ (= N_1 - 1)$$
$$q = 1,2,3,4 \ (= N_2 - 1)$$

$$\left[C_{pq} \right] = \begin{bmatrix} C_{11} & C_{12} & C_{13} & C_{14} \\ C_{21} & C_{22} & C_{23} & C_{24} \\ C_{31} & C_{32} & C_{33} & C_{34} \end{bmatrix}$$

$$C_{pq} = \sum_{I=1}^{N_1=4} \sum_{J=1}^{N_2=5} \lambda^{(1)}_{pi} \lambda^{(2)}_{qj} S_{ij}$$

MATRIX OF ORTHOGONAL POLYNOMIAL COEFFICIENTS OF m th FACTOR F_m.

$$\left[\lambda^{(m)}_{i_m j_m}\right] = \begin{bmatrix} \lambda^{(m)}_{11} & \lambda^{(m)}_{12} & \cdots & \lambda^{(m)}_{1,N_m} \\ \lambda^{(m)}_{31} & \lambda^{(m)}_{32} & \cdots & \lambda^{(m)}_{3,N_m} \\ \lambda^{(m)}_{31} & \lambda^{(m)}_{32} & \cdots & \lambda^{(m)}_{3,N_m} \\ \cdots & \cdots & \cdots & \\ \lambda^{(m)}_{N_m-1,1} & \lambda^{(m)}_{N_m-1,2} & \cdots & \lambda^{(m)}_{N_m-1,N_m} \end{bmatrix} = \begin{bmatrix} \text{LINEAR CONTRAST COEFFICIENTS} \\ \text{QUADRATIC} \quad '' \quad '' \\ \text{CUBIC} \quad '' \quad '' \\ \cdots\cdots\cdots\cdots \\ (N_m-1)\text{th} \quad '' \quad '' \end{bmatrix}$$

MATRIX METHOD FOR OBTAINING A-EFFECT IN A TWO FACTOR DESIGN

$N_A = 4$ (FOUR LEVELS OF A)

$N_B = 5$ (FIVE LEVELS OF B)

$$\underset{(K=1,2,3)}{\left[C_K(A)\right]} = \begin{bmatrix} \lambda_{11}^{(1)} & \lambda_{12}^{(1)} & \lambda_{13}^{(1)} & \lambda_{14}^{(1)} \\ \lambda_{21}^{(1)} & \lambda_{22}^{(1)} & \lambda_{23}^{(1)} & \lambda_{24}^{(1)} \\ \lambda_{31}^{(1)} & \lambda_{32}^{(1)} & \lambda_{33}^{(1)} & \lambda_{34}^{(1)} \end{bmatrix} \begin{bmatrix} S_{11} & S_{12} & S_{13} & S_{14} & S_{15} \\ S_{21} & S_{22} & S_{23} & S_{24} & S_{25} \\ S_{31} & S_{32} & S_{33} & S_{34} & S_{35} \\ S_{41} & S_{42} & S_{43} & S_{44} & S_{45} \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} \lambda_{11}^{(1)} & \lambda_{12}^{(1)} & \lambda_{13}^{(1)} & \lambda_{14}^{(1)} \\ \lambda_{21}^{(1)} & \lambda_{22}^{(1)} & \lambda_{23}^{(1)} & \lambda_{24}^{(1)} \\ \lambda_{31}^{(1)} & \lambda_{32}^{(1)} & \lambda_{33}^{(1)} & \lambda_{34}^{(1)} \end{bmatrix} \begin{bmatrix} \Sigma S_{1j} \\ \Sigma S_{2j} \\ \Sigma S_{3j} \\ \Sigma S_{4j} \end{bmatrix}$$

MATRIX METHOD FOR OBTAINING B-EFFECT IN A TWO-FACTOR DESIGN

$N_A = 4$ (FOUR LEVELS OF A)
$N_B = 5$ (FIVE LEVELS OF B)

$$\left[C_k(B)\right] = \begin{bmatrix} \lambda_{11}^{(2)} & \lambda_{12}^{(2)} & \lambda_{13}^{(2)} & \lambda_{14}^{(2)} & \lambda_{15}^{(2)} \\ \lambda_{21}^{(2)} & \lambda_{22}^{(2)} & \lambda_{23}^{(2)} & \lambda_{24}^{(2)} & \lambda_{25}^{(2)} \\ \lambda_{31}^{(2)} & \lambda_{32}^{(2)} & \lambda_{33}^{(2)} & \lambda_{34}^{(2)} & \lambda_{35}^{(2)} \\ \lambda_{41}^{(2)} & \lambda_{42}^{(2)} & \lambda_{43}^{(2)} & \lambda_{44}^{(2)} & \lambda_{45}^{(2)} \end{bmatrix} \left\{ \begin{bmatrix} 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} S_{11} & S_{12} & S_{13} & S_{14} & S_{15} \\ S_{21} & S_{22} & S_{23} & S_{24} & S_{25} \\ S_{31} & S_{32} & S_{33} & S_{34} & S_{35} \\ S_{41} & S_{42} & S_{43} & S_{44} & S_{45} \end{bmatrix} \right\}^T$$

$(k = 1, 2, 3, 4)$

$$= \begin{bmatrix} \lambda_{11}^{(2)} & \lambda_{12}^{(2)} & \lambda_{13}^{(2)} & \lambda_{14}^{(2)} & \lambda_{15}^{(2)} \\ \lambda_{21}^{(2)} & \lambda_{22}^{(2)} & \lambda_{23}^{(2)} & \lambda_{24}^{(2)} & \lambda_{25}^{(2)} \\ \lambda_{31}^{(2)} & \lambda_{32}^{(2)} & \lambda_{33}^{(2)} & \lambda_{34}^{(2)} & \lambda_{35}^{(2)} \\ \lambda_{41}^{(2)} & \lambda_{42}^{(2)} & \lambda_{43}^{(2)} & \lambda_{44}^{(2)} & \lambda_{45}^{(2)} \end{bmatrix} \begin{bmatrix} \Sigma S_{i1} \\ \Sigma S_{i2} \\ \Sigma S_{i3} \\ \Sigma S_{i4} \\ \Sigma S_{i5} \end{bmatrix}$$

THE INTERACTION CONTRAST MATRIX FOR A TWO FACTOR EXPERIMENT

$N_1 = 4, \quad N_2 = 5$

$p = 1, 2, 3, = N_1 - 1$

$q = 1, 2, 3, 4 = N_2 - 1$

$$
\left[C_{pq} \right] = \begin{bmatrix}
\Sigma\Sigma \lambda^{(1)}_{1i} \lambda^{(2)}_{1j} S_{ij} & \Sigma\Sigma \lambda^{(1)}_{1i} \lambda^{(2)}_{2j} S_{ij} & \Sigma\Sigma \lambda^{(1)}_{1i} \lambda^{(2)}_{3j} S_{ij} & \Sigma\Sigma \lambda^{(1)}_{1i} \lambda^{(2)}_{4j} S_{ij} \\
\Sigma\Sigma \lambda^{(1)}_{2i} \lambda^{(2)}_{1j} S_{ij} & \Sigma\Sigma \lambda^{(1)}_{2i} \lambda^{(2)}_{2j} S_{ij} & \Sigma\Sigma \lambda^{(1)}_{2i} \lambda^{(2)}_{3i} S_{ij} & \Sigma\Sigma \lambda^{(1)}_{2i} \lambda^{(2)}_{4j} S_{ij} \\
\Sigma\Sigma \lambda^{(1)}_{3i} \lambda^{(2)}_{1j} S_{ij} & \Sigma\Sigma \lambda^{(1)}_{3i} \lambda^{(2)}_{2j} S_{ij} & \Sigma\Sigma \lambda^{(1)}_{3i} \lambda^{(2)}_{3j} S_{ij} & \Sigma\Sigma \lambda^{(1)}_{3i} \lambda^{(2)}_{4j} S_{ij}
\end{bmatrix}
$$

WHERE $\Sigma\Sigma \equiv \sum_{i=1}^{N_1=4} \sum_{j=1}^{N_2=5}$

MATRIX METHOD FOR OBTAINING INTERACTION CONTRASTS FOR A TWO FACTOR EXPERIMENT

$N_1 = 4 \qquad N_2 = 5$

$$\underset{(K=1,2,\ldots 12)}{\left[C_K(AB)\right]} = \begin{bmatrix} \Sigma L_i^{(1)} S_{i1} & \Sigma L_i^{(1)} S_{i2} & \Sigma L_i^{(1)} S_{i3} & \Sigma L_i^{(1)} S_{i4} & \Sigma L_i^{(1)} S_{i5} \\ \Sigma Q_i^{(1)} S_{i1} & \Sigma Q_i^{(1)} S_{i2} & \Sigma Q_i^{(1)} S_{i3} & \Sigma Q_i^{(1)} S_{i4} & \Sigma Q_i^{(1)} S_{i5} \\ \Sigma C_i^{(1)} S_{i1} & \Sigma C_i^{(1)} S_{i2} & \Sigma C_i^{(1)} S_{i3} & \Sigma C_i^{(1)} S_{i4} & \Sigma C_i^{(1)} S_{i5} \end{bmatrix} \begin{bmatrix} L_1^{(2)} & Q_1^{(2)} & C_1^{(2)} & q_1^{(2)} \\ L_2^{(2)} & Q_2^{(2)} & C_2^{(2)} & q_2^{(2)} \\ L_3^{(2)} & Q_3^{(2)} & C_3^{(2)} & q_3^{(2)} \\ L_4^{(2)} & Q_4^{(2)} & C_4^{(2)} & q_4^{(2)} \\ L_5^{(2)} & Q_5^{(2)} & C_5^{(2)} & q_5^{(2)} \end{bmatrix}$$

WHERE $\Sigma \equiv \sum_{i=1}^{N_1=4}$

INTERACTION CONTRAST MATRIX FOR A TWO-FACTOR EXPERIMENT

$N_1 = 4$

$N_2 = 5$

$p = 1,2,3, \; (= N_1 - 1)$

$q = 1,2,3,4 \; (= N_2 - 1)$

$$\begin{bmatrix} C_{pq} \end{bmatrix} = \begin{bmatrix} C_{11} & C_{12} & C_{13} & C_{14} \\ C_{21} & C_{22} & C_{23} & C_{24} \\ C_{31} & C_{32} & C_{33} & C_{34} \end{bmatrix}$$

$$C_{pq} = \sum_{I=1}^{N_1=4} \sum_{J=1}^{N_2=5} \lambda^{(1)}_{pi} \lambda^{(2)}_{qj} S_{ij}$$

COMPUTATIONAL SCHEME FOR A TWO FACTOR CONTRAST

$N_1 = 4, \quad N_2 = 3$

$$
\begin{aligned}
C_{22} = A_Q\, B_Q &= \sum_{j_1=1}^{4} \sum_{j_2=1}^{3} \lambda^{(1)}_{2j_1} \lambda^{(2)}_{2j_2} S_{j_1 j_2} \\
&= \lambda^{(1)}_{21} \lambda^{(2)}_{21} S_{11} + \lambda^{(1)}_{21} \lambda^{(2)}_{22} S_{12} + \lambda^{(1)}_{21} \lambda^{(2)}_{23} S_{13} \\
&+ \lambda^{(1)}_{22} \lambda^{(2)}_{21} S_{21} + \lambda^{(1)}_{22} \lambda^{(2)}_{22} S_{22} + \lambda^{(1)}_{22} \lambda^{(2)}_{23} S_{23} \\
&+ \lambda^{(1)}_{23} \lambda^{(2)}_{21} S_{31} + \lambda^{(1)}_{23} \lambda^{(2)}_{22} S_{32} + \lambda^{(1)}_{23} \lambda^{(2)}_{23} S_{33} \\
&+ \lambda^{(1)}_{24} \lambda^{(2)}_{21} S_{41} + \lambda^{(1)}_{24} \lambda^{(2)}_{22} S_{42} + \lambda^{(1)}_{24} \lambda^{(2)}_{23} S_{43}
\end{aligned}
$$

A COMPLETE CONTRAST MATRIX FOR A 5 X 4 X 3 DESIGN
(INCLUDING THE SUM TOTAL C_{000})

$$\left[C_{i_1 i_2 i_3} \right] \quad \text{WHERE} \begin{cases} i_1 = 0,1,2,3,4 \\ i_2 = 0,1,2,3 \\ i_3 = 0,1,2 \end{cases}$$

$$= \begin{bmatrix} 000 & 001 & 002 & 010 & 011 & 012 & 020 & 021 & 022 & 030 & 031 & 032 \\ 100 & 101 & 102 & 110 & 111 & 112 & 120 & 121 & 122 & 130 & 131 & 132 \\ 200 & 201 & 202 & 210 & 211 & 212 & 220 & 221 & 222 & 230 & 231 & 232 \\ 300 & 301 & 302 & 310 & 311 & 312 & 320 & 321 & 322 & 330 & 331 & 332 \\ 400 & 401 & 402 & 410 & 411 & 412 & 420 & 421 & 422 & 430 & 431 & 432 \end{bmatrix}$$

ABOVE MATRIX CONTAINS 5 X 4 X 3 = 60 TRIPLETS, THESE TRIPLETS REPRESENT THE CONTRASTS; e.q., 312 $\equiv C_{312}$. C_{000} IS NOT A CONTRAST.

COMPUTATIONAL SCHEME FOR A THREE FACTOR CONTRAST

$$N_1 = 4, \quad N_2 = 3 \quad N_3 = 2$$

$$C_{121} = (A_L B_Q C_L) = \sum_{p,q,r,}^{4,3,2} = \lambda^{(1)}_{1p} \; \lambda^{(2)}_{2q} \; \lambda^{(3)}_{1r} \; S_{pqr}$$

$$= \lambda^{(1)}_{11} \lambda^{(2)}_{21} \lambda^{(3)}_{11} S_{111} + \lambda^{(1)}_{11} \lambda^{(2)}_{21} \lambda^{(3)}_{12} S_{112} + \cdots + \lambda^{(1)}_{11} \lambda^{(2)}_{23} \lambda^{(3)}_{12} S_{132}$$

$$+ \lambda^{(1)}_{12} \lambda^{(2)}_{21} \lambda^{(3)}_{11} S_{211} + \lambda^{(1)}_{12} \lambda^{(2)}_{21} \lambda^{(3)}_{12} S_{212} + \cdots + \lambda^{(1)}_{12} \lambda^{(2)}_{23} \lambda^{(3)}_{12} S_{232}$$

$$+ \lambda^{(1)}_{13} \lambda^{(2)}_{21} \lambda^{(3)}_{11} S_{311} + \lambda^{(1)}_{13} \lambda^{(2)}_{21} \lambda^{(3)}_{12} S_{312} + \cdots + \lambda^{(1)}_{13} \lambda^{(2)}_{23} \lambda^{(3)}_{12} S_{332}$$

$$+ \lambda^{(1)}_{14} \lambda^{(2)}_{21} \lambda^{(3)}_{11} S_{411} + \lambda^{(1)}_{14} \lambda^{(2)}_{21} \lambda^{(3)}_{12} S_{412} + \cdots + \lambda^{(1)}_{14} \lambda^{(2)}_{23} \lambda^{(3)}_{12} S_{432}$$

NOTE: C_{121} CONTAINS $4 \times 3 \times 2 = 24$ TERMS.

COMPLETE CONTRAST MATRIX FOR AN $N_1 \times N_2 \times N_3$ DESIGN

$$\left[C_{i_1 i_2 i_3} \right] \qquad \text{WHERE} \begin{cases} i_1 = 0, 1, \ldots N_1 - 1 \\ i_2 = 0, 1, \ldots N_2 - 1 \\ i_3 = 0, 1, \ldots N_3 - 1 \end{cases}$$

$$= \begin{bmatrix} 000 & 010 \ldots 0,N_2-1,0 & 001 \ldots 00,N_3-1 & 001 \ldots 01,N_3-1 \ldots & 0,N_2-1,N_3-1 \\ 100 & 110 \ldots 1,N_2-1,0 & 101 \ldots 10,N_3-1 & 111 \ldots 11,N_3-1 \ldots & 1,N_2-1,N_3-1 \\ 200 & 210 \ldots 2,N_2-1,0 & 201 \ldots 20,N_3-1 & 211 \ldots 21,N_3-1 \ldots & 2,N_2-1,N_3-1 \\ - - - & - - - & - - - & - - - & - - - \\ N_1-1,00 & N_1-1,10 \ldots N_1-1,N_2-1,0 & N_1-1,01 \ldots N_1-1,0N_3-1 & N_1 1,1 N_3-1 \ldots & N_1-1,N_2-1,N_3-1 \end{bmatrix}$$

ABOVE MATRIX CONTAINS $N_1 N_2 N_3$ TRIPLETS, THESE TRIPLETS REPRESENT THE CONTRASTS e.g. $211 \equiv C_{211}$. C_{000} IS NOT A CONTRAST.

INTERACTION CONTRASTS OF ORDER LESS THAN THE HIGHEST

$$C_{i_1 0 i_3 0 0 i_6 \ldots 0} = \sum_{j_1 j_2 \ldots j_m}^{N_1, N_2, \ldots, N_m} \lambda^{(1)}_{i_1 j_1} \lambda^{(2)}_{0 j_2} \lambda^{(3)}_{i_3 j_3} \lambda^{(4)}_{0 j_4} \lambda^{(5)}_{0 j_5} \lambda^{(6)}_{i_6 j_6} \cdots \lambda^{(m)}_{0 j_m} S_{j_1 j_2 \ldots j_m}$$

$$= \sum_{j_1 j_2 \ldots j_m}^{N_1, N_2, \ldots, N_m} \lambda^{(1)}_{i_1 j_1} \lambda^{(3)}_{i_3 j_3} \lambda^{(6)}_{i_6 j_6} S_{j_1 j_2 \ldots j_m},$$

SINCE $\lambda^{(k)}_{0 j_k} = 1$ FOR ALL J_k AND $1 \leq k \leq m$

MAIN EFFECTS TOTALS

$$C_{i,00\ldots0} = \sum_{j_1, j_2, \ldots, j_m}^{N_1, N_2, \ldots, N_m} \lambda^{(1)}_{ij_1} \lambda^{(2)}_{0j_2} \cdots \lambda^{(m)}_{0j_m} S_{j_1 j_2 \ldots j_m}$$

$$= \sum_{j_1, j_2, \ldots, j_m}^{N_1, N_2, \ldots N_m} \lambda^{(1)}_{ij_1} S_{j_1 j_2 \cdots j_m}, \quad \text{SINCE } \lambda^{(k)}_{0 j_k} = 1 \text{ FOR ALL } j_k;\ 1 \leq K \leq m$$

= MAIN EFFECT (TOTAL) OF FACTOR F_1

SIMILARLY FOR MAIN EFFECTS (TOTALS) OF FACTORS $F_2, F_3, \ldots, F_m$

NOTE: $i = 1, 2, 3, \ldots, N_1 - 1$ YIELDS THE LINEAR, QUADRATIC, CUBIC, ..., $(N_1 - 1)$ TH CONTRAST COMPONENT RESPECTIVELY OF THE MAIN EFFECT TOTAL FACTOR F_1.

MAIN EFFECTS TOTALS OF THE K-TH FACTOR F_k

$$C_{00\cdots 0 i_k \cdots 0} = \sum_{J_1 J_2 \cdots J_m}^{N_1, N_2 \cdots, N_m} \lambda^{(1)}_{0J_1} \lambda^{(2)}_{0J_2} \cdots \lambda^{(k-1)}_{0J_{k-1}} \lambda^{(k)}_{i_k j_k} \lambda^{(k+1)}_{0j_{k+1}} \cdots \lambda^{(m)}_{0j_m} S_{J_1 J_2 \cdots J_m}$$

$$= \sum_{J_1 J_1 \cdots J_m}^{N_1, N_2, \cdots, N_m} \lambda^{(k)}_{i_k J_k} S_{J_1 J_2 \cdots J_m}$$

TOTAL SUM OF OBSERVATIONS

$$C_{00\ldots 0} = \sum_{J_1 J_2 \cdots J_m}^{N_1, N_2, \ldots, N_m} \lambda^{(1)}_{0J_1} \lambda^{(2)}_{0J_2} \cdots \lambda^{(m)}_{0J_m} S_{J_1 J_2 \cdots J_m} \quad \text{WITH NO REPLICATIONS}$$

$$= \sum_{J_1 J_2 \cdots J_m}^{N_1 N_2 \cdots N_m} S_{j_1 j_2 \cdots j_m} \equiv T$$

$C_{00\ldots 0} = rT = T'$, WITH r REPLICATIONS PER TREATMENT

$C_{00\ldots 0}$ IS <u>NOT</u> A CONTRAST

SUM OF SQUARES (SS) OF $C_{i_1 i_2 \cdots i_m}$ WITH REPLICATIONS. INTERACTION OF FACTORS $F_1, F_2, \ldots, F_m$

$$SS(C_{i_1 i_2 \ldots i_m}) = \frac{1}{r}\left[\sum_{j_1 j_2 \cdots j_m}^{N_1, N_2, \ldots, N_3} \Lambda^{(1)}_{i_1 j_1} \Lambda^{(2)}_{i_2 j_2} \cdots \Lambda^{(m)}_{i_m j_m} S_{j_1 j_2 \cdots j_m}\right]^2$$

WHERE $$\Lambda^{(k)}_{i_k j_k} \equiv \frac{\lambda^{(K)}_{i_k j_k}}{\sqrt{\sum_{j_k=1}^{N_k}\left[\lambda^{(k)}_{i_k j_k}\right]^2}}; i_k \neq 0$$

r = NUMBER OF REPLICATIONS PER CELL (TREATMENT)

SUM OF SQUARES (SS) WITH REPLICATION
SPECIAL CASES

A. INTERACTION OF FACTORS F_1 AND F_3

$$SS(C_{i_1 0 i_3 0 \ldots 0}) = \frac{N_1 N_3}{r \prod_{k=1}^{m} N_k} \left[\sum_{j_1 j_2 \ldots j_m}^{N_1, N_2 \ldots, N_m} \Lambda^{(1)}_{i_1 j_1} \lambda^{(2)}_{0 j_2} \Lambda^{(3)}_{i_3 j_3} \lambda^{(4)}_{0 j_4} \cdots \lambda^{(m)}_{0 j_m} S_{j_1 j_2 \ldots j_m} \right]^2$$

$$= \frac{N_1 N_3}{r \prod_{k=1}^{m} N_k} \left[\sum_{j_1 j_2 \ldots j_m}^{N_1, N_2 \ldots, N_m} \Lambda^{(1)}_{i_1 j_1} \Lambda^{(3)}_{i_3 j_3} S_{j_1 j_2 \ldots j_m} \right]^2$$

SINCE $\lambda^{(k)}_{0 j_k} = 1$, $1 \leq K \leq m$

B. INTERACTION OF FACTORS F_2 F_5 F_8

$$SS(C_{0 i_2 0 0 i_5 0 0 i_8 0 \ldots 0}) = \frac{N_2 N_5 N_8}{r \prod_{k=1}^{m} N_k} \left[\sum_{j_1 j_2 \ldots j_m}^{N_1, N_2 \ldots, N_m} \Lambda^{(2)}_{i_2 j_2} \Lambda^{(5)}_{i_5 j_5} \Lambda^{(8)}_{i_8 j_8} S_{j_1 j_2 \ldots j_m} \right]^2$$

FORMULAS FOR GENERATING THE ELEMENTS OF AN ORTHOGONAL POLYNOMIAL CONTRAST MATRIX

(1) $C_{i_1 i_2 \ldots i_m}$

$$= \sum_{j_1=1}^{N_1} \sum_{j_2=1}^{N_2} \cdots \sum_{j_m=1}^{N_m} \lambda^{(1)}_{i_1 j_1} \lambda^{(2)}_{i_2 j_2} \cdots \lambda^{(m)}_{i_m j_m} S_{j_1 j_2 \ldots j_m}$$

$$\equiv \sum_{j_1 j_2 \ldots j_m}^{N_1 N_2 \ldots, N_m} \lambda^{(1)}_{i_1 j_1} \lambda^{(2)}_{i_2 j_2} \cdots \lambda^{(m)}_{i_m j_m} S_{j_1 j_2 \ldots j_m}$$

where we have written

$$\sum_{j_1}^{N_1} \sum_{j_2}^{N_2} \cdots \sum_{j_m}^{N_m} \equiv \sum_{j_1, j_2, \ldots, j_m}^{N_1 N_2 \ldots, N_m}$$

The single summation symbol will be employed in the formulas which follow.

(2) Values of the Subscripts

$i_1 = 0, 1, \ldots, N_1 - 1$ $\quad$ $j_1 = 1, \ldots, N_1$

$i_2 = 0, 1, \ldots, N_2 - 1$ $\quad$ $j_2 = 1, \ldots, N_2$

. . . . $\quad$ . . . .

$i_m = 0, 1, 2, \ldots, N_m - 1$ $\quad$ $j_m = 1, \ldots, N_m$

(3) Levels of the Factor F_m

N_1 = number of levels of F_1

N_2 = number of levels of F_2

. . . .

N_m = number of levels of F_m

(4) Factors. $F_1, F_2, \ldots, F_m$

(5) $C_{i_1 i_2 \ldots i_m}$ represents the general element of the orthogonal contrast (interaction) matrix of the factors F_1, $F_2, \ldots, F_m$.

(6) Index Notation

$i_m j_m$ refers to the element in the i^{th} row and j^{th} column of the orthogonal polynomial coefficient matrix for the m^{th} factor F_m.

$i_m = 1, 2, 3, \ldots, N_m - 1$ yields the linear, quadratic, cubic, ..., $(N_m - 1)$ contrasts respectively.

$j_m = 1, 2, 3, \ldots, N_m$ refers to the first, second, third, ..., N_mth levels for the factors $F_1, F_2, F_3, \ldots, F_m$ respectively.

$\lambda^{(m)}_{i_m j_m}$ represent row vectors of the transpose matrix of the Fisher orthogonal polynomial coefficients of factor F_m with the restriction that $i_m \neq 0$. For example, let us write down the matrix for $\lambda^{(2)}_{i_2 j_2}$:

$\Lambda^{(m)}_{i_m j_m}$ represent the normalized form of $\lambda^{(m)}_{i_m j_m}$

$$\left[\lambda^{(2)}_{i_2 j_2}\right] = \begin{bmatrix} \lambda^{(2)}_{11} & \lambda^{(2)}_{12} & \lambda^{(2)}_{13} & \lambda^{(2)}_{14} \\ \lambda^{(2)}_{21} & \lambda^{(2)}_{22} & \lambda^{(2)}_{23} & \lambda^{(2)}_{24} \\ \lambda^{(2)}_{31} & \lambda^{(2)}_{32} & \lambda^{(2)}_{33} & \lambda^{(2)}_{34} \end{bmatrix} = \begin{bmatrix} \text{Linear contrast coefficients} \\ \text{Quadratic contrast coefficients} \\ \text{Cubic contrasts coefficients} \end{bmatrix},$$

where

$$\begin{cases} j_2 = 1, 2, 3, 4 \\ i_2 = 1, 2, 3. \end{cases}$$

Note that the first row of the matrix gives the linear contrast coefficients, the second row gives the quadratic contrast coefficients, and the third row gives the cubic contrast coefficients. Note also the $\lambda^{(2)}$ - matrix refers to factor F_2 possessing 4 levels. To summarize, $\lambda^{(2)}_{i_2 j_2}$ represents a 3×4 matrix for factor F_2 of 4 levels with 3 contrasts: linear, quadratic, cubic. The extension to $\lambda^{(m)}_{i_m j_m}$ is obvious.

(7) <u>The Special Case</u> $\lambda^{(m)}_{0j_m}$. Define

$\lambda^{(m)}_{0j_m}$ = 1 for all m's and j_m's . This symbol is used in a two-fold way: (a) to obtain the total sum of the observations; (b) to obtain interactions of lower than the highest order. It should be noted that the row vector consisting of 1's is <u>not</u> orthogonal to the row vectors of the orthogonal coefficient matrix $\left[\lambda^{(m)}_{i_m j_m}\right]$ where $i_m \neq 0$. The row vector $\left[\lambda^{(m)}_{0j_m}\right]$ will <u>not</u> be subject to normalization.

(8) <u>Interaction Contrasts of Order Less Than the Highest</u>

a. $C_{i_1 0 i_3 0 0 i_6 0 \ldots 0}$

$$= \sum_{j_1 j_2 \ldots j_m}^{N_1, N_2, \ldots, N_m} \lambda^{(1)}_{i_1 j_1} \lambda^{(2)}_{0 j_2} \lambda^{(3)}_{i_3 j_3} \lambda^{(4)}_{0 j_4} \lambda^{(5)}_{0 j_5} \lambda^{(6)}_{i_6 j_6} \lambda^{(7)}_{0 j_7} \cdots \lambda^{(m)}_{0 j_m} S_{j_1 j_2 \ldots j_m}$$

$$= \sum_{j_1 j_2 \ldots j_m}^{N_1, N_2, \ldots, N_m} \lambda^{(1)}_{i_1 j_1} \lambda^{(3)}_{i_3 j_3} \lambda^{(6)}_{i_6 j_6} S_{j_1 j_2 \ldots j_m}, \text{ since } \lambda^{(k)}_{0 j_k} = 1 \text{ for } 1 \leq k \leq$$

b. Main Effects Total of the $k^{\underline{th}}$ Factor F_k

$$C_{0\,0\,\ldots\,i_k\,0\,\ldots\,0}$$

$$= \sum_{j_1 j_2 \cdots j_m}^{N_1, N_2, \ldots, N_m} \lambda^{(1)}_{0j_1} \lambda^{(2)}_{0j_2} \cdots \lambda^{(k)}_{i_k j_k} \lambda^{(k+1)}_{0j_{k+1}} \cdots \lambda^{(m)}_{0j_m} S_{j_1 j_2 \cdots j_m}$$

$$= \sum_{j_1 j_2 \cdots j_m}^{N_1, N_2, \ldots, N_m} \lambda^{(1)}_{i_k j_k} S_{j_1 j_2 \cdots j_k}, \text{ since } \lambda^{(k)}_{0j_k} = 1 \text{ for } 1 \leq k \leq m .$$

(9) The Special Case $C_{00\ldots0}$. Total Sum of Observations $S_{j_1 j_2 \cdots j_m}$

$$C_{00\ldots0} = \sum_{j_1 j_2 \cdots j_m}^{N_1, N_2, \ldots, N_m} \lambda^{(1)}_{0j_1} \lambda^{(2)}_{0j_2} \cdots \lambda^{(m)}_{0j_m} S_{j_1 j_2 \cdots j_m}$$

$$= \sum_{j_1 j_2 \cdots j_m}^{N_1, N_2, \ldots, N_m} S_{j_1 j_2 \cdots j_m}, \text{ since } \lambda^{(k)}_{0j_k} = 1,\ 1 \leq k \leq m$$

$= T$ (Total sum of observations with one unit per cell).

Now assume there are r replications per cell (treatment). Then

$$C_{00\ldots0} = rT = \sum_{k=1}^{r} T_k = T' .$$

$$\text{The G. A. (Grand Average)} = \frac{rT}{rN_1N_2\ldots N_m} = \frac{T'}{r\prod_{k=1}^{m} N_k} ,$$

$$\text{Without replicates, G. A.} = \frac{C_{00\ldots0}}{\prod_{k=1}^{m} N_k}$$

$$= \frac{T}{\prod_{k=1}^{m} N_k} ,$$

(10) Sum of Squares (SS) of $C_{i_1 i_2 \ldots i_m}$ with Replication.

Interaction of Factors $F_1, F_2, \ldots, F_m$.

$$SS\,(C_{i_1 i_2 \ldots i_m})$$

$$= \frac{1}{r}\left[\sum_{j_1 j_2 \ldots j_m}^{N_1, N_2, \ldots, N_m} \frac{\lambda^{(1)}_{i_1 j_1}}{\sqrt{\sum_{j_1=1}^{N_1}\left[\lambda^{(1)}_{i_1 j_1}\right]^2}} \cdot \frac{\lambda^{(1)}_{i_2 j_2}}{\sqrt{\sum_{j_2=1}^{N_2}\left[\lambda^{(2)}_{i_2 j_2}\right]^2}} \cdots \frac{\lambda^{(m)}_{i_m j_m}}{\sqrt{\sum_{j_m=1}^{N_m}\left[\lambda^{(m)}_{i_m j_m}\right]^2}}\, S_{j_1 j_2 \ldots j_m}\right]^2$$

$$= \frac{1}{r}\left[\sum_{j_1 j_2 \ldots j_m}^{N_1, N_2, \ldots, N_m} \Lambda^{(1)}_{i_1 j_1} \cdot \Lambda^{(2)}_{i_2 j_2} \cdots \Lambda^{(m)}_{i_m j_m}\, S_{j_1 j_2 \ldots j_m}\right]^2$$

$$= \frac{1}{r}\left[\sum_{j_1 j_2 \ldots j_m}^{N_1, N_2, \ldots, N_m} \prod_{k=1}^{m} \Lambda^{(k)}_{i_k j_k}\, S_{j_1 j_2 \ldots j_m}\,.\right]^2.$$

It should be pointed out that the orthogonal contrast coeffieients $\lambda^{(k)}_{i_k j_k}$ (with the restriction that $i_k \neq 0$) are normalized as follows:

$$\Lambda^{(k)}_{i_k j_k} \equiv \frac{\lambda^{(k)}_{i_k j_k}}{\sqrt{\sum_{j_k=1}^{N_k} \left[\lambda^{(k)}_{i_k j_k}\right]^2}} .$$

(ii) Special Case. Sum of Squares (SS) of $C_{i_1 0 i_3 00\ldots 0}$, with Replication, Interaction of First and Third Factors F_1 and F_3.

$$SS(C_{i_1 0 i_3 0\ldots 0})$$

$$= \frac{N_1 N_3}{r \prod_{k=1}^{m} N_k} \left[\sum_{j_1 j_2 \ldots j_m}^{N_1, N_2, \ldots, N_m} \frac{\lambda^{(1)}_{i_1 j_1}}{\sqrt{\sum_{j_1=1}^{N_1} \left[\lambda^{(1)}_{i_1 j_1}\right]^2}} \cdot \lambda^{(2)}_{0 j_2} \cdot \frac{\lambda^{(3)}_{i_3 j_3}}{\sqrt{\sum_{j_3=1}^{N_3} \left[\lambda^{(3)}_{i_3 j_3}\right]^2}} \cdot \lambda^{(4)}_{0 j_4} \ldots \lambda^{(m)}_{0 j_m} S_{j_1 j_2 \ldots j_m} \right]^2$$

$$= \frac{N_1 N_3}{r \prod_{k=1}^{m} N_k} \left[\sum_{j_1, j_2, \ldots, j_m}^{N_1, N_2, \ldots, N_m} \frac{\lambda^{(1)}_{i_1 j_1}}{\sqrt{\sum_{j_1=1}^{N_1} \left[\lambda^{(1)}_{i_1 j_1}\right]^2}} \cdot \frac{\lambda^{(3)}_{i_3 j_3}}{\sqrt{\sum_{j_3=1}^{N_3} \left[\lambda^{(3)}_{i_3 j_3}\right]^2}} S_{j_1 j_2 \ldots j_m} \right]^2 ,$$

since $\lambda^{(k)}_{0j_k} = 1, \quad 1 \leq k \leq m$

$$SS(C_{i_1 0 i_3 0 \ldots 0}) = \frac{N_1 N_3}{r \prod_{k=1}^{m} N_k} = \left[\sum_{j_1 j_2 \ldots j_m}^{N_1, N_2, \ldots, N_m} \Lambda^{(1)}_{i_1 j_1} \Lambda^{(3)}_{i_3 j_3} \; S_{j_1 j_2 \ldots j_m} \right]^2$$

As pointed out in Paragraph 7, it should be observed that the $\lambda^{(m)}_{0j_m}$ are not normalized. When $i_m \neq 0$, the $\lambda^{(m)}_{i_m j_m}$ are normalized thus:

$$\Lambda^{(m)}_{i_m j_m} = \frac{\lambda^{(m)}_{i_m j_m}}{\sqrt{\sum_{j_m=1}^{N_m} \left[\lambda^{(m)}_{i_m j_m} \right]^2}} .$$

g. The Application of the Algorithm to a 5 x 4 x 3 x 2 Design. What follows is the application of the general formulas and the index-positional algorithm to the 5 x 4 x 3 x 2 design given in the data-matrix shown in Chart VIII. There are 120 (5 x 4 x 3 x 2) treatments (without replications).

Chart VI exhibits the index-presentation of this data-matrix.

The matrices (of orthogonal polynomial values) associated with each factor are given in Chart I.

The orthogonal coefficient-multipliers corresponding to each element of the data-matrix are computed for the effect total of the contrast

C_{3121} $(\equiv A_C B_L C_Q D_L)$. These coefficients are given in Chart II. (The complete set is given in Chart IV.) Those for the effect totals of the contrast C_{3020} $(\equiv A_C C_Q)$ are given in Chart III. The complete set is given in Chart V.

The effect totals of some of the contrasts of this $5 \times 4 \times 3 \times 2$ design are given in Tables 5, 6, and 7.

FORMULAS AND ORTHOGONAL CONTRAST VECTORS FOR TWO SPECIFIC EFFECTS OF A 5 X 4 X 3 X 2 DESIGN

GENERAL:

$$C_{i_1 i_2 i_3 i_4} = \sum_{J_1}^{N_1} \sum_{J_2}^{N_2} \sum_{J_3}^{N_3} \sum_{J_4}^{N_4} \lambda^{(1)}_{i_1 j_1} \lambda^{(2)}_{i_2 j_2} \lambda^{(3)}_{i_3 j_3} \lambda^{(4)}_{i_4 j_4} S_{j_1 j_2 j_3 j_4}$$

A FOUR FACTOR INTERACTION:

$$C_{3121} = \sum_{J_1=1}^{N_1} \sum_{J_2=1}^{N_2} \sum_{J_3=1}^{N_3} \sum_{J_4=1}^{N_4} \lambda^{(1)}_{3J_1} \lambda^{(2)}_{1J_2} \lambda^{(3)}_{2J_3} \lambda^{(4)}_{1J_4} S_{J_1 J_2 J_3 J_4}$$

A TWO FACTOR INTERACTION:

$$C_{3020} = \sum_{J_1=1}^{N_1} \sum_{J_2=1}^{N_2} \sum_{J_3=1}^{N_3} \sum_{J_4=1}^{N_4} \lambda^{(1)}_{3J_1} \,(1)\, \lambda^{(3)}_{2J_3} \,(1)\, S_{J_1 J_2 J_3 J_4}$$

WHERE THE λ-VECTORS ARE:

$$\left[\lambda^{(1)}_{3\,j_1}\right] = [-1 \;\; +2 \;\; 0 \;\; -2 \;\; +1] \qquad \left[\lambda^{(3)}_{2\,j_3}\right] = [+1 \;\; -2 \;\; +1]$$

$$\left[\lambda^{(2)}_{1\,j_1}\right] = [-3 \;\; -1 \;\; +1 \;\; +3] \qquad \left[\lambda^{(4)}_{1\,j_4}\right] = [-1 \;\; +1]$$

$N_1 = 5$; $N_2 = 4$; $N_3 = 3$; $N_4 = 2$; (e.g. $J_1 = 1, 2, 3, 4, 5,$)

CHART I

MATRICES OF ORTHOGONAL POLYNOMIAL VALUES FOR A 5X4X3X2 FACTORIAL DESIGN

FACTOR	NR OF LEVELS	TYPE OF CONTRAST	POLYNOMIAL VALUES	CORRESPONDING SYMBOLS	GENERAL SYMBOL
F_1 (≡A)	5	LINEAR QUADRATIC CUBIC QUARTIC	$\begin{bmatrix} -2 & -1 & 0 & +1 & +2 \\ +2 & -1 & -2 & -1 & +2 \\ -1 & +2 & 0 & -2 & +1 \\ +1 & -4 & +6 & -4 & +1 \end{bmatrix}$	$\begin{matrix} \lambda^{(1)}_{11} & \cdots & \lambda^{(1)}_{15} \\ \cdot & \cdots & \cdot \\ \cdot & \cdots & \cdot \\ \lambda^{(1)}_{41} & \cdots & \lambda^{(1)}_{45} \end{matrix}$	$\lambda^{(1)}_{i_1 j_1}$
F_2 (≡B)	4	LINEAR QUADRATIC CUBIC	$\begin{bmatrix} -3 & -1 & +1 & +3 \\ +1 & -1 & -1 & +1 \\ +1 & -3 & -3 & +1 \end{bmatrix}$	$\begin{matrix} \lambda^{(2)}_{11} & \cdot\ \cdot & \lambda^{(2)}_{14} \\ \cdot & \cdot\ \cdot & \cdot \\ \lambda^{(2)}_{31} & \cdot\ \cdot & \lambda^{(2)}_{34} \end{matrix}$	$\lambda^{(2)}_{i_2 j_2}$
F_3 (≡C)	3	LINEAR QUADRATIC	$\begin{bmatrix} -1 & 0 & +1 \\ +1 & -2 & +1 \end{bmatrix}$	$\begin{matrix} \lambda^{(3)}_{11} & \cdot & \lambda^{(3)}_{13} \\ \lambda^{(3)}_{21} & \cdot & \lambda^{(3)}_{23} \end{matrix}$	$\lambda^{(3)}_{i_3 j_3}$
F_4 (≡D)	2	LINEAR	$\begin{bmatrix} -1 & +1 \end{bmatrix}$	$\lambda^{(4)}_{11} \quad \lambda^{(4)}_{12}$	$\lambda^{(4)}_{i_4 j_4}$

GENERAL FORMULA FOR FOUR-FACTOR ORTHOGONAL CONTRASTS

$$C_{i_1 i_2 i_3 i_4} = \sum_{j_1=1}^{N_1} \sum_{j_2=1}^{N_2} \sum_{j_3=1}^{N_3} \sum_{j_4=1}^{N_4} \lambda^{(1)}_{i_1 j_1} \lambda^{(2)}_{i_2 j_2} \lambda^{(3)}_{i_3 j_3} \lambda^{(4)}_{i_4 j_4} S_{j_1 j_2 j_3 j_4}$$

CHART 11

GENERATION OF ELEMENTS OF ORTHOGONAL CONTRAST MATRIX BY A GENERAL FORMULA (EFFECT TOTAL C_{3121})

		C_1		C_2	C_3
		D_1	D_2	D_1	D_2
A_1	B_1	$(-1)(-3)(+1)(-1)=-3$ S_{1111}	$(-1)(-3)(+1)(+1)=+3$ S_{1112}	$(-1)(-3)(-2)(-1)= +6$ S_{1121}	$(-1)(-3)((1)(1)= +3$ S_{1132}
	B_2	$(-1)(-1)(+1)(-1)= -1$ S_{1211}	$(-1)(-1)(+1)(+1)= +1$ S_{1212}	$(-1)(-1)(-2)(-1)= +2$ S_{1221}	$(-1)(-1)(1)(1)= +1$ S_{1232}
	B_3	$(-1)(+1)(+1)(-1)= +1$ S_{1311}	$(-1)(+1)(+1)(+1)= -1$ S_{1312}	$(-1)(+1)(-2)(-1)= -2$ S_{1321}	$(-1)(+1)(1)(1)= -1$ S_{1332}
	B_4	$(-1)(+3)(+1)(-1)= +3$ S_{1411}	$(-1)(+3)(+1)(+1)= -3$ S_{1412}	$(-1)(+3)(-2)(-1)= -6$ S_{1421}	$(-1)(+3)(1)(1)= -3$ S_{1432}
A_2	B_1	$(+2)(-3)(+1)(-1)= +6$ S_{2111}	$(+2)(-3)(+1)(+1)= -6$ S_{2112}	$(+2)(-3)(-2)(-1)= -12$ S_{2121}	$(+2)(-3)(1)(1)= -6$ S_{2132}
	B_2	$(+2)(-1)(+1)(-1)= +2$ S_{2211}	$(+2)(-1)(+1)(+1)= -2$ S_{2212}	$(+2)(-1)(-2)(-1)= -4$ S_{2221}	$(+2)(-1)(+)(-1)= -2$ S_{2232}
A_5	B_3	$(+1)(+1)(+1)(-1)= -1$ S_{5311}	$(1)(1)(1)(1)= 1$ S_{5312}	$(1)(1)(-2)(-1)= +2$ S_{5321}	$(1)(1)(1)(1)= 1$ S_{5332}
	B_4	$(+1)(+3)(+1)(-1)= -3$ S_{5411}	$(+1)(+3)(+1)(+1)= +3$ S_{5412}	$(1)(3)(-2)(-1)= +6$ S_{5421}	$(1)(3)(1)(1)= 3$ S_{5432}

CHART III

GENERATION OF ELEMENTS OF ORTHOGONAL CONTRAST MATRIX BY A GENERAL FORMULA (EFFECT TOTAL C_{3020})

		C_1			C_3
		D_1	D_2		D_2
A_1	B_1	(-1)1(+1)1= -1 S_{1111}	(-1)1(+1)1= -1 S_{1112}		(-1)1(+1)1= -1 S_{1132}
	B_2	(-1)1(+1)1= -1 S_{1211}	(-1)1(+1)1= -1 S_{1212}		(-1)1(+1)1= -1 S_{1232}
	B_3	(-1)1(+1)1= -1 S_{1311}	(-1)1(+1)1= -1 S_{1312}		(-1)1(+1)1= -1 S_{1332}
	B_4	(-1)1(+1)1= -1 S_{1411}	(-1)1(+1)1= -1 S_{1412}		(-1)1(+1)1= -1 S_{1432}
A_2	B_1	(2)1(1)1= +2 S_{2111}	(2)1(1)1= 2 S_{2112}		(+2)1(+1)= 2 S_{2132}
	B_2	(2)1(1)1= 2 S_{2211}	(2)1(1)1= 2 S_{2212}		(+2)1(+1)= 2 S_{2232}
A_5	B_3	(+1)1(+1)= 1 S_{5311}	(+1)1(+1)1= 1 S_{5312}		(+1)1(+1)1= 1 S_{5332}
	B_4	(+1)1(+1)1= 1 S_{5411}	(+1)1(+1)1= 1 S_{5412}		(+1)1(+1)1= 1 S_{5432}

CHART IV

COMPLETE SET OF ORTHOGONAL COEFFICIENTS CORRESPONDING UNIQUELY TO ELEMENTS OF DATA MATRIX (EFFECT TOTAL C_{3121})

		C_1		C_2		C_3	
		D_1	D_2	D_1	D_2	D_1	D_2
A_1	B_1	-3	+3	+6	-6	-3	+3
	B_2	-1	+1	+2	-2	-1	+1
	B_3	+1	-1	-2	+2	+1	-1
	B_4	+3	-3	-6	+6	+3	-3
A_2	B_1	+6	-6	-12	+12	+6	-6
	B_2	+2	-2	-4	+4	+2	-2
	B_3	-2	+2	+4	-4	-2	+2
	B_4	-6	+6	+12	-12	-6	+6
A_3	B_1	0	0	0	0	0	0
	B_2	0	0	0	0	0	0
	B_3	0	0	0	0	0	0
	B_4	0	0	0	0	0	0
A_4	B_1	-6	+6	+12	-12	-6	+6
	B_2	-2	+2	+4	-4	-2	+2
	B_3	+2	-2	-4	+4	+2	-2
	B_4	+6	-6	-12	+12	+6	-6
A_5	B_1	+3	-3	-6	+6	+3	-3
	B_2	+1	-1	-2	+2	+1	-1
	B_3	-1	+1	+2	-2	-1	+1
	B_4	-3	+3	+6	-6	-3	+3

CHART V

COMPLETE SET OF ORTHOGONAL COEFFICIENTS
CORRESPONDING UNIQUELY TO ELEMENTS OF DATA MATRIX
(EFFECT TOTAL C_{3020})

		C_1		C_2		C_3	
		D_1	D_2	D_1	D_2	D_1	D_2
A_1	B_1	-1	-1	+2	+2	-1	-1
	B_2	-1	-1	+2	+2	-1	-1
	B_3	-1	-1	+2	+2	-1	-1
	B_4	-1	-1	+2	+2	-1	-1
A_2	B_1	+2	+2	-4	-4	+2	+2
	B_2	+2	+2	-4	-4	+2	+2
	B_3	+2	+2	-4	-4	+2	+2
	B_4	+2	+2	-4	-4	+2	+2
A_3	B_1	0	0	0	0	0	0
	B_2	0	0	0	0	0	0
	B_3	0	0	0	0	0	0
	B_4	0	0	0	0	0	0
A_4	B_1	-2	-2	+4	+4	-2	-2
	B_2	-2	-2	+4	+4	-2	-2
	B_3	-2	-2	+4	+4	-2	-2
	B_4	-2	-2	+4	+4	-2	-2
A_5	B_1	+1	+1	-2	-2	+1	+1
	B_2	+1	+1	-2	-2	+1	+1
	B_3	+1	+1	-2	-2	+1	+1
	B_4	+1	+1	-2	-2	+1	+1

CHART VI

COMPLETE MATRIX OF INDICES
CORRESPONDING UNIQUELY TO ELEMENTS
OF DATA MATRIX

		C_1		C_2		C_3	
		D_1	D_2	D_1	D_2	D_1	D_2
A_1	B_1	1111	1112	1121	1122	1131	1132
	B_2	1211	1212	1221	1222	1231	1232
	B_3	1311	1312	1321	1322	1331	1332
	B_4	1411	1412	1421	1422	1431	1432
A_2	B_1	2111	2112	2121	2122	2131	2132
	B_2	2211	2212	2221	2222	2231	2232
	B_3	2311	2312	2321	2322	2331	2332
	B_4	2411	2412	2421	2422	2431	2432
A_3	B_1	3111	3112	3121	3122	3131	3132
	B_2	3211	3212	3221	3222	3231	3232
	B_3	3311	3312	3321	3322	3331	3332
	B_4	3411	3412	3421	3422	3431	3432
A_4	B_1	4111	4112	4121	4122	4131	4132
	B_2	4211	4212	4221	4222	4231	4232
	B_3	4311	4312	4321	4322	4331	4332
	B_4	4411	4412	4421	4422	4431	4432
A_5	B_1	5111	5112	5121	5122	5131	5132
	B_2	5211	5212	5221	5222	5231	5232
	B_3	5311	5312	5321	5322	5331	5332
	B_4	5411	5412	5421	5422	5431	5432

CHART VII

MATRIX OF TREATMENTS BY YATES REPRESENTATION

		C_1		C_2		C_3	
		D_1	D_2	D_1	D_2	D_1	D_2
A_1	B_1	(1)	d	c	cd	c^2	c^2d
	B_2	b	bd	bc	bcd	bc^2	bc^2d
	B_3	b^2	b^2d	b^2c	b^2cd	b^2c^2	b^2c^2d
	B_4	b^3	b^3d	b^3c	b^3cd	b^3c^2	b^3c^2d
A_2	B_1	a	ad	ac	acd	ac^2	ac^2d
	B_2	ab	abd	abc	abcd	abc^2	abc^2d
	B_3	ab^2	ab^2d	ab^2c	ab^2cd	ab^2c^2	ab^2c^2d
	B_4	ab^3	ab^3d	ab^3c	ab^3cd	ab^3c^2	ab^3c^2d
A_3	B_1	a^2	a^2d	a^2c	a^2cd	a^2c^2	a^2c^2d
	B_2	a^2b	a^2bd	a^2bc	a^2bcd	a^2bc^2	a^2bc^2d
	B_3	a^2b^2	a^2b^2d	a^2b^2c	a^2b^2cd	$a^2b^2c^2$	$a^2b^2c^2d$
	B_4	a^2b^3	a^2b^3d	a^2b^3c	a^2b^3cd	$a^2b^3c^2$	$a^2b^3c^2d$
A_4	B_1	a^3	a^3d	a^3c	a^3cd	a^3c^2	a^3c^2d
	B_2	a^3b	a^3bd	a^3bc	a^3bcd	a^3bc^2	a^3bc^2d
	B_3	a^3b^2	a^3b^2d	a^3b^2c	a^3b^2cd	$a^3b^2c^2$	$a^3b^2c^2d$
	B_4	a^3b^3	a^3b^3d	a^3b^3c	a^3b^3cd	$a^3b^3c^2$	$a^3b^3c^2d$
A_5	B_1	a^4	a^4d	a^4c	a^4cd	a^4c^2	a^4c^2d
	B_2	a^4b	a^4bd	a^4bc	a^4bcd	a^4bc^2	a^4bc^2d
	B_3	a^4b^2	a^4b^2d	a^4b^2c	a^4b^2cd	$a^4b^2c^2$	$a^4b^2c^2d$
	B_4	a^4b^3	a^4b^3d	a^4b^3c	a^4b^3cd	$a^4b^3c^2$	$a^4b^3c^2d$

CHART VIII
DATA FOR THE 5x4x3x2 ORTHOGONAL DESIGN

		C_1		C_2		C_3	
		D_1	D_2	D_1	D_2	D_1	D_2
A_1	B_1	2	2	2	2	3	8
	B_2	1	2	2	3	4	4
	B_3	2	4	6	2	4	4
	B_4	1	2	2	4	3	2
A_2	B_1	3	4	5	7	9	10
	B_2	2	2	3	7	7	8
	B_3	3	6	6	6	8	10
	B_4	1	4	3	6	5	7
A_3	B_1	4	6	8	10	12	14
	B_2	2	4	6	10	10	12
	B_3	6	8	8	10	14	16
	B_4	2	6	6	8	8	10
A_4	B_1	5	8	11	13	15	18
	B_2	2	6	9	13	13	17
	B_3	9	10	10	14	20	22
	B_4	3	7	9	10	11	13
A_5	B_1	6	10	14	18	21	20
	B_2	3	6	10	17	16	19
	B_3	10	12	10	18	24	28
	B_4	3	11	10	12	13	18

$C_{1000}(\equiv A_L)$ 652	$C_{0100}(\equiv B_L)$ −120	$C_{0010}(\equiv C_L)$ 290	$C_{0001}(\equiv D_L)$ 817
$C_{2000}(\equiv A_Q)$ 000	$C_{0200}(\equiv B_Q)$ −60	$C_{0020}(\equiv C_Q)$ 10	
$C_{3000}(\equiv A_C)$ -14	$C_{0300}(\equiv B_C)$ −340		
$C_{4000}(\equiv A_{Qt})$ 000			

TABLE 5

ALL THE MAIN EFFECT TOTALS

OF THE 5X4X3X2 DESIGN

$C_{1100}(\equiv A_L B_L)$ -52	$C_{2300}(\equiv A_Q B_C)$ -6	$C_{0220}(\equiv B_Q C_Q)$ -30	
$C_{1300}(\equiv A_L B_C)$ -234	$C_{3200}(\equiv A_C B_Q)$ 24	$C_{0310}(\equiv B_C C_L)$ -30	$C_{0011}(\equiv C_L D_L)$ -10
$C_{1010}(\equiv A_L C_L)$ 224	$C_{1001}(\equiv A_L D_L)$ 96	$C_{0120}(\equiv B_L C_Q)$ 30	$C_{0210}(\equiv B_Q C_L)$ -30
$C_{0110}(\equiv B_L C_L)$ -90	$C_{3001}(\equiv A_C D_L)$ 18		

TABLE 6

TWO-FACTOR EFFECT TOTALS

OF THE 5X4X3X2 DESIGN

(13 OUT OF 35)

THREE-FACTOR EFFECTS

$C_{0111}(\equiv B_L C_L D_L)$ −30	$C_{0221}(\equiv B_Q C_Q D_L)$ 30	$C_{0311}(\equiv B_C C_L D_L)$ −10
$C_{0211}(\equiv B_Q C_L D_L)$ −10	$C_{0321}(\equiv B_C C_Q D_L)$ −50	$C_{0121}(\equiv B_L C_Q D_L)$ 50

(6 OUT OF 50)

FOUR-FACTOR EFFECTS

$C_{2121}(\equiv A_Q B_L C_Q D_L)$ 414	$C_{1111}(\equiv A_L B_L C_L D_L)$ 438

(2 OUT OF 24)

TABLE 7

THREE AND FOUR FACTOR EFFECTS

TOTALS OF THE 5X4X3X2 DESIGN

h. <u>An Algorithm for Computing Elements of an Orthogonal Contrast</u>.

The $5 \times 4 \times 3 \times 2$ factorial design shown in Chart VI of the preceding section will be used to great advantage, it is hoped, to explain the technique of computing the elements of an orthogonal contrast.

Let us consider the general formula for obtaining the elements of the contrast matrix for the $5 \times 4 \times 3 \times 2$ design discussed in Section g:

(3h-1)

$$C_{i_1 i_2 i_3 i_4} = \sum_{j_1=1}^{5} \sum_{j_2=1}^{4} \sum_{j_3=1}^{3} \sum_{j_4=1}^{2} \lambda^{(1)}_{i_1 j_1} \lambda^{(2)}_{i_2 j_2} \lambda^{(3)}_{i_3 j_3} \lambda^{(4)}_{i_4 j_4} S_{j_1 j_2 j_3 j_4} .$$

As shown in <u>Chart I</u>, an orthogonal λ-matrix of polynomial values is associated with each of the four factors. Thus the matrices

(3h-2)

$$\left[\lambda^{(1)}_{i_1 j_1}\right], \quad \left[\lambda^{(2)}_{i_2 j_2}\right], \quad \left[\lambda^{(3)}_{i_3 j_3}\right], \quad \left[\lambda^{(4)}_{i_4 j_4}\right]$$

are associated respectively with factors F_1, F_2, F_3, F_4 respectively.

The superscripts of the elements of the λ-matrix associated with each factor, indicate the number of the factor with which that λ-matrix is associated. For example the superscript (3) indicates that this λ-matrix is associated with factor F_3, etc.

The number of rows of each λ-matrix is <u>one less</u> than the <u>number of levels</u> of the factor with which it is associated. The <u>number of columns</u> of each λ-matrix is equal to the number of levels of the associated factor. Thus the $\left[\lambda^{(1)}_{i_1 j_1}\right]$ matrix associated with F_1 has <u>four rows</u> and <u>five columns</u>.

The <u>row-number</u> indicates either the linear, or, quadratic, or cubic, or quartic component of the contrast; <u>row number two</u> indicates the quadratic component. (Either-or is meant in the <u>exclusive</u> sense.)

Each row in a λ-matrix is a row-matrix with its proper number attached to it. For example, the third row of polynomial values $(-1 \; +2 \; +0 \; -2 \; +1)$ in the λ-matrix in Chart I associated with F_1 is symbolized by:

$$\left[\lambda^{(1)}_{31} \quad \lambda^{(1)}_{32} \quad \lambda^{(1)}_{33} \quad \lambda^{(1)}_{34} \quad \lambda^{(1)}_{35}\right].$$

The first subscript of each element in this matrix is the number $\underline{3}$, which alludes to the cubic component of the contrast.

Let us explain how to compute an element of the contrast matrix for this design, say, $C_{\underline{3121}}$. The procedure is the same for computing any element of a contrast matrix:

The following set of four ordered doublets is formed:

$$\left[\underline{3}j_1 \; ; \underline{1}j_2 \; ; \underline{2}j_3 \; ; \underline{1}j_4\right]. \tag{3h-3}$$

These doublets have a positional order: $3j_1$ is the first doublet, $1j_2$ the second doublet, etc. The number of doublets in a set is equal to the number of factors.

The first numbers of the first, second, third and fourth doublets in (3h-3) are taken from the first, second, third, and fourth subscripts respectively in $C_{\underline{3121}}$. These are underlined in (3h-3).

The second members of the doublets in (3h-3) are taken from the (positional) subscripts of the elements of the data-martix, $S_{j_1 j_2 j_3 j_4}$, namely j_1, j_2, j_3, j_4; with j_1 going to the first doublet, j_2 to the second, etc. Thus the subscripts of the general contrast symbol $(C_{i_1 i_2 i_3 i_4})$ furnish the first members of the doublets; the subscripts of the elements of the data-matrix furnish the second members of the doublets.

Each doublet of a set represents the symbol of a polynomial value of a λ-matrix associated with a particular factor. Thus in a set of doublets:

[31; 11; 21; 12] , the third doublet, 21, corresponding to the third factor in Chart I, symbolizes the polynomial value taken from the second row, first element of the matrix $\left[\lambda^{(3)}_{i_3 j_3}\right]$. This value is (+1).

Let us consider Charts II and IV in the computation of the contrast C_{3121}. It is to be noted that every element of the data-matrix is utilized in the computation of a single element of the contrast-matrix, such as C_{3121}, or say C_{4211}.

The computation of C_{3121}: Each element of the data-matrix is multiplied by the product of four polynomial values, one from each λ-matrix associated with its corresponding factor. For example, let us consider the data-element S_{1332}. For this purpose the following set of four ordered doublets is formed:

$$[3\underline{1};\ 1\underline{3};\ 2\underline{3};\ 1\underline{2}]. \tag{3h-4}$$

This is the same as in (3h-3) except that j_1, j_2, j_3 and j_4 take on the values of 1, 3, 3, and 2 respectively. These are underlined in (3h-4). The expression (3h-4) serves as a command in computer logic. It says in effect:

a. First doublet, 31: select from the λ-matrix corresponding to the first factor, the polynomial value from the third row and first column. This value is (- 1). (See Chart I for values of doublets.)

b. Second doublet, 13: select from the λ-matrix corresponding to the second factor the polynomial value from the first row and third column. This value is (+1).

c. Third doublet, 23: select from the λ-matrix corresponding to the third factor the polynomial value from the second row and third column. This value is (+1).

d. Fourth doublet, 12: select from the λ-matrix corresponding to the fourth factor the polynomial value from the first row and second column. This value is (+1). The product of these four polynomial values corresponding to the data (matrix) element S_{1332} is (- 1). This product is then multiplied by the value of the data (matrix) element S_{1332} (=4 from Chart VIII). The result is (-4) for this data element. For the computation of C_{3121}, see Chart IV. It contains the complete set of the products of quadruples (four polynomial values) associated with each element of the data-matrix given in Chart VIII.

The sum of the products of the appropriate polynomial values multiplied by each data (matrix) element yields the value of the contrast C_{3121}; this value is 414. Note that the first members of all the four ordered doublets formed for computing the contrast C_{3121} have the values 3, 1, 2, 1 in the natural order 1, 2, 3, 4 respectively. This is indicated in expression (3h-3) where the first members of the doublets are fixed (namely, 3, 1, 2, 1) and j_1, j_2, j_3 and j_4 vary to generate the indices of (the elements of) the data matrix.

This procedure is valid for computing contrasts of an m-factor orthogonal design. The validity is assured by the general formula given in the preceding pages. This formula represents the sum of products of each element of a data-matrix by m appropriately (uniquely) selected polynomial values.

To compute a particular contrast $C_{i_1^* i_2^* \cdots i_m^*}$ we form the set of doublets:

(3h-5)
$$\left[i_1^* \, j_1; \; i_2^* \, j_2; \; i_3^* \, j_3; \; \cdots; i_m^* \, j_m \right],$$

where $i_1^*, i_2^*, \cdots, i_m^*$ are fixed values for the particular contrast and $j_1, j_2, \cdots, j_m$ vary, taking on the values of the indices (subscripts) of the elements of the data-matrix.

Suppose one wishes to compute, say, the contrast C_{3020} ($\equiv A_C C_Q$). We form the following set of doublets:

(3h-5a)
$$\left[3j_1; \; 0j_2; \; 2j_3; \; 0j_4 \right]$$

where j_1, j_2, j_3, j_4 take on the values of the subscripts of the elements of the data-matrix.

Now $0j_k$ is the symbol representing $\lambda^{(k)}_{0j_k}$; and since $\lambda^{(k)}_{0j_k} = 1$ for $1 \leq k \leq m$, $0j_k$ represents the value 1. For example, in computing C_{3020} consider the data-element, S_{5412}. (See Chart III). The set of doublets for this data-element is:

$$\left[35;\ \underline{04};\ 21;\ \underline{02}\right] \quad . \tag{3h-6}$$

Since 04 and 02 each represent the value 1, the coefficient-multipliers are:

$$(+1),\ (\underline{1}),\ (+1),\ (\underline{1}). \tag{3h-7}$$

The value of the product of these coefficients is +1; this value is in turn multiplied by the data-element represented by 5412 of the data-matrix of Chart VIII, namely the number 11. The result is +11 for the element S_{5412}. Similarly this is done for every data-element and the results are summed to yield the value of C_{3020}.

Note that the underlined second and fourth doublets, in (3h-6), namely $\underline{04}$ and $\underline{02}$ and their respective referents (1) and (1) in (3h-7) do not represent polynomial values. They merely indicate that the value 1 is to be used as a multiplier in lieu of the ordered polynomial values for the factors (F_2 and F_3) that are not being considered, i.e., are eliminated, in the calculation of this particular contrast, namely C_{3020} ($= A_C C_Q$). Factors, F_2(=B) and F_4(=D) are not taken into account in this contrast.

In general an ordered doublet such as $0j_k$ whose first member is 0 represents the value 1; and 1 is to be used as a multiplier in lieu of the positionally corresponding polynomial value for the missing k-th factor. Thus for the computation of the contrast, $C_{i_1^* i_2^* 0i_4^* \cdots 0i_k^* \cdots i_m^*}$ the following sets of ordered doublets are generated:

(3h-8) $$\left[i_1^* j_1;\ i_2^* j_2;\ 0j_3; i_4^* j_4; \cdots; 0j_{k-1}; i_k^* j_k; \cdots; i_m^* j_m\right]$$

since $0j_3 = 0j_{k-1} = 1$ (3h-8) becomes:

(3h-9) $$\left[i_1^* j_1;\ i_2^* j_2; 1\ ; i_4^* j_4; \ldots; 1; i_k^* j_k\ ; \ldots; i_m^* j_m\right],$$

where i_1^*, i_2^*, i_4^*, etc. are fixed values for the particular contrast and j_1, j_2, ..., j_m vary by taking on the values of the subscripts of the elements of the data-matrix.

4. CONCLUSIONS. The algorithm and general formulas developed and presented in this paper for obtaining and computing orthogonal contrasts are ideally suited for a computer. The factors, their levels, the associated λ-matrices of polynomial values, and the elements of the data-matrix, are represented by unique sets of positionally ordered numerical indices (subscripts and superscripts).

What is of equal significance is that the orthogonal contrasts are obtained by virtue of a unique correspondence and relationship between the positionally ordered numerical indices of the contrasts and those of the λ-matrices and of the elements of the data matrix. This situation is amply conducive to machine computations that involve sums of products.

To Summarize: There are three principal entities required for the computation of orthogonal contrasts. These are: the factors and their levels, their corresponding λ-matrices of polynomial values, and the data-matrix.

The computer identifies the factors by means of a set of numbers: $\{1, 2, \ldots, m\}$.

The levels of each factor are identified in the computer by means of a set of positionally ordered numbers: $\{j_1, j_2, \ldots, j_m\}$,

where:

$$\begin{array}{ll} j_1 = 1,\ 2,\ \ldots,\ N_1 & i_1 = 1,\ 2,\ \ldots,\ N_1 - 1 \\ j_2 = 1,\ 2,\ \ldots,\ N_2 & i_2 = 1,\ 2,\ \ldots,\ N_2 - 1 \\ \text{- - - - - - - - - - - -} & \text{- - - - - - - - - - - - - -} \\ j_m = 1,\ 2,\ \ldots,\ N_m & i_m = 1,\ 2,\ \ldots,\ N_m - 1\ . \end{array}$$

The computer stores one λ-matrix with the proper tag-number corresponding to each factor:

$$\left[\lambda^{(1)}_{i_1 j_1}\right],\quad \left[\lambda^{(2)}_{i_2 j_2}\right],\quad \ldots\ ,\ \left[\lambda^{(m)}_{i_m j_m}\right].$$

The tag-numbers are the superscripts: (1), (2), ..., (m). These correspond (1-1) to the factor numbers: $\{1,\ 2,\ \ldots,\ m\}$.

The rows and columns of the λ-matrices are identified by the sets of positionally ordered numbers: $\{i_1,\ i_2,\ \ldots,\ i_m\}$ and $\{j_1,\ j_2,\ \ldots,\ j_m\}$ respectively.

The elements of the data-matrix are numbered by varying the positional subscripts:

$$\{j_1,\ j_2,\ \ldots,\ j_m\}\ .$$

To obtain a specific contrast $C_{i_1^* i_2^* \ldots i_m^*}$ each element of the data-matrix is multiplied by the product of m appropriately ordered polynomial values:

$$\lambda^{(1)}_{i_1^* j_1}\ ,\quad \lambda^{(2)}_{i_2^* j_2}\ ,\ \ldots,\lambda^{(m)}_{i_m^* j_m}\quad .$$

(Here $S_{j_1 j_2 \ldots j_m}$ is a particular element of the data-matrix.)

The results for each data-element are summed to obtain the effect total of the desired contrast:

$$C_{i_1^* i_2^* \ldots i_m^*} = \sum_{j_1=1}^{N_1} \sum_{j_2=1}^{N_2} \cdots \sum_{j_m=1}^{N_m} \lambda^{(1)}_{i_1^* j_1} \lambda^{(2)}_{i_2^* j_2} \cdots \lambda^{(m)}_{i_m^* j_m} S_{j_1 j_2 \ldots j_m} .$$

To compute contrasts in which one or more of the subscripts $\{i_k\}$ $(1 \leq k \leq m)$ is zero the following procedure is adopted:

The value 1 is substituted in the proper positional order for the polynomial value corresponding to the factor(s) that is being eliminated in computing a desired contrast. As was noted on several occassions, $C_{000\ldots0}$, the sum total of all the observations is not a contrast.

5. SUMMARY OF FORMULAS. What follows is a summary of the formulas for obtaining contrasts in orthogonal designs; and an explication of the symbolic terms and of the index notation used in this report.

SUMMARY OF FORMULAS

SUMMARY OF FORMULAS FOR GENERATING THE ELEMENTS OF AN ORTHOGONAL POLYNOMIAL CONTRAST MATRIX

1. General Element of Orthogonal Contrast (Interaction) Matrix for Factors $F_1, F_2, \ldots F_m$.

$$C_{i_1 i_2 \ldots i_m} = \sum_{j_1=1}^{N_1} \sum_{j_2=1}^{N_2} \cdots \sum_{j_m=1}^{N_m} \lambda^{(1)}_{i_1 j_1} \lambda^{(2)}_{i_2 j_2} \cdots \lambda^{(m)}_{i_m j_m} S_{j_1 j_2 \ldots j_m}$$

$$\equiv \sum_{j_1 j_2 \ldots j_m}^{N_1, N_2, \ldots, N_m} \lambda^{(1)}_{i_1 j_1} \lambda^{(2)}_{i_2 j_2} \cdots \lambda^{(m)}_{i_m j_m} S_{j_1 j_2 \ldots j_m}$$

where

$$\sum_{j_1=1}^{N_1} \sum_{j_2=1}^{N_2} \cdots \sum_{j_m=1}^{N_m} \equiv \sum_{j_1 j_2 \ldots j_m}^{N_1, N_2, \ldots, N_m}$$

a. Elements of Data (Observations) Matrix: $S_{j_1 j_2 \ldots j_m}$

b. Factors: $F_1, F_2, \ldots, F_m$

c. Levels of Factors $F_1, F_2, \ldots, F_m$

N_1 = Number of levels of F_1

N_2 = " " " " F_2

....

N_m = " " " " F_m

d. Values of Subscripts

$i_1 = 0, 1, 2, \ldots, N_1 - 1$	$j_1 = 1, 2, \ldots, N_1$
$i_2 = 0, 1, 2, \ldots, N_2 - 1$	$j_2 = 1, 2, \ldots, N_2$
....	....
$i_m = 0, 1, 2, \ldots, N_m - 1$	$j_m = 1, 2, \ldots, N_m$

Summary of Formulas for Generating the Elements of an Orthogonal Polynomial Contrast Matrix (Cont'd)

e. Index Notation

$i_m j_m$: element in the $i^{\underline{th}}$ row and $j^{\underline{th}}$ column of the orthogonal polynomial coefficient matrix for the $m^{\underline{th}}$ factor F_m.

i_m = 1, 2, 3, ..., $N_m - 1$ yields the linear, quadratic, cubic, ..., $(N_m - 1)$th contrasts respectively.

j_m = 1, 2, 3, ..., N_m refers to the first, second, third, ..., N_mth levels of the factors F_1, F_2, F_3, ..., F_m.

$\lambda^{(m)}_{i_m j_m}$: row vectors of the transpose matrix of the Fisher orthogonal coefficients of factor F_m, with the restriction that $i_m \neq 0$.

$\Lambda^{(m)}_{i_m j_m}$: normalized form of $\lambda^{(m)}_{i_m j_m}$

(2) Formula for Matrix $\left[\lambda^{(m)}_{i_m j_m}\right]$

$$\left[\lambda^{(m)}_{i_m j_m}\right] = \begin{bmatrix} \lambda^{(m)}_{11} & \lambda^{(m)}_{12} & \cdots & \lambda^{(m)}_{1,N_m} \\ \lambda^{(m)}_{21} & \lambda^{(m)}_{22} & \cdots & \lambda^{(m)}_{2,N_m} \\ \lambda^{(m)}_{31} & \lambda^{(m)}_{32} & \cdots & \lambda^{(m)}_{3,N_m} \\ \cdots & \cdots & \cdots & \cdots \\ \lambda^{(m)}_{N_m-1,1} & \lambda^{(m)}_{N_m-1,2} & \cdots & \lambda^{(m)}_{N_m-1,N_m} \end{bmatrix} \quad \begin{bmatrix} \text{Linear contrast coefficients} \\ \text{Quadratic contrast coefficients} \\ \text{Cubic contrast coefficients} \\ \cdots\cdots\cdots \\ (N_m-1)\text{th contrast coefficients} \end{bmatrix}$$

Summary of Formulas for Generating the Elements of an Orthogonal Polynomial Contrast Matrix (Cont'd)

(3) Formula for the Special Case

$\lambda^{(k)}_{0j_k}$ = 1 for $1 \leq k \leq m$ and for all j_k's . Symbol used to obtain the total sum of the observations; also used to obtain interactions of order lower than the highest order.

(4) Formula for Normalized Form, $\Lambda^{(k)}_{i_k j_k}$, of $\lambda^{(k)}_{i_k j_k}$, $1 \leq k \leq m$

$$\Lambda^{(k)}_{i_k j_k} = \frac{\lambda^{(k)}_{i_k j_k}}{\sqrt{\sum_{j_k=1}^{N_k}\left[\lambda^{(k)}_{i_k j_k}\right]^2}}, \quad i_k \neq 0$$

(5) Special Cases of Orthogonal Contrasts

a. $$C_{i_1 0 i_3 0 0 i_6 0 \ldots 0} = \sum_{j_1 j_2 \cdots j_m}^{N_1, N_2, \ldots, N_m} \lambda^{(1)}_{i_1 j_1} \lambda^{(2)}_{0 j_2} \lambda^{(3)}_{i_3 j_3} \lambda^{(4)}_{0 j_4} \lambda^{(5)}_{0 j_5} \lambda^{(6)}_{i_6 j_6} \lambda^{(7)}_{0 j_7} \cdots \lambda^{(m)}_{0 j_m} S_{j_1 j_2 \cdots j_m}$$

$$= \sum_{j_1 j_2 \cdots j_m}^{N_1, N_2, \ldots, N_m} \lambda^{(1)}_{i_1 j_1} \lambda^{(3)}_{i_3 j_3} \lambda^{(6)}_{i_6 j_6} S_{j_1 j_2 \cdots j_m}, \text{ since } \lambda^{(k)}_{0 j_k} = 1 \text{ for } 1 \leq k \leq m$$

b. Main Effects

$$C_{i_1 0 0 \ldots 0} = \sum_{j_1 j_2 \cdots j_m}^{N_1, N_2, \ldots, N_m} \lambda^{(1)}_{i_1 j_1} \lambda^{(2)}_{0 j_2} \cdots \lambda^{(m)}_{0 j_m} S_{j_1 j_2 \cdots j_m}$$

Summary of Formulas for Generating the Elements of an Orthogonal Polynomial Contrast Matrix (Cont'd)

(5) Special Cases of Orthogonal Contrasts (Cont'd)

b. Main Effects (Cont'd)

$$C_{i,00\cdots0} = \sum_{j_1 j_2 \cdots j_m}^{N_1, N_2, \ldots, N_m} \lambda_{i,j_1}^{(1)} S_{j_1 j_2 \cdots j_m}$$

= Main (Total) effect of factor F_1 of order one

Similarly for main effects of F_2, F_3, ..., F_m.

c. Total Sum of Observations $S_{j_1 j_2 \cdots j_m}$

$$C_{00\cdots0} = \sum_{j_1 j_2 \cdots j_m}^{N_1, N_2, \ldots, N_m} \lambda_{0j_1}^{(1)} \lambda_{0j_2}^{(2)} \cdots \lambda_{0j_m}^{(m)} S_{j_1 j_2 \cdots j_m}$$

$$= \sum_{j_1 j_2 \cdots j_m}^{N_1, N_2, \ldots, N_m} S_{j_1 j_2 \cdots j_m}, \text{ since } \lambda_{0j_k}^{(k)} = 1 \text{ for } 1 \le k \le m$$

= T (Total sum of observations with one unit per cell)

$$C_{00\cdots0} = rT = \sum_{k=1}^{r} T_k = T',$$ assuming r replications per cell (treatment)

d. Grand Average (G.A.)

(1) Without Replication

$$G.A. = \frac{C_{00\cdots0}}{\prod_{k=1}^{m} N_k} = \frac{T}{\prod_{k=1}^{m} N_k}$$

(2) With Replication

$$G.A. = \frac{rT}{r N_1 N_2 \cdots N_3} = \frac{T'}{r \prod_{k=1}^{m} N_k}$$

Summary of Formulas for Generating the Elements of an Orthogonal Polynomial Contrast Matrix (Cont'd)

(6) **Sum of Squares (SS) of** $C_{i_1 i_2 \ldots i_m}$, **with Replication**

Interaction of Factors $F_1, F_2, \ldots, F_m$.

$$SS(C_{i_1 i_2 \ldots i_m}) = \frac{1}{r}\left[\sum_{j_1 j_2 \ldots j_m}^{N_1, N_2, \ldots, N_m} \Lambda^{(1)}_{i_1 j_1} \Lambda^{(2)}_{i_2 j_2} \cdots \Lambda^{(m)}_{i_m j_m} S_{j_1 j_2 \ldots j_m}\right]^2.$$

$$= \frac{1}{r}\left[\sum_{j_1 j_2 \ldots j_m}^{N_1, N_2, \ldots, N_m} \prod_{k=1}^{m} \Lambda^{(k)}_{i_k j_k} S_{j_1 j_2 \ldots j_m}\right]^2$$

Where

$$\Lambda^{(k)}_{i_k j_k} \equiv \frac{\lambda^{(k)}_{i_k j_k}}{\sqrt{\sum_{j_k=1}^{N_k}\left[\lambda^{(k)}_{i_k j_k}\right]^2}}, \quad i_k \neq 0$$

Summary of Formulas for Generating the Elements of an Orthogonal Polynomial Contrast Matrix (Cont'd)

(7) Special Case. Sum of Squares (SS) of $C_{i_1 0 i_3 0 \cdots 0}$, with Replication. Interaction of First and Third Factors F_1 and F_3.

$$SS(C_{i_1 0 i_3 0 \cdots 0}) = \frac{N_1 N_3}{r \prod_{k=1}^{m} N_k} \left[\sum_{j_1 j_2 \cdots j_m}^{N_1, N_2, \ldots, N_m} \frac{\lambda^{(1)}_{i_1 j_1}}{\sqrt{\sum_{j_1=1}^{N_1} [\lambda^{(1)}_{i_1 j_1}]^2}} \cdot \lambda^{(2)}_{0 j_2} \cdot \frac{\lambda^{(3)}_{i_3 j_3}}{\sqrt{\sum_{j_3=1}^{N_3} [\lambda^{(3)}_{i_3 j_3}]^2}} \cdot \lambda^{(4)}_{0 j_4} \cdots \lambda^{(m)}_{0 j_m} S_{j_1 j_2 \cdots j_m} \right]^2$$

$$= \frac{N_1 N_3}{r \prod_{k=1}^{m} N_k} \left[\sum_{j_1 j_2 \cdots j_m}^{N_1, N_2, \ldots, N_m} \frac{\lambda^{(1)}_{i_1 j_1}}{\sqrt{\sum_{j_1=1}^{N_1} [\lambda^{(1)}_{i_1 j_1}]^2}} \cdot \frac{\lambda^{(3)}_{i_3 j_3}}{\sqrt{\sum_{j_3=1}^{N_3} [\lambda^{(3)}_{i_3 j_3}]^2}} S_{j_1 j_2 \cdots j_m} \right]^2, \text{ since } \lambda^{(k)}_{0 j_k} = 1 \text{ for } 1 \leq k \leq m$$

$$= \frac{N_1 N_3}{r \prod_{k=1}^{m} N_k} \left[\sum_{j_1 j_2 \cdots j_m}^{N_1, N_2, \ldots, N_m} \Lambda^{(1)}_{i_1 j_1} \Lambda^{(3)}_{i_3 j_3} S_{j_1 j_2 \cdots j_m} \right]^2$$

Summary of Formulas for Generating the Elements of an Orthogonal Polynomial Contrast Matrix (Cont'd)

Likewise, for the interaction of factors F_2, F_5 and F_8,

$$SS(C_{0i_2 00 i_5 00 i_8 0 \ldots 0}) = \frac{N_2 N_5 N_8}{r \prod_{K=1}^{m} N_k} \left[\sum_{j_1 j_2 \cdots j_m}^{N_1, N_2, \ldots, N_m} \Lambda^{(2)}_{i_2 j_2} \Lambda^{(5)}_{i_5 j_5} \Lambda^{(8)}_{i_8 j_8} S_{j_1 j_2 \cdots j_m} \right]^2$$

6. ACKNOWLEDGEMENT. The author is glad to have the opportunity of expressing his gratitude and deep appreciation to his colleague George Millman for the constructive and helpful suggestions as well as for the aid in the preparation of this paper.

BIBLIOGRAPHY

Anderson and Bancroft: "Statistical Theory in Research", 1952, McGraw-Hill (New York, Toronto, London)

Bennett and Franklin: "Statistical Analysis in Chemistry and the Chemical Industry", 1954 Wiley (New York and London)

Cochran and Cox: "Experimental Designs", 2nd Ed., 1957 Wiley (New York and London)

Cox, D. R.: "Planning of Experiments", 1958 Wiley (New York and London)

Davies, O. L.: "The Design and Analysis of Industrial Experiments", 1954, Oliver and Boyd (London and Edinburgh)

Fisher, R. A. and Yates, F.: "Statistical Tables for Biological, Agricultural and Medical Research", 1949, Oliver and Boyd (London and Edinburgh)

Kempthorne, O.: "The Design and Analysis of Experiment", 1952, Wiley (New York and London)

Scheffé, H.: "The Analysis of Variance", 1959, Wiley (New York and London)

A SEMIAUTOMATIC WAR GAMING SYSTEM

John L. Donaldson
Research Analysis Corporation
and
Thomas R. Shaw
Operations Research, Inc.

SUMMARY

PROBLEM. To devise a method whereby a digital computer can be used to support war gaming activities and to design a set of computer programs which will accomplish this objective.

FACTS. War gaming is currently receiving much attention as a methodology in the field of operations research. However, the hand-played war game involves many disadvantages which tend to restrict the desirability of of its application to the solution of many problems which otherwise lend themselves well to resolution by gaming techniques. Basically it is an extremely costly venture. A large group of experienced players and skilled controllers is required to operate a game. Further, a great amount of time can be spent in repetitive, laborious calculations which can affect both the accuracy and timeliness of the results.

On the other hand, the high speed digital computer can perform certain data processing functions with great rapidity and accuracy. There are, however, limitations to the usefullness of computers particularly in the performance of functions that require a high degree of human insight and decision. The simulation of decision processes by computers is possible only when the criteria for judgment and the alternatives for action can be adequately described in a quantitative fashion. Also the speed and accuracy of computers can be compromised by the need for frequent human intervention. The computer, therefore, can be used to best advantage in those areas where present human knowledge and ability permits.

DISCUSSION. The semiautomatic war gaming system has two aspects - man and machine. In the system the computer has three functions: performing the assessment calculations, maintaining the quantitative records, and displaying the results, while the man portion of the system is concerned primarily with the problems of decision-making. These two aspects of the system are treated in this paper in terms of the flow of information. The operations which comprise the system are discussed according to input, function, and output. For the human aspect of the system these items are

related to the responsibilities which they entail; in the computer portion of the system these items are examined quantitatively in terms of data sets.

The system was designed for the THEATERSPIEL Study (35.10) and was used in the play of POMEX I. From this initial attempt, much experience was gained leading to the projection of certain general consequences of such an application of computer techniques. Some degree of insight was realized concerning the multi-faceted problem of coordinating the many functions which comprise the system. Finally the learning process enabled other uses for this system to become evident and suggested ways in which further refinements could be incorporated conducive to greater applicability to other problems.

INTRODUCTION

The THEATERSPIEL Study was established in October, 1959, with Mr Richard E. Zimmerman (Chairman), BGen John G. Hill, USA (Ret.), and Capt. J. O. F. Dorsett, USN (Ret.), being assigned as the original members of the study group. As originally conceived, the purpose of the project was to be an outgrowth of the FAME game[1]; it was to develop a theater level war game that would indicate the Army's need in future military operations paying particular attention to the requirements placed on the development of TO/Es. Further, the study was to draw on the resources provided by computer usage.

In the beginning the members of the study were aware of the advantages and disadvantages inherent in various gaming methodologies as applied to the study at hand. Drawing on their previous experiences in simulations and war gaming, they felt that it would be undesirable to program all facets of the gaming environment, e.g., certain of the presently unquantifiable aspects of combat decision-making, and yet there was a desire to avoid the tedium and repetition of much of the assessment phase of the gaming operation. It was this particular thinking that led to the design of the semiautomatic gaming system.

However the semiautomatic system described in this paper did not immediately result from this initial attitude. It was necessary first to determine to what extent the game should be automated; that is, which should be the human functions in the game and which the computer's. One of the difficulties concerned with defining the programming the human decision function, a basic characteristic of the gaming approach, was expressed by the late John Von Neuman in The Computer and the Brain:[2]

A SEMIAUTOMATIC WAR GAMING SYSTEM

John L. Donaldson
Research Analysis Corporation
and
Thomas R. Shaw
Operations Research, Inc.

SUMMARY

PROBLEM. To devise a method whereby a digital computer can be used to support war gaming activities and to design a set of computer programs which will accomplish this objective.

FACTS. War gaming is currently receiving much attention as a methodology in the field of operations research. However, the hand-played war game involves many disadvantages which tend to restrict the desirability of of its application to the solution of many problems which otherwise lend themselves well to resolution by gaming techniques. Basically it is an extremely costly venture. A large group of experienced players and skilled controllers is required to operate a game. Further, a great amount of time can be spent in repetitive, laborious calculations which can affect both the accuracy and timeliness of the results.

On the other hand, the high speed digital computer can perform certain data processing functions with great rapidity and accuracy. There are, however, limitations to the usefullness of computers particularly in the performance of functions that require a high degree of human insight and decision. The simulation of decision processes by computers is possible only when the criteria for judgment and the alternatives for action can be adequately described in a quantitative fashion. Also the speed and accuracy of computers can be compromised by the need for frequent human intervention. The computer, therefore, can be used to best advantage in those areas where present human knowledge and ability permits.

DISCUSSION. The semiautomatic war gaming system has two aspects - man and machine. In the system the computer has three functions: performing the assessment calculations, maintaining the quantitative records, and displaying the results, while the man portion of the system is concerned primarily with the problems of decision-making. These two aspects of the system are treated in this paper in terms of the flow of information. The operations which comprise the system are discussed according to input, function, and output. For the human aspect of the system these items are

related to the responsibilities which they entail; in the computer portion of the system these items are examined quantitatively in terms of data sets.

The system was designed for the THEATERSPIEL Study (35.10) and was used in the play of POMEX I. From this initial attempt, much experience was gained leading to the projection of certain general consequences of such an application of computer techniques. Some degree of insight was realized concerning the multi-faceted problem of coordinating the many functions which comprise the system. Finally the learning process enabled other uses for this system to become evident and suggested ways in which further refinements could be incorporated conducive to greater applicability to other problems.

INTRODUCTION

The THEATERSPIEL Study was established in October, 1959, with Mr Richard E. Zimmerman (Chairman), BGen John G. Hill, USA (Ret.), and Capt. J. O. F. Dorsett, USN (Ret.), being assigned as the original members of the study group. As originally conceived, the purpose of the project was to be an outgrowth of the FAME game[1]; it was to develop a theater level war game that would indicate the Army's need in future military operations paying particular attention to the requirements placed on the development of TO/Es. Further, the study was to draw on the resources provided by computer usage.

In the beginning the members of the study were aware of the advantages and disadvantages inherent in various gaming methodologies as applied to the study at hand. Drawing on their previous experiences in simulations and war gaming, they felt that it would be undesirable to program all facets of the gaming environment, e.g., certain of the presently unquantifiable aspects of combat decision-making, and yet there was a desire to avoid the tedium and repetition of much of the assessment phase of the gaming operation. It was this particular thinking that led to the design of the semiautomatic gaming system.

However the semiautomatic system described in this paper did not immediately result from this initial attitude. It was necessary first to determine to what extent the game should be automated; that is, which should be the human functions in the game and which the computer's. One of the difficulties concerned with defining the programming the human decision function, a basic characteristic of the gaming approach, was expressed by the late John Von Neuman in The Computer and the Brain:[2]

> Hence it is to be expected that an efficiently organized large natural automation (like the human nervous system) will tend to pick up as many logical (or informational) items as possible simulataneously, and process them simultaneously, while an efficiently organized large artificial automaton (like a large modern computing machine) will be more likely to do things successively - one thing at a time, or at any rate not so many things at a time. . . natural automata are likely to be highly <u>parallel</u>, while. . . artificial automata will tend to be. . . <u>serial</u>.

Establishing thus the area for one boundary, to maintain certain decisions as human responsibility, the question next to be faced was how far to push this boundary.

Soon after RAdm. Marion N. Little, USN (Ret.), Mr. William H. Sutherland, and Mr. Billy L. Himes joined the study in the spring of 1960, work was commenced on the SANDWAR series of games.[3] It had been decided that in this case the computer would be used primarily to maintain the records of the play. In accordance with this decision, a system was devised by which all units being played in the game were recorded by computer methods on a file (see Chapters 2 and 3), and the results generated by the hand-played assessment models were incorporated into this file providing a current record on the status of all units. It was to this effort that the Chief of the Strategic Division, Dr. Joseph O. Harrison, Jr., first introduced the support of the Computing Laboratory; Maj. R. G. Williams, USA (Ret.) of the COMPLAB together with Mr. Himes were for the most part responsible for much of the computer work at this time. Shortly thereafter one of the authors of this paper, Mr. Donaldson, joined the study and began working with Maj. Williams and Mr. Himes on the records system.

During the play of the SANDWAR games it soon became apparent that the computer could perform additional functions in supporting game play. As had been earlier realized, "it is. . . advantageous, as far as possible, to remove the human element from any elaborate chain of computation, and only to introduce it where it is absolutely unavoidable, at the very beginning and at the very end."[4] It was on the basis of thinking along this line that it was decided in late 1960 to mechanize the greater part of the assessment phase of the gaming operation. Four of the previously hand-played models were to be programmed for the POMEX series of games: one for air combat, one for support weapons effect, one for ground combat, and finally one for logistics. The records system which had been used during the SANDWAR games was to form the starting point for the development of the system to incorporate the four models. Mr. George E. Clark, Jr., CAMPLAB Division Chief, made additional personnel available for the undertaking, which was begun early in 1961.

Thus it was the other author of this paper, Mr. Shaw, was assigned to work on the master program along with Maj. Williams and Mr. Donaldson. Continuing in the direction indicated by the records system, it was realized that this to a great extent determined what the requirements were for certain aspects of the input and the output of the proposed system. It suggested a method of programming. While the work was being done on the master program, concurrent with this effort the above mentioned members of the THEATERSPIEL group and Mr. David B. Webster, who joined the study at this time, together with support from COMPLAB, programmed the four assessment models.* The design and structure of these models is reported in a series of papers.[5, 6, 7] POMEX I was played in the latter part of July and early August of 1961 using the new system. A complete report on the play of this game is in preparation.[7] It is the purpose of the present paper to describe only the master program and the context in which it was placed and to discuss some of the consequences of using such an approach.

That this approach, the semiautomatic gaming system, is of current interest and of significant value is demonstrated in a paper recently prepared for the Defense Atomic Support Agency discussing the need for a gaming system which will meet the gaming requirements of the Joint Chiefs of Staff. In the paper much attention has been focused on the problem of devising a computer-assisted war gaming capability.[8] The present paper presents one such system.

---

* Air: Capt. Dorsett, programming support, Miss Arla E. Weinert; Support Weapons and Ground Combat Models: Mr. Webster and Mr. Sutherland, programming support, Mrs. Barbara Fain and Mr. J. B. Creegan; Logistics Model: RAdm. Little and Mr. Himes, programming support, Mr. Donaldson.

Chapter 1

GAME ENVIRONMENT. In the semiautomatic system discussed in this paper, the digital computer fulfills two functions: first, it performs the game assessment calculations, relieving the control group of this tedious, time-consuming responsibility; and second, it serves as a bookkeeper, providing complete numerical records of the play, interval by interval, in a form suitable for post-game analysis. To appreciate this application of a computer and its consequences, the reader must first be familiarized with the system within which the computer operates.

The war gaming system can be considered as a sequence of related events, the relationship being what might be termed an "information flow". Thus for each event there is an input (which is the result of some prior event), some function which prescribes the manner in which this input is to be processed, and an output which is the result of this function (which will be input to the next event). By defining all events individually with regards to their inputs, functions, and outputs, the system as a whole is described. This chapter will examine the system in this manner, with one exception: The function of the computer operation, and its execution, will be the subject of a detailed discussion in the second chapter; in the present chapter computer input-output will be discussed only to the extent necessary for the continuity of development.

The events occurring within the system can be divided into three phases: pre-game planning, game play, and post-game analysis. Although each of these phases will be examined separately, it should be remembered that in reality they do not operate independently, since they too are related by an information flow, or input-output process.

PRE-GAME PLANNING. Once the study directive has been received, and it has been decided that war gaming is an appropriate method of solution of the problem, the pre-game planning phase is begun. The initial effort of this phase is to obtain a satisfactory statement of the problem together with specification of the purpose and objectives of the game. This does not preclude the possibility that in the later stages of this phase it may be necessary to redefine the problem and objectives repeatedly, however at the outset at least some general statement of purpose is a prerequisite to further development.

After the purpose has been determined, preparations for the game proceed along two parallel paths. Both the substantive and methodological aspects of play must be described. Consistent with the outlined objectives the game environment must be established. This includes choosing a locale, developing a scenario, and collecting pertinent data. The choice of local consists of selecting the geographical sector in which the game is to be played, of a size commensurate with the level of aggregation desired. The scenario includes the description of the political, economic, and cultural aspects of the environment leading up to the conflict. Also a part of the scenario are the TO/Es of the forces to be engaged in the conflict. In addition there arises the need for many other quantitative factors describing the geographic region, weapons; capabilities, and many similar data as required by the particular objectives of the study.

While this work is being done, attention must also be focused on developing rules and procedures for the play phase. This includes the rules according to which the players will make their decisions and the procedures by which control will implement the players' orders. Establishing procedures also includes the development of the assessment models since these models and the way in which they are programmed will reflect the decisions made with respect to procedures. In the semiautomatic system this is perhaps the most time-consuming element of the preparations and also the most critical. Efficient rules and procedures together with realistic models are among the most important aspects of the system.

As the mechanized components of the system are defined, and after the quantitative factors have been obtained, some time must be spent in putting these data in a form consistent with the input requirements of the models. Here again effective procedures will ensure less time being wasted during play of the game due to improper or inaccurate data.

Prior to the play of the game some time must be devoted to player orientation. The players must be briefed on the scenarios so as to become familiar with the environment for the game so that they might learn what is expected of them. They must be given their game objectives. Secondly they must be instructed on the rules of play. So that they might better expedite the system and use it to its full potential, they must also be given a good understanding of the mechanics of the play. Finally, they must be provided with a record of the status of their forces and all relevant data.

Orientation of the players is the final step prior to the play of the game. Once this has been accomplished, the second phase of the system can be initiated.

GAME PLAY. The game play phase can be considered as a repetitive cycle of events; the cycle being repeated until the pre-stated objectives of the play have been realized, i.e., until one of the player teams has been successful in achieving its predetermined goal. In some cases however, this may not be possible, and it then becomes the responsibility of the control group to terminate play.

The player teams initiate play be determining what tactics or strategies they wish to employ in acheiving their goals. On the basis of their mission and available forces, the players generate orders which are communicated to the control group. The control group then takes the orders issued by both player teams and integrates them, judging as to their relative feasibilities. Control, in rendering these decisions, considers such aspects as whether or not one side's forces can execute their orders without exposing themselves to enemy action, or whether or not a move is logistically feasible. Once control has evaluated the orders, it is necessary to specify the interactions that will result. Viewing the execution of both teams' orders with respect to one another, the control group is able to establish what interactions will occur.

As the battle situations become evident, control translates a description of these interactions into appropriate machine language. When all the battles have been so defined, it is then possible to feed this information into the computer. The computer, on the basis of the models programmed during the pre-play phase, then assesses the outcomes of the interactions of the opposing forces. It determines what has been gained and lost by the two sides. Upon completing these calculations, the computer then generates output which consists of the results of the assessment in terms of casualties, moves, and other similar information. These results are distributed to the two player teams and to the control group. On the basis of the results, the control group prepares a summary of the action for the players to supplement the machine results.

The cycle then begins again with the players weighing the results against the achievement of their objectives. Based on the current status of their forces and whatever intelligence estimates they may have received, they generate a new set of orders, and the cycle is repeated. This repetition occurs until, as said above, either the control group halts play, or a player team realizes its goal. When play is stopped, the last phase of the system begins.

POST-GAME ANALYSIS. Analysis of the game is perhaps the least defined aspect of the system. It can follow a number of different courses dependent upon the original intent of the study; nevertheless, there is a very general pattern which this phase might follow. First many questions must be asked, such as what were the critical aspects of the game, what caused the turning points of the action, how did the initial situation as defined in the pre-play phase affect the outcome of play? It must be determined what the essential elements of the game were that influenced the consequent action and how they affected that aspect of the play relevant to the stated problem. Analysis for these factors can be both quantitative and qualitative. The former lends itself well to being resolved on the computer, whereas the latter most generally is handled by the control group with support from the players. It is important to realize the potentiality of computer analysis of the results. Since the results have all been generated by the machine and complete records kept in machine lanquage, all the data required for a quantitative analysis of results are already in a form suitable for immediate machine analysis.

Interpreting the quantitative and qualitative analysis leads to the conclusions to be drawn from the game. From such a system both substantive and methodological conclusions may result. The methodological conclusions are then incorporated into the system improving it for the next play, while the substantive conclusions are either held until numerous repetitions of the game can further substantiate them, or else they are used to infer possible recommendations with regards to the original study directive.

Figure 1 summarizes the material presented in this chapter; each aspect of the system which is a separate event is enclosed within a rectangle; also included, in some cases, is a brief indication of the activities performed during the event. The diagram also serves to demonstrate the principle of information flow. In the following chapter the reason for the emphasis on this principle will become evident. The computer programs, the subject of the next chapter, have the primary function of providing for the proper flow of the information necessary for assessment calculations during the computer operation.

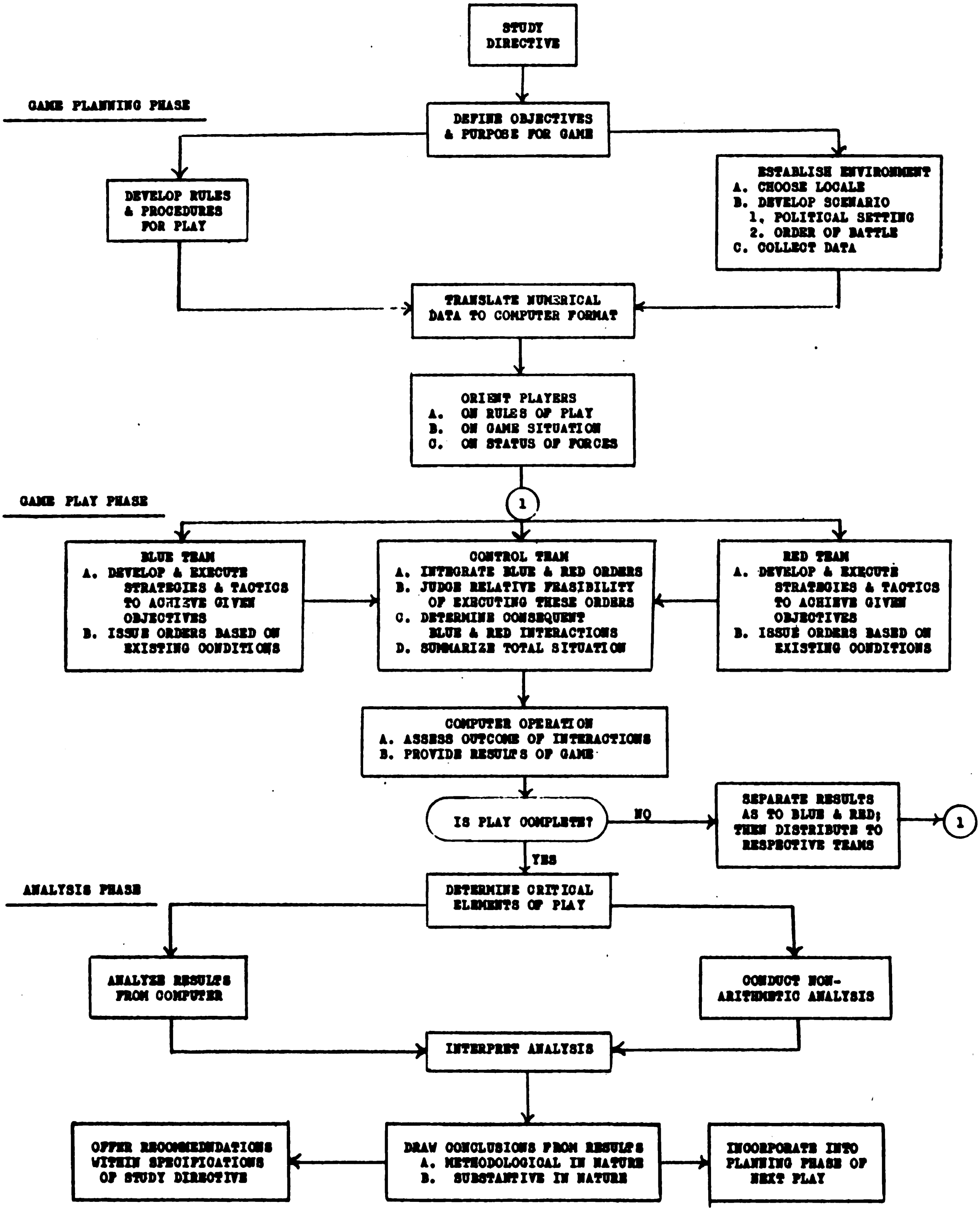


FIG. 1 - SEQUENCE OF EVENTS IN A SEMIAUTOMATIC GAMING SYSTEM

Chapter 2

<u>THE MECHANICS OF THE COMPUTER OPERATIONS.</u> The two basic purposes for using the computer in the semiautomatic system have been indicated in Chapter 1 - assessment and bookkeeping. The assessment function is accomplished through the application of various models, defined by the type of function they perform. For example, an air model assesses the interactions occurring during various phases of air operations, such as escort missions, interceptor missions, reconnaissance, interdiction, and the like; there will be as many models as there are well-defined, distinct assessment operations. The need to interconnect these models generates the requirement for some master program that provides the medium in which these models can operate. There is the further stipulation that this master program will be responsible for maintaining accurate and up-to-date records, with the provision for automatic changes to these records.

Thus it is the intent of this chapter to enable the reader to understand what are the requirements for a master program, its inputs, its operations, and its outputs. The objective is to describe these characteristics of the master program in a way conducive to other applications, i.e., so that others may find use for it.

<u>OBJECTIVES</u>. The specifications that are placed on the design of the master program are as follows:

(1) To require a minimum control effort in composing input to the computer.

(2) To establish an input format which is meaningful to control (a minimum of symbolism).

(3) To include the means for processing, routing, and storing data sets for use by assessment models.

(4) To allow for the operation of logically distinct models.

(5) To provide a method whereby accurate records may be maintained with the capability for their alteration.

(6) To enable results to be displayed in an understandable form.

<u>INPUT.</u> The input to the computer falls into three categories. There is that input which results from control definition of the combat interactions. There is also the status of forces file which includes all units being played in the game and their attributes. Finally, there are those inputs from control which do not result from any defined interaction but are changes to the status of forces file; these include such changes as increasing the number of men in a unit when reinforcements are introduced by the control group or specifying a new location when a unit is to have its assigned location changed. (These examples assume that strength and location are attributes of a unit and are recorded in the status of forces file.)

The basic principle involved in the input that defines the interactions is that all units participating in a given combat situation will comprise what is termed a "battle group", and the information for each battle group will be recorded on punched cards, one card per unit. All such units must be designated explicitly to be considered by the assessment models. In addition to naming the units, it is assumed that there would be certain factors included which describe the conditions of the battle and influence its outcome. Such factors as posture, terrain, and type of engagement might be included.

Each interaction defined as a separate battle group is processed as a separate engagement within the computer assessment of the outcome. The control group has the responsibility of specifying the different battle groups and parameters involved for each play; the master program maintains each as a separate entity in referencing the assessment models.

One of the fundamental elements in the system is the status of forces file. It is prepared initially during the pre-game planning phase by the control group; all relevant data for each unit to be played in the game are placed on standard forms, and they are then translated and processed onto magnetic tape. This is the only non-mechanized, or non-automatic, aspect of maintaining the status of forces file. It then serves as input to the initial interval of play, after which it is automatically revised consequent to the assessments of outcomes, and any new values for the characteristics of units are then incorporated into it. The characteristics of the units contained in the status of forces file are an integral part of the determination of the outcomes of the interactions. These characteristics are the factors plugged into the formulae of the models. The emphasis placed on the processing of these data will be seen later in this chapter.

The last type of input to the computer system is related to the status of forces file. As has been explained, the status of forces file exists on magnetic tape and is automatically processed and changed by the master program as a result of changes to unit characteristics as generated by the models. However the possibility for non-machine generated changes must be acknowledged. For this reason, provision is included within the master program to incorporate changes to unit characteristics issuing directly from the control group. Thus by control decision whole units, or parts thereof, can be eradicated or revised automatically as indicated by the changes recorded on punched cards.

DESCRIPTION OF THE MASTER PROGRAM. In Chapter 1 the principle of information flow was emphasized. In terms of the master program it is of equal importance; however in the medium of the computer the information assumes the form of data sets. The input information in "raw" form is organized by the master program into logically distinct data sets. The master program is then concerned with the ordering and storing of these data sets. When this has been accomplished, the master program references the relevant models which are to operate on the data sets. As changes to data pieces within sets occur, it is the responsibility of the master program to incorporate these changes into the data sets. Finally when all the changes have been affected, the master program provides the means whereby the revised data sets are edited and dumped as output from the computer. The master program is composed of a number of basic routines which enable it to accomplish these functions. There are six such routines:

(1) Read battle group cards.

(2) Select and store status of forces data.

(3) Reference models and adjust data.

(4) Edit assessment results.

(5) Update status of forces file.

(6) Edit status of forces file.

Each one will be discussed in terms of data sets with regard to the procedures to be followed in executing its operations, the input required, internally stored data necessary for execution, and the results of the operation.

The general flow of operations performed by the master program is presented by the flow diagram in Figure 2. The diagram is a much simplified one; the more specific details of the operation have been excluded. In determining what should be included in the diagram, the authors have attempted to present only those relationships such that to change them would, in effect, create a different program. It is felt that changes within any one of the individual boxes would not appreciably affect the over-all program; but to change the relationship among the operations illustrated would be such a significant alteration that it would be more advantageous to design a new program.

READ BATTLE GROUP CARDS ROUTINE. The master program "starts" by reading in the control information defining the battle groups, or combat interactions. This information has been recorded on cards punched in a specific format designed for the problem. Each card contains on it the designation of a unit involved in the interaction, in addition to parameters relating the unit to the battle situation. The data are extracted from the cards and are converted from the input code to the internal language of the computer. The process is continued until all the cards for the units being played during the present interval have been read. The names of these units are stored in a list which is to serve as a key for model routing. The control data for these units are then stored to be integrated at a later stage with the information extracted from the status of forces file. When all the battle groups have been read into the computer and processed in this fashion, the functions of the first routine have been accomplished, and the computer system is ready for the second routine to begin operation.

SELECT AND STORE ROUTINE. The select and store routine also performs an input function. This routine reads in the status of forces file (from magnetic tape). Contained on this file, as has been mentioned above, is a record of all the units and data describing these units. The routine in reading the file checks the name of each unit against the list of names made from the control input, and when a match is found, the data for this unit are extracted from the file.

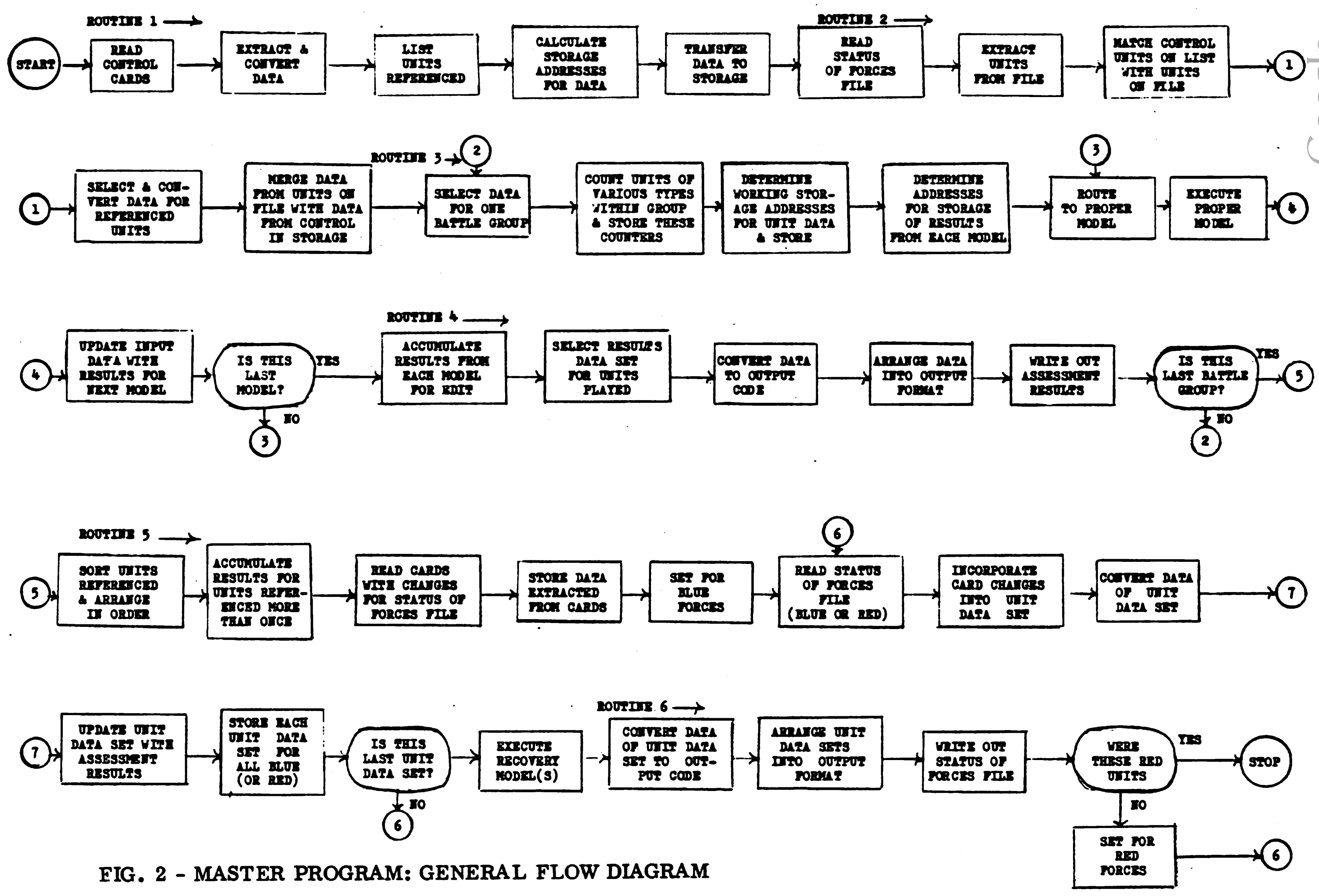


FIG. 2 - MASTER PROGRAM: GENERAL FLOW DIAGRAM

The data are then converted from the tape code to the internal language of the computer in the same fashion as were the card input data. Once the data have been extracted and converted, they are stored with the card input for the unit. The process is repeated until all the units contained in the control list have been matched with data taken from the status of forces file. When all such information has been stored, the storage region will be organized in the form of the following example (in consecutive machine cells):

Name of 1st Unit Specified	3rd Div
1st Control Input Parameter (posture)	Defend
2nd Control Input Parameter (terrain)	Flat
1st File Datum (location)	Berlin
2nd File Datum (strength)	9,000
3rd File Datum (armament - %)	100
etc., for all specified units.	

Each unit and its corresponding input data organized in this manner within the computer are referred to as a "unit data set". The remainder of the explanation of the computer system will be focused on the processing of this basic entity, this process being analogous to the principle of information flow in the non-automated portion of the system.

MODEL SELECTOR AND DATA ADJUSTOR ROUTINE. The central routine of the master program is the model selector and data adjustor program. The other routines of the system merely supplement the functions of this routine. Its purposes are to reference the appropriate model and to provide it with the unit data sets necessary for its calculations. To accomplish this the routine first selects the unit data sets comprising one battle group and transfers these to a working area. Next it determines what type (and how many) units are represented in the group; by doing this the routine is then able to determine what models should be called in to assess the outcome of the interaction. At this point a slight digression is warranted to make explicit the assumptions underlying this approach and what it requires.

The obvious premise is that the type of unit involved in an interaction entails what model should assess its effect on the outcome. Specifically it implies that a battle group composed only of air units, perhaps squadrons or wings, should be processed by an air model. This is obvious; however, what is not so clear is the procedure to be followed when the battle group is composed of a mixture of types of units, i.e., a battle group containing air, artillery, armor, and other dissimilar units. What procedure is to be applied must be decided early in the pre-game planning phase and requires what might be considered simply a delegation of responsibility - which models should assess what portion of the interactions.* The approach agreed upon by the control group is arbitrary as far as the master program is concerned; regardless of what decision is reached, however, some means of specifying the unit type is necessary. Thus the two requirements for the master program are that first a doctrine be defined, and second a means be provided whereby it is possible to differentiate between the types of units.

With this in mind the reader can now better understand the function of the master program to determine what types of units are present in the battle group. Before the models can be executed, however, there are still two operations which the master program must perform. It prepares a list of machine addresses, which are the first cells of each type of input unit data sets, and it also calculates the amount of storage necessary for results of the assessment and assigns storage addresses for this purpose. At this point the master program is ready to reference each model in turn in accordance with the procedure established in the planning phase.

The master program thus provides each model with the following four items: 1) the input data sets, 2) the addresses of the locations of these data sets, 3) the number of the various types of units within each battle group, and 4) the first addresses of the storage areas where the results are to be placed. After each model has assessed the interaction and has stored its results in the results region, the master program revises the input unit data sets with respect to these results, so that as each subsequent model operates, it is then provided with an updated data set. In this way there is established an interconnection between the various models of the system. This also demonstrates the importance attached to the procedure to be followed concerning the order in which the models are to be referenced. Since this is a fixed system, i.e., the logical order of the models never varies, emphasis should be placed on selecting that order which most nearly represents the usual sequence of events in reality.*

---

* See Chapter 3

* It is acknowledged that in reality sometimes events occur simultaneously; however, reality must be compromised to be made compatible with the fact that the digital computer operates sequentially.

After the models have assessed the outcome of a particular battle situation, the entire assessment operation is repeated for each of the remaining battle groups. At the completion of each cycle, the input data sets for the processed battle groups are discarded, while the data sets of results are stored for the later phases of the operation.

RESULTS EDIT ROUTINE. It is the function of the results edit routine to provide the output from the assessments. It first selects a unit data set of results. Next it converts the data into the output code, arranges them according to the output format, and writes them on magnetic tape. The results indicate all those items of the status of forces file which have been altered by the models and are the actual changes, not the result of these changes. For example, given an infantry division which has suffered heavy losses in combat, the results from this might be the number of casualties suffered to personnel and losses of equipment. The results output then consists of the name of the unit and changes to that unit, and these are given for each model and in total for all models. The routine continues in this manner until the results for all units played during the interval have been edited.

UPDATE ROUTINE. It was pointed out at the beginning of the discussion that the status of forces file was automatically maintained, and it is the function of the last phase of the system to accomplish this task. The first part of this operation is the update routine. The routine sorts all the data sets of results and arranges them in the same order as they appear in the status of forces file. This generates the requirement for a definite order for the units in the file. This could be done in either of two ways; either a list of the order of the units in the file could be stored within the routine, or the units could be arranged in some logical pattern in the file. The latter choice is the one incorporated in the computer system; it is assumed that all units are recorded in the file by number and that these numbers are in ascending sequence. Thus the routine is able to order all the data sets of results in ascending sequence to facilitate the updating process. While ordering these data sets, the routine checks for units have been referenced more than once. Where a unit does appear more than once in the data sets, the results are accumulated forming just one data set for each unit further facilitating the update process.

An auxilary function of the routine is to provide the capability for making changes to the status of forces file which are not machine generated, i.e., those that directly reflect a control decision. To execute this, it is possible to introduce such changes by punched cards. So that computer storage restrictions would impose no limitation on the number of units that could be changed in this manner, the card changes are read for only one unit at a time; the next set are read in after the first set of changes have been made. These changes, therefore, must be in the same order as the units on the file. Any datum for a unit can be changed except the identification number. The number of such data changes is unrestricted so that, for example, control could revise the number of personnel assigned to a unit to reflect a decision regarding reinforcements, or it could alter all the data attributes if necessary.

After the results of the model assessments have been ordered and a set of control changes for one unit read in, the routine begins to read in the status of forces file from magnetic tape. As each unit is extracted from the file, a check is made to determine whether any of its data attributes are to be replaced by those data of the control cards. (In the present system two cards are required per unit.) If there are any, the new data are substituted for the corresponding data comprising the file unit data set. Next the unit data set is converted to the internal language of the computer, and a check is made for the existence of any assessment results for the unit. When such results are present, the file unit data set is updated with this information, and the revised unit data set is stored within the machine. This process is continued until all the units for one side (Blue or Red) have been transferred from the file into the computer at which point the integration of all changes for these units from control and the models should have been completed.

The reason for storing all the data sets of just one side is to provide what is required for the execution of models that do not perform interaction assessment calculations, but rather that accomplish what might be called "recovery procedures". It is assumed that such models do not, therefore, require access to unit data sets for both sides; as a consequence only the data sets representing either Blue or Red units, respectively, are stored at any given time for these models. By this approach a more effective utilization of storage space is accomplished. (In the THEATERSPIEL application of the system, a logistics model was included at this point to perform consumption and resupply calculations for all units in the theater of operations.)

<u>OUTPUT GENERATOR</u>. After the model, or models, have been executed, the system has only to generate the output. Output is generated after each pass through the update-recovery portion of the system, i.e., after both the

Blue and Red units have been processed. This involves selecting, in turn, each of the unit data sets, converting the data into the output code, arranging the data sets into the output format, and writing this material on magnetic tape; in doing this the revised status of forces file is produced.

From the system then there results two forms of output: assessment results and a revised status of forces file. In addition all input to the system has been placed on punched cards. Thus all the quantitative material of the play exists in machine language. These three items can be retained for the work on the post-game analysis phase and provide the initial means whereby this analysis can be efficiently executed by the computer, thus accruing an important additional benefit from a computer supported gaming system.

CHAPTER 3

THE THEATERSPIEL COMPUTER SYSTEM. The computer system described in the previous chapter has been designed for the THEATERSPIEL Study (35.10, Strategic Division), for its POMEX series of war games. The development of the system was accomplished through the joint work of this study group and the Computing Laboratory staff; as such many of the decisions concerning critical aspects of this development reflect the efforts and decisions of both groups.

POMEX I was played during the latter part of July and the early part of August, 1961. In preparing for play (Phase I) and during play (Phase II) a great deal of attention was directed towards providing for the efficient employment of the computer system; at the same time much was learned in applying the system. It is the purpose of the present chapter to present some of the methods devised for the application of the computer system by THEATERSPIEL, and in the following chapter to give some indication of what one computer-oriented experience has gained for the study group, so that other studies with a similar orientation may benefit from this first attempt.

PHASE I: THEATERSPIEL COMPUTER USAGE PREPARATIONS. It was decided that the computer-oriented objectives for the play of POMEX I would be to mechanize four separate models: an air model, a support weapons model, a ground combat model, and finally a logistics model. The substance of the models, coupled with the over-all objectives of the study, determined the level of aggregation of play, i.e., the amount of detail to be included. As a consequence the size of units to be played was that of division level. Further the choice of the particular theater to be played affected the decision as to what types of units were played. Finally the data required by the models for each unit determined what characteristics were used to describe the units. It was in this manner that the specifications placed on the design of the status of forces file became evident.

The preparation of the status of forces file involved several stages of development. First a unit designation system by which the units could be identified was devised consistent with the various classifications of units to be considered in the game; it was based on the use of five digits, the pattern and use of which is illustrated in Table 1. It can be noted that there can be no more than 100 units of any given type and nationality with the use of this symbolic system, however the addition of one digit would increase this number to 1,000 and would remain compatible with the system in its present form. The unit designation number as such was used throughout the game by both the player teams and the control group when referring to specific units. For map purposes, only the last two digits were used on the unit symbols. However the color and shape of the symbol determined the rest of the designation, and thus identification of units was accomplished.

After the scheme for unit designations had been developed, it was necessary to specify what characteristics would form the various unit data sets for each type of unit. Since each model required certain pieces of data for each unit, all that was needed was to accumulate these requirements. To be kept in mind in doing this, however, was the goal of compactness, i.e., where possible to have the data pieces serve more than one model's needs. Table 2 contains the result of this effort. As can be seen, there are six unit types represented for which in most cases the characteristics are the same.

Table 1

EXPLANATION OF UNIT DESIGNATION NUMBER

First Digit: Allegiance

B = Blue
R = Red

Second Digit: Type

0 = AIR	Air unit
1 = SAM	SAM unit
2 = SPT	Support Weapons unit
3 = GND	Ground Combat unit
4 = LOG	Logistics unit
5 = STN	Supply Point

Third Digit: Nationality

0 = US	United States
1 = BR	Great Britain
2 = FR	France
3 = WG	West Germany
4 = BE	Belgium
5 = NE	Netherlands
6 = SR	Soviet Russia
7 = EG	East Germany
8 = PO	Poland
9 = CZ	Czechoslavakia

Fourth and Fifth Digits: Identification Number

As desired

EXAMPLES:

B1234 =	34th French SAM unit
R0601 =	1st Soviet Air unit

AIR UNITS	SAM UNITS	SUPPORT UNITS	GROUND UNITS	LOGISTICS UNITS	SUPPLY UNITS	
Name of Unit	Name of Unit	Name of Unit	Name of Unit	Name of Unit	Name of Unit	
Unit Desig	Unit Desig	Unit Desig	Unit Desig	Unit Desig	Unit Desig	
Location	Location	Location	Location	Location	Location	
Activity	Activity	Activity	Activity	Activity	Activity	
Road In	Road In	Road In	Road In	Road In	Road In	
Priority	Priority	Priority	Priority	Priority	Priority	
Max Input Cp	Max Input Cp	Max Input Cp	Max Input Cp	Max Input Cp	Max Input Cp	
Pres Inpt Cp	Pres Inpt Cp	Pres Inpt Cp	Pres Inpt Cp	Pres Inpt Cp	Pres Inpt Cp	
Tons/100 Men	Tons/100 Men	Tons/100 Men	Tons/100 Men	Tons/100 Men	Max Sup Strd	
Pres Str	Pres Str	Pres Str	Pres Str	Pres Str	Prs Sup Strd	
Prior Str	Prior Str	Prior Str	Prior Str	Prior Str	Road Out	1
Auth Str	Auth Str	Auth Str	Auth Str	Auth Str		2
OH Supply 1	OH Supply 1	OH Supply 1	OH Supply 1	OH Supply 1		3
2and4	2and4	2and4	2and4	2and4		4
3	3	3	3	3		5
5	5	5	5	5		6
Airfld Cap		Weapons	Combat Pot			7
Numb Sorties	Prev Firings					8
						9
						10

For example, each unit data set except for supply points, which are a special type of "unit", has a characteristic labeled "pers str" which is the abbreviation used for "present strength"; this has a different meaning depending on the type of unit described; for air units it describes the number of planes, for SAM units the number of launchers, and for the rest the number of men. This demonstrates what was achieved in striving for compactness and applies to many of the other characteristics. An explanation of the meaning of the other abbreviations can be found in Appendix A.

In the process of preparing the status of forces file the next step was to design a suitable format. It has been previously indicated that the file exists on magnetic tape, however the actual working file, which the player teams and the control group use, is the listing made from the magnetic tape on the High Speed Printer; the format was designed with this fact in mind considering the size of the sheets of printer paper, the amount of material on the status of forces file, and the clarity of presentation. The result was a printer page containing, at most, eighteen unit data sets arranged in two tables of nine units each with the data pieces of each unit data set placed vertically with respect to one another. An example of this format is provided in Table 3.

For the other input to the computer two forms were used. Table 4 provides a sample card format for the battle groups. Sheets similar to this one were filled out by the control group to be key-punched on cards to be read into the machine. The first six columns of the card contained the unit designation number; the next six were used to specify the percentage of the unit being employed in the case of air units or, in the case of ground units, to indicate the type of terrain in which the engagement was to take place; the third set of six columns was used to specify the type of engagement being fought, and the last six either the target for an air unit or an alternative location for a ground unit. The remaining fifty-six columns were reserved for comments, and although these were not a necessary part of the computer input, they were used by the control group to supplement the records. The repetition of the number six is significant with regard to the input and output of the system, and there is a reason for it. The input-output is written in specific code, or language, which is required by the High Speed Printer and which permits the intermixing of alphabetic and numeric characters; as such it is only necessary for working with the magnetic tapes (the status of forces file and the assessment results), but the consistency it was decided to employ this same code when using cards. In this way the entire input-output medium is written in the one language; all pieces of data included can consist of no more than six characters due to translation of these data pieces into internal machine language and the given work size within the computer.

This same idea applies to the changes to the status of forces file made by the control group. These changes are punched on cards, and each six columns on the card corresponds to one data piece of the unit data set. Since each unit is described by some 20 data pieces, two cards are used per unit (columns 79 and 80 are omitted). For example, columns 13 through 18 would contain whatever new data piece that should replace the third data piece of the unit data set (counting the unit designation number as the first). Referring to Table 3 suppose that control desires to change the road in to the Russian armored division designated R3601, a card would be prepared with the designation punched in columns one through five and the new road in number, 512, in columns 16, 17 and 18. The revised status of forces file for D+9 would then reflect the change .

In addition to designing the input formats, one output format had to be prepared for the results of the assessments. A sample format can be seen in Table 5 in which the results for three units are shown. For each unit there are contained both its name and designation number, the date, its new location (if it has been moved, as each of the three have), and the various losses of the unit. The casualties are given in four columns; the first three show the losses calculated by the assessment models (air, support and combat) and the fourth the total casualties. Each page contains units of only one side, either Blue or Red; this is done so that the results can be separated and distributed to the respective player teams. Every unit specified within the battle groups in any given interval of play will be included on the results sheets in this fashion.

During Phase I the four models were programmed and integrated with the master program. A great deal of care was taken in the coordination of this task to insure internal consistency within the resulting computer system. Each model was written as a separate program which, when finally a part of the complete system, could be entered from the master program and exited as an internally logically distinct and independent program, requiring only the input data external to itself, i.e., the unit data sets, the initial addresses thereof, and the initial addresses for the results storage. As this programming effort and the data preparation were completed, attention was directed to determining what procedures would be followed in the execution of play.

SR GND STATUS REPORT POMEX I D + 8

NAME OF UNIT	1 TK DIVISN	2 TK DIVISN	3 TK DIVISN	4 TK DIVISN	5 TK DIVISN	6 TK DIVISN	7 TK DIVISN	8 TK DIVISN	9 TK DIVISN
UNIT DESIG	R3601	R3602	R3603	R3604	R3605	R3606	R3607	R3608	R3609
LOCATION	PE11	NA44	NB53	PD49	PD44	PD47	PC08	NB80	NB94
ACTIVITY									
ROAD IN	411	604	414	51	194	52	294	419	418
PRIORITY	1	1	1	1	1	1	1	1	1
MAX INPUT CP	7649	7649	7649	7649	7649	7649	7649	7649	7649
PRES INPT CP	5480	4297	4133	2132	2231	3065	4265	1799	4245
TONS/100 MEN	372	372	372	372	372	372	372	372	372
PRES STR	10524	9434	7881	9993	6214	7146	5330	5609	8777
PRIOR STR	10524	10524	7808	10252	6439	7262	5330	6378	8777
AUTH STR	10524	10524	10524	10524	10524	10524	10524	10524	10524
ON SUPPLY 1	96	85	71	96	56	65	48	51	79
2AND4	43411	38529	31979	40813	25381	29228	21555	22906	35501
3	1641	1602	1330	1697	1056	1215	896	953	1471
5	3684	3269	2914	3495	2154	2480	1829	1943	3013
COMBAT POT	495	398	181	424	91	188	86	63	296

NAME OF UNIT	10 TK DIVISN	1 MTRZ DIVISN	2 MTRZ DIVISN	3 MTRZ DIVISN	4 MTRZ DIVISN	5 MTRZ DIVISN	6 MTRZ DIVISN	7 MTRZ DIVISN	8 MTRZ DIVISN
UNIT DESIG	R3610	R3620	R3621	R3622	R3623	R3624	R3625	R3626	R3627
LOCATION	PC15	NE71	PE31	PD00	NA76	NA78	NA79	NB92	NB91
ACTIVITY									
ROAD IN	192	53	54	191	648	644	649	416	412
PRIORITY	1	1	1	1	1	1	1	1	1
MAX INPUT CP	7649	5044	5044	5044	5044	5044	5044	5044	5044
PRES INPT CP	3221	5030	5032	4080	1219	3028	3351	3798	3850
TONS/100 MEN	372	180	180	180	180	180	180	180	180
PRES STR	3746	6111	11064	4215	6531	6821	6063	8684	11190
PRIOR STR	3746	6111	11064	4215	7403	7561	6834	8684	11190
AUTH STR	10524	13018	13018	13018	13018	13018	13018	13018	13018
ON SUPPLY 1	34	56	109	38	59	62	64	78	100
2AND4	15151	11953	21640	3246	12901	13476	13960	16985	21888
3	630	698	1264	482	654	688	815	892	1279
5	1287	1466	2653	1011	1582	1652	1711	2082	2683
COMBAT POT	7	0	342	1	36	40	45	342	354

TABLE 3

SAMPLE STATUS OF FORCES FILE PAGE

DESIGNATION NO. OF UNIT

PERCENT OF UNIT ATTACKING (AIR) OR TERRAIN FACTOR (GROUND)

TYPE OF ENGAGEMENT

OBJECTIVE: TARGET (AIR) OR NEW LOCATION (GROUND)

COMMENTS, IF ANY

	DESIGNATION NO. OF UNIT (1–6)	PERCENT OF UNIT ATTACKING (AIR) OR TERRAIN FACTOR (GROUND) (7–12)	TYPE OF ENGAGEMENT (13–18)	OBJECTIVE: TARGET (AIR) OR NEW LOCATION (GROUND) (19–24)	COMMENTS, IF ANY (25–80)
1	B0001	100	4	R3608	BLUE FTRBMR WING INTERDICTS RED ARMORED UNIT
2	B3001	1	2	AB55	RED ATTACKS BLUE INF IN PREPARED POSITION
3	R3608	1	12	AB23	RED ADVANCE TO AB23 IF SUCCESS
4	END GROUP				ALL UNITS OF INTERACTION INCLUDED
5					
6					

TABLE 4

SAMPLE CARD FORMAT FOR BATTLE GROUND INPUT

CASUALTY ASSESSMENTS POMEX I

	8 TK DIVISN		R3608	
	NEW LOCATION		NB80	D + 8

CAUSE OF CAS	AIR	ARTY OR SSM	GROUND	TOTAL
INPUT CAP	2365	0	0	2365
PERSONNEL	163	131	539	833
COMBAT POT	7	6	23	36

	4 MTRZDIVISN		R3623	
	NEW LOCATION		NA76	D + 8

CAUSE OF CAS	AIR	ARTY OR SSM	GROUND	TOTAL
INPUT CAP	1503	0	0	1503
PERSONNEL	166	152	628	946
COMBAT POT	3	3	13	19

	5 MTRZDIVISN		R362[illegible]	
	NEW LOCATION		NA78	D + 8

CAUSE OF CAS	AIR	ARTY OR SSM	GROUND	TOTAL
PERSONNEL	N	159	656	815
COMBAT POT	O	3	15	18
	N			
	E			

TABLE 5

ASSESSMENT RESULTS FORMAT

PHASE II: THEATERSPIEL GAME PLAY. The Blue and Red player teams were each provided with a copy of the game scenario, their respective status of forces file, and large maps of the theater of operations depicting the distribution of their units. Using these three sources of information, the player teams developed their tactics and communicated the orders to execute these tactics to the control group. This was accomplished in two ways. Large acetate overlays indicating the movement of the units by the players were transferred into the control room; this was supplemented by written material stating the players' general objectives for the current interval.

After the overlays from both player teams had been merged with the control map, the control group proceeded to define the resulting interactions. The assistant controller for logistics judged the extent to which the (player) indicated moves could be executed within the given time interval of play, based on such factors as availability of transport and troop readiness. In the process of thus advancing the units, both Blue and Red, it was the responsibility of the controller and assistant controller for ground combat to determine in what cases combat interactions would result and what units would be involved in the action. Figure 3 displays a general diagram of the combat interactions defined on the basis of indicated movement; Red (darkened symbols) had indicated a thrust over the MLR (main line of resistance) thereby placing his units in the proximity of the fixed Blue positions. This results in two separately defined interactions, the engagements being grouped with the assumption that there will be no direct interaction among the units of the two different battle groups. Next the assistant controller for air operations committed air units for support of the respective forces, in addition to stipulating any air interactions which were to take place, both of these actions reflected the decisions of the players with respect to their orders for the employment of air units.

As each of the battle groups was thus defined, the information was recorded on the control input sheets, each assistant controller being responsible for filling out his own sheets. The input sheets were then given to the assistant controller for machine operations who was responsible for making a comprehensive check of all input sheets. Proper designation of units, consistent interaction numbering, and completeness were among the matters checked. After the examination of the input sheets was completed, the information recorded was punched on cards which were then checked and sorted for errors, and finally the computer system was run.

One set of output from the computer was separated to include only Blue results and one only Red results and distributed to the respective player

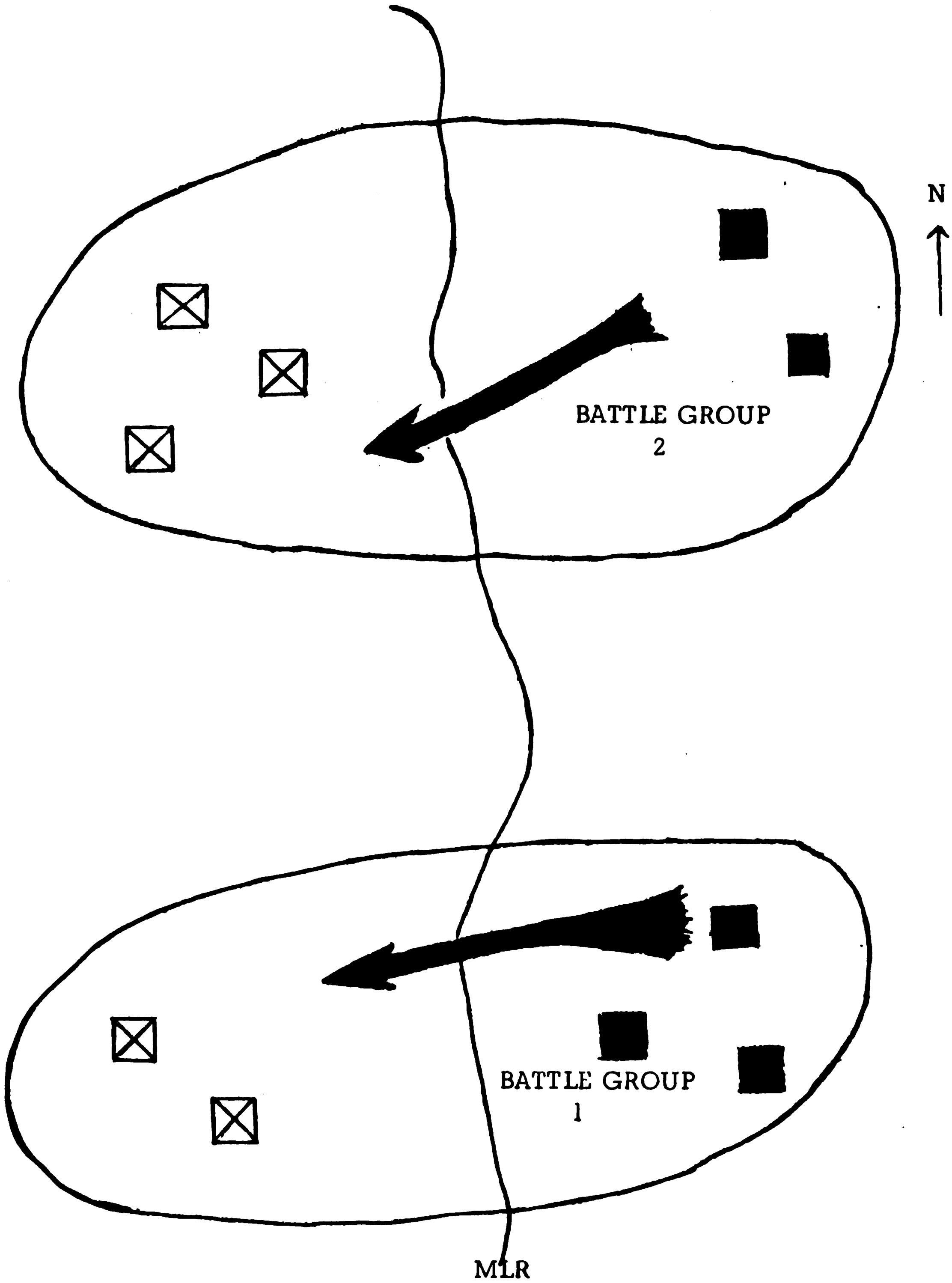


Fig. 3 - Battle Groups Defined

teams. Duplicate sets of output were distributed among the members of the control group, who then composed reports to supplement the machine record. At this time, too, the assistant controller for intelligence used the results to generate the intelligence estimates provided for the players.*

The play continued in this manner with the players each time using the revised status of forces file together with the other information given them by the control group to generate their new set of orders, until the controller judged that the combat had achieved that degree of resolution which had been initially desired, at which time the game was ended.

This short summary of the THEATERSPIEL computer play of POMEX I has been included to serve as an indication of how the play of a semiautomatic system can be implemented; in addition it has been written to provide the background for a discussion of some of the problems realized in such an attempt. Those readers more interested in the substantive aspects of play are referred to the paper which has been written summarizing the history of play of POMEX I.[7]

* In the first play of the system the intelligence model had not been programmed, and thus this operation was performed manually. It is anticipated, however, that it will be programmed and incorporated into the computer system for the next play.

Chapter 4

SOME CONSIDERATIONS ON THE SUBJECT OF COMPUTER USAGE

EFFECTS ON GAME ORGANIZATION. The effects on organization are felt in several ways. First it becomes necessary to define precisely the rules by which play is to be conducted. The rules must be well defined so that the programming can reflect these rules; since the nature of programming is the expression in symbolic language of a logical progression of operations, the rules must be stated in a form conducive to the accomplishment of this task. Secondly the procedures to be followed in implementing the computer system during play must be equally explicit. There must be a specific delegation of responsibility within the control group, and each member of this group must understand at least the fundamental principles of the operation of the computer and of the specific programming involved. This is of importance if the maximum potential of the computer is to be realized. If the operation of the system is to proceed efficiently, the members of the control group should have a working knowledge of the programs used. Superficially this may not appear to be very necessary, however during the progress of play many unexpected problems will arise, and the individuals concerned must be prepared to cope with these rapidly. Furthermore an attempt should be made to provide simple and concise forms for each step of the control operations in order to avoid delays arising from errors and from misunderstandings.

There are two further implications which, broadly speaking, can be considered a part of the organizational aspects of a semiautomatic system. As the computer system increases in complexity, it becomes more difficult to revise it. However if its development proceeds in a logical and orderly fashion, this will tend to alleviate the severity of this problem. Care must be taken to avoid the creation of a "black box" which becomes unmanageable. Finally it must be realized that the use of the computer introduces new responsibilities into the control room. One reason for using the computer is to absolve the control group of many repetitive and tedious functions usually associated with control procedures in a hand-played game. Nevertheless in diminishing the magnitude of these efforts, it is possible to create new and more tedious difficulties related to the use of the computer, since it creates the requirement for high standards of accuracy.

THE QUESTION OF ACCURACY. The matter of accuracy in working with a computer is two-sided. First information prepared for the computer must attain high standards of accuracy. Errors made in preparing input for the computer can cause the system to fail in its operation, resulting in unnecessary delay, or the errors can go unnoticed with the consequence that they are perpetuated into successive stages before they are detected. Here again proper organizational procedures can alleviate the difficulty; it must be realized, however, that the final responsibility in this area lies with the personnel of the control group, further emphasizing the need for their proper understanding and knowledge of the system.

Secondly the computer provides the means for greater reliability in the accuracy of the results of the game. (This is not to be confused with validity.) Once the programs have been properly checked out, there need be no concern for errors in the computations. Furthermore the speed and capacity of the computer allows for more comprehensive calculations. Not only does the computer enable the system to include more elaborate methods of calculation that would be infeasible when performed by hand, but it also allows many more factors to be considered, and in greater detail. Of course, this opportunity should not be unnecessarily exploited; it is possible to design a system which provides too much detail. If a player team is given an overabundant amount of information including much irrelevant material, some of it will tend to be ignored and will be of no use. Another aspect to be treated in a cautious manner is the temptation to compromise the game rules, or the corresponding calculations within the models, to adjust to the requirements of the computer. Frequently times will occur in which certain calculations present a problem in their translation to a programmed sequence. In resolving the difficulty the programmer must avoid an arbitrary compromise for the sake of programming clarity. In addition to this, certain operations may arise for which the programming approach is not immediately evident. Approximating the operation must be done with an appreciation for the error introduced; consideration must be given to the fact of whether or not this error will tend to cancel out during the remainder of the calculations or be intensified. From this discussion the reader should be aware that the problem of accuracy can work for or against the system, although it will generally be positive factor if the proper attitude is adopted when designing the system with regards to organization and procedures.

In summary, there is a certain danger to be avoided in the use of a computer in the gaming environment. It must be remembered that the function of the computer in this environment is to support the control operations. If the computer receives too great an emphasis, its very advantage can be turned into a detriment with too little thought being afforded to substance and too much to method. Yet if concentrated attention is directed towards the design of the system in the pre-game planning phase, the degree to which the game play becomes subordinated to the computer operation is reduced to a level at which the efficient relationship between man and computer is achieved.

APPLICATIONS. The system described in this paper could be used to provide a satisfactory approach to any gaming environment similar to the one outlined in Chapter 1. This is one in which there is a basic concern for the quantitative assessment of the interaction among some fundamental entities, and one in which there is a need for the consideration of a great number of these entities and quantitative factors describing them. Further there should be a requirement for logically distinct operations to be performed in the calculation of these interactions which must be repeated a sufficient number of times to warrant their mechanization. Finally there should be a desire to maintain the separation of the human decision functions and the quantitative analysis resulting from these decisions, and yet the desire to maintain the interrelationships involved.

If these conditions are met, then the semiautomatic system discussed could be utilized in any one of three ways. At the first level of utilization, the method of approach might be applied to other studies, thus substantially reducing the necessary planning effort involved. That is, it could be applied as a logical system. At the second level of application, the interpretation of this logical system, the master program, could be adapted with a few slight revisions to other systems into which the pertinent models could be incorporated. In this case all that would be necessary would be to construct these models in a fashion consistent with the requirements of the master program maintaining the basic operations outlined in Chapter 2. Finally the third level of utilization would be one in which the whole system would be used in toto including the models programmed by THEATERSPIEL. This, however, requires a more detailed understanding of the substance of the system and entails reference to the papers describing these models.[5,6,7]

LIMITATIONS. There are two major limitations to be considered when discussing the feasibility of this approach to war gaming. There was some mention made in Chapter 2 of the problem of storage within the computer and in the first section of this chapter of the difficulty involved in making changes to the routines as the programming becomes progressively more complex. The latter is by far the more important. For example, if it were desired to revise and improve one of the models, the changes necessary would more than likely affect the other models. This snowballing effect would vastly increase the time required to make the alteration. The greater the departure of the new requirements from the original ones, the more difficult will be the task of adjusting the system to suit these new requirements.

The problem of internal storage can be solved, although at the expense of operating efficiency. Extensive use of magnetic tapes can provide an almost unlimited storage capacity for the system. In doing so, however, it must be realized that the speed with which the operations can be accomplished will consequently be compromised, since tape storage, as opposed to internal storage, has a much slower access time for computational purposes. As presently designed the THEATERSPIEL semiautomatic system's use of tape storage is minimal with the result that one complete computer run of the system is accomplished in about fifteen minutes. It can be anticipated that a significant increase in the use of tape would double or triple this time. Moreover the reliance on greater usage of tape storage would entail revising the logic of the presently programmed system which in itself would require some time (in the order of months) to accomplish.

EVALUATION. In general there are two major criticisms made of war gaming as a means of problem solution; one is with respect to timeliness and the other with regards to cost. It is the purpose of this section to demonstrate that in terms of these two criticisms, the semiautomatic system is a significant advancement in the state of the art. There are frequent references in the literature about the expense gaming entails, and furthermore that it is such an extensive and time-consuming operation that by the time the game is finished and the study completed, there is no longer a requirement for the results. THEATERSPIEL's first play with the system in POMEX I would tend to support the view that by the use of the semiautomatic system, this need not be the case.

After the initial period of familiarization and orientation to the use of the system the last 10 intervals of play of POMEX I were completed over a span of 2 weeks. It is estimated that to play a hand game with a similar degree of complexity and detail which the use of the computer permitted would require approximately 3 weeks of real time for each interval of game play. Assuming in both cases the need for the full-time efforts of 10 analysts, together with any necessary additional support, the cost per game interval if played by hand would be about $30,000; whereas using the semiautomatic system, the cost per game interval would be about $2,100.

Thus the total cost for 10 intervals played by hand would be $300,000 and would take about 30 weeks, whereas with the semiautomatic system the 10 intervals were completed in 2 weeks (10 work days) at a total cost of $21,000. Of course, this description is incomplete unless the time spent in developing the system is considered. The preparation of the computer gaming system took about 6 months; this estimate includes design planning, data collection, programming, debugging and education. This information is summarized in Table 6.

Table 6

TIME AND COST OF HAND PLAY VERSUS THE SEMIAUTOMATIC SYSTEM

	Hand-Played Game	vs	Computer-Assisted Game
Real Time/Game Interval (Wks)	3.0		0.2
Estimated Cost/Interval	$30,000		$2,100
Time for Preparation and Play of POMEX I (Wks)	30		28
Cost of POMEX I	$300,000		$294,000

But this table too is incomplete, in two ways. First the estimate of time for preparation to play the game by hand should be revised with the consequent effect on cost. Much of the same work required in preparation for play with the computer-assisted system would also be required in preparing for hand play. The same data would have to be obtained, the same models prepared for use (though in a different form), and many of the same procedures would have to be devised. Thus in effect this would increase substantially the time and cost estimates for a hand played game.

The second area in which the presentation of Table 6 is deficient is with respect to the ideal of future plays. The saving is demonstrated most dramatically when considering future uses of the two approaches. Speculating as to a future play the results might appear as presented in Table 7.

Table 7

TIME AND COST FOR A TWENTY INTERVAL FUTURE PLAY

	Hand Played Game	vs	Computer Assisted Game
Real Time/Game Interval (Wks)	3.0		0.2
Estimated Cost/Interval	$30,000		$2,100
Time to Play POMEX II (Wks)	60		4
Cost to Play POMEX II	$600,000		$42,000

From this it can be seen that each repeated usage of the system will increase the practicality of its development. However, this also demonstrates the fact that where only one game play with the system is desired, the merits of the computer-assisted game are not sufficiently obvious. Nevertheless each future use of the system adds to the merit of its original design and development.

There is another significant difference to be noted between these two approaches which has a bearing on the criticism of timeliness. It has been pointed out earlier in this paper that the post-game analysis phase can be made far more efficient and profitable if the computer is put to good use. This is especially true when the computer-assisted system has been used, since all the data to be analyzed already exist in machine form on punched cards and magnetic tape. Thus the problem of organizing the game data for analysis purposes can greatly be diminished by the use of the computer-assisted system.

In conclusion, the advantages gained during the game play phase and post-game analysis phase with regards to time and cost would seem more than to compensate for any lengthening of the pre-game planning phase. Finally in those cases where a number of plays of the system are desired, the computer-assisted game seems to be a vast improvement over the hand-played game in terms of both time and cost.

Appendix A

TABLE OF MEANINGS FOR UNIT ATTRIBUTES

NAME OF UNIT:	Actual name of unit.
UNIT DESIG:	Designation number of unit (explained in text) used in game system.
LOCATION:	Geographic position of unit.
ACTIVITY:	Mission of unit in given interval of play.
ROAD IN:	Number indicating supply point from which supplies are to be obtained.
PRIORITY:	Number indicating relative importance of unit in obtaining supplies.
MAX INPUT CP:	The upper limit (or original value) of a unit's capacity to receive supplies.
PRES INPT CP:	Capacity of unit to receive supplies during given interval.
TON/100 MEN:	Total authorized weight of unit per each 100 men.
PRES STR:	Number of men in unit available for combat at end of given interval.
PRIOR STR:	Number of men in unit at beginning of given interval.
AUTH STR:	Original number of men assigned to unit.
OH SUPPLY I:	Total weight (in tons) of Class I supplies unit has available.
II AND IV:	Total weight (in tons) of Class II and IV supplies unit has available.
III:	Total weight (in tons) of Class III supplies unit has available.

OH SUPPLY V:	Total weight (in tons) of Class V supplies unit has available.
AIRFLD CAP:	Number representing the capacity of an airfield to expedite air operations.
NUMB SORTIES:	Number of planes flown during last interval.
PREV FIRINGS:	Number of missiles launched during last interval.
WEAPONS:	Index of combat value for support weapons units.
COMBAT POT:	Index of combat value for ground combat units.
MAX SUP STRD:	The upper limit on amount of supplies a supply point can store.
PRS SUP STRD:	The amount of supplies a supply point has stored during given interval.
ROAD OUT: 1-10	A number indicating which units can obtain supplies from given supply point.

REFERENCES

1. Operations Research Office, "FAME: A War Game for Testing Division Organizations (U)," ORO-T-383, Dec 60. SECRET-RESTRICTED DATA.

2. Von Neuman, John, The Computer and the Brain, Yale University Press, 1958.

3. Operations Research Office, "SANDWAR I and IA, A Theater Level War Game in Iran (U)," RAC-TP-47, October 1961. CONFIDENTIAL.

4. Weiner, Nobert, Cybernetics, The Technology Press, 1949.

5. Research Analysis Corporation, "Game Models for Tactical Air Operations in Theater War Games (U)," RAC-TP-52, December 1961. SECRET.

6. Research Analysis Corporation, "Logistics Model in POMEX I," Draft TP in preparation.

7. Research Analysis Corporation, "A Developmental War Game Set in Europe Using Computer-Assisted Assessments (U)," Draft TP in preparation.

8. Defense Atomic Support Agency, Joint War Gaming Feasibility Study, Contract No. DA-49-164-SZ-077, Report No. TO-B61-23, 28 April 61.

TRANSIENT NUCLEAR RADIATION EFFECTS ON ELECTRON TUBES AND TRANSISTORS

Richard G. Saelens
U. S. ArmySignal Research and Development Laboratory
Fort Monmouth, New Jersey

INTRODUCTION. The magnitude of the nuclear radiation effects program at USASRDL has grown considerably during the past year. Although considerable exposure-type investigations have been performed at other facilities, approximately one and one-half years have elapsed since a Godiva-type reactor was available. This paper will describe a large-scale exposure at the Godiva reactor. The magnitude of this effort was made possible by the 16-month delay. By capitalizing on the interim period between experiments, a program was designed to permit the exposure of approximately 600 electron tubes and solid state devices. The objective of this experiment was to obtain information on electron devices which could be utilized by electronic equipment designers. In order to provide these data, sample sizes were selected which would provide statistically valid information. Although some question may arise as to the sample sizes chosen, other factors including economics and availability of exposure space and test equipment were also considered. Approximately 20 types of transistors were exposed (sample size approximately 25 each), and 10 types of electron tubes (sample size approximately 10 each). The experiment design included data on controls not exposed to the radiation environment, data acquisition on other environments, parameters which could possibly have an effect, such as ambient temperature, and controls over as many conditions as possible. In addition, controls were exercised over the operation condition and capability of the test equipment. The experiment was performed in late August 1961 instead of May 1961 as originally planned and, therefore, the complete analysis of the vast amount of data could not be accomplished. The data obtained during this experiment are still in the process of being reduced. Complete data analysis is being performed at USASRDL, and also will be performed under contract.

RADIATION EFFECTS MOBILE LABORATORY (REML). The magnitude of this experiment necessitated the design and instrumentation of a special mobile laboratory. This consisted of a 28' trailer equipped with energizing circuitry, FM tape recorders, oscilloscopes, oscilloscope cameras, temperature recorders, and ambient temperature controls. The interior of the trailer showing the magnetic tape recorders and electron tube circuitry can be seen in Fig. 1. (Figures are at the end of this article.) The temperature recorder, transistor circuitry, power supplies, and digital voltmeter are shown in Fig. 2. A special rack designed to accommodate six oscilloscopes is shown in Fig. 3. In order to minimize circuit changes between exposures, printed circuit boards were utilized for the electron tubes. Through the proper selection of dropping resistors, the correct voltage is applied to each tube type without changing the supply voltage (Fig. 4).

INSTRUMENTATION. The recording instrumentation of the REML consisted of five 14-channel high-speed magnetic tape recorders and four dual-beam oscilloscopes. The oscilloscopes were used to simultaneously monitor certain tape inputs. This information was used to verify the validity of the tape data, and provided immediate information at the site on the transient changes occurring in the devices under test.

A block diagram of a typical exposure during the experiment is shown in Fig. 5. Cables from the exposure head are connected to a junction box placed inside the Kiva (building housing the reactor). Typical exposure heads are shown in Figs. 6, 7, 8, and 9. Note the sulphur pellets mounted directly in the potting material on the transistor test heads. The junction box is shown in Fig. 10. Cables from the junction box feed into a patch panel on the side of the trailer and are, in turn, connected to the tube and transistor circuitry. Transistor outputs are fed into differential preamplifiers before they are recorded on magnetic tape.

A digital voltmeter with printout was utilized in the transistor monitoring system. Through a series of stepping switches, it is possible to monitor 100 channels of information. The stepping rate is 0.5 sec. The information recorded in digital form is used for pre- and post- radiation measurements.

EXPOSURE OF DEVICES.

a. Solid State Devices:

The following is a list of solid state devices which were exposed during the experiment:

Special Computer Devices: Germanium 2N1304, 2N1305, 2N1306, 2N1307, 2N710, 2N101, 2N100

Silicon Switching Devices: 2N706

Power Transistors: Germanium 2N797, 2N1309, 2N1043, 2N1046
Silicon: 697

General Purpose Devices: Germanium 2N1406

Epitaxial Devices: 2N743, low signal silicon 2N335 and 2N726

Several diodes were also exposed, such as the 1N752, XR-39, 1N652, and a GaAs Varactor diode. Parameters were monitored on all devices during and after the radiation pulse.

b. Electron Tubes:

The following electron tube types were exposed: 2582 and 7457 power tubes, Nuvistor triodes and tetrodes, 6J6, 12AT7, 6AQ5, 6943, 7244, 2146, 1724, and a Z-2352. Plate current, I_b, grid leakage current, I_c, and a-c signal gain, e_p, were measured on these devices before, during, and after the radiation pulse.

SANDIA PULSE REACTOR FACILITY (SPRF). Briefly, the characteristics of the SPRF are as follows:

60 μs pulse width at half height

3×10^{16} leakage neutrons/burst

3×10^{20} neutrons/sec peak leakage rate

10^{13} n/cm^2 E > 1 Kev

2×10^{17} n/cm^2. sec peak intensity

2000 - 3000 rads gamma dose

2-3 x 10^7 rad sec

DOSIMETRY. Fast neutron dose measurements were made with S^{32} pellets. S^{32} pellets from USASRDL were exposed in pairs during several shots in order to obtain corroboratory information. Additional neutron dose measurements were made with Pu^{239} ($E > 1$ Kev), Np^{237} ($E > 0.7$ Mev), and U^{238} ($E > 1.5$ Mev) foils furnished by Sandia. Gamma dose measurements were made with USASRDL NBS film-badge type dosimeters and microdosimeters (glass rods). Gamma dose rate was measured with the MgO-RAD* and SEMIRAD.

DATA ANALYSIS. The complete plan of analysis requires the data to be treated as follows:

Transfer the information from the magnetic tapes to a visual form. This may be accomplished by means of a visicorder.

Compare this information with the data which were taken simultaneously on oscilloscopes during the radiation pulse. The oscilloscope pictures and the visicorder information should be compared for any discrepancies. If discrepancies are observed (due to bandwidth limitations in the magnetic tape system), correlations should be made to compensate for any errors which were introduced.

The data should then be normalized in order to eliminate the measured parameter variations due to variation in burst yields.

Representative curves are to be plotted for each type of electron device, and confidence levels indicated.

Adequate statistical methods will be employed during this and other phases of the analysis.

The method of data acquisition (i.e., recording presentation) utilized in the experiment will be compared with the methods used by other investigators. An optimum method of data acquisition will then be proposed.

Based on the results of the above analysis, recommendations will be presented for the design of future experiments.

---

*Magnesium Oxide Radiation Detector

This plan could not be completely accomplished in the short period of time between the conclusion of the experiment, the return of the mobile laboratory, and the present time. What is reported here are the results obtained to date.

RESULTS.

a. General:

Equipment failures during the experiment resulted in a partial loss of information during two shots. Three tape recorders malfunctioned during one exposure, and one recorder during a second exposure. These malfunctions resulted in a total of less than four percent of the data.

Electron devices were exposed during 16 shots. The information recorded on magnetic tape was translated into visual form by playing the tape data into oscilloscopes at the laboratory. Polaroid pictures were taken of the waveforms. The magnetic tapes will also be played back into a visicorder in order to determine long-term recoveries on certain transistors. Recording at 60 inches per second (ips) and playing back at 1-7/8 ips produces a 32:1 reduction in tape speed, and is compatible to the bandwidth of the visicorder. A comparison was also made of the waveforms which were recorded directly on the oscilloscope during a burst to the data which were played back from the tapes. The waveforms were identical.

b. Electron Tubes:

Representative results obtained on electron tubes are shown in the following figures. The bottom trace of Fig. 11 depicts the a-c output of the 7244 ceramic tube. The top trace represents grid leakage current. The internal construction of this tube is identical to the 1724 ceramic tube except that the active internal elements are mounted in a T-6 glass envelope. Fig. 12 is the same as Fig. 11 except that the horizontal sweep time on the oscilloscope is 0.2 ms/cm vs. 0.5 ms in Fig. 11. The a-c output of a conventional 6J6-type electron tube is shown in Fig. 13. The 7244 and 1724 have the same electrical characteristics as the 6J6.

The a-c output of a 1724 ceramic dual triode is shown in Fig. 14. As stated previously, this tube type is identical to the 7244 except for the ceramic envelope.

Direct-record amplifiers, although offering a higher frequency response than FM, introduce a phase shift. This phase shift, using a direct-record amplifier, is shown in Fig. 15. The signal in Fig. 14 was recorded directly onto an oscilloscope. Fig. 15 was recorded on tape through a direct-record amplifier. The phase shift is caused by the low frequency response of the direct-record amplifier. This does not occur in the FM record-reproduce system.

The a-c output of the 12AT7 is shown in the bottom trace of Fig. 16. The top trace is the grid leakage current. The 2225 is the ceramic equivalent of a 12AT7 glass triode. The a-c output and grid current signal of a 2225-type electron tube are shown in Fig. 17. The signal was recorded through a direct-record amplifier. Several direct-record amplifiers were utilized because a limited number of FM amplifiers were available. During future experiments, the tape system will be completely FM.

The a-c output of the 6AQ5 is shown in Fig. 18. The 2146 metal-ceramic tube is the equivalent of a 6AQ5 glass-type tube. The a-c output of the 2146 is shown in Fig. 19.

The a-c output of two 6943-type electron tubes is shown in Fig. 20. The 6943 is a subminiature glass pentode.

The a-c output of a Z-2352 stacked ceramic triode is depicted in Fig. 21.

Nuvistor triodes and tetrodes were irradiated at several positions. Fig. 22 represents the a-c output of a Nuvistor triode located at the reactor scree. A Nuvistor triode positioned 8-1/2" from the screen is shown in Fig. 23. The Nuvistor triode a-c output at a distance of 4-1/2" from the screen is shown in Fig. 24. The Nuvistor tetrodes exposed at the screen and 8-1/2" from the screen are shown in Figs. 25 and 26 respectively.

Before the data on solid state devices is presented, I would like to report the results obtained on the MgO-RAD detector. It is presented at this point because it is a vacuum device. Essentially, the active element of the MgO-RAD consists of a titanium cylinder approximately 1/2" in length and 1/4" in diameter. Inside the Ti envelope or cylinder is a collector electrode concentric with the outer envelope. A ceramic spacer is used as an insulator between the emitter and collector. A negative potential of 300 V is applied to the emitter while the collector is at +300 V.

The inside of the Ti envelope is coated with a layer of specially processed MgO. Secondary electrons (Compton and photoelectrons) ejected from the Ti envelope cause a multiplication of electrons through the MgO layer. It is estimated that a multiplication factor as high as 1000 may ensue.

The MgO-RAD operates on the SEMIRAD principle, i.e., secondary electrons; however, the multiplication which is produced permits a small detector size with a high output. A typical radiation pulse detected with the MgO-RAD is shown in Fig. 28.

c. Solid State Devices:

The data compiled on solid state devices are in the process of further analysis. The following information is presently available:

Fig. 29 shows the transient change in I_{CO} in a germanium (Ge) n-p-n developmental transistor. The peak leakage current is 82 μa. The neutron dose during this shot was 1×10^{12} NVT. Although the gamma dose rate was recorded on one channel on each tape recorder, the SEMIRAD detector which was utilized failed to function properly. Additional data were obtained on an oscilloscope with another gamma detector, and an attempt will be made to correlate the peak gamma dose rate with the peak I_{CO} for various devices.

The top trace in Fig. 30 shows the change in H_{FE} of a silicon 2N706 n-p-n transistor. This corresponds to a 25% decrease in H_{FE}. The bottom trace represents the transient increase in I_{CO} during the burst.

The change in H_{FE} on a 2N1039 germanium p-n-p power transistor is shown in the top trace of Fig. 31. The bottom trace is the change in I_{CO}.

The response of a 2N710 germanium p-n-p type transistor is shown in Fig. 32. Trace A is the change in H_{FE}, while Trace B is the transient change in I_{CO}. The neutron dose is approximately 10^{12} NVT.

Fig. 33, Trace A shows the change in H_{FE} of a 2N1305 germanium p-n-p transistor. The peak leakage current is shown in Trace B.

Fig. 34, Trace A shows the permanent damage in a 2N406 germanium p-n-p transistor after a neutron dose of 1.1×10^{12} NVT. The bottom Trace B is the transient change in I_{CO}.

CONCLUSIONS.

a. Transistors:

The transient and permanent changes in I_{CO} and H_{FE} shown for the transistors are representative of each type. Upon completion of the analysis of these data, confidence limits for each type of device will be available. Variations in I_{CO} and H_{FE} will be correlated to gamma rate, neutron dose, and type of device - base width, alpha cutoff, etc.

In general, the transient change in I_{CO} is proportional to the gamma dose rate, and the permanent change in H_{FE} is proportional to the integrated dose.

b. Electron Tubes:

Transient changes occurring in electron tube operation appear to be caused by electron emission (photoelectrons, Compton electrons) from the elements in the tube structure. Under static conditions (no input signal), an increase in plate current would be observed during the duration of the radiation pulse. Under normal a-c signal operation, this increase in plate current causes the tube to operate or amplify on the non-linear portion of the i_b vs. e_g curve. After the radiation pulse, the tube resumes normal operation with no permanent damage.

Circuit time constants associated with different tube types will be a contributing factor to the time interval before the tube will resume normal operation. Materials in the tube and type of tube geometry will also be a significant factor.

In general, both the ceramic and glass tubes appear to exhibit the same transient effects during the radiation pulse. The ceramic-metal tube in which the metal envelope is internally connected to the plate exhibited the greatest effect. This would be consistent with the afore-mentioned theory, inasmuch as additional electrons would be liberated because the physical cross-section is greater.

The Nuvistor triode showed almost negligible effects during the radiation burst, which is attributable to the type of construction and size. The Nuvistor tetrode showed greater effects during the burst. This can be related to the external cap which is the plate connection.

Other effects, such as gas liberation, changes to insulation resistance are possible in certain types of tubes. Additional studies will be performed in order to evaluate different types of electron tube construction and to investigate the mechanisms which produce the transient effects.

c. Dosimetry:

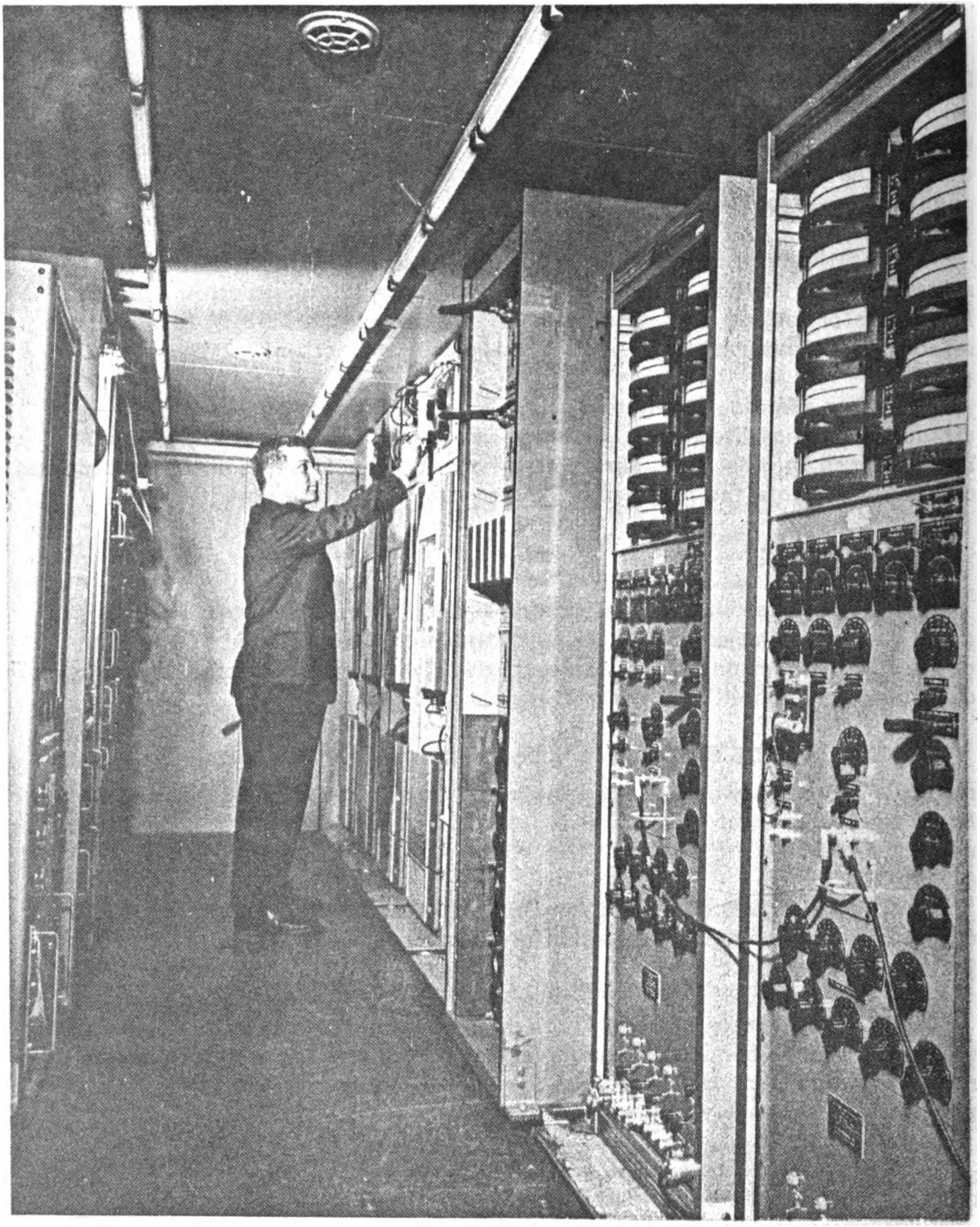
The results obtained with the NBS film badge dosimeters were inconsistent with the data obtained with the micro-dosimeters. Darkening of the film by neutrons causes this error. The correction factor for neutron darkening was larger than the actual gamma dose.

The USASRDL S^{32} neutron dose measurements were higher by approximately a factor of two over the Sandia S^{32} neutron dose. This was attributed to the fact that a different cross-section and calibration source were used by USASRDL.

The MgO-RAD compared favorably with the Sandia photodiode. The output of the MgO-RAD was proportional to the ΔT (temperature rise or yield) of the reactor.

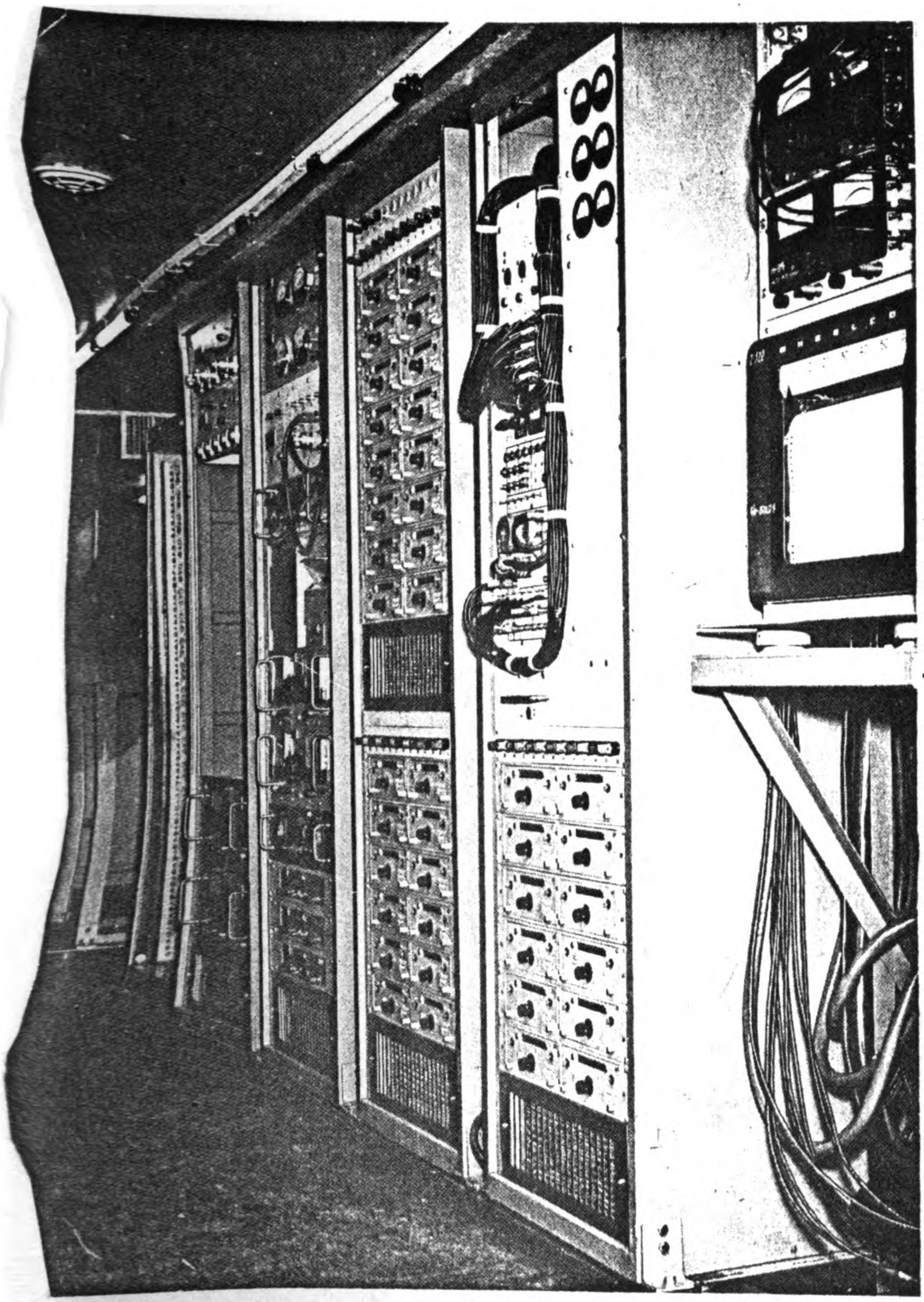
In conclusion it may be said that the Radiation Effects Mobile Laboratory will permit multi-parameter measurements during radiation effects experiments. In addition, the design of the instrumentation in the REML has proven to be extremely rugged.

Further experiments will be conducted after a complete analysis of the data obtained on the 600 electron devices. The results of the analysis will determine the design of the next experiment.



FM TAPE RECORDERS AND ELECTRON TUBE CIRCUITRY

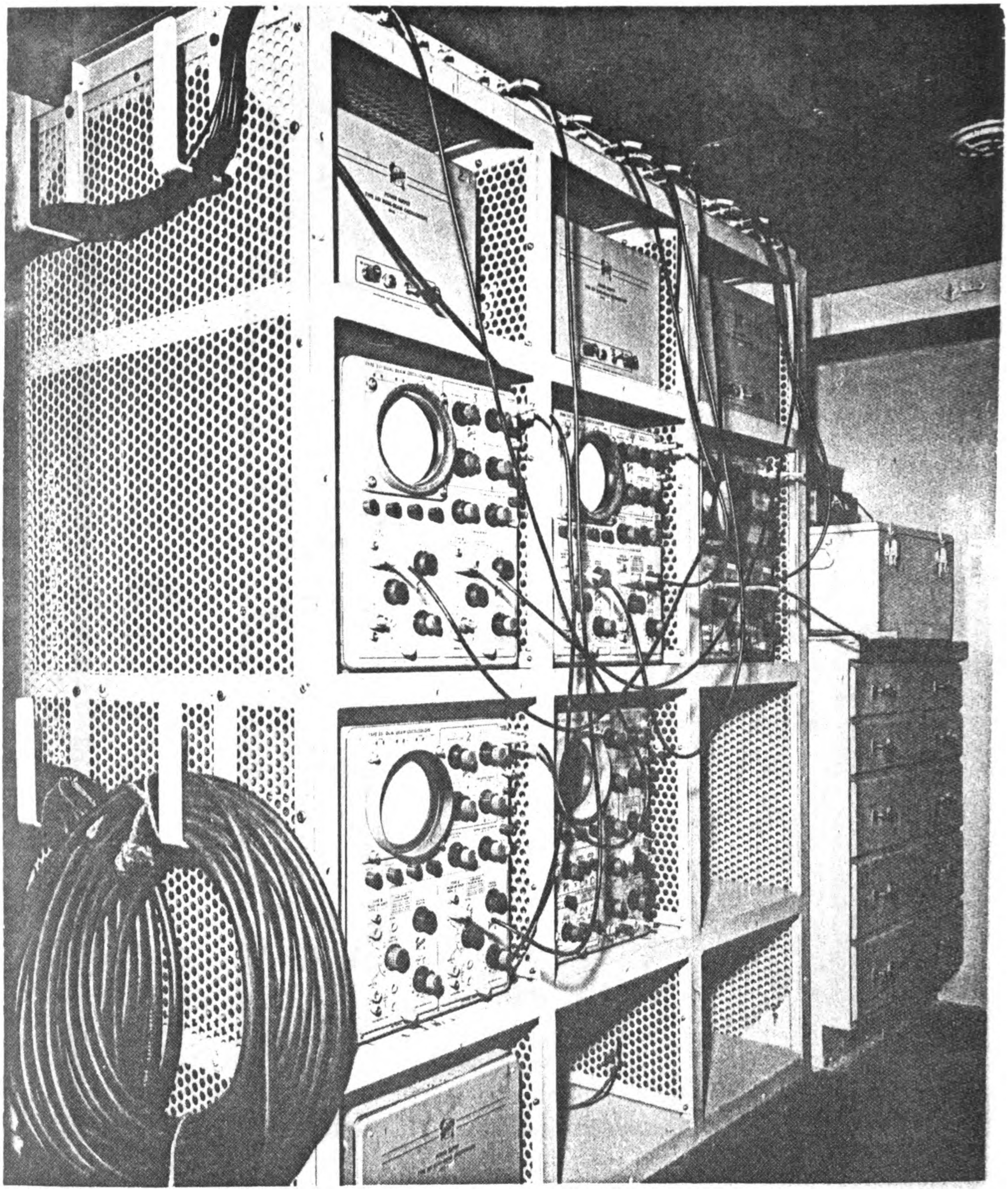
FIG. 1



RADIATION EFFECTS MOBILE LABORATORY

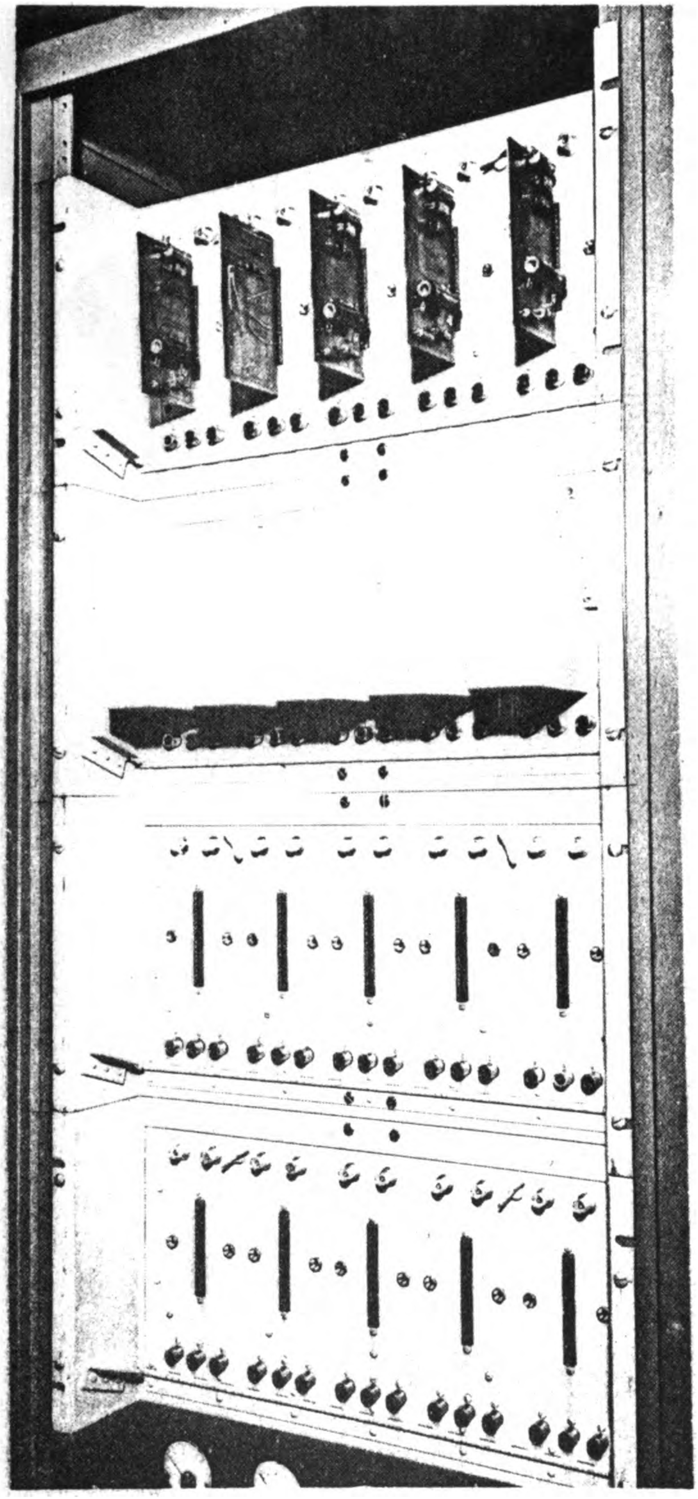
TEMPERATURE RECORDER, TRANSISTOR CIRCUITRY, DIFFERENTIAL PRE-AMI
POWER SUPPLIES, DIGITAL VOLTMETER AND PRINTOUT

FIG. 2



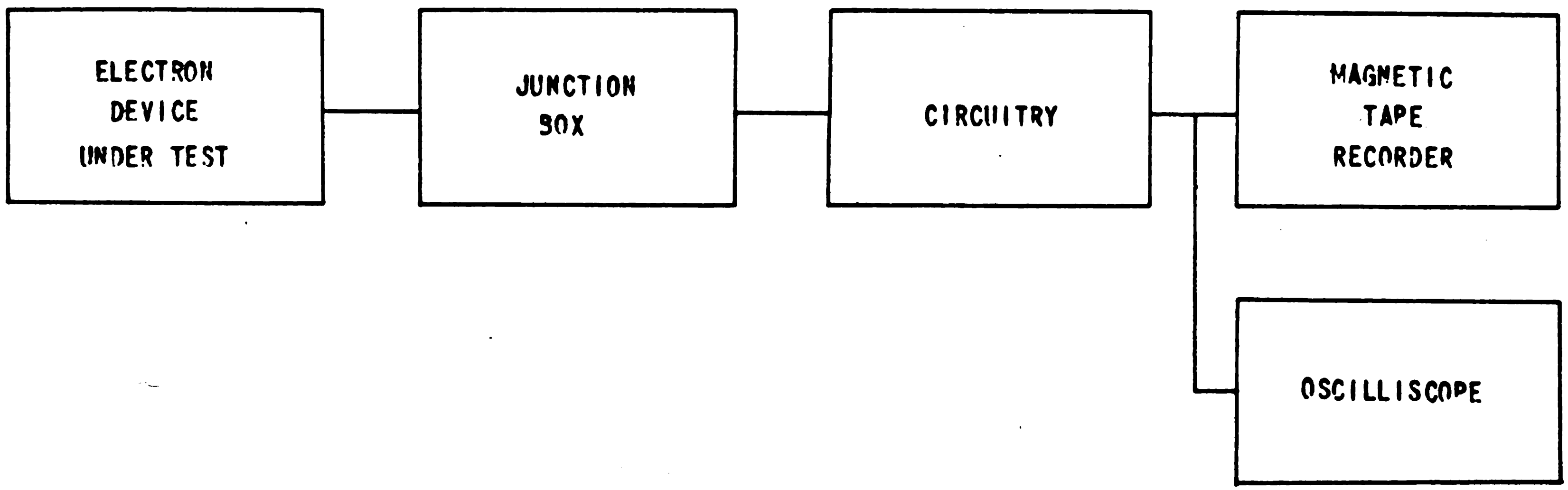
RADIATION EFFECTS MOBILE LABORATORY
UNIT HOUSING SIX OSCILLISCOPES AND POWER SUPPLIES

FIG. 3



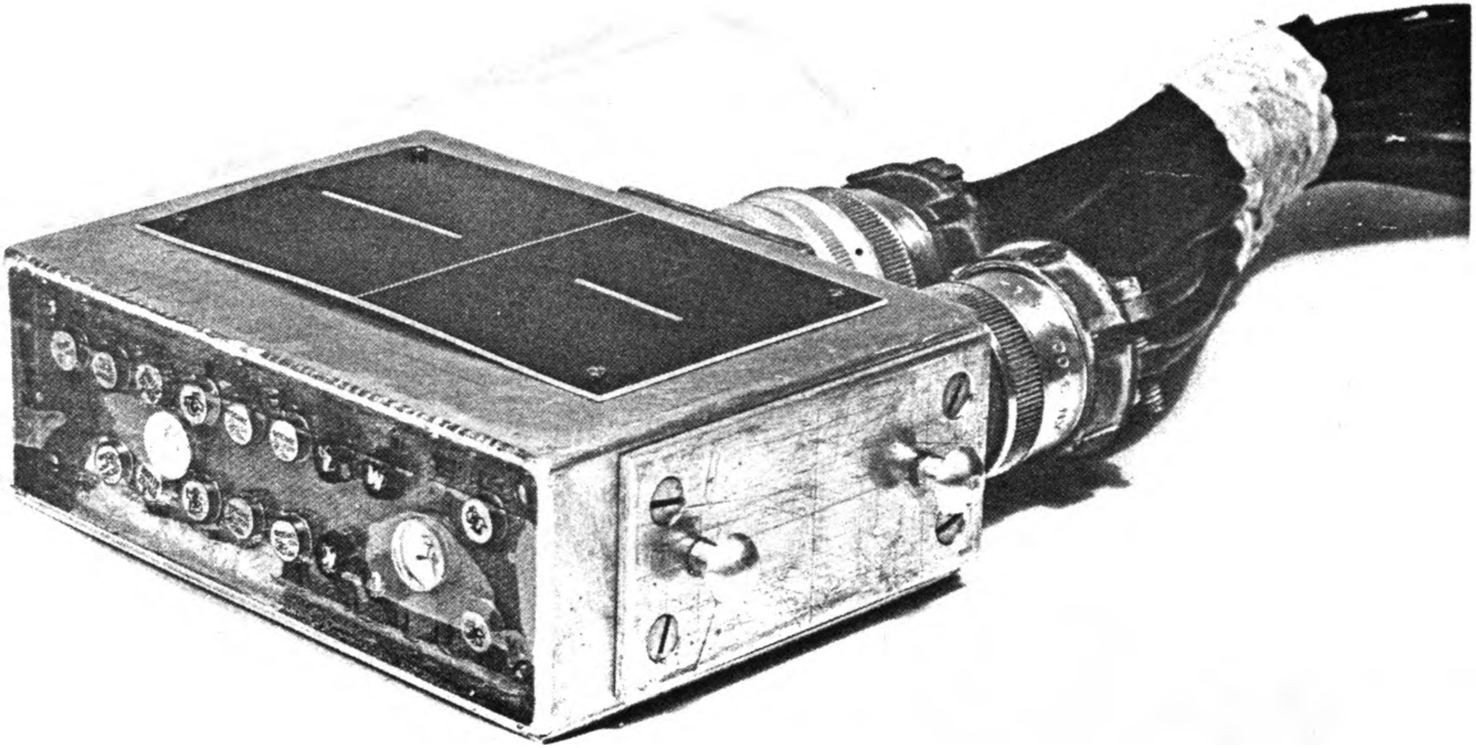
ELECTRON TUBE INSTRUMENTATION UTILIZING PRINTED CIRCUIT BOARDS

FIG. 4



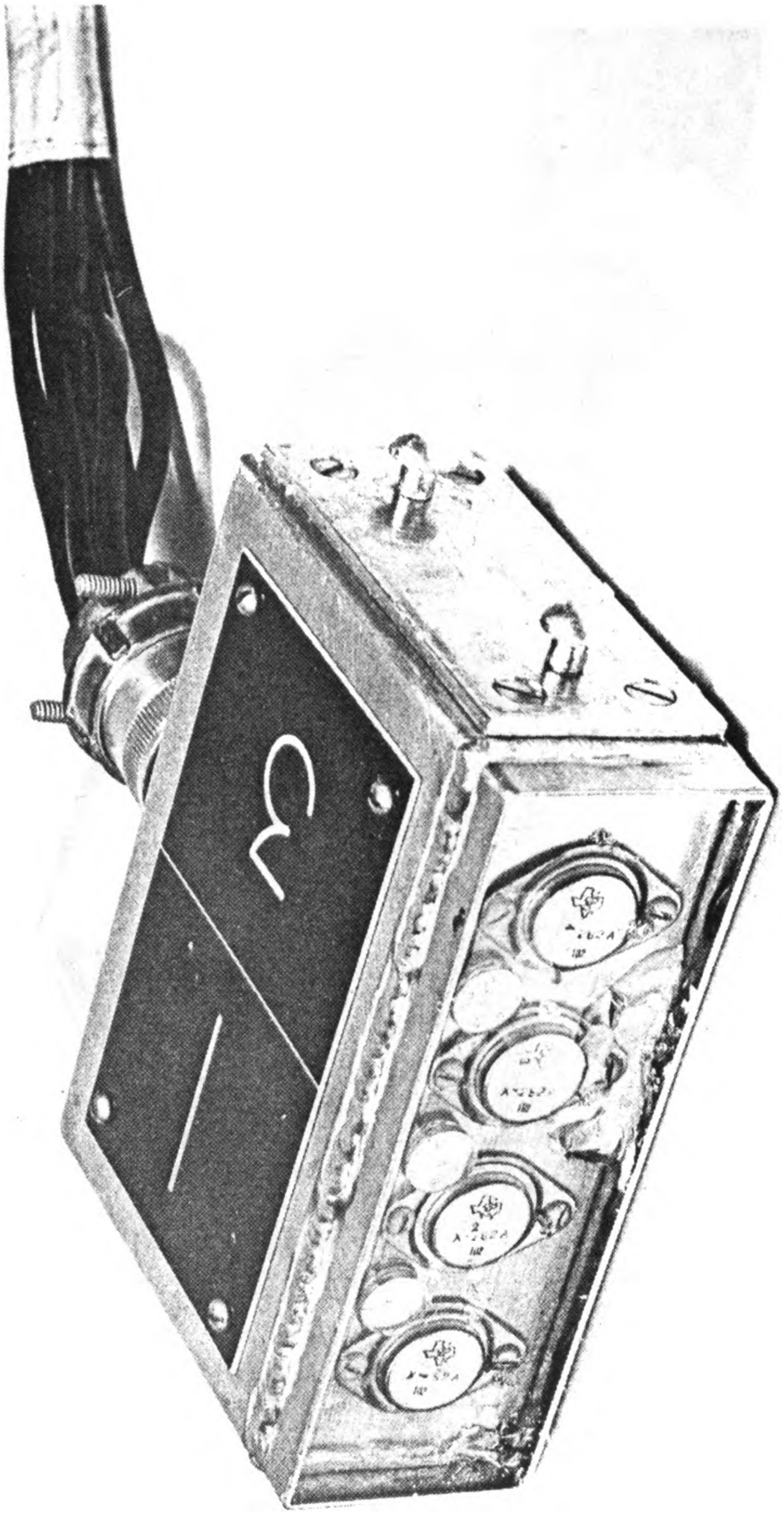
BLOCK DIAGRAM FOR TYPICAL EXPOSURE

FIG. 5



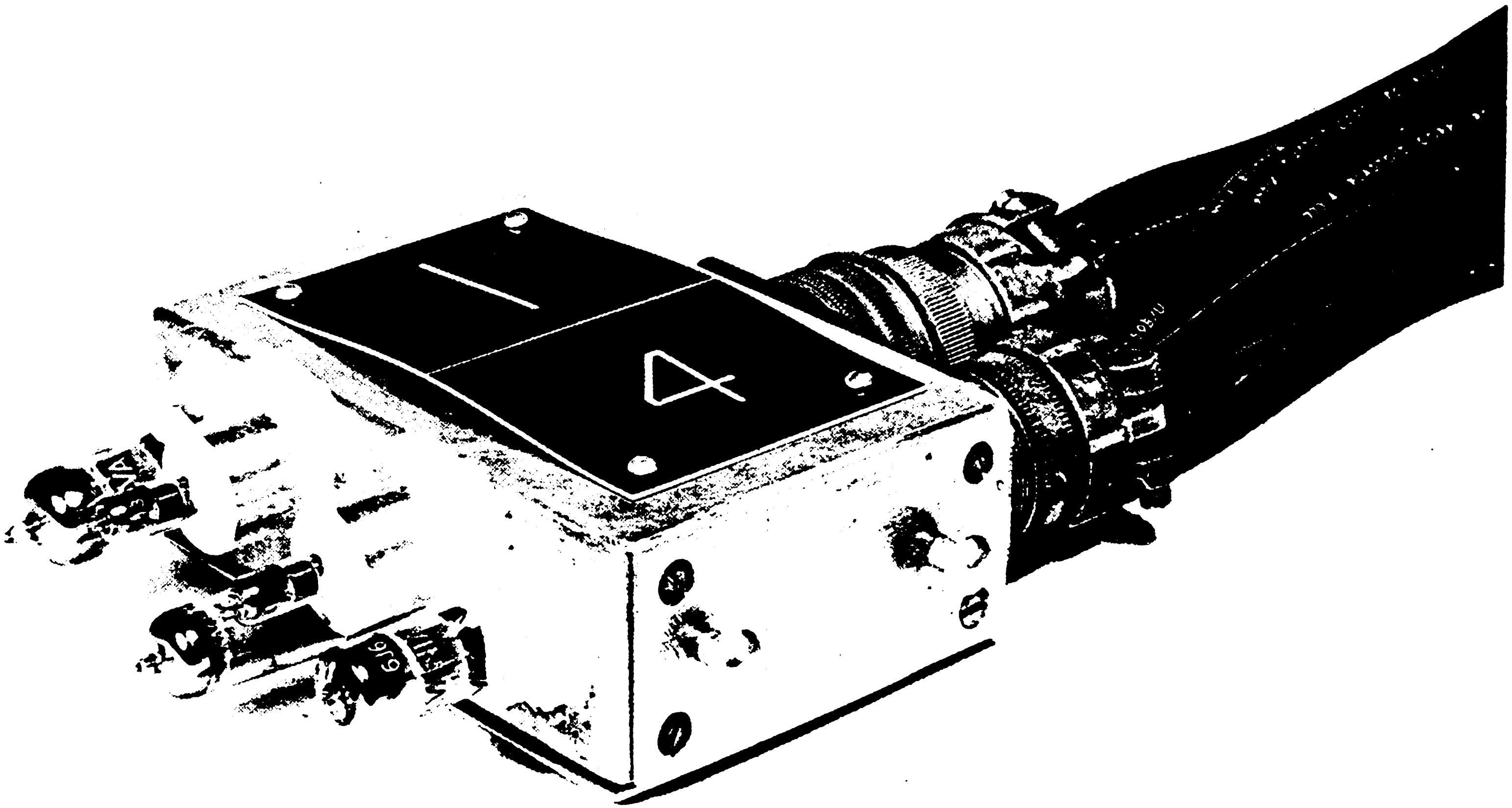
DIFFERENT TYPES OF TRANSISTORS MOUNTED ON CHASSIS WITH DOSIMETRY

FIG. 6



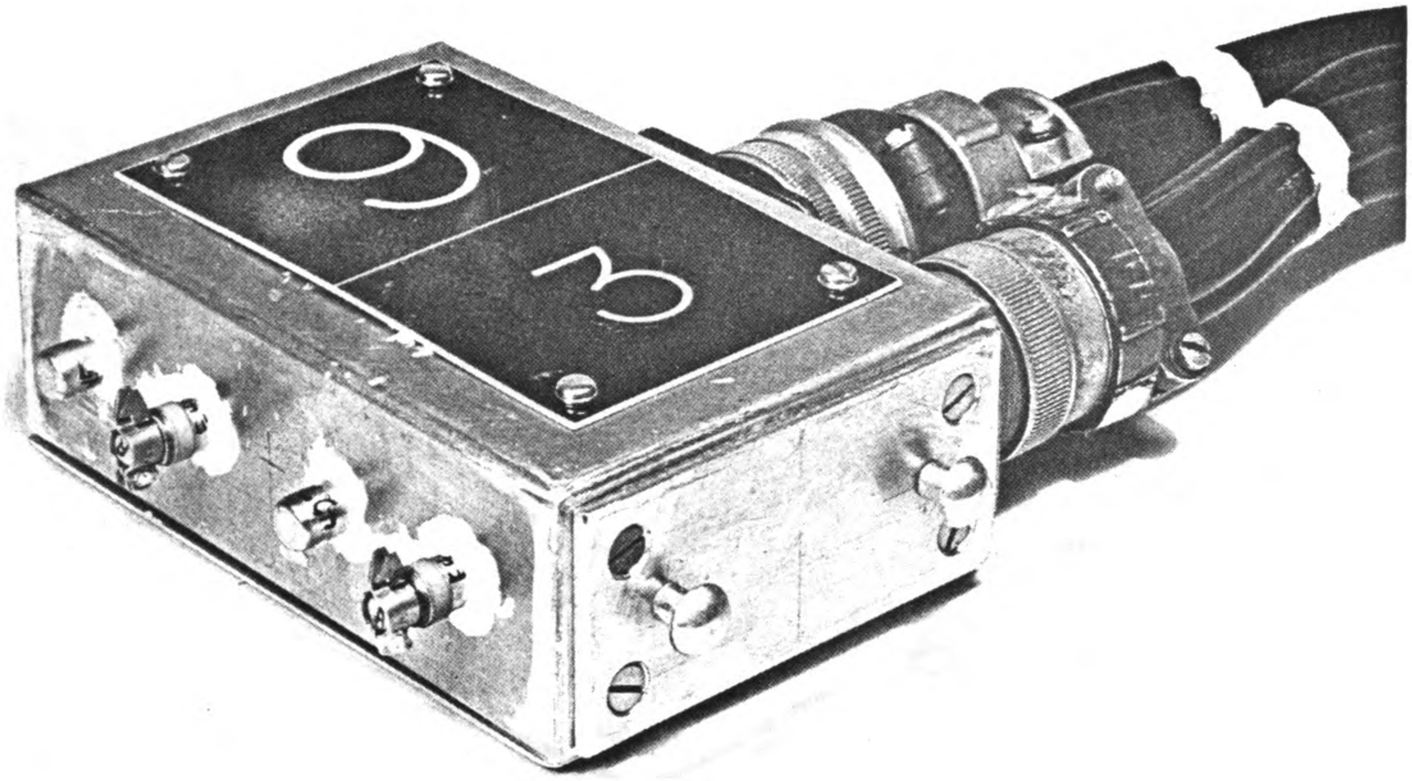
POWER TRANSISTORS DOSIMETRY MOUNTED ON CHASSIS

FIG. 7



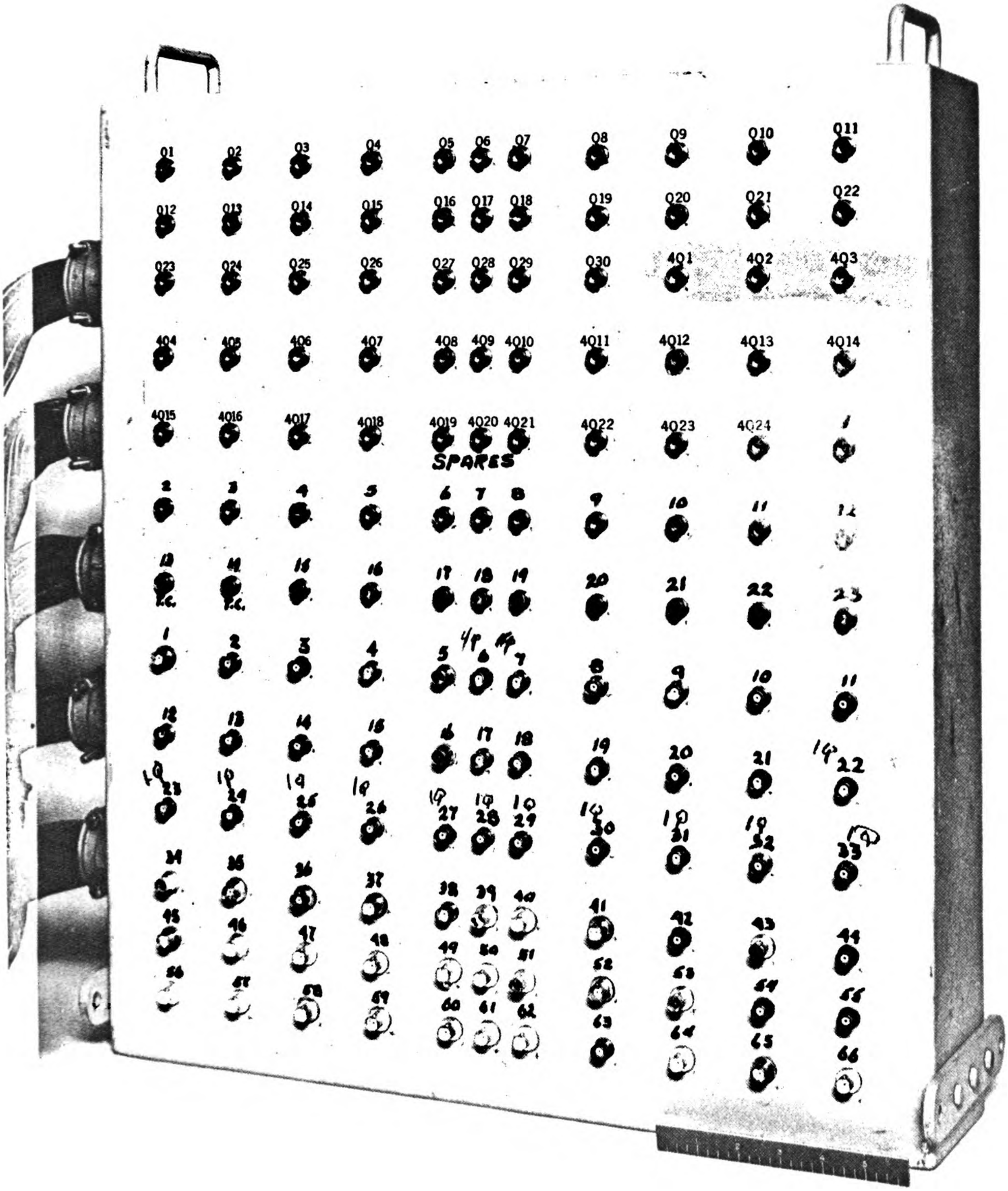
ELECTRON TUBES MOUNTED ON CHASSIS

FIG. 8



NUVISTOR TYPE ELECTRON TUBES MOUNTED ON CHASSIS

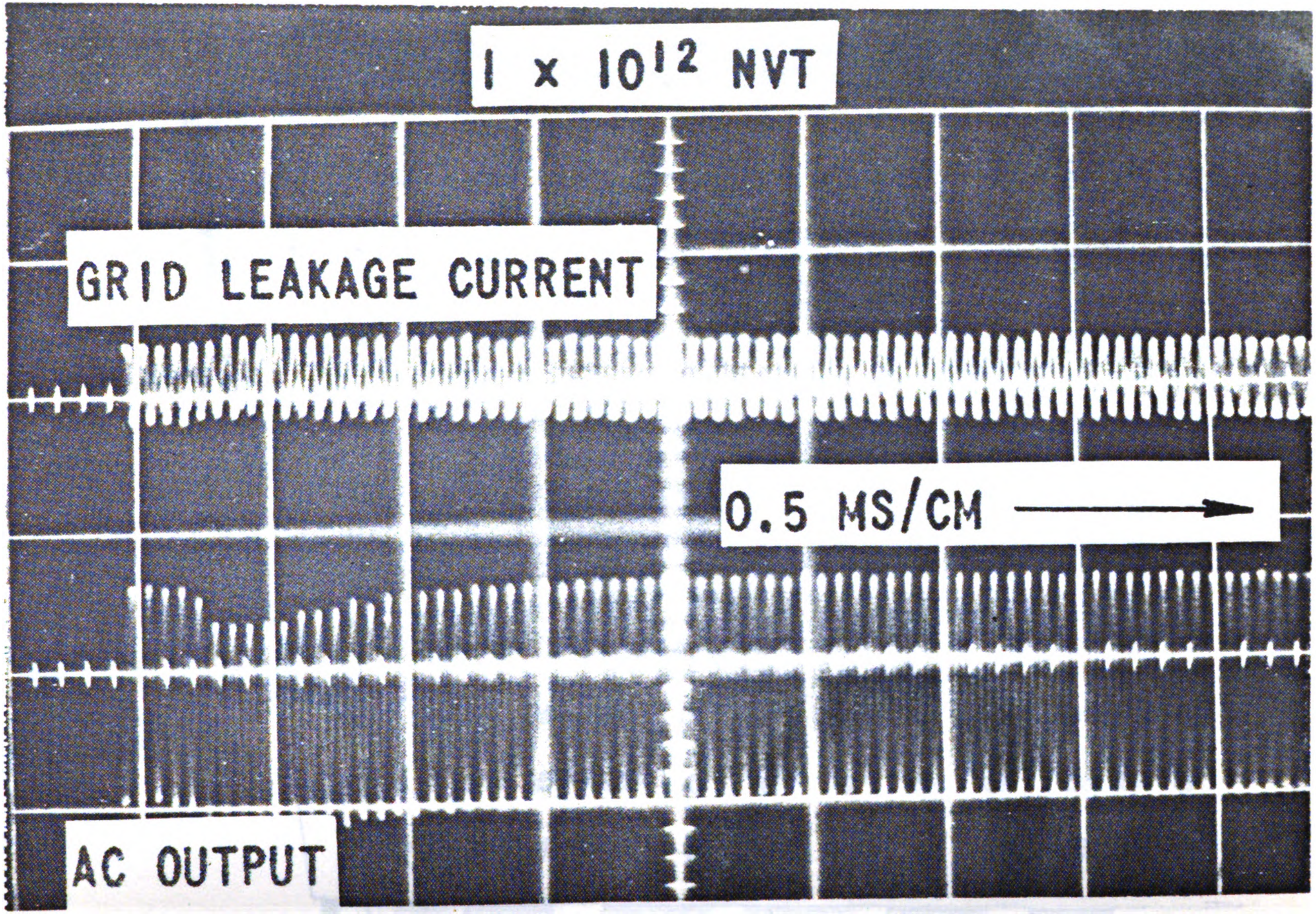
FIG. 9



JUNCTION BOX

FIG. 10

Fig. 11



7244 ELECTRON TUBE

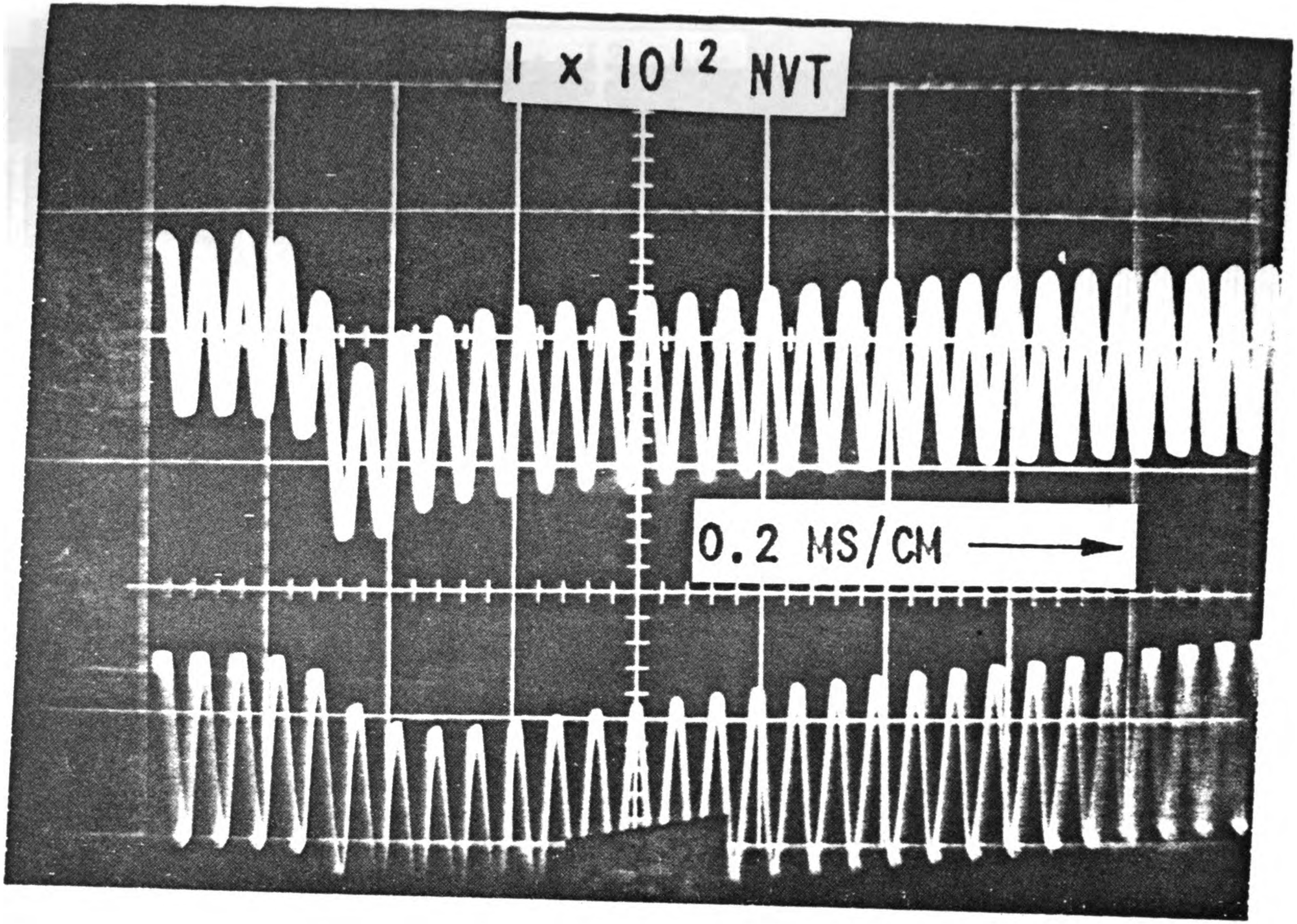
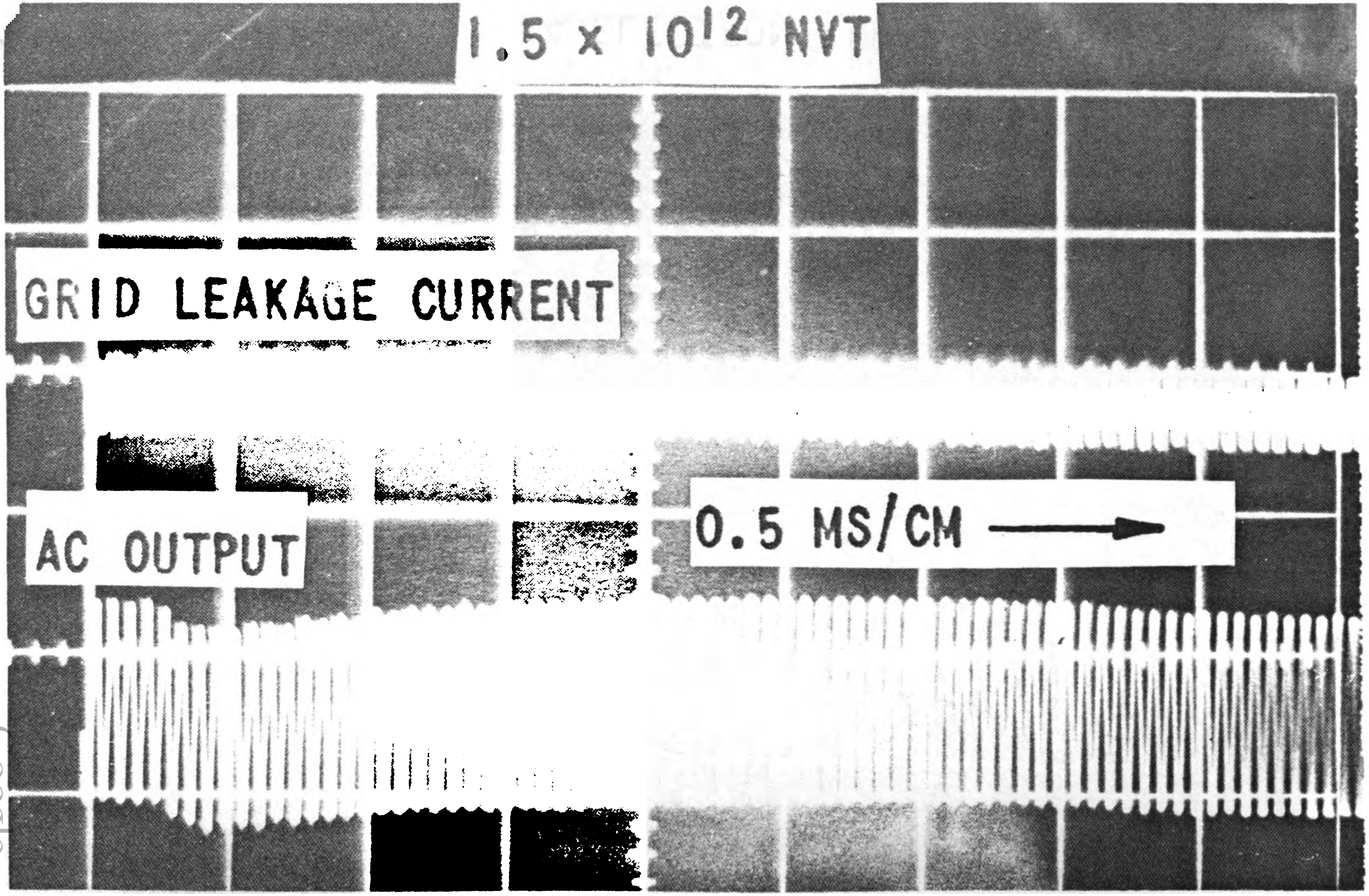


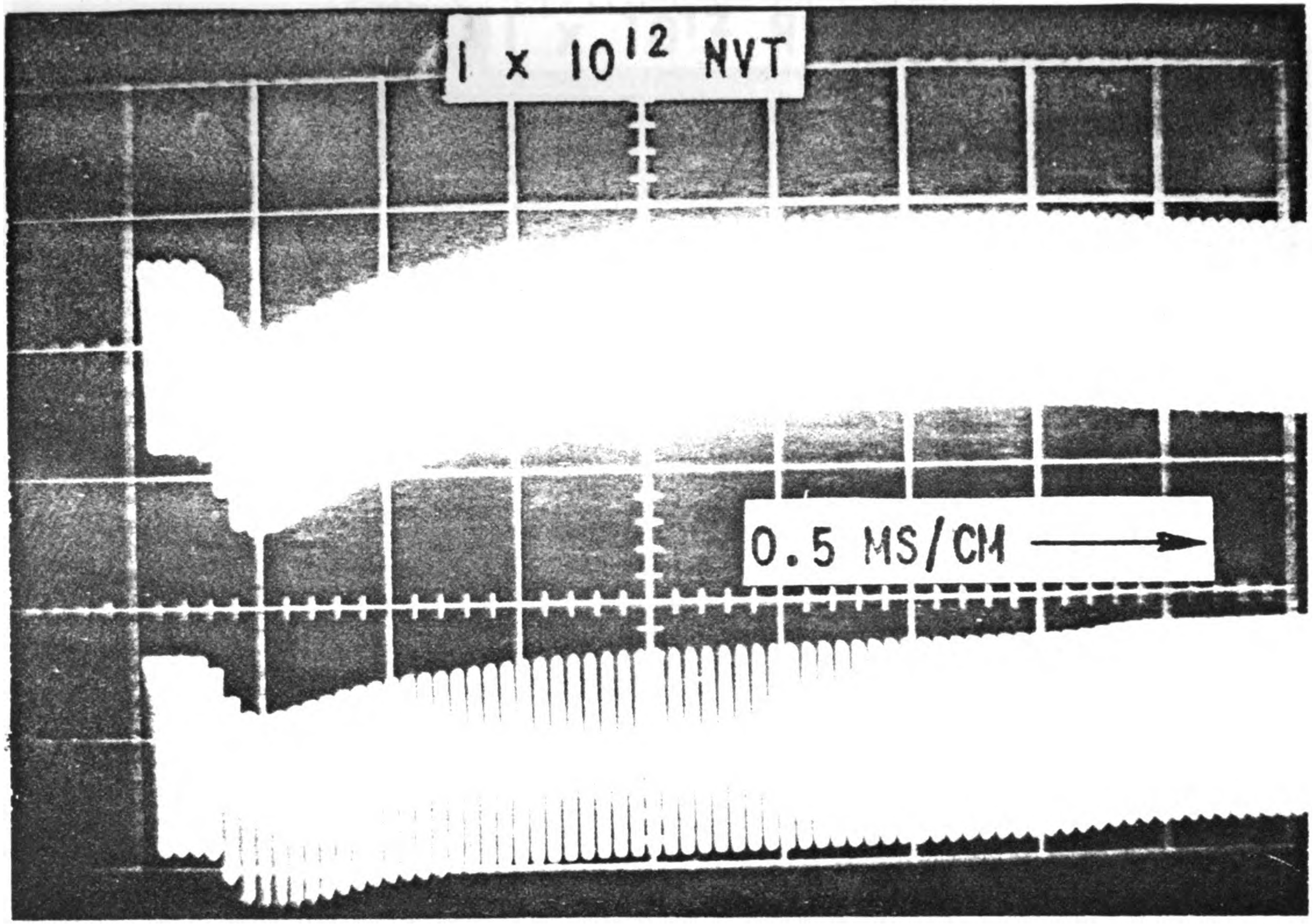
Fig. 12

7244 ELECTRON TUBE

Fig. 13



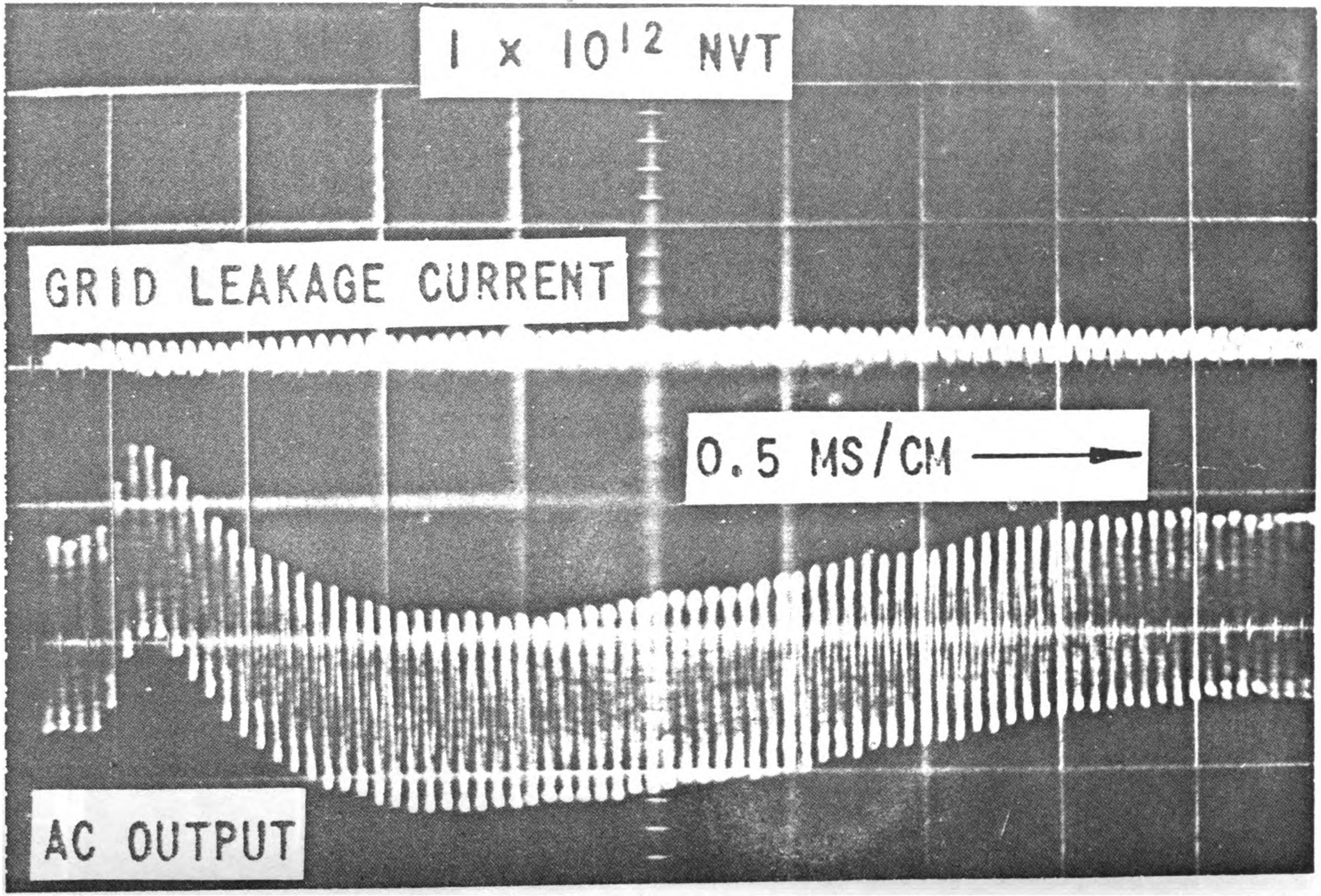
6J6 ELECTRON TUBE



1724 ELECTRON TUBE

FIG. 14

Fig. 15



1724 (DIRECT RECORD)

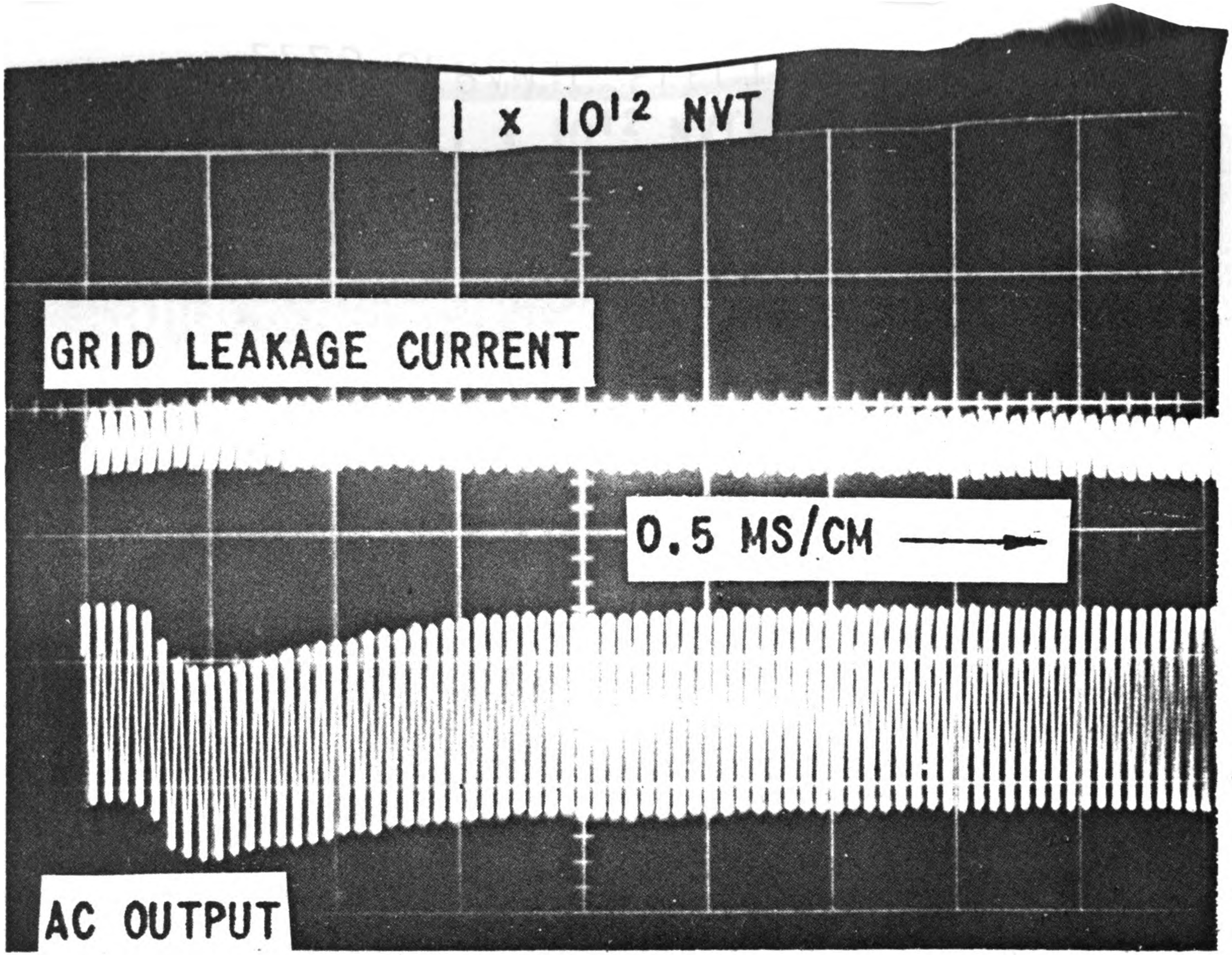
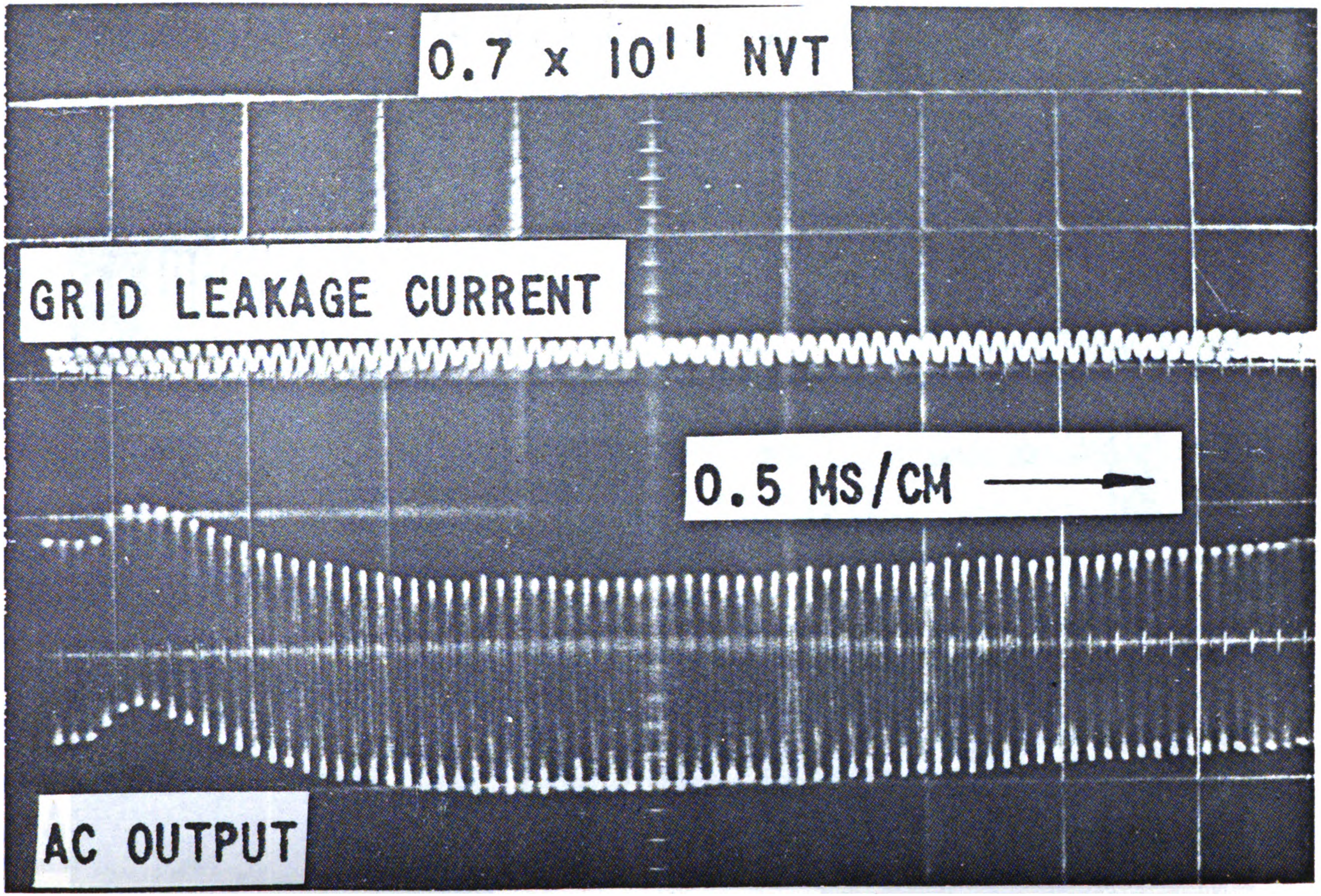
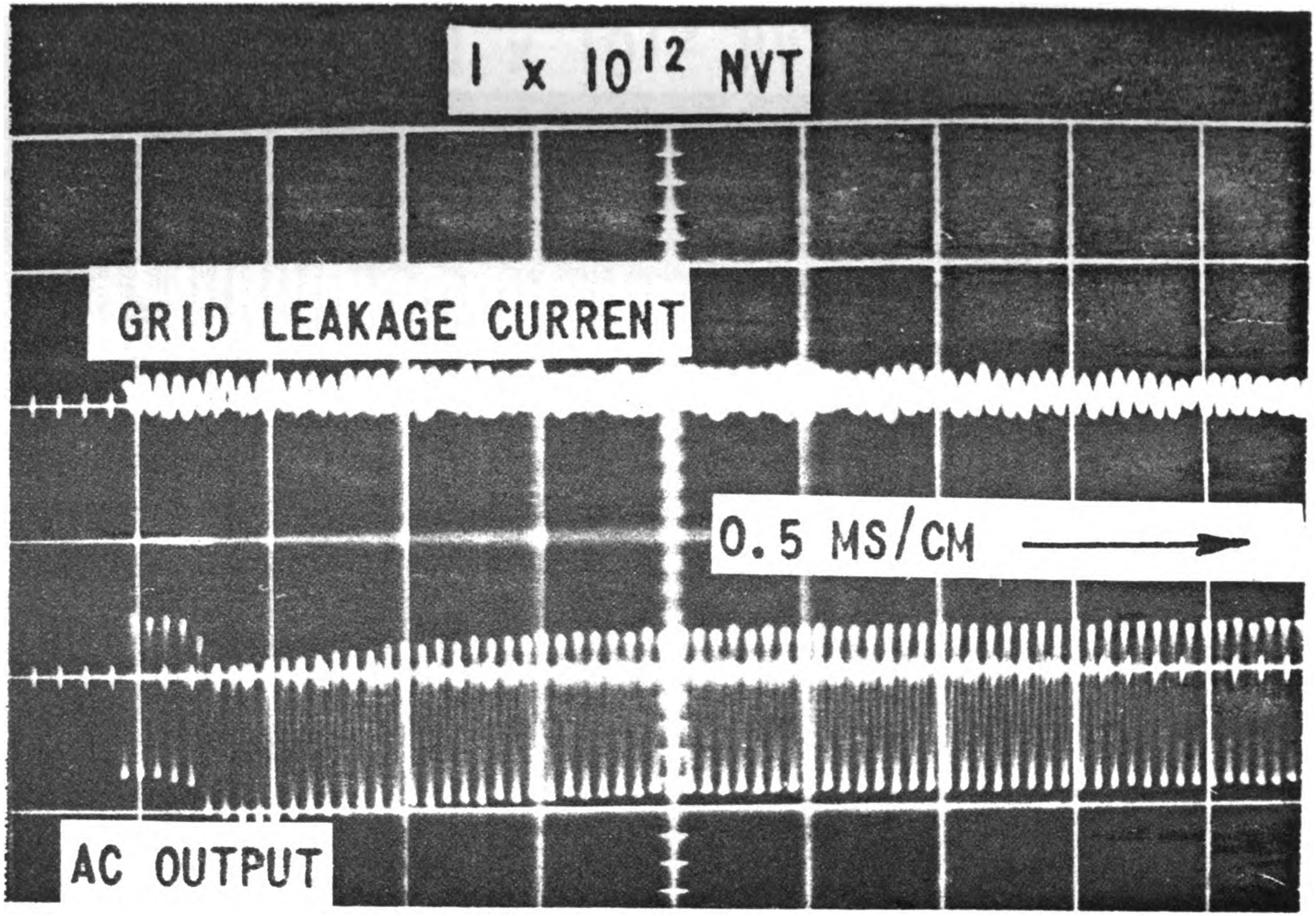


Fig. 16

12 AT7 ELECTRON TUBE



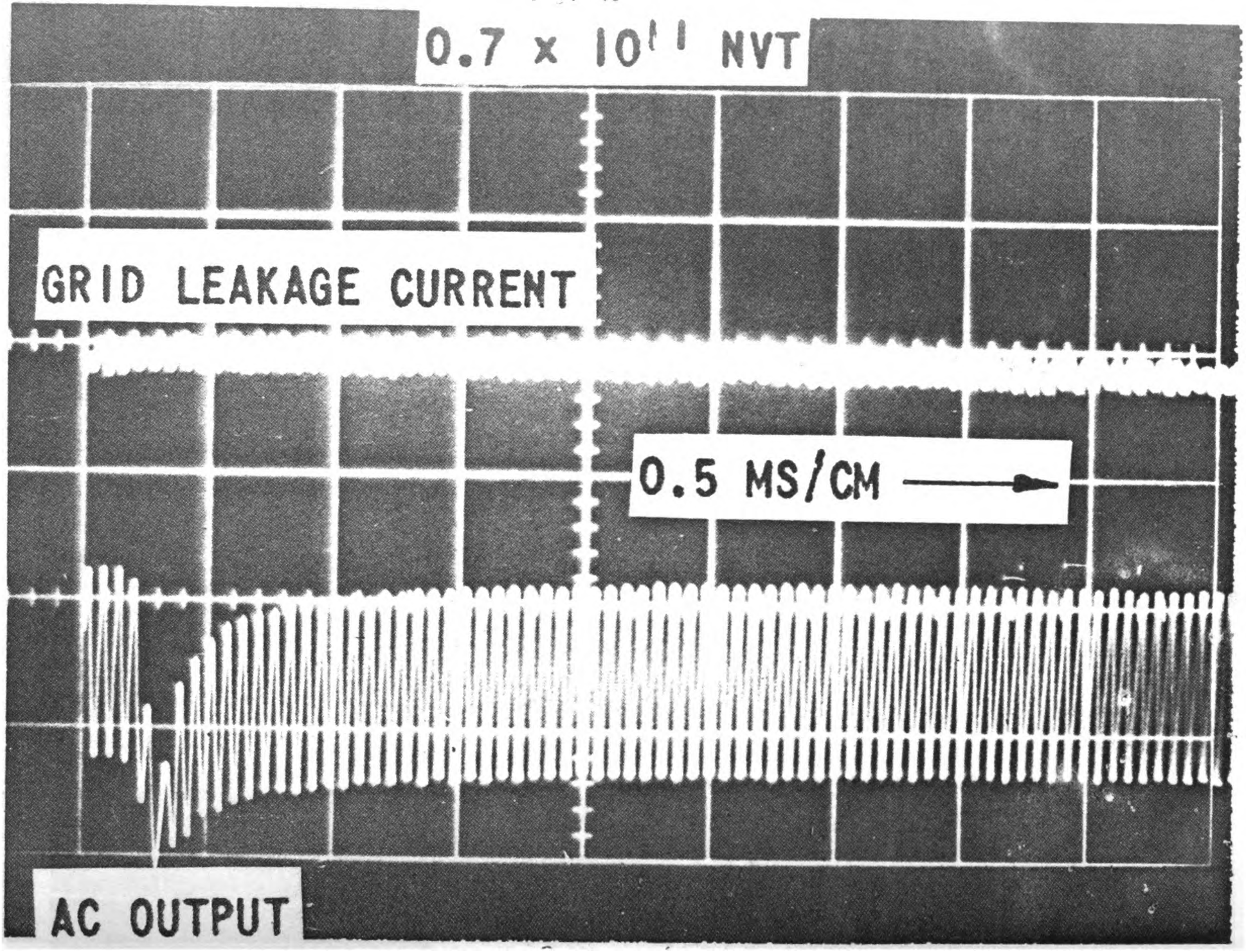
2225 CERAMIC ELECTRON TUBE



6AQ5 ELECTRON TUBE

FIG. 18

Fig. 19



2146 METAL CERAMIC ELECTRON TUBE

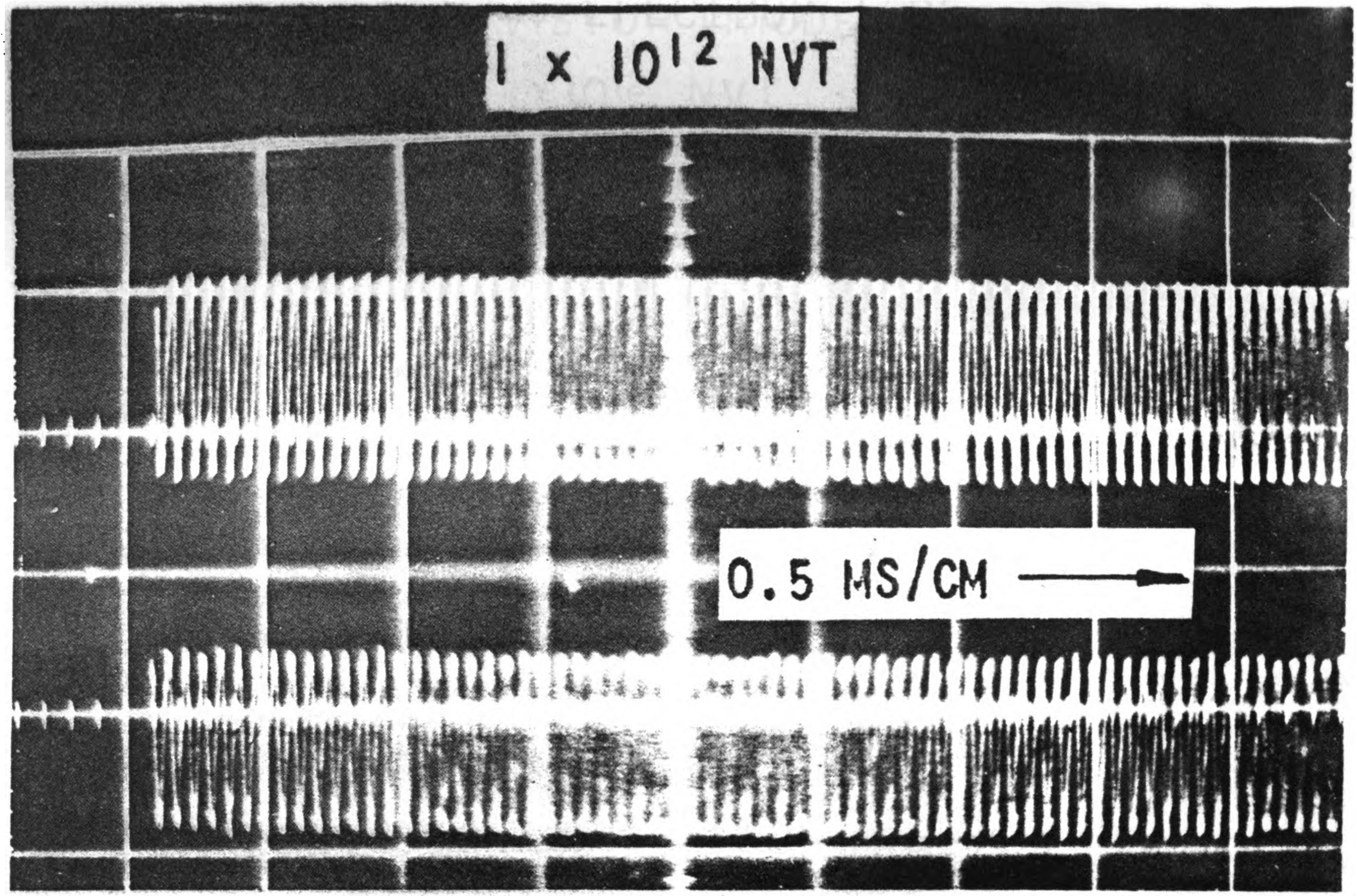
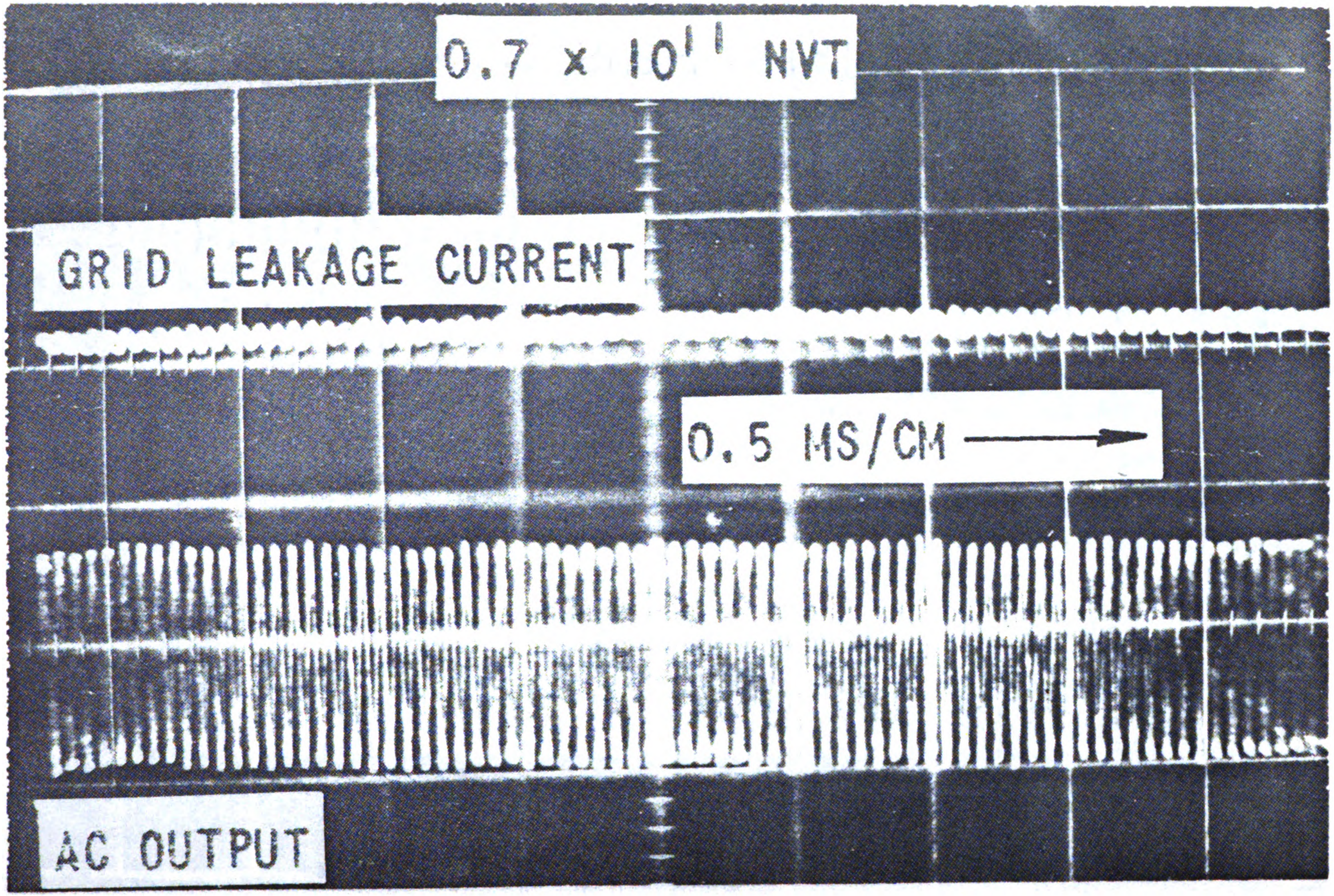


Fig. 20

AC OUTPUT 6943 SUB-MINIATURE PENTODE

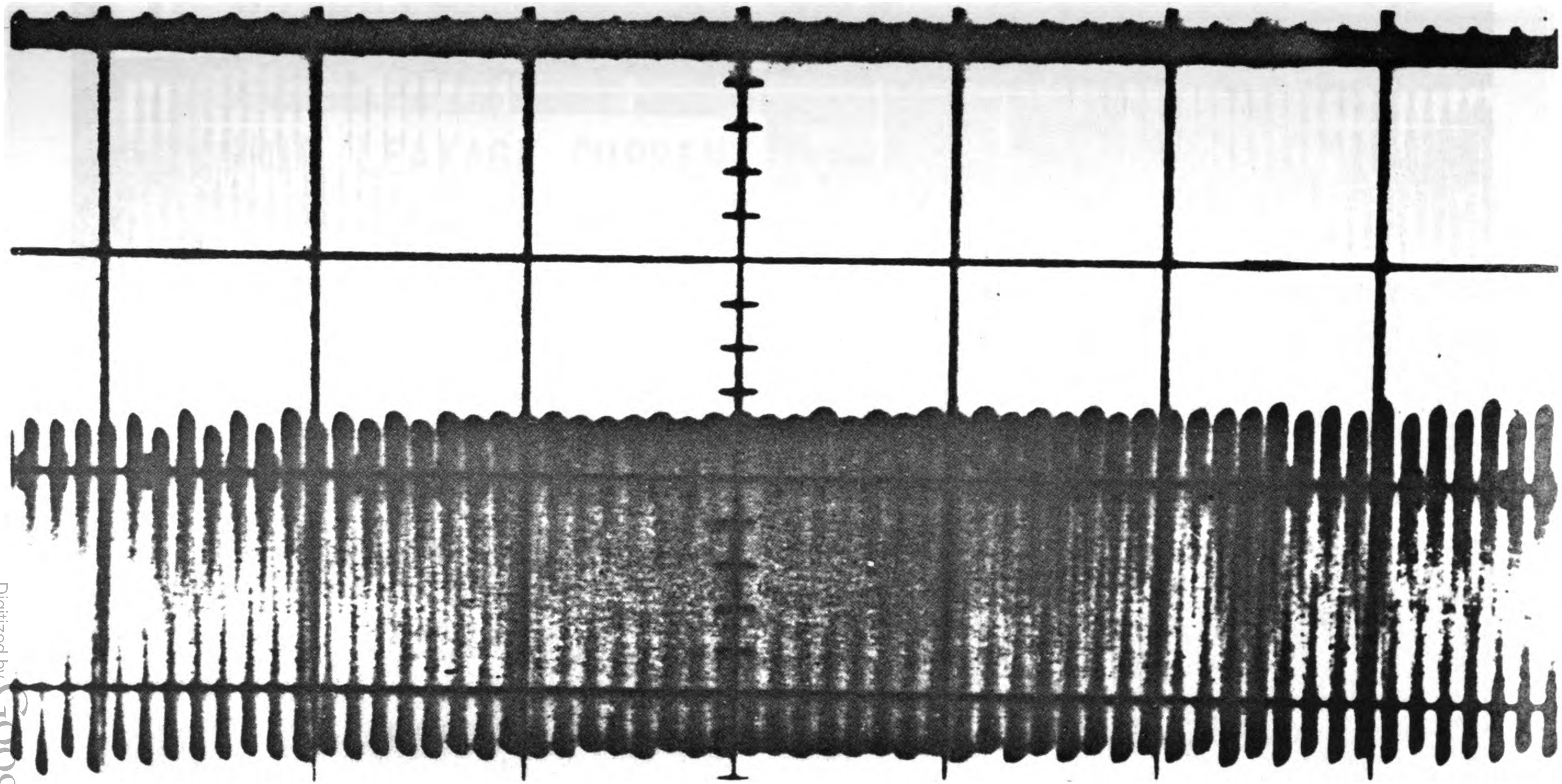
Fig. 21



Z-2352 ELECTRON TUBE

NUVISTOR TRIODE

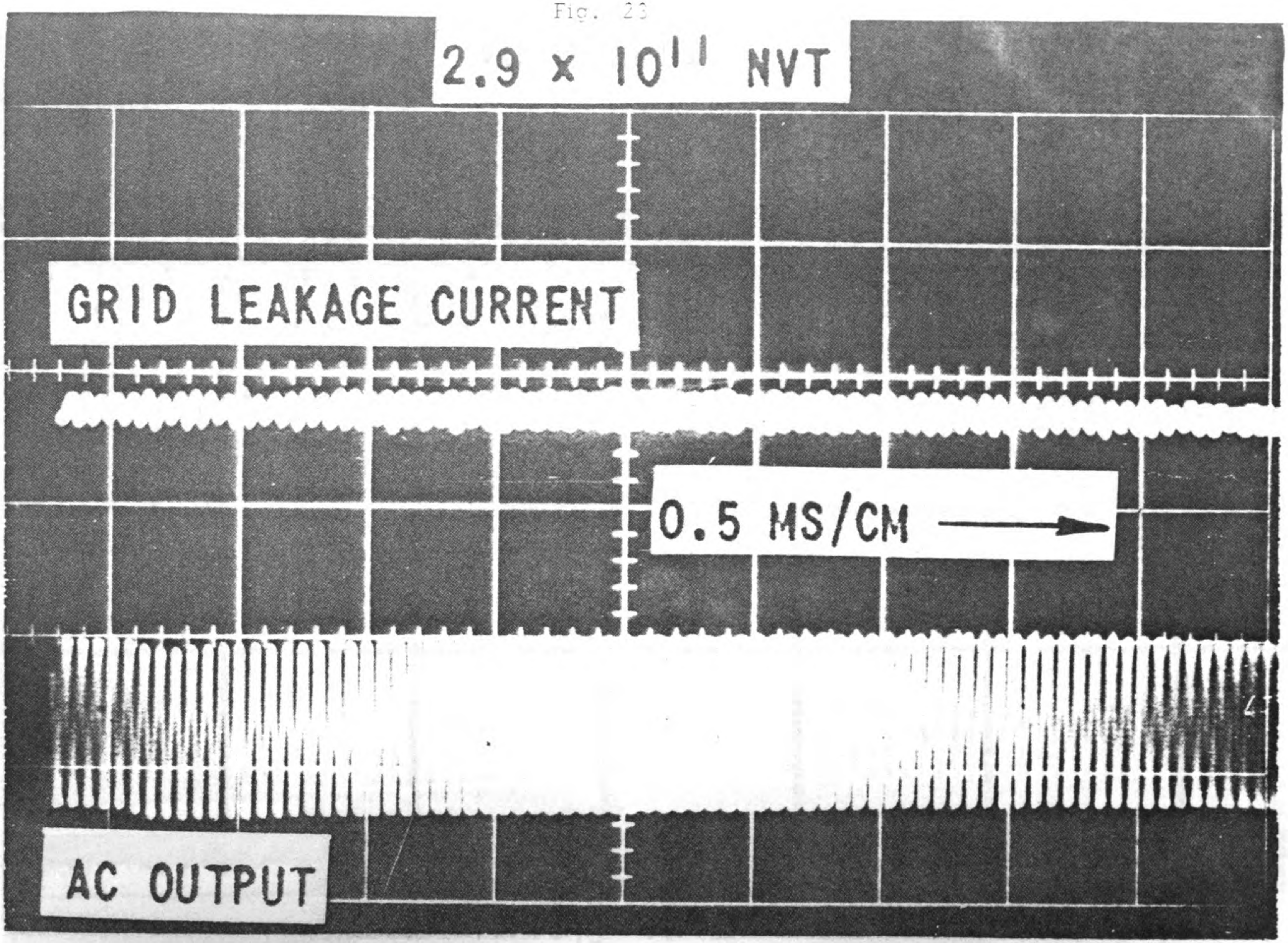
1×10^{12} NVT



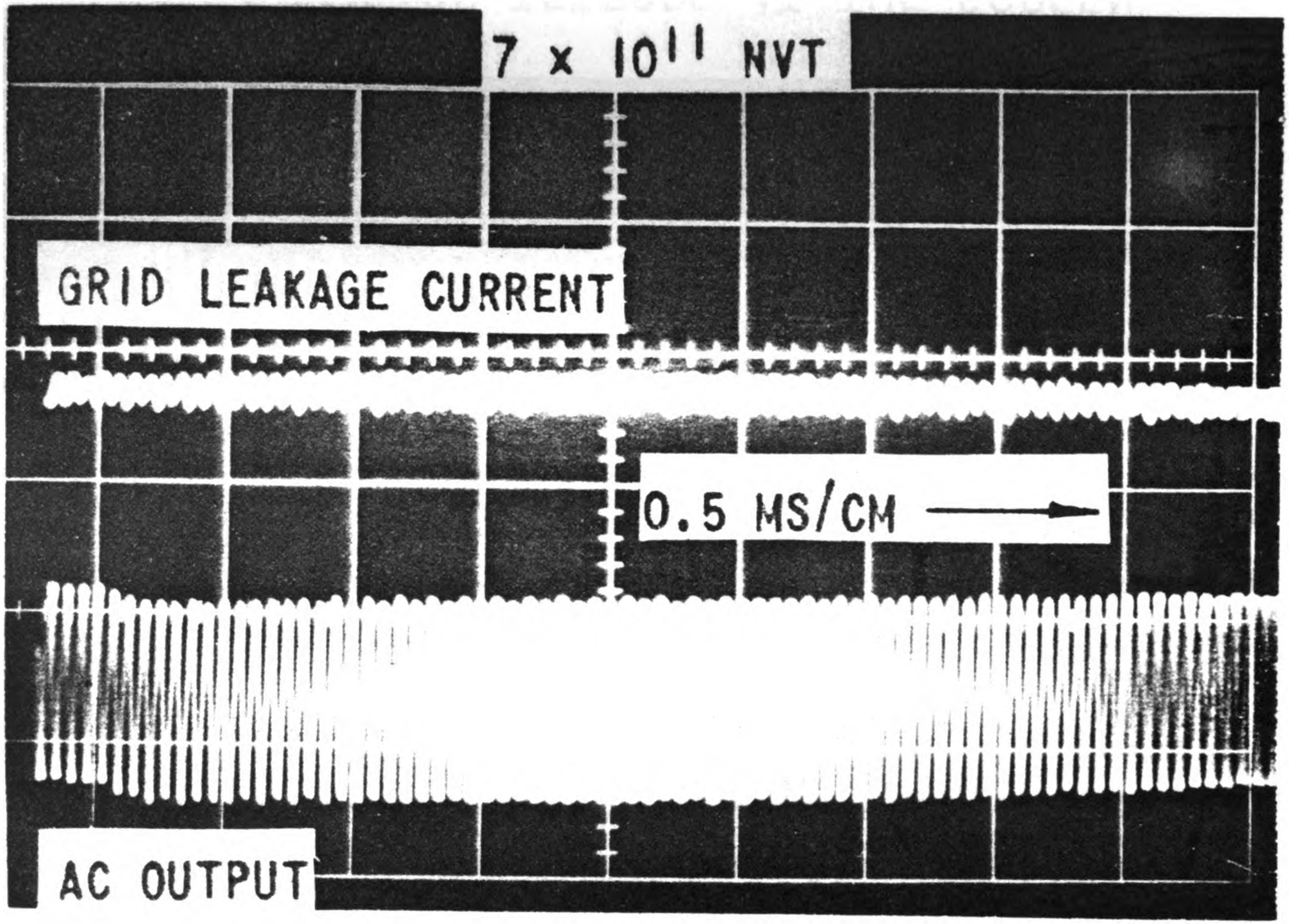
AC OUTPUT

FIG. 22

Fig. 23



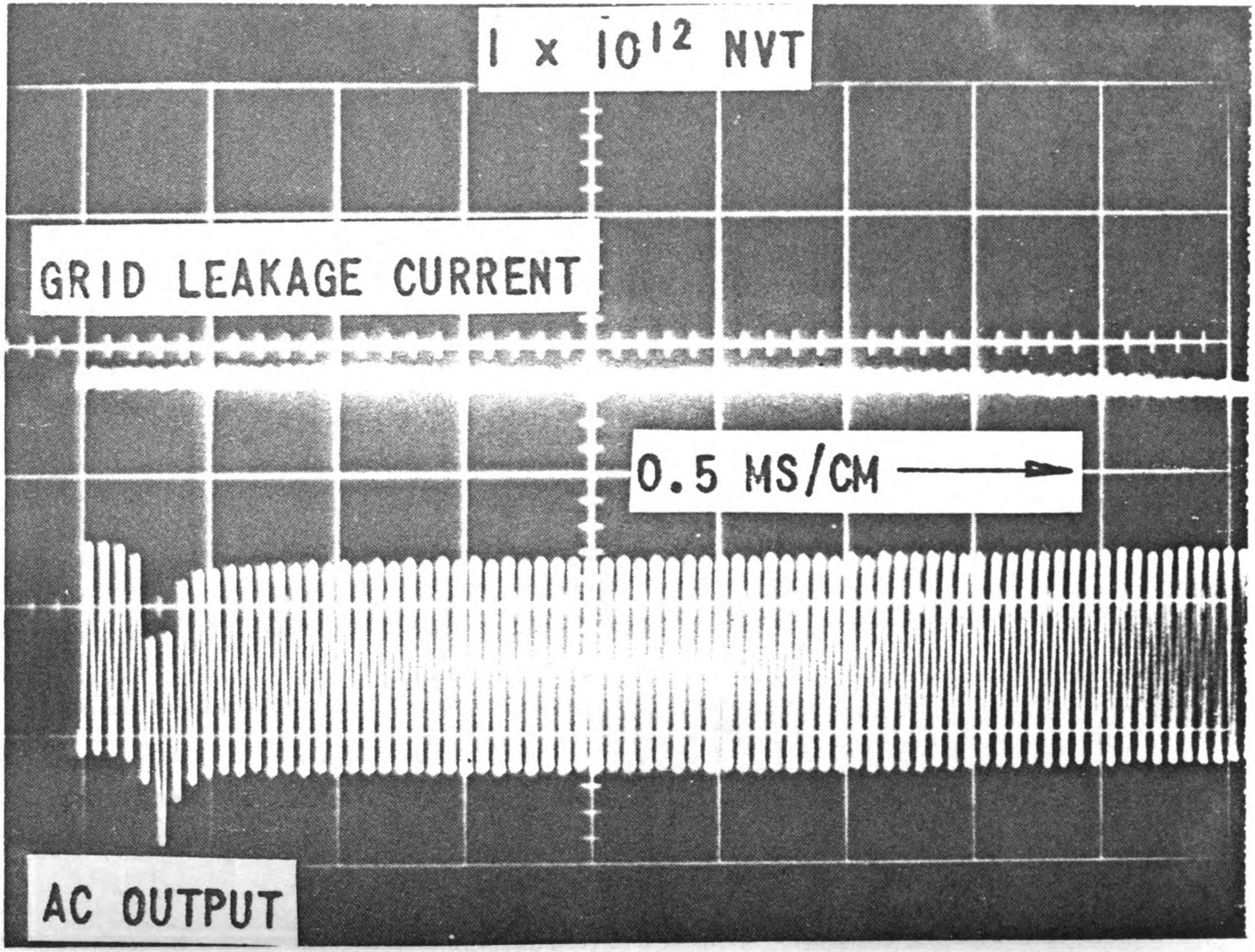
NUVISTOR TRIODE 8 1/2" FROM REACTOR SCREEN



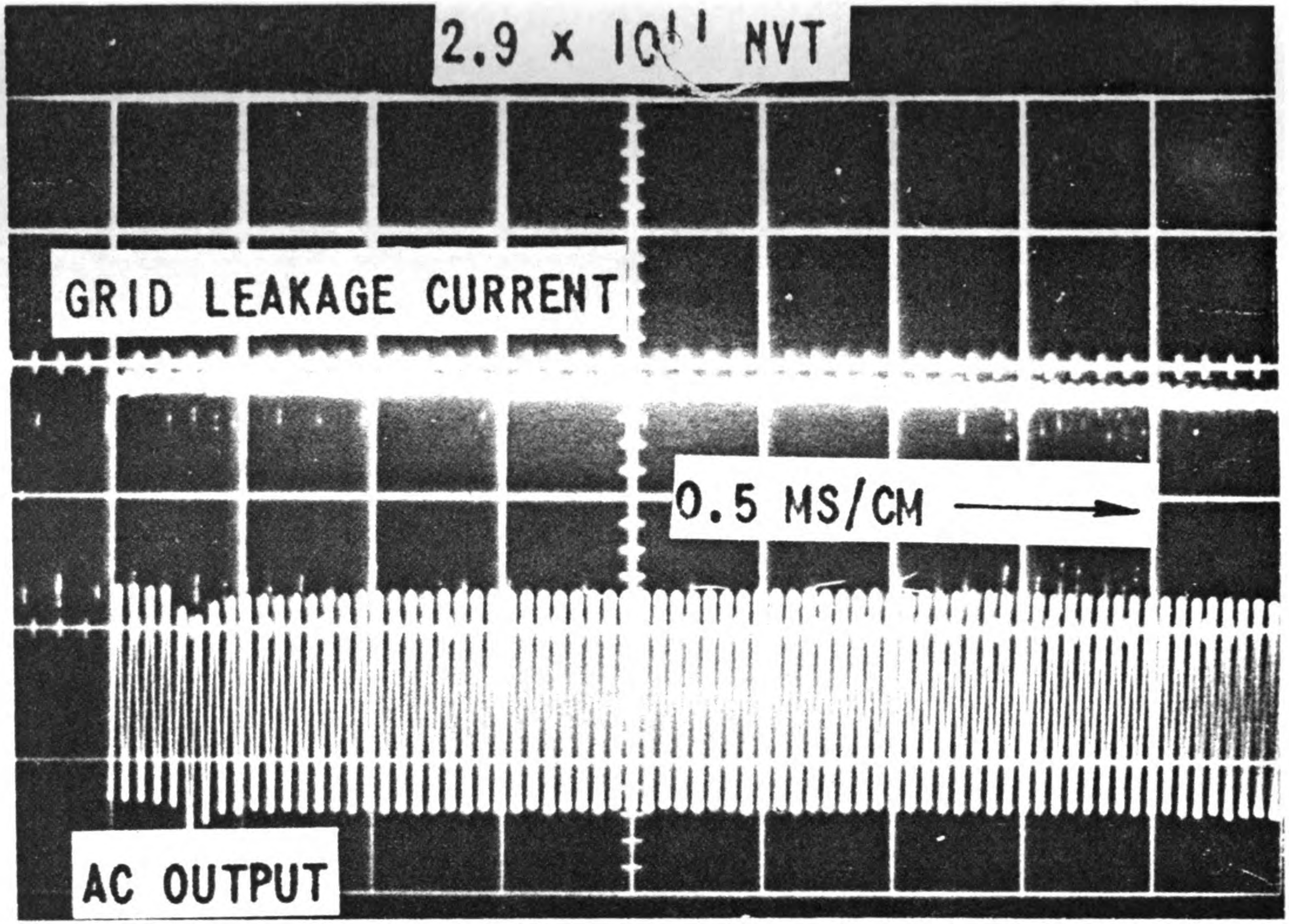
NUVISTOR TRIODE 4 1/2" FROM REACTOR SCREEN

FIG. 24

Fig. 25



NUVISTOR TETRODE AT THE SCREEN



NUVISTOR TETRODE 8 1/2" FROM REACTOR SCREEN

FIG. 26

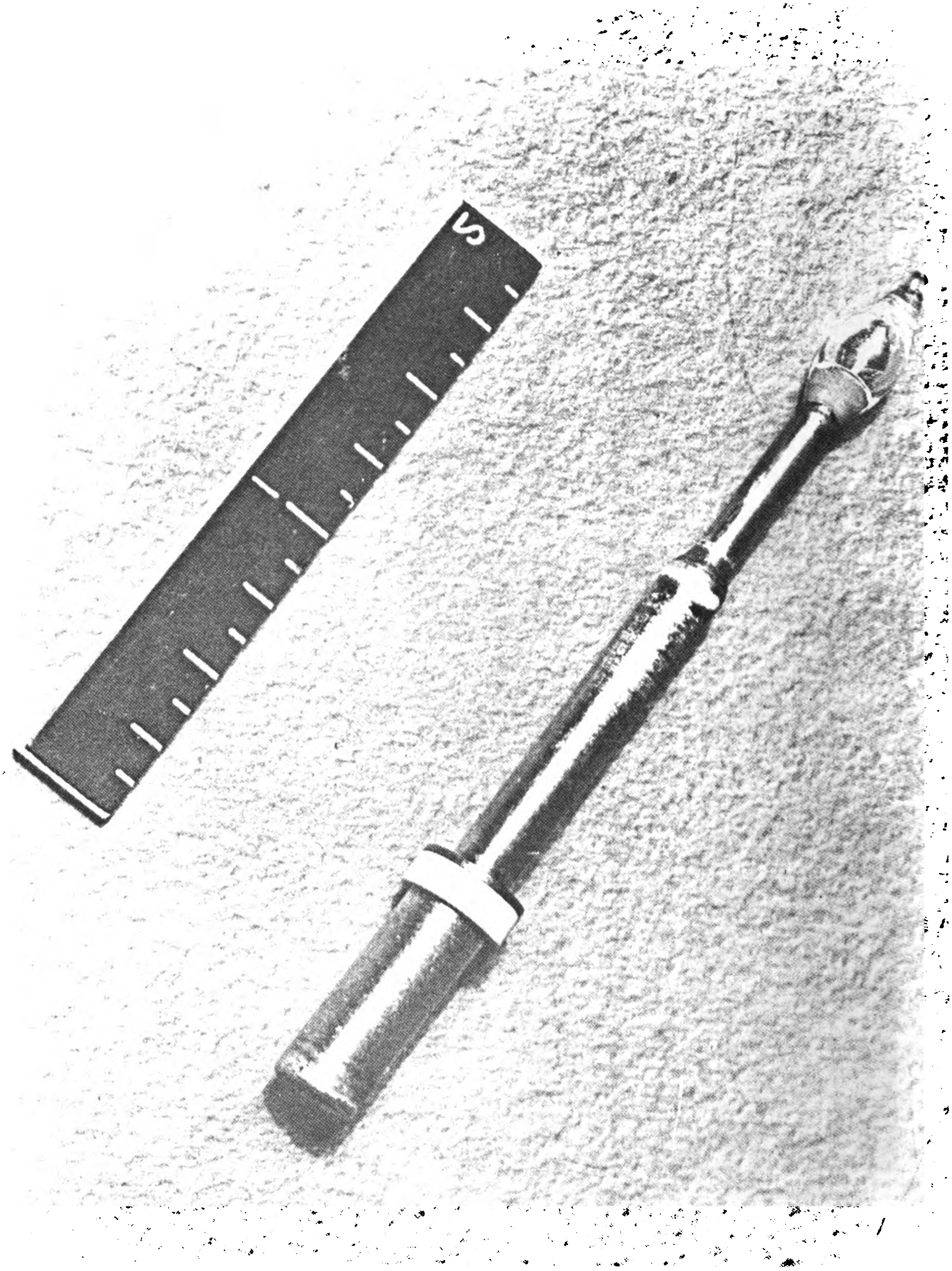
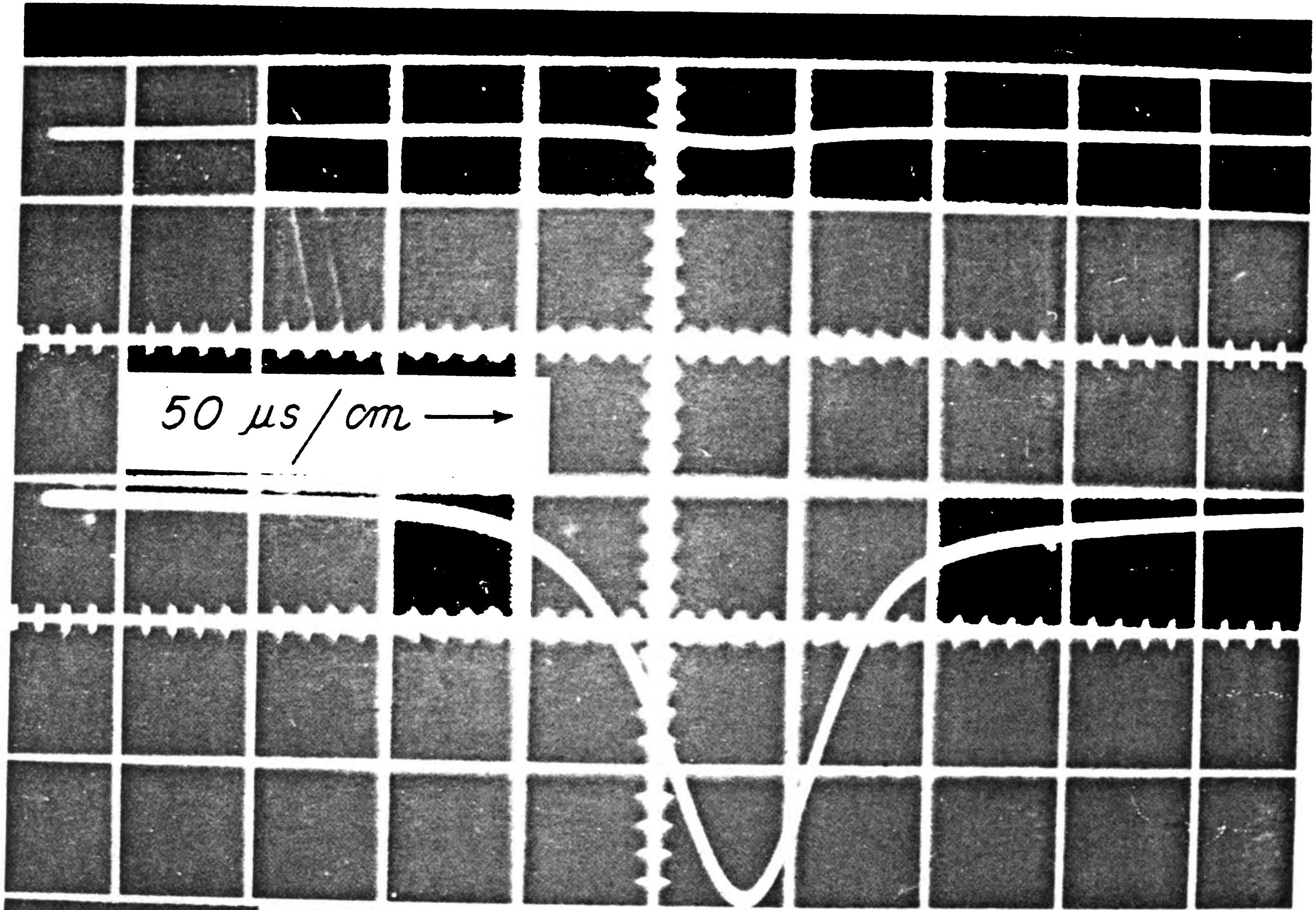
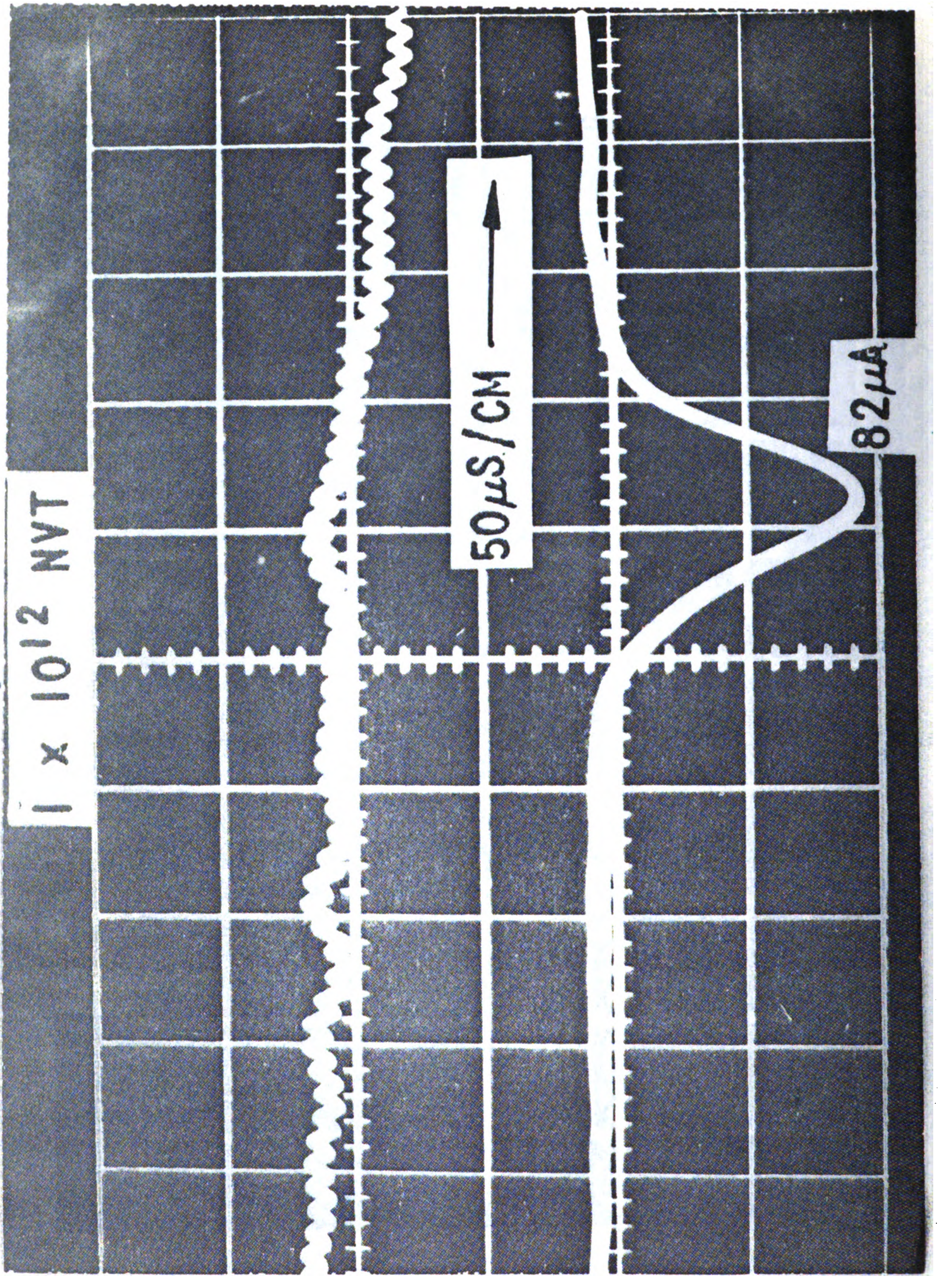


Fig. 27



OUTPUT OF Mg O RAD DURING RADIATION PULSE AT SANDIA REACTOR

Fig. 28



2N797 GE N-P-N

Fig. 29

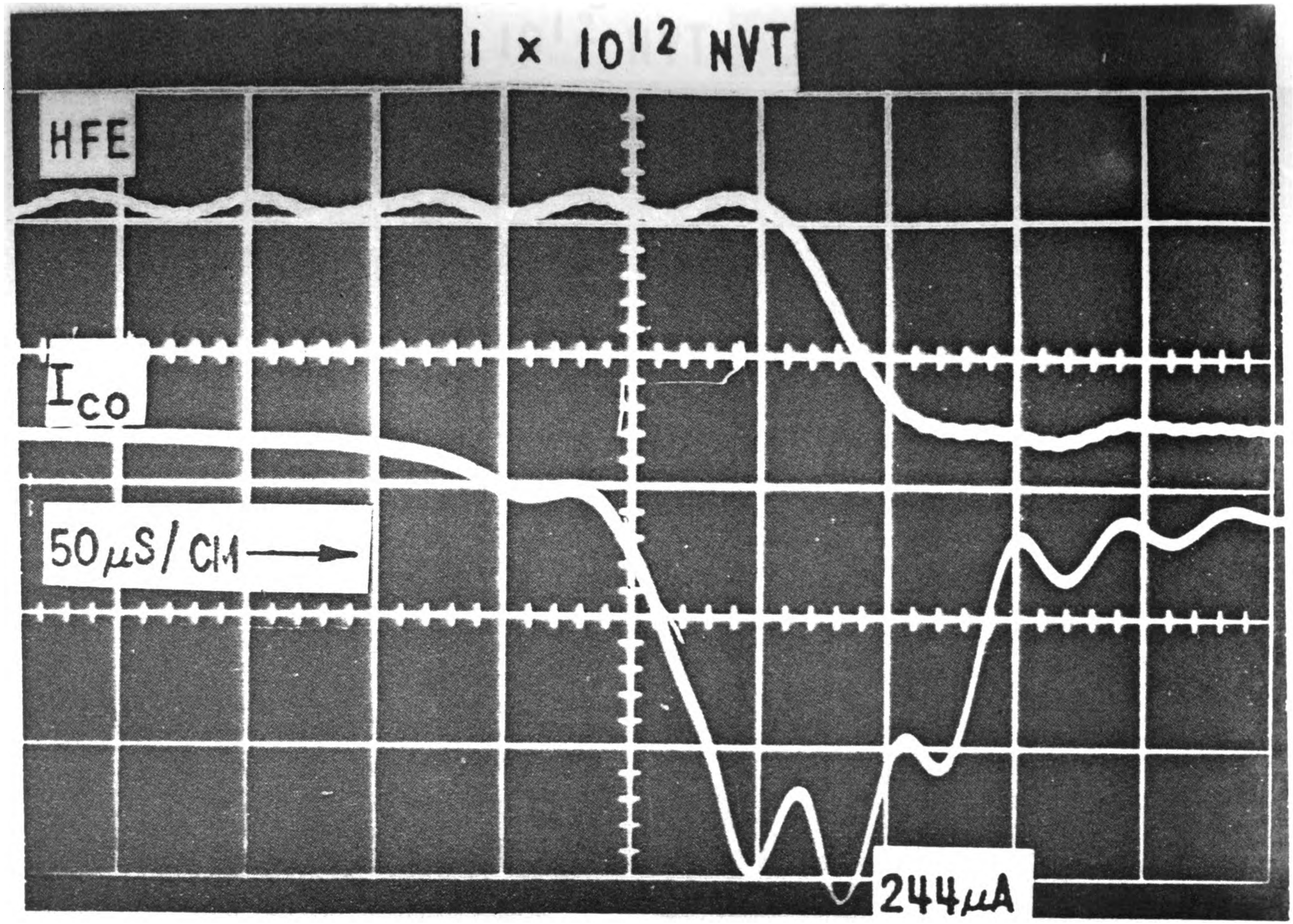
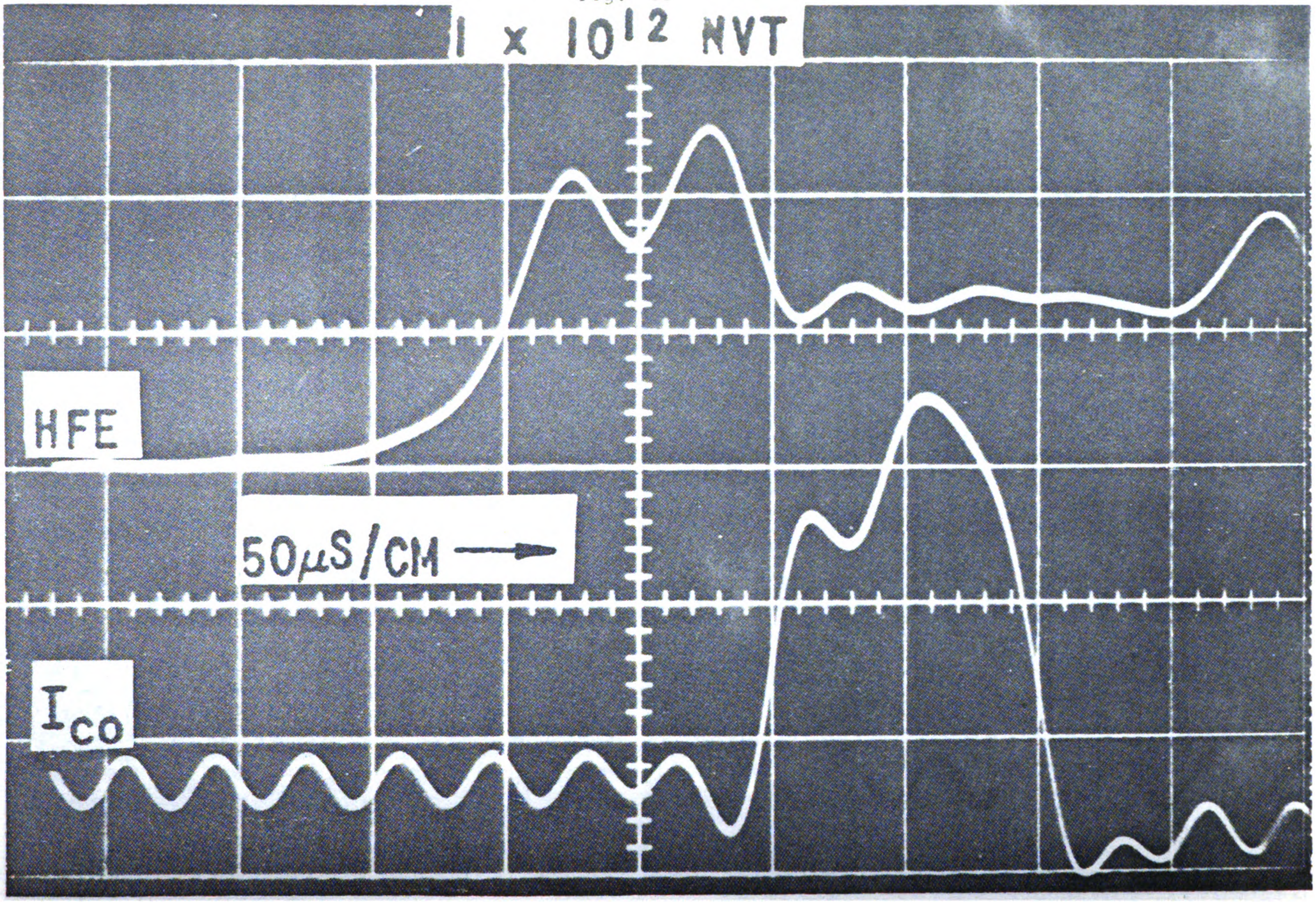


Fig. 30

2N706 Si N-P-N

Fig. 31



2N 1039 Ge p-n-p POWER

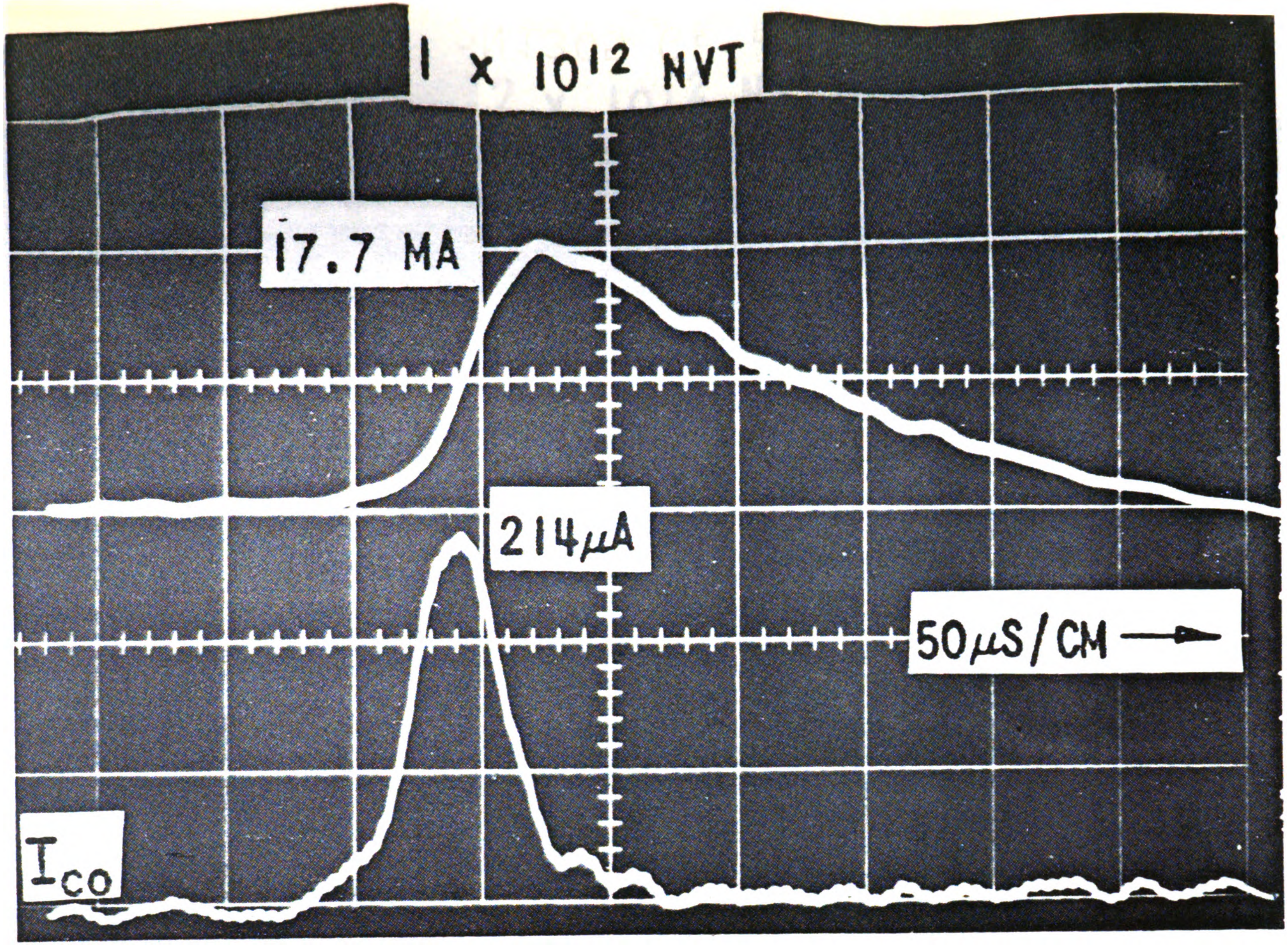
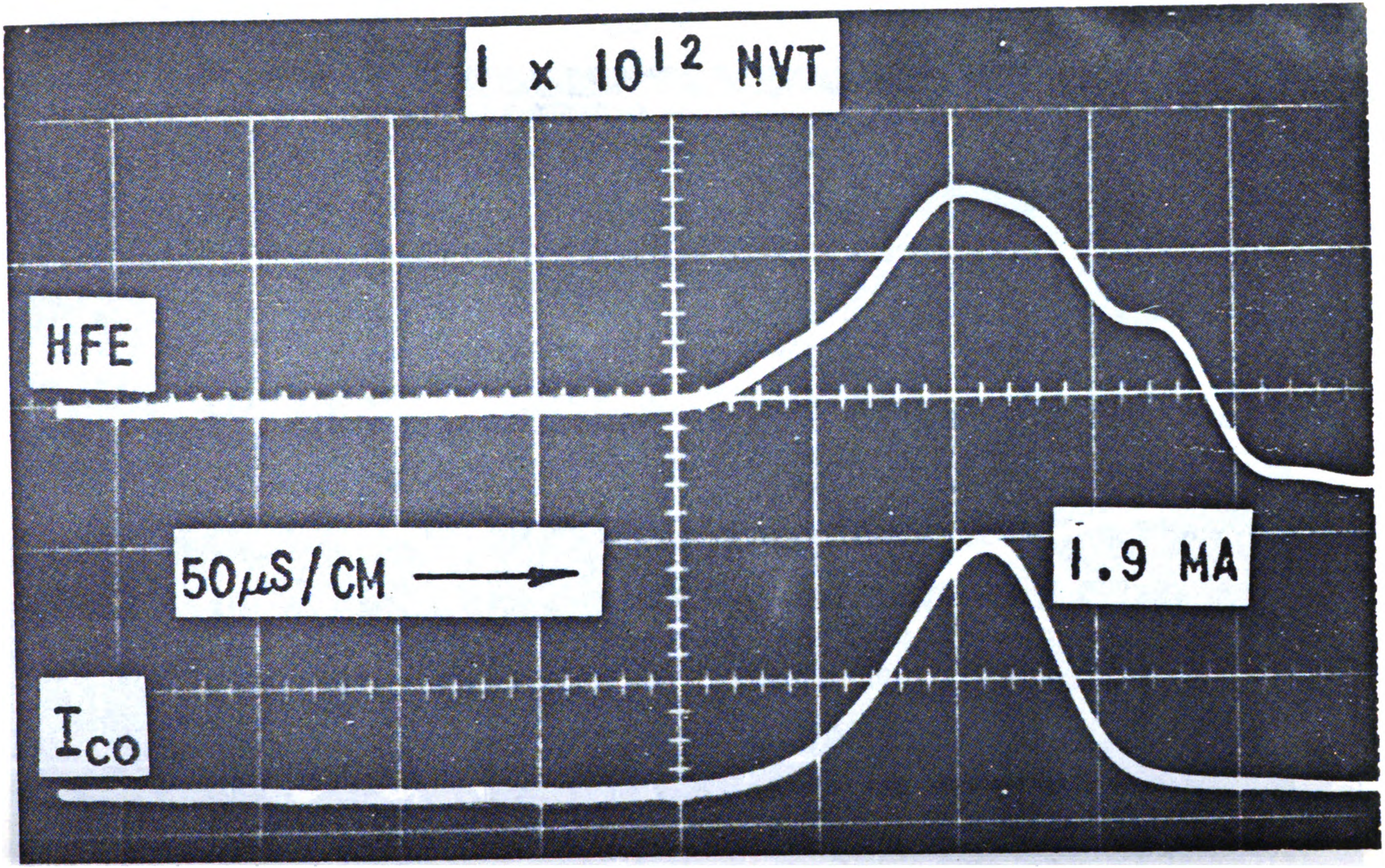


Fig. 32

2N710 Ge PNP

Fig. 33



2N1305 Ge p-n-p

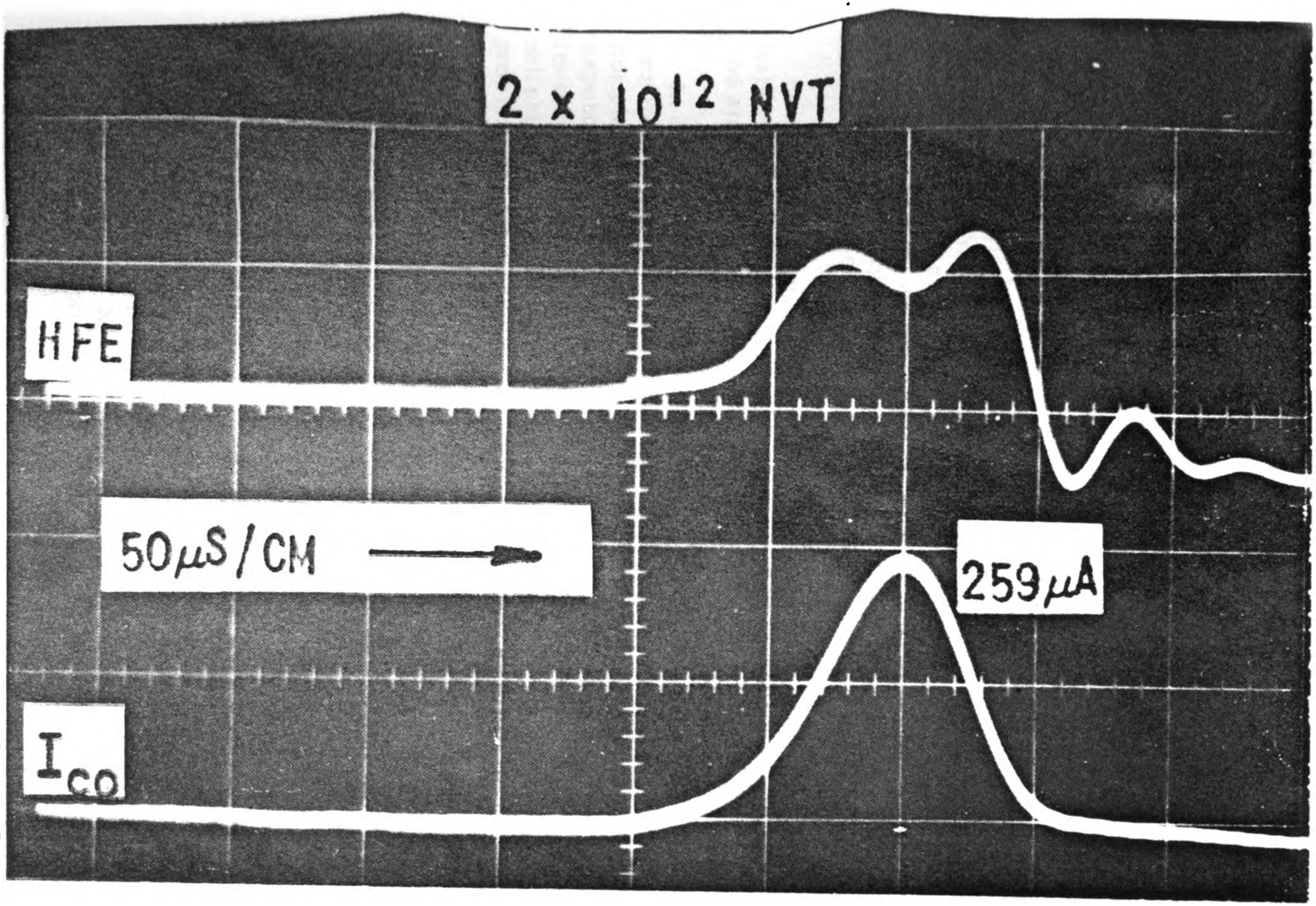


Fig. 34
2N1406 GE P-N-P

RELIABILITY TESTING AND ESTIMATION FOR SINGLE AND MULTIPLE ENVIRONMENTS (Preliminary Report)

S. K. Einbinder
Picatinny Arsenal
Ingram Olkin
Stanford University

1. INTRODUCTION. In this age of complex missiles and costly weapon systems, reliability has become an important objective. Programs for assuring high reliability are now considered a basic part of the development plan for new warheads and fuzes at Picatinny Arsenal. The environmental factors or stresses, such as temperature, vibration, acceleration, rough handling, etc., to which a weapon is subjected are many, and vary widely in level of severity. In addition they may be encountered singly and multiply, simultaneously or in sequence.

Our problem is concerned with the testing and estimation of weapon reliability. The term "weapon" may refer to a "warhead," a "fuze," a "safing and arming mechanism (S and A)," etc.

In order to establish high reliability, large sample sizes are generally required, greater than are usually available in a development program for a complex and expensive item. To obtain the most information with the least expenditure of samples and funds, research is being conducted by many investigators into new and improved statistical methods for solving the reliability or failure estimation problem. There are many phases of the problem that still require a realistic solution.

We first define the basic problem and the quantities we are trying to estimate and then indicate procedures for making point and interval estimates. These procedures will assume that some estimates of the failure distribution parameters are available.

Lastly, a ray method for estimating the distribution parameters for the multivariate stress case is described.

2. UNIVARIATE CASE. We first present the estimation problem for the univariate case. The distribution of failure stress of a weapon or other item may be estimated by testing to failure. The cumulative distribution function of the failure stress X represents the probability that the failure stress is less than x or the proportion of the population whose failure stress is below the value x (Fig. 1) [Figures start on page 275].

If the distribution of applied stresses actually encountered in use, $h(x)$, is known, then the average probability of failure in use is given by:

$$E_h\left[F(X)\right] = \int_R h(x)\, F(x)\, dx \quad ,$$

where $R = \{x : -\infty < x < c\}$ is the region over which the use distribution ranges. In general, the use distribution $h(x)$ is not known. However, upper and sometimes lower limits of the applied stress which the weapon must withstand are generally specified in military specifications. If c represents the upper limit of some stress variable, say temperature, then by the mean value theorem for integrals, it is evident that

$$F(c) \geq E_h\left[F(X)\right] \quad ,$$

i.e., the failure probability at the upper limit of the stress variable is an upper bound for the average probability of failure in use. The point c will be referred to as a critical point or stress. It should also be noted that Reliability = 1 - Probability of Failure. Thus our objective is to determine or estimate

$$F(c) = \int_{-\infty}^{c} dF(x) = \Pr\{X \leq c\} \quad ,$$

where c is a known critical stress. For the normally distributed univariate distribution with mean μ and variance σ^2, the proportion of the population below c is given

$$F(c) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{c} \exp\left[-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right] dx \equiv g_1(\mu, \sigma^2; c) \; .$$

Thus our problem is to estimate g, which is a function of the population parameters, based upon a sample of size N from a normal population.

3. MULTIVARIATE CASE. The problem may be generalized to the multivariate case where the failure distribution is a function of more than one stress variable. A geometrical interpretation is shown in Figure 2 for the bivariate case.

In the case of p variables, where the use distribution is $h(x) \equiv h(x_1, \ldots, x_p)$ and the cumulative failure distribution is $F(x_1, \ldots, x_p) \equiv F(x)$, then the average probability of failure in use under joint action of the p stresses is given by

$$E_h[F(X)] = \int_{-\infty}^{c_1} \cdots \int_{-\infty}^{c_p} h(x)\, F(x)\, dx,$$

where $(c_1, \ldots, c_p)$ represents the upper limit or maximum level of each of the p applied stresses. As before,

$$F(c) \geq E_h[F(X)] \quad .$$

For the multivariate normal distribution, with mean vector $\mu = (\mu_1, \ldots, \mu_p)$ and (positive definite) covariance matrix $\Sigma = (\sigma_{ij}) : p \times p$, our problem then is to estimate

$$F(c) = \frac{1}{(2\pi)^{p/2}|\Sigma|^{1/2}} \int_{-\infty}^{c_1} \cdots \int_{-\infty}^{c_p} \exp\left[-\frac{1}{2}(x-\mu)\Sigma^{-1}(x-\mu)'\right]dx \equiv g_p(\mu, \Sigma; c).$$

Thus, the general problem may be summarized as follows: Based upon a sample of size N from a normal population, it is required to estimate the g functions defined for the univariate and multivariate cases both by point estimation and by confidence limits.

The g functions defined so far were all one sided, i.e., they represent the proportion of the population in one tail of the normal distribution. In reliability work, we are primarily concerned with the one sided case. However, the results can be extended to the two sided case which is of interest in other applications.

4. POINT ESTIMATION. We now consider the problem of obtaining point estimates of the g functions for the univariate and multivariate cases. In Section 4.1 we consider the use of the maximum likelihood estimates of μ and σ^2, and in Section 4.2 give a discussion of uniformly minimum variance unbiased (UMVU) estimators. In order to facilitate the presentation, the mathematical details and derivations are deferred to the Appendix.

4.1 Maximum Likelihood Estimates. Since the sample mean $\bar{x} = \sum_1^N x_i/N$ and sample variance $s^2 = \sum_1^N (x_i - \bar{x})^2/N$ is a maximum likelihood estimate of (μ, σ^2), it is intuitively reasonable to consider $g_1(\bar{x}, s^2; c)$ as an estimator of $g_1(\mu, \sigma^2; c)$, and the following asymptotic results provide a more tangible justification.

Since $g_1(\bar{x}, s^2; c)$ is a function of the sample moments, it follows that $g_1(\bar{x}, s^2; c)$ is asymptotically normally distributed, namely,

$$\text{(1)} \qquad \frac{\sqrt{N}\left[g_1(\bar{x}, s^2; c) - g_1(\mu, \sigma^2; c)\right]}{s\,\varphi\left(\frac{c-\bar{x}}{s}\right)\cdot\left[1 + \frac{N-1}{2N}\left(\frac{c-\bar{x}}{s}\right)^2\right]^{1/2}} \longrightarrow N(0,1) \quad ,$$

where

$$\varphi\left(\frac{z-a}{b}\right) = \frac{1}{\sqrt{2\pi}\, b} \exp\left[-\frac{1}{2}\left(\frac{z-a}{b}\right)^2\right] \qquad \text{(see Appendix).}$$

The latter result may be used for obtaining asymptotic confidence intervals.

In the multivariate case the expressions are more complicated. Let $\bar{x} = (\bar{x}_1, \ldots, \bar{x}_p)$ be the vector of sample means, $S = (s_{ij})$ be the $p \times p$ sample covariance matrix. Then $g_p(\bar{x}, S; c)$ is asymptotically normally distributed, namely,

$$(2) \qquad \frac{\left[g_p(\bar{x}, S; c) - g_p(\mu, \Sigma; c)\right]}{\sqrt{V_\infty(\bar{x}, S)}} \longrightarrow N(0, 1) \ ,$$

where

$$(3) \quad V_\infty(\mu, \Sigma)$$

$$= \frac{1}{N}\sum_{i,j=1}^{p} H_i H_j \sigma_{ij} + \frac{(N-1)}{N^2}\sum_{\substack{i \le j, \\ k \le \ell}} H_{ij}H_{k\ell}(\sigma_{ik}\sigma_{j\ell} + \sigma_{i\ell}\sigma_{jk}) \ ,$$

and where, for example,

$$(4)\quad H_1 = -\varphi\left(\frac{c_1-\mu_1}{\sqrt{\sigma_{11}}}\right)\frac{|\Lambda_{22}|^{1/2}}{(2\pi)^{(p-1)/2}}\int_{-\infty}^{c_2-a_2}\cdots\int_{-\infty}^{c_p-a_p}\exp\left(-\tfrac{1}{2}z\Lambda_{22}z'\right)dz\,,$$

where $a_j = \mu_j + (c_1 - \mu_1)\,\sigma_{ij}/\sigma_{11}$, $j = 2, \ldots, p$,

$$(5)\quad (1+\delta_{ij})H_{ij} = -\lambda_{ij}\,g_p(\mu, \Sigma\,;\,c)$$

$$+\frac{|\Lambda|^{1/2}}{(2\pi)^{p/2}}\int_{-\infty}^{c_1-\mu_1}\cdots\int_{-\infty}^{c_p-\mu_p}\left(\sum_{\alpha} z_\alpha\lambda_{\alpha i}\right)\left(\sum_{\beta} z_\beta\lambda_{\beta j}\right)\exp\left(-\frac{1}{2}z\Lambda z'\right)dz\,,$$

δ_{ij} is the Kronecker delta,

$$\Sigma = \begin{pmatrix}\sigma_{11} & \sigma_1' \\ \sigma_1 & \Sigma_{22}\end{pmatrix} \equiv \begin{pmatrix}\lambda_{11} & \lambda_1' \\ \lambda_1 & \lambda_{22}\end{pmatrix}^{-1} = \Lambda^{-1}$$

(see Appendix). The difficulty in actually carrying out the computations lies in the computations of H_i and H_{ij}, since these involve the tails of a multivariate normal distribution. For $p = 2, 3$, tables are available which permit the computations of the H_i. The determination of the H_{ij} may have to be carried out by Monte Carlo methods. This investigation is still incomplete.

4.2 Uniformly Minimum Variance Unbiased Estimation. UMVU estimators of the g functions have been obtained for a number of cases, and we now present a summary of these results.

When σ^2 is known, the UMVU estimator of $g_1(\mu, \sigma^2; c)$ is

$$\int_{-\infty}^{\sqrt{N/(N-1)}\,(c-\bar{x})/\sigma} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2} t^2\right) dt ,$$

and for σ^2 unknown, the result is

$$\int_0^{\left[\max\ 0,\ \frac{1}{2} - \frac{(\bar{x}-c)\sqrt{N}}{2v\sqrt{N-1}}\right]} \frac{t^{[(N-2)/2]-1}(1-t)^{[(N-2)/2]-1}}{B\left(\frac{N}{2}-1,\ \frac{N}{2}-1\right)}\, dt ,$$

where $v^2 = Ns^2 = \sum (x_i - \bar{x})^2$. These results were obtained by Kolmogorov [1], and Lieberman and Resnikoff [2]. The two-sided univariate case is also given in [2]. Washio, Morimoto, and Ikeda [3] consider the exponential family, rather than just the normal distribution, and give a number of results concerning unbiased estimators. Schmetterer [4] considered a more general type of problem but in the framework of the univariate normal distribution, and this was extended to the multivariate normal case, as well as to other families of distributions, by Ghurye and Olkin [5]. The UMVU estimator for the p-variate normal distribution was obtained in [5] and by Lieberman [6].

When Σ, the population covariance matrix, is known, the UMVU estimator of $g_p(\mu, \Sigma; c)$ is given by

$$\frac{|\Lambda|^{1/2}}{(2\pi)^{p/2}} \int_{-\infty}^{(c_1 - \bar{x}_1)\sqrt{N/(N-1)}} \cdots \int_{-\infty}^{(c_p - \bar{x}_p)\sqrt{N/(N-1)}} \exp\left(-\frac{1}{2}\, t \Lambda t'\right) dt \ ,$$

and when Σ is unknown, the result is

$$\frac{\Gamma\left(\frac{N-1}{2}\right)}{\Gamma\left(\frac{N-p-1}{2}\right)} |V|^{-1/2} \int_R (1 - tV^{-1}t')^{[(N-p-1)/2]-1} dt \ ,$$

where $V = NS$, $R = \left\{ t : -\infty < t_i < (c_i - \bar{x}_i)\sqrt{N/(N-1)},\ i = 1, \ldots, p,\ tV^{-1}t' < 1 \right\}$.

Whereas for the univariate case the UMVU may easily be found from tables of the normal distribution and the incomplete Beta distribution; the integrals for the multivariate case are more troublesome, and numerical methods and approximations may have to be used.

5. CONFIDENCE INTERVALS. Next, we consider the problem of obtaining confidence intervals for the g functions. In the univariate case,

$$g_1(\mu, \sigma^2; c) = \int_{-\infty}^{(c-\mu)/\sigma} \frac{\exp\left(-\frac{1}{2} t^2\right)}{\sqrt{2\pi}} dt \ .$$

From the fact that $\sqrt{N-1}\ (c - \bar{x})/s$, or equivalently, $\sqrt{N(N-1)}\ (c - \bar{x})/v$, where $v^2 = NS^2$, has a non-central t-distribution with $N - 1$ degrees of freedom, and the monotonic nature of the function, we can obtain exact confidence limits using the tables of the non-central t-distribution, [7] or [8].

Resnikoff [9] presents tables for the univariate case, for both one and two-sided tails, based on the UMVU estimator, and gives both point estimates and confidence belts. The confidence coefficients are .90, .95 and .99 for sample sizes 3, 4, 5, 7, 10(5)40, 50, 75, 100, 150, 200.

For the two-sided case and the multivariate one or two-sided case, no direct method is available. One procedure is to use a large sample approximation. We have observed that the use of the maximum likelihood estimates yields asymptotic normality, so that from (1) we obtain the confidence interval

$$g_1(\bar{x}, s^2; c) \pm \frac{z}{\sqrt{N}} s \varphi\left(\frac{c-\bar{x}}{s}\right) \sqrt{\left[1 + \frac{N-1}{2N}\left(\frac{c-\bar{x}}{s}\right)^2\right]} ,$$

with confidence coefficient $1-\alpha$, where z is the $100\alpha\%$ double-tail point of the $N(0,1)$ distribution.

A similar development can be made for the multivariate case by using (2), but the results are more complicated.

Two other procedures have been suggested for the two sided-case, one by Wolfowitz [10], and one by Arnold, which appears in a paper by Wallis [11]. General descriptions of some of these methods may be found in Bowker and Goode [12, Chapter 11], and in [11]. Various computational procedures are outlined in [11].

5.1 <u>Open Problems</u>. In the cases for which alternative procedures of estimation are available, comparisons of the techniques need to be made. Only a first step has been taken in the multivariate case, and more work is required. In particular, where expressions are available, appropriate tables should be prepared.

A second phase is to consider cases where the underlying distribution is not normal. Some alternative distributions have been considered in [3] and [5].

6. RAY METHOD FOR ESTIMATING THE MULTIVARIATE NORMAL DISTRIBUTION. So far this presentation has been concerned essentially with the problem of estimating the tail probabilities of a normal distribution assuming that estimates of the parameters of the normal distribution are available or can be obtained from sample data. In order to use these results in making reliability estimates, it is necessary to be able to estimate the distribution parameters. In the case of the normal distribution, we need estimates of the mean vector and the covariance matrix. Practical and efficient methods are required for obtaining estimates of the distribution parameters which do not involve excessively large samples.

We next describe some preliminary results of a ray method for estimating the parameters of the multivariate normal distribution, which may permit fewer observations to be made. There are still a number of open questions, and we do not know how good the method is.

To simplify the discussion, let us first confine our remarks to the bivariate normal case. The extension to the multivariate case will be described later.

First, consider the model for the case where an object is subjected jointly to two stresses, for example temperature and vibration. We assume that the random stresses (X_1, X_2) at which failure occurs for the population of objects has a bivariate normal distribution $BVN(\mu, \Sigma)$. The failure stress may also be called the strength of an item. Let (ξ_1, ξ_2) be the levels of the applied stresses X_1 and X_2, respectively. Assume that all items in the population whose strength X_1 or X_2, is less than the respective applied stresses ξ_1, ξ_2 will fail under these loading conditions. Then the proportion of the population that will fail under the applied stress ξ_1, ξ_2 is given by

$$F(\xi_1, \xi_2) = 1 - \int_{\xi_1}^{\infty}\int_{\xi_2}^{\infty} \frac{e^{-\frac{1}{2}(x-\mu)\Sigma^{-1}(x-\mu)'}}{2\pi|\Sigma|^{1/2}} \, dx \, ,$$

which also represents the probability of failure for the given applied stresses.

Figure 3 depicts a correlated BVN failure stress distribution which we wish to estimate. The distribution is defined by five parameters: two means, μ_1 and μ_2, and three covariances, σ_{11}, σ_{12}, and σ_{22}. By testing items to failure along the horizontal ray w_0 which is selected to lie essentially below the entire distribution function, we get the marginal distribution of X_1 under our failure definition. Similarly, if the vertical ray $w_{\pi/2}$ is properly selected sufficiently to the left of the distribution, we get the marginal distribution of X_2. Thus, for the BVN case, we can obtain estimates of the marginal means and variances of X_1 and X_2, respectively, which are the unknown means and variances of the failure distribution function. With two rays, therefore, it is possible to estimate four of the required five parameters. In order to estimate the remaining parameter, the covariance σ_{12}, tests along another ray are required. The best ray w along which to test appears to be the one passing through the mean of the distribution. By testing-to-failure along the ray w, we mean that the applied stresses are increased along w until failure occurs. According to our definition, failure will occur when either of the applied stresses exceeds its respective strength as defined by the BVN distribution.

If we start at $w = -\infty$ and increase the stress along w until failure occurs, the probability of failure in the region Δw is given by

$$p(w)\,\Delta w = \Delta x_1 \int_{w \sin\alpha + x_{20}}^{\infty} \frac{e^{-\frac{1}{2}(x-\mu)\Sigma^{-1}(x-\mu)'}}{2\pi|\Sigma|^{1/2}}\, dx_2$$

$$+ \; \Delta x_2 \int_{w \cos\alpha + x_{10}}^{\infty} \frac{e^{-\frac{1}{2}(x-\mu)\Sigma^{-1}(x-\mu)'}}{2\pi|\Sigma|^{1/2}}\, dx_1 \,.$$

This equation represents the proportion of the population strengths or items whose failure stresses lie in the shaded region shown in Fig. 4.

Solution of the above equation (see Appendix) results in the following expression for the probability density function of the failure stress

$$
\text{(6)} \qquad p(w) = \cos\alpha \cdot \varphi\left(\frac{w\cos\alpha + x_{10} - \mu_1}{\sqrt{\sigma_{11}}}\right) \cdot \left[1 - \Phi(d_1)\right]
$$

$$
+ \sin\alpha \cdot \varphi\left(\frac{w\sin\alpha + x_{20} - \mu_2}{\sqrt{\sigma_{22}}}\right) \cdot \left[1 - \Phi(d_2)\right] ,
$$

where $\varphi\left(\frac{z-a}{b}\right) = \frac{1}{\sqrt{2\pi}\, b} \exp - \frac{1}{2}\left(\frac{z-a}{b}\right)^2$, $\Phi(z) = \int_{-\infty}^{z} \varphi(\tau)\, d\tau$,

$$
d_1 = \frac{w\sin\alpha + x_{20} - \mu_2}{\sqrt{\sigma_{22}(1-\rho^2)}} - \rho\, \frac{(w\cos\alpha + x_{10} - \mu_1)}{\sqrt{\sigma_{11}(1-\rho^2)}}
$$

$$
d_2 = \frac{w\cos\alpha + x_{10} - \mu_1}{\sqrt{\sigma_{11}(1-\rho^2)}} - \rho\, \frac{(w\sin\alpha + x_{20} - \mu_2)}{\sqrt{\sigma_{22}(1-\rho^2)}}
$$

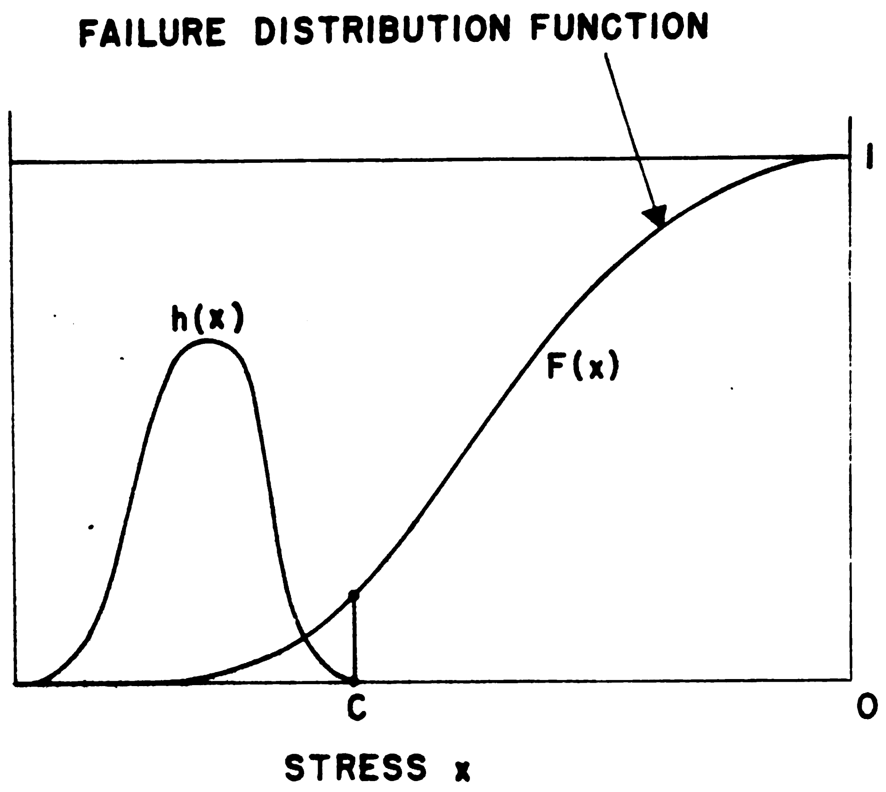
Figure 4 shows $p(w)$ for several values of the covariance σ_{12} while holding the mean vector and the variances constant at the values indicated.

It is evident that p(w) is asymmetric, the amount of asymmetry depending upon the correlation between the variables X_1, X_2. The mode of the distribution also depends on the correlation between the variables. The next question is to determine the best estimator for ρ or σ_{12} based upon a sample of the failure strengths W along the ray w_α and assuming that the other distribution parameters μ_1, μ_2, σ_{11}, σ_{22} are known or their estimates are available. This problem is still unsolved, but it does appear that ρ should be estimable from tests along the ray w. After we estimate ρ, the entire BVN distribution will be defined and may be used for making estimates of the g functions as described earlier.

Suppose we have a multivariate normal (MVN) distribution in p variates, then we will have p unknown means μ_i and $p(p + 1)/2$ unknown covariances σ_{ij}. By using $p(p + 1)/2$ rays we will be able to estimate all of the parameters. We use p rays, which we call principal rays, to estimate the p marginal means μ_i and variances σ_{ii} and $p(p - 1)/2$ rays to evaluate all of the covariances $\sigma_{ij} (i \neq j)$. Just as in the BVN case, it only appears necessary to consider two stress variables at a time in order to estimate the covariance of these two variables. The procedure is repeated for all possible combinations of the variables.

Thus, assuming that the failure model described is valid, an attempt has been made to estimate the MVN distribution by testing along rays with the expectation that this procedure may be more efficient in general than procedures which involve mapping out the failure distribution surface.

It was assumed that the failure distribution along rays would be obtained by testing to failure. In many practical applications, testing to failure is impossible, especially where more than one applied stress or environment is involved. Thus only success or failure in functioning properly may be observed after subjection to the environmental conditions. Under these circumstances, testing along rays is desirable because it may permit use of some sort of sensitivity type of experiment for estimating the failure distribution along the ray. This information, in turn, may then be used to estimate the desired MVN distribution parameters as described. The distribution functions derived in this paper are based on testing-to-failure along a ray w. If this procedure is not followed, the distributions may be different from those shown. Again, our work in connection with this problem

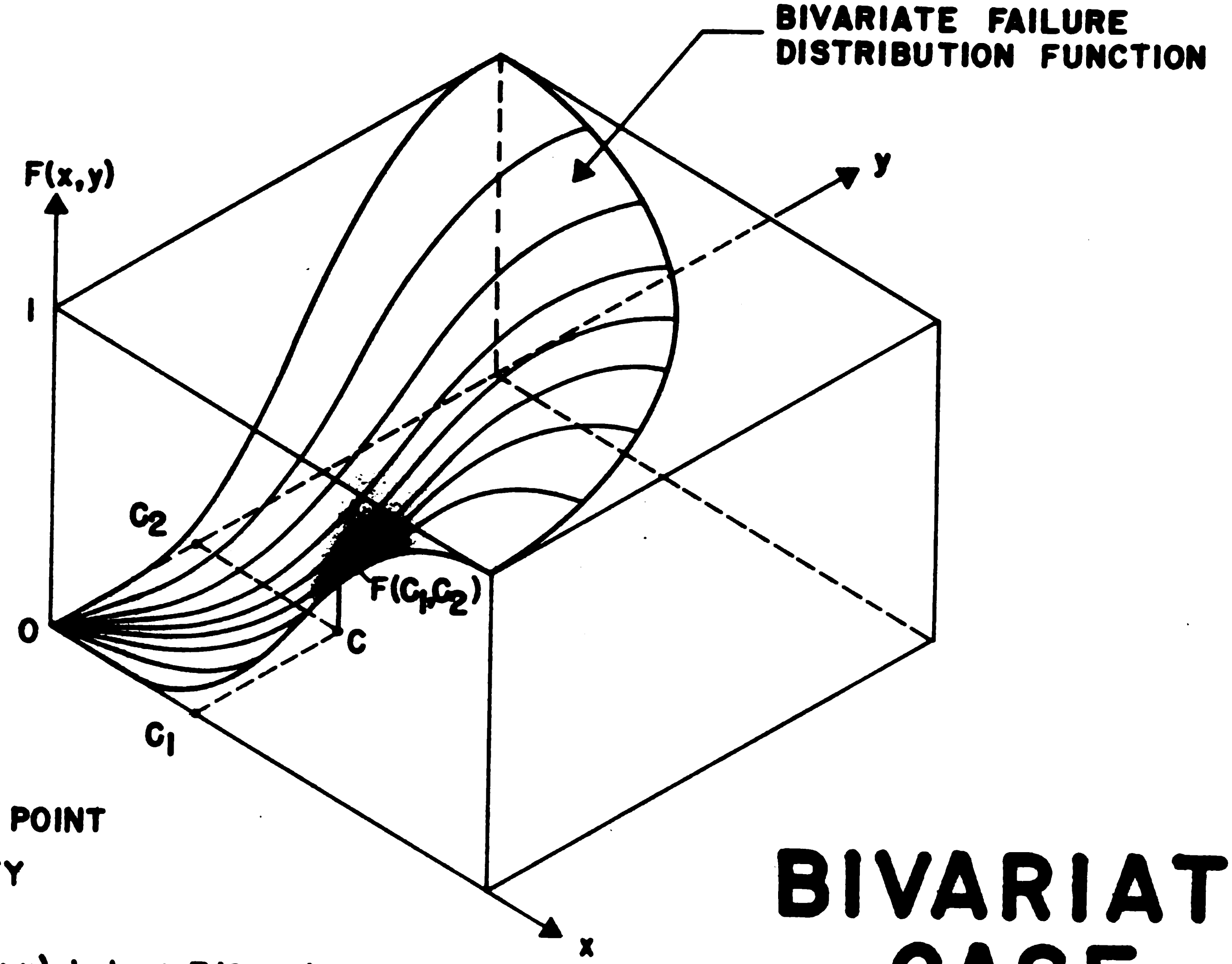


AVERAGE PROBABILITY
OF FAILURE IN USE:

$$E_h[F(X)] = \int_{-\infty}^{C} h(x)\,F(x)\,dx \leq F(C)$$

UNIVARIATE CASE

FIG 1



C = (C1, C2) = CRITICAL POINT

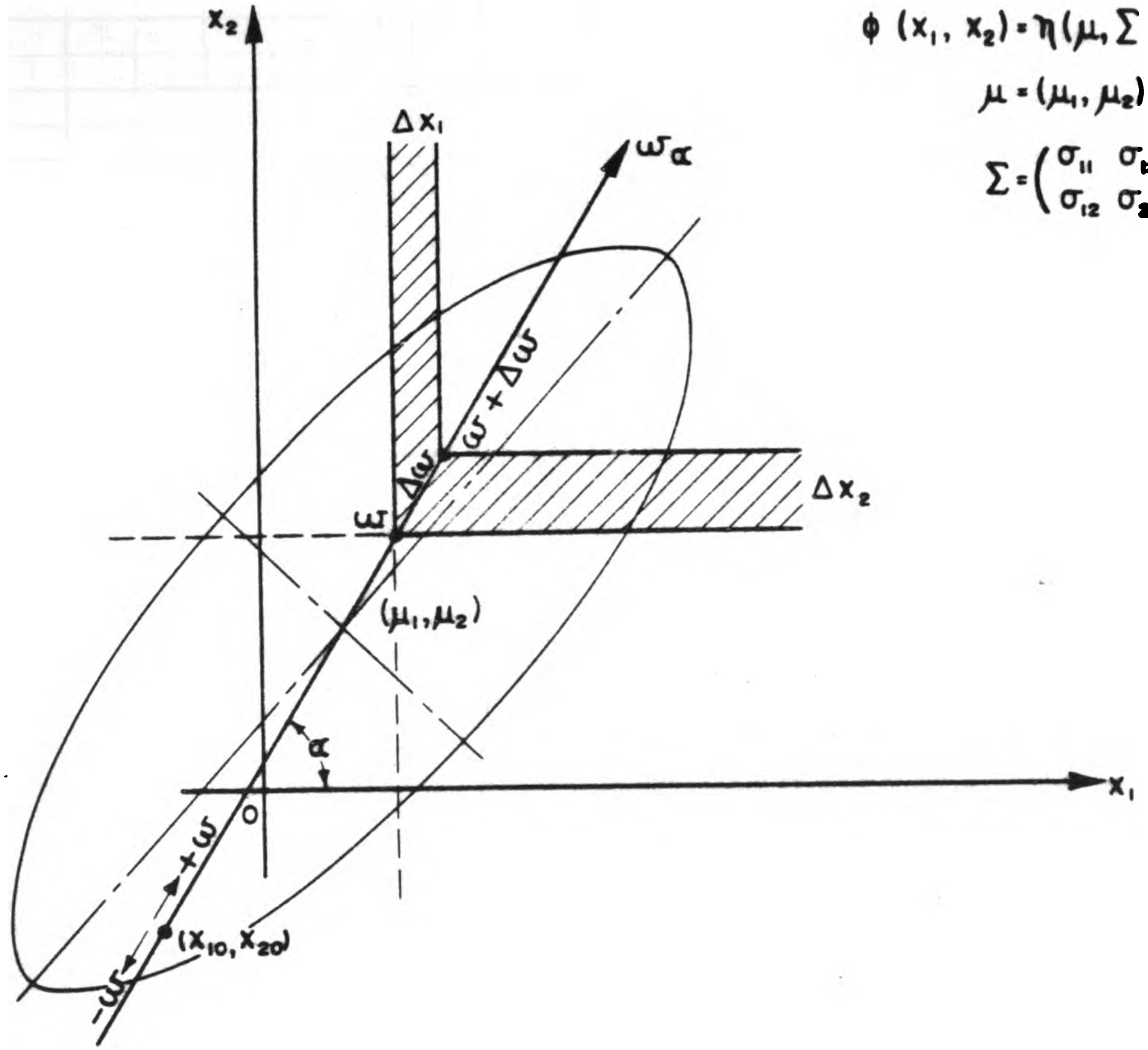
AVERAGE PROBABILITY OF FAILURE IN USE:

$$E_h[F(x,y)] = \int_0^{C_1}\int_0^{C_2} h(x,y)\,F(x,y)\,dx\,dy \leq F(C_1,C_2)$$

BIVARIATE CASE

FIG 2

TESTING ALONG RAY, ω_α

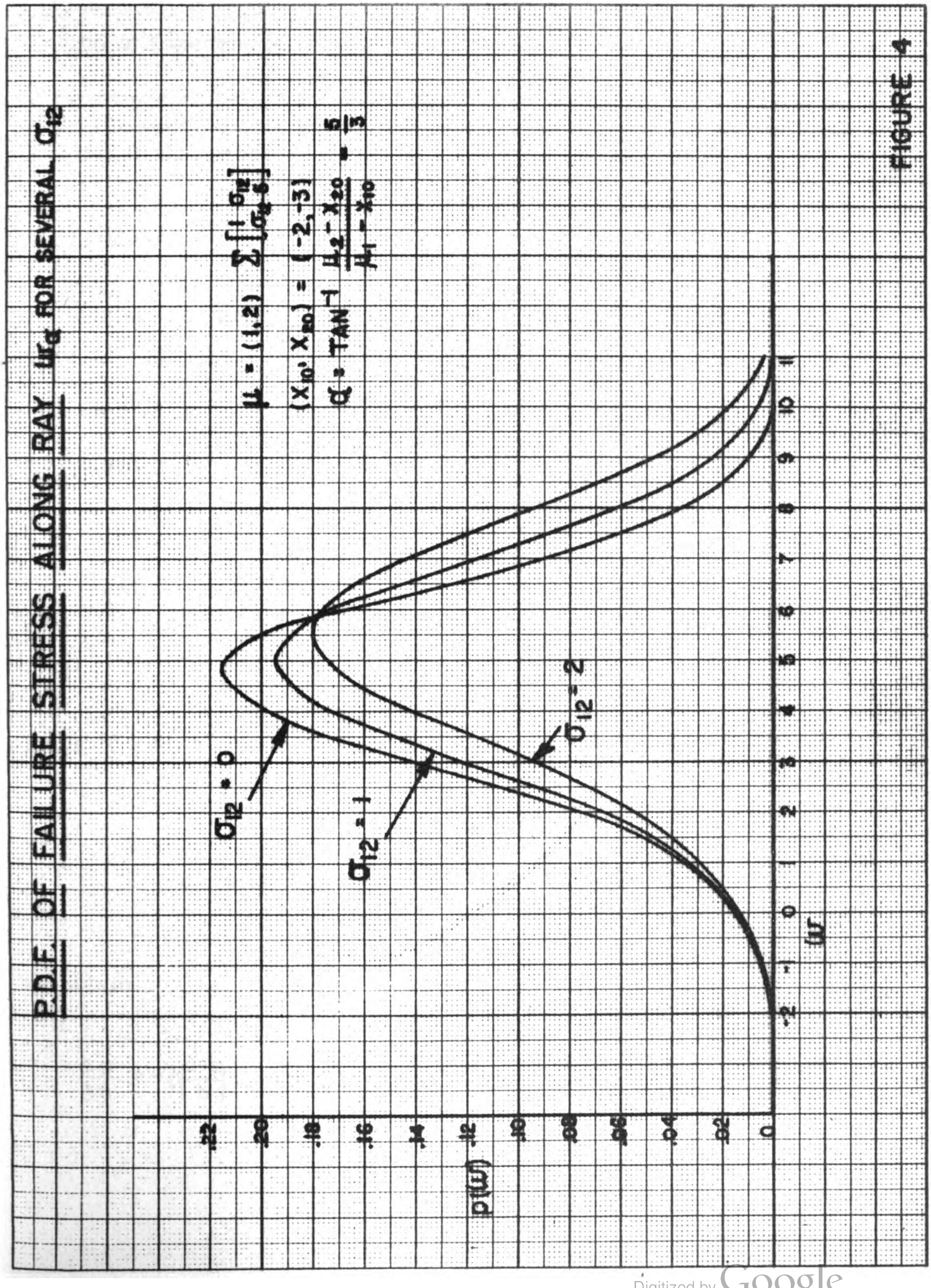


$$\phi(x_1, x_2) = \eta(\mu, \Sigma)$$

$$\mu = (\mu_1, \mu_2)$$

$$\Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{pmatrix}$$

FIG 3



P.D.F. OF FAILURE STRESS ALONG RAY ω_α FOR SEVERAL σ_{12}

FIGURE 4

is far from complete. We have attempted to summarize the status of our efforts toward a practical solution of the reliability testing and estimation problem in this paper.

Appendix A

To simplify the presentation, the univariate and multivariate results are treated separately. We first obtain the asymptotic distribution of $g_1(\bar{x}, s^2; c)$. Since $g_1(\bar{x}, s^2; c)$ is a function of the sample moments it follows, [13, p. 354, 366], that $g_1(\bar{x}, s^2; c)$ is asymptotically normal with mean $g_1(\mu, \sigma^2; c)$ and variance

$$\left(\frac{\partial g}{\partial \bar{x}}\bigg|_{\mu,\sigma^2}\right)^2 \cdot \mathrm{Var}(\bar{x}) + \left(\frac{\partial g}{\partial s^2}\bigg|_{\mu,\sigma^2}\right)^2 \cdot \mathrm{Var}(s^2) + 2\left(\frac{\partial g}{\partial x}\bigg|_{\mu,\sigma^2}\right) \mathrm{Cov}(\bar{x}, s^2).$$

Since $\bar{x}$ and s^2 are independent, $\mathrm{Cov}(\bar{x}, s^2) = 0$. From

$$g_1(a,b;c) = \frac{1}{\sqrt{2\pi}\, b} \int_{-\infty}^{c} \exp\left[-\frac{1}{2}\,\frac{(t-a)^2}{b^2}\right] dt ,$$

we obtain

$$\frac{\partial g}{\partial a} = -\varphi\left(\frac{c-a}{b}\right), \quad \frac{\partial g}{\partial b^2} = \frac{\partial g}{\partial b}\,\frac{1}{2b} = -\frac{1}{2b}\left(\frac{c-a}{b}\right)\varphi\left(\frac{c-a}{b}\right) ,$$

where

$$\varphi\left(\frac{u-a}{b}\right) = \frac{1}{\sqrt{2\pi}\, b} \exp\left[-\frac{1}{2}\left(\frac{u-a}{b}\right)^2\right].$$

Also, $\mathrm{Var}(\bar{x}) = \sigma^2/N$, $\mathrm{Var}(s^2) = 2(N-1)\sigma^4/N^2$, and hence the asymptotic variance is

$$V_\infty(\mu, \sigma^2) = \varphi^2\left(\frac{c-\mu}{\sigma}\right)\frac{\sigma^2}{N}\left[1 + \frac{N-1}{2N}\left(\frac{c-\mu}{\sigma}\right)^2\right].$$

Since $V_\infty(\bar{x}, s^2)$ is a rational function of the sample moments $\bar{x}$, s^2, it follows by Slutsky's Theorem [13, p. 255], that $V_\infty(\bar{x}, s^2)$ converges in probability to $V_\infty(\mu, \sigma^2)$, and hence, [13, p. 254], that

$$\frac{g_1(\bar{x}, s^2; c) - g_1(\mu, \sigma^2; c)}{V_\infty(\bar{x}, s^2)} \to N(0,1).$$

We now consider the multivariate case, and adopt the notation $\mu = (\mu_1, \ldots, \mu_p)$, $\Sigma = (\sigma_{ij}) : p \times p$, $\Lambda = \Sigma^{-1}$, $\bar{x} = (\bar{x}_1, \ldots, \bar{x}_p)$, $S = (s_{ij}) : p \times p$. Here $\bar{x}$ is the vector of sample means, and S is the matrix of sample covariances, obtained from a sample of size N.

We now have

$$g_p(\bar{x}, S; c) = \frac{1}{(2\pi)^{p/2}|S|^{1/2}} \int_{-\infty}^{c_1} \cdots \int_{-\infty}^{c_p} \exp\left[-\frac{1}{2}(t-\bar{x})S^{-1}(t-\bar{x})'\right] dt .$$

Define

$$H_i = \frac{\partial g}{\partial \bar{x}_i}\bigg|_{\mu,\Sigma}, \quad H_{ij} = \frac{\partial g}{\partial s_{ij}}\bigg|_{\mu,\Sigma};$$

then by the same argument as the univariate case, $g_p(\bar{x}, S; c)$ is asymptotically normal with mean $g_p(\mu, \Sigma; c)$ and variance

$$V_\infty(\mu, \Sigma) = \sum_{i,j} H_i H_j \operatorname{Cov}(\bar{x}_i, \bar{x}_j) + \sum_{i \le j, k \le \ell} H_{ij} H_{k\ell} \operatorname{Cov}(s_{ij}, s_{k\ell}) .$$

The terms involving $\operatorname{Cov}(x_i, s_{jk})$ are zero and have been omitted. We first note that

$$\operatorname{Cov}(\bar{x}_i, \bar{x}_j) = \sigma_{ij}/N, \ \operatorname{Cov}(s_{ij}, s_{k\ell}) = (\sigma_{ik}\sigma_{j\ell} + \sigma_{i\ell}\sigma_{jk})(N-1)/N,$$

[14, p. 161].

The evaluation of H_1 yields

$$H_1 = -\frac{1}{(2\pi)^{p/2} |\Sigma|^{1/2}}$$

$$\times \int_{-\infty}^{c_2-\mu_2} \cdots \int_{-\infty}^{c_p-\mu_p} \exp\left[-\frac{1}{2}(c_1-\mu_1, t_2, \ldots, t_p)\Lambda(c_1-\mu_1, t_2, \ldots, t_p)'\right] dt_2, \ldots, dt_p .$$

If we partition $\Lambda = \begin{pmatrix} \lambda_{11} & \lambda_1 \\ \lambda_1' & \Lambda_{22} \end{pmatrix}$, $\lambda_{.1} = (\lambda_{12}, \ldots, \lambda_{1p})$, complete the square, note that

$$\lambda_{11} - \lambda_1 \Lambda_{22}^{-1} \lambda_1' = 1/\sigma_{11}, \ |\Sigma| = \sigma_{11} \left| \Sigma_{22} - \frac{\sigma_1' \sigma_1}{\sigma_{11}} \right| = \sigma_{11} |\Lambda_{22}|^{-1},$$

and simplify, we obtain

$$H_1 = -\frac{e^{\left[-\frac{1}{2}\frac{(c_1-\mu_1)^2}{\sigma_{11}}\right]}}{\sigma_{11}^{1/2}\sqrt{2\pi}} \cdot \frac{|\Lambda_{22}|^{1/2}}{(2\pi)^{(p-1)/2}} \int_{-\infty}^{c_2-a_2} \cdots \int_{-\infty}^{c_p-a_p} \exp(-\tfrac{1}{2} z \Lambda_{22} z')\, dz,$$

where $a_j = \mu_j + (c_1 - \mu_1)\sigma_{1j}/\sigma_{11}$, $j = 2, \ldots, p$, i. e.,

$$H_1 = -\varphi\left(\frac{c_1-\mu_1}{\sqrt{\sigma_{11}}}\right) g_{p-1}\left((a_2, \ldots, a_p), \Lambda_{22}^{-1}; (c_2, \ldots, c_p)\right).$$

We now evaluate H_{ij}, but first write

$$g_p(\mu, \Sigma; c) = \frac{1}{(2\pi)^{p/2}} \int_{-\infty}^{c_1} \cdots \int_{-\infty}^{c_p} \exp\left[-\frac{1}{2} \operatorname{tr} \Lambda B + \frac{1}{2} \log|\Lambda|\right] dt,$$

where $B = (t-\mu)'(t-\mu)$. Also

$$\frac{\partial g}{\partial \sigma_{ij}} = \sum_{\alpha \le \beta} \frac{\partial g}{\partial \lambda_{\alpha\beta}} \frac{\partial \lambda_{\alpha\beta}}{\partial \sigma_{ij}} .$$

But

$$\frac{\partial \lambda_{\alpha\beta}}{\partial \sigma_{ij}} = \begin{cases} -(\lambda_{\alpha i}\lambda_{j\beta} + \lambda_{\alpha j}\lambda_{i\beta}), & i \neq j \\ -\lambda_{\alpha i}\lambda_{i\beta}, & i = j, \end{cases}$$

and

$$\frac{\partial g}{\partial \lambda_{\alpha\beta}} = \frac{|\Lambda|^{1/2}}{(2\pi)^{p/2}} \int_{-\infty}^{c_1} \cdots \int_{-\infty}^{c_p} e^{\left[-\frac{1}{2}\operatorname{tr}\Lambda B\right]} \frac{1}{2}\left[-b_{\alpha\beta} + \sigma_{\alpha\beta}\right] dt,$$

$$\frac{\partial g}{\partial \lambda_{\alpha\alpha}} = \frac{|\Lambda|^{1/2}}{(2\pi)^{p/2}} \int_{-\infty}^{c_1} \cdots \int_{-\infty}^{c_p} e^{-\frac{1}{2}\operatorname{tr}\Lambda B} \left[-b_{\alpha\alpha} + \sigma_{\alpha\alpha}\right] dt.$$

After collecting terms we obtain

$$(1 + \delta_{ij})H_{ij} = -\lambda_{ij}\, g_p(\mu, \Sigma; c)$$

$$+ \frac{|\Lambda|^{1/2}}{(2\pi)^{p/2}} \int_{-\infty}^{c_1-\mu_1} \cdots \int_{-\infty}^{c_p-\mu_p} \left(\sum_{\alpha} t_{\alpha}\lambda_{\alpha i}\right)\left(\sum_{\beta} t_{\beta}\lambda_{\beta j}\right) e^{\left[-\frac{1}{2}t\Lambda t'\right]} dt,$$

where δ_{ij} is the Kronecker delta.

Appendix B

In the present section we derive formula (6). Consider a BVN(μ, Σ), and a ray $w = x_1 + x_2$, where the slope is given by $\tan\alpha = (x_2 - x_{20})/(x_1 - x_{10})$ (see Figure 3). The probability density of failures along this ray is given by

$$p(w)dw = dx_1 \int_a^{\infty} \frac{e^{-\frac{1}{2}(x-\mu)\Lambda(x-\mu)'}}{(2\pi)|\Sigma|^{1/2}} dx_2 + dx_2 \int_b^{\infty} e^{-\frac{1}{2}(x-\mu)\Lambda(x-\mu)'} dx_1$$

$$\equiv \quad dx_1 A + dx_2 B ,$$

where $a = w \sin\alpha + x_{20}$, $b = w \cos\alpha + x_{10}$. Let $y_j = x_j - \mu_j$, $j = 1, 2,$ then by completing the square,

$$A = \frac{\exp\left(-\frac{1}{2} y_1^2 \frac{|\Lambda|}{\lambda_{22}}\right)}{(2\pi)|\Sigma|^{1/2}} \int_{a-\mu_2}^{\infty} \exp\left[-\frac{1}{2}(y_2 \sqrt{\lambda_{22}} + y_1 \lambda_{12}/\sqrt{\lambda_{11}})^2\right] dy_2$$

$$= \frac{\exp\left(-\frac{1}{2} \cdot \frac{y_1^2}{\sigma_{11}}\right)}{\sqrt{2\pi}\,\sigma_{11}} \int_d^{\infty} \exp(-\tfrac{1}{2} z^2)\, dz,$$

where

$$d = (a - \mu_2)\sqrt{\lambda_{22}} + y_1 \lambda_{12} / \sqrt{\lambda_{11}}$$

$$= \frac{(w \sin\alpha + x_{20} - \mu_2)}{\sqrt{\sigma_{22}(1-\rho^2)}} - \frac{\rho(w \cos\alpha + x_{10} - \mu_1)}{\sqrt{\sigma_{11}(1-\rho^2)}} .$$

Hence

$$A = \varphi\left(\frac{w \cos\alpha + x_{10} - \mu_1}{\sqrt{\sigma_{11}}}\right) \cdot \left[1 - \Phi(d)\right] .$$

By a similar reduction we obtain the expression for B. Formula (6) follows by combining results.

REFERENCES

[1] Kolmogorov, A. N., "Unbiased Estimates," Izv. Akad. Nauk SSSR. Ser. Mat., 14 (1950) 303-326. Amer. Math. Soc. Trans. No. 98.

[2] Lieberman, Gerald J. and Resnikoff, George J., "Sampling Plans for Inspection by Variables," J. Amer. Statist. Assoc., 50 (1955) 457-516.

[3] Washio, Y., Morimoto, H., and Ikeda, N., "Unbaised Estimation based on Sufficient Statistics," Bull. Math. Statist., 6 (1956) 69-93.

[4] Schmetterer, L., "On a Problem of J. Neyman and E. Scott," Ann. Math. Statist., 31 (1960) 656-661.

[5] Ghurye, S. G., and Olkin, I., "Unbiased Estimators of Densities," Abstract Ann. Math. Statist., 32 (1961) 1346.

[6] Lieberman, Gerald J., "Contributions to Sampling Inspection," Doctoral dissertation, Stanford University, 1953.

[7] Johnson, N. L. and Welch, B. L., "Applications of the Non-Central t-distribution," Biometrika, 31 (1940) 362-89.

[8] Resnikoff, George J. and Lieberman, Gerald J., "Tables of the Non-Central t-distribution," Technical Report No. 32, Applied Mathematics and Statistics Laboratory, Stanford University, 1957.

[9] Resnikoff, George J., "Interval Estimation of Proportion Defective in Sampling Inspection by Variables," Technical Report No. 21, Applied Mathematics and Statistics Laboratory, Stanford University, 1955.

[10] Wolfowitz, J., "Confidence Limits for the Fraction of a Normal Population which lies between two given limits," Ann. Math. Statist., 17 (1946) 483-488.

[11] Wallis, W. Allen, "Use of Variables, in acceptance inspection for percent defective," Chapter 1 in Selected Techniques of Statistical Analysis, ed. by Eisenhart, Hastay and Wallis, New York: McGraw-Hill, 1947.

[12] Bowker, A. H. and Goode, H., Sampling Inspection by Variables, New York: McGraw-Hill, 1952.

[13] Cramér, Harald, Mathematical Methods of Statistics, Princeton: Princeton University Press, 1946.

[14] Anderson, T. W., Introduction to Multivariate Analysis, New York: John Wiley and Sons, 1958.