AD NUMBER: AD0837141

LIMITATION CHANGES

TO:

Approved for public release; distribution is unlimited.

FROM:

Distribution authorized to US Government Agencies only; Administrative/Operational Use; 1 Jan 1968. Other requests shall be referred to Office of Chief, Research and Development, Washington, DC 20310.

AUTHORITY

OCRD, D/A ltr dtd 14 Jul 1972

Army Science Conference Proceedings

18-21 JUNE 1968

Volume II

Principal Authors L through Z.

OFFICE, CHIEF OF RESEARCH AND DEVELOPMENT
DEPARTMENT OF THE ARMY

DISTRIBUTION STATEMENT

Each transmittal of this document outside the agencies of the U.S. Government must have prior approval of the Chief, Research and Development, Department of the Army, ATTN: CRDSTI.

P R O C E E D I N G S

O F T H E

1968 ARMY SCIENCE CONFERENCE

UNITED STATES MILITARY ACADEMY, WEST POINT, N. Y.

18 - 21 JUNE 1968

V O L U M E II

(Principal Authors L thru Z)

D R A F T

HEADQUARTERS
DEPARTMENT OF THE ARMY
OFFICE OF THE ADJUTANT GENERAL
WASHINGTON, D. C. 20315

SUBJECT: 1968 Army Science Conference

TO: See Distribution

1. Inclosed for your information and use is Volume two of the 1968 Army Science Conference Proceedings, comprised of the unclassified papers by principal authors L thru Z which were presented at the conference 18-21 June, US Military Academy, West Point, New York.

2. Volumes one and three of the Proceedings are being distributed separately.

By Order of the Secretary of the Army:

1 Incl
Vol II, 1968 Army Science
Conference Proceedings

A MATHEMATICAL MODEL FOR TRAVERSAL OF RIGID OBSTACLES BY A PNEUMATIC TIRE

A. S. LESSEM AND A. J. GREEN
MOBILITY AND ENVIRONMENTAL DIVISION
U. S. ARMY ENGINEER WATERWAYS EXPERIMENT STATION
VICKSBURG, MISSISSIPPI

PART I: INTRODUCTION

Background

The U. S. Army Engineer Waterways Experiment Station (WES) is active in a program of research in vehicle dynamics, as part of the vehicle mobility research program under the sponsorship and guidance of the Development Directorate, U. S. Army Materiel Command. One of the principal tools used in vehicle dynamics research is the mathematical model, whose main purpose is to predict the dynamic performance of vehicles. Ideally, this prediction takes place in a computing laboratory. The prediction may suggest certain vehicle design modifications, and, in addition, can serve to guide the conduct of field and laboratory tests.

The mathematical model for predicting the performance of a particular vehicle can be organized in several complementary sections. One may be a model of the terrain profile and its capacity for deformation; another, the mechanical interaction of the various members of the vehicle body and suspension system; and a third, the dynamic properties of the traction elements--tires, tracks, or other devices. This paper is concerned with the modeling of traction elements. Heretofore,* in mathematical models of vehicles, pneumatic tires have been represented by a simple, single spring, or a spring and viscous damper. Results have shown that these are inadequate.

Purpose and Scope

The purpose of this work was to develop and verify a mathematical model that simulated the ability of a pneumatic tire to envelop obstacles. The model was to be able to predict both the vertical and horizontal forces transmitted from the terrain through the tire to the vehicle axle. A realistic tire-terrain contact mechanism

* For example, (1), (2).

that considers terrain features and vehicle motions was to be made a part of the model. Only nondeforming terrain was considered. The pneumatic tire model was to be verified by simulating laboratory tests in which a tire was towed over a series of rigid obstacles, and comparing the output with the actual test results.

PART II: TEST PROGRAM

The laboratory program consisted of two series of tests: one, to obtain quasi-static load-deflection characteristics for pneumatic tires under several conditions of ply rating (PR) and inflation pressure; the other, to generate dynamic performance data for a pneumatic tire towed across a series of rigid obstacles, under different conditions of load, inflation pressure, and speed.

Load-Deflection Tests

The tires used in the load-deflection tests were mounted on a rigid axle-hub assembly and loaded by raising a motor-driven platform against the stationary tire, at various speeds. Quasi-static load-deflection curves were obtained for 9.00-14, 2-, 4-, and 8-PR, treadless pneumatic tires at inflation pressures of 0, 10, 20, 30, and 40 psi. A representative set of curves is shown in fig. 1. The curves represent several cycles of loading and unloading and show pronounced hysteresis. It became apparent early in the test series that there was no significant difference between the load-deflection curve recorded at the slowest vertical speed (0.022 in./sec) of the platform and that recorded at the fastest speed (0.143 in./sec), so all subsequent tests were conducted at the fastest platform speed.

Rigid-Obstacle Tests

Under several combinations of load, inflation pressure, and speed, a pneumatic tire was towed over a series of rigid obstacles on a nondeformable base. The test carriage and a portion of the obstacle course are shown in fig. 2. Tests were conducted using a 9.00-14, 2-PR tire inflated to 10, 20, and 30 psi. The tire was loaded to 500, 1000, and 1500 lb and was towed at speeds of 1, 2, and 3 ft/sec. Vertical axle acceleration, vertical displacement, and horizontal restraining force were recorded.

PART III: MATHEMATICAL MODEL

The model that has been used often in vehicle dynamics studies to represent a pneumatic tire is shown in fig. 3. The spring element represents the ability of the tire to recover its original shape in the absence of deforming forces; the dashpot element (when it is included) represents the dissipation of energy in the tire by means of viscous damping. Forces are transmitted through the tire to the axle in the vertical direction only. Contact between tire and terrain takes place at a single point.

The basic idea for the tire model described herein is to represent the tire by many spring elements, as shown in fig. 4. This accounts for the physically significant contributions of horizontal forces transmitted to the axle. It was anticipated that a model composed of only spring elements would be adequate for high inflation pressures and low speeds, since accompanying dashpot elements, although important to a truly realistic representation of tire compliance, were not included in the present model. It may be necessary to include dashpot elements in the model to achieve the desired degree of model accuracy for low inflation pressures and high speeds.

When the pneumatic tire is represented by many spring elements, the displacement forcing function for one spring element need not be the same as that for the neighboring spring. This feature permits small-size terrain features with abrupt changes in slope to be "enveloped" by the model. As each spring element is deflected, the spring force produced and transmitted to the vehicle axle is determined. Knowledge of the magnitude and location of this force permits a force vector to be computed and resolved into horizontal and vertical components. The contributions of the individual elements are summed vectorially and used as inputs to the mathematical model that represents the vehicle dynamic characteristics.

To convert the basic idea for the tire model into practical form, procedures were developed for division of the pneumatic tire into segments, calculation of spring coefficients for the segments, and conversion of terrain profiles into separate displacement forcing functions for the segments.

Division into Segments

As a vehicle traverses irregular, obstacle-ridden terrain, the tires contact the ground through "contact areas" that change in size as the tire-vehicle system goes through dynamic displacements. That portion of the tire carcass that undergoes flexing and deformation at any instant is identified as the "active region" (see fig. 5). The active region was assumed constant in size and large enough to include most anticipated fluctuations in contact area. In reality, the size of this active region varies with time just as does the tire contact area. As the tire rotates, different portions of the carcass are swept into, and out of, this active region. Thus, instead of focusing attention on a particular portion of the tire as it rotates about the axle, a nonrotating area between the terrain and axle is monitored. It is this nonrotating active region that is divided into segments (see fig. 5). These segments are fixed in position with respect to one another, and may be regarded as boundaries through which the physical tire passes during rotation.

Calculation of Segment Spring Coefficients

The load-deflection curves shown in fig. 1 represent conditions for which large-scale carcass flexing occurs, so that as more load is applied, a larger area of carcass is involved in restraining

this load. The load-deflection curves can be regarded as the result of joint contributions of many segments acting together. Thus, from these curves the small-scale flexural properties of any one segment in the active region can be inferred.

A concept of "effective radial deflection" was developed as an aid in calculating segment spring coefficients. In this concept, illustrated in fig. 6, the actual segment deflection encountered during a load-deflection test is replaced by a fictitious uniform radial deflection, Δ . The effective radial deflection of each segment varies as the deflection varies at the vertical reference position (VRP).

The flexure property of the tire carcass contained within each segment boundary is represented by a linear spring with coefficient K that is the same for all segments. The spring is positioned along the radial center line of each segment. With the location of each segment with respect to the VRP known, the magnitude and line of action of the force due to deflection of each spring can be determined. In the experimental setup for recording quasi-static load-deflection curves, the load cell registered the total vertical force exerted by the deflected tire carcass. The link between the analytic representation of the tire and the physical reality of its load-deflection curve can be made by using an equation of static equilibrium as follows:

$$F = 2 \sum_{i=1}^{n} K \Delta_i \cos \phi_i \qquad (1)$$

where

F = Load cell reading, lb
K = Segment spring coefficient, lb/in.
Δ_i = Effective radial deflection, i^{th} segment, in.
ϕ_i = Force direction angle, i^{th} segment, radians
n = Number of segments each side of VRP.

To apply this analytic representation to a specific tire, one point on the load-deflection curve for that tire must be used to calculate K in equation 1. If a value of carcass deflection at the VRP, δ , is selected (for instance, δ = 1 in.), and if the load cell output corresponding to this selection, F_1 , is read from a load-deflection graph, then K can be computed as:

$$K = \frac{F_1}{2 \sum_{i=1}^{n} \Delta_i \cos \phi_i} \qquad (2)$$

In this equation, the values of Δ_i corresponding to δ = 1 in. are read from a graph (fig. 7) of effective radial deflection versus carcass deflection, and the values of $\cos \phi_i$ are fixed by section geometry.

Once K is calculated, this value can be used in equation 1 to define the analytical load-deflection curve. Thus, any other δ may be selected and corresponding Δ_i obtained from the effective radial deflection graph. These values of Δ_i, together with the known values of $\cos \phi_i$, are put into equation 1 to compute the load corresponding to the selected deflection. Used in this manner, equation 1 produces a concave-upward load-deflection curve typical of pneumatic tires. In fig. 8 an experimental load-deflection curve is compared to several points produced by equation 1 using a K value determined at $\delta = 1$ in. This value was chosen on the ascending portion of the experimental load-deflection curve. The use of a linear, constant-coefficient spring as representative of the flexing of a segment of tire carcass appears to be reasonable. The basically hysteretic characteristic of the deflecting tire has been ignored for the present.

Segment Displacement Functions

With the rotating tire mathematically represented by a stationary segmented active region, the nondeformable terrain profile encountered by the tire is represented as a displacement function that traverses the active region. The segments are deflected, in sequence, by the displacement function. These sequential deflections are illustrated in fig. 9. The time of application and the deflection amplitude are different for each segment. The application times shown in the figure are calculated using dimensions of the active region and number of segments.

The computer implementation of the model must provide for generation of a terrain profile function. This function must be shifted in time to account for its sequential encounters with the segments, and it must be changed in amplitude to account for dynamic motions of the tire-vehicle system. Computer capacity allocated to the tire model determines how accurately these requirements are met.

PART IV: APPLICATION TO TRAVERSAL OF RIGID OBSTACLES

Obstacle Test Simulation

The tire model was combined with a model for a test carriage, and the overall model was used to predict the responses recorded in the obstacle tests. Study of the physical structure of the test carriage suggested the composite tire-carriage model shown in fig. 10.

The model of the pneumatic tire was organized with the active region divided into 10 segments, and the segment spring coefficients were calculated for several inflation pressures.

Numerical values for the carriage spring and damping coefficients were obtained by repeating the obstacle tests using a rigid aluminum wheel, thus revealing the dynamic properties of the carriage alone. The values obtained were reasonably independent of carriage speed.

In the writing of the equations of motion, it was assumed that (a) the obstacle does not deform, (b) there is no slip between tire and obstacle under towed conditions, (c) no forces are generated parallel to the wheel axle, (d) the carriage towing speed remains constant, and (e) the pneumatic wheel load remains constant. The nomenclature used in the equations is as follows:

B_H Horizontal carriage damping coefficient, lb/in./sec
B_V Vertical carriage damping coefficient, lb/in./sec
F_o Pneumatic load applied to tire in excess of, or in opposition to, deadweight load
g Acceleration of gravity
H_{ri} Threshold height of i^{th} segment in equivalent rigid wheel, in.
K Segment spring coefficient, lb/in.
K_H Horizontal carriage spring coefficient, lb/in.
m Inertial mass of test carriage, lb sec^2/in.
x Horizontal axle displacement, in.
Y_i Vertical obstacle height beneath i^{th} segment, in.
z Vertical axle displacement, in.
δ Center-line static deflection, in.
Δ_i Deflection of i^{th} segment, in.
ϕ_i Location angle for i^{th} segment

The equations of motion are as follows:

$$m\ddot{z} = \sum_{i=1}^{10} K \Delta_i \cos \phi_i - B_V \dot{z} - g \left(m \pm \frac{F_o}{g} \right) \tag{3}$$

$$m\ddot{x} = \sum_{i=1}^{10} K \Delta_i \sin \phi_i - B_H \dot{x} - K_H x \tag{4}$$

where

$$\Delta_i = \begin{cases} Y_i - H_{ri} - z \ , & Y_i - H_{ri} - z \geq 0 \\ 0 \ , & Y_i - H_{ri} - z < 0 \end{cases} \tag{5}$$

Equation 3 is used to calculate the resultant vertical force, equation 4 the resultant horizontal force, and equation 5 the deflection of each segment. This segment deflection, Δ_i, is permitted to have positive values only; negative values are replaced by zero. This corresponds to an assumption that the tire may only be compressed by the obstacle and not stretched by it.

The rigid-wheel threshold height, H_{ri}, is illustrated in fig. 11. Each H_{ri} gives the height of the i^{th} segment contact point above the ground when the tire has no static deflection. The H_{ri} concept serves two purposes: first, to permit the tire model to display realistic static deflections by invoking as many segments as needed to restrain an applied load, and second, to modify the

height of obstacle displacement functions as required by the height of the contact point of each segment.

Computer Implementation

The tire-carriage model was run on a general-purpose analog computer. The overall organization of the computing elements is shown in block-diagram form in fig. 12. The segmented spring computer diagram is shown in fig. 13; the vertical and horizontal computer diagram for the carriage dynamics, in fig. 14.

The generation of segment obstacle functions required the shifting process discussed previously under the heading "Segment Displacement Functions." This was accomplished by using the track-hold capability of the analog computer. The shifting interval controls the accuracy of the quantizing of the input obstacle function. Hence, a trade off between allocated computer space and desired quantizing accuracy is generally necessary for this type of problem. A chain of 34 track-hold amplifier pairs was used in this study; the organization is shown in fig. 15. A representative computer output is shown in fig. 16. The upper trace shows the time history of obstacle heights as seen by the active region of the tire.

Comparison of Computed and Observed Results

Some computed dynamic motions and their laboratory counterparts are compared in fig. 17. This figure shows responses to traversal of a 2- by 8-in. rigid rectangular obstacle. The particular responses shown are the horizontal restraining force exerted on the axle by the carriage and the vertical displacement of the axle.

The comparison of computed and observed responses was essentially qualitative. This was consistent with the use of the crude linear representation of the test carriage. The most desired feature of the composite tire-carriage model, the ability to reproduce the basic features of the response wave shapes as seen in the laboratory, was realized; the composite model, through the segmented representation of the pneumatic tire, was capable of producing responses with realistic wave shapes. The positively and negatively directed horizontal restraining forces and the vertical axle displacements were displayed with basically correct form. If a point-contact model had been used, the response time would have been identical with the time for the point to traverse the obstacle. The multiple-contact model under discussion produces responses whose durations are realistically extended beyond the corresponding obstacle durations. A point-contact model would have been incapable of producing the horizontal forces; the multiple-contact model does so in a natural way.

A quantitative comparison of the computed and observed responses revealed several discrepancies that were related to the simple linear model of the test carriage and to the finite number of segments in the tire model. These can be overcome by more refined modeling (3).

PART V: SUMMARY

The data shown indicate that a mathematical representation of a pneumatic tire, in terms of the deflections of many radial segments, successfully displays the essential feature of horizontal and vertical forces transmitted through the tire. It is indicated further that the segmented tire model enables realistic predictions to be made of the displacement and force time histories for a pneumatic tire towed over a rigid obstacle.

The success of the segmented wheel model suggests that the energy dissipation properties of the tire can be segmented in the same manner as the deflection properties can be segmented, that the model can be refined and enlarged to properly represent a powered wheel traversing deformable obstacles, and that representation of the compliance of other types of traction elements (e.g. tracks) by a realistic series of segments may be possible.

REFERENCES

1. Prasiloski, J. C., and Heal, S. F., "M52-XM682 Suspension and Fifth Wheel Study," Report No. RRC-6, May 1964, U. S. Army Tank Automotive Center, Warren, Michigan.
2. FMC Corporation, "A Computer Analysis of Vehicle Dynamics While Traversing Hard Surface Terrain Profiles," Contract Report No. 3-155, Feb 1966, U. S. Army Engineer Waterways Experiment Station, CE, Vicksburg, Mississippi.
3. Lessem, A. S., "Dynamics of Wheeled Vehicles; Report 1, A Mathematical Model for the Traversal of Rigid Obstacles by a Pneumatic Tire," U. S. Army Engineer Waterways Experiment Station, CE, Vicksburg, Mississippi. (Publication pending.)

ACKNOWLEDGMENT

Acknowledgment is made to personnel of the Simulation Laboratory of the George C. Marshall Space Flight Center, National Aeronautics and Space Administration, Huntsville, Alabama, for assistance and use of their computing facilities.

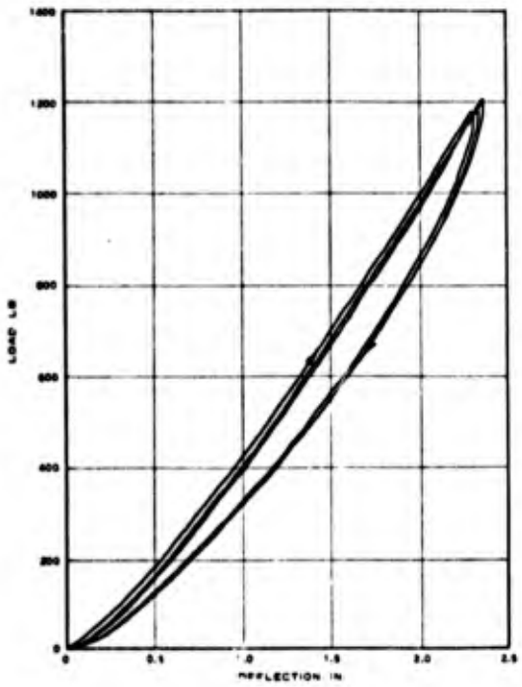


Fig. 1. Representative load-deflection curves; 9.00-14, 8-PR tire, 10-psi inflation pressure

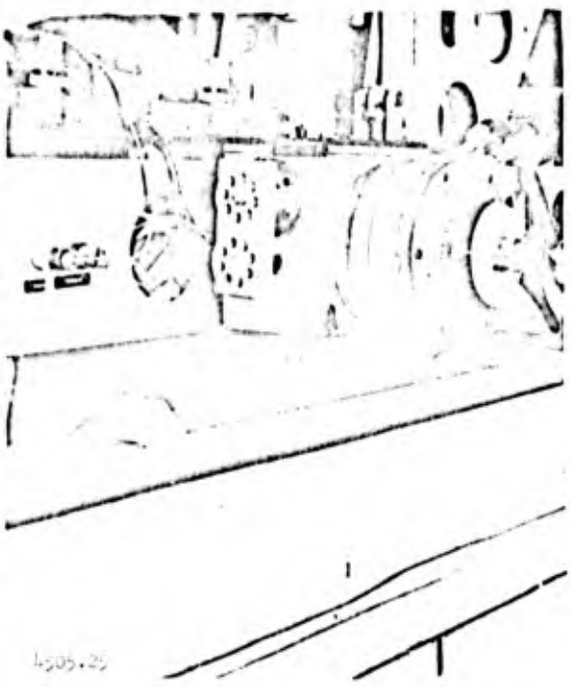


Fig. 2. Test carriage and portion of obstacle course

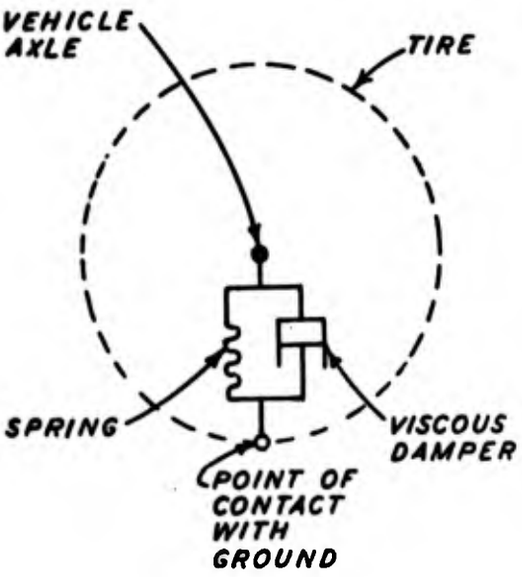


Fig. 3. Point-contact model

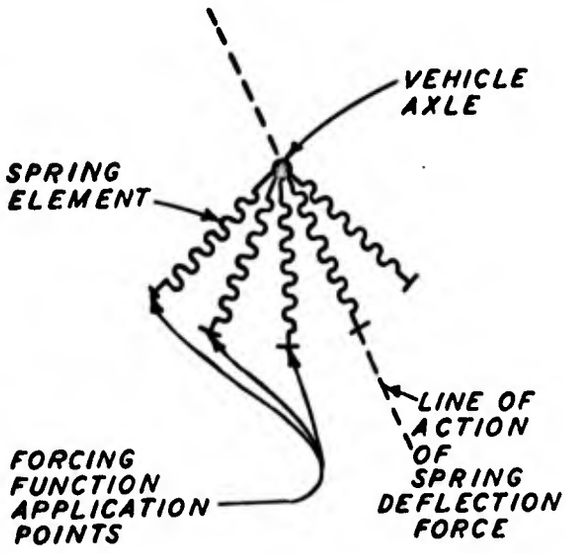


Fig. 4. Extended-contact model

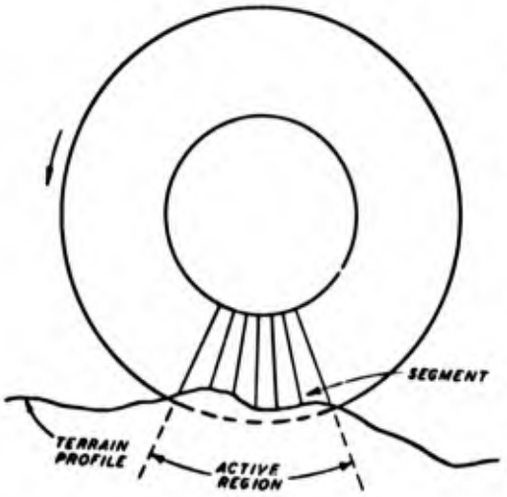


Fig. 5. Segmented, nonrotating, active region

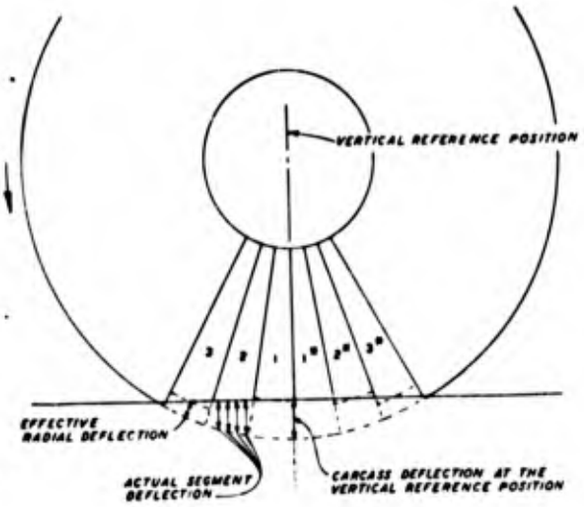


Fig. 6. Illustration of concept of effective radial deflection

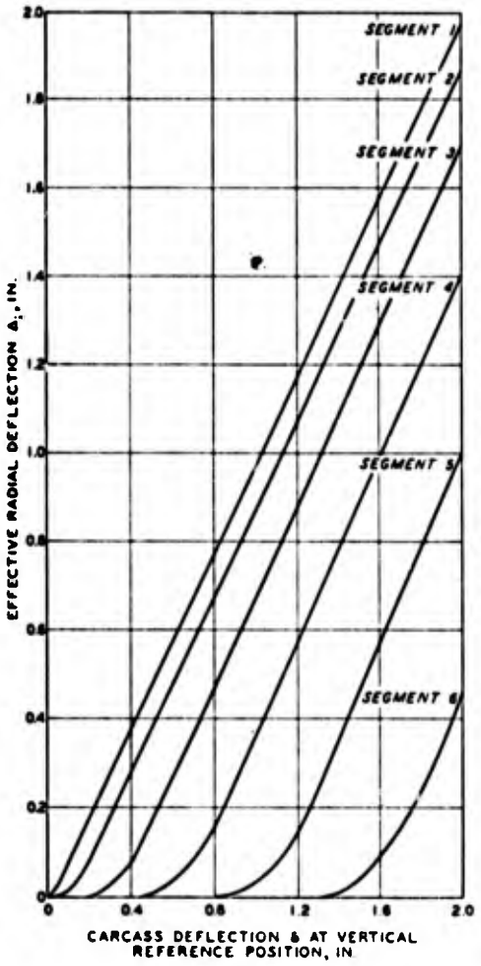


Fig. 7. Representative radial deflection curves

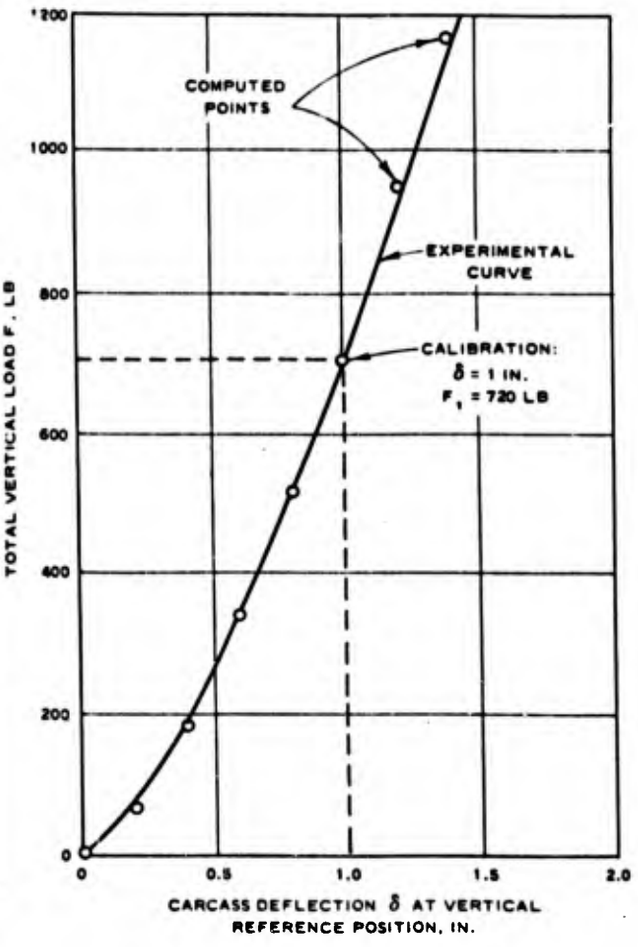


Fig. 8. Experimental load-deflection curve

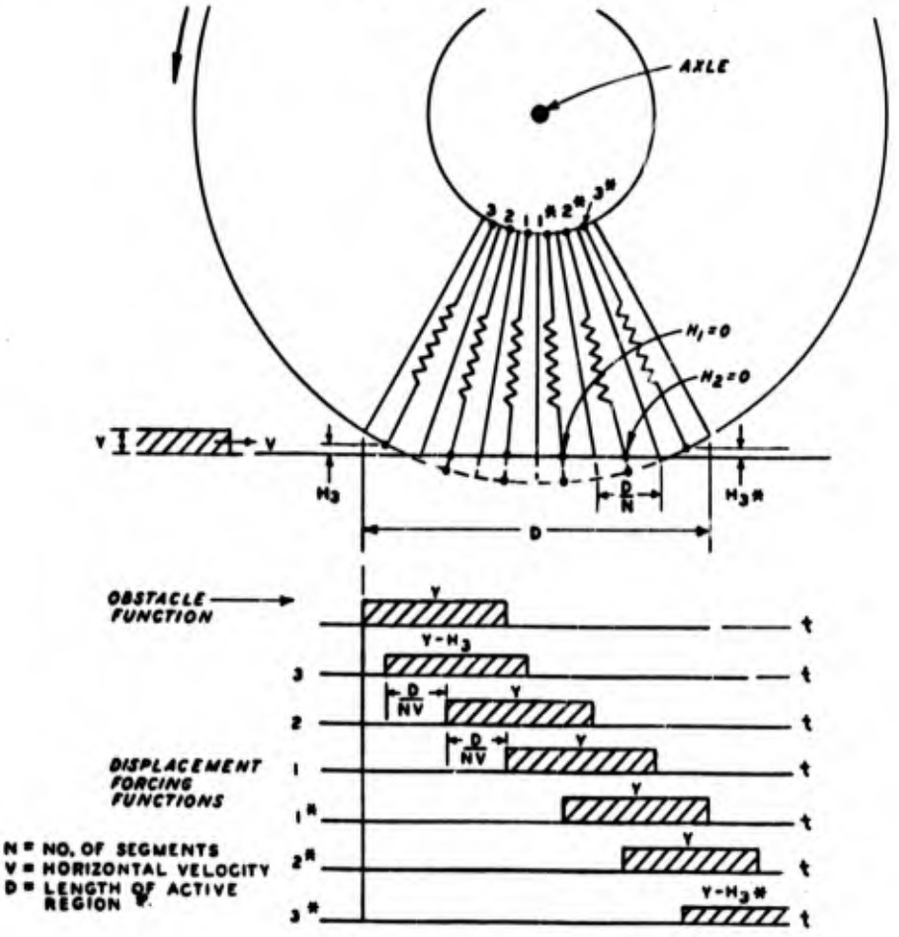


Fig. 9. Sequential deflections

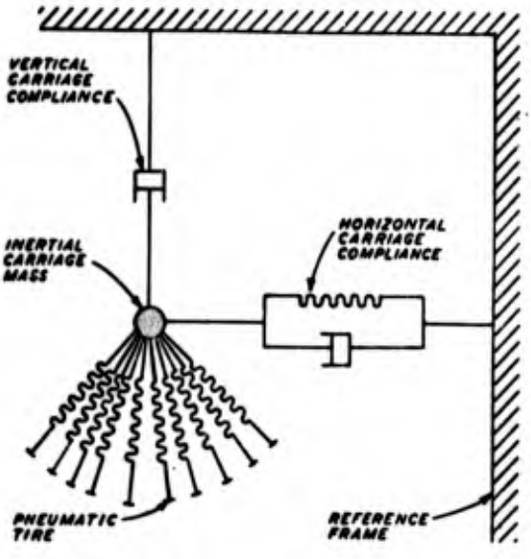


Fig. 10. Composite tire-carriage model

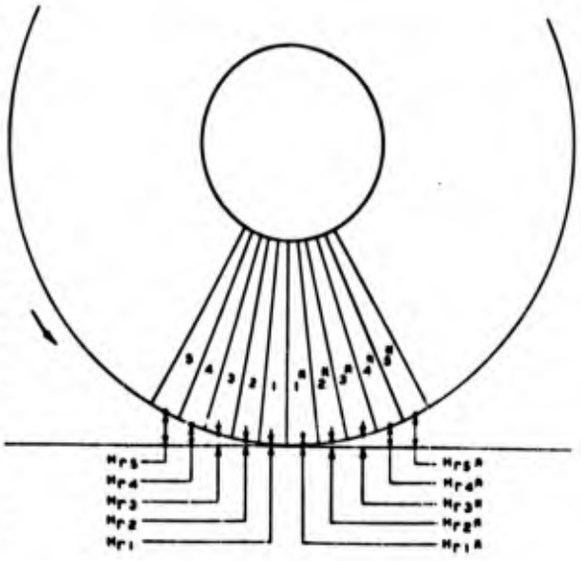


Fig. 11. Rigid-wheel threshold height

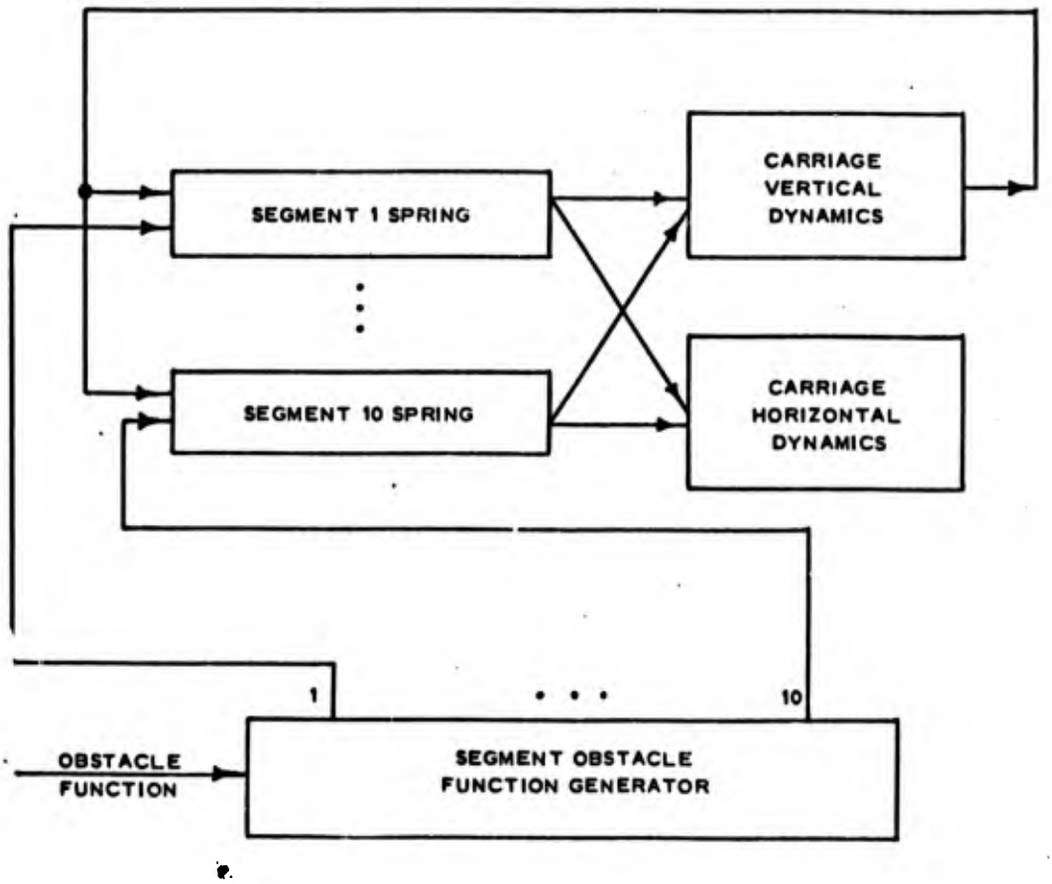


Fig. 12. Organization of computing elements

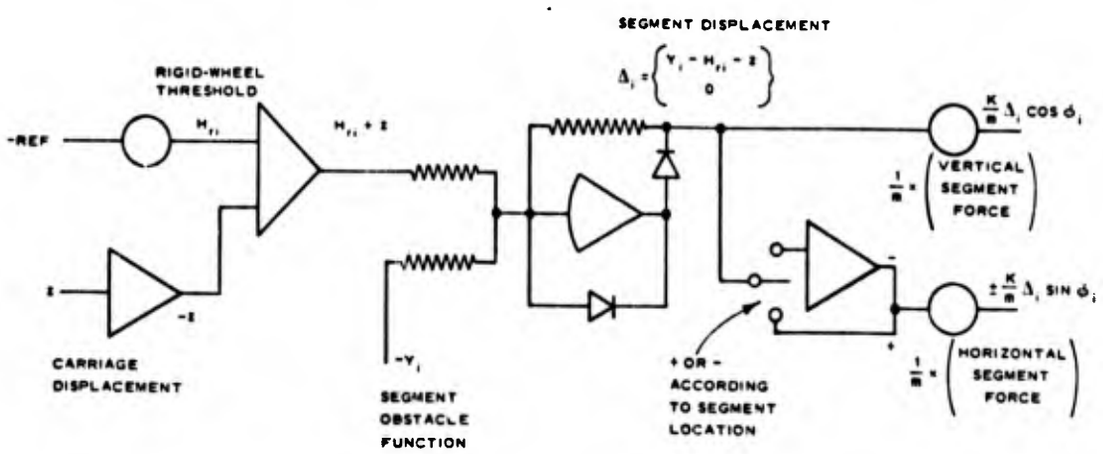


Fig. 13. Segmented spring computer diagram

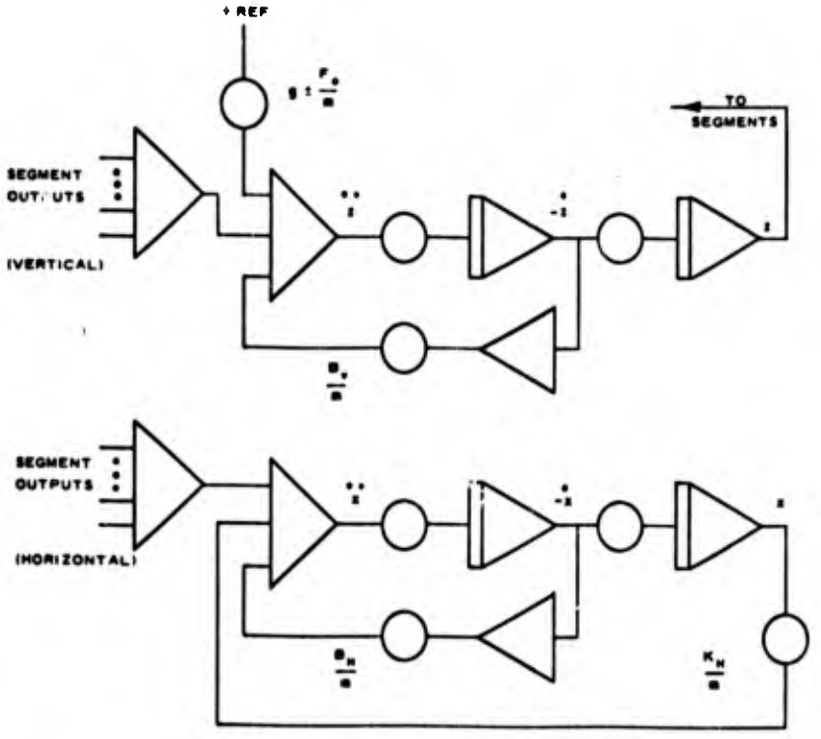
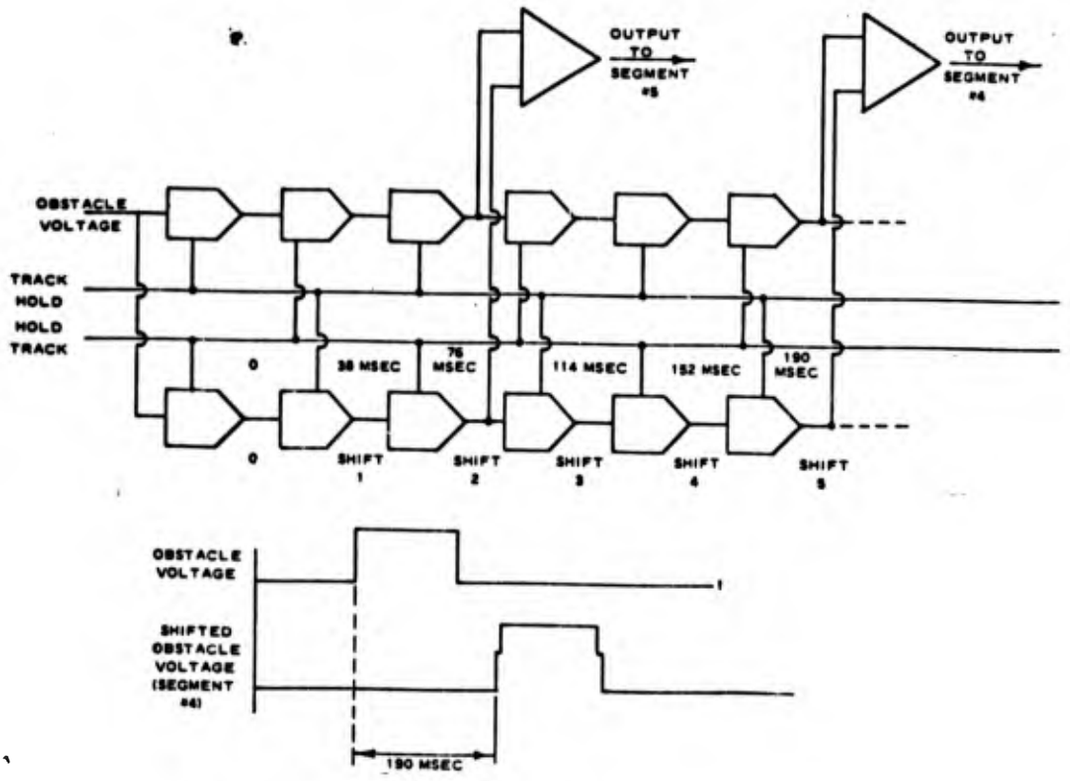


Fig. 14. Computer diagram for carriage dynamics



• Fig. 15. Organization of track-hold chain

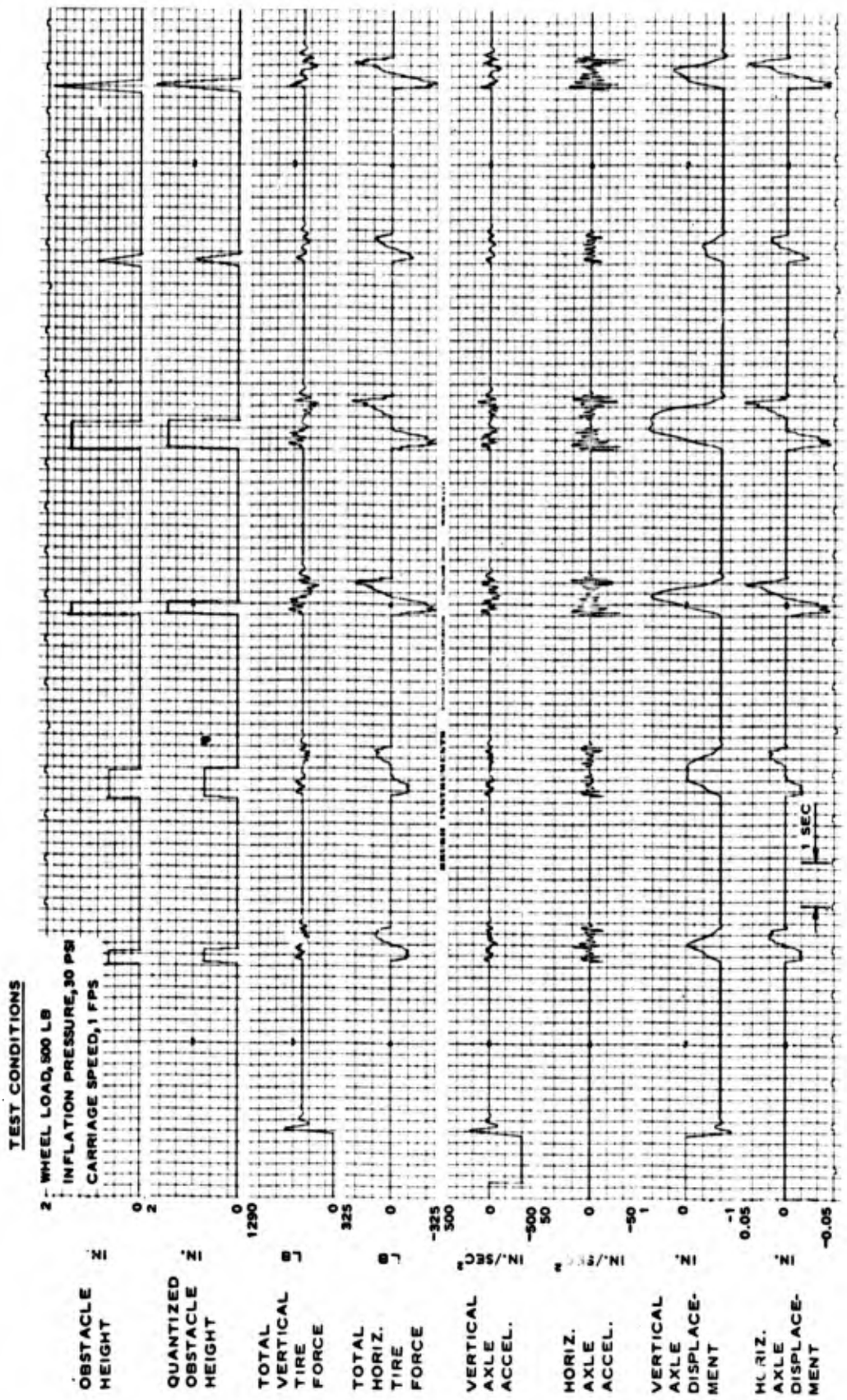
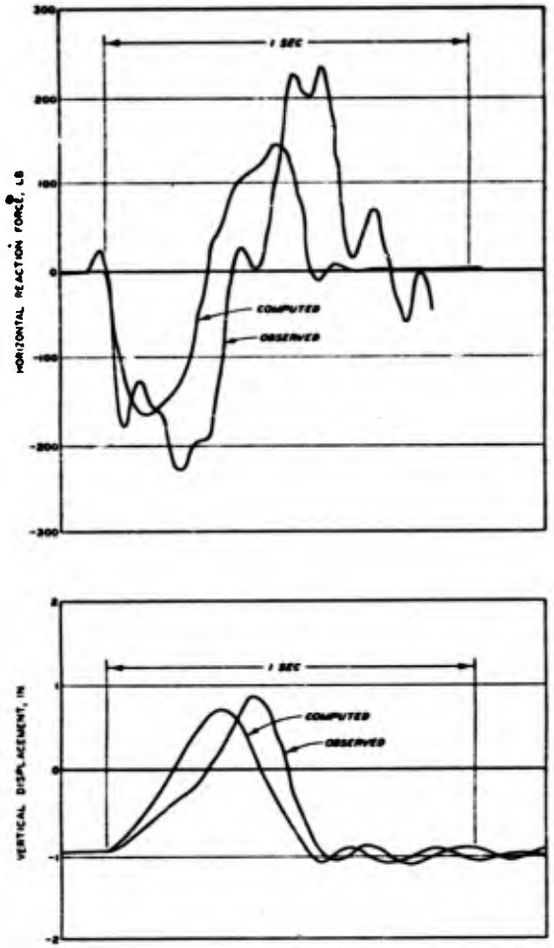
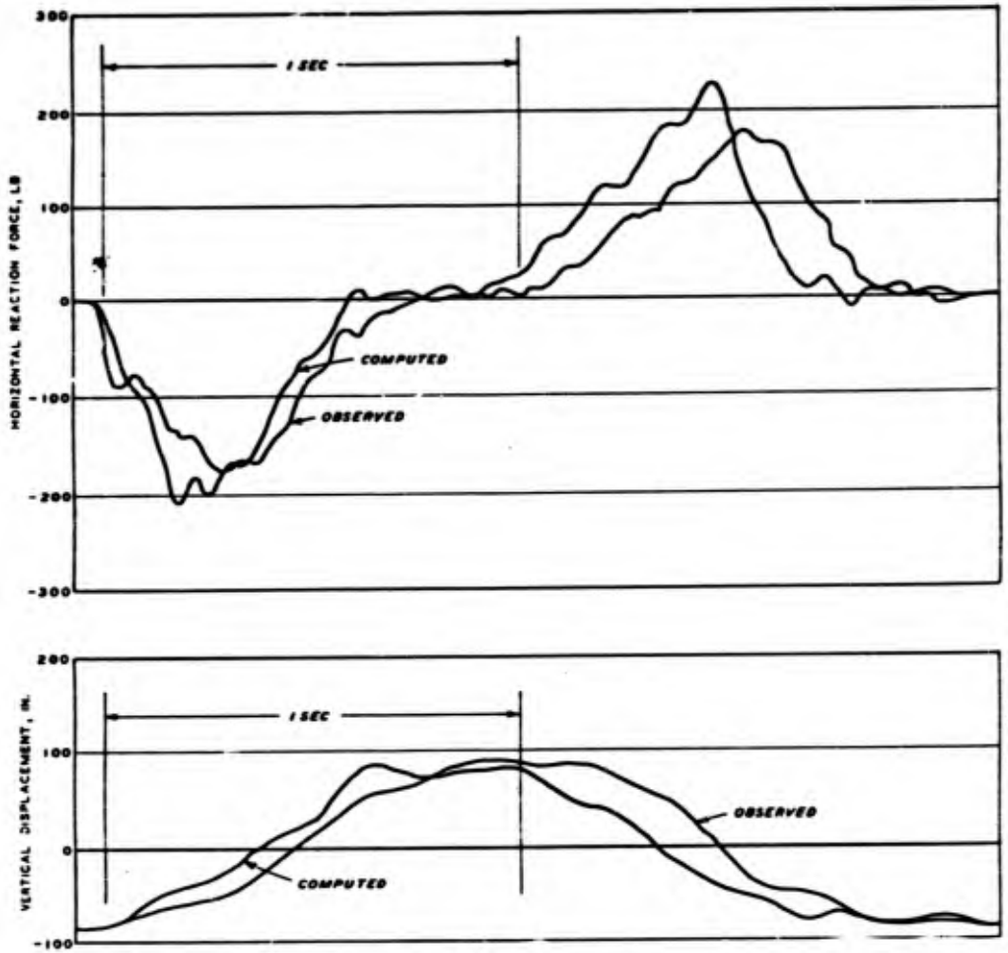


Fig. 16. Representative computed results



a. 1000-lb load, 20-psi inflation pressure, 3-fps carriage speed



b. 500-lb load, 30-psi inflation pressure, 1-fps carriage speed

Fig. 17. Comparison of empirical and analytical results (traversal of 2- by 8-in. rigid obstacle)

EMOTIONAL RESPONSE OF RHESUS MONKEYS TO CHRONIC PSYCHOLOGICAL STRESS

MURRAY D. LEVINE and THOMAS P. GORDON
BEHAVIORAL RESEARCH LABORATORY
USA HUMAN ENGINEERING LABORATORIES
ABERDEEN PROVING GROUND, MD.

This paper will summarize the results of some recent investigations in the area of emotional responsiveness of rhesus monkeys to psychological stress. The similarity between this data and that obtained in recent work with humans under stress has led to an examination of the significant variables accounting for the data, and to a proposal for a multidisciplinary research approach in this area.

Shock avoidance has become a standard psychological stress paradigm. In this situation, an animal is required to press a bar in order to avoid a shock which cycles at constant intervals. Most subjects (S's) learn this response quickly and after several hours avoid all but a few shocks. Bar press rate becomes stable for any individual S and can be maintained for long periods of time.

This paradigm has been accepted as a psychological stress because the S's are avoiding (anticipating) a painful stimulus, and they continue to respond (are motivated), and because many S's have developed lasting physiological changes as a consequence of being exposed to the paradigm. These psychosomatic symptoms include such diverse effects as experimentally induced obesity (25), prolonged increase in alcohol ingestion (5), the development of peptic ulcers (3, 24) and intestinal intussusceptions (8), as well as the exacerbation of experimentally induced diseases (13).

Some experiments, however, have failed to replicate these psychosomatic effects of chronic shock avoidance (2, 7, 19). There is even some evidence that monkeys on an avoidance schedule show a marked resistance to the effects of injections of polio virus when compared with controls (13). Because of such contradictary results, some experimentors have attempted to verify that this paradigm is a stressor by making concomitant measures of adrenal cortical excretion during the avoidance sessions (16). Adrenal cortical excretion estimated by the urinary or plasma 17-hydroxycorticosteroid (17-OHCS)

level has been shown to be associated with what could be termed emotional arousal. For instance, the 17-OHCS level rises sharply when a monkey is first housed in a new cage (17), presented with a snake in a plastic box (14), during the initiation of a shock avoidance schedule (18), and during a conditioned emotional response paradigm (15). Humans, before a college test (6), prior to elective surgery (20), or during life crises (4, 9, 10) also show a sharp rise in adrenal cortical output.

The above are acute stressors and this is reflected in the transitory nature of the 17-OHCS elevation. More chronic stressors, however, add a new level of complexity. For instance, Friedman (9) studying parents whose children were in the terminal stages of leukemia found large individual differences in adrenal cortical output in response to the various crisis periods. The authors have interpreted those cases where there was a crisis with no concomitant 17-OHCS increase as instances where the individual's psychological defences were effective in shielding him from the potential stressor. Adrenal cortical elevations are thus viewed as a response to a threat for which the subject is not adequately defended. Reviewing his own extensive research, Mason (15) notes that large individual differences in response to a potentially stressful situation have been "...a monotonously regular observation in psychoendocrine research".

Research in this laboratory has produced similar individual differences in responsiveness. In a previous study (11) using rhesus monkeys on a free-operant avoidance schedule, with urinary 17-OHCS level as an index of stress, we concluded that for most S's the initial introduction into the apparatus and the training to the operant task were indeed stressful. During the experimental period (6 hour work/rest cycle for 30-45 consecutive days) however, there were large individual differences in 17-OHCS output; some S's remaining elevated for several weeks, others returning quickly to quite low levels. One subject remained elevated for the entire period (45 days) and on autopsy had a frank gastric ulcer. All other subjects showed no evidence of pathology. The differences in 17-OHCS output can be seen in Figure 1 which shows this animal (P30) and a subject who adapted to the paradigm (SL). Figure 2 illustrates the possible irrelevance of the experimental variable. These two Ss are a yoked experimental (778) and control (774) pair and their 17-OHCS levels are virtually indistinguishable. Certainly the avoidance program alone cannot necessarily be considered as a stressor, psychological or otherwise. Further interpretation of these results was difficult because of the lack of a clear baseline level of 17-OHCS for most of the S's (11 of the 14 S's had higher hormone levels during the adaptation period than during the stress period).

In a second study (12), we attempted to develop basal environmental conditions and baseline 17-OHCS levels, as well as to demonstrate further these large individual differences in responsiveness to the stress paradigm. Eight male rhesus monkeys were placed

in a familiar metabolism cage situated in the colony room and their urine was collected for one week. The S's were then placed in the experimental booths and left undisturbed until their 17-OHCS levels returned to the colony room level (usually three weeks). The S's were then trained for 5 days and returned to the colony room where a dominance hierarchy was established. All S's were ranked on three measures: 17-OHCS level during training, time to learn the avoidance task, and aggressiveness as estimated from position in the dominance hierarchy. The two subjects who were maximally different on these measures were chosen for further testing. This was done to maximize the possibility of obtaining differences in adaptation. These 2 animals were returned to the booth and maintained on the shock avoidance schedule for 60 consecutive days.

In Figure 3, the mean (n=8) 17-OHCS level in the colony room (metabolism cage) is 0.53 mg/day. At the end of three weeks in the experimental apparatus (booth adaptation), all S's returned to a level below 0.50 mg/day. The two S's chosen for further testing (Figure 4) have met cage and booth adaptation levels of 17-OHCS below 0.40 mg/day, but show remarkably different levels during the experimental period (60 days shock avoidance). Subject 180 was the most agressive animal of the 8, took longest to learn the avoidance task (10 hours), and had the highest 17-OHCS response to training. Animal 153 was the least aggressive, took the shortest time to learn the task (3 hours), and had the lowest 17-OHCS response to training. The minimal responsiveness of animal 153 suggests a possible hypoadrenal-cortical or hypopituitary state. The 17-OHCS response to ACTH (Acthar-Armour), and Metyrapone (Metopirone-CIBA), (Figure 4) indicates a normal adrenal cortex and adequate pituitary ACTH reserve. Animal 153 is just as physiologically capable of elevated 17-OHCS levels as S 180 is capable of minimal output.

These large differences in chronic 17-OHCS levels are not attributable to differences in performance on the schedule. These S's have virtually the same response rate and receive few ($\bar{x}$ = 22/day) shocks . The 17-OHCS differences seem to be primarily a function of the psychological state of the S's. These data, of course, are biased since the S's were initially chosen for their differences, but the difference does illustrate what can be found in even a small group of S's.

In order to further clarify the individual difference phenomenon, an additional group of subjects was maintained for 30 days on essentially the same schedule. Of the total of 19 S's, 8 exhibited high 17-OHCS level and 6 were low throughout the experimental period. The remaining 5 S's were of mixed type and were placed in a questionable category. The mean 17-OHCS levels for the high and low groups are shown in Figure 5. The groups also differ on the three criteria used in the previous study. The high group has a higher initial 17-OHCS response to training, takes longest to learn to avoid the task ($\bar{x}$ = 10 hrs.), is characterized as

aggressive. The low group, on the other hand, has substantially lower 17-OHCS levels during training, takes less time to learn the avoidance task ($\bar{x}$ = 5 hrs.), and is characterized as non-aggressive. Most of the S's in this group were given a free shock (F. S.) period (36 F. S. per 6 hour avoidance period) for some index of lability. In Figure 6, the mean 17-OHCS level for the 2-day periods, before, during and after the free shock, are shown. The S's all respond, but again, a large difference between the groups is evident.

Several conclusions seem to emerge from this experimental series.

1. Most S's show an elevation in 17-OHCS during the initial phases of adaptation to a new environment and during training. We have interpreted this rise as indicative of a rise in emotionality.

2. While the initial stages of the shock avoidance schedule usually prove stressful for all subjects, the day-to-day confrontation with the schedule is not necessarily stressful. Although not all S's are emotional in this chronic phase, all are motivated enough to maintain a stable response and a low shock rate. Those S's who do not adapt to the situation even though they have control of the experimental contingencies present a problem.

3. The non-adaptors, we hypothesize, are aggressive, highly dominant S's who have been socially reinforced for aggression. Since this (presumably) well learned response to a threat is relatively ineffective in the unnatural experimental situation, the S's remain emotional even though they are able to exert control over the shock contingency. This interpretation is partially supported by the anecdotal evidence obtained during training which indicates that the differences in learning time were probably a function of the S's emotionality. The high, non-adapting S's were agitated during the training period, spending a majority of their time attacking the shock grid and cage. The adapting S's, however, were characterized as quite unemotional during the training period and spent much of their time examining the cage. These behavioral differences, of course, enhanced the chance that the adapting S's would find the bar.

What we are suggesting here is that the interaction between a subject and a potential stressor can be predicted if one knows something about the S's past experiences and habits. This necessity for a more historical approach is shown very clearly in work with human subjects.

A psychiatric population could be considered a prime example of a group under chronic psychological stress and one would predict adrenal cortical activation as a predominant syndrome. Sachar, using both schizophrenic (22) and severely depressed patients (23), noted that the adrenal cortical output was not generally

elevated. However, there was a rise in 17-OHCS for those patients who were having most difficulty defending against their environment or for those whose defenses were challenged through therapy. Conversely, the patients with well-entrenched psychotic symptomatology showed the lowest adrenal cortical levels.

There are several additional studies with normal subjects which tend to reinforce the hypothesis that there is not a one-to-one correlation between potential stressors and a stress response. In one study (1), 8 helicopter medics were followed for several months during actual combat duty in Vietnam. Urine was collected on flying and non-flying days for adrenal cortical comparisons. Surprisingly, the 17-OHCS levels for these 2 periods showed no differentiation. In fact, the overall 17-OHCS level was generally depressed from normal values. The subjects were observed to have well-ordered psychological defenses which served to minimize the potential threat in the environment. As the authors note "...flying vs. non-flying cannot be interpreted alone as stress vs. nonstress, nor can the threat of death and mutilation be interpreted as stress without consideration of the manner in which each individual perceived the threat." In another study, Rose et al (21) attempted to use personality characteristics to predict the 17-OHCS excretion levels of human subjects exposed to a chronic stress situation (basic training). Using personal history questionnaires, psychiatric interviews and field observation, the authors made judgments about the emotional state of each of the subjects as well as about the manner in which each typically coped with potentially stressful situations. Using these variables, they correctly predicted adrenal cortical response levels for 70% (19 of 27) of the subjects. They concluded that the results support the hypothesis that 17-OHCS secretion is elevated when individuals are unable to cope with a situation, and low when psychological defenses are adequate.

The human and monkey data seem to be leading in similar directions: Objectively stressful situations (the threat of death or mutilation) or previously stressful situations (shock avoidance) seem to lose, under some circumstances, their ability to provoke an adrenal cortical response. This is not automatic, however. Some monkeys and some humans seem to have difficulty in achieving this disassociation. In humans, it would seem that a rise in 17-OHCS is associated with a cognitive evaluation which leaves the subject feeling helpless to control the situation. The monkey data can be interpreted similarly. We hypothesize that the monkeys' well-ordered and heavily reinforced habits of aggressiveness in the face of threat are of little use during chronic shock avoidance. Their consequent non-adaptation to the program is analogous to a "failure of defenses". In both instances, the physiological response to a stimulus is modified by past associations and habits. This is certainly not a novel possibility, but, interestingly, both the monkey and human data suggests that the outcome of this response-response interaction can be monitored via the adrenal cortical

response.

This close correspondence between the human and animal work has led to a joint research project between this laboratory and the Neuroendocrine Department of Walter Reed Army Institute of Research. This multidisciplinary approach will attempt to test experimentally some of the hypotheses generated by this research.

A laboratory for human research is being developed which will allow careful control and study of subjects for a prolonged (3-5 month) interval. The S's will be assessed via psychiatric interviews and psychological tests in order to generate testable hypotheses concerning the relationship between selected personality variables and response to chronic and acute stresses. Performance in specially designed laboratory tasks and potentially threatening field tests will be our main dependent behavioral variables. Physiological performance will be measured in terms of the effects of prolonged stress on endocrine function, specifically cortisol and its major metabolites, and testosterone, as well as on selected autonomic measures. Endocrine state during more acute stress will be evaluated via plasma growth hormone, free fatty acids, plasma cortisol, and urinary epinephrine.

Future investigations using monkeys will be designed to assess the effects of variables associated with experimental designs not practicable with human subjects. A large free roaming colony is being established where dominance and aggressiveness can be observed in a social living environment. Experimental environments will be designed to establish passive and aggressive avoidance tasks. Further studies are being designed to investigate the physiological effects of prolonged 17-OHCS suppression in the face of chronic "stress", as well as to chart the behavioral and physiological consequences of prolonged 17-OHCS elevations.

This dual subject, multidisciplinary approach will, we hope, help gain some insight into the process called psychological stress.

BIBLIOGRAPHY

1. Bourne, P. G., R. M. Rose, and J. W. Mason. Urinary 17-OHCS levels data on seven helicopter ambulance medics in combat. Arch. Gen. Psychiat. 17: 104-110, 1967.

2. Brady, J. V. Behavioral stress and physiological change. A comparative approach to the experimental analysis of some psychosomatic problems. Trans. N. Y. Acad. Sci. 483-496, 1965.

3. Brady, Joseph V., Robert W. Porter, Donald G. Conrad, and John W. Mason. Avoidance behavior and the development of gastroduodenal ulcers. J. Exp. Analysis Behav. 1: 69-72, 1958.

4. Bunney, W.E., J. W. Mason, J. F. Roatch, and D. A. Hamburg. A psychoendocrine study of severe psychotic depressive crises. Am. J. Psychiat. 122: 72-80, 1965.

5. Clark, R. and E. Polish. Avoidance conditioning and alcohol consumption in rhesus monkeys. Science. 132: 223-224, 1960.

6. Fishman, J. R., D. A. Hamburg, J. H. Handlon, J. W. Mason, and E. Sachar. Emotional and adrenal cortical responses to a new experience. Arch. Gen. Psychiat. 6: 271-278, 1962.

7. Foltz, E. L., and F. E. Millet. Experimental psychosomatic disease states in monkeys. I. Peptic ulcer - "Executive Monkeys". J. Surg. Res. 4: 445-453, 1964.

8. Foltz, E. L., F. E. Millett, D. E. Beber and J. F. Alksne. Experimental psychosomatic disease states in monkeys. II. Gut hypermobility. J. Surg. Res. 4: 454-464, 1964.

9. Friedman, S. B., J. W. Mason, and D. A. Hamburg. Urinary 17-hydroxy-corticosteroid levels in parents of children with neoplastic disease. Psychosom. Med. 25: 364-376, 1963.

10. Hamburg, D. A. Plasma and urinary corticosteroid levels in naturally occurring psychologic stresses. Proc. Publ. Ass. Res. Nerv. Ment. Dis. 40: 406-413, 1962.

11. Levine, M. D., T. P. Gordon, W. J. Johnson, and R. M. Rose. Chronic free operant avoidance as a psychological stress: A re-evaluation. Submitted to journal of Physiology and Behavior.

12. Levine, M. D., T. P. Gordon, and R. M. Rose. Individual differences in urinary 17-OHCS levels during chronic free operant avoidance. In preparation.

13. Marsh, J. T. Experimental studies on response to viral infection. In: Symposium on Medical Aspects of Stress in the Military Climate. Washington, D. C., Walter Reed Army Institute of Research. 1964, pp. 489-501.

14. Mason, J. W. Psychological influences on the pituitary-adrenal cortical system. In: Recent Progress in Hormone Research, edited by C. Pincus. New York: Academic Press Inc., 1959.

15. Mason, J. W. Psychoendocrine approaches in stress research. In: Symposium on Medical Aspects of Stress in the Military Climate, Washington, D. C. Walter Reed Army Institute of Research, 1964, pp. 375-421.

16. Mason, J. W., J. V. Brady, W. W. Tolson, J. A. Robman, E. D. Taylor, and E. H. Mougey. Patterns of thyroid, genadal, and adrenal hormone secretion related to psychological stress in the monkey. Psychosom. Med. 23: 446-454, 1961.

17. Mason, J. W., C. T. Harwood, and R. R. Rosenthal. Influence of some environmental factors on plasma and urinary 17-hydroxy-corticosteroid levels in the rhesus monkey. Am. J. Physiol. 190: 429-433, 1957.

18. Mason, J. W., G. T. Mangon, Jr., J. V. Brady, D. Conrad, and D. McK. Rioch. Concurrent plasma epinephrine, norepinephrine and 17-hydroxy-corticosteroid levels during conditioned emotional disturbances in monkeys. Psychosom. Med. 23: 344-353, 1961.

19. Polish, E., J. V. Brady, J. W. Mason, J. S. Thach, and W. Niemeck. Gastric contents and the occurrence of duodenal lesions in the rhesus monkey during avoidance behavior. Gastroenterology. 43: 193-201, 1962.

20. Price, D. B., M. Thaler, and J. W. Mason. Preoperative emotional states and adrenal cortical activity. Studies on cardiac and pulmonary surgery patients. A.M.A. Archives Neurol. Psychiat. 77: 646-656, 1957.

21. Rose, R. M., R. O. Poe, and J. W. Mason. Psychological state and body size as determinants of 17-OHCS excretion. Arch. Int. Med. In press.

22. Sachar, E. J., J. W. Mason, H. S. Kolmer, and K. L. Artiss. Psychoendocrine aspects of acute schizophrenic reactions. Psychosom. Med. 25: 510-537, 1963.

23. Sachar, E. J. Corticosteroids in depressive illness. II. A longitudinal psychoendocrine study. Arch. Gen. Psychiat. 17: 554-567, 1967.

24. Sawrey, W. L. and J. D. Weisz. An experimental method of producing gastric ulcers. J. comp. physiol. Psychol. 49: 269, 1956.

25. Teitelbaum, P. and P. Derks. The effect of amphetamine on forced drinking in the rat. J. Comp. Physiol. Psychol. 51: 801-810, 1958.

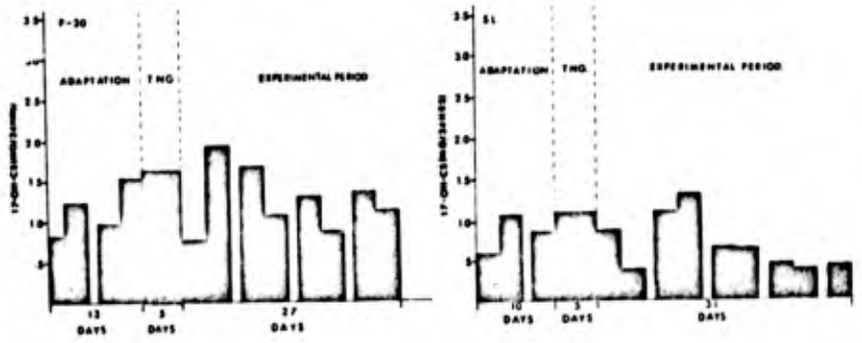


Fig. 1. Urinary 17-OHCS (mg/day) for one adapted S (P30) and one non-adapted S (SL) during booth adaptation, training (Tng) and experimental period.

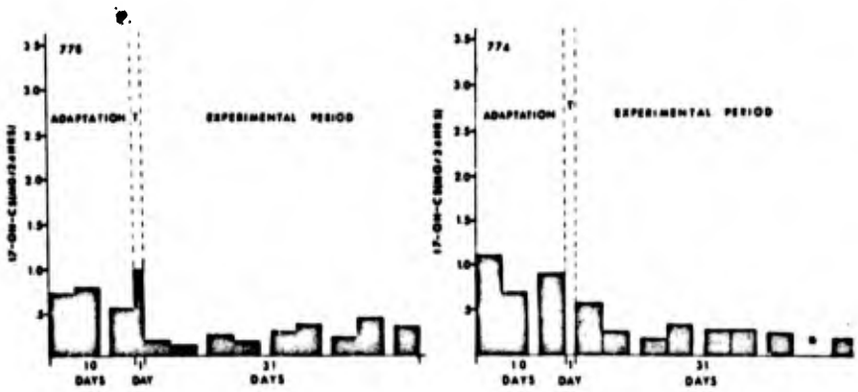


Fig. 2. Urinary 17-OHCS (mg/day) for a yoked experimental (778) and control (774) pair during adaptation, training (T) and experimental periods.

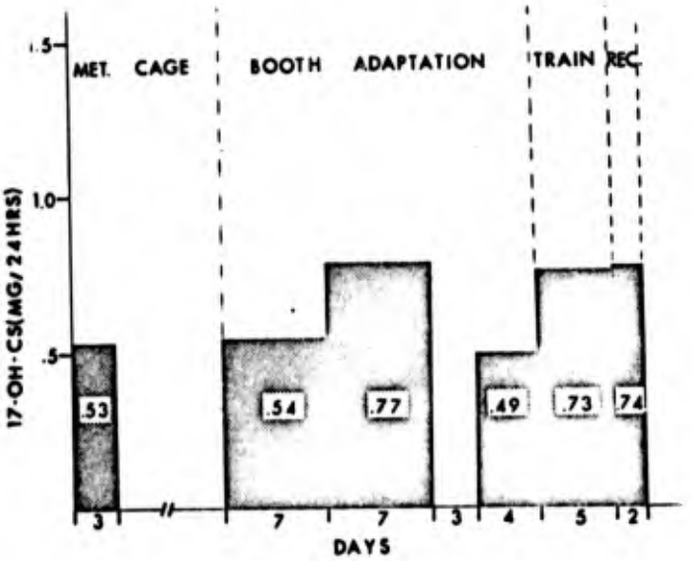


Fig. 3. Mean urinary 17-OHCS excretion (mg/day) for all S's (N=8) during metabolism cage adaptation (met. cage) booth adaptation, training and recovery (Rec.) periods.

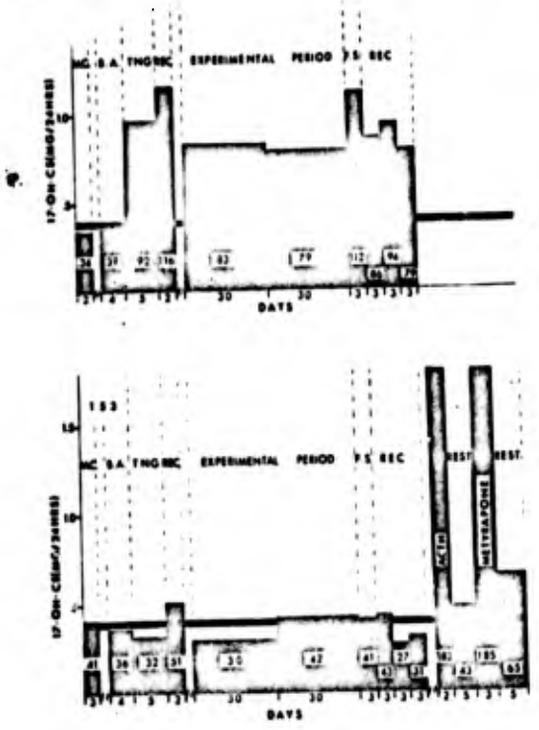


Fig. 4. Mean urinary 17-OHCS excretion in 2 S's preselected for their behavioral and physiological differences. Abbreviations: Metabolism Cage (M.C.). Booth Adaptation (B.A.). Training (TNG.). Free Shock (F.S.). Recovery Period (REC.).

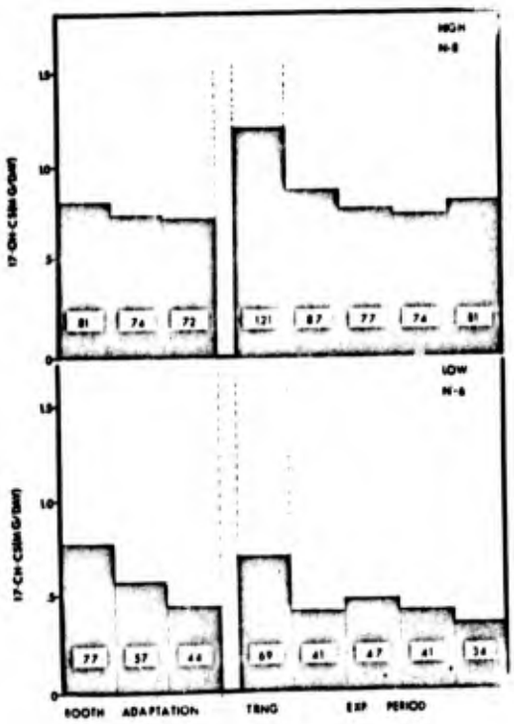


Fig. 5. Mean urinary 17-OHCS levels (mg/day) during booth adaptation, training, and experimental period for 8 non-adapting (high) and 6 adapting (low) subjects.

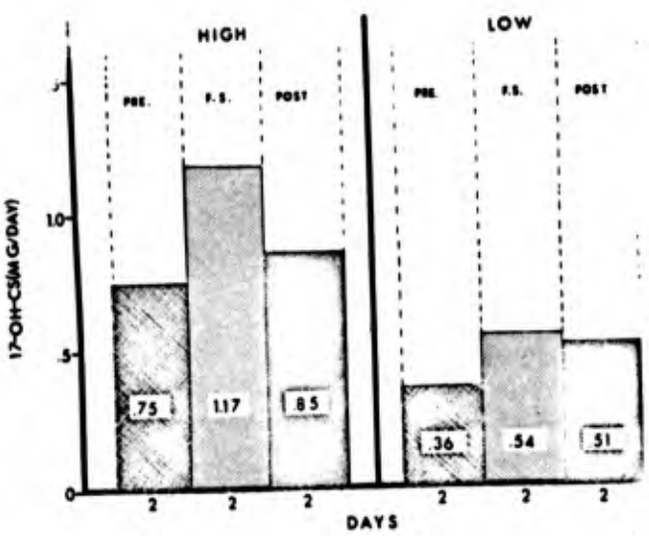


Fig. 6. Mean urinary 17-OHCS levels (mg/day) for high and low responders before, during, and after two days of free shock administration.

STUDIES RELATING TO THE USE OF PROPELLANT GAS

CHARLES J. LITZ, Jr.
U.S. ARMY FRANKFORD ARSENAL
PHILADELPHIA, PENNA.

INTRODUCTION

Although great strides have evolved in the design philosophy of Propellant Actuated Devices (PAD) for aircrew escape systems and other military applications, most, if not all, operate by the expansion of propellant gases either totally confined within closed ballistic cylinders (such as removers or thrusters) or semiconfined (such as rocket-assisted type catapults). These PAD are ideal for their intended function since they represent the best available compromise between personnel safety, aircraft performance, and existing technology.

A new use for PAD has been conceived - Boundary Layer Control (BLC). In fluid flow about a solid body, there is a very thin air layer attached to the body, called the "boundary layer" (BL), where friction plays an essential part. At the wall of the solid body, the fluid adheres to it and the friction forces retard the motion of the fluid. As one means of boundary layer control, a relatively high velocity jet of gases directed along the surface of an aerodynamic body can prevent separation of the boundary layer, reattach the laminar flow, and reduce the drag (1).* With the advent of the PAD Boundary Layer Control concept and the ejection of gas directly into the airstream, the thinking of aircrew survival expanded to that of saving the entire aircraft and crew intact, particularly at conditions of stall during take-off or landing.

These studies with the PAD/BLC concept indicated a means to reduce the skin friction drag of ballistic projectiles, and it

*See REFERENCES.

was also pointed out by the U.S. Air Force* that, as another possible use, the PAD/BLC technique could be applied to the reduction of the afterbody drag (boattailing and base) on high L/D reentry vehicles during the landing maneuver. Maintaining attached flow results in improved landing and an increase in control surface effectiveness.

As part of the "In-house Laboratory Independent Research Program" at Frankford Arsenal, a project was initiated wherein studies and tests were performed using the PAD/BLC system. The application was unique and heretofore untried, as evidenced by the issuance of several basic patents to the investigator. Chronologically, the first study related to the restoration of circulating airflow about stalled airfoils, to function, in effect, as an anti-stall device for fixed wing aircraft. The second study considered injecting propellant gas into atmospheric air boundary layers from flat plates and cones, for reduction of skin friction drag.

Continued studies and tests, with emphasis on skin friction drag reduction of projectiles, are underway at Frankford Arsenal. The following is a summary of the work accomplishments completed through December 1967.

SUMMARY OF STUDIES AND TESTS

1. Emergency Control of Boundary Layer on Aircraft Wings by Propellant Energy - "PAD Antistall"

Initial effort** considered the transient phenomena associated with the establishment of restoration of circulation flow by blowing gases from either the leading and/or trailing edges of airfoils (Figure 1). High speed motion pictures were used to study the rate of change of two-dimensional smoke flow patterns about two airfoils (NACA 23015 and 65006). The time required for the start of circulation build-up after blowing-jet initiation was found to be 0.1 second, and for circulation decay after blowing-jet cessation, 0.2 second.

The above results (of some 32 separate test runs) represent the first known record of actual measurement of the transient interval of time for circulation flow to break down and subsequently be restored on two-dimensional airfoil profiles. In fact, in one of the reports (Reference 2, written by Professor D. C. Hazen of Princeton University, where the tests were conducted) it was noted that ..... "The measurements have just scratched the

*"Comments on Technical Aspects of the Proposal," Atch to ltr, FDMM, AFDL (AFSC), to FA (SMUFA J5100), Subj: "Report P67-1-1, 'Emergency Control of Boundary Layer on Aircraft Wings by Propellant Energy - PAD Anti-Stall,'" 13 Oct 1967.

**Sponsored by U.S. Army Research Office-Durham, in 1963.

surface of an important and almost completely unknown segment of aerodynamics."

In April 1965, seven separate tests were conducted with a PAD antistall unit in a subsonic two-dimensional wind tunnel. The unit consisted of a high pressure plenum chamber (Figure 2) fitted to the leading edge of an airfoil (Figure 3). The free stream air velocity selected for each test run was in the range of take-off and landing speed of Army fixed wing aircraft; i.e., 40 to 100 miles per hour. The 16-inch cord, 2-foot span metal airfoil with leading edge blowing slots was positioned at a stalled angle of incidence of 18° within the wind tunnel and, after the wind in the tunnel was brought up to speed, the gas generator was initiated (Figure 4).

The main charge of 100 grams of composite type propellant burned for approximately 4.0 seconds (much like a cigarette) at a rate of about 1/16 inch per second at 1000 psi. The seven firings were convincing demonstration of proven feasibility of the basic concept, and reattachment of the air circulation flow about the airfoil was accomplished beyond any question. It was found that the solid propellant gas generator was capable of producing adequate mass flow and pressure characteristics in order to completely unstall the profile at an extremely high angle of attack.

Figure 5 presents the basic lift curve for the airfoil model and the results of runs 4, 5, and 7. At the 18° angle of attack, separation of airflow about the airfoil resulted and stall occurred. Runs 4, 5, and 7 achieved reattachment for the entire burning time of the gas generator. With the gas generator delivering gas to the airfoil, the lift coefficient was increased from 1.2 on the basic lift curve to 1.925 for run number 4. This increase of 0.725 in the lift coefficient represents an increase of 60.4 percent in lift of the airfoil model. Thus, the PAD/BLC not only reattached the airflow about the stalled airfoil model in the wind tunnel, it increased the lift substantially.

A system such as the PAD/BLC in an aircraft would make it possible for the pilot to regain control of the stalled aircraft with the reattached flow and, with the increased lift, make a more rapid ascent.

The local premature separation at 8° angle of attack, noted on the basic lift curve (Figure 5), was due to a discontinuity in the upper surface curvature of the airfoil model, which occurred during fabrication. If the chamber and slot were properly faired into the upper surface, a basic lift curve as indicated by the dotted line in Figure 5 would have resulted. In spite of this handicap, the gas generator did supply enough mass flow to reattach the airstream about the airfoil model.

X

The Smoke and Wind Tunnels located at the James Forrestal Research Center, Princeton University, were employed for these studies, under the direction of Professor D. C. Hazen and Mr. R. F. Lehnert.

Of the several factors considered in order to evaluate the feasibility of a PAD-operated BLC system, one was the range of conditions (or missions) from which the BLC operates. Data relative to the proposed BLC conditions, time of operation, and type of control established (Table I) were formulated after several service and commercial pilots were interviewed by personnel of Princeton University.

TABLE I. Results of BLC Survey

BLC Function (Condition)	Time of Operation (sec)	Type of Control
Emergency (landing)	30	Manual
Safety device (to permit recovery from a dropped wing)	5	Automatic
Short take-off or landing (STOL)	60 total (40 take-off 20 landing)	Manual

The times listed in Table I are within the range of current state-of-the-art gas generator design. For the emergency condition caused by either partial or total power failure of the thrusting engine, sufficient BLC power is supplied to the wing to allow the aircraft to land at reduced speeds, requiring less kinetic energy. Based on the results of a preliminary aerodynamic study (Reference 3), it was determined that a gas generator weighing 3.5 lb (2.1 lb of propellant) and delivering gas for 30 seconds would permit the safe landing of an aircraft (wing loading of 10 lb/ft^2) without engine power.

Thus, the advantage of the PAD antistall system is apparent for, in an emergency and safe landing requirement for BLC, this system does not require any of the aircraft engine power for its operation. The PAD unit is self-contained; it functions automatically; and it can be located in the wing of the aircraft to provide weight and space savings (Figure 6).

Other studies considered retrofitting the DeHavilland CV-2 Caribou aircraft with a simple lift BLC system, such as PAD-energized blowing flaps, to further improve its short take-off and

landing (STOL) characteristics (Figure 7). The findings indicated that the PAD/BLC system would occupy less volume and weigh less, by at least an order of magnitude (x 1/10) than the CV-2 motor drive air pumping system. However, due to the limitations of the CV-2 tail design, which could not trim out the expected induced pitching moments associated with the BLC gas-blowing system, the PAD/BLC was not considered feasible for this particular application.

Complete documentation of these PAD/BLC studies and tests is presented in a series of technical reports (2, 3, 4, and 5). In addition, an outline for additional tasks relating to three-dimensional model wind tunnel and full-scale free-flight tests of the PAD antistall unit was prepared (6). Such a program received favorable comments.* The following is a direct quote which assessed the feasibility of applying the concept to a U.S. Air Force flight system.

"The propellant actuated device (PAD) antistall unit has the apparent advantage of being a relatively lightweight, low-volume means for obtaining BLC airflow requirements independent of the main propulsion system. This is an important consideration in the sudden loss of or reduction in propulsive system thrust during take-off. As indicated, the rate at which the PAD system could be initiated (and high-lift wing flow conditions could be achieved and maintained) is of great importance. The proposed effort would include an experimental investigation of the dynamic characteristics of the PAD system during Phase III of the proposed study."

At this time, current funding levels could not support the suggested effort as outlined in Reference 6; however, in the event of budgetary authorization, the proposed tests will be considered.

2. PAD Two-phase Boundary Layer and its Drag Reduction Characteristics

This study considers coexisting propellant gas and atmospheric air boundary layers which occur when propellant gas is injected into the free stream air flow at the surface of a flat plate and cone. Two different and diverse methods of solution were employed to obtain mathematical solutions.

---

*a. "Comments on Technical Aspects of the Proposal," Atch to ltr, FDMM, AFDL (AFSC), to FA (SMUFA J5100), Subj: "Report P67-1-1, 'Emergency Control of Boundary Layer on Aircraft Wings by Propellant Energy - PAD Anti-Stall,'" 13 Oct 1967.
b. Ltr, Hq, USAMC (AMCSA-E) to FA (SMUFA J5100), Subj as above, 29 Aug 1967.
c. Ltr, USARO-Durham to FA (SMUFA J5100), Subj as above, 30 Aug 1967.

The results of the first method indicated reductions in drag on a 1-1/2 by 2 ft flat plate at zero incidence in a free stream airflow of 100 ft/sec would be in the order of 33 percent or more when propellant gas is injected at the surface of the plate with a velocity of 1 ft/sec.

In the second study (or method) of solution, the overall frictional drag value was calculated to be reduced by 84 percent for a 20° solid cone with a wetted area of 3 ft^2, in a free stream air velocity of 1000 ft/sec. The amount of propellant (2.65 lb) required for gas discharge for a typical mission was calculated to occupy approximately 6.7 percent of the internal cone volume. The propellant considered was a double base type which burns linearly, much like a cigarette.

One means of illustrating the amount of decrease in the drag where the boundary layer is kept laminar by propellant gas discharge is illustrated in Figure 8, which is a plot of the skin friction resistance formula.

From Figure 8 it is possible to deduce the variation of the coefficient of skin friction for the flat plate under the conditions of optimum blowing where it is noted that the distance between the curve marked "laminar" and that marked "turbulent" corresponds to the saving in drag effected by the application of PAD-gas blowing. The relative saving in drag calculated with respect to turbulent drag increases somewhat as the Reynolds number is increased. It varies from 80 to 94 percent in the range of Reynolds numbers 10^6 to 10^8.

For the cone, the duration of gas discharge was determined as that interval of time required for the projectile to travel distances of from 1000 to 6000 feet with a span of velocities of 500, 1000, and 1500 feet per second. In this analysis, the influence of Mach number and the associated compressible turbulent boundary layer was considered since it does, indeed, have an effect on the skin friction coefficient and, hence, the drag.

Consideration was also given to the effects of the annular grooved gas discharge slot and the boundary layer of the rotating cone on the aerodynamic stability and drag characteristics. Also, as noted in Refernce 7, the data obtained from the firings with a 20 mm cone-cylinder projectile indicated no discernible effect of grooves on the drag coefficient or on other aerodynamic coefficients.

With respect to rotation, the Coriolis forces in the boundary layer give rise to an additional acceleration in the direction of flow for rotating bodies in an axial air stream, and the added acceleration has the same effect as a favorable pressure

gradient (Reference 1). In addition and to a lesser extent, the centrifugal forces acting in the boundary layer carried with the body exert a beneficial influence by delaying separation. However, actual testing, soon to be initiated, will make the final determination of the effects, if any, of the groove and Coriolis forces on the drag and stability of the cone.

A complete summary of the above studies is documented in two technical reports - References 8 and 9.

Based on the promising results of the mathematical study, the 20° cone with a "wetted" area of three square feet and a circular gas discharge slot 25 percent aft of the apex was designed (Figure 9). The cone was designed with a peripheral gas port opening of 0.010 + 0.001 inch and houses the propellant charge internally. Also, the gas can be discharged either perpendicularly or at an angle of 45° to the centroidal axis by inserting different parts in the basic assembly. Fabrication of the cone is scheduled for completion by the third quarter FY 68, with laboratory test and evaluation to follow.

SUMMARY

In Summary, the research studies presented in this paper represent a unique and heretofore untried manner for the use of propellant gas to prevent aircraft stalling.

Results of the wind tunnel tests were convincing demonstrations of proven feasibility of the basic concept in reattaching the air circulation flow about the airfoil (5).

Considering the mathematical studies of gas discharge from flat plates and 20° cones, substantial skin friction drag reductions in the order of 80 to 86 percent are indicated (6).

REFERENCES

1. H. Schlichting, Boundary Layer Theory; New York, N. Y.: McGraw Hill Book Company, 1960.

2. D. C. Hazen and F. C. Karins, "A Preliminary Study of the Rate of Growth and Decay of Blowing-Jet-Induced Circulation on Two-dimensional Profiles," Princeton University Report No. 632, December 1962.

3. C. J. Litz, Jr., "Emergency Control of Boundary Layer on Aircraft Wings by Propellant Energy - PAD Antistall. Preliminary Aerodynamic Study," Frankford Arsenal Report M66-13-3, March 1966.

4. C. J. Litz, Jr., and A. E. Larsen, "Emergency Control of Boundary Layer on Aircraft Wings by Propellant Energy," Frankford Arsenal Report R-1675, April 1963.

5. R. F. Lehnert, "A Preliminary Study of a Solid Propellant Gas Generator Supplying a Leading Edge Blowing Slot for Flow Reattachment," Princeton University Report No. 758 (Frankford Arsenal Report M66-13-1), October 1965.

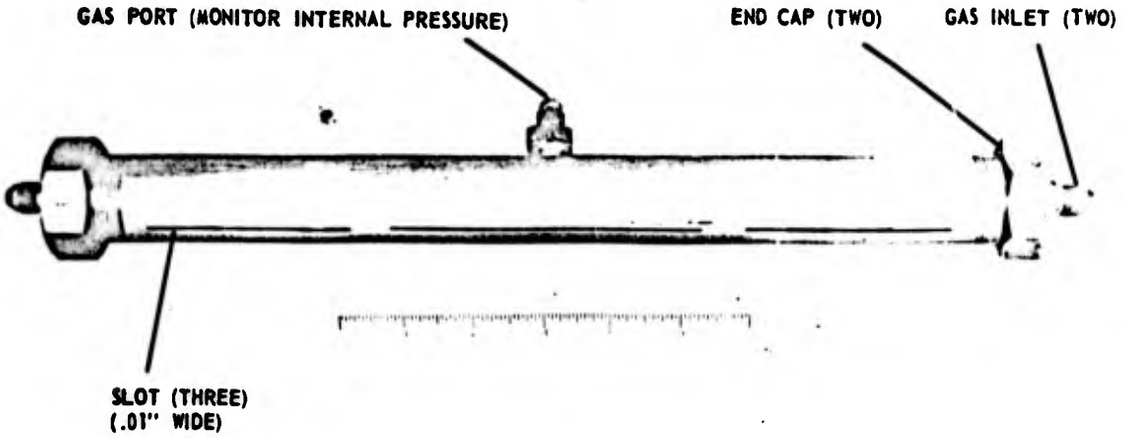
6. C. J. Litz, Jr., and A. E. Larsen, "Proposal for Continued Study of Emergency Control of Boundary Layer on Aircraft Wings by Propellant Energy - PAD Antistall," Frankford Arsenal Proposal P67-1-1, March 1967.

7. E. R. Dickenson, "The Effects of Annular Rings and Grooves, and of Body Undercuts on the Aerodynamic Properties of a Cone-Cylinder Projectile at M = 1.72," Ballistics Research Laboratory Report BRL-1284, June 1960.

8. C. J. Litz, Jr., "Emergency Control of Boundary Layer on Aircraft Wings by Propellant Energy - PAD Antistall. PAD Two-phase Boundary Layer and its Drag Reduction Characteristics," Frankford Arsenal Report M66-13-2, January 1966.

9. C. J. Litz, Jr., "Emergency Control of Boundary Layer on Aircraft Wings by Propellant Energy - PAD Antistall. PAD Two-phase Boundary Layer and its Drag Reduction Characteristics (Part II)," Frankford Arsenal Report M66-13-4, March 1967.

Figure 1. Emergency Control of Boundary Layer on Airfoil Wing by Propellant Energy



SLOT AREA = AJ = (.01 X 10.5) - 0.105 SQ. IN.

Figure 2. High Pressure Plenum Chamber

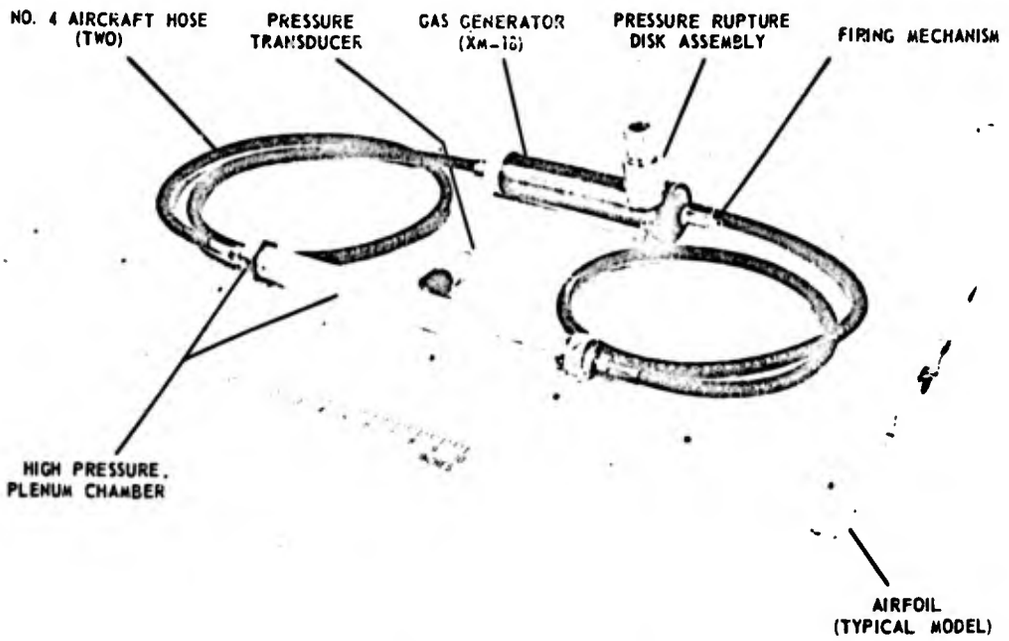


Figure 3. PAD Antistall System

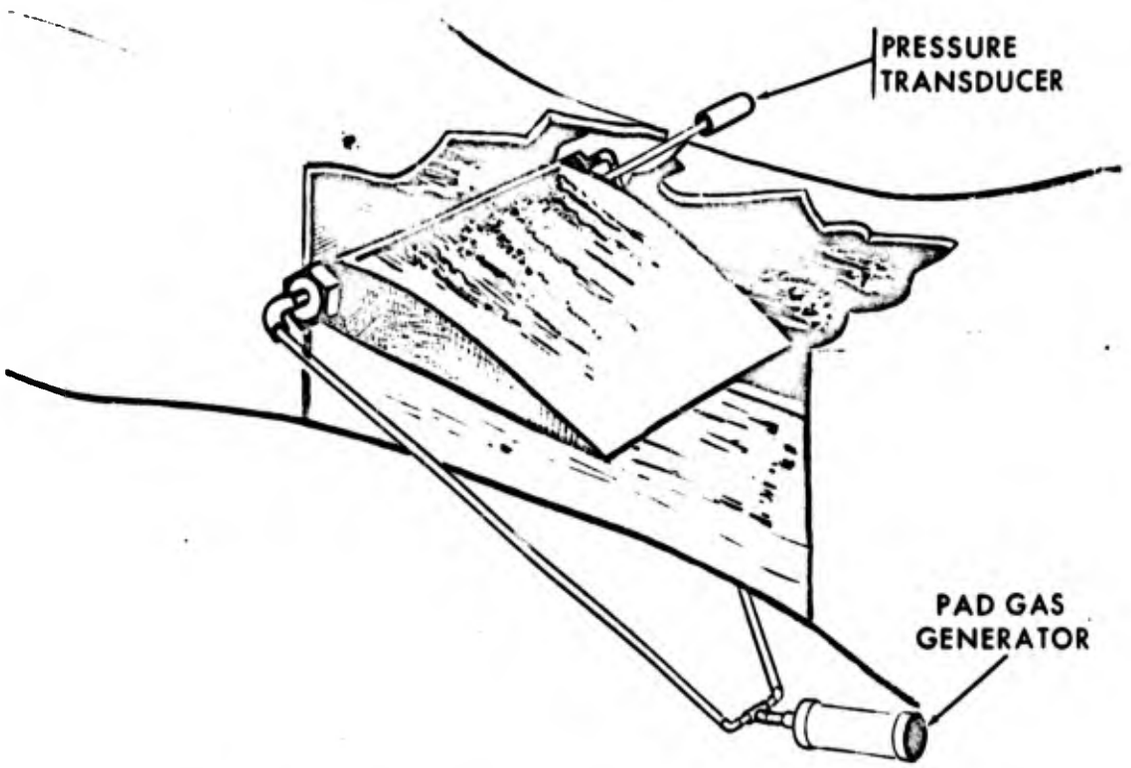


Figure 4. Schematic of Wind Tunnel Test of PAD Antistall Unit

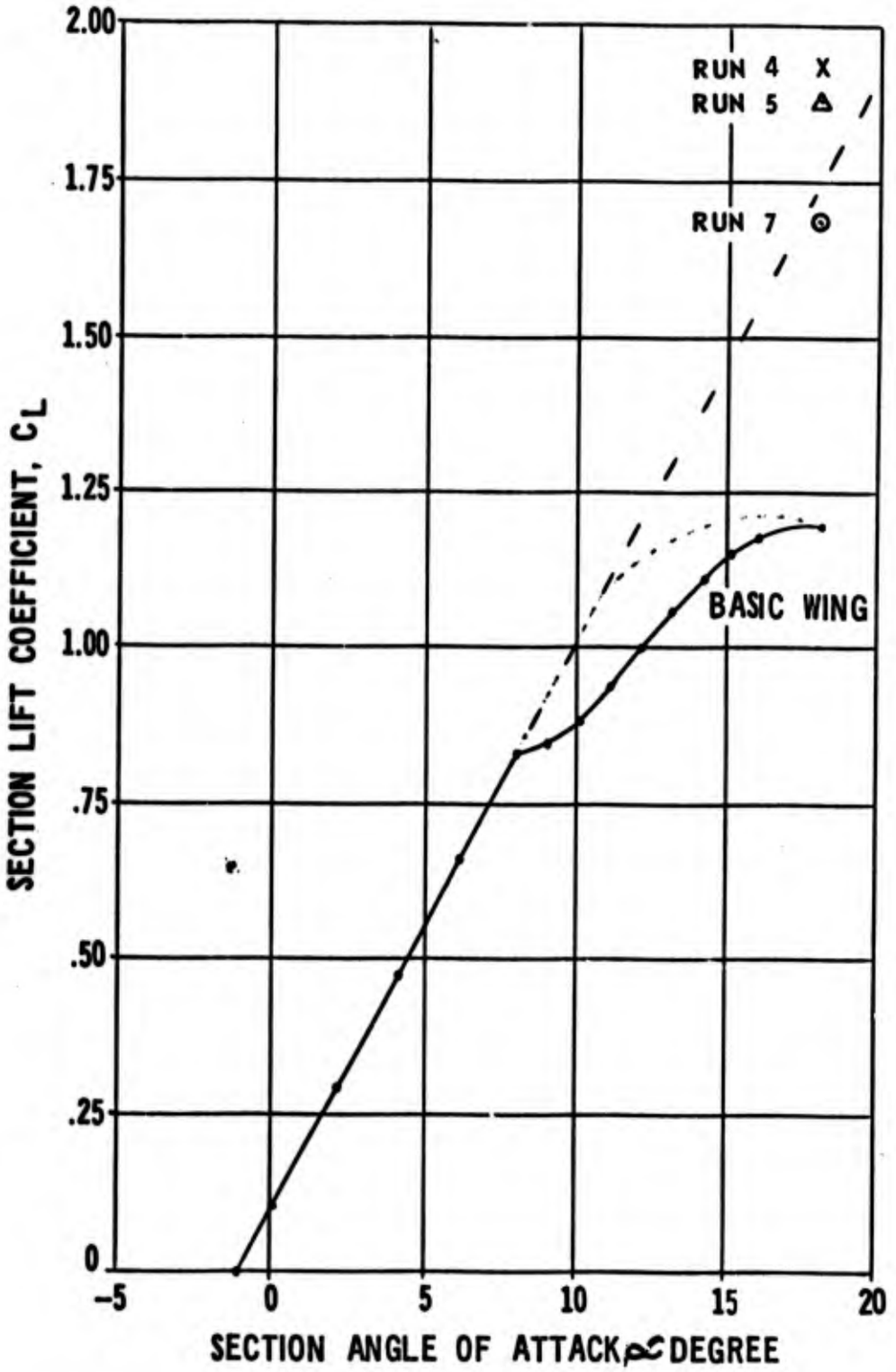


Figure 5. Characteristic of the 18 inch Chord Metal Airfoil at High Reynolds Number - PAD Antistall

[LEADING EDGE BLOWING SHOWN]

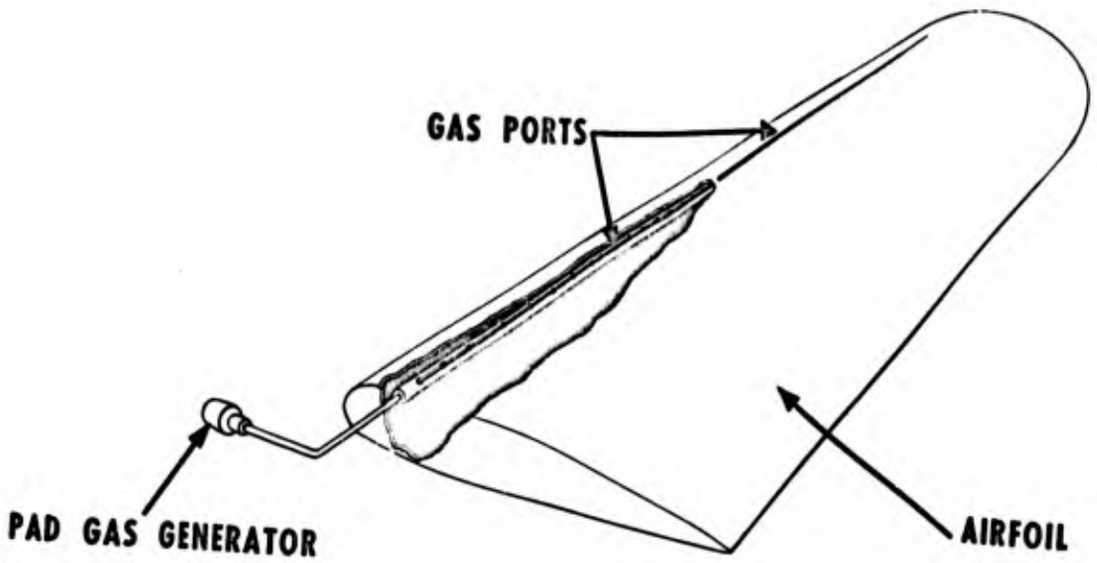


Figure 6. Airfoil with PAD Antistall Unit

[STALLED CONDITION ARISES AT TAKE OFF OR LANDING]

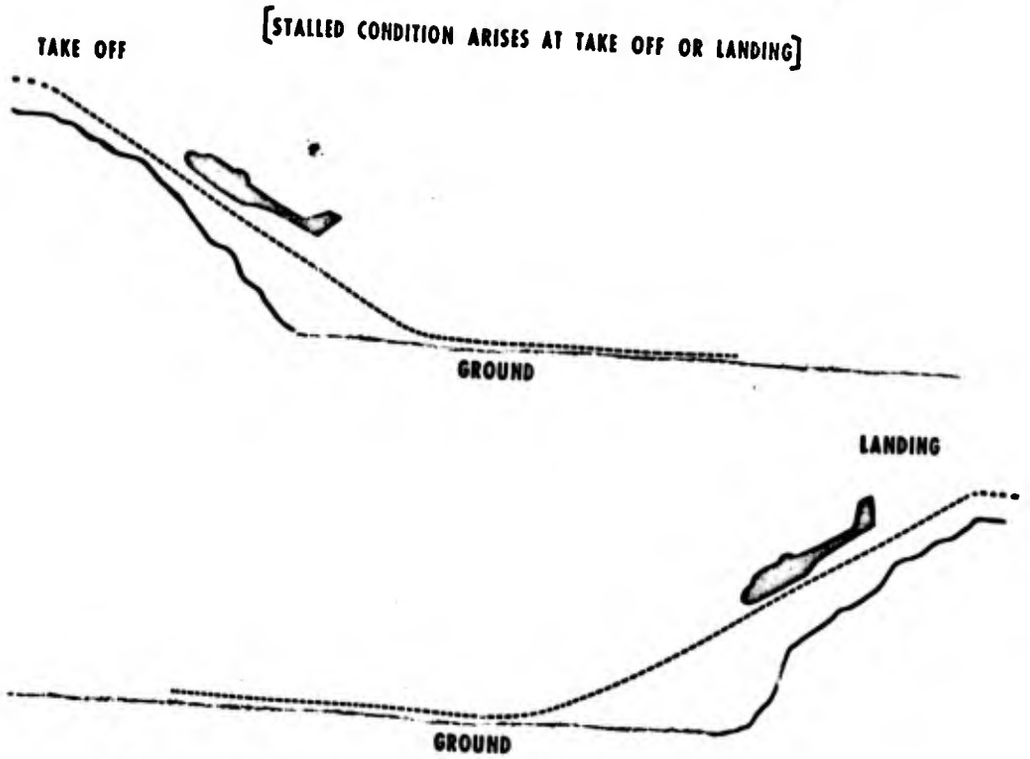
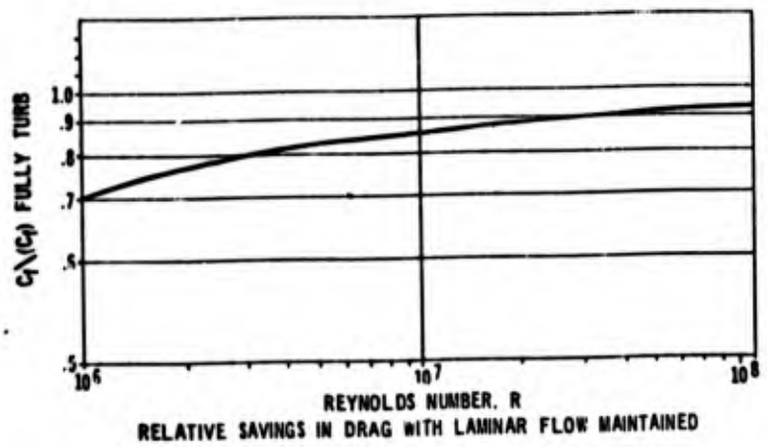


Figure 7. PAD Antistall Applied to the Caribou



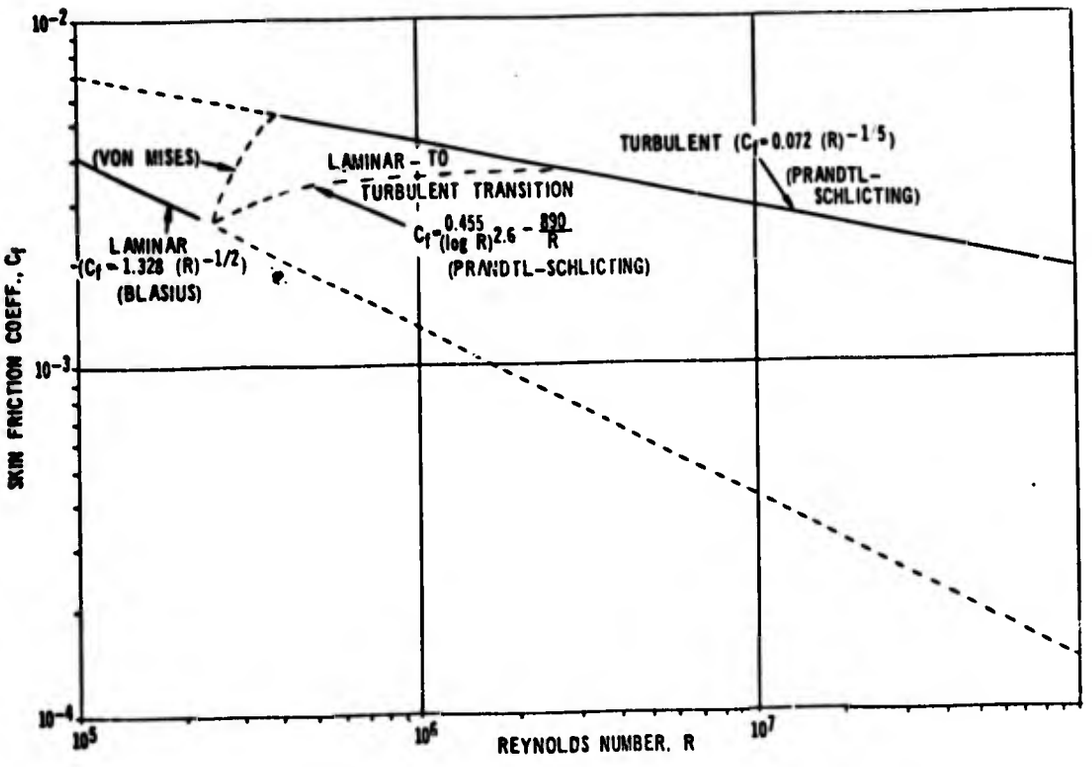


Figure 8. Coefficient of Skin Friction for a Flat Plate at Zero Incidence

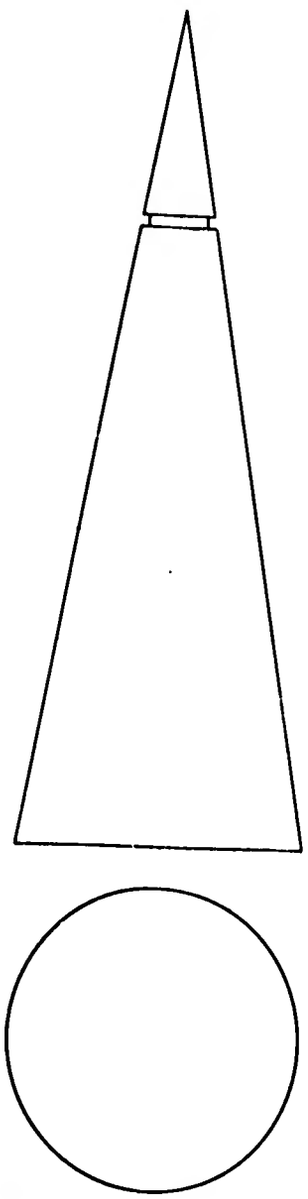


Figure 9. Design Layout of 20° PAD-Cone (external view)

THE BALLISTIC AND MECHANICAL PROPERTIES OF POLYMERS

DAVID M. MARTIN, CAPTAIN, CM
ROBERT W. LEWIS, 1ST LIEUTENANT, CM
GEORGE R. THOMAS
U. S. ARMY NATICK LABORATORIES
NATICK, MASSACHUSETTS

I. INTRODUCTION

Transparent armor for applications such as vision blocks, wind screens, and face shields rely on polymers either completely or as a part of a ceramic-plastic composite. Transparent polymeric armor systems are not currently as efficient on a weight basis as opaque armor systems. This work was undertaken to find ways to increase the efficiency of polymeric armor systems.

Orientation and processing variations are known to have a large effect on the static mechanical properties of polymers and spurious data have indicated increases in ballistic resistance for hot stretched homogeneous armor. The influence of hot stretching and other processing variables on ballistic resistance were studied. The effect of orientation on fracture surface energy was studied in some detail to provide insight into the fundamental processes involved in polymer failure.

It would be highly desirable to be able to correlate ballistic behavior with the material and physical properties of the polymeric system to enable rapid predictions of ballistic resistance as well as to lead to improved transparent armor. Up to the present time no such correlation has been known. It was one of the purposes of this study to determine what properties, if any, would consistently be indicative of the ballistic strength of the polymer.

II. INFLUENCE OF PROCESSING VARIABLES

A study of the effects of orientation, thermal treatment, and water content was carried out to establish the effects of processing variations on the static and ballistic behavior of polymers. The effects of orientation and gammairradiation induced crosslinking on the fracture surface energy calculated by direct measurement of energy input to a growing crack was studied. Dramatic changes in the static rate mechanical properties occurred (see Table I) but the magnitude of those effects was not carried over to the ballistic resistance.

Hot stretching of a wide variety of polymers was attempted. Various polymers were hot stretched at temperatures in excess of the glass transition. The ease of hot stretching was found to vary widely both because of chemical decomposition (e.g., polysulfone) and necking behavior (e.g., polypropylene). Polycarbonate, poly (methyl methacrylate), XT-500, polystyrene, and polyphenylene oxide were successfully hot stretched.

Samples of the hot stretched polymers were tested for ballistic and mechanical properties at the U. S. Army Natick Laboratories. The existence and extent of orientation of the hot stretched polymers were established by means of optical birefringence. Fracture surface energy was determined by direct measure of energy input to a stable fracture in a triangular cross section bending sample(1). The fracture energy for biaxially oriented polystyrene as a function of degree of hot stretching is shown in Figure 1. Figure 2 shows the results of gamma irradiation on the fracture energy of biaxially oriented polystyrene. The large anisotropy induced by orientation is apparent in Figure 1. Inhomogeneous orientation can degrade the ballistic performance of polymers by introducing planes of low fracture surface energy which leads to splitting and lath-like failure of the plate. Homogeneously oriented polymers did not show any change in ballistic resistance compared to the non-oriented state when tested with a .22 cal fragment simulator.

Stretched polymers do perform better when utilized in composite armor as back up plates for ceramic materials. The reason for this difference in behavior is most likely due to the increase in toughness in the plane of the plate exhibited by oriented polymers In the case of direct interaction with a missile, the armor plate experiences large shear stresses under the missile, while for a back up material the stresses are largely tensile acting in the plane of the plate.

When stretched polycarbonate is used as a back up plate in a glass-polymer transparent armor system, weight savings on the order of 15 percent can be achieved. Further weight reductions are possible through optimization of design of composite armor systems.

Polycarbonate samples were quenched from above the glass transition, annealed below the glass transition, and held at 90°C in both very low and very high humidity atmospheres. None of the treatments gave rise to detectable changes in the ballistic properties of polycarbonate, when tested with a .22 cal fragment simulating projectile. Oriented poly(methyl methacrylate) samples were water saturated and dried in controlled temperature and humidity conditions. No changes in the ballistic resistance was observed even though large changes in the static rate properties have been reported by Ender(3) for similar treatments. The insensitivity of the ballistic resistance to the treatments described is somewhat surprising in light of the large changes which occur in static properties. The insensitivity to mechanical state also implies an attendant large dependence of the ballistic resistance on the fundamental building blocks of the polymer rather than the details of their arrangement.

III. TRANSITION TEMPERATURE CORRELATION

An extensive search for a correlation between material properties and ballistic resistance was conducted in order to provide direction in the development of new materials for armor use. In the course of this search, several suspected pertinent properties were investigated. Table II shows a sample of these variables. No real correlation was found to exist between the ballistic properties and the properties shown, as with all other commonly observed properties.

Several investigators, notably Nielsen(4), Matsuoka(5,6), and Boyer(7), have hinted that in order for a polymeric material to have good impact strength, it must have at least two transitions, one occurring above ambient temperature (the glass transition temperature, Tg) and one below ambient temperature (a "brittleness" temperature). Figure 3 and Table III give the results obtained when such a correlation is attempted in terms of the relative difference between the upper and lower transitions.

As can be seen from the graph, the correlation is remarkably good. It shows that as the spread between the two transitions increases, the better the polymer performs ballistically. For the first time a property is known to exist which can predict ballistic performance.

A low temperature transition in a glassy amorphous polymer is not in itself indicative of good impact behavior. It appears that this transition must be specific in its source; e.g., preliminary data show that the low temperature transition must be caused by motion in the main chain of the polymer, not by side group movement. This restriction is not necessarily adverse. It can be quite useful from a molecular standpoint, resolving hitherto controversial transitions into main chain or side chain motions.

The most exciting aspect of the transition temperature correlation is that it obviously depends on molecular parameters such as chain stiffness. That thermal and mechanical behavior for a polymer should be so intimately connected opens new vistas of investigation. Parameters such as chain stiffness and glass temperatures, calculable from first principles, have ultimate influence on the behavior of bulk polymer systems. Screening of candidate polymer armors by simple thermal measurements as well as prediction of the effects of molecular changes on the ballistic resistance should now follow. The correlation should allow for the synthesis of new polymers designed specifically for high impact resistance.

IV. CONCLUSIONS

We have studied the effects of hot stretching induced orientation on fracture energy and found large changes both perpendicular and parallel to the orientation directions. Orientation does not improve the ballistic resistance of homogeneous armor but can increase the efficiency of composite armor. Processing variables such as thermal history and relative humidity at working temperatures were shown to have little effect on the ballistic resistance of polymers, a fact which is important in design considerations as well as from a theoretical standpoint.

The establishment of the relationship between thermal and ballistic properties is of great importance both from the point of view of the armorer in the selection of materials and for the polymer physicist and synthetic polymer chemist in their understanding of the processes that occur in polymers as well as in the molecular design of materials to accomplish specific tasks. The backbone composition of the polymer molecule was demonstrated to be of overwhelming importance in comparison to the more delicate configurational changes induced by processing of the bulk polymer.

BIBLIOGRAPHY

1. Nakayama, J., "Direct Measurement of Fracture Energies of Brittle Heterogeneous Materials," J. Am. Ceramic Soc., 48, No. 11, p. 583, November 1965.

2. Hsiao, C. C., "Effects of Atomic Irradiation and of Molecular Orientation on the Mechanical Behavior of Polymers," Report No. TID-21874, Atomic Energy Commission, 1965.

3. Ender, D. H., "Study of Yielding Phenomena," Technical Report 68-7-CM, U. S. Army Natick Laboratories, July 1967.

4. Nielson, L. E., Mechanical Properties of Polymers, Reinhold Publishing Company, New York City, New York, 1962.

5. Matsuoka, S. and Ishida, Y., "Multiple Transitions in Polycarbonate," J. Polymer Sci., C-14, 247, 1966.

6. Matsuoka, S., Aloisio, C. J. and Daane, J. H., "Some Aspects of Brittle Failure in Polymers," Report, Bell Telephone Laboratories, Inc., Murray Hill, New Jersey.

7. Boyer, R. F., "Multiple Transitions and Relaxations of Polymers in Relation to Mechanical Properties," Conference on Polymer Structure and Mechanical Properties, U. S. Army Natick Laboratories, April 1967.

TABLE I

Effect of Uniaxial Elongation on Physical Properties of Polymers

Polymer	Relative Fracture Strength	Relative Extension Modulus
Polystyrene(2) parallel to orientation	1.2	1.2
(100% elongation) perpendicular to orientation	0.5	0.5
Poly(Methyl Methacrylate)(2) parallel to orientation (100% elongation)	1.25	-
Polyethylene tetraphthalate(3) parallel to orientation	-	17.8
(High elongation) perpendicular to orientation	-	0.67
Nylon 66(3) parallel to orientation	-	19.2
(High elongation) perpendicular to orientation	-	0.67
Polyacrylonitrile(3) parallel to orientation	-	5.8
(High elongation) perpendicular to orientation	-	0.67

TABLE II

Ballistic Performance vs Material Properties

Polymer	V_{50} (ft/sec)	Izod Impact (ft #/in. of notch)	Flex Modulus (psi x 10^5)
PPO	1250	1.5 - 1.9	10.5
Polysulfone	1130	1.3	3.8
Noryl	1120	1.3	3.6
Polycarbonate	980	12.0 - 17.5	3.4
Nylon (6-6)	1000	1.0 - 2.0	1.8 - 4.1
Polyurethane	930	No Break	3.2 - 3.5
Nylon 6	930	1.0 - 5.5	1.4 - 4.0
PMMA	600-900	0.4 - 0.5	3.9 - 4.7
PCTFE	680	2.5 - 2.7	---
Polystyrene	580	.25 - .40	---
Teflon	470	3.0	---

TABLE III

V_{50} Ballistic Limits of Polymers
(17 Grain Fragment Simulator 0° Obliquity 30 oz/sq ft Areal Density)
and Corresponding Transition Temperatures

Polymer	V_{50} (ft/sec)	T_b(1) °K	T_{HD}(2) °K	T_b/T_{HD}
PPO(3)	1250	103(4)	464(4)	.22
Cycoloy 800(5)	1130	—	394(4)	—
Polysulfone	1130	172(4)	447(4)	.38
Noryl(6)	1120	127-130(7)	403(4)	.32
Surlyn A(8)	1030	166(9)	311(4)	.53(9)
Polycarbonate	950-1010	172(4)	410(4)	.42
XT-500(11)	1000	188(12)	359(12)	.52
Nylon (6-6)	990-1010	148(10)	330(10)	.45
Polypropylene	940	178(10)	275(10)	.64
Polyurethane	930	140(10)	310(10)	.45
Nylon 6	930	148(10)	330(10)	.45
Celcon,Delrin(13)	820-880	233(4)	383(4)	.61

TABLE III (cont'd)

Polymer	V_{50} (ft/sec)	T_b(1) °K	T_{HD}(2) °K	T_b/T_{HD}
ABS(14)	850	—	372(4)	—
PMMA(15)	600-900(16)	313(14)	375(4)	.83
CTFE(Kel-F)(17)	680	248(10)	320(10)	.78
Polystyrene	580	333(4)	372(4)	.90
TFE(Teflon)(18)	460-480	---(19,10)	150-180(19,10)	1.0(19)

Notes on Table:

(1) - Brittleness Temperature
(2) - Heat Distortion Temperature at 264 psi fiber stress
(3) - Polyphenylene oxide
(4) - Values obtained from 1968 Edition of Modern Plastics Encyclopedia, McGraw-Hill Publishing Co., New York City, New York
(5) - Marbon's combination of ABS and polycarbonate
(6) - Modified PPO(3)
(7) - Value obtained from direct communication with G.E.
(8) - DuPont's Ionomer
(9) - T_b reported as less than 166°K
(10) - These values obtained from Wada, Y., "Mechanical Dispersions and Transition Phenomena in Semi-crystalline Polymers", J. Phys. Soc. Japan, 16, 6, June 1961 (1226). The T_b is here a low temperature amorphous transition. The T_{HD} is really T_g, but the values are very nearly the same.

TABLE III (cont'd)

(11) - American Cyanamid combination of PMMA and a rubbery phase
(12) - Obtained from dilatometric data obtained from direct communication with American Cyanamid
(13) - Polyoxymethylene
(14) - Poly(acrylonitrile-butadiene-styrene)
(15) - Polymethylmethacrylate
(16) - Wide variation in ballistic data believed to be due to processing and fabrication variables
(17) - Polychlorotrifluoroethylene
(18) - Polytetrafluoroethylene
(19) - Values obtained from Wada did not show any amorphous transition other than the T_g. Therefore, the ratio becomes $T_g/T_g = 1$.

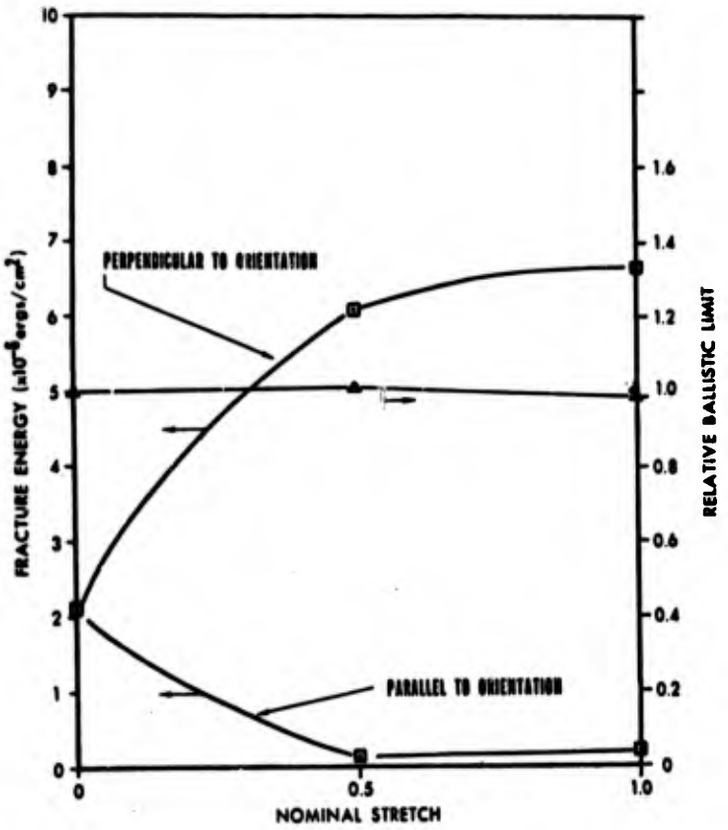


Figure 1. The fracture energy and relative ballistic limit of biaxially oriented polystyrene

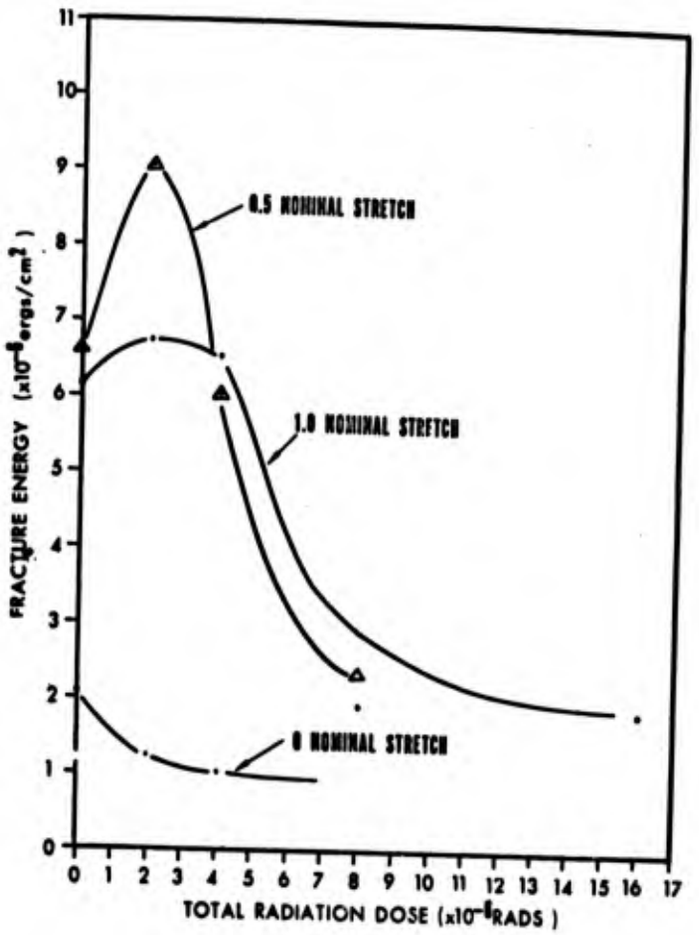


Figure 2. The fracture energy of biaxially oriented polystyrene in the orientation direction as a function of gamma irradiation dose

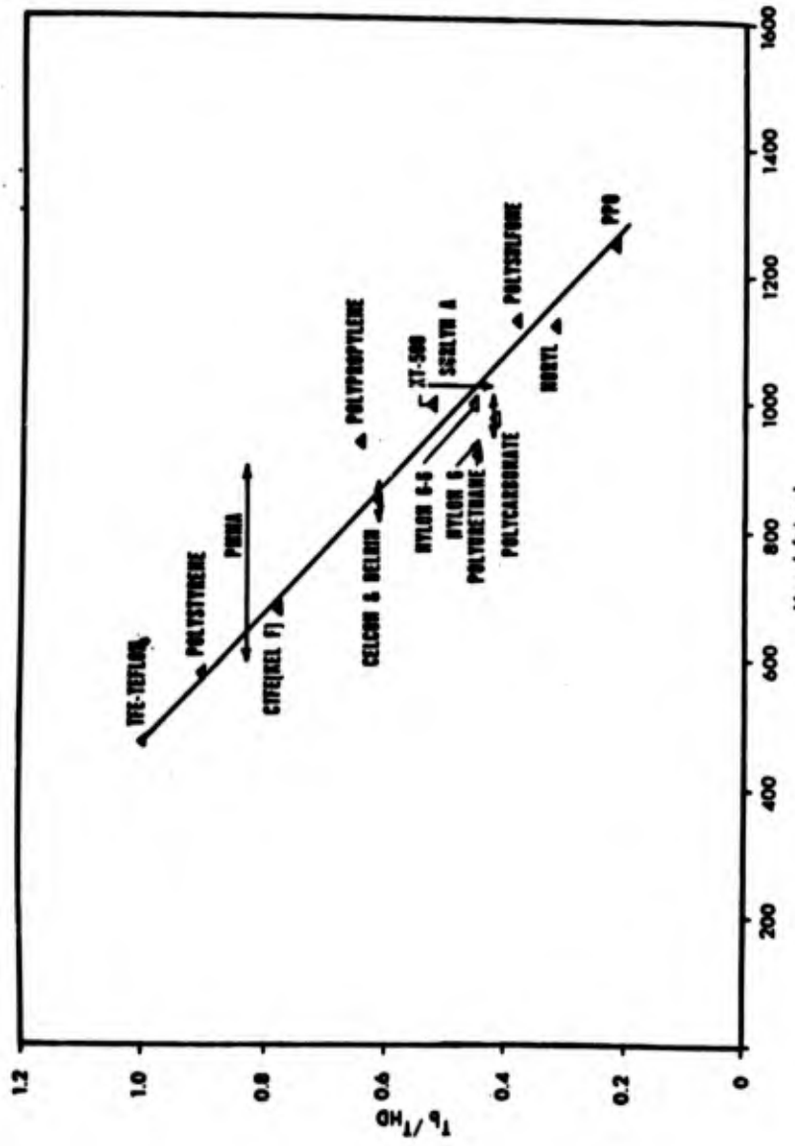


Figure 3. The relationship of the ratio of brittleness and heat distortion temperatures to the ballistic limit at constant areal density (30 oz per square foot of various polymers)

HILIS - A HIGH INTENSITY LIGHT SYSTEM FOR ALGAE FOOD PRODUCTION

ROBERT O. MATTHERN
U. S. ARMY NATICK LABORATORIES
NATICK, MASSACHUSETTS

In Japan today algae, grown in coastal waters much as grass is grown on land, sells as garnish for eight dollars a pound (dry weight) (1). The dense civilization of the Aztec Indians at the time of the Spanish conquest could probably not have existed except for the utilization of a bluegreen algae as food (2). From these two instances alone, it can be seen that the vanguard of food research lies not in the use of algae as a modern food but in the application of modern technology to finding practical economical methods for producing it. The algae which have been used as food are either filamentous or leaf-like in form. Such forms are easily harvested by manual methods but are not well suited to the unit operations of modern mass fermentation. As a consequence, the engineer and biologist have turned to the production of unicellular strains which are small rugged individuals (3 to 20 microns in diameter or maximum dimension) and have high growth rates. The Chlorella and Scenedesmus strains have been the most widely studied.

The selection of these strains required extensive study for wholesomeness. Results of these studies indicate that these strains with their natural symbiotic array of microflora could safely furnish 25% of a total caloric intake for man (3). As progress was made in strain selection, mass culture, nutritional evaluation, processing and food preparation, four problems became apparent: acceptability, nutritional quality, large culture volume requirements, and poor energy conversion efficiencies. With hundreds of thousands of strains to select from, with application of genetic technology and with processing methods already developed for flavor improvement (4), acceptability and nutritional quality offer no real obstacles. The problems of large culture volume requirements and poor energy conversion efficiencies are much more complex and formidable.

Mass culture under natural sunlight rarely yields concentrations as high as 1.5 g/L dry weight under purely photosynthetic

processes (5). The reasons for this are that sunlight is not continuous for 24 hours, intensity does not exceed 10,000 lumen and only a small portion of the radiant energy can be efficiently utilized by the photosynthetic processes. In our laboratory, batch systems utilizing fluorescent illumination at 2000 lumen produced 0.5 g/L/day. Early attempts to utilize high intensity illumination in excess of 10,000 lumen failed due to lack of understanding of the importance of light-dark phasing, nutrient requirements and heat transfer. Unfortunately, many erroneously concluded that the problem was one of a maximum permissible light intensity between 1000 and 2000 lumen. The successful operation of a four lamp (Quartzline) 2.7-liter continuous culture system (6, 7) demonstrated that intensities up to 52,000 lumen were feasible and that the dense culture of 9 g/L under continuous culture reduced the culture volume to 37.3 L/man as compared to 1000 L/man for the fluorescent batch system (based on a total food requirement of 500 g/day/man). Unfortunately, the geometric configuration of this early high intensity light system, though a breakthrough, would permit neither variation of growth parameters nor further increase in light intensity due to a problem of heat transfer. The same limitations apply to the system developed by P. J. Hannon and C. Patouillet (8) which is essentially a scale up of the four lamp system to six lamps. Comparisons of other systems are discussed by R. L. Miller and C. H. Ward (9).

Approach: A definitive study of the effects of variations in each of the environmental growth parameters was considered as serving several useful purposes: (1) reduce the culture volume, (2) improve the energy conversion efficiencies, (3) determine what, if any, effect a variation in the growth parameters might have upon quality and flavor of the food produced, and (4) describe the environmental growth parameters and their control thereby affording rational design criteria for scale up. The growth parameters include temperature, light intensity, light quality, light-dark phasing, gas exchange, heat transfer, availability of nutrients, pH, and maximum and minimum concentrations of specific nutrients. If these growth parameters could be controlled in one system then their interrelationships could also be determined. The high intensity light system (HILIS) was designed to do this. A description of the design, operation and function of the components to meet the study requirements will follow and the results of studies of the first three parameters presented. Results demonstrate a reduction in culture volume to 21.6 L/man and improved efficiency of 267% over the 4 lamp system.

DESCRIPTION OF HILIS

The HILIS will be described following the classical unit operations, processes and flow model concepts of chemical engineering indicating how these concepts permit a definition of the environmental parameters. Where applicable pertinent design considerations will also be discussed under each of these headings.

Figure 1 shows a schematic diagram of the principal functional components of the HILIS. Figure 2 is a photograph of the system less the heat exchangers and powerstat control to the lamps.

Illumination: Illumination is provided by six 1500 watt Quartzline lamps. Each lamp is jacketed with both a 22 mm and a 38 mm pyrex tube. Between these tubes chilled distilled water is circulated to absorb heat and to filter out ultraviolet and infrared light. Of the total energy input, 12% is emitted as visible light and 88% as energy at the ultraviolet and infrared wavelengths. The visible light is low in the blue and rich in the red wavelengths. For these lamps the ratio of blue at 440 nm to red at 680 nm wavelengths (wavelengths at which chlorophyll exhibits maximimum absorption) is 1 to 4.5. The use of filters was considered as a means of preventing the infrared light from reaching the culture and of striking a better balance between the blue and red spectrum, but was not found necessary. Operational procedures for maintaining temperature control and heat transfer were developed which eliminated the need for filtration of the infrared. Recent developments in the plasma generation of light with metal halide arc lamps make it possible to design a light source with a specific desired spectrum eliminating requirements for filters yet fitting configuration of lamp jackets.

The total illumination emitted from the Quartzline lamps is controlled through the regulation of the applied voltage using a powerstat. The total illumination may be computed by the following equation furnished by the lamp manufacturer where V is voltage and I is total illumination in lumens per lamp,

$$I = 33{,}000 \left(\frac{V}{277}\right)^{3.54}.$$

Heat transfer and temperature control: The design of the water jackets around the Quartzline lamps was determined experimentally based on heat transfer limitations. In the operation of an earlier high intensity system it was found that when the culture temperature rose at a rate of 1°C in 20 sec., the culture was adversely affected (6, 7). This rise in temperature was calculated to be an increase of 2000 calories per min. per lamp. Using this value as a limiting factor, several sizes of lamp jackets were tested, and a 38 mm diameter selected. It appears that a major portion of the infrared and visible spectrum pass successively through the inner 20 mm diameter pyrex tube surrounding the lamp, the distilled water layer, the outer 38 mm pyrex tube, and the medium where it is absorbed by the algal cells and converted into heat. This heat is then transferred by conduction through the surrounding medium, through the outer pyrex tube and back to the distilled water. Chilled distilled water is pumped to the lamp jackets and the heated water is returned to a heat exchanger for cooling. The use of copper sulfate for

absorbance of the infrared was tried but found to materially reduce the energy in the visible spectrum at concentrations effective for filtration of the infrared. The cooling water circulated around the lamps serves as the primary cooling system.

A 3/8" 304 series stainless steel tubing coil through which cold water is circulated on demand serves as the secondary cooling system. A solenoid value actuated through a temperature sensing device regulates the flow of cold water through the cooling coil.

The design approach was to circulate cooling water from a single heat exchanger through both the primary and secondary cooling systems. This was found to give too large a variation in temperature ($\pm$ 2.5°C). A procedure of using dual heat exchangers was worked out which permits a maximum temperature variation of $\pm$ 0.2°C from the mean. Water entering the lamp jackets is maintained at a maximum of 27°C while the cooling water delivered to the stainless steel coil is maintained at 4°C. The function of mixing in heat transfer is discussed below under the respective operation.

Safety temperature controls are an essential part of the equipment. An exterior surface mounted thermistor attached to the bottom flange sensesthe liquid temperature while a second surface mounted thermistor attached to the top flange senses overheating of the top portions due to an accidental drop in liquid levels below the lamps. These two sensing devices are wired in series to turn off the power to the lamps at preset maximum temperatures. Autoclavable thermistor probes are immersed in the liquid culture suspension to measure the culture temperature and to sense temperature for control of the secondary cooling system.

Gas transfer: Inlet gas flows from an air and carbon dioxide mixing facility through a metering valve, rotameter and Millipore filter to the culture vessel where it is injected through a 1/16" orifice directly below the center of the turbine impeller. A 1/16" stainless steel plunger is incorporated in the gas injection unit to facilitate cleaning the orifice without shutting down the system. Handling of outlet gases is described under flow model section.

Gas transfer, primarily from the standpoint of maintaining a supply of carbon dioxide to the organisms may be controlled by several methods: concentration of carbon dioxide in influent gases (air), flow rate of gas, and mixing.

Mixing: Mixing also serves as the means of controlling light-dark phasing, heat transfer, and availability of nutrients. While it is possible to determine relative importance to mixing upon each parameter, what is truly required is a definition of the effects of mixing variables upon the combination of the four

parameters in terms of net growth response. Light-dark phasing and heat transfer are flow dependent operations which give optimum results at low mixing speeds and large ratios of impeller to tank diameters (10). Gas exchange is a turbulence dependent operation which gives optimum results for high mixing speeds and small ratios of impeller to tank diameters. Nutrient availability is more than likely a function of both flow and turbulence. In the HILIS the rate of mixing may be varied from 0 to 1200 rpm and the impeller diameters from 1-1/2" to 6". The culture vessel has an equivalent diameter of 9.64 in. The culture vessel has an 8-3/4" square section with the corners beveled to improve the mixing configuration. The coils, piping, probes, and lamp jackets appear to serve as adequate baffles.

Flow model: Operation of the HILIS as a continuous steady state system will permit an accurate determination of the maximum and minimum concentration of nutrients and pH control. In a batch system the nutrient concentrations are continually changing from the time of seeding to the time of harvesting. In a continuous system, operating under steady state conditions, the nutrient feed rate is equal to the rate of harvest resulting in a constant concentration and composition of both nutrients and organisms. A change in concentration of a specific nutrient in the nutrient feed results in a change in concentration of that nutrient in the culture vessel. Any effect of a change in nutrient concentration is reflected directly by a change in the concentration of organisms in the culture and harvest.

Nutrient is fed to the culture chamber using a peristaltic pump, the rate being controlled by a timer. The harvest can be handled using an overflow weir tube or a pump controlled by an electronic level control. The overflow weir tube has been found to be the more dependable method of harvest. The harvest overflow tube also serves as the gas outlet. The liquid harvest suspension is trapped in a 20 liter carboy and the gas passes through a second trap followed by a 20% acid wash to reduce cross contamination in the laboratory. When the liquid harvest is electronically controlled, a separate gas outlet is provided from the top of the gas head space.

Sterilization: The HILIS was designed and built to permit in-place steam sterilization of all metal and glass components for pure culture work. Stainless steel of 304, 308 and 316 series has been used where contact with liquid and moist gases occur. Viton gaskets are used throughout for seals between metal parts and between glass and metal parts. Valving arrangements were devised to permit steam sterilization of pipe junctions between culture chamber and complementary vessels and equipment. For work with a single algal strain and its symbiotic microorganisms, primarily bacteria, a simpler valving arrangement has been used.

Foam control: An electronic system senses the presence of foam through a probe and actuates a peristaltic pump to deliver a limited quantity of antifoam. The period between additions may also

be controlled as desired by a timer.

ENVIRONMENTAL PARAMETER STUDIES

The basis for selecting the order of environmental parameters to be studied were: (1) independence from influence of other parameters, (2) influence upon the other parameters, (3) operational characteristics of HILIS, and (4) importance with respect to research aims. The planned sequence for the study of the parameters is as follows: (1) temperature, (2) light intensity, (3) mixing as an independent variable of gas exchange, light-dark phasing, heat transfer and nutrient availability, (4) light quality, (5) gas exchange, (6) determination of maximum and minimum concentration of nutrients, and (7) pH.

Work on temperature and light intensity have been completed. These completed studies and their significance are presented to demonstrate that the rationale of the approach is correct and that the HILIS can perform as designed. Chlorella 71105 was chosen for these studies as more is known about its physiology than any other alga and has served as the model organism in our previous studies. As the Chlorella 71105 with the natural occurring symbiotic microorganisms were used for these studies, aseptic techniques were not followed. A selection of operating conditions within the limits of the environmental conditions for algal growth was made based on previous experience in operation of earlier illuminated fermentors and actual experimentation during the long period required for the adjustment of mechanical and electrical functioning units. The conditions set for these studies have been summarized in Table 1.

Temperature parameter: Under low intensity light (less than 5000 lumen) the optimum temperature for growth for Chlorella 71105 has been determined to be 39°C. As explained earlier, the problem of heat transfer under very intense illumination is a complicated one. The algal cells as they pass the light source are undoubtedly hotter than the surrounding medium and are nearly as cool as the medium as the cells pass by the secondary cooling coil. To place the temperature sensing probe at the correct point to measure the precise average temperature of the cells or the medium would be a major undertaking of little value with relation to the immediate aims of the project. The same end was more easily attained by placing the sensing probe in a position least affected by either the light source or the cooling systems and then measuring growth response at specific measured temperature to obtain the range for optimum growth. The values for dry weight, packed cell volume and cell count gave curves similar to that shown in Figure 3 for the dry weight versus temperature.. A study of the curves indicate that good growth occurs within a range of 35°C and 39°C with a maximum growth occurring at 35°C. A T-6 medium was used for this study (6, 7). As cell concentrations at 200,000 lumen and 35°C were high (18.2 g/l), there was concern that nutrients might be

limiting growth. A new T-30 medium was designed to support 30 g/l following a procedure described in a previous publication (6, 7), and was tested under the same conditions as the T-6 medium. As no increase in dry weight, packed cell volume, or cell count was encountered using the T-30 medium, the nutrients were not considered to be limiting growth.

Light intensity: Operating at 35°C, the effect of light intensity upon growth was studied. Results are given in Table 2. A plot of this data (Figure 4) indicates that total yield in g/day dry weight may be expressed as a linear function of the log of total illumination in lumens when feed rate is constant.

It appears that the maximum values for cell concentration and total illumination have not been reached at the largest measured value of 25.5 g/l dry weight under 300,000 lumen (this illumination is equivalent to 30 times the intensity of sunlight as measured over a square foot of the earth's surface at noon). At a total illumination of 300,000 lumen, the surface intensity at the outside of the 38 mm jacket is 189,000 lumen. This intensity is 90 times greater than the value of 2000 lumen used in the batch system in 1960. An arithmetic plot of the data shows that the optimum utilization of light energy occurs in a range between 100,000 and 200,000 lumen total illumination under the environmental conditions studied. This is not the ultimate; it is but the first approximation. Before any meaningful attempt at precise definition of total illumination and light intensity can be attempted, the other environmental conditions need be brought into the ranges for optimum growth. These results do offer a basis for predicting the effect of total illumination upon growth and demonstrate that neither maximum intensity nor maximum total illumination have been reached.

In a practical sense, the results are even more encouraging. The first aim of this research has been achieved with a major reduction in volume requirements. In 1961, 400 liters were considered a basis for the design of algal systems to meet the food and gas exchange requirements of one man. The present attainment demonstrates that only 21.6 liters are required (based on 500 g/man/day intake).

With respect to the second aim there has been a significant increase in energy conversion efficiency over our earlier work. A comparison of the results of the HILIS with the earlier 4 lamp high intensity system (6, 7) may be made on the basis of total energy input measured as lumens of visible light when the dilution rates are equal. On the basis of total energy at 52,000 lumen, total daily yields from the HILIS are 262% of those of the 4 lamp system.

FUTURE

When all nine environmental conditions can be brought into range for optimum growth the volume requirements are expected to approach 5 liters per man. At this time overall energy conversion efficiencies in converting electrical energy into a high energy biomass will probably approach 40%. High energy conversions will be attained through matching light sources with light requirements. Initial studies on light quality offer exciting prospects of a major breakthrough in energy conversion when coupled with present-day attainments by lamp manufacturers in plasma generation of light. The metal halide arc lamps are 5 times as efficient as incandescent lamps and, in addition, permit selection of desired wavelengths.

Light quality: Although much work has been reported as to the effect of certain narrow band spectra upon the photosynthetic process, none of it could be correlated to permit even a reasonable selection of specific monochromatic lights, or their relative energies for the design of light quality experiments in the HILIS. Work was conducted, under contract, by Dr. Thomas E. Brown of The Charles F. Kettering Institute to make a definitive study of the monochromatic light requirements of 17 strains of algae representing ten taxonomic divisions and known differences in photoactive pigmentation (11). Each strain was individually grown at equal energy levels in white light and light of nine separated 10 nm bandwidths corresponding to the major absorption peaks of known pigments. Measurements were made of growth, pigmentation, photosynthesis, respiration, and where possible, morphology and structure. These studies indicate: (1) while light of one specific wavelength is required for the formation of a pigment, usually another wavelength is necessary for the efficient utilization of the pigment in oxygen production, (2) broad spectrum light is many times more important in pigment formation than in pigment utilization, (3) light specificity has a significant effect upon the respiration of organisms which are predominately heterotrophs (this effect is primarily a function of light adaptation), and (4) there is a synergistic effect resulting from compounded wavelengths. These results not only offer a firm basis for wavelength selection based on pigment action but suggest methods for controlled ripening of green fruit at time and place of use, control of growth of heterotrophic organisms, and control of mode of reproduction in algae.

SUMMARY

Under the mass culture phase two problems have presented major obstacles to the ultimate establishment of a practical culture system - large culture volume requirements and poor energy conversion efficiencies. An investigation of the parameters influencing growth has been the approach to eliminating these obstacles. The HILIS was designed as a functional experimental system to permit the definition of the environmental parameters for optimum production

of an algal biomass offering rational engineering design criteria basic to scale up. These parameters include temperature, light intensity, light quality, maximum and minimum concentration of nutrients, light-dark phasing, gas exchange, heat transfer, availability of nutrients, and pH. Results of light intensity studies demonstrate that the first goal - high yield in a small space - is attainable. The HILIS has been operated as a continuous system with a dilution rate of 0.91 per day (7L/day) under 300,000 lumen of incandescent illumination (30 times intensity of sunlight) yielding 25.5 gm/L dry weight of Chlorella 71105. The increased culture density attained through light reduces the culture volume required for the support of one man from the often stipulated 400 liters to 21.6 liters.

Initial results in light quality parameter studies make it possible to select the spectrum for maximum utilization of light by an algal strain. This, coupled with plasma generation of light offer prospects of a major breakthrough for greatly increasing the energy conversion efficiencies.

These initial attainments not only confirm the rationale of the approach to define the environmental parameters in one system but also demonstrate the means for doing so. It would appear that the HILIS is to become a prototype of systems for the mass culture of algae, a food of the future.

ACKNOWLEDGEMENTS

Grateful acknowledgement must be made to Dr. Mary Mandels, Mr. John A. Kostick, and SP/4 Edward Allen for their assistance and suggestions which materially aided in the present successful operation of the HILIS and to many others whose assistance and encouragement made the conception of an idea a reality.

BIBLIOGRAPHY

1. Tamiya, H. Role of algae as food. In Reports from the Japan Microalgae Research Institute, 1 (1): 9-24 (1959).

2. Farror, W. V. Tecuitlati: A glimpse of Aztec food technology. Nature, 211: 341-342 (1966).

3. Powell, R. C., Nevels, E. M. and McDowell, M. E. Algae feeding in humans. J. Nutri., 75: 7-12 (1961).

4. Matthern, R. O. The potential of algae as a food. Activities Report, 18 (2): 101-109 (1966).

5. Arthur D. Little Inc. Pilot-plant studies in production of Chlorella. In Burlew, J. S. (ed.). Algal culture from laboratory to pilot plant, pp. 234-272. Washington, D. C.: Carnegie Institution of Washington Publication 600. (1953).

6. Matthern, R. O., and Koch, R. B. The continuous culture of algae under high light intensity. Presented at AIBS Symposium, Space Biology, Oregon State Univ, Corvallis, 28 Aug. 1962 (Published Am. Biol. Teach., 25 (6 & 7): 502-511 (1963).

7. Matthern, R. O., and Koch, R. B. Developing an unconventional food, algae, by continuous culture under high light intensity. Food Technol., 18 (5): 58-62, 64-65 (1964).

8. Hannon, P. J. and Patouillet, C. Gas exchange with mass culture of algae. Appl. Microbiol., 11: 446-452. (1963).

9. Miller, R. L. and Ward, C. H. Algal bioregenative systems. SAM-TR-66-11, USAF School of Aerospace Medicine, Brooks Air Force Base, Texas. (Feb, 1966).

10. Rushton, J. H. and Oldshue, J. Y. Mixing of liquids. Chem. Eng. Progr., Symposium Series, 55 (25): 181-198 (1959).

11. Brown, T. E. Spectral requirements of algae. Final report in preparation. Project No. 1C014501A71C, Contract No. DA 19-129-AMC-565 (N), U. S. Army Natick Laboratories, Natick, Mass.

Table 1 - Conditions for Operation of HILIS for temperature and light intensity parameter studies.

Condition	Units	Setting for Study of Temperature	Setting for Study of Light Intensity
Total Illumination	Lumen	200,000	25,000 to 300,000
Gas Flow Rate	l/hr	60	60
Conc. of CO_2 in Air	% of total	10	10
Impellar Dia.	inches	4	4
Impellar/tank	diameter ratio	.43	.43
Impellar Speed	rpm	450	450
Nutrient Feed Rate	ml/hr	284 ± 5	295 ± 10
Dilution Rate	per day	~0.9	0.91
Temperature	°C	31 to 40	35
Organism		Chlorella 71105	Chlorella 71105
Nutrient (6)		T-6 & T-30	T-6
Primary cooling inlet	°C	~ 21	21
Secondary cooling inlet	°C	4	4

Steady State Condition - Above conditions maintained with ± 0.2 gm/L dry weight for a minimum period of 48 hr.

Table 2 - Summary of data taken measuring growth as a function of light intensity at dilution rate of 0.21/day in HILIS.

Total Illumination Lumens	Run	Length Run hours	Dry wt. g/l	PCV ml/ml	Total Yield g/day
25,000	1	72	5.20	.021	36.4
	2	48	5.25	.021	
	3	72	5.15	.021	
50,000	1	91	9.8	.039	68.6
	2	48	9.8	.038	
	3	192	9.7	.038	
100,000	1	96	15.6	.052	107.8
	2	72	15.4	.051	
	3	72	15.7	.051	
200,000	1	264	19.8	.092	138.6
	2	115	19.8	.093	
300,000	1	144	25.5	.132	178.5

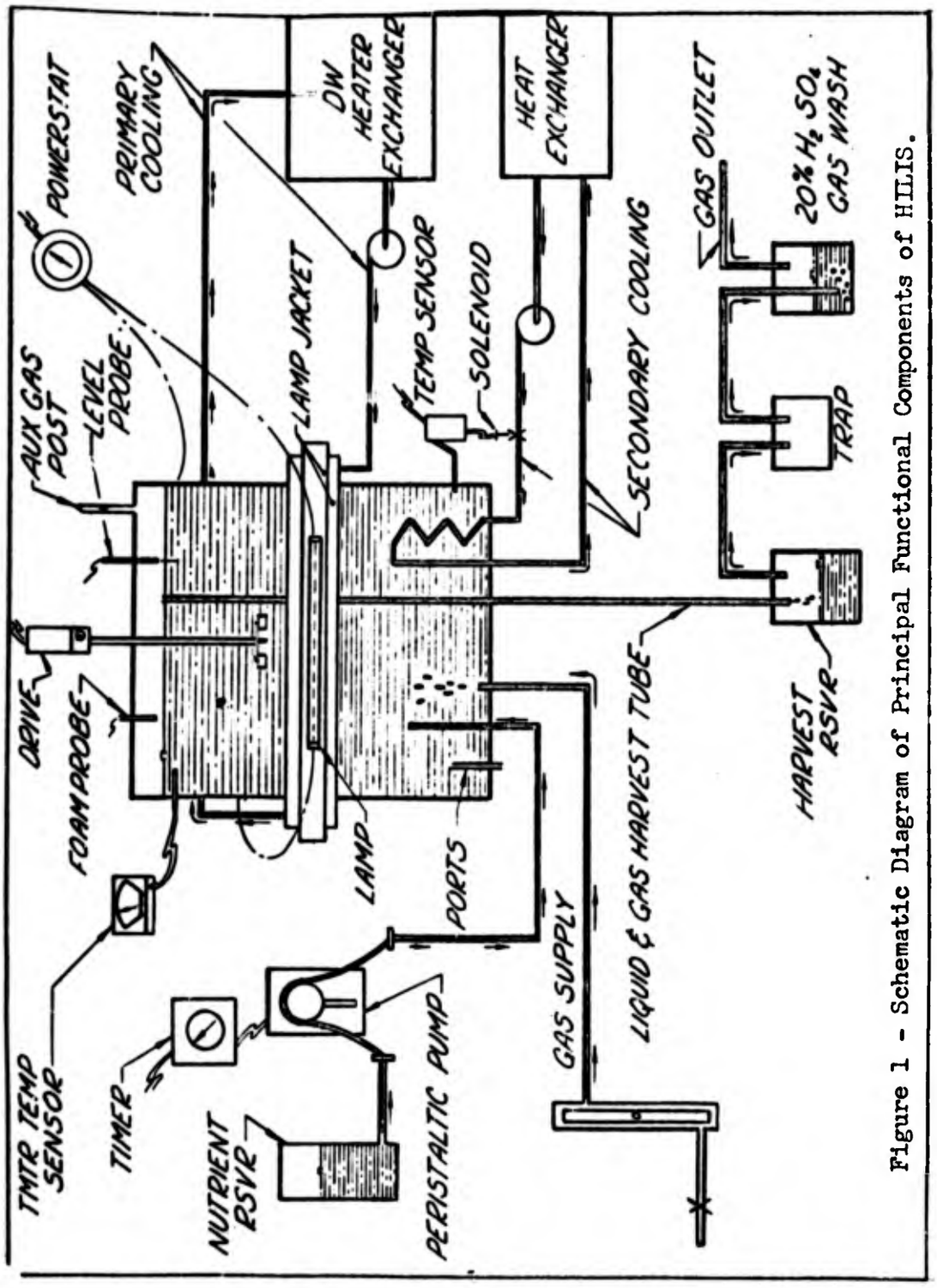


Figure 1 - Schematic Diagram of Principal Functional Components of HILIS.

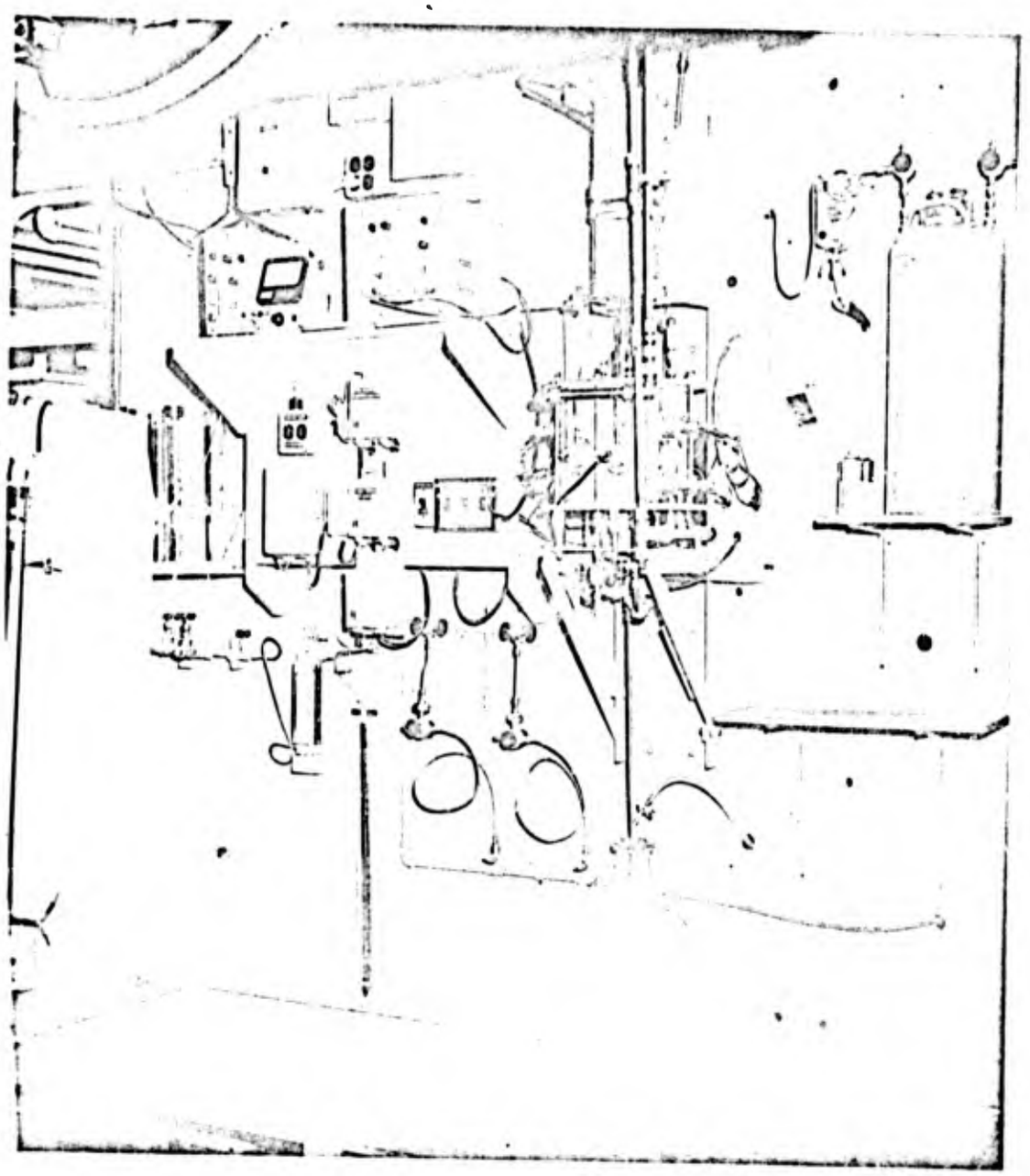


Figure 2 HILIS (High Intensity Light System) set up for continuous culture of Chlorella 71105. On left side are all electronic and electrical switching, antifoam, temperature indicator, level, timers, and temperature relay controls mounted on offset panel below which is gas metering and filtration line. In center, from top to bottom are variable drive to impeller, culture chamber and harvest reservoir. On right are nutrient reservoir and peristaltic pump. Heat exchangers, gas trap and wash, and powerstat to lamps are not shown in photograph.

MATTHERN

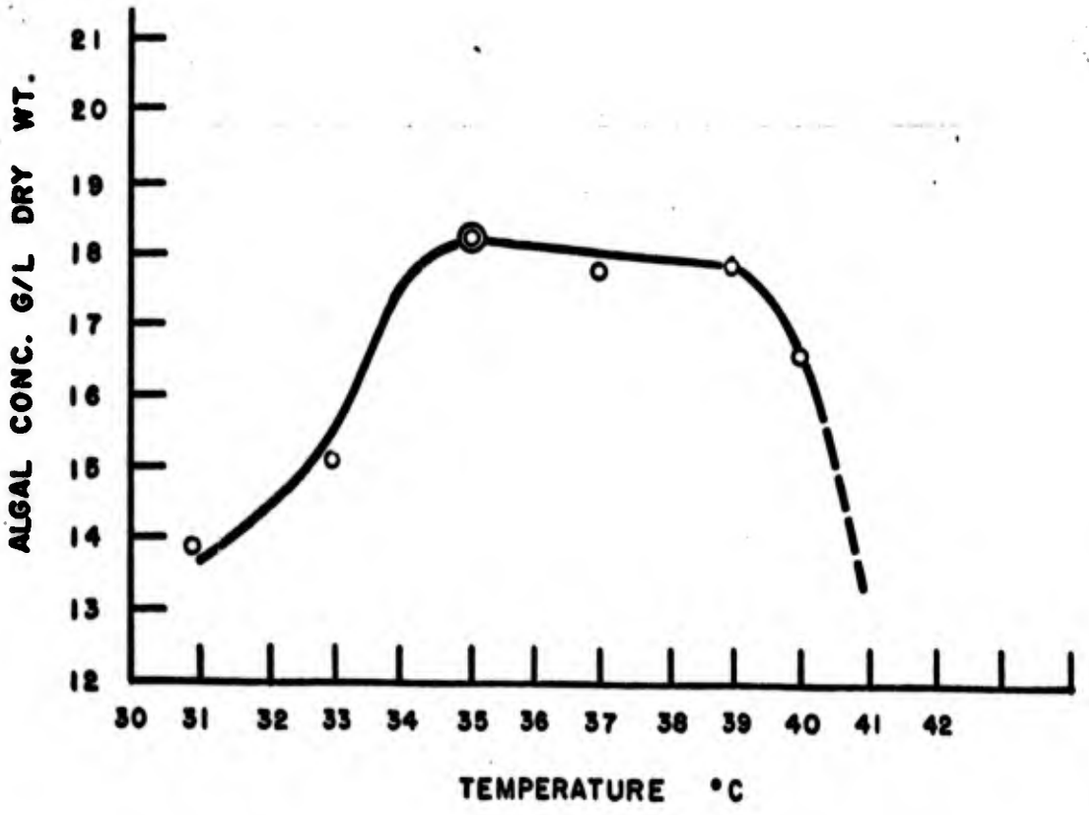


FIGURE 3 - Effect of Temperature on Growth of Chlorella 71105 in HILIS

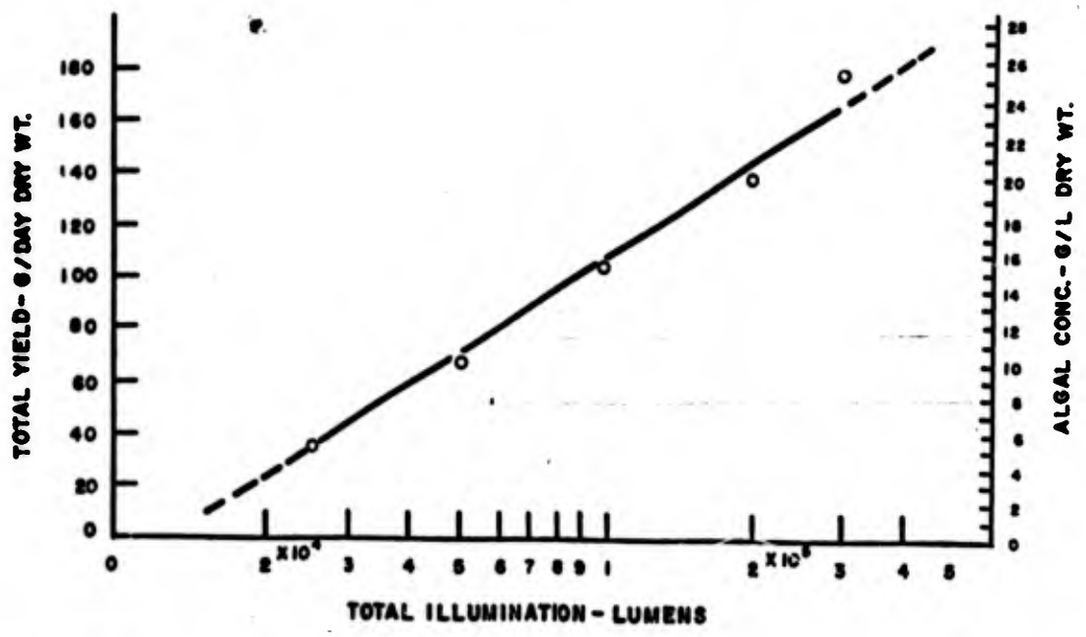


FIGURE 4 - Effect of Illumination upon Production of Chlorella 71105 During Continuous Operation of HILIS, Dilution Rate 0.91/Day

15

BOUNDARY-LAYER STUDIES ON HELICOPTER ROTORS IN FORWARD FLIGHT

W. J. McCROSKEY, CAPT., ORDC
U. S. ARMY AERONAUTICAL RESEARCH LABORATORY
AMES RESEARCH CENTER, MOFFETT FIELD, CALIF., 94035

The helicopter has proven to be one of the most useful pieces of equipment the Army has in the field. This success has stimulated numerous new concepts for more advanced aircraft, such as the AH-56 compound helicopter, and most of these new designs draw heavily from rotary-wing technology. Therefore, the Army's continuing interest in basic rotary-wing aerodynamics research can be expected ultimately to produce even better flight vehicles for supporting field operations.

From the point of view of the research scientist, the aerodynamics of helicopter rotors provides some of the most complex problems to be found in the field of fluid mechanics. These problems have been sadly neglected heretofore, and the available theories are quite incapable of fully describing the flow phenomena associated with high-speed forward flight. Therefore, the design engineer who seeks to improve the performance of contemporary helicopters must turn to a backlog of experience, empiricism, limited experimental data, and analyses of highly simplified special cases. In this paper, we shall seek to improve the basic understanding of the aerodynamic factors that limit the performance of rotors, particularly with regard to maximum lift capability and flight speed. For the most part, these limiting factors can be traced to the separation characteristics and skin friction drag of the viscous boundary layers on the rotor blades, and these boundary layers are more complex than the steady, approximately two-dimensional ones that exist on fixed-wing aircraft. We are led, then, to address these two basic questions: First, "What is the actual structure of the viscous flow?" and second, "Do three-dimensional and unsteady effects in the boundary layer significantly affect the performance of the rotor?" If they do, then we would expect the classical methods for designing fixed-wing aircraft to be inadequate for high-performance helicopters.

In attempting to answer the preceding questions, we must first develop a realistic theoretical model for the viscous flow over a rotor blade so that we can formulate a well-posed mathematical

problem. Next we shall solve the governing equations for several idealized examples of laminar, incompressible flow. These examples are carefully chosen to represent the essential features of situations encountered in practice, and at the same time, to correspond in some respects to classical two-dimensional flows that have already been studied extensively in connection with fixed-wing aircraft. Then we shall discuss these solutions in comparison with the results of some preliminary experiments designed to check both the assumptions of the theory and the extrapolation of the results from laminar to turbulent flows. Finally, we shall mention how our results can be utilized in practice and indicate the directions in which additional research in this area should proceed.

GENERAL FORMULATION

Our basic procedure is to perform a regular perturbation expansion of the unsteady, three-dimensional Navier-Stokes equations of motion for an incompressible boundary layer with small crossflow. We shall refer to the flow normal to the rotor blade as the primary flow, denoted by u and v in Fig. 1. The crossflow, w, is in the spanwise direction. It is advantageous to adopt Cartesian axes (x,y,z) fixed to the leading edge of the moving blade, as shown in Fig. 1, so that the flight velocity, V_∞, approaches the blade at an angle ψ. The y-axis is the axis of rotation, and so long as x/z remains small, z can be interpreted as the radial distance. This is not the case, however, for the crossflow velocity, w. The radial velocity plotted in the subsequent figures is given by $V_r \simeq w + (x/z)u$ where, generally, w and $(x/z)u$ are comparable in magnitude.

For the coordinate system we have chosen, the appropriate equations are those found in the standard texts on boundary-layer theory, such as Schlichting [1], with the addition of Coriolis and centrifugal forces, namely,

$$\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} + \frac{\partial w}{\partial z} = 0 \tag{1}$$

$$\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} + v\frac{\partial u}{\partial y} + w\frac{\partial u}{\partial z} = 2\Omega w + \Omega^2 x - \frac{1}{\rho}\frac{\partial p}{\partial x} + \nu\frac{\partial^2 u}{\partial y^2} \tag{2}$$

$$\frac{\partial w}{\partial t} + u\frac{\partial w}{\partial x} + v\frac{\partial w}{\partial y} + w\frac{\partial w}{\partial z} = -2\Omega u + \Omega^2 z - \frac{1}{\rho}\frac{\partial p}{\partial z} + \nu\frac{\partial^2 w}{\partial y^2} \tag{3}$$

where ρ is the density, ν the kinematic viscosity, and p the static pressure.

In general, these equations are too complex to solve analytically, so we must seek special cases for which certain terms play secondary roles. One important case has been explored by the author in Ref. 2, where it was shown that the condition $w/u \ll 1$ affords

enormous simplifications with little loss in generality for a wide range of practical applications. The portion of the rotor disc where this assumption is valid is indicated by the unshaded region in Fig. 1; in addition, we must exclude stagnation and separation lines on the blade, where $w \geq u$. Under these restrictions, the independent variables can be written in the form of a regular perturbation expansion as

$$u = u^{(0)} + \epsilon u^{(1)} + \epsilon^2 u^{(2)} + \ldots \tag{4}$$

$$v = v^{(0)} + \epsilon v^{(1)} + \epsilon^2 v^{(2)} + \ldots \tag{5}$$

$$w = \epsilon w^{(1)} + \epsilon^2 w^{(2)} + \ldots \tag{6}$$

where the small parameter ϵ is approximately equal to x/z. This expansion gives the following set of first-order equations:

$$\frac{\partial u^{(0)}}{\partial x} + \frac{\partial v^{(0)}}{\partial y} = 0 \tag{7}$$

$$u^{(0)} \frac{\partial u^{(0)}}{\partial x} + v^{(0)} \frac{\partial u^{(0)}}{\partial y} = \nu \frac{\partial^2 u^{(0)}}{\partial y^2} - \frac{1}{\rho} \frac{\partial p}{\partial x} \tag{8}$$

$$u^{(0)} \frac{\partial w^{(1)}}{\partial x} + v^{(0)} \frac{\partial w^{(1)}}{\partial y} = \nu \frac{\partial^2 w^{(1)}}{\partial y^2} - 2\Omega u^{(0)} - \frac{1}{\rho} \frac{\partial p}{\partial z} \tag{9}$$

These reduced equations contain all the essential qualitative features of the flow, erring only in quantitative details that can be estimated by solving the higher order equations.

To complete our mathematical formulation of the problem, we need to specify the boundary conditions for equations (7) to (9). At the surface of the blade, $u = v = w = 0$. At the outer edge of the boundary layer, $u = u_i$ and $w = w_i$, where the subscript i denotes the local inviscid flow just outside the boundary layer.* In the most general case, determining u_i and v_i is a formidable task, but here we shall utilize the results from Ref. 2 for a generalized inviscid solution for rotors with constant circulation** in the spanwise direction,

$$u_i = (\Omega z + V_\infty \sin \psi) \frac{\partial \Phi}{\partial x} \tag{10}$$

$$w_i = \Omega(\Phi - 2x) + V_\infty \cos \psi \tag{11}$$

---

*Here, as in classical boundary-layer theory [1], we cannot specify an outer boundary condition for v.

**Constant circulation implies a lift distribution that increases linearly with z.

where $\psi = (\Omega t + \psi_o)$ is the angle defined in Fig. 1, Ω is the rotational speed of the rotor, and $\Phi(x,y)$ is the two-dimensional potential for uniform flow past the airfoil section in question, placed in a two-dimensional stream of unit speed in the x direction. In other words, equations (10) and (11) enable us to construct a family of inviscid solutions for rotors from classical airfoil theory.

The significance of the preceding development is threefold. In the first place, the first-order primary flow over most of the rotor is governed by equations (7) and (8) and is independent of both crossflow and unsteady effects. With the aid of equation (10), we can construct the primary flow point by point around the rotor disc. Secondly, we can use the higher order solutions, $u^{(1)}$, $u^{(2)}$, etc., to determine the trends that will occur in the regions where our assumption that $w/u \ll 1$ begins to break down. Thirdly, the crossflow equation for w, equation (9), is linear. Therefore, we can split w into a particular solution describing the effects of rotation, or hover, and a complementary solution describing the effects of translation, or forward flight. The particular solution, w_a, satisfies equation (9) and the following inhomogeneous boundary condition at the edge of the boundary layer:

$$w_a(y \to \delta) = \Omega[\Phi(x,0) - 2x] \tag{12}$$

The complementary solution, w_b, is governed by

$$u \frac{\partial w}{\partial x} + v \frac{\partial w}{\partial y} = \nu \frac{\partial^2 w}{\partial y^2} \tag{13}$$

$$w_b(y \to \delta) = V_\infty \cos \psi \tag{14}$$

Of course, w_a and w_b, as well as u, vanish at $y = 0$.

REPRESENTATIVE EXAMPLES

From the general formulation outlined above, we can solve the boundary-layer equations for arbitrary rotor blades if we are furnished the pressure distribution over the blade, or alternatively, the potential function $\Phi(x,y)$. This is generally a formidable task, however, and to avoid becoming lost in a great forest of numerical calculations, let us examine some representative examples that yield analytical solutions.

Flat Plate

The simple case of a flat-plate blade at zero angle of attack, for which $\Phi = x$ and $p =$ constant, provides a great deal of insight into the relative importance of the translational and rotational motion. In the next example, we shall examine the additional

effects of pressure gradients, but for the moment the primary quantities of interest are the crossflow velocity profiles, the influence of the crossflow upon the second-order primary flow, and the skin friction.

If we define the boundary-layer similarity variable as

$$\eta = y\sqrt{(\Omega z + V_\infty \sin\psi)/2\nu x} \tag{15}$$

and write the velocity profiles in the form

$$u^{(0)} = (\Omega z + V_\infty \sin\psi)F_0'(\eta) \tag{16}$$

$$\epsilon u^{(1)} = \left(\frac{\Omega x V_\infty \cos\psi}{\Omega z + V_\infty \sin\psi}\right) F_1'(\eta) \tag{17}$$

$$\epsilon^2 u^{(2)} = \left(\frac{\Omega x}{\Omega z + V_\infty \sin\psi}\right)^2 \left[F_{20}'(\eta) + \left(\frac{V_\infty \sin\psi}{\Omega z + V_\infty \sin\psi}\right) F_{21}'(\eta) + \left(\frac{V_\infty \cos\psi}{\Omega z + V_\infty \sin\psi}\right)^2 F_{22}'(\eta)\right] \tag{18}$$

$$\epsilon w^{(1)} = V_\infty \cos\psi\, G_0'(\eta) + \Omega x\left[-G_{10}'(\eta) + \left(\frac{V_\infty \sin\psi}{\Omega z + V_\infty \sin\psi}\right) G_{11}'(\eta) + \left(\frac{V_\infty \cos\psi}{\Omega z + V_\infty \sin\psi}\right)^2 G_{12}'(\eta)\right] \tag{19}$$

$$\epsilon^2 w^{(2)} = \Omega x\left(\frac{x}{z}\right)\left\{G_{20}'(\eta) + \left[\frac{\Omega z V_\infty \cos\psi}{(\Omega z + V_\infty \sin\psi)^2}\right] G_{21}'(\eta) + \left[\frac{\Omega z V_\infty^2 \sin\psi\cos\psi}{(\Omega z + V_\infty \sin\psi)^3}\right] G_{22}'(\eta) + \left(\frac{\Omega z}{\Omega z + V_\infty \sin\psi}\right)\left(\frac{V_\infty \cos\psi}{\Omega z + V_\infty \sin\psi}\right)^3 G_{23}'(\eta)\right\} \tag{20}$$

then equations (1) to (3) reduce to a set of third-order, ordinary differential equations in which F_i' and G_j' are universal functions of η. These equations have been solved numerically and the results are presented in Fig. 2.

A point of current interest and controversy in the field of rotary-wing aerodynamics is the magnitude and direction of the crossflow. From equation (19), we see that the main parameter that

determines the relative size of the crossflow due to rotation compared to that due to translation is $(V_\infty \cos \psi)/\Omega x$. Now, a typical rotor might operate at Ω = 300 rpm and V_∞ = 100 ft/sec; then, at 1 foot from the leading edge and 10 feet from the hub, $x/z = 0.1$ and $V_\infty/\Omega x \simeq 3$. Therefore, let us look at the radial velocity profiles for these representative conditions, plotted in Fig. 3 for eight points around the rotor disc. In this and subsequent figures, the curves are drawn to the point where $u/u_1 = 0.99$, which is taken to be the edge of the boundary layer.

The most striking feature of these results is that the crossflow due to translation dominates the radial velocity over most of the rotor disc for these conditions. For $\psi = 90^\circ$ and 270°, the crossflow is primarily due to rotation, and the principal effect of the translational motion is to thin or thicken the boundary layer. However, for points only 45° away, the translational contribution clearly predominates. This effect is further magnified as we increase the flight velocity or decrease the distance from the leading edge, but it is virtually independent of z.

Let us now examine the effect of the crossflow upon the primary flow. Several velocity profiles are plotted in Fig. 4 for the same conditions as before, except that x is larger so as to accentuate the effects of rotation. We note that although the higher order corrections are quite small compared to the first-order (Blasius) solution, they generally "accelerate" the primary flow, thereby tending to delay separation and transition to turbulence. The function F_1 represents both unsteady effects and the coupling of the yawed wing crossflow with the primary flow. It is "stabilizing" for $90^\circ < \psi < 270^\circ$ and "destabilizing" over the rear half of the disc; the dominant physical mechanisms are the inertia forces and the momentum flux $w(\partial u/\partial z)$. Rotational effects comprise F_{20}, which is always stabilizing by virtue of the Coriolis and centrifugal accelerations. On the other hand, the destabilizing term F_{22} depends primarily upon the momentum fluxes and unsteady effects of forward flight. The purely unsteady term, F_{21}, lags F_1 by 90° and is destabilizing for $0 < \psi < 180^\circ$ and stabilizing for $180^\circ < \psi < 360^\circ$.

The x and z components of the skin friction are given by

$$\tau_x = \rho(\Omega z + V_\infty \sin \psi)^{3/2} \left(\frac{\nu}{x}\right)^{1/2} \left\{0.3321 - 0.06784 \left(\frac{\Omega x V_\infty \cos \psi}{\Omega z + V_\infty \sin \psi}\right) + \left(\frac{\Omega x}{\Omega z + V_\infty \sin \psi}\right)^2 \left[1.2791 - 0.7299 \left(\frac{V_\infty \sin \psi}{\Omega z + V_\infty \sin \psi}\right) - 0.3015 \left(\frac{V_\infty \cos \psi}{\Omega z + V_\infty \sin \psi}\right)\right]\right\} + \cdots \qquad (21)$$

$$\tau_z = \rho(\Omega z + V_\infty \sin\psi)^{3/2}\left(\frac{\nu}{x}\right)^{1/2}\left\{0.3321\left(\frac{V_\infty\cos\psi}{\Omega z + V_\infty\sin\psi}\right)\right.$$

$$+\left(\frac{\Omega x}{\Omega z + V_\infty\sin\psi}\right)\left[0.3331 - 0.8756\left(\frac{V_\infty\sin\psi}{\Omega z + V_\infty\sin\psi}\right)\right.$$

$$\left.\left. - 0.2782\left(\frac{V_\infty\cos\psi}{\Omega z + V_\infty\sin\psi}\right)^2\right]\right\} + \cdots \qquad (22)$$

An interesting result can be derived from these expressions if we consider the torque or power required to overcome skin-friction drag on a rotor with flat-plate blades. The average power required over one revolution is

$$P = \frac{1}{2\pi}\int_0^{2\pi}\int_A \bar{\tau}\cdot\bar{V}\,dA\,d\psi \qquad (23)$$

where A represents the area of the blades and $|\bar{V}| = (\Omega z + V_\infty \sin\psi)$ is the magnitude of the instantaneous velocity with which the element dA moves through space. Of particular interest are the contributions of the higher order terms in equations (21) and (22), in comparison with first-order results. By substituting equations (21) and (22) into equation (23), we can readily show that the percentage increase in P, above that calculated with two-dimensional strip theory, is a strong function of the "advance ratio," $V_\infty/\Omega R$, and a weak function of the "aspect ratio," R/c, where R and c are the span and chord of the blade, respectively. Accordingly, results to third order, but neglecting the reversed flow region where $(\Omega z + V_\infty \sin\psi) < 1$, are shown in Fig. 5 for $c/R = 0.1$. This type of substantial increase in P with advance ratio has been postulated in the past [3], but only empirically.

Airfoils With Pressure Gradients

With the flat-plate results as a frame of reference, let us turn now to more realistic cases involving pressure gradients. Extensions of the well-known Falkner-Skan solutions, for two-dimensional flows with inviscid velocities $u_i \propto x^n$, are useful for assessing the effects of pressure gradients upon the crossflow. We begin by writing the following transformation of the independent variables:

$$\xi = \frac{x}{c}\,; \quad \eta = y\left[\frac{\beta(n+1)(\Omega z + V_\infty\sin\psi)x^{n-1}}{2\nu c^n}\right]^{1/2} \qquad (24)$$

$$\Phi(\xi) = \frac{c\beta\xi^{n+1}}{n+1} \qquad (25)$$

where c is the chord of the blade, β is an arbitrary constant that can be adjusted to approximate more closely an actual rotor blade, and Φ is written in a form that allows us to use the powerful mathematical technique known as separation of variables. Then the appropriate forms of the velocity profiles are

$$u^{(o)} = (\Omega z + V_\infty \sin \psi)\beta \xi^n f'(\eta) \tag{26}$$

$$\epsilon w^{(1)} = \Omega c \left[\frac{\beta \xi^{n+1}}{n+1} g_1'(\eta) - 2\xi g_2'(\eta) \right] + V_\infty \cos \psi \, g_3'(\eta) \tag{27}$$

As in the case of the flat plate, substituting these expressions into equations (1) to (3) gives a set of ordinary differential equations for the universal functions f, g_1, g_2, and g_3.

We can use the f and g solutions to approximate the boundary layer on an actual blade by patching together the results for various values of n and β. One such example is shown in Fig. 6, where we use $n = 1/3$, $\beta = 2.4$ at the leading edge and $n = -0.065$, $\beta = 1.0$ downstream to simulate the inviscid flow over a standard NACA 0012 airfoil at small angles of attack.

Crossflow velocity profiles for this composite "airfoil" are shown in Fig. 7 for the same flight conditions as before (i.e., $V_\infty = 100$ fps, $\Omega = 300$ rpm, and $z = 10$ ft), at two chordwise stations. The results at $x/c = 0.1$, just upstream of the point of minimum pressure, are interesting because here the boundary layer is very thin and the skin friction is high. We see that the effects of rotation are completely negligible here; virtually all the crossflow is due to the translational motion, V_∞.

On the other hand, at $x/c = 0.7$ the pressure gradient is adverse, the boundary layer is thicker, and separation is more likely to occur. These results (Fig. 7(b)) show that the crossflow due to rotation, w_a, is even more important than on the flat plate (cf. Fig. 3). If the pressure gradient were more severe, as on the top of airfoils at large angles of attack, we would find that the centrifugal effects are sufficiently large to force the fluid near the wall radially outward at all values of ψ around the rotor disc. Thus we see that the nature of the chordwise pressure gradient and the distance from the leading edge affect the detailed structure of the crossflow. In turn, these factors will affect the influence of the crossflow upon the primary flow, since, in general, the greater the crossflow due to rotation, the more separation will be delayed.

EXPERIMENTS

The preceding analysis was developed for laminar flow under the assumption of small crossflow. Questions arise as to whether this

assumption is valid and to what extent the features of laminar flow carry over to the turbulent boundary layers often encountered in practice. Preliminary experiments were performed on the two rotor configurations shown in Fig. 8, using the evaporating chemical film technique described by Tanner and Yaggy [4]. They were interested in streamline directions and transition from laminar to turbulent flow, and their results confirm qualitatively our assumption of small crossflow, in the case of hovering rotors. Here we shall look at results for two identical blades, with the same pressure gradients and inviscid flows, but with laminar flow on one and turbulent flow* on the other. Since the boundary conditions and the centrifugal, Coriolis, and pressure forces are virtually the same in both cases, these experiments should illustrate the importance of wide variations in the momentum fluxes and viscous forces upon the average crossflow in the boundary layer.

The results for the lower surface of a conventional rotor hovering at 10° collective pitch are shown in Fig. 9, where the streaks behind small disturbances indicate the streamline directions at the surface of the blade. Also shown is the theoretical result for laminar flow, which agrees well with the data. It is interesting that the streamline directions are almost identical in both cases, indicating that the laminar theory adequately predicts the magnitude of the crossflow even in turbulent flows. This result is somewhat surprising, in view of significant differences we would expect in the viscous stresses for the two cases. On the other hand, this seems to indicate that the laminar theory can serve as a useful guide in formulating subsequent analyses of turbulent flows.

Results for the hovering offset rotor configuration (cf. Fig. 8) are shown in Fig. 10. The purpose of the offset is to simulate the inviscid crossflow due to forward flight, $V_\infty \cos \psi$, by the crossflow due to the offset, Ωd. This analogy between V_∞ and Ωd was suggested in Ref. 2 as a means of obtaining meaningful measurements without sophisticated dynamic instrumentation. As in the case of the conventional rotor, the streamline patterns do not depart greatly from circular arcs, and the laminar and turbulent results are virtually identical. These results should lend additional support to the conclusions we can draw from the laminar theory.

SUMMARY AND CONCLUSIONS

In our analysis of the boundary layers on rotors in forward flight, we have demonstrated most of the essential features of the flow by means of relatively simple analytical solutions. Within the framework of the small crossflow and quasi-steady approximations, which are valid over most of the rotor, the laminar flow is now well understood. An important exception is the detailed structure of the flow near the separation lines, where recourse to numerical analysis of the full three-dimensional, unsteady Navier-Stokes equations will

*The turbulent flow was achieved by means of a boundary-layer trip at the leading edge.

be required if we are to predict accurately the maximum lift that can be produced by a rotor.

From the present investigation, we can draw the following specific conclusions:

1. The effect of rotation can be important, and helpful in delaying separation of the primary flow, in regions of adverse pressure gradients. The crossflow due to rotation increases with increasing x/z, so that in practice the maximum benefits would be attained by developing airfoils that tend to stall at the trailing edge first. Also, maximum lift coefficients should occur on the portions of the rotor blade closest to the hub.

2. Because there is substantial crossflow due to translation, or forward flight, the boundary layer on a typical rotor generally resembles the viscous flow over a swept wing. On the basis of detailed analysis of the flow over a flat plate, the effects upon the primary flow appear to be favorable at some azimuthal positions of the blade relative to the direction of flight and unfavorable at others. The maximum benefits accrue in the quadrant $180^\circ < \psi < 270^\circ$, where "retreating-blade stall" is commonly presumed to begin. This may be one reason that actual rotors have been observed to perform better than would be expected on the basis of steady-state, two-dimensional section characteristics of the blades (cf. Harris [3]).

3. On the other hand, the stabilizing influence of the crossflow increases the power required to overcome skin friction drag. For a flat-plate blade with an aspect ratio of 10, the increase varies with flight speed from about 2% in hover to about 36% at $V_\infty/\Omega R = 1$. This may partially explain why the minimum drag coefficients of actual rotors tend to exceed those for fixed wings.

4. The experiments show that the streamline directions are nearly identical for laminar and turbulent flows. Therefore, our laminar analysis should provide valuable guidelines in formulating approximate theories for turbulent boundary layers on rotors.

5. The present study is preliminary; there appear to be numerous areas for significant additional research in connection with boundary layers on rotors. Theoretical studies should include (a) numerical analysis of equations (1) to (3) near separation lines and at frequencies of oscillation significantly larger than Ω, (b) the development of simplified theoretical models for which the turbulent boundary-layer equations can be analyzed, (c) a brief examination of the effects of compressibility, (d) an extension of the present theory to include the effects of boundary-layer blowing and circulation control, and (e) an overall program to synthesize the results of these studies into specific recommendations for improved airfoil sections and blade planforms.

6. To complement these theoretical studies most efficiently, future experiments should be directed primarily at defining the locations and patterns of separated flow, at measuring the direction of streamlines and the magnitude of the local skin friction, and at developing a better physical model of the turbulent crossflow. These measurements are difficult and challenging because they must be performed in a rotating, unsteady environment. However, the technological advances that can come from further increases in our basic understanding of rotary-wing aerodynamics are well worth the effort. Hopefully, the present investigation will stimulate new approaches in this area.

REFERENCES

1. Schlichting, H. Boundary Layer Theory (McGraw-Hill, New York, 1960), pp. 191-193.

2. McCroskey, W. J. and Yaggy, P. F. "Laminar Boundary Layers on Helicopter Rotors in Forward Flight," AIAA Paper No. 68-70.

3. Harris, F. D. "Preliminary Study of Radial Flow Effects on Rotor Blades," J. Am. Helicopter Soc., Vol. 11, No. 1, p. 1 (1966).

4. Tanner, W. H. and Yaggy, P. F. "An Experimental Boundary-Layer Study on Hovering Rotors," J. Am. Helicopter Soc., Vol. 11, No. 1, p. 22 (1966).

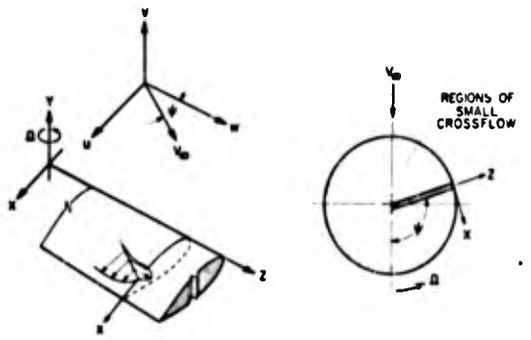


Figure 1.- Coordinates in the rotating system.

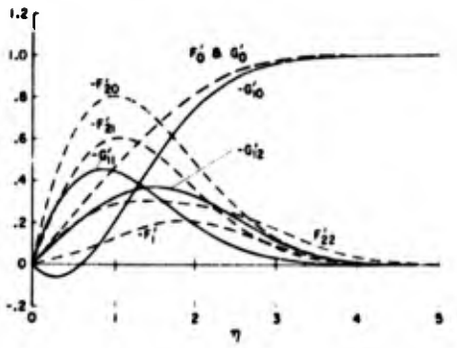


Figure 2.-Universal velocity functions for a flat-plate blade.

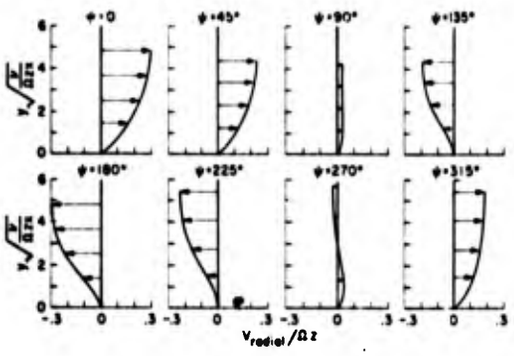


Figure 3.- Radial velocity profiles on a flat-plate blade $V_\infty/\Omega z = 0.3$, $x/z = 0.1$.

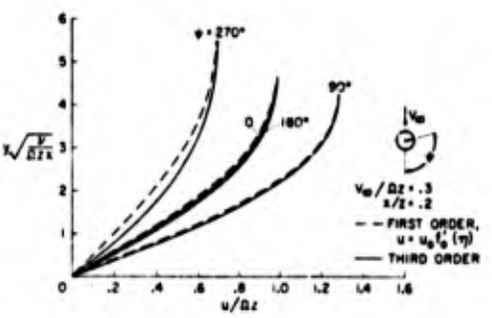


Figure 4.- Primary flow velocity profiles on a flat-plate blade.

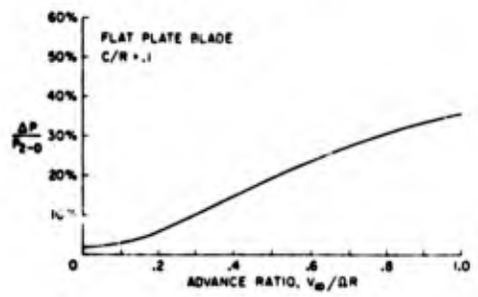


Figure 5.- Effect of crossflow on power required to overcome skin-friction drag.

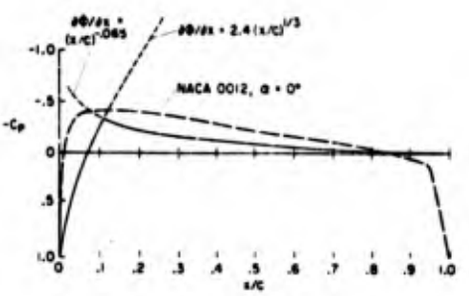
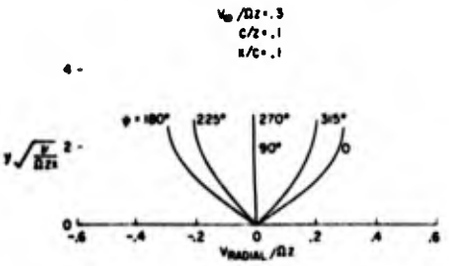
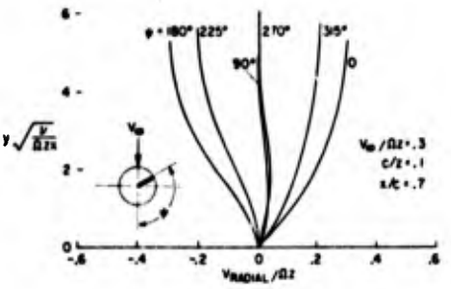


Figure 6.- Pressure distribution on composite airfoil.



(a) Near the leading edge.



(b) Near the trailing edge.

Figure 7.- Radial velocity profiles on composite airfoil.

Figure 7.- Concluded.

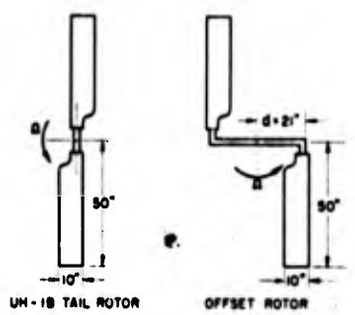


Figure 8.- Rotor configurations used in experiments.

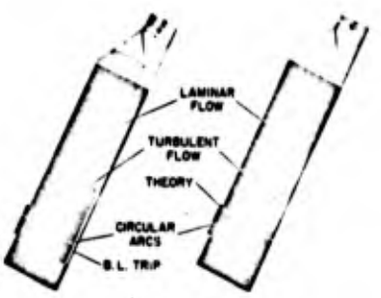


Figure 9.- Surface streamline patterns on pressure side UH-1B tail rotor.

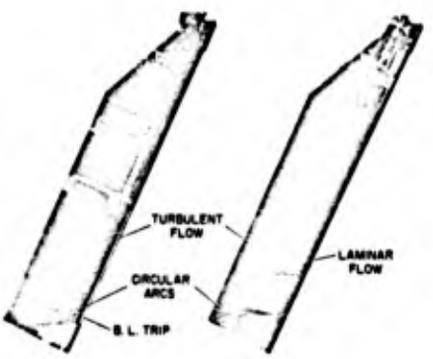


Figure 10.- Surface streamline patterns on pressure side of offset rotor.

BLANK

BOND CHARACTER AND THE METASTABILITY OF PLOSOPHORIC GROUPS OF MILITARY SIGNIFICANCE

ROBERT C. McMILLAN
U. S. Army Mobility Equipment Research and Development Center
Fort Belvoir, Virginia

INTRODUCTION: The usual calculation of chemical bond energy from electronic orbitals is difficult and subject to great variation in accuracy. Appendix A describes a simple algorithm for calculating bond energy. The algorithm utilizes the experimental interatomic distances, the assumed electronic orbitals, and the change in energy for an atom determined by the amount of charge localization. Variables may be introduced into the electronic orbitals and charge distributions varied to obtain minimum energy conditions. Similar calculations give the bond energies for molecular ions. Then the ionization potentials of molecules may be determined. Table I compares calculated and experimental values of bond energies and ionization potentials for several diatomic molecules.

TABLE I

Comparison of Calculated and Experimental Bond Energies and Ionization Potentials for Some Diatomic Molecules

	Bond Energy		Ionization Potential	
Molecule	Calculated	Experimental	Calculated	Experimental
H_2	4.809	4.476	15.64	15.6
N_2	9.845	9.756	15.34	15.51
NO	6.482	6.487	9.69	9.5
O_2	4.952	5.08	-	-

For the original calculations, the atomic electronic energy was determined using Slater's [1] rules. However, difficulty was encountered when heavy atoms were incorporated into heteronuclear molecules. Investigations revealed that although Slater's rules may be used to approximate the total energy of the atoms, they do not accurately give the first ionization potential for many atoms.

This problem has been resolved by retaining the form of the energy terms used by Slater, but determining the constants from experimental atomic ionization potentials of the valence electrons.

Although the procedure for calculating the molecular bond energy is of interest, in this paper primary emphasis will be given to the results obtained for the group of compounds known as the azides. Sodium azide (NaN_3) and potassium azide (KN_3) are relatively stable compounds. Lead azide ($Pb(N_3)_2$), used extensively as a primary explosive, hydrazoic acid (HN_3), and silver azide (AgN_3) are representatives of the less stable azides.

The bonding molecular orbitals in the azide group and the electronic charge distribution on the atoms change depending on the nature of the bond to the cation and the ionization state of the azides. Usually this distinction is made on the basis of ionic or covalent bonding to the cation. Evans and Yoffe[2] pointed out the influence of the ionization potential of the metal, or cation, on the nature of the bond to the azide. The results of the present work indicate that in addition to the effect of ionization potential, a covalent bond has characteristics that extend beyond the nearest neighbor atoms. This extension of the covalent bond over several atoms is a result of the delocalization of the electronic orbitals over the entire molecule.

The discussion that follows will cover first N_3^-, its ionization to form N_3, and further ionization to form N_3^+. Next, the effect of atomic ionization potentials and the requirements necessary to form ionic or covalent bonds will be investigated. Then it will be possible to compare different mechanisms of decomposition.

The Azide Ion, N_3^-. The sixteen electrons in the valence shell for N_3^- must be assigned to electronic orbitals. From the experience that has been gained in working with this method of calculating bond energies, it has been found that arbitrary assignment of the density of electrons in some orbitals may be offset by providing variability in other orbitals. The assignment for the electronic density on each atom for different molecular orbitals is given in Table II.

TABLE II

Distribution of Electrons for Calculating the Bond Energy for N_3^-

Type of Orbital	Electron Density		No. of Electrons
	End Atoms	Center Atom	
σ_g	c_1^0	1	2
σ_u	c_1	$1-2c_1$	2
π_u	c_2	$1-2c_2$	4
π_g	1/2	0	4
σ_g	1/2	0	2
σ_u	1/2	0	2

The bond energy calculated for N_3^-, using an assumed N-N distance of 1.16Å, is 299.8 kcal/mole. This value is in excellent agreement with the enthalpy of formation for N_3^- given by Gray[3] of 34.8 kcal/mole. Converting the enthalpy of formation to bond energy for the N_3^-, one obtains 302.7 kcal/mole.

The electronic charge distribution on the N_3^- is also obtained from the minimization of the energy. It is found that the central nitrogen atom has a charge of $+0.92|e|$. Bonnemay and Daudel[4] give $+0.66|e|$ for the value of this charge.

The Azide Radical, N_3. The removal of a single electron from N_3^- will produce N_3. Since the π_g orbital is non-bonding, this is a likely orbital for losing the electron. The calculation of the bond energy for this configuration, using the same interatomic distances as for N_3^-, yields 235.8 kcal/mole, or an electron affinity for N_3 of 64 kcal/mole. This value for the electron affinity agrees with Gray's[3] value of 69 ± 7 kcal/mole.

The Positive Azide Ion, N_3^+. As in proceeding from N_3^- to N_3 it was expected that an electron would be removed from the π_g orbital, so in going from N_3 to N_3^+ the single electron should be removed from the π_g orbital. When the interatomic distance of 1.16Å is used also for N_3^+, the calculated ionization potential is 266.5 kcal/mole compared to Gray's value of 283 ± 6 kcal/mole. Although the calculation appears to be low, it is in excellent agreement with some experimental results of Franklin, Dibler, Reese, and Krauss[5].

Nature of the Cation-Azide Bond. When will a metal azide be ionic? Some attempts have been made to answer this question by considering the ionization potential of the metal as the determining factor. In the following discussion the relationship between the metallic ionization potential and ionic character will be derived.

Consider an azide ion and a metallic positive ion. Assume no polarization of the azide ion then a minimum energy (maximum bonding) will be achieved for a linear configuration.

M	N	N	N
+1.	-.86	+.96	-.86

For this configuration the energy will be given by

$$E_{Ionic} = I_p - E_A + E_c$$

where I_p = ionization potential of the metal

E_A = electron affinity of the azide radical, and

E_c = coulombic energy of the system.

If the azide is to be ionic, this energy must be less than the energy that can be obtained from covalent bonding. The pure covalent bond energy may be approximated by the method of Appendix A to be:

$$E_{covalent} = \frac{3.598}{r}$$

where r is the metal-nitrogen bond distance in angstroms and the energy is given in electron volts (ev). Thus, an azide is ionic when

$$I_p < I_{Max}$$

when $I_{Max} = E_A - E_c + E_{Covalent}$. (I)

Since both E_c and $E_{Covalent}$ are functions of r, the maximum ionization potential, I_{Max}, that a metal may have and still form an ionic bond is determined by the metal-nitrogen bond distance. Some values for I_{Max} as a function of r are given in Table III.

TABLE III

Maximum Ionization Potential for Ionic Azides as a Function of the Metal-Nitrogen Distance

r.(Å)	I_{Max}(e.v.)
1	11.2
1.5	8.45
2.0	7.12
2.5	6.34
3.0	5.82

Although pure ionic or pure covalent bonding seldom exists, the mixing of the two types of bonding will be most noticeable when the cation ionization potential is approximately equal to I_{Max}.

To use Equation I, let us assume a radius of 0.58Å for the end nitrogen atom of the azide ion, use the ionic radius for the cation, and compare the calculated maximum ionization potential with experimental ionization potential for the metal. The results of these calculations are given in Table IV for several metals which form monovalent azides.

TABLE IV

Comparison of I_{Max} and Experimental Ionization Potentials

Metal	Assumed r (Å)	Equation I I_{Max}(ev)	Experimental I_p (ev)	Difference $I_{Max}-I_p$
Ag	1.84	7.50	7.542	-0.042
Tl	2.05	7.05	6.07	0.98
Cu	1.54	8.30	7.68	0.62
Na	1.55	8.26	5.12	3.14
K	1.91	7.33	4.318	3.012
C_s	2.25	6.70	3.87	2.83

These results may be divided into two categories. The first group consists of those metals with the difference, $I_{Max}-I_p$, less than one electron volt. In this group we have Ag, Tl, and Cu.

The second group contains the alkali metals with a difference of approximately three electron volts. The alkali metals bonds are ionic, but the metals of the first group may form bonds that are ionic, mixed, or covalent depending on actual bond length. The mixed bonding mechanism is the most probable.

A similar derivation may be made for the divalent metals using their first two ionization potentials. The maximum sum of these two ionization potentials will be given by:

$$I_{Max} = 2E_A - E_c + 2E_{Covalent}$$

A factor, two, occurs since there are two azide ions and two covalent bonds. The coulombic energy must be calculated for a given configuration. Calculations based on a linear molecule and the doubly ionized atomic radii yield the following results.

Metal	Calculated	Experimental Sum	Difference
	I_{Max} (ev)	$I_{p1} + I_{p2}$	I_{Max} - Sum
Ba	20.7	15.14	5.56
Sr	22.6	16.55	6.05
Ca	24.4	17.91	6.49
Mg	29.1	21.57	7.53
Pb	21.9	22.34	- .44
Cd	24.6	25.8	-1.2
Hg	23.1	29.08	-5.98

If the ionization potential were considered as the only criterion for the nature of the bond, Mg and Pb would be expected to have much the same bond characteristics. However, the results in the last column indicate that the bonding should be different.

The above calculations were made on the basis of the radii of doubly ionized atoms. For PbN_6 and SrN_6 the metal nitrogen bond length was taken to be 1.78Å and 1.7Å, respectively. For each of these azides the bond length in crystals is approximately 2.6Å (6-8). This new length corresponds to a value of 16.6 ev for I_{Max}. For PbN_6 this means that the molecule becomes more covalent. For SrN_6 the trend is toward a more covalent bond also, but since SrN_6 was originally considered to be an ionic bond, the actual bond has a mixed ionic and covalent character.

Bond Energy of HN_3: The molecular orbitals used in the calculation of the energy of N_3 leave an unpaired electron in a Π_g orbital on the end nitrogen atoms. This orbital may be mixed with the 1s orbital of the hydrogen. The coefficients of the molecular orbitals may be varied to give a maximum bond energy which is 290.7 kcal/mole, when the experimental values are used for bond lengths.

Another mode of bonding the hydrogen to the azide is to assume a localized covalent bond to one of the end nitrogen atoms. For the experimental bond lengths this yields a value 18.1 kcal/mole more stable than the delocalized model. Since the localized bonding is more stable, it will be taken as the electronic configuration for HN_3. Then the estimates of the individual bond energies in kcal/mole are:

$H - N_3$	82.3
$HN - N_2$	73.9
$HN_2 - N$	152.6

It should be pointed out that these bond energies are not necessarily the energy required to separate the molecule into two parts at the given bond. These energies are the calculated values for the given bond. In those cases where electrons are shared between two non-adjacent atoms, the energy is assigned to adjacent atom bonds.

The total bond energy of 308.8 kcal/mole is to be compared to the value given by Gray[3] of 318 kcal/mole. The calculations are based on the formation of a gas molecule, but at 25°C HN_3 is a liquid. If it is assumed that Trouton's rule[9] may be used for HN_3, then the heat of vaporization at the melting point, 37°C, is 6.8 kcal/mole. Gunther, et al,[10] measured the heat of vaporization at 12.4°C to be 7.3 kcal/mole. This correction brings the calculated value to within one percent of Gray's value.

Bond Energy of PbN_6: The bond energy calculations for lead azide do not permit a clear choice between the localized and the delocalized models for the covalent bond. The ambiguity in bonding is not present in the calculation for HN_3. Two characteristics of PbN_6 which differ from HN_3 may account for this ambiguity. The first difference is a longer bond between the metal and the azide. The second difference is the coulombic repulsion of charges on the two azide groups of PbN_6.

Consider the longer bond length between the metal and the azide. For the calculations of bond energy this length for PbN_6 was taken as 2.6Å. This value is slightly longer than the shortest value for the bond length of 2.48Å measured by Glen.[6] However, it is considerably shorter than his longest value of 2.97Å. When a localized bond is changed to a delocalized bond, the decrease in bond strength of the metal-azide bond is proportional to the original localized bond strength. Since the delocalization of the electrons strengthens the azide bonds independent of the metal-azide bond length, total bonding of the molecule by delocalization may increase even though the metal-azide bonds are weaker. The long bond of PbN_6 favors the delocalization of the bonding electrons over the entire azide group.

Coulombic repulsion between the two azide groups must be considered for both the localized and the delocalized models of

PbN_6. Such coulombic repulsion does not have to be considered for HN_3 since there is only one azide group per molecule. In the delocalized model the repulsion between groups is offset by coulombic attraction within the azide groups. The localized model does not have this additional coulombic attraction within the azide groups. Thus the total strength of the bonds within the azide will be stronger for the delocalized model.

The energy calculations for both models of PbN_6 yield 538 kcal/mole which is equivalent to a heat of formation of 137 kcal/mole. Since the calculations are based on the formation of a gas molecule, it is not surprising that the experimental heat of formation for crystalline PbN_6, 115 kcal/mole,[3] is less than the calculated value. This difference of 22 kcal/mole represents the additional bonding that results from crystallization.

Although both models yield the same total bond energy, the distribution of energy between the bonds is different as shown in Table V.

TABLE V
Bond Energies of PbN_6 (kcal/mole)

Bond	Localized Model	Delocalized Model
Pb - N_3	42.4	26.6
PbN - N_2	73.9	121.2
PbN_2 - N	152.6	121.2

In the calculation of bond energy for PbN_6, the 1/2 hν terms have been assumed to be the same as the similar terms in N_3^-. This accounts for 8 of the 15 vibrations. Since the Pb-N_3 bond is weak, these vibrations should be of low energy and weaken the bond only slightly. However, it should be remembered that if $h\nu < kT$ then each vibration will have an energy, RT, which will weaken the bonding. At 300°K this weakens each Pb-N_3 bond by 2.1 kcal/mole.

Choi and Boutin[8] have studied the crystal structure of alpha lead azide. They found that the azide groups are linear. Some of the groups are symmetric while others are asymmetric. This agrees with the fact that both models for PbN_6 give the same bonding energy. In each case the azide is linear, but only for the delocalized model of bonding will the azide group be symmetric.

The amount ionic bonding in PbN_6 is indicated by the amount charge moved from the lead to the azides. For the localized model this is 0.53 electrons and for the delocalized model it is 0.64 electrons. Since 2 electrons must move to be fully ionic, it is clear that PbN_6 has mixed ionic and covalent bonding.

Decomposition Kinetics: The covalent bonding of HN_3 and PbN_6 discussed in the preceeding sections make it possible to propose

7

mechanisms for the decomposition of these azides. It is clear that the weakest bond for HN_3 is the weaker of the two bonds in the azide group whereas for PbN_6 the bonding of the azide to the lead is the weakest bond. The decomposition of these azides should proceed by the rupture of these weak bonds.

For HN_3 in the liquid state the reaction $HN_3 \rightarrow HN + N_2$ is endothermic by 9 kcal/mole and for the gas phase it is endothermic by 2 kcal/mole. These energies represent the difference between the energies of the ground states of the molecules. It should not be concluded that this is the activation energy for the reaction. Although the reaction in the gas phase is a forbidden combination, in a liquid the forbidden condition may be relaxed. The activation energy for the unimolecular reaction is that energy necessary to break the bond plus the endothermicity or 83 kcal/mole.

A bimolecular reaction will proceed with a lower activation energy. An estimate[11] of this energy is 39 kcal/mole for the gas phase and 46 kcal/mole for the liquid. Gray and Waddington[12] studied the spontaneous detonation of HN_3 as a function of temperature and pressure. An analysis[13] of this data yielded an activation energy of 37 kcal/mole for an assumed bimolecular reaction process.

For PbN_6 the reaction will occur on a surface. Since this is not a gas phase reaction, the energy of sublimation must be included in any energy calculations. The weak bond occurs between the lead and the azide. Although this is the bond which first breaks in a bimolecular reaction, at least two azide radicals will continue to react forming three nitrogen molecules and releasing sufficient energy to make the total reaction exothermic. The estimated activation energy for a bimolecular reaction is 15 kcal/mole when the delocalized bonding model is assumed, and 23 kcal/mole when the localized bonding model is assumed. Baum[14] found an activation energy of 26 kcal/mole for thermal decomposition of PbN_6. Although this energy seems to be in agreement with the localized bonding model, a recent theory[15] discussing the initiation of PbN_6 shows that an activation energy of about 18 kcal/mole can yield under certain conditions an apparent activation energy of 26 kcal/mole.

Summary: The algorithm, described in the appendix, for calculating bond energies has been used successfully to calculate bond energies for the azides. Not only can the total bonding of the molecule be calculated, but the individual bonds within the molecule may be estimated. These individual bond energies provide a basis for estimating the activation energy for decomposition kinetics. Two mechanisms of reactions in the azides are illustrated by HN_3 and PbN_6. In the molecule of HN_3 the weak bond is in the azide group, but for PbN_6 the weak bond is the metal-azide bond.

APPENDIX A

ALGORITHM FOR CALCULATING BOND ENERGIES

In the "atoms in molecules" approach to molecular bonding, Moffitt[16] used the atomic electronic energy as the major contribution to the total molecular energy. Ferreira[17] used somewhat the same approach, but considered only the contributions to the molecular bonding energy introduced by the ionization state of the atoms. He assumed that the energy was a continuous and differentiable function of the charge on the atom.

In this paper we follow this assumption, but the function will be derived from energy terms similar to those used by Slater[1] with constants derived from experimental ionization potentials.

$$E_{sj} = -(N + q_j)(A - Bq_j)^2 + NA^2$$

N = Number of valence electrons in the atom

q_j = Excess number of electrons on the atom

A, B = Constants to be determined

The molecular orbitals for the molecule have the form

$$\psi_i = \sum_j C_{ij} \phi_{jk}$$

C_{ij}, constants

ϕ_{jk} an atomic orbital on j'th atom.

It will be assumed that the orbitals are normalized such that

$$\sum_j C_{ij}^2 = 1.$$

Then the number of electrons on an atom is given by

$$n_j = \sum_i C_{ij}^2$$

and the number of excess electrons is given by $q_j = n_j - Z_j$ where

Z_j is the nuclear charge.

The molecular bonding is then[18]

$$E_B = \sum_j E_{sj} + \frac{1}{2} \sum_j \sum_{l \neq j} \frac{q_j q_l}{R_{jl}} + E_C + \frac{1}{2} \sum_s h\nu_s$$

where:

$$E_C = -\frac{1}{4} \sum_i \sum_j \sum_{l \neq j} \frac{C_{ij}^2 C_{jl}^2 + \sum_k{}' C_{il} C_{kl} C_{ij} C_{kj}}{R_{il}}$$

$\sum_j$, $\sum_l$ are sums over the atoms

$\sum_i$ is a sum over the valence electrons

$\sum_k{}'$ is a sum over electrons with same spin as "i" and filling a molecular orbital utilizing the same atomic orbitals as "i"

V_s is a molecular vibration

R_{jl} is the experimental equilibrium distance between atoms j and 1.

It has been found that E_B gives values which are in good agreement with experimental bond energies. When E_B is calculated for a molecule and also for its ionized form, the difference between the two energies gives the ionization potential.

REFERENCES

1. Slater, J. C., Phys. Rev., 36, 57 (1930)

2. Evans, B. L., and Yoffe, A. D., Proc. Roy. Soc., 238A, 568 (1457)

3. Gray, P., Quarterly Reviews, 17, 441 (1963)

4. Bonnemay, A. and Dandel, M. R., Comp. Rend., 230, 2300 (1950)

5. Franklin, J. L., Dibler, V. H., Reese, R. M., and Krauss, M., J. Amer. Chem. Soc., 80, 298 (1958)

6. Glen, G. L., J. Amer. Chem. Soc., 85, 3892 (1963)

7. Rossenwassen, H., Report 1551-TR, USAERDL, Fort Belvoir, Va. (1958)

8. Choi, C. S., Bontin, H. P., Technical Report 3572, Picatinny Arsenal (1967)

9. The Condensed Chemical Dictionary, 7th Edition, Reinhold Publishing Corp., New York (1966) Article: "Trouton's Rule"

10. Gunther, Meyer, and Muller-Skjold, Z. Physik. Chem., A175, 154 (1935)

11. Eyring, H. Private Communication on Method of Estimating Activation Energies

12. Gray, P., and Waddington, T. C., Nature, 179, 576 (1957)

13. McMillan, R. C., Unpublished Analysis

14. Baum, F. A., Stanyukovich, K. P., and Shekhter, B. I., Physics of an Explosion (1951), Russian Book

15. Mueller, H. J., and McMillan, R. C., To be published

16. Moffitt, W., Proc. Roy. Soc., 210A, 245 (1951)

17. Ferriera, R., J. Phys. Chem., 68, 2240 (1964)

18. McMillan, R. C., Proceedings of the Fifteenth Basic Research Laboratory Contractor's Conference and Symposium (1966). The derivation is discussed in more detail in this report.

A FERROELECTRIC GAMMA-RAY SPECTROMETER

JOHN H. MC NEILLY
US ARMY NUCLEAR DEFENSE LABORATORY
EDGEWOOD ARSENAL, MARYLAND
DONALD D. GLOWER

I. Introduction

One of the highest priority requirements in the field of initial radiation measurements is the development of instrumentation to measure the initial gamma-ray spectrum. The need for spectral measurements is paramount for the evaluation and correlation of various systems response. Due to the existence of extremely high dose rates and intensities (such as those encountered during underground weapon tests), previous attempts to make such measurements have met with limited success. Instruments with the capability of operating in such extreme environments are encumbered by drawbacks such as large masses that negate their usefulness.

However, recent developments in the use of ferroelectric materials as radiation detectors have made them prime candidates for the solution of this problem. Specifically, ferroelectric radiation detectors exhibit such desirable characteristics as dose-rate independence, simplicity, small size, ruggedness, ability to withstand very high total doses, and, except for readout, they do not require auxiliary electronic equipment such as amplifiers and power supplies.

II. Ferroelectric Gamma-Ray Detector

Many dielectric materials, of which barium titanate and lead zirconate titanate are examples, exhibit a remarkable set of ferroelectric properties, where at a certain critical temperature, the dielectric constant becomes very large. Below the critical temperature, the material can retain a permanent electric polarization. The internal polarization is due to a relative displacement of positive and negative ions within the crystal which is caused by a slight elongation and deformation of the unit cell. For example, Figure 1 shows the polarized state of barium titanate with oxygen atoms displaced along the (100) direction and the barium and titanium displaced in the opposite direction. Such materials exhibit a polarizability

far greater than that in the normal dielectric. Furthermore, once the material is polarized in a particular (100) direction, it retains this polarized condition indefinitely, provided the crystal remains below the critical temperature. Above the critical temperature, called the Curie temperature, the crystal loses almost all of its remnant polarization. In their polarized state, ferroelectric materials also exhibit strong piezoelectric and pyroelectric properties.

The ferroelectric material used throughout this investigation is ceramic lead zirconate titanate [$Pb\ (Zr_{0.65}Ti_{0.35})\ O_3$ + 1 wt% Nb_2O_5] which for convenience shall be referred to as PZT. The crystal structure of PZT is tetragonal below the critical temperature with lattice constants of approximately 4.0 Å and a Curie temperature of about 365° C. The remnant polarization of this material is about 36 $\mu c\ cm^{-2}$ compared with that of about 7 $\mu c\ cm^{-2}$ for barium titanate.

The radiation detector consists of a suitable geometry of ferroelectric material with parallel electroded surfaces. The material has been "poled" so that its internal electric dipoles are oriented in a common direction perpendicular to the electrode faces; the result is a permanent remnant polarization internal to the crystal. With the ferroelectric material in this condition, an incident radiation pulse causes charge to flow into an external circuit, Figure 2.

The application of the ferroelectric device to radiation detection and dosimetry has been studied by Hester and Glower at Sandia, who subsequently worked out the theoretical details of the charge-release phenomenon. (1-4) These investigators have shown that the charge-release process is a surface pyroelectric effect. Radiation absorbed in the detector produces a temperature change, and even though the temperature change may be very small, the stored charge flows onto the electrode surface. This is due to the strong pyroelectric effect of such material. Thus, for a ferroelectric material, the equation describing the charge release is

$$\frac{1}{A}\frac{dQ}{dt} = K\frac{d\gamma}{dt} \qquad (1)$$

where A is the area of the electroded faces of the detector and $\frac{d\gamma}{dt}$ is the gamma-ray dose rate. The constant, K, is physically equal to the pyroelectric coefficient divided by the specific heat of the material and, for thin samples of PZT, K has been experimentally measured to be about 0.60 $pC\ cm^{-2}Rad^{-1}(PZT)$.

Experiments performed on irradiated samples of PZT of equal area have shown that the magnitude of the charge output remains constant for sample thicknesses as small as 0.3 mm. Therefore, it is concluded that the total charge output depends not upon the total energy absorbed in the entire crystal but only upon the energy absorbed in some thin layer on the surface of the detector. Charge liberated in the center of the crystal probably would not reach the

electrode because of the very low material conductivity. On this basis, it has been postulated that charge is released onto the crystal surface from only a thin layer near the surface, the charge being due to the energy absorbed in this layer. The thickness of the charge layer in a polarized ferroelectric has been previously estimated to be in the range from about 10^{-4} to 10^{-6} cm. This layer is called the active charge layer, and it has an effective thickness labeled X_0.

A ferroelectric detector may be constructed to produce any desired response time when inserted into its measurement circuit. If the terminating resistance of the measurement circuit is fixed, then an appropriate choice of detector area and thickness can be made to provide the required time response.

An indication of the relative neutron response of the detector was experimentally obtained by an exposure at the Sandia Pulse Reactor Facility (SPRF). In this experiment, various measurements were made with lead and polyethylene between the detector and radiation source. These data, shown in Figure 3, indicate that only about 2 percent of the detector output was due to the neutrons for this particular environment.

Thus, the ferroelectric radiation detector, a rate-measuring device capable of measuring very high dose rates and total doses can be easily fabricated to provide any desired time response and is quite insensitive to neutrons. Therefore, this detector is an ideal candidate for the construction of a gamma-ray spectrometer.

III. The Energy Spectrometer

The spectrometer system itself employs an array of ferroelectric detectors positioned behind a corresponding set of metal filters, as shown in Figure 4. The pulsed radiation source shown at the left emits a spectrum of photons that passes through the filters and subsequently produces a charge pulse at each detector. If each metal shield has a unique thickness, then the magnitude of the charge generated at each detector will also be unique. The result is the measurement of the filtration curve, that is, the charge output of the detector as a function of shield thickness. When information relative to the spectrum to be measured is available, then an experimenter may choose the filters in such a manner that he can achieve the maximum sensitivity in each channel. Let the filtration curve be designated $Q(x)$, where x is the filter material thickness; and let the incident radiation spectrum be $N(E)$, where $N(E)dE$ gives the number of photons per unit area emitted by a plane monodirectional source which has energies within the interval E to $E+dE$.

To develop the equation which related the response of any given detector in the spectrometer system to the radiation spectrum $N(E)$, consider the equation

$$\int_0^{E_{max}} E\,N(E)dE = \text{the energy emitted per unit area by the radiation source, which is incident upon a given filter.} \quad (2)$$

Then refer to Figure 1,

$$\int_0^{E_{max}} e^{-\mu(E)x}\,E\,N(E)dE = \text{the energy which is transmitted through a filter of thickness x and absorption coefficient } \mu(E), \text{ per unit area.} \quad (3)$$

Therefore, Figure 3 describes the total energy incident upon the surface of the ferroelectric detector. The charge released when this energy is absorbed in the detector is given by

$$Q(x) = AK \int_0^{E_{max}} (1-e^{-\sigma(E)X_D})e^{-\mu(E)x}\,E\,N(E)dE, \quad (4)$$

where A is the detector area, K is the energy-to-charge conversion factor previously discussed, $\sigma(E)$ is the energy absorption coefficient of the ferroelectric material, and X_D is an effective thickness related to the surface phenomenon of the pyroelectric effect. This equation may be written more simply as

$$Q(x) = \int_0^{E_{max}} G(x,E)\,N(E)\,dE, \quad (5)$$

where G(x,E) is a function equal to the known quantities on the right hand side of Figure 4. This figure is the well known Fredholm equation of the first kind, involving the unknown N(E), where G(x,E) is the Kernel and Q(x) is the experimentally measured filtration curve. The solution of the integral equation is the classical unfolding problem.

IV. The Unfolding Technique

Since no explicit solution of the equation for any arbitrary functions Q(x) and G(x,E) exists, approximation methods are usually employed. The approach taken here was suggested by Gold of the Argonne National Laboratory (5,6). This particular numerical technique replaces the integral equation with a set of approximate linear simultaneous equations, which are then put in the appropriate matrix form for computer analysis. The solution for the unknown spectrum then proceeds by applying knowledge of the physics of the measurement such as the fact that the spectrum can never become negative. Since Gold's technique is an iterative one, this is accomplished by initiating the iterative process with an assumed spectrum that is positive. Gold's method also guarantees a positive solution

for a positive value of the measured parameter, Q, and for a response matrix that is positive definite in character, which is the case under consideration. The particular calculations involved in the unfolding employ the response matrix in the forward direction only, thus ensuring that convergence to the appropriate solution, if it occurs at all, must proceed within the positive range.

The iterative unfolding method has produced meaningful solutions in this application and has been incorporated into a computer program called GISMO (for Gold's Iteration Scheme for Matrix Operations).

V. Experimental Results

Experimental verification of the energy spectrometer concept and design was accomplished by use of the Martin Marietta 0.6 Mvp flash x-ray machine and a single ferroelectric radiation detector which was connected directly to an oscilloscope. The detector and shield configuration illustrated in Figure 5 was rigidly mounted in a fixed position relative to the flash x-ray target. The x-ray pulse was directly incident upon the sensitive area of the ferroelectric radiation detector.

The detector was shielded in a prearranged manner so that each change of shield effectively increased the number of energy channels. This shielding arrangement was necessary since the radiation beam from the flash x-ray encompassed a relatively small area and as a consequence, it was not practical to attempt simultaneous measurements of several energy channels. Data points were recorded for total detector response to the radiation from each flash x-ray pulse which traversed the shield. The polyethylene shield was designed to stop photo and Compton electrons that were ejected from the metal shield.

The experimental procedure was designed to minimize experimental error due to fluctuations in the flash x-ray machine output. The procedure consisted of recording data points for each shield thickness of tin, copper, and lead and then repeating each measurement. Sufficient flash x-ray pulses were received to produce five data points for each shield configuration.

The data are shown graphically in Figure 6. The data points show a scatter of approximately ±10 percent in the output of the radiation detector when the five data points for each configuration are considered. The actual scatter as measured is shown by the error bars. The actual data in this figure were reduced by the GISMO computer program, and the energy spectrum shown in Figure 7 represents the spectrum for an average radiation pulse as measured over many pulses.

The experiment demonstrates that an energy spectrum such as shown in Figure 7 can be obtained for a single burst from a flash x-ray machine if the radiation covers sufficient area to expose simultaneously the various detector-shield combinations. The fact that approximately seventy-five pulses were used requires that the spectrum as measured represent some kind of average. Extreme changes in spectra for different pulses are not anticipated, however, so it is most likely that the spectrum as measured is quite representative.

A comparison of the measured and normalized energy spectrum with the spectrum measured by Bouchard (7) at Sandia and calculated for a bremsstrahlung response reveals that the general shape is as expected. These spectra are graphed in Figure 7. The bremsstrahlung spectrum as drawn is not attenuated through filters, as was the flash x-ray exposure. All flash x-ray machines have metal shields between the electron target and the external experimental area since noise attenuation is particularly important. The difference in the low-energy tail can be ascribed at least in part to the different physical characteristics of these flash x-ray machines. The higher energy maximum of about 270 keV for the Martin machine compared with 225 keV for the Bouchard spectrum can also be explained in a similar manner. The high-energy tail measured by the ferroelectric spectrometer system reveals a slight "knee" at approximately 425 keV. The Bouchard and bremsstrahlung spectra do not show this.

The data of Southward (8) at Berkley were analyzed and found to contain a structure similar to that measured for the Martin Company flash x-ray. Southward's integrated energy spectrum was plotted as a differential spectrum and compared with the Martin spectrum. The shapes of the two spectra at energies greater than the peak were similar in that an inflection point or "knee" is present in both. An energy spectrometer containing more energy channels should give better detail of the structure in this "knee."

In the analysis of the experimental data, it became apparent that one could analyze the time dependence of the flash x-ray spectrum. If we assume that the spectrum changes with time, this change may be described by rewriting Equation 5 as

$$Q(x,t) = \int_0^{E_{max}} G(x,E)\, N\,(E,t)dE. \tag{6}$$

If the flash x-ray pulse is divided into time intervals of Δt, then the spectrum for each interval can be found by using the same unfolding technique that was used to solve for the time-integrated spectrum. In this case, the response of the spectrometer system for the time interval Δt is given by

$$Q(x,\Delta t) = \int_0^{E_{max}} G(x,E)\, N(E,\Delta t)dE, \tag{7}$$

where $Q(x,\Delta t)$ and $N(E,\Delta t)$ are the charge and spectrum functions integrated over the time interval Δt.

Analysis of the data proceeded by dividing the charge outputs of the spectrometer system into two time intervals. The spectrum for each of the two time intervals was unfolded and the results are presented in Figure 8. Here, $\alpha(\Delta t_1)$ is the time average spectrum over the first 30 ns of the radiation pulse and $\alpha(\Delta t_2)$ is the spectrum over the remaining 50 ns. The third spectrum is the total spectrum which was measured for the entire radiation pulse. These results indicate a softening of the flash x-ray as a function of time.

VI. Summary

The effort to date has produced a radiation detector that has many characteristics necessary to measure the gamma spectra of pulsed radiation sources. In addition, it has been demonstrated that a spectrum from a single pulse can be obtained.

REFERENCES

1. Hester, D. L., and Glower, D. D.; "The Charge Release Mechanism for Ferroelectric Radiation Detectors," Nuclear Applications, 2, 41 (1966).

2. Hester, D. L., Glower, D. D., and Overton, L. J.; "Use of Ferroelectrics for Gamma-Ray Detection," Trans. Prof. Tech. Group on Nuc. Sci. IEEE, NS-11, 145 (1964).

3. Glower, D. D., and Hester, D. L.; "Steady State Gamma Ray Detection Using Ferroelectrics," Trans. Am. Nuc. Soc., 8, 61 (1965).

4. Glower, D. D., and Warnke, D. F.; Nuclear Rad. Eff. Conf., IEEE Summer General Meeting, CP-63-1137, June (1963).

5. Gold, R., and Scofield, N. E.; "Iterative Solution for the Matrix Representation of Detection Systems," Bull. Am. Phys. Soc., 2, 276 (1960).

6. Gold, R.; "An Iterative Unfolding Method for Response Matrices," ANL-6984 (1964).

7. Bouchard, G. H.; "Measurement of Bremsstrahlung Dose and Spectrum from a 600 KVP Pulsed X-Ray Generator Using Photographic Film," Sandia Corporation Reprint, June (1962).

8. Southward, H. D.; "Measurement of the Continuous Spectrum of a Flash X-Ray Machine," University of New Mexico Engineering Station Report, September (1964).

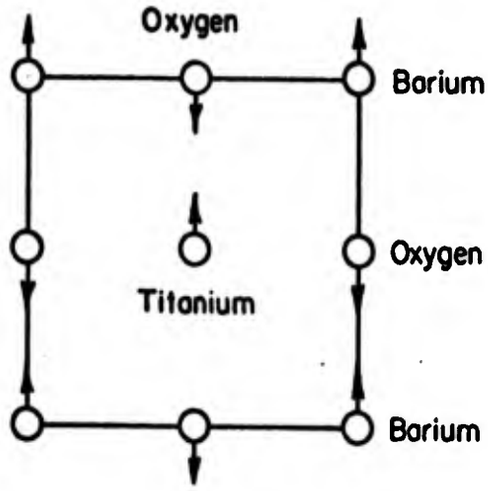


Figure 1. A two-dimensional representation of the polarized state in a Perovskite-type material.

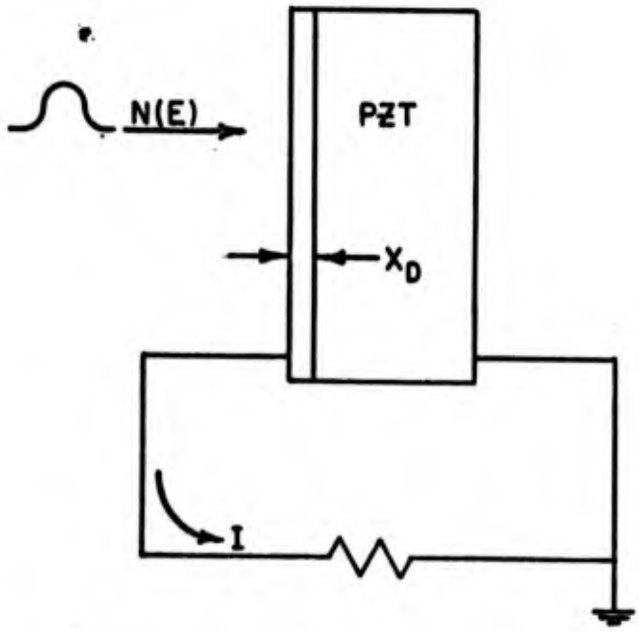


Figure 2. A schematic representation of a ferroelectric radiation detector.

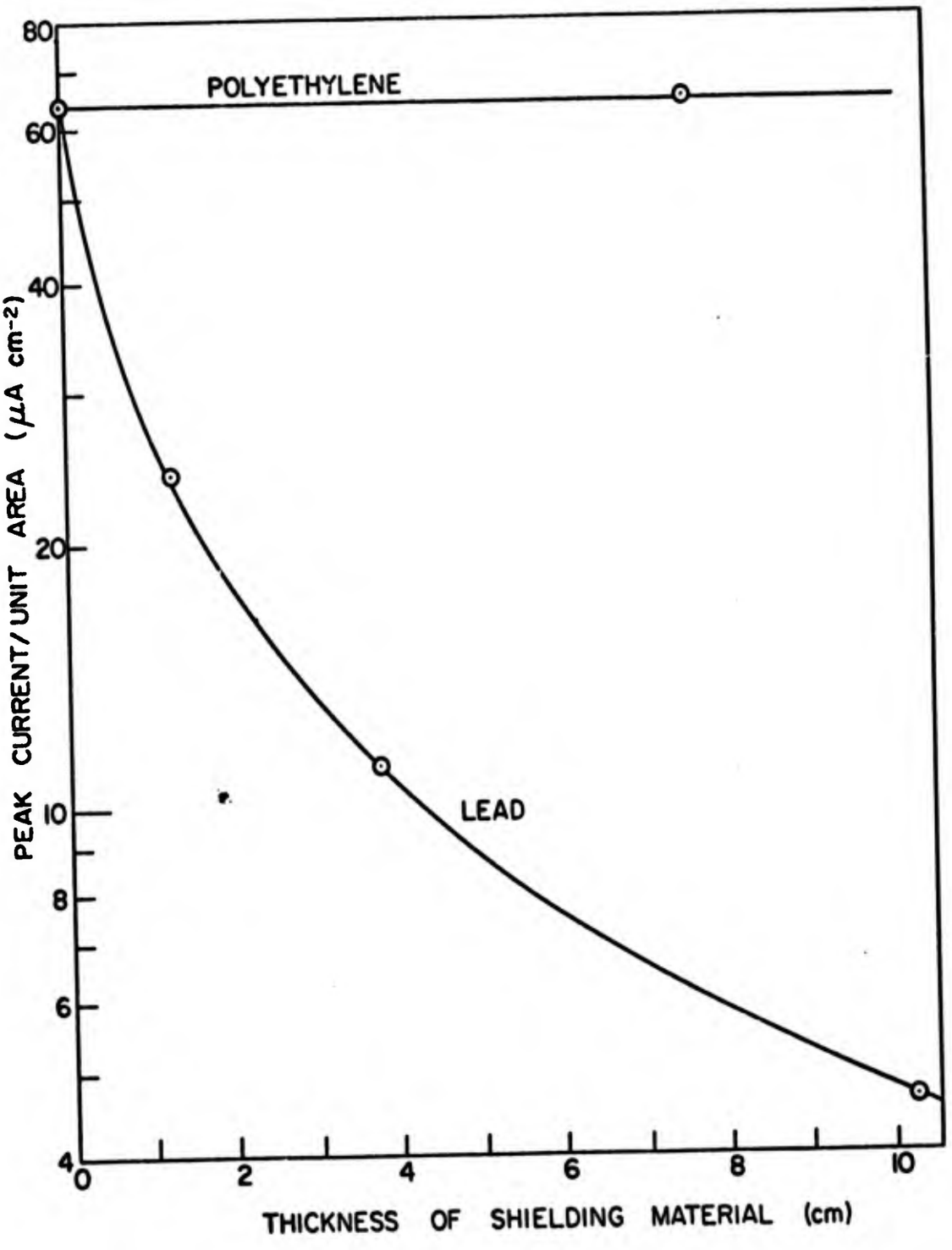


Figure 3. Peak current as a function of thickness of polyethylene and lead shielding for the Sandia Pulse Reactor.

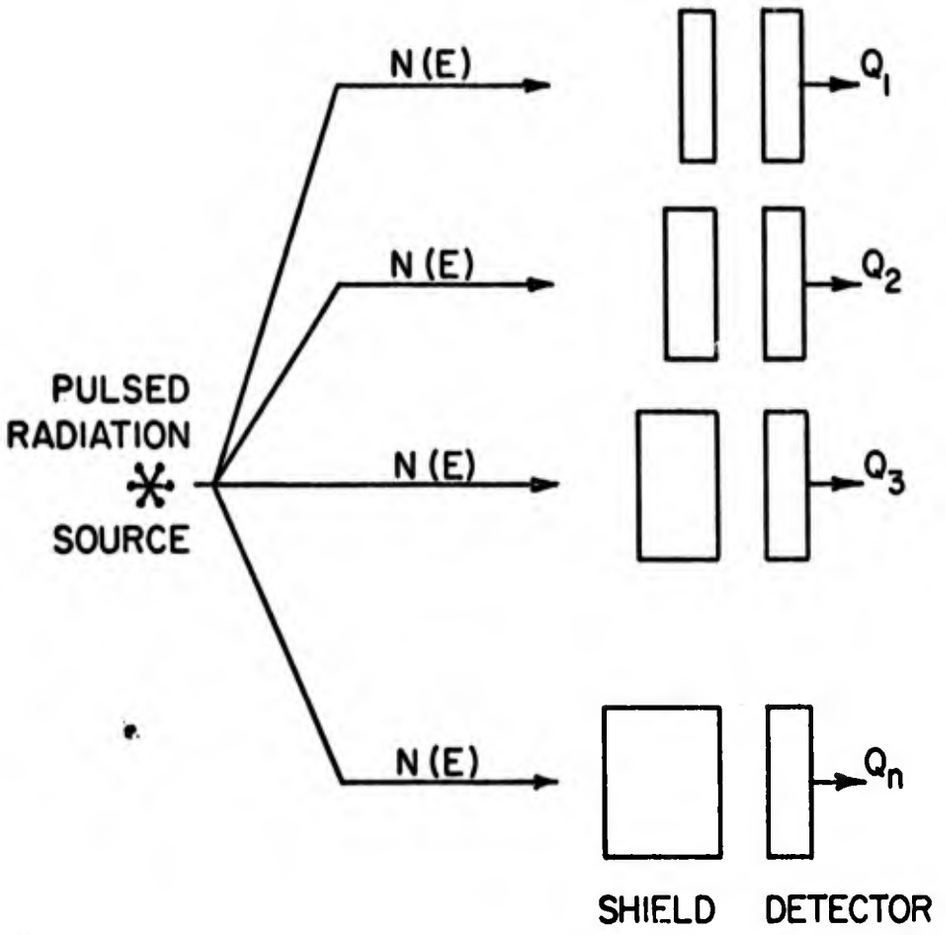


Figure 4. A schematic of the spectrometer system showing the source, shield system, detectors and charge output.

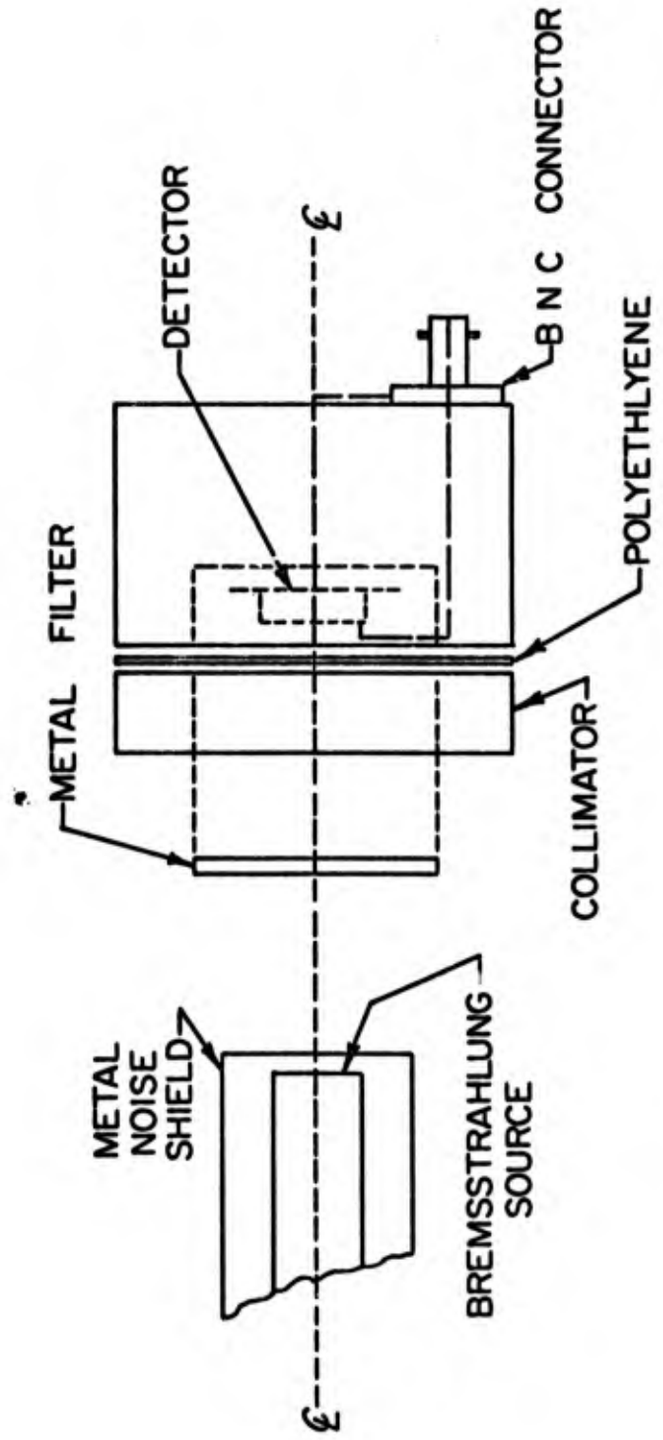


Figure 5. Experimental arrangement of detector and filter configuration at the Martin-Marietta X-ray Facility.

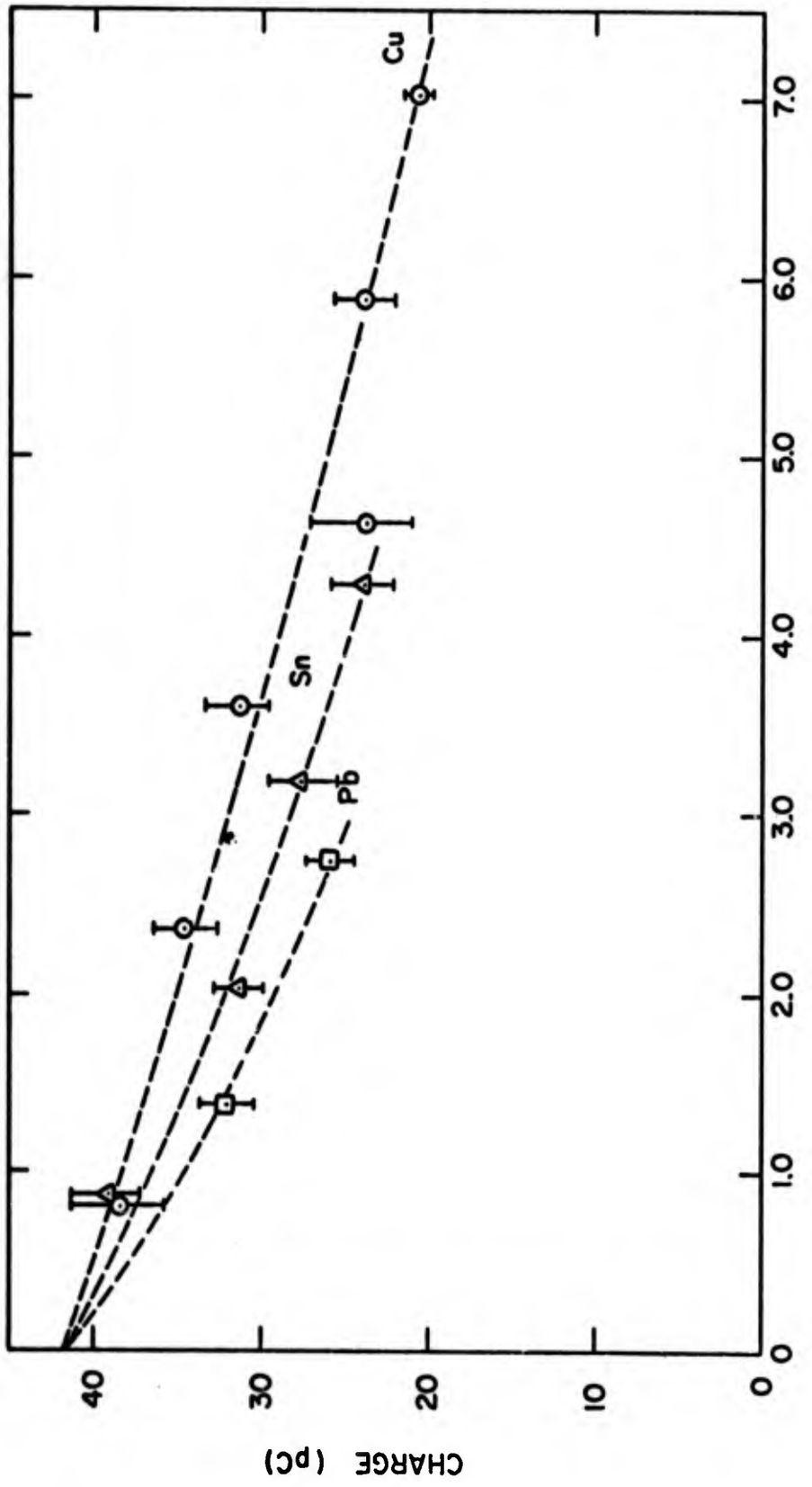


Figure 6. Charge filtration curve in picocoulombs as a function of areal density.

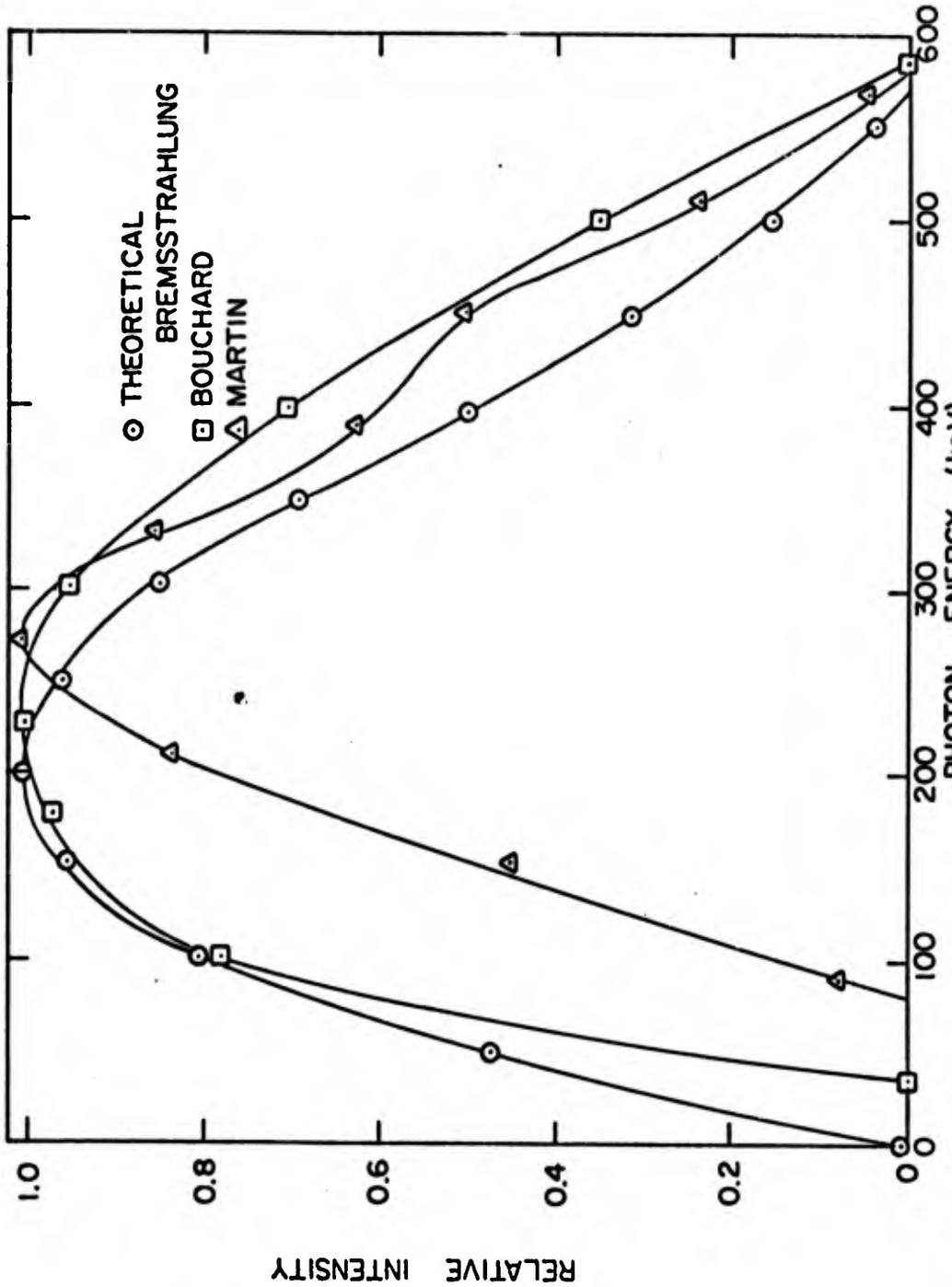


Figure 7. Comparison of theoretical bremsstrahlung spectrum, flash x-ray spectrum as measured by Bouchard and the Martin Co. flash x-ray spectrum.

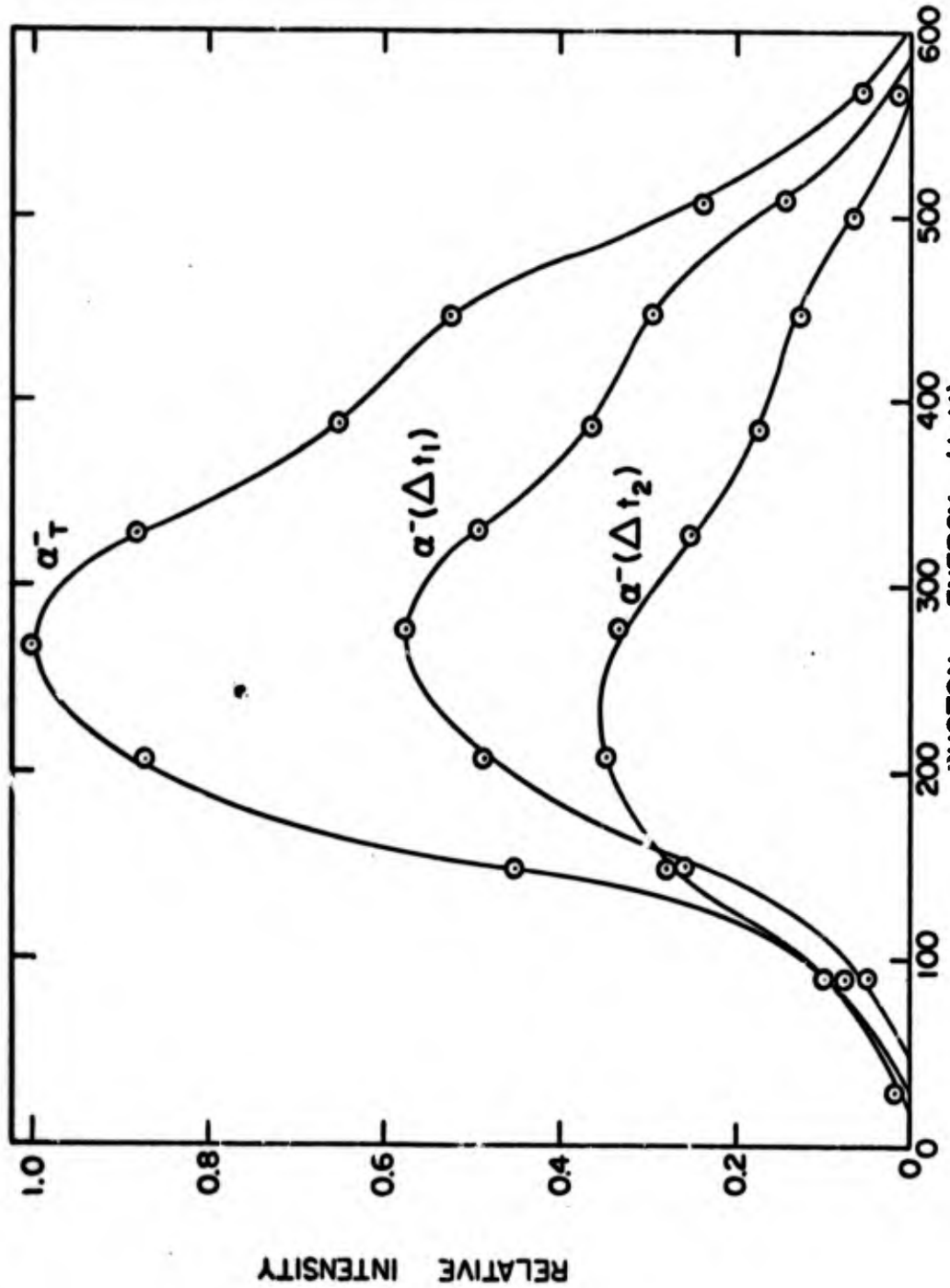


Figure 8. A graph showing the time dependent nature of the Martin Co. flash x-ray spectrum.

GASEOUS HALIDE LASERS

JAMES A. MERRITT and GEORGE J. DEZENBERG
U. S. ARMY MISSILE COMMAND
REDSTONE ARSENAL, ALABAMA

The purpose of this paper is to review the published results of gaseous halide lasers and to present the work done in this area in the Physical Sciences Laboratory of the U. S. Army Missile Command, Redstone Arsenal, Alabama.

Exothermic interhalogen reactions of the type

$$A + BC \rightarrow AB^* + C + \text{Energy}$$

are capable of producing population inversion in AB*. Some typical reactions of this type are

(i) $Cl + HI \rightarrow HCl^* + I + 32$ kcal/mole
(ii) $Cl + HBr \rightarrow HCl^* + Br + 16$ kcal/mole
(iii) $Br + HI \rightarrow HBr^* + I + 16$ kcal/mole
(iv) $F + HCl \rightarrow HF^* + Cl + 32$ kcal/mole
(v) $F + HBr \rightarrow HF^* + Br + 47$ kcal/mole
(vi) $F + HI \rightarrow HF^* + I + 63$ kcal/mole
(vii) $H + F_2 \rightarrow HF^* + H + 98$ kcal/mole

In the reactions producing population inversion, the reaction product AB* is formed directly in the vibrationally excited state. This is to be contrasted with existing gas lasers where population inversion is produced by using collisions of the second kind, disassociative collisions, etc. It seems possible, using an exothermic reaction to produce population inversion, that a laser could be made requiring no additional excitation source.

The potential power capabilities and efficiencies of chemical lasers of this type are much greater than any presently

operating laser. J. C. Polanyi (1) calculated that, using the above systems, in a 1-liter cavity at 100 to 1000 mm Hg pressure, a laser pulse of 10^{-6} sec. duration, containing 10^3 joules of energy, to 10^{-7} sec. duration, containing 10^4 joules of energy, could be produced. For this calculation it was assumed that A and BC reacted suddenly with an exothermic energy of reaction of 50 kcal/mole and that this energy preferentially excited one vibrational level of AB*. For the assumed conditions, the calculation establishes an upper limit for the peak output power to be 10^9 to 10^{11} watts. J. C. Polanyi (1) also calculated that with a pressure of 0.1 to 1 mm of Hg in a 1-liter cavity, 10^2 to 10^3 watts continuous output power is possible.

A pulsed HCl laser using an exothermic chemical reaction was first reported by J. V. V. Kasper and G. C. Pimentel (2). In their experiment a 60-cm, 14 mm i. d. quartz laser tube was filled with a 1:2 mixture of Cl_2 to H_2. This mixture was then exposed to the flash from a xenon-filled quartz flash tube to produce the following reactions:

$$Cl_2 + h\nu \rightarrow 2Cl$$

$$Cl + H_2 \rightarrow HCl + H$$

$$H + Cl_2 \rightarrow HCl^* + Cl$$

Lasing occurred on several lines in the V(2 → 1) band of HCl*. The output wavelengths were in the 3.8 μm. region. The peak power was 10 watts, and the pulse duration was 20 μsec. or less.

Subsequently, pulsed lasing has been obtained in HCl, HF and DF using mixtures of Cl_2 - HI (3), UF_6 - H_2 (4), and UF_6 - D_2 (4). A photolysis flash was used to initiate the chemical reaction in all of the above systems. The Cl_2 - HI mixture seems to be the most promising of this group. Laser pulses with peak powers in the 10^3 watt region have been obtained by J. R. Airey (3). The output pulse is typically of μsec. duration and contains in the order of 10^{-3} joules of energy.

Electrically pulsed discharges have been used to produce lasing in the following series of diatomic molecules: HF, DF, HCl, DCl, HBr, and DBr (5-7).

K. G. Anlauf, et al. (8) have maintained, continuously, a vibrational population inversion in the molecular product of the

following three reactions:

$$(a)\ H + Cl_2 \rightarrow HCl^* + Cl$$

$$(b)\ H + Br_2 \rightarrow HBr^* + Br$$

$$(c)\ Cl + HI \rightarrow HCl^* + I$$

Continuous population inversion was obtained in reactions (a) and (c) with gains of 1.6%/meter and 5%/meter, respectively. The above results were obtained at low total pressure (approximately 10^{-3} torr) with the longitudinal walls of the vessel maintained at liquid nitrogen temperature.

In the Physical Sciences Laboratory of the U. S. Army Missile Command, pulsed lasing has been obtained using $H_2 + Cl_2$ and $UF_6 + H_2$ mixtures using photolysis to initiate the reactions. However, the emphasis has been placed on obtaining a continuous gaseous halide laser.

In light of the results of K. G. Anlauf, et al. (8), an attempt was made to construct a continuous wave HCl laser using a flowing mixture of Cl_2 and HI. The experimental apparatus consisted of a 18 mm i. d., 0. 9 m. long laser tube equipped with NaCl Brewster angle windows. Two ports were provided at one end to allow the HI and Cl_2 to be introduced into the tube. Separate inputs were used to prevent the gases from reacting before they entered the laser tube. The opposite end of the tube had one port which was connected to a liquid nitrogen cold trap and a 425 liters/min. vacuum pump. Gold coated front surface spherical reflectors, in a confocal configuration, were used to form the resonator. The diameter of the reflectors was 2. 54 cm. and their radius of curvature was 1 meter. A 2 mm diameter hole in the center of one mirror was used to couple the energy out of the cavity.

In order to disassociate the Cl_2 and initiate reaction (c), the following excitation techniques have been used: an axial D. C. discharge, and axial R. F. discharge, an axial C. W. optical source, and an off-axis microwave discharge in the Cl_2 input tube. To date, repeatable C. W. lasing has not been obtained. The best result obtained with each source is given below.

No lasing was observed when an axial 30 to 100 ma. D. C. discharge was used. An excessive amount of iodine was deposited in the tube with this method of excitation.

A 18 MHz., 200 watt transmitter was used to produce an axial R. F. discharge. C. W. lasing was obtained with this method of excitation. With a flowing mixture of 2 torr HI and 4 torr Cl_2, 0.3 watts was obtained. The output power decayed to zero in approximately 3 minutes.

A 2,450 MHz., 125 watt microwave source was used to produce a discharge in the Cl_2 input tube. Several 1 watt pulses were obtained when the pressures were 3 torr HI and 6 torr Cl_2. However, no C. W. lasing was obtained with this method of excitation.

Two 400 watt Hg. street lamps, with the outer jacket removed, were used to produce axial optical excitation. A pyrex laser tube was used in this experiment to prevent disassociation of the HI. (The maximum of the Cl_2 absorption curve occurs at 3,300 A and the maximum absorption in HI occurs at 2,500 A. Pyrex does not transmit wavelengths shorter than 3,000 A. Hence, pyrex is an ideal filter to use in conjunction with an optical source in the Cl_2 - HI system.) When only Cl_2 and HI were used only a few small pulses were obtained. However, 20 mW was obtained when 2 torr HI, 2 torr Cl_2, and 5 torr He were used.

Although repeatable C. W. lasing has not been achieved, the theoretical potential of gaseous halide lasers makes this area worth pursuing.

(1) J. C. Polanyi, Appl. Optics Supplement 2 of Chemical Lasers, pp. 109-127 (1965).

(2) J. V. V. Kasper and G. C. Pimentel, Phys. Rev. Letters 14, 352-354 (1965).

(3) J. R. Airey, IEEE J. Quantum Electronics QE-3, 208 (1967).

(4) K. L. Kompa and G. C. Pimentel, J. Chem. Phys. 47, 857-858 (1967).

(5) T. F. Deutsch, Appl. Phys. Letters 10, 234-236 (1967).

(6) T. F. Deutsch, Appl. Phys. Letters 11, 18-20 (1967).

(7) T. F. Deutsch, IEEE J. Quantum Electronics QE-3, 419-421 (1967).

(8) K. G. Anlauf, D. H. Maylotte, P. D. Pacey, and J. C. Polanyi, Phys. Letters 24A, 208-210 (1967).

LARGE NONSYMMETRIC DEFLECTIONS OF THIN SHALLOW SHELLS

JOHN F. MESCALL

ARMY MATERIALS AND MECHANICS RESEARCH CENTER
WATERTOWN, MASSACHUSETTS

INTRODUCTION

The physical problem we wish to discuss is that of the large nonsymmetric deflection of thin shells. To give some background material, a shell is considered thin when its radius of curvature is large compared to its thickness. For such a thin shell deflections are considered large when they reach the order of magnitude of the shell thickness. Mathematically speaking, the governing differential equations become nonlinear at this stage, and this has some interesting consequences including a loss of uniqueness which can give rise to instability in a physical sense - viz., buckling of the shell.

Perhaps the clearest frame work in which to describe these nonlinear effects is that of a load - deflection curve. If we plot the load, P, acting on a shell versus a corresponding deflection, δ, then linear theory predicts a straight line behavior indefinitely, as in Figure 1. On the other hand a nonlinear P-δ relationship can indicate the emergence of local maxima and such maxima define critical values of load, P_{cr}, at which a shell can buckle. In Figure 1, if the load P is incremented even slightly above the maximum on the nonlinear curve, then since the only equilibrium position available for this load is on the right hand branch, the inference is the shell will jump to this branch. Thus, for a small increase in load one sustains a large increase in deflection - i.e., the shell buckles.

Nearly all theoretical studies of this nonlinear behavior have focused on the axisymmetric case and for two very good reasons. First, because one would expect a priori that these equations would be adequate for many problem classes (e.g., concentrated loads at the apex of a sphere or uniform pressure on a spherical cap) - and secondly, the corresponding ordinary nonlinear differential equations are in themselves difficult enough to solve. However, the complete solution to these equations is now obtainable [2] and a comparison of

theory and experiment reveals that these equations are not adequate to explain the physical phenomenon for certain cases - most notably uniform pressure on a spherical cap. Axisymmetric theory is too conservative in this problem class and predicts buckling loads ranging from 25% to 75% above the best experimental data. There appear to be two avenues of approach to explain this discrepancy - first, imperfections in the shell manufacture or testing procedures (regarding which we have nothing to say here) - and secondly the possibility that in trying to find a passage from the prebuckled to the postbuckled state the physical shell chooses a path through nonsymmetric states of equilibrium. For these reasons and because there is a very wide class of problems involving nonsymmetric loads, we are led to consider the nonlinear equations governing large nonsymmetric deflections.

GOVERNING EQUATIONS

For shallow spherical cap we employ the Maguerre equations expressed in polar coordinates (x,θ) as follows:

$$\nabla^4 w = \nabla^2 \psi + (\psi'/x + \ddot{\psi}/x^2)\, w'' + (w'/x + \ddot{w}/x^2)\psi''$$

$$-2\,(\dot{\psi}'/x - \dot{\psi}/x^2)(\dot{w}'/x - \dot{w}/x^2) + 4\,p(x,\theta)$$

$$\nabla^4 \psi = -\nabla^2 w + (\dot{w}'/x - \dot{w}/x^2)^2 - (w'/x + \ddot{w}/x^2)w''$$

where $(\,)' \equiv \partial(\,)/\partial x$ $\quad (\dot{}) \equiv \partial(\,)/\partial\theta$

$$\nabla^2(\,) \equiv (\,)'' + (\,)'/x + (\ddot{}\,)/x^2$$

$$0 < x < \lambda \qquad\qquad 0 < \theta < 2\pi\ ,$$

and $p(x, \theta)$ is an applied pressure

These then are a coupled pair of 4th order nonlinear partial differential equations for the normal displacement w and the stress function ψ. Once w and ψ have been determined, all other

quantities of physical interest may be obtained by suitable differentiation or integration.

Boundary conditions are specified at the edge of the shell $x = \lambda$. For a clamped edge, the conditions on w and ψ would be

$$w(\lambda , \theta) = w'(\lambda , \theta) = o$$

$$[\psi'' - \nu\psi'/x - \nu\ddot{\psi}/x^2]_{\lambda,\theta} = o$$

$$\lambda[\psi'' - \nu\psi'/x - \nu\ddot{\psi}/x^2]' - \psi^1/\lambda - \ddot{\psi}/\lambda^2 + \nu\ddot{\psi} + 2(1+\nu)(\ddot{\psi}/x)']_{\lambda,\theta} = o$$

At $x = o$ we require the solution to meet conditions of finite stresses and displacements. These imply the following "limiting form" of the differential equations which we use at $x = o$:

$$\nabla^2\eta = 2\eta'' = \zeta + 2w''\psi'' + 4p$$

$$\nabla^2\zeta = 2\zeta'' = -\eta - (w'')^2$$

$$\eta = 2w'' \qquad \zeta = 2\psi''$$

NUMERICAL PROCEDURES

What we propose to do is to outline an effective numerical technique for the solution of these equations. Since each point on a load deflection curve such as Figure 1 requires its own solution to the nonlinear equations, it is clear that a complete load deflection curve requires that the equations be solved many times. Furthermore, a change in the geometry parameter, λ, requires a new load deflection curve, as does a change in loading conditions or boundary conditions. Thus, we desire a numerical technique which is sufficiently accurate to correlate with experiment yet sufficiently rapid to permit determination of solutions for a wide range of geometric, load and boundary parameters.

MESCALL

It is convenient to rewrite the differential equations in the form:

$$\nabla^2 \eta = \zeta + F_1(w,\psi) + 4p$$

$$\nabla^2 \zeta = -\eta + F_2(w)$$

$$\nabla^2 w = \eta$$

$$\nabla^2 \psi = \zeta$$

where

$$F_1(w,\psi) \equiv (\psi'/x + \ddot{\psi}/x^2)w'' + (w'/x + \ddot{w}/x^2)\psi'' - 2(\dot{\psi}'/x - \dot{\psi}/x^2)(\dot{w}'/x - \dot{w}/x^2)$$

$$F_2(w,\psi) \equiv (\dot{w}'/x - \dot{w}/x^2)^2 - (w'/x + \ddot{w}/x^2)w''$$

To treat the nonlinear aspect of the problem we shall use a Newtonian iterative scheme which consists in replacing the nonlinear system by a sequence of linear systems of partial differential equations. The elements of this latter sequence are determined by assuming a trial solution (w_o,ψ_o) and expanding our dependent variables in Taylor series about this solution. For example the nonlinear term $F_1(w, \psi)$ is expanded thus:

$$F_1(w,\psi) = F_1(w_o,\psi_o) + (\partial F_1/\partial w')\delta w' + (\partial F_1/\partial w'')\delta w'' + \ldots + (\partial F_1/\partial \psi')\delta\psi' + \ldots + (\partial F_1/\partial \dot{\psi}')\delta\dot{\psi}'$$

We neglect nonlinear terms in δw and $\delta\psi$. Inserting this and similar expansions into the governing equations leads to the following set of linear correctional equations for $\delta\eta$, $\delta\zeta$, δw, and $\delta\psi$:

$$\nabla^2(\delta\eta) - \delta\zeta - [(\psi'/x + \ddot{\psi}/x^2)\delta w'' + (\delta\psi'/x + \delta\ddot{\psi}/x^2)w''$$

$$+ (w'/x + \ddot{w}/x^2)\delta\psi'' + (\psi''/x)\delta w' + (\psi''/x^2)\delta\ddot{w}$$

$$- 2(\delta\dot{\psi}'/x - \delta\dot{\psi}/x^2)(\dot{w}'/x - \dot{w}/x^2)$$

$$- 2(\dot{\psi}'/x - \dot{\psi}/x^2)(\delta\dot{w}'/x - \delta\dot{w}/x^2)]$$

$$= \zeta - \nabla^2\eta + F_1(w,\psi) + 4p(x,\theta)$$

$$\nabla^2(\delta\zeta) + \delta\eta - [2(\dot{w}'/x - \dot{w}/x^2)(\delta\dot{w}'/x - \delta\dot{w}/x^2)$$

$$- (w'/x + \ddot{w}/x^2)\delta w'' - (\delta w'/x + \delta\ddot{w}/x^2)w'']$$

$$= -\nabla^2\zeta - \eta + F_2(w,\psi)$$

$$\nabla^2(\delta w) - \delta\eta = \eta - \nabla^2 w$$

$$\nabla^2(\delta\psi) - \delta\zeta = \zeta - \nabla^2\psi$$

This system looks more difficult than the original, since now the coefficients depend upon w and ψ, but in fact it is simpler because it is linear. To recapitulate, we assume we have an estimate of the starting solution: $T^o = (\eta^o, \zeta^o, w^o, \psi^o)$. (This may, e.g., be obtained from the linear equations for small values of load.) . From this we evaluate coefficients of the linear

system and the right hand sides. If we can solve these equations, we have a correction δT to T^o. With this new estimate of T, we resolve the equations for δT, and continue the iterative process until it converges or fails to converge in a specified number of trials.

There are alternative procedures one might adopt to treat the nonlinear aspect of the problem. For example, Bauer and Reiss[1] have recently considered a similar nonlinear problem for a flat rectangular plate and in their approach use an iterative procedure which amounts to successive approximation - i.e., nonlinear terms are put on the right hand side, a solution found by solving the resulting linear system. Since the coefficient matrix in this case is independent of the current estimate of the solution, they also use a direct procedure for inversion of the large system of equations which form the finite difference analogue of the linear system. The advantage of this procedure is that having obtained the inverse and stored it, successive iterates can be obtained by matrix multiplication, with the saving of a considerable amount of running time. A major requirement for this approach is of course a large memory or frequent tape transport.

On the plus side for a Newtonian iterative procedure are the facts that a smaller requirement is made on memory, and the rate of convergence is quadratic. However, the most telling circumstance seems to us to be the following. Consider the problem of a spherical cap subjected to a concentrated load at the apex. This is an axisymmetric problem governed by a set of ordinary nonlinear equations which we have discussed earlier [2]. Archer [3] was among the first to discuss this problem using nonlinear equations and finite deflections. In so doing, he used an iterative scheme which amounts to successive approximation as described above. It turns out that in this problem such a method will work only for the relatively early stages of the nonlinearity in the load - deflection curves, after which convergence stops. At this point one is uncertain whether the failure to converge is due to arrival at a maximum on a $P-\delta$ curve and thus the non-existence of a solution for higher P in the immediate vicinity of the prior solution, or due to the failure of the numerical procedure to converge. If one believes the former to be the case, then the last value of P at which a solution is found is taken to be the buckling load. In point of fact, however, the actual load - deflection curve does not reach a maximum until P is several times larger than that predicted by Archer's results. We have been able to obtain these higher solutions [2] using a Newtonian iterative scheme rather than successive approximation.

Thurston [4], another investigator in this area raised the question of whether Archer could have obtained solutions at higher P levels by taking smaller increments in P. To answer this question, we reverted to use of a successive approximation scheme and verified that this procedure broke down precisely where Archer

reports. We took very small increments of P and yet could not produce a solution at a higher value of P . Furthermore, since we had already obtained the solution at higher P levels (from the Newtonian procedure) we inserted this "starting solution" into the method of successive approximation and essentially asked it to find itself. Oddly enough, the procedure failed to converge even under these conditions. The successive "corrections" oscillated about the correct solution, but with amplitudes which were outside the prescribed limits of error. For these reasons we consider it important to use a Newtonian iterative scheme for the nonlinear aspect of the nonsymmetric problem.

To solve the linear partial differential equations which determine each component in the Newtonian iteration we use a finite difference scheme which results in a nine point star centered about each mesh point. Since we must determine four unknowns (η, ζ, w, ψ) at each point we have 4 'N' M linear algebraic equations to solve. Experience with the axisymmetric case and experimental evidence regarding asymmetric buckling patterns suggests that N and M must each be of the order of fifty. Fortunately, the coefficient matrix of this large system is very sparse, and the numerical scheme should capitalize on this property.

In general there are three techniques for solving such large systems of algebraic equations, viz., direct, iterative or step-by-step methods. Our first choice was to use an iterative scheme which amounts to a line iteration - i.e., we imporved an initial estimate of the solution along a whole meridian at a time. One advantage of this approach is that for each meridian one has a tri-diagonal set of linear equations which is compactly and efficiently solved by Gaussian elimination. Assuming a trial solution over the finite difference mesh, one corrects it meridian by meridian, comparing successive iterates at the end of a sweep over the entire grid, until the process converges to a specified degree of accuracy, or fails to converge after a specified number of iterations.

Thus one has the interesting situation of two iterative procedures, one inside the other, so to speak, going simultaneously. It is interesting to ask how their rates of convergence affect one another. What we offer here are some early experimental (in a numerical sense) results which bear on this question.

One fixes at the outset the accuracy required for the final solution, but since one has two iterative sequences proceeding at once, the total amount of computation required may depend upon how one obtains successive iterates. Thus, suppose we describe the outer iteration as:

$$T = T^{o} + \delta T^{(1)} + \delta T^{(2)} + \delta T^{(3)} \ldots\ldots$$

and set as a criterion for convergence that

$$\left|\left(T_{ij}^{(n)} - T_{ij}^{(n-1)}\right)/T_{ij}^{(n)}\right| < \varepsilon_1$$

while the inner iteration is described by:

$$\delta T_o^{(k)} , \delta T_1^{(k)} , \delta T_2^{(k)} \ldots\ldots\ldots \rightarrow \delta T^{(k)}$$

and one requires for convergence that

$$\left|\left(\delta T_m^{(k)} - \delta T_{m-1}^{(k)}\right)/\delta T_m^{(k)}\right|_{i.j} < \varepsilon_2$$

One might anticipate that it is worthwhile determining $\delta T^{(k)}$ very accurately since this presumably would improve the convergence of the outer iteration and require determination of fewer $\delta T^{(k)}$. Our experience with these equations indicates that this is not the case. We found that we could arrive at the same final solution for $\varepsilon_1 = .001$ with $\varepsilon_2 = .001, .050$ or $.010$. The number of additional $\delta T^{(k)}$ required to be computed about doubled - but the total volume of computation was considerably smaller when the larger value of ε_2 was used. The total computation time was reduced by a factor of from two to five depending upon the geometric parameter λ and the value of the applied load.

The iterative procedure encounters convergence difficulties of two distinct types. The first stems from the fact that the boundary conditions as formulated give rise for "adequate" mesh spacing to an off-diagonal term which is dominant in the system of algebraic equations. This causes the inner iterative procedure to fail to converge. One alternative was to simulate the clamped edge with the boundary conditions

$$w(\lambda , \theta) = o \qquad w'(\lambda , \theta) = o$$

$$\ddot{n}(\lambda , \theta) = C_1 M_o \qquad \psi'(\lambda , \theta) = C_2 X_o$$

where M_o and X_o are the meridional moment and shear as determined by an asymptotic analysis for the linear axisymmetric case. The argument tentatively is that for the uniform pressure loading case the difference between this set of boundary conditions and the more general statement of a clamped edge will not propagate very far into the shell. We have obtained results for certain values of λ, and these results agree quite well (as they should) with the axisymmetric theory for early stages of deformation.

The second type of convergence difficulty for the iterative procedure is of a different type and can be shown to be dependent upon both the value of applied load and the mesh size employed. For example, consider a clamped spherical cap under uniform pressure and having a geometry parameter $\lambda = 12$. With a mesh spacing of $N_x = 20$, $N_\theta = 10$, we found solutions for nondimensional values of pressure $p = .1$ to $p = .9$ in increments of $.1$. It was anticipated that nonsymmetric solutions would begin to emerge at about $p = .75$, but they failed to do so. Arguing that they might emerge when more meridians were used, we repeated the calculations for $N_x = 20$, $N_\theta = 30$. This time, we found symmetric solutions but only up to $p = .52$, after which convergence could not be obtained. In these calculations the starting solution for a new value of p was the previously converged solution for a lower value of p . (To initiate the whole process we used the linear solution which is obtainable directly for small values of load). To test the hypothesis that there may be a numerical instability present, we began using as starting solutions for higher values of load the converged solutions for $N_x = 20$ and $N_\theta = 10$, with the same value of p . We were then able to find solutions for $N_x = 20$, $N_\theta = 30$ up to $p = .7$ but not at $p = .8$. These convergence difficulties of the iterative procedure are believed to be due to the fact that in using a Newtonian iterative procedure to treat the nonlinear aspect of the problem we have a situation in which the coefficients of the linear partial differential equations to be solved at each step of the nonlinear iteration depend upon the most recent estimate of the solution. Thus the convergence qualities of this numerical method depend upon the value of the solution and this can result in the situation where the iterative procedure converges for certain loads and mesh spacing but not for others. It was decided therefore to examine the question of whether a direct method which does not depend upon the convergence of the inner iteration might be more suitable for these nonlinear equations.

In using a direct procedure for the solution of the large matrices corresponding to the finite difference analogue of the linear differential equations encountered in the Newtonian iteration, we can capitalize on the sparse nature of these matrices by partitioning the "supermatrix" U into the form:

$$U = \begin{bmatrix} B_1 & C_1 & & & A_1 \\ A_2 & B_2 & C_2 & & \\ & A_3 & B_3 & C_3 & \\ & & \cdots\cdots\cdots\cdots & & \\ C_n & & & A_n & B_n \end{bmatrix}$$

where A_i B_i C_i are themselves matrices. It turns out that these submatrices are of the form:

$$A_i = \begin{bmatrix} \beta_1 & \gamma_1 & & & \\ \alpha_2 & \beta_2 & \gamma_2 & & \\ & \alpha_3 & \beta_3 & \gamma_3 & \\ & & \cdots\cdots\cdots\cdots & & \\ & & & \alpha_m & \beta_m \end{bmatrix}$$

where α_i β_i γ_i are in turn submatrices. Efficient methods already exist for inverting a tridiagonal matrix such as A_i, but the presence of the off-diagonal terms A_1 and C_n invalidate the application of those methods for inverting the"supermatrix" U. Since this type of matrix should emerge in other contexts and applications involving periodic phenomena, it was thought worth-while to present a method for solving a system of equations (or matrices) of the form:

MESCALL

$$B_1 T_1 + C_1 T_2 + A_1 T_n = D_1$$

$$A_2 T_1 + B_2 T_2 + C_2 T_3 = D_2$$

$$A_3 T_3 + B_3 T_3 + C_3 T_4 = D_3$$

$$\cdots\cdots\cdots\cdots\cdots\cdots$$

$$C_n T_1 \qquad A_n T_{n-1} + B_n T_n = D_n$$

where $A_i\ B_i\ C_i$ may be matrices and T_i, D_i vectors of order m.

If we define:

$$W_1 = B_1 \qquad S_1 = W_1^{-1} D_1 \qquad R_1 = W_1^{-1} C_1$$

$$Q_1 = W_1^{-1} A_1 \qquad P_1 = R_1$$

$$W_i = B_i - A_i R_{i-1} \qquad S_i = W_i^{-1} (D_i - A_i S_{i-1}),\ R_i = W_i^{-1} C_i$$

$$Q_i = W_i^{-1} A_i Q_{i-1},\ P_i = P_{i-1} R_i$$

$$(i = 2, 3 \ldots\ldots\ldots n-1)$$

$$W_n = B_n - A_n R_{n-1} + (-1)^{n-1} Q_{n-1} \qquad S_n = W_n^{-1} (D_n - A_n S_{n-1})$$

then

$$T_n = M^{-1} \left(S_n - R_n \left(S_1 + \sum_{i=1}^{n-2} (-1)^i P_i S_{i+1}\right)\right)$$

where

$$M = \left(1 - R_n \left(Q_1 + \sum_{i=1}^{n-2} P_i Q_{i+1} + (-1)^{n-2} P_{n-1}\right)\right)$$

and

$$T_i = S_i - R_i T_{i+1} + (-1)^i Q_i T_n \qquad i = n-1, n-2 \ldots 1$$

The above algorithm should substantially increase the efficiency of a direct approach to the problem. It is presently being implemented via a computer program in which it was found that this algorithm substantially reduces memory requirements.

DISCUSSION OF RESULTS

As mentioned earlier in this paper, there are two general approaches to explaining the premature buckling of thin shells observed experimentally. The first considers the possibility of the emergence of alternative (nonsymmetric) paths to larger deformation states. Several writers have attempted to determine the existence of such paths without actually solving the complete set of governing equations. Such attempts concentrate on determining "bifurcation points" from the load-deflection curve predicted by nonlinear axisymmetric theory. By considering only "infinitesimal" perturbations about the basic (symmetric) equilibrium state one may linearize the nonsymmetric part of the problem, and in fact reduce it to an eigenvalue problem. Such "bifurcation analyses" have been made for the case of uniform pressure on the clamped spherical shell, and have predicted buckling loads below those of the nonlinear symmetric theory. Presumably the numerical solutions obtained for the complete nonlinear nonsymmetric equations should reveal the spontaneous emergence of these nonsymmetric states. However this was not found to be the case. When symmetric boundary and load conditions were employed, the asymmetric equations yielded symmetric deformation states which were in close agreement with those of the nonlinear

symmetric equations, and did so for loads up to the symmetric buckling load.

The inference of this result is that for the case of uniform pressure, at least, the supposition of a spontaneous emergence of nonsymmetric states at loads below the critical value of the nonlinear symmetric theory is not well founded. The cause of premature buckling is much more likely to be an imperfection of some sort. As a preliminary result along these lines, and as an example of a solution for a nonsymmetric loading condition, we considered the clamped spherical cap under the load $p = p_o (1 + \alpha \cos \phi)$, where p_o and α are constants regulating the overall and the nonsymmetric load intensities. Such a loading could also be interpreted as providing some idea of the response of a shell under uniform pressure but whose thickness varied in the circumferential direction. It was found, e.g., that for $\lambda = 12$ a five percent deviation from uniform pressure ($\alpha = .05$) could produce departures in deflections on the order of twenty percent, even for the very early load-deformation history. (See Figure 2). Such sensitivity to "imperfections" could easily provide a mechanism for premature collapse. More detailed study of such nonsymmetric imperfections is presently underway.

REFERENCES

[1] Bauer, L., and Reiss, E. L., "Nonlinear Buckling of Rectangular Plates", J. Society Indust. Appl. Math., Vol. 13, No. 3, Sept. 1965 pp. 603-626.

[2] Mescall, J., "Numerical Solution of the Nonlinear Equations for Shells of Revolution", U. S. Army Materials Research Agency, AMRA TR 66-11 (May 1966).

[3] Archer, R. R., "On the Numerical Solution of the Nonlinear Equations for Shells of Revolution", J. Math. Physics, 41, pp. 165-178 (1962).

[4] Thurston, G. A., and Penning, F. A., "The Stability of Shallow Spherical Shells Under Concentrated Load", NASA Contract Report 265 (1965).

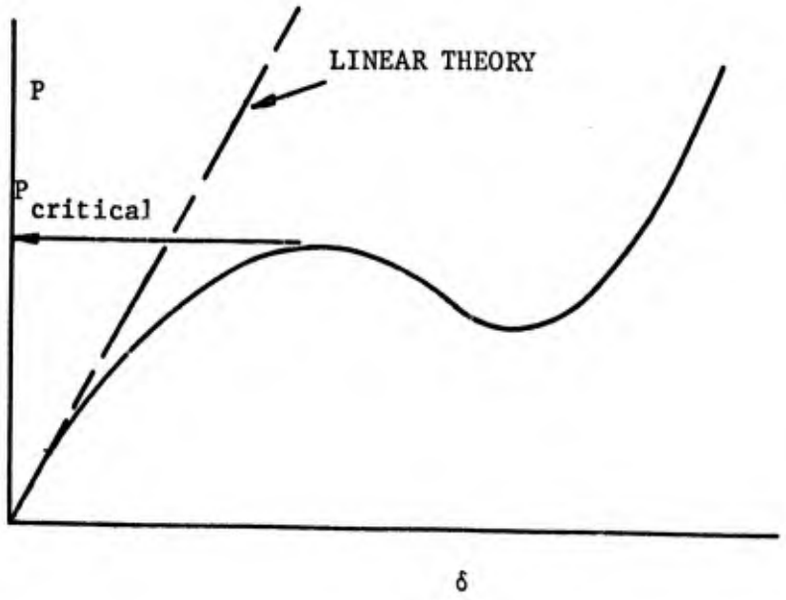


FIGURE 1

TYPICAL LOAD DEFLECTION CURVE

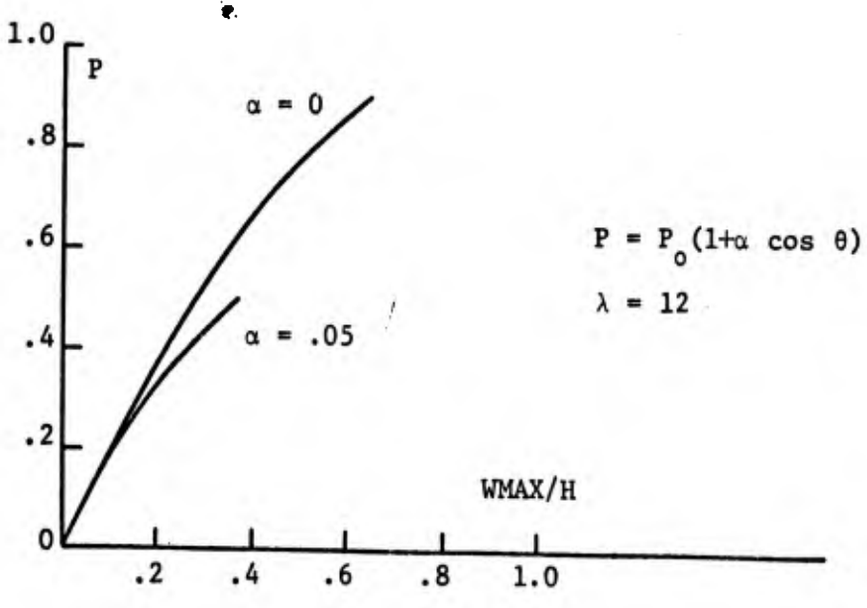


FIGURE 2

MAXIMUM DEFLECTION VS LOAD

IONIC SENSORS

JOHN N. MRGUDICH
U. S. ARMY ELECTRONICS COMMAND
FORT MONMOUTH, NEW JERSEY

I. INTRODUCTION

It is well known that aqueous solutions of many acids, bases and salts are fairly good electrical conductors. This "electrolytic" conductivity is due entirely to movement of ions in solution. There is no "electronic" conductivity such as that observed with metals and semiconductors.

It is also known that a compressed pellet of silver iodide powder is an ionic conductor. The only charge carriers are silver ions relatively free to move through an open but rigid lattice of iodide ions. There is no experimentally detectable electronic conductivity through silver iodide. It is legitimate to think of this pellet of silver iodide as a "solid electrolyte" having many of the electrical properties of an aqueous solution of silver nitrate. Thus, a metallic silver electrode dipping into a solution of silver nitrate exhibits an electrode potential: a metallic silver electrode pressed against one face of a solid pellet of silver iodide should also exhibit an electrode potential.

This paper describes work that has been carried out to explore possible application areas that might emerge from a study of how such solid-electrolyte electrode potentials can sense and respond to changes in the external environment.

II. DEFORMATION SENSING

Consider the case where we have identical silver electrodes in intimate contact with opposite faces of an AgI pellet. The voltage of such an Ag-AgI-Ag array will obviously be zero since the opposing electrode potentials of the two silver electrodes are presumed to be equal.

However, if we bend the array, one electrode will be subjected to a stretching deformation while the other will be subjected to a compressive deformation. These different types of electrode stress (one tensile and the other compressive) should result in different changes in the two electrode potentials involved, and the voltage of the deformed array (reflecting the difference between deformation-induced changes in electrode potentials) should no longer be zero.

A. The First Experiment

The schematic of Figure 1 illustrates the simple experimental setup used for initial demonstration of the existence of deformation-induced voltages.

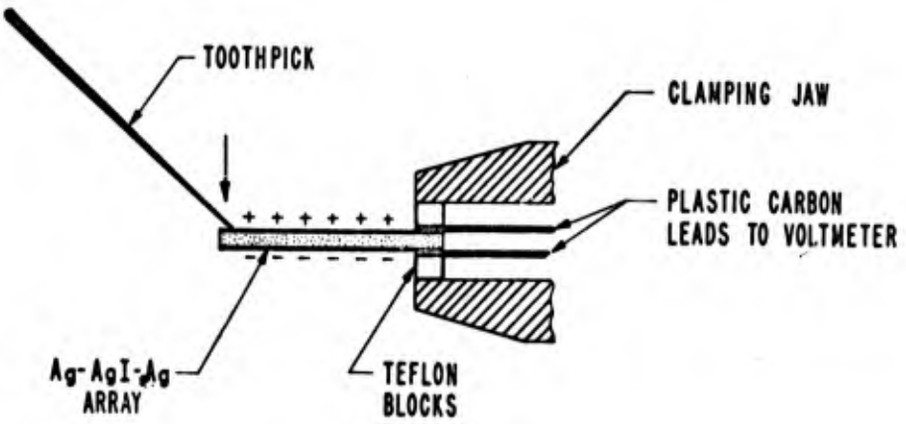


Figure 1. A slight bending torque generates a DC voltage across an Ag-AgI-Ag pellet

This first work used rather thin AgI pellets (0.25 mm thick, 1.27 cm diameter, compressed under 3500 kg load). The silver electrodes were sputtered films approximately 1500 Angstroms thick. Voltage output was read on the one-millivolt scale of a 610B Keithley electrometer (10^{14} ohms input impedance).

It was found that a gentle, steady downward force applied, for example, by the thinner tip of a tapered wooden toothpick to the overhanging edge of the clamped array caused the top silver electrode to exhibit a substantially constant voltage which was positive with respect to the bottom electrode. Removing the toothpick caused the generated voltage to drop quickly to zero. Using the same pick, but applying a steady upward force against the bottom silver electrode, the generated voltage was of about the same magnitude (about one millivolt maximum) but of reversed polarity; that is, the bottom electrode became positive.

Subsequent work showed that the electrode in tension was always positive compared to the electrode in compression. The substantially constant output voltage seemed to be linear with applied force. Response time was apparently quite fast since quick, short taps with the pick resulted in immediate, but not maximum, voltage response.

B. Other Early Experiments

Using substantially the same experimental setup as that shown in Figure 1, it has been found that:

1. Substituting sputtered platinum electrodes for sputtered silver electrodes resulted in arrays which exhibited bending-induced voltages. The effect with platinum, however, was only about half as pronounced.

2. Using soft, electronically conductive plastic electrodes instead of silver resulted in arrays which did not exhibit bending-induced voltages.

3. Use of silver wire electrodes wrapped with absorbent thread and saturated with silver nitrate solution resulted in arrays which did not exhibit induced voltages.

4. Arrays made by compressing powdered silver or powdered platinum electrodes onto the AgI pellet faces exhibited bending-induced voltages.

5. Using powdered silver electrodes, increasing the AgI pellet thickness resulted in increased bending-induced voltage output for the same bending curvature.

We interpret the above observations, and expecially Items 2 and 3, as indicating that the source of the deformation-induced voltage is within the stiff metallic electrodes rather than the solid electrolyte.

C. A Possible Strain Gage Application

1. Experimental approach

The bending-induced voltage effects described above suggest a strain gage application and, with the help of Mr. Frank Priebe and Mr. Leslie Wilcox of USAECOM, we carried out various experiments to explore this possibility. One phase of the work used one millimeter thick arrays made by pressing (3500 kg load) thin layers (0.1 gm) of Engelhard silver powder onto opposite faces of a

1.27 cm diameter pellet of Mallinckrodt AgI (0.5 gm). The experimental set up, illustrated in Figure 2, firmly bonds the Ag-AgI-Ag array near one end of a 46 x 2.54 x 0.12 cm tool steel test bar,

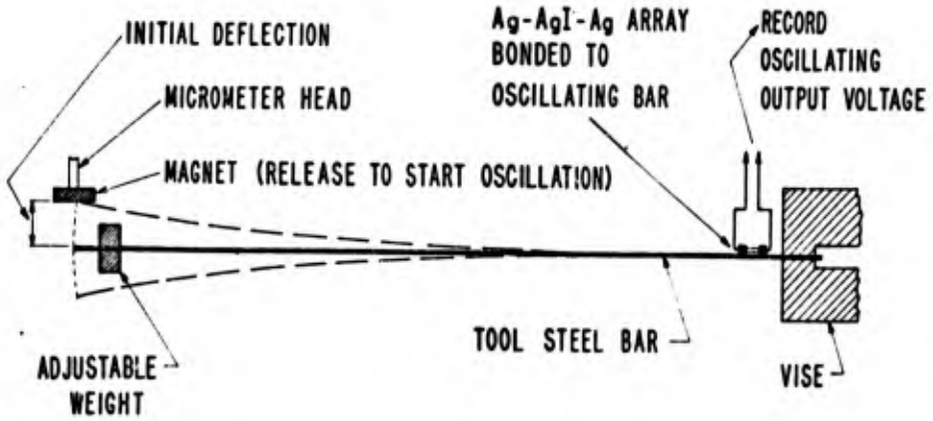


Figure 2. Use of Ag-AgI-Ag arrays in a typical strain gage testing procedure

clamps this end of the bar in a vise, adjusts the bar weight to yield the desired oscillating frequency, deflects and holds the free end of the bar a measured distance from its rest point, starts the Electronics Associates Model 1110 Variplotter, releases the bar by snapping away the holding magnet, and records the voltage response of the array to the pendulum-like flexing forces imposed upon it by the now oscillating bar.

2. Reproducibility

Figure 3 illustrates three repeat runs made recently using a test bar assembly made some eight months earlier. The bar weight was 363 gms; this gave an oscillating frequency of 1.7 cps. The nearly sine-like nature of the recorded voltage output can be taken as strong evidence of faithful response of the array to the simple harmonic flexing motion of the oscillating bar.

3. Electrical "cleanliness"

Of probably greater potential significance is the absence of detectable noise in the traces of Figure 3 despite the 100X amplification applied by the plotter to the array output.

This electrical cleanliness made it possible to record the rather interesting pattern shown in Figure 4. This was made by setting a

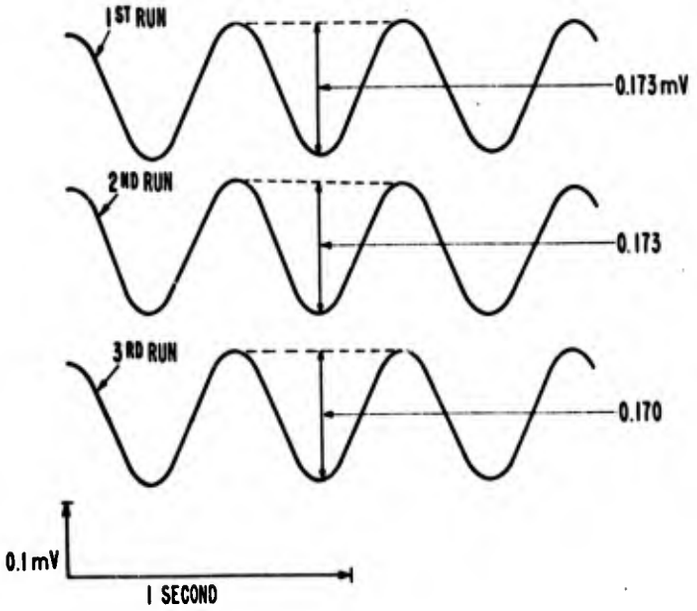


Figure 3. Reproducibility check

heavily weighted bar into oscillation and by adjusting the plotter to a repeat sequence whereby the plotter arm automatically returned to its left starting position after each complete sweep to the right. The maximum amplitude trace was the first sweep, the plotter arm then returned to the left and swept out the second highest amplitude trace, etc. for the fifteen recorded traces, each with a steadily decreasing amplitude as the continually oscillating bar slowly decreased its excursions.

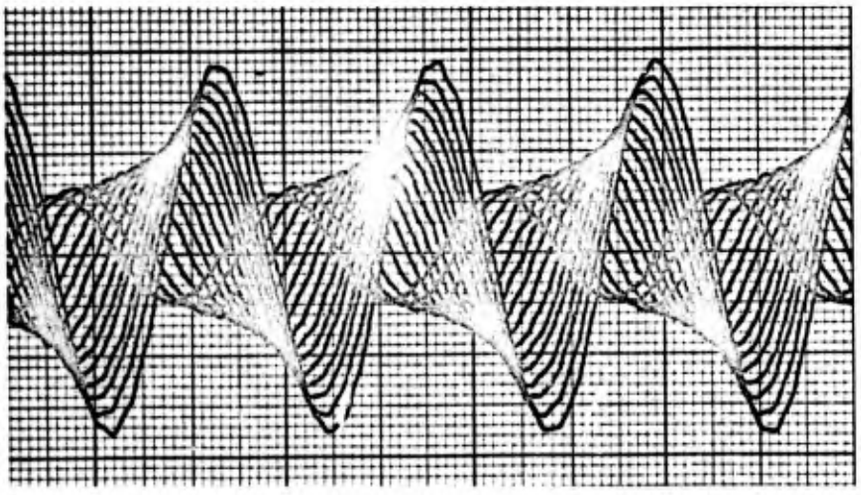


Figure 4. Electrically clean voltage outputs

4. Sensitivity of response

With the present plotter and present array dimension an initial deflection of 0.1 mm of the free end of the 46-cm test bar is just barely detectable as a two-microvolt movement of the plotter pen. Several alternatives involving array shape and dimensions together with a feasibly higher amplification of generated output are available to improve overall sensitivity. Another way to increase sensitivity is to use several series-connected, simultaneously responding arrays.

5. Linearity of response

The plot of Figure 5 illustrated the linear manner in which array output increased with increasing initial bar deflection when bar frequency was 1.7 cps (i.e., a bar weight of 363 gms as in Figure 3). Note that the last point at 0.171 volt/7 mm deflection checks well with the values of Figure 3.

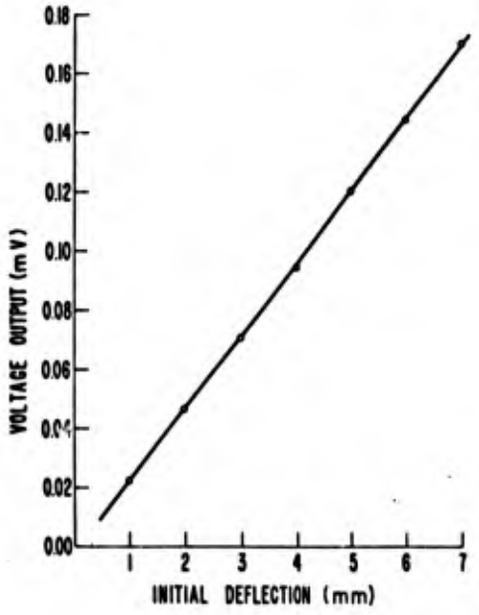


Figure 5. Linearity of response at 1.7 cps

The same straight-line relationship was obtained at 2.4 cps (bar weight reduced from 363 to 150 gms). At higher frequencies (up to 5.8 cps with no added bar weight) the sine-like traces of Figure 3 became increasingly saw-tooth in character and the output versus initial deflection curve of Figure 5 became S-shaped, leveling off at a maximum output of 0.11 millivolt.

It seems to us logical to assume that the array begins to respond the moment the test bar is deflected, but that it takes about

one-half second for this particular array-bar-recorder combination to show maximum output. Dr. Kurt Ikrath of USAECOM, eliminating the test bar and applying an oscillating bending force directly onto the array, has recorded array response at 80 cps.

6. The effect of a temperature change on array behavior

Some preliminary experiments in which the entire bar assemblage was placed in a controlled temperature room have shown that the oscillating array voltage is not drastically temperature dependent over a range extending, at least, from -20°C to 40°C, provided that both electrodes are at the same temperature.

This condition of equal electrode temperatures is stringent since it is known (Katayama, 1908; Reinhold, 1928; Mrgudich, 1965) that a difference of but one degree Centigrade between the temperatures of the two silver electrodes will yield an "ionic thermocouple" voltage of about 0.6 to 0.8 millivolt. The warmer electrode is negative (i.e., warming a silver electrode increases its tendency to ionize by pumping positive silver ions into the AgI electrolyte and charge-compensating electrons into the external circuit).

There are, however, two much more worrisome aspects of the effect of temperature changes on array behavior. One of these revolves about the fact that silver iodide is an example of a very rare phenomenon: in the temperature range 0°C to 145°C, silver iodide contracts on heating and expands on cooling. Since the pressed-on silver electrodes behave normally (expand on heating), it is obvious that temperature changes, expanding one side of the array interface and contracting the other side, will tend to disrupt the silver/silver iodide contact with probably increased tendency to electrical noise. In addition, all electrical contacts to the array should be of a mechanical type since the heat of soldering would be definitely harmful. The second worrisome temperature effect is that silver iodide undergoes a phase change at 145°C which results in disintegration of the compressed pellet.

Both of these effects (contraction on heating and disintegration at 145°C) can probably be neatly avoided by going to some other silver halide (AgBr) or through use of some of the newer alkali halide/silver halide double salts (MAg_4I_5) reported by Bradley and Greene (1966, 1967) and Owens and Argue (1967).

7. Use of Ag-AgI-Ag arrays as strain gages

Present wire or semiconductor strain gage technology is very well advanced from the important standpoints of sensitivity, reliability, versatility, miniaturization and cost. All such gages, however, require accessory power to energize bridge circuits needed

to measure changes in resistance associated with stress-induced deformation. The chief, and at the moment possibly only, advantage of the solid-electrolyte strain gage is that its output is self-generating so that operating power requirements can be much less. This opens the door to applications involving unattended, long-term monitoring of deformable sensing elements such as diaphragms, bellows, Bourdon tubes, etc. Another possible applications area may involve extremely high temperature strain measurements using as electrolyte any of various oxides, salts, or ceramics which become stable, ionic conductors at elevated temperatures.

8. An interesting rotation effect with weighted arrays

It is not necessary that we limit ourselves to strain gage applications. Figure 6 is an exploded cross-section of an ionic device which uses deformation sensing to measure rotation.

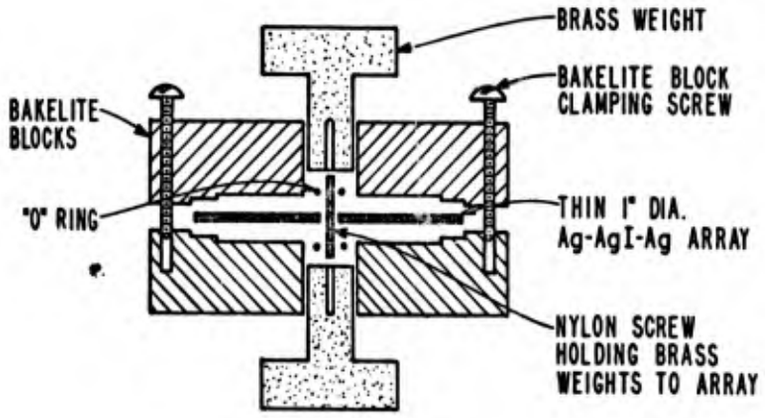


Figure 6. An ionic device to sense rotation

The two brass weights (approximately 63 gms each) are screwed onto the nylon screw and held snugly across the small "O" rings to the thin (0.9 mm) Ag-AgI-Ag array (2.54 cm diameter). The array is clamped between bakelite blocks machined to hold the weighted array by its rim. Not shown are small ports in the bakelite blocks which permit insertion of small clips to make electrical contact between array and plotter.

In the vertical position as shown in Figure 6, the weight of both brass blocks deforms the array downwards, yielding a measured output voltage of about .07 millivolt. When turned through 90° about a horizontal axis parallel to the plane of the array, the voltage output is zero since the weight vectors are parallel to the ar-

ray and bending torques cancel. An additional 90° rotation, which reverses the positions of the two weights, deforms the array in the opposite direction resulting in a voltage output of 0.07 millivolt of reversed polarity. The voltage output followed array rotation in an apparently faithful manner, yielding a repeating sine wave as the device was rotated at a uniform rate.

9. A possible light weight seismograph

It was noticed throughout all this work on deformation sensing, and expecially with work using weighted arrays, that voltage response was quite sensitive to vibration. The small irregularities in the sine-like traces of Figure 4 are very probably due to pickup of stray vibrations. This sensitivity to vibration, easily enhanced by use of suitable mechanical lever arms, might be exploited as the basic principle of a light weight seismograph with direct electrical readout. If, as seems likely, electrolytes stable over a wider range of temperatures could be perfected, it is quite possible that a seismograph particularly suitable for long-term, unattended monitoring of remote terrestial stations, or even lunar stations, could be developed.

10. Summary of deformation sensing

The wisdom of pursuing efforts to develop other possible applications areas -- accelerometers, microphone pickups, safety devices, alarm systems, traffic monitoring, arterially or abdominally activated pacemakers, etc. -- should be evaluated in terms of the apparently unique properties of the solid electrolyte approach:

a. At rest the array is a passive, low-impedance, zero-volt component.

b. Arrays are stable, inexpensive and amenable to miniaturization (including thin-film fabrication).

c. Arrays are sensitive to applied deformation, quickly responding by generating a small but noise-free, fairly stable dc voltage.

d. Arrays can be wired in series to increase voltage output or in parallel to increase response reliability.

This listing of advantages must be followed by a listing of disadvantages and possible pitfalls:

a. Present arrays using AgI as the electrolyte have a rather narrow operating temperature range.

b. An Ag-AgI-Ag array, or any other array using different electrolytes, is a solid-electrolyte battery whose normal open-circuit voltage is zero, but which can be "charged" by mechanical deformation. Development of new battery systems is normally fraught

with many unexpected difficulties.

III. TEMPERATURE SENSING

A. Introduction

Mention has been made above that a dc voltage is generated by a temperature difference between the metallic electrodes of symmetrical solid-electrolyte arrays such as Ag-AgI-Ag, Pb-$PbCl_2$-Pb, Ag-AgBr-Ag, etc. Temperature coefficients are generally rather high (0.1 to 1.0 millivolt per degree Centigrade compared to conventional metallic thermocouples at 0.01 to 0.05 millivolt per degree). Reference (3) describes such an effect as the basis of an ionic thermogenerator.

We will here be concerned with the use of this property as the basis of an ionic temperature sensing device.

B. An Ionic Typewriter Keyboard

It was while working with Ag-AgI-Ag arrays as deformation sensors that it was noticed that even momentary, warming fingertip contact with one face of such arrays resulted in a voltage response. An intriguing aspect of this was the speed of response (0.1 to 0.2 seconds) to even the lightest finger touch. Indeed, the array began to respond even before the finger touched it (apparently due to sensing of radiated finger heat). This fast response is probably due to the low heat capacity of the thin silver electrode and its relatively large surface area (approximately 1 sq-cm).

Mr. Paul Bramhall and Mr. Bernard McLaughlin of USAECOM have studied the possibility that one could build an "ionic" typewriter keyboard in which fast finger contact (typing) could yield voltage output signals for subsequent processing. They assembled a 4-key bank of series-connected, one millimeter thick Ag-AgI-Ag arrays To minimize deformation response when touched, each array was firmly supported in recesses milled into a wooden block. The device performed as expected: a quick finger tap on the top (exposed) electrode gave a quick voltage response; using two fingers and simultaneously touching two arrays approximately double the voltage output, etc. Voltage output varied from person to person, depending on individual fingertip temperatures. Some of our more facetious friends have extrapolated this observation to a requirement that typists for this application should have "hot" fingertips.

As laymen in medicine, we were rather surprised to find two persons (a man and a woman) whose fingertips were practically invariably cooler than room temperature (implying the presence of a cooling, heat-absorbing biological mechanism). We were also surprised

to note that exercise tended to lower the fingertip temperature of all individuals tested.

C. An Ionic Surface Temperature Probe

Figure 7 is an exploded view of the head of a surface temperature probe. The Ag-AgI-Ag array was made by lightly precompressing a relatively large amount (5 gms) of silver powder in a 1.27 cm

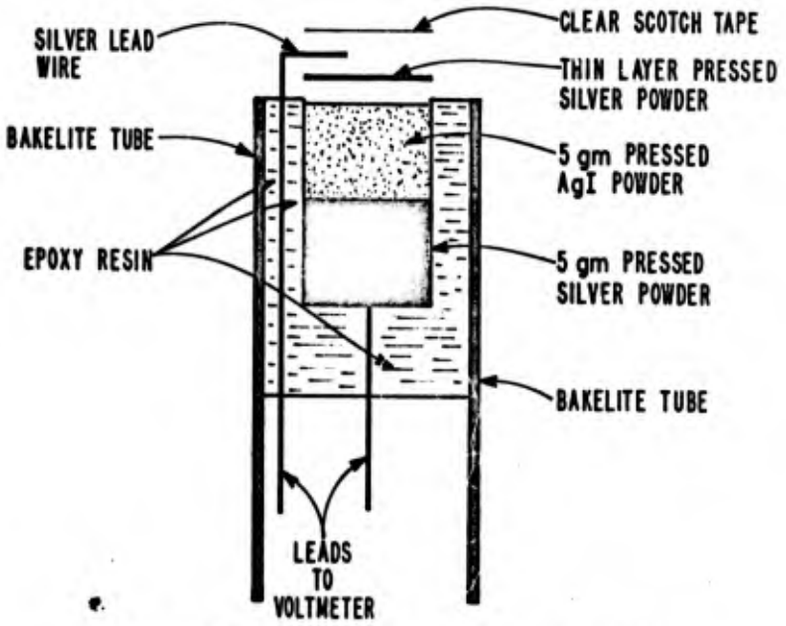


Figure 7. Surface temperature probe

cavity, adding and lightly precompressing a thick (5 gm) block of AgI powder, adding and lightly precompressing a small amount (0.025 gm) of silver and then completing array compression by pressing at 3500 Kg load. A lead was soldered to the thick block of silver (its thickness prevented harmful heating of the AgI). The array was then positioned in a bakelite tube (2.54 cm diameter), a flat silver lead bent and positioned to contact the top, thin silver electrode and the assembly encapsulated with epoxy resin. A strip of clear scotch tape was placed across the top electrode to maintain contact between lead and electrode and to protect the tip of the probe against gross contamination.

The thin top electrode has a low heat capacity so that its temperature response will be fast; the thick silver iodide block tends to lower heat transport between the top sensing electrode and the bottom, thick reference electrode. The large size of the bottom electrode tends to minimize the heating effect of what little heat is transported from the sensing electrode.

The probe worked as expected in that response was fast and held to a steady value for up to two minutes. Probe calibration was by means of large rubber erasers preheated to known temperatures and held against the sensing electrode. The temperature coefficient was about 0.4 millivolt per degree Centigrade. This is somewhat lower than expected, probably because of the insulating effect of the protective scotch tape.

IV. TIME SENSING: SOLID-ELECTROLYTE TIMERS

An earlier report on "A Rechargeable Thin-Film Solid-Electrolyte Battery" was presented at the 1966 Army Science Conference (6). This described the battery characteristics of an Ag-AgI-Pt array operating as a solid-electrolyte concentration cell in which the internal driving force was the nigh concentration of silver in the silver anode (negative) tending to force positively charge silver ions through the AgI electrolyte to reach a balance with the low concentration of silver in the platinum cathode (positive).

The significant feature of that report is that the closed-circuit voltage of the discharging Ag-AgI-Pt concentration cell dropped exponentially with time of discharge.

The discharge curves of Figure 7 (Figure 5 of the reference report) illustrate the exponential closed-circuit voltages when the battery, recharged between discharges, was discharged across different resistors.

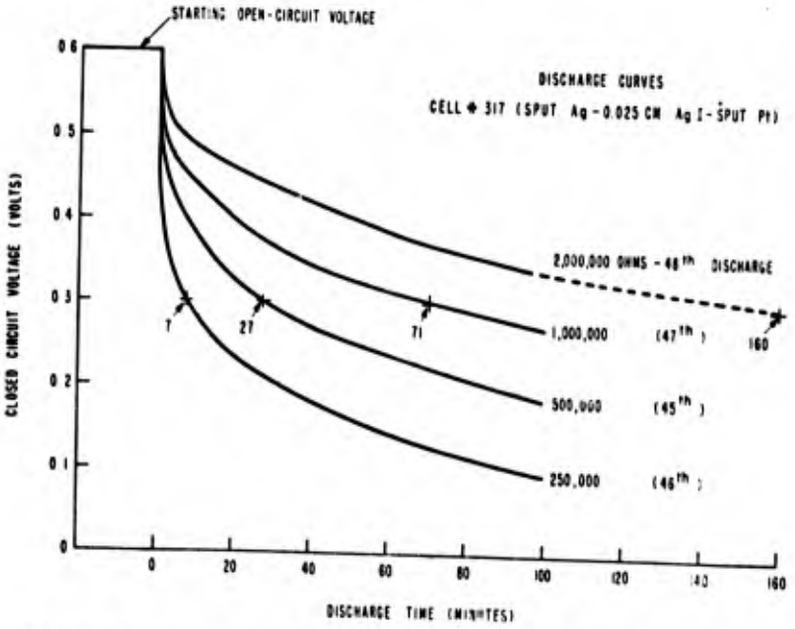


Figure 7. Time dependence of the closed-circuit voltage of a discharging Ag-AgI-Pt battery

Given an accessory circuit which would switch when the closed-circuit voltage reached some predetermined value (say 300 millivolts) and given that we want this switching to take place after a given time lapse (say 27 minutes), it seems likely that the cell, discharge across 500,000 ohms, could serve as the timing device for these particular conditions.

This "timer" application has not been explored: it remains as a possible applications area to be evaluated as and if a need arises. Potential advantages include ease of fabrication, low cost, rugged construction, reasonable reliability, ease of presetting to yield variable time through adjustment of the external resistance, reuse through recharging. The major anticipated problem area is temperature dependence which could possibly be alleviated by suitable compensation using two batteries of different output capacities and subsequent monitoring of the <u>difference</u> in closed-circuit voltages.

V. CURRENT SENSING

Consider the 3-pellet, 4-terminal Ag_1-AgI_1-Pt_1-AgI_2-Pt_2-AgI_3-Ag_2 solid-electrolyte array of Figure 8. The Ag_1-AgI_1-Pt_1 segment is a rechargeable solid-electrolyte battery that can be charged to, say, 0.60 volt at which point Ag_1 will tend to pump silver ions internally and electrons externally into Pt_1. Similarly, the Ag_2-AgI_3-Pt_2 segment can be charged to exactly 0.60 volt so that Ag_2 will tend to pump silver ions internally and electrons externally into Pt_2.

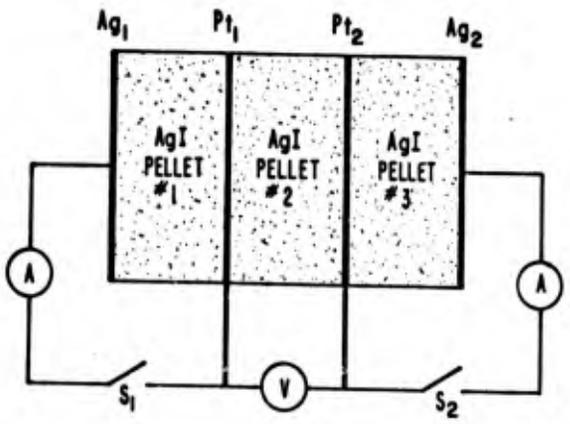


Figure 8. An ionic device showing a high voltage readout for a low current readin

After both segments are charged to the same voltage (easily realized by charging in parallel from a common charging source), the voltage across Pt_1 and Pt_2 will be zero.

Now, with switch S_2 open, momentarily close S_1. The left segment will be momentarily discharged and this discharge will deliver a small amount of silver from Ag_1 to Pt_1. The voltage between Pt_1 and Pt_2 will no longer be zero since Pt_1, now contaminated with some silver, will tend to pump silver ions internally and electrons externally from Pt_1 to Pt_2. Pt_1 will become negative with respect to Pt_2.

If both Pt_1 and Pt_2 are extremely thin sputtered films, a minute amount of transported silver will result in a fairly large percentage contamination of Pt_1 and the voltage between Pt_1 and Pt_2 will be quite high. We have observed cases where a current of only 0.01 microampere flowing for one second through S_1 increased the voltage between Pt_1 and Pt_2 from a starting zero value to over 100 millivolts.

An interesting additional feature of this 4-terminal solid-electrolyte device is that we can easily lower the Pt_1-Pt_2 voltage (originally generated by closing S_1 momentarily) by closing S_2 momentarily. The current thus permitted to flow across Ag_2-AgI_3-Pt_2 delivers silver from Ag_2 to Pt_2. When the silver so delivered equals that originally delivered from Ag_1 to Pt_1, the voltage between Pt_1 and Pt_2 will again be zero. If S_2 is closed again, Pt_2 becomes richer in silver than Pt_1 and Pt_2 then becomes negative with respect to Pt_1.

Thus it is that one can alternately tap S_1 and S_2 and adjust the "readout" Pt_1-Pt_2 voltage to any desired value between rather wide limits. Note that the original starting conditions can be restored by recharging from a 0.6 volt charging source.

At one time we felt that this device might be useful as an adaptive memory element, but it turned out that a resistance-change readout was easier to work with than our voltage-change readout.

At the moment it seems that the device can be useful as a simple but very, very sensitive coulometer.

VI. EPILOGUE

The driving force behind this effort was an inner conviction that control of ionic charge transport phenomena offered potential application opportunities equal to if not greater than those already developed through control of conventional electronic charge transport.

It was realized from the very start that ionic charge transport involved simultaneous transport of charge and atomic mass. Such transport of matter as an inevitable consequence of transport of

charge can be viewed, with considerable justification, as a disastrous shortcoming. More optimistically, it can be considered as a sort of second degree of freedom which, if properly turned, may permit us to accomplish a multiplicity of objectives with but a single transport. It should be recalled that biological systems have associated with their often fantastically high levels of sensitivity, versatility, adaptability, responsiveness and retentivity many easily detected electrical effects based solely upon ionic charge transport through membranes of variable permeability.

The device applications described above are but an early effort in a field that modern technology has substantially ignored.

ACKNOWLEDGEMENT

In addition to the help of those mentioned in the text, the author would like to thank Mr. Alan Fein and Mr. Edward Fister of USAECOM; the former for laboratory assistance and the latter for continuing interest and constructive criticism. He is also indebted to Dr. Jerome Catz of the University of Miami, Coral Gables, Florida.

REFERENCES

1. M. Katayama, Zeit. f. Phys. Chem., 61, 566 (1908)
2. H. Reinhold, Zeit. anorg. u. allgem. Chem., 171, 181 (1928)
3. J. N. Mrgudich, U. S. Pat 3,170,817, February 23, 1965.
4. J. N. Bradley and P. D. Greene, Trans. Faraday Soc., 62, 2069 (1966); 63, 424 (1967).
5. B. B. Owens and G. R. Argue, Science, 157, 308 (1967).
6. J. N. Mrgudich, P. J. Bramhall and C. M. Schwartz, Army Science Conference Proceedings, Office, Chief of Research and Development, Department of the Army, Washington, D. C. (1966).

THYROIDAL DYSFUNCTION DURING SIMULATED ALTITUDE CONDITIONS*

PHILIP F. MULVEY, JR.
U. S. ARMY RESEARCH INSTITUTE OF ENVIRONMENTAL MEDICINE
NATICK, MASSACHUSETTS

Previous studies have indicated various degrees of thyroidal dysfunction in rats during exposure to high altitudes usually by simulated techniques (1-8). The present study was undertaken to determine whether acute exposure to hypoxic conditions of simulated 15,000 ft. (429 mm Hg), exerts its effect on the trapping of ^{131} iodine and/or the processes of organification in the biosynthesis of thyroxine and its release from the thyroid gland of the rat. The data obtained in the present study indicated that the conversion of radiomonoiodotyrosine (MIT-^{131}I) to radiodiiodotyrosine (DIT-^{131}I), and the synthesis of radiothyroxine (T_4-^{131}I) were significantly inhibited by hypoxia. In addition, total plasma radioactivity (^{131}I) and newly formed protein bound iodine (PB^{131}I) from hypoxic rats were significantly lower compared with control animals.

MATERIALS AND METHODS

Sprague-Dawley rats of the Charles River strain with an initial body weight of 150-200 gm were employed for all studies. Animals were fed a low iodine diet either as pelleted or powdered test form for 12 - 49 days. All animals used in a specific experiment were on the diet for the same length of time. The powdered iodine test diet contained 0.18 ug of iodine per gram of diet**(Low Iodine Test Diet, Nutritional Biochemical Corp., Cleveland, Ohio) and was used for all studies with the exception of the 23.5 hour and 28 hour studies. The pelleted form was not chemically analyzed and was used in the 23.5 hour and 28 hour studies. Water was allowed ad libitum. For each experiment, 4-8 rats were placed in individual

*The principles of laboratory animal care as established by the National Society for Medical Research were observed.
**Assayed by Graham Laboratories, Inc., Washington Street, Brookline Massachusetts.

chambers at a barometric pressure of 429 mm Hg (equivalent to 15,000 feet altitude) ventilated with a regulated air flow of 1.5 liters per minute. Equal number of rats were placed in similar chambers maintained at atmospheric pressure. After a specific interval of time (1 to 24 hours), the experimental animals were returned to atmospheric pressure. All animals then were given $Na^{131}I$ (1.1 to 2.6 uc) intraperitoneally. The experimental rats were brought back to 429 mm Hg for an additional 16 hours. All animals were sacrificed by exsanguination under light ether anesthesia. The thyroids were removed, weighed, counted in a Packard Auto-Gamma Spectrometer and hydrolyzed at pH 7.4 for 24 hours with pancreatin in 1 ml of Krebs-Ringer phosphate medium with added propylthiouracil (10^{-3}M). Twenty-five (25) microliter aliquots of digest were chromatographed in the following solvent systems: butanol-acetic acid-water (75-10-15) ascending, t-amyl alcohol saturated with 2N NH_4OH descending and t-butanol-2N NH_4OH (4-1) descending. The strips were radioautographed for identification of the ^{131}I containing spots. Each radioactive area was cut from the strip and counted in a NaI (TL) well system. A sample of serum from each animal was treated with 10 volumes of 15% cold trichloroacetic acid (TCA) to precipitate the proteins. This was then centrifuged and washed twice with 1% TCA and the precipitate assayed for radioactivity (protein bound ^{131}iodine-PB^{131}I). A sample of untreated serum was assayed to determine the total plasma ^{131}I radioactivity. Analysis of variance and/or student's t test, non-paired, were used to analyze the statistical significance of the data (9).

RESULTS

I. Thyroidal uptake of ^{131}I (Table 1): Rats exposed to simulated altitude for 1 to 24 hours prior to receiving ^{131}I and re-exposed to hypoxic conditions for an additional 16 hours showed no statistically significant alteration in uptake of ^{131}I compared with ambient pressure controls.

II. Chromatographic analysis of ^{131}I in thyroid digest (Table 2): Chromatographic results indicated that the thyroid glands from rats at altitude for 23.5 hours had a statistically significant increase in the relative formation of radiomonoiodotyrosine (48.7 $\pm$ 1.5% vs 43.0 $\pm$ 1.2%; $P<0.001$) compared with controls. A statistically significant reduction in radiothyroxine synthesis was also observed in the digest from the hypoxic animals (2.8 $\pm$ 0.52% vs 6.6 $\pm$ 0.33%; $P<0.001$) following 28 hours at altitude. After 40 hours of exposure the synthesis of radiotriiodothyronine appeared to be significantly enhanced (12.6 $\pm$ 0.9% vs 9.0 $\pm$ 1.3%; $P<0.05$) compared with the ambient pressure controls. The chromatography data did not indicate any statistically significant change in the level of gland iodide between control and experimental animals following various exposure periods.

IIa. Effect of altitude on the ratio of MIT-^{131}I/DIT-^{131}I and T3-

$^{131}I/T_4$-^{131}I (Table 3): Hypoxic animals indicated an enhanced MIT-^{131}I/DIT-^{131}I ratio following 23.5 hours of exposure (1.79 vs 1.47; $P<0.001$). The ratio of T_3-$^{131}I/T_4$-^{131}I was also increased in the experimental glands after 28 hours at altitude (3.97 vs 1.17; $P<0.001$).

III. Effect of hypoxia on plasma radioactivity (^{131}I) and protein bound iodine -PB^{131}I. Table 4: Exposure to altitude for 17 hours resulted in a significant reduction in total plasma radioactivity and protein bound radioiodine (0.83 ± 0.15 vs 2.13 ± 0.39; $P<0.01$) and 59.8 ± 7.5 vs 81.2 ± 4.7; $P<0.02$) respectively. In the present study changes in plasma radioiodine and in PB^{131}I levels in the hypoxic animals were detected prior to alterations in thyroxine biosynthesis by the thyroid gland.

DISCUSSION

The elevated thyroidal MIT/DIT ratio as seen in the control animals is probably due to iodine deficiency. Normally, one would expect the thyroidal MIT/DIT ratio in rats on an adequate iodine diet to be less than one (10). In addition, it has been reported (11, 12, 13) that iodine deficiency *in vivo* leads to the formation of more MIT and DIT and more T_3 relative to T_4 (11, 12). The increase in uptake of radioiodine and the enhancement in the MIT-^{131}I/DIT-^{131}I and T_3-$^{131}I/T_4$-^{131}I ratios, tables 1 and 3, as seen in both groups of animals used for the 23.5 hour, 28 hour, and 40 hour studies were probably due to the length of time on the iodine poor diet. These animals were on the diet for a longer time compared with the rats used for the 17 hour, 19 hour, and 20 hour and 21 hour studies. Accordingly, the former group of animals had an increased gland weight compared with the latter group. Assuming, therefore, in an iodine deficient rat that nearly all the radioiodine taken up by the thyroid gland is ultimately converted to hormonally active material, an increase in uptake would be reflected by an increase in the formation and release of thyroid hormones. Since within each experiment the length of dietary control was constant, comparisons in the effect of hypoxia on the thyroid gland and plasma radioactivity levels are appropriate only within each experimental period. Therefore, it is important to realize that the rats used in the present study did appear to show signs of thyroidal iodine deficiency, prior to exposure to simulated altitude. The changes reported in this study in the distribution of radioiodotyrosines and radioiodothyronines within the thryoid gland are relative changes and not absolute ones.

Since hypoxia did not result in significant changes in ^{131}I accumulation or in chromatographically identified inorganic ^{131}I but was associated with an increase in MIT-^{131}I, no depressing effect of hypoxia on the conversion of radioiodide to MIT-^{131}I was observed in the present study. However, the study did indicate an effect of hypoxia on the relative conversion of radiomonoiodotyrosine to

radiodiiodotyrosine. This observation suggested that an initial effect of hypoxia in the thyroid gland was on the biosynthesis of thyroxine and was evident in the conversion of radiomonoiodotyrosine to radiodiiodotyrosine. This step in thyroxinogenesis appeared to be more sensitive to hypoxia than the conversion of iodide to monoiodotyrosine. Nelson and Anthony (14) have reported that following 32 hours at 18,000 ft they observed an increased MIT/DIT ratio along with a decreased T_4 level in hypoxic rats.

As the period of exposure was lengthened a change in the ratio of radiotriiodothyronine to radiothyroxine was observed. The mechanism responsible for this shift in relative distribution of T_3 to T_4 might be associated with the change in MIT/DIT ratio. Slingerland et al (10) using an antithyroid drug, induced an increased MIT/DIT ratio and suggested that in presence of an increased MIT plus a reduced DIT, the synthesis of T_3 is relatively facilitated over that of T_4 and the T_3/T_4 ratio is increased. It should be mentioned, however, that the relative reduced T_4 levels observed might be partially associated with duration of fasting. Grossie and Turner (15) have reported that both the degree and duration of food restriction affects T_4 secretion in the rat. Thus, the relative alteration in the biosynthesis of thyroxine under hypoxic conditions must be carefully interpreted in light of associated changes which could occur by other means than by hypoxia alone.

Although there was no detectable change in thyroidal ^{131}I accumulation during 17 hour exposure to altitude, there was a significant decrease in plasma radioactivity and in $PB^{131}I$. Of particular interest to this observation is the published work of Keating et al (16) and Rawson (17). They showed that thyroid stimulating hormone (TSH) increased both the uptake of radioiodine in the thyroid and the release of the hormone from the gland. However, they regard the release of the hormone as the primary event and the increase of radioiodine uptake as a secondary phenomenon induced by this release. Thus, should hypoxia moderate the rate of release of TSH, the initial effect would be seen on the rate of hormonal release from thyroid gland. This would result in a reduced plasma $PB^{131}I$ level in the animals and could occur prior to any detectable alteration in thyroidal ^{131}I uptake or organification. The concentration of $PB^{131}I$ in the plasma is a complex function which depends not only on levels of TSH but also on the following factors: the proportion of dose accumulated by the gland, the rate of secretion of hormonal iodine, the amount of organic iodine in the thyroidal and extrathyroidal pools and the rates of degradation and fecal excretion of organic iodine. An increased plasma volume in the hypoxic animals would not be expected to contribute materially to the observed difference in plasma radioactivity and $PB^{131}I$ values. Thus, the mechanism responsible for the decrease in plasma radioactivity and $PB^{131}I$ following acute exposure remains to be determined.

ACKNOWLEDGMENTS

The author is indebted to Dr. Milton Landowne for his helpful suggestions and to D. Green, SP-5, for excellent general assistance; and to Dr. Ralph Dusek and Miss Ella Munro for their assistance regarding the statistical analysis of the data.

REFERENCES

1. van Middlesworth, L., Science 110:120, 1949.
2. _____ and M. M. Berry, Amer. J. Physiol. 167:576, 1951
3. DeBias, D. A., and W. Yen, USAF School of Aerospace Medicine, Technical Documentary Report No. SAM-TDR-63-101, 1963.
4. Nelson, B. D., and A. Anthony, Proc. Soc. Exp. Biology and Med. 121:1256, 1966.
5. Harclerode, J. E., R. T. Houlihan and A. Anthony, Proc. of the Penna. Acad. of Science, 38:47, 1964.
6. Verzar, F., E. Sailer and V. Vidoric, J. Endocrin. 8:308, 1952.
7. Surks, M. E. Endocrinology 78:307, 1966.
8. LaRoche, G., and C. L. Johnson, Aerospace Medicine 38:499, 1967.
9. Snedecor, G. W. Statistical Methods Applied to Experiments in Agriculture and Biology, Iowa State College Press, 1959.
10. Slingerland, D. W., D. E. Graham, R. K. Josephs, P. F. Mulvey, Jr., A. P. Trakas and E. Yamazski, Endocrinology 65:178, 1959.
11. Gross, J., and R. Pitt-Rivers, Recent Progress in Hormone Research 10:109, 1954.
12. Leloup, J., and F. Lacliver, C. E. Acad. Science 241:509, 1955.
13. Querido, A., K. Schut and J. Terpstra, J. Ciba Foundation Colloquia on Endocrinology 10:24, 1957.
14. Nelson, B. D., and A. Anthony, Doctoral dissertation, Physiology-Biophysics Laboratories, The Pennsylvania State University, Penna., September, 1964.
15. Grossie, J. and C. W. Turner, Proc. Soc. Exp. Biology and Med. 110:631, 1962.
16. Keating, R. F., R. W. Rawson, W. Peacock and R. D. Evans, Endocrinology 36:137, 1945.
17. Rawson, R. W. Ann. N. Y. Acad. Sci, 50:491, 1949.

TABLE 1. EFFECT OF SIMULATED ALTITUDE ON THYROIDAL UPTAKE OF ^{131}I.

Time at Altitude			Thyroidal Uptake
Pre-^{131}I	Post-^{131}I		% of dose/mg thyroid
Hrs.			
1	16	C(13)	1.60 ± 0.11*
		A(13)	1.43 ± 0.18
			N.S.
3	16	C(8)	1.50 ± 0.13
		A(8)	1.20 ± 0.19
		P	N.S.
4	16	C(12)	1.06 ± 0.13
		A(12)	1.01 ± 0.26
		P	N.S.
5	16	C(16)	0.85 ± 0.11
		A(16)	1.24 ± 0.09
		P	N.S.
7.5	16	C(32)	1.78 ± 0.13
		A(32)	1.77 ± 0.11
		P	N.S.
12	16	C(16)	1.75 ± 0.19
		A(16)	1.55 ± 0.08
		P	N.S.
24	16	C(7)	1.80 ± 0.35
		A(7)	1.21 ± 0.10
		P	N.S.

C = control animals. The figure in parenthesis represents number of animals per group
A = simulated altitude (hypoxic) animals
P = probability that the mean of the altitude group is not statistically different than control group ($P > 0.05$)
* = mean ± standard error of mean

TABLE 2. EFFECT OF SIMULATED ALTITUDE ON DISTRIBUTION OF ^{131}I IN IODOTYROSINES AND IODOTHYRONINES OF HYDROLYZED THYROID GLANDS

Altitude-Hrs ^{131}I Pre	Post		% Total ^{131}I on Chromatogram					
			Butanol-Acetic Acid - H_2O*				T-Butanol - 2N NH_4OH**	
			I	MIT	DIT	Front	T_3	T_4
1	16	C(7)	16.1 ± 1.3	38.5 ± 0.46	27.3 ± 1.9	15.3 ± 1.1	13.3 ± 2.5	19.7 ± 1.9
		A(7)	18.4 ± 1.4	33.2 ± 1.0	27.9 ± 1.4	20.4 ± 0.07	20.1 ± 0.06	21.8 ± 2.0
		P	N.S.	N.S.	N.S.	N.S.	N.S.	N.S.
3	16	C(8)	7.8 ± 1.1	42.0 ± 3.6	27.7 ± 2.3	21.4 ± 4.1	9.3 ± 1.7	12.3 ± 1.2
		A(8)	10.3 ± 1.3	34.3 ± 3.3	31.3 ± 1.1	24.1 ± 3.6	7.8 ± 1.8	12.2 ± 1.9
		P	N.S.	N.S.	N.S.	N.S.	N.S.	N.S.
4	16	C(8)	10.9 ± 1.4	44.3 ± 5.4	23.1 ± 1.4	21.8 ± 5.7	5.9 ± 0.84	7.9 ± 1.2
		A(8)	9.5 ± 2.3	43.8 ± 4.6	20.5 ± 1.9	26.1 ± 4.6	4.8 ± 1.1	6.1 ± 1.4
		P	N.S.	N.S.	N.S.	N.S.	N.S.	N.S.
5	16	C(16)	8.5 ± 1.6	44.2 ± 2.0	31.9 ± 1.9	15.3 ± 2.5	6.2 ± 1.1	14.9 ± 1.9
		A(16)	6.6 ± 0.6	46.2 ± 1.3	35.3 ± 2.0	11.8 ± 1.8	5.8 ± 1.1	10.6 ± 0.9
		P	N.S.	N.S.	N.S.	N.S.	N.S.	N.S.
7.5	16	C(32)	10.6 ± 0.8	43.0 ± 1.2	30.3 ± 1.0	16.0 ± 1.2	13.2 ± 1.7	8.3 ± 0.6
		A(30)	10.1 ± 0.7	48.7 ± 1.5	28.0 ± 0.8	13.5 ± 1.0	16.8 ± 3.0	7.2 ± 0.6
		P	N.S.	< 0.001	N.S.	N.S.	N.S.	N.S.
12	16	C(16)	12.6 ± 1.1	41.6 ± 0.91	25.2 ± 1.3	20.6 ± 1.1	7.5 ± 0.66	6.6 ± 0.33
		A(16)	14.6 ± 1.1	51.1 ± 1.6	23.7 ± 1.0	10.5 ± 1.1	8.9 ± 0.84	2.8 ± 0.52
		P	N.S.	< 0.001	N.S.	< 0.001	N.S.	< 0.001
24	16	C(7)	12.8 ± 1.2	45.8 ± 2.1	25.5 ± 1.5	15.9 ± 1.51	9.0 ± 1.3	7.9 ± 1.2
		A(7)	13.7 ± 0.5	48.9 ± 2.5	22.5 ± 1.7	14.8 ± 1.46	12.6 ± 0.9	4.1 ± 0.5
		P	N.S.	N.S.	N.S.	N.S.	< 0.05	< 0.02

*I^-, MIT, DIT, Front determined in Butanol - Acetic Acid - H_2O system.
**T_3, T_4 determined in either t-Amyl-2N NH_4OH or t-Butanol-2N NH_4OH system.

MULVEY

158

TABLE 3. EFFECT OF SIMULATED ALTITUDE ON THE RATIO OF RADIOMONOIODOTYROSINE TO RADIODIIODOTYROSINE AND RADIOTRIIODOTHYRONINE TO RADIOTHYROXINE

Time at Altitude Pre-^{131}I Hrs	Post-^{131}I	MIT-^{131}I/DIT-^{131}I	T_3-^{131}I/T_4-^{131}I
1	16	C(7) 1.41	0.66
		A(7) 1.19	0.87
		P N.S.	N.S.
3	16	C(8) 1.65	0.78
		A(8) 1.12	0.67
		P N.S.	N.S.
4	16	C(8) 1.97	0.78
		A(8) 2.25	0.86
		P N.S.	N.S.
5	16	C(16) 1.44	0.41
		A(16) 1.34	0.54
		P N.S.	N.S.
7.5	16	C(32) 1.47	1.74
		A(30) 1.79	2.33
		P < 0.001	N.S.
12	16	C(16) 1.76	1.17
		A(16) 2.24	3.97
		P < 0.05	< 0.001
24	16	C(7) 1.86	1.31
		A(7) 2.25	3.49
		P N.S.	= 0.01

TABLE 4. EFFECT OF HYPOXIA ON PLASMA RADIOACTIVITY (^{131}I), AND PROTEIN BOUND IODINE ($PB^{131}I$)

Time at Altitude Pre-^{131}I (Hrs)	Post-^{131}I (Hrs)		Total Plasma Radioactivity (^{131}I) % dose/10 ml plasma	% of Total Plasma Radioactivity as: $PB^{131}I$
1	16	C(12)	2.13 ± 0.39	81.2 ± 4.7
		A(12)	0.83 ± 0.15	59.8 ± 7.5
		P	< 0.01	< 0.02
4	16	C(8)	4.43 ± 0.53	93.6 ± 0.80
		A(8)	1.15 ± 0.33	78.8 ± 5.2
		P	< 0.001	< 0.02
5	16	C(16)	3.06 ± 0.38	87.1 ± 4.5
		A(16)	0.80 ± 0.11	62.4 ± 7.3
		P	< 0.001	< 0.01
7.5	16	C(31)	5.78 ± 0.71	88.8 ± 2.6
		A(31)	2.64 ± 0.26	70.4 ± 3.9
		P	< 0.001	< 0.001
12	16	C(16)	6.68 ± 0.81	92.3 ± 0.82
		A(16)	3.67 ± 0.76	74.9 ± 4.3
		P	< 0.01	< 0.001
24	16	C(7)	2.76 ± 0.60	90.6 ± 3.7
		A(7)	2.56 ± 0.74	93.6 ± 0.9
		P	N.S.	N.S.

ION MIGRATION IN FROZEN SOIL

RICHARD P. MURRMANN, CAPTAIN, CmlC., and PIETER HOEKSTRA
U.S. ARMY COLD REGIONS RESEARCH AND ENGINEERING LABORATORY
HANOVER, NEW HAMPSHIRE

Radioactive waste management frequently involves placement of radioactive material into the ground either for storage or ultimate disposal (19). Extensive research on the behavior of radionuclide ions in soil coupled with experience at disposal sites has shown that this practice has not created a health hazard. However, an increasing population, the expansion in atomic energy utilization, and our concern over environmental pollution would indicate that site selection will become increasingly more difficult. Certainly, a continuing need exists for investigation of the suitability of potential locations for ground disposal of radioactive waste.

Radionuclides are transported away from the disposal site mainly by the flow of groundwater. The radionuclides dissolve in water and infiltrate through the soil. The ions normally move at a slower rate than the water due to chemical interactions of the ions with the soil constituents. The dispersion of radionuclides in soil and rock formations is thus highly dependent on the permeability of the site, and the amount and direction of ground water flow. With time, the radioactivity deposited at a site decreases due to radioactive decay and dilution by dispersion in the soil volume. For a good site the radioactivity in the ground should be reduced to a safe level before pollution of a potable water supply can occur.

Underground compartments in permafrost or artificially frozen ground may provide storage sites for radioactive waste (5). A similar procedure has already been utilized for storage of liquid natural gas (8). Since frozen ground is essentially impermeable to water, the migration of radionuclides due to ground water movement is eliminated. However, thermal degradation from heat produced during radioactive decay, and the possibility of ionic diffusion are both processes which should be investigated when considering frozen ground for the containment of radioactive waste. The object of this paper is to summarize our recent research results on ion migration in frozen ground.

Much of the present understanding of frozen ground (2) has been gained through extensive study of montmorillonite, a well characterized clay mineral found in soil. Montmorillonite crystals have a plate-like appearance with a high specific surface area of 800 m^2 per g which allows observation of maximal interfacial effects. The mineral has an exchange capacity of about one meq per g which means that one meq of cations is adsorbed by one g of soil. The negative charge is the result of crystal imperfections in the montmorillonite mineral. The electrical charge on montmorillonite and other soil minerals is, in part, responsible for the impedance in movement of radionuclide ions through soil.

Anderson and Hoekstra (4) determined the interlamellar spacing, the distance between the faces of adjacent clay plates, in Na-montmorillonite-water paste at different subzero temperatures. As shown in Figure 1, the interlamellar spacing remained constant at about 23 A (Angstroms) with decreasing temperature to about -5C where a rapid decrease in spacing to about 9 A occurred. The interlamellar spacing then continued to decrease slowly with decreasing temperature. Ice was first observed to form in the sample at a temperature of about -6C. Upon warming, the spacing increased slowly to 10 A at -1C and then expanded rapidly to the initial value of 23 A where ice was no longer present.

These results show that supercooling of water occurs in a clay-water system. When a clay-water paste freezes the water migrates to the extralamellar pore space where, as illustrated in Figure 2, ice forms. The exchangeable ions are concentrated in the thin films having a thickness equivalent to that of several monolayers of water. The concentration of ions as well as interfacial effects give the water in the unfrozen films a large freezing point depression. Further decreasing the temperature causes the migration of additional amounts of water and the formation of more ice. Upon heating, ice melts and water reoccupies the interlamellar region.

The total amount of liquid (unfrozen) water in frozen Na-montmorillonite paste at different temperatures has been determined by a number of investigators (1, 6, 10, 14, 18) but only the results of Nersesova and Tsytovich (18) cover the temperature range of interest. Their data are represented in Figure 3. Also shown for comparison is the temperature dependence of the liquid water content as calculated by Low *et al.* (16) from water adsorption isotherms and swelling pressure data. Although the two curves indicate a slightly different liquid water content at any given temperature, they are similar in that both show a rapid decrease in water content in the temperature range from -0.5C to -3C with a less rapid decrease in water content between -3C and -15C. The interlamellar water content, calculated using the interlamellar spacing as suggested by Anderson and Hoekstra (4), is also shown. In this calculation, it is assumed that the specific surface area of the clay is 800 m^2 per g and that the density of water is one g per cc. Thus, the estimated interlamellar water content is numerically equal to the interlamellar volume at any given temperature. By comparing the interlamellar and total liquid water contents, it is clear that most of the liquid water is present as thin films in the interlamellar region between clay plates. Some liquid water probably also exists in the extralamellar regions of the clay paste which include clay edge-clay edge, clay edge-clay face, and clay-ice interfaces, and also grain boundaries between ice crystals.

Recently, we determined the apparent self-diffusion coefficient of Na ions in frozen Na-montmorillonite using a radioactive tracer technique (17). Experimental conditions were selected which conformed to the model of diffusion in an infinite medium from an instantaneous planar source. The specific solution of the diffusion equation for these conditions can be written as,

$$-\log \frac{A}{M} = \log 2(\pi Dt)^{1/2} + \frac{X^2}{9.2Dt} \qquad (1)$$

where t is the time in sec, X is the distance in cm from the place of application of Na22, M is the total activity of Na22 deposited on the sample, A is the activity of Na22 as a function of X and t, and D is the apparent diffusion coefficient in $cm^2\ sec^{-1}$. The value of D was calculated from the slope of a plot of $-\log(A/M)$ versus X^2. The diffusion coefficients reported in this paper are termed "apparent" because the value is dependent on the sample matrix through which the ions must migrate. For example, both ion-surface interactions and an increased path length due to the presence of mineral particles lead to a lower value of the diffusion coefficient than would be observed for ion diffusion through bulk solution having properties identical to that of the solution near particle surfaces.

The value of the diffusion coefficient of Na ions in the frozen Na-montmorillonite paste at different temperatures is shown in Figure 4. The diffusion coefficient decreased from 8.8×10^{-7} cm^2 sec^{-1} at -0.6C to 4.3×10^{-7} cm^2 sec^{-1} at -3C and, subsequently, decreased less rapidly with decreasing temperature to 1.7×10^{-7} cm^2 sec^{-1} at -15C. Also shown in Figure 4 is the average value of 6 determinations made at a temperature of -3.1C during another experiment. This value of $4.02 \pm 0.16 \times 10^{-7}$ cm^2 sec^{-1} is slightly lower but still in good agreement with that of 4.25×10^{-7} cm^2 sec^{-1} at -3.2C obtained during the present experiment. The diffusion coefficient of Na ions in a comparable sample of montmorillonite has been reported to be 1.4×10^{-6} cm^2 sec^{-1} (21) at 25C which, surprisingly, is only a factor of ten more than that found for the frozen sample at -15C. Thus, the rate of diffusion in the frozen samples is relatively high.

Lattice diffusion coefficients of ions in both clay minerals, $D = 10^{-23}$ cm^2 sec^{-1} (7), and ice, $D = 10^{-11}$ cm^2 sec^{-1} (20), are much lower than those reported in this study. Therefore, Na ions did not migrate through either the clay particles or ice crystals present in the frozen clay paste. Rather, the ions must have migrated through the films of unfrozen water. Comparison of the temperature dependence of the diffusion coefficient of Na ions (Fig. 4) to that of the liquid water content of Na-montmorillonite (Fig. 3) indicates that the value of the diffusion coefficient is strongly related to the unfrozen water content or, more correctly since most of the water is located between clay particles, on the film thickness of the unfrozen water. As the film thickness decreases with decreasing temperature, the ions are confined closer to the clay surface where mobility is impeded (9, 15). Since temperature not only determines the thermal energy of the ion but also the thickness of the film through which the ion migrates, no activation energy can be assigned to the diffusion coefficient of Na ions in frozen Na-montmorillonite. To apply the theory of rate processes, the properties of the system through which the ion migrates must remain constant. The results of this experiment in which ions diffused over the entire length of the samples (5 cm) substantiate Hoekstra's (11, 12, 13) earlier conclusion that the films of unfrozen water are continuous throughout the frozen clay paste.

Electrical conductance measurements offer another method for investigating ion mobility in soil. The electrical conductance of the Na-montmorillonite paste used in the diffusion study was determined at a frequency of 1000 hertz after placing the clay paste between stainless steel electrodes in a glass capillary tube. The sample was first frozen at -16C after which the temperature was increased to 8C and, subsequently, again lowered to -10C. The conductance of the sample was determined frequently during the warming and cooling cycles. The specific weight conductance, $\overline{L}$, was calculated from,

$$\overline{L} = \frac{\ell}{BW} L_T \qquad (2)$$

where B is the cross-sectional area of the electrode in cm^2, ℓ is the distance between electrodes in cm, W is the weight of dry clay in g per cm^3 of clay-water paste, and L_T is the observed total conductance in ohm^{-1}.

The temperature dependence of the specific weight conductance is shown in Figure 5. Upon warming the sample from -15C, the conductance increased continuously until a temperature of about -0.5C, the freezing point (16), was reached. Thereafter, the conductance increased much more slowly. Cooling the sample resulted in a decrease in conductance until at a temperature of about -6C, the conductance decreased sharply to near the value obtained during the warming cycle. From the previous discussion on the nature of a frozen clay-water system, we deduce that the increase in conductance during the warming cycle is related to an increasing film thickness of unfrozen water until all ice melted at about -0.5C. The further increase in conductance at temperatures above -0.5C undoubtedly corresponds to an increase in thermal energy of the charge carrier. Finally, the marked hysteresis in the cooling curve between -0.5C

and -6C can be accounted for by the supercooling of water and the eventual nucleation of ice. Thus, like the diffusion coefficient of Na ions, the electrical conductance of frozen Na-montmorillonite paste is highly dependent on temperature.

We now wish to inquire about the nature of the current carrying element in the frozen clay paste. Earlier, Hoekstra (11) described a model for use in calculating electrical conductance from the ionic mobility in a frozen clay-water system. He considered the different components of the system, clay-ice-unfrozen water, to be parallel conductors and assumed an arrangement of clay plates parallel to the current flow with individual clay plates spaced by continuous film of unfrozen water of equal thickness at a given temperature. This arrangement is the same as depicted in Figure 2. The combined conductance of the ice and clay particles was considered to be negligible compared to that of the films of unfrozen water containing counter ions. The specific weight conductance was found to be related to the ionic mobility by an equation equivalent to

$$\overline{L} = \frac{\mu \bar{\sigma} F}{2} \qquad (3)$$

where μ is the ionic mobility in $cm^2\ sec^{-1}\ volt^{-1}$, $\bar{\sigma}$ is the exchange capacity of the clay in eq g^{-1}, and F is the Faraday in coul eq^{-1}. Unfortunately, there are no data available on the mobility of ions in a frozen clay-water system. The ionic mobility, however, can be calculated through use of the Nernst-Einstein equation,

$$\mu = \frac{zFD}{RT} \qquad (4)$$

where z is the formal charge per ion, R is the gas constant in joule $mole^{-1}\ deg^{-1}$, and T is the absolute temperature. Combining equations 3 and 4,

$$L = \frac{z \bar{\sigma} F^2 D}{2RT} . \qquad (5)$$

Thus, the specific weight conductance of frozen Na-montmorillonite paste can now be computed using the Na ion diffusion coefficients shown in Figure 4. The results of this calculation have been included in Figure 5 for comparison with the experimental conductance.

The experimental specific weight conductance of the frozen Na-montmorillonite paste is somewhat higher at any given temperature than that calculated from the Na ion diffusion coefficients. This may be due to the presence of charge carriers other than Na ions such as protons. Certainly, when the ideality of equation 5 used in the calculation is considered, the agreement between the observed and calculated values is reasonably good. This result would support the conclusion that the mobility of Na ions was high enough to account for much of the conductance of the frozen clay paste. Conductance measurements do offer a simple method for approximating diffusion coefficients in such a system.

So far only frozen clay-water systems have been discussed. However, we have also studied ion migration in frozen soil. Using a procedure similar to that already described (17), the apparent diffusion coefficient of Na ions in both frozen Fairbanks and Barrow silt was determined at different temperatures. These soils are found in permafrost regions, in interior and northern Alaska, respectively. As shown in Figure 6, the diffusion coefficient for Na ions was virtually the same in both soils at any given temperature. The value of the diffusion coefficient decreased from $10 \times 10^{-7}\ cm^2\ sec^{-1}$ at -0.6C to about $1 \times 10^{-7}\ cm^2\ sec^{-1}$ at -2C and, thereafter, decreased less rapidly with decreasing temperature to about $0.5 \times 10^{-7}\ cm^2\ sec^{-1}$ at -6C. Qualitatively, the temperature effect on the diffusion coefficient of Na ions in the soils was remarkably similar to that observed for the frozen montmorillonite-water system. This resemblance is

not surprising since the film thickness of unfrozen water through which ions diffuse is primarily dependent on temperature (3). Moreover, both soils contain clay minerals of the type similar to montmorillonite. The value of the diffusion coefficient of Na ions was lower in the soils than in the clay at comparable temperatures. Although this may be due to several factors, the larger particle size of the silt soils is thought to result in a longer path length for ion diffusion and, as explained earlier, a smaller apparent diffusion coefficient.

The observations reported in this paper would support the conclusion that ions migrate in frozen soil by diffusion through continuous thin films of unfrozen water located adjacent to particle surfaces. Although the rate of diffusion is lower in frozen than in unfrozen soil having the same total moisture content, this rate is still quite high when compared to solid state diffusion. Thus, frozen soil cannot be considered as an impermeable storage site for radioactive waste. However, since solution does not infiltrate through frozen soil, frozen soil offers a distinct advantage over unfrozen soil for radioactive waste containment.

LITERATURE CITED:

(1) Anderson, D.M. (1966) Phase composition of frozen montmorillonite-water mixtures from heat capacity measurements, *Soil Sci. Soc. Amer. Proc.*, 30:670-675. Also USA CRREL Research Report 218.

(2) Anderson, D.M. (1967) The interface between ice and silicate surfaces, *Jour. Coll. and Interface Sci.*, 25:174-191. Also USA CRREL Research Report 219.

(3) Anderson, D.M. (1967) Ice nucleation and the substrate-ice interface, *Nature*, 216:563-566.

(4) Anderson, D.M. and Hoekstra, P. (1965) Migration of interlamellar water during freezing and thawing of Wyoming bentonite, *Soil Sci. Soc. Amer. Proc.*, 29:498-504. Also USA CRREL Research Report 192.

(5) Brown, J. (1966) Massive underground ice in northern regions, *Army Sci. Conf. Proc.*, 2:89-102.

(6) Buehrer, T.F. and Aldrich, D.G. (1946) Studies in soil structure. VI: Water bound by individual soil constituents as influenced by puddling, University of Arizona Tech. Bul. No. 110.

(7) de Haan, F.A.M., Bolt, G.H., and Pieters, B.G.M. (1965) Diffusion of potassium-40 into an illite during prolonged shaking, *Soil Sci. Soc. Amer. Proc.*, 29:529-530.

(8) de Nevers, N., Liquid natural gas, *Scientific American*, p. 30-37, Oct. 1967.

(9) Gast, R.G. (1963) Relative effects of tortuosity, electrostatic attraction, and increased viscosity of water on self-diffusion rates of cations in bentonite-water systems. New York: Pergamon Press, *Proc. 1963 Int. Clay Conf.*, 1:251-259.

(10) Hemwall, J.B. and Low, P.F. (1955) The hydrostatic repulsive force in clay swelling, *Soil Sci.*, 82:135-145.

(11) Hoekstra, P. (1965) Conductance of frozen bentonite suspensions, *Soil Sci. Soc. Amer. Proc.*, 29:519-522.

(12) Hoekstra, P. and Chamberlain, E. (1964) Electro-osmosis in frozen soil, *Nature*, 203:1406.

(13) Hoekstra, P. and Keune, R. (1967) Pressure effects on conductance of frozen montmorillonite suspensions. New York: Pergamon Press, *Clays and Clay Minerals*, 15th Conf., p. 215-225.

(14) Kolaian, J.H. and Low, P.F. (1960) Thermodynamic properties of water in suspensions of montmorillonite. New York: Pergamon Press, *Clays and Clay Minerals*, 9th Conf., p. 71-84.

(15) Low, P.F. (1962) Influence of adsorbed water on exchangeable ion movement. New York: Pergamon Press, *Clays and Clay Minerals*, 9th Conf., p. 219-228.

(16) Low, P.F., Anderson, D.M., and Hoekstra, P. (1967) Some thermodynamic relationships for soils at or below the freezing point: I. Freezing point depression and heat capacity. *Jour. Geophys. Res.* (In press). Also USA CRREL Research Report 222, Parts 1 and 2.

(17) Murrmann, R.P., Hoekstra, P., and Bialkowski, R.C. (In press) Self-diffusion of Na ions in frozen Wyoming bentonite-water paste, *Soil Sci. Soc. Amer. Proc.*

(18) Nersesova, Z.A. and Tsytovich, N.A. (1966) Unfrozen water in frozen soils, *Proc. 1st Int. Conf. Permafrost,* NAS-NRC, Pub. No. 1287, p. 230-237.

(19) Parsons, P.J. (1965) Radioactive waste disposal into the ground, *Safety Series Publication No. 15,* IAEA, Vienna.

(20) Ramseier, R.O. (1967) Self-diffusion of tritium in natural and synthetic ice monocrystals, *Jour. App. Phys.*, **38**:2553-2556. Also USA CRREL Research Report 232.

(21) van Schaik, J.C., Kemper, W.D., and Olsen, S.R. (1966) Contribution of adsorbed cations to diffusion in clay-water systems, *Soil Sci. Soc. Amer. Proc.*, **30**:17-22.

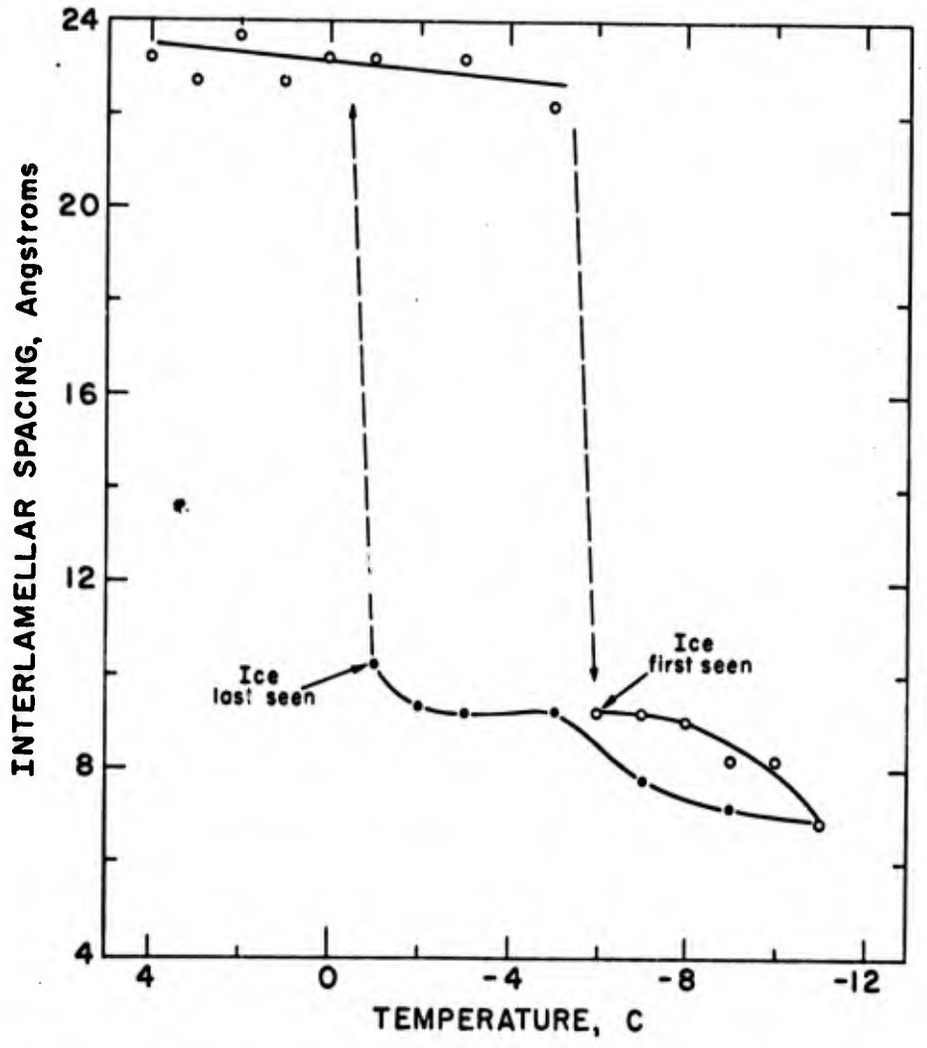


Figure 1. Interlamellar spacing of Na-montmorillonite as a function of temperature.

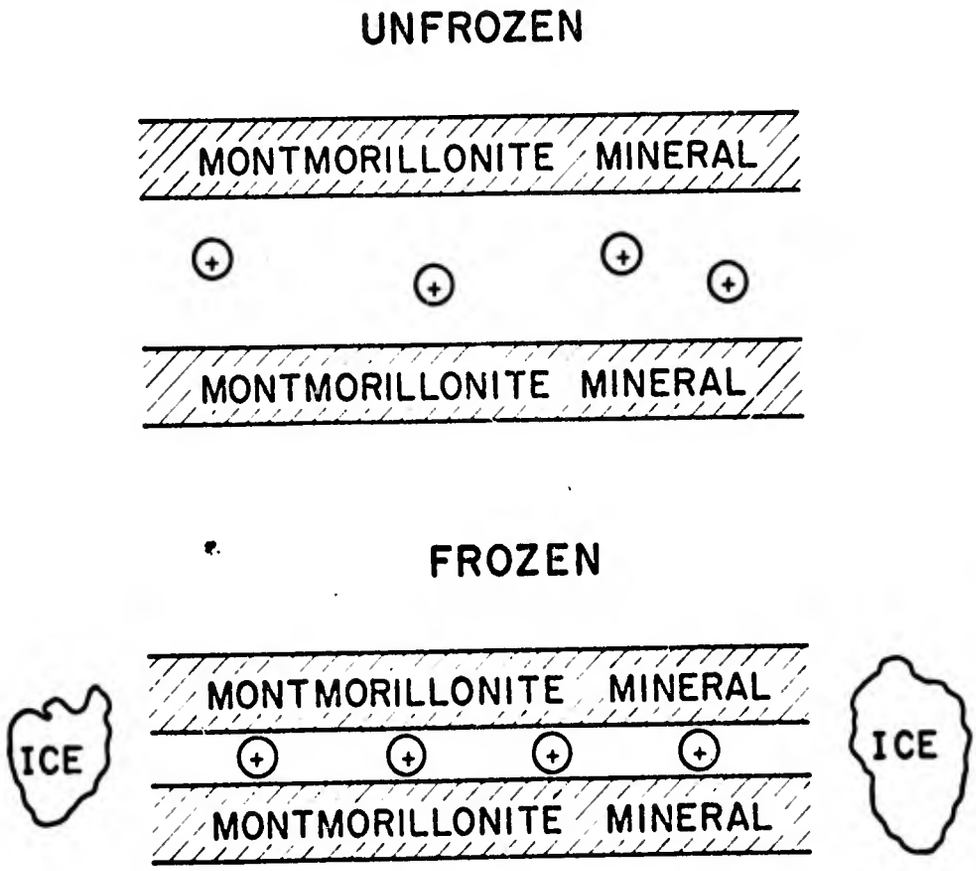


Figure 2. Schematic of a montmorillonite-water system upon freezing.

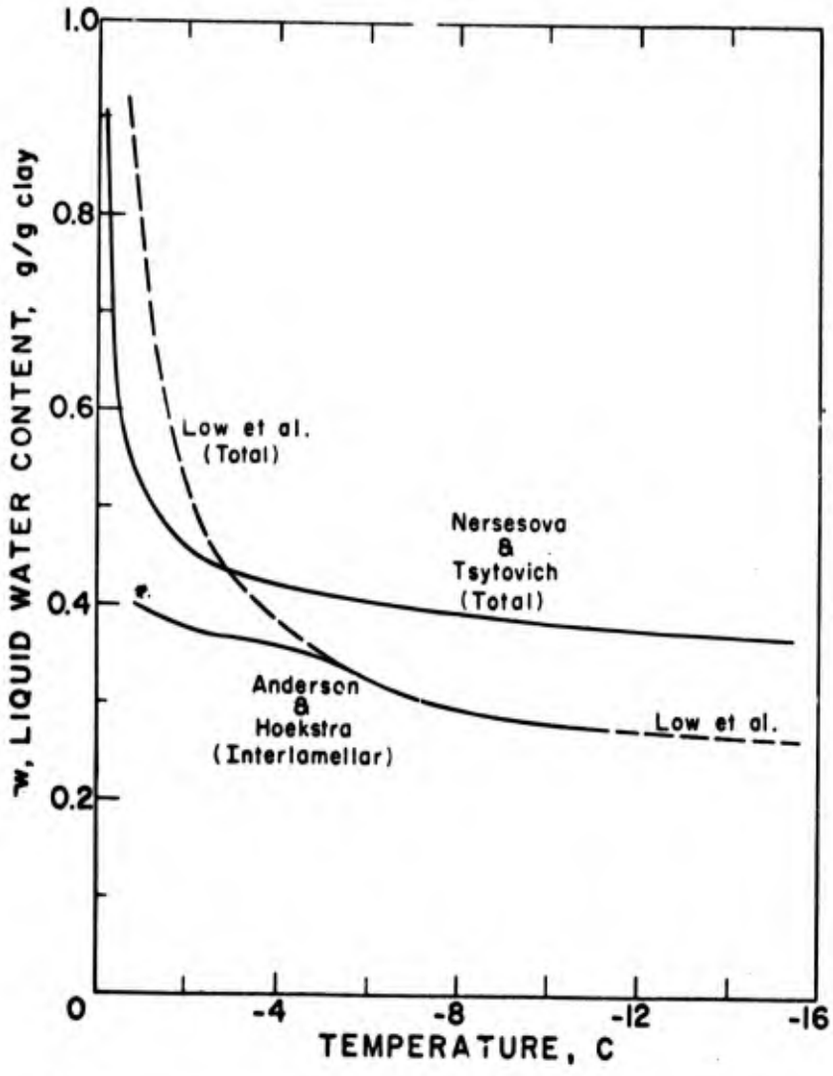


Figure 3. Liquid water content of frozen Na-montmorillonite as a function of temperature.

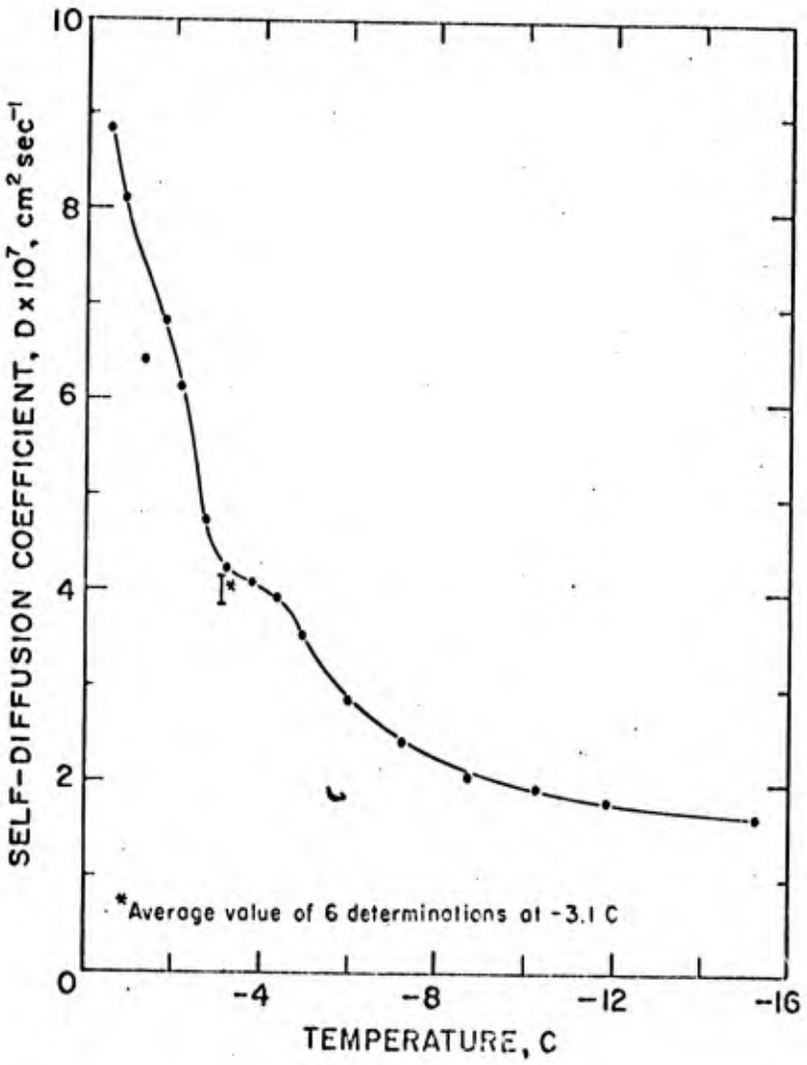


Figure 4. Self-diffusion coefficient of Na ions in frozen montmorillonite-water paste as a function of temperature.

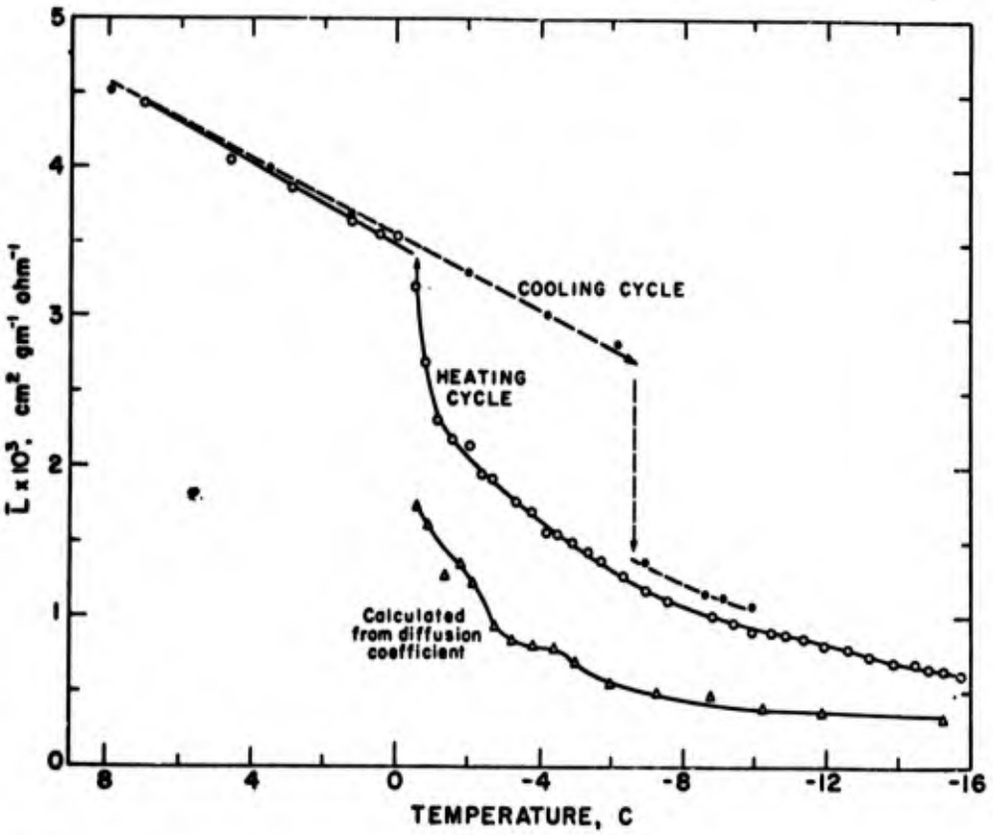


Figure 5. Experimental and calculated specific weight conductance of Na-montmorillonite as a function of temperature.

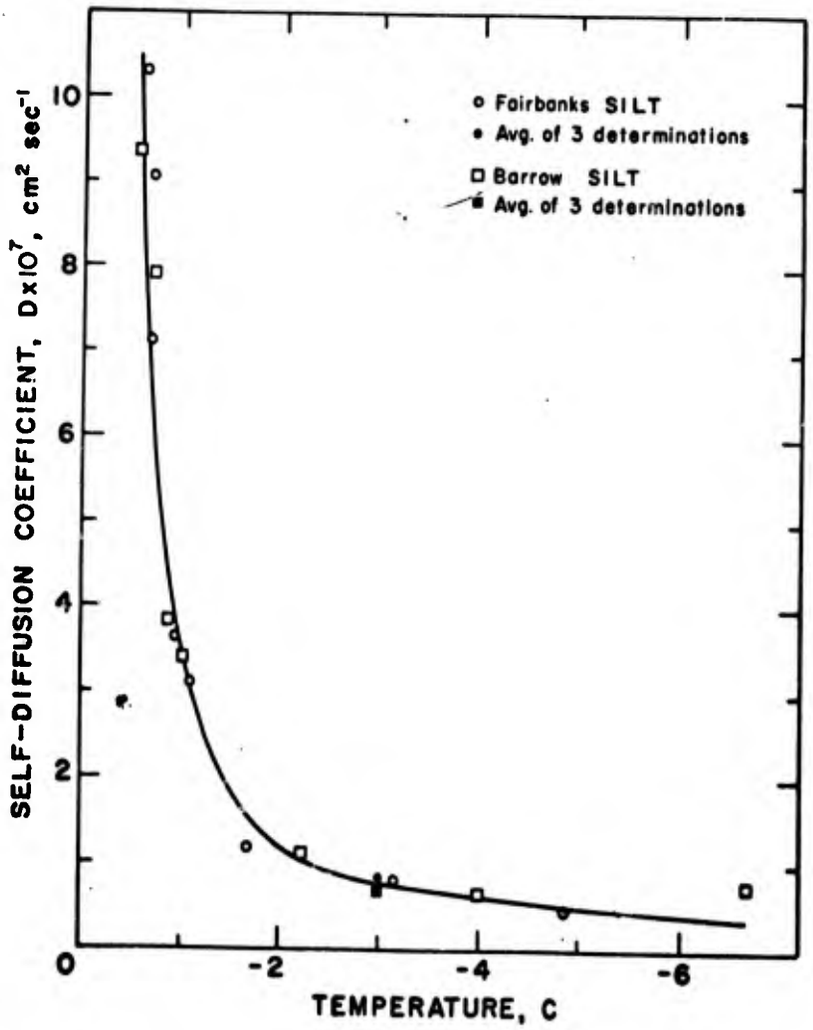


Figure 6. Diffusion coefficient of Na ions in frozen Fairbanks and Barrow silt as a function of temperature.

NEMARICH

ASSOCIATED YTTERBIUM AND NIOBATE IMPURITIES IN CALCIUM TUNGSTATE

JOSEPH NEMARICH
HARRY DIAMOND LABORATORIES
WASHINGTON, D. C. 20438

Calcium tungstate containing trivalent neodymium ions is one of the few optically pumped laser materials that exhibits continuous laser action at room temperature. In addition, laser emission has been obtained from $CaWO_4$ containing various other rare earth ions.(1) A good understanding of the behavior of rare earth impurity ions in this crystal is desirable since many of the problems encountered are common to all laser host crystals. A foremost problem for these materials is that of understanding how the crystal interacts with the impurity ions to produce the observed optical properties. Crucial to calculations bearing on this problem is a knowledge of the location of the impurity ion in the crystal and in particular the relative location of the ion with respect to other impurities. From the electron paramagnetic resonance spectra of the trivalent ytterbium ion in single crystals of $CaWO_4$ we have been able to identify the location of Yb^{3+} with respect to certain other impurities introduced into this crystal. In this paper we will discuss the details of the spectra obtained, and how the results are used to identify the location of associated impurities. The relation of these results to the problems of the optical spectra of laser crystals is also discussed.

When a rare earth impurity ion is present in an insulating crystal, the usual approximation made is that the host crystal acts as an array of point charges which creates an inhomogeneous electric field at the impurity ion site. The electrons of the impurity ion have a potential in this crystal field which may be expanded in terms of spherical harmonics.(2) The coefficients of this expansion are usually referred to as crystal field parameters. The energy states of the ground 4f configurations of the trivalent rare earth ions are highly degenerate, and the crystal field acts to remove much of this degeneracy. The extent to which the various degenerate energy states are split depends strongly on the magnitudes of the individual crystal field parameters. In most cases the crystal field will also mix energy states from other configura-

tions into the ground 4f configuration. The effect of this mixing is that previously highly forbidden 4f to 4f optical transitions now become allowed transitions. The crystal field perturbation thereby results in a much richer optical spectrum for ions in crystalline hosts.

The energy states of trivalent rare earth ions with an odd number of 4f electrons have a residual degeneracy that can be removed by a magnetic field. Crystals containing these ions will in most cases show electron paramagnetic resonance absorption signals in the microwave region of the spectrum.(3) If the symmetry of the crystal field potential is lower than cubic, these EPR signals will be highly anisotropic. That is, for a given microwave frequency, the value of the magnetic field at which a resonance occurs will depend on the orientation of the magnetic field with respect to the crystalline axes. The symmetry of the field for resonance will reflect the symmetry of the crystal field at the impurity ion and much may be inferred about the position of the ion in the crystal from a study of its paramagnetic resonance spectrum. An effective g-factor may be calculated for the magnetic field oriented along different crystal directions from the ratio of the microwave frequency and the magnetic field for resonance. These g-factors may then be used together with optical spectra (or in many cases, independently) to identify the ground state of the ion. The effective g-factor of the ion depends on the magnitude of the crystal field parameters and can be used together with the optical spectrum of an ion to obtain a set of crystal field parameters that best describes the optical and EPR spectra.

A trivalent rare earth ion will often enter a crystal substitutionally for a divalent ion. This is the situation in $CaWO_4$ where trivalent rare earth ions substitute for divalent calcium ions.(4) In these cases, the crystal will generate charge defects during growth, or charge compensation may be accomplished by adding other impurity ions to the crystal. The rare earth ion may then be located in several types of crystallographic sites. For example, in $CaWO_4$, the ion may be located at a calcium site with either no other charge defects nearby or a charge defect located in one of several possible nearby sites. Since the crystal field seen by the ion depends on its local environment, in general a different set of crystal field parameters will pertain to each type of crystal site. The resultant optical and EPR spectra for the crystal will then be a superposition of spectra from the ions in the various types of sites. The interpretation of optical spectra can be much more difficult when this occurs since it is not always possible to assign a set of spectral lines to ions in a particular type of site. Also, as a practical matter, when an ion can be located in several types of sites the threshold energy for laser action for the crystal may be raised.

However, if one can unambiguously determine the local environment of several types of sites for an ion in a crystal, it may prove beneficial. A comparison of the crystal field parameters for the various sites may aid in obtaining a better method for calculating the crystal field parameters from first principles. The need for improved calculations stems from the fact that although the point charge model for the crystal field gives qualitatively correct results, detailed calculations for the crystal field parameters from the properties of the host lattice usually differ considerably from calculation of these parameters from optical spectra.(5) Whereas one may not be able to calculate correctly the effect of the whole lattice on the impurity ion, calculation of a differential change in crystal field parameters may be possible. The higher resolution obtained in EPR spectra, as well as the dependence of the symmetry of the field for resonance on the site symmetry of the ion, suggests that the local environment of an ion can best be obtained from EPR spectra. We have used this approach in the investigation of Yb^{3+} in $CaWO_4$.

The EPR spectrum of Yb^{3+} in $CaWO_4$ was obtained with an x-band paramagnetic resonance spectrometer constructed at this laboratory. Samples used for the measurement were grown by the Czochralski method and cut in blocks 5 x 5 x 3 mm or larger. Prior to cutting, the samples were oriented by X-ray methods to $\pm$ 1 deg or better. The samples were mounted on a sapphire post on the axis of a TE_{011} mode cavity, which had an unloaded Q of about 24,000. The cavity was tuned to near critical coupling by a vane tuner, and the signal klystron was locked to the sample cavity resonant frequency. All measurements were made at a temperature of 4.2°K. To avoid saturation of the resonance signals, very low input power (about 10^{-8} W) and superheterodyne detection were used. The magnetic field was obtained from a 12-in. electromagnet and could be rotated in a plane perpendicular to the axis of the sample cavity. Field modulation at 36 Hz was employed, and the static field was measured with a nuclear magnetic resonance magnetometer and electronic counter.

The EPR spectrum of Yb^{3+} in $CaWO_4$ containing niobium is shown in figure 1. The spectrum was obtained with the magnetic field parallel to the crystal [110] axis and at a microwave frequency of 9.3 GHz. The sample was prepared with 0.05 atomic percent ytterbium and 0.10 atomic percent niobium. Resonance lines marked Tetragonal Sites are due to Yb^{3+} in calcium sites with no other impurities nearby.(6) These lines have the tetragonal symmetry characteristic of the point symmetry of the calcium site. The strong central line of the tetragonal spectrum is due to the Yb^{3+} ions with nuclear spin zero and the remaining eight lines are due to Yb^{3+} isotopes with nuclear spins 1/2 and 3/2. The other strong lines in the spectrum appear only in samples prepared with

niobium.(7) To insure that these lines are not due to another impurity, crystals were prepared with spectroscopic grade niobium and found to have the same strong additional lines. These lines were presumed to be due to $(NbO_4)^{3-}$ complexes which substituted for $(WO_4)^{2-}$ complexes near the ytterbium ions. The resultant perturbation of the crystal field by the charge defect causes a shift in the effective g-factor of the ytterbium ion.

To determine the location of the $(NbO_4)^{3-}$ complexes responsible for the individual additional lines in the spectrum, the fields for resonance in principal crystallographic planes were examined. As the field was moved away from the [110] axis, it was found that many of the lines split into pairs and some of the strong lines became weak and some weak lines became strong. Many of the lines were too weak to follow for more than a few degrees from the principal crystal axes, but it was possible to trace the fields for resonance of six of the additional lines shown in figure 1. It was found that these six lines could be grouped into three sets, which we refer to as spectra I, II, and III. As the magnetic field was moved out of the crystal (001) plane, each of the six lines split into a pair. The splitting for spectrum I can be seen in figure 2, which shows the angular dependence of the magnetic field for resonance as the field is varied in the $(1\bar{1}0)$ plane. Two of the lines are each degenerate when the field is along the [110] direction and split into pairs of lines as the field moves toward the crystal c-axis. When the field is along the c-axis, all four lines coalesce into a single line. From figures 3 and 4 it is seen that spectra II and III have a similar behavior. The data shown in figures 2, 3, and 4 were taken at a frequency of 8.8 GHz. The variation of the resonance field in the plane perpendicular to the c-axis is shown for all three spectra in figure 5. Here it is seen that each pair of lines coalesce at approximately 10 deg from the crystal [100] axis and are maximally separated at about 55 deg from the [100] axis. These were taken at a frequency of 8.6 GHz. The effective g-factors along the [001] and [110] directions for the spectra I, II, and II and for the tetragonal sites are summarized in table I. The resonance for the tetragonal sites is isotropic in the (001) plane.

	I	II	III	Tetragonal
$g_{[001]}$	1.90	1.51	1.37	1.06
$g_{[110]}$	3.45	2.04	2.78	3.92
	4.42	5.99	5.08	

Table I. Effective g-factors along crystal [001] and [110] directions for various spectra in $CaWO_4$: Yb,Nb.

The relative location of Yb^{3+} and $(NbO_4)^{3-}$ can be inferred from these spectra by considering the arrangement of atoms in a $CaWO_4$ crystal. Figure 6 shows two adjacent unit cells of $CaWO_4$ with a Yb ion replacing a calcium ion at the center of one of the unit cells. $CaWO_4$ has a tetragonal structure with the unit cell having dimensions of 5.24 Å along the a- and b-axes and 11.38 Å along the c-axis. Each tungsten atom is at the center of a tetrahedron of oxygen atoms, which together form a covalently bonded $(WO_4)^{2-}$ complex. For clarity most of the oxygen tetrehedra have been omitted from the figure. When niobium is added to the crystal for charge compensation, it substitutes for tungsten to form a $(NbO_4)^{3-}$ complex. Consider substitution of a $(NbO_4)^{3-}$ complex for one of the first near neighbor $(WO_4)^{2-}$ complexes such as that marked 1 in figure 6. There are four such first near neighbor positions, all of which lie in a plane perpendicular to the crystal c-axis and containing the Yb ion. It is to be expected that such a substitution would perturb appreciably the crystal field potential of the Yb^{3+} ion and give rise to a g-shift. In general the charge associated with the oxygens in different covalently bonded complexes such as $(WO_4)^{2-}$ and $(NbO_4)^{3-}$ will vary. (8) Since the crystal field parameters will go approximately as some inverse power of the distance to the perturbing charge, (2) the nearest oxygens to the Yb^{3+} can be expected to have the greatest effect. In figure 7 the nearest oxygen in each of the first nearest neighbor $(WO_4)^{2-}$ is shown projected onto the (001) plane containing the Yb ion. Since substitution of a $(NbO_4)^{3-}$ in any one of the first near neighbors is equally likely, the composite spectrum from all such first near neighbor associated ytterbium and niobate impurities will have four-fold symmetry in the (001) plane. Because the field for resonance is not dependent on the sense of the magnetic field, the four types of associates will give rise to only two lines when the field is in a (001) plane. Also from figure 7, one can see that if indeed the main perturbation arises from the nearest oxygen, all four first near neighbor types of associates will be equivalent when the field is a few degrees from the [100] or the [010] direction. When the magnetic field is moved in the $(1\bar{1}0)$ plane, since the nearest oxygens are not located in the (001) plane, all four of the first near neighbor type associates will be inequivalent and give rise to four resonance lines. When the field is along the c-axis, all four lines should coalesce since all four types of sites would again appear equivalent.

A similar argument can be made for $(NbO_4)^{3-}$ complexes substituting for second nearest neighbor $(WO_4)^{2-}$ (marked 2a and 2b in figure 6), and for alternate fourth nearest neighbor $(WO_4)^{2-}$. There are eight fourth nearest $(WO_4)^{2-}$, but half of these have their nearest oxygen at one distance and the other half at a

farther distance. Only the closest oxygen from the fourth nearest neighbors $(WO_4)^{2-}$ is shown projected in figure 7. Substitution of a $(NbO_4)^{3-}$ for a $(WO_4)^{2-}$ in either a first, second, or fourth nearest neighbor position might be expected to give qualitatively the type of spectra shown in figures 2 to 5. Since the perturbing charges in the three possible types of sites are at varying distances, one would expect the magnitude of the g-shift to vary for each type of $(NbO_4)^{3-}$ substitution.

No strong spectrum was observed that could be correlated with a $(NbO_4)^{3-}$ substituting in the third near neighbor $(WO_4)^{2-}$ located either axially above and below the Yb ion. A substitution of a $(NbO_4)^{3-}$ complex in this position would give a g-shift, but the spectrum should be axially symmetric.

The many additional weak lines seen in figure 1 are no doubt due in part to $(NbO_4)^{3-}$ complexes substituting in more distant $(WO_4)^{2-}$ sites (including the alternate fourth near neighbor position discussed above). Some of these lines, however, are hyperfine lines associated with the central lines of spectra I, II, and III.

Although the proposed model gives the correct qualitative characteristics of the observed spectra, a quantitative verification would also be desirable. Using the reported crystal field parameters for Yb^{3+} in tetragonal sites, it should be possible to make some simplifying assumptions about the charge distribution in the $(NbO_4)^{3-}$ and $(WO_4)^{2-}$ complexes and compute the resultant change in crystal field parameters and g-factors. In this way the various assumed nearby $(NbO_4)^{3-}$ substitutions could be correlated with spectra I, II, and III. To pin down some of the many degrees of freedom allowed in assigning a charge distribution to the $(WO_4)^{2-}$ and $(NbO_4)^{3-}$ complexes, a comparison of the optical spectra of Yb^{3+} in $CaWO_4$ with and without niobium might also be made. A calculated change in the crystal field parameters could then be checked against observed changes in both the EPR and the optical spectra. A preliminary look at the optical spectra indicates that this may be possible.

In summary, the paramagnetic resonance spectrum of Yb^{3+} in $CaWO_4$ containing niobium has been obtained. In addition to the tetragonal spectra of Yb^{3+} usually observed in $CaWO_4$, we have identified three additional spectra, which are due to Yb^{3+} in lower symmetry sites. The spectra have the behavior expected if a Yb^{3+} substitutes for a Ca ion and a $(NbO_4)^{3-}$ complex substitutes for a $(WO_4)^{2-}$ complex in either a first, second, or alternate fourth nearest neighbor position, if one assumes that the main perturbation of the $(NbO_4)^{3-}$ is effected by its oxygen nearest

the Yb^{3+} ion. This identification of three separate ytterbium-niobate associates in $CaWO_4$ may allow a correlation of EPR and optical spectra with the charge distribution in nearby $(NbO_4)^{3-}$ and $(WO_4)^{2-}$ complexes. A calculation of changes in crystal field parameters that explain both spectra may lead to improved methods for calculating crystal field parameters.

We would like to thank W. Viehmann and R. M. Curnutt for preparing the crystals used for this work, A. J. Edwards for X-ray orientation of the crystals, and J. P. Sattler for assistance in obtaining the data.

REFERENCES:

1. W. V. Smith and P. P. Sorokin, The Laser (McGraw Hill Book Co., New York, 1966), p. 124.

2. B. G. Wybourne, Spectroscopic Properties of Rare Earths (Interscience Publishers, New York, 1965), chapter 6.

3. W. Low, Paramagnetic Resonance in Solids (Academic Press, New York, 1960).

4. K. Nassau and A. M. Broyer, J. Appl. Phys. 33, 3064 (1962).

5. G. Burns, J. Chem. Phys. 42, 377 (1965).

6. U. Ranon and V. Volterra, Phys. Rev. 134, A 1483 (1964).

7. J. Nemarich and W. Viehmann, J. Phys. Chem. Solids 29, 57 (1968).

8. S. Karavelas and C. Kikuchi, Molecular Orbital Calculation of Vanadium in Tungsten Site of $CaWO_4$, Univ. Mich. Technical Report 06029-3-T (September 1965).

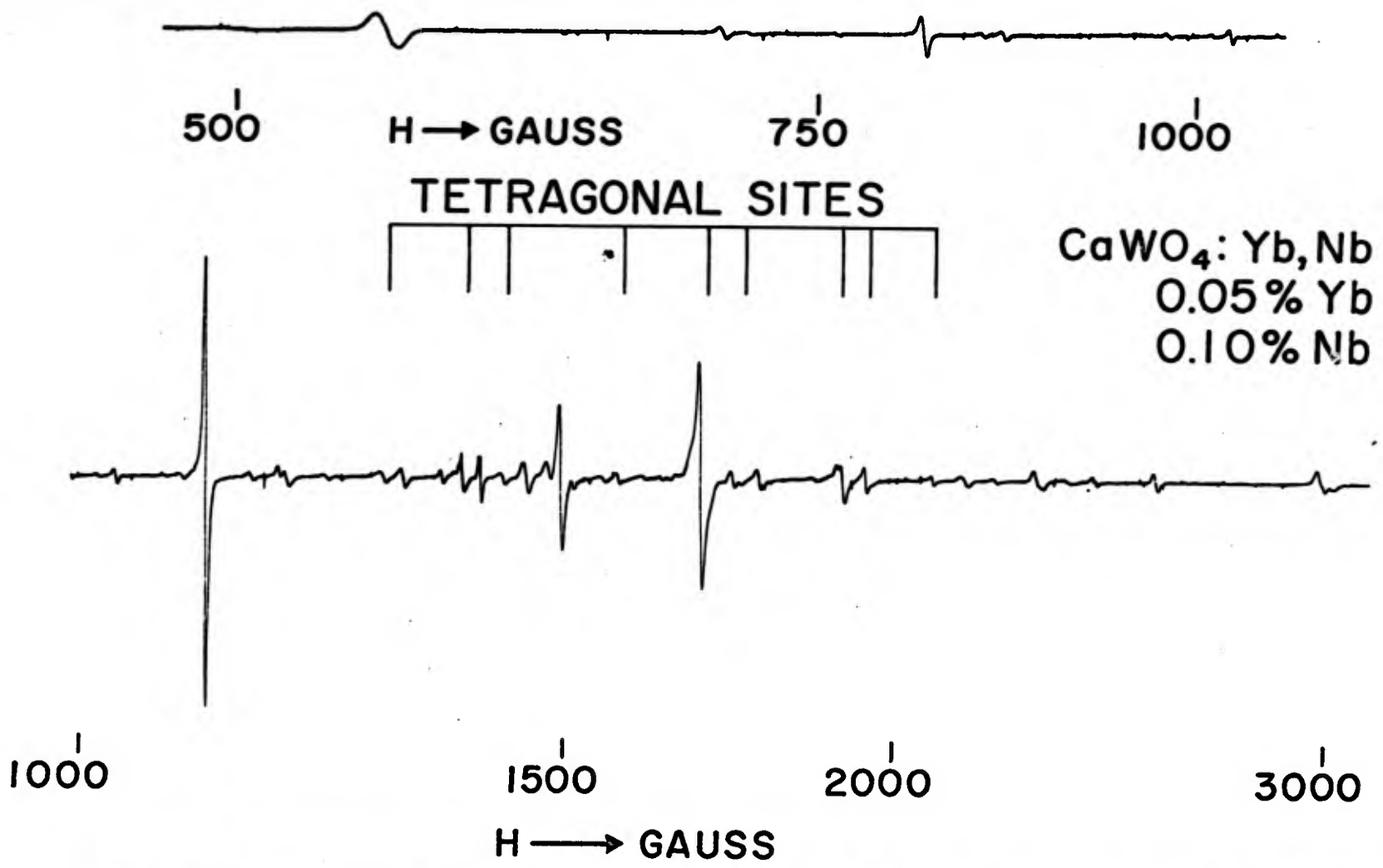


Figure 1. Paramagnetic resonance spectrum of $CaWO_4$:Yb,Nb with the static field along [110]. The low field portion of the spectrum is at the top of the figure.

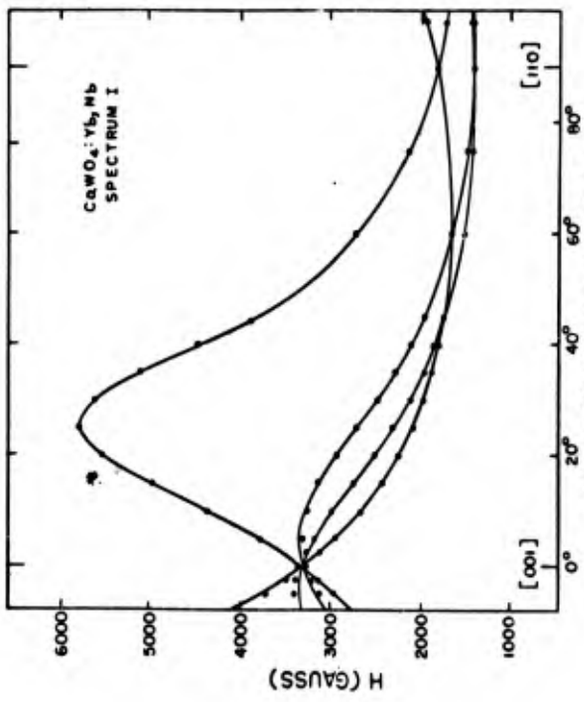


Figure 2. Angular dependence of the magnetic field for resonance of spectrum I in $CaWO_4$:Yb,Nb when the field is varied in the (1$\bar{1}$0) plane.

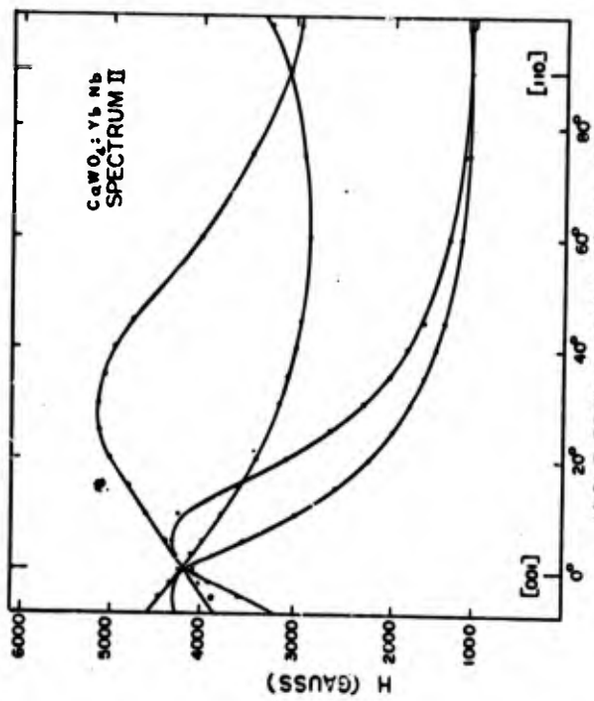


Figure 3. Angular dependence of the magnetic field for resonance of spectrum II in $CaWO_4$:Yb,Nb when the field is varied in the (1$\bar{1}$0) plane.

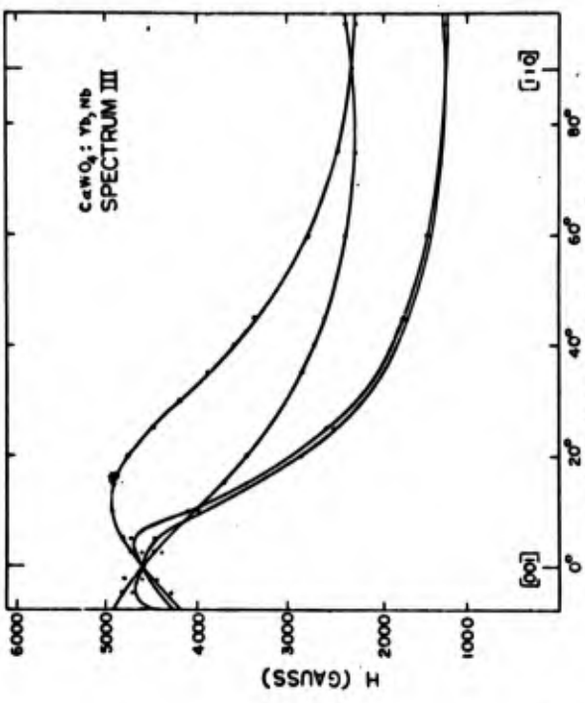


Figure 4. Angular dependence of the magnetic field for resonance of spectrum III in $CaWO_4$:Yb,Nb when the field is varied in the $(1\bar{1}0)$ plane.

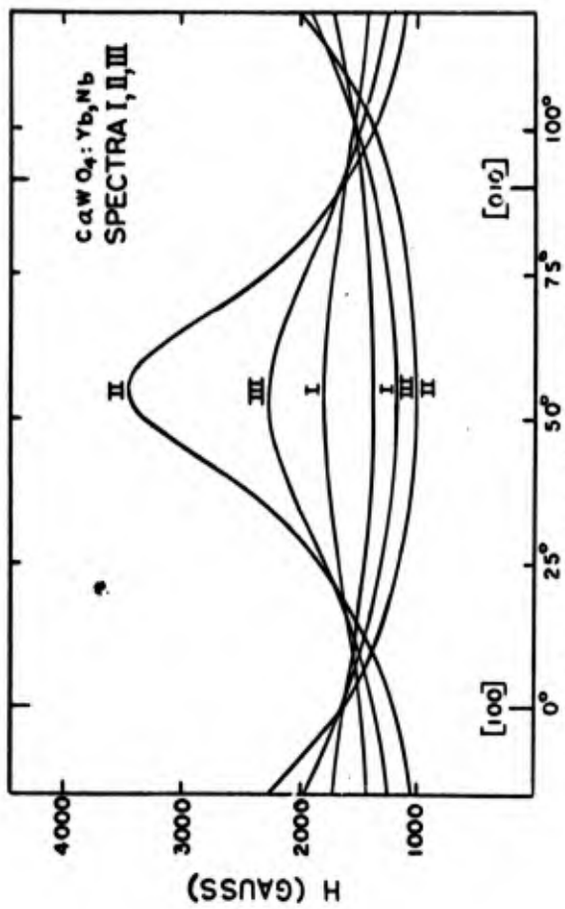


Figure 5. Angular dependence of the magnetic field for resonance of spectra I, II, and III in $CaWO_4$:Yb,Nb when the field is varied in the (001) plane.

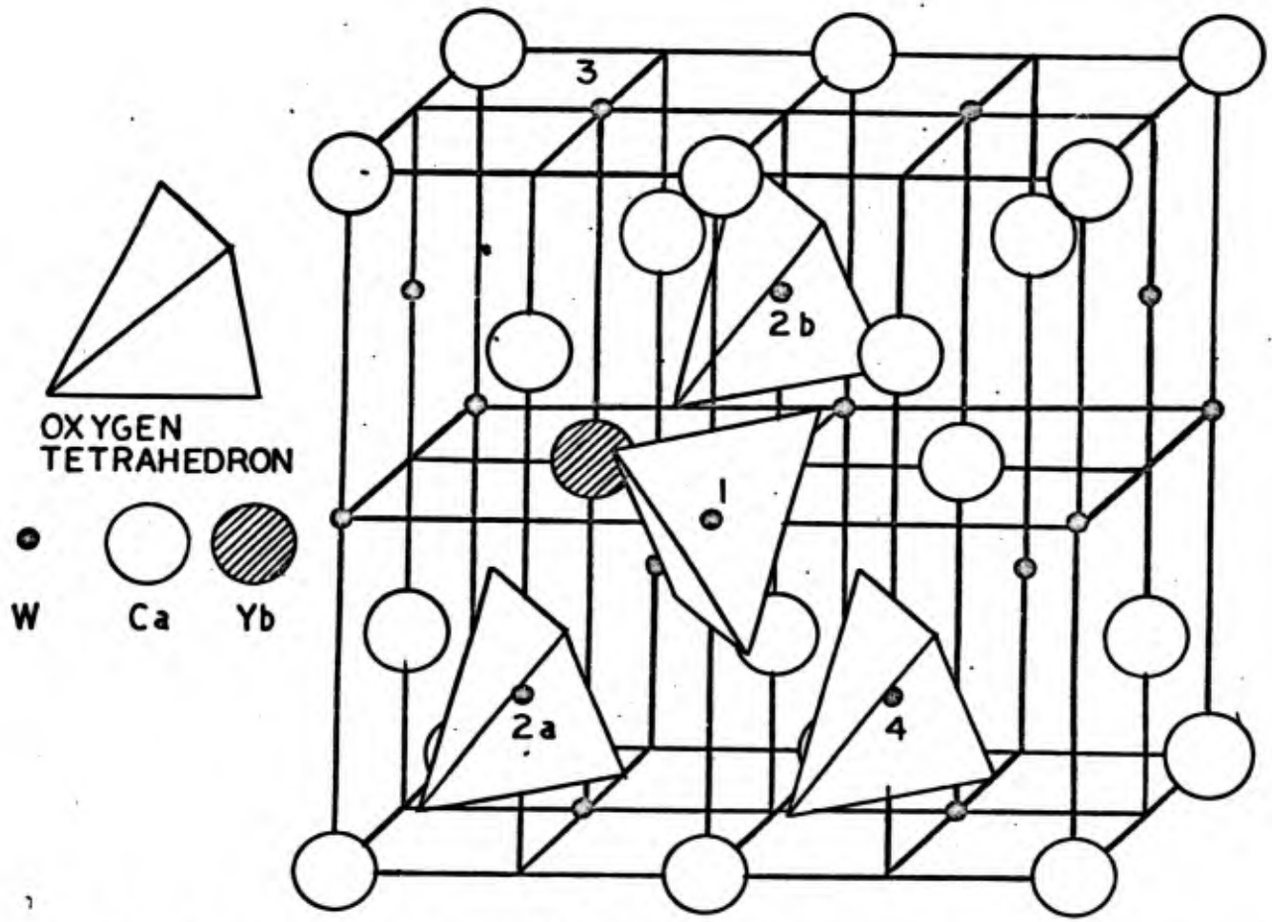


Figure 6. Two unit cells of $CaWO_4$ with a Yb ion shown substituted for a Ca ion. The oxygen tetrahedra for typical first, second, and fourth nearest neighbor tungstate complexes are shown schematically.

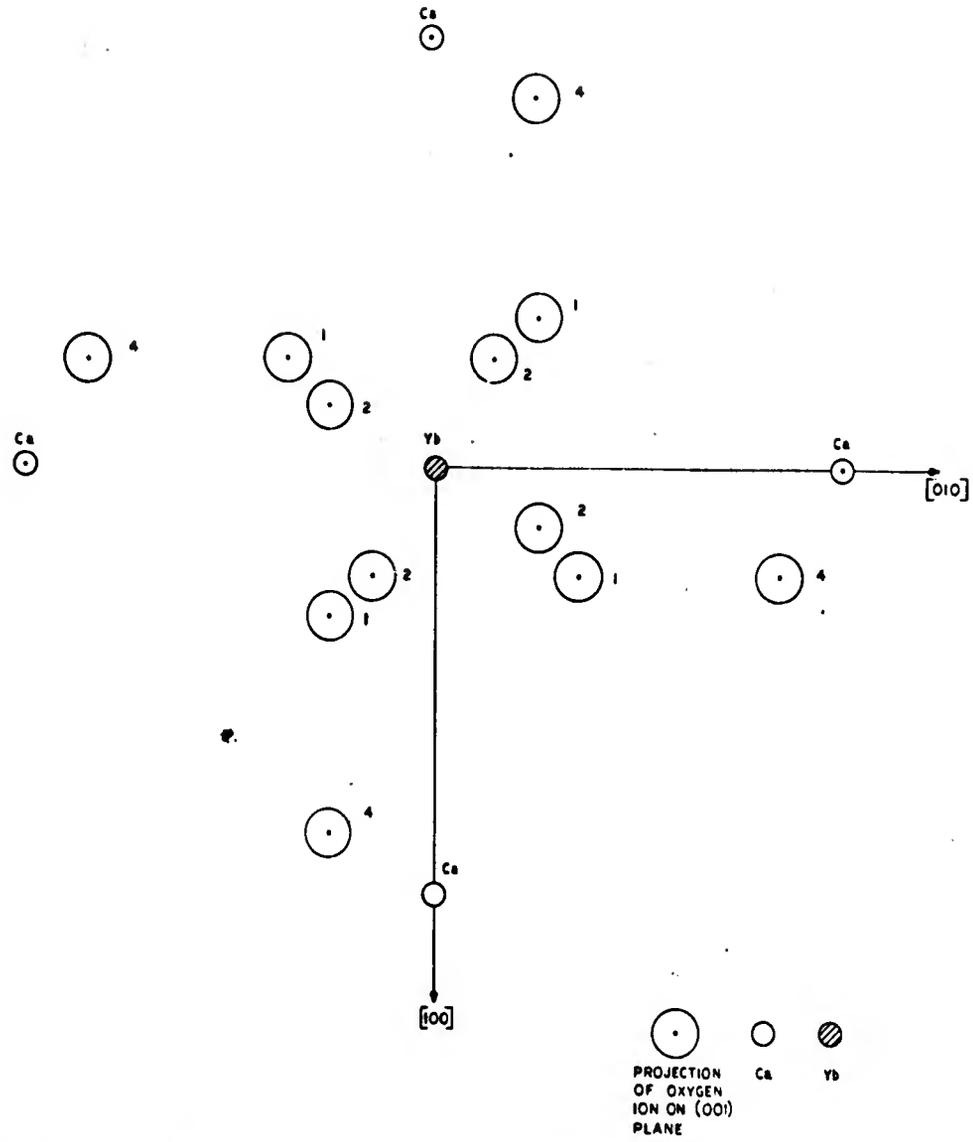


Figure 7. Calcium ions and the projection of oxygen ions in the (001) plane of $CaWO_4$. A Yb ion is shown substituted for a Ca ion and only the nearest oxygen of the first, second, and fourth nearest neighbor tungstate complexes are shown.

A NEW METHOD OF DEPOSITING SILICON CARBIDE FOR HIGH-MODULUS FIBERS

VERNON A. NIEBERLEIN
U. S. ARMY MISSILE COMMAND
REDSTONE ARSENAL, ALABAMA

INTRODUCTION

Fiber glass has been in use for about 30 years and was quick to find its way into aircraft structures and later into missiles. Solid propellant motor cases, nose cones, nozzles, and other structural components which are made from fiber glass composites are now common items. In recent years attention has been turned to materials other than glass, principally ceramics, from which fibers with improved properties could be made. Thus, with the proper fiber-matrix combination and proper design, aerospace components with very impressive modulus-to-density ratios or superior high-temperature properties can be made.

Ten of the favorite candidate materials for structural fibers are shown in Table I arranged in order of their modulus-to-density ratio.

TABLE I

DATA ON BULK MATERIALS

MATERIAL	DENSITY LBS./IN.^{3}	MODULUS OF ELASTICITY 10^6 LBS./IN^2	MODULUS TO DENSITY RATIO 10^8 INCHES	MELTING TEMPERATURE °F
BORON CARBIDE	0.091	65	7.15	4500
BERYLLIUM	0.066	44	6.65	2340
SILICON CARBIDE	0.115	70	6.10	4350
BORON	0.083	50	6.03	4200
BERYLLIUM OXIDE	0.103	55	5.33	4580
BERYLLIUM CARBIDE	0.088	45	5.08	3800
ALUMINUM OXIDE	0.137	60	5.00	3660
CARBON-GRAPHITE	0.076	14	1.84	7500
QUARTZ	0.083	11	1.32	3150
E-GLASS	0.092	10	1.10	1400

The materials shown have a range of properties, rather than a single sharply defined value, depending on the source of the material and the processing variables, and thus the order of merit is not rigidly fixed. This is especially true of carbon/graphite. The value in Table I is conservative for this material in that modulus values greater than 60×10^6 have been reported (Ref. 1).

A great deal of experience has been acquired in recent years in the technology of boron filaments and to a somewhat lesser extent in the technology of silicon carbide filaments and boron carbide filaments (Ref. 2). Silicon carbide has been chosen for experimentation by the Army Missile Command because of its very favorable potential in aerospace applications. SiC not only has a good modulus-to-density ratio, but also has excellent oxidation resistance up to 1500°C. This material is relatively inert as far as reacting with the most of the commonly proposed matrix materials. As in the case of boron filaments, silicon carbide filaments are normally prepared by vapor deposition.

Starting with Pring and Fielding in 1909 (Ref. 3) vapor deposition has been extensively explored as a means for coating objects with SiC. These early investigators deposited a coating onto a carbon rod using silicon tetrachloride and hydrogen or silicon tetrachloride, benzene and hydrogen. Their attempts with trichlorosilane were not successful. Moers (Ref. 4) coated a variety of substrates with SiC by passing silicon tetrachloride and toluene vapors over a heated surface. Kendall and Yeo (Ref. 5) successfully pyrolyzed substituted silane vapors, namely CH_3SiCl_3 and $(C_2H_5)_2SiCl_2$, to produce SiC coatings. They also obtained deposits from mixtures of hydrogen and the vapors of silicon tetrachloride and toluene. Silicon tetrachloride, hydrogen and toluene has also been used by Powell (Ref. 6) and Merz (Ref. 7). The latter investigator also used CH_3SiCl_3 and hydrogen or $SiCl_4$ + CCl_4 + H_2 at 1400°C. Susman, Spriggs, and Weber (Ref. 8) explored the pyrolysis of methylchlorosilane vapors. These included $(CH_3)_3SiCl$, $(CH_3)_2SiCl_2$, CH_3SiCl_3 and $(CH_3)_4Si$, and it was found that $(CH_3)_3SiCl$ was the most attractive of the group. Hough and Earley (Ref. 9) also investigated the first three of this group along with $CH_3Si(OCH_3)_3$. They found $(CH_3)_3SiCl$ to have a much higher deposition rate at equivalent temperatures. Huggins and Pitt (Ref. 10) also used $(CH_3)_3SiCl$ in their study of vapor deposited SiC, where they determined the effect of operating parameters on the character of the deposit. McCandless, Withers and Davies (Ref. 11), Withers and Redmond (Ref. 12), and McCandless and six co-workers (Ref. 13) tried silicon halides and various hydrocarbons,

and substituted silanes plus hydrogen for the production of continuous SiC-on-tungsten composite filaments by vapor deposition. Most of their work has centered about the use of $C_2H_5SiCl_3$ and hydrogen.

EXPERIMENTAL WORK AND RESULTS

Since the vapor deposition approach to SiC filaments had already been well explored to some depth by capable investigators, it was decided to try a liquid approach. Because deposition temperatures for any type of SiC coating process are far above the boiling points of liquid organic or inorganic compounds, any liquid chemical process would therefore involve film-boiling. A film-boiling process might be defined by some to be basically a vapor deposition process also, since the heated substrate is encased in a vapor sheath and out of contact with any liquid. However, in its engineering aspects a film-boiling process is treated as a liquid process since all of the bulk fluids are liquids.

An apparatus was constructed for performing film-boiling experiments on a stationary fiber as shown in Figure 1. The horizontal plating cell itself was constructed of quartz walls and brass end caps sealed with O-rings. A glass water jacket surrounded the cell to permit visibility and minimize liquid evaporation. A 3-inch tungsten wire was stretched between contacts, one of which was spring loaded to compensate for thermal expansion of the wire. The tungsten was resistively heated by AC as shown, or DC through Kovar seals in one end. A schematic diagram of the complete apparatus is given in Figure 2. The plating liquid was led into the cell from a reservoir until the liquid level was about $\frac{1}{4}$ to $\frac{1}{2}$ inch above the tungsten wire. The unfilled space above the liquid surface was connected to a surge tank through which helium flowed. A mercury U-tube manometer indicated pressure in the system.

Liquids tried in this system in film-boiling experiments were $SiCl_4$ + toluene, $SiCl_4$ + dichloroethylene, $SiCl_4$ + cyclohexane, SC-77 (a proprietary organosilane), $(CH_3)_3SiCl$, and CH_3SiCl_3. The best of these was $(CH_3)_3SiCl$ which produced a smooth fiber with a rapid plating rate and only very slight solution discoloration. A picture of the surface is shown in Figure 3. Table II shows the plating rates. A 2-mil tungsten wire could be built up with a silicon carbide coating to the diameters shown, in 30 seconds.

TABLE II

FIBER DIAMETERS AFTER 30 SECONDS PLATING

Temperature	Finished Diameter
925°C	0.004 inch
995°C	0.007 inch
1050°C	0.011 inch
1135°C	broke

These deposits were eccentric and contained cracks and voids and were excessively brittle. A marked improvement in the deposit was obtained if the $(CH_3)_3SiCl$ were diluted with 70% to 80% of inert solvent. Diluents tried were cyclohexane, trichloroethylene, 1,2-dichloroethane, acetone, diethyl ether, methyl alcohol, amyl acetate, and an equal-volume mixture of cyclohexane and acetone. The best of these was cyclohexane. While the use of a diluent slowed down the plating rate, it greatly increased the strength and modulus as well as minimizing cracks and voids. A photomicrograph of a fiber plated in this apparatus using 80% cyclohexane and 20% $(CH_3)_3SiCl$ is shown in Figure 4. This picture clearly shows the layered SiC structure and the eccentricity of the deposit. The system was placed under positive pressures of 10, 20, 40, 80, and 160 torr without affecting this eccentricity or the plating rate.

To demonstrate the usefulness of the process, the system was made continuous. The plating cell was redesigned to accommodate a moving vertical filament. A schematic diagram is shown in Figure 5. A tungsten wire enters the top of the cell and passes over the pegs as shown, three of which are Teflon and the third is platinum-clad brass. Electrical contact is made with the platinum peg only, and film boiling takes place from this point downward. Electrical contact anywhere above the platinum peg causes the filament to break. The moving filament proceeds down through three cone shaped baffles or bubble strippers which aid in suppressing side forces caused by the vigorous film boiling. Figures 6 and 7 show film boiling with and without the bubble strippers, respectively. The lower electrical contact is made by a mercury well through which the filament travels. Mercury is held in the well by a neoprene slab with a 10-mil hole in the center which permits passage of the fiber without any mercury leakage. It is necessary to use DC for resistive heating in this apparatus, as AC causes the filament to break at the upper contact. The filament goes directly from the plating cell to a take-up reel which supplies the necessary tension to pull it through the apparatus at a uniform rate.

In this setup the heated portion of the moving filament is about 2 inches long and the rate of travel is about 2 inches per minute. Using 1-mil tungsten wire substrate, the DC potential is 40 volts and current is 0.72 ampere. This gives a filament temperature of about 900°C. A plating solution of 30% trimethylchlorosilane and 70% cyclohexane under these conditions produces a finished diameter of 4.5 mils, and the deposit is concentric as shown in Figure 8.

SUMMARY AND CONCLUSIONS

1. A process has been developed which is capable of depositing silicon carbide on a heated substrate by means of liquid film-boiling on the submerged hot object.

2. The film-boiling process, when adapted to a continuously moving filament, will build a 1-mil diameter substrate up to 4.5 mils with a silicon carbide coating in one minute of plating time.

3. The process produces a layered concentric deposit.

4. Scale-up to production quantities appears feasible.

ACKNOWLEDGEMENTS

The author acknowledges the help received from Clarence W. Austin of the U. S. Army Missile Command for photomicrography, and Iqbal Ahmed of Watervliet Arsenal for mechanical property measurements.

REFERENCES

1. STANDAGE, A. E., and PRESCOTT, R., "High Elastic Modulus Carbon Fibers", Nature, vol. 211(5045), p. 169, (July 9, 1966).

2. RANCH, H. W., SUTTON, W. H., and McCREIGHT, L. R., "Survey of Ceramic Fibers and Fibrous Composite Materials", General Electric Company Tech. Rept. AFML TR-66-365 (Oct. 1966).

3. PRING, J. N. and FIELDING, W., "Preparation of Some Refractory Metals From Their Chlorides", Jour. Chem. Soc., vol. 95, Part II, pp. 1497-1506 (1909).

4. MOERS, K., The Preparation of Pure Refractory Carbides, Nitrides, and Borides by Transport Growths and a Description of a Few of Their Properties: II, Z. anorg. allgem. chem., vol. 198, pp. 243-261 (1931).

5. KENDALL, J. T. and YEO, D., "Preparation of Pure Silicon Carbide", Proc. XI The Intern. Cong. Pure Appl. Chem., London, vol. 1 (7), pp. 171-175 (1947).

6. POWELL, C. F., OXLEY, J. H., BLOCHER, J. M., Jr., "Vapor Deposition", John Wiley & Sons, Inc., New York, p. 367 (1966).

7. MERZ, K. M., "Crystal, Whisker, and Microcrystalline Forms of Silicon Carbide", Edited by O'Connor, J. R. and Smiltens, J., "Silicon Carbide-A High Temperature Semi-Conductor", Symposium Publications Division, Pergamon Press, New York, p. 76 (1959).

8. SUSMAN, S., SPRIGGS, R. S. and WEBER, H. S., "Vapor Phase Growth of β-Silicon Carbide Single Crystals", Edited by O'Connor, J. R. and Smiltens, J., "Silicon Carbide-A High Temperature Semiconductor", Symposium Publications Div., Pergamon Press, New York, pp. 94-109 (1959).

9. HOUGH, R. L. and EARLY, D. E., "Silicon Carbide from Pyrolysis of Nonstoichiometric Organosilanes", Air Force Materials Laboratory Technical Report AFML-TR-66-174 (1966).

10. HUGGINS, H. W. and PITT, C. H., "Characteristics of Silicon Carbide Produced by Thermal Decomposition of Trimethylchlorosilane", Ceram. Bull., vol 46, No. 3, pp. 266-269.

11. McCANDLESS, L. C., WITHERS, J. C., and DAVIES, L. G., "High Modulus-to-Density Reinforcements for Structural Composites", General Technologies Corporation AFML-TR-65-265 (1965).

12. WITHERS, J. C., and REDMOND, J. P., "Synthesis by Vapor Deposition of New Filament Materials", General Technologies Corporation, AFML-TR-66-329 (1966)

13. McCANDLESS, L. C., BAJEFSKY, R. D., BRUMMETT, C. R., DAVIES, L. G., REDMOND, J. P., SHAVER, R. G. and WITHERS, J. C., "High Modulus-to-Density Reinforcements for Structural Composites", General Technologies Corp., AFML-TR-65-265, Part II (1966).

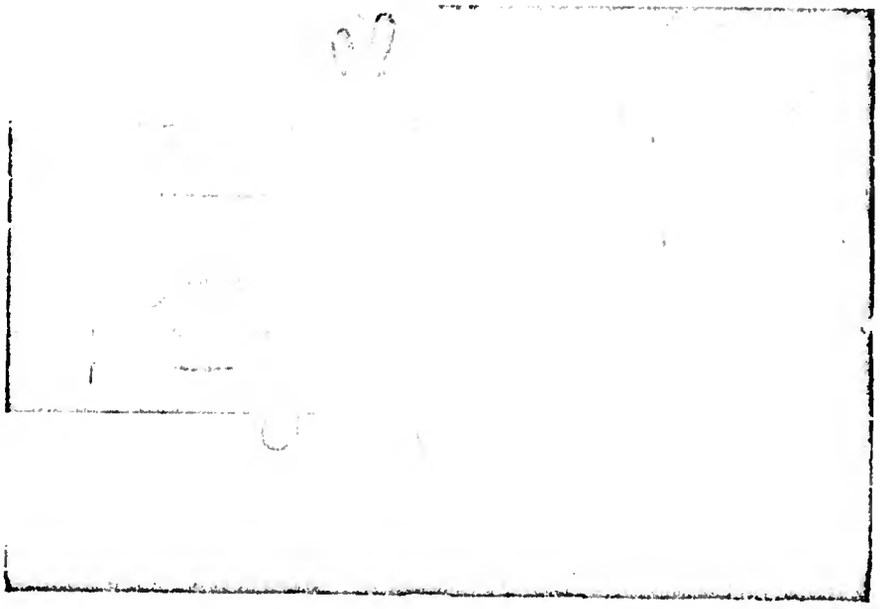


Figure 1. Horizontal Film Boiling Cell

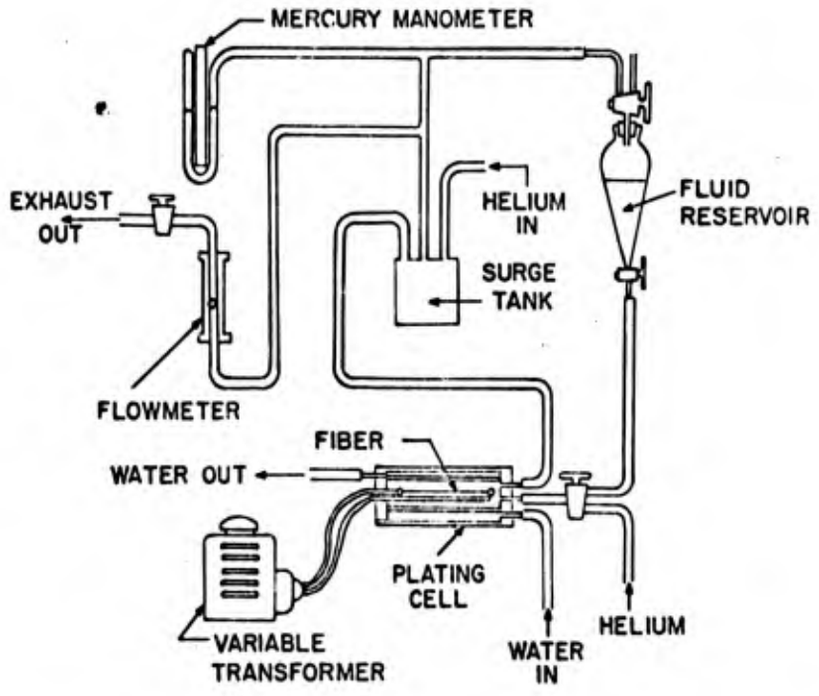


Figure 2. Apparatus for Horizontal Film Boiling

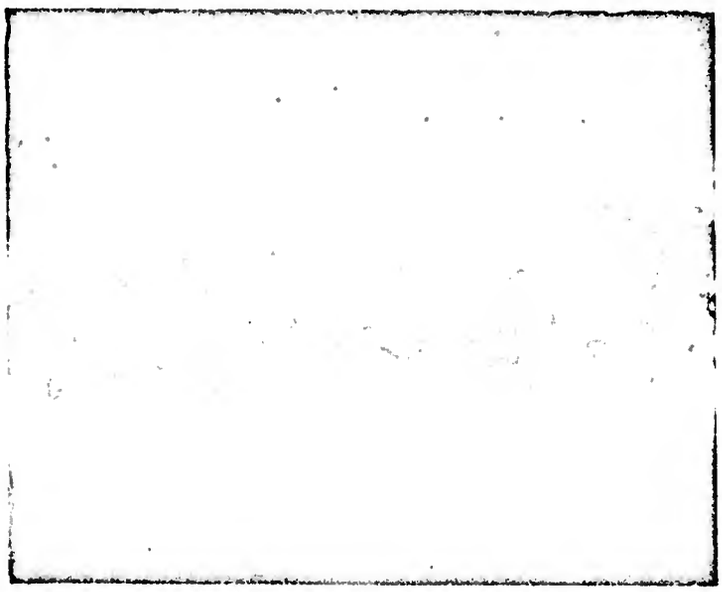


Figure 3. Surface of the SiC Deposit
Magnification 1000X

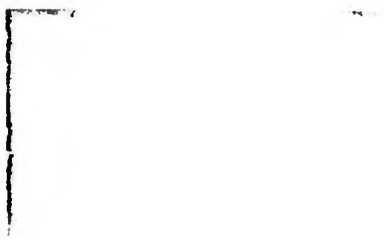


Figure 4. Cross Section of Fiber Plated Horizontally
Magnification 250X

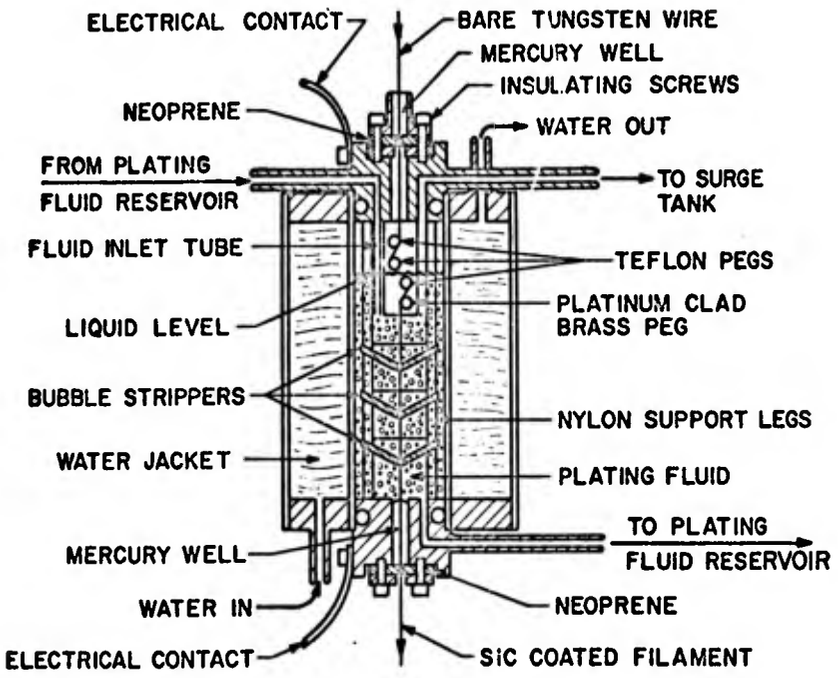


Figure 5. Vertical Continuous Plating Cell

Figure 6. Closeup of Continuous Plating with Bubble Stripper

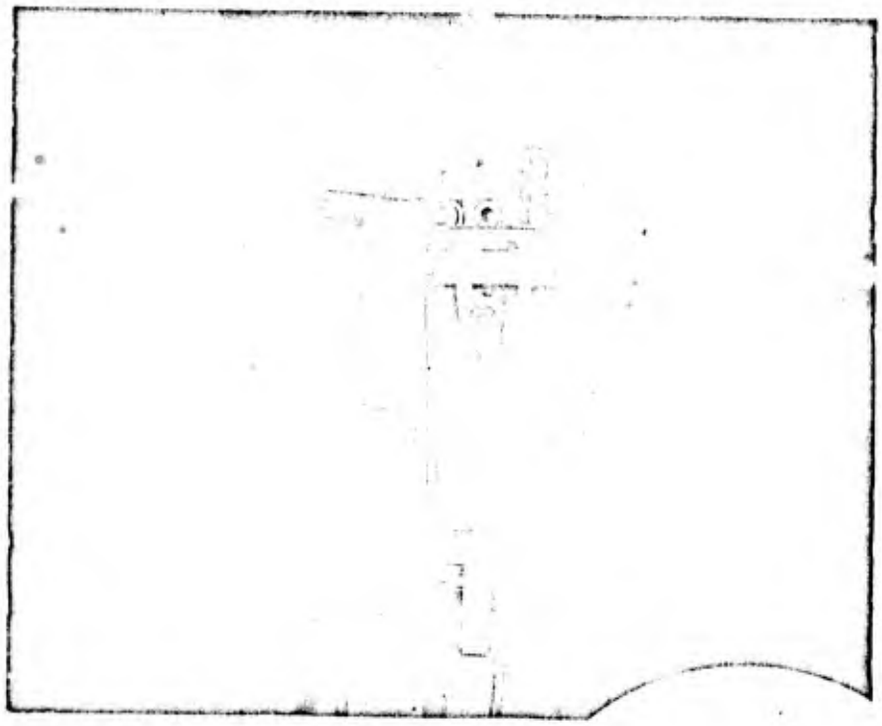


Figure 7. Continuous Plating Without Bubble Stripper

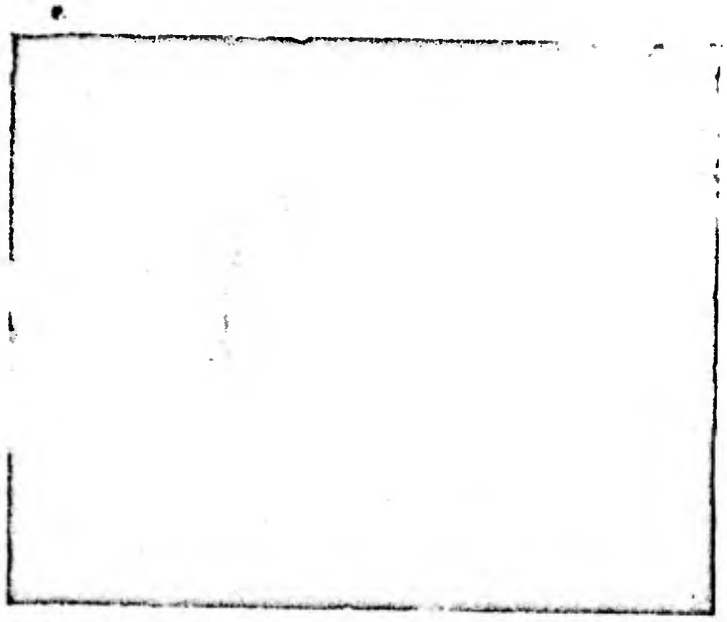


Figure 8. Cross Section of Fiber Plated Vertically
Magnification 500X

THE IN VIVO LOCALIZATION OF STAPHYLOCOCCAL ENTEROTOXIN B

SIGURD J. NORMANN, CAPTAIN, MC and ROBERT F. JAEGER
U. S. ARMY MEDICAL UNIT
FORT DETRICK, MARYLAND

Staphylococcal enterotoxin B is a low molecular weight protein exotoxin capable of producing acute staphylococcal food poisoning in experimental animals.(1) In addition, the purified toxin injected intravenously into rabbits and monkeys produces lethargy, fever, shock and death (2) while the same toxin given to rats and mice appears to have little clinical effect. Since there is no evidence of humoral immunity, the reasons for this difference in species susceptibility as well as the mechanism of action of the toxin itself remain unknown.

Although staphylococcal enterotoxin B (SEB) produces death in some laboratory animals, the clinical course as well as the pathology shows little to explain the cause of death. Apart from emesis, pyrexia, and hypotension of unknown etiology, clinical manifestations include mild hypoglycemia, elevation of blood urea nitrogen, thrombocytopenia, and transient leukopenia.(2-4) At necropsy the principal lesion is pulmonary interstitial edema, (5) but its occurrence is variable and not necessarily present in all animals.

Since neither morbid pathology nor the clinical course suggest a mechanism for the toxemia, we have been concerned with the fate of enterotoxin within the body in an attempt to locate and define the cellular sites of toxin action. It is the purpose of the present communication to describe means to identify the toxin in cells, as well as to measure its removal rate from blood. By combining fluorescent labeling of the toxin with fluorescent antibody methods, we have been able to correlate the vascular clearance of the toxin with its cellular localization in both rats and monkeys. Such studies have shown that the proximal convoluted tubules of the kidney are the principal sites of toxin removal and, furthermore, suggest that the toxin gains access to these tubules by a process of glomerular filtration and tubular reabsorption.

Enterotoxemia was induced in rats and monkeys by intravenous injection of the purified SEB. The biologic distribution of the unlabeled toxin was then assessed by fluorescent antibody methods one hour after challenge (Table I). No significant differences in distribution were observed at this time period between the two animal species for the tissues examined. In both species, a minimal positive reaction was observed in spleen and gastrointestinal tract; whereas, the lung and multiple sections at various levels of the brain were negative. The liver of the monkey had occasional but infrequent fluorescence whereas the liver of the rat failed to show significant localization. In contrast to these organs, the kidneys of both species exhibited a brilliant green luminescence. The location of this fluorescence appeared to be confined to the proximal convoluted tubules as little fluorescence was observed in distal tubules and glomeruli were consistently negative.

The results of two other methods for evaluating SEB distribution in the intact animal are also summarized in Table I. When the toxin was labeled with the fluorescent reagent fluorescein isothiocyanate (SEB-FIT), its distribution could be observed directly in tissue sections examined in ultraviolet light. However, there were no significant differences in distribution between the labeled toxin and that of unlabeled toxin identified by the direct fluorescent antibody technique. On the other hand, radioactive iodinated toxin did show some preference for liver which was not appreciated with use of the other two methods. All methods indicated a high concentration of toxin in the kidney.

Although labeling of the toxin with either a fluorochrome or radioactive iodine did not destroy its toxicity, the labels did have an effect on the rate of clearance of the toxin from the circulation (Table II). Thus, when SEB labeled with radioactive iodine was injected into rats its rate of vascular clearance was significantly faster than the clearance of SEB-FIT. Since the rate of clearance was proportional to the toxin concentration, the clearance rate constant or K was obtained from the slope of the plot of the natural logarithm of the toxin concentration against time. Following bilateral nephrectomy, there was a dramatic retardation in the clearance rate of SEB but the clearance remained exponential. This reduction in clearance rate was used to calculate the amount of toxin removed by the kidney, since the difference in clearance rates before and after nephrectomy represented the rate of toxin accumulation by the kidney. At 60 minutes after injection between 75 and 80% of the toxin regardless of label in both rats and monkeys theoretically should have been cleared by the kidney. This observation underscored the importance of the kidney in the removal of intravenously injected toxin and confirmed the impression of the distribution of SEB as indicated by fluorescence microscopy and reported in the previous table.

The question next arose as to the mechanism of renal clearance. Therefore, the following sets of experiments were performed in order to determine whether the SEB localized in the renal tubules by a process of glomerular filtration and tubular reabsorption or by pinocytosis from the efferent glomerular blood.

The objective of the first set of experiments was to determine the earliest localization of the toxin in the renal tubules and to correlate its subsequent distribution with the time after challenge. It was reasoned that if the kidney samples were taken at sufficiently close intervals after challenge it might be possible to observe (a) whether the toxin was initially present within the tubular lumen and (b) the direction of migration of toxin across the cell. The experiments were done with both unlabeled SEB detected by fluorescent antibody methods and with SEB conjugated with fluoresceine isothiocyanate (FIT). Both methods gave essentially similar results.

The earliest time interval in which the toxin could be visualized in the renal tubules was 15 seconds after intravenous challenge. At this time the toxin was clearly present on the luminal side of the proximal convoluted tubules as a distinct thin layer of fluorescence adjacent to the cell surface. In some sections, it was possible to observe that the toxin was present at the glomerular tubular junction. However, it was confined to the tubular epithelium and was not observed adherent to the parietal layer of Bowman's capsule nor to glomerular capillary endothelium. Indeed, the toxin appeared to show an affinity only for the brush border of tubular epithelium. From 15 seconds to one minute, the intensity of the fluorescence increased, but its distribution relative to the cell did not change. Figure 1 is a representative section taken at one minute after challenge showing the toxin confined to the tubular lumen.

Five minutes after injection, the toxin was still located predominantly on the luminal surface of the tubular cell, but the fluorescence intensity had increased. By 15 minutes, a definite gradient of fluorescent intensity from the luminal side of the cell towards the base of the cell was observed. At 30 minutes the entire tubular cell was fluorescent (Figures 2, 3). In all sections examined the presence of fluorescence was confined to the proximal tubules and did not involve the glomerulus or the distal segments.

It was concluded that the earliest appearance of either unlabeled SEB or SEB labeled with FIT was on the luminal side of the tubular cell presumably adherent to the brush border of the proximal convoluted tubules. The SEB then migrated from the luminal side to the basal portions of the cell so that by 30 minutes the entire cell was fluorescent. The above findings strongly indicated that SEB was filtered by the glomerulus and reabsorbed by the proximal tubules. If this were so, then it ought to be possible

to prevent the localization of SEB in the kidney by a change in tubular pressure. Such a change can be induced by ureteral ligation. For this experiment, a unilateral ligation was performed so that the opposite kidney could serve as a control. SEB-FIT was then administered at 6, 12 and 24 hours after ligation and tissue was obtained 30 minutes later.

In kidneys obtained at 6 hours after ureteral ligation, the ligated and control kidneys showed an equal intensity of fluorescence. However, by 12 hours there was a marked reduction in fluorescence intensity in those kidneys whose ureter had been ligated. This change was more pronounced at 24 hours (Figures 4, 5), at which time most of the renal cortex was devoid of identifiable SEB. No appreciable pathologic changes were evident by light microscopy. Thus, a change in tubular hydrostatic pressure could prevent the localization of SEB in the kidneys.

If SEB were filtered by the glomerulus, then some relationship should exist between the rate of clearance of SEB and the rate of glomerular filtration. The calculation of glomerular filtration rate was made assuming that all SEB reaching the glomerulus was filtered and that SEB was not removed by the kidneys in any other manner. Under these conditions glomerular filtration rate was equal to the product of the clearance rate constant of the kidney (i.e., the difference in rate constants before and after nephrectomy) and the animal's blood volume. Table III presents the results. It is observed that in both rat and monkey there is a remarkable similarity between the accepted value for glomerular filtration rate and that derived from SEB clearance.

From the foregoing, it is proposed that the toxin gains access to the renal tubules by a process of glomerular filtration and tubular reabsorption. The evidence which supports this conclusion is the following: (a) the earliest appearance of the toxin following intravenous challenge is on the luminal surface of the proximal tubules; (b) in subsequent time periods, the toxin appears to move from the luminal side of the cell to involve the entire cell; (c) ureteral ligation, which produces a gradual increase in tubular hydrostatic pressure ultimately stopping glomerular filtration, prevents SEB from appearing in the kidney; and, (d) the rate of SEB clearance by the kidney is compatible with the rate of glomerular filtration.

The low molecular weight of the toxin (35,300) would be consistent with glomerular filtration. However, Crawley and co-workers,(2) who investigated the blood binding properties of SEB by precipitation methods, reported that 90% of intravenously injected SEB was transported in blood as a complex with serum albumin. From the study presented here, however, it now appears to be firmly established that intravenously administered SEB is rapidly removed from the circulation with concomitant accumulation

in proximal convoluted tubules. Since it would be difficult to equate rapid vascular clearance with glomerular filtration - if indeed the toxin were bound to a circulatory transport protein - the next set of experiments was initiated to re-examine the question of whether staphylococcal enterotoxin exhibited any affinity for blood protein.

Immunoelectrophoresis was performed using both rat and monkey plasma and results were similar. A typical pattern for the rat is presented in Figure 6. Electrophoresis of SEB in plasma (100 µg/ml mixed at 37° C. for 30 minutes) was first performed and compared with an equal concentration of SEB in saline. In both instances, only a single precipitin band was found to rabbit anti-SEB and it was apparent that plasma had not altered the mobility of the identified fraction of SEB. Since on serial dilution, the SEB band in plasma disappeared at approximately the same dilution as SEB in saline (approximately 25 µg/ml) it was unlikely that any appreciable SEB complex had been formed in plasma whose immunological identity had not been detected. In order to observe whether the location of the SEB coincided with a plasma protein component, immunoelectrophoresis of plasma containing SEB was performed against rabbit anti-rat plasma and in the other direction against rabbit anti-SEB. As seen in Figure 6, there was no identity between the SEB band and any of the identifiable components of plasma. A similar result was obtained regardless of whether SEB was mixed with rat plasma in vitro or added to plasma by intravenous challenge (blood sample obtained 4 minutes after injection of 500 µg SEB/100 gm. body weight). Although the addition of SEB in monkey plasma in vitro (100 µg/ml) yielded a pattern similar to that of the rat, the concentration of SEB 4 minutes after intravenous injection in the monkey (1000 µg/kg) was not sufficient to resolve clearly the location of SEB by immunoelectrophoresis. However, the presence of SEB could be detected by Ouchterlony gel diffusion by using a larger sample volume.

Analysis by Ouchterlony gel diffusion failed to reveal any evidence of association between SEB and plasma proteins. A typical experiment using monkey plasma is presented in Figure 7. Only a single precipitin line was observed between plasma SEB and rabbit anti-SEB and this line was continuous with the precipitate formed with SEB in saline. The fact that the precipitin line appeared at the same distance between SEB in saline and anti-SEB as between SEB in plasma and anti-SEB, when equal concentrations of SEB were used, indicated that the coefficient of diffusion of SEB alone and SEB in plasma was identical. Furthermore, when the plasma components were identified with goat anti-monkey, it was apparent that the SEB precipitin line extended completely across the lines of all major plasma components.

Finally, since the molecular weight of SEB at 35,300 is less than that of most plasma proteins, it was possible to separate SEB from plasma using molecular sieve chromatography. When SEB alone was applied to a Sephadex G-75 column, it was eluted in a volume significantly greater than that of the protein peak of plasma applied alone. Therefore, it would be expected that when plasma and SEB were present together and if SEB were bound to a blood protein, it should migrate at a faster rate and appear in the plasma protein peak. However, when the SEB was added to plasma either by injection in vivo or by incubation in vitro, there was no significant change in the elution volume of SEB. Similar results were obtained with both rat and monkey plasma.

Whereas failure to identify a complex by immunologic methods might occur if the character of the SEB were altered by complex formation, this possibility was excluded by the chromatographic experiments and by the fact that on serial dilution all SEB in plasma could be accounted for as a single band on immunoelectrophoresis. Despite the inability to clearly locate SEB on immunoelectrophoresis pattern from the monkey following intravenous challenge, it could be identified on Sephadex G-75 and Ouchterlony gel diffusion. These experiments, together with observations made when SEB was added to monkey plasma in vitro support the contention that the behavior of SEB in monkey plasma is similar to that of rat plasma. Therefore, the present studies demonstrate the unlikelihood of association between SEB and the plasma proteins of a nonimmune animal and imply the restriction that if such binding does occur it must be of weak character and easily dissociable. This conclusion implies that the behavior of the toxin in the circulation and its biologic fate following intravenous injection probably reflect the intrinsic properties of the SEB molecule itself.

Although the majority of the toxin is rapidly removed from the circulation by the kidney, it was important to establish not only the duration of renal toxin localization but also whether sites other than the kidney might become involved late in the course of toxemia. Accordingly, the distribution of fluorescein-labeled toxin was examined in the hours following challenge in both monkeys and rats. Table IV presents the findings.

In the normal rat the toxin remained in the kidneys without apparent diminution of fluorescence intensity for 4 to 8 hours postchallenge. Thereupon a decrease was observed so that by 12 hours visible fluorescence was half that observed at one hour. Nevertheless, fluorescence was still present at 24 hours. The only site other than the kidney that concentrated sufficient toxin for visualization was the liver at 2 to 4 hours postchallenge. The importance of the liver can be demonstrated further by preventing the toxin from gaining access to the kidneys. Thus, in the nephrectomized rat, the liver fluorescence was intensified

revealing it to be a site of toxin concentration and an alternative site to the kidney. The majority of the toxin was confined to Kupffer cells. This observation was compatible with data obtained using radio-iodinated labeled toxin in which it was shown that the highest concentration of toxin in the nephrectomized rat occurred in the liver although significant radioactivity also appeared in the lung, skin, and gastrointestinal tract. When the monkey was examined instead of the rat a major difference appeared in the duration of the toxin localization in the kidney as may be seen in Table IV. While considerable toxin was present in the rat kidney at 8 hours postinjection, it was effectively gone from the monkey kidney in the same time interval.

In summary, the present experiments based on morphologic examination and clearance kinetics clearly indicate that the kidney is the predominant site of toxin localization after intravenous injection of SEB. Furthermore, our experiments show that the majority of the toxin found in the kidneys of both rat and monkey is located in the proximal convoluted tubules. Evidence has been advanced to substantiate the hypothesis that toxin gains access to the renal tubules by glomerular filtration and tubular reabsorption. Having gained access to proximal tubules, the toxin remains there for several hours. The possibility that SEB recycles through the kidney by being transported across renal cells and returned to the circulation is not excluded by the present experiments. However, significant redistribution of labeled toxin from the kidney to other sites did not appear to occur despite the increased fluorescence noted with time in liver Kupffer cells as well as a slight increase in radioactively labeled toxin in the lung. Of considerable interest is the fact that the alternative site to toxin localization in the kidney is the liver and that at no time could evidence be obtained for cerebral localization of the toxin. The absence of toxin in the brain is consistent with previous findings that there is no electro-encephalographic alteration in toxemic animals (6) and that the initiation of vomiting depends on local stimulation of a reflex arc originating in the gastrointestinal tract.(7) It is concluded that the kidney may play a major role in the pathogenesis of enterotoxemia.

In conducting the research described in this report, the investigators adhered to the "Guide for Laboratory Animal Facilities and Care," as promulgated by the Committee on the Guide for Laboratory Animal Facilities and Care of the Institute of Laboratory Animal Resources, National Academy of Sciences -- National Research Council.

REFERENCES

1. SCHANTZ, E. J.; ROESSLER, W. G.; WAGMAN, J., SPERO, L., DUNNERY, D. A. and BERGDOLL, M. S. Purification of staphylococcal enterotoxin B. Biochemistry 4:1011-1016, 1965.

2. CRAWLEY, G. J.; GRAY, I.; LeBLANG, W. A. and BLANCHARD, J. W. Blood binding, distribution and excretion of staphylococcal enterotoxin in monkeys. J. Infect. Dis., 116:48-56, 1966.

3. CRAWLEY, G. J.; BLACK, J. N.; GRAY, I. and BLANCHARD, J. W. Clinical chemistry of staphylococcal enterotoxin poisoning in monkeys. Appl. Microbiol., 14:445-450, 1966.

4. LAMARRA, C. and CARR, C. J. The botulinal, tetanal, and entero-staphylococcal toxins: a review. Clinical Pharmacology & Therapeutics 8:286, 1967.

5. FINEGOLD, M. J. Interstitial pulmonary edema; An electron microscopic study of the pathology of staphylococcal enterotoxemia in rhesus monkeys. Lab. Invest. 16:912, 1967.

6. HODOVAL, L. F.; MORRIS, E. L.; ELSBERRY, D. C. and BEISEL, W. R. Effect of staphylococcal enterotoxin B on the electroencephalogram of monkey. Applied Microbiology 15:403, 1967.

7. SUGIYAMA, H., and HAYAMA, T. Abdominal viscera as site of emetic action for staphylococcal enterotoxin in the monkey. J. Infectious Disease 115:330, 1965.

8. BENDITT, E. P.; STRAUBE, R. L. and HUMPHREYS, E. M. The determination of total circulating serum proteins and erythrocyte volumes in normal and protein depleted rats. Proc. Soc. Exp. Biol. Med., 62:189-192, 1946.

9. SMITH, H. W. The Kidney. Oxford University Press, New York, 1955, p. 534.

10. DITTMER, D. S. (ed.). Blood and Other Body Fluids. Fed. Amer. Soc. Exp. Biol., Washington, p. 4, 1961.

TABLE I

Biologic Distribution of SEB One Hour After Challenge.

	Animal	Tissue Method	Kidney	Liver	Lung	Spleen	GI	Brain
SEB	Rat	Fl. Ab.	++++	±	-	±	±	-
SEB-FIT	Rat	Fl. Micros.	++++	±	-	+	±	-
SEB-^{125}I	Rat	Counts % Dose	56.7	5.4	0.68	1.22	4.49	0.03
SEB	Monkey	Fl. Ab.	++++	±	-	±	±	?
SEB-FIT	Monkey	Fl. Micros.	++++	±	-	±	±	-
SEB-^{131}I (Ref. 13)	Monkey	Counts % Dose	42	25	5	0.15	6	0.2

Abbreviations: Staphylococcal enterotoxin B (SEB); fluoresceine isothiocyanate labeled toxin (SEB-FIT); radioactive iodine tagged toxin (SEB-^{125}I or SEB-^{131}I); fluorescence (Fl); antibody (Ab); microscopy (micros.). Distribution by radioactivity is expressed as per cent of injected dose counted in the entire organ.

TABLE II. Clearance Data and Toxin Distribution in Kidney (see text). The clearance rate constant is represented by K.

	Animal	Intact Animal K	Post Nephrectomy K	Theoretical % Injected Dose Kidney
SEB-FIT	Rat	0.082	0.015	81.8
SEB-^{125}I	Rat	0.121	0.031	74.3
SEB-^{125}I	Monkey	0.096	0.022	76.8

TABLE III. Calculation of Glomerular Filtration Rate (GFR)

Toxin	Animal	Kidney Clearance Constant	Blood Volume	GFR	
				SEB	Accepted
SEB-FIT	Rat	.067	6.2 ml/100g	0.42	0.60 (8,9)
SEB-^{125}I	Rat	.090	6.2 ml/100g	0.56	0.60 (8,9)
SEB-^{125}I	Monkey	.074	44.7 ml/kg	3.31	3.66 (10)

TABLE IV. Extent of Fluorescence by Time After Injection of SEB-FIT

		5 min	30 min	1 hr	2 hr	4 hr	8 hr	12 hr	24 hr
Normal Rat	Kidney	++	++++	++++	++++	+++	++	++	±
	Liver	-	-	±	+	+	±	±	-
Nephrecto-mized Rat	Liver	±	+	+	++	+++	++	++	±
Normal Monkey	Kidney	++	++++	++++	+++	++	±	±	-
	Liver	-	-	-	-	-	-	-	-

Sections of lung, brain, and heart were devoid of specific fluorescence at the time periods examined in both rat and monkey. Questionable fluorescence was observed on occasion in spleen and gastrointestinal tract. No definite localization of toxin was found for the nephrectomized monkey.

Fig. 1. Localization of SEB in rat kidney 1 minute after injection. Fluorescent antibody method. Fluorescence remains on the luminal border of the cell. X 820

Fig. 2. Localization of SEB in rat kidney 30 minutes after injection. Fluorescent antibody method. Fluorescence is now seen throughout the cell. X 1440

Fig. 3. Localization of SEB in monkey kidney 60 minutes after injection. Fluorescent antibody method. This picture shows tubular fluorescence adjacent to a glomerulus which is devoid of fluorescence. X 1440

Fig. 4. Localization of SEB in rat kidney 30 minutes after injection. SEB-FIT with fluorescence microscopy. This kidney is from the control side of a rat whose opposite kidney had its ureter ligated. Compare this picture with Fig. 5. X 520

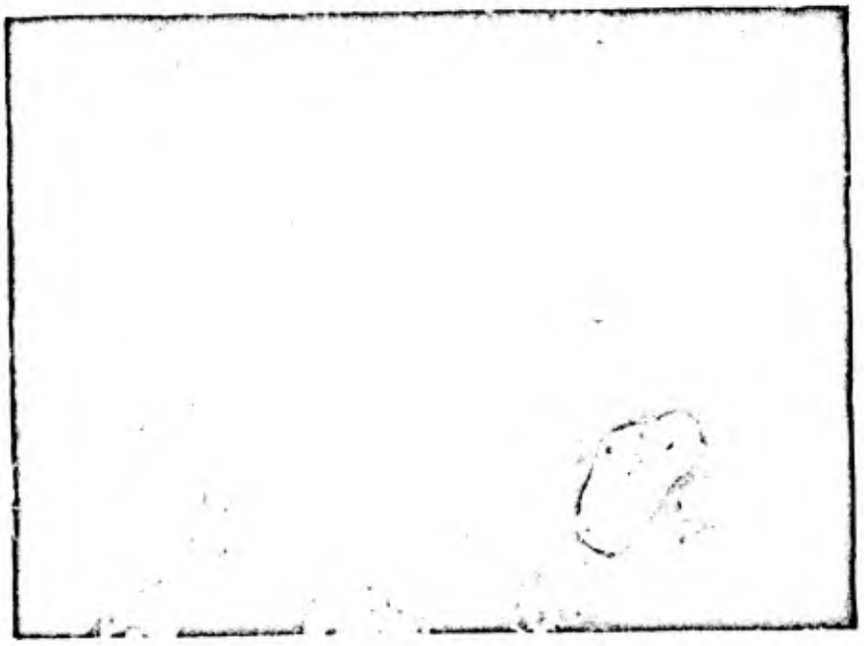


Fig. 5. Localization of SEB in rat kidney 30 minutes after injection with ureter ligation 24 hours prior to injection. SEB-FIT with fluorescence microscopy. Note the marked decrease in fluorescence and in the number of tubules involved. X 520

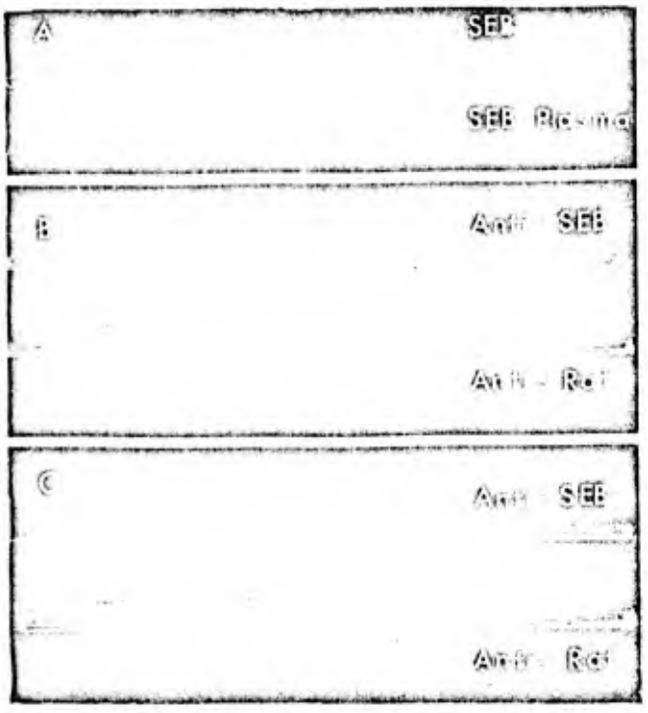


Fig. 6. Immunoelectrophoresis of staphylococcal enterotoxin B (SEB) and plasma proteins. The anode is on the left. Electrophoresis for 45 min at 12 ma per strip, using 0.02 M phosphate buffer, pH 7.4, with sodium chloride added to bring ionic strength to 0.100. Experiment A shows location of SEB in rat plasma identical to that of SEB in saline (immunodiffusion against rabbit anti-SEB). In experiment B, the electrophoretic sample was SEB-incubated with plasma in vitro; in Experiment C, the sample was obtained from a rat 4 min after SEB injection. In both experiments, the location of the SEB band is identical to that shown in Experiment A, and the SEB band is not associated with any identified plasma protein.

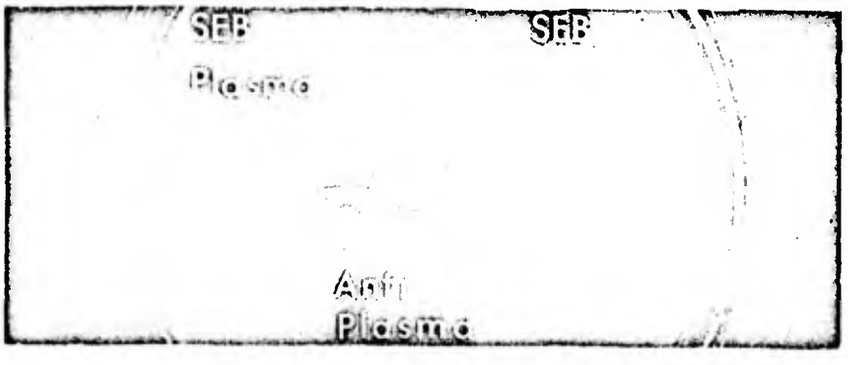


Fig. 7. Ouchterlony gel-diffusion analysis of staphylococcal enterotoxin B (SEB). The center well is rabbit anti-SEB. The other wells contain SEB in saline, SEB in monkey plasma, and goat antimonkey sera.


Blank

PROTECTION AND REVERSAL OF LETHAL MUSTARD DAMAGE RESULTING IN RECOVERY OF CELL VIABILITY

BRUNO PAPIRMEISTER, CLAIRE L. DAVISON AND CLARK L. GROSS
MEDICAL RESEARCH LABORATORY
EDGEWOOD ARSENAL, MARYLAND 21010

Sulfur mustard (H) is one of the oldest of chemical warfare agents. It acts on skin to produce vesication and slow healing, incapacitating, easily infected sores, and systemically, to inhibit cell division, bring about a pronounced leucopenia and to cause hemorrhages and edema in mucosal linings of the intestine and lungs. Also, at somewhat higher doses, H initiates significant pharmacological effects in the central nervous system (1-4). Although many man-years of research on mustard have been done, no effective prophylaxis or therapy is currently available. Because of this glaring deficiency in defensive capability a program was instituted approximately 10 years ago to investigate and elucidate the mechanism of action of H with a view toward establishing a rationale for prophylaxis and therapy. As a result of this program we are optimistic that the destructive effects of H may soon be overcome or modified.

The present article has a two-fold purpose: (a) to define, describe and analyze the mechanism of lethal H action, and its prevention and reversal, at the molecular, subcellular, cellular and tissue levels and (b) to illustrate the value of basic and fundamental research as a practical approach for solving practical problems.

DNA IS THE SENSITIVE TARGET

H is known as a cytotoxic alkylating agent which mimicks the effects of radiation. It has its most pronounced lethal action on proliferating cells and viruses, causing them to become "sterilized" (i.e. to lose their capacity to undergo division and produce an unlimited number of progeny). Nondividing cells are considerably less sensitive to H. They require much larger, often non-physiological, doses to inhibit or interfere with their normal functions. Investigators generally agree that this high susceptibility of reproducing cells is due to alterations produced within DNA, although H also reacts extensively with most other cellular constituents (5-9). The evidence which supports the notion that DNA, the genetic

substance of the cell, represents the principal and most critical target is: (a) H at cytotoxic doses caused a rapid and long lasting inhibition of cellular and viral DNA synthesis, but affected other biochemical functions (i.e. RNA synthesis, protein synthesis, respiration etc.) to a much lesser extent or not at all (10). (b) H had its biologically most intense effects when acting on DNA-containing viruses or on purified, genetically competent transforming DNA (5-7, 11). (c) H produced chemical changes in DNA, both in vivo and in vitro, which could readily account for the observed biological effects. These DNA alterations, which will be dealt with in greater detail in a later section, included the formation of covalent, interstrand crosslinks. These crosslinks interfered with the separation of the two complementary strands and prevented ordinary semiconservative DNA replication (6, 12, 13). (d) H initiated the degradation of cellular DNA, caused the induction of prophage in lysogenic bacteria and, generally, produced effects which were qualitatively similar to those brought about by other DNA - damaging agents (i.e., ultraviolet- and x-irradiations, mitomycin, decay of radioactivity incorporated in DNA, etc) (14).

In spite of this imposing evidence, the idea that DNA is the principal target to lethal H action suffered from some serious shortcomings. How is it possible that certain cell types are more sensitive than others? Why and how does DNA synthesis recover in some strains, but fail to do so in others? How can single-step mutations confer resistance or, conversely, hypersensitivity? Why are cells at certain stages of their division cycle more prone to lose their viability than are the same cells at a different stage? Why and how do some additives to the culture medium enhance or reduce the lethal outcome? Since similar phenomena occurred when employing x-rays, ultraviolet rays, etc., where the agent's permeability was not a factor, one must conclude that either DNA was not the primary target to lethal H action or, that the actual sensitivity of cells to H depended on both the nature of the primary DNA lesion and metabolic events which could either enhance or repair the primary damage. Whatever the answer, we felt that an investigation to determine how resistant cells were able to cope with H-induced lesions should help provide a rationale for eventual therapy.

The results of such studies were unequivocal, and provided compelling evidence that DNA was indeed the most critical target in lethal H-action (15, 16). Furthermore, the experiments pointed to the H-induced crosslinks in DNA as the alterations which were primarily responsible for those observed lethal effects. Utilizing a variety of bacterial mutants, which were carefully chosen to provide a wide spectrum of sensitivities to H, we could show that survival after H exposure was possible only in those strains which eliminated the guanine-containing crosslinks from their DNA (GHG). The speed and specificity of this elimination was directly related to eventual survival. Those bacteria (E. coli, strain B/r) capable of eliminating GHG most rapidly (Fig. 1) and most specifically (Fig. 2) had the highest resistance to H (Fig. 3). The more sensitive strains showed a delayed and less specific elimination of GHG

(E. coli B) or no elimination of GHG at all (E. coli Bs-$_1$). The sensitivities of the excision capable strains E. coli B/r and E. coli B, were subject to wide variation, depending on both pre- and post-treatment growth conditions (14). Similar changes failed to affect the high susceptibility of the excision incapable mutant, E. coli Bs-$_1$, (Fig. 4). These results suggested that resistance to H was occasioned by the functioning of a repair mechanism which enabled cells to excise from their DNA those moieties which would otherwise block replication. This conclusion was also supported by results obtained with H-treated bacteriophage T_1. After having been exposed to H prior to infection (to preclude any effects of the agent on the cell) these bacterial viruses had considerably higher survival in the repair capable mutants, E. coli strain B and B/r than in the repair inapt mutant, E. coli Bs-$_1$ (Fig. 5). The DNA repair inhibitors caffeine and proflavine readily negated this difference. This indicated that different susceptibilities were attributable to the repair system.

Time and space does not permit the citing of many additional studies which further attested to the fact that DNA damage was primarily responsible for the lethal action of H. The main concerns in building a rationale for therapy are to assess the importance of DNA repair in human cells, to determine what types of DNA damage are repairable, to assess what other types of damage and what other systems for repair are available and to discover how DNA repair may be stimulated and inhibited. This report will discuss some of these questions.

TYPES OF DNA DAMAGE

What is the nature of the lesions sustained in DNA? What are the biological consequences? To answer these questions we considered two general classes of DNA damage: (I) Distortion of DNA structure. This damage affected large regions of the genetic macromolecule and resulted in immediate and wide-spread biological disturbances. And (II) distortion of DNA information. This damage which was initiated by chemical modification of a base, altered biological functions in more subtle ways. The final result was determined by the essentiality of the protein specified and by the type of coding error incurred. Only brief consideration will be given here to type (II) damages because they do not appear to be responsible for the high cytotoxicity associated with bifunctional alkylating agents. Similar, or even identical, damage was produced by single-armed mustard analogs.

Type (II) damage - which may be compared to the alteration of words in an essay - could be inconsequential or give rise to nonsense or even fraudulent information. In a biological context, type (II) damages, which could be either lethal or sublethal, were likely responsible for the mutagenic, carcinogenic and teratogenic properties of alkylating agents (8, 17). They were probably produced by reactions of the agent with guanine. This reaction, by virtue of subsequent elimination or affecting the coding affinity for cytosine

during replication, resulted in heritable mutations. Furthermore, the base modification could have a transcription effect which caused the formation of a faulty or non-functional enzyme. These interpretations are consistent with our previous reports on the effects of various mustards on the survival, mutability and enzyme-forming capacity of T_2 bacteriophage (6).

The lethal action of H was primarily attributable to type (I) DNA damage and, more specifically, to those alterations in the topology of DNA initiated by bifunctional alkylation. The covalent crosslinking of the twin strands of DNA interfered with strand separation and the process of ordinary semiconservative replication. The following experimental evidence indicated that this crosslink was the most important factor. (a) The immediate, highly effective inhibition of DNA synthesis and viability by relatively few molecules of bifunctional agents bound to viral or cellular DNA. Many more molecules of the monofunctional alkylating agent were required to produce similar effects (15). (b) The isolation of covalently crosslinked DNA from H-exposed cells (15). The interstrand nature of this crosslink was deduced from chemical indentification of the crosslinked moiety (2 guanine molecules alkylated through their 7N positions to one H molecule -GHG) and from the ease and spontaneity with which the alkylated DNA renatured following denaturation by heat or alkali. (a return to the native conformation was expected if the bases on the complementary twin strands were kept in register by a few or even a single covalent bond) (12). (c) The recovery of DNA synthesis and viability of H treated cells and viruses followed the elimination of interstrand crosslinks from their DNA (10, 15).

Recent evidence obtained in our laboratory indicated that 3 other type (I) DNA damages also seemed to contribute significantly to the lethal action of H: (a) breakage of the phosphodiester backbone resulted in the sensitization of DNA to the degradative action of an exonuclease (Fig. 6). It is of interest to note that single strand breaks were also produced by the single armed sulfur mustard (semi H). Also, the resistance of 3 bacterial mutants to semi H appeared to correlate well with levels of "DNA ligase", an enzyme which repairs single strand DNA interruptions (18). (b) a large increase in the hydration of DNA around the positively charged quaternary nitrogen atoms produced by the alkylation of guanine. This "swelling" of the DNA structure decreased the average density of DNA (Fig. 7), interfered with the rapid injection of DNA through the thin tail section of T_4 bacteriophage and was readily prevented by Mg^{++} or other polyvalent (but not monovalent) cations (Fig. 8). The degree of hydration-induced swelling of DNA was significantly enhanced by polyfunctional alkylating agents which were capable of creating several positively charged centers in very close proximity along the DNA structure. The biological importance of the hydration effect in H-treated cells remains to be explored. (c) intrastrand crosslinking which interfered with the folding and proper packaging of H-treated DNA within the protein envelope during phage maturation. These effects were previously described (6).

The type (I) DNA damage that predominates may well depend on the nature and sequence of the biological system(s) under investigation (e.g. DNA replication, cell division, folding of DNA in chromosome, movement of DNA across a membrane, maturation of a virus, etc). While it is of some scientific importance to further elucidate the lethal mechanism, our interest is the establishment of a rationale for therapy against H. Therefore, we next asked how resistant cells are able to cope with lethal DNA lesions and recover.

REPAIR OF LETHAL DNA DAMAGE

Does lethal DNA damage invariably culminate in cell death? The answer is an emphatic "no". Many cells are equipped to handle these insults to their vital molecules. The survival of H-treated bacteria depended on the cell's inherent ability to remove defects in its DNA and on environmental factors which allowed the moribund cell to perform this job quickly and efficiently (Figs. 1-4).

The repair of damaged DNA as visualized at present (Fig. 9), is carried out by a battery of enzymes which act sequentially to (a) recognize the distortion of the two-stranded molecule, (b) make an incision in the vicinity of the defect by an endonucleolytic attack on the phosphodiester backbone, (c) eliminate the damage by sequential exonucleolytic excision of nucleotides which include the damaged moiety, (d) replace the excised region by utilizing the intact complementary strand to faithful reproduction, and (e) rejoin the newly synthesized sections to the older parts of the chromosome.

Distortions of the DNA structure by H could be recognized by default, that is the inability of the faulty parts to form appropriate complexes with histones and other basic cellular components. This association of DNA with the positively charged compounds is a dynamic feature of the cell cycle and appears to be essential in protecting the chromosome from degradation by cellular nucleases. The operation of this two-component system with normal and H-treated DNA was demonstrated (Fig. 10). At constant concentration of radioactively labeled DNA, increasing the quantity of a partially purified bacterial extract caused progressively smaller amounts of DNA to be degraded. A marked preferential degradation of the H-treated DNA, compared to normal DNA, was noted with larger levels of extract. The addition of excess unlabeled calf thymus DNA resulted in complete hydrolysis of both the normal and H-treated DNA. These findings were best explained by postulating that the extract contained: (a) a limiting quantity of a DNA-complexing component, which was able to combine fully with normal DNA, but only partially with H-damaged DNA, and (b) an excess of a non specific nuclease which degraded all DNA's to which it had access. The DNA complexing component reaction could have had precedence over the nuclease action and thus rendered DNA non-susceptible to the latter's action. Experiments showed that the competitive effect of adding an excess of unlabeled DNA caused both normal and H-damaged DNA to be degraded extensively and without preference. The complexing component present in E. coli bacteria was purified and is currently under investigation.

The recognition step of DNA repair was also studied *in vitro* using a system designed to emulate the *in vivo* situation. The components of the system were DNA which had been previously treated with S^{35}-labeled H, DEAE dextran (a large molecular weight carbohydrate obtained from Pharmacia Co. which contained many positively charged diethylaminoethyl moieties) and pancreatic deoxyribonuclease. Using various combinations of the above components, we demonstrated a specific elimination from DNA of the H-induced distortions, analagous to those previously described for the bacterial extract (Figs. 11,12). The ease with which partial repair of DNA was effected suggested that a chemotherapeutic approach against lethal H damage is feasible.

Considerable evidence was obtained to validate the existence and explain the mechanism of the 3 additional steps of DNA repair-excision, replacement and rejoining. Among these was the isolation of several hypersensitive mutants which were unable to carry out one or another of the postulated repair steps. Furthermore, enzymes were discovered which are exquisitely suited to perform the presumed functions (e.g. exonuclease III, DNA polymerase, DNA ligase) (18-21). However, it is beyond the scope of the present paper to consider all the details. Let it suffice to say that the overall function of these activities was to restore the original molecule, so that it could resume its normal replication and function. This would permit the cell to recover from the lethal insults to its DNA.

The discussion of DNA repair would be incomplete without briefly alluding to at least two additional aspects: (a) The great versatility of the error-correcting mechanism, and (b) the evolutionary implications of tinkering with the genetic material. The former aspect allows the living cell to deal with a large number of chemically different lethal alterations of its DNA (e.g. those produced by ultraviolet and x-radiations, mitomycin, etc) (22-24). The repair process could be largely responsible for the cross-resistance or cross-sensitivity of cells to a number of different DNA-damaging agents. The latter aspect (b) relates to the cell's ability to cope with other stresses induced by the environment. The elimination of sublethal defects from DNA enhanced the genetic stability and permitted organisms to survive in environments which produced high mutation rates (25). Furthermore, repair enzymes were involved in performing genetic recombination, a process which likely played a significant role in evolution (26).

DNA repair appears to be an active cellular process which ordinarily operates at only a fraction of its capacity. What happens to a cell when the repair system is taxed beyond this optimum (i.e. following exposure to large doses of H)? This was the next subject of concern when we asked: "How do cells die from DNA damage?"

HOW CELLS DIE FROM DNA DAMAGE

The study of cell death is complicated for at least 2 reasons: (a) lack of unanimity as to what constitutes death and (b) the great difficulty in unraveling the very large number and complex cause-and-effect relationships which occur in moribund cells.

For our purposes, cell death due to DNA damage was defined as "reproductive cell death", i.e. the inability of the cell to divide and produce an unlimited number of progeny. This was reasonable, since non-dividing cells did not readily succumb to H, and, more or less continue to perform their functions normally.

H sterilization of a cell was primarily due to lethal alterations within DNA. However, cell death could be averted since many cells had the facility to repair defects in their DNA and recover. On the other hand, if overtaxed, the repair metabolism could itself greatly upset the cell's homeostatic mechanisms and cause its demise. If such secondary actions occurred, they could activate autolytic enzymes, release toxins and inhibitors and thus greatly contribute to the tissue-damaging and pharmacological effects of H. Our studies with H-treated skin in fact, indicated that the vesicant action of H primarily resulted from secondary tissue damage occasioned by the cell's unsuccessful attempts to repair its DNA. These results are discussed later.

TABLE 1. DIFFERENCES IN THE PHYSIOLOGICAL RESPONSES AFTER H-TREATMENT BETWEEN THE REPAIR-CAPABLE E. COLI B/r AND THE REPAIR-INCAPABLE E. COLI Bs-1. Bacteria were exposed to 3×10^{-4}M H and then incubated for various times.

Response Measured	Method	Results for	
		E. coli B/r	E. coli Bs-1
Survival	Colony-forming	5×10^{-2}	$< 10^{-7}$
DNA Synthesis	Incorporation of labeled precursors	Temporarily inhibited, partially recovered	Permanently inhibited
RNA Synthesis	Incorporation of labeled precursors	Partially inhibited	Partially inhibited
Extractability of DNA	Detergent phenol	Largely lost at interphase	Efficiently recovered
Breakdown of DNA	Acid-solubility of prelabeled DNA	Extensive	Small
Polyribosomes	Sucrose Gradients	Rapidly lost	Significantly increased
Breakdown of RNA	Acid-solubility of prelabeled RNA	Extensive after 1 hour lag	Very small
Respiration	$^{14}CO_2$ production from glucose-U^{C14}	Ceased after 1 hour	Continued more than 2 hours
Cell lysis	turbidity change, cell contents in medium	Significantly large	Not significant

How DNA repair metabolism affected the physiology of the H-treated cell was investigated in several bacterial mutants. Some of the results obtained with E. coli B/r, a mutant which possessed an efficient repair system, and with E. coli Bs-1, a hypersensitive mutant which lacks the repair capability, are presented. In the Table 1 experiments a fairly high dose of H was used, sufficient to kill over 95% of E. coli B/r cells, to insure that biochemical changes could be attributed to non-viable organisms. The most surprising feature noted was the large extent of ancillary damage (much of it unrelated to DNA) sustained by the repair - apt B/r cells. When cultured after H-treatment, B/r bacteria rapidly activated the excision mechanism (Fig. 1). However, at this dose level, the H-induced DNA damage was too extensive to be fully repaired and consequentially a large proportion of the H-sensitized DNA was degraded. This initiated a sequence of well-defined catabolic biochemical responses which eventually culminated in cell death: (a) entanglement of the partially single-stranded DNA with other cell constituents (loss of extractability), (b) manufacture of deficient mRNA (loss of polyribosomes), (c) activation of cell's ribonuclease (degradation of RNA to acid soluble components), (d) release of positively charged ribosomal proteins and enzymes which alter cell permeability, cause leakage of cellular constituents and interfere with respiration, and finally (e) gross autolytic manifestations, such as cell lysis, which were incompatible with life.

These changes were initiated by H-damage to the cell's DNA and by DNA repair rather than by actions of the agent on other cell components (e.g. ribosomes, membranes, enzymes. etc). This conclusion was substantiated by the fact that similar catabolism did not take place in E. coli Bs-1. These cells, although completely sterilized by H, not only failed to degrade their DNA but also essentially remained physiologically intact. Thus the physiological response of E. coli Bs-1 to H was more akin to that noted for normally non-dividing cells than to the more rapidly proliferating types.

The idea that DNA-repair can have deleterious consequences reminds one of a statement attributed to Molière who said: "Most people die of their remedies, not their maladies". We must take account of this admonition to succeed in developing a rationale for therapy. It must be determined if more is to be gained by enhancing the efficiency of repair or by blocking it at some early step. Also, while the former approach may be more appropriate for reversing the lethal effects of H in one system (e.g. preventing leucopenia) the latter could be more important or practical in another system (e.g. preventing tissue damage). Only a detailed, critical assessment of the mechanism of H-induced vesication will enable us to make the logical choice in the case of exposed skin. Therefore, we next turned our attention to skin and asked: "what is the relevance of DNA damage to the development of skin lesions?"

PAPIRMEISTER, DAVISON and GROSS

RELEVANCE OF DNA DAMAGE TO THE DEVELOPMENT OF SKIN LESIONS

Are skin lesions* produced by H a consequence of DNA damage or is it perhaps due to some other type of damage? How could this be determined experimentally? The answers to these questions would be important in establishing a rationale for therapy against the vesicant effects of H.

To resolve a potential relationship between DNA damage and vesication, we used several ideas developed in microbial systems; namely, (a) repair of DNA was essential for viability to recover, (b) certain compounds inhibited repair and thereby sensitize cells to H (28), and, (c) an overtaxed DNA repair system itself brought about extensive cellular damages. If these concepts also apply to skin and if H induced DNA damage and its repair play an analagous role in producing skin lesions, then DNA repair inhibitors should significantly affect the development of the lesions.

Caffeine and proflavine, two DNA repair inhibitors, significantly influenced the vesicant action of H. When applied to rabbit skin immediately after H, both inhibitors increased the initial severity of the lesion produced at each dose level employed and decreased the minimal dose required to produce a visible lesion (Figs. 13a and 14a). A notable feature of these experiments was the great stimulation of healing and hair regrowth at later stages. This surprising stimulation of hair growth always began at the periphery of the H-lesion and was especially noticeable in the areas that had been exposed to the repair inhibitors (Figs. 13b and 14b).

The findings suggested the following conclusions about H-exposed skin: (a) The lesion was due to an initial damage to DNA. (b) Rabbit skin possessed a very active, but inhibitable, DNA repair capability. (c) Healing and hair growth were stimulated by cellular breakdown products.

Based on these results and interpretations, we have postulated the following model (Fig. 15): When exposed to H, skin cells sustained damages to their DNA. These damages, especially the interstrand crosslinks, prevented normal DNA replication, (point 1) which is a prominent activity of proliferating cells. Because of this inhibition, the equilibrium which normally regulated the relative concentrations of free DNA, histone and deoxyribonucleoprotein (DNP) was disturbed. In addition, damaged DNA was degraded by the nucleases of the repair cycle, which further aggrevated the regulatory mechanisms. This resulted in an accumulation of an abnormal concentration of free histones within the nucleus. The histones altered the nuclear membrane, in effect destroying it, which released the nuclear contents into the cytoplasm, where they acted on other cellular, membranes. When the lysosomal membranes were destroyed, the autolytic enzymes were released, which caused cell lysis. The enzymes were

*Following exposure to even minute amounts of this alkylating agent, skin undergoes, after a relatively uneventual latent period, these progressive changes: increased vascular permeability, erythema, inflammatory reactions, edema, and finally, blister formation. This often culminates in an incapacitating, slow-healing, easily infected sore (27).

then free to act on surrounding cells and tissues, which caused the release of toxins, inhibitors, mediators, etc. - effects which bring on the inflammatory responses (increased vascular permeability, erythema, edema, etc) which eventually lead to the visible lesion.

This model readily explained the time lapse usually required for production of skin lesions in cells sustaining DNA damage. The rate at which the lesion developed depended on the concentration of free histones. This, in turn, was governed by the comparative rate of DNA degradation to DNA resynthesis. On the other hand, heat and proteolytic enzymes produced vesication in minutes (29). These agents acted much faster than H because they bypassed the DNA repair cycle. This caused direct membrane damage, release of lysosomal enzymes and promoted autolysis (point 2).

Many components of damaged cells continued to accumulate in exposed sites. Some of these, especially the deoxyoligonucleotides, are known to stimulate mitosis and cell division (30). This stimulation could account for the fact that hair growth was greater in some of the H-treated areas than in untreated regions. Additional support for this view was the finding that the highest rates of hair growth occurred in those areas which were also treated with DNA repair inhibitors. The largest concentration of DNA fragments should be found here, since recent evidence indicates that proflavine and caffeine interfere with resynthesis necessary for completion of the repair process (31). This suggests that these inhibitors acted at steps b or c in the model. The possibility that the mitotic rate might be artificially stimulated by deoxyoligonucleotides could have profound implications for promoting healing of many different types of injury.

It should be noted that we do not attribute the pathogenesis of H-induced vesication to the inhibition of cell division per se, but rather to aborted metabolic processes which are peculiarly characteristic of dividing cells. Since non-proliferating cells confine their DNA activities primarily to transcription events rather than replication events, many of the reactions which caused tissue damage do not occur. As a consequence, non-proliferating cells and tissues were not seriously affected by DNA-damaging agents.

The foregoing model is entirely consistent with all known facts and current concepts in biochemistry and physiology.

ACKNOWLEDGMENTS: We thank Dr. William C. Bradbury and E-5 Andrew W. Westling for their conceptual and technical contributions during parts of the investigation, E-2 Jerry R. Schroer and E-5 Andrew W. Westling for their invaluable aid in the preparation of the manuscript, Captain Michael J. Dowler for his critical reading of the report and Miss Shirley D. Hinegardner for her outstanding editorial effort and for typing the paper.

LITERATURE CITED

1 Philips, E.S., J. of Pharmacol. Exptl. Therap., Part II, 99:281, (1950).
2 Wheeler, G.P., Cancer Research, 22:651 (1962).
3 Timmis, G.M., Biochem. Pharmacol., 4:49 (1960).
4 Ross, W.C.J. Biological Alkylating Agents, Butterworth & Co., London, 1962.
5 Herriott, R.M., J. Gen. Physiol., 32:221 (1948).
6 Papirmeister, Bruno, CRDL Special Publication 2-45, Army Chemical Center, Maryland, (1961).
7 Loveless, A. and Stock, J.C., Proc. Roy. Soc., B, 150:423 (1959).
8 Lawley, P.D. and Brookes, P., Exptl. Cell Res., Suppl. 9:512 (1963).
9 Crathorn, A.R. and Roberts, J.J., Nature 211:150 (1966).
10 Harold, F.M. and Ziporin, Z.Z., Biochim. et Biophys. Acta, 28:482 (1958).
11 Loveless, A., Proc. Roy. Soc., B, 159:348 (1964).
12 Lawley, P.D. and Brookes, P., J. Mol. Biol., 25:143 (1967).
13 Doskočil, J. and Šormová, Z., Coll. Czechoslov. Chem. Commun., 30:481 (1965).
14 Papirmeister, B. and Davison, C.L., Biochim. et Biophys. Acta, 103:70 (1965).
15 Papirmeister, B. and Davison, C.L., Biochem. Biophys. Res. Commun., 17:608 (1964).
16 Lawley, P.D. and Brookes, P., Nature, 206:480 (1965).
17 Auerbach, C. and Robson, J.M., Nature, 157:302 (1946).
18 Hurwitz, J., et al, J. Cell. Physiol., Suppl. 1, 70, No.2, pp.181, (1967).
19 Richardson, C.C. and Kornberg, A., J. Biol. Chem., 239:242 (1964).
20 Richardson, C.C., et al, Cold Spring Harbor Symp. Quant. Biol., 28:9 (1963).
21 Olivera, B.M. and Lehman, I.R., Proc. Natl. Acad. Sci. US., 57:1426 (1967).
22 Setlow, R.B. and Carrier, W.L., Proc. Natl. Acad. Sci. U.S. 51: 226 (1964).
23 Boyce, R.P. and Howard-Flanders, P., Proc. Natl. Acad. Sci. U.S., 51:293 (1964).
24 Hanawalt, P.C. and Haynes, R.H., Biochem. Biophys. Res. Commun., 19:462 (1965).
25 Bohme, H., Biochem. Biophys. Res. Commun. 28:191 (1967).
26 Clark, A.J. and Margulies, A.D., Proc. Natl. Acad. Sci. U.S., 53:451 (1965).
27 Warthin, A.S. and Weller, C.V., J. Lab. Clin. Med. 3:447 (1918).
28 Harm, W., Mutation Res., 4:93 (1967).
29 Kahl, F.R. and Pearson, R.W., J. Investig. Derm., 49:43 (1967).
30 Braun, W. and Firshein, W., Bacteriol. Rev., 31:83 (1967).
31 Wragg, J.B. and Legator, M.S., Chem. Eng. News, Nov 20 (1967) p.19.

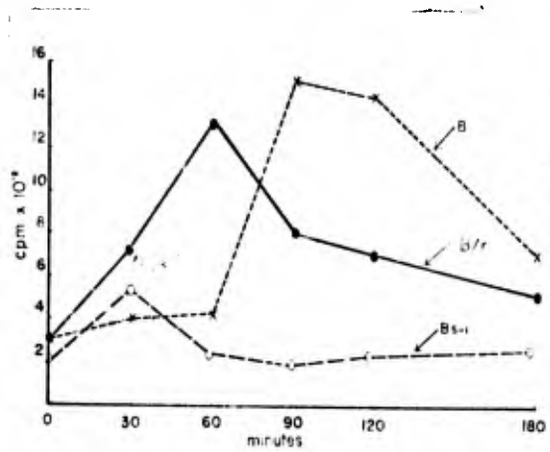


FIG 1 RATE OF ELIMINATION OF H-INDUCED CROSSLINKS FROM DNA E. coli mutants previously exposed to S³⁵ labeled H were incubated for various times. Crosslinks (GHG) isolated from growth medium were counted. *

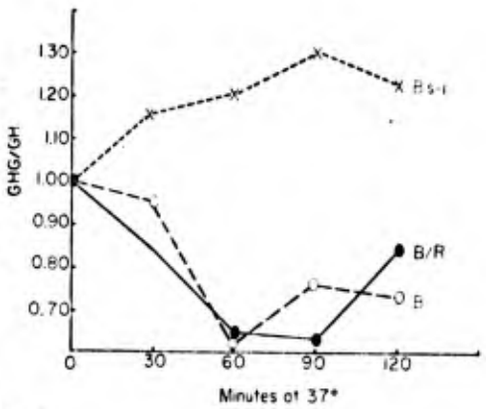


Fig 2 COMPARATIVE RATES OF ELIMINATION OF H-INDUCED CROSS-LINKS AND SINGLE-REMOVAL ALKYLATIONS FROM DNA E. coli mutants previously exposed to ³⁵S labeled H were incubated for various times. The ratio of crosslinks (GHG) to monofunctional alkylations of guanine (GH) remaining in isolated DNA was determined

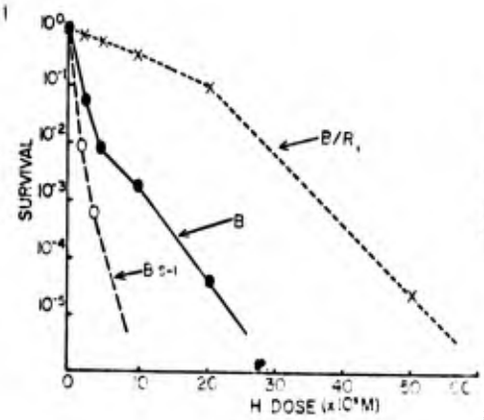


FIG 3 SENSITIVITY OF E. Coli MUTANTS TO H Log phase cells obtained from a tris-buffered synthetic medium were exposed to various concentrations of H. Appropriate aliquots were plated on nutrient agar for determination of colony forming ability

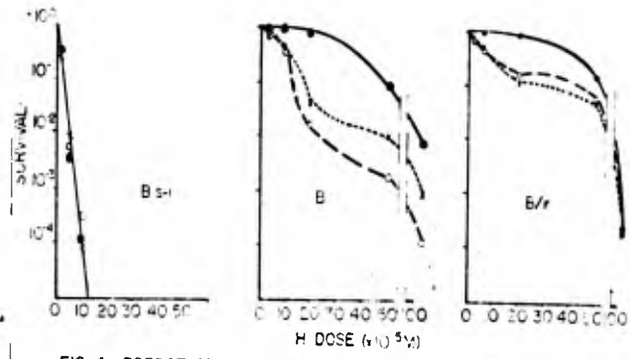


FIG 4 EFFECT OF PRE-GROWTH CONDITIONS ON SENSITIVITY OF E. Coli MUTANTS TO H Bacteria were grown to a limit aerobically in three media with final pH 8.4 (○), pH 7.2 (x), and pH 4.8 (●). After H-treatment, bacteria were plated on nutrient agar for determination of colony forming ability.

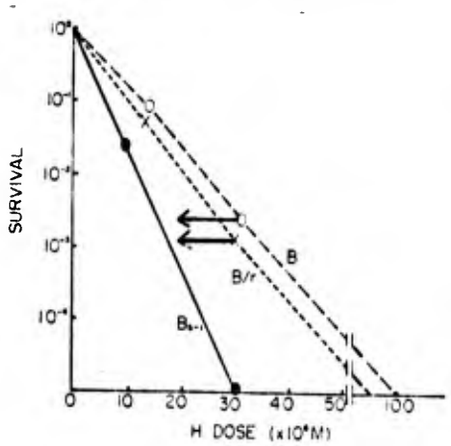


Fig 5 HOST CELL REACTIVATION OF H-TREATED T1 BACTERIOPHAGE BY REPAIR CAPABLE E COLI MUTANTS After exposure to H, viruses were assayed for survival of plaque forming ability, on the various strains. Arrows indicate effects of caffeine and proflavine

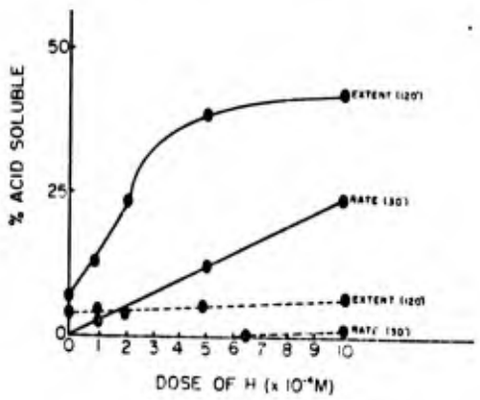


Fig 6. BREAKAGE OF DNA PHOSPHODIESTER BACKBONE BY H Following exposure to various doses of H, aliquots of native (●—●) and heat-denatured (●---●) bacterial DNA were incubated with exo-nuclease for 30 minutes to determine the rate of hydrolysis and for 120 minutes to measure the extent. The enzyme, which acts only on double-stranded DNA possessing interruptions in the backbone, was isolated from E coli 15 TAU.

*For additional details of experimental method, consult authors.

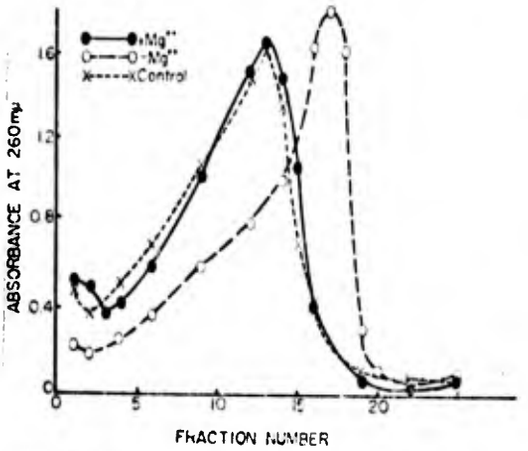


Fig 7 DECREASE IN DNA DENSITY BY H-TREATMENT AND PREVENTION OF THIS DECREASE BY Mg^{++}. Salmon sperm DNA in buffered saline was treated with 5×10^{-4}M H in the presence (●—●) or absence (o--o) of 10^{-2}M $MgCl_2$ and placed on a sucrose gradient of 5% to 20%. After centrifuging for 15 hours at 53,000 x g, fractions were analyzed for UV absorbance at 260mμ. Lowest fraction number had the highest density.

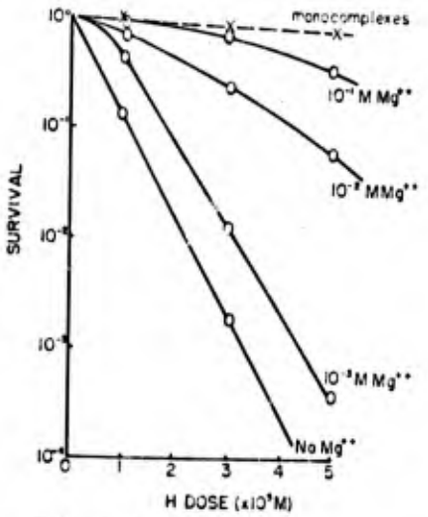


Fig. 8 PRODUCTION OF AN EARLY-STEP DAMAGE TO T1 BACTERIOPHAGE BY H AND PREVENTION OF THIS DAMAGE BY Mg^{++}. Suspensions of purified virus in buffered saline were exposed to H in the presence of indicated concentrations of $MgCl_2$. Aliquots were plated on E coli B_s, for determination of plaque forming ability (solid lines). Note that the survival of particles treated with H in the presence of 10^{-1}M $MgCl_2$ approaches that of monocomplexes formed by preinfecting E. coli B_s, with virus in the presence of chloramphenicol (broken line).

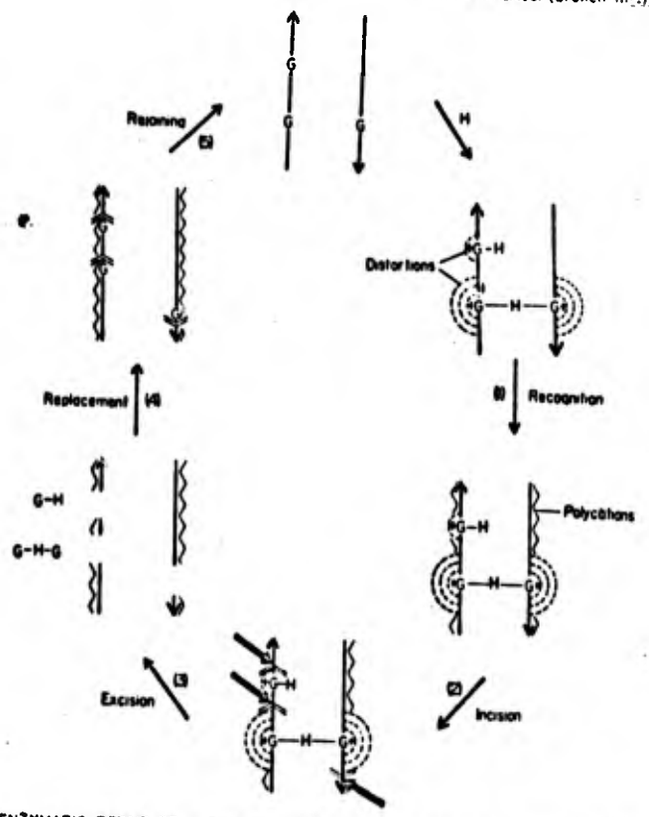


Fig 9 ENZYMATIC REPAIR OF H-DAMAGED DNA. H produces distortions in DNA by reacting with guanine (G). Distortions are greater for crosslinks (GHG) than for single-armed alkylations (GH). Repair of damaged DNA is accomplished by the following major steps: (1) recognition of defects from inability to form protective complexes with polycations (which may be an integral part of a biological membrane), (2) incision at unprotected sites by endonuclease (arrows), (3) excision of defects by exonuclease, (4) replacement synthesis by DNA polymerase utilizing complementary strand as template, and (5) rejoining of ends via phosphodiester bonds by DNA ligase to restore the original molecule.

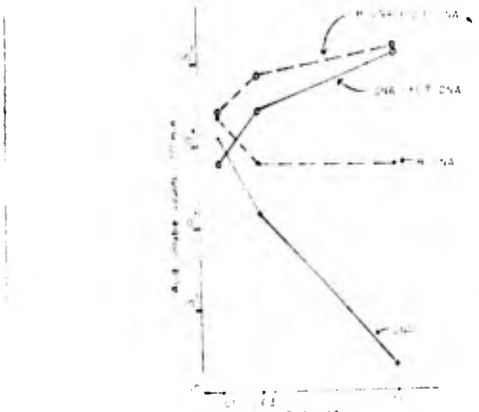


FIG 10. RECOGNITION AND PREFERENTIAL DEGRADATION OF H-TREATED DNA IN AN ARTIFICIAL SYSTEM. Untreated and H-treated radioactively labeled bacterial DNA was incubated with isolated rat liver polyribosomal extract activated by Mn++. At various times of incubation aliquots were withdrawn and the acid-soluble radioactivity was determined. Note that the extract preferentially degraded H-treated DNA. Addition of DNAase increased the rate of solubilization of H-treated DNA.

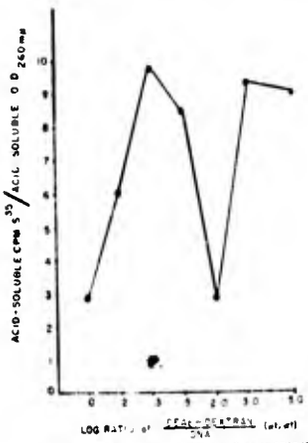


Fig 11. RECOGNITION AND PREFERENTIAL DEGRADATION OF H-TREATED DNA IN AN ARTIFICIAL SYSTEM. After treatment with 5 x 10⁻⁴ M 35S-labeled H. coli thymus DNA was purified, mixed with indicated amounts of DEAE dextran, and incubated with pancreatic deoxyribonuclease (1 μg/ml). The productions of acid-soluble radioactivity and materials absorbing at 260 mμ were compared.

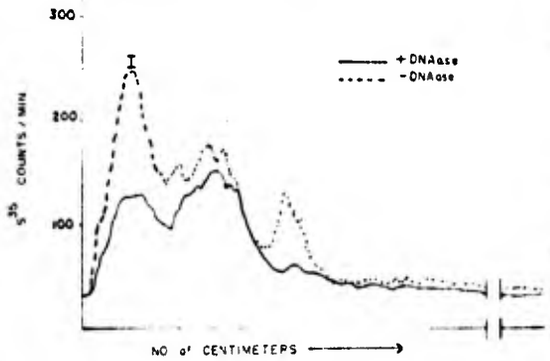


Fig 12. ELIMINATION OF H-INDUCED CROSSLINKS IN AN ARTIFICIAL SYSTEM. The acid insoluble fraction obtained from the Fig 11 experiment (DEAE dextran/DNA = 0.3) was hydrolyzed by heating with 1N HCl, chromatographed, and the distribution of radioactivity determined. Note that DNAase decreased significantly the radioactivity corresponding to crosslinks (peak I).

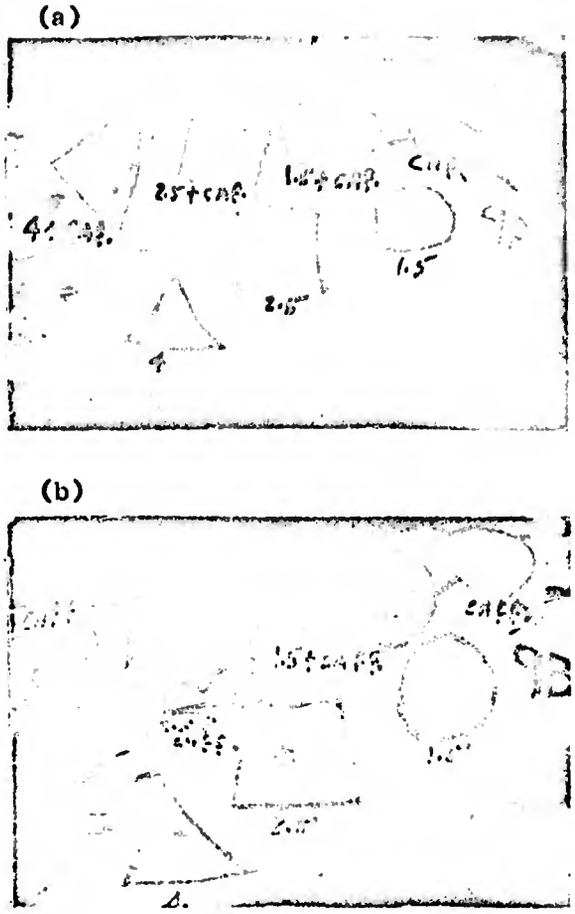


FIG 13. ACCENTUATION OF THE H-INDUCED SKIN LESION BY CAFFEINE. Depilated rabbit skin was exposed in duplicate areas to various doses of H and one set (upper row) was then covered for 48 hours with an ointment base containing caffeine, a DNA repair inhibitor. The second set (lower row) was not further treated. Note that the inhibitor increased the severity of the H lesion after 4 days (a) and enhanced the regrowth of hair after 19 days (b) Caffeine and ointment controls produced no permanent effects.

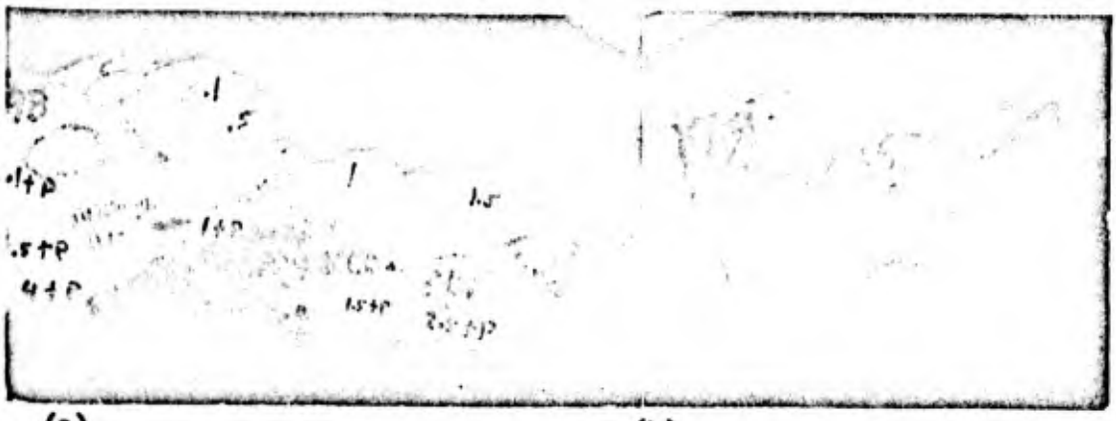


FIG 14. ACCENTUATION OF THE H-INDUCED SKIN LESION BY PROFLAVINE. Procedure was similar to the caffeine experiment (Fig. 13) except that the lower row was treated with the DNA repair inhibitor, while the upper row was treated with only H.

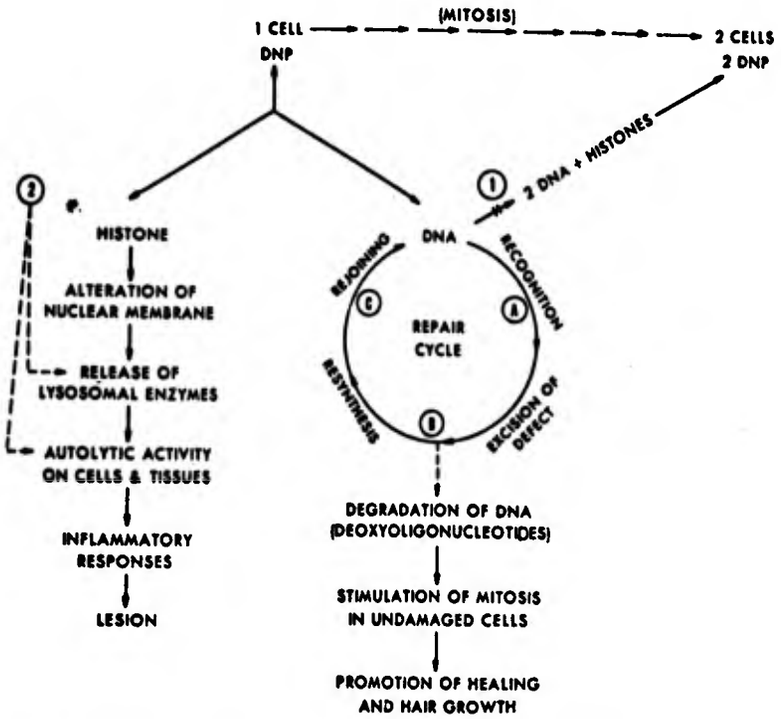


Fig. 15. MODEL ILLUSTRATING RELATIONSHIPS BETWEEN DNA DAMAGE AND DNA REPAIR IN H-TREATED SKIN

(1) SITE OF ACTION OF H; (A) PT. WHERE REPAIR INCAPABLE CELLS (E. COLI B_{s-1}) BLOCKED; (B) + (C) PTS. WHERE REPAIR INHIBITORS BLOCK CYCLE.

(2) PROBABLE SITES OF ACTION OF HEAT AND PROTEOLYTIC ENZYMES.

VISCOELASTIC RHEOLOGY BY ULTRASONIC INTERFEROMETRY

J. G. PARKS
U. S. Army Tank-Automotive Command
Warren, Michigan

INTRODUCTION

Maxwell (1) predicted from a theoretical model of the liquid state that if a liquid were stressed rapidly enough, it would show an elastic response to stress. In other words, a liquid possesses a relaxation frequency below which the liquid acts as a viscous medium and above which elastic behavior is dominate. When a material exhibits a combination of these responses it is known as a "viscoelastic" substance.

Viscoelastic behavior is evident in compressional as well as shear stress. Any complete rheological investigation, therefore, of viscoelastic fluids involves the determination of four basic parameters: shear viscosity and rigidity and bulk viscosity and rigidity. Almost all the information available on the moduli comes from a comparison of the velocity and attenuation of shear waves (governed by the shear parameters) and compressional waves (governed by shear and bulk moduli).

A knowledge of the physical properties of viscoelastic fluids has immediate value in the design of recoil mechanisms, suspension systems, and high speed hydraulic systems. With this last objective in mind, a research project on viscoelastic fluids was initiated by the author in 1967 at the Physical Science Laboratory, USATACOM, to determine the effects of fluid additives on transmission efficiencies. The first phase of that effort, namely the measurement of compressional wave velocities and attenuation factors, is present in this report.

EXPERIMENTAL PROCEDURE

The standard ultrasonic interferometer consists of a source of longitudinal sound waves (normally a piezoelectric crystal) mounted at one end of a fluid (or gas) column and a flat metallic reflector at the other end. By the use of suitable electronic circuits, one may observe the variation of circuit parameters as the

standing wave pattern within the chamber is altered. This pattern change can be affected by either moving the reflector along the principal axis of the chamber or by varying the frequency at which the transducer oscillates: in both cases the ratio of the path length to the wave length of sound in the fluid is changed.

The movable plate, single-crystal interferometer was first constructed by Pierce (2) for the study of gases. The application of this device to liquids was first made by McMillan and Langemann (3). Increased sensitivity, especially in the presence of fluids with large absorption coefficients, can be obtained by replacing the reflecting plate by a second crystal, identical to the transmitting crystal. Changes are detected in the circuit-associated with the second crystal. The two-crystal interferometer was first analyzed by Fry (4).

The variable path interferometer, when applied to liquids, requires adequate seals, accurate linear motion, and sensitive displacement measurements; conditions which complicate an already difficult measurement problem. The physical configuration of an ultrasonic interferometer can, however, be greatly simplified by making the frequency of the applied signal the variable of interest. The chamber-crystal system can then be accurately positioned in a permanent configuration and displacement measurements replaced by frequency determinations.

Care must be exercised in using the fixed-geometry ultrasonic interferometer with fluids which show dispersion. Observations made by Parthasarathy (5) failed to detect evidence of a dependence of velocity on frequency in many organic liquids. The selection of liquids for use in this experiment was largely based on Parthasarathy's results. The work of Rupvano (6) was also reviewed for evidence of dispersion in the standard liquids tested.

Several fixed-geometry, two-crystal ultrasonic interferometers were designed and built at the Physical Science Laboratory, USATACOM, during the Summer of 1957. The chamber material was transparent plastic. This substance was selected for its excellent machining properties and for its value as an aid to the observer in the detection of bubbles during the filling operation. The separations between transducers were on the order of one centimeter. The transducers used in this experiment possessed a natural resonant frequency of one megacycle. The experimental configuration in schematic form is shown in Figure 1.

EXPERIMENTAL CONFIGURATION

The experimental approach was twofold. In the first method the amplitude of the received rectified signal was recorded on an X-Y Plotter (X axis operating in time mode) as the frequency of the oscillator was changed with the aid of a mechanical drive. The frequencies of the extrema of the resulting plot were subsequently measured with the frequency counter. This method will henceforth be referred to as the amplitude approach; it is capable of providing information on both the velocity and the attenuation of the compressional sound wave in the interferometer.

The second measurement technique consisted of an examination of the phase angle difference between the initial and transmitted signals. This frequency variable proved to be relatively insensitive to spurious reflections from the walls of the chamber. This approach was used only to measure velocity and it served as a check on the data obtained in the amplitude procedure.

The accurate determination of crystal separation distance was achieved by using a liquid of known velocity as a calibration standard. Distilled water was used extensively as a calibration liquid because the speed of sound as a function of temperature in this liquid is well known.

THEORETICAL ANALYSIS

Consider the problem of finding the steady state motion of a fluid between two infinite plane parallel boundaries where one of the boundaries is given a prescribed periodic motion normal to its surface and the other boundary is free to move. We may ignore the effects of diffraction provided that, compared with the linear dimensions of the planes, the length of the compressional wave in the liquid is small, and second, that the path length, L, is not large.

The boundaries may be considered as harmonic oscillators with mass M, resistance R, stiffness S, and effective area A. The transmitting crystal is acted upon by a periodic force of the form

$$F(t) = F_o e^{i\omega t} \tag{1}$$

where ω is 2π times the frequency of the applied force and the variable t represents time. It is assumed that the face of the transmitting crystal also behaves in a periodic fashion given by

$$X_1 = |X_1| \, e^{i\omega t} \, e^{i\delta_1} \tag{2}$$

$|X_1|$ is the amplitude of the crystal's oscillation while δ_1 represents the phase difference between the crystal's motion and the periodically applied force.

The equation of motion for the transmitting crystal has the form

$$\ddot{MX}_1 + R\dot{X}_1 + SX_1 + Ac\rho(\dot{\xi}_T - \dot{\xi}_R)_{y=0} = F_o e^{i\omega t} \tag{3}$$

where c is the velocity of propagation of sound in the liquid, ρ, is the liquid's density, and ξ_T and ξ_R are the particle displacements in the medium due to a sound wave traveling from and towards the transmitting crystal, respectively. The variable y denotes the distance measured from the transmitting crystal (at y = 0) to the receiving crystal (at y = L).

Particle displacements have the following form:

$$\xi_T = \xi_o e^{-\alpha y} e^{i\omega t} e^{-i\beta y} \tag{4}$$

$$\xi_R = \xi_o e^{-\alpha(2L-y)} e^{i\omega t} e^{i\beta y} \gamma_L e^{i\delta_L} \tag{5}$$

where α is the attenuation factor, $\beta = \omega/c$, γ_L, the amplitude loss factor due to reflection, and δ_L is the phase change due to reflection.

The receiving crystal has a displacement from equilibrium of X_2.

$$X_2 = |X_2| e^{i\omega t} e^{i\delta_2} \tag{6}$$

and in analogy to Equation (3) the equation of motion for this crystal is

$$\ddot{MX}_2 + R\dot{X}_2 + SX_2 + Ac\rho(\dot{\xi}_T - \dot{\xi}_R)_{y=L} = 0 \tag{7}$$

Equations (2), (4), (5), and (6) may be used to simplify the equations of motion:

$$[(S - M\omega^2) + i\omega R] |X_1| e^{i\delta_1}$$

$$+ i\omega Ac\rho\xi_o [1 - \gamma_L \phi_L \psi_1^2] = F_o \tag{8}$$

$$[(S-M\omega^2) + i\omega R] |X_2| e^{i\delta_2} + i\omega Ac\rho\xi_o [1-\gamma_L \phi_L \psi_2^{-2}] \psi_1 \psi_2 = 0 \tag{9}$$

where $\psi_1 \equiv \exp(-\alpha L)$; $\psi_2 \equiv \exp(-i\beta L)$; $\phi_L \equiv \exp(i\delta_L)$

The boundary conditions needed to solve Equations (8) and (9) amount to synchronous motion between the crystal movement and the total particle displacement. At the transmitting crystal we have

$$\xi_o [1 + \gamma_L \phi_L \psi_1^2] = |X_1| e^{i\delta_1} \tag{10}$$

and at $y = L$ (the receiving crystal) we obtain

$$\xi_o [1 + \gamma_L \phi_L \psi_2^{-2}] \psi_1 \psi_2 = |X_2| e^{i\delta_2} \tag{11}$$

A relationship of interest is obtained from Equations (9) and (11).

$$\gamma_L \phi_L = \psi_2^2 \frac{\omega(Ac\rho + R) - i(S-M\omega^2)}{\omega(Ac\rho - R) + i(S-M\omega^2)} \tag{12}$$

Equations (10), (11), and (12) yield the transfer function

$$\frac{|X_1|}{|X_2|} e^{i(\delta_1 - \delta_2)} = \frac{\psi_1^{-1} + \psi_1 \gamma_L \phi_L}{\psi_2 + \psi_2^{-1} \gamma_L \phi_L} \tag{13}$$

from which the phase difference between the initial and transmitted signal can be extracted.

$$\delta_1 - \delta_2 = \mathrm{Tan}^{-1} [\mathrm{Tanh}(\ln\sqrt{\gamma_L} - \alpha L)\tan(\delta_L)] - \mathrm{Tan}^{-1} [\mathrm{Tanh}(\ln\sqrt{\gamma_L}) \mathrm{Tan}(\delta_L + \beta L)] \tag{14}$$

Reduction of the previous equations provides the following expression for X_2.

$$|X_2|e^{i\delta_2} = \frac{2\omega Ac\rho F_o}{B_+ B_-} (\psi_1^{-1} \psi_2^{-1} - \psi_1 \psi_2)^{-1} \tag{15}$$

where $B\pm = \omega[Ac\rho \pm R] \pm i(S - M\omega^2)$

EXPERIMENTAL RESULTS

1. AMPLITUDE METHOD

The amplitude $|X_2|$ described in Equation (15) has a frequency dependence which, aside from circuit parameter influence, has the following form:

$$|X_2|^2 \propto [\cosh(2\alpha L) - \cos(2\beta L)]^{-1} \tag{16}$$

Expression (16) shows that when $2\beta L$ is equal to even integer multiples of π radians the transmitted signal amplitude will reach a maximum value. Hence amplitude extrema show a periodicity indicated by the expression

$$\Delta(2\beta L) = \frac{4\pi}{c}(\Delta f)L = 2\pi \tag{17}$$

where the symbol Δ represents the change in angular measure between successive maxima or minima.

Several factors tend to destroy the idealization described above. Liquids with low values of attenuation display reflection effects. Attenuation and circuit parameter influences shift the extrema frequency values in certain regions of the usable frequency range. The former problem can best be solved by selecting highly attenuating liquids as the calibration fluid and then examining low attenuation data for systematic patterns. The latter problem can be somewhat reduced if the same frequency range is employed for measurements of all liquids.

Table I shows the results of the measurement of velocity at ambient temperatures for several liquids using the periodicity relationship (17).

TABLE I

AMPLITUDE DATA FOR VELOCITIES

Liquid	Density (gm/cm^3)	Temp. (^oC)	Δf (kc)	Velocity (m/sec)	%Ref.
Glycerol	1.260	25	95.38	1905.7	0.09
Ethylene Glycol	1.113	26	83.01	1658.5	0.09
Distilled Water	0.998	26	75.24	*	-
5606 A Hydraulic Oil	.889	25	68.63	1371.4	0.30
5606 A					
1/2% additive	.891	24	69.09	1380.5	-
1% additive	.893	25	68.95	1377.6	-
2% additive	.898	26	68.98	1378.2	-
Butyl Alcohol	.810	26	62.52	1242.9	0.03
Turpentine	.880	$26\frac{1}{2}$	63.87	1268.0	-
Methanol	.791	25	55.42	1107.3	0.39
2 Propanol	.781	26	57.28	1144.4	-

*Distilled water was used as the calibration liquid.

The percentage errors denoted by "% Ref." refer to the velocities found in Mason's (7) text corrected to the temperatures indicated in the table. The average standard deviation associated with Δf is 0.24%.

The measurement of the attenuation factor demands a more careful examination of X_2. The assumption of identical crystals does not seriously affect the analysis of velocity measurements. In the case of α, however, this assumption should be revoked and one should return to the equations of motion for further analysis. Considerable calculation yields the following relationship.

$$2\alpha L + \phi = 2\mathrm{Tanh}^{-1}\left[\frac{(X_2)_-}{(X_2)_+}\right] \tag{18}$$

where $(X_2)_+$ and $(X_2)_-$ are two points, at the same frequency, on the maxima and minima envelopes of the amplitude plot, respectively. ϕ is a function of frequency and acoustical impedance, $c\rho$. The behavior of ϕ, at a fixed frequency, was found to be a linear function of acoustical impedance.

$$\phi = a(c\rho) + b \tag{19}$$

The evaluation of α then consists of determining the values of a and b in Equation (19) at a particular frequency by examining several liquids with known attenuation factors and then using the density and velocity values of a liquid to be tested to subtract out the contribution due ϕ. Table II represents the results of such an analysis on various liquids at ambient temperatures.

TABLE II

AMPLITUDE DATA ON ATTENUATION
(FREQUENCY= 1.16MC)

Liquid	Cρ (ohms/cm^2)	Zα$_L$ +ϕ (radians)	ϕ (radians)	$\frac{\alpha}{f^2} \times 10^{15}$ (nepers/cm)
Distilled Water	1.495×10^5	.575	.573	.25
Methanol	0.872	.314	.310	.90
Turpentine	1.105	.446	.442	1.50
Castor Oil	1.430	.557	.393	60.0
Butyl Alcohol	1.002	.384	.374	3.7
Ethylene Glycol	1.843	.864	.749	42.7
2-Propanol	.893	.530	.326	7.6
5606 A	1.219	.440	.409	11.5
Ethanol	0.957	.406	.354	19.3

The experimental error associated with attenuation factors presented above is 6% as determined from calibration studies.

2. PHASE DIFFERENCE METHOD

Equation (14) yields a relatively simple expression for the phase difference between input and output signals when the assumptions of zero attenuation and identical crystals are observed.

$$\mathrm{Tan}(\delta_2 - \delta_1) = \frac{\omega R \ \mathrm{Tan}\ \beta L}{1 - (S - M\omega^2)\ \mathrm{Tan}\ \beta L} \tag{20}$$

Several features of Equation (20) are worth noting at this point. First, phase angle differences equal to integer multiples of π radians are coincident with identical values for βL. Second, the right-hand side of Equation (20) becomes infinite when $\tan(\beta L) = (S - M\omega^2)^{-1}$. These features must be modified when the transducers are non-identical (both mechanically and electronically) or when significant attenuation is present. There are, therefore, irregularly spaced frequency positions corresponding to integer multiples of $\pi/2$ radians for the phase difference values.

The first derivative of Equation (20) with respect to ω yields the following expression.

$$\frac{d(\delta_1 - \delta_2)}{d\omega} = \frac{(S+M\omega^2)\mathrm{Tan}^2(\beta L) - \beta L \sec^2 \beta L - \mathrm{Tan}\beta L}{\omega^2 R^2 \ \mathrm{Tan}^2\ \beta L + [1 - (S - M\omega^2)\mathrm{Tan}\ \beta L]} \tag{21}$$

Maxima for Equation (21) occur when either condition described in the preceding paragraph is satisfied. When

$$|\omega_o^2 - \omega^2| \gg 1 \ ; \quad \omega_o^2 = \frac{S}{M} \tag{22}$$

one should expect the maximum values of $d(\delta_1 - \delta_2)/d\omega$ to have the following periodicity.

$$\Delta(\beta L) = \frac{2\pi}{C} \Delta f \cdot L = \pi \tag{23}$$

The quantity selected for measurement in this experiment was $d(\delta_1 - \delta_2)/d\omega$ because (1) it was a variable compatible with the most accurate measurement procedure and (2) the influence of circuit phenomena was reduced in this variable as compared to $(\delta_1 - \delta_2)$.

The transducer resonance region, characterized by the condition

$$\omega_o^2 - \omega^2 \approx 0$$

was avoided and maximum values of $d(\delta_1 - \delta_2)d\omega$ were used to obtain the velocity data. Table III shows the results of this measurement.

TABLE III

VELOCITY DATA FROM PHASE DIFFERENCE MEASUREMENTS

Liquid	Temp. (°C)	Δf (kc)	%Stat.	%Ref.	Velocity (m/sec)
Glycerol	23	96.33	1.23	0.68	1895
Ethylene Glycol	23	84.34	1.11	0.18	1695
Distilled Water	24	76.05	0.16	*	1496
Castor Oil	25 1/2	76.79	0.87	0.20	1472
5606 A	24	68.82	0.53	-	1354
Benzene	24	66.06	0.66	0.30	1300
Turpentine	24	64.43	0.73	0.71	1267
Propanol	22 1/2	57.15	0.10	-	1124
Silicone Oil(5000 Centistokes)	23	49.85	0.07	-	981

*Distilled water used as calibration fluid.

DISCUSSION AND CONCLUSIONS

The velocity measurements obtained using the amplitude method were nominally accurate to within 0.5% for fluids with attenuation factors greater than 0.015 nepers per centimeter. For fluids with attenuation factors less than 0.015 nepers per centimeter, scattered waves accounted for most of the difficulty in measuring velocities with the amplitude method.

The phase angle measurements were made using a dual-beam oscilloscope. The signals were matched in amplitude, and displacements along the zero voltage level were used as indications of phase difference. Such measurements are subject to observer error in reading the oscilloscope trace. This technique of measurement was employed since commercially available phase angle measuring equipment with the sensitivity required for this experiment is limited to frequencies below twenty kilocycles.

These problems possibly could be eliminated if the measurement procedure consisted of examining the Lissajou pattern formed by two signals at points where the phase difference between them was an integral multiple of π radians. Since these positions correspond to maxima in both $d(\delta_2 - \delta_1)/d\omega$ values and signal amplitudes for frequencies differing from the transducer resonance region, they could be used to determine the periodicity. This technique would require, however, the insertion of impedance matching elements to extend this method into the range of frequencies where harmonic generation is presently observed.

BIBLIOGRAPHY

1. Maxwell, J. C., Philosophical Transactions, 162, 49, 1866.

2. Pierce, R. M., Proceedings of the American Academy of Arts and Sciences, Vol. 60, No. 5, 1925.

3. McMillan, D. R., and Langemann, R. T., Journal of Acoustical Society of America, Vol. 19, 1947.

4. Fry, J. W., Journal of Acoustical Society of America, Vol. 25, 1953.

5. Parthasarathy, F., Sci. Industrial Research, Vol. 3, 299, 1943.

6. Rupuano, J., Physical Review, Vol. 12, 79, 1947.

7. Mason, W. P., "Piezoelectric Crystals and Their Application to Ultrasonics", Van Nostrand, 1952.

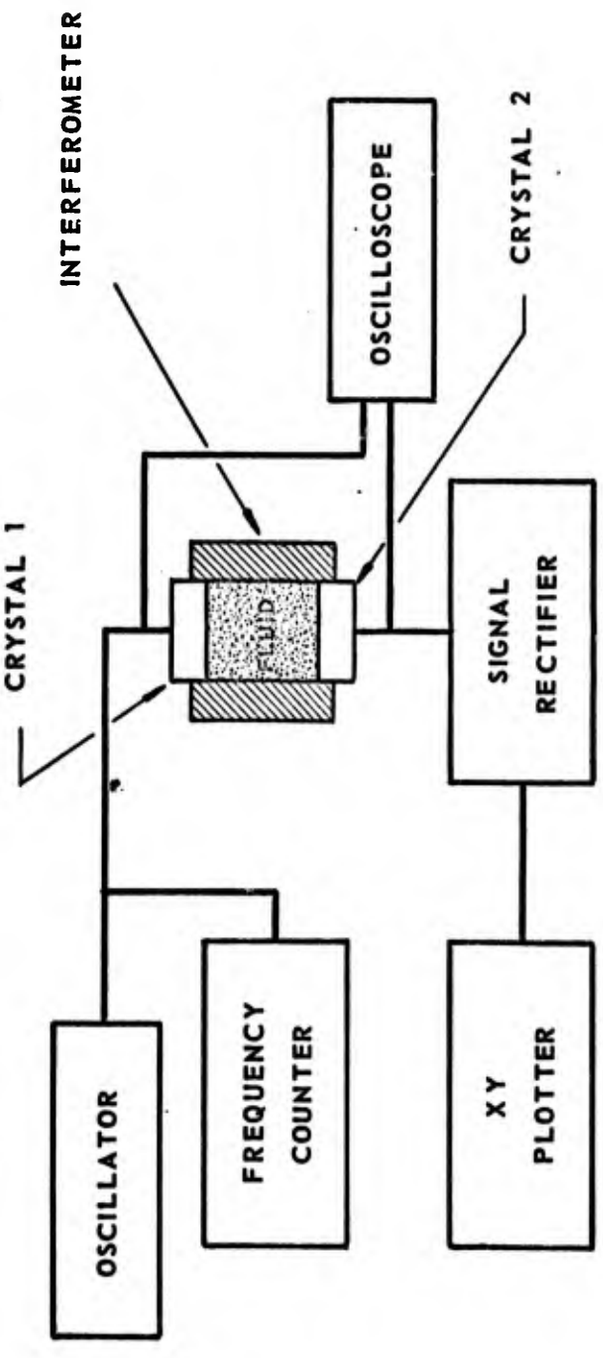


FIGURE 1
EXPERIMENTAL CONFIGURATION

ELECTROLESS DEPOSITION OF NICKEL AND COBALT BASED ALLOYS

FRED PEARLSTEIN AND ROBERT F. WEIGHTMAN
PITMAN-DUNN RESEARCH LABORATORIES, FRANKFORD ARSENAL
PHILADELPHIA, PA. 19137

Introduction

The electroless deposition of nickel and cobalt from hypophosphite-based solutions has found numerous applications by virtue of characteristics unmatched by electrodeposition. For example, since deposition is produced by contact of the catalytic surface with the solution, exceptionally good deposit uniformity may be obtained. Also, the presence of phosphorus (from hypophosphite) in the deposits provides exceptional hardness (1) and corrosion resistance (2) to nickel and valuable magnetic properties (3) to cobalt. Electroless deposits may, in addition, be directly applied to non-conductive materials (4). Electroless deposition of copper (5) and palladium (6) has also been successfully applied utilizing formaldehyde and hydrazine reducing agents respectively. In spite of intensive investigations in this area, the electroless deposition process, at present, is limited to the four metals mentioned above. It was felt that electroless deposits with a wider variety of chemical, physical, and magnetic properties might be achieved by means of codeposition of other metals during electroless nickel or cobalt deposition. Although hydrazine, borohydride, or amine borane may be utilized as reducing agents for electroless nickel or cobalt deposition, this investigation is limited to the more commonly used hypophosphite solutions.

Nickel-tungsten and nickel-molybdenum electrodeposits (7) are well known even though neither tungsten nor molybdenum can be electrodeposited alone from aqueous solutions. Although the mechanism is not clear, the electrolytic reduction of nickel ions apparently induces reduction of tungsten and molybdenum salt to the metal. Cobalt behaves similarly to nickel in this regard. It was reasoned that the catalytic chemical reduction (electroless deposition) of nickel or cobalt might induce simultaneous reduction of tungsten and molybdenum as well as other metals.

PEARLSTEIN, WEIGHTMAN

Electroless Nickel Alloys

Experimental

Additions of sodium tungstate or molybdate were made to Brenner's acid-citrate electroless nickel solution (8):

Nickel sulfate $.6H_2O$ - 35 g/l
Sodium citrate $.2H_2O$ - 8.4 g/l
Sodium hypophosphite $.H_2O$ - 10 g/l

The effects of citrate concentration and solution pH on deposit composition were determined. Electroless deposits were produced by immersing activated* platinum (50 cm^2) in 500 ml of the test solution at 98 ± 1°C for one hour. Deposit composition was determined by wet chemical analytical procedures.

The feasibility of producing other electroless nickel alloy deposits by induced chemical reduction was investigated by adding selected metal salts to each of three types of electroless nickel solutions: acid citrate, acid-acetate and ammoniated alkaline-citrate. Deposits were produced on activated (4) glass microscope slides and analyzed spectrographically for presence of elements codeposited with electroless nickel. Solutions indicated as promising for deposition of electroless nickel alloys were used for deposition on activated platinum as described above. The deposits were then analyzed for composition.

The efficiency of utilization of hypophosphite was determined during several electroless nickel and alloy deposition tests. Hypophosphite concentrations were determined using the McCloskey (9) method.

Alloy deposits on steel (as-deposited or after heat treatment at 400°C for one hour) were cross-sectioned and tested for diamond pyramidal hardness.

Selected alloy deposits were produced on activated glass microscope slides and were flaked off for determination of melting point. The alloy deposits were heated in a furnace until molten and then gradually cooled with the melting point determined by the isothermal interval during solidification.

Alloy deposits were applied to activated platinum and tested for resistance to 60 per cent (volume) hydrochloric acid or concentrated nitric acid.

---

* Two minutes in 5 g/l $SnCl_2$-10 ml/l HCl; rinse;
Two minutes in 0.2 g/l $PdCl_2$-1 ml/l HCl; rinse.

PEARLSTEIN, WEIGHTMAN

Alloy deposits of eight micron thickness were applied to steel panels and exposed to five per cent neutral salt spray for determination of corrosion resistance and protection.

Results and Discussion

The addition of sodium tungstate to the acid-citrate electroless nickel solution in the usual pH range of four to six resulted in little more than trace amounts of tungsten in deposits. However, when the pH was raised above seven (with reduced nickel content to prevent precipitation of hydroxides) substantial tungsten was produced in resultant deposits.

The effects of citrate concentration in the following evolved solution were determined in regard to electroless deposition rate and deposit composition:

Nickel sulfate $.6H_2O$ - 7 g/l
Sodium tungstate $.2H_2O$ - 35 g/l
Sodium hypophosphite $.H_2O$ - 10 g/l

The solutions were adjusted to a pH of 8.2 at room temperature prior to use. The results are shown in Figure 1. The tungsten content of deposits reached a maximum level of about 20 per cent by weight with 40 g/l sodium citrate dihydrate present in the solution. The deposition rate steadily decreased with increasing citrate concentration.

The effect of solution pH on electroless nickel-tungsten deposit composition is shown in Table I. It is evident from the deposit compositions that the tungsten is present in elemental form and not as occluded oxides since the entire deposit is accounted for by the determined quantities of nickel, phosphorus and tungsten. The phosphorus content of deposits is lowest and tungsten content highest at solution pH 8.2.

The composition of deposits from the solutions at pH 8.2 containing 20 or 40 g/l sodium citrate dihydrate were determined at 68, 78 and 88 as well as 98°C. The solution temperature had little effect on deposit composition although, of course, the deposition rate decreased with decreasing temperature.

The addition of molybdate to the acid citrate electroless nickel solution prevented formation of any electroless deposits apparently by poisoning of the catalytic surface.

The results of spectrographic analysis of deposits produced from electroless nickel solution to which metal salts were added are shown in Table II. Codeposition of palladium, rhenium, tungsten, zinc or tin with electroless nickel is indicated from the ammoniated alkaline-citrate solution.

The alloy composition of deposits produced from the ammoniated alkaline-citrate solution with various metal salt additives are shown in Figure 2. Increasing zinc sulfate concentration of solution resulted in decreasing deposition rates but increasing zinc content of deposits. It is surprising that as much as 15 per cent zinc was produced in deposits considering the high electronegativity of zinc. On the other hand, tin which is much more noble than zinc and readily forms alloy electrodeposits with nickel, was not codeposited in substantial quantities with electroless nickel when sodium stannate was added to the solution. However, it has not yet been established that zinc or tin is present in deposits in the elemental form.

Increasing ratio of tungsten to nickel ion in the ammoniated alkaline-citrate solution resulted in deposits of increasing tungsten content to a maximum of approximately 12 per cent. Increasing perrhenate concentration in the electroless nickel solution resulted in increasing rhenium content of deposits to a maximum of 46 per cent at a concentration of potassium perrhenate corresponding to 1.54 g/l. It is also noteworthy that the deposit contained only two per cent phosphorus compared to about five per cent phosphorus in the electroless nickel deposit.

Rhenium is preferentially deposited to nickel in that deposits contain a larger proportion of rhenium than is present in the solution. Yet attempts to produce electroless rhenium deposits without presence of nickel ions were fruitless.

Electroless nickel-rhenium deposits were also produced containing the additional element of tungsten, zinc, or tin by addition of the appropriate metal salt to the solution that deposits a nickel - 46 per cent rhenium alloy. The following are examples of such alloys:

55.5% Ni - 32.7% Re - 5.1% P - 6.6% Zn
53.0% Ni - 44.1% Re - 1.4% P - 1.1% Sn
55.7% Ni - 31.5% Re - 2.6% P - 10.1% W

Electroless palladium deposition. Additions of palladium chloride to the ammoniated alkaline-citrate electroless nickel solution caused the bath to decompose after short use at the high temperature (98°C) of operation. Palladium nuclei are produced throughout the solution which provide a large surface area for rapid catalytic decomposition of the hypophosphite. This led to attempts to effect stabilization of an electroless palladium solution which might be useful in pursuing the development of electroless palladium alloys with nickel or other metals. The following solution, capable of electroless palladium deposition with good solution stability at temperatures up to 70°C was developed:

$PdCl_2$	2 g/l
NH_4OH	160 ml/l
NH_4Cl	26.8 g/l
$NaH_2PO_2.H_2O$	10 g/l

The solution was used in one case, for depositing palladium on a small specimen. This was done over a six day period at 40°C after which time a deposit approximately 175 microns thick resulted. The Vickers hardness number of the deposit was about 165 kg/mm^2. Addition of 30 g/l nickel sulfate hexahydrate was made to the above electroless palladium solution and the deposit analyzed for nickel content. Palladium was preferentially deposited and only 6.0 per cent nickel was produced in the deposit along with 2.6 per cent phosphorus. Palladium alloys may be useful as diffusion barrier layers and as coatings for electronic components. The possibility of producing other electroless palladium alloys is currently under investigation.

Spectrophotometric studies. The mechanism of induced codeposition of metal with nickel is not known. It was considered that the alloy may be produced by discharge of a complex ion containing both elements of the alloy. Attention was directed to the electroless nickel-rhenium plating solution from which deposits containing 46 per cent rhenium were obtained. Visible and ultraviolet spectrophotometric analyses were made of the solution with and without perrhenate present in order to ascertain presence of nickel-rhenium complexes. No evidence of such complexes was found from spectrophotometric studies.

Hypophosphite utilization. It was considered of interest to determine the efficiency of utilization of hypophosphite during deposition of the nickel-rhenium alloy compared to electroless nickel alone. A search of the literature revealed that about 30 to 40 per cent of the hypophosphite comsumed is utilized for reduction of nickel from acid electroless nickel plating solutions with the deposit assumed to be solely nickel (10). No references were found for hypophosphite utilization from alkaline electroless nickel solutions. Perhaps one reason for others not considering phosphorus reduction in calculations is that the phosphorus reduction mechanism has not yet been established. However, it has been shown that phosphorus in deposits may originate from both hypophosphite and phosphite (11) and thus may involve a one or three electron change. For the sake of calculations of hypophosphite utilization in this report a two electron change is assumed.

An acid-citrate electroless nickel solution was tested for hypophosphite utilization and found to be 42.1 per cent efficient. (See Table III.) When no account is taken of phosphorus content but the entire deposit assumed to be nickel, the efficiency of hypophosphite utilization is 39.3 per cent which is in the range of reported values. Hypophosphite utilization tests were made with

three alkaline electroless nickel and three electroless nickel-rhenium plating solutions. The results are shown in Table III. The alkaline electroless nickel solution averaged about 46.6 per cent efficiency of hypophosphite utilization or about 4.5 per cent higher than the acid solution. The higher deposition efficiency from the alkaline solution is probably due to decreased hydrogen ion concentration available to compete with nickel for reduction. However, there may also be a factor of chemical dissolution of nickel from acid solutions with an apparent decrease in reduction efficiency. The nickel-rhenium plating solution operated at even higher efficiency, averaging about 55.7 per cent. It is thus indicated that the nickel-rhenium alloy is more readily deposited than electroless nickel alone. The presently most acceptable theory of electroless deposition mechanism is the hydride theory (12) which predicts that deposition efficiency cannot exceed 50 per cent. For this reason, the experimental evidence of deposition at greater than 50 per cent efficiency has theoretical significance and may indicate a need for modification of the present theory.

Properties of Deposits

Electroless nickel deposits containing zinc, tin, tungsten or rhenium were in general not as hard as conventional nickel deposits from the ammoniated alkaline-citrate solution either in the as-plated or heat treated condition.

The melting point of electroless nickel deposits from the ammoniated alkaline-citrate solution was found to be 885°C. A nickel-tungsten deposit of about 13 per cent tungsten had a melting point of about 1400°C and a nickel-rhenium alloy of 46 per cent rhenium of about 1700°C. Thus, there may be high temperature applications for the alloy deposits.

Electroless nickel - 20% tungsten deposits were quite resistant to concentrated nitric acid being virtually unattacked after 24 hours' immersion. Ordinary electroless nickel deposits are rapidly attacked by concentrated nitric acid. Although electroless nickel is only slowly attacked by 60 per cent (volume) hydrochloric acid, the tungsten alloy was even more resistant to attack. Electroless nickel-zinc deposits were on the other hand rapidly attacked by 60 per cent hydrochloric acid.

Eight micron deposits on steel of electroless nickel deposits containing 11 per cent zinc, two per cent tin, 13 per cent tungsten or 46 per cent rhenium were exposed to five per cent neutral salt spray. Deposits with zinc provided little corrosion protection. The other alloy deposits were comparable to electroless nickel in protective value with the rhenium alloy somewhat superior and the tungsten alloy somewhat inferior. However, the tungsten alloy deposit was somewhat more resistant to tarnishing than electroless nickel. A duplex deposit of electroless nickel followed by

electroless nickel - 20% tungsten alloy appears promising for providing both corrosion protection and tarnish resistance.

Electroless Cobalt Alloys

Experimental

Additions of tungstate, perrhenate, stannate, zinc or nickel ions were made to Brenner's (8) ammoniated alkaline-citrate electroless cobalt solution:

Cobalt chloride $.6H_2O$	- 30 g/l
Sodium citrate $.2H_2O$	- 80 g/l
Ammonium chloride	- 50 g/l
Sodium hypophosphite $.H_2O$	- 20 g/l
Ammonium hydroxide (29%)	- 60 ml/l

Electroless deposits were produced by immersing activated platinum (50 cm^2) in 500 ml of test solution at 95 ± 1°C for one hour. Wet chemical analytical procedures were used for determination of deposit composition.

The principle application of electroless cobalt is for memory storage devices by virtue of the magnetic properties of the deposits. Thus electroless cobalt alloys were examined for coercivity and remanence compared to conventional electroless cobalt deposits using a Sperry-Rand B-H Loop Tester with drive field of 60 cycles/sec and field intensity of 1500 oersteds (13). The electroless deposits for magnetic property determinations were produced upon chemically polished copper strips (1.1 cm x 10.2 cm x 0.08 cm) activated by 30 second immersion in 0.1 g/l $PdCl_2$ - 0.1 ml/l HCl at 25°C.

Results and Discussion

The addition of zinc sulfate or sodium stannate to the electroless cobalt solution failed to produce significant zinc or tin in resultant electroless deposits. The effect of tungstate, perrhenate or nickel ion additions to electroless cobalt solution on the composition of deposits are shown in Figures 3, 4 and 5. The first two are examples of induced codeposition while nickel is presumably codeposited by independent ability for electroless deposition. Electroless cobalt alloy deposits were produced containing as much as nine per cent tungsten. As much as 30 per cent rhenium was produced in deposits by addition of only 0.8 g/l potassium perrhenate to the electroless cobalt solution. Higher perrhenate concentrations caused the deposition rate to decrease to very low values. The phosphorus content and deposition rates generally decreased with increasing tungsten or rhenium content of deposits. Increasing concentrations of nickel sulfate to the electroless cobalt solution resulted in increasing deposition rates and increasing phosphorus content of deposits.

It was found that the addition of trace quantities of certain organic compounds affected the composition of electroless cobalt deposits. For example, 0.9 mg/1 phenylthiourea added to the electroless cobalt solution resulted in a decreased phosphorus content of deposits from 4.6 to 3.3. Some sulfur may also be incorporated into deposits but analysis was not made for this element.

Properties of Electroless Cobalt Alloys

Additions of one to ten g/1 sodium tungstate dihydrate to electroless cobalt solution resulted in deposits with coercivity only moderately increased over electroless cobalt alone. The remanence of deposits was not affected.

Additions of 0.3 g/1 potassium perrhenate to electroless cobalt solution provided deposits of moderately decreased coercivity but of considerably reduced remanence.

Marked increase in coercivity and remanence was provided by electroless cobalt deposits from solution containing 1 to 3 g/1 nickel sulfate hexahydrate. However, when 10 g/1 nickel sulfate was added, the coercivity of deposits became very low while remanence was retained. Such deposits may be useful for switching mechanisms since reversal of polarity may be effected with little energy consumption.

The hysteresis loops of some of the electroless cobalt alloys discussed were examined. These are shown in Figure 6 for deposits in the thickness range of 11,000 to 13,000 Å. The hysteresis loop was little affected by the deposition of cobalt-tungsten alloy. The hysteresis loop for deposits from solution containing 0.3 g/1 potassium perrhenate reflects the considerably decreased remanence. Addition of 3 g/1 nickel sulfate hexahydrate provided deposits of increased squareness as well as increased coercivity and remanence. The addition of 10 g/1 nickel sulfate provided deposits of very low coercivity but considerable remanence. Addition of 0.9 mg/1 phenylthiourea to solution produced an electroless cobalt deposit of the hysteresis loop shape shown in Figure 6 reflecting increased remanence and decreased coercivity. Thus electroless cobalt alloys may be produced to provide a variety of magnetic properties.

Little attention has been given to the corrosion protective properties of electroless cobalt. Recent studies by the authors indicate that electroless cobalt deposits provide potentially valuable protective properties to steel particularly when preceded by an electroless nickel deposit. However, electroless cobalt deposits tarnish readily when exposed to saline conditions to form a most unpleasant-appearing mottled brown condition. It was found that electroless cobalt deposits from solution containing ten g/1 sodium tungstate dihydrate or 0.9 mg/1 phenylthiourea were notably improved in resistance to tarnishing.

Conclusions

The autocatalytic chemical reduction (electroless deposition) of nickel or cobalt from alkaline solutions induces simultaneous reduction of other metal ions such as tungsten, rhenium, tin, or zinc.

Electroless nickel-tungsten alloy deposits containing up to 20 per cent tungsten are more resistant than electroless nickel to acid attack and more resistant to tarnishing during salt fog exposure. Electroless deposits of nickel-rhenium alloy contain as much as 46 per cent rhenium. The rhenium alloys are much higher in melting point than conventional electroless nickel making them potentially useful for high temperature applications. In spite of the high electronegativity of zinc, electroless nickel deposits containing as much as 15 per cent zinc are produced. Tin is also codeposited during electroless nickel deposition but only to the extent of about two per cent.

An electroless palladium-nickel alloy containing six per cent nickel is produced by addition of nickel sulfate to a newly developed electroless palladium solution.

Electroless cobalt solutions to which were added tungstate, perrhenate, nickel salt or phenylthiourea provided deposits with a wide variety of magnetic properties.

The efficiency of hypophosphite reducing agent increases during electroless deposition in the following order: acid electroless nickel, alkaline electroless nickel, alkaline electroless nickel-rhenium.

References

1. Randin, J. P., Hintermann, H. E., Plating, 54, 523 (1967).

2. Minjer, C. H., Brenner, A., Plating, 44, 1297 (1957).

3. Sallo, J. S., Plating, 54, 257 (1967).

4. Pearlstein, F., Metal Finishing, 53, 59 (1955) Aug.

5. Cahill, A. E., Proceedings Amer. Electroplaters' Soc., 44, 130 (1957).

6. Rhoda, R., U. S. Patent 2,915,406 (1959).

7. Holt, M. L., Metal Finishing, 54, 48 (1956) Sept.

8. Brenner, A., Riddell, G., Proceedings Amer. Electroplaters' Soc., 34, 156 (1947).

9. McCloskey, J. P., Plating, 51, 689 (1964).

10. Gorbunova, K. M., Nikiforova, A. A., Physiocochemical Principles of Nickel Plating, p. 16, (translated from the Russian), Israel Program for Scientific Translations, Jerusalem (1963).

11. Gorbunova, K. M., Nikiforova, A. A., Zhurnal Fizicheskoi Khimii, 31, 1687 (1957).

12. Lukes, R. M., Plating, 51, 969 (1964).

13. Pearlstein, F., Weightman, R. F., Plating, 54, 714 (1967).

Acknowledgement

The authors wish to acknowledge with thanks the many helpful suggestions of A. Gallaccio of Frankford Arsenal.

Table I. Effect of Solution pH on Composition of Electroless Nickel-Tungsten Alloy Deposits

Solution pH	Deposit Composition, per cent by weight (20 g/l sodium citrate.$2H_2O$) Nickel	Tungsten	Phosphorus	(40 g/l sodium citrate.$2H_2O$) Nickel	Tungsten	Phosphorus
5.5	-	-	-	87.50	2.88	9.48
7.0	87.50	6.68	4.80	74.64	16.58	8.62
8.2	83.75	12.68	4.38	73.50	20.48	6.53
9.5	-	-	-	72.00	18.19	9.80

Table II. Spectrographic Analysis of Deposits from Electroless Nickel Solutions to Which Various Metal Salts Were Added

Metal Salt Added	Concentration	Results of Spectrographic Tests: Acid-citrate	Acid-acetate	Ammoniated Alkaline-citrate
$AuCl_3.3H_2O$	0.2 g/l	(unstable solution)	(unstable solution)	(no deposit)
$PdCl_2$	"	(unstable solution)	(unstable solution)	*
$H_2PtCl_6.6H_2O$	"	(no deposit)	(no deposit)	Trace
$RhCl_3.3H_2O$	"	None	None	None
$KReO_4$	"	Trace	Trace	*
$CuSO_4.5H_2O$	0.005 moles/l	(no deposit)	(no deposit)	(no deposit)
$Na_2MoO_4.2H_2O$	"	(no deposit)	(no deposit)	(no deposit)
$Na_2WO_4.2H_2O$	"	(no deposit)	(no deposit)	*
$ZnSO_4.7H_2O$	"	(no deposit)	(no deposit)	*
$Na_2SnO_3.3H_2O$	"	-	-	*

* Detection of substantial quantities of metal codeposited with electroless nickel.

Table III. Efficiency of Utilization of Hypophosphite

Solution Type	Deposit Composition: Nickel	Phosphorus	Rhenium	Per Cent Utilization of Hypophosphite
Acid-citrate	92.58	7.88	-	42.1
Alkaline-citrate	95.72	5.26	-	47.3
"	95.48	4.94	-	44.5
"	96.20	4.62	-	48.5
"	55.71	2.23	43.77	51.7
"	56.21	2.26	45.73	58.8
"	52.61	1.71	46.17	56.6

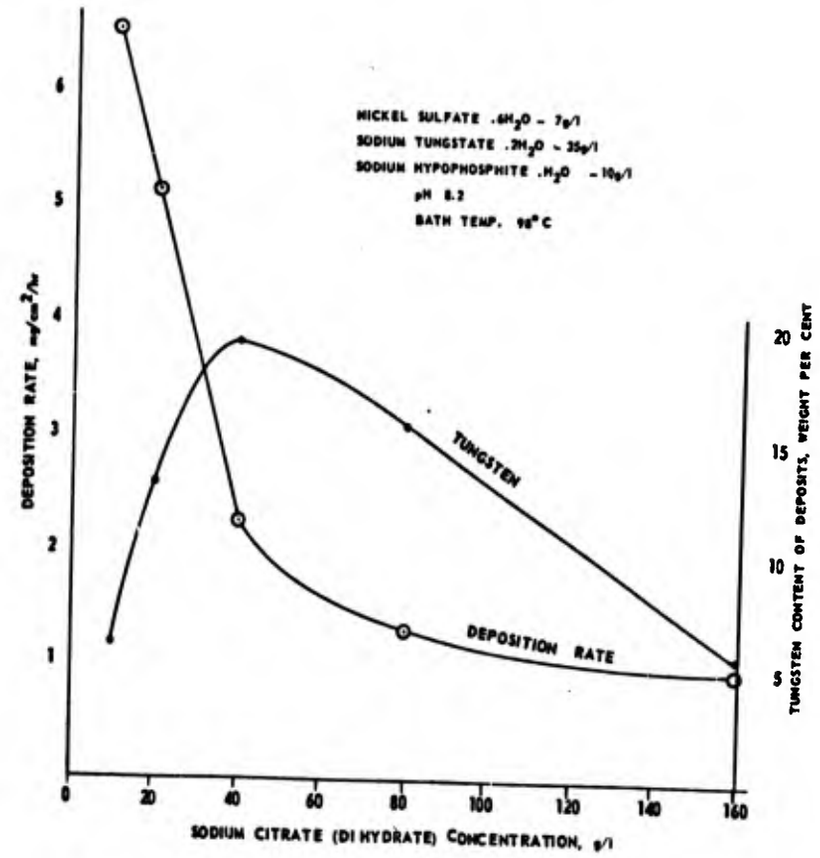


FIGURE 1. Effect of citrate concentration on electroless deposition of nickel-tungsten alloys.

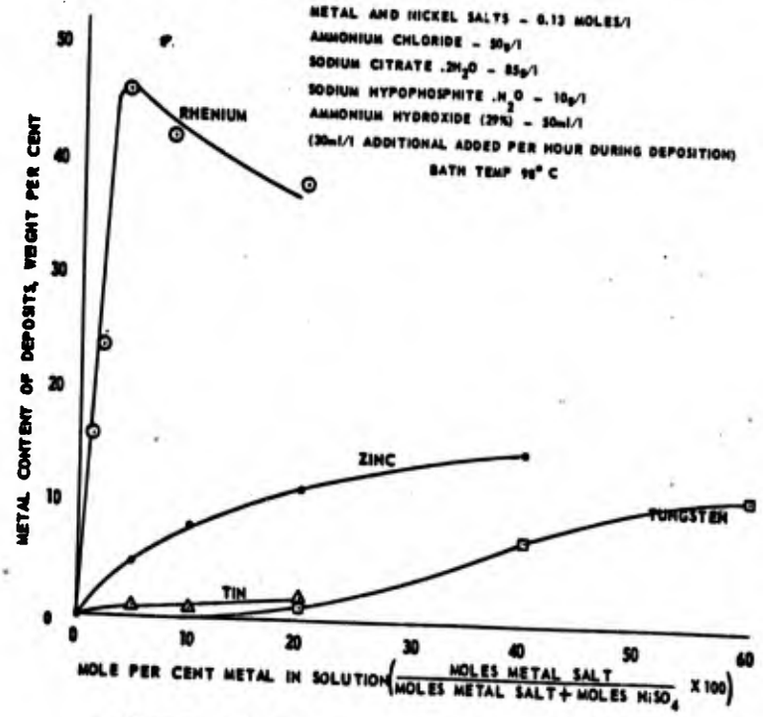


FIGURE 2. Effect of additions of perrhenate, tungstate, stannate or zinc ions to ammoniated alkaline-citrate electroless nickel solution on composition of deposits.

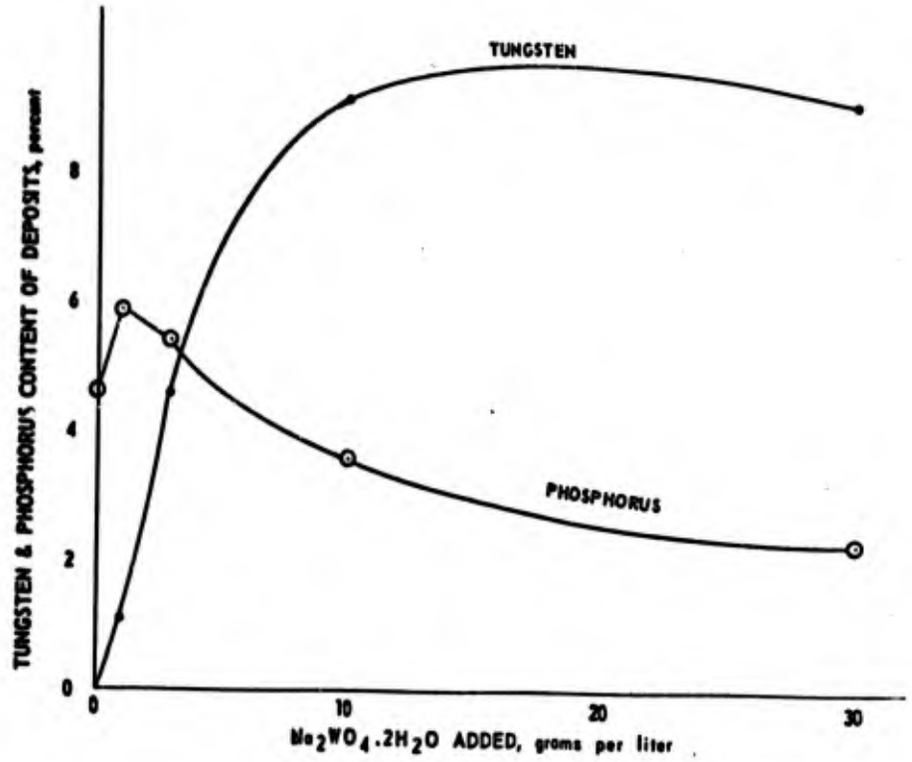


FIGURE 3. Effect of sodium tungstate additions to electroless cobalt solution on the composition of deposits.

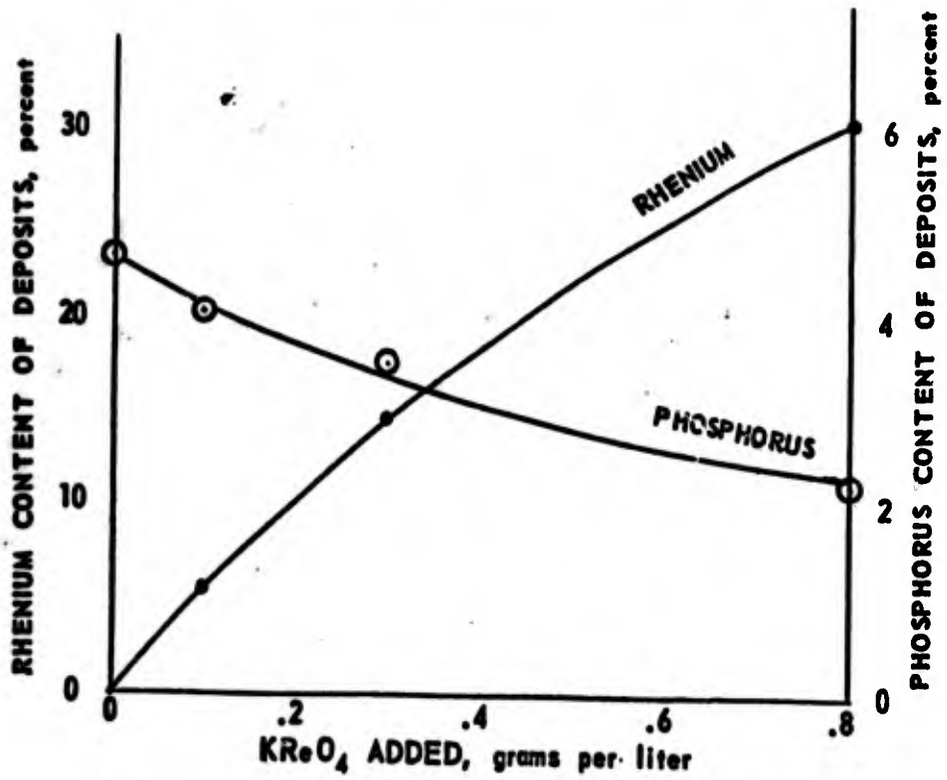


FIGURE 4. Effect of potassium perrhenate additions to electroless cobalt solution on the composition of deposits.

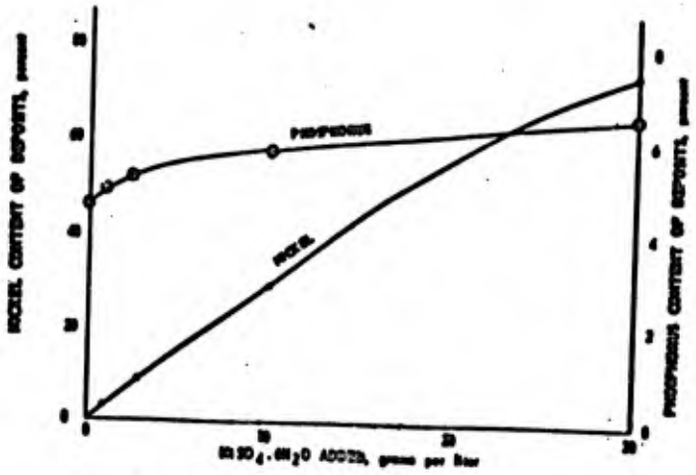


FIGURE 5. Effect of nickel sulfate additions to electroless cobalt solution on the composition of deposits.

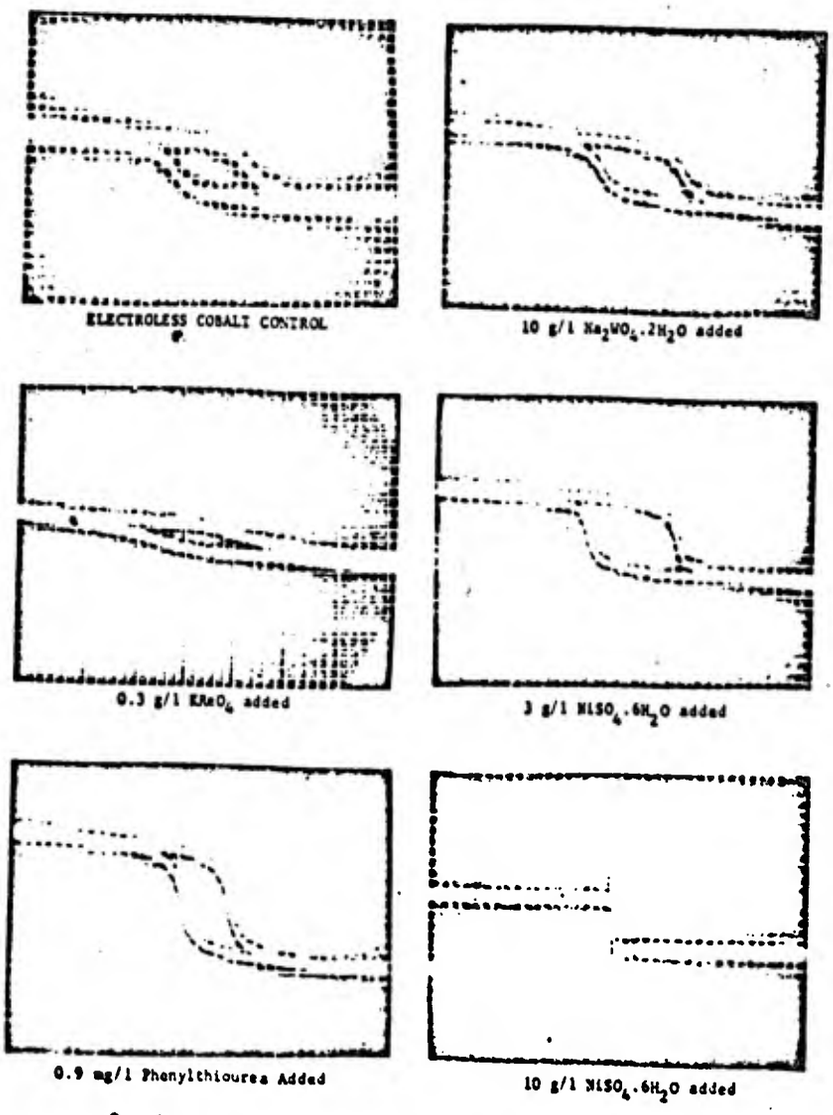


FIGURE 6. Effect of metal salt additions to electroless cobalt solution on the hysteresis loops of deposits (11,000 to 13,000 Å thick).

IN VITRO GROWTH OF P. KNOWLESI: A MODEL TO STUDY METABOLISM OF MALARIAL PARASITES AND ACTION OF ANTIMALARIAL DRUGS

HERMAN POLET, M.D.
WALTER REED ARMY INSTITUTE OF RESEARCH
WASHINGTON, D. C.

Malaria can cause greater loss of soldiers than casualties during military operations in areas where the disease is endemic (11). The incidence of malaria and man days lost can be greatly reduced by the administration of effective drugs. Recently, however, certain strains of malarial parasites have been found in Southeast Asia which are resistant to all known antimalarial drugs (10). It is the emergence of malarial strains resistant to the known drugs and the necessity to find new drugs effective against the resistant malarias that are the basis of the present Army Malaria Program. As part of this effort we have developed a method (8) to grow in test tubes a malarial parasite of monkeys, P. knowlesi, for one life cycle. This parasite is infective to man (2) and causes a disease similar to human malaria. This culture method has proven to be valuable in studying the metabolism of malarial parasites; such investigation may give a lead to designing new antimalarial compounds. The method further provides a model to test new potential antimalarials and to study the mode of action of antimalarial drugs.

MATERIALS AND METHODS

Parasites. Rhesus monkeys (Macaca mulatta), weighing 4-6 lbs were inoculated intravenously or intraperitoneally with erythrocytic forms of P. knowlesi. Monkey blood, of which 7-25% of the erythrocytes were parasitized with ring forms, was used to prepare cultures. During the course of the 24 hour in vivo schizogonic cycle, it was often observed that the parasitemia decreased, sometimes as much as 10%. This decrease was most marked during the first five hours of the cycle. Therefore, parasites used for biochemical and pharmacological studies were harvested from the monkey after the first 4 hours of the cycle had elapsed. By that time parasites had developed to a stage we called "middle aged" ring forms. Blood was drawn from the femoral vein in a syringe containing 100-200 units of heparin. Erythrocytes were separated from the serum by centrifugation

for 10 minutes at 700 x G at 4° C. Subsequently, the percentage of parasitized erythrocytes was adjusted to 7-8%, if necessary, by adding appropriate quantities of normal fresh monkey erythrocytes to the parasitized cells. Blood with more than 25-30% parasitized cells and subsequently diluted to 7% or less yielded poor in vitro growth. Finally, 0.02 ml of packed cells were inoculated into each culture tube containing 11 ml of medium. Thus, each culture contained about 16-18 million parasitized erythrocytes. Inoculation of the cultures can be facilitated by preparing a concentrated cell suspension (1.5 ml packed cells in 8.5 ml of medium) and dispensing this suspension into culture tubes with a 0.2 ml automatic pipet.* Gentle shaking kept the cell suspension homogenous, insuring an equal distribution of parasitized cells into cultures during pipetting. Less than one hour was allowed to elapse between harvesting the parasites from the monkeys and the onset of in vitro indubation. During the interval, parasitized cells were cooled in an ice bath.

Nutrient medium. Eagle's basal medium (3) with a 0.2 mM concentration of each amino acid was supplemented with 0.2mM of the following amino acids: glycine, l-alanine, l-glutamic acid, l-aspartic acid, l-proline, and l-serine.** Ten percent fresh human AB serum was added to the medium and the pH adjusted to 7.4 - 7.5 with a $NaHCO_3$ solution. Polypropylene tubes of 12 ml content+ were filled with 11 ml of nutrient medium and subsequently inoculated with parasitized erythrocytes, as described above. Then, the tubes were rotated in a roller at a speed of 1/5 rpm at 37° ± 1° C. Continuous rotation of the culture tubes keeps the cells continuously suspended, a condition necessary for optimal parasitic growth. Occasional AB sera agglutinated monkey erythrocytes, causing a slight inhibition of the growth rate.

Experiments on the extracellular amino acid requirements of P. knowlesi. Thirteen lots of Eagle's basal medium (3) were made, differing from each other only in that one of the following thirteen amino acids was missing: l-cystine, l-arginine, l-tryptophane, l-histidine, l-leucine, l-isoleucine, l-lysine, l-methionine, l-glutamine, l-tyrosine, l-threonine, l-valine, l-phenylalanine. A fourteenth medium containing all of the mentioned amino acids served as a control. All media contained 0.2mM of each of the following amino acids: l-alanine, l-aspartic acid, l-glutamic acid, l-proline, l-serine, glycine. The media were supplemented with 10 per cent fresh human AB serum, dialysed against saline until free from amino acids as tested by complete disappearance of l-leucine-C^{14}, l-isoleucine-C^{14} or l-histidine-C^{14} added to the serum prior to dialysis.

---

*Baltimore Biological Lab., Baltimore, Md., Cat. No. 05-719

**Nutritional Biochemical Corp., Cleveland, Ohio

+Federal Scientific Co., Kensington, Md., Cat. No. 3111 for tubes and 3110 for closures.

Determination of DNA. In a second set of three experiments 1 uc orotic-6-C^{14} acid* hydrate (specific activity 3.88 mc/mM) was added to each culture at the onset of incubation and DNA was extracted at the end of incubation by the method of Smith, Tannhauser and Schneider as outlined by Hutchinson and Munro (5):

Cells were dissolved in a solution of 0.05% sodium lauryl sulfate in water at 3° C., 0.1 mg. carrier DNA (sperm) per culture was added and trichloroacetic acid (TCA) was added to a final concentration of 10 per cent. The cell fragments were centrifuged to a pellet at 46,000 x G for 30 minutes at 3° C. The pellet was washed twice with 10 per cent TCA, twice with ethanol, twice with a 3:1 mixture of ethanol and ether. RNA was extracted with 0.3 N KOH at 37° C. for one hour followed by precipitation of the rest of the cells in 5 per cent TCA at 46,000 x G for 30 minutes at 3° C. The pellet was washed once with 5 per cent TCA. All solutions used had a temperature of 3° C. or less. The pellet was dissolved in 0.1N NaOH, plated in duplicate, dried and counted to at least 1000 counts in a low background gas flow counter.**

RESULTS

Cultivation. Parasitized blood from 54 monkeys has been used to prepare a total of over 5,000 cultures. These cultures were used in experiments designed for various purposes. Ring forms from all but six monkeys developed to immature schizonts and segmenters after incubation for 20-24 hours at 37° C. Results of 10 experiments, chosen at random, are shown in Table 1. Plasmodial growth was synchronous: at intervals during incubation most of the parasites were observed to be in the same developmental stage. This is similar to *in vivo* growth of *P. knowlesi*. This synchrony was not absolute; sporulation of segmenters with reinvasion of red blood cells by daughter parasites lasted *in vivo* about 1-3 hours, when the parasitemia was 7-25%, so that some overlapping among different developmental stages was observed. A similar degree of asynchrony occurred in the *in vitro* cycle. The pH of the growth medium shifted only 0.2 - 0.3 units during a 20-24 hour incubation period, provided the inoculum did not exceed 16-18 million parasitized cells. This was important because parasitized red cells are readily and visibly injured in an acid medium, resulting in a marked inhibition of the parasitic growth rate.

Extracellular amino acid requirements. After ring forms or trophozoites had been incubated for 18-20 hours, immature schizonts and segmenters appeared in all cultures except in cultures where l-isoleucine had been deleted from the growth medium. In the latter cultures all of the parasites failed to grow beyond the ring or trophozoite stage in all experiments performed.

---

*New England Nuclear Corp., Boston, Mass.

**Model 4342, Nuclear Chicago Corp., Des Plaines, Ill.

Using the number of immature schizonts and segmenters present at the end of incubation as a parameter of growth in the cultures missing one of the other amino acids tested, it was found that the results varied from one experiment to the other. Therefore, the amount of orotic-6-C^{14} acid incorporated into plasmodial DNA during an 18 hour incubation period was used as a measure of growth. The results, illustrated in Table 2, show that orotic-6-C^{14} acid incorporation into DNA during an 18 hour incubation period was significantly inhibited in the absence of l-isoleucine, l-methionine, l-cystine, l-tyrosine, l-arginine, l-glutamine, l-histidine or l-lysine; this inhibition was particularly marked in the absence of the first three amino acids. Orotic-6-C^{14} acid incorporation into DNA in the absence of l-threonine, l-phenylalanine, l-leucine, l-tryptophane and l-valine was not significantly different from the control.

Initial experiments had shown that l-alanine, l-aspartic acid, l-glutamic acid, l-proline, l-serine and glycine had no effect on parasitic growth.

DISCUSSION

Cultivation of malarial plasmodia. The methods of cultivation of erythrocytic forms of malarial plasmodia of Trager (13,14) and Geiman (1,4) have been known for over 20 years. They have the advantage that plasmodia can be grown in vitro for several growth cycles and have been extensively used to investigate nutritional and biochemical characteristics of plasmodia and to study the influence of antimalarial drugs on plasmodial growth. The unique advantage of the new method described here is that a large number of cultures (several hundreds) can be prepared from one monkey by a simple and rapid procedure. This advantage provides the opportunity to collect a great number of observations from a single well planned experiment.

The method offers a reproducible model to study the plasmodial schizogony of a primate malaria. Consistent results have been obtained in studies of the metabolism of P. knowlesi. Initial results have shown that the method is suitable for the testing of new compounds for possible antimalarial activity and for the study of the mode of action of antimalarial drugs.

Extracellular amino acid requirements for in vitro growth of P. knowlesi. The failure of ring forms or trophozoites to develop in the absence of l-isoleucine indicates that this amino acid is essential for growth of erythrocytic forms of P. knowlesi (9).

In the absence of l-methionine, growth of P. knowlesi is reduced to 54 per cent; this is in accord with the findings of McKee and Geiman, who showed that P. knowlesi requires an extracellular source of methionine for in vitro (6) as well as in vivo (7) growth.

Significant inhibition of growth of P. knowlesi in the absence of l-cystine, l-tyrosine, l-arginine, l-glutamine, l-histidine or l-lysine suggests that P. knowlesi requires an extracellular source of these amino acids for optimal in vitro development during one schizogonic cycle.

Finally, when the degree of inhibition of DNA synthesis in the absence of the amino acids which had a significant effect on parasitic growth is correlated with the molar amounts of the same amino acids in adult human hemoglobin (15), then it can be seen that inhibition of DNA synthesis is most marked in the absence of these amino acids of which hemoglobin contains the least (Table 2). The amino acid composition of human rather than monkey hemoglobin is compared because of its availability. The composition of monkey hemoglobin is believed to be similar (16). The observation that l-isoleucine and l-methionine are essential for plasmodial growth suggests that analogs of the two amino acids might have antimalarial activity.

SUMMARY

A simple and rapid method for the preparation of large numbers of short term cultures of erythrocytic forms of P. knowlesi is described. This method has shown itself to be valuable in testing nutritional and biochemical aspects of P. knowlesi, in testing compounds for antimalarial activity and in studying the mode of action of antimalarial drugs.

l-isoleucine and l-methionine are essential for growth of erythrocytic forms of P. knowlesi.

An extraerythrocytic source of l-cystine, l-tyrosine, l-arginine, l-glutamine, l-histidine and l-lysine is required for optimal in vitro development of P. knowlesi during one schizogonic cycle.

REFERENCES

1. Ball, E. G., Anfinsen, C. B., Geiman, Q. M., McKee, R. W. and Ormsbee, R. A., 1945. In vitro growth and multiplication of the malaria parasite Plasmodium knowlesi. Science 101:542.

2. Chin, W., Contacos, P. G., Coatney, G. R., and Kimball, H. R., 1965. A naturally acquired quotidian type malaria in man transferable to monkeys. Science, 149:865.

3. Eagle, H., 1955. Nutrition needs of mammalian cells in tissue culture. Science, 122:501.

4. Geiman, Q. M., Anfinsen, C. B., McKee, R. W., Ormsbee, R. A. and Ball, E. G., 1946. Studies on malarial parasites. VII. Methods and techniques for cultivation. J. Exp. Med. 84:583.

5. Hutchinson, W. C., and Muncro, H. N., 1961. The determination of nucleic acids in biological materials. The Analyst, 86:768.

6. McKee, R. W., Geiman, Q. M., and Cobbey, T. S., 1947. Amino acids in the nutrition and metabolism of malarial parasites. Fed. Proc. 6:276.

7. McKee, R. W., and Geiman, Q. M., 1948. Methionine in the growth of the malaria parasite, Plasmodium knowlesi, Fed. Proc. 7:172.

8. Polet, H., 1966. In vitro cultivation of erythrocytic forms of Plasmodium knowlesi and Plasmodium berghei. J. Mil. Med. 131: 1026.

9. Polet, H., and Conrad, M. E., 1968. Malaria: extracellular amino acid requirements for in vitro growth of erythrocytic forms of P. knowlesi. Proc. Soc. Exp. Biol. Med., 127:251.

10. Powell, R. D., 1966. The chemotherapy of malaria. Clin. Pharmacol. Therap., 7:48.

11. Sadun, E. H., 1964. International Panel Workshop on cultivation of plasmodia and immunology of malaria. Introduction. Am. J. Trop. Med. Hyg. 13:145.

12. Snedecor, G. W., Ed., 1956. Statistical Methods. Iowa State College Press, Ames, Iowa, p. 251.

13. Trager, W., 1947. The development of the malaria parasite, Plasmodium lophurae in red blood cell suspension in vitro, J. Parasit. 33:345.

14. Trager, W., 1950. Studies on the extracellular cultivation of an intracellular parasite (avian malaria). J. Exp. Med. 92:349.

15. Wintrobe, M. M., 1961. Clinical Hematology, Lea & Febiger, Philadelphia, Pa., p158.

16. Zuckerkandl, E., Jones, R. T., Pauling, L., 1960. A comparison of animal hemoglobins by tryptic peptide pattern analysis. Proc. Soc. Nat. Acad. Sc., 46, 1349.

Table I

Percentage of different developmental stages of erythrocytic forms of *P. knowlesi* after 20 hours incubation of middle aged ring forms.

	Trophozoites	Immature Schizonts	Segmenters
Experiment 1	6	81	13
2	6	61	33
3	4	88	8
4	20	77	3
5	18	80	2
6	13	83	4
7	6	74	20
8	20	80	
9	11	87	2
10	13	60	27

Note: Degenerated parasites are not included in the table and amount to an average of 4%.

Table II

Comparison of the growth rate of P. knowlesi in culture media depleted of various amino acids and the number of moles of these amino acids in adult hemoglobin.+

Amino Acid Depleted from Culture Media	Orotic-6-C^{14} Acid Incorporated into DNA of P. knowlesi	No. of Moles of Amino Acid per Human Hemoglobin (15)
Control	100 ± 13	
Lysine*	88 ± 11	44
Histidine*	87 ± 5	38
Glutamine*	87 ± 7	30
Arginine*	84 ± 8	12
Tyrosine*	83 ± 8	12
Cystine**	72 ± 13	6
Methionine**	54 ± 10	6
Isoleucine	9 ± 4	0

+Data are expressed as percentage of the control ± SD. Values are the average of three experiments with duplicate analyses. One asterisk denotes a value which is significantly different from the control value at the 5 percent level. The double asterisk denotes values which are significantly different from the control value at the 1% level. Comparison among the means was performed by a modification of the method of J. W. Tukey (12). Net counts of radioactivity were used for statistical comparison. The amino acid composition of adult human hemoglobin is given because of its availability. The composition of monkey hemoglobin is believed to be similar(16).

A NEW APPROACH TO PROTECTION AGAINST INTENSE LIGHT ENERGY

R.A. Prosser and J.T. Stapler
U.S. Army Natick Laboratories
Natick, Massachusetts 01760

INTRODUCTION Approximately one-third of the energy of a thermonuclear explosion is emitted in one minute or less in the form of light and heat rays. This electromagnetic energy is capable of causing skin burns and starting fires at considerable distances from the point of burst; hence, it is referred to as "thermal radiation". For a one megaton thermonuclear weapon, for example, the radii for first and second degree burns are about 15 and 11 miles, respectively, and the radius in which newspapers will ignite is about 9 miles (1).

This report is concerned with showing that certain polymers automatically generate a significant measure of thermal protection on exposure to intense thermal radiation by the following mechanism: 1. the polymer decomposes yielding a gas and compounds containing chromophoric groups; 2. the gas propels the compounds containing chromophoric groups into the region between the polymer and the source of the radiant energy; 3. the compounds containing chromophoric groups vaporize, intercept the thermal photons, and reradiate or dissipate the energy to the air as heat. This constitutes the new approach to protection against intense light energy.

THERMAL BEHAVIOR OF POLYVINYLIDENE FLUORIDE In the search for means of protection from the intense thermal energy of a thermonuclear explosion, it was established by Barnes and Yelland (2) that of the polymers listed in Table I, polyvinylidene fluoride, PVF_2, provided the best protection. This was determined by preparing 10 discs of each polymer, each containing 5 percent of a colored additive to make them all opaque, and each 0.060 inches thick. These discs were then exposed to a single-elliptical, carbon arc-image furnace at an irradiance level of 23 cal/cm^2 for one second. The temperature

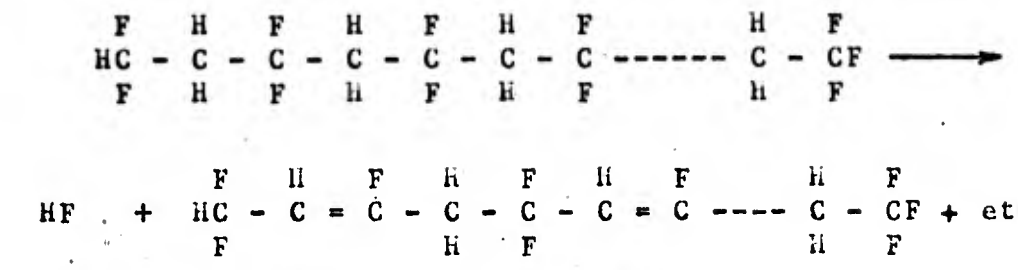
rise, ΔT (in degrees C), above ambient of a thermocouple placed on the rear surface of the disc was measured, the ten values were averaged, and the frequency of ignition of the decomposition products computed (2). This information is listed in Table I together with the polymer abbreviations to be used.

TABLE I

THE EFFECT OF HIGH-INTENSITY THERMAL RADIATION ON VARIOUS VINYL POLYMERS

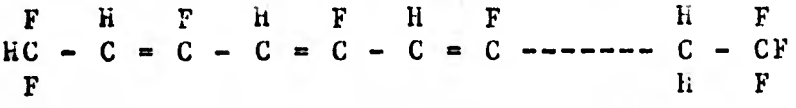
Polymer	Repeating Unit	Avg. ΔT (°C)	Frequency of Ignition (%)
Polyvinylchloride PVC	$-\mathrm{C}(\mathrm{H})(\mathrm{H})-\mathrm{C}(\mathrm{H})(\mathrm{Cl})-$	22.3	100
Polychlorotrifluoro-ethylene, Kel-F	$-\mathrm{C}(\mathrm{F})(\mathrm{F})-\mathrm{C}(\mathrm{F})(\mathrm{Cl})-$	13.0	0
Polyvinyl Fluoride PVF	$-\mathrm{C}(\mathrm{H})(\mathrm{H})-\mathrm{C}(\mathrm{H})(\mathrm{F})-$	16.8	100
Polyethylene PE	$-\mathrm{C}(\mathrm{H})(\mathrm{H})-\mathrm{C}(\mathrm{H})(\mathrm{H})-$	19.2	100
Polytrifluoro-ethylene, PVF_3	$-\mathrm{C}(\mathrm{F})(\mathrm{F})-\mathrm{C}(\mathrm{H})(\mathrm{F})-$	11.3	0
Polyvinylidene Fluoride, PVF_2	$-\mathrm{C}(\mathrm{H})(\mathrm{H})-\mathrm{C}(\mathrm{F})(\mathrm{F})-$	10.8	50
Teflon	$-\mathrm{C}(\mathrm{F})(\mathrm{F})-\mathrm{C}(\mathrm{F})(\mathrm{F})-$	36.3	0

When pyrolyzed PVF_2 yields copious quantities of HF and a yellow to amber wax. A representative reaction is



gas yellow to amber, waxy fraction

Now, since the allylic position is relatively weak, under the intense heat more HF should split off so that the above reaction will proceed readily and some of the molecules with two double bonds should lose HF to produce



which is a molecule with three conjugated double bonds. This process can continue producing conjugation of longer wavelength anywhere in the degraded polymer chains (yellow wax).

Those molecules containing conjugated unsaturation together with the vast majority of molecules which undoubtedly do not, are propelled into the region between the polymer disc and the carbon arc-image furnace by the HF. Here the molecules with conjugated unsaturation intercept those light photons which lie within their absorption spectrum, go to a higher energy level, dissipate the energy via successive vibrational and rotational states while returning to the ground state, pick up a second light photon and continuously repeat the cycle provided they do not degrade or ignite. The half-life of the cycle will "generally vary down to 10^{-8} to 10^{-9} seconds" (3). Many chromophores consist of linear, conjugated unsaturation, called polyenes, and are quite efficient in intercepting thermal photons. As an example, consider the cyanogen bands of the carbon arc-image furnace (and peculiar thereto) at about 360, 390, and 420 mμ. The combined bands at 360 and 390 mμ are 40-50 percent more intense than any other span of equal width in the carbon arc emission spectrum. The molecule, $H-(CH=CH)_6-H$, a polyene, has a λ max at 364mu and its absorption spectrum overlaps the two lower cyanogen bands. It has a molar absorbtivity (e) of 138,000 (4). Assuming that $-(CH=CF)_6-$, a structure which could occur in the yellow, waxy fraction, also has these properties, the amount of material required to reduce the light intensity to 1/10,000 of its initial

value (an absorbance or optical density of 4) can be computed using the Lambert-Beer Law, $\log_{10}(Io/I) = ecb$ = absorbance = optical density, where Io is the initial light intensity, I is the intensity of transmitted light, e is the molar absorbtivity, c is the concentration in moles per liter, and b (= 1 cm) is the path length through the sample in cm.

$$\log_{10}(10{,}000/1) = 138{,}000 \times c \times 1$$
$$c = 2.9 \times 10^{-5} \text{ moles/liter}$$
$$c = 2.9 \times 10^{-8} \text{ moles/cc}$$

Since the molecular weight of $H - (CH = CF)_6 - F$ is 284 grams, the weight of this material required to block 99.99 percent of the light in the 364 mμ region and thereby protect 1.0 cm^2 of the polymer disc from light in that region is

$$2.9 \times 10^{-8} \times 284 = 8.22 \times 10^{-6} \text{ grams}$$
$$= 8.22 \text{ micrograms.}$$

Since the weight of the yellow, waxy fraction produced during a one second exposure to the arc is about 26 mg (5), it is entirely possible for such a small quantity of polyene linkages to arise, and even more probable that shorter polyene segments also will be produced and in greater amounts. Consequently, once the waxy fraction appears in the region between the PVF_2 disc and the light source, practically all of the energy below about 400 mμ and about 1/3 of the light energy above 400 mμ is intercepted and reradiated or dissipated to the air. Curves showing the carbon arc emission spectrum and the percent transmission of a 5 percent solution of the waxy fraction in spectroscope grade p-dioxane are shown in Fig. 1.

From Fig. 1. it is apparent that the energy of the combined cyanogen bands at 360 and 390 mμ is less than 10% of the energy emitted by the arc, or, for one second, less than 2.3 cal. The amount of energy penetrating the 2, 4, 6, 8, 10, 12, 12 heptafluorododeca- 1, 3, 5, 7, 9, 11 hexaene, $H-(CH=CF)_6-F$, is therefore less than 0.00023 cal. On a bright day the intensity of sunlight is about 0.02 cal/cm^2 sec.

EXPERIMENTAL RESULTS There is considerable evidence for the presence of unsaturation in the pyrolysis products of PVF_2. Mass spectroscopy of the gaseous pyrolysis products of PVF_2 indicated the presence of

$$CH_2F - CH = CF - CH = CH - CF = CH - CF = CH_2$$

Note the presence of four conjugated double bonds. Also, practically all of the gaseous decomposition products contained unsaturation (9).

Examination of the yellow, waxy fraction showed:

1. Assuming the formation of one double bond for every 10 carbon atoms, the elemental analyses calculated and found are:

	C%	F%	H%
Calculated	40.0	57.0	3.0
Found	39.5	57.5	3.02
Found	40.7	56.1	3.17

Since there is about one multiple bond for 10 carbon atoms, there is a good chance for a small amount of conjugation.

2. IR absorption in the 1590 and 1680 cm^{-1} region which is characteristic of conjugated unsaturation (11) is strong.

3. UV transmission, Fig. 1. shows less than 1 percent transmission from 200 to 400 mu and then the transmission rises rather rapidly. Polyenes (hydrocarbon) show intense absorption up to 370 mu and then the absorption falls off rapidly (12). Thus the behavior of the amber waxy fraction in the UV conforms to what would be expected.

4. As stated earlier, as the number of conjugated double bonds in the polyene decreases the relative amounts of that polyene should increase, ie., there should be more polyenes containing 3 double bonds than there are containing 6 double bonds. Hence, the UV absorbance curve should peak where butadiene peaks, 217 mu (13), and it does.

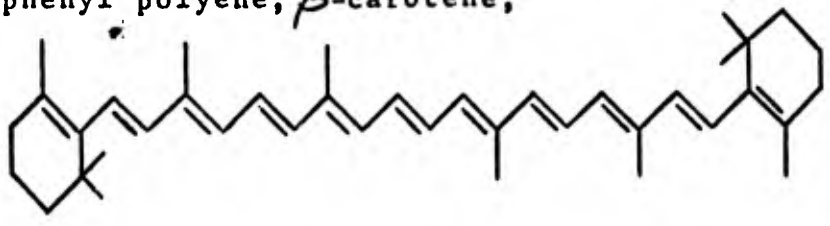
5. Since the wax is yellow it must absorb blue light, and, consequently, chromophores, which could contain conjugated unsaturation, must be present.

6. On standing for a period of time the wax hardens, just as paint hardens or "dries" when exposed to air. However, since paint "dries" by oxidation at double bonds, this implies the presence of double bonds in the waxy fraction.

For a given concentration of the polyene groups there is some thickness of the waxy fraction at which 99.9 percent of the thermal energy that lies within the absorption spectra of the polyene groups is intercepted.

Increasing the thickness so that absorption in the same region is 99.999 percent will not change the amount of light transmitted very much. The point is that above some thickness the light transmitted will be substantially independent of thickness. The approximate amount of light that should be transmitted by a "thick" layer of the amber, waxy fraction can be estimated from the curves of Fig. 1. At each point on the abscissa the product of the percent transmission and emission at that point on the abscissa can be plotted. The area under the resulting curve is a measure of the energy transmitted and amounts to about 50%. Since the concentrations of the various polyenes in the amber, waxy fraction is unknown, one cannot be any more specific. In any case, if one is to accept that the polyenes in the amber, waxy fraction are responsible for most of the thermal protection afforded, a "thick" layer of the fraction should intercept a reasonable portion of the incident light based on the known behavior of polyenes (see para. 3, Experimental Results, above). When tested, using a thickness which should provide absorption above the 99.9 percent level (approximately 0.040 inches), 55 percent of the light from the arc was absorbed. Thus the yellow, waxy fraction absorbs a significant portion of the thermal energy and conjugated unsaturation is therefore a major factor in the thermal behavior of PVF_2.

As a check of this concept a saturated solution of a diphenyl polyene, β-carotene,



in a polyacrylate containing a solvent was prepared. The solvent was evaporated leaving an amber disc about 1/16" thick and 2" in diameter containing about 2.4 mg solute. β- carotene absorbs strongly from about 390 to 500 mμ (14), and weakly at bordering wavelengths. The wavelength at which maximum absorption occurs (λmax.) is 460 mμ, and the molar absorbtivity (e max.) at the wavelength of maximum absorption is 150,000. From Fig. 1, one would estimate that less than 50% of the light should be absorbed. On testing, using a clear control, about 40% of the light of the carbon arc was absorbed. There is then no question but that a polyene, protected from oxidation and dispersed as molecules, will absorb a significant fraction of the light energy of the arc.

DISCUSSION For PVF_2, the manner in which polyenes can be generated, their existence in the decomposition products, and their presence there in an amount sufficient to absorb a significant fraction of the incident light energy has now been established. It will now be shown, where applicable, that the energy attenuation by chromophoric groups (primarily polyenes) operating according to the mechanism postulated is the controlling factor determining the relative ΔT values occurring in Table I.

Since PVF_2 produces the most HF and, therefore, probably the greatest quantity of molecules containing conjugated unsaturation of all the hydrofluoropolymers, and of course PE and Teflon, and if such molecules are a major factor in energy dissipation, then PVF_2 should have the lowest ΔT value and it does. The fact that the ΔT value was the lowest despite the fact that the frequency of ignition was 50%, strongly indicates that the conjugated sequences survived the flames.

Turning now to PVF_3, it is apparent that HF can also split off in the same manner as for PVF_2. Consequently PVF_3 should have a low ΔT and it does. After removal of HF the conjugated sequences consist solely of carbon and fluorine. These should not ignite; the frequency of ignition should be low, and it is.

When HF splits off from PVF, unsaturated hydrocarbon segments are formed which should ignite readily destroying the conjugation, and generate a lot of smoke. Consequently the ΔT for PVF should be above that for PVF_2 and PVF_3, the frequency of ignition should be high, and a lot of smoke should be produced, all of which occur.

Thermal protective creams containing halogenated chemicals compounded at the U.S. Army Natick Laboratories have shown excellent thermal protection when exposed to the highest intensity of the electric arc for periods up to 6 seconds (15). Copious quantities of decomposition products were generated which expanded towards the light source. It was felt at the time that there might be some connection between thermal protection and these gaseous decomposition products (15). One of the ingredients of these creams is polychlorotrifluoroethylene, Kel-F (20). On pyrolysis, Kel-F, yields the same yellow-amber, waxy fraction (6) as does PVF_2, PVF_3 and PVF (7) which indicates the presence of chromophoric groups and a gas, monomeric trifluorochloroethylene (6). Since fluorine decolorizes the waxy fraction (6) the chromophoric groups

probably consist of conjugated unsaturation. With a zero frequency of ignition, the ΔT value should be below that of PVF and it is.

Since PE, on pyrolysis, would not yield significant amounts of conjugated unsaturation compared to PVF_2 and PVF_3 its ΔT value should be higher and it is. In this case the usual energy sinks such as endothermic decomposition, loss of high temperature material, etc. are the controlling factors.

It is known that PVC loses HCl very readily on pyrolysis (much more readily than PVF_2 loses HF). The amount of conjugated unsaturation produced should be much larger (8) and its ΔT value the very lowest of the group. There is a fundamental difference between the pyrolysis behavior of the two polymers, however. In the case of PVC the HCl comes off so readily (9) that relatively few carbon-carbon bonds are broken. Thus most of the conjugated sequences remain on the surface of the disc absorbing the thermal energy very efficiently which results in a high ΔT value. Here again the conjugated sequences are hydrocarbons so even if any are ejected in front of the disc they will provide little protection because they will be destroyed by the flames as in the case of PVF.

The surface of a Teflon disc after arc pyrolysis is brown to black. Chromophores are therefore present on the surface of the disc and the energy of the arc will be absorbed efficiently. Since the extent of decomposition is quite low, energy attenuation or dissipation by endothermic decomposition etc. is not available. The main factor here is probably reflection (only 9 percent because of the colored additive (5)) until the surface discolors. The ΔT value should be high and it is.

Since the halogen gases are colored in those cases where they are formed they will contribute to thermal protection (a low ΔT) just as the polyenes do.

Other channels which the incident thermal energy might take are: 1. reflection, 2. absorption causing a rise in temperature, 3. endothermic decomposition, 4. a change in phase, 5. loss of high temperature material, 6. conduction to the surroundings, 7. re-radiation, and 8. scattering by particulate matter (smoke). Although these various channels can and do serve as energy sinks for a considerable portion of the incident energy, none,

when considered individually, could explain the data of Table I. There is no doubt that energy attenuation by the mechanism described is the dominant factor in these experiments.

It is interesting to note that although blisters were raised on the surface of the plastic containing the β carotene by the energy absorbed, no char was formed. This indicates that thermal protection is not necessarily a function of char formation. This conclusion is supported by the fact that PVC, PE, Kel-F, PVF, PVF_2, and PVF_3 all show charred surfaces after pyrolysis on the carbon arc-ima-ge furnace but widely different ΔT values. The temperature reached by the material which actually absorbs the light photons is probably so high that the composition of this material for each polymer is practically the same, 90% carbon or better. Since char formation probably takes place in a fraction of a second in all cases, if it were a dominating factor in these experiments the behavior toward the impinging light energy should be substantially the same resulting in ΔT values which are close together. This is not the case. There are two other polymers which show extensive char formation on pyrolysis. These are the phenolic and the epoxy polymers. The phenolics yield xanthene (16), which on oxidation gives xanthone (17). Xanthone is stable up to 860 degrees centigrade and when it decomposes dibenzofuran results (18). All three compounds absorb strongly in the UV and dibenzofuran fluoresces (19). An epoxy was pyrolyzed at the U.S. Army Natick Laboratories and its UV spectrum is very close to the one for PVF_2 shown in Fig. 1. Hence, in addition to the char both polymers yield compounds containing chromophoric groups. Both polymers have been used in nose cones and motors of rockets. Also, it appears that even when the life of the motor is exhausted, the char remains. It is suggested that one of the reasons for the good performance of these polymers is that the decomposition products contain chromophoric groups which act as a "heat shield".

Polyenes transmit much of the light of wavelength greater than 500 mμ. To block this energy and simultaneously reduce the system to its essentials one therefore needs a carrier and several pure compounds containing chromophoric groups which will together efficiently absorb thermal radiation of injurious wavelengths. When intense thermal radiation strikes this system it will be absorbed by the compounds containing chromophoric groups vaporizing them and decomposing the carrier mainly

into gases. The gases will propel the compounds into the region between the target and light source where they will absorb energy as described earlier.

A rough estimate of the amount of material required per square centimeter for actual use can be computed. Earlier it was calculated that 8.22×10^{-6} gms/cm^2 = 2.42×10^{-3} oz/yd^2 was required to block out the cyanogen bands. Allowing for a carrier and compounds containing chromophoric groups which will absorb outside of the cyanogen bands, the amount required for substantial protection should not exceed 0.2 oz/yd^2 (2 oz/yd^2 is considered a light coating on fabrics). This small weight of material required constitutes an important advantage of this approach. In fact on a protection afforded per gram of material ratio basis, it is difficult to even conceive of an alternate system that is better.

CONCLUSIONS The thermal protective mechanism deduced from the pyrolysis behavior of the polymers tested appears to be correct, and, used intelligently, should lead to optimum protection against the intense light energy from a thermonuclear burst or any other source.

In general, any combination of compounds containing chromophoric groups or which will react to provide such groups, dissolved or dispersed in a carrier which will generate or provide a propelling gas when exposed to thermal evergy should act as a "heat shield", ie, provide a measure of thermal protection.

The degree of protection depends on the temperature of decomposition or vaporization of the carrier, the amount of propellant produced by the carrier (conversion should be close to 100% to avoid char formation which absorbs energy efficiently), the volatility of the compounds containing chromophoric groups, the thermal stability of these compounds, their concentration in the carrier (the higher the concentration the more heat generated in a given layer of carrier and the sooner the compounds containing chromophoric groups are released), and the extent to which the absorption spectrum of the dyes covers the range of injurious thermal radiation.

ACKNOWLEDGEMENT The authors wish to thank Dr. J. Bornstein, Dr. W.E.C. Yelland, Dr. W.J. Barnes, and Mr. D. Feldman for many stimulating discussions during this investigation.

BIBLIOGRAPHY

1. Glasstone, S., (edit), The Effects of Nuclear Weapons, U.S. Government Printing Office, Washington, D.C., (Revised 1964), page 333 and 572.

2. Barnes, W.J. and Yelland, W.E.C., Degradation of Halopolymers by Exposure to High-Intensity Thermal Pulses, Materials Research and Engineering Report #65-4, Clothing & Organic Materials Division, U.S. Army Natick Laboratories, Natick, Mass., 01760, (1965).

3. Jaffe, H.H. and Orchin, M., Theory and Application of Ultraviolet Spectroscopy, John Wiley and Sons, N.Y., (1962), p. 533.

4. Ibid, p. 228.

5. Barnes, W.V. and Yelland, W.E.C., Thermal Protective Materials Evaluation Report #14, Clothing & Organic Materials Division, U.S. Army Natick Laboratories, Natick, Mass. 01760, (6 Oct 66), p. 8.

6. U.S. Patent #2,543,530.

7. Madorsky, S.L., Hart, V.E., Straus, S., and Ledlak, V.A., J. of Res., NBS, 51, (6): 331, (1953).

8. U.S. Patents #2,754,210; #2,772,158; #2,772,159.

9. Madorsky, S.L., Thermal Degradation of Organic Polymers, Interscience, N.Y., (1964), p. 161.

10. Prosser, R.A., J. of Analytical Chemistry, 39: 1128, (1967).

11. Bellamy, L.J., The Infra-Red Spectra of Complex Molecules, J. Wiley, N.Y. (1962) p. 40.

12. Naylor, P. and Whiting, N.C., J. Chem. Soc., (1955) p. 3038.

13. Jaffe, H.H. and Orchin, M., op. cit., p. 228.

14. Ibid, p. 233.

15. Kapala. T. and Oesterling, J.F., Hands and Face Creams for Protection Against Thermal Radiation, (Confidential), Quartermaster Research and Development Command, Natick, Mass. 01760, (26 April 57).

16. Megson, N.J.L., J. Soc. Chem. Ind., 58: 131, (1939).

17. Elderfield, R.C., Heterocyclic Compounds, Vol. 2, (1951) p. 454.

18. Ibid p. 442.

19. Dictionary of Organic Compounds, Vol. 2, p. 894, Oxford University Press, N.Y. (1965).

20. Sperandio, G.J. and Jenkins G.L., Improvement of Physical Properties of Creams for the Protection of Normally Exposed Film to Thermal Radiation, Contract DA19-129-QM-1079; 1 Feb 68 - 31 July 69, Quartermaster Research and Engineering Command, Natick, Mass. 01760.

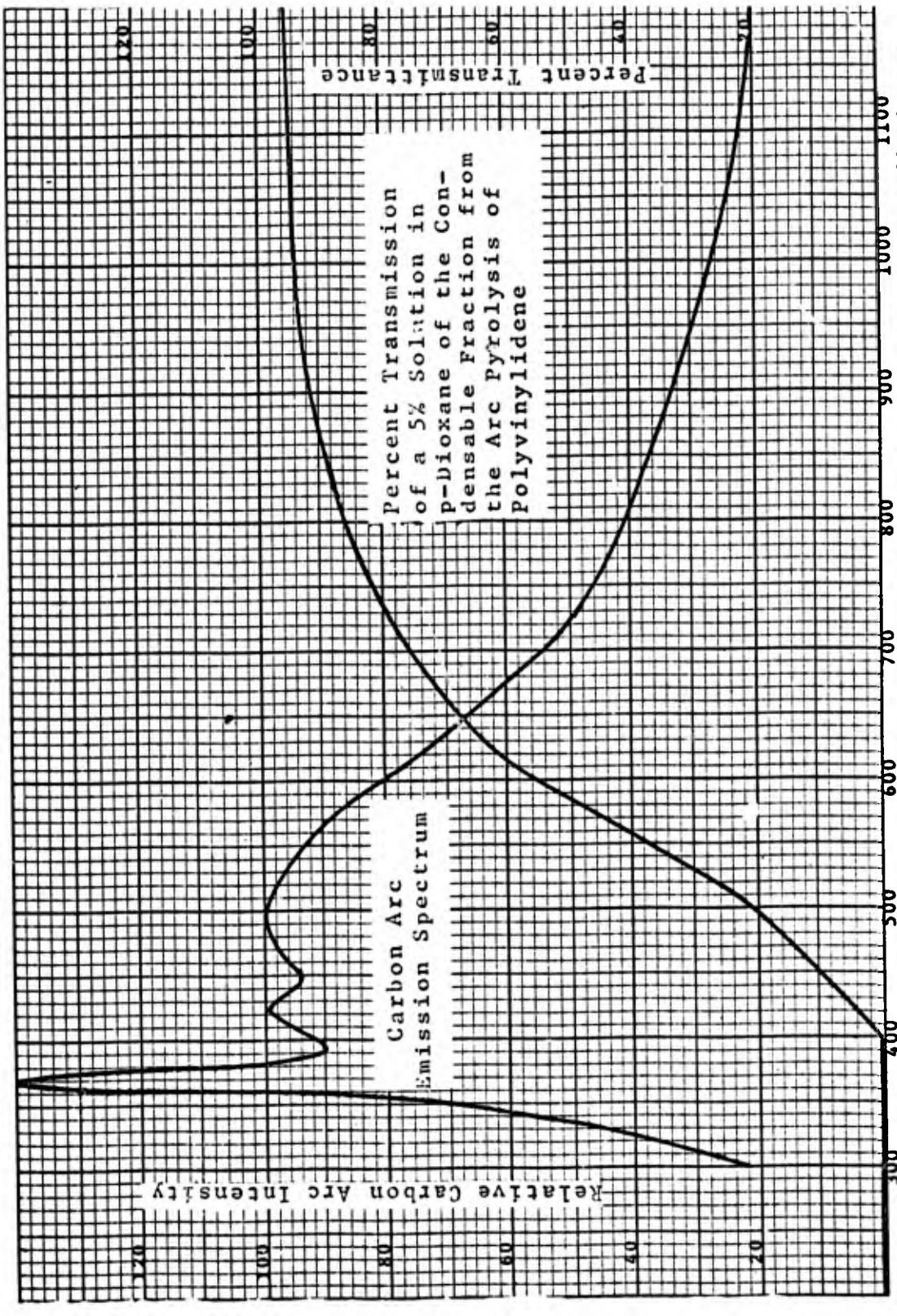


Fig. 1. Relative Carbon Arc Intensity and Percent Transmittance vs. Wavelength in Millimicrons

STRESS REVERSAL IN BONDING MATERIALS BETWEEN FIBERS

S. L. PU and M. A. SADOWSKY
U. S. ARMY WATERVLIET ARSENAL
WATERVLIET, NEW YORK

INTRODUCTION

Plane elastostatic analysis of an infinite plate containing a circular hole under uniform tension at infinity was first given by Kirsch (1). The solution for the case of a row of circular holes in an infinite plate was obtained by Sadowsky (2) and Howland (3). Details on stresses in a plate containing two holes were discussed by Ling (4) for holes of equal size and by Davies and Hoddinott (5) for holes of unequal size. The analogous problem for a circular inclusion was solved by Goodier (6). This paper is to discuss in some detail the stresses in an infinite plate containing two rigid circular inclusions of the same size. Extension to cases of elastic inclusions, of two inclusions of unequal size* and of an inclusion near the straight boundary of a semi-infinite plane under tension parallel to this boundary will be published elsewhere.

STATEMENT OF THE PROBLEM

Shown in Fig. 1 are two parallel identical microfibers. They will be treated as rigid bodies embedded in a matrix and bonded to it. The matrix is subject to one of the following two uniaxial stresses at infinity. Case 1, the longitudinal tension case:

$$\sigma_x = 1 \quad , \quad \sigma_y = 0 \tag{1}$$

Case 2, the transverse tension case:

$$\sigma_x = 0 \quad , \quad \sigma_y = 1 \tag{2}$$

*After acceptance of this paper, Goree published "In-Plane Loading in an Elastic Matrix Containing Two Cylindrical Inclusions," J. Comp. Mat'ls, Vol. 1, No. 4, 1967. He used the complex variable method.

The region showing peculiarities of stress distribution is that between the fibers. At a certain degree of proximity stress reversal takes place, contrary to superficial expectations. The conditions at the point x = y = 0 should suffice to represent the region. Consequently, the problem is to determine the stresses σ_x and σ_y at x = y = 0 in both cases.

THE BIPOLAR SYSTEM OF COORDINATES

The natural geometric pattern connected with Fig. 1 is that of bipolar coordinates. A general discussion of stress and strain in bipolar coordinates was given by Jeffery (7).

The orthogonal curvilinear bipolar coordinates (α, β) are defined by the conformal transformation

$$x + iy = i\,c\,\mathrm{ctn}\,\frac{\alpha + i\beta}{2} \tag{3}$$

in which c is a real positive constant and (x, y) are the cartesian coordinates of the point (α, β). Introducing the abbreviation

$$\mu = \cosh\beta - \cos\alpha \tag{4}$$

we may separate Eq. (3) into two real equations:

$$x = \frac{c\sinh\beta}{\mu} \quad , \quad y = \frac{c\sin\alpha}{\mu} \tag{5}$$

The cartesian equations of the parametric α- and β- curves, Fig. 2, are

$$(x - c\cdot\mathrm{ctnh}\,\beta_1)^2 + y^2 = \frac{c^2}{\sinh^2\beta_1} \tag{6}$$

$$x^2 + (y - c\cdot\mathrm{ctn}\,\alpha_1)^2 = \frac{c^2}{\sin^2\alpha_1} \tag{7}$$

Eq. (6) is the parametric curve $\beta = \beta_1$ which is a complete circle of radius $|c/\sinh\beta_1|$ with its center located at $x = c\cdot\mathrm{ctnh}\,\beta_1$, $y = 0$. For $\beta_1 = 0$ we obtain y-axis. For $\beta_1 = -\infty$ we obtain a point circle at O_1 while the point circle at O_2 corresponds to $\beta_1 = \infty$. From Eq. (4) and (5) follows that the sign of y is the same as the sign of $\sin\alpha_1$. Consequently, the parametric β-curve, $\alpha = \alpha_1$ given by Eq. (7) is not the entire circle but only an arc of it joining O_1 and O_2 . The radius of the circle is $|c/\sin\alpha_1|$ and its center is located at $x = 0$, $y = c\cdot\mathrm{ctn}\,\alpha_1$. The selection of the arc which constitutes the complete parametric β-curve depends on the value of α_1 , Fig. 2.

Let us use the following notations:

$$x_{,\alpha} = \frac{\partial x}{\partial \alpha} \quad , \quad x_{,\alpha\alpha} = \frac{\partial^2 x}{\partial \alpha^2} \quad , \quad y_{,\alpha\beta} = \frac{\partial^2 y}{\partial \alpha \partial \beta} \quad , \text{ etc.}$$

Eq. (5) yield

$$\left.\begin{aligned} x_{,\alpha} &= y_{,\beta} = -c\,\frac{\sin\alpha \sinh\beta}{\mu} \\ x_{,\beta} &= -y_{,\alpha} = c\,\frac{1-\cos\alpha\cosh\beta}{\mu} \end{aligned}\right\} \tag{8}$$

The metric scale coefficient h of the bipolar system is given by

$$\frac{1}{h^2} = x_{,\alpha}^2 + y_{,\alpha}^2 = \frac{\mu}{c} \tag{9}$$

The direction cosines of the bipolar pattern are

$$\left.\begin{aligned} \cos(\alpha, x) &= \cos(\beta, y) = -\sin\alpha \sinh\beta \,/\mu \\ \cos(\beta, x) &= -\cos(\alpha, y) = (1-\cos\alpha\cosh\beta)/\mu \end{aligned}\right\} \tag{10}$$

BIHARMONIC STRESS FUNCTIONS

The stress function $\phi = \phi(\alpha,\beta)$ in 2-dimensional elasticity problems is a biharmonic function, satisfying

$$\nabla^4 \phi = 0 \tag{11}$$

In the bipolar system it is convenient to use the funccion $\omega(\alpha,\beta)$ which is related to $\phi(\alpha,\beta)$ as follows:

$$\phi = \omega/\mu \tag{12}$$

Substituting Eq. (12) in the biharmonic equation (11) we have the fourth order partial differential equation:

$$\omega_{,\alpha\alpha\alpha\alpha} + 2\omega_{,\alpha\alpha\beta\beta} + \omega_{,\beta\beta\beta\beta} + 2(\omega_{,\alpha\alpha} - \omega_{,\beta\beta}) + \omega = 0 \tag{13}$$

It is obvious for the problem at hand that the stress distribution in the matrix is symmetric with regard to both the x-axis and the y-axis. It follows that both the stress function $\phi(\alpha,\beta)$ and the omega function $\omega(\alpha,\beta)$ are even functions both in α and β. By the method of separation of variables we are left with two types of ω-functions which are solutions of Eq. (13) even in α and β. Assigning catalog numbers $10n+2$ and $10n+4$ to them, we write

$$\omega_{10n+2} = \cosh(n+1)\beta \cos n\alpha \tag{14}$$

$$\omega_{10n+4} = \cosh(n-1)\beta \cos n\alpha \tag{15}$$

For reasons of intended Fourier analysis in α, the range of n is 0, 1, 2,...

It is an established experience that biharmonic analysis in orthogonal curvilinear coordinates leads to general formulas for biharmonic functions which are incomplete for the initial values n = 0 and n = 1. This requires filling of the gaps by specially designed replacement biharmonics which would naturally fit into the places of the failed functions.

In our case the two functions for n = 0 in Eq. (14) and (15)

$$\omega_2 = \cosh\beta \quad , \quad \omega_4 = \cosh\beta \tag{16}$$

are to be dropped. It is obvious to drop one of these two functions. Furthermore, we have

$$\phi_2 - \phi_{14} = \frac{\omega_2 - \omega_{14}}{\mu} = \frac{\cosh\beta - \cos\alpha}{\mu} = 1 \tag{17}$$

The stress function $\phi_2 - \phi_{14} = 1$ contributes nothing to the stresses and displacements. To keep $\omega_{14} = \cos\alpha$, we drop both ω_2 and ω_4.

Two biharmonics have to be formed to compensate the loss. We form

$$\phi = \log\mu \tag{18}$$

$$\phi = (1 - \cos\alpha\cosh\beta)/\mu^2 \tag{19}$$

The corresponding ω-functions are designated by

$$\omega_{4''} = \mu\log\mu \quad , \quad \omega'' = (1 - \cos\alpha\cosh\beta)/\mu \tag{20}$$

A general solution of Eq. (13), denoted by [S], is given by the following superposition of particular solutions

$$[S] = c''[S''] + d'[S_{4''}] + \sum_{n=1}^{\infty} c_n[S_{10n+2}] + \sum_{n=1}^{\infty} d_n[S_{10n+4}] \tag{21}$$

where [S''], [$S_{4''}$], ... denote solutions generated by ω-functions ω'', $\omega_{4''}$, ..., and c'', d', ... are superposition coefficients to be determined from the boundary conditions.

BOUNDARY CONDITIONS

On account of the assumed rigidity of the microfibers and due to symmetries of shape and loading, the boundary conditions are:

1. Conditions on $\beta = \beta_0 > 0$ for all values of α :

$$u_x = u_1 \quad , \quad u_y = 0 \qquad \text{for case 1} \tag{22}$$

$$u_x = u_2 \quad , \quad u_y = 0 \qquad \text{for case 2} \tag{23}$$

2. Regularity conditions at infinity:

$$\sigma_x = 1 \ , \ \sigma_y = \tau_{xy} = 0 \qquad \text{for case 1} \tag{24}$$

$$\sigma_y = 1 \ , \ \sigma_x = \tau_{xy} = 0 \qquad \text{for case 2} \tag{25}$$

Values u_1 and u_2 are still unknown, the superposition coefficients in Eq. (21) would be found in terms of u_1 , in case 1 and u_2 in case 2. To avoid this inconvenience in computation, we may change the original problems to replacement problems. The replacement problems are formulated as follows:

1. Boundary conditions on $\beta = \beta_0 > 0$ for all values of α :

$$4Gc\,u_x = 1 \ , \ u_y = 0 \qquad \text{for cases 1 and 2} \tag{26}$$

2. Regularity conditions at infinity:

$$\sigma_x = \sigma_1 \ , \ \sigma_y = \tau_{xy} = 0 \qquad \text{for case 1} \tag{27}$$

$$\sigma_y = \sigma_2 \ , \ \sigma_x = \tau_{xy} = 0 \qquad \text{for case 2} \tag{28}$$

where σ_1 and σ_2 , stresses at infinity, are not prescribed and are subject to computation. The solutions of the original problems, denoted by $[S_1]$ and $[S_2]$ for case 1 and case 2 respectively, are found by dividing the solutions of the replacement problems, $[S_1^*]$ and $[S_2^*]$, by σ_1 and σ_2. Symbolically,

$$[S_1] = \frac{1}{\sigma_1}[S_1^*] \quad , \quad [S_2] = \frac{1}{\sigma_2}[S_2^*] \tag{29}$$

The replacement solutions $[S_1^*]$ and $[S_2^*]$ remain the same superposition type given by Eq. (21) but the coefficients are to be determined from the appropriate replacement boundary conditions (26), (27) and (26), (28).

STRESSES AND DISPLACEMENTS

Once a stress function $\phi = \phi(\alpha, \beta)$ has been chosen, the computation of stresses and displacements follows an established pattern. Details in the bipolar system may be found in (8). The displacement components in Fourier series in α for particular solutions are as follows:

Solution [S"]:

$$\left.\begin{aligned} 4G\mu c\, u_x &= 2 \sinh \beta \\ 4G\mu c\, u_y &= -2 \sin \alpha \end{aligned}\right\} \tag{30}$$

Solution [$S_{4''}$]:

$$\left.\begin{aligned} 4G\mu c\, u_x &= -\sinh\beta + \sinh 2\beta \cos\alpha - \sinh\beta \cos 2\alpha \\ 4G\mu c\, u_y &= (1+\cosh 2\beta)\sin\alpha - \cosh\beta \sin 2\alpha \end{aligned}\right\} \tag{31}$$

Solution [S_{10n+2}], n = 1, 2,...

$$\begin{aligned} 4G\mu c\, u_x = {} & (n+1)\sinh n\beta \cos(n-1)\alpha + \\ & + [4(1-\nu)-2(n+1)]\sinh(n+1)\beta \cos n\alpha + \\ & + [(-3+4\nu)\sinh n\beta + n\sinh(n+2)\beta]\cos(n+1)\alpha \end{aligned} \tag{32}$$

$$\begin{aligned} 4G\mu c\, u_y = {} & (n+1)\cosh n\beta \sin(n-1)\alpha - \\ & - [4(1-\nu)+2n]\cosh(n+1)\beta \sin n\alpha + \\ & + [(3-4\nu)\cosh n\beta + n\cosh(n+2)\beta]\sin(n+1)\alpha \end{aligned} \tag{33}$$

Solution [S_{10n+4}], n = 1, 2, ...

$$\begin{aligned} 4G\mu c\, u_x = {} & [n\sinh(n-2)\beta + (3-4\nu)\sinh n\beta]\cos(n-1)\alpha - \\ & - [4(1-\nu)+2(n-1)]\sinh(n-1)\beta \cos n\alpha + \\ & + (n-1)\sinh n\beta \cos(n+1)\alpha \end{aligned} \tag{34}$$

$$4G\mu c\, u_y = [n\cosh(n-2)\beta - (3-4\nu)\cosh n\beta]\sin(n-1)\alpha -$$
$$-[-4(1-\nu)+2n]\cosh(n-1)\beta \sin n\alpha +$$
$$+ (n-1)\cosh n\beta \sin(n+1)\alpha \qquad (35)$$

The stress components at points on the y-axis ($\beta = 0$) for the solution [S] are:

$$2c^2\sigma_\alpha = -2c'' + (1 - 2\cos\alpha + \cos 2\alpha)\,d' +$$
$$+ \sum_1^\infty c_n [-n(n+1)\cos(n-1)\alpha + 2(n+1)^2\cos n\alpha -$$
$$- n(n+3)\cos(n+1)\alpha] +$$
$$+ \sum_1^\infty d_n [-n(n-3)\cos(n-1)\alpha + 2(n-1)^2\cos n\alpha -$$
$$- n(n-1)\cos(n+1)\alpha] \qquad (36)$$

$$2c^2\sigma_\beta = 2c'' + (-1 + 2\cos\alpha - \cos 2\alpha)\,d' +$$
$$+ \sum_1^\infty c_n [n(n+1)\cos(n-1)\alpha - 2(n^2-1)\cos n\alpha +$$
$$+ n(n-1)\cos(n+1)\alpha] +$$
$$+ \sum_1^\infty d_n [n(n+1)\cos(n-1)\alpha - 2(n^2-1)\cos n\alpha + \qquad (37)$$
$$+ n(n-1)\cos(n+1)\alpha]$$

$$\tau_{\alpha\beta} = 0 \qquad (38)$$

Putting $\alpha = \pi$ in Eq. (36)(37), we have stresses at x = y = 0.

MATRICES FOR REPLACEMENT PROBLEMS

The coefficients involved in the solutions $[S_1^*]$ and $[S_2^*]$ are obtainable in explicit forms, Ref. (4). Due to the present advances in high speed digit computers, the work of evaluation of coefficients is left to the computer. The conditions expressed by Eq. (27), (28) can be written as

$$\mp\, c'' + \sum_{1}^{\infty} (c_n + d_n) = 0 \tag{39}$$

This constitutes the first rows in the matrices, with the upper sign referring to $[S_1^*]$, the lower sign referring to $[S_2^*]$. The boundary conditions in the form of Eq. (26) determine the rest of the matrices (identical in both cases) following the sequence 1, $\sin\alpha$, $\cos\alpha$, $\sin 2\alpha$, $\cos 2\alpha$, ... with elements supplied by Eq. (30)-(35). The systems of equations for superposition coefficients are given by Eq. (40).

$$\begin{bmatrix}
A_{11} & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & \cdots \\
A_{21} & A_{22} & A_{23} & A_{24} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \cdots \\
-2 & A_{32} & A_{33} & A_{34} & A_{35} & A_{36} & 0 & 0 & 0 & 0 & 0 & 0 & \cdots \\
0 & A_{42} & A_{43} & 0 & A_{45} & A_{46} & 0 & 0 & 0 & 0 & 0 & 0 & \cdots \\
0 & A_{52} & A_{53} & 0 & A_{55} & A_{56} & A_{57} & A_{58} & 0 & 0 & 0 & 0 & \cdots \\
0 & A_{62} & A_{63} & 0 & A_{65} & A_{66} & A_{67} & A_{68} & 0 & 0 & 0 & 0 & \cdots \\
0 & 0 & 0 & 0 & \checkmark & \checkmark & \checkmark & \checkmark & \checkmark & \checkmark & 0 & 0 & \cdots \\
0 & 0 & 0 & 0 & \checkmark & \checkmark & \checkmark & \checkmark & \checkmark & \checkmark & 0 & 0 & \cdots \\
0 & 0 & 0 & 0 & 0 & 0 & \checkmark & \checkmark & \checkmark & \checkmark & \checkmark & \checkmark & \cdots \\
0 & 0 & 0 & 0 & 0 & 0 & \checkmark & \checkmark & \checkmark & \checkmark & \checkmark & \checkmark & \cdots \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \checkmark & \checkmark & \checkmark & \checkmark & \cdots \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \checkmark & \checkmark & \checkmark & \checkmark & \cdots \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \cdots
\end{bmatrix}
\cdot
\begin{bmatrix} c'' \\ d' \\ c_1 \\ d_1 \\ c_2 \\ d_2 \\ c_3 \\ d_3 \\ c_4 \\ d_4 \\ c_5 \\ d_5 \\ \vdots \end{bmatrix}
=
\begin{bmatrix} 0 \\ \cosh\beta_0 \\ 0 \\ -1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ \vdots \end{bmatrix}
\tag{40}$$

In Eq. (40), $\vee$ denotes non-vanishing elements of $[A_{ij}]$ for $i > 6$, which are given in Table I.

$$\left.\begin{aligned} A_{11} &= -1 \text{ for replacement problem corresponding to case 1} \\ A_{11} &= +1 \text{ for replacement problem corresponding to case 2} \end{aligned}\right\} \quad (41)$$

The remaining non-vanishing elements are omitted here and may be found in Ref. (9). For a given value of β, Eq. (40) and (41) yield two sets of coefficients evaluated on a computer.

TABLE I. NON-VANISHING ELEMENTS OF $[A_{ij}]$ for $i > 6$

j \ i	$2n+1$	$2n+2$
$2n-1$	$(3-4\vee)\cosh(n-1)\beta_0 + (n-1)\cosh(n+1)\beta_0$	$(-3+4\vee)\sinh(n-1)\beta_0 + (n-1)\sinh(n+1)\beta_0$
$2n$	$(n-2)\cosh(n-1)\beta_0$	$(n-2)\sinh(n-1)\beta_0$
$2n+1$	$-2(n+2-2\vee)\cosh(n+1)\beta_0$	$-2(n-1+2\vee)\sinh(n+1)\beta_0$
$2n+2$	$-2(n-2+2\vee)\cosh(n-1)\beta_0$	$-2(n+1-2\vee)\sinh(n-1)\beta_0$
$2n+3$	$(n+2)\cosh(n+1)\beta_0$	$(n+2)\sinh(n+1)\beta_0$
$2n+4$	$(n+1)\cosh(n-1)\beta_0 - (3-4\vee)\cosh(n+1)\beta_0$	$(n+1)\sinh(n-1)\beta_0 + (3-4\vee)\sinh(n-1)\beta_0$

RESULTS ON THE ORIGINAL PROBLEMS $[S_1]$ AND $[S_2]$

Referring to Fig. 1, the natural choice for the characterization of the configuration is obviously the diameter 2r of the microfibers and the partition thickness 2δ between the microfibers. The values of c and β_o corresponding to a given set of values of r and δ are determined by

$$r = c / \sinh \beta_o \quad , \quad r + \delta = r \cosh \beta_o \tag{42}$$

The characteristic shape ratio of the configuration is

$$(\delta / r) = \cosh \beta_o - 1 \tag{43}$$

The following table shows a few corresponding values of (δ/r) and β_o.

δ/r	0.002	0.01	0.1	0.2786	1	2	4.037	9.068	26.31
β_o	0.0632	0.1413	0.444	0.7301	1.317	1.7627	2.3	3.0	4.0

The stresses along x-axis between microfibers and those on the boundary $\beta = \beta_o$ can be easily computed once the coefficients involved in Eq. (21) have been determined. However we are interested in the maximum normal stress in the matrix for fibers in close proximity, which is believed to be σ_x at x = y = 0. Therefore the computation is concentrated on stresses at that point and the results are plotted in Figs. 3 and 4 for $[S_1]$ and $[S_2]$ respectively.

The stress concentrations for close proximity of both microfibers were to be expected, but the stress reversal of σ_y from positive 1 at infinity to negative values near the origin present an unexpected surprise. The reversal is conspicuous in case 2: pulling in the y-direction produces a region of biaxial compression in the matrix between the microfibers. Explanation: the two fibers amount (almost) to a rigid body shaped like ∞ preventing the lateral Poisson contraction in the x-direction, hence a squeeze (pressure) in matrix with a compressive σ_x. From $\epsilon_z = 0$ (plane strain) and $\epsilon_y = 0$ (or nearly so: adhesion of matrix to fiber) follows by Hooke's law $(1-\nu)\sigma_y = \nu \sigma_x$, whence a compressive σ_y. Fig. 5 shows σ_y plotted for points along the y-axis for $\nu = 0.4$, $\beta_o = 0.444$. The curve shows stress reversal.

For large values of β_o the stresses σ_y at x = y = 0 in $[S_2]$ are positive, Fig. 4. The critical value of β_o which gives $\sigma_y = 0$ at the point x = y = 0 is found to be 0.7301 for $\nu = 0.4$. Stress reversal appears as β_o becomes smaller than the critical value.

The stress concentration increases as β_o decreases. For $\nu = 0.4$, the increase is in a more gradual manner for $\beta_o > 1$. It becomes very rapid for $\beta_o < 0.5$. A conclusion can be drawn that the volume percentage of the embedded fibers has an upper limit over

which the increase of fiber contents might weaken the composite material due to the rapid increase in stress concentration. Assuming the embedded fibers are uniformly distributed in a hexagonal pattern, the volume percentages of fibers are 56% and 80% respectively for $\beta_o = 1$ and 0.5. This suggests that the increase in fiber contents will strengthen the composite when the fibers are 56% or less in volume. But at high fiber contents, say at 80%, destructive stress concentrations in the matrix might prevail.

REFERENCES

1. Kirsch, G., "Die Theorie der Elastizität und die Bedürfuisse der Festigkeitslehre," ZVD42, 1898.

2. Sadowsky, M. A., "Two-Dimensional Problems in the Theory of Elasticity," ZAMM, Vol. 8, 1928.

3. Howland, R. C. J., "Stresses in a Plate Containing an Infinite Row of Circular Holes," Proc. Roy. Soc. London A148, 1935.

4. Ling, C. B., "On the Stresses in a Plate Containing Two Circular Holes," J. of Appl. Phys., Vol. 19, No. 1, 1948.

5. Davies, G. A. O. and Hoddinott, J. R., "Stresses in a Plate Pierced by Two Unequal Circular Holes," J. Roy. Aero. Soc., Vol. 67, 1963.

6. Goodier, J. N., "Concentration of Stress Around Spherical and Cylindrical Inclusions and Flaws," Trans. ASME, Vol. 55, 1933.

7. Jeffery, G. B., "Plane Stress and Plane Strain in Bipolar Coordinates," Phil. Trans. A221, 1921.

8. Sadowsky, M. A. and Pu, S. L., "Effect of Interface Surface Tension Arising Between Parallel Microfibers at Close Distance," Watervliet Arsenal Tech. Rpt, WVT-6725, 1967.

9. Sadowsky, M. A. and Pu, S. L., "Stress Reversal in Bonding Material Between Fibers," Watervliet Arsenal Tech. Rpt, WVT-6736, 1967.

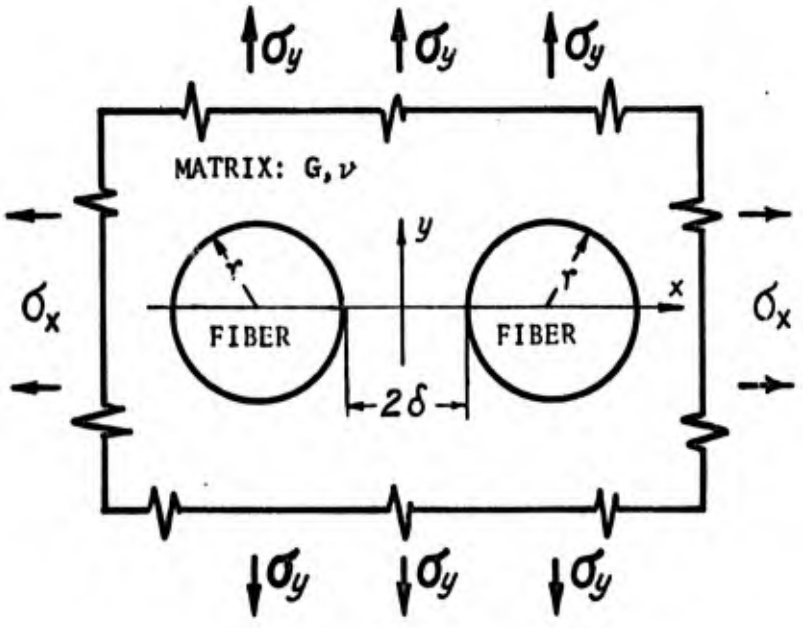


FIGURE 1. TWO RIGID FIBERS EMBEDDED IN A MATRIX

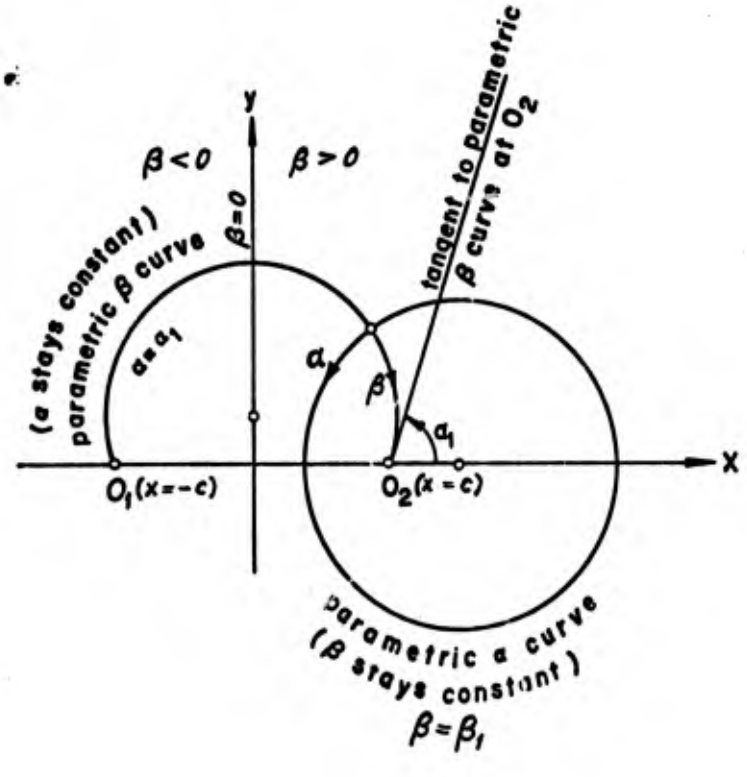


FIGURE 2. BIPOLAR SYSTEM OF COORDINATES

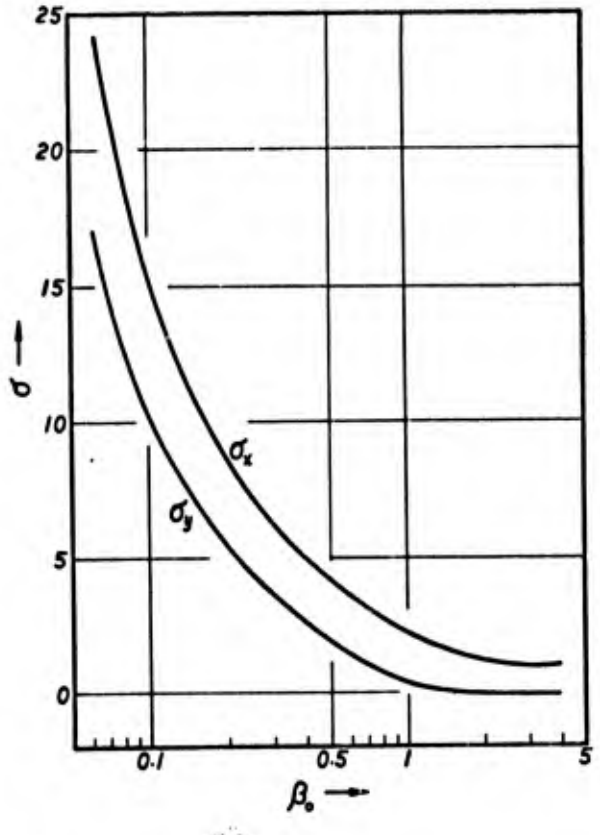
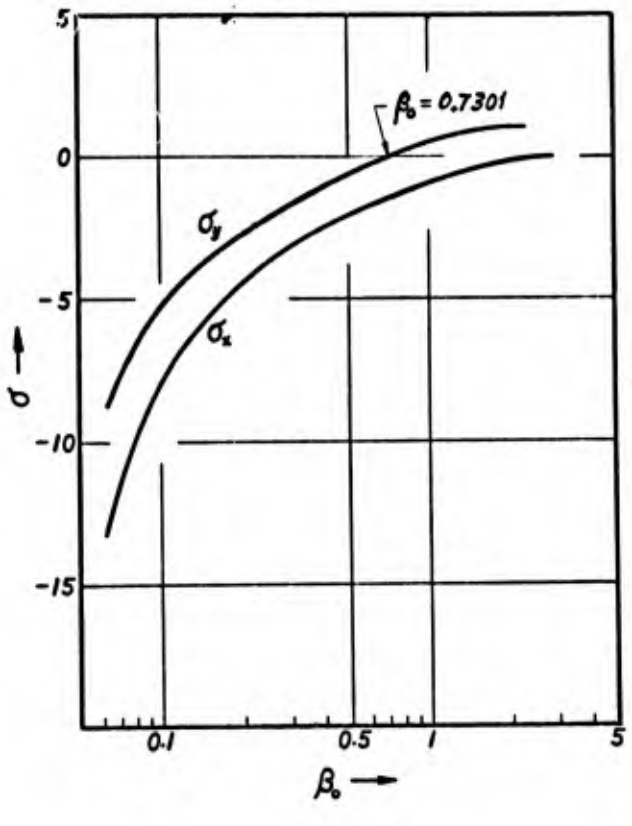


FIGURE 3.

NORMAL STRESSES AT x = y = 0

IN [S_1] FOR ν = 0.4



FIHURE 4.

NORMAL STRESSES AT x = y = 0

IN [S_2] FOR ν = 0.4

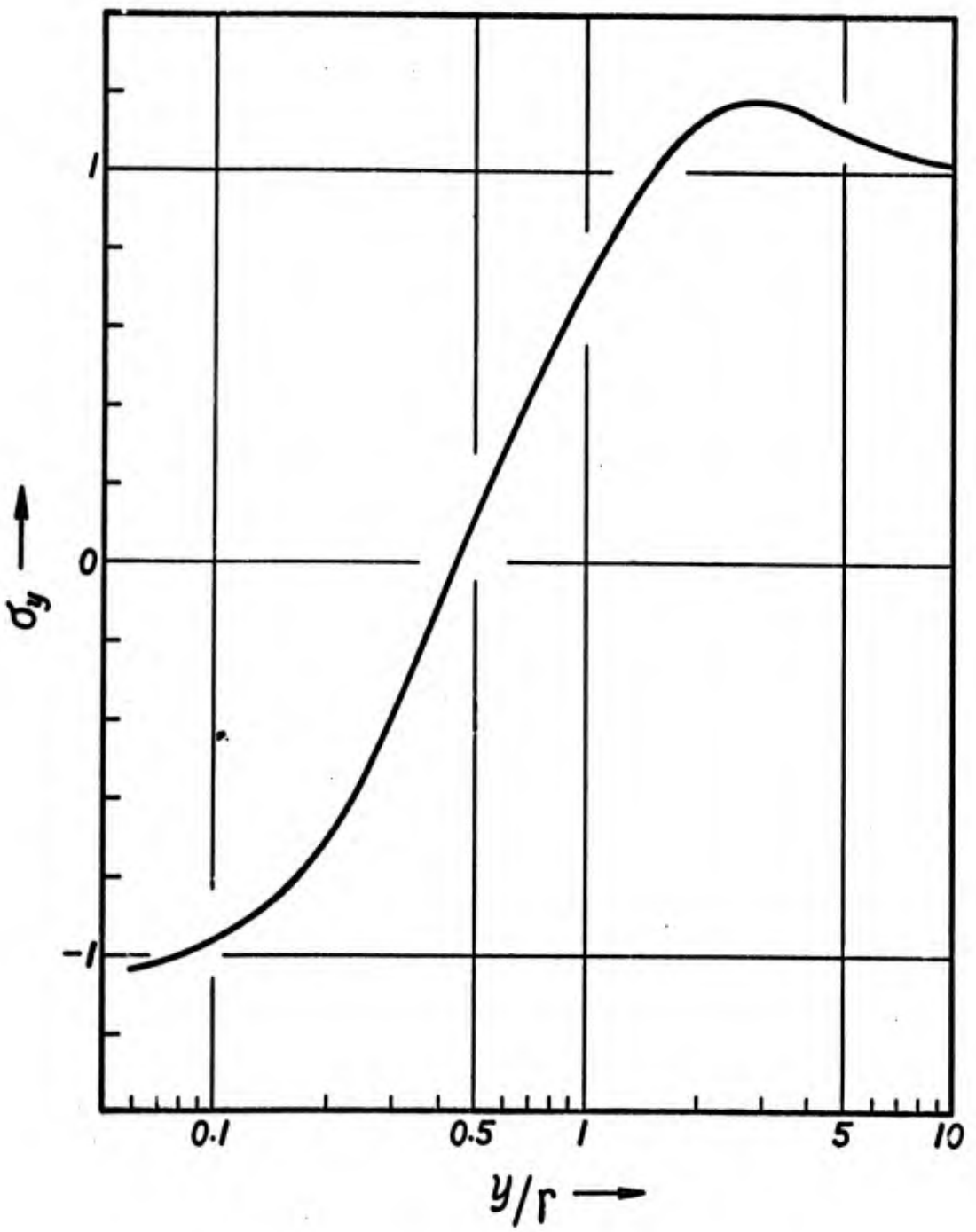


FIGURE 5. THE STRESS σ_y ALONG THE y-AXIS IN [S_2] FOR ν = 0.4, β_o = 0.444

NEW KNOWLEDGE ON THE FUNDAMENTAL LIMITATIONS TO OPERATION OF FIELD-EFFECT TRANSISTORS

KEATS A. PULLEN, JR., D.E.
U. S. ARMY BALLISTIC RESEARCH LABORATORIES
ABERDEEN PROVING GROUND, MARYLAND

INTRODUCTION

The field-effect transistor is an active electron device of the charge-control type. In it, the control of current flow is obtained by control of carriers in a channel of semiconductor material. Control is obtained by immobilizing free electrons in crystal lattice sites which are otherwise capable of aiding the movement of charge. The basic device consists of a source of charge carriers (an emitter in the parlance of bipolar transistors), a channel through which more or less carriers may move, and a drain for removal of the carriers (called a collector with bipolar transistors). The structure which controls the immobilization of electrons in lattice sites is called the gate structure.

A graphical representation of the structure of the field-effect transistor is shown in Figure 1. An accelerating voltage is applied between source and drain; this voltage provides the stress required to draw charges from the source to the drain. A control voltage is applied to the gate structure, or control electrode; this voltage controls the number of lattice sites in which carriers are immobilized and thereby controls the total current capacity of the channel. The channel therefore includes two regions, the conducting region, in which the carriers are free to move, and the depletion region, in which the carriers are immobilized. Consequently, the flow of current through the device may be changed by varying the relative sizes of the depletion regions and the conduction region.

Field-effect transistors are typical active or amplifying devices in that they provide a means whereby one kind of an applied variable can be used to control the behavior of another kind of variable. In terms of general vector or tensor theory, this is stated as an "across" or covariant variable controlling a "through" or contravariant variable; typically, both an across and a through

variable are required to be present in any path or circuit for energy to be transferred or power to be dissipated.

Curiously, all presently known solid-state devices capable of amplification involve the interaction of an across variable controlling a through variable. Whereas it appears to be possible theoretically to build active devices involving the control of an across variable by an appropriate through variable, the only device presently known to approach this definition, the transformer, is passive or bilateral (operates in either direction).

The theory for the field-effect transistor was originally defined by Shockley[1] and was extended by Dacey and Ross.[2] This theory is generally satisfactory in most respects, but it was found not to explain a second mode of operation described in a letter to the *Proceedings of IEEE* by Evans and Pullen.[3] It is the purpose of this paper to establish the modifications required to explain the recently reported alternate mode of operation and to examine the implications of the new knowledge in solid-state electronics.

ANALYTICAL CONSIDERATIONS

The equations for the normal operating mode for the field-effect transistor, based on the derivations of Shockley and by Dacey and Ross, lead to the following relations:

$$g_m = g_{mo}\left[1 - \sqrt{(-V_a/W_o)}\right] \tag{1}$$

$$I_d = (g_{mo}W_o/3)\left[1 + (V_a/W_o)(3 - 2\sqrt{(-V_a/W_o)}\right] \tag{2}$$

$$g_m/I_d = (3/W_o)\left[1 - \sqrt{(-V_a/W_o)}\right]/\left[1 + (V_a/W_o)\right.$$

$$\left.(3 - 2\sqrt{(-V_a/W_o)})\right] \tag{3}$$

Whether the quotient of the resulting indeterminate expression obtained as $V_a \to -W_o$ is finite or infinite depends on the proper evaluation of a radical, but clearly it does not appear to approach a basic physical parameter in the limit. (Practically, the quotient of Eq. 3 becomes unbounded as I_d approaches zero.)

The disclosure that the value of g_m/I_d for a field-effect transistor has a maximum value less than the Fermi constant, $\Lambda = q/kT$, over from two to five orders of magnitude of drain current introduces a new aspect to the concept of operation of these devices. Here Λ^{-1} = 26 millivolts or 26 ohm-milliamperes, q is the charge on the electron, k is Boltzmann's constant, and T is the absolute temperature. Clearly, the Shockley and the Dacey-Ross theory

of operation of the devices needs extension to include a range of operation in which the value of g_m/I_d is not explicitly dependent on gate voltage.

Examination of Shockley's original paper shows that the effect of a transition region in which the magnitude of the bound charge changes from zero in the channel to ρ has been neglected in Shockley's analysis. This transition region, of thickness roughly comparable with one "Debye length," experiences a potential jump of Λ^{-1} volts. If, as may occur in this case, the transition region is of very small depth, the electric field gradient across the boundary can become rather large. In this instance, the effect of the boundary is by no means negligible.

It should be noted that if the density of free carriers increases from channel center toward the gate the effect is to decrease the effective width of the transition region. The result of such a change would be to increase the importance of the Fermi or diffusion mode of operation of the field-effect transistor and correspondingly decrease the importance of the conventional square-law mode.

ELECTRIC FIELD CALCULATIONS

It is enlightening to calculate the electric field in the ideal channel with a variety of charge site distributions. For the purpose of this analysis, it is necessary to make the following assumptions:

1. Equal total number of charge-migration sites distributed from gate to gate across the channel leads to equal maximum current capacity.

2. Conventional Poisson theory techniques can be applied.

3. Diffusion of current in the channel is maximum when the cross-channel potential gradient is a minimum. (In other words, one looks for a diffusion mode of operation by searching for conditions of minimum potential gradient across the channel.)

4. Channel current in an infinitesimal layer of channel is proportional to the carrier site density in that infinitesimal layer. (This is reasonable since charges have to diffuse or drift through these sites.)

5. Effects of discrete charge size may be neglected.

These conditions are generally compatible with the conditions chosen by Shockley and by Dacey and Ross, and permit study of channel distributions not considered by them.

The channel doping distributions normally considered for field-effect transistors include constant doping across the channel and a negative parabolic doping having maximum density of carriers in the center and zero at either gate. Calculation of the binding voltage required to control channel flow for different kinds of channel doping leads to the results shown in Figure 2 for power-law doping, negative parabolic, and hyperbolic doping. The equations for the total channel potential difference as a function of exponent in the power-law case, for the parabolic, and the hyperbolic cases are:

$$V_d/Ke_o = 1/(n + 2) \qquad \text{Power Law} \qquad (4)$$

$$V_d/Ke_o = 5/8 \qquad \text{Parabolic} \qquad (5)$$

$$V_d/Ke_o = (\coth n - \operatorname{csch} n)/n \qquad \text{Hyperbolic} \qquad (6)$$

This latter function may be approximated by the expansion:

$$V_d/Ke_o = 1/2 - n\ /24 + n\ /240 - 17n\ /\ 8! + \ldots \qquad (7)$$

The following table indicates the relative value of the various channel-closing potentials for different values of n:

n	0	1	2	3	4	Parabola	cosh y - 1
Power series	0.50	0.33	0.25	0.20	0.17	0.62	--
Hyperbolic	0.50	0.46	0.38	0.30	0.24	--	0.245

These data show that the binding voltage required to "close" the channel is significantly reduced when the carrier sites in the channel are concentrated on the outer edges of the channel and relatively little current flow can occur in the center.

The advantage of the power-law distribution over the hyperbolic is possibly surprising. Two factors contribute to this, however. First, for $n > 0$, the carrier concentration in the channel center is zero with the power-law distribution, whereas it approaches a small, but finite, value for the hyperbolic. The second feature is that the correct format for the power-law distribution is really $|y|^n$, not y^n. Probably the hyperbolic distribution would be more readily obtained with diffusion or epitaxial growth techniques, however. A doping structure based on (cosh y - 1) is significantly better than the power distributions for $n \leq 2$, showing the importance of minimum doping at channel center.

A little reflection on the meaning of the above potential distributions can lead to a better understanding of expected device

behavior. Whereas a channel having an inverted parabolic distribution of charge sites can experience relatively substantial current flow, only a small part of the current flowing is actually controlled by the electric field in the transition region, and the balance streams across uncontrolled. Under this condition, the boundary region will be comparatively broad, and the normal Shockley and the Dacey-Ross theory will apply except when the depletion region almost closes the channel.

Examination of the curves in Figure 3 showing simulated constant-potential contours in the channel gives particular emphasis to this situation. As can be seen, the region of near-zero gradient is crowded to the channel center with the parabolic (center peak value) distribution and is only a little less crowded in the center for the uniform distribution. With the power-series distributions particularly, however, the region of near-zero gradient widens as the value of the exponent n is increased.

THE DIFFUSION MODEL

Based on a diffusion model of channel current control at very small currents, it is possible to calculate the change in gate voltage required to change channel current by a given number of orders of magnitude. For example, a change of channel current from one nanoampere (10^{-9} amperes) to one milliampere, a current ratio of one million to one, requires a minimum change in potential as defined by the equation:

$$\Delta V_g = \ell n\ 10^6 = \ell n(I_2/I_1) \qquad (8)$$

From this equation, and the fact that the natural logarithm of 10^6 is 13.8155, the required minimum potential change, V_g, is readily seen to be 0.36 volts. Examination of a typical set of curves for a field-effect transistor shows that the gate voltage change required to produce a given channel current change greatly exceeds the value required in a diffusion model. In fact, it may be ten or more times the value given by theory.

A similar calculation may be made for the voltage change required to produce a two-to-one change in the current in the channel; the resulting change is approximately 18 millivolts. This represents the minimum voltage change which can be expected to lead to the corresponding current change. Values between 18 and 20 millivolts for a two-to-one change in current were found to exist over several orders of magnitude of channel current, as was reported by Evans and Pullen.[3]

If a channel system is considered in which only a channel of infinitesimal thickness is opened by the action of the gate, then

the current change introduced by a change of gate voltage is simply the current flow in the sheet of infinitesimal thickness. If the assumption is made that as the gate is opened the current carried by each new sheet element is substantial or large compared to the previous total current, a current equation of the following form results:

$$I_T = \sum_{u}^{j-1} \Delta I_u + \Delta I_j \tag{9}$$

where ΔI_j is the current increment in the sheet element j. If the current increments increase as assumed, then:

$$I_T = \sum_{u}^{j-1} + \Delta I_j = \sum_{u}^{j} \Delta I_u \triangleq \Delta I_j > \sum_{u}^{j-1} \Delta I_u. \tag{10}$$

Now, the partial of I_T with respect to V_g may be obtained in terms of the quotient of partials:

$$\partial I_T/\partial V_g = (\partial I_T/\partial y)/(\partial V_g/\partial y) \tag{11}$$

But the value of $\partial I_T/\partial y$ is ΔI_j, the current increase in the incremental distance y, and the value of $\partial V_g/\partial y$ is a function of the field intensity and also of the parameter $(\Lambda)^{-1}$. As a result, the expression for the transconductance per unit current takes the approximate form:

$$(1/I_d) \cdot (\partial I_d/\partial V_g) = [f(\Lambda)\Delta I_j]/[\sum_{u}^{j} I_u] \longrightarrow \kappa\Lambda \tag{12}$$

In this equation, the function of lambda multiplied by the current ratio has been represented as a linear function of lambda, where the coefficient kappa has a value in the range $0 < \kappa < 1$, and lambda again is the Fermi constant. The coefficient κ measures the control efficiency of the gate field. Clearly, under diffusion conditions in the channel, the value of kappa is relatively fixed and can be expected to be in the range between 0.4 and 1.0. Under ideal conditions, where the summation term is very small compared to the incremental term, a value of almost exactly unity might be expected. Naturally, it is impossible to achieve this condition.

If the value of the current increment approximately equals the value of the uncontrolled current in the channel, then the value of kappa can be expected to be one half, since only half of the current flow can be controlled by the gate field. As the fraction of the channel current which is uncontrolled increases, the value of kappa becomes very small, and the Shockley and the Dacey-Ross mode of operation results.

In typical devices, the maximum limiting value of kappa need not approach unity. In fact, with metal-insulator-semiconductor devices (MISFETs or IGFETs) the limiting value is more nearly one half, again over several decades of current. The reason for this condition is still obscure, but it may be that imperfect control cannot be avoided with these devices. Typically, they have a control field applied from only one side, and there may be substantial partially controlled leakage current on the "back side" in spite of the extremely low conductivity of the semiconductor material. The control of conduction takes place by the capacitive binding of charges across the capacitor or by attraction of mobile charges to the neighborhood of the insulator, and the channel is unbounded on one side unless a diode is built into the chip to provide a definite boundary.

The idea is often expressed that the ideal structure for a field-effect transistor is one having a channel of infinitesimal thickness. The theory herein developed in fact confirms the "infinitesimal thickness" theory and suggests that possibly the optimum approach to achieving it may be by developing a "two-sheet" channel, or one consisting of two very thin current sheets.

The two-current-sheet structure described can be obtained only by designing the channel to concentrate current flow against the gates. Such a channel design must be based on a variable doping density going from zero (intrinsic material) to a high doping density against the gates as indicated by the channel calculations already discussed.

OTHER CHARGE DISTRIBUTION PHENOMENA

Just as the transverse field in the channel is a function of y^{n+1}, the total charge between channel center and the infinitesimal thickness plane at point y is also a function of y^{n+1}. This means that the boundary point at which a given total mobile charge is reached is determined by the same function of y as is the transverse field. Superficially, this might seem to indicate that no benefit could be gained from use of a graded channel doping distribution. However, such is not necessarily the case. In the first place, the width of the transition region between channel and depletion zone becomes substantially smaller as the value of n is increased. This means that the point of transition between square-law and diffusion control in the channel will occur at a higher value of $\partial V/\partial y$ in a channel doped to provide a higher exponential distribution of doping. In addition, over the range of values of $\partial V/\partial y$ of concern here, the critical value of the slope of $\partial V/\partial y$ at the point of transition between the square-law and the diffusion modes normally decreases as the value of n is increased. All these conditions lead to a higher transconductance per unit current as the exponent n is increased.

The curves for electric field for a partly open channel may be obtained from the closed-channel curves in Figure 4. Since the channel is opened by the removal of the electric field within the open part of the channel, examination of the equations shows that the field becomes zero between $y = 0$ and $y = y_o$, and follows the contour shown in Figure 4 between $y = y_o$ and $y = 1$. As a result, the section of the curves between $y = 0$ and $y = y_o$ should be deleted for a channel open to $y = y_o$, and the curves should be translated vertically downward so that the field is exactly zero for each curve at $y = y_o$. Consequently, these curves can be used in the study of either closed or open channels by the use of the described replotting techniques.

Examination of the slope of the various curves shows that over the central part of the channel, for $y < (n + 1)^{-1/n}$, the curves for distributions having higher numerical exponents have lower slopes than those for the corresponding lower exponents. Consequently, as shown in Figure 5, the diffusion mode of operation will be the mode of operation to increasingly high channel current as the exponent value is increased. This can be expected to yield higher transconductances per unit current at normal operating currents as well.

In general terms, then, the equation for the transconductance per unit current for the field-effect transistor takes the form:

$$g_m/I_d = (\partial I_d/\partial y)/I_d[(\partial V_g/\partial y) + (1/\Lambda)(\partial\{f(y_o)\delta(y-y_o)\}/\partial y_o)] \quad (13)$$

where the first denominator term is the field gradient term and the second, the transition field term. Strictly, this second term takes the form of a derivative of a set of functionals, because the transition is not either "infinitely sharp" nor uniform. Because of the nature of the transition and the way it moves with the channel depletion boundary, however, the value of the term is always non-zero, and its magnitude may be approximated as a function of y_o in a manner such as has been chosen.

Because of the nature of the charge control problem encountered with the field-effect device, it is really remarkable that values of g_m/I_d approaching the value of Λ can be obtained at all. This condition can only exist if the magnitude of $\partial I_d/\partial y$ is approximately equal to I_d. This leads to a current equation in the form:

$$dI/I = \kappa\Lambda\tau\, dy \quad (14)$$

where τ is a scaling factor. Integrating gives:

$$I/I_o = \exp(\kappa\Lambda\tau y) \quad (15)$$

Evidently, an increase in doping as a function of distance from the channel center tends to maximize the transconductance per unit current toward the value called for by the Fermi constant.

APPLICATION CONSIDERATIONS

Bipolar transistors have proven to be of considerable use in tuned amplifier circuits for low-power, high-frequency applications, but they have not proven suitable replacements for electron tubes in high-power applications. The reason for this is that the intrinsic transconductance per unit current of 39 mhos per ampere existing in the bipolar transistor introduces very difficult design compromises.

The transconductance within the intrinsic transistor, namely from base to collector directly across the junction, generates voltage gain both within the device and at the device terminals. The total impedance in the output circuit, the load impedance, collector spreading impedance, and even emitter spreading impedance, all contribute to the overall voltage gain in the forward direction. It is therefore necessary to make certain that all internal "loop gains" of the amplifier, or the open-loop gain around all critical paths within the amplifier be substantially less than unity under any possible combination of test or operating conditions. Typically, for the bipolar transistor, the path from the intrinsic base-to-emitter junction to the collector-to-base junction through the forward conductance and back to the base-emitter junction through the parasitic admittances of the device is the most critical one. Practically, the voltage gain must be less than 20 across the intrinsic junction from base to collector, or less than 200 from emitter to collector, to assure that an internal loop gain of less than unity is reasonably probable.

To see how this limitation affects the conventional transistor amplifier and how the field-effect transistor can alleviate the gain problem, consider the limitations on the design for a solid-state power amplifier subject to the following requirements:

Input power from mains	100 watts
Supply voltage	28 volts
Intrinsic g_m/I for bipolar unit	Λ
Effective g_m/I for FET	$\Lambda/10$
Peak output current	5 times average

Now, based on an average current of four amperes, a peak of 20 amperes will flow, and the maximum permitted output impedance is therefore 1.4 ohms. This impedance, which will apply under mismatch conditions, sets the maximum value which will allow the required peak current to flow. As the output is loaded properly, this impedance will decrease and, as long as the input signal power is sufficiently large, voltage gain will decrease and efficiency will increase.

Assuming that the amplifier has a common-base configuration, the maximum instantaneous voltage gain MUST be less than 200 (many

engineers limit it to 100 in circuits having particularly critical stability requirements), and the maximum permitted intrinsic transconductance then is 140 mhos, or about seventeen percent of that intrinsically available.

Present-day field-effect transistors have transconductances per unit current less than about fifteen percent of this value, whereas the intrinsic value for the bipolar device is nearly six times too large. Whereas experience with electron tubes has shown that an increase in supply voltage makes it possible to use devices with comparatively poor g_m/I efficiencies, there appears to be no way in which an excessive g_m/I ratio can be controlled reliably. For this reason, there is a developing interest in the application of field-effect devices with RF tuned amplifiers.

A review of the conventional theory of RF amplifiers based on bipolar devices indicates that insufficient attention has been given to the multiplicity of internal feedback paths existing within bipolar transistors, and, as a result, the problem of the high intrinsic gain present from junction to junction has been overlooked. The field-effect device has a substantially simpler representation and has losses introduced in a way which leads to reduced g_m/I efficiency. This limits the peak available voltage gain in a very effective way. As a result, solid-state amplifiers based on field-effect devices should show greater power handling ability than similar units based on bipolar devices.

RADIATION TOLERANCES

The ability of a device to tolerate the application of nuclear radiation depends on the nature of the structural reaction to the radiation and on the relation of the resulting damage to its mechanism of operation. Several modes of damage, some temporary in nature and some permanent, can be created by radiation. The temporary damage is usually a result of ionization and consists of ejection of normally bound electrons through the impact of high-energy photons. Permanent damage can result from the impact of high-energy photons on the nucleus. It is also caused by structural disarrangement of the lattice or nuclei by neutrons.

Temporary damage often manifests itself as a current pulse which continues to flow until the freed charges have either recombined or been swept out of the devices. This temporary damage can lead to permanent damage if the current flow is sufficient to lead to thermal runaway before completion of the removal of the excess charges. For this reason, it is better that circuits intended for a radiation environment be designed to be power-limited when devices which are subject to radiation-induced conduction are used.

Protection against thermal destruction of semiconductor devices which are subjected to transient radiation may be achieved by the use of properly placed wire-wound resistors for current and power limitation. The resistance stability of wire-wound resistors in a nuclear environment is extremely high (a few percent change at most), and the value of resistance may easily be selected to limit energy dissipation to the rated values for the device.(4)

Damage encountered with bipolar transistors initially is a function of minority carrier lifetime, or diffusion length. Since the beta of a transistor is an inverse function of the square of the diffusion length and the diffusion length is an inverse function of total radiation exposure, rapid degradation of beta can be expected. As the dosage is increased, a second degrading factor enters, namely, a reduction of mobility with increased total dose.

The reduction of diffusion length, as noted, decreases the beta of the bipolar transistor in accord with a square-law relation. Since the carriers in a field-effect device are majority carriers, however, the damage leads to recombination in the channel. As a result, the initial effect on the FET is a reduction of current-carrying capacity due to recombinations, and little effect on either leakage currents or on the available g_m/I is either expected or observed.* At very high dose levels, the mobility reduction effect can be expected to degrade behavior, but this is above the normal range of usefulness of many electronic components.

No component of the Fermi "constant," the electron charge, Boltzmann's constant, or the absolute temperature has any direct dependence on nuclear radiation conditions encountered. With the bipolar transistor, both the intrinsic forward admittance and the intrinsic input admittance are dependent on the Fermi constant also, but the division of current between base and collector is dependent on the ratio of base width to diffusion length. This makes the collector current dependent on a function of the form:

$$1 - 0.5(W/L_p)^2 \qquad (16)$$

where W is the base width and L_p the diffusion length. In a similar way, the base current is dependent on:

$$(W/L_p)^2. \qquad (17)$$

As a consequence, the dependence of collector current is small until large exposure is encountered, but for the base current it is quite rapid. The absence of current division in the FET makes it free of this difficulty.

---

*This has been clearly shown in Report No. 1, Contract DA-04-200-AMC-725(X), issued by Siliconix, Inc.

When the channel of an FET is doped according to a power-law distribution, irradiation can be expected to increase the efficiency of the device and have relatively minor effect on the maximum current capacity. It also appears that a maximum breakdown voltage rating may result for a given total number of carrier sites in the channel, since the high-conductivity part of the channel is concentrated against the gate electrode where the immobilizing field is most effective.

CONCLUSION

It has been shown that a transconductance per unit current limitation exists at low current levels in the field-effect transistor. This limitation is consistent with the theory of these devices as developed by Shockley. By proper design, the transition current between the diffusion mode and the Shockley mode of operation may be reached at significantly higher currents than observed in present devices. The resulting devices offer promise of improving circuit design possibilities and also of leading to more radiation-resistant circuits.

REFERENCES

1. Shockley, W.: A Unipolar "Field-Effect" Transistor, Proc. IRE, Vol. 40, November 1952.

2. Dacey, G. C., and Ross, I. M.: The Field-Effect Transistor, BSTJ, November 1955.
   Dacey, G. C., and Ross, I. M.: Unipolar "Field-Effect" Transistor, Proc. IRE, Vol. 41, August 1953.

3. Evans, L., and Pullen, K. A.: Letter to the Editor, Proc. IEEE, Vol. 54, January 1966.

4. See Chapter 7, Pullen, K. A.: Handbook of Transistor Circuit Design, Prentice-Hall, 1962.

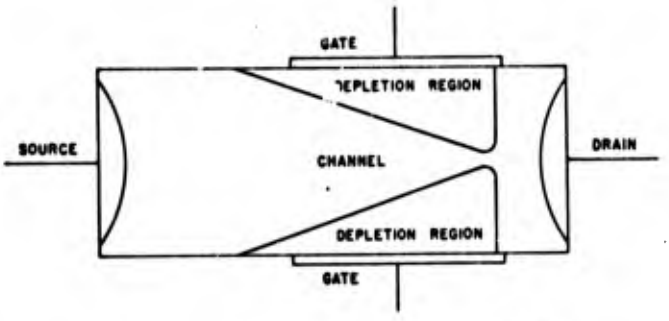


Figure 1. Idealized Field-Effect Transistor

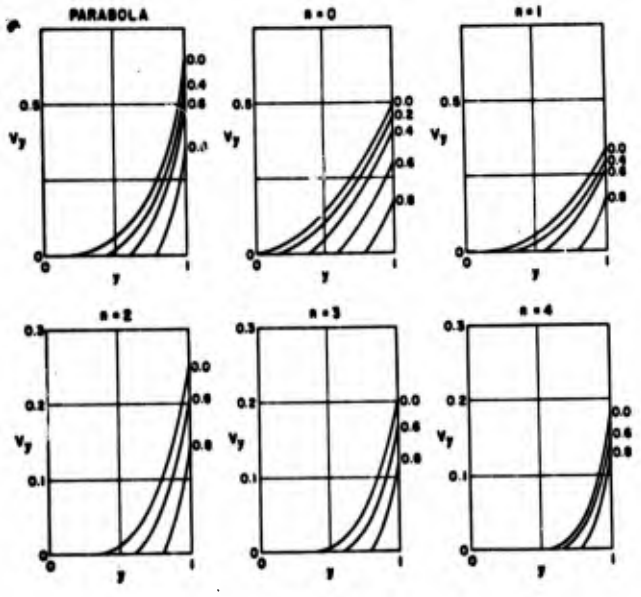


Figure 2. Channel Voltage Contours for Different Doping Distributions

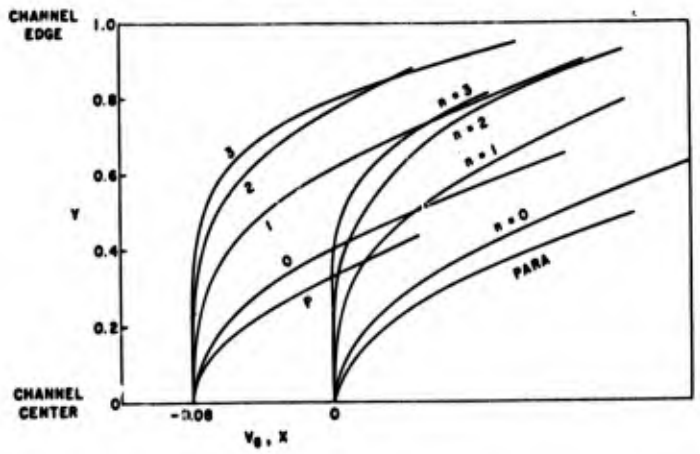


Figure 3. Constant Potential Contours for Channel - Different Doping Distributions

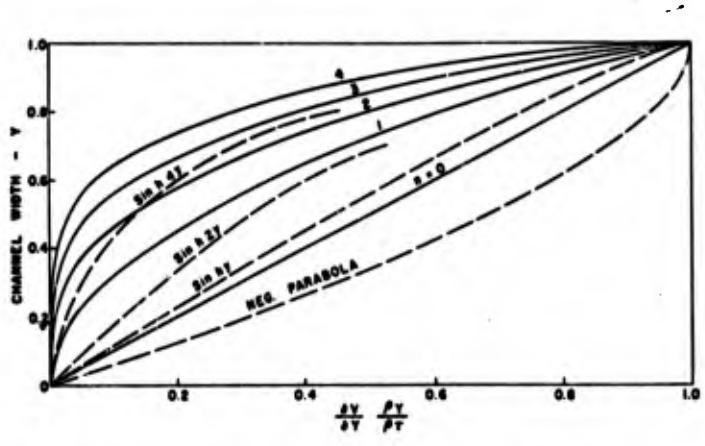


Figure 4. Channel Field Intensity - Different Doping Distributions

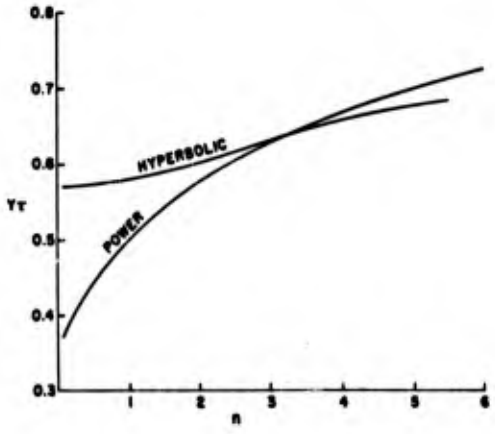


Figure 5. Points of Equal Slope of Electric Field for Power Law and Hyperbolic Channel Charged Distributions

PROBLEMS OF CONTROLLED BIRD-FLIGHT

JOHN J. ROMBA
U. S. ARMY LIMITED WAR LABORATORY
ABERDEEN PROVING GROUND, MARYLAND

In a general way, a bird's flight space can be described in terms of a complex set of conditions and relations that influence the bird's behavior at a given time. Within the framework of these conditions and relations, some of which may be amenable to experimental manipulation, the bird's behavior may be brought under some degree of control. Conditioned bird flight can be described as a succession of learned responses in which one response produces conditions essential to the next, until the last response terminates in a reinforcement.

Simple flight control is demonstrated when a bird learns to fly a familiar course to reach a place of food. The bird's flight domain may reach several miles. Controlled long distance outflights in unfamiliar surroundings, however, have been difficult to attain.

A sufficient condition for obtaining outflight is reinforcement provided at the end of a correctly made flight. In one procedure, the bird flies out to turn on a signal, such as the sound of a horn, which it has learned to associate with food reward. The horn comes to evoke a conditioned response, namely, return to the launch point and food. An effective way to develop this performance is initially to establish discrimination learning by rewarding key pecking on a small platform when a tone is present, but not when the tone is absent. Then the bird learns to turn on the tone by moving away from the platform. Guidance of flight is unnecessary with this procedure - outflight in any direction can be reinforced. In another procedure, reinforcement is obtained by the bird at a target station which is some predetermined distance from the launch point. With this procedure, flight guidance is essential. The bird can learn to guide on such terrain features as a trail, tree-line, stream, or by flying toward a landmark. It may also guide on telemetered signals. For this theme, consider that the guiding stimuli for the bird is a road.

Apparently, long distance flight in strange territory could be obtained if the bird found the "down-range" food station on every trial. In the event that the station is not found on some trials, long distance flight is still possible if an alternative procedure is provided to terminate the trial and have the bird return. An acoustic recall signal could serve as the controlling stimulus. The latter contingency without recall control places the burden for return on the bird, invariably resulting in the bird turning back "sooner than later" from outflight.

A study was made to find out how to obtain long distance outflight in strange territory with pigeons when the target station is absent from the field of flight on a portion of the trials.

THE EXPERIMENT

The setting for the field trials consisted of a dovecote (loft) mounted on a truck. The road ahead of the truck presented a continuum of potentially rewarding stimuli to the birds. The script called for the birds to leave the truck upon being released and then search along the road for the target, defined as a human whose grain containing hand was outstretched after the bird alighted nearby.

Recall learning occurred first. Training began with a schedule under which birds learned to fly to a compartment where they had to peck a key that brought on an audible recall signal. A transition was then made from this situation to one where the bird received the recall signal in flight. The feed station training was finally introduced after the birds learned to react to the recall signal.

Subjects. The subjects of the study were 22 racing homing pigeons, all of which were about 6 months old at the time of purchase. These birds had not been loft trained, and they had no previous outdoor flying experience. Their contact with humans had been restricted primarily to times when feeders and watering devices were filled. During the time the birds were used in this investigation, their laboratory home space consisted of an indoor community-type coop and an outdoor flypen.

When hunger is used as the motivating force to obtain learning or to maintain performance in the laboratory, bird weights are usually reduced to between 70 and 80 per cent of normal before experiments are begun. Hunger is further increased by depriving the birds of all food before each test. Birds that are severely deprived of food may not have the required stamina and strength to make long-distance flights without alighting occasionally to rest in places other than at the target. Or, severely deprived birds may return prematurely to the loft. In the present experiment, the birds were maintained at near normal weight levels as much as possible. They were deprived of food only to the extent necessary to provide

motivation for learning.

Conditioned movement. Initially, in the recall training phase, birds were taught to walk between two points on signal. The vehicle used to obtain the conditioned movement was food reinforcement given at both positions. Two test compartments were separated by a 3 ft. long, wire enclosed, runway in such a way that a bird could walk from one into the other. Work panels, containing a key for pecking, a stimulus light, and a recessed food hopper, were placed at each end of the two-compartment configuration. An alternation schedule regulated the on and off time of each work box. After five reinforcements were completed in the first compartment on a continuous schedule (the hopper opened for 3 seconds every time the key was pecked), a bright panel light came on several inches to the side and above the key. The signal light indicated that no more reinforcements could be obtained in that compartment and the other work panel was programmed to operate. After five reinforcements in the second compartment, an interrupted tone sounded to signal a time-out. The bird could then return to the first compartment and begin another work cycle. At first, the alternation did not occur when the light or sound came on, but when the pecking responses resulted in no further reinforcements. With further training, the birds were observed to turn away from a work panel almost immediately after the signal came on, with no more key peck responses. The movement between boxes was relatively swift with little break in stride going through the gates.

Loft training. Until this time, when the birds were about 8 months old, they had never experienced free flight away from the loft. The birds' acquaintance with their loft surroundings was provided by an aviary located above the flypen. From the aviary, the birds could see all sides of their external environment. Now, the gate opening to the outside from the flypen was removed each day after all birds had been fed. After about a week, most of the birds in the colony had experienced flying outside the coop. One bird was eventually lost on a flight made from the coop.

Flight trials in indoor cage space. In this phase, the birds were required to fly from one open-space position to another on cue. Flight trials were begun in a laboratory test area. The test compartments were raised four feet from the floor and placed four feet apart, gate to gate. A landing platform was attached to each box. Trials were conducted within a chicken wire-covered wood frame structure, 4 ft. wide, 6 ft. high and 8 ft. long. Programming was the same as in the preceding phase. By the third day, all but two birds completed their task within the 30 minute session time limit.

Flight trials in outdoor space. The same wire-covered enclosure with work boxes, that was used in the laboratory, was reconstructed out-of-doors. No change was made in the trial procedures. When transfer of training was accomplished to the earlier

level of performance, the enclosure was removed, but the boxes were kept at the same height and distance from each other. On the first session, every bird flew off either before doing any work or after one or several alternations were made between the compartments. Occasionally, the birds returned to work for several reinforcements but left again. The hungry birds also learned to obtain food outside the test situation, because grain could be found easily in the fields surrounding the laboratory. Thereafter, all testing conducted near the laboratory was done in the flypen and under a large net 20 ft. wide by 100 ft. long and 10 ft. high.

Changes in trial conditions and events. Several changes were made in the programmed trial events and procedures at this point. One box rested on the back of a truck and was designated the "home" position for convenience. The other box was placed on a small platform, and it was called the "away" position. The gate was removed from the away box. A time-out period was included between each outflight.

Removal of the gate was part of the transition which could eventually take the away box from the field altogether. The away box was merely a vehicle used initially to obtain flight away from the home position and to provide a place where the signal could be turned on for the return flight. Gate removal at the away box had no significant effect on the birds' behavior. No unusual delays in entering or leaving the box were observed.

Heretofore, it was convenient to have the birds work without pause within a session. More flights were thereby obtained and more birds tested per unit time. The single behavior chain which comprised most sessions was now broken up. To be consistent with the final task procedure, a time-out from reinforcement would need to be programmed for the home position. In the procedure that was now adopted, the gate on the home box was locked at the fifth reinforcement, the bird removed from the box and a wait of 5 seconds made before the bird was hand lofted again into outflight. The interrupted trials procedure initially had disruptive effects on the birds' behavior. When they returned to the home box, they typically interrupted the work schedule there by leaving the box, pausing for a few seconds on the platform, and entering it again. This behavior occurred one or more times until the gate was locked just before the last reinforcement was delivered. After about ten sessions with the new time out and hand lofting procedures, most of the birds again were working well.

Change in recall contingency. The away box had been operating on a continuous reinforcement schedule in which the food tray opened for three seconds every time the key was pecked. After the tray closed on the fifth reinforcement, the recall sounded. Now the number of reinforcements was gradually reduced until the acoustic signal was turned on in the away box by a single key peck.

Until the time that food reinforcement was removed from the away box, the birds were clearly responding to the recall signal. They were observed to emerge from the box at the sound of the signal without testing the key for further reward. As the number of food reinforcements per trial became smaller, more and more key responses were made after the signal onset. When all food reinforcement stopped at the target, the birds showed a tendency to remain at the away box. Usually, they would peck at the key several times, pause, peck again, leave the box, walk back in, peck some more, etc. Almost all the birds stopped working during the first or second session with this procedure, even though they earlier had shown good discriminatory reaction to the auditory signal and food could be obtained at the home position as usual. Evidently, when reinforcement at the away position was removed from the trial events, a breakdown occurred in a strongly chained habit sequence with the result that the birds stopped working. The next ten sessions were spent in strengthening the habits of turning on the signal and returning from outflight. Flight distance was also increased from 5 m. to 30 m.

Demonstration of stimulus control by the acoustic signal. At about this time the key that turned on the signal went out of adjustment slightly. Previously, the switch on the key was so sensitive that only a slight peck on the key was needed to effect closure. Now considerably greater pressure was needed to operate the key. This occurrence accidentally provided an experimental demonstration of stimulus control by the auditory signal.

The behavior displayed by the birds subsequent to the accidental happening in the equipment can be conveniently classified in five categories: (1) A few birds pecked hard enough to bring on the signal with one try, and their behavior showed no change; (2) about 6 birds were observed to peck once at the key without effecting closure, leave the box, pause on the landing platform and then return to peck the key again - these birds did not leave the away box until the auditory signal came on; (3) several other birds continued to peck at the key until the signal sounded - they did not leave the box without the signal; (4) some birds returned to the home position without the signal being sounded, paused, did not enter the home box, and flew back to the away position to peck again at the key - generally once - and entered the away box to peck again at the key there (since no cues, other than the sound of the signal, were available to the bird to inform it when rewards could be obtained at the home position, most birds had now demonstrated that their behavior was under the desired stimulus control); (5) those birds that returned and entered the home cage without the signal. Birds that showed this sort of behavior were for a while not permitted to make another trip until after a 10 second delay. Later these birds were allowed to make as many trips as necessary to turn on the signal and receive reinforcement.

Training and adaptation to humans. The birds had not adapted well to humans even after several months of almost daily handling. It was apparent even before the experiment that the birds were very shy of people. The negligible contact the birds had had with humans during their first 6 months of life probably had a lot to do with their resistance to adaptation. Explicit taming and adaptation procedures were then used to make approach to the human invariably rewarding. This training also involved easy capture of the birds by the human target.

In a confined space, a bird was launched from one person to another over distances ranging from 1 to 20 m. The persons acting as targets were changed daily and used only once for this learning. Each target remained still until the bird landed on him, then he extended his hand containing grain and permitted the bird to feed for about two seconds. After this, the target moved his hand to the open front of a box 10 inches along all sides, and permitted the bird to obtain several more grains, then the hand was closed. The bird then entered the box, which had grain scattered on its floor, and it could eat for about 5 seconds. The bird was removed and another trial begun.

After two weeks of target behavior training, some birds preferred to alight on the ground and walk to the target rather than land directly on the person. After alighting on the ground, a bird didn't always approach the target immediately, and occasionally, the food-containing hand had to be outstretched to entice the bird to come closer.

During the first flight trials under the net, the bird had an unobstructed view of the human target for the entire flight distance of 30 meters. The bird was hand-launched. Once captured in the trap box, the bird was transferred to a community carrying case. When 6 to 8 birds were collected in the carrying case they were returned to the launch point for another round of trials.

Prior to beginning open-field work, the net was removed from the test site and human-target trials were continued there. The target distance began at about 5 meters and increased by ten meter increments with each trial until a distance of 80 meters was achieved. Some birds flew away from the test site, but usually returned to land at the target, especially when another bird was eating there. Recall trials were also given. For the first time, the recall signal was turned on while the bird was in outflight. In several such trials, some birds turned at the sound of the signal without touching down at the away box. Most of the birds though, continued to the box, pecked the key, and returned to the launch point.

Open-field trials. In this final phase of the training program, the birds learned to follow a road, and sometime in outflight, they either found a human target or received a recall signal. All flights were begun with hand lofting from the truck-mounted loft.

When the human target was found at some roadside position, the trial ended there with reward and capture. If no target was found, the bird was recalled to the home box at the truck. There the bird was rewarded, then removed from the box and placed in one of the side cages on the truck. If the bird's flight failed to bring on one of the the rewarding contingencies with a reasonable time period, usually several minutes, another bird was lofted.

On the first day, the flight distance to the human target began at 3 meters and was extended to 70 meters by the end of the session. While the birds were learning road-following behavior during the early field sessions, the away box continued to be used for turning on the recall signal. The novelty of the open field setting obviously had a disruptive effect on the birds' behavior. They spent the first hour mainly in exploring the terrain surrounding the test site. Only 6 birds worked to the full distance of 70 meters, which was attained after 10 trials. Ten other birds flew off before or during the session and did not return to the site. All of the birds exhibited a great deal of high altitude circling behavior, along with frequent flights along the tree line that paralleled the road about 400 meters away. By the end of the day, 7 of the 10 birds that had left the test site had returned to the home loft.

By the third day, the farthest human target distance attempted was 250 meters, and was reached by three birds in 10 trials. Flight behavior was less variable by this day, and birds that flew away from the test area did so mainly near the end of the session. The same site was used as on the first two days.

On the fourth day, testing was moved to a black-top road with a high tree line close to each side. The new site appeared well-suited for obtaining road-following behavior. It was assumed that the tree line would keep the birds in bounds along the road. More straight-line flights were obtained on this day, although some circling did occur at altitudes above the tree line. Several birds reached the target positioned at a distance of 200 meters.

Eleven birds had been lost in the first 4 days of open field training. To reduce further losses, the conditions were simplified. Only human target trials were run for several sessions and the targets were stationed no more than 50 meters from the launch point.

The birds had now begun to make straight-line flights consistently to the target, though it couldn't yet be determined that road-following behavior <u>per se</u> was taking place. There was a sharp decrease in incidental circling and general exploratory behavior.

By the ninth day of field work, the flight distance was extended to 300 meters. Recall trials were now programmed with target trials on a random basis. A box was no longer used in the recall trials to obtain the signal, which now was sounded while a bird was

in outflight. Once turned on, the signal remained on until after the bird returned to the starting position, entered the home cage, and received its reward.

The birds at first tended to ignore the recall signal and to continue in outflight after the sound came on. After repeated trials, however, some birds began to show a sharp turning response a moment after the signal came on.

Daily change of test site. The test site was changed daily beginning during the 4th week of open-field trials. The position of the most distant target found on any test day usually became the launch point for the following day's trials, with flights continued in the same direction along the road. The principal cues for directionality of flight were the direction in which the birds were lofted from the launch point, and the position of the truck in relation to the road. Targets were positioned several feet off the road in the brushline. Each target was hidden to an extent that made detection easy only when a bird had flown to within about 30 meters of the target. Recall trials were not allowed to exceed 100 meters in distance. A marine horn that could be heard beyond a half mile was used to give the recall signal. Its effective range for a bird in flight may have been somewhat less than 1/2 mile.

The greatest flight distances achieved by all birds on each day for the 8 days were 300 m, 300 m, 400 m, 600 m, 700 m, 900 m, 1100 m and 1200 m respectively. The starting distance for each session was usually 200 to 300 m. The distance was increased for every trial in varied increments until the final distance was reached by about the tenth trial. Only 1 bird was lost in the last 16 sessions.

Behavioral anomalies. The behavior of two of the birds in target trials during this phase differed from that of the other birds in a predictable way. One bird, number 57, continued to be reluctant to approach humans. Typically, it made several round trips to the target position without landing at the target or anywhere else. The bird continued to fly until another bird was released, whereupon no. 57 joined the second bird in flight to the target, landed with it to be fed, and trapped. Another bird, number 65, frequently flew only a short distance, perched on an overhead telephone cable, and waited until another bird came along. It joined the second bird in flight to the target, landed and approached the target normally.

The random distribution of target and recall trials posed a problem for the birds. Only if a continuous series of recall trials was run, would the birds begin to turn sharply at the first sound of the recall signal. In a continuous series of target trials the birds readily made long outflights. A random mixing of the two types of trials invariably resulted in short outflights on the target trials. Only after the bird had made one or more such short round trips,

would it then make a longer one to the target position.

Flights in the wrong direction along the road were made in about 10 to 20 per cent of all trials per session. These wrong-way flights usually occurred on the first trial or two. A bird would fly about 20 m, turn, then fly past the truck in the wrong direction, and almost always go out of sight. On returning from wrong-way flights, the birds frequently went on to the target without landing at the truck.

Orientation of the roadway. Two sessions were conducted on a road that ran at a right angle to the one on which a number of sessions had just been completed. The new road had a treeline close to one side and another running parallel to it and about 100 m away. In initial trials some birds made repeated flights toward the distant treeline and back to the truck loft. Perhaps significantly, the compass direction of these outflights was the same as that of the road that had been used during the preceding several sessions. Recall trials were not attempted because of the difficulty experienced in obtaining good road-following behavior. The most distant target position reached in the 1st session was 200 m. Birds that more readily flew toward the target would lead a confused bird in the proper direction. Thus, in most of the successful trials - those that ended at the target - 2 or more birds flew together. The same kind of behavior occurred during the 2nd session on this road. Flights of 300 m were obtained in the 2nd session.

Social facilitation of learning. At this time a bird (no. 7), that had left the training site on the first day of open field trials, returned to the home loft. This bird was one of the shyest of the group, never having become well adapted to people. It was also a bird that had found ways of escaping from under the net occasionally during intermediate training. In general this bird had shown little promise for the desired kind of conditioned behavior. Now, of course, because of its long absence, no. 7 was relatively naive compared with the other birds. Partly because of the shortage of birds, no. 7 was returned to an active training status.

On the day following its return, no. 7 was given several trials, not exceeding 10 m. overall, both with a target and without a target (recall signal). One of the more experienced birds was lofted with no. 7 for each trial. On each trial no. 7 performed surprisingly well.

On the next day, trials with no. 7 achieved a distance of 50 m. to the target. A more sophisticated bird again accompanied no. 7 on each flight. In recall trials, when the experienced bird turned sharply at the sound of the recall signal, no. 7, which was following closely, also turned. In target trials, the 2 birds appeared to compete for food at the target. After the 2nd day, no. 7 received the same training as the other birds.

Social facilitation and performance in unfamiliar terrain. Following reintroduction of no. 7 to the training program, all birds were tested on a road that had not been used before. The road was about 1300 m in length and it was bordered on either side by high, dense foliage. The birds were lofted in pairs. Performance in both target and recall trials was markedly superior to that of single birds flying over an unfamiliar road. Launch-to-target distance of 800 m was achieved.

Performance on a road that had been used earlier. It had become increasingly difficult to find suitable lengths of available road within Aberdeen Proving Ground that had not been used before. Finally, a 3½ mile stretch of road that had been used in earlier trials was used again for the next 3 sessions. The direction of flight in these sessions was the reverse of that of the earlier trials along this road.

Initially, the birds showed a tendency to fly the wrong way with respect to the direction of the target from the launch point. In the first session the birds would fly about 400 m in the wrong direction to a large body of water at which the road ended. Upon reaching the water, the birds then retraced their flight path and flew past the truck in the direction of the target. Paired birds, lofted together, usually kept within about 10 m of each other in flight. The first day's maximum launch-to-target distance was 600 m and was reached by all birds.

On the second day, the birds had flown a mile to the target by the end of the 6th trial. Their flight speed was measured on the last two trials. The birds were clocked at between 30 and 35 miles per hour. On the third day, the first trial launch-to-target distance was ½-mile; the distance was increased in ½-mile increments until 1½ miles was reached in 4 trials. During the three sessions, all birds consistently showed good road-following performance, and no undesirable behavior was observed. Bird no. 7 followed the other test birds to the target on every trial.

Performance on an unfamiliar road surrounded by large open fields. The final training sessions were conducted along a 3-mile length of road through grassy fields several hundred m. wide on each side. The terrain thus was different from that of any other site that had until now been considered suitable for training purposes. The initial launch-to-target distance was 100 m. The target was hidden just inside the grass line off the road. It could probably be seen by the birds from a distance of about 30 m. Unless otherwise noted, the birds were flown in pairs. By the end of the first session the launch-to-target distance had been increased to 600 m. The birds again tended to fly in the opposite direction from the target immediately after being lofted. Their flight behavior was essentially similar to that observed in the preceding sessions. For the next 5 sessions, launch-to-target distances of 800 m, 1000 m, 1200 m, 1300 m,

and 1200 m, respectively, were achieved. A new stretch of the road was used for each session. The first trial of a session usually began with a launch-to-target distance of 400 m. In the last session, the birds were lofted singly rather than in pairs, and only one trial was conducted. Bird no. 7 flew to the target, but did not land there. It turned around, flew toward the truck, and then flew out of sight. Several minutes later, another bird was lofted. Somewhere in flight, it was joined by no. 7 and both birds then landed at the target. One other bird lofted singly did not reach the target on the first try, but did later after joining another bird in flight. The remaining birds showed good target-reaching behavior.

At the end of the 3-mile distance, three sessions were conducted over the same road in the reverse direction. The launch-to-target distance in all trials was one mile. In the first session, the birds were flown in pairs in each of three trials. All birds reached the target in each trial. In one trial, a pair of birds joined another pair in flight after one of the pairs had flown away from the site. Road-following behavior in general was excellent. One pair landed on the road about half-way to the target, remained there for about 30 seconds, then resumed flight to the target. Single bird flights were made in the second session. Bird no. 7 reached the target in each of two trials and landed on the road near the target. This bird did not approach the target, however, until another bird joined it. In other trials, 2 birds lofted singly reached the target only after each joined another bird in flight. On the third session, all birds (7) were released at the same time and all 7 flew directly to the target.

DISCUSSION

Some significant problems relating to the organismic and environmental conditions and events that play a part in obtaining long distance search flights by birds have now been studied experimentally. The results of the present investigation leave no doubt that a major problem with respect to using pigeons as subjects is obtaining reliable flight behavior in unfamiliar surroundings.

The birds in the present study performed well even in "unfamiliar" terrain in both target and no-target trials when a series of either one or the other kind was given.* When flight procedures were randomly intermixed in a series of trials, the birds tended to

*The terrain was judged to be unfamiliar at a given training site if training had not been conducted at that specific site before. Since all training was conducted within the confines of a portion of Aberdeen Proving Ground, and all birds at one time or another had flown away from their training and/or home loft, the possibility must be considered that the overall area was in a general way familiar to the birds.

carry over their responses from one kind of trial to the other kind - that is, changes in procedure were usually not accompanied by an immediate shift in the appropriate behavior. The dual procedures were discontinued before an adequate test of their combined use could be made. In most out-of-door field trials the target procedure was used exclusively to obtain information on whether birds can be guided into long distance flights in "strange" terriroty. For these trials, the targets were hidden only to the extent that the birds did not know where the targets were until they had flown some distance along the road.

The birds clearly generalized their behavior to the several roads used. Had the study continued, the number of trials needed to reach an acceptable level of performance would undoubtedly have become less with each new "unfamiliar" road. The results suggest that the birds did learn the general idea of "road" and, with enough experience, might have responded correctly to this terrain feature in whatever new and novel setting it appeared, with perhaps little or no adaptation. Obtaining the required distance then would be mainly a function of the degree of motivation.

There is little assurance that good control over performance developed in a restricted field can readily be transferred to an open-field environment. The great variation of stimuli found in the open-field space is accompanied by more response alternatives. Thus, a variety of uncontrolled and undesirable responses may occur randomly.

It was discovered that birds performed better when they were accompanied in flight by one or more other birds. The term "social facilitation" has been used to describe motivated behavior enhanced by social stimulation. The acquisition of a new response brought on by association with another animal engaged in the same behavior is also called "imitative learning." One bird in the group used in this study adapted to open field controlled flight in about one-third the usual time, when it was accompanied on every outflight by a more experienced bird.

Competition for food may be a major part of the social facilitation effect. There have been several previous reports that social factors can produce increments of individual activity. Hake and Laws in a recent study found that behavior suppressed by aversive stimulation can be increased by the presence of a second animal. The same principle seems to apply to the behavior of birds in unfamiliar surroundings.

The successful use of older birds in the present study indicates that the age of a bird, at least within fairly wide limits, is not critical for loft learning or the acquisition of new habits. The resistance to taming that was experienced may have been associated with early lack of contact with humans.

Strong motivation for learning was established while maintaining body weight at or near normal levels. An effective motivating level of hunger was achieved when the birds were deprived of food for 48 hrs. preceding each training session. Temporary weight loss occurred during periods of new learning, but as performance improved, the losses were usually regained within several sessions.

Automated procedures can be used even in open-field training. An 18-position dovecote was designed to permit automation of field training procedures. The dovecote was arranged in three levels with six compartments in each level. Each compartment had two gates, one through which the bird could exit, and another to enter. A feeder was arranged so that a predetermined amount of grain could be dropped into a cup in each unit. Continuous access was provided to a watering device. A sequence of events could be programmed to begin in the first compartment, and continue in succession automatically, one compartment at a time, until all 18 compartments had participated.

CONCLUSIONS

Indications are that the flight behavior of pigeons in unfamiliar surroundings can be progressively modified toward an acceptable level of performance by repetitive trials in which the test site is changed daily.

Training pigeons to perform a required task can be expedited by exploiting the phenomenon of social facilitation, i.e., by pairing an experienced bird and a naive bird in training. Even experienced birds work better in pairs than they do singly.

Unwanted behaviors can easily be developed inadvertently through accidental chaining. Close and precise control of conditioning stimuli must be maintained throughout training. Unwanted behaviors, accidentally chained, can have disruptive effects on overall performance, and may be difficult to eliminate.

Birds as old as 6 months, with little prior contact with humans, can be trained to perform all tasks required in a human detection system.

Birds need not be maintained at reduced body weight in order to perform training tasks. Sufficient motivation can be provided by 24 to 48 hrs. deprivation before each training session.

BLAST PHENOMENA FROM EXPLOSIONS AT THE WATER SURFACE

A. SAKURAI, J. M. PINKSTON, JR.,* AND J. N. STRANGE
NUCLEAR WEAPONS EFFECTS DIVISION
U. S. ARMY ENGINEER WATERWAYS EXPERIMENT STATION
VICKSBURG, MISSISSIPPI

Introduction

The principal objective of this study was to evaluate the blast and shock phenomena (in air and underwater) resulting from a pure surface explosion (i.e. the situation where the charge is half in and half out of the water).

The particular case of an explosion at an air-liquid (water, in most cases) interface is considered in this paper because this case is easier to handle both analytically and experimentally than an explosion at an air-solid interface, and thus permits one to compare more readily analytical and experimental results. Moreover, it should be noted that solids under sufficiently strong impulsive loads are known to behave much as liquids. The general results of the air-liquid interface case can be utilized for any air-solid case by substituting the appropriate material constants to represent the liquid properties of the solid under consideration. The problem of predicting surface burst phenomena involves particular difficulties when compared with the so-called free-space cases in which the explosion occurs entirely within a given surrounding medium. The most serious difficulty arises from the fact that the general characteristics of the shock waves vary depending on the nature or the type of the explosion, since the majority of the features of the shock waves are controlled by the initial stages, where the characteristics of the individual explosion predominate. This situation makes it difficult to establish any simple scaling rule independent of the individual nature of the explosives used. For this reason it is usually difficult to deduce anything of a general nature from the results of individual tests or step-by-step numerical solutions starting from specified initial input conditions. A less serious difficulty is found in the magnitude of the density ratio of the two interface media involved. Although this ratio for air to water is much less than 1, it is not small enough to warrant neglecting it. Thus, the motions in the different media cannot be treated separately, as in the case of air to ground.

Aside from existing literature on the air-ground cases, only a few previous studies on the air-water case are available for either experimental or theoretical investigations (references 1 through 9). All these data are based on rather small chemical explosions, and the relation of these results to nuclear burst cases is very uncertain, especially for this particular configuration (i.e. a spherical charge half submerged in water).

This situation thus implies a definite need for a thorough study of the problem, especially for nuclear burst cases. It is essential to study the initial and later stages separately, so that the proper approaches can be sought for handling each of these different stages, whose characteristic features are depicted in fig. 1.

As mentioned earlier, there may be no general solution that will represent all the initial stages caused by different kinds of explosions, such as chemical, nuclear, electric spark, etc., so that they must be studied individually. On the other hand, some of the important characteristics of the flow fields in the later stages can be expressed in a general term regardless of the nature of the explosions, mainly because only a small interaction exists between the flow fields in water and those in air regions above the water; thus they can be formulated separately in the following manner. A linear acoustic approximation is used to satisfy the water environment, while a modified free airblast approximation is used to express the flow field in air. This model leads to a set of air and water shock pressure formulas, which contain some unknown parameters and functions to be determined from the characteristics of the initial stages, depending on the type of explosive.

Two different types of initial stages were considered: a point source model of a finite amount of mechanical energy released instantaneously at a point on the interface assuming this to be realistic for the wide varieties of nuclear explosions; and a general chemical explosion of bare, spherical HE.

These two cases were utilized for determination of the unknown parameters and functions appearing in the pressure formulas in the later stages. The pressure values given by the formulas thus obtained for the HE cases were compared with the test data from the explosion of forty 10-pound spherical TNT charges conducted simultaneously with the theoretical investigation. The comparisons of the theoretical values with experimental data and with test data available from charge weights as high as 10,000 pounds show good agreement.

Solution of Later Stage

General characteristics of the flow based on acoustic theory

Assume here linear acoustic properties throughout the entire space. This assumption is obviously not valid near the explosion source, especially in the air environment. Nevertheless this serves to give an overall insight into the understandings of the problem

which can thus lead to the development of modifications necessary to obtain more realistic solutions.

As illustrated in fig. 2, designate $p(r,z,t)$ and $p'(r,z,t)$ as the water and air overpressures at the time t. Let r and z be the cylindrical coordinates; p_o, ρ_o, and c_o are the pressure, the density, and the sound velocity in the undisturbed water region, and $\bar{p}_o$, $\bar{\rho}_o$, and $\bar{c}_o$ are the corresponding values in air.

General solutions for p and p', expressed in the Laplace-Bessel integral forms, are fitted together at the interface to satisfy the boundary conditions, and then are simplified, by utilizing the small value of the density ratio ε^2 $(\varepsilon^2 \equiv \bar{\rho}_o/\rho_o)$, to yield

$$p(r,z,t) = \frac{z}{q^3} F_o\left(t - \frac{q}{c_o}\right) H\left(t - \frac{q}{c_o}\right) + p_a \qquad (1)$$

$$p'(r,z,t) = \frac{1}{q} \bar{\rho}_o W_o\left(t - \frac{q}{\bar{c}_o}\right) H\left(t - \frac{q}{\bar{c}_o}\right) \qquad (2)$$

where $q = \sqrt{r^2 + z^2}$, F_o and W_o are the input functions related to the singularities of $p - p'$ and $\partial/\partial t(v' - v)$ at the center. Here v and v' are the z components of the velocities in the water and air regions, respectively. The input functions F_o and W_o are given as

$$\left.\begin{aligned} F_o(t) &\equiv \int_0^\infty (p' - p)_{z=0}\, r\,dr \\ W_o(t) &\equiv \int_0^\infty \frac{\partial}{\partial t} (v' - v)_{z=0}\, r\,dr \end{aligned}\right\} \qquad (3)$$

H is the Heaviside step function, and p_a represents the term corresponding to the water pressure induced by the air pressure p' given by Equation 2. Equations 1 and 2 exhibit the following important general characteristics of the flow field. First, the airblast field p' is represented by a spherical wave from a point source; second, the water shock p consists of two parts: one caused by a doublet source and the second one, p_a, induced by the air pressure. While these qualitative descriptions of the flow field are generally true, since they are based primarily on the small value of ε, it is only necessary to modify Equations 1 and 2 so that the resulting formulas can be used.

Modification of the formulas

First, Equation 2 for the air pressure p' should be replaced by the equivalent free airblast characteristics caused by an explosion in free air with an equivalent charge weight W_{ef} which is larger than the actual charge weight W, and the ratio β defined by

$$\beta \equiv \left(\frac{W}{W_{ef}}\right)^{\frac{1}{3}} \tag{4}$$

Here β is a parameter that depends on the nature of the explosion. Second, the airblast field is not exactly in spherical symmetry but should be distorted due to the interaction effect with the water shock field. Since this distortion is rather small, this can be incorporated in the formula by introducing a parameter S, which will be called the "Skewness factor." With use of these parameters β and S, the characteristics of the air pressure p' such as its peak pressure $P(r,z)$ and its duration $D(r,z)$ at (r,z) are determined from the corresponding free-air peak pressure $P_f(\lambda_e)$ and the duration $D_f(\lambda_e)$ at an equivalent reduced distance λ_e as

$$P(r,z) = \left(\frac{1 + S\cos\Theta}{1 + S}\right)^2 P_f(\lambda_e) \tag{5}$$

$$D(r,z) = \beta D_f(\lambda_e) \tag{6}$$

where $\tan\Theta = \frac{r}{z}\left(0 \leq \Theta \leq \frac{\pi}{2}\right)$

$$\lambda_e = \beta q W^{-\frac{1}{3}} \frac{1 + S}{1 + S\cos\Theta} \tag{7}$$

Using this airblast pressure on the interface as an input, the air-induced pressure term p_a can be estimated numerically by utilizing the scheme given in reference 10; here alternatively the following approximation formula is developed by introducing various minor assumptions

$$p_a(r,z,t) = \frac{U_o}{m'_o c_o} \frac{1}{q}\left(\frac{z}{q}\right)^{\nu} \bar{\rho}_o W_o(t - \tilde{t})H(t - \tilde{t}) \tag{8}$$

where U_o is the initial airblast velocity, m'_o is its initial decay rate, ν is a parameter whose value is about 0.1, and $\tilde{t}(\equiv\tilde{t}(r,z))$ is the arrival time of the wave at a position (r,z). The resulting pressure $p(r,z,t)$ consists of two waves represented by the two terms in Equation 1 complemented by Equation 8. In principle, the second wave induced by the airblast propagates ahead of the direct wave of the first term. But the difference is very small for most of the cases (is zero for $r = 0$) and can be noticeable only in the region close to the surface, where the pressure-time histories of test data indeed show two peaks. It is also noticed that the first term is proportional to $z \cdot q^{-3}$ while the second is roughly proportional to q^{-1}. Thus the pressure is dominated by the second term with increasing distance q. But the retaining of the first term is indispensable to represent the importance of the direct

effects near the source. Now, the input functions $F_o(t)$ and $W_o(t)$ can be estimated approximately as

$$\left.\begin{aligned} F_o(t) &= \left(P_s R_s - P'_s R'_s\right)_{r=0} \cdot a \\ \overline{\rho}_o W_o(t) &= \left(m' p'_s - \varepsilon^2 m P_s\right)_{r=0} \cdot a \end{aligned}\right\} \qquad (9)$$

where P_s and R_s are the water shock front pressure and the shock front position, P'_s and R'_s are the corresponding values for the air shock front, m and m' are the decay rates of the water and air shock velocities, and a is the effective radius of the spread of these inputs.

Further details of the pressure formulas depend on the nature of an explosion through the determination of those constants and functions introduced above.

Initial Stage and Pressure Formulas

Bare, spherical chemical explosive

Consider here the case of the spherical chemical explosive whose center is positioned at the water surface. The magnitude of a in this case can simply be

$$a = a_o \equiv \text{"the charge radius"}$$

and one may postulate that

$$\left.\begin{aligned} \left(P_s R_s - P'_s R'_s\right)_{r=0} &= \left[\left(P_s\right)_{r=a_o} - \left(P'_s\right)_{r=a_o}\right] a_o \\ \left(m' P'_s - \varepsilon^2 m P_s\right)_{r=0} &= \left(m' P'_s\right)_{r=a_o} - \varepsilon^2 \left(m P_s\right)_{r=a_o} \end{aligned}\right\} \qquad (10)$$

Now those pressure values at $r = a_o$ can be computed using the free airblast and free-water shock data in the following manner

$$\left.\begin{aligned} \left(P_s\right)_{r=a_o} &= P_{so} e^{-t/D_{so}} \\ \left(P'_s\right)_{r=a_o} &= P'_{so} e^{-t/D'_{so}} \end{aligned}\right\} \qquad (11)$$

where P_{so} and D_{so} are, respectively, the peak overpressure and the duration of the free-water shock at the charge surface, and P'_{so} and D'_{so} are the corresponding values for the free airblast. In practice, further simplifications can be introduced because of the fact that $P_{so} \gg P'_{so}$ and $P'_{so} \gg \varepsilon^2 P'_{so}$, and one will get finally from Equations 9, 10, and 11

$$\left.\begin{aligned} F_o(t) &= P_{so} a_o^2 e^{-t/D_{so}} \\ \bar{\rho}_o W_o(t) &= m' P'_{so} a_o e^{-t/D'_{so}} \end{aligned}\right\} \qquad (12)$$

Substituting Equation 12 into Equations 1 and 8, and utilizing the free air and free-water shock values at the charge surface (references 11 and 12), one obtains the following formula for the water shock:

$$p(\lambda_r,\lambda_z,\tau) = 8.4 \times 10^3 \frac{\lambda_z}{\lambda^3} e^{-3.9(\tau-\tau')} H(\tau-\tau') + 1.1 \times 10^4 \frac{1}{\lambda}\left(\frac{\lambda_z}{\lambda}\right)^{0.1} e^{-26\tau} H(\tau) \text{ (psi)} \qquad (13)$$

where

$$\lambda = qW^{-\frac{1}{3}}, \quad \lambda_r = rW^{-\frac{1}{3}}, \quad \lambda_z = zW^{-\frac{1}{3}} \left(\text{ft-lb}^{-\frac{1}{3}}\right)$$

$$\tau = (t - \tilde{t})W^{-\frac{1}{3}}, \quad \tau' = (q/c_o - \tilde{\tau})W^{-\frac{1}{3}} \left(\text{msec-lb}^{-\frac{1}{3}}\right)$$

Regarding the parameters β and S in the airblast formulas (Equations 5 and 6), the value of β can be determined from the energy consideration ($\beta = 0.87$ in this case). The parameter S is more or less related to the overall geometry of the shock wave, and there is no clear way to estimate its magnitude beyond the fact that it is small. Thus $S = 0.15$ was found to give a better fit to the test data, although its variation is not sensitive to the resulting pressure values.

Those pressure formulas derived in the case represented by Equations 5, 6, and 7 with $\beta = 0.87$, $S = 0.15$ for the airblast, and Equation 13 for the water shock are plotted in figs. 3-5 and compared with various test data (references 1 through 7, 13, and 14). The good agreement between the theoretical values and the test data in these figures establishes the usefulness of these formulas for spherical HE charges. This agreement also indicates the overall validity of the various assumptions introduced in the process of analysis in the previous section. This fact is important since this gives some assurance of reliability for the general formulas given by Equations 5 through 9, thereby establishing the validity of the formulas in the next section for the nuclear explosion cases, for which any direct verification is not feasible.

Point source model

With particular attention to the case of stronger

explosions, consider here their initial stages, as described by a similarity solution, which can be derived as a result of the following three major assumptions.

First, a point source input is assumed in the sense that a finite amount of mechanical energy is released instantaneously at a point on the air-water interface.

Second, the pressures at the shock fronts (in air and water) are assumed to be much higher than the ambient pressure in the undisturbed region, so that the latter quantity is negligible.

Last, it is necessary for the similarity relation to assume that both air and water behave as polytropic substances, or in other words, the product of (pressure) $\times$ (density)$^{-\gamma}$ is a function of entropy only, where the exponent γ takes different values for different media. The validity of using this assumption in the airblast region has been well established. However, its use for the water environment in connection with the second assumption above should be regarded with caution. This fact is important to recognize when taking into account the difference between the initial stages in weak and intense explosions.

In any case, the assumptions above lead to a similarity of flow field and reduce the basic system of equations to a simpler system by use of similarity variables, and furthermore, the similarity makes it possible to determine a definite energy partitioning ratio of the airblast and water shock in terms of the released energy, which is a constant in the solution considered herein.

By utilizing the small value of ε, a successive approximation scheme for the solution of the similarity equations was developed and the system of equations for the first approximation solution was found numerically. This provides an almost spherically symmetric solution (reference 15) for the airblast region with its distortion in the order of ε, and a very oblate water shock front, which is almost a plane wave parallel and close to the intersection surface. This solution was used to compute the energy ratio, which gives $\beta = 0.79$ and $S = 4 \times 10^{-3}$.

The input functions F_o and W_o of Equation 9 are found (for this approximation) from

$$F_o(t) = P'_s(R'_s)R'^2_s$$

$$\bar{\rho}_o W_o(t) = \frac{3}{2} P'_s(R'_s)R'_s$$

Since the function $R'_s(t)$ as well as the pressure $P'_s(R'_s)$ are available from the free airblast data, these can be utilized here.

Conclusions

General pressure formulas for the region not so close to the explosion source were derived and their output calculations were compared with HE test data; the comparison showed very good

agreement except for a small region on each side of the interface, wherein both measurements and formulas appear less reliable. A point source model was also developed to furnish the necessary initial conditions to the general formulas for nuclear explosion cases.

References

1. Niffenegger, C. R. and Heathcote, T. B., "Shockwave Propagation in Shallow Water, III - Pressures, Durations, Impulses, and Craters from 100-lb TNT Charges (U)," NAVORD 6077, Feb 1958, U. S. Naval Ordnance Laboratory, Silver Spring, Md., CONFIDENTIAL.
2. "Explosion Effects Data Sheets," NAVORD 2986 (superseded by NOLTR 65-218), June 1955, U. S. Naval Ordnance Laboratory, Silver Spring, Md., CONFIDENTIAL.
3. Strange, J. N. and Halper, S. H., "A Quantitative Evaluation of the Underwater Shock Wave Resulting from Surface and Underwater Explosions (U)," Technical Report No. 2-615, Feb 1963, U. S. Army Engineer Waterways Experiment Station, CE, Vicksburg, Miss., CONFIDENTIAL.
4. Pittman, J. F., "Air Blast Field Above TNT Spheres Fired Half Immersed in a Water Surface (U)," NAVORD 6801, June 1960, U. S. Naval Ordnance Laboratory, Silver Spring, Md., CONFIDENTIAL.
5. Strange, J. N., "Air Blast Along the Surface from Surface Bursts," Sept 1961, U. S. Army Engineer Waterways Experiment Station, CE, Vicksburg, Miss., CONFIDENTIAL. (Unpublished paper prepared for internal use.)
6. Belliveau, L. J., Rogers, W. C., Jr., and Lehto, D. L., "The Prediction of the Shock Overpressures from Surface-Burst Nuclear Weapons Along the Surface (U)," NAVORD 6240 (DASA-1139), Jan 1959, U. S. Naval Ordnance Laboratory, Silver Spring, Md., CONFIDENTIAL.
7. Pearce, N. and Potter, R., "Measurements of the Air Blast from a Charge Detonated near a Water Surface," Foreign Weapon Effects Report No. AWRE H15/51, FWE-49, Sept 1955, Atomic Energy Commission, Washington, D. C., FOR OFFICIAL USE ONLY.
8. Christian, E. A. and Rosenbaum, J. H., "On the Oblique Reflection of Underwater Shockwaves from a Free Surface; Part IV, Charges at the Surface," NAVORD 3680, Apr 1954, U. S. Naval Ordnance Laboratory, Silver Spring, Md., CONFIDENTIAL.
9. Collins, R., "Intense Explosions at the Ocean Surface," Report No. AS-66-9, Aug 1966, College of Engineering, University of California, Berkeley, Calif.
10. Sakurai, A. and Pinkston, J. M., "Water Shock Waves Resulting from Explosions Above an Air-Water Interface; Results of a Theoretical Investigation," Technical Report No. 1-771, Report 1, Apr 1967, U. S. Army Engineer Waterways Experiment Station, CE, Vicksburg, Miss.
11. Moulton, J. F., Jr., "Nuclear Weapons Blast Phenomena," DASA-1200, Vol 1, Mar 1960, U. S. Naval Ordnance Laboratory, Silver Spring, Md., SECRET.

12. Cole, R. H., Underwater Explosions, Princeton University Press, Princeton, N. J., 1948.
13. Strange, J. N. and Miller, L., "Blast Phenomena from Explosions at an Air-Water Interface (U); A Preliminary Study of Water Shock Waves Directly Beneath a Surface Explosion (U)," Miscellaneous Paper No. 1-814, Report 1, June 1966, U. S. Army Engineer Waterways Experiment Station, CE, Vicksburg, Miss.
14. Strange, J. N. and Miller, L., "Blast Phenomena from Explosions at the Air-Water Interface, Small-Scale Test Results," U. S. Army Engineer Waterways Experiment Station, CE, Vicksburg, Miss. (Unpublished.)
15. Miller, L., "Water Shock Pressures Resulting from 10,000-lb TNT Tests in Mono Lake," Dec 1967, U. S. Army Engineer Waterways Experiment Station, CE, Vicksburg, Miss. (Unpublished paper prepared for internal use.)
16. Sakurai, A., "Blast Wave Theory," Basic Developments in Fluid Dynamics, Vol 1, Academic Press, New York, 1965, pp 309-375.

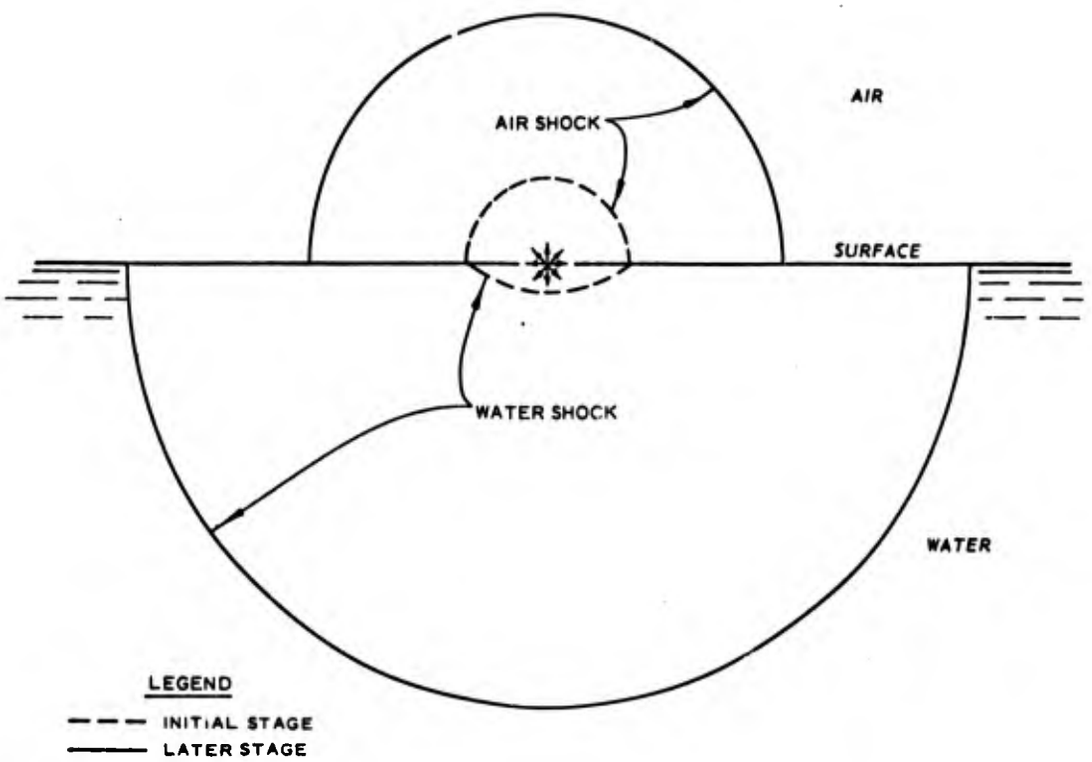


Fig. 1. Characteristics of initial and later stages resulting from explosion at water surface (charge half submerged)

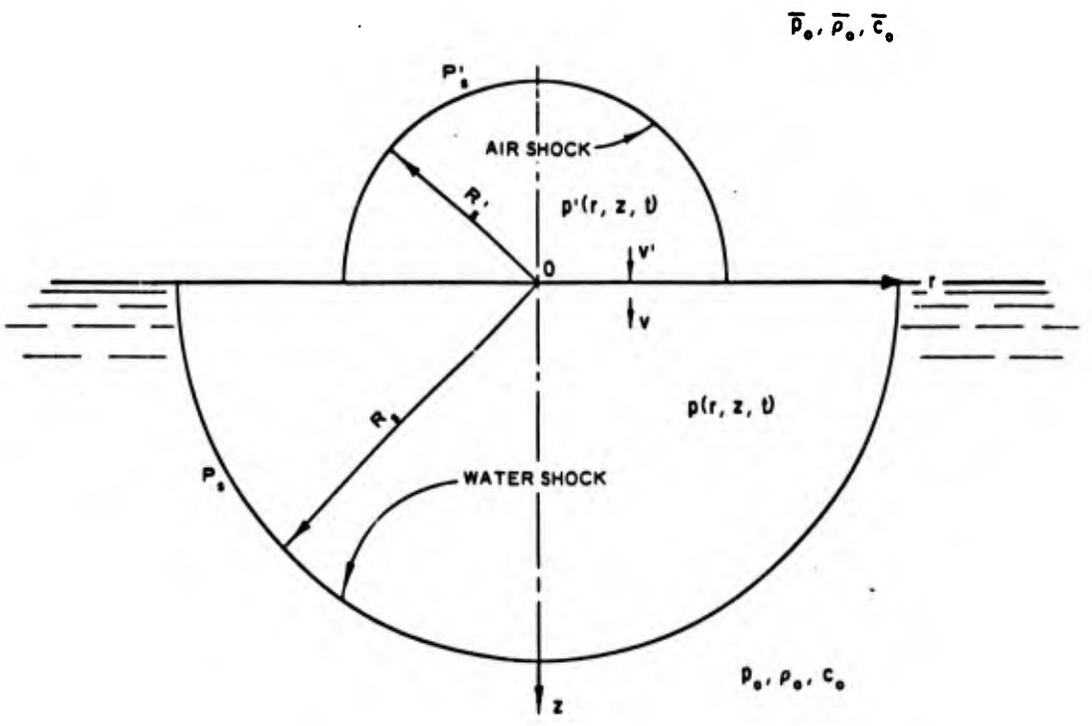


Fig. 2. Illustration of coordinate system and notation

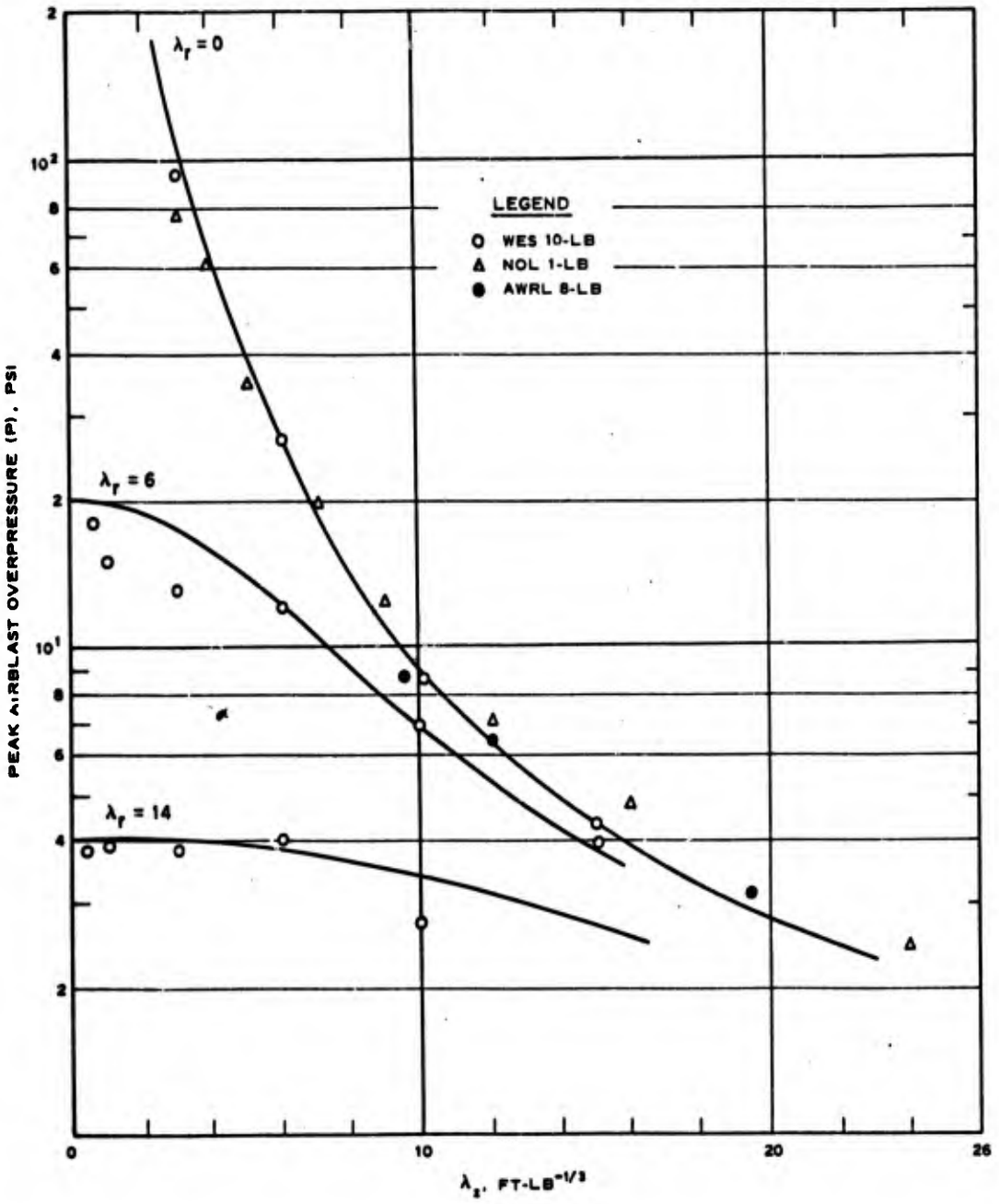


Fig. 3. Peak airblast overpressures from Equation 5 compared with test data; λ_r = 0, 6, and 14 ft-lb$^{1/3}$

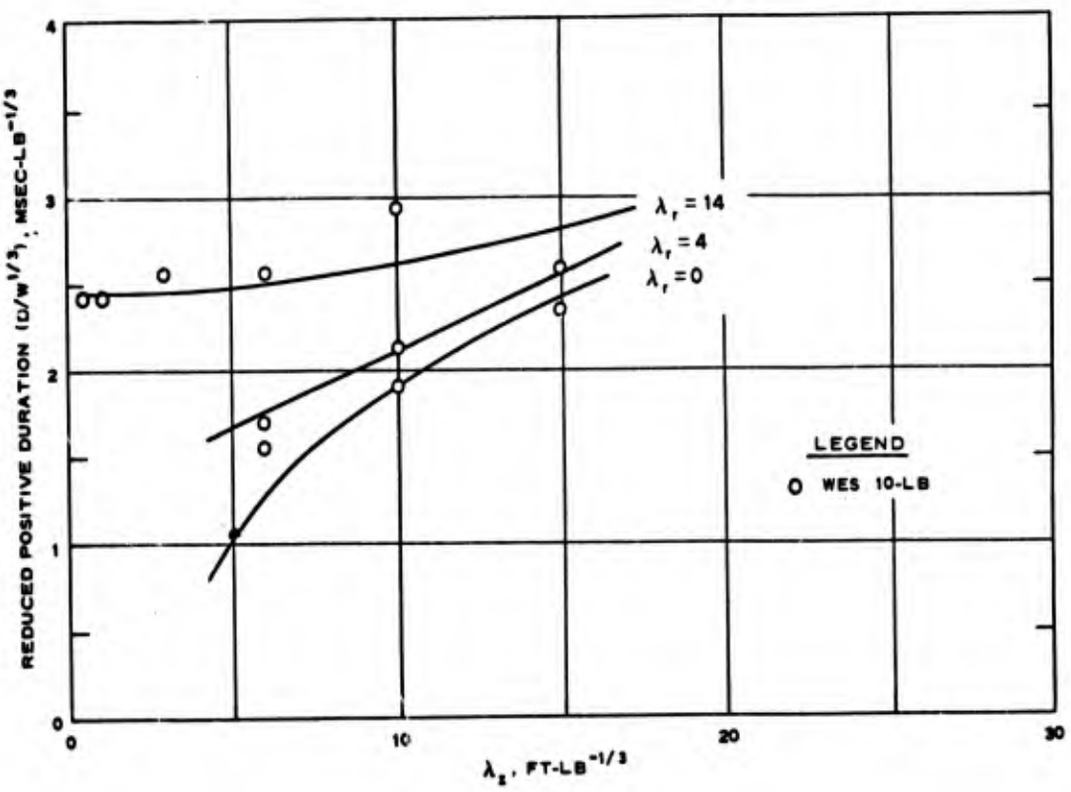


Fig. 4. Reduced positive duration from Equation 6 compared with test data; λ_r = 0, 4, and 14 ft-lb$^{-1/3}$

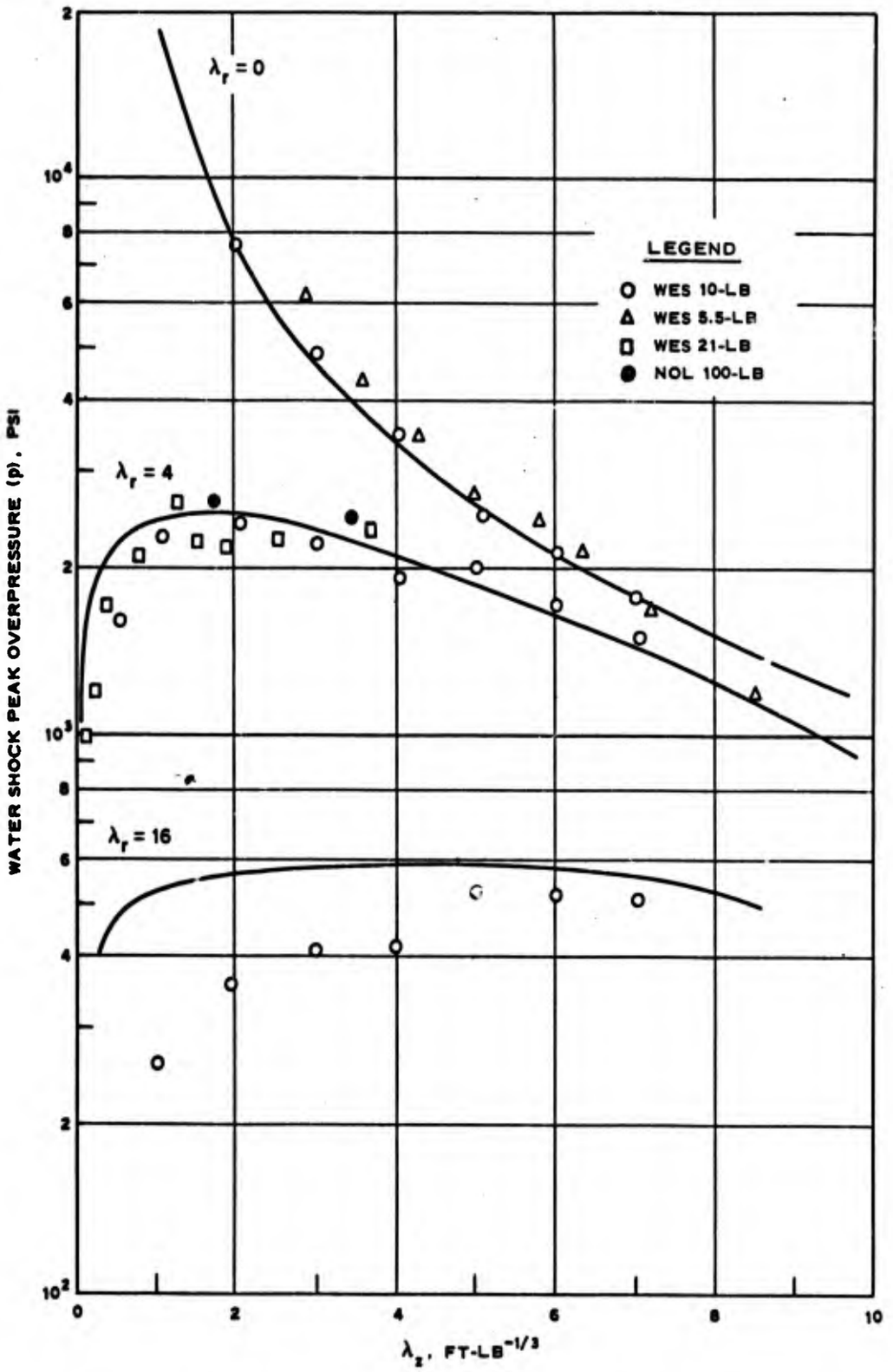


Fig. 5. Water shock peak overpressure from Equation 13 compared with test data; λ_r = 0, 4, and 16 ft-lb$^{-1/3}$

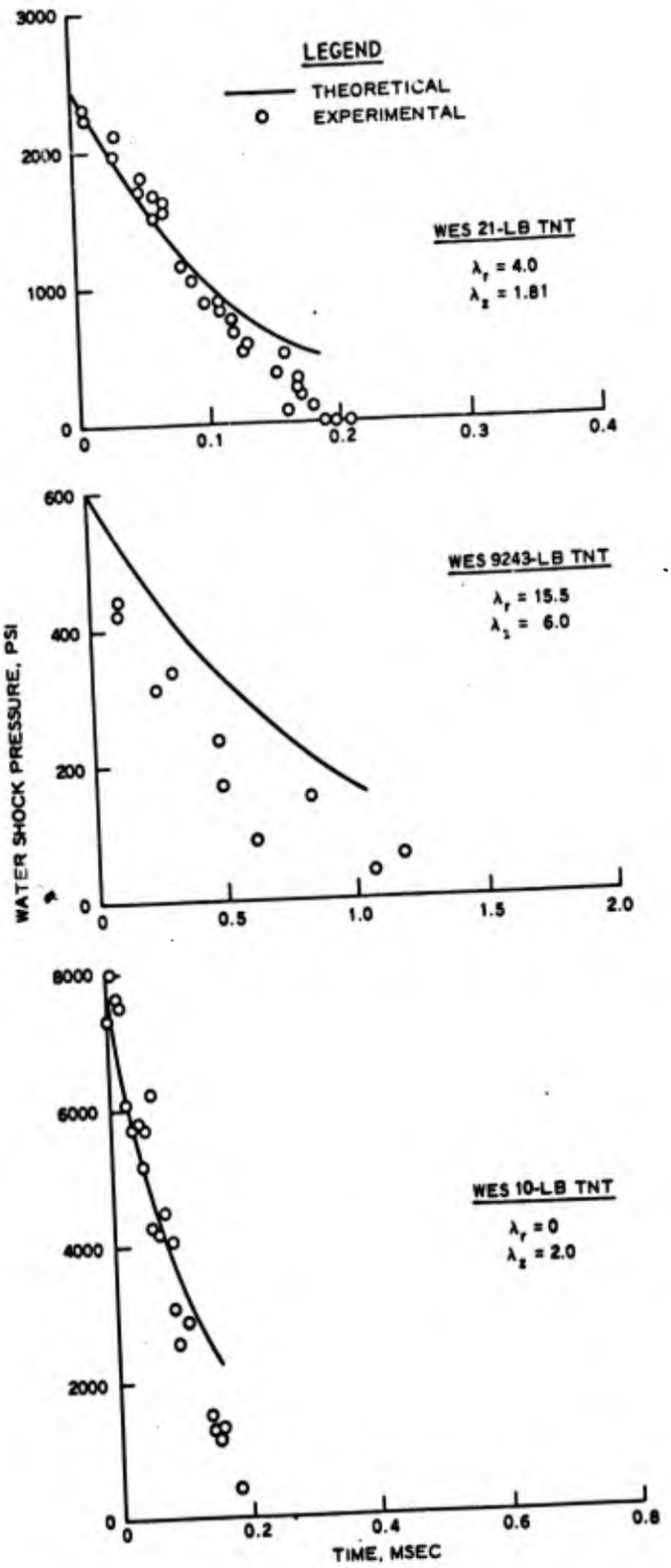


Fig. 6. Water shock pressure time history from Equation 13 compared with test data

ATMOSPHERIC PROPAGATION OF INJECTION LASER RADIATION FOR HIGH DATA RATE TRANSMISSION

ERNST J. SCHIEL, RUDOLPH R. GAMMARINO, EDWARD ARAS
ELECTRONIC COMPONENTS LABORATORY
U. S. ARMY ELECTRONICS COMMAND
FORT MONMOUTH, NEW JERSEY

The application of lasers to communication systems in the atmosphere restricts the otherwise useful characteristics of a laser, particularly its coherence properties. In spite of this fact the available high bandwidth and high directionality of optical systems can be used advantageously without the constraints imposed by the coherence aspects of lasers.

In previous investigations[1] it was found that amplitude modulation of a laser beam is impractical for transmission of information over distances greater than 1 km because of the severity of the air scintillations. However, various pulse code modulation techniques which incorporate the use of threshold detection, i.e., PFM, PPM and PCM provide a practical means to overcome the random amplitude fluctuations due to the atmospheric scintillations. Fried,[2] and others, have demonstrated the need of achieving a large signal-to-noise ratio for optical communication systems. In addition to overcoming the air scintillations, a large signal-to-noise ratio is required to overcome transmission losses due to inclement weather conditions such as fog, rain and snow.

Injection laser communication systems offer many advantages over other laser communication systems because of their operational characteristics such as high peak power, high efficiency, small size, ease of modulation and fast response time.

Injection lasers, per se, are inferior to gas and crystal lasers with respect to beam spread. However, the small size of the emitting junction of the injection laser is the best known approximation to a powerful point source and with proper collimation optics, beam spreads of less than one milliradian can easily be achieved.

The emphasis in this program was the development of a high data rate system for use in transmission through the atmosphere. This system was designed to have a large signal-to-noise ratio for distances up to 20 km and can, therefore, tolerate large insertion losses due to unfavorable weather conditions. This system has a maximum capacity of 10 megabits/sec and either PFM, PPM, or PCM can be used. With presently available military PCM equipment, simultaneous

transmission of 196 audio channels is feasible. Specifically presented in this paper are: the design and construction of the transmitter and receiver, signal-to-noise considerations, and operational characteristics of the system.

An experimental injection laser communication link was set up over a 13 km range. The transmitter was located at Fort Monmouth, N. J. The path of the laser beam traversed various terrain including a wide river, parking lots, woods and surburban areas. The average height of the laser beam above the ground was approximately 25 meters and its path across the water was approximately 1 km.

OPERATIONAL CHARACTERISTICS OF THE INJECTION LASER

The spectral output of the GaAs injection laser at threshold consists of an envelope of spectral peaks representing the allowable modes of oscillation. As the driving current is increased these modes spread in width because of the junction heating. At three times threshold the spectral emission becomes approximately Gaussian in shape with the individual modes barely discernable. This "smeared" spectra is approximately 16 Å wide at the 3 dB points. The peak emission is at 8450 Å. This peak shifts approximately 80 Å when the duty cycle approaches 50% because there is a substantial rise in the average temperature of the injection laser at this high duty cycle.

The radiation emitted by the injection laser has a beam divergence of 14 degrees. The intensity distribution versus angle can be approximated by a Gaussian function except in the plane perpendicular to the emitting junction. In this plane a diffraction pattern similar to that caused by a straight edge is superimposed on the Gaussian intensity distribution.

INJECTION LASER DRIVING CIRCUIT

The injection laser* used in the transmitter has a threshold current of 0.7 A at 77°K, and the output power at 4 A is 0.8 W. Because of the high average driving power required for the transmission of high data rate information, a special driving circuit was designed and built. A schematic diagram of the circuit is shown in Fig. 1. This driving circuit can deliver 4 A pulses with a pulse width of 35 ns and can operate at repetition rates up to 10 MHz/sec.

A special high-power, high-frequency transistor** is used as the switching element. Because of the high power dissipated in the transistor it was mounted on a heat sink with radiating fins through which forced air is directed. A Zener diode is placed across the power supply line to protect the switching transistor against accidental over-voltage. The driving current is monitored by measuring

*The injection laser used in these experiments was received from IBM Corp., under Contract No. DA28-043-AMC-01272(E).

**RCA n-p-n silicon planar epitaxial overlay transistor TA275B. Maximum power rating is 100 watts; f_t = 30 MHz.

the voltage across a 1 ohm resistor, which is put in series with the laser diode. A H-P model 216-A pulse generator is used to trigger the driving circuit. The driving circuit was placed as close as possible to the electrical feedthroughs in the dewar to minimize the loop inductance.

COOLING OF THE LASER DIODE

To achieve the high repetition rates needed in this laser communication link, the GaAs injection laser must be cooled to take advantage of the high quantum efficiency and low-threshold currents presently available only at lower temperatures. A specially designed dewar was constructed to maintain the injection laser at a temperature of 77°K (liquid nitrogen). To insure maximum heat transfer, both sides of the diode junction must be cooled. The injection laser is held in position by a "double cold finger." One side of the cold finger is mounted directly to the main dewar (Fig. 2). The other side of the cold finger is electrically insulated from the main dewar, but is connected to the liquid nitrogen reservoir by a teflon tube which allows the liquid nitrogen to come in contact with this side of the cold finger, but isolates it electrically from the dewar.

The lower section of the dewar is removable to allow easy access to the injection laser. A quartz window is mounted in this section for the transmission of the laser radiation. The collimation lens is placed directly in front of this window. Two electrical feedthroughs are placed above the injection laser in the upper section of the dewar and copper strips connect these to the laser heat sink, as can be seen in Fig. 2.

TRANSMITTER OPTICS

Because of the small physical size of the emitting junction, beam collimation is easily achieved by a single lens. When the injection laser is located exactly at the focal point of the collimation lens, the far field pattern is fan shaped, but a small amount of defocusing causes the far field pattern to become nearly circular in cross section. The beam divergence can be varied by adjusting the distance between the focal point of the lens and the emitting junction. The beam divergence (θ) is given by the following equation:

$$\theta = \frac{Dx + Af}{(x + f)\, f} \quad , \tag{1}$$

where x is the distance between the emitting junction and the focal point f. D is the diameter of the lens and A is the largest lateral dimension of the injection laser. A can be taken as the junction length. The lens diameter required for collimation of the radiation is given by

$$D = 2f \tan \frac{\alpha}{2} \quad , \tag{2}$$

α is the beam divergence at the junction.

It was found that the practical lower limit of the beam divergence for use in communication systems is 1 milliradian. If a smaller beam divergence were used, refraction and deflection caused by air turbulence over long distances would cause intolerable large signal fluctuations. Furthermore, if a very small beam divergence (less than 1 milliradian) was used the stability of the mounting platform, and even the building in which the receiver and transmitter are located, becomes extremely critical.

In the experimental transmitter, a Zeiss-Sonar lens (f = 180 mm, F = 1:2.8) is used for beam collimation. This lens is ideally suited for this purpose because it has a relatively long focal length and sufficient movement in the direction of the optical axis to permit the beam divergence to be varied from less than 1 milliradian to 5 milliradians. Figure 3 shows the transmitter with the collimation optics and sighting telescope. The transmitter is placed on a 12" milling machine rotary table which has fine vertical and horizontal movements and a minimum amount of backlash to insure stable operation.

RECEIVER STATION

The receiver station (as shown in Fig. 4) is located at a distance of 13 km from the transmitter. Two separate receiver systems are employed in this station. One receiver uses a Questar telescope with a $3\frac{1}{2}$" aperture, as seen on the left, and the other receiver uses a 12" Cassegrainian telescope, as seen on the right. Variable field stops are placed at the focal point of the telescope and narrow bandpass filters are used in these receivers to suppress the background radiation. The field stops allow the field of view to be varied from 1-5 milliradians. Interference filters with a bandpass of 100 Å and centered at 8500 Å are used. Field stops and interference filters are placed as close to the focal points of the main mirror as possible. The detector, a 7102 photomultiplier tube, is mounted directly behind the interference filter and field stop. The received signal was monitored on a Tektronix 585A oscilloscope. Background current measurements were taken before each test. The quantities measured were the dc current of the photomultiplier tube and its RMS noise. A H-P model 425A dc microvolt-micrometer was used to measure the dc current. It was connected directly to the photomultiplier tube output. In measuring the RMS noise of a photomultiplier tube its output was first amplified 100 times (40 dB) and then measured with a H-P model 411A RMS voltmeter.

For error rate measurements, a 100 kHz frequency standard was used to trigger the laser driving circuit of the transmitter. This frequency was chosen because it is much higher than the highest frequency of the signal fluctuations caused by the scintillation effects in the atmosphere. The received signal was amplified 100 times and then used to trigger a H-P model 214A pulse generator. The adjustable trigger level of this generator allows the setting of a defined detection threshold. It was adjusted by the operator to the optimum level for detection of the received signal pulses.

The equation for the received power is given by:

$$P_{rec} = \frac{P_o T_{rec} T_{tr} A_{rec}}{\Omega_{tr} R^2} C \tau_1 \exp - [\sigma (t) R]. \quad (3)$$ *

In this equation, a very simple expression is used to represent the time fluctuation of the signal, i.e., $C \tau_1 \exp - \sigma(t) R$. The attenuation factor could be considered to be a product of two functions, one which is slowly varying (attenuation due to fog or rain) and one which varies in the low audio range which is due to air scintillations. This time varying attenuation factor allows the $\frac{S}{N}$ equation to be written in a very meaningful form for error rate analysis.

The $\frac{S}{N}$ is given by:

$$\frac{S}{N} = \frac{P_{rec} S_c}{[4e (i_d + i_{bgr}) \Delta f]^{\frac{1}{2}}} . \quad (4)$$

The values of the background current is determined by the following equation:

$$i_{bgr} = A_{rec} T_{rec} \Omega_{rec} \Delta\lambda \; b \; S_c . \quad (5)$$

For daylight operation a value of $1 \times 10^{-6} \frac{W}{cm^2 \, \mathring{A} \, sr}$ was chosen for the spectral irradiance (b). This value represents the spectral irradiance for a clear bright day.

Considering the operational characteristics of the receiver and transmitter to be constant in comparison to $C \tau_1 \exp(-\sigma R)$, we can express Eq. (4) in the following simple form:

$$\frac{S}{N} = K C \tau_1 \exp(-\sigma R). \quad (6)$$

Where K is a constant determined for a particular receiver station at a fixed distance from the transmitter. The received power for various distances is given in Figs. 5 and 6 for the $3\frac{1}{2}$" receiver and the 12" receiver, respectively. Also shown in these figures is the equivalent noise input power for day and night operation. Assuming the signal-to-noise must be greater than 17 dB for acceptable reception[3] (an error rate of less than 10^{-7}), the value of the minimum acceptable power at the receiver can be calculated. In this form the error rate measurements can easily be correlated to C, τ, and σ. From Figs. 5 and 6, we can determine the acceptable transmission losses for various distances. This value in dB is the difference between the received power and the minimum acceptable power level. As an example, consider the 12" receiver system operating in daylight with the transmitter 13 km from the receiver. From Fig. 6, we find that the transmission loss must be less than 22 dB for acceptable reception.

*See Appendix for list of symbols and numerical values.

ATTENUATION OF RADIATION IN THE ATMOSPHERE

The atmospheric insertion loss in the propagation of infrared radiation is mainly due to three distinct effects: (1) water vapor absorption, (2) Rayleigh and Mie scattering, i.e., $\exp(-\sigma R)$ and, (3) scintillations. All three effects are dependent on weather conditions. The losses due to water absorption and scattering remain constant for a comparatively long time (minutes to hours). Scintillations on the contrary cause signal fluctuations up to 1 kHz.

Water Vapor Absorption

The wavelength of the injection laser radiation ($\lambda = 8500$ Å) is at the center of the first infrared transmission window (7200 - 9400 Å). The transmission (τ_1) can be computed by using the emperical expression developed by Langer,[4]

$$\tau_1 = K_1 \left(\frac{w_1}{w} \right)^{\beta_1} \tag{7}$$

The dependence of the absorption on the amount of precipitable water in the atmosphere has been tabulated by Langer.[4] For a relative humidity of 100% and a temperature of 90°F, the amount of precipitable water in the 13 km path is approximately 500 mm. The transmission for this extreme condition is 62.3% and the corresponding attenuation is less than 3 dB.

Scattering

For a wavelength of 8500 Å, Rayleigh scattering can be completely neglected. Mie scattering, however, which is caused by particles in the atmosphere of size comparable to the wavelength has a very large effect in the transmission. The transmitted intensity can be expressed as $\exp(-\sigma R)$. Direct measurements of σ are difficult, but a good estimate can be obtained by relating it to the visual range V. The relations,[5]

$$\sigma = + \frac{3.91}{V} \left(\frac{\lambda}{0.55} \right)^{q} \tag{8}$$

and

$$\tau_s = \exp(-\sigma R) \tag{9}$$

make it possible to compute the transmission losses for a given wavelength ($\lambda = 0.85$ μ) and range R. A typical value for q during average viewing conditions is 1.3. When the visual range is limited to 6 km or less due to haze, a value for q, suggested by Lohle[6] is $q = 0.585\ V^{1/3}$. Figure 7 shows the dependence of σ on the visual range, and Fig. 8 the scattering loss for different values of σ. From Fig. 6 we find that the maximum loss must be less than 22 dB for acceptable reception. Thus, it can be calculated that the maximum acceptable value of for transmission over the 13 km injection laser line is 0.39 km^{-1}. This value corresponds to a visual range of 6 km.

Scintillations

A fundamental limitation to communication by coherent light transmission through the atmosphere is the unavoidable effects created by normal atmospheric turbulence. The presence of random inhomogenieties in the atmosphere causes a random fluctuation of the refractive index, which in turn, distorts, bends and impresses an undesirable modulation on the laser radiation. The effects on the received signal are very clearly shown in Fig. 9, which shows this random modulation. (The output of the transmitter consists of pulses of equal intensity.) All pulses that are above the detection threshold can be used to carry information in PCM, PFM, or other pulse modulation methods. If, however, a "deep fade" occurs, i.e., the received signal falls below the detection threshold, errors in a data channel are introduced. It is usually assumed and has also been experimentally verified for short ranges that scintillations have a log-normal distribution. Fried's analysis[2] of the effects of scintillation on the error rate have shown that the source of the very large loss factor C is the high probability in a log-normal distribution of values much smaller than the mean value. It is, therefore, necessary that a very large signal-to-noise ratio be achieved in order to obtain a low error rate.

In our measurements, scintillation strength was directly observed on the oscilloscope, but because of the rather extensive instrumentation required, no attempts were made to determine the numerical value of the refractive index structure constant. It was found that the 22 dB "fade margin" was sufficient to suppress the effects of scintillations to the extent that an error rate of less than 10^{-5} could be attained for all conditions except where the laser radiation was severely attenuated by scattering in dense haze or fog (visual range less than 6 km).

EXPERIMENTAL RESULTS

Extensive tests were conducted over the injection laser link during the spring, summer and fall of 1967. During a 24-hour test run on 14-15 June 1967, the measured error rate did not exceed 2.5×10^{-5} except for a three-hour period at dawn when ground fog limited the visibility to less than 6 km. Since it was not possible to present all the data taken, error rate measurements for four typical conditions are given:

Time:	25 May 1967, 1600 hours
Weather:	Rain, visibility 5 km
Scintillations:	weak
Error rate:	2.6×10^{-5}

Time:	14 June 1967, 1100 hours
Weather:	light haze, overcast, visibility 10 km
Scintillations:	medium
Error rate:	1.1×10^{-5}

Time:	14 June 1967, 2115 hours
Weather:	light haze, visual range 10 km
Scintillations:	weak
Error rate:	2.5×10^{-7}
Time:	15 June 1967, 0515 hours
Weather:	ground fog, visual range 4 km
Scintillations:	strong
Error rate:	6.8×10^{-3}.

CONCLUSIONS:

The injection laser link has been designed and constructed for transmission of data at a rate of 10 megabits/sec. Special consideration was given in the design to achieve the highest possible signal-to-noise ratio. This link was ideally suited for error rate measurements. The high permissive atmospheric insertion loss could compensate to a high degree for the influence of adverse weather conditions. An error rate in the range of 10^{-4}-10^{-5} could be maintained for this 13 km link even when the visual range diminished to 6 km. Since the operation over a range of 13 km constitutes a severe test, it is expected that operation over shorter distances (8 km "down the hill") will be greatly improved and that this link will then be available for all but the most adverse weather conditions (heavy fog).

REFERENCES

1. E. J. Schiel, J. J. Bolmarchich, "Direct Modulation of a He-Ne Gas Laser," Proc. IEEE, June 1963.

2. D. L. Fried and R. A. Schmeltzer, "The Effect of Atmospheric Scintillation on an Optical Data Channel--Laser Radar and Binary Communications," North American Aviation, Inc., Anaheim, California, Rpt. C7-805/401, Optical Effects of Atmospheric Turbulence, Compiled as Lecture Notes Under Contract NAS-8-18035, pp. 207-244, March 1967.

3. W. R. Bennett, "Methods of Solving Noise Problems," Proc. IRE, May 1956.

4. R. M. Langer, J. R. M. Bege Co., Arlington, Mass., "Atmospheric Transmission in the Near Infrared," U. S. Signal Corps Contract No. DA36-039-SC-72351, Part I, Final Report, May 1957.

5. P. W. Kruse, Elements of Infrared Technology, John Wiley & Sons, New York, N. Y., 1962.

6. F. Löhle, Phys. Zeitschrift, 45 (1944).

APPENDIX

LIST OF SYMBOLS

Symbol	Description	Values
A_{rec}	Effective area of receiver aperture (12" telescope) Effective area of receiver aperture ($3\frac{1}{2}$" Questar)	700 cm^2 57.8 cm^2
i_{bgr}	Background current 12" telescope) Background current $3\frac{1}{2}$" Questar)	3.24×10^{-10} A 2.66×10^{-11} A
i	Dark current (25°C)	1.8×10^{-11} A
P_o	Output of the transmitter	1.0 W
P_{rec}	Incident power on receiver aperture	80 µW
R	Distance between the transmitter and receiver	13 km
S_c	Cathode sensitivity of the phototube	2.7×10^{-3} A/W
$\frac{S}{N}$	Signal-to-noise ratio	---
T_{rec}	Transmission coefficient of receiver	0.5
T_{tr}	Transmission coefficient of transmitter	0.5
Δf	Amplifier bandwidth	20 MHz
Ω_{tr}	Beam spread of transmitter	10^{-6} sr
Ω_{rec}	Field of view of receiver	4×10^{-6} sr
$\Delta\lambda$	Bandpass of the interference filter	100 Å
b	Spectral irradiance	$10^{-6} \frac{W}{cm^2 \text{ Å sr}}$
τ_1	Water vapor transmission factor	---
σ	Scattering coefficient (Mie)	---
V	Visual range	---
C	Scintillation factor	---

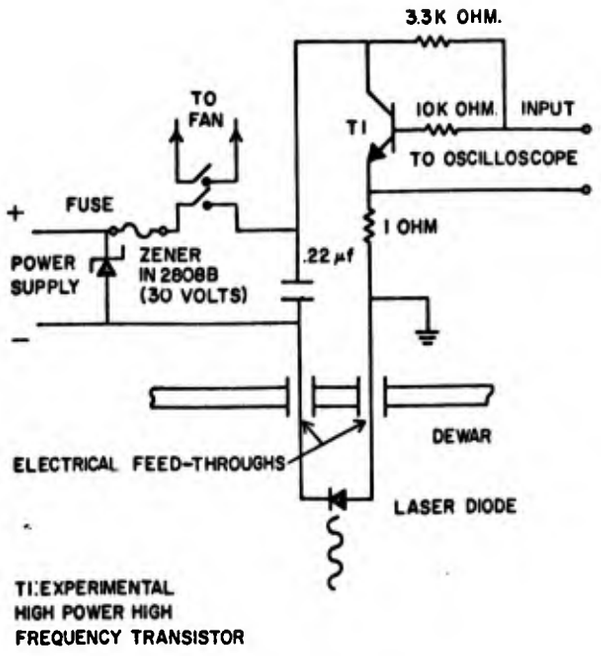


Fig. 1 Schematic Diagram of Driving Circuit

Fig. 2 Injection Laser Mount in Dewar

Fig. 3 Transmitter

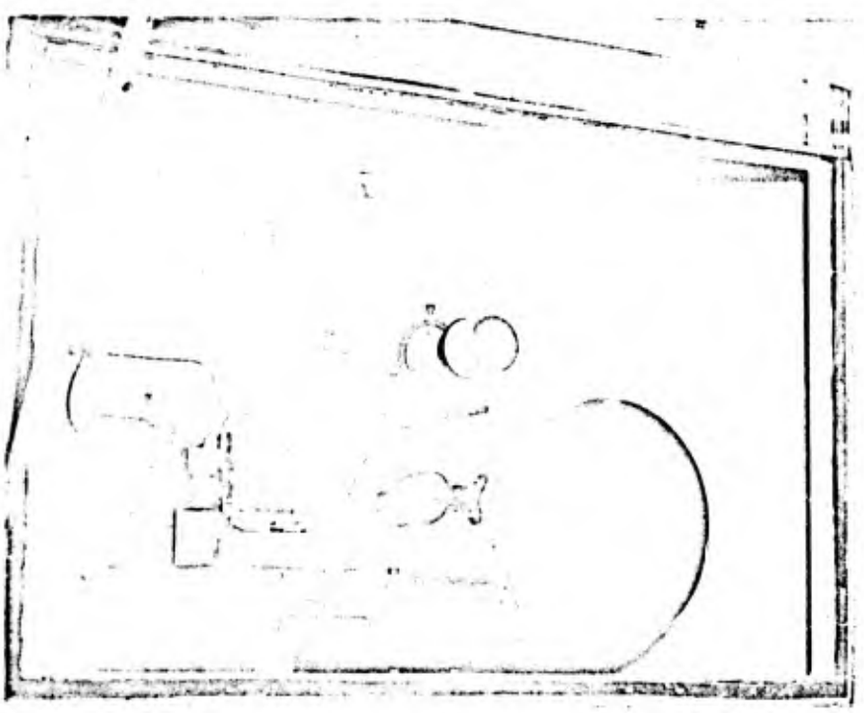


Fig. 4 Receiver Station

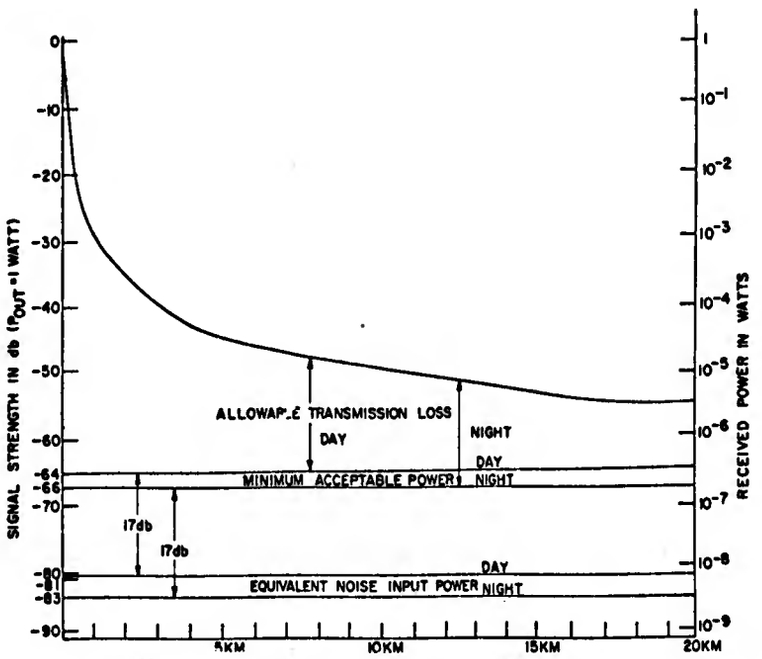


Fig. 5 Received Power as a Function of Distance ($3\text{-}\frac{1}{2}$" Receiver)

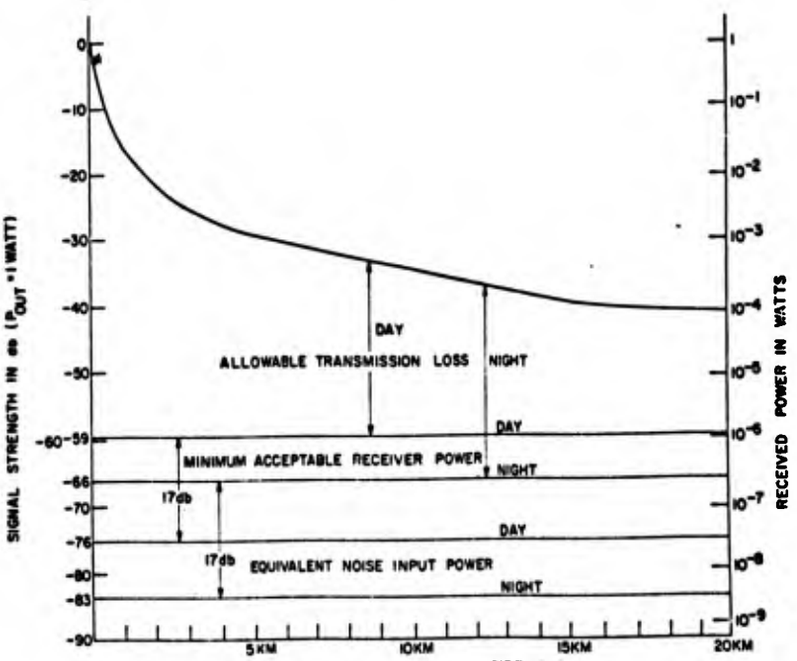


Fig. 6 Received Power as a Function of Distance (12" Receiver)

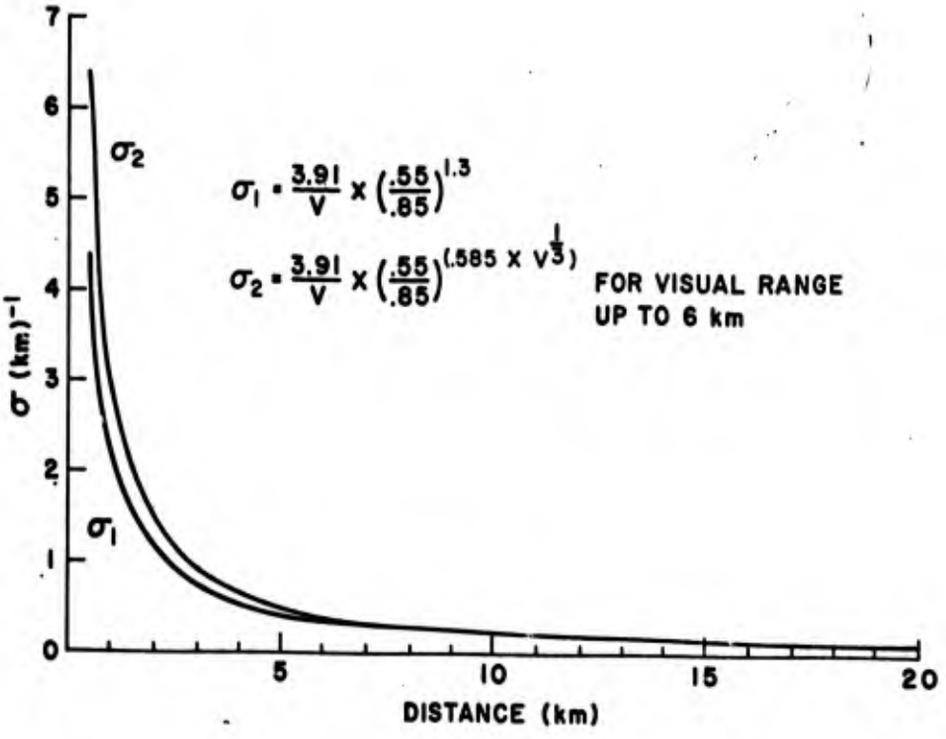


Fig. 7 σ versus Visual Range

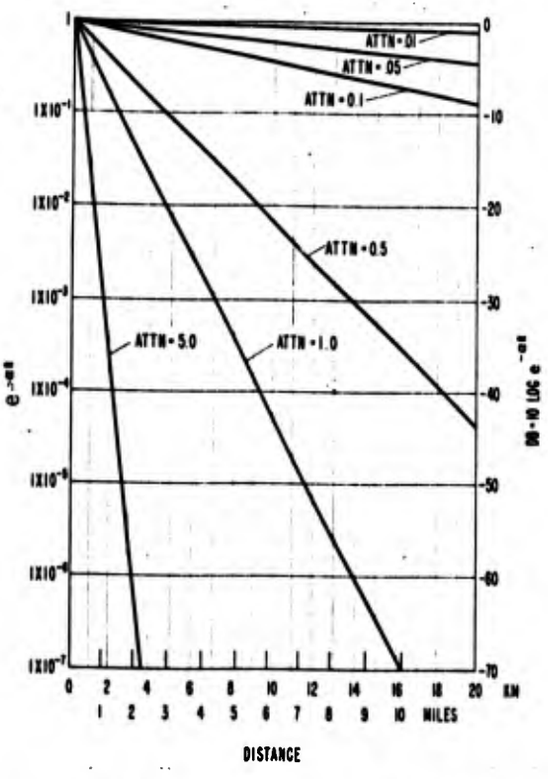


Fig. 8 Scattering Loss for Different values of σ

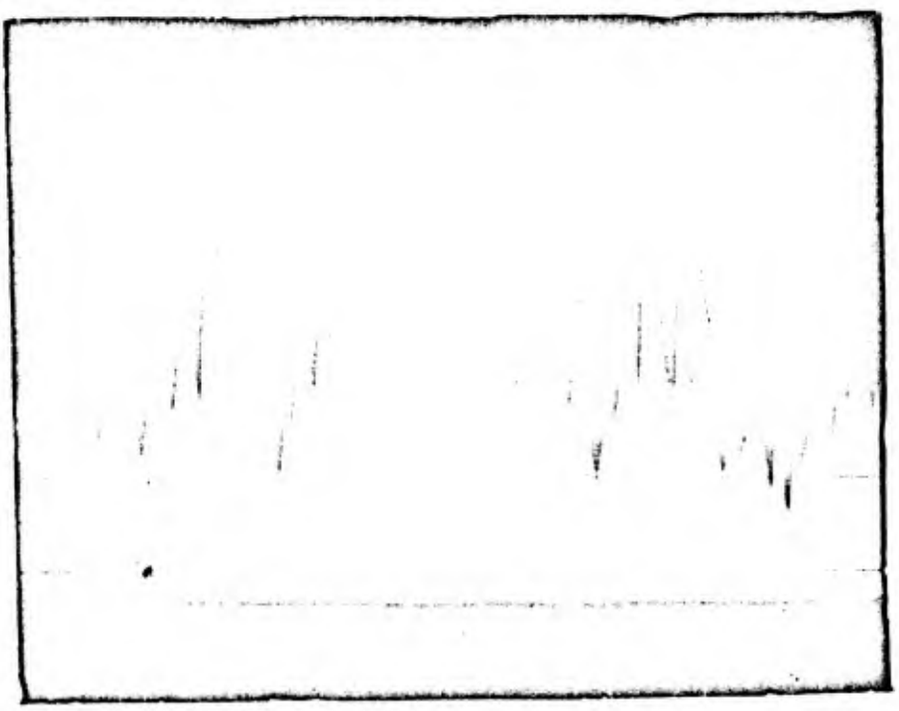


Fig. 9 Signal Fluctuations Due to Scintillations

WATER VAPOR PERMEABILITY IN A POLYESTER BASED POLYURETHANE

N.S. Schneider, L.V. Dusablon and L.A. Spano
U.S. Army Natick Laboratories
Natick, Massachusetts 01760

INTRODUCTION

The sorption, diffusion and interactions of water vapor with polymers is of widespread importance to Army material requirements. Environmental water can have a profound effect on the performance of polymers through its influence on mechanical behavior alone or its contribution to hydrolytically induced polymer chain scission. Moreover, the mechanism of water transmission is important in a variety of protective applications such as packaging films, corrosion resistance coatings and the like. In the present work, the main emphasis has been on the study of the mechanism of water vapor transport in polymers with high moisture vapor transmission rates (MVTR's) that might serve as physiologically comfortable barrier films. Such polymers would be useful as fabric coatings for rainwear or for Chemical-Biological (CB) agent protection if the coating were resistant to agent vapors. The importance of such a development hardly needs to be emphasized since it would open the door to a significant new approach to achieving chemical protection as part of the combat clothing issue.

To date there has been little progress in the development of polymers having MVTR's which are sufficiently high to serve as physiologically acceptable barrier films. However, certain commercial polyurethane elastomers offer MVTR's which are on the threshold of the range needed for physiological benefit. The benefits conferred even by the limited MVTR of these polyurethanes have been demonstrated in climatic chamber tests at Natick Laboratories. Volunteer subjects wearing air impermeable coveralls coated with neoprene rapidly succumbed to heat

exhaustion whereas test subjects in polyurethane coated coveralls survived 2 hours at rest followed by a 50 minute walk under tropical conditions. The latter subjects were uncomfortable but remained active.

The work presented in this paper was undertaken to gain an improved understanding of the factors controlling water vapor transmission in polyurethanes. Polyurethanes were chosen as candidate materials because of their favorable MVTR's. In addition, since polyurethanes are block copolymers, they offer numerous possible variations in the nature, length and proportions of the alternating long flexible and short hard segments that constitute the polymer chain. The study was expected to provide a basis for determining whether and by what means polyurethanes with the required high MVTR's could be designed. In addition, the study was expected to shed light on the general mechanism of water vapor sorption and diffusion which is important to other polymer applications.

THEORY

Consider a polymer film of uniform thickness 1 with a difference in the concentration of penetrant molecules $\Delta c = c_2 - c_1$ across the film. According to Fick's first law[1,2] the flux of penetrant molecules J in grams or $cm.^3$ per unit area and per unit time is proportional to the concentration gradient dc/dl

$$J = - D \frac{dc}{dl}$$

The constant of proportionality D is known as the diffusion constant. If the film is exposed to vapor at a pressure p_2 on one side and $p_1 < p_2$ on the other side and if the solubility of the vapor in the polymer is governed by Henry's law

$$c = kp$$

then the flux can be expressed in terms of the pressure gradient

$$J = -Dk \frac{dp}{dl}$$

Thus, the rate of transmission of penetrant per unit pressure gradient is governed by the product of a thermodynamic factor and a kinetic factor. The thermodynamic factor is the Henry's law solubility constant k which depends solely on the polymer-penetrant interactions. The kinetic factor is the diffusion constant D which is governed by the rate of the polymer segmental motion and the size of the penetrant molecules. The product Dk is termed the permeability coefficient P and in an ideal system would be independent of the experimental conditions.

In attempting to understand the relation between the polymer structure and moisture vapor transmission rates it will be necessary to resolve P into its component factors and relate these separately to the polymer structure.

EXPERIMENTAL METHODS

Steady state transmission rates were determined using cup cells containing water and sealed with the polymer film. The cells were placed in an air stream maintained at 70°F and 50°% R H and the loss in weight due to evaporation of water across the film was measured at timed intervals. The results were converted into values of the flux or the permeability coefficient as required from a knowledge of the film area, film thickness and relative humidity difference. The water solubility was determined from the weight gain of the sample immersed in water. In combination with P this permits a calculation of the effective D as the ratio P/k.

While the parameters obtained from these measurements are useful in comparing different polymers, this method does not provide data which are needed to understand the complex behavior exhibited by water vapor. In particular, it is not possible to vary the concentration of water in the polymer as required to determine the sorption isotherm or the dependence of the diffusion constant on concentration. For this purpose a vacuum-sorption system was used in which the sample could be exposed to a selected vapor pressure of water and the weight gain measured with a Cahn electro-balance or quartz spring. Diffusion constants were calculated by standard methods[1,2,3] from the rate of water uptake and the solubility was obtained from the final constant weight of the sample at the particular vapor pressure.

SAMPLES

An initial survey was conducted on several commercial or exploratory polyurethanes of incompletely specified composition but the main body of the work was concerned with a series of structurally varied polyurethanes. All were composed of a polyester or polyether (molecular weight 2,000), (1,4)-butanediol and MDI in the mole ratio 1:3.2:2. They differed only in the nature of the flexible segment which included polybutylene adipate and three polyethers representing an increasing frequency of ether linkages in the order polytetramethylene oxide PTMO, polypropylene oxide PPO, and polyethylene oxide PEO.

MOISTURE VAPOR TRANSMISSION RATES OF TYPICAL POLYURETHANES

The moisture vapor transmission rates of three commercial and one exploratory polymer (LD - 550) appear in Table I expressed as the flux for a one mil thick film. Since all of these samples have a water solubility of about 2%, the nearly eight-fold variation in permeability which occurs in this group of polymers must arise solely from differences in the apparent diffusion constant for water.

The diffusion of penetrant molecules in a rubbery polymer matrix can be represented as occurring by a process of jumping into holes created by the cooperative motion of the polymer segments. In effect, the penetrant diffusion constant is determined by the rate of polymer segmental motion. It is well established that the rate of segmental motion for all polymers is described by a nearly universal function which depends only on the difference between the experimental temperature and the glass transition temperature $\Delta T = T-Tg$[5]. (The glass transition temperature is the point at which the the sample transforms from a rubbery to a glassy solid due to the freezing out of segmental motion.) In turn, this implies that in comparing polymers at a fixed temperature, the diffusion constant, like the rate of segmental motion, will be larger if the glass transition temperature is lower. The comparision in the last column of Table I shows that this prediction is true, suggesting that for polymers with equal water solubilities, the polymer with the lower Tg will have the higher permeability. However useful such a rule proves to be on a qualitative basis, it fails quantitatively since theory predicts a one-hundred fold rather than an eight-fold range in MVTR for the differences in Tg corresponding to these polymers.

RESULTS ON STRUCTURALLY VARIED POLYURETHANES

The relevant parameters characterizing the steady state transmission of water vapor in the series of structurally varied polyurethanes are listed in Table II. The water solubility changes very little in the first three samples but undergoes a spectacular increase in the PEO type of polyurethane.

Taking the results for the PBA sample as the basis of comparison, consider first the behavior of the PTMO sample. Since PTMO-PU has a much lower Tg, it would have been expected that the permeability would be

much higher than for the PBA sample, in keeping with the comparative behavior of the commercial polyurethanes. In fact, the permeability of the PBA and PTMO samples are nearly the same, whereas theory predicts a 35 times higher value for the latter sample. There is another inconsistancy in this set of results. In the PEO sample the water concentration is 80 times higher than in the PBA sample and yet the permeability has increased only 7 times. Since Tg is the same for these two polymers, it would have been expected that the increase in permeability would equal the 80 times increase in water concentration. These two comparisons indicate that the water permeability cannot be predicted simply from a knowledge of the concentration of water or the glass transition temperature. The behavior is complicated by additional considerations which can be understood only by examining tne component factors of the permeability coefficient in more detail.

Before proceeding further it would be useful to compare the MVTR of these films with values of practical interest. An MVTR of about 40g./100 in.^{2}/24 hours is required for perceptible heat relief while the MVTR for standard cotton poplin (5.3oz./yd.^{2}) is about 650g./100 in.^{2}/24 hours. For all the polymers the values of J1, which represent the MVTR for a one mil film, exceed the minimum required value. The value of the unnormalized flux J for the PEO sample exceeds even that for cotton poplin. The usual normalization to one mil thickness does not hold for this sample and is omitted from the Table II.

POLY(BUTYLENE ADIPATE) POLYURETHANE

The sorption isotherm obtained by the vacuum sorption method is shown in Figure 1. The water concentration (g. water/cm.^{3} polymer) against the partial pressure or activity of water (p/po) is initially linear, approximating Henry's law behavior. However, above an activity of 0.6 the isotherm turns concave upward in a manner typical of polar sorbents in polar substrates. Unexpectedly it was found that the isotherms determined at different temperatures were superimposable. This signifies that the heat of mixing for water with the polyurethane is zero[1]. Accordingly, it appears that the environment of water molecules dissolved in the polymer is energetically equivalent to that of liquid water. The dependence of the diffusion constant on the water concentration in the polymer is shown in Figure 2. The diffusion constant decreases with increasing water concentration in marked contrast to the usual strong increase

in the diffusion constant that occurs with the swelling of polymers by organic vapors[1].

The concave upward behavior of the sorption isotherm, the zero heat of mixing, and the decrease in the diffusion constant of water can all be attributed to the self association of the water molecules dissolved in the polymer matrix. The reconstitution of intermolecular hydrogen bonding among water molecules in clusters is responsible for the zero heat of mixing; the diminished activity of the increasingly clustered water, as compared to randomly distributed water molecules, increases the concentration of dissolved water and produces the concave upward isotherm[5]: and the immobilization of water in clusters reduces the effective diffusion constant[6]. Thus water clustering dominates the overall process of water vapor transmission in this polymer.

POLY (TETRAMETHYLENE OXIDE) POLYURETHANE

The question to be answered is whether clustering also occurs in this polyurethane and, if so, whether it accounts for the much lower than expected MVTR. Both the sorption isotherm and the curves representing the concentration dependence of the diffusion constant for this polymer so closely resemble those for the PBA-PU that it is unnecessary to reproduce them. This not only shows that water clustering dominates the MVTR of this polymer but also indicates that the extent of clustering is no greater than in PBA-PU. Another mechanism is required to explain the abnormally low apparent diffusion constant.

There is evidence that the low diffusion constant arises from the presence of organized amorphous structure in the PTMO-PU which results from association of the urethane regions of the polymer chains[7,8]. Low angle x-ray scattering patterns provide direct evidence that such domain structure is present in the PTMO-PU but not in the PBA-PU[9] and this conclusion is supported by several lines of more indirect evidence[8]. If these domains are impermeable to water, the diffusing molecule would be required to travel around the impediment, thus lengthening the diffusion path and decreasing the effective diffusion constant[11].

POLY (ETHYLENE OXIDE) POLYURETHANE

The sorption isotherm for this polymer (Figure 3) differs significantly from the behavior of the other two polymers and is marked by an almost exponential increase in concentration above an activity of 0.5. This

strong concave upward curvature indicates that a larger fraction of the water exists in the form of clusters than occurs in the other two polymers. The radical increase in the extent of clustering is also reflected in the precipitous drop in the diffusion constant with increasing water concentration (Figure 4). Thus, it can be concluded that the relatively limited increase in MVTR accompanying the large difference in water concentration, as compared to PBA-PU, is due to the immobilization of a much larger fraction of the water in clusters. However, it should be pointed out that in this instance the values of the diffusion constants obtained from the time dependence of the sorption are influenced by other effects[11,12] which tend to overemphasize the decrease in D with increasing concentration.

CONCLUSIONS

These studies provide considerable insight on the mechanism of water vapor transport in polyurethanes and open the way to the rational design of polymers with high moisture vapor transmission rates (MVTR's). Certainly the major conclusion revealed by this work is that the clustering of water dominates all aspects of the water transport process. In particular, clustering is responsible for the decrease in the effective diffusion constant which tends to offset the gain in permeability due to the increased water uptake. Behavior of this sort is undoubtedly not peculiar to the polyurethanes but probably occurs to a significant degree in other polymers as well. The previously unsuspected presence of domain structure, which occurs to varying degrees depending on the composition of the particular polyurethane, also acts to reduce the diffusion constant and the MVTR. These two factors, clustering and structure, complicate any simple correlation between MVTR's and such characteristic polymer parameters as the water solubility and the glass transition temperature.

Despite these effects it is evident that major changes in the MVTR can be achieved by suitable alteration of the polyurethane structure, even within the narrowly prescribed limits of the present study. For the most hydrophilic polymer, poly (ethylene oxide) polyurethane, the MVTR is competitive with the values for textile fabrics. However, in any practical application it would be necessary to limit the extent of swelling of the polymer by water. Fortunately, due to the counterbalancing effects of clustering on the diffusion constant and water uptake, it should be possible to design polymers with lower degrees of swelling at still acceptable

moisture vapor transmission rates. There are several alternative ways of modifying PEO-PU. Films of suitable composition could be made from physical mixtures of the PEO- and PTMO- polyurethanes or mixing could be accomplished on a molecular basis using block or random copolymers incorporating the ethylene oxide repeating unit. In addition, high permeabilities might be achieved with polymers of another structure. In particular, ionic groups are known to destructure water and could minimize the intrusion of clustering. Although not included in this presentation, results have been obtained on polyion graft copolymers that do, in fact, show higher water vapor permeabilities at substantially lower swelling ratios.

TABLE I

MOISTURE VAPOR TRANSMISSION RATES FOR SEVERAL COMMERCIAL POLYURETHANES

Polyethers	J1 $\frac{\text{g-mil}}{100\text{in}^2\text{-24hrs.}}$ *	Tg°c
Du Pont LD-550	190	-60
Estane 574-140	84	-51
Polyesters		
Mobay PBA	68	-45
Estane 574-100	26	-18

Conditions for steady state transmission measurements: cup cells; inside, water in contact; outside, 40% RH, T= 90°F.

*The saturation water concentration was about 2% for all samples.

TABLE II

WATER SORPTION AND PERMEABILITY IN STRUCTURALLY VARIED POLYURETHANES

Sample	Thickness (mil)	Water Conc. g/100g	J (a)	J_1 (b)	Tg °c
PBA	11.2	1.6	6.9	78	-45
PTMO	12.5	1.8	4.3	64	-85
PPO	13.0	2.8	6.0	78	-50
PEO	14.0	126.0	780.0	--	-49

Conditions for steady state transmission measurements: cup cells; inside, water in contact; outside, 50% RH; T=70°F.

(a) g./100 in.^{2}/24hrs.

(b) g.-mil/100in.^{2}/24hrs.

ACKNOWLEDGEMENTS

The authors wish to acknowledge the work of Elizabeth Snell and Robert Prosser.

REFERENCES

1. Rogers, C. E. in Engineering Design for Plastics, E. Baer (Ed), Reinhold Publ. Corp, New York, (1964), Chap. 9.
2. van Amerongen, G., J. Rubber Chem. Tech., 37, 1067 (1964).
3. Crank, J., The Mathematics of Diffusion, Oxford, London (1956), Chap. 1.
4. Ferry, J. D., Viscoelastic Properties of Polymers, John Wiley, New York (1961), pp 215 and ff.
5. Zimm, B. H., and J. L. Lundberg, J. Phys. Chem., 60, 425 (1956).
6. Barrie, J. A., and B. Platt, Polymer 4, 303 (1963).
7. Cooper, S. L., and A. V. Tobolsky, J. Appl. Polymer Sci., 10, 1837 (1966).
8. Clough, S. B., and N. S. Schneider, Polymer, in press.
9. Clough, S. B., A. King and N. S. Schneider, in preparation.
10. Michaels, A. S., and H. J. Bixler, J. Polymer Sci., 50, 393 (1961).
11. Rogers, C.E., Op. Cit. Chapter 8.
12. Fujita, H., Fortschr, Hochpolym.Forsch, 3, 1 (1961).

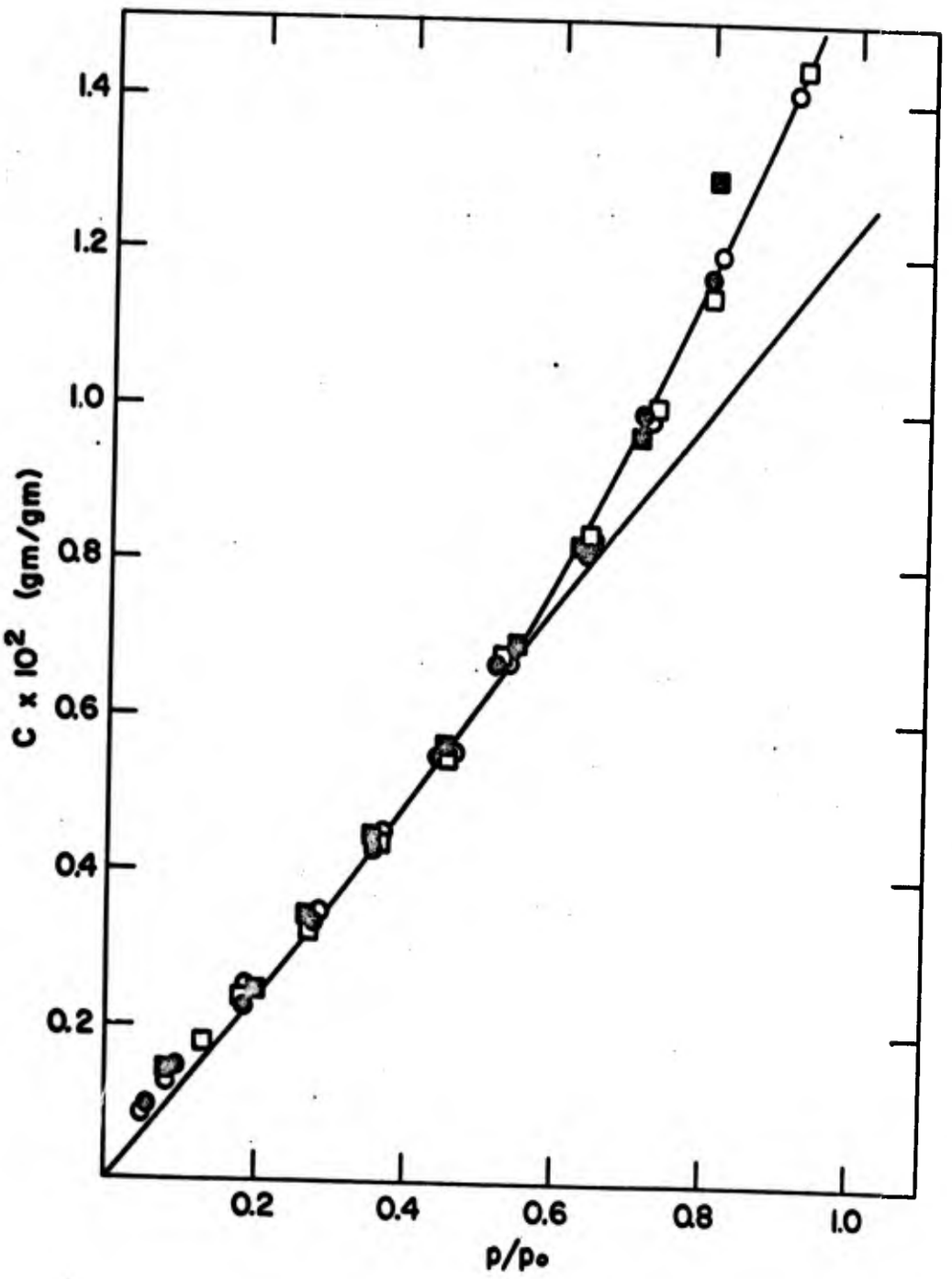


FIGURE 1 Sorption isotherm for poly(butylene adipate) at 40°C; unfilled points sorption, filled points desorption.

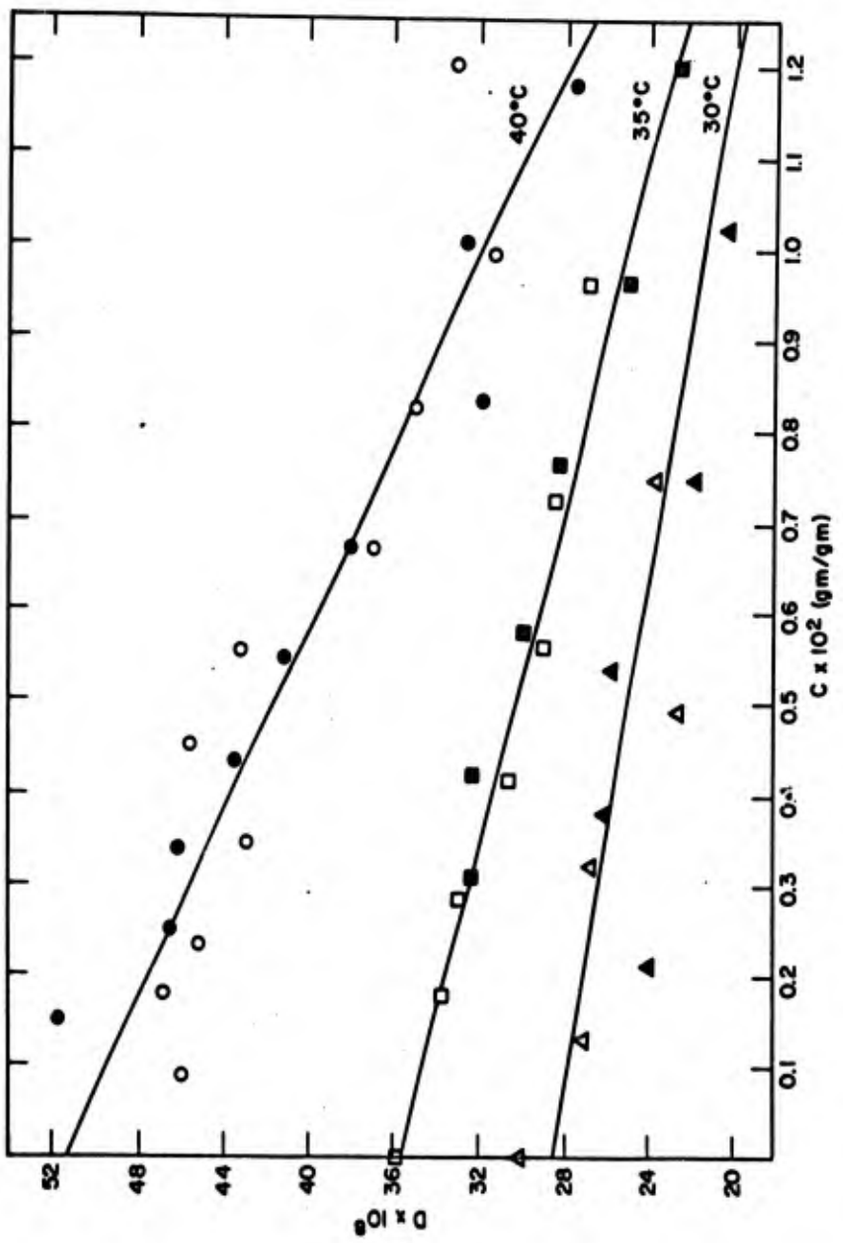


FIGURE 2 Dependence of the diffusion constant for poly(butylene adipate) on water concentration; unfilled points sorption, filled points desorption.

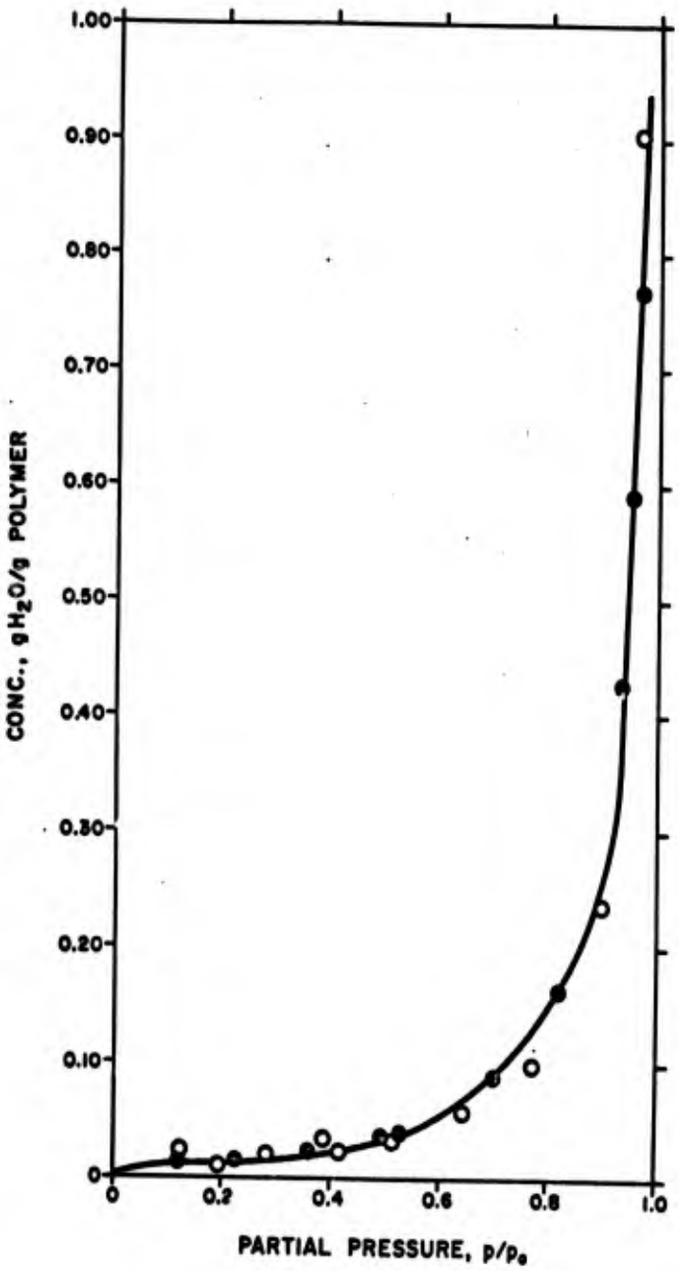


FIGURE 3 Sorption isotherm for poly(ethylene oxide) at 35°C; unfilled points sorption, filled points desorption.

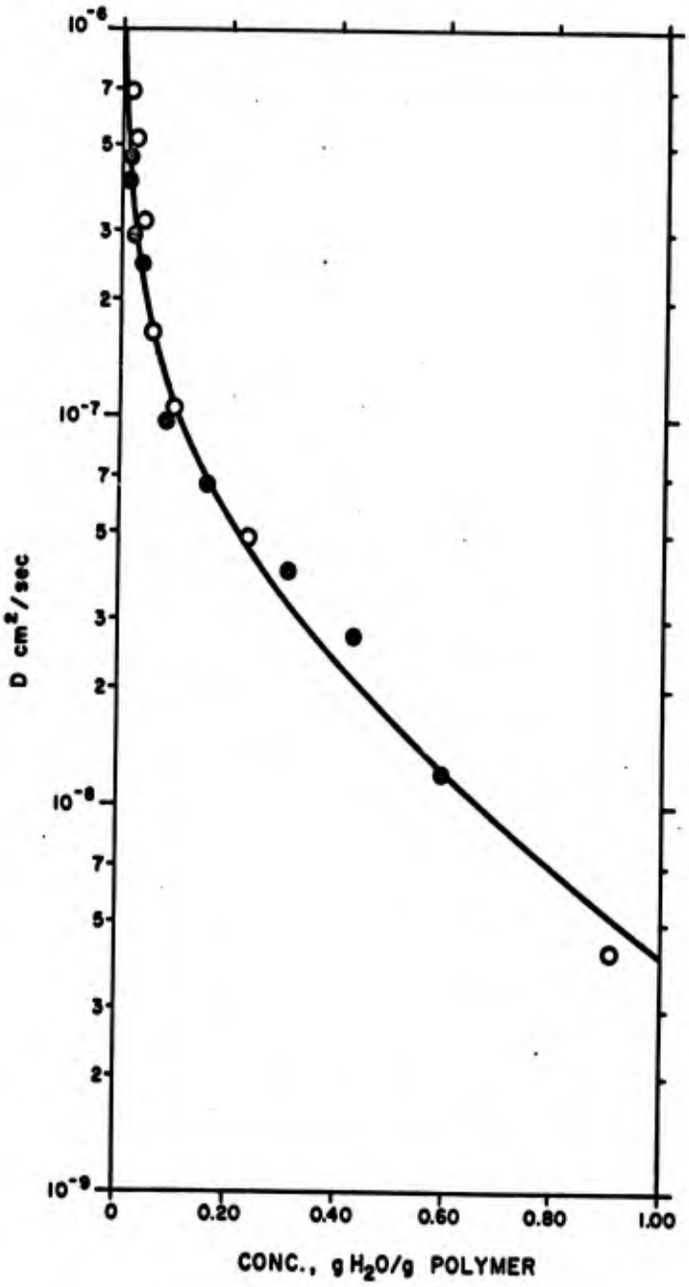


FIGURE 4 Dependence of the diffusion constant for poly(ethylene oxide) on water concentration at 35°C; unfilled points sorption, filled points desorption.

MEASUREMENT OF ORGANIZATIONAL COMMUNICATIONS AND PROGRAM EFFECTIVENESS

HERMAN I. SHALLER
CHIEF, PROGRAMS BRANCH
POLICY AND ANALYSIS DIVISION
CIVIL WORKS - OFFICE, CHIEF OF ENGINEERS
WASHINGTON, D.C.

I. INTRODUCTION

The problem of planning national programs is a formidable task replete with dichotomies, paradoxes, incompatibilities, semantic and measurement barriers, and logical and psychological difficulties. Some of the broad subject areas which are applicable here are:

1. Philosophy of Planning
2. The Role of Psychology in Planning
3. The Role of Politics in Planning
4. The Role of Science and Technology
5. Socioeconomic and Technological Aspects
6. The Role of Economic and Technological Forecasting
7. Education and Training
8. Utilization of Existing Resources
9. Facilities as a Factor in Planning
10. Compatibility Problems of Diverse Goals and Objectives at the Various Levels; i.e.,individual, organizational, inter-organizational, and national
11. Qualitative Methods for Planning; i.e., intuition and judgment
12. Quantitative Tools and Techniques for Planning

One way of testing organizational plans, including those at the national level, is by inquiring into how the expenditure of resources is associated with desired contributions to a mission. At the national level, for example, we may be interested in and concerned with program contributions to such things as

1. the national economy
2. the national employment level
3. the national level of health and welfare
4. the national military posture
5. the national educational level

Part of the total problem, therefore, revolves around the extent to which a determination can be made concerning how well an ongoing program correlates with a stated mission -- not only in qualitative terms but in a quantitative way.

The central difficulty in attacking the problem of planning comprehensive scientific and technological programs at the agency levels in government, for example, stems from the absence of a comprehensive display of the ongoing national economic and social program. This difficulty, in turn, results from basic inadequacies of understanding, appreciation, knowledge, and wisdom, all of which are required for the development and display of an appropriate set of national goals and objectives. These, in turn, must be understandable and acceptable if we are to develop comprehensive national programs which are compatible with the ideals of democratic government.

At this juncture, when all the wisdom of the past has yet to be suitably interpreted and appropriately utilized with the problems as presently stated and understood, we must anticipate some faltering which will undoubtedly occur during any effort to cope with difficult and often unprecedented problems.

One possible approach to the above stated problem may be via a concept for projecting "program vectors" upon "mission vectors." This concept, along with a demonstration of methodological application, is the subject of this paper. Since the concept and proposed methodology represent a projection of work programs upon missions or program requirements, they may also, in effect, serve as a tool for measuring the degree to which total organizations communicate. In this connection, the paper demonstrates via a hypothetical experiment how one may determine in a quantitative manner the degree to which communications exist between the "management" of an organization and the "organizational performers."

II. BASIC CONCEPTS

1) The Mission Vector

Consider first any organization and a characterization of that organization (i.e., a family, community, business enterprise, government agency, a state, or the nation).

One way of characterizing the organization is by describing its mission and associating the description with some magnitude or quantitative measure such as resources, for example. This kind of characterization will be denoted by $\vec{M}$ and will be called the mission vector.

Associated with any $\vec{M}$ is a set of attributes (i.e., goals, objectives, functions, policy statements, aspirations, etc.), which, when clearly, appropriately, and properly weighted according to relative importance may be said to describe $\vec{M}$. Ideally, the unweighted set of attributes will form a logically exhaustive finite set of independent unit vectors,

$$\vec{a}_j, \quad (j = 1, 2, \ldots\ldots, n),$$

such that the summation of the weighted set defines the mission as follows:

$$(1) \quad \sum_{j=1}^{n} a_j \ \vec{a}_j = \vec{M}$$

where a_j is the magnitude of $\vec{a}_j$.

2) The Program Vector

Consider now the collection of activities associated with the organization (i.e., tasks, projects, areas of effort, etc.). Any arbitrary task, collection of tasks, project or collection of projects, etc., will be called a category and denoted by

$$\vec{C}_i, \quad (i = 1, 2, \ldots, m)$$

The i^{th} category then is simply an arbitrary activity or arbitrary collection of activities which, it is assumed for the moment, will behave as a vector. The magnitude of any $\vec{C}_i$ will be denoted by C_i and any category will be completely described by $C_i \ \vec{C}_i$.

Now consider the problem of describing the organizational program. Any organizational program will be defined as the aggregate or collection of all the activities associated with the organization or the enterprise and will be denoted by $\vec{P}$, the program vector. We now have the following additional relationships:

$$(2) \quad \vec{P} = P(\vec{M})$$

$$(3) \quad \vec{P} = \sum_{i=1}^{m} C_i \ \vec{C}_i$$

3) Program Effectiveness

A measure of effectiveness, E of $\vec{P}$ relative to $\vec{M}$ is obtainable if we define

$$(4) \quad E = \frac{\vec{P} \cdot \vec{M}}{\vec{M} \cdot \vec{M}}$$

If, as is usually the case, we claim to be rational and we say that our programs are responsive to the needs, goals, and objectives of our mission, then the projection of programs upon missions appears to be a natural measure of program effectiveness.

This definition of effectiveness should prove equally applicable and suitable for measuring performance relative to requirements if and when requirements are stated and specified in terms of a vector space.

4) Variability of $\vec{M}$ and $\vec{P}$

Changes in either $\vec{M}$ or $\vec{P}$ or both require the introduction of some function which will characterize the time rate of change of $\vec{M}$ or $\vec{P}$ or both. Accordingly, let $\vec{\emptyset}$ be some function which defines the time rate of change of $(\vec{M} - \vec{P})$ as follows:

$$(5) \quad \vec{\emptyset} = \frac{d}{dt} (\vec{M} - \vec{P})$$

The concepts concerning the existence of a mission vector, $\vec{M}$, a set of attributes, $\vec{a}_j$, a set of categories $\vec{C}_i$ and a program vector, $\vec{P}$ appear to be potentially useful ingredients for management research studies because, if it turns out that the quantities discussed do indeed behave as vectors in practice, then everything known about the behavior of vectors is immediately applicable to the behavior of missions and programs as defined in the preceding sec-

tions. The extremely powerful techniques of the algebra and calculus of vectors and tensors would then be at our disposal for pertinent studies along with the tools of "Boolean Algebra" and "Control Theory."

I am firmly convinced that any serious studies in depth in the area of management research will involve many other tools normally associated with the mathematical, engineering, and social sciences. As a matter of fact, management science can never advance beyond the frontiers of the social sciences.

III. A HYPOTHETICAL EXPERIMENT

A. Purpose of the Experiment

One aspect of the concepts under discussion is a kind of game which involves the director or chief executive of an organization vs. the staff of that organization. The gaming aspect stems from the fact that the chief executive is concerned primarily with the mission vector while the staff (as a whole) is concerned with the program vector. The problem here is that each individual staff member is concerned with only one, or at most, several components of the program vector. Furthermore, each staff member is concerned with his mission vector as he perceives it, which may or may not be consistent with the organization's mission vector. At best, each staff member contributes, via his efforts in a program, to only some of the components of the mission vector as perceived by the chief executive.

The question therefore arises as to how well the chief executive communicates the mission vector to the staff (as an entity), and how well each individual staff member communicates to the chief executive the fact that his effort contributes to specific components of the mission vector.

The experiment I propose, therefore, is one which will test the degree to which a chief executive and his staff are communicating. The organization here will be a hypothetical National Office for Research in Group Psychology.

B. Outline of the Experiment

1) The Mathematical Model

There are three basic ingredients to the proposed research program planning model. The first is the concept of a category-attribute matrix which describes numerical relationships between the attributes and the categories of research. The matrix yields a numerical output called "effectiveness." The second ingredient is a predetermined collection of classification systems which suggest a set of constraints for a balanced research program.

Classification systems offer an input in terms of which part of the planning is actually done. Hence, the classification systems are explicit interpretations of sets of boundary conditions. The third ingredient is a perturbation technique which is used to answer questions posed to the system. Changes in the state of a program may be made by perturbing the existing state and evaluating the new state for desired program balance and effectiveness.

The essentials of the proposed model are described below.

Nomenclature

C_i = the i^{th} category ($i = 1, 2, \ldots, m$)

A_j = the j^{th} attribute ($j = 1, 2, \ldots, n$)

$W(A_j)$ = relative importance of the jth attribute

$W(C_i)$ = relative importance placed upon the ith category

c_{ij} = relative contribution of the ith category to the jth attribute as viewed by the organizational manager

X = the budget

U_i = number of units of research in the ith category

K_i = cost per unit of research for the ith category

$1/K_i'$ = effectiveness of the ith category per dollar

x_i = budget allocation for the ith category

U_{ij} = number of units of research in the ith category associated with the jth attribute

a_{ij} = relative contribution of the ith category relative to the jth attribute as viewed by the organizational performer

e_{ij} = relative effectiveness of the ith category relative to the jth attribute

$\dot{E}_i$ = effectiveness of the ith category relative to all the attributes per unit of research

E_i = effectiveness of the ith category relative to all the attributes

E = effectiveness of the research program

Definitions and Ground Rules

$$\sum_i x_i = X$$

$$\sum_j W(A_j) = 1$$

$$a_{ij} = \frac{U_{ij}}{U_i}; \quad \sum_j a_{ij} = \sum_j \frac{U_{ij}}{U_i} = 1$$

$$W(A_j) = \frac{\sum_i U_{ij}}{\sum_i U_i}$$

$$W(C_i) = \frac{U_i}{\sum_i U_i}; \quad \sum_i W(C_i) = 1$$

$$c_{ij} = \frac{U_{ij}}{\sum_i U_{ij}}; \quad \sum_i c_{ij} = 1$$

$$e_{ij} = a_{ij}\, W(A_j)$$

$$\dot{E}_i = \sum_j e_{ij}$$

$$E_i = U_i \dot{E}_i$$

$$E = \sum_i E_i$$

The Concept of a Unit of Research

By definition, and for each category of research, let

u = 1 unit of research = 1 scientist-year

α = cost/scientist-yr/unit of research

v = other personnel required/unit of research

β = cost of v/unit of research

w = facilities required/unit of research (i.e., square feet of office space, square feet of lab space, etc.)

γ = cost of w/unit of research

y = required equipment allocations/unit of research (computers, radio telescopes, electron microscopes, ships, planes, etc.)

δ = cost of y/unit of research

Postulates

1. $x_i = f(E_i) = K_i' E_i$, states that the amount of allocation of funds for any category of research is a function of the effectiveness of that category of research.

2. $x_i = K_i U_i$, where $K_i = f(C_i)$, states that the allocation of funds for the ith category is determined by the number of research units and the cost per unit of research associated with that category.

3. $K_i = f(C_i) = \alpha u + \beta v + \gamma w + \delta y + \ldots$, states that there is a cost per unit of research associated with each category.

Relationships between K_i, K_i', and E_i

From Postulates 1 and 2, we have $K_i' E_i = K_i U_i$ or

$$\frac{K_i}{K_i'} = \frac{E_i}{U_i} = \frac{U_i \dot{E}_i}{U_i} = \dot{E}_i$$

where

$\dot{E}_i$ = effectiveness per unit of research of the ith category

K_i = cost per unit of research for the ith category

$1/K_i'$ = effectiveness of the ith category per dollar

Theorems (Proofs follow directly from definitions, ground rules, and postulates)

I. $$E_i = \frac{x_i}{K_i} \sum_j e_{ij} .$$

II. $$E = \sum_i \sum_j \frac{x_i}{K_i} e_{ij} .$$

III. $$x_i = \frac{XK_i\, W(C_i)}{\sum_i K_i\, W(C_i)} .$$

IV. $$a_{ij}\, W(C_i) = c_{ij}\, W(A_j)$$

V. $$W(A_j) = \sum_i a_{ij}\, W(C_i) .$$

Model of the State of a Research Program

Figures 1-4 are the arrays which may be used to characterize the nature and the state of a research program at any time t.

X	K	U	C	A_1	A_2	...	A_j	...	A_n
x_1	K_1	U_1	C_1	U_{11}	U_{12}	...	U_{1j}	...	U_{1n}
x_2	K_2	U_2	C_2	U_{21}	U_{22}	...	U_{2j}	...	U_{2n}
.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	...	.	...	.
.	.	.	.	.	.	.	.	.	.
x_i	K_i	U_i	C_i	U_{i1}	U_{i2}	...	U_{ij}	...	U_{in}
.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	...	.	...	.
.	.	.	.	.	.	.	.	.	.
x_m	K_m	U_m	C_m	U_{m1}	U_{m2}	...	U_{mj}	...	U_{mn}

Fig. 1 - The U_{ij} array: a display of the distribution of units of research and dollars at any time t

W(C)	A / W(A) / C	A_1 / $W(A_1)$	A_2 / $W(A_2)$	...	A_j / $W(A_j)$	...	A_n / $W(A_n)$
$W(C_1)$	C_1	a_{11}	a_{12}	...	a_{1j}	...	a_{1n}
$W(C_2)$	C_2	a_{21}	a_{22}	...	a_{2j}	...	a_{2n}
.	.	.	.	.	.	.	.
.	.	.	.	...	.	...	.
.	.	.	.	.	.	.	.
$W(C_i)$	C_i	a_{i1}	a_{i2}	...	a_{ij}	...	a_{in}
.	.	.	.	.	.	.	.
.	.	.	.	...	.	...	.
.	.	.	.	.	.	.	.
$W(C_m)$	C_m	a_{m1}	a_{m2}	...	a_{mj}	...	a_{mn}

Fig. 2 - The a_{ij} array: characterizes the nature of the research under way in each category at any time t

C	A_1	A_2	...	A_j	...	A_n	$\dot{E}$	E
C_1	e_{11}	e_{12}	...	e_{1j}	...	e_{1n}	$\dot{E}_1$	E_1
C_2	e_{21}	e_{22}	...	e_{2j}	...	e_{2n}	$\dot{E}_2$	E_2
.	.	.	.	.	.	.	.	.
.	.	.	...	.	...	.	.	.
.	.	.	.	.	.	.	.	.
C_i	e_{i1}	e_{i2}	...	e_{ij}	...	e_{in}	$\dot{E}_i$	E_i
.	.	.	.	.	.	.	.	.
.	.	.	...	.	...	.	.	.
.	.	.	.	.	.	.	.	.
C_m	e_{m1}	e_{m2}	...	e_{mj}	...	e_{mn}	$\dot{E}_m$	E_m

Fig. 3 - The e_{ij} array: may be characterized as an "effectiveness" measurement model

C	A_1	A_2	...	A_j	...	A_n
C_1	c_{11}	c_{12}	...	c_{1j}	...	c_{1n}
C_2	c_{21}	c_{22}	...	c_{2j}	...	c_{2n}
.	.	.	.	.	.	.
.	.	.	...	.	...	.
.	.	.	.	.	.	.
C_i	c_{i1}	c_{i2}	...	c_{ij}	...	c_{in}
.	.	.	.	.	.	.
.	.	.	...	.	...	.
.	.	.	.	.	.	.
C_m	c_{m1}	c_{m2}	...	c_{mj}	...	c_{mn}

Fig. 4 - The c_{ij} array: characterizes the nature of the research under way at any time t by attributes and any other classifications as may be required for program balance

2) Categories, Attributes, and Classification Systems

(a) Categories.

A category describes an area of activity or effort. It may consist of a single task or a collection of tasks which have a common center of interest. The category should provide a convenient unit of effort for planning purposes at the working level. One possible set of categories for use in describing the state of a program is indicated below:

Assume first the existence of an organization with a research mission such that the organization contains a physical sciences division composed of several branches. Let one of these branches be electronics composed of sections identified as Electronic Systems, Physical Electronics, Electromagnetic Wave Propagation and Radiation, Electronic Components, and Communications Theory.

The above sections may constitute convenient units of effort for planning purposes for some organizations, but too general or not specific enough for others. Assuming the latter, then each of the sections may be further subdivided into more convenient or smaller units of effort. For example, communications theory may be factored into smaller areas of activity and identified as circuit analysis and synthesis, information theory and coding, data processing, and bionics.

The limit of this factoring process is approached as we begin to identify precisely defined tasks or activities.

(b) Attributes.

An attribute is a "reason" for support of a category. The attribute should describe a characteristic property of a portion of the research mission in such a manner that the summation of all attributes, appropriately weighted, describes the mission in practice. One possible set of attributes may consist of terms or expressions like scientific value, technological value, military value, coordination of effort, dissemination of information, support of a profession, prestige, and pressures and/or interests of customers.

(c) Classification Systems.

The concept of a "Balanced Program" is considered to be related to sets of classification systems. If we define a balanced program as one which includes an appropriate amount of effort in those areas or fields of service in which the organization may reasonably be expected to engage, then the use of appropriate classification systems offers a technique for determining whether a program is adequately balanced. Appropriate and acceptable classification systems, therefore, suggest a set of constraints or a set of guidelines which offer an input in terms of which part of the planning may actually be done. Examples of some possible classification systems are as follows:

1. The set of attributes listed in paragraph (b) above
2. Chronological
   a. Short range
   b. Intermediate range
   c. Long range

3. Environmental
   a. Undersea
   b. Earth's surface
   c. Aerospace
4. Operations
   a. Surveillance
   b. Command control
   c. Intelligence
   d. Logistics
5. Spectrum of Scientific and Technological Effort
   a. Pure research
   b. Basic research
   c. Applied research
   d. Exploratory development
   e. Development and design
   f. Production
   g. Operations
   h. Maintenance
6. Customer Interest (as appropriate and/or applicable)

(d) For our experiment with the hypothetical National office for Research in Group Psychology we will concern ourselves with only two sets of classification systems:

The attributes which describe the mission at the level of the chief executive.

The categories which describe the program from the program manager's points of view.

3) The Procedure for Data Collection

Based upon personal interviews with the chief executive and the various program managers, the following information and data are considered obtainable.

(a) The chief executive describes his mission vector -- this is the set of attributes which he believes describes his mission in practice, along with the relative importance which he assigns to each of the attributes (i.e., he weights the attributes).

(b) Each of the program managers describes his program vector -- this is the set of categories for which each manager has responsibility as he perceives it. He also reveals the amount of money presently committed to each category, along with the number of units of research (scientist-years) associated with each category.

4) Data Collection

The nature and kind of information which may be received follows:

ORGANIZATIONAL MISSION VECTOR

Listing and Weighting of Attributes

Generate principles which apply to behavior of groups varying in size from two people to those of large organizations (15).

Generate principles which apply to groups in special situations (confinement, isolation, stresses of cold war, guerilla warfare, combat) (20).

Develop techniques useful in composition, training, and management of groups (15).

Develop techniques useful for special purpose group behavior (conflict resolution, negotiation, group therapy, resistance to "brain washing," etc.) (20).

Dissemination of information to the appropriate scientific communities (15).

Dissemination of information to the appropriate military channels (15).

ORGANIZATION PROGRAM VECTOR

List of Categories	Dollars Committed to Program Categories	Units of Research Associated with Program Categories
Individual behavior in the group	1,500,000	100
Group effectiveness	1,200,000	86
Interaction among groups	1,150,000	85
Group reinforcement	730,000	47
Motivation	420,000	28
Behavioral aspects of conflict and warfare	920,000	51
Special methods for behavior modifications	860,000	57
Behavioral aspects of power	450,000	32
Group cooperation and power	430,000	25
Cross-cultural research	450,000	60

5) Data Assembly, Computations, and Production of Information

Utilizing the foregoing data, the mathematical model may be implemented as shown in the following tables.

Table I - is the a_{ij} array developed in accordance with Figure 1. It characterizes the nature of the research under way at the present time. This table also shows the attribute weightings as given by the chief executive vs. the weightings as computed from the nature of the ongoing research program.

Table II - is the u_{ij} array developed in accordance with Figure 2. It displays the distribution of units of research and dollars at the present time.

Table III - is the e_{ij} array developed in accordance with Figure 3. It measures (by definition) the "effectiveness" of each category relative to each attribute, the effectiveness rate, E Dot, of each category relative to the mission per unit of research, and the "total current effectiveness" of each category relative to the total mission.

Table IV - is the c_{ij} array developed in accordance with Figure 4. It characterizes the nature of the research under way by attributes, which is the display most desired by the chief executive.

Table I. a_{ij} Array

		Given	.1500	.2000	.1500	.2000	.1500	.1500
		Computed	.1316	.2180	.1863	.1906	.1481	.1254
C	W (C)		W(A1)	W(A2)	W(A3)	W(A4)	W(A5)	W(A6)
1	.1751	W(C 1)	.0500	.1500	.3000	.2000	.2000	.1000
2	.1506	W(C 2)	.2000	.2000	.2000	.1000	.2000	.1000
3	.1489	W(C 3)	.2000	.3000	.1500	.1500	.1000	.1000
4	.0823	W(C 4)	.1000	.3000	.2000	.2000	.1000	.1000
5	.0490	W(C 5)	.2000	.2000	.2000	.1000	.1000	.2000
6	.0893	W(C 6)	.1000	.3000	.1000	.3000	.1000	.1000
7	.0998	W(C 7)	.0500	.2000	.2000	.3000	.1000	.1500
8	.0560	W(C 8)	.1000	.2000	.2000	.2000	.1500	.1500
9	.0438	W(C 9)	.1000	.2000	.1000	.3000	.1500	.1500
10	.1051	W(C 10)	.2000	.1500	.1000	.1500	.2000	.2000

Table II. u_{ij} Array

X	K	U	C	A1	A2	A3	A4	A5	A6
1,500,000	15000	100	1	5.0	15.0	30.0	20.0	20.0	10.0
1,200,000	13953	86	2	17.2	17.2	17.2	8.6	17.2	8.6
1,150,000	13529	[illegible]	3	17.0	25.4	12.8	12.8	8.5	8.5
730,000	15532	47	4	4.7	14.1	9.4	9.4	4.7	4.7
420,000	15000	28	5	5.6	5.6	5.6	2.8	2.8	5.6
920,000	18039	51	6	5.1	15.3	5.1	15.3	5.1	5.1
860,000	15088	57	7	2.9	11.3	11.4	17.1	5.7	8.6
450,000	14063	32	8	3.2	6.4	6.4	6.4	4.8	4.8
430,000	17200	25	9	2.5	5.0	2.5	7.4	3.8	3.8
450,000	7500	60	10	12.0	9.0	6.0	9.0	12.0	12.0
8,110,000		571		75.2	124.3	106.4	108.8	84.6	71.7

Table III. e_{ij} Array

C	A1	A2	A3	A4	A5	A6	E Dot	E(I)
1	.0075	.0300	.0450	.0400	.0300	.0150	.1675	1.68
2	.0300	.0400	.0300	.0200	.0300	.0150	.1650	1.42
3	.0300	.0600	.0225	.0300	.0150	.0150	.1725	1.47
4	.0150	.0600	.0300	.0400	.0150	.0150	.1750	.82
5	.0300	.0400	.0300	.0200	.0150	.0300	.1650	.46
6	.0150	.0600	.0150	.0600	.0150	.0150	.1800	.92
7	.0075	.0400	.0300	.0600	.0150	.0225	.1750	1.00
8	.0150	.0400	.0300	.0400	.0225	.0225	.1700	.54
9	.0150	.0400	.0150	.0600	.0225	.0225	.1750	.44
10	.0300	.0300	.0150	.0300	.0300	.0300	.1650	.99

Table IV. c_{ij} Array

C	A1	A2	A3	A4	A5	A6
1	.0665	.1205	.2821	.1837	.2365	.1397
2	.2289	.1382	.1617	.0790	.2034	.1201
3	.2262	.2048	.1199	.1171	.1005	.1187
4	.0625	.1133	.0884	.0864	.0556	.0656
5	.0745	.0450	.0527	.0257	.0331	.0782
6	.0679	.1229	.0480	.1406	.0603	.0712
7	.0379	.0916	.1072	.1571	.0674	.1194
8	.0426	.0514	.0602	.0588	.0568	.0670
9	.0333	.0402	.0235	.0689	.0444	.0524
10	.1597	.0723	.0564	.0827	.1419	.1676

V. DISCUSSION OF RESULTS

A central question -- one of continuous concern -- is, to what extent does an ongoing program contribute to an entire mission? Alternatively stated, one may ask how well the chief executive communicates with his staff, since the ongoing program is bound to reflect the extent to which such communicability exists in practice.

Table I provides a quantitative determination to the above question by comparing the GIVEN data with the COMPUTED data. For the hypothetical organization shown, the good correlation between the given and computed data (lines 1 and 2 of Table I) indicates that the ongoing program closely matches the stated mission. It may therefore be concluded that communications between the chief executive and his staff are good.

The assumption in this exercise is that we have a very closely knit group (all are concerned with and are interested in group psychology). What kind of results would be obtained where the chief executive is the President of the United States and the programs are those of all Federal Agencies and Departments? What would the results be at even much lower national organizational levels, or at various corporate and other institutional levels? How would the mission vector be stated and how well would ongoing programs match organizational missions where the missions consist of a diversity of goals, objectives, and policy statements?

Computation of effectiveness in accordance with equation (5) indicates an effectiveness number close to unity for the hypothetical case considered. This is because no account was made for how efficiently each program was being conducted. The computations for this simplified illustration assumed 100 percent efficiency for each contribution from each category to each attribute. For real programs, efficiency would, of course, have to be determined and the summation of contributions from each category to each attribute would be something less than unity. The principle of entropy, whether in thermodynamics or communications theory, and as may be utilized in social or political systems theory, precludes any real process, as we know it, from operating at 100 percent efficiency.

VI. CONCLUSION

1. Equations (1), (2), and (3) describe basic relationships

associated with all rational mission-program structures. Solution of these equations yields a description of the present state of a program relative to a specified set of goals and objectives.

2. Equation (4) represents a precise and unambiguous definition of program effectiveness. It offers a basic standard for program effectiveness measurement and permits rational comparisons of alternative mission-program combinations. The process of effectiveness measurement, in view of the nature of the process, includes a measure of the extent to which organizational communications exist between the "organizational management" and the "organizational performers."

3. Equation (5) defines a function for measuring the relative effects of rates of change in missions or programs or both. An arbitrary datum for $\vec{\emptyset}$ is the zero vector. Obviously, when $\vec{P} = \vec{M}$, $E = 1$ and $\vec{\emptyset} = \vec{0}$. This state, which is considered unattainable in practice, represents an ideal analytical condition. It defines a point where stable equilibrium exists between missions and programs. By definition, it is at this utopian point that programs may be thought of as being conducted at 100 percent efficiency at a particular point in time.

References

1. Churchman, C. W., Ackoff, R. L., and Arnoff, E. L., Introduction to Operations Research. Wiley, 1957.

2. Gibbs, J. Willard, and Wilson, Edwin Bidwell, Vector Analysis. Yale University, 1901.

3. Ledley, R. S., Shaller, H. I., Rotolo, L. S., Wilson, J. B., "Methodology to Aid Research Planning," IEEE Transactions on Engineering Management, June 1967.

4. National Academy of Sciences, Basic Research and National Goals. March, 1965.

5. National Academy of Sciences, Applied Science and Technological Progress. June, 1967.

6. National Commission on Technology, Automation, and Economic Progress, Technology and the American Economy, February, 1966.

7. Shaller, H. I., "An Exploratory Study in Research Planning Methodology," ONR Report ACR/NAR-27, September, 1963.(AD 421895)

8. _____., "Management Vectors," Copyright 1965.

9. _____., "A General Theory and Technique for Measurement of Organizational Communications and Program Effectiveness," Copyright 1967.

RESEARCH GOALS FOR IMPROVED BLOOD LOGISTICS

CHARLES E. SHIELDS, LT COLONEL, MC;
FRANK R. CAMP, JR., LT COLONEL, MSC;
and GEORGE H. SEEGER
US ARMY MEDICAL RESEARCH LABORATORY
FORT KNOX, KENTUCKY

In the military supply chain, "blood, bullets, and bread" can be considered to represent the main essential areas of both practical as well as emotional concern. Blood and bread--both are necessary for life--but it is the blood that must be replaced the fastest, and it is blood that is the most critically perishable. What is critical is:

1. the blood can only be obtained from our fellow man,
2. the cells in the blood have a definite life span so that there is a continual loss of the living cells,
3. the injuries of war are usually considerably more severe compared to civilian life and the medical treatment will often involve massive transfusions--thirty, forty, or even more pints for one man, and
4. the blood itself has certain characteristics called blood groups that must be matched to the man receiving the transfusion or this, in itself, will result in reactions, or even death.

Thus, the blood must be collected carefully from a donor and quickly transported to the side of the casualty in good, safe condition for the transfusion.

BACKGROUND

Shortly after the turn of the century, the importance of the blood type and the need to match this type between the donor and recipient was recognized by Karl Landsteiner (1), (this year representing the 100th anniversary of his birth). This major step, implementing the fundamental realization that man's blood could only be replaced by another man's, laid the groundwork for present-day transfusion practices. However, one important aspect remained as

a problem. Freshly drawn blood was very useful for transfusion, but deterioration occurred very rapidly and, to avoid this, most early transfusions were direct from person-to-person which was often hazardous to the donor. Indirect transfusion became possible when blood could be collected in an anticoagulant solution, and by World War II the most promising was acid-citrate-dextrose, or ACD. Specific comparison testing of several solutions, particularly the survival of the stored red blood cells in the recipient, indicated the value of ACD (2, 3, 4, 5). Blood in ACD had usually better than 70% 24-hour posttransfusion survival even after 21 days of storage; therefore, 21 days was accepted as a practical shelf life which allowed the development of blood banks.

Storage of blood still had difficulties when military needs were included. Bank blood could be retained in good condition through the 21-day storage period if carefully refrigerated and left essentially undisturbed during storage (6). Thus, overseas blood supply route requirements had to include: 1) fast delivery to meet the 21-day shelf life limitation, 2) refrigeration, and 3) protection against the physical stresses of shipment, logistical conditions that were still not completely met in World War II and in Korea.

Despite considerable scientific advance in other fields since that time, shipped blood is still a logistical problem. The blood is drawn from donors, using ACD collection solution, and stored in blood banks until used, or for not more than 21 days. The only real changes have been in collection containers which originally were vacuum-type glass bottles and now are gravity-filled plastic bags; and the shipping box of World War II, though gradually improved over the years, is still under evaluation. Despite faster flight times, this factor may seldom be usable under battle conditions.

The Blood Transfusion Division, USAMRL, Fort Knox, Ky., has been evaluating new blood banking developments to anticipate new concepts. The present report combines recent research in which a normal body chemical added to stored blood has led to better preservation and longer shelf life, with the possible effects this may have on military blood logistics.

INCREASING SHELF LIFE

Though no major change has been made in the collection solution since World War II, many efforts to improve the solution have been under study. Most of these have concentrated on protecting the red blood cell by either improving the ionic content of the

solution, or by adding some type of a chemical preservative. In the first category, ACD has been modified into a new form of solution, citrate-phosphate-dextrose (CPD). The CPD solution has been reported to provide blood with survival levels at 28 days that are comparable to levels seen with 21-day-old ACD units (8). In the second category, most chemicals added to blood have been related to their possible ability to maintain a better environment for the red blood cell, particularly involving adenosine-tri-phosphate (ATP), the main energy source for the cells (9). First on the list was adenosine (10), but on clinical trial it caused lowering of the blood pressure in recipients (11). Other similar chemicals have been tried including inosine (12), guanosine (13), and adenine (14-18). Adenine, initially tested in 1959 by Nakao (14), and by Simon (15) in 1962, showed that blood collected in ACD with adenine had survival levels above 70% even after 46 days of storage. A similar pattern of improved survivals over a longer shelf life has been found by other laboratories (16,17).

A group from Sweden (18) has been using ACD-adenine solution in their blood bank and have transfused over 5000 units, at least 500 of which were transfused after storage over 21 days. Some reactions have occurred, but the incidence rate was not different from that reported with ACD units.

The Army Adenine Program has also been evaluating ACD-adenine at WRAIR, Ft Knox, Germany, Puerto Rico, and in cooperation with civilian laboratories as well. These have consisted of tests for metabolic changes, determination of possible clinical toxicity in various animals including man, as well as detailed evaluation of the blood transfused with the following program being an example.

Most of the programs relied upon the in vivo 24-hour post-transfusion survival of the stored blood as the primary guideline, with various in vitro tests included for secondary evidence. At Ft Knox, the in vivo survival was carried out using two isotopes. Phosphorus-32 was used to label fresh cells drawn from the recipient on the day of transfusion (19). The unit of stored blood was labeled with chromium-51 (20). Both isotopes were infused and the fresh cell activity permitted a correction calculation of the initial activity of the stored cells. The radioactivity of frequent samples drawn from the recipient over the next two to three weeks was measured and compared to the initial level and expressed as a per cent on semilog paper against time. The points were joined by a linear approximation and used to indicate the estimated 24-hour

SHIELDS, CAMP, and
SEEGER

posttransfusion survival percentage.

The in vitro tests included plasma hemoglobin, bilirubin, potassium, hematocrit, and osmotic fragility (21), in particular, with uric acid and other chemical tests and urinalysis being included as needed. The subjects were evaluated clinically for possible toxic effects. The stored blood was sampled at the time of transfusion, and at other sampling times when available. Blood from the stored unit was cultured at three temperatures--0°, 25°, and 37° C to rule out pathological bacterial contamination, with only two out of over 600 cultures being found positive and these units were not given.

The basic design of the experiment consisted of units collected in one of four types of solutions: ACD, ACD with adenine, CPD, or CPD with adenine (adenine concentration .5 mM, prepared sterilely by Fenwal). Three time periods were used--28, 35, or 42 days of storage. To reduce variability, the original donor received his own blood in an autologous transfusion.

The results of the in vivo survival tests are shown in the graph (Fig. 1), and these group survival averages tended to fit a curve. All survivals at 28 days were over 70%. Those units with either ACD or CPD alone were associated with average survival value of less than 70% by 35 days or longer. In contrast, the average survival value did not drop below 70% even at 42 days in blood collected in adenine-supplemented ACD or CPD. The in vitro studies also indicated a greater protective effect with adenine than in the anticoagulant alone.

Other programs are in progress in both our laboratories and with others using patients in which adenine has been shown to be associated with good survivals after long storage. Most of the transfusions have been less than four units or pints; however, so far no side effect has been associated with adenine.

IMPACT OF INCREASED SHELF LIFE

With the scientific evidence accumulated, it is becoming more obvious that adenine or some other material will provide a practical increase in the shelf life of stored blood. As far as can be predicted on the basis of present data, adenine supplementation will probably provide an extension of the shelf life to at least 35 days and possibly 42 days in an emergency. It is then necessary to consider the secondary effects of such an extension.

SHIELDS, CAMP, and
SEEGER

First is the matter of crossmatching the blood. The present Army blood program procedure does require the collection of two extra tubes of clotted blood, then attaching these to the bag for use in crossmatching, even though many follow the procedure used in civilian banks in which part of the collection tubing is divided into segments and these are used for crossmatching (Fig. 2). Because these remain attached to the unit and are clearly identified with the unit, many obvious clerical errors can be avoided. On serial testing, the crossmatch with the clotted specimen, whether in the test tube or collection tubing, began to be erratic and essentially unusable by 28 days. By some adjustments in the procedure, it was possible to crossmatch plasma rather than the serum from the clotted blood. Thus, by using unclotted blood in the non-segmented collection tubing, which would be as preserved as the blood in the adenine unit, it would be possible to obtain suitable samples for crossmatching for as long as the unit is stored.

Second, the actual storage time presents a mixed blessing. The shelf space required to store a wide variety of different blood types in a civilian bank would have to be increased if the shelf life is lengthened, since rare blood types would be kept until the last possible moment. For the military, this may be a major problem in the field hospital with limited space. One expedient may be useful, however. Since low-titer group O negative blood can be called universal donor type because it can be used for emergency transfusion of all blood types, it would be possible for any forward medical installation to stock their bank with only this type of blood. In the rest of the evacuation chain, more specific types may be stored, but again the rotation of stock may depend upon other medical problems. For example, the blood stored for 30-40 days may not tolerate such mechanical stresses as occur in the artificial lung or heart pump (22). Because there is a natural loss of many other components during storage, including platelets and labile clotting factors, major evacuation hospitals dealing with the unusual bleeder may require either the specific components, if known, or have to use fresh blood in an emergency. Thus, it is apparent that simply increasing the shelf life does not make the blood bank operation simpler. For this reason, another program is under way to determine the usage probability factor. Basically, the program is attempting to define the source and type of demand which may permit establishing a central source for all blood types with secondary distribution channels designed to meet both sporadic unusual needs as well as critically heavy loads in multiple battle areas.

Third, considering that the supply would be partially

dictated by the usage, it still remains a problem in transportation. The present shipping box could hardly be used in the following mission: a small isolated unit requires four units of blood for casualties until they can be carried to an evacuation area--deliver units within one hour--avoid enemy fire and try to avoid landing because of unstable ground conditions. We would like to illustrate such a mission on film using a simulated flight and drop.

Film strip: These are illustrative film clips of air deliveries of blood using standard blood boxes and new types of containers. Some were released in free fall from heights of 60 to 500 feet with forward speed of 120 knots. The severe trauma of free fall can be overcome by a new container, but this can only hold 1 - 5 blood bags and does not have refrigeration. The cargo chute can handle most of the containers, but does present a problem if the air delivery needs to avoid enemy observation.

During this study, complaints were received that individual plasma units containing fresh frozen plasma were frequently breaking during shipping (Fig. 3). Fortunately, a new type of packing material was found which could be wrapped around individual units and satisfactorily prevent the breakage (Fig. 4). The material is made of two sheets of plastic so joined as to form bubbles of air (Air Cap Co.). The material also provided sufficient additional protection in the standard boxes to permit satisfactory air drop. There still remains the development of an adequate temperature-controlled shipping box, as well as a recording temperature monitor. Investigation along this line is in progress including designing a box in which the refrigerant could be changed without involving the basic blood units. Then, once the shipping container is considered adequate, the problem of identifying shipments as blood or other critical medical or technical supplies could be eased by a cargo coding system using a sign system or prominent use of colors.

SUMMARY

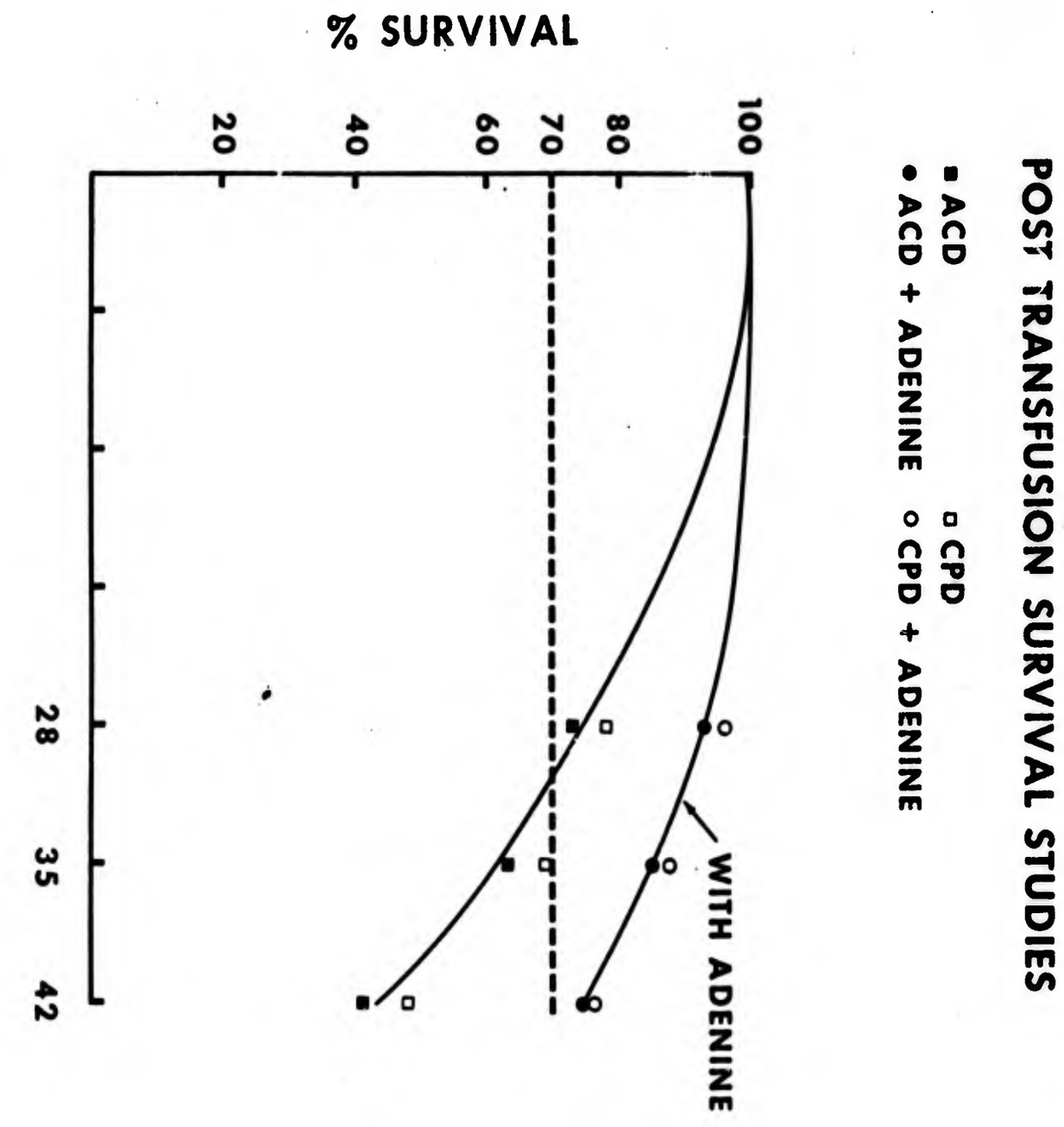
The need for blood for the battle casualty has created many problems. The natural perishability, the limited human source for blood, and the need for a safe therapeutic agent are the major areas. The changes in collection methods and solutions used for blood storage are being studied and the results of one particular chemical, adenine, were reported. Based on these results and other reports, the length of useful blood storage may be lengthened from 21 days to perhaps 42 days. As better collection and preservation methods

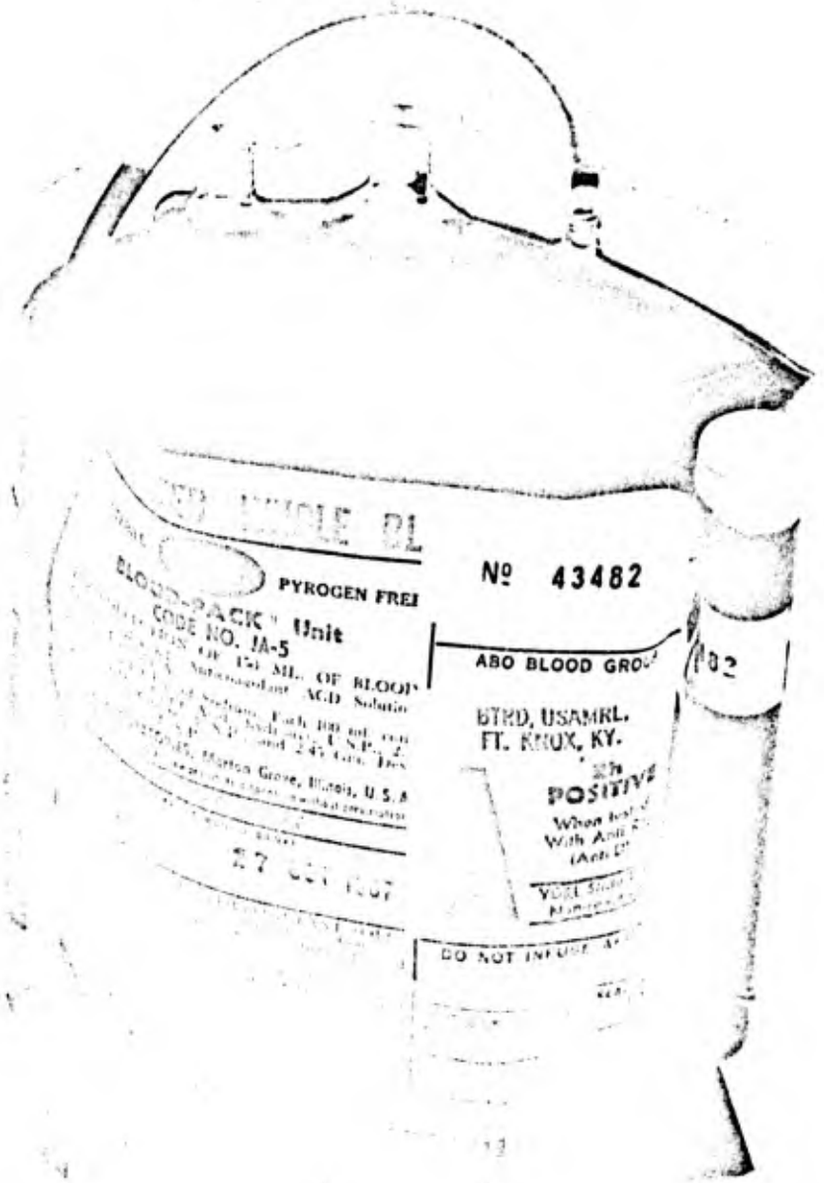
evolve, these, in turn, require better shipping and transportation schemes. Some of these areas under active study were discussed. The crossmatching procedure can already be improved by existing civilian practice, and by using unclotted blood in the new collecting solution, the test still can be performed even over the longer storage periods. The increased shelf life will require better bank inventory procedures which may be aided by a better understanding of the usage probability. Transporting the blood requires safe and refrigerated containers and the use of a variety of forms of delivery. Newly designed containers and packing materials have been studied as a basis for these new requirements, and the movement can be expedited by a cargo coding system. Thus, each logistical phase must be improved to meet the overall goal of efficiently supplying the highest quality blood.

REFERENCES

1. Landsteiner, K.: Contribution to the knowledge of the anti-enzymatic, lytic and agglutinating effects of blood serum and lymph. (Zur Kenntnis der antifermentativen, lytischen and agglutinierenden Wirkungen des Blutserums und der Lymphe.) Centr. Bakt. Orig. 27: 357-366, 1900.
2. Mollison, P.L.: Blood Transfusion in Clinical Medicine, 4th ed., Blackwell Scientific Pub., Oxford, England, 1967.
3. Gibson, J.G., 2nd, R.D. Evans, J.C. Aub, T. Sack, and W.C. Peacock: The post-transfusion survival of preserved human erythrocytes stored as whole blood or in re-suspension, after removal of plasma, by means of two isotopes of radioactive iron. J. Clin. Invest. 26: 715, 1947.
4. Rapoport, S.: Dimensional, osmotic, and chemical changes of erythrocytes in stored blood. II. Evaluation of several acid and neutral preservation mixtures. Effect of storage at 25° C. in Alsever's solution. J. Clin. Invest. 26: 616, 1947.
5. Strumia, M.M., A.D. Blake, Jr., and J.J. McGraw, Jr.: An acid-citrate-dextrose solution with low water volume and low dextrose concentration. J. Clin. Invest. 26: 678, 1947.
6. Ross, J.F., C.A. Finch, W.C. Peacock, and M.E. Sammons: The *in vitro* preservation and post-transfusion survival of stored blood. J. Clin. Invest. 26: 687, 1947.
7. Kendrick, D.B., Brigadier General, MC, USA: Blood Program in World War II. Office of The Surgeon General, Department of the Army, Washington, D. C., 1964.
8. Gibson, J.G., C.B. Gregory, and L.N. Button: Citrate-phosphate-dextrose solution for preservation of human blood. A further report. Am. J. Clin. Path. 28: 569, 1957.

9. Simon, E.R.: Adenine and purine nucleosides in human red cell preservation. A review. Transfusion 7: 395, 1967.
10. Gabrio, B.W., D.M. Donohue, and C.A. Finch: Erythrocyte preservation. V. Relationship between chemical changes and viability of stored blood treated with adenosine. J. Clin. Invest. 34: 1509, 1955.
11. Clarke, D.A., J. Davoll, F.S. Philips, and C.B. Brown: Enzymatic deamination and vasodepressor effects of adenosine analogs. J. Pharmacol. Exp. Ther. 106: 291, 1952.
12. Gabrio, B.W., D.M. Donohue, F.M. Huennekens, and C.A. Finch: Erythrocyte preservation. VII. Acid-citrate-dextrose-inosine (ACDI) as a preservative for blood during storage at 4° C. J. Clin. Invest. 35: 657, 1956.
13. Prankerd, T.A.J.: Revival of stored blood with guanosine and its successful transfusion. Lancet 1: 469, 1956.
14. Nakao, M., T. Nakao, M. Tatibana, H. Yoshikawa, and T. Abe: Effect of inosine and adenine on adenosine triphosphate regeneration and shape transformation in long-stored erythrocytes. Biochim. Biophys. Acta 31: 564, 1959.
15. Simon, E.R., R.G. Chapman, and C.A. Finch: Adenine in red cell preservation. J. Clin. Invest. 41: 351, 1964.
16. de Verdier, C.H., L. Garby, M. Hjelm, and C. Hogman: Adenine in blood preservation: Posttransfusion viability and biochemical changes. Transfusion 4: 331, 1964.
17. Bishop, C.: Some in vitro effects of adenine added to stored blood. Transfusion 4: 265, 1964.
18. Akerblom, O., C.H. de Verdier, M. Finnson, L. Garby, C.F. Hogman, and S.G.O. Johansson: Further studies on the effect of adenine in blood preservation. Transfusion 7: 1, 1967.
19. Mollison, P.L., M.A. Robinson, and D.A. Hunter: Improved method for labeling red cells with radioactive phosphorus. Lancet 1: 766, 1958.
20. Gibson, J.G., and W.A. Scheitlin: Method employing radioactive chromium for assaying the viability of human erythrocytes returned to the circulation after refrigerated storage. J. Lab. Clin. Med. 46: 679, 1955.
21. Shields, C.E., T.E. Allen, and B.J. Fogel: Evaluation by continuously recording hypotonic hemolysis of various normal and pathological erythrocytes. Am. J. Med. Sci., Vol 251, No. 6: (62/656), June, 1966.
22. Bernstein, E.D.: Sublethal damage to the red blood cell from pumping. Circulation, Supp I 36: 1-266, 1967.


POST TRANSFUSION SURVIVAL STUDIES
■ ACD
□ CPD
● ACD + ADENINE
○ CPD + ADENINE
% SURVIVAL
100
80
70
60
40
20
WITH ADENINE
28
35
42
DAYS OF STORAGE


PYROGEN FREE
Nº 43482
Unit
CODE NO. JA-5
ABO BLOOD GRO
BTRD, USAMRL.
FT. KNOX, KY.
POSITIVE

CODE NO.
TA-1

Fenwal

600 ML.
CAPACITY

TRANSFER PACK

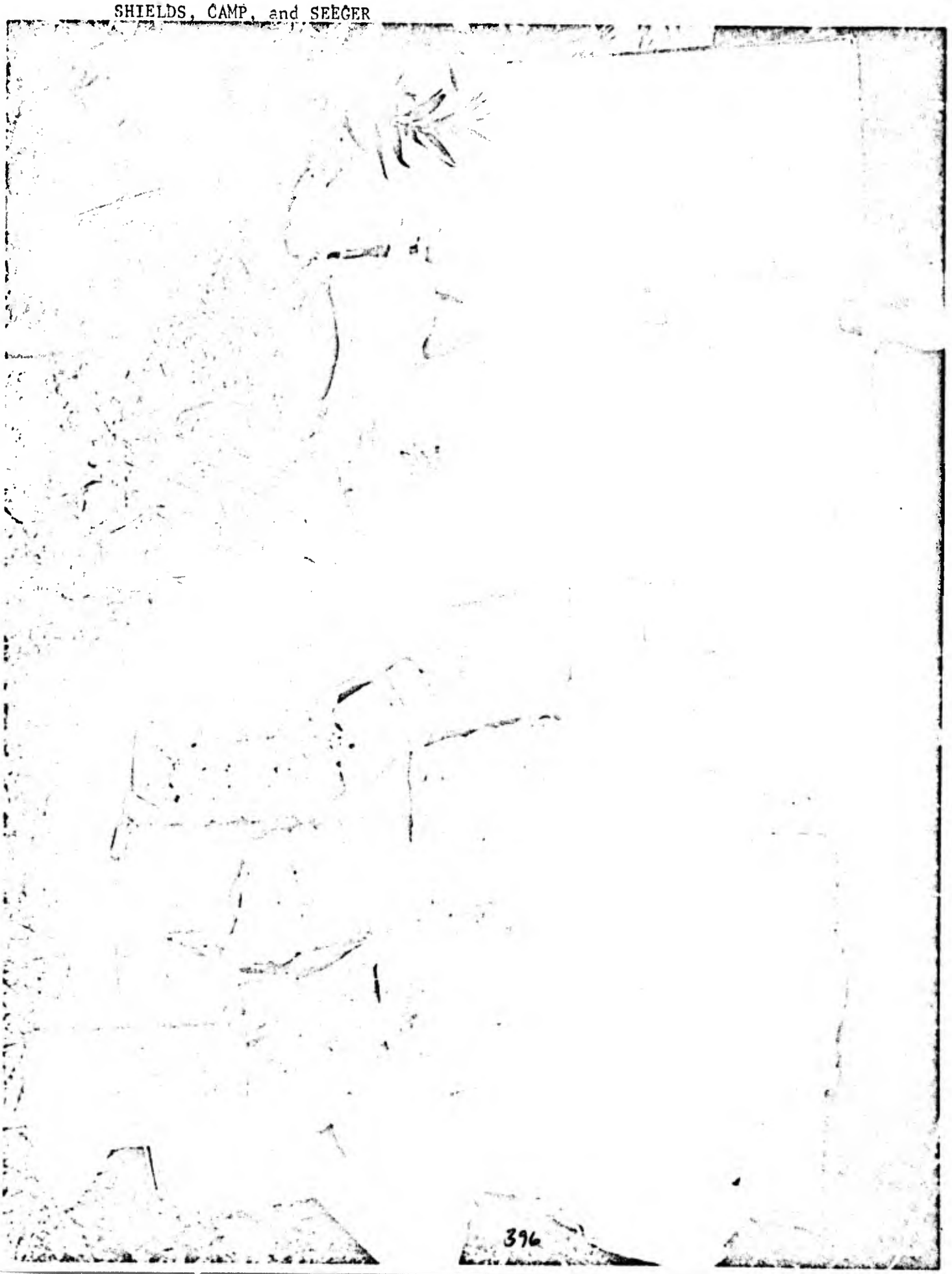
Container is STERILE, PYROGEN FREE

CAUTION: Federal (U.S.A.) law restricts this device to sale by or on order of a physician

FENWAL LABORATORIES, Morton Grove, Illinois U.S.A.

...D FOREIGN PATENTS ...

DO NOT VENT



SLEEMAN, DIGGS, HAMIT

THE EFFECTS OF ANTIBIOTICS, CORTICOSTEROIDS, AND PERITONEAL LAVAGE ON EXPERIMENTAL PERITONITIS

H. KENNETH SLEEMAN, PhD, JOHN W. DIGGS, AND HAROLD F. HAMIT, COL, MC
WALTER REED ARMY INSTITUTE OF RESEARCH
WASHINGTON, D. C. 20012

The prevention and treatment of bacterial shock remains a problem in military medicine. Shock caused by Gram-negative microorganisms frequently becomes refractory to therapy, and despite recent advances in treatment, the mortality rate remains high. Current trends in the therapy of bacterial shock, and the related field of endotoxin shock, have been the subject of several comprehensive reviews(1,2,3,4). Antibiotic therapy, while not a panacea, was most effective when in vitro sensitivity tests were performed prior to treatment(3,5). Occasionally, the use of antibiotics intensified the state of shock, presumably through the sudden destruction, lysis, and release of bacterial endotoxin(6). However, treatment with kanamycin has been successful in both experimental peritonitis in dogs and peritonitis from perforated appendices in humans(7,8) and a peritoneal lavage with kanamycin in saline was effective in the treatment of late peritoneal sepsis(9). In addition, large doses of the corticosteroids have been used to protect and treat humans and experimental animals in both bacterial(3) and endotoxin shock(2,10). Also, the combination of dexamethasone and kanamycin in large doses increased survival of dogs following endotoxin or hemorrhagic shock(11). However, the effectiveness of large doses of corticosteroids in the treatment of late bacterial shock remains a controversial matter(4,12).

The present study was undertaken primarily to study the effects of antibiotics, dexamethasone 21-phosphate, and peritoneal lavage on the lethality of a standardized Escherichia coli-hemoglobin peritonitis in the rat. Our experiments were designed to study the effectiveness of these therapeutic regimes from the initiation of the infection to a point near the expected time of death. At the time when the animals become refractory to one treatment, a combination of the treatments was evaluated. The effects of dexamethasone on peritoneal absorption also were investigated.

METHODS AND MATERIALS: Experimental Model - White, male Sprague-Dawley rats weighing 275-300 grams were used throughout the study. The principles of laboratory animal care as promulgated by the National Society for Medical Research were observed. Bacterial shock was produced by the intraperitoneal injection (5 ml/kg body weight) of a mixture of E. coli 0111:B4 (10^8/ml) and hemoglobin (4 gram percent). The preparation of the inoculum was described previously (13,14). Under these experimental conditions, about 70 percent mortality occurs within 24 hours after the inoculation. Rats were randomly assigned to treatment groups or untreated control groups. Treatment was initiated at selected times from one hour prior to the inoculation to 16 hours following the inoculation. The number of deaths occurring in 24 hours were compared in the treated and untreated groups. Where indicated, chi-square tests for significance were performed.

Treatments-1. Dexamethasone 21-phosphate (Decadron phosphate-Merck, Sharp, and Dohme) was given intramuscularly 1 hour prior to the E. coli-hemoglobin inoculation, immediately after the inoculation, and at 4, 8, 12 and 16 hours following the inoculation.

2. Antibiotics-Chloramphenicol (25 mg/kg), tetracycline (25 mg/kg), neomycin (40 mg/kg), and kanamycin (40 mg/kg) were injected intraperitoneally at the following times, immediately after the E. coli-hemoglobin inoculation and at 4, 8, 12 and 16 hours following the inoculation. The amounts of each antibiotic used was the largest dose easily tolerated by the rat.

3. Peritoneal Lavage. Animals were anesthetized with sodium pentobarbital and abdominal punctures were made about 2 inches to the left and right of the midline. Polyethylene tubing (PE 240) notched near the end, was inserted through one puncture site, gently manuvered to push aside the omentum, and guided into the pelvic gutter. The other end of the tubing was attached to a bottle of 0.9 percent saline suspended above the level of the animal to permit gravity flow. Polyethylene tubing also was inserted into the other abdominal puncture in a similar manner, and the free end placed in a collection container. The peritoneal cavity was lavaged with 100 ml of normal saline over a 1 hr period during which time the abdomen was manipulated frequently to prevent channeling of the flow. A single lavage was performed at 4, 8, and 12 hours following the E. coli-hemoglobin inoculation.

4. Combined Dexamethasone and Antibiotics-Single intramuscular injections of dexamethasone were given immediately after the E. coli-hemoglobin inoculation and at 4 and 8 hours following the inoculation. Then, single intraperitoneal injections of kanamycin or neomycin were given at 8 or 12 hours following the inoculation.

5. Combined Peritoneal Lavage and Dexamethasone-A saline peritoneal lavage was performed at 12 hours following the E. coli-hemoglobin inoculation and a single intramuscular injection of dexamethasone was given immediately after the lavage.

6. Combined Peritoneal Lavage and Antibiotic-A saline peritoneal lavage was performed at 12 hours following the E. coli-hemoglobin inoculation and a single intraperitoneal injection of kanamycin was given immediately after the lavage.

Absorption of Dye-Tagged Protein-Large molecular substances and collodial particles, such as bacteria, are absorbed similarly through the diaphragmatic lymphatics(15). Therefore, a study of the peritoneal absorption of dye-tagged protein would provide information as to the rate of absorption of bacteria from the peritoneal cavity (16). Normal rat plasma was tagged with Evans blue dye, T-1824, as described by Courtice and Steinbeck(15). Rats were given dexamethasone intramuscularly 1 hour prior to the E. coli-hemoglobin inoculation. Two hours after the dexamethasone injection, dye-tagged plasma, 5 ml/kg was given intraperitoneally. Two hours following the dye tagged plasma injection the animals were anesthetized with Nembutal and 5 ml of blood drawn from the aorta. The concentration of Evans blue dye in the plasma of the recipient animal was determined spectrophotometrically at 625 mμ. Milligrams of dye per liter of plasma were determined from a standard curve.

Reagent Sensitivity Test-E. coli 0111:B4 was sensitive to 30 μg chloramphenicol, 30 μg tetracycline, 5 μg kanamycin, 5 μg neomycin as determined by the standard impregnated disk technique. Dexamethasone could not be tested in this system because of the presence of bacteriostatic agents in the commercial solution. However, E. coli growth in tube tests was not inhibited by dexamethasone.

RESULTS: The effectiveness of the various treatments on the survival of rats subjected to E. coli-hemoglobin peritonitis is shown in Table 1. The value of each treatment varied depending upon the time of initiation after the E. coli-hemoglobin inoculation. All treatments except kanamycin, decreased deaths substantially below that found in untreated controls when therapy was begun up to 8 hours after the inoculation. At 12 hours after the inoculation, kanamycin and peritoneal lavage administered separately were ineffective, and the effectiveness of the other antibiotics was signficantly reduced. Only dexamethasone significantly ($p < .001$) reduced mortality when given 16 hours after the E.coli-hemoglobin inoculation. Also, but not shown in Table 1, dexamethasone given 1 hour prior to the E. coli-hemoglobin inoculation provided the same protection, 3 deaths in 32 animals, as when given up to 4 hours after the inoculation. The untreated control group shows the reproducibility of this model, that is about 70 percent mortality in 24 hours after inoculation.

Since kanamycin and neomycin lost effectiveness when administered 8 to 12 hours and peritoneal lavage was ineffective at 12 hours after the original inoculation, combinations of dexamethasone, antibiotics and peritoneal lavage were studied. Table 2 shows the results when decadron was given at 0, 4, and 8 hours and kanamycin and neomycin at 8 and 12 hours after the inoculation. The combined treatment with dexamethasone and antibiotics did not increase survival significantly over that found with dexamethasone alone. In fact, the use of antibiotics at these times following dexamethasone at 8 hours may be contraindicated, since the mortality associated with the combination of dexamethasone and antibiotics (25 and 33 per cent) was higher than with dexamethasone alone (18 per cent).

Table 3 compares the effectiveness of dexamethasone and kanamycin immediately following a saline peritoneal lavage performed 12 hours following the E. coli-hemoglobin inoculation. The combined use of a peritoneal lavage and dexamethasone did not increase survival over that found with dexamethasone alone. However, the use of kanamycin, following the peritoneal lavage, increased survival over that found with either treatment alone (Table 1). These results would suggest that once the concentration of bacteria had been reduced by peritoneal lavage, the antibiotic, kanamycin, was quite effective and also that the mechanism by which dexamethasone protected differed from that of the antibiotics and peritoneal lavage.

Previous studies in our laboratory showed a direct relationship between the rate of absorption of dye-tagged protein from the peritoneal cavity and mortality(16). A plasma dye concentration of 10 mg/liter or less, two hours after the intraperitoneal injection of dye tagged protein, was a good indicator of reduced clearance and eventual death. Dexamethasone was given at 1 hour prior to the E. coli-hemoglobin inoculation, immediately following the inoculation, and 2 and 4 hours after the inoculation. The average plasma dye concentrations following treatment are shown in Table 4. Dexamethasone increased the absorption of dye tagged protein from the peritoneal cavity over that of the untreated control. Although the increase in the amount of dye absorbed was small in the groups treated prior to and immediately following the inoculation, it does exceed the 10 mg/liter of plasma which was related to survival. The results would suggest that some of the beneficial effect of dexamethasone could result from improved bacterial clearance from the peritoneal cavity.

DISCUSSION: The efficacy of antibiotics, corticosteroids, and peritoneal lavage in the treatment of experimental peritonitis in the rat depended on two factors, the therapy employed and the time of initiation. In all instances, treatments initiated early after the E. coli-hemoglobin inoculation were more effective than those given later in the infection (Table 1). In contrast, the animals became refractory to both antibiotics and peritoneal lavage

at about 12 hours following the intraperitoneal injection of E. coli-hemoglobin. Since this corresponded to the time when maximum hematologic and metabolic changes occurred, and considerable tissue necrosis was found (17), one could postulate that a state of irreversible shock was produced. However, dexamethasone provided protection even at 16 hours after inoculation, and therefore each treatment must be considered individually as to its effectiveness and mechanism of action.

The ineffectiveness of antibiotics beyond 8 and 12 hours after the inoculation probably results from the sudden destruction of bacteria and the release of toxins at a time when the resistance of the animal was lowered by the infection (3,6). This would be substantiated by several findings: 1. The antibiotics were effective early in the infection at which time the bacterial count was high and resistance was maximal, 2. The antibiotics were ineffective when bacterial counts were high and extensive metabolic changes had occurred, 3. The most effective in vitro antibiotics, kanamycin and neomycin, were the least effective late in the infection, and 4. Immediately following the removal of bacteria by a peritoneal lavage at 12 hours, antibiotic therapy reduced mortality at a time when neither treatment alone was effective. The beneficial effects of kanamycin therapy previously reported (7,8,9) apparently resulted from its use early in the infection and at a time when the host was able to detoxify the products which resulted from bacterial lysis.

The primary benefit realized from a saline lavage was the physical removal of viable organisms from the peritoneal cavity. A secondary benefit could result from the removal of toxins by peritoneal dialysis during the lavage. The failure of peritoneal lavage to protect the host at 12 hours would indicate that sufficient physiologic and/or metabolic changes had occurred so that the ultimate outcome did not depend on a high concentration of viable organisms in the peritoneal cavity. However, at that time the animal was not refractory to additional therapy, and both dexamethasone and kanamycin decreased the number of deaths over that found with peritoneal lavage alone (Table 3). This finding suggests that a peritoneal lavage containing kanamycin would be beneficial at this time and supports the work of Schumer and associates (9).

The use of large doses of corticosteroids in both clinical studies and experimental shock models have yielded conflicting results when given at various times after the initiation of shock (2,3,10,11). However, in this model, dexamethasone offered the best protection against bacterial shock, and was effective even at 16 hours after the E. coli-hemoglobin inoculation. Combined therapy with dexamethasone and kanamycin were no more beneficial than dexamethasone alone, although this combination was effective in decreasing mortality in hemorrhagic and endotoxin shock in dogs (11). At this time, discrepancies in results found with corticosteroid

therapy can be attributed only to variability in the response of a species to a particular steroid, the time of administration, the concentration used, and the route of administration.

The mechanisms by which the corticosteroids confer protection are not known. Dexamethasone and the corticosteroids in general may protect by, 1. maintaining normal cell structure and thereby preventing the release of lysosomal enzymes(18,19), 2. inducing liver enzymes which are involved in the synthesis of proteins and nicotinamide(20), or 3. increasing cardiac output and decreasing peripheral resistance(3). Our studies showed that dexamethasone increased the membrane permeability of the peritoneum and permitted the more rapid clearance of dye-tagged protein. Since the rate of absorption of colloidal substances and bacteria from the peritoneal cavity can be equated(16), the beneficial effect of dexamethasone may be due to increased clearance of bacteria.

SUMMARY: The effectiveness of antibiotics, corticosteroids and peritoneal lavage on experimental peritonitis in the rat depends upon the time of initiation of treatment. Compared with untreated controls, treatment with antibiotics was ineffective after 8 hours or 12 hours, and peritoneal lavage was totally ineffective 12 hours after the E. coli-hemoglobin inoculation. In contrast, dexamethasone provided protection even at 16 hours after inoculation. Combined treatments of dexamethasone and antibiotic or peritoneal lavage was no more effective than dexamethasone alone. However, the combined peritoneal lavage and kanamycin treatment at 12 hours after inoculation, increased survival significantly over a peritoneal lavage alone. The beneficial effects of dexamethasone could be attributed in part to increased peritoneal absorption.

TABLE 1
Effectiveness of treatments (deaths/group) when given at times (hours) following E. coli-hemoglobin inoculation

Treatments	Amount	Hours after inoculation 0	4	8	12	16	Untreated Controls	12 hr deaths vs untreated controls*
Tetracycline	25 mg/kg	1/12 (8%)	3/18 (17%)	2/18 (11%)	5/12 (42%)	8/12 (67%)	12/18 (67%)	n.s.
Chlorampehnicol	25 mg/kg		0/12 (0%)	1/12 (8%)	5/12 (42%)	7/12 (58%)	9/12 (75%)	n.s.
Neomycin	40 mg/kg	2/24 (8%)	2/24 (8%)	7/24 (29%)	13/24 (54%)	15/24 (62%)	16/24 (67%)	n.s.
Kanamycin	40 mg/kg	3/10 (30%)	7/22 (32%)	10/22 (45%)	20/28 (71%)	16/22 (72%)	20/26 (72%)	n.s.
Dexamethasone	8 mg/kg	2/26 (8%)	4/42 (10%)	6/34 (18%)	6/48 (13%)	22/56 (40%)	104/129 (80%)	p <.001
Peritoneal lavage	100 ml saline		0/6 (0%)	1/6 (16%)	6/6 (100%)		14/20 (70%)	

*Chi-square statistics; n.s. = not significant

TABLE 2
Effectiveness of Combined Dexamethasone and Antibiotic Treatment (deaths/group) When Given At Times (hours) Following E. coli-hemoglobin Inoculation

Antibiotic		0 hrs	4 hrs	8 hrs	Untreated Controls
Kanamycin	8 hrs	1/12 (8%)	1/12 (8%)	4/12* (33%)	9/12 (75%)
	12 hrs	0/12 (0%)	1/12 (8%)	4/12 (33%)	
Neomycin	8 hrs	0/12 (0%)	2/12 (17%)	3/12* (25%)	7/12 (58%)
	12 hrs	0/12 (0%)	3/12 (25%)	3/12 (25%)	

*Mortality was not significantly (chi-square test) reduced when compared with dexamethasone alone (Table 1).

TABLE 3
Effectiveness of Combined Peritoneal Lavage and Dexamethasone or Antibiotic Treatment (deaths/group) at 12 hours Following E. coli-hemoglobin Inoculation

Treatment	Peritoneal Lavage 12 hrs	Untreated Controls
Dexamethasone 12 hrs	3/12 (25%)*	7/12 (58%)
Kanamycin 12 hrs	2/7 (33%)*	8/12 (67%)

*Using the chi-square test with a correction for small groups, mortality was not significantly reduced when the combined treatment was compared with dexamethasone alone (n.s.), or kanamycin alone ($p < .10 > .05$); mortality was significantly ($p < .05$) reduced when compared with peritoneal lavage alone (Table 1).

TABLE 4

Effects of Dexamethasone on the Absorption of Dye-Tagged Protein from the Peritoneal Cavity in *E. coli*-hemoglobin Peritonitis

Treatment	No. of Animals	Average dye conc. mg/liter plasma
1. Dexamethasone - 1 hr before inoculation	12	16
2. Dexamethasone- Immediately after inoculation	12	14
3. Dexamethasone - 2 hrs after inoculation	12	36
4. Dexamethasone - 4 hrs after inoculation	6	27
5. Inoculation only	12	7

REFERENCES

1. Martin, D.S., Cassisi, N.J., and Pickens, J.L.: Endotoxin Shock: A Collective Review. Review Surg., 22:311-319, 1965.

2. Hamit, H.F.: Current Trends of Therapy and Resarch in Shock. Surg. Gyn. Obst., 120:835-854, 1965.

3. Weil, M.H., and Shubin, H.: Shock Caused by Gram-negative Microorganisms. Ann. Int. Med., 60:384-400, 1964.

4. Shoemaker, W.C.: Shock: Chemistry, Physiology, and Therapy. Charles C. Thomas, Springfield, Ill., 1967.

5. Altemeir, W.A., Todd, J.C., and Inge, W.W.: Gram-negative Septicemia: A Growing Threat. Ann. Surg., 166:530-542, 1967.

6. Spink, W.W., Braude, A.I., Castendea, M.R., and Silva, G.R.: Aureomycin Therapy in Human Brucellosis Due to Brucella melitensis. J.A.M.A., 138:1145-1148, 1948.

7. DiVencenti, F.C., and Cohn, I., Jr.: Intraperitoneal Kanamycin in Advanced Peritonitis. Amer. J. Surg., 111:147-153, 1965.

8. Cohn, I., Jr., Cotlar, A.M., and Richard, L., Jr.: Intraperitoneal Kanamycin-Clinical Experience. Am. Surgeon, 29: 756-760, 1963.

9. Schumer, W., Lee, D.K., and Jones, B.: Peritoneal Lavage in Postoperative Therapy of Late Peritoneal Sepsis. Surg., 55: 841-845, 1964.

10. Lillehei, R.C., Longerbeam, J.K., Bloch, J.H., and Manax, W.G.: The Nature of Irreversible Shock: Experimental and Clinical Observations. Ann. Surg., 160:682-708, 1964.

11. Ashford, T., Palmerio, C., and Fine, J.: Structural Analogue in Vascular Muscle to the Functional Disorder in Refractory Traumatic Shock and Reversal by Corticosteroids. Ann. Surg., 164-575-585, 1966.

12. Thomas, C.S., and Brockman, S.K.: The Role of Adrenal Corticosteroid Therapy in Escherichia Coli Endotoxin Shock. Surg. Gyn. Obst., 126:61-69, 1968.

13. Davis, J.H., and Yull, A.B.: A Toxic Factor in Abdominal Injury. II. The Role of the Red Cell Components. J. Trauma, 4:84-90, 1964.

14. Filler, R.M., and Sleeman, H.K.: Pathogenesis of Peritonitis. I. The Effect of Escherichia coli and Hemoglobin on Peritoneal Absorption. Surg., 61:385-392, 1967.

15. Courtice, F.C., and Steinbeck, A.W.: The Rate of Absorption of Heparinized Plasma and 0.9 per cent NaCl from the Peritoneal Cavity of the Rabbit and Guinea-pig. Austral. J. Exper. Biol., 33:171-182, 1950.

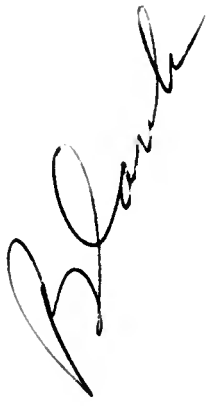
16. Sleeman, H.K., Diggs, J.W., Hendry, W.S., and Filler, R.M.: Pathogenesis of Peritonitis. II. The Effect of Escherichia coli and Adjuvant Substances on Peritoneal Absorption. Surg. 61:393-398, 1966.

17. Hendry, W.S., Sleeman, H.K., Diggs, J.W., and West, R.L.: Pathogenesis of Experimental Peritonitis in the Rat. Exper. Med. Surg., 24:303-312, 1966.

18. Riggs, T.R.: Hormones and the Transport of Nutrients, in Action of Hormones on Molecular Processes. Eds., Litwack, G., and Kritchewsky, D., John Wiley and Son, Inc., New York, 1964. pp 25-28.

19. Weissman, G., and Thomas, L.: The Effects of Endotoxin, Endotoxin Tolerance, and Cortisone on the Release of Acid Hydrolases from a Granular Fraction of Rabbit Liver. J. Exper. Med. 116: 433-450, 1962.

20. Feigelson, P., and Feigelson, M.: Studies on the Mechanism of Cortisone Action, in Action of Hormones on Molecular Processes. Eds. Litwack G., and Kritchevsky, D., John Wiley and Sone, Inc., New York, 1964, pp 218-233.



FREEZE DESICCATION - A NEW METHOD OF FOOD PRESERVATION

JURGEN STRASSER, JOHN G. KAPSALIS* and J. WALTER GIFFEE
U. S. ARMY NATICK LABORATORIES
NATICK, MASSACHUSETTS

INTRODUCTION

Most dehydrated foods, when compared with other kinds of processed foods, offer considerable logistic advantages for military feeding. Dried products are lightweight, easy and quick to prepare, and are stable when held without refrigeration over an extended length of time. As a result of intensive research performed in recent years, these products have reached a high quality standard. The main breakthrough in acceptance occurred when the food was vacuum freeze dried, a process by which the water vapor sublimes from the frozen product under low total pressure directly into the gaseous phase. By this method migration of solutes, shrinking, case hardening, loss of flavor volatiles, and other undesirable effects of air drying were greatly reduced.

Since the inception of this process about 15 years ago much pioneering development work has been performed at the Quartermaster Food and Container Institute in Chicago and its successor, the Food Laboratory of the U. S. Army Natick (Mass.) Laboratories.

In conventional freeze drying the water subliming from the food is condensed on refrigerated surfaces at sub-zero temperature. In 1956 work at the Western Regional Research Laboratory and in collaboration with the Biological Warfare Laboratories at Fort Detrick, Maryland, showed that frozen pellets of bacterial cultures could be vacuum freeze dried using silica gel (1). Recently we have investigated the use of solid desiccants similarly for drying foods. We termed this process "Freeze Desiccation". Preliminary studies indicated that the retention of quality of the food processed by freeze desiccation was high and that the equipment required was less intricate and less expensive than that needed for the conventional method. The new process offered certain additional advantages which are discussed later in this paper. Although our

experiments were performed on a laboratory scale and in batches, the results suggest that industrial application in the form of a continuous freeze-desiccation plant is a distinct possibility.

PROCEDURES

The differences between laboratory-scale freeze drying and freeze desiccation are indicated in Fig. 1. The food is prepared in the same way for both processes. In conventional freeze drying a condenser to trap the water vapor and heating plates are necessary to provide the heat for sublimation. In freeze desiccation, as performed in our laboratories, the frozen food was wrapped with a protective fabric and surrounded by a solid desiccant inside a drying chamber. Silica gel, which is granular, amorphous silicon dioxide, was found to be suitable for this purpose. Water vapor penetrates its submicroscopic pores and is held by surface absorption and capillary condensation. The action of silica gel on water vapor is reversible by heat. The desiccant will totally desorb at 350°F, and is then capable of adsorbing to its full capacity again. In repeated cycles there is some loss of sorption capacity of the gel due to occlusion of small amounts of volatile organic materials from the foods and their alteration by the heat applied to dry the gel. For some purposes other desiccants are sometimes useful. For example, when low moisture levels in the dried food were desired, a mixture of silica gel with molecular sieves was found to be better than silica gel alone. The molecular sieves are crystalline complex silicates containing aluminum and one or more other metallic elements.

The desiccation chamber, packed with the frozen food samples as described above, was evacuated to about 0.3 mm Hg, at which point the valve to the vacuum pump was closed. The chamber was then placed in a refrigerated room (5°C). The most suitable wall temperature of the chamber was found to be between 0°C and 10°C. The moisture of the food is sublimed from the solid into the vapor phase, and is absorbed by the desiccant at a rate sufficient to keep the frozen food from melting. During the early stages of dehydration, the heat evolved due to the adsorption of water by the desiccant is utilized as the primary energy source for sublimation. As drying progresses, the heat conducted to the product from the chamber walls becomes increasingly important. Thus the size and shape of the chamber are also important. When the ice is completely sublimed, the walls of the chamber may be brought up to room temperature to hasten desorption of bound water and attainment of the ultimate moisture level.

STRASSER, KAPSALIS* and GIFFEE

STUDIES OF THE FREEZE-DESICCATION PROCESS UNDER LABORATORY CONDITIONS

1. Relationships Between The Weight of The Food And The Weight of The Desiccant.

Preliminary studies to determine the relationship between the initial weight of the product and the weight of the desiccant needed for drying were conducted with high protein foods (precooked beef and raw codfish). Differences in the initial moisture content of the frozen samples were considered to be negligible. Experiments with beef promptly revealed that the ratio of food to desiccant regulates the degree to which the product is dried (Fig 2). The sharp increase in the amount of desiccant necessary to obtain a low moisture content in the meat reveals the limitations of a one-step batch-type freeze-desiccation process. It should be noted that a smaller amount of molecular sieve than of silica gel is necessary to obtain the same final moisture content. The difference is caused by the strong affinity between molecular sieves and water in the low moisture region.

2. Influence of The Location of The Desiccant.

In order to investigate the influence of the heat transport from the desiccant to the product, two experiments were performed. In one experiment codfish samples (discs 16mm diameter, 6mm thickness) were packed in silica gel and the moisture content was determined after different drying times. The weight ratio of the frozen fish to silica gel was 1:5. The chamber wall temperature was kept at +10°C. In the other experiment, the sample and the desiccant were located in connecting chambers. The weight ratio and the chamber wall temperature were the same as in the first experiment. The drying times for samples in contact with the desiccant were about 50% less than those for those not in contact because of the better conditions for heat and mass transfer (Fig. 3). The experiments also showed that the final moisture content was independent of the location of the food relative to that of the desiccant. The final moisture content depends upon the weight ratio of frozen sample to dry desiccant, final chamber temperature, and the sorption characteristics of both the food and the desiccant.

3. Correlation Between The Desorption Isotherm of The Food And The Adsorption Isotherm of The Desiccant.

At the end of the freeze-desiccation process, an equilibrium water vapor pressure has been established in the chamber. This equilibrium can be best understood if the adsorption isotherm of the desiccant and the desorption isotherm of the food at the final drying temperature are known. For isotherm measurements the instrumentation shown in Fig. 4 was used. Precise water vapor pressures were generated in the system by means of a

thermostatically controlled water bath. This method of measuring isotherms is superior to the usual technique of using saturated salt solutions since closely spaced points over the entire relative humidity (R.H.) range can be obtained and controlled. A front view of the apparatus is shown in Fig. 5.

The moisture sorption isotherms of raw codfish and silica gel are plotted in Fig. 6. On the basis of these isotherms, one can calculate the amount of desiccant necessary to reach a certain desired moisture content in the food. The following equation was derived:

$$R = \frac{m_d \times (1 + \frac{m_i}{100})}{m_i - m_e}$$

R = ratio of weight of wet food to weight of dry desiccant

m_d = final equilibrium moisture content of desiccant (% dry basis)

m_i = initial moisture content of food (% dry basis)

m_e = desired final moisture content of food at end of drying process (% dry basis)

The sorption characteristics of molecular sieves are compared with the sorption behavior of silica gel in Fig. 7. Below about 30% relative humidity, molecular sieves of the 4A type have a considerably higher affinity for water vapor. Above 30% R.H., silica gel is the more hygroscopic. Therefore, substitution of a molecular sieve for the silica gel below about 30% R.H. would be suitable when low final moisture contents are desired, especially in a continuous freeze-desiccation apparatus. Such a plant would be almost mandatory for large scale production.

It is conceivable that eventually a continuous freeze-desiccation process may become competitive with the freeze-drying process now in use. It might even be considerably less expensive, since high-cost refrigeration equipment would be displaced by low-cost desiccants, which can be used repeatedly following simple heat reactivation. The solution of all problems concerning the feasibility of a continuous system must await the information that can only be supplied by pilot-plant studies.

ADVANTAGES

The freeze-desiccation process has several advantages. No sophisticated freeze-drying equipment with refrigerated condenser and heating plates is necessary. A simple laboratory desiccator or any other sealable container may be used for the drying

chamber. Vacuum pumping is necessary only at the beginning of the process. Therefore, a single vacuum source can be used for a whole series of chambers which are simply sealed off after evacuation and transferred to a cooler.

The advantages of predetermining a uniform terminal moisture content in the freeze-desiccated foods are manifold. In fact, this feature is as yet one of the unachieved goals for the conventional freeze-drying method. It has been known for several years that some foods undergo oxidative deterioration more rapidly when excessively dry than when they contain the so-called monolayer amounts of water (4, 3). Still higher levels of moisture promote discoloration and off-flavor development through amino-carbonyl and other types of reactions. Retention of desirable texture also requires optimal amounts of water in the stored product, and it was recently demonstrated (2) that this optimal moisture range can be accurately defined by suitable instrumentation.

There is a need by the military for a number of compressed food bars for both tactical and logistic purposes. Freeze-dried foods are particularly suitable for compression because of the voids remaining after removal of water. A certain amount of residual moisture in the dried product contributes a desirable degree of plasticity for compression purposes while an overly dry product tends to crumble. Without the advantage of freeze desiccation, products are often overdried and an extra step is introduced wherein the products are moistened before compression. This constitutes an added cost.

Freeze desiccation is a very gentle process; the mild heat of adsorption released by the desiccant is supplied to the food by conduction without perceivable damage to the product. The high survival rates achieved with bacterial cells by Graham et al. (1), is evidence that even subtle damage is unlikely.

SUMMARY

Freeze desiccation of foods is based on the use of a solid desiccant for the adsorption of water vapor which sublimes from the frozen food at a reduced pressure. Desorption isotherms of the food and adsorption isotherms of the desiccant were used for the calculation of the quantities of desiccant necessary to obtain different pre-determined residual moisture contents of the food. An equation and a curve were developed for this purpose. The new method compares favorably with conventional freeze dehydration. It offers the specific advantages of accurate control of the end point residual moisture in the food. For moisture contents above about 5%, the use of silica gel as a desiccant was found to be very suitable. For moisture contents below this limit, substitution of molecular sieves for the silica gel in a two-stage desiccation process was necessary in order to desorb the more firmly bound water.

REFERENCES

1. Graham, R. P., A. H. Brown, and L. F. Ginnette. 1956. Development of Silica Gel Drying Process for Vegetative Bacteria. Contract Research Report, Western Regional Research Laboratory, Albany, California, Interagency Agreement No. CD3-2460. Also: US Patent No. 2853707 and US Patent No. 2897600.

2. Kapsalis, J. G. and B. Drake. 1967. Hygroscopic equilibrium and textural parameters of special freeze-dried foods. Presented at 27th Ann. Meeting, Inst. Food Technol., Minneapolis, Minn. 15-19 May.

3. Maloney, J. F., T. P. Labuza, D. H. Wallace and M. Karel. 1966. Autoxidation of methyl linoleate in freeze-dried model systems. I. Effect of water on the autocatalyzed oxidation. J. Food Sci., 31, 878.

4. Salwin, H. 1959. Defining minimum moisture contents for dehydrated foods. Food Technol., 13, 594.

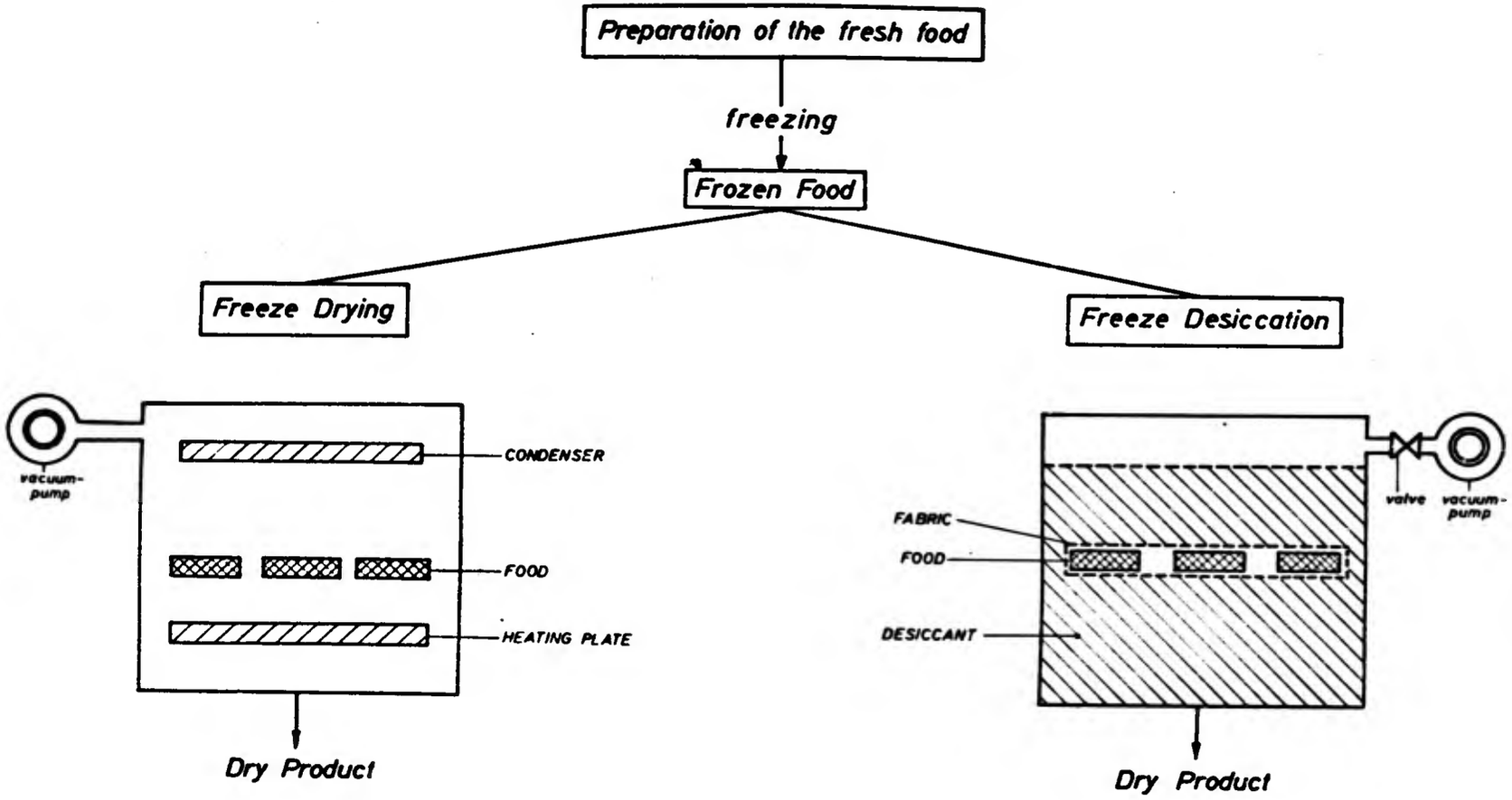


Fig. 1. Comparison of commonly used freeze drying with freeze desiccation

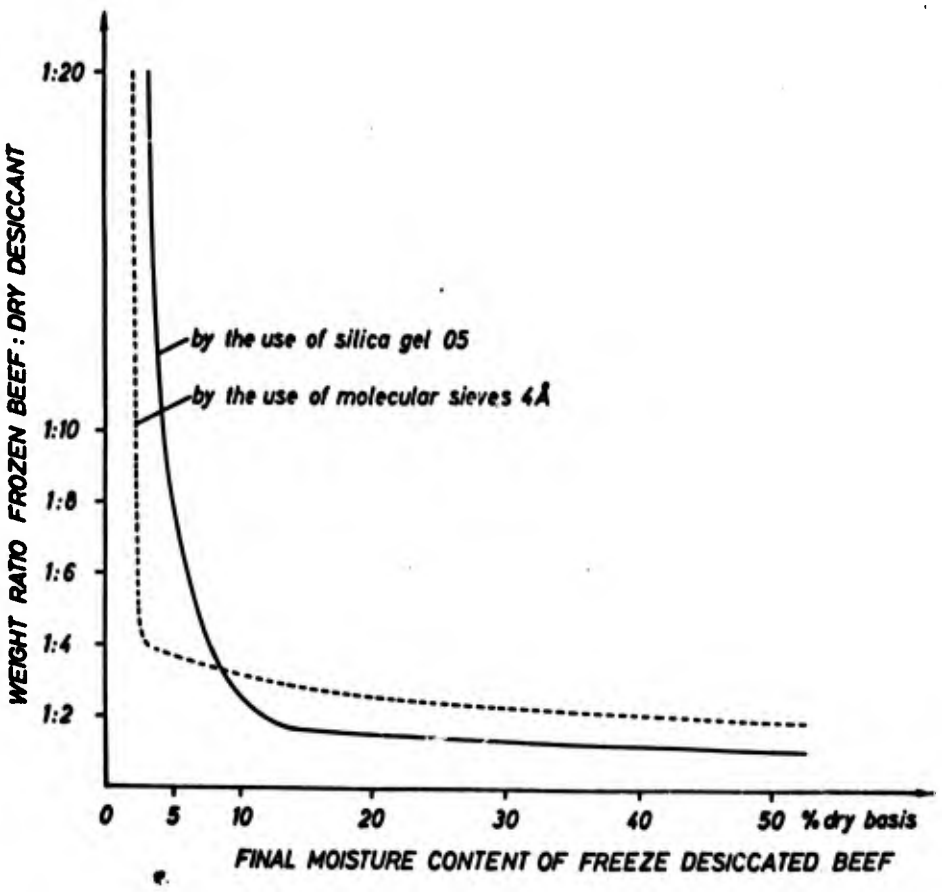


Fig. 2. Relationship between food to desiccant weight ratio and final moisture content in beef.

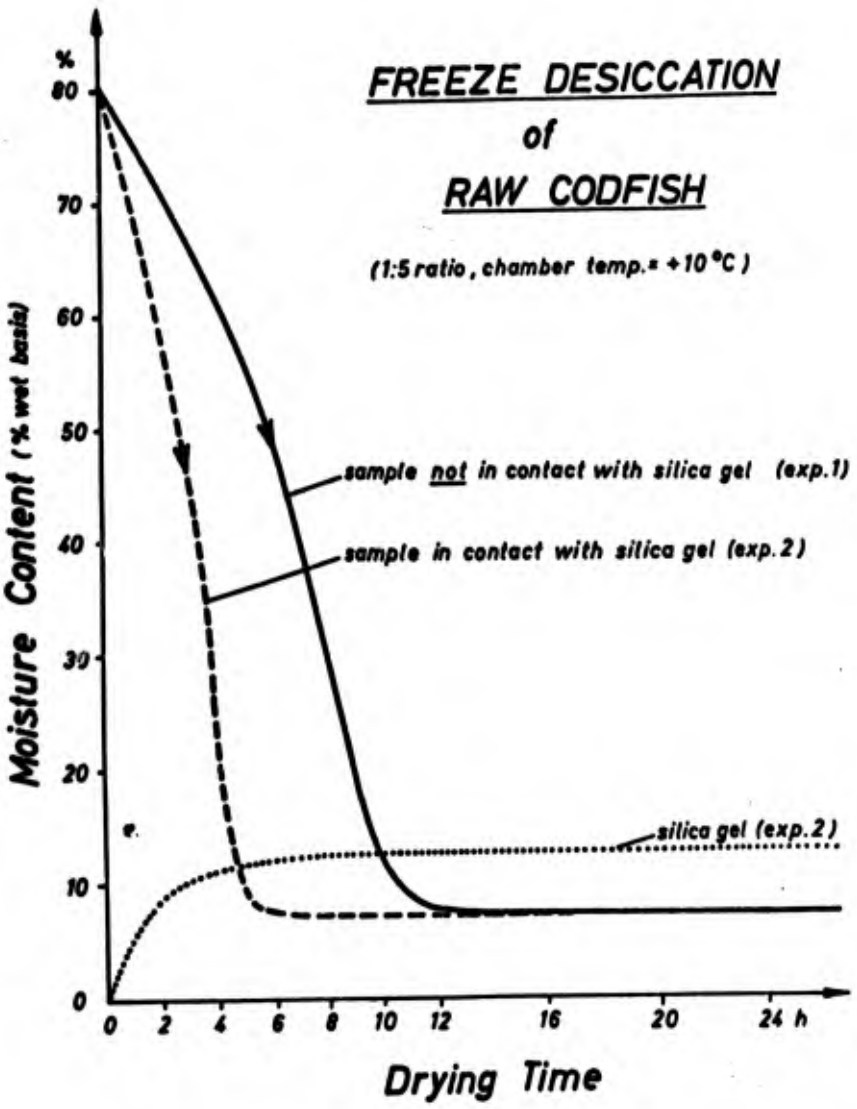


Fig. 3. Comparison of different methods of freeze desiccation. The dotted line indicates the moisture uptake of silica gel during experiment 2.

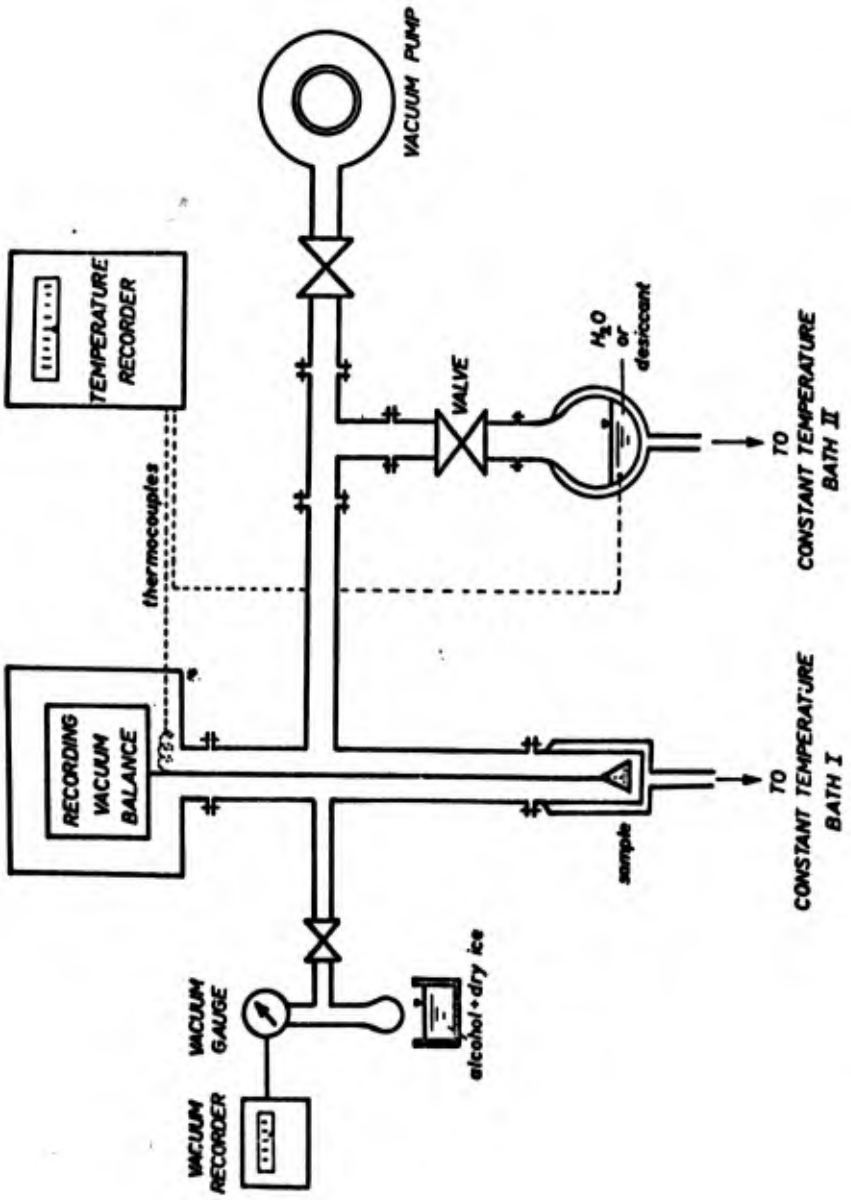


Fig. 4. Apparatus for the measurement of sorption isotherms or freeze-dehydration rates. The bath with alcohol and dry ice is used for the determination of the partial pressure of "non condensible gases" in the system.

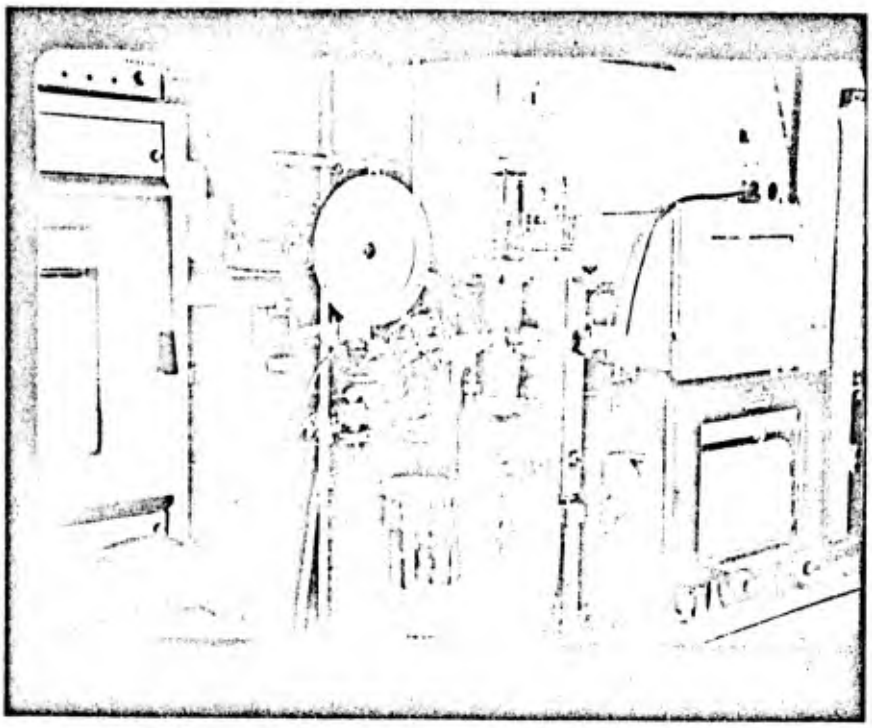


Fig. 5. Apparatus for determining sorption isobars.

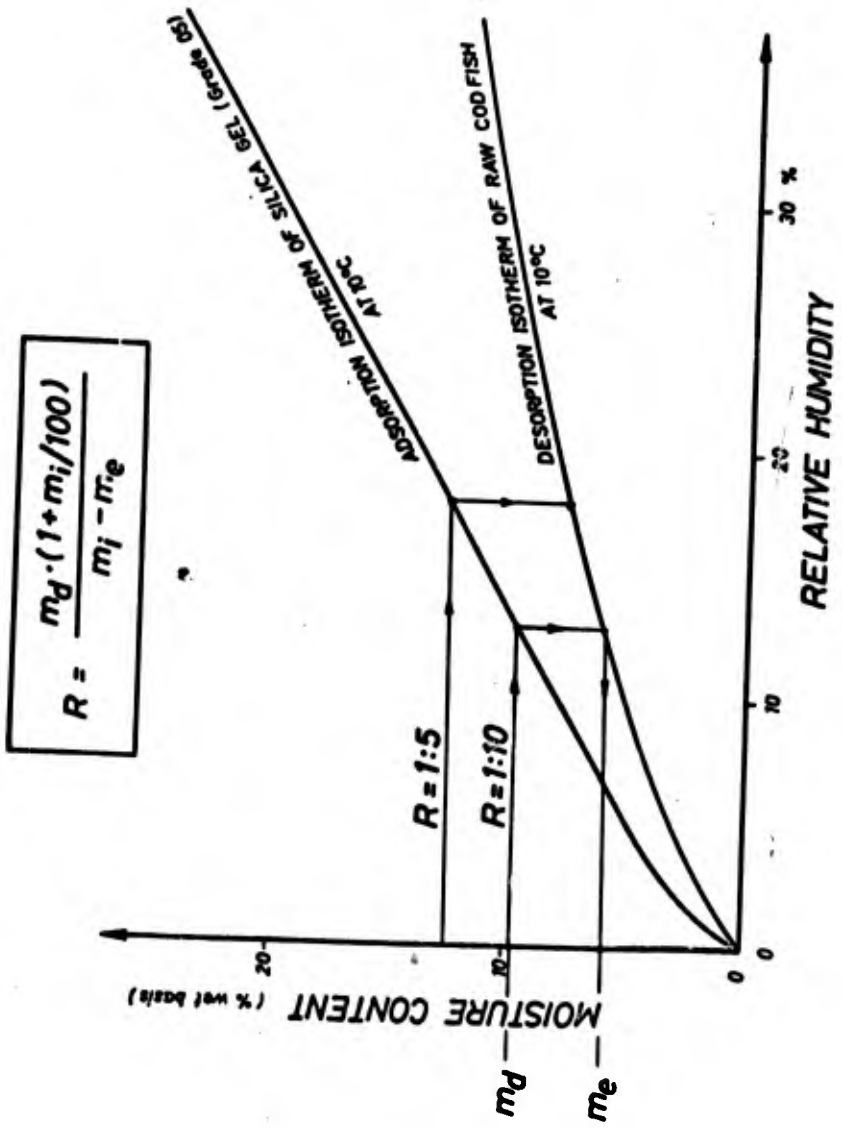


Fig. 6. Moisture sorption isotherms of raw codfish and silica gel.

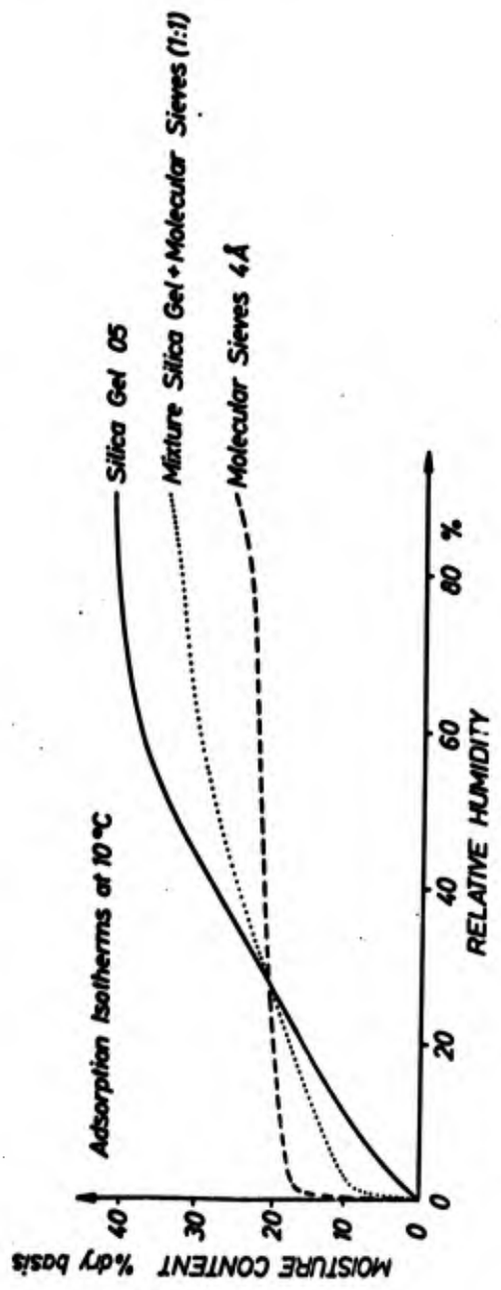


Fig. 7. Moisture sorption isotherms of silica gel and molecular sieves

MECHANISM OF THERMAL DECOMPOSITION OF HMX (1,3,5,7-TETRANITRO-1,3,5,7-TETRAZACYCLOOCTANE)

BULUSU SURYANARAYANA, JOSEPH R. AUTERA and RICHARD J. GRAYBUSH
EXPLOSIVES LABORATORY, FRL, PICATINNY ARSENAL
DOVER, NEW JERSEY

INTRODUCTION: The basic mechanisms by which certain substances can undergo explosive decomposition are not well understood. No generalized models are available to explain the mechanism of explosion in terms of primary electronic and molecular processes. One useful approach has been to study the slow decomposition of metastable materials induced by various stimuli such as heat, light and ionizing radiation. Knowledge concerning these processes such as rate constants and identity of the reactive intermediates, should be helpful for a mechanistic understanding of the fast reactions in explosives.

HMX (1,3,5,7-tetranitro-1,3,5,7-tetrazacyclooctane) has been selected in this work for a detailed study of its thermal decomposition for two reasons. Firstly, it is a chemical close relative and a higher melting homologue of the widely used explosive, RDX (1,3,5-trinitro-1,3,5-triazacyclohexane). Secondly, there is much

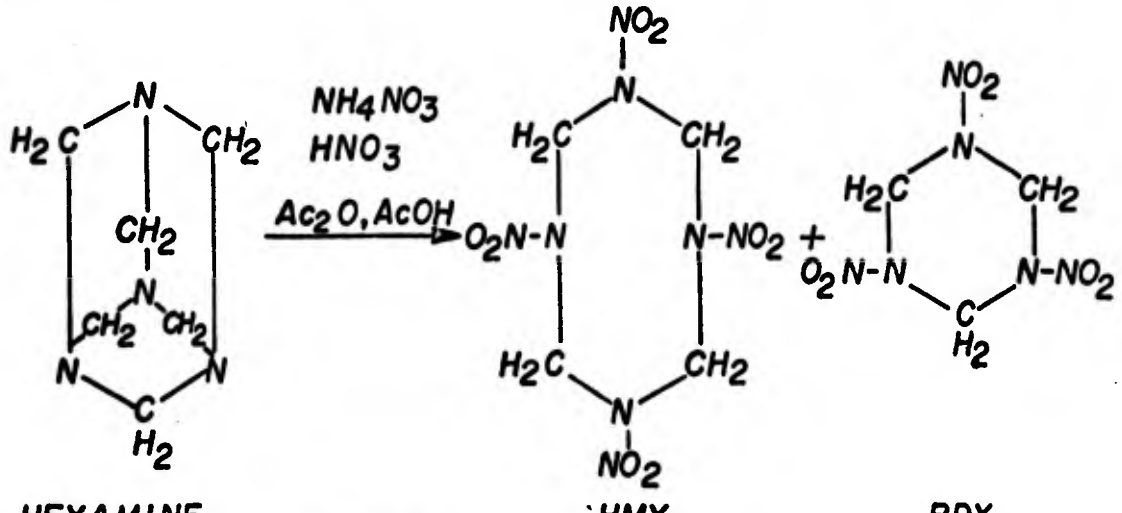


FIGURE 1

current interest in its four crystalline forms which show somewhat different impact sensitivities (1). A variety of their properties, e.g., crystal structures (2a-c), stereoisomerism (3), phase transitions (1,4), electrical polarizations (3), piezoelectric behavior (5) and U.V. and I.R. absorption (3,6,7) have been investigated recently. Thermal decomposition kinetics of RDX (8) and alkaline decomposition of both RDX and HMX (9) have also been studied before to some extent. However, detailed studies of thermal decomposition of HMX, especially in the solid state, have not been reported earlier by other laboratories. The general chemistry of organic nitramines of which HMX and RDX are examples, was reviewed recently by Urbanski (10). HMX and RDX can be readily synthesized (11) by nitration of hexamine (Figure 1) with ammonium nitrate and nitric acid.

As a part of this continuing mechanism study, the products of thermal decomposition of δ-HMX[1] were determined by a mass spectrometric analysis (12). The decompositions were carried out isothermally at several temperatures in the range 225-285°, since earlier kinetic data (13) indicated a possible change in the composition and/or ratio of the products. In addition, to help with understanding the breakdown of HMX molecule to the observed products the origin of nitrogen in the latter was traced by isotopic nitrogen (N^{15}) labeling (14). There are two chemically different types of nitrogens in HMX, viz., the nitro nitrogens and the amino nitrogens in the ring. For the purpose of this work HMX with 99% N^{15} in the nitro groups was used. The data on the chemical composition of the gas evolved during thermal decomposition and the isotopic composition of the components containing nitrogen are presented in this paper. It will be shown that the formation of the observed major species can be rationalized by assuming that the N-N bonds in HMX do not break. The discussion will include a consideration of possible mechanisms and new experiments with carbon-13 and oxygen-18 labeled HMX to test them further.

EXPERIMENTAL

Materials: Technical grade HMX obtained from Holston Defense Corporation contained about 1% RDX as impurity. It was purified by an initial soxhlet extraction with ethylene dichloride to remove the more soluble RDX (15). This was followed by three crystallizations from acetone. This procedure gives the β-polymorph.

[1] Of the four polymorphs (β, α, γ and δ) of HMX the β-form is the stable one at room temperature. However, upon heating, β-HMX undergoes a phase transition to δ-HMX at about 200°C. As the temperature is increased above 200° it begins to decompose rapidly and at 280° it melts. In all the experiments described here the isothermal decomposition of both labeled and unlabeled HMX, were carried out above 225°. At the temperatures used in these experiments, therefore, HMX exists as the δ-form.

SURYANARAYANA, AUTERA
and GRAYBUSH

Nitro-N^{15}-labeled HMX: This was synthesized by nitrolysis of hexamine with a mixture of $NH_4N^{15}O_3$ and $HN^{15}O_3$, both 99% enriched. The N^{15}-labeled RDX, a byproduct of the reaction, was separated from HMX by ethylene dichloride extraction. The nitro-N^{15}-labeled HMX was then purified further by successive precipitations with water from solutions in dimethyl sulfoxide and acetone, respectively. Finally, two recrystallizations were done from acetone. The sample was dried in vacuo at 78° for 24 hours before use.

The RDX formed as byproduct carries an identical "label"— nitro-N^{15}. It was recovered by evaporation of the ethylene chloride solution and purified by precipitation from acetone solution with petroleum ether followed by recrystallization from ethylene dichloride.

C^{13}-Labeled HMX and RDX: Hexamine-C^{13} was first prepared via $C^{13}H_3OH$ and $C^{13}H_2O$. The nitrolysis of hexamine -C^{13} and the separation-purification steps were similar to those of N^{15}-materials.

O^{18}-Labeled HMX and RDX: Since an anhydrous mixture of $NH_4NO_3^{18}$ and HNO_3^{18} was required to synthesize O^{18}-HMX a special vacuum line was designed to make it from KNO_3^{18} as the starting material. The remaining synthesis and purification steps were similar to those of N^{15}-materials.

All the isotopic starting materials were obtained from Isomet Corporation, Palisades Park, New Jersey. Details of the above procedures will be published elsewhere.

Analysis of the Labeled HMX and RDX: All the carbons and oxygens in HMX are chemically equivalent. Therefore, only the labeled nitrogen requires a determination of both location in the molecule and the amount. It was assumed that there would be no exchange between the nitro-nitrogen and the amino-nitrogen during the synthesis resulting in incorporation of N^{15} in the ring. Proof that all the N^{15} was in the nitro groups was obtained from the mass spectral fragmentation patterns of HMX/RDX. The latter were obtained on a CEC21-104 mass spectrometer which was equipped with a solid sample introduction probe. Some peaks in the mass spectra may be used to locate as well as estimate the isotope content in the molecule. The "base" peak in these spectra at m/e = 46 was presumably due to the ion, $(NO_2)^+$. Many nitro compounds give rise to this peak in their mass spectra (16). In the case of the present nitramines a process such as

$$\text{Nitramine} \xrightarrow{-e^-} R_1R_2N{-}NO{-}O^{+\bullet} \longrightarrow R_1R_2N\cdot + NO_2^+$$

would give $(NO_2)^+$ at m/e = 46. This was confirmed by the appearance of a peak at m/e = 47 $(N^{15}O_2)^+$ in the spectra of the nitro-N^{15} labeled materials. Also, the latter spectra showed less than 1%

abundance of m/e 46 relative to 47. This demonstrated that all the N^{15} was in the nitro groups of HMX.

The C^{13} and O^{18} materials were of lower enrichment than N^{15}. The enrichments calculated from the isotope peak ratios of various fragments were: C^{13}, 42% and O^{18}, 5.4%, respectively.

Thermal Decompositions: 50 mg. samples of β-HMX (both labeled and unlabeled) were outgassed at pressures of about 10^{-4} mm for 15-20 hours and then heated at constant temperature under an argon pressure of 40 mm in a closed vacuum system. The inert gas, argon, was intended for preventing sublimation of the sample. The constant volume (380 ml) vacuum system was provided with a 1 ml sampling volume for removal of samples for analysis and a propeller type stirrer to help mixing of the gases. The entire assembly was attached to a CEC Model #21-620A mass spectrometer. In the case of unlabeled HMX, 1 ml samples of the gas mixture obtained by decomposition were injected directly into the mass spectrometer at different reaction times. But, in the case of the N^{15}-labeled HMX the sample gas mixture was separated into three fractions: volatile at -195°, at -80° and at room temperature. Each of these was analyzed in the mass spectrometer separately.

The constancy of the mass spectra of successive samples of gas mixture indicated the completion of decomposition in each run. The runs were repeated at different temperatures as shown in Tables 3 and 4. The procedure for RDX was identical except for the temperature. Only one run was made with RDX at 190°C. At this temperature the complete decomposition would take several hours. Samples of the gaseous products were, however, taken for analysis only during the first two hours at an estimated 10-20% decomposition.

Analysis of the Mass Spectra: The major products were initially identified as N_2O, HCHO, N_2, NO, CO and CO_2 with small amounts (<1%) on HCN, $(CN)_2$ and H_2. From the known fragmentation patterns (17) they were expected to contribute to various m/e peaks as

Table 1

m/e	Contributed by
12	HCHO, CO, CO_2 and HCN
14	N_2O, NO, N_2, CO and HCHO
22	CO_2
28	N_2O, N_2, CO, CO_2 and HCHO
29	HCHO, CO and N_2
30	N_2O, NO and HCHO
44	CO_2 and N_2O

shown in Table 1. Using the previously determined calibration factors for the various components and the equations of Table 1 the concentrations of the individual products were computed.

In the case of N^{15}-labeled HMX and RDX the mass spectra were more complicated and needed fractionation of the products to simplify the analysis. The patterns obtained from the decomposition products of RDX and HMX were very similar. The marked differences observed between the mass spectra of the decomposition products from the unlabeled and the nitro-N^{15}-labeled HMX are illustrated in Table 2. The intensities (chart divisions) of only the important peaks (m/e) are shown here. The striking increase in the intensities of peaks at m/e = 45, 31 and 29 and a decrease in the relative intensities of 44, 30 and 28 are noteworthy. Detailed procedures to analyze the mass spectra obtained from the decomposition products of labeled and unlabeled HMX will be reported elsewhere.

Table 2

Mass Spectra (Peak Heights[a] Versus M/E) Obtained from the Thermal Decomposition Products of Nitro-N^{15}-Labeled and Unlabeled HMX

M/E	Products From Unlabeled HMX at 285°			Products From Nitro-N^{15}-Labeled HMX at 230°		
	-195°	-78°	RT	-195°	-78°	RT
28	7350	1550	273	618	688	167
29	104	135	570	1470	2280	560
30	2900	5740	820	110	1300	375
31	20	40	18	890	1890	198
32	20	40	27	9	9	7
44	-	12,800	40	8	1357	37
45	-	110	-	7	4517	127
46	-	40	-	-	61	-

[a] Chart divisions corrected for background but not normalized; only ratios of important peaks are to be considered significant in this table since the absolute peak heights depend upon many experimental variables.

Results: The chemical composition of the decomposition products of δ-HMX is shown in Table 3. It can be seen from this data that:

1. Formaldehyde and nitrous oxide account for about 65% of the products, i.e., these are the major products.

2. The amount of formaldehyde observed varies by a large magnitude.

3. Up to ∿260° the composition appears to be constant within experimental errors, but at 285° (near the m.p. of HMX) the yields of N_2 and NO increase at the expense of N_2O. Further work is required to explain this difference at higher temperatures satisfactorily.

Formaldehyde: Although not shown in the data presented here, it was observed (12) that during a given thermal decomposition run the formaldehyde pressure rose to a maximum and then decreased by some 20-30%. It seems reasonable to assume that the decrease in the amount of HCHO after reaching a maximum was due to polymerization. The variation in the amounts of HCHO can be attributed to loss by polymerization to some extent. Also, due to this loss, the actual amount of formaldehyde formed in the decomposition of HMX appears to be greater than what was observed by the method of analysis used here. This emphasizes the point that HCHO and N_2O are the major products of decomposition of HMX.

Distribution of N^{15} Among the Decomposition Products of Nitro-N^{15}-Labeled δ-HMX: The isotopic species and the chemical constituents of the decomposition products from HMX and RDX, both nitro-N^{15} labeled, are shown in Table 4. The three runs with HMX and one with RDX, all show very similar isotopic (N^{15}) distribution among the products. A few uncertainties in the data indicated in the

Table 3

Thermal Decomposition Products[(a)] of δ-HMX
(moles per mole of HMX)

Temp, °C	N_2O	N_2	NO	HCN[(b)]	CO	CO_2	CH_2O[(c)]
226	2.52	0.47	0.42	0.02	0.40	0.56	1.60
234	2.63	0.54	0.37	0.01	0.32	0.48	1.20
242	2.58	0.45	0.28	0.03	0.46	0.56	1.15
250	2.58	0.46	0.50	0.04	0.47	0.41	0.87
258	2.57	0.47	0.55	0.07	0.47	0.54	0.47
285	1.60	0.93	1.20	0.52	0.42	0.30	0.86

(a) After complete decomposition a polymeric residue is left, amounting to 2-4% by weight of original HMX. This was not yet identified.
(b) In some later experiments at lower temperatures up to 4-5% HCN was observed.
(c) The variation in these amounts was in part due to polymerization of formaldehyde.

footnotes to the table were caused by slight cross-contamination of the three fractions into which the mixture of products was separated. The isotopic species of nitrous oxide was clearly distinguishable from a large peak at mass-to-charge ratio, m/e = 45 in the mass spectrum which could arise only from $NN^{15}O$ or $N^{15}NO$. Of these two species, $N^{15}NO$ could be eliminated on the basis of its cracking pattern. For, on the basis of the known cracking pattern of N_2O,

this would be expected to show the parent ion at m/e, 45 and a peak at m/e, 30 (relative abundance, 31) corresponding to the ion, NO^+. It was found, however, that in the mass spectrum there was no contribution to the peak at m/e, 30 from $N^{15}NO$ after subtracting other contributions to the same peak. On the other hand, a prominent peak at m/e, 31 which is consistent both with the cracking pattern of $NN^{15}O$ and the parent ion of nitric oxide-N^{15} ($N^{15}O$) was found. Thus, there is little doubt that the isotopic species of nitrous oxide was almost entirely $NN^{15}O$. This also confirms that the N^{15} in the original HMX molecule was in the nitro groups, only.

It can be seen from Table 4 that nearly all the nitrogen also consisted of the mixed isotopic species NN^{15}. The nitric oxide and hydrogen cyanide consisted predominantly of N^{15} and N^{14}, respectively. HCN, therefore, was derived mainly from the ring nitrogen in HMX.

Table 4

Nitrogen Containing Thermal Decomposition Products of Nitro-N^{15}-Labeled δ-HMX and RDX

Isotopic	Species	Percent Composition			
		δ-HMX			RDX
		230°	254°	285°	190°
N_2O[a]	$N^{14}N^{14}O$	-	-	-	-
	$N^{14}N^{15}O$	98-99	98	98	98-100
	$N^{15}N^{14}O$	-	-	-	-
	$N^{15}N^{15}O$	1-2	2	2	0-2
N_2[b]	$N^{14}N^{14}$	-	-	-	-
	$N^{14}N^{15}$	93-100	92	95	98-100
	$N^{15}N^{15}$	0-7	8	5	0-2
NO[b]	$N^{14}O$	0-7	0-5	0-5	0-5
	$N^{15}O$	93-100	95-100	95-100	95-100
HCN[c]	HCN^{14}	100	100	100	100
	HCN^{15}	-	-	-	-

(a)Both CO_2 and $N^{14}N^{14}O$ give a peak at m/e = 44; but CO_2 can be determined from the peak at m/e = 22. However, the presence of $N^{14}N^{14}O$ up to an estimated maximum of 3% was possible. Also up to about 5% $N^{15}N^{14}O$ was possible in some runs.

(b)About 5% $N^{15}N^{15}$ and 10% $N^{14}O$ are possible in all cases; their interference with each other and a contribution to mass 30 from HCHO made their determination uncertain.

(c)Presence of up to 15% HCN^{15} was possible but could not be detected under our experimental conditions.

One experiment with nitro-N^{15}-RDX was done as mentioned earlier and the results included in Table-4 in order to see if the N^{15} distribution in the products would be similar. It was apparent that identical isotopic species were formed. The chemical composition of the products was also similar to the products from HMX -decomposition. However, Robertson (8) did not observe hydrogen cyanide among the products.

DISCUSSION

Bond Breaking Processes in HMX: The exclusive appearance of nitrous oxide as $NN^{15}O$ strongly suggests that the reaction path to it does not involve fission of an N-N bond. If any other isotopic species ($N_2^{15}O$, $N^{15}NO$ or N_2O) had been observed in substantial amounts, one could visualize a route to N_2O which does require such fission. For example fission of an N-N bond in HMX produces NO_2 which may undergo reduction to N_2O by the formaldehyde observed in the products. In the case of nitro-N^{15}-labeled HMX this gives $N^{15}N^{15}O$. This conclusion is consistent with the preliminary observations on photolytic (wavelength, 3100-3300Å) decomposition of nitro-N^{15}-labeled HMX also being studied in this laboratory. The nitrous oxide obtained here consisted of 60-70% $NN^{15}O$ and 30-40% $N_2^{15}O$. It appears, therefore, that two processes are occurring simultaneously in the photolysis one of which involves breakage of N-N bonds. The fact that very little $N_2^{15}O$ was formed in the thermal decomposition and also, that all the nitrogen consisted of the mixed isotopic species, NN^{15} support the above conclusion that an N-N bond fission does not occur to significant extent. Breakage of one or more C-N bonds is of course, necessary to account for the production of $NN^{15}O$ and CH_2O as the major products (Table 3). However, the fact that the nitrogen in HCN consisted entirely of the unlabeled species (N^{14}) would imply that the C-N bond was intact, contrary to the conclusion from the N_2O and N_2 data. But, HCN is produced in too small an amount (Table 3) in the thermal decomposition to be indicative of a major process unless it is consumed rapidly in a secondary reaction. Based on the major products, N_2O, N_2 and NO these experiments with N^{15}-labeled HMX show, therefore, that the thermal decomposition involves a predominant breakage of the C-N bonds (69.7 kcal/mole) rather than the weaker N-N bonds (38.4 kcal/mole) (18a). This applies to the decomposition at all three temperatures used for δ-HMX although the ratio of the products at 285° appears to be different (Table 3) from that at lower temperatures. Furthermore, the preference for the C-N bond fission appears to be general to the secondary nitramines since nitro-N^{15}-RDX gave identical N^{15} distribution in its products.

Experimental Activation Energies: The available figures of activation energy (summarized in Table 5) for the thermal decompositions of HMX and RDX vary over a wide range and continue to be modified as new data are obtained. Experiments in our laboratory show that the activation energy for HMX decomposition varies in the three temperature regions as shown in the table indicating a change

Table 5

		Temperature Range	Activation Energy
This work[a] (12):	δ-HMX (m.p. 275°)	220-245°	46 kcal/mole
		245-270°	18.5
		270-285°	52
Robertson (8):	RDX (m.p. 205°)	213-299°	47.5
Rauch (19):	RDX	207-217°	52(b)

(a) Full discussion of the kinetic.date will be presented in a future paper.
(b) Quoted to be a tentative value.

in the relative importance of different reactions. However, all the values are lower than the energy of the C-N bonds (69.7 kcal/mole). This suggests a catalytic process which lowers the activation energy for the reaction. There seems to be an uncertainty about the N-N bond energy. The earlier value of 38.4 kcal/mole was modified recently (20) to 55 kcal/mole. The resonance of the nitramino group, presumably, results in a considerable stabilization of the N-N bonds. The existence of resonance contribution which will have the effect of decreasing the N-N bond distance was supported by previous structural studies of three nitramines: HMX (2), ethylenedinitramine (21) and dimethyl nitramine (22). A recent neutron diffraction study of the structure of HMX by Choi (2c) of our laboratory confirmed this. All these values, collected together in Table 6, are consistent with a partial double bond character in the N-N bonds. This implies that the N-N bonds in HMX and RDX are considerably stronger than the normal N-N single bonds. Although the present N^{15} tracer experiments strongly suggest that the fission of the N-N bonds does not occur in the thermal decomposition of HMX and RDX, they do not permit clear conclusions about a C-N bond fission being the rate determining step.

Table 6

Normal N-N Bond Distances and Those Found in Nitramines

	Bond length, Å	Ref
Normal N-N Single bond	1.48	(18b)
" N=N Double bond	1.24	(18b)
HMX Average N-N length	1.39	(2a)
	1.39	(1b)
	1.37	(2c)
Dimethylnitramine	1.26	(22)
Ethylenedinitramine	1.33	(21)

Mechanism of Decomposition: Several mechanisms may be considered for the thermal degradation of HMX which are consistent with the above conclusions regarding the bond-breaking processes and with the fact that CH_2O and N_2O are the major products. Figure 2 shows

FIGURE 2

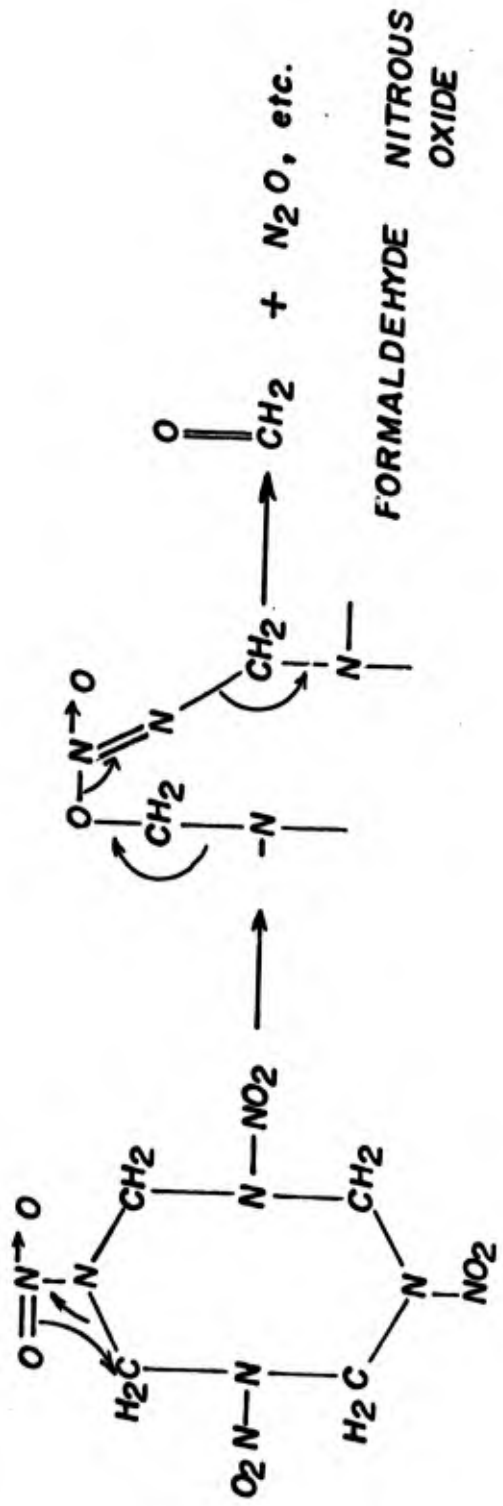


FIGURE 3

a concerted mechanism in which the intermediate, (A) undergoes a simultaneous intramolecular oxygen transfer to the carbon and a fission of the C-N bond to give CH_2O and N_2O. Robertson (8) proposed a somewhat similar mechanism for the decomposition of RDX, shown in Figure 3. Both of them require an intramolecular transfer of an oxygen atom to the adjacent carbon.

While it is not possible to distinguish the two mechanisms experimentally, the intermediate (A) (CH_2:NNO_2) in the first mechanism (Figure 1) merits consideration on the basis of the mass spectra of HMX and RDX. They showed peaks at m/e, 222 and 148, respectively,which shifted to 225 and 150 in the spectra of the N^{15}-labeled materials. These fragment ions (odd-electron) contain, therefore, 3 and 2 nitro groups, respectively. The neutral fragments eliminated must be of mass 74 for the case of unlabeled compounds and 75 for the labeled compounds. This mass number corresponds to CH_2:NNO_2 which can be reasonably expected according to the mechanism shown in Figure 4. The intermediate (A) (Figure 1) would be identical to the neutral fragment, 74 formed in this breakdown of HMX and RDX by electron impact.

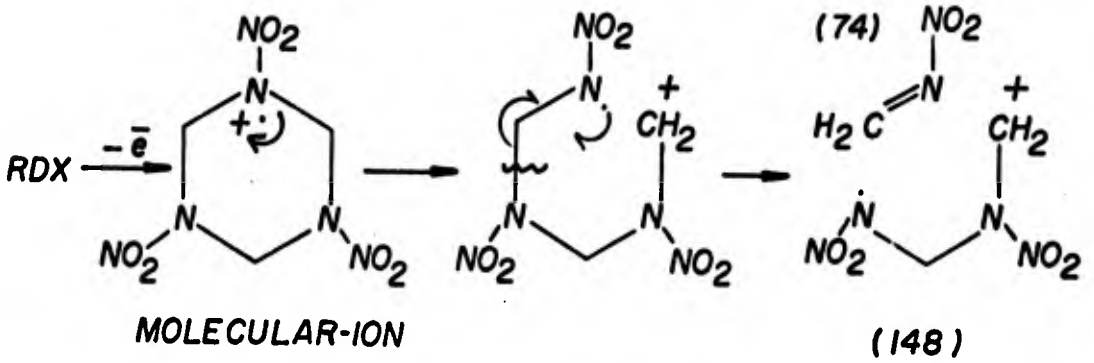


FIGURE 4

A third mechanism for the thermal decomposition of HMX which can be described as a stepwise dissociative process (i.e., not a concerted process) is the excision of an $\dot{N}$-NO_2 moiety. The latter may rearrange itself and loose an oxygen atom to give N_2O. Presumably the oxygen will be lost to a methylene fragment giving CH_2O, now by a process which may or may not be intramolecular.

Experiments with C^{13}- and O^{18}-HMX: It would be useful to determine if the decomposition reaction is strictly intramolecular, to check the validity of the above mechanisms. In addition, to distinguish between primary processes which may be inter or intramolecular is important since a process characteristic of the crystalline state is more likely to be an intermolecular one. An intramolecular process will be a characteristic property of the molecule. Current work in this laboratory is designed to determine

this by the use of pairs of isotopic tracers.

One such attempt consists of decomposing a coprecipitated mixture of C^{13}-labeled and O^{18}-labeled HMX. If formaldehyde is formed by transfer of the oxygen atom to an adjacent carbon in the molecule then, C^{13} and O^{18} would not appear together. On the other hand, if the reaction is intermolecular there will be some isotope crossover. $C^{13}H_2O^{18}$ with mass number 33 will be an unusual peak in the mass spectrum of HMX decomposition products distinguishable from the other possible isotopic species:

CH_2O	m/e = 30
$C^{13}H_2O$	= 31
CH_2O^{18}	= 32
$C^{13}H_2O^{18}$	= 33

In a preliminary experiment the m/e = 33 peak was found, surprisingly, in greater proportion than required on a random mixing basis. Among many possible reasons for such intermixing, the most important one seems to be a secondary reaction such as polymerization in which CH_2O^{18} and $C^{13}H_2O$ would exchange. Attempts are being made to devise experiments in which the formaldehyde can be mass analyzed rapidly before a secondary exchange reaction can interfere with the isotopic composition. Another approach under investigation is a chemical trapping of CH_2O which will preserve the original isotope distribution.

Intermolecular processes may also be recognized by mixing of N^{15} and O^{18} isotopes. Work along these lines is in progress.

References

1a. H. H. Cady and L. C. Smith, "Studies on the Polymorphs of HMX", LAMS-2652, Los Alamos Scientific Laboratory, Los Alamos, N.M. (1962).

1b. H. H. Cady, A. C. Larson and D. T. Crommer, Acta Cryst., 16, 617 (1963).

2. (a) P. R. Eiland and R. Pepinsky, Z. Kristallogr., 106, 273 (1955)
(b) W. C. McCrone, Analy. Chem., 22, 1225 (1950).
(c) C.S. Choi and H. P. Boutin, unpublished work.

3. M. Bedard, H. Huber, J. L. Meyers and G. F. Wright, Can. J. Chem. 40, 2278 (1962).

4. B. Suryanarayana, J. R. Autera and R. J. Graybush, Molecular Crystals, 2, 373 (1967).

5. J. N. Maycock and D. E. Grabenstein, Science, 152, #3721, 508 (1966).

6. (a) R. N. Jones and G. D. Thorn, Can. J. Res., 27, Sec B, 828 (1949).
(b) R. Schell, Picatinny Arsenal Semi-Annual Progress Report, June 1965.

7. (a) J. L. Myers and A. M. Bedard, "The Infrared Studies of the HMX Polymorphs and their Estimation of RDX", CARDE Tech. Memo. #139 (1957).
(b) H. P. Boutin, S. Trevino and H. Prask, Picatinny Arsenal Tech Report #3370, September 1966.

8. A. J. B. Robertson, Trans. Faraday Soc., #313, XLV, Pt 1, 85 (1949).

9. (a) S. Epstein and C. A. Winkler, Can. J. Chem., 29, 731 (1951).
(b) W. Jones, J. Am. Chem. Soc., 76, 829 (1954).

10. (a) T. Urbanski, Chemistry and Technology of Explosives, Vol 3., Chapters 1 to 4, Pergamon Press, (1967).
(b) A. H. Lamberton, Quar. Rev., 5, 75, (1951).

11. W. E. Bachmarr and J. C. Sheehan, J. Am. Chem. Soc., 71, 1842 (1949).

12. B. Suryanarayana,and R. J. Graybush, Proceedings of the 39th Congress on Industrial Chemistry, Brussels, Belgium (1966).

13. B. Suryanarayana, et.al. to be published.

14. B. Suryanarayana, R. J. Graybush and J. R. Autera, Chem. & Ind., 2177 (1967).

15. Physical and Chemical Properties of RDX and HMX, Holston Defense Corp, Kingsport, Tenn. Control #20-P-26, Series B, (1962).

16. F. W. McLafferty, "Interpretation of Mass Spectra", Benzamin, Inc., New York (1967).

17. Mass Spectral Data, American Petroleum Institute Research Project #44, (1959).

18. L. Pauling, The Nature of the Chemical Bond, 3rd Edn., Cornell (1960) (a) p. 85 (b) p. 228.

19. F. C. Rauch, American Cyanamid Co., Progress Report #5, Contract No. DAAA21-68-C-0334 (1968).

20. R. C. Cass, et. al., J. Chem. Soc., 958 (1958).

21. F. J. Llewellyn and F. E. Whitmore, J. Chem. Soc., 1315 (1948).

22. W. Costain and E. G. Cox, Nature 160, 826 (1947).

THE ENZYMOLOGY OF BALLISTIC WOUNDING

SAMUEL A. TISDALE JR. AND CHARLES C. BERDJIS, LTC, MC
BIOPHYSICS LABORATORY, RESEARCH LABORATORIES
EDGEWOOD ARSENAL, MARYLAND

Introduction:

For many years, methods have been sought that would allow valid comparison of the wounding capabilities of various kinetic energy missiles. That this has not been entirely satisfactory is attested to by the large numbers of parameters that have been studied in the past. Recent comparisons have used pine boards, the "58 foot-lb" criterion[1], wound tract measurements, gelatin-block tissue substitutes, etc. It was felt that perhaps more physiological methods would be helpful in supplementing the customary techniques in use, and towards this end, we have been studying several such subsystems. One of these is the change in serum enzyme levels consequent to ballistic wounding.

When tissues are injured by a variety of means, the cells are either directly damaged or subsequently undergo degeneration secondary to impairment of their blood supply or other factors. Intracellular components are released into the surrounding tissues by leakage or actual cell disruption. They can then gain access to the intravascular space, where they can be serially sampled and measured. Previous human experience (e.g. myocardial infarction) tends to show that, within limits, the amount of enzyme released (and hence measured) may be proportional to the amount of tissue damaged or destroyed.

A number of enzymes were studied in an attempt to ascertain those which would be most useful in studying wounds of goats. The enzymes evaluated included glutamic-pyruvic transaminase (GPT), alkaline phosphatase (AP), creatine phosphokinase[4] (CPK) aldolase[5] (ALD) and glutamic-oxaloaetic transaminase[6] (GOT). Since goat tissues were found to contain little if any GPT, and AP levels were not found to change with a reasonable degree of trauma. Because of this, both were dropped from consideration in these studies. Studies with CPK are still in progress.

Materials and Methods:

1. Animals: The "standard" animal for ballistic research in our laboratory is the castrated Angora Goat (Capra Hircus). Weights varied from 33.6 to 48.2 kilograms.

2. Enzyme Activity Assays: (a) GOT - The colorimetric Sigma-Frankel[7] method was employed. A Coleman Jr. spectrophotometer was used at 505 mu and calibration was performed using the Sigma 505-10 calibration standard solution. Clearance studies were performed using the Karmen ultraviolet assay[8] measured at 340 mu on a Beckman DB spectrophotometer. (b) ALD - The Sigma modification[9] of the Sibley and Lehninger[10] method was used with readings at 540 mu on a Beckman DB spectrophotometer.

3. Exogenous Enzyme: Activity was obtained in two ways: (a) 200-250 ml of whole blood was removed by venapuncture from each of 4 animals, frozen at 20°C, and thawed to produce hemolysis. The blood was centrifuged at 4°C (250 xg for 20 min) and the supernatent stored frozen. (b) 16-58 gm of Brachial triceps muscle were surgically removed from each of nine animals, minced in nine parts of isotonic saline, and homogenized in a Waring Blendor for 10 min. The supernatent was centrifuged at 4°C (350 xg for 20 min) and filtered through 7 u (5 animals) or 3 u (4 animals) Seitz filters and frozen.

4. Clearance Studies: Enzyme preparations (as above) were thawed, diluted as necessary with distilled water to maintain isotonicity, and rapidly infused intravenously into the goat using the jugular vein. Serum enzyme activity was measured before and serially after infusion for 72 hours.

5. Ballistic Studies: The goats were weighed, sheared, and the jugular vein cannulated with polyethylene tubing for subsequent sampling. They were then anesthetized intravenously with pentobarbital sodium diluted 1:2 to prevent hemolysis, positioned 1 meter from the gun muzzle, and shot with 1/4 inch steel spheres (weighing 1.04g) at various velocities. Entrance and exit velocities were determined electronically. Hind legs were the targets, the largest muscle mass 1.5 cm posterior to the femur being used.

Blood for GOT and ALD enzyme activity was drawn prior to wounding, and at 5, 22, 30, 46, 52, 70 and 94 hours afterwards. The animals were sacrificed; the wound tracts were then measured and photographed.

Controls were subjected to all of the above measures except that the missile was fired to miss the animal. Enzyme levels did not rise at all in these animals.

Goats wounded and immediately sacrificed have been evaluated in an effort to ascertain any change in wound tracts with elapse of a week.

Results:

A. Normal Serum Enzyme Levels:

1. GOT: 71 determinations had a mean of 62 Sigma-Frankel (SF) units with a standard deviation (SD) of 17 units.

2. ALD: 16 determinations had a mean of 13 Sibley-Lehninger (SL) units with a SD of 5 units.

B. Normal Tissue Enzyme Levels:

Studies to establish normal enzyme levels are continuing. Liver GOT levels were found to be 45,500 to 48,200 SF units per gram of wet tissue. GOT in skeletal muscles from various body sites have varied from 36,500 to 39,300 SF units per gram of wet tissue. ALD from skeletal muscle varied from 11,000 to 19,600 SL units per gram of wet tissue.

C. Clearance of Endogenous Enzyme:

The rates of enzyme clearance, represented by the slopes of the lines, are seen in Figure 1. GOT was cleared at rates varying from 16 to 44% each 8 hours (mean 24%, SD $\pm$ 10%). Neither the source of the enzyme (i.e. RBC or muscle) nor the initial level significantly effected the rate of clearance.

ALD was cleared at rates of 17-38% per 8 hours (mean 26%, SD + 8%) and was similarily independent of enzyme source or initial level. When calculated for biological half-life, GOT had mean serum half-life of 24 hrs (SD $\pm$ 9 hrs) while ALD had a half-life of 20 hrs (SD $\pm$ 8 hrs).

D. Ballistic Studies:

Figure 2 graphically represents typical results from one animal, GMZ-4. Kinetic energy (KE) in joules imparted to the tissue has been calculated from the formula: $KE = \frac{1}{2} mv^2$. The dimensions of the wound, as determined by detailed dissection, are the basis of the illustrated two dimensional reconstructions. By using the formula:

$$V = \pi \frac{h}{3} (r_1^2 + r_1 r_2 + r_2^2)$$

where V = volume of wound-tract segment in cc.

h = length of wound-tract segment in cm.

r_1 = radius of entrance wound in cm.

r_2 = radius of exit wound in cm.

the volumes of each segment, the total wound volume (ΣV), and the muscle wound volume (ΣMV), represented by the shaded areas, have been computed.

Figures 3 and 4 show the increment of enzyme level over preshoot levels plotted as a function of time. Since the "area under the curve" is enzyme contributed by the damaged tissue, comparisons of this "area" with kinetic energy, wound volumes have been made. Computer run variance analysis shows no significant fit between enzyme areas and wound volume or ALD area and kinetic energy, but there is a linear fit significant at the 95-99% level between SGOT area and KE (see Figure 5).

Comment:

The evaluation of ballistic missile damage is made difficult by the presence of both temporary and permanent wound tracts. At best, the relative contribution of each of these phenomena to the ultimate morbidity and mortality of the wound is speculative. In addition, the measurement of wound tracts is subject to great inaccuracies, although it has been possible to predict the volumes of the permanent and temporary cavities in terms of the energy expended in the wound[2].

Perhaps more important than the permanent wound tract is the extent of damage caused by the temporary cavity. Surrounding the permanent tract a zone of muscle with marked extravasation of blood and some cellular swelling and separation is usually found. Under ideal conditions, survival studies[3] and tissue culture studies[11] have demonstrated only partial irreversible damage in this area, but under field conditions, this damaged tissue may form the background for bacterial growth, resulting in far greater damage.

The use of indirect means of measuring tissue damage, as described in this report, is an attempt to deal with this problem. Although the preliminary studies are too limited to draw far reaching conclusions, there appears to be an encouraging relationship between energy input and GOT release. At the same time, no relationship was found between energy input and permanent wound tract measurements in this study.

A limiting factor is the possibility that the tissue removed in the formation of the permanent wound tract is no longer available to contribute to the rise in enzyme activity. Furthermore, enzyme activity could be directly destroyed by the energy released. The lack of relationship between energy and ALD levels is unexplained, but work is in progress evaluating other sized missiles at varied velocities and energies to clarify these results.

References:

1. Beyer, J.C., editor: Wound Ballistics. Washington, D.C., Office of the Surgeon General, Department of the Army, p93, 1962.

2. ibid, pp 140-141.

3. ibid, p134.

4. Kleine, T.O. Localization of Creatine Kinase in Microsomes and Mitochondria of Human Heart and Skeletal Muscle and Cerebral Cortex. Nature, 207:1393, 1965.

5. Hess, B. Enzyme in Blutplasma. G. Thieme, Stuttgart, 1962.

6. Schmidt, F.W., quoted in Methods of Enzymatic Analysis. Academic Press, New York, p837, 1963.

7. Sigma Technical Bulletin No. 505, Sigma Chemical Company, St. Louis, Missouri, 1964.

8. Karmen, A. A Note on the Spectrophotometric Assay of GO-T in Human Blood Serum. J. Clin. Invest., 34:131, 1955.

9. Sigma Technical Bulletin No. 750, Sigma Chemical Company, St. Louis, Missouri, 1961.

10. Sibley, J.A. and Lehninger, A.L. J. Biochem., 177:859, 1949.

11. Wood, W.O. The Pathology of Wounding: in vitro Regeneration of Goat-Liver Tissue Wounded by a Ballistic Missile, Edgewood Arsenal Technical Report, Edgewood Arsenal, Maryland, EATR 4141, November 1967.

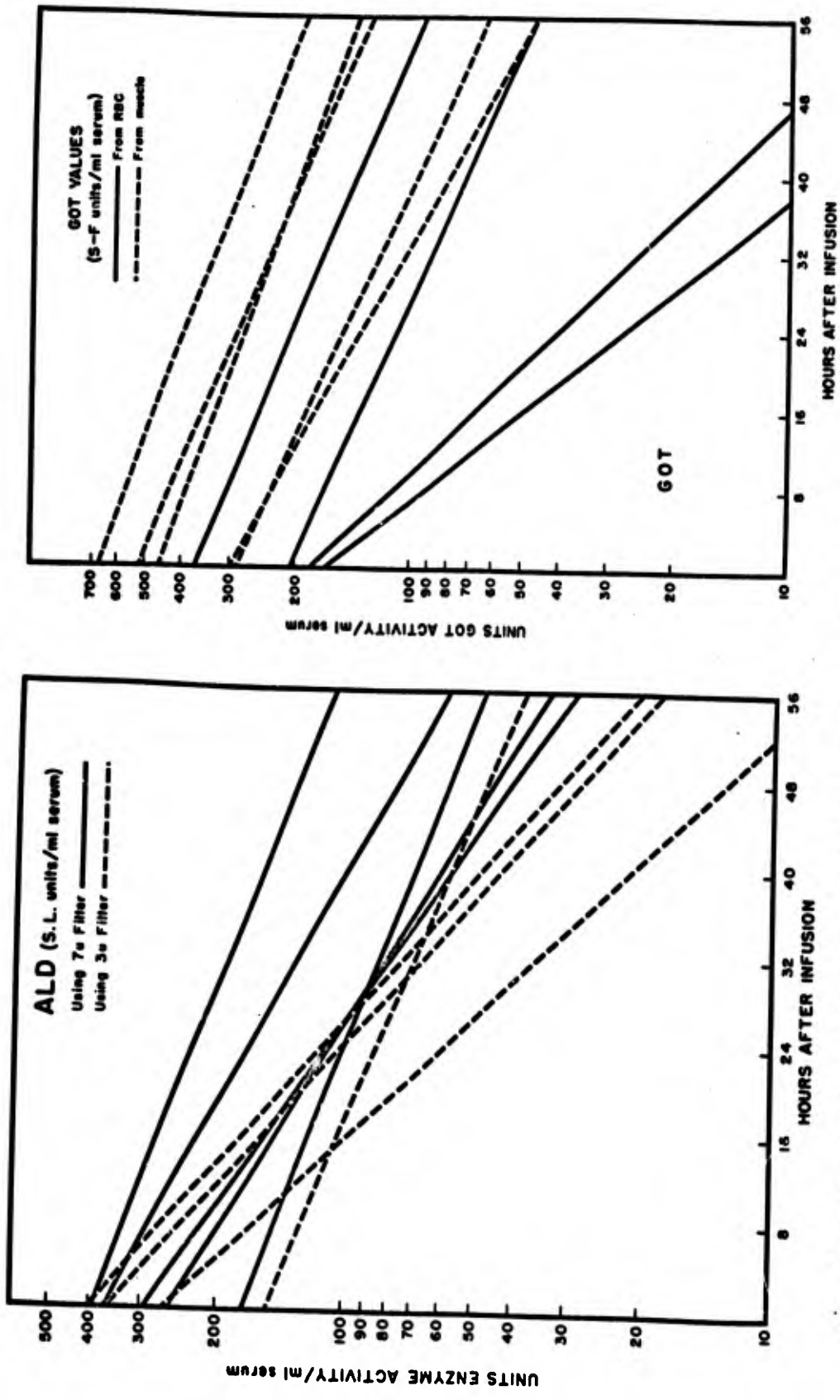


Fig. 1: Clearance of exogenous ALD (left) and GOT (right) from the circulation.

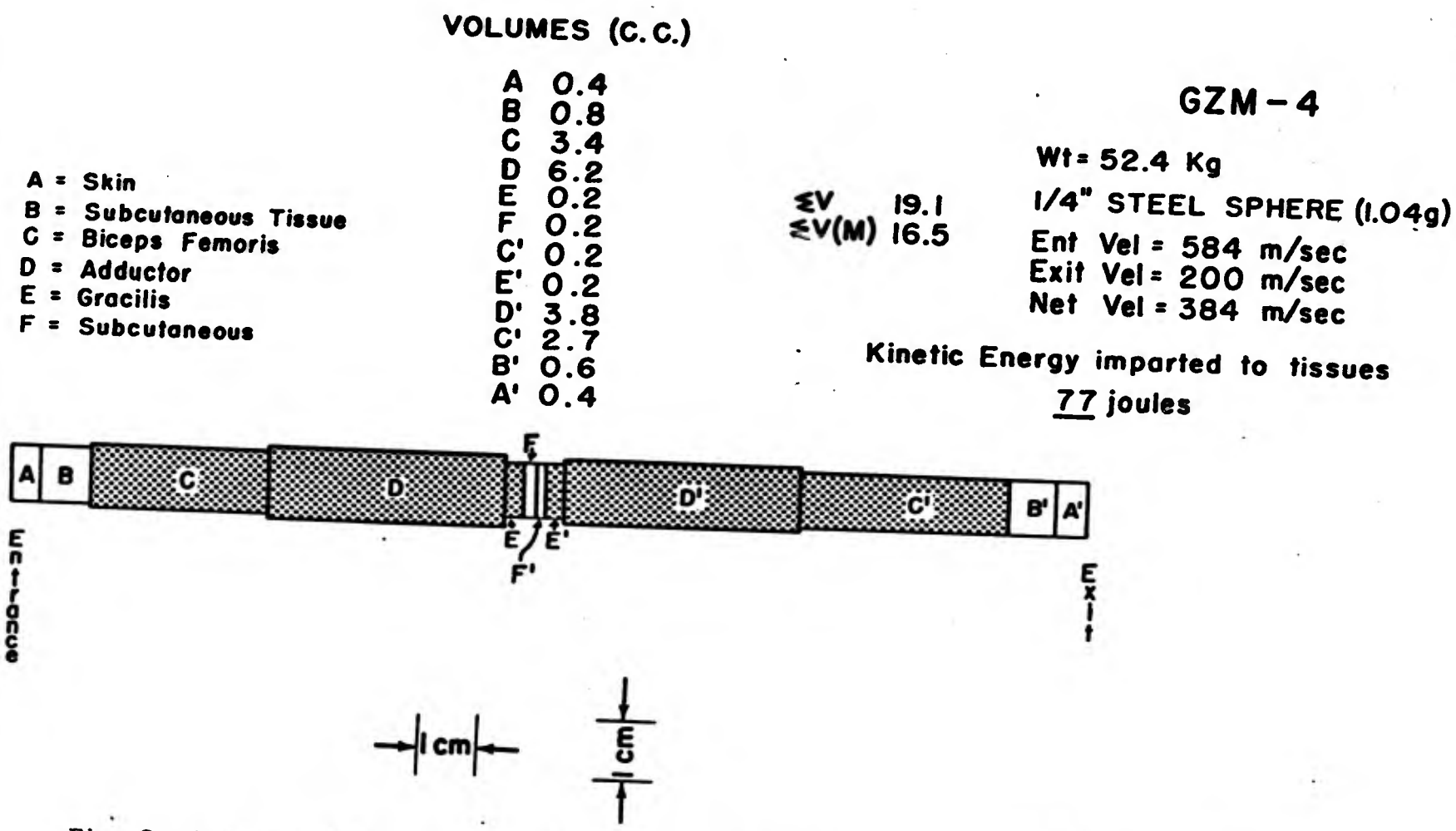


Fig. 2: Diagram of wound tract inflicted by $\frac{1}{4}$" steel sphere in thigh muscle of goat GZM-4.

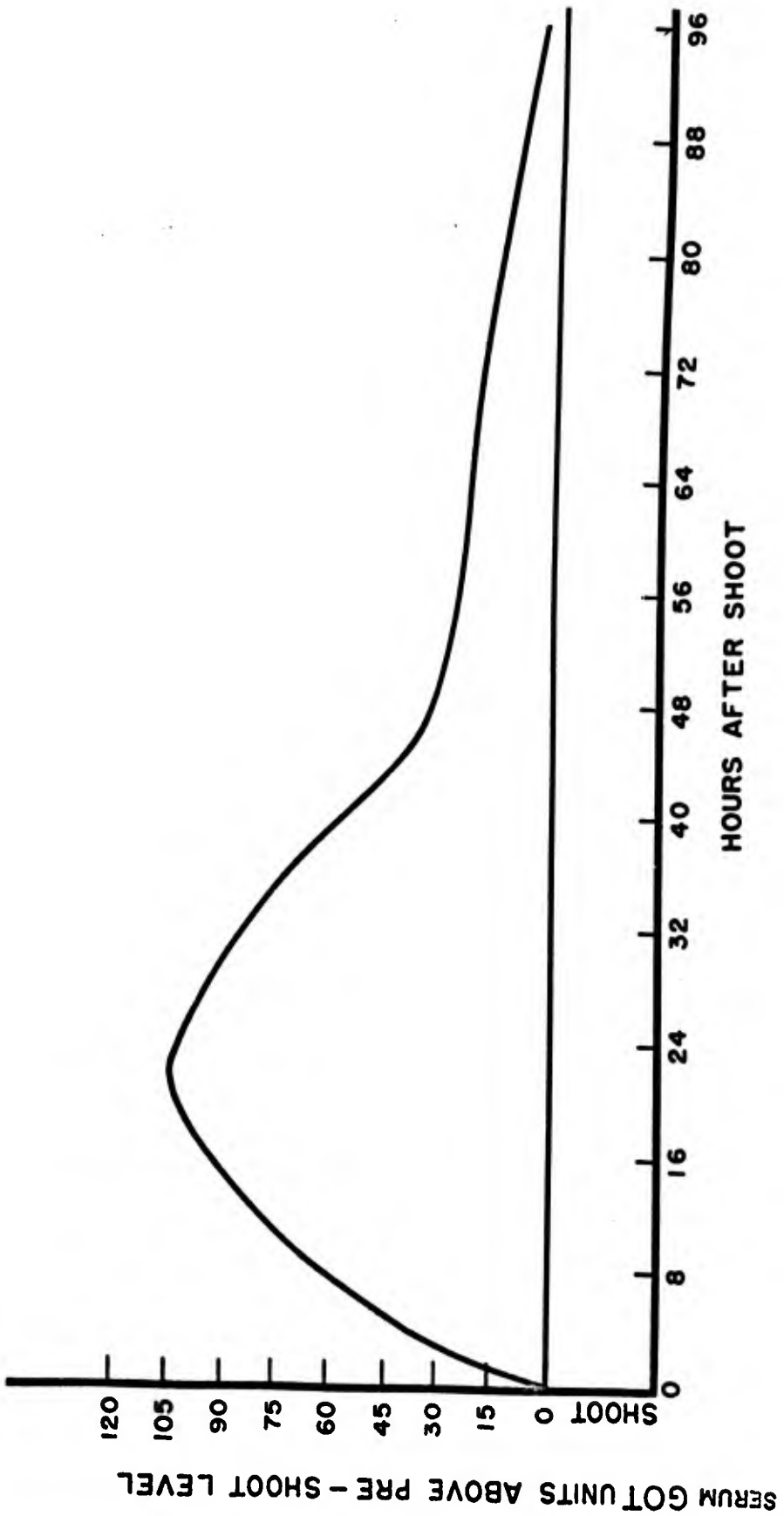


Fig. 3: Sequential record of serum GOT levels for goat GZM-4 after wounding.

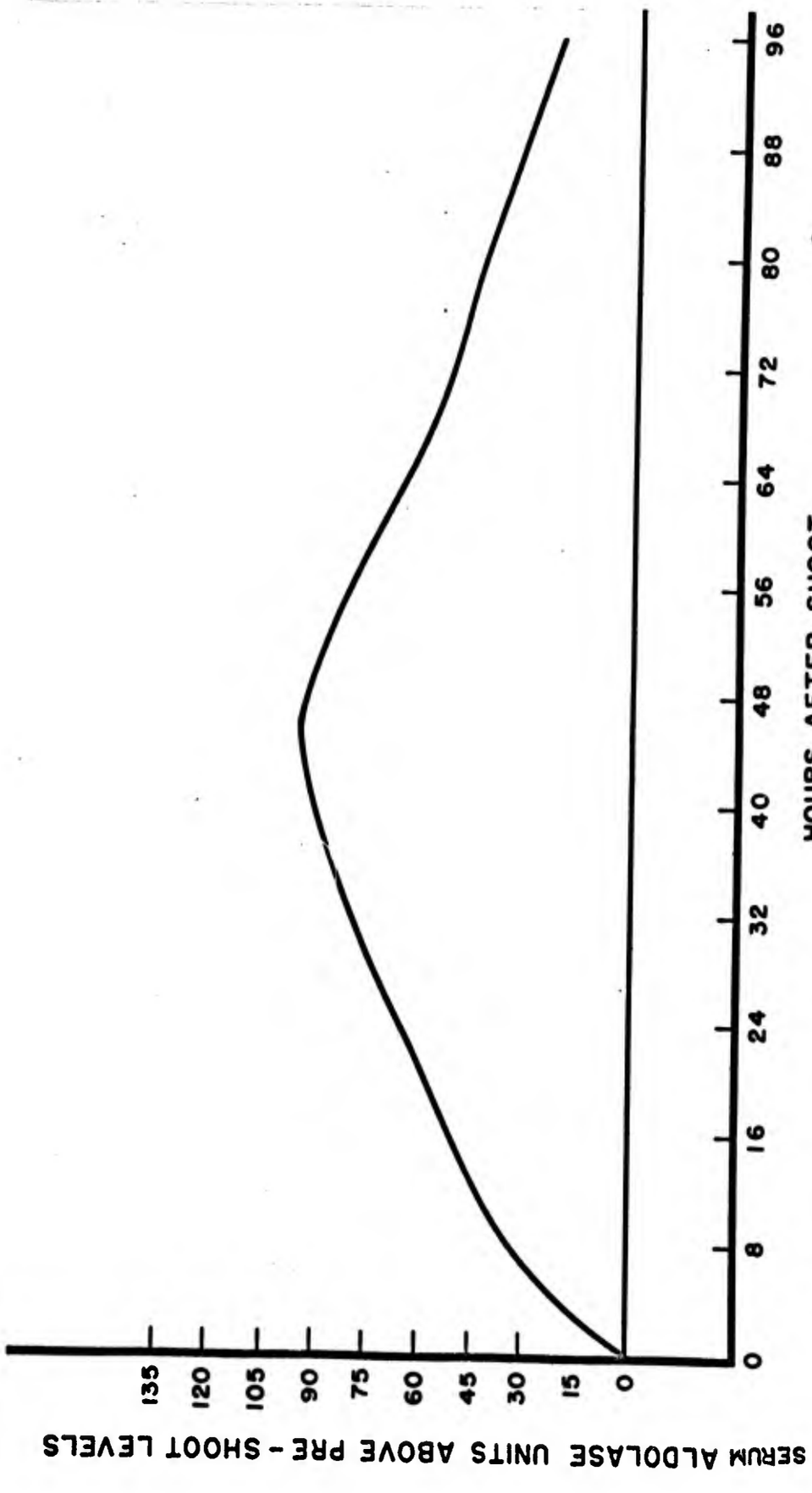


Fig. 4: Sequential record of serum ALD levels for goat GZM-4 after wounding.

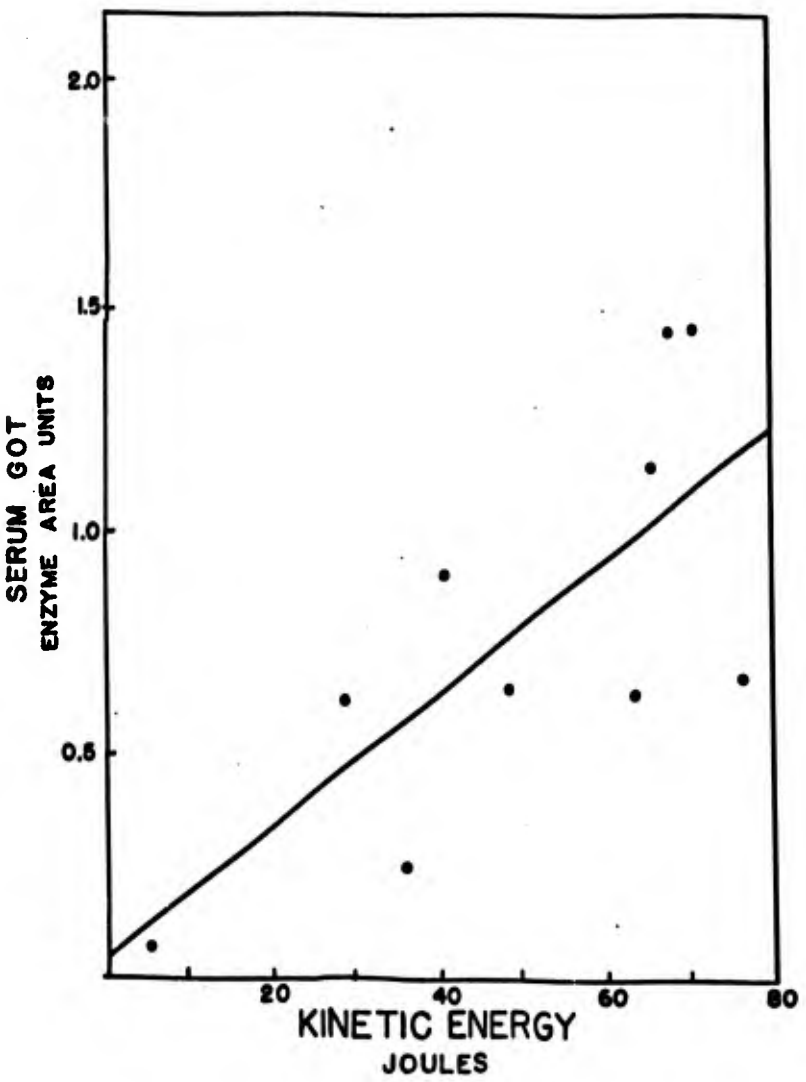


Fig. 5: Relationship between areas under GOT curves and kinetic energy input.

THE DEVELOPMENT OF A LABORATORY MODEL FOR THE TRANSMISSION OF AFRICAN TRYPANOSOMIASIS

RONALD A. WARD*
WALTER REED ARMY INSTITUTE OF RESEARCH
WASHINGTON, D. C.

The laboratory colonization of tsetse flies and subsequent development of mass rearing procedures (NASH et al, 1964, 1965a, 1965b, 1966) have provided the foundation for conducting trypanosome transmission experiments in temperate laboratories. These studies were initiated to develop a model for the transmission of African trypanosomiasis with Trypanosoma brucei Plimmer & Bradford as the parasite, Glossina austeni (Newstead) as the invertebrate host, English lop-eared rabbits as maintenance animals and guinea pigs as donor and recipient hosts. Glossina austeni was selected as the tsetse host due to its availability in culture and the Lugala I strain of Trypanosoma brucei as the parasite due to its non-infectivity to man and close relationship to T. rhodesiense. The feeding procedure for the flies required the use of lop-eared rabbits and the guinea pig is susceptible to this strain of T. brucei.

Materials and Methods

All flies in these experiments were reared from pupae produced by the Glossina austeni colony at the Tsetse Research Laboratory, Department of Veterinary Medicine, University of Bristol, England. Groups of 200, 17-19 day old pupae were packed in expanded polystyrene containers, shipped to London by overnight express and placed in groups of 50 in $\frac{1}{2}$" dry sand in special plastic containers for maturation. The jars were placed on a rack over a pan of water in a vault maintained at 24°C with a relative humidity of 80% and a

---

* These studies were sponsored by a U.S. Secretary of the Army's Research and Study Fellowship to conduct research on arthropod-borne diseases at the London School of Hygiene and Tropical Medicine.

light cycle of twelve hours darkness and twelve hours subdued light. Each jar had a cloth bottom and screened top to permit free passage of humid air to the incubating pupae. Pupal containers were examined daily and when emergence started (day 30), adults were transferred daily between 0900 - 1100 hours to Geigy-type cages. The cages are constructed of $\frac{1}{8}$" diameter stainless steel rod welded to form a frame with external dimensions of 6" x 3.4" x 2". A rectangular bag of black Terylene netting is stretched over the frame and is closed with a rubber band. Flies were transferred to the cage by prying up one corner of the pupal container and holding the Geigy cage above it or releasing new adults into a 2' x 2' x 2' mosquito cage and collecting individuals by hand in 3" vials. Usually ten flies of the same age were placed in each Geigy cage. Groups of five to six cages were placed in enamel pans on a rack in the vault. The bottom of each pan was covered with paper toweling to collect fecal material and tsetse pupae. No deliberate attempt was made to mate the flies. Approximately one-third the cages contained mixed sexes of flies so some mating and pupal production occurred. As fly emergence from any single batch of pupae occurred over a span of one week it was necessary to have a supply of infected donor animals prepared for feeding these flies.

In December 1966, three guinea pigs were inoculated with Trypanosoma brucei (Lugala I= EATRO 207) which had been frozen 10/12/65 after three consecutive blood passages from a cyclical transmission through tsetse flies in Africa. Infected blood from two of these guinea pigs was pooled in a stabilate, placed in ampules in 1 ml. aliquots and frozen at -70°C (CUNNINGHAM and HARLEY, 1962) to provide a uniform inoculum for the transmission experiments. At monthly intervals, beginning in January 1967, three to four guinea pigs were inoculated with 0.1 - 0.2 ml. of the thawed stabilate by the intraperitoneal route. Abundant trypanosomes were present in the peripheral circulation five to ten days after inoculation. Infective feeds could be made during a three week period when stumpy forms were always observed. Later experiments used a strain transmitted by Glossina austeni during this period.

Prior to fly feeding, the hair was clipped from each side of a guinea pig, the animal anesthetized with pentobarbital sodium and placed in a Geigy restraining board. Four guinea pigs could be placed on each board with two fly cages placed between each guinea pig (GEIGY, 1948). Flies of known age groups were fed on an infected animal in darkness for 15-20 minutes. Afterwards, the number of fed flies was recorded, and the unfed flies were transferred to a separate container. On the second and third day after the infective feed, the flies were fed on a normal guinea pig as a precaution against mechanical transmission of trypanosomes from the proboscis of the recently fed fly. For the next 7-14 days the flies were given blood meals from the ears of lop-eared rabbits (NASH et al, 1966). Lop-eared rabbits are the preferred maintenance host due to the extremely large vascular surface of the ear that is available for

tsetse feeding. In general, flies will take a greater quantity of blood from these rabbits than from guinea pigs; hence, survival is enhanced. At the end of this period the possibility existed that the flies harbored infective metacyclic trypanosomes in their salivary glands. To prevent infection of the rabbits, the flies were then transferred to new indicator guinea pigs for subsequent feeds in groups of four to eight cages. Blood smears were taken from all experimental animals to determine if transmission of trypanosomes occurred. Tsetse flies were fed six days weekly.

The principles of laboratory animal care as promulgated by the National Society for Medical Research were observed.

A detailed record was kept of the feeding of each cage of flies and the fly mortality. The flies were killed and dissections were made of the salivary glands, proventriculus, midgut and hindgut to determine the presence of trypanosomes 21 - 59 days after an infective meal.

Results

Relation of sex and age of fly at infective feed to infection of fly

In a series of six experiments tsetse flies were segregated into age groups of day one flies (3 - 27 hours after emergence from pupae) and day two to three flies (28 - 72 hours old) and offered an infective meal. After subsequent maintenance and donor blood meals were completed the flies were sexed and dissections made for trypanosome infections. The results of the midgut infections are summarized in Table I.

Day one male flies had a mean infection rate of 21.9% while the day one females were slightly higher with 29.2% showing midgut infections. As the difference between the means was not statistically significant, the two groups could be pooled to give an average infection rate of 26.7% for day one flies. Only a single female from 173 day 2-3 flies was infected. The infection rate in the older flies was 0.6%.

Infection rates of day-old flies and parasitaemia levels of guinea pigs

Although fly infectivity experiments were only made on guinea pigs with overt parasitaemia and stumpy forms present, there was considerable variation among these parasite counts. An attempt was made to determine if a correlation existed between the percentage of flies developing midgut infections and the trypanosome density in the peripheral circulation of the guinea pigs at the time of fly feeding. Differential counts of the stumpy forms were made as these are considered to be the stage infective to the fly. Table II presents the data from experiments where at least four surviving

flies were available for dissection. A statistical analysis of the data indicated that no correlation existed between total trypanosome counts or numbers of stumpy forms and midgut tsetse fly infections within the levels of parasitaemia of these experiments.

Distribution and nature of trypanosome infections in tsetse flies

A number of the fly dissections were studied in detail to determine the localization of trypanosomes and their relative abundance within various organs of the digestive system. The dissections were made at periods ranging from 21-59 days after a meal on an infected guinea pig. Flies were usually starved for a day prior to dissection so the partially digested blood meal would not obscure the dissection. The alimentary tract was removed intact, washed in 0.85% saline solution, placed in a clean saline drop and loosely covered with a coverslip for observation under the microscope. If trypanosomes were observed, the digestive system was severed at the level of the salivary glands, proventriculus, midgut and hindgut and each portion crushed on a separate slide. Some slides were stained with giemsa solution for further study. A rough quantitation of the degree of infection used the following scale:

\- No trypanosomes observed
\+ 1 - 10 parasites
++ 11 - 100 parasites
+++ 101 - 1000 parasites
++++ 1001 or more parasites

A summary of the dissection of 21/27 infected tsetse flies is shown in Table III. Six flies which were examined for midgut infections are not listed in the Table. Infections of the midgut and proventriculus were fairly uniform. No extremely low (+) and very few moderate (++) infections were encountered. This may be related to the mature age of the infection in the fly at the time of dissection or the inherent susceptibility of the fly. Only three salivary gland infections were observed. Two of these occurred in females from cage #5 which were infective to guinea pigs. The proportion of tsetse flies with salivary gland infections was 3/97 or 3.1%.

The presence of trypanosomes in the hindgut of some flies was rather unexpected. The possible significance of this observation will be discussed.

Metacyclic and crithidial trypanosomes were seen in salivary gland smears. Proventricular stages were abundant in the proventriculus and were often present in the midgut. Typical trypanosome stages were abundant in the midgut. The trypanosomes observed in the hindgut and rectum smears were the typical trypanosome stage with an occasional proventricular stage.

Cyclical transmission of Trypanosoma brucei to guinea pigs

An outline of the cyclical transmission of the Lugala I strain to guinea pigs and back to flies is presented in Table IV. Infection of the first recipient guinea pig, C-964, by fly bite apparently occurred no later than day 40 after the initial infective meal of the tsetse. The second transmission was more rapid with infection to H-908 becoming patent on the sixth or seventh day after feeding of infected tsetses. H-910, the final animal in the series, became patent between the eighth and twelfth day after feeding infective flies. The subsequent reinfection of new flies from H-908 and H-910 completed the cycle of transmission.

Discussion

These studies were originally conceived as an analysis of the genetic factors of the tsetse fly that are related to fly infectivity with polymorphic trypanosomes. Shortly after the project started it was evident that many of the African observations of WIJERS (1958), WIJERS and WILLETT (1960) on G. palpalis and T. gambiense and BAKER and ROBERTSON (1957) on G. morsitans and T. brucei and T. rhodesiense might not be applicable to the Glossina austeni - Trypanosoma brucei host - parasite system. Hence, the orientation in this work was directed towards an evaluation of certain non-genetic parameters of this system relating to the infectivity of the invertebrate host.

The age of the tsetse fly at the time of the infective blood meal appears to be the most important physiological condition of the vector for successful laboratory infections with T. brucei. Day old tsetse flies had a midgut infection rate of 27% while in older flies the level of infection was less than 1%. This is similar to the observation of WIJERS (1958) in G. palpalis.

WILLETT (1966) presents detailed morphological evidence that the peritrophic membrane in Glossina austeni and other Glossina species is incompletely developed in 24 hour old flies. This permits trypanosomes to readily reach the ectoperitrophic space in and near the proventriculus or easily penetrate through the peritrophic membrane. In older flies the peritrophic membrane appears to act as an impenetrable barrier except for a few rare trypanosomes which go through the mid- and hindgut until the open end of the peritrophic membrane is attained. They pass forward through the ectoperitrophic space to the proventricular region.

The role of temperature on the susceptibility of the tsetse to infection is not clear. It is known that the incubation of tsetse pupae at 29°C as compared with 22°C can accelerate the developmental cycle of T. vivax in G. palpalis (DESOWITZ and FAIRBAIRN, 1955). A suggestion was made that G. austeni might become

more susceptible to infection by incubating the pupae at a higher temperature. Two groups of 50 pupae were incubated at 30°C and 80% relative humidity for five and ten days respectively. In the former group fifteen adults emerged. Most of these were deformed or stunted. Seven of these fed on day one on an infected host and one fly developed a midgut infection. No flies emerged from the group incubated for ten days. This experiment indicates that the pupa of Glossina austeni cannot tolerate higher incubation temperature for an extended period.

ROBINSON (1912) and other workers have stated that the presence of stumpy forms is necessary for the infection of tsetse flies. WIJERS and WILLETT (1960) presented statistical evidence that the infection rate of flies was influenced by the absolute number of stumpy forms present at the time of feeding. On the other hand, BAKER and ROBERTSON (1957) found no differences in G. morsitans fed on guinea pigs infected with various numbers of stumpy forms of T. brucei and T. rhodesiense. A recent study indicated that G. palpalis could become infected from a blood meal of T. gambiense apparently containing only long thin forms (MSHELBWALA, 1967). The present study tends to agree with that of BAKER and ROBERTSON. No feeds were made of Glossina austeni on guinea pigs which lacked stumpy forms so the absolute role of this stage in G. austeni infectivity still requires confirmation.

The distribution of trypanosome stages within the salivary glands, proventriculus and midgut were in agreement with previous studies of the polymorphic trypanosomes in tsetse flies. The unaccounted presence of trypanosomes in the hind gut of certain flies requires clarification. The most logical explanation is that of contamination during dissection of the fly. Partially digested blood from the midgut may have been expressed through the pyloric valve into the hindgut dislodging some trypanosomes. This can be determined by histological study of infected flies. A remote possibility exists that in an abnormal host-parasite system polymorphic trypanosomes may rarely undergo multiplication in the hindgut.

Although cyclical transmission was clearly demonstrated of the Lugala I strain of T. brucei to guinea pigs further studies are required to determine the factors responsible for the difference between midgut and salivary gland infection rate. Glossina pallidipes (LEWIS and LANGRIDGE, 1947) and G. morsitans (BAKER and ROBERTSON, 1957) have both been shown to have midgut and salivary gland infections at higher levels than the G. austeni observed here. This could be related to differential susceptibility of T. brucei strains to tsetse species, inherent genetic differences in fly susceptibility, use of probing procedures in detecting fly infections as contrasted to salivary gland dissections and use of different donor species as trypanosome hosts.

Summary

Glossina austeni can be readily infected with the polymorphic trypanosome, T. brucei (Lugala I) using guinea pigs as donor and recipient hosts and lop-eared rabbits as maintenance hosts. The age of the fly at the time of the infective meal is the most important factor in successful laboratory infections. Day old flies had a midgut infection rate of 27% while less than 1% of older flies became infected. 3% of day old flies developed salivary gland infections. A complete cyclical transmission of T. brucei was demonstrated. Sufficient replicates of the infection procedures have been performed to indicate that this system should be a useful model for the study of African trypanosomiasis.

References

1. BAKER, J.R. & ROBERTSON, D.H.H. (1957). Ann. trop. Med. Parasit., 51, 121.

2. CUNNINGHAM, M.P. & HARLEY, J.M.B (1962). Nature, 194, 1186.

3. DESOWITZ, R. & FAIRBAIRN, H. (1955). Ann. trop. Med. Parasit., 49, 161.

4. GEIGY, R. (1948). Acta trop., 5, 201.

5. LEWIS, E.A. & LANGRIDGE, W.P., (1947). Ann. trop. Med. Med. Parasit., 41, 6.

6. MSHELBWALA, A.S. (1967). Nature 215, 441.

7. NASH, T.A.M. & KERNAGHAN, R.J. (1964). Ann. trop. Med. Parasit., 58, 168.

8. _____ & _____ (1965a). Ibid., 59, 266.

9. _____ & _____ (1965a). Bull. ent. Res., 56, 65.

10. _____ & JORDAN, A.M. & BOYLE, J.A. (1966). Trans. R. Soc. trop. Med. Hyg., 60, 183.

11. ROBINSON, M. (1912). Proc. R. Soc. (B), 85, 527.

12. WIJERS, D.J.B. (1958). Ann. trop. Med. Parasit., 52, 385.

13. _____ & WILLETT, K.C. (1960). Ibid., 53, 314.

14. WILLETT, K.C. (1966). Exp. Parasit., 18, 290.

TABLE I. *Trypanosoma brucei* midgut infections in *Glossina austeni*

Experiment number	Day 1 flies		Day 2+ flies	
	Males	Females	Males	Females
	No. infected / No. dissected	No. infected / No. dissected	No. infected / No. dissected	No. infected / No. dissected
1	-	3 / 5	-	0 / 16
2	1 / 3	3 / 7	0 / 5	0 / 13
3	2 / 10	8 / 27	0 / 7	0 / 37
4	1 / 10	2 / 15	0 / 19	1 / 34
5	1 / 1	0 / 4	0 / 13	0 / 29
6	2 / 8	3 / 7	-	-
Total	7 / 32	19 / 65	0 / 44	1 / 129

TABLE II. Trypanosome counts of guinea pigs infected with *Trypanosoma brucei* and midgut infection rates in *Glossina austeni* which fed on these animals

Guinea pig no.	Date	Total no. trypanosomes per c.mm.	No. stumpy forms per c.mm.	No. flies dissected	% infected flies
C-955	23 Jan.	176,600	47,700	4	75
C-959	1 Feb.	126,700	8,900	8	25
C-959	2 Feb.	150,000	21,000	10	10
H-924	11 Mar.	323,300	109,900	19	16
H-924	14 Mar.	699,900	105,000	6	0
H-908	24 Mar.	220,000	68,000	9	11
H-918	4 Apr.	156,600	36,000	5	60
H-918	7 Apr.	93,300	20,500	4	0

TABLE III. Distribution of trypanosomes in Glossina austeni infected with T. brucei. (- = no parasites, + = 1 to 10 parasites, ++ = 11 to 100, +++ = 101 to 1000, ++++ = more than 1001. A blank space indicates no observation.)

Tsetse no.	Age (days)	Presence of trypanosomes in: Salivary glands	Proventriculus	Midgut	Hindgut
3	35	-		++++	-
4	46	-		+++	-
5a	57	+++	++++	+++	-
5b	57	+++	++++	++++	-
12	48	-	+++	++++	
19a	39	-	++++	+++	
19b	39	+	++++	+++	++++
21	58	-		++	++++
23a	56	-		++	-
23b	56	-		++	+++
25	59	-	++++	++	++
29	47	-		++++	
36	44	-	+++	+++	
40a	37	-	++++	+++	+
40b	37	-	+++	+++	++
41	31	-	+++	+++	-
43	31	-	+++	+++	
52	33	-	++++	++++	
57	28	-	++++	++++	-
58a	30	-	++++	++++	-
58b	30	-	+++	++	-

TABLE IV. Cyclical transmission of Trypanosoma brucei to Guinea Pigs by Glossina austeni and tsetse reinfection

Day	Observation
1	6/10 female day 1 tsetse flies fed on guinea pig C-955 infected with blood passaged T. brucei (Lugala I)
2 - 3	Flies fed on uninfected guinea pig C-975
4 - 9	Flies fed on lop-eared rabbit
10 - 41	Flies fed on uninfected guinea pig C-964 Blood films from C-964 negative on days 21, 27, 34 and 41
42 - 49	Flies transferred to uninfected guinea pig H-908
49	Blood film C-964 strongly positive, H-908 with low trypanosome infection
50 - 56	Flies transferred to uninfected guinea pig H-910
55	C-964 and H-908 both positive
57	4 surviving flies dissected; 2 with salivary gland & midgut infections; 1 with midgut infection Blood film H-910 negative
64	H-908 and H-910 positive
59 - 62	Uninfected day 1 flies fed on H-908 (see experiment 5, Table I)
69 - 73	Uninfected flies fed on H-910 (see experiment 6, Table I)
92	Flies from experiments 5 & 6 with midgut infections

ATMOSPHERIC ELECTRICAL STRUCTURE

WILLIS L. WEBB
ATMOSPHERIC SCIENCES LABORATORY
U. S. ARMY ELECTRONICS COMMAND
WHITE SANDS MISSILE RANGE, NEW MEXICO

1. INTRODUCTION

Man's knowledge of the earth's electromagnetic structure has developed piecemeal, in part because technological progress has dictated such an approach, but more importantly because a truly comprehensive electromagnetic framework has eluded the geoscientist. The fact that most geoelectromagnetic phenomena exhibit strong diurnal features has, from the beginning, pointed toward a tidal origin for some of the more exotic phenomena. Lack of correlation with obvious gravitation tidal inputs forced the consideration of tides generated by thermal effects. Lack of experimental and theoretical support for such tidal systems limited progress in this area until the discovery in 1964 by Beyers and Miers (1, 2), of the Army's Atmospheric Sciences Laboratory at White Sands Missile Range, of strong diurnal temperature variations with associated tidal type wind variations at the stratopause level.

The discovery of these stratospheric tidal circulations (3) came as a result of an Army-initiated synoptic rocket exploration of the atmosphere above levels to which balloons can effectively penetrate. The Meteorological Rocket Network (MRN) (4) has uncovered certain other important aspects of the earth's atmospheric structure. These include interchange of mass between the hemispheres across equatorial regions and a profusion of detail structure which increases in significance with increasing altitude. This synoptic MRN has a basic aim of relating the physical processes of this middle atmospheric region with the "weather" of the lower atmosphere and the electrical structure of the ionosphere region. This paper concerns the impact of these tidal circulations on our knowledge of the physical processes which produce the extremely important electrical state of the atmosphere.

It has long been known that the observed structure of the ionosphere could not be simply derived from static considerations of a solar flux into a spherically stratified uniform absorber (i.e., the typical Chapman layer). Discovery of these stratopause tidal circulations of hemispheric extent, which have no previous counterpart

in our experience, has provided a new opportunity to consider physical processes which may provide a unified and comprehensive picture of the earth's electromagnetic structure. These processes center on differential forces imposed on electrons and positive ions by neutral vertical winds of low latitude tidal circulations in the presence of the earth's magnetic field. The electric fields which are produced by this mechanism result, through the Hall effect, in strong dynamo electric currents the boundary effects of which produced many of the electrical phenomena with which we are familiar.

A very coarse picture of the earth's electromagnetic generating system is presented here. The details await acquisition of more adequate synoptic data on the mechanical and electrodynamical structure of the upper atmosphere.

2. DIURNAL CIRCULATIONS.

The diurnal circulations which are indicated by the MRN data to be a prominent feature of the stratopause region have been outlined in general by Webb (3). Additional observational data have shown the importance of the stratospheric tidal jet (STJ) to the structure of the summer easterly circulation (5), and data from Tartagal, Argentina presented by Webb and Tabanera (6) have demonstrated that the summer tidal circulation extends, in its well organized form, to equatorward of 22° latitude. A summer diurnal study at Ascension Island (7) has shown that the tidal circulation is markedly different at lower latitudes.

The observed oscillation of stratopause air away from the equator during the morning hours indicates the formation of a longitudinal trough of reduced pressure in the leading edge of the diurnal heat wave at the equatorial boundary with an associated downward motion of mesospheric and lower ionospheric air on each side of the equator. Conversely, the return tidal motion toward the equator during the afternoon and evening produces, as a result of the boundary conditions, a high pressure ridge over the equator with upward motions in the same lower latitude locations. The above concepts, when considered with available ionospheric information, indicate the probability that downward motions exist in the lower ionosphere at noontime in the organized mode indicated in Figure 1 at low geographic latitudes (5-15°) between the trough and crest of the diurnal heat wave.

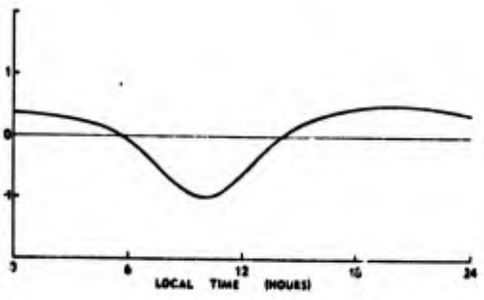


Figure 1. Normalized diurnal character of vertical motions across the mesopause surface at low geographic latitudes

The speeds of these vertical winds and their geographic

extent are of prime interest. The narrow latitudinal extent of this vertical circulation which has been demonstrated lends evidence for stronger winds than could have earlier been predicted, with speeds in the meters per second range to be expected in the E region. In addition, all of the data point toward the summer hemisphere's tidal circulation being stronger than that of the winter hemisphere.

3. ELECTRICAL PROCESSES

Motion of a gas containing a mixture of electrons and ions through a magnetic field will result in production of an electromotive force oriented in the direction of flow as a result of differing mobilities of the charge carriers. Fejer (8) has stated the general relation

$$q(\vec{E} + \vec{u} \times \vec{B}) + M\vec{g} + M \nu \vec{v} - \vec{\nabla} p n^{-1} = 0, \qquad (1)$$

where $\vec{E}$ is the electric field, $\vec{u}$ the velocity of the charged ($q = 1.6 \times 10^{-19}$ coulombs) particles, $\vec{B}$ the magnetic field, M the mass of the particles, $\vec{g}$ the gravity, ν the collision frequency, $\vec{v}$ the velocity of the charged particles relative to the neutral gas and $\vec{\nabla} p n^{-1}$ is the pressure gradient of the charged molecules of n particles m^{-3}. He derived the vertical component solution for the case of a vertical wind exerting a force (F) across collisions between neutral and charged molecules to obtain the charged particle speed (v) relative to the neutral gas:

$$v = -\frac{F}{M}\left(\frac{\omega}{\nu^2 + \omega^2}\right), \qquad (2)$$

where ω is the gyrofrequency for the particular charged particles involved. The gyrofrequency is given by $\omega = qBM^{-1}$, which indicates that the gyrofrequency of electrons will be significantly greater than that for the ions as a result of the difference in mass (9.1×10^{-31} kg for electrons and 1.67×10^{-27} kg or greater for ions). When the gyrofrequency is greater than the collision frequency, the charged particle will be effectively bound to the magnetic field, while in the opposite case, the charged particle will move with the neutral wind. The gravitational and pressure gradient terms of Eqs. (1) and (2) can be considered negligible in the 70-150 km region.

Collision frequencies are significantly different for ions (assumed to be molecular oxygen) and electrons as is indicated in Figure 2. In both cases the collision frequency decreases with increasing altitude in accord with the usual density scale height. These data would then indicate that in the 70-150 km altitude region particularly, electrons would be restrained in their motion while their partner positive ions would move with the neutral wind ($\vec{w}$). The force producing this separation will increase with decreasing height as a result of an increasing collision frequency until the collision frequency becomes high enough to carry the electrons along with the wind also. The current which would result from this impressed force, if no limiting impedance were involved, can be obtained by evaluating the difference in relative velocities in this height region of electrons ($\vec{w}_e = 0$) and ions ($\vec{w}_i = \vec{w}$) from Eq (2) and using this

difference ($\vec{w}$) as a measure of the rate of charge separation. Thus, if a zero impedance return current path is available, a velocity of separation ($\vec{w}$) of the positive ion and electron pairs yields a current density ($\vec{J}$)

$$\vec{J} = nq\vec{w}. \qquad (3)$$

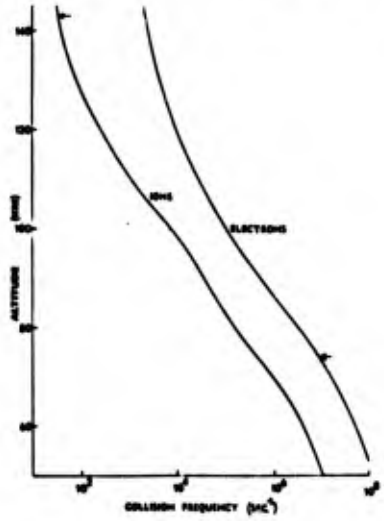


Figure 2. Typical collision frequencies versus altitude for ions (oxygen molecules) and electrons. Arrows indicate gyrofrequencies

In the open circuit case where a current cannot flow the electric field ($\vec{E}$) required to prevent further charge separation (that is, to force the positive ions also to move upstream with a speed of -w) can be calculated approximately from the relation

$$q\vec{E} = -M\nu\vec{w} \qquad (4)$$

Generally in the atmosphere a circuit is neither absolutely open nor closed. The presence of mobile charge carriers affords the atmosphere a certain conductivity through a unit volume which is defined by the field form of Ohm's law

$$\vec{J} = \sigma\vec{E} \qquad (5)$$

where $\vec{J}$ is the current density produced by an electric field $\vec{E}$ and σ is the conductivity in mhos per meter through a column of one square meter cross section. The resistivity (ohm meters) is the reciprocal of the conductivity.

4. ELECTRICAL STRUCTURE

The electrical structure of the atmosphere varies with time and space in a complex fashion and thus can only be represented in a limited fashion by a particular model or set of data. While some general features of electron structure are adequately portrayed by the Chapman theory (9) of layer formation, it is apparent that many important features can only be explained by significant deviations from this assumed radiational and static equilibrium condition. It is thus desirable to make gross generalizations relative to the general electrical structure to obtain a first look at the nature of important physical processes which may occur, and then to compare the expected results of these processes with deviations from the static electrical structure which are observed to exist. This procedure is reasonable as long as it is clearly remembered that a uniformity has been assumed which does not exist.

A magnetic field exerts a complicating influence on the conductivity structure of an ionized gas. Typical vertical distributions of specific, Pederson and Hall conductivities for a mean midlatitude noontime atmospheric model are illustrated in Figure 3.

The very low conductivity of the lower atmosphere ($\sim 2 \times 10^{-14}$ mhos per meter) results from the high collision frequency (and thus low mobility) produced by high air density and the small number

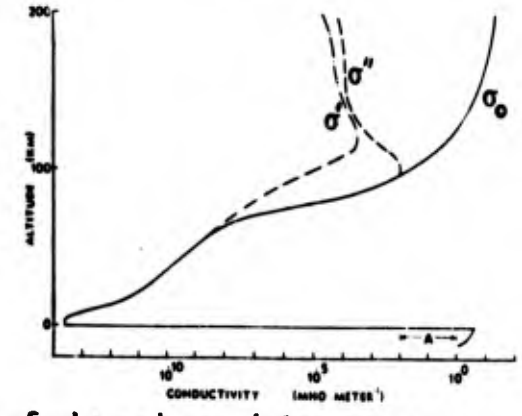


Figure 3. Typical vertical distributions of the specific (σ_o), Pederson (σ') and Hall (σ'') conductivities for mid-latitude nocn conditions after Cole and Pierce (10) from the surface of the ocean to 100 km and Hanson (11) from 100 km upward. The symbol A refers to the range of variable conductivity of the earth's crust.

of charged particles. The conductivity increases rapidly with altitude to values of the order of 10^{-9} mhos per meter at the base of the D region, partly because of an increase in number density of charged particles. Above about 50 km the presence of free electrons becomes important as a result of gross differences in mass and collision cross sections of the charge carriers and the conductivity becomes nonisotropic, with a component structure which is illustrated in the upper portions of Figure 3. At low latitudes the specific conductivity (σ_o) is most applicable meridionally, the Pederson conductivity (σ') is most applicable vertically and the Hall conductivity (σ'') is most applicable zonally.

Vertical downward motions produced by the tidal circulations will, in low latitudes of the morning and early afternoon sunlit hemisphere, result in electrical charge separation in the vertical direction with positive charges forced downward by collision between those ions and molecules of the neutral flow. Using the open circuit approximation of Eq. (4)(in the absence of an effective return circuit path) with vertical speeds of one meter per second at 80 km and 10 mps at 100 km acting on ions of molecular weight 32, a vertical upward-directed equilibrium electrical field of the order of 0.066 V m^{-1} at the 80 km level is obtained, decreasing in strength below and above that level with a value of 0.015 at the 100 km level as illustrated in Figure 4. The charge separation process and its opposing electric field will rigidly maintain a positive space

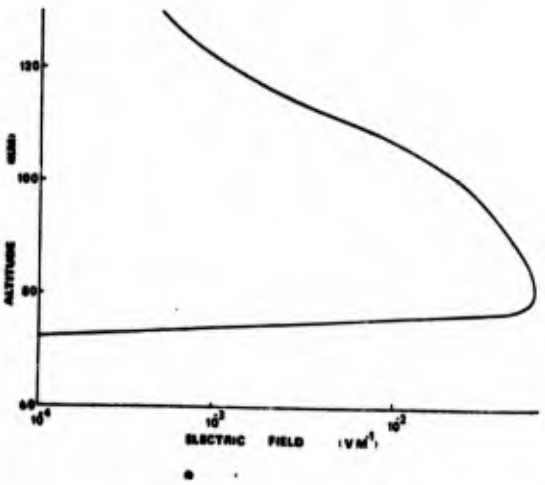


Figure 4. Vertical distribution of the electric field generated by a noontime mesopause downward motion of the tidal circulation of one meter per second at 15 degrees latitude.

charge in the stratopause region and a negative space charge in the ionosphere during the morning and early afternoon. This region of separated charge will exhibit a diurnal structure under control of the tidal circulation which continuously rotates around the earth at a speed of approximately 460 mps at 15 degrees latitude.

The same reasoning may be applied to evaluate the electrical structure which will result from the tidal upward flow as the heat wave recedes locally from 2 PM until sunrise. Using vertical circulation values of 0.4 of the morning values (from mass continuity considerations) for the late afternoon period from 2 PM to sunset, mean values of the vertical electric field will be approximately one fourth the noontime values illustrated in Figure 4, with the principal additional difference being that the direction of the resulting electric field will be reversed. The efficiency of this source of electric field generation will decrease during the evening as a result of the diurnal decrease in electron density.

Juxtaposition of electric and magnetic fields in an ionized gas will result in imposition of forces on charged particles in the direction of the cross product of the electric and magnetic vectors (8). In the D and E regions the ions are immobilized (relative to motions through the gas) by collisions so that electrons are the principal carriers of this "Hall" current. Velocity of electron motion (unhampered by collisions) constituting such a current is given by

$$\vec{v} = \frac{\vec{E} \times \vec{B}}{B^2} \qquad (6)$$

Using values of $B = 0.34 \times 10^{-4}$ webers m^{-2} directed northward and $E = 0.066$ V m^{-1} (Figure 4) directed upward as obtained during the morning hours a westward velocity of electron motion of approximately 2×10^3 mps is obtained for 80 km altitude.

The current density carried by such a system can be calculated by use of Eq. 3, which yields an 80 km value of 2.6×10^{-7} a m^{-2} as is illustrated in Figure 5 when a concentration of 10^9 electrons per m^3 is used with the motion derived above. Similar calculations at other altitudes indicate that the vertical distribution of the midday Hall current of low latitudes is of the type illustrated in Figure 5. Integration of this current density over the suspected cross section yields a total dynamo current of the order of 10^5 amperes, which is in general agreement with estimates Chapman and Bartels (12) obtained from variations in the magnetic field.

The component of the dynamo currents produced by the general circulation (horizontal) can be estimated from the closed circuit assumption of Eq. 3 for a representative 100 mps wind to be of the order of 10^{-8} a m^{-2} and will reverse direction seasonally with the monsoon. This latter effect should exhibit an influence on the upper E region electrical structure where the tidally produced structure is small and at other locations where the dominant tidally produced electrical structure is reduced.

Now the electric fields associated with these dynamo

electric currents can be calculated from Eq. 5, using values of conductivity obtained from Figure 3. This calculation yields an eastward-directed horizontal electric field at 80 km of 0.043 V m^{-1} which, over the longitudinal span of 1.5×10^7 in which the downward

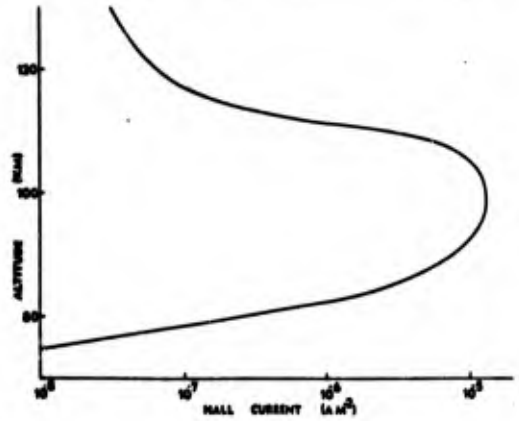


Figure 5. Hall current produced in low latitudes by the tidal downward motion as the heat wave approaches. The direction of current flow is to the east.

flow occurs, indicates a potential difference of 650,000 V between sunrise and early afternoon along the longitude circle at 15 degrees latitude. These considerations lead to the conclusion that in a longitudinal belt at low latitudes (∿ 15 degrees) there will be a region of electromotive force (emf) generation with a potential difference of the order of 6×10^5 V at the base of the E region between the early morning sector and the early afternoon sector (Figure 6) for each meter per second of vertical down wind of the tidal circulation at 80 km altitude. The direction of this potential gradient is such that the early morning sector (B) is depressed in potential relative to the early afternoon sector (A).

In the afternoon and evening the tidally generated electric fields which drive the Hall currents will reverse direction and be oriented downward. The Hall current will then flow westward after 2 PM. The electron density of the E region decreases throughout

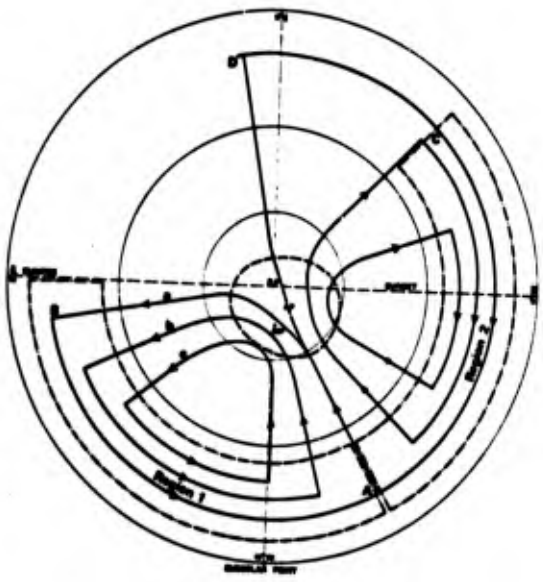


Figure 6. Equatorial plane projection of the dynamo current system. Regions 1 and 2 (inclosed by dash-dot curves) represent regions of electromotive force generated by vertical tidal motions. Positions of the rotational (RP), magnetic (MP) and auroral (AP) poles for the Northern Hemisphere are indicated. The dashed circular curve represents the centerline of the auroral zone.

this period to make the Hall current generating process ineffective during the late nighttime (smaller than noontime by two orders of magnitude). It is clear, then, that the sunset dynamo current is weaker than the morning current.

Flow of these dynamo currents in the E region ($\sim$ 100 km) of the atmosphere will result in development of a diurnal pattern of electrical potential distribution in that region. The above analysis indicates that the maximum potential will be located geographically near the point marked A in Figure 6 (near 2 PM at low latitudes). Minimum potential should be located near the point marked B. Significant diurnal differences will be induced in the hemispheric dynamo currents even at equinox times as a result of asymmetries between the rotational and magnetic axes, and as a result of hemispheric separation of the vertical tidal circulations over the rotational equator (3) there will be a region of reduced potential between the low latitude emf regions of the two hemispheres. More substantial hemispheric differences will develop as the subsolar point moves away from the equator, with the summer hemisphere being favored with a stronger tidal circulation and enhanced electron concentrations, and thus with stronger dynamo currents and greater potential differences. All of these factors indicate that potentials in equatorial regions will be different between hemispheres for similar geomagnetic latitudes and make it probable that trans-equatorial currents of diurnal character will flow between the hemispheric emf zones, probably with an annual bias from the summer toward the winter hemisphere.

Such currents will flow along magnetic field lines, so the specific conductivity (solid curve) of Figure 3 would be applicable. These currents would start near the 100 km level of one hemisphere on a magnetic field line, progress upward in a divergent field to cross the magnetic equator at a few hundred (and/or thousand) kilometers altitude, and converge back to the 100 km level at the same low latitudes of the other magnetic hemisphere.

This current will have the principal characteristics of a conduction current in the middle and upper ionosphere where charge carriers are plentiful, and will be characterized by fast individual transients of a few particles along magnetic field lines in the relatively collision-free exosphere. This low latitude interhemispheric dynamo current circuit is in the location of the inner Van Allen radiation belt (13), and presumably provides the basic energy for that phenomenon. The inner radiation belt electric current path then smooths inequalities in the hemispheric tidally produced emf regions. There should be marked diurnal variations in the basic low energy currents flowing in the region of the inner radiation belt. Some of the electrons and ions taking part in these exospheric currents will, because of their favored initial trajectory angles and kinetic energies, be trapped in the magnetic field for short or long periods. These trapped particles will drift longitudinally (electrons drifting to the east and positive ions to the west) along L-shells (14) to envelop the entire low latitude global region in a few tens of minutes in the case of particles with high energies.

Thus, the daytime dynamo regions supply the inner Van Allen radiation belt with ions and electrons which then interact with the nighttime upper atmosphere of middle and low latitudes as they drift around the globe, resulting in enhanced ionization and heating in low and middle latitudes of the nighttime ionosphere.

In order for these hemispheric generators of dynamo currents in the 100 km region to be effective in current transport, it is necessary that there be return current paths in the E region. Such paths will be minimum resistance paths from the high to low potentials regions. At low latitudes resistivity which results from a high collision frequency of neutral ions in the D region will limit ionic currents which could constitute return flows at altitudes below the emf zone, and rigidity of the magnetic field will preclude an effective electron current flow above. Positive and negative ions in low latitudes will move longitudinally as a result of this horizontal electric field in the 100 to 140 km region, and this current flow will weaken the tidally generated emf.

An important return current path is available in the direction of the magnetic field lines. Along the field lines the meridional conductivity will be that indicated by the solid curve of Figure 3, decreasing with increasing magnetic latitude until in polar regions the conductivity will be reduced to essentially the Pederson values (σ') which are representative of those regions. In high latitudes the dynamo current paths will be along the high conductivity zones of the auroral ovals. These current paths are illustrated by curves a,b and c of Figure 6. The impedance of these paths will be of the order of 10^{10} ohms m^{-2}, an order of magnitude less than that offered by the Pederson conductivity (Figure 3) of possible longitudinal paths.

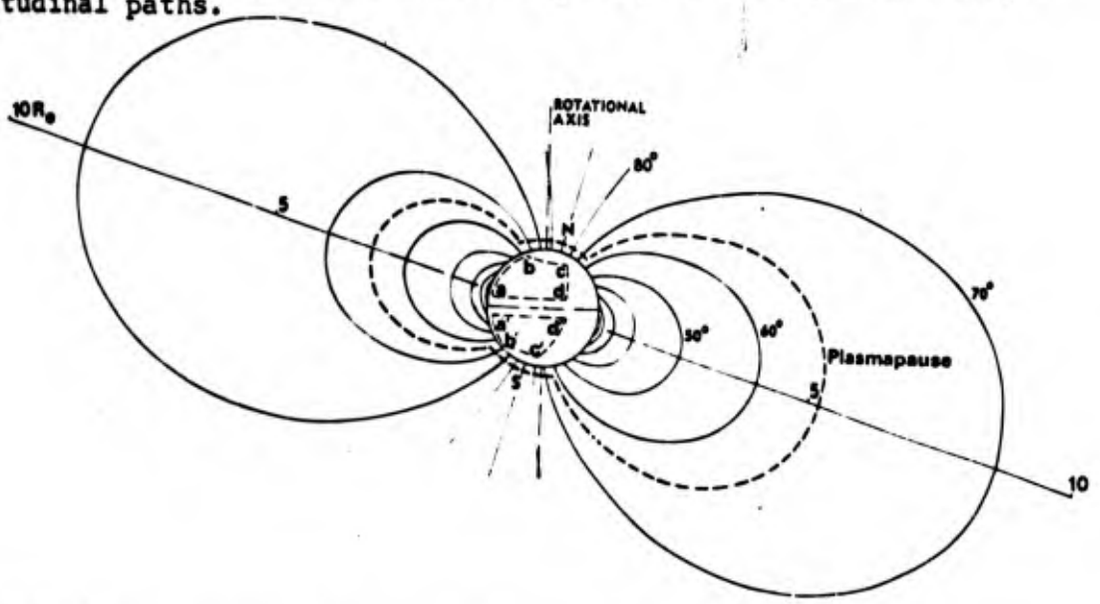


Figure 7. Equinoctial cross section of the earth's geomagnetic field when the sun is over 160 degrees west longitude. The light dashed curves represent projections of hemispheric circuit elements of the tidally generated dynamo current. The heavy dashed curves represent the plasmapause.

It is clear that the electric potential distribution of the lower ionosphere which is established by the dynamo currents will, in general, not be the same in the two auroral zones. Asymmetry of magnetic and geometric axes will produce a diurnal variation in the direction and intensity of these polar potential differences for various conjugate locales. This important diurnal variation will be further modified by the seasonal structure of the dynamo currents. Differences in potential of as much as 10^4 volts are to be expected at opposite ends of some high latitude magnetic field lines.

The results of these high latitude electric potential differences are immediately apparent on inspection of Figure 7. Currents will flow along magnetic field lines through the exosphere in response to these potential differences, forming additional circuits through the opposite hemisphere for return flow of currents generated by the tidally driven dynamo emfs. As an example, in Figure 7 the potential of point b of the Northern Hemisphere dynamo circuit (dashed curve) will be higher than the potential of b' because b is electrically closer to the high potential at d than is b' to the high potential at d' . Using a potential difference of 10^4 volts with an assumed path impedance of 10^5 ohms m^{-2}, a current flow of the order of .1 ampere m^{-2} should result along the magnetic field lines connecting b and b' . Precipitation currents in excess of this value have been observed in the auroral zones (15, 16).

Twelve hours later in the situation illustrated in Figure 8, the electric potential positions of b and b' will be reversed, with b' at a higher potential than b . The exospheric current flow in the outer Van Allen belt region will thus exhibit a marked diurnal reversal in direction as the asymmetric positions of the auroral zones alter the relative potentials of the ends of high latitude magnetic field lines. These currents apparently form the basic power source for the outer Van Allen belt.

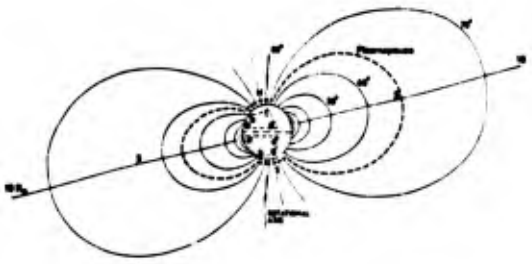


Figure 8. Similar to Figure 7 except that the sun is over 20°E longitude.

It is important to note that the regions involved in the current of both radiation belts will be firmly tied to the earth's rotation. Since most particles are charged above an altitude of roughly one earth's radii and thus participate in the currents, the effective captive mass of the earth will then be enclosed in the volume contained within the plasmapause surface and high latitude lower ionosphere indicated by the heavy dashed curves of Figures 7 and 8. These charged particles (electrons and ions) will always be transient, so that the electrical structure of the upper ionosphere

and exosphere will be the result of dynamic processes.

Geomagnetic field lines of polar regions extend beyond the plasmapause to sufficient distances from the earth that their energy density ($\sim B^2/8\pi$) is reduced to near the order of the solar atmospheric plasma ($\sim 2nkT$). Interaction of high latitude field lines with the solar wind will thus become important in the relevant physical processes. This extraterrestrial mode of electrical processes would appear from experimental evidence to occur poleward from a variable region near 70 degrees magnetic latitude (beyond the 8 L-shell). The magnetic field lines in these regions will be tied to the electric potential of the lower ionosphere in the region through which they pass. This will result in the earth exhibiting important nonsymmetrical electrical aspects to the solar wind. The north polar region will generally be at a higher potential (maximum of the order of 10^4 volts) when the sun is over the Western Hemisphere (30°E to 150°W longitude, Figure 7), with a maximum where the difference in latitude of the magnetic and rotational equators is maximum. The reverse polarity will be observed in the case illustrated in Figure 8.

Under such conditions an external electric current will flow through the earth system under the control of the dynamo circuits. The internal portion of this current will flow from south to north when the sun is over the Western Hemisphere (Figure 7) and in the opposite direction during the following diurnal period (Figure 8). The internal path segment of this external electric circuit will almost surely be through the low impedance path of the outer radiation belt. The reason for the fast response of the outer belt to variations in the solar wind is thus obvious, and the gross source of charged particles of various energies from the solar wind which are observed in that region is more easily understood. It is important to note, however, that the impetus for this external current system is from the potential field set up by the dynamo currents.

The kinetics of charge particle transport between the solar and earth environments should exhibit interesting physical characteristics. Positively charged particles (protons) precipitating from the solar wind will, as a result of their greater mass, penetrate deeper into the earth's atmosphere than will electrons. The magnetic pole at a lower potential (the winter pole) will then be the recipient of a flux of these solar protons, the "q" of which will be deposited below the dynamo currents. The electrons involved in this current will, on the other hand, interact (arrive or depart) at a higher altitude as a result of their small mass. These factors are well illustrated by winter Polar Cap Absorption events (17). A segment of this external circuit must be established in the earth's near space, probably through the action of a Hall current produced by the ambient magnetic and electric fields.

It is interesting to note, then, that the current flow discussed above represents the earth's electrical connection with the solar environment. In the same way that the tropospheric region of high impedance evidences electrostatic phenomena, the earth-solar plasma circuits may cause the internal electrical system of the earth to operate at a different potential from that of the external

portions of the circuit. This would be caused by differing mobilities of the charge carriers which take part in the polar segments of the external circuit. A downward flux of electrons into the positive pole should be observed. Over the negative pole, however, it is possible that the current is constituted by a downward flux of protons from the solar wind and an upward flux of ambient electrons from the ionosphere. This is so because of gross differences in mass, gyrofrequency and collision frequency of the two carriers.

It may well be, then, that the earth's integral rotational system (inside the plasmapause) is maintained at some electrical potential different from the solar plasma of the earth's near space by this differential impedance. If so, that difference would be such that the earth would be negatively charged and thus at a reduced electrical potential. It should be again noted that while the earth's orbital motion and flow of the solar plasma past the earth may supply some of the energy required for this charge separation, the principal source of energy and the control agent will be the tidal circulation which derives its energy from solar heating of the ozonosphere.

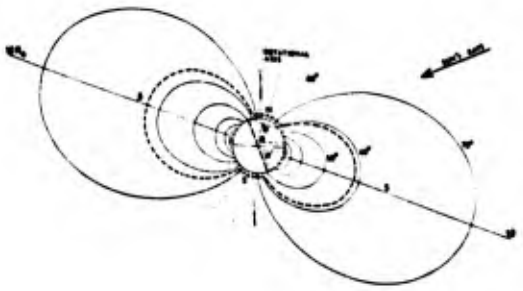


Figure 9. Similar to Figure 7 in that the observation point is over 160°W longitude, except that the local time is sunrise and the subsolar point is at 23.5°N (summer solstice in the Northern Hemisphere).

An intriguing aspect of upper atmospheric electrical and magnetic phenomena concerns the annual cycle. Magnetic and electrical phenomena generally exhibit maximum intensity at equinox times and minimum intensity at solstice time. The cause of this is apparent on inspection of Figure 9. The conductivity of the winter high latitude ionosphere is reduced in the absence of photoionization so that the usual low impedance of high latitude field lines will be high and these paths are essentially opened. In addition, the dynamo currents of the winter hemisphere do not extend to high latitudes, with a subsequent reduction in the potential difference achieved between the poles.

The conductivity of the lower atmosphere and earth (see Figure 3) beneath the dynamo circuits offers additional circuit paths which may serve to relax the E region electrical potential distribution. A high path impedance of the order of 10^{17} ohms m^{-2} in the lower troposphere will limit the current flow in this circuit to a small fraction of the total current flowing in the dynamo circuits. Observations show that the earth is the continuous recipient of a fair-weather electric current of the order of 10^{-12} amperes m^{-2}, which integrates over the globe to be approximately 1500 amperes. This

influx of positive ions is driven by a positive electric potential gradient which has a surface value of roughly 100 volts m^{-1} and decreases with altitude, indicating an overall potential difference between the earth and the upper atmosphere of more than 10^5 volts. With an overall system relaxation time of a few tens of minutes, it is clear that these phenomena can be maintained only if there is a continuous flow of current to the upper atmosphere to prevent growth of the approximately 10^5 coulombs of negative charge which is resident on the earth (18).

Such a return current path is believed to exist in thunderstorm regions, particularly in the lightning discharge path. Positive charge transport upward of an average of 20 coulombs per stroke through a global average of approximately 100 strokes per second which is indicated by observational data will provide adequately for the return current. Greater than order of magnitude reductions in lower tropospheric path impedance will be effected through these thunderstorm regions, so that effectively the earth is tied through lightning discharges to the potential of the dynamo circuit over their locale. A maximum in thunderstorm occurrence is observed late in the day (~7PM), and it is postulated that the fair-weather electric field (~10^5 volts) results from the mean potential drops in the dynamo currents between the 2PM (Figure 6) maximum and the thunderstorm locations. Tropospheric electrification thus appears to be a dynamic phenomenon which is an integral part of the global electrical system.

Transport of electricity invariably results in heating of the medium. This heating is the familiar ohmic process if collisions occur, such as is the case in the ionosphere. In the exosphere the energy of electrified particles will be governed by the electrical structure of the environment. The thermal structure of the upper atmosphere will then partly depend on the flow of these complex currents, and a cursory examination of the magnitudes involved indicates that this source of energy will be important, particularly at the dynamo levels and above, where the heat capacity is relatively low.

5. CONCLUSIONS

The earth's general electric structure is found to have its basic origin in charge separation in the lower ionosphere which is produced by differential transport of charge carriers by low latitude vertical winds of the stratospheric tidal circulations. Vertical electric fields are generated in the E region in low latitudes of each hemisphere which, in conjunction with the earth's permanent magnetic field, power Hall current circuits which flow longitudinally at low latitudes toward the early afternoon sector and have their principal return paths through high latitudes.

These primary dynamo currents produce gross structure in the horizontal electrical potential distribution of the E region that is not symmetrical between the magnetic hemispheres, with a resulting development of interhemispheric currents along magnetic field lines which provide the basic particle fields and power sources forming the radiation belts. Variations in these potentials caused

by irregularities in the tidal circulations and by inhomogeneities in charged particle densities provide charged particle accelerating mechanisms which are a principal feature of the radiation belts. The inner radiation belt current equalizes hemispheric differences between the emf regions, and the outer radiation belt current equalizes differences in potential between auroral zones.

An external current circuit is indicated by these considerations to exist through polar regions of the upper atmosphere which may result in a net negative charge on the earth system, with a corresponding positive charge in the near space of the solar atmosphere. The internal path of this current will be through the outer radiation belt, and control of the current will be exercised by the dynamo circuit potentials of high latitudes and the conductivity of the solar wind.

An electrical circuit is formed between the lower ionosphere and the earth's surface by the relative short circuit of thunderstorms and their lightning discharges which connect the earth's surface to the lower potentials of the midlatitude afternoon and tropical nighttime equatorial dynamo circuits so that a small part (~1%) of the dynamo current flows in the tropospheric circuit.

The electron density of the ionosphere and the exosphere is in large part controlled by the electrical structure of the E region. The charged particles of the exosphere and the nighttime upper ionosphere are generally in transit under the direction of the electric fields of the primary dynamo circuits. It is necessary to this concept that these exospheric currents exhibit a diurnal reversal in direction.

The thermal structure of the upper atmosphere will necessarily be strongly dependent on the electrical structure, with ohmic dissipation of electrical energy providing a principal source of thermal energy in certain regions.

This initial look at the diurnal electrical structure of the earth's atmosphere has clarified some aspects of atmospheric physical processes. It has pointed the way toward investigations that should illuminate some of the points which, for experimental or theoretical reasons, remain as difficulties in the clear understanding of the earth's electrical structure. Progress has come from the application of synoptic principles, emphasizing the fact that in a complex system such as the earth's atmosphere, isolated investigations depend on sheer luck for success.

REFERENCES

1. Beyers, N. J., and B. T. Miers, 1965. Diurnal Temperature Change in the Atmosphere between 30 and 60 kilometers over White Sands Missile Range, J. Atmos. Sci., 22, 262-266.
2. Miers, B. T., 1965. Wind Oscillations between 30 and 60 kilometers over White Sands Missile Range, New Mexico, J. Atmos. Sci., 22, 382-387.
3. Webb, W. L., 1966. Stratospheric Tidal Circulations. Proceedings

of the 1966 Army Science Conference, U. S. Military Academy, West Point,N. Y., Vol. II, 465-479. AD 634 650.

4. Webb, W. L., 1966. Structure of the Stratosphere and Mesosphere, Academic Press Inc., New York, 382 p.
5. Morris, J. E., 1968. Circulation of the Arctic Mesosphere in Summer, ECOM-5181, Atmospheric Sciences Laboratory, U. S. Army Electronics Command, White Sands Missile Range, New Mexico.
6. Webb, W. L., and T. M. Tabanera, 1968. Stratopause Hemispheric Similarities. Planetary and Space Sciences. Vol. 10.
7. Beyers, N. J., and B. T. Miers, 1968. A Tidal Experiment in the Equatorial Stratosphere over Ascension Island, J. Atmos. Sci., 25, 155-159.
8. Fejer, J. A., 1965. Motions of Ionization. Physics of the Earth's Upper Atmosphere, Chapter 7, 157-175; eds., C. O. Hines, I. Paghis, T. R. Harty and J. A. Fejer, Prentice-Hall, Edgewood Cliffs, N. J., 434 p.
9. Chapman, S., 1931. The Absorption and Dissociative or Ionizing Effect of Monochromatic Radiation in an Atmosphere on a Rotating Earth. Proc. Phys. Soc., 43, 26, 483-497.
10. Cole, R. K., Jr., and E. T. Pierce, 1965. Electrification in the Earth's Atmosphere for Altitudes between 0 and 100 kilometers. J. Geophys. Res., 70, 12, 2735-2749.
11. Hanson, W. B., 1965. Structure of the Ionosphere. Satellite Environmental Handbook., ed., F. S. Johnson, Standford University Press, p. 23-49.
12. Chapman, S., and J. Bartels, 1940. Geomagnetism, Clarendon Press, Vols. I and II, Oxford.
13. Van Allen, J. A., 1959. Radiation Observations with Satellite 1959 Epsilon. J. Geophys. Res., 64, 271-286.
14. Heikkila, W. J., and W. I. Axford, 1965. The Outer Ionospheric Regions. Physics of the Earth's Upper Atmosphere, Chapter 5, p. 96-133, eds., C. O. Hines, I. Paghis, T. R. Harty and J. A. Fejer, Prentice-Hall, Edgewood Cliffs, N. J., 434 p.
15. Ulwick, J. C., W. P. Reidy and K. D. Baker, 1967. Direct Measurements of the Ionizing Flux in Different Types of Auroral Forms. Space Research VII, North-Holland Publishing Company, Amsterdam, ed., R. L. Smith-Rose, Vol. I, 656-664.
16. Hinteregger, H. E., L. A. Hall and G. Schmidtke, 1965. Solar XUV Radiation and Neutral Particle Distribution in July 1963 Thermosphere. Space Research V, North-Holland Publishing Company, Amsterdam, eds., D. G. King-Hele, P. Muller and G. Righini, 1175-1190.
17. Reid, G. C., 1965. Solar Cosmic Rays and the Ionosphere. Physics of the Earth's Upper Atmosphere, Prentice-Hall, Edgewood Cliffs, N. J., eds., C. O. Hines, I. Paghis, T. R. Harty and J. A. Fejer, 245-270.
18. Chalmers, J. A., 1967. Atmospheric Electricity, Pergamon Press, London, 515 p.

HURRICANE SURGE DETERMINATIONS ON THE TEXAS COAST AND IN GALVESTON BAY

JERRY W. WOODWARD
U. S. ARMY ENGINEER DISTRICT, GALVESTON
GALVESTON, TEXAS

Prediction of storm surges for hurricanes of various degrees of severity is required for both the design and economic analyses of protective structures proposed for the prevention of hurricane flood damages. This paper presents a summary of work done on hurricane surges by Marinos and Woodward (1967)[1] and Reid and Bodine (1967)[2]. The surge hydrograph is first determined for the open coast and then is propagated through the entrances of Galveston Bay where hydrographs are determined at various locations. Due to the differences in computations of the two methods, they will be described separately.

In the past, the numerical determination of storm surges for the open coast has been primarily confined to a one-dimensional problem. This was accomplished by taking in account only the wind acting in the direction of the coast. The state of the art was advanced by Freeman, Baer, and Jung (1957)[3] when they introduced a method for determining the setup caused by the current parallel to the bottom contours (bathystrophic flow) in addition to the setup normal to the bottom contours.

The phenomenon of storm tide rise along the coast, as described by the "Bathystrophic Storm Tide" theory, is caused by: (1) the direct stress of the wind on the surface of the water (combined wind stress effect) and (2) the additional effect created by the wind stress due to the Coriolis force (Bathystrophic effect). The alongshore wind component or wind blowing parallel to the coast causes a rise in the water level at the coast as a result of the Coriolis effect, which produces a deflection of flow to the right in the northern hemisphere. Hence, as the cyclonic counterclockwise wind system approaches the Texas coast a rise in the mean water level is experienced. As the hurricane eye and region of maximum winds pass the coast, the water level begins to recede back to normal due to the lesser velocity wind acting on the water surface.

In this investigation, it is assumed that: (1) the onshore wind setup responds instantaneously to the onshore wind stress, (2) there is no volume transport toward shore across the bottom contours, (3) the advection of momentum is negligible, (4) alongshore sea surface

height is uniform, and (5) space derivatives of the current speed can be neglected with respect to the Coriolis parameter.

The volume transport differential equations of motion and continuity appropriate for the open coast surge problem are taken as follows:

$$\frac{\partial F_y}{\partial t} = k\ U\ U_y - K\ F_y^2\ D_T^{-7/3} \tag{1}$$

$$g\ D_T \frac{\partial S}{\partial x} = k\ U\ U_x + f\ F_y. \tag{2}$$

$$\frac{\partial S}{\partial t} = 0 \tag{3}$$

where g is the acceleration of gravity, t is time, x and y are the cartesian coordinates in the horizontal plane, F_y is the vertically integrated flux parallel to depth contours, k is the dimensionless wind stress coefficient, K is the seabed resistance coefficient having dimensions of length$^{1/3}$, S is the setup of the water elevation relative to the prescribed datum, U is the actual wind speed, U_x and U_y are the x and y components of the wind speed, D_T is the mean total water depth including S, and f is the Coriolis parameter. The terms $k\ U\ U_x$ and $k\ U\ U_y$ are the kinematic wind stress divided by the density of the water (density assumed constant throughout the spacial and temporal ranges). The Coriolis parameter can also be expressed as

$$f = 2\omega \sin \phi \tag{4}$$

where ω is the angular velocity of the earth or $2\pi/24$ radians per hour, and ϕ is the latitude.

The following relationships can be obtained from equations (1) and (2) for resolving the problem by numerical integration:

$$\Delta S = \Delta S_1 + \Delta S_2 \tag{5}$$

and

$$\Delta F_y = \Delta F' - \Delta F'' \tag{6}$$

where,

$$\Delta F' = k\ U\ U_y\ \Delta t \tag{7}$$

$$\Delta F'' = C_1 \frac{K\ F_y^2}{D_T^{7/3}}\ \Delta t \tag{8}$$

$$\Delta S_1 = C_2 \frac{k\ U\ U_x}{D_T}\ \Delta x \tag{9}$$

$$\Delta S_2 = C_3 \frac{F_y \sin \phi}{D_T}\ \Delta x \tag{10}$$

where Δx and Δt are the prescribed distance along a traverse and increment of time, respectively; ΔS_1 is the computed wind setup perpendicular to the bottom contours; and ΔS_2 is the computed wind setup parallel to the bottom contours. The constants C_1, C_2, and C_3 are merely conversion factors.

The above equations do not give the total surge buildup at the coast since there is an initial rise of the water level along the coast caused by sea, swell, and initial circulation, and also a component of the surge due to the difference in atmospheric pressure from the periphery of the hurricane to its eye. Since the astronomical tide is assumed to be linearly additive to the storm surge, the total surge may be expressed as

$$S_T = S_1 + S_2 + S_{\Delta p} + S_i + S_A \qquad (11)$$

where $S_{\Delta p}$ is the surge due to the difference in atmospheric pressure, S_i is the initial surge, and S_A is the astronomical tide.

The locations selected to determine surge elevations are cities or highly populated areas where the prediction of storm surge heights are of the utmost importance. The surge elevations are computed along imaginary lines referred to as "traverses," which extend from beyond the edge of the continental shelf (depth of 600 feet) to the desired location on the coast to the bottom contours. The computations are initially commenced at the 600-foot depth in the vicinity of the continental shelf because the edge of the shelf represents a point of an unlimited influx of water to replenish the volume being set up.

For reproducing hydrographs of record, or for application to the synthetic hurricanes, the method is considered valid only along traverses located near or to the right of the hurricane's region of maximum winds.

Traverses are divided into incremental reaches (ΔX) and setup computations are performed for each increment. As the traverse nears the shoreline, the increments become smaller since the average depth over the shallow reaches is increasingly more significant in the overall hurricane surge prediction.

The locations along the coast where hurricane surge was computed for calibration purposes are as follows: Port O'Connor, Colorado River, Freeport, Galveston, and Sabine Pass, Texas, and Cameron, Mermentau River, and Eugene Island, Louisiana. These locations are shown on Fig. 1.

The hurricanes used with this study for calibration purposes moved on paths located well to the right of Port Isabel and Mustang Island traverses. Consequently, calibration of the method for these traverses was accomplished by interpolation from values derived for other locations.

In applying the computational scheme along the traverses, the initial step in the calibration involved using the data and physical characteristics of hurricanes of record to reproduce the observed surge hydrographs as nearly as possible. The three hurricanes used for calibration were, "Carla," 9-12 September 1961; "Audrey," 27-28 June 1957; and the "Hurricane of 3 October 1949." These three hurricanes were

selected because of the availability of observed data on surges, wind, and other characteristics.

Calibration, using a given hurricane, involved plotting the traverses on each isovel pattern for each time interval. Beginning at the shoreline, the traverse was divided into incremental reaches (ΔX) measured in nautical miles with points at 0, 1-1/4, 2-1/2, 5, 10, 15, 20, 25, 30, 40, and at every additional 10-nautical mile reach to a point on the traverse just beyond the continental shelf. The wind velocity and direction was read graphically at each point along the traverse for the purpose of computing the wind velocity components. Wind velocities were interpolated for points between the isovels and the wind directions were averaged between the directional arrows shown on the isovel pattern. The depths of water were determined along the traverse at the beginning and end of each incremental reach.

Computations were commenced at the most distant incremental reach from the coastline for each hourly time increment. It can be seen that the surge is computed for each distance increment and is accumulated shoreward to obtain the total surge at the shoreline for the desired hour. A complete hydrograph of the storm was obtained by repeating the computation for each time increment.

To obtain proper calibration, reliable records regarding pressure difference, initial surge, and the astronomical tide were necessary.

During the calibration process, two coefficients had to be solved. The first being the wind stress coefficient "k" and the second being the relative bottom friction coefficient "K". Wind stress coefficients derived by Van Dorn (1953)[4] that were based on a variance of wind speed were used for this investigation. This procedure produced better results than the 3.0×10^{-6} wind stress coefficient that was determined by the Corps of Engineers' Lake Okeechobee studies (1955)[5] and Reid (1957)[6]. Van Dorn recognized that when precipitation was occurring over the water, higher coefficients could be possible. For this reason, Van Dorn's equation was increased by ten percent and better correlation was obtained in reproducing the recorded surge hydrographs.

Once the method for determining the wind stress coefficient was decided upon, a trial-and-error procedure was used to determine the best bottom friction coefficient, "K," for each traverse. Values of 0.01, 0.02, and 0.03 were used at each location. A definite correlation could be seen on the wide shelf region along the west Louisiana and east Texas coast with the higher 0.03 "K" value. As the shelf narrowed down in the Galveston and Freeport area, the 0.02 and 0.01 coefficients produced the closer reproduction of the observed surges, respectively. Since storm and high tide observed data were not available for the narrow shelf region on the lower Texas coast, it was necessary to devise a mathematical expression that could be used to extrapolate bottom friction values for any shelf width.

A multiple linear correlation analysis was made using the data that best reproduced the observed hydrographs. As a result it was found that both the shelf width and slope were needed to determine

the coefficient. The following relationship gave the better coefficients:

$$K = \frac{4.58 \times 10^{-6} \, W^{1.85}}{S^{1.3}} \qquad (13)$$

where W is the shelf width, nautical miles; and S is the shelf slope, minutes.

A check of the method was made using Hurricane Debra (1959) and Hurricane Hilda (1964). "Debra" was a relatively small hurricane that crossed the Texas coast near Galveston, while Hilda was a rather severe hurricane affecting the Louisiana coast near New Orleans. The small size of the wind field of Debra and the effect of the Mississippi River delta may have had on Hilda indicated they were not as good a check as could have been expected.

Figs. 2 through 4 show the results of the calibration for Hurricane Carla at Sabine Pass, Galveston, and Freeport, respectively.

After satisfactory calibration and checking were accomplished, the procedure was applied to synthetic hurricane isovel patterns and characteristics for predicting the surge expected to be produced by a certain size or type storm at a particular location.

The meteorological aspects of hurricanes considered in these computations are central pressure; wind speed, direction, pattern, and dimensions; and forward speed of translation. These synthetic hurricane parameters are given in the "National Hurricane Research Project Report No. 33 (1959)[9]." Synthetic hurricanes are classified as large, mean, or small radius to the region of maximum winds; each with high, moderate, or slow forward speed of translation.

The categories for each set of hurricane parameters are given in Report No. 33. It should be noted that these parameters are characteristic of an extreme hurricane for the Texas coast. The central pressure index (CPI) for each location represents an average return period of 100 years as shown in Report No. 33. The isovel patterns, available from Report No. 33, of the various synthetic hurricanes are based on a CPI pertaining to the upper Texas coast. These patterns may be applied to specific locations having other CPI's by adjusting the winds by the ratio given in Report No. 33.

The traverses used for applying the calibrated equations to the synthetic hurricane parameters were Sabine Pass, Galveston, Freeport, Port O'Connor, Mustang Island, and Port Isabel.

The synthetic hurricane isovels were oriented so as to have the traverse passing through the center of the maximum wind region, which represents the critical path of the hurricane.

The procedure for determining wind component values was greatly simplified over that used for calibration because only one isovel pattern was used for the entire duration of the storm. Each pattern was oriented so that the zero isovel, the traverse, and the coastline all intersected at a common point. This point represents zero time, or a point where the hurricane's wind pattern is assumed to be just beginning. Wind speed and direction were determined at selected points along the traverse for the purpose of computing the

wind velocity components, U_x and U_y. For each of these points the wind velocity components were multiplied by the true wind velocity, U, and a curve was plotted of UU_x and UU_y versus distance from the zero point. The curves, called wind stress curves, are for convenience in determining the wind stress, UU_x and UU_y, at the beginning and end of each incremental reach for each time interval. Typical wind stress curves for a hurricane with large radius, high speed of translation parameter are shown on Fig. 5.

Each time interval point represents the coastline at that particular time. A horizontal scale, graduated in distances, representing each incremental reach from the shore to the continental shelf, is applied to determine the wind stress values at the beginning and end of each incremental reach along the traverse. Fig. 8 represents the relative position of the coastline and shelf in the wind field hour 10, or 10 hours after the winds started acting over the shelf.

A typical curve showing the rise of the water surface due to the pressure reduction is also shown on Fig. 5.

An initial surge value of 2.50 feet was found to be representative for most locations along the Texas coast.

Since the phasing of the astronomical tide and hurricane surge cannot be predicted, a constant average value of astronomical tide must be used. Although a constant average value seems reasonable to use in developing the surge hydrographs, an appropriate amount should be added to the computed surge crest since the peak storm surge and the high tide could occur simultaneously.

Fig. 6 shows the peak surge heights in feet above the mean sea level datum for the 100-year CPI frequency, large and mean radius, each with high and moderate speeds of translation for the location indicated.

Once the synthetic hurricane surge heights were determined for the open coast, a numerical scheme of computation of the two-dimensional, non-linear storm surge was devised to predict surge heights in Galveston Bay, Texas.

As in the open coast surge investigation, it was assumed that the advection of momentum is considered negligible. Because of the small scale and shallow depths of the bay, the effect of the earth's rotation is neglected. The differential equations of motion and continuity appropriate to the bay problem are as follows:

$$\frac{\partial U}{\partial t} + gD\frac{\partial H}{\partial x} = X - f\,Q\,U\,D^{-2} \qquad (14)$$

$$\frac{\partial V}{\partial t} + gD\frac{\partial H}{\partial y} = Y - f\,Q\,V\,D^{-2} \qquad (15)$$

$$\frac{\partial H}{\partial t} + \frac{\partial U}{\partial x} + \frac{\partial V}{\partial y} = R, \qquad (16)$$

where x and y are the horizontal cartesian coordinates; U and V are the vertically integrated x and y components, respectively, of transport per unit width, t is time; g is gravity; H is the water level elevation, mean sea level datum; D is the depth of water at position x,y at time t; Q is the magnitude of the transport per unit width; f is the bed resistance coefficient; R is the rainfall rate; and X and Y are the x and y components of the kinematic wind stress divided by the density of water (the density being assumed constant). Mannings' relation implies that

$$D^{-n} = D^{-7/3} \tag{17}$$

but, in the interest of saving computing time, one can employ the simpler form

$$D^{-n} = D^{-2}$$

provided that the appropriate associated bed resistance coefficient, f, is selected.

The value of Q is obtained from U and V by the relation

$$Q = \sqrt{U^2 + V^2} \tag{18}$$

which is a positive quanity.

The variables H and D are related by the expression

$$D = H - Z \tag{19}$$

where Z is the elevation of the seabed relative to the mean sea level datum. It was presumed that Z was a function of x and y only and that any scour of the bottom was ignored.

The effect of atmospheric pressure was neglected in the above equations because the pressure was considered relatively minor in a small shallow bay. However, it was considered through the specification of the appropriate surge hydrograph in the adjoining sea, and is entered as a boundary condition.

In the present system of computation, four different types of boundary conditions are employed. All of these conditions relate the normal component of flow at the boundary to the state of the water level at the boundary.

To set up the two-dimensional system, the bay was represented in terms of a discrete grid. From knowledge of the topography of the bay, each grid square elevation is regarded as uniform, thus forming a two-dimensional, stair-step type approximation of the actual topo. The boundary condition on the normal component of flow (Q_n) at the juncture of a flooded square and a dry square is taken as

$$Q_n = 0 \tag{20}$$

if the elevation of the water ($\bar{H}$) is less than that of the adjacent dry land. Another boundary condition concerning land and water is if the water level is greater than the dry land. The rate of flooding (Q_n) is given by the relation

$$Q_n = \pm C_o D_b \sqrt{g D_b} \tag{21}$$

where C_o is a non-dimensional overflow coefficient and D_b is the depth of water over the crest of the barrier.

In the case where the water elevation on both sides of a

barrier exceeds the crest elevation of the barrier, the flow may be taken as that of a submerged wier, namely

$$Q_n = \pm C_s D_b \sqrt{g|H_1 - H_2|} \qquad (22)$$

where D_b is again the depth of the water over the barrier, H_1 and H_2 are the water levels on both sides of the barrier, and C_s is the discharge coefficient including the $\sqrt{2}$.

A boundary condition frequently employed for oscillations in a bay open to the sea is to prescribe the value of the water elevation at the mouth. Due to the dampening of the system by radiation of energy to the sea, a more appropriate boundary condition for the case of free oscillations is

$$Q_n = \pm \sqrt{gD}\ |H_g - H| \qquad (23)$$

where $\sqrt{gD}$ is an admittance coefficient (with dimensions of velocity).

Since emphasis will be placed on the evaluation of flow and water elevations in a bay which is connected to the sea, another boundary condition is required to specify the outer limits of the computational grid in the sea. This boundary is shown in Fig. 7 by the bottom of the grid itself. 7

The initial conditions within the system can be somewhat arbitrary since the scheme includes an allowance for frictional dissipation and radiation of energy. Therefore, as in laboratory model experiments, it is reasonable to start the system from a state of equilibrium in which U and V are zero and H is a uniform height. However, a reasonable time period must be allowed for the scheme to reach a state where its response reflects only the effect of the forcing functions.

In a general problem, the values of H are evaluated on a uniform cartesian mesh at spacing ΔS for uniform time steps Δt. The values of H are centered in each grid square at time $n\Delta t$. Values of Z are specified as permanent storage for the same locations as H so that D may be evaluated. Values of U are evaluated at even half-steps of x, odd half-steps of y, and odd half-steps of t, while V is evaluated at odd half-steps of x, even half-steps of y, and odd half-steps of t.

The variables X and Y are supplied at spatial locations consistent with U and V, respectively, but at even half-steps of t. Values of the gulf level, H_g, are supplied for locations and times on the outer boundary of the bay consistent with the locations and times for the H values on that line. Values of the rainfall rate, R, are supplied at positions that are consistent with H but one-half time step out of phase with H. After the values of X and Y are generated from the coarse spatial and temporal array, they are evaluated by linear interpolation for the detailed mesh.

The field of Z, along with information pertinent to the positions, elevations, and discharge coefficients for barriers, is stored as permanent storage.

In order to assure stability of the numerical system, the value of Δt must not exceed a certain limit which is dictated by the choice of spacing, s, and the depth field. The condition may be expressed as

$$\Delta t < \frac{\Delta s}{\sqrt{2\, g\, D_{max}}}$$

where D_{max} is the maximum depth to be encountered in the system.

Computation of U and V is carried out row by row and sequentially from left to right along a given row, starting with the application of the seaward boundary condition for V on the first pass and with application of the appropriate end conditions for each subsequent row.

The adopted grid scheme and indexing system for blocks is shown superimposed over the Galveston Bay configuration in Fig. 12. The x-axis is aligned parallel to the coastline with the y-axis directed inland. The grid interval, Δs, was selected as two nautical miles. In Fig. 10 the shaded area represents the region that is above mean sea level. The cross hatched row and column represents an artifax of extremely high elevation so that the normal component of flow automatically vanishes. For the left boundary, $j=4$ to 19, the flow is simply set equal to zero. The alongshore flow is allowed seaward of the barriers along the coast. The specified forcing function, H_g, is supplied on the outer row of the Gulf of Mexico. The alignment of barriers was selected in such manner that the resulting flows would be similar to that of the prototype.

Wind stress components were supplied at spacing of $5\Delta s$ over the bay at time intervals corresponding to that of the meteorological charts. A linear interpolation program is used to convert this information into a detailed array at spacing Δs.

For the purpose of initial calibration, a spring astronomical tide, which occurred September 2 to 4, 1964, was selected. This tide was selected because of its high surge oscillations, negligible wind speeds, and availability of records. The time increment between successive iterations, Δt, was chosen as three minutes for stable calculations and the gulf water levels, H_g, were linearly interpolated for each iteration. An initial static situation with $U, V = 0$ and $H = H_g$ for all i,j was taken to begin the computations.

The discharge coefficients, C_s and C_o, were selected as 0.2 and 0.4, respectively, with the exception of the Galveston Entrance Channel, which was given a C_s value of 0.6 initially. Due to the geometry of this pass, a high flow coefficient was used to permit a greater degree of flow efficiency. In the final calibration, it was found the above C_o and C_s values gave reasonable results, but the C_s value at the entrance channel had to be altered. This value will be discussed in a later paragraph.

Several computer runs were made in an effort to establish the proper bottom friction factor (f) for the astronomical tide. The friction factor was varied between 0.01 and 0.001. From trial and error, it was found that a value of 0.0025 gave best results.

All final calibration factors for the astronomical tide were employed for the numerical computation of Hurricane Carla (September 9-12, 1961) with the additional parameter of wind stress. The first computations revealed that the water level response in the bay was low; the hydrographs showed that there was a deficiency of water in the bay. This indicated that the flow coefficient for the Galveston Entrance Channel was too low. About this time, the Waterways Experiment Station in Vicksburg, Mississippi, had just completed a series of tests on this pass using a flume model. The results of these tests indicated that the gravity flow coefficient was about 0.7. If this value is used, along with the $\sqrt{2}$ in the equation, the coefficient in the submerged barrier equation, C_s, would be of the order of 1.0. It was also concluded that when the channel was simulated in the numerical model, it had been rotated 45 degrees from its actual alignment, and that no consideration was given to flow attributed to wind acting over the channel. With these changes incorporated in the numerical scheme, the computations were repeated in order to make a comparison of the computed and observed hydrographs. Fig. 8 shows the results of the calibration.

In order to check the effects of these changes, the same astronomical tide hydrograph was recomputed, and the results indicated that even better correlation was obtained.

Using the same coefficients as for Carla and the recomputed astronomical tide, calculations were made to verify the model with Hurricane Cindy (September 16-17, 1963). The overall correlation obtained for this storm was considered sufficient for assuming that the numerical model was adequately calibrated.

Having achieved calibration of the numerical model, a study was then initiated to determine the effects of synthetic storms acting over Galveston Bay. It was concluded that the maximum surge elevations within the bay would be dependent upon the path of the hurricane and translation speed.

Rainfall and rainfall runoff, having been neglected in the calibration because of the relatively negligible amounts that fell, had to be included in this phase. To simulate this phenomena, a prescribed rainfall rate was considered uniformly distributed over all watersheds.

Surge heights (H_g) that were used for input values were obtained from the open coast surge prediction method, as previously described. To commence computations, the values of U,V, and H were set equal to the values found for the spring astronomical tide. This method was used in an effort to simulate prototype conditions by having the water in motion, as opposed to a static condition.

Fig. 9 shows the effects of one plan of improvement being studied, which provides a gate opening for the Galveston Entrance Channel.

The gate has a top elevation of +5.0 mean sea level and a

width of 400 feet. The only water which is permitted to enter the system is the overtopping of the gate during periods of high storm surges. It can be seen that the water elevation is significantly depressed at the gate, but the heights at I=15, J=19 (Baytown) are not reduced in comparison to those of actual conditions. This indicates that the wind stress over the bay plays a major role in producing the water elevations in upper regions of the bay.

REFERENCES

1. Marinos, G. and Woodward, J. W., "A Method for Estimating Hurricane Surge Hydrographs," Journal of the Waterways and Harbors Division, ASCE (to be published).

2. Reid, R. O. and Bodine, B. R., "Numerical Computation of Tide and Storm Surges in Galveston Bay," Journal of the Waterways and Harbors Division, ASCE, Vol. 94, No. WW1, February, 1968.

3. Freeman, J. C., Jr., Baer, L., and Jung, G. H. (1957), "The Bathystrophic Storm Tide," Journal of Marine Research, Vol. 16, No. 1.

4. Van Dorn, W. C. (1953), "Wind Stress on an Artificial Pond," Journal of Marine Research, Vol. 12.

5. U. S. Army Corps of Engineers (1955), "Waves and Wind Tides in Shallow Lakes and Reservoirs," Summary Report, Project CW-169, U. S. Army Engineer District, Jacksonville, Florida.

6. Reid, R. O. (1957), "Modification of the Quadratic Bottom - Stress Law for Turbulent Channel Flow in the Presence of Surface Wind - Stress," U. S. Army Corps of Engineers, Beach Erosion Board, Tech. Memo. No. 93, Feb., 1957.

7. Graham, H. E. and Nunn, D. E., "Meteorological Considerations Pertinent to Standard Project Hurricanes, Atlantic and Gulf Coast of the United States," National Hurricane Research Project Report No. 33, 1959.

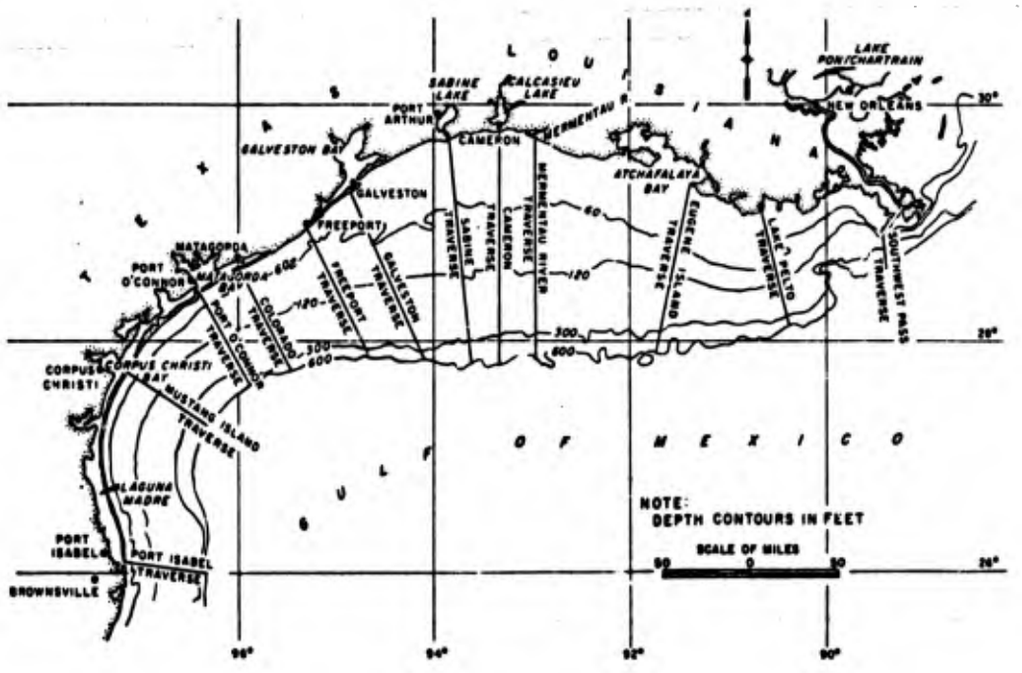


FIG. 1 - TEXAS-LOUISIANA GULF COAST

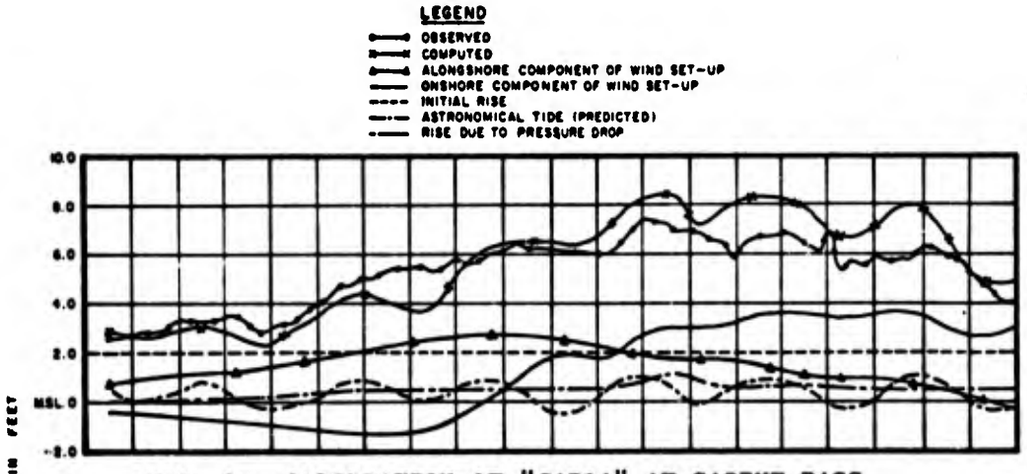


FIG. 2 - CALIBRATION OF "CARLA" AT SABINE PASS

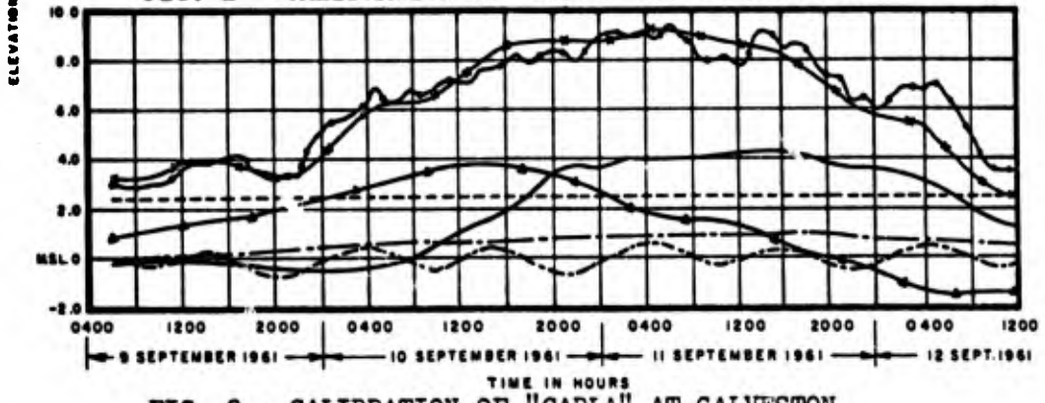


FIG. 3 - CALIBRATION OF "CARLA" AT GALVESTON

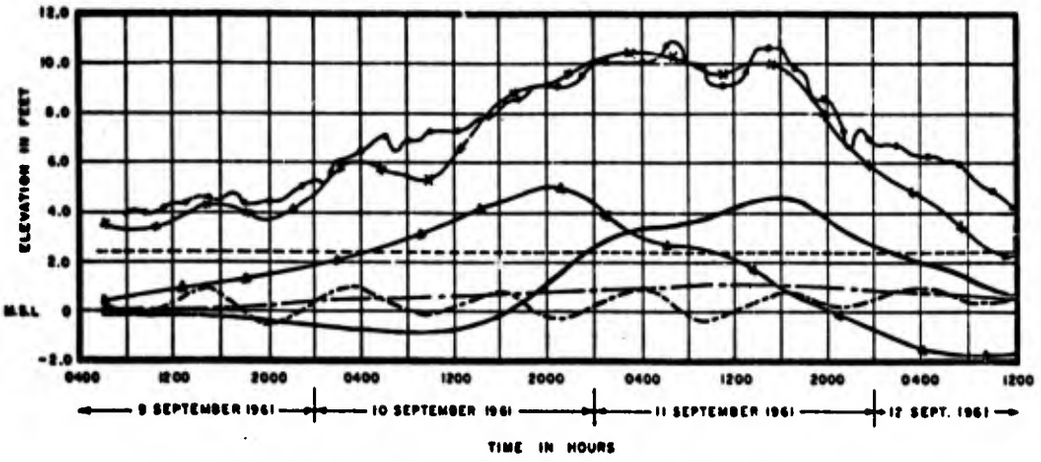


FIG. 4 - CALIBRATION OF "CARLA" AT FREEPORT

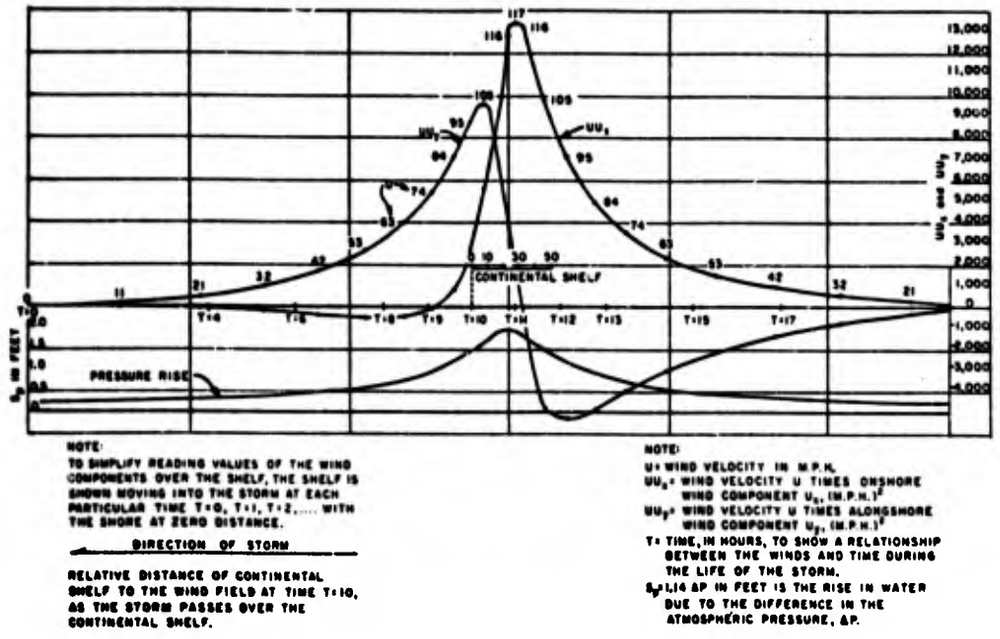


FIG. 5 - WIND STRESS CURVES FOR LARGE RADIUS, HIGH SPEED OF TRANSLATION SYNTHETIC HURRICANE

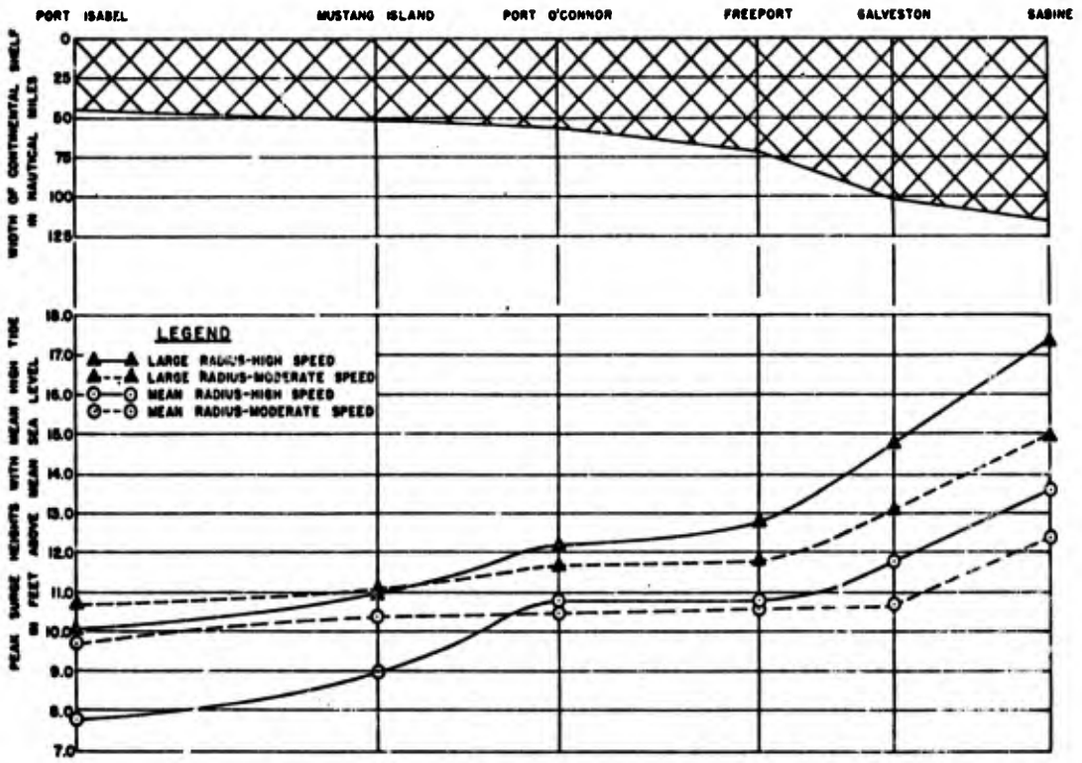


FIG. 6 - PEAK SURGE HEIGHTS AT VARIOUS LOCATIONS ALONG TEXAS COAST

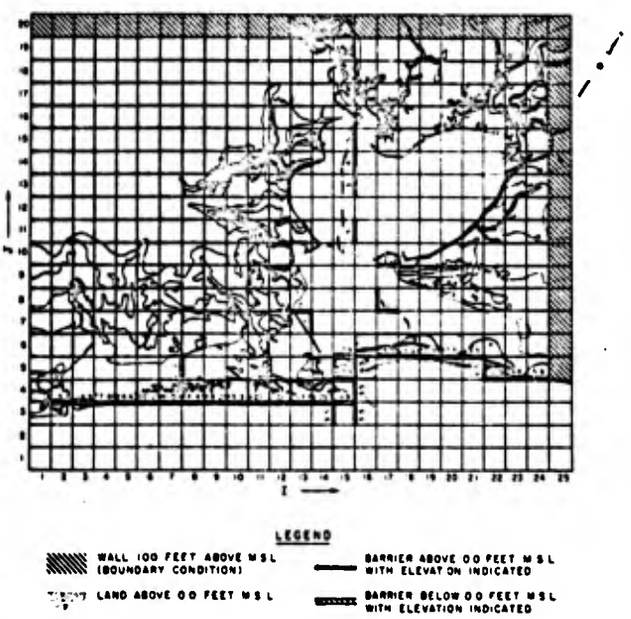


FIG. 7 - GALVESTON BAY GRID SCHEME

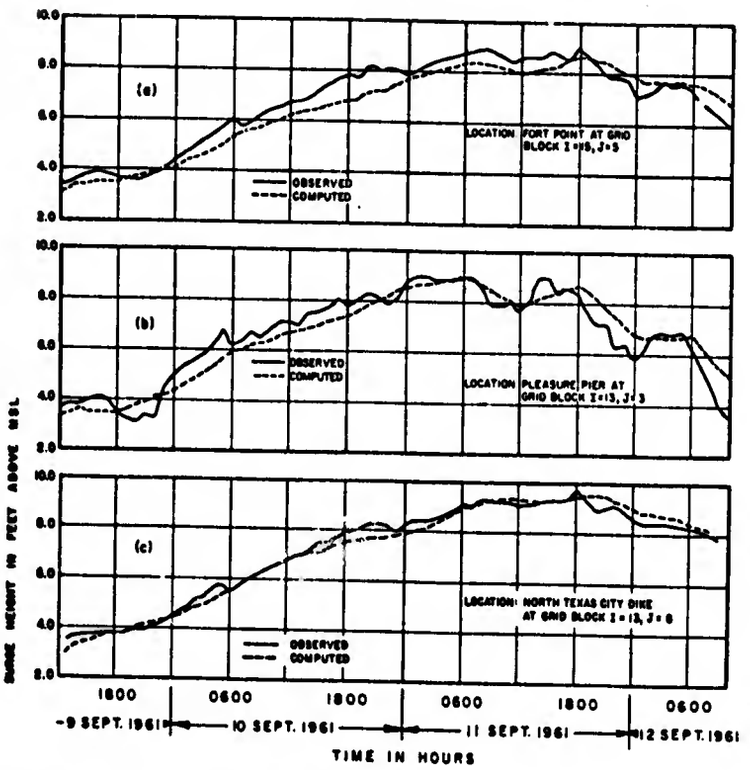


FIG. 8 - CALIBRATION OF "CARLA" AT VARIOUS LOCATIONS

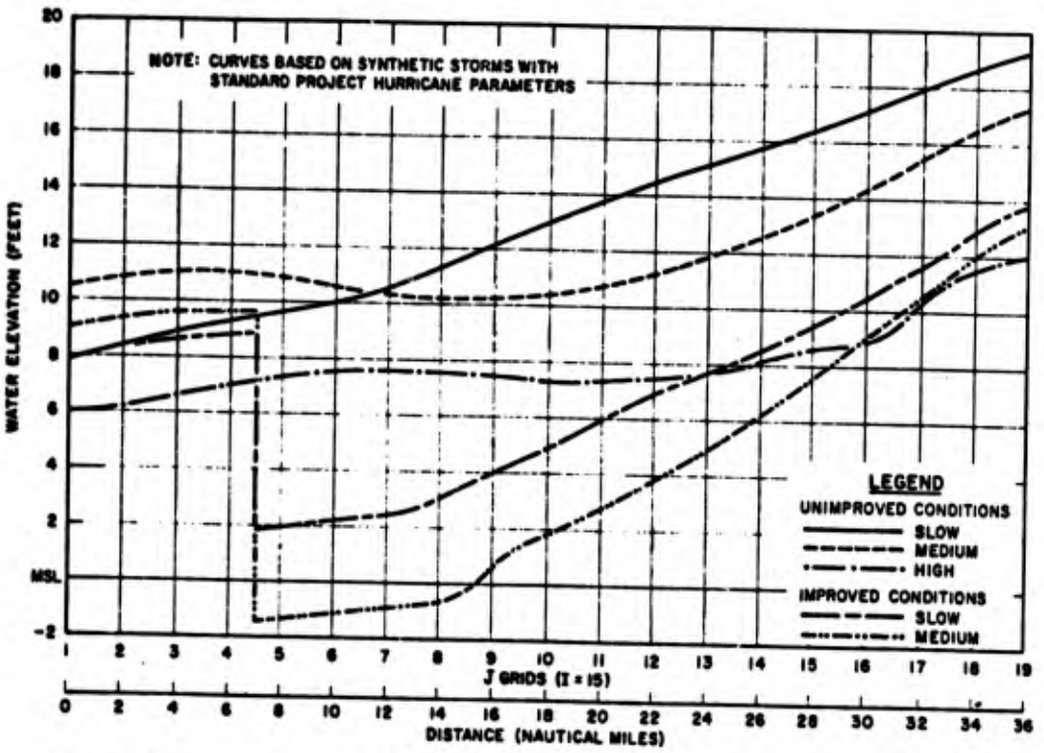


FIG. 9 - SURGE PROFILE FROM THE OPEN COAST TO BAYTOWN

ADVANCES IN PHOTO-ELECTROLYTIC IMAGING SYSTEMS

MICHAEL C. ZERNER, CPT., SIGC; JAMES F. SOBIESKI,* CPT., SIGC
AND HARVEY A. HODES
COMBAT SURVEILLANCE AND TARGET ACQUISITION LABORATORY
FORT MONMOUTH, NEW JERSEY

INTRODUCTION

Since the problems of military air reconnaissance have not been completely solved by conventional silver halide photography, military laboratories have had considerable interest in novel techniques of image recording.[1,2] Photopolymerization, or polymerization initiated by light energy, has been an especially attractive new technique since it promises to provide high resolution images with little or no processing. As presently constituted, however, photopolymerization systems have several defects. A major disadvantage is a lack of sensitivity compared with that found in silver halide photography. Another defect in some photo-initiated systems is degradation of the coated films in storage, either through autopolymerization or through decomposition of the initiating ingredients.

Efforts to improve the sensitivity of the photopolymerization process have led several investigators to begin examining photo-electrically initiated polymerization.[3] In a photo-electrolytic imaging system light is first converted into a proportional amount of current through the use of a photoconductive layer. The resultant flow of current initiates polymer formation in an image recording layer. In practice, a sandwich arrangement is used as shown in Figure 1. The image recording layer containing polymerizable monomers and initiators is sandwiched between two conductive supports, one of which, in this case a Nesa coating, is light transparent. This transparent electrode allows light to strike a photoconductive layer in contact with the monomer coating. The two possible configurations, with the image recording layer either below or above the photoconductive layer, are shown in the figure. In the dark, and with an applied voltage, little current flows through the photoconductor; when given an image-wise exposure, however, an image-wise conductivity pattern forms.[4] The

amount of current which flows is proportional to the intensity of light striking the photoconductor. This flow of current is used to produce free radicals through electrolysis of materials contained in the image recording layer. The amount of polymerization is controlled by the electrolytic production of these free radicals. Each free radical initiates a chain reaction involving thousands of monomer units;[5] the chemical combination of these monomers to form polymer creates the image. Thus, this technique differs from photopolymerization in that the image recording layer is sensitive only to current and can be made totally insensitive to light in the absence of an applied potential.

A photographic system which is to have any degree of light sensitivity must amplify the interaction of radiation with matter. Unlike silver halide photography which obtains this sensitivity through post exposure chemical development,[6] amplification in both photopolymerization and photo-electrolytic polymerization is obtained through the free radical initiated chain reaction just described. However, one of the advantages of the photo-electrolytic procedure is that amplification can also be obtained from the photoconductor. It has been well established that one photon of light striking a photoconductor can lead to the flow of many electrons. Thus one photon can electrolytically produce many free radicals, each one capable of initiating a polymer chain. Indeed, photoconductors with quantum efficiencies (electrons/photon) of tens, hundreds, and even thousands will be described and offer a great opportunity to improve upon the final light sensitivity of polymerization schemes used in image formation.

The problem of developing a successful photo-electrolytic imaging system readily breaks down into two parts; that of finding an effective photoconductive layer that is capable of converting light into a proportional amount of current when a potential is applied, and that of finding a compatible image recording layer sensitive to the passage of current.

THE PHOTOCONDUCTIVE LAYER

A great number of materials show photoconductivity.[4] The photoconductors used in most of our investigations have been zinc oxide, cadmium sulfide and cadmium sulfo-selenide. Emulsions of dye sensitized zinc oxide in dielectric binders, such as silicone resins, have good moisture stability enabling them to withstand moist contacts with the image recording layer; they have been used to produce images of some quality. A definite advantage of zinc oxide photoconductors is that, by incorporation of specific dyes in the emulsion, it is relatively easy to adjust the frequency of light to which they respond. Small amounts of fluorescein or mixtures of bromophenol blue and rose bengal give zinc oxide photoconductors fairly uniform response throughout the visible

spectrum. Although zinc oxide photoconductors can be made that pass little current in the dark, they respond slowly to the presence of light and none have been made that have quantum efficiencies much greater than unity. Although systems based on zinc oxide might be useful for copying and reproduction, it is, at present, difficult to visualize their use in photographic systems of "camera" speed.

Both cadmium sulfide and cadmium sulfo-selenide photoconductors possess high sensitivity. Layers of these materials have been made that pass little current in the dark, react rapidly to visible radiation, and have very high quantum efficiencies. We have focused most of our attention on these materials as they show the greatest promise in creating a sensitive image recording system.

Methods of fabricating cadmium sulfide and cadmium sulfo-selenide photoconductive layers range from spray deposition, to dispersion in various organic dielectric binders, and to sintering. As is to be expected, the technique of preparation greatly influences the photoconductive properties.

Spray deposited cadmium sulfide is made by spraying a fine mist of cadmium chloride and thiourea onto a heated electrically conductive substrate.[7] The major advantage of this method of preparation is the rapid light response shown by the resultant photoconductive layer. Although spray deposited cadmium sulfide has many interesting possibilities, such as fast response, good quantum efficiency, and ease of fabrication, its rapid deterioration in the presence of moisture is a serious drawback. Overcoating the photoconductor with silicone resins has partially solved this problem, although at the expense of quantum efficiency.[8]

Dispersions of cadmium sulfide and cadmium sulfo-selenide in organic binders eliminate the moisture deterioration exhibited by spray deposited layers. This type of photoconductor is made by intimately mixing the cadmium sulfide or cadmium sulfo-selenide with a dielective binder, usually epoxy or silicone resins diluted with a solvent. A thin coating is laid on a conducting substrate such as Nesa glass. Evaporation of the solvent and heat curing of the epoxy resin leaves a uniform dispersion of photoconductor in binder. Epoxy bound cadmium sulfide and cadmium sulfo-selenide layers show little deterioration with moisture and are mechanically quite tough. Response time is somewhat quicker for cadmium sulfo-selenide than for cadmium sulfide, and is more rapid for thin layers, higher applied voltages, and greater light intensities. At 225 volts a five mil thick cadmium sulfo-selenide photoconductive layer responds well to a 1/100 second shutter passing 10,000 foot-candles. The same layer reproduces with equal quality a 1/50 second shutter passing only 100 foot-candles. The spectral sensitivity of this photoconductor in the visible region is illustrated in Figure 2. Cadmium sulfo-selenide has its maximum response more in the red than does cadmium sulfide.

Figure 3 gives the quantum efficiency of this photoconductor as a function of light intensity and the applied voltage. Again, as in the case of spray deposited layers, the quantum efficiency is greater for lower light levels. The light current to dark current ratio at 100 foot candles for this photoconductor is approximately 10^5, far superior to the corresponding ratio for the much thinner spray deposited layers.

The most promising photoconductors we have examined to date have been made by post-treating sintered photoconductors with dielectric resins. The effect on the photocurrent of such post-treated photoconductive layers is given for a typical example in Figure 4. As the figure indicates, the improvement in photoconduction is more dramatic for lower light levels. This post-treatment also renders the sintered photoconductor almost waterproof and makes a normally soft sintered layer tough and abrasive resistant. Figure 5 shows that quantum efficiencies as high as 10^5 are obtainable, representing considerable amplification for the composite imaging system. The light current to dark current ratio at 100 foot-candles illumination is 10^4 for epoxy treated sintered cadmium sulfo-selenide and 10^5 for sintered photoconductors post-treated with silicone resins. The spectral response of these photoconductors is similar to that given in Figure 2. As in the case of the epoxy bound photoconductors, the photo-response is more rapid for cadmium sulfo-selenide than for cadmium sulfide, and depends on the layer thickness, the applied voltage, and the light intensities. A five mil epoxy treated sintered cadmium sulfo-selenide photoconductive layer at 45 volts reproduces a 1/75 second shutter passing 100 foot-candles with accuracy.

Before beginning a discussion of the image recording layer, it should be pointed out that many photoconductors are rectifiers and only behave as successful photoconductors under a specific polarity. Because of this, the choice of which configuration should be used in the image recording sandwich, as shown in Figure 1, is not only a function of the type of polymer initiation used, whether cathodic or anodic, but is also a function of the specific nature of the photoconductor.

THE IMAGE RECORDING LAYER

The major problem in finding a successful photo-electrolytic imaging system is the fabrication of the image recording layer. In the systems described in this paper, image formation is dependent upon a rapid and efficient electrolytic initiation of polymerization. The image recording layer, then, must contain initiating species which react upon electrolysis, monomer or monomers easily polymerized by these initiators, electrolytes which aid in the passage of current, compatible binders or gelling agents, humectants, stabilizers, if required, and other ingredients necessary for the functioning of the particular system of interest.

Polymerization using aqueous solutions of monomers seems to be the most promising to date, although other polar solvents containing electrolytes which enable them to conduct current have been used. Aqueous systems have an important advantage in that they are compatible with gelatin and agar binding materials. With glycerine or other humectants, and with carefully chosen electrolytes, gels of polymerizable systems can be made using gelatin or agar that readily conduct current. A "dry" image recording layer, as pictured in Figure 1, is a desirable goal in any convenient photographic system, a consideration which will permeate much of this work, and a point to which we will repeatedly return.

N,N'-Methylenebisacrylamide is easier than most water soluble monomers to polymerize, and is commercially available. However, its limited water solubility (4gms./100gms.H_2O) restricts its utility somewhat. N,N'-Trimethylenebisacrylamide is much more water soluble (50gms./100gms.H_2O) and is even more easily polymerized. Mixtures of N,N'-trimethylenebisacrylamide and w-acrylimidocaproic acid polymerize extremely easily and more rapidly than either component alone. Most of our exploratory work has been done with mixtures of N,N'-trimethylenebisacrylamide, w-acrylamimdocaproic acid, and the very soluble acrylamide, which yield a tough solid copolymer.[10]

The very heart of the photo-electrolytic polymerization system is the initiation mechanism. Most known electrolytical methods of initiating polymerization are quite slow and inefficient, and could not be successfully utilized. In the following discussion we will focus our attention on three general schemes which we have developed that show promise of rapidly and efficiently initiating polymerization.

The first scheme is quite simple. Mixtures of water soluble monomers with acid and hydrogen peroxide yield flawless polymer at the cathode.[11] The nature of the mechanism believed to be effective in this reaction is given in Equation 1.

1. A $H_2O_2 + H^+ \rightarrow H_3O_2^+$

B $H_3O_2^+ + e \rightarrow H_2O + HO\cdot$

The speed of this reaction and the fact that polymeric images are formed with no post-development are encouraging. Unfortunately, the strong acid concentrations necessary for this system to work efficiently are sufficient to destroy some photoconductors with which the mixture is placed in contact.[8] Although the mixture containing acid and peroxide is quite stable as a liquid for long periods of time, "dry" emulsions coated on metal substrates appear to autopolymerize upon standing several days.

A more promising technique of initiating polymerization electroyltically which shows none of the tendencies of autopolymer-

ization that the acid peroxide system shows, is a system of metal ion initiation. The image recording layer is coated on a metal substrate, which is made the anode. Upon electrolysis, the anode is partially oxidized and metal ions are ejected into the emulsion upon passage of current. Equation 2A shows this for an iron anode.

$$2.\ A \quad Fe \rightarrow Fe^{+3} + 3e$$

$$B \quad Fe^{+3} + R \rightarrow Fe^{+2} + R^{+}$$

$$C \quad Fe^{+2} + H_2O_2 \rightarrow Fe^{+3} + HO^{-} + HO\cdot$$

$$D \quad H_2O_2 + R \xrightarrow{Fe^{+3}} R^{+} + HO^{-} + HO\cdot$$

Many other materials work; this particular example is typical for a metal which dissolves primarily in the higher of two readily available valence states. Certain anions, such as chlorides in the case of a steel anode, or nitrates in the case of a copper anode, aid this process of metal oxidation. Reducing agents, such as hydroxylamine hydrochloride or diphenylcarbazide, present in the solution or emulsion, then reduce the metal ions to their lower valence state, Equation 2B. The developing step then involves adding hydrogen peroxide or washing the exposed emulsion in dilute hydrogen peroxide solution to yield the hydroxyl free radicals capable of initiating polymerization. Equation 2D, obtained by the addition of 2B and 2C, suggests that the overall reaction is one catalyzed by dissolved metal ions. This suggests that one metal ion can be used to form many free radicals, another source of amplification in the final image forming process. Careful studies of the concentrations of reducing agents and metal ions in 10% acrylamide, 1% N,N'-trimethylenebisacrylamide solutions indicate that, at least in the case of iron, an amplification factor of ten or greater can be expected; that is, Reaction 2D occurs ten or more time before the metal ion is so trapped by polymer that it can react no more.

The amplification demonstrated in Equation 2D is somewhat counteracted by a loss of efficiency in 2A, representing the conversion of current to oxidized metal ions. The maximum efficiency that could be expected is 0.33 since three electrons are required for the formation of a single ferric ion. A careful study indicates that potassium and sodium chloride are the most efficient electrolytes for aiding in the dissolution of steel; a 0.1 normal solution giving a yield of about 0.13.

Equation 3 shows that the metal ion initiating system works as well in principle if the anode dissolves in the lower of two available valence states. In this case a silver anode dissolves in its common +1 valence. Persulfate oxidizes the silver ion, as shown in Equation 3B, and forms free radicals capable of initiating vinyl polymerization. Mild reducing agents reduce the Ag^{+2} back to Ag^{+1}. Equation 3D again demonstrates that the silver ions

act much as a catalyst for the direct reaction of persulfate with the reducing agent. In this manner one silver ion is capable of initiating more than one polymer chain.

3. A $Ag \rightarrow Ag^{+} + e$

B $Ag^{+} + S_2O_8^{-2} \longrightarrow Ag^{+2} + SO_4^{-2} + (SO_4^{-})\cdot$

C $Ag^{+2} + R \rightarrow Ag^{+} + R^{+}$

D $S_2O_8^{-2} + R \xrightarrow{Ag^{+}} SO_4^{-2} + R^{+} + (SO_4^{-})\cdot$

E $(SO_4^{-})\cdot + H_2O \rightarrow SO_4^{-2} + H^{+} + HO\cdot$

Equation 3E indicates that, in aqueous media, polymerization may be initiated by hydroxyl free radicals rather than those formed directly from persulfate.

Some careful studies have been made on the metal ion initiating system. On a steel plate in which the monomer and initiating materials are incorporated in a "dry" gelatin coating, polymer can uniformily be deposited over $1cm.^2$ area upon passage of 2×10^{-4} coulombs, or 20 milliamps of current for 1/100 of a second. However, this current sensitivity appears somewhat dependent on the type of binder used. Using the same mixture in liquid experiments, without the gelatin, it is apparent that differential polymerization[12] occurs for solutions containing as little as 10^{-5} molar ferric ion. Electrolysis experiments using steel electrodes demonstrate that the passage of 10^{-5} coulombs is sufficient to generate enough ferric ion to polymerize a disc $1cm.^2$ and 0.1mm thick. (Solution efficiency experiments have indicated that the exact amount of current which converts metallic iron to ferric ions is a function of the potential.) The passage of 1.0 milliamps for 1/100 of a second might thus be considered a limit to the sensitivity of this system for image formation. Comparing these figures with the figures given previously for polymerization on a gelatin plate makes it apparent that a great deal of current sensitivity is lost in going from liquid systems to gelled layers. We are presently extensively investigating different binders to minimize this loss of sensitivity. Gelatin is one of the poorest binders. Since acrylamide and acrylates seem to affect the gelling strength of gelatin, high concentrations of this binder are required. There is also evidence that gelatin copolymerizes, shortening the polymer chains. The use of agar as a binding material seems to hold much promise, eliminating the two difficulties just mentioned. Preliminary experiments show that agar layers have a current sensitivity almost ten times greater than that of gelatin.

A third technique for initiating vinyl polymerization that we would like to discuss utilizes aqueous solutions of zinc salts. The mechanism we believe to be active is that given in Equation 4, which reminds one of the mechanisms often associated with the photo-

conduction of zinc oxide.[14] Zinc ions complex with oxygen dissolved in the media.

4. A $Zn^{+2} + O_2 \rightarrow ZnO_2^{+2}$

B $ZnO_2^{+2} + e \rightarrow ZnO_2^{+}$

$ZnO_2^{+} \rightarrow Zn^{+2} + O_2^{-}$

B^1 $ZnO_2^{+2} \overset{\text{or}}{+} 3e \rightarrow Zn + O_2^{-}$

C $O^{-}_2 + H_3O^{+} \rightarrow H_2O + HO_2\cdot$

D $H_2O + HO_2\cdot \rightarrow H_2O_2 + HO\cdot$

E $Zn + 2H_3O + 2H_2O_2 \rightarrow Zn^{+2} + 4H_2O + 2HO\cdot$

Upon electrolysis this charged complex migrates to the cathode where it is reduced, releasing O_2^{-} according to Reaction 4B or $4B^1$. O_2^{-} then reacts to yield perhydroxyl free radicals, Equation 4C, or these radicals further react to form hydroxyl radicals, Equation 4D. Either perhydroxyl or hydroxyl radicals can initiate polymerization. Equation 4E is added to demonstrate that in slightly acidic media any zinc deposited according to Equation $4B^1$ can be used to reduce the H_2O_2 formed, again resulting in hydroxyl free radicals.

The polymer formed by zinc salt initiation adheres well to the cathode. The system works equally well with steel, aluminum and lead electrodes; there seems to be some advantages to initiation on copper.

An immediate advantage of the zinc salt initiating technique over that of metal ion injection is that polymer is formed immediately, with no post-processing. Since initiation is at the cathode we are able to examine photoconductors that have rectifying properties which make cathodic initiation desirable. Utilizing the zinc salt initiation system agar layers have been made with about the same current sensitivity as that demonstrated in the metal ion injection system; that is, the passage of 20 milliamps for 1/100 of a second will leave a readily observable $1cm.^2$ polymer image in an agar bound layer coated on steel.

Another advantage of the zinc salt initiation scheme is that the polymer formed on the cathodic plate upon electrolysis is formed directly at the interface of this electrode and the image recording layer, resulting in a tight polymer buildup adhering to the plate. This polymer of limited thickness and independent of the thickness of the image recording layer, is capable of yielding high resolution. The metal ion initiating system ejects metal ions into the bulk of the image recording layer, away from the interface. Polymer buildup, then, is throughout the thickness of the layer, resulting in thicker polymer images. The thickness of the

layer in the metal ion initiation scheme thus becomes another facet affecting resolution. We shall return to further considerations of resolution in the discussion.

DISCUSSION

Vinyl polymerization usually involves the chain reaction of between 10^3 and 10^6 monomers, although it is certainly doubtful that we can easily obtain the latter figure. Considering 100 percent conversion of light to electricity, one photon could thus cause between 10^3 and 10^6 events. Since photoconductors with high quantum efficiencies have been devised, additional amplification can be obtained above that provided by the polymerization itself.

As Figure 5 shows, photoconductors suitable for use in photo-electrolytic imaging systems can be made with quantum efficiencies of between 10^2 and 10^4. The coupled system utilizing these photoconductors thus represents an imaging system with a net possible gain of 10^5 - 10^{10}. Silver halide is generally quoted to have an amplification upon development of between 10^6 and 10^9, and there are occasional reports of greater amplification. On this basis, imaging systems based on photo-electrolytic polymerization might compete. But whether one electron can lead to the formation of a chemical species that will initiate polymerization is a question not yet answered. Such a sensitivity would be the theoretical limit of the acid-peroxide system, and, under favorable conditions, the zinc salt initiating systems. Although metal anode solution is a process with an efficiency less than one, the fact that one metal ion could lead to the formation of many free radicals, each capable of initiating polymerization, gives the metal ion injection scheme the possibility of efficiencies greater than one.

Before ending any comparison of sensitivity between silver halide and photo-electrolytic polymerization it is essential to point out that the image amplification in the former case is immediately and effectively made apparent upon development. The silver particles that constitute the image have considerable optical density. No polymer of similar effect has yet been found. Intense dyes, as nigrosine black, can be included in the image recording layer, but this necessitates a post-exposure wash away of unpolymerized material.

The resolution of a photographic system is, of course, another important consideration. Under the most favorable conditions the size of the polymer molecule, itself, limits this resolution. The largest polymer molecule conceivable in these systems, a "cubic" molecule 10^2 monomer units on a side, is easily subtended by any visible wavelength and is thus no limiting factor on resolution. The accuracy of the photoconductor in reproducing with current the exact location of the focused image is a function of the photoconductive material, its grain size, the thickness of the photoconductive layer, the methods of fabrication and the applied electric field.

Sintered photoconductors, which show the greatest quantum efficiency, have relatively poor resolution capabilities because of their porous structure. Experiments adjusting the composition of the sintering mixture, the heating cycle and the maximum sintering temperature indicate that better results are obtainable. Recent disclosures indicate that inclusion of cadmium carbonate improves image quality considerably.[15] Zinc oxide photoconductors bound in silicone resins have been used which give greater resolution, but these photoconductors are limited by their low light sensitivity.

Many problems remain to be solved before photo-electrolytic imaging systems realize their full potential. Resolution is one such problem, as our best sintered photoconductors have resolved, to date, only 4 line pairs per millimeter. The attainment of a "camera speed" light sensitivity is another desirable feature of this new system. Pictures of resolution patterns have been taken with an illumination of 1.3 foot-candles in 0.1 second, which is roughly equivalent to an A.S.A. of 6. We feel that photo-electrolytic polymerization shows great promise of achieving more "speed". Picture contrast is another area of investigation which we have only begun. Contrast depends in a complex manner on nearly all the parameters of the system including the nature of the photoconductor, the initiation scheme, the type of polymer, the type of binder used in the image recording layer and the type of development used to make the image apparent.

REFERENCES

1. For a good review of "New Approaches in Photography", see, for example, J. J. Robillard, PSE, 8, 18 (1964)

2. A. Sheep, S. Chaberek, I. Kafsky, et.al. (Technical Operations, Inc.) "Evaluation of Unconventional Photographic Processes", Final Report 23 February 1960; Air Force Contract AF 33 (616)-5717

3. E. Cerwonka, A. S. Deutsch, and S. Levinos, (General Anailine and Film Corporation) "Investigation of Polymerization Processes, Final Report, 13 March 1966; US Army Electronics Command, Contract DA-36-039-AMC-00119(E)

4. See for example, R. M. Schaffert, Electro-photography, The Focal Press, London and New York (1965)

5. For a discussion of polymerization see, for example, J. C. Bevington, Radical Polymerization, Academic Press, New York (1961)

6. C. E. K. Mees and T. H. James, The Theory of the Photographic Process, 3rd Edition, The Macmillan Company, New York (1966)

7. R. R. Chamberlin and J. S. Skarman, J. Electrochem. Soc. 113, 86 (1966)

8. The photoconductor can be isolated from the polymerization system by an intermediate layer which conducts current greater in a direction perpendicular to its surface than parallel, Figure 1. An example of such an intermediate might be Corning's "microlead" glass, a glass containing small wires imbedded to carry current only perpendicular to its major surfaces.

9. M. Berkenblit, G. Cheroff, F. Hochberg and A. Reisman, U.S. Patent 3,333,985 (1967)

10. Many other considerations effect the choice of monomers used. A more thorough discussion of these factors is in preparation.

11. Other per-compounds, such as persulfates, perborates and water soluble organic peroxides and hydroperoxides work equally as well.

12. Depending on the reducing agent, there may be a slow reaction between the reducing agent and the per-compounds used, releasing free radicals and initiating polymerization even in the absence of metal ions. For a given concentration of per-compound the rate of this direct oxidation-reduction reaction is directly proportional to the concentration of the reducing agent. If it takes X time for this direct reaction to initiate polymerization,

then we say "differential polymerization" occurs with sufficient metal ion concentration to initiate polymerization in less than X/4.

13. In evaluating the sensitivity of layers using gelatin as a binder it is necessary to exercise some care, if only for the sake of completeness, in discriminating between polymerization and gelatin tanning. The injection of some metal ions, as, for example, those of Cr from steel, can be very effective in rendering gelatin insoluble.

14. A. Bernas, J. Phys, Chem., 68, 2047 (1964)

15. Fuji Photo Film, Co. Ltd., Brit. Patent 1,064,140 (1967)

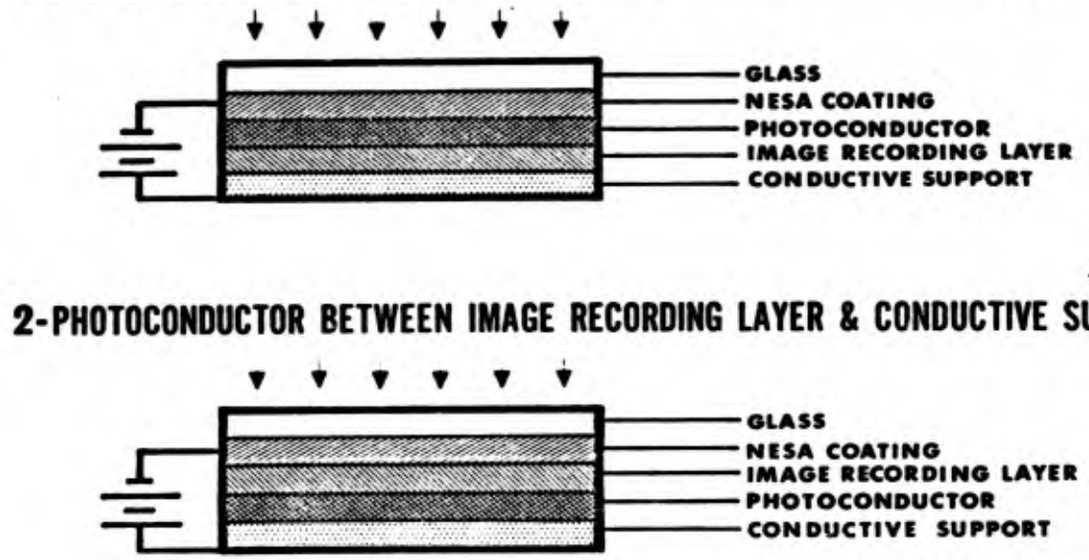

1-PHOTOCONDUCTOR BETWEEN LIGHT SOURCE & IMAGE RECORDING LAYER
GLASS
NESA COATING
PHOTOCONDUCTOR
IMAGE RECORDING LAYER
CONDUCTIVE SUPPORT
2-PHOTOCONDUCTOR BETWEEN IMAGE RECORDING LAYER & CONDUCTIVE SUPPORT
GLASS
NESA COATING
IMAGE RECORDING LAYER
PHOTOCONDUCTOR
CONDUCTIVE SUPPORT

Figure 1.

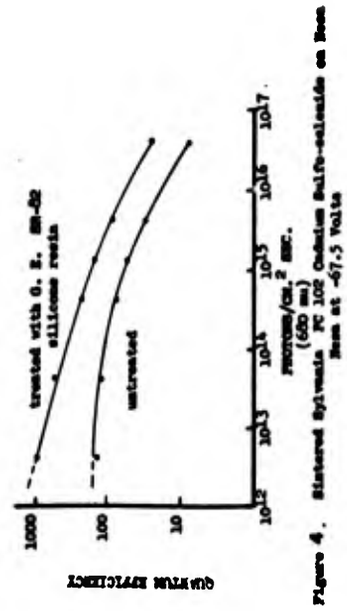


Figure 4. Sintered Sylvania PC 102 Cadmium Sulfo-selenide on Nesa
Nesa at -67.5 Volts

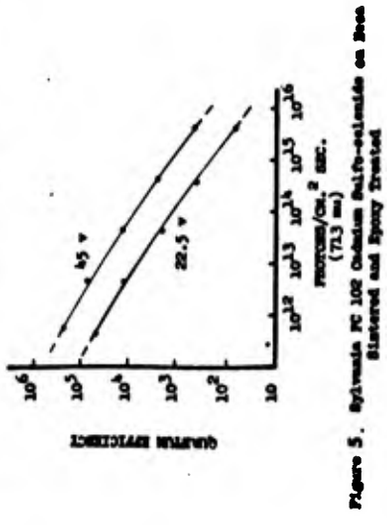


Figure 5. Sylvania PC 102 Cadmium Sulfo-selenide on Nesa
Sintered and Epoxy Treated

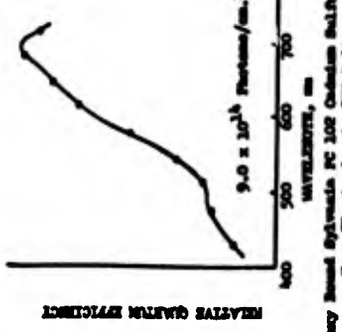


Figure 2. Epoxy Bound Sylvania PC 102 Cadmium Sulfo-selenide on Nesa
Nesa Electrode at -225 Volts

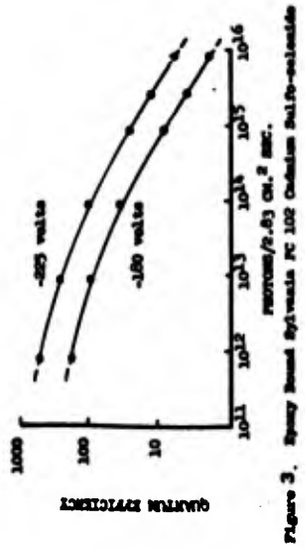


Figure 3. Epoxy Bound Sylvania PC 102 Cadmium Sulfo-selenide on Nesa