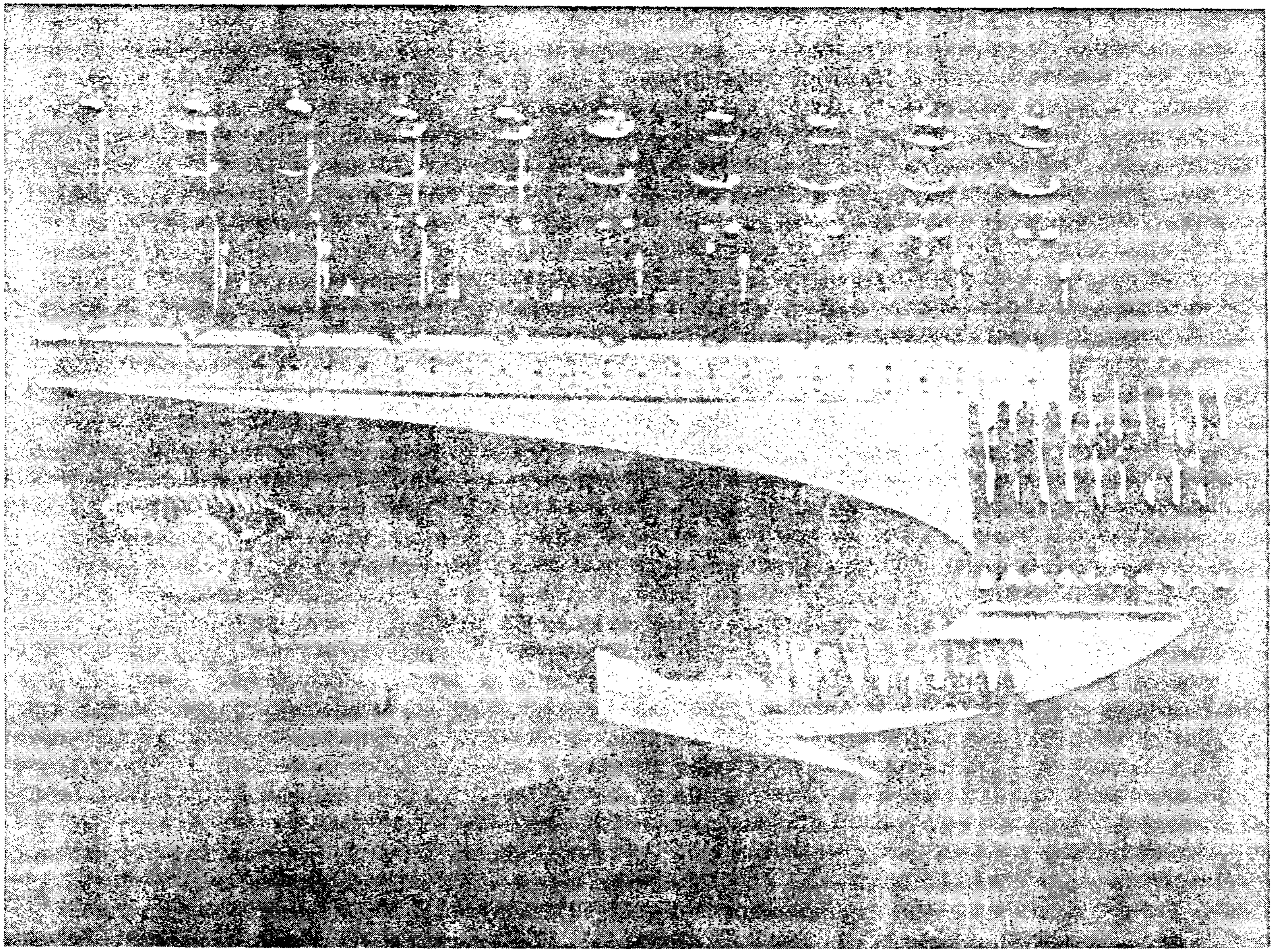
Slide 27

Binary Computer



Slide 27

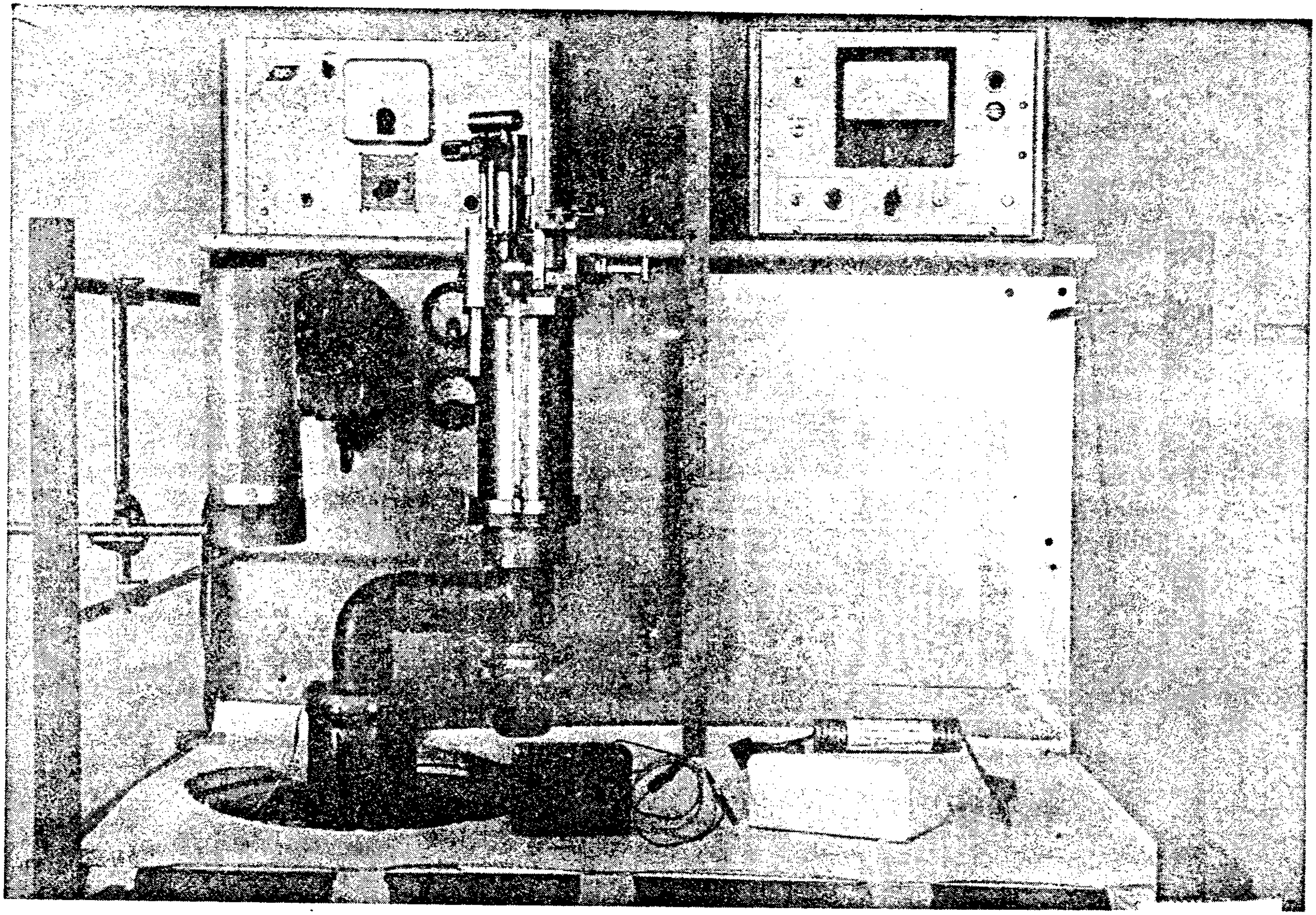
The following slide shows the progress over a number of years - essentially a decade - in a computer. This little "J" down on the front patch represents the item that now does what all of the other ones preceding it did in the past.

I have given you a brief view of the contributions in communications and electronics. I think that one can say equally well that major advances have occurred in radar, in television, and certainly in many fields of science. Much has come from the work being done by the Defense Department, of which the Army is a member.

Certainly one very important field is that of the MASER.

Slide 28

Gaseous Maser



Slide 28

This is a gaseous maser. It came about during the process of studying the structure of hydrogen. For those who might not know, "maser" means "microwave amplification and stimulation of emission radiation." The MASER is an extremely low-noise device and has now made possible developments in radar and communications which are rather phenomenal.

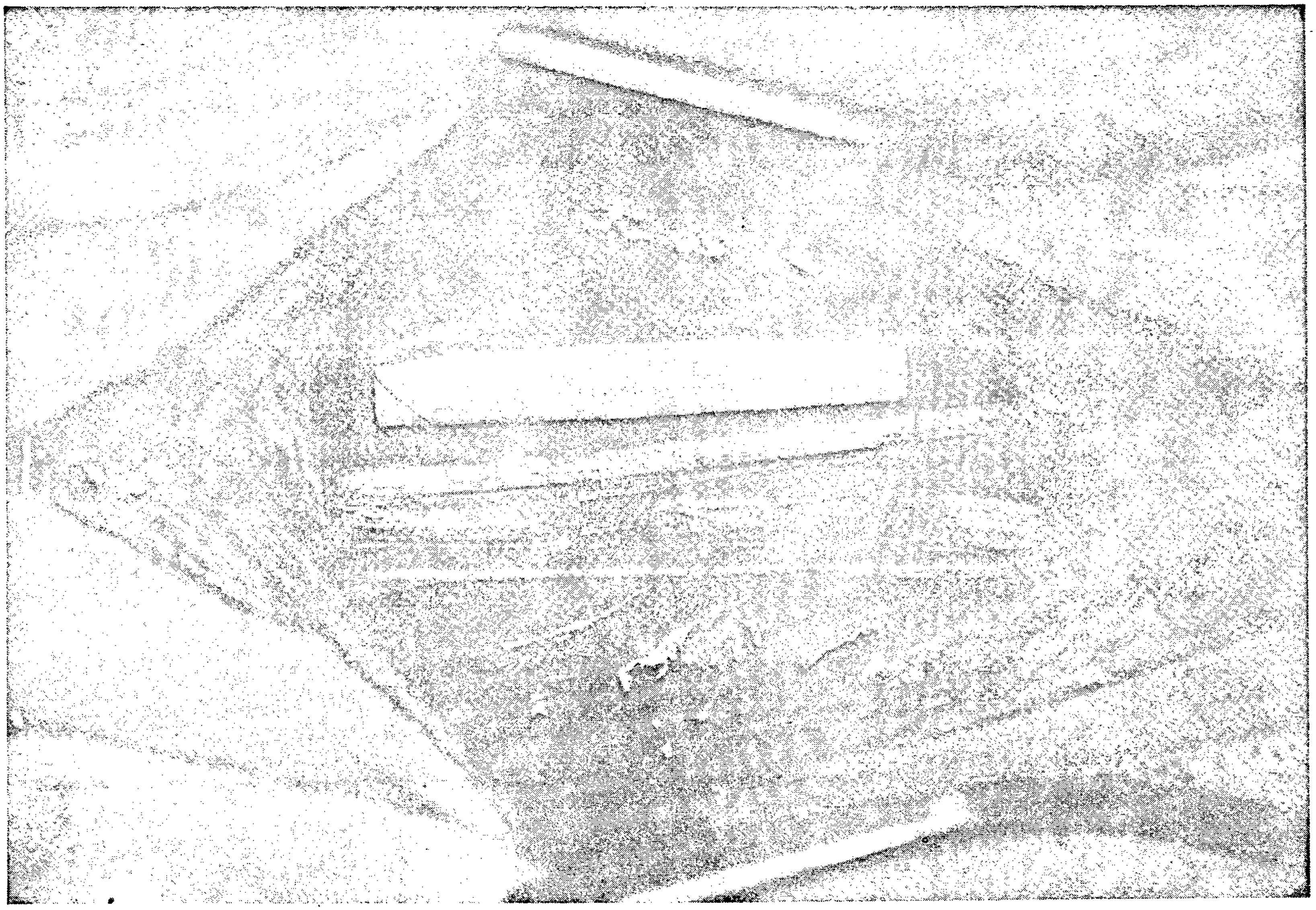
The signal to noise ratio is so high that the maser promises to reduce power requirement from 10 megowatt to 30 kilowatts in a radar application, or reduce antenna size from a 250 ft dish to a 60 ft dish; or increase the range from 3,000 miles to 12,000 miles; or reduce the target cross-section from 150 sq in to about a half sq in; or reduce false alarm rates from one per day to one per nine months. These are all tied in with the fact that the noise in these oscillators or amplifiers is extremely low. Used as a frequency device, the gaseous maser can yeild precisions in the order of one part in 10^9, or reflected in something more popular, would maintain time constant to within one second in 300 years.

And now a few words about the Medical Corps. As you know, they have had a long and rather eminent career in the field of medical research, starting with Walter Reed. They are concerned with medical problems wherever our own American soldiers are. Mass immunization methods are being worked on, as well as yellow fever control. You would be interested to know that when the Asian flu hit this country the Army medical research units had already isolated the virus some three years previously and had determined what were the necessary anti-toxins that would be used to check it. The results of this work were applied to checking the virus when the country was exposed to it.

Important work is being done in the field of nerve repair, using monomolecular films of millipore (and I think this probably means many, many pores). They actually have been able to surround a severed spinal nerve or an optic nerve in an animal with this particular material. There are holes it in which enable nutrients to penetrate through and feed the growing nerves, and actually make it possible for the severed nerves to bridge gaps of the order of several millometers and unite with their proper partners. It has resulted in considerable decrease in scar tissue and should have exceptional success in the repair of many severed nerves.

Slide 29

Nerve Repair



Slide 29

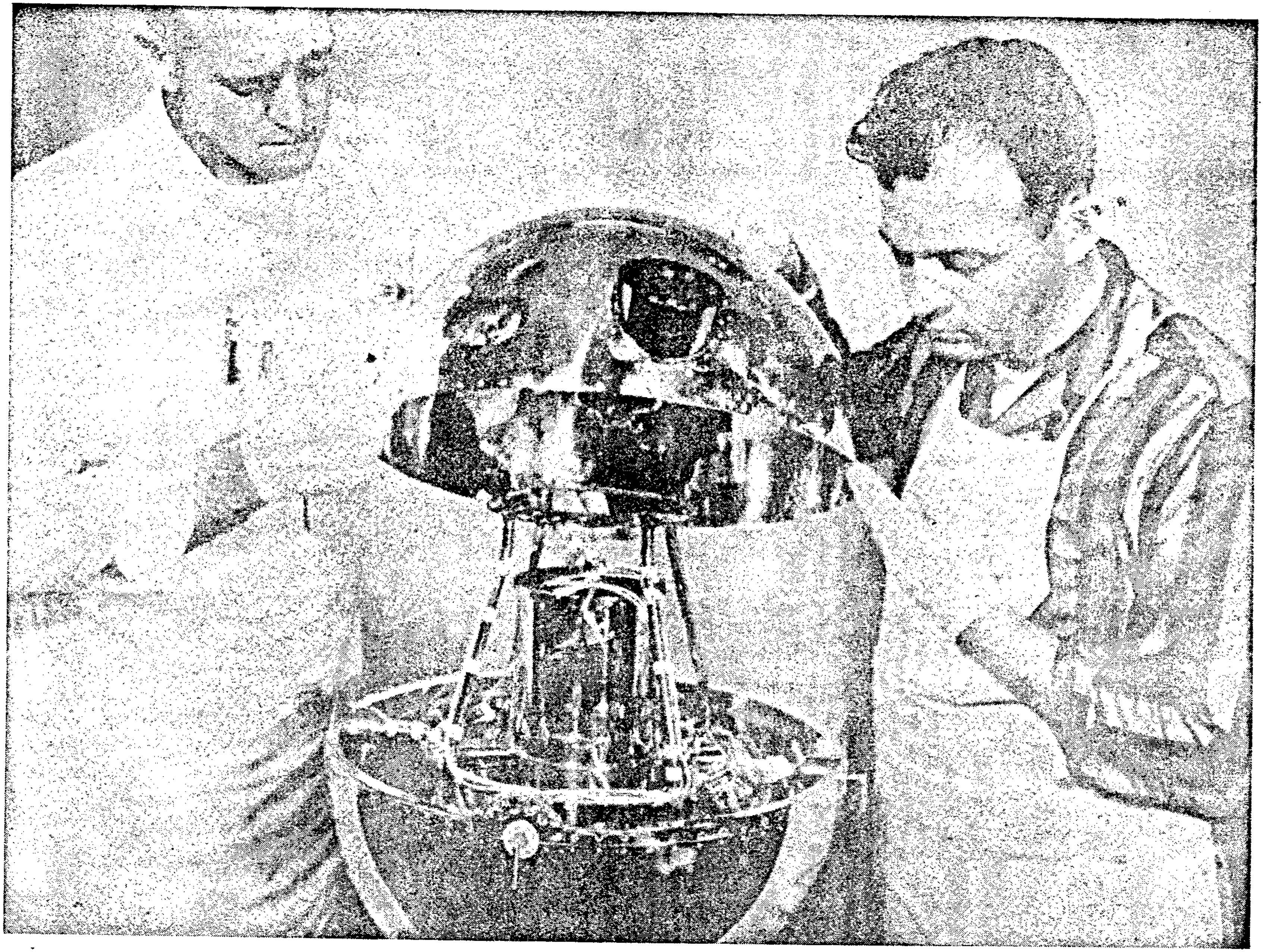
This slide shows a picture of that. There you see the severed nerve, and that white material will go around this part and will stay there. The wound will be sewed up and in a short period of time the nerve fibers will reunite.

There is a counterpart to this. This has to do with a new type of bone glue which has been developed by the Walter Reed Army Institute of Research in conjunction with the Hahnaman Medical College and Hospital. This is a polyurethane foam. Actually what happens is when a bone is broken, the break is essentially set, a two-bladed circular saw separated by a proper distance is used to saw out two channels and a piece of bone is taken out of the injured bone, the polyurethane foam is packed into the space, and then the bone is put back in place. In several minutes the glue is hard enough so that it can be chiseled away with a hammer and after sewing up the wound it is possible, within a period of 48 hours, for the patient to walk away. This has been done. So, you see if we combine the gigantic stapling machines which the Russians have developed with this, we have a real good do-it-yourself technique of repair.

In several general areas you know the Army's interest. In the field of infra-red, control instrumentation and photography, much work has been done. Weather prediction comes in for a rather major share of Army research.

Slide 30

Weather satellite



Slide 30

The next slide shows a weather satellite.

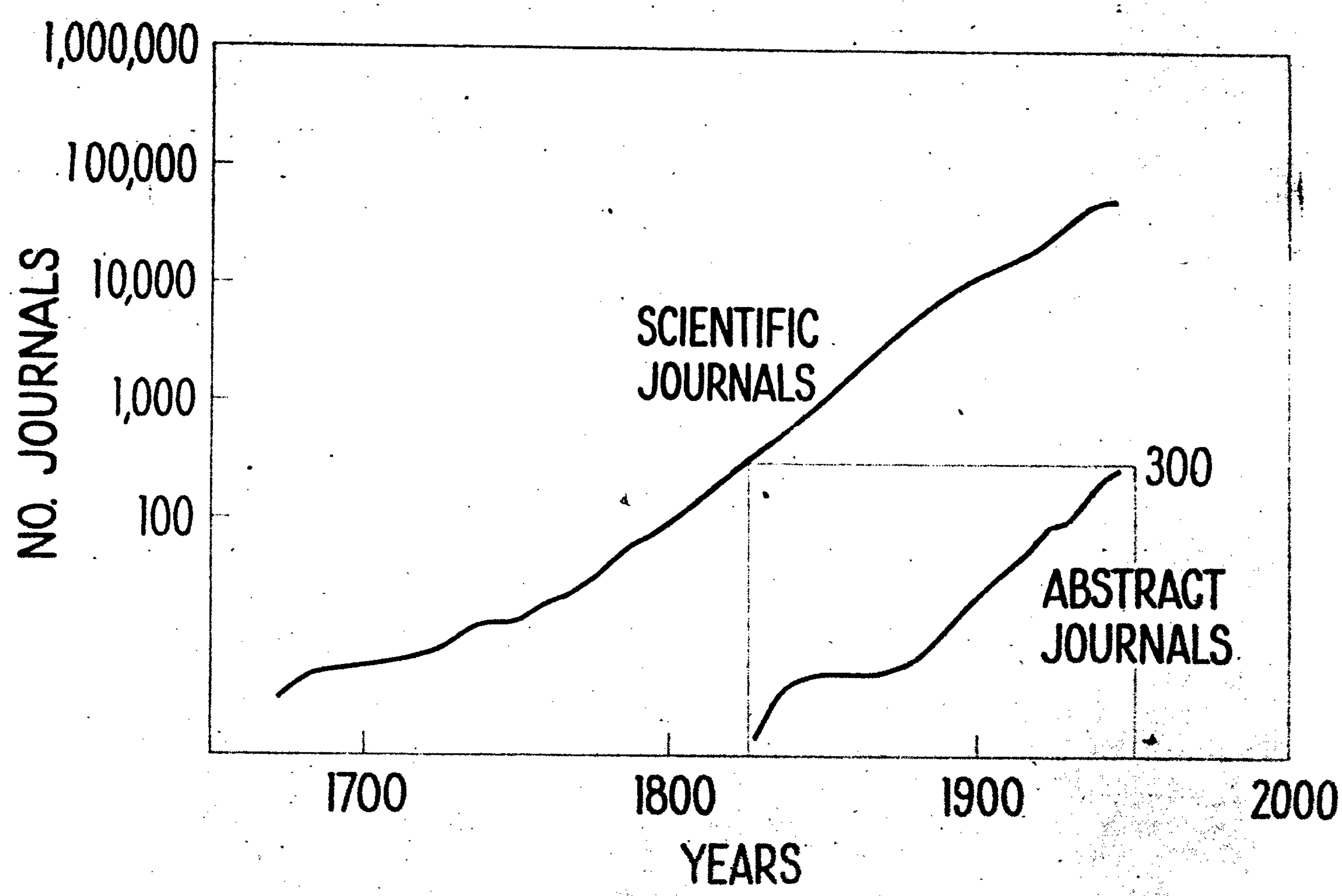
Just a brief summary; I think that one might gather, and I don't have to tell you people, because not only are you representatives of the Army but of industry, and you know that much of the research dollar that goes into defense ultimately finds useful outlets into civilian economy. The taxpayer certainly gets his dollar's worth, we think.

As to the future position of government in research development, it appears as though it will be in it for a long time; first, because it is necessary; secondly, it is part of the way we do things; and thirdly, the growth of knowledge is going along at such a terrific rate that it doesn't appear as though small units in our economy can support the demands that are placed on them. I have two examples here:

Slide 31

Growth of Knowledge

GROWTH OF SCIENTIFIC JOURNALS SINCE 1700

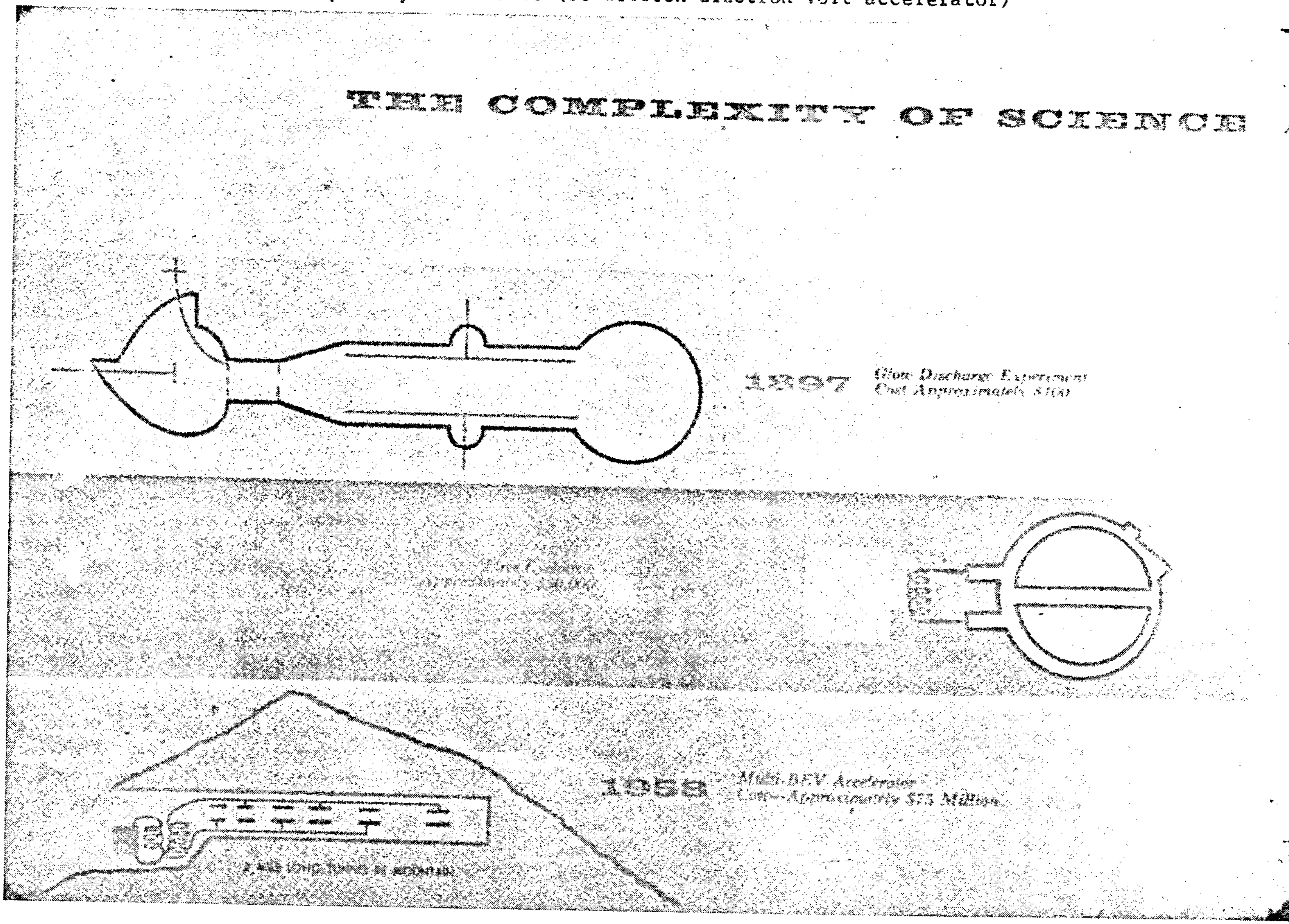


Slide 31

This first slide is a picture of the growth of knowledge, plotted on, as you observe, a semi-log scale from 1700 to the year 2000, and if you look at the scientific journals, you find that the slope is about equivalent to a doubling of knowledge every twelve years. This is on the assumption that there is a one-to-one equivalence between new knowledge and the new data published in the scientific journals. In the field of physics, I understand, it doubles every six years. In the abstract journal field you can see that here the slope of the curve is the same, so that our knowledge is growing so rapidly that even the abstract journals that just report on what is in the scientific journals are experiencing similar problems. Someone said that we renovate our society every two and a half decades, and when one thinks that about 90 percent of the brains that ever existed on the face of the earth are here today, you can well understand it.

Slide 32

Complexity of Science (50 billion electron volt accelerator)



Slide 32

This slide shows the 50 billion electron volt accelerator which is being planned at Stanford Research Institute and is being supported by funds which I understand the President approved; I would like to give you just an indication here of actually what has happened in this complexity of science. In 1897, I think the first experiments by Crooks in a glow-discharge tube cost about $100. In 1934 the first cyclotron, having greater resolution, cost approximately $50,000 and in 1958, the Stanford multi-billion electron volt linear accelerator cost approximately 100 million dollars and required a tunnel two miles long through one of the mountains in California to house it. So you can see tremendous increase in the cost of research. Rather interestingly, the cost per electron volt remained essentially constant through this period. It is 1/10th of a cent per electron volt when you figure it out.

I would like to terminate my discussion by saying that the Army is preparing to meet this rather large challenge of increasing technology and will expect, in the process of doing it, to add even more contributions to the general welfare of the nation.

PREDICTION OF THE RELIABILITY OF COMPLEX SYSTEMS

N. E. Golovin
Advanced Research Projects Agency

The purpose of the following remarks is to outline a point of view toward the reliability of complex systems which we have been developing in ARPA. In so doing, we shall attempt to describe why the problem of predicting the reliability of complex systems is such a difficult one, and hazard some suggestions as to lines of effort which perhaps have not been adequately emphasized because of extensive and somewhat fruitless searches for simple solutions.

First, it is probably advisable to start by defining a few principal terms, some of which have already been used.

By <u>part</u> we will mean the simplest constituent of a group of objects in an assembly of interest. Generally, it is an object which is not normally considered disassemblable into simpler elements. An electronic tube, a transistor, or a capacitor are examples. By a <u>component</u> will be meant an integrated group of parts performing, generally, a simple function in a grouping of similar objects. An instrument, such as a voltmeter or a complete radio receiver or transmitter, can be considered as a component. By a <u>subsystem</u> will be meant an aggregation of components performing a major function in a <u>system</u>. For example, if the <u>system</u> in question is a group of satellites to be used for navigational purposes, a <u>subsystem</u> would be the group of shipborne receivers, computers, and other similar components which transform satellite signals into a ship's latitude and longitude.

The term <u>reliability</u> has been defined in various ways. The following definition is essentially that first introduced by Carhart [1] and seems to have fairly wide acceptance. The term <u>reliability</u> of a system, subsystem, component, or part will be taken to mean <u>the probability that it will perform its required functions, under defined conditions, for a specified operating time</u>. This definition requires that the measure of reliability is to be a <u>number</u>. It presupposes, therefore, that the required functions of the object whose reliability we seek to establish are quantitatively relatable in some way to the numerical measure of its probability for performing them. It also presupposes that means exist for connecting, again quantitatively, the performance of the object to the environmental conditions under which it will operate. As we will see, an important aspect of the difficulty in establishing reliability lies in establishing such quantitative relationships.

Now we are interested here principally in the immediate problem of <u>predicting</u> rather than in <u>evaluating</u> reliability. The former is concerned with assigning a performance probability to a system before it is built, while the latter can be carried out only when at least a system prototype is available for testing. <u>Prediction</u>, therefore, requires estimating the performance probability of a system under conditions when not even a complete design may be available. The importance of prediction is associated generally with managerial judgments as to a proposed system's practicality or operational usefulness. In major programs, such as the NIKE-ZEUS Missile Defence System, Project Mercury, or a Communications Satellite System, prediction of expected operational reliability must be an integral part of the initial design feasibility study, and, therefore, an essential part of the decision to build or not to build a system prototype for further study. For example, if a communications satellite were to have a predicted mean life (a term which will be defined later) of two months instead of twelve, and its price in orbit runs into the tens of millions, then the associated estimates of the costs of establishing and maintaining a system of say four satellites in effective condition, may well be so great as to cast some doubt on the merits of even a large scale research and development effort. The large costs of such space systems further underline the importance of reasonably accurate reliability prediction because even relatively small differences in expected reliability will correspond to large absolute cost differences. Moreover, systematic reliability analysis in the initial stages of design produces additional engineering inputs for consideration of alternative approaches to an over-all system design. It will be particularly useful for guiding choice of acceptable trade-offs since generally performance, weight, space, cost, and operational reliability have all to be jointly manipulated in attaining an optimized design for the system.

Let's then address ourselves to the situation in which we have a detailed system design before us and see how far we can get in developing a general technique for predicting its reliability. My procedure will be to develop a theoretical approach to the problem interspersed with some comments and comparisons related to current methods in handling arbitrarily complex systems.

From some points of view, the crux of the problem in such an analysis is two-fold: First, the matter of how one defines "failure," and second, how one attempts to construct an expression for the over-all reliability of a system.

Conventionally one considers two types of failure, the so-called "catastrophic" and "degradation" kinds. The first is associated with the sudden, total failure of an object of interest, breakage of the heater element in an electronic tube being an example. The "degradation" type corresponds to gradual deterioration of one or more of an object's characteristic parameters to the point in time where an essential function can no longer be fully performed; for example, the gradually decreasing rate of cathode emission or, more generally, the drift in time of any electronic tube characteristic. In a general analysis, it is difficult to maintain a continuing distinction between these two types of failure, nor is it really necessary. In the subsequent remarks, we will combine these two physically distinct types of failure into one; we will say that an object fails at the time that any of its relevant physical characteristics attain values outside a specified range. Our analysis will try to show how this range must be determined for the general method to be consistent and useful.

As to the manner of constructing an expression for the over-all reliability of a system, the usual procedure is to begin with failure studies of parts and to construct from such data, successively, estimates for the reliability of components, subsystems, and finally of the system as a whole. We will reverse this usual procedure and start with a definition of failure for the system, and then work back through subsystems and components to the data on parts failures. The basic reason for this reversed approach is a somewhat theoretical one; namely, the fact that a part cannot logically be said to have failed unless the over-all system has done so. This means that the definition of part failure must be completely implied by the quantitative definition of what constitutes system failure. This point of view, it should be mentioned, is adopted in MIL-STD-441 for Reliability of Electronic Equipment [2], which suggests that the required performance of system details should be obtained by working back from over-all system functional requirements.

Now the over-all system design must specify how its outputs must fall within certain specified ranges of values if the system's objectives are to be met. The failure of a system to meet design objectives can thus be always unambiguously and quantitatively defined. For example, the transmitter power level in a communications system must be above a definite minimum value, if a specified receiver, at a given location, is to insure a specified, minimum usefulness of delivered information.. Furthermore, considering the assembly of distinct subsystems which interact to insure the output characteristics of the system, we can also take as given a set of mathematical relationships which allow calculation of over-all system outputs from the characteristic outputs of the constituent subsystems. This is not an unreasonable assumption. For a design to be at all realizable, such mathematical relationships are either deducible from applicable physical theory or have been empirically established from related prior experience with similar equipments. This is necessarily the case if subsystem, and lower order, nominal performance specifications are to have a rational scientific foundation. As a matter of fact, if such is not clearly the case, it can probably be cogently argued that the state of the applicable theoretical and practical arts does not justify a major system development program.

A key initial point from the reliability analysis point of view is the existence of such a mathematical representation, theoretical or empirical, as a foundation for rational, nominal design specifications. This is the case because such a mathematical representation can be used for constructing, on a computer of adequate capacity, a system simulation program in terms of the output characteristics of all of the system's subsystems. Computer-based system simulation will then allow the systematic study of the effects on over-all system outputs of arbitrary variations in the structure of subsystem output characteristics. The results of this type of investigation, ideally, will be the unambiguous specification of quantitative ranges for subsystem outputs, individually and/or in interdependent groups, which must be maintained if over-all system outputs are to be within the ranges defined by the tolerance requirements for system nonfailure. In this type of Monte Carlo simulation experiments, efficient conduct of the studies would no doubt be aided by experience with statistical experimental design techniques in other fields.

To emphasize the point, the importance of proceeding from the tolerance limits on over-all system outputs to the mathematically implied maximum ranges of allowed variation in subsystem outputs is, principally, that one thereby obtains a valid quantitative definition of subsystem failures. Furthermore, these definitions then permit equally valid specifications of the probabilities of failure of particular subsystems, or of combinations of these into groups, if some are found not to be individually independent with respect to failure. Thus, the probability of failure of a particular independent subsystem is the likelihood that one or more of its outputs fall outside the tolerance limits established as acceptable by such a computer-based simulation. Similarly, the probability of failure of statistically interdependent groups of subsystems is the likelihood that the structure of the groups' outputs to other individual subsystems or groups falls outside the ranges specified by the simulation study. Aside from the quantitative definition of what constitutes subsystem failure, such investigation will thus also have as an inescapable by-product the quantitatively justified grouping of subsystems into statistically independent entities whose probabilities of success or failure can then be validly multiplied together to obtain a measure for the probability of success or failure of the over-all system.

The remainder of the argument should now be clear. In similar fashion, one next treats each subsystem as a mathematically structured assembly of component outputs, and then each component as a mathematically related group of parts outputs. Employing computer simulation, there are then developed quantitative criteria for failures of components and parts, as well as their valid groupings for purposes of combining probabilities of success or failure.

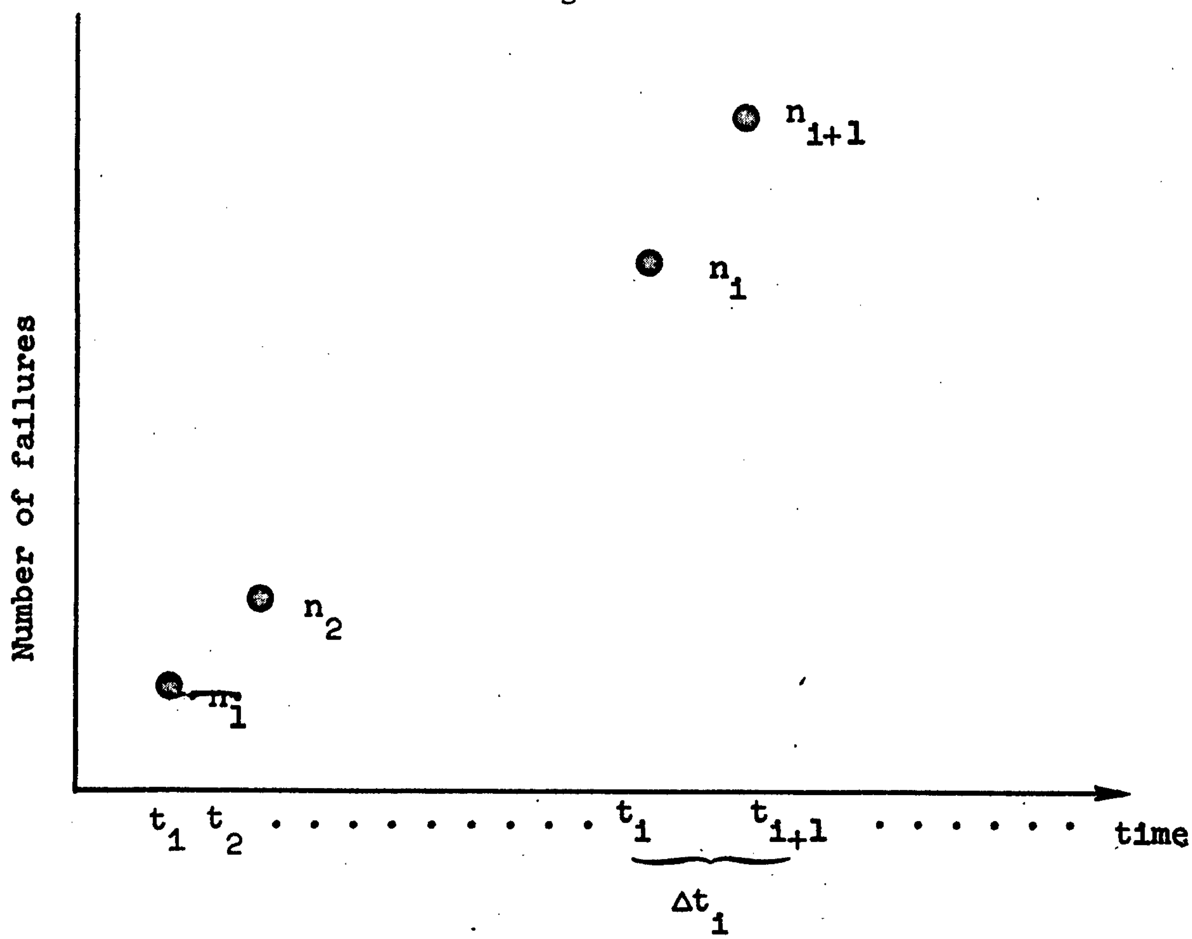
The essential product of these successive simulation studies is then two-fold: (1) We have estimates of the permitted range of parameter values for each part in the system as required for over-all system output acceptability; and (2) we have a quantitatively justified rather than an arbitrary basis for combining part failure probabilities to obtain, successively, such probabilities for components, subsystems, and the over-all system.

The approach I have outlined has been developing fruitfully, particularly during the last two or three years, in a number of other organizations concerned with complex systems. The White Sands Missile Range, the Rand Corporation, Chance-Vought Aircraft, Convair-Pomona, and the Autonetics Division of North American Aviation have each, in some measure, adopted this approach. At the White Sands Missile Range, for example, the current effort consists of constructing a probabilistic simulation model for the NIKE-HERCULES System in order to develop the technique fully with the view of applying it subsequently to Army missile systems at the design stage [3]. At Autonetics, the technique has been, in part, employed for the analysis and prediction of the reliability of the guidance system for MINUTE-MAN. While I am not familiar with any system to which the approach has been applied in its entirety, it seems that it has sufficiently solid logical merit to grow in importance, particularly in the case of systems in which the commitment to invest in a proposed design is greatly dependent on a realistic and objectively founded prediction of its reliability, and, furthermore, where the anticipated investments are so great that thorough, and therefore costly, reliability analyses have unquestionable managerial justification.

Let's complete our analysis by turning next to the question of how one develops a generally valid expression for the reliability of an individual part. For each part of the system our computer-simulation investigations have resulted in a quantitative specification of the ranges of acceptable variation in its parameters. Let's assume that we have as many parts as are required for an adequate sample, that we have a clearly defined environment in which the part will be required to maintain its characteristics, and that adequate facilities exist for carrying out a life test of the sample in such an environment. Incidentally, of all the assumptions so far made in this discussion, these last are among the most unreasonable. Usually, at the design stage of a system, many parts do not yet exist, the environment in which they must operate is not clearly established, and testing facilities allowing study of their performance under a realistic reproduction of the anticipated environment are almost never available.

We can test this sample until all of its members fail, and accumulate our results in the way shown in figure #1.

Figure 1



N = Number of parts initially in sample

$\Delta n_i = n_{i+1} - n_i$ = Number of failures in Δt_i

$N - n_i$ = Number of parts operational at time t_i

$\dfrac{\Delta n_i}{N-n_i}$ = Probability of failure during Δt_i

$1 - \dfrac{\Delta n_i}{N-n_i}$ = Probability of survival during Δt_i

With such information available, we can then carry out the calculation shown in the next two figures:

Figure 2

$R(t_i) \equiv$ Probability that a part survives time t_i

$R(t_i+\Delta t_i) \equiv$ Probability that a part survives time $t_i+\Delta t_i$

Then

$$R(t_i+\Delta t_i) = R(t_i)\left[1 - \frac{\Delta n_i}{N-n_i}\right], \text{ or}$$

$$-\Delta R(t_i) = R(t_i)\left(\frac{\Delta n_i}{N-n_i}\right), \text{ or}$$

on dividing through by Δt_i ,

$$\frac{-R(t_i)}{\Delta t_i} \cdot \frac{1}{R(t_i)} = \left(\frac{\Delta n_i}{N-n_i}\right)\frac{1}{\Delta t_i}$$

Introduce the definition of $\lambda'(t_i)$:

$$\left(\frac{\Delta n_i}{N-n_i}\right)\frac{1}{\Delta t_i} \equiv \lambda'(t_i) \quad ,$$

where $\lambda'(t_i)$ is the probability of failure per unit time given by the sample for the interval Δt_i. We can then write:

$$-\frac{\Delta R(t_i)}{R(t_i)} = \lambda'(t_i)\ \Delta t_i$$

Figure 3

Making the usual assumptions we can then pass to a differential relationship of the form:

$$-\frac{dR(t)}{R(t)} = \lambda(t)dt ,$$

which, on integration becomes:

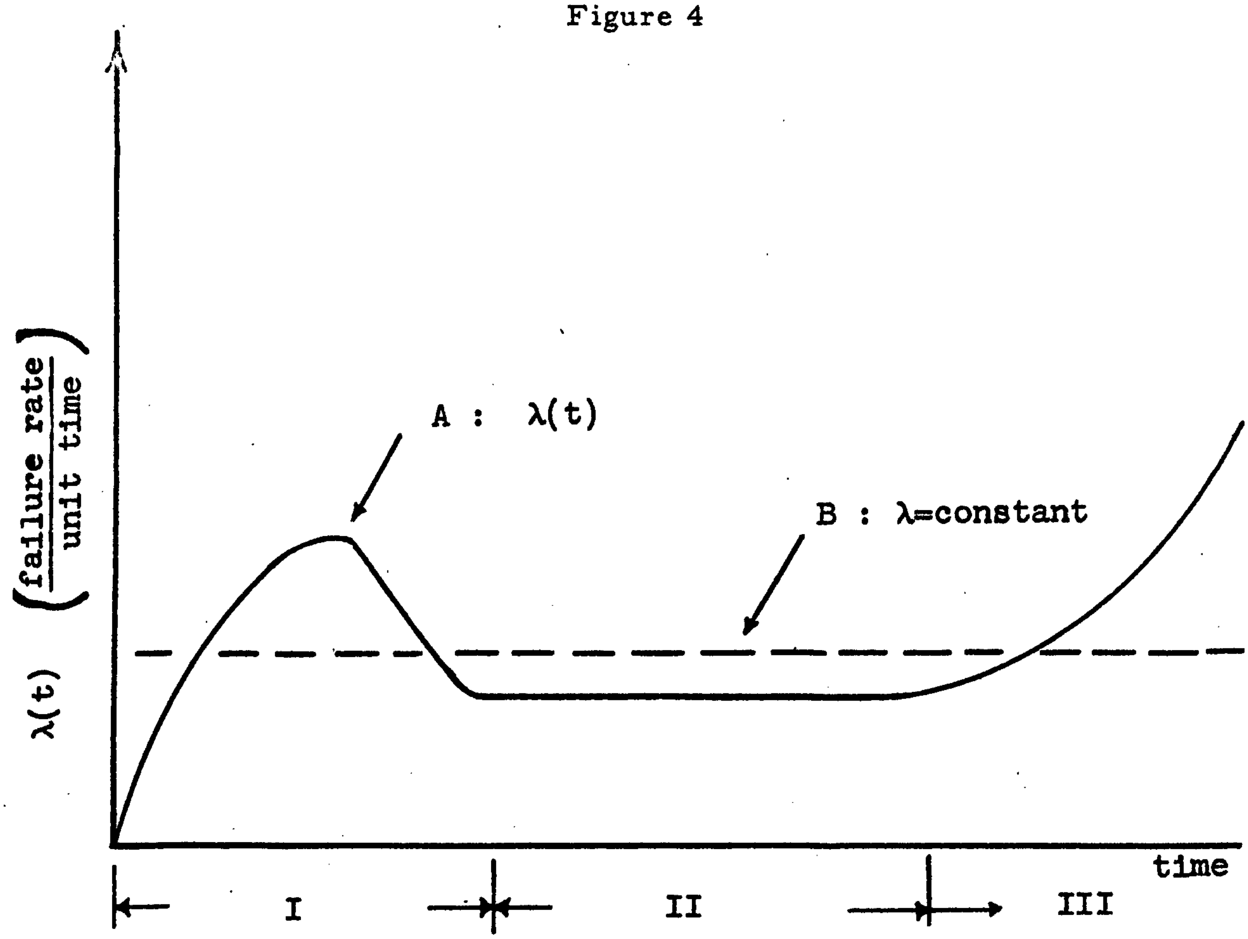
$$R(t) = R(o)e^{-\int_0^t \lambda(t)dt}$$

Here R(o) is the probability that the part is functional at the time the system begins its operation. Practically speaking this can hardly ever be taken as unity, but may be assumed close to this value. So we usually write:

$$R(t) = e^{-\int_0^t \lambda(t)dt}$$

The function $\lambda(t)$, let's call it "the part λ characteristic" may be arbitrary in character. In general, it is supposed to have the form of curve "A" in the following figure:

Figure 4



Region I : "Infant mortality" period

II : Mature operating life period

III : Rapid deterioration period

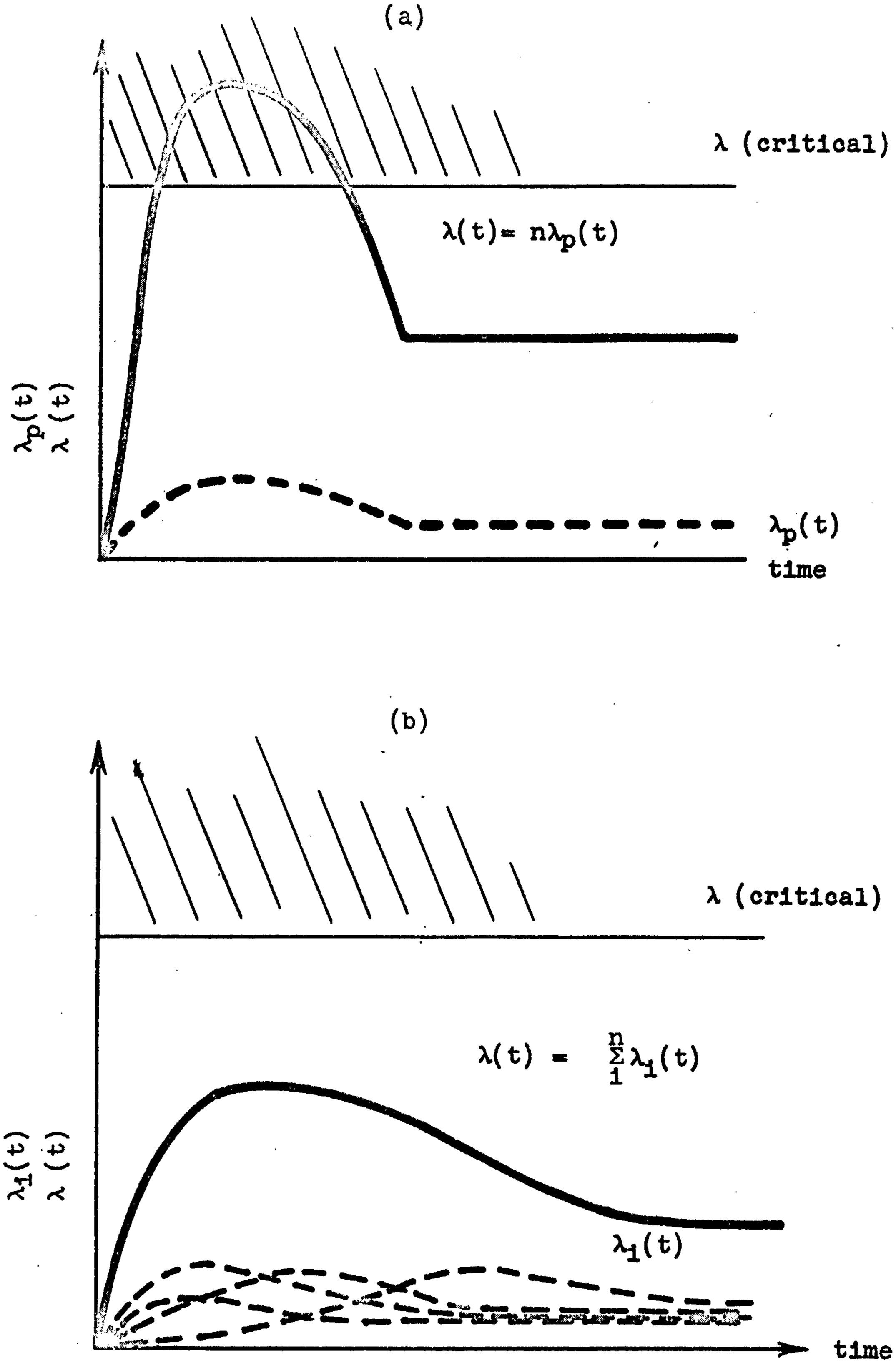
The straight line "B" is an average constant λ characteristic which, largely for the sake of simplicity in applications, is frequently assumed as applicable to most parts and components, for the purpose of taking, so to speak, a "first cut" at describing the corresponding reliability functions.

Substantial effort has gone into finding analytically tractable expressions for the part λ characteristic, or for its reciprocal defined as "the mean time to failure." The usual practices are to assume either that λ is constant, as has been mentioned, or that the part's mean time to failure is normally distributed about some average value with an appropriately chosen variance. There are many applications in which such simplified distributions are useful. However, it must be kept in mind that when many failure rates have to be added together to get a composite rate, the <u>errors</u> in such rates are also added. Accordingly, particularly in the case of complex systems, numerical methods allowing the use of actual rather than assumed part failure characteristics should be employed if at all possible. Additional reasons for care in this connection are suggested by the following:

1. In the first place, if a part is to be employed, without a prior "burn-in" period, in a component in which it is duplicated a large number of times, the superposition of the "infant mortality" periods may result in a total failure rate, for an appreciable time, which is not acceptable for the component or subsystem. This effect is shown in figure #5(a).

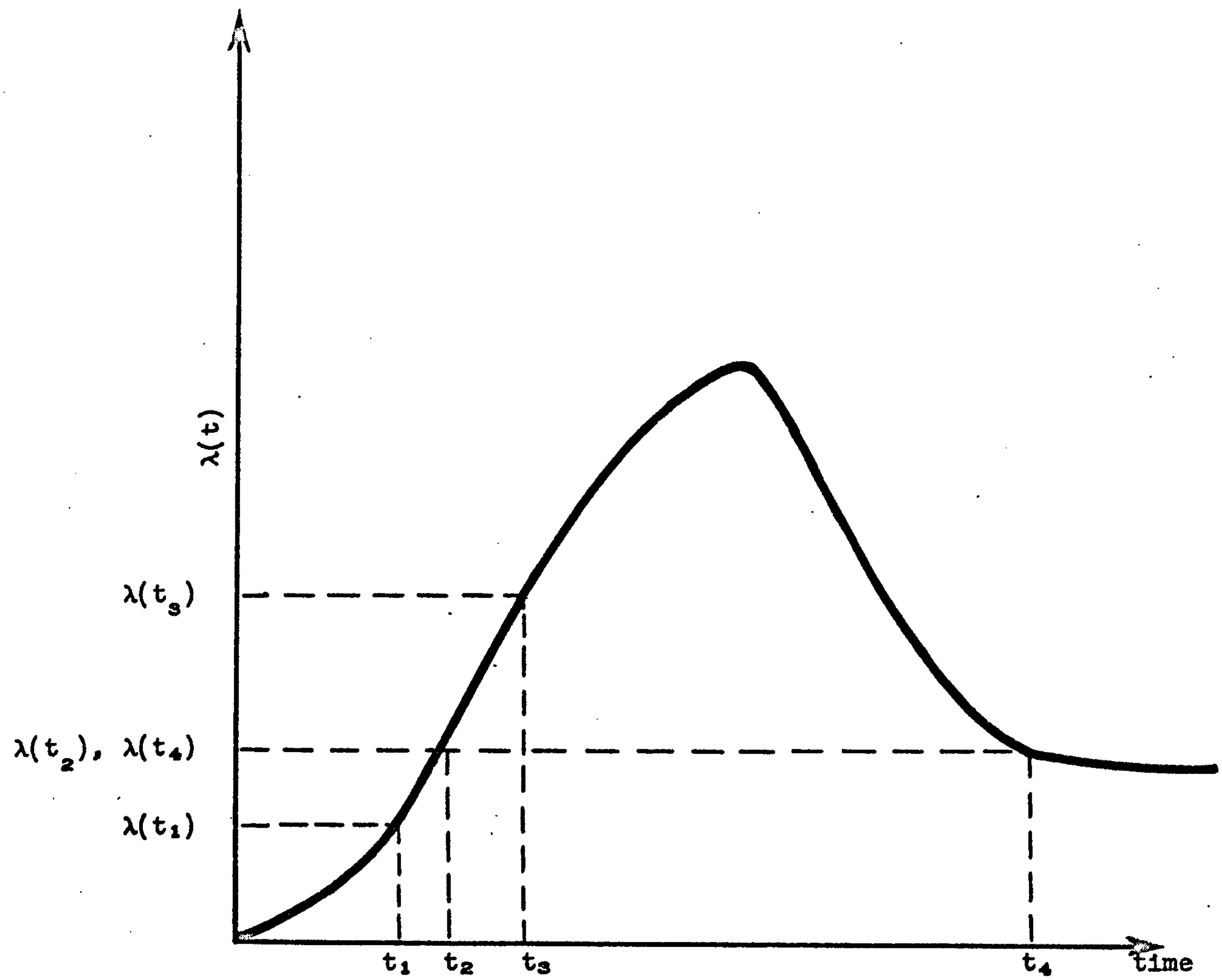
2. Secondly, if a number of different parts, with varying λ characteristics, are appropriately employed in a single component, it may be possible to design easily arrangements in critical circumstances that lead to a component λ characteristic which has desired form or a maximum desired failure rate level. How this can be done is shown in figure #5(b).

Figure 5


(a)
λ (critical)
λ(t)= nλp(t)
λp(t)
λ(t)
λp(t)
time
(b)
λ (critical)
λ(t) = Σ λi(t)
λi(t)
λ(t)
λi(t)
time

3. Thirdly, if a part is to be "burnt-in" prior to use, the required "burn-in" period cannot really be adequately established without constructing the time dependent λ characteristic. This can be seen readily in the next figure (#6). If the system operating period is t_1, for example, a "burn-in" period in unnecessary and even harmful since $\lambda(t_1) < \lambda(t_4)$. If the system operating period is t_2 or greater, a "burn-in" period is desirable.

Figure 6



The last three figures, and some of the associated arguments, have been taken from Druzhinin's article [4] in the book Reliability of Radio-Electronic Apparatus, published in 1958 by "Soviet Radio." This book, incidentally, is the first of promised annual publications of collections of research papers in the field. Apparently in this, as in so many other fields, U.S.S.R. technical organizations have initiated a systematic, broad-based approach. The National Bureau of Standards, in general, and Joan Rosenblatt, in particular, are to be thanked for their initiative in providing translations of some of the more important Russian papers in the reliability area.

Having established λ characteristics for all parts in the system, we can then directly employ the results of the previous analysis for systematic construction of the reliability functions for components, subsystems, and the over-all system.

The procedure can be illustrated by the argument on the next figure (#7), where $R_p(t)$ is the "part" reliability function:

Figure 7

$$R_p(t) = e^{-\int_0^t \lambda_p(t)dt}$$

For a component of n "independent" parts:

$$\lambda_c(t) = \sum_{i=1}^{n} \lambda_i(t) \quad \text{, and}$$

$$R_c(t) = e^{-\int_0^t \left(\sum_{i=1}^{n} \lambda_i(t)\right)dt}$$

If there are K identical parts in a simple redundant arrangement, the group's reliability function is:

$$R_g(t) = 1 - \left[1 - R_p(t)\right]^k \quad ,$$

where $R_p(t)$ has the form shown above.

The component reliability function, $R_c(t)$, [assuming (n-k) "independent" parts and a single group of k parts in a redundant arrangement] is then:

$$R_c(t) = R_g(t)e^{-\int_0^t \left(\sum_{i=1}^{n-k} \lambda_i(t)\right)dt}$$

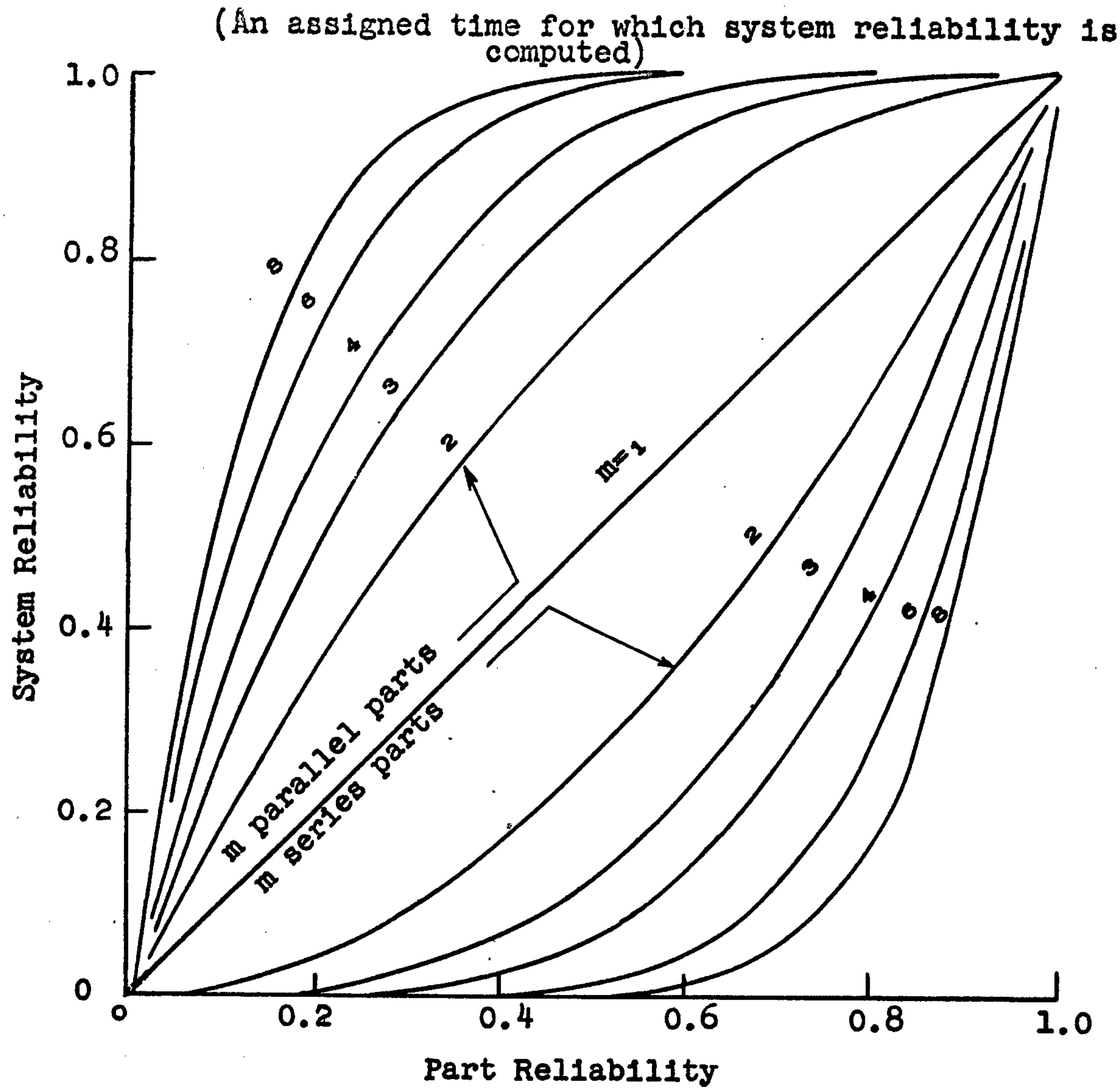
We are now familiar with the reliability function for a single part. If a component, for example, consists of n parts, which our analysis has shown to have independent failure probabilities, then the λ characteristic for it is simply the sum of the part characteristics and its reliability function is as shown on the figure. In the general case, $\lambda_c(t)$ must be obtained by the detailed superposition process of the type previously shown in figure #5.

On the other hand, returning to figure #7, if the component has been found, through our analysis, to have a group of parts which must be treated as an entity with respect to independence of failure probability in relation to the other parts, the reliability function for the group must be built up in accord with the logical relations found for the parts in the group. Probably the simplest case of this sort occurs when the group's parts merely provide functional redundancy. In such a case, the group and component reliability functions can be obtained as is shown on the figure.

The argument for other components, for the subsystems, and the over-all system then proceeds in an analagous way.

Here it should also be mentioned that another important by-product of the general method outlined is that if the resultant over-all system reliability is found to be, for example, unsatisfactorily low, a firm basis has been established already for evaluating the regions of the system where increased part or component reliability, or the employment of redundancy, will be most effective in raising the over-all reliability of the system. Incidentally, the relative values of improving part reliability and redundance, as well as the reliability degradation due to multiplying parts in series, can be inferred from the following figure (#8). In computing these curves, the parts are assumed identical and their reliability is assumed to follow the exponential law.

Figure 8

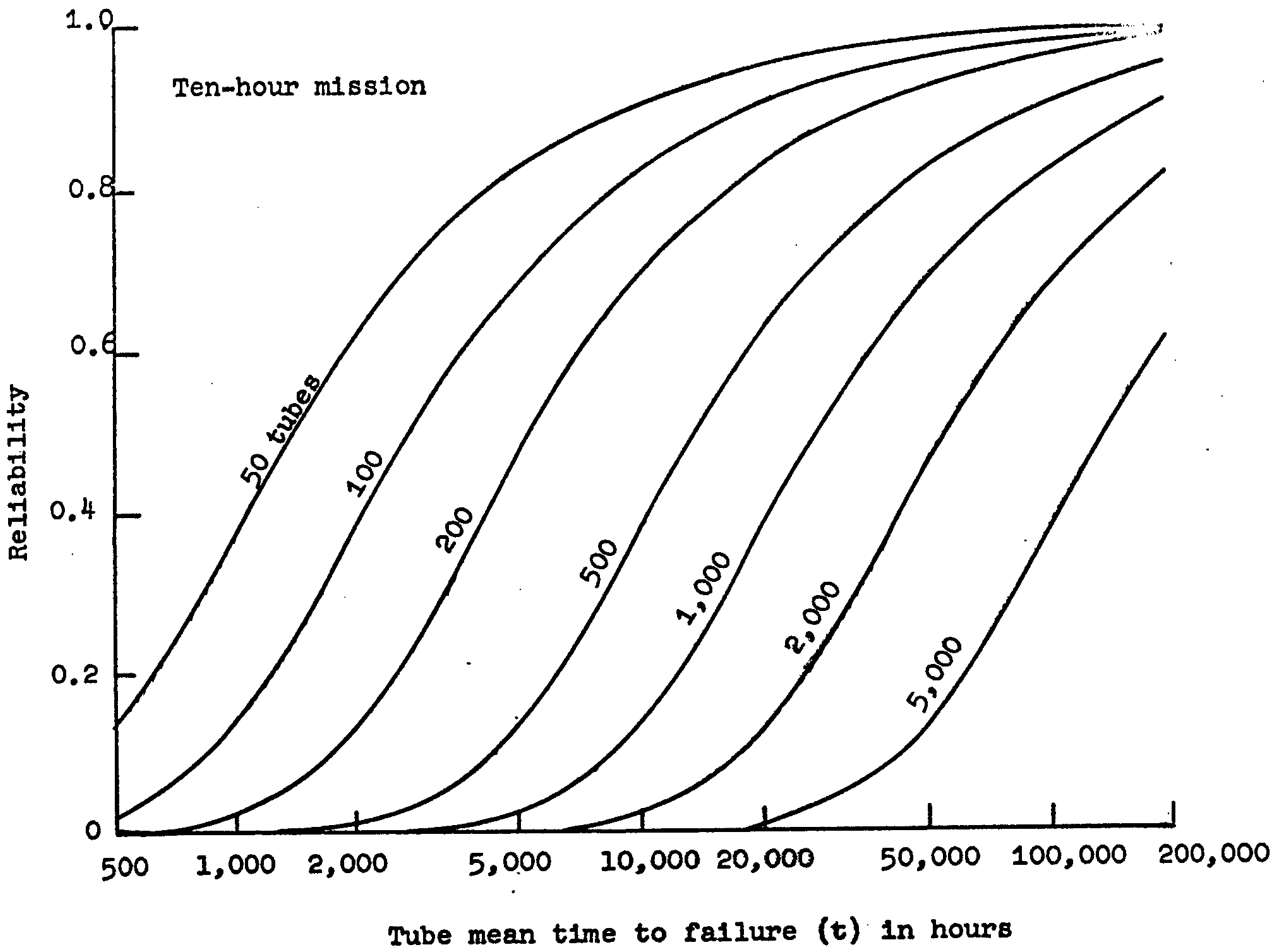


Reliability of a system of m identical parts

The interaction of increasing complexity and part reliability is striking, as shown in the next figure (#9). This is based again on the assumption that all parts have identified constant λ characteristics and are independent in their effect on over-all system reliability.

The last two figures are taken from R. R. Carhart's Rand Corporation Report, "A Survey of the Current Status of the Electronic Reliability Problem."

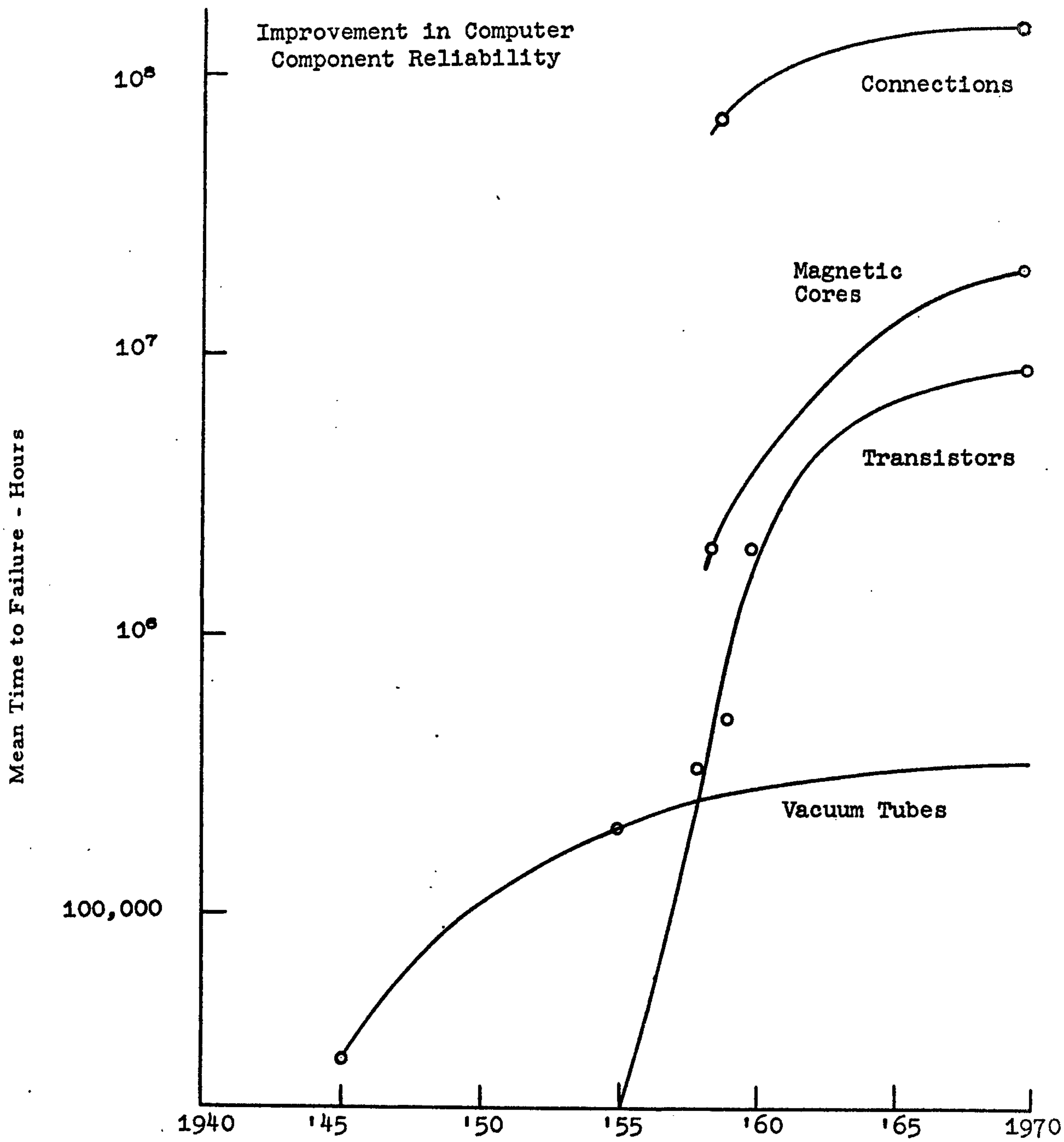
Figure 9



Reliability for 10-hour mission vs tube mean time to failure

In discussing part testing as a part of the job of predicting the reliability of systems in early stages of development, it was implied that the parts to be used in the system may not actually be available for tests to determine their λ characteristics. Of course, if such restrictions exist, there is little choice but to use available failure rate information for similar parts previously tested or used under closely related environmental conditions. However, this is an extremely dangerous procedure, at least for systems whose development cycles extend over several years. This is the case because technology is advancing so rapidly in some fields that errors of several orders of magnitude are possible unless careful and explicit allowance is made for the changing state-of-the-art. The following figure (#10), taken from a recent IBM report [5], shows what is expected to happen in a field of major importance to military applications -- that of computers.

Figure 10


Improvement in Computer Component Reliability
Mean Time to Failure - Hours
10^8
10^7
10^6
100,000
Connections
Magnetic Cores
Transistors
Vacuum Tubes
1940
'45
'50
'55
'60
'65
1970

It should be pointed out here that the indicated procedure is not offered, of course, as a panacea. It is being merely suggested that its employment will furnish a very useful and powerful tool for integration into conventional design practices. Nor has any mention been made of the pervasive and insidious influence on system reliability of various human factors throughout both the developmental and operational phases of a system's life.

Looking back over what has been said, it is clear that an obvious aspect of the problem of reliability prediction has been omitted; namely, a discussion of who is it that is going to make the analysis and take the responsibility for the ad hoc assumptions and simplifications that usually need to be made in applying any theoretical structure to a practical situation. The question is far from trivial because it is possible to get from highly responsible and competent groups, as has been already mentioned, estimates of reliability which differ by a factor of more than 10. An adequate treatment of this question might well warrant a time comparable to that which we have already spent. However, a number of assertions are rather readily in order, particularly for the type of approach which has been outlined.

1. In the first place, a key requirement in our point of view is clear articulation of the system design into associated mathematical structures and computer-simulation schemes. Such an undertaking must necessarily be undertaken either by the design staff itself or by a group otherwise living with the job. The requirement for this association is not just a matter of the complexity of the task, but also that the design program itself will benefit enormously from a thorough-going application of the suggested procedure.

2. In the second place, adequate resources must be provided for part development, procurement, and testing; sufficiently realistic environmental conditions must be available for the program, and means must exist for taking advantage both of information available on part and component performance from other contexts, and of changing technology which rapidly makes most extant information quickly obsolete. These tasks, most conveniently, must also be closely associated with the engineering group responsible for design.

3. In the third place, strong motives must exist for a realistic approach to the problem of reliability prediction if useful results are to be made available. This is clearly the case because of the frequency of situations, particularly in the case of complex systems, where no useful data is available and judgments must be substituted. Strong motivations to conservative and realistic prediction obviously can be found only if financial losses rather than gains are to be expected from lack of realism. These considerations suggest that the guidance and technical direction necessary for an adequate and realistic reliability prediction program cannot be

expected usually to be found ready-made in the group responsible for justifying the feasibility and usefulness of a design. This is particularly the case with complex systems for which it is out of the question to require a serial post-review of the designer's estimates -- simply because the task is too big and will probably require more time than can be afforded in postposing decisions to accept or not accept a given design.

There is, of course, no question that the system design contractor must have a competent reliability analysis group and that it should, at least administratively, not be under the direct management control of the design group itself. However, the above remarks suggest, further, particularly in the case of government procurement actions for new complex systems in which matters of operational reliability are of basic importance, that somewhat more attention than may have been customary in the past be given to developing reliability analysis and prediction programs coincident in time with the beginning of a system design. As a matter of fact, it can probably be persuasively argued that a thorough-going, coincident effort is likely to be not only more fruitful but also, in the final summing up, including operational phases, much less costly.

Also, it seems fairly clear that such reliability analysis and prediction programs should proceed either under the direct, in-house guidance of the government or be conducted under such guidance assisted by adequately motivated contractors not themselves committed to major R&D management or hardware programs in the systems being so studied.

REFERENCES

[1] R. R. Carhart, A Survey of the Current Status of the Electronic Reliability Problem, Rand Corporation Report, RM-1131, August 14, 1953.

[2] MIL-STD-441, Reliability of Military Electronic Equipment, 20 June 1958, Office of the Secretary of Defense, Supply and Logistics.

[3] Exploratory Conference on Missile Model Design for Reliability Prediction, Report of the 3rd Meeting, White Sands Missile Range, 23 April 1959.

[4] G. B. Druzhinin, On Methods of Calculating the Reliability of Systems, Soviet Radio, Moscow, 1958.

[5] Digital Computer Characteristics for Space Applications, IBM Corporation, Federal Systems Division, Oswego, New York, June 9, 1959.

ON THE REPEATED-MEASUREMENTS DESIGN IN BIOLOGICAL EXPERIMENTS

Ardie Lubin
Walter Reed Institute of Research

SOME DIFFICULTIES IN USING THE REPEATED MEASUREMENTS DESIGN. The phrase "repeated measurements design" is used to characterize those experiments where each subject* is tested more than once. Usually this is done to increase the precision of the experiment by eliminating the between-subjects deviance from the estimate of error deviance. Often it is done to avoid multiplying the number of subjects used in the experiment.

The main emphasis of this paper will be on the design where each subject receives only one treatment, applied repeatedly over a period of time, and the chief interest is in the chronic effect of the treatment. An example of such a design would be a drug experiment where each subject is given a constant drug dose every day and tested periodically.

The rest of the paper will discuss the multiple treatment cross-over design in which each subject receives a single treatment for a fixed unit of time, but is changed to a different treatment wherever a new time unit starts. A common example would be a drug experiment where a subject might be on drug A the first week, drug B the second week, and so on. The separate effect of each drug is then estimated from the results.

The purpose of this paper is to point out that: a) any repeated measurements on the same organism will in general exhibit statistical dependence; therefore multivariate analysis of variance rather than univariate analysis of variance is appropriate, and b) all standard cross-over designs assume that the carryover effect of a treatment on a succeeding treatment is constant and does not depend on the nature of the succeeding treatment, i.e., carryover is additive and does not interact with succeeding treatments.

Most of this paper is concerned with possible experimental and statistical answers to the questions which arise when dependent measures are used in a continuous treatment design. The problem of carryover effects that interact with subsequent treatments is quite different. No answers to this problem are given here; instead we ask if there is, in fact, any way of preserving the advantages of a cross-over design and obtaining unbiased estimates of the treatment effects when carryover interaction is present.

Let us take a hypothetical psychiatric experiment with a repeated treatment design. Say that a psychiatrist thinks slow reaction times are characteristic of paranoid schizophrenics and he wishes to alleviate this symptom by chronic administration of some tranquillizing drug. He selects a sample of N paranoid schizophrenics, puts each patient on a

*The word subject is used here as a general synonym for the experimental unit of observation.

maintenance dose and starts testing reaction time once a week. At the end of k weeks, the reaction time scores can be arranged as a rectangle, N rows by k columns. The statistical analysis indicated by such tests as Edwards (1950), Lindquist (1953) and McNemar (1949), would be a two-way analysis of variance, with k-1 degrees of freedom for the effect of weeks, N-1 degrees of freedom for the between-subjects effect, and (k-1)(N-1) degrees of freedom for the subject-by-week interaction effect. Then the significance of the differences between the k weekly means would be assessed by an F ratio using the subject-by-week interaction as the error term. Let us call this ratio the "univariate F."

One of the basic assumptions for the use of subject-by-week interaction as the error term, is that all ovserved scores are statistically independent of one another. However, in this hypothetical experiment, it is almost certain that the scores on the first week will have a positive correlation with the scores on the second week, third week, etc.

In 1948, Kogan suggested that if the assumption of independence is not met, the univaraite F ratio overestimates the significance of the difference between the k-means. In 1954, G.E.P. Box, in a brilliant article, gave a general technique assessing the effect of departures, from independence and from equal variances, on the univariate F. In general, his conclusions substantiate Kogan's guess; when the null hypothesis is correct and the observations are dependent, the univariate F will exceed the tabled significance levels more often than it should. Roughly speaking - the effect of correlation between the weeks (i.e., treatments) is to reduce the apparent number of degrees of freedom in the numerator and denominator of the F ratio.

Box's model, and the conclusions he drew, are worth sketching here since they demonstrate why multivariate analysis of variance, rather than univariate analysis of variance is most generally appropriate for correlated observations. Two assumptions are made:

a) The vector of scores for any subject is statistically independent of the score vector for any other subject, under the null hypothesis.

b) Each vector is a sample from the same multivariate normal population.

In terms of our hypothetical psychiatric study, this means that the N paranoids are randomly selected and the relation between the scores of any two weeks, say week s and week t, is bivariate normal. The variance of week t, v_{tt}, need not equal v_{ss}; v_{et} does not necessarily equal the correlation between any other pair of weeks.

C. R. R. Rao in 1952 (pp. 239-244) showed how Hotelling's T^2 could be adapted to give an exact test of the differences between correlated means. Basically, Rao takes a linear function of the k scores and compares the mean of this linear function to the variance of the linear function. (A convenient computation routine for this test is given by T. W. Anderson in his 1958 text par. 5.3.5).

Using an exact multivariate approach, Box shows that, under the null hypothesis, the true distribution of the univaraite F with (k-1) over (k-1)(N-1)d.f. can be approximately represented by the same F value with the degrees of freedom reduced by a fraction, ε. This fraction, epsilon, is a function of the k by k covariance matrix.

$$(1)\quad \varepsilon = k^2(\bar{v}_{tt} - \bar{v}_{..})^2/(k-1)\left[\sum_{t=1}^{k}\sum_{s=1}^{k} v_{ts}^2 - 2k\sum_{t=1}^{k}\bar{v}_{t.}^2 + k^2\,\bar{v}_{..}^2\right]$$

where v_{ts} is the covariance of the N pairs of scores from week t and week s, and $\bar{v}_{tt}$ is the average variance for the k weeks.

The maximum value of epsilon is one, and this is reached only when the k variances are equal and the $\frac{k(k-1)}{2}$ correlations are constant. In this case, Box's approximation gives the exact results; when the correlations are constant and the variances are equal, then the univariate F ratio can be used to give the exact significance level of the differences between the k correlated means.

Geisser and Greenhouse (1958) have shown that the lowest value that epsilon can take is 1/(k-1). They argue that since no one has shown what sample estimate of epsilon is most appropriate, and the robustness of epsilon has not been investigated, it is best to use the minimum value of epsilon for a conservative test. This conservative test consists of computing the univariate F, and entering the tabulated F distribution with 1 over N-1 d.f. If the result is significant, there is no need to go further; the exact test would be significant. However, if the conservative test is not significant, one can now make an upper-limit test of the univariate F (setting epsilon equal to unity). If an assumed epsilon value of unity gives a non-significant result, then the null hypothesis can be accepted, since no calculated value of epsilon can give a more significant result. However, if using the full degrees of freedom gives a significant result, then the research worker is in a dilemma. Geisser and Greenhouse apparently would next try Box's approximate test, using a sample estimate of epsilon. I would recommend an exact multivariate test such as Rao's.

You can see that the Geisser-Greenhouse approach allows one to bracket the significance level of F with the same amount of computation that is used in the usual two-way analysis of variance. The laborious computations for an exact multivariate A. of V. include the data necessary for a two-way A. of V. Therefore, it will always be profitable to try the Geisser-Greenhouse approach first, before proceeding to the rest of the distasteful arithmetic necessary for multivaraate analysis.

Here it is essential to stop and point out that Box's model explicitly assumes multivariate normality. What alternatives do we have if multivariate normality does not hold or can not be forced by a transformation? As we mentioned previously, the Rao exact multivariate test for differences between correlated means essentially compares the mean of a linear function to the variance of that linear function. The question of multivariate normality can therefore be posed as the question of whether the scores produced by the linear function have a normal distribution. When k is large and correlations are near-zero, we know that the linear function will yield a near-normal distribution of scores. However, if the linear function scores are not normally distributed, the means will have a near-normal shape, assuming the samples of N subjects to be large and selected at random. Therefore the Rao multivariate test will be robust to deviations from normality when N is large or when k is large and the correlations are small.

In those cases where robustness is in question because of small N, high correlation, or other characteristics of the data it seems to me that the basic strategy should be to resort to the randomisation test introduced by R. A. Fisher (1935, par. 21). If we use Box's first assumption, that each subject's vector of scores is independent, and change Box's second assumption to read "each vector is a sample from the same symmetric multivariate distribution" then we will meet Fisher's requirement that the scores for the treatments be drawn from the same population. Since the problem is whether the means differ significantly, it seems reasonable to use the usual univariate "between treatment means" deviance as the criterion. However, E. S. Pearson (1937) has pointed out that the most powerful criterion depends upon the form of population distribution. For example, when the population distribution is rectangular, midpoints rather than means should be used. The null hypothesis here is that the k scores for any subject are completely interchangeable and any permutation of the k scores can be substituted for the original vector. Since there are N subjects there are $(k!)^N$ sets of scores. Each set is a possible sample from the original finite set of scores. The between-treatments deviance can be computed for each permutation and we can ascertain where our observed between-treatment deviance falls in the frequency distribution of all possible values from this finite sample. If our observed sample value equals or exceeds the assigned significance level, the means can be judged to be significantly different.

This permutation test preserves one of the advantages of the univariate A. of V. approach, N can be less than k. (The multivariate methods cannot be applied routinely for N less than k since the inverse of the k by k covariance matrix does not exist). One disadvantage of the permutation test for differences between means is the requirement that all treatments have identical distribution moments (except fot the means). However, the identical distribution assumption apparently is made in every parametric or non-parametric statistical test, of the difference between two or more samples. The assumption of identical distributions seems to be necessary for generating any statistical test of differences. Some empirical results I have seen suggest that if the distributions are symmetric about their midpoints, they need not be identical; the permutation test is presumably robust to non-identical distributions in these cases.

The basic disadvantage of the permutation test is the extraordinary amount of labor required for even moderate values of N and k.

Suppose, instead of asking if the means are different, we ask if the scores for one week tend to be higher than the scores for other weeks. Then the hypothesis concerns the equality of the rank order averages.

As is well known, Kendall's W, or concordance coefficient, is a simple easily-computed test of this hypothesis. (1948).

Wallis and Friedman independently, and about the same time as Kendall, devised statistics that are algebraically equivalent to Kendall's W.

Essentially, Kendall's W is a permutation test on scores that have been transformed into rankings. The basic assumptions are - score vector independence and identical treatment distributions, exactly the same as those made for Fisher's randomization test, but the laborious computations have disappeared. However, it should be noted that we are now asking a different question - whether the average rank differs significantly between treatments. Does inequality of the average rank imply inequality of the means and vice versa? I have found several empirical examples where Kendall's W was significant but the univariate and multivariate A. of V. tests fell below significance.

Generally, one assumes that the rank order statistic and the A. of V. statistic are testing the same thing, but that the rank-order test is less powerful. However, the discovery of empirical examples where Kendall's W was significant and the F ratio wasn't, shook my faith in this proposition. Since then, I have learned how to construct examples where the means are exactly identical but the average rank differs significantly. However, in the construction of these counter-examples, I found it necessary to introduce non-identical distributions, to violate one of the two basic assumptions.

Therefore, I would like to raise the explicit question: What are the necessary and sufficient conditions such that rank-order tests are less powerful versions of the analogous A. of V. tests? This problem transcends the context of repeated measurements. Perhaps situations can be devised such that any rank-order statistic will be more significant than its metric analog. I raise this question - I hope some statistician can answer it.

I am saying that sometimes rank-order tests answer a different question than their metric analogs do. I am not saying that rank-order tests should be abandoned. There may be many occasions when the A. of V. test is not quite the right way to answer the question - when the major interest is in whether one treatment differs from another treatment, and the amount of the difference is irrelevant. There are other situations where it is not clear that the units of measurement are all equal, as in psychological test scores, so that equal metric differences may not be of equal importance. In these and other cases, the experimenter, upon reflection, may discover that he is more interested in rank-order than in metric differences.

Let us now come back to our psychiatric example. You will recall that in our example the psychiatrist had placed his schizophrenic patients on a tranquilizer in the hope that the reaction times would be shortened. Time is a natural unit of measurement and there is little ambiguity there. If he is primarily interested in the therapeutic value of the drug, then the exact amount of decrease is important. Presumably, any improvement which is insignificant for practical purposes, say a decrease of 1/100 of a second, would be of little therapeutic interest, even if it were statistically significant. However, if his interest is primarily theoretical, for example, he hopes to find whether the delay is at the nerve-muscle junction or is caused by central factors, then <u>any</u> decrease in reaction time will be of interest to him.

Even if he knows that relative and not absolute differences are his main interest, should the psychiatrist use a general test of differences such as Kendall's W, or a test which specifies an a priori rank-order; for decrease in reaction time should be a monotonic function of number of weeks on the drug. Whenever a set of correlated means has a predicted rank-order, each subject's obtained rank-order can be correlated with the predicted rank-order and the average of all N rank-order correlations can be tested for significance. In 1954 Jonckheere presented an explicit test of this sort, using Kendall's tau. Lyerly (1952) has discussed the distribution of the average Spearman rank-order coefficient, rho.

Jonckheere's average tau test (as well as the equivalent Spearman form) is unique among non-parametric tests in that there is no parametric analog. So far as I know, there is no regression procedure or Hotelling T^2 criterion that can be applied to test for monitonicity. Any metric technique needs a formal specification of the exact mathematical relation between reaction time and weeks, before such a relationship can be tested.

This brief survey of the statistical tests appropriate to a continuous treatment design does not, of course, cover all the relevant topics, but it does show there are rational procedures for treating the data which differ considerably from those found in many statistical text-books.

So, to summarize the statistical recommendations in our hypothetical experiment, the psychiatrist might use the Geisser-Greenhouse multivariate A. of V. approach or he might use Jonckheere's average rank-order coefficient, but he should not make a routine application of the usual two-way A. of V.

Let me deal briefly with some of the experimental problems raised by repeated measurements. Almost certainly there will be an improvement in reaction time, whether or not the drug is used. The very act of measuring reaction time gives the patient practice on this task, allows him to adjust to the situation, and so on. This quasi-Heisenberg effect is very common with most kinds of repeated measurements. The blood pressure of a subject is usually higher during the first few determinations than on subsequent occasions. The prick of the hypodermic needle can cause significant changes in blood composition until the subject becomes habituated.

One common way of dealing with the problem is to run a control group. This allows us to estimate the trend, without the drug. Another way is to run each patient through the measurement procedure until he reaches a steady state. Control groups are, of course, almost always necessary because of vagaries in the experimental situation, apparatus, etc., but even when controls are used, I advocate running each subject to a steady state. Not only do you eliminate any complex trend that may exist, but the intra-subject variation usually decreases markedly. This makes it particularly advantageous to use the intra-subject rather than the inter-subject variance as error.

But this raises the question of what part of the performance we want to measure. Perhaps it is exactly the factor in learning, habituation, practice, etc., which the experimenter wants to study. In this case, a control group will enable him to assess the effect of a drug on the initial rate of change. In most situations we are interested in the performance of the Subject on a well-learned routine task. When this is, in fact, true, then we may be measuring some factor which is irrelevant to our question when we include measurements taken at a time of rapid learning or habituation.

Let me hasten now to my final point, a sweeping generalized warning against the use of crossover designs.

If you wish to assess the separate effect of two or more treatments, don't apply the treatment to the same organization. A brief logical justification is as follows: if you're trying to assess the effect of a treatment by itself, then almost certainly you do not have enough previous data to estimate the carryover effect and in particular the interaction of the carryover effect with other treatments. But all designs using two or more treatments on the same organism assume that there is no interaction of the carryover effect with preceding or subsequent treatments.

Another way of looking at it is to consider the rotation experiment. Here the treatments are applied in predetermined sequence and the problem is the effect of the sequence of treatments on the subject rather than the effects of the individual treatment.

There are countless examples in medicine where the order is all-important, e.g., when weak and strong bacterial strains are injected in an organism. The enormous difference in the effect of the two rank-orders is the basis for vaccination.

If the experimenter who proposes to use a cross-over design thinks that a rotation experiment with the same treatments would also yield important information, he is assuming that carryover interaction can exist; that treatment A can inhibit or potentiate treatment B. In this case, his estimates of the effect of each treatment from the cross-over design will be hopelesssly enmeshed with the carryover interaction effects.

REFERENCES

1. Anderson, T. W. (1958). Introduction to multivariate statistical analysis. New York, Wiley.

2. Box, G. E. P. (1954). "Some theorems on quadratic forms applied in the study of analysis of variance problems. II. Effects of inequality of variance and correlation between errors in the two-way classification." Ann. Math. Statist., 25, 484-498.

3. Edwards, A. L. (1944) Statistical Analysis. New York, Rinehart.

4. Fisher, R. A. (1935) Statistical methods for research workers. New York, Stechart.

5. Friedman, M. (1937). "The use of ranks to avoid the assumption of morality implicit in the analysis of variance." J. Am. Statist. Assn. 32, 675.

6. Geisser, S. and Greenhouse, S. W. (1958). "An extension of Box's results on the use of the F distribution in multivariate analysis." Ann. Math. Statist., 29, 885-891.

7. Geisser, S. and Greenhouse, S. W. (1959). "On methods in the analysis of profile data." Psychometrika, 24, 95-112.

8. Jonckheere, A. R. (1954a). "A distribution-free k-sample test against ordered alternatives." Biometrika, 41, 133-145.

9. Jonckheere, A. R. (1954b). "A test of significance for the relations between m rankings and k ranked categories." Brit. J. Statist. Psychol., 7, 93-100.

10. Kendall, M. G. (1938). "A new measure of rank correlation." Biometrika, 30, 81.

11. Kendall, M. G. (1948). "Rank correlation method." London, Charles Griffin & Co., Ltd.

12. Kogan, L. S. (1948). "Analysis of variance - repeated measurements." Psychol. Bull., 45, 131-143.

13. Lindquist, E. F. (1953). Design and analysis of experiments in psychology and education. New York, Houghton Mifflin.

14. Lyerly, S. B. (1952). "The average Spearman rank correlation coefficient." Psychometrika, 17, 421-428.

15. McNemar, Q. (1955). Psychological Statistics. New York, Wiley.

16. Pearson, E. S. (1937). "Some aspects of the problem of randomisation." Biometrika, 29, 53-64.

17. Rao, C. R. (1952). Advanced statistical methods in biometric research. New York, Wiley.

18. Wallis, W. A. (1939). "The correlation ratio for ranked data." J. Am. Statist. Assn., 34, 533.

THE GERMFREE LABORATORY AT THE WALTER REED ARMY INSTITUTE OF RESEARCH: Design of Experiments using Germfree Animals.

Ole J. Malm
Stanley M. Levenson
Captain Richard E. Horowitz
Departments of Germfree Research and Surgical Metabolism and Physiology
Walter Reed Army Institute of Research, WRAMC

Germfree rats, mice, guinea pigs and chicks are now routinely available in special laboratories like the Walter Reed Department of Germfree Research. The germfree animal has become a research tool, uniquely suited to provide answers which cannot be obtained by the use of conventional animals alone.

By the use of germfree animals, certain problems can be readily and equivocally answered in simple experiments which do not involve large numbers of animals and statistical evaluation of the experimental data. A fundamental question, asked by Louis Pasteur (1885) was whether life without bacteria was possible. This question has been answered in the affirmative by the successful rearing of a number of animal species over long periods of time by the pioneer laboratories in germfree research, (goat, rabbit, monkey, rat, mouse, guinea pig, fowl and fish).

Many metabolic processes occurring in the animal organism may be dependent upon enzyme systems of commensal bacteria rather than on endogenous enzymes in the animal. The germfree animal lends itself superbly for the study of these problems. It is possible, through a few well designed experiments, to obtain definite answers to a problem which requires a great number of complicated experiments when undertaken with conventional animals as exemplified in the following study of urea metabolism accomplished at the WRAIR Germfree Laboratory (1). The metabolism of urea, the first organic compound to be synthesized (Wohler, 1828), has always interested biologists and physicians. Considerable time and effort has been expended by large numbers of investigators in laboratories all over the world attempting to determine whether the metabolism of urea in mammals was under endogenous or bacterial control. In a review of this problem published in Physiologic Reviews, Kornberg listed over 50 investigations, yet the precise role of the intestinal bacterial flora remained equivocal and inferential. Indeed, as recently as 1956, Conway, a leading Irish biochemist, presented evidence before the 20th International Physiological Congress, which he interpreted as showing that the gastric urease of mice was intracellular rather than bacterial.

The problem of the bacterial origin of urease was clearly susceptible to test in the germfree animals. Accordingly the metabolic unit of the Germfree Laboratory, WRAIR, injected subcutaneously C^{14} urea into two conventional and three germfree rats, and administered it orally to one germfree rat. Each rat was then immediately placed in a metabolic apparatus and its urine, stools, and expired air were collected. Any hydrolysis of urea to ammonia and carbon dioxide would be readily detectable, since the CO_2 formed from the administered urea would contain radioactive carbon.

The conventional animal's expired air contained 100 times as much radioactivity as the germfree animal's. The pattern of urea hydrolysis in the germfree rats was the same whether the urea was given subcutaneously or intragastrically.

The very small fraction of the injected C^{14} (0.02%) expired by the germfree rat is due to spontaneous hydrolysis of urea, not to enzymic breakdown.

These results, conclusively, demonstrate that the enzymic hydrolysis of urea by the rat is effected only by the urease of its bacteria. Moreover these results provide the experimental answer to the clinical observation that certain oral antibiotics effectively control ammonia toxicity of patients with liver dysfunction. With a few germfree animals and in a very short period of time, an unequivocal answer to this problem which had been inconclusively worked on by many investigators for over 75 years was obtained.

Unfortunately, many experiments in which germfree animals can be of singular value, involve a more complicated design due to some special problems in germfree research. These special problems fall into two main categories:

1. The special environment in which the germfree animal lives, and

2. Peculiarities inherent in the germfree animal itself.

In the discussion to follow we will define some of these environmental and biological factors peculiar to germfree research. The main problem is to devise the proper control for the germfree animal when the control is to be his normal or conventional laboratory counterpart. This is a vital question since a well controlled experiment, properly planned, will save time, work and money by reducing the number of animals necessary to obtain statistically significant results and obviate repetitions.

THE "GERMFREE" ENVIRONMENT. The germfree environment is potentially the most controllable of any now available in which to conduct animal research. Ideally, in any experimental study, the investigator would strive at following "the dictum of the single variable." In order to do so, he must know his experimental system, including the environmental conditions of his animals, in every detail and duplicate the conditions to which the experimental animals are subjected as closely as humanly possible in the controls.

Diet, temperature, humidity, ventilation, illumination, caging, noise, handling and gentling of animals are factors which should be under continuous control in acute as well as chronic type experiments. One must have constancy of the exterior milieu so as not to disturb the homeostasis of the internal milieu, except by the experimental variable under study.

We do not know to what extent minor and uncontrolled variations in one or more of the environmental factors mentioned, may influence the performance of animals in a given experiment. It is because of this lack of specific information on several counts that one should control all known variables in the experiment. Otherwise, differences found between experimentals and controls may be ascribed to the experimental variable, while in reality the observed difference was mainly due to uncontrolled variations in one or more environmental factors.

Let us first consider housing and caging of germfree animals. The Reyniers type steel tanks (Figures 1, 2 and 3) used in our laboratory provide protection of the animals to air-contamination through a filter system in the inlet air and a germicidal trap for the outlet air. However, there is a rather brisk and steady flow of air (5 cfm) under slight positive pressure, which affects temperature, humidity and barometric conditions in the tank. Furthermore entry into the tank is limited to the glove ports and the autoclave route. The animals can thus only be handled by hands protected by thick rubber gloves plus cotton work gloves. The handling and fondling aspects and their possible influence upon the reactions and emotions of the animal are largely unknown as experimental parameters. We should recognize this fact and equalize conditions whenever possible.

With regard to caging, the limited space in each tank might tempt the investigator to use small restraining cages, and even to cram two or more animals into each cage. This is of course only permissible if controls are housed in an identical way, although there is usually no need for such extreme space economy in our animal rooms.

In many experiments, especially where influences of dietary factors are under study in germfree versus conventional animals, the temptation to house more than one animal in a cage should be overcome. If one animal dies and is cannibalized by the survivor, the experiment may be ruined. If the cage of the germfree animal is of a type which limits coprophagia, the cage of the control animal should be identical. The feces eaten by the conventional animal are not the same as those eaten by the germfree. The conventional feces contain bacterial body constituents, but even more important, vitamins synthetized by the bacteria of which the vitamin B-group may be the most important.

With regard to temperature inside the germfree tank, this is a function of seven factors: The temperature of the inlet air, the rate of air flow and the humidity, the temperature of the room in which the tank is located due to ready convection of room temperature through the steel walls, the illuminating lights, the animals own heat production and last, but not least, to heating incident to operation of the autoclave attached to the tank when entry or exit of material is necessary.

The tank temperature can be controlled within rather narrow limits by special devices; the point is that temperature variations induced in the germfree tank should be duplicated for the control animals at the same time. The marked influences of environmental temperature on a great number of biological phenomena are well known and need not be detailed here. It suffices to mention as examples (2) that growth rates, dietary requirements, physical activity, sexual cycles and functions, mitotic activity and renewal rate of the epidermis are all markedly influenced by the environmental temperature. Environmental temperature also affects survival rates following different types of trauma, like hemorrhage shock, tourniquet shock and burns. (3)

Although the sensitivity of animal functions is not as pronounced to changes in humidity as in temperature, major and uncontrolled variations should be avoided. The requirements for optimal levels of humidity, as for temperature, vary with age and species of animals. Temperature and humidity affect energy exchange in all warm-blooded animals. Particularly in the stressed animal and perhaps especially in burn studies, humidity and temperature control are mandatory. (4)

The illumination requirements of animals cannot be accurately defined today. A constant day-night cycle seems to be particularly important for rodents. Thus seasonal variation in breeding can be reduced or eliminated. (5) Illumination for paired experiments must be of the same intensity and wave length for it is known that light of different wave lengths has profound influences on adrenal functions. (6)

Noise as a potentially important factor is not well understood in its disturbing effect on animals in a secluded environment like the steel tank. All we can do, is to equalize the noise factor by the simple rule: if you bang the experimental tank A, bang tank B, housing the controls.

Diet is another very important factor needing control due to the special processing needed for germfree animals. It is evident that equal conditions for experimental animals and controls imply that both get the same diet. The diet for germfree animals is autoclaved prior to entry, and when distributed in the tank, it is not subject to attack and alteration by bacterial contamination. Not so for the conventional animals. Even if the diet is autoclaved under the same conditions as for the germfree animals, the similarity may end here. As soon as the diet is cooled and distributed to the conventional animals contamination with its manifold implications will take place. We do not know how to completely equalize the factors influencing the diet in experiments involving both germfree and conventional animals. To illustrate our attempt towards this end, some details from a current series of long-term experiments carried out in collaboration with NIAMD on germfree and conventional rats on a choline-deficient, cirrhosis producing diet will be briefly summarized.

The diet is mad up identically by the same person for three groups of rats, (1) germfree in sterile tanks, (2) conventional rats in nonsterile tanks, and (3) conventional rats in our ordinary rodent room. Ingredients, weighing and mixing, and sterilization procedures are identical. The water supply is identical for all groups, only canned U. S. Coast Guard Emergency water is used. Food is offered in equal amounts to all groups on Mondays, Wednesdays and Fridays.

It is evident that identical environmental conditions for germfree experimental animals and their conventional controls, apart from the presence of bacteria in the environment of the latter, necessitate that the controls are also kept in tanks in the same room. Air flow rate, pressure, temperature, humidity, illumination, handling and noise can thus with proper care be canceled out as experimental variables. A third group of animals, conventionals in ordinary animal rooms, should ideally be set up to distinguish differences in reaction to an experimental variable between conventional controls housed inside tanks versus controls in the animal rooms. Only by careful analysis of such triple-phased experiments can we learn more about the relative importance of the environmental factors discussed previously.

Now to the germfree animal itself. The main know physiological differences between the germfree and the conventional animal involve the cellular and humoral defense mechanisms, especially the roticuloendothelial system (RES), and as a corollary, certain of the plasma proteins; also the gut, especially the cecum of the rat and the guinea pig. (7,8,9)

1. <u>THE STATE OF THE CELLULAR AND HUMORAL DEFENSE MECHANISMS IN THE GERMFREE ANIMAL</u>. By definition, the germfree animal is free of demonstrable bacterial and fungal infections by the culture techniques used to establish germfreeness. The animal does not harbor parasites, as determined by fecal screening for eggs and parasites and careful autopsy. While most workers probably feel that exogenous viruses are not present in germfree animals, the situation is not clear with regard to viruses which may be transferred to the fetus in utero or (possibly) through the milk in suckling rats and mice born of germfree parents. This unsettled status of the germfree animal with regard to viruses is unfortunate if germfree animals are used in experiments designed to study development of tumors in cancer research, and of course, in experiments with viral agents in a presumably virgin organism. The absence of a live micro flora accounts for the unstimulated state of the lymphoid tissue and particularly for the low numbers of plasma cells seen in the tissues of the entire gut of germfree rodents and birds.

It is, however, important to realize that the germfree animal is exposed to antigenic challenge by foreign materials and that while his RES is underdeveloped anatomically and possibly functionally, it is certainly not dormant. Bacteria, and maybe viruses, are always present in the diet when prepared. Infectious agents are killed by autoclaving, but lipopolysaccharides and heat-coagulated bacterial proteins may enter the germfree organism and act as antigens. Protein material from the food itself is another source of antigens. While the supply of bacterial antigenic material must be substantially less in the gut of the germfree animal, the situation is not different with regard to antigens offered with the food itself. The underdevelopment of the RES refers particularly to the lack, or scarceness of, <u>nodular</u> lymphoid structures in the gut, while "free" or scattered RES elements, including plasma cells, are always found in the mucosa and submucosa to an extent of 10 to about 30 per cent of that seen in conventional animals of the same species. The status of the RES elements in the respiratory tract of the germfree animals remaind to be studied in detail.

The low intensity of challenge by RES-stimulating antigens must be kept in mind in the design of, and especially in the interpretation of, experiments involving traumatic procedures like hemorrhage, traumatic shock, radiation injury and burns. At our present state of knowledge, it is naive to interpret differences in survival or tolerance to any one of these procedures between germfree and conventional animals solely to the presence or absence of bacteria. In any situation involving tissue injury, the germfree animal must presumably be in a different position to take care of the consequences of massive cellular destruction.

At the present time, the conventional animal serves as a control for his germfree counterpart, or vice versa, only if a marked difference in two variables is accepted and taken into account in the interpretation of experiments involving tissue injury:

a) the conventional animal contains bacteria and,

b) has a normally developed RES.

If in the future one could achieve a "normal" development, at least in terms of tissue mass, of the RES in the germfree animal by nonspecific stimulation with one or more injected or fed antigens, experiments involving tissue injury may become more meaningful with regard to the effects of the crucial variable - the presence or absence of bacteria.

Another project which, if successful, will enhance the usefulness of the germfree animal as a tool of research, is the production of a nutritionally complete, wholly synthetic diet which is hypo-allergenic or, ideally, non-antigenic. Several laboratories, including our own, are presently engaged in this work starting from the soluble, synthetic diet of Greenstein and Birnbaum. (10) This type of diet will permit basic studies of immunologic and defense mechanisms, including the physiology and biochemistry of the RES under completely controlled conditions.

THE PLASMA PROTEINS. The concentration of gamma globulins and the carbohydrate-rich alpha globulins are lower in germfree animals than in their conventional controls. These proteins are synthetized mainly or exclusively, by cells belonging to the RES. When the germfree animal is challenged with antigenic material, especially live bacteria, the RES is activated, and in some weeks the plasma protein spectrum cannot be distinguished by ordinary chemical and electrophoretic methods from that of a conventional animal. (11,12,13)

THE GUT AND CECUM. Smaller villi and very scant development of lymphatic structures are characteristic of the germfree state. The most striking difference, however, is the markedly increased cecal volume in germfree mammals. The cecum with contents weighs on the average 3 - 5 times as much in most germfree guinea pigs. The cecal wall structures seem underdeveloped, thinner, and the water content of the cecal contents is higher. This finding has tentatively been interpreted to indicate active transfer of water from the plasma to the cecal contents, since the water content of the lower ileum fluid entering the cecum, is less, and not different from that found in conventional animals on a similar type diet. The large cecum gives rise to a rather high incidence of fatal volvulus, especially in the guinea pig. The trapping of substantial amounts of total body water in the cecal fluid is a complication in all experiments which will induce shock, for example, hemorrhage, tourniquet, and burns. An added control in this type of experiment may be cecectomized animals. Such preparations have been made successfully at the Lobund Institute by Dr. Gordon and his associates.

By way of closing the discussion of the many factors which need control in germfree research, we will give a brief account of an experiment which may turn out to be crucial in clarifying the alleged role of bacterial endotoxins as the agents which may be responsible for so-called "irreversible" hemmorhagic shock.

In every war, shock has been the major emergency complication of the wounded solider, and this problem will be even greater in any future war. Considerable delays in the treatment of civilian and military casualties caused by thermonuclear warfare must be anticipated. "Irreversible" shock will become a clinical problem of a magnitude never before encountered. (Irreversibility is a state of refractoriness to treatment in which the best available treatment fails to prevent or only delays circulatory failure and death). During the past 25 years, circumstantial but impressive evidence has accumulated which suggests that while lessened blood volume is the primary cause of shock, the development of irreversibility after severe hemorrhagic or traumatic shock is due to the entry of bacterial endotoxins into the circulation. This hypothesis, championed by Fine and his group in Boston (14,15), states that severe hypoxia in the bled, hypotensive and shocked animal, will lead to a breakdown of the normal gut-blood barrier to bacteria and endotoxins and allow absorption or entry of bacterial endotoxins into the circulation. Endotoxins from gram-negative bacteria normally present in the intestinal flora, will, when introduced into the circulation, augment the already severe arterial hypotension by vasodilatory effects and result in collapse of the circulation, followed by death. Furthermore, the RES in the shocked animal has a markedly reduced phagocytic capacity towards potentially harmful macromolecules like bacterial lipopolysaccharides. Therefore, circulating endotoxin in amounts which in the non-shocked animal will be readily taken care of by the RES, is now free to exert its deletary effect in the shocked animal.

Obviously, this hypothesis could be put to test in the germfree animal. Such experiments have been reported by McNulty and Linares, (16) at Walter Reed and Zweifach, et al. (17) working at Lobund. Both groups used the germfree rat and found no significant differences in survival rates of germfree and conventional rats subjected to identical surgical procedures. In other words, germfree rats subjected to a bleed-out procedure and maintained at a fixed low level of arterial blood pressure for four hours, will upon retransfusion of the shed blood recover or die in numbers which are no different from that observed in the conventional rats. Taken at face value, the inference would be that bacteria on their endotoxins are not involved in the irreversibility of hemorrhagic shock and death in the germfree rat, which dies with the same gross and microscopic anatomical lesions found in the conventional animals. Hemorrhage into the small gut and injury to the mucosa, are characteristic features of the autopsy findings in irreversibly shocked rats. On the basis of these experiments in germfree rats, one cannot, however, discard the endotoxin-hypothesis as disproved. Small amounts of bacterial endotoxins, arising from heat-killed bacteria in the diet, are undoubtedly present in the germfree rat, some may be stored in the RES elements in the mesenteric lymph nodes. This endotoxin may be released during hypoxia, and additional small amounts may be absorbed from the intestinal contents during the hypotensive period and not be taken care of by the RES.

The argument is that these small amounts of circulating endotoxin are enough to precipitate irreversibility because the germfree animal with his anatomically underdeveloped RES is less resistant to endotoxin.

Experiments by Dr. Einheber in our laboratory with injection of a purified E. coli endotoxin which in a sufficient dose will kill the normal and the germfree mouse and in lesser doses induce a period of prostration and hypotension, showed, however, that this germfree animal is no more sensitive to endotoxin than his conventional control. The matter rests here at the present time. Definitive experiments to test the hypothesis must await production of a truly endotoxin-free, germfree animal, maintained on the synthetic hypo-allergenic diet, with and without an artificially stimulated RES.

SUMMARY. The design problems inherent in research with germfree animals have been described, specifically in regard to peculiarities of the germfree environment and animals. Methods for control of environmental and physiological peculiarities which permit investigators to follow "the dictum of the single variable" are discussed.

Figure 1

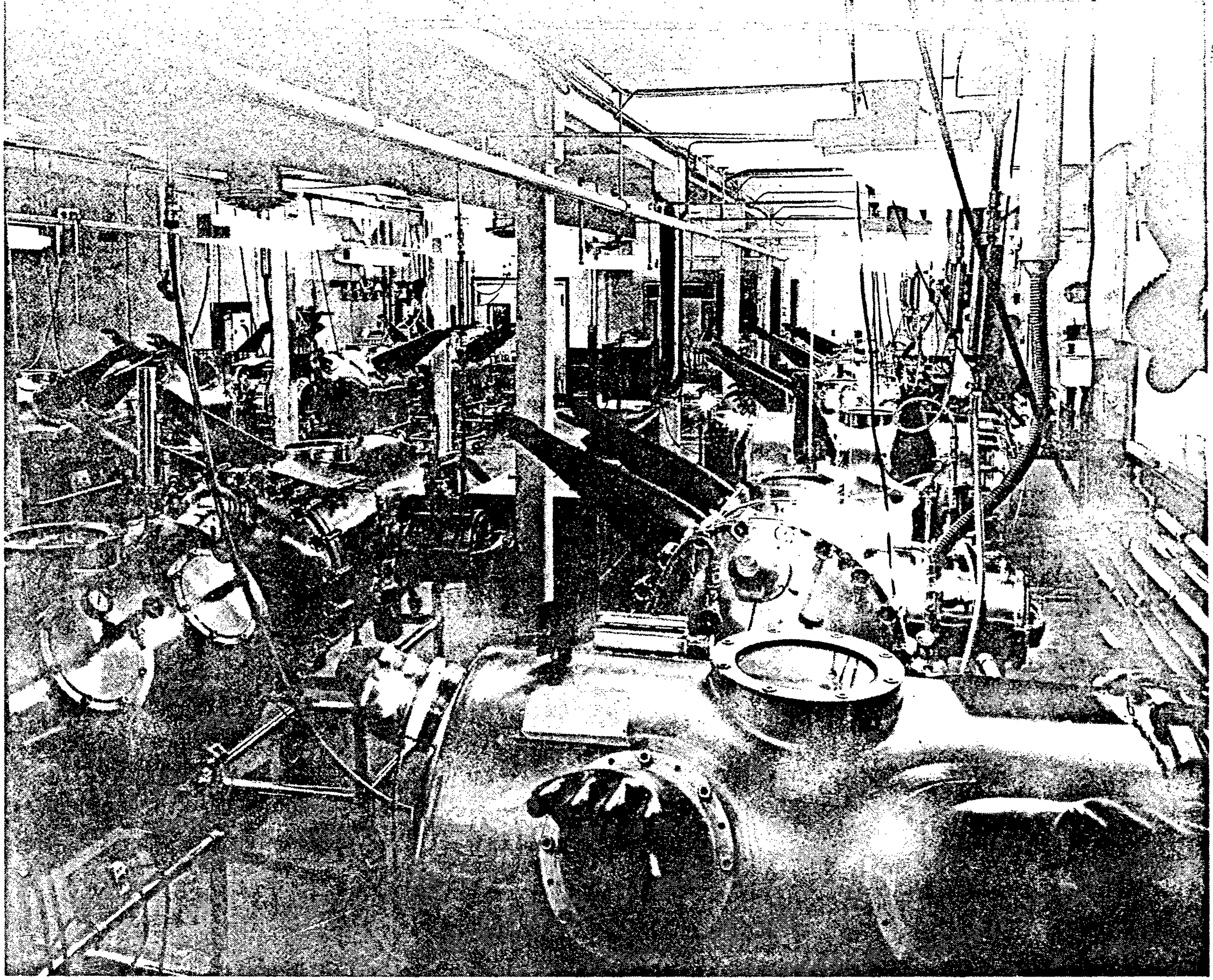


Figure 1

The tank room of the Department of Germfree Research at the Walter Reed Army Institute of Research.

Figure 2

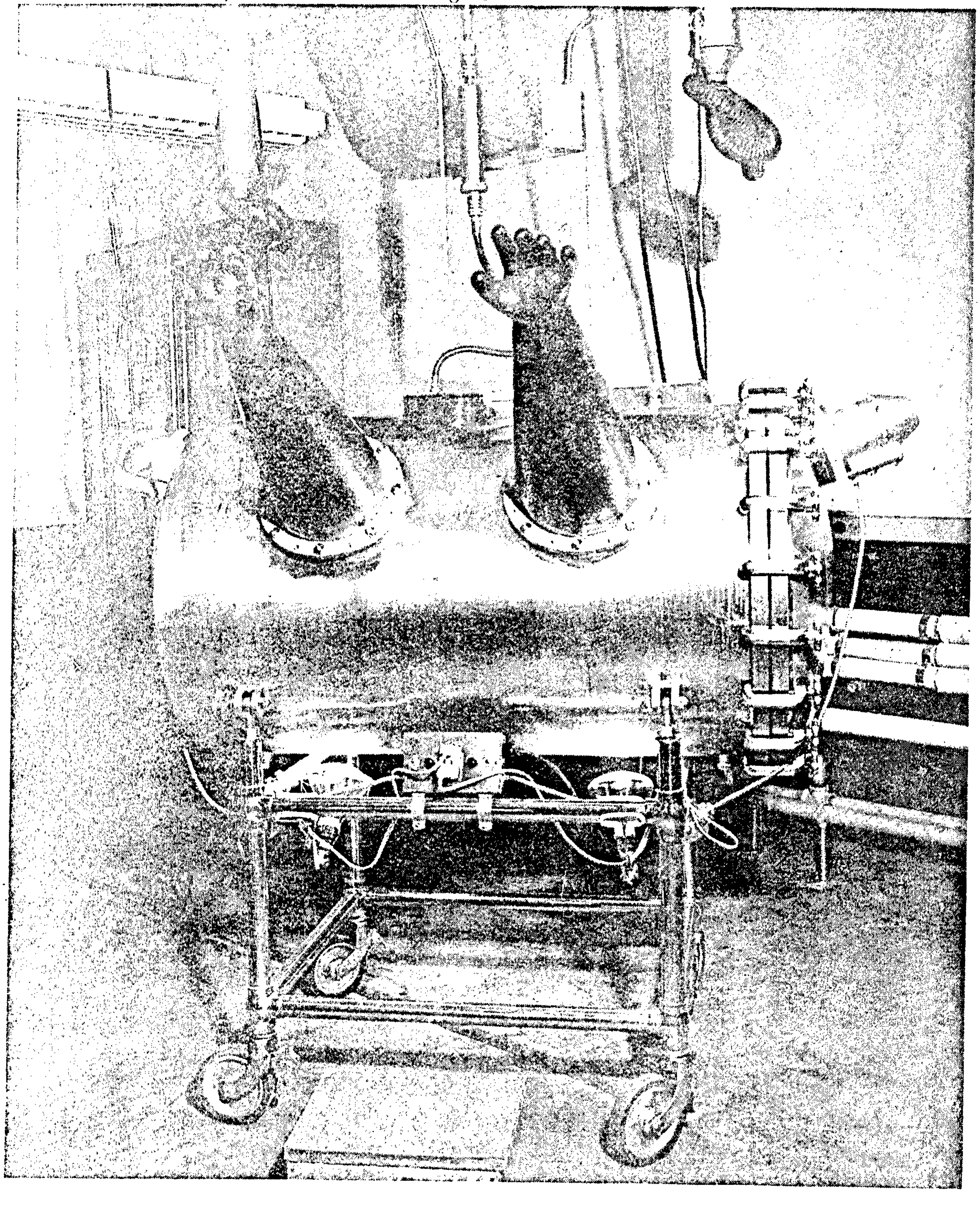


Figure 2

A Reyniers type heavy stainless steel germfree tank. The Department of Germfree Research at the Walter Reed Army Institute of Research uses 16 such one-man tanks as well as 4 similar tanks designed for two-man operation.

Figure 3

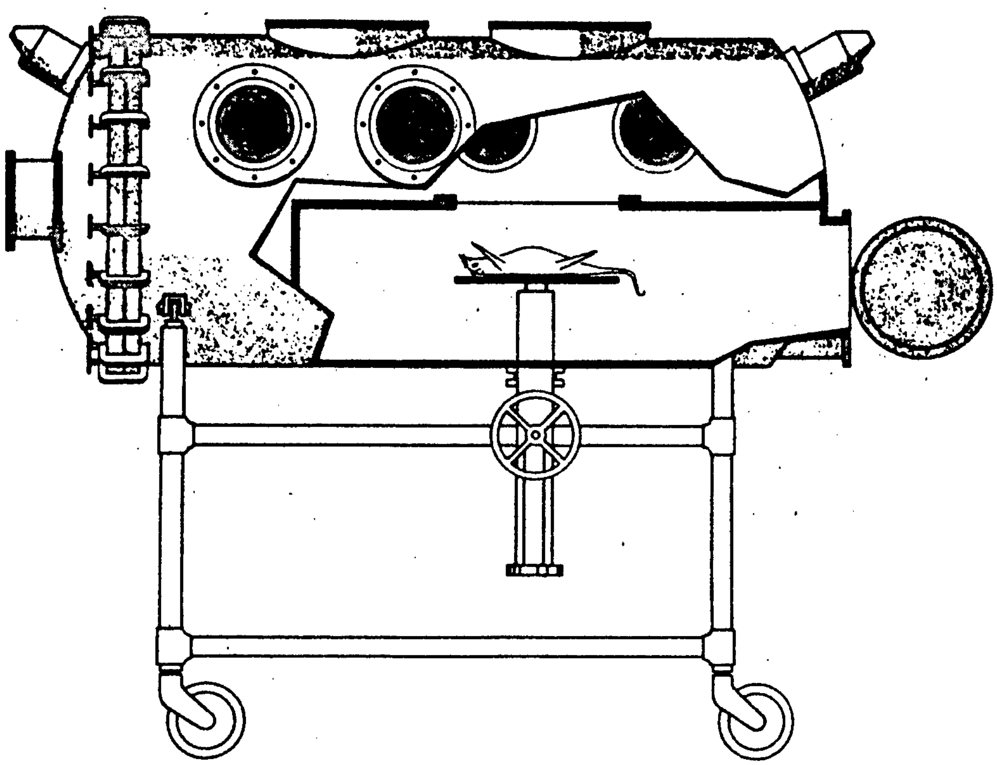


Figure 3

Diagrammatic cross-section of the germfree operating unit. The pregnant animal is introduced through the lower door into the lower section and is placed on the operating table. An elevator raises the table to the opening, which is covered by a sterile plastic sheet. Operator's hands are introduced through the glove ports while the view ports permit observation of the operative field. Caesarean section is performed by incision with a cautery through the plastic sheet, the heat of the cautery fusing the plastic to the skin. The uterus containing the young is excised and pulled into the sterile upper section of the tank, where the young are removed from the uterus, stimulated and breathing provoked.

REFERENCES

1. Levenson, S. M., Crowley, L. V., Horowitz, R. E. and Malm, O. J. The metabolism of carbon-labeled urea in the germfree rat. J. Biol. Chem., 234:2061, 1959.

2. Mills, C. A. Influence of environmental temperatures on warm-blooded animals. Ann. N. Y. Acad. Sci. 46:97, 1945.

3. Rosenthal, S. M. Experimental chemotherapy of burns and shock. Publ. Health Rep., 57:1923, 1942.

4. Pence, D. and Lindsey, D. Effect of high environmental humidity on survival of mice following an experimental burn. Am. J. Physiol., 195:719, 1958.

5. Alexander, D. P. and Frazer, J. F. D. Interchangeability of diet and light in rat breeding. J. Physiol., 116:50, 1952.

6. Gillissen, G. Uberblick und Versuch einer morphologischen Objektivierung sensorischer Einflusse auf dem, Organismus Arch. Hyg. 137:335, 1953.

7. Gordon, H. A., Doll, J. P., and Westmann, B. S. Effects of the "normal" bacterial flora on various morphological characteristics of the animal host: A comparative study of germfree and normal stock chickens, rats and mice. Anat. Rec., 130:307, 1958.

8. Wostmann, B. S. Serum proteins in germfree vertebrates. Ann. N. Y. Acad. Sci., 78:254, 1959.

9. Thorbecke, G. J., Gordon, H. A., Wostmann, B. S., Wagner, M. and Reyniers, J. A. Lymphoid tissue and serum gamma globulin in young germfree chickens. J. Infect. Dis., 101:237, 1957.

10. Greenstein, J. P., Birnbaum, F. M., Winitz, M. and Otey, M. C. Quantitative nutritional studies with water-soluble, chemically defined diets. I-V. Arch. Biochem. Biophys., 72:396-456, 1957.

11. Wostmann, B. S. and Gordon, H. A. Changes in the serum protein pattern of germfree rats upon exposure to a conventional bacterial flora. IV. Internat. Congr. Biochem., Vienna, Sept. 1958.

12. Gustafsson, B. E. and Laurell, C. -B. Gamma globulins in germfree rats. J. Exp. Med., 108:251, 1958.

13. Gustafsson, B. E. and Laurell, C. -B. Gamma globulin production in germfree rats after bacterial contamination. J. Exp. Med., 110:675, 1959.

14. Fine, J., Frank, E. C., Ravin, H. A., Rutenberg, S. H. and Schweinburg, F. B. The bacterial factor in traumatic shock. New Engl. J. Med., 260:214, 1959.

15. Fine, J., Rutenberg, S. and Schweinburg, F. B. The role of the reticuloendothelial system in hemorrhagic shock. J. Exp. Med., 110:547, 1959.

16. McNulty, W. P., Jr. and Linares, R. Hemorrhagic shock of germfree rats. Am. J. Physiol., 198:141, 1960.

17. Zweifach, B. W., et al. Irreversible hemorrhagic shock in germfree rats. J. Exp. Med., 107:437, 1958.

THE DEVELOPMENT OF PARAMETERS FOR DETERMINING THE RESISTANCE OF SELECTED MISSILES COMPONENTS TO MICROBIOLOGICAL DETERIORATION

C. Bruce Lee
Physical Sciences Laboratory, Research and Engineering Directorate, OTAC

The recent development of the Military Missiles Program in this country has necessitated a re-evaluation of the procedures in practically all phases of microbiological research, development and testing. This fact has been brought about by the number and complexity of the new materials employed in missiles, the peculiar designs and engineering of the components, and the problems of storage and ultimate operational requirements.

Deterioration microbiologists engaged in military activities realize the importance of the preceding statements and have found it necessary to develop for missiles research new parameters for testing. Further, there has been a need to adapt and re-evaluate those already in use, and to undertake research in order to assure to the manufacturers of missiles and missiles components that microbiological deterioration, specifically fungus action, will not be a factor in the malfunction of missiles once they are operational.

Missiles, missiles systems and their components are unusual in that there is a unique interdependence of items upon each other and all materials incorporated into a system must be verified absolutely reliable if the missile is to be, and remain, a tactical item. Thus, assurances of reliability must be secured by undertaking microbiological aging and deterioration testing in order to assure that there will be no difficulties traceable to the deteriorating effects of fungus action in the manufacture, storage and operation of the items.

For those present who may be unaware of the national program on military microbiological deterioration, a brief recase since its inception in the early period of World War II will be given. With the outbreak of hostilities, and movement of conflict to the tropics of the world, the military establishment suddenly found itself confronted with a monstrous problem; biological in origin, in which fungi, minute plants, were actually ruining and rendering unserviceable millions of dollars worth of critical materials by a natural ability to utilize in their metabolism the substrates supplied in the composition of military materials.

These fungi, the minute plants, are incapable of performing photosynthesis and, thus, they differ from the large familiar green plants which can make their own food. The species of fungi that are of concern to the military are usually microscopic or barely macroscopic in detail and all must secure their nutrition from pre-formed sources. I imagine that there are many here this morning who have vivid memories of food and clothing spoilage during tours of duty in the tropical areas of the world. The majority of the deterioration fungi reproduce most commonly by spores. These are shed into the atmosphere, the soil or water and, being easily air borne, they come to rest on a host of materials. If the material is susceptible to fungus growth, growth will proceed from the substance of the material substrate which is the pre-formed food necessary for fungus metabolism.

Usually, hydrocarbons and various minerals are the most easily metabolized sources of nutrition. Whatever the available nutrition, however, fungus growth on materials partially or completely destroys the material. Growth may be surface, or it may proceed internally, with the fungi producing thread-like mycelium, the first indications of fungus growth and presence.

During the war, fungus action was reported on a large number of items ranging from optical instruments to textiles. Most important, the spectrum of materials available for attack was almost entirely natural-in-origin, and included items which were cellulosic, proteinaceous or possessed animal or vegetable fats in their compositions. Control, was perforce, expedient and necessitated the complete discard and replacement of affected components, or the application of crude, surface-applied fungicides which often ruined serviceability of items by altering the physical or chemical properties to such an extent that the concerned material was rendered useless for military applications.

As a result of these experiences, the government entered the field of microbiology and sponsored basic and applied research on the control of fungus attack of military materials with the result that over the years, numerous tests have been developed which are capable of laboratory application in specification procedures. Particular efforts have been made to include specific tests for specific items. The resulting specifications invariably designate certain strains of species of fungi which have proven superior ability in degrading particular types of materials on the basis of origin and composition. For example, reference to any fungus test specification will reveal the prescription for the use of a single or species in tandem, and which may be cellulolytic, proteolytic or lipidophylic in degrading ability.

During the war years the employment of various synthetics in military items was begun, and since the cessation of hostilities, this use has expanded until currently, the role of synthetics in military goods far exceeds the natural-in-origin products in many items. At the beginning of the government efforts in the control of fungus deterioration, there was scant concern with possible fungus utilization of ths synthetic products; it being assumed generally, that these were inherently resistant. However, it was not long before the first incorporate uses of synthetics into military items that testimonies from the services revealed the fallacies of this assumption. Evidence was presented that synthetics were often excellent metabolic sources of nutrition for fungi with resulting alterations in physical and chemical properties. Conferences by government microbiologists on this problem resulted in a common approach to the entire field of fungus attack on materials and resulted in the following conclusions for national use:

1. In instances where materials have been found susceptible to utilization by fungi, such should be withdrawn from use and substitution accomplished employing funginert materials. The wide diversity of presently-available synthetics makes this possible.

2. Superior design and engineering of items must be employed from the concept stage until final manufacture and should take advantage of primary and continuing advice and suggestions of the microbiologist in order to eliminate loci of possible fungus utilization in any part of the completed assembly. (See Illustration 1 on the following page.)

Basically, these two suggestions have been followed by military microbiologists and the most important and pressing problems have been mainly solved or controlled.

Aside from basic and applied research, the military microbiologist functions currently as a consultant in development and testing. Further, suggestions are made concerning materials use and advice is given on design of components to withstand microbiological attack. In instances where there is no inert substitute for a susceptible item, advice is rendered regarding the use of possible fungicides. The present list of these chemicals is immense compared to years past and many have been developed for particular needs and uses. Contemporary fungicides rely on incorporation or compounding into a product, as well as on surface application. They take advantage of chemical and physical properties with a minimum of alterations to an item's characteristics.

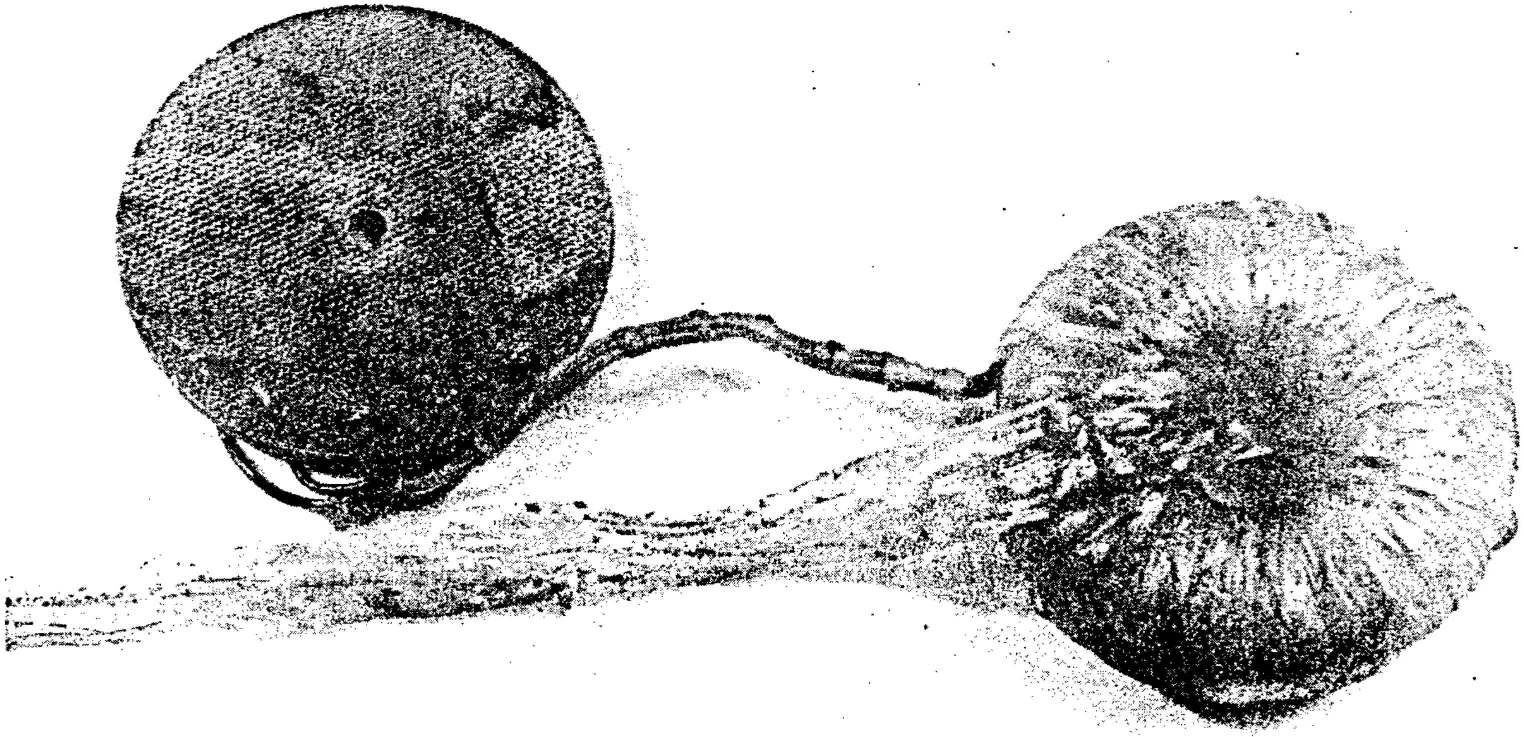
The requests of missiles manufacturers to our installation for information relative to their products' fungus resistance introduced a new phase of testing. Previous activities had been concerned with our mission for tank and tank-automotive vehicles and equipment. Usually, these materials were tested in part and a whole assembly was rarely submitted, although our facilities are geared to accommodate a six-by-six truck. Because of a strategic location in the automotive development center of Detroit, our organization was confronted suddenly with missiles measuring nearly sixty feet and with diameters from five to eight feet. The speaker is still amazed with the first request from a missile manufacturing service, "Can you expose this to fungus attack." <u>This</u>, being a missile nearly sixty feet long!!!!! (Illustration 2).

Missiles and missiles components submitted to fungus exposure were the Redstone, parts of the Jupiter and the entire Honest John. In addition, this organization had knowledge of research performed on the NIKE at a private installation.

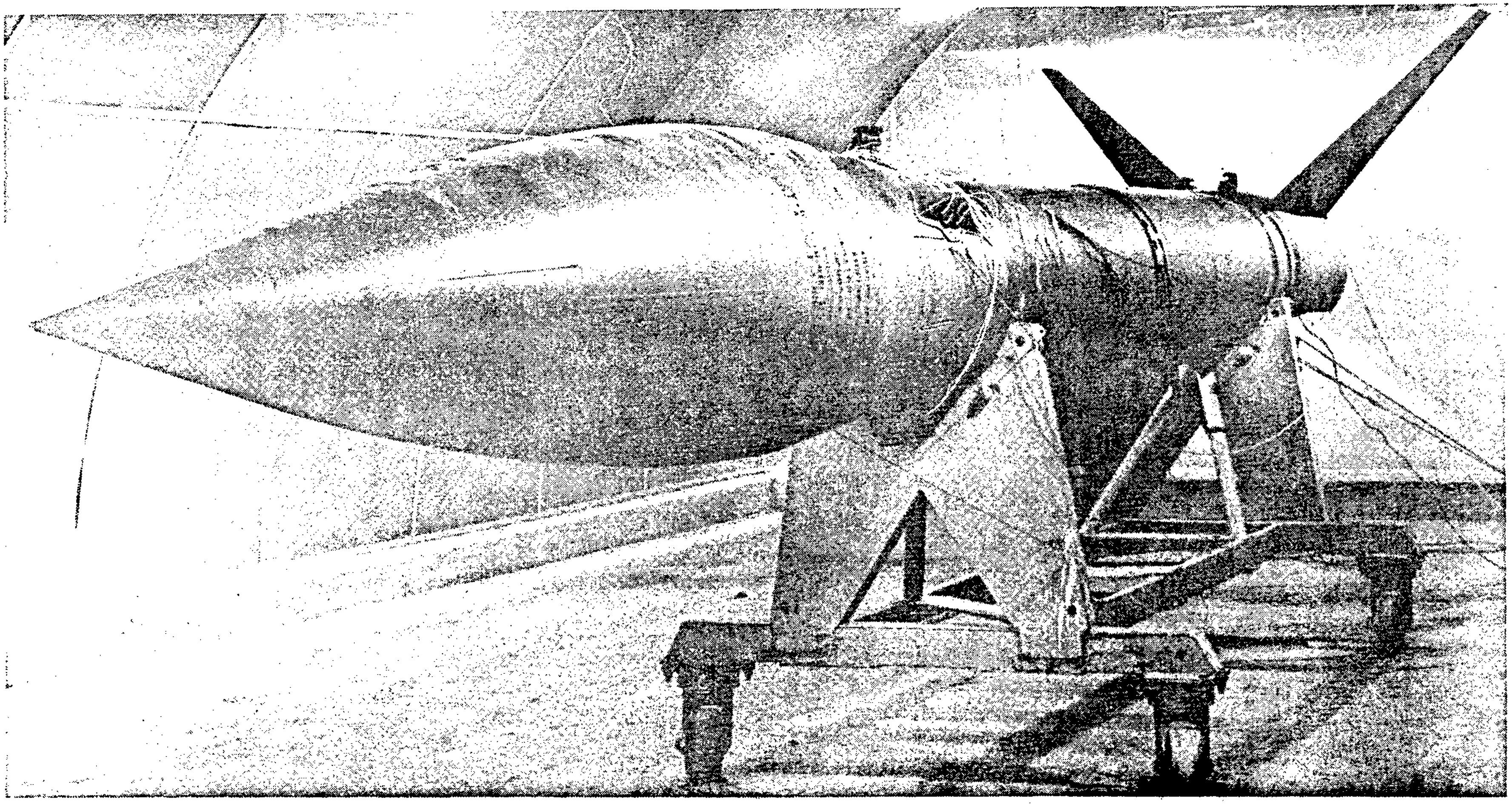
Literature surveys of the available specifications revealed no references specifically concerned with missiles, or the great variety of unique materials incorporated into these tactical weapons. Thus, our problem for the past few years was clearly indicated; the development of test parameters that would define the behavior and resistance of missiles and their components to fungus attack.

In addition to the whole or disarticulated missiles, information on fungus resistance was also desired for a large heterogeneous selection of materials, parts, partially-assembled systems and standard and specialty items. Many of these materials are synthetic in origin and prime in their use on military missiles components and were originally employed as it had

Illustration I



T-2 transformer from the 1800 VA inverter from the Redstone Missile, CCMD. Silk winding removed in part to reveal the presence of fungi on the silk.



Honest John Rocket, Douglas, positioned within the Tropical Chamber at the White Sands Missile Range, New Mexico. Photograph included gives an idea of the size of missiles submitted for microbiological exposure.

been found that the natural-in-origin materials did not, and could not be expected to function adequately under the special conditions of temperature, humidity, etc., introduced by the storage and intended operation of the missiles.

To determine the procedures necessary for performing microbiological research on missiles, certain objectives were proposed for the investigations and these are:

1. What is the overall role of fungi in the utilization of missiles materials?

2. Is the missile item upon which fungi are growing actually being degraded and rendered unfit for service, or is the growth adventitious?

3. What items, or parts, are susceptible to fungus utilization and what are resistant?

4. What species of the lower fungi are to be indicated as materials degraders and are the materials susceptible to only one species, or is their utilization by several or more species?

5. What methods can be devised to demonstrate the effectiveness of protection and corrective measures necessary?

The narrative of the paper this morning will interpret the experience in the laboratory on missiles research in the light of the preceding points.

Further, in order to establish parameters for undertaking the research it was necessary first to arbitrarily limit the parts of the missiles which would concern the microbiologist. This was accomplished by an overall inspection at the site of manufacture. Some of the missiles tested were large and others easily accommodated into our testing facilities.

The inspection established the first parameter of testing; the fact that our research would be limited and conducted on the tail sections of the missiles. These are the parts containing the motor and control instrumentation, as well as the electrical connections. Samples of materials used in other parts of the missiles were requested and conclusions also submitted on their behavior to microbiological attack. Limitations in the size of the missile parts for testing were dictated by the accommodations available.

One of the important parameters in the fungus investigations was the choice of the testing situation. In previous experiences involving tank and tank-automotive components, all testing was accomplished in the laboratory and involved various pieces of environmental equipment. Owing to the size and diversity of materials used in missiles, decision was necessary as to testing site. Previous experiences of the speaker have indicated that wherever possible, it is more advantageous to employ the natural situation. According to location and program financing, various installations of the country have secured data on the microbiological resistance of components in the field using such places as the tropical rain forest, the savannah, the desert or shore locations; places in which the temperatures and humidities are varying optimal for the development of the lower fungi in demonstrating their ability to degrade materials. When the missile research at Detroit first commenced, it was decided to press for running the research in the actual tropical rain forests available in Puerto Rico, or the Panama Canal zone. However, as a result of financial limitations on funding, the Detroit group was forced to confine work to the Detroit area and to arbitrarily choose a parameter of our own devising, the simulated tropical conditions afforded in the use of the tropical room.

The Detroit tropical room has been employed over a period of eight years for automotive testing and has been developed to attain conditions of humidity and temperature that are simulations of nature in offering optimal conditions for fungus development within the room and on materials placed into the room for evaluation. The room is a large structure, 20 feet long by 15 feet wide and with 9 foot ceilings and 8 foot access doors.

The conditions of temperature and humidity are original with the Detroit group and were determined on the basis of data available from the meteorological records of the rain forests of the world.

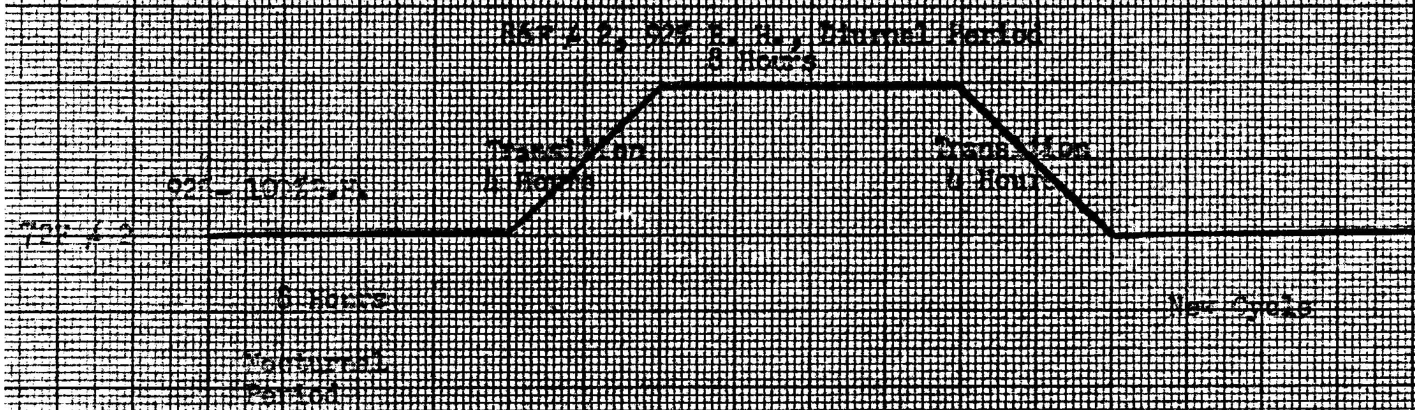
The simulation of conditions in the room, a phase of the parameter of the testing situation, resulted in a four cycled 24 hour day. There were eight hours of diurnal conditions with the temperature at 86 F $\pm$ 2 and the relative humidity at 92%; a four hour crepuscular period for transition during which the temperature and humidity were altered to assume the nocturnal conditions of the tropical rain forest and the temperature at 72 F $\pm$ 2 and the relative humidity 92% to saturation. The nocturnal period was followed by another transition crepuscular interval and the cycle resumed. (Illustration 3).

Fungus population and activity within the tropical room was assured using banked beds of soil, decaying leaves, rotting cardboard, rotted equine and bovine feces and the walls were hung with untreated canvas duck. The atmosphere was circulated using fans and examined bi-monthly employing petri dishes of nutrient agar to define species population. The choice and adaptation of the room to missiles investigations was supported by data from previous experimental testing at Detroit and also from information forwarded to this installation from other places with similar equipment. (Illustration 4).

The use of the cycled atmosphere was evaluated by comparison and it was determined that with its use a greater number of species would be noted than using the room with constant temperatures and humidities.

In addition to the Tropical room, moist chamber cabinets were also used in the investigations because of the large amount of materials received. The constant situation substantiated the employment of cycled conditions with the wider spectrum of species and deterioration results than would be noted under the stable situations.

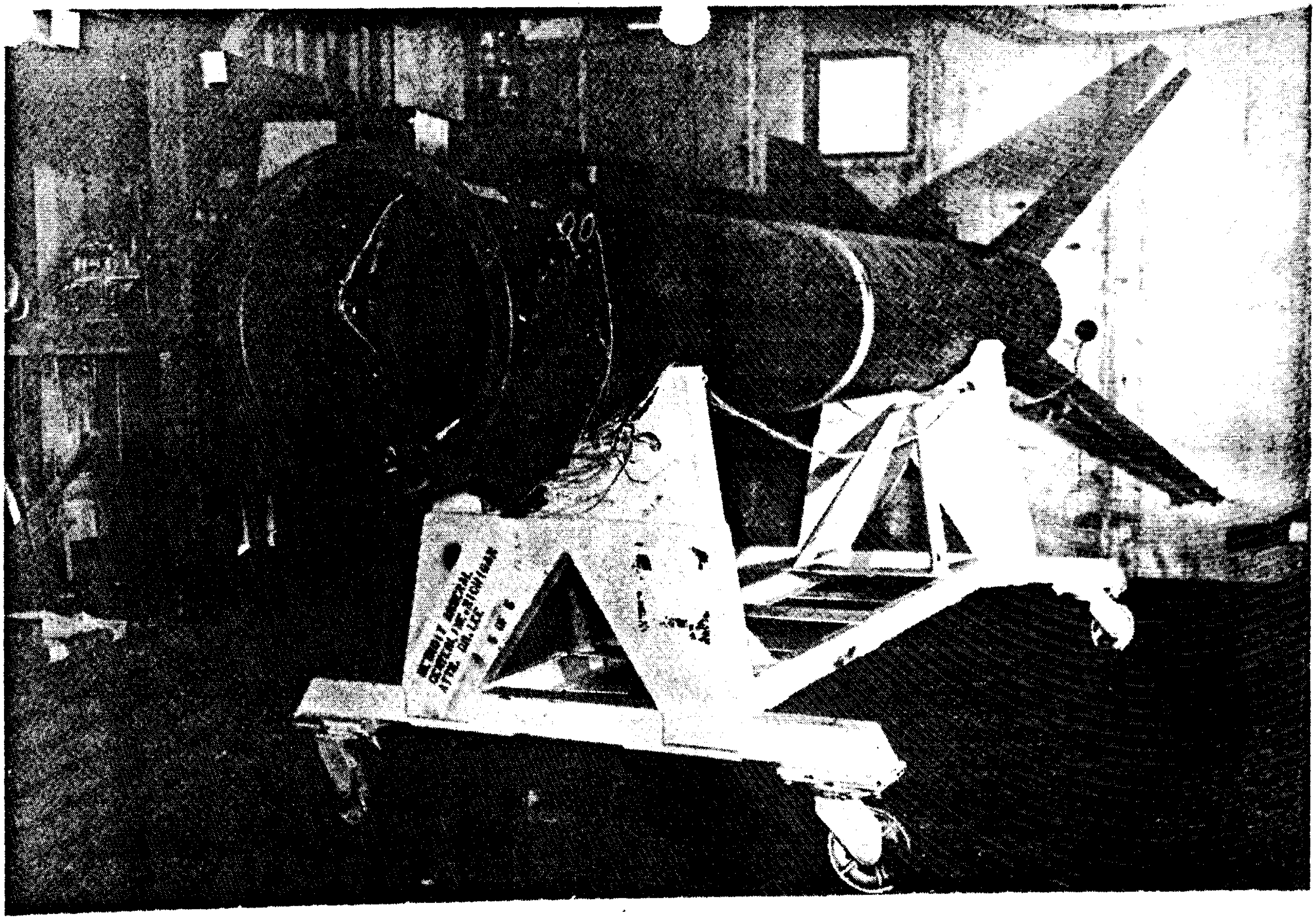
Illustration 3



24 Hours Temperature and Humidity Cycle in Tropical Room

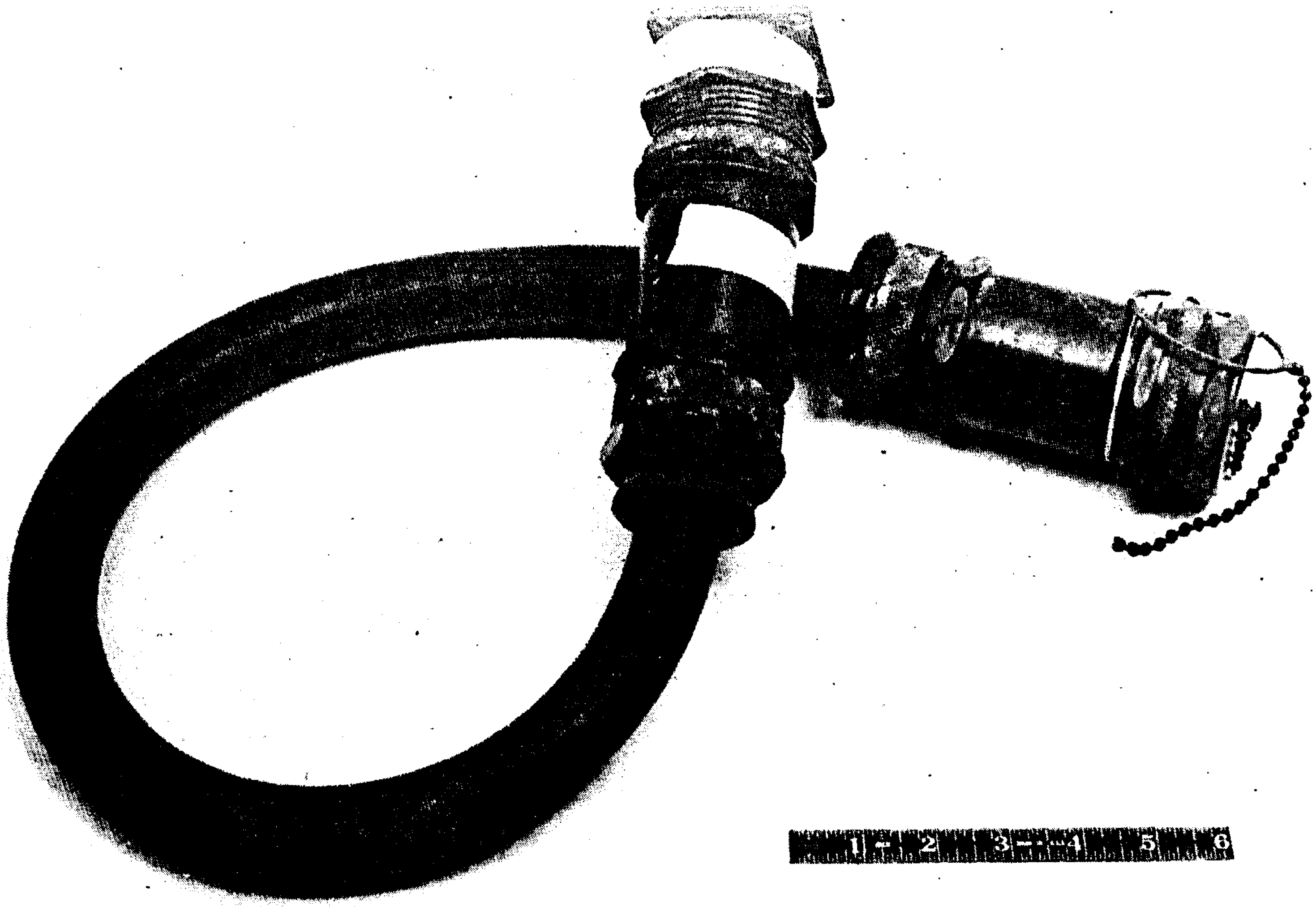
Graph of the twenty-four cycle of temperature and humidity conditions in the Tropical Room of the Detroit Arsenal - OTAC organization. An illustration of the diurnal, crepuscular and nocturnal conditions.

Illustration 4



The Honest John Rocket positioned for microbiological exposure in the Tropical Room at Detroit Arsenal - OTAC. Illustration includes racks holding components of the Redstone and also shows soil banked against the walls of the room.

Illustration 5



Multiconductor cable, FT-1, showing the importance of choosing funginert materials for inclusion in missiles. Cable photographed at the conclusion of 90 days of exposure within the Tropical Room at Detroit Arsenal - OTAC. There were no variations in pre-fungus and post-fungus exposure performance ratings.

The choice of testing situation indicated the third parameter to be followed in the microbiological testing of missiles; the importance of non-treatment of materials prior to testing. Past and present specifications often required a pre-treatment of materials be washing, placement into water baths with adjusted pH and temperatures, and the use of chemical cleaning, etc. All of these presented artificial barriers to securing accurate estimations of materials to fungus action. Would not it be more realistic and revealing of the actual resistance of materials to fungi if there was no pre-treatment of surfaces and compositions in any manner? Thus, for the eleven proposals of testing, no pre-treatment was employed, the materials being placed into the testing situation as received. (Illustration 5).

This parameter was not unique with missiles, but it was made official by inclusion into the missiles testing proposals and was chosen from data secured from testing on tank and tank-automotive vehicles. Employment of this parameter goes back to the idea of simulating in the laboratory as closely as possible, the conditions that would actually attain in storage or ready operation.

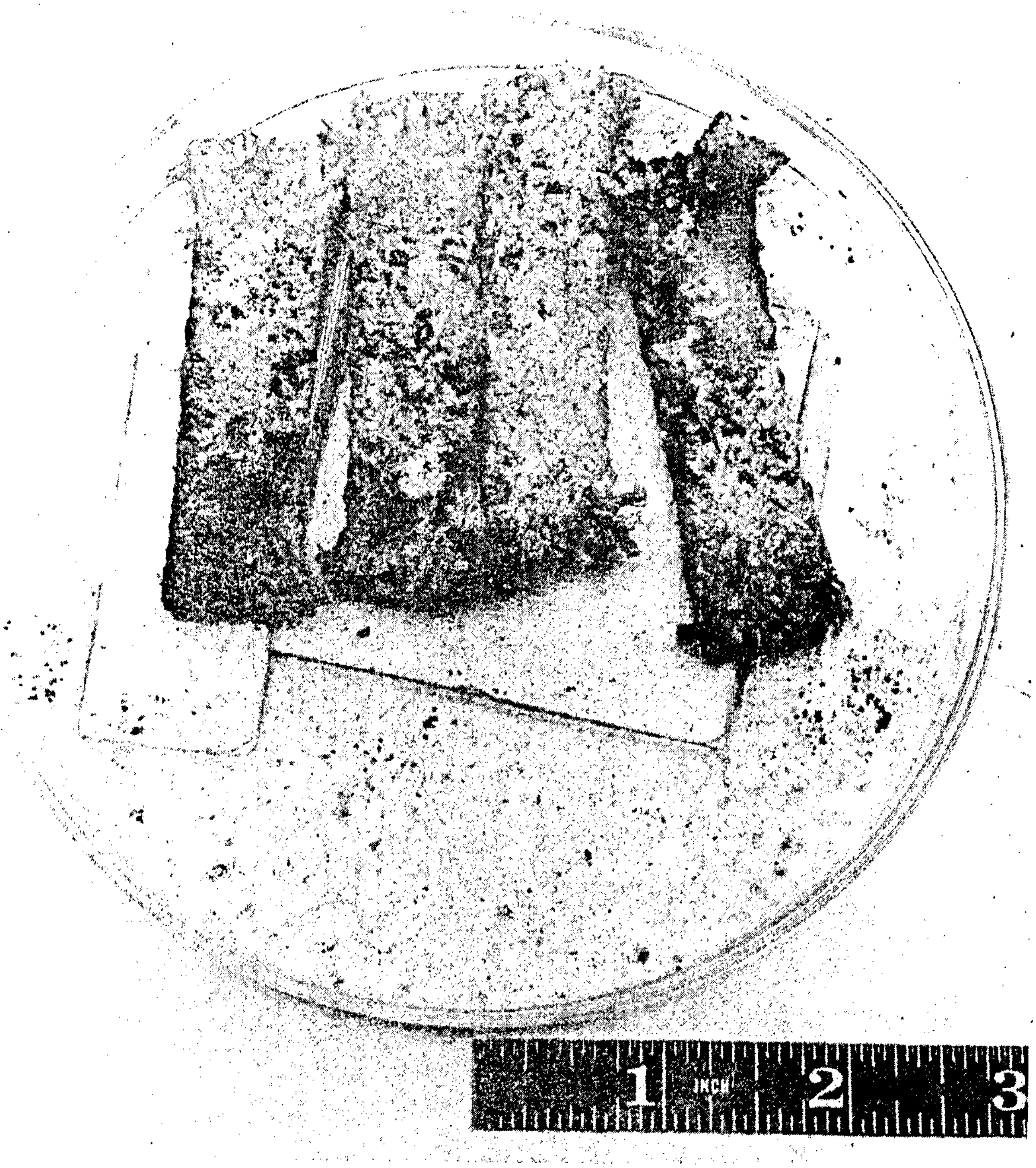
The fourth parameter developed for missiles testing was the choice of species of fungi. Again, in this matter, we relied on the data from past research and testing. However, since the majority of missiles items submitted for testing were new and unique, efforts had to be expended in securing information on composition of materials. Those which were cellulosic were inoculated with cellulolytic fungi; those proteinaceous with proteolytic species, etc. In the use of the various synthetics, the knowledge of the chemical composition was germane. In instances where it was impossible to define a material as to composition, a wide spectrum of fungus species was employed and the species mixture inoculated onto the material under test. At the conclusion of testing, observations were conducted to identify the fungi still evident and this information served as supporting evidence of actual utilization of the material. (Illustration 6).

The fifth parameter developed for the missiles research was the determination of the testing time. This is a crucial point and has been a concern of the Detroit organization ever since microbiology was established as a function.

Early specification tests prescribed a testing period of seven days. Later ones called for fourteen, with rarely twenty-one days. Since the first published specifications, testing time has gradually lengthened to ninety days. At Detroit, the shorter periods were viewed as unrealistic for producing data for estimating the effective resistance of materials' microbiological deterioration. Accordingly, over the years this laboratory has extended gradually the testing time of all components from thirty days to forty-five, to sixth, and currently ninety days. Only with the use of the longer period, it is felt, that we shall have a parameter to determine sufficiently the true behavior of materials to fungus attack. (Illustration 7).

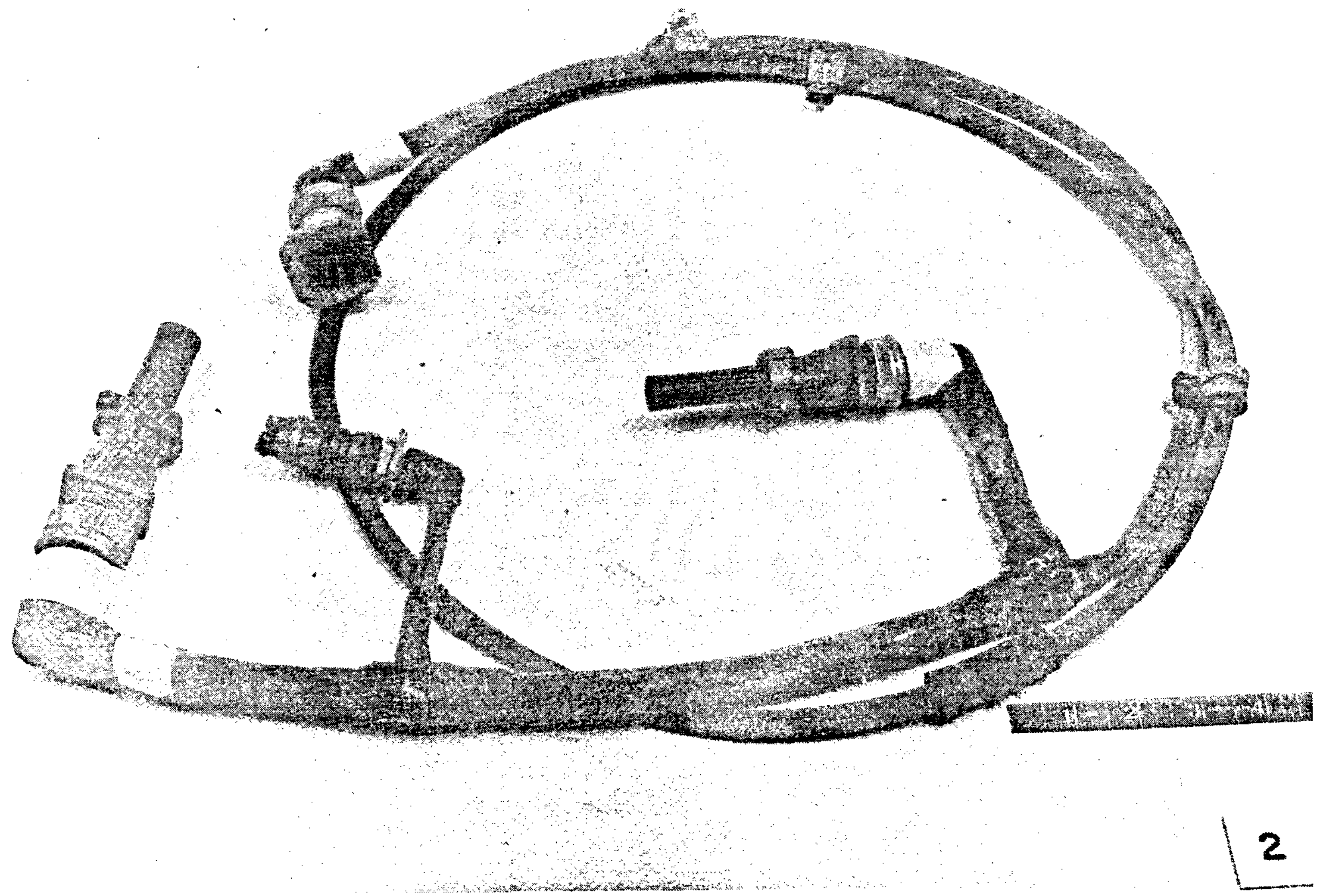
Support for the longer missile testing period relied on data secured on many dissimilar materials. In missiles work, it was noted that many of the items required a period for becoming conditioned to the atmosphere of the

Illustration 6



Asbestos samples adulterated with cotton. The cotton, being cellulose, supports a heavy growth of fungi while the funginert asbestos remains free of microbiological growth. Photograph taken at the conclusion of 90 days exposure in the Tropical Room at the Detroit Arsenal - OTAC.

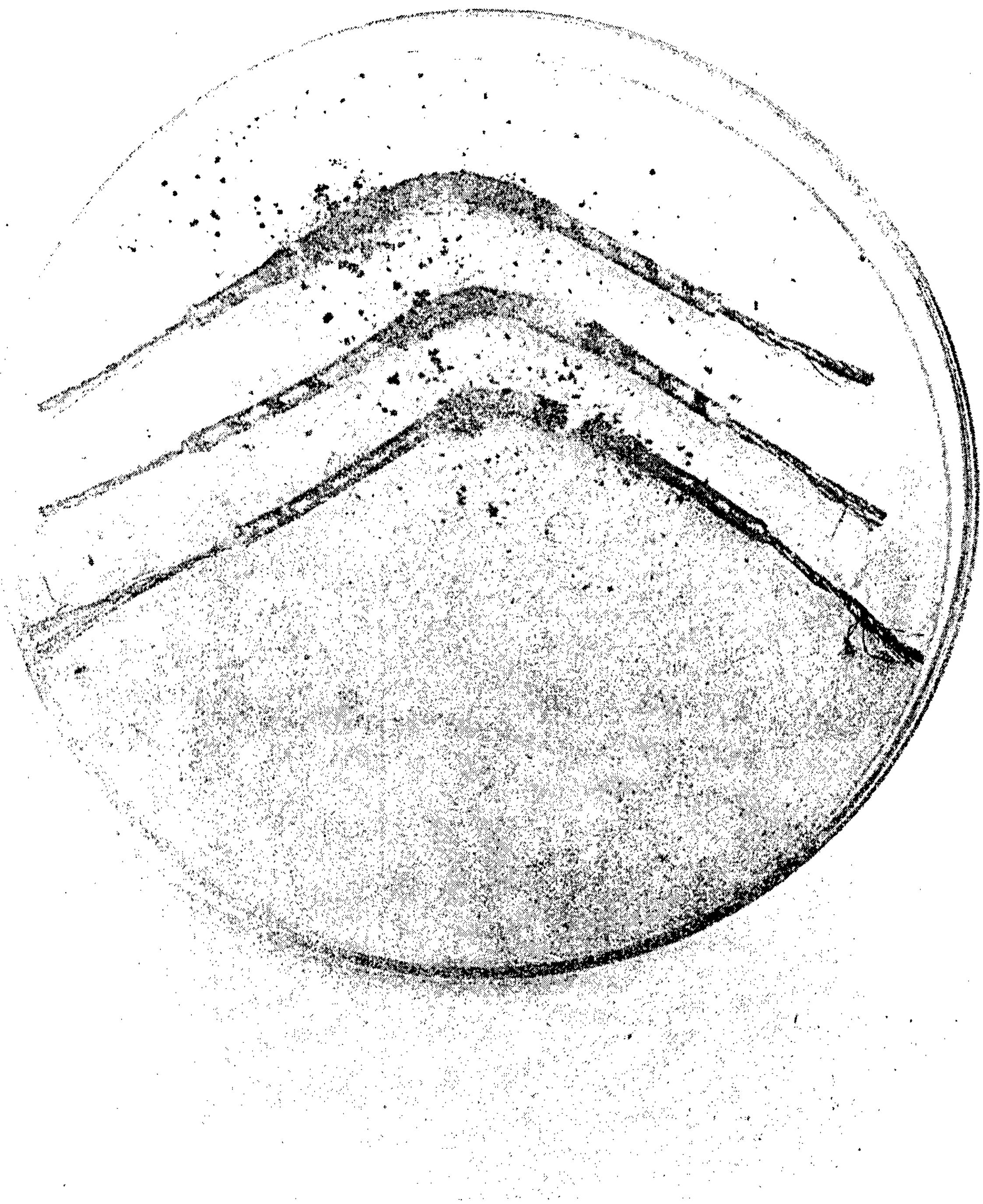
Illustration 7



2

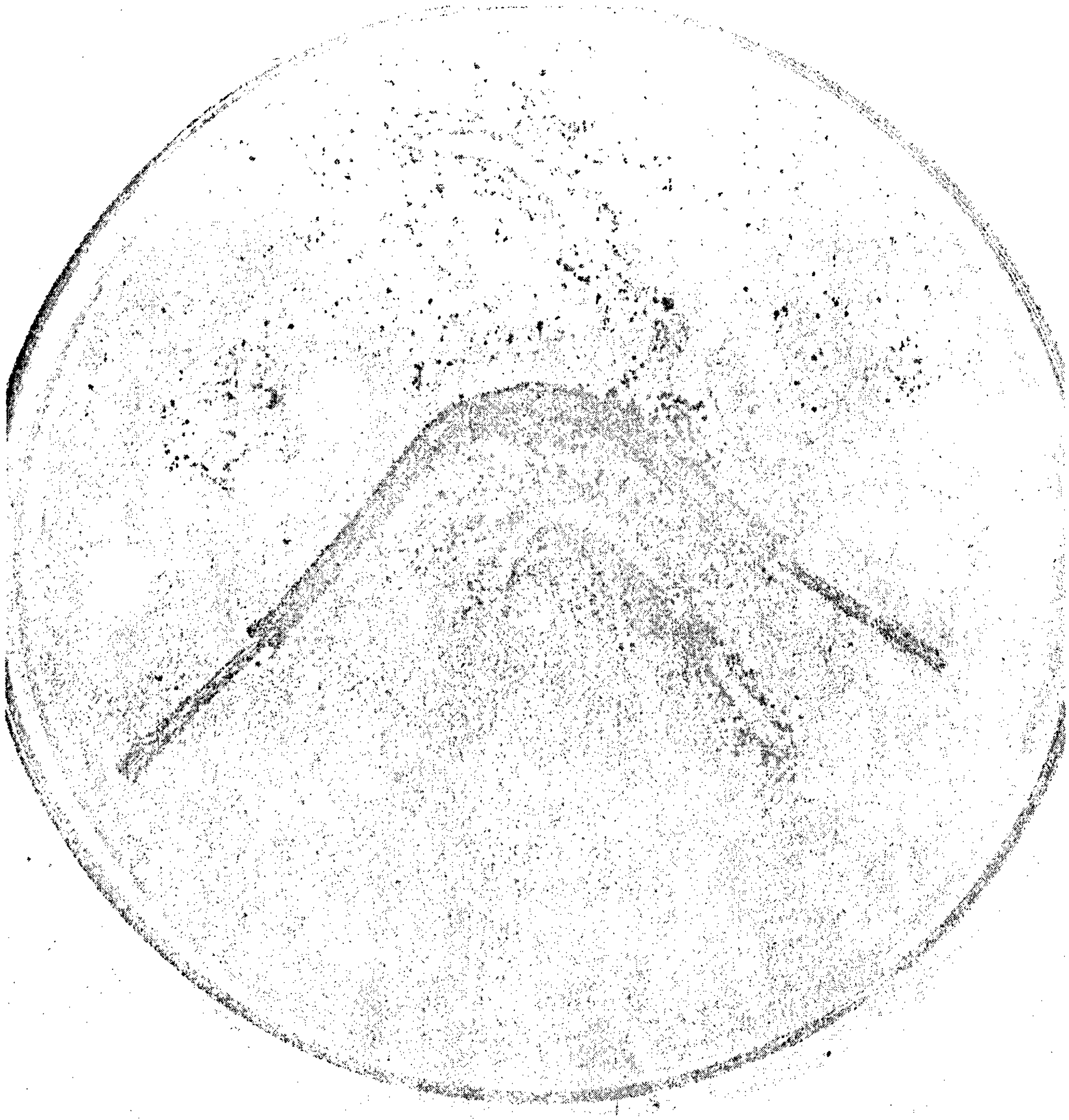
Cable harness following ninety days incubation with fungi in the Tropical Room of the Detroit Arsenal - OTAC. Cable insulation completely degraded and performance rating failed to meet the requirements for the component. Choice of inert materials is requisite.

Illustration 8



Cable leads following thirty days incubation in the Tropical Room of the Detroit Arsenal - OTAC organization. Both material resistance to fungus growth and performance ratings failed to meet minimum standards.

Illustration 9



Cable lead at the end of ninety days incubation in the Tropical Room of the Detroit Arsenal - OTAC organization. Insulation completely degraded and performance failed to meet minimum requirements for the material.

Illustration 10

Summary of the weekly reports of a selected group of miscellaneous components used in the Redstone Missile, CCMD. Report covers a three months period and demonstrates the weekly changes in observations. Visual observations were also supported by performance ratings, where possible, at the thirty and sixty day periods.

testing chamber. Examination of missiles and components in storage for fungus development supported this contention. This adjustment period varied from thirty to sixty days and in that time there was often little development of fungi on missiles surfaces. However, once the materials were conditioned to the testing chamber, fungus growth proceeded rapidly and often apparently, instantaneously. This was particularly true of rubbers, plastics, and some of the miscellaneous components, assembled and disassembled components. The facts of the longer testing period accounted for the discrepancies also noted in our results as compared to other installations also performing tests on the same materials, but using shorter testing times. Fungi are living organisms and they all possess a threshold, above which, they cannot be stimulated to grow more rapidly. (Illustrations 8 and 9).

An adjunct of this longer testing period was the information found at the conclusion of the ninety day testing period. Materials removed from the testing chambers and placed onto tables out in the laboratory responded by developing different growth patterns, species of fungi developing, and loci on materials that were being utilized. This would not have been attained in the shorter testing periods.

An additional factor developed for demonstrating missiles resistance to fungus was the writing of the actual testing procedures. This was done whether the missiles were tested in toto or, disarticulated with parts disassembled. Testing followed basically the steps outlined in a specification developed in our laboratory and modified specifically for missiles and taking into consideration the parameters mentioned this morning. Consideration also had to be given whether missiles components were supplied as sealed or unsealed in an effort to eliminate a consideration for corrosion damage owing to moisture and which might have been primarily determined microbiological. (Illustration 10).

The use of the performance tests at this installation has been a parameter that has been pioneered at this place. All missiles materials, as received, were inspected for applicable, possible performance tests. Basically, these tests are demonstrations of the physical, chemical or mechanical properties and included data on strength, electrical conductance, depolymerization and visual evidence of changes such as complete or incomplete rotting, embrittlement, softening, bubbling, bleeding out of chemicals, crystallization of materials' surfaces, etc.

The use of the performance test was augmented and verified by the use of the periodic performance ratings secured from materials over the period of ninety days. These periodic tests were conducted within the tropical chamber so as to take advantage of the temperatures and humidities that would be found in the storage situations in the field. Further, these periodic tests verified the parameter of increased testing time. Often a material would fail within thirty days, while another would fail in sixty or at the terminal ninety days. The use of the periodicity in testing allowed for savings in money and testing time by defining the exact time in which a material failed. (Illustrations 11 and 12).

This installation always requires a sufficient number of samples in order to allow for the periodic performance testing of items from preinoculation to final evaluation.

Fungus attack of missiles and missiles components is usually evident as surface growth on the various components. At our laboratory, the materials were separated prior to testing into coarse assemblages based on common characters. For example, we received:

natural and synthetic rubbers

electrical components, assembled

electrical components, unassembled

miscellaneous components containing plastics, finishes and textiles

coverings, insulations and gasketing

metallic units with organic-in-origin parts

single and multiconductor cables.

Fungus growth was noted on many of the preceding. However, it was not employed as a definitive parameter without the supporting data from other parameters previously mentioned. Visual evidence is deceptive and decision is required whether the growth is adventitious or deleterious. Using the performance test, the latter is easily accomplished and data based on changes in physical properties such as losses in tensile strength, powdering of surfaces, scuff resistance alterations, loss or increase in adhesion; chemical tests with alteration in composition or electrical measurements of changes in current carrying capacity. All of the preceding, of course, require a comparison with pre-fungus test data in order to have a comparison with the post test ratings.

All of the information presented this morning has been considered in forming conclusions on the importance of fungi as deteriorating agents on missiles and missiles components. Further, the results of our investigations indicated that control of microbiological deterioration is necessary in order to eliminate fungi as possible causative factors of malfunction from the manufacture to ready storage and ultimate operation.

Illustration 11

Table I – SUMMARY OF WEEKLY REPORTS (Continued)

		Asbestos Tape	Cotton-Vinyl Tape	Wool Felt	Wool Felt	Asbestos Binders	Insulation Sleeving	Silicone Compound	Pressure-Sensitive Tape	Silicone Compound	Cotton Tape	Vinyl Tape	Foam Rubber	Vinyl Coating
Sample Number		14	15	16	17	18	19	20	21	22	23	24	25	26
Report Period	Exposure (Days)													
14-18 Jul	7	*	S	*	M	*	M	*	*	*	*	*	M	XX
21-25 Jul	14	*	SM	*	SM	M	SV	*	*	*	S	S	MS	XX
28 Jul 1 Aug	21	*	*	S	SV	M	E	*	*	*	S	S	SV	*
4-8 Aug	28	*	*	S	SV	M	E	*	*	*	S	S	E	S
8-14 Aug	35	*	SM	S	E	M	E	*	*	*	S	S	E	M
14-21 Aug	42	*	M	M	H	M	E	*	S	*	S	S	E	SV
21-28 Aug	49	*	H	M	H	M	E	*	S	*	S	S	E	E
29 Aug 4 Sept	56	*	H	H	H	M	E	*	SV	*	S	MS	E	E
5-11 Sept	63	*	H	E	H	M	E	*	SV	*	S	E	E	E
12-19 Sept	70	*	H	E	H	M	E	*	SV	*	S	E	E	E
19-26 Sept	77	*	H	E	H	M	E	*	SV	*	S	E	E	E
27 Sept 3 Oct	84	*	M	E	H	*	E	*	SV	*	SM	E	E	E
4-10 Oct	91	*	M	E	H	*	E	*	SV	*	SM	E	E	E

XX - Sample not available for starting date; however, growth was extensive at the end of the test.

- * No fungus growth
- S Slight fungus growth
- SM Slight-to-moderate growth
- M Moderate growth
- MS Moderate-to-severe growth
- SV Severe growth
- E Extensive growth

Magnetic counter, solenoid coils, from improperly sealed component. Photograph taken at the conclusion of ninety days and indicates the importance of correct sealing of components in the elimination of fungus attack of material.

Illustration 12

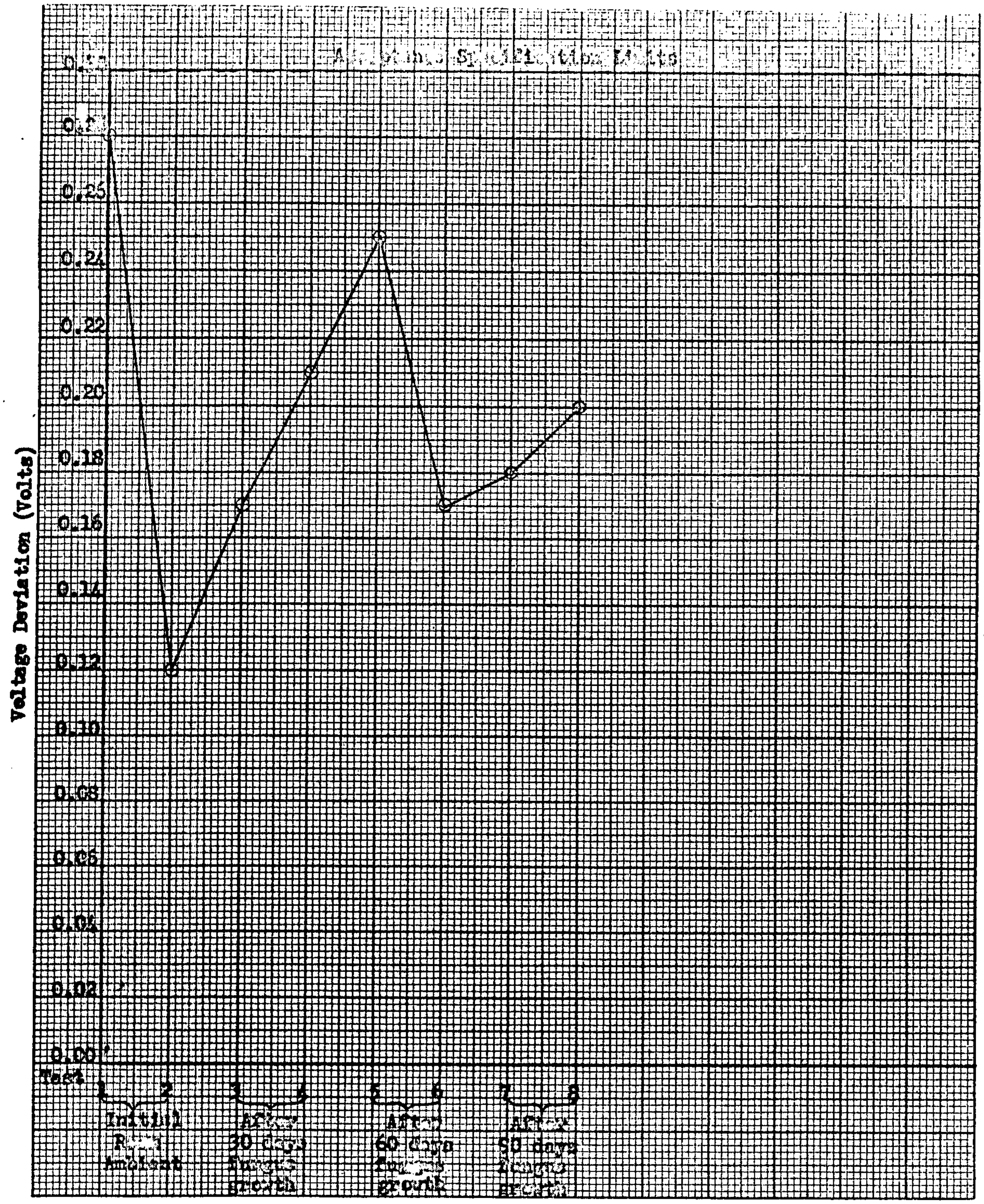


Figure 7 - Voltage deviation for constant-input, constant-load test (two-hour period)

The importance of performance testing before, during and at the completion of microbiological exposure. Voltage deviation

DESIGN OF ENVIRONMENTAL EXPERIMENTS AND RELIABILITY PREDICTION

A. Bulfinch
Feltman Research and Engineering Laboratories
Picatinny Arsenal

. . . ———————— . . .

PREFACE. This paper demonstrates how accepted statistical techniques stand to reduce the cost of testing missiles and missile components. These techniques are not treated in full; however, examples of their use and references are given.

The subject matter covered is oriented towards practising reliability engineers. It is hoped that some of their troublesome problems have been clarified.

ABSTRACT. Describes the need for, and use of factorial designs in surveying the separate effects of a large number of environmental treatments with maximum reliability and a minimum number of test specimens. Shows how information from the factorial experiment can be used to define reliability, and how this information can be used in tests of increased severity to predict "reliability-in-use."

SUMMARY. Reliability is defined.

The need for testing to failure is emphasized by comparing construction engineering problems with missile reliability problems. Methods are given to convert "safety margins" to measures of probability which can be used to predict "reliability-in-use."

The advantages of using factorial designs to survey the separate effects of a large number of environmental conditions are described. Information obtained from factorial experiments is used to formulate the definition of reliability in terms of the severest environments. These environments are used in tests of increased severity. Tests of increased severity are used to establish the relationship between use conditions and test conditions in order to predict "reliability-in-use."

Detailed examples are given of methods of predicting high "reliabilities-in-use" with small sample sizes.

CONCLUSIONS.

1. Any particular component can have many reliability values simultaneously. There is one reliability value associated with each possible combination of environmental condition and measurable functioning characteristic.

2. The results of factorially designed environmental experiments should be used in defining component reliability.

3. Tests of increased severity should be used in combination with factorial designs to predict "reliability-in-use" from test results.

4. Tests of increased severity can demonstrate high "reliabilities-in-use" with small sample sizes.

RECOMMENDATIONS.

1. Factorial designs should be used in combination with tests of increased severity to predict "reliability-in-use."

2. The test condition used should be experimentally determined.

3. Military standards should be revised to permit the experimental determination of the test conditions to be used.

4. The terms on which reliability is defined should be experimentally determined.

. . . ______________ . . .

I INTRODUCTION

A General

The statistical aspects of reliability are not new. All of the necessary concepts are adequately treated in modern statistical literature. The lack of information about measurable characteristics of the missile system and the environment it experiences in use, as well as the high cost of test specimens, have created the current problems.

Urgently needed are highly efficient, experimental techniques that can be uniformly applied by various segments of the same organization and by different organizations. Highly efficient techniques are required because of the need to demonstrate very high reliabilities with very small sample sizes. High reliabilities are required, of course, to assure succesful functioning of complex systems composed of many components. Only small samples can be used because of the high cost and/or scarcity of test speciment. Uniform, standardized procedures are required for the collection of comparable data.

B Purpose

This paper purposes to do the following:

1. Define component reliability.

2. Describe how factorial designs can be used to survey the effects of several environmental conditions (with minimum sample size) preparatory to defining reliability in terms of these conditions.

3. Describe how "reliability-in-use" can be predicted from results of laboratory tests of increased severity.

4. Describe how tests of increased severity can use information from factorial experiments.

C Scope

The accepted statistical definition of reliability follows: Reliability is "the probability of successful functioning in use." This is a general definition that is applicable to any operating system. To define component reliability, the general definition must be modified to include:

a. Environmental conditions under which successful functioning took place.

b. The component characteristics that functioned successfully.

This means that every component can have as many reliabilities as the total number of possible combinations of environmental conditions and component characteristics. To have the weapon system reliability meaningful, the component reliabilities should be defined in terms of the most severe conditions so that the stated component reliability will be the minimum.

D Background

The techniques described, the most efficient known, are designed to maximize the amount of information obtainable from a given sample size. In addition, these techniques are definitive enough to serve as standard procedures throughout the same or different organizations over an extended period of time.

The uniform applicability of these techniques is as important as the efficiency. A large part of the value of experimentally determined reliability data is their scope of applicability. That is, reliability data collected by means of standardized procedures are cumulative in the mathematical sense. Hence, the precision with which reliability values are known can be improved with time as additional data are collected. This makes it possible to collect a reference file of reliability data on a variety of standard components.

II STATEMENT OF PROBLEM

In the usual case, the development engineer has one or more items to test under many different environmental conditions. The items to be tested may have two or more properties that must be evaluated.

Objectives of the usual reliability experiments for newly developed items follow:

1. Determine how well the engineer has succeeded in developing a reliable item.

2. Obtain an unbiased estimate of the "strength" (i.e., ability to withstand stress) of the item with minimum cost.

3. Determine the separate and combined effects of the environmental treatments on the reliability of each property measured.

4. Determine the effect of the length of time under the separate and combined effects of the environmental treatment on the reliability of each property measured.

5. Predict the "reliability-in-use" from the test results.

Objective 4 requires life-testing techniques which have been treated extensively elsewhere. This paper describes other techniques for increasing test severity.

The multiple properties of an item can be measured simultaneously. This poses no particular problem. The real problem confronting the engineer derives from his having different components that must be treated with a large number of different environments on a very limited budget.

Since most components are unique, they must be treated as separate problems. There is no known way to combine different kinds of components into a single integrated experiment. However, a single integrated experiment can be designed for several kinds of environments. As a result, the problem is one of designing the most efficient experiment for a single type of item and repeating the process for each type.

It is assumed that in every missile component there exists a true but unknown "strength," created by the particular design developed and used by the engineer in building the component. It is further assumed that the true "strength" is a constant and not a random variable for any particular design over short periods of time.

The present practice of designing components to pass the current military standards without failure does not attain the intended objectives for the following reasons:

1. An unbiased estimate of the true strength of the item cannot be obtained unless some items fail.

2. Cost prohibits testing all items at the same level of severity.

3. Reliabilities are demonstrated only in proportion to the number of items tested.

Testing without failure all items at the same level of severity can lead to completely erroneous conclusions. For example, in comparing two designs, there may be available more test specimens of the poorer design. Tests by military standards result in an equal number of failures for the two designs. Under these conditions, experimental evidence favors the poorer design.

Instead of subjecting all test specimens to the same test conditions regardless of their intended use, the level of severity of the test should be progressively increased until a failure is obtained. This procedure will lead directly to an unbiased estimate of the true strength with a minimum number of test specimens and establish the correct level of severity for each type of component at which the failure rate should be determined. This procedure will correctly differentiate between different designs.

Experience in the construction engineering field has shown that assurance that an item will not fail in use requires a large "safety factor" to be built into the item. To determine the "safety margin," the load applied must be increased until the test item breaks, or fails in some other manner. This is, in effect, a test of increased severity that leads directly to an estimate of the <u>true</u> breaking load the engineer is seeking to determine. The average, and standard deviation of only a few (3-6) such results are all that are required, because each value so obtained is an estimate of the true value. The difference between the "observed average breaking load" and the "load expected in use" divided by the "load expected in use" is called the "safety margin" or margin of safety. The larger this value is, the "safer" the engineer feels in predicting that the item will do the intended job without failing.

The construction engineer could have elected to load each test specimen with only the load expected in use, but because he has designed the item to withstand the load expected in use, this procedure tells him nothing about the true breaking load. All he learns is what he already knows--that it will not fail! Now if he wants to "feel safe" in predicting invariably successful functioning of the item, he must test <u>many</u> items. Asked to conduct his test in this manner, the engineer would rebel because he knows, as we do, that--it is far too costly!

In missile component testing, we should simulate the procedure used by the construction engineer-"load the item until it breaks" and then calculate the "safety margin." To do this, we must shift our attention from finding the number of items to be tested without failure at a single level of severity to <u>finding the level of severity that will cause failure</u>, and then finding the <u>failure rate</u> at that level. That is, to find the reliability (the probability of success) we must first find the failure rate (the probability of failure). To do this with small sample sizes, we must use a test of increased severity to find the level that will cause failure.

Robert Lusser, formerly of Redstone Arsenal, has advocated this approach of "testing to failure" for some time (Ref. 1). However, he neither showed <u>how</u> to "load" missile components until they fail, nor predicted the "reliability-in-use" by means of the laws of probability. He was satisfied with using "margins of safety" (Ref. 2).

III APPLICATION OF PROBABILITY LAWS

An item will not fail until the applied stress exceeds the item's "strength." If the "strength" is much greater than the stress expected to be experienced in use, the chance (probability) of failure in use is very small, and the chance of success (reliability) is very high. It is in this sense that "high reliability" is defined. That is, high reliability means high probability of successful functioning under actual use conditions; it does not mean high reliability under the test conditions.

To translate the reliability demonstrated under test conditions to a "reliability-in-use" value, the relation between the "use and "test" conditions must be known. Experience has shown that this relationship can be adequately represented by frequency distributions. This places the relationship on a probabilistic basis, and makes possible the use of the laws of probability.

When the average of the conditions in use is known, the level of severity required at which items must be tested to demonstrate any given reliability can be calculated in advance. As a result, reliabilities can be correctly predicted with small sample sizes without testing to failure. Alternatively, when the ultimate strength of the item is desired, the first failure method described below can be used. Both of these procedures predict "reliabilities-in-use" with small sample sizes through the use of the multiplication law which states that the probability of simultaneous occurrence of two independent events equals the product of the probabilities of separate occurrences of the events. Examples of both of these procedures are given below.

In reliability testing, the two simultaneous events referred to are test specimen failure and test condition causing that failure. Both the "failure rate" and the "chance" of the test condition's occurrence in "use" can be considered probabilities. By the above law, the predicted failure rate in use will be the product of the failure rate obtained in testing and of the chance that the test conditions could occur in use. The predicted "reliability-in-use" will then be equal to one minus this product.

When nothing is known about the environmental conditions expected in use, or when these conditions may vary in an unpredictable way, no prediction can be made about a unique "reliability-in-use." However, these methods of testing to failure can still be used to advantage. Knowing how much punishment components can take before failing reduces the number of unknowns, can be valuable in deciding how, and under what conditions, a particular missile can be used. This information can be useful in choosing between missiles for particular purposes. Moreover, where nothing is known about the conditions in use, the "most severe use condition" the item can be subjected to for any specified reliability can be calculated. An example is given below.

The methods described below show how missile components can be loaded (tested) to failure through the use of tests of increased severity. These methods also show how safety margins can be used to predict "reliability-in-use."

IV LABORATORY TEST METHODS

It is assumed in these methods that the test item can fail in but one way; that is, the binomial distribution is applicable.

A Factorial Designs

Plans should be made to conduct the laboratory experiments in two stages. First, survey the separate effects of the several environmental conditions of interest in one integrated experiment. The two-to-the-n^{th} factorial designs or their optimized modifications are the most efficient for this purpose. These designs can be used to select the treatments causing the highest failure rates. These treatments can then be used to define the reliabilities of the test item. If the reliabilities determined in terms of these treatments are acceptable, the reliabilities of the test item in terms of any of the other treatments will also be acceptable. This procedure will reduce the magnitude and complexity of the experiments conducted to determine and predict reliability. More importantly, component reliabilities obtained in this manner will furnish a more realistic basis for calculating systems reliability.

See references 5 and 6 for available designs. These designs are the most efficient known. Experiments based on these designs may be conducted without changing the treatment procedure except to arrange for the test specimens to receive the number and kind of treatments required by the particular design used. However, the best differentiation among treatments is obtained when the level of severity used will cause 50 percent of the test specimens to fail.

For the purpose of this application, the two levels of each treatment can be the <u>presence</u> and <u>absence</u> of the treatment. Alternatively, any two levels of the treatments can be used.

The number of test specimens required in the optimized designs is one more than the total number of treatments used (Ref. 5). The more versatile fractional factorial designs (Ref. 6) require at least 16 items for experiments containing from five through eight treatments, and at least 32 items for nine through 13 treatments. With twice these numbers of items, the latter type designs can also measure interactions--how the effect of any one environment depends upon the others. Interactions among treatments cannot be measured by any design except the factorial.

Factorial designs permit a type of statistical analysis that distinguishes between variations due to chance and variations having assignable causes, and produces more information from a given number of items than any other known procedure. These designs actually increase the effective sample size by making it possible to use each observation (or measurement) for more than one purpose. In fact, each treatment effect is determined as though the entire experiment is conducted to determine that particular treatment effect alone. As a result, the reliability with which <u>each</u> treatment effect is determined can be based on the total number of items used in the epxeriment. The three-treatment-design example described below demonstrates this point.

Further advantages in using factorial designs in environmental testing experiments follow:

a. No control groups are required.

b. Each treatment effect can be determined independently of all the others. That is, unambiguous conclusions can be drawn about each treatment effect.

c. Complex experiments involving a large number of treatments can be easily handled with the factorial procedures.

d. This is the <u>only</u> experimental design in which the relationship among the treatments can be measured. That is, the factorial design can determine whether the effect of one environmental treatment depends upon <u>any</u> of the others. These effects are called interactions.

e. The probability of being right or wrong can be controlled.

f. When the number of treatments used becomes large (three or more), only a fraction (1/2, 1/4, 1/8, etc.) of the total number of combinations in the factorial need be used.

When multiple replications cannot be used and only attribute (go, no-go) data are available, these designs can still be used to take advantage of their efficiency. However, in cases of this kind, the usual analysis of variance cannot be made. Instead, the usual summations are made to obtain and compare two binomial proportions (by the Fisher exact method) to determine the effect of each treatment. See Example No. 1 below.

Results of factorial designs are used as a guide in determining how to define reliability prior to conducting the test of increased severity. That is, the factorial experiment surveys all of the environmental treatments of interest (with a minimum number of test specimens) to determine the difference, if any, among the environmental effects. A decision is then made whether to redesign the item. If the item is considered acceptable at this time, reliability is defined in terms of the environmental treatment or treatments found to be most severe. If no differences are found among the effects, reliability can be defined in terms of a combination of several of the treatments considered most important from an engineering point of view. If reliability is defined in terms of the most severe treatments, the reliability values obtained will be smaller than those obtained with the other treatments. This is a necessary condition if the system's reliability derived from the component's reliabilities is to have meaning.

Tests of Increased Severity

Results obtained from the factorial experiments can be used to determine which of the environmental treatments will be used in the following procedures to predict "reliability-in-use."

B First-Failure Method (Single Factor)

Increase the level of severity after each test result is obtained until the test item fails. If the test destroys the item, increase the level of severity used with each succeeding item tested until a failure is obtained.

The level of severity can be increased in a variety of ways, such as the following:

1. Using more extreme treatments (e.g., higher temperatures or higher G-values).

2. Using two or more treatments on each test specimen.

3. Repeating the same treatment or set of treatments on the same item.

NOTE: Increasing the length of time an item is subjected to a particular treatment is not used here as a means of increasing the degree of severity.

By starting at, or near the level of severity expected in use, a failure should be obtained within five or six trials (or items). After the level of severity has been found that will cause failure, three or more items should be tested to estimate the failure rate at this level.

To determine the predicted "reliability-in-use," find the probability of occurrence in use of the test condition (at which the failure rate was measured) from a table of individual terms of the Poisson distribution (such as Table 39 of reference 9), where "m" is the expected use condition used. This in effect determines the probability associated with the "safety margin." The product of this probability value and the failure rate found under the test condition is the predicted failure rate in use. The predicted "reliability-in-use" is equal to one minus this product.

V EXAMPLES

A Example No. 1

This example demonstrates how factorial designs can be used in combination with tests of increased severity. A simple three-treatment-experiment example is given below. The treatments used in this example are identified and defined as follows:

Identification	Treatment
A	Trans. Vib.
B	Flight Shock
C	Waterproofness

For purposes of the factorial design, each treatment is considered to have two levels:

1. Lower level or absence of the treatment (designated by subscript 1).

2. Higher level or presence of the treatment (designated by subscript 2).

The total number of possible combinations of three treatments, each at two levels, is two cubed or 8. These 8 combinations can be written in the following pattern:

	A_1		A_2	
	B_1	B_2	B_1	B_2
C_1	(1)	b	a	a + b
C_2	c	b + c	a + c	a + b + c

A minimum of 8 items would be required for this plan, each receiving different treatment combinations as follows:

Item Number	Treatment Combinations
1	None (1)
2	B only
3	A only
4	A + B
5	C only
6	B + C
7	A + C
8	A + B + C

By using the same letters (A, B, and C) and symbol (1) to represent the results obtained from testing the eight items, it can be shown symbolically that the treatment effects can be independently determined, using the total number of items in the entire experiment for each treatment as follows:

EFFECT OF TREATMENT A

$$A + (A + b) + (A + c) + (A + b + c) - \left[(1) + b + c + (b + c)\right] = 4A$$

EFFECT OF TREATMENT B

$$B + (B + c) + (a + B) + (a + B + c) - \left[(1) + c + a + (a + c)\right] = 4B$$

EFFECT OF TREATMENT C

$$C + (b + C) + (a + C) + (a + b + C) - \left[(1) + b + a + (a + b)\right] = 4C$$

One-fourth of these differences equals the average effect of the respective treatments. From the above equations, it can be seen that the results obtained from the eight items have been used three times--once for each treatment. This produces an effective sample size equal to 3 X 8, or 24 items; yet, each treatment has been determined independently of the others.

The above three-factor factorial can be used as an example of a fractional factorial design as follows:

	A_1		A_2	
	B_1	B_2	B_1	B_2
C_1	-	b	a	-
C_2	c	-	-	a + b + c

A minimum of four items is required in this design. As before, the separate effects can be determined by a process of summation and subtraction as follows:

EFFECT OF TREATMENT A

$$A + (A + b + c) - (b + c) = 2\,A$$

EFFECT OF TREATMENT B

$$B + (a + B + c) - (a + c) = 2\,B$$

EFFECT OF TREATMENT C

$$C + (a + b + C) - (a + b) = 2\,C$$

One-half of these differences equals the average effect of the respective treatments.

When there is only one item available for each treatment combination, and only success and failure data are available, the usual analysis of variance cannot be used. However, the above differences, which will be binomial proportions in this case, can be compared by the Fisher exact method for 2 x 2 tables (Ref. 7) to determine the treatment effects. A very convenient set of tables for this purpose can be found in Ref. 8, which contains tables of minimum contrasts based on Fisher's exact method.

When it can be determined, from the results of the factorial esperiment, which environmental conditions will be used to define reliability, the level (or severity) of the condition required to demonstrate a given "reliability-in-use," with a small sample, can be calculated in advance of testing, on the assumption that no failures will be obtained; if the average condition in use is known.

The test conditions required can be calculated as follows:

$$R = 1 - P_t \text{ (UCL)}$$

where:

R = the specified "reliability-in-use."

P_t = the probability of test conditions occurring in use.

UCL = the upper confidence limit (associated with the specified confidence level) of the failure rate expected under the test conditions to be calculated below.

When R and UCL are known, P_t can be calculated from the above formula. Given P_t (the probability) and the average use condition (m), the required test condition (i) can be found from Table 39 of reference 9, or from the following formula:

$$P_t = m^i e^{-m}/i\,!$$

where (e) is the base of natural logarithms.

1 Sample Calculations

Using the same three-factor-experiment example as above gives the following typical set of results, when one is entered as a "failure" and zero is entered as a "success." It is assumed that a knowledge of the item being tested has led to the decision that transportation vibration, flight shock, and waterproofness, in that order, are the three environmental conditions most likely to affect the important functioning characteristic of this item; this characteristic is contact resistance.

The treatment procedure and work-sheet (to record results) for this experiment would be the following two-entry table. An "X" in the item column means that the item receives the corresponding treatment, while a blank means that the item does not receive the treatment.

Treatment Procedure

Order of Treatment	Item No.							
	1-4	5-8	9-12	13-16	17-20	21-24	25-28	29-32
Trans. Vib. (A)			X	X			X	X
Flight Shock (B)		X		X		X		X
Waterproofness (C)					X	X	X	X
Results: Replication 1	0	0	1	1	1	1	0	1
2	0	0	1	0	1	1	0	0
3	0	1	1	1	1	1	0	1
4	0	1	0	1	0	1	0	0
Totals	0	2	3	3	3	4	0	2

The results of one complete replication should be obtained under a single set of controlled conditions (e.g., in the same day, same operators, same instruments, etc.), before going to the next replication. This will make it possible mathematically to subtract out of the results the effect of changing conditions.

By placing these results in the usual factorial matrix, the following table would be obtained:

	A_1		A_2	
	B_1	B_2	B_1	B_2
C_1	0	0	1	1
	0	0	1	0
	0	1	1	1
	0	1	0	1
	0	2	3	3
C_2	1	1	0	1
	1	1	0	0
	1	1	0	1
	0	1	0	0
	3	4	0	2

In preparation for analyzing these results, the usual summing process would give the following series of two-factor tables:

Summing over A

	B_1	B_2	Row Totals
C_1	3	5	8
C_2	3	6	9
Column Totals	6	11	17

Summing over B

	A_1	A_2	Row Totals
C_1	2	6	8
C_2	7	2	9
Column Totals	9	8	17

Summing over C

	A_1	A_2	Row Totals
B_1	3	3	6
B_2	6	5	11
Column Totals	9	8	17

Each one of the marginal totals is the sum of 16 observations. The results can now be analyzed and interpreted as follows:

Source	Effects	Test of Significance [a]
Main Effects		
Trans. Vib. (A)	9/16 vs 8/16	non-significant
Flight Shock (B)	9/16 vs 11/16	non-significant
Waterproofness (C)	8/16 vs 9/16	non-significant
Replication		
1	5/8	non-significant
2	3/8	non-significant
3	6/8	non-significant
4	3/8	non-significant
Interactions		
A x B	8/16 vs 9/16	non-significant
A x C	4/16 vs 13/16	significant
B x C	8/16 vs 9/16	non-significant
A x B x C	7/16 vs 10/16	non-significant

[a] By the Fisher exact method for the 95% (two-sided) confidence level

2 Interpretation (when the above order is used)

(a) None of the effects is significant except the AC interaction. This means that items which have received transportation vibration treatment are significantly less waterproof than those not receiving transportation vibration.

(b) None of the treatments taken alone is significant, although the flight shock effect approaches significance. This result suggests the need for additional flight-shock tests if this treatment is considered important from an engineering point of view.

(c) The fact that the three-factor (ABC) interaction is not significant shows the following:

(1) Waterproofness does not change the effect of transportation vibration on flight shock (AB interaction).

(2) Flight shock does not change the effect of transportation vibration on waterproofness (AC interaction).

(3) Transportation vibration does not change the effect of flight shock on waterproofness (BC interaction).

(d) The fact that replication is not significant means that conditions were under a state of control throughout the experiment.

These results show clearly that the effect of transportation vibration on waterproofness is the most severe combination. It would appear from the results that a decision to improve the waterproofness characteristics is required. After this has been done, reliability must be defined. The results of this experiment show that reliability should be defined in terms of contact resistance (the functioning characteristic of interest) under the following environmental conditions:

(1) Transportation vibration followed by waterproofness (since these two conditions interact).

(2) Flight shock (since this treatment effect approaches significance).

If the reliability of the contact resistance under these conditions is acceptable, the reliability of the contact resistance under the other conditions will also be acceptable.

If the average transportation vibration condition in use is assumed to be 5 G's and the required reliability is 0.995, the test condition required to demonstrate this reliability with a sample of 5 test specimens can be calculated as follows:

when: $R = 1 - P_t \text{ (UCL)}$

then: $P_t = \dfrac{1 - R}{\text{UCL}}$

when: $R = .995$

$\text{UCL} = .52$ (the upper confidence limit at the two-sided 95% confidence level for testing five items and obtaining no failures)

then: $P_t = \dfrac{1 - .995}{.52} = .0096$

From Table 39 of reference 9 the test condition (i) associated with a use condition (m) of 5 G's and a probability (P_t) of .0096 is found to be equal to 10.9 G's. This is the level of transportation vibration required, followed by the waterproofness test to demonstrate a contact resistance reliability of 0.995 if no failures are obtained. If failures are obtained, the test conditions required to demonstrate a reliability of 0.995 with a sample of five test specimens are as follows:

Observed No. of Failures in Sample of 5 Test Specimens	% Failure	Test Condition Required, in G's
0	0	10.9
1	20	11.3
2	40	11.5
4	80	11.7

NOTE: It is evident from the above sample calculations that, for small-sample sizes, the difference in test conditions between zero and anything less than 100% failures is insignificant. This means that the required test condition can be conservatively estimates by expecting a high failure rate.

B Example No. 2

When the average of the conditions in use is known, very high values for the "reliability-in-use" can be correctly predicted with very small sample sizes, if the level of severity is increased until a failure is obtained:

Given:

Average use conditions (m) = 5 G's

Found (Using First - Failure Method):

Number of items used to find test condition to cause first failure	5
Test condition found (i)	18 G's
Number tested at 18 G's	5
Number of failures at 18 G's	2

The probability (P_t) of the test condition's occurring in use, from Table 39 of reference 9, when m = 5 and i = 18, is found to be $P_t = 0.000004$. The upper confidence limit (UCL) of the observed failure rate (2/5) for the two-sided 95% confidence level, from Table V of reference 8, is found to be UCL = 0.8534.

Since:

$$R = 1 - P_t \text{ (UCL)}$$

Then:

$$R = 1 - (0.000004)(0.8534) = 0.9999966,$$ which is the predicted "reliability-in-use."

C Example No. 3

When nothing is known about the expected conditions in use, the "most severe condition in use" an item can withstand can be calculated:

Given:

Required "reliability-in-use" = 0.9999

Test condition used (i) = 10 G's

Number of items tested = 10

Number of failures obtained = none

Since:

$$R = 1 - P_t(UCL)$$

$$P_t = \frac{1 - R}{UCL}$$

From Table IX of reference 8, UCL for no failures in 10 trials equals 0.3085, for the two-sided 95% confidence level.

Then:

$$P_t = \frac{1 - 0.9999}{0.3085} = 0.000324$$

From Table 39 of reference 9, the "most severe condition in use" (m) for i = 10 and P_t = 0.000324 is found to be:

m = 2.6 G's

With the given test result and test condition, this is the "most severe condition in use" under which the item will have 0.9999 reliability.

REFERENCES

1. Lusser, Robert, "Unreliability of Electronics - Cause and Cure," Redstone Arsenal, Huntsville, Ala., Nov. 1957.

2. Lusser, Robert, "Reliability Through Safety Margins," Redstone Arsenal, Huntsville, Ala., Oct. 1958.

3. Dixon and Massey, "Introduction to Statistical Analysis," McGraw-Hill Book Co., Inc., NYC, 1957, Second Edition.

4. Davies, O. L. (Editor), "Statistical Methods in Research and Production," Hafner Publishing Co., NYC, 1957, Third Edition.

5. Plackett and Burman, "The Design of Optimum Multifactorial Experiments," Biometrika, Vol 33, 1946, page 305.

6. "Fractional Factorial Experiments Designs for Factors at Two Levels," NBS Applied Math. Series No. 48.

7. Fisher, R. A., "Statistical Methods for Research Workers," Hafner Publishing Co., NYC, 1950.

8. Mainland, Herrera, and Sutcliffe, "Tables for Use With Binomial Samples," Department of Medical Statistics, NY University College of Medicine, NYC, 1956.

9. Pearson, E. S., and Hartley, H. O., "Biometrika Tables for Statisticians," Volume I, Cambridge University Press, Cambridge, England, First Edition, 1956.

MULTI-DIMENSIONAL STAIRCASE DESIGNS FOR RELIABILITY STUDIES

David R. Howes
U. S. Army Chemical Corps Engineering Command

This paper suggests the need for a sequential staircase procedure whereby a given contour of response could be traced experimentally without necessarily defining the entire response surface. Such a method would have important application in reliability studies, in design and engineering, etc. where the intent is to hold malfunctioning of some type at some predetermined level.

As an example, suppose that an artillery shell is to be filled with poison gas and closed with a burster tube (Figure 1, on the next page). It has been found that the leakage of these shells is affected by two variables, the interference of the burster tube (Variable A) and a variable B which is a structural characteristic of the shell body.

Although it may be possible to specify levels of A and B which will be satisfactory, it is also necessary to know the threshold of leakage in order to set manufacturing tolerances, filling procedures, etc. This involves the response surface generated in leakage in the A-B space.

Possible Methods

1. Factorial

It is possible to fit a response surface to the results of the experiment shown in Figure 2 (see next page) by well-known methods. The sample size, 1800, may seem excessive for accuracy which cannot exceed 2%.

2. Confined Factorial and Staircase Method

Staircase methods have been described which permit the isolation of percentage points of a problem with only one variable.[1,2,3] For a two variable case it would be possible to treat one variable by a staircase method, and the other factorially. (See Figure 3).

3. Multi-Dimensional Staircase Method

An extension of the staircase method to N variables is possible, although the methods have not been produced yet, since the theory doesn't exist. We would assume that a smooth response surface existed and that it would be possible, staircase-wise to follow some response contour on that surface say 90%, 93%, 99%, etc. Interaction of the variables is also a possibility.

The factorial approach may be inefficient here since it collects data not needed merely to trace a contour. It may also be wasteful of time, materials, and manpower, since the experiment cannot proceed until a large predetermined number of items are available. It should also be mentioned that it runs against the strong desire usually found

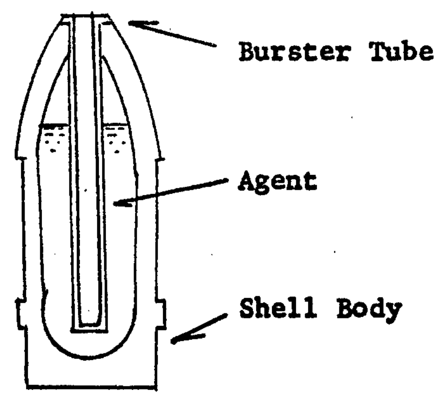


Figure 1.

Agent Filled Shell

Number of Shells

Variable A Level	Variable B Level 1	2	3	4	5	6
6	50	50	50	50	50	50
5	50	50	50	50	50	50
4	50	50	50	50	50	50
3	50	50	50	50	50	50
2	50	50	50	50	50	50
1	50	50	50	50	50	50

Figure 2.

93% Points for Staircase Method

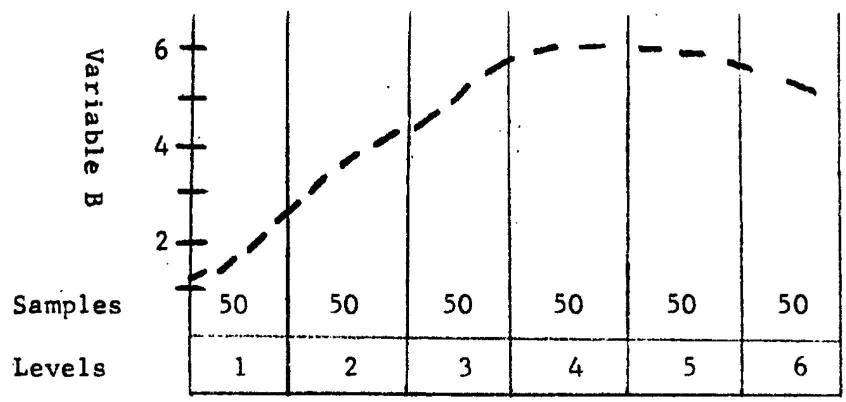


Figure 3.

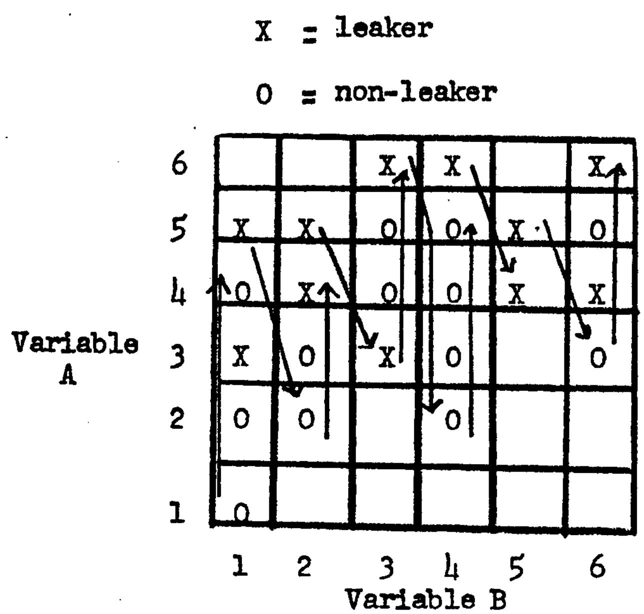


Figure 4.

among reliability engineers to try something, then try something else. Since the statistician must live with this, it would seem most desirable to adopt procedures which resemble to as great an extent feasible, those used by the reliability and test engineers. The sequential nature of the procedures has also the advantage of allowing experimentation to continue, while new experimental vehicles are being fabricated.

Example of a Multi-Dimensional Method:

In this procedure we might try a single sample at a time, and move over the response surface in accordance with certain rules based on the previous test result. (See Figure 4).

In this design, we proceed upward one step at a time in A in level No. 1 until two leakers have been found, then proceed to the next higher level of B at a level of A, one level below that on which the first leaker occurred. Then proceed upward in A until again two leakers have been found and proceed as before. Take the mid-point on A between the two leakers at each level of B as a point on the 50% leakage contour. Repeat this experiment as many times as necessary to get the desired precision of estimate. Instead of taking a single sample, we might take a sample of n and consider it "reliable" if no more than c leakers are found. This would, I suppose, lead to the tracing of percentage lines. Using a table of random numbers, I was able to get fairly good results in tracing a 10% contour over a bi-variate surface; good enough results to suggest the desirability of answering the following questions:

(1) What are the most effective rules to follow when traversing the response surface?

(2) What confidence can be placed in the results of K trials, using a sample size n in each trial, with an allowance of c defects as an estimate of the c/n fraction line?

(3) What method of computation or statistic would be used to obtain the c/n fraction line estimate from the data?

REFERENCES

1. National Defense Research Council Applied Mathematics Panel Report No. 101.1

2. Bureau of Ordnance, U. S. Navy, NAVORD Report No. 65-46, 21 March 46 - Anderson, McCarthy, & Tukey

3. Dixon, W. J. and Mood, A. M., "A Method of Obtaining and Analyzing Sensitivity Data" Journal of the American Statistical Association. Vol 43, No 241 (1948)

A PROPOSED RESEARCH PROGRAM FOR PROVIDING A QUANTITATIVE BASIS FOR PREVENTIVE MAINTENANCE POLICIES ON ORDNANCE EQUIPMENT

Walton M. Hancock
and
Randall E. Cline

This talk contains an outline of a proposed program which can be used to aid in the establishment of preventive maintenance policies. The program represents a general approach applicable to both existing systems and systems yet to become operational. It is anticipated that this approach should ultimately lead to the simplification of the maintenance of Ordnance equipment.

Since the Ordnance Corps has many different types of equipment, which vary both in complexity and density, no one PM policy can be applicable to all types of equipment. The research effort must therefore be oriented towards developing the proper general approach to the establishment of preventive maintenance policies for a variety of weapons systems. The applicability and usefulness of the approach, then, can be demonstrated by selecting a limited number of weapons systems and evaluating different preventive maintenance policies for them.

The talk is presented as follows: Part 1 contains a discussion of work done by others and comments on methods used to develop general solutions. Part 2 presents a general approach to the preventive maintenance problem, and Part 3 is a mathematical formulation which has been used to present, in a compact form, the ideas developed in the general approach. The fourth part contains a proposed program which will provide information for the evaluation of preventive maintenance policies on specific weapons systems in order to serve as examples of the application of this general approach to establishing preventive maintenance policies for other weapons systems.

1. <u>CURRENT STATUS OF RESEARCH ACTIVITY CONCERNING MAINTENANCE POLICIES</u>. An extensive library search and a number of visits have been made in order to take advantage of work performed by other research groups in the maintenance area.[1] One finds that there has been quite a bit of effort put into the specific details of establishing preventive maintenance policies for a particular type of equipment; there are, however, relatively few people that are concerned with a general approach.

The areas in which much work has been done include the military electronics field and the civilian trucking industry. An examination of the approaches used in these areas has proved quite helpful.

In the military electronics field, investigations have been made into the problems of reliability. The reliability of equipment is directly related to the amount and type of maintenance. Methods of analysis used in evaluating and improving reliability can also be used in developing maintenance programs. Briefly, the methods have been as follows:

---

1. A list of the most pertinent books and articles is given in the bibliography.

a. Rather than conduct broad scale studies, particular using units have been selected for detailed study.

b. Only new equipment or equipment put through major overhaul is issued to the units to be studied.

c. Detailed records are kept on the life history of the equipment. The amount of scheduled and unscheduled maintenance, the time required for repairs, basic causes of failure, parts usage and the time interval between failures are all carefully documented.

d. The reliability of the equipment is related to the mission that is to be performed, and the mean time to failure is frequently used in deriving an expression of reliability. Since the missions are usually expressed in terms of the number of hours of use per mission, then the probability that the equipment will perform the expected mission can be predicted.

e. Emphasis has been placed in the classification of the types of failures by basic components of the system such as by tube types, types of resistors, and capacitors. The reliability of the equipment is then expressed in terms of the reliability of its components.

Considerable success has been attained by the use of the above methods. The most notable of these have been realized by classifying the basic causes of failure for design purposes.

Since maintenance is one of the principal costs in the civilian trucking industry, effort has gone into the solution of their maintenance problems. The following represents the general methods used by the industry:

a. A detailed life history is kept on each vehicle. These data contain a record of all repairs, including parts usage and costs, and a record of all maintenance performed. Incidence of breakdowns, associated costs, and the mileage of the vehicle are recorded for such events in the life history.

b. By analysis of the life histories, norms are established for the expected life of each major component. Careful investigations are then made to determine the causes of failures which seem premature. Failures are also classified as to design deficiencies, improper maintenance or poor driving. Those that are in the design deficiency category are used to change the specifications on new equipment. Those caused by improper maintenance are used to modify the amount, kind, and time interval of scheduled maintenance. The failures caused by poor driving are analyzed for improved driver education.

c. The "cost of maintenance per mile" is also derived from the detailed life history. This is a control technique used by administrative personnel to see if the total cost of maintenance is kept within prescribed limits.

Proper feedback of information and the proper analysis of this information is considered by most trucking firms to be an absolute necessity. This applies to commercial trucking fleets, bus fleets, truck and car rental fleets, and users of off-the-road equipment. The same approach prevails regardless of the actual use to which the equipment is put.

2. A GENERAL APPROACH TO THE MAINTENANCE PROBLEM. In order to develop a general approach that can be applied to all types of weapons systems, it is first necessary to classify weapons systems in such a way that their common characteristics as well as their differing characteristics are evident. Preventive maintenance policies will then be related to these characteristics. It appears that all weapons systems can be classified (for purposes of establishing maintenance policies) in terms of the following basic parameters:

a. Complexity

b. Density

c. Mission

Each of these three basic classifications is a vector quantity, or stated more simply, each can be described in terms of a number of factors. For instance, the complexity of a weapons system may be defined in terms of the crew requirements, the number of components, the total cost of the system, the average amount of time required to locate troubles, the ratio of time to check out the system compared to the time to complete a mission, etc. Similarly, density may be expressed in terms of geographical dispersion of equipment, travel time from support unit to supported units, total number of units in the field, etc. Missions can be defined in terms of the time equipment is required to be operable, the movements of operations which must be accomplished, the precision with which operations must be performed, etc.

Examples of the way these classifications are related to maintenance policies are as follows: Experience has shown that for electronic equipment an increase in complexity increases the maintenance requirements. The density of weapons systems affects the organization of maintenance crews and supporting test equipment. The mission also affects the type of maintenance. Since many weapons systems are required to perform a number of different missions, to achieve simplicity of maintenance at a minimum cost, the maintenance requirements may also vary. For example: trucks that are used on hard surface roads will require different maintenance than trucks used off the road. One trucking firm that was visited had different maintenance schedules for long distance vehicles than for local haul equipment, because the cost of a breakdown of a vehicle some distance from maintenance support was many times higher than for a vehicle used locally.

3. MATHEMATICAL FORMULATION OF A GENERAL APPROACH. Using the classification of weapons systems introduced in Part 2, a general approach to establishing preventive maintenance policies will now be considered. In the analysis of failure data for electronic equipment, the term reliability of a system or a component of a system has been generally used to denote the probability that a system or component will perform its required mission under given conditions for a specified operating time. The reliability of Ordnance equipment can be defined similarly. In developing a model which relates maintenance and reliability, the following assumptions are made:

a. Reliability is dependent upon the age of the equipment. For example, tanks, trucks, missiles, etc., tend to fail more frequently as the equipment becomes older.

b. The reliability is also dependent upon past usage of the equipment. For example, it is expected that the number of failures in trucks increase as the number of miles traveled increases. Similarly, the number of times a missile is checked is believed to affect the probability that the missile will fire.

Since both age and past usage are assumed to affect the reliability of equipment, it is useful to redefine reliability. Consequently, the following notation will be introduced. For any given system, let t = calendar age of the system, i.e., the number of years since the equipment was issued (new) to the user, and let $x(t)$ = usage of the system prior to time t, i.e., $x(t) = \int_0^t f(\tau)d\tau$, where $f(\tau)$ is some measure of usage.

The reliability of the system will be defined as follows:

The reliability of a system is the probability that the system will perform its required mission (which includes a specified operating time), given that the age of the system, t, and the past usage, $x(t)$, and the preventive maintenance policies are known. The reliability of a system will be designated by the symbol $r[t, x(t)]$.

It has been suggested that for wheeled and tracked vehicles, $x(t)$ = mileage, and for missiles, $x(t)$ = number of times certain checks have been made on the system.

Both scheduled and unscheduled maintenance are related to the reliability of a system. The purpose of maintenance is to increase $r[t,x(t)]$, and hence to maintain the reliability above some predetermined minimal level. The policies regarding scheduled and unscheduled maintenance are expected to be influenced by the complexity and by the density of the system. Visual inspections and operational check-out procedures are designed to ascertain whether the reliability of the system is above this predetermined level. Since a weapons system may be required to perform several different missions, we must consider all operations which the system may be required to perform and the associated performance times. Classify these missions (that is operation-time combinations) in groups in such a way that all missions in any given group are roughly equivalent in terms of requirements on the system. Such a group of missions will be called a task. Now order these tasks in such a way that if the system can perform any given task, then it can also perform all simpler tasks. Designate these tasks by $T_1, \ldots, T_\ell$, $(\ell \geq 2)$, where T_1 indicates the simplest task and T_ℓ corresponds to the most difficult task. Having specified a given task, say T_k, there exists a corresponding probability that the system can perform this task. Denote this probability by $r_k[t,x(t)]$.

Then for fixed t and x(t), and all $k = 2, \ldots, \ell$,

$$r_k[t,x(t)] \geq r_{k+1}[t,x(t)] .$$

Now for each task T_k and any given maintenance policy, there exists a surface $r_k[t,x(t)]$ which represents the reliability of the system relative to T_k. Such a surface is illustrated in Figure 1.

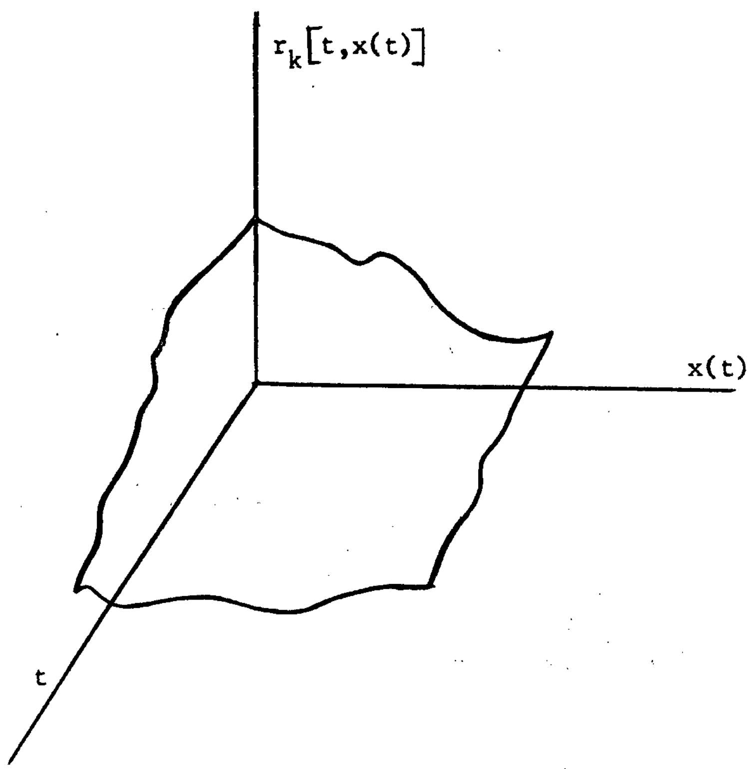


Figure 1

Theoretically, a maintenance policy should be designed in such a way that for any value of t, $0 \leq t \leq t'$, where t' is that time at which the system is discarded, and for any usage x(t), the reliability of the system should be maintained in such a way that for some α, $0 < \alpha \leq 1$,

$$r_k\left[t,\ x(t)\right] \geq \alpha \tag{1}$$

for some task T_k. Since it is assumed that the reliability of a system decreases with both time and usage, then maintenance must be performed in an attempt to satisfy equation (1). The physical situation is illustrated in Figure 2 in which the lower surface represents the actual system reliability and the upper surface represents the desired goal of maintenance. It is to be noted that $r_k\left[t,x(t)\right]$ may actually exceed α. Conversely, the effect of inappropriate design on the system may be such that $r_k\left[t,x(t)\right]$ never attains the desired goal. Finally, it is observed that the goal may not be constant over the entire expected life of the system and may be lowered as equipment is phased out.

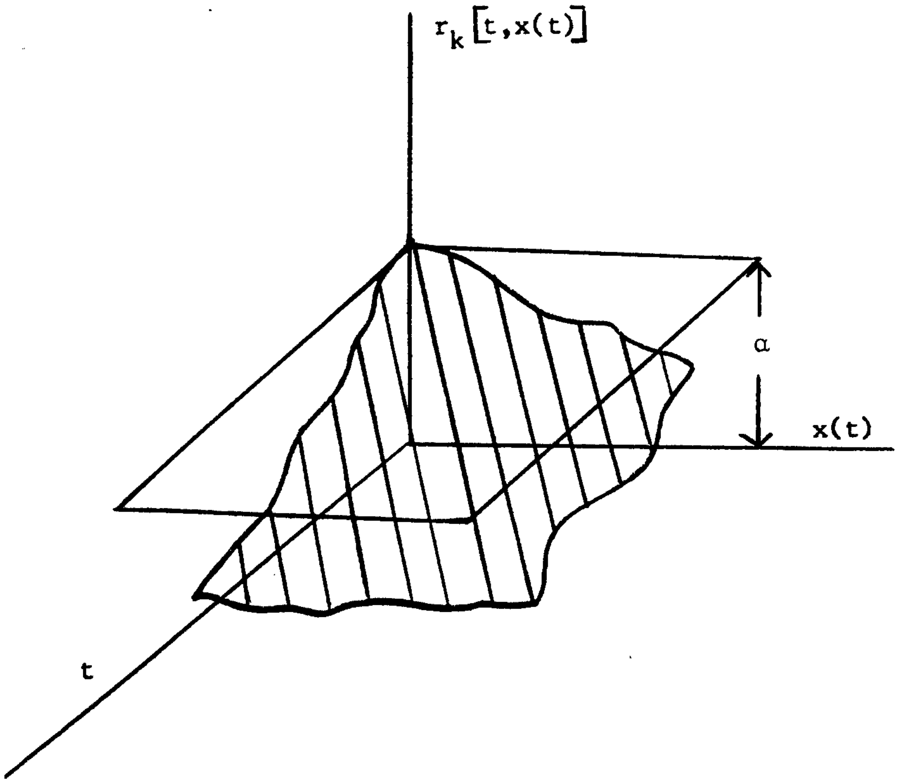


Figure 2

The life history of a system can be represented by a curve in the $t,x(t)$ plane. Associated with this curve is the corresponding reliability. Such a curve is shown in Figure 3.

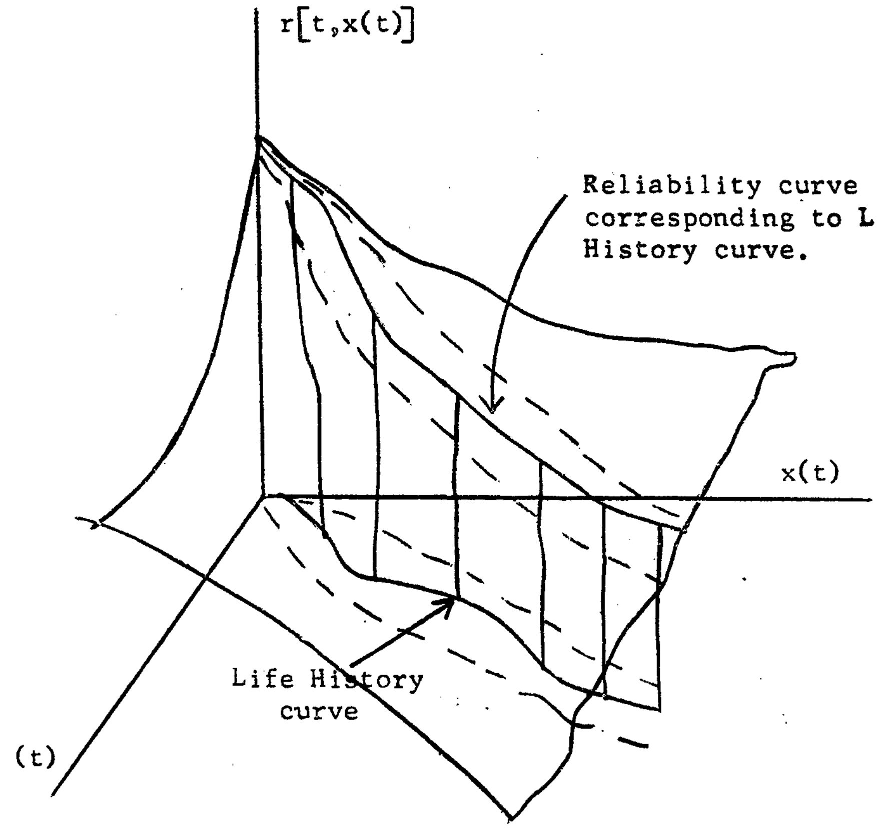


Figure 3

Observe now that since a collection of similar weapons systems in an Army unit will not be used in identical amounts, then observations taken from a particular unit will form a wedge as illustrated by the broken lines in Figure 3.

As mentioned above, maintenance of any type is intended to increase the reliability of the system. Thus, for any maintenance performed at a given point $[t,x(t)]$, let:

$$\Delta r_k \left[t,x(t)\right]$$

denote the change in $r_k[t,x(t)]$ obtained by performing the maintenance. The amount of this change is a random variable dependent upon the type of maintenance performed, the level of skill of the technician performing it, and the tools or test equipment available to him. Graphically, this may be illustrated as in Figure 4 for a particular system having maintenance performed at points $[t,x(t)]$, $[t_2,x(t_2)]$, ..., $[t_n,x(t_n)]$, where the jumps in the curve indicate the corresponding changes in reliability resulting from the maintenance.

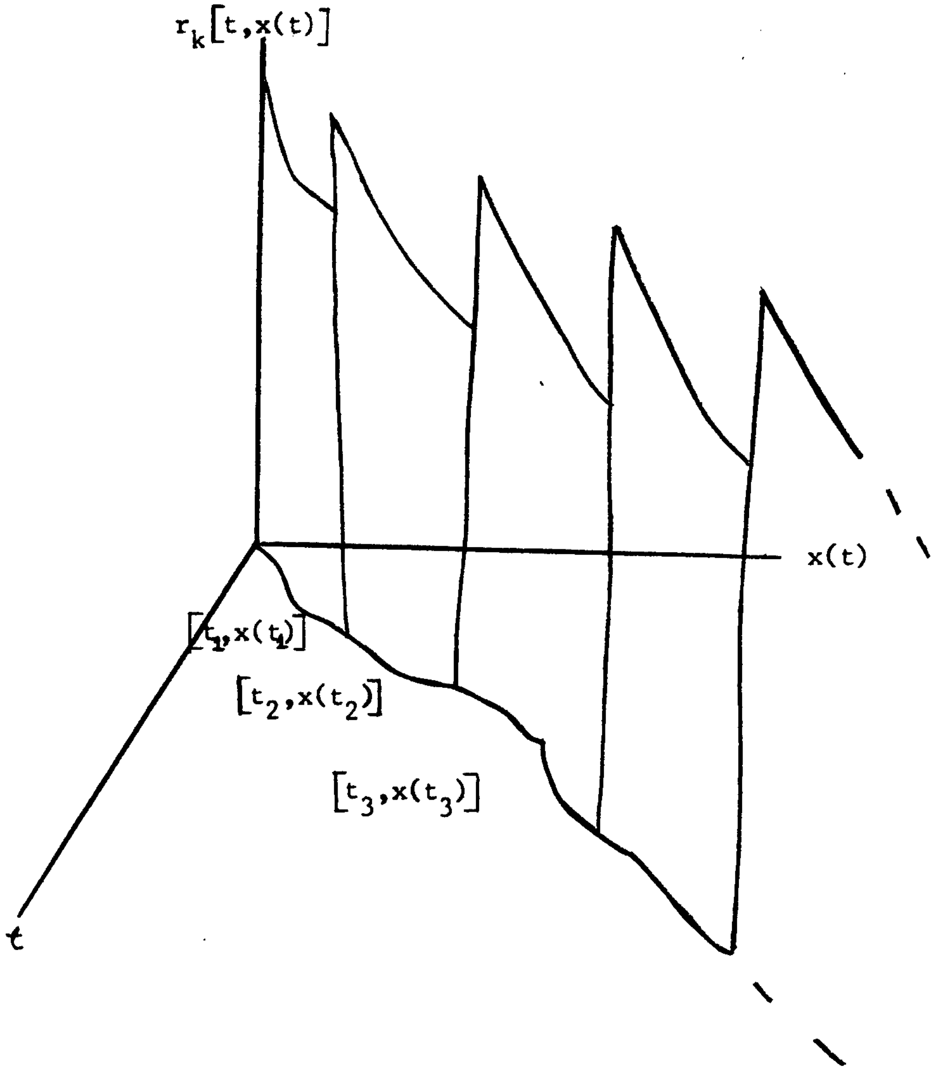


Figure 4

It is to be observed that if not only $r_k[t,x(t)]$ and $\Delta r_k[t,x(t)]$ are known, but also various costs of maintenance associated with $\Delta r_k[t,x(t)]$, then measures can be developed which relate costs of maintaining equipment to the expected loss from having equipment inoperative. Consequently, to use this approach for evaluating the effects of various preventive maintenance policies in terms of the change of the reliability of the system, it will be necessary to develop techniques for estimating the forms of $r_k[t,x(t)]$ and $\Delta r_k[t,x(t)]$ for various tasks and maintenance practices.

To aid in the estimation of these functions, it is useful to consider the effects of various components of the system on overall system performance. Now any given system can be represented as a collection of major components (or subsystems). Associated with each component is a corresponding reliability surface again a function of t and x(t). Assuming these major components are serially connected and statistically independent, then at any point $t, x(t)$,

$$r_k\left[t, x(t)\right] = \prod_{j=1}^{n} r_k^{(j)}\left[t, x(t)\right]$$

where there are n major components with reliabilities $r_k^{(j)}\left[t, x(t)\right]$, $(j=1, \ldots, m)$. By estimating not only $r_k\left[t, x(t)\right]$ and $\Delta r_k\left[t, x(t)\right]$, but also the corresponding quantities for major components, those components requiring the most maintenance will be apparent. Thus, concurrent with the collection of data to be used in estimating the overall system reliability, data on various components will also be collected. It is to be observed that this additional component analysis is essential for systems in which major components have been replaced.

Continued work on the general formulation will be directed toward relating the function $r_k\left[t, x(t)\right]$ and $\Delta r_k\left[t, x(t)\right]$ both to the organization of maintenance in terms of costs, skills, tools, etc., and to the complexity, density and task classifications of weapons systems.

4. PROPOSED PROGRAM FOR FIELD OBSERVATION AND DATA ANALYSIS. To obtain the information required to develop the curves and surfaces discussed in Part 3, a field observation program will be necessary. It is necessary that at least two, and preferably three, weapons systems be selected that represent different points in the complexity-density range. Since life histories are to be collected, new weapons systems should be selected whenever practicable. The specific weapons systems that are to be studied have not been chosen at the present time. A wheeled or tracked vehicle, a missile, and possibly a hand weapon will probably be selected. In selecting the weapons system and the units to be observed, proper consideration also must be given to the missions that are being performed.

Since a main purpose of the research is to quantitatively evaluate the effect of preventive maintenance practices to permit determination of the proper amount of scheduled maintenance that should be performed, it will be necessary, after an initial observation period, to slightly modify the existing maintenance practices and observe their effects.

Specifically, we will attempt to collect the following information:

A. Scheduled Maintenance

For each scheduled maintenance or operational check performed, the following information is desired:

1. Type of scheduled maintenance

2. Frequency of scheduled maintenance

3. Equipment usage between scheduled maintenance periods.

4. Deficiencies found during scheduled maintenance.

5. Parts replaced during scheduled maintenance.

6. Time to repair deficiencies found during scheduled maintenance periods.

7. Time to perform scheduled maintenance.

8. The echelons that perform the scheduled maintenance.

B. Unscheduled Maintenance

For each failure requiring unscheduled maintenance the following information is desired:

1. Frequency of failure

2. Basic cause of failure

3. Elapsed time since specific scheduled maintenance.

4. Usage since specific scheduled maintenance.

5. Parts needed for repair of failure

6. Parts available

7. Time to repair failure

8. Echelon performing the repair

C. Maintenance Organization

For the unit being observed the following information is desired:

1. Skills and equipment available at using unit and supporting units.

2. Inspection criteria at each echelon

3. Work load at each echelon.

We plan to initially place technically qualified field engineers on a full time basis with the units to assist user personnel in recording the above information. It is hoped that after approximately three to six months the cooperating units will be able to provide the necessary information, and the field staff will be required only to monitor the data collection program on a part-time basis. This would also free the field staff to initiate a similar observation program with an additional using unit.

As discussed in Part 3, the data collection program will be initiated concurrently with the continuation of the development of the general approach. As the program proceeds, the level, type and method of data collection may, of course, require modification. The initial information obtained will aid in selecting the most pertinent of many possible characteristics of weapons systems for first consideration in the mathematical formulation, and in defining the groups of missions required of the different types of systems studied. Such effort will be needed to further refine the general formulation to insure that the results will have practical significance to the particular weapons systems under consideration.

BIBLIOGRAPHY FOR APPENDIX III

1. "Effects of Maintenance on System Reliability," ARINC Research Corporation, Publication No. 101-16-144, Sept., 1959.

2. "An Analysis of the Effects of Maintenance on Parts Replacements," Ralph L. Madison, AERO. Radio Inc., Publication No. 112, Jan. 6, 1958.

3. "A Survey of the Current Status of the Electronic Reliability Program," R. R. Carhart, The Rand Corporation, RM 1131, Aug., 1953.

4. "Maximizing Expected Machine Up Time," Jay B. Heyne, Systems Development Laboratory, Hughes Aircraft Co., July, 1959.

5. "Evaluation and Prediction of Circuit Performance by Statistical Techniques," J. Marini, H. Brown, R. Williams, ARINC Inc., Publication No. 113, Feb., 1958.

6. "The Condition of 60 Farm Tractors Point Ways to Improve Maintenance," J. A. Weber, Society of Automotive Engineers, Paper No. 181, Sept., 1957.

7. "Problems Relating to Systems Checkout and Final Acceptance of Production Missiles," M. A. Stephens, Society of Automotive Engineers, Paper No. 48B, April, 1958.

8. "A Bibliography of Life Testing and Related Topics," William Mendenhall, Biometrika, Vol. 45, Parts 3 and 4, pp. 521-543, Dec., 1958.

9. "Maintenance Management--Operational Readiness Plan for the Redstone Missile System," Lt. Col. Brian O. Montgomery, Redstone Arsenal, Alabama, Sept., 1958.

10. "Optimum Component Redundacy for Maximum System Reliability," Robert Gordon, Operations Research Society Journal, Volume 5, No. 2, April, 1957. (pp. 229-243).

11. "Reliability of Military Electronic Equipment," The Advisory Group on Reliability of Electronic Equipment, Office of the Assistant Secretary of Defense (Research and Engineering), 4 June, 1957.

12. "Queues, Inventories, and Maintenance," Philip M. Morse, John Wiley and Sons, 1957.

13. "Reliability Factors for Ground Electronic Equipment," Henney, Lopatin, Zimmer, Poler, and Naresky, McGraw Hill, New York, 1956.

14. "A Good Preventive Maintenance Program," SAE Journal, Vol. 65, p. 88, July, 1957.

15. "Automatic Transmissions Make Some Truck Operator's Happy, Others Sad," SAE Journal, Vol. 65 pp. 56-58, Spetember, 1957.

16. "Simplicity - Key to Military Design," SAE Journal, Vol. 67, pp. 34 and 35, July, 1959.

17. "Air Force Model Shop Sets New PM Pace," Commercial Car Journal, Vol. 90, pp. 74-75, December, 1955.

18. "Maintenance Yardsticks Show Bus PM Efficiency," Commercial Car Journal, R. W. Ziffle, Vol. 90, pp. 68, 128-132, December, 1955.

19. "Systems Checkout Insure Missile Reliability," M. A. Stevens, SAE Journal, Vol. 66, pp. 80-81, August, 1958.

20. "Do Road Calls Measure PM Efficiency?" Commercial Car Journal, Vol. 90, pp. 71 and 186, September, 1955.

A STATISTICAL ANALYSIS OF VARIOUS PARAMETERS AFFECTING THE BURNING CHARACTERISTICS OF FLARE SYSTEMS

Bossie Jackson
Pyrotechnics Chemical Research Section
Pyrotechnics Laboratory
Picatinny Arsenal

The development of pyrotechnic flare compositions involves the investigation of a number of different variables. From the many investigations conducted previously numerous hypotheses were formed concerning the relationship of such variables as candlepower and burning rate with flare case coating, loading pressure, and the particle size of the fuel employed in the system. A better knowledge of the basic factors governing the burning characteristics of solid mixtures was desired. It was expected that the results of this investigation would tend to substantiate the various hypotheses.

Previous data tends to show that candlepower and burning rate are depended upon particle size, loading pressure, and are not affected by flare case coating. The analysis of these relationships was based on data obtained from this study using two statistical methods, the test for least significant differences, and in particular the analysis of variance. The experimental design for these studies is given in Figure 1. (See next page.) This configuration was used for four (4) levels of magnesium, M_1, M_2, M_3, and M_4. The flare case coatings are shown by C_1, C_2, C_3, and C_4. Seven (7) levels of loading pressure were studied ranging from P_1 to P_7. Five samples were utilized for each combination of pressure, case coating, and magnesium particle size.

A standard flare composition (Table I) [Tables can be found at the end of this article] was used for this experiment. This composition contains 48% magnesium, 42% sodium nitrate, 2% polyvinyl chloride, and 8% Laminac resin. Flare compositions are consolidated in a variety of cylindrical cases under a specified pressure to obtain a cigarette-type propogating composition rather than one which flashes or explodes. The use of self-hardening polyester resins in flare compositions eliminated the need for consolidating them at very high pressures. Most flare compositions are presently loaded at pressures ranging from 4000 pounds per square inch (psi) to 10,000 psi.

Standard flare compositions each containing different mesh sizes of magnesium were evaluated in this study. These magnesium granulations together with their average particle diameter are given in Table II. The mesh sizes of magnesium are 20/50, 30/50, 50/100, and 100/200 with particle sizes varying from 437 microns to 110 microns. The compositions were consolidated at 2000, 4000, 7000, 10,000, 15,000, 20,000 and 25,000 pounds per square inch.

The effect of loading pressure on candlepower can be observed in Figure II. The mean candlepower values vary from 201,000 to 223,000 which approximates an eleven (11) percent change. It is apparent from this graph that a definite trend exists with candlepower increasing with increased loading pressure. Table III summarizes the luminous intensity values observed at each pressure level. The least significant difference value at the ninety-five (95) percent confidence level is also given for these values. Despite the apparent trend of candlepower values, it will be observed that the difference in light output from 4,000 psi to 25,000 psi are not large enough to be significant based on the least significant difference value. This conclusion also holds for the candlepower at 2,000 and 4,000 pounds per square inch. However, the differences between the values at 2,000 psi and those at 7,000 psi and above are large enough to be termed significant. The appearance of this trend may be attributed to the relationship between porosity and heat conduction. A more porous column will conduct heat at a slower rate as a result of the air pocket acting as insulators giving slow burning rates and low candlepower. Conversely, the less porous column will conduct heat at a faster rate giving higher candlepower and burning rates.

Higher candlepower values are obtained from smaller particle diameters of magnesium as shown in Figure III. Candlepower plotted as a function of average particle diameter decreased with increasing particle size. The higher candlepower value obtained for the 168 micron magnesium compared to the finer 110 micron magnesium may be attributed to the distributional effect of particle size. As evidenced by the low average particle diameter, the 50/100 mesh magnesium contained a large percentage of fines which placed the material in the upper range of the finer 100/200 mesh magnesium which may account for the higher light output. The candlepower values for each mesh size magnesium are tabulated in Table IV. It is immediately apparent that these intensity values are significantly different from each other on application of the least significant difference value. It is also observed that the 50/100 mesh magnesium gave significantly higher candlepower than the 100/200 mesh fuel. This result reflects the necessity for reducing the tolerance limits of particle size for the different mesh sizes of magnesium. It is believed that data can be accumulated from studies conducted previously to show that candlepower definitely decreases with increasing particle size.

It was previously mentioned that any of the investigators in the field of pyrotechnics believed that candlepower was unaffected by case coatings. This was verified by the results obtained from this experiment as shown in Table V. The flare case coatings studied were Amberlac 292, Laminac resin 4116, Polyethylene 617, and paraffin wax. The candlepower values vary from 210 to 218 and are essentially the same based on the least significant difference value of 8.8. Flare case coatings are especially necessitated with compositions containing self-hardening resins. Since these resins undergo considerable shrinkage on curing, voids and air pockets are created as a result of the composition separating from the flare case wall. Such a condition gives rise to possible detonations or increased burning rates.

Just as candlepower shows an insignificant trend resulting from increased loading pressure, burning rate values show a parallel trend. Figure IV illustrates this trend as the loading pressure is increased from 2,000 psi to 25,000 psi. The burning rate values show a trend towards reduction with increasing loading pressure. Table VI tabulates the mean burning rate values obtained at each pressure level. It also summarizes that the differences in burning rates are not large enough to be significant. Based on this method of analysis, it can be concluded that burning rate is not affected by increasing loading pressure. As a result of the oppositely parallel trends shown by candlepower and burning rate, it can be hypothesized that these two variables are interrelated. This hypothesis is corroborated and borne out when considering the effect of particle size of magnesium on burning rate.

Figure V shows burning rate plotted as a function of average particle diameter. It can be seen that burning rate decreases with increasing average particle size. As shown on Table VIII the differences between these values were found to be significant based on the least significant difference value of 0.05. It was previously observed that a corresponding effect was obtained with candlepower values, except in the result for 50.100 mesh magnesium. To further complicate the picture it was observed that significant differences in burning rates existed for the various flare case coatings. The results are given in Table VIII. The burning rates vary from a slow 3.61 inches per minute for the polyethylene to a fast 4.52 inches per minute for paraffin wax with Amberlac and Laminac resins yielding values in the middle. These differences based on the test for least significant difference indicate that the minimum and maximum values here are significantly different from the intermediate ones which are essentially the same. The significant effects of flare case coatings are undoubtedly due to their variation in binding strength, rate of thermal degradation, and end-products produced on combustion.

Burning rates of pyrotechnic compositions are also derived from the weight composition undergoing combustion per unit time. Figure VI illustrates the effect of loading pressure on the grams of composition per second from 2,000 pounds per square inch to 25,000 pounds per square inch. It is shown that grams of composition per second tends to increase with increasing loading pressure. The change in pressed density from 2,000 psi to 25,000 psi approximated twenty (20) percent. As shown in Table IX the differences in grams of composition per second are not significant based on the least significant difference value of 0.60. It was previously observed that the linear burning rate was not significantly affected by loading pressure.

In direct contrast to the above result, it was determined that the average particle diameter of magnesium had a significant effect on the weight of composition consumed per unit time. Based on the previous effects of particle size on both candlepower and burning rate this result could be anticipated. Figure VII shows grams of composition per second as a function of average particle diameter. It is observed that the number of grams burned per unit time varies inversely as the average particle diameter. The mean burning rate values tabulated in Table X are shown to be significantly different from each other as a result of the test for least significant difference.

By observing Table XI, it can be seen that flare case coatings do not significantly affect the grams of composition per second. This conclusion results in the fact that the average particle diameter of magnesium is the only parameter that significantly affects candlepower, burning rate (inches per minute), and burning rate (gram per second). The only other parameter contributing a significant effect was flare case coating on the linear burning rate.

Summarizing the results previously discussed, the analysis of variance table for candlepower is observed (Table XIII). This table gives the main effects, first and second order interactions of the parameters under consideration. Where the calculated F-ratio exceeds the critical F-ratio the effect of the parameter is considered significant. The main effects of magnesium particle size and case coatings are in accord with the results based on the test for least significant difference. The main effect from loading pressure observed here is significant as opposed to its insignificant effect based on the least significant difference value. Considering that such large changes in loading pressure (92%) results in very small changes in light output, it may indicate that the increasing trend of candlepower is insignificant. The first order interactions of magnesium loading pressure and case coating-loading pressure show a significant effect on candlepower, while magnesuim-case coating is insigificant. Evidently, the flare case coating cancels out the effect of the magnesium particle size. The second order interactions of these parameters are shown to be insignificant.

Table VIII outlines the analysis of variance table for burning rate (inches)per minute). It is observed that main effects resulting from magnesium and case coating are significant and loading pressure insignificant. This corroborates the results based on the test for least significant difference. The first order interactions of magnesium-case coating and magnesium-loading pressure are shown to be significant, while case coating-loading pressure is insignificant. The second order interactions are significant.

The results of the analysis of variance for burning rate (grams per second) is given in Table XIV. Except for the fact that flare case coating is shown to be insignificant here, the results parallel exactly those obtained for the analysis of variance on the linear burning rate. This table also substantiates the results from the test for least significant difference.

TABLE I

COMPOSITIONS EVALUATED

Ingredients	Percent by Weight			
Magnesium, Atomized, 20/50, 437 microns	48			
Magnesium, Atomized, 30/50, 322 microns		48		
Magnesium, Atomized, 50/100, 168 microns			48	
Magnesium, Atomized, 100/200, 110 microns				48
Sodium Nitrate, 34 microns	42	42	42	42
Polyvinyl Chloride, 27 microns	2	2	2	2
Laminac Resin 4116*	8	8	8	8

* Laminac Resin 4116 - 98.5%

Lupersol ddm - 1.0%

Nuodex - 0.5%

TABLE II

MAGNESIUM GRANULATIONS

Mesh Size	Tapped Density, gm./cc.	Average Particle Diameter, Microns
20/50	1.07	437
30/50	1.08	322
50/100	1.14	168
100/200	1.08	110

TABLE III

TESTS FOR LEAST SIGNIFICANT DIFFERENCE OF LOADING PRESSURE VS. AVERAGE CANDLEPOWER

Least Significant Difference - 13.8

Level of Confidence, % - 95.0

Loading Pressure Psi	Average Candlepower
2,000	201.0
4,000	213.0
7,000	215.0
10,000	217.0
15,000	218.0
20,000	222.0
25,000	223.0

TABLE IV

TESTS FOR LEAST SIGNIFICANT DIFFERENCE
OF MAGNESIUM MESH SIZE VS. AVERAGE CANDLEPOWER

Least Significant Difference - 6.6

Level of Confidence, % - 95.0

Magnesium Mesh Size	Trapped Density gms/cc	Average Particle Size, Microns	Average Candlepower
20/50	1.07	437	130.0
30/50	1.08	322	154.0
50/100	1.14	168	293.0
100/200	1.08	110	285.0

TABLE V

TESTS FOR LEAST SIGNIFICANT DIFFERENCE
OF FLARE CASE COATING VS. AVERAGE CANDLEPOWER

Least Significant Difference - 8.8

Level of Confidence, % - 95.0

Flare Case Coating	Average Candlepower
Amberlac Resin 292	218.0
Laminac Resin 4116	215.0
Polyethylene 617	218.0
Paraffin Wax	210.0

TABLE VI

TESTS FOR LEAST SIGNIFICANT DIFFERENCE

Least Significant Difference - 0.61

Level of Confidence, % - 95.00

Loading Pressure, Psi	Average Burning Rate, Inches Per Minute
2,000	4.41
4,000	4.54
7,000	4.29
10,000	4.23
15,000	4.03
20,000	4.17
25,000	4.06

TABLE VII

TESTS FOR LEAST SIGNIFICANT DIFFERENCE

Least Significant Difference - 0.05

Level of Confidence, % - 95.00

Magnesium Mesh Size	Average Burning Rate, Inches Per Minute
20/50	2.62
30/50	3.01
50/100	5.66
100/200	5.84

TABLE VIII

TESTS FOR LEAST SIGNIFICANT DIFFERENCE

Least Significant Difference - 0.14

Level of Confidence, % - 95.00

Flare Case Coating	Average Burning Rate, Inches Per Minute
Amberlac Resin 292	4.13
Laminac Resin 4116	4.15
Polyethylene 617	3.61
Paraffin Wax	4.52

TABLE IX

TESTS FOR LEAST SIGNIFICANT DIFFERENCE

Least Significant Difference - 0.60

Level of Confidence, % - 95.00

Loading Pressure Psi	Burning Rate Grams Composition/second
2,000	5.06
4,000	5.37
7,000	5.38
10,000	5.49
15,000	5.48
20,000	5.58
25,000	5.57

TABLE X

TESTS FOR LEAST SIGNIFICANT DIFFERENCE

Least Significant Difference - 0.06

Level of Confidence, % - 95.00

Magnesium Mesh Size	Burning Rate Grams Composition per second
20/50	2.41
30/50	2.82
50/100	6.11
100/200	6.33

TABLE XI

TESTS FOR LEAST SIGNIFICANT DIFFERENCE

Least Significant Difference - 0.23

Level of Confidence, % - 95.00

Flare Case Coating	Burning Rate Grams Composition/second
Amberlac Resin 292	4.36
Laminac Resin 4116	4.32
Polyethylene 617	4.50
Paraffin Wax	4.50

TABLE XII

ANALYSIS OF VARIANCE TABLE FOR CANDLEPOWER VALUES

ANALYSIS OF VARIANCE

Source Main Effects	Sum of Squares	Degree of Freedom	Mean Squares	F - Ratios Calculated	Critical F - Ratios 95% Level of Confidence
Magnesium (M)	3,076,232.87	3	1,025,410.956	2220	2.60
Pressure (P)	26,818.30	6	4,469.716	5.18	2.66
Coating (C)	5,592.08	3	1,864.0266	2.29	3.86
Interactions, First Order					
M + C	7,334.05	9	814.8944	1.77	1.88
M + P	15,535.53	18	863.085	1.88	1.57
C + P	9,805.12	18	544.7289	1.90	1.66
Interactions, Second Order					
M + P + C	15,415.97	54	285.6661	<1	1.32
Residual Error	206,239.40	448	460.3558		
Total Sum of Squares	3,362,972.94	559			
Pooled Variance	460,355,800				
Standard Deviation	$\sqrt{460,355,800} = 21,455$				

TABLE XIII

ANALYSIS OF VARIANCE TABLE FOR BURNING RATE (INCHES PER MINUTE) VALUES

				BURNING RATE, INCHES PER MINUTE	
Source	Sum of Squares	Degrees of Freedom	Mean Square	F - Ratios Calculated	Critical F - Ratios 95% Level of Confidence
Main Effects					
Magnesium (M)	12,718,235	3	4,272,745	19,600	2.60
Pressure (P)	160,662	6	26,777	1.94	2.66
Coating (C)	61,301	3	20,651	15.3	3.86
Interactions First Order					
M + C	12,163	9	1,351	6.2	1.88
M + P	248,338	18	13,797	63.1	1.57
C + P	30,731	18	1,707	1.3	1.66
Interactions Second Order					
M + P + C	70,522	54	1,306	6.0	1.32
Residual Error	97,992	448	219		
Total Sum of Squares	13,399,945	559			
Pooled Variance	0.0219				
Standard Deviation	$\sqrt{0.0219} = 0.15$				

TABLE VIV

ANALYSIS OF VARIANCE TABLE FOR BURNING RATE (GRAMS PER SECOND) VALUES

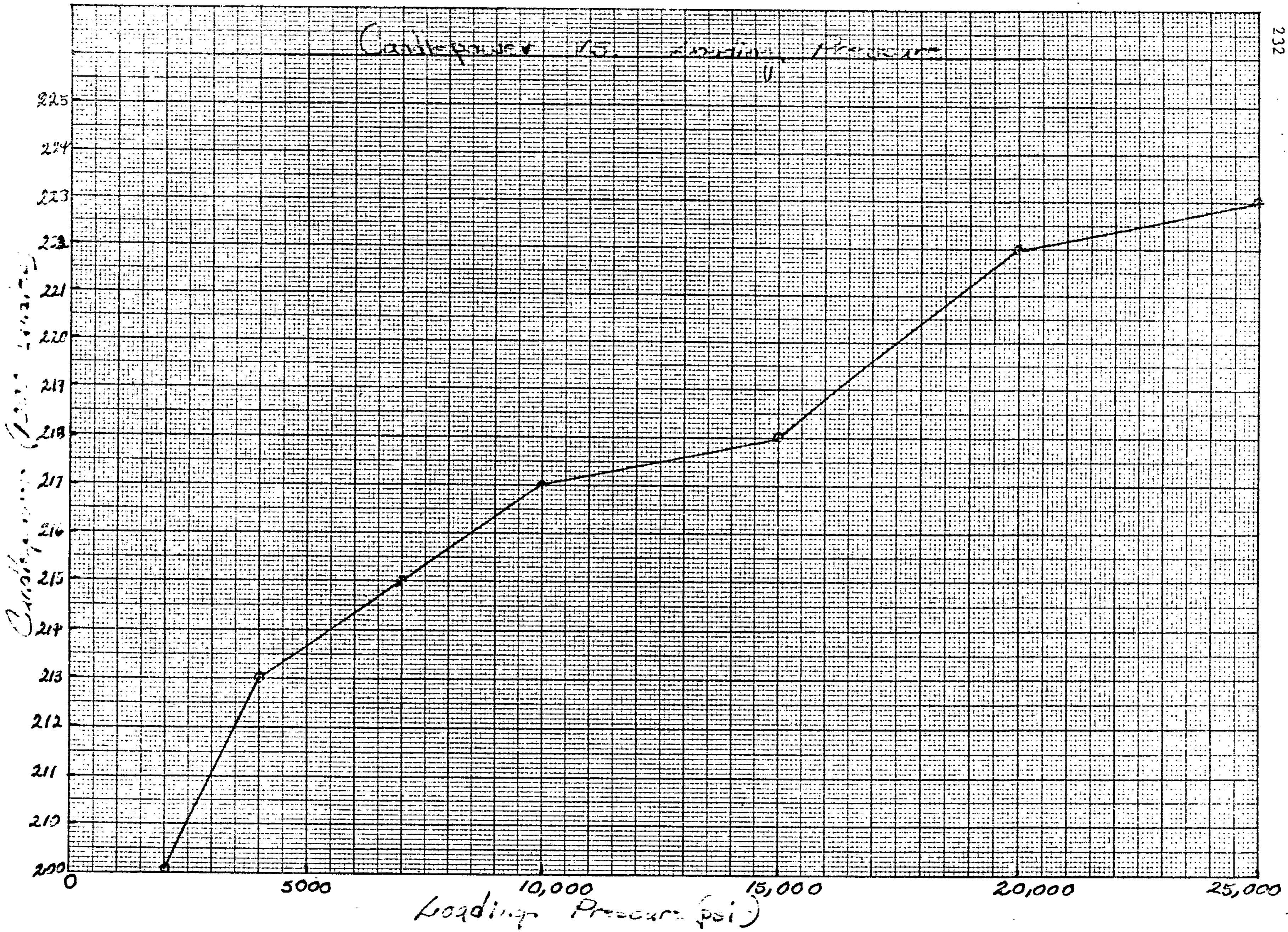
Analysis of Variance				Burning Rate, Grams of Composition/Sec.	
Source	Sum of Squares	Degree of Freedom	Mean Square	F - Ratio Calculated	Critical F - Ratio 95% Level of Confidence
Main Effects					
Magnesium (M)	18,333,445	3	6,111,148	176,000	2.60
Pressure (P)	150,184	6	25,031	1.89	2.66
Coating (C)	35,199	3	11,733	2.13	3.86
Interactions, First Order					
M + C	34,800	9	3,866	11.2	1.88
M + P	238,236	18	13,235	38.4	1.57
C + P	24,573	18	1,363	1.07	1.66
Interaction, Second Order					
M + P + C	68,748	54	1,273	369	1.32
Residual Error	154,606	448	345		
Total Sum of Squares	19,039,724	559			
Pooled Variance	0.0345				
Standard Deviation	$\sqrt{0.0345} = 0.19$				

Figure I

EXPERIMENTAL DESIGN

	M_1				M_2				M_3				M_4			
	C_1	C_2	C_3	C_4	C_1	C_2	C_3	C_4	C_1	C_2	C_3	C_4	C_1	C_2	C_3	C_4
P_1																
P_2																
P_3																
P_4																
P_5																
P_6																
P_7																

Figure II


225
224
223
222
221
220
219
218
217
216
215
214
213
212
211
210
200
0
5000
10,000
15,000
20,000
25,000
Loading Pressure (psi)