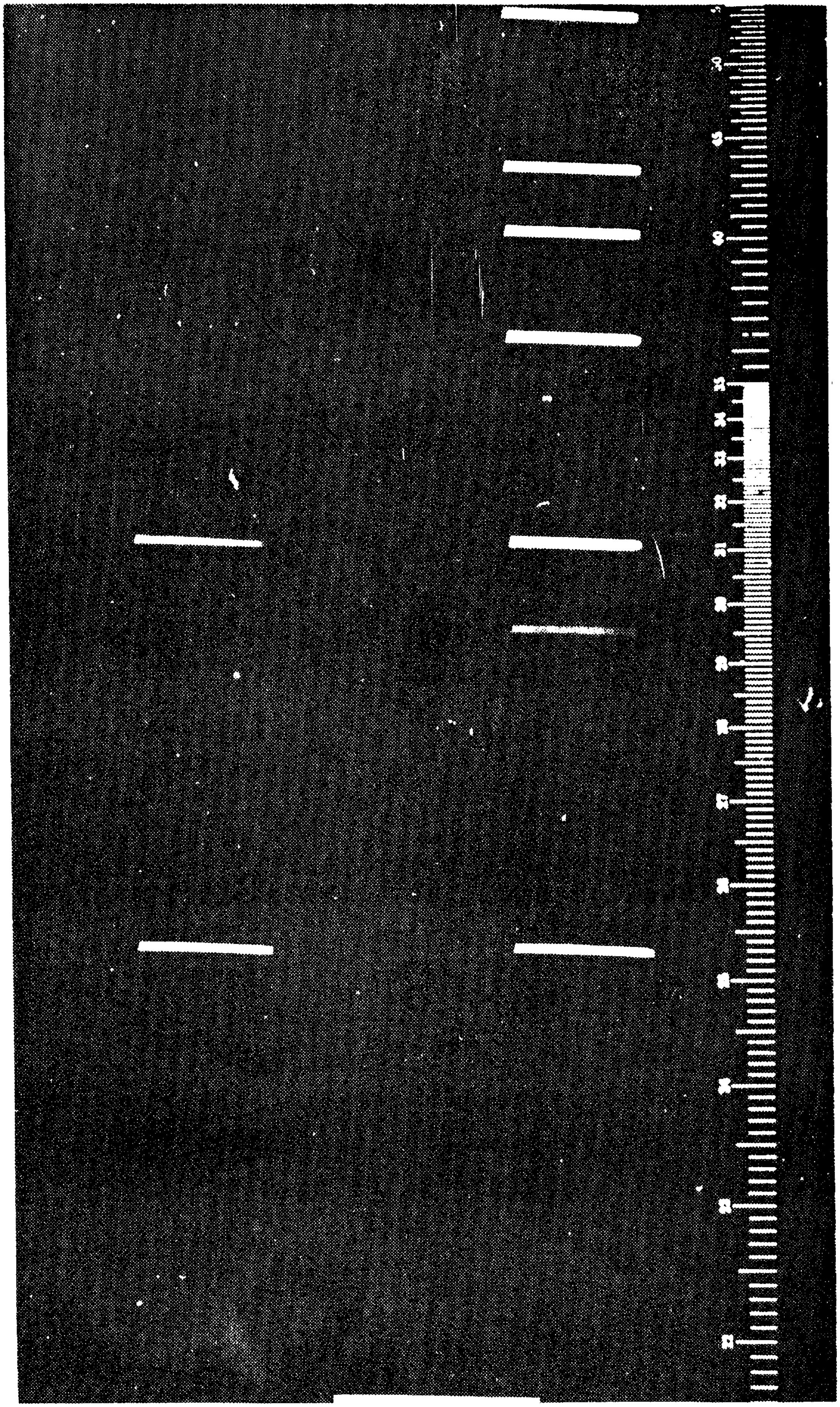


Plate No. 4

This page intentionally left blank.

Plate No. 5

This page intentionally left blank.

Plate No. 6

This page intentionally left blank.

Plate No. 7

This page intentionally left blank.

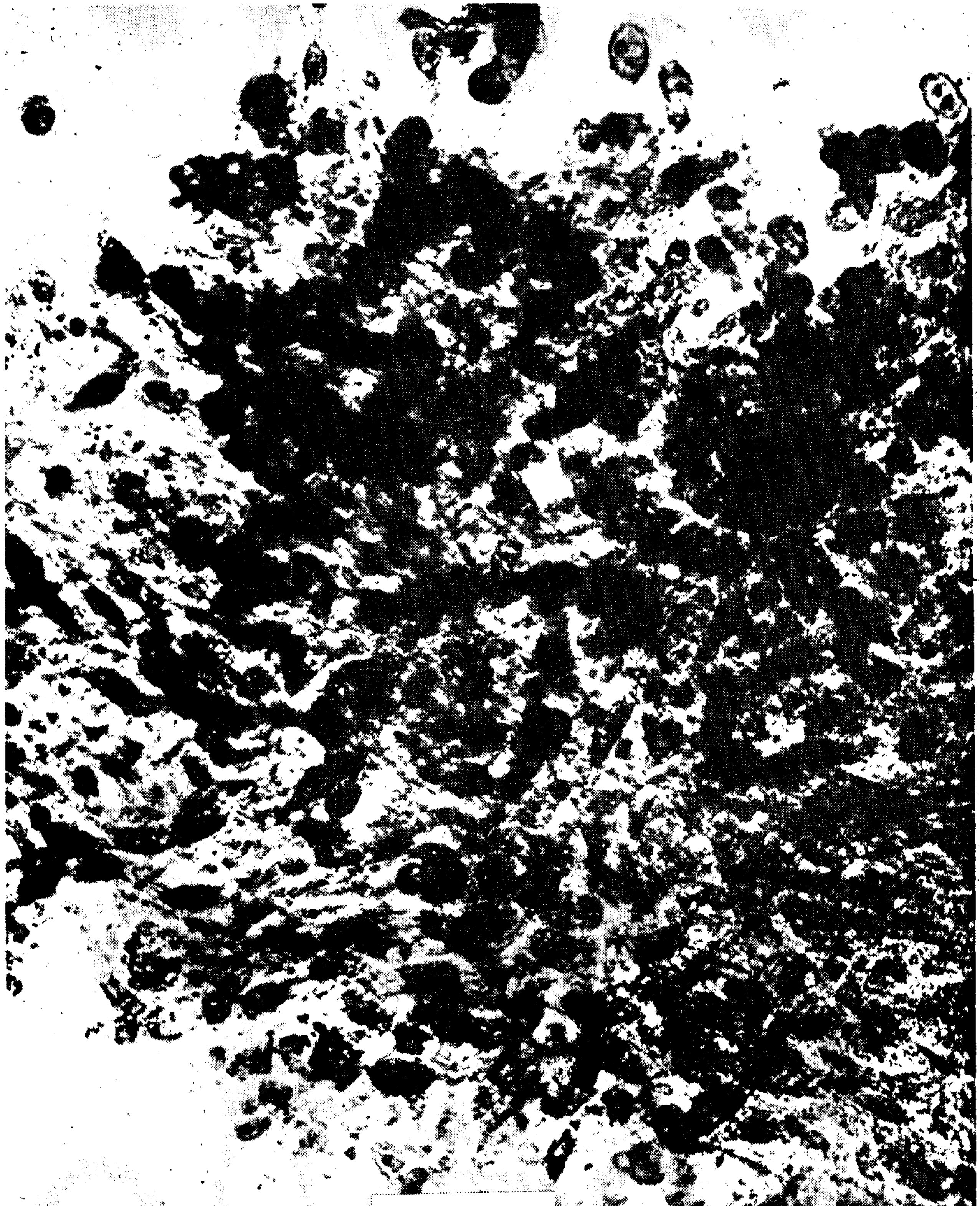


Plate No. 8

This page intentionally left blank.

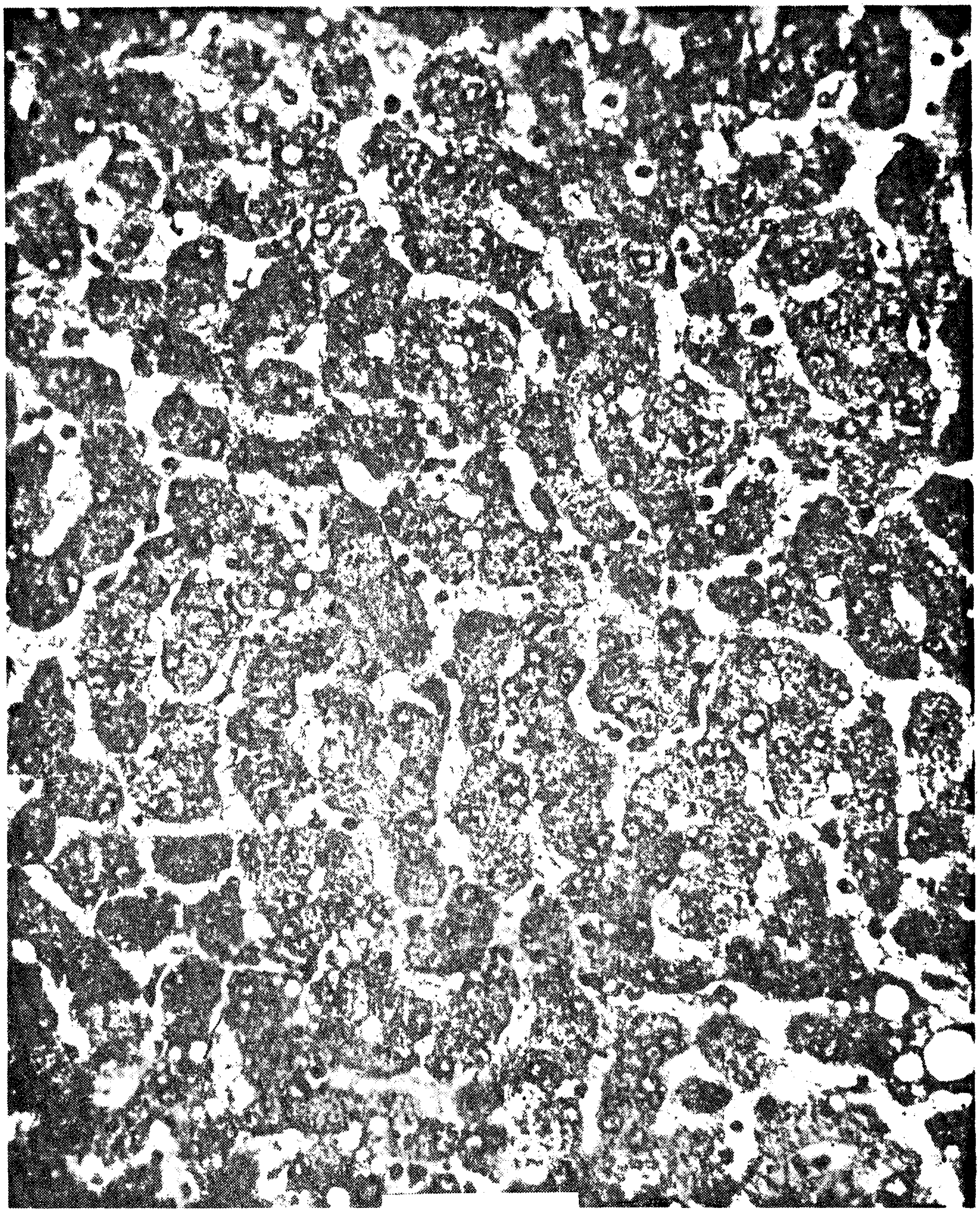


Plate No. 9

This page intentionally left blank.

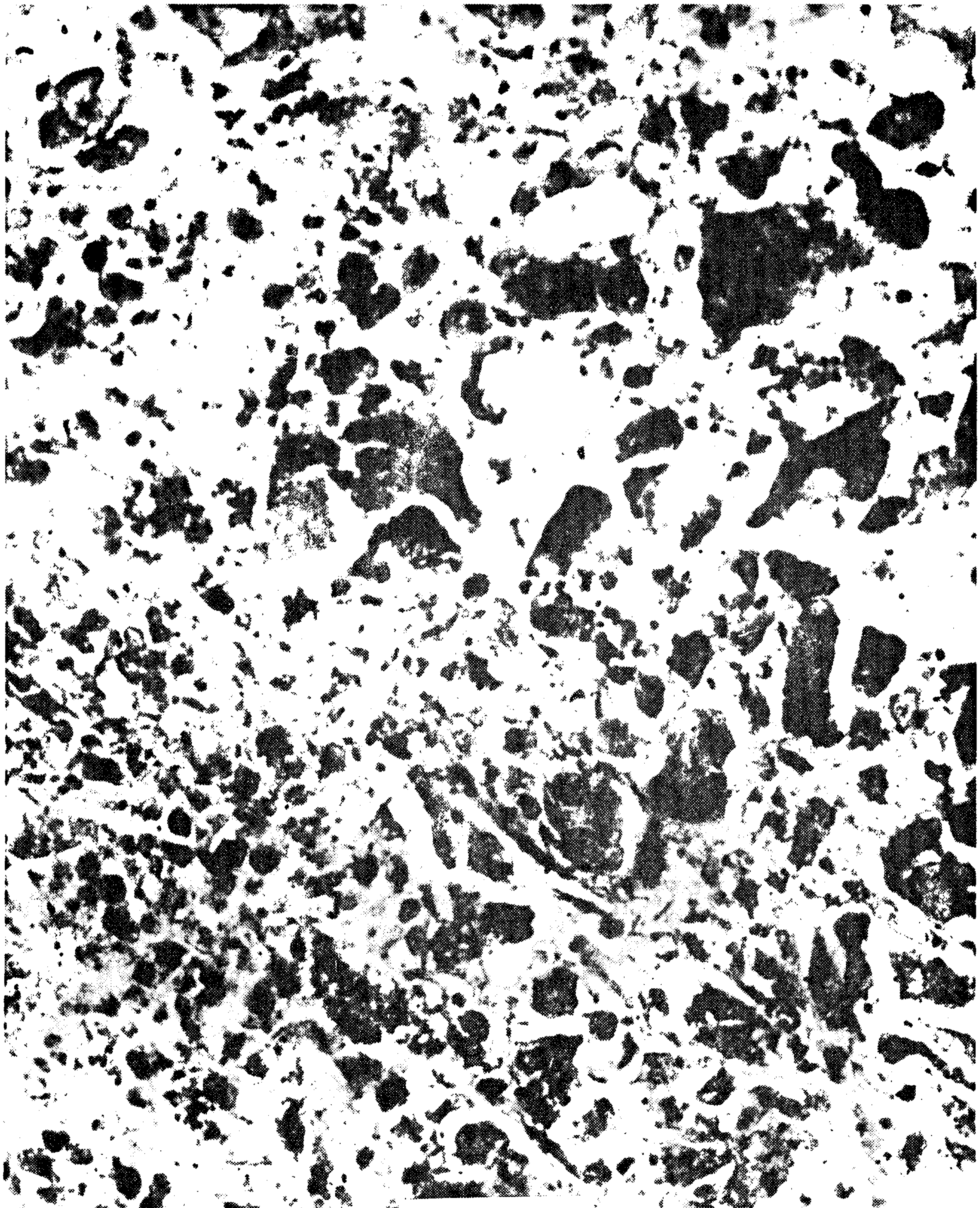


Plate No. 10

This page intentionally left blank.

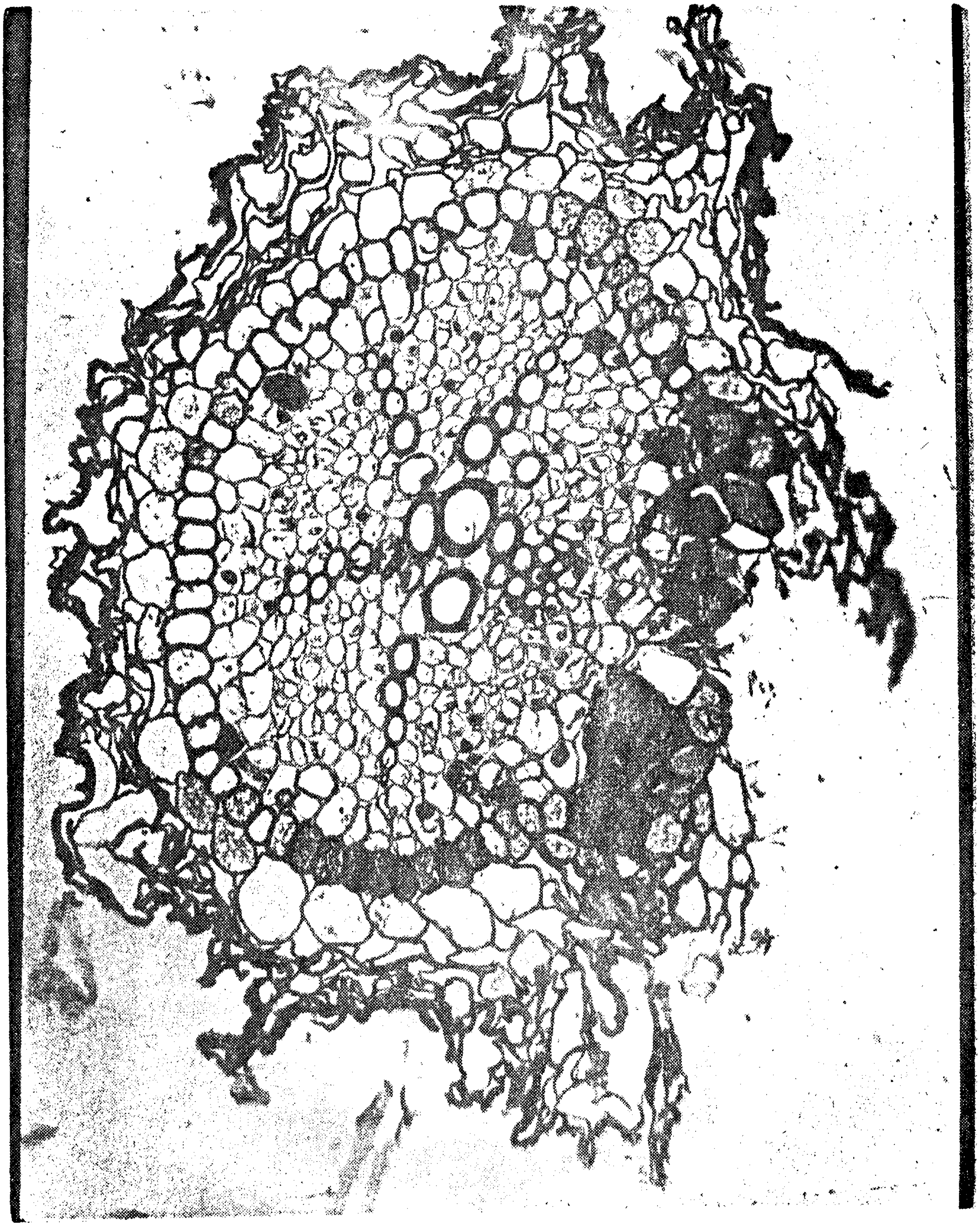


Plate No. 11

This page intentionally left blank.

Plate No. 12

This page intentionally left blank.

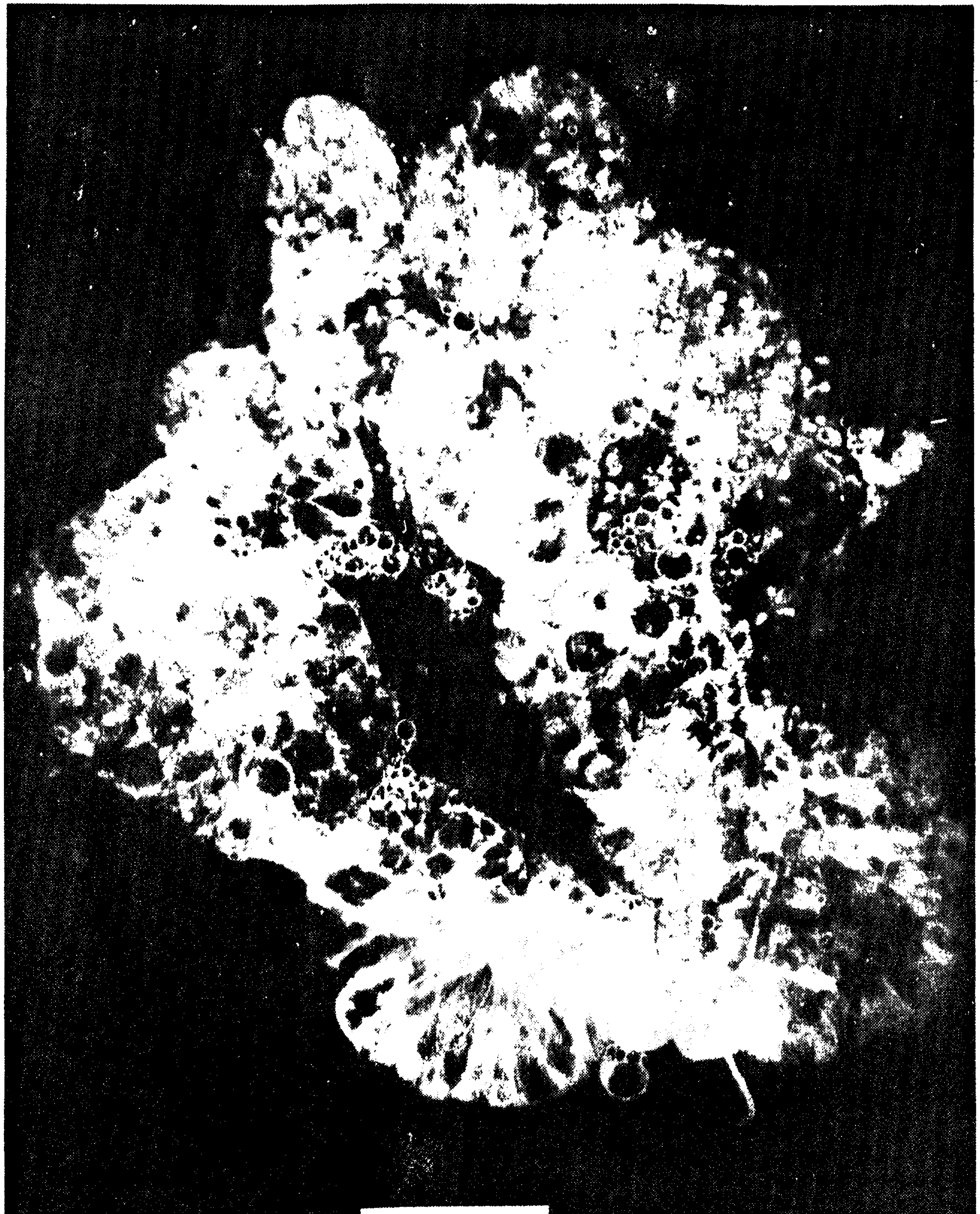


Plate No. 13

This page intentionally left blank.

Plate No. 14

This page intentionally left blank.

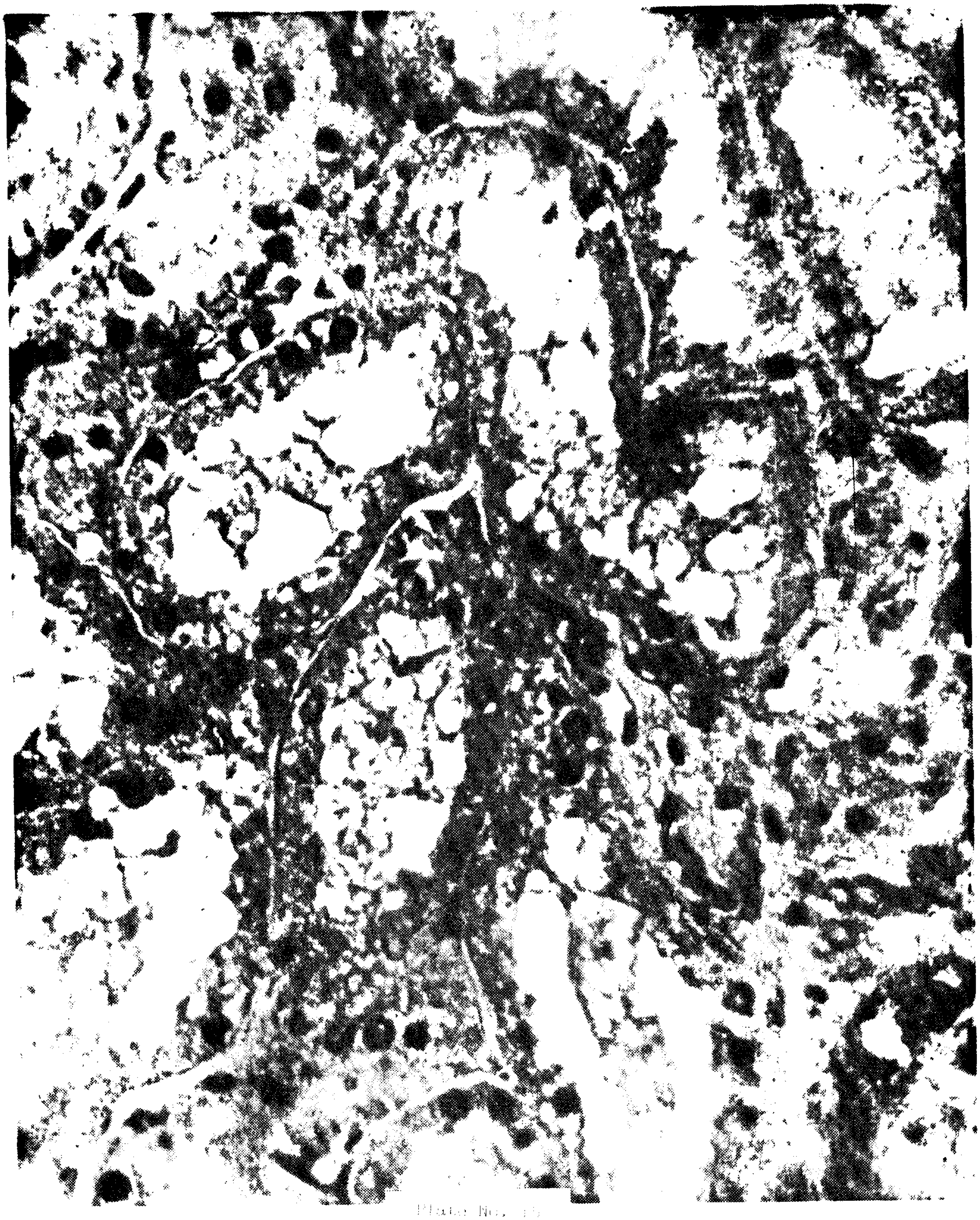


Plate No. [illegible]

This page intentionally left blank.

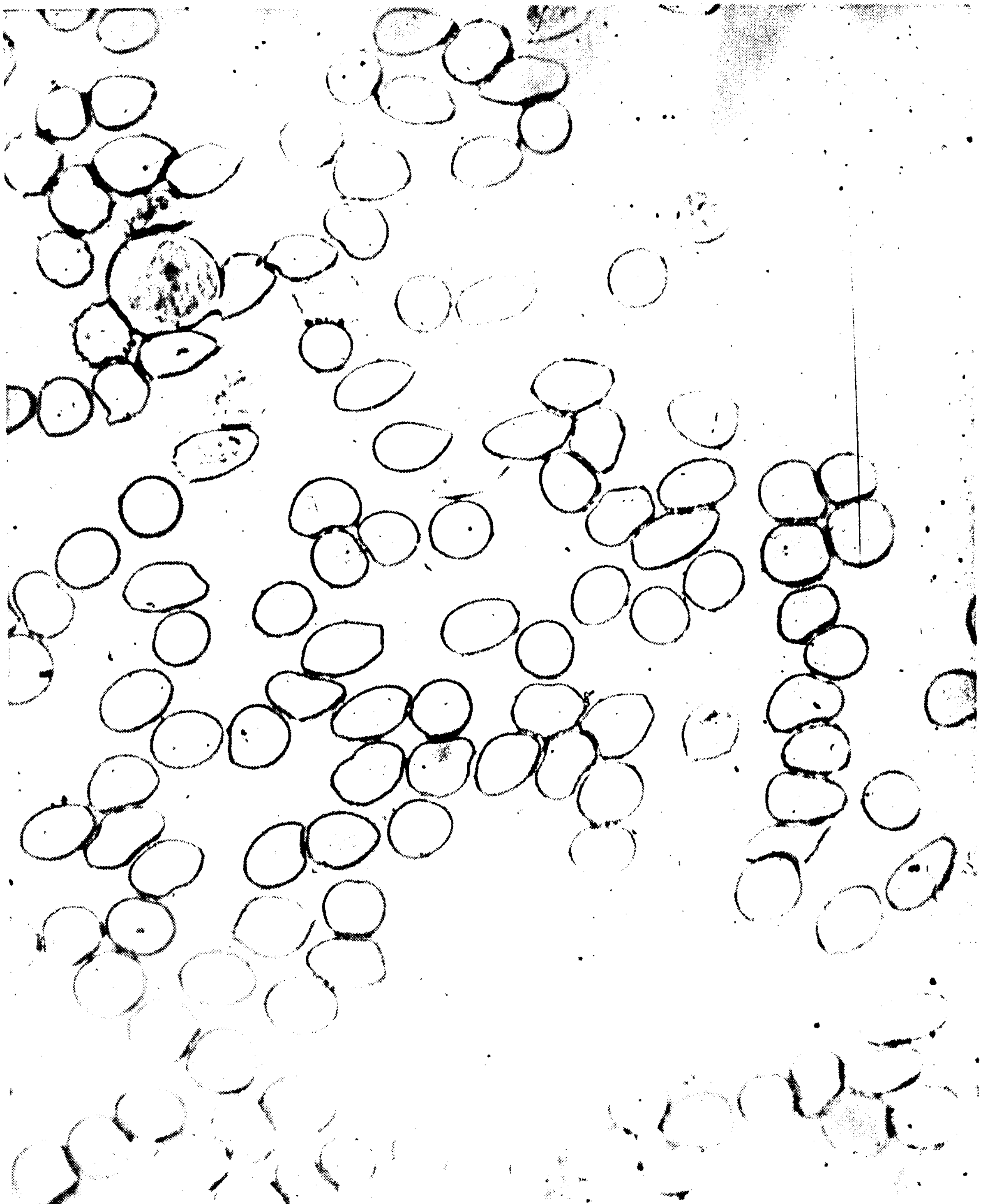


Plate No. 16

This page intentionally left blank.

Plate No. 17

This page intentionally left blank.

Plate No. 18

This page intentionally left blank.

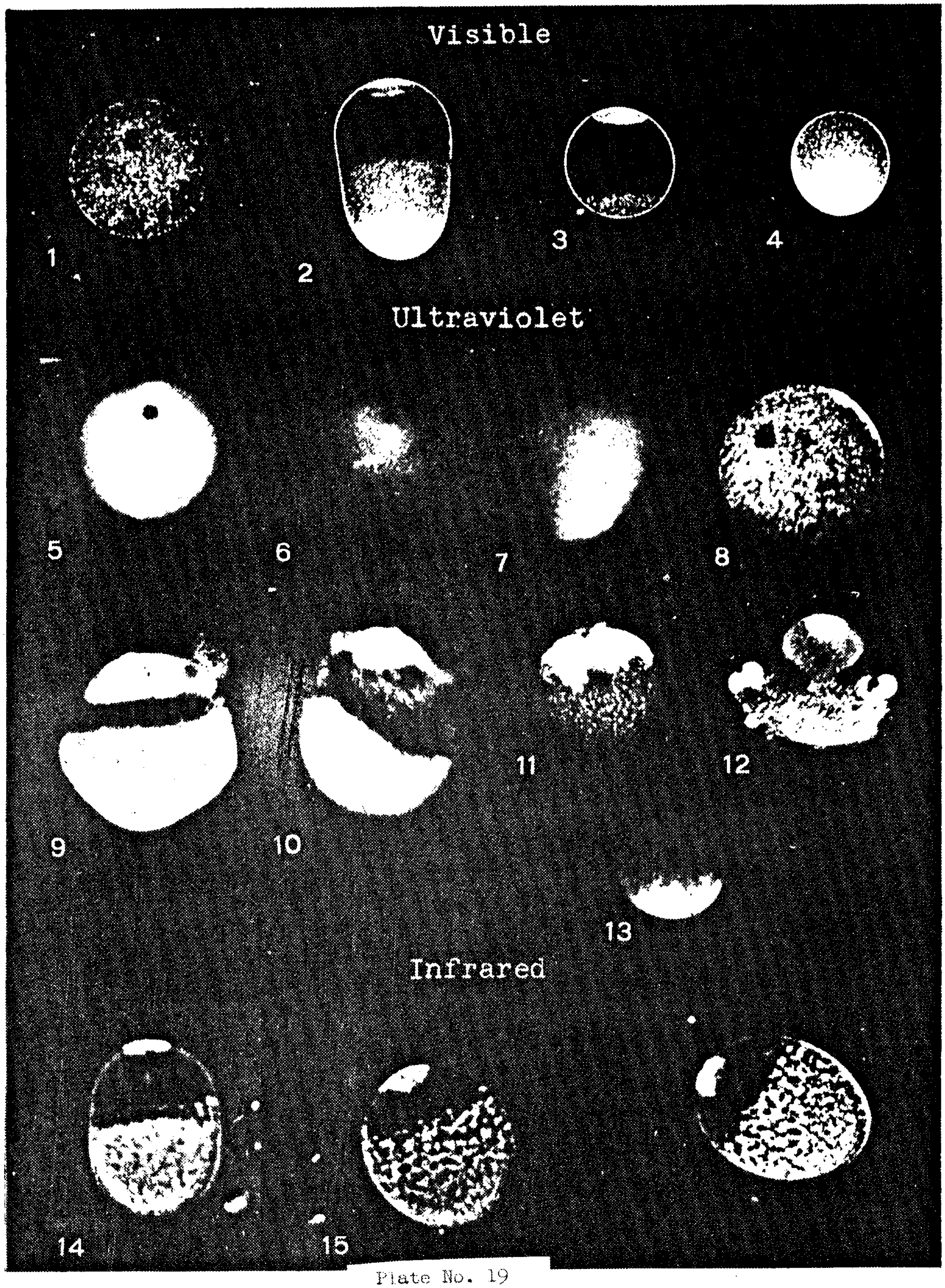


Plate No. 19

This page intentionally left blank.

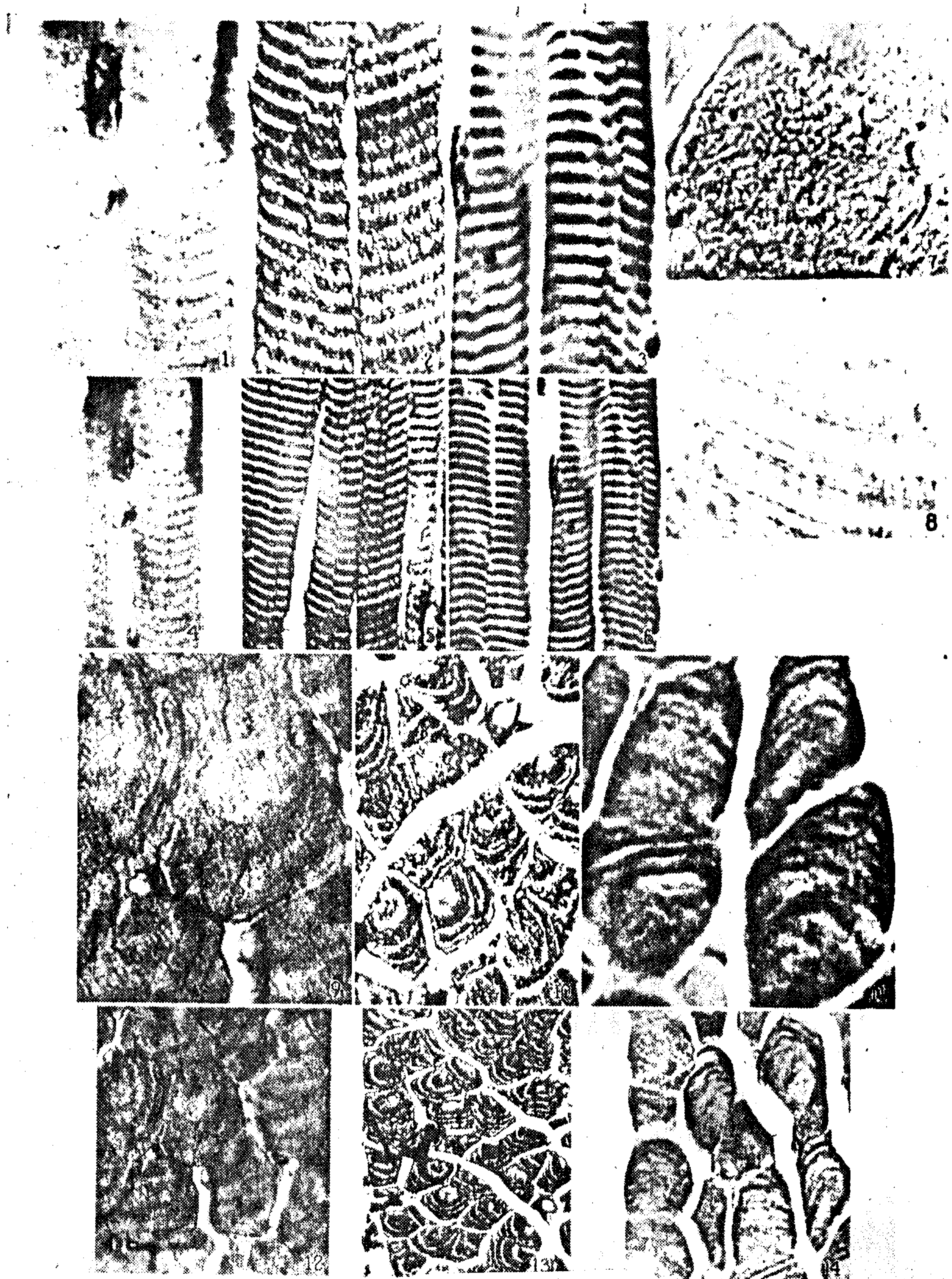


Plate No. 20

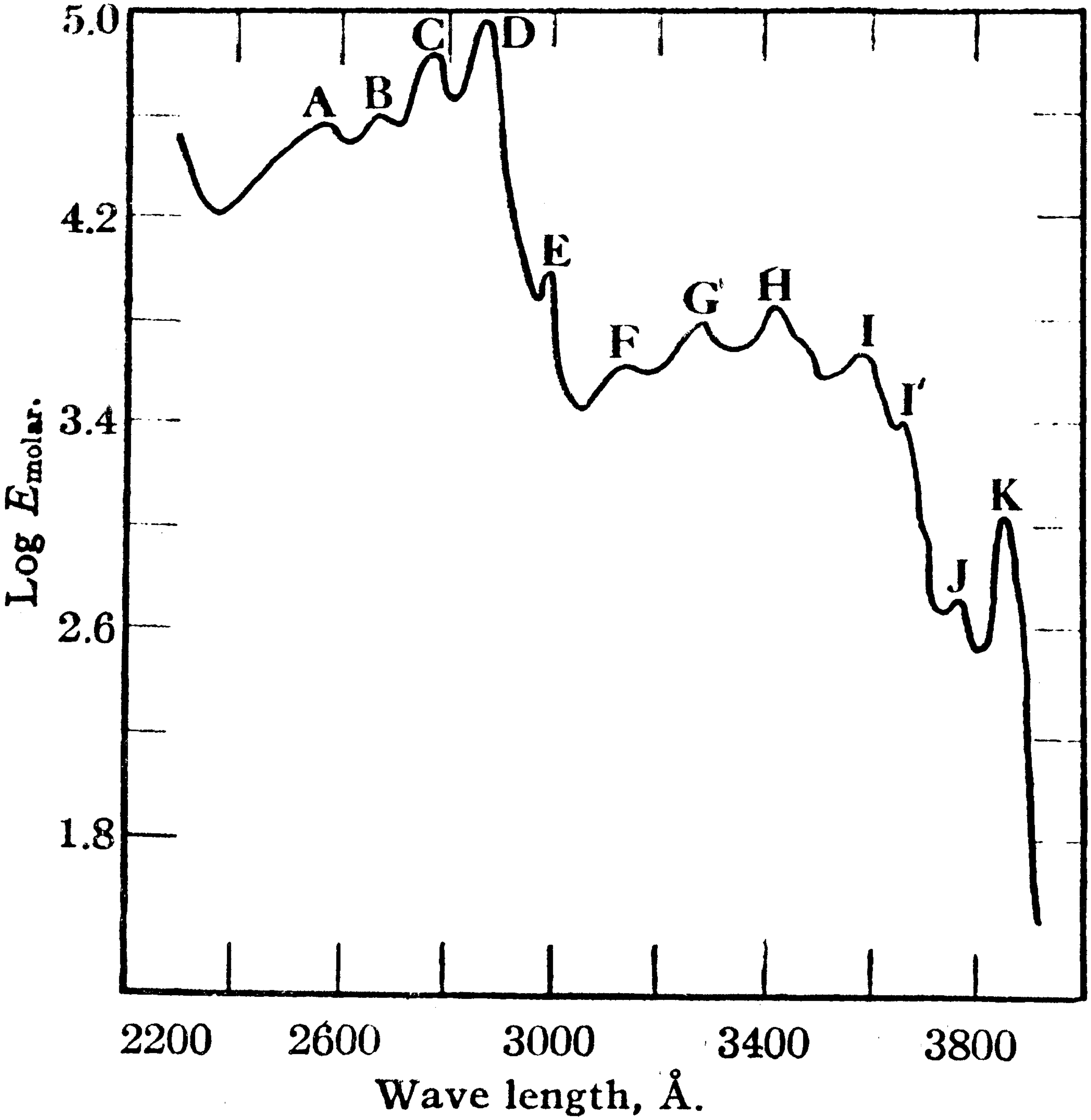


Fig. 1.—1,2-Benzanthracene.

Plate No. 21

This page intentionally left blank.

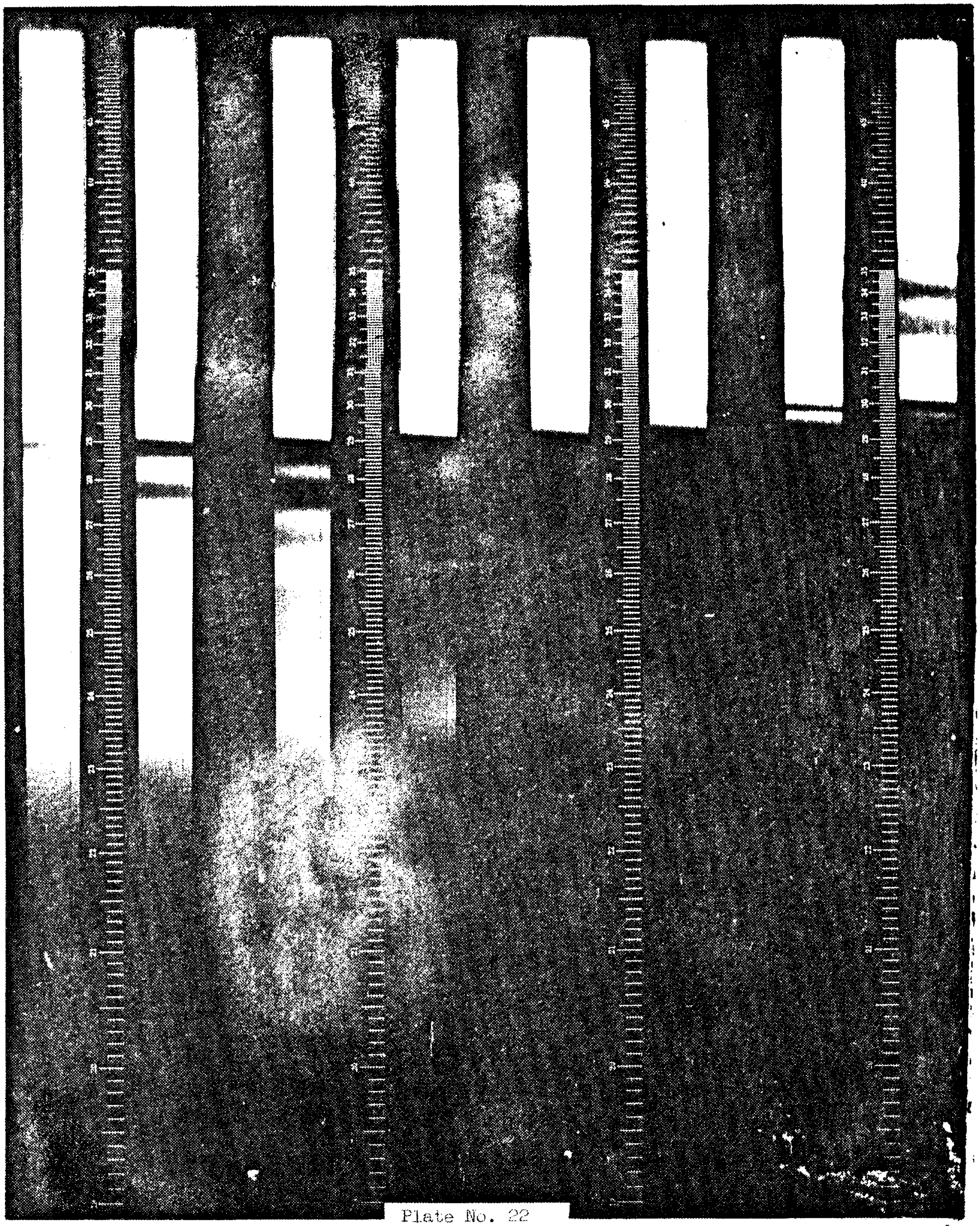


Plate No. 22

This page intentionally left blank.

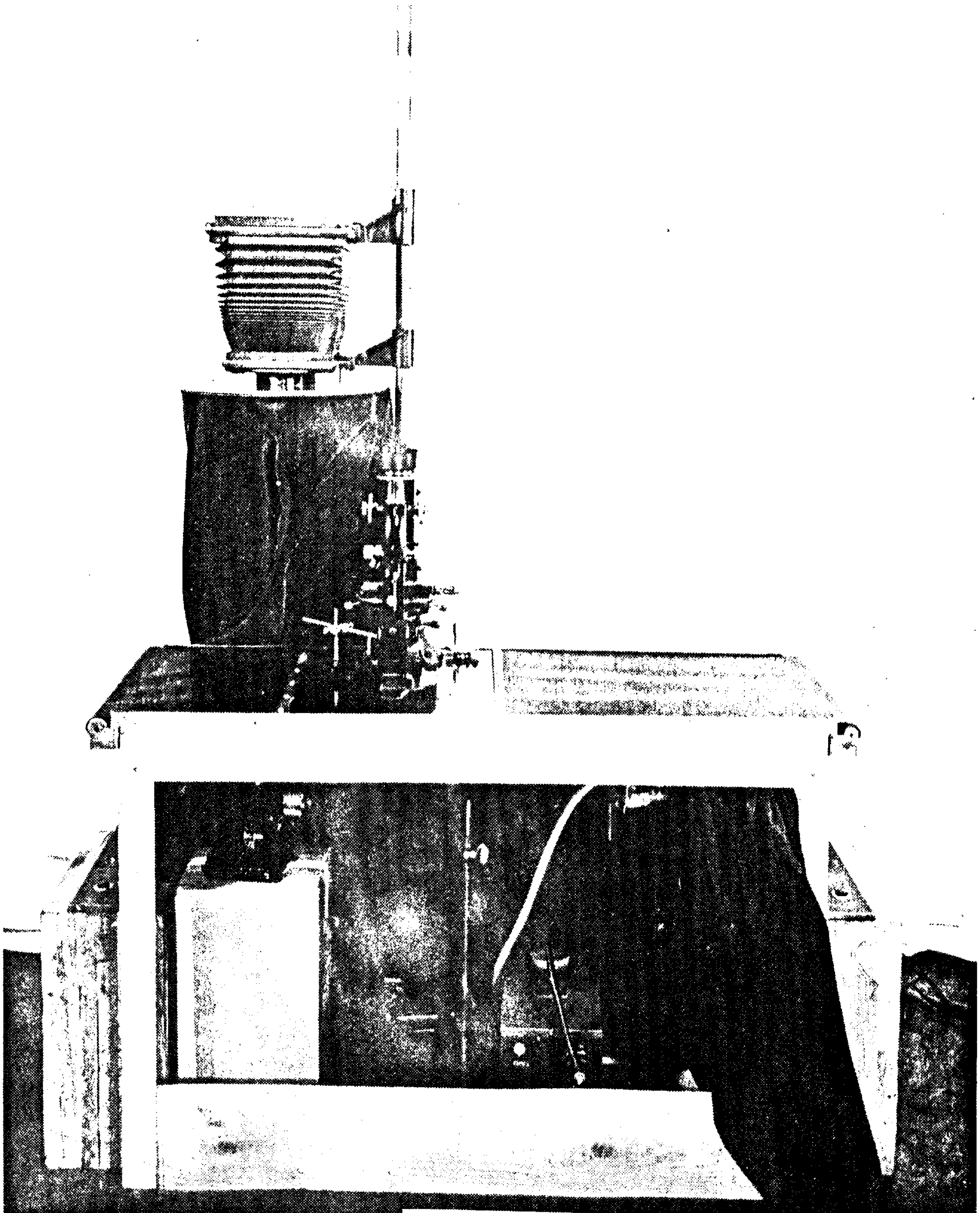


Plate No. 23

This page intentionally left blank.

REDUNDANCIES IN HUMAN BIOMECHANICS AND THEIR APPLICATION IN ASSESSING MILITARY MAN-TASK DISABILITY PERFORMANCE RESULTING FROM BALLISTIC AGENTS

William H. Kirby, Jr.
Ballistic Research Laboratories

ABSTRACT. Nervous system redundancies have a counterpart in human biomechanics which are evident when analyzing man-task performance using simulations of functional disabilities. These multiple choice pathways available for task accomplishment could explain some of the surprisingly small drops in performance in the presence of otherwise significant tissue trauma especially in cases of highly motivated soldiers. Such findings are thought to be useful in a bioengineering approach for optimizing man-machine or man-task relationships and deriving technical rules for the design of body armor and protective clothing for minimizing personal injury from ballistic agents.

INTRODUCTION. This is a discussion concerning a problem which associates engineering and medicine in a military context. As such it is a cross-application of exact and inexact sciences. With the advent of more intensive interest in mathematical applications to biological problems and the maturation of systems analysis, one now has more formidable tools with which to attack such problems. However, this discussion is restricted to an identification of the natural conditions allowing for multiple choice pathways in task accomplishment. No mathematical model is proposed.

Engineering and mathematical applications to systems behavior, particularly "living-systems", have brought about many new ideas regarding complex problem analysis. An effect of this has been and is, increasingly, that of providing us with more understanding of the nature of these systems. This encourages us to go back to our natural systems for re-examination but with more enlightenment on what we are really looking for.

This morning attention is directed to the response of complex living systems, specifically, humans, under a very large spectrum of traumatic or wounded states. Such interactions are of interest in a setting of military stress situations. Furthermore, it is important to correlate these responses with performance associated with defined tasks (or military occupational specialties) as they appear in the various tactical roles.

In our first approximations, we assume the presence of sufficiently high motivation such that if, despite the presence of trauma, a human biomechanical capability to perform in any fashion exists, it will be used. Consideration today will be restricted to qualitative structure-function relationships and useful methods of analysis. There is significant starting data relating to gross mechanisms for uninjured human function in terms of anatomical and physiological knowledge. One useful approach to this kind of problem is the use of simulation procedures and gaming. This could give us some idea of the value of a loss of function regardless of cause. Obviously wounds cannot be directly simulated, but many endpoint effects in terms of functional disabilities can be. Finally, another essential is the ability to study man-task interactions as they relate to accepted performance levels. In this respect industrial engineers, ergonomists, and applied experimental psychologists are evolving much useful information.

DISCUSSION. Ballistic agents directed at potential enemy soldiers are assessed in terms of the incapacitating effects which they bring about. A wound per se may have little meaning in a military stress situation unless it is matched against task requirements. Incapacitation is defined in this frame of reference as the "diminished capability to perform a defined task." Specifically, a wounded enemy soldier may have an otherwise serious wound in an upper limb, but should his task (military occupational specialty) require the use of only one limb to handle his weapon or machine for counter use during the stress engagement, and thereby effectively incapacitate his enemy (us), then the incapacitating effect of our ballistic agent is not great. It may not even be significant. Of course, the converse may be true. A small but deeply penetrating wound may be very incapacitating. An example of the latter could be one in which a soldier is required to have extremely fine motor coordination in both limbs simultaneously in order to perform the task adequately. In this case, quite a large variety of small penetrating wounds at various anatomical locations could be cited in which motor (muscle) control would be lost resulting in either an inability to control fine movement or an inability to manipulate the entire limb into position for accomplishing the fine movement. The point, of course, in this definition of incapacitation is the fact that an evaluation concerns, not man alone, but man coupled to his assigned tasks in prescribed environments.

As physicians we think in terms of pathological mechanisms pertaining to injury. Such understanding is predicated, in turn, on knowledge from the basic medical sciences. Functional disabilities ascribable to any spectrum of injuries are assumed to be variable and time dependent. Functional deficits are related in this human system analysis to the final common motor pathway which is a clinical way of referring to one's re-active muscle mechanics. This applied concept is useful inasmuch as the interlock existing between man and his external physical mission or mechanical work is to be accomplished is one of dynamic character. At times this interlocking dynamics is characterized by numerous physical actions while at other times, such actions are numerically minimal but the physiological requirements are intense. In any event, it is obvious that in studying human actions we must deal with a complex integration of both physical and physiological factors.

As engineers we express work in terms of force times distance. Physiologically we have often considered work on the part of a living system as an expenditure of metabolic energy. In this sense we are, of course, in error in our definition of work. We need to differentiate between metabolic energy utilized for life support or maintenance of homeostatis and that metabolic energy available and employed to accomplish "useful" external work. In engineering we sometimes use the expression "useful work" when discussing thermal efficiency of a machine. For example, the thermal efficiency of a locomotive may be 8 per cent, meaning that 8 per cent of the available energy is converted to so many foot-pounds of output and the remainder to heat. However, in handling the combination of inanimate and animate (man and weapon), some practical criteria for NORMAL man-task behavior are needed against which one can assess performance for incapacitated states.

In our first approximations, it appears practical to use a qualitative form of biomechanics providing we can correlate this satisfactorily with appropriate performance levels as derived from experimental analysis. Anthropometrical, physiological, and psychological variations in the human are, of course, an integral part of any appraisal of human function and behavior in the final analysis. Hopefully one looks forward to mathematical expressions for causal relations that are valid and reliable.

Man is endowed with considerable versatility. The use of this versatility is influenced by the mind (brain), body, and environment. The need to relate wound tract pathology with the internal biomechanics of body actions is apparent. However, we also find it useful to study overall functional behavior while in action in order to get impressions of the sequential dynamics that differentiate one kind of task from another. These applied human mechanics can be studied in terms of their mechanistic factors ... i. e., bone-muscle function. In the long run our model must include input factors such as sensing and/or kinesthetic items as well as kinematics, time-varying factors such as endurance and fatigue, and system alterations due to environment in this total biological-task-environmental stress system problem.

The biophysics and pathological dynamics of wounding are matters of great importance but beyond our scope this morning. We will attempt here to explore the problem in reverse. We may begin with certain functional (motor) deficits which could conceivably be derived from one or more wounds. In the future we expect to assess other factors in the human circuitry such as loss of sensory function, kinesthetic changes, and trauma to the body subsystems such as the nervous system, cardio-vascular system, and respiratory system. We hope to be able to draw from others doing human research in the life sciences and biotechnology. As mentioned, we will try to relate traumatic effects to that final common motor pathway. By simulating motor deficits we are, in effect, working backwards. The more general motor deficits are not difficult to simulate and as such enable us to make at least first approximations of man-task (soldier-weapon or soldier-task) performance behavior as this occurs in simulated tactical role exercises. Incidentially, it is felt that this approach will allow us to do as Professor Pearson outlined yesterday ... ... weld together the important elements including mathematical and war and computer games on the one hand and laboratory experiments, range trials, and Army exercises on the other.

Our view of a human-task model includes the following factors:

1. The Human or Living Machine
2. Tasks
3. Interactions of the Living Machine and Task Accomplishment
4. Environment (not discussed in this paper)

The Human or Living Machine.

In our present work, the human body is being considered as a living machine endowed with a natural structural system capable of serving its internal biological needs while performing external tasks. Many sources are available for descriptions of human mechanics. The Hall of Biology of Man of the American Museum of Natural History is one of these. Anatomists, particularly J. C. Grant (A Method of Anatomy) and G. B. Duchenne (The Physiology of Motion, more recently translated by E. B. Kaplan of Columbia University), offer very detailed work along these lines. Your attention may have been attracted recently to the present series of articles in LIFE Magazine which discusses some of the bio-mechanics of the human body.

While many people have not thought about human function in terms of mechanics, they are not particularly surprized when human activity is conservatively compared with mechanical devices. It is also not difficult to see that some overlap or redundancy must exist in a system in which the parts are so intricately intertwined as in the human.

Studies of normal (non-injured) human subjects performing given tasks show that different people use different methods in performing them. In addition, these methods are modified as subjects become more accustomed to the tasks. We say that workers become conditioned or skilled in their tasks. Handicapped workers are often forced to use even radically different methods. In the few simulated incapacitation studies that we have made, we have found a considerable variety of methods employed by our limited number of subjects. In a real sense, then, we can consider this multiplicity of methods in terms of bio-mechanical redundancies. As a matter of fact, one can often explain the adaptation or describe the compensatory pathway chosen on the part of a disabled individual when a more natural one is not available for performing a given task.

Tasks.

Turning briefly to the applied or task side of the picture, i. e., the nature of the things that impose motor output requirements on the human machine, we find an almost endless array of situations. Knobs have to be grasped, turned, and released; levers (such as rifle trigger, bolt, etc.) have to be pulled and pushed; things have to be grasped, raised, lowered and released; buttons have to be pushed, tools have to be handled, vehicles guided, etc. All such physical tasks may be arbitrarily viewed as being accomplished by composites of elementary actions performed by the human (machine).

Interactions of the Living Machine and Physical Tasks.

Viewing the human machine in action shows a link-linkage system including its affixed but integrated motors which raise and lower limbs (for walking, lifting objects, etc.), raise, lower, turn, and tilt the head, etc. Identification of these muscle-motor systems, their sequence(s), and associated force factors can be studied extensively. Muscle and joint functions are quite well known. Their motion and force vectors can be studied. The kinematic processes activated by the human motors can be at least approximated by various methods and are very informative from both qualitative and quantitative points of view. Data are being generated in many of the associated sciences involved in the type of cross-application presented here, and the ways and means are evolving rapidly for handling such massive bits and pieces of acquired information. The need for the systems analysis has already been mentioned. Information handling, cybernetic modeling using acceptable analogs, computer technology, statistical and adaptive control techniques are included in the knowledge sources useful in understanding complex system behavior.

Today I would like to call attention to a very crude experiment which indicates the multiplicity of motor pathways available to the human in performing given tasks. Variations in weapon design, man-task functional modes, objective definitions, parameters and boundary conditions have not been specified from a systems viewpoint. We have used very simple immobilization techniques for inhibiting certain motor functions in order to enforce and observe the use of alternative motor pathways. This experimental exercise has to do with rifle firing and reloading. The target is about 2 feet by 2 feet at a distance of approximately 40 yards on a flat terrain. One of the first questions of interest to us was, "Can one perform under functional disability conditions, rather than how well"? However, as you will see from the film, we get cues as to how well one does and can perform.

In order to be a little more inclusive, the rifle experiment was run using two different weapons, namely, the U. S. M-14 and the Russian AK both of which are standard items. We will not speculate on weapon differences as such, but we will mention a few things about a man-rifle relationship with and without functional losses. You will observe in the film that a soldier can fire his weapon quite effectively in all firing positions without either upper limb. Pistol grips and monopods seem to be quite helpful to a soldier so disabled. Now let us view the film.

A MAN-TASK DISABILITY EXERCISE: RIFLE FIRING. Time will allow for only a few general remarks. I would like to say that the more one reviews and studies motion pictures of man-task behavior the more one can comprehend the biomechanics taking place in such exercises. The anatomist may be the first to observe the variations in anatomical mechanics; the statistician may quickly detect probabilities in terms of cause and effect relationships; the abstract mathematician may see early cues for a stochastic model; the mechanical engineer may be the first to note the vectorial mechanics in 3-dimensional space; the physiologist may immediately observe the abrupt discontinuities in the human functional activities and feel more sensitive about the corresponding metabolic requirements involved; etc.

At this time, we have no valid statistical data. In our very crude investigation, we have observed with caution, of course, the following:

1. Partial losses up to and including either total limb did not prevent the subjects from firing and maneuvering.

2. Target scores did not decrease much below the 65% - 75% accuracy range even for the most severe simulated disability that we employed . . . the inability to use the trigger arm.

3. Firing rates and reload times were only a few seconds longer in the absence of a total upper limb.

4. The motions exhibited by the subjects are somewhat influenced by size, shape, and weight characteristics of the weapons. However, required body positions. . . . .standing, prone, etc. . . . naturally influence the biomechanical adjustments which must take place.

MOTION STUDY OF HUMAN ACTION. Since human work is accomplished by means of body actions, the study of body movements has evolved as one of the principal approaches to the problem of finding more effective ways of performing tasks. From such empirical studies, rules have been developed which are available for more effective application in planning and designing tasks, machines, and weapons. Skillful application of these principles diminishes fatigue. It is interesting but, perhaps, not surprising that the methodology proposed for the study of incapacitation i. e., functional deficit simulation, is useful in studying normal man-task phenomena.

In recent years, physicians, applied experimental psychologists, physiologists, and engineers have been studying performance factors and have developed considerable empirical data especially in regard to environmental perception and functional response. Others giving much thought to human anatomical function include (in addition to anatomists) orthopedists, physical therapists, and designers of prostheses. Functional anthropologists, physicists and advanced systems engineers are also becoming essential participants in this area of activity. However, we have not fully recognized the capabilities of these professions in cross-discipline or cross-professional applications. A sizeable effort has been directed in more recent years, to a so-called ergonomic approach to man-task analysis especially in Europe. This approach places more emphasis on the integration of anatomy, psychology and physiology as well as economics for solving problems in human performance.

The Gilbreths, pioneers in the development of human motion principles, devised a list of 17 so-called "elements" or "therbligs" as they have been called. Such elements have been considered as basic units of motion and apply for the most part to the upper limb functions. They include such terms as "transport empty" meaning moving the hand from one position at a work place to another in that vicinity; "transport loaded" meaning the same thing except that the hand is now carrying an object in which case the characteristics of the object are specified; "grasp" meaning a securing of an item either by a pinch-grasp performed by the fingers or a palmar-grasp as performed by enclosing the hand about the item; "hold" meaning that the hand in question is maintaining an object in a fixed position while, perhaps, the opposite hand is doing something to the object as may occur in an assembly operation; etc. In each case, the elements are timed. Therbligs are usually measured to one-thousandth of a minute. Stop-watch time and motion study engineers usually measure their elements in terms of one-hundredth of a minute. Instead of using therbligs, they use descriptive terms which are more general, such as "pick up hammar", "tap dowel to flush position", "place hammar aside", "place assembly in tote box", and "measure outside diameter with micrometer". As long as the the motions are defined and measured consistently, they are useful in the analysis and the synthesis of operations.

Very quickly I would like to show you the motion study and analysis technique for a given operation which is described on the following slides (figures). Our slides show several motion study charts based on a hypothetical exercise. In slide 1 (Figure 1) we show a "Right-hand: Left-hand Chart" and for simplicity only the actions of the hands are described.

MOTION ANALYSIS - STEP I

OPERATION: BRING RIFLE TO BEAR ON TARGET FROM THE *PORT ARMS* POSITION TO THE *SHOULDER AIM* POSITION, FIRE SIX ROUNDS, AND RETURN RIFLE TO THE *PORT ARMS* POSITION.

RIGHT HAND	LEFT HAND
1. *HAND* ASSISTS IN BRINGING THE RIFLE (REAR STOCK) TO RIGHT SHOULDER WHILE *INDEX FINGER* IS LOOSELY POSITIONED ON THE TRIGGER. THE *HAND* MAINTAINS A MODERATELY FIRM GRIP ON THE STOCK.	1. *HAND* ASSISTS IN BRINGING RIFLE (FRONT STOCK) TO THE SHOULDER LEVEL, WITH *PALM* AND *FINGERS* MAINTAINING A FIRM SUPPORT AND GRASP.
2. THE REAR STOCK IS POSITIONED AGAINST THE FRONT OF THE SHOULDER; THE *HAND* INCREASES ITS GRIP ON THE STOCK EXCEPT FOR THE *INDEX FINGER* WHICH MAINTAINS AN ESSENTIALLY NEUTRAL POSITION.	2. THE FRONT STOCK IS POSITIONED IN CONJUNCTION WITH THE REAR STOCK AND THE *HAND* TAKES ON A MODERATELY SEVERE SUPPORT FUNCTION.
3. WHILE THE *HAND* (AND *FINGERS*, EXCEPT THE *INDEX FINGER*) MAINTAINS A FIRM GRIP ON THE REAR STOCK, THE *INDEX FINGER* PRESSES AGAINST THE TRIGGER, INCREASING THE PRESSURE UNTIL IT TRIPS THE FIRING PIN.	3. *HAND* SUPPORTS THE FRONT STOCK.
4. IMMEDIATELY AFTER ABSORTION OF THE RECOIL, THE *INDEX FINGER* IS SHIFTED (MEETING NO RESISTANCE) TO THE ORIGINAL POSITION OF THE NEXT ROUND.	4. *HAND* SUPPORTS THE FRONT STOCK.
5. SAME AS 3.	5. *HAND* SUPPORTS THE FRONT STOCK.
6. SAME AS 4.	6. *HAND* SUPPORTS THE FRONT STOCK.
7. SAME AS 3.	7. *HAND* SUPPORTS THE FRONT STOCK.
8. SAME AS 4.	8. *HAND* SUPPORTS THE FRONT STOCK.
9. SAME AS 3.	9. *HAND* SUPPORTS THE FRONT STOCK.
10. SAME AS 4.	10. *HAND* SUPPORTS THE FRONT STOCK.
11. SAME AS 3.	11. *HAND* SUPPORTS THE FRONT STOCK.
12. SAME AS 4.	12. *HAND* SUPPORTS THE FRONT STOCK.
13. SAME AS 3.	13. *HAND* SUPPORTS THE FRONT STOCK.
14. SAME AS 4.	14. *HAND* SUPPORTS THE FRONT STOCK.
15. OPPOSITE OF 1.	15. OPPOSITE OF 1.

FIGURE 1

This page intentionally left blank.

You observe that the wording in this first slide is in lay language and is lengthy, but this was done purposely for our presentation. The motions of each hand are described singly, and then listed synchronously, one beside the other. You will see more clearly in the next slide the relationship of the activities of each hand. However, this first slide does show how motions are described in terms of elements and a method for analyzing them.

In Slide 2 (Figure 2) you see a condensation in the language as compared to Slide 1. This terminology is often employed by motion analysts. Note how quickly you can get a visual time-history of the events on the part of each hand. In order to emphasize the more important factors in such an operation, one usually includes a clear descriptive summary showing numerical relationships between active and inactive motions.

This same operation cycle has been made a little more sophisticated in Slide 3 (Figure 3) by adding a graphical representation for the various kinds of elements. One purpose of this format is to emphasize the idle or inactive elements. I should point out that ordinarily these graphical representations are in terms of time per element giving a quantitative as well as a qualitative measure. Our example here is hypothetical and admittedly would be more effective if a full motion-time study were made. However, the purpose is to point out a methodology useful in studies relating body mechanics to machine and/or task characteristics.

By applying this technique to human tasks and translating such observations into anatomical mechanics, one may be able to specify in more detailed fashion the various biomechanical pathways employed by humans for given functional disabilities. We all know of unusual accomplishments on the part of some handicapped workers. The same is true for some accident victims immediately following trauma. Certainly the time parameter and undoubtedly a host of others are important, such as environmental conditions, task nature, human motivation, etc., and should be specified in modeling man-task behavior under conditions of disability. In this way, redundant bio mechanical networks available to the human may be weighted and probabilitistic methods applied. It is, of course assumed to be necessary to have a definition for normal performance for any given task as a baseline.

This page intentionally left blank.

MOTION ANALYSIS - STEP II

OPERATION: BRING RIFLE TO BEAR ON TARGET FROM THE *PORT ARMS* POSITION TO THE *SHOULDER AIM* POSITION, FIRE SIX ROUNDS, AND RETURN RIFLE TO THE *PORT ARMS* POSITION.

RIGHT HAND	LEFT HAND
1. TRANSPORT LOADED	1. TRANSPORT LOADED
2. POSITION	2. POSITION
3. PULL LOADED (INDEX FINGER) & HOLD	3. SUPPORT & HOLD
4. PUSH EMPTY (INDEX FINGER) & HOLD	4. SUPPORT & HOLD
5. PULL LOADED (INDEX FINGER) & HOLD	5. SUPPORT & HOLD
6. PUSH EMPTY (INDEX FINGER) & HOLD	6. SUPPORT & HOLD
7. PULL LOADED (INDEX FINGER) & HOLD	7. SUPPORT & HOLD
8. PUSH EMPTY (INDEX FINGER) & HOLD	8. SUPPORT & HOLD
9. PULL LOADED (INDEX FINGER) & HOLD	9. SUPPORT & HOLD
10. PUSH EMPTY (INDEX FINGER) & HOLD	10. SUPPORT & HOLD
11. PULL LOADED (INDEX FINGER) & HOLD	11. SUPPORT & HOLD
12. PUSH EMPTY (INDEX FINGER) & HOLD	12. SUPPORT & HOLD
13. PULL LOADED (INDEX FINGER) & HOLD	13. SUPPORT & HOLD
14. PUSH EMPTY (INDEX FINGER) & HOLD	14. SUPPORT & HOLD
15. TRANSPORT LOADED	15. TRANSPORT LOADED

SUMMARY

	TOTAL NO. OF ELEMENTS	NO. OF ACTION ELEMENTS	NO. OF PASSIVE SUPPORT ELEMENTS
RIGHT HAND	15	15 (100%)	0 (0%)
LEFT HAND	15	3 (20%)	12 (80%)

FIGURE 2

This page intentionally left blank.

MAN-TASK CHART

OPERATION: BRING RIFLE TO BEAR ON TARGET FROM THE *PORT ARMS* POSITION TO THE *SHOULDER AIM* POSITION, FIRE SIX ROUNDS, AND RETURN RIFLE TO THE *PORT ARMS* POSITION.

RIGHT HAND	LEFT HAND	RIFLE
TRANSPORT LOADED	TRANSPORT LOADED	IDLE
POSITION	POSITION	IDLE
PULL LOADED & HOLD	SUPPORT & HOLD	INSTANT OF FIRE
PUSH EMPTY & HOLD	SUPPORT & HOLD	IDLE
PULL LOADED & HOLD	SUPPORT & HOLD	INSTANT OF FIRE
PUSH EMPTY & HOLD	SUPPORT & HOLD	IDLE
PULL LOADED & HOLD	SUPPORT & HOLD	INSTANT OF FIRE
PUSH EMPTY & HOLD	SUPPORT & HOLD	IDLE
PULL LOADED & HOLD	SUPPORT & HOLD	INSTANT OF FIRE
PUSH EMPTY & HOLD	SUPPORT & HOLD	IDLE
PULL LOADED & HOLD	SUPPORT & HOLD	INSTANT OF FIRE
PUSH EMPTY & HOLD	SUPPORT & HOLD	IDLE
PULL LOADED & HOLD	SUPPORT & HOLD	INSTANT OF FIRE
PUSH EMPTY & HOLD	SUPPORT & HOLD	IDLE
TRANSPORT LOADED	TRANSPORT LOADED	IDLE

SUMMARY

	TOTAL NO. OF ELEMENTS	NO. OF ACTION ELEMENTS	NO. OF PASSIVE SUPPORT ELEMENTS	NO. OF IDLE ELEMENTS
RIGHT HAND	15	15 (100%)	0 (0%)	0 (0%)
LEFT HAND	15	3 (20%)	13 (80%)	0 (0%)
RIFLE	15	6 (40%)	0 (0%)	9 (60%)

KEY: VERTICAL HATCHING = ACTION ELEMENTS
HORIZONTAL " = SUPPORT "
CROSS " = IDLE "

FIGURE 3

You can see the additional work required to correlate motion mechanics with anatomical causes since we must concurrently consider supporting postural states. The need for unique mathematics is apparent for we are dealing with complex functions. For example, we are interested in the role or contribution of an individual muscle or muscle group in its relationship to an array of essential tasks. This has not been done in the past either by engineers or psychologists to the best of our knowledge for tasks at large.

As stated, it is of primary interest to try and study the relationships that might exist between anatomical structures and their resulting human motion complexes. It becomes obvious that in spite of one or even more than one imposed functional disability, one may perform a defined task satisfactorily. Case studies have been reported where in times of actual injury, humans perform miraculously. Extreme motivation, fear, pain, and a host of other psychological and physiological factors are deeply imbedded. However, if the structure-function pathways did not exist, high motivation alone could not create new physical anatomical entities. Another important spectrum of feasible alternatives derives from the environment not considered in this paper.

In an anatomical review of the upper limb musculo-skeletal system, it is observed that for given muscle dysfunctions, others are often available to assist in the accomplishment of a specified task. Your attention is directed to Table I in which you observe an anatomical matrix of upper limb muscles. The columns are anatomical actions. The rows are the muscles arranged in order from top to bottom to correspond to the proper regions from shoulder to fingers. Notice that in almost all cases, more than one muscle contributes to a given anatomical motion. This is shown by the presence of more than one x in any column.

It is believed that we are getting closer to technical explanations as to why some perform "miraculously" under handicap. It is the motivation that is mirculous, we can do pretty well from the technical or biotechnical point of view.

We are giving serious thought to the development of a functional anatomical-physiological model of man considering all body systems, or subsystems, such as the musculo-skeletal, cardiovascular, nervous, respiratory, gastrointestinal, and genito-urinary, using what physiological feedbacks we know about and can use. Ordinarily we would avoid such complexity, but you must remember that our analysis actually begins with wound tract pathology or tissue trauma assignable to given fragment physics. A task need not suffer if its imposed demands are not beyond certain limits. Table I shows the structure-function relationships for the human upper limb. Functionally, redundancies are present. This is indicated as mentioned in Table I by examining anatomical functions performed by more than one muscle. In order to be more correct the x's should be of variable size to indicate in a more quantitative sense, weighted contributions on the part of the respective muscles. The point is that redundancies in terms of anatomical capability to perform exists to accomplish external actions. They exist in a given anatomical region. They exist also because of anatomical duality since many tasks are not fully demanding in terms of both limbs simultaneously. Postural changes may compensate satisfactorily for such losses.

RELIABILITY OF SYSTEMS. The reliability of a system's performance may be related to the objective which the system is expected to attain. Since we are concerned with the complex living-system coupled to inanimate components, nature's solutions to reliability phenomena are of special interest. This is especially true if the objective function of a given ballistic fragment is to maximize the disruption of this living-performing arrangement. At the biological-cell level the answer to unreliability is the presence of many more cells than are required, most of which are in parallel linkages. According to W. S. McCulloch; Von Foerster expresses the view that whatever else it (an ordered 'living' system) contributes, a redundancy of structure is fundamental.*,** McCulloch also states . . . "that which is redundant is, to the extent that it is redundant, stable."**

Such phenomena, including information, channels, and structure are important if additional insight is to be attained concerning explanations for the adaptation and compensation powers so often demonstrated by human body behavior.

---

*"On Self-Organizing Systems and Their Environments," by H. Von Foerster. Fron "Self-Organizing Systems Proceedings of an Interdisciplinary Conference." Pergamon Press. 1960.

**"The Reliability of Biological Systems," by W. S. McCulloch, same publication.

Reliability can be defined statistically as the probability "P" that an element or a system will perform satisfactorily for a given period of time. Unreliability is the probability of a failure during a specified time and is given as "q" = 1 - P. Here independence assumes that the failure of one element does not affect the probability of failure of any other element. Trauma from various causes has not yet been studied in this sense, but it is being proposed. However, the lack of knowledge of complex biological system behavior would seem to favor the application of stochastic processes at the present time.

If redundancies exist in human biomechanics as well as for the many supporting subsystems, the act of purposely inhibiting any single channel, even if it were feasible to identify it, may not yield applied information most needed at the present time. Other channels, especially those we are not yet aware of, may take over. However, "groups" of redundant channel inhibition may yield significant responses.

If we study task requirements in terms of biomechanical redundancies, we can begin to describe these pathways. It is well known that no task requires all available channels. Hypothetically, then, performance decrement due to a specific channel deficit may not be significant. Our introductory observations regarding given functional losses for defined tasks show some influence on performance behavior. We are probably dealing in some measure with numerical discontinuities and/or non-linear systems.

But what are the boundary conditions? How will these vary from task to task? How much subsystem understanding is required or when can it simply be "blackboxed"? We feel, as others do, that in complex systems in which there is a problem to be solved, it makes sense to speak of the quality of a solution. The method of solution is determined to a large extent by the problem objective. Herein may lie the reason for some of the differences obtained when many views are taken of the same problem. The approach may be identical but inputs may be biased in various ways. Inputs and criteria obviously influence output numbers.

It is interesting to note that research programs in the life sciences are increasingly accentuating the need for increased knowledge concerning fundamental processes, principles, and mechanisms found in biological systems. For the present, we must formulate and use what basic homeostatic mechanisms known by which body processes are regulated. Studies are being made of adaptive and regulatory processes; neural network theory; reflexes and other feedback systems; random redundant processes;

sensing and transducer mechanisms; the specificity of sensory and motor phenomena and other such functions. It has only been in recent years that interest in the study of physical and mathematical principles relating to control systems has become intensified.

Two general approaches to incapacitation assessment are of immediate interest to us, namely, (1) the medical and biophysical aspects of pathological dynamics for certain fragment physics, and, (2) the probable performance alterations for specified tasks due to the traumatic effects of such traumata.

SUMMARY. In summing, it is believed that the following ideas are important:

1. Humans possess a general-purpose type of anatomical structure-function arrangement. Overlap and redundance, indicated by the fact that more than one muscle contributes to the same anatomical function and the presence of duality particularly in the upper and lower limbs, are factors contributing to the nature of this capability. This same idea may be carried down to the cellular level for a single muscle inasmuch as muscle cells exist in a large number of subsets many of which are in parallel. Finally, we might add that even a single living cell may have different capacities depending on many physiological factors including conditioning.

2. In this context it is felt that the load or demand on the general-purpose human is a significant function of task requirements. Quality of performance, imposed physical forces, variable time durations, environmental factors, and motion behavior in 3-dimensional space are items that may vary radically from one task to another.

3. In assessing disability performance a definition of non-disability performance is essential. We recognize the need for unique combinations of talents in this interdisciplinary problem area.

4. A large amount of experimental information concerning man-task dynamics can be generated for specified tasks using conventional work measurement techniques in conjunction with physiological instrumentation and control system knowledge now available.

Appendix

UPPER LIMB MUSCLE CODE

Code	Muscle Name
Shoulder:	
A_1	Trapezius
A_2	Serratus anterior
A_3	Subclavius
A_4	Pectoralis minor
A_5	Pectoralis major
A_6	Subscapularis
A_7	Supraspinatus
A_8	Infraspinatus
A_9	Teres minor
A_{10}	Teres major
A_{11}	Biceps brachii
A_{12}	Coracobrachialis
A_{13}	Triceps brachii
A_{14}	Deltoid
Arm:	
A_{11}	Biceps brachii
A_{13}	Triceps brachii
A_{12}	Coracobrachialis
B_1	Brachioradialis
B_2	Brachialis
B_3	Anconeus
Forearm:	
A_{11}	Biceps brachii
A_{13}	Triceps brachii
B_1	Brachioradialis
B_3	Anconeus
C_1	Supinator
C_2	Pronator quadratus
C_3	Pronator teres
C_4	Flexor carpi radialis
C_5	Extensor carpi radialis longus
C_6	Flexor digitorum sublimis
C_7	Flexor carpi ulnaris
C_8	Extensor carpi radialis brevis
C_9	Extensor carpi ulnaris
C_{10}	Flexor digitorum profundus
C_{11}	Extensor digitorum communis
C_{12}	Palmaris longus
C_{13}	Abductor pollicis longus
C_{14}	Flexor pollicis longus
C_{15}	Extensor indicis proprius
C_{16}	Extensor digiti quinti proprius
C_{17}	Extensor pollicis longus

Code	Muscle Name
Hand:	
C_6	Flexor digitorum sublimis
C_{10}	Flexor digitorum profundus
C_{11}	Extensor digitorum communis
D_1	Extensor pollicis brevis
C_{17}	Extensor pollicis longus
D_2	Abductor pollicis brevis
C_{13}	Abductor pollicis longus
C_{14}	Flexor pollicis longus
D_3	Flexor pollicis brevis
C_{15}	Extensor indicis proprius
C_{16}	Extensor digiti quinti proprius
D_4	Flexor digiti quinti brevis
D_5	Abductor digiti quinti
D_6	Abductor pollicis
D_7	Palmaris brevis
D_8	Opponens pollicis
D_9	Opponens digiti quinti
D_{10}	Lumbricales
D_{11}	Interossei dorsales
D_{12}	Interossei volares

UPPER LIMB MOTOR FUNCTIONS AND CAUSAL FACTORS

ANATOMICAL REGIONS (WITH CAUSAL FACTORS)

	SHOULDER	ARM Fl	ARM Ex	ARM Ab	ARM Ad	ARM Ro M	ARM Ro L	FOREARM Fl	FOREARM Ex	FOREARM Su	FOREARM Pr	HAND AND DIGITS Ha Fl	Ha Ex	Ha Ab	Ha Ad	Fi Fl 2 pr ph	Fi Fl ds ph	Fi Ex ph	Th Fl ls ph	Th Ex ls ph	Th Fl 2d ph	Th Ab	Fi Ex In	Fi Ad In	Fi Ex 5th	Fi Ex 2 ds ph	Th Ex 2d ph	Th Ab mc	Fi Fl 5th mc	Th Fl mc ph	Th Ad	Fi Fl 5th	Fi Ab 5th	Th Ab mc	Th Ro mc	Fi Ab mcmc 5th	Fi Fl mc ph	Fi Ro mc 5th	Fi Ab 4	Fi Ad 4	
SHOULDER:																																									SHOULDER:
A_1	X																																								A_1
A_2	X																																								A_2
A_3	X																																								A_3
A_4	X																																								A_4
A_5	X	X		X																																					A_5
A_6	X	X	X	X	X	X																																			A_6
A_7	X	X		X		X	X																																		A_7
A_8	X			X	X		X																																		A_8
A_9	X				X		X																																		A_9
A_{10}			X		X	X																																			A_{10}
A_{11}	X		X					X		X																															A_{11}
A_{12}		X		X																																					A_{12}
A_{13}			X					X																																	A_{13}
A_{14}		X	X	X																																					A_{14}
ARM:																																									ARM:
A_{11}	X	X						X		X																															A_{11}
A_{13}			X					X																																	A_{13}
A_{12}		X		X																																					A_{12}
B_1		X						X																																	B_1
B_2								X																																	B_2
B_3									X																																B_3
FOREARM:																																									FOREARM:
A_{11}	X	X						X		X																															A_{11}
A_{13}			X					X																																	A_{13}
B_1		X						X																																	B_1
B_2								X																																	B_2
B_3									X																																B_3
C_1										X																															C_1
C_2											X																														C_2
C_3											X	X																													C_3
C_4												X		X																											C_4
C_5													X	X	X																										C_5
C_6												X		X	X																										C_6
C_7												X			X																										C_7
C_8													X		X																										C_8
C_9												X			X																										C_9
C_{10}												X				X	X	X																							C_{10}
C_{11}													X					X																							C_{11}
C_{12}																						X																			C_{12}
C_{13}														X	X			X																							C_{13}
C_{14}																			X		X		X	X																	C_{14}
C_{15}																									X																C_{15}
C_{16}																											X														C_{16}
C_{17}														X																											C_{17}
HAND:																																									HAND:
C_6												X			X																										C_6
C_{10}												X				X	X																								C_{10}
C_{11}													X													X															C_{11}
D_1														X						X																					D_1
C_{17}														X													X														C_{17}
D_2																						X																			D_2
C_{13}														X						X		X																			C_{13}
C_{14}																				X		X							X												C_{14}
D_3																			X		X							X													D_3
C_{15}																							X	X																	C_{15}
C_{16}																									X																C_{16}
D_4																																X									D_4
D_5																																	X								D_5
D_6																															X										D_6
D_7																																									D_7
D_8																														X				X	X						D_8
D_9																													X							X		X			D_9
D_{10}																										X											X				D_{10}
D_{11}																										X											X		X		D_{11}
D_{12}																										X											X			X	D_{12}

TABLE I

(See Appendix for identification of code)

This page intentionally left blank.

HALF-NORMAL PLOTS
FOR MULTI-LEVEL FACTORIAL EXPERIMENTS

S. A. Krane
C-E-I-R, Inc., Dugway Field Office

1. INTRODUCTION. Half-normal plots for the interpretation of 2^p factorial experiments have been developed and popularized largely through the work of Cuthbert Daniel (see Daniel [1956] and [1959]). In this method the $2^p - 1$ main effects and interactions are estimated from observations on the 2^p treatment combinations. The empirical cumulative distribution of these estimates is then graphically compared with a cumulative distribution derived from a normal population. A rationale for this procedure is found in the approximate normality of the null distribution of the estimates, based upon normality of experimental errors or upon the tendency embodied in the Central Limit Theorem. According to Daniel, the half-normal plot permits the analyst to judge the reality of the largest main effects and interactions and serves to indicate bad values, heteroscedasticity, dependence of variance on mean and some types of defective randomization. The object of the present paper is to indicate and illustrate possible applications of half-normal plots to balance multi-level factorial experiments in general.

2. AN EXAMPLE. It appears easiest to introduce the technique of half-normal plotting for balanced multi-level factorial experiments in the context of a particular example. For this purpose we shall employ Example 8.1 of Davies [1954]. According to the authors (p. 291, "the data . . . are taken from the results of an investigation into the effects on the physical properties of vulcanized rubber of varying a number of factors, the property recorded being the wear resistance of the samples, and the factors being:

A five qualities of filler
B three methods of pretreatment of the rubber
C four qualities of the raw rubber . . ."

The data are reproduced in Table 1. From the data, the author develops the usual analysis of variance as shown in Table 2. The interpretation (Davies [1954, p. 296]) notes the significance of all main effects and two-factor interactions when tested against the three-factor interaction as error.

Table 1. (Table 8.1 of Davies [1954])
DATA OF A 5 x 3 x 4 FACTORIAL EXPERIMENT
WEAR RESISTANCE OF VULCANISED RUBBER

Level of factor A	Level of factor C: 1			2			3			4		
	Level of factor B: 1	2	3	Level of factor B: 1	2	3	Level of factor B: 1	2	3	Level of factor B: 1	2	3
1	404	478	530	381	429	528	316	376	390	423	482	550
2	392	418	431	239	251	249	186	207	194	410	416	452
3	348	381	460	327	372	482	290	315	350	383	376	496
4	296	291	333	165	232	242	158	279	220	301	306	330
5	186	198	225	129	157	197	105	163	190	213	200	255

Table 2. (Table 8.16 of Davies [1954])
ANALYSIS OF VARIANCE OF TABLE 8.1

Source of Variation	Sum of squares	Degrees of freedom	Mean square	Variance ratio
Between levels of factor A ..	478,463	4	119,616	374+
B ..	52,794	2	26,397	82.5+
C ..	150,239	3	50,080	156+
Interactions AB	16,807	8	2,101	6.57+
AC	53,890	12	4,491	14.0+
BC	6,416	6	1,069	3.34*
Remainder = interaction ABC ..	7,688	24	320	
Total	766,297	59		

* Denotes significant, that is $\geq$ 5% value but < 1% value.

+ Denotes highly significant, that is $F \geq 1\%$ value.

In order to analyze the given experimental data by half-normal plotting, we shall reduce the data to single degree of freedom sums of squares. The method to be used depends upon the definition of complete sets of orthogonal contrasts for each of the factors A, B, and C. This definition generally is somewhat artibrary, but it is our experience that an experimenter familiar with the nature of the factor levels and the purpose of the experiment can, in most instances, provide sufficient justification for the prior definition of a meaningful complete set of single degree of freedom orthogonal contrasts among the levels. The use of orthogonal polynomials for quantitative levels is often indicated, while for qualitative levels, meaningful comparisons among certain levels are often obvious. On occasion, only a partial set of orthogonal comparisons will appear to be of intrinsic value and it may be necessary to complete the orthogonal set by adding contrasts of no apparent importance. In the absence of useful information on the nature of the levels (except that they are all qualitative) in the present example, we shall be totally arbitrary in defining the contrasts, but will attempt to indicate their potential interpretations. These contrasts are shown in Table 3. For factor A, contrast A_0, is the "null" or "average" contrast*, while A_1 compares the average of levels 1 and 2 against the average of levels 3, 4 and 5, A_2 compares level 1 vs. level 2, A_3 compares level 3 against the average of levels 4 and 5 and A_4 compares level 4 with level 5. The orthogonality of the set is evident in that the coefficients sum to zero for all contrasts except the null contrast and the sum of products of coefficients is zero for all pairs of contrasts. For factor B, the non-null contrasts compare level 1 with level 2 with the average of levels 1 and 3. (In another context, B_1 and B_2 are the orthogonal polynomials for three equally spaced levels, B_1 being the linear contrast and B_2 the quadratic.) For factor C, the contrasts C_1, C_2 and C_3 make the following comparisons among levels, respectively: (1 and 2) vs. (3 and 4), (1 and 3) vs. (2 and 4) and (1 and 4)

---

* Daniel 1962 has suggested the term "null" is inappropriate because of the generally positive expectation of this contrast. We chose the term because (i) it is connoted by our zero subscript notation, (ii) this contrast is not a comparison among levels, and (iii) this contrast is generally "of no consequence" in the analysis.

Table 3. ORTHOGONAL CONTRASTS EMPLOYED

Factor A

Contrast	Level 1	2	3	4	5	Sum of Squares
A_0	+1	+1	+1	+1	+1	5
A_1	+3	+3	-2	-2	-2	30
A_2	+1	-1	0	0	0	2
A_3	0	0	+2	-1	-1	6
A_4	0	0	0	+1	-1	2

Factor B

Contrast	Level 1	2	3	Sum of Squares
B_0	+1	+1	+1	3
B_1	-1	0	+1	2
B_2	-1	+2	-1	6

Factor C

Contrast	Level 1	2	3	4	Sum of Squares
C_0	+1	+1	+1	+1	4
C_1	-1	-1	+1	+1	4
C_2	-1	+1	-1	+1	4
C_3	+1	-1	-1	+1	4

vs. (2 and 3). These contrasts would be of interest, e. g., in the event that factor C incorporated two subfactors, say D and E, where levels 1 and 2 are at the low level of D and levels 3 and 4 at the high level of D, while levels 1 and 3 are at the low level of E and 2 and 4 at the high level of E. Then C_1 is the effect of D, C_2 is the effect of E and C_3 is the interaction of D and E.

The three sets of contrasts (A_1, A_2, A_3, A_4), (B_1, B_2) and (C_1, C_2, C_3) will provide a basis for reducing the sums of squares for factor A (4 d. f.), factor B (2 d. f.) and factor C (3 d. f.) to independent single degree of freedom sums of squares. It remains to develop such a basis for the two- and three-factor interactions. A natural method for accomplishing this is the extension of the original single factor contrast sets to interaction contrast sets. This method is exemplified in Table 4 for Factors B and C. All possible combinations of the levels of B and C are employed as columns, while rows are contrasts. For any combination of a particular level, say i, of B with a particular level, say j, of C, the coefficient in the contrast B_qC_r is obtained by multiplication of the coefficient of level i of B in the contrast B_q by the coefficient of level j or C in the contrast C_r. Sums of squares of the B and C contrasts may be obtained by multiplication of the corresponding sums of squares for B and for C.

Of the 12 orthogonal contrasts in Table 4, B_0C_0 is the null contrast while the contrasts B_0C_1, B_0C_2, B_0C_3, B_1C_0 and B_2C_0 are simply the original contrasts C_1, C_2, C_3, B_1 and B_2, respectively, averaged over all levels of the other factor. The six contrasts B_1C_1, B_1C_2, B_1C_3, B_2C_1, B_2C_2, B_2C_3 are new and constitute a basis for partitioning the BC interaction sum of squares (6 d. f.) into orthogonal single degree of freedom sums of squares. Application of this method will likewise produce bases for partitioning the sums of squares for AB (8 d. f.), AC (12 d. f.) and ABC (24 d. f.).

The above method of defining interaction contrasts is incorporated in the method we employ for calculating half-normal variates by desk

Table 4. ORTHOGONAL CONTRASTS FOR FACTORS B AND C

Level of C	1			2			3			4			Sum of Squares
Level of B	1	2	3	1	2	3	1	2	3	1	2	3	
Contrast B_0C_0	+1	+1	+1	+1	+1	+1	+1	+1	+1	+1	+1	+1	12
B_0C_1	-1	-1	-1	-1	-1	-1	+1	+1	+1	+1	+1	+1	12
B_0C_2	-1	-1	-1	+1	+1	+1	-1	-1	-1	+1	+1	+1	12
B_0C_3	+1	+1	+1	-1	-1	-1	-1	-1	-1	+1	+1	+1	12
B_1C_0	-1	0	+1	-1	0	+1	-1	0	+1	-1	0	+1	8
B_1C_1	+1	0	-1	+1	0	-1	-1	0	+1	-1	0	+1	8
B_1C_2	+1	0	-1	-1	0	+1	+1	0	-1	-1	0	+1	8
B_1C_3	-1	0	+1	+1	0	-1	+1	0	-1	-1	0	+1	8
B_2C_0	-1	+2	-1	-1	+2	-1	-1	+2	-1	-1	+2	-1	24
B_2C_1	+1	-2	+1	+1	-2	+1	-1	+2	-1	-1	+2	-1	24
B_2C_2	+1	-2	+1	-1	+2	-1	+1	-2	+1	-1	+2	-1	24
B_2C_3	-1	+2	-1	+1	-2	+1	+1	-2	+1	-1	+2	-1	24

calculator. A sample work sheet for this method is shown in Table 5a and 5b*. Section I of this table is merely a recopying of the original data from Table 1. In this section a column of five numbers represents the observations for the five levels of A over a particular one of the twelve combinations of levels of B and C represented by columns. Section II is computed by operating on these columns with the contrasts A_0, A_1, A_2, A_3, A_4. For example, products of the coefficients of A_0 with the corresponding elements of a particular column are formed and these five products are summed and entered in Section II in the first row of that column. Similarly, sums of products of coefficients of A_1 with corresponding elements of columns are entered in the second row of Section II, and so forth. Thus each element of Section II is formed as a sum of products of coefficients of, say, A_p with corresponding observations and is entered on the ($p+1$)-th row of the appropriate column.

Section II may be visualized as an aggregation of 20 rows of three elements each, where each row corresponds to a particular contrast of A and a particular level of C. The three elements of each row correspond to the three levels of B. Now Section III is formed from Section II by summing products of coefficients of the B contrasts and corresponding elements of each row of three. Each such sum of products is entered in the corresponding row, with the sum of products from coefficients of B_q entered as the ($q + 1$)-th element of that row.

Section III may be visualized as comprising four sub-sections of 15 elements each, with the elements of a sub-section corresponding to a particular contrast of AB (i. e., a particular combination of a contrast of A and a contrast of B) and the four sub-sections corresponding to the four levels of C. Section IV is formed from Section III by summing products of coefficients of the C contrasts and identically placed elements from the four corresponding sub-sections. The sum of products is entered in the corresponding place of one of the sub-sections of Section IV, with the sum of products from coefficients of C_r entered in the ($r+1$)-th sub-section.

---

* This calculation method is essentially the same as that given in Appendix 8G of Davies [1954, pp. 363-6]. In some instances the format in Davies (with the addition of final column of half-normal variates) may be preferred.

Table 5a. COMPUTATION OF SINGLE DEGREES OF FREEDOM

I	404	478	530	381	429	528	316	376	390	423	482	550
	392	418	431	239	251	249	186	207	194	410	416	452
	348	381	460	327	372	482	290	315	350	383	376	496
	296	291	333	165	232	242	158	279	220	301	306	330
	186	198	225	129	157	197	105	163	190	213	200	255
II	1626	1766	1979	1241	1441	1698	1055	1340	1344	1730	1780	2083
	728	948	847	618	518	489	400	235	232	705	930	844
	12	60	99	142	178	279	130	169	196	13	66	98
	214	273	362	360	355	525	317	188	290	252	246	407
	110	93	108	36	75	45	53	116	30	88	106	75
III	5371	353	-73	4380	457	-57	3739	289	281	5593	353	-253
	2523	119	321	1625	-129	-71	867	-168	-162	2479	139	311
	171	87	9	599	137	-65	495	66	12	177	85	21
	849	148	-30	1240	165	-175	795	-27	-231	905	155	-167
	311	-2	-32	156	9	69	199	-23	149	269	-13	49
IV	19083	1452	-102	-419	-168	158	863	168	-518	2845	-40	-550
	7494	-39	399	-802	-19	-101	714	59	81	2510	555	865
	1442	375	-23	-98	-73	89	110	69	-65	-746	-31	83
	3789	441	-603	-389	-185	-193	501	199	-81	-281	165	209
	935	-29	235	1	-43	161	-85	21	1	225	-1	-201

Table 5b. COMPUTATION ON SINGLE DEGREES OF FREEDOM (Cont'd)

V	$A_0B_0C_0$	$A_0B_1C_0$	$A_0B_2C_0$	$A_0B_0C_1$	$A_0B_1C_1$	$A_0B_2C_1$	$A_0B_0C_2$	$A_0B_1C_2$	$A_0B_2C_2$	$A_0B_0C_3$	$A_0B_1C_3$	$A_0B_2C_3$
	$A_1B_0C_0$	$A_1B_1C_0$	$A_1B_2C_0$	$A_1B_0C_1$	$A_1B_1C_1$	$A_1B_2C_1$	$A_1B_0C_2$	$A_1B_1C_2$	$A_1B_2C_2$	$A_1B_0C_3$	$A_1B_1C_3$	$A_1B_2C_3$
	$A_2B_0C_0$	$A_2B_1C_0$	$A_2B_2C_0$	$A_2B_0C_1$	$A_2B_1C_1$	$A_2B_2C_1$	$A_2B_0C_2$	$A_2B_1C_2$	$A_2B_2C_2$	$A_2B_0C_3$	$A_2B_1C_3$	$A_2B_2C_3$
	$A_3B_0C_0$	$A_3B_1C_0$	$A_3B_2C_0$	$A_3B_0C_1$	$A_3B_1C_1$	$A_3B_2C_1$	$A_3B_0C_2$	$A_3B_1C_2$	$A_3B_2C_2$	$A_3B_0C_3$	$A_3B_1C_3$	$A_3B_2C_3$
	$A_4B_0C_0$	$A_4B_1C_0$	$A_4B_2C_0$	$A_4B_0C_1$	$A_4B_1C_1$	$A_4B_2C_1$	$A_4B_0C_2$	$A_4B_1C_2$	$A_4B_2C_2$	$A_4B_0C_3$	$A_4B_1C_3$	$A_4B_2C_3$
VI	60	40	120	60	40	120	60	40	120	60	40	120
	360	240	720	360	240	720	360	240	720	360	240	720
	24	16	48	24	16	48	24	16	48	24	16	48
	72	48	144	72	48	144	72	48	144	72	48	144
	24	16	48	24	16	48	24	16	48	24	16	48
VII	6069348	52708	87	2926	706	208	12413	706	2236	134900	40	2521
	156000	6	221	1787	2	14	1416	15	9	17500	1283	1039
	86640	8789	11	400	333	165	504	298	88	23188	60	144
	199396	4052	2525	2102	713	259	3486	825	46	1097	567	303
	36426	53	1151	0	116	540	301	28	1	2109	0	842
VIII	2464	230	-9	-54	-27	14	111	27	-47	367	-6	-50
	395	-2	15	-42	-1	-4	38	4	3	132	36	32
	294	94	-3	-20	-18	13	22	17	-9	-152	-8	12
	447	64	-50	-46	-27	-16	59	29	-7	-33	24	17
	191	-7	34	0	-11	23	-17	5	1	46	-0	-29

The method of calculation of Sections II, III and IV may not be apparent at first glance, but verifying part or all of the data in Table 5a from the description above should help to clarify the process. Computing clerks will find it helpful to write the coefficients of each contrast on a strip of paper, appropriately oriented vertically or horizontally and spaced so that when overlaid on the worksheet each coefficient appears adjacent to the element to be multiplied.

Section V (Table 5b) merely identifies the elements of Section IV and subsequent sections according to the contrasts they represent. This identification is, of course, highly systematic and might well be omitted when familiarity with the method is attained.

Section VI contains the "divisors", obtained by multiplying the sums of squares of the coefficients of the contrasts A_p, B_q, C_r appropriate to each element, as found in Table 3.

Section VII contains the single degree of freedom sums of squares corresponding to each contrast. Each element is obtained by squaring an element of Section IV, dividing by the corresponding element of Section VI and entering in the corresponding place of Section VII.

Section VIII contains the half-normal variate values, each of which is computed as the square root of the corresponding element of Section VII, positive or negative according to the sign of the corresponding element of Section IV. (It would perhaps have been advisable to include the first decimal of each of these values in order to discriminate more fully among them.)

Certain check computations in the method have been omitted, but an over-all check can be readily obtained from Section VII by comparing sums of these single degree of freedom sums of squares with the usual analysis of variance of Table 2. These checks are indicated in Table 6. It will be noted that all sums of squares agree with Table 2 within the expected rounding error accumulated from Section VII.

The half-normal variates must now be ordered by magnitude before plotting. This ordering is shown in Table 7, along with an identification of the contrast represented (letters with subscripted zeroes have been dropped) and the appropriate quantile of the empirical distribution, defined by

Table 6. DEVELOPMENT OF USUAL ANALYSIS OF VARIANCE

Source	d.f.	Contrasts	S.S.
A	4	$A_1B_0C_0$ $A_2B_0C_0$ $A_3B_0C_0$ $A_4B_0C_0$	478462
B	2	$A_0B_1C_0$ $A_0B_2C_0$	52795
C	3	$A_0B_0C_1$ $A_0B_0C_2$ $A_0B_0C_3$	150239
AB	8	$A_1B_1C_0$ $A_2B_1C_0$ $A_3B_1C_0$ $A_4B_1C_0$	
		$A_1B_2C_0$ $A_2B_2C_0$ $A_3B_2C_0$ $A_4B_2C_0$	16808
AC	12	$A_1B_0C_1$ $A_2B_0C_1$ $A_3B_0C_1$ $A_4B_0C_1$	
		$A_1B_0C_2$ $A_2B_0C_2$ $A_3B_0C_2$ $A_4B_0C_2$	
		$A_1B_0C_3$ $A_2B_0C_3$ $A_3B_0C_3$ $A_4B_0C_3$	53890
BC	6	$A_0B_1C_1$ $A_0B_1C_2$ $A_0B_1C_3$	
		$A_0B_2C_1$ $A_0B_2C_2$ $A_0B_2C_3$	6417
ABC	24	$A_1B_1C_1$ $A_2B_1C_1$ $A_3B_1C_1$ $A_4B_1C_1$	
		$A_1B_1C_2$ $A_2B_1C_2$ $A_3B_1C_2$ $A_4B_1C_2$	
		$A_1B_1C_3$ $A_2B_1C_3$ $A_3B_1C_3$ $A_4B_1C_3$	
		$A_1B_2C_1$ $A_2B_2C_1$ $A_3B_2C_1$ $A_4B_2C_1$	
		$A_1B_2C_2$ $A_2B_2C_2$ $A_3B_2C_2$ $A_4B_2C_2$	
		$A_1B_2C_3$ $A_2B_2C_3$ $A_3B_2C_3$ $A_4B_2C_3$	7690
Total	59		766301
Mean	1	$A_0B_0C_0$	6069348
Raw total	60		6835649

Table 7. HALF-NORMAL VARIATES

Order k	Variate X_k	Contrast	Quantile P_k	Order k	Variate X_k	Contrast	Quantile P_k
60	2464	Null		30	24	$A_3B_1C_3$	.5000
59	447	A_3	.9915	29	23	$A_4B_1C_1$	.4831
58	395	A_1	.9746	28	22	A_2C_2	.4661
57	367	C_3	.9576	27	20	$-A_2C_1$	.4492
56	294	A_2	.9407	26	18	$-A_2B_1C_1$	.4322
55	230	B_1	.9237	25	17	$A_3B_2C_3$	.4153
54	191	A_4	.9068	24	17	$-A_4C_2$	.3983
53	152	$-A_2C_3$	.8898	23	17	$A_2B_1C_2$	.3814
52	132	A_1C_3	.8729	22	16	$-A_3B_2C_1$	.3644
51	111	C_2	.8559	21	15	A_1B_2	.3475
50	94	A_2B_1	.8390	20	14	B_2C_1	.3305
49	64	A_3B_1	.8220	19	13	$A_2B_2C_1$	.3136
48	59	A_3C_2	.8051	18	12	$A_2B_2C_3$	.2966
47	54	$-C_1$	.7881	17	11	$-A_4B_1C_1$	.2797
46	50	$-A_3B_2$	.7712	16	9	$-A_2B_2C_2$	.2627
45	50	$-B_2C_3$	.7542	15	9	$-B_2$	.2458
44	47	$-B_2C_2$	.7373	14	8	$-A_2B_1C_3$	.2288
43	46	A_4C_3	.7203	13	7	$-A_4B_1$	.2119
42	46	$-A_3C_1$	.7034	12	7	$-A_3B_2C_2$	.1949
41	42	$-A_1C_1$	.6864	11	6	$-B_1C_3$	.1780
40	38	A_1C_2	.6695	10	5	$A_4B_1C_2$	.1610
39	36	$A_1B_1C_3$	.6525	9	4	$A_1B_1C_2$	.1441
39	34	A_4B_2	.6356	8	4	$-A_1B_2C_1$	.1271
37	33	$-A_3C_3$	.6186	7	3	$-A_2B_2$	.1102
36	32	$A_1B_2C_3$	.6017	6	3	$A_1B_2C_2$	.0932
35	29	$-A_4B_2C_3$	.5847	5	2	$-A_1B_1$	.0763
34	29	$A_3B_1C_2$	.5678	4	1	$-A_1B_1C_1$	.0593
33	27	$-A_3B_1C_1$	.5508	3	1	$A_4B_2C_2$	.0424
32	27	$-B_1C_1$	.5339	2	0	$-A_4B_1C_3$	.0254
31	27	B_1C_2	.5169	1	0	A_4C_1	.0085

$$P_k = \frac{2k - 1}{2n}$$

where k is the rank order and n is the number of variates. Here, as in most instances, it seems appropriate that the null contrast be excluded from the variates to be examined. The sign of the contrast is now attached to the label and only positive variates are plotted.

The variate values and quantiles are next plotted on half-normal probability paper (as in Figure 1) for interpretation. Discussion of the interpretation phase of the analysis of this example will be deferred to a later section.

3. SOME THEORY*. At this point we shall touch briefly on some theoretical aspects of the development of half-normal variates from multi-level factorial experiments. To simplify the discussion we shall assume that we are concerned with a three-factor experiment, although it should be remembered that the theory and methodology apply with equal validity to any number of factors.

We denote by y_{hij} the observation obtained with factor A at level h, factor B at level i and factor C at level j, where h = 1, 2, . . . , a; i = 1, 2, . . . , b; j = 1, 2, . . . , c. The coefficients of the orthogonal contrasts for factor A will be indicated by a_{ph}, denoting the coefficient for level h in the p-th contrast. Similarly the coefficients of the contrasts for factors B and C are denoted b_{qi} and c_{rj}, respectively. We assume that for each factor there is a null contrast, these being denoted A_0, B_0, C_0 and defined by

$$a_{oh} = b_{oi} = c_{oj} = 1; \text{ all } h, i, j.$$

---

* This section is based on well-known results concerning distributions of linear functions of random variables and may be verified by reference to standard introductory texts on mathematical and theoretical statistics.

Furthermore, by the definition of orthogonal contrasts,

$$\sum_h a_{ph} = \sum_i b_{qi} = \sum_j c_{rj} = 0$$

$$p = 1, 2, \ldots, a-1; \quad q = 1, 2, \ldots, b-1; \quad r = 1, 2, \ldots, r-1;$$

and

$$\sum_h a_{ph}\, a_{p'h} = \sum_i b_{qi}\, b_{q'i} = \sum_j c_{rj}\, c_{r'j} = 0$$

$$p \neq p'; \quad q \neq q'; \quad r \neq r'.$$

The three-factor contrasts are defined by

$$(A_p B_q C_r) = \sum_h \sum_i \sum_j a_{ph}\, b_{qi}\, c_{rj}\, y_{hij};$$

$$p = 0, 1, \ldots, a-1; \quad q = 0, 1, \ldots, b-1; \quad r = 0, 1, \ldots, c-1.$$

Suppose that there are no treatment effects*, i.e.,

$$E\{y_{hij}\} = \mu; \text{ all } h, i, j;$$

and that the experimental errors are independent and have constant variance for all observations, i.e.,

$$E\{(y_{hij} - \mu)^2\} = \sigma^2; \text{ all } h, i, j.$$

---

*The symbol E { } denotes the mathematical expectation operator.

Then

$$E\{A_p B_q C_r\} = 0,$$

unless

P = 0, q = 0 and r = 0, in which case

$$E\{A_0 B_0 C_0\} = abc\ \mu .$$

Furthermore*,

$$V\{A_p B_q C_r\} = (\sum_h a_{ph}^2)(\sum_i b_{qi}^2)(\sum_j c_{rj}^2)\ \sigma^2 .$$

Denote by Y_{pqr} the variate defined by

$$Y_{pqr} = (A_p B_q C_r) \Big/ \sqrt{(\sum_h a_{ph}^2)\ (\sum_i b_{qi}^2)\ (\sum_j c_{rj}^2)} .$$

Then

$$E\{Y_{000}\} = \sqrt{abc}\ \mu ;$$

$$E\{Y_{pqr}\} = 0, \text{ unless } p = 0,\ q = 0,\ r = 0;$$

$$V\{Y_{pqr}^2\} = \sigma^2 .$$

If the experimental errors are normally distributed, then the Y_{pqr} are normally distributed. (Under fairly weak assumptions the Y_{pqr} will tend to be normally distributed in large experiments even for non-normal distributions of experimental error.) Then the non-negative half-normal variates,

*The symbol $V\{\ \}$ denotes the variance operator, $V\{X\} = E\{(X-E\{X\})^2\}$.

$$X_{pqr} = \left| Y_{pqr} \right| ,$$

$$= \sqrt{(A_p B_q C_r)^2 / (\sum_h a_{ph}^2) (\sum_i b_{qi}^2) (\sum_j c_{rj}^2)}; \; p, q, r \neq 0, 0, 0;$$

are indeed distributed according to the half-normal density

$$f(x) = \sqrt{2/\pi \sigma^2} \quad \exp(-X^2/2\sigma^2), \qquad X \geq 0$$
$$X \leq 0.$$

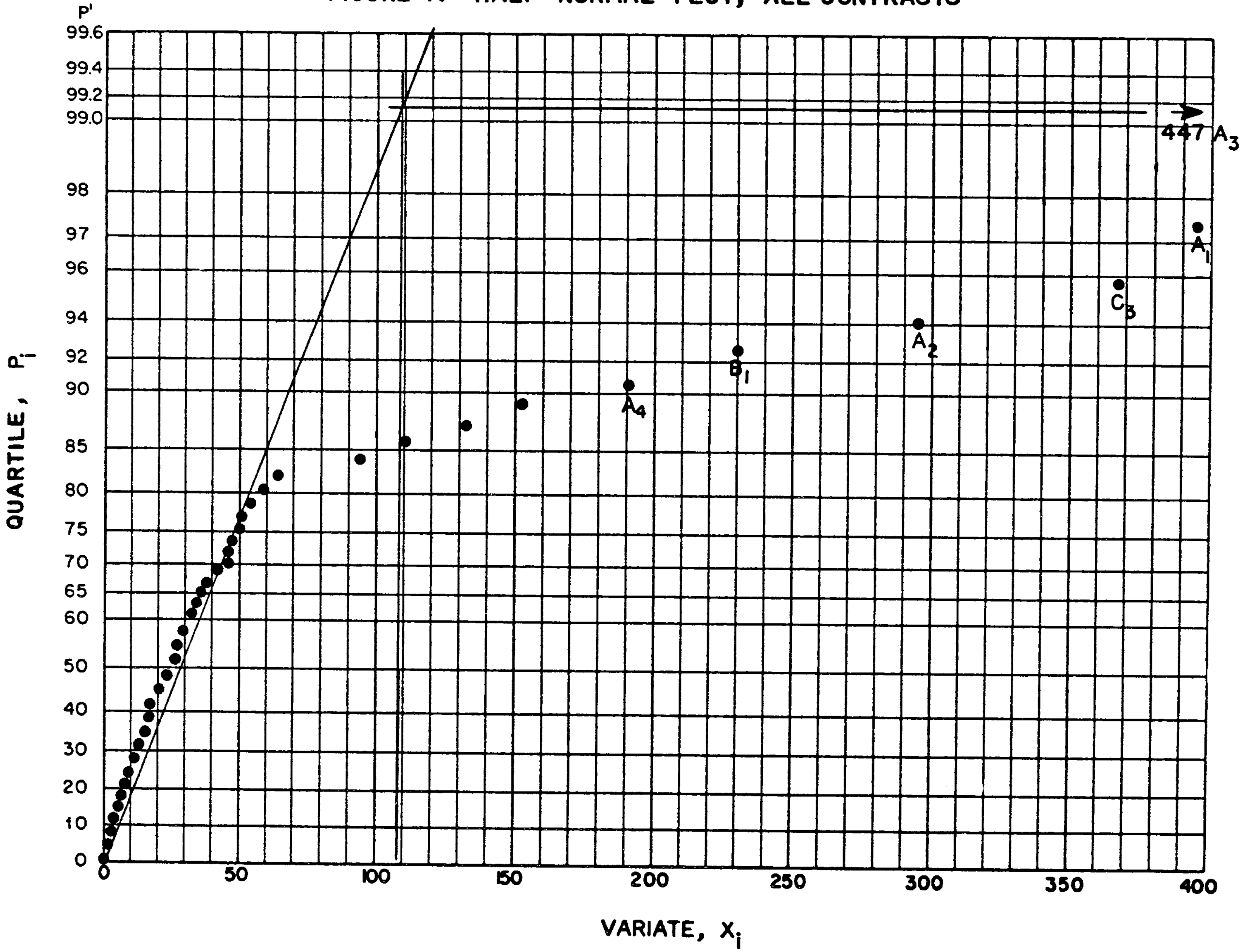
From this result, the half-normal variates for multi-level experiments may be seen to be essentially equivalent to those for 2^p experiments, making the work of Daniel 1959 and Birnbaum 1959 relevant to the interpretation.

4. INTERPRETATION OF EXAMPLE. We shall turn now to the interpretation of the example given earlier. Some difficulty will be experienced because of our ignorance of the precise nature of the factors and their levels, the experimental techniques and the observations themselves, but we shall attempt to proceed along lines suggested by Daniel for 2^p experiments.

To recapitulate the results of Section 2, we have, in Table 7, 59 ordered variates $X_k \equiv X_{pqr}$ whose empirical cumulative distribution should resemble the cumulative half-normal distribution under the hypothesis that there are no treatment effects. We have plotted these values against their quantiles in Figure 1, where they should be approximately linear under the null hypothesis.

We note at a glance that the plotted points are markedly and systematically non-linear. In fact, a little preliminary geometrical construction leads us to believe that a number of the variates are too large to have arisen by chance under the null hypothesis. The rationale for this belief is as follows. Under the null hypothesis the standard deviation, σ, is

FIGURE 1. HALF-NORMAL PLOT, ALL CONTRASTS



directly approximated by the value of X_m, where

$$m = (0.683\,n + 0.5),$$
$$= 41, \text{ approximately.}$$

From Table 7,

$$X_{41} = 42.$$

Then, under the null hypothesis, the plotted points should lie near a straight line through the origin and the point (X_m, P_m), indicated in Figure 1. Should the largest X lie "far enough" to the right of this line it is reasonable to presume that it did not arise by chance under the null hypothesis. It may then be taken as real and the next largest X promoted to the largest. This is roughly equivalent to increasing the ordinate of the second point to that of the first point. Should this replotted point also lie "far enough" to the right of the line, it too may be judged real and excluded, promoting the next X to the largest, etc. In Figure 1, we make a crude test of the largest values by constructing a horizontal through the largest point to intersect the previously constructed empirical cumulative distribution line. From this intersection we drop a vertical line and observe that all contrasts represented by points lying to the right of this vertical would have to be excluded before the largest X would lie on or above the original c. d. line. In this crude manner we judge from Figure 1 that six to ten of the largest values of X would be unlikely to occur under the null hypothesis. This graphical construction is no "exact" test; in fact it is rather likely that one or more contrasts would be judged "real" in this manner even if the null hypothesis did, in fact, hold. There is one element of conservativism in this procedure, in that the plotted c. d. line is based upon <u>all</u> contrasts, while a c. d. line based only on contrasts not judged "real" at this stage would lie to the left of the original line.

Let us tentatively suppose that the six largest contrasts (A_3, A_1, C_3, A_2, B_1, A_4) are real, considering (after Daniel [1959, p. 315]) their simple <u>names</u>, as well as their magnitudes relative to the rest of the set. We plot anew the 53 remaining contrasts in Figure 2. Actually, in addition to the ten largest remaining contrasts, only a fraction of the points are plotted, together with the c. d. line through (X_m, P_m), where

$$m = (0.683)(53) + 0.5,$$

$$= 37, \text{ approximately.}$$

The values of P_k are, of course, recalculated for $n = 53$. It appears reasonable to judge from this plot that the four largest contrasts (A_2C_3, A_1C_3, C_2, A_2B_1) are real.

A final plot of the values obtained after eliminating the ten largest values is shown in Figure 3. It appears in this plot that all real effects have been removed, with a residual error standard deviation approximately equal to

$$X^2_{34} = 841$$

(The actual mean square of the 49 residual contrasts is 816.)

Some further details of interpretation might be attempted. For example, there is a suggestion in Figure 1 and in Table 7 that there may have been plot-splitting, with factor B applied within plots. This also appears plausible from the rudimentary information given as to the nature of this factor. A further plotting, not shown here, in which contrasts including B_1 or B_2 were separated from those containing B_0 suggests a whole plot standard deviation of about 50-60 and a split-plot standard deviation of about 20-25.

5. COMPUTER USE. We have used half-normal plots for multi-level factorial experiments for almost two years. Our first major attempt to employ this technique was in the analysis of an unreplicated 10 x 5 x 3 x 2^2 experiment. The factor levels in this experiment were applied in a split-split-split plot design and certain problems of variance heterogeneity were apparent. The half-normal plotting of this data was sufficiently informative that it appeared worthwhile to develop a program for the IBM 1620 to be employed in computing half-normal variates from multi-level factorial data. This program, Single Degree of Freedom Analysis of Variance (SIDOF), has a capacity of eight factors, each at two to ten levels. It requires as input the observations and normalized vectors of contrast coefficients α_p, β_q, γ_r, etc., where

FIGURE 2. HALF-NORMAL PLOT, SIX LARGEST CONTRASTS OMITTED

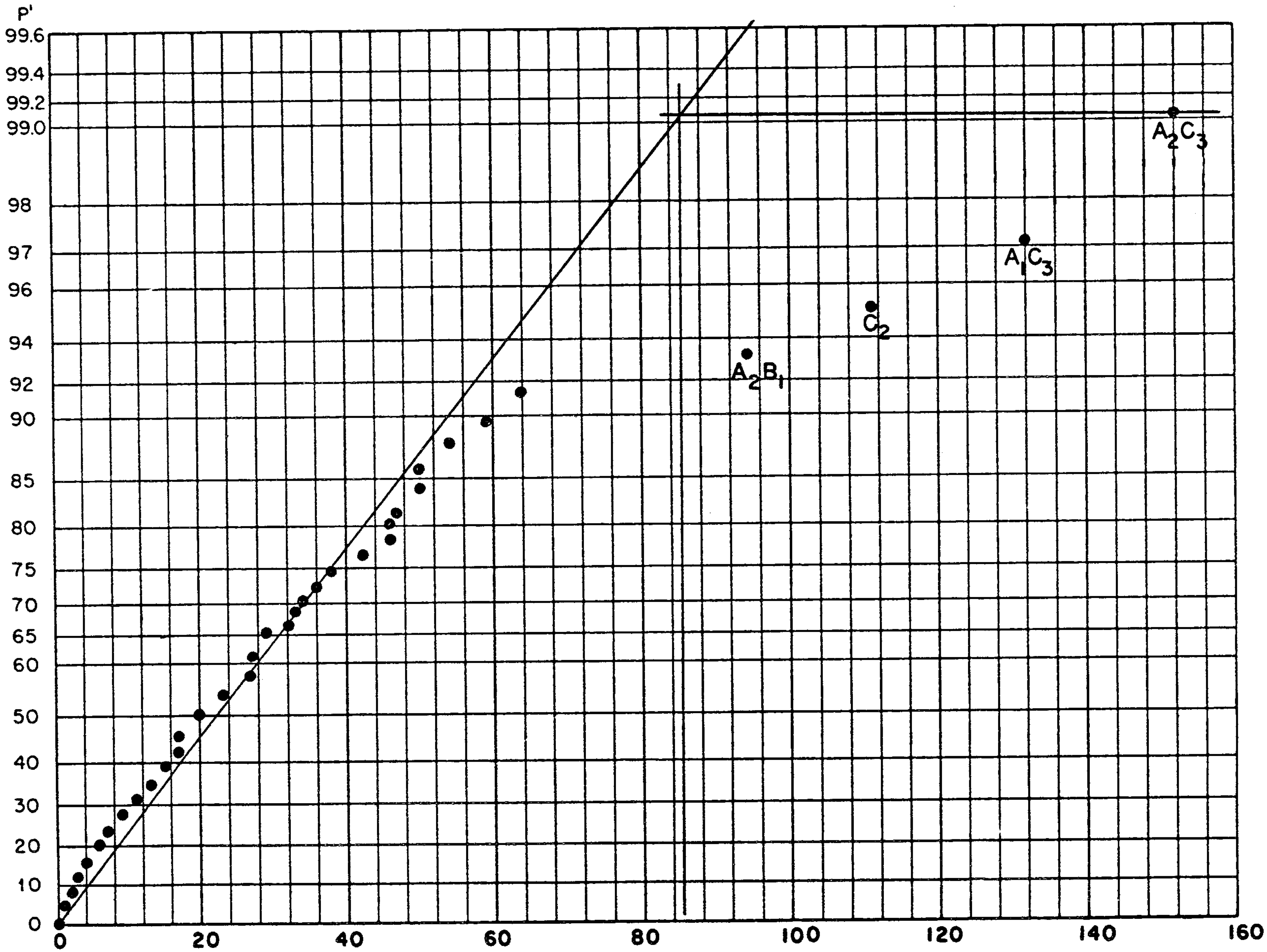
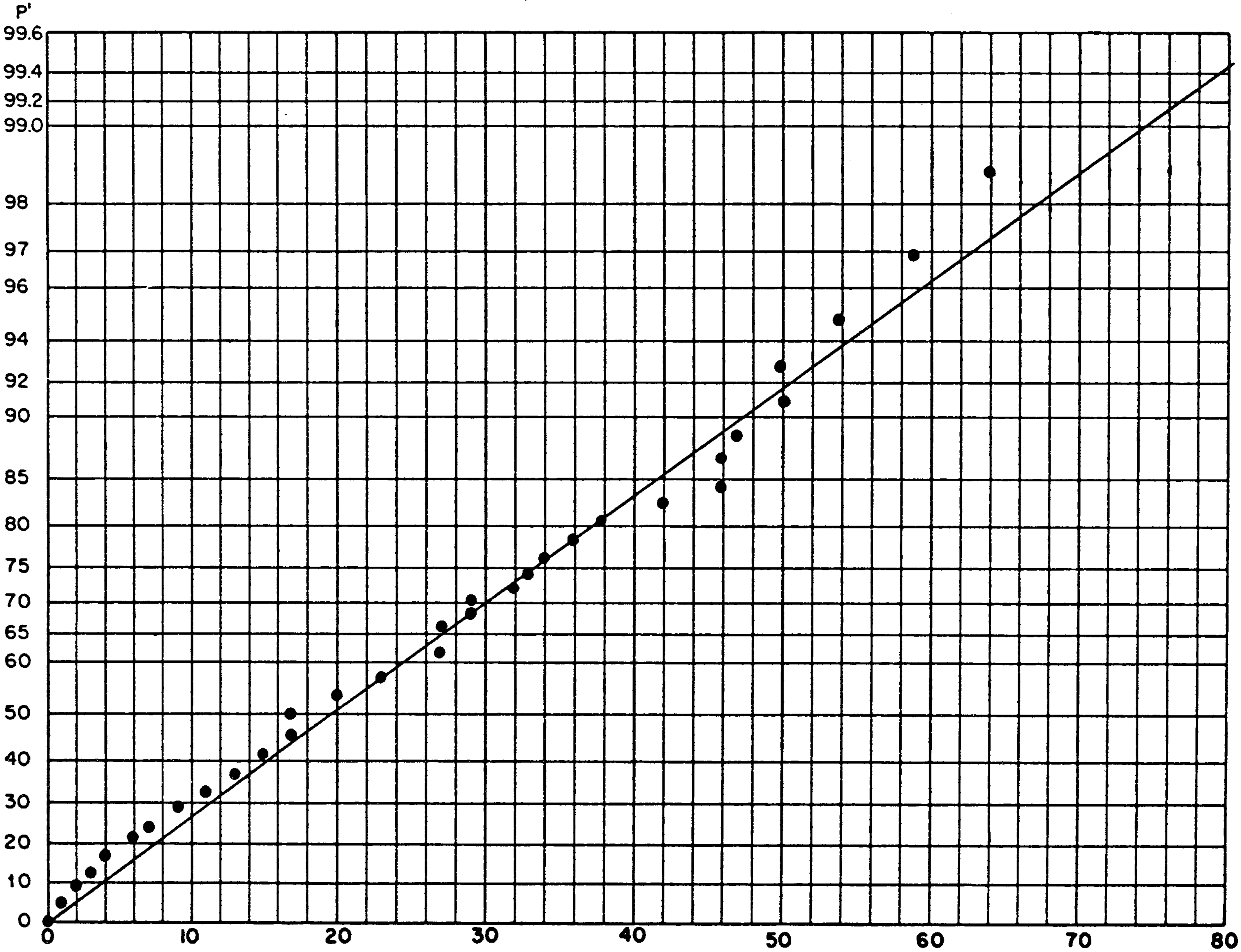


FIGURE 3. HALF-NORMAL PLOT, TEN LARGEST CONTRASTS OMITTED



$$\alpha_{ph} = a_{ph} / \sqrt{\sum_h a_{ph}^2}, \qquad h = 1, 2, \ldots, a;$$

$$\beta_{qi} = b_{qi} / \sqrt{\sum_i b_{qi}^2}, \qquad i = 1, 2, \ldots, b;$$

$$\gamma_{rj} = c_{rj} / \sqrt{\sum_j c_{rj}^2}, \qquad j = 1, 2, \ldots, c;$$

etc.

Each factor requires an additional "pass" through the machine. On the first pass, the machine computes the quantities (assuming three factors),

$$(A_p)_{ij} = \sum_h \alpha_{ph} \, y_{hij}.$$

On the second pass are computed the quantities.

$$(A_p B_q)_j = \sum_i \beta_{qi} \, (A_p)_{ij}$$

and on the third pass the quantities

$$(A_p B_q C_r) = \sum_j \gamma_{rj} \, (A_p B_q)_j.$$

At each pass the output includes both the (signed) contrasts developed and their squares. This program was one of the first developed for the IBM 1620 at Dugway Proving Ground and consequently was employed for a short period of time as a general-purpose analysis of variance. (It is, of course, much slower than other general-purpose programs available.)

6. EXPERIENCE. Some general comments on our experiences with half-normal plots for multi-level factorials may be in order. We shall be guided in this commentary largely by the approach of Daniel [1959].

a. Graph Sheets. We have generally used half-sheets of the Probability Scale x 90 Divisions paper available from Keuffel and Esser (Nos. 358-23 and 359-23).* Similar papers are available from several other sources. These papers are not particularly well-suited to the purpose. It would appear that special half-normal paper might be commercially feasible, but it is not, to our knowledge, currently available.

b. Birnbaum's test statistic. The test statistic developed by Allan Birnbaum [1959] has been used for our purposes. Birnbaum's work has been particularly oriented toward 2^p experiments and studies of the behavior of this statistic in multi-level factorials would be useful.

c. Defective values. Daniel indicates the utility of half-normal plotting in 2^p experiments for detecting defective values. For multi-level factorials the presence of defective values appears more difficult to diagnose, particularly with unrestricted sets of orthogonal contrasts. The isolation of the particular defective values is also more difficult.

d. Plot-splitting. The effect of plot-splitting upon the half-normal plots for multi-level experiments is similar to that described by Daniel. We have some reservations concerning indiscriminate searches for plot-splitting, however. It is generally accepted that in most experiments two-factor interactions tend to be smaller than main effects, three-factor interactions tend to be smaller than two-factor interactions, etc. (Here we are speaking of real effects and interactions, though perhaps of negligible magnitude.) Thus in actual experiments the slope of half-normal plots may be expected to increase with the relative number of high order interactions included. The plotted results of an experiment involving a number of small but real interactions may appear very similar to the results induced by plot-splitting, since split plot error contrasts invariably contain a relatively larger number of the higher order

---

*The graph sheets used in Figures 1, 2, and 3 were reproduced from a master kindly provided by Mr. Daniel. It is hoped that such sheets will soon be published.

contrasts. Our practice is generally to employ a split plot analysis only when knowledge of the experimental techniques indicates its propriety.

e. Convexity of plots. The detection of antilognormal distribution of error by downward convexity of half-normal plots appears difficult, as indicated by Daniel [1959, p. 336]. Most of our analysis work is, however, based on transformed data and we have seldom experienced this particular anomaly. In any event, the averaging effect of the contrasts would presumably minimize the effects of non-normality of error. On the other hand, we have noted that the removal of a moderate number of points representing apparently real effects often results in a downward convexity of the upper portion of the plot. We generally attribute this appearance to the inadvertent removal of one or more points representing error contrasts, for the result looks very much like the plot of a normal distribution with truncated upper tail.

7. REFERENCES.

Birnbaum, A. [1959], On the analysis of factorial experiments without replication, Technometrics I, No. 4, 343-57.

Davies, O. L., ed. [1954], The Design and Analysis of Industrial Experiments, Oliver and Boyd, London and Edinburgh.

Daniel, C. [1956], Fractional replication in industrial research, Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, Vol. V, 87-98, Univ. Of California Press, Berkeley and Los Angeles.

Daniel, C. [1959], Use of half-normal plots in interpreting factorial two-level experiments, Technometrics I, No. 4, 311-41.

Daniel, C. [1962], Personal Communication.

This page intentionally left blank.

PROPORTIONAL FREQUENCY DESIGNS

Sidney Addelman
Research Triangle Institute

CONDITION OF EQUAL FREQUENCIES. In 1945 Finney [5] introduced the procedure, known as fractional replication, which permitted the uncorrelated estimation of some of the effects and interactions when only a fraction of the full factorial arrangement was used. The standard method of constructing fractional replicate plans is to first choose an identity relationship and then deduce from this relationship the appropriate treatment combinations. By utilizing the assumption that the higher order interaction effects are negligible this standard procedure permits the estimation of the remaining effects. For the symmetrical factorial structure (all factors having the same number of levels) the standard procedure yields uncorrelated estimates due to the condition of equal frequencies of the factor levels. If the treatment combinations of the 2^5 factorial plan were inspected one would find that

(1) Each level of every factor occurs exactly eight times with every level of any other factor.

(2) Each combination of levels of any factor occurs exactly four times with every combination of levels of each pair of factors.

(3) Each combination of levels of any pair of factors occurs exactly two times with every combination of levels of any other pair of factors.

(4) Each level of any factor occurs exactly two times with every combination of levels of any three factors.

(5) Each level of any factor occurs exactly once with each combination of levels of any four factors.

(6) Each combination of levels of any pair of factors occurs exactly once with every combination of levels of any three factors.

Because the condition of equal frequencies is satisfied for all six of the above cases uncorrelated estimates of all effects can be obtained.

Now consider a 1/4 replicate of the 2^5 factorial structure defined by the identity relationship

$$I = ADE = BCD = ABCE$$

and consisting of the following treatment combinations:

A	B	C	D	E
0	0	0	0	0
0	0	1	1	1
0	1	0	1	1
0	1	1	0	0
1	0	0	0	1
1	0	1	1	0
1	1	0	1	0
1	1	1	0	1

One can verify that in this plan each level of any factor occurs exactly twice with every level of any other factor, and hence uncorrelated estimates of all main effects are obtainable, if all interactions are negligible. It can also be verified that each level of a factor does not occur the same number of times with every combination of levels of those pairs of factors with which it is aliased. Hence not all main effect estimates are uncorrelated with two-factor interaction estimates.

When one wishes to construct fractional replicate plans for symmetrical factorial arrangements one need only satisfy the appropriate equal frequency conditions to obtain uncorrelated estimates of the effects. However, in the construction of fractional replicate plans for asymmetrical factorial arrangements (all factors not having the same number of levels) the condition of equal frequencies requires more treatment combinations than are necessary to yield uncorrelated estimates.

CONDITION OF PROPORTIONAL FREQUENCIES. Although the equal frequency condition is sufficient to guarantee orthogonality of factors it is not a necessary condition. It was proved by Addelman and Kempthorne [3] that a necessary and sufficient condition that the main effect estimates of two factors be uncorrelated is that the levels of one factor occur with each of the levels of the other factor with proportional frequencies. Consider two factors, A and B, occurring at r and s levels respectively. Let

N = number of treatment combinations in the plan

$n_{i.}$ = number of times the i level of factor A occurs

$n_{\cdot j}$ = number of times the j level of factor B occurs

n_{ij} = number of treatment combinations in which the i level of factor A occurs with the j level of factor B.

The above necessary and sufficient condition for orthogonality can be displayed mathematically as

(1) $$n_{ij} = n_{i\cdot} n_{\cdot j}/N \; .$$

The condition of proportional frequencies can be generalized so that plans may be constructed with permit uncorrelated estimates of two-factor interactions as well as main effects. Consider three factors A, B and C. In order that the interaction AB can be uncorrelated with C, each combination of the levels of A and B must occur with the levels of C with proportional frequencies, that is

(2) $$n_{ijk} = n_{ij\cdot} n_{\cdot\cdot k}/N$$

Since it is desirable that A be uncorrelated with B

(3) $$n_{ij\cdot} = n_{i\cdot\cdot} n_{\cdot j\cdot}/N$$

and hence

(4) $$n_{ijk} = n_{i\cdot\cdot} n_{\cdot j\cdot} n_{\cdot\cdot k}/N^2 \; .$$

This condition which assures that AB is uncorrelated with C also implies that AC is uncorrelated with B, BC uncorrelated with A, and hence AB, AC and BC are pairwise uncorrelated. If a plan contains four or more factors condition (4) must be replaced by

(5) $$n_{ijkm} = n_{i\cdot\cdot\cdot} n_{\cdot j\cdot\cdot} n_{\cdot\cdot k\cdot} n_{\cdot\cdot\cdot m}/N^3$$

which is the necessary and sufficient condition that a plan permit uncorrelated estimation of all main effects and two-factor interaction effects.

<u>COLLAPSING OF LEVELS</u>. A factor at s_1 levels may be collapsed to a factor at $s_2 < s_1$ levels by making a many-one correspondence of the set of s_1 levels to the set of s_2 levels. If $s_1 = s_2^m$ then the s_1 levels can be collapsed to $(s_1-1)/(s_2-1)$ factors each having s_2 levels. Some illustrations of typical correspondence schemes are presented below.

Three-level factor		Two-level factor
0	$\longrightarrow$	0
1	$\longrightarrow$	1
2	$\longrightarrow$	0

Four-level factor		Three-level factor		Two-level factors
0	$\longrightarrow$	0	$\longrightarrow$	000
1	$\longrightarrow$	1	$\longrightarrow$	011
2	$\longrightarrow$	2	$\longrightarrow$	101
3	$\longrightarrow$	1	$\longrightarrow$	110

Five-level factor		Four-level factor		Three-level factor		Two-level factor
0	$\longrightarrow$	0	$\longrightarrow$	0	$\longrightarrow$	0
1	$\longrightarrow$	1	$\longrightarrow$	1	$\longrightarrow$	1
2	$\longrightarrow$	2	$\longrightarrow$	2	$\longrightarrow$	1
3	$\longrightarrow$	3	$\longrightarrow$	2	$\longrightarrow$	1
4	$\longrightarrow$	0	$\longrightarrow$	0	$\longrightarrow$	0

An orthogonal main-effect plan for the $2^2 \times 3^2$ experiment which permits uncorrelated estimates of all main effects with only nine treatment combinations is now constructed to illustrate the technique of collapsing levels. First construct an orthogonal main-effect plan for four factors, each having three levels with nine treatment combinations, namely

0	0	0	0
0	1	1	2
0	2	2	1
1	0	1	1
1	1	2	0
1	2	0	2
2	0	2	2
2	1	0	1
2	2	1	0

If each of the first two factors are collapsed to two-level factors, the resulting treatment combinations constitute an orthogonal main-effect plan for the $2^2 \times 3^2$ experiment and are displayed below.

0	0	0	0
0	1	1	2
0	0	2	1
1	0	1	1
1	1	2	0
1	0	0	2
0	0	2	2
0	1	0	1
0	0	1	0

The smallest plan which yields uncorrelated estimates of the main effects of the $2^2 \times 3^2$ experiment and which also satisfies the equal frequency condition would require 36 treatment combinations.

It should be mentioned that the proportional frequency condition will be satisfied no matter what type of correspondence scheme is used to perform the collapsing procedure. However, the efficiency of the estimates depends upon the particular correspondence scheme chosen.

If the $(s_i - 1)$ degrees of freedom for each of the t_i factors at s_i levels are represented by $(s_i - 1)$ orthogonal contrasts among the s_i levels, the estimates obtained by these contrasts will be uncorrelated with the estimates obtained with the contrasts for any other factor, because the correspondence scheme automatically guarantees proportional frequencies of the levels of each factor.

REPLACING FACTORS. The collapsing procedure given above can be reversed so that a factor at s^m levels can replace $(s^m - 1)/(s - 1)$ factors, each at s levels. The replacement procedure can be illustrated by the construction of an orthogonal main-effect plan for the 3×2^4 experiment with eight trials. First construct an orthogonal main-effect plan for the 2^7 experiment with eight trials. The seven two-level factors can be represented by X_1, X_2, X_1X_2, X_3, X_1X_3, X_2X_3 and $X_1X_2X_3$.

The treatment combinations for this plan are

0	0	0	0	0	0	0
0	0	0	1	1	1	1
0	1	1	0	0	1	1
0	1	1	1	1	0	0
1	0	1	0	1	0	1
1	0	1	1	0	1	0
1	1	0	0	1	1	0
1	1	0	1	0	0	1

It is known that there exists an orthogonal main-effect plan for the 2^3 experiment with four trials. The treatment combinations for this plan are 000, 011, 101, and 110. Thus, by choosing three factors of the 2^7 plan whose X representations are such that the generalized interaction of any two of the three factors is the third factor, three two-level factors can be replaced by a four-level factor, according to the following correspondence scheme:

Two-level factor				Four-level factor
0	0	0	$\longrightarrow$	0
0	1	1	$\longrightarrow$	1
1	0	1	$\longrightarrow$	2
1	1	0	$\longrightarrow$	3

Since the X representations of the first three factors of the above plan are X_1, X_2 and X_1X_2, these three factors can be replaced by a four-level factor and the orthogonal main-effect plan for the 4×2^4 experiment in eight trials is given by the following treatment combinations:

0	0	0	0	0
0	1	1	1	1
1	0	0	1	1
1	1	1	0	0
2	0	1	0	1
2	1	0	1	0
3	0	1	1	0
3	1	0	0	1

The plan for the 3×2^4 experiment is then obtained by collapsing the four-level factor to a three-level factor by the correspondence

Four-level factor		Three-level factor
0	$\longrightarrow$	0
1	$\longrightarrow$	1
2	$\longrightarrow$	2
3	$\longrightarrow$	1

The smallest plan which yields uncorrelated estimates of main effects in the 3×2^4 experiment and which also satisfies the equal frequency condition would require 24 treatment combinations.

The procedure for constructing plans which permit uncorrelated estimates of all main effects and some or all of the two-factor interaction effects for asymmetrical factorial arrangements consists of first constructing the corresponding plan for a symmetrical factorial arrangement and then utilizing the collapsing or replacing techniques to obtain the desired plan. Whereas a plan permitting uncorrelated estimates of all main effects and all two-factor interactions among the two-level factors in the $2^3 \times 3^4$ experiment would require 72 treatment combinations to satisfy the condition of equal frequencies it would only require 27 treatment combinations to satisfy the proportional frequency condition.

BLOCKING. Even though the proportional frequency designs are highly fractionated they may still require more trials than can be carried out under uniform conditions. Thus, it would be desirable to divide the experimental data into smaller blocks in such a manner that the main effects may still be estimated without correlation. In order to perform an experiment in blocks one may utilize one or more of the factors of an orthogonal main-effect plan for the $4 \times 3^2 \times 2^6$ experiment with sixteen trials. The following plans may be derived from this one by using various factors as blocking factors:

(i) $4 \times 3^2 \times 2^5$ in 2 blocks of 8 treatment combiantions,

(ii) $4 \times 3^2 \times 2^3$ in 4 blocks of 4 treatment combinations,

(iii) $3^2 \times 2^6$ in 4 blocks of 4 treatment combinations,

(iv) $4 \times 3 \times 2^6$ in 4 blocks of 4 treatment combinations.

ORTHOGONAL POLYNOMIALS. The orthogonal contrasts which define effects and interactions in an equal frequency design can be readily determined from a table of orthogonal polynomials. The advantage of using orthogonal contrasts to define effects and interactions arises from the fact that orthogonal polynomials are so constructed that any term of the polynomial is independent of any other term. This property of independence permits one to compute each regression coefficient independently of the others and also facilitates testing the significance of each coefficient.

Tables of orthogonal polynomials for the case of equally spaced levels are readily available, e.g. Fisher and Yates [6], Anderson and Houseman [4]. It would be an impossible task to compile a general table of orthogonal polynomials for unequally spaced levels. However a simple procedure for computing these orthogonal polynomials is available and will be presented below. If equally spaced levels do not each occur in a plan an equal number of times the published tables of orthogonal polynomials are not appropriate. The orthogonal polynomials for equally spaced levels which do not occur in a plan with equal frequency can be computed by the following method for unequally spaced levels.

For any set of orthogonal polynomials the linear contrast is of the form $\Sigma(\alpha + \beta x)y_x$, where α and β are constants, x is the level at which the factor occurs, y_x is the response to the treatment combination with the factor at the x level and the summation is over every value of x which is presented. The quadratic and cubic contrasts are of the form $\Sigma(\alpha + \beta x + \gamma x^2)y_x$ and $\Sigma(\alpha + \beta x + \gamma x^2 + \delta x^3)y_x$, respectively. The extension to higher order contrasts is obvious. Two contrasts are orthogonal if the coefficients of each contrast sum to zero and the sum of products of the corresponding coefficients of the two contrasts is zero.

Table 1

Coefficients of Orthogonal Contrasts

Level of x	Linear	Quadratic	Cubic
0	α	α	α
1	$\alpha + \beta$	$\alpha + \beta + \gamma$	$+\beta + \gamma + \delta$
2	$\alpha + 2\beta$	$\alpha + 2\beta + 4\gamma$	$+ 2\beta + 4\gamma + 8\delta$
4	$\alpha + 4\beta$	$\alpha + 4\beta + 16\gamma$	$+ 4\beta + 16\gamma + 64\delta$

We will illustrate the procedure for obtaining orthogonal polynomials for unequally spaced levels with an example.

Consider an independent variable x with levels 0, 1, 2 and 4. The coefficients of the linear, quadratic and cubic contrasts for this example are displayed in Table 1. The coefficients of the linear contrast must sum to zero. Thus,

$$4\alpha + 7\beta = 0.$$

Setting $\beta = 1$ we find that $\alpha = -7/4$. In order that the coefficients of the orthogonal contrasts be integers reduced to lowest terms we multiply these coefficients by 4 to obtain $\beta = 4$ and $\alpha = -7$. Substituting $\alpha = -7$ and $\beta = 4$ in the linear contrast given in Table 1, gives the linear coefficients.

Level of x	Coefficient of linear contrast
0	-7
1	-3
2	1
4	9

The coefficients of the quadratic contrast must sum to zero. Hence,

$$4\alpha + 7\beta + 21\gamma = 0$$

The sum of products of the corresponding coefficients of the linear and quadratic contrasts must also equal zero. Thus,

$$35\beta + 145\gamma = 0$$

Solving these two equations to obtain integral values for α, β and γ we obtain $\alpha = 14$, $\beta = -29$ and $\gamma = 7$.

If we substitute these values in the quadratic contrast and reduce the resulting coefficients to lowest terms the coefficients of the quadratic contrast is given by

Level of x	Coefficient of Quadratic contrast
0	7
1	-4
2	-8
4	5

Similarly the sum of the coefficients of the cubic contrast and the sum of products of the corresponding coefficients of the linear and cubic contrasts must each equal zero. Hence,

$$\begin{aligned} 4\alpha + 7\beta + 21\gamma + 73\delta &= 0 \\ 35\beta + 145\gamma + 581\delta &= 0 \\ 44\gamma + 252\delta &= 0 \end{aligned}$$

Solving these equations to obtain integral values for α, β, γ and we obtain $\alpha = -36$, $\beta = 392$, $\gamma = -315$ and $\delta = 55$. If we substitute these values in the form of the coefficients of the cubic contrast given in Table 1 and reduce the resulting coefficients to lowest terms, the coefficients of the cubic contrast are given by

Level of x	Coefficients of Cubic contrast
0	-3
1	8
2	-6
4	1

The orthogonal polynomials are presented in the following table.

Table 2
Orthogonal Polynomials

Level of x	Linear	Quadratic	Cubic
0	-7	7	-3
1	-3	-4	8
2	1	-8	-6
4	9	5	1

The symbol β represents one unit of the linear effect of a factor when set equal to unity. In order to obtain integral coefficients β was set equal to 4 and hence $(1/4)\beta$ represents one unit of the linear effect. Consequently the linear contrast with coefficients given in Table 2 represents the estimate of 1/4 the linear effect of the factor. It is easily verified that the coefficients of the quadratic contrast are given by

$$7 - \frac{29}{2}\,x + \frac{7}{2}x^2$$

where x = 0, 1, 2 and 4 respectively. Thus the symbol $\frac{2}{7}\gamma$ represents one unit of the quadratic effect, and the linear contrast with coefficients given in Table 2 represents the estimate of $\frac{2}{7}$ the quadratic effect of the factor.

Similarly it may be demonstrated that the cubic contrast with coefficients given in Table 2 represents the estimate of 12/55 the cubic effect of the factor.

This constant which is multiplying each effect will be denoted by $1/\lambda$ and in the tables of orthogonal polynomials presented by Addelman and Kempthorne [3] the value of λ and the sum of squares of the coefficients were both given. Any contrast defined by the coefficients given in the tables of orthogonal polynomials represents $1/\lambda$ times the appropriate effect of the factor.

EFFICIENCIES. Although any many-one correspondence of the set of s_1 levels to the set of s_i levels will yield proportional frequencies of the levels, there arises the problem of which correspondence is "best" in some sense. The problem may be solved by determining the efficiencies of the main-effect estimates obtained using proportional frequencies relative to the estimates which would result from using equal frequencies of the levels of each factor.

As an illustration we will calculate the relative efficiency of a three-level factor in a main-effect plan with twenty-five trials.

Assume the correspondence scheme used to collapse a five-level factor to three levels is as follows:

Five-level factor		Three-level factor
0	$\longrightarrow$	0
1	$\longrightarrow$	1
2	$\longrightarrow$	2
3	$\longrightarrow$	2
4	$\longrightarrow$	0

The levels 0, 1, and 2 occur in the ratio's 2: 1: 2. Thus for this factor the 0 level occurs in ten treatment combinations, the 1 level occurs in five treatment combiantions and the 2 level occurs in ten treatment combinations.

The variance of the linear effect estimate of this factor is equal to

$\sigma^2/20$ and hence the information on a unit basis is equal to $20/25\sigma^2 = 4/5\sigma^2$. The variance of the linear effect estimate of a three-level factor in 3^n trials is equal to $\sigma^2/2.3^{n-1}$ and the information on a unit basis is $2.3/^{n-1}/3^n\sigma^2 = 2/3\sigma^2$. Hence the relative efficiency of the linear effect estimate is equal to $(4/5)x(3/2) = 6/5$.

The variance of the quadratic effect estimate for the three-level factor in twenty-five trials is equal to $\sigma^2/4$ and the information is then equal to $4/25\sigma^2$. The variance of the quadratic effect estimate with 3^n trials is equal to $\sigma^2/2.3^{n-2}$ and hence the information on a unit basis is equal to $2/9\sigma^2$. The relative efficiency of the quadratic effect estimate is therefore equal to $(4/25)x(9/2) = 18/25$.

Table 3

Relative Efficiencies of Proportional Frequency Estimates

Level	0 : 1 Proportional frequency	Efficiency
	1 : 2	8/9
	2 : 3	24/25
	1 : 4	16/25
	3 : 4	48/49
	2 : 5	40/49
	1 : 6	24/49

Level / Contrast	0 : 1 : 2 Proportional frequency	Efficiency
Linear	1 : 2 : 1	3/4
Quadratic	1 : 2 : 1	9/8
Linear	2 : 1 : 2	6/5
Quadratic	2 : 1 : 2	18/25
Linear	1 : 3 : 1	3/5
Quadratic	1 : 3 : 1	27/25
Linear	2 : 3 : 2	6/7
Quadratic	2 : 3 : 2	54/49
Linear	3 : 1 : 3	9/7
Quadratic	3 : 1 : 3	27/49
Linear	1 : 5 : 1	3/7
Quadratic	1 : 5 : 1	45/49

The relative efficiencies of the estimated effects are presented for various proportional frequencies in Table 3. One would choose the proportional frequencies which give the greatest efficiency of estimates. Thus for example, if an experiment in twenty-five trials involved two-level factors the two levels should occur in the ratio 2 : 3 rather than in the ratio 1 : 4 because the efficiency of the 2 : 3 ratio is 24/25 whereas the efficiency of the 1 : 4 ratio is only 16/25.

Bibliography

[1] Addelman, S. Orthogonal main-effect plans for asymmetrical factorial experiments. Technometrics, 4:21-46, 1962.

[2] Addelman S. Symmetrical and asymmetrical fractional factorial plans. Technometrics, 4:47-58, 1962.

[3] Addelman, S. and O. Kempthorne. Orthogonal main-effect plans. ARL 79, United States Air Force, November 1961.

[4] Anderson, R. L. and E. E. Houseman. Tables of orthogonal polynomial values extended to N=104. Agricultural Experiment Station Research Bulletin 297. Iowa State College, Ames, Iowa, 1942.

[5] Finney, D. J. The fractional replication of factorial arrangements. Annals of Eugenics, 12:291-301. 1945.

[6] Fisher, R. A. and F. Yates. Statistical tables for biological, agricultural and medical research. Oliver and Boyd, Edinburgh, 1938.

OPTIMAL DESIGN OF EXPERIMENTS

Herman Chernoff [1/]
Stanford University

1. INTRODUCTION. I would like to discuss some aspects of the theory of optimal design of experiments with particular emphasis on its relevance to the practice of statistics. There are two major branches of classical statistics, Estimation and Testing of Hypotheses, for which the theory of optimal design yields different results. Because of the time limitation, I shall confine my attention to certain results and examples in the theory of estimation.

2. SOME EXAMPLES. To illustrate the theory let us consider three examples. The first example is a well known one with a trivial solution. That is the one of estimating the slope of a regression (straight line). More specifically we have

Example 1.

The experimenter may choose any number y between -1 and +1. This number y designates an elementary experiment which corresponds to observing

$$Z = \alpha + \beta y + u$$

where u is normally distributed with mean 0 and variance 1 and α and β are unknown parameters. The experimenter is permitted to select a design consisting of n values $y_1, y_2, \ldots, y_n$, with possible repetitions. The design corresponds to performing the n designated experiments independently. It is desired to select a design which will yield the best possible estimate of the slope β.

It is well known and it is intuitively obvious that the best design consists of selecting $y = -1$ and $y = +1$ each half the time (providing n is even).

1/ This work was supported in part by Office of Naval Research Contract Nonr-225(52) at Stanford University. Reproduction in whole or in part is permitted for any purpose of the United States Government.

Another example which is of some current interest, having been discussed in yesterday's paper by Mr. Langlie [5] on a problem in reliability, and which is also relevant to the problem of Probit Analysis, may be expressed as follows:

Example 2.

A device, which may be used only once, can operate successfully under a stress s with probability

$$p = \int_{\frac{s-\mu}{\sigma}}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} \, dt \, .$$

In other words one may say that the strength of the device, as measured by the maximum stress under which it will operate successfully, is normally distributed with unknown mean μ and variance σ^2. It is desired to select a design consisting of the choice of stress levels $s_1, s_2, \ldots, s_n$ which will yield an optimal estimate of $\mu - k\sigma$. The elementary experiment, designated s, consists of course of observing the success or failure of the device when used under stress s.

Finally a third problem which was discussed in detail in a recent paper of mine [2] deals with accelerated life testing. Here we wish to estimate the mean life time of a device when used under an environment of ordinary stress conditions. If, this mean lifetime is great and it is desired to have the estimate soon, then it is necessary to accelerate. The device is subjected to a much larger than ordinary stress. The results of such accelerated life testing can be relevant only if one assumes some form of relationship connecting the mean lifetime under various stresses. As an approximation we shall assume a quadratic relationship for some limited range. In addition since time is of the essence we shall assume that the cost of observing a device under stress s is proportional to the mean lifetime under that stress. Let us be more specific.

Example 3.

A device under stress environment s has lifetime T with an exponential distribution with failure rate (reciprocal of mean) given by

$$\varphi = \theta_1 s + \theta_2 s^2 \qquad \text{for } 0 \leq s \leq s^*$$

where θ_1 and θ_2 are unknown parameters. It is desired to estimate the failure rate under the ordinary stress s_o. This is

$$\varphi_o = \theta_1 s_o + \theta_2 s_o^2 .$$

An elementary experiment designated by s consists of observing the lifetime T of a device subjected to the environment s. The cost of the experiment s is

$$C(s) = c(\theta_1 s + \theta_2 s^2)^{-1} .$$

It is desired to select a design consisting of experiments $s_1, s_2, \ldots;$ $0 \leq s_i \leq s^*$, so as to obtain an optimal estimate of φ_o for a specified total cost.

Each of these examples has certain elements in common. Each may be regarded as a special case of the following general formulation. There is a set $\mathcal{E}$ of available elementary experiments e. In each case the distribution of the data of an experiment depends on the experiment and on k unknown parameters represented by $\theta = (\theta_1, \theta_2, \ldots, \theta_k)$. We wish to estimate some function $g(\theta_1, \theta_2, \ldots \theta_k)$ of the parameters. A design consists of the independent performance of experiments $e_1, e_2, \ldots$ with possible repetitions. It is desired to find a design which yields the best possible estimate of $g(\theta_1, \theta_2, \ldots, \theta_k)$ for a specified total cost or for a specified number of observations.

3. THE LINEAR REGRESSION MODEL. In 1952, Elfving [4] derived an elegant geometric solution to the optimal design problem for a special but important case of the above general formulation. As we shall see this result is applicable to a large variety of problems. Let $\mathcal{E}$ be a set of experiments e denoted by (y_1, y_2). The experiment e consists of observing

$$Z = \theta_1 y_1 + \theta_2 y_2 + u$$

where u is normally distributed with mean 0 and variance 1. It is desired to obtain an optimal estimate of $a_1 \theta_1 + a_2 \theta_2$ using a design consisting of n observations. The first example of estimating the slope of a straight line is a special case of Elfving's linear regression model where $\mathcal{E}$ is the set of points $(1, y)$ with $-1 \leq y \leq 1$, and $(a_1, a_2) = (0, 1)$.

Elfving's solution consists of constructing a set S which is the smallest convex set containing the points (y_1, y_2) of $\mathcal{E}$ and their negatives $(-y_1, -y_2)$. Then extend the vector from $(0, 0)$ to (a_1, a_2) until it penetrates the set S. The point of penetration (w_1, w_2) represents the optimal design. If this point is one of the original points (y_1, y_2) or $(-y_1, -y_2)$ the optimal design consists of repeating (y_1, y_2) n times. Otherwise the point of penetration is on a line segment connecting points corresponding to two of the original experiments (or their negatives). Then the optimal design consists of repeating these two experiments in proportions given by the distances from (w_1, w_2) to the two points. The greater proportion corresponds to the experiment closer to (w_1, w_2). Finally the variance of the least squares estimate based on this design is

$$\sigma_\phi^2 = \left[n(w_1^2 + w_2^2)\right]^{-1}(a_1^2 + a_2^2) = a_1^2/nw_1^2 = a_2^2/nw_2^2$$

This solution can be illustrated with example 1. Here S is the square whose corners are $(1, 1)$ and $(-1, -1)$ corresponding to $y = 1$ and $(1, -1)$ and $(-1, 1)$ corresponding to $y = -1$. The line from $(0,0)$ through $(a_1, a_2) = (0, 1)$ penetrates S at $(0, 1)$ which is halfway between $(1, 1)$ and $(-1, 1)$. Thus the optimal design consists of repeating the experiments corresponding to $y = 1$ and $y = -1$ each half the time (as was well known). Furthermore the variance of the estimate of β should be $1/n$.

Elfving's result applies in the obvious fashion to experiments involving k parameters. Here we need repeat at most k of the available experiments in certain proportions to obtain the optimal estimate.

4. RESULTS FOR THE MORE GENERAL PROBLEM. As mentioned in the preceeding section the problem treated by Elfving is a special case of the more general one formulated in section 2. For this more general problem, related results have been obtained [1]. These results concern designs which are asymptotically locally optimal. We shall defer the interpretation of these adjectives until the discussion of Example 2 in section 5.

It was shown that asymptotically locally optimal designs depend on the form of the matrix $J(e)$ which is defined as Fisher's information matrix divided by the cost of the experiment e. In other words if experiment e has cost $C(e)$ and yields data X with probability distribution $f(x, \theta, e)$, Fisher's information matrix is

$$I(e) = \left\| E\left\{ \frac{\partial \log f(X, \theta, e)}{\partial \theta_i} \frac{\partial \log f(X, \theta, e)}{\partial \theta_j} \right\} \right\|$$

and the information per unit cost is

$$J(e) = I(e)/C(e).$$

Clearly if the cost of experimentation is constant one need concern oneself only with $I(e)$. The relevance of Fisher's Information derives from its well known additive properties and the fact that the maximum-likelihood estimate $\hat{\theta}_n$ based on the outcome of n independent repetitions of e, has an approximately normal distribution with mean θ and covariance matrix $[nI(e)]^{-1}$ for large n.

When it is desired to estimate one function of the k parameters, there are asymptotically locally optimal designs which involve at most k of the experiments of ξ in certain proportions. This result which corresponds to one of Elfving's results, together with the use of Fisher's Information, permits one to reduce the calculation of optimal designs to the maximization of a function of a fixed number of variables.

In the linear regression problem of Elfving, the information matrix for $e = (y_1, y_2)$ is

$$I = \| y_i y_j \| = J.$$

Since asymptotically optimal designs are determined by the information per unit cost it follows that for any problem where $J(e)$ can be put in the above form, the solution is the same as Elfving's with a_i replaced by $\frac{\partial g}{\partial \theta_i}$.

The illustration of the next section will help clarify the meaning of these results. In the meantime it may be remarked that if for each experiment the distribution of the outcome depends on only one function of the parameters, $J(e)$ can be put in the above form and Elfving's results are applicable. In particular they are applicable to both examples 2 and 3.

5. ILLUSTRATION. We shall find it informative to illustrate the method with example 2. Here the outcome of the experiment s is success or failure where the probability of success is

$$p(s, \mu, \sigma) = \int_{\frac{s-\mu}{\sigma}}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} \, dt = 1 - \Phi\left(\frac{s-\mu}{\sigma}\right)$$

where Φ is the normal cdf. In other words the role of the density $f(X, \theta, e)$ is played by

$$f = p^X(1 - p)^{1-X}$$

where $X = 1$ for success and 0 for failure.

$$\log f = X \log p + (1-X) \log(1 - p)$$

$$\frac{\partial \log f}{\partial \mu} = \frac{X - p}{p(1 - p)} \frac{\partial p}{\partial \mu}$$

$$\frac{\partial \log f}{\partial \sigma} = \frac{X - p}{p(1 - p)} \frac{\partial p}{\partial \sigma} .$$

Since $E\{(X-p)^2\} = p(1-p)$,

$$J(s) = I(s) = \left[p(1 - p)\right]^{-1} \left\| \begin{matrix} \left(\frac{\partial p}{\partial \mu}\right)^2 & \frac{\partial p}{\partial \mu}\frac{\partial p}{\partial \sigma} \\ \frac{\partial p}{\partial \mu}\frac{\partial p}{\partial \sigma} & \left(\frac{\partial p}{\partial \sigma}\right)^2 \end{matrix} \right\|$$

$$J(s) = \left\| y_i y_j \right\|$$

where

$$y_1(s) = \left[p(1-p)\right]^{-1/2} \frac{\partial p}{\partial \mu} = \left[2 \pi p(1-p)\right]^{-1/2} \sigma^{-1} \exp\left[-(s-\mu)^2/2\sigma^2\right]$$

and

$$y_2(s) = \left[p(1-p)\right]^{-1/2} \frac{\partial p}{\partial \mu} = \left[2 \pi p(1-p)^{-1/2} (s-\mu)\, \sigma^{-2} \exp\left[-(s-\mu)^2/2\sigma^2\right] .$$

Next we plot the set of points $\left[y_1(s),\ y_2(s)\right]$ in Figure 1. We add the negatives of these points and construct S the smallest convex set containing them. We note that for $s = \mu + t\sigma$, $y_2(s)/y_1(s) = t$. We also note the curve of $\left[y_1(s),\ y_2(s)\right]$ reaches its maximum and minimum at $s = \mu \pm k_o\sigma$ where $k_o = 1.57$. Finally, since we wish to estimate $\mu - k\sigma$, we draw the vector from (0, 0) through (1, -k), i. e. the line through the origin with slope -k, and note where it penetrates the convex set S.

Clearly there are two cases.

Case 1. $|k| \leq k_o$. Here the vector penetrates S at one of the original $\left[y_1(s), y_2(s)\right]$ points. In fact this point corresponds to $s = \mu - k\sigma$ and hence the optimal design consists of using $s = \mu - k\sigma$ for all observations.

Case 2. $|k| > k_o$. Here the vector penetrates S at the straight line section of the boundary. The optimal design consists of applying the stress levels $\mu - k_o\sigma$ and $\mu + k_o\sigma$ in proportions $k+k_o$ to $k-k_o$. In cases 1 and 2 the formal application of the formula for the variance of the maximum likelihood estimate of $\mu - k\sigma$ based on the optimal design is given by

$$2\pi\sigma^2 \Phi(k)\left[1-\Phi(k)\right] \; e^{k^2} n^{-1}$$

in case 1, and

$$2\pi\sigma^2 \Phi(k_o)\left[1-\Phi(k_o)\right] \; e^{k_o^2} k_o^{-2} k^2 n^{-1} = 1.64\,\sigma^2 k^2 n^{-1}$$

in case 2.

6. THE RELEVANCE OF OPTIMAL DESIGN. Now we shall find the illustrative example helpful in interpreting the results of the theory of optimal design of experiments and in understanding its relevance in practical applications. For simplicity let us confine our attention to case 2 at first.

First we note one very peculiar aspect of the optimal design. Since it involves using stress levels $\mu - k_o\sigma$ and $\mu + k_o\sigma$, to apply it one must know μ and σ. But if one knew μ and σ, there would be no need to experiment. While this seems to be ridiculous, a glance at figure 1 indicates that if one used an approximation to $\mu \pm k_o\sigma$, one would have a rather good approximation to the optimal design. Thus there is surprisingly little loss of efficiency when one is not certain about μ and σ. It is this property that the word local is used to describe. In other words our design would be efficient if we knew the parameters and

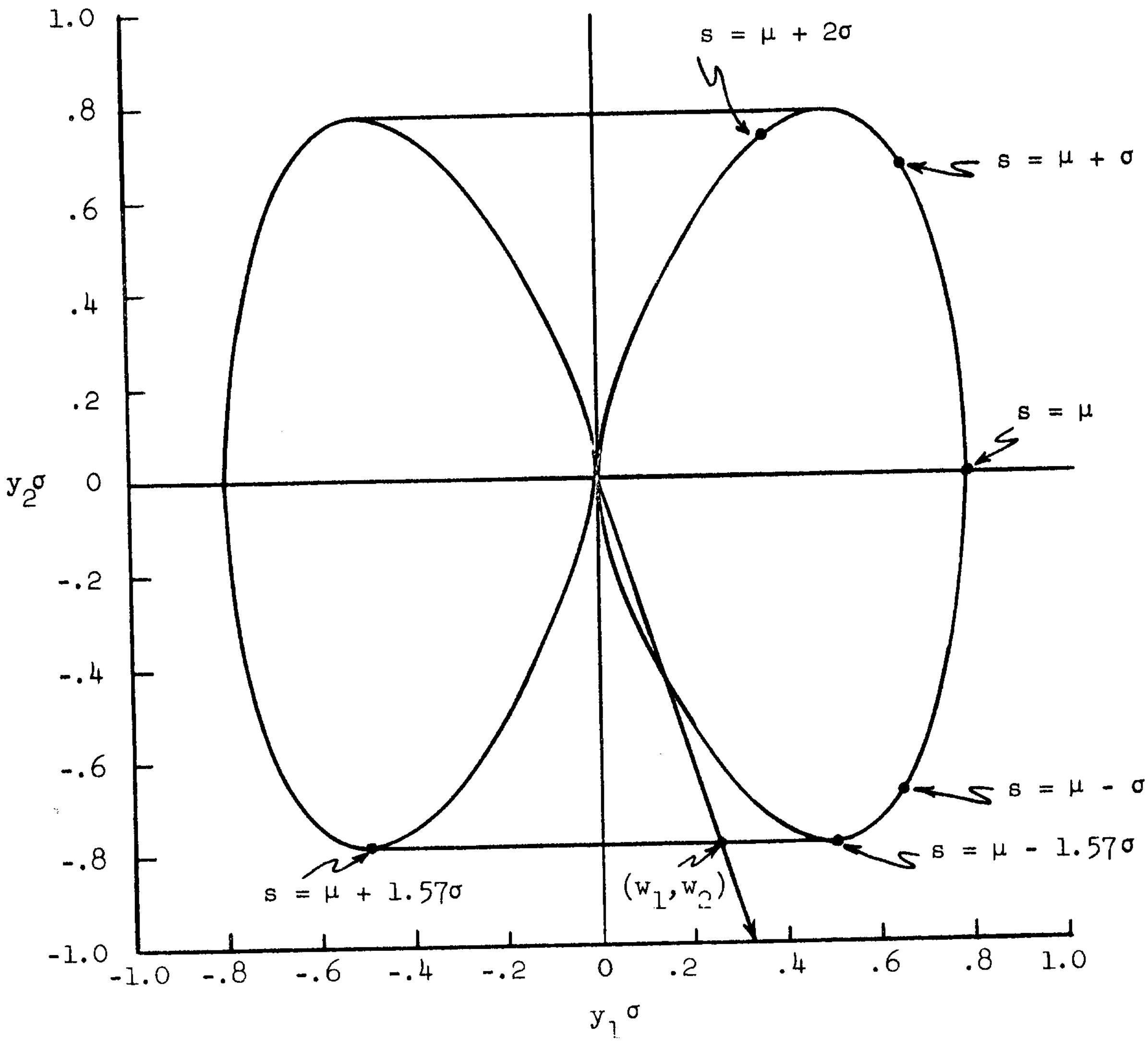

1.0
.8
.6
.4
.2
$y_2\sigma$ 0
-.2
-.4
-.6
-.8
-1.0
-1.0 -.8 -.6 -.4 -.2 0 .2 .4 .6 .8 1.0
$y_1\sigma$
$s = \mu + 2\sigma$
$s = \mu + \sigma$
$s = \mu$
$s = \mu - \sigma$
$s = \mu - 1.57\sigma$
$s = \mu + 1.57\sigma$
(w_1, w_2)

Figure 1

This page intentionally left blank.

is approximately efficient if we use an approximation to the unknown parameters.

This raises the issue of the adjective asymptotic. If one had a large sample available, one could use some of the initial observations to derive an initial estimate of and on which to base an approximation to the optimal design. Furthermore the qualification asymptotic derives from a couple of other aspects. First, the properties relating the variance of the approximate distribution of the maximum likelihood to the information matrix and giving the efficiency of this estimate is based on asymptotic theory assuming large sample size. A second and relatively minor point, is illustrated by example 1 if an odd number of observations are available. The optimal design calls for putting half the observations at +1 and half at -1. This is impossible in a trivial way when n is odd. On the other hand the effect of this impossibility is negligible when n is large.

Having seen how we must qualify the tern optimal by the adjectives local and asymptotic, we can now consider a more fundamental issue. Briefly, our optimal design is simply impractical. Only in the rather unrealistic context where I had absolute faith in the model would I consider this as a solution. In fact, any reasonable statistician would insist on using several other stress levels at least to check on the model.

Another unreasonable aspect of our optimal design arises from its derivation based on the single minded purpose of obtaining a good estimate of one function $g(\theta_1, \theta_2, \ldots, \theta_k)$ of the parameters. In many practical problems, experimentation is used to serve several purposes simultaneously.

One may reasonably inquire about what function does the theory of optimal design serve, if (1) the optimality must be qualified as locally asymptotically optimal and (2) the designs it yields are unreasonable. Basically the functions are the following. First, the theory provides a yardstick for comparison purposes. If the designs proposed yesterday by Mr. Langlie, or the Up and Down Method [3, p. 319], or some other practical design turns out to be relatively efficient compared to our solution (as measured by asymptotic variance) then clearly there is no point in attempting to improve on this aspect of these methods. If, on the other hand, one of these methods were to have a low efficiency, then one is forced to delve deeper to see what, if anything, can be done to improve the design.

Second, theory not only presents an optimal design but indicates rather clearly how this design can be modified with relatively low loss of efficiency. The theory serves to direct the attention of the practical statistician toward designs which combine relatively high efficiency with practical utility when robustness and multi-purpose considerations are taken into account.

7. MISCELLANEOUS COMMENTS. I would like to conclude this paper with a few assorted comments. First, the proposed solution to example 2 in case 1 when $|k| \leq k_0$ consists of repeating one experiment n times. Not only is this solution impractical, but from a theoretical point of view it represents a degenerate situation. When a single level s is used, one can use the data to estimate only

$$p(s, \mu, \sigma) = \int_{\frac{s-\mu}{\sigma}}^{\infty} (2\pi)^{-1/2} e^{-t^2/2}\, dt$$

or functions of $p(s, \mu, \sigma)$. Then one can check whether $\frac{s-\mu}{\sigma}$ is in fact close to k (as it should be if the design were optimal). But not knowing σ, one can not estimate $\mu - k\sigma$. Thus the formula for the asymptotic variance presented at the end of section 5 is meaningful only as an approximation to the case where several levels of stress close to the optimal one were used. Alternatively one could regard $p(1-p)n^{-1}$ as the asymptotic variance of the estimate of p.

For a large sample sequential procedure, it seems clear that our theory is applicable. If one were to reestimate the parameters after each observation, and use these estimates to derive approximations to the optimal design, the resulting procedure should be asymptotically optimal in the sequential version and the adjective local need not be applied.

What is more interesting, perhaps, is the study of the "not so large" sample sequential case. Here even the following seemingly simple problem proposed by Harold Gumbel does not have a simple solution. Suppose that experiment e_i yields observation X_i which is normally distributed with unknown mean μ and unknown variance σ_i^2, $i=1, 2$, and it is desired to estimate μ. In other words, two measuring instruments of unknown accuracy are available. How should one select between the two experiments _sequentially_ so as to obtain a good estimate efficiently when the sample size is not necessarily very large?

BIBLIOGRAPHY

(1) Chernoff, H. (1953), Locally optimal designs for estimating parameters, _Ann. Math. Statist._ **24** 586-620.

(2) Chernoff, H. (1962), Optimal accelerated life designs for estimation, _Technometrics_ **4**, 381-408.

(3) Dixon, W. J., and Massey, F. J. (1957), _Introduction to Statistical Analysis_, (2nd ed.) McGraw Hill, New York.

(4) Elfving, G. (1952), Optimal allocation in linear regression theory, _Ann. Math. Statist._ **23**, 255-262.

(5) Langlie, H. J. (1962), A Reliability Test Method for "One-Shot" Items; presented at the Eighth Conference on The Design of Experiments in Army Research, Development and Testing.

This page intentionally left blank.

VIBRATION EXPERIMENTS

F. Pradko
Dynamic Simulations Laboratories
U. S. Army Ordnance Tank-Automotive Command

The design of military vehicles is a rather complicated mixture of many technical activities. In each new development of a tank, truck or jeep, a substantial amount of engineering ideas are required to be blended together to bring forward vehicles possessing features and merit of advanced capability.

While each vehicle development is a separate and distinct program, there are development goals and problems that continually reappear and appear to be common. For example, it is always important to create a good suspension system and it is equally important to provide a dynamically stable vehicle.

The Suspension System is vital for it determines the vehicle ride and vibration behavior and it also establishes the tolerable speed limit of travel over various terrain surfaces, both for the man and the machine.

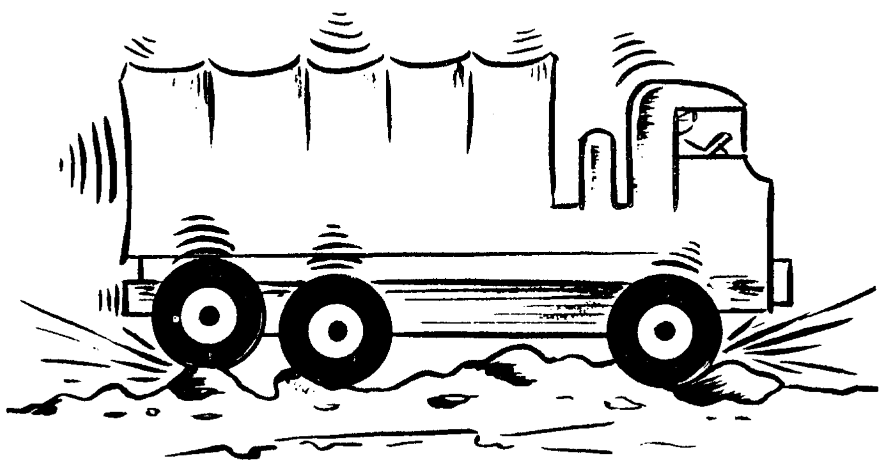


FIG. 1. STEP ONE

Vehicle stability is a design area that also receives considerable attention, particularly in combat vehicle programs where large caliber weapons are expected to be fired from chassis of reduced weight and decreased size. Within military circles reference to vehicle stability differs from the usual automotive connotation. In place of steering behavior or directional control, vehicle stability pertains to pitch and roll movement, resulting from the gun recoil forces.

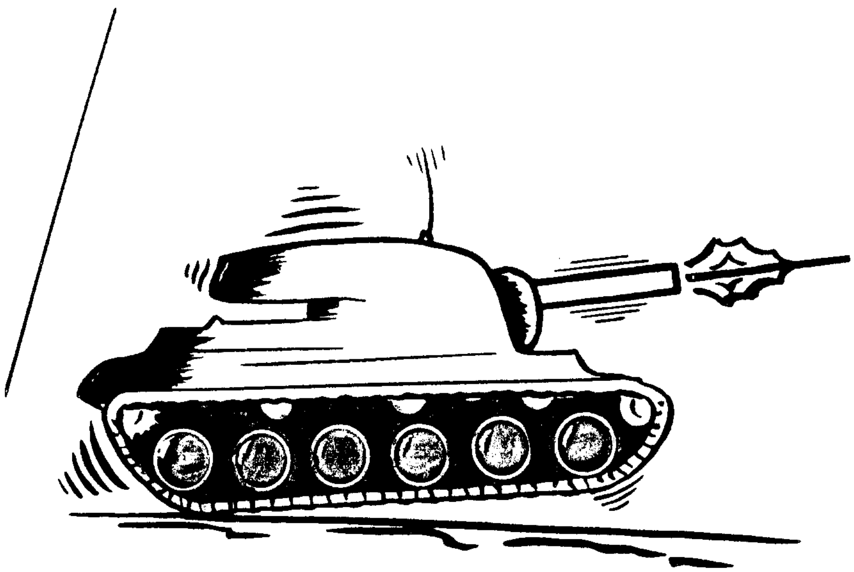


FIG. 2. STEP TWO

These two elements of vehicle design can be considered as perennials. They are always around and unfortunately they are rather "tough nuts" to handle. Individually they present major stumbling blocks to design engineers. Normally, useful study of these problems is beyond the level of developing a new layout or "cranking" through several equations on a desk calculator. As a result, the magnitude of each task has produced design specialists. These people, however, are not medicine men who consistently are able to generate successful answers.

They need help; they need either physical or analytical means to guide and measure their design approaches. Consequently, knowledge of these systems must be consistently bolstered and expanded. This demand requires unique capability:

First, since all wheeled and tracked vehicles are earth bound, knowledge of road profiles or terrain contours upon which they move must be secured.

Next, detailed mathematical models are necessary that describe the vehicle and how it reacts to external disturbances or internal design changes.

Then, an accurate recording procedure is essential to transmit results from high speed computers in such format that their meaning may be assessed graphically, visually or physically.

The benefits of integrating these steps would be a complete capability for realistic design evaluation comparable to controlled tests at a proving ground.

Accordingly, the Army Tank-Automotive Command sought a means to physically simulate the suspension performance and stability dynamics of a design while it was still in the blueprint stage, and to bring together the theory, mathematics and computer machinery to study each of these problems indoors in the laboratory. It was also the feeling that if this program was to be successful, it must produce the ability to predict behavior and allow practicing engineers the opportunity to pretest their designs before commitment to fabricate expensive wood mock-ups, experimental test rigs or engineering prototypes.

Suspension Simulation:

Simulation of vehicle suspension systems is basically the task of predicting the motion response of the vehicle to disturbances from the road. A thorough understanding of the elements that constitute this system and their interrelationship is essential to such study. Beginning with the road is perhaps logically the first step. The basic requirement is to present to the wheels, springs, and shock absorbers, the vertical displacement and frequency identical to those existing in road or terrain surfaces. For this purpose it is possible to construct a computer model of a road profile using either digital or analog computer techniques. The basic data profile information may be secured by conventional rod and transit means or automatic measuring instruments.

In real life the road is a stationary wave form over which the vehicle travels. The relationship between car velocity and the static road generates the suspension dynamics.

In a computer simulation the vehicle's forward movement cannot be faithfully reproduced. To conveniently maintain an order of reality the road is moved instead. The road profile is presented to the suspension components, as a continuous rearward velocity that normally would be road speed.

To accurately duplicate the physical case the road is presented to each wheel separately, properly phased so that the rear wheels "see" the same road irregularities as the front wheels, although at a later time. This phasing is governed by the wheel base and vehicle speed.

One successful procedure of road profile generation utilizes a digital computer and a digital-to-analog converter. The digital machine stores the road profile data in elevation increments. It selects the elevation that each wheel requires at a particular time and generates the time between elevations.

Several preliminary considerations which must be resolved before the construction of such a digital road function include:

1. The number of vehicle wheels.

2. The spacing between wheels.

3. The starting road level.

4. The number of computer cells per road increment.

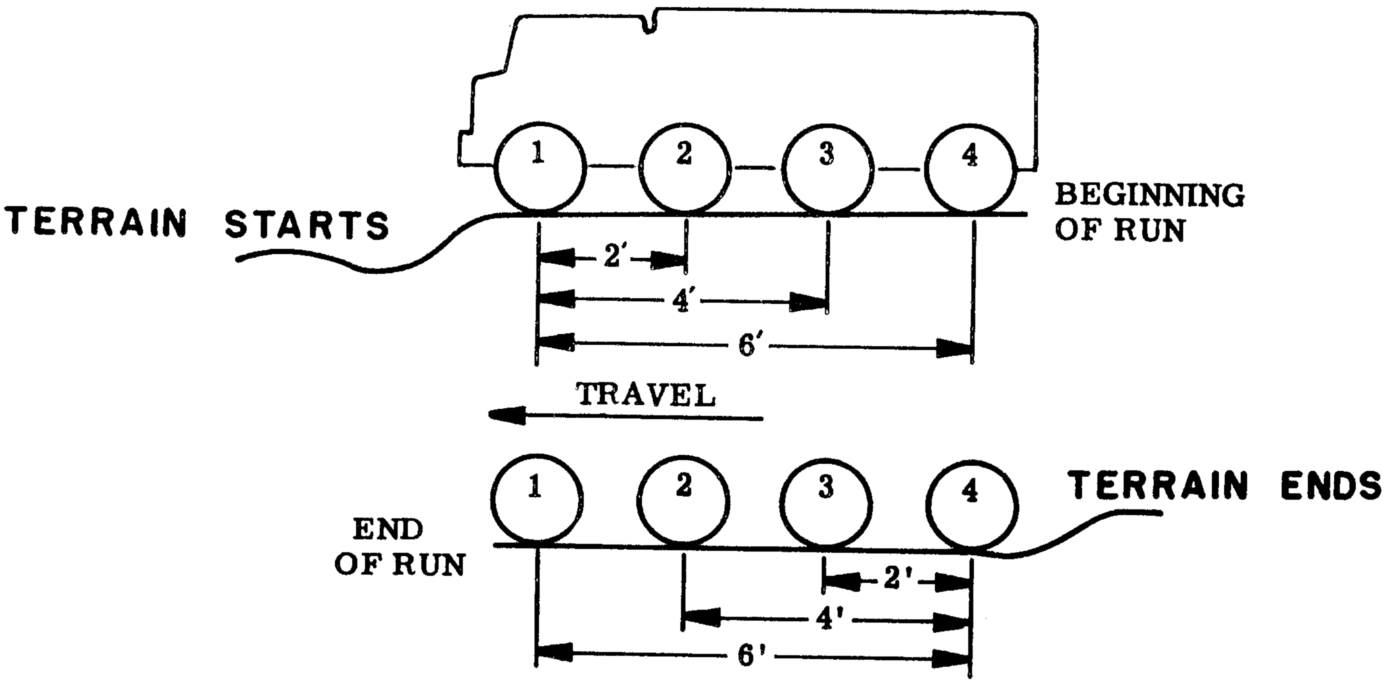
5. Overall length of the road profile.

The time at which a particular point on the road will arrive at each wheel is determined by the wheel spacing. This spacing is also considered in deciding how many elevation values will be equivalent to one linear foot of road. The following example will illustrate these points.

A vehicle suspension is set up on the analog computer; this simulation is for one side of the vehicle only, it being assumed that the other side is identical. The vehicle has four wheels on a side, spaced two feet apart. It will be driven over a Belgian Block type road. The road consists of 307 elevations spaced one foot apart. These things being known, it is possible to prepare a scheme for generating the road function which will pass under each wheel in sequence. To rerun the road after once traversing it, a starting road level must be assumed, usually the initial starting elevation, or very near to it. For the conditions just outlined a situation similar to that shown in Figure 3 will exist.

Preliminary assumptions:

1. Vehicle is sitting on road level.
2. Start of road strikes 1st wheel
3. One computer word = 1 linear ft.



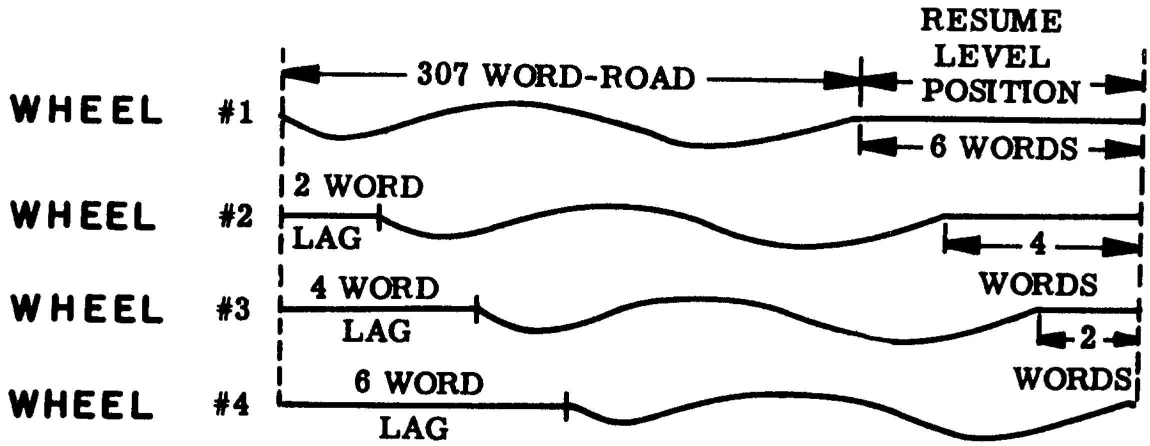


FIG. 3. ROAD PROFILE

The best procedure in this instance is to let one foot of road be represented by one computer word. However, if the spacing between wheels is uneven, a scheme utilizing several words to the linear foot would be required.

After the preliminary road function details have been accounted for, the actual generation of the road function can be undertaken. This naturally divides into the following steps:

1. Preparation of the data tape.

2. Placing road level data into computer memory.

3. Generation of the road function tape.

4. Transfer of the road function data to the digital-to-analog converter, etc.

Each computer "word" of information contains five channels, four of which are used for terrain simulation. Each of these channels may represent data. An algebraic representation of such a word is 1 aa bb cc dd 00, where each pair of letters represents one channel in the output of the Digital-Analog conversion system, while the number (1) in the sign position designates a particular group of D-A Converters, (Note that since the last channel is not used it is represented by 00, i.e. no information present). With proper scaling and programming, each channel can become a road-profile-wheel-terrain-function-generator (RPWTFG). Hence, there are four RPWTFG's per computer word.

The time between data increments on the computer is generated by using a time control subroutine which increases or decreases the time between "calling up" the data increments. The speed of the road function is determined from the recorder tracings as follows:

$$\text{mph} = \frac{\text{actual length of road (ft)}}{\text{length of converted road (mm)}} \times \text{paper speed} \left(\frac{\text{mm}}{\text{sec.}}\right) \times \frac{30\ \text{mph}}{44\ \text{ft/sec}}$$

By this procedure vehicle road speed is established. Maximum road speed is only limited by the computers ability to call up data. Utilization of a time delay subroutine within the computer facilitates decreasing the speed. Also, greater speed can be attained by shortening the road, i.e. picking up every second or every third road elevation. Oscillograph recordings, as generated by this system, are illustrated in Figure 4.

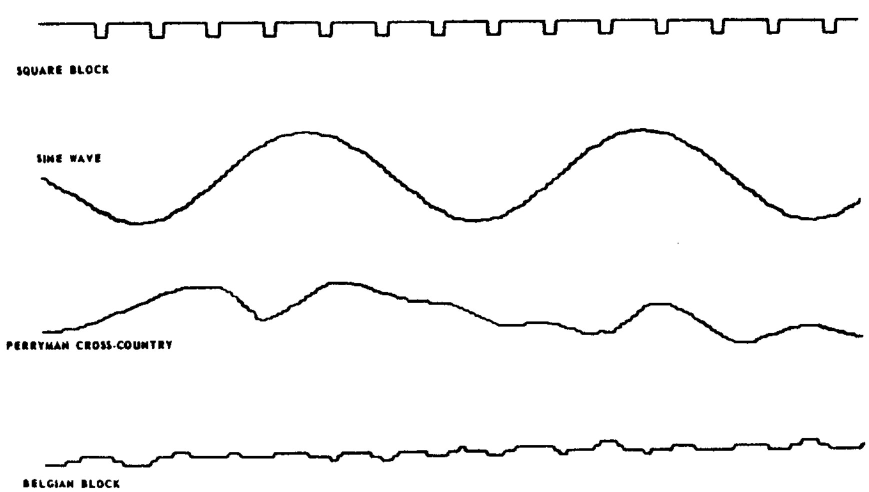


FIG. 4. ROAD PROFILE

A magnetic tape recorder and reproducer add convenience and efficiency to the system. By recording successive speeds on magnetic tape, the digital computer is used only once for a particular vehicle. In addition, velocity multiplication may be obtained by recording at one speed and reproducing at higher speeds.

Thus, any terrain that can be numerically described as increments of elevation with respect to horizontal distance, can be simulated with a digital computer for the analysis of vehicle behavior.

With the road prepared the simulation requires a model of the suspension system. For this purpose the Analog computer is best suited. The computer, as used, provides an accurate representation of the design. In a true sense the computer is an electronic model of the vehicle. The degree of realism achieved is naturally governed by the quantity of vehicle characteristics simulated.

As in any simulation, the system is first described by a mathematical model. The equations represent the dynamic system - the vehicle chassis, the suspension system and the road surface input.

Vehicular vibration components include the mass and inertia of the sprung components, the suspension springs, shock absorbers, road wheel masses, and the spring and damping characteristics of the wheel assembly.

The typical method of describing a vehicle to be simulated is shown in Figure 5 and Figure 6. From the diagrams, the equations of motion may be stated. These expressions are written as common linear differential equations with non-linear coefficients.

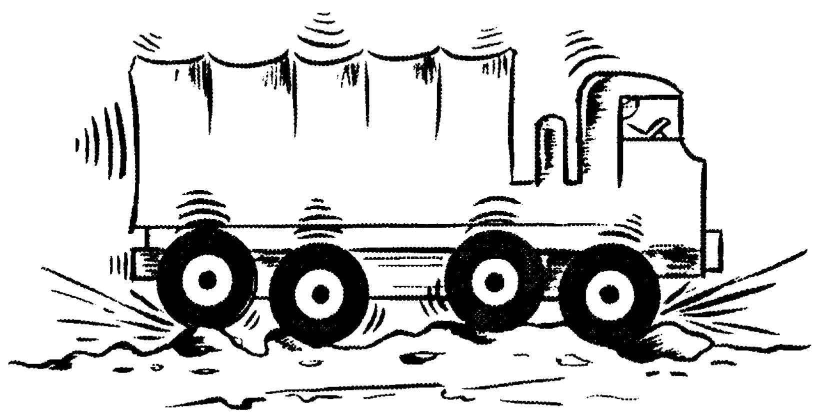


FIG. 5. VEHICLE CONFIGURATION

Non-linearities in the simulation exist normally due to non-linear spring characteristics, double acting shock absorbers, and the fact that wheels may leave the road surface.

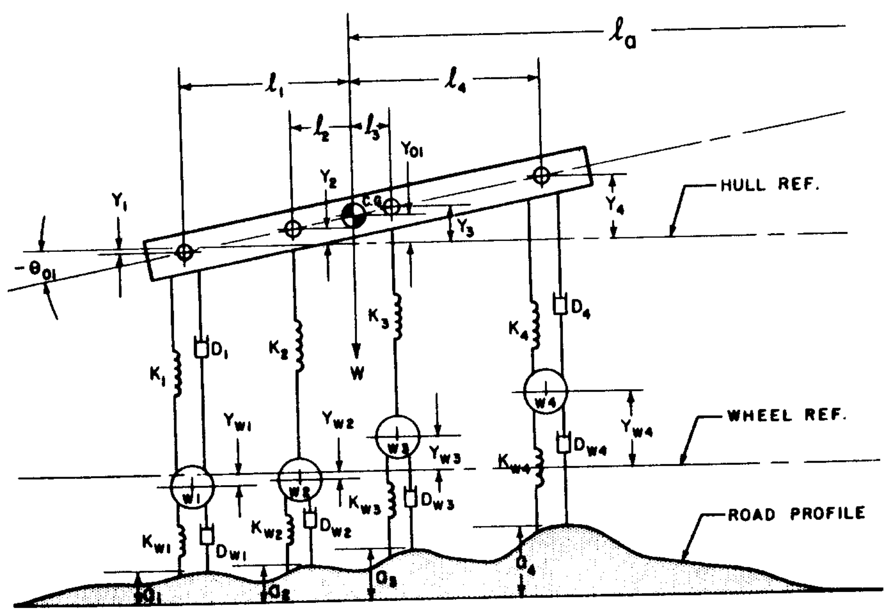


FIG. 6. VEHICLE SCHEMATIC

The equations result from linear and angular counterparts of Newton's second law of motion. The summation of the vertical forces on the chassis equals the mass of the chassis times its vertical acceleration; the summation of the torques about the center of gravity is equal to the polar moment of inertia times the angular acceleration of the hull. The forces and torques result from relative displacement and velocity of the springs and shock absorbers respectively. For example, the vertical force of the front wheel spring is equal to the spring constant (K_1) times the relative displacement between the wheel and the chassis just above the wheel. Similarly, the torque is a product of this force times the distance from the center of gravity. The shock absorber force is a product of the damping coefficient and the relative velocity between the wheel and the chassis above the wheel.

Chassis, pitch and bounce equations may be developed using these relationships. Separate differential equations are written to describe the motion of each wheel. Auxiliary equations are written to relate the pitch motion to the vertical motion so that the displacement and velocity of the chassis at each wheel station may be found.

Each wheel of the vehicle is considered a separate mass, spring, and damper system connected to the ground and to the chassis, the link to the chassis being the suspension spring and shock absorber. The force exerted on the wheel by the ground is equal to a product of the wheel rubber displacement and spring constant. This force may have only one sign since the ground cannot "pull" down on the wheel. The suspension spring force is also exerted on the wheel.

Simulation Equations:

Chassis Vertical Motion:

$$\ddot{Y}_o = \frac{\sum F_y}{M_o} \quad \text{(c.g. Bounce Acceleration)}$$

$$\ddot{Y}_o = -\frac{K_1}{M_o}(Y_1 - Y_{w1}) - \frac{K_2}{M_o}(Y_2 - Y_{w2}) - \frac{K_3}{M_o}(Y_3 - Y_{w3})$$

$$-\frac{K_4}{M_o}(Y_4 - Y_{w4}) - \frac{D_1}{M_o}(\dot{Y}_1 - \dot{Y}_{w1}) - \frac{D_4}{M_o}(\dot{Y}_4 - \dot{Y}_{w4}) + g$$

Chassis Pitch Motion:

$$\ddot{\theta} = \frac{\sum T}{J_o} \quad \text{(c.g. Angular Acceleration)}$$

$$\ddot{\theta} = -\frac{K_1 l_1}{J_o}(Y_1 - Y_{w1}) - \frac{K_2 l_2}{J_o}(Y_2 - Y_{w2})$$

$$+ \frac{K_3 l_3}{J_o}(Y_3 - Y_{w3}) + \frac{K_4 l_4}{J_o}(Y_4 - Y_{w4})$$

$$- \frac{D_1 l_1}{J_o}(\dot{Y}_1 - \dot{Y}_{w1}) + \frac{D_4 l_4}{J_o}(\dot{Y}_4 - \dot{Y}_{w4})$$

Vertical Wheel Motion:

$$\ddot{Y}_w = \frac{\sum F_w}{M_w}$$

$$\ddot{Y}_{w1} = -\frac{K_{w1}}{M_{w1}}(Y_{w1} - a_1) - \frac{D_{w1}}{M_{w1}}(\dot{Y}_{w1} - \dot{a}_1) + \frac{K_1}{M_{w1}}(Y_1 - Y_{w1})$$

$$+ \frac{D_1}{M_{w1}}(\dot{Y}_1 - \dot{Y}_{w1}) + g$$

$$\ddot{Y}_{w2} = -\frac{K_{w2}}{M_{w2}}(Y_{w2} - a_2) - \frac{D_{w2}}{M_{w2}}(\dot{Y}_{w2} - \dot{a}_2) + \frac{K_2}{M_{w2}}(Y_2 - Y_{w2}) + g$$

$$\ddot{Y}_{w3} = -\frac{K_{w3}}{M_{w3}}(Y_{w3} - a_3) - \frac{D_{w3}}{M_{w3}}(\dot{Y}_{w3} - \dot{a}_3) + \frac{K_3}{M_{w3}}(Y_3 - Y_{w3}) + g$$

$$\ddot{Y}_{w4} = -\frac{K_{w4}}{M_{w4}}(Y_{w4} - a_4) - \frac{D_{w4}}{M_{w4}}(\dot{Y}_{w4} - \dot{a}_4) + \frac{K_4}{M_{w4}}(Y_4 - Y_{w4})$$

$$+ \frac{D_4}{M_{w4}}(\dot{Y}_4 - \dot{Y}_{w4}) + g$$

Auxiliary Chassis Equations:

$$Y_{1-4} = Y_o + \ell_{1-4}\ \mathrm{SIN}\ \theta \qquad \dot{Y}_{1-4} = \dot{Y}_o + \ell_{1-4}\ \mathrm{SIN}\ \theta$$

The non-linearities are best described in graphical form, as is shown in Figures 7 and 8.

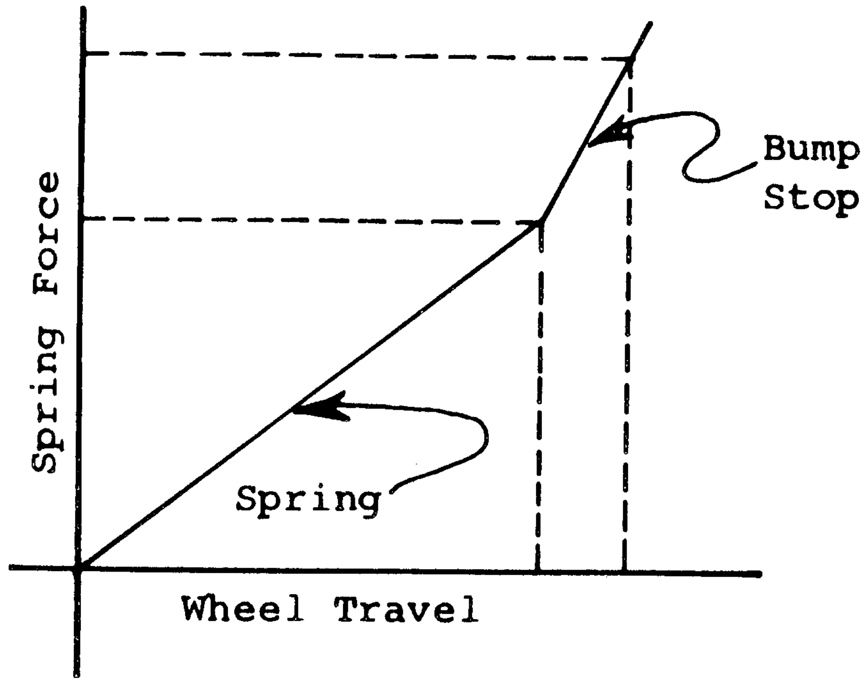


FIG. 7. SPRING LOAD VS WHEEL DEFLECTION

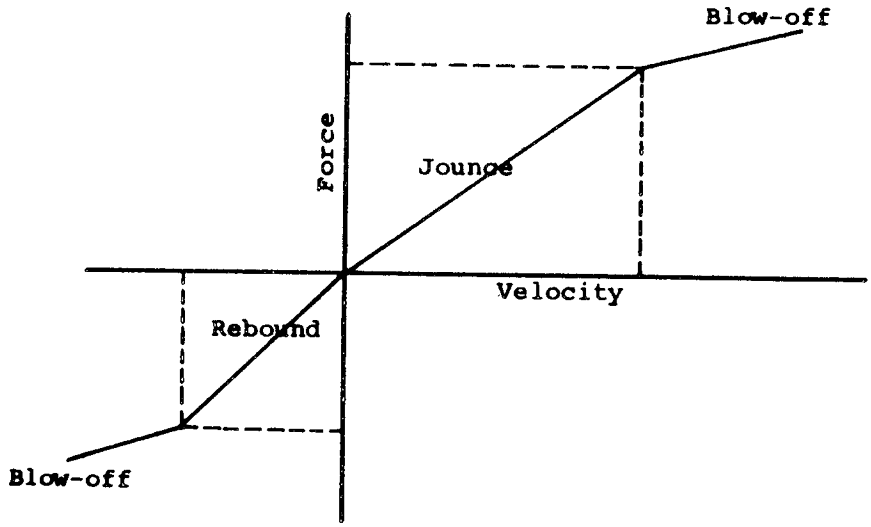


FIG. 8. SHOCK ABSORBER VS VELOCITY AT WHEEL

The non-linear suspension springing is composed of two linear segments, the one of lesser slope being the suspension spring, and the other the bump stop. The shock absorber non-linearity is shown in Fig. 8, which has four linear segments simulating different rates in compression and expansion with blow-off valves. The circuitry for creating the significant segments of a suspension simulation are shown in Figures 9 - 13. If a wheel leaves the ground, no spring force can exist between the ground and the wheel. To provide for this realistic action, a diode representing a unidirectional spring force is put in series with the wheel feedback loop, as is shown in the wheel circuit, Figure 11.

$$\dot{Y}_1 - \dot{Y}_{W1} = \dot{Y}_0 + \ell_1 \dot{\Theta} - \dot{Y}_{W1}$$

$$\dot{Y}_4 - \dot{Y}_{W4} = \dot{Y}_0 - \ell_4 \dot{\Theta} - \dot{Y}_{W4}$$

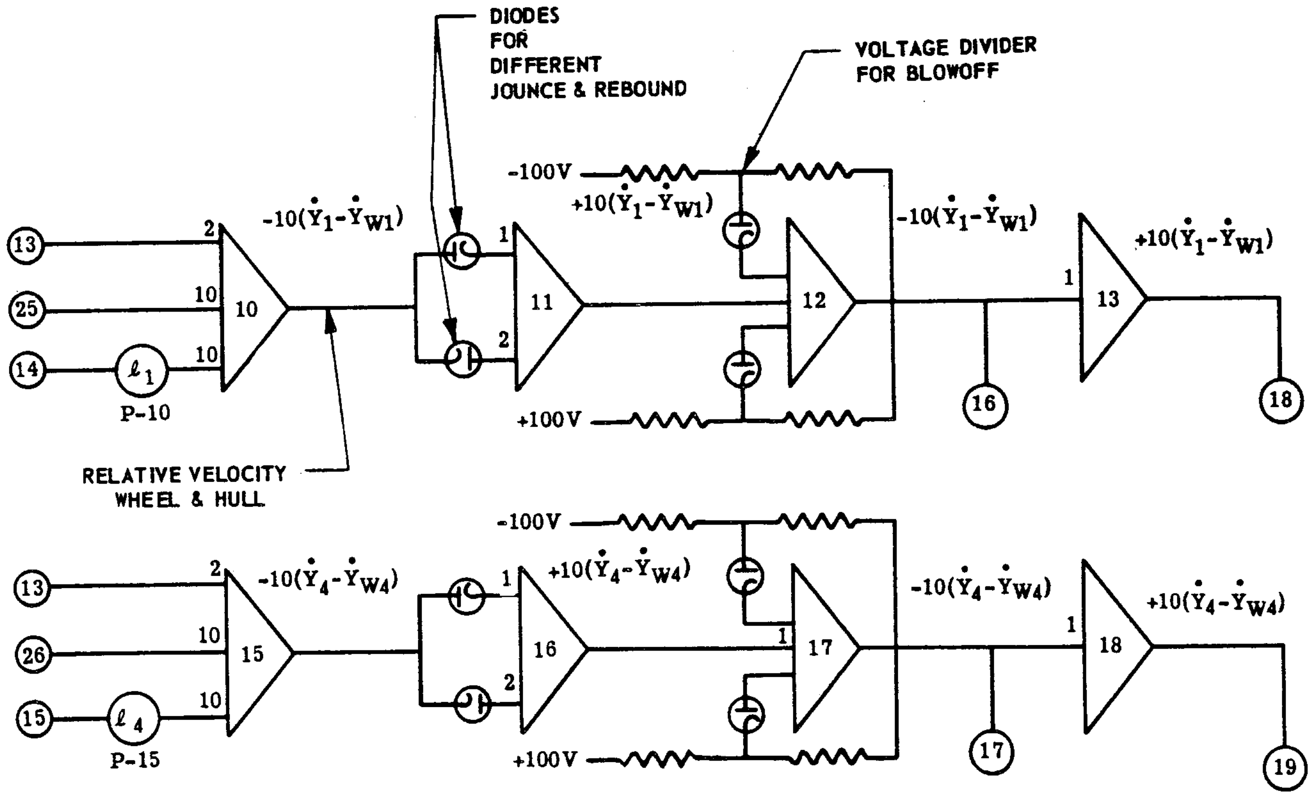


FIG. 9 SHOCK ABSORBER CIRCUITS

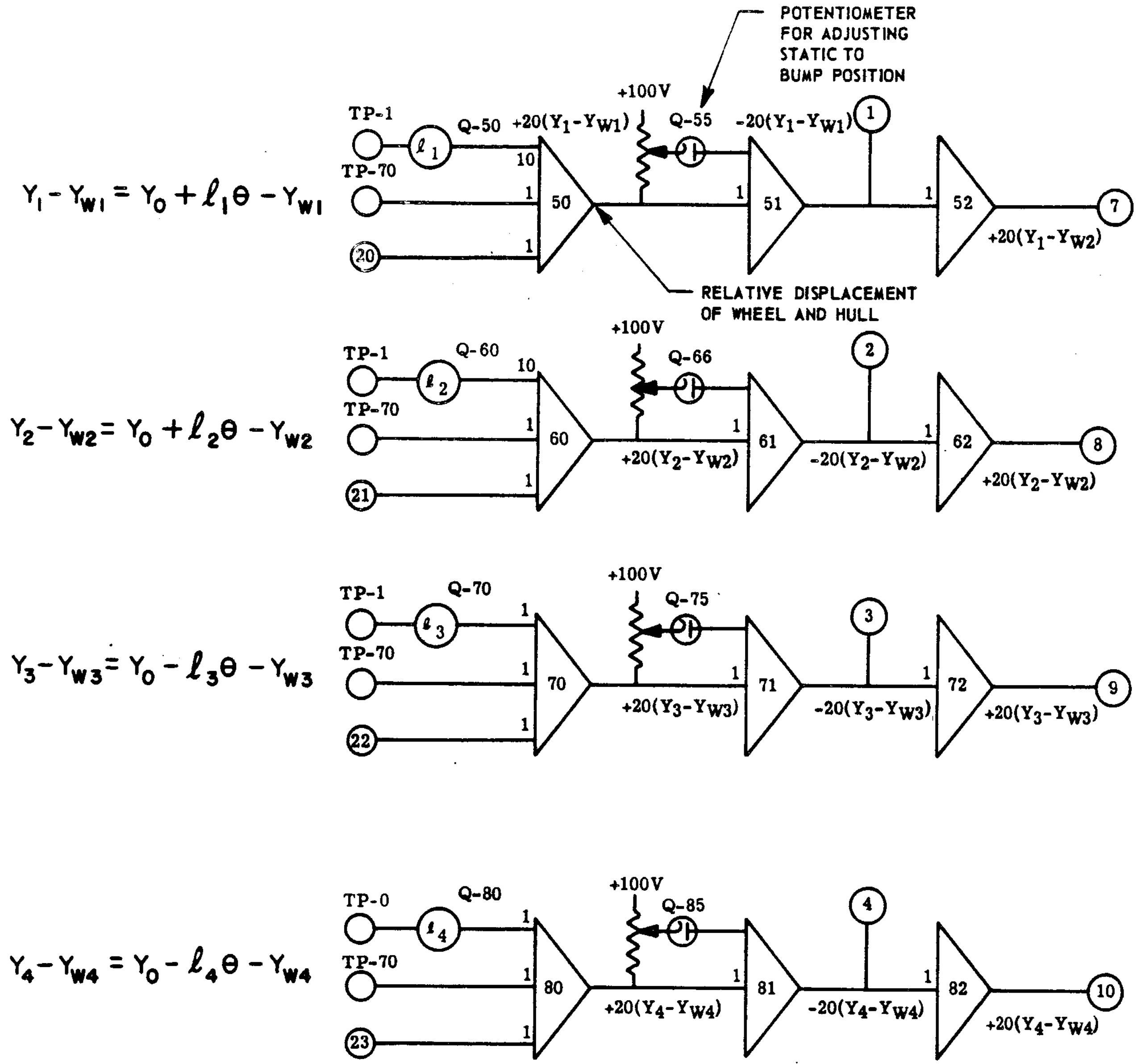


FIG. 10 SPRING CIRCUITS

$$\Sigma \frac{F_w}{M_w} = \ddot{Y}_w$$

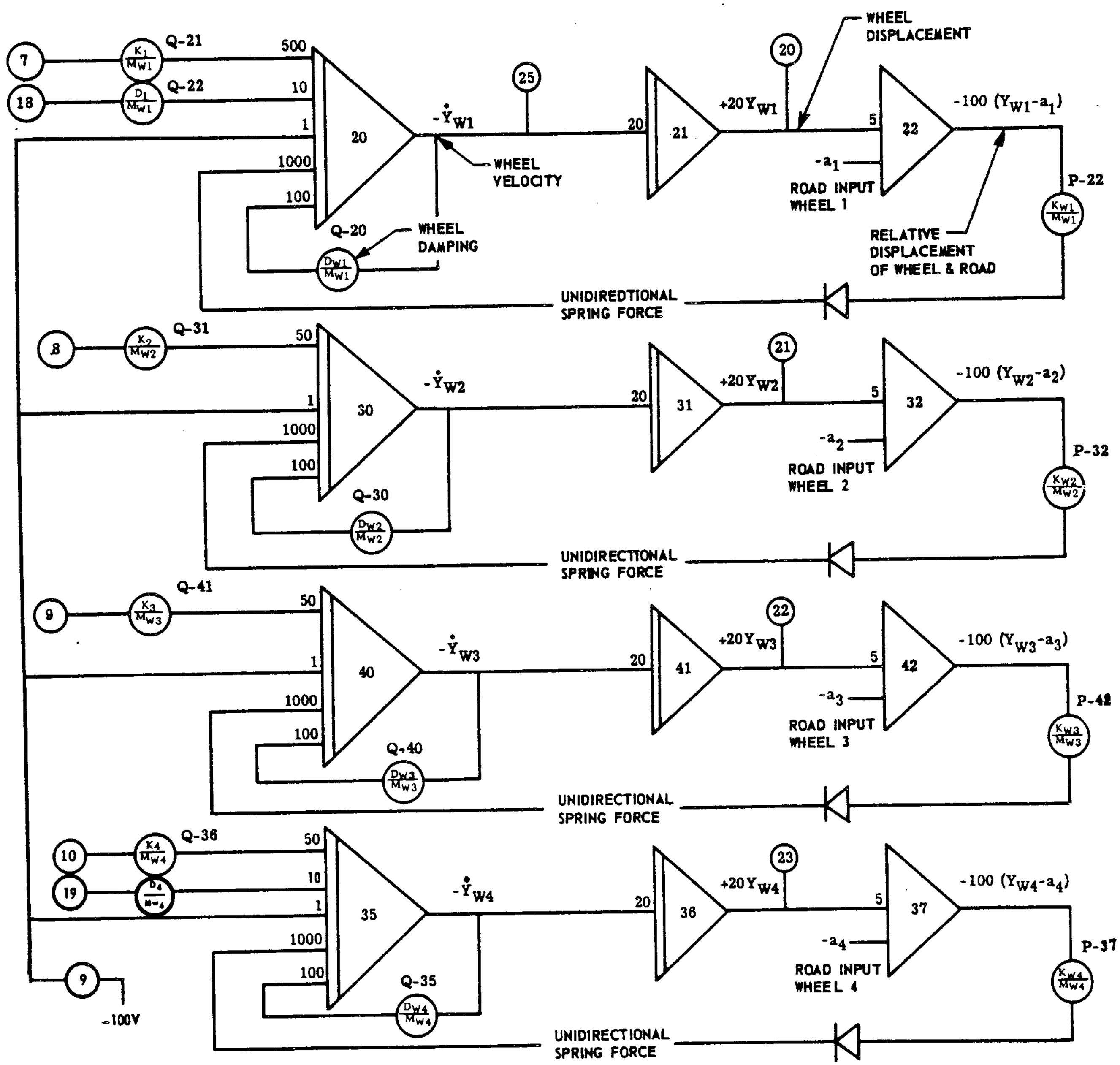


FIG. 11 WHEEL VELOCITY, AND DISPLACEMENT

$$\frac{\Sigma F_Y}{M_O} = \ddot{Y}_O$$

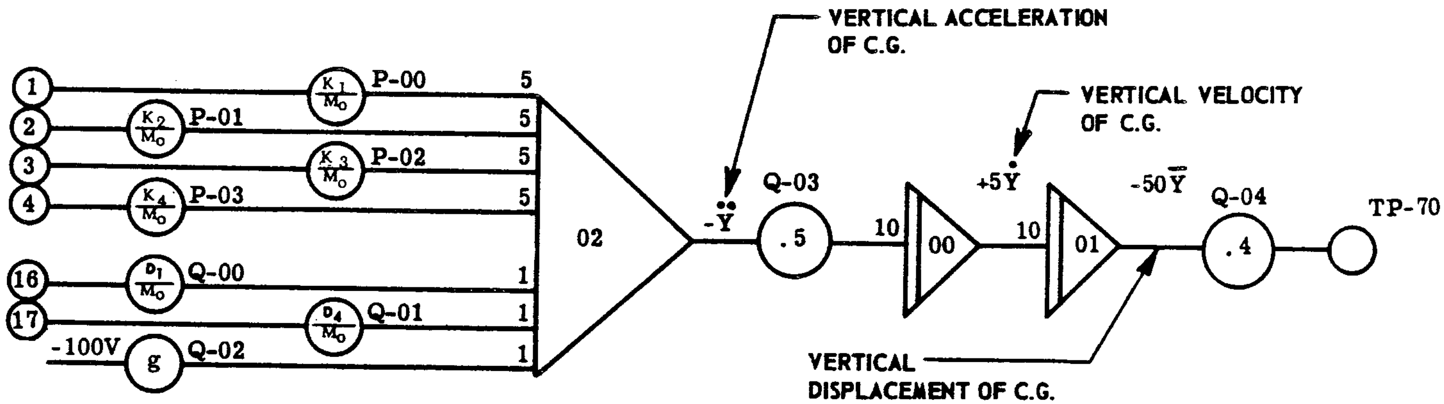


FIG. 12 C.G. VERTICAL ACCELERATION, VELOCITY, DISPLACEMENT

$$\frac{\Sigma T}{J_o} = \ddot{\theta}$$

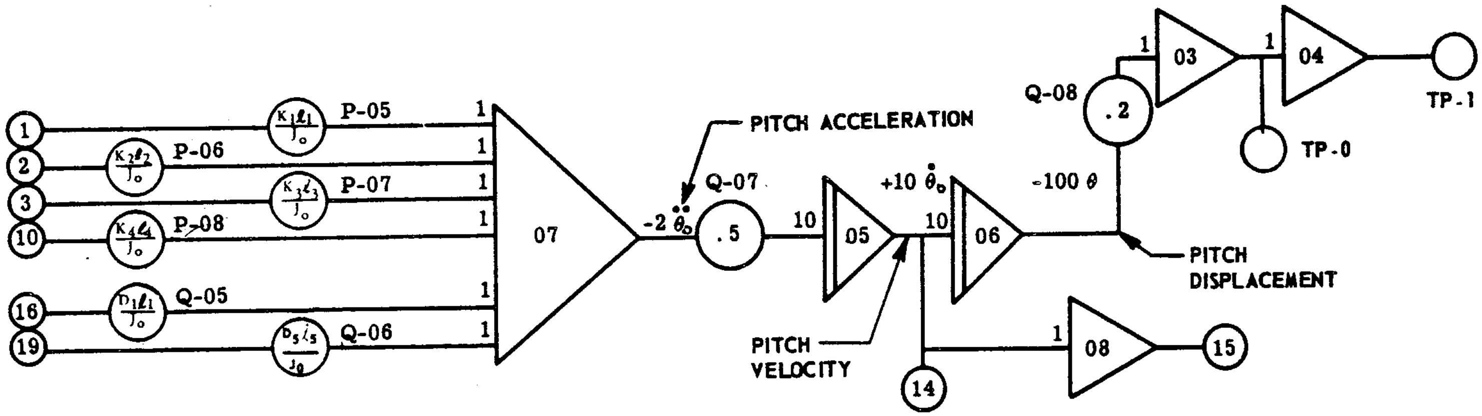
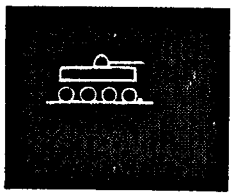
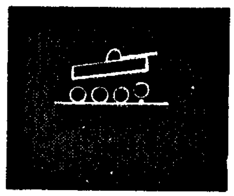


FIG. 13 C.G. PITCH ACCELERATION, VELOCITY, DISPLACEMENT

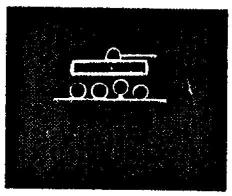
Simulation begins when the electronic version of a road and suspension system are brought together. Results are best analyzed using oscillographic or pen recorder output devices. Recorded paper tracings provide an excellent permanent record for lengthy detailed analysis. The oscillograph display system offers an opportunity to observe the simulation visually as an animated presentation. The dynamics of a complete vehicle or any component thereof may then be studied. This display system is used in conjunction with the analog computer. A cathode-ray-tube is used to convert the output voltages of the computer into a direct pictorial representation. Application of this system is shown in Figure 14. The series of photographs describe the motion of a tank that would be seen on the tube. The vehicle is shown negotiating at successive instances a 4" x 4" square obstacle.



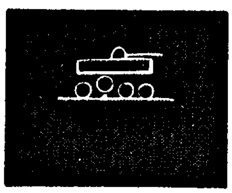
VEHICLE APPROACHING OBSTACLE



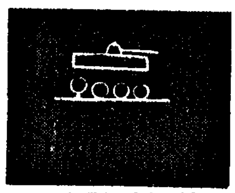
WHEEL ONE PASSING OVER OBSTACLE



WHEEL TWO PASSING OVER OBSTACLE



WHEEL THREE PASSING OVER OBSTACLE



WHEEL FOUR PASSING OVER OBSTACLE

FIG. 14. VISUAL DISPLAY SYSTEM

This visual display provides a quick and easy method of conducting a preliminary analysis of new suspensions. It is also a good means of debugging a new simulation setup.

<u>Simulator</u>:

Results of computer simulations may also be studied with the aid of a motion simulator. The value of the Simulator, Figure 15, lies in its ability to physically reproduce realistic "ride motion" that may be predicted by a computer simulation. Thus, by combining computer studies of new concepts with a simulator analysis, design merits may be judged in the laboratory by engineers, designers, and administrative people before a design is considered for fabrication. Each individual may ride a new suspension in the Simulator and personally evaluate his area of interest firsthand.

The Simulator described below is a four degree of freedom machine capable of providing bounce, pitch, roll and yaw motions.

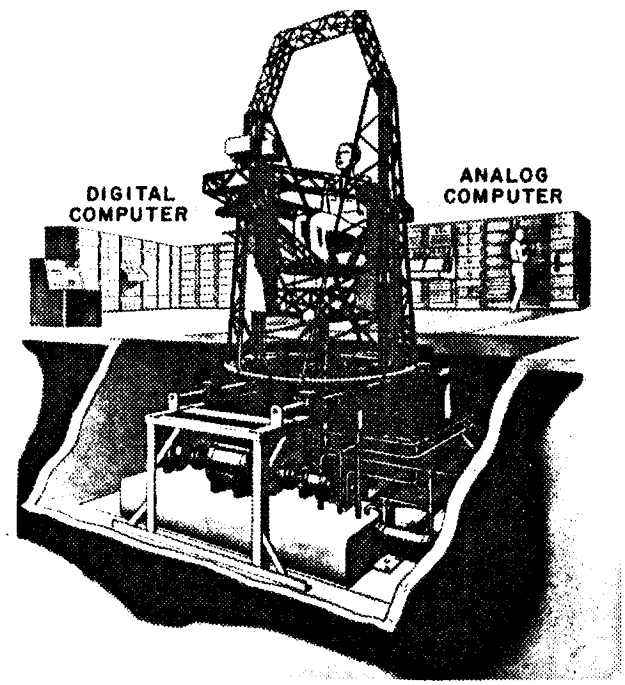


FIG. 15. SIMULATOR

The control of the machine is optional. The Simulator may be controlled from an instrument panel to produce either sine, square or triangular motions. Random motion inputs may be fed directly into the Simulator from an Analog computer simulation or by reproducing information previously recorded on magnetic tape.

This machine is hydraulically driven and electronically controlled. Each of the four motions may be used individually or simultaneously.

Motion	Max. Tot. Travel	Max. Frequency	Acceleration
Bounce	3 ft	10 cps	2 g's
Roll	40 deg	10 cps	30 radians/sec^2
Pitch	40 deg	10 cps	30 radians/sec^2
Yaw	20 deg	3 cps	15 radians/sec^2

Perhaps the most significant claim that can be broadcast for the Simulator, at this time, is that it will make possible performance trials of designs prior to building of a design. In some instances it is the only economical approach, considering time and cost, particularly, where a new design is being investigated using many alternatives.

The creation of this Simulator provides the Army Tank-Automotive Command with a design tool that has been sought for some time. The need for an instrument of this kind has been in continuous demand for military suspension studies and other shock and vibration programs.

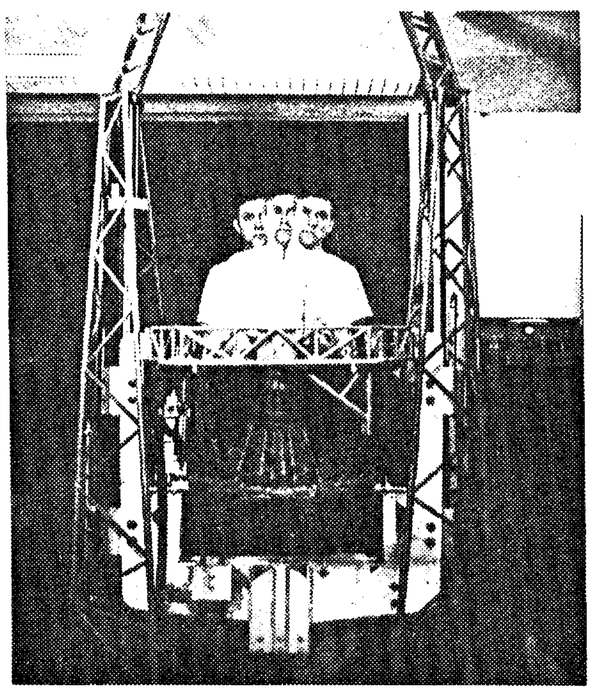


FIG. 16. DYNAMIC SIMULATION

The immediate response to the Simulator was generally favorable, but reserved. Comments usually indicated that the vibratory motions were good. However, it was repeatedly stated that the laboratory environment around the Simulator degraded the intended realism. The common complaint was that the "out of doors" atmosphere seemed to be missing.

To compensate for this a visual display was created providing a 180 degree field of view horizontally and 48 degrees in the vertical plane. A 35mm motion picture format was used to produce a "three screen" presentation. This method was selected, based upon successful tryouts of a unique projection system developed and tailored to the Simulator.

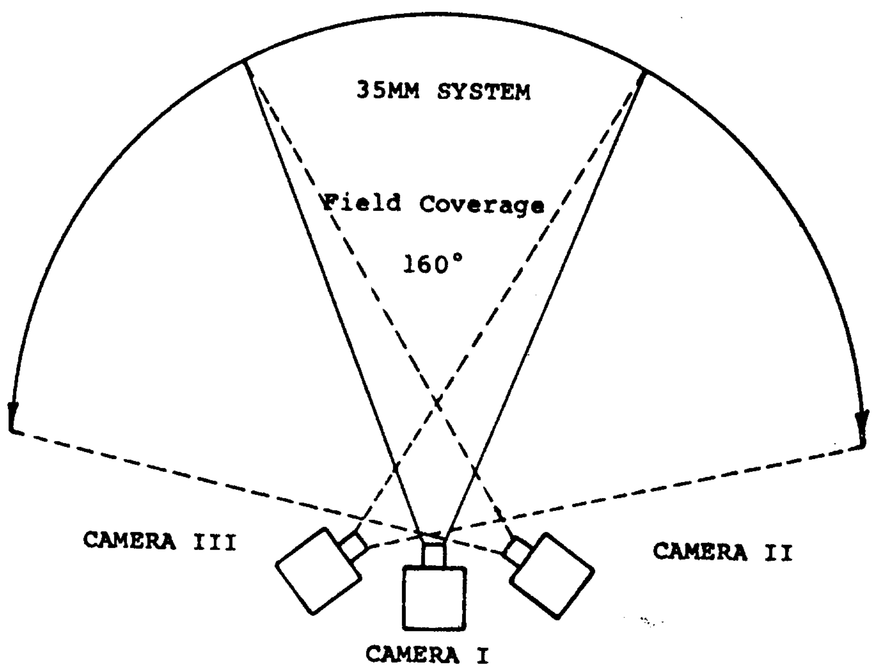


FIG. 17. CAMERA SYSTEM

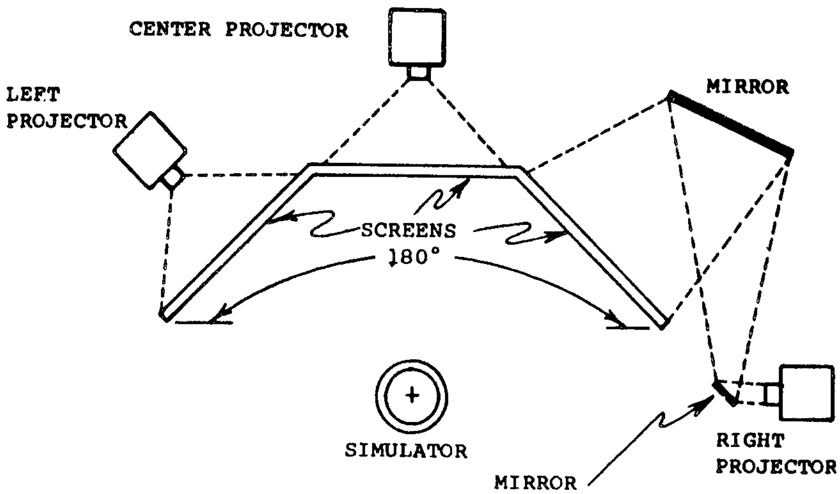


FIG. 18. PROJECTION SYSTEM

The activity scene is photographed by three cameras and backprojected to the subject in the Simulator by three synchronized-interlocked projectors. This system presents to the observer a scene that compares favorably to a view from within a moving vehicle.

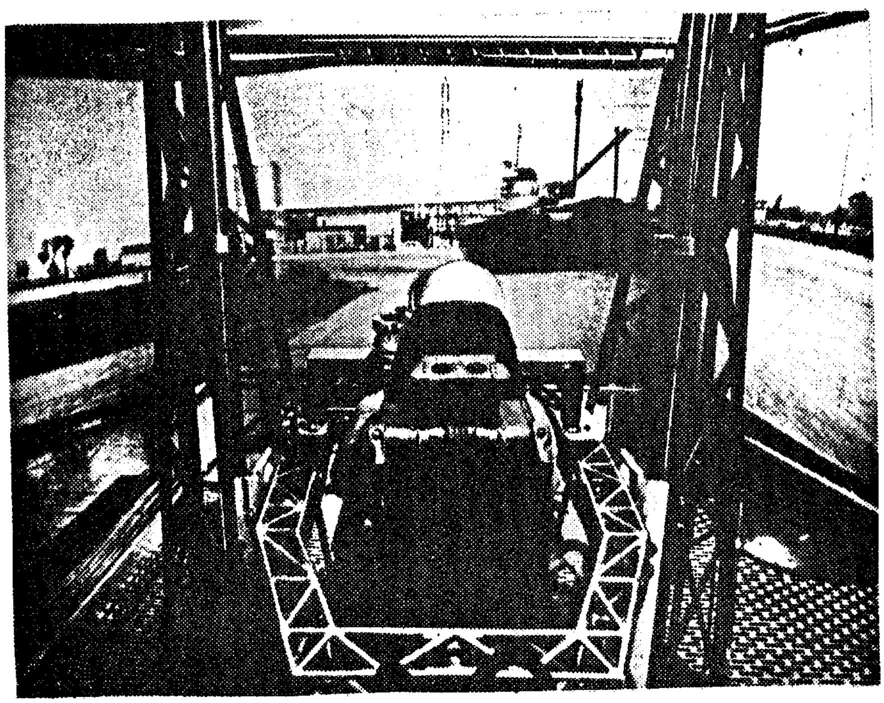


FIG. 19. DISPLAY SYSTEM

<u>Vehicle Stability</u>:

The Simulator was also used to simulate vehicle firing stability dynamics. The starting cue for this program was very forceably observed in vehicles like the Self-Propelled Artillery Weapon shown in Figure 20.

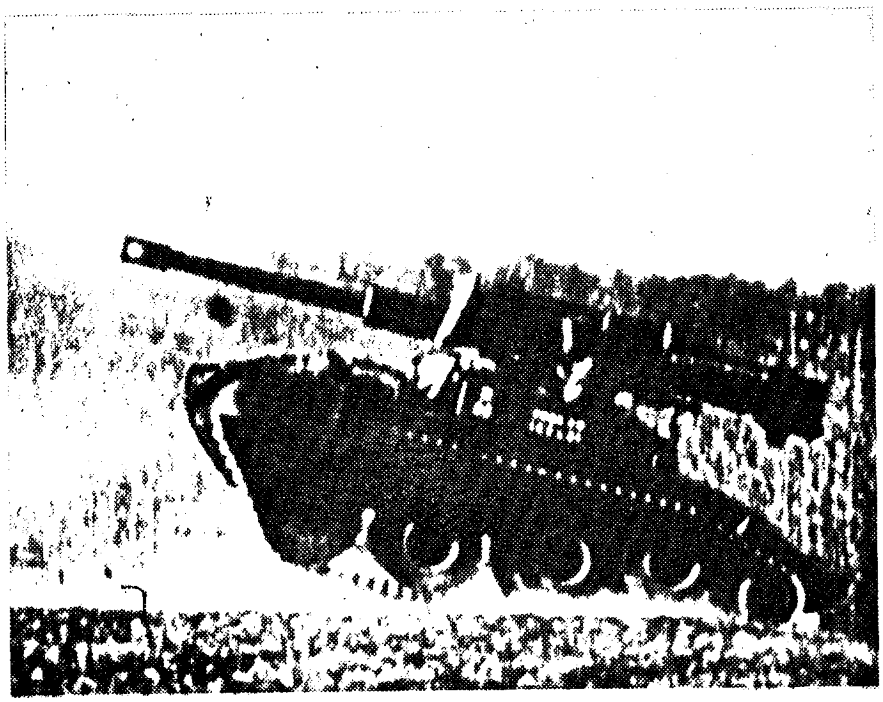


FIG. 20. SELF-PROPELLED ARTILLERY WEAPON

In keeping with this indicated trend of big guns on small chassis platforms it was necessary to establish with greater accuracy, the stability of contemplated designs.

The characteristic events describing fire stability were analyzed by charting the flow of events and establishing the equations of motion.

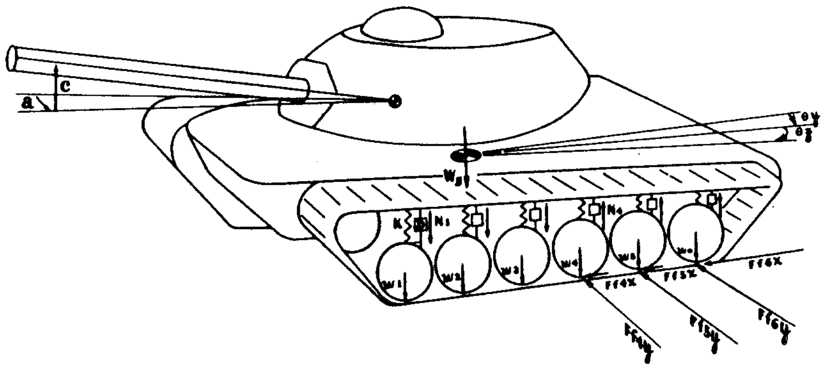


FIG. 21. VEHICLE SCHEMATIC

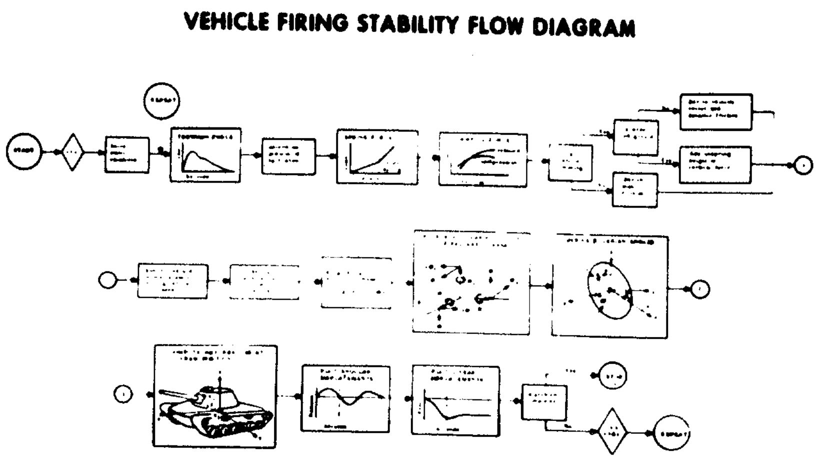


FIG. 22. FLOW DIAGRAM

Equations of Motion:

LONGITUDINAL TRANSLATION: GUN FORCE + GROUND FRICTIONAL FORCE = MASS X ACCELERATION $F_{gx} + F_{fx} = m\ (\ddot{x} - y\ \dot{\theta}_z + z\dot{\theta}y)$

LATERAL TRANSLATION: GUN FORCE + GROUND FRICTIONAL FORCE = MASS X ACCELERATION $F_{gy} + F_{fy} = m\ (\ddot{y} - \dot{z}\theta_x + \dot{x}\theta_z)$

BOUNCE: GUN FORCE - SPRUNG WEIGHT + SUSPENSION FORCES = MASS X ACCELERATION $F_{gz} - W_s c_z Z + \sum_{i=1}^{n} N_i = m\ (\ddot{z} - \dot{x}\theta_y + \dot{y}\dot{\theta}_x)$

ROLL: GUN MOMENT + GROUND FRICTIONAL MOMENT + SUSPENSIONAL MOMENT = ANGULAR ACCELERATION X MOMENT OF INERTIA $-\bar{z}F_{gy} + \bar{y}F_{gz} + (z + z_0)\ F_{fy}c_yY + \sum_{i=1}^{n} N_iY_i = \ddot{\theta}_xI_x - \ddot{\theta}_zI_{xz} +$

$$(I_z - I_y)\ \theta_y\dot{\theta}_z - I_{xz}\dot{\theta}_x\dot{\theta}_y$$

PITCH: GUN MOMENT - GROUND FRICTIONAL MOMENT - SUSPENSIONAL MOMENT = ANGULAR ACCELERATION X MOMENT OF INERTIA $-\bar{x}F_{gz} + \bar{z}F_{gx} - (z + z_0)\ F_{fx}c_xX - \sum_{i=1}^{n} N_ix_i = \ddot{\theta}_yI_y + \dot{\theta}_x\dot{\theta}_x\ (I_x - I_z)$

$$+ (\dot{\theta}_x^2 - \dot{\theta}_z^2)\ I_{xz}$$

YAW: GUN MOMENT + GROUND FRICTIONAL MOMENTS = ANGULAR ACCELERATION X MOMENT OF INERTIA $-\bar{y}F_{gx} + \bar{x}F_{gy} + \sum_{i=1}^{n}$

$x_iF_{fiy}\ c_yY - \sum_{i=1}^{n} y_iF_{fix}c_xX = \ddot{\theta}_zI_z - \ddot{\theta}_xI_{xz} +$

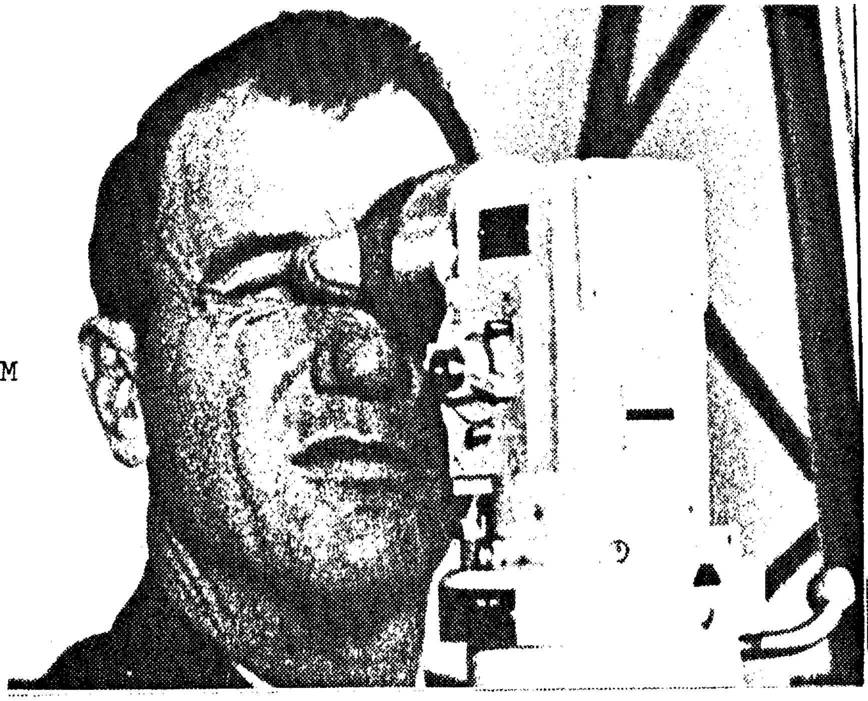
$(I_y - I_x)\ \dot{\theta}_x\dot{\theta}_y + I_{xz}\dot{\theta}_y\dot{\theta}_z$

The derived statements were for weapon systems free to move in three degrees of angular freedom - roll, pitch and yaw; and three degrees of translational freedom - fore and aft movement, bounce, and lateral slip. The equations define vehicle motion as effected by interrelated factors of gun firing force, gravity, terrain influence, and the resisting forces of the suspension.

The dynamics of firing stability are calculated on a digital computer. When this program is used in conjunction with the Simulator, the results are stored in computer memory. The physical arrangement of the Simulator permits the occupant

to fire any weapon by merely pulling the usual trigger. The command to fire is completely controlled by the man in the seat.

AIM



FIRE

FIG. 23. FIRING SIMULATION

The inputs to this problem consist of various vehicle measurements, weights, moments of inertia, gun recoil force, type of soil, and type of suspension (active or locked-out). The output consists of detailed information in tabular or graph form showing angular and translational disturbances and their respective displacements, velocities, and accelerations with respect to time. This information describing gun firing force impact on the vehicle and resultant vibrations is available for the C.G. of the vehicle with reference to earth-fixed axes and for any other point on or within the vehicle, such as gun muzzle, engine mounts and crew stations with reference to the vehicle axes.

Summary:

The combination of computers and simulation techniques at the Army Tank-Automotive Command has provided an effective and versatile designer's tool. Informative preliminary studies have been conducted of new suspension systems and stability characteristics, without the use of hardware units of the design. Probes of unique approaches have quickly established design direction and payoff areas.

SUSPENSION NOMENCLATURE

$\ddot{Y}_o$ = Vertical acceleration of the center of gravity.

$\dot{Y}_o$ = Vertical velocity of C.G.

Y_o = Vertical displacement of C.G.

$\ddot{\theta}_o$ = Pitch acceleration about C.G.

$\dot{\theta}_o$ = Pitch velocity about C.G.

θ_o = Pitch displacement at C.G.

$(Y_1 - Y_{w1})$ = Relative displacement between hull and wheel at wheel 1.

$(\dot{Y}_1 - \dot{Y}_{w1})$ = Relative velocity of the hull and wheel at wheel 1.

$(Y_{w1} - a_1)$ = Relative displacement between wheel and input bump at wheel 1.

J_o = Pitch Moment of Inertia.

M_o = Sprung mass.

M_w = Wheel mass.

ℓ = Distance from wheel centerline to C.G.

K_{1-4} = Suspension spring constant.

D_{1-4} = Shock absorber damping constant.

K_w = Spring constant of road wheel rubber.

D_w = Damping constant of road wheel rubber.

a_{1-4} = Road inputs to wheel No. 1-4.

Y_{1-4} = Chassis displacement.

g = Acceleration of gravity.

VEHICLE STABILITY NOMENCLATURE

F_g = Gun force.

F_x = Frictional force.

W_s = Sprung weight.

M = Sprung mass.

$\bar{x}$, $\bar{y}$, $\bar{z}$ = Position of trunnion centerline.

z_o = Static height of C.G.

n = Number of wheels.

$\ddot{y}$, $\ddot{x}$, $\ddot{z}$ = Translational acceleration.

$\ddot{\theta}_x$ $\ddot{\theta}_y$ $\ddot{\theta}_z$ = Angular acceleration.

$\dot{y}$ $\dot{x}$ $\dot{z}$ = Translational velocity.

$\dot{\theta}_x$ $\dot{\theta}_y$ $\dot{\theta}_z$ = Angular velocity.

I = Moment of Inertia.

K = Suspension spring constant.

D = Shock absorber damping.

a = Gun azimuth.

e = Gun elevation.

ACKNOWLEDGMENTS

Acknowledgment is extended to the past and present members of the Dynamic Simulations Laboratory at the Army Tank-Automotive Command for their individual contributions which led to the completed development described in this paper, and to Mr. O. K. Till, F. Anderson and R. Armstrong for development of the experimental equipment and instrumentation. I am also indebted to Mr. J. Stanton and J. Lynch, Jr. for their art and photographic work in preparing the illustrations.

REFERENCES

1. S. Heal, "Suspension Analysis", Army Tank-Automotive Command, Research Report No. RR-38.

2. M. Archambault, S. Heal, S. Staich, "Generation of Road Profiles for Vehicle Ride Simulations", Army Tank-Automotive Command, Research Report No. RR-44.

3. C. Fischer, "Mathematical Model and Digital Computer Program for Vehicle Firing Stability Analysis", Army Tank-Automotive Command, Research Report No. RR-27.

4. I. J. Sattinger, S. Sternick, "An Instrumentation System for the Measurement of Terrain Profile", University of Michigan Report No. 2948-36-F, December 1961.

5. "Operating and Maintenance Instructions for Simulation Display System", University of Michigan Report No. 2023-518-M, October 1955.

6. I. J. Sattinger, D. F. Smith, "Computer Simulation of Vehicle Motion in Three Dimensions", University of Michigan, May 1960.

7. R. Dean - Averns, "Automobile Chassis Design".

8. V. Kowachek, "Firing Stability", Technical Notes.

This page intentionally left blank.

SIZE EFFECTS IN THE MEASUREMENT OF SOIL STRENGTH PARAMETERS

Ben Hanamoto
Land Locomotion Laboratory
Army Tank and Automotive Center

and

Emil H. Jebe
Institute of Science and Technology
The University of Michigan

INTRODUCTION. The main concern of the Land Locomotion Laboratory, ATAC, is the relationship of a vehicle to the terrain over which it travels. Once an insight into this relationship has been attained, the problems encountered from the initial design of a vehicle to its ultimate use in the field can be more easily grasped and rationally solved. The solution of engineering problems depends upon the selection of the relevant variables and a description of the functional relationship among these variables. The selection of the vehicle variables are within broad limits at the disposal of the designer, but for the terrain or soil, the selection of suitable variables becomes much more complicated. Soil is probably one of the most complex of all engineering materials [1]. Researchers in soil mechanics have added much to the knowledge of the mechanical and physical characteristics of soils, but as yet, no fully satisfactory general theory is available.

Land locomotion is an engineering application of soil mechanics to off-road vehicular operation. Its objective is to determine the relationships of vehicles, or more precisely, of the wheel and track, to the strength properties of the soil. In land locomotion research, one of the important questions is the nature of the pressure-sinkage relationship. The Bekker equation

$$p = \left(\frac{k_c}{b} + k_\phi \right) z^n \tag{1}$$

represents a family of curves with three unknown constants, k_c, k_ϕ, and n, and two variables, p the pressure under a loaded area, and z the sinkage* [2].

---

* b is a known constant, the plate width. p is per unit area (i. e., square inch).

A set of these constants which will approximately describe pressure-sinkage observations can be obtained [3]. To determine these constants, pressure-sinkage experiments were performed in the laboratory with footings of various sizes and shapes. The constants obtained were used to predict the sinkages for other loaded areas. It was found that the predictions were adequate for tracked vehicles, where the relative shapes of the loaded areas were similar. Rectangular test footings with a length/width ratio greater than 5 were used for determining the soil parameters used in the prediction equations. The basic equation includes only the width term b. For cohesionless soils, the ultimate bearing strength is dependent on the width only for long loaded areas [4]. To determine what the minimum length had to be before a footing was not considered long, laboratory tests were conducted with footings of varying ℓ/b ratios. Acceptable results for the pressure-sinkage relationship were obtained when the ℓ/b ratio was greater than 5. Consequently, all pressure-sinkage measurements were taken with footings of at least an ℓ/b ratio of 5 or greater.

The ultimate equations in which these soil strength parameters are to be used apply to the general case of predicting vehicle performance for both tracked and wheeled vehicles. As noted above, the predictions of tracked vehicle sinkage and motion resistance have been generally satisfactory. For wheeled vehicles, however, improvements are needed. One of the differences to be noted between a tracked and wheeled vehicle is the shape of the loaded area. In most cases, tracked vehicles have a contact area of relatively long length as compared to width. Such a length-width ratio is not the usual situation for wheeled vehicles at moderate sinkages. Most tires will have a contact area where the ℓ/b ratio is close to 1 or 2.* Therefore, it was thought that the shape of the loaded area when ℓ/b was less than 5 might have effects on the pressure-sinkage relationship. A clearer understanding of the pressure-sinkage relationship in this region would provide us with an improved soil-vehicle model with broader and more useful applications. Consequently, a test program was undertaken by the Land Locomotion Laboratory to investigate further the pressure-sinkage relation.

*For example, the Army 6 x 6 5-ton truck normally carries an 11.00 x 20 tire. At one inch sinkage, ℓ/b - 1 and at eight inch sinkage ℓ/b = 2.33.

DESCRIPTION OF TEST PROGRAM. The test program to study the effect of plate size on the vertical soil strength or sinkage parameters was divided into two parts. The first part comprised a study of the reliability or repeatability of the test results for the equipment and mixing techniques that were to be used in the tests. The second part covered the measurement of the load-sinkage curves while using different sized plates.

In carrying out the experiments a laboratory model bevameter is used to drive a constant speed hydraulic piston arrangement which pushes a plate into the soil in a bin. The depth of sinkage and the force on a load cell are plotted electrically by an X-Y plotting device. The curve is traced on linear graph paper. From this graph values for p and z may be read off and plotted on log-log paper. The slope of the least square fitted line on the double-log plot gives an estimate of the parameter n. The constant term in such a fitted equation is the logarithm of

$$\frac{k_c}{b} + k_\phi$$

where b is the width of the plate used. Thus it is seen that k_c and k_ϕ cannot be estimated by any straight forward statistical technique. Use of two different b values, however, will permit setting up two simultaneous equations in k_c and k_ϕ.

PART I. We wished to determine the maximum number of penetrations that could be made with one preparation of the soil bin. For this purpose we set up a uniformity trial using only a 2" x 10" plate. Orientation and location of the penetrations was arranged as indicated in Figure 1. The three orientations shown were carried out in a randomized block design with six replicates.

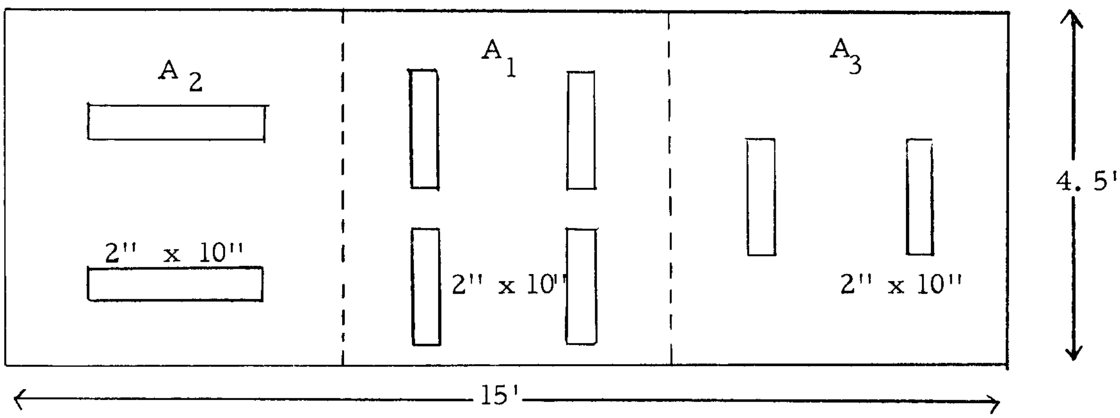


FIGURE I: SAMPLE BLOCK LAYOUT FOR UNIFORMITY TRIALS

RESULTS FOR PART I: Mean pressures were computed for the orientations at 1", 2" and 3" depths of sinkage. Mean differences for the orientations at a given depth were found to be homogeneous (i.e., not significant). At the 2" depth of sinkage, the coefficient of variation was about 6 percent, quite a satisfactory value. Examination of the variability within the orientations showed that the variation within the arrangement A_1 was significantly greater than that for the other arrangements.

CONCLUSIONS FOR PART I: The experimental procedure would yield results of adequate reliability. The orientation A_1 appeared undesirable for taking pressure-sinkage measurements. Therefore, we decided to make the spacings between determinations as similar to the A_2 arrangement as practicable. Since a rectangular plate of size 3" x 10" was to be used in Part 2, it appeared necessary to "beef up the apparatus" to handle the greater pressures required to sink such a large plate.

PART 2. The second phase of the test program comprised measurement of the load-sinkage relation in dry sand with plates of varying sizes. Analysis of test data would provide estimates of the three parameters in the Bekker equation and their variations for the sizes and shapes of plates used. The results should show the dependence, if any, of the parameters on the size and shape of plate.

Many problems arose in the consideration of the second part of the test program. One point was that all plates should be tried in one mix of the bin. Thus, a complete block design would be preferred.

A <u>second point</u> is that the sinkage equation is a stress-strain relationship. It may also be described as a functional relation [5]. The problems of estimation which arise for the functional relation have been resolved by Dr. Joseph Berkson by introducing the "control variable" concept [6]. It appears that z may be taken as the control variable in our problem. This approach is contrary to the usual dependent-independent variable point of view, but Berkson has shown that the method is unbiased for estimating the functional relation if the errors in p are unbiased.

<u>Thirdly,</u> it was clear that a statistical analysis of the estimation procedure for the Bekker equation was needed. With transformation to logarithms of p and z it is assumed that the standard linear regression assumptions are valid in the transform [7]. Thus, estimation of the parameter n is quite straightforward. When n has been obtained, the procedure takes $z = 1$, hence, $\log z = 0$, and predicts a value of $\log p$, say P_o. Now, anti-log $P_o = p^* = k_\phi + k_c/b$. By taking two values of the plate width, b_1 and b_2, and corresponding p_1^* and p_2^* values, the estimation equations for k_c and k_ϕ become

(2) $$k_c = \frac{(p_1^* - p_2^*)b_1b_2}{b_2 - b_1}$$

and

(3) $$k_\phi = \frac{b_2p_2^* - b_1p_1^*}{b_2 - b_1}$$

These results pointed out two things:

1. Widely spaced b values to permit use of large $b_2 - b_1$ would reduce the variance of the estimates of k_c and k_ϕ,

2. A formula for the variance of p^* is needed.

The variance formula for p^* presents some difficulty because p^* is a nonlinear function of P_o. We recall, however, that the choice of an appropriate experimental design will give us direct estimates of the experimental error for our estimated k_c and k_ϕ values so that we can bypass the variance formula problem.

DESIGN OF THE PART 2 EXPERIMENTS. From the statistical analysis it was clear that the largest possible difference in plate widths should be used to estimate the parameters k_c and k_ϕ. This consideration along with the desire for a complete block experiment already mentioned resulted in a revision of the choice of dimensions for the rectangular plates to be used in the experiment. The plate sizes actually selected are given in Table 1.

TABLE 1
LIST OF PLATE DIMENSIONS FOR PART 2 EXPERIMENTS

(in inches)

Rectangles:			
	1 x 4	2 x 4	3 x 4
	1 x 6	2 x 6	3 x 6
	1 x 8	2 x 8	3 x 8
	1 x 10	2 x 10	3 x 10
Circles:			
Diameters:	2 and 4.		

The entire set of plates thus provided 14 treatments to be set up in a completely randomized block experiment where one block equals a mix of the soil bin. Each plate was to be used in a randomly selected plot of size about 2' x 2' to make a single measurement of the pressure-sinkage curve. It was further decided to complete six replicates which would yield 24 independent pairs of estimates of k_c and k_ϕ, but would, of course, yield 84 estimates of the parameter n, one estimate being obtained from each pressure-sinkage curve.

ANALYSIS OF RESULTS FOR PART 2. Analysis of the estimated values for n is straightforward and results are readily interpreted. For the k_c and k_ϕ values some difficulties arise. Plotting the means is a most useful device for aid in understanding the effects indicated by the analysis.

Figures 2 and 3 show the width and length effects on n without considering the interaction. In order to present the interaction effect, we show the usual type of plot for displaying an interaction. Parallel lines for the various lengths of plate would be indicative of no interaction.

Thus, the lack of parallelism exhibited in Figure 4 is indicative of the nature and source of the interaction. The major pattern, however, is still that shown in Figures 2 and 3. There appears to be a decrease in the n values with an increase in area of plate whether the area increase is due to change in length or change in width. The width effect is much smaller in going from 2" to 3", but the length effect is about the same at all widths.

In considering the analysis of the estimated k_c and k_ϕ values it will be useful to recall equations (2) and (3) which show how the estimates are obtained. There are some obvious points to be noted from these equations, but we shall defer them until later. Each experimental unit or plot in the soil bin yields a p versus z relation as drawn by the x, y plotter while a single plate is sunk into the soil. Then results from two different plate widths have to be combined to obtain a single estimate of k_c or k_ϕ . We may combine 1" and 2" widths or widths of 2" and 3" or 1" and 3". As shown above, the latter is the best choice, but we see that a single replicate or "set" will only give us four such independent estimates.

By keeping the width difference constant and varying the length we can pair within one set these four pairs of plates:

1 x 4	and	3 x 4	1 x 8	and	3 x 8
1 x 6	and	3 x 6	1 x 10	and	3 x 10

to give us four pairs of estimates for k_c and k_ϕ . Utilizing the six "sets" gave us 24 values for analysis.* Estimates obtained from these pairings may be analyzed for the length effect alone. But we would also like to study the width of plate effect on these two parameters and look for interaction, if any, as we did for the parameter n.

First, we consider the estimates of k_c. The mean values obtained for the pairs just listed were:

---

*A set consists of six replicates carried out at the same depth. Eighteen replicates were actually completed at each of two depths. Aggregation over six replicates forms a set. Thus, there are a total of six sets.

This page intentionally left blank.

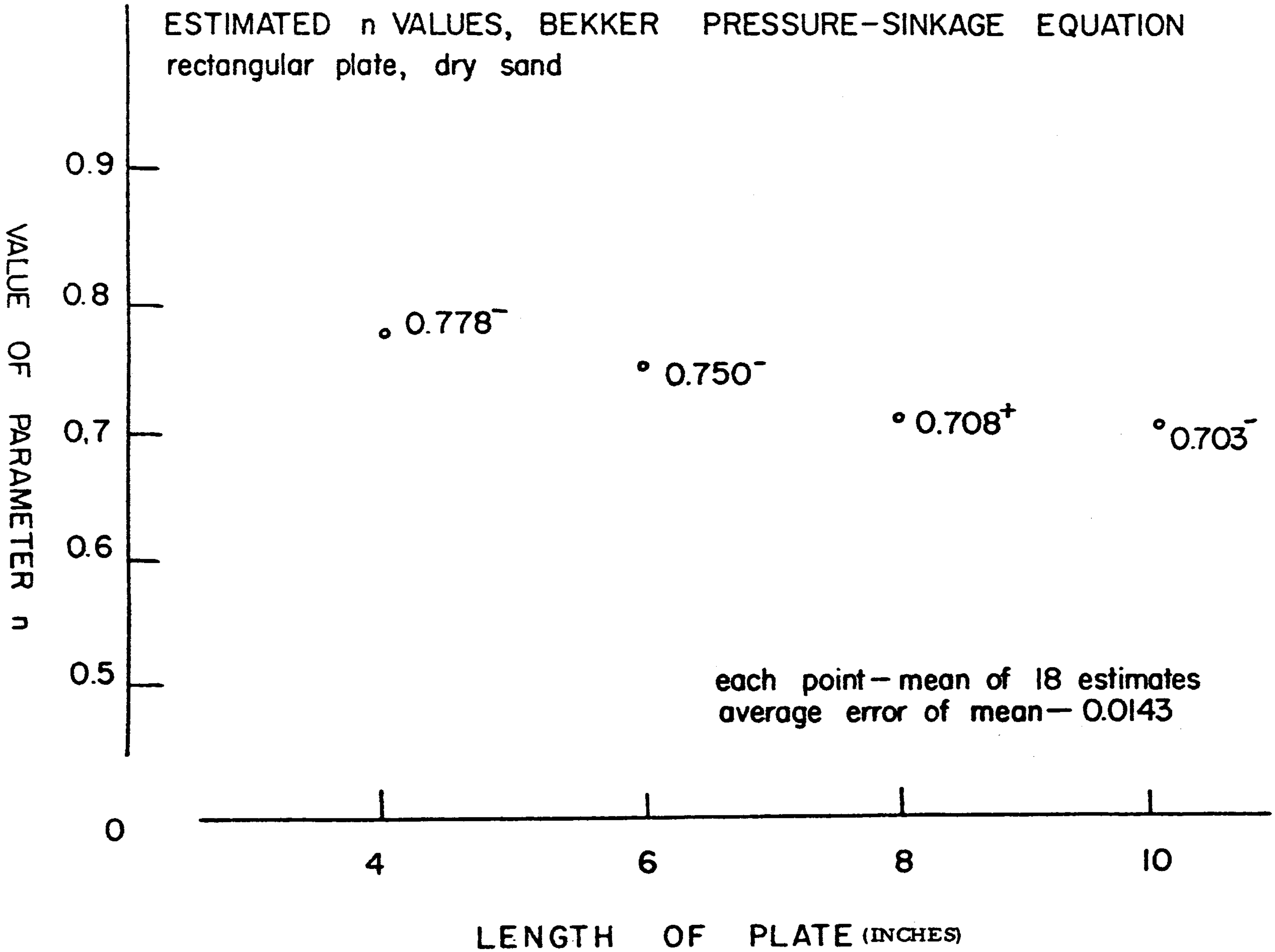

ESTIMATED n VALUES, BEKKER PRESSURE-SINKAGE EQUATION
rectangular plate, dry sand
VALUE OF PARAMETER n
0.9
0.8
0.7
0.6
0.5
0
0.778⁻
0.750⁻
0.708⁺
0.703⁻
each point – mean of 18 estimates
average error of mean – 0.0143
4
6
8
10
LENGTH OF PLATE (INCHES)

Figure 2

This page intentionally left blank.

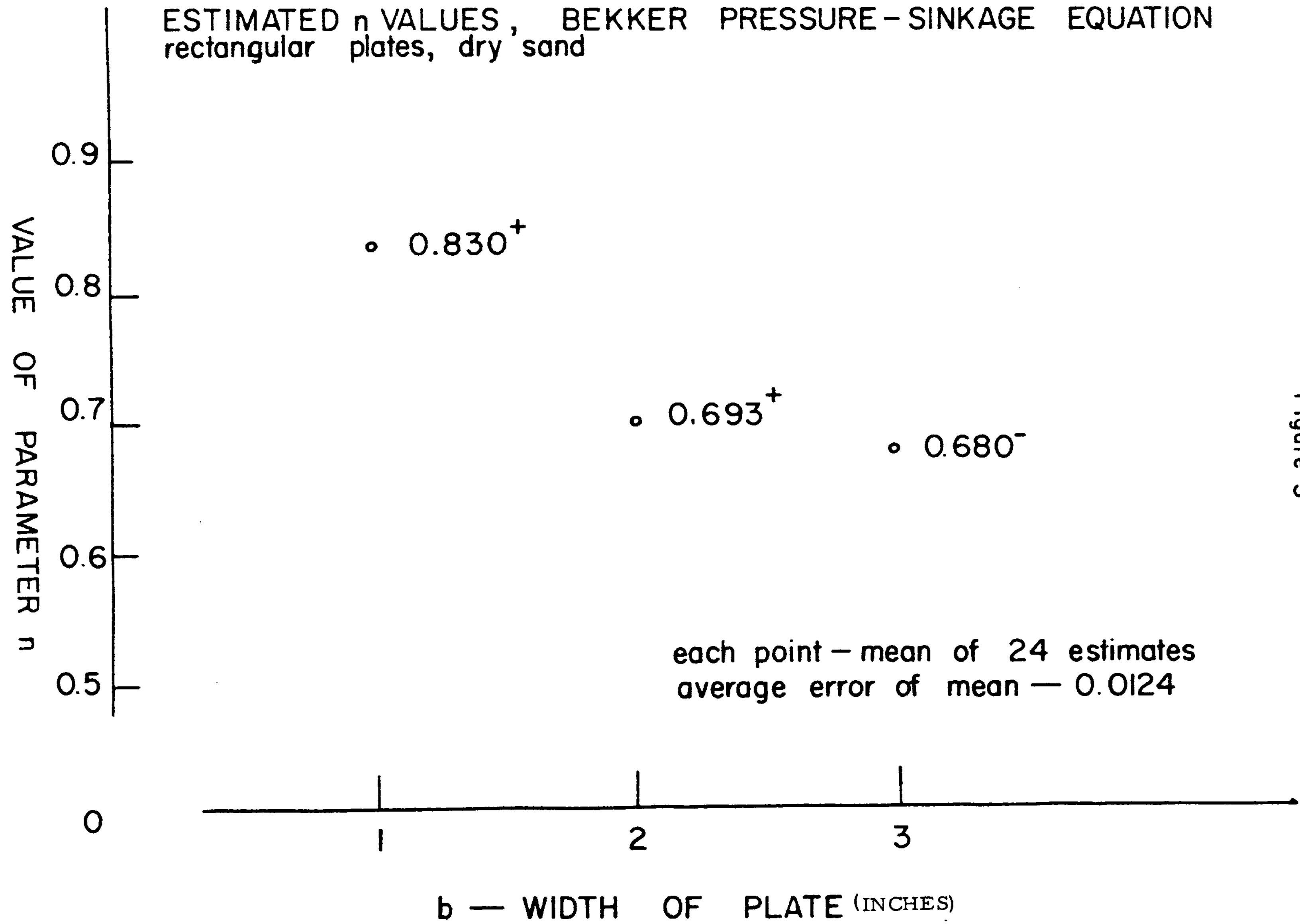

ESTIMATED n VALUES, BEKKER PRESSURE-SINKAGE EQUATION
rectangular plates, dry sand
0.9
0.8
0.7
0.6
0.5
0
VALUE OF PARAMETER n
0.830+
0.693+
0.680-
each point – mean of 24 estimates
average error of mean — 0.0124
1
2
3
b — WIDTH OF PLATE (INCHES)

Figure 3

This page intentionally left blank.

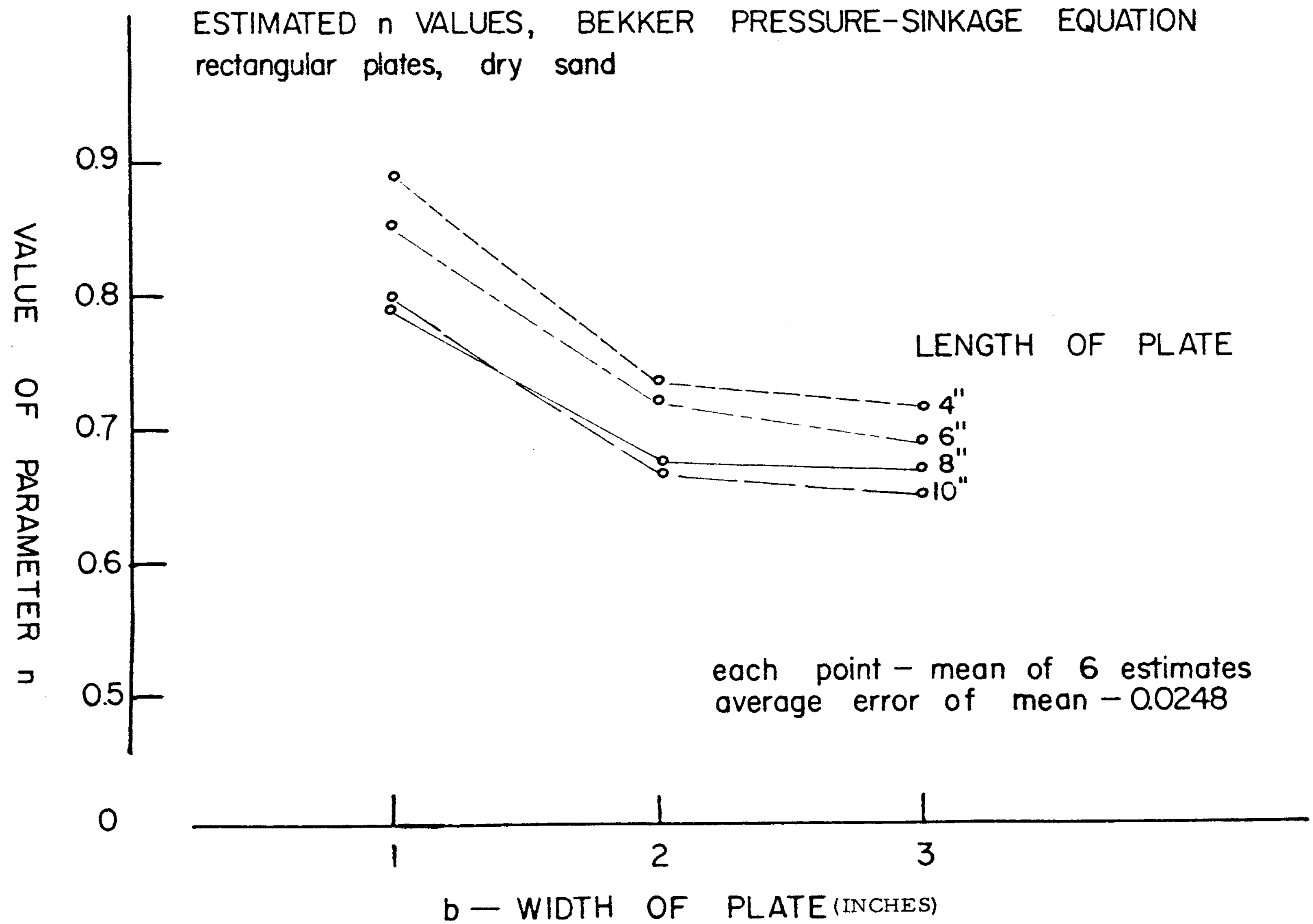

ESTIMATED n VALUES, BEKKER PRESSURE-SINKAGE EQUATION
rectangular plates, dry sand
VALUE OF PARAMETER n
0.9
0.8
0.7
0.6
0.5
0
LENGTH OF PLATE
4"
6"
8"
10"
each point – mean of 6 estimates
average error of mean – 0.0248
1
2
3
b — WIDTH OF PLATE (INCHES)

Figure 4

This page intentionally left blank.

Pairs:

1 x 4	1 x 6	1 x 8	1 x 10
and	and	and	and
3 x 4	3 x 6	3 x 8	3 x 10

Mean Value of k_c

-3.305	-4.030	-4.000	-4.166

Relevant mean squares abstracted from the analysis of variance [8] for comparing these means are:

Source	d. f.	M. S.
Length of Plate	3	0.8977
Length of Plate by Depth	3	1.0526
Length of Plate by Sets within Depth	12	0.2262

Since the $P(F_{3,12} \geq 3.89) = 0.05$, we may conclude that there is a length effect on k_c and that the length effect is not uniform at the two depths studied (i.e., $F = 3.97 = 0.8977 0.2262$).

To extend our analyses other plate pairings were studied. Finally, it seemed in our judgment that the following pairs were most useful for securing information on the width effect plus a little more length information:

Pairs:

1 x 4	1 x 6	2 x 4	2 x 6
and	and	and	and
2 x 8	2 x 10	3 x 8	3 x 10

Mean Value of k_c

-2.463	-2.495	-5.169	-5.063

Note that the pairs selected show a constant length difference of 4 inches. Furthermore, the length differences are balanced for comparing 1 and 2 inch widths with 2 and 3 inch widths for the study of the plate width effect in estimating k_c. It appears that the latter effect is large. Relevant mean squares from the analysis of variance are as follows:

Source	d. f.	M. S.
Width Pairs	1	41.717
Width Pairs by Depth	1	7.183
Length Differences within Width Pairs		
Length in 1 and 2	1	0.003
Length in 2 and 3	1	0.034
Length Difference by Depth		
in 1 and 2	1	0.006
in 2 and 3	1	5.914
Experimental Error	12	1.529

For comparing the width pairs we obtain an F ratio $= 41.717/1.529 > 27$. We conclude that the parameter k_c depends on the choice of plate widths. In estimating k_c, pairing of 1 and 2 inch widths gives a larger value (algebraically) than pairing 2 and 3 inch widths. Looking at it another way the data show that changing the length/width ratio for the plates from the range 4 to 6 to a range of 2 to 3 materially alters the k_c value obtained.

Similar analyses for the estimates of the parameter k_ϕ were carried out as just described for k_c. We note that the k_ϕ estimates were positive, mostly in the range +7 to +10. Our conclusions about the effect of plate width on the parameter k_ϕ were the same as reached in regard to k_c, but no length effect was detected.

With these results available we may return to the further consideration of the method used to obtain the estimates. First, we observe that if there were no width of plate effect in the p versus z relation, then k_ϕ would be equal to, say, p^*. In the notation used above, no width of plate effect would mean $P_{01} = P_{03}$, if one inch and three inch plate widths were used, and, hence, $p_1^* = p_3^*$ = a common value, p^*. Observed p_1^* and p_3^* values would differ, of course, due to experimental variation, but average values would be equal. Further, when $k_\phi = p^*$, then the solution for k_c is zero. The parameter k_c was once regarded as a measure of cohesiveness of the soil and, hence, might be zero for sand [3]. The present experiments, in which mason's sand was used certainly do not agree with this point of view about the parameter k_c.

Recognizing then that the present model for the p versus z relation does admit a width of plate effect our analyses so far have indicated the magnitude of this effect. In addition, we have obtained some indication of a length effect on k_c although none appeared for k_ϕ.

Second, with respect to the method of obtaining the estimates of k_c and k_ϕ, we point out that the estimates obtained are correlated. This correlation cannot be avoided because the values are based on the solution of two simultaneous equations. In fact, the estimates also would have been correlated if we had been able to obtain them by a direct least squares procedure. In the least squares case, however, it is usually easy to write down the covariance of the estimates and, hence, to find the correlation if desired. For the two simultaneous equations in k_c and k_ϕ, matrix methods can be applied to find the covariance matrix for k_c and k_ϕ. The result obtained for the covariance of k_c and k_ϕ is

(4) $$\mathrm{cov}(k_c, k_\phi) = (b_2 - b_1)^{-2}\,(b_1^2 b_2\, V(p_1^*) + b_1 b_2^2 V(p_2^*))$$

where b_2 is the width of the larger plate and V( ) denotes variance of the enclosed quantity.

It is interesting to examine the consequences for different choices of plate widths in estimating k_c and k_ϕ. As a first approximation it may be adequate to assume $V(p_1^*) \cong V(p_2^*)$. If this assumption is made, the following correlation values are obtained:

Plate widths paired: (widths in inches)		
1 and 3	1 and 2	2 and 3
Value of correlation of k_c and k_ϕ		
-0.8944	-0.9487	-0.9805

Observed values of the correlation between k_c and k_ϕ will, of course, differ from these theoretical values because of experimental variation and lack of equivalence of $V(p_1^*)$ and $V(p_2^*)$. It is clear, however, that the above results further support the use of maximum difference in plate widths for estimation of these parameters. Actual plots of the k_c and k_ϕ pairs support the high correlation of these estimates.

These considerations and the analyses for k_c and k_ϕ made it seem desirable to return to an analysis of the 84 p* values which correspond to the 84 n values analyzed earlier. Such an analysis should give a clearer assessment of the length, width, and interaction effects than we have obtained from analysis of the k_c and k_ϕ values. Any effects found in this analysis must then be present in the k_c and k_ϕ values because of the method of derivation of these estimates from the p* values.

The means for the p* values are presented in Table 2. Table 3 following gives the analysis of variance for the p* values. From the table of means it is seen that the width-of-plate effect is much larger than the length-of-plate effect. Both effects, however, may be judged significant from the analysis of variance results. It is interesting to note that the length by width interaction mean square is so small; no interaction is indicated. This result is in contrast to the analysis of the n values, Table 4, for which the interaction effect was judged significant.

CONCLUSIONS FOR PART 2. A large series of experiments have been conducted to study the pressure versus sinkage relation and to estimate the parameters in this relation. The experiments were carried out using a relatively homogeneous soil material, dry mason's sand. Bearing plates used comprised two circles of 2 and 4 inch diameters and 12 rectangular plates varying in size from 1 x 4 to 3 x 10 inches (refer Table 1 for list of plate sizes.) Thus, the length over width ratio of the plates ranged from a maximum of 10 to 1 down to 4 to 3, or from long narrow plates to almost square plates. This range of length over width ratios was selected to cover the range from tracked vehicles to wheeled vehicles with tires.

From the estimates of k_c, k_ϕ, and n plus the estimated pressures for one inch sinkage our analyses show that:

(1) The parameter n varies with the width and length of plate. For one inch width of plate the n value was 0.83; at three inch width, the value was 0.68. Between 2" and 3" width there was little change. With length, n varied from 0.78 to 0.70 over lengths of 4" and 10". The decrease was nearly uniform over the 6" interval.

(2) The estimated pressures for one inch of sinkage (the p* values) show variation with both length and width of plate but no interaction of the factors is indicated. (Refer to Tables 2 and 3.)

(3) The parameter k_c decreases algebraically with increase in plate length. The algebraic change, however, was much greater when estimates were compared from pairing of plates of 1" and 2" widths with estimates obtained by pairing 2" and 3" widths.

(4) The parameter k_ϕ showed little or no response to length of plate but a large response to width of plate.

(5) The estimates of k_c and k_ϕ are highly correlated. This correlation is negative so that when k_c increases in value k_ϕ decreases in value.

From these analyses it appears that the p versus z relation in the general form $p = (k_\phi + k_c/b)z^n$ is inadequate to predict the pressure-sinkage response. Although not pointed out previously in this paper, it should be mentioned that our analyses to date are based only on the experimental results for sinkage in the range 0.6" to 2". Perhaps it should be added that Dr. Bekker has not claimed that his equation would be adequate in the entire ℓ/b region we have studied.

FUTURE WORK. While much has been learned, there is clearly need for the following:

a. Similar laboratory experiments in other soil media.
b. Experiments with greater depth differences in the soil bins to assess the depth effect, if any, on the parameters under homogeneous soil conditions.
c. Revision of the model to take account of the dimensions of the bearing surface.
d. Improvement of the model to cover a wider range in depth of sinkage, say, from at least 0.5 to 5.0 inches.

Table 2

TABLE OF MEANS FOR p^* VALUES

(estimated pressure for one inch sinkage) summarized by plate dimensions (inches)

SIZE OF PLATE	p^*	COMBINED MEANS	p^*
Circles		all Circles	6.865
2" diameter	5.509		
4" "	8.222		
Rectangles			
1 x 4	5.188	Lengths (over all Widths)	
1 x 6	4.889	4	6.429
1 x 8	4.701	6	6.378
1 x 10	4.737	8	6.296
2 x 4	6.706	10	6.141
2 x 6	6.671		
2 x 8	6.420	Widths (over all Lengths)	
x 10	6.170	1	4.928
3 x 4	7.392	2	6.492
3 x 6	7.576	3	7.512
3 x 8	7.568		
3 x 10	7.515		

Table 3

ANALYSIS OF VARIANCE OF p^* VALUES
(estimated pressure for one inch sinkage)

SOURCE OF VARIATION	DEGREES OF FREEDOM	MEAN SQUARE
Plates	(13)	8.4927
Circles vs Rectangles	1	2.5279
Between Circles	1	22.0784
Among Rectangles	(11)	7.6181
Widths	2	40.9546
Lengths	3	0.4964
Length by Width	6	0.0668
Plates by Depth	(13)	0.2290
Experimental error	52	0.1793

Table 4

ANALYSIS OF VARIANCE OF n VALUES

(estimated from six replicates grouped together for fitting the pressure-sinkage equation in double-log form)

SOURCE OF VARIATION	Degrees of FREEDOM	MEAN SQUARE
Plates (13)		
Circles vs Rectangles	1	0.0112
Among Circles	1	0.1248
Among Rectangles (11)		
Widths	2	0.1113
Lengths	3	0.0228
Lengths by Widths	6	0.0193
Plates by Depth	13	0.0010
Experimental Error	52	0.0037

REFERENCES

1. Karafiath, L., "Classification of Soils and the Significance of Their Parameters in the Theory of Land Locomotion," Research Report No. 5, 1958, Land Locomotion Research Laboratory, OTAC, Centerline, Mich.

2. Bekker, M. G., "A Practical Outline of the Mechanics of Automotive Land Locomotion," 1955, Land Locomotion Research Laboratory, Detroit Arsenal, Centerline, Mich.

3. Bekker, M. G., "Theory of Land Locomotion, The Mechanics of Vehicle Mobility," 1956, The University of Michigan Press, Ann Arbor, Mich.

4. Taylor, D. W., Fundamentals of Soil Mechanics, John Wiley and Sons, Inc., New York, 1948.

5. Lindley, D. V., "Regression Lines and the Linear Functional Relationship," Jour. of the Royal Stat. Soc., Supplement, 9, 218-244, 1947.

6. Berkson, Joseph, "Are There Two Regressions," Jour. Amer. Stat. Assn., 45, 164-180, 1950.

7. Mood, A. M., Introduction to the Theory of Statistics, 1950, McGraw-Hill, New York.

8. Scheffe, H., The Analysis of Variance, 1959, J. Wiley and Sons, Inc., New York.

EFFECTIVENESS OF CERTAIN EXPERIMENTAL PLANS UTILIZED IN SENSORY EVALUATIONS

J. Wayne Hamman and Jan Eindhoven
Armed Forces Food and Container Institute
Chicago, Illinois

First, I would like to present some specific purposes of sensory testing at the Armed Forces Food and Container Institute. This will be followed by a discussion of the experimental results obtained from the sensory evaluation of four meat products.

SOME PURPOSES OF SENSORY TESTING. You are aware of the numerous food items developed, purchased, stored and consumed by the Armed Forces. A continuous program exists at the Institute to determine whether or not differences in quality or stability exist between different samples of food. Here are some of the most common requirements for conducting these sensory tests:

1. Pre-award evaluation for intent to purchase. When a certain food item (such as peanut butter) is required by the Army, it advertises for bids from manufacturers. Those manufacturers who are interested submit samples of their products for preference evaluation. These samples are taste-tested, and those that are reliably poorer than our standard products are rejected. In this way, sensory testing screens out lower quality products that are relatively unacceptable to the soldier-consumer.

2. Storage stability. Since foods may not be used for several years after they are packed, a considerable amount of research is devoted to extend the shelf-life of a food. Sensory tests are concerned with the preference or intensity of off-flavor changes that take place during storage. Sensory tests are made on foods stored at different temperatures over time up to two years, and more.

3. Packaging studies. Often, a flexible package may be desired for use in the field. However, the relative storage life of food in such a plastic package must be considered if a change is to be made from a canned food.

4. Processing variables. New processing and preservation methods of foods, such as freeze-dehydration of meats, offer new problems in flavor and texture for evaluation. It must be determined whether or not this new product is as desirable as the existing food prepared by other methods.

5. Special Purpose Foods. Recent evaluations have included novel preparations of foods designed specifically for space flight. For example, meat dishes that may be consumed through a straw. Also the new Quick-Serve Meals have been developed which consist largely of pre-cooked dehydrated foods. Preliminary testing is done first at the Institute to determine whether or not these foods are satisfactory enough for further testing in the field among astronauts and soldiers.

SENSORY EVALUATION LABORATORY. In the sensory evaluation laboratory careful attention is given to assure that each sample of food is treated in the same way as every other one in an evaluation. Some of the procedures followed include

1. The random assignment of code numbers to the samples so that subjects will not be biased.

2. In order for the individual to regain sensitivity, that is to get the flavors of a previous sample out of his system, a 30-second time interval is specified between the time that a subject returns the rating of his previous sample and when he receives his next one. Automatic timers are used.

3. In a sensory test each sample is served first, second, third, etc., an equal number of times to minimize position effect. When the number of subjects permits, all possible serving sequences of samples are used to reduce both serving position and sequence biases that might exist.

4. The number of samples that a subject receives is normally limited to four in order to minimize effects of fatigue and to maintain interest in the evaluation.

PURPOSE OF EXPERIMENT. This experiment is concerned with the effect on sensory results when meat samples are presented to subjects in different combinations and sequences. Specific topics considered are

1. Sequence effects. How is the rating of a sample influenced by the quality of a preceding sample? It is hypothesized that when more highly preferred samples precede those of relatively low preference, the difference between them is emphasized. It is further hypothesized that when the more highly preferred samples follow those of relatively low preference, the difference between them is reduced.

2. Position effects. How is the rating of a sample influenced by the number of preceding samples he has evaluated?

3. Matnitude of error term. How is the size of the error term affected by the quality and sequences of the samples presented?

MATERIALS AND METHODS. Four meat products were evaluated in this experiment: ham, pork, chicken (white) and chicken (dark). Four samples of each of the meat products consisted of two control and two treated samples. An additional preparation variable was included for each meat product which causes both the two control samples and the two treated samples to be considered as non-duplicates. However, the determination of the effect of the additional variable is not the intent of this experiment and will not be specifically considered in this paper.

The subjects sat in a semi-enclosed testing booth for privacy in making evaluations. Each individual received four samples of one of these meat products. These samples were presented one at a time through a turn-table in a wall separating the booth from the serving and preparation area. The subject was asked to state his preference for each sample on a nine-point rating scale. The terminology on this hedonic scale(5) ranged from dislike extremely, coded 1, to like extremely, coded 9, and is shown on the illustrated EAM card (Figure 1) [Figures and Tables can be found at the end of this article] which is used for rating and mechanical data reduction.

Subjects were selected at random from a pool of about 450 employees.

EXPERIMENTAL PLANS. Five experimental plans are considered:

(1) 4! : Conventional plan with all sequences of serving orders. Twenty-four subjects are required for this plan in order to encompass all sequences.

(2) cccc: Two control samples balanced over four serving positions. The two control samples, you will recall, are the non-treated samples and differ by a preparation variable.

(3) tttt: Two treated samples balanced over four serving positions.

(4) cctt: Two control samples followed by two treated samples. The two control samples were served equally often in positions 1 and 2; the two corresponding treated samples were served equally often in positions 3 and 4.

(5) ttcc: Two treated samples followed by two control samples. The two treated samples were served equally often in positions 1 and 2; the two corresponding control samples were served equally often in positions 3 and 4.

Twenty-four subjects were selected at random for each of these plans; all plans were carried out for each of the four products.

EXPERIMENTAL DESIGN. A latin square experimental design[1] was utilized in this study in order to

a. determine the effect, if any, of the sequence of the presentation on the rating for a sample, and

b. reduce the experimental error, if a position effect was present.

Six replications of a 4x4 latin square design were utilized in each plan.

Figure 2 illustrates the allocation of samples for the 4! conventional plan which has all possible sequences. Subjects 1, 2, 3, 4, constitute the first replication, then, and the four samples A, B, C, D, all occur once in each order in a replicate and a subject receives all four samples. The basic Analysis of Variance components, before isolating certain individual degrees of freedom is also given as a part of Figure 2. Since the subjects in a replicate were selected at random, no difference was anticipated between replicates. In the Analysis of Variance treatment x replication and order x replication interactions were pooled into the error term, since there is no reason to expect that these interactions are real.

SEQUENCE RESULTS. Results for the conventional plan (4!) which had all serving orders are shown in Table 1. Mean preference scores for controls were higher than treated for all four of the meat products. The mean differences ranged from 0.56 to 0.79 scale points and significance values were no larger than $P = .06$ in testing the null hypothesis, namely, that the control mean and treated mean are the same.

Table 2, which presents results for the plan with two control samples served first followed by two treated samples, shows a substantial increase in the discrimination between control and treated samples. The combined mean difference increased from 0.69 for the conventional plan to 1.00 for this cctt plan and individual probability values declined with the exception of pork which remained about the same (P = .01 vs .02). This phenomenon has been described in previous studies.

In a study with soups and beverages [4] the situation of poor samples following good ones was termed "contrast", that is, one of emphasizing differences. An explanation hypothesized for the phenomenon of "contrast" was that the positive qualities of the good sample are either noticed to be absent or bad qualities are noticed as present in the poor one, thus emphasizing in either case the short-comings of the less preferred one.

Results of the alternative situation where relatively poor samples precede the good ones are given in Table 3. You will notice that the direction of the differences determined in the conventional 4! and contrast plans are not found here. The combined treated samples were rated 0.06 scale points higher than the control in this ttcc plan and none of the individual product differences were statistically significant. This situation where poor samples preceded good ones was termed "convergence"[4], although with these liquids, convergence effects were not found to be statistically significant. An explanation hypothesized is that the presentation of a "poor" sample increases an individual's awareness of the presence of some of the negative characteristics in a "good" sample[4].

Conclusions drawn from these results might be modified somewhat, after a consideration of the position of presentation. We might ask the question: what part of the observed contrast and convergence effects might be due to the fact that samples were presented in positions 3 and 4? We will now proceed to an examination of the positional effect of presentation.

POSITIONAL RESULTS. An examination of the effect of the order in which the sample was received on its rating has been made, considering all five plans. Mean scores by position are given in Table 4. Combined positional means are given in the lower part of this Table for each type of plan. Hypothesis tested were