Proceedings

13 TH ANNUAL AIR FORCE SCIENCE & ENGINEERING SYMPOSIUM

Vol. II

Arnold Engineering Development Center
Arnold AFS, Tennessee

University of Tennessee Space Institute
Tullahoma, Tennessee

Sponsored by
Air Force Systems Command
Office of Aerospace Research

27-29 September 1966

PROCEEDINGS

of the

13th ANNUAL AIR FORCE SCIENCE & ENGINEERING SYMPOSIUM

27-28-29 September 1966

ARNOLD ENGINEERING DEVELOPMENT CENTER
ARNOLD AIR FORCE STATION, TENNESSEE

VOLUME II

UNCLASSIFIED, UNCONTROLLED DISTRIBUTION VOLUME

Sponsored by

AIR FORCE SYSTEMS COMMAND

and

OFFICE OF AEROSPACE RESEARCH

BLANK PAGE

TABLE OF CONTENTS

VOLUME I

UNCLASSIFIED, UNCONTROLLED DISTRIBUTION

VOLUME II

UNCLASSIFIED, UNCONTROLLED DISTRIBUTION

Contents

(U) CONFRONTING THE HEAT PROTECTION PROBLEM IN HYPERSONIC NOZZLES USING HIGH-ENTHALPY AIR

By

Gordon M. Gray, Lt Colonel, USAF

Arnold Engineering Development Center
Arnold Air Force Station,
Tennessee

Lt Col Gordon M. Gray

BIOGRAPHY

Gordon M. Gray, LtCol USAF

Prior to recent reassignment to the Air War College as a student, LtCol Gray was chief of the TRIPLTEE Program Office at Arnold Engineering Development Center. The TRIPLTEE large hypersonic wind tunnel is now under design and will provide a Mach 3-7 true flight environment with a 10 ft diameter usable test section.

LtCol Gray has taught in many fields of engineering science, serving on the Air Force Academy faculty for five years and as exchange professor at the Philippine Military Academy for two years. He has also been a project officer at the Air Force Special Weapons Center, and has participated in the development of a jet fuel pumping system which serves all of Formosa.

LtCol Gray received his BS degree at the U. S. Naval Academy in 1950. He was granted an MS at the University of Illinois in 1954, and a PhD in Structures and Applied Mechanics at the same institution in 1958. He is a registered engineer in Tennessee, and a member of NSPE, ASEE, ASCE and Sigma Xi.

TABLE OF CONTENTS

ABSTRACT

The process characteristics involved in producing hypersonic flow with high enthalpy air are examined to identify the major factors which bear upon the design of the nozzle. High pressure, high heat flux and high oxidation potential of the gas stream emerge as the critical elements for consideration. The combined effect of these three characteristics is shown to restrict the usable cooling techniques when provision is made for adequate gas stream containment.

Current studies and experimentation with film cooling are described and a computer program is utilized to develop coolant characteristic curves which establish mass-flow ratios and orifice geometry based upon nozzle dimensions and gas stagnation conditions.

Self-contained throat protection methods are described and analyzed on a basis of both material properties and system behavior. The constant profile ablators are identified for their significant potential in throat cooling. For the matrix, foamed zirconia and sintered tungsten powder are finding use while the phenolics provide the desired volatilizing properties of the sacrificial phase.

Among the uncooled refractories, a hypereutectic hafnium carbide-carbon system offers a stable liner material in which the carbon platelets act as stress sinks for the high transient thermal compressive stresses. Pre-heated, yttria-stabilized zirconia offers another solution for overcoming hot-face thermal shock. Composite throat section designs are shown which optimize the best thermal properties of each material.

In conclusion, an economic comparison is made among available and future throat protection techniques and Mach 13 is shown to be the approximate limit for true-flight-environment testing in large blow-down facilities.

BLANK PAGE

LIST OF FIGURES

NOMENCLATURE

a - sonic gas stream velocity (fps)

f - pressure stress (psi)

g - gravitational constant

k - c_p/c_v

ℓ - axial flow dimension (in.)

p - internal pressure (psia)

r - radius (in.)

t - orifice gap width (in.)

v - gas stream velocity (fps)

C - general coefficient

D - nozzle diameter (in.)

E - modulus of elasticity (psi)

K - configuration sensitivity factor

M - Mach number

N - temperature proportionality

R - gas constant

T - absolute temperature (^{o}R)

α - coefficient of linear thermal expansion

μ - Poisson's ratio

σ - temperature stress (psi)

$\dot{w}$ - unit mass flow (lb/sec)

SUBSCRIPTS

a - inner radius

b - outer radius

c - coolant

ci - coolant injection station

e - atmospheric

m - main stream

o - stagnation

* - throat

θ - tangential direction

BLANK PAGE

1. INTRODUCTION:

The trend toward higher flight-vehicle speeds in the hypersonic regime has continued unabated since man first accomplished controlled supersonic flight in the mid-40s. This trend must continue to be matched in ground test facilities that we are planning if those facilities are to be useful in simulating the actual flight environment. Facility size is also important. The larger engines of the future will demand higher mass flows and longer test durations, which will consequently result in higher heat fluxes in the tunnel components. In Figure 1, the energy equation (modified for stagnation conditions) has been plotted to show the true-flight temperature--pressure relationships just mentioned. The diminishing returns in Mach number from boosts in temperature and pressure are evident from this figure. For example, at 60,000 ft the pressure must be raised five orders of magnitude to increase the Mach number by one order, and it takes twice the temperature rise to increase the Mach number from 10 to 12 as it does from 2 to 4.

Recent accomplishments in high-pressure electric arc technology now permit the continuous generation of high-enthalpy air (to 15,000°F) to meet these test requirements. However, it is proving more difficult to contain and control these hot gases than to produce them, particularly where large mass flows are involved. The major design limitations are clearly the result of high-temperature material properties. For example, in an air environment, the oxidation rate of most materials increases with both temperature and pressure. Figure 2 shows these effects with respect to aluminum and zinc. Oxide build-up directly effects heat transfer and strength properties and can be additionally detrimental in wind tunnels if particles are carried down stream to impinge on the test article. Another generally adverse characteristic of materials at high temperature is viscoelastic behavior and the associated decrease in strength. Tungsten is the best of the commercially available refractory metals in this respect but, as depicted in Figure 3, it too loses strength and displays viscoelastic characteristics when the temperature gets sufficiently high.

In hypersonic nozzles, most of the high-temperature shortcomings of materials become evident at or near the minimum cross-section or throat area. In that region, the gas density and heat transfer characteristics are maximized, resulting in the severest thermal problems. Notwithstanding these problems, much successful hypersonic testing has been done in facilities that can accommodate small-scale test

articles; where the economics of fabrication are not too restrictive and the possibility of catastrophic failure from high pressures and temperatures is remote. However, to scale up such test facilities, even one order of magnitude requires the designer to face materials problems and safety considerations which are steeply disproportionate. It is primarily this view that is addressed in the following discussion -- nozzle throat heat protection in large blow-down tunnel facilities. This paper discusses recent work and planned research at Arnold Center on different approaches to protect the throat from deterioration, with particular attention being given to refinements of gas film cooling methods and recent developments in the use of refractory materials.

2. FILM COOLING:

Film cooling has been selected for heat protection in the 10-inch throat of the Mach 7 nozzle for the TRIPLTEE hypersonic facility currently under design at AEDC. This specification resulted largely from the results of sponsored research by H. C. Roland, et al (1) in which a comprehensive theoretical analysis of nozzle throat heat transfer was developed and programmed for computer solution. This analysis specifically treats the region between film coolant injection and nozzle throat, and has provisions for transpiration cooling, back-side cooling and multiple tangential coolant injection stations. A finite number of gas-stream layers are considered in both the main stream and boundary. Consecutive, axial increments made up of these layers are then balanced with respect to momentum and heat by an iterative process. Outputs include gas-stream temperature and velocity profiles and wall temperatures between the point of coolant injection and throat, for the gas properties and coolant mass ratio specified. Experimental verification of this theoretical work is continuing in a cooperative effort between NASA-Ames and Arnold Center and also under contract with the University of Tennessee. Certain parametric calculations of Roland's program have also been confirmed experimentally by Lewis and Horn (2).

In order to make the versatility of this program more readily available to the designer for specific coolant system proportioning, the writer has studied the interrelationship of coolant injection geometry, gas properties, main stream flow characteristics, and throat wall temperature, using Roland's program in modified form. It was first assumed that coolant injection would be accomplished under theoretically optimum conditions--sonic, tangential flow at the orifice, with the coolant and main-stream velocities and

pressures matched to provide a condition of minimum boundary layer disturbance. Also, the axial pressure profile, from stagnation reservoir to throat, was taken to vary linearly, and the sonic condition was assumed to exist at the throat. Figure 4 shows an idealized axisymmetric nozzle profile with film-cooling orifice for reference. For these conditions, with isentropic flow, it may be shown that the location of the injection orifice is defined, for given stagnation conditions and throat diameter, as follows:

From the energy equation at stagnation

$$\frac{a_{ci}^2}{k-1} + \frac{v_{ci}^2}{2} = \frac{a_o^2}{k-1} \qquad (a)$$

but $v_{ci} = a_c$ and $a = \sqrt{kgRT}$

substituting these relationships into (a), we obtain

$$a_{ci}^2 = kgR \left[T_o - T_c (k-1)\right]$$

by definition

$$M_{ci} = \frac{v_{ci}}{a_{ci}} = \sqrt{\frac{T_c}{T_o - T_c(k-1)}} \qquad (b)$$

now, from the state and energy equations

$$\frac{D_*^2}{D_{ci}^2} = M_{ci} \left[\frac{\frac{k+1}{2}}{1 + \frac{k-1}{2} M_{ci}^2}\right]^{\frac{k+1}{2(k-1)}} \qquad (c)$$

If we substitute (b) into (c) and rearrange, an expression is obtained in which

$$D_{ci} = CD_* \qquad (d)$$

where C is a function only of k, T_c and T_o.

Similarly, from the energy equation,

$$\frac{D_*^2}{D_{ci}^2} = \left(\frac{p_{ci}}{p_o}\right)^{1/k} \sqrt{\left[\frac{\left(\frac{k+1}{2}\right)^{\frac{k+1}{k-1}}}{\frac{k-1}{2}}\right]\left[1 - \left(\frac{p_{ci}}{p_o}\right)^{\frac{k-1}{k}}\right]} \qquad (e)$$

Recalling now that the pressure distribution between the stagnation reservoir and throat is assumed to vary linearly, and remembering that for sonic conditions

$$p_* = 0.528\ p_o,$$

we can develop relationship

$$\frac{p_{ci}}{p_o} = 0.472\ \frac{\ell_{ci}}{\ell_o} + 0.528 \qquad (f)$$

Substituting (f) into (e), an expression is obtained relating D_{ci} to ℓ_{ci}, the axial location of the coolant injection orifice with respect to the throat. However, D_{ci} is already available from (d), hence, with only the initial throat diameter, the dimension between reservoir and throat and the stagnation temperature, the orifice position can be established. Since the relationship given in (b) is relatively insensitive to variation in T_c, so long as T_c is small in comparison to T_o, we can reasonably take T_c equal to $0.1(T_o)$, and utilize a value of $k = 1.4$, for air to find that

$$D_{ci} = 1.45\ D_*$$

and

$$\ell_{ci} = 0.88\ \ell_o \qquad (g)$$

Equation (g) indicates the desirability of having the injection orifice close to the stagnation reservoir for the purpose of matching coolant and main stream flow velocities. In reality, however, the optimum location would be somewhat closer to the throat where actual boundary layer build-up (including velocity mismatch turbulence) would be minimized by the suppressing effect of the injected coolant. Hence, the longitudinal location of the orifice ring contains a judgment factor on the part of the engineer.

Still remaining to be determined is the orifice slot width and coolant mass flow. These quantities are related to the wall temperature at the throat through the physical properties of the gases, and have been calculated with the aid of Roland's computer program in modified form. The resultant data is displayed in Figures 5 and 6 in such a manner that the designer may select a mass-flow ratio and orifice slot width to provide different predetermined throat wall temperatures. Since the heat-transfer program calculates minimum coolant requirements, Figures 5 and 6 are not conservative, and their use should be accompanied by adequate consideration of mass flow and temperature fluctuations during operation. The curves are shown for air as main-stream and coolant gas, but the concept and computer program are easily adaptable to other flow media.

3. REFRACTORY RESEARCH:

The use of film, transpiration or back-side cooling for nozzle throat protection is inevitably accompanied by problems of the complexity and expense of a coolant supply system. Consequently, alternative designs which reduce or eliminate the need for a separate coolant supply system and provide mechanical simplicity in operation are always attractive to the engineer. The coming generation of large blow-down test facilities such as Arnold Center's TRIPLTEE (Mach 3-7) and LORHO (Mach 8 and higher) have stimulated research in this direction, much of which is currently underway and unreported. However, two philosophies of design which have innate simplicity will be discussed:

a. The Constant-Profile Ablators. The two-material ablator system is essentially a self-contained, one-shot transpiration cooling system which can be recharged for subsequent runs if the volatile-phase material is judiciously chosen. The concept relies for throat protection upon a sacrificial loss of mass by one phase of throat material while the other phase maintains the aerodynamic contour of the nozzle. In wind tunnel design, the divergent nozzle geometry must be closely controlled to maintain a high quality of flow at the test section. Hence, it is important to insure that the throat section does not erode under the effect of the high-energy gas stream. One critical dimension is the throat diameter. To show the sensitivity of test-section Mach number to changes in throat diameter, Figure 7 sets forth a plot of the differential equation

$$\frac{dM}{dD_*} = K$$

where the sensitivity factor

$$K = \frac{4M\left(1 + \frac{k-1}{2} M^2\right) D_*^{\frac{3(k-1)}{k+1}}}{(k+1)\left[(D^2M)^{\frac{2(k-1)}{k+1}} - M^2\right]}$$

D and M refer to the nozzle exit. It may be noted that K increases rapidly with increasing Mach number and decreasing throat diameter, which emphasizes the requirement for dimensional integrity at the throat throughout the test run. The skeletal matrix is characteristically a dense refractory metal or oxide of carefully controlled porosity and permeability. Sintered tungsten powder, Figure 8, (3), and foamed, yttria-stabilized zirconia, Figure 9, (4), are finding successful application.

Because of the flexibility available in the technology of grading and controlling the porosity of the matrix, it is possible to tailor the gasification rates to the protection requirements at different stations in the nozzle throat. This has resulted in the fabrication of complex built-up throat structures to satisfy severe heat transfer conditions. In high-heat-flux regions, the permeability should be high to permit good coolant flow rates. In the high-void-ratio "reservoir" areas, the large porosities are induced by incorporating in the green bodies, a volatile pore former such as ammonium carbonate which gasifies and escapes with presintering or sintering heats. Dimensional instability of the porous matrix can be a problem both during sintering and during nozzle operation. However, with accurate molding and machining of the green compacts and attention to the control of pre-sintering compaction pressures and temperatures and sintering sensitivities of the material, the regions of desired porosity may be specified with a precision equivalent to other design accuracies.

Interstitial voids in this skeletal structure are filled with the organic material which must volatilize at high temperatures of the throat region and display protective properties consistent with the thermal environment it is to experience. Two major processes come into play as this sacrificial phase receives a thermal load. First, it passes from solid to gaseous state and the latent heats of fusion and vaporization absorb thermal energy. The gas then continues to provide a protective film over the nozzle surface downstream of the throat. The effectiveness of this protection varies directly with the Prandtl number of the gas film.

A prime consideration in selecting an ablatant is the potential interaction between gas products and the high-energy air passing through the nozzle. Boundary layer combustion for example, can distort profiles of velocity, pressure and temperature in the test section. On the other hand, it is equally undesirable for the gasified material to remain inert but to recondense downstream with resultant damage to model instrumentation or aerodynamic profiles. Still another consideration pertains to possible interaction between ablatant and refractory matrix. With carbon present from the pyrolizing ablator, a tungsten throat matrix can be easily lost by the formation of an eutectic in the tungsten-carbon system.

The phenolic nylons and laminating resins have many physical characteristics that lend themselves to satisfactory use in these "infiltrated" ablators (5). They volatilize in a clean, non-condensing fashion, appear to remain in the boundary layer because of their low densities, have a reasonable Prandtl number and do not create undesirable ion concentrations at hypersonic testing temperatures. The following extract from the test program report on phenolic-impregnated porous oxides (4) is quoted as indicative of the behavior of these materials:

"Performance of the phenolic impregnated specimens was excellent. The front face of the specimen increased in temperature to about 4200°F in a few seconds, with the phenolic ablating and charring progressively without damage to the foam. No foam material was lost from the specimen during test.

"During the first phase of testing, the phenolic near the surface rapidly chars and ablates, cooling the specimen during the first few seconds of test. The char layer formed progresses backward, allowing a gradual thermal gradient and preventing build-up of dangerous thermal stress. The areas retaining virgin and slightly decomposed phenolic appear to be stronger than the unimpregnated foams, and the char also appears to be contributing to the strength of the system, while increasing thermal conductivity in the areas where it is completely carbonized. The foam structure limits available oxygen to the char, so recession of the char is very slow. These interactions of mechanisms provide the excellent performance observed."

Figure 10 shows an example of a compound, refractory metal powder throat insert. Low permeability upstream and down from the throat permits only enough gasification to preclude burn-out. High permeability at the throat itself, and medium

porosity in the primary substrate allow passage of sufficient coolant to protect the high-heat-flux area. The rate of heat transfer to the gas-liquid interface in the infiltrant is also controlled by the primary substrate. Backing up the surface layers is the high porosity reservoir in which is stored a mass of organic volitant sufficient to provide throat protection for one operational cycle. The inserts are designed for rapid removal so that different combinations of porous surfaces or different infiltrants can be used for varying test conditions. After use, the insert must be leached, pressure cleaned of oxidized residue and inspected for micro-cracks prior to reimpregnation. Among the oxides, foamed zirconia can now be fabricated so inexpensively that single-run inserts are an economic possibility though they are not yet capable of providing the design sophistication of graded, porous powder metallurgy.

b. The Uncooled Refractories. The attractiveness for wind tunnel application of the simple, completely uncooled nozzle throat has been enhanced by recent technological advances in materials preparation and research. Furthermore, this concept offers increased aerodynamic efficiency by eliminating heat transfer to a wall coolant. Recent work by J. L. Jacocks (6) has shown that boundary layer growth can be suppressed by running hot walls in the throat, which indicates that better test cores are obtainable with uncooled inserts.

Much of the knowledge gained in rocket nozzle research on refractory behavior is also applicable to wind tunnel throat protection. Nevertheless, the particular requirements that the wind tunnel throat material endure many temperature excursions and a severely oxidizing hot air atmosphere; and the essentiality that nozzle contour integrity be maintained, are differences which present acute problems to the hypersonic nozzle designer. The destructive deformations associated with thermal shock are indicative of these design problems. Many refractory oxides and composites which show good initial thermal shock resistance, deteriorate rapidly after a few thermal cycles because of loss of essential constituents through diffusion or chemical reaction. Furthermore, the pressure--temperature cycle in the throat region is not readily amenable to prestressing, since adding compression to the liner only aggravates the already high compressive thermal stresses there.

A simplified thermoelastic stress analysis results in the following expression for tangential stress at the throat:

$$\sigma_\theta = \frac{\alpha E}{1-\mu}\left(1 + \frac{r_a^2}{r^2}\right)\left[\frac{Nr_b}{2} - \frac{N}{3}\left(\frac{r_b^3 - r_a^3}{r_b^2 - r_a^2}\right) + \frac{T_b}{2}\right] +$$

$$\frac{\alpha E}{1-\mu}\left(1 - \frac{r_a^2}{r^2}\right)\left[\frac{Nr_b}{2} - \frac{N}{3}\left(\frac{r^3 - r_a^3}{r^2 - r_a^2}\right) + \frac{T_b}{2}\right] -$$

$$\frac{\alpha E}{1-\mu}\left[N(r_b - r) + T_b\right] \qquad (h)$$

where $N = \dfrac{T - T_b}{r_b - r}$

At the inner and outer radii this reduces to

$$\sigma_{\theta a} = \frac{\alpha E N}{1-\mu}\left[r_a - \frac{2}{3}\left(\frac{r_b^3 - r_a^3}{r_b^2 - r_a^2}\right)\right]$$

$$\sigma_{\theta b} = \frac{\alpha E N}{1-\mu}\left[r_b - \frac{2}{3}\left(\frac{r_b^3 - r_a^3}{r_b^2 - r_a^2}\right)\right]$$

These stresses have been calculated for reasonably thick liners of dense material with steep temperature gradients, and were found to be more than an order of magnitude above the ultimate strength of any refractory (compressive at the inner surface and tensile at the outer surface). Using these expressions, the relationship between stress and temperature differential across the throat wall was also plotted, and it was found that benefits which could be gained in stress reduction by pre-heating the material were not significant until the temperature differential was reduced to a few hundred degrees. With gas stagnation temperatures commonly above 4000°F in hypersonic nozzles, maintaining a high pre-heat can become a technical problem.

patterns to reduce compressive stresses and to provide better expansion characteristics.

For the tensile region, the high stresses predicted by equation (h) may be counteracted with reinforcement in the form of pre-fired ceramic rods or compatible refractory metal fibers (9). Success with vacuum hot pressing and firing random distributions of ceramic rods and other forms of large grog in the base material has been reported (10). When utilizing the metallic fiber--ceramics in high temperature applications, fabrication techniques must be carefully controlled and the design must be closely related to the intended temperature environment, for mismatches in thermal expansion coefficients between the composite phases can be destructive. The expansion coefficient curves shown in Figure 12 identify material pairs which have reasonable compatibility and have performed successfully in this capacity. Also shown in Figure 12 is a micrograph of a typical metal fiber--ceramic two-structure system. For tungsten-mullite composites, Miller (11) reports rupture strengths increased up to 90% over original base material strength, using a metal fiber volume of 20%. However, oxidation resistance of the metal phase of this composite was generally poor above 2000°F, which indicates that applications would require tailoring to other than high temperature environments.

Another approach which has shown much promise for throat protection to 4000°F, but which is not yet economically feasible for large nozzle applications, utilizes a cladding or slurry coating process for application of a hafnium-tantalum alloy over different, compatible substrata (12, 13). A nozzle processed in this fashion is shown in Figure 13, together with micrographs of the grain structure that allowed this material to perform effectively in oxidizing, thermal shock environments. After the coating process is completed, the alloy is pre-oxidized in a controlled atmosphere to build up an arbitrarily thick layer of dual-component oxide. In this oxide formation, which is fairly complex and temperature sensitive, an outer scale of compound oxide first forms. This is followed by development, below the outer scale, of a two-phase system composed of a hafnium-rich α phase and a tantalum-rich β phase which form a microcomposite of alternating platelets or stringers. Oxygen diffusion into this layer preferentially oxidizes the hafnium-rich stringers leaving the β material intimately enmeshed in the oxide. There results a system having a tenatious coating that acts as a thermal barrier with good shock resistance. The cumulative effect of successive exposures to high temperature causes gradual recession of the Hf-Ta system, and eventually the substrate must be "recapped" to prevent its degradation.

The tangential throat stress associated with the gas stream pressure is, from the elastic theory,

$$f_{\theta_a} = p \frac{r_a^2}{r_b^2 - r_a^2} \left(\frac{r_b^2}{r^2} + 1\right)$$

and is, of course, superimposed upon the thermal stress during blowdown. Figure 11 illustrates the interaction of thermal and pressure stresses at the beginning and end of flow. Several of the major design problems are evident from this figure:

a. Prestressing the throat could help to counteract the high outer tensile stresses from thermal and pressure loads but would aggravate the compression at the inner face and is therefore impractical. Furthermore, high temperature relaxation of the liner would soon eliminate any intended prestress.

b. High thermal gradients result in thermal shock problems, though these may be mitigated by balancing the tensile pressure stress with the compressive thermal stress at the hot face.

c. Preheating can be effective in overcoming thermal shock problems though it complicates the throat fabrication.

d. When flow is stopped, the pressure stress disappears leaving the temperature stress unabated, though distributed somewhat because of its transient nature.

If we add to these the need for oxidation resistance and gas-flow shear resistance, the designer is left with no single-material candidates which can satisfy all requirements associated with the throat region.

Among the composite structures, many steel shell--alumina liner combinations have seen use to temperatures of 3000°F, particularly for protection of the convergent throat section. At higher temperatures, stabilized cubic zirconia must be used to avoid catastrophic volume changes accompanying the monoclinic--tetragonal and tetragonal--cubic crystalline transformations. Unfortunately, zirconia is quite sensitive to thermal shock and serves best when preheated, as discussed above. This prevents hot surface spalling--a manifestation of compressive or shear failure near the inner radius. Some work with zirconia has also been done (7, 8) using groove

Iridium, electro-deposited over metallic tungsten to a thickness of several mils, has shown promise for throat protection to 5600°F with an additional factor of design confidence provided by metallic ductility (14). The iridium bonds firmly to the tungsten surface and creates an excellent oxygen diffusion barrier. During exposure to high temperature flow fields, a volatile oxide of iridium slowly forms and is swept away in the gas stream. Though iridium recession rates of only 1 mil/hour of operation are common, this technique has not been translated to large nozzle systems because of unfavorable initial costs.

In reaching beyond the refractory oxides to the higher melting points of borides, nitrides and carbides, the problem of high temperature chemical reaction between throat and air stream becomes increasingly significant. One of the foremost contributions to wind tunnel throat protection using these materials has resulted from work with hafnium carbide in nose cone studies. In research reported by Kendall, et al, (15) it was found that in high-enthalpy air environments, hafnium carbide reacts to form a thin surface oxide coating of hafnia which provides a diffusion barrier against further carbide decomposition. In this research, attention was given to the hypereutectic carbide--carbon system in which the carbon phase forms a random distribution of minute platelets that act as built-in stress sinks for the high thermal compressive loads on the throat surface (see Figure 14a). Thermal shock can thus be suppressed by adjusting carbon content in this HfC+C system. Further development work is going on in refining the techniques of forming large shapes from this type of material. Hot isostatic pressing and sintering of hafnium carbide and graphite powders have been successful. Attempts are also being made to eliminate the undesirable structural weakening which is caused by oxygen diffusion along the carbon boundaries that results at points where carbon platelets intersect the HfC-HfO_2 surface. This preferential oxidation is shown in Figure 14b.

To review the material properties available for use as uncooled refractories, two nozzle throat designs are given in Figure 15. These configurations attempt to optimize the use of materials, both from a standpoint of performance acceptibility and cost effectiveness. Each cross section shows a steel retainer having a shell thickness calculated to resist both elevated working temperatures and deformations from the blow-down pressure pulse. In the unpreheated concept, shown on the left in Figure 15, the layer beneath the steel shell is designed to serve both as the insulation for the steel and as the tensile structural element in the liner.

Since this material will not experience the peak temperatures of the inner regions of the liner, working stress levels may be fairly high and the metallic fiber--ceramic material pairs of Figure 12 or the rod-reinforced ceramics would find application. The inner most ring of the unheated throat is fabricated of the hypereutectic carbide, HfC+C, to provide needed compressive stress sinks within the material. As previously mentioned, the layer of hafnia which develops at the exposed surface serves to control oxygen diffusion into the carbide. Gas-stream temperatures to 7000°F are feasible using this arrangement.

The preheated throat cross section on the right in Figure 15 overcomes thermal shock in the dense oxide liner by "anticipating" the temperature pulse peak. To reduce the gas-stream liner temperature differential carbon or silicon carbide heating elements are available with performance capabilities up to 4000°F. The refractory pack surrounding the heating elements is of graded density with the lighter material adjacent to the steel shell where the greatest thermal barrier is needed. Yttria-stabilized zirconia is considered best for maintenance-free endurance in both the dense and light oxides. Acceptable temperature rise rates indicate a preheat conditioning cycle of approximately 20 minutes for the zirconia. Gas-stream temperatures to 5000°F are feasible using this concept.

4. CONCLUSION:

In general, the same thermal and structural limitations of materials are faced by the designers of both flight vehicles and the test facilities for those vehicles. Flight weight restrictions on the one hand, or precision and repetitive durability of facility operation on the other may magnify certain aspects of design criteria, but the goal of higher temperature regimes and dynamic pressures is common. As heat environments for testing are increased, cooling must eventually be applied to even the most refractory materials that can be used in the throat. Unfortunately, cooling is not an open-ended application for nozzles. The best refractories remain in the solid phase only to 7000°F. If film or transpiration cooling factors for high-shear-flow test environments are applied to this material limit, stream temperatures of 10,000°F should be possible in large facilities. This is also about the limit that small-scale, back-side-cooled nozzle throats have endured using high-conductivity copper liners. (Wall thickness structural requirements on throats in larger facilities remove back-side cooling from consideration at much lower temperatures.)

Thus, one may conclude that for blow-down tunnels, the practical limit for true simulation is about Mach 13 (extrapolated from Figure 1, circa 10,500°R). Beyond this, either test duration or degree of simulation, including size must be sacrificed. Graphically, the developmental milestones discussed in this paper are shown in Figure 16 with their design amortization costs. The upward trend of these costs for future nozzle heat protection is evident, and at Mach 13, our present technology is exhausted. Clearly, if we are to have the heat protection needed to extend testing beyond this point, the challenge for a new level of ingenuity squarely confronts the wind tunnel designer.

REFERENCES

1. Roland, H. C., Pasqua, P. F. and Stevens, P. N.: Film and Transpiration Cooling of Nozzle Throats, AEDC TR 66-88.

2. Lewis, H. F. and Horn, D. D.: Some Considerations in the Design of Water-Cooled Nozzle Throat Sections for High-Temperature, High-Pressure Operations. AEDC TDR 64-51, AD 437973.

3. Schwarzkopf, P. and Weisert, E.D.: Self-Cooled Rocket Nozzles. RTD TDR 63-4046, Vol I-III.

4. Breslich, F. N. and Stejskal, L. M.: Evaluation of Impregnated Porous Refractory Oxide Composites. AFML-TDR 64-185.

5. Schmidt, D. L.: Interdisciplinary Research on Ablative Materials. AFML-TR-65-62.

6. Jacocks, J. L.: The Aerodynamic Design of Axi-symmetric Wind Tunnel Nozzles. MS Thesis, Dec 65.

7. Licciardello, M. R.; Ohnysty, B; and Stetson, A.R.: Development of Frontal Section for Super-Orbital, Lifting, Reentry Vehicle, FDC-TDR-64-59.

8. Stratton, W. K., et al: Advances in the Materials Technology Resulting from the X-20 Program, AFML TR-64-396.

9. Swica, J. J., et al: Metal Fiber Reinforced Ceramics, WADC TR-58-452.

10. Yavorsky, P., Unpublished data, Zirconia Corporation of America, 1966.

11. Miller, D. G., Singleton, R. H., and Wallace, A. V.; Metal Fiber Reinforced Ceramic Composites, Ceramic Bulletin, Vol. 45, No. 5 (1966).

12. Hill, V. L. and Rausch, J. J.: Protective Coatings for Tantalum-Base Alloys. AFML TR 64-354.

13. Marnoch, K: High Temperature Oxidation Resistant Hafnium-Tantalum Systems. AFML TR 65-240.

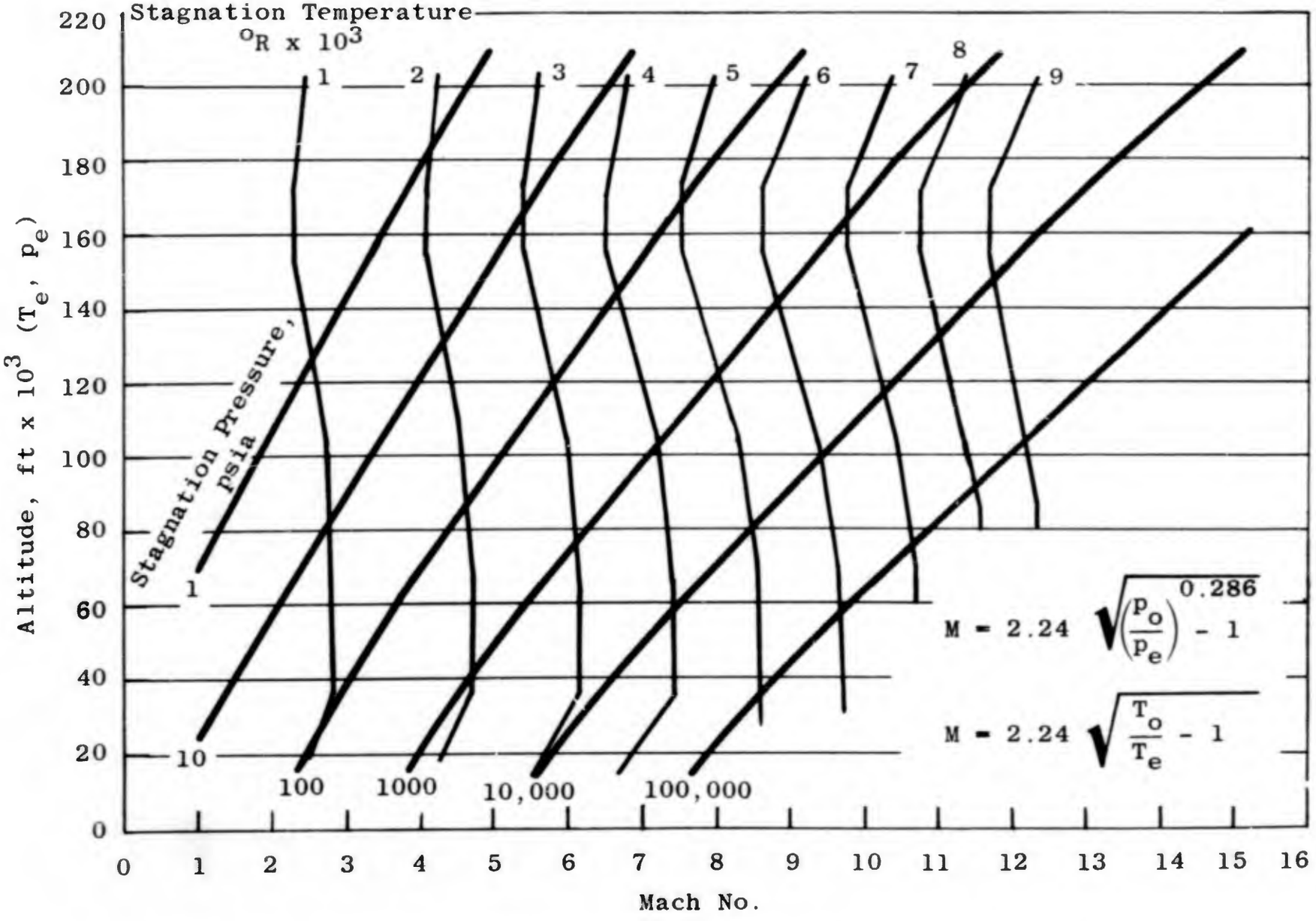
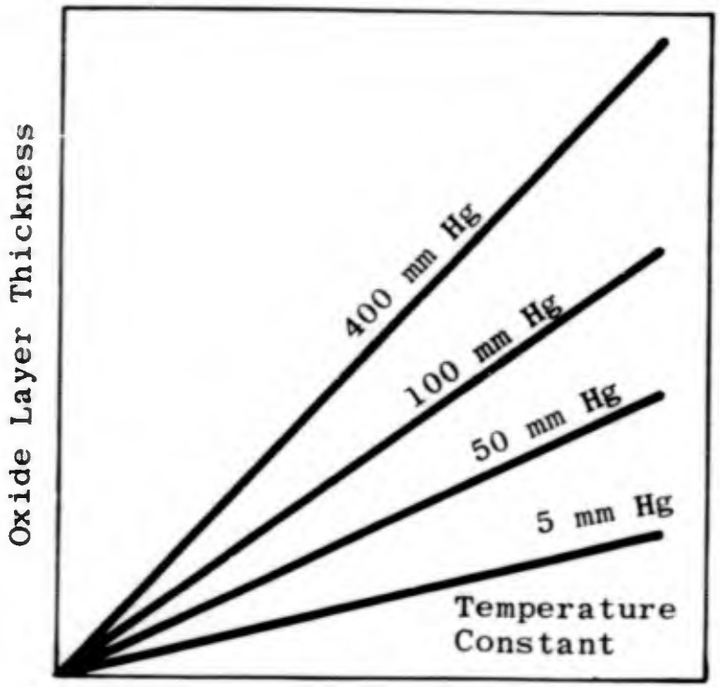


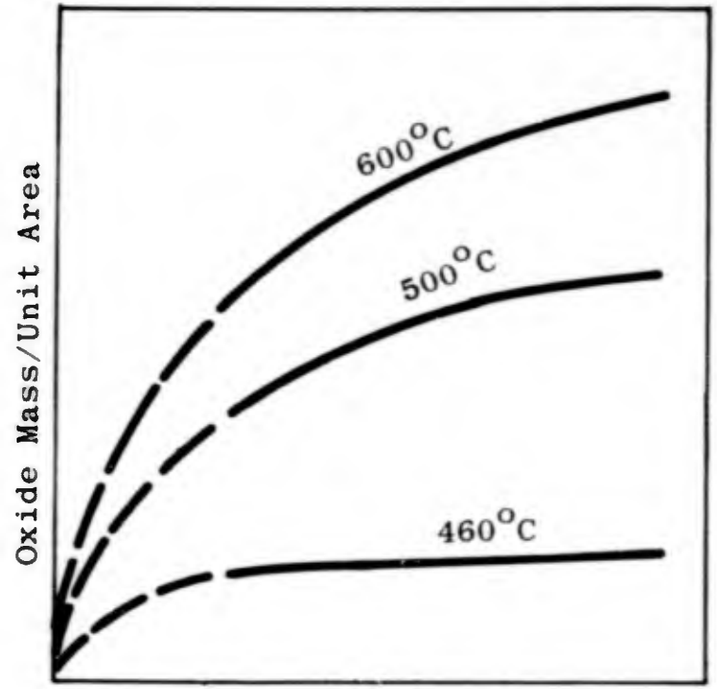
Fig. 1 Pressure-Temperature Requirements for True Flight Simulation

14. Brett, J. and Seigle, L. L.: Experimental Study of Factors Controlling the Effectiveness of High-Temperature Protective Coatings for Tungsten. AFML TR 64-392.

15. Kendall, E. G., Slaughter, J. I. and Riley, W. C.: A New Class of Hypereutectic Carbide Composites. AIAA Journal, Vol. 4, No. 5 (1966).

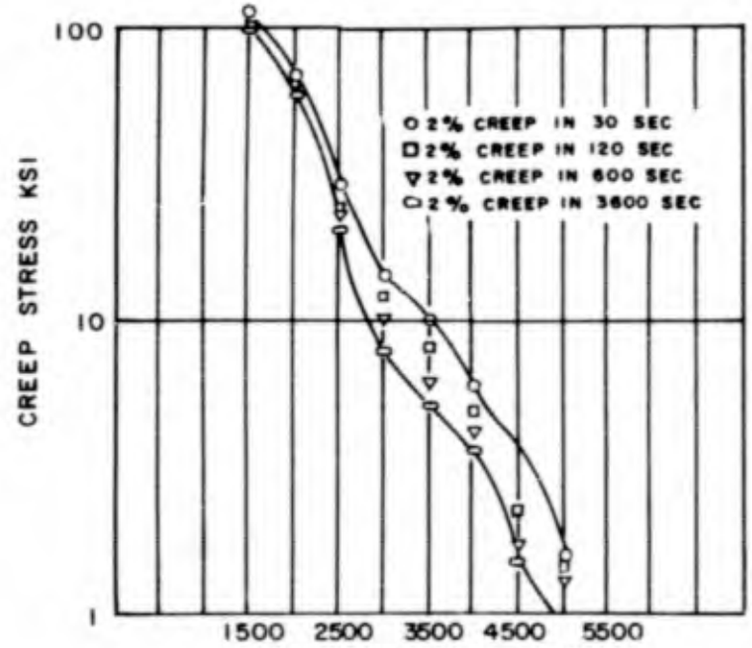


ZnO_2 Growth for Different Oxygen Partial Pressures

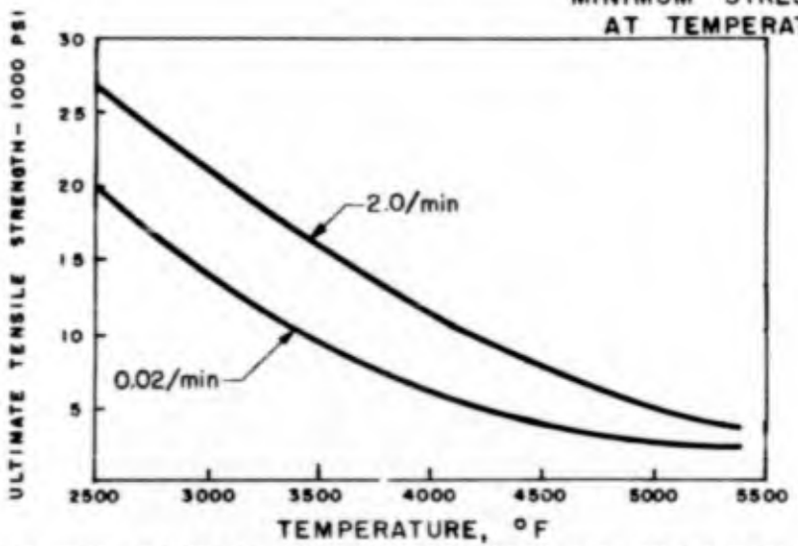


Oxide Buildup on Aluminum in Air

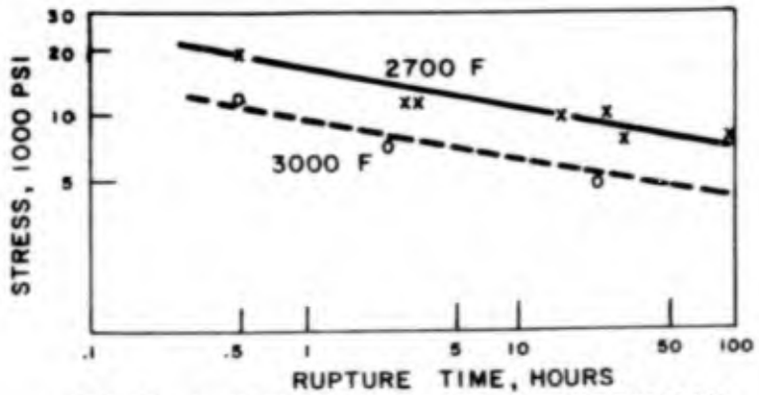
Fig. 2 Oxidation Behavior of Materials



MINIMUM STRESS REQUIRED TO REACH 2% CREEP AT TEMPERATURES FROM 1500°F TO 5000°F (ASD-TDR-63-585)



STRENGTH – TEMPERATURE RELATIONSHIP FOR RECRYSTALLIZED TUNGSTEN POWDER (TWO STRAIN RATES) (LOR & BOONE)



STRESS – RUPTURE DATA FOR COMMERCIALLY PURE TUNGSTEN ROD (26) (DMIC REPORT 191)

Fig. 3 High-Temperature Properties of Tungsten

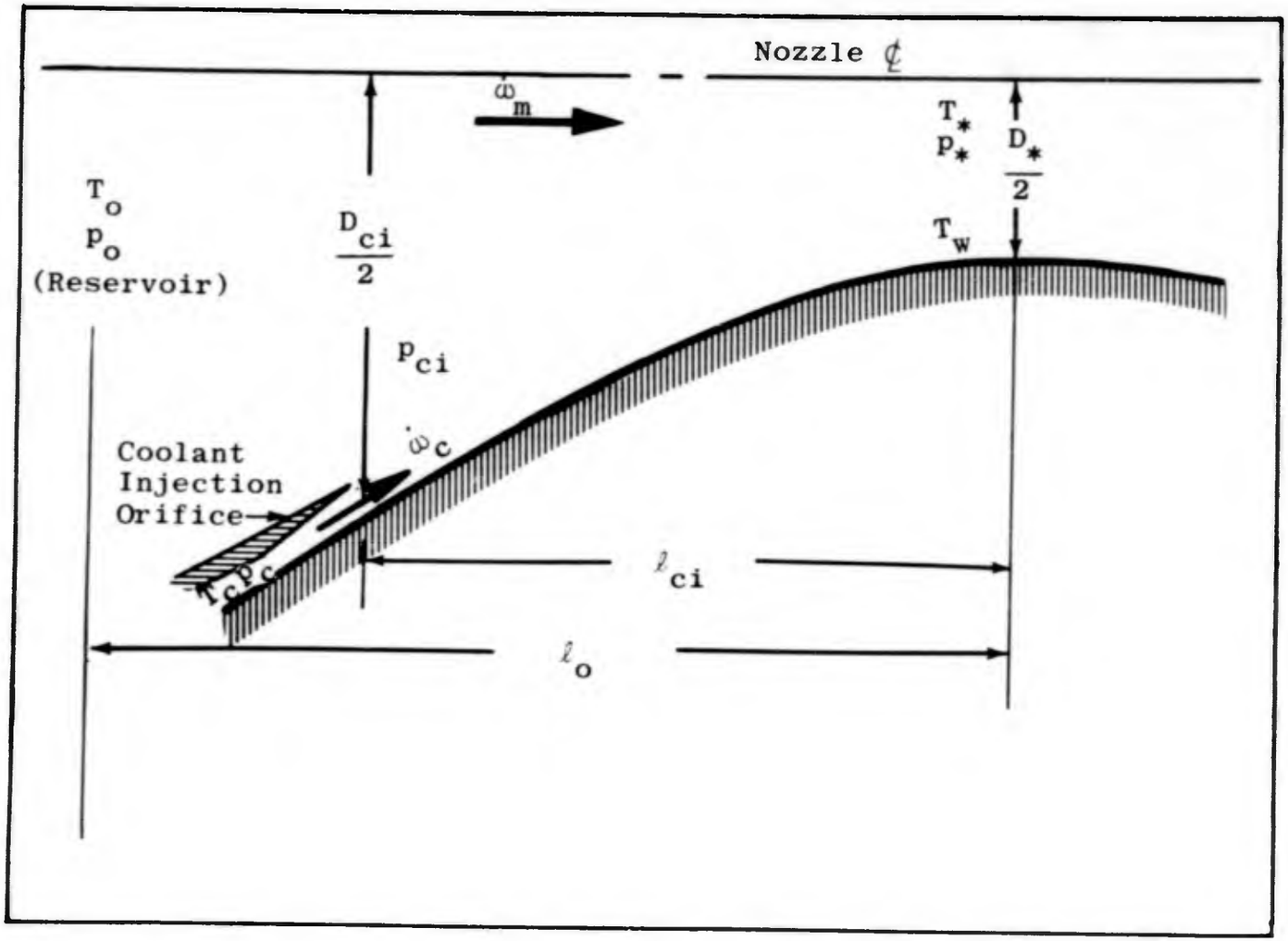


Fig. 4 Nozzle Throat Film-Cooling Configuration

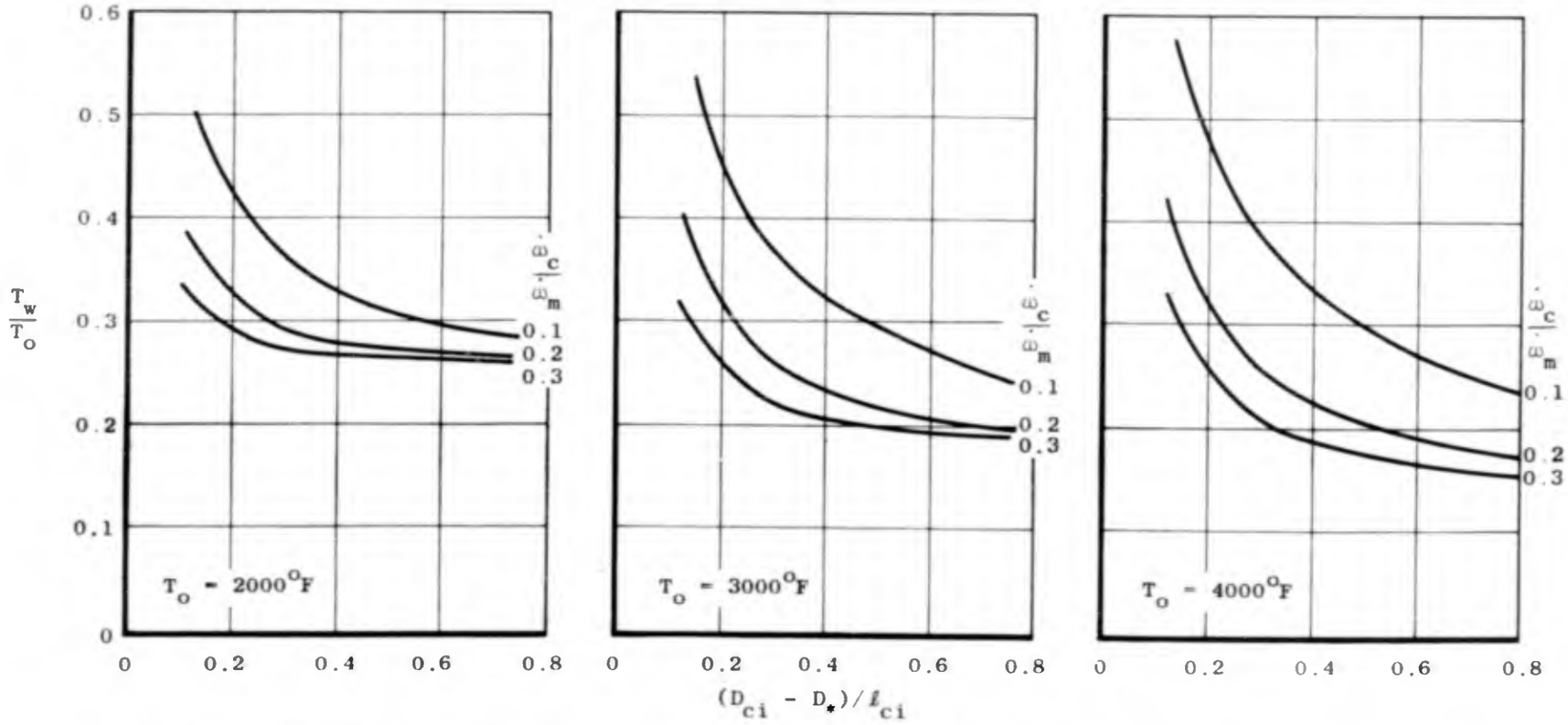


Fig. 5 Mass Flow Requirements for Adiabatic Throat Cooling with Air

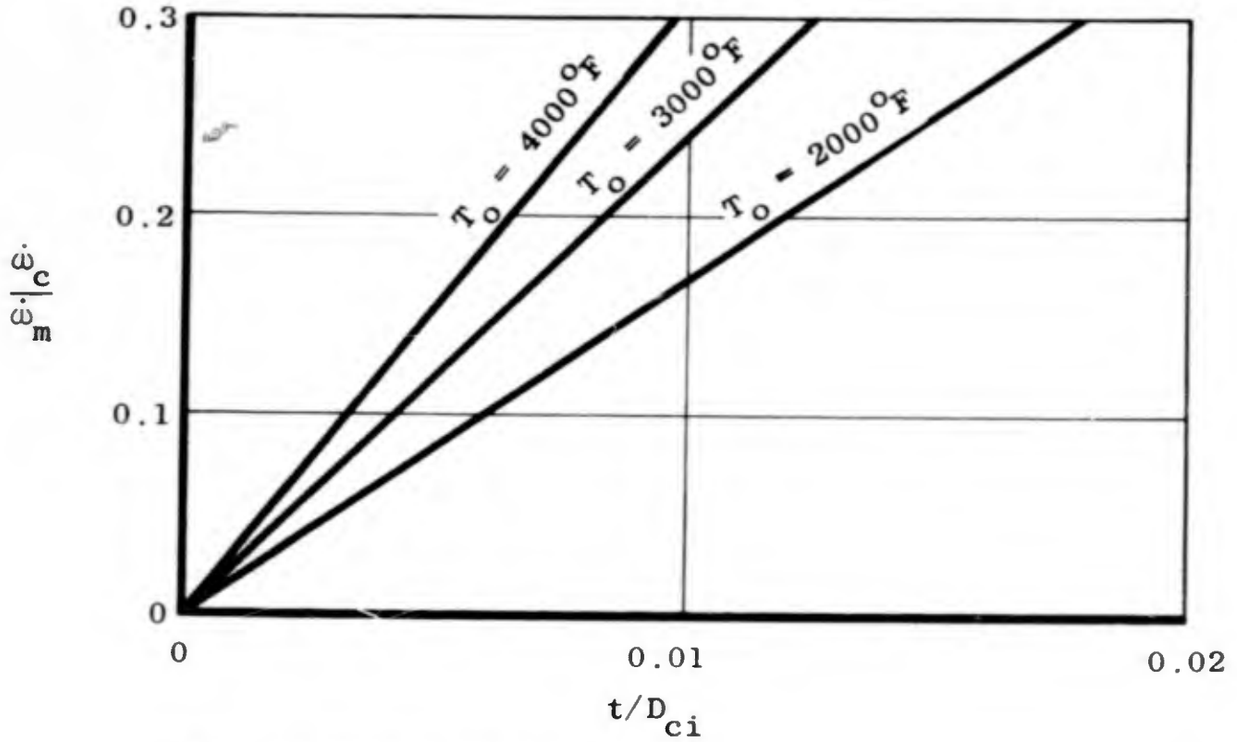


Fig. 6 Mass-Flow Ratio - Coolant-Orifice Dimensions for Air

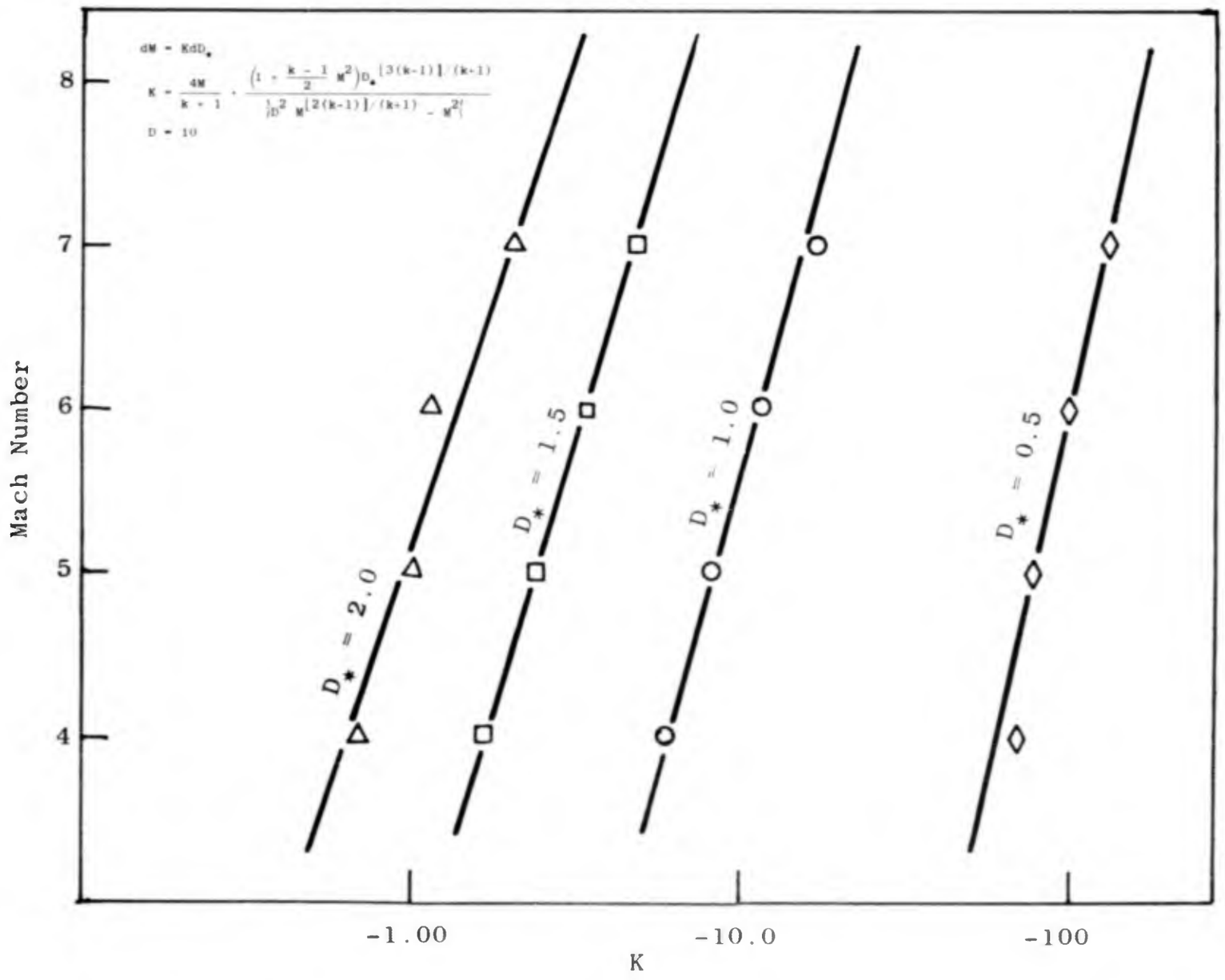
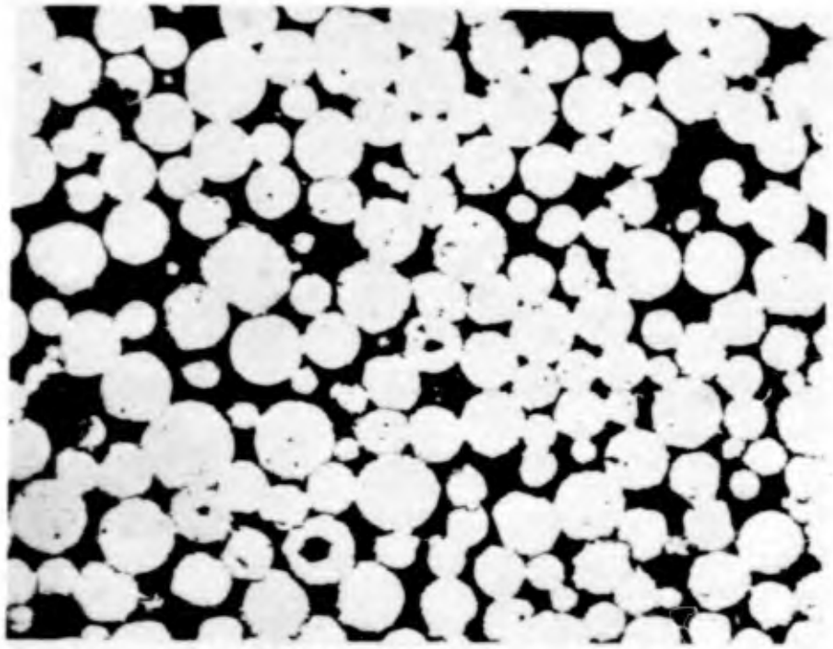
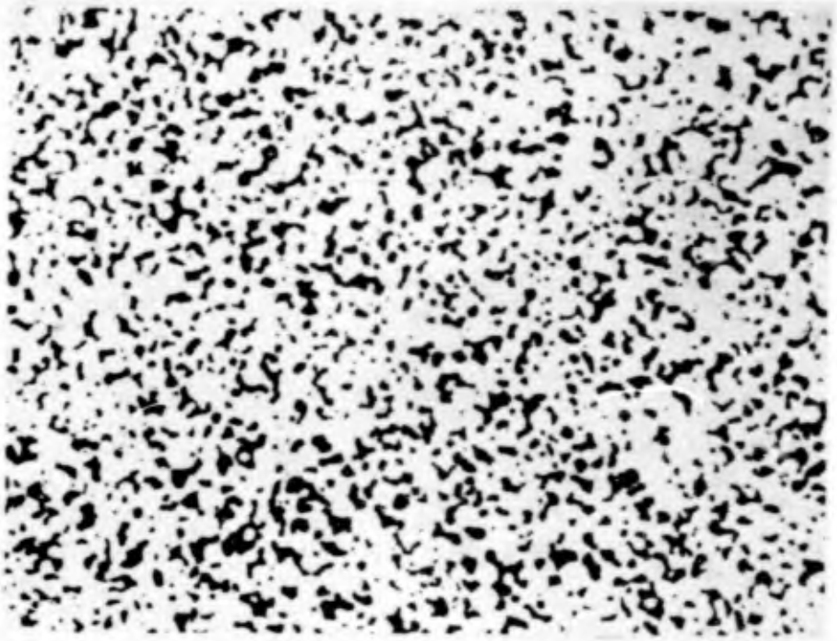


Fig. 7 Mach Number-Throat Diameter Sensitivity Factor, K



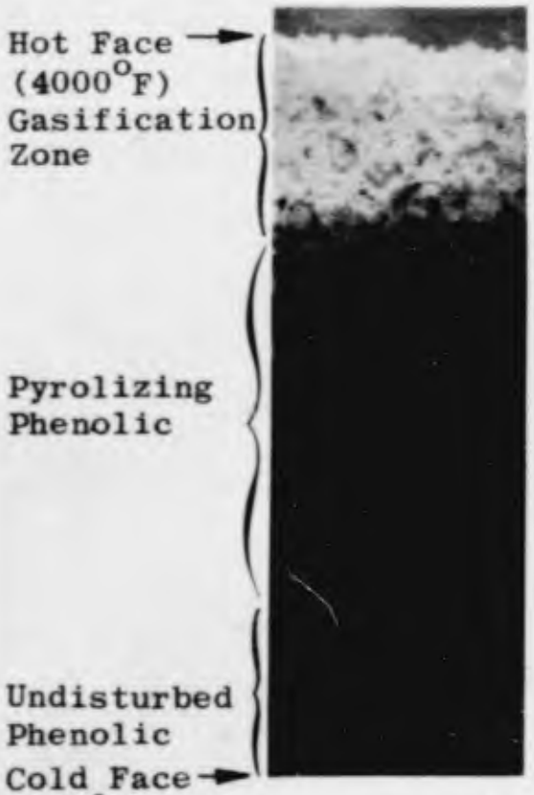
40-micron, 25%-Porous, Spherical Tungsten Powder Matrix (Unetched 250X)



10-micron, 19%-Porous, Spherical Tungsten Powder Matrix (Unetched 250X)

(From RTD-TDR 63-4046)

Fig. 8 Porous Tungsten Microstructure

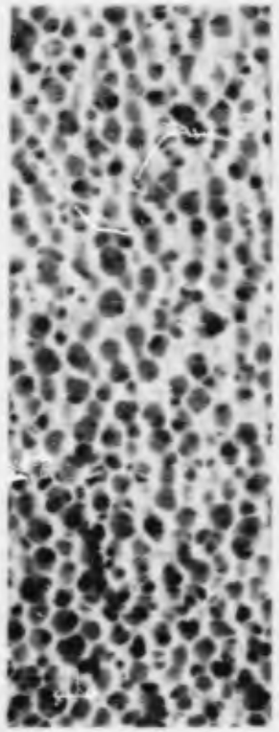


(Phenolic Filled Matrix)

X7

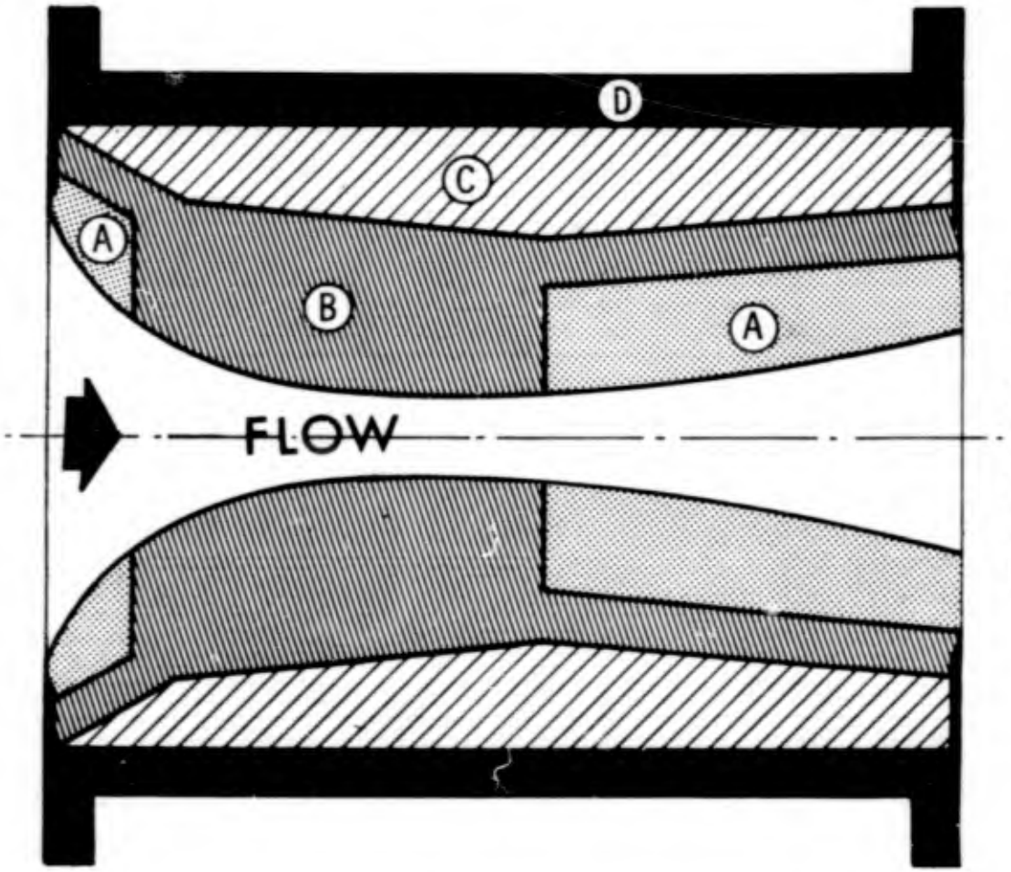
Double Pore-Size Structure

X30



Single Pore-Size Structure
(Unfilled Matrix)
X5

Fig. 9 Foamed Stabilized Zirconia



(A) Low Permeability Entrance and Exit Material for Minimum Gasification

(B) High Permeability Throat Material for Maximum Cooling

(C) High Porosity Reservoir for Maximum Gasifier Containment

(D) Steel Liner for Pressure Containment

Fig. 10 Compound Self-Cooled Nozzle Insert

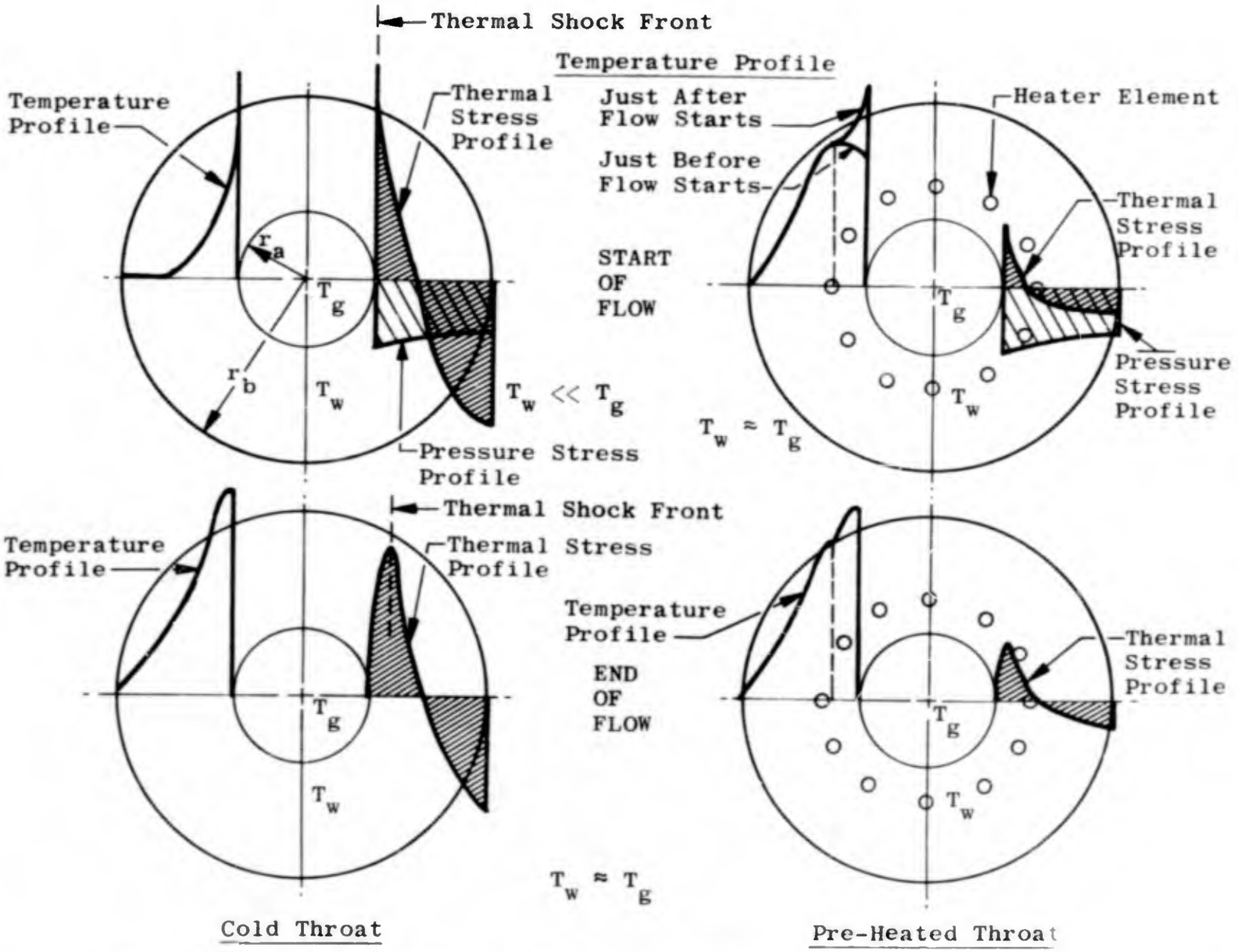
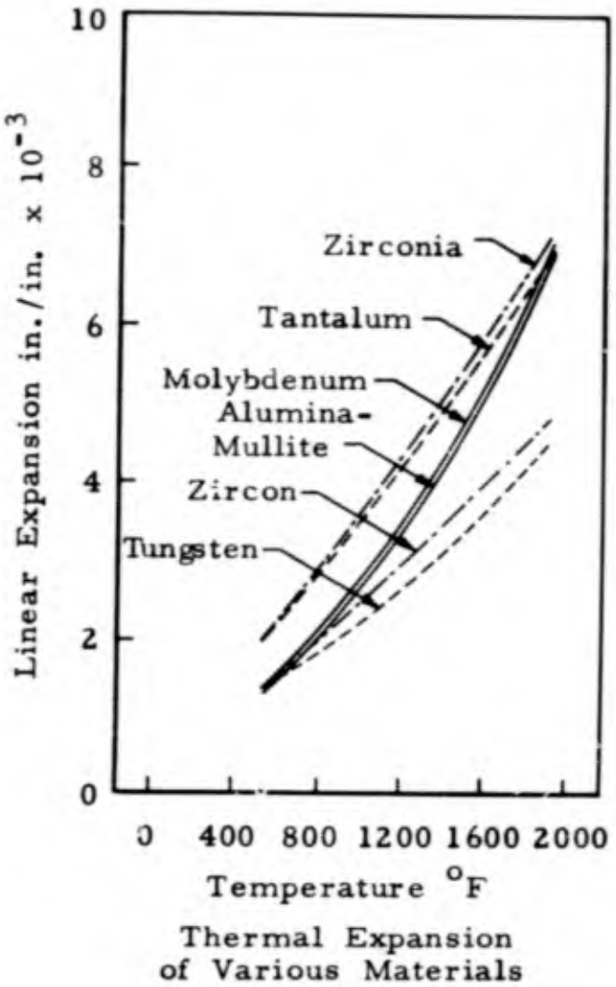
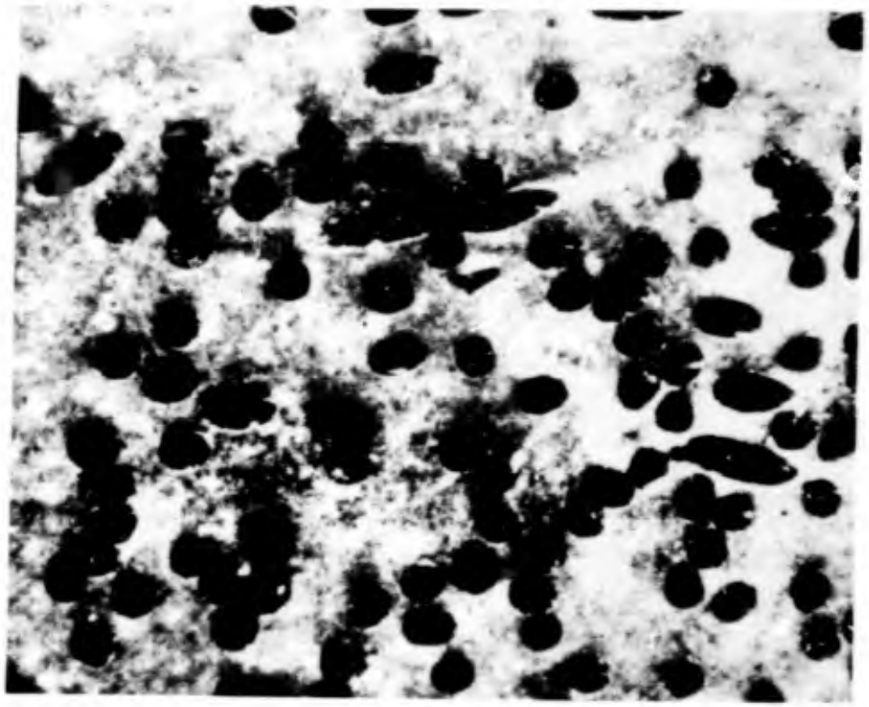


Fig. 11 Profiles of Nozzle Throat Temperature and Stress Transients



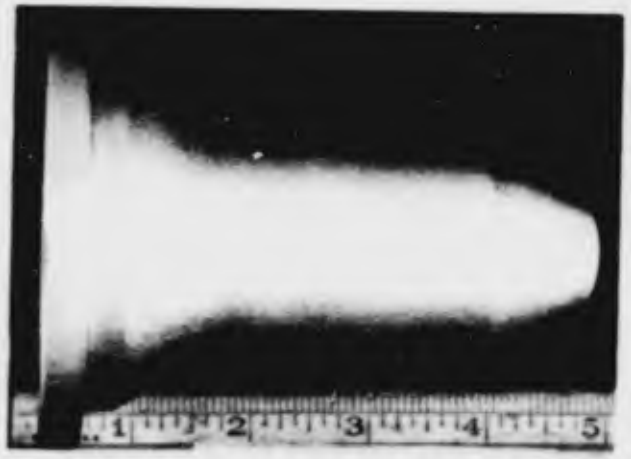
Thermal Expansion
of Various Materials

(From WADC TR 58-452)

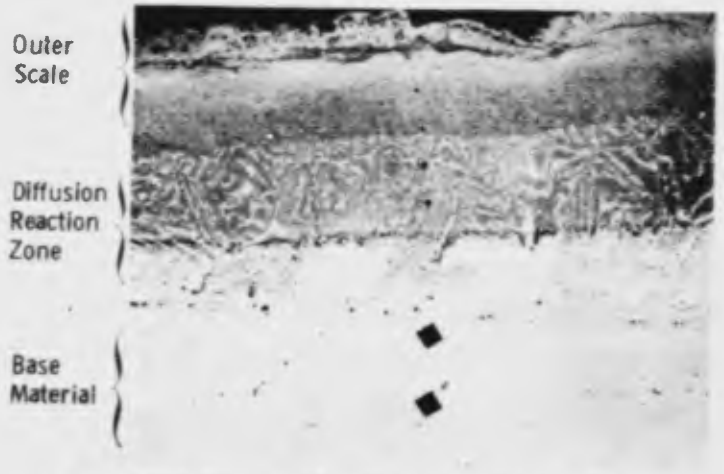


Metallic Fiber Reinforced Ceramic
(X80)

Fig. 12 Properties of Metallic Fiber-Ceramic Composites



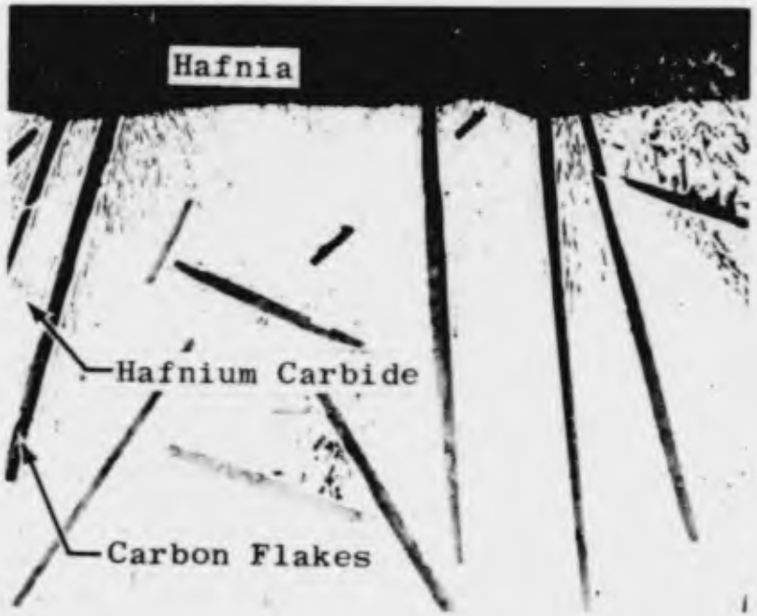
a. Small Nozzle Slurry-Coated with Hafnium-Tantalum Alloy (Sylvania Sylcor Div.)



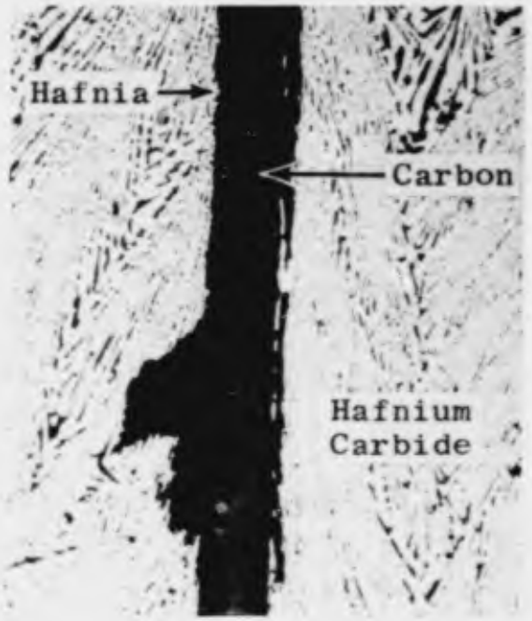
b. Oxide - Hf-Ta Reaction Zone (X100) (AFML-TR-64-354)

c. Oxide - Hf-Ta Reaction Zone (X3800) (AFML-TR-65-250)

Fig. 13 High-Temperature Oxidation Resistant Hafnium-Tantalum System



a. Surface Oxide Layer on HfC + C
(X100)



b. Oxide Penetration along Graphite
(X250)

(From Battelle Memorial Institute)

Fig. 14 Hypereutectic Hafnium Carbide-Carbon Micro-Structure

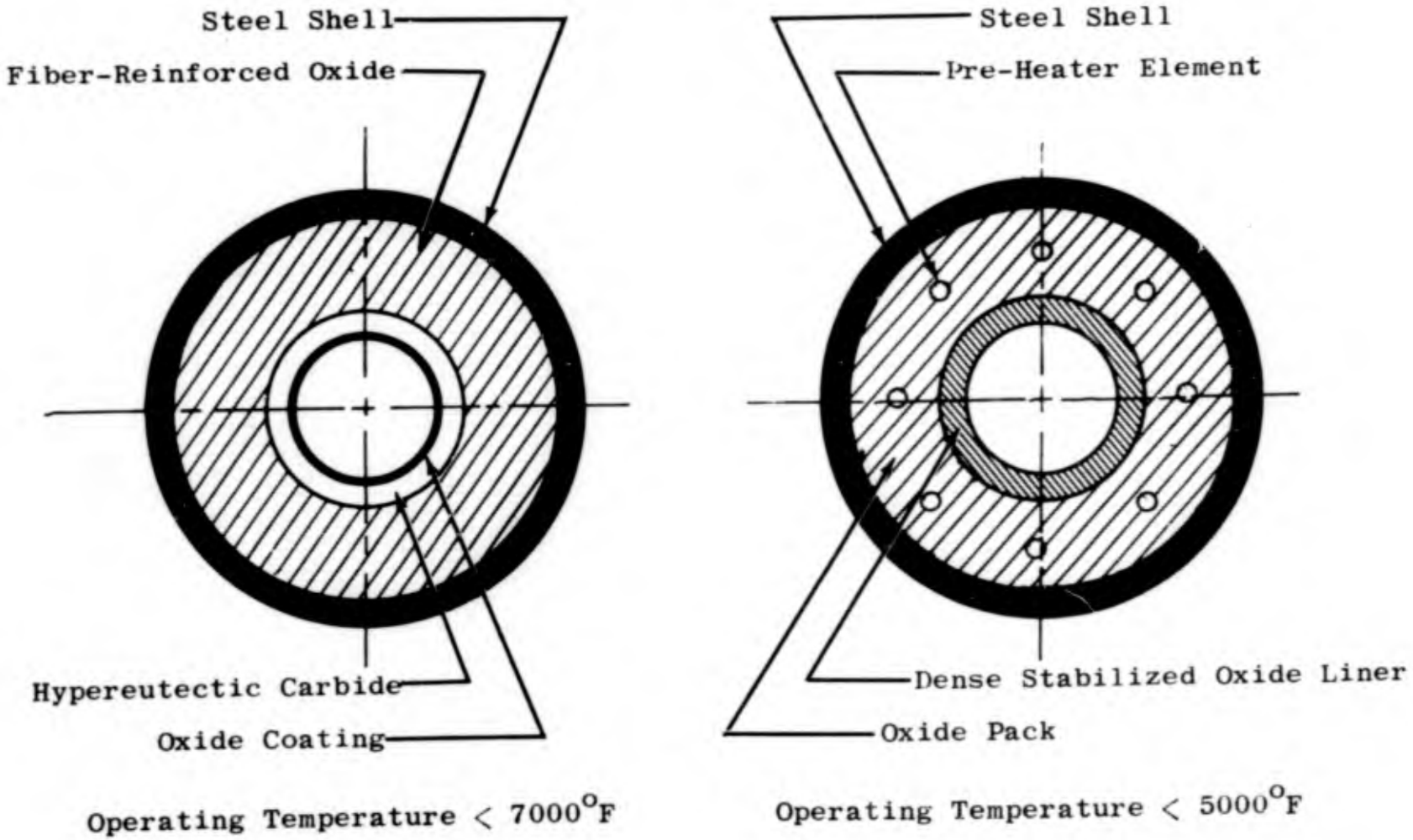


Fig. 15 Uncooled Composite Nozzle Throat Designs

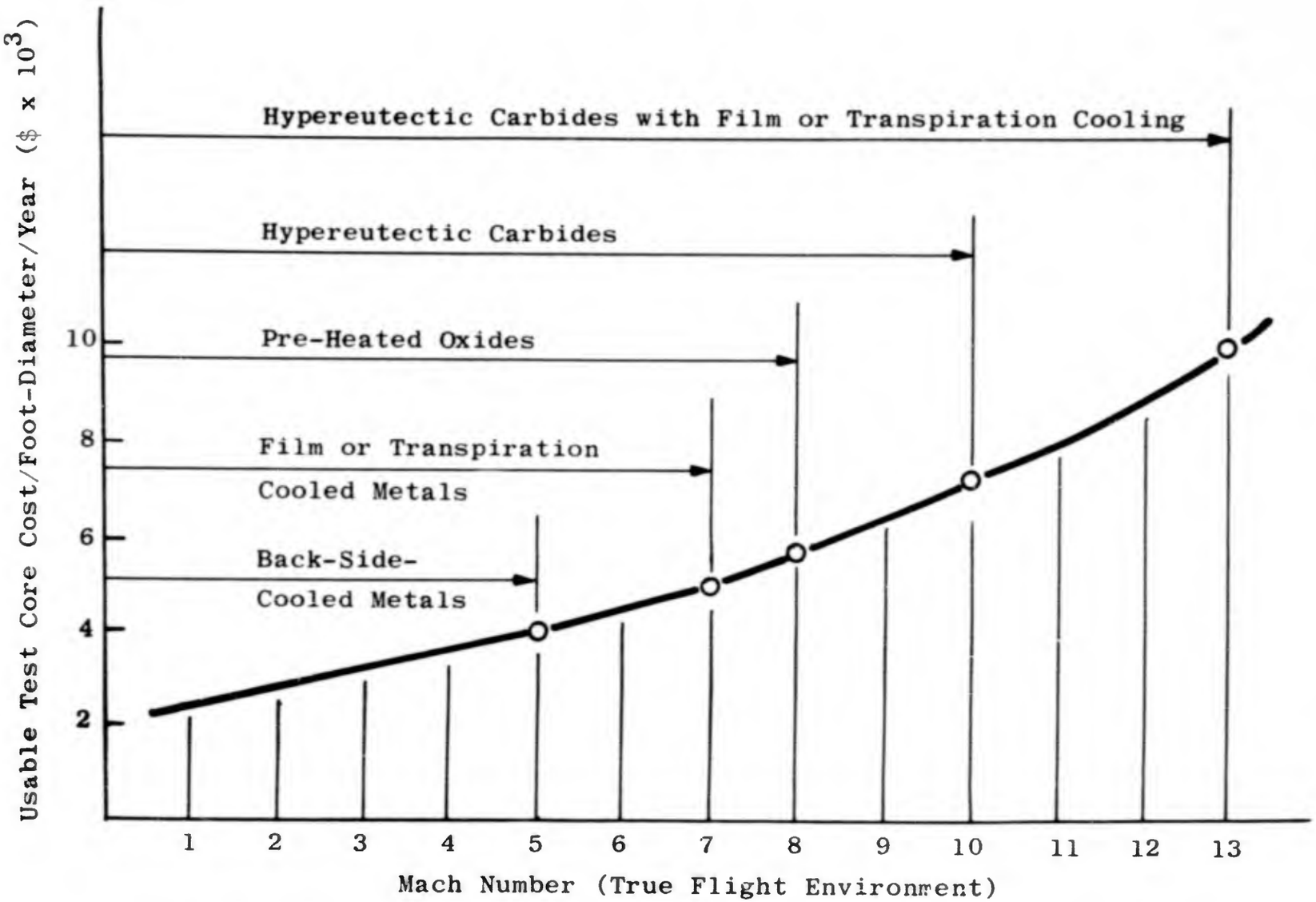


Fig. 16 Nozzle Throat Amortization Costs in Large Blow-Down Facilities

THE USE OF FLIGHTPATH ACCELEROMETERS IN PERFORMANCE FLIGHT TESTING

WILLIE L. ALLEN
Chief, Trainer & Misc
Engrg Section

ROBERT H. WEIGHT
1st Lt, USAF
Aerospace Engineer

AIR FORCE FLIGHT TEST CENTER
EDWARDS AIR FORCE BASE, CALIFORNIA
FLIGHT TEST ENGINEERING
PERFORMANCE ENGINEERING BRANCH

Willie L. Allen

WILLIE L. ALLEN
Air Force Flight Test Center

Mr Willie L. Allen is a supervisory aerospace engineer with the Flight Test Engineering Division, Air Force Flight Test Center. He currently supervises the Trainer and Miscellaneous Engineering Section in the flight testing of conventional and VSTOL aircraft.

Mr Allen was graduated from the University of Washington in 1952 with a Bachelor of Science in Aeronautical Engineering. He was employed as a flight test engineer at Edwards AFB and served as project engineer on various performance and stability flight test programs until he became chief of the Fighter Engineering Section in 1958. In that position he supervised numerous flight test programs of Air Force fighter aircraft before assuming his present position in 1966.

Robert H. Weight

ROBERT H. WEIGHT
1st Lt, USAF
Air Force Flight Test Center

Robert H. Weight, a native of Syracuse, New York, was graduated from the United States Air Force Academy with a BS Degree in Engineering Sciences in 1962. He attended the University of Michigan from 1962 to 1963, and was graduated with an MSE Degree in Aeronautical Engineering in December 1963. He has been assigned to the Flight Test Engineering Division, Air Force Flight Test Center, Edwards Air Force Base, since January 1964. He has worked on the F-5 Category II Stability and Control Program, F-4C Category II Performance Program, and was project engineer on the RF-4C Category II Performance and Stability Program. Lt Weight is presently the project engineer on the F-111A Category II Stability and Control Tests.

ABSTRACT

The development of a useful Flightpath Accelerometer System has taken place over the past ten years. In present systems an extremely sensitive accelerometer is mounted in the nose boom of an aircraft and aligned to the flightpath through the use of an angle of attack vane. The resolution of the accelerometer is increased through the use of a Vernier range extender. By directly measuring acceleration along the flightpath, excess thrust can be calculated from Newton's second law of motion. These techniques are presently in use at the Air Force Flight Test Center to gather range, drag, climb and maneuverability data from flight tests. Future developments and applications are discussed.

INTRODUCTION

Performance flight testing at the Air Force Flight Test Center is concerned with the in-flight determination of aircraft performance parameters such as drag, thrust, and fuel flow. Accurate determination of these basic parameters allow a complete description of the aircraft's performance characteristics and limitations.

The accurate determination of fuel flow does not present many problems to the flight test engineer with modern volumetric or mass flowmeters installed in the aircraft. Proper data handling methods will allow a time integration of fuel flow to determine the total fuel used from engine start. With an accurate engine start gross weight, it is possible for the engineer to accurately fix the aircraft gross weight at any given time.

The determination of drag and thrust generally speaking, is a much more difficult problem. Thrust cannot be measured directly without expensive and sometimes impractical pressure and temperature probes in the exhaust stream. Normally, these probes would be mounted on a swinging rake which could be passed through the exhaust stream at intervals. This type of an installation reduces the temperature problems involved in taking measurements in a hot gas flow. Much work has been accomplished in the field of direct thrust measurements, but many problems remain. For further information see reference 13.

Thrust can be calculated from pressure and temperature measurements at selected engine stations and engine performance curves furnished by the manufacturer. This method can in some cases furnish adequate thrust data for flight test purposes, and is much more easily implemented than the measurement of exhaust gas parameters.

Drag data can be obtained through stabilized level flight techniques. In stabilized level flight, the thrust component along the flightpath is assumed equal to the drag. Steady state flight test techniques are time consuming and therefore expensive. Quasi-steady state techniques such as constant altitude, and power setting accelerations can be used to obtain data throughout the flight envelope of a high performance aircraft. The assumption that thrust equals drag no longer holds when airspeed and/or altitude are varying. The difference between the actual thrust available and the thrust required for steady state level flight is called excess thrust.

Three methods of measuring the acceleration due to excess thrust are available to the flight test engineer; accelerometers, time rate of change of airspeed, and radar tracking (time rate of change of ground speed).

Several attempts have been made in the past to determine acceleration along the flightpath through the use of an accelerometer mounted to the airframe with its sensitive axis along the longitudinal axis of the aircraft. This approach presented several problems. It required a coordinate transformation from a body axis system to a wind axis system. Accurate determination of angle of attack was necessary. A complete description of the body axis accelerometer method is available in reference 12.

Mr. Willie L. Allen, then Chief of the Fighter Engineering Branch of the Air Force Flight Test Center, Edwards Air Force Base, California, carried this idea one step further. An attempt was made to align the sensitive axis of the accelerometer to the flightpath of the aircraft through the use of an angle of attack vane mounted on a nose boom. In 1960 this approach was used in the Category II performance tests on the F-106A airplane. The results showed that a true flightpath accelerometer was feasible, but the accelerometer used at that time did not have the required resolution.

With the advent of small high resolution accelerometers, a usable flightpath accelerometer system was realizable. It is the purpose of this paper to discuss the uses of these high resolution devices in performance flight testing.

CONSTRUCTION AND OPERATION

The flightpath accelerometer system presently in use at the Flight Test Center was procured from the Systron-Donner Corporation for use on the Category II performance tests on the F-4C and RF-4C aircraft. This system consists of three subassemblies; the sensor, the sensor electronics and filter, and the Vernier range extender.

The sensor is a Systron-Donner Model 4310 linear force balance servo accelerometer with a symmetrical full range of ±0.5g. This accelerometer functions as a miniature servo system responsive to input accelerations along the sensitive axis. As illustrated in figure 1, input acceleration causes a force which tends to move the seismic mass (1). This movement, which upsets the servo's balance, is detected by the position error detector (2). The servo nulling amplifier (3), requires a given preset input from the position error detector. The amplifier, therefore, sends a feedback current through the restoring coil which is located in a permanent magnetic field (4) causing a force exactly equal and opposite to the input force caused by the acceleration. The restoring force returns the seismic mass to its original position and in turn reduces or increases the output of the position error detector until it reaches null. The servo system is now in a force balance condition. The current required to achieve this balance is directly proportional to the input acceleration. An analog voltage output directly proportional to input acceleration can be obtained by measuring the drop this current produces across a known resistor (5).

Normally the sensor electronics (position error detector electronics and the servo amplifier) are enclosed in the same case as the torque coil and seismic mass and position error detector. Due to space limitations within the nose boom the electronics were enclosed in the same case as the signal filter. The analog voltage output from the sensor is filtered in two active second order filters. The frequency response of the filter is flat from zero to one cycle per second within ±0.5 percent and has an attenuation greater than 30 db from ten cycles per second out to infinity. The filtering network is designed to take out any high frequency engine vibration or nose boom vibration.

The third subassembly is the Vernier range extender. This range extender consists of a Vernier channel amplifier, a base channel amplifier, stepping circuitry and a reference voltage circuit. It converts the normal analog output voltage of the acceleration sensor into a two-channel output based upon the expanded scale concept. It also provides a direct reading output by recombining the base and Vernier output voltages.

The basic operation of the Vernier range extender can be understood by reference to figure 2. The Vernier amplifier amplifies the filtered acceleration signal until its output reaches five or zero volts, depending on whether the acceleration is increasing or decreasing, respectively. This voltage level (5 or 0 volts) is detected and the base amplifier is increased or decreased by a discrete voltage increment by the stepping circuitry. This voltage step, equivalent to a fixed increment of acceleration, is fed back to the input of the Vernier amplifier to offset the incoming acceleration signal. This negative feedback resets the Vernier amplifier and a further change in acceleration can be sensed at the Vernier amplifier. This system provides a total of 24 steps. There are three outputs from the range extender. The Vernier amplifier output, the base amplifier output, and a recombined analog output. Each output has a range of zero to five volts for an acceleration input from -0.5 to +0.5 g. Therefore, each output voltage is biased by +2.5 volts at zero g.

The packaging of the three subassemblies is shown in figure 3. The sensor is attached to an angle of attack vane and mounted in the nose boom of the aircraft (see figure 4). The electronics and filter were placed in a temperature controlled environment, either an electronics bay or aircraft cockpit.

ACCURACY AND RELIABILITY

Although no formal accuracy investigation has been accomplished, several factors are known to affect the accuracy of the system. Among these are: sensor temperature, electronics temperature, vibration, vane construction, and cross axis sensitivity. The temperature effect is probably the largest so it will be discussed first.

Temperature changes at the sensor affect both the scale factor (V/g) and the null of the sensor. The variation in voltage scale factor is approximately .01 percent of full range per degree Fahrenheit. This effect is almost totally due to the effect of temperature on the permanent magnetic field in the sensor. This effect can be compensated for by trimming the load resistor with a thermistor network. However, because of space limitations within the nose boom the load resistor and other sensor electronics are not located in the same temperature environment as the sensor magnet. The compensation can only be accomplished when the electronics and the magnet temperature environments track together. Therefore, there is a scale factor shift due not only to the temperature in the nose boom, but also due to a differential temperature between the sensor and the sensor electronics.

A null shift because of temperature changes is due to the hairsprings which carry the feedback current to the torqueing coil to restore the seismic mass. As the sensor temperature varies these hairsprings exert a residual torque on the seismic mass, causing null variation with temperature. The null can vary as much as .001 percent of full range per degree Fahrenheit.

The actual temperature environment of the sensor is not known. However, total temperature at the nose boom can vary from -42 degrees at .8 Mach number above the troposphere to 250 degrees centigrade at Mach 3 above the troposphere. Current studies are underway to determine what the temperature is inside the nose boom of an F-4 aircraft at various flight conditions. Until these studies are complete, no attempt will be made to correct for the temperature effects on the sensor. It is believed that the data from flights at over Mach 2 will not be reliable. The accuracy specifications on the sensor are only guaranteed up to 200 degrees Fahrenheit. Therefore, it may be necessary to restrict the flightpath accelerometer from certain areas of the flight envelope of high performance aircraft or provide a controlled environment for the sensor.

High amplitude, asymmetric vibration can saturate the accelerometer and cause a shift in the filtered signal. A symmetrical vibration will not shift the signal even though the accelerometer is saturated. It is best to keep vibration

inputs to a minimum since no correction factor for vibration is available. On systems currently in use vibration effects are neglected.

Cross axis sensitivity can cause errors at other than 1-g flight. Problems with cross axis sensitivity will appear while trying to determine thrust limited maneuvering capability. Depending upon the aircraft, normal load factors as high as 7 or 8 g's can be encountered. The Donner Model 4310 accelerometer has a cross axis sensitivity of .002 g per g. Therefore, if 8 g were applied at 90 degrees to the flightpath, an additional .016 g would be sensed along the flightpath. This error can easily be corrected, and presents no problem if properly handled.

Other sources of error such as nonlinearity and hysteresis can be accounted for in the calibration of the system.

To determine zero longitudinal g, it is necessary to align the vane perpendicular to the local vertical. The present method employs a pendulum which is hung on the vane and allowed to swing until it seeks the local vertical (see figure 5). The pendulum is reversible so that the effects of misalignment can be neglected from the preflight zero.

Up-wash effect on the angle of attack vane also will influence the accuracy of the data. Investigation has shown (reference 12) that a certain vane and nose boom geometry is necessary to minimize the up-wash effects on an angle of attack vane. It is advisable to place the accelerometer and vane as far forward on the nose boom as possible to reduce these local flow effects.

For the reasons discussed above, differential acceleration data obtained over a very short time period should be considerably more accurate than absolute values of acceleration measured with this system.

APPLICATIONS

The advantages of the flightpath accelerometer system can best be seen through a description of other methods for obtaining excess thrust. The most common of these is the airspeed-time method. To illustrate the principles involved in this method, analysis techniques for a typical performance test, the level acceleration will be described.

A level acceleration test is conducted by accelerating and decelerating the aircraft throughout the speed range at constant altitude and power setting. To obtain excess thrust data the airspeed indicator is read at time intervals throughout the acceleration. The indicated airspeeds are instrument corrected and converted to calibrated and then to true airspeed. The conversion from calibrated airspeed to true airspeed can involve appreciable corrections. As an example an error of one knot in calibrated airspeed amounts to almost a two knot error in true airspeed at high altitudes. The true airspeed reading is then time differentiated which further amplifies errors in the airspeed readings. A longer time interval between data points minimizes the effect of measurement errors; however, it also decreases definition of variations in the computed excess thrust. A small interval between data points on the other hand allows for sharp variations in computed excess thrust, but any errors in airspeed will be amplified which will result in an erroneous value of the computed excess thrust. Usually a small time interval is selected and data smoothing techniques are applied to the resultant dV/dt curve to minimize the effect of airspeed errors. Judgment must be applied in the use of the smoothing routines so that significant variations in excess thrust are not smoothed out of the data.

There are many problems inherent in the airspeed-time method of calculating excess thrust. However, it is still the most reliable method and will probably be used in conjunction with other methods in the future.

Radar or phototheodolite tracking will also produce data which the engineer can use to calculate excess thrust. However, all of these methods, radar tracking, phototheodolite tracking, and airspeed-time require a time averaging process to generate curves of dV/dt. One of the main advantages of the flightpath accelerometer is that it has a continuous output and calculated values of excess thrust are independent of the sampling rate. The need for careful selection of time intervals and/or smoothing techniques is eliminated with the use of the flightpath accelerometer.

Figure 6 shows airspeed-time and FPA excess thrust data obtained from the same level acceleration. The data scatter in each figure gives an indication of the improved

quality of the FPA data. One of the reasons for the improved quality is the resolution available from the FPA. The specifications for the Model 4310 accelerometer quote a resolution of less than 10^{-5} g which is equivalent to a change in airspeed of approximately .01 knots per minute. Resolution of this magnitude is impossible with the airspeed-time method.

From the excess thrust data it is quite easy to calculate the climb performance of an aircraft.

It will be noticed from figure 7 that the rate of climb curves are more clearly defined by the FPA method than the airspeed-time method. This is due to the fact that the smoothing routines used in the airspeed-time method to reduce scatter also have a tendency to round off any changes in slope.

For reasons previously discussed the FPA is more accurate in determining a differential acceleration rather than an absolute acceleration. This feature adopts itself quite readily to the determination of the drag of any configuration change such as landing gear, flaps, or external stores. It is believed that this feature will prove particularly useful for the determination of the drag of external stores. An actual store or a dummy shape can be dropped from stabilized level flight to determine the drag of the store. For accurate determination of this drag it is now necessary to conduct extensive stabilized level flight (thrust = drag) tests for each loading. If the store drop method proves successful, the amount of flight test time required to determine external store drag will be drastically reduced.

The success of the store drop method will be dependent in a large part upon the knowledge of the induced drag of the aircraft, both with and without the external stores. Dynamic flight test techniques such as constant altitude accelerations and decelerations and varying load factor maneuvers are under development at the Flight Test Center for determination of aircraft drag characteristics. Although much remains to be accomplished in this area the results of preliminary tests indicate that these dynamic test methods will prove to be very useful.

Another useful application of the flightpath accelerometer is in the determination of thrust limited maneuvering capability, and energy maneuverability. These data are gathered by performing level accelerations to the thrust limited Mach number. Then at constant power setting and altitude a constant load factor turn is initiated, allowing the airspeed to bleed off to a minimum Mach number. These decelerating turns are performed at several

load factors. Excess thrust data is obtained from the FPA during these maneuvers analyzed in the usual manner. With this technique a definition of the maneuvering capability throughout the flight envelope can be obtained in a very short time.

The use of these dynamic flight test techniques has been feasible only since the advent of the flightpath accelerometer. The transient characteristics of the airspeed indicator and altimeter are not adequate to give the resolution required by these methods. Other techniques such as stabilized turns verify the data obtained from the windup turns described above.

For accurate definition of aircraft performance it is necessary to accurately determine the acceleration normal to the flightpath as well as acceleration along the flightpath. With the present single axis flightpath accelerometer systems, a normal accelerometer is mounted to the airframe and a coordinate transformation is used to obtain acceleration normal to the flightpath. Future efforts will include an investigation of the feasibility of including a normal acceleration sensor in the same package with the longitudinal accelerometer.

A visual indicator has been connected to the flightpath accelerometer system as a pilot aid in stabilizing the aircraft. Initial results show that the time required to stabilize the aircraft has been cut in half through the use of the visual readout. It is anticipated that this visual indicator will permit development of test methods that will provide even greater savings in flight test time.

CONCLUSIONS

All aircraft performance flight test techniques in common use today require a knowledge of aircraft acceleration or deceleration along the flightpath. The application of high resolution accelerometers has made possible the direct measurement of these accelerations and has greatly enhanced the Air Force Flight Test Center's capability of defining aircraft performance using both steady state and dynamic flight test techniques.

REFERENCES

1. Weight, R., 1st Lt, USAF, Gentry, J., Capt, USAF, Rogers, R., Major, USAF, RF-4C Category II Performance and Stability Test, FTC-TR-65-40, Air Force Flight Test Center, January 1966, UNCLASSIFIED.

2. Yancey, M., Twinting, W., Capt, USAF, F-4C Category II Performance Test, FTC-TR-65-41, Air Force Flight Test Center, January 1965, UNCLASSIFIED.

3. Systron-Donner Corp., Specifications for the Model 4310 Accelerometer, Unpublished.

4. Aldin, A., Novikoff, H., Accelerometer Fundamentals - Part I, Systron-Donner Corp., 1 November 1963, UNCLASSIFIED.

5. Clampitt, R., Novikoff, H., Accelerometer Fundamentals - Part II, Systron-Donner Corp., 1 November 1963, UNCLASSIFIED.

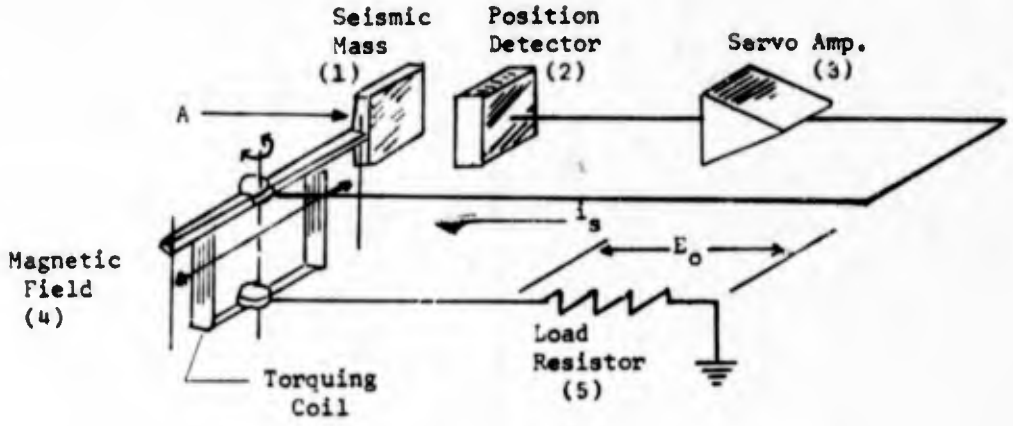
6. Wong, G., Report on Flightpath Accelerometer, TIM-65-1016, Air Force Flight Test Center, July 1965, UNCLASSIFIED.

7. Walker, R., Capt, USAF, Calculation of the Zero Lift Drag Coefficient of External Stores Using Wind Axis Accelerometers, FTC-TIM-64-1010, Air Force Flight Test Center, UNCLASSIFIED.

8. Walker, R., Capt, USAF, A Comparison of Several Techniques for Determining Aircraft Test Day Climb Performance, Air Force Flight Test Center, June 1965, UNCLASSIFIED.

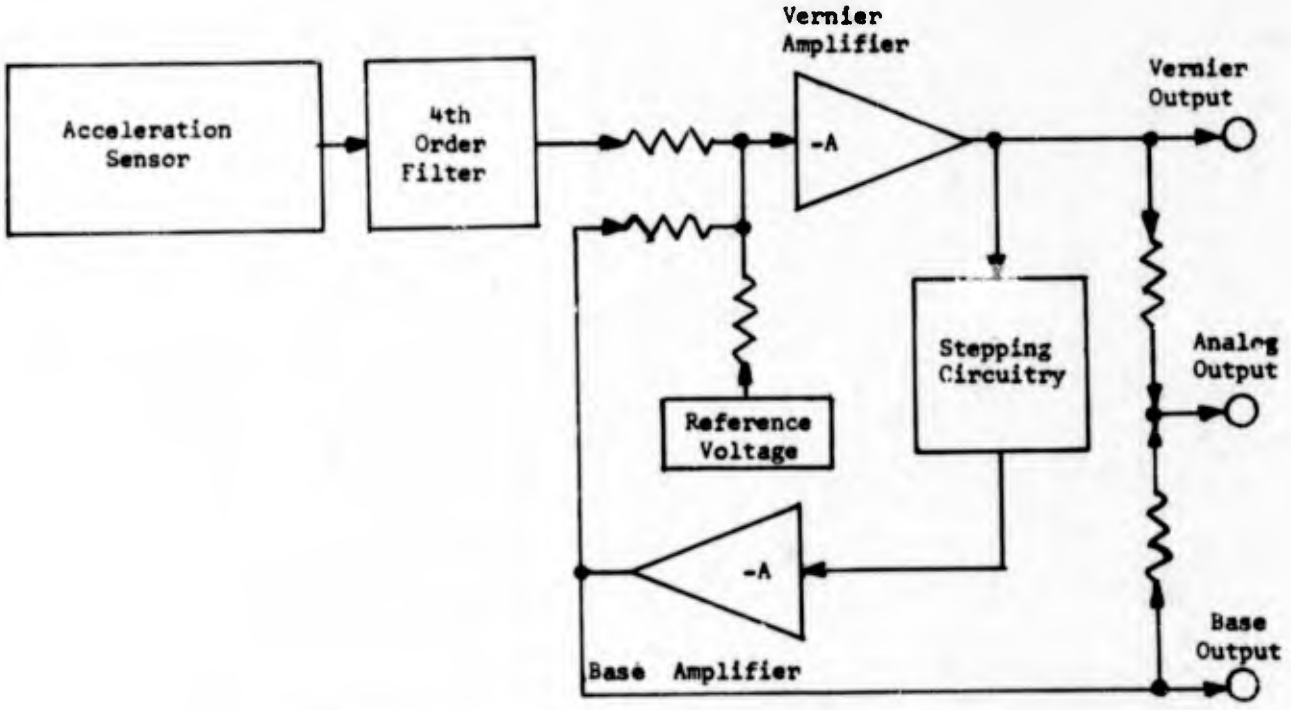
9. Beeler, E., Beelman, F., Saltzman, E., Flight Techniques for Determining Airplane Drag at High Mach Numbers, NACA-TN-3821, NACA, August 1965, UNCLASSIFIED.

10. Davidson, T., Measurement of Net Thrust in Flight, Journal of Aircraft, Volume I, No. 3, May-June 1964, UNCLASSIFIED.



Functional Diagram for Model 4310 Accelerometer

Figure 1



Simplified Block Diagram for the Vernier Range Extender

Figure 2

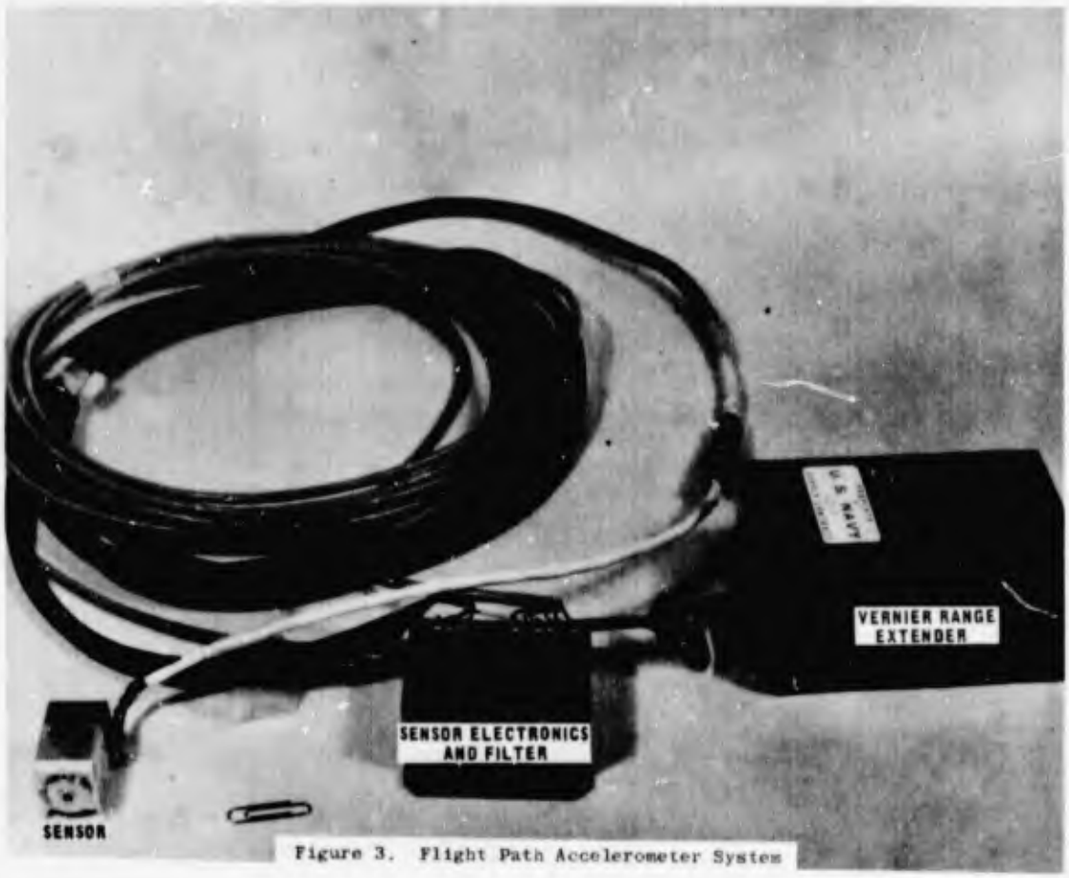


Figure 3. Flight Path Accelerometer System

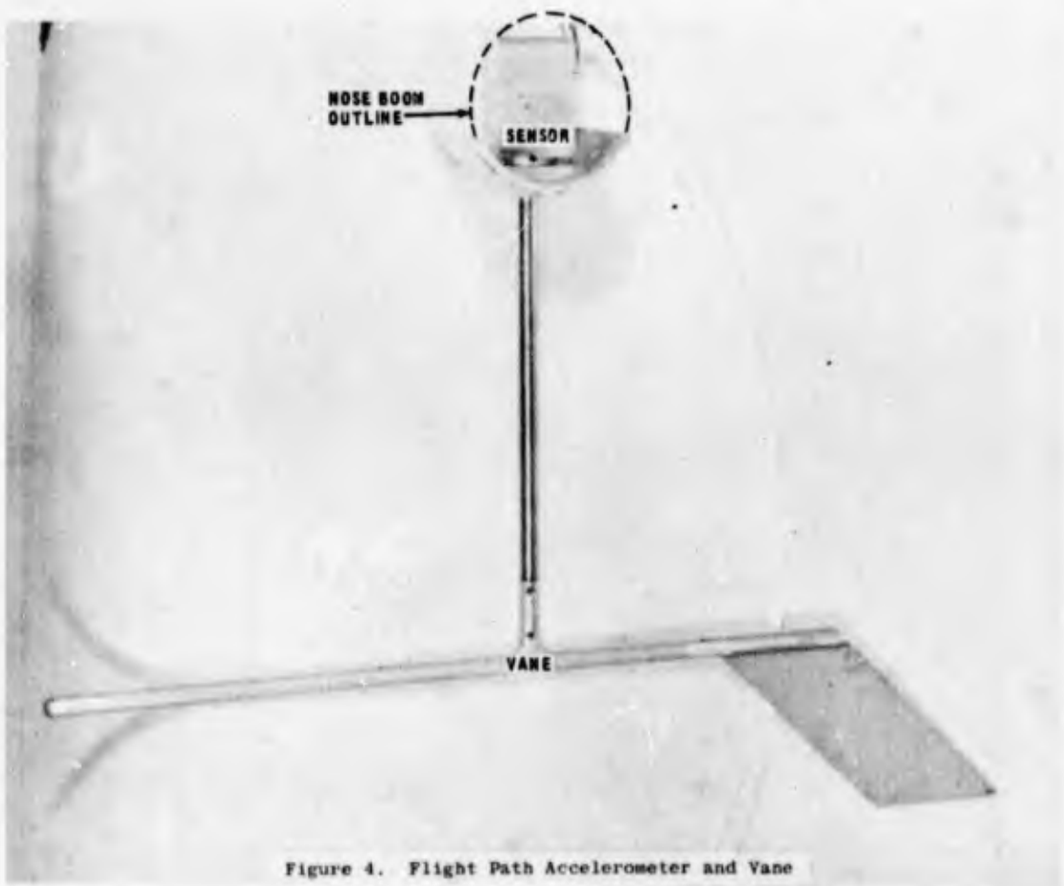


Figure 4. Flight Path Accelerometer and Vane

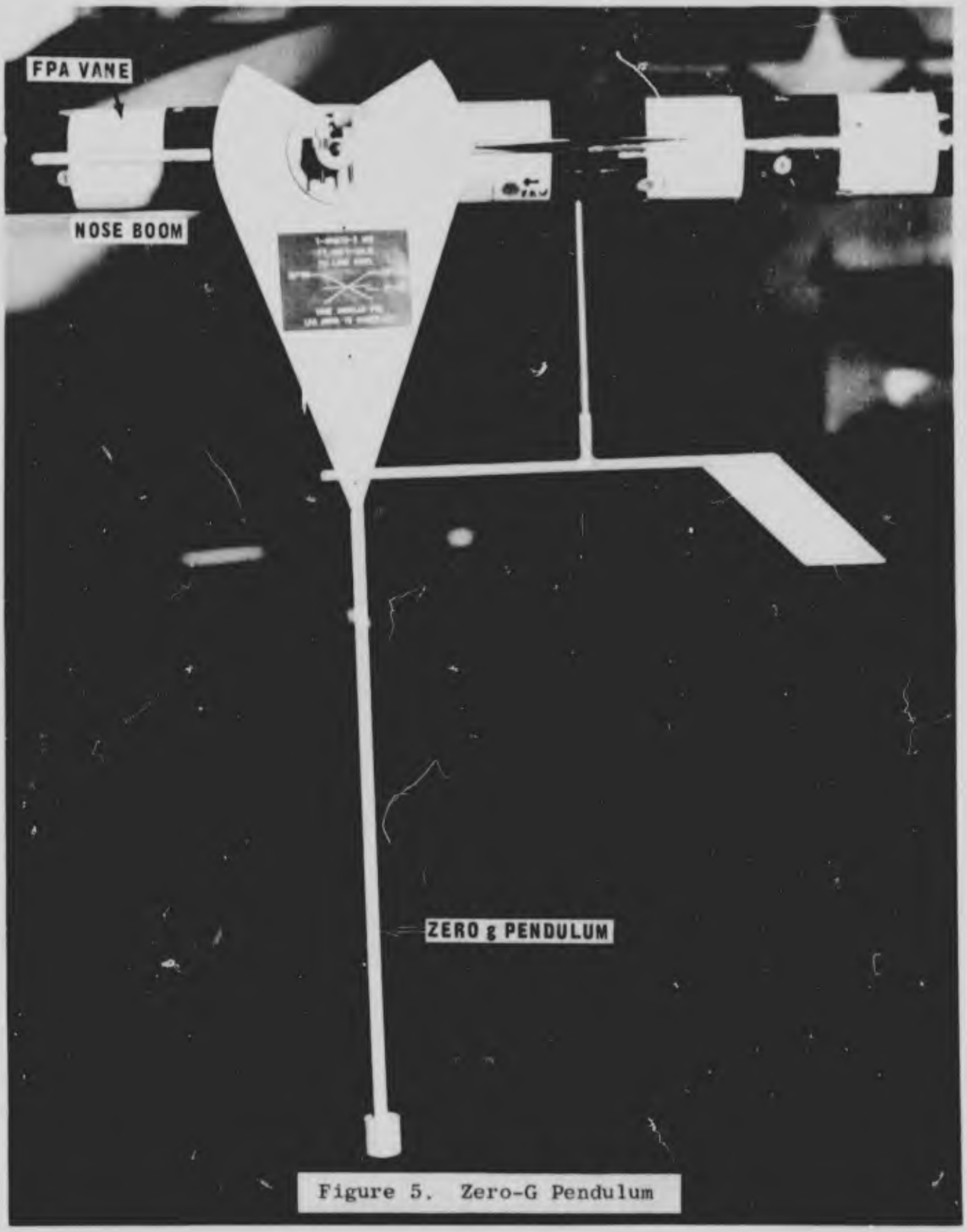
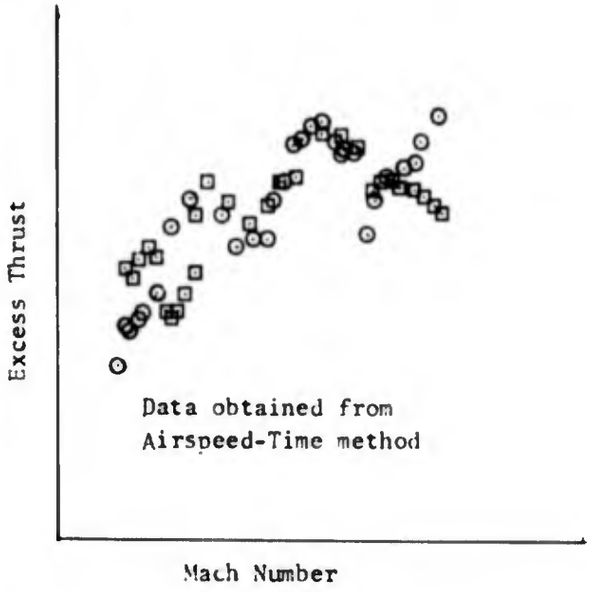


Figure 5. Zero-G Pendulum



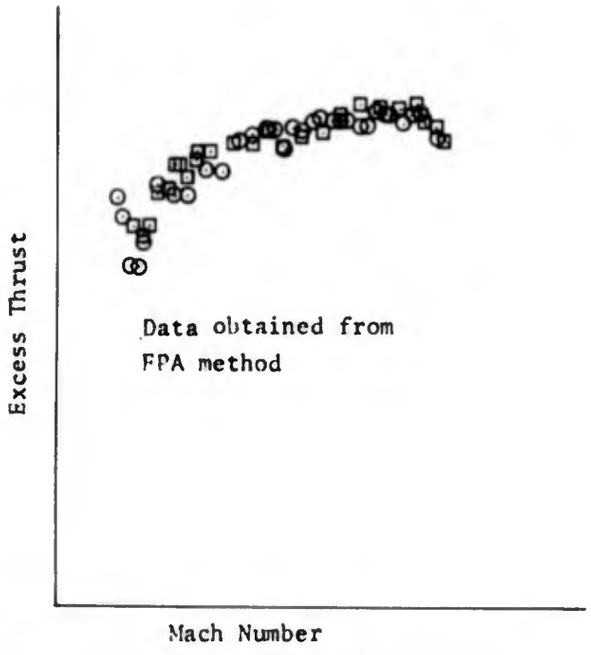
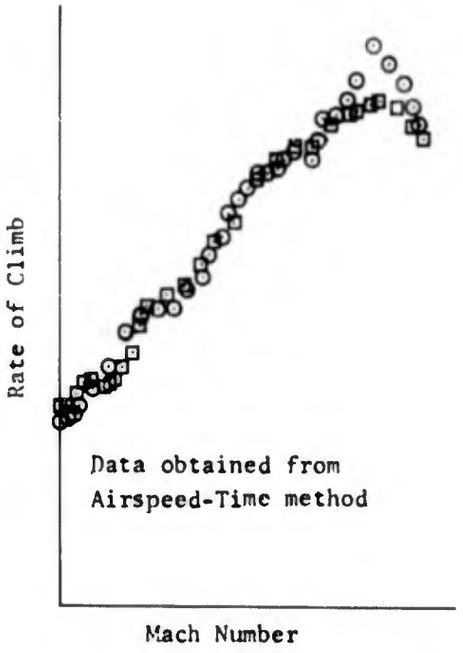


Figure 6. Excess Thrust



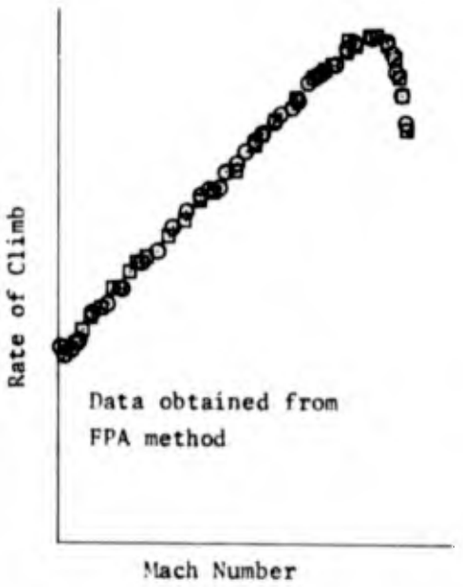


Figure 7. Rate of Climb Potential

(U) APPLICATION OF NUCLEONIC MASS MEASUREMENT TECHNIQUES FOR MEASUREMENT OF TWO-PHASE FLUID CONDITIONS EXISTING IN AIRCRAFT FLUID SYSTEMS

by

Harry W. Schmidt, GS-12

Systems Engineering Group
Research & Technology Division
Wright-Patterson AFB, Ohio

Harry W. Schmidt

BIOGRAPHY

Harry W. Schmidt

Mr. Schmidt is an Aerospace Engineer assigned to the Systems Engineering Group, Wright-Patterson Air Force Base, Ohio. He received a BS Degree in Mechanical Engineering at the University of Wisconsin in 1960. Since his graduation, Mr. Schmidt has been employed by the Air Force at Wright-Patterson Air Force Base. Since his initial assignment he has served as project engineer on various projects including the development of new ice detection techniques for aircraft surfaces, cryogenic pump developments, and general aircraft fuel system projects. He is presently assigned to the Directorate of Propulsion and Power Subsystems Engineering of the Systems Engineering Group as a fuel system project engineer, primarily in support of military cargo type aircraft.

FOREWORD

This paper describes the application of nucleonic mass measurement techniques to the measurement of two-phase flow conditions existing in aircraft fluid systems. Although the paper primarily describes oil aeration measurement, the techniques involved are applicable to aircraft fuel and hydraulic fluid flow measurements. The discussion of radiation and radiation mass measurement techniques is presented from the viewpoint of application engineering and is not on a physicist's level as commonly associated with nucleonics. Naturally, the success of utilizing proven techniques in new applications depends upon a basic understanding of the techniques. The objective of this paper is to present the success of a nucleonic mass measurement application and to present basic design considerations from a users standpoint in order to promote acceptance and new applications of nucleonics within the Air Force.

ABSTRACT

APPLICATION OF NUCLEONIC MASS MEASUREMENT TECHNIQUES FOR MEASUREMENT OF TWO-PHASE FLUID CONDITIONS EXISTING IN AIRCRAFT FLUID SYSTEMS (U)

The difficulty of accurately measuring two-phase liquid flow (with air entrainment) in actual aircraft fluid systems has resulted in serious deficiencies in the design and testing of aircraft systems. The need for accurate measurement of air entrained fluid flow was highlighted by critical bearing failures in the gear boxes of T-34 engines. The analysis and investigation which lead to the necessity for accurate measurement of the amount of aeration existing in the T-34 lubrication system is described. An investigation of various measurement techniques revealed the potentiality of utilizing nucleonic mass measurement techniques for aeration detection. A description of the principles of radiation attentuation of radiation in matter, detection techniques, and system accuracy is described. A description of in-house testing of available nucleonic hardware to determine the practicality and accuracy of nucleonic detection techniques for aeration measurement is given. In-house efforts proved the validity of the measurement technique, but proved the available hardware inadequate for actual flight test aeration measurements. The urgency of the T-34 gear box problem necessitated a search for companies capable of supporting an immediate aeration flight test program. The combined efforts of the Air Force and industry which resulted in the modification of industrial nucleonic type equipment for aeration testing is described. A basic description of the equipment and the test results are discussed. The increasing acceptance of nucleonics as a useful instrumentation tool by

the Air Force and aircraft companies in conjunction with the potential applications realized from testing has generated the need for the development of a nucleonic instrument suitable for across-the-board evaluation of aircraft fuel and oil systems. The development goals and applications of the across-the-board type nucleonic instrument are described. The paper describes the anticipated benefits that nucleonic mass measurement technique will have on future design and testing of aircraft fluid flow systems. Specific benefits described include specification updating as a result of the increased knowledge of fluid flow characteristics, and improved assurance that new aircraft designs will meet oil aeration and fuel V/L (vapor-liquid ratio by volume) specification requirements. It is anticipated that the knowledge gained from continued testing with nucleonic mass measurement systems and advanced fuel and oil system requirements will lead to operational in-flight monitoring of fluid flow at critical points.

I. INTRODUCTION

The difficulty of accurately measuring two-phase flow in aircraft fluid systems has resulted in serious deficiencies in the design and testing of aircraft systems. The need for accurate measurement of air entrained in lubricating oil flow was highlighted by critical bearing failures in the reduction gear boxes of T-34 engines. The T-34 turbo-prop engines are installed in C-133 aircraft which are being used by the Military Airlift Command (MAC). The T-34 engine has been plagued with a multitude of gear box failures many of which resulted in extensive engine damage. The T-34 turbo-prop engine has a unique reduction gear box design, in its horsepower class, in that the high speed pinion bearings of the reduction gears are sleeve bearings instead of roller or ball bearings. Extensive redesign efforts, including increasing the size of the bearing and changing the material had not reduced the number of gear box failures.

1.1 Investigation of T-34 Lubrication System

The T-34 engine lube flow system is illustrated in Figure 1. A high pressure pump receives oil from the airframe supply tank and delivers it to the various bearings in the engine and reduction gear box. Oil is collected in sumps located in the engine and gear box where scavenge pumps, as the name implies, continuously scavenge oil and air from the sumps and return the aerated oil to the oil tank. The oil tank must control the amount of air entrainment by deaeration devices and dwell time.

An investigation of the history of the gear box failures revealed one similarity in all failures. The high speed pinion bearings were either completely seized and could not be examined or contained pitting on the bearing surface. Three possible causes for the pitting considered were: (1) corrosion, (2) aeration - the pitting was similar to that found in pumps due to cavitation, (3) a combination of corrosion and aeration. The investigation also revealed that there had never been any tests to demonstrate that the engine lubrication system was capable of meeting aeration specification requirements. The specification requirement is that the maximum aeration in the oil supplied to the engine shall not exceed 10% by volume. Another interesting fact was that the engine manufacturer had not been able to duplicate the gear box failures by testing on the engine test stand. Later, a significant difference between the aircraft lubrication system and the engine test stand was realized: The engine test stand was equipped with a 55 gallon oil tank, whereas the capacity of the aircraft oil tank is 13 gallons.

Air entrainment in lubricating systems had been considered detrimental for two reasons. Both lubricating and heat rejection properties of oil are degraded by the entrainment. "Galling" or smearing of the bearing surface may be the evidence of improper lubrication. High localized temperatures are evidence of poor heat rejection.

T-34 ENGINE LUBRICATION SCHEMATIC

NOTE: THE SCHEMATIC DOES NOT REPRESENT ACTUAL SPATIAL & DIMENSIONAL RELATIONSHIPS BETWEEN COMPONENTS

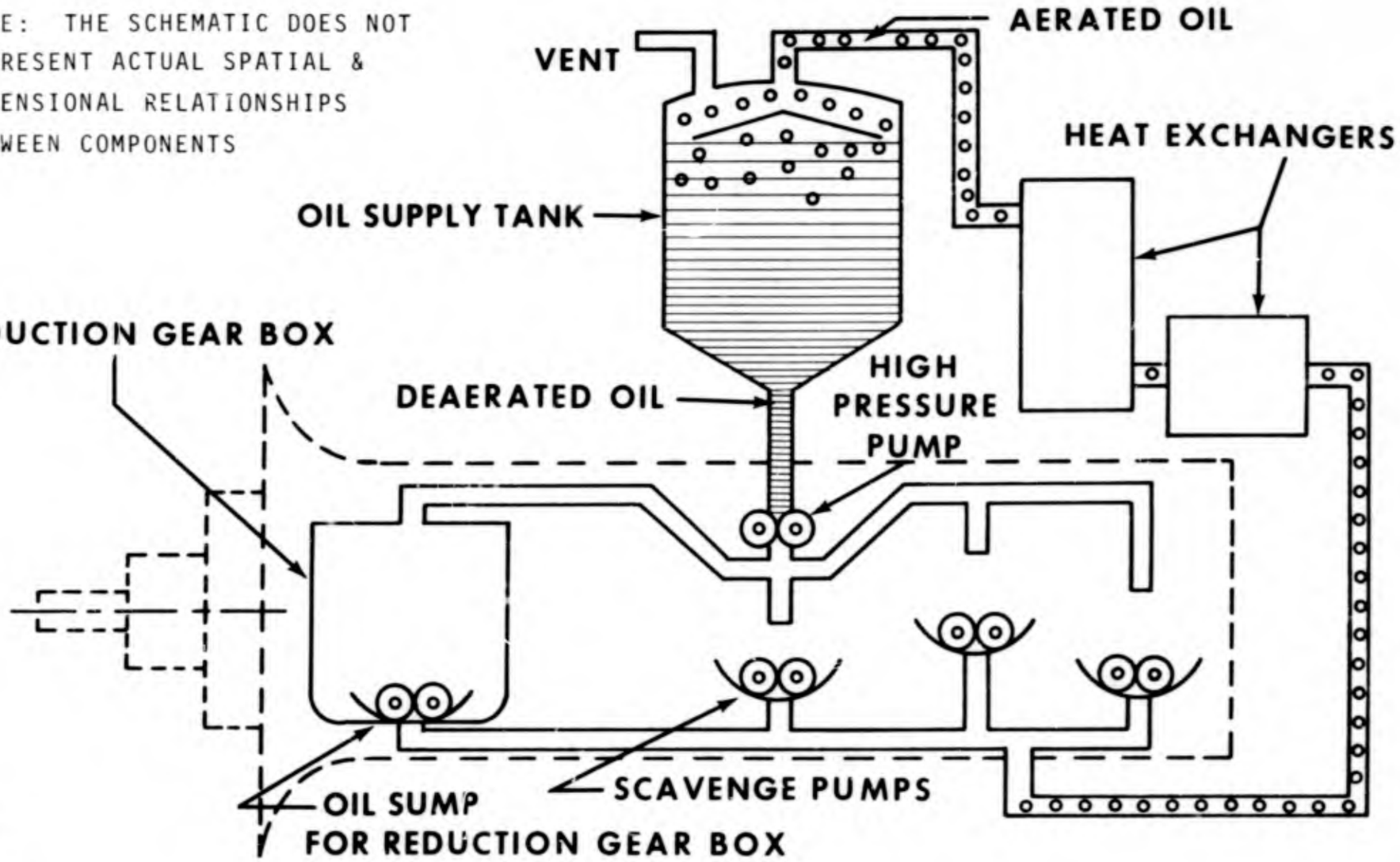


Fig. 1

Since these common signs of excessive aeration had not been present with the failures, the possibility of aeration as a contributing cause to the gear box failures had been discounted. The striking similarity between the sleeve bearing pitting and cavitational type pitting introduced the possibility of a new type of harmful effect of aeration; the harmful effect being an erosive or cavitational type pitting caused by air entrainment and/or air and oil vapor. A sleeve bearing produces a pumping action with high pressure gradients and contains soft bearing materials. These are conditions conducive to pitting by gas (air and oil vapor) entrainment.

II. THE NEED FOR AN ACCURATE AERATION METER

In order to determine the relationship between aeration and the bearing pitting, an accurate measurement meter is required. The minimum requirements for an aeration meter for C-133 testing were as follows:

a. Provide continuous monitoring of the oil flow condition in a 2 inch I.D. oil supply line with a maximum flow rate of 300 ppm.

b. Be capable of functioning during flight testing, without creating a flight hazard.

c. The installation of the meter must not change the characteristics of the fluid to the degree that erroneous aeration readings will result.

d. The minimum accuracy for test purposes shall be $\pm$10% for a gaseous volume fraction range from 0 to .25 (span), i.e., for 10% actual aeration, the allowable range of readout is 9% to 10%.

Previous techniques of measuring oil aeration were investigated and found to be unacceptable. One method was simply to install a sightglass in the oil supply line and estimate the amount of aeration by visual observation and comparison to past experience. The accuracy and limits of the sightglass technique are obvious. The other technique consists of a trapping system, which taps and traps a sample of the oil. The volume of the trapped oil is measured before and after the air has settled out in order to determine the percent aeration. The trapping system certainly is not a continuous monitoring system, it does affect the flow characteristics, the inaccuracy exceeds that desired, and it is not practical for flight testing.

The criticality of the gear box failures led to a search of various mass measurement techniques which could be readily adapted to oil aeration measurement and provide the accuracy needed. The success of using nucleonic mass measurement techniques for tank fluid quantity gaging systems led to the consideration of applying the nucleonic techniques to oil aeration measurements. The basic aeration application is shown in Figure 2.

AERATION MEASUREMENT

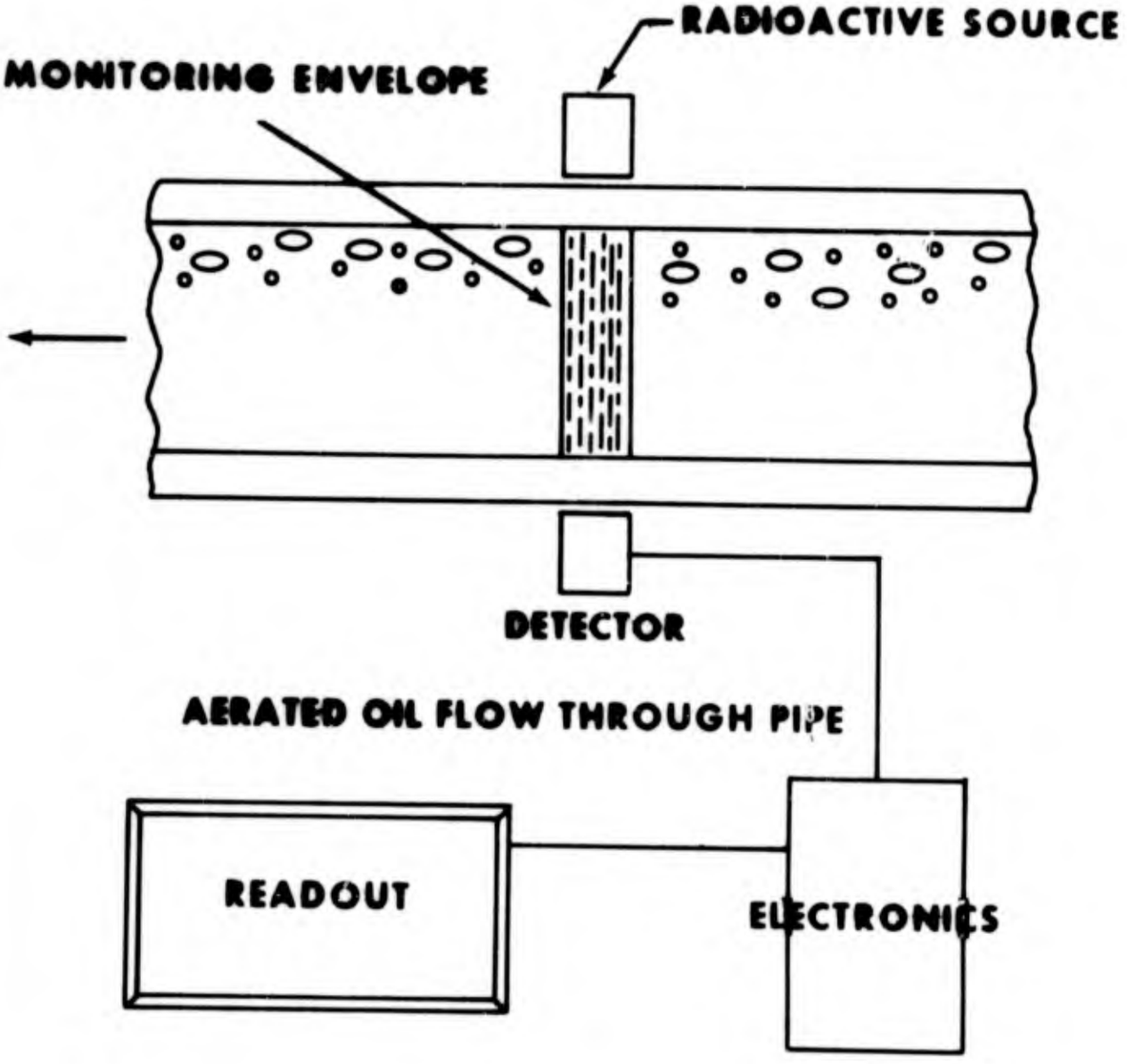


Fig. 2

Referring to Figure 2, the measurement system consists of a radioactive isotope and a detector mounted on the supply line from the oil tank together with the electronics required to record and interpret the radiation energy detected. Basically, a substance, such as the lubricating oil shown in Figure 2, attenuates the radiation energy emitted from a source. As the amount of oil in the pipe increases, the energy counts/time or energy level detected decreases. Since the measurement system is monitoring a constant envelope in the flow path, the oil mass - energy detected relationship may be used to determine the percent oil mass under flow conditions. By knowing the percent oil mass, the percent aeration is known. Note that the system is a gas-to-liquid ratio meter and is not a mass flow meter. Before any nucleonic and design details are considered, it can be seen that good accuracy will depend upon the following general requirements:

a. Good repeatability or uniformity in radiation.

b. An adequate spread in the energy levels detected over the range of aeration to be measured.

c. Fast response to changing flow conditions.

III. PRINCIPLES OF NUCLEONIC MASS MEASUREMENT

In order to successfully use the nucleonic measurement technique for aeration and other two-phase fluid flow measurements, a basic understanding of nucleonics is required. The following areas are discussed in this section.

a. The characteristics of radiation and selection of isotopes.

b. The attenuation of radioactivity in matter.

c. Detection techniques.

d. System accuracy.

3.1 Types of Radiation

Radiation is a process of atomic disintegration of unstable atoms. This radioactive decay is a statistical process. The repetitiveness of energy emissions per a given time can be calculated and the average count rate determined. The radioactive isotopes used in mass measurement systems are normally artificially made in atomic reactors or accelerators.

There are three forms of radiation of interest in the field of mass measurement. They are alpha (α) particles, beta (β) particles, and gamma (γ) rays. Alpha particles have a positive charge and a mass of 2 protons and 2 neutrons (neuclei of helium atom), therefore have very little penetrating power. Alpha particles can be stopped by a thin sheet of paper. Beta particles have a negative charge, but very little mass.

Actually they are high speed electrons; therefore, they have greater penetrating powers than alpha particles, but still can only penetrate very thin sections of metal. When beta particles are used for mass measurement of fluids flowing in pipes, special thin sections or "windows" are required in the pipe to permit adequate penetration of the rays into the fluid. Gamma rays have no charge, no mass, and are a form of photon energy having a high frequency and short wave length. Gamma rays possess relatively high penetrating power. By comparing the design requirements of the aeration meter and the penetrating abilities of the different types of rays, it becomes evident that the use of gamma energy is best suited for the application. Since the design of an aeration meter for the given application would most likely utilize gamma rays, the remaining discussion on nuclear techniques will be limited to gamma radiation.

3.2 Gamma Radiation

Typical gamma ray sources used in mass measurement techniques are shown in Table I.

TABLE I. Typical Sources (Active Nuclei)

Energy Range	Isotope	Energy	Half-Life
Low	American (Am^{241})	60 KEV	470 yrs
Intermediate	Barium (Ba^{133})	357 KEV	10.7 yrs
	Cesium (Cs^{137})	652 KEV	30 yrs
High	Cobalt 60 (Co^{60})	1.17 MEV 1.13 MEV	5.3 yrs

The high energy of Co^{60} versus the small monitoring envelope (2 inch I.D. cross-section of tube) makes Co^{60} an impractical isotope for the application. Gamma radiation is classified as monoenergetic since every gamma ray released from a given isotope has essentially the same level of energy.

It was stated that gamma rays have relatively high penetrating powers compared to alpha and beta particles. From Table I we see that gamma rays exist over a considerable energy level range. The relationship between electron volt (EV) energy and penetrating power can be realized from Figure 3. From the curves it can be seen that a ¼ inch thickness of aluminum stops 12% of 660 KEV γ radiation, and stops 37% of 60 KEV γ radiation. The mono-energetic nature of gamma radiation is helpful in detection techniques. It is evident that energy level, or type of γ radiation source to be used is an important design requirement.

3.3 Passage of Gamma Rays Through Matter

The selection of the gamma ray type radiation and the general

$I/I_o = e^{-\mu\rho x}$

GRAPH SHOWS THE TRANSMISSION COEFFICIENTS FOR 60KEV & 660KEV γ RADIATION AS A FUNCTION OF ALUMINUM THICKNESS

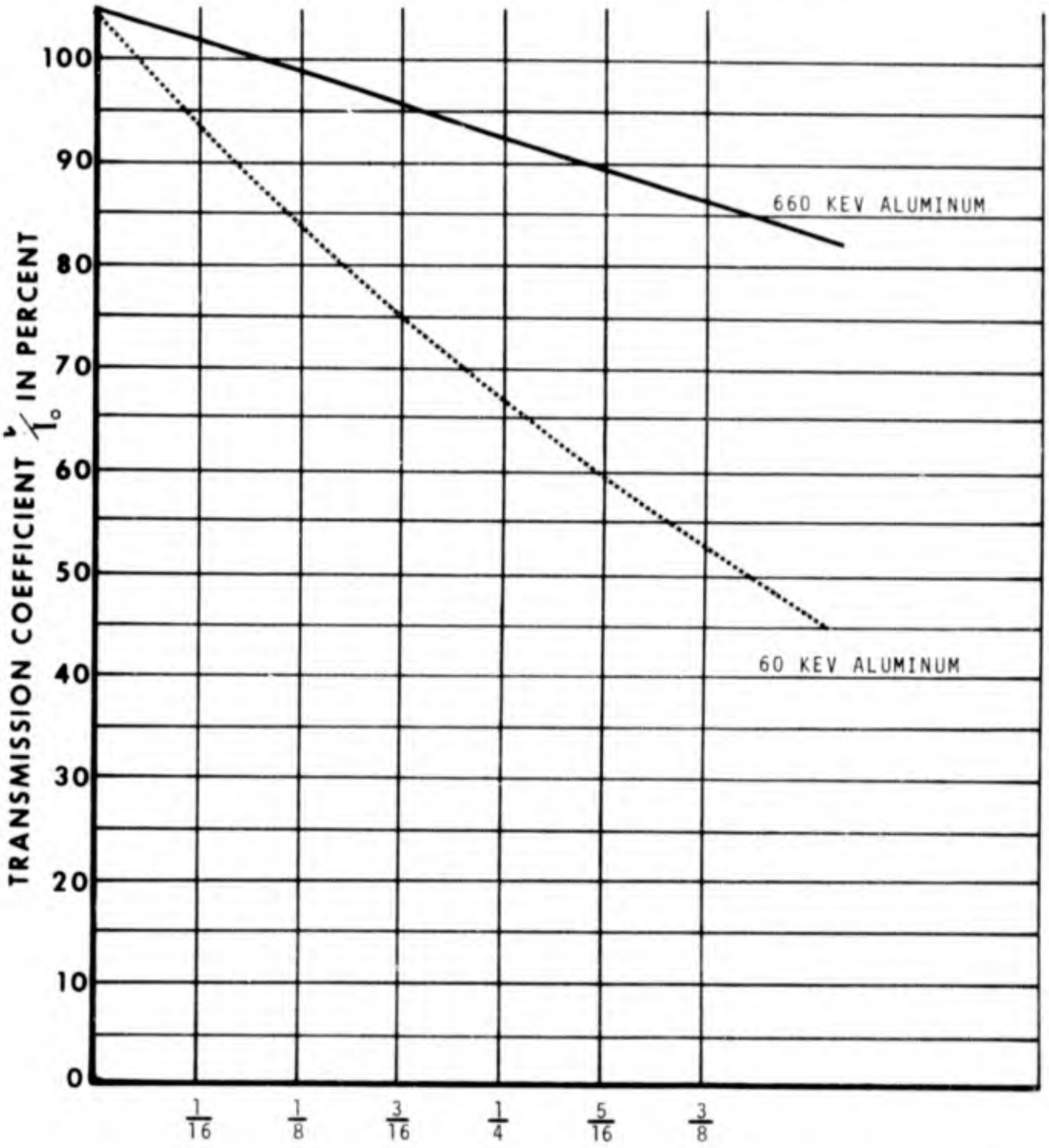


Fig. 3

energy range of low to intermediate (as shown in Table I) brings the design considerations to the step where knowledge of the interaction between gamma rays and matter is required. Figure 4 illustrates the interaction between gamma rays and a hydrocarbon lubrication oil. Only a carbon atom and a hydrogen atom need be shown to illustrate the interactions. The γ source emits gamma rays of energy level $h\nu$. Some rays will not collide with the fluid atoms and will strike the detector. These rays are called primary radiation. For the gamma energy level range (less than 1 MEV) selected there are two main interaction processes. They are: (1) Compton scattering effect and (2) photo-electric absorption. Compton scattering is the process in which a gamma ray collides with an electron and as a result the electron is driven out of its bound position in the atom. The collision deflects the gamma ray and lowers its energy level to $E\gamma'$, which equals the original gamma energy minus the total energy imparted to the electron (orbital binding energy plus energy of electron in ejected state). The deflected gamma ray is known as scattered radiation. The ejected electron in Compton scattering is of such a low energy level that its chance of reaching the detector can be disregarded. In the photo-electric absorption interaction the entire energy of a gamma ray is absorbed by an electron. In this case the ejected electron usually contains considerably more energy (energy of gamma ray - minus the orbital binding energy) than the Compton scattered electron.

Due to the main interactions described, it is evident that there are primarily three categories of radiation which can reach the detector. These energies are: primary gamma radiation, scattered gamma radiation, and electron radiation. Before it can be determined how to utilize the interaction processes in the detection of fluid mass changes, the probabilities of the two main interactions must be known. Figure 5 shows the relationship between the gamma ray energy level and the "Z" (atomic number) of the absorber (in our case hydrocarbon oil). Since hydrocarbons have low "Z" numbers, and as already decided, the energy level of the gamma rays will be less than 1 MEV, it can be seen that the Compton scattering effect is the predominant interaction. Therefore, this interaction becomes the working tool for detection of hydrocarbon oil mass changes.

In order to detect the interaction processes that take place within the fluid, the detector must efficiently detect the radiation reaching the detector. Therefore, a high "Z" number gas filled detector will be used to efficiently stop the radiation (high percent of interactions) in order to measure it. The two basic types of detector systems are analog and digital. The analog type provides a current output proportional to the radiation rays penetrating the detector for a given time period. The digital detector provides an energy pulse for each radiation ray detected.

3.4 Measurement Systems

There are two basic techniques to utilize the Compton

ATTENUATION OF GAMMA RADIATION IN MATTER

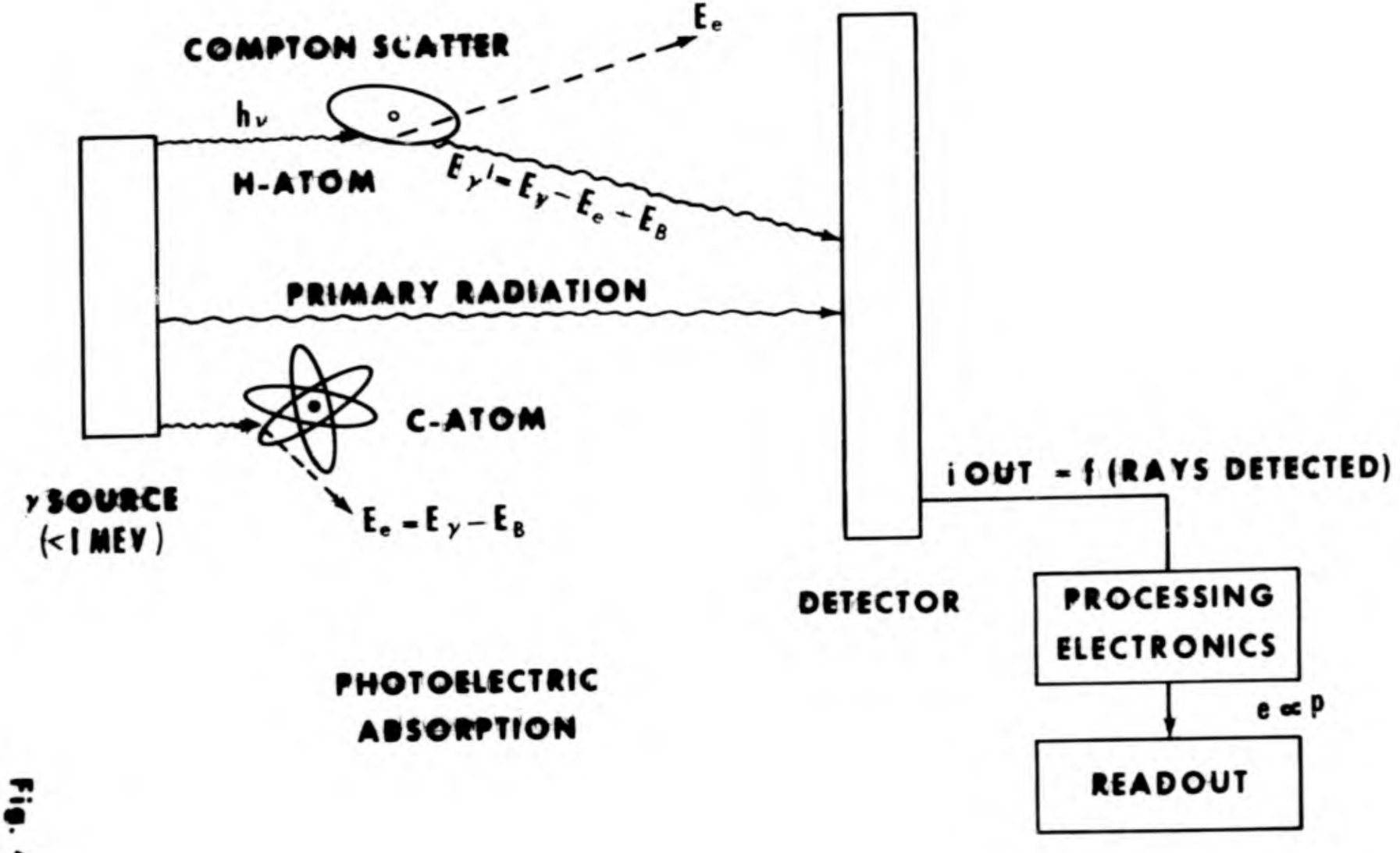


Fig. 4

RELATIVE IMPORTANCE OF PHOTOELECTRIC & COMPTON SCATTERING FOR GAMMA RAY ENERGY BELOW 1 MEV

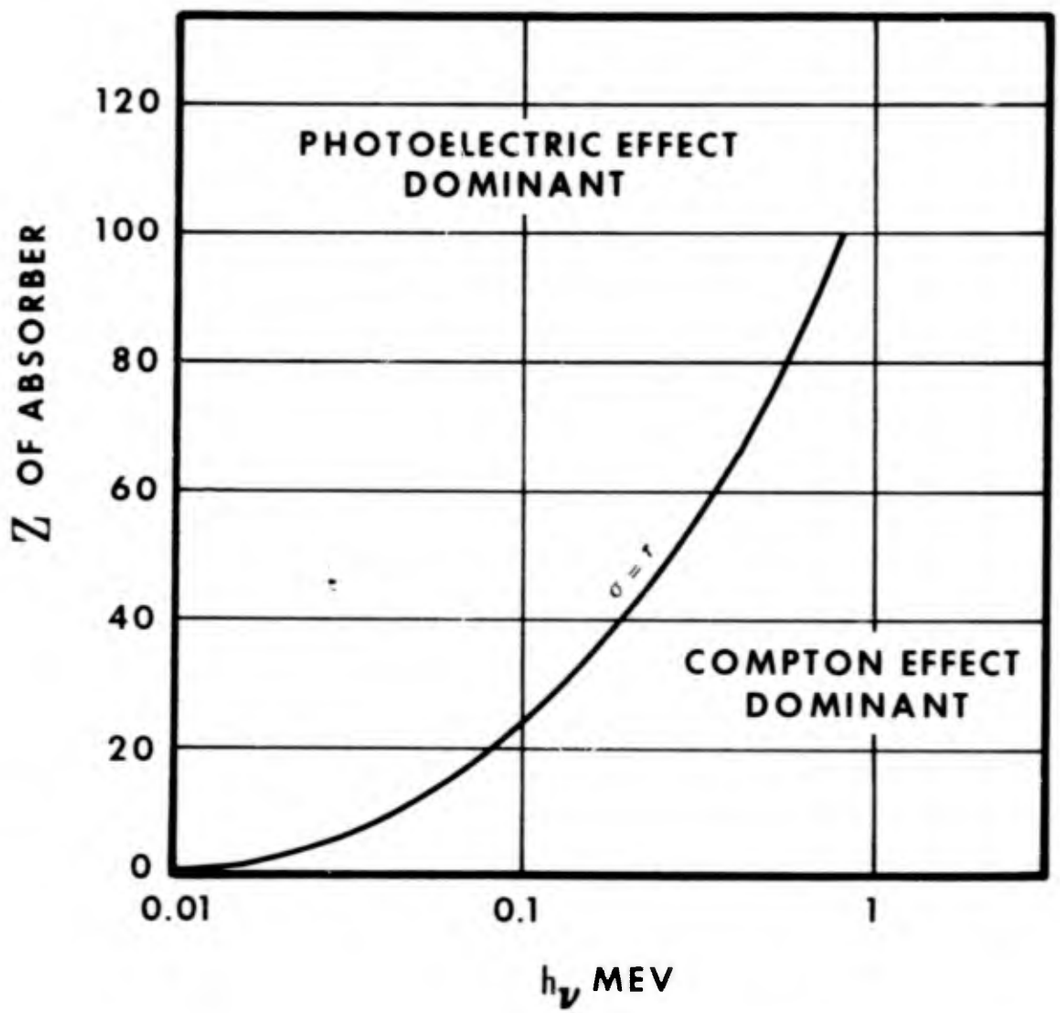


Fig. 5

scatter interaction. They are simply to measure the primary radiation not scattered (transmission technique), or to measure the radiation that is scattered (scattering technique).

The basic transmission equation is: $I_{(T)} = \frac{I_o e^{-\mu\rho X}}{4 \pi X^2}$ (1)

Where: $I_{(T)}$ = unscattered radiation through the path length

I_o = radiation at the source incident to the fluid

μ = the attenuation coefficient which is a constant for the fluid being irridiated by a given source energy. It is a function of the summation of the interaction processes.

ρ = fluid density

X = the radiation path length

e = 2.718

The portion of the equation $e^{-\mu\rho X}$ is the mass attenuation characteristic of radiation. The fact that radiation is directly responsive to mass is an inherent feature that makes a radiation measurement of mass very desirable. The $1/4 \pi X^2$ portion of the equation is the geometric attenuation factor and becomes a constant when source-detector geometry is fixed for a given design. The transmission energy available for mass measurement techniques then becomes:

$$I_T = I_o e^{-\mu\rho X} \qquad \text{(primary radiation)} \qquad (2)$$

Scattered radiation is simply: $I_S = I_o - I_T$ (scattered radiation) (3)

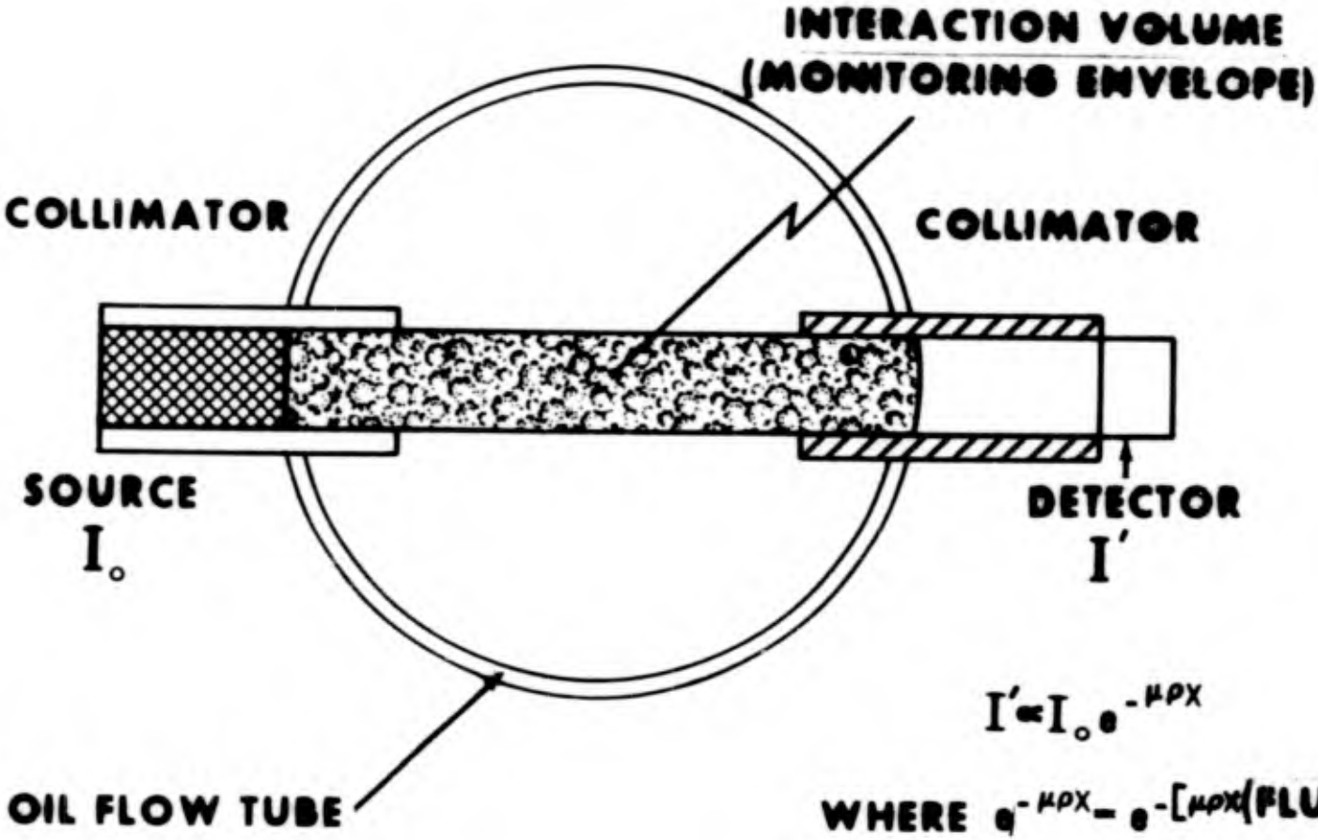
By substitution: $I_S = I_o (1-e^{-\mu\rho X})$ (4)

By comparison of equations (2) and (3) it is realized that the two forms of radiation I_T and I_S are complementary. This implies that one or the other should be measured but not both, yet they offer two distinct detection techniques, both having advantages for specific applications.

3.5 Typical Transmission Measurement

Figure 6 illustrates a simple transmission measurement set-up, and as previously stated, the accuracy of transmission measurement depends upon the ability to measure only primary radiation (equation No. 2). The collimators shown at the source and detector, shield the detector from Compton scattering that results from interactions in both the fluid and the tube wall. The collimators also shield the detector from electron bombardment due to photoelectric absorption, which may be the predominant interaction in the tube wall, especially if it has a high "Z" number, such as stainless steel. Another method of shielding the

TRANSMISSION MEASUREMENT



$$I' \propto I_o e^{-\mu\rho x}$$

WHERE $e^{-\mu\rho x} = e^{-[\mu\rho x(\text{FLUID}) + \mu\rho x(\text{TUBE WALL})]}$

Fig. 6

detector from electron bombardment can be achieved by a thin metallic shield on the detector sensing head, which easily attenuates electron rays. Depending upon the source and detector face geometry the collimators consist of concentric circular or rectangular grids of high "Z" material. The purpose of the high "Z" material is to absorb essentially all of the scattered gamma rays (photoelectric effect) that get to the grid. The accuracy required of the system depends upon the grid spacing. It is obvious that the interaction volume or monitoring envelope shown in Figure 6 would not provide accurate aeration detection due to the large portions of the flow stream cross section that are out of the monitoring envelope. The illustrated inaccuracy may be overcome by installing sufficient source-detector pairs to achieve adequate cross section monitoring, yet maintaining transmission measurement per source detector pair. In the case of multiple pairs the aeration or mass measurement readout would be derived from the summation of the radiation detected. The collimators shown in Figures 6 and 7 do not show the individual grids except for the outer envelope grids.

3.6 Typical Scattering Measurement

Figure 7 illustrates a simple scattering measurement set-up. Again the accuracy depends upon measuring only the wanted rays (scattered) $I \propto I_o (1-e^{-\mu\rho X})$. The detector is positioned to avoid transmission rays, and scattering calculations for the scattering angles permit positioning of the detector in the field of highest scattering intensity. The interaction volume consists of the volume common to both detector and source collimator projected envelopes. The ideal interaction envelope is one that includes all of the fluid cross section but does not extend into the tube wall. Scattering measurement techniques may also require more than one source detector pair.

3.7 Transmission Versus Scattering

A discussion of the design details that are required for the optimum selection of the measurement technique will not be attempted in this paper. The complementary nature of scattering and primary radiation also applies to technique selection in that if good resolution with one technique is difficult to achieve, resolution should be relatively easy to achieve with the other technique. Resolution is a major fact in the selection of the mass measurement technique. For an actual flow cross section to be monitored, the total mass attenuation factor for the cross section is $e^{-[\mu\rho X(\text{Fluid}) + \mu\rho X(\text{Tube Wall})]}$. If the $\mu\rho X$ for the fluid is less than the $\mu\rho X$ for the tube wall, the resolution of the transmission technique will be low and the scattering technique may be the best selection. The resolution (or sensitivity) can be expressed as the count rate or energy level spread over the maximum $\Delta\rho$ to be measured. A factor to consider in the technique selection is the fact that good geometry is difficult to achieve for the scattering technique, and is relatively easy to achieve for the transmission technique.

3.8 Accuracy

The accuracy of detecting incident radiation (as a means of

TYPICAL SCATTERING MEASUREMENT

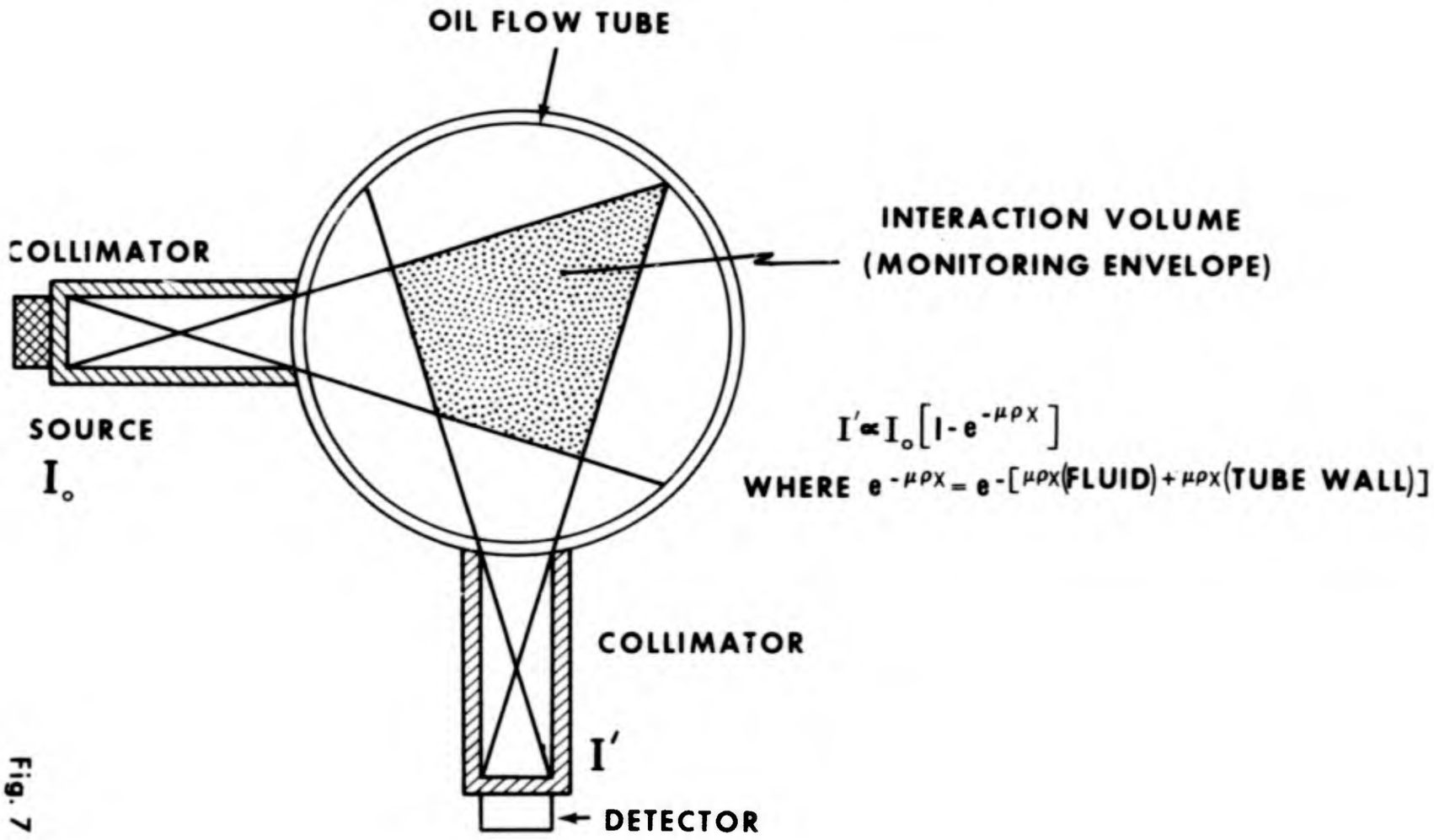


Fig. 7

detecting density changes) depends upon statistical probabilities. It is an accepted fact that for a count rate of N counts/second, the accuracy is $\pm(N)^{\frac{1}{2}}$. Taking the time constant (counting interval) into consideration, the statistical accuracy for the counts sensed (Nt) is $\pm(Nt)^{\frac{1}{2}}$. The percent statistical error then becomes $(Nt)^{\frac{1}{2}}/Nt \times 100\%$. It can be seen that either the radiation intensity incident on the detector or the counting interval (t) may be increased to improve the accuracy. A curve showing the relationship of Nt and % error is shown in Figure 8. The curve shows that excellent statistical accuracies can be achieved; however, both counts/second (N) and counting interval (t) have practical limitations. Let us consider the limitations of t first. For an aeration measurement meter or any flow measurement requirement the counting interval (t) represents the system response to condition transients. For example if a detection system using a two second counting interval for energy level readouts was required to sense a slug of air 50% by volume passing through the monitoring section for a duration of .5 seconds, the detector would erroneously sense the slug of air as approximately 10% by volume existing for two seconds. Increasing the counting interval (t) increase the statistical accuracy, but may decrease system accuracy depending upon the design requirements for response to fluid transients as illustrated in the above example. To illustrate the practical limitations of counts/second (N), assume a hypothetical case where the response to transients requirement and other factors has established a t of one second. The real statistical error required is .20%; however, someone decides to reduce the statistical error to .05%. In Figure 8, for t = one second it can be seen that required radiation intensity at the detector and hence the source strength for .05% error is 16 times the source strength for .20% error. The increase in shielding material weight and volume for human safety is the obvious penalty. In addition to the statistical inaccuracy, there are the common inaccuracies due to non-linearity, environment effects, calibration, etc. The total inaccuracy (σ) equals $(\sigma_1^2 + \sigma_2^2 + \cdots + \sigma_n^2)^{\frac{1}{2}}$ where σ_1, σ_2, etc. are the individual inaccuracies. The state-of-the-art overall accuracy is .5% to 2% for analog type detection systems and approximately .1% for digital detection systems.

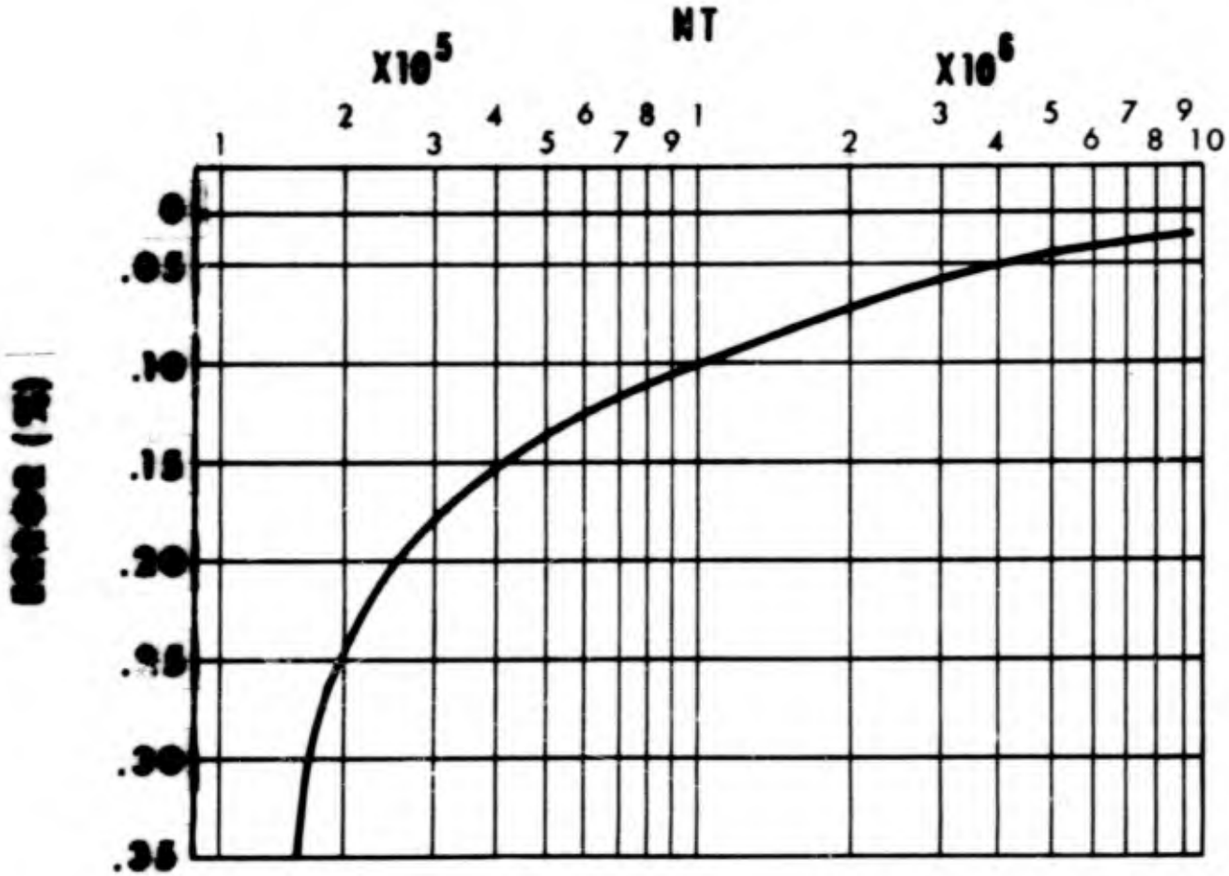
3.9 Calibration of Aeration Meter

In accordance with Air Force policy, any newly developed system requires thorough testing to demonstrate performance. The nucleonic aeration measurement meter requires a rather complex test set-up since the flow stream must consist of known fluid and air ratios in order to verify the validity of the meter read-outs. Fortunately, once the performance of a nucleonic system has been demonstrated, a very simple yet adequate calibration technique may be employed for each aircraft or test stand installation. The simple calibration consists only of establishing the end points of the radiation energy-aeration relationship curve. For a given meter design the shape of the curve will remain the same, so it is merely a matter of adjusting the curve so

STATISTICAL RELATIONSHIP BETWEEN RADIATION COUNTS AND ERROR

N = COUNTS/SEC

T = TIME (SECONDS)



the end points coincide with the full and empty conditions of the flow cross section. A couple of convenient points such as 50% and 75% aeration conditions may be set-up to verify the design accuracy of the system. Installation type calibration may be achieved under static (non-flow) conditions. If the fluid temperature drifts from the calibrated fluid temperature, the aeration readings may easily be adjusted to correct for change in specific gravity of the fluid.

3.10 Human Safety

The safe handling of any form of energy, requires a proper knowledge of the characteristics of the energy and adequate safety rules or regulations. The handling of radiation requires knowledge of the potentially harmful biological effects of radiation and the safety regulations established by the Atomic Energy Commission (AEC). The emphasis of human safety, in peacetime applications of radioactive isotopes, has resulted in an outstanding record of safety. Human safety is a design requirement for nucleonic mass measurement meters to the degree that workers inexperienced with the harmful effects of radiation cannot get a harmful dose of radiation, unless, of course, they disregard warning signs and break into a meter.

IV. UTILIZATION OF NUCLEONIC AERATION METERS

4.1 In-House Testing of Technique

Coincident with the search for an aeration meter were plans for imminent C-133 flight testing. The nucleonic mass measurement technique was selected as the best approach; however, the existence of such a meter was unknown. An in-house effort was expedited in an attempt to verify the validity of the nucleonic approach and to provide flight test equipment for the C-133 flight test scheduled within a few weeks. Available equipment consisted of 100 and 300 micro-Curie amounts of Co^{60} isotopes, Geiger Meuller type detectors, electronic count scalers, and a nucleonic oil quantity gage modified for aeration measurements. The makeshift designs did not provide an adequate aeration meter for flight testing within the available time period, but the effort expended did verify the validity of the measurement technique. The limitations of the makeshift equipment did prove to demonstrate the importance of some of the design considerations previously discussed. Some lessons learned by testing were:

a. Working with Co^{60} demonstrated the importance of proper energy level radiation. As was previously stated the energy of Co^{60} is too great for the application. Poor resolution resulted and attempts to improve resolution by increasing path length "x" degraded the accuracy. In addition, the source intensity (I_o) was too low for good accuracy.

b. The importance of proper geometry was realized. Trial and error source-detector geometrics demonstrated the importance of

monitoring an adequate percent of the events. Events are described as aeration events (location of the bubbles) and nucleonic interaction events. A good geometric design must monitor an adequate portion of the fluid cross section to ensure that entrained air is not escaping detection. Good geometry also requires that the accuracy of radiation measurement will not be affected by the geometric position of the interactions within the interaction volume. A simple method of checking the geometric accuracy of a nucleonic aeration meter may be conducted as follows:

Select a convenient length of the flow pipe with the aeration meter installed on the pipe. The aeration meter should be calibrated. Fill the length of pipe to 50% by volume of oil and cap off the ends. Position the pipe in normal installed position and record aeration readout. Rotate the pipe about its longitudinal axis and record aeration readouts at various angular positions. If the aeration readouts for the actual 50% aeration condition are within design accuracy, the geometric design of the meter is adequate.

4.2 Ground and Flight Testing

A search for companies capable of supporting the immediate aeration test program was conducted. It was discovered that Industrial Nucleonics Corporation, located in Columbus, Ohio, manufacture nucleonic mass measurement systems. The company had demonstrated the practicality and accuracy of the nucleonic mass measurement technique, by a record of several years of successful equipment operation in a wide variety of industrial applications. The outstanding support of Industrial Nucleonics personnel resulted in the modification of one of their industrial nucleonic systems so it could be used as an aeration meter in the immediate test program. Success of the modified meter in both ground and flight tests led to the modification of another system to enable concurrent testing by industry and the Air Force. The details and chronological order of test events is not considered significant for this paper; however, extensive testing has proven that the nucleonic mass measurement techniques for aeration measurement are accurate and practical for ground and flight testing. The significant T-34 engine gear box information obtained by the tests included the following:

a. The most significant information obtained was the verification that aeration is a contributing factor to the bearing surface pitting and subsequent gear box failures. Pratt & Whitney, the engine manufacturer of the T-34 engines, conducted a 150 hour endurance test using the C-133 engine lubrication system, including the 13 gallon oil tank. One of the Industrial Nucleonics modified meters was used in the test to continuously monitor the aeration. The aeration readings were approximately 2½% throughout the sea level test, and would normally be considered well within safe operating limits (10% is specification limit). However, the 150 hours of 2½% aeration resulted in serious

bearing pitting to the extent that bearing seizures were imminent. The bearing pitting had the same appearance as the failed bearing removed during engine overhauls. Previous 150 hour endurance testing identical to the one described above, except with a large test stand oil tank (55 gallons) had never resulted in bearing pitting. Aeration readings with the 55 gallon oil tank were 0%.

b. Testing revealed that aeration can be decreased approximately 50% by lubrication oil with superior anti-foaming and deaeration characteristics.

c. Aeration is so critical in the C-133 engine lubrication system that it essentially has to be non-existent in order to solve the bearing problems. A larger aircraft oil tank with improved deaeration is needed.

d. A small leak in the oil supply line results in high aeration readings due to the fact that the oil supply line pressure is below atmospheric.

e. To date there is an unknown factor, other than apparent leaks in the oil supply line, which causes some engine lubrication systems to aerate excessively. Optimum oil selection and oil tank design cannot protect the bearings from the excessive aeration created by the unknown cause.

f. The fact that such a very small amount of aeration (2½%) can cause bearing pitting, in conjunction with the unknown cause of higher aeration readings in some engines, brings to attention the possibility of the need for a flight instrument that would provide an aeration warning signal and/or continuous aeration recordings for overhaul information.

V. DEVELOPMENT OF AN UNIVERSAL TWO-PHASE FLUID MEASUREMENT METER

There is no substitute for testing under actual operating conditions. To date little is known about the two-phase fluid condition existing in aircraft fuel and oil systems during actual operation. The lack of knowledge has been primarily due to inadequate measurement methods. The fact that the T-34 engine lubrication system had not been subjected to aeration tests is not unique, but actually common practice. The same lack of knowledge applies to fuel two-phase conditions (vapor/liquid ratio) existing in aircraft fuel systems during actual operation. Currently V/L measurements are achieved by an elaborate test set-up, the use of empirical equations, and are conducted only during pump performance testing. Although aircraft are tested to demonstrate climb to altitude performance with boost pumps off, the actual V/L conditions existing are usually not known. The actual amount of air entrainment in hydraulic systems during operating conditions is also generally unknown.

BIBLIOGRAPHY

1. Industrial Nucleonics Corporation, Design and Development of an Isotopic Altimeter Setting Indicator, Part I, Technical, dated 23 March 1966.

2. Industrial Nucleonics Corporation, Cryogenic Density Technology State-of-Art With Nucleonic Techniques, Report dated 20 January 1966.

3. Price, W. J., Nuclear Radiation Detection, McGraw-Hill Book Co. Inc. 1958.

4. Ethesington, H., Editor, Nuclear Engineering Handbook, McGraw-Hill Book Co. Inc., 1958.

5. Raimondi, E. J., Development of a Continuous Nucleonic Oil-Quantity Gaging System, SEG Technical Documentary Report No. SEG TDR 64-45, August 1964.

There is a definite need to develop a nucleonic meter which can be readily adapted to a multitude of aircraft fluid systems, in order, to provide test data for existing problems and to obtain more knowledge about the fluid condition of critical aircraft flow systems. As was pointed out previously, the aeration meters successfully used to date are modified industrial type meters and as a result are too large for most aircraft installations and have certain limitations due to the fact that they were not designed for aircraft use. In addition to the general specification requirements for accuracy, performance, and resistance to environmental conditions, there are two unique requirements for the universal type meter:

a. A "clamp-on" system to enable measurement of two-phase conditions of fluid flow in a variety of pipe sizes, without "breaking" into the fluid system to be monitored.

b. The capability of measuring fuel V/L conditions and oil aeration with the same unit by calibration adjustments only.

VI. CONCLUSION

The practicality and accuracy of nucleonic mass measurement techniques for oil aeration detection have been well proven. Both the Air Force and industry have accepted the nucleonic aeration meter as standard engine test stand equipment. It is anticipated that continued use of nucleonic measurement techniques will provide valuable information of the two-phase fluid conditions existing in critical aircraft flow systems. The capabilities of nucleonic mass measurement will result in updating specification requirements to insure that new aircraft fluid flow systems actually meet two-phase fluid requirements. The nucleonic mass measurement technique may be applied to the development of economical and accurate cockpit instrumentation for the monitoring of critical flow systems for existing and advanced aircraft.

BLANK PAGE

THIN-FILM HEAT TRANSFER GAGES (U)

by

John W. Frye, Jr.
Captain, USAF

and

Milton E. Franke
Associate Professor

Department of Mechanical Engineering
Air Force Institute of Technology
Air University
Wright-Patterson AFB, Ohio

Capt. John W. Frye, Jr.

BIOGRAPHY

John William Frye, Jr. was born 29 December 1934 in Salisbury, North Carolina. He attended North Carolina State College of Agriculture and Engineering, Raleigh, North Carolina, and in May 1958 he was graduated with honors, receiving a degree of Bachelor of Mechanical Engineering, Aeronautical Option. A Distinguished Military Graduate of the Air Force R.O.T.C. program, he was commissioned a Second Lieutenant in the USAF Reserves. From college he went to work as an apprentice engineer for Lockheed Aircraft Corporation, Marietta, Georgia. He was subsequently called to active duty in March 1959. Since coming on active duty, he has attended pilots school, Instructor Pilots School, Instrument Instructor Pilots School. Navy Instructor Pilots School, Academic Instructors School, and Programmed Instruction School. When at his assigned base, Laredo AFB, Texas, Captain Frye performed instructor pilot duties; taught Aerodynamics and Flight Instruments in the classroom; and wrote a textbook on flight instruments for the Air Training Command. Captain Frye also participated as a crew member in a fourteen day simulated Gemini flight conducted at Brooks Aerospace Medical Center, San Antonio, Texas. Captain Frye received a Master of Science degree in Aerospace Mechanical Engineering from the Air Force Institute of Technology, Wright-Patterson AFB, Ohio, March 1966.

Milton E. Franke

BIOGRAPHY

Milton Eugene Franke was born on 7 April 1931 in Springfield, Illinois. In June 1952 he was graduated from the University of Florida with the degree of Bachelor of Mechanical Engineering with Honors. At the same time he was designated a Distinguished Military Graduate of the AFROTC and commissioned a Second Lieutenant in the Air Force Reserves. After working as a Graduate Engineer at Westinghouse, he was called to active duty by the Air Force in September 1952 and sent to graduate school. In March 1954 he received the degree of Master of Science in Mechanical Engineering from the University of Minnesota. He then was stationed at Wright-Patterson AFB and employed as a project engineer and later as a senior project engineer in the Propulsion Laboratory. From 1957 to 1959 he was a Research Engineer with the duPont Company. Since 1959 he has been a member of the faculty at the Air Force Institute of Technology, and presently he is Associate Professor of Mechanical Engineering. He is a member of Sigma Tau, Tau Beta Pi, American Society of Mechanical Engineers, and American Society of Engineering Education. He has completed all requirements, except dissertation, for the PhD degree at The Ohio State University.

BLANK PAGE

THIN-FILM HEAT TRANSFER GAGES (U)

by

John W. Frye, Jr., Capt, USAF
and
Milton E. Franke, Associate Professor

ABSTRACT

Current Air Force applications require considerable research in the field of gas dynamics. The shock tube and other pulsed flow generators provide the means of simulating many of the regions of interest. Unfortunately, pulsed flow test facilities require special rapid response instrumentation due to the very brief testing times.

For more than a decade, fast response pressure transducers and high speed photography have made gas dynamic studies in devices such as shock tubes relatively routine. More recently, projected high altitude and high Mach number regimes have specifically focused attention on the pulsed flow generator, since the desired regimes can currently be simulated no other way. Two of the main areas of interest in this flight regime are surface temperatures and rates of heat transfer. Thin-film resistance type gages (thermometers) are sensitive and rapid enough to meet experimental needs in both of these areas. Surface coatings have extended the thin-film gage into the realm of ionized flow.

This paper discusses methods and techniques for fabricating, calibrating, and using thin-film gages. The gages were fabricated using liquid platinum paint applied to a quartz backing material and were calibrated using the electrical pulse method. The wall surface temperature variation and rates of heat transfer were determined for a station

on the wall of a shock tube for the laminar boundary layer flow condition immediately after shock wave passage for shock Mach numbers between 1.15 and 2.50. The results were correlated with theoretically predicted values and other published experimental data.

NOMENCLATURE

A	Area, in^2 or ft^2
(A)	Ammeter
B_L	Laminar boundary layer heat transfer parameter, $qt^{\frac{1}{2}}$, $Btu/ft^2 sec^{\frac{1}{2}}$
B_L'	$q(t/P_2)^{\frac{1}{2}}$, $Btu/ft^2 sec^{\frac{1}{2}} atm^{\frac{1}{2}}$
C_p	Specific heat at constant pressure for air, Btu/lb$_m$F
c	Specific heat for the backing material, Btu/lb$_m$F
E	Voltage, volt
E_b	Bridge voltage, volt
I	Current, amp
I_o	Gage current, amp
k	Thermal conductivity, Btu/sec ft F
K	Thermal diffusivity $(k/\rho c)$, ft^2/sec
L	Thickness of the film, in
M_s	Shock wave Mach number
P	Pressure, atm
q	Heat transfer rate per unit area, $Btu/ft^2 sec$
q_L	Laminar boundary layer heat transfer rate, $Btu/ft^2 sec$
R	Resistance, ohm
R_o	Room temperature resistance of the gage, ohm

R_2 Bridge resistor (in series with R_o), ohm

R_v Variable resistor, ohm

T Temperature, F or R

T_r Recovery temperature, F or R

t Time, sec

U Velocity, ft/sec

x Distance along wall, in

y Distance normal to the wall, in

α Temperature coefficient of resistance, ohm/ohmF

λ Dummy variable of integration

ρ Density, lb_m/ft^3

μsec Microseconds, 10^{-6}

Subscripts

e refers to velocity in shock wave coordinates

o refers to gage

s refers to shock wave

w refers to the wall or properties evaluated at the wall

wg refers to gas properties evaluated at the wall

1 refers to conditions upstream of the shock wave or resistor R_1

2 refers to conditions behind the shock wave or resistor R_2

BLANK PAGE

INTRODUCTION

In order to simulate the flight regimes of interest today elaborate and extensive ground facilities such as shock tubes and pulsed flow generators are required.

Although the shock tube has been in existence for about sixty-seven years, only since World War II has it been widely used in organizations engaged in research and development. The use of the shock tube grew slowly at first, and the slow growth was due, in part, to the lack of adequate instrumentation. Instruments normally considered ideal for steady-state gas dynamic studies were found to be grossly inadequate for use in the shock tube since the response times of the instruments were relatively long compared with the shock tube flow time. The improvements in fast response pressure transducers, schlieren equipment, and high speed photography greatly accelerated the use of short duration test devices.

More recently, surface temperature and heat transfer measurements have become practical and relatively routine by use of the thin-film resistance thermometer and the thick resistance thermometer or calorimeter [1-4]* The thin-film gage is basically a thin (1 micron or less) metal film fired onto the surface of an insulator such as quartz or Pyrex. It measures temperatures which are essentially those of the insulator surface. On the other hand, the calorimeter gage has a relatively thick metal film; all of the heat transferred to it is retained essentially by the film itself. For high heat transfer rates, Rose [5] and Hartunian and Varwig [6] have shown

---

*Numbers in brackets designate references at end of paper.

steady-state temperatures to 1000F [24]. Recently, much effort has been directed toward calibration procedures [24-27]. A calibration circuit which gives reported accuracy to $\pm$ 1 percent is described by Seginer et al. [25]. The circuit was also used to calibrate an analog network that electrically provided the heat transfer rate history from the surface temperature history.

The study described herein was concerned with developing instrumentation for measuring wall temperatures and heat transfer rates in the AFIT shock tube. The objective of the study was to design, fabricate, and calibrate a set of thin-film gages; to assemble the associated equipment required to operate the gages; and to make representative velocity, temperature, and heat transfer measurements.

GAGE CONSTRUCTION AND OPERATING CIRCUIT

Gage Construction

A photograph of the thin-film gage designed for use in this study is shown in Figure 1. The gage consisted of a platinum film fired onto a $\frac{1}{2}$-inch diameter, 1/8-inch thick disk of polished quartz. The quartz disk was glued to a lucite plug, which, in turn, was mounted in a brass adapter for positioning in the shock tube. The installation details are described in the Apparatus and Procedure section.

There are three basic methods for applying the film to the insulator: painting, using a brush or drafting pen [9, 12]; vacuum evaporation [10]; and glow discharge sputtering [7, 11]. Although the evaporation and sputtering techniques probably provide the most uniform films, painting is the simplest technique and requires less complicated equipment and procedures. For this reason, the gages used in this study were deposited on the

that the calorimeter gage has somewhat of an advantage since, for large temperature changes, large corrections are required to account for changes in the thermal properties of the thin film and the insulator.

These gages operate on the basic principle that a change in temperature causes a change in electrical resistance. Thus, the gages are connected electrically so that, under approximately constant current conditions, a change in temperature will generate a voltage signal which can be measured. Temperature changes and heat transfer rates are determined by use of appropriate equations. Due to their extremely fast response times (less than one microsecond) [7], thin-film gages are also well suited for determining shock wave velocity.

The development of the thin-film gage has been rather rapid. In 1953 the Cornell Aeronautical Laboratory began an internal research program to develop a fast response resistance thermometer [8], and in 1954 the Air Force sponsored a continuation of the program. The results of this program, reported by Vidal [9], established the feasibility of using thin-film gages in shock tunnels. In addition, Vidal presented a rather complete summary of the theory, construction details, operation, and performance of the gages. In the same time period Chabai and Emrich [10] designed a thin gold foil gage, and, shortly thereafter, Rabinowicz et al. [7, 11] developed a platinum gage. In 1959 Taylor [12] described the characteristics of a gage that was used in a combustion-driven shock tube. Since 1959 numerous studies have been conducted on thin-film gages and their applications [13-21]. The gages have been used effectively for studies of high temperature reactions and kinetics [22], and, with suitable coatings, they can be operated under ionizing flow conditions [23] and at

polished quartz substrate using a drawing pen. The gages made in this manner were found to operate satisfactorily so that no other fabrication procedure was used. In fact, some investigators [2, 9] have preferred the painted film since they found that the evaporated film did not adhere or bond satisfactorily to the insulator surface.

After the film was applied, the disk was heated to 700F in a ventilated oven to burn off the metal carrier and then baked at 1300F for about five hours. Copper lead wires, which were inserted through small holes drilled in the quartz disk, were attached to the film using conducting silver epoxy, Figure 2. This method of attachment was found to provide a reliable means of supporting and electrically connecting the lead wire.

Details of the gage construction are given in Appendix I since the simplicity of the process is misleading. The successful and accurate fabrication of a thin-film gage requires patience, practice, and strict adherence to a few simple, but required, procedural steps.

Typically, the films were approximately 1/16-inch wide and 5/16-inch long with a probable thickness of less than one micron. Although the film thickness was not measured, Vidal [9] indicated that Hanovia Liquid Bright Platinum 05-X when painted and baked on glass can be expected to have a thickness on the order of 0.1 micron. Gage resistances were typically 40 to 50 ohms. The thermal properties of the gage are discussed in the Calibration and the Results and Discussion sections.

Operating Bridge Circuit

Although a number of circuit arrangements are possible, the one used in this study is shown in Figure 3. By following the usual bridge circuit analysis [12, 24, 25] and referring to Figure 3, the current through the

gage is

$$I_o = \frac{E_b}{R_2 + R_o} \tag{1}$$

Under balanced conditions, $\Delta E = 0$, and

$$E_1 = E_2 = \frac{R_1 E_b}{R_1 + R_v} = \frac{R_2 E_b}{R_2 + R_o} \tag{2}$$

Changes in gage resistance will cause the bridge to be unbalanced. Assuming that $\Delta R/R_o$ and $\Delta I/I_o$ are small, the bridge unbalance can be approximated by

$$\Delta E(t) = \frac{E_b R_2 \Delta R}{[R_2 + R_o]^2} \tag{3}$$

By substituting for E_b from equation (1), the unbalance can be expressed as

$$\Delta E(t) = \frac{I_o R_2 \Delta R}{R_2 + R_o} \tag{4}$$

Equation (4) can then be rearranged to give

$$\Delta R = \frac{[R_2 + R_o] \Delta E}{R_2 I_o} \tag{5}$$

For small temperature changes, a linear relation between temperature and resistance such as

$$R = R_o [1 + \alpha (T - T_o)] \tag{6}$$

can be assumed without loss of accuracy.

Thus,

$$\Delta R = \alpha R_o \Delta T \tag{7}$$

than 100 microseconds. This is intuitively acceptable because the thermal conductivity of the film is much greater than that of the quartz. With this assumption the gage may be treated as a homogeneous solid, and the solution for the surface temperature is [9, 12, 28]

$$T(t) = \frac{1}{(\pi k \rho c)^{\frac{1}{2}}} \int_0^t \frac{q(\lambda)}{(t - \lambda)^{\frac{1}{2}}} d\lambda \qquad (10)$$

which, for a constant rate of heat transfer q, integrates to

$$T(t) = \frac{2qt^{\frac{1}{2}}}{(\pi k \rho c)^{\frac{1}{2}}} + T(0) \qquad (11)$$

After rearranging,

$$q = \frac{(\pi k \rho c)^{\frac{1}{2}} \Delta T(t)}{2t^{\frac{1}{2}}} \qquad (12)$$

By substituting for $\Delta T(t)$ from equation (8), the heat transfer can be specified in terms of the voltage signal as

$$q = \frac{(\pi k \rho c)^{\frac{1}{2}}}{2\alpha} \left[\frac{R_2 + R_o}{R_2 R_o I_o} \right] \frac{\Delta E}{t^{\frac{1}{2}}} \qquad (13)$$

The equations for the case where q is not constant, which occurs during the first few microseconds following passage of the shock wave, are given by Vidal [9] and Taylor [12].

Boundary Layer Theory

The theory of the heat transfer and the boundary layer development along the wall of a shock tube has been presented by Rott and Hartunian [29], Mirels [30], and Bershader and Allport [31]. Hartunian et al. [32]

By combining equations (5) and (7), the temperature change of the gage is given by

$$\Delta T = \frac{[R_2 + R_o]\,\Delta E}{\alpha\, R_o\, R_2\, I_o} \tag{8}$$

THEORY

The unsteady heat conduction theory must be considered in order to obtain heat transfer rates from the signals generated by the gages. Furthermore, since gas flows occur in the shock tube, the boundary layer equations are also involved in the theoretical analysis used for correlation with experiment.

Heat Conduction Theory

A one-dimensional analysis is justified because, as previously indicated, the thickness of the film is small compared with the length and width of the film. To further simplify the problem, the insulator may be treated as a semi-infinite solid due to the very short test times in the shock tube [7, 9]. The heat conduction model is shown in Figure 4. The unsteady heat conduction equation

$$\frac{\partial T}{\partial t} = K \frac{\partial^2 T}{\partial y^2} \tag{9}$$

and the appropriate boundary conditions may be specified for regions 1 and 2. A solution to the set of equations has been worked out [9, 12]; however, considerable simplification is obtained if the insulator surface temperature and the mean temperature of the film are assumed to be equal. Vidal [9] has shown that the error is usually less than 10 percent if the films are relatively thin ($\frac{1}{2}$ micron or less) and the test times greater

summarized the theory of laminar and turbulent boundary layers. To illustrate the features involved, the typical boundary layer development and the wall temperature history are sketched in Figure 5. As shown, the unsteady flow condition has been transformed to one of steady flow by use of a shock-fixed coordinate system, i.e., one in which the wall moves at U_s while the shock wave is stationary. In this case, the fluid and the wall have the same velocity upstream of the shock wave, whereas behind the shock wave the fluid and the wall no longer have the same velocity and a boundary layer develops. The boundary layer will exhibit the usual characteristics of laminar, transition, and turbulent flow, and the wall temperature will indicate these flow regimes.

It has been shown [29, 33] that the heat transfer rate through the laminar boundary layer behind a shock wave is given by

$$q_L = B_L \, t^{-\frac{1}{2}} \tag{14}$$

where

$$B_L = (T_r - T_w) \sqrt{\frac{(k\rho C_p)_w}{\pi} \left[1 - (\sqrt{2} - 1) \frac{U_2}{U_s} \right]} \tag{15}$$

The theoretical prediction that the rate varies inversely as the square root of time is the same as predicted by equation (12) for a constant rate of heat conduction into the backing material. Thus, the surface temperature should remain constant when the boundary layer is laminar.

GAGE CALIBRATION

Gage calibration involves measuring α, $(k\rho c)^{\frac{1}{2}}$, and R_0 for use in equations (8) and (13). Although the value of $(k\rho c)^{\frac{1}{2}}$ is approximated by using the property values of the backing material, direct experimental measurement is preferred. The value of $(k\rho c)^{\frac{1}{2}}/\alpha$ is usually obtained by

applying a known constant heat flux to the gage in the form of an electrical pulse.

Two electrical pulse techniques are used: the single pulse method [7, 12] and the more recent double pulse method [16, 24, 27]. The single pulse method offers "in use" calibration. On the other hand, the double pulse method does not require a measurement of the film surface area; it involves measuring $(k\rho c)^{\frac{1}{2}}$ first in air and then immersed in another medium such as distilled water or acetone.

The single pulse method was used in this study and yielded an accuracy and repeatability consistent with the other measured variables. The gage calibration circuit is shown in Figure 6. A step in heat flux was applied to the gage by the discharge of a capacitor which had a relatively long time constant compared with the test time. The voltage unbalance, caused by gage heating and resistance change, was recorded, and the value of $(\pi k\rho c)^{\frac{1}{2}}/2\alpha$ was calculated using equation (13). Details of the gage calibration are given in Appendix II.

Values of α and R_o were obtained using a resistance bridge, galvanometer, and a constant temperature water bath.

APPARATUS AND PROCEDURE

Shock Tube

The AFIT shock tube is a conventionally-designed, variable-length device with a constant 4 by 8 in. cross-sectional area. Instrumentation includes oscilloscopes, pressure transducers, amplifiers, schlieren equipment, time-delay circuits, and trigger circuits.

The air-operated, diaphragm-rupturing probe can be actuated either manually or electrically.

Gage Installation and Instrumentation

Figure 7 shows three gages installed in the upper (side) wall of the shock tube with the heat transfer gage approximately 12 feet from the diaphragm. The gages were flush mounted to the inside surface with the longest dimension of the film normal to the direction of motion of the shock wave. The timer trigger gages, which do not require bridge circuits, were spaced one foot apart and were used to determine shock wave velocity by starting and stopping a time interval meter. The meter, shown in Figure 8, indicated the elapsed time between the pulses generated by the two gages. In order to operate the time meter, the output from the gages was amplified by a 1000X solid-state pulse amplifier, Figure 7, and by a 20X amplifier, Figure 8.

The components of the operating and calibration bridge circuits for the gage used for temperature and heat transfer measurements were assembled on the control panel, Figure 8. An oscilloscope, shown in Figure 8, was used to display the unbalance of either the operating bridge, Figure 3, or the calibration bridge, Figure 6. The oscilloscope was set to single sweep and was triggered by one of the velocity measuring gages. A Polaroid camera was used to record the oscilloscope trace. Just prior to breaking the diaphragm the operating bridge was balanced as closely as possible.

RESULTS AND DISCUSSION

The thermal properties (in special form) and the resistance of the gages were obtained. Velocity, temperature, and heat transfer measurements were made, and the results were correlated with the theoretical predictions and the experimental results of other investigators.

Gage Characteristics

Typical values of α, $(\pi k\rho c)^{\frac{1}{2}}/2$, and R_o are given in Table I.

TABLE I

Thin-Film Gage Set III

GAGE No.	R_o ohm	α ohm/ohmF	$(\pi k\rho c)^{\frac{1}{2}}/2$ Btu/ft^{2}Fsec$^{\frac{1}{2}}$	AREA 10^{-4}ft^2
1	63.76	0.000987	0.067	0.59
2	32.90	0.001758	0.069	0.83
3	47.69	0.000905	0.066	0.57
4	49.73	0.001065	0.071	0.66

R_o depends on room temperature.

The values obtained for $(\pi k\rho c)^{\frac{1}{2}}/2$ agree quite well with the values for Pyrex (0.066 to 0.068 Btu/ft^{2}F sec$^{\frac{1}{2}}$) reported by Hartunian and Varwig [6], Vidal [9], Vidal and Hilton [8], and Bogdan [24]. The maximum deviation from the average was only 0.004 Btu/ft^{2}F sec$^{\frac{1}{2}}$ or about 6 percent, assuming the properties are the same for quartz and Pyrex. It is interesting to note that the calculated value of $(\pi k\rho c)^{\frac{1}{2}}/2$ for Pyrex, based on the values of k, ρ, and c, is also approximately 0.066 Btu/ft^{2}F sec$^{\frac{1}{2}}$.

Velocity Measurements

The results of the shock wave velocity measurements are shown in Figure 9. For velocity measurements, it is advantageous to use very narrow films. In this study the film widths were about 1/16-inch, and the gages were spaced one foot apart. Thus, the separation was large compared with the width, and the measurements should have an accuracy within one percent. The theoretical curve shown in Figure 9 was taken from Egan and Foster [34],

and the experimental results indicate the possibility of shock wave attenuation. The reasons for the differences between experiment and theory are described in detail by Glass [35] and Hall [36].

Temperature and Heat Transfer Measurements

The temperature histories along the shock tube wall were obtained for shock wave Mach numbers between 1.15 and 2.50. A typical oscilloscope trace of the gage response or temperature history is shown in Figure 10. The trace is quite similar to the typical one sketched in Figure 5. All of the data taken in the study exhibited the characteristics illustrated in Figure 10, except for shock Mach numbers less than 1.25. In these traces a small temperature step occurred, but the transition from a laminar to a turbulent boundary layer was not distinguishable.

The results of the laminar boundary layer temperature measurements are shown in Figure 11. The temperature step was calculated using equation (8) and then divided by the square root of the pressure behind the shock wave $P_2^{\frac{1}{2}}$. This eliminated the pressure dependence, since the wall surface temperature rise varies directly as $P_2^{\frac{1}{2}}$ [32]. The results were correlated with the theoretical curve given by Hartunian [32] and indicated a temperature rise consistently lower than that predicted by theory. Nevertheless, the deviation was not significantly different from the experimental results reported by others. However, improvements in the design and components of the gage circuits should improve the sensitivity and balance capability, and probably increase the accuracy of the results.

The laminar boundary layer heat transfer results are shown in Figure 12. To obtain an over-all impression of the data, values of $qt^{\frac{1}{2}}$ were calculated using equation (12) and compared with the curves of B_L given by Weatherston

et al. [33]. This proved to be rather difficult because the data were obtained for a range of pressures. After analyzing the graphs of Weatherston and considering that the temperature rise varies directly as $P_2^{\frac{1}{2}}$, it was found to be again advantageous to divide by $P_2^{\frac{1}{2}}$ and reduce all the B_L curves to a single curve for purposes of comparison. Therefore, a new parameter was defined as

$$B_L' = B_L/P_2^{\frac{1}{2}} = q\,(t/P_2)^{\frac{1}{2}} \tag{16}$$

The experimental results agree quite well for Mach numbers less than about 1.75, but deviate somewhat at higher Mach numbers. A possible explanation is that the runs above M = 1.9 were made with a lower initial driven gas pressure. In this case, the pressure behind the shock wave P_2 was about 0.1 atm, whereas the results reported by Hartunian [32] and Weatherston [33] were for P_2 ranging from 0.2 to 5.0 atm. Recently, Chen and Emrich [37] have shown experimentally that the flow behind a shock wave is not entirely steady and that the laminar velocity profiles deviate somewhat from theoretical predictions. This may explain some of the deviations found in this study; however, the test conditions used by Chen and Emrich were considerably different from those used in this study. Their observations were made at a position 30 cm from the diaphragm in a 1/4 by 3-inch shock tube with the low-pressure end of the tube open to atmosphere. In any case, the results agree qualitatively and are considered reasonably good in view of the rather complicated flow characteristics involved. Improvements to the gage circuits as discussed previously concerning the temperature measurements should provide more accurate data.

The details of the calculations involved have been reported by Frye [38].

Quartz Disks

The quartz disks were cut, drilled, and polished to their final dimensions prior to applying the film. The holes were drilled using an ultrasonic vibration drill. To prevent the drill from breaking out the back surface of the disk, the disks were fastened to a sheet of soda glass with high temperature wax prior to drilling. After drilling, they were thoroughly washed in acetone and then "burned" (just short of melting) with an oxygen rich torch to remove all surface residues. The disks were then handled with tweezers until after the film was applied.

Platinum Fil-

A film of Hanovia Liquid Bright Platinum 05-X was applied to the quartz disk with an old-fashioned dip point (size 1) pen. Since the gage resistance is controlled by the length and cross-sectional area of the film, considerable practice was required in applying the film to obtain the desired resistance and uniformity. The pen was dipped in the liquid platinum and all excess liquid was rubbed off on the bottle lip. The film was then drawn across the disk between the two holes with the aid of a straight edge. A generous coat of the liquid platinum was then applied around the lead attachment holes.

The freshly painted gages were then placed in an unheated oven and raised in temperature to 700F with the door open approximately $\frac{1}{2}$ inch to provide ventilation. The oven door was closed after about an hour when the oven had stabilized at 700F. The temperature was then increased to 1300F and maintained for 4 to 5 hours. Finally, the closed oven was allowed to cool slowly to room temperature. This procedure was quite successful in achieving a good bond between the film and the quartz.

CONCLUSIONS

1. With a reasonable amount of practice, thin-film gages can be manually constructed using liquid platinum paint.
2. The gages appear to have a reasonably long life under the conditions tested since, after more than 100 runs, no film erosion or property charges were detected.
3. The thin-film gages are quite convenient and accurate when used for shock wave velocity measurement.
4. The thin-film gages indicated surface temperature histories under shock tube flow conditions that agreed qualitatively with the theoretically predicted histories and reasonably well with the experimental results of others. Indirectly, the gages indicated the boundary layer characteristics.
5. The gages provided a qualitative indication of the rate of heat transfer to the wall. However, reported heat transfer rates, both theoretical and experimental, must always be used carefully, particularly under shock tube flow conditions.

APPENDIX I

Gage Construction

The purpose of this appendix is to provide a detailed description of the procedures used to construct the thin-film gages used in this study. The process described herein reflects the influence of references and discussions with others who have made and used thin-film gages. Over 100 gages were made before sufficient experience was accumulated to permit rapid, reliable gage construction. The gages were made in sets of five or more and required approximately two days to complete.

Electrical Leads

Several methods were successfully used to attach the leads to the film. The most popular method described in the literature consisted of baking a layer of liquid silver over the ends of the film and then soldering lead wires directly to the cured silver using low temperature solder. However, the slightest overheating during the soldering operation would ruin the gage. For temperatures up to 300F, the epoxy method seemed to be superior for the following reasons:

1) The epoxy firmly anchored the lead wires in the holes.
2) The cured epoxy had a negligible resistance.
3) The application of the epoxy did not involve any critical steps, and mistakes could be washed off with acetone (prior to curing).

Solid copper lead wires were prepared as for conventional soldering. The two parts (body and hardener) of the Epoxy Products E-Solder Nr. 3021 Silver Epoxy were mixed on a glass dish, and the dressed end of each wire was generously coated with epoxy and inserted into one of the holes in the disk. A sufficient amount of epoxy was applied around the hole to assure a good contact with the film. The epoxy was allowed to harden for about two hours at room temperature, and then the gages were placed in the oven. The temperature was raised to 200F and maintained for 3 to 4 hours. This second baking promoted the epoxy hardening and stress relieving of the film. The closed oven was then allowed to cool to room temperature. Occasionally, it was desirable to remove some of the epoxy which was applied around the holes. A piece of masking tape was placed over the film, and the epoxy was removed by filing until the connection was flush with the tape.

APPENDIX II

Gage Calibration

Calibration data for the thin-film gages were obtained using the calibration circuit shown in Figure 6.

Initially, a fixed resistor was placed in the circuit in place of the gage to obtain the bridge response to the capacitor discharge. With the bridge minutely, but intentionally, out of balance ($\pm$ 0.0002 volts), the steady-state response of the bridge was obtained on the oscilloscope and photographed. Without disturbing the circuit, the mercury switch was opened and the capacitor was allowed to charge. Then the mercury switch was closed and the bridge response was again obtained on the oscilloscope and photographed. A typical result is shown in Figure 13, and, as can be observed, an essentially constant voltage step was obtained for approximately the first 1000 μsec.

The above procedure was then repeated with the thin-film gage in the calibrating circuit. Typical results are shown in Figure 14. The traces obtained indicated that the $\Delta E(t)$ variation and, consequently, the surface temperature variation were parabolic functions of time. With the ammeter in the circuit, a slight hook was obtained, whereas with the ammeter switched out of the circuit the hook was not observed. The hook was therefore attributed to the meter. In either case, the quantity $\Delta E/t^{\frac{1}{2}}$ was found to be essentially constant for values of t between 100 and 1000 μsec. A typical plot is shown in Figure 15.

The rate of heat transfer to the gage is the power dissipated divided by the area of the film, i.e.,

$$q = \frac{I^2 R_o}{A} = \frac{E_b^2 R_o}{A(R_g + R_o)^2} \qquad (17)$$

where equation (1) was used to eliminate the current. Equation (17) assumes that the film thickness and width are uniform and that $\Delta R/R_o$ is small. In this study ΔR was approximately 0.5 percent of R_o. By substituting in equation (17) for q from equation (13), using equation (1) and rearranging, one obtains

$$\frac{(\pi k \rho c)^{\frac{1}{2}}}{2\alpha} = \frac{R_2 \, R_o^{\,2} E_b^{\,3}}{A(R_2 + R_o)^4} \, \frac{t^{\frac{1}{2}}}{\Delta E} \tag{18}$$

Values of $(\pi k \rho c)^{\frac{1}{2}}/2\alpha$ were calculated using the values of $\Delta E/t^{\frac{1}{2}}$ obtained typically as indicated in Figure 15. Details of the calculations are given by Frye [38].

ACKNOWLEDGEMENTS

The authors take this opportunity to thank Doctors G. W. Mueller and U. Grimm of the Aeronautical Research Laboratories and Captain R. M. Bowman, AFIT, for their council on thin-film gages and Mr. Richard Brown of the Mechanical Engineering Laboratory, AFIT, for his assistance with the shock tube apparatus.

REFERENCES

1. Rose, P. H., and J. O. Stankevics, "Stagnation-Point Heat-Transfer Measurements in Partially Ionized Air," AIAA Journal, 1, No. 12, 2752-2763 (1963).

2. Rose, P. H., and W. I. Stark, "Stagnation Point Heat-Transfer Measurements in Dissociated Air," Journal of the Aeronautical Sciences, 25, No. 2, 86-97 (1958).

3. Hertzberg, A., "The Application of the Shock Tube to the Study of the Problems of Hypersonic Flight," Jet Propulsion, 26, No. 7, 549-568 (1956).

4. Hall, J. Gordon, and A. Hertzberg, "Recent Advances in Transient Surface Temperature Thermometry," Jet Propulsion, 28, 719-723 (1958).

5. Rose, P. H., "Development of the Calorimeter Heat Transfer Gauge for Use in Shock Tubes," The Review of Scientific Instruments, 29, No. 7, 557-564 (1958).

6. Hartunian, R. A., and R. L. Varwig, "On Thin-Film Heat-Transfer Measurements in Shock Tubes and Shock Tunnels," The Physics of Fluids, 5, No. 2, 169-174 (1962).

7. Rabinowicz, J., M. E. Jessey, and C. A. Bartsch, Resistance Thermometer for Heat Transfer Measurement in a Shock Tube, GALCIT Memorandum No. 33, California Institute of Technology, 1956.

8. Vidal, R. J., and J. H. Hilton, The Construction and Application of a Rapid Response Resistance Thermometer Probe, Cornell Aeronautical Laboratory Report No. IM-1062-A-1, 1956.

9. Vidal, R. J., Model Instrumentation Techniques for Heat Transfer and Force Measurements in a Hypersonic Shock Tunnel, Cornell Aeronautical Lab. Report AD-917-A-1, 1956, WADC TN 56-315, AD 97238.

10. Chabai, Albert J., and R. J. Emrich, "Measurement of Wall Temperature and Heat Flow in the Shock Tube," Journal of Applied Physics, 26, 779-780 (1955).

11. Rabinowicz, J., M. E. Jessey, and C. A. Bartsch, "Resistance Thermometer for Transient High-Temperature Studies," Journal of Applied Physics, 27, 97-98 (1956).

12. Taylor, B. W., Development of a Thin Film Heat-Transfer Gage for Shock Tube Flows, UTIA TN No. 27, Institute of Aerophysics, University of Toronto, 1959, AD 230 721.

13. Fabian, G. J., Hypersonic Research Summary, Cornell Aeronautical Lab. Report No. AD-1118-A-11, 1960, AFOSR TR 60-58, AD 238 152.

14. Collins, D. J., R. Greif, and A. E. Bryson, Jr., "Measurement of the Thermal Conductivity of Helium in the Temperature Range 1600-6700°K," International Journal of Heat and Mass Transfer, 8, 1209-1216 (1965).

15. Friedman, S., and J. A. Fay, "Heat Transfer from Argon and Xenon to the End Wall of a Shock Tube," The Physics of Fluids, 8, No. 11, 1968-1975 (1965).

16. Konopka, W., Heat Transfer Instrumentation for the Grumman Hypersonic Shock Tunnel, Grumman RM-287, Research Dept., Grumman Aircraft Engineering Corp., 1965.

17. Lauver M. R., "Shock Tube Thermal Conductivity," The Physics of Fluids, 7, 611-612 (1965).

18. Nagamatsa, H. T., "Shock Tube Technology and Design," Fundamental Data Obtained from Shock Tube Experiments, Edited by A. Ferri: 86-183, Pergamon Press, New York, 1961.

19. Scagnetti, M., and J. Crabol, "Fast-Response, Platinum-Film Temperature Probes," La Recherche Ae'rospatials, 97, 23-30 (1963).

20. Sturtevant, B., and E. Slachmuylders, "End-Wall Heat-Transfer Effects on the Trajectory of a Reflected Shock Wave," The Physics of Fluids, 7, No. 8, 1201-1207 (1964).

21. Weisblatt, H., and A. Clemente, Development of a 0.005-inch Wide Gold Thin Film Resistance Thermometer Heat Transfer Gage for Use in Shock Tubes, AVCO Tech. Memorandum, Research and Advanced Development Division, AVCO Corporation, 1960, AD 243 225.

22. Bradley, John N., Shock Waves in Chemistry and Physics, John Wiley & Sons Inc., New York, 1962.

23. Marrone, P. V., and R. A. Hartunian, "Thin-Film Thermometer Measurements in Partially Ionized Shock Tube Flows," The Physics of Fluids, 2, 6, 719-721 (1959).

24. Bogdan, L., High Temperature, Thin-Film Resistance Thermometers for Heat Transfer Measurement, NASA CR-26, Cornell Aeronautical Laboratory, Inc., 1964.

25. Seginer, A., A. Cohen, and J. Rom, "Calibration of Thin Film Resistance Thermometers for Heat Flux Measurements in the Shock Tube," Israel Journal of Technology, 3, No. 1, 25-30 (1965).

26. Skinner, G. T., "Calibration of Thin-Film Backing Materials," ARS Journal, 31, No. 5 (1961).

27. Skinner, G. T., A New Method of Calibrating Thin Film Gage Backing Materials, CAL-105, Cornell Aeronautical Laboratory, Inc., 1962, AD 282 409.

28. Carslaw, H. S., and J. C. Jaeger, Conduction of Heat in Solids, Oxford University Press, New York, 1949.

29. Rott, N., and R. Hartunian, On Heat Transfer to the Walls of a Shock Tube, OSR-TN-55-422, Cornell Aeronautical Laboratory, Inc., 1955, AD 79113.

30. Mirels, H., The Wall Boundary Layer Behind a Moving Shock Wave, Presented at the 1957 Boundary Layer Research Symposium, Freiburg, Springer-Verlag, Berlin, 1958.

31. Bershader, D., and J. Allport, On The Laminar Boundary Layer Induced By a Travelling Shock Wave, TR II-22, Princeton University Department of Physics, 1956.

32. Hartunian, R., A. Russo, and P. Marrone, Boundary-Layer Transition and Heat Transfer in Shock Tubes, Cornell Aeronautical Lab. Report No. AD-1118-A-3, 1959, AFOSR TN 59-564, AD 234 728; also 1958 Heat Transfer and Fluid Mechanics Institute, Stanford University Press, 1958.

33. Weatherston, R. C., A. L. Russo, W. E. Smith, and P. V. Marrone, Gasdynamics of a Wave Superheater Facility for Hypersonic Research and Development, AFOSR TN 59-107, Cornell Aeronautical Laboratory, Inc., 1959, AD 210 223.

34. Egan, D. S., and R. A. Foster, Gas Dynamics Research with the Air Force Institute of Technology Shock Tube, AFIT Thesis, Wright-Patterson AFB, Ohio: Air Force Institute of Technology, 1956.

35. Glass, I. I., Shock Tubes, Part I: Theory and Performance of Simple Shock Tubes, UTIA Review No. 12, Institute of Aerophysics, University of Toronto, Toronto, Canada, 1958.

36. Hall, J. Gordon, Shock Tubes, Part II: Production of Strong Shock Waves; Shock Tube Applications, Design and Instrumentation, UTIA Review No. 12, Institute of Aerophysics, University of Toronto, Toronto, Canada, 1958.

37. Chen, C. J., and R. J. Emrich, "Investigation of the Shock-Tube Boundary Layer by a Tracer Method," The Physics of Fluids, 6, No. 1, 1-9 (1963).

38. Frye, J. W., Jr., Thin-Film Heat Transfer Gages, MS Thesis, GAM/ME/66A-3, Wright-Patterson Air Force Base, Ohio: Air Force Institute of Technology, 1966.

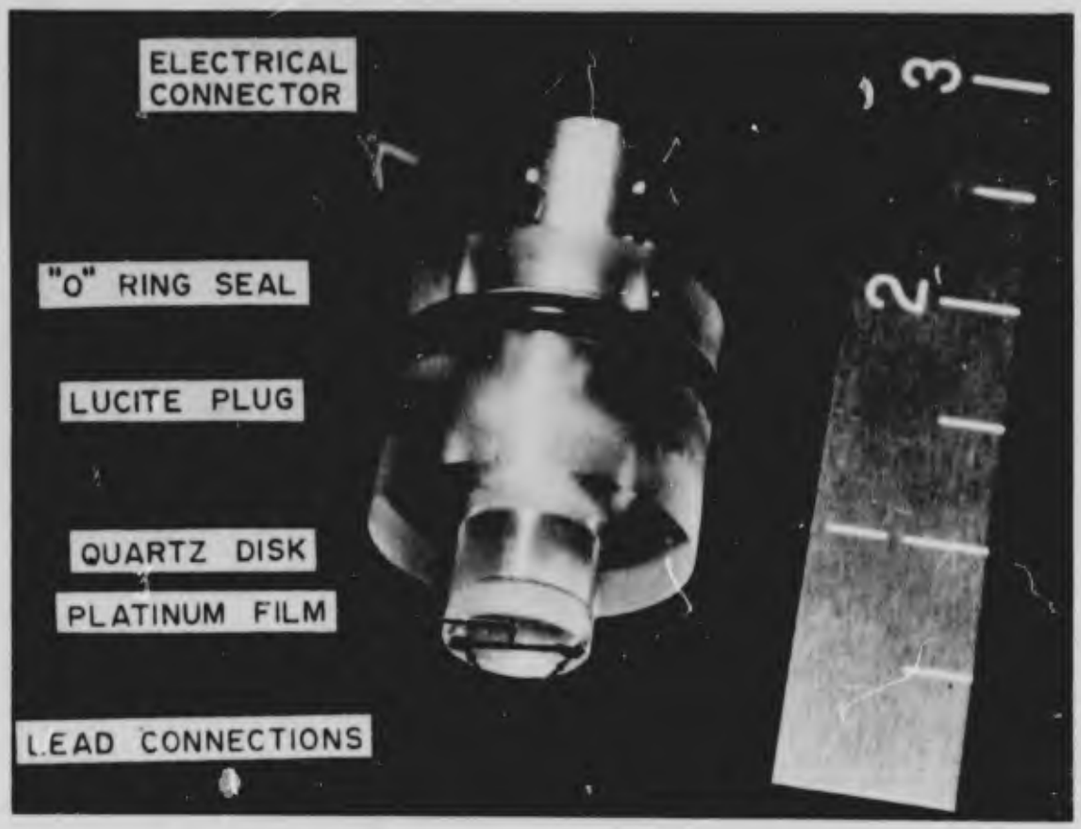


Figure 1 A Typical Thin-Film Gage

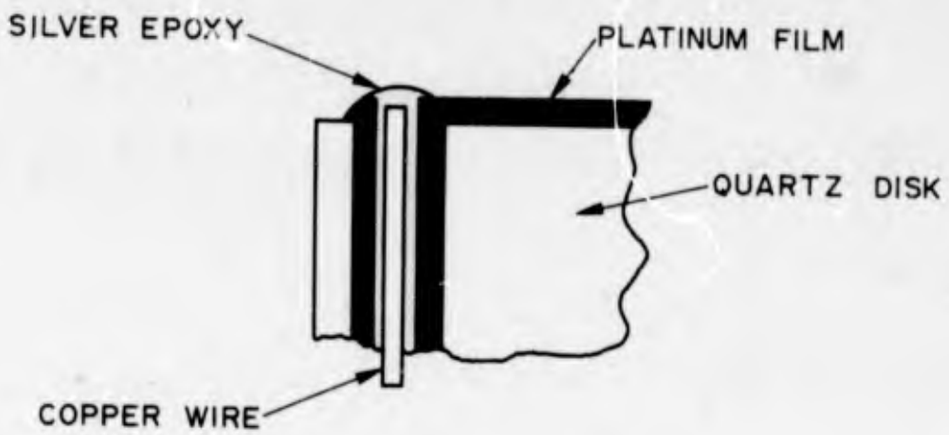


Figure 2 Schematic of the Electrical Lead Attachment

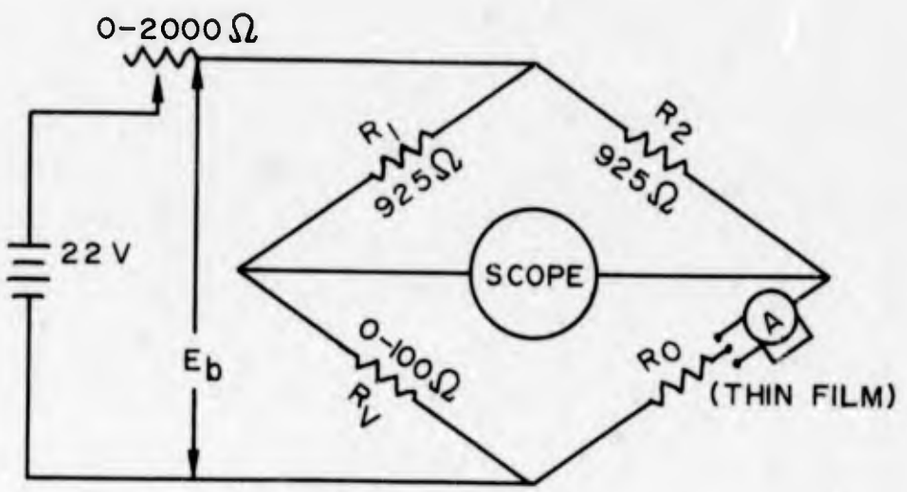


Figure 3 Gage Operating Circuit

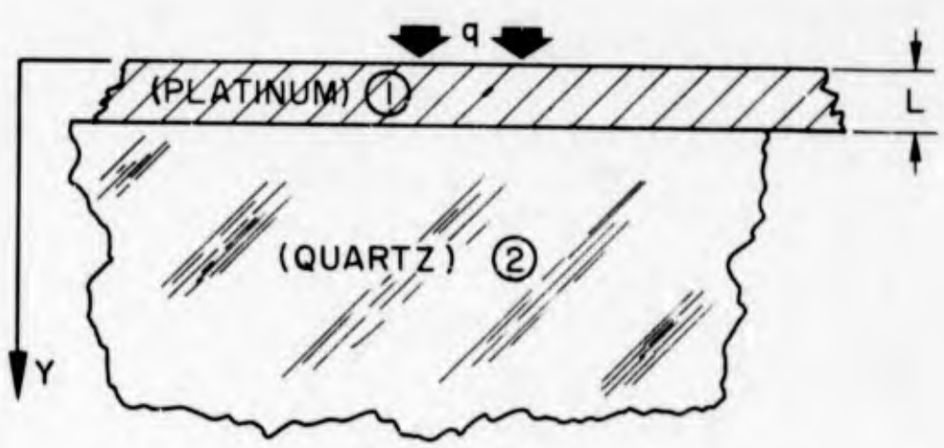


Figure 4 Heat Conduction Model

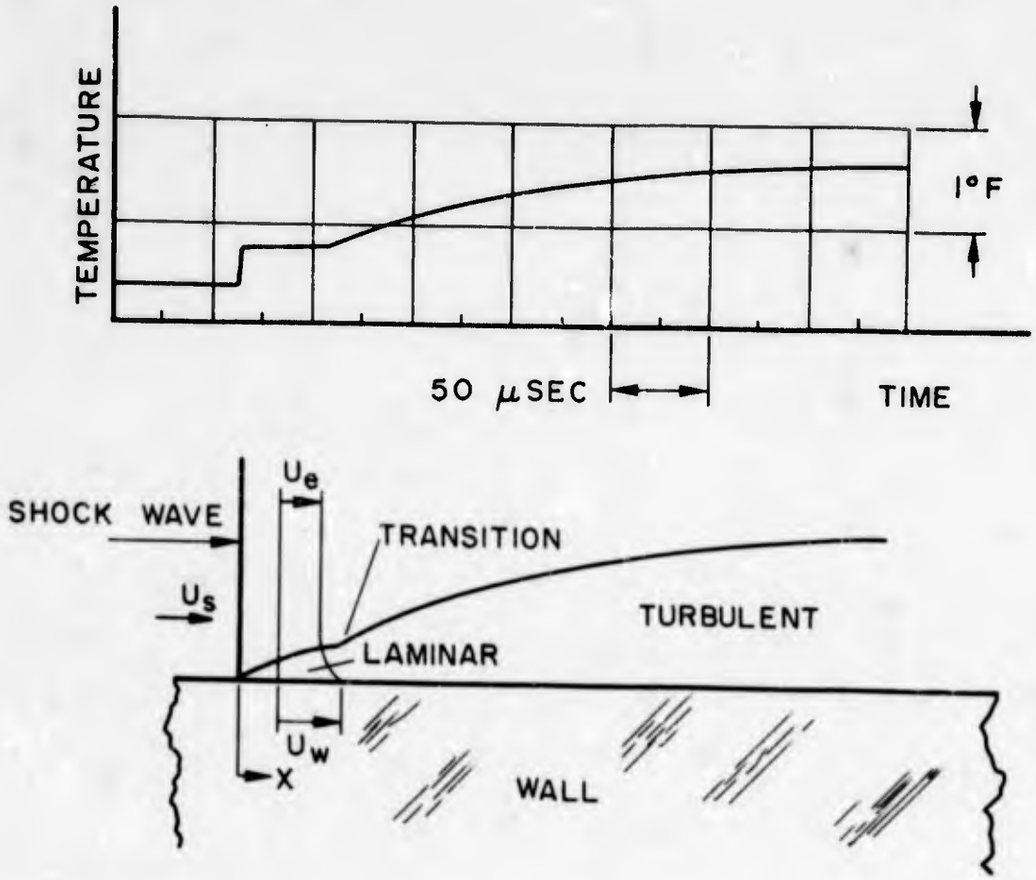


Figure 5 Typical Boundary Layer Development in Shock-Fixed Coordinates and the Corresponding Wall Temperature History Trace for $M_s = 1.77$, $P_1 = 0.15$ atm.

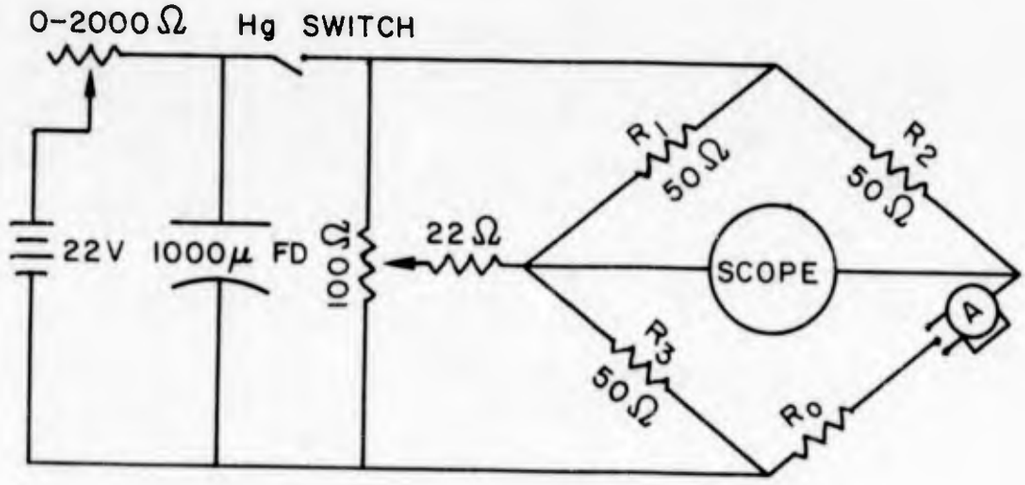


Figure 6 Gage Calibration Circuit

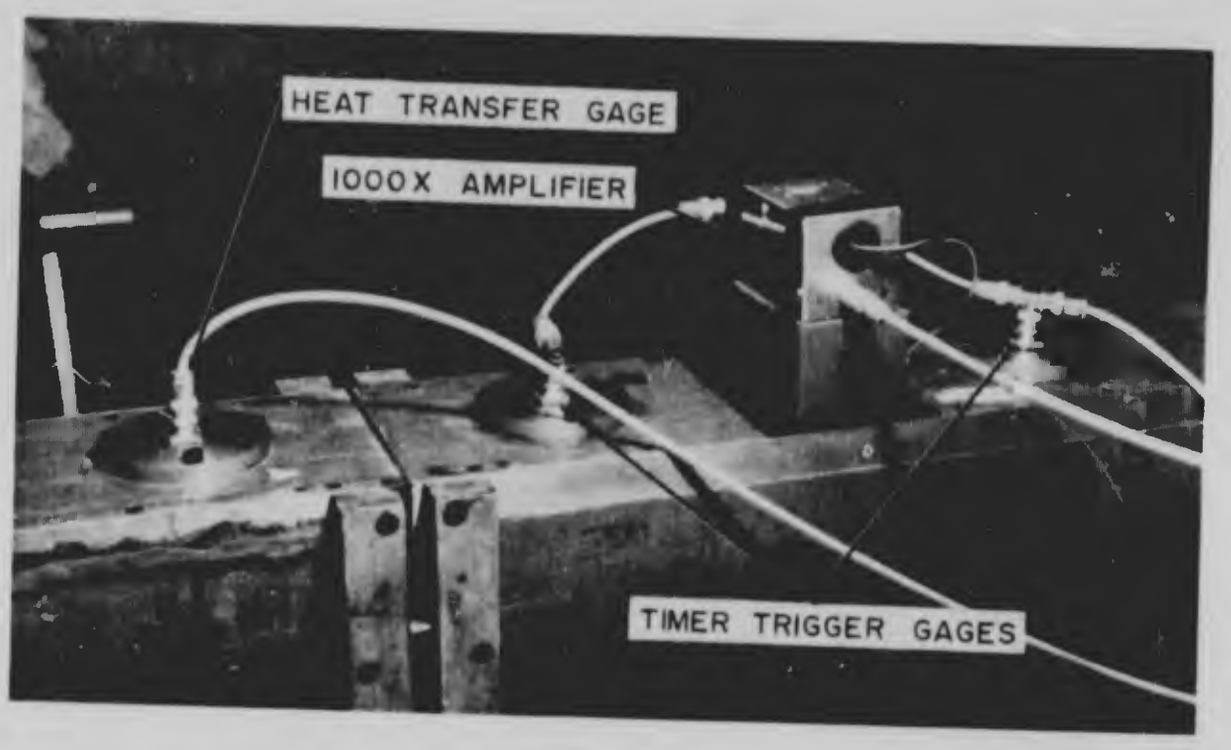


Figure 7 Side Wall Gage Installation

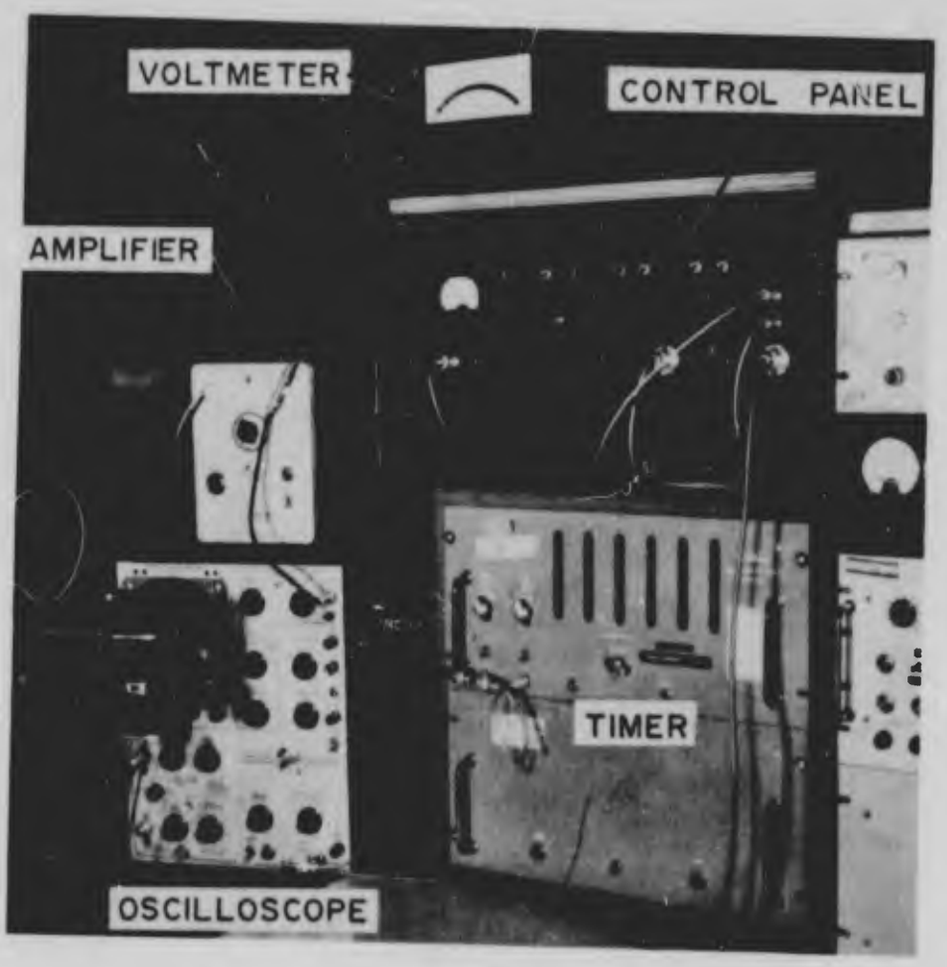


Figure 8 Instrumentation

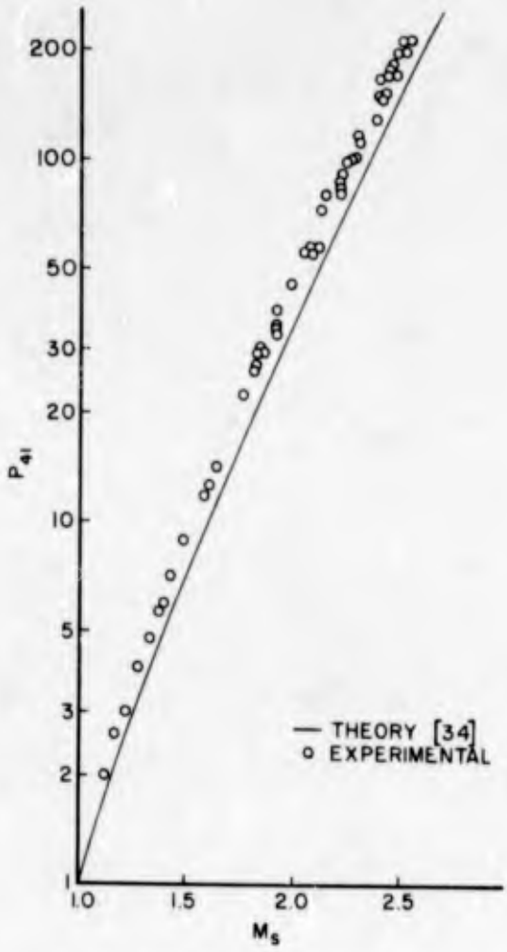


Figure 9 Effect of Diaphragm Pressure Ratio on the Shock Wave Mach Number in the AFIT Shock Tube

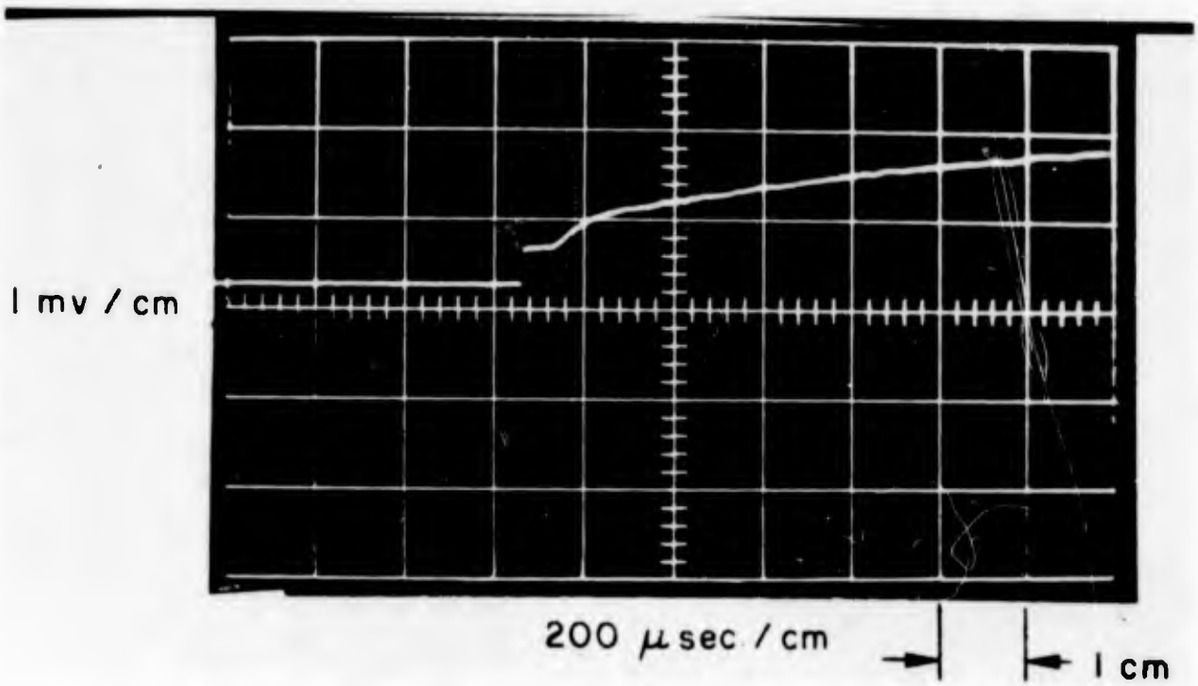


Figure 10 A Typical Oscilloscope Trace Generated by a Thin-Film Gage

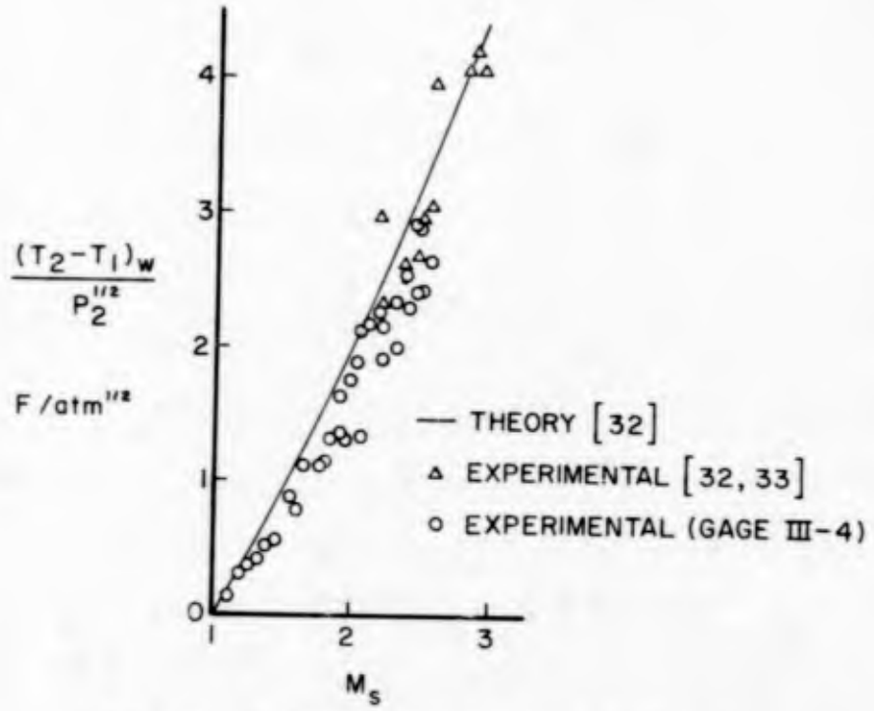


Figure 11 The Effect of Shock Mach Number on the Step in Wall Temperature of an Insulated Surface Behind a Moving Shock Wave

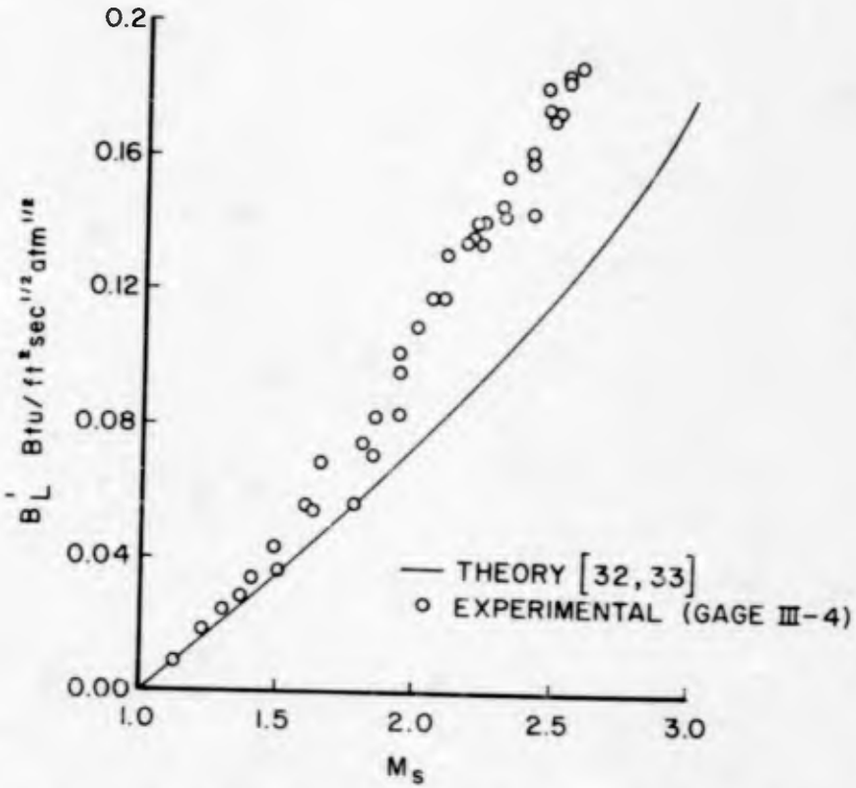


Figure 12 The Effect of Shock Mach Number on the Laminar Boundary Layer Heat Transfer

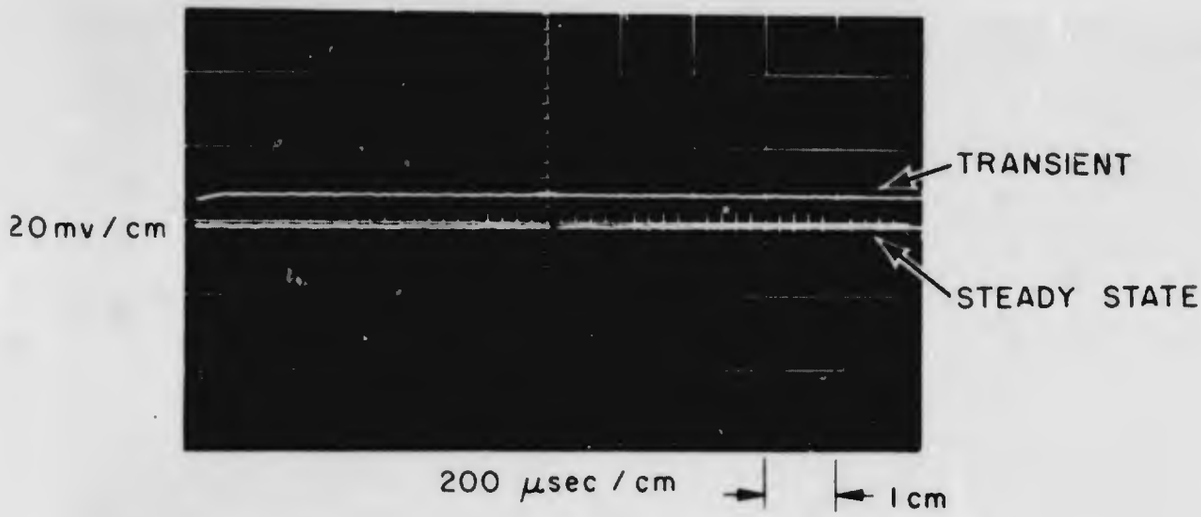


Figure 13 A Typical Calibration Response Using a Fixed Carbon Resistor in Place of the Gage

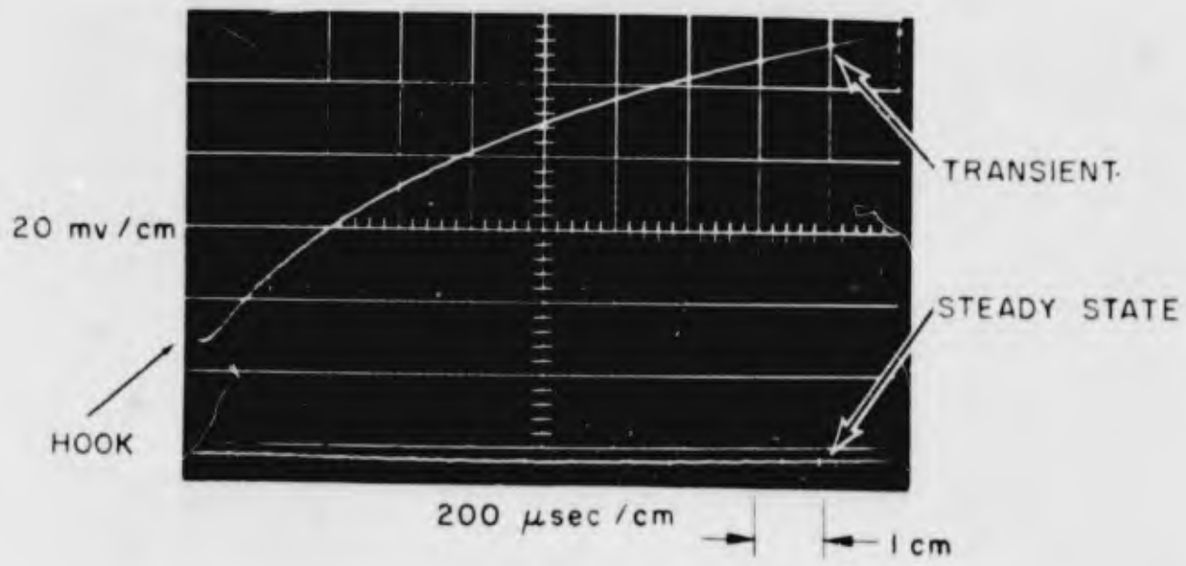


Figure 14 A Typical Calibration Response of a Thin-Film Gage

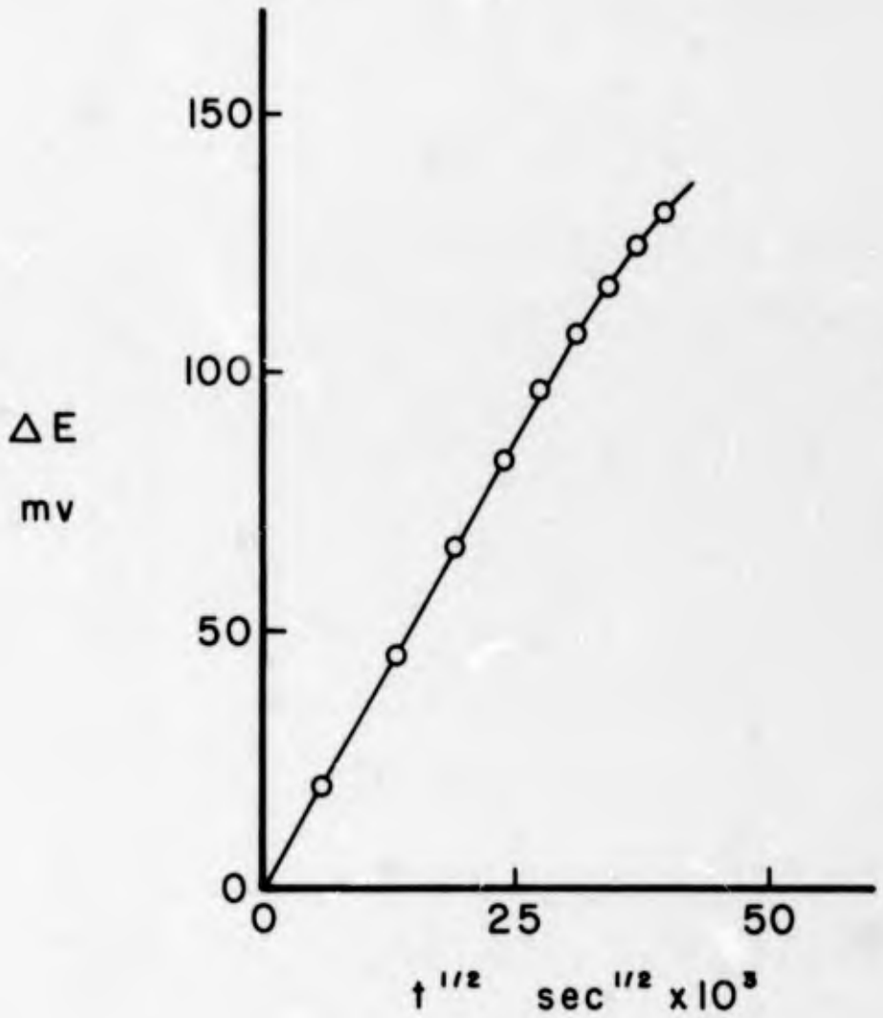


Figure 15 Converted Calibration Curve (Gage III-4)

BLANK PAGE

PERFORMANCE ANALYSIS OF AIRCRAFT SUBSYSTEMS BASED ON AUTOMATIC AIRBORNE DATA ACQUISITION (U)

By

Donald M. Caldwell, Jr.

Lee Winograd

Flight Research Division
Directorate of Flight Test Engineering
Air Force Flight Test Center
Edwards AFB, California

Donald M. Caldwell, Jr.

DONALD M. CALDWELL, JR.
Air Force Flight Test Center

Mr Donald M. Caldwell, Jr., received his BA Degree in Physics from the University of California at Berkeley in 1952. He has also done graduate work at U.S.C. and U.C.L.A. Following commissioned service with the Navy, he joined the Air Force Flight Test Center in 1955 as a physicist, primarily responsible for development of calibration procedures, equipment and facility design involved in the establishment of the Center Precision Measurement Equipment Laboratory.

Since 1961, as Aerospace Engineer with the Flight Research Branch, he has performed trajectory analysis on land recovery of lunar vehicles and standardization of aircraft flight performance. Currently he is Project Engineer on analysis of aircraft subsystem performance. He is a member of AIAA.

Lee Winograd

LEE WINOGRAD
Air Force Flight Test Center

Mr Lee Winograd served as a naval aviator aboard the USS Saratoga and also as a seaplane pilot. At the Nave Bureau of Aeronautics, Washington, D. C., he was involved in the long-range planning of Naval and Marine Corps engine requirements and also in the technical management of the design and development of turboprop and helicopter turbine engines.

He currently holds a Commercial Pilot License with flight instructor, ground instrument instructor, and glide ratings.

He is a graduate of both the University of Southern California's College of Mechanical Engineering and the College of Aeronautics. He received his Master of Science Degree in Mechanical Engineering from the University of California, Berkeley, and has completed forty-nine (49) units toward a Doctorate in Engineering. He attended Southeastern University, Washington, D. C., where he studied for a MBA in Administration.

He is currently assigned to the Flight Research Branch of the Air Force Flight Test Center, Edwards AFB, California.

BLANK PAGE

ABSTRACT

A number of measurements have been made of parameters of subsystems' operation on the C-141A aircraft in order to analyze and predict degraded performance, malfunctions and failures. A method of approach is described for analyzing subsystem functions and correlating such measurements to reveal trends in the electrical and mechanical condition of subsystems and their major components.

Principles are presented for identifying characteristic performance parameters for three functional classes of subsystems and for selecting measurable variables of subsystem operating state, control inputs and environmental stresses. The feasibility of measuring and recording the variables with sufficient accuracy for detection of performance trends is discussed.

Means of calculating trends in the measurements are given. These trends provide a tool for gleaning from many flight hours of data a few numbers which determine if subsystem performance is changing significantly and which help to isolate the cause. Such information enables the subsystems engineer to discriminate, for example, between instrumentation malfunction and trends in monitored subsystems.

The report concludes with a discussion of statistical measures of regression and correlation of flight data giving information on present and predicted subsystem "health" and subsystem interactions.

TABLE OF CONTENTS

LIST OF SYMBOLS

SYMBOL	DEFINITION
A	Cross-section area of generator core; area available for heat transfer
B_{max}	Maximum flux density
C	Specific heat capacity
C_1, C_2, etc.	Constants
C_p	Specific heat at constant pressure
d	Thickness of laminations in generator
E	Received energy; conversion factor
E_a	Fuel cell terminal voltage
E_n	Noise power per cycle per second
E_{o_ℓ}	Fuel cell maximum voltage
E_p	Polarization voltage
E_{rms}	Root mean square output voltage
f	Frequency; Faraday constant
f_{o_ℓ}	Force, loaded
f_{o_u}	Force, unloaded
F	Force dimension; weighting factors
h_{CD}	Enthalpy at compressor discharge
h_{CI}	Enthalpy at compressor inlet
h_{TD}	Enthalpy at turbine discharge
h_{TI}	Enthalpy at turbine inlet

SYMBOL	DEFINITION
I_o	Optimum control current
I_m	Maximum control current
K	Boltzmann's constant; conductivity; loss coefficient
K_e	Eddy current loss per unit volume
K_h	Hysteresis loss per unit volume
ℓ	Length dimension
L	Ohmic losses
L_r	Losses between antenna and receiver
n	Normal load factor
N	Number of turns in windings
NF	Noise figure
P	Maximum average power rating; pressure
P_c	Power loss due to copper resistance
P_{CD}	Pressure at compressor discharge
P_{CI}	Pressure at compressor inlet
P_{TD}	Pressure at turbine discharge
P_{CI}	Pressure at turbine inlet
P_e	Power loss due to eddy currents
P_h	Power loss due to hysteresis
P_{TD}	Pressure at turbine discharge
P_{TI}	Pressure at turbine inlet
q	Deterioration coefficient

SYMBOL	DEFINITION
r	Correlation coefficient
R	Universal gas constant; effective winding resistance
R_{th}	Thermal resistivity
S	Entropy; total erosion
S_{g_i}	Input stiffness
S_{g_o}	Output stiffness
S_K	Skewness coefficient
t	Time dimension
T	Effective noise temperature; lifetime of subsystem
T_A	Antenna noise temperature
T_{CD}	Temperature at compressor discharge
T_{CI}	Temperature at compressor inlet
T_o	Reference temperature for noise figure measurements
T_t	Thermal temperature
T_{TD}	Temperature at turbine discharge
T_{TI}	Temperature at turbine inlet
U_e	Energy available for external work
U_P	Energy made unavailable by polarization
v	Specific volume
V	Volume; true airspeed
W	Johnson noise power
x	Exponent in Steinmetz hysteresis law

SYMBOL	DEFINITION
Z	Number of electrons in reaction
α	Thermal diffusivity
β	Bandwidth
Δa	Change in availability of energy
ΔF	Free energy change in reaction
ΔH	Altitude difference
ΔP	Pressure overshoot
ΔS_C	Change in entropy through compressor
ΔS_T	Change in entropy through turbine
ΔT	Attenuator temperature minus ambient
ϵ	Effectiveness
η	Efficiency
η_C	Compressor efficiency; fuel cell efficiency
η_e	Electrical efficiency
η_T	Turbine efficiency
ρ	Mass density
σ	Standard deviation
$\dot{\psi}$	Rate of change of aircraft heading

PREFACE

In May of 1964, work was begun by the Flight Test Engineering Research Division on a study entitled "Improved Systems Analysis and Data Processing Techniques." These techniques were to be applied to measuring and analyzing the performance of aircraft subsystems in test flight environments. Data acquisition was based on an automatic airborne instrumentation and recording system. The objectives of this study were to develop:

1. Techniques to determine if subsystems conform to specifications.

2. Means of analyzing degraded performance, malfunctions and failures and identifying the causes.

3. Techniques for detecting performance trends and predicting degradation and catastrophic failure.

This report is intended to describe the approach to analysis of subsystem performance and the techniques developed thus far. An integrated aircraft pulse code modulation data system was developed using modified hardware from the Turbojet Engine Analyzer System project of Systems Engineering Group, Research and Technology Division. The system is described in reference 1. The application of the data system in Category II Systems Evaluation Tests of the C-141A aircraft by the Air Force Flight Test Center from May 1965 to January 1966 is outlined in references 2 and 3.

Subsequent reports will describe the results of applying these techniques to analysis of actual subsystem performance on the C-141A and the relation of these techniques to assessment of aircraft reliability and maintainability.

INTRODUCTION

It is shown herein that any subsystem in an aircraft can be categorized by its basic function of generation,[1] transformation or configuration of energy. Each subsystem, by virtue of this function, has certain parameters which characterize how well it is performing. Examples of such parameters are efficiencies, electrical signal to noise ratios, elastic moduli and other parameters that tell the extent to which the subsystem degrades the availability of the energy with which it deals.

These performance characteristics are affected primarily by three sources of stress:

1. The design of the subsystem.

2. The environment in which it is required to function.

3. The manner in which it is controlled.

[1] Strictly speaking, conversion of the mode in which energy is made available, e.g., generation of electrical energy by conversion from mechanical rotation, by an alternating current generator. These functions are more fully defined in the next section.

The design, the environment and the operating mode differ somewhat in the types of stresses they apply to the subsystem. But the ways in which the subsystem experiences stress are primarily these:

1. Being forced to operate outside of designed operating limits. (example: shaft overspeeds)

2. Being subjected to excessive rates of change. (example: abrupt pressure losses)

3. Enduring out of tolerance conditions for too long a time. (example: engine turbine section deterioration by prolonged overheat)

This report describes typical measurable variables from which performance characteristics (present health) can be computed for any subsystem in the three functional classes. It describes quantities that can be measured to specify the stresses on the subsystem.

The trends in these measured variables are described, which tell how much and how fast the performance is being degraded by the stresses imposed. Trend (predicted health) indicators are given, in the form of statistical measures of regression and correlation. These serve to condense the trend information to meaningful plots which can be reviewed in a short time, so that incipient malfunctions can be identified.

The mathematical and physical developments are presented in terms of their very basic concepts, equations and statistical measures. Subsequent reports will extend the scope and depth of the development.

In summary, this report proposes a way of categorizing the functional characteristics of subsystems which yields insight into what measurable parameters best reflect the current and anticipated status of the subsystem.

CLASSIFICATION OF SUBSYSTEMS BY FUNCTION

DEFINITIONS

The following terminology is used uniformly throughout the report. These terms may not be used in the same sense in other literature on systems analysis.

SYSTEM: A strategic, tactical or support aircraft. Examples are the B-52H, F-111A and C-5A.

PRIMARY SUBSYSTEM: A principal subdivision of a system that performs energy generation, transformation or configuration. Examples of primary subsystems performing these functions respectively are the engines, the flight control subsystem and the airframe.

SECONDARY SUBSYSTEM: A subdivision of a primary subsystem performing the same function or a different function. Examples are a fuel control, an

amplifier power supply or a control surface. A primary subsystem may be composed of many secondary subsystems.

ENERGY GENERATION: Conversion of energy from one form to another to make it more useful. The change of the chemical potential energy of fuel to the kinetic energy of hot gases in a turbojet engine or the change of the energy of a rotating shaft into electrical power by a generator are examples.

ENERGY TRANSFORMATION: Generally, the function of changing the magnitude, frequency or phase of an applied process. This function can be further categorized into control or information processing, i.e., those functions which enable the crew to operate or monitor another function. Examples of subsystems which perform this function are the environmental control (air conditioning) subsystem or the communication and navigation radars.

ENERGY CONFIGURATION: The function of maintaining subsystems in a given spatial relation one to another, statically or dynamically, or changing the place of application of mechanical, electrical or other forces. This function is exemplified in the action of structural members, airfoils, control cables, electrical wiring and the like.

MODULE: A subdivision of a subsystem having an envelope within which a certain function or group of functions can be considered isolated for purposes of evaluating environmental, operational or design stresses. Amplifiers, servomechanisms, sections of an engine, displays, etc., can be regarded as modules. Many modules can be removed from the system for checkout or maintenance. Each module is itself a secondary subsystem or contains other subsystems.

The basic reason for classifying subsystems into the categories defined above is that subsystems or modules having the same function have in common certain performance characteristics, or parameters. These enable the analyst to determine how well the subsystem is performing its function (present health) and how long it can be expected to continue to do so at a given level of confidence (predicted health).

It is postulated that what a given subsystem is really designed to do is to minimize the degree to which it degrades the availability of the energy it deals with during its operation. In the case of an energy generation subsystem, the goal is to generate efficiently. The objective for a transformation subsystem is maximum power transfer, best signal to noise ratio, minimum control error or the like. Proper operation of the configuration subsystem is evidenced by preservation of its elastic properties, surface finish, electrical continuity, or similar characteristics.

It is clear that the subsystem must operate to achieve minimum degradation of performance under certain constraints. These are imposed by its environment (principally through exchange of heat and work) and by control inputs. Another set of constraints is imposed in two ways by the design of the subsystem. If the design did not provide for sufficient resistance to known environmental or control stresses, subsystem performance may degrade more rapidly than optimum. If the range of expected stresses was imperfectly known

or operational constraints are accidentally exceeded, malfunctions or failures can be expected.

The next sections of this report specify the types of characteristic performance parameters for each functional class of subsystem. These characteristic performance parameters are of great utility because they are generally functions (in the sense of being expressible mathematically) of measurable variables giving the state of the subsystem at a given time. The values of environmental stress variables and control input variables and control input variables also can be measured in most cases.[1]

CHARACTERISTIC PERFORMANCE PARAMETERS

THE GENERATION SUBSYSTEM:

The concept of the efficiency of a process or effectiveness of a power cycle is certainly basic in propulsion or electrical engineering. The concept is formulated typically as follows:

$$\text{Efficiency } \eta = \frac{\text{useful energy}}{\text{applied energy}}$$

$$\text{Mechanical } \eta = \frac{\text{useful work}}{\text{useful work + losses}}$$

$$\text{Electrical } \eta = \frac{\text{output}}{\text{input}} = \frac{\text{output of generator}}{\text{output + losses}} = \frac{\text{input - losses}}{\text{input of motor}}$$

$$\text{Thermal } \eta = \frac{\text{work output}}{\text{heat added}} \text{ (for a closed system).}$$

Since the basic characteristic which enables us to diagnose subsystem health is the change in energy availability, the efficiency must be formulated to give loss terms explicitly, if possible. This will now be done in some detail for the engine as a gas generator and for an electrical generator and a fuel cell. The method of analysis can be used for other generation subsystems or modules. Since the engine and electrical subsystems are vital to system effectiveness and the former gives rise to many maintenance actions, a detailed treatment is justified.

TURBOJET ENGINE:

The function of the turbojet engine is to convert the chemical potential energy of hydrocarbon fuels and air into the kinetic energy of a jet exhaust stream. If this function is viewed as an adiabatic steady flow process with real losses, an overall thermal efficiency based on fuel heat content can be assigned (reference 5, p. 66).

---

[1] The feasibility of measuring state and control variables and specifying constraints precisely enough to determine characteristic performance is the subject of a subsequent report. For such cases as the adaptive control (autopilot) and crew member plus transformation subsystem, the problem of measurement, with minimum perturbation of the measured system is paramount, and is considerably difficult to analyze.

$$\eta_{thermal} = \eta_{propulsive}\eta_{cycle}\eta_{combustion}$$

The propulsive efficiency is a function of the thrust power output and the relative kinetic energy of ram air (as measured by airspeed and altitude) and jet exhaust. The combustion efficiency is a function of burner design and the chemical kinetics of fuel combustion processes. Only cycle efficiency will be considered in this discussion because primary interest is in gas generator health and its trends. Although these are obviously correlated with aircraft maneuver and environmental stresses, the engine exhibits degraded performance even in the test cell. By focusing on the loss mechanisms, we can understand better the significance of the correlation between stress and degraded ability to function.

A good characteristic parameter for a power cycle (conversion of heat to work) is the effectiveness, elucidated by Obert in reference 6, pages 431 - 437:

$$\varepsilon = \frac{\Sigma W}{\Delta a} = \frac{\text{Work delivered by power cycle}}{\text{Work delivered by cycle} + \text{Heat delivered by surroundings}}$$

The algebraic sum of numerator and denominator in this equation represents the loss of available energy arising because of the irreversibilities of the cycle. It can be assumed for steady flight at subsonic speed that the transfer of heat from the surroundings to the cycle is reversible. Hence, irreversibilities arise only from processes within the subsystem.

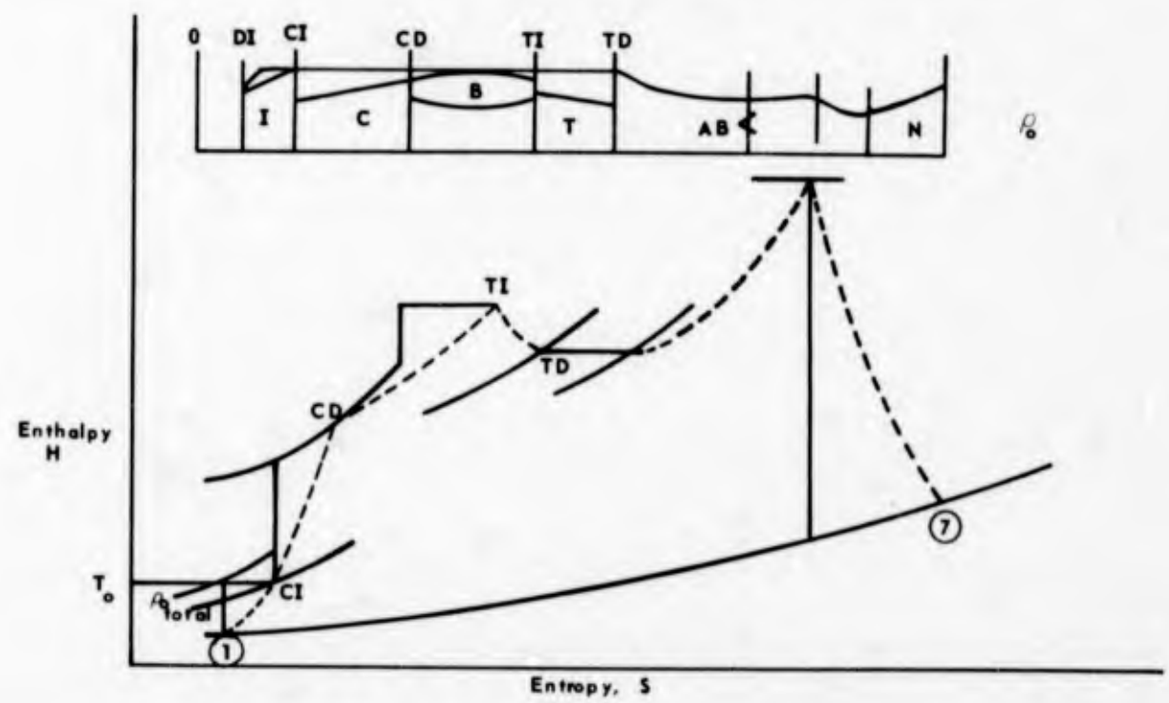


FIGURE 1. BRAYTON CYCLE FOR TURBOJET ENGINE

$$\eta_T = \frac{(h_{TD} - h_{TI})\text{real}}{(h_{TD} - h_{TI})\text{constant S}}$$

The quantities (derived in the appendix):

$$\Delta S_C = C_p \ell n \frac{T_{CD}}{T_{CI}} - R \ell n \frac{P_{CD}}{P_{CI}}$$

and $$\Delta S_T = C_p \ell n \frac{T_{TD}}{T_{TI}} - R \ell n \frac{P_{TD}}{P_{TI}}$$

are the increases in specific entropy of the air (for unit mass) involved in the operation of these subsystems.

If measurements are made of the pressure and temperature ratios involved, entropy increases can be computed as a function of time. Consider the possibility that the increase may be linear (figure 2):

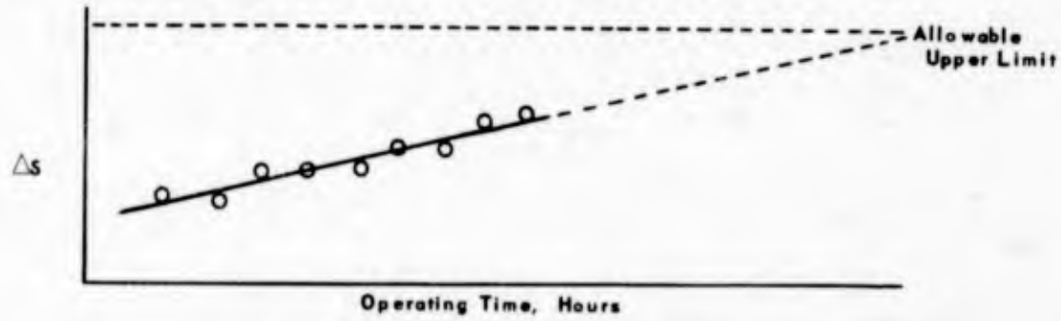


FIGURE 2. LINEAR ENTROPY INCREASE

Upper limits can be placed on the loss of subsystem efficiency that can be tolerated before maintenance action is taken. The trend shown above can indicate when this limit is expected to be reached. Moreover, if the rate of loss accumulation increases (figure 3), this could be indicative of rapid deterioration in the mechanical condition of engine components.

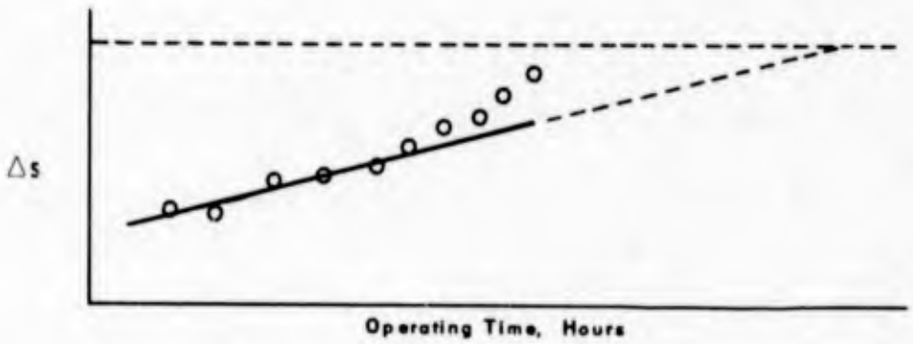


FIGURE 3. ACCELERATED LOSS ACCUMULATION

The effectiveness of the Brayton Cycle for air can be formulated for the engine as a whole and an overall entropy increase calculated. The cycle is illustrated in figure 1 (reference 7, p. 161); the entropy increase is given by S at point 7 minus S at point 1. It is most useful, however, to calculate module (engine section) efficiencies and the associated losses. This is done in appendix I for compressor and turbine.[1] The cycle efficiency is the product of module efficiencies:

$$\eta_{cycle} = \eta_{ID}\eta_{C}\eta_{B}\eta_{T}\eta_{AB}\eta_{N}$$

where the subscripts refer to modules, as follows:

ID	Inlet diffuser
C	Compressor
B	Burner
T	Turbine
AB	After-burner
N	Nozzle

Other symbols are as follows:

O	Ambient conditions
DI	Diffuser inlet station
CI	Compressor inlet
CD	Compressor discharge
TI	Turbine inlet
TD	Turbine discharge

The losses are identified explicitly by means of the first and second laws of thermodynamics and can be shown to arise from certain fundamental mechanisms involved in the basic function of the subsystem. For example, the engine compressor's function is to raise the pressure and temperature levels (enthalpy) of the air by performing work on it (ideally) isentropically and adiabatically. The basic function of the turbine is to extract sufficient work from the hot gases emerging from the burner to provide for compression, again without increase of entropy if possible. Because of irreversible processes (friction, thermal and mechanical distortions, etc.) involved in the shafts, discs and blades, the actual enthalpy changes will be accompanied by changes in entropy. The efficiency equations can be expressed as:

$$\eta_{C} = \frac{(h_{CD} - h_{CI})\text{constant S}}{(h_{CD} - h_{CI})\text{real}}$$

[1] A complete discussion of a turbojet thermodynamic model and factors affecting module efficiencies is given in reference 4.

From the mechanism of the generator, the flux density can be expressed as

$$B_{max} = \frac{E_{rms}}{4.44 \times 10^{-8} ANf}$$

where A = cross-section area of core

N = number of turns in windings

the losses can then be written:

$$P_c = C_1 E_{rms}^2$$

$$P_e = C_2 E_{rms}^2$$

$$P_h = C_3 E_{rms}^2$$

where

$$C_1 = R_{effective}^{-1}$$

$$C_2 = \frac{K_e V}{(4.44 \times 10^{-8})^2 A^2 N^2}$$

and

$$C_3 = \frac{K_h V}{(4.44 \times 10^{-8})^2 A^2 N^2 f} \qquad \text{for } x = 2$$

the total losses become:

$$P_{losses} = C_4 E_{rms}^2 = (C_1 + C_2 + C_3) E_{rms}^2$$

In general $R_{effective}$, K_e and K_h change with time as the generator is subject to stress, such that C_4 and P_{losses} will increase. For complete evaluation of the overall efficiency, the mechanical friction losses must be included.

Another type of generator is the chemical cell, where losses come about from progressive polarization (reference 8). For the fuel cell (typical reaction: $H_2 + \frac{1}{2} O_2 \rightarrow H_2O + 2e$) one can define a cell efficiency:

Another subsystem whose efficiency of generation may be evaluated is an electrical generator. Electrical losses occur in the windings of the generator due to copper resistance, the maintenance of eddy currents and alterations of magnetic domains (hysteresis). If the electrical efficiency is expressed as:

$$\eta_e = \frac{\text{output power}}{\text{output power + losses}}$$

the loss terms consist of:

$$P_c = \frac{E_{rms}^2}{R} \qquad \text{copper loss}$$

$$P_e = K_e V f^2 B_{max}^2 d^2 \qquad \text{eddy current loss}$$

$$P_h = K_h V f B_{max}^x \qquad \text{hysteresis loss}$$

where

R = effective winding resistance at frequency f

E_{rms} = output voltage

V = volume of core

f = operating frequency (e.g., 400 cps)

B_{max} = maximum flux density

d = thickness of laminations in core

K_e = eddy current loss per unit volume, dimensions of conductivity $[m^{-1}\ell^{-3}tq^2]$, a function of material

K_h = hysteresis loss per unit volume, dimensions of reluctivity $[m^{-1}\ell^{-1}q^2]$, a function of material, thermal and mechanical history, and range of B_{max} considered

x = an exponent generally between 1.6 and 2.5, a function of material and range of B_{max} (Steinmetz Empirical Law)

$$\eta_c = \frac{U_e}{U_e + U_p}$$

where

U_e = energy available for external work

U_p = energy made unavailable by polarization

$U_e + U_p = \Delta F$, the free energy change from the chemical reaction

The energies are related to the voltages produced in the cell as follows:

$$U_e = ZfE_a$$

$$U_p = ZfE_p$$

$$\Delta F = ZfE_o$$

thus

$$E_p = \frac{\Delta F}{-Zf} - E_a$$

where

E_a = cell terminal voltage

E_p = polarization voltage

E_o = theoretical maximum available voltage

Z = number of electrons produced in the reaction (2 in the example)

f = Faraday constant (calories per electron volt)

The terminal voltage can be measured continuously; initial ΔF can be found from electrochemical tables (calories per mole). Thus polarization losses will change E_a and ΔF as a function of time and the time history of E_p will show the loss in cell performance.

The reactions of these generation-type subsystems to stress has been shown to appear in the form of losses which can be measured in terms of the parameters of subsystem operation.

THE TRANSFORMATION SUBSYSTEM:

This is a large class of airborne subsystems whose functions can be considered in the following categories:

Transformation
- Control
- Information Processing
  - Communication
  - Computation
  - Measurement - Monitoring

With the exception of the mechanical portions of flight control subsystems, fuel controls, constant speed generator drives and the like, the bulk of transformation systems are electrical and electronic. They can be further characterized by which variable of an electrical signal they modulate (amplitude, frequency or phase) and in what mode they deal with the variable (continuous, discrete or hybrid).

Characteristic performance parameters are of the form of impulse-response transformation functions, such as gain-bandwidth products, noise figures, stability margins, time constants, or measures of computational errors. These characterize the degradation in the overall organization of the available energy (information loss) or in the overall amount of available energy (power loss) with which the system deals.

In general, a complex avionic subsystem (e.g., terrain following radar) can be broken down to a number of modules according to function, such as power supplies, (Klystron) modulators, attenuators, couplers and antennas. Each module is in turn composed of a number of circuits employing separate or integrated components. At each level, a performance characteristic can usually be identified; it then becomes a question of the technical and economic feasibility of measuring the quantities necessary to calculate performance.

An example of an overall performance characteristic parameter is the general radar performance equation developed by W. M. Hall (reference 9), of the following form:

$$\frac{E}{N} = \frac{\text{Received energy}}{\text{Noise power per cycle per second}} = \frac{\text{Applied energy}}{\text{Applied - Useful energy}}$$

$$\frac{E}{N} = \frac{\text{Received power X time of observation}}{\text{Boltzmann's constant X effective noise temperature}}$$

Dimensionally,

$$\frac{E}{N} = \frac{[\text{power}]\ [\text{time}]}{[\text{entropy}]\ [\text{temperature}]} = \frac{[FLt^{-1}]\ [t]}{[FL\theta^{-1}]\ [\theta]}$$

The term of interest is the effective noise temperature:

$$T = T_a/L_r + T_t(1 - 1/L) = (NF - 1)T.$$

where T_a = Antenna noise temperature (composite of radiating sources in a field of view of the antenna)

L_r = Loss between the antenna and receiver, decibels

T_t = Thermal temperature of transmission line

L = Ohmic losses in waveguide, antenna and duplexing equipment, decibels

$(NF-1)T_o$ = Effective receiver noise temperature

NF = Receiver noise figure

T_o = Reference temperature for NF measurement

The various terms in the equation can be evaluated in general only from a complete calibration of the radar system using a standard target and noise figure determination. As shown in reference 9, E/N can be expressed as a function of transmitter power, transmitter and receiver gains, and look times based on antenna pattern geometry. If calibrations are made with constant propagation factors and ranges, the change of N with time will indicate changes in the factors, which can be correlated with other performance trends to isolate the faulty module.

In practice, one factor which could be measured periodically in flight is the radar receiver noise figure, NF, defined as the ratio of input signal-to-noise ratio to output signal-to-noise ratio. It can be determined by measuring the receiver noise power and comparing it with the power calculated on the basis of its measured temperature. Noise power can be measured by comparing detector load current due to receiver noise alone with current due to added input from a noise generator substituted for the receiver input. Under a test computer control, the necessary switching, gain adjustment and load current measurement could be made a few times per hour of operation, and current ratio, together with receiver temperatures, recorded for ground analysis. For modern radars, although noise powers are on the order of a picowatt, the noise figure is typically 6 to 18 decibels. Limits can be set for a given receiver. Trends in measured NF toward excessively high values would indicate degradation in crystal mixer, intermediate frequency pre-amplifiers or the local oscillator. The noise power calculated for the receiver on the basis of the temperature, T, of its input circuitry is a function of the noise voltage generated by various random processes such as thermal agitation, emission and flow of electron streams, contact variations in junction devices, etc. These processes are generally known as Johnson noise, shot noise, flicker and grain-size noises. For example, if R is the real component of the effective impedance of the input, the Johnson noise power is given by:

$$W = \frac{E_n^2}{R} = \frac{2}{\pi} KT\beta \quad \text{watts}$$

where

E_n = Noise voltage

K = Boltzmann's constant (1.37×10^{-23} watt-sec(°K)$^{-1}$)

R = Resistance in ohms

and

β = Bandwidth of operation of component, radians sec^{-1}

The Johnson noise power is a very small power which is a measure of the deterioration of information in the function of the receiver input. Note that W is a small fraction of the energy supplied to the receiver which causes a degradation in information, R is a property of the materials of its components and β represents the number of cycles of operation. The significance of these terms will be discussed in the summary of this section.

Considerable power will be dissipated in the operation of a device whose function is to attenuate power. As used in a radar receiver-transmitter, dissipative coaxial or waveguide attenuators have a maximum average power rating P, in watts, which is a function of their attenuation ratio, α, and the temperature rise, ΔT, allowed in their operation. α is defined as follows:

$$\alpha \equiv 10 \log_{10}(P_{in}/P_{out}) \quad \text{decibels}$$

Dissipated power can be expressed by:

$$W = EP\left[1 - \frac{1}{\text{antilog}_{10}(\alpha/10)}\right] = \frac{\Delta T}{R_{th}} \quad \text{Btu hr}^{-1}$$

where

P = Maximum average power rating — watts

R_{th} = Thermal resistivity of the attenuator material — $\text{Btu}^{-1}\text{hr°F}$

$$R_{th} = \frac{A}{V\alpha\rho C} = \frac{A}{VK}$$

A = Area available for heat transfer — ft^2

V = Volume in which dissipation occurs — ft^3

α = Thermal diffusivity — $\text{ft}^2\text{hr}^{-1}$

ρ = Mass density — LBM ft^{-3}

C = Specific heat capacity — Btu $\text{LBM}^{-1}\text{°F}$

K = Conductivity — Btu $\text{hr}^{-1}\text{ft}^{-1}\text{°F}^{-1}$

E = Conversion factor — Btu hr^{-1}watt

ΔT = Attenuator temperature minus ambient temperature — °F

Assuming equilibrium and some knowledge of the thermal environment of the device, a continuous measurement of ΔT as a function of time should reveal changes in α. Trends in the attenuation ratio can indicate shifts in setting, if it is a variable attenuator, or changes in material properties for a fixed attenuator. The mechanism of dissipation is dielectric attenuation, a vector combination of direct current resistivity and quadrature

hysteresis due to high frequency molecular polarization shifts.

The transformation characteristic parameters discussed so far, signal-to-noise ratio or noise figure and power attenuation, are typical of those appropriate to information handling and power transfer subsystems respectively. Figures of merit must be identified which are directly related to the basic functions of the subsystem. It is obvious, for example, that a video amplifier will convert essentially the same amount of its input power to heat whether its gain-bandwidth product is correct or not. An airborne digital computer will exhibit the same gross losses whether or not it has several gates or flip flops disabled.

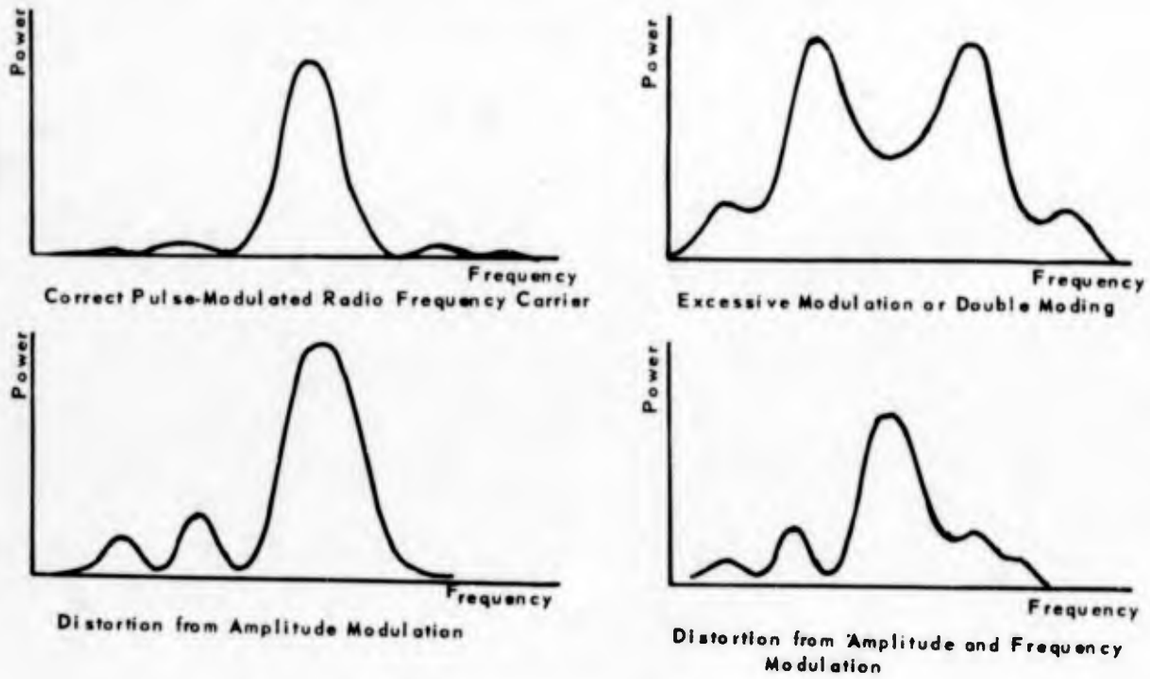


FIGURE 4. RADAR TRANSMITTER SPECTRAL DISPLAYS

For the information handling subsystem, one of the most useful parameters can be the overall output spectrum in response to some input modulation. The response of a radar transmitter to modulation by short rectangular pulses at the basic pulse repetition frequency is a representative case, treated in detail in reference 10, sections 11-723 to 11-727. Figure 4 shows an ideal response spectrum and incorrect spectral displays, which can indicate magnetron pulling[1] or pushing,[2] magnetron double-moding, mode shifting or jumping, and improper pulse width.

For the power transfer subsystem, exemplified by radio frequency transmission lines or microwave waveguides, a useful figure of merit is the voltage standing wave ratio, VSWR, which is a measure of reflections occurring in a line of finite length due to line terminations in other than its characteristic impedance.[3] Changes in VSWR reveal changes in condition of both the line and its terminations.

[1]Reflections from nearby objects.
[2]Improper pulse shape or amplitude
[3]Impedance giving maximum power transfer.

THE CONFIGURATION SUBSYSTEM

The basic functions of members of this class of subsystems are to maintain spatial relationships and to transmit forces. The characteristic parameters should therefore describe the ability of a subsystem (e.g., a structural member, cable, antenna or control surface) to maintain its dimensions, contour and mechanical strength, statically and dynamically. A measure of its performance could thus involve a strain in response to a process of inelastic deformation by cyclic stress, i.e., structural fatigue. The loss mechanisms involved in such a process are principally dissipation of internal potential energy from a position of rest, known as relaxation and dissipation of kinetic energy into heat by internal solid friction (reference 13, p. 194-5). The increase of entropy during the subsystem's function could be related to the irrecoverable work represented in the area of a hysteresis loop on a stress-strain diagram.

However, solid materials undergoing a forced change of mechanical state experience changes of internal structures.[1] These can be induced by forces or temperature gradients. The rates of such changes depend on the previous history of deformations and their stress-time relation. The degree of relaxation and friction is determined, in turn, by internal structure. So the entropy gain cannot, in general, be deduced from observations of state variables such as strain at discrete times, without knowledge of the stress history and continuous state of the material.

Even if continuous observations could be made, it is questionable whethe assessment and prediction of structural fatigue is yet theoretically possible The scatter in predicted fatigue lives (cycles to failure) of aircraft structures is often one to two orders of magnitude (reference 14, p. 119). According to Freudenthal, even the concept of a limit of dissipated energy as a predictor of fatigue life is invalid. "Phenomenological theories [of fatigue] based on the assumption of a casual relation between mechanical damping as an expression of 'internal friction' and progressive fatigue damage, specifying or implying the existence of a critical amount of dissipated energy associated with final fatigue failure, appear to be as far from representing the significant aspects of fatigue as structural theories based upon the assumption that work hardening can be considered a measure of accumulation of fatigue damage" (reference 14, p. 128). The author goes on to state that the idea of a limit of dissipated energy is not only a priori untenable, but has been experimentally disproved.

The extrapolation of fatigue tests in the prediction of cumulative damage has been done for many years on the basis of some form of "Miner's Law," i.e., that damage in the specimen may be expressed as the ratio of the number of cycles at a given stress level to the number of cycles to failure at that level. This linear cumulative damage theory has been applied, for example, in aircraft structure fatigue damage calculations based on flight load spectra and fatigue strength limits expressed as stress ratios (reference 15).

---

[1] e.g., by crystallization, precipitation, fragmentation, etc., for metallic materials.

P = Average power output per cycle of period dt

Ruderfer extended this very general representation to cases where q, S and K were functions of available power and time; his approach to the analysis of aging is felt to merit consideration by subsystems analysts.

SOURCES OF STRESS

The performance characteristics discussed thus far are measures of the deterioration resulting from normal function of the subsystem due to unavoidable irreversible processes. Stresses are imposed on the subsystem which maintain or accelerate the rate of deterioration and which are a function of its environment, the control inputs to it and the adequacy of its design.

THE ENVIRONMENT

Subsystems experience stress by exchange of heat, mass or work with an atmosphere (uncontrolled or controlled) or with other subsystems. In many cases, the generation and configuration-type subsystems encounter an uncontrolled atmosphere directly (e.g., engines, airframe, control surfaces and cables) while transformation-types (computers, radars, communication sets) are maintained in a controlled environment. The principal mechanisms of exchange are temperature differences, mechanical loads and particles in periodic, aperiodic or random modes. Examples are the very low frequency thermal cycling from ground to high altitude temperatures, turbulence and gust loads and icing of control surfaces. Subsystems create environmental stresses for one another: engine bleed air presents high temperatures and gradients to the environmental control system, cyclic loads are transmitted from rotating and vibrating equipment to other structures, transients in one electronic subsystems cause disturbances in another.

CONTROL INPUTS

Subsystems experience stress by response to correct and incorrect actuation of controls. Major sources of stress are normal starting and stopping transients, particularly for subsystems that operate at temperatures and pressures far from ambient, or which handle large amounts of power. An example is the energizing of a radar transmitter. Without proper warm-up to raise effective input resistance, excessive current can be drawn with destructive power dissipation in magnetron circuits.

Perhaps the major sources of stress in the airframe are maneuver loads. Excess attitude angles, rates and accelerations can all increase stress onset. In the engine, throttle bursts are the source of high shaft torques and pressure and temperature gradients, all of which are deleterious.

"STRESSES" FROM DESIGN

There are numerous cases in which the designer of a subsystem was unable to anticipate stresses from the above sources, particularly those involving subsystem interaction. The Lockheed Electra crashes of 1959 are an illustration. Complex elastic interactions of the propeller-engine-nacelle-wing combined structure permitted aerodynamic forces on the propeller to excite wing

However, recent experimental work by Lipsett on crack propagation shows failure of Miner's Law in circumstances where high and low stress levels are applied alternately, the order of alternation being significant (reference 16); precisely such circumstances are encountered in random gust and maneuver loads on the airframe.

The above considerations make the choice of a characteristic performance parameter for the configuration subsystem not definitive. A reasonable approach to performance assessment for this type of subsystem would be empirical correlation of readily measurable variables with results of mechanical inspections and failure data from flight operations experience. A typical measurand would be the root mean square amplitude of vibration at a point on a structural member or rotating shaft. This is obtainable as the integrated output of a velocity pickup (reference 17). It is a measure of the exchange of energy of the structure with its environment. The change of intensity as a function of time, or the relative amplitudes of vibration at various points in the structure can be a very sensitive indicator of abnormal behavior (reference 18). If the output of the velocity pickup is applied to a spectrum analyzer (perhaps a set of fixed filters at discrete frequencies throughout the range of vibration frequencies present) changes in the relative energy present at each frequency can be characteristic of deterioration in the structure.

The characteristic performance parameters discussed to this point suggest that performance degradation of any type of subsystem can be viewed as a function of:

1. The fraction of the available energy produced by the subsystem's operation which goes into irreversible processes.

2. Some material properties which are measures of loss mechanisms.

3. The operating time of the subsystem, or, for a cyclic process, the number of cycles per unit time.

These concepts were developed by Ruderfer (reference 19, p. 361) into an equation for the lifetime of a subsystem in terms of uniform erosion per cycle of operation, proportional to energy irreversibly dissipated. The total erosion of the system, given by the following equation, can serve as a generalized characteristic performance parameter:

$$S = \int_0^T qKP\,dt$$

where

T = Lifetime of subsystem

q = Deterioration coefficient (property of material)

K = Loss coefficient (fraction of available energy which produces erosion)

flutter, resulting in failure in torsion.

It is fairly common in the test instrumentation of aircraft to find that an instrumentation subsystem of excellent overall design which has been thoroughly bench-tested and run through combined-environment tests, will exhibit malfunctions when installed due to interactions with other subsystems, e.g., noise due to transient fields of electrical machinery, caused by inadequate shielding.

DETERMINATION OF MEASURABLES

Thus far it has appeared that the condition of a subsystem depends upon basic mechanisms of loss of available energy implicit in its functions. These determine its response to stresses from its environment (including interactions with other subsystems), from manipulation of its controls and from design deficiencies.

Measurements must be made to determine the subsystem's operating state, environmental conditions and control variables for some reference period of time and to detect meaningful trends indicating excessive deterioration or incipient malfunction. The measured variables should therefore include:

1. Loss terms in characteristic parameter equations.

2. Manual or automatic control settings.

3. Environmental stresses, primarily temperatures, pressures and mechanical loads.

It may not be feasible in all cases to measure variables in the loss terms directly, so condition may have to be inferred from measurements of known inputs and outputs or from responses to stress, such as component temperatures, strains, dimensional changes, etc.

Subsystem performance can be revealed by two distinctive techniques of data acquisition:

1. Passive monitoring - recording of state, control and environmental measurands during normal aircraft operation. This approach is applied in maintenance, malfunction detection or crash recorders. Sampling rates may be adjusted to particular flight regimes such as take-off or cruise. Data may be screened so that only off-nominal readings are recorded.

2. Diagnostic monitoring - recording while subsystems are subjected to special control settings, varied power supply voltages, calibrated inputs, etc. This approach is applied in test recorders such as the AFFTC integrated aircraft data system, C-141 lead-the-force performance recorders and similar devices for analysis.

FEASIBILITY OF MEASUREMENTS

When any transducer is inserted into a subsystem to measure variables of its operation, energy must be withdrawn from the subsystem to obtain the information; such loading affects the measured variable. A measurement can be made with sufficient precision to establish changes in the variable in question only if the loading error is less than some fraction of the expected change. The transducer and measured subsystem must be analyzed to determine static and dynamic loading errors. The principles of such analysis are exemplified in reference 11, where it is shown that the loading error of a force transducer measuring forces on a structural element is given by

$$\frac{f_{o_\ell}}{f_{o_u}} = \frac{1}{1 + \frac{S_{g_o}}{S_{g_i}}}$$

where f_{o_ℓ} = Actual measured variable static force, loaded by transducer.

f_{o_u} = Measured variable static force, which would exist if the transducer did not load the device being measured.

S_{g_o} = Output stiffness of measured system, ratio of applied force to displacement.

S_{g_i} = Input stiffness of force transducer.

Such a representation specifies that loading error can be minimized if S_{g_i} is made high, and allows estimation of the error on the basis of S_{g_o} and S_{g_i}, which can be computed or determined by calibration. The input stiffness of the transducer is directly analogous to the impedance presented by an eletrical pickoff to a circuit in which current is being measured. The optimum situation would be a complete mismatch, i.e., no power transferred into the transducer to obtain an indication. In practice, the highest impedance is obtained by use of "null balance" devices which use an internal source of energy to develop an effect in opposition to the effect being measured, detect equality very sensitively, and read out by the setting of the internal control.

TRENDS

When functional analysis has been carried out to establish characteristic performance parameters and feasible measurands have been identified and recorded, a large amount of data is accumulated. Some basic techniques of

analysis can then be applied to establish subsystem performance compared with specifications, present rates of deterioration and trends indicating excessive deterioration, malfunction or incipient failure.

The measurands may be examined for individual behavior with reference to nominal operating points, limits, rates of change with respect to time, or some integral function. They may be correlated with measurands from the same subsystem or another related subsystem. Trends can be characterized as transient (aperiodic) or secular (long term). The following examples illustrate typical behavior and its significance.

NOMINAL OPERATING POINT OR LIMITS

A characteristic parameter of any subsystem whose function is to control a given variable at a set operating point is obviously the instantaneous value of the variable itself, for example, the frequency of the output voltage of an A.C. generator controlled by a constant speed drive (CSD). Assuming this to be sampled once per second for a wide range of engine speeds and that the frequency is a random variable with a normal or Gaussian distribution, a plot of say 1200 observations would show a typical distribution (figure 5):

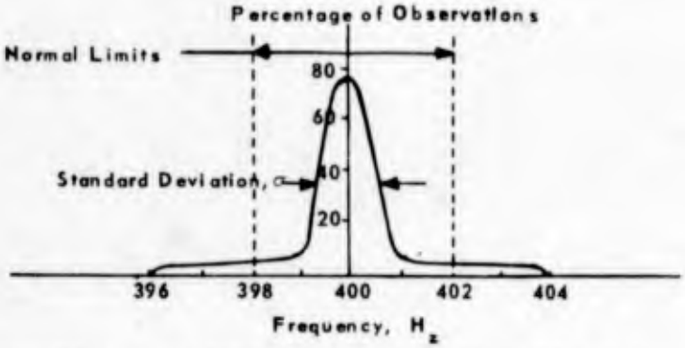


FIGURE 5. DISTRIBUTION OF FREQUENCY READINGS

Several characteristics of this distribution are significant: the mean value, F, (400 Hz), the standard deviation σ (±1 Hz) and the nominal range (±2 Hz). If a subsequent distribution representing samples, say from 600 seconds to 1800 seconds shows a change of characteristics, thus (figure 6)

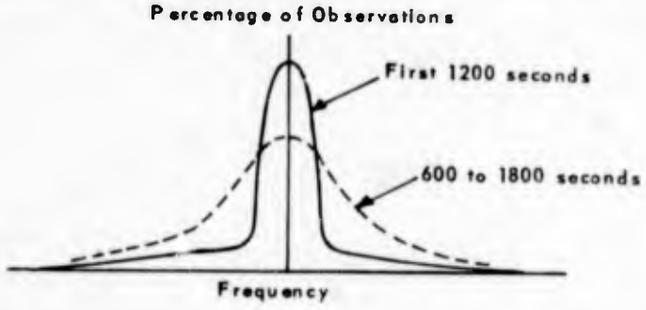


FIGURE 6. INCREASED DEVIATION IN DISTRIBUTION, WITH PASSAGE OF TIME

this could indicate a malfunction in the CSD drive transmission, speed sensor or servo valve. Such a malfunction might be expected to be accompanied by an increase in the temperature rise in the CSD lubricating oil. If the

temperature rise has been continuously recorded, its distribution would show a shift toward higher values or development of positive skew (a measure of asymmetry of the distribution) (figure 7).

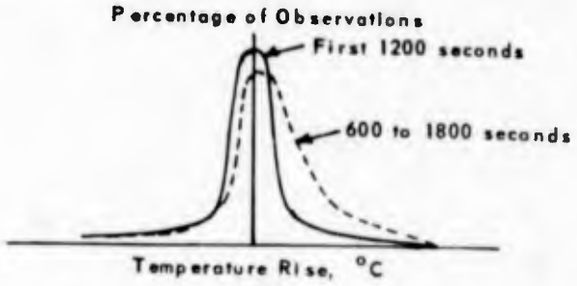


FIGURE 7. SHIFT TOWARD EXTREME VALUES, WITH PASSAGE OF TIME

In some passive monitoring systems, limit checks are made and only those readings outside the nominal range are recorded. In these cases, an increase in the number of out of limit readings per unit operating time is the trend indicator. It is neither necessary nor desirable to plot the distributions; only the values of F, σ and skew of temperature rise need to be plotted, as functions of operating time (figure 8).

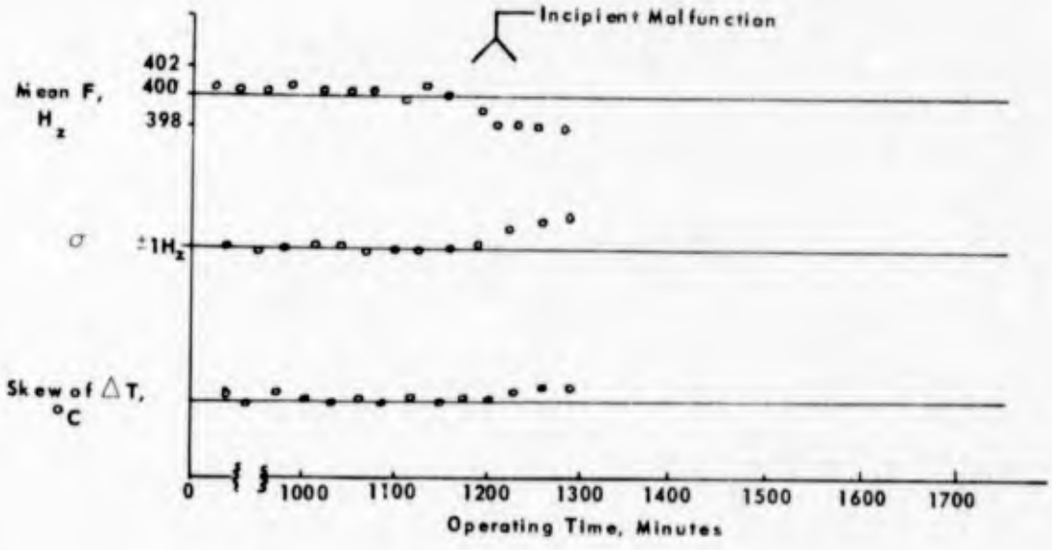


FIGURE 8. RELATED TRENDS IN DISTRIBUTION STATISTICS

The same trend indicators can be plotted for noise figures, VSWR's and other characteristic parameters whose values are expected to remain constant, in the case of diagnostic monitoring.

RATES OF CHANGE

The response time and settling characteristics of subsystems whose function is to control rates of change are indicators of proper function and health of the system. One example is the response of the environmental control system automatic pressure controllers in establishing and holding pressurization schedules during aircraft climb and descent. Here diagnostic procedures call for subjecting the system to various rates of climb and descent for various control settings. Measurements are made of aircraft pressure altitude and cabin pressure altitude, and cabin rate of climb is computed. Typical climb performance is illustrated (figure 9).

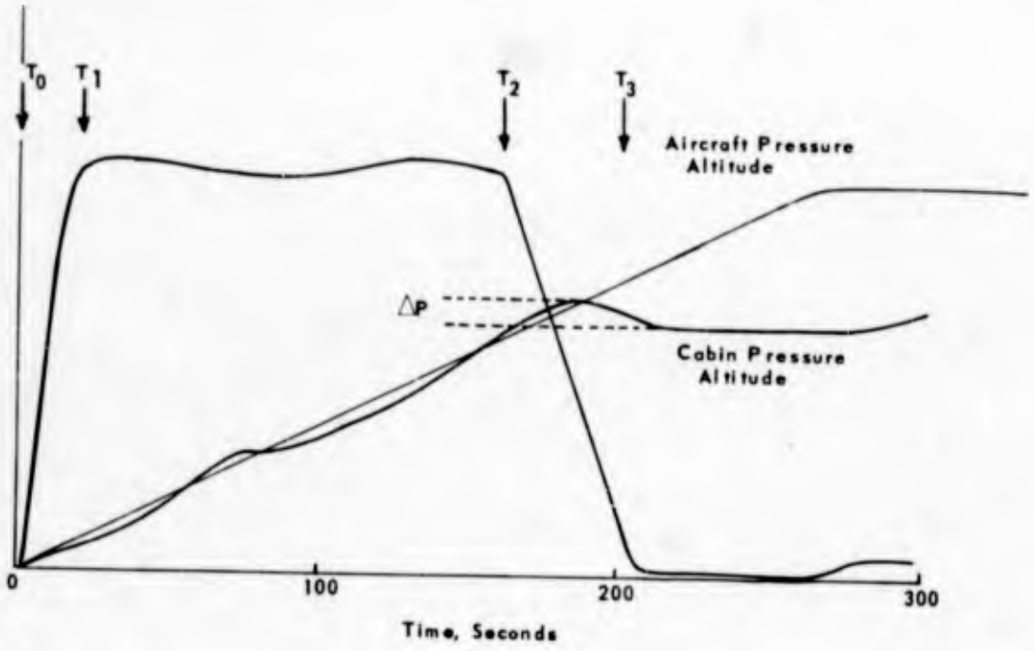


FIGURE 9. DYNAMIC CHARACTERISTICS OF PRESSURIZATION SUBSYSTEM

Here the characteristic parameters response time ($T_1 - T_0$), initial cabin rate of climb, anticipation time ($T_3 - T_2$), overshoot (ΔP) and leakage rate (subsequent cabin climb at altitude) can be compared with nominal values and trends correlated with pressurization subsystem module health.

Other types of indicators yielding information on the response of the subsystem to transient stress are the starting and stopping values of state and control variables. If the rate of change of engine speed and exhaust gas temperature are computed for the same environment conditions (e.g., ground starts at the same field elevation and ambient temperature), the trends in these variables can indicate degraded modules or components in the engine. For engine starts, the slope of the RPM vs time curve has small variations.

A significantly lower slope, described as a slow start (reference 12, p. 2-54), can indicate excessive rotor friction, faulty ignition, faulty fuel control, burner pressure problems, improper power lever rigging, etc., (figure 10).

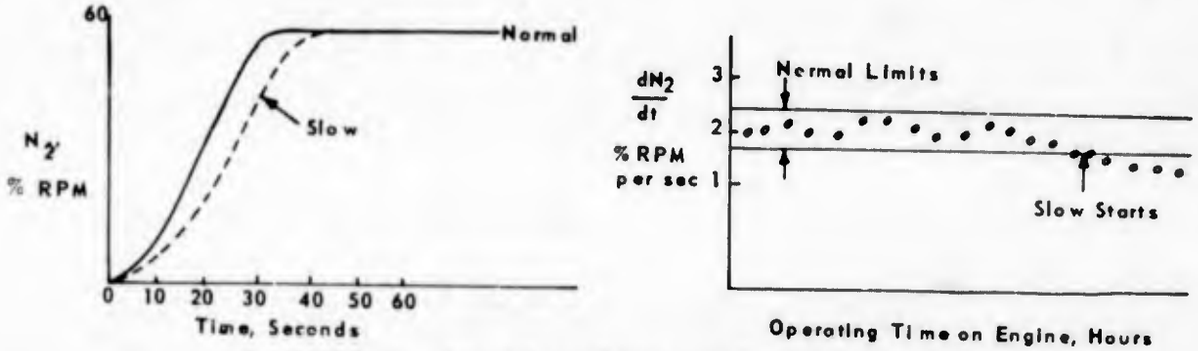


FIGURE 10. ENGINE STARTING CHARACTERISTICS

During engine shut-down, a fast stop can be indicative of excessive rotor friction or failure in accessory drive mechanisms. If starting and stopping angular accelerations are regularly computed, the trend in the mean and measures of variance as functions of time can be used as malfunction indicators.

INTEGRAL FUNCTIONS

Deterioration of subsystems as a result of cumulative strains can be revealed by measurement of control, state or environmental variables and computation of the time integral of stress. An example of a control stress occurs in the operation of navigation search radars over water. If the radar employs a magnetron oscillator, when the range is switched from medium to long or vice versa, many operators tune the magnetron control current to obtain optimum fidelity in the scope display. The average current, which controls the amplitude of the high voltage pulses from the modulator that reach the magnetron cathode, has a maximum value above which internal arcing causes cathode and insulator damage. As the magnetron ages the tolerable maximum de creases. The current is often adjusted just below the value which produces a transient meter indication of arcing, thus the total current applied during operation of the radar is a measure of cathode condition. (figure 11)

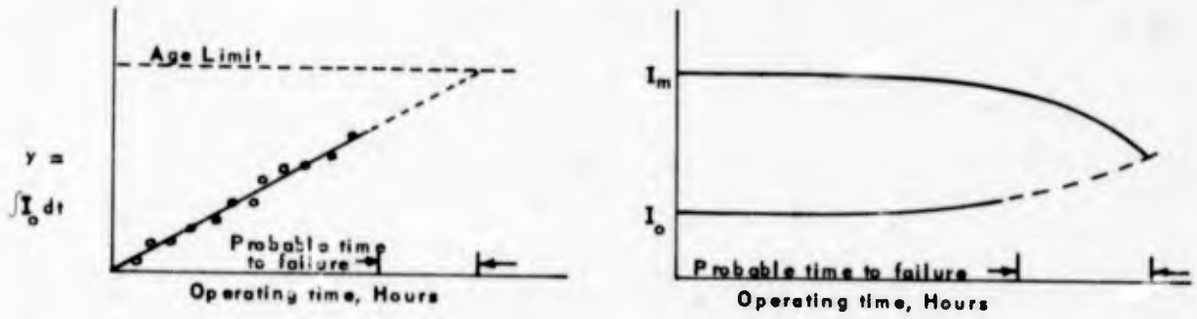


FIGURE 11. MAGNETRON AGING CHARACTERISTICS

Here I_o is the control current used to obtain an "optimum" display (for a given radar receiver gain). If I_m, the maximum current without arcing, is

known as a function of magnetron operating time (from laboratory tests for a given model), the plot shown at the right in figure 11 can be constructed.

A similar indicator of stress is the state variable exhaust gas temperature (EGT) for a turbojet engine. It is well known that high temperatures, along with high centrifugal loads, decrease the strength of turbine blades and wheels, causing a gradual permanent stretch and elongation, which can lead to intergrain cracking and rupture. Even normal EGT over a long period of time produces this "creep" effect (reference 13, p. 174-176. Extremely high EGT for a very short time can cause a rapid onset of creep. The strain is proportional to the integral of thermal stress, with higher EGT's having more severe effects. The following trend indicator can be applied:

$$y = \int F \times EGT \, dt$$

Where F is a weighting factor representing creep rate and is a function of EGT and thrust (by virtue of RPM and gas loads or turbine pressure) (figure 12).

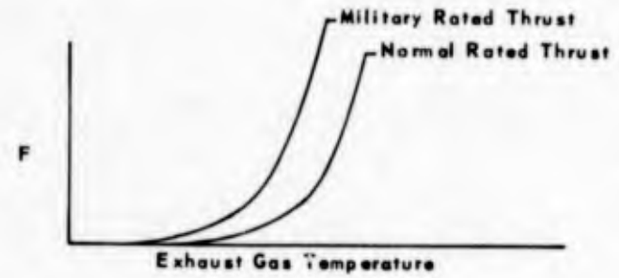


FIGURE 12. CREEP RATE WEIGHTING FACTORS

The following computed quantities could be plotted: (figure 13)

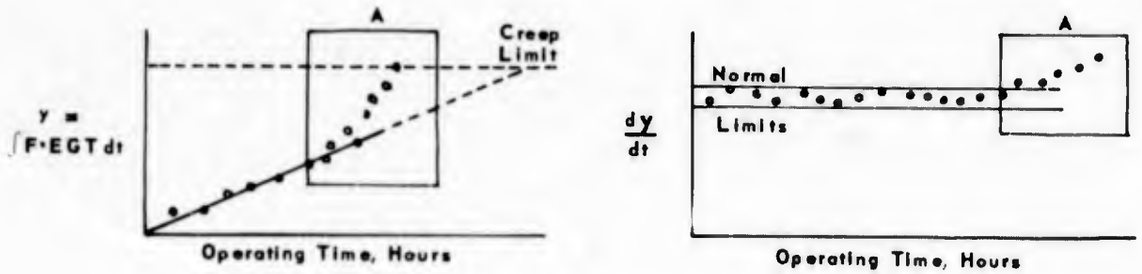


FIGURE 13. TRENDS IN THERMAL STRESS

Behavior as shown in region A could be associated with a faulty fuel control or improper engine operation.

A similar characteristic parameter is lubricating oil pressure for a rotating machine, the trends of which can give valuable information about bearing and oil seal conditions. There is an absolute lower limit on oil pressure below which the lubricating film thickness and cooling effect is insufficient to prevent bearing damage, seizure and severe damage to the

machine. Before this limit is reached, there is excessive bearing wear which is a function of the pressure drop below normal operating level, the time this condition lasts, bearing condition and loads. There is also an upper limit on oil pressure above which excess stress is placed on oil seal structures; similar conditions on magnitude and time apply. A parameter whose trend can be used as a condition indicator is thus:

$$y = \int F_1 F_2 F_3 P_{oil} dt$$

Where F_1 is a weighting factor dependent on pressure drop, F_2 depends on bearing age and F_3 depends on applied loads. These weighting factors could have typical functional forms shown below; the exact expressions must in general be determined empirically. (figure 14)

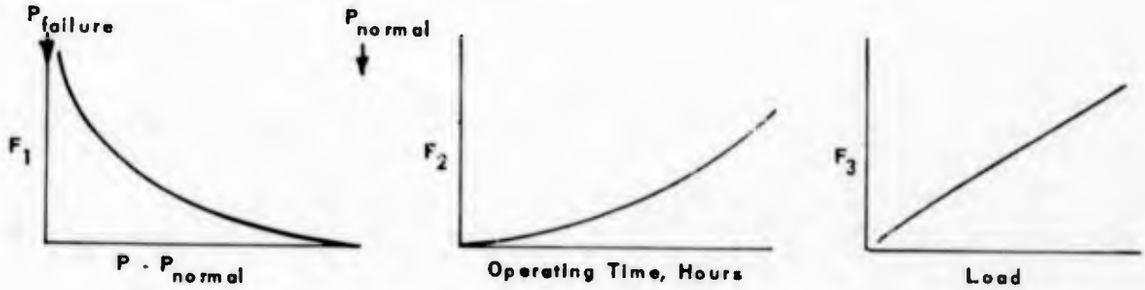


FIGURE 14. WEIGHTING FUNCTIONS IN BEARING WEAR

The behavior of Y and its rate of change as functions of time can reveal bearing condition in a manner similar to that of EGT and magnetron current. When excess friction develops because of oil starvation, the temperature rise of the lubricating oil can be expected to increase over its normal mean value; this trend substantiates the malfunction indication given by the stress integral.

SUBSYSTEM INTERACTIONS

When two or more subsystems are known to be functionally related, there are certain trends in the measured variables which can reveal changes in the interaction. An example can be taken from the operation of a terrain-following radar (TFR). The desired offset clearance (absolute altitude above terrain) is set in the operating controls of the TFR, as a nominal value. The actual offset clearance is measured by the low altitude radar (LARA). This measured value is affected by the electronic parameters influencing the LARA calibration and the mechanical factors involved in antenna alignment (as determined by boresight calibrations). If the difference, ΔH, between the offset clearance setting and actual measured altitude is taken as a random variable, properties of its distribution function can reveal changes in the influence factors. The skewness of the distribution, for example, may be computed as the coefficient:

$$S_k = \frac{\text{mean - mode}}{\text{standard deviation}}$$

flow (WF) and exhause gas temperature (EGT) for a turbojet engine are plotted against engine pressure ratio (turbine discharge over compressor inlet), over a fairly wide range of pressure ratios in steady-state operation, the relationships are nearly linear. A straight line can be fit by methods of regression from which estimates the dependent variable, Y, can be computed from values of the independent variable, X. The coefficient of correlation, r, can then be computed which is a measure of the exactness of the assumed functional dependence.[3] One form of r is given by

$$r = \pm \left[\frac{\Sigma (Y_{est} - \bar{Y})^2}{\Sigma (Y - \bar{Y})^2} \right]^{1/2}$$

where $\bar{Y}$ is the arithmetic mean value of Y and Y_{est} is the value of Y estimated from the regression equation linear in X. The numerator of the ratio expresses the explained variation in Y and the denominator expresses the total variation in Y. If r is 1, there is perfect (linear) correlation in the behavior of the two variables; if r is zero there is no direct functional relationship. If the correlation coefficients of each of RPM, WF and EGT versus EPR are compared as functions of time, a change in all simultaneously could indicate pressure transmitter malfunction (reference 20). A change in WF versus EPR while other correlations remained unchanged could indicate engine module deterioration or fuel flow instrumentation problems.

Another sort of correlation trend can reveal performance changes in essentially unrelated subsystems if the same quantity can be computed from different data sources. For example, in a coordinated level turn of an aircraft, the normal load factor at the center of gravity should be a function of the true airspeed V and rate of turn, $\dot{\psi}$, given by

$$n(\text{g-units}) = \left[\left(\frac{V\dot{\psi}}{g} \right) + 1 \right]^{1/2}$$

The normal load factor can also be measured by an accelerometer located at the center of gravity of the aircraft. If the true airspeed computed by the aircraft's air data computer and the compass heading are recorded, n can be computed for each turn. If the correlation coefficient between the computed n and the accelerometer reading changes significantly, performance degradation is indicated in either the compass system, air data computer or acceleration instrumentation.

[3] See any standard work on statistical analysis.

where mean = arithmetic average of readings

mode = most frequently occurring reading in a large group

standard deviation = root-mean-square of deviations of all readings from their mean

Since generally only the mean is affected by the occurrance of extreme values, larger values of ΔH than normal will result in a positive S_k. If the value of S_k is continuously computed for recordings of ΔH (say on the basis of readings 0 to 200, 50 to 250, 100 to 300, etc.), the trend in S_k can be associated with shifts in LARA calibration (slow drifts toward larger ΔH's), shifts in antenna alignment (sudden displacement of the distribution or discontinuity in the mean) or changes in the offset setting circuit (irregular changes in the distribution). (figure 15)[1]

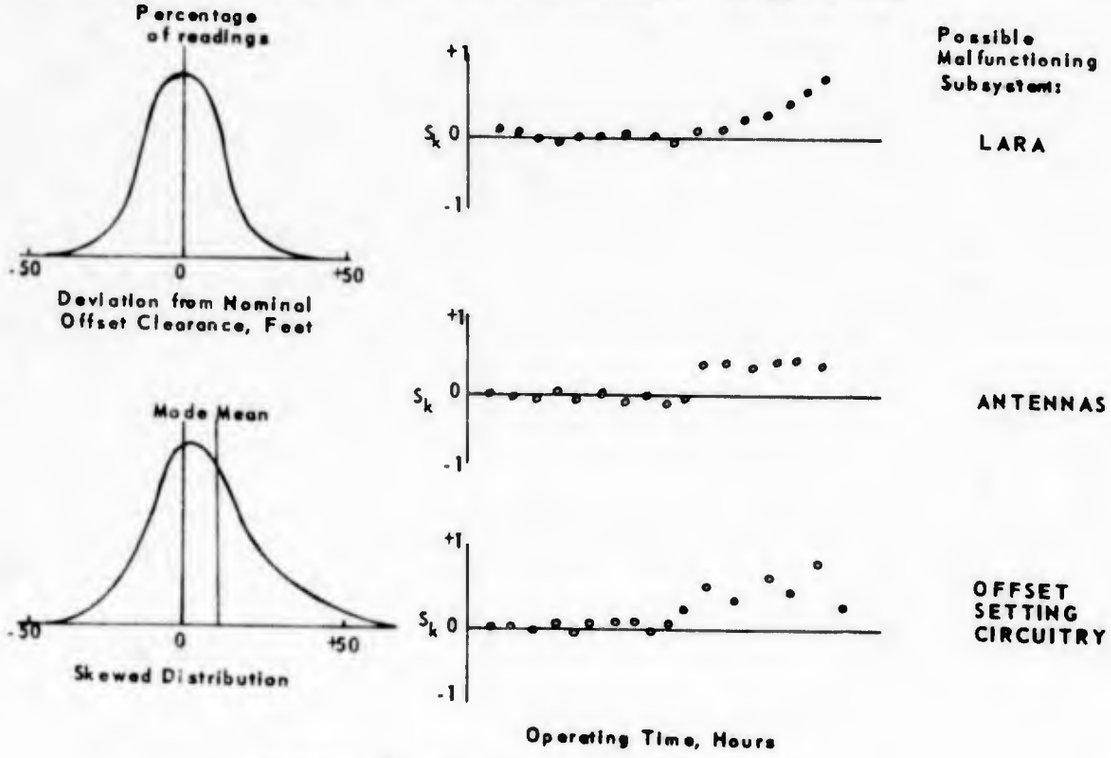


FIGURE 15. TRENDS IN SKEWNESS

CORRELATION TRENDS

Where certain variables measured in subsystem operation are functionally related to other variables (e.g., propertied of state such as gas temperatures and pressures), this functional relationship can be used to reveal not only trends in the subsystem performance but in the instrumentation used for monitoring. For example, if the values of referred[2] spool speed (RPM), fuel

[1]There is always the possibility of compensating shifts, which are undetectable by this trend technique.

[2]Adjusted to standard conditions of ambient temperature and pressure.

SUMMARY

The development of instrumentation and recording systems capable of continuous monitoring of hundreds of subsystems' measurements on an aircraft provides a basic tool for analysis and prediction of subsystem performance. The concepts and techniques described in this paper are believed to provide a unified approach to applying this tool in category II systems evaluation testing. A rational basis has been identified for selecting measurands for recording and typical diagnostic trends have been set forth. The approach is outlined below:

1. Classify the subsystem or module as to basic functions.
2. Derive characteristic performance parameters for the functions.
3. Identify the factors in the performance parameters which are associated with loss mechanisms.
4. Establish the measurable variables which make up the loss terms.
5. Determine the feasibility of recording the measurand with the requisite precision.
6. Measure, record and compute loss terms as functions of time.
7. Compute and plot variance and correlation statistics whose time trends reveal performance changes.

Trend analysis techniques are in process of being formulated in a set of digital computer programs for processing C-141A flight records and for application to F-111 and subsequent aircraft. Since a variety of correlation relations need to be explored, the analysis of subsystem evaluation recorder data is at present confined to ground-based computer operations.

APPENDIX

DERIVATION OF ENTROPY CHANGES FOR COMPRESSOR AND TURBINE

GENERAL EQUATION

From the first and second laws of thermodynamics, the following equation may be derived for the compression and expansion processes (reference 6, p.287):

$$TdS = C_p dT - vdP$$

where

S = Entropy
T = Temperature
P = Pressure
v = Specific volume
C_p = Specific heat at constant pressure

Substitution for v from the perfect gas relation

$$Pv = RT \quad (R = \text{universal gas constant})$$

yields:

$$dS = C_p \frac{dT}{T} - R\frac{dP}{P}$$

which, upon integration, is:

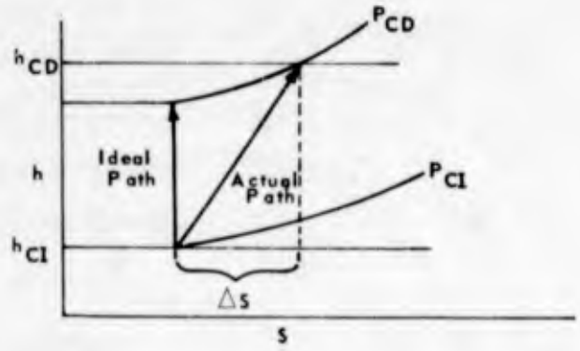
$$\Delta S_{0-1} = C_p \ln\left(\frac{T_1}{T_o}\right) - R \ln\left(\frac{P_1}{P_o}\right)$$

where the subscript o refers to inlet conditions and the subscript 1 to discharge conditions.

RELATIONSHIP OF ENTROPY CHANGE TO EFFICIENCIES

COMPRESSOR

Although the compressor work may be considered to be done adiabatically, it cannot be considered reversible or frictionless. The conventional method of expressing the compressor efficiency, η, is the ratio of the reversible (isentropic or ideal) work of the compressor to the real work required. The ideal and real paths are shown in the following enthalpy-entropy diagram:



The compressor efficiency is related to the changes in enthalpy by:

$$\eta_c = \frac{(h_{CD} - h_{CI})_{\text{constant S}}}{(h_{CD} - h_{CI})_{\text{real}}}$$

The increase in entropy may be computed as a measure of efficiency, since it is a point property dependent only on the value of limits, and can be evaluated even if the actual path is irreversible.

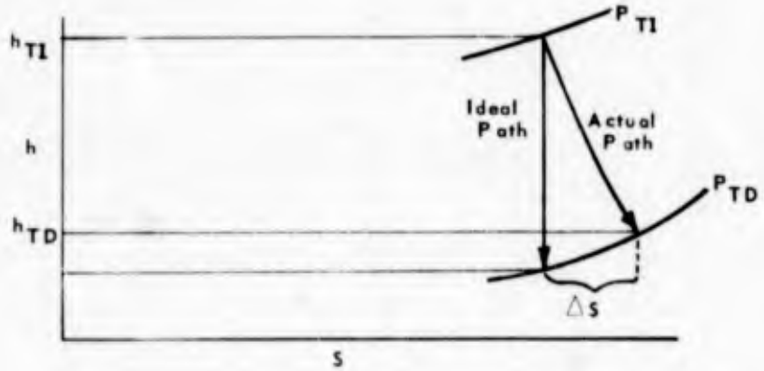
Substituting the inlet and discharge pressure and temperature ratios in the general equation yields:

ΔS_c = entropy change through compressor

$$= C_p \ln\left(\frac{T_{CD}}{T_{CI}}\right) - R \ln\left(\frac{P_{CD}}{P_{CI}}\right)$$

TURBINE

Similar considerations apply to the ideal and real enthalpy changes in the turbine expansion process:



The turbine efficiency is related to the changes in enthalpy by:

$$\eta_T = \frac{\left(h_{TI} - h_{TD}\right)_{real}}{\left(h_{TI} - h_{TD}\right)_{constant\ S}}$$

Substituting the inlet and discharge property ratios in the general equation yields:

ΔS_T = entropy change through turbine

$$= C_p \ln\left(\frac{T_{TD}}{T_{TI}}\right) - R \ln\left(\frac{P_{TD}}{P_{TI}}\right)$$

REFERENCES

1. Huppert, L., Integrated Aircraft PCM Data System for the C-141 Aircraft, FC-4556, AiResearch Manufacturing Division, Garrett Corporation, June 1965, (U), (paper presented at the Twenty-Sixth Meeting, AGARD Flight Mechanics Panel, Paris, France, June 1965)

2. Grissman, R., Capt, USAF, Flight Test Application of a PCM Data System, (U), (paper presented at the USAF Flight Data Recorder Conference, Wright-Patterson Air Force Base, August 1965).

3. Grissman, R., Capt, USAF, Flight Test Application of an Integrated Aircraft PCM Data System, (U), (paper presented at the Twenty-Sixth Meeting, AGARD Flight Mechanics Panel, Paris, France, June 1965).

4. Rumer, W. I. et al, Turbojet Engine Analyzer System Study, Final Report, ASD-TDR-62-842, AD411117, Aeronautical Systems Division, June 1963, (U) (AiResearch Manufacturing Division, Garrett Corporation Document No. FC-4186-R).

5. Durham, F. P., Aircraft Jet Powerplants, New York, Prentice-Hall, 1951.

6. Obert, E. F., Concepts of Thermodynamics, New York, McGraw-Hill, 1960.

7. Hill, P. G. and Peterson, C. R., Mechanics and Thermodynamics of Propulsion, Reading, Addison-Wesley, 1965.

8. Adams, D. R., et al, Fuel Cells - Power for the Future, Cambridge, Fuel Cell Research Associates, 1960.

9. Hall, W. M., "General Radar Equation", Space/Aeronautics R & D Handbook 1962-1963, pp D-11 to D-21, New York, Conover-Mast Publications, 1963.

10. Basic Electronics Technology and Testing Practices, Chapter II: Testing Techniques and Practices, Air Force Technical Order 31-1-141-12, 1 May 1965.

11. Chapin, W. E. and Mitchell, R. K., A Summary of Force Transducer Technology: Measurement Techniques, Calibration, Element Design, Availability and Research Activity, Transducer Information Center Summary Report No. 2, Columbus, Batelle Memorial Institute, April 1966.

12. Turbojet Engine Analyzer System Development Program, Monthly Progress Report No. 8, (U), FC-4334-8, AiResearch Manufacturing Division, Garrett Corporation, July 1964.

13. Freudenthal, A. M., The Inelastic Behavior of Engineering Materials and Structures, New York, John Wiley & Sons, Inc., 1950.

14. Freudenthal, A. M. and Gumbel, E. J., "Physical and Statistical Aspects of Fatigue", Advances in Applied Mechanics, Vol. IV, New York, Academic Press, Inc., 1965.

15. Fatigue Design Criteria and Life Prediction Computer Program for Aircraft Structures, (U), FDL-TDR-64-56, Flight Dynamics Laboratory, 1965.

16. Lipsitt, H. A., Crack Propagation in Cumulative Damage Fatigue Tests, (paper presented at the 11th Annual Air Force Science and Engineering Symposium, 1964).

17. Durbin, E. J., Head, Instrumentation and Control Laboratory, Princeton University: private communication.

18. Malfunction Detection with Vibration Instrumentation, Aircraft Gas Turbine Operation Information Letter No. 14, Pratt and Whitney Aircraft Co.,1962.

19. Ruderfer, M., "Thermodynamics of Failure and Aging", Physics of Failure in Electronics, Vol. 2, Rome Air Development Center Series in Reliability, 1964.

20. Driessen, E. A. et al, "Use of Recorders in Future Aircraft Operation", AIAA J. Aircraft, Vol. 2, No. 3, May-June 1965.

BLANK PAGE

(U) AN ANALYSIS OF RADIOISOTOPE DYNAMIC POWER SYSTEMS FOR FUTURE MILITARY SPACE PROGRAMS

by

David Kauffman, Lt, USAF

and

John T. Piker, Lt, USAF

The contents of this paper reflect the views of its authors and do not imply Department of Defense endorsement of factual accuracy or opinion.

Headquarters, Space Systems Division
Los Angeles AFS, California

Lt David Kauffman

BIOGRAPHY

a. Lt Kauffman is a project officer in aerospace propulsion and space power systems at Hq, Space Systems Division, Air Force Systems Command. He is 25 years old and was born in Glenside, Pennsylvania. He received his bachelor's degree, with honor, in chemical engineering from the California Institute of Technology in June 1962 and was commissioned a second lieutenant in the Air Force reserve. He entered active duty in September 1962, returning to California Institute of Technology, where he received his master's degree in chemical engineering in June 1963. He has been in his present position at Space Systems Division since that time.

b. Lt Kauffman is responsible for the identification and documentation of requirements for exploratory and advanced development in the areas of spacecraft electrical power systems, electric and advanced propulsion, and aerospace support techniques. Lt Kauffman is a member of the American Institute of Aeronautics and Astronautics, the American Institute of Chemical Engineers, Tau Beta Pi, and Sigma Xi.

Lt John T. Piker

BIOGRAPHY

a. Lt Piker is a project officer in space weapons and weapons effects at Hq Space Systems Division, Air Force Systems Command. He is 25 years old and was born in Conemaugh, Pennsylvania. He received an Associate of Arts degree from Potomac State College, West Virginia in 1961. He enlisted in the Air Force in 1962. Following basic training he was selected for the Airman Education and Commissioning Program. He received his bachelor's degree in aerospace engineering from the University of Arizona in June 1965. He received his commission as a second lieutenant in September 1965 and has been assigned to Space Systems Division since then.

b. Lt Piker is responsible for the identification and documentation of requirements for exploratory and advanced development in the areas of space weapons and weapons effects. He formerly was an assistant project officer responsible for similar activities in the areas of spacecraft electrical power systems and electric and advanced propulsion.

BLANK PAGE

ABSTRACT

Studies indicate that the next step in the nation's space program - beyond the current programs - may be longer duration, earth orbiting satellites manned by two to six astronauts. Since these missions would not be carried out until the 1970s, the exact nature of the spacecraft, the missions to be performed, and other parameters have not been, and would not be, defined for several years. The electrical power requirements for such missions can, however, be estimated with reasonable confidence. Average power consumption for such missions will be between 2 and 10 kilowatts; peak loads for on-board equipment may be considerably higher.

Four possible energy sources are examined for this class of missions: chemical, solar, nuclear reactor, and radioisotope. Chemical and nuclear reactor sources are eliminated because of excessive weight. Radioisotope energy sources are selected over solar energy.

Thermoelectric, thermionic and dynamic energy conversion are compared. Dynamic conversion is selected based on its overall greater potential.

A study is presented comparing five candidate dynamic energy conversion systems: the Rankine cycle using mercury, water, and organic working fluids; the Brayton cycle and the Stirling cycle. The thermodynamics of each cycle are presented and performance characteristics are given parametrically. Each cycle is examined in light of other factors, including flexibility, launch-pad and on-orbit maintainability, simplicity, weight, reliability, development status, and development problem areas. Costs are considered by examining development costs, mission-peculiar modification costs, and operational costs. The Brayton cycle is selected as optimum for this class of missions.

CONTENTS

TABLES

FIGURES

Electrical power systems for spacecraft have always been extremely important elements in our nation's space programs. Electrical power is a fundamental requirement for almost everything that can be done in space, and the limited amounts of electrical power available for spacecraft have always been among the most critical design constraints imposed. During the past few years, a great deal of effort and money have gone into developing better, more efficient, lighter weight space power systems for use in our future space programs. With eight years of practical experience in the space business, it is now possible to analyze future space missions realistically and determine their electrical power requirements.

POWER REQUIREMENTS:

In the area of unmanned satellites, we have considerable experience. An examination of all the unmanned spacecraft that have been flown to date indicates that roughly one-tenth to one-half of an electrical watt is required for each pound of total payload placed in orbit. (There have, of course, been exceptions to this rule-of-thumb, such as the Echo balloon satellites which required no power; but the number of exceptions is extremely small.) Although spacecraft electronic components are getting more complex, the power required per pound of payload in orbit has not changed significantly for some time. The power systems for all unmanned satellites launched to date have consisted of batteries, solar cells, and a few radioisotope thermoelectric generators.

With the advent of the Titan and Saturn class space boosters, larger satellites can be put into orbit. In general, however, most of the planned unmanned satellites are remaining fairly small (less than 2,000 pounds). The added booster capability is being used to place several satellites in orbit on the same launch, rather than to increase the size of individual satellites. Emphasis will be placed on increasing satellite lifetime and reliability. Some satellites now in orbit have been operating successfully for several years, using solar cell and radioisotope thermoelectric generator power systems. It is logical to assume that during the next ten years there will be further advances in unmanned spacecraft reliability and lifetime. It is not likely, however, that the electrical power requirements for single spacecraft will exceed one or two kilowatts of power with the majority of the satellites requiring only a few hundred watts.

In the area of manned spacecraft, we have considerably less experience to draw on; however, significant conclusions can still be reached. The Mercury spacecraft, which carried only one man, required less than one kilowatt of power, averaged out over the duration of the mission. Peak power requirements at launch and just prior to re-entry were considerably higher. This power was supplied entirely by batteries carried on board the spacecraft.

The Gemini spacecraft is considerably more complex in its complement of electronic equipment. In addition, it has double the crew size of

higher. This paper will now examine all the possible ways of providing the required power for such future spacecraft. Energy sources will be examined first, then energy conversion methods. Emphasis will be placed on obtaining the maximum capability with the minimum total program cost.

ENERGY SOURCES:

There are four possible energy sources that can be used to provide electrical power to satellites. These are chemical energy, solar energy, nuclear reactor energy, and radioisotope decay energy. Each type has certain advantages and disadvantages for application to the class of missions under consideration: 2 to 10 kilowatts for 30 days to a year. Each energy source must be examined in terms of weight, cost, vehicle integration, reliability, and safety. Table 1 presents a brief summary of the findings for each energy source.

First, the weights of the various energy sources will be compared. Chemical energy can be converted to electrical energy most efficiently in a fuel cell. Good fuel cell systems are capable of converting up to 65% of the total chemical thermal energy into useable electrical energy. The combination of chemicals with the highest power density (watt-hours per pound) is hydrogen and oxygen. Theoretically, 1.56 kilowatt-hours of energy is available per pound of reactants (equivalent to the total available free energy rather than the thermal energy). In practice, the maximum attainable is about 1.2 kilowatt-hours per pound. In addition, the hydrogen and oxygen must be stored in insulated tanks. The state-of-the-art of insulated cryogenic tankage for hydrogen and oxygen indicates that a minimum of one-third of a pound of tankage is required for each pound of reactants. This reduces the available energy to 0.9 kilowatt-hours per pound, neglecting entirely the weight of the conversion device. For a thirty-day, two kilowatt mission, then, a total of 1600 pounds of reactant and tankage is required. For a one-year ten kilowatt mission, this figure jumps to about 100,000 pounds.

Solar energy is available constantly at no cost in weight. It must be converted to electricity, however, either by a photovoltaic process (solar cells) or in a thermal conversion device of some sort. In the photovoltaic case, no weight will be assigned at all to the energy source. In the thermal conversion case, it is necessary to provide a solar collector, usually a parabolic mirror, and a thermal energy storage device in order to provide a constant heat input to the conversion device over the entire orbit. This weight is, of course, independent of mission duration. Detailed weight estimates have been carried out at the 20 kilowatt level, indicating that the collector and thermal storage unit will weigh roughly 1000 to 1200 pounds. Scaling down to ten kilowatts, it is estimated that these two components would weigh about 600 to 700 pounds.

Nuclear reactors offer extremely high power densities (watts per pound). The SNAP 10A reactor, representative of the type required for these missions, provided 30 thermal kilowatts and weighed just 250 pounds.

the Mercury spacecraft. Nevertheless, its total electrical power requirements range from 0.6 kilowatts in stand-by mode on orbit to 1.5 kilowatts during periods of peak activity. The maximum duration for the Gemini missions is fourteen days. The electrical power for the first few Gemini flights was provided by on-board batteries. Hydrogen/oxygen fuel cells were added for the longer missions, starting with Gemini V.

The Apollo Command and Service Module electrical power system also consists of hydrogen/oxygen fuel cells. For this three-man mission lasting a total of fourteen days, a minimum of 1.2 kilowatts is required with frequent peaks as high as 3.0 kilowatts. Batteries will be used for the Lunar Excursion Module, which carries two men to the surface of the moon for a short time.

Following the presently planned programs, there are no firm, approved manned space missions in the nation's space program. It is relatively easy, however, to postulate what some features of the next generation of missions will be. First of all, from our experience with unmanned satellites, we know that we must push for longer duration missions in order to accomplish the most possible with the smallest outlay of resources. Unless there are significant physiological limitations preventing longer stays in space for men, we can anticipate manned programs gradually growing from the currently planned 30 days to 45 days, 90 days, and a year or more, probably with crew rotation in the last case. From our experience in manning complex aircraft and our experience in unmanned satellites, it is probable that we will keep the size of spacecraft crews fairly small. Various studies have indicated that anywhere from two to six men would be optimum for a variety of space tasks. Several independent studies have indicated that from 0.5 to 0.8 kilowatts of electrical power is required continuously for life support of one man in a spacecraft. Electronic equipment on board spacecraft may require a great deal of power, or very little at all. For most concepts studied to date, the electronic equipment in manned spacecraft appears to require an average amount of power approximately equal to the life support requirement. We can then postulate that the total average power requirements will be between 1.0 and 1.6 kilowatts per man. If crew size varies from two to six men, it then appears that the power requirements will fall between 2.0 and 9.6 kilowatts, or roughly 2 to 10 kilowatts.

It is apparent, then, even though we do not know the specific missions to be accomplished, that the next generation of manned spacecraft, whether they be developed by the National Aeronautics and Space Administration or by the Department of Defense, will have average power requirements somewhere in the range between 2 and 10 kilowatts and mission durations of 30 days to a year. There may, of course, be peak power requirements during periods of maximum activity that are considerably

In order to use such a reactor on board a manned spacecraft, however, a large, heavy shield must be placed between the reactor and the rest of the spacecraft. It is estimated that a shadow shield weighing 4000 to 7000 pounds would be required to provide adequate protection to men on board the same satellite as an operating SNAP 10A-type reactor.

Radioisotopes offer a relatively lightweight energy source. Like reactors, however, they require shielding to protect crew members from radiation. By selecting the proper radioisotopes, however, the amount of shielding required can be kept to a minimum. Two candidate radioisotopes are considered for manned space missions, polonium 210 and plutonium 238. The polonium isotope is much less expensive and more readily available, but it is limited to missions of roughly 90 days in length because of its short half-life (0.38 years). Plutonium 238, on the other hand, is more expensive and much scarcer, but has a very long half-life (87 years). Recent estimates of radioisotope heat source and shield weights for the class of missions under consideration indicate that a total weight of about 15 to 35 pounds per thermal kilowatt will be required for polonium heat sources (varying with geometry of the heat source and shield), and from 30 to 60 pounds per thermal kilowatt will be required for plutonium. This is equivalent to 375 to 875 pounds for a five kilowatt polonium system using a 20% efficient conversion device, and 750 to 1500 pounds for a similar plutonium heat source and shield.

On a weight basis alone, chemical energy sources must be eliminated for missions much in excess of thirty days. The exact trade-off point between chemical sources and others must, of course, be determined by careful study for any missions falling in the 20 to 45 day range. Solar energy sources for photovoltaic or thermal conversion methods are the lightest available. Nuclear reactors are attractive, but have excessively heavy shielding requirements. Radioisotope heat sources require considerably less shielding than nuclear reactors, although still weighing more than solar energy sources. Chemical sources and nuclear reactors will be eliminated from further consideration in this paper because of their excessive weight compared to other energy sources.

When compared on a cost basis, radioisotopes are less attractive than solar energy sources. The components needed to convert solar energy to heat are relatively inexpensive. Various cost estimates have centered around $250,000 per system. The cost for photovoltaic systems, like the weight, will not be considered here. Radioisotope energy source costs can be determined by calculating the amount of radioisotope needed at end of mission, allowing for decay during the mission and from time of encapsulation to launch, and multiplying this figure by the cost per thermal watt. Atomic Energy Commission cost estimates for radioisotopes vary with the quantity required and the time period in which it is required. For sake of comparison, however, the cost of polonium 210 can be considered to be $15 per thermal watt and the cost of plutonium 238 to be $1000 per thermal watt. Using these figures, a five kilowatt, 90 day mission using polonium 210 and a 20% efficient conversion device would require a heat source costing $750,000. A heat source for a longer mission using

plutonium 238 and a 20% efficient conversion device would cost $25,000,000. Some of this cost would be recovered if the isotope were returned and reprocessed for other uses.

In the area of vehicle integration, the simpler the operational mode of an energy source, the better. For solar photovoltaic systems, the solar panels must be kept within ± 15° of a normal to the sun in order to operate effectively. (Within a wide range, photovoltaic power falls off directly as the cosine of the angle to the solar normal.) For systems converting the sun's energy into heat with a parabolic mirror, orientation must be maintained within ± $\frac{1}{4}$ to $\frac{1}{2}$°. This imposes severe limits on the spacecraft operations. In addition, for both conversion methods, the large area exposed to the sun will create additional aerodynamic drag problems in low orbits. This will either limit the lifetime of the spacecraft or require that additional chemical propellants be stored on board to boost the spacecraft back up as needed. The radioisotope energy sources, on the other hand, can be designed into the vehicle, take up very little space, will not interfere with operating the vehicle, and impose no additional aerodynamic drag.

Both the energy output of the sun and the energy provided by decay of radioisotopes can be considered to be 100% reliable. Reliability considerations, then, come into count only in regard to those items in series between the energy source and the electrical power bus. For all solar energy systems, both photovoltaic and thermal, the principal reliability stumbling blocks are the deployment mechanicms and the orientation mechanisms. In both cases large structures must be erected in space and kept aimed toward the sun for the duration of the mission. Reliability of such operations is not high. With radioisotope energy sources, no such reliability problems arise. The heat is supplied to the energy conversion device regardless of what else happens.

From a safety standpoint, solar energy sources are much more attractive than radioisotope sources. Both the crew and the general population must be protected against any possible radiation hazard imposed by the radioisotopes. The problems in providing safe, routine space operations with radioisotope fuels have not been completely solved for the large amounts of radioisotopes required for manned missions. These problems are being attacked, however, and the results appear favorable. A detailed analysis of the safety and hazards aspects of polonium 210 and plutonium 238 is beyond the scope of this paper. Information on these problems is available elsewhere.

The cost, weight, and safety factors tend to favor solar energy. The vehicle integration and reliability factors favor use of radioisotopes. The final choice, of course, can depend on many factors. It must be made, however, by the program director for each mission in the class under consideration. From a program director's viewpoint, reliability for operation during the entire length of the mission must be the single most important criteria. The second most important criteria is minimum interference with the primary mission. Weight and cost are important

but relatively minor factors compared to the first two. Neglecting the safety problem, he would choose radioisotopes. The safety problem, however, is one that may not be in the program director's hands. If techniques are developed to provide for safe use of radioisotope fuels, then he can ignore this factor in making his choice. This factor will, in essence, be redistributed among the others, since safe operational use of radioisotopes will mean weight penalties here, costs there, and possible vehicle integration problems of some kind. On the other hand, if the techniques are not developed, the program director has no choice but to use solar energy.

In summary, then, the four possible energy sources have been examined. Chemical energy was ruled out because of its excessive weight for long duration missions. Nuclear reactors were ruled out because of the excessive weight of shields required for manned missions. After considering many factors, radioisotopes were chosen over solar energy based on reliability and vehicle integration advantages. It is assumed that continuing work on hazards and safety problems will permit use of large radioisotope energy sources in space.

CONVERSION DEVICES:

There are three means of converting thermal energy to electrical energy which can be considered for application to spacecraft electrical power systems in the near future. They are thermoelectric conversion, dynamic conversion, and thermionic conversion. Each of these methods will be examined in light of development cost and risk, operational cost, reliability and vehicle integration.

First, a brief description of each of these energy conversion techniques is in order.

Thermoelectric energy conversion makes use of semiconducting materials, such as lead-telluride and silicon-germanium. When these materials are joined in couples, and a temperature difference is maintained between the junctions of the couples, electrical power is generated. The power generated is proportional to the temperature difference maintained and the total heat flux passing through the semiconductors. It also is dependent on the properties of the specific semiconductor materials used. For most space applications, temperature differences of several hundred degrees Fahrenheit can be obtained. Overall efficiency of thermoelectric power systems is low, usually about five percent (electrical power out divided by end-of-life thermal power in). Advances in the state-of-the-art may increase overall efficiency to six to eight percent in the next few years. For the sake of comparisons in this paper, the conservative value of five percent efficiency will be used. For relatively large thermoelectric power systems, power to weight ratios of five watts per pound, excluding the heat source and shield, can be expected.

Dynamic energy conversion systems consist of a turbine and a generator,

together with the necessary fluid loops and heat exchangers necessary to convert thermal energy into electrical energy. A variety of designs, using various working fluids have been proposed. (The next section of this paper will treat these systems in more detail.) Using the best of these proposed systems, overall efficiencies as high as 27 percent have been calculated Actual efficiency and weight varies considerably with the different dynamic system concepts. For the purposes of comparisons in this section, it is assumed that a system will be developed with an overall efficiency of 20 percent and a weight varying linearly from 850 pounds at two kilowatts to 1500 pounds at ten kilowatts. (The next section will show that such a system is well within the state-of-the-art for development in the very near future.)

Thermionic energy conversion is very simple in concept. The common vacuum tube is the most familiar example of thermionic power. An emitter is heated to very nigh temperatures, typically above 2000°F. Electrons are boiled off and collected on a cooler element. In the process, a voltage is produced and power is generated. Efficiencies of about 15 percent and power to weight ratios of ten watts per pound appear reasonable for the near future, provided that the extremely high heat source temperatures required can be made available.

Table 2 summarizes the merits and demerits of the three conversion schemes relative to development cost, development risk, operational cost, reliability, and vehicle integration problems.

The present state-of-the-art in thermoelectric devices is sufficiently advanced that development costs would be relatively low. Thermoelectric elements were used for power conversion on the SNAP-10A nuclear reactor space test and worked satisfactorily. Both lead-telluride and silicon-germanium thermoelectric elements have been used in space. It is estimated that approximately $9 million would be required for the additional development program needed to produce five percent efficient systems utilizing radioisotope heat sources. It is estimated that the development cost of the radioisotope heat source for thermoelectric power conversion will be about $11 million, making a total power system development cost of $20 million. The radioisotope heat source required would be larger than any built to date for space use, but would require only a moderate development program since the capsule temperatures are not extremely high (1600 to 1700°F).

Development costs for dynamic systems will be somewhat higher than for thermoelectric systems. A variety of estimates have been made to determine the cost of completely developing, testing and qualifying a low-kilowatt dynamic space power conversion system. Although there are wide variations in these estimates, approximately $29 million is reasonable (11). The heat source development would be similar to that for thermoelectrics, since the temperature range involved is comparable. Thus, total development costs would be $40 million for a dynamic system.

The costs for developing a thermionic system are much less certain

since the state-of-the-art is less advanced. Various estimates indicate that $25 million should be sufficient to develop a thermionic system capable of using high temperature radioisotope heat sources. The heat source, however, represents a considerably more difficult development than the heat source for thermoelectric or dynamic systems. It is estimated that at least $35 million would be required to develop, test, and qualify a radioisotope heat source capable of providing the necessary heat output at over 2000°F for extended periods of time. The total development cost, then, comes to $60 million.

It is clear that development of thermoelectric systems is less expensive than development of dynamic systems or thermionics. The cost estimates cited above have been determined using the same bases. There may be some absolute errors involved in them, but the relative order of development cost appears to be quite firm.

The relative risk involved in developing, testing and qualifying a system cannot be defined quite so easily as development cost. In general, however, the two are related. The risk in being able to successfully develop a thermoelectric system is quite small, principally because other thermoelectric systems have been built in the past for a variety of uses, including spacecraft power systems. The materials problems involved are well known. The development program would consist mainly of applying known technology to a new design and qualifying it for new uses.

With dynamic systems, the development risk is slightly higher. No dynamic power system has yet been used in space, although there are several programs in being for the development of dynamic space power systems. There appear to be no fundamental problems in the way of successfully designing, testing, and qualifying such a system; but the job has not been done in the past. Problems in bearings, for both Rankine and Brayton cycle turbines, and in zero-g two phase heat transfer, for Rankine systems only, remain as engineering problems today. Their solutions appear to be feasible, but they have not yet been proven.

The development risks involved with thermionic systems are considerably higher. Individual thermionic diodes have been built and tested for periods of several thousand hours. Total experience with thermionic converters, however, is still quite limited. The basic problems are those of extremely high temperature materials. Proper material strength and sealing characteristics must be maintained under continuous operation at well over 2000°F. The problems can probably be solved, given sufficient time and resources; but the development risk remains considerably higher than for thermoelectric or dynamic systems.

In assessing the relative development risks, the heat source must also be considered. For both the dynamic and thermoelectric systems, only moderate heat source temperatures are required. If the capsule surface operates at 1600 to 1700°F, this is sufficient. For thermionics, the capsule must operate at 2000°F or higher. Taking into account all

possible accident modes, it can be readily shown that the abort criteria are limiting for the thermoelectric and dynamic systems, whereas the operating criteria tend to be limiting in the thermionic case. (The worst credible accident is usually one in which the radioisotope capsule is buried in an insulating layer of earth for long periods of time, resulting in heating of the capsule and build-up of internal pressure.) The development risk for a thermionic heat source, therefore, is somewhat greater than for a dynamic or thermoelectric heat source since more severe materials problems arise.

Considering both the conversion device and the heat source, then, the relative order of development risk involved in a program to provide two to ten kilowatts of power for long durations in space is thermoelectric, dynamic, and thermionic, with the least risk system given first.

Unit cost, weight, and reliability criteria can all be reduced to dollar figures. It is significant to determine the total operational cost for the various conversion devices. Actual cost of the conversion hardware and the heat source can be estimated easily. The weight of the entire power system can be translated into a dollar figure by determining the cost of putting it into orbit. For this purpose, a figure of $500 per pound in orbit is used. This figure is based on use of Titan-III boosters and low-earth orbits. Finally, reliability goals in all three of the systems being considered can be met by use of redundant components. In the case of thermoelectrics and thermionics, these consist of addition of extra converters. In the case of dynamic systems, reliability is obtained by providing redundant turbines and alternators. The weight of these redundant components can be determined and the cost of boosting them into orbit added to the total operational cost.

Table 3 presents both the development and operational costs for a typical series of ten five-kilowatt missions, each lasting 90 days, fueled with polonium 210. Such a series of missions could extend over a period of four to ten years. Redundant component weights are included. It is apparent that the lower development costs of thermoelectric are more than offset b y the much higher unit costs of the heat source and the heavier overall system weight when compared to the other two choices. The edge which dynamic systems hold over thermionic systems lies principally in the lower development costs. Table 4 presents a similar comparison for two five-kilowatt missions, each lasting two years, using plutonium 238 as the fuel. It is assumed that the same plutonium 238 is used in both missions (i.e., unit costs of the heat source have been reduced by 50 percent). Again, the cost of the heat source predominates.

The combined factors of unit operational cost, unit weight, and reliability strongly favor dynamic systems. Even though individual components of dynamic systems may cost more, be heavier, or less reliable, the overall system will cost considerably less after just a few launches.

The vehicle integration problems involved tend to offset one another.

No one of the three conversion systems appears to offer any distinct advantage over the others. With thermoelectric systems, the shear volume and mass of the system present considerable design problems to the spacecraft developer. In using dynamic systems, problems involving start-up and shut-down must be faced. In addition, valuable crew time is required for replacement of conversion units. With thermionic systems, the most significant problem is the very high temperature heat source which must be carried and the high temperature radiator which must be built into the spacecraft. In all three cases, engineering solutions can be found to solve the vehicle integration problems.

The basic choice of conversion mechanism then boils down to a choice between a lower cost, lower risk development of thermoelectric systems, with inherently high operating costs; a moderate cost, moderate risk development program for dynamic systems leading to extensive operational savings after a few missions; and a high risk, high cost development program for thermionics with marginal advantages, if any, over dynamic systems. Based on the extensive overall cost savings for the same utility in space, dynamic systems are chosen over thermoelectrics.

DYNAMIC ENERGY CONVERSION SYSTEMS:

There are five candidate dynamic energy conversion system concepts which may be considered for applications in the low-kilowatt range: the Rankine cycle, using water, mercury, or organic working fluids; the Brayton cycle; and the Stirling cycle. Following a brief description of each of these concepts, they will be compared in some detail and one system selected. The criteria on which the selection will be based include flexibility, maintainability, simplicity, weight, reliability, development status, anticipated development problems, and operational cost.

In all five of these power system concepts, the heat from the radioisotope heat source is transferred to the working fluid, which in turn drives a turbine or engine. The turbine or engine is connected mechanically to an alternator, which produces electrical power. The waste cycle heat must be dissipated through a radiator, either directly to space or through a secondary cooling loop.

The power output of an ideal turbine is given by the enthalpy change in the fluid as it passes through the turbine. This can be expressed as

$$P = \dot{H} = \underline{H}\dot{m}$$

where P = the ideal turbine output power

$\dot{H}$ = the enthalpy change in the fluid as it passes through the turbine

$\underline{H}$ = the enthalpy per unit mass of fluid

$\dot{m}$ = the mass rate of flow of fluid

It is clear from this equation that to vary the output of a given ideal turbine, either the mass rate of flow or the enthalpy per unit mass of fluid must be changed. The efficiency of a real turbine is determined, in part, by the ratio of the inlet to outlet pressures. This ratio must be maintained nearly constant to obtain maximum efficiency from a given design.

The ideal Rankine cycle is characterized by boiling the working fluid at constant pressure, expanding the vapor through a turbine at constant entropy, condensing the vapor at constant pressure, and pumping the liquid back to the original state at constant entropy. In practice, losses are induced by turbines and pumps which are not entirely isentropic and by fluid flow friction losses.

In the water Rankine system, the fluid is superheated considerably into the vapor region to obtain reasonable specific volume (cubic feet per pound of fluid). Typical turbine inlet conditions are 1200°F at 1200 psia. Outlet conditions for a typical system would be 870°F at 186 psia. The thermodynamic properties of water lead to use of very small turbines for low-kilowatt power systems. In order to accomodate proper volumetric flow rates and maintain reasonable efficiency, extremely small turbine blades and blade clearances must be used. This accounts for the relatively low efficiency attainable in the two to ten kilowatt range. (See Table 5)

Using mercury instead of water in a Rankine cycle system solves some problems and creates others. The specific heat of mercury vapor is much less than that of water, thus higher flow rates and greater machine tolerances can be permitted. As a material, however, mercury is quite corrosive. As with water, some superheating of the vapor is required, principally to insure that there will not be excess condensation of mercury as it passes through the turbine. Typical operating parameters for the turbine are 1250°F, 350 psia inlet, and 605°F, 7 psia outlet. Efficiencies of mercury systems fall between 11 to 13.5 percent over the two to ten kilowatt power output range.

With organic Rankine systems, the operating characteristics change somewhat. Organic fluids, such as biphenyl and Dowtherm-A, have a positive saturated-vapor curve on a temperature-entropy diagram. The result of this is that the saturated vapor can be expanded directly through the turbine with no superheat. The vapor expands into a region of superheat, eliminating the possibility of condensing in the turbine, one of the principal problems in turbine design and lifetime. Organic systems are limited principally by the inability of working fluids to withstand high temperatures for extended periods of time without decomposing. For Dowtherm-A or biphenyl, 700°F appears to be maximum temperature limit for practical systems. Typical turbine conditions are 700°F, 110 psia inlet, and 580°F, 0.5 psia outlet. Efficiency ranges from 13 to 17 percent for the entire system.

In both the water and organic Rankine cycles, a regenerator is often

used for heat transfer from the superheated vapor leaving the turbine to the liquid leaving the pump. In this manner, some of the waste cycle heat is recovered. Theoretically, only the heat of condensation at the turbine outlet pressure need be radiated to space; all sensible heat in the vapor can be recovered.

The Brayton cycle uses a non-condensing inert gas working fluid, typically argon or helium/xenon mixtures. In an ideal cycle, the gas is heated at constant pressure, expands through a turbine at constant entropy, gives up its waste heat at constant pressure and is compressed to the starting pressure at constant entropy. In virtually all Brayton cycle systems, a recuperator is used to recover much of waste heat from the turbine exhaust gas. In a typical system using a radioisotope heat source, the turbine inlet conditions will be 1500°F, 45 psia, and outlet will be 1100°F, 24 psia.

The Stirling cycle system employs a reciprocating engine instead of a turbine. It consists of two pistons synchronized within the same cylinder to provide constant volume heat transfer, both input and output, and constant pressure volume change in the working fluid. Ideally, the machine closely approximates a Carnot cycle. In practice, of course, it fails to perform quite so well because of finite heat transfer rates and internal friction. Light-weight inert gases are used as the working fluid.

Power system flexibility is defined as the ability to operate effectively at other than design-point conditions. All five of the dynamic concepts under consideration can be operated over the entire two to ten kilowatt range of interest when coupled with the proper alternator. In some cases, however, the overall system efficiency falls off rapidly with deviations from design-point operation. Returning to the earlier discussion of turbine output, it can be readily shown that a turbine will operate near its efficiency peak only when the mass flow rate is varied, but the volumetric flow rate and the inlet/outlet pressure ratio are maintained nearly constant. This cannot be done with Rankine cycle systems operating anywhere near the saturated vapor line since pressure and specific volume cannot be regulated independently. For the Brayton cycle turbine and the Stirling cycle reciprocating engine, however, this is not the case. Operation is entirely within the gaseous regime; therefore, pressure ratios and volumetric flow rates can be maintained constant while changing the mass flow rate. By this means, it is quite easy to obtain a five to one variation in power level with Brayton and Stirling systems while still maintaining fairly high efficiency. With any of the Rankine systems, however, if efficiency is to remain reasonably constant, the maximum change in power level permissable is approximately two and one-half to one. (See Table 5)

Ease of maintainability on orbit and on the launch pad is determined principally by the problem of replacing a turbine/alternator combination, or, in the case of the Stirling engine, an engine/alternator combination. Since the Brayton and Stirling cycle systems would require changing only

inert gas and electrical connections, they would be much easier to maintain than the Rankine cycle systems, which would require changing liquid and gas lines as well as electrical connections. Further, in the case of the mercury system, the fluid is highly toxic; and in the organic system, the most common fluids have extremely unpleasant odors, even though they are not toxic in small amounts.

The Brayton cycle system is a considerably more simple system than any of the others in its operation. There are no problems of zero-g, two-phase fluid flow and zero-g, two-phase heat transfer as are encountered in the boilers and condensers of all three Rankine cycle systems. The complexity of the Stirling engine is its single greatest drawback. A series of pistons, rapid-rate heat exchangers, and precisely operating valves must all be carefully synchronized.

There have been a number of detailed weight estimates for the five systems being compared. In general, the weight of the power conversion equipment (turbine or engine and alternator) does not vary considerably from system to system. The weights of the heat source, heat exchangers, and radiator, however, vary inversely with the overall system efficiency. On a weight basis, therefore, the Brayton and Stirling systems are advantageous.

Reliability of the five systems must be estimated, since no in-space operating data is yet available for any of them. Based on ground tests results and the relative complexity of the systems, it appears that all except the Stirling concept would have comparable high reliability. All four of the other systems would probably require the use of a redundant turbine/alternator unit in order to meet the reliability criteria usually established for manned spacecraft.

To date, far more effort and resources have been placed on the mercury Rankine system than on any other. This system has been developed through ground test in conjunction with the Sunflower, SNAP-2 and SNAP-8 programs, with the principal development work carried out by TRW and Aerojet General. Turbine/alternator units have been built, run and tested for over 3000 hours in endurance runs and for tens of thousands of hours total running time. The development problems appear to be well in hand, with the exception of basic knowledge in zero-g boiling and condensing. The organic Rankine system has been developed through component and breadboard testing, principally by Sundstrand Aviation. Results have been favorable, but there are still basic questions involving long-term working fluid stability, off-design operation, and zero-g boiling and condensing. The water Rankine system for space has received comparatively little attention except in a few studies. Although extensive experience in large water Rankine systems for ground power stations exists, very little space-oriented development has been accomplished. The Brayton cycle has been receiving considerable interest in the past two to three years. The NASA Lewis Research Center has sponsored a broad exploratory development program, involving their own personnel and facilities as well as several contractors - Garrett

Corporation, Pratt and Whitney, and others. In addition, Garrett Corporation has done extensive development and testing of Brayton cycle components under contract to the Air Force and under their own sponsorship. The principal uncertainty still remaining is in the long term performance of gas bearings under load. The Stirling engine for space use has been investigated by Allison Division of General Motors under Air Force sponsorship. It has generally been found to be too complex a device for long-term operation in the hostile space environment.

On the basis of flexibility, maintainability, and weight, the Brayton and Stirling cycle concepts appear to have a significant edge overthe Rankine cycle concepts. In considering system simplicity, the Brayton cycle is simplest, followed by the three Rankine cycle concepts, which are in turn followed by the complex Stirling cycle concept. Reliability for any of the four turbine systems is about the same, with the Stirling engine trailing behind. The mercury Rankine system is the furthest developed today, with the Brayton cycle and organic Rankine cycle systems falling next. Very little applicable development of Stirling engines or water Rankine systems has been performed. The foreseeable development problems for the Brayton cycle and the organic and mercury Rankine cycles are comparable. It is safe at this point, then, to rule out the Stirling cycle and water Rankine systems as contenders for two to ten kilowatt space power systems for development in the near future.

Figure 1 presents the influence of overall system efficiency on a typical series of missions. Again, the model of ten 90-day five-kilowatt missions using Po^{210} fuel has been used. It is evident that a considerably savings in dollars will be realized by choosing the most efficient system practical. In using Pu^{238} as a fuel, the cost/efficiency curve is much steeper and favors the Brayton cycle even more. Using the same criteria as Table 3, it is interesting to note that, near 20 percent overall efficiency, a gain of one percent in system efficiency at a cost of up to 100 pounds in weight will cost less after just ten 90-day, Po^{210} missions. For this reason, the Brayton cycle system is selected as having the best overall potential from a development and use standpoint.

SUMMARY:

In this paper, we have tried to show briefly that there is a need for spacecraft electrical power systems capable of providing from two to ten kilowatts of power for extended durations. For reasons of reliability and system integration, radioisotopes are chosen as the best energy source. After taking into consideration weight, development cost, operational cost, reliability, and many other factors, it is shown that the best means for converting the energy to useable electrical power is with a Brayton cycle dynamic energy conversion system.

Of necessity, many salient points have been touched on only very briefly. This paper is meant to show the basic methodology involved and the series of decisions which must be made, rather than to include all the detail necessary to make such decisions. The interested reader is referred to the many excellent system comparisons and trade-off studies contained in the references.

REFERENCES

1. A. Lowi, Jr., "Solar Dynamic Power Systems Study, Vol II, Solar Dynamic Power Systems Evaluation," Aerospace Corporation, SSD-TDR-63-115, Vol II, 15 September 1965 (Unclassified)

2. L. M. Hedgepath, G. D. Huffman, and F. R. Ostdiek, "Comparisons of Weights and Performances of Solar Dynamic Energy Conversion Systems," Air Force Aero Propulsion Laboratory, AFAPL-TR-65-44, September 1965 (Unclassified)

3. T. E. Williams, et al, ""Organic/Rankine Cycle Auxiliary Power Systems, Technical Seminar," Sundstrand Aviation, August 1965 (Unclassified)

TABLE 1

ENERGY SOURCE COMPARISON

Energy Source \ Criteria	Weight	Cost	Vehicle Integration Problems	Reliability	Safety
Chemical	High	Low	Low	Medium	High
Solar	Low	Low	High	Medium	High
Nuclear Reactor	High	Medium	Medium	Medium	Medium
Radioisotope	Medium	Medium to High	Low	High	Low

TABLE 2

ENERGY CONVERSION DEVICE COMPARISON

Conversion Device \ Criteria	Development Cost	Development Risk	Operational Cost	Reliability	Vehicle Integration Problems
Thermoelectrics	Low	Low	High	High	Medium
Dynamic Systems	Medium	Medium	Low	Medium	Medium
Thermionics	High	High	Low	High	Medium

TABLE 3

TOTAL POWER SYSTEM COSTS

10 5-KILOWATT 90-DAY MISSIONS USING Po^{210}

(Millions of Dollars)

	Thermoelectric	Dynamic	Thermionic
R&E Costs			
Energy Source	11.0	11.0	35.0
Conversion Device	9.0	29.0	25.0
Total	20.0	40.0	60.0
Operational Costs Per Mission			
Energy Source	3.00	0.75	1.00
Conversion Device	0.50	0.50	0.50
Launch (at $500/#)	2.50	1.05	0.92
Total	6.00	2.30	2.42
Total Operational Costs for 10 Missions	60.0	23.0	24.2
Total Power System Costs	80.0	63.0	84.2

TABLE 4

TOTAL POWER SYSTEM COSTS

2 5-KILOWATT 2-YEAR MISSIONS USING Pu^{238}

(Millions of Dollars)

	Thermoelectric	Dynamic	Thermionic
R&D Costs			
Energy Source	11.0	11.0	35.0
Conversion Device	9.0	29.0	25.0
Total	20.0	40.0	60.0
Operational Costs Per Mission			
Energy Source (Re-used)	50.00	12.50	16.65
Conversion Device	0.50	0.50	0.50
Launch (at $500/#)	3.00	1.18	1.08
	53.50	14.18	18.23
Total Operational Costs for 10 Missions	107.0	28.4	36.5
Total Power System Costs	127.0	68.4	96.5

TABLE 5

DYNAMIC ENERGY CONVERSION SYSTEMS

Criteria / Cycle	Efficiency (%)	Power Range Flexibility	Development Risk	Maintainability
Mercury Rankine	11 - 13.5	2.5:1	Medium	Low
Water Rankine	7 - 11	2:1	High	High
Organic Rankine	13 - 17	2.5:1	Medium	Medium
Brayton	21 - 26	5:1	Medium	High
Stirling	21 - 27	5:1	High	Medium

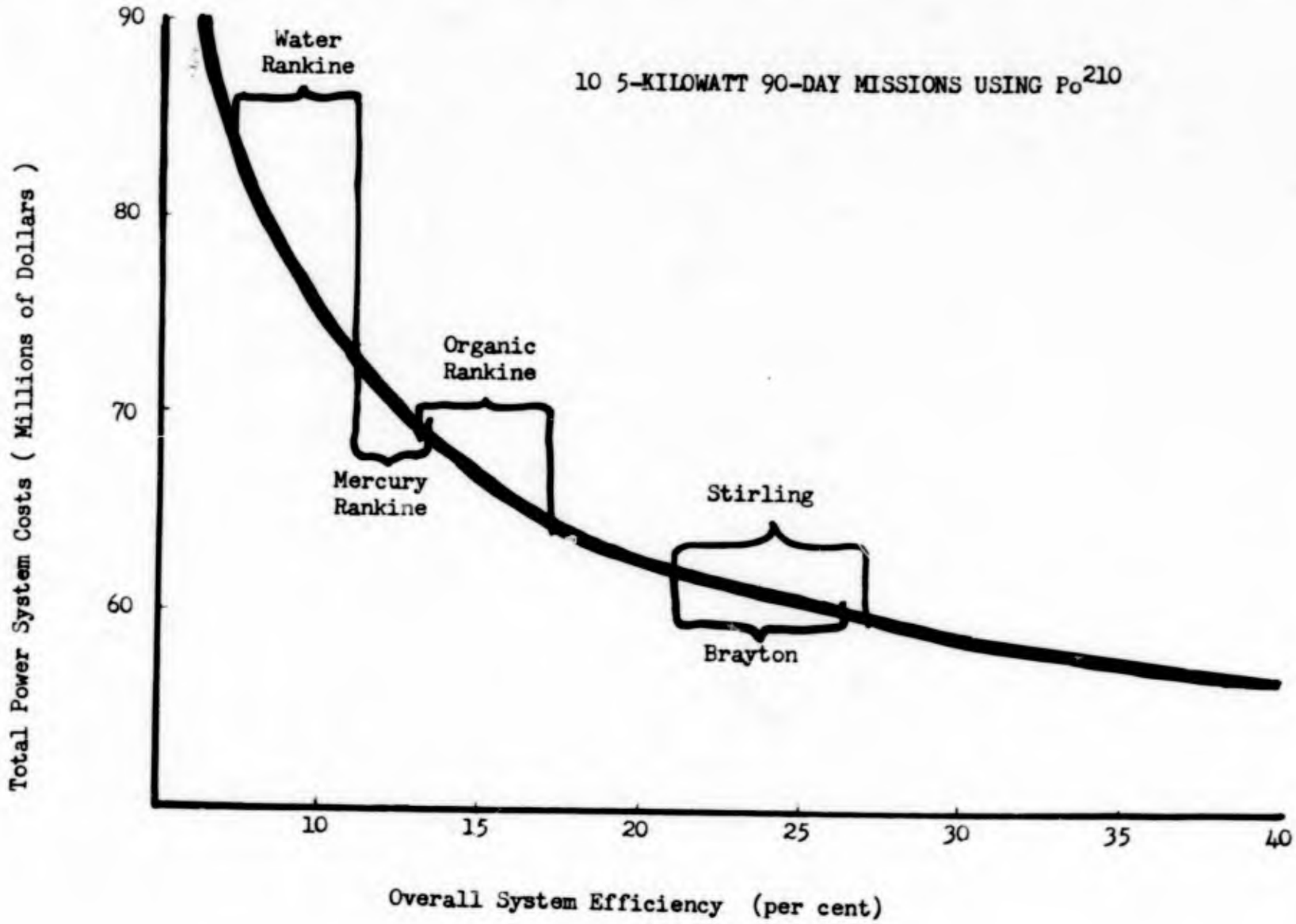


FIGURE 1 - POWER SYSTEM COST AND EFFICIENCY

(U) GEMINI PERFORMANCE OPTIMIZATION

by

John D. Regenhardt, Capt, USAF

ABSTRACT

Gemini Launch Vehicle
System Program Office
Space Systems Division
Los Angeles AF Station, Calif.

Capt. John D. Regenhardt

BIOGRAPHY

Capt Regenhardt was born 31 years ago in Marion, Illinois; after attending elementary and high schools in Cape Girardeau, Missouri, he received an appointment to the U.S. Naval Academy in 1953. He graduated with a BS degree in general engineering and was commissioned in the Air Force in June 1957. He spent the next year at an Air Force Guided Missile Maintenance Officer School. From this school he was assigned to the United States' first operational ICBM squadron, the 576th Strategic Missile Squadron at Vandenberg AFB, California. He participated in both Thor/Discoverer and Atlas operations while serving in various organizations at Vandenberg. In 1960 Capt Regenhardt began graduate studies at MIT under the sponsorship of the Air Force Institute of Technology. He was awarded an MS degree in Aeronautics and Astronautics in 1962. His thesis was titled, "Star Occultation Measurements as an Aid to Navigation in Cir-Lunar Space." Following five months as an in-service engineer at Vandenberg AFB, Capt Regenhardt was transferred to the Space Systems Division of the Air Force Systems Command. From December 1962 until the present he has been a project officer in the Operations Directorate of the Gemini Launch Vehicle System Program Office.

ABSTRACT

Unburned propellants carried into orbit in Stage II of the Gemini booster represent wasted payload. With each mission costing many millions of dollars, the Air Force and NASA are obligated to maximize the productivity of each flight. To this end a new technique has been developed on the Gemini program. This technique is known as Real Time Performance Computation (RTPC). The purpose of the RTPC system is to provide accurate booster performance information during the countdown. This information is used in several ways. First, it makes accurate performance information available to the Mission Director and the Launch Director. Thus, their decisions concerning performance are based on the most accurate data available rather than estimates.

In addition, the RTPC system generates data that enables the launch crew to take appropriate action to improve performance. The idea in this case is to regulate propellant temperatures by using insulating wraps so that the optimum temperature combinations exist at lift-off. Finally, the RTPC system keeps track of the ullage volumes. Again, positive action can be taken to prevent the violation of an ullage limit.

In summary, the RTPC system has proved very beneficial to the Gemini program. It has provided the equivalent of significant performance improvements without adding to the program cost. Based on the Gemini experience, all large liquid space boosters would benefit from a technique similar to RTPC.

Chapter 1

Introduction

Unburned propellants carried into orbit in Stage II of the Gemini booster represent wasted payload. At the current cost of orbiting payloads it is imperative that this waste be minimized. All booster programs have had to concern themselves with payload capability. This concern usually brings about four distinct phases of action. First, there is the design phase. This involves the well understood area of sizing the booster to meet the payload requirements. At this point some allowance is usually made for payload growth. The second phase is the engineering change phase. This phase is necessary only if payload growth uses up the original allowance and then requires additional booster capability. In this case, weight saving and similar performance improvements are obviously required. That payload will grow to meet the booster capability must be a corollary to the axiom that expenditures will rise to meet income. This change phase was required on the Gemini program. Another possibility during the change period is that a design change, such as a structural beef-up, will degrade performance and then require a companion performance improvement change. The third phase is the pre-flight phase. With actual values of payload weight, booster dry weight, engine performance and other data available a final check is made to insure that booster performance is adequate. Finally, flight data is used to reconstruct the trajectory and to determine achieved booster performance. To these four phases, the Gemini program added a fifth. This is the Real Time Performance Computation (RTPC) Phase. RTPC and its benefits are the subject of this paper. RTPC is a function that takes place from the time propellant loading is completed (approximately lift-off - 9 hours) until launch. During this period more and more parameters become fixed. This makes it possible for the RTPC team to provide increasingly accurate performance data. The need for this data and its uses will be described later.

Even on a manned program such as Gemini costs play a major role in decision making. In the payload area cost considerations were especially prominent. Figure 1 compares the costs of three booster programs. These three were chosen because they are all liquid propellant space boosters that were developed as ballistic missiles. This allows me to treat each the same way in arriving at cost per payload pound figures. The ballistic missile development costs are not considered. The costs used to develop Figure 1 are based on prices for the production of these space boosters. Launch services are also included.

As mentioned previously, the Gemini booster required engineering changes in order to meet the demands of increasing spacecraft weight. Figure 2 illustrates the history of spacecraft weight and booster payload capability.

Although booster capability never drops below spacecraft weight, it is apparent that action was taken in 1963 which resulted in significant

performance improvements by late 1964. Had this not been done, the spacecraft weight (which grows at an average rate of 34 lbs/month) would have exceeded the average booster capability by mid-1964. When it became evident that performance improvement engineering changes were required to meet the spacecraft requirement, the Martin Company was asked to perform a weight savings study. The schedule and reliability aspect of the proposed changes will not be discussed in this paper. A total of thirty-eight changes were studied. The estimated cost per pound of payload improvement for the changes ranged from 850 dollars per pound to 77,500 dollars per pound. It should be noted that the cost per pound rated for changes is computed on the basis of cost of the change for all boosters and payload gain per booster. Since the costs used to devise Figure 1 are costs per booster the cost per pound for booster procurement is not directly comparable with the cost per pound ratio for changes. Providing schedule and reliability effects were not prohibitive, Gemini performance improvement changes were selected on the basis of their cost per pound ratio. Figure 3 compares the cost per pound ratio of some of the changes selected for implementation on the Gemini booster.

To this point no mention has been made of the statistical nature of booster performance estimates and the variable nature of booster performance requirements. In reality there is a statistical distribution which describes the booster performance. This distribution is nearly normal, and where I have used the terminology "booster performance" or "booster performance capability" the three sigma low value of booster performance is meant. Quoting a manned booster's three sigma low performance or its performance capability began with the Mercury program. This procedure was carried over into the Gemini program.

The performance requirements of the Gemini booster vary from mission to mission for reasons other than payload weight changes. The Gemini V mission required an apogee altitude of 189 nautical miles. The Gemini VI mission called for an apogee of 146 nautical miles. From a performance standpoint, using the 146 nautical mile apogee as a reference, this increase in apogee altitude of 43 nautical miles is equivalent to 112 pounds of additional payload. To provide a common base for comparison, mission changes are usually converted into their payload equivalent. Launch azimuth is another mission parameter which affects the booster payload capability.

There is another factor which has a profound effect on payload capability. This factor is yaw steering or "doglegging." It is a yaw maneuver designed to insert the spacecraft directly into the target orbit. Since the spatial relationship between the launch site and the target orbit varies continuously during the launch window, the amount of yaw steering varies and thus the payload capability varies with time. This effect on payload ties in closely with the RTPC effort and has a marked effect on the RTPC results.

Chapter II

Real Time Performance Computation

It has been pointed out that orbiting a manned payload is an expensive undertaking. It logically follows that any unused booster performance capability constitutes an expensive waste. Thus the goal of those responsible for the Gemini booster's performance has been to obtain the maximum booster performance, consistent with spacecraft requirements, cost and booster reliability. RTPC is an important tool in the achievement of this goal. It provides accurate, current performance data during the countdown. RTPC has two obvious benefits. First, it provides data which enables the launch crew to take measures to improve performance by adjusting erector curtains and/or wrapping portions of the booster with polyethylene. This will be described later. The second advantage of RTPC is that it lets us reduce the safety factor we allow between spacecraft weight and booster performance. The easiest way to explain these advantages is to explain the RTPC system. I will do this by describing the steps that take place in their chronological order.

A. Determine specific booster design payload capability - This is accomplished by digital computer. The program incorporates the six degree of freedom differential equations of motion of the booster. Nominal Stage I open loop guidance is simulated in the program. This simulation is tempered with test table data on the specific vehicle's flight control and guidance components. The Stage I simulation runs until the nominal outage conditions (based on a nominal propellant loading for the specific vehicle) are reached. Thrust, specific impulse and mixture ratio test stand data on the installed Stage I and Stage II engines is factored into the computations. At this point the effects of the Stage I shutdown transient and the Stage II start transient are computed. During the Stage II portion of the computation the effects of closed loop guidance are included. Performance is determined by the amount of propellants available above nominal Stage II outage. The mission payload weight is used during this run. This solution is checked by converting the excess performance into equivalent payload and making a new run with this maximum payload. The result of this step is a value for a specific booster's predicted, nominal payload capability, P_N(lbs).

B. Determine specific booster three sigma dispersed performance - The following is a list of parameters which make significant contributions to the overall performance dispersion.

(1) Thrust (Stages I and II)
(2) Specific Impulse (Stages I and II)
(3) Outage (Stages I and II, includes mixture ratio effect)
(4) Dry Weight (Stages I and II)
(5) Atmospheric Density (Stages I and II)
(6) Winds (Stages I and II)
(7) Thrust Vector Misalignment, Pitch (Stages I and II)

(8) Thrust Vector Misalignment, Yaw (Stages I and II)
(9) Thrust Vector Misalignment, Roll (Stages I and II)
(10) Pitch Programmer Error (Stage II)
(11) Pitch Gyro Drift (Stage I)

The distribution of the above parameters is assumed to be normal except for outage. The three sigma value for each parameter is determined except for outage where the 0.99 value is used. The effect of these dispersions is determined by making a performance computation as described in A above for each parameter with the parameter set at its three sigma value and the proper correlation introduced in the remaining parameters. In this way the resulting effect on performance for each parameter is determined. These results are then root sum squared to give the overall three sigma value for booster performance. This is normally done in terms of the difference from nominal performance. While not mathematically rigorous this technique has proved adequate. It has the advantage of taking much less computer time than the more rigorous Monte Carlo technique. It should be noted that a $-2^{o}(F)$ temperature tolerance is assumed in the propellant loading and that this effect on outage is included. The result of this step is the specific booster's three sigma performance difference with respect to normal performance, $\Delta P_{3\sigma}(lbs)$.

C. Determine specific booster performance margin - In order to accomplish this step the spacecraft weight, $W_{S/C}(lbs)$, for the specific booster must be known. The result of this step is performance margin, $P_M(lbs)$. It is determined as follows.

(1) $P_M = P_N - \Delta P_{3\sigma} - W_{S/C}$

The statement was made in Chapter I that the term "booster performance" is used throughout this paper to mean three sigma low booster performance, $P_{-3\sigma}(lbs)$. It follows that

(2) $P_{-3\sigma} = P_N - \Delta P_{3\sigma}$

It should be noted that P_M is negative whenever the spacecraft weight exceeds the booster performance capability.

D. Make final spacecraft weight adjustment - The value of $P_{-3\sigma}$ is submitted to the National Aeronautic and Space Administration's Gemini Program Office (NASA's GPO) forty-five days prior to each mission. Using this information NASA can make a final adjustment to $W_{S/C}$. This could be done by changing the number of experiments carried or by varying the amount of Orbital Attitude and Maneuvering System (OAMS) propellant carried. This T - 45 day transmission of $P_{-3\sigma}$ enables NASA GPO to take full advantage of the booster's payload capability.

E. Launch operations

1. Condition propellants - The Gemini booster engines are volumetric devices, i.e., they consume fuel and oxidizer at a fixed volumetric ratio.

Therefore it is most efficient to load usable propellants in this same volumetric ratio. Since the propellants are normally loaded at less than ambient temperature, heating takes place from the time the propellants leave the Ready Storage Vessels (RSVs) and the associated refrigerating system until they are consumed in the engines. It has been our practice to optimize the propellant temperatures for a launch in the middle of the launch window, i.e., at this time the loaded propellant temperatures are such that the usable volumes match the engine mixture ratios. (I have not considered the Stage II fuel bias which exists to protect against a Stage II fuel depletion shutdown). RSV temperatures can be determined which will result in optimum loaded temperatures at the specified time. The initial RSV temperatures are compensated for the heat transfer which takes place as the propellants flow from the RSVs to the booster tanks and that which takes place during the approximately eight hours that the loaded booster sits on the launch pad. Therefore, the first step in the launch operation is to condition propellants to the proper temperatures in the RSVs.

2. Load propellants - Propellant loading normally begins at T - 12 hours. The entire operation takes three to four hours. It is apparent that conditioning propellants to the proper temperature serves no useful purpose if the propellants are not loaded accurately. The Gemini program uses modified petroleum industry flowmeters (volumetric devices) checked by tank liquid level sensors on the booster to accomplish the loading. The system specification for accuracy is 0.3 percent.

3. Measure loaded temperatures - Each booster propellant tank is equipped with a temperature transducer. As soon as the propellants are loaded, temperature readings are taken. From this point temperature readings are taken every half hour until launch. These temperatures are both tabulated and plotted. The use of these tabulations and plots will be described later. Figure 4 is a plot of temperature vs. time in one of the tanks.

4. Perform digital computation - Tank temperatures are relayed to the Martin Company's Baltimore, Maryland, plant. The Cape Kennedy weather forecast is sent to Baltimore by Datafax. In addition to the preceding data, the boost phase flight parameters and specific engine data are loaded into the Martin Company's computer. Two of the program's outputs are transmitted to a Martin Company engineering team at Cape Kennedy. Predicted propellant temperatures are sent to "the Cape" so they can be compared to actual propellant temperatures. This allows the Cape engineering group to check the program's temperature prediction solution. Predicted temperatures are plotted along with measured temperatures in Figure 4. This temperature information determines the ullage limit timer. This will be discussed later. Predicted temperatures at the beginning and end of the launch window are plotted on a mixture ratio graph. Figure 5 is an example of such a graph. Due to the specific heats of the fuel and oxidizer and their exposed heat transfer areas, the slope of the line which connects the window opening and window closing temperatures is greater than the slope of the optimum mixture ratio line. In general the practice on

the Gemini program has been to target the RSV temperatures so that the mixture ratio will be optimized at the middle of the window. The type of presentation shown in Figure 5 makes it easy to see what to do if the temperatures are not in the desired relationship. The target line represents a good temperature relationship assuring that optimum mixture ratio is desired at mid-window. The Gemini IX line represents the case where the fuel temperature is too high relative to the oxidizer temperature. In this case insulating the fuel tank and opening the Stage II erector curtain to increase the oxidizer heating rate would improve performance. The other information transmitted to the Cape engineering group is the end product of the RTPC effort. It is a plot of performance margin, P_M vs. time in the launch window. Figure 6 is an example of the data which is datafaxed from Martin/Baltimore to the Cape engineering group.

The dashed line represents a 105 degree launch azimuth. This information is used at the NASA Houston Mission Control Center (MCCH) by the slow malfunction monitors. The positive slope of the 105 degree line demonstrates the temperature effects when the mission parameters are fixed. Figure 6 corresponds to the case illustrated by the Gemini IX line in Fig. 5, i.e., the fuel temperature is high relative to the oxidizer temperature. The solid line in Figure 6 represents the actual P_M vs. time relationship. Here the effects of launch azimuth and yaw steering are evident. At the opening of the window the launch azimuth is 97.8 degrees. The 43 lbs difference between the actual P_M and the 105^o case is mainly due to the fact that the 97.8 degree launch azimuth makes use of a larger component of the earth's velocity. The rapid drop in P_M to -305 lbs is due to an increasing launch azimuth and a rapid increase in the amount of yaw steering required. At approximately fifteen minutes into the window the range safety limit of 105 degrees is reached. From this point the guidance targeting equations constrain the launch azimuth to 105 degrees but continue to provide for yaw steering parallel to the target plane until such time as a 105 degree launch provides an insertion parallel to the target plane with no yaw steering. The effect of decreasing the amount of yaw steering is reflected in P_M from 15 minutes until one hour into the window. At this point the mission parameters have become fixed and the major factor affecting P_M is propellant temperature. The RTPC program is run at intervals during the countdown to take advantage of later tank temperature data, later weather predictions and current erector curtain and tank insulation configurations.

5. Target orbit determination - The target spacecraft is launched approximately 100 minutes prior to the Gemini spacecraft. At approximately T - 35 minutes in the Gemini count, the actual target ephemeris is computed at MCCH and relayed to the Martin Company's performance engineering group at Cape Kennedy.

6. P_M adjustment based on actual target ephemeris - Performance calculations to this point have been based on a set of nominal target ephemeris data. The target ephemeris affects P_M through its effect on launch azimuth and the amount of yaw steering required. Upon receipt of the actual target ephemeris, corrections are made to the P_M vs. time plot.

This correction is made by manual interpolations on pre-prepared graphs.

7. Final report to the Mission Director at MCCH and the Launch Director in the blockhouse - Plots similar to Figure 6 are transmitted by datafax to MCCH at regular intervals during the countdown. The last datafax transmission is normally made at T - 90 minutes in the Gemini count. The blockhouse is kept advised of the performance situation during the count by hot-line. A final update, including the correction for actual target ephemeris is given to MCCH and the blockhouse at approximately T - 30 minutes. Two comments are appropriate at this point. First, this paper has employed $P_{-3\sigma}$ as a reference point. This technique is consistent with the actual method employed on the Gemini program. However, a stated value for P_M means very little to most people in terms of mission success probability. The mission director and the launch director decide to launch or not to launch based on the value of P_M for the time that they intend to launch. They are concerned with a successful insertion by the booster. There is a relationship between P_M and the probability of a successful insertion by the booster. This relationship is shown in Figure 7. A good case can be made for reporting to the Mission Director and the Launch Director in terms of insertion success probability vs. time instead of P_M vs. time. However, graphs similar to Figure 7 are available at locations to which the P_M vs. time data is transmitted and the conversion to a success probability can be made following receipt. Thus, the commitment to launch is made with full knowledge of odds as far as performance is concerned. The second comment concerns the individual propellant tank ullage limits. A minimum ullage volume is established for each tank. As the propellants warm up the resultant expansion will eventually violate the ullage limits. Since an adequate ullage volume is mandatory to prevent tank collapse during the engine start transient, a violation of the ullage limit results in a booster "No-Go." The RTPC system predicts the time at which the ullage limits will be violated. This information is also transmitted to the Mission Director and the Launch Director as illustrated by Figure 6. It could be very important in determining any allowable "hold duration." This is especially important during the summer months when the propellant heating rate is high.

Chapter III

Results and Conclusions

Results

The results of incorporating a Real Time Performance Computation system on the Gemini program are not precisely measurable. However, it is apparent that the system has benefited the program in several ways.

A. Less "payload pad" - It is only natural to protect against performance uncertainties by creating a large "payload pad". Since booster performance is practically fixed during a program's flight phase, the only way to get this safety factor is to decrease the spacecraft weight. This is done by deleting experiments or by only partially loading spacecraft propellant tanks. Either solution reduces the utility and productivity of the associated mission. The RTPC system has given the program managers such confidence in our knowledge of booster performance that they have reduced the "payload pad" to a reasonable minimum. This, in turn, has increased the productivity of each flight.

B. Actively improve performance during the countdown - The RTPC system provides the necessary information for applying insulation to specific propellant tanks when such action will be beneficial. This action has been taken during a number of Gemini launch countdowns.

C. Actively avoid ullage limits - The RTPC system indicates when insulation is needed to avoid exceeding ullage limits during the launch window.

D. Provide information for slow malfunction monitoring - The RTPC system provides information to MCCH so that the appropriate performance constraints can be employed in the slow malfunction monitoring operation.

E. Display effects of rendezvous mission - The varying values of launch azimuth and the varying amount of yaw steering required by the rendezvous mission have a profound effect on booster performance capability. Since the P_M vs. time curve is displayed at MCCH, these effects are quantitatively displayed to all interested parties. Thus, decisions involving booster performance are based on facts rather than estimates.

Conclusions

The RTPC system has been a valuable and useful tool on the Gemini program. At a point in time when booster performance improvements became prohibitively expensive, significant additional performance (reduction of the payload pad) was brought about by RTPC. Thus, it can be concluded that the RTPC technique used on Gemini will be applicable to any large liquid booster program where there is little or no excess booster performance. Since payloads do characteristically grow to meet or exceed their booster performance capability, the RTPC technique should see a great deal of use in the future.

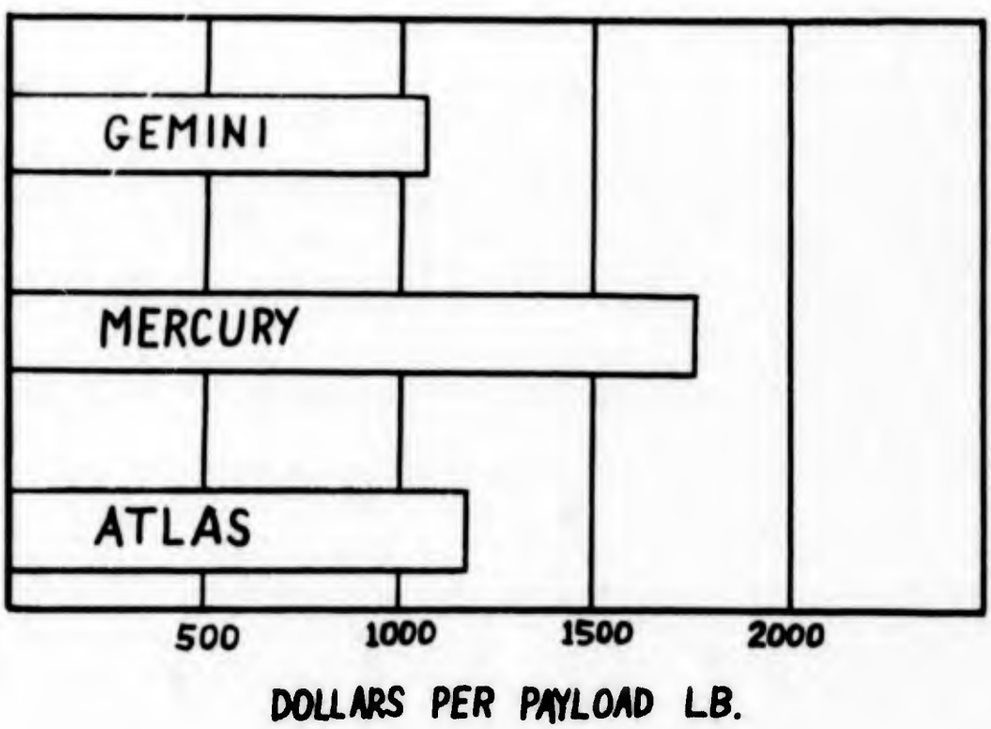


Figure 1 - Cost per Payload Pound Ratios for Low Earth Orbits

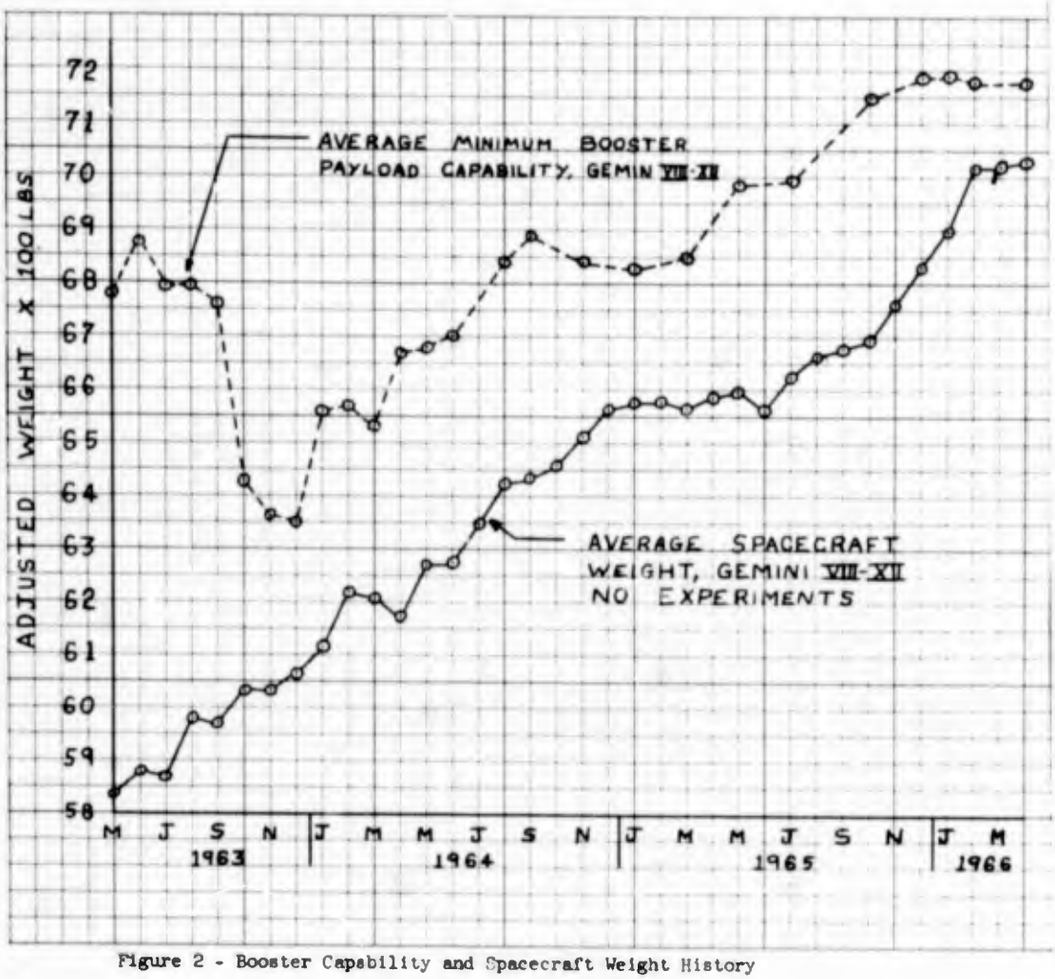


Figure 2 - Booster Capability and Spacecraft Weight History

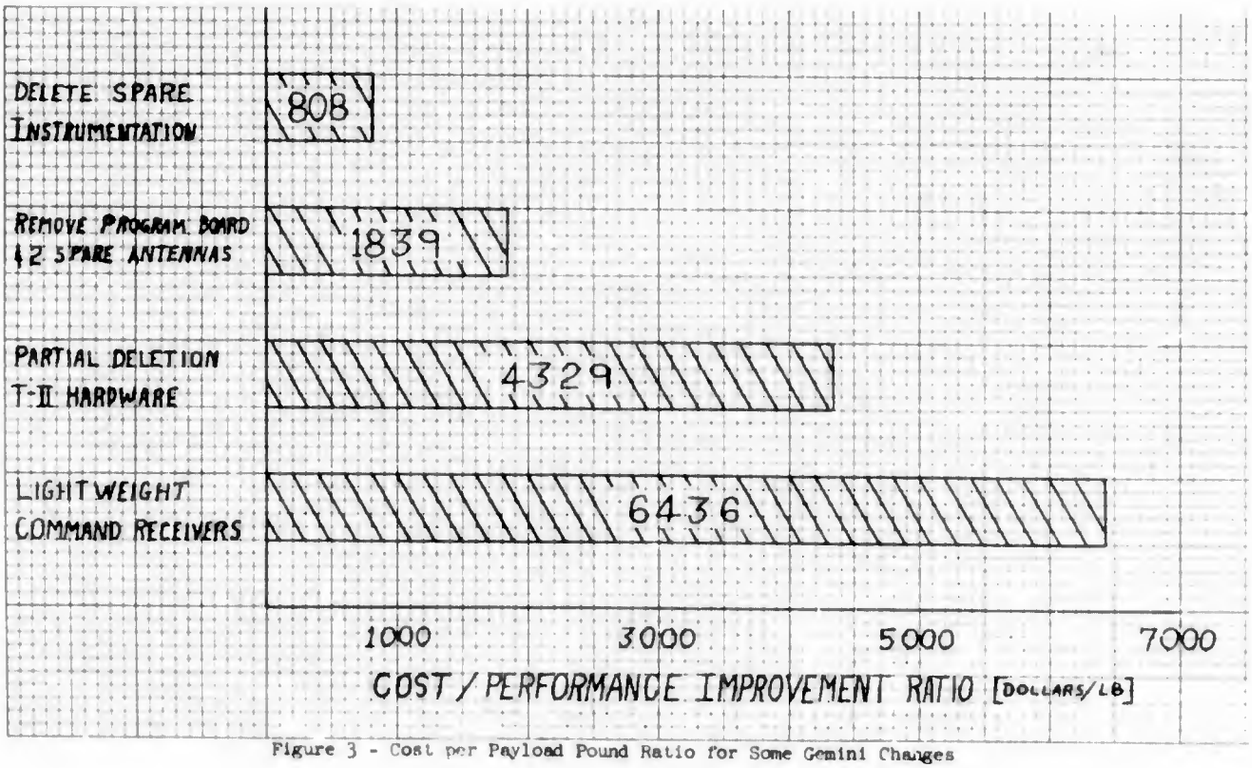


Figure 3 - Cost per Payload Pound Ratio for Some Gemini Changes

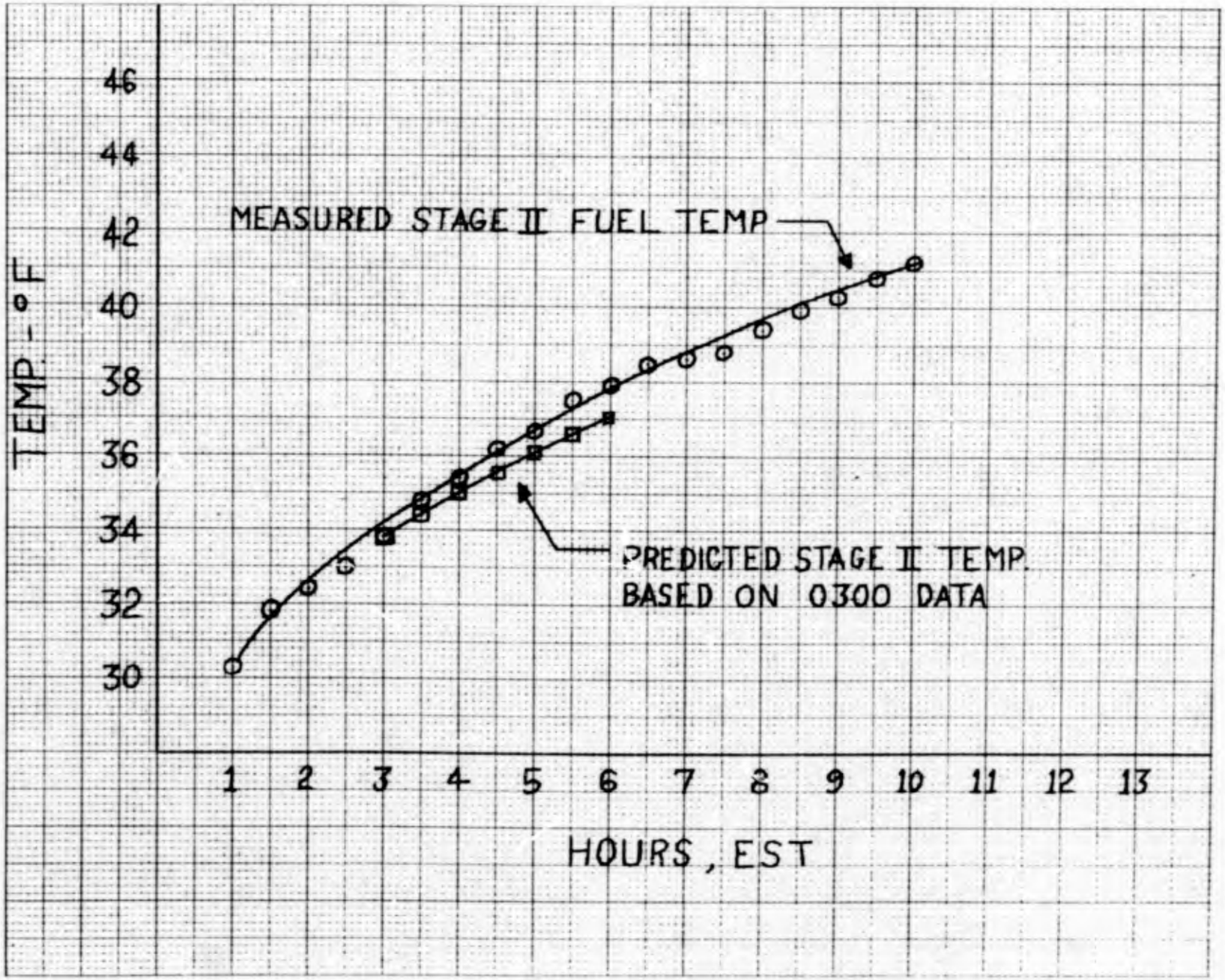


Figure 4 - Gemini IX Stage II Fuel Tank Temperature History

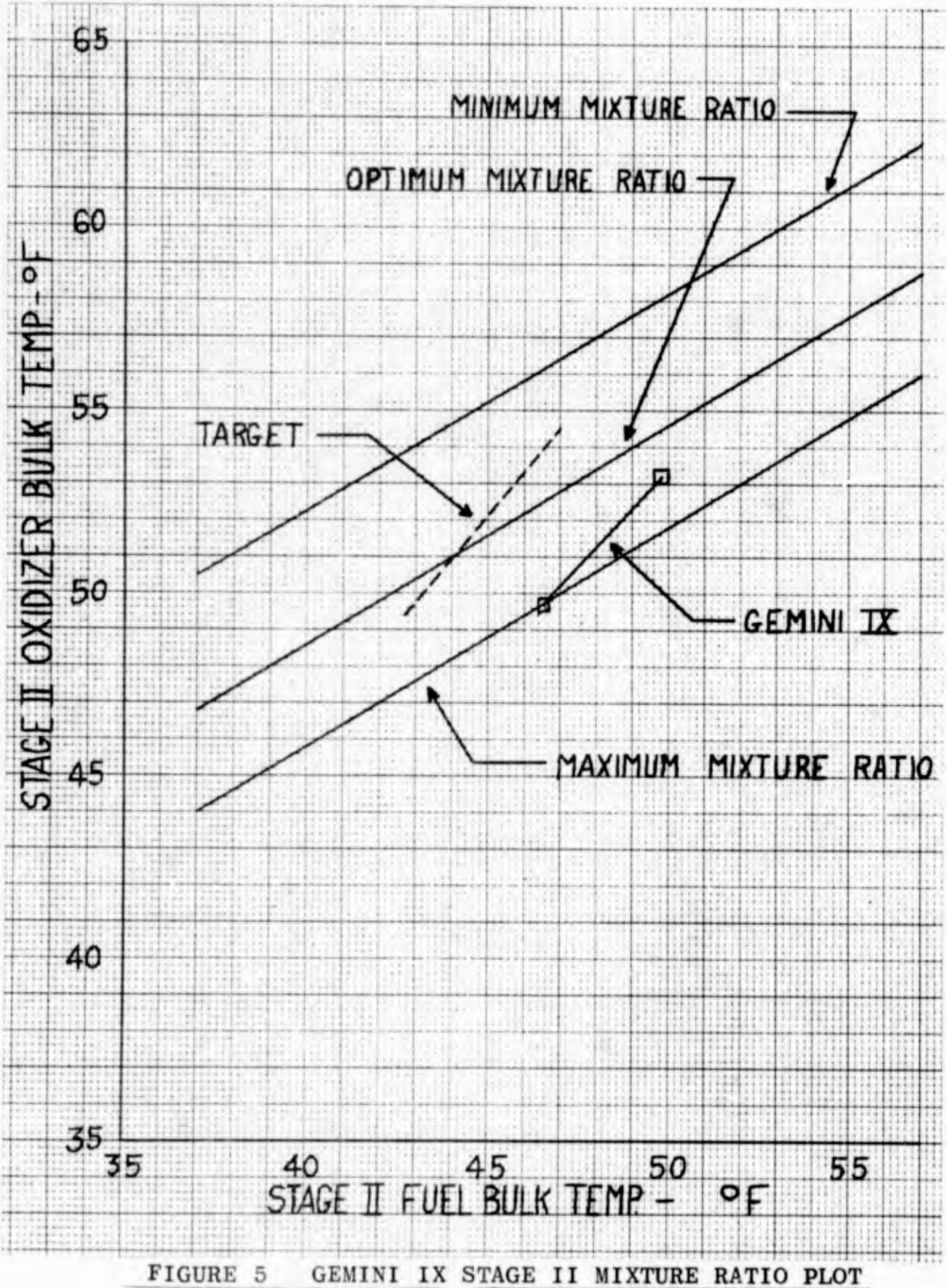


FIGURE 5 GEMINI IX STAGE II MIXTURE RATIO PLOT

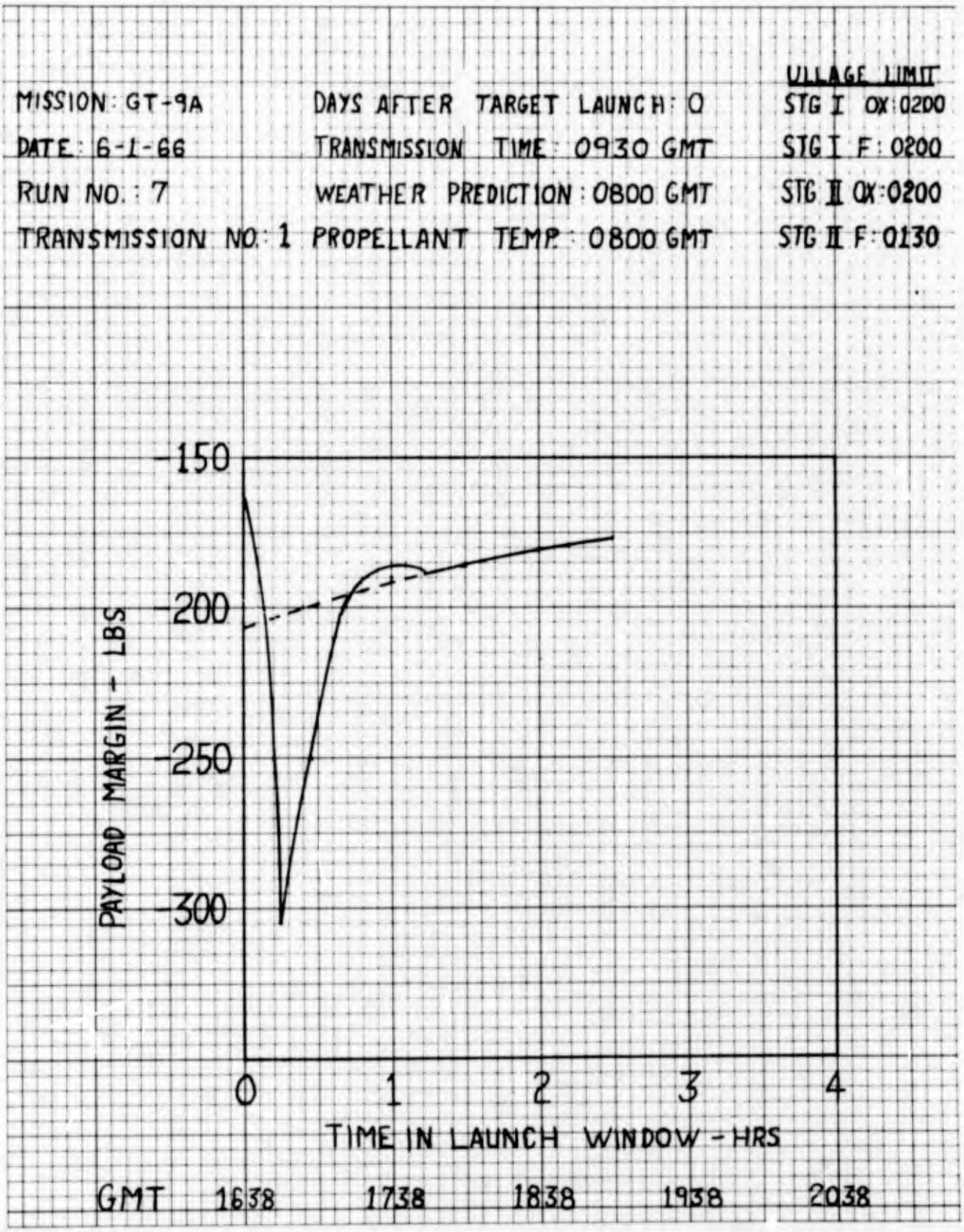


FIGURE 6 GEMINI IX PERFORMANCE MARGIN CURVE

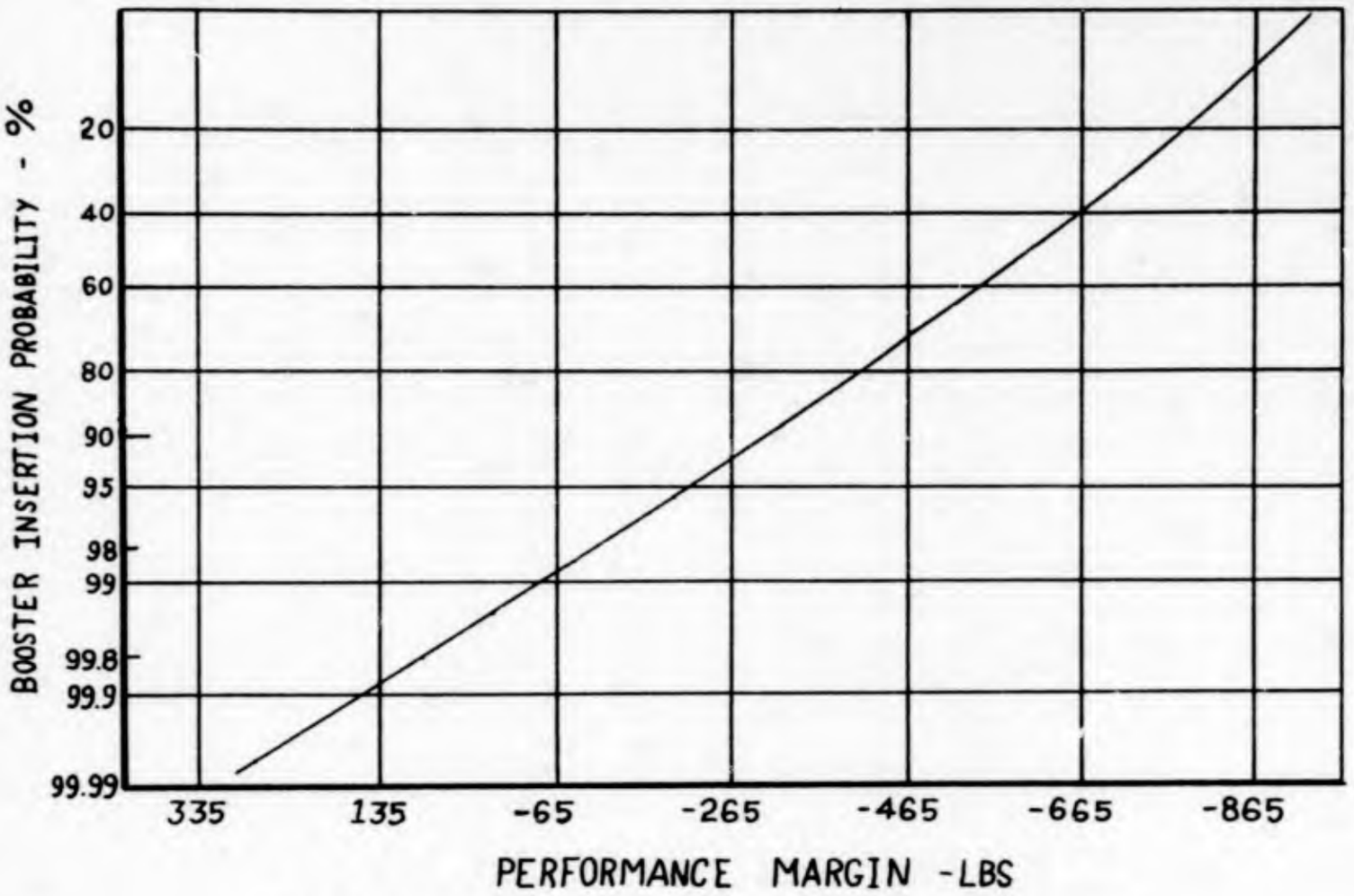


Figure 7 - Gemini IX Booster Insertion Probability Curve

DESIGN, FABRICATION AND TESTING OF A PROTOTYPE FLY-AWAY SATELLITE TRIANGULATION SYSTEM (U)

PREPARED FOR THIRTEENTH ANNUAL AIR FORCE SCIENCE AND ENGINEERING SYMPOSIUM

BY
LT COL MARTIN SELINFREUND & GEORGE P. MUSANTE

1381st Geodetic Survey Squadron
1370th Photomapping Wing (MAC)
UNITED STATES AIR FORCE

Lt Col Martin Selinfreund

BIOGRAPHY

Martin Selinfreund entered the Glider Corps of the Army Air Corps in Sept 1942. Upon termination of the Glider Program he entered the Aviation Cadet program and graduated in December 1943 as a 2nd Lt Bombardier at Carlsbad, New Mexico. After graduation he joined the 711 Bomb Sq of the 8th Air Force in England, where he was awarded the Distinguished Flying Cross and the Air Medal with 3 Oak Leaf clusters for his participation in 33 World War II daylight bombardment sorties.

He left the Army Air Corps in Dec 1946 to enter the Latin American Institute. Soon after graduation he was recalled to active duty with the Unites States Air Force. From 1952 to the present he has worked in mapping and surveying with the 1371st Mapping and Charting Squadron, Aeronautical Chart and Information Center and the 1381st Geodetic Survey Squadron with time out to attend the Ohio State University under the AFIT program to obtain a Bachelor of Science Degree on Geodesy, and 4-1/2 months on assignment as Representative to the Chilean Antarctic Expedition of 1962-1963 under the Scientist Exchange Program. In 1959 he earned his Master Navigator Wings and in 1962 he received the Missile Badge for work on the Minuteman survey design. He is presently Chief of the Stellar Camera Division of the 1381st Geodetic Survey Squadron of the 1370th Photo Mapping Wing (MAC).

George P. Musante

BIOGRAPHY

George P. Musante received his Bachelor of Engineering degree from Rensselaer Polytechnic Inst. in 1946. He was self-employed as Consulting Engineer thru 1957. He then joined Spellman Engineering Inc., Orlando, Florida as Director of Engineering from 1957 thru 1959, where he was responsible for missile complex structural and cryogenic system design. He joined Instrument Corporation of Florida in 1959 as Chief Systems Engineer and developed the first Geodetic Stellar Camera System. He was also instrumental in the development of the PO-1000 camera. In January 1963, he was selected at the Instrument Corp. of Florida Technical Representative with the 1381st Geodetic Survey Squadron in Orlando, Florida. As Technical Representative to the AF he was concerned with camera system operation, maintenance and personnel training. He joined Civil Service in May 1964 as General Engineer with the 1381st Geodetic Survey Squadron where he has been concerned with system operation, maintenance and new system configuration and development.

ABSTRACT

On 31 October 1962, the ANNA IB geodetic satellite was launched and the 1381st Geodetic Survey Squadron embarked upon a program of testing the Air Force Geodetic Stellar Camera System. The results of the tests demonstrated that the system is capable of extending geodetic control to a proportional accuracy of better than one part in 200,000.

With the tests successfully concluded, efforts were directed toward developing a lightweight portable system to replace the 15,000 pound system. Simultaneously, interest was generated in the use of passive satellites as space targets in addition to the active satellite such as ANNA. An additional component of the PC-1000 camera was required to interrupt the continuous trace of a passive satellite in order to provide the satellite position at a specific time. This generated the requirement for timing accuracy far in excess of that required in the present system.

This paper discusses the background of the Air Force Geodetic Stellar Camera System and presents the details of the design and fabrication of a prototype fly-away Satellite Triangulation System. The development of a chopping shutter and accurate timing equipment is also discussed.

TABLE OF CONTENTS

DESIGN, FABRICATION AND TESTING OF A PROTOTYPE FLY-AWAY SATELLITE TRIANGULATION SYSTEM

BLANK PAGE

I. INTRODUCTION

Satellite triangulation, a sophisticated relatively new survey method, has opened new doors to accurately locating points on the earth's surface. For the first time in history remote islands can be tied accurately to continents. Isolated datums can be interrelated and intercontinental ties can be made.

Satellite triangulation, using geodetic stellar cameras, is accomplished by photographing satellites against a star background. The directions to the stars are known, and the coordinates of the star and satellite images can be measured. The directions from the camera station to the satellite positions can then be determined through mathematical reduction of the measured image coordinates. By simultaneously observing a satellite from different stations, the individual directions, or lines in space, from camera station to satellite can be combined to form a system of triangles. These triangles can then be solved for the position of the unknown station or stations.

Measurable dot images are produced from each star by using a series of short shutter openings. The time of these openings must be known to within 10 milliseconds to obtain the desired accuracy in the camera orientation. Measurable dot images of a satellite are produced in two ways. The first is by using an active satellite equipped with a strobe light. The light flashes for only a few milliseconds and therefore produces a dot image rather than a streak. The second method is to photograph a passive or reflecting satellite, such as ECHO, through a rapidly opening and closing shutter. This chopping shutter technique produces a series of dots, and the times of the shutter openings and closings must be determined to better than a millisecond to achieve the desired accuracy.

The geodetic stellar camera capability in the 1381st Geodetic Survey Squadron was developed in late 1959 in order to participate with the Air Force Cambridge Research Laboratories in stellar triangulation experiments, using flares carried aloft by rockets. Several disappointing failures were encountered due to rocket and flare package malfunctions. An attempt to use a strobe light on an aircraft also failed due to difficulties with the light. In October 1962, the ANNA 1B satellite, which contained a flashing strobe light for use with geodetic stellar cameras, was successfully launched. Results of tests conducted with ANNA demonstrated that the system was capable of extending geodetic control with a proportional accuracy of better than one part in 200,000. More recent tests, conducted with the GEOS A satellite, have resulted in typical internal consistencies of one part in 320,000, although data reduction is not complete.

The systems employed to make the first efforts were primitive by present day standards and costly by virtue of their one of a kind status. They were

heavy and bulky, resulting in high transportation costs and were restricted to use with active or flashing light satellites.

Air Force and DIA interest in a capability to perform with either active or passive (reflecting) satellites and in reducing transportation costs provided the impetus for developing a fly-away system. This paper is concerned with the development, by the 1381 Geodetic Survey Squadron, of a prototype system which is readily transportable and is capable of observing both active and passive satellites. We would like to acknowledge AFCRL's encouragement, advice, and financial aid in making this prototype a reality.

First, a brief description of an Optical Satellite Triangulation System, including present MOD I and MOD II configurations, will be given. Then the fly-away, or MOD III system, will be discussed from the standpoint of design, procurement and fabrication, and testing.

II. AN OPTICAL SATELLITE TRIANGULATION SYSTEM

BASIC ELEMENTS Any stellar camera system consists of basic elements as follows:

1. A device capable of determining its own direction and a direction to a target in space -- a long focal length high resolution camera.

2. A device capable of controlling the opening and closing of the internal shutter for timed sequences to expose the star field to ascertain the camera direction -- an electronic controller that opens and closes the shutter.

3. A device to record shutter open and close events against a time base -- time recorder.

4. A device to store time and shutter event data until the time recorder can print it -- time data storage unit.

5. Devices to obtain timing information transmissions from available reliable sources -- VLF, WWV, LORAN "C" Receivers.

6. A device to maintain time of day and supply it to the time recorder -- timing generator and frequency standard.

7. Devices to supply regulated AC and DC power as well as automatic switching of emergency power to the system -- AC generators with automatic transfer switch, AC/DC regulated power supplies, battery packs and battery chargers.

8. A device to shelter the components from the elements -- a dome, trailer, van, available building, or tent.

9. A device to synchronize internal timing equipment to transmitted or external time reference -- dual trace oscilloscope.

MODEL I SYSTEM DESCRIPTION The Model I systems were designed and manufactured by a commercial contractor at a cost of $85,000 each and were delivered in mid year 1959. The electronic components are housed in an air-conditioned van, and the camera is operated on an external tripod. The system occupies 1385 cubic feet and weighs 10,000 pounds. It is air transportable by C-124 or C-130 aircraft (Figure 1).

The camera controller is an electro-mechanical device; the pre-selected exposure sequence is machined into a 14" diameter aluminum disk. The program disk is rotated at .5 RPM allowing voids of specific size to pass between a light source and a photo diode assembly. This permits an

electrical pulse to command the shutter. Any variations in the program require costly engineering and fabrication of additional disks.

The timing equipment consists of an electro-mechanical clock with a drift rate of 4 to 5 milliseconds per hour, making it necessary to record a pre and post operation clock vs. time station transmission comparison. The data recording system is a two track magnetic tape recorder that records shutter function on one track and the 1 pulse per second output of the clock on the second track. The shutter return is represented as a 300 cycle tone in the closed position and 3,000 cycles in the open position. The clock is represented by a second pulse. After mission, the magnetic tape is reproduced on recording oscillograph paper and manual mechanical measurements are made. Time reduction tolerance is at best ± 3 milliseconds. With the necessity of having a time transmission signal available before and after each mission, an unmodified MOD I system cannot be deployed on a world-wide basis, but only to locations where radio time signals are available.

MOD II SYSTEM DESCRIPTION The Model II system was designed and manufactured by a commercial contractor at a cost of $225,000. The systems were delivered in June 1962. Optical and electrical equipment is permanently housed in a transportable trailer type shelter. Both sections of the shelter have environmental control. The timing components of this system have internal accuracies of one part in 10^{10} that could be suitable for passive satellite data acquisition. However, it does not have the capability of precision synchronization to a time station, or the necessary recording speed for chopping shutter operation. The system occupies 2,020 cubic feet and weighs 13,460 pounds. It is air transportable by C-124 or C-130 aircraft (Figure 2).

The MOD II camera controller is programmed by 5 channel teletype codes, pre-punched on paper tape, thus eliminating the possibility of program variation by the operator in the field. The controller is limited to a maximum speed of 700 milliseconds, repetition rate which is not sufficient for a chopping shutter mode of operation.

The time of day recording system produces data on paper tape using 5 channel teletype code. It has a maximum printing speed of 9 columns every 750 milliseconds and requires operator training in punch tape reading. Environmental control of $\pm 2^{\circ}F$ is necessary to insure dependable operation of the sensitive electronics equipment associated with the system, requiring approximately 10kw of electrical power on a non-interrupted 24-hour basis.

The Model II system has proven itself a very useful and dependable geodetic tool over the past years and is capable of world-wide operation. Its major limitations are its vast size and weight and adaptability to active satellites only.

III. DESIGN OF THE MOD III OR FLY-AWAY SYSTEM

CRITERIA With the primary objectives in mind (dual capability for active and passive satellites and reduction of size and weight of our equipment), the following criteria were used in designing the MOD III or fly-away system:

1. Active and passive satellite capability.

2. Time keeping accuracy to 1 part in 10^{10}, (an error of one second in 300 years).

3. Time recording accuracy ± 5 microseconds.

4. Minimum recording repetition rate, 20 print-outs per second (a time print-out at open and close of shutter 10 times per second).

5. Solid state.

6. No requirement for environmental control.

7. Transportability (readily loaded and unloaded in aircraft, helicopter, etc.).

8. Light weight.

9. Small, compact.

10. Interchangeability of parts from system to system.

11. Ruggedness.

12. Standard modules (plug in circuit modules).

13. Inexpensive.

14. Availability of replacement parts.

15. Easily housed.

With the criteria identified, 2 man (Officer/NCO) teams, to serve as project officer and project technician, were selected from within the Stellar Camera

Division of the 1381st to be responsible for writing the initial specifications and researching off-the-shelf units which would meet the specifications.

IDENTIFICATION OF REQUIRED COMPONENTS The combined experience of people long associated with stellar camera activities at the 1381st GSS and Air Force Cambridge Research Laboratories was used to select the general components needed for the MOD III system. An engineering meeting was held at Orlando, Florida, in June 1965, at which time the requirement and general specifications for the new system were discussed. The list of general components was selected to conform to these specifications. The next step was to assign initial product research in the twelve categories listed below to one of the two-man research teams:

1. Chopping shutter
    a. Programmer
    b. Recorder
2. Timing system
    a. Frequency standard
    b. Timing generator
3. Recording system
    a. Time recorder
    b. Controller
    c. Data reduction of time information
4. Power supplies
    a. Power generators
    b. AC/DC power supplies
    c. Battery pack back ups
    d. Automatic switching circuits
5. Packaging
    a. Interface

b. Transportability

c. Housing

6. Cameras

a. Rebuilding

b. New design

c. Shutter research

7. Acceptance testing

a. Component

b. Final

8. Communication support

a. Transceiver (SSB)

b. Linear amplifiers

9. New equipment operator training

10. Time receivers

a. WWV

b. VLF

c. LORAN "C"

11. Test equipment

a. Optic

b. Electronic

12. Cold storage for photographic plates

IV. PROCUREMENT AND FABRICATION

PRODUCT RESEARCH Extensive investigation was conducted during the preliminary phase of the project to identify suitable components available

for use with the light weight camera system. The first phase consisted of 400 to 450 inquiries mailed to possible suppliers, requesting specifications relative to their products. The replies and specifications were studied and evaluated during a 7-8 month period for desirability and compatibility with our engineering concepts. Phase two consisted of personal visits by our officer/technician teams to evaluate their products and assess their ability to produce to specifications. Some equipment was acquired on a no-obligation loan basis for evaluation in our laboratory. We also contacted agencies and companies that had components in operation, for their opinions and comments. Five months were required to complete phase two.

RESULTING COMPONENTS Rigid application of the criteria was applied during product research for off-the-shelf components and the following components were assembled:

1 Frequency standard type 1115-B, General Radio Company (Figure 3)
1 Timing Generator (digital synchronometer) type 1123-ASL General Radio Company (Figure 4)
1 High speed digital printer, Series 1200, Franklin Electronics, Inc. (Figure 5)
1 WWV Receiver, Model WVTR A-Mark II, Specific Products

1 VLF Receiver/Phase Comparator/Chart recorder Model 1880, Electronics Engineering Company
1 Loran C Receiver, Model 601-A, Pickard and Burns Electronics (Figure 6)
1 Oscilloscope, Type 422, Tektronix, Inc. (Figure 7)
1 Standby Power Supply 24V DC/Battery Charger, Model 724BR, Hewlett Packard Company (Figure 8)
2 - 3kw Power Generators 115V, 60 cps, Model 3 MM 25, Kohler Co.
1 Power Transfer Panel, Kohler, Model MDT A-246501

DEVELOPMENT OF IN-HOUSE COMPONENTS Three major components, due to their specialized nature, were not available commercially as off-the-shelf items. These were a high speed data storage, a digital camera controller, and a portable cold storage unit. Inquiries convinced us that the high cost and long lead time required for feasibility studies and engineering models required we build or design the units in-house to our own standards and specifications. In addition, a Rapidyne aerial shutter FSN 6760-035-6487 was modified in-house for use as a chopping shutter.

After three months of designing, redesigning, building and testing circuitry using commercial digital circuit modules, a high speed data storage component which met all of our specifications was readied. The modules consisted of 2 1/2 x 5 1/2 printed circuit boards of glass epoxy with 18 pin connectors and solid state components. Each printed circuit or module

houses one or more logic circuits as required. The completed storage assembly required 92 modules and a suitable power supply for its operation.

In choosing our Binary Code Decimal format, consideration was given the following: ease in performing counting and dividing operations as required in the camera controller, economy of storage and gating operations and error detection in our high speed storage system. Several possible codes were evaluated and the 8421 format was selected.

Employing only two state devices, with a memory, we constructed a storage system with the capacity to retain 24 digital numbers, that is the storing of 96 bits, in two 48 bit shift registers using 1248 BCD format. Each register is capable of accepting information in less than 100 nano seconds and is cleared by a print complete pulse from the recorder.

The camera controller was also designed using commercial digital circuit modules. It has the capability to accommodate any anticipated sequence of shutter commands for the internal or external shutters in a fully automatic or manual mode of operation. It has provisions for fiducial flash, shutter operation test and indicator lights to identify malfunction of the shutter system.

The third item, a suitable refrigerator system for storing photographic plates was built commercially to our specifications. It has the capability of obtaining -14^{o} in six hours of operation and will permit no more than $1^{o}F$ of temperature rise in 24 hours at $75^{o}F$ ambient temperature, thus providing appropriate conditions for photo plate shipment and storage at remote sites, where commercial power is not available and it is necessary to rely on the portable power generators (Figure 9). These generators are run only during the mission period but the period is adequate to retain the storage temperature in the refrigerator.

A dependable high-speed aerial camera shutter was located in AF stock that was capable of producing the short exposures required for passive satellite observation. However, this shutter was commanded by a control system in a mode completely unrelated to our requirements, and consequently required in-house modification to suit our needs.

The shutter contains two spring driven leaf assemblies "A" and "B" which are wound by continually running DC motors and mechanically locked in the cocked position and controlled by solonoid tripping devices. Exposure, or one cycle is performed by "A" assembly function to open and "B" assembly to close. Originally the aerial control unit would command the rewind by pulsing "A" closed and "B" open. We made the rewind function automatic by installing a mechanical assembly internally in the shutter that commands the rewind cycle when "B" assembly closes, allowing the unit to

rewind itself at completion of each exposure cycle. We also had to know when the shutter function had taken place and required an electrical pulse to command recording of time of day relative to open and close of each exposure. This was accomplished by installing mechanical cam assemblies and microswitches adjusted to provide a pulse from the "A" assembly when it is fully open and from the "B" assembly when it is fully closed. In order to insure that mean shutter function time recording is accurate and to compensate for mechanical misadjustment and individual switch dwell time each shutter is calibrated to determine actual delay time. This is done by using calibrated photo diodes, a light source and the systems time recording capability. The photo diodes are positioned on one side of the shutter assembly, one in the center of aperture and one on the outside edge of aperture. The light source is placed on the opposite side. The photo diodes are patched into the print command circuits. The shutter is then operated for several cycles. The shutter return switch vs photo diode energizing is recorded by the printer, calibrating the signal delay to ± 5 microseconds of actual shutter function. The calibration is performed on all shutters after maintenance and at random times during field operations.

In order to position the Rapidyne shutter in front of the camera lens and at the same time isolate it from the camera, a special mount was also designed and build in-house (Figure 10).

INTERFACE AND PACKAGING Component selection was made with interface considerations. The high-speed data storage unit was designed to accommodate the negative going level circuits of the timing generator by using converter modules. To insure intelligent, well shaped pulses, Schmitt trigger modules were included in the incoming side of the data storage system. The recording unit, however, would receive only positive going pulses without extensive modification. It was therefore necessary to employ negative to positive converters with pulse amplifier modules to allow for printer operation at a greater distance from the timing generator and data storage units.

The camera controller was designed using a 2mc frequency oscillator and modular dividing circuit. The solenoid drivers are capable of 10 amperes loading at 30 V DC. Fiducial circuits are capable of very fine adjustments to insure optimum exposure. The camera controller is capable of operating any known ballistic camera and associated equipment.

All components will be mounted in aluminum racks of special design in custom molded fiberglas designed particularly for our operational conditions. The entire electronic system consists of 3 light weight portable cases. No single package exceeds 190 pounds. Electrical interconnections between units are weather proof, durable and externally available for operation and application of electrical power to maintain the equipment in a

ready mode during transit or storage periods. The custom molded cases also include automatic pressure relief valves to insure safety during air shipment at high altitudes. In anticipation of the requirement for operation in extremely cold sections of the world, we are including strip heaters, thermostats and insulation to the fiberglas cases that will allow system operation in temp of -40° without external heating equipment or environmental control.

V. MODEL III SYSTEM TEST

All single components were inspected by our technical personnel for compliance with manufacturer's specifications and our minimum requirements. The equipment was subjected to cold tests in refrigeration cold boxes and heat and humidity test under controlled conditions. Electronic equipment in all cases remained operational in the 0°F to 140°F range. However, some problems developed in the mechanical portions of the system. When subjected to a temperature of less than 20°F, the paper strip recorder used with the VLF and phase comparator would display intermittent operation and the recording printer, driven by a synchronous motor, lost driving speed. To eliminate this problem we included insulation and thermostatically controlled heating elements to our packages, allowing for proper operation at -40° without environmental control. The equipment as a packaged system was subjected to commercial transportation and transit over rough terrain in a truck. The equipment received normal handling by our camera team personnel in off-loading and setting up with no interruption to the timing system and remained in operational condition. Although testing is incomplete at this time, we do not anticipate any significant problems from normal handling and shipping.

The one component found inadequate during our test phase was the standby battery power supply. It was required to supplement the internal batteries in the frequency standard and timing generator and to operate the VLF Receiver/Phase Comparator/Chart Recorder during periods when no generator or commercial power was available.

Our original remote site concept required that we operate the frequency standard, Timing Generator and VLF Receiver/Phase Comparator/Chart Recorder for 24 hours on the standby power supply without external power applied. The original unit selected would operate under optimum conditions for only 6 hours and at temperatures below freezing for much less. As a result, we have selected a larger Standby Power Supply which operates with either lead-acid or nickel cadmium batteries with a charge rate of 100 amp capacity, and supplies 24 volt DC and additionally 110 volt AC regulated output sufficient to operate the entire system 24 hours or longer in the event commercial power and generator power should fail.

The final configuration of the MOD III system is as follows (Figure 11):

Weight - 1750 pounds
Cube - 120 cubic feet

Eight to ten systems, plus operators, can be transported in one C-130 aircraft (Figure 11).

The timing rack contains equipment capable of maintaining time to one part in 10^{10} world-wide. The control rack provides commands for the internal and external shutters, dial selected programming in 1 millisecond increments thru 3 minutes 59 seconds, and selection of automatic or manual operation.

The recording equipment will print 12 columns at 20 lines per second. This allows for printing an identifier plus hours, minutes, seconds, milliseconds, 10s of microseconds.

The system is capable of conducting triangulation using either active or passive satellites.

VI. CONCLUSION

Although the exact savings in transportation cost cannot be readily assessed, eight systems may now be transported by one aircraft instead of only one system. It may be assumed that transportation will be reduced by 88%. For example, a project in South America cost $85,000.00 in transportation funds to move the MOD I and MOD II. With the Fly-Away system (MOD III), the cost would have been reduced to approximately $10,625.00. (Figure 12)

The MOD I system cost $85,000.00 per system. The MOD II system cost $225,000.00 per system. These were realistic costs in terms of the state of the art when the systems were built and these systems served their purpose well. However, present advances in semiconductors, electronic logic, smaller components with insignificant heat sensitivity characteristics permitted us to dispense with the requirement for heat dispersion. With environmental control no longer a requirement, the air conditioner, large costly housing and large heavy power generating equipment could be ignored. These factors plus in-house component design and interfacing combined to lower the cost of the Fly-Away system to $25,000.00 with chopping shutter, less camera.

In addition to lower initial cost and lower transportation costs, the new system will effect an important third savings. Since it can operate with either active or passive satellites, more observations will be possible

and data collection time should be cut by one-half to two-thirds. The MOD III is also more accurate and more readily maintained and operated.

Based on the test results of the prototype of the Fly-Away system, which performed beyond expectations, the 1381st GSS has ordered enough components to make eight complete systems for operational use. In short, we believe the MOD III to be a significant advance in Air Force Geodetic Stellar Camera Systems.

BLANK PAGE

FIGURE 1 MOD I SYSTEM WITH POWER GENERATOR (MA112) WITH CAMERA AND CHOPPING SHUTTER. CUBE, 1381 CUBIC FEET, WEIGHT 10,000 POUNDS, AIR TRANSPORTABLE IN C-124 ONLY.

FIGURE 2 MOD II SYSTEM WITH MB5 POWER GENERATOR. CUBE, 2,020 CUBIC FEET, WEIGHT 13,460 POUNDS. TO THE REAR OF THE VAN IS AN ELECTRONICS COMPARTMENT WHICH TAKES UP NEARLY 1/2 OF THE VAN. THE FRONT PORTION IS THE CAMERA COMPARTMENT WHICH HOUSES THE PC-1000 CAMERA AND TRIPOD. BOTH COMPARTMENTS (CAMERA AND ELECTRONICS) ARE AIR CONDITIONED AND HEATED FOR ENVIRONMENTAL CONTROL. AIR TRANSPORTABLE IN C-124 ONLY.

FIGURE 3 FREQUENCY STANDARD MANUFACTURED BY GENERAL RADIO CORPORATION 35 HOURS OF INTERNAL BATTERY AVAILABLE. THE CRYSTAL OSCILLATOR STABILITY IS ONE PART IN TEN TO THE TENTH POWER (10^{10}) AFTER 30 DAYS OF OPERATION.

FIGURE 4 TIMING GENERATOR. READ OUTS AVAILABLE FROM TEN MICROSECONDS TO 24 HOURS IN TEN MICROSECONDS INCREMENTS. ADJUSTABLE DELAY FROM ONE MICROSECOND TO 999.9 MILLISECONDS IN ONE MICROSECONDS INCREMENTS. VISIBLE READ OUT IN HOURS, MINUTES AND SECONDS. IT HAS 24 HOURS OF INTERNAL BATTERIES AVAILABLE AND IS MANUFACTURED BY GENERAL RADIO CORPORATION.

FIGURE 5 HIGH SPEED DIGITAL PRINTER MANUFACTURED BY FRANKLIN ELECTRONICS CORPORATION. THE UNIT IS CAPABLE OF PRINTING OUT 12 DIGITS EVERY 50 MILLISECONDS (240 CHARACTERS PER SECOND).

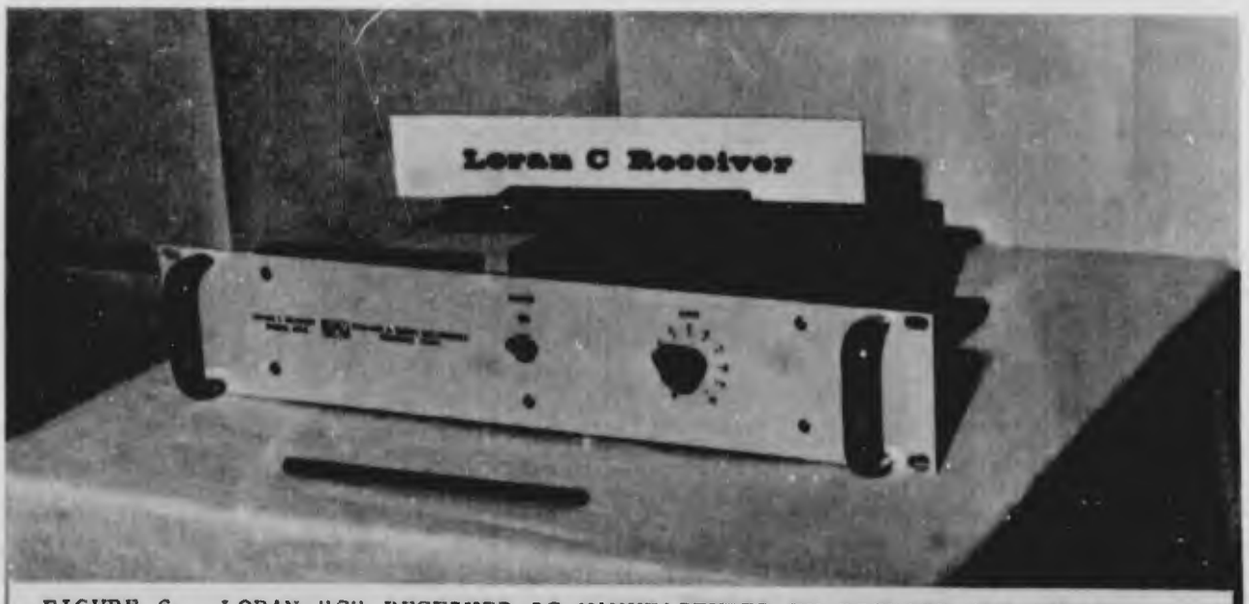


FIGURE 6 LORAN "C" RECEIVER IS MANUFACTURED BY PICKARD AND BURNS ELECTRONICS. CAPABLE OF RECEIVING 100 KC AT A DISTANCE OF APPROXIMATELY 2,000 MILES. REQUIRES SPECIAL ANTENNA ASSEMBLY.

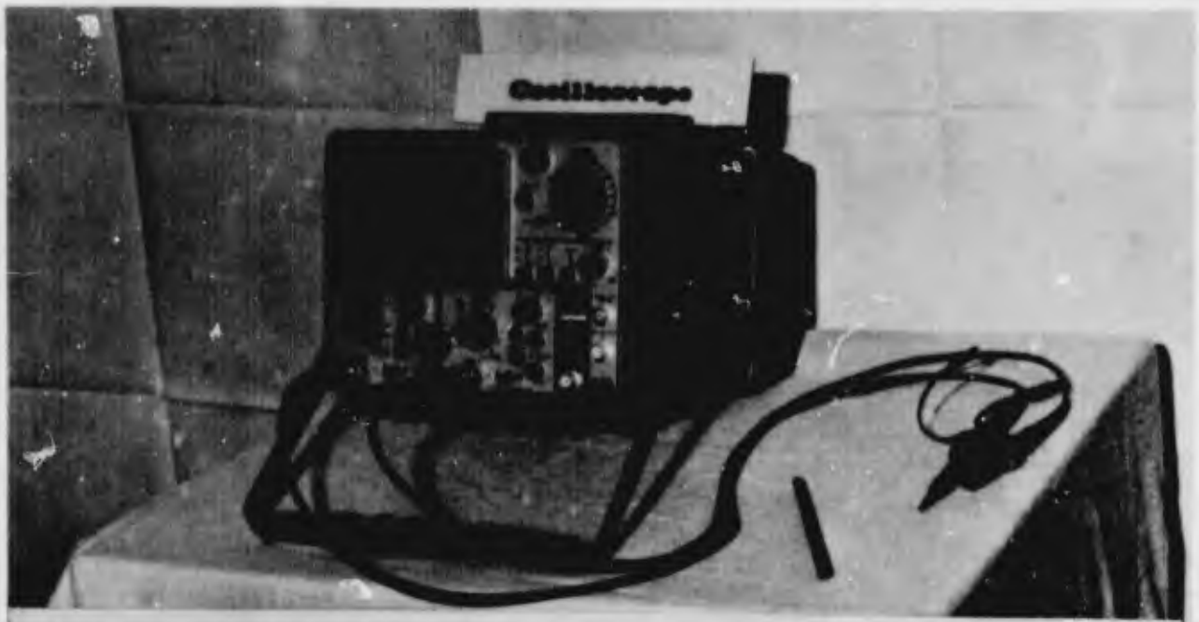


FIGURE 7 OSCILLOSCOPE IS MANUFACTURED BY TECHTRONICS CORPORATION. THIS OSCILLOSCOPE IS CAPABLE OF MEASURING DOWN TO 1/2 MICRO-SECOND PER CENTIMETER AND IS USED IN SETTING AND CHECKING THE TIMING OF THE TIMING GENERATOR. WITH THIS OSCILLOSCOPE IN CONJUNCTION WITH THE WWV RECEIVER AND LORAN "C" RECEIVER, WE CAN SET THE TIME OF THE TIMING GENERATOR WITHIN FIVE MICRO-SECONDS OF REAL TIME.

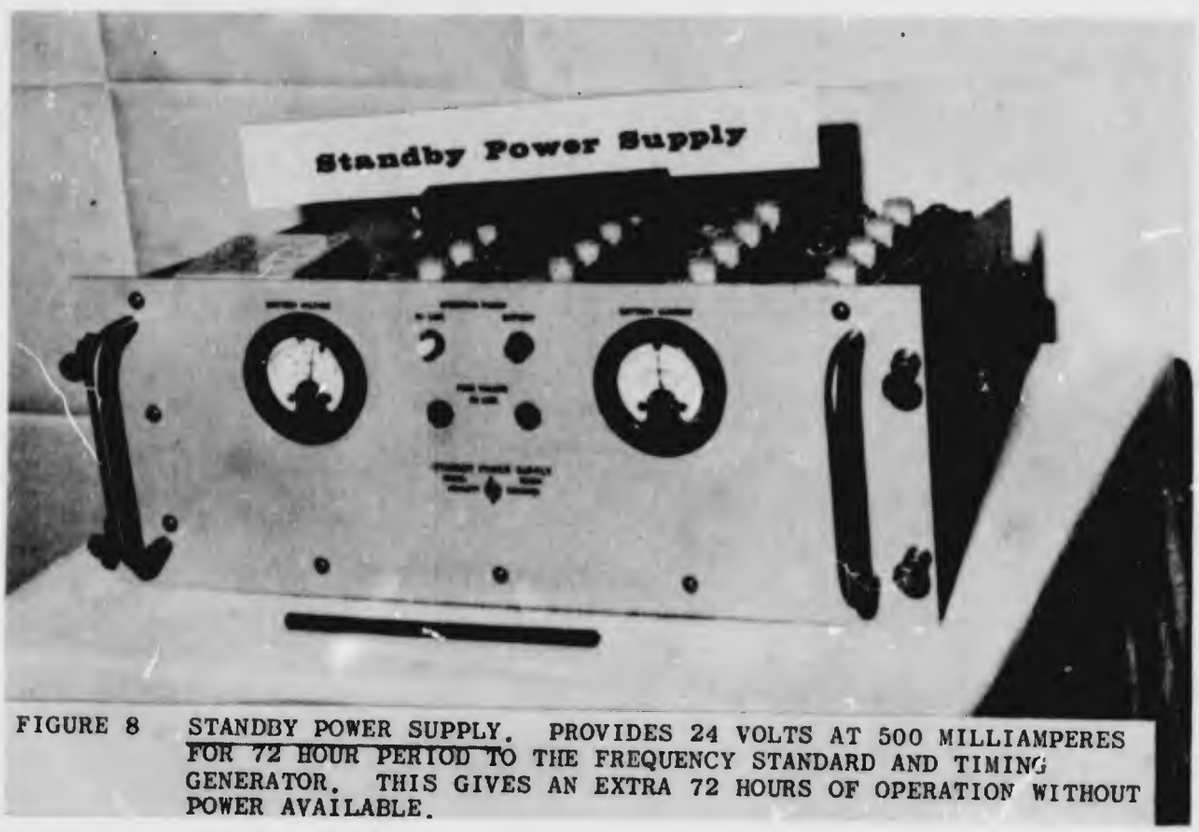


FIGURE 8 STANDBY POWER SUPPLY. PROVIDES 24 VOLTS AT 500 MILLIAMPERES FOR 72 HOUR PERIOD TO THE FREQUENCY STANDARD AND TIMING GENERATOR. THIS GIVES AN EXTRA 72 HOURS OF OPERATION WITHOUT POWER AVAILABLE.

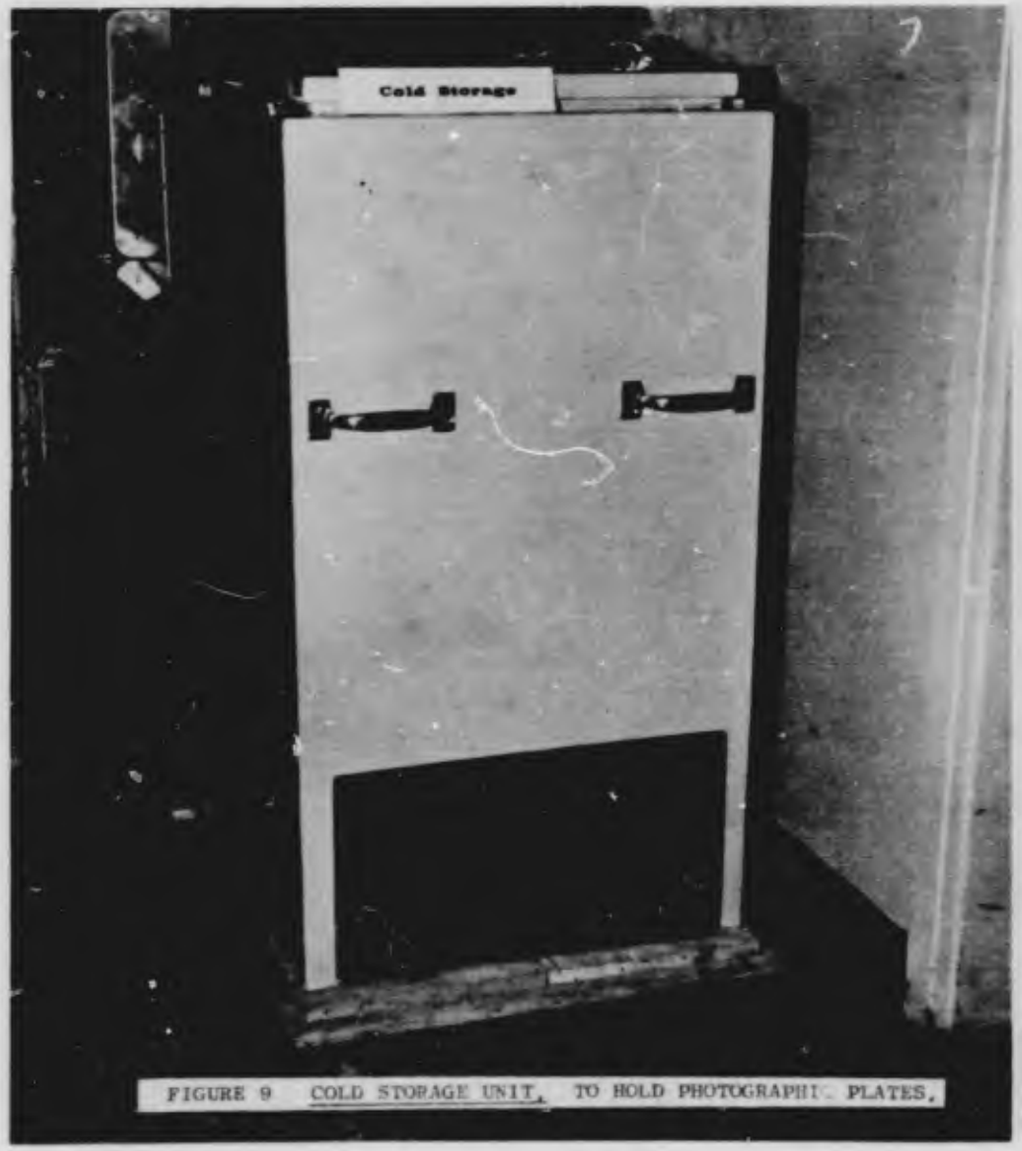


FIGURE 9 COLD STORAGE UNIT, TO HOLD PHOTOGRAPHIC PLATES.

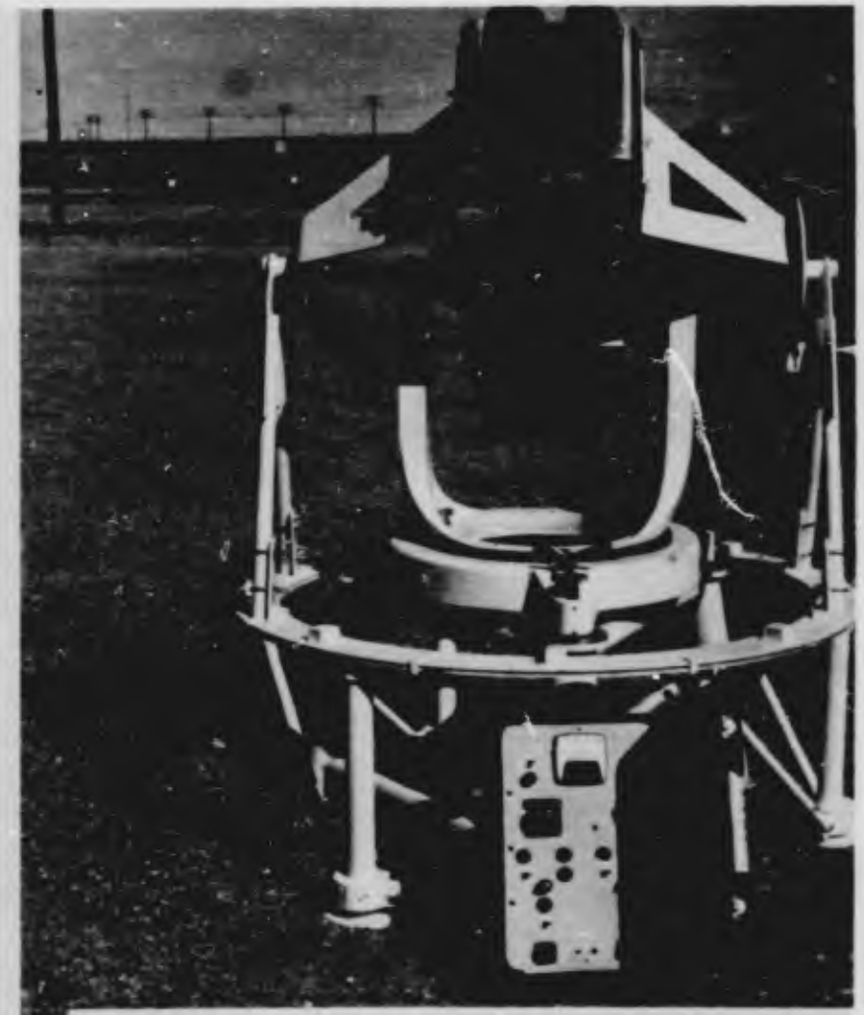


FIGURE 10 PC-1000 CAMERA WITH CHOPPING SHUTTER AND MOUNT
A. PC-1000 IN CHOPPING SHUTTER CONFIGURATION W
PROTOTYPE CONTROLLER IN THE FOREGROUND. TH
CONTROLLER WILL BE REPLACED BY A RACK MOUNT
INTERNAL - EXTERNAL SHUTTER CONTROLLER

FIGURE 11 MOD III SYSTEM WITH 120 CUBIC FEET CUBE, 1,550 POUNDS. TO THE RIGHT ARE THE TWO RACKS SITTING ON TOP OF EACH OTHER, THE LOWER RACK IS THE RECEIVER AND THE UPPER RACK IS THE TIMER WHICH HOUSES ALL TIMING EQUIPMENT. THE RACK TO THE LEFT IS THE CONTROL RACK WHICH CONTROLS THE CAMERA SYSTEM. PC-1000 CAMERA WITHOUT CHOPPING SHUTTER ADAPTER. TWO POWER GENERATORS WITH TRANSFER PANELS.

FIGURE 12 MOD III COMPONENTS (1750 lbs) ON THE LEFT AND THE MOD II SYSTEM (6 1/2 TONS) ON THE RIGHT. TEN MOD III CAMERA SYSTEMS ARE READILY TRANSPORTABLE IN C-130 AIRCRAFT ALONG WITH PERSONNEL REQUIRED TO OPERATE SAME.

Current and Light Storage Effects in Lithium and Sodium Doped Crystals of ZnO

by

Y. S. Park and C. W. Litton

Aerospace Research Laboratories
Wright-Patterson AFB, Ohio

Y. S. Park

BIOGRAPHY

Dr Yoon Soo Park

Dr Park joined the Solid State Physics Research Laboratory of ARL on July 2, 1963, where he is conducting research on the optical and electrical properties of II-VI compound semiconductors. Currently he is investigating the exciton fine structure of II-VI compounds such as CdS and CdSe at low temperatures, thereby contributing to a better understanding of the energy band structure of compound semiconductors. Areas of interest encompassed are crystal growth, cryogenic and optical techniques, photoconductivity and luminescence.

His experience includes, Instructor in Physics, Pusan University, Teaching Assistant, University of Alberta, 1952-1955, Research Physicist, D. H. Baldwin Company, 1956-1962, where he worked on research and development of CdSe film type photoconductors and analog digital converters. Part-time faculty in Physics, University of Dayton, 1965-1966, where he taught "Modern Physics" and "Nuclear Physics" to graduate and upper class students.

C. W. Litton

BIOGRAPHY

Mr Cole W. Litton

Mr Cole W. Litton, Solid State Physics Research Laboratory, was born on September 29, 1930 in Memphis, Tennessee. He came to the Aerospace Research Laboratories in 1957, after assignments at Rocket and Ballistic Research and Electronic Analog Computer Work for WADC at Wright-Patterson AFB, Ohio. Prior to this he was a research assistant at the University of Tennessee Experiment Station and Physical Optics Laboratory at the University performing Analytical Optical Work.

Current and Light Storage Effects in Lithium and Sodium Doped Crystals of ZnO

Abstract

The phenomena of mechanical excitation of fluorescence and current storage have been observed in Li- and Na-doped ZnO single crystals following optical or thermal stimulation of a sample at 77°K. Na-doped crystals possessed a good current storage property and gave mechanically excited emission only after optical stimulation; Li-doped crystals, on the other hand, did not show a current storage property and responded to mechanical tapping only after thermal stimulation. The threshold for optical stimulation of hole storage was found at a wavelength of 3680Å. The depth of the hole storage level lies 0.06 eV above the valence band. Strong, blue-green edge fluorescence can also be triggered by infrared light in an optically stimulated sample.

I. Introduction

Current and light storage effects are the properties revealed by a certain class of crystals, the so-called "tap-effect" crystals, found in the family of II-VI compound semiconductors. The "tap-effect" was first observed in CdS crystals and reported by Warschauer and Reynolds.[1] When "tap-effect" or "tap" crystals are cooled to liquid nitrogen temperature (77°K) in the dark, then stimulated with light and tapped gently with a hard object, they emit a flash of luminescence known as "edge emission." At the same time that the light is given off (due to the mechanical tapping) a decrease in conductivity of the crystal is produced. These two events, optical and electrical, indicate the fact that current and light are indeed stored in these crystals and that the energy is stored in the form of light and current (at a given applied potential) and is released by mechanical means. The current storage effect is more dramatically demonstrated in the following observation: When the "tap" crystals are cooled to 77°K in the dark, they increase in conductivity by as much as eleven orders of magnitude after the crystal is exposed to stimulating radiation; when the light is removed, the conductivity remains high, as if electrons were "stored" in the conduction band, hence the name "storage" crystals. It has also been determined that ultraviolet or visible light is not necessary for stimulation of the crystal. The crystal can be stimulated even thermally. A crystal of this type, cooled to 77°K, then warmed in total darkness to approximately 200°K, with subsequent immersion in liquid nitrogen, shows flashes of luminescence upon mechanical tapping. Thus the phenomenon of "mechanically excited emission" is considered to be one in which radiative emission is obtained by mechanical excitation of optically or thermally stimulated "tap" crystals.

Ever since the discovery of the "tap-effect", both the optical and electrical properties of these crystals have been the subject of considerable study in our laboratory.[2, 3, 4, 5] The "tap crystals" are characterized by several interesting properties which are suggestive of applications: The capacity of these crystals to store light, their highly persistent conductivity, and their unique ability to give off light after thermal stimulation are extremely interesting from the point of view of practical applications. Recently Litton and Reynolds[3] have proposed that the "tap-effect" in CdS crystals is due to the presence of impurities from Group I of the Periodic Table. Armed with this knowledge, we have been able to produce large CdS single crystals that can be mechanically excited to give off bright green flashes of the phonon-assisted edge emission by simply doping the crystals with Li or Na. This technique has now been extended to ZnO from which blue to blue-green emission is observed under mechanical excitation.

It has been known for sometime that the current storage property is not shown by all "tap" crystals. Subsequent to thermal stimulation, crystals having a good storage property showed little or no emission upon tapping, even though they showed a strong response after optical stimulation. Conversly, those crystals that show good mechanically stimulated emission after optical stimulation are good current storage crystals and show little or no "tap" emission after thermal stimulation. During the course of the present experiments, we have established that Na-doped crystals of both CdS and ZnO possessed a good current storage property; these crystals also showed a good "tap" emission subsequent to optical stimulation. On the other hand, Li-doped crystals showed, in general, a good "tap" emission subsequent to thermal stimulation. Thus, in this work, we have established that the Li and Na dopants are definitely responsible for the "tap-effect" and that they play distinctly different roles in "tap" crystals.

Zinc oxide, known since the time of the ancients, occurs naturally as the mineral zincite and has been the subject of extensive investigation for many years. In fact, over a period of more than 25 years, many of the electrical and optical properties of ZnO crystals (of both natural and synthetic origin) have been measured because of their potential application. In fact, investigations of ZnO have revealed many interesting and unusual properties, not the least of which have been the current and light storage properties found in a certain class of crystals. The purpose of this paper is to introduce the experimental facts of mechanically-excited luminescence and current storage in a certain class of ZnO crystals and to interprete these observations in terms of the recently proposed[6] energy band structure; a further purpose is to discuss the emission in terms of trapping levels which lie within the forbidden-energy-gap region.

Like several of the other II-VI compounds, ZnO crystallizes in the hexagonal modification or wurtzite structure. A recent detailed study[6] of the absorption and reflection of ZnO crystals proposes the ZnO band structure shown in Fig. 1. This model is in accordance with the original group symmetry studies of Birman.[7] The fundamental exciton absorption bands at $k = o$ consist of the s-like conduction band and the p-like valence band. The valence band is split into three doubly degenerate states due to the spin-orbit and crystal field effects. On the basis of group-theoretic arguments, it can be shown, in the wurtzite lattice, that the conduction band is of Γ_7 symmetry, that the top valence band (A-band) is of Γ_9 symmetry, and that the two lower valence bands (B and C-bands) are each of Γ_7 symmetry. The selection rules allow a $\Gamma_9 \rightarrow \Gamma_7$ transition for light polarized in the orientation $E \perp C$: $\Gamma_7 \rightarrow \Gamma_7$ transitions are allowed for both modes of polarization ($E \perp C$ and $E \parallel C$).

II. Experimental Arrangement

A. Crystal Preparation.

The single crystals used in these investigations were grown from the vapor phase by a method due to Park et al.[8] A previously reported technique for growing ZnO single crystals consists of the vaporization of metallic zinc and subsequent oxidization of the Zn vapor in a reaction zone separated from the vaporization chamber.[9] The vapor phase technique we have employed here uses sintered powders of ZnS, ZnSe and ZnTe doped with Li or Na as the starting materials. Fig. 2 shows the arrangement of the growth tube used. The sintered powders of ZnS, ZnSe or ZnTe doped with Li or Na are heated at 1375°C in a quartz boat. The argon gas served as a carrier gas and the rate of the gas flow was maintained at 0.5 ft^3/hr. As the material is sublimed, the vapor is carried forward by the carrier gas into the reaction zone, where oxygen is introduced at the rate of 0.1 ft^3/hr.; the reaction zone temperature is maintained in the range of 1300-1350°C. It is important to introduce the oxygen gas after the sublimation has begun. The reaction proceeds as follows and goes to completion: $2\,ZnS + O_2 \rightarrow 2\,ZnO + S_2$. Fig. 3 shows some of the crystals grown from ZnSe powder as a starting material. The crystals obtained are in the form of needles, prisms and plates. Prisms, having a cross section in the range of 1-5 mm and a length of the order 1 ~ 2 cm, are obtained. The plate-like crystals have a cross-section as large as 1 cm^2.

B. Measurements.

For the u-v excited emission spectra, the crystals were irradiated with monochromatic light from a high pressure Hg arc lamp, using a quartz prism monochromator as the filter. The exciting wavelength was set at 3400Å and the spectral-band-pass of the monochromator was sufficiently narrow to preclude overlap with the crystal emission. The fluorescence from the crystals was focused upon the entrance slit of a 1/2 - meter, f/4.4, Bausch & Lomb grating monochromator, equipped with a 1200 lines/mm grating, blazed to 3000Å, giving a linear dispersion of 16 A/mm in first order. The monochromator was equipped with a 1P28 photomultiplier detector. For conductivity measurements, the electrodes were applied to opposite faces of the prism-type crystals. Indium solder (Indalloy No. 1) was used for the electrodes and was applied with a 25-W Sonobond ultrasonic soldering gun. The current was measured with a Keithly Model 610R electrometer at field strengths ranging from 1 to 10 volt/cm. The potential gradient was developed in the direction of the c-axis. In the fluorescence measurements, the crystals were cooled by immersion in either

liquid nitrogen (77°K) or liquid helium (4.2°K); the crystals were glued with rubber cement to one end of a glass rod and placed in the dewar. In the low temperature electrical conductivity measurements, the crystals were thermally anchored to the cold finger of a cryostat and temperatures were measured by a thermocouple that was in thermal contact with the sample.

III. Experimental Observations

A. Mechanism of Light Storage and Release.

First a Na-doped ZnO crystal was cooled by immersion to the temperature of liquid nitrogen. Then the crystal was stimulated with light of wavelength shorter than 3680Å. After stimulation, the crystal was completely shielded from light. If the crystal is now tapped in the dark in the direction of the c-axis, one finds that a blue-green flash of fluorescence known as "edge-emission" is emitted from the body of the crystal. From a reasonably well stimulated crystal, several hundred flashes at the rate of one flash per tap can be obtained.

Next a Li-doped ZnO crystal was subjected to optical stimulation. After stimulation, the crystal was tapped; however no "tap" emission was observed. When a crystal was cooled to 77°K, then warmed in total darkness to approximately 200°K, with subsequent immersion in liquid nitrogen, it showed flashes of the blue-green emission when mechanically tapped; in fact, it was observed that the crystal glowed more uniformly after thermal stimulation than after optical stimulation. Furthermore, only the gentlest tapping (sometimes even the rattling due to the boiling) was sufficient to excite emission. However, the emission is completely exhausted after only a few taps. On the other hand, Na-doped ZnO crystals did not show mechanically excited emission subsequent to thermal stimulation.

B. Characteristics of the Mechanically Excited Emission.

After optical stimulation at low temperatures, the mechanically excited emission from a Na-doped sample is highly directional. It is found that the impact is most effective when applied parallel to the c-axis of the crystal, with excitation perpendicular to the c-axis having little or no effect. Moreover, the emission is observed to radiate in streamers from the point of impact in the direction of c-axis. In addition, it is found that the emission is polarized with the E-vector perpendicular to the c-axis, in accord with the direction of polarization observed for normal edge emission, indicating that a $\Gamma_7 - \Gamma_9$ transition is operative. The flashes observed after thermal stimulation in Li-doped ZnO are not as localized as those obtained in

the optically-stimulated case but are observed more uniformly throughout the crystal.

In order to have a better understanding of the nature of the mechanically excited emission, uv-excited edge-emission of "normal" and "tap" crystals was taken both at 77 and 4.2°K.

Illustrated in Fig. 4 are the general features of the "normal" ZnO fluorescence spectrum, showing several peaks of the broad-band, phonon-assisted, edge emission (near ultraviolet) and the well-known broad green band which peaks at about 5500Å. Several investigators[10, 11] have studied the phonon-assisted peaks as well as the broad green peak in ZnO, and it has been shown that the relative intensities of these two peaks depend markedly on the electrical conductivity of the crystals, with the intensities of the two peaks becoming nearly equal at reasonably high conductivities; it has also been shown that the conductivity of ZnO is affected strongly by excess Zn in the crystals and that the conductivity increases rapidly with increasing Zn concentration. The characteristic, phonon assisted peaks are separated by the longitudinal optical phonon energy, $\hbar\omega$ = 0.073 eV.[12] The inset diagram shows the highly resolved and detailed spectrum which appears between the absorption edge (ground state exciton Γ_5 state) and the phonon assisted emission; this detailed spectrum comprises the bound exciton lines whose characteristics have been studied in detail by Reynolds, et al.[13]

The spectra of the uv-excited emission from "normal" crystals at 90°K[11] and "tap" crystals at 77°K are shown in Figs. 5(a) and (b), respectively. Here, the bound exciton emission, which is actually comprised of the line structure as indicated in the inset diagram of Fig. 4, appears as a broad emission. This is because the spectra in the inset diagram of Fig. 4 was taken photographically with a spectrograph having a higher resolution than the monochromator used in taking the spectra of Fig. 5 (A linear dispersion of 2 A/mm compared with 16 A/mm). In "normal" crystals, the line at 3700Å (I) is the most intense, while the intensity of the I line in "tap" crystals appears greatly reduced. In both "normal" and "tap" crystals, the zero-phonon peak corresponds to the bound exciton emission lines. This peak is followed, at successively longer wavelengths, by a series of equally spaced peaks, each of which is less intense than its predecessor; successive peaks are separated by approximately 0.073 eV, the longitudinal phonon energy.[12] It is also to be noted that the bound exciton emission in "tap" crystals is more enhanced than in "normal" crystals. Fig. 6 shows the spectrum of the uv-excited fluorescence from "tap" crystals at 4.2°K. The spectrum at this temperature is quite similar to that observed at 77°K, except that the emission peaks are sharper and better resolved at 4.2°K than at 77°K.

C. Mechanism of Current Storage and Release.

When the "normal" ZnO crystals are grown from "pure" starting materials, the resistivities are in the range of a few Ω-cm at room temperature (this is usually attributed to excess Zn). When doped with lithium or sodium, the crystals have high resitivities, on the order of 10^4 Ω-cm, and are photoconducting. This indicates that the Li dopant acts like an acceptor in these crystals. When these tap crystals are cooled in the dark, the conductivity goes to a very low value of 10^{-11} $(\Omega\text{-cm})^{-1}$, or less, then upon exposure to light of wavelength shorter than 3680Å, the conductivity increases to values as high as 10^{-3} $(\Omega\text{-cm})^{-1}$. When the light is removed the conductivity remains high. At low light levels, the time for equilibrium current to be established is very long. Similar observations were also made on CdS storage crystals.[1,4]

Fig. 7 shows the current through one of the tap crystals (Na-doped ZnO) at 77°K as a function of time under illumination by light of 3600Å wavelength at intensities in the ratio of 1:2:4. With maximum stimulating intensity, this crystal reached a maximum current of 3×10^{-3} amperes and stayed there indefinitely. Moreover, the crystal could be made to stay in the high conductivity state at any current above 3×10^{-7} amperes, by simply varying the time of exposure to light of a given intensity or by varying the light intensity. Thus it can be seen that there is a considerable integration effect over a large range of currents and times of exposure to light. These facts alone indicate that a complex trap distribution is involved in this type of crystal.

Fig. 8 shows the current through a tap crystal as a function of temperature following stimulation at 77°K with u-v light. The crystal was placed in a state of maximum stored conductivity at 77°K. Then the temperature was raised at the rate of 2°K/min. As can be seen, pronounced decay of the current begins at about 167°K and four distinct traps are noticed at the temperatures 176, 185, 206 and 235°K.

IV. Discussion

In order to explain successfully the optical effects such as luminescence, photoconductivity or "tap effect", one must introduce a set of trapping levels and recombination centers in the "forbidden-gap" region of the energy bands of a semiconductor, as described earlier. Then, the optical phenomena can be interpreted in terms of transitions between the Γ_9 and Γ_7 valence bands, and a set of these trapping levels and recombination centers. Though a fairly detailed measurement of the ZnO edge fluorescence spectrum has been made previously,[11] no unique model has been proposed to account for the observed edge emission as was done for CdS.[14]

This is partly because of complexity and scarcity of the data due to poor quality of ZnO crystals in comparison with CdS.

To qualitatively explain the observed tap effect in Li doped ZnO, we make the tentative assignment of trapping levels shown in Fig. 9. Here the conduction and valence bands have been represented with the multiplicity of valence bands and the symmetry assignments of the bands have been omitted for simplicity.

The uppermost level is assumed to be an electron trap level, taken to be 0.083 eV below the conduction band. Electron transitions are assumed to originate with this level. The intense emission line at 3700Å (I), appearing near the absorption edge at 90°K (Fig. 5(a)), is considered to arise from recombination of a trapped electron in this level and a free hole, in accordance with the Lambe-Klick-Dexter model.[15] The intensity of this line is greatly reduced in "tap crystals" as a result of Li or Na doping (Fig. 5(b)). A threshold wavelength for stimulation of the tap effect in Na-doped ZnO crystals has been found at 3680Å, indicating that there is a hole trap level 0.065 eV above the valence band.

Optical stimulation of the crystal can be accomplished by production of hole-electron pairs and subsequent trapping of the holes in this level, or it can be achieved by light stimulation of electrons from the hole trap level (which in the dark is an electron trap), using light of wavelength shorter than 3680Å. The transition back to this level does not take place. Thus holes trapped at this level provide the storage effect.

The mechanically excited edge emission is a result of the recombination from a free hole and a trapped electron, the hole in this case being freed from the trap responsible for the storage effect by mechanical energy. The fact that the mechanically excited emission can be produced thermally in a Li-doped ZnO crystal indicates there exists a different hole trap somewhere close to the conduction band. Also it is not unreasonable that, if there are traps near enough above the storage level, electrons can be thermally stimulated out of the storage level into the traps as the crystal is being cooled. Thus, holes trapped in the storage level are made available by mechanically freeing them into the valence band to recombine with trapped electrons to give off the blue-green fluorescence.

It may well be that the trap distribution in the ZnO crystals is far more complex than that which is indicated in the thermally stimulated current data of Fig. 8; and, it may even be more complex than the defect structure ascribed to the CdS tap crystals.[4] But, the above picture describes qualitatively, at least, our experimental observations. It is also

noted that the storage level in CdS is an order of magnitude deeper than the corresponding level in ZnO, the distance from the valence band in CdS being 0.75 eV in comparison with 0.065 eV found for ZnO.

In both CdS and ZnO, the mechanically excited emission is composed of most of the emission lines which comprise their respective edge emission spectra. However, there is a marked difference between the uv-excited edge emission spectra of ZnO and CdS crystals that show the "tap effect." Phonon-assisted green edge emission in normal CdS is located at approximately 0.14 eV below the absorption edge, while phonon-assisted edge emission in ZnO appears at about 0.016 eV below the absorption edge. In CdS, the first or zero-phonon peak is always the most intense in spectra of crystals that do not show mechanically excited emission. However, in both the mechanically excited and uv excitted spectra of CdS "tap" crystals, the zero-phonon peak is always less intense than the second peak; moreover, the linewidths are appreciably wider in the edge emission structure.[2] Fig. 10 illustrates the emission spectra of CdS tap crystals.

As shown Fig. 5(a), the zero-phonon peak in ZnO corresponds to bound exciton emission. This indicates that the primary bound exciton transition is accompanied by phonon cooperation. The intense emission line at 3700Å (I), observed between the bound exciton emission and the 1st phonon (n= 1) peak, is considered to be due to an optical transition involving band-to-center. In "tap" crystals (See Fig. 5(b))) the intensity of the 3700Å emission line is greatly reduced and the 1st phonon (n= 1) peak appears intense. Bound exciton emission is also enhanced.

The depth of the storage or hole trap level from the valence band, on the order of 0.06 eV, is determined from the fact that the threshold for optical stimulation of hole storage occurs at 3680Å. The 0.06 eV corresponds to infrared light of 19μ wavelength. Therefore, an optically stimulated sample can be caused to emit radiation by triggering it with infrared light of 19μ; and infrared triggering was indeed observed when a cooled, optically-stimulated, crystal was brought near a heated soldering gun. The crystal glowed brightly and emitted the blue-green fluorescence observed under mechanical excitation. Thus, the lower activation energy for stimulation of hole storage in ZnO extends the infrared detection capability farther beyond the long-wavelength-limit set by CdS "tap" crystals (1.65μ corresponding to an activation energy of 0.75 eV for CdS, as compared to 19μ for ZnO).

V. Potential Application

The properties revealed by the "tap-effect" such as the crystals capacity to store light energy, highly persistent conductivity, and the

ability to emit light energy, following thermal stimulation, are extremely interesting from the point of view of application.

Attractive possibilities for applications are an infrared detector, an image intensifier, a storage device, a radiation dosimeter, an accelerometer, an optical exposure meter, a triggering mechanism as in high speed photography (ballistics, e.g.), a crystal oscillator in resonant circuits, negative resistance circuit elements, and a diode laser of the current injection type, to name a few. Storage properties also appear to have a great many possibilities in computor technology as in memory devices or yes-no switches. As an illustration (See Fig. 11), consider an energy storage device used as an accelerometer. Here an ultraviolet light is used to stimulate the crystal. A storage crystal is located within a container on a block of copper. The crystal is cooled by liquid nitrogen. The crystal is stimulated by irradiating it with ultraviolet. After the stimulation, the crystal is tapped on the side by a mass which is suspended from a support. Movement of the mass can be done by acceleration from mechanical means or from the action of an electromagnet. When the mass strikes the side, a blue-green light is given off. The light is sensed by a sensing unit which could be a photomultiplier. The mechanical tapping can also be produced by an electromechanical transducer such as a piezoelectric transducer. In other circumstances, the stimulation may be done by thermal energy or very long wavelength infrared energy for a time sufficient to heat the crystal to stimulate the storage unit. Or the stored energy can be triggered by infra-red radiation.

The fact that the stored energy can be triggered by infra-red radiation offers a good possibility of making the crystal into an infra-red detector useful for military purpose. A crystalline plate made up of a mosaic of crystals will act as an image intensifier upon irradiation with infrared energy. First the large crystal plate is put into the stored state by illuminating with ultraviolet. A portion of the entire region will glow upon absorption of the signal from the infrared target.

The frequency of light output from the storage crystals can also be varied by use of the proper doping agents, or by use of the storage crystals from other crystals of II-VI family such as CdS, CdSe, CdS:Se and ZnS which cover to whole range of fluorescent colors from the infrared, to the ultraviolet. It is known, from previous measurements,[3] that the CdS "tap" crystals (crystals unintentionally doped with both Na and Li) show a negative resistance characteristic at low temperatures; moreover, there is every reason to believe that the Na-doped ZnO crystals also show this characteristic. A negative resistance characteristic simply means that the crystals show a region of negative resistance ($-E/R$) in their current-voltage characteristics. It is this region which figures importantly in the

diode laser action of compound semiconductors from the III-V groups, such as GaAs, since such lasers are usually operated at currents in the region of negative resistance. As a consequence of the negative resistance, there is every reason to believe that lasing diodes can also be made from the alkali-metal-doped II-VI compounds, provided the crystal resistivities can be reduced to sufficiently low values at low temperatures. The crystals which show a negative resistance characteristic can also be made to oscillate at predetermined frequencies in a tuned RLC circuit and such oscillators are frequently useful in many applications.

A further application of the light storage property of these crystals might be found in high speed photography such as impact photography in ballistics. Since the rise time of the light emission pulse is very fast ($<10^{-8}$ sec), such crystals might be used to trigger the impact of a bullet or a projectile by means of this ability to give off a fast light pulse upon mechanical excitation (shock).

References

1. D. M. Warschauer and D. C. Reynolds, J. Phys. Chem. Solids 13, 251 (1960).

2. C. W. Litton and D. C. Reynolds, Phys. Rev. 125, 516 (1962).

3. C. W. Litton and D. C. Reynolds, Phys. Rev. 133, A536 (1964).

4. B. A. Kulp, J. Appl. Phys. 36, 553 (1965).

5. B. A. Kulp, K. A. Gale and R. G. Schulze, Phys. Rev. 140, A252 (1965).

6. Y. S. Park, C. W. Litton, T. C. Collins and D. C. Reynolds, Phys. Rev. 143, 512 (1966).

7. J. L. Birman, Phys. Rev. Letters 2, 157 (1959).

8. J. DeWitt, Y. S. Park and D. C. Reynolds, Bull. Am. Phys. Soc. 11, 179 (1966).

9. E. Scharowsky, Z. Physik 135, 318 (1953); G. Bogner and E. Mollwo, J. Phys. Chem. Solids 6, 136 (1958).

10. B. Andress and E. Mollwo, Naturwiss. 46, 623 (1959).

11. B. Andress, Z. Physik 170, 1 (1962).

12. R. J. Collins and D. A. Kleinman, J. Phys. Chem. Solids 11, 190 (1959).
S. S. Mitra, Bull. Am. Phys. Soc. 10, 333 (1965).

13. D. C. Reynolds, C. W. Litton and T. C. Collins, Phys. Rev. 140, A1726 (1965).

14. L. S. Pedrotti and D. C. Reynolds, Phys. Rev. 120, 1664 (1960).

15. J. J. Lambe, C. C. Klick and D. L. Dexter, Phys. Rev. 103, 1715 (1956).

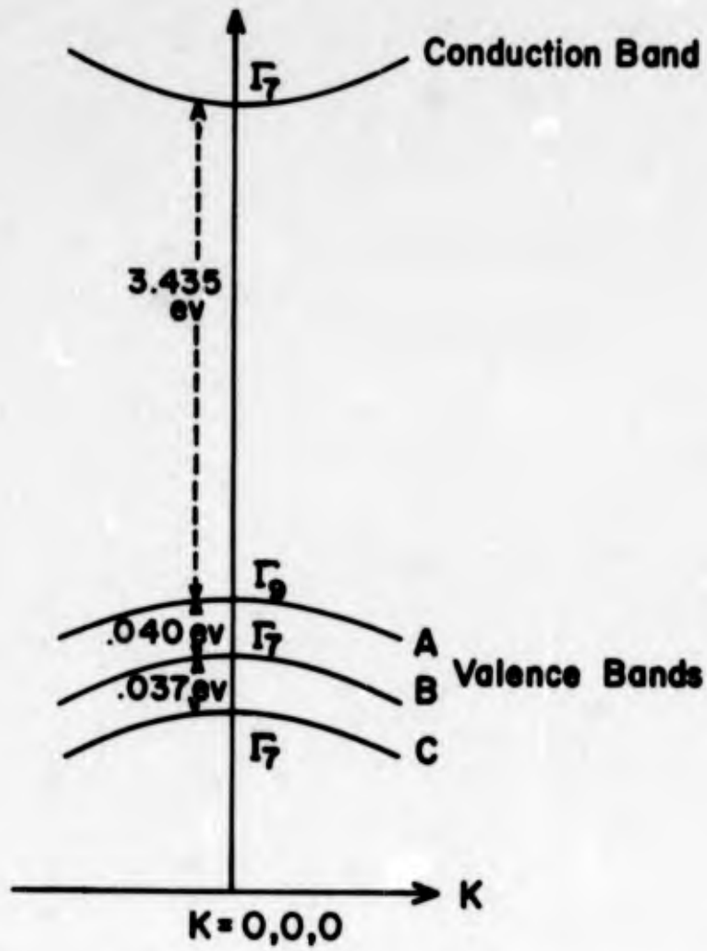


Fig. 1. Band structure of ZnO at 4.2°K. The selection rules allow a $\Gamma_7 \rightarrow \Gamma_9$ transition for light polarized in the orientation $E \perp C$; $\Gamma_7 \rightarrow \Gamma_7$ transitions are allowed for both modes of polarization ($E \perp C$ and $E \parallel C$).

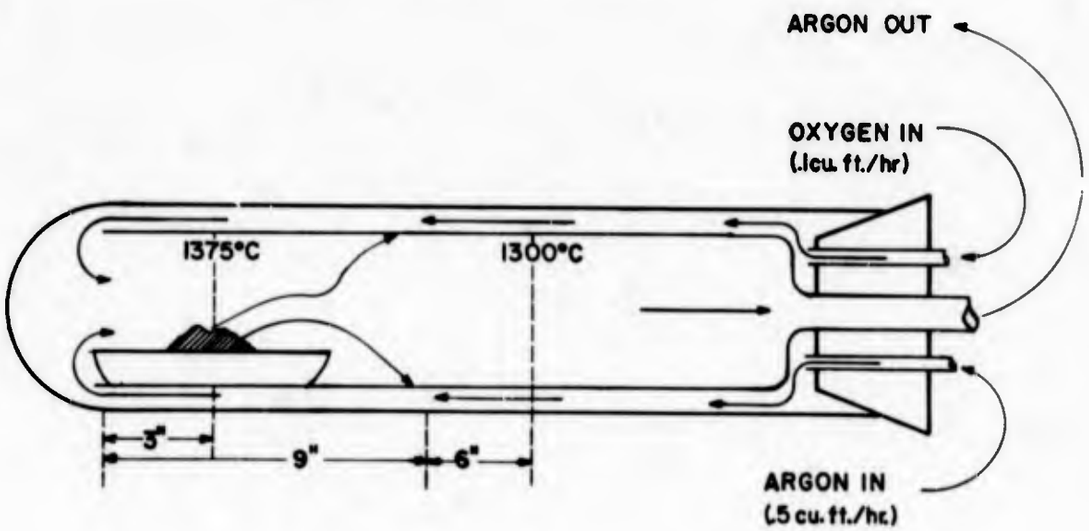


Fig. 2. ZnO Single Crystal Growth Tube.

Fig. 3. Li- and Na-doped ZnO Single Crystals grown from ZnSe powder as a starting material.

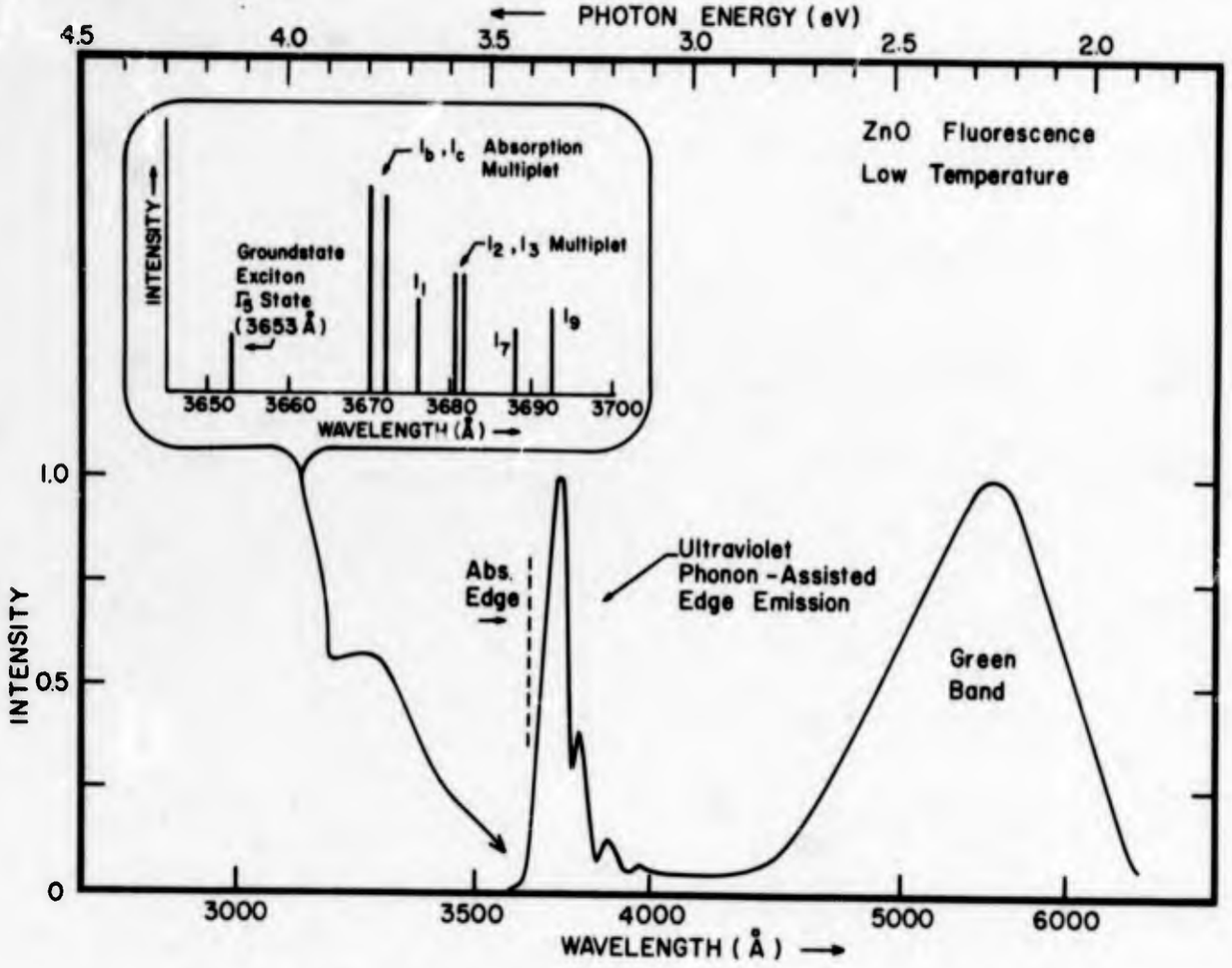


Fig. 4. A diagram of the ZnO fluorescence spectrum at low temperature. Shown in the spectrum is the well-known broad "green" band, as well as the characteristic, phonon assisted edge emission which appears in the near ultraviolet and whose peaks are separated by the longitudinal-optical phonon energy, $\hbar\omega \approx$ 0.073eV. The inset diagram shows bound exciton emission lines which appear between the absorption edge and the phonon-assisted emission.

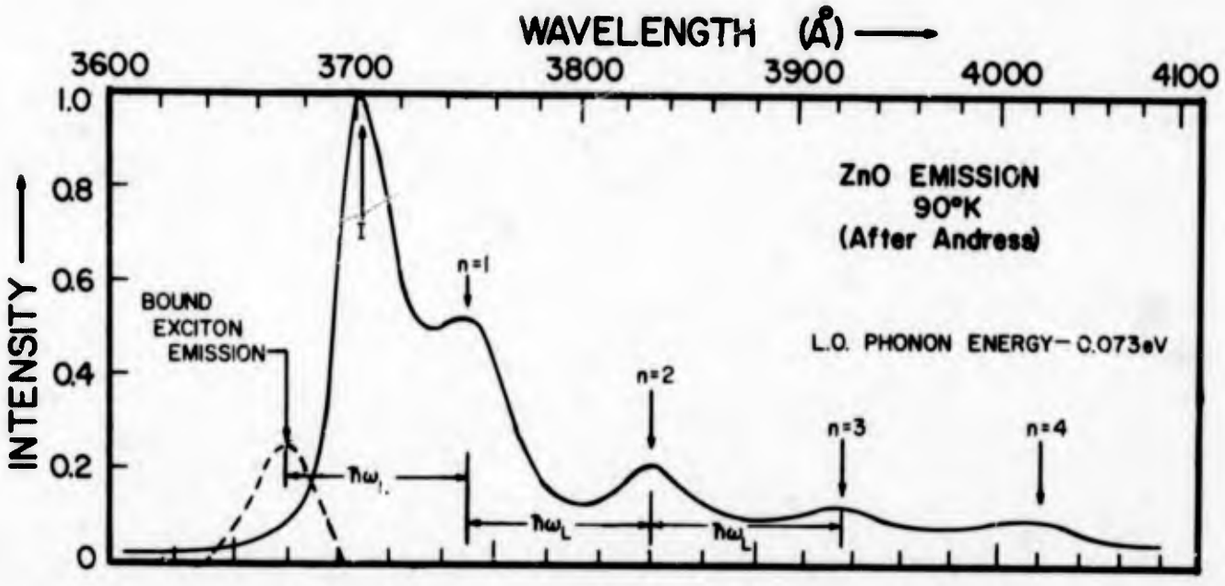


Fig. 5(a). Fluorescence spectrum of a "normal" ZnO crystal taken at 90°K (after Andress).

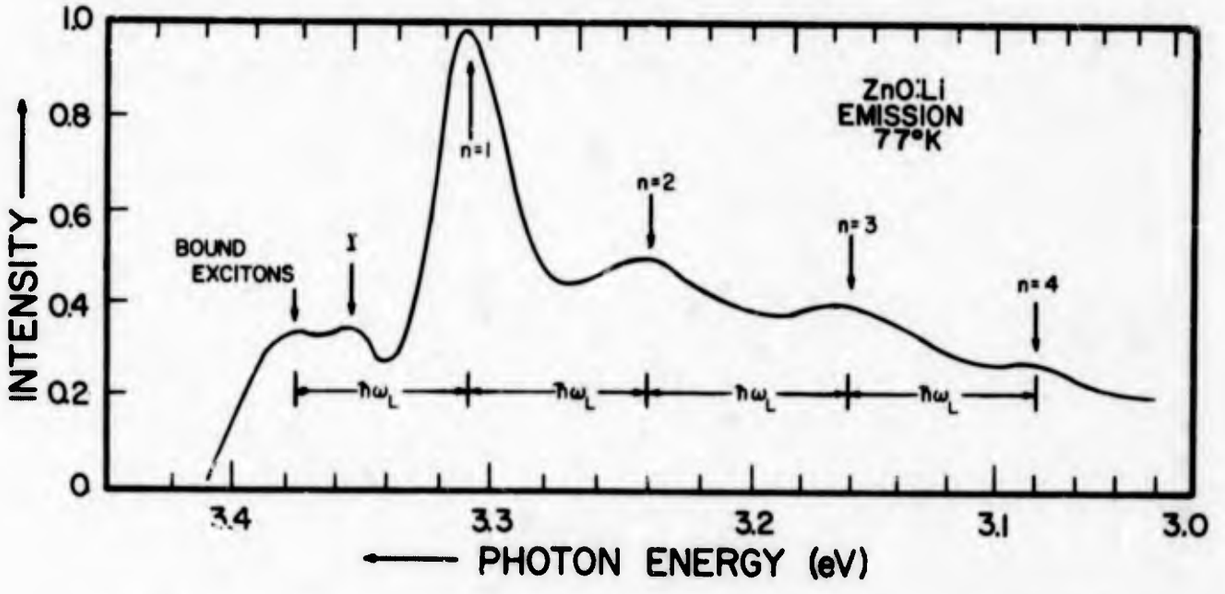


Fig. 5(b). Fluorescence Spectrum of a Li-doped ZnO "tap" crystal at 77°K. Note that the intensity of the I line is greatly reduced than in "normal" ZnO. In both "normal" and "tap" crystals the zero-phonon peak corresponds to the bound exciton emission lines.

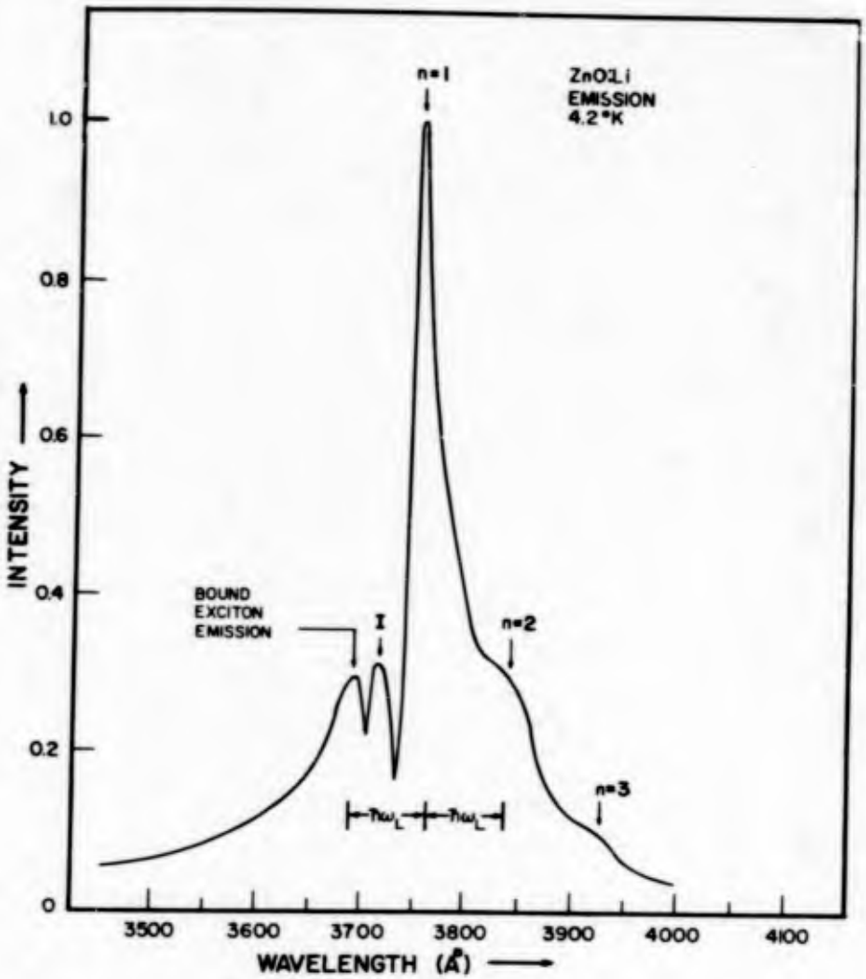


Fig. 6. Fluorescence Spectrum of a Li-dope ZnO "tap" crystal at 4.2°K.

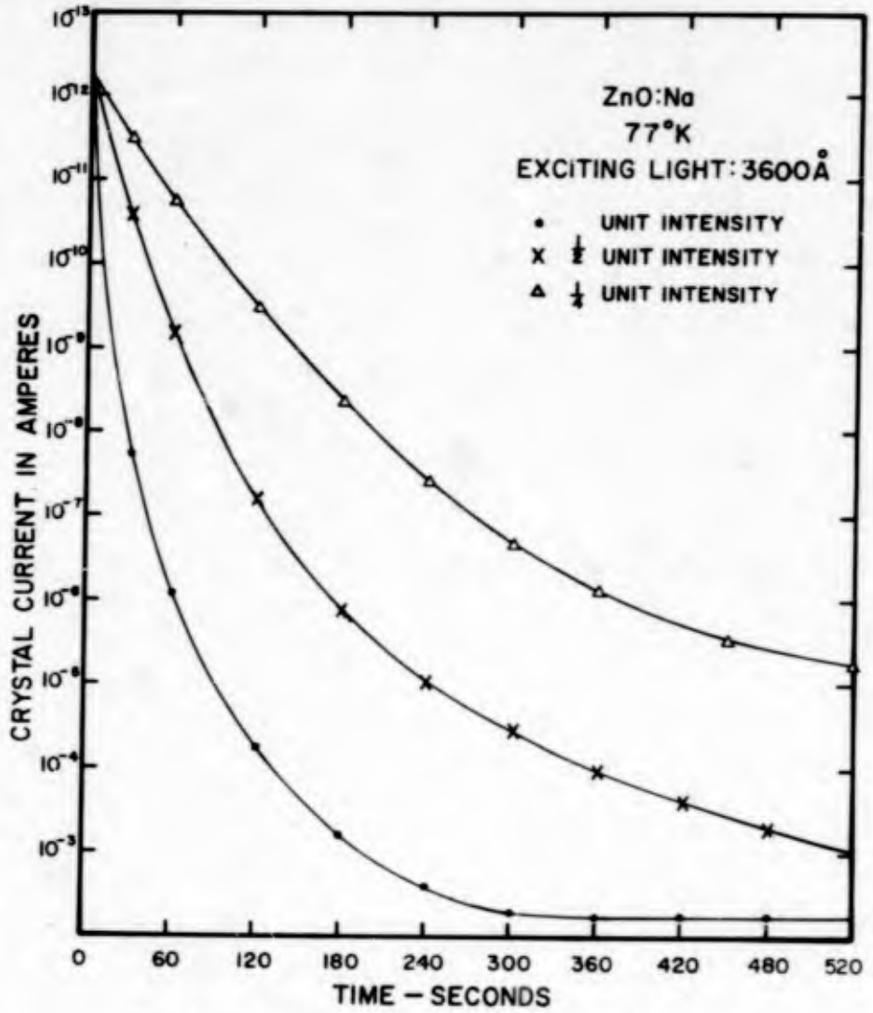


Fig. 7. The build-up of storage current in a Na-doped ZnO crystal as a function of time at various illumination levels.

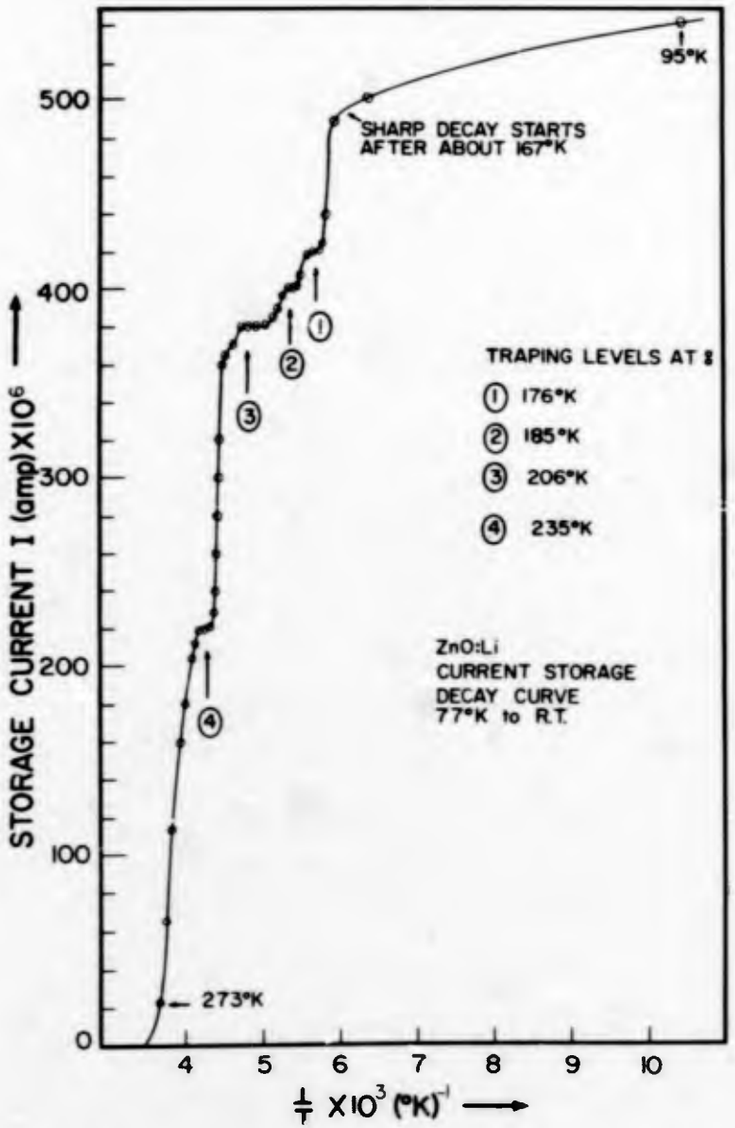


Fig. 8. Thermally stimulated current in a Na-doped ZnO crystal. Pronounced decay of the current begins at about 167°K. Four distinct traps are noticed at the temperatures 176, 185, 206 and 250°K.

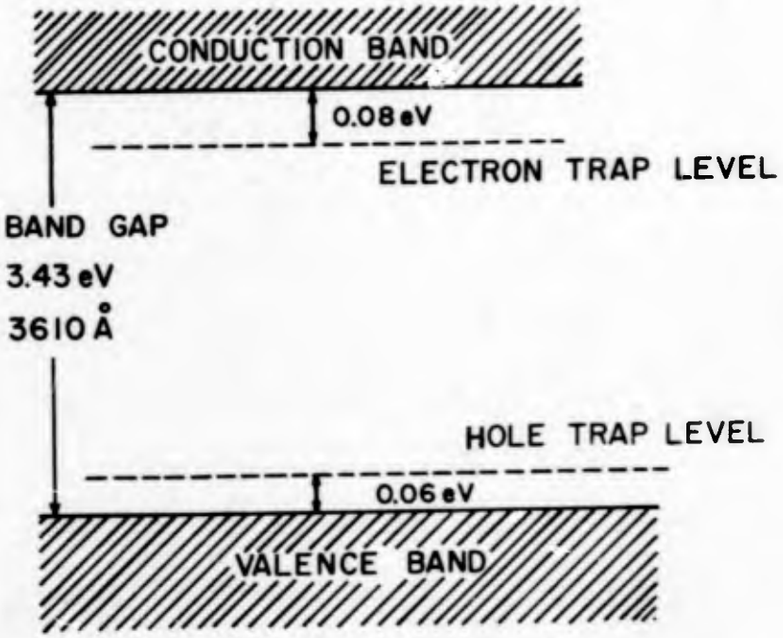


Fig. 9. Energy model for edge emission in ZnO "tap" crystals.

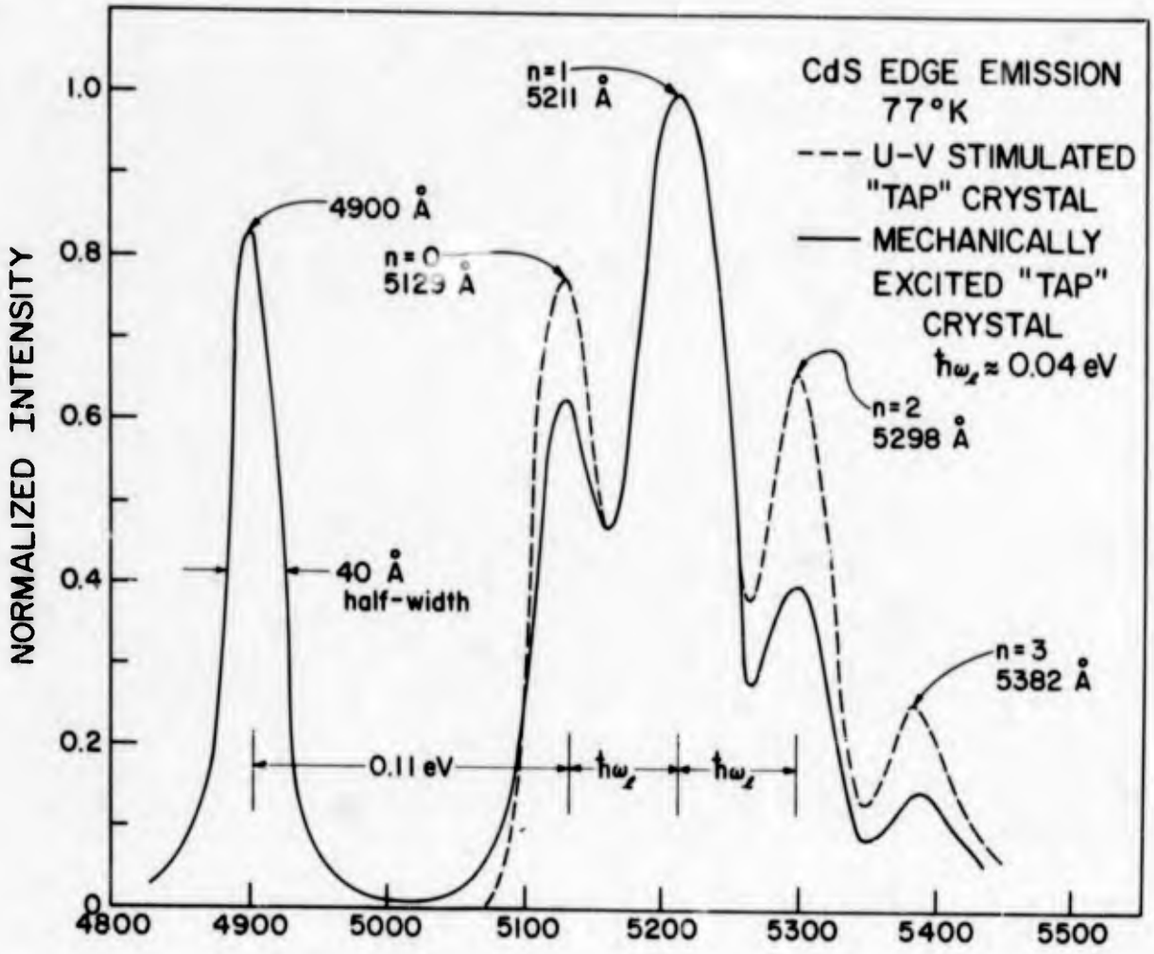


Fig. 10. CdS "Tap Emission." Note that both the mechanically excited and uv simulated spectra are normalized to unity at n = 1 peak (5211Å) of the phonon-assisted-series in both spectra.

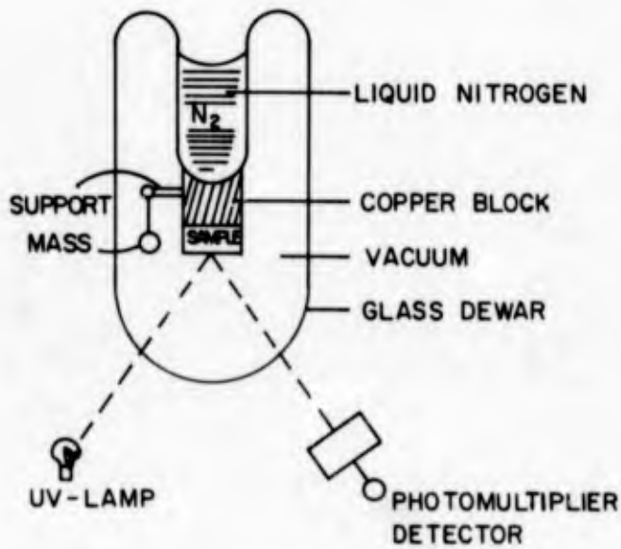


Fig. 11. Energy Storage Device.

ELECTRICAL AND MAGNETIC PROPERTIES OF (U) SINGLE CRYSTALS OF RARE EARTH TUNGSTEN BRONZES AND URANIUM TUNGSTEN BRONZES

by

Charles V. Collins, Captain, USAF

Air Force Institute of Technology
Air University
Wright - Patterson AFB, Ohio

BIOGRAPHY

Charles V. Collins was born on 18 April 1938 in Gary, Indiana, the son of Richard Carter Collins and Martha Parry Collins. After completing his work in 1956 at Crown Point High School, Crown Point, Indiana, he entered the United States Naval Academy. In June of 1960, he was graduated with the degree of Bachelor of Science in Engineering and received a commission of Second Lieutenant, Regular Air Force. He was assigned to the 566th Strategic Missile Squadron at Francis E. Warren AFB, Wyoming, where he served as a missile launch officer on an Atlas "E" ICBM combat crew. In April of 1964, he attended the 64B class of Squadron Officer School at Maxwell AFB, Alabama. Upon graduation in August of 1964, he was reassigned to the Air Force Institute of Technology at Wright-Patterson AFB, Ohio. In June of 1966, he completed graduate work in Space Physics and graduated with a Master of Science degree. He was then assigned to the United States Air Force Academy, Colorado, where he is presently serving as an Instructor of Physics.

Capt Charles V. Collins

BLANK PAGE

ABSTRACT

Rare earth tungsten bronze single crystals have been prepared for the first time by the method of electrolysis. $Sm_{.09}WO_3$ and $Gd_{.19}WO_3$ crystals were grown from a melt of rare earth chloride and tungsten trioxide at temperatures ranging from 1050° C to 1300° C. Sodium tungsten bronze single crystals, Na_xWO_3, with x ranging between .37 and .81 were also prepared. These crystals were grown from melts of fused sodium tungstate, tungsten trioxide, and tungsten at temperatures from 700° C to 900° C. Powder samples of a new tungsten bronze, uranium tungsten bronze, have been synthesized by a thermal reaction of tungsten trioxide, uranium oxide, and tungsten. All samples were investigated by x-ray diffraction methods. Magnetic measurements using a Faraday balance indicate that uranium is tetravalent in the tungsten bronze compound. Magnetic measurements also provided a means of analysis for the rare earth bronze crystals whose magnetic moment is due to the paramagnetism of the trivalent rare earth ions. Electrical conductivity measurements were made on all samples using a four point potential probe technique. A special conductivity cell was designed for this purpose. The metallic nature of the cubic rare earth and uranium tungsten bronzes was observed during these measurements. Because the tungsten bronzes are light in weight and have electrical properties that can be varied, they may have application in airborne equipment where weight is an important consideration.

TABLE OF CONTENTS

INTRODUCTION

The term "tungsten bronze" refers to a series of non-stoichiometric compounds whose general formula can be represented by M_xWO_3. In this formula, M stands for one of several metallic elements and x has some value between zero and one The tungsten bronzes can be thought of a solid state defect structures where WO_3 forms a host lattice into which M atoms are introduced interstitially. These M atoms randomly occupy lattice positions and are dissociated into ions and free electrons. To date, tungsten bronzes have been prepared with sodium (Reference 3), lithium (Reference 21), potassium, cesium, and rubidium (Reference 12), silver (Reference 19), thallium (Reference 18), lead (Reference 2), barium (Reference 6), and the rare earths (Reference 15). The tungsten bronzes are of interest because of their unusual electrical properties such as exhibiting semiconducting as well as metallic behavior, their wide range of homogeneity, the symmetry of their crystal structures, and their resistivity against strong mineral acids. In addition, they may have application in airborne equipment because of their light weight. Therefore, in the past, most of the work with tungsten bronzes has been done or supported by the United States Air Force or the United States Atomic Energy Commission.

In general, the structure of the tungsten bronze varies with the size of the introduced M atom and with the value of x. In the simplest case, the structure is cubic, but hexagonal, tetragonal, orthorombic, and monoclinic bronzes are also known. The unit cell of a cubic tungsten bronze is shown in Figure 1. The introduced metal atom is located at the center of a perovskite-type structure. At the corners of the cell are tungsten atoms with an oxygen atom located midway between the tungsten atoms on each edge

Previous investigations (References 3,7,8) have shown the metallic nature of the alkali tungsten bronzes. In order to explain the origin of this metallic conductivity, threetheoretical models based on crystal field theory are discussed These models differ in their basic bonding mechanisms. The three difference mechanisms are:

1) the direct overlap of M-M orbitals
2) the overlap of W-W $5d_\epsilon$ orbitals
3) the overlap of W-O-W orbitals

The first mechanism suggest that the conductivity of the bronzes results from the formation of a conduction band when the wave functions of the introduced metal atoms overlap. Mackintosh (Reference 11) has proposed a model of the Na_xWO_3 based on this mechanism. The conduction band is supplied with electrons from the valence electrons of the metallic atoms. This model, therefore, assumes that the introduced metal atoms are responsible for both the formation of the conduction band and the supply of the conduction electrons.

The second mechanism proposes the formation of a conduction band by the overlap of the tungsten $5d_\epsilon$ orbitals. According to Morin (Reference 13), the $5d_\epsilon$ orbitals of W(VI) extend extremely far and overlap, forming a conduction band in WO_3. Normally, this band is empty and WO_3 is an insulator. However, when the metal atoms are introduced, donor levels are added in the forbidden energy gap. These levels are then ionized and the conduction band is supplied with electrons. The result is the observance of an increased conductivity.

The third mechanism, proposed by Goodenough (Reference 9), is known as the superexchange model. It is based on an overlap of tungsten $5d_\epsilon$ wave functions with the p wave functions of the oxygen atoms. They form a delocalized molecular orbital with a π bond character. This mechanism also assumes conductivity occurs when the introduced atoms donate their valence electrons to the normally empty conduction band found in WO_3. The essential difference between the last two models is the mechanism by which the conduction band is formed in WO_3.

Until now very little attention has been paid to the relationship between the electron concentration and the structural symmetry of the tungsten bronzes. For instance, the structure of sodium tungsten bronze changes from tetragonal to cubic (around $x = 0.25$) as more sodium atoms are introduced into the WO_3 lattice. This might be interpreted to be solely the result of the number of introduced atoms. If this were true, identical structures of the tungsten bronzes could be formed by adding the same number of metallic atoms to the tungsten trioxide. This is not the case, however, for the rare earth bronzes experience the transformation from tetragonal to cubic at $x = 0.085$. The ratio of x-value of the trivalent rare earth atoms to the x-value of the monovalent sodium atoms suggests the transformation occurs in the tungsten bronzes when a certain electron concentration exists. The concept of electron number controlled crystalline transformations is supported by the fact that thermal excitation of WO_3 causes the transition from monoclinic to orthorombic to tetragonal (Reference 16).

The purpose of this investigation was to make conductivity measurements on the rare earth tungsten bronzes to determine if they also display metallic conductivity in their metal richest phase. These measurements were to determine the value of resistivity versus x-value, resistivity versus temperature, the electron concentration, and the temperature coefficient. As the result of the measurements, it was hoped to find which mechanism most accurately explained the metallic nature of the tungsten bronzes. Because measurements on powders suffer from poor irreproducibility and result in inaccurate absolute values of conductivity due to grain boundary resistances in the powder, single crystals of the rare earth bronzes were grown and used to obtain the conductivity data.

A second part of this work was concerned with the investigation of a new tungsten bronze: the uranium tungsten bronze. This bronze was prepared to check a concept which relates the electron concentration to the crystallographic phases in M_xWO_3 systems. The results of structure investigation and magnetic measurements on the uranium tungsten bronze are discussed in this paper.

SAMPLE PREPARATION

Rare earth tungsten bronzes are prepared as powders by a thermal reaction of an appropriately proportioned mixture of rare earth oxide, tungsten trioxide, and tungsten metal. As an example, $Gd_{0.1}WO_3$ is formed according to the following equation:

$$Gd_2O_3 + 19\ WO_3 + W \rightarrow 20\ Gd_{0.1}WO_3$$

The starting materials shown on the left of the equation are ground together, placed in a silica ampule which was sealed off while its contents were in vacuo, and annealed for approximately 120 hours at 1000°C. The ampule and its contents were then quenched in cool water. To insure reaction completeness, the contents of the ampule were reground and reheated by the process just described. As a further check, the product of the reaction was investigated by x-ray diffraction methods. Uranium tungsten bronze powders were prepared in a similar manner.

Single crystals of rare earth tungsten bronzes were prepared by the method of electrolysis. A suitable starting mixture was heated in a cubicle which was positioned in the center of a cylindrical furnace. When the mixture was molten, two electrodes were placed in the melt and current passed through it. The result was an oxidation-reduction process. The positive ions migrate to the cathode, forming a crystal around it. The negative oxygen ions go to the anode and are liberated.

Sodium tungsten bronze crystals were grown before attempting to prepare rare earth tungsten bronze single crystals. A starting mixture of Na_2WO_4 and WO_3 in the appropriately proportioned molar ratio was used in conjunction with platinum electrodes and a platinum crucible. The crystals obtained were varied in composition, structure, and size by changing the molar ratio of Na_2WO_4 to WO_3, the temperature, the current, and the time of electrolysis. They were grown at temperatures ranging from 700°C to 900°C with 10-30 milliamperes of current flowing through the melt. The time of electrolysis was from four to 48 hours. The Na_xWO_3 crystals obtained were all sturdy, well formed crystals, the largest of which were one centimeter or more on each side. The color of the crystals varied from violet at $x = 0.37$ to bright yellow at $x = 0.81$. Figure 2 is a photograph of single crystals of $Na_{.85}WO_3$.

The rare earth tungsten bronze single crystals were grown using the same basic procedure, but from different starting mixtures and at higher temperatures. For instance, $GdCl_3$ was mixed with the WO_3 in a molar ratio of 1:3 which reacted as follows:

$$2GdCl_3 + 6WO_3 \rightarrow Gd_2(WO_4)_3 + 3WO_2Cl_2$$

As the mixture is heated, the WO_2Cl_2 comes off as a gas leaving a melt of the gadolinium tungstate to be electrolized. The first gadolinium crystals were grown with the temperature at 1050° C and 10 milliamperes of current flowing through the melt. Electrolysis was performed for 16 hours. For subsequent crystals, the time of electrolysis, temperature, and current were all varied. The crystals obtained were irregularly shaped and blue in color, the largest being about eight millimeters in length and two millimeters in diameter at the thickest point. Figure 3 is a photograph of single crystals of $Sm_{.09}WO_3$.

X-RAY INVESTIGATION

All samples were examined by x-ray diffraction techniques using nickel filtered copper radiation and a Phillips 114 mm diameter Debye Scherrer camera in conjunction with the standard Norelco x-ray equipment. All patterns showed the sharp lines of a single phase and features typical of body centered cubic samples. The films were indexed and corrected for shrinkage, and the lattice constants for all cubic samples were determined by the Nelson-Riley extrapolation method (Reference 14). Table 1 shows the results of indexing one of the Na_xWO_3 crystals.

TABLE 1
Powder Diffraction Data of Cubic $Na_{.00}WO_3$
$a_o = 3.842$

d_{obs}	I	hkl	d_{calc}
1.0649	M	023	1.0655
1.0259	S	123	1.0267
.9607	W	004	.9604
.9318	VS	014	.9317
.9057	S	033	.9055

The columns of Table 1 show the observed and calculated d-spacings, the intensity of the diffraction lines, and the Miller indices. From the x-ray investigation, the uranium tungsten bronze was found to be cubic at x = 0.078 and tetragonal at x = 0.075. All rare earth tungsten bronze crystals were found to be cubic. Figure 4 shows the lattice parameter a_o as a function of the x-value for Gd_xWO_3.

MAGNETIC MEASUREMENTS

The magnetic susceptibility of several of the cubic rare earth tungsten bronze samples and the cubic uranium tungsten bronze samples was determined with a Faraday balance using the horizontal force method (Reference 10). The susceptibility of the sample can be determined from the force using the following relationship:

$$\chi_g = F/(mH \cdot \frac{dH}{dx})$$

where χ_g is the gram susceptibility, m is the mass of the sample in grams, and $H \cdot \frac{dH}{dx}$ is the intensity of the field dotted with the field gradient.

The magnetic measurements on cubic $U_{.08}WO_3$ show that this tungsten bronze is paramagnetic. Table II shows various magnetic quantities for $U_{.08}WO_3$ over a range of temperatures.

TABLE II
Magnetic Properties of $U_{.08}WO_3$

T(°K)	$\chi_g \cdot 10^6$	$\chi_m \cdot 10^6$	$1/\chi_g \cdot 10^6$	μ
285	1.20	3764	83	3.46
243	1.35	4234	.74	3.47
226	1.40	4391	.71	3.45
196	1.55	4861	.64	3.46
164	1.70	5331	.59	3.44
132	1.95	6116	.51	3.46
110	2.15	6743	.47	3.46
89	2.40	7527	.42	3.48
85	2.45	7684	.41	3.48

The effective magnetic moment is calculated from the following equation (Reference 17):

$$\mu_{eff} = 2.84[\chi_m (T-\theta)]^{\frac{1}{2}}$$

where μ_{eff} is the effective moment, χ_m is the molar susceptibility, T is the temperature in °K, and θ is the Weiss constant. The Weiss constant comes from the Curie-Weiss law ($\chi_m = C/T-\theta$) which gives the temperature dependence of the susceptibility of a compound. θ is determined graphically from a plot of the reciprocal of gram susceptibility against temperature. When the plot is extrapolated to cross the abscissa, the point of intersection determines θ in degrees from the origin. Figure 5 shows the determination of θ for $U_{.08}WO_3$ (θ = -113) while Figure 6 shows θ for $Gd_{.19}WO_3$ (θ = 0).

The paramagnetism of the rare earth elements made it possible to analyze the rare earth tungsten bronze single crystals by magnetic measurements. A calibration curve of susceptibility versus x-value was constructed for the Gd_xWO_3 and Sm_xWO_3 compounds By entering the curve for gadolinium shown in Figure 7 with the value of susceptibility that was measured, each of the gadolinium tungsten bronze single crystals was analyzed.

CONDUCTIVITY MEASUREMENTS

The conductivity measurements were made using a four point potential probe technique with a conductivity cell designed for this purpose. The cell allowed accurate resistivity measurements to be made with a minimum of error due to contact resistance. This cell consists of a cylindrical block of annealed copper upon which rests the crystal to be measured. The crystal sits in the center of the block, insulated from it by two layers of sheet mica which are attached to the copper by epoxy glue. The specimen is situated between two copper electrodes, one fixed and one springloaded to exert pressure for better contact. These electrodes, through which the current passes, are also insulated from the main block. The voltage drop across the crystal is picked off by two tungsten probes which contact the crystal from above. They are also springloaded and insulated from the block. To insure good contact between the crystal and the current electrodes, black platinum paste was smeared on the faces of the electrodes.

The circuit in Figure 8 was assembled to complete the conductivity measurements. The power supply was a Hewlett-Packard constant voltage/constant current supply. A differential voltmeter, made by John Fluke Company, was one means used to measure the voltage drop across the crystal. Another instrument used was a Rubicon 5 dial potentiometer. One instrument was used to cross check the other so that precise readings of the voltage drop could be obtained. A Hewlett-Packard microvoltmeter was used to measure the thermocouple voltage. The variable resistor was adjusted to set the amount of current flowing in the circuit after the power supply was set at a desired voltage. A Weston milliammeter, accurate to one ampere, displayed the value of current directly. Once both the current flowing in the circuit and the voltage drop across the crystal were recorded, the resistivity of the crystal was computed.

Measurements of resistivity were made along a temperature range of 100°K to 350°K. Figures 9 and 10 show plots of resistivity against temperature for $Na_{.65}WO_3$, $Gd_{.19}WO_3$, and $Sm_{.09}WO_3$. In all three plots, the resistivity is seen to increase with rising temperature. Also the points fall approximately in a straight line in each case. Figure 11 shows the resistivities of the gadolinium tungsten bronze powders that were measured as a function of x-value. This graph shows that as the electron concentration increases, the resistivity of gadolinium tungsten bronze at room temperature decreases. Preliminary measurements on the uranium tungsten bronze powders indicate metallic behavior for this cubic tungsten bronze also. However, since single crystals of the bronze were not available, the absolute value of conductivity was not obtained. Still, powder measurements do reflect the electrical nature of the compound to be metallic, as shown in Figure 12.

The results of the conductivity calculations for the three single crystals measured along the full temperature range are summarized in Table III. All three crystals showed positive temperature coefficients of resistivity. In the table, the observed conductivity is merely the reciprocal of the resistivity measured at room temperature. The calculated conductivity is determined by the equation $\sigma = neu$ where σ is the conductivity, n is the number of conduction electrons per cc., e is the electronic charge, and μ is the mobility of the conduction electrons.

Table IV
Conductivity of Various Bronzes

Bronze	σ obs	σ calc
$Na_{.85}WO_3$	$2.91 \cdot 10^4$	$2.97 \cdot 10^4$
$Gd_{.19}WO_3$	$6.81 \cdot 10^3$	$6.93 \cdot 10^3$
$Sm_{.09}WO_3$	$3.65 \cdot 10^3$	$3.74 \cdot 10^3$

DISCUSSION

Sample Preparation

The method of preparing the rare earth tungsten bronzes as powders was well established prior to the start of this investigation. The same general procedure also proved to work for the preparation of the uranium tungsten bronze powders which had not been prepared previously. However, the growth of rare earth tungsten bronze single crystal presented some problems. The main cause of these problems was the high temperature that was required because of the high melting point of the rare earth compounds being used. This high temperature caused various reactions between the melt and the crucible, and between the melt and the electrodes. Another problem caused by the high temperature involved keeping the composition of the melt constant as WO_3 becomes highly volatile above 800°C. Other problems not specifically related to the temperature included coordinating the time of electrolysis with the current and also finding the optimum molar ratio for the starting mixture. The best results were obtained when electrolysis was carried out for approximately two hours with the use of medium amounts of current (around 40 milliamperes). The best starting mixture proved to be about 1:3 molar ratio of rare earth chloride to tungsten trioxide. The rare earth chloride reacted quickly with the WO_3 to form rare earth tungstates, which are the lowest melting compounds. The tungstate then electrolyzes quickly, forming crystals of up to one centimeter in length and several millimeters in diameter.

X-ray Investigation

As shown in Figure 2, the lattice constant versus x-value

curve levels off rapidly as the value of x approaches the limiting value for the cubic phase. It is assumed that this leveling off is caused when the crystal becomes saturated with quasifree electrons so that the unit cell dimensions stabilize near some value of a_o. A similar effect is noted at high x-values for the sodium tungsten bronzes (Reference 4).

The x-ray investigation does seem to support a concept relating the electron concentration to phase. If the transformation from cubic to tetragonal depends on a certain number of free electrons, then this transformation can be roughly predicted for a certain x-value of a particular tungsten bronze by knowing the valency of the introduced metal atoms. This assumes that all electrons dissociated from the introduced atoms are part of the free electron gas. Based on this electron concentration-structure hypothesis, a calculation was made to determine what sample composition should be prepared to obtain the cubic structure of uranium tungsten bronze. From the experience with the sodium tungsten bronzes (these bronzes crystallize cubic at about x = .3 with sodium having a valency of +1) and the rare earth bronzes (these bronzes form cubic at x = .1 when the rare earth atoms show a valency of +3), it appears that the product of x-value and valency is nearly constant at .3 for the cubic tungsten bronzes. Therefore, the tetravalent uranium tungsten bronzes were predicted to be cubic at approximately x = .075. From the preparation of these bronzes and the results of the subsequent x-ray investigation, the transformation from cubic to tetragonal was found to actually take place near x = .078. The range of homogeneity for the cubic phase was not completely determined but extends beyond x = .1 at least.

Magnetic Measurements

Table II indicates that $U_{.08}WO_3$ has an effective magnetic moment of 3.46 Bohr magnetons. This moment compares well with the moments of uranium sulfate, uranium oxalate, and uranium acetylacetonate which have moments of 3.52, 3175, and 3.21 respectively. Because uranium is tetravalent in these three compounds, it is suggested that uranium in the tungsten bronzes is present in the tetravalent state. The presence of tetravalent uranium can also be determined by considering the possible values of the moments for various other valence states of uranium. For trivalent uranium a higher moment has to be expected than the moment observed for $U_{.08}WO_3$. Hexavalent uranium would exhibit no moment. Kossel's rule states that if two ions are isoelectronic, they should display the same effective magnetic moment. Therefore, since pentavalent uranium is isoelectronic with hexavalent neptunium, it should show a moment of 1.7 Bohr magnetons. This same effect is observed for Eu^{+2} and Gd^{+3} which both have seven 4f

electrons and both exhibit an effective moment of 7.9 Bohr magnetons. Since the other valence states of trivalent, pentavalent, and hexavalent uranium would result in an effective moment that differs greatly from the observed moment, they may be ruled out as possible configurations of uranium in the tungsten bronzes. The elimination of U(III), U(V), and U(VI) as possible valence states leaves only U(IV) which must be present.

The remaining, and often discussed, question is whether uranium acts as a transition d-element or as an f-element (rare earth). If it is considered a d-element, then the spin alone accounts for the moment of the ion and the orbitals are assumed to be quenched. A magnetic moment of about 2.7 Bohr magnetons would have to be expected for a "transition metal U(IV)." If uranium acts as a rare earth (5f element), the orbitals would resist quenching (due to the L-S coupling) and uranium would have an effective moment of 3.6 Bohr magnetons. Uranium (IV), acting as a rare earth, can be compared to praseodynium (III) which also has two f electrons, and which has an effective moment of 3.6 Bohr magnetons. The observed moment of 3.46 Bohr magnetons for U(IV) leads to the conclusion that uranium, at least in the tetravalent state, acts as a rare earth.

As an analysis tool for the rare earth tungsten bronzes, the magnetic measurements proved useful. The nondestructive analysis is based on the assumption that tungsten in the bronzes is hexavalent and that the paramagnetism, therefore, results from the introduction of the rare earth atoms. The construction of a calibration curve allowed the determination of the x-value of each of the crystals from its measured susceptibility.

Conductivity Measurements

The plots of resistivity versus temperature for the $Sm_{.09}WO_3$, $Gd_{.19}WO_3$, and $Na_{.65}WO_3$ single crystals show two important characteristics. First, as the temperature is increased, the resistivity of the crystals increases. Also, the dependence of electrical resistance on temperature is nearly linear. These two points imply a metallic nature for the sodium and rare earth bronzes studied in this investigation. The metallic behavior of sodium tungsten bronze has been established previously (References 3, 7, 8) and is confirmed by the present work. All the samples measured displayed an inverse relation between electron concentration and resistivity and had positive temperature coefficients of resistivity.

The effects of grain resistance on conductivity measurements are shown when the resistivity at room temperature for powder samples is compared to the resistivity of a single crystal.

For instance, at room temperature the resistivity of $Gd_{.20}WO_3$ powder is $2.1 \cdot 10^{-3}$ while that of $Gd_{.19}WO_3$ single crystal is much lower, having a value of $1.4 \cdot 10^{-4}$ ohm-cm. The main disadvantage of powder measurements is that the absolute value of resistivity cannot be determined when grain boundaries are present. Single crystals do not have grain boundaries and in addition, if cubic, do not have a preferred orientation with respect to conductivity.

Sienko (Reference 5) found a very low magnetic moment for the sodium tungsten bronzes indicating the absence of W(V) ions (if present, W(V) would exhibit a moment of 1.5 Bohr magnetons). On the basis of these measurements, Sienko assumed that all tungsten atoms in the bronzes are best treated as hexavalent. This assumption was adopted during this work to allow the analysis of the $Gd_{.19}WO_3$ and $Sm_{.09}WO_3$ crystals as previously mentioned. While Sienko's conclusion is the logical one, it is possible for pentavalent tungsten to have antiparallel spins showing an antiferromagnetic contribution to the effective moment and resulting in a moment near zero. This may give the impression that all tungsten are hexavalent. However, the present conductivity measurements show that Sienko is correct. From the formula for conductivity ($\sigma = ne\mu$), the number of conduction electrons present in the bronzes can be estimated. To determine n, the observed conductivity was used and the carrier mobility was calculated from the Hall mobilities reported in previous investigations (References 3, 8, 20) by constructing a calibration curve of mobility versus x-value. The calculation shows that the number of conduction electrons in the rare earth tungsten bronzes is approximately three times the number of rare earth atoms present as determined from the molecular weight and density of the crystal. Since gadolinium and samarium are both trivalent, it is assumed that all donated electrons are conduction electrons and that none are trapped with the tungsten. Therefore, the tungsten must display a valency of $^{+}6$ in the tungsten bronze compounds.

The formula $\sigma = ne\mu$ can be manipulated another way to show the characteristics of the tungsten bronzes. If three electrons are assumed to be donated by the rare earth atoms, the mobility of the rare earth bronze can be calculated from the observed conductivity. This mobility decreases with rising temperature and increases with rising electron concentration. Sienko and Truong offer possible explanations for these changes (Reference 20).

The decrease in mobility with rising temperature can be explained by considering the thermal motion of the atoms and electrons in the bronzes. As the temperature is increased, there is

greater scattering of the electron carriers which effectively lowers the mobility. The reason for the increase in mobility with increasing electron concentration is, however, not immediately apparent. There are three possible explanations to account for this increase. One explanation could be that there is an impurity banding in the bronzes. A second possibility is that the alkali ions introduce point charge defects in the WO_3 structure. These extra positive charges break up the band structure at low electron concentrations, but have less effect when high electron concentrations are present. In other words, the electrons remain localized about the positive charges at low x-value, but as more electrons are introduced, the effects of the positive charges are smoothed out and the electrons at high x-value become quasifree. A third possibility is that ionic scattering may account for relaxation in the conduction process. As WO_3 is doped, the conduction band is postulated to be raised to higher energies, thus effectively making the energies of the conduction electrons higher as x increases. Since the ionic scattering cross section for electrons with high energies is less than for lower energy electrons, the mobility would be greatest in the metal richest state.

Sienko and Truong favor the second explanation of delocalization of electrons as the electron concentration increases. Impurity banding is doubtful since no Knight shift is detected in the nuclear magnetic resonance studies of Na^{23} in the Na_xWO_3 (Reference 1). The ionic scattering mechanism looks incorrect since it predicts a decrease in mobility as the3/2 power of temperature while a linear decrease with temperature is observed.

Three basic mechanisms which attempt to explain the origin of the conductivity of the tungsten bronzes were mentioned in the INTRODUCTION. Of these three, the metallic behavior of the bronzes appears to be best explained by theW-O-W superexchange as proposed by Goodenough. The W-W overlap mechanism seems doubtful since the distance of 5.4 A across the unit cell seems to be too great for a pure W-W exchange. The direct M-M overlap is possible for the alkali bronzes where the separation between metal atoms is not too great, but in the rare earth bronzes, only about one unit cell in ten have an introduced rare earth atom. In spite of this large separation which could preculde M-M overlap, the present measurements show that metallic conductivity is still present. Therefore, it is probable that both the M-M overlap and W-W overlap mechanisms can be ruled out as the principal mechanism of the conductivity.

Error in the Measurements

The error in the magnetic measurements is ± 2% of the recorded susceptibility as determined by Lieutenant James (Reference 10).

The error in the present conductivity measurements was calculated by the normal method of indirect measurements. The relative errors were assumed independent of each other with the exception of the length between probes and the test area since they were measured with the same instrument. The error in resistivity is due to:

1) error in distance between voltage pickoff probes
2) error in crystal area exposed to the current
3) error in potentiometer readings
4) error in milliammeter readings
5) error in temperature readings

At room temperature, the overall error is estimated to be 4.5% of the reported resistivities. This moderate error, comparable to the errors previously reported, is mainly due to the difficulty in determining accurately the dimensions of the crystal.

CONCLUSIONS

Several results of this work appear helpful in understanding the nature of the tungsten bronzes. These results include the synthesis of powder samples of a new tungsten bronze, U_xWO_3, which was prepared for the first time. Magnetic susceptibility measurements confirmed that uranium is present in the tetravalent state in the tungsten bronze compound. These measurements further indicate that U(IV) tends to act as a rare earth, or f, element. Because the size of the uranium atom is comparable to the alkali and rare earth atoms, the transformation from cubic to tetragonal was predicted to occur at $x = .075$ The fact that this transformation was observed at about $x = .078$ supports the concept that electron concentration (not the number of introduced atoms) is the dominant factor in this phase transformation.

Single crystal preparation of the rare earth tungsten bronzes has also been accomplished for the first time. The electrical conductivity measurements on these crystals and the uranium bronze powders show clearly that these cubic tungsten bronzes are metallic. The inserted trivalent rare earth atoms donate three electrons to the conduction band. In the case of uranium tungsten bronzes, four electrons become conduction electrons. The results of the conductivity measurements also suggest that Goodenough's model of the tungsten bronze is best. The conductivity of the rare earth bronzes in spite of large separations between W-W and M-M atoms supports the W-O-W-superexchange concept.

BIBLIOGRAPHY

1. Barnes, R. G. et al. "Nuclear Magnetic Reasonances in Sodium Tungsten Bronzes." Bulletin of the American Physical Society, Series II, 4:166 (1959).

2. Bernoff, R. A. and L. E. Conroy. "Lead Tungsten Bronzes." Journal of the American Chemical Society, 82:6261 (1960).

3. Brown, B. W. and E. Banks. "Conductivity of the Sodium Tungsten Bronzes." Physical Review, 84:609 (1951).

4. Brown, B. W. and E. Banks. "The Sodium Tungsten Bronzes." Journal of the American Chemical Society, 76:963 (1954).

5. Conroy, L. E. and M. J. Sienko. "Magnetic Susceptibility of the Lithium Tungsten Bronzes." Journal of the American Chemical Society, 74:3520 (1952).

6. Conroy, L. E. and T. Yokokawa. "The Preparation and Properties of a Barium Tungsten Bronze." Inorganic Chemistry, 4:995 (1965).

7. Ellerbeck, L. D. et al. "Electrical Resistivity of Cubic Sodium Tungsten Bronze." The Journal of Chemical Physics, 35:298 (1961).

8. Gardner, W. R. and G. C. Danielson. "Electrical Resistivity and Hall Coefficient of Sodium Tungsten Bronze." Physical Review, 93:46 (1954).

9. Goodenough, J. B. "Transition-metal Oxides with Metallic Conductivity." Colloque Internatonaux du Centre National de la Recherche Scientifique, 149:162 (1964).

10. James, L. T., Jr. "A Magnetic Balance for Measuring Magnetic Susceptibilities." M. S. Thesis, Air Force Institute of Technology, 1965.

11. Mackintosh, A. R. "The Electronic Structure of Metal Tungsten Bronzes." Contribution No. 1209, Institute for Atomic Research and Department of Physics, Ames, Iowa, 1962.

12. Magneli, A. "Studies on the Hexagonal Tungsten Bronzes of Potassium, Rubidium and Cesium." Acta Chemica Scandinavica, 7:315 (1953).

13. Morin, F. J. "Electrical Properties of Fe_2O_3 and Fe_2O_3 Containing Titanium." Physical Review, 83:1005 (1951).

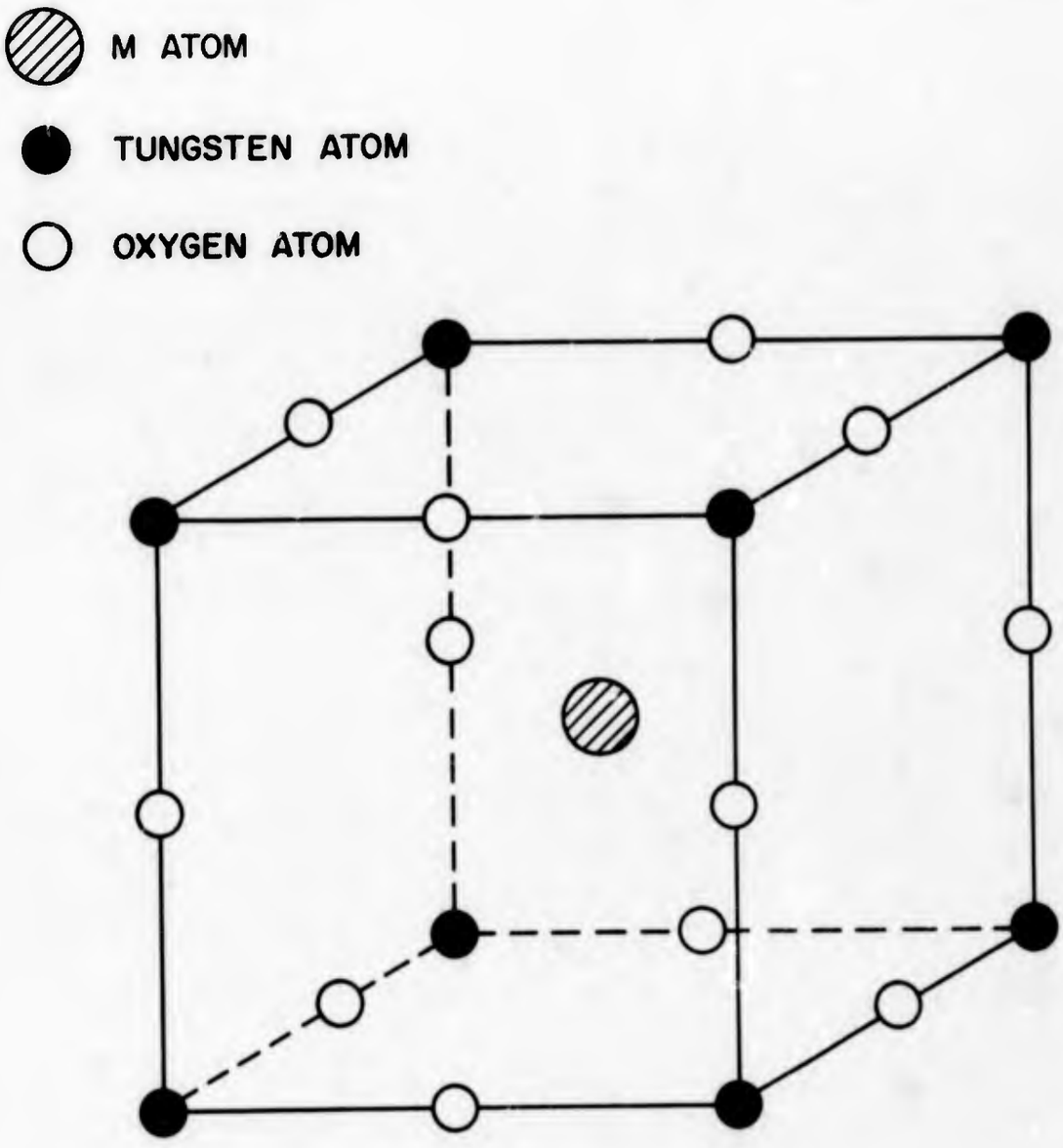


Figure 1. Unit Cell of M_xWO_3

14. Nelson, J. B. and D. P. Riley. "An Experimental Investigation of Extrapolation Methods in the Derivation of Accurate Unit - Cell Dimensions of Crystals." Physical Society of London 57:160 (1945).

15. Ostertag, W. "Rare Earth Tungsten Bronzes." Inorganic Chemistry 5:758 (1966).

16. Ribnick, A. S. et al. "Phase Transitions in Sodium Tungsten Bronzes." Nonstoichiometric Compounds. Washington, D. C.: American Chemical Society, 1963.

17. Selwood, P. W. Magnetochemistry. New York, N. Y.: Interscience Publishers, Inc., 1956.

18. Sienko, M. J. "Thallium-Tungsten Bronze: A Solid State Defect Structure." Journal of the American Chemical Society, 81:5556 (1959).

19. Sienko, M. J. and B. E. Mazumder. "Some Solid State Studies of Silver-doped WO_3." Journal of the American Chemical Society, 82:3508 (1960).

20. Sienko, M. J. and T.B.N. Truong. "Studies in Non-Stoichiometry. Electrical Conductivity and Carrier Mobility in Lithium Tungsten Bronzes." Journal of the American Chemical Society, 83:3939 (1961).

21. Straumnis, M. E. and S. S. Hsu. "The Lithium Tungsten Bronzes." Journal of the American Chemical Society, 72:4027 (1950).

ACKNOWLEDGMENTS

This investigation was conducted at the Air Force Materials Laboratory, Wright-Patterson AFB, Ohio, under Project No. 7360 and Task No. 736005.

Figure 2. Single Crystals of $Na_{.68}WO_3$

Figure 3. Single Crystals of $Sm_{.09}WO_3$

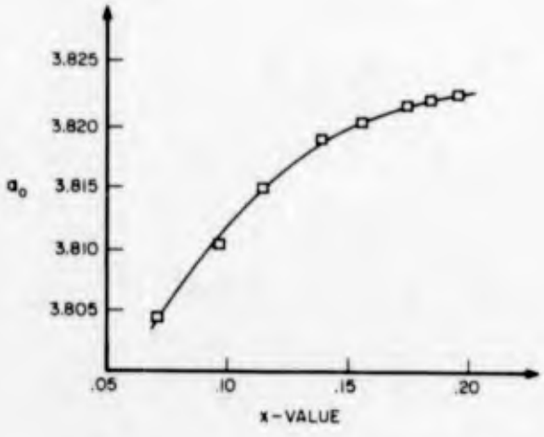


Figure 4. a_o versus x for Gd_xWO_3

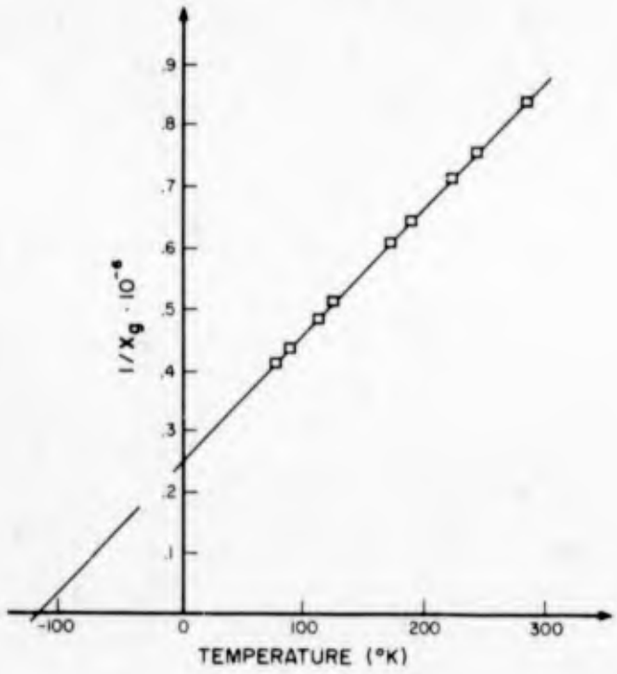


Figure 5. Determination of θ for $U_{1-x}WO_3$

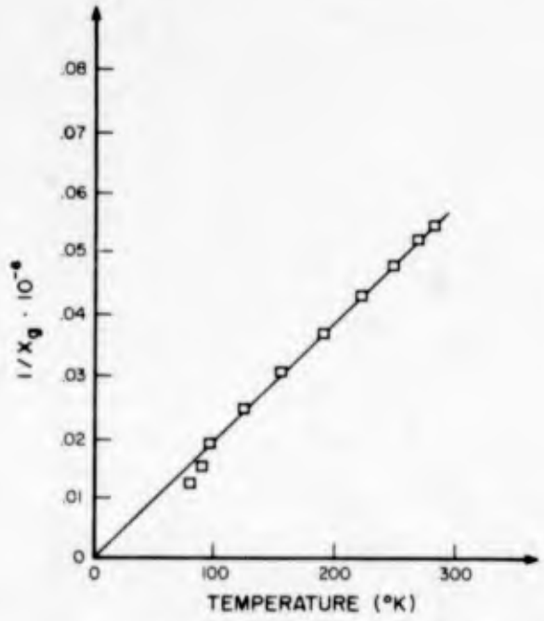


Figure 6. Determination of θ for $Gd_{.19}WO_3$

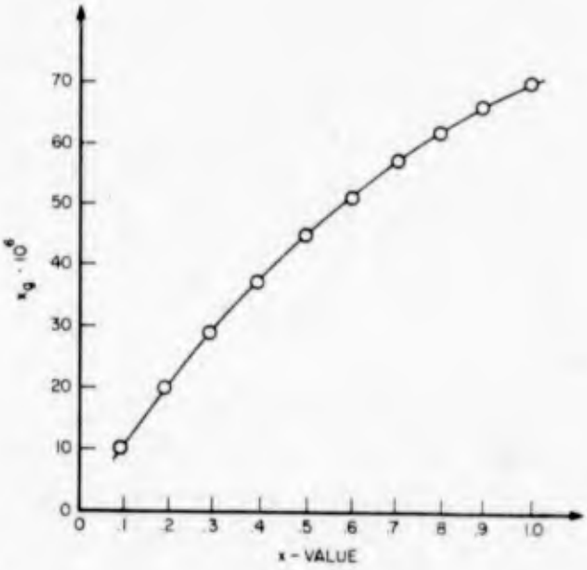


Figure 7. Calibration Curve for Gd_xWO_3

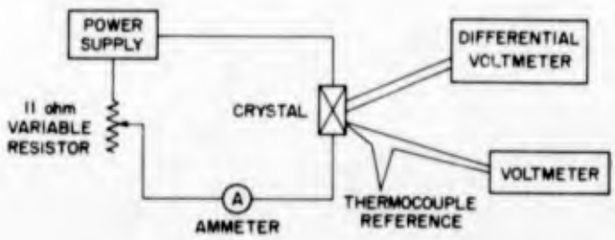


Figure 8. Conductivity Measurement Circuit Diagram

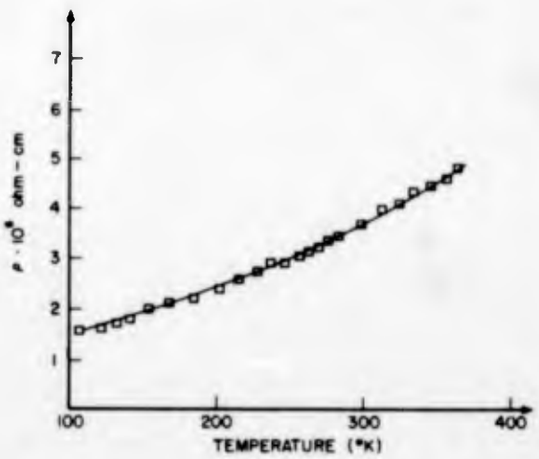


Figure 9. Resistivity versus Temperature for $Na_{.68}WO_3$

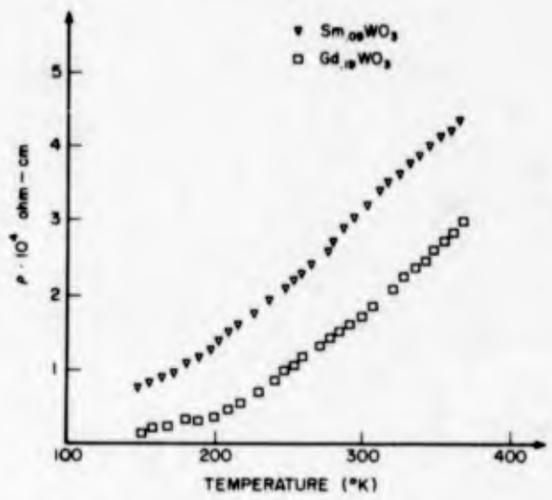


Figure 10. Resistivity versus Temperature for $Gd_{.19}WO_3$ and $Sm_{.08}WO_3$

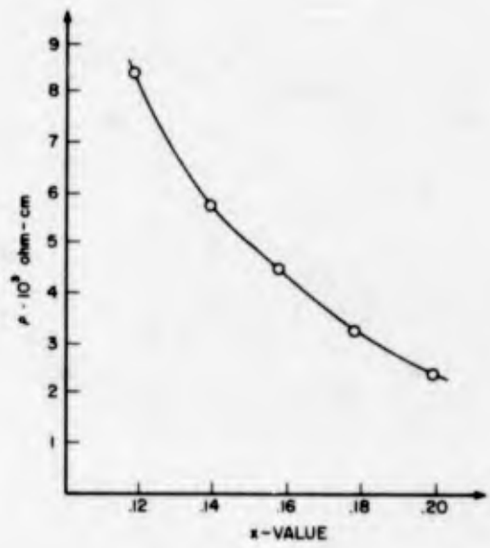


Figure 11. Resistivities of Gd_xWO_3 Powders versus x-value

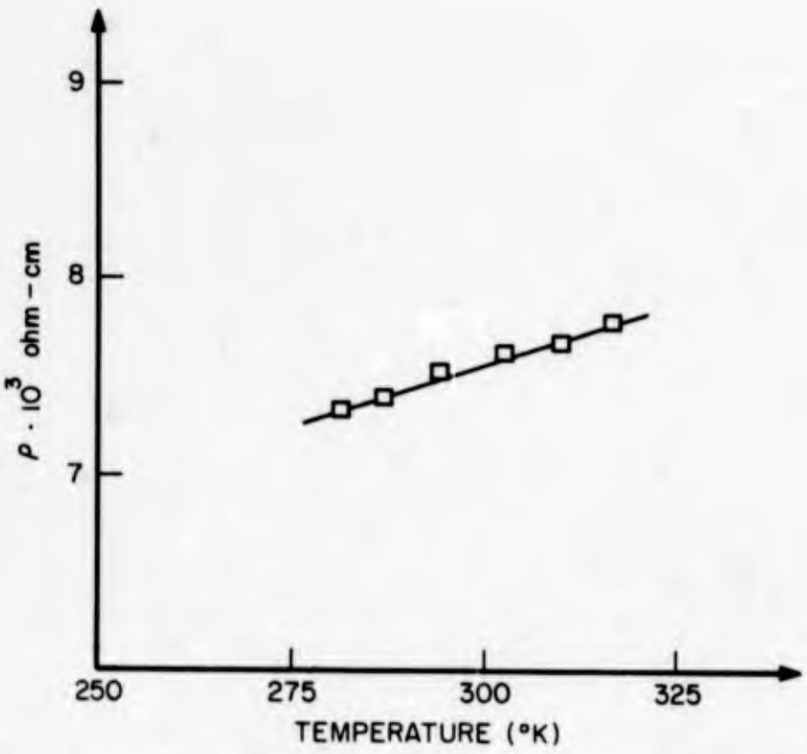


Figure 12. Resistivity versus Temperature for $U_{.08}WO_3$

DIELECTRIC RESONATORS

by

James C. Sethares
Martin R. Stiglitz

Air Force Cambridge Research Laboratories
Office of Aerospace Research
L. G. Hanscom Field, Mass.

James C. Sethares

BIOGRAPHY

James C. Sethares was born in Hyannis, Massachusetts, U.S.A. on December 13, 1928. He received the B.S. degree in electrical engineering from the University of Massachusetts in 1959 and the S.M.degree in electrical engineering from the Massachusetts Institute of Technology in 1962. He then joined the Microwave Physics Laboratory at AFCRL as a research engineer to investigate fundamental interactions between electromagnetic waves and matter. He has specialized in the study of microwave properties of single crystal solids. He is a member of the Institute of Electrical Engineers.

Martin R. Stiglitz

BIOGRAPHY

Martin R. Stiglitz was born on March 24, 1920 in Vienna, Austria. He came to the United States in 1939 and served in the U. S. Parachute Troops from 1942 to 1945, with service in New Guinea and the Philippines. He graduated from Northeastern University in 1959, with a BBA degree in Engineering and Management. He is presently attending the Boston University Graduate School where he is majoring in Physics. He has been with the Air Force Cambridge Research Laboratories since 1954 as an electronics engineer with the Microwave Physics Laboratory. He is engaged in investigation of interactions of acoustic energy with electromagnetic waves in solids. He is the co-author of the first prize winning paper "Phonon Interactions" presented at the 1961 Science and Engineering Symposium in San Francisco, California.

DIELECTRIC RESONATORS

James C. Sethares and Martin R. Stiglitz

Air Force Cambridge Research Laboratories
Office of Aerospace Research
L. G. Hanscom Field, Bedford, Massachusetts

ABSTRACT

High-dielectric-constant, low-loss dielectric crystals, such as rutile and strontium titanate, can be made to resonate at microwave frequencies, and thus find application in microwave technology, wherever conventional metallic microwave resonators are used.

This paper deals with rutile and strontium titanate crystals used as microwave filters, and the frequency tuning and stabilization of dielectric resonators. An analogy between high dielectric crystals and a water drop is presented for the purpose of possible rain drop size measurements. An analysis of the filter and tuned resonators, based on equivalent electrical networks, is made and the predictions of the analysis supported by experimental evidence.

INTRODUCTION

This paper concerns the use of low loss, high dielectric materials as microwave resonant structures. Solid state high dielectric resonators have certain advantages over conventional cavity resonators, such as miniature size and extremely high Q's, which make them attractive from a practical point of view. Suitable materials, such as rutile, strontium and barium titanate, are presently available for the purpose.

Some of the electromagnetic characteristics of high dielectric resonators have been known for many years. One of the earliest papers dealing with low loss high dielectric materials is that of Richtmyer's (1939). In his paper he shows that a high dielectric material ($\epsilon' >> 1$) in free space resonates in an electromagnetic mode. The mode amplitude decreases with time because of dielectric and radiation losses. Most dielectric materials, whose properties are found in tables or graphs (except in recent years), have small dielectric constants and/or relatively large dielectric losses. In this sense they are not suitable as electromagnetic resonant structures. Renewed interest in high dielectrics is in part due to the present availability of materials with relatively low dielectric losses, and, with dielectric constants two and three orders of magnitude above the free space permittivity. Radiation losses can be essentially eliminated by electromagnetic shielding.

More recently, in 1960, Ohaya[2] reported X band Q's $\sim 10^3 - 10^4$ and $\epsilon' \sim 10^2$, for single crystal rutile at 300°K and 4.2°K. Bell and Rupprecht[3], in 1961, reported similar Q's and $\epsilon' \sim 10^2 - 10^3$ for strontium titanate crystals at K band. The list of such materials seems to be growing rapidly.

The basic electromagnetic mode configurations for different shaped dielectric resonators are discussed in various sources. Richtmyer[1] investigated some of the resonant characteristics of toroidal, spherical, and ring shaped samples; Okaya and Barach[4] investigated the resonant modes of an anisotropic rectangular resonator; and, Yee[5] calculated the exact modes for a sphere, and the approximate ones for circular cylindrical and rectangular resonators.

Here, the $TE_{10\delta}$ mode - for rectangular and cylindrical resonators, - and the H_{110} dominant mode - for spherical resonators - are of primary interest. The $TE_{10\delta}$ mode is dominant, (lowest frequency) for the cylindrical case if its axial length is $\leq$ its diameter; and, for the rectangular case if its height is < both its cross sectional dimensions. Qualitatively, H_{110} and $TE_{10\delta}$ modes are quite similar. Field configurations of both modes resemble those of a magnetic dipole. (See Figures 5a and 13).

When discussing dielectric resonators we have in mind one that is shielded against radiation losses by conducting walls. The walls are good conductors and placed sufficiently far from resonator surfaces so that mode configurations are determined, essentially, by the resonator alone. In practice, wall losses and wall tuning effects are observed and should be important in most device applications. An investigation of this

phenomenon is being carried out by Cohn et al[6]. In this paper we have neglected wall effects.

In Section I, frequency tuning of dielectric resonators is considered. Of the presently available tuning techniques the one discussed there has been found the most useful - at least until now. In Section II, a method of heat sinking dielectric resonators for CW operation is presented. Sections III and IV describe applications of the resonant characteristics of high dielectric systems in microwave filters and as a possible means of rain drop measurements.

SECTION I

FREQUENCY TUNING OF DIELECTRIC RESONATORS

Introduction

There are at least three methods by which dielectric resonators may be tuned. One method involves the variation of relative dielectric constant, another the alteration of resonator dimensions, and the third the perturbation of electromagnetic fields of the resonator. The dielectric constant may be varied by changing the operating temperature, or for some materials, by the application of an electric dc biasing field. The dependence of resonant frequency on dielectric constant and resonator dimensions has been reported elsewhere[7].

In this paper frequency tuning employing field perturbing techniques is considered. Such a technique, using two closely spaced dielectric resonators, was originally suggested by Shaw[8] and experimentally verified by Yee[8]. They also investigated frequency tuning using more than two resonators.

In general, the techniques of tuning resonant structures by perturbing electromagnetic fields of a resonator are well established. In particular, frequency tuning by introducing small perturbations can be handled analytically and is more or less well understood. For larger perturbations the inclusion of higher order correction factors presents analytical difficulties. For still larger perturbations, say when perturbing energies are on the order of system energies, the analysis may (for simple geometries) simplify once again. It is generally the intermediate case which presents the most difficulty. Some insight into this intermediate range has been made by studying idealized electrical analogues of coupled resonators and relating some of their general features to the more complicated coupled dielectric resonators. We discuss this further after presentation of some experimental results.

Experiment

A block diagram of the experimental arrangement is shown in Figure 1. The BWO supplies a sweep signal which passes through a 10 db coupler to a

circulator and on through to a waveguide section. The waveguide is terminated with an adjustable short. The generator can sweep over a frequency range from 6.5 to 11.5 GC or over any selected portion of this range. The power tapped off at the directional coupler passes through a shunt type wavemeter to a crystal detector. The detected signal is fed to the input of Channel No. 1 of the dual trace oscilloscope. The return signal, reflected from the sliding short, passes through the circulator to a second crystal detector. This detected signal is fed to the input of Channel No. 2 of the oscilloscope. The oscilloscope horizontal sweep is calibrated in units of megacycles per centimeter. By adjusting the sweep range and the coupling of electromagnetic energy into the dielectric crystals, a sharp absorption dip was observed. The sharpness of the dip corresponds to the loaded Q of the resonator. The dip was identified as absorption in the dominant ($TE_{10\delta}$) resonator mode. This dip was aligned with the dip produced by the precision wavemeter, on the other channel of the oscilloscope, and the frequency was recorded.

The waveguide assembly housing the crystals and sliding short are shown in Figure 2. A dial rotation of 5 degrees results in a lateral movement of 0.000434 inches. The Teflon rods with the rutile crystals are inserted into the waveguide assembly and the two crystals brought into contact in the center of the waveguide. When the two rutile crystals are in contact with each other they form essentially a single resonator whose volume is the sum of the individual ones. Tuning was accomplished by separating the resonators in increments of .434 mils and measuring the new resonant frequency after readjusting the short for maximum coupling. This was achieved by moving the short until a maximum dip was observed on the oscilloscope screen. Maximum coupling occurs when the sliding short is $\frac{\lambda_g}{2}$ from the resonator axis.

The samples* used in these experiments are cut from the same boule. Orientation is such that disk axes coincide with the C crystallographic axis.

Figures 3 and 4 show some experimental results of tuning rutile resonators of various shapes and sizes. In all cases the resonator to the left (facing the paper) is tuning the one to the right. Asymptotic frequencies, for large separation, are characteristic of the resonator to the right. Table I summarizes the percent tuning range, mode of operation, and separation sensitivity (over the linear portion near zero separation) for the four cases.

* Rutile and Strontium Titanate samples were supplied by C. Sahagian, Crystal Physics Branch, AFCRL.

TABLE I

Tuning characteristics of the pairs of resonators of Figures 3 and 4. Tuning ranges refer to the complete curves. Separation sensitivity refers to the linear portion of each curve near zero separation.

Curve	Tuning range (percent of mean frequency)	Separation Sensitivity MH_Z per mil	Mode of Tuned Resonator
A	22	35	Dominant TE_{106}
B	10	18	Higher Order
C	10	162	Dominant TE_{106}
D	8	12	Higher Order

Electrical Analogue

The resonant frequencies of a pair of coupled electrical resonators, as shown in Figure 5d with $R_1 = R_2 = 0$, are found to be as follows:

$$\omega_{\pm}^2 = \frac{\frac{1}{2}(\omega_2^2 + \omega_1^2) \pm \sqrt{\frac{1}{4}(\omega_2^2 - \omega_1^2)^2 + K^2 \omega_1^2 \omega_2^2}}{(1 - K^2)} \qquad (1)$$

where $\omega_n \equiv (L_n C_n)^{-\frac{1}{2}}$ are the individual resonances

and $K \equiv + \frac{M}{\sqrt{L_1 L_2}}$ is the coupling coefficient. If the uncoupled resonant frequencies ω_1 and ω_2 are equal, Equation (1) reduces to Equation (2).

$$\omega_{\pm} = \frac{\omega_\infty}{\sqrt{1 \mp K}} \qquad (2)$$

where $\omega_1 = \omega_2 \equiv \omega_\infty$

In addition, if the total energy

$$U = \left[U_{L_1}\right] + \left[U_{C_1}\right] + \left[U_{L_2}\right] + \left[U_{C_2}\right] + \left[U_{L_{12}}\right] \qquad (3)$$

in the lossless circuit is kept constant as the mutual inductance is varied, we find that K satisfies the following relation. (Here, the brackets indicate time average values.)

$$K = \left(\frac{U}{2\left[U_M\right]} - 1 \right)^{-1} \qquad (4)$$

where $\left[U_M\right] \equiv \left[U_{L_{12}}\right]$

$\left[U_M\right]$ is the average energy stored in the mutual inductance. It lies in the range $0 \leq \left[U_M\right] \leq U/4$ such that $0 \leq K \leq 1$, and is equal to the capacitive electric energy minus the inductive energy of L_1 and L_2. Upon combining Equations (3) and (4) we obtain,

$$\omega_{\pm}^2 = \frac{\omega_{\infty}^2 \left(1 - \frac{2\left[U_M\right]}{U} \right)}{1 - \frac{2\left[U_M\right]}{U}(1 \pm 1)} \qquad (5)$$

Equation† (5) has been derived without reference to geometry. It is valid for any two coupled resonators provided the total stored energy remains constant as mutual coupling is varied. Differences between resonators, other than frequency, are absorbed or accounted for in the dependence of mutual energy on geometry, orientation angle, and separation distance. (For example, $\left[U_M\right]$ may depend on separation distance S and orientation angle θ_2 (Figure 5) according to $\left[U_M\right] = e^{-\alpha(S + S_o)} \sin^2 (\theta_2 + \beta_o)$ where parameters S_o, α, and β_o are constant for a particular set of coupled resonators.)

Consider next two almost identical dielectric resonators operating under the following conditions.

a. Both are driven in their dominant $TE_{10\delta}$ mode at frequency ω.

b. They are magnetically coupled.

c. Their independent uncoupled resonant frequencies are identical.

d. They are separated and oriented as in Figure 5a.

e. The total stored energy U is constant and given by

$$U = \left[U_{H_1}\right] + \left[U_{E_1}\right] + \left[U_{H_2}\right] + \left[U_{E_2}\right] + \left[U_{H_{12}}\right] .$$

Each energy term has its analogue in the corresponding circuit energy expression Equation (3).

The pair of resonators shown in Figure 5a satisfy conditions a through e above, and, we expect Equations (4) and (5) to apply to this pair of resonators.

If the above analogy is correct, the maximum possible tuning range is determined by the condition $|K| \leq 1$, or $\left[U_M\right] \leq U/4$. This requires the resonant frequencies to be in the following ranges.

†It is interesting to compare the expression for ω_- with Ginzton's[9] perturbation equation for frequency.

$$\omega_\infty \leq \omega_+ < \infty \tag{6}$$

$$\frac{\omega_\infty}{\sqrt{2}} \leq \omega_- \leq \omega_\infty$$

In Figures 3 and 4 only the ω_- modes have been plotted. According to Equation (6) the maximum possible percent tuning range, for the ω_- mode, is 30%. Both modes, whose field configurations are indicated in Figure 5c, are observed experimentally except when the resonators are close together; then the ω_+ mode "moves" outside the range of our measuring equipment and in addition the ω_+ mode is difficult to excite.

An attempt was made to find a relationship between frequency (or equivalently $[U_M]$) and separation distance, which agrees with experiment - for the case $\theta_1 = \theta_2 = 0$ and $\theta_1 = \pi/2$, $\theta_2 = 0$. (By proper choice of parameters Equation (9) can be made to fit, possibly fortuitously, some of the experimental curves.) We are guided by the following observations:

a. As $S \to \infty$ the resonant frequency asymptotically approaches the natural frequency of one of the resonators.

b. As $S \to 0$ the resonant frequency is linearly related to S. (This relation follows from a perturbation analysis.)

c. The fields outside of a single resonator decrease exponentially with distance.

A relationship which shows some promise is the following one.

$$\left[U_m\right] = \frac{U}{4} e^{-\alpha S} \tag{7}$$

where $S = x + L$
Distance x is the physical separation distance (= 0 when the resonators are in contact) and L is the sum of the half heights of the two resonators. If Equation (7) is substituted into Equations (4) and (5) they become

$$K = (2 e^{\alpha S} - 1)^{-1} \tag{8}$$

$$\omega_\pm^2 = \frac{\omega_\infty^2 \left(1 - \frac{e^{-\alpha S}}{2}\right)}{1 - \frac{e^{-\alpha S}}{2}(1 \pm 1)} \tag{9}$$

Equations (8) and (9) are sketched in Figure 6.

Equation (9) predicts that for the ω_- mode a plot of $\ell n\ (\omega_\infty^2 - \omega_-^2)$ versus x should produce a straight line. Results of such plots for Curve A (Figure 3) and for Yee's[10] curve are nearly straight lines over most of the tuning range. (Both curves are for idential resonators.) Deviations from straight lines occur near the asymptotic frequencies. Straight lines are not obtained for Curves B, C, D (Figures 3 and 4) in which the resonator frequencies are different and/or higher order modes are involved. We note finally, that for weak coupling ($\alpha S \to \infty$) Equation (8) predicts $K \sim e^{-\alpha S}$. This same exponential dependence for K was obtained by Cohn[11] for the case in which resonators are placed in a waveguide operated below cutoff.

In this section the general technique of tuning dielectric resonators by perturbing external fields has been discussed. Some experimental results of the resonant frequencies of coupled resonators were presented and an intuitive understanding of some of the results was obtained with the aid of an electrical equivalent circuit. Further study and the acquisition of more experimental data are in progress so as to clarify and extend the preceeding results.

SECTION II

TEMPERATURE STABILIZATION OF DIELECTRIC RESONATORS

Continuous wave (CW) operation of dielectric resonators has some severe limitations unless some form of temperature compensation is employed. The reason for this is the strong dependence of permittivity (dielectric constant) on temperature. Heating of the sample arises from losses in the dielectric. In rutile ($Ti\ O_2$), the relative change of the dielectric constant is 800 parts per million per degree centigrade (800 ppm per degr. C). Since the resonant frequency is proportional to $\frac{1}{\sqrt{\epsilon'}}$, the relative frequency change is 400 ppm per degree C. Higher dielectric constant materials, such as strontium titanate ($\epsilon' = 240$) and barium titanate ($\epsilon' > 1000$) have greater temperature sensitivities.

A heat-sinking technique has been developed[12] which results in temperature stabilization at power levels exceeding 100 milli-watts of CW power. Boron nitride (BN) is used as a heat conducting path from the dielectric to the waveguide walls which act as a heat sink. See Figure 7. The room temperature heat conductivity of BN is 250 $\frac{(BTU/hr)\ ft}{ft^2\ F^o}$, and its dielectric constant (~ 4.4) is small compared to that of rutile (~ 75).

Figure 8 shows frequency shift versus input power for a $Ti\ O_2$ resonator of dimensions 116 x 122 x 122 mils for the three following conditions.

Curve 1. Complete system as shown in Figure 7.

Curve 2. Complete system with pressed BN slabs removed.

Curve 3. Complete system with pressed BN slabs and powder removed.

Figure 9 shows the measured resonant frequency versus time for two cases. Curve A corresponds to the conditions for Curve 2 of Figure 8, while Curve B corresponds to the conditions for Curve 1 of Figure 8. (In both cases the incident power level was 10 mw). It was not possible to obtain a similar curve for the dielectric alone because the frequency drifted too rapidly to measure. After 12 minutes the heat-sinked resonator reached thermal equilibrium. The resonator in BN powder alone did not stabilize even after one hour of operation. The difference in resonant frequency between curve A and B is attributed to the frequency-tuning effect of the BN slab.

The heat sinking technique described here has been used in some of our experiments. The system is relatively easy to assemble, BN is

readily available and inexpensive, and no serious deterioration of resonator characteristics has been noted.

SECTION III

BAND PASS FILTER

In this section we describe a microwave band pass filter which uses the resonant characteristics of high dielectric materials. The filter has been tested experimentally and its performance is in agreement with theoretical predictions based on the analysis of an electrical equivalent circuit.

Figure 10 shows the complete band pass filter. A cylindrical strontium titanate resonator (dimensions 2.54 x .635 mm) is inserted into an inductive waveguide iris. A TE_{10} waveguide mode, incident from the left, excites the $TE_{10\delta}$ magnetic dipole mode of the resonator. Part of the incident signal is dissipated as heat in the resonator and the remaining is reradiated into the guide to the right of the iris.

An electrical equivalent circuit for this filter is shown in Figure 11. Inductor, L, represents the iris inductance in the absence of the resonator. Mutual inductance, M, represents the number of TE_{10} magnetic field lines passing through the resonator. The unloaded resonator Q_o is $\frac{\omega_o L_o}{R_o}$ and the resonant frequency is $\omega_o = (L_o C_o)^{-\frac{1}{2}}$.

A straightforward analysis using circuit and transmission line theory gives the following expressions for voltage transmission (τ) and reflection (ρ) coefficients.

$$\tau = \frac{-\Omega K^2}{(1 + j\Omega)\left[\Omega K^2 + (1 + j\Omega)\left(\frac{1}{Q} + j\left[1 - \frac{\omega_o^2}{\omega^2}\right]\right)\right]} \qquad (10)$$

$$\rho + \tau = \frac{j\Omega - 1}{j\Omega + 1} \qquad (11)$$

where $\Omega \equiv \frac{\omega L}{Z_g}$

$$K \equiv \frac{2M^2}{L L_o}$$

$$Y_g^{-1} = Z_g = \frac{\eta}{\sqrt{1 - \frac{\omega_c^2}{\omega^2}}}$$

$$Q = \frac{\omega L_o}{R_o}$$

and where ω_c is the guide cutoff frequency, η the free space wave impedance. Equations (10) and (11) have been programmed on a computer to check the filter performance for various values of resonator Q's and iris widths. Preliminary computer results show good agreement between theory and experiment. Complete results will be published later as an AFCRL report.

Partial conditions for optimizing filter performance may be deduced analytically for the following case - which corresponds to actual experimental conditions.

a. $f_o \approx 9\ GH_Z$

b. $Q_o \gtrsim 500$

c. frequency range $f_o \pm 5$ times filter bandwidth

d. $\frac{\delta}{a} = \frac{\text{iris width}}{\text{guide width}} \leq .1\ (\Omega << 1)$

Under these conditions the maximum transmission coefficient and the 3db bandwidth are given by the following expressions.

$$|\tau_o| \cong \frac{\Omega_o Q_o K^2}{\Omega_o Q_o K^2 + 1} \tag{12}$$

$$\text{B.W.} \cong \frac{\omega_o}{Q_o} (\Omega_o Q_o K^2 + 1) \tag{13}$$

Here, the zero subscript means the quantity is to be evaluated at the resonant frequency ω_o.

The gain bandwidth product from Equations (12) and (13) is as follows,

$$(\text{Gain Bandwidth Product})_{\text{small iris}} = \omega_o \Omega_o K^2 \tag{14}$$

The quantity Ω_o is important because it is directly related to iris width. If Ω_o is known, the iris width may be found and vice versa.

We find for the gain bandwidth product, under conditions a, b, and c above, but with a large iris ($\Omega >> 1 => \frac{\delta}{a} \to 1$), the following expression.

$$(\text{Gain Bandwidth Product})_{\text{large iris}} = \frac{\omega_o K^2}{\Omega_o (1 - K^2)} \tag{15}$$

Here Ω_o appears in the denominator whereas in Equation (14) Ω_o is in the numerator. Hence, an important result is that <u>an optimum value of Ω_o exists such that the gain bandwidth product is maximum</u>.

There are two special cases involving equations (12) and (13) which are of some practical interest. For high Q resonators, $Q_o >> (\Omega_o K^2)^{-1}$, the transmission coefficient is approximately unity and the bandwidth $\omega_o \Omega_o K^2$. For low Q resonators, $Q_o << (\Omega_o K^2)^{-1}$, the transmission coefficient is approximately $\Omega_o Q_o K^2$ and the bandwidth $\frac{\omega_o}{Q_o}$ which is, as expected, just the bandwidth of the resonator alone. The preceding results

should be useful for design purposes.

A photograph of the frequency response of the filter is shown in Figure 12. Data related to the photo are given in Table II.

TABLE II

Data for Strontium Titanate Band Pass Filter
for two values of iris width.

Iris Width	2.8 mm	2.4 mm
Disc diameter	2.54 mm	2.54 mm
Disc thickness	0.635 mm	0.635 mm
Insertion loss at f_o	2.5 db	4.5 db
f_o	8.672 GH_Z	8.672 GH_Z
B.W. (3 db)	30 MH_Z	20 MH_Z
Rejection	> 40 db	> 60 db
Dielectric Constant	240	240

Experimental results with these filters have exceeded initial expectations and have, in general, been very encouraging. Performance characteristics are comparable to those of other, more complicated, microwave filters. One of the problems encountered in initial experiments with rutile filters was the generation of spurious modes. No such problems are encounted with strontium titanate filters. Subsequent experiments at 77 degrees K have reduced insertion losses to 0.5 db. Investigation of this filter is still in progress.

SECTION IV

WATER DROP DIAMETER MEASUREMENT WITH A SWEPT FREQUENCY MICROWAVE SIGNAL[13]

When a plane polarized wave impinges on a water droplet electric and magnetic dipoles are induced within the drop[14]. Energy is taken from the incident wave, part of it is absorbed as heat and part is reradiated at the frequency of the incident wave. If we view a water-drop as a high dielectric resonator we may draw some analogies between high permittivity crystalline resonators, such as rutile and strontium titanate, and water-drop resonators. Water has a high dielectric constant (over some temperature and frequency ranges - Table III)[15], however, its dielectric losses are large compared to those of the aforementioned crystals. It is, therefore, not possible to achieve the high loaded Qs of 3000 or more, at X-band frequencies, as it is with crystals. Still, loaded Qs of 200 at 3 cm wavelength are sufficiently sharp to allow measurement of the resonant

frequency with better than 1% accuracy.

The resonant frequency of a homogeneous, isotropic, dielectric resonator depends on its size, geometry and relative dielectric constant. The boundary value problem of such a dielectric resonator of spherical shape has been solved by H. Y. Yee[5]. The lowest resonant mode is found to be the H_{110} mode, shown in Figure 13, with resonant frequency approximately given by Equation (16).

$$f = \frac{C}{2\, a \sqrt{\epsilon'}} \qquad (16)$$

Here $C = 3 \times 10^{10}$ cm/sec and a is the radius of the sphere.

If an electromagnetic wave is incident on a water drop and excites the H_{110} mode, then, the size of the drop may be calculated using Equation (16). The dielectric constant is known from Table III and the frequency at which an absorption peak occurs can be measured. Experimental evidence in support of the preceding discussion is given below.

The experimental setup is similar to the one depicted in Figure 1 but with the rutile disc resonators replaced by water drop resonators mounted on a polystyrene slide. The slide contains machined spherical cavities of 0.185 and 0.190 cm radii. The cavities are filled with distilled water through a small hole in the slide. The incident sweep frequency microwave signal passes through the circulator into the waveguide section housing the polystyrene slide with the water-drops. The reflected signal is fed to one channel of the dual trace oscilloscope. An adjustable waveguide short is tuned to couple the maximum power into the water-drop (critical coupling). The wavemeter measures the resonant frequency of the sample when the absorption peak of the wavemeter on the second channel of the CRO coincides with the absorption peak of the sample.

The spheres of radii 0.185 and 0.190 cm showed resonances at 9.944 GH_Z and 9.848 GH_Z respectively. Theoretical frequency values according to Equation (16) are 9.968 GH_Z and 9.705 GH_Z respectively. The difference frequencies between resonators are, 267 MH_Z theoretical, and 96 MH_Z experimental. A large portion of this discrepancy may be attributed to dielectric losses which were not taken into account in deriving Equation (16). For practical applications of this method, a correction factor may have to be determined experimentally.

In this section it has been shown that a water drop enclosed in a waveguide behaves as a dielectric resonator at X band frequencies. Besides water drop measurements which may be of interest to the meteorologist, another possible application may be fog dispersal by high powered microwave sources. In any application, however, where the resonator is not enclosed by metal, radiation losses will play a significant role and their resulting effects will have to be determined.

SECTION IV

CONCLUSION

A physical understanding of the resonant frequency and mode configuration behavior for arbitrarily coupled magnetic dipole modes associated with high dielectric resonators has been achieved. It has been found that under certain conditions, a fundamental tuning range limit of 30% may exist. A heat sinking technique has been developed which allows for CW operation of high dielectric resonators. A novel, simple, efficient, microwave band pass filter employing high dielectric materials has also been developed. Finally, a possible technique for the measurement of rain drop size has been suggested and partially verified under laboratory conditions.

REFERENCES

1. R. D. Richtmyer, "Dielectric resonator", J. Appl. Phys., Vol. 10, p.391, (June 1939)
2. A. Okaya, "The rutile microwave resonator", Proc. IRE, vol. 48, p1921 (1960)
3. R. O. Bell and G. Rupprecht, "Measurement of small dielectric losses in material with a large dielectric constant at microwave frequencies", IRE Trans. on Microwave Theory and Techniques, vol. 9, pp. 239-242, (May 1961)
4. A. Okaya and L. F. Barash, "The dielectric microwave resonator", Prcc. IRE, vol. 50, p. 2081 (Oct 1962)
5. H-Y. Yee, "An investigation of microwave dielectric resonators", Hansen Labs., Stanford Univ., Calif., M. L. Rept. 1065, (July 1963)
6. S. B. Cohn and E. N. Torgov, "Investigation of Microwave Dielectric-Resonant Filters", Rantec Corp., Report No. 6, Contract # DA-36-039-AMC-02267 (E), (1 Dec 64 to 28 Feb 65)
7. J. C. Sethares and S. J. Naumann, "Design of Microwave Dielectric Resonators", IEEE Trans. on MTT vol. 14 (Jan 66) 2.
8. H. J. Shaw, W.W.Hansen Labs, Stanford Univ., "Research on Selected Microwave Properties of Ferrites", Contract # AF19(628)-342,(1963)
9. Ginzton, E. L., Microwave Measurements, McGraw-Hill, (1957) 439
10. W.W.Hansen Labs, Stanford Univ., "Research on Selected Properties of Ferrities", M.L.Report 1150, April 1964, Contract # AF19(628)-342,p.5
11. S.B.Cohn, and K.C.Kelley, "Microwave Dielectric-Resonator Filters", Rantec Corp., Report # 2, Contract # DA-36-039-AMC-02267(E), (1 Oct 63 - 31 Dec 63) p 32
12. M. Stiglitz and J. C. Sethares, "Frequency stability in dielectric resonators", Proc. IEEE, 53 p 311
13. M. Stiglitz, "Swept Frequency Measurements of Water drop Diameter", to be published in Proc. IEEE.
14. L. J. Battan, Radar Meteorology, The University of Chicago Press, (1960) 26 - 29.
15. C. R. Burrows, S. S. Attwood, "Radio Wave Propagation", Academic Press, Inc., New York, N. Y. 1949, p. 276 and 300

TABLE III

RELATIVE DIELECTRIC CONSTANT ϵ' OF WATER AT DIFFERENT TEMPERATURES AND WAVELENGTHS

Temperature	ϵ'			
t° C	$\lambda =$ 0.5 Cm	$\lambda =$ 1.58 Cm	$\lambda =$ 3.2 Cm	$\lambda =$ 10 Cm
0	7.0	19.0	42	78.7
5	8.1	25.3	51.3	80.4
10	9.4	32.0	58.6	80.6
15	11.0	38.1	63.6	79.7
20	12.8	44.0	66.4	78.5
25	14.9	49.0	68.0	77.20
30	17.3	52.7	68.7	75.7
35	19.9	55.5	68.9	74.1
40	22.7	57.7	68.7	72.6

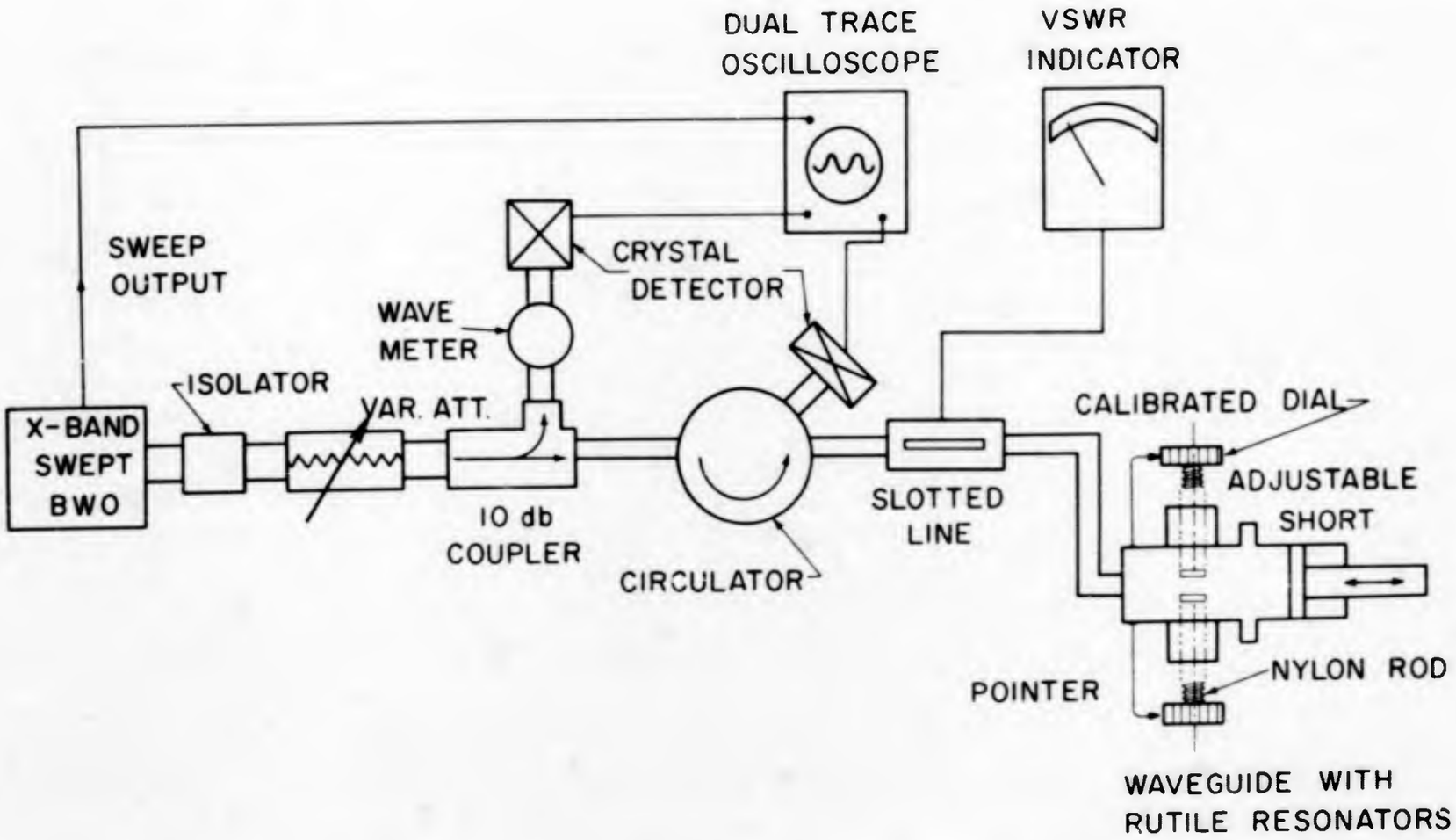


FIGURE 1 BLOCK DIAGRAM OF DIELECTRIC RESONATOR TUNING EXPERIMENTS. FOR OTHER EXPERIMENTS DESCRIBED IN THIS REPORT THE WAVEGUIDE SECTION HOUSING THE RUTILE RESONATORS WAS APPROPRIATELY MODIFIED.

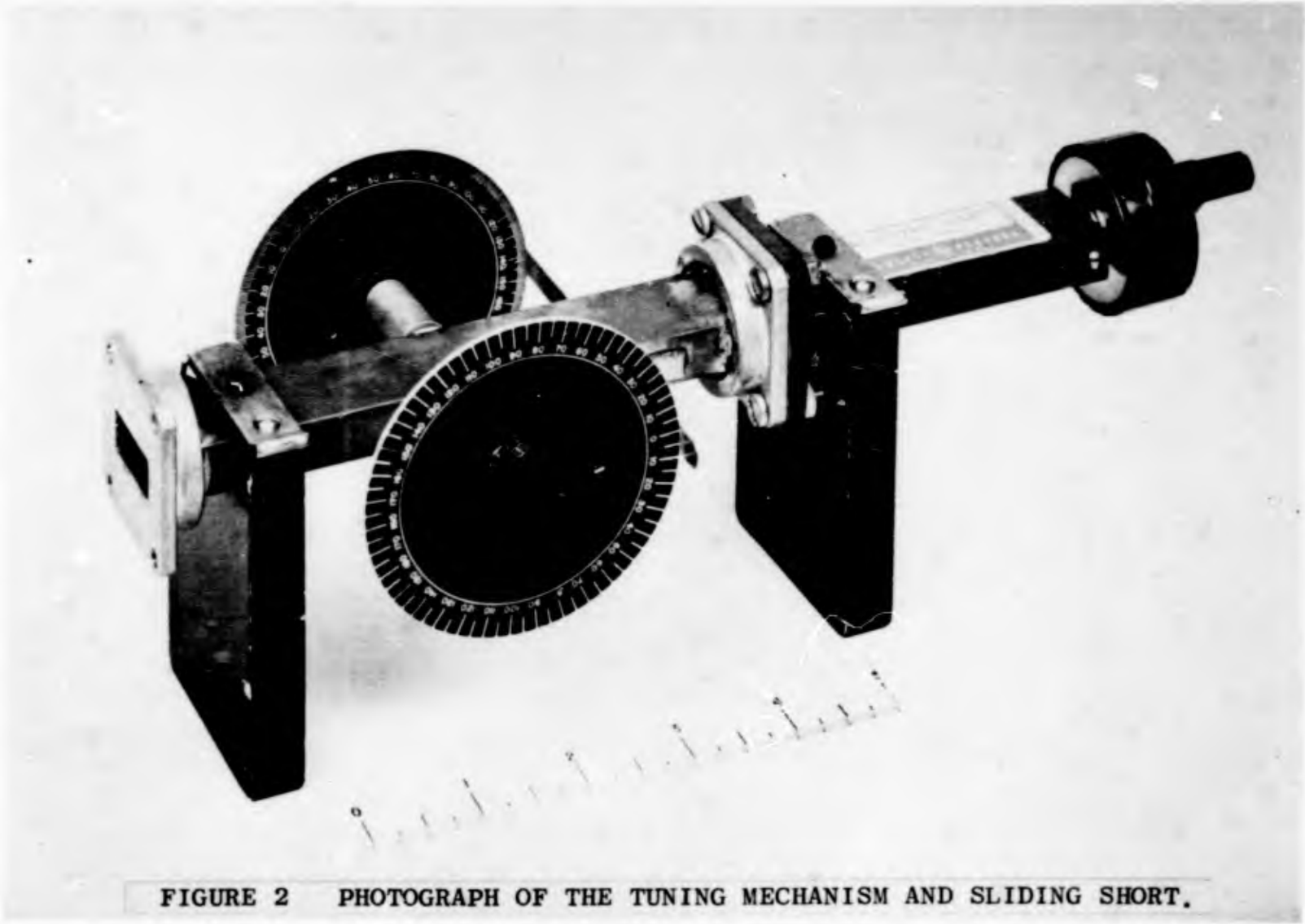


FIGURE 2 PHOTOGRAPH OF THE TUNING MECHANISM AND SLIDING SHORT.

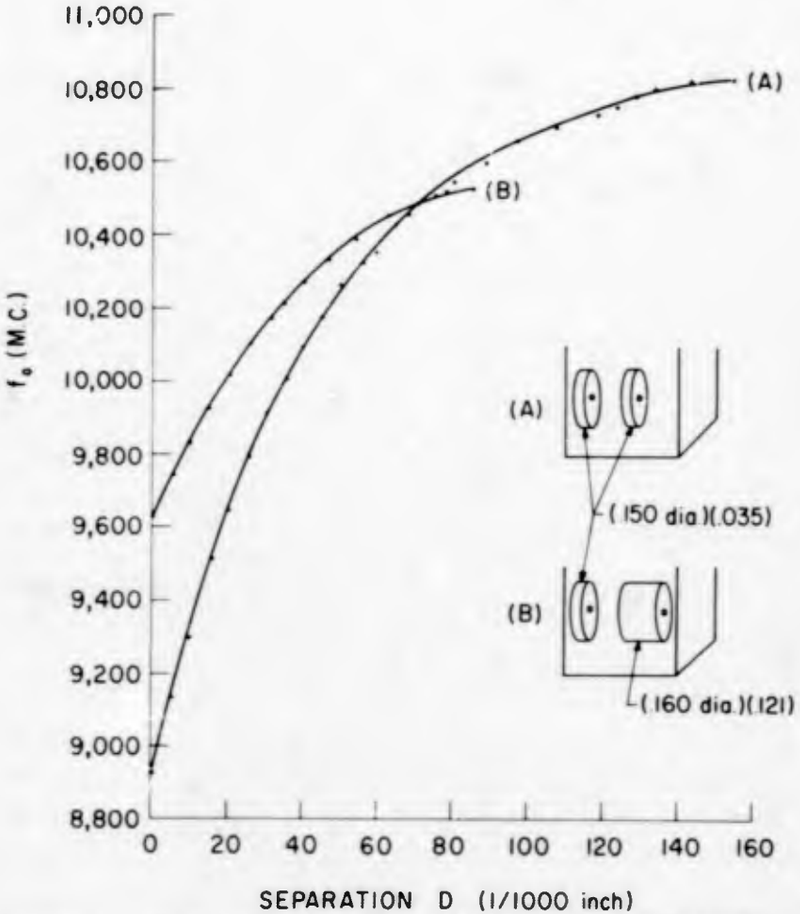


FIGURE 3 TUNING CURVES OF RUTILE RESONATORS - RESONANT FREQUENCY IN MEGACYCLES VERSUS SEPARATION DISTANCE IN MILS. RESONATOR DIMENSIONS ARE IN INCHES.

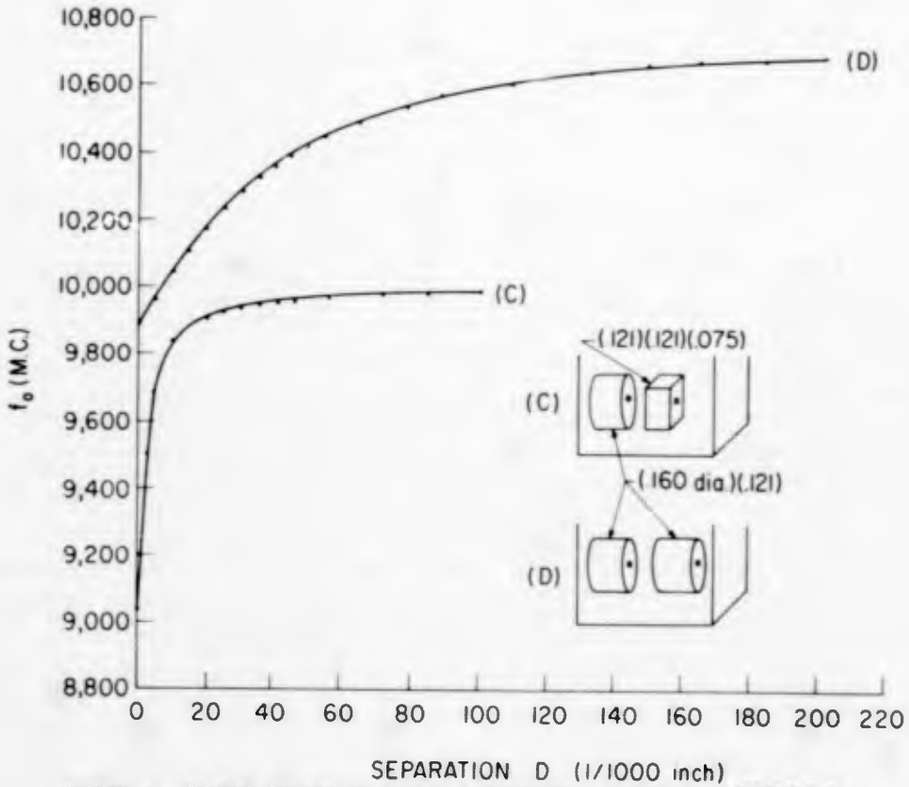


FIGURE 4 TUNING CURVES OF RUTILE RESONATORS - RESONANT FREQUENCY IN MEGACYCLES VERSUS SEPARATION DISTANCE IN MILS. RESONATOR DIMENSIONS ARE IN INCHES.

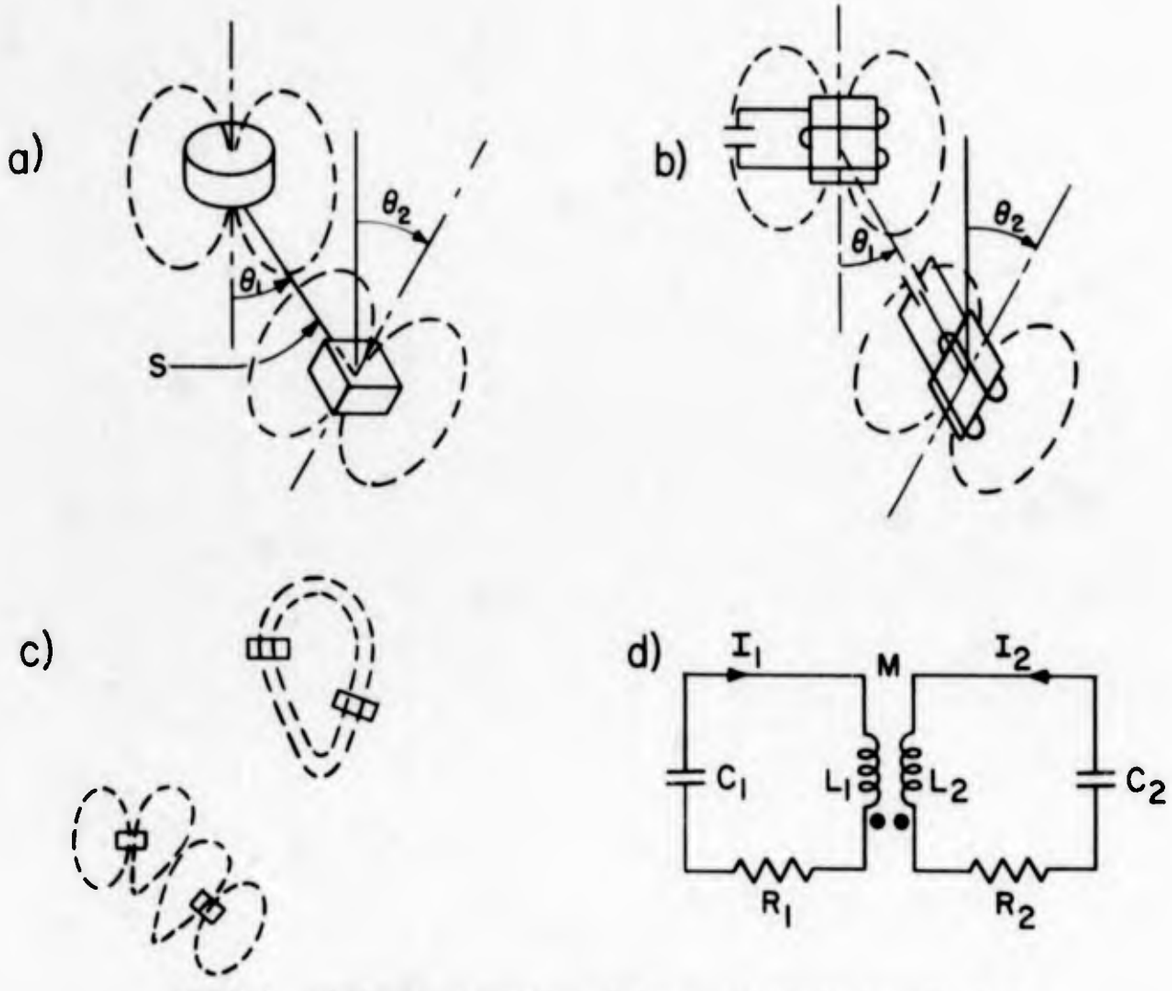


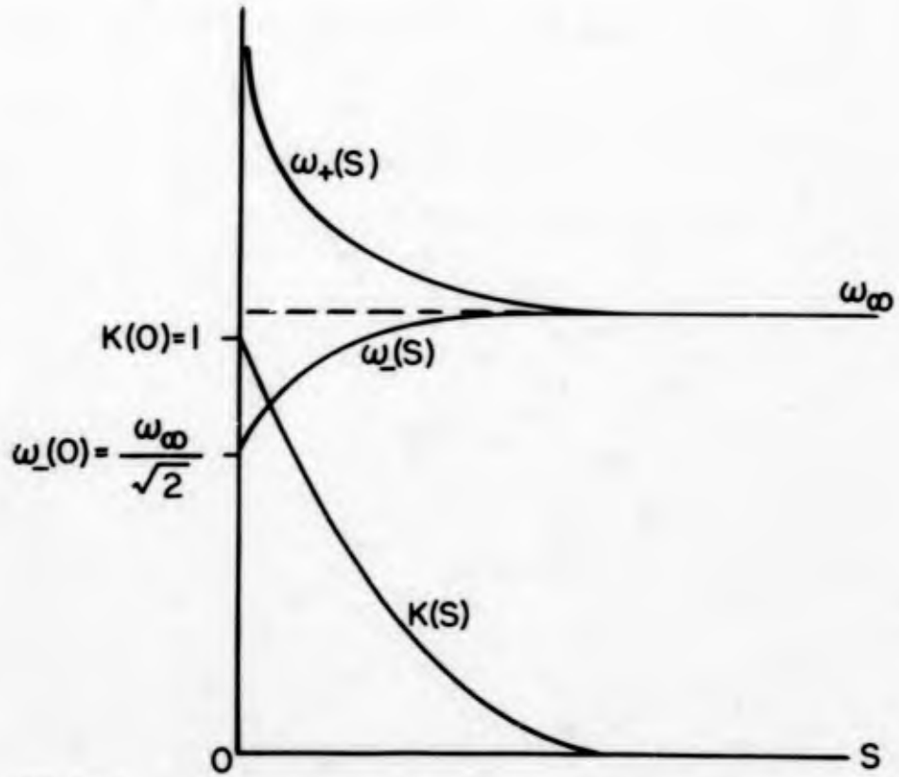
FIGURE 5 MAGNETICALLY COUPLED DIELECTRIC RESONATORS EXCITED IN TE_δ MAGNETIC DIPOLE MODES AND EQUIVALENT COUPLED ELECTRICAL RESONATORS.

A. ANGLES θ_1 AND θ_2 ARE MEASURED IN A COMMON PLANE WHICH INCLUDES THE AXES OF BOTH RESONATORS. THE ELECTRIC FLUX DENSITY IS SOLENOIDAL AND LIES ALMOST ENTIRELY WITHIN THE RESONATOR.

B. ROUGH ANALOGY TO COUPLED RESONANT CIRCUITS.

C. TWO POSSIBLE MODE CONFIGURATIONS. MODES AT THE UPPER RIGHT AND LOWER LEFT CORRESPOND TO POSITIVE AND NEGATIVE COUPLING RESPECTIVELY.

D. ELECTRICAL EQUIVALENT CIRCUIT.



AFCRL PHOTO

FIGURE 6 RESONANT FREQUENCY OF THE ω_+ AND ω_- MODES, AND COUPLING COEFFICIENT, VERSUS SEPARATION DISTANCE ACCORDING TO EQUATIONS (8) AND (9).

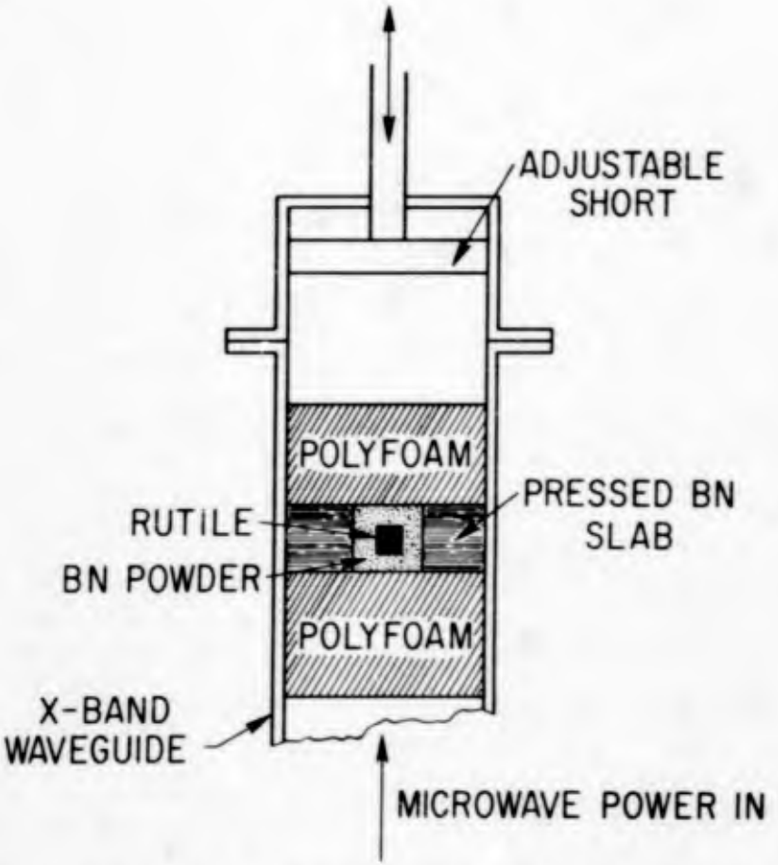


FIGURE 7 DETAILS OF RUTILE RESONATOR IMBEDDED IN BORON NITRIDE (BN) POWDER WITH PRESSED BN SLABS FOR HEAT CONDUCTION TO WAVEGUIDE WALLS.

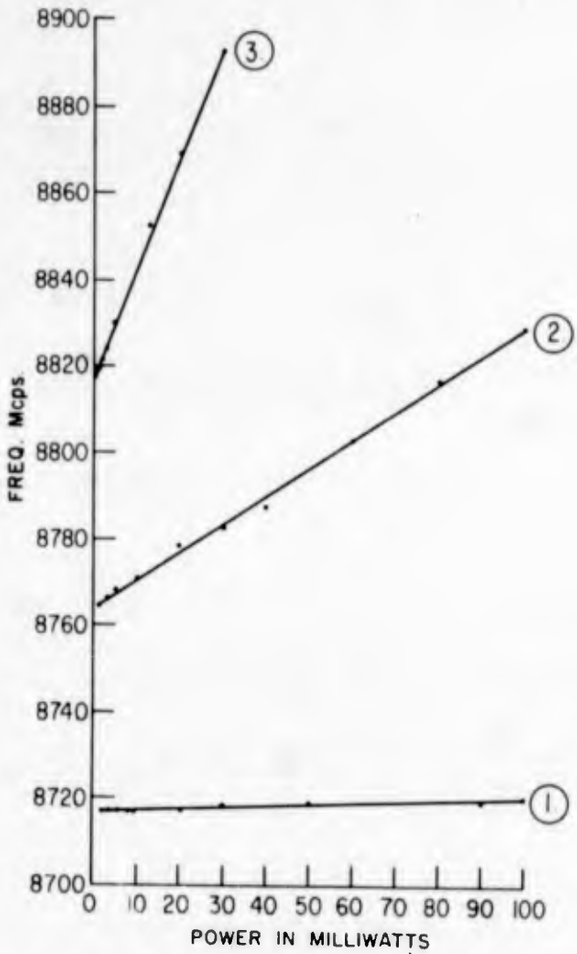


FIGURE 8 FREQUENCY SHIFT VERSUS INPUT POWER FOR THREE CONDITIONS TO DEMONSTRATE EFFECTS OF HEAT SINKING. (3) RESONATOR IN AIR, (2) RESONATOR SURROUNDED BY BN POWDER BUT WITH NO HEAT PATH TO WAVEGUIDE WALLS, (1) RESONATOR WITH COMPLETE HEAT SINKING AS SHOWN IN FIGURE 7.

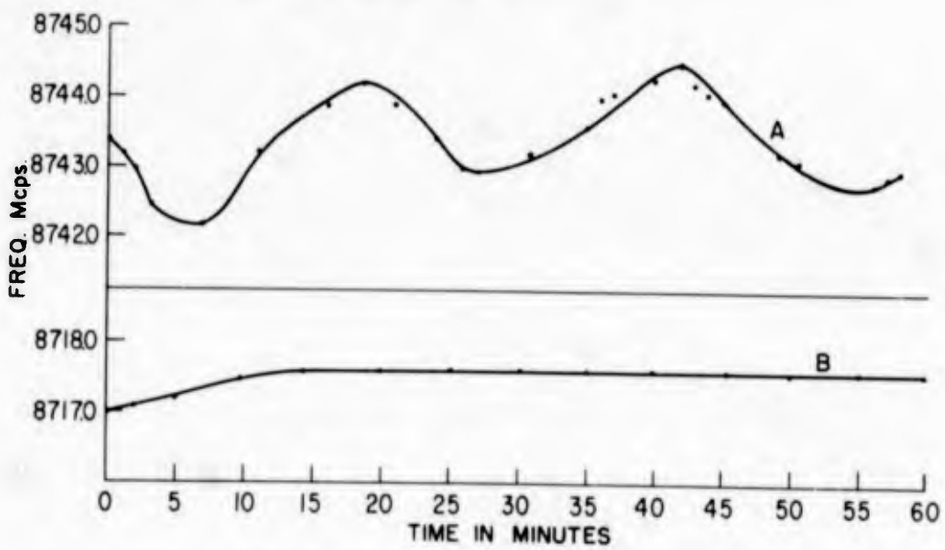


FIGURE 9 RESONANT FREQUENCY VERSUS TIME WITH 10 MW CW POWER INPUT TO RESONATOR. OPERATING CONDITIONS FOR CURVES A (BN POWDER, NO HEAT SINK) AND B (BN POWDER AND HEAT SINK) CORRESPOND TO THOSE FOR CURVES (2) AND (1) OF FIGURE 8. IT WAS NOT POSSIBLE TO OBTAIN A SIMILAR CURVE FOR A RESONATOR IN AIR.

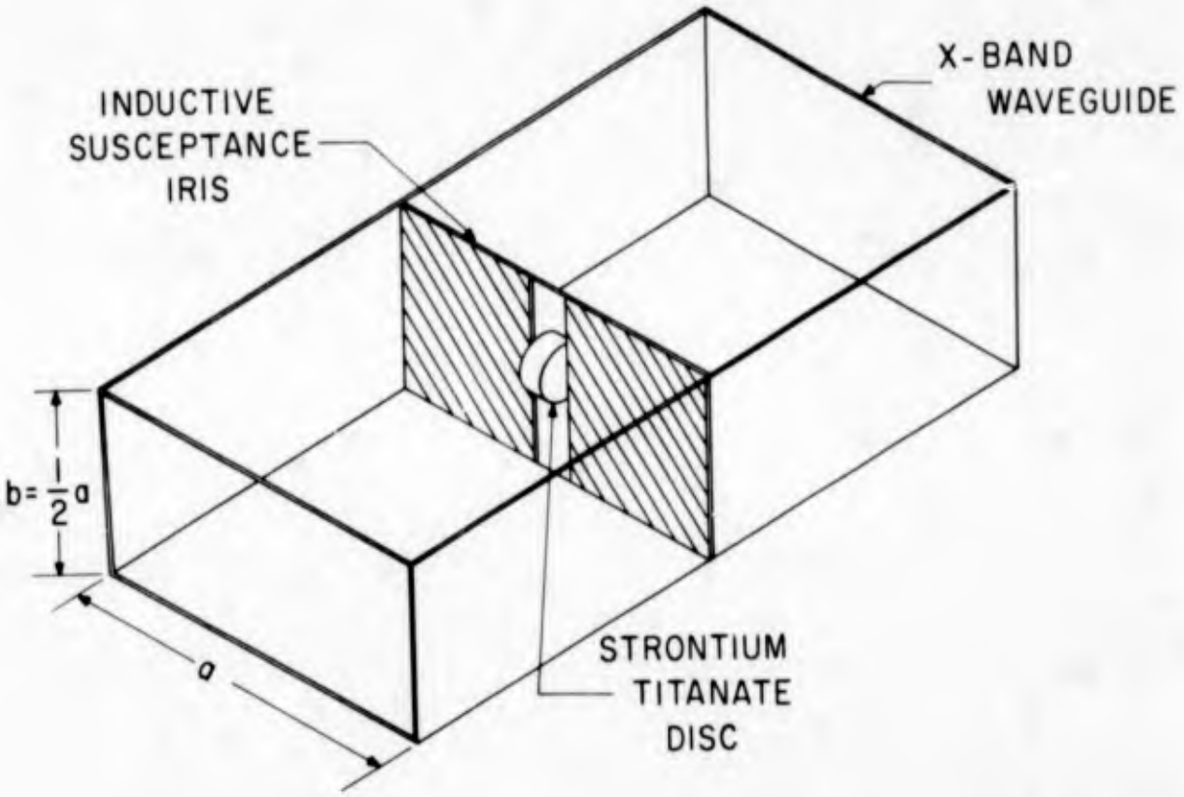


FIGURE 10 MICROWAVE BAND PASS FILTER EMPLOYING THE MAGNETIC DIPOLE MODE OF A STRONTIUM TITANATE DIELECTRIC RESONATOR.

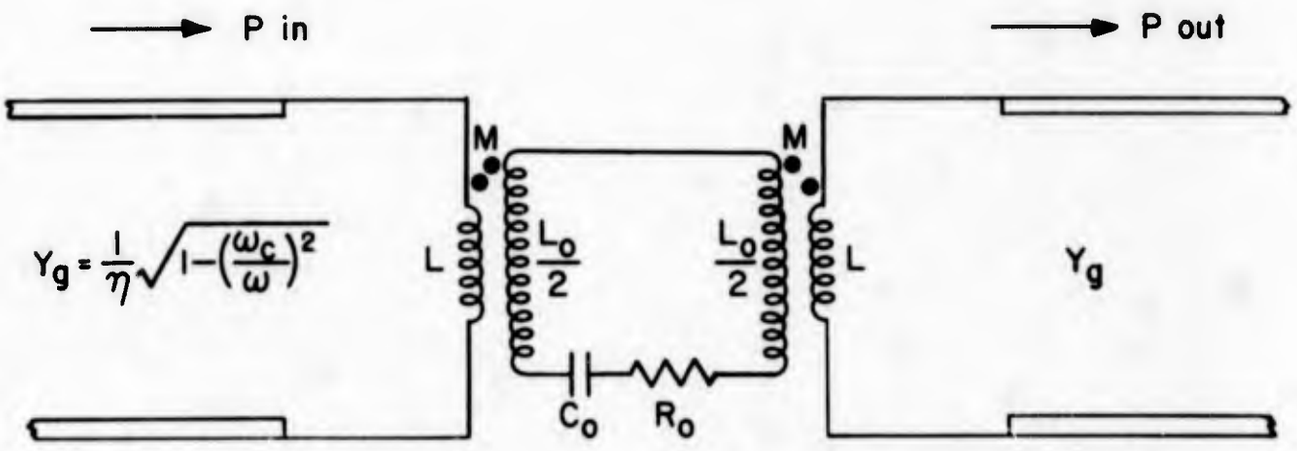


FIGURE 11 EQUIVALENT CIRCUIT FOR THE BAND PASS FILTER OF FIGURE 10. INDUCTOR L IS THE IRIS INDUCTANCE IN THE ABSENCE OF THE RESONATOR.

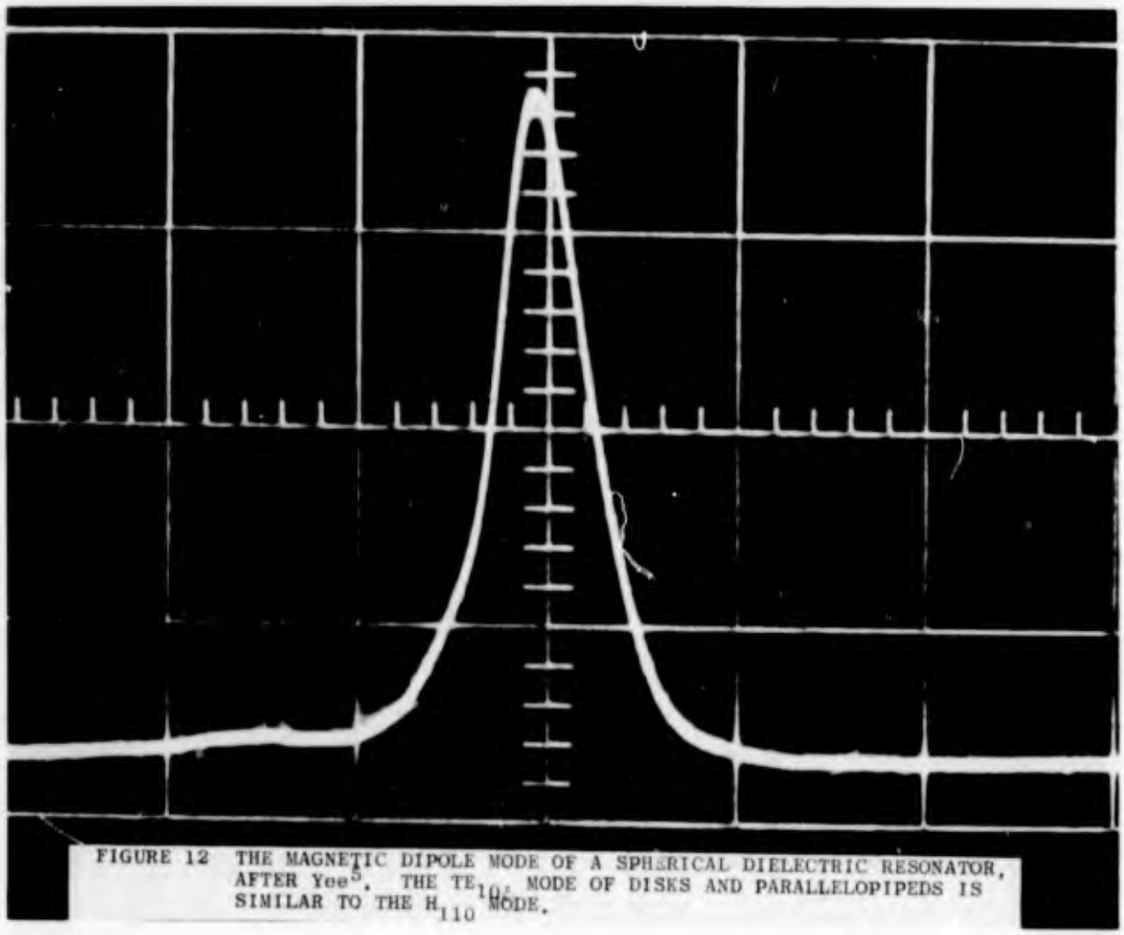


FIGURE 12 THE MAGNETIC DIPOLE MODE OF A SPHERICAL DIELECTRIC RESONATOR, AFTER Yee[5]. THE $TE_{10\delta}$ MODE OF DISKS AND PARALLELOPIPEDS IS SIMILAR TO THE H_{110} MODE.

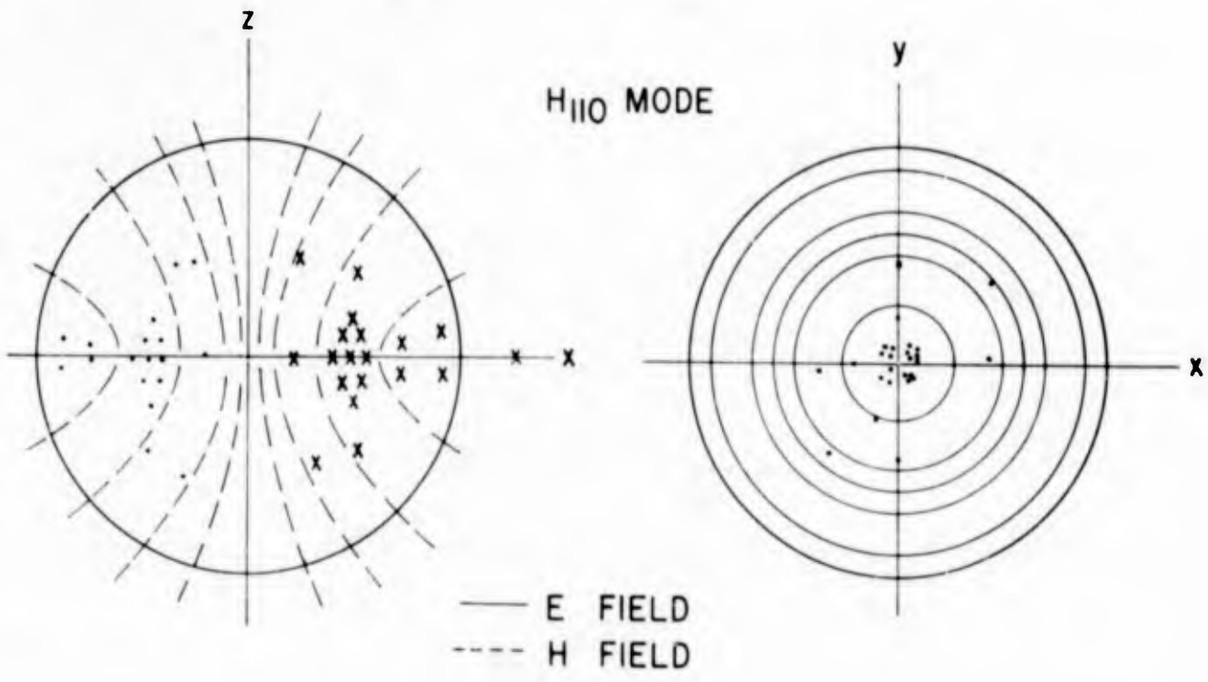


FIGURE 13

(U) SOLID STATE DISPLAY TECHNIQUES

by

Carlton J. Peterson, Captain, USAF

Air Force Flight Dynamics Laboratory
Research and Technology Division
Wright-Patterson AFB, Ohio

Capt. Carlton J. Peterson

CAPT. CARLTON J. PETERSON

Capt. Peterson received his BSEE from Washington University, St. Louis, Missouri, in 1959; and his MSEE from the Air Force Institute of Technology, Dayton, Ohio, in 1963. His six years of duty with the Air Force have been performed as a Ground Electronics Officer with the Air Defense Command SAGE System, an AC&W Radar Maintenance Officer with the Air Defense Command Pine Tree Line, a graduate student, and a Task Engineer with the Air Force Flight Dynamics Laboratory. In his present position, he is responsible for applied research and advanced technology efforts on aircraft display mechanization techniques. He served as the organizer for the Display Technology Session at the recent 18th Annual National Aerospace Electronics Conference and has authored articles on ECCM signal processing and display mechanization techniques.

BLANK PAGE

(U) SOLID STATE DISPLAY TECHNIQUES

ABSTRACT

During the past two years significant improvements have been made on the basic properties of materials used in solid state displays. The capabilities of increased electroluminescent (EL) phosphor maintenance, new high contrast filter techniques, and mass produced control circuitry arrays have greatly motivated Air Force interest in advanced display techniques.

Critical to the analysis of such displays is the understanding of electrical and optical phenomena resulting in the microscopic and macroscopic characteristics of solid state displays. Due to the extremely short persistance of EL emission, the only realistic way of generating a useful level of display intensity for cockpit applications has been to drive the EL phosphor continuously. This requires the combination of EL phosphor layers with mass produced control circuitry. Such circuitry must accept and store high speed input signals, and switch apply or restrict EL drive signals such the legible presentations are made to appear on the face of the display.

Display requirements are first determined by the pilot's basic range of perceptual sensitivities and second by operational needs. Considerable effort has been directed towards both of these areas to achieve a practical solid state display capability. The most important result of such efforts has been a breakthrough in display filtering techniques which has permitted the fabrication of displays reflecting 2 percent of all incident light and transmitting 35 percent of all emitted light. Combining this capability with that of improved high intensity zinc sulfo-selenide EL phosphors, it has been possible to fabricate and flight test solid state displays that can be clearly seen under bright ambient conditions.

SOLID STATE DISPLAY TECHNIQUES

INTRODUCTION

Air Force interest in the development of improved aircraft displays has motivated a research program aimed at the perfection of solid state display techniques. This interest stems from a realization of the advantages to be gained by the development of nonmechanical displays and the resulting display capability provided for the Air Force. Advantages of solid state over electromechanical or other electronic generation techniques can be measured in terms of digital control, improved reliability, automatic scale factor variability, minimized display weight, reduced size and power requirements, and ultimately the reduction of display cost (Figure 1). Achievement of such advantages and of the goals of the solid state display effort cannot be guaranteed, but it is possible to state with confidence that previously nonexistent control-display equipments resulting from this program will have an important place in future aircraft. This confidence is based on the fact that Air Force Flight Dynamics Laboratory (AFFDL) solid state displays have been successfully flown in T-39 and KC-135 aircraft at Wright-Patterson AFB during the past year and a half (Figure 2). Similar displays using AFFDL developed techniques are scheduled for use by NASA as guidance and navigation readouts for Apollo and LEM spacecraft. These applications are the direct result of improvements that have been made on the basic properties of materials used in solid state displays (9).

EMISSION TYPE DISPLAYS

The requirement for nonmechanical displays dictates the need for either cathodoluminescent or electroluminescent (EL) emission for the conveyance of display information. Due to numerous considerations, electroluminescence, the generation of nonthermal visual radiation by excitation of an electric field, is considered to be the most suitable phenomenon for use in solid state displays. An electric field of 35 volts per millimeter is normally required to generate light in an EL lamp; however, the actual drive voltage required to display information is typically in the region of one or two hundred volts. Since the emitted light must pass

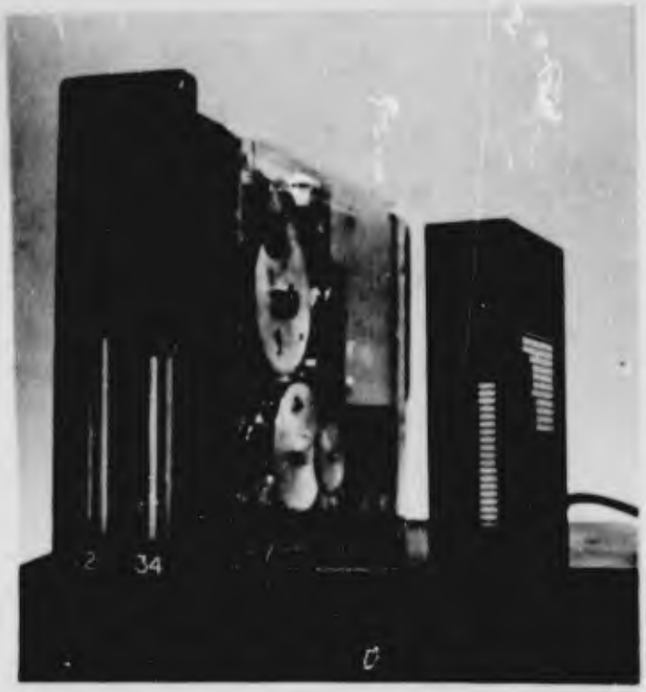


Figure 1. Electromechanical vs Solid State Displays

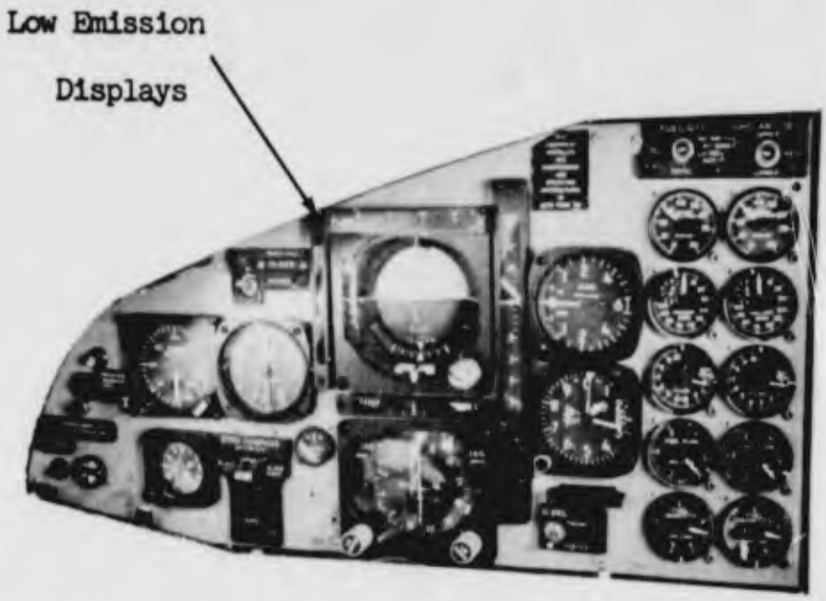


Figure 2. Modified T-39 Instrument Panel

from beneath the electrodes placed on both sides of the EL layer, at least one of these electrodes must transmit light. Such electrodes can be placed on glass, mylar, or on top of the phosphor dielectric layer itself. Many factors are involved in the fabrication of good EL displays, as primary concern must be given to the optimization of trade-offs between basic electrical and optical properties (Reference Appendix I for an analysis of electroluminescence).

Numerous research efforts have been performed to perfect EL lamp fabrication techniques, so as a result it is now possible to construct high quality EL lamps in-house (3). This capability has been invaluable in gaining a better insight into the potential of using EL materials in solid state displays and in assisting in the identification of meaningful and practical research programs. The advantages of EL displays are truly unlimited as they depend primarily on the imagination of the design engineer. It will no longer be necessary to consider electromechanical limitations before the needs of the pilot. By patterning one or both of the two electrodes, light can be made to appear where the electrically selected electrodes overlap. Assuming many electrodes exist, patterns of light can be made to appear to move smoothly making pointer or column presentations possible. As can be seen in Figure 3, the numeric readout also exists as an extremely simple and rugged display device. Complex X-Y matrix displays have also been developed and evaluated for presentation of multi-sensor radar, forward looking infra-red, and TV information; however, application of such displays cannot be considered possible without severe compromises until at least 1969. The capability of being able to present continuous, one dimensional information using solid state display techniques does exist and is the reason for current Air Force interest in applying this capability to the cockpit.

SOLID STATE DISPLAY DEVELOPMENT

The development of practical solid state display techniques has not come without its own set of problems; however, these problems were primarily those being studied by several research groups throughout the world. As a result, initial AFFDL efforts addressed themselves to EL lamp technology and solid state display control circuitry. These efforts were successfully completed and provided the

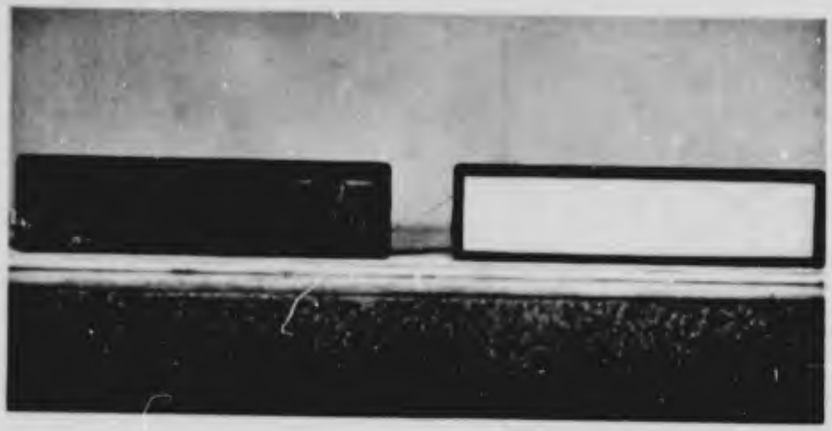


Figure 3. EL Numeric Readouts

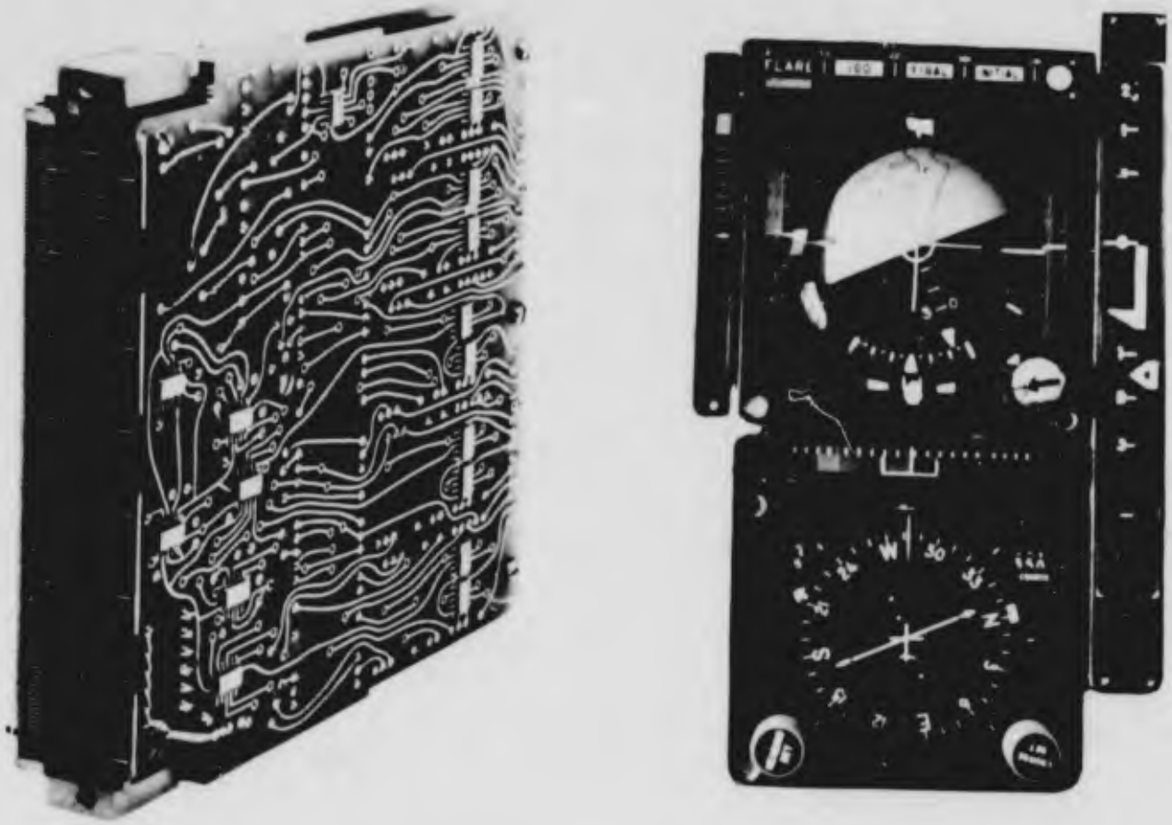


Figure 4. Experimental Solid State Displays

capability of fabricating experimental solid state displays shown in Figure 4. At that time considerable emphasis was placed on control circuitry, as the problem of continuous display control, the characteristics of basic control mechanisms, and the design of advanced control circuitry were considered to be of significant importance to the development of solid state displays (8). This still left one major deficiency, the intensity and effective life of EL phosphors. The lack of success in this very critical area made it quite apparent that it would not be possible to depend on EL phosphor improvements to achieve acceptable readability of solid state displays under daylight ambient conditions. Awareness of this problem motivated a study of human perception where a simple but important determination was made that low emission displays could be seen in daylight ambients if they had acceptable contrast. As it is only permissible to control ambient illumination at the surface of a display panel, reduction of background reflectance would normally be achieved through filtering techniques. This would in turn either restrict the viewing angle or reduce the already low emission intensity of the display (9).

In-house experimentation with neutral density, polarized, and micro-mesh filters in conjunction with high intensity EL lamps re-emphasized a well known fact that filters can improve emission type display legibility under high ambient lighting conditions (Figure 5). Such filtering techniques were not already in use because research oriented personnel were striving to increase, not decrease, the emission intensity of EL displays, even at the expense of display legibility. Since that time, filtering techniques have been developed under the direction of the AFFDL that permit the absorption of a high percentage of ambient light without greatly reducing emitted light of the EL display. This development has resulted in an extremely significant breakthrough in solid state display technology. Appendix II identifies the characteristics of high contrast (HI-CON) techniques that have permitted the fabrication of low-emission 2 percent reflective - 35 percent transmissive solid state displays that can be clearly seen under bright ambient conditions.

The benefit of HI-CON filtering techniques can be realized in a variety of ways, as such displays can either

Figure 5. Filter Enhancement

Figure 6. Legibility Study Panel

be seen under ambient conditions never before considered possible, or they can be used at reduced operating levels and result in extremely low power dissipation and long operating life. Figure 6 shows numeric readouts that were used in cockpit legibility studies in which the improved HI-CON display required 1.3 foot lamberts (ft-l) out (4.3 generated), the original HI-CON display required 9.6 ft-l out (32 generated) and the unfiltered EL display required 36 ft-l out (36 generated) for accurate and immediate viewing. Conditions were such that subjects were preadapted to 5,000 ft-l for 30 seconds and then told to read the display under an incident panel illumination of 1400 foot candles (ft-c). Accuracy of reading was the determining factor in the identification of intensity thresholds (10). The improved HI-CON display can also be seen under illumination levels approaching that of direct sunlight when the emission intensity is increased to 30 ft-l. This filter development is also being applied to storage tubes to help solve the severe daylight legibility problem associated with current and future TV air-to-ground tactical weapon delivery systems (Figure 7).

Achievement of practical solid state displays is far from a one breakthrough effort, as it is actually an effort that must be aimed at many research and development domains. In order to give a true perspective to the reader, the following topics will be discussed; the psychophysical limits and capabilities of pilots using HI-CON displays; the improvement of solid state display tolerance to high temperature; and the requirements for solid state display logic and control circuitry.

PSYCHOPHYSICAL LIMITS AND CAPABILITIES OF PILOTS

The ability of a pilot to read a display is normally dependent on cockpit illumination, the reflective and emissive characteristics of the display, preadaptation and display fixation time, plus work load, fatigue, vibration, hypoxia, noise and other less severe effects. In addition, the use of segmented solid state displays required that consideration be given to the effect of apparent motion of discretely displayed information. Characteristics such as informational brightness, visual angle of the information, background brightness of the informational surroundings, fixation time, informational background color, adaptation color and brightness, and location of informational

Figure 7. HI-CON Storage Tube Comparison

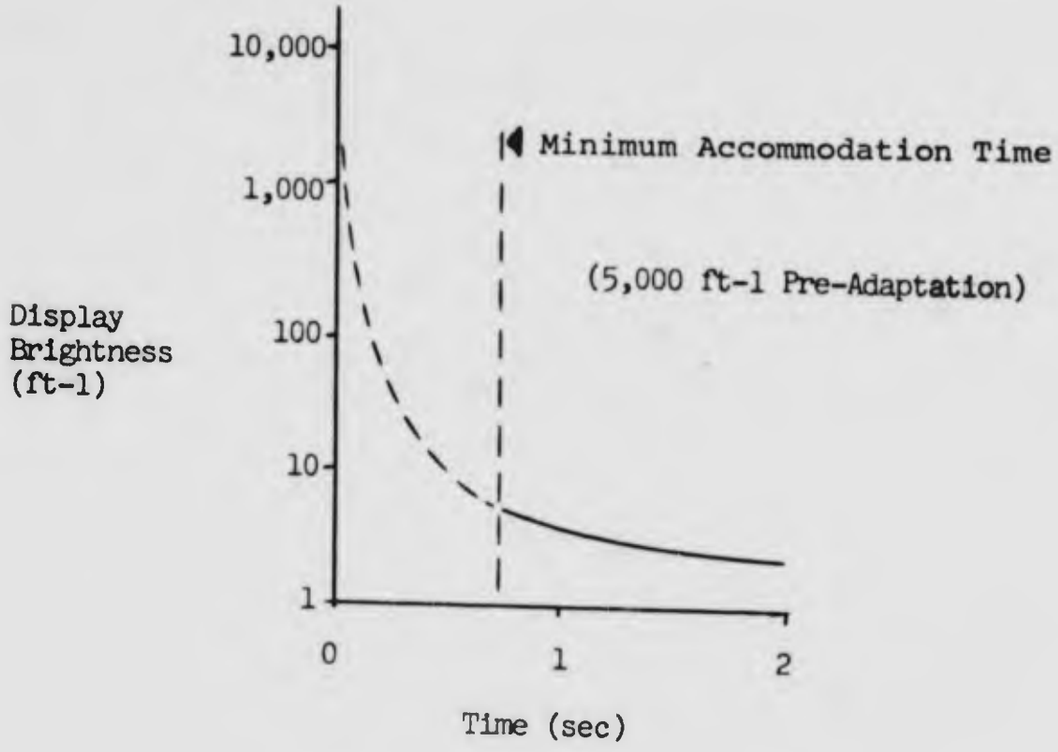


Figure 8. Adaptation Response

stimulus on the retina are also of considerable importance.

Considering the fact that the useful life of an EL lamp depends largely on emission intensity, it is desired to operate EL displays at the lowest intensity compatible with good performance. A lower eimssion intensity limit does exist for the situation where a pilot has been adapted to a skyline of 3,000 to 5,000 ft-l for a long period of time, and where he immediately fixates on the display (2). Under these conditions he would require approximately 3-5 ft-l of emitted light against a dark background to see the display (Figure 8). Since 2 percent reflective HI-CON displays will have to be seen under conditions where incident ambient illumination varies from less than 1 ft-c to the extreme condition of over 10,000 ft-c, background reflectance of the display can not always be assumed negligible. Obviously, there will be some point at which the minimum intensity of 3-5 ft-l would have to be increased (9). Such a situation is more complex but can be effectively expressed in terms of acceptable contrast ratios (a ratio of emitted light to background reflected light). Ignoring color contrast, information stroke width, and a variety of other factors, it is possible to generalize and give an estimate that a contrast of 30 percent would be sufficient for a pilot adapted to 100 ft-l to accurately see displayed data. For the condition where the pilot is preadapted to a high level and the background is above the negligible amount, some contrast level above 30 percent will have to be maintained through increased display emission intensity.

Since the pilot is not always looking outside the cockpit and does not normally look at a panel that has an incident illumination much above 3,000 ft-c, the actual display intensity requirements are reduced. The use of sunglasses (normally 85 percent absorptive) does not apparently change this situation as the relative ratios of the skyline, display, and background do not change, unless of course the glasses are of a type that absorb more light through the top than the bottom. The perceptual capabilities and limits of man determined under a controlled laboratory situation cannot be directly applied to the design of aircraft displays as the actual visual tasks of the pilot flying an aircraft cannot be simply duplicated. Even if a "complete" moving base and visual aid simulator were used with an actual cockpit it would not be possible to accurately simulate the color, intensity and distribution of light

energies entering the cockpit from the sun, the skyline and surrounding clouds. Obviously, inflight evaluation would appear to be a solution to this problem; however, this results in complications with instrumentation and tabulation of statistically valid data. As the lesser of two evils inflight evaluation, by a limited number of pilots and engineers, is by necessity the preferred approach.

To provide a better understanding of the real time perceptual requirements of a pilot, inflight data has been accumulated on KC-135, T-39, C-123, CH-3, and an Apache light aircraft to better determine the intensity and effect of cockpit panel illumination on the legibility of HI-CON EL displays (Figure 9). The data reported on in Appendix III has helped considerably to identify why earlier versions of the HI-CON displays could not be seen as well as expected from laboratory studies. The results of these findings, along with experiences gained with a variety of HI-CON numeric readouts, bargraphs, aircraft displays, and test lamps, has resulted in considerably improved low emission HI-CON displays that can be seen under all ambient conditions experienced in a T-39 cockpit.

Information pertaining to display legibility under low ambient conditions will not be discussed, except for the comment that cockpit night lighting schemes have as goals uniformity and minimized light scattering, both of which are inherent characteristics of EL displays (Figure 10). More important is the consideration for resolution and display continuity. Solid state displays must, by necessity, present data by the motion of light across the face of a display. This can only be achieved by selection and excitation of discretely electroded segments within the display. If for instance, it were desired to have a 1/2 inch by 1/10 inch pointer move over a 5 inch expanse to present localizer displacement information, it could be accomplished by exciting three adjacent 1/2 inch by 3/100 inch segments (leaving 3/1000 inch for electrical isolation between segments) such that a bar of light 1/2 inch by 1/10 inch would appear. By turning off the "right" segment and turning on a segment to the "left" the 1/10 inch wide pointer would appear to move in 1/30 inch steps (13). Depending on requirements for electrical isolation this number could be increased to five or ten segments using high quality fabrication techniques - or by using certain unique techniques could result

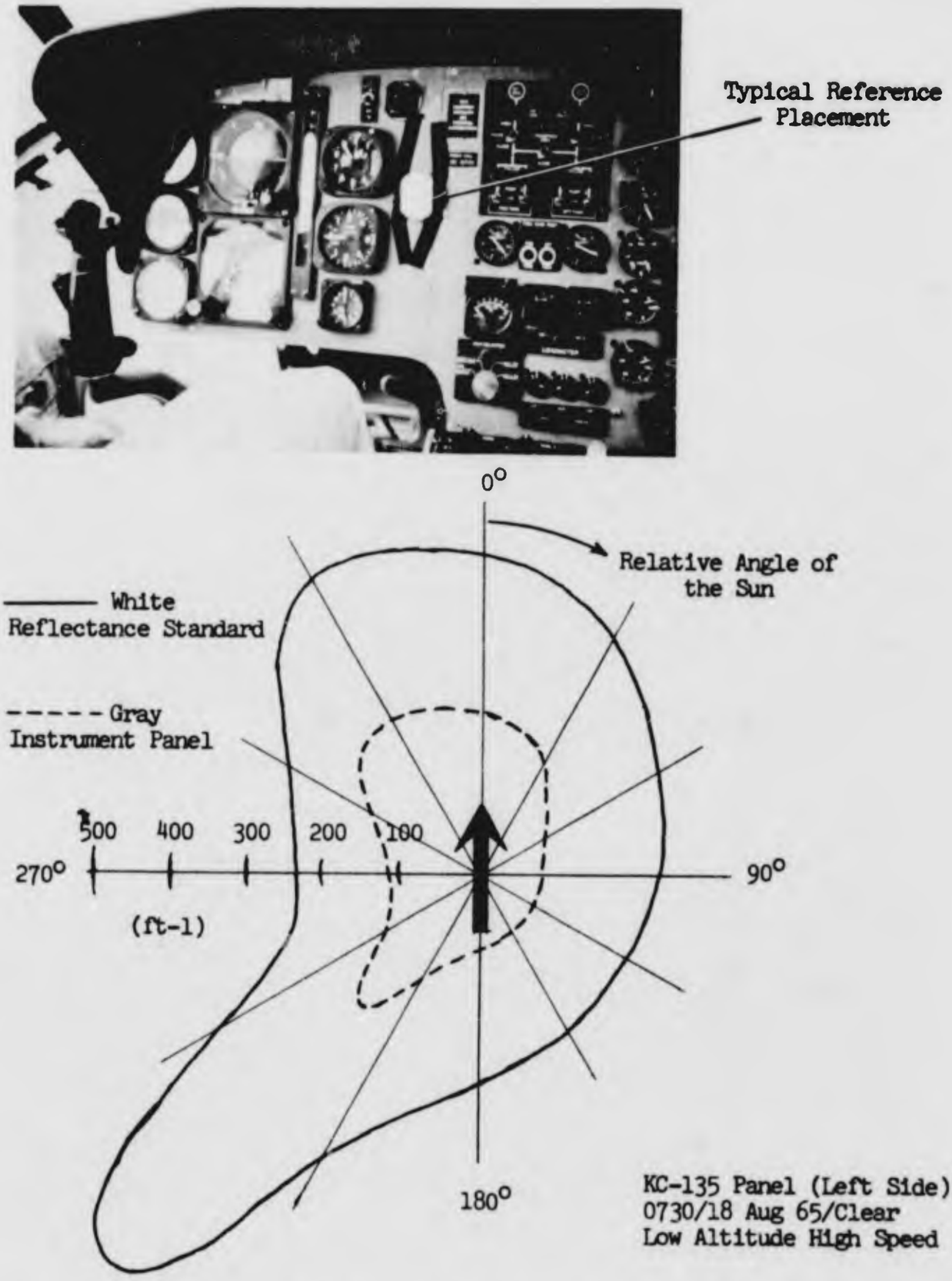


Figure 9. Cockpit Panel Reflectance Characteristics

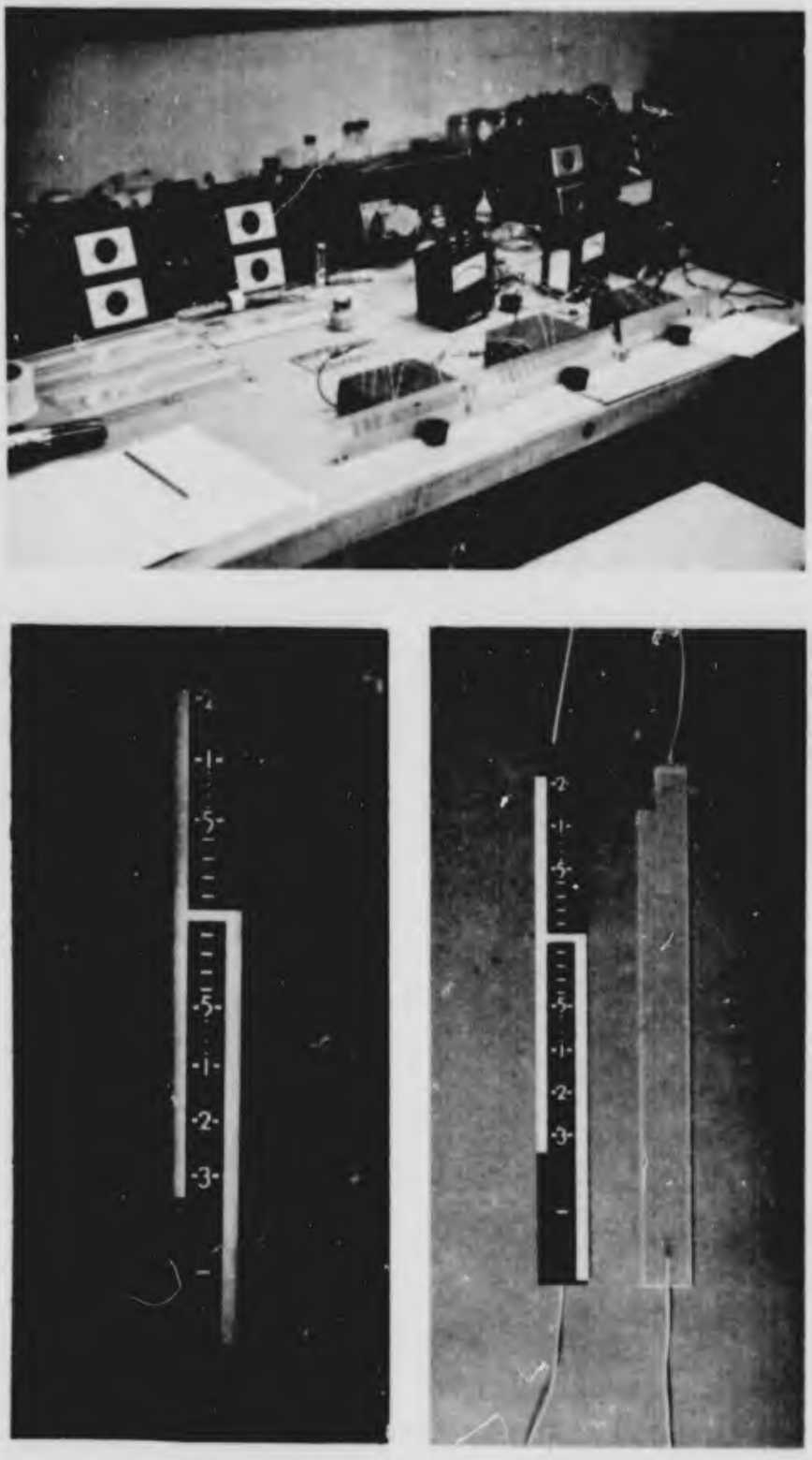


Figure 10. In-House EL Instrument Lighting

in 50 light lines per 1/10 inch. There is a point at which a pilot cannot resolve information without magnification. Further, under normal conditions, the pilot views his displays at from 30 to 36 inches so display resolution achievable with 50 closely spaced segments per inch is considered to be the most that will ever be required. Realistically due to the widely varying dynamic conditions experienced in an aircraft, and due to the limited time devoted to seeing each display, the requirement for less than 50 emitting segments per inch is quite reasonable. It is desirable to present information with the lowest acceptable resolution in a solid state display as less segments per inch means a considerable reduction in the complexity of driving electronics. There is an indication that gapless displays (elimination of the non-emitting regions between segments) will be achieved this year through AFFDL efforts, and will further reduce overall resolution requirements. This is due to the fact that minimum resolution is dictated by the condition in which the aircraft is stable and the pilot is able to see the pointer step from one position to another. Perception of this stepping is highly dependent on the existence of a nearby reference, which is exactly what is provided by the emission gaps found in regular EL displays. Elimination of these gaps will remove the reference, and overall resolution requirements will be reduced.

Directly involved with resolution is a consideration for display continuity. When high rates of stepping occur (1/10 inch per human perceptual "frame time"), the brain integrates the discrete data presented, and for a certain band of rates conveys a sense of motion to the observer. This is an advantage to solid state display design as resolution requirements actually decrease in the presentation of dynamic information. Considering the fact that at any one moment in time, cockpit displays only present a fraction of the data required by a pilot to effectively perform his mission, the analysis of perception must at some point be generalized and put into a proper perspective with other equally important display design factors.

SOLID STATE DISPLAY TOLERANCE TO HIGH TEMPERATURE

Next to legibility, one of the most important considerations pertaining to the use of solid state displays is the ability of such displays to operate in the widely varying thermal ambients existing in aircraft cockpits. This consideration may not be as great in the future as it is now if all of the primary sources of heat, the conventional electromechanical instruments, are removed from the instrument panel and replaced with low power solid state displays. More realistically the introduction of EL displays will be gradual, and will therefore require improved high temperature operating characteristics. The best EL phosphors available have constant voltage and frequency half lives of from 1,000 to 10,000 hours at room temperature, while the same phosphor will reach half brightness an order of magnitude sooner at 90°C (14) It should be recognized though, that aircraft panel temperatures are not normally as high as 90°C and that aircraft displays are operated on the average only a few hours of each day. As a result, useful life even at today's higher panel temperatures, is compatible with the mean time between failure rates experienced with the more complex electromechanical displays.

During the past 18 months the effect of emission degradation on T-39 landing instruments (which surround the ADI, the hottest instrument on the panel) has been so small as to be undetectable. In addition, the use of EL displays in one KC-135 and three T-39's during this same period has not resulted in even one partial display failure. Even though it is possible to rationalize and state the fact that the problem is not as severe as one might be led to believe, there still exists a need for improvement of high temperature operation of EL phosphors. One of the reasons for this need is the suspicion that the degradation effect occurs at elevated temperatures. This means that solution of the high temperature problem could also extend the room temperature life of EL displays to beyond the practical life of any application conceived. This could possibly extend the mean time between failure of the EL portion of solid state displays to beyond the useful life of the vehicle to which they were applied.

As stated earlier in this paper, achievement of the goals of the solid state display program cannot be guaranteed; however, displays emitting a scant 15 ft-l are being effectively used today during all visual and thermal ambient conditions in a T-39, a feat that until recently was predicted impossible without a major breakthrough in phosphor technology. It is considered equally possible that a current AFFDL program to improve the performance of EL lamps at elevated temperatures will help to solve the last major problem confronting the use of EL displays in the cockpit.

Naturally, in order that such a program be successful it is necessary to fully understand the phenomenon of electroluminescence and the associated mechanisms responsible for reduced light output (5). Referring again to the analysis of electroluminescence which is presented in Appendix I, it has been calculated that the imperfection lines, along or adjacent to which the light emission takes place, constitute only about 10^{-5} of the total particle volume. As the energy conversion produces 99 percent heat and for the most part occurs in such a tiny fraction of the particle's volume, it must be concluded that the temperature of these lines is significantly higher than that of the surrounding crystal (4). Apparently this heating accelerates the copper ion drift which in turn results in a blunting of the emitting heads by the screening of local fields. Other factors such as reduced luminescent efficiency and hole storage are also likely the result of localized heating and a possible cause of frequency saturation of EL brightness. From unpublished work of Mr. Petertyl it is possible to assume that phosphor particles have an average temperature which is a function of ambient temperature and the microscopic heat pulses generated within the imperfection lines (11). As a result, the equilibrium temperature of a particle is dependent on:

> "The efficiency of heat transfer within and between a particle and its embedding medium; the thermal conductivity of the embedding medium; the surface-to-volume ratio of the particle; the temperature of the embedding medium (dielectric losses plus the ambient); the voltage and frequency of the exciting field; the volume fraction of phosphor particles in the lamp."

The equilibrium temperature may be more important to the rate of copper ion diffusion and lamp deterioration than the momentary peak temperature when one considers that diffusion is a time-temperature process. Optimized thermal design of EL lamps should permit the lowest equilibrium temperature in the phosphor particle and the least chance for diffusion.

It is a known fact that as the volume fraction of phosphor is increased initial brightness goes up but at the expense of higher power dissipation (heat) and life (Figure 11). The degree of increase in brightness by increasing the volume fraction of phosphor, and therefore the number of light emitting particles, is apparently only limited by the requirement for electrical isolation between particles (6). It has been observed that larger phosphor particles have a more gradually sloped decay curve than do smaller particles (14). The reason for this could be that small particle lamps of the same volume fraction of phosphor would emit more light, generate more heat, and deteriorate faster; or the smaller diameter particles require shorter diffusion distances for copper ions to screen emission; or the smaller particles having a higher total surface area may be more vulnerable to light absorption effects of electrolytically produced zinc.

As it seems unlikely that a new class of high temperature long life EL phosphors will become available soon, an effort has been initiated to utilize the already improved characteristics of zinc sulfo-selenide phosphors but with a modification to improve the efficiency of heat transfer between the phosphor particles and embedding medium by encapsulating the particles in a transparent, high resistivity oxide film. This film, should have a dielectric constant that is equal to or greater than the phosphor dielectric constant; it should have a low dielectric loss; be impervious to minute quantities of water vapor; have a refractance that will permit maximum light coupling; and have as high a thermal conductivity as possible. Further, a dielectric medium should be employed that is the best tradeoff between high thermal and dielectric conductivity, high dielectric constant, low loss, high resistivity, and low water vapor permeability. Finally a high surface to volume phosphor ratio should be used to gain heat transfer area, while using the encapsulation technique to protect the larger expanse of phosphor surface from the harmful effects of water vapor.

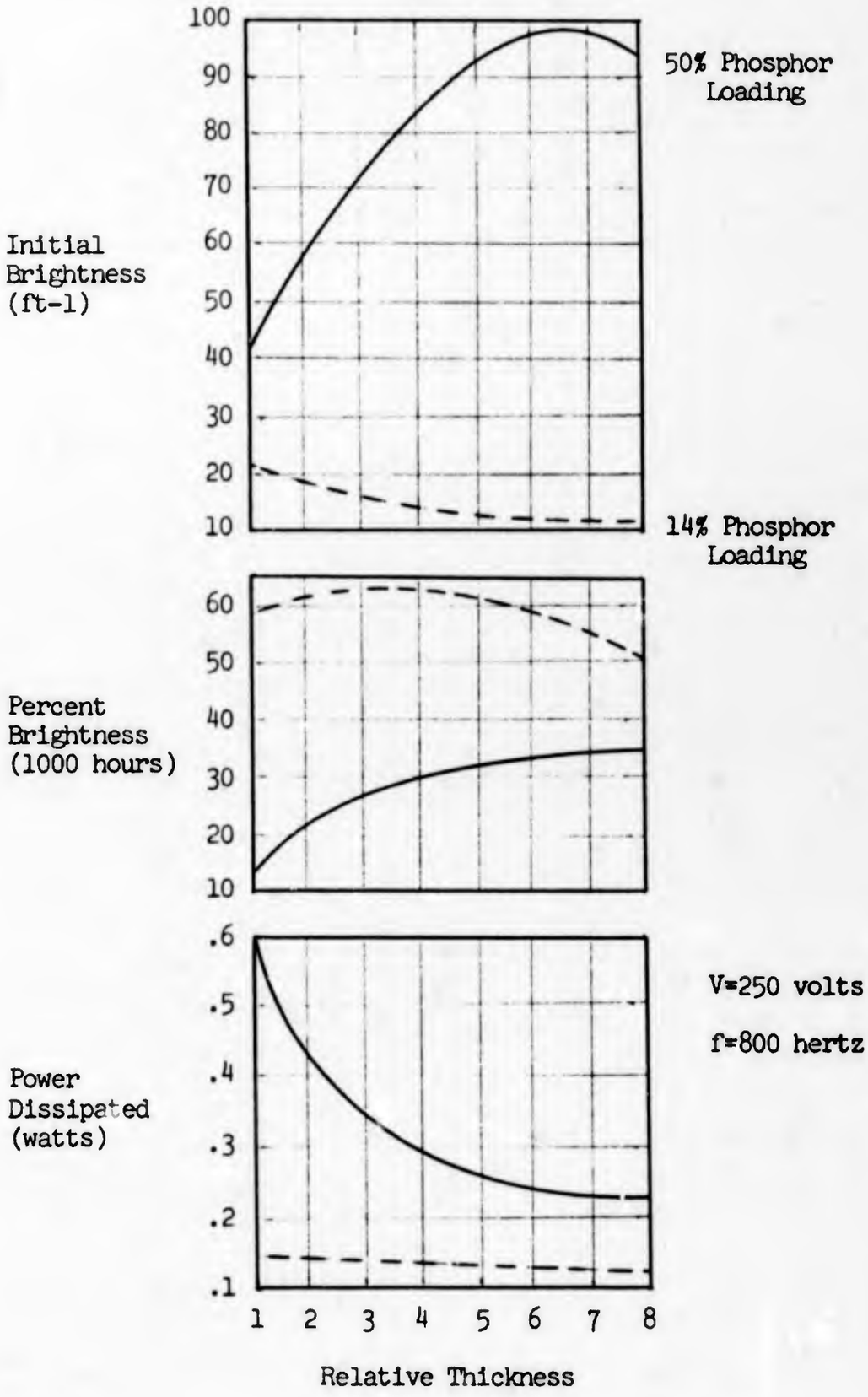


Figure 11. Phosphor Volume Fraction Loading Effects

The results of this and other related studies are expected to permit the fabrication of improved EL lamps by the spring of 1967. Considering the level of success experienced already with the advent of HI-CON techniques and the development of sulfo-selenide EL phosphors, improved temperature operation should guarantee the application of solid state displays to the cockpit.

SOLID STATE DISPLAY LOGIC AND CONTROL CIRCUITRY

Past efforts in the development of solid state displays have primarily centered on the fabrication of control and logic circuitry. This is reasonable as the solid state display concept evolved a number of years ago from the idea that if mass produced circuitry could be developed a display technique could be perfected that would have broad commercial and military applications. A variety of control techniques such as nonlinear resistive, series ferroelectric, and photoconductive have been worked on during the past few years but with limited success (Figure 12). In most cases however, the efforts helped to identify the real problems associated with manufacturing solid state displays. Uniformity, thermal effects, time response, encapsulation problems, and nonrepeatability, quickly curtailed thoughts of complete solid state display panels for aircraft application and thin wall TV sets.

More recently, a number of new approaches to the mass produced control circuitry problem have been supported by the Rome Air Development Center, the Avionics Laboratory, AFFDL, and other Government agencies. AFFDL efforts have been oriented towards the use of uniform large area ferroelectric ceramic strips, and more recently towards the development of bulk phenomena semiconductor control circuits. The goal of the first effort, which resulted in the recent delivery of a 1200 element line-at-a-time coincidently selected video display was to combine the best characteristics of EL phosphors with those of ferroelectric ceramics (7). The materials cost for control of an EL segment has been reduced by this program from about $15.00 to $.30 per element (Figure 13). Unfortunately, the technique did not permit the use of an EL and ferroelectric layer alone but also required a diode at each segment location. Such a

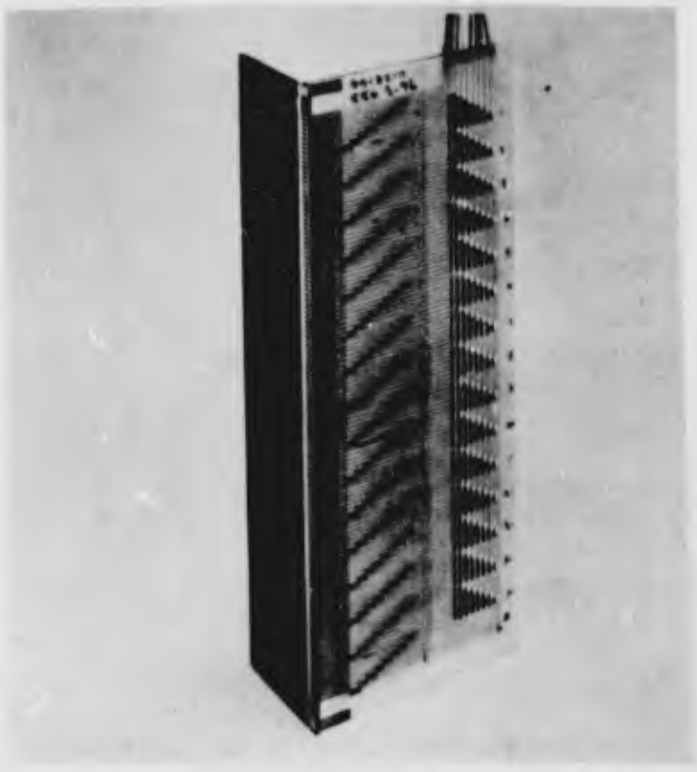


Figure 12. Photoconductive Bargraph Display Control

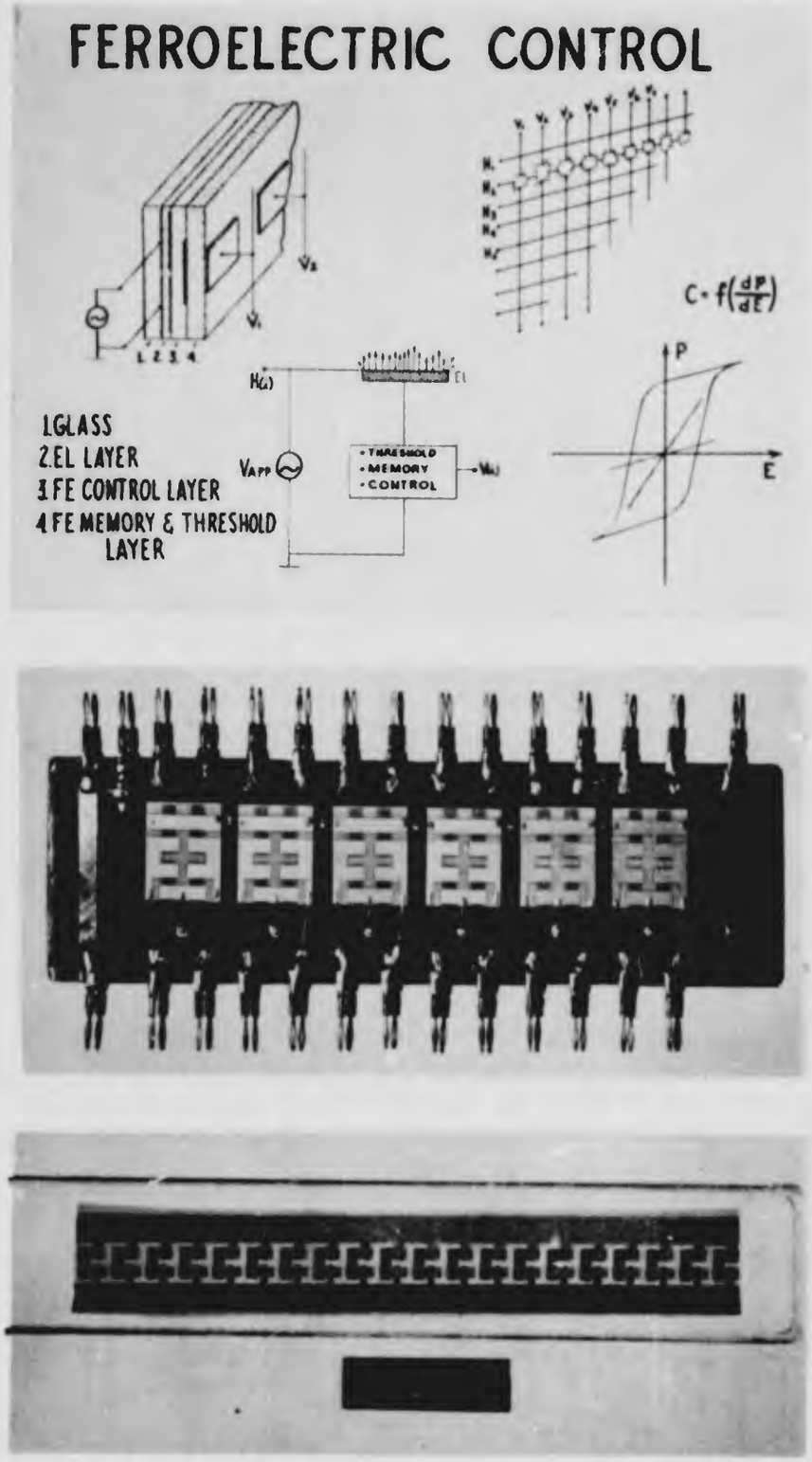


Figure 13. Ferroelectric Ceramic Control Circuitry

requirement is acceptable for many displays but the advantages are certainly less than anticipated. Application of the ferroelectric ceramic techniques to matrix displays where space is limited, is at present the most advantageous application for ferroelectric ceramic control circuits. The second effort, initiated last May, has been concerned with the development, fabrication and packaging of large arrays of bulk phenomen control circuits (Figure 14). Such an approach was to optimize recently improved semiconductor phase change materials for the control of EL lamps (1). The materials used are of such a nature that they can be changed from a high impedance nonconductive state to a low impedance fully conductive state in less than a microsecond. Resistance of the material can be varied over seven orders of magnitude (from 100 megohms to 10 ohms). As bulk semiconductor devices do not require a depletion layer for operation, they are insensitive to changes in polarity and are thereby ideally suited for controlling A.C. driven solid state displays. There are many other features of the material such as insensitivity to impurity levels, simple fabrication techniques, characteristic structuring and restructuring electronically, wide operating temperatures, and the predicted capability of reliably fabricating thousands of such devices on a single substrate at an extremely low cost. This effort was initiated to provide practical cost competitive solid state displays for the cockpit by the end of 1967.

Regardless of the breakthroughs achieved in the legibility, maintainability and reliability of EL displays; the full capabilities of nonmechanical solid state displays cannot be achieved until suitable control circuits have been developed. These circuits will not only be required to achieve basic capabilities such as distributed threshold, memory and high voltage control in order to be suitable; but they must also have the capability of being fabricated in highly reliable, inexpensive and sufficiently large arrays. The significance of such a requirement must be emphasized as it is presently quite expensive and difficult to construct the control circuitry for solid state displays. Due to the limited availability of ferroelectric ceramic and bulk semiconductor control circuitry, it is still necessary to utilize silicon control rectifiers for the control of light emission in solid state displays (12).

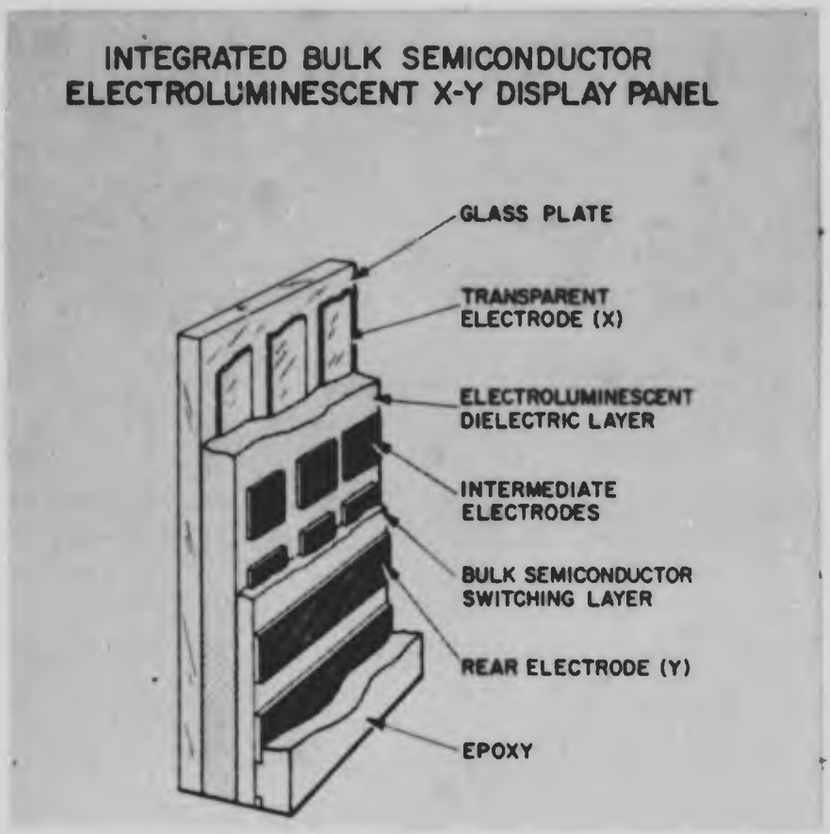


Figure 14. Bulk Phenomena Control Circuitry

CONCLUSION

In summary, the effort described is one part of a program entitled "Advanced Display Generation Techniques" which in turn is an integral part of the Air Force Flight Dynamics Laboratory "Control-Display" effort. Successful completion of this effort is expected to permit the development of a family of completely electronic solid state displays capable of effectively presenting flight control information. Solid state displays may very well be the display of the future. They may permit significant improvements in the cost, reliability and accuracy of aircraft displays or they may even be the basis of tomorrow's commercial video displays. Elimination of the apparently insurmountable requirement for higher brightness of EL phosphors has already had a strong and favorable impact on the future of solid state displays. Even so, full achievement of the end goals of the solid state display program will only be possible when key accomplishments are made in: the development of bulk-phenomena materials for display-integrated electronics, the improvement of solid state display tolerance to temperature extremes, and the utilization of solid state display mechanization concepts in the development of meaningful and useful displays.

In a slightly different context, mention should be made of the potential of applying EL to cockpit panel lighting. The requirement for EL panel lighting is valid, as it is based on improved reliability; guaranteed emission uniformity; no heat operation; low power operation; if necessary, programmed rather than unprogrammed replacement; extended freedom of panel design; smaller size, less weight, less interconnections; better appearance and less clutter; minimized panel wiring; and vibrationless high "G" and high shock operation. In the past, EL panel lighting suffered from many difficulties. The primary shortcoming was the fault of the sulfide class of EL phosphors. Use of such phosphors resulted in low output brightness, unacceptable halflife and poor operating characteristics at elevated temperatures. As a result of recent phosphor development programs these problems have been effectively overcome for the panel lighting application. Operation of sulfo-selenide multi-color phosphors at low intensity levels can result in extremely uniform brightness characteristics for thousands of hours. Use of constant current rather than constant voltage operation further improves uniformity to a level comparable to that achieved with

incandescent bulbs, but without the possibility of catastrophic failure. Achievement of specific emission colors for the life of the panel is probably the only characteristic open to question, as a slight color shift sometimes occurs with extended EL operation; however, indications are that the problem will either be solved under the temperature study program, or that in fact the effect is so slight that the benefits of EL lighting will greatly offset the disadvantages of any such characteristic. It should be mentioned that the Air Force has successfully used EL panel lighting in the rear cockpit of the YF-12A for a number of years.

Since the progress of the AFFDL solid state display program has been based entirely on an attempt to understand all aspects of the control-display problem, it has been considered necessary to perform research to study future display requirements. One such effort has been to design an experimental solid state altimeter display (Figure 15). The possibility exists that such a display could be developed as a result of this study; however, the primary goal of the effort has been to understand and identify the problems associated with complex solid state displays. Obviously, there are still many engineering and psychophysical questions that must be answered about the display shown in Figure 15.

Previous efforts in the area of solid state displays have been hindered by two limitations: the emission intensity of field effect electroluminescent phosphors and the complex control circuitry required in high resolution solid state displays. As a result of efforts on both EL phosphors and high contrast techniques, the first limitation has been dramatically overcome. Equally significant, the work performed on control circuitry has provided the indication that simple and reliable control circuits will soon be available for Air Force application.

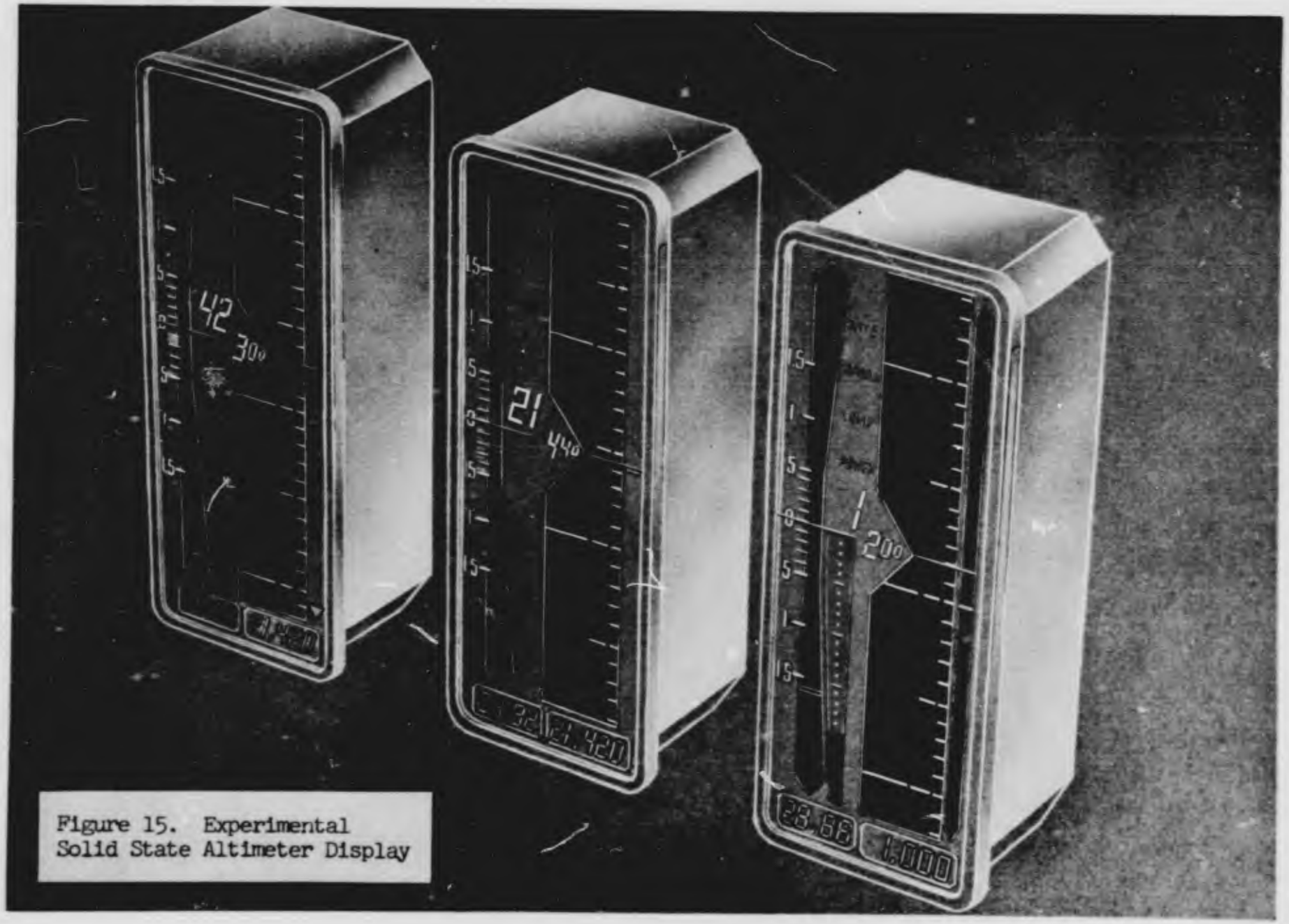


Figure 15. Experimental Solid State Altimeter Display

BIBLIOGRAPHY

1. Babb, B.A., New Switches and Circuits for Control of Electroluminescent Arrays. Proceeding of the National Aerospace Electronics Conference, 135 (1966).

2. Blackwell, H.R., "Contrast Thresholds of the Human Eye." J. Opt. Soc. Amer., 36, 624-643 (1946).

3. Calicchia, Richard, Laboratory Processing of EL Panels. Technical Report. (Rome Air Development Center) Griffiss Air Force Base, New York (1963).

4. Fischer, A.G., Investigation of Carrier Injection Electroluminescence. AFCRL-62-588. Air Force Cambridge Research Laboratory, Bedford, Massachusetts (15 August 1962)

5. Henisch, H.K., Electroluminescence. The MacMillan Co. New York. (1962).

6. Katona, Gabriel P., "The Effect of the Phosphor-Embedding Medium on the Performance of Electroluminescent Cells." J. Electochem, Soc., 8, 109 (August 1962).

7. Lechner, Bernard J., Solid-State Raster Scanning for Display. AFAL-TR-65-106. Air Force Avionics Laboratory, Wright-Patterson AFB, Ohio (May 1965).

8. Peterson, Carlton J., An Investigation of Electroluminescent Video Displays. AFIT Thesis. Wright-Patterson AFB, Ohio (September 1963).

9. Peterson, Carlton J., and Smith, Hugh A., Development of High Contrast Electroluminescent Techniques for Aircraft Displays. AFFDL-TR-66-6 (Air Force Flight Dynamics Laboratory) Wright-Patterson AFB, Ohio (1966).

10. Petertyl, S.V., Fuller, P.R., Improved Contrast And Visibility for Electroluminescent Displays. Technical Paper presented to the National Aerospace Electronics Conference, (paper distributed during conference) (1966)

11. Petertyl, S.V., Improving the Performance of Electroluminescent Lamps at Elevated Temperatures. AFFDL-TR-(Forth coming under Air Force Contract AF33(615)-3872.) (1967).

12. Wade, James R., An Investigation of Solid State Display Control Circuitry. AFIT Thesis. Wright-Patterson AFB, Ohio (June 1965).

13. Development of a Solid State Display Device. Technical Documentary Report No. RTD-TDR-63-4163 (Research and Technology Division) Air Force Flight Dynamics Laboratory, Wright-Patterson AFB, Ohio (1964).

14. Improved Electroluminescent Phosphors. Technical Documentary Report No. RADC-TDR-64-159 (Rome Air Development Center) Griffiss AFB, New York (1964).

NOTE

The following appendices have been deleted to minimize the length of this paper:

I. Electroluminescence

II. HI-CON Displays

III. Reflective Measurements of Cockpit Panels

Copies of these supporting appendices will be made available by the author to interested personnel.

BLANK PAGE

(U) USING THE ELECTRONIC COMPUTER TO DEFINE AND IMPLEMENT POLICY

by

Raymond E. Christal

Personnel Research Laboratory
Aerospace Medical Division
Air Force Systems Command
Lackland Air Force Base, Texas

Dr. Raymond E. Christal

DR. RAYMOND E. CHRISTAL

Chief, Occupational Structures Research Division

After being released from commissioned service with the US Navy at the end of World War II, Mr. Christal returned to North Texas State College where he received his BS (History and Mathematics) and MS (Psychology) degrees in 1948. Immediately after graduation he accepted a position as Test Psychologist with the Personnel Research Laboratory. He has served continuously with the Laboratory since that time, becoming Chief of the Officer Section of the Selection and Classification Division in 1954, and Assistant Chief of that Division in 1957. In 1958 he was appointed as Chief of the Occupational Structures Research Division and has served in that position to this date.

Dr. Christal received his PhD in Psychology and Educational Psychology from the University of Texas in 1956, having completed all requirements during off-duty hours. He served as a member of the night school faculty of Trinity University during the 1950-58 period.

ABSTRACT

Using an electronic computer, the fixed-X regression model can be applied to determine how relevant variables must be weighted in order to reproduce a sample of judgments rendered by an individual or committee. The resulting equation can simulate the individual or committee in judging new stimuli or situations of the same type. The fixed-X model permits introduction of interaction, non-linear, and categorically-coded predictor variables which often are necessary to account for an individual's "rating policy." Equations thus far developed are highly valid, and have held up on cross-application. Similarities and differences in "policies" can be detected by clustering judges in terms of the homogeneity of their regression equations. In a recent study, a Hq USAF board of senior officers reviewed descriptions of 3575 representative officer positions and made decisions concerning the appropriate grade level to be associated with each. In an effort to identify the factors considered by these board members in making their judgments, over a hundred variables were hypothesized and evaluated. Eventually, a nine-predictor equation was developed which accurately reproduced the board's actions. Subsequently, this equation was applied by the computer to determine appropriate grades for an additional 10,000 officer positions. In other applications, the model has been used to study promotion board policies, and the policies of supervisors in evaluating their subordinates. The model is being used to develop a mechanized airmen assignment system which will duplicate the actions currently performed by assignment specialists. A study is planned to develop a reassignment model which will give appropriate consideration to job and personnel characteristics. While previous applications have been in the military setting, the model may be applied to study individual and group judgments concerning such diverse properties as the beauty of pictures, effectiveness of workers, quality of English compositions, or acceptability of job applicants.

Using the Electronic Computer to Define and Implement Policy (U)

The fixed-X linear multiple regression model can be used to determine how relevant variables must be weighted in order to reproduce a sample of judgments rendered by an individual or committee. The resulting equation can be applied to simulate the individual or committee in judging new stimuli or situations of the same type. The fixed-X model permits introduction of interaction terms, powered terms, and categorically-coded predictors which are often necessary to account for an individual's "rating policy."

I've prepared a short fairy tale to illustrate how the model is used.

Once upon a time, there was an Oriental king who was concerned as to how he might make a name for himself in history. "I know," he said, I'll select me a harem larger than King Solomon's."

So the word went out, and soon thousands of young girls were arriving from the various provinces to seek the king's approval.

Early one morning the king began his selection process. As each girl filed by, the king carefully scrutinized her and then expressed his judgment.

"Excellent!" he would say. "This one is very pleasing to my eye." Or perhaps he would hum and haw. "This one's not too bad. Maybe yes, maybe no." Many times he would show his disapproval in no uncertain terms. "Never!" he would say. "Pass on! Pass on!"

In each instance, the Court Recorder attempted to quantify the king's degree of approval by checking the appropriate level on a 9-point scale which had been devised especially for the occasion by the Chief of the Royal Psychometricians.

By suppertime the king had considered some 300 girls. His eyes and imagination were beginning to tire.

"Most High First Counselor," he said, "you've been watching me all day and by now you should know my likes and my dislikes. I've decided to leave the selection of my harem in your hands. But take care! If your choices do not please me, it will be your head!"

After the king retired, the Most High First Counselor summoned the Chief of the Royal Psychometricians. "I'm passing the job on to you," he said. "If you fail to please the king, your head will roll along with mine."

The Chief of the Royal Psychometricians called his staff together and explained the situation.

"We must not fail," he said, "or it will be all of our heads."

"How shall we proceed?" asked one of the young staff members, who was fresh out of the Royal Academy.

"Well," responded the Chief, "we know how the king rated the first 300 girls. Right?"

"Right!"

"And we can see everything the king saw when he looked at the girls."

"Right?"

"Right!"

"Then all we have to do is to uncover the girly characteristics considered by the king and determine how he weighted them in his final judgment. This is a natural for the Multiple Linear Regression Model."

"But how do we know which characteristics he considered?" asked the neophyte.

"We don't, you fool! Didn't they teach you anything in school? That's what the regression model is for. If a girly characteristic adds to our ability to predict the king's ratings, we may assume he gave it consideration. Now let's get on with the business."

"How about height?" asked one of the staff members. "Does the king like short girls or tall girls?"

"Neither," replied another. "I would guess that the relationship between height and the king's preference is curvilinear. Some girls are too tall, while others are too short."

"Well," responded the neophyte, "if the relationship is curvilinear, then we cannot use the linear regression model. If you were to plot the curve between height and acceptability, I think you would find it to be parabolic."

"They really didn't teach you very much in that school, did they?" commented the Chief. "What is the general equation for a parabola?

"$ax^2 + bx + c$," responded the neophyte.

"Bravo!" declared the Chief. "Now let X be a vector of heights. If we square each value in the height vector, we generate a new vector X^2. Now, if we introduce these two predictors in the regression model, we can solve for the set of least squares weights a and b, plus a regression constant c. There's no problem in fitting curvilinear relationships with the linear regression model as long as the proper power terms are introduced as predictors. The linear restriction is on the weights and has nothing to do with form of the predictors."

"How about eye color?" asked one of the other staff members who was eager to move on. "I'm sure the king looked at the color of each girl's eyes."

"Fine," said the Chief, "we will consider eye color in our equations. Since eye color is not an ordered variable, we must introduce a separate categorically-coded predictor for each color."

"What the Chief means," whispered one of the staff members to the neophyte, "is that for a variable associated with a particular eye color, each girl will be assigned a value of 1 if her eyes are that color, and will be assigned a value of Ø if her eyes are not that color."

"It's been my observation," said one of the group, addressing the Chief, "that the king likes blue eyes on blondes, but not on brunettes."

"That's easily handled," responded the Chief. "First we will introduce categorical predictors for each hair color; then we can cross-multiply eye color and hair color variables in order to generate the appropriate interaction predictors."

"I don't understand," confessed the neophyte.

"Well," replied the Chief, "if we multiplied the value of a blue-eyed blonde on the blue eye variable times her value on the blonde hair variable, what would be the result?"

"Oh, I see!" said the neophyte. "She would get a value of 1 on the interaction predictor, while the product for girls with any other eye-hair color combination would be Ø."

"Exactly," said the Chief, "you're learning."

"I thought," said one of the group, "that the regression model assumes predictors to be normally distributed and also that their joint-distribution is normal. We certainly can't meet this assumption using powered terms, interaction terms, and categorically-coded predictors."

"You're right," said the Chief, "if you're thinking about the multinormal model. But we're going to use the fixed-X model, which makes no assumptions. We would be stupid to restrict ourselves to normally distributed predictors. It would force us to omit most of the variables which we know the king considered."

"But," objected the staff member, "if we use the fixed-X model, we cannot generalize beyond the computing sample."

"Who can't?" responded the Chief. "Just because we can't talk about standard errors and make probability statements about population parameters, let's not assume our equation will fail to hold up."

"Well, I'm from Outer Missourivich," said the staff member.

"Very well," replied the Chief, "if it will make you feel better, we will develop our equation on the first 150 girls and cross validate on the second 150."

And so went the conference into the wee hours of the morning. Over a hundred predictors were eventually defined, each representing a girly characteristic which might have influenced the king's judgments. The time had now come for the acid test. Could they produce an equation which would simulate the king?

Some of the royal guards had to be called in to help measure and evaluate each girl on the predictor variables. It was a madhouse with 20 guards checking eye colors, weighing, and measuring the 300 girls.

By late afternoon, the raw data had been accumulated. All that night and throughout the next day and night one could hear the constant clicking of the abaci beads coming from the Royal computing shop. Then came the answer:

"We got an R^2 of .87 and it held up in the cross-application sample," reported a messenger to the Chief of the Royal Psychometricians, who was with his staff in the coffee room anxiously awaiting the results.

"Humm," said the Chief, "that's pretty good. But it's not good enough for me to risk my head on it. There must be some variable we failed to consider."

"Maybe the king likes girls who look like his mother," offered the neophyte. "Men often do."

"You're a genius," said the Chief. "It's certainly worth a try."

"How can we quantify that?" asked a staff member. "You certainly can't measure it with a yardstick."

"We'll establish a rating board," responded the Chief. "Each board member will judge how much each girl looks like the king's mother, and we will use an average of their ratings for each girl as our new predictor."

When the new variable was introduced into the king's policy equation, the R^2 jumped to .94. Everyone now felt confident. The rest was routine. By the end of the week, all of the 8,000 girls in the applicant pool had been evaluated by the equation, and those with the highest composite scores were selected.

The king was very pleased with the results, and as a reward, he gave the Most High First Counselor and the Chief of the Royal Psychometricians their choice of the leftovers.

In the preceding fairy tale, I have described capturing the policy of a single judge. However, if there is high interrater agreement among members of a policy board, the mean values can be used as the criterion vector to represent the entire board. If interrater agreement is low, it may be

that the raters can be divided into two or more groups within each of which there is high agreement. This can be accomplished through application of a hierarchical grouping technique which clusters criteria in terms of the homogeneity of their prediction equations. Thus, if more than one policy exists among board members, each such policy can be identified and described. Differences in policies are thereby pinpointed for arbitration.

At the Personnel Research Laboratory we have found the policy-capturing model to have many applications. Equations thus far developed are highly valid and have held up on cross-application.

For example, we have developed equations to simulate officer promotion boards and have been able to predict, with a high degree of accuracy, the ratings that a board will give to officers under consideration. At the present time we are developing an equation to simulate the actions of career counselors who make the initial assignments of airmen graduating from basic training at Lackland Air Force Base. In the near future we expect to use these equations in a computer-assisted assignment system.

In one study we wanted to determine the relative importance of certain variables in determining the value of airmen working in a given Air Force specialty. Some of these variables related to specific knowledges and skills, some to attitudes and motivation, some to supervisory talents, and so on. We planned to have a number of supervisors in the field evaluate a sample of workers described in terms of these characteristics, and then develop equations expressing their policies.

Now the main expense in conducting a study of this nature is associated with the collection of predictor information. The method involves the administration of tests and the collection of ratings on hundreds of workers at dozens of Air Force bases. So instead of using data collected on real live people, we decided to develop samples by ascribing scores to simulated workers. We created several samples in this manner, each containing 250 cases. It can be demonstrated that such a procedure will yield exactly the same policy equations for this type of study as one would get using real people, provided two conditions are met. First, every case which is generated must be within bounds of reasonableness to the judge. That is, you are not permitted to describe a case as being 18 years old and having had 25 years of education. Second, the scores must be ascribed in a manner which will assure a reasonable variance on each predictor. Not one of the two hundred judges in this study realized that he was dealing with artificial workers. Furthermore, many of the obtained policy equations correlated in the high 90's with the ratings from these judges.

This brings me to a discussion of the Officer Grade Requirements Study, which to my knowledge is the largest effort on record involving policy capturing and implementation in an operational setting.

The Director of Air Force Manpower and Organization asked us in the Personnel Research Laboratory if we could conduct a study to determine the appropriate distribution of grades for jobs in various officer specialties and utilization fields. We said, "Fine--if you will make a policy decision concerning the appropriate grade levels for a sample of

jobs which we will select, then we will provide you with an equation to determine the appropriate grade levels for the rest of the jobs in the Air Force."

The agreement was made, and as a first step we collected comprehensive job descriptions from 85,000 officers. We selected from these a sample of 3,575 descriptions representing jobs at all levels in all specialties. Then a policy board composed of 22 colonels was called at HQ USAF to determine the appropriate grades for jobs in this sample.

Five members of the board provided independent grade ratings for each of the 3,575 jobs. Board members had access to any information needed about the nature of a job being rated. First, they had the job description. If they needed more information, they could consult the members of the panel identified as being knowledgeable in the relevant career area. If that was not enough, they could call special air staff consultants available at HQ USAF. Finally, if needed, they could call the supervisor of the incumbent in the job being rated.

Ratings were independent. Board members were not allowed to have knowledge of the current UMD grade authorization for the job. They were not informed of the grade held by the incumbent or by his supervisor. Members were not permitted to ask other board members or consultants concerning the appropriate grade for the job being rated.

Since ratings provided by the policy board were to be used for establishing Air Force grade requirements, it was important to demonstrate that these ratings were stable; that there was high agreement among board members concerning grade requirements for particular jobs; that the raters were not biased for or against jobs in various specialties or commands. To this end, a series of analyses of the policy board ratings was accomplished. We were prepared to abort the entire study if for any reason the board ratings could not be defended.

The tables and figures in your handout show the results of these analyses. First, the level of interrater agreement was .92, which was considered acceptable for the proposed application. Second, the raters expressed a high level of confidence in their grade ratings. Third, the raters were not biased toward jobs in their own specialty or command. Finally, the analyses indicated that raters were not simply giving us back the grade levels currently authorized for jobs; and they did not recommend a wholesale across-the-board increase in grade authorizations.

Once the policy board's grade ratings for jobs in the criterion sample had been accepted, we began our efforts to develop a policy equation. This turned out to be a challenging bit of detective work, because we had no information indicating what factors were considered by board members in making their decisions about grade requirements. During the next eight months we hypothesized and evaluated over a hundred variables which might have influenced their judgments.

Table 4 in the handout describes the variables in the final equation. The first five variables are job evaluation factors rated by lieutenant colonels and majors in the field and may be considered as being primary

definers of officer grades. According to the policy equation, the grade of a job is determined first by the complexity, variety, and level of activities managed; second, by the scope and significance of work for which planning is done; third, by the requirements for special training courses and on-the-job experience; fourth, by the importance and independence of judgments and decisions required by the job; and fifth, by the level of agencies and individuals to which or to whom the incumbent must communicate. In addition to the five job evaluation factors, the organizational level of the job and the level of the job within this organization help to determine its appropriate grade level. It was found that supervisor's judgments concerning the appropriate grade level for jobs tended to be inflated. However, the equation automatically adjusts supervisory ratings to bring them into line with the policy board's recommendation, and this variable was found useful for inclusion in the equation.

After we were satisfied that we had an appropriate policy equation, we reproduced copies of the descriptions for an additional 10,000 jobs and sent them to the field for rating on-the-job evaluation factors. Altogether we collected some 840,000 factor-by-job ratings from 2,500 majors and lieutenant colonels. A total of 1,750 jobs from the original criterion sample were included in this administration so that we could revalidate the policy equation. Correlation between predicted policy board ratings in this sub-sample using data collected at time 1 and again at time 2 was .93, and the two distributions of predicted scores had approximately equal means and standard deviations. Thus, the policy equation was shown to be stable across time, and again shown to be stable across judges.

At this stage of the project, appropriate grade requirements had been determined for 11,825 jobs. This sample was used as a base to determine the appropriate distribution of grades for various specialty and specialty groups.

I've included representative tables in your handout to indicate the extent and nature of the changes in grade authorizations which would take place if the policy of the board were to be implemented. I think you can see that implementation would not be easy. For example: the board's policy recommends only 93 lieutenant colonel jobs in the Navigator-Observer area, rather than the 885 currently authorized. Many of these lieutenant colonel jobs would be downgraded all the way to the lieutenant or captain level. General downgrading would also occur in the pilot area.

At the same time, gross upgrading would occur in Scientific and Development Engineering, in Aircraft Control, Operations, and Communications-Electronics.

One of the surprising outcomes of the study is that a single policy equation is applicable for determining grade requirements in all officer areas. Recommendations for upgrading in the S&E area is due to requirements for management, planning, decision-making, and communication skills. Downgrading in the aircrew area is due to lower requirements for these same skills. It is interesting that formal education did not enter into

the policy equation, even though it was included in the analysis.

It should be pointed out that the Officer Grade Requirements Study considered only the appropriate grade levels for specific jobs. It did not give consideration to appropriate pay. It may well be that the aircrew jobs recommended for downgrading from lieutenant colonel to captain should be paid lieutenant colonel wages. Also, the study did not concern itself with the number of grades which must be authorized at each level to support orderly career progression of officers. Nor did it consider the importance of keeping pilots and navigators in the cockpit in order to amortize the Air Force's training investment.

The recommendation for upgrading thousands of jobs in the S&E area to the major level looks good; but these positions cannot be filled with qualified majors unless more lieutenants and captains are retained.

The results of the OGR study must be only one of many items to be considered in establishing the Air Force's total requirements for officer grades. My purpose in describing this study today is to illustrate an application of the policy capturing model. In this study the model permitted us to express the policy of a board in a definite and precise manner, and the meaning of grade to this board is defined.

The Air Force is constantly calling boards together to determine how variables should be weighted together for making decisions. Who shall be promoted? Who shall be taken off flying status? Which specialties shall be given proficiency pay? Which officers shall be integrated into the regular Air Force? Which individuals shall be selected as astronauts? The solution of all such problems involves the weighting together of factors judged to be relevant for achieving agreed-upon goals. If these factors are made explicit, then the multiple linear regression analysis model can be applied to derive a precise statement of the factors and weights to be used in carrying out the board's recommendations.

And while the applications I have described have been mostly in the military setting, I'm sure you can see how the model can be applied to study such diverse properties as the quality of beefstock, the beauty of pictures, the effectiveness of workers, the quality of English compositions, or the acceptability of applicants for a king's harem. Thank you.

REFERENCES

Bottenberg, R. A. & Ward, J. H., Jr. Applied multiple linear regression. PRL-TDR-63-3, AD-413 128. Lackland Air Force Base, Tex.: Personnel Research Laboratory, Aerospace Medical Division, March 1963.

Bottenberg, R. A., & Christal, R. E. An iterative technique for clustering criteria which retains optimum predictive efficiency. WADD-TN-61-30, AD-261 615. Lackland Air Force Base, Tex.: Personnel Research Laboratory, Wright Air Development Division, March 1961.

Christal, R. E. Officer grade requirements I. Overview. PRL-TDR-65-15. Lackland Air Force Base, Tex.: Personnel Research Laboratory, Aerospace Medical Division, September 1965.

Chrsital, R. E. JAN: A technique for analyzing group judgment. PRL-TDR-63-3, AD-403 813. Lackland Air Force Base, Tex.: Personnel Research Laboratory, Aerospace Medical Division, February 1963.

Ward, J. H. Jr., & Davis, Kathleen. Teaching a digital computer to assist in making decisions. PRL-TDR-63-16. Lackland Air Force Base, Tex.: Personnel Research Laboratory, Aerospace Medical Division, June 1963.

BLANK PAGE

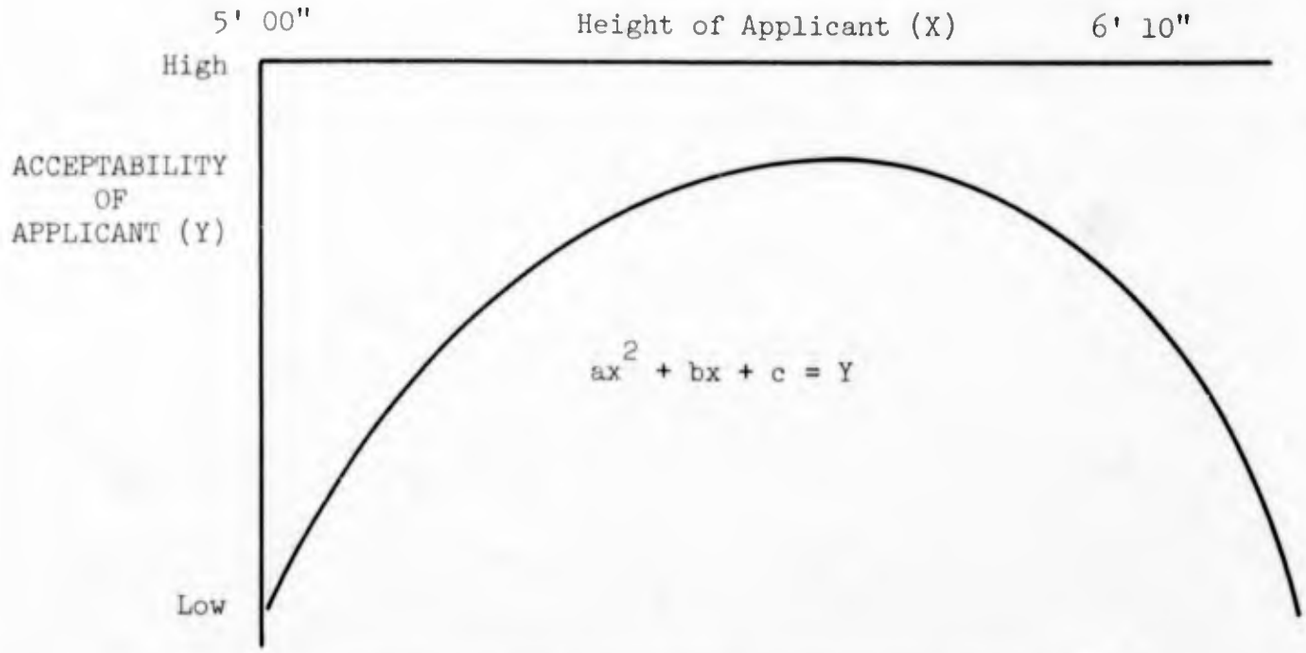


FIGURE 1 RELATIONSHIP BETWEEN APPLICANT HEIGHT AND JUDGED ACCEPTABILITY

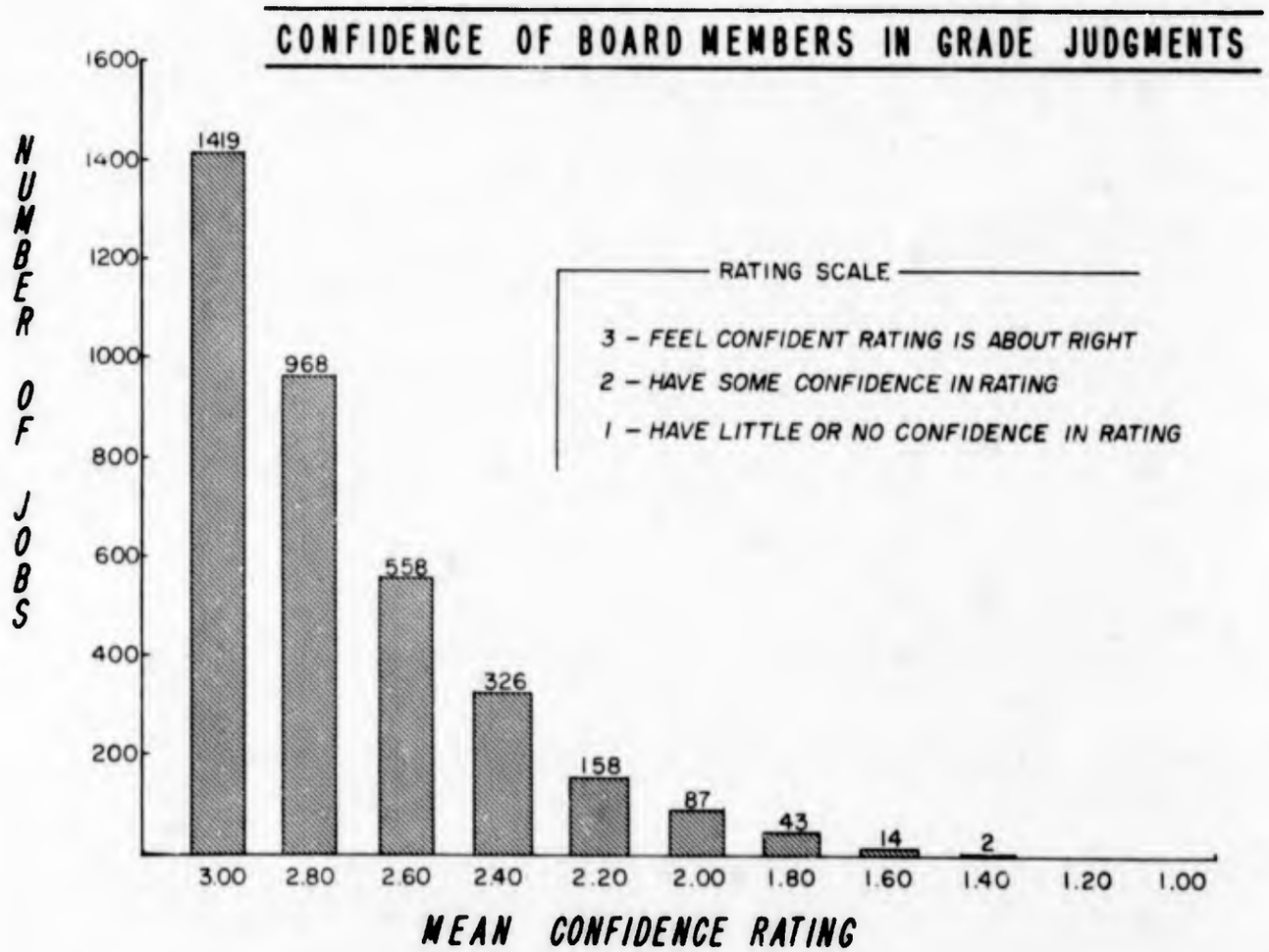


FIGURE 2 CONFIDENCE OF BOARD MEMBERS IN GRADE JUDGMENTS

Table 1

EXAMPLES OF CATEGORICALLY-CODED AND INTERACTION PREDICTORS

Applicant Number	Predictor Vectors X_1 Blue Eyes	X_2 Brown Eyes	X_3 Brown Eyes	X_4 Blonde Hair	X_5 (X_1 X_3)	X_6 (X_1 X_4)
1	1	0	1	0	1	0
2	0	1	1	0	0	0
3	0	1	1	0	0	0
4	1	0	0	1	0	1
5	0	1	0	1	0	0
6	1	0	0	1	0	1
.	.	.	.	.	.	.
.	.	.	.	.	.	.
.	.	.	.	.	.	.
N	1	0	1	0	1	0

Table 2

ACCURACY OF POLICY EQUATION IN PREDICTING GRADES AWARDED BY THE HQ USAF POLICY BOARD TO JOBS IN THE CRITERION SAMPLE

Degree of Accuracy	Cumulative N	Cumulative % N
Exactly on	1292	36.1
Within 1/3 of a grade	2931	82.0
Within 2/3 of a grade	3457	96.7
Within 1 grade	3557	99.5
Within 1 1/3 or a grade	3572	99.9
Within 1 2/3 of a grade	3575	100.0

Table 3. **Average Deviation of Each Board Member's Ratings by Job Category[b]**

Job Category[a]	Rater Number																					
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
All jobs	-0.4	0.9	-0.0	1.2	0.2	-0.2	-1.0	0.2	-0.5	0.6	1.0	-0.5	-0.2	-0.4	-0.4	0.3	0.4	-0.2	0.2	-0.1	-0.8	-0.2
SAC	-0.7	1.4	-0.2	1.1	0.2	0.2	-1.0	0.1	-0.2	1.7	0.9	-1.0	-0.1	-0.8	-0.7	0.3	0.1	-0.4	0.2	-0.2	-0.9	-0.3
TAC	-0.2	1.0	-0.1	1.4	0.6	-0.2	-1.1	0.1	-0.7	0.4	1.3	-0.8	-0.3	-0.7	-0.1	0.4	0.2	-0.2	0.6	-0.4	-1.0	0.1
PACAF, USAFE, USAFSS	-0.2	0.8	-0.2	1.3	0.1	-0.2	-0.8	0.4	-0.9	0.7	1.0	-0.8	-0.4	-0.3	-0.1	0.2	0.1	-0.1	0.2	-0.1	-0.6	-0.0
Hq Comd, Hq USAF	-0.4	0.8	0.3	0.8	-0.3	-0.3	-1.0	-0.2	-0.3	-0.1	0.7	0.1	-0.2	0.1	0.1	0.2	0.9	0.2	-0.5	-0.1	-0.4	-0.3
AFSC, OAR	-0.2	0.5	0.8	1.3	-0.1	-0.8	-1.4	0.1	-1.1	-0.2	1.1	0.8	-0.5	0.0	-0.5	0.2	1.5	-0.3	-0.0	-0.1	-0.6	-0.6
AFLC, AFAFC, AFCS	-0.3	0.2	-0.3	1.1	-0.1	-0.6	-0.7	0.5	-0.5	-0.1	0.9	-0.1	-0.0	-0.3	-0.5	0.3	0.8	-0.3	-0.1	-0.2	-0.3	0.5
AAC, SOU, ADC, CONAC	-0.1	0.8	-0.4	1.4	0.0	-0.1	-0.7	0.5	-0.7	0.5	0.9	-0.8	-0.3	-0.4	-0.2	0.1	0.4	-0.2	-0.0	-0.0	-0.7	-0.1
ATC, AU, USAFA	-0.4	0.6	-0.4	1.4	0.6	-0.6	-0.9	0.2	-0.7	-0.5	1.0	-0.1	-0.1	-0.7	-0.4	0.4	0.1	0.1	0.8	0.5	-0.8	-0.1
Other Commands	-0.4	0.9	0.2	1.2	0.3	0.1	-1.0	0.5	-0.3	0.4	1.0	-0.7	-0.4	-0.3	-0.4	0.4	0.5	-0.2	-0.1	-0.5	-1.3	-0.1
Pilot and Navigator	-0.4	1.7	-0.2	1.2	0.6	-0.3	-1.2	0.4	-0.4	1.0	1.3	-0.6	-0.3	-0.7	-0.3	0.7	-0.1	-0.1	-0.5	-0.6	-1.3	-0.2
Air Operations	-0.8	0.6	0.0	1.1	-0.3	0.4	-0.9	0.2	-0.3	1.6	1.0	-0.8	-0.3	-0.2	-0.3	-0.0	0.3	-0.3	-0.0	-0.5	-0.7	0.1
Scientific and Eng.	-0.4	0.8	0.3	1.3	-0.3	-0.4	-1.2	0.1	-0.6	0.6	0.8	0.0	-0.3	-0.4	-0.7	0.2	0.9	-0.3	0.2	-0.0	-0.5	-0.1
Materiel and Compt.	-0.1	0.6	-0.4	0.9	-0.1	-0.4	-0.8	0.4	-0.3	0.1	0.5	-0.8	-0.1	-0.2	-0.3	0.0	1.0	-0.1	0.2	0.2	-0.1	-0.3
Professional	-0.1	1.2	0.1	1.0	1.1	-0.6	-1.2	0.0	-1.4	-0.8	1.5	-0.6	-0.0	-0.4	-0.5	0.5	0.6	-0.4	1.0	0.0	-1.0	-0.4
Admin. and Support	-0.4	-0.1	-0.2	1.7	0.1	0.1	-0.3	0.1	-0.6	0.3	0.7	-0.5	-0.3	-0.8	-0.3	0.0	0.3	-0.1	0.8	0.5	-0.7	-0.3

[a] Pilot and Navigator AFSCs: 10XX, 11XX, 12XX, 13XX, and 15XX, except omit 1515.

Air Operations AFSCs: 14XX, 1515, 16XX, 17XX, 18XX, and 19XX.

Scientific and Engineering AFSCs: 25XX, 26XX, 27XX, 28XX, 30XX, 31XX, 32XX, 43XX, 47XX, 55XX, and 57XX.

Materiel and Comptroller AFSCs: 60XX, 63XX, 64XX, 65XX, 67XX, 68XX, and 74XX.

Professional AFSCs: 88XX, 89XX, 90XX, 91XX, 92XX, 97XX, and 99XX.

Administrative and Support AFSCs: 02XX, 03XX, 23XX, 70XX, 73XX, 75XX, 79XX, 80XX, 81XX and 82XX.

[b] The values in this table were computed by taking the difference between the average of ratings assigned by a rater (on the 16 point scale) to jobs in a particular category from an average of ratings assigned by all raters to jobs in that category. Since three points on the 16-point scale represents one grade level, a value of 3.0 in Table 1 would indicate that a board member rated jobs in a given category approximately one grade higher than other board members. Similarly, a value of -3.0 would indicate judgments averaging approximately one grade lower than those of other members. The highest reported value is only 1.7, and most of the values are less than 1.0. The largest values tend to be associated with judges who rated all categories somewhat high or low, and these judges did not show a bias toward jobs in particular categories.

Table 4

Definition of Variables Included in Officer Grade Requirement Policy Equation

Variable Number	Variable Name	Variable Definition	Source of Data
1	Management	The level of executive and managerial skills required by the job. The complexity, variety and level of the activities which are directed, organized, coordinated, controlled, commanded, or evaluated.	Mean ratings for each job obtained from five field judges
2	Planning	The extent to which planning is required by the job. The scope and significance of work for which planning is done. The longer the time span for which planning is done, the higher the rating.	Mean of ratings for each job obtained from five field judges
3	Special Training and Work Experience	The extent to which the job requires knowledges and skills which must be acquired through special training courses or on-the-job experience. Does not include general courses given by Squadron Officer School, Air Command and Staff College, or Air War College.	Mean of ratings for each job obtained from five field judges
4	Judgment and Decision Making	The importance and independence of Judgments and decisions required by the job. The nature, variety, and possible impact of decisions. The less well defined the guidance for decisions, the higher the rating; while the more specific and detailed the guidance, the lower the rating.	Mean of ratings for each job obtained from five field judges
5	Communications Skills	The extent to which the job requires skill in oral and written communication as well as the level of the individuals and agencies involved.	Mean of ratings for each job obtained from five field judges
6	Level of Organization in Which Job Occurs	DOD or Hq USAF = 9 Hq Major Air Command = 8 Numbered AF or equivalent = 7 Air Division or equivalent = 6 Wing or equivalent = 5 Group or equivalent = 4 Squadron or equivalent = 3 Detachment or equivalent = 2 Other = 0	Data from Job Description Form
7	Level of Job Within Organization	Command Element = 7 Directorate, Department, Office or equivalent = 6 Division or equivalent = 5 Branch or equivalent = 4 Section or equivalent = 3 Unit or equivalent = 2 Other = 0	Data from Job Description Form
8	Field Grade Rating	Rating of appropriate grade for job using the 16-point OGR Grade Rating Scale.	Mean ratings for each job obtained from five field judges
9	Supervisor's Grade Rating	Supervisor's rating of appropriate grade for job using a 7-point scale.	Data from Job Description Form

Table 5. **Example Results from Officer Grade Requirements Study[1]**

Navigator-Observer Jobs

(AFSC 15XX)

ROGR	UMD				
	Lt/Capt	Major	Lt Colonel	Colonel	Total
Colonel	--	--	--	--	--
Lt Col	21	60	12	--	93
Major	922	849	280	--	2051
Lt/Capt	7432	1180	593	--	9205
Total UMD (A)	8375	2089	885	--	11349
Total ROGR (B)	9205	2051	93	--	11349
Change (A-B)	-830	38	792	--	--

Multi-Engine Jet Jobs

(AFSCs 106XC, 123X with shreds)

ROGR	UMD				
	Lt/Capt	Major	Lt Colonel	Colonel	Total
Colonel	--	--	18	--	18
Lt Col	--	25	218	--	243
Major	677	493	525	--	1695
Lt/Capt	3444	796	133	--	4373
Total UMD (A)	4121	1314	894	--	6329
Total ROGR (B)	4373	1695	243	18	6329
Change (A-B)	-252	-381	651	-18	--

Operations Jobs

(AFSC 14XX)

ROGR	UMD				
	Lt/Capt	Major	Lt Colonel	Colonel	Total
Colonel	--	89	228	96	413
Lt Col	243	579	914	39	1775
Major	1678	1736	229	--	3643
Lt/Capt	1154	225	114	--	1493
Total UMD (A)	3075	2629	1485	135	7324
Total ROGR (B)	1493	3643	1775	413	7324
Change (A-B)	1582	-1014	-290	-278	--

[1] "ROGR" is the recommended Officer Grade requirements.
"UMD" is authorization Unit Manning Document.

Table 5 (*Continued*)

Scientific Jobs

(AFSC 26XX)

ROGR	UMD				
	Lt/Capt	Major	Lt Colonel	Colonel	Total
Colonel	5	6	3	--	14
Lt Col	76	60	24	--	160
Major	386	98	15	--	499
Lt/Capt	290	13	--	--	303
Total UMD (A)	757	177	42	--	976
Total ROGR (B)	303	499	160	14	976
Change (A-B)	454	-322	-118	-14	--

Development Engineering Jobs

(AFSC 28XX)

ROGR	UMD				
	Lt/Capt	Major	Lt Colonel	Colonel	Total
Colonel	--	30	15	1	46
Lt Col	240	298	89	1	628
Major	1736	413	24	--	2173
Lt/Capt	743	20	--	--	763
Total UMD (A)	2719	761	128	2	3610
Total ROGR (B)	763	2173	628	46	3610
Change (A-B)	1956	-1412	-500	-44	--

Research & Development Management Jobs

(AFSC 27XX)

ROGR	UMD				
	Lt/Capt	Major	Lt Colonel	Colonel	Total
Colonel	--	4	97	228	329
Lt Col	5	48	231	9	293
Major	27	55	26	--	108
Lt/Capt	7	4	--	9	20
Total UMD (A)	39	111	354	246	750
Total ROGR (B)	20	108	293	329	750
Change (A-B)	19	3	61	-83	--

(U) A STUDY OF JOB PREFERENCES OF GOVERNMENT AND NON-GOVERNMENT PERSONNEL

by

Robinette E. McCabe, Mgmt. Evaluation Specialist, USAF

Air Force Plant Representative Office
Lockheed Missiles & Space Company
Sunnyvale, California

Robinette E. McCabe

BIOGRAPHY

Robinette E. McCabe

A career civil service employee, Mrs. McCabe began working for the federal government during World War II at, what was then, Wright Field, Dayton, Ohio. She began as a clerk-steno in 1944 when her husband was stationed there as a Lieutenant in the Army Air Corps. Subsequent Air Force positions included being secretary to General Nathan F. Twining and Lt. General William E. Kepner.

During five years spent in Alaska following World War II, she began free-lance writing and has had numerous articles published in engineering, architectural, and construction magazines. A six-year period with the USAF Installations Representative Office, South Pacific Region, in San Francisco, broadened her writing experience in this field.

For the past three years, she has been with the USAF Plant Representative Office, Lockheed Missiles and Space Company, Sunnyvale, California, performing management analysis work. Her present position is that of Management Evaluation Specialist with the office's Industrial Management Evaluation Division.

An enthusiastic believer in life-long education, she has continually sought to up-date her skills. Attending college principally in the evenings, she earned an A.A. degree (1961) in mathematics and science from San Mateo College, and a B.A. degree (1963) and M.S. degree (1964) from California State College at San Jose, California. She also completed the requirements for the Junior College Teaching Credential (1966). Her next goal is a doctoral degree in Business Administration, and she has been enrolled for the past year and a half in the University of Santa Clara's evening program, with this objective in mind.

This fall she returned to San Mateo College not as a student, but as a member of the faculty, and is teaching two courses in the Evening College.

ABSTRACT

(U) A STUDY OF JOB PREFERENCES OF GOVERNMENT AND NON-GOVERNMENT PERSONNEL

Executives in government, business, and industry are interested in increased work output, improved quality, and decreased costs. Managers have become increasingly aware that achievement of these goals is dependent upon effective motivation. This research suggests that motivating incentives vary in effectiveness and preference from one group to another, and that preferences can be shown to be related to age, sex, education, marital status, occupation, and employer.

One hundred forty-five USAF officers and civil service employees of the San Francisco Bay Area, and two hundred employees of the Lockheed Missiles and Space Company, participated in the research. Government men differed from non-government men on all incentives except advancement. Government males rated pay, type of work, and "employer" higher than did the non-government men. Contrary to expectations, they assigned a lower value to security and benefits. Hours and working conditions were also seen as less important to government men.

Sharp differences were encountered between male and female preferences. Security was given top place in both government and non-government groups of women. Of less importance to the female than to her masculine work-mates are pay, advancement, and working conditions. The incentive which singularly revealed the greatest difference in male and female preferences was type of work. Although males gave it first place, it was of only median concern to the female.

The most noticeable changes appeared as a function of education. Pay, advancement, and type of work all gained marked importance as the educational level went up.

(U) A STUDY OF JOB PREFERENCES OF GOVERNMENT AND NON-GOVERNMENT PERSONNEL

Executives in government, business, and industry are interested in increased work output, improved quality, and decreased costs. In recent years, managers have become increasingly aware that achievement of these goals is dependent upon effective motivation. The basic problem is to find an incentive, or combination of incentives, capable of getting employees to participate willingly and fully in furthering the goals of the organization. Management generally practices the philosophy that increased pay elicits increased production, efficiency, and identity with the organization. Application of this philosophy has resulted, among other things, in the expenditure of much time, effort, and money, in the development of complex wage-incentive systems.

Certainly the potential of the financial reward as an incentive should not be underestimated. Day-to-day experience in business, government and in industry, as well as investigations bearing upon wage-incentive plans, show that the pay incentive can often serve to increase earnings, raise production, and reduce costs. However, in no instance have experimental data been presented to support the implication that the pay envelope is the primary incentive for all individuals regardless of sex, intelligence, age, educational level, occupation, etc. Conversely, there is considerable evidence which suggests that these characteristics probably affect not only the importance an individual will place on pay, but the relative importance of various other incentives as well.

We are all familiar with the sharp difference in preferences that people exhibit in their free time. Some like to fish, others spend long hours maintaining a boat, racing car, or hunting equipment. Still others may prefer to paint or work in their backyard. Some pursue a higher education, while others spend their free hours in community activities, clubwork, or in a variety of other ways. We have all marveled at the great lengths to which people will go, and the unlimited amount of energy and money they will spend, in pursuing their off-the-job preferences. It is the thesis of this paper that people have equally distinct, and individually motivating preferences on the job. The need for accurate information relative to incentives that can meet the needs and interests of individuals working together is central to the problem of enlarged productivity, improved quality, reduced costs, and job satisfaction.

BACKGROUND OF THE PROBLEM

There is little in the literature dealing with experimental investigations of employee wants or preferences prior to the 1930's. Since that time, although the total number of studies is extremely small, when compared with many other areas of investigation, the number of studies has been increasing. The probable explanation of this is the changing philosophy of American management. However grudging, there has been an increased tendency on the part of management to admit that employees should be treated as human beings, that is to say, presumably, as individuals with dignity, aspirations, ideas, and preferences of their own.

Pressures that facilitated the change in industry have been labor shortages, governmental intervention in labor-management relations, the increased strength of unions, increasing size of the business enterprises, and separation of ownership and management which has evolved a professional managerial group. Difficulties experienced by this new management in tapping the 'will-to-work', coupled with a growing awareness that full utilization of human resources can be essential to an organization's success, have provided the climate for the increasing attention being given to employee wants and preferences.

A review of the literature on this subject is beyond the scope of this paper, but mention of selected investigations is necessary to provide adequate perspective for this work. Working independently in the 1930's, Chant in the United States and Wyatt and Langdon in Great Britain, conducted studies of job incentives which yielded remarkably similar results. Chant (1932) found credit for all work done, and interesting work, as being more important than pay, and gave job security as being the least important of all.

Centers (1945) obtained somewhat similar data on job security when he took a sampling of the complete occupational stratification of adult white males. In asking those that were satisfied with their work what it was that they liked about their jobs, he found that only slightly more than six percent of the men sampled replied "security, steady work."

An investigation by Wilkins (1949) lends support to the suggestion that educational background, or level of intelligence, may affect attitudes toward incentives. He compared rank-order ratings of eight incentives with intelligence measures of the young men making the ratings. The more intelligent were found to be more interested in prospects, security, variety, and efficient organization, while the less intelligent were more interested in work-mates, hours, pay, and leave.

One limitation of Wilkins' work is that his study was conducted in England; thus generalization of his conclusions to the American situation may not be appropriate. However, Jurgensen's work (1947; 1948; 1949; 1961) does not suffer from this limitation, and it, too, implies that educational level is one of the prime factors affecting job preferences.

Jurgensen's first study, which appeared in 1947 and involved an analysis of the job preferences of 1189 men and 150 women who were job applicants of the Minneapolis Gas Company, concluded that:

> "... Job preferences are affected more by extent of education than by most of the other variables. Advancement, type of work, pay, and working conditions became more important as extent of education increases, and security, company, coworkers, supervisors, hours, and benefits became less important. Changes are not always gradual, the points of high school graduation and college attendance being particularly important. ..."

For this work, Jurgensen devised the Jurgensen Job Preference Blank, a brief questionnaire listing ten employee incentives and involving the rank-order technique. Classification information on age, sex, marital status, dependents, monthly salary, education, and main occupation was also requested on the questionnaire, and the data obtained was analyzed in terms of these various sub-categories.

During the past 19 years, Jurgensen has continued his research with the JJPB by conducting one of the longest investigations of job preferences on record. In 1961 he reported on the trends he had observed during the previous 15 years and noted that the pattern as a whole had remained remarkably consistent. For example: Differences between men and women remained clear during the entire period.

Jurgensen concluded that job preferences have important implications in selecting, training, and supervising employees, and that the information may be used even more valuably in determining personnel policies of an organization, and in conducting union negotiations.

Several other investigators, among them Thompson and Goad (1949), and Hardin, Reif, and Heneman (1951), have used the JJPB to study job preferences.

OBJECTIVES OF THE STUDY

The purpose of this research was to investigate the importance of various employee incentives, using the Jurgensen Job Preference Blank (JJPB) as the measuring device. See Figures 1 and 2 for a copy of the questionnaire. The main objectives were: (1) to present the relative ratings, given the incentives listed on the JJPB, by the individuals selected for study; (2) to show the relation between ratings and such factors as age, sex, educational level, number of dependents, salary, employing agency, and occupation; (3) to compare ratings given by individuals concerned in this study, all employed at the time of the study, with ratings obtained by other investigators administering the same questionnaire to unemployed and/or other job applicants; and (4) to investigate the reliability of the JJPB, i.e., the consistency or stability of responses when individuals were asked to repeat their preferences after a given interval of time.

Four hypotheses were put forth to assess the variables of the study:

1. It was expected that there would be a significant difference between the ratings made by personnel employed by the government and non-government employees, with the latter assigning higher values to "Pay" and government employees assigning higher values to "Security" and "Benefits." Believing that people tend to emphasize what they believe they possess (or, at least what they hear most about), it was reasoned that this difference would be found because government personnel are continually confronted with communications emphasizing security and benefits.

2. Despite a possible difference in the woman's role in the highly-industrialized site selected for this study and the midwest and southern locales where the JJPB has been tested previously, it was expected that there would still be significant differences in the way women rate the incentives and the way men rate them. That is, it was expected that the results obtained in this study would corroborate the findings of previous investigators using the JJPB relative to differences between women and men. Even though women have entered the labor market at increasing rates during the past few years, it was reasoned that women, in general, still have only short-range work plans, and, as such, are not as interested in the long-range work factors as are men.

3. It was anticipated that there would be material differences between the men and women involved in this study, all employed at the time of the study, and the thousands of job applicants involved in the Jurgensen (1961) summary. Because of the higher level of work skills sampled in this study, and the difference in work status, it was expected that Security would get a lower rating than observed in Jurgensen's work.

4. It was presumed that individuals engaged in different functional areas of work would rate incentives differently, with professional and managerial people placing greater emphasis on pay and security than those engaged in clerical, manual, sales and technical work. It was reasoned that these individuals, as evidenced by their climb to their present positions, would have internalized the upper-middle-class values (which emphasizes pay and security) to a greater degree than individuals engaged in clerical, manual, technical, and sales work, and that these values would be reflected in their higher ratings of pay and security.

DESIGN JUSTIFICATION

I. SELECTION OF THE MEASURING DEVICE

Previous investigators have sought to discover what the worker wanted from his job by three different methods: (1) Multiple-item inventories, or questionnaires, have been used. From a statistical analysis of the data, it has been possible to derive factors whose content could be identified from a comparison of the interrelationships among the items. (2) Employees have been interviewed, and asked what they liked or disliked about their jobs. This method usually involved some form of weighting of the intensity of the like or dislike. (3) A previously selected list of incentives (either positive or negative) have been presented to workers, who have been asked to study the list and rate the factors as to desirability.

It is worthy of note that a number of investigators have concluded that, although the methodology is different, the factors that emerge are essentially the same. Among those taking this position have been Viteles (1953), Herzberg, et al. (1959), and Gilmer (1961). Thus, it appears possible that the incentives listed on the Jurgensen Job Preference Blank are essentially the same, or have an equivalent relationship, to the positive incentives identified by other investigators.

The paired-comparison, or forced-choice, technique was considered, but it was rejected because, in order for all possible comparisons to be made with ten factors, 45 pairs would have to be considered by the rater. This would not only be time-consuming, but confusing to the participant as well. Consequently, the rank-order presentation of items appeared to be the most advantageous technique for this investigation.

The limitations of the design are acknowledged, however. It is readily apparent that the differences between ranks do not indicate differences of equal magnitude. However, since information on the relative value of the incentives is needed for practical application (i.e., the absolute value of an incentive does not necessarily have to be obtained), the methodology and measuring device used by Jurgensen appeared to be suitable for this research, also.

One other item prompted selection of the JJPB. A number of investigators have offered either objective evidence or subjective opinion relative to the attitudes of government employees, and the writer has been particularly interested in obtaining valid data on differences. Hardin, et al (1951) used the JJPB to study the stability of job preferences of midwest department store employees and indicated that the JJPB possesses the capability for reliably measuring organizational preferences. Therefore, the JJPB was chosen as a device capable of measuring any organizational preferences that might exist between government and non-government personnel.

II. GROUPS STUDIED

Non-Government personnel. Two hundred employees of the Lockheed Missiles and Space Company (LMSC), located at Sunnyvale, California, were asked to participate in the study. Employing slightly over 30,000 people at the time of the study, this Company had approximately 21,000 people located at their Sunnyvale facility. What resulted as representative of LMSC were 199 useable questionnaires, 170 from men, and 29 from women. Age range varied from 9 individuals who were less than twenty-four years of age, to 18 individuals 50 years of age or more.

Government personnel. These individuals were from the Air Force Plant Representative Office, Lockheed MSC, Sunnyvale, and Air Force Regional Civil Engineers' Office, South Pacific Region, San Francisco. It was felt that the personnel of these offices could be combined for this research inasmuch as they had comparable types of manpower. Homogeneous personnel skills were involved; both offices were staffed by management, engineering, technical, and clerical personnel. The entire staff of both offices participated in the study, making a combined total of 145 individuals in the government group. The percentage of females encountered was noticeably higher in the government offices than in the industrial sample (of the 145 individuals, there were 36 females and 109 males in the government group). Accordingly, 25% of the government people were female, versus 15% of the non-government group. There were no material differences in age, however.

.....

JURGENSEN JOB PREFERENCE BLANK

(What makes a job good or bad?)

Decide which of the following is most important to you and place 1 on the line in front of it. Then decide which is second in importance to you and place a 2 in front of it. Continue listing the items in order of importance to you until the least important is ranked 10.

All the items are important, but people differ in the order in which they rank them. Answer according to how you think, not how you believe others think.

_____ ADVANCEMENT (Opportunity for promotion).

_____ BENEFITS (Vacation, sick pay insurance, etc.).

_____ COMPANY (Employment by a Company/or Government organization/that you are proud to work for).

_____ CO-WORKERS (Fellow workers who are pleasant, agreeable, and good working companions).

_____ HOURS (Good starting and quitting time, good number of hours per day or week, day or night work, etc.).

_____ PAY (Large income during year).

_____ SECURITY (Steady work, no lay-off, sureness of being able to keep your job).

_____ SUPERVISION (A good boss who is considerate and fair).

_____ TYPE OF WORK (Work which is interesting and well liked by you).

_____ WORKING CONDITIONS (Comfortable and clean; absence of noise, heat, cold, odors, etc.).

BE SURE AND FILL IN OTHER SIDE

Figure 1.

The following data about yourself are desired for research purposes:

1. SEX:

_____ Male
_____ Female

2. MARITAL STATUS

_____ Single
_____ Married
_____ Widowed
_____ Divorced
_____ Separated

3. DEPENDENTS (besides yourself)

_____ None
_____ One
_____ Two
_____ Three
_____ Four
_____ Five
_____ More than five

4. AGE (of self)

_____ Under 20
_____ 20-24
_____ 25-29
_____ 30-34
_____ 35-39
_____ 40-44
_____ 45-49
_____ 50-54
_____ 55-59
_____ 60 or Over

5. MONTHLY SALARY:

_____ Less than $200
_____ $200 - $299
_____ $300 - $399
_____ $400 - $499
_____ $500 - $599
_____ $600 - $699
_____ $700 - $799
_____ $800 - $899
_____ $900 - $999
_____ $1,000 or over

6. EDUCATION:

_____ 8th grade or less.
_____ Some H.S., but not completed.
_____ H.S. or vocational school diploma.
_____ Diploma plus technical or business school.
_____ Some college, but not completed.
_____ College or University degree.
_____ Advanced University degree.

7. MAIN OCCUPATION _____

Please check the functional area in which you feel your occupation fits:

_____ Clerical
_____ Manual
_____ Sales
_____ Technical
_____ Professional
_____ Manager*

*For the purpose of this research, a manager is defined as one who supervises one or more supervisors (that is, a manager is one level above the first-line supervisor). If you are a manager, and your occupation also fits in one of the other functional areas, please check BOTH blanks.

Figure 2.

III. ADMINISTRATION OF QUESTIONNAIRE

As was mentioned previously, the Jurgensen Job Preference Blank (JJPB) was used to obtain data. The basic question that was asked of the participants was, "What makes a job good or bad?" The front side of the form (see Figure 1) requested the participants to rank-order the incentives.

The reverse side of the questionnaire (see Figure 2) asked the rater to fill in data relative to sex, marital status, dependents, age, monthly salary, education, and main occupation.

Participants were not asked to put their names on the forms since it was felt that if they were not asked to identify themselves, more valid data would result. Raters rendering what they feel to be the 'expected reply' rather than their own personal evaluation is a commonly-reported phenomena in the behavioral sciences, and it was hoped that this procedure would elicit truer evaluations. However, it was clear to the investigator that omission of data on individual reliability would be sure to introduce difficulties in the interpretation of findings, and data on reliability could only be obtained through a retest. Therefore, the questionnaires were numbered, and as each questionnaire was returned to the investigator by the participants, the participant's name was recorded on a list opposite the appropriate JJPB number. This posed only minor difficulty for all personnel tested at the LMSC Sunnyvale Complex wore security badges in plain sight, and all personnel tested in San Francisco were known personally by the author.

Two weeks after the initial ratings had been obtained, a random sample of 60 individuals, drawn from the 344 participants, were asked to fill out the questionnaire again (see Figure 3). This constituted the retest, for individuals were asked to rate the same incentives again in order of personal preference.

First Administration	Time Lapse	Follow-up Administration
Initial test. Three hundred forty-four (344) people were asked to rate the incentives listed on the JJPB.	Two weeks	Retest. Sixty (60) individuals, randomly drawn from the total group of 344, were again asked to rate the incentives listed on the JJPB.

Figure 3. Block Diagram of Research Design.

Of the 200 questionnaires filled in by Lockheed employees, 199 proved to be useable (one person checked all of the incentives 1, and stated that he could not put them in any meaningful rank order). However, two people who were asked to participate in the study refused to do so.

Administration of the forms began in March and was completed in April 1964. The data were placed on IBM cards and a Control Data Corporation computer was used to compute the Means and Standard Deviations for the sub-groups studied. Since a total of 2,100 Means and Standard Deviations were involved, this appeared to be the only practical method of handling the data.

Further statistical analysis was considered for the mean ranks. Since it was apparent from the design of the study that unequal subgroups would be involved, and since a number of authorities (Dixon & Massey, 1957; Li, 1961; Edwards, 1960) have pointed out that it is not proper to assume the underlying variance is equal if the sample sizes are not equal, it was evident that the underlying variance would have to be tested prior to investigating mean-rank differences.

The F-test was used to analyze whether the underlying variances were equal. Where the variances were found to be statistically different, no further test was required to conclude that the sub-groups were really different relative to the incentive being rated. Where the hypothesis that the variances were equal was accepted, the variances were pooled to provide a "weighted" variance and the differences between the means were tested. The 95% level of confidence was chosen for testing the significance of the differences. For further details on the statistical treatment, see Appendix E.

RESULTS

Presentation of all statistical data that resulted from this study is beyond the scope of this paper. However, group and sub-group means in support of the major trends and differences that will be discussed appear in the Appendixes. The reader's attention is invited to the fact that the incentive considered to be the most important was given a rank of 1, and the least important was given a rank of 10. Thus, the smaller the Mean that is shown, the more important the factor was considered to be.

Government personnel versus private industry employees. Some surprising results were obtained between the groups. Male preferences, which proved to be statistically different on all ten incentives, are shown in Appendix A. The female differences failed to reach statistical significance (perhaps because of the small number of individuals involved), but they followed the same pattern as the men. Generally, it can be said that the data indicates government personnel rate pay, type of work, and company higher than industrial employees. Contrary to popular opinion, they assign a lower value to benefits. Hours and

working conditions are also seen as less important by government employees. The only incentive on which the males did not differ was advancement. In rank order preference, this was seen as the second most important incentive for both groups. Type of work was the most important factor so far as men were concerned in both government and non-government work situations.

Male versus female preferences. Women in both the government and industrial groups gave security top place. Differences on co-workers, advancement, pay, security, supervisor, type of work, and working conditions, all reached statistical significance at the 95% level of confidence. Of less importance to the female than to her masculine working companions are pay, advancement, and working conditions. Of more importance to her are hours, co-workers, and security. The most significant difference between the two groups was encountered on type of work. For the male, type of work was in first place, but it only received fourth place rating from the women.

Educational differences. Noticeable changes occur with an increase in education. If the reader will follow in a linear manner through the various educational-level sub-groups shown in Appendix C, the differences will be apparent for several of the incentives. Advancement, pay, type of work, and security show the effect. Mean ranks for advancement, type of work, and pay move from a median or low preference for those with some High School to a very high preference for those individuals with an advanced degree. However, there is a decreasing value placed on security with an increase in educational level.

The preference for security shown at the lower educational level may explain the high value placed on security by the women in this study. Since the educational level in the female group did not climb proportionately to that found in the male group, the high value placed on security by women may not be as much a function of sex as it is a function of educational level. Further research, employing a factor-analysis type of design, should be capable of isolating the determining factor and answering the question.

Reliability of the Measuring Device.

Individual reliability. Individual retest results ranged from a positive correlation of 1.0 to a zero correlation. Spearman's coefficient of rank-order correlation (Rho) was used to study the individual measure. In considering the distribution of the correlation coefficients over the range encountered, it was determined that mode correlation was less than 0.50, whereas, in the case of ten ranks, a coefficient of 0.63 is needed at the 5% level of confidence. In considering the distribution of the ratings, it appeared that the greatest lack of reliability came from the fourth through the tenth rank. The first three preferences given by the participants generally showed a high degree of consistency.

Group reliability. The reliability of the instrument in measuring

group preferences was also investigated. A comparison was made of the mean ranks resulting from the initial administration of the JJPB, and the mean ranks resulting from the retest. Here a more encouraging picture was encountered. The Means and Standard Deviations obtained from the initial rankings and the retest rankings were tested for statistical difference. Only one incentive, Benefits, proved to be different, at the 95% level of confidence. These results suggest that, while the individual retest reliability of the instrument is low, the instrument does reliably measure the means of groups. That is, the instrument's reliability relative to the measurement of group preferences is relatively good.

It should be noted that the reliability results of this study replicate those of Hardin, Reif, and Heneman (1951), who concluded that group means obtained from the JJPB are reliable, but that the individual-retest reliability was low. Jurgensen's 1961 summary of the stability of JJPB medians over a 15-year period further indicates that the JJPB is not only a reliable measurement of group means relative to the job preferences, but that these group preferences remain stable over a rather long period.

DISCUSSION

The failure to show industrial employees placing a higher value on pay and government workers placing a higher value on security and benefits, required rejection of the first hypothesis. The data showed that the government workers of this study valued pay, type of work, and company a little more than industrial employees. They assigned a lower value to benefits, and a lower value to security. Thus, the rationale that was behind the first hypothesis, i.e., that people tend to emphasize what they have, or those factors that they hear most about, must be rejected. The lower preferences indicated by the government workers for benefits and security may result from the government workers feeling that they now have these incentives. This may result from a tendency not to rate as highly preferred something which they feel is already "built-into" their job. However, this is only one possible explanation.

The second hypothesis, that there would be significant differences in the way women rated the incentives and the way men rated them, was accepted. Women reported preferences higher than those expressed by men on hours, co-workers, and security. Men reported higher preferences on pay, type of work, advancement, and working conditions. The most outstanding difference in the mean ranks was on type of work, with this being of primary concern to the male, but of only median concern to the female.

On the basis of other studies, most of these differences between the sexes were expected. However, other differences expressed were surprising. The very low ratings given working conditions and hours were somewhat different than was anticipated from the women. Working conditions and hours appear to be not nearly as important to women as the

emphasis that is generally placed on them. These findings are in direct contrast to the popularly expounded belief that women are particularly interested in short-range or temporary factors which increase the pleasantness of work -- whereas the male is given credit for making long-range work plans, and being much less interested in working conditions.

Also, the women encountered in this study were as equally desirous of "getting ahead" as the male, for advancement was in second place in both groups.

The findings of this study were also different from the findings of other studies relative to female preferences (Jurgensen, 1961; Thompson and Goad, 1949; Hardin, et al, 1951). For example, Jurgensen reported that women in the midwest locale rated type of work as most important, whereas women in this study rated security as the most important, with type of work being fourth in rank-order presentation.

Some of this difference might be cancelled by a comparison of the means. Inasmuch as Jurgensen used medians in his 1961 summary, such a comparison is not possible. Only a rank-order comparison can be made, and different ranks can sometimes be assigned to two means having no statistically-significant difference. Conversely, the same rank can be given to means which, when tested, prove to be statistically different.

Despite this limitation, some valid conclusions can be drawn from a comparison of the results of both studies (see Appendix D). For example, in both studies working conditions and hours were seen as being of very low importance to all participants. Interestingly enough, advancement was seen as being of about the same importance in the midwest as in the San Francisco Bay Area (it was in second place in both studies). Pay, however, appears to have a much higher value for workers on the West Coast than it has in the midwest. In the San Francisco Bay Area, both men and women placed pay in relatively high preferences (third place). In contrast, it received a relatively low rating in the midwest (6th place for men and 7th place for women). These differences are great enough to warrant serious consideration despite the mean-median limitation on comparison mentioned above.

The third hypothesis predicted that security would get a lower rating in this study than it received in Jurgensen's work. There are material differences in the mean ranks obtained in this study and Jurgensen's median ratings published in 1961 and security is one of the most baffling switches. Men in his summary of 24,343 preferences listed security in first place; men in this study gave it fourth place. Women in the Jurgensen summary gave security second place; women in this study gave it first place. Thus, the third hypothesis appears to be true for men, but must be rejected in terms of female preferences.

It was reasoned that the higher level of occupational skills, and the obvious difference in job status between job applicants and on-board employees, would result in the individuals in this study possessing a

greater feeling of security and thus considering it of less importance in terms of preference. That is, it was felt that the individuals in this study would not be as anxious about security, and would not be considering it in as desirable a frame-of-reference, and would thus not give it as high a rating on the job preference choices.

An environmental variable entered this study which may have affected the rating given to security. The mean rank for all females on this incentive was derived from both government, and industrial employees of Lockheed Missiles and Space Company. An inspection of the data showed that the highest preference for this incentive came from the industrial group. At the time of the study, Lockheed Missiles and Space Company was undergoing a reduction and there was much indication that this reduction was affecting, to a great extent, the lower occupational groups which were heavily populated with females. On one occasion when the writer handed the JJPB to a woman and explained the purpose of the study, the woman responded with the comment, "I can tell you what every woman in this place wants first right now, and that's security." An inspection of the JJPB (see Figure 1) will show that the incentive security if further defined with, "steady work, no lay-off, sureness of being able to keep your job."

The clear trends observed in technical, professional and managerial personnel preferences show that these are distinct groups. The trend of their preferences was in the direction predicted by the fourth hypothesis with respect to pay. The importance of advancement, as well as the importance of pay became greater for professional men than it was for those engaged in technical areas, and it was most important of all to managers.

Hypothesis number four cannot be accepted in its entirety, however, because the hypothesis included security as well as pay. It was contended that upper-middle-class values emphasize pay and security. There is much clear evidence from this study that security is valued to individuals engaged in lower occupations, and it is also clear that they place a higher value on security that do professional and managerial personnel. It must be then, that upper-middle-class values emphasize pay and advancement -- and that security is viewed as a corollary of pay and advancement. In any event, it appears not to be as potent an incentive in the upper groups as type of work, pay and advancement. Yet the fact that it is in one of the first four places in all rank-order ratings published to date, including this study, shows that it should still be given an important place in all managerial considerations and personnel planning.

This study corroborates the findings of Hardin, et al (1951) that the JJPB is capable of differentiating broad sub-group preferences and organizational preferences, with considerable reliability. That is, it has differentiated between the sub-groups of this study to a statistically significant degree. Thus, it is felt that the JJPB has potential as a management tool. What an employee can and will do depends on the extent to which his needs and wants are satisfied by management. Whether a Department of Defense organization, or a private company, if management

knows its organizational preferences, and can adequately fulfill the wants of its employees, it has an enormous advantage over its competitors, who fail to do so.

The writer must agree that individual reliability is highly desirable. However, individual reliability is only one measure of an instrument's value. The ultimate value of an instrument is its validity, i.e., how well it measures what it is designed to measure. Without further research, this question cannot be settled relative to the JJPB. However, it is believed that the JJPB's reliability in measuring group differences, and organizational preferences, has been sufficiently replicated to warrant further use of the instrument as a measure of these.

Admittedly, however, the JJPB is only an indirect measure of motivation, with all data limited to:

(1) What individuals were consciously aware of as motivators, and

(2) The degree to which the rater may choose to reveal his preferences.

REFERENCES

Blum, M. L. and Russ, J. J. A study of employee attitudes toward various incentives. Personnel, 1942, 19, 438-444

Brayfield, A. H. and Crockett, W. H. Employee attitudes and employee performance. Psychol. Bull., 1955, 5, 396-424

Centers, R. Motivational aspects of occupational stratification. J. soc. Psychol., 1945, 38, 187-217

Chant, S. N. F. Measuring factors that make a job interesting. Personnel Journal, 1932, 11, 1-4

Dixon, W. J. and Massey, F. J. Jr. Introduction to statistical analysis. New York: McGraw-Hill, 1957

Edwards, A. L. Experimental design in psychological research. New York: Holt, Rinehart, and Winston, 1960

Gilmer, B. Industrial psychology. New York: McGraw-Hill, 1961

Goad, R. O. and Thompson, A. S. What workers want from their jobs. Employment Security Review, 1949, 16, 11-13

Hardin, E., Heneman, H. G. Jr., and Reif, H. G. Stability of job preferences of department store employees. J. appl. Psychol., 1951, 35, 256-259

Herzberg, F., Mausner, B., and Snyderman, Barbara The motivation to work. New York: John Wiley, Inc., 1959

Herzberg, F., Mausner, B., Peterson, R. O., and Capwell, Dora F. Job attitudes: review of research and opinion. Pittsburgh: Psychol. Svs. of Pittsburgh, 1957

Jurgensen, C. E. Trends in job preferences over a 15-year period. Paper read at American Psychol. Assoc. Meeting, New York, Sept 1961

Jurgensen, C. E. Selected factors which influence job preferences. J. appl. Psychol., 1947, 31, 553, 564

Jurgensen, C. E. What do job applicants look for in a company? Personnel Psychol., 1948, 1, 433-445

Jurgensen, C. E. What do job applicants want? Personnel, 1949, 25, 352-355

Li, J. C. R. Introduction to statistical inference. (3rd ed.) Michigan: Edwards Bros., 1961

Likert, R. Motivation: the core of management. American Management Association Personnel Series No. 155, 1953, 3-21

McCabe, Robinette E. A Comparative Investigation of Job Preferences of Selected Sub-Groups of Employees. Unpublished Master of Science thesis presented to faculty of San Jose State College, June 1964

Wyatt, S. and Langdon, J. N. Fatigue and boredom in repetitive work. Industrial Health Research Board Report No. 77, 1937, 43-46

Viteles, M. S. Motivation and morale in industry. New York: Norton Company, 1953

APPENDIX A

MEAN RANKS ASSIGNED JOB PREFERENCE FACTORS BY MALES WORKING FOR GOVERNMENT VERSUS PRIVATE INDUSTRY

	GOVT		NON-GOVT			
	Mean	Std Dev	Mean	Std Dev	F-Value	T-Value
Advancement	3.18	1.95	3.52	2.24	1.308	-1.810
Benefits	6.44	1.94	5.94	2.49	1.655*	
Company	5.68	2.78	6.28	2.92	1.103	-2.442**
Co-Workers	6.74	2.46	6.79	2.28	0.855	-0.225**
Hours	8.28	1.81	7.30	2.33	1.664*	
Pay	3.51	2.33	3.92	2.47	1.121	-2.490**
Security	4.94	2.60	4.66	3.08	1.399*	
Supervisor	5.65	2.18	6.29	2.29	1.102	-3.370**
Type of Work	2.94	2.42	3.48	2.64	1.186	-2.511**
Work. Conditions	7.58	2.20	6.81	2.33	1.118	4.044**
NUMBER	109	170				

CRITICAL REGIONS (for $\alpha = .05$)

$F > .0250\ (169, 108) = 1.340$, or $F < .9750\ (169, 108) = 0.830$

For 277 degrees of freedom, $t > \pm 1.970$

* Value obtained indicates there is a significant difference in the underlying variance.

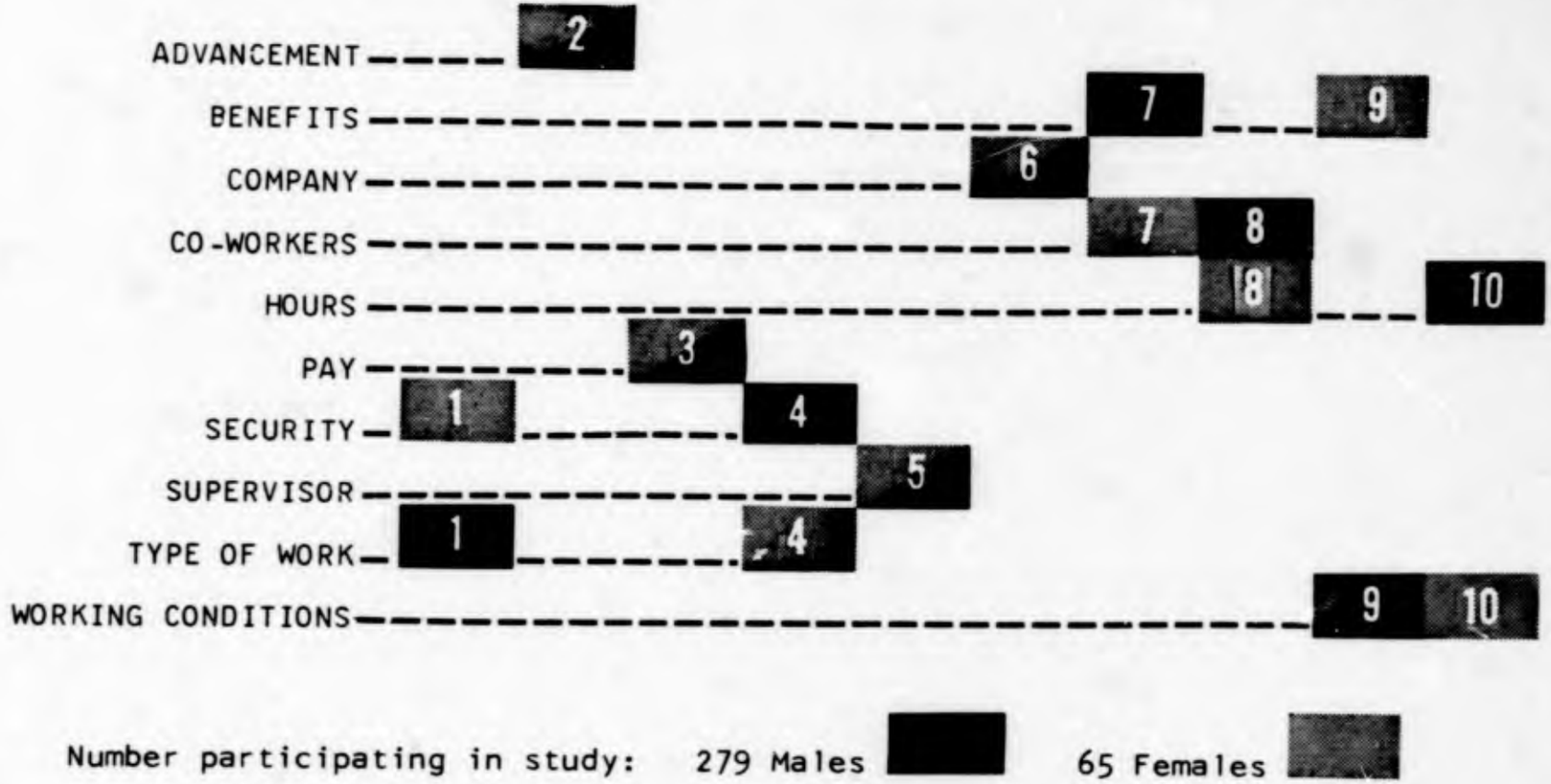
** Value obtained indicates there is a significant difference in the two Means relative to the incentive indicated.

APPENDIX B

MALE vs FEMALE PREFERENCES

JOB PREFERENCE FACTORS - RANK ORDER OF MEANS

WHERE 1 EQUALS HIGH PREFERENCE AND 10 EQUALS LOW PREFERENCE



APPENDIX C

EFFECT OF EDUCATION ON JOB PREFERENCES

MEAN* RANKS ASSIGNED JOB PREFERENCE FACTORS

*Where 1.0 equals High Preference and 10.0 equals Low Preference

	GOVERNMENT PERSONNEL[1]						NON-GOVERNMENT PERSONNEL[2]					
	8th Grade, some High School	High School Diploma	Diploma, plus Tech School	Some College, but Not Completed	College or University Degree	Advanced Degree	8th Grade, some High School	High School Diploma	Diploma, plus Tech School	Some College, but Not Completed	College or University Degree	Advanced Degree
[redacted]	5.0	3.8	3.8	3.0	2.9	2.1	3.9	3.5	3.8	3.8	3.4	2.3
[redacted]	5.0	5.7	6.1	6.0	6.5	6.8	5.3	5.0	5.7	6.3	6.9	7.3
COMPANY	2.0	6.2	4.7	5.7	5.9	5.1	5.7	6.6	5.9	6.2	6.8	5.0
CO-WORKERS	2.9	6.2	4.7	5.7	5.9	5.1	5.7	6.6	6.9	6.2	6.8	5.0
HOURS	9.0	8.2	8.7	7.9	8.4	8.4	8.2	6.8	7.1	7.2	7.3	8.0
[redacted]	10.0	6.0	5.3	4.4	4.3	3.1	5.6	3.8	5.1	3.7	2.9	2.5
SECURITY	1.0	4.7	6.3	4.6	4.7	6.7	4.4	3.8	3.5	4.1	5.8	8.4
SUPERVISOR	4.0	4.5	4.4	5.9	5.7	6.6	6.4	7.0	6.4	6.7	5.5	5.0
[redacted]	7.0	3.5	4.3	3.1	2.6	2.1	4.4	4.7	4.2	2.9	2.6	1.4
WORKING CONDITIONS	6.0	8.7	6.3	8.0	7.5	6.9	6.1	7.1	6.8	6.6	6.8	7.5
Number of Men	3	4	9	29	57	7	16	35	26	43	40	9
					Total 109						Total 170	

[1] GOVERNMENT PERSONNEL: 109 Males on duty with or employed by, U.S. Air Forces.

[2] NON-GOVERNMENT PERSONNEL: 170 Males Employed by Lockheed Missiles and Space Company, Sunnyvale, California.

[redacted] shows EFFECTIVE INDICATOR: (Trend correlating with years of education apparent in both groups.)

APPENDIX D

COMPARISON OF RANK-ORDER RATINGS RESULTING FROM
15-YEAR SURVEY OF JOB PREFERENCES AND RESULTS OF THIS STUDY

	RANKS RESULTING FROM 15-YEAR SURVEY *		RANKS RESULTING FROM THIS STUDY	
	Men	Women	Men	Women
Advancement	2	3	2	2
Benefits	8	10	7	9
Company	4	4	6	6
Co-Workers	5	5	8	7
Hours	9	9	10	8
Pay	6	7	3	3
Security	1	2	4	1
Supervisor	7	6	5	5
Type of Work	3	1	1	4
Working Conditions	10	8	9	10
NUMBER IN STUDY	24,343	8,245	279	65

* Jurgensen, 1961

APPENDIX E

STATISTICAL EQUATIONS USED

Means were derived from the following formula:

$$\bar{y} = \frac{y_1 + y_2 + \ldots + y_n}{n} \qquad \text{(Equation 1)}$$

where $\bar{y}$ is the sample mean, y_1 through y_n are the rank-order numbers assigned any particular incentive by the particular individuals encountered in any specific group. The n observations were added together, and the sum divided by n.

Standard deviations were derived from the square-root of the following formula for sample variance:

$$s^2 = \frac{\sum (y - \bar{y})^2}{n - 1} \qquad \text{(Equation 2)}$$

where s^2 is the sample variance, and n is the sample size. As will be obvious to any student of statistics, the Sum of Squares was substituted for the numerator in the right-hand portion of Equation 2. Thus, the standard deviation for any particular incentive's mean was derived by taking the square root of the following equation:

$$s^2 = \frac{\sum y^2 - \frac{\left(\sum y\right)^2}{n}}{n-1} \qquad \text{(Equation 3)}$$

After means and standard deviations were derived from the computer sheets, comparisons were made of the underlying variance for the major sub-groups relative to each of the ten incentives. The 95% level of confidence was chosen for testing the significance of the differences.

The method described by Li (1961), using the sample variances, s_1^2 and s_2^2, and the degrees of freedom corresponding to the variances, with the statistic F, was the procedure used to test the hypothesis that the underlying variances were equal.

Where the variances were found to be statistically different, no further test was required to conclude that the sub-groups were really different relative to the incentive being rated. Where the hypothesis that the variances were equal was accepted, the variances were pooled (to provide a "weighted" variance), and the differences between the means were tested. The pooled variance was determined by the following formula:

$$s_p^2 = \frac{s_1^2 (n-1) + s_2^2 (n-1)}{(n-1) + (n-1)} \qquad \text{(Equation 4)}$$

where n is the sample size, and n-1 the degrees of freedom for the sample variance being considered.

The statistic t was used to test the hypothesis that mean ranks relative to any particular incentive were equal. Li (1961) shows that, when testing the hypothesis that two population means are equal from samples derived from the population, the statistic t becomes:

$$t = \frac{\bar{y}_1 - \bar{y}_2}{\sqrt{s_p^2 \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \qquad \text{(Equation 5)}$$

Where the means being considered were shown, relative to tabled t values corresponding to the appropriate degrees of freedom, to be statistically different, no further test was required to conclude that

the sub-groups being considered were truly different relative to the particular incentive for which the means were computed.

Individual reliability of the JJPB was measured with Spearman's Rank-Order Coefficient of Correlation (Rho), which was computed from:

$$\text{Rho} = \frac{6 \left(\sum D^2 \right)}{n \left(n^2 - 1 \right)} \qquad \text{(Equation 6)}$$

The above Equation 6 provided an opportunity to test how well rank-order responses made on the initial test correlated with response made on the retest using the same instrument.

BLANK PAGE

MAXIMIZING PROTECTION FROM EOQ SAFETY LEVELS

by

IRVING KATZ
Chief, Operations Analysis Office

and

VICTOR J. PRESUTTI, JR.
Operations Analyst

Operations Analysis Office
Directorate of Operations
Headquarters, Air Force Logistics Command

Irving Katz

Irving Katz

Irving Katz, Chief of the Operations Analysis Office of Hq Air Force Logistics Command, was born in New York City in 1918. He received his B.S. degree in 1938 and his M.B.A. in 1951, both from C.C.N.Y., and has taken graduate studies at George Washington University, Omaha University, and Johns Hopkins University. Through correspondence courses, he has completed all the major Air Force education series, including the Air War College and the Industrial College of the Armed Forces.

Mr. Katz has served with several government agencies, including the War Production Board, the Office of the Housing Expediter, the Strategic Air Command, the Army Chemical Corporation Research and Engineering Command, and has been with the Air Force Logistics Command since 1956. During World War II he served in the Army Ordnance Corps, in grades from private to captain, and is now a colonel in the Air Force Reserve. He has received outstanding performance ratings in both SAC and AFLC, and was awarded the Air Force Decoration for Exceptional Civilian Service for his work in SAC. He has been a member of the Council of Air Force Scientists, and is now the Chairman of AFLC's Logistics Research Advisory Committee, a member of the Commander's Senior Civilian Advisory Group, a member of the Executive Committee of the Military Operations Research Symposia, and an active participant in community affairs.

Mr. Katz belongs to Phi Beta Kappa, the American Association for the advancement of Science, the American Statistical Association, the Air Force Association, the Reserve Officers Association, the Operations Research Society of America and the Dayton Council on World Affairs. He has published many operations research studies and made presentations at numerous symposia. He is listed in American Men of Science and other Who's Who publications.

Victor J. Presutti, Jr.

Victor J. Presutti, Jr.

Victor J. Presutti, Jr., of the Operations Analysis Office of Hq Air Force Logistics Command received his B.S. degree in 1961 and his M.S. in 1966, both from Syracuse University. While at Syracuse he was elected to Pi Mu Epsilon National Mathematics Honorary Society.

Mr. Presutti served as an Air Force Lieutenant in the Operations Analysis Office from October of 1962 to May of 1965. For his contributions in Economic Order Quantity Item problem areas he was awarded the Air Force Commendation Medal and in 1964 was nominated for AFLC Distinguished Company Grade Officer of the year.

BLANK PAGE

Abstract

Maximizing Protection from
EOQ Safety Levels

A Safety Level is the quantity of materiel, in addition to the operating level of supply, required to be on hand to permit continuous operations in the event of minor interruption of normal replenishment or unpredictable fluctuations in demands. In an effort to increase effectiveness without increasing cost, the Air Force Logistics Command conducted research leading to new formulations of Safety Levels for bit and piece items (otherwise known as Economic Order Quantity or EOQ items), at the wholesale depot echelon.

Various Safety Level alternatives were tested, using two independently developed large-scale Monte Carlo computer simulation models. This paper treats the results from only one of these models, which was developed in the Operations Analysis Office for use under conditions of random item behavior with unusually high variability.

Both simulations demonstrated the superior cost-effectiveness of the new Safety Level formulation proposed by the Operations Analysis Office. The indicated benefits to be derived are very large, including simultaneous improvement in support effectiveness and a reduction in inventory of millions of dollars.

TABLE OF CONTENTS

I. Introduction

The Economic Order Quantity (EOQ) Items are the expendable, consumption type, or bit and piece type items in the Air Force inventory. These typically low cost items represent about 85 to 90 percent of the total items in the Air Force inventory.

The Operations Analysis Office of Headquarters Air Force Logistics Command (AFLC) first became involved in the EOQ area in December of 1962, when we were asked by what is now the Requirements Management Division of the Directorate of Supply for ideas about the proper base period of time to use in calculating EOQ Issue Rates. Since then we have worked closely with the Directorate of Supply in an effort to improve AFLC management of EOQ items. Our work in this area led to two previous Operations Analysis publications, Some Comments Regarding the EOQ Concept of Inventory Management, and Practical Aspects of Economic Order Quantities.

As a result of the conceptual picture of an optimal Safety Level in the latter paper, in February of 1965, the Vice Commander of AFLC instructed the Directorate of Supply and the Operations Analysis Office to design and test several Safety Level alternatives "with the objective of increasing effectiveness without increasing cost." This report describes the results of our efforts to meet that objective.

II. Safety Levels

According to AFLCR 57-42 the Safety Level is "the quantity of material required to be on hand to permit continued operations in the event of minor interruption of normal replenishment or unpredicted fluctuation in issue demand." In other words, the Safety Level is protection, or insurance, against uncertainty. Like any other insurance it ought to be purchased in a manner that offers the most return for the money spent.

The current[1] policy sets the Safety Level at 25 percent of the units forecast to be issued during the Procurement Lead Time or PLT (the time from the release of a Purchase Request to the delivery of the first production quantity).

Before discussing the alternative Safety Level policies tested, there is a point to be made about the current policy. Among the criticisms of this policy is the claim that it is not a variable Safety Level. This is not literally true; it is variable, and is a function of both Issue Rate and PLT. Thus if the PLT is 6 months and the issue rate is 10 per month, the Safety Level is 25 percent of 60 units or 15 units; if the PLT is 8 months and the issue rate is 1 per month the Safety Level is 2 units.

[1] "Current" means the time when this study was being performed, namely the latter part of 1965 and early months of 1966.

The issue is not the absence of a variable Safety Level, but rather one of choosing the correct variables, and function of those variables, to make the Level most effective.

In a previous paper[2] we expressed the conviction that the unit price of an item should be a key variable in the Safety Level calculation. This conviction has now gained widespread acceptance. As a result, all of the new alternatives tested had a common difference from the current approach - the Safety Levels of these alternatives reflect the unit price of the item in an effort to obtain the most protection per dollar invested, or the most return for the money spent.

III. The Alternatives

A. The Operations Analysis (OA) Alternatives

The alternatives designed by the Operations Analysis Office are shown on the next page. Most of these OA alternatives were designed for maximum efficiency, so that as the unit cost increases the Safety Level is decreased until it disappears completely. Two of the OA alternatives (OA1d and OA2d) were designed as compromise alternatives because our associates in the Directorate of Supply believed that all items, regardless of unit cost, should have a Safety Level. The OA1 alternatives are more austere, and consequently give less protection, then the OA2 alternatives. This methodology was employed to give the manager a broad range of alternatives to choose from, depending on the situation and upon explicit or implicit values for stockout penalties. For example, if

---

[2] Presutti, Victor J. and Irving Katz, Practical Aspects of Economic Order Quantities, Operations Analysis Report No. 5, Hq Air Force Logistics Command (MCOA), Wright-Patterson Air Force Base, Ohio, May 1965

Table 1

OA ALTERNATIVES

	Unit Cost Range ($)				
Alter-native	0-4.99	5.00-9.99	10.00-14.99	15.00-24.99	25.00 or more
OA1a	5.00	2.00	0.00	0.00	0.00
OA1b	6.25	2.50	0.00	0.00	0.00
OA1c	7.50	3.00	0.00	0.00	0.00
OA1d	5.00	2.00	0.50	0.50	0.50
OA2a	7.00	5.00	3.00	1.00	0.00
OA2b	8.75	6.25	3.75	1.25	0.00
OA2c	10.50	7.50	4.50	1.50	0.00
OA2d	7.00	5.00	3.00	1.00	0.50

K = table value

Safety Level (S/L) in units = (K) • (Monthly Issue Rate) $\cdot\sqrt{\frac{PLT}{9}}$

where PLT = Procurement Lead Time in months

Ex. Suppose we have a line item as follows:

Unit Cost = $12.00

Monthly Issue Rate = 10 units

Procurement Lead Time = 9 mos.

Now if we are going to use OA2a:

S/L = (3) • (10) $\cdot\sqrt{\frac{9}{9}}$ = 30 units

the financial situation were extraordinarily tight, the manager could select a more austere alternative than he might otherwise consider desirable. On the other hand, if effectiveness were below acceptable levels a "richer" alternative might be selected:

B. Other AFLC Alternatives

The other alternatives tested (See Tables 2 thru 5) originated in AFLC's Directorate of Supply. In these alternatives the Safety Level was made a function of the unit cost and of the value of the annual issues for the item in question. In all cases, the Safety Level formula cited for these AFLC alternatives is modified as follows:

a. If (PLT) • (Monthly Issue Rate) $\leq$ 20,

$$K = \text{minimum} \begin{cases} \text{table value} \\ 1 \end{cases}$$

b. If (PLT) • (Monthly Issue Rate) $>$ 20 and issues in all quarters of "history" used do not vary from the quarterly average by more than 10%, $K = \text{minimum} \begin{cases} \text{table value} \\ 3 \end{cases}$

Table 2

ALTERNATIVE AFLC 1

Dollar Value of Annual Issues	Unit Cost Range ($)					
	.00-.99	1.00-9.99	10.00-24.99	25.00-99.99	100.00-499.99	500.00 or more
.00-99.99	2	1	1	0	0	0
100.00-499.99	3	2	1	1	0	0
500.00-1,499.99	4	3	2	2	1	0
1,500.00-4,999.99	4	4	3	3	2	0
5,000.00-9,999.99	5	4	3	3	2	1
10,000.00-49,999.99	6	5	4	4	3	2
50,000.00 or more	7	6	5	5	3	2

K = table value

$$S/L = K \cdot \sqrt{(PLT) \cdot (\text{Monthly Issue Rate})}$$

Table 3

ALTERNATIVE AFLC 2

Dollar Value of Annual Issues	Unit Cost Range ($)					
	.00-.99	1.00-9.99	10.00-24.99	25.00-99.99	100.00-499.99	500.00 or more
.00-99.99	3	2	1	0	0	0
100.00-499.99	4	3	1	1	0	0
500.00-1,499.99	5	4	2	2	0	0
1,500.00-4,999.99	6	5	3	3	1	0
5,000.00-9,999.99	7	6	4	3	2	1
10,000.00-49,999.99	8	7	5	4	3	2
50,000.00 or more	10	8	6	5	3	2

K = table value

$$S/L = K \cdot \sqrt{(PLT) \cdot (\text{Monthly Issue Rate})}$$

Table 4

ALTERNATIVE AFLC 3

Dollar Value of Annual Issues	Unit Cost Range ($)					
	.00-.99	1.00-9.99	10.00-24.99	25.00-99.99	100.00-499.99	500.00 or more
.00-99.99	2	1	1	0	0	0
100.00-499.99	3	2	1	1	0	0
500.00-1,499.99	3	2	2	1	0	0
1,500.00-4,999.99	7	5	3	2	1	0
5,000.00-9,999.99	11	9	5	5	3	2
10,000.00-49,999.99	15	11	9	7	5	3
50,000.00 or more	19	16	13	9	7	3

K = table value

$$S/L = K \cdot \sqrt{(PLT) \cdot (\text{Monthly Issue Rate})}$$

Table 5

ALTERNATIVE AFLC 4

Dollar Value of Annual Issues	Unit Cost Range ($)					
	.00-.99	1.00-9.99	10.00-24.99	25.00-99.99	100.00-499.99	500.00 or more
.00-99.99	2	1	1	0	0	0
100.00-499.99	3	2	1	1	0	0
500.00-1,499.99	3	2	2	1	0	0
1,500.00-4,999.99	7	5	3	2	1	0
5,000.00-9,999.99	15	12	8	3	2	1
10,000.00-49,999.99	20	18	12	7	3	1
50,000.00 or more	25	20	16	9	4	2

K = table value

$$S/L = K \cdot \sqrt{(PLT) \cdot (\text{Monthly Issue Rate})}$$

C. Hq USAF Alternative

In addition to the above alternatives, there was one other tested. This one, designed at Hq USAF,[3] was the most sophisticated of the alternatives considered. In it the Safety Level is a function of the variance to mean ratio, as well as the unit cost and the annual demands.

Conceptually, reflection of each item's variability is a desirable attribute. However, the added complexity of such an approach could be quite costly in computer time when applied to approximately one million EOQ items. In order to determine the true worth of this alternative, one would have to balance its added cost against the improvement it might afford.

IV. The Simulation

A simulation model, which had previously been constructed in the Operations Analysis Office (with the assistance of UNIVAC personnel) for the test of trend analysis procedures, was adapted for use in testing the Safety Level alternatives. The model is fully described in Operations Analysis Report No. 5, Practical Aspects of Economic Order Quantities.

In testing the Safety Level alternatives only Group 4 (10,000 line items selected at random) of the input data was used. Furthermore, although the simulation provides for other measures of effectiveness, the Fill Rate (units issued at time of demand/units demanded) is the measure of effectiveness employed in this evaluation of the Safety Level alternatives. This is not the same Fill Rate as is used in routine AFLC operations. The operational Fill Rate, during a given time interval, is defined as the units shipped divided by the units demanded during that time interval. This formula yields an unrealistically high Fill Rate because it allows credit for items that were backordered originally and shipped after some waiting period. If the time interval considered is long enough, this operational Fill Rate can approach 100 percent, no matter how long the requisitioners are actually waiting for their supplies.

It is noted that as a result of a recent review by the Comptroller of AFLC,[4] "off-the-shelf" Fill Rates will be added to the AFLC management tools, as another measure of effectiveness of the supply system.

It should be noted that because of the short test period in the simulation, as limited by the availability of data, the difference in Fill Rates between alternatives is probably smaller than would be revealed over a longer time interval. Thus alternatives giving an improved Fill Rate

---

[3] Kay, Lt Col Ira M., Safety Level Cost/Effectiveness Study, Supplement 1 to: Research Report, Annex 2 to Study Directive AFSLP 64-1, Hq USAF, 31 August 1965.

[4] AFLC Management Review, 25 February 1966, Management Analysis Division, Headquarters AFLC.

in the one year test in the simulation can be expected to give an even greater improvement over a period of five years. This disadvantage of our simulation will gradually disappear as we enlarge our real-life data base (a project currently in progress), and is viewed as a worthwhile sacrifice in order to include the dynamics of changing item demand patterns which our input data carry from the real world.

By the time we began the Safety Level portion of our studies, another EOQ simulation became available.[5] This simulation gets its demand inputs from a data generator, based upon empirically derived statistical distributions. In this respect it appeared to be less desirable than a simulation using live demand data as inputs. On the other hand, it does provide a ten year test period (which can be further expanded). The Directorate of Supply requested that the Safety Level alternatives be tested on this simulation also. Although we do not report the results of the SGOS simulation in this paper, it is significant to note that both simulations ranked the alternatives in about the same preferential sequence.

V. The Results of the Simulation

A. Discussion

Before actually evaluating the alternatives, there is a point to be made regarding our processing of cost and effectiveness. If two alternatives are being considered, and if A costs less than B and provides better effectiveness, then clearly A is preferred to B. (In our step-by-step effort to eliminate candidate alternatives which are not preferred, we shall say that one "dominates" another if it is less costly and also more effective.) However, if A costs less but B affords better effectiveness, then a cost-effectiveness trade-off must be made. This is an extremely difficult trade-off to make because the cost of a backorder is not as yet unequivocally specifiable.

B. Results

The Operations Analysis simulation results are presented in Table 6. The left side of the table ranks alternatives by cost (least costly ranks highest), while the right side ranks alternatives by effectiveness (best Fill Rate ranks highest). The Safety Level cost is the inventory investment in Safety Levels during the test year.

Since all the OA alternatives dominate both the current method and AFLC alternatives 1 thru 4, we can eliminate the latter from consideration. We can also eliminate OAld from consideration since it is dominated by OAlc

[5] Specifics regarding this model can be obtained from the Logistics Simulation Center (SGOS) at AFLC.

In addition, we can eliminate OA1a, OA1b, OA2b, and OA2d. The basis for these eliminations is presented in Appendix A.

We can therefore conclude that the best alternative, regardless of the cost of a backorder (or how much it is worth

SIMULATION RESULTS

Cost		Effectiveness	
Alternative	S/L Cost ($)	Alternative	Fill Rate
No S/L	0	OA2c	57.6
OA1a	206,833	OA2b	57.1
OA1b	245,834	OA2d	57.0
OA1c	252,790	OA2a	57.0
OA1d	299,703	OA1c	55.8
OA2a	498,256	OA1b	55.4
OA2d	557,905	OA1d	55.3
OA2b	583,906	OA1a	55.0
OA2c	683,937	AFLC4	54.9
AFLC1	729,516	Current	54.5
AFLC2	730,460	AFLC3	54.3
Current	742,897	AFLC1	54.2
AFLC4	967,182	AFLC2	54.0
AFLC3	1,054,687	No S/L	54.1

Table 6

to raise the Fill Rate), is either OA1c, OA2a, OA2c, or no Safety Level at all.

In Figure 1 the simulation results are depicted graphically. The point values are the specific alternatives tested. The cost-effectiveness curve is an envelope of the best possibilities recognized. Note that although the only points on the curve are those alternatives that have not

already been eliminated, all the OA alternatives are considerably closer to the curve than any of the others.

The Hq USAF designed alternative reflecting the variance/mean ratio was not tested in the Operations Analysis simulation, but was tested in the SGOS simulation (See IV). In that test all of the OA alternatives were again closer to the optimum cost-effectiveness curve than all other alternatives, including the Hq USAF alternative. In fact, two of the OA alternatives dominated the Hq USAF alternative.

VI. Investigation of OA2d

From Figure 1 it seemed clear that OA1c and OA2a gave attractive combinations of cost and effectiveness. However, we had been advised that the Directorate of Supply could not, at this time, endorse any alternative that gave no Safety Level to items with a unit cost of $25.00 or more. Consequently, we decided to investigate OA2d, which differs from OA2a only by allowing a Safety Level of approximately one-half month for all line items with a unit cost of $25.00 or more.

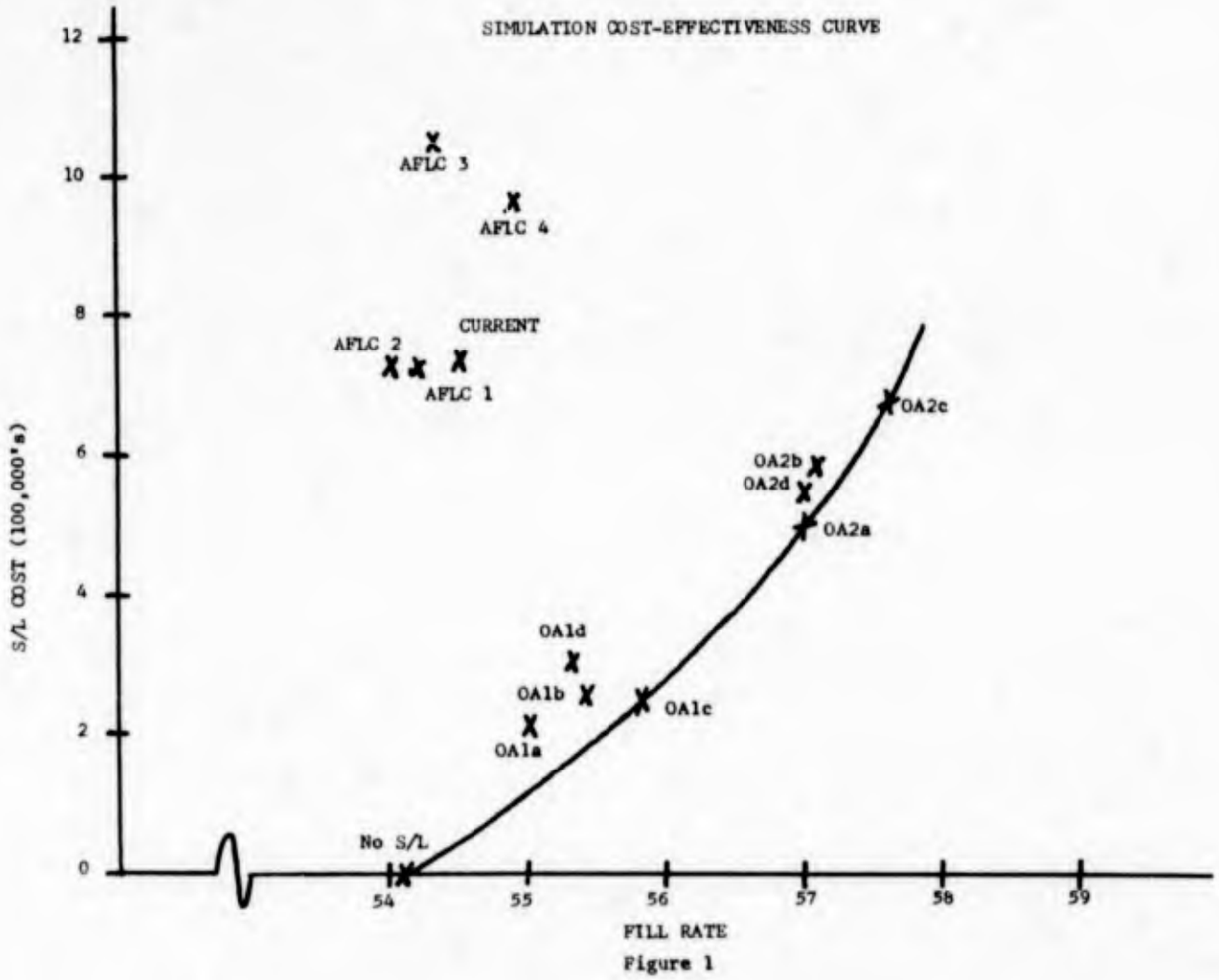


Figure 1

From Table 6 we noted that OA2d is dominant to the current policy. In Figure 1 we noted that OA2d, although not on the optimal curve, was very close to it.

Our next step was an analysis of actual inventory data. (See Appendix B.) Table 7 gives a summary of our findings. It shows, on the basis of data from over 80% of the EOQ items, that under the current policy in 1965 we required a Safety Level investment of $90 million. If in 1965 OA2d had been implemented, it would have required a Safety Level investment of $73 million for the same items. Furthermore, if in 1965 we had extended the current policy to give the same protection as OA2d, it would have required an investment of $179 million, or $106 million more than OA2d.

Table 7

OA2d VS. CURRENT POLICY

	Safety Level Investment (millions of dollars)
Current Policy	$90
OA2d	73
Cost Reduction	$17
Extended Current Policy*	179
OA2d	73
Cost Avoidance	$106

* Fixed percent of Procurement Lead Time necessary to obtain same protection as OA2d.

In view of the great gains apparently achievable through use of OA2d in lieu of the current Safety Level, and since OA2d seemed to be a gradual way of learning operationally whether "protection per dollar" principles would not lead to serious shortages on higher cost items, we concluded that it was a sound approach to adopt, at least on an interim basis.

VII. Recommendation

We recommended implementation of OA2d at the earliest possible date.

VIII. Status and Prospects

A. Status

Headquarters Air Force Logistics Command and Headquarters United States Air Force have approved our recommendation. Work is beginning on the programming of Safety Level alternative OA2d, on a high priority basis.

B. Prospects

We are continuing our work in Safety Levels and in the EOQ area in general. One facet that is receiving particular attention is that of procurement and procurement strategies.

APPENDIX A

I. The Elimination of OA1a

Assume OA1a is the preferred alternative.

a. OA1a is preferred to No S/L. Since OA1a costs $206,833 more than No S/L and increases the Fill Rate .9%, it must be worth more than $\frac{\$206,833}{.9}$ or $229,814 to increase the Fill Rate 1%.

b. OA1a is preferred to OA2c. Then it must be worth less than $\frac{\$477,104}{2.6}$ or $183,502 to increase the Fill Rate 1%.

But a and b cannot both be true. Hence our assumption is false, and so OA1a cannot be the preferred alternative.

Note: It can be argued that the "worth" of a unit of fill is not constant. However, looking back at our optimum cost-effectiveness curve (Figure 1) we note that as the Fill Rate increases the cost of each unit of fill increases. Therefore if we are willing to pay X dollars to increase our fill one unit from A to B, we should expect to pay more than X dollars to increase our fill one unit from B to C. Hence, the argument for the elimination of OA1a is even stronger than as presented.

II. The Elimination of OA1b

Assume OA1b is the preferred alternative.

a. OA1b is preferred to No S/L. Therefore it is worth more than $\frac{\$245,834}{1.3}$ or $189,103 to increase the Fill Rate 1%.

b. OA1b is preferred to OA2a. Therefore it is worth less than $\frac{\$252,422}{1.6}$ or $157,764 to increase the Fill Rate 1%.

Since a and b cannot both be true, our assumption is false and OA1b cannot be the preferred alternative.

III. The Elimination of OA2b

Assume OA2b is the preferred alternative.

a. OA2b is preferred to OA2a. Therefore it is worth more than $\frac{85,650}{.1}$ or $856,500 to increase the Fill Rate 1%.

b. OA2b is preferred to OA2c. Therefore it is worth less than $\frac{100,031}{.5}$ or $200,062 to increase the Fill Rate 1%.

Since and b cannot both be true our assumption is false and OA2b cannot be the preferred alternative.

IV. The Elimination of OA2d

Assume OA2d is the preferred alternative.

a. OA2d is preferred to OA2a. Therefore it is worth more than $\frac{59,649}{.08*}$ or $745,613 to increase the Fill Rate 1%.

b. OA2d is preferred to OA2c. Therefore it is worth less than $\frac{126,032}{.6}$ or $210,053 to increase the Fill Rate 1%.

Since a and b cannot both be true, our assumption is false and OA2d cannot be the preferred alternative.

* Fill Rate for OA2d = [illegible].04
Fill Rate for OA2a = 56.96

V. Elimination Theorems

Situations can arise in which the Measure of Effectiveness,[1] the basis for choosing between alternative procedures, consists of a linear combination of component measures. For such cases, the rationale used in the preceding elimination of alternatives can be formalized and extended. The two Elimination Theorems that follow are illustrative of this type of extension.

The following definitions hold for the remainder of this Appendix:

Positive measure: a measure in which an increase is desirable (For example, Fill Rate)

Negative measure: a measure in which a decrease is desirable (For example, cost)

A_k = alternative k (k = 1, 2, 3)

$P_i(A_k)$ = the number of units of positive measure i of alternative k

$N_j(A_k)$ = the number of units of negative measure j of alternative k

$\alpha_i (> 0)$ = the weight assigned to positive measure i

$\beta_j (> 0)$ = the weight assigned to negative measure j

$M(A_k)$ = the Measure of Effectiveness of alternative A_k

$$= \sum_{i=1}^{n} P_i(A_k) \cdot \alpha_i - \sum_{j=1}^{m} N_j(A_k) \cdot \beta_j$$

$|X|$ = absolute value of X

---

[1] Measure of Effectiveness in this Appendix has a broader meaning than in the body of the report. The measure of effectiveness used earlier corresponds to a positive measure in this Appendix.

Elimination Theorem 1:

Let i (i = 1, 2) represent a positive measure and j (j = 1) a negative measure of alternative k (K = 1, 2, 3)

If a. $P_1(A_1) > P_1(A_2)$

b. $P_1(A_2) \neq P_1(A_3)$

c. $P_2(A_3) > P_2(A_2)$

d. $N_1(A_2) > N_1(A_3)$

e. $$\left|\frac{P_2(A_1) - P_2(A_2)}{P_1(A_1) - P_1(A_2)}\right| < \left|\frac{P_2(A_2) - P_2(A_3)}{P_1(A_2) - P_1(A_3)}\right|$$

f. $$\left|\frac{N_1(A_1) - N_1(A_2)}{P_1(A_1) - P_1(A_2)}\right| < \left|\frac{N_1(A_2) - N_1(A_3)}{P_1(A_2) - P_1(A_3)}\right|$$

Then $M(A_1) > M(A_2)$ or $M(A_3) > M(A_2)$ (i.e., A_2 can not be the preferred alternative) regardless of the value of α_1, α_2, and β_1.

Proof: i. Assume $M(A_2) \geq M(A_1)$

Then $P_1(A_2)\alpha_1 + P_2(A_2)\alpha_2 - N_1(A_2)\beta_1 \geq$

$$P_1(A_1)\alpha_1 + P_2(A_1)\alpha_2 - N_1(A_1)\beta_1$$

$$\left[P_1(A_2) - P_1(A_1)\right] \cdot \alpha_1 + \left[P_2(A2) - P_2(A_1)\right] \cdot \alpha_2 + \left[N_1(A_1) - N_1(A_2)\right] \cdot \beta_1 \geq 0$$

$$\left[P_1(A_1) - P_1(A_2)\right] \cdot \alpha_1 \leq \left[P_2(A_2) - P_2(A_1)\right] \cdot \alpha_2 + \left[N_1(A_1) - N_1(A_2)\right] \cdot \beta_1$$

$$\left|P_1(A_1) - P_1(A_2)\right| \cdot \alpha_1 \leq \left|P_2(A_1) - P_2(A_2)\right| \cdot \alpha_2 + \left|N_1(A_1) - N_1(A_2)\right| \cdot \beta_1$$

$$\alpha_1 \leq \left|\frac{P_2(A_1) - P_2(A_2)}{P_1(A_1) - P_1(A_2)}\right| \cdot \alpha_2 + \left|\frac{N_1(A_1) - N_1(A_2)}{P_1(A_1) - P_1(A_2)}\right| \cdot \beta_1$$

(1) $\alpha_1 < \left| \frac{P_2(A_2) - P_2(A_3)}{P_1(A_2) - P_1(A_3)} \right| \cdot \alpha_2 + \left| \frac{N_1(A_2) - N_1(A_3)}{P_1(A_2) - P_1(A_3)} \right| \cdot \beta_1$

ii. Assume $M(A_2) \geq M(A_3)$

Then $\left[P_1(A_2) - P_1(A_3) \right] \cdot \alpha_1 + \left[P_2(A_2) - P_2(A_3) \right] \cdot \alpha_2$

$$+ \left[N_1(A_3) - N_1(A_2) \right] \cdot \beta_1 \geq 0$$

$$\left[P_1(A_2) - P_1(A_3) \right] \cdot \alpha_1 \geq$$

$$\left[P_2(A_3) - P_2(A_2) \right] \cdot \alpha_2 + \left[N_1(A_2) - N_1(A_3) \right] \cdot \beta_1$$

$$\left| P_1(A_2) - P_1(A_3) \right| \cdot \alpha_1 \geq$$

$$\left| P_2(A_2) - P_2(A_3) \right| \cdot \alpha_2 + \left| N_1(A_2) - N_1(A_3) \right| \cdot \beta_1$$

(2) $\alpha_1 \geq \left| \frac{P_2(A_2) - P_2(A_3)}{P_1(A_2) - P_1(A_3)} \right| \alpha_2 + \left| \frac{N_1(A_2) - N_1(A_3)}{P_1(A_2) - P_1(A_3)} \right| \cdot \beta_1$

Hence our assumptions lead to a contradiction (See equations (1) and (2)) and the Theorem is true.

Therefore $\left[P_2(A_2) - P_2(A_1)\right]\alpha_2 + \left[N_1(A_1) - N_1(A_2)\right]\beta_1$

$+ \left[N_2(A_1) - N_2(A_2)\right]\beta_2 \geq 0$

$\left[P_2(A_1) - P_2(A_2)\right]\alpha_2 \leq$

$\left[N_1(A_1) - N_1(A_2)\right]\beta_1 + \left[N_2(A_1) - N_2(A_2)\right]\beta_2$

$\left|P_2(A_1) - P_2(A_2)\right| \cdot \alpha_2 \leq$

$\left|N_1(A_1) - N_1(A_2)\right| \cdot \beta_1 + \left|N_2(A_1) - N_2(A_2)\right| \cdot \beta_2$

$$\alpha_2 \leq \left|\frac{N_1(A_1) - N_1(A_2)}{P_2(A_1) - P_2(A_2)}\right| \cdot \beta_1 + \left|\frac{N_2(A_1) - N_2(A_2)}{P_2(A_1) - P_2(A_2)}\right| \cdot \beta_2$$

$$(1)\quad \alpha_2 < \left|\frac{N_1(A_2) - N_1(A_3)}{P_2(A_2) - P_2(A_3)}\right| \cdot \beta_1 + \left|\frac{N_2(A_2) - N_2(A_3)}{P_2(A_2) - P_2(A_3)}\right| \cdot \beta_2$$

ii. Assume $M(A_2) \geq M(A_3)$

Then $\left[P_1(A_2) - P_1(A_3)\right]\alpha_1 + \left[P_2(A_2) - P_2(A_3)\right]\alpha_2$

$+ \left[N_1(A_3) - N_1(A_2)\right]\beta_1 + \left[N_2(A_3) - N_2(A_2)\right]\beta_2 \geq 0$

Therefore $\left[P_2(A_2) - P_2(A_3)\right]\alpha_2 + \left[N_1(A_3) - N_1(A_2)\right]\beta_1$

$+ \left[N_2(A_3) - N_2(A_2)\right]\beta_2 \geq 0$

$\left[P_2(A_2) - P_2(A_3)\right]\alpha_2 \geq \left[N_1(A_2) - N_1(A_3)\right]\beta_1 + \left[N_2(A_2) - N_2(A_3)\right]\beta_2$

$\left|P_2(A_2) - P_2(A_3)\right| \cdot \alpha_2 \geq$

$\left|N_1(A_2) - N_1(A_3)\right| \cdot \beta_1 + \left|N_2(A_2) - N_2(A_3)\right| \cdot \beta_2$

$$(2)\quad \alpha_2 \geq \left|\frac{N_1(A_2) - N_1(A_3)}{P_2(A_2) - P_2(A_3)}\right| \cdot \beta_1 + \left|\frac{N_2(A_2) - N_2(A_3)}{P_2(A_2) - P_2(A_3)}\right| \cdot \beta_2$$

Hence our assumption leads to a contradiction (See equations (1) and (2)) and the Theorem is true.

Elimination Theorem 2:

Let i (i = 1, 2) represent a positive measure and
j (j = 1, 2) represent a negative measure of alternative
k (k = 1,2, 3)

If a. $P_1(A_1) > P_1(A_2)$

b. $P_2(A_1) > P_2(A_2)$

c. $N_1(A_3) < N_1(A_2)$

d. $N_2(A_3) < N_2(A_2)$

e. $P_2(A_2) \neq P_2(A_3)$

f. $P_1(A_2) < P_1(A_3)$

g. $$\left|\frac{N_1(A_1) - N_1(A_2)}{P_2(A_1) - P_2(A_2)}\right| < \left|\frac{N_1(A_2) - N_1(A_3)}{P_2(A_2) - P_2(A_3)}\right|$$

h. $$\left|\frac{N_2(A_1) - N_2(A_2)}{P_2(A_1) - P_2(A_2)}\right| < \left|\frac{N_2(A_2) - N_2(A_3)}{P_2(A_2) - P_2(A_3)}\right|$$

Then $M(A_1) > M(A_2)$ or $M(A_3) > M(A_2)$ (i.e., A_2 can not be the preferred alternative) regardless of the values of α_1, α_2, β_1, and β_2.

Proof: i. Assume $M(A_2) \geq M(A_1)$

$$P_1(A_2)\alpha_1 + P_2(A_2)\alpha_2 - N_1(A_2)\beta_1 - N_2(A_2)\beta_2 \geq$$
$$P_1(A_1)\alpha_1 + P_2(A_1)\alpha_2 - N_1(A_1)\beta_1 - N_2(A_1)\beta_2$$
$$\left[P_1(A_2) - P_1(A_1)\right]\alpha_1 + \left[P_2(A_2) - P_2(A_1)\right]\alpha_2 + \left[N_1(A_1) - N_1(A_2)\right]\beta_1$$
$$+ \left[N_2(A_1) - N_2(A_2)\right]\beta_2 \geq 0$$

While the preceding arguments do not constitute a thorough study of the area of "simulation result evaluation" they do provide us with at least two insights in this area:

a. Given a number of component measures, under certain conditions, there is a significant amount of information that can be gained regarding the merits of the alternatives under consideration even when information about the relative importance of the measures is lacking.

b. As the number of component measures increases there is a growth in the amount and the complexity of information needed to eliminate alternatives.

APPENDIX B

I. The Data

In this Appendix, using actual data, we will cost the Safety Level approaches of the OA2d and of the current method, for the items considered.[1] In addition, we will cost an extension of the current method that affords the same protection (i.e., same average Safety Level per line item[2]) as OA2d.

The entries in Table B1 (page B2) represent the number of line items in the designated unit cost-procurement lead time areas. This table was used solely to compute the average lead time in each Cost Group.

In Table B2 (page B3) we have the number of line items, the percentage of line items, the dollar value of Annual Issues, the percentage of Annual Issues ($), and the dollar value of the Monthly Issues in the various Cost Groups. These figures will be used in determining the cost of the alternatives considered in this Appendix.

---

[1] The data were taken from the D062 (EOQ) System as of 31 Dec 65 and represent a summary of over 80% of the total EOQ items. The items excluded were exception items, obsolete items, MAST System items, SAC Nuclear Ordnance items, and F-104, F-105 items at SAAMA.

[2] Three ways of defining "same protection" were considered for use: a) same Fill Rate; b) same number of units of Safety Levels; c) same average Safety Level per line item. We would have liked to use the first definition, but it is impossible to determine the Fill Rate given by a specified Safety Level without simulating or making certain dubious assumptions about the demand patterns of the items. The second definition would require the use of an improbable assumption about the average unit price of a line item for each Cost Group. In addition to the fact that the third definition did not require simulation or assumptions, it is also the most conservative of the three. That is, by using the first definition and simulating or by using the second definition and making the necessary assumption we would arrive at a larger estimate of savings for OA2d than by using the third definition.

BREAKOUT OF ITEMS BY UNIT COST AND ANNUAL ISSUES

Unit Cost	Number of Line Items	% of Line Items	Annual Issues ($)	% of Annual Issues ($)	Monthly Issues ($)
.01-.99 (Cost Group 1)	158,536	19.37	18,007,000	3.30	1,500,583
1.00-4.99 (Cost Group 2)	201,730	24.64	41,104,000	7.52	3,425,332
5.00-9.99 (Cost Group 3)	109,380	13.36	41,658,000	7.62	3,471,499
10.00-24.99 (Cost Group 4-5)	132,585	16.20	77,458,000	14.18	6,454,831
25.00-99.99 (Cost Group 6)	142,409	17.40	154,111,000	28.21	12,842,578
100.00-499.99 (Cost Group 7)	62,150	7.59	142,375,000	26.06	11,864,579
500.00 and over (Cost Group 8)	11,844	1.45	71,632,000	13.11	5,969,331
Total	818,634	100.00	546,345,000	100.00	45,528,732

Table B2

Table B1

BREAKOUT OF ITEMS BY UNIT COST AND PROCUREMENT LEAD TIME

PLT (mos)	Unit Cost ($) .01-.99 (Cost Group 1	1.00-4.99 (Cost Group 2)	5.00-9.99 (Cost Group 3)	10.00-24.99 (Cost Groups 4 - 5)*	25.00-99.99 (Cost Group 6)	100.00-499.99 (Cost Group 7)	500.00 or more (Cost Group 8)	Total
1	33	113	79	78	56	18	4	381
2	1,062	1,031	414	363	345	99	11	3,325
3	3,325	3,137	1,262	1,230	1,032	268	23	10,277
4	8,817	12,189	5,405	5,513	4,723	1,379	196	38,222
5	9,914	12,125	6,132	6,880	7,182	2,388	461	45,082
6	72,305	52,513	21,955	23,988	23,037	8,341	1,551	203,690
7	13,381	26,236	15,710	18,672	19,838	7,933	1,594	103,364
8	8,459	16,490	10,630	14,762	18,068	9,190	1,593	79,192
9	37,737	70,265	41,929	51,477	55,152	24,688	4,205	285,453
10	1,759	3,842	2,984	4,744	5,507	2,718	593	22,147
11	569	1,420	1,059	1,825	2,495	1,945	581	9,894
12	1,049	1,986	1,447	2,296	3,768	2,539	854	13,939
>12**	126	383	374	757	1,206	644	178	3,668
Average PLT	6.75	7.27	7.54	7.71	7.87	8.17	8.38	7.47

*Cost Group 4 is all line items with a unit cost of $10.00-$14.99.
Cost Group 5 is all line items with a unit cost of $15.00-$24.99.
**If PLT > 12 it was assumed PLT = 14.

II. Definitions

The following definitions will be used throughout the remainder of this Appendix:

$S_{i,j}$ = Safety Level in units for line item i of Cost Group j

$C(S_{i,j})$ = Safety Level Investment for line item i of Cost Group j

N_j = Number of line items in Cost Group j

F_j = Fraction of line items in Cost Group j

T_j = Procurement Lead Time (in months) of each line item in Cost Group j (it was assumed that all line items in a given Cost Group had the average PLT for the Group, in order to facilitate computation)

$R_{i,j}$ = Monthly Issue Rate for line item i of Cost Group j

$K_1 = 7$

$K_2 = 7$

$K_3 = 5$

$K_4 = 3$

$K_5 = 1$

$K_6 = .5$

$K_7 = .5$

$K_8 = .5$

$U_{i,j}$ = Unit cost of line item i of Cost Group j

S_j = Safety Level in months for Cost Group j

$C(S_j)$ = Safety Level Investment for Cost Group j

$C(S)$ = Total Safety Level Investment

S = Average Safety Investment in months for all line items

III. Alternative OA2d

$$S_{i,j} = R_{i,j} \cdot K_j \cdot \sqrt{\frac{T_j}{9}}$$

$$C(S_{i,j}) = R_{i,j} \cdot K_j \cdot \sqrt{\frac{T_j}{9}} \cdot U_{i,j}$$

$$C(S_j) = \sum_{i=1}^{N_j} C(S_{i,j})$$

$$= \sum_{i=1}^{N_j} R_{i,j} \cdot K_j \cdot \sqrt{\frac{T_j}{9}} \cdot U_{i,j}$$

$$= K_j \sqrt{\frac{T_j}{9}} \sum_{i=1}^{N_j} R_{i,j} \cdot U_{i,j}$$

Note that $\sum_{i=1}^{N_j} R_{i,j} \cdot U_{i,j}$ is just the Monthly Issues ($) for Cost Group j which can be read directly from the last column of Table B2.

Now, taking PLT from Table B1, and Monthly Issues ($) from Table B2 we get:

$C(S_1) = (7/3 \quad \sqrt{6.75}) \quad \$1,500,583 = \$9,095,034$

$C(S_2) = (7/3 \quad \sqrt{7.27}) \quad \$3,425,332 = \$21,545,338$

$C(S_3) = (5/3 \quad \sqrt{7.54}) \quad \$3,471,499 = \$15,889,051$

$C(S_4)^* = (3/3 \quad \sqrt{7.71}) \quad \$3,471,499 = \$\ 9,640,353$

$C(S_5)^* = (1/3 \quad \sqrt{7.71}) \quad \$2,983,332 = \$\ 2,761,571$

---

*It was impossible for us to calculate $C(S_4)$ and $C(S_5)$ directly, since we had only the number of line items and Dollars Issued for Cost Groups 4 and 5 combined. Since (in OA2d) items in Cost Group 4 will receive three times the Safety Level of Cost Group 5 we deliberately under-estimated the number of items in Group 4 and thereby under-estimated the overall protection

$$C(S_6) = \left(\frac{.5}{3}\sqrt{7.87}\right)\ \$12,842,578 = \$6,003,905$$

$$C(S_7) = \left(\frac{.5}{3}\sqrt{8.17}\right)\ \$11,864,579 = \$5,651,099$$

$$C(S_8) = \left(\frac{.5}{3}\sqrt{8.38}\right)\ \$\ 5,969,331 = \$2,680,202$$

Therefore $C(S) = \sum_{j=1}^{8} C(S_j) = \$73,466,553$ under OA2d

---

Also,

$$S_j = K_j\sqrt{\frac{T_j}{9}} = \frac{K_j}{3}\sqrt{T_j}$$

$$S = \sum_{j=1}^{8} F_j S_j$$

Since

$F_1S_1 = .1937(6.061) = 1.174$

$F_2S_2 = .2464(6.290) = 1.550$

$F_3S_3 = .1336(4.577) = .611$

$F_4S_4 = .0540(2.777) = .150$

---

* (Contd)
of OA2d. Since (in OA2d) for each dollar issued in Cost Group 4 the Safety Level investment is three times the investment for each dollar issued in Cost Group 5, we deliberately overestimated the number of dollars issued in Group 4, and thereby overestimated the total cost of OA2d. As can be seen from Table B2, the line item density (the number of line items in a Cost Group divided by the difference between the maximum and minimum unit costs in the Group) and the Monthly Issues ($) density (the Monthly Issues ($) in a Cost Group divided by the difference between the maximum and minimum unit costs in the Group) decrease as the unit cost increases. Thus, to accomplish our under-estimation of items and overestimation of dollars issued in Group 4 we assumed the line item density for Group 4 is equal to that of Group 5, and the Monthly Issues density for Group 4 is equal to that of Group 3.

$F_5S_5 = .1090(.926) = .100$

$F_6S_6 = .1740(.468) = .081$

$F_7S_7 = .0759(.476) = .036$

$F_8S_8 = .0145(.482) = .007$

Therefore S = 3.709 under OA2d

IV. The Current Method

$$S_{i,j} = .25 \cdot R_{i,j} \cdot T_j$$

$$C(S_{i,j}) = .25 \cdot R_{i,j} \cdot T_j \cdot U_{i,j}$$

$$C(S_j) = \sum_{i=1}^{N_j} C(S_{i,j})$$

$$= \sum_{i=1}^{N_j} .25 \cdot R_{i,j} \cdot T_j \cdot U_{i,j}$$

$$= .25 \cdot T_j \sum_{i=1}^{N_j} R_{i,j} \cdot U_{i,j}$$

So

$C(S_1) = .25 \cdot (6.75)\ (\$1,500,583) = \$2,532,234$

$C(S_2) = .25 \cdot (7.27)\ (3,425,332) = 6,225,541$

$C(S_3) = .25 \cdot (7.54)\ (3,471,499) = 6,543,776$

$C(S_4) + C(S_5) = .25 \cdot (7.71)\ (6,454,831) = 12,441,687$

$C(S_6) = .25 \cdot (7.87)\ (12,842,578) = 25,267,772$

$C(S_7) = .25 \cdot (8.17)\ (11,864,579) = 24,233,403$

$C(S_8) = .25 \cdot (8.38)\ (5,969,331) = 12,505,749$

Therefore $C(S) = \sum_{j=1}^{8} C(S_j)$ = \$89,750,162 under the current method; more than one-fifth greater than the cost of OA2d.

Also,

$S_j = (.25)(T_j)$

$$S = \sum_{j=1}^{8} F_j S_j$$

$$= \sum_{j=1}^{8} F_j \cdot .25 \cdot T_j = .25 \sum_{j=1}^{8} F_j \cdot T_j$$

But $\sum_{j=1}^{8} F_j \cdot T_j$ is just the average PLT which has already been computed and shown in Table B1.

Therefore S = (.25) (7.47) = 1.868 under current method; approximately one-half as great as the average safety level in months in OA2d.

V. An Extension of Current Methodology

Suppose now we wanted to increase the protection to that of OA2d (i.e., go from an average S/L of 1.868 months to 3.709 months) by extending the current methodology (i.e., keep Safety Level a fixed percentage of the Procurement Lead Time).

Then we would have the following situation:

$S_j = X \cdot T_j$ (where X = a fixed fraction of PLT)

$$S = \sum_{j=1}^{8} F_j \cdot X \cdot T_j = X \sum_{j=1}^{8} F_j \cdot T_j =$$

$X \cdot (7.47)$

Since $X(7.47) = 3.709$, $X = .497$.

Now since the Safety Level for every item is increased by $\frac{.497}{.250}$ or 1.99 the total Safety Level Investment Cost is also increased by 1.99.

Hence C(S) = 1.99($89,750,162) = $178,602,822 under an extension of the current methodology designed to give the same protection as OA2d; almost two and one-half times the cost of OA2d, or over a hundred million dollars more.

BIBLIOGRAPHY

Kay, Lt Col Ira M., Safety Level Cost/Effectiveness Study, Supplement 1 to: Research Report, Annex 2 to Study Directive AFSLP 64-1, Hq USAF, 31 August 1965.

Presutti, Victor J. and Irving Katz, Practical Aspects of Economic Order Quantities, Operations Analysis Report No. 5, Hq Air Force Logistics Command (MCOA) Wright-Patterson AFB, Ohio, May 1965.

Management Analysis Division, AFLC Management Review, Headquarters, AFLC, 25 February 1966.

AFLC Regulation 57-42, Operational Requirements - Economic Order and Stockage Program Requirements Policy, Hq Air Force Logistics Command, Wright-Patterson AFB, Ohio, 13 August 1965.

CORRELATION OF ANALYTICAL AND EMPIRICAL TECHNIQUES FOR DESIGNING SUPERSONIC AND HYPERSONIC DECELERATORS (U)

By

W. R. Pinnell, Air Force Flight Dynamics Laboratory
and
Fredrick Bloetscher, Goodyear Aerospace Corporation

William R. Pinnell

BIOGRAPHY

PINNELL, WILLIAM R. (USAF Flight Dynamics Laboratory)

Education:

B. S. in Aerospace Engineering, 1961, West Virginia University, Morgantown, West Virginia.

Experience:

Mr. Pinnell has conducted and been associated with subsonic and supersonic decelerator free flight and wind tunnel test programs since 1961. The range of decelerators tested extends from various types of subsonic parachutes through supersonic parachutes and ballutes. He has been instrumental in the design of wind tunnel test apparatus for both subsonic and supersonic test programs and in the calibration of two subsonic in-house wind tunnel facilities. Included in current responsibilities is the monitoring of Air Force contracts under ADDPEP and similar Flight Dynamics Laboratory sponsored programs for advancing decelerator technology.

Fredrick Bloetscher

BIOGRAPHY

BLOETSCHER, FREDRICK (Goodyear Aerospace Corporation)

Education:

B. S. in Aeronautical Engineering, 1947, University of Michigan, Ann Arbor, Mich,; graduate courses in compressibility and supersonics, 1948 to 1949, NASA Langley Field, Va. (University of Virginia)

Experience:

Mr. Bloetscher is responsible for the decelerator design analysis, design, fabrication, and data evaluation under the current ADDPEP program. Previously, he was a staff section head in the Programs Department of the Space Systems Division and was responsible for projects in re-entry and recovery vehicles. He has had charge of several escape and recovery programs and has held project engineering responsibilities on the Dyna-Soar escape system program. He has directed several projects in the development of escape capsules, thrust reversers, and recovery devices that required design analysis, extensive model testing, and full-scale hardware testing. He also has worked on high-speed internal flow problems and model testing at the NASA Langley Facility.

BLANK PAGE

ABSTRACT

Recent programs investigated and evaluated design techniques for three classes of aerodynamic decelerators. The results provide engineering design capability and reduce the effort required for future applications. Test results and correlations are presented for the following classes:

1. Small supersonic parachutes

   Analytical approaches for establishing configurations for the supersonic regime to Mach 5.6 are presented. First-order correlations between the results of analytical approaches, wind-tunnel tests, and flight tests behind symmetrical bodies are discussed.

2. Reefed, high dynamic pressure parachutes

   A low-porosity configuration established by wind-tunnel tests having predictable drag loads when highly reefed is described. Free-flight tests have demonstrated excellent configurations stability and that initial loads for a reefed heavy duty parachute can be predicted. Resulting design and supporting data are illustrated to Mach 2.6.

3. Inflatable decelerators

   Free-flight and full-scale wind-tunnel tests of flight hardware have demonstrated the ability to design both textile and metal cloth inflatable devices for operating behind both symmetrical and asymmetrical payloads. Heat-protected and cooled Nomex units are discussed for testing from Mach 5 to 10.

NOMENCLATURE

A_i/A_e	Inlet to exit area ratio
C_DA	Effective Drag area (drag coefficient-area product)
C_{Do}	Drag coefficient based on nominal decelerator area
D	Diameter
D_e	Diameter Roof Element
D_o	Nominal decelerator design diameter
g	Dynamic pressure
K_{RIA}	Reefing coefficient based on reefed inlet area
K_{RLL}	Reefing coefficient based on reefing line length (feet)
λ	Geometric porosity
M	Mach number
Psf	Pressure in lbs/ft^2
R_e	Reynolds number
x/d_b	Ratio of decelerator trailing distance to forebody diameter
ξ	Inlet reefing ratio (reefed inlet area to unreefed inlet area)

SECTION I - INTRODUCTION

The advent of high -speed and high-altitude flight necessary for both manned and unmanned flight vehicles has resulted in a requirement for supersonic and hypersonic first-stage decelerators. The role of the first-stage decelerator in supersonic/hypersonic recovery is one of stabilization and initial deceleration to flight conditions in which the final recovery device or method can be effective. The Air Force Flight Dynamics Laboratory has sponsored in-house and contracted efforts for developing full scale deployable decelerators for performance at supersonic and hypersonic velocities and dynamic pressures as high as 10,000 psf. The establishment of design data and performance of three decelerator types in these environments has been attempted and is reported herein. For high supersonic and hypersonic stabilization and first stage deceleration small parachutes and ram air-inflated balloon decelerators (ballutes) have been investigated. For possible supersonic high dynamic pressure applications the large reefed heavy duty ribbon parachute canopy has been considered. Areas of investigation for these decelerator types have included fabrication techniques, materials capabilities, wind tunnel testing and free flight testing.

SECTION II - SMALL SUPERSONIC PARACHUTES

1. GENERAL

The problems associated with providing a capability for first-stage deceleration and stabilization in supersonic flows using a parachute canopy have been studied by both contractors and the government. Early attempts consisted primarily of extending the capability of existing parachute configurations. When even the most promising of these configurations was found to be limited to operation at or below Mach 2, experimentation produced the HYPERFLO[a] family of parachute canopies. These are comprised of a solid conical skirt and a flat, porous roof. After extensive development, the HYPERFLO-type parachute was demonstrated close to Mach 4, with varying degrees of success. This canopy type has been refined further under the sponsorship of AFFDL and is known as the PARASONIC[b] canopy (see Figures 1A and 1B). The PARASONIC canopy is considered a hybrid of the HYPERFLO since it is comprised of a solid skirt and a porous roof. In lieu of the flat roof and conical skirt or inlet, the PARASONIC canopy has a shape very similar to the inflated shape of the HYPERFLO. It is this HYPERFLO hybrid or PARASONIC design that is being advanced by current programs for free-flight demonstration up to Mach 5.6.

In designing small supersonic parachutes, both analytical and empirical approaches have been attempted. The latest empirical approach, refinements, and correlation with experimental results are presented in Items 2 through 5, below.

---

[a] TM, Cook Electric Company, Chicago, Illinois.

[b] TM, Goodyear Aerospace Corporation, Akron, Ohio

2. CONFIGURATION SELECTION

a. General

Small-scale and full-scale wind-tunnel tests,[1-5,a] sled tests,[6] and free-flight tests[7] have indicated that some HYPERFLO configurations have inflation and aerodynamic stability behind forebodies over a range of Mach numbers. Small-scale models (D_o = 0.42 to 0.68 ft) constructed with roof panels of very fine mesh (Perlon) have shown satisfactory aerodynamic performance in the wind tunnel at low temperatures to Mach 5.5[1] Similar roof construction has been satisfactory in full-scale (D_o = 4 ft) wind-tunnel tests.[4,5] Aerodynamic heating precludes using the fine mesh nylon and Perlon materials at Mach numbers above 2.1.[7]

b. Roof Materials

Metal meshes and cloths have been investigated as roof materials, but fabrication and joining techniques are not sufficiently developed to maintain structural integrity. A Nomex mock leno mesh[8] roof material several times as coarse as Perlon mesh was developed for higher temperature operation. Coatings for this Nomex material were developed to raise temperature capability and to control porosity. Canopies (D_o = 4 ft) constructed with the coated and uncoated Nomex mesh were tested at the Arnold Engineering Development Center (AEDC) in the propulsion wind tunnel facility.[4] Performance of the canopy with the uncoated Nomex mesh was unsatisfactory because of excessive porosity, while a canopy using the coated Nomex mesh roof at the Mach number tested (M = 2.6) operated properly. During a later test program, two full-scale parachutes with coated Nomex mesh roofs were tested to Mach 3.5 These tests confirmed that canopies constructed with the coated Nomex roof performed equally or better aerodynamically than the prior conventional HYPERFLO canopies constructed with low-temperature Perlon roof mesh. These tests also confirmed the success of a refined parachute gore pattern that eliminated surplus fabric and wrinkles near the intersection between the skirt and roof.[8]

c. Wind-Tunnel Tests

Full-scale, wind-tunnel test results have shown that inflation stability is very Mach number dependent on the ratio of the measured inlet area to the measured equivalent exit area A_i/A_e. From the test data, the value of the area ratio that ensures inflation stability appears to be near the isentropic value. This relationship has been confirmed to Mach 3, which is the limit of full-scale wind-tunnel testing.[5] Figure 2 shows the relationship between area ratios designed into successful canopies and isentropic values of area ratio that yield sonic flow at the exit of an analogous duct.

Small-scale wind-tunnel tests to Mach 5.5 have covered a range of model sizes, forebody sizes, mesh materials, mesh locations, and canopy locations with respect to the forebody. Reduced porosity was shown to improve

---

[a]Superior numbers in the text refer to items in the List of References.

inflation characteristics from Mach 4 to 5.5; however, when this figure was attained by reducing the diameter of the roof mesh area, oscillations were seen to increase. Other units indicated reasonable performance with total porosities between 5 and 15 percent, with roofs constructed from 1/4-in. ribbon grids, and with selected forebody and downstream canopy locations. These tests indicated the feasibility of using a more coarse roof material. Inlet area to exit ratios between 2.58 and 7.74 were seen to be effective over the Mach range from 2 to 5.5.

d. Free-Flight Data

Table I summarizes the design and flight data for the four HYPERFLO-type Nomex parachutes that survived flights at Mach numbers greater than 2.1. Size, porosity, x/d_b, line length, and ribbon size were variables within the group of HYPERFLO test items. Between Tests 3 and 4, only the ribbon spacing was changed. The fourth HYPERFLO, tested at Mach 3.97, used 3/8-in. -wide webbing for the roof construction and had the lowest porosity of any parachute tested (7.5 percent compared with 10 percent and 14 percent for the other canopies.) This 7.5 percent porosity canopy was considered to perform better than the other free flight-tested canopies. The test item that gave the poorest performance of the four listed in Table I has a high porosity, was the smallest canopy (D_o = 2.71 ft), had 1.25-in. roof ribbons, and had an area ratio of 2.75 vs 3.6 suggested by the isentropic line of Figure 2 and 5.6 for the most successful free-flight item. These and other free-flight data[7] have suggested that the downstream position of the canopy skirt should be approximately seven diameters from the forebody base.

Review of the free-flight data indicates that fair to good flight performance can be achieved from Mach 2.84 to 4 with low porosity, 3.% to 4.12 D_o HYPERFLO-ribbon-roof parachutes. Low porosity and small ribbon size appear to improve the performance of the HYPERFLO. Low porosity is an important factor with increasing Mach numbers. Free-flight tests also have demonstrated the feasibility of using hemisflo canopies of conventional hemisflo ribbon design to Mach 3.44.[7]

Full-scale wind-tunnel tests to Mach 3 have shown that excellent performance is attained at an A_i/A_e of 4, and free-flight tests to Mach 4 report good performance is attained at an A_i/A_e near 5.6. It is concluded, therefore, that still lower porosities will be required at higher Mach numbers for good performance.

TABLE I - FREE-FLIGHT DATA SUMMARY OF SUCCESSFUL HYPERFLO-TYPE NOMEX PARACHUTES

No.	Test M	Psf	D_o(ft)	Total	Mesh	Ribbon width (in)	Line length	x/d_b	Performance	A_i/A_e
1	2.84	120	2.71	14.0	28.5	0.750	2D	7.80	Poor/fair	2.75
2	2.83	61	3.69	13.7	27.5	1.250	2D	10.40	Fair	2.85
3	3.22	345	3.69	10.8	22.0	0.375	2D	8.45	Fair	3.55
4	3.97	113	3.69	7.5	14.9	0.375	2D	8.45	Good	5.60

3. PARACHUTE CANOPY DESIGN CONSIDERATIONS

The goal of the current ADDPEP[a] programs in the small parachute area is to demonstrate free-flight capability to Mach 5.6 at 121,000 ft and at a dynamic pressure of 200 psf. A PARASONIC canopy has been designed for this test and the same design extended to three other test conditions (see Table II).

TABLE II - CURRENT PARASONIC PARACHUTE DESIGN CONDITIONS UNDER ADDPEP

Test M	Altitude (thousands of ft)	q (psf)	Re (millions)	C_{D_o} *
2.50	85.3	200	1.340	0.25
5.60	160.0	43	0.100	0.17
3.35	98.0	200	0.998	0.22
5.60	121.2	200	0.550	0.17

* Predicted nominal drag coefficient based on results of wind-tunnel and free-flight tests

The first step in the analytical design is the calculation of a point mass trajectory to establish time histories of the parametric values that describe the environment in which the decelerator must operate. Parameters of interest are dynamic pressure, Mach number, Reynolds

[a] AFFDL sponsored three phases entitled, "Aerodynamic Deployable Decelerator Performance Evaluation Program.

number, and temperature. Time histories of these parameters for the test points listed in Table II are included in Reference 9.

Small-and large-scale wind-tunnel tests and free-flight tests indicate that an A_i/A_e ratio of 7.74 provides good inflation to Mach 5.5 and reasonably good performance as low as Mach 2. With this in mind, an A_i/A_e of approximately 7 was chosen for the canopy design to meet the test objectives outlined in Table II.

During the first phase of ADDPEP, an analytical method was developed for predicting pressure distribution and resultant loading on PARASONIC-type parachute canopies.[8] This method considers the canopy analogous to a one-dimensional duct, with normal shock relationships in front of the inlet and sonic flow through the roof. The flow conditions ahead of the inlet are free stream values. The pressure in the roof or mesh area can be estimated based on the sonic condition present there. The pressures acting on the solid shirt area are estimated by considering the area as a ring airfoil as explained in Reference 8. The static pressure inside the canopy is determined from the local flow conditions. The effect A_i/A_e on the base pressure values behind closed roof areas is estimated as outlined in Reference 9.

Static (fully inflated) structural analysis of flexible structures is dependent on the distribution of applied loads and the resulting decelerator shape. The final shape for a PARASONIC parachute design is obtained from an iteration process that matches both aerodynamic shape requirements and the aerodynamic load distribution with a structurally stable shape. This process was developed under programs sponsored by the AFFDL and is included in Reference 8 as Appendixes II and III.

4. THERMAL ANALYSIS

A thermal analysis of the parachute to be deployed at Mach 5.6 at 121,000 ft. was conducted to determine the temperature rise that can be expected on an element of yarn positioned in the roof of the parachute under the most severe flight conditions. A temperature evaluation of the skirt and lines was not conducted. Instead, a comparison of thermal environment and the thickness of the material with the roof-panel conditions showed that the temperature along the skirt would not reach a critical value over the deceleration time period. The procedures used to analyze the roof element are outlined in Reference 8. Near stagnation conditions are considered to exist inside the parachute canopy subject only to the mass flow rate of the air flowing from the parachute envelope through the orifices in the porous roof. Since the ratio between the total pressure inside the parachute and the local pressure on the back surface of the roof is greater than critical, sonic flow is considered to exist through the porous roof.

The elements in the parachute canopy are shown in Figure 3. The roof is composed of a matrix of yarn elements so the porosity can be controlled by the amount of yarn elements used. The lower sketch of Figure 3 is an enlarged view of the roof surface and is used as a flow model for analysis

SECTION III - LARGE REEFED SUPERSONIC PARACHUTES

1. GENERAL

a. Goal

The goal of the large supersonic parachute effort was establishing the parachute configuration, loadings, design, and structural materials for attaining an initial load of 200,000 lb at supersonic speeds by appropriately reefing the canopy. The final canopy size was chosen as 16 ft. D_o. The performance range of interest extended from transonic speeds to Mach 3 and a dynamic pressure of 10,000 psf. The flight test efforts were conducted in two series, the first series of three in 1965 and the second series of three in 1966. The first series extended to Mach 1.8 and the second series to Mach 2.6.

The state of the art at the program's start was to consider a 25 to 30 percent porosity hemisflo or reefed conical ribbon parachute for the transonic and low supersonic flight regime. However, flight-test and wind-tunnel data indicated that these parachutes had inflation instability problems and gross ribbon flutter when highly reefed at the higher Mach numbers. Because of inflation instability, drag prediction was difficult above transonic speeds, and the average drag load was considerably less than the drag load peaks that design the system.

b. Investigation of Canopy Configurations

Tests were conducted in the propulsion wind-tunnel facility at AEDC to investigate different canopy configurations to Mach 3 while increasing the degree of reefing as the Mach number was increased, 4,5 These tests indicated that inflated reefed shapes could be attained with higher degrees of reefing by decreasing canopy porosity. These data also indicated that a low-porosity, extended-skirt hemisflo with $2D_o$ lines and midgore reefing had the best inflation stability and predictable drag coefficient values at supersonic speeds (Mach 1.8 to 3.0) with appropriate reefing sizes. Since the overall program goal was to extend the reefed canopy operation to Mach 3, the low porosity hemisflo was chosen as the configuration for all tests. In addition one canopy construction, set of materials, and strengths were selected based on ultimate testing at Mach 3 and 10,000 psf. This led to the use of Nomex materials for all ribbons, verticals, and sewing thread. Nylon was retained for the suspension lines, skirt band, and vent band due to its better-known energy absorption characteristics. The relatively large thickness of these members minimized heating effects at the initial peak loads.

To obtain good strength-to-weight ratios and a flexible canopy for good packing densities, a continuous horizontal ribbon design was used with only one joint per ribbon. The ribbons were sandwiched between dual, continuous-suspension/radial lines that ran from one missile pin across the canopy vent to the opposite missile pin. Ribbon strength was reduced

An equation developed by Bartz[10] is used for estimating the heat transfer coefficient in the roof orifice, assuming there is turbulent flow. The range of this coefficient is calculated as a function of the trajectory. The heat flux rate to the yarns surrounding the roof orifice is then calculated.9 Once the heat flux rate is known, Reference 8 of Appendix I can be used to estimate the required diameter of the roof elements. Reference 8 reports the results of testing various coatings and materials. The most efficient thermal protective coating evaluated was Dyna-Therm D-65, a product of the Dyna-Therm Chemical Co. If the coating test results and the cold wall heat flux are used, the diameter of a roof element can be calculated using Equation I-6 of Reference 8, Appendix I. This diameter and the heat flux rate to a roof element are then combined to calculate the temperature response of the roof on the basis of transient radial heat flow into a solid cylinder. The results of this temperature response evaluation for deployment conditions of Mach 5.6 at 121,000 ft and for a Nomex roof element 0.1 in. in diameter coated with Dyna-Therm D-65 to 0.15 in. in diameter are given in Figure 4.9 The curves in Figure 4 result from a computer iteration process described in the reference. The surface temperature of the roof element rises to 1000 F during the first second after deployment and then continues to rise at a much slower rate to about 1120 F at 2.5 sec after deployment. Thereafter, the temperature decreases gradually to approximately 600 F at 20 sec after deployment. The temperature felt by the Nomex material, which is protected by the Dyna-Therm D-65 coating, rises gradually during the first 10 sec to about 520 F. The temperature then continues to rise very slowly while the trajectory continues into less dense air. This study indicated that the critical temperature for the Nomex material was not reached.

Earlier parachute designs [11,12] indicated that the skirt section, leading edge lip, and riser lines should adequately withstand the local heat input provided they are locally coated with a thin layer of protective coating. The buildup of material at the inlet lip alleviates much of the heating problems in this area.

5. PARACHUTE CONSTRUCTION AND DESIGN REFINEMENTS

Fabrication of parachute roof panels from coated yarns 0.15-in. thick presents mesh or grid weaving problems as well as problems in seaming between gores and between the roof and skirt portions of the canopy. Attempts to stitch components woven from this stiff and heavy yarn have proved futile. Interweaving of the roof gores by looping the yarns together at the gore edges also was unsatisfactory. The final approach was to weave, by hand, the entire roof as a unit. This hand weaving used a continuous yarn that was always oriented at 45 F to the gore centerline. A three-dimensional surface was developed upon which the roofs were woven.

The full-scale wind-tunnel tests conducted in August 1965 indicated that failures in the confluence point for the suspension lines occurred due to high-frequency variation of suspension-line force. This problem has been overcome by installing a confluence keeper ring approximately 3 in. in diameter at a point 20 in. from the confluence point.

in groups as the distance from the canopy vent increased.

The reefing line lengths were determined for the first series of tests by analytical approaches and wind-tunnel data. The analytical approach was used because a 2 to 1 ratio of static drag coefficient values existed between available wind-tunnel data and the corresponding opening shocks when associated with infinite mass conditions.[a] Drag traces also indicated a 50 percent variation from the average static values. The analytical approach was evaluated first by comparing its predictions for conical ribbon canopies at deployment conditions similar to the first test point. The correlation of predicted and actual loads was good, and the analytical approach was used for predicting the reefing line lengths for the first two ADDPEP tests. The results from the first flight test confirmed the selected reefing line length of 20.5 ft. This analytical approach will be rigorously presented in the final reports of the ADDPEP Phase II and III programs. The approach predicts filling time, reefing line length, and total force. The drogue parachute selected for deploying the system for this first series was a four-foot D_o hemisflo similar in geometry to the test parachute. Different degrees of permanent reefings were used in an attempt to maintain drogue deployment forces.

Results of each test in the first series were evaluated to establish any changes necessary for the following tests. No changes were required relative to the canopy design, materials, or constructions. Changes were made to increase the reefing line strength and to add protection locally to the suspension lines to prevent line damage during deployment.

The second series of tests incorporated these changes, and the drogue parachute was reduced in size from 4 ft-D_o to 2.5 ft-D_o. It was reefed in an attempt to maintain drogue deployment forces at the higher dynamic pressures. No changes were made relative to the canopy during the test series. However, reefing line strength was increased, additional line protection was added, and the strength of the deployment bag was increased for the highest dynamic pressures.

2. TEST DATA

a. Configuration Stability

Canopy inflation stability and system stability are demonstrated in high-speed color movies taken at 500 and 700 frames per second by two test vehicle-mounted cameras viewing directly to the rear. The camera with the slower film speed is mounted near the test vehicle centerline. The higher speed camera was mounted near the test vehicle centerline for the first flight. It was moved to the outside of the test vehicle for all later flights. As can be seen from the first test film, the canopy has a firm shape when reefed and when disreefed in the transonic range. During later flights, the reefed configuration is shown throughout supersonic and transonic regime. A firm canopy shape is maintained again throughout the test regime.

---

[a] Weight of entire system greater than 30 times the decelerator drag area (C_dA)

b. Loads

The reefing line length for the desired peak load was originally estimated using past data and the calculated approach, which preserved continuity during the inflation process. The flow-in was based on stream conditions, an instantaneous inlet area, and a canopy exit area. The exit area was based on the canopy porosity and the surface area of the inflated hemisphere at that instant. A conical shape from the hemisphere to the inlet with surface lengths preserved formed the canopy configuration. Static drag values were estimated and the force associated with the acquired mass term was calculated. Based on the calculations, a reefing line length of 20.5 ft was established for the first test. The smooth load traces for the first and following tests indicate the deployment, opening, and inflated load characteristics of the canopy (see Figures 5 and 6).

During the program the measured loads were used to aid in a more accurate prediction of reefing line lengths for a given load at more severe test conditions. Three types of curves were considered, one of which is a collection of flight-and wind-tunnel data and is plotted as C_D versus a reefing parameter (see Figure 7). The second is coefficients based on flight test peak load/q values versus Mach number and uses reefing line lengths and reefed inlet areas as references (see Figure 8). The third is predicted peak loads/q calculated from extrapolated coefficients (see Figure 9). The test points are indicated on the figures. The curves indicate the same shape of the C_D versus Mach number for flight and tunnel tests. The free-flight data indicate higher values that may be due to the better inflation stabilities observed in free flight.

Peak loads, opening times, and opening shock factors are presented in Table III.

c. Design

The canopy is composed of 32 gores and has a 16-ft D_o. It is a 10 percent extended skirt hemisflo, 14 percent nominal porosity canopy with the equivalent of $2D_o$ lines. Midgore reefing[a] was used for all tests. Nomex ribbon strengths and spacings are presented in Table IV.

[a] Reefing Rings placed on gore centerline at canopy skirt as opposed to conventional location at suspension line.

TABLE III - LOAD DATA

Test	Reefing line lengths (ft)	Deployment		Opening time (line stretch to load max), sec	Max load (lb)	Opening shock factor	Comments
		M	q, psf				
LP-3	20.50	1.50	1988	0.130	154,500	1.48	Broke 12,000 lb reefing line 0.10 sec after peak load; no other damage
LP-4	14.83	1.63	2118	0.160	108,850	1.18	Line twist near missile, several lines broken; no other damage
LP-5	14.83	1.84	3383	0.200	152,600	1.26	Line twist near missile, several lines damaged; no other damage
LP-7	13.10	2.22	3697	0.120	112,000	. . .	Broke 18,000 lb reefing line and separated canopy near inflated reefed shape
LP-8	11.00	1.20	1514	0.200	25,700	. . .	No damage; reefed for a higher test condition
LP-9	11.00	2.80	5640	0.095	120,000	. . .	Broke 24,000 lb reefing line and separated canopy upon inflation

TABLE IV - RIBBON STRENGTHS AND SPACING

Ribbon no.	Strength (lb)	Ribbon spacing (in.)
1 (vent)	2 - 8700 Nylon	. . .
2 - 8	2 - 3000 Nomex	0.465
9 - 14	1 - 4000 Nomex	0.465
15 - 20	1 - 3000 Nomex	0.465
21	1 - 3000 Nomex plus 1 - 6000 Nylon reinforcement	0.465
22 - 28	1 - 3000 Nomex	0.430
29	1 - 2000 Nomex plus 1 - 6000 Nylon reinforcement	0.430
30 - 41	1 - 2000 Nomex	0.430
42 (skirt)	2 - 10,000 Nylon	0.430

d. Significant Observations

Based on the free-flight tests, the following results were obtained:

Configuration

1. Confirmation of small and large-scale wind-tunnel results that this extended skirt hemisflo canopy when reefed has excellent inflation stability at supersonic speeds

2. Confirmation of small and large-scale wind-tunnel results that this canopy when reefed provides good to excellent system stability at supersonic speeds

3. That this canopy when fully open has excellent inflation stability at transonic and subsonic speeds

4. That this canopy when fully open provides excellent system stability at transonic and subsonic speeds

5. That this canopy has minimal ribbon flutter and canopy oscillation in either the reefed or fully open position over the test range.

Loads

1. Confirmation of predictable loads and smooth load traces with a canopy with excellent inflation stability

2. Establishment of loads, opening times, and opening shock factors for this extended skirt hemisflo canopy at deployment speeds to Mach 1.8; partial data to Mach 2.6

3. That line protection is required in the region of the locking loop to prevent several individual lines from breaking during deployment at Mach 1.8; increased line protection is required at Mach numbers greater than Mach 1.8 for high loads

Design Approach

1. That this canopy design is adequate for measured loads of 155,000 lb at Mach 1.8; inspection indicated that higher loads are within the canopy's capability at this or lower Mach numbers

2. That the basic canopy materials, seams, joints, and attachments are adequate for the loads and the environment to Mach 2.4

3. That Nomex materials can withstand the high loadings and loading rates associated with canopy verticals and ribbons

4. That to attain the high loads above Mach 1.8 the suspension lines require additional protection from deployment damage

5. That the cutter system is adequate for reefing line strengths to 18,000 lb (N21979 reefing cutter assembly, Sandia Corp, Albuquerque, N.M., with an OA-A7 firing mechanism formerly manufactured by Ordnance Associates and now fabricated by Hercules Powder Co., Wilmington, Del.)

6. That the original deployment bag and tie system, described in the ADDPEP Phase II Progress Report, is adequate to Mach 1.8.

7. That the increased strength deployment bag, also described in the ADDPEP Phase II Progress Report, is adequate to Mach 2.6

SECTION IV - BALLUTES

1. GENERAL

The goal of the BALLUTE (see Figure 10) effort was to establish the configuration, loadings, design, and structural materials for flight speeds to Mach 10 and for dynamic pressures to 500 psf. Flight test efforts were conducted in 1965 and 1966.

Early efforts established analytical techniques for determining the shape and static loadings for a pressurized membrane structure.[8] The areas for further investigation included determining:

1. Internal pressure coefficients

2. Means for extending the performance limits of Nomex BALLUTES

3. Structural designs of metal BALLUTES and means to reduce deployment dynamics

In-plant, wind tunnel, heat tunnel, and flight tests were conducted in conjunction with analysis during these investigations. Internal pressure coefficients were of basic interest, since the most efficient BALLUTE structure design considers the internal and external pressures. It was

found for trailing decelerators that, as the Mach number was increased the internal pressure coefficient increased and reached a plateau above Mach 4 for the inlet locations tested (see Figure 11).

These data were obtained using solid models and two deployable flight test units. The solid models were tested at the AEDC and at NASA Langley. Data from the flight test units indicated higher pressure coefficient values were obtained than the values obtained by using the smaller models behind simulated test missiles. Based on these results, present trailing BALLUTE configurations have higher than necessary internal pressures for attaining their design shapes at Mach numbers greater than 2.5. Conversely, the pressure distribution data also indicates that more blunt BALLUTE shapes are usable as the flight Mach number is increased from 2.5 to 4.

For BALLUTES alone, the pressure coefficients are lower, more nearly approaching the values desired for present shapes and minimum weight structure. Larger weight savings are possible if the performance limits of Nomex BALLUTES are extended by using techniques that keep the fabric temperatures in a high strength-to-weight ratio during critical loads. Techniques investigated included coatings and internal cooling by fluids.

Some coatings investigated gave nearly an order-of-magnitude increase in cloth life times when subjected to a Mach 8 stream in the Goodyear Aerospace heat tunnel. The test conditions and test results for a typical BALLUTE material and several candidate coatings are presented in Table V.

found for trailing decelerators that, as the Mach number was increased the internal pressure coefficient increased and reached a plateau above Mach 4 for the inlet locations tested (see Figure 11).

These data were obtained using solid models and two deployable flight test units. The solid models were tested at the AEDC and at NASA Langley. Data from the flight test units indicated higher pressure coefficient values were obtained than the values obtained by using the smaller models behind simulated test missiles. Based on these results, present trailing BALLUTE configurations have higher than necessary internal pressures for attaining their design shapes at Mach numbers greater than 2.5. Conversely, the pressure distribution data also indicates that more blunt BALLUTE shapes are usable as the flight Mach number is increased from 2.5 to 4.

For BALLUTES alone, the pressure coefficients are lower, more nearly approaching the values desired for present shapes and minimum weight structure. Larger weight savings are possible if the performance limits of Nomex BALLUTES are extended by using techniques that keep the fabric temperatures in a high strength-to-weight ratio during critical loads. Techniques investigated included coatings and internal cooling by fluids.

Some coatings investigated gave nearly an order-of-magnitude increase in cloth life times when subjected to a Mach 8 stream in the Goodyear Aerospace heat tunnel. The test conditions and test results for a typical BALLUTE material and several candidate coatings are presented in Table V.

Cooling techniques were developed during the metal BALLUTE fabrication investigation. Based on the results of the coating and cooling investigations, Nomex cloth BALLUTES coated with one of these materials and with an internal cooling system will be flight tested to Mach 10 during the summer of 1966.

2. METAL CLOTH BALLUTES

During the fabrication investigation of metal cloth BALLUTES, many specimens and two full-scale prototype models were fabricated. The welded specimens ranged from simple lap seams to joints to simulated subassemblies. Some subassemblies were tested in a subsonic wind tunnel while attached to a cloth flag to establish structural integrity of the system during the early inflation process. Wind-tunnel tests of the first two BALLUTES made of stainless steel cloth B[a] indicated that any local damage during deployment can lead to total damage.

Several factors caused the BALLUTE design to be critical during inflation with this material at minimum tunnel test conditions (Mach 2.8 at 120 psf). These factors included marginal cloth strength, lower cloth strength-to-weight ratios than textiles (approximately 1/5), and the inability of metals to withstand repeated high local elongations. Design improvements were made as a result of wind-tunnel tests of the two prototype and the flag-mounted structures. These included mechanical design improvements and developing a rapid partial inflation technique to reduce flagging loads. The rapid partial inflation technique utilizes a water/alcohol mixture that quickly vaporizes at test altitudes and fills the BALLUTE from 1/2 to 2/3 its full size. With this enclosed volume, the metal cloth surfaces are prevented from gross movements and inversions after line stretch.

During the development of the rapid partial inflation technique, several vacuum chamber tests and one flight test were made using textile BALLUTES. A rapid transfer of heat occurred between the BALLUTE fabric and the vapor. Results from flight testing a Nomex BALLUTE with thermocouples imbedded in the cloth indicated a significant drop in cloth temperatures upon deployment. The flight test conditions were Mach 4 at 98,500 ft on an ascending trajectory (γ = 80 F). Three thermocouples measured the cloth temperature. They were located along a gore, with the third thermocouple located at a position corresponding to 1/2 of the BALLUTE diameter and maximum surface slope. The values versus time starting at deployment are presented in Figure 12. The liquid solution used was 12 oz of methyl alcohol and 4 oz of water. Since the flight was short and the BALLUTE was packed in a dual wall container, it was anticipated that the contained fluid would maintain its temperature until deployment. The energy available for vaporization is (1) the sensible heat in the solution itself, as a liquid and with the water becoming ice; (2) heat in the fabric; and (3) heat in the incoming air. The heat required to vaporize methyl alcohol is 482 Btu's/lb, or 12/16 X 482 = 362 Btu's for test. To vaporize water,

---

[a] Cloth B - 304 stainless steel, 100 X 100 strands per inch; each strand consists of seven filaments of 0.0016-in. diameter wire. Cloth strengths: 150 lb/in. at 70 F and 45 lb/in. at 1000 F.

1000 Btu's/lb are required, or 4/16 X 1000 = 250 Btu's. The exact sequence of vaporization has not been established; however, the following action sequence is postulated (see Figure 13):

1. Alcohol vaporizing absorbs heat from alcohol liquid (A toward B)

2. Alcohol vaporizing and alcohol liquid absorbs heat from water (A to B)

3. Alcohol vaporizing and alcohol liquid absorbs heat from water freezing (A to B to C)

4. Vapor and liquid absorb heat from the fabric by convention (B-C to D); all alcohol is vaporized at 70 deg

5. Convection heats 12 oz of alcohol vapor, and BAL LUTE pressure becomes 460 psf at 600 deg

The same type of temperature decrease was demonstrated during the inflation of the metal BALLUTE during the tunnel tests.

3. RESULTS

BALLUTE designs have been completed for a broad spectrum of deployment conditions ranging from Mach 2 at 4000 ft to Mach 10 at 225,000 ft. Testing has been accomplished in the performance range between Mach 1.5 at 4000 ft and Mach 4 at 98,500 ft. The remaining regime is scheduled for free-flight tests this year.

Fabrication techniques for coated textile construction and stainless steel have been proved by wind-tunnel or free-flight tests. Supersonic drag data for a BALLUTE with a burble fence (see Figure 10) indicate a free stream C_D of one based on a nominal BALLUTE diameter. Values of $C_D \cong 0.8$ are obtained at x/d suitable for stable performance behind a variety of symmetrical and unsymmetrical forebodies.

Development of the BALLUTE concept has led to its application in high-performance trajectory control for the Prime and ALARR projects. Further developments in metal cloth application, flexible coatings, and chemical cooling show great promise in solving thermal problems associated with higher performance. Dynamic design of BALLUTES up to six feet total diameter has been demonstrated.

REFERENCES

1. Deitering, J. S., and Hilliard, E. E.: Wind Tunnel Investigation of Flexible Aerodynamic Decelerator Characteristics at Mach Numbers 1.5 to 6. AEDC-TR-65-110, June 1965.

2. Deitering, J. S.: Wind Tunnel Investigation of Flexible Parachute Model Characteristics at Mach Numbers 1.5 to 5. AEDC-TR-63-263, January 1964.

3. Deitering, J. S.: Performance of Flexible Aerodynamic Decelerators at Mach Numbers 1.5 to 6. AEDC-TDR-63-119, July 1963.

4. Lowry, J. F.: Aerodynamic Characteristics of Various Types of Full Scale Parachutes at Mach Numbers 1.8 to 3.0. AEDC-TR-64-120, June 1964.

5. Reichenan, D. E. A.: Aerodynamic Performance of Various Hyperflo and Hemisflo Parachutes at Mach Numbers from 1.8 to 3.0. AEDC-TR-65-57, March 1965.

6. Pedersen, P. R.: Study of Parachute Performance and Design Parameters for High Dynamic Pressure Operation. AFFDL-TDR-64-66, May 1964.

7. Nickel, W. E., and Sims, L. W.: Study and Exploratory Free-Flight Investigation of Deployable Aerodynamic Decelerators Operating at High Altitudes and at High Mach Numbers. AFFDL-TDR-64-35, Vol. 1, July 1964.

8. Nebiker, F. R.: Aerodynamic Deployable Decelerator Performance Evaluation Program. AFFDL-TR-65-27, May 1965.

9. GER-12286: Analysis and Design of Small Parachute Test Items. Akron, Ohio, Goodyear Aerospace Corporation, September 1965.

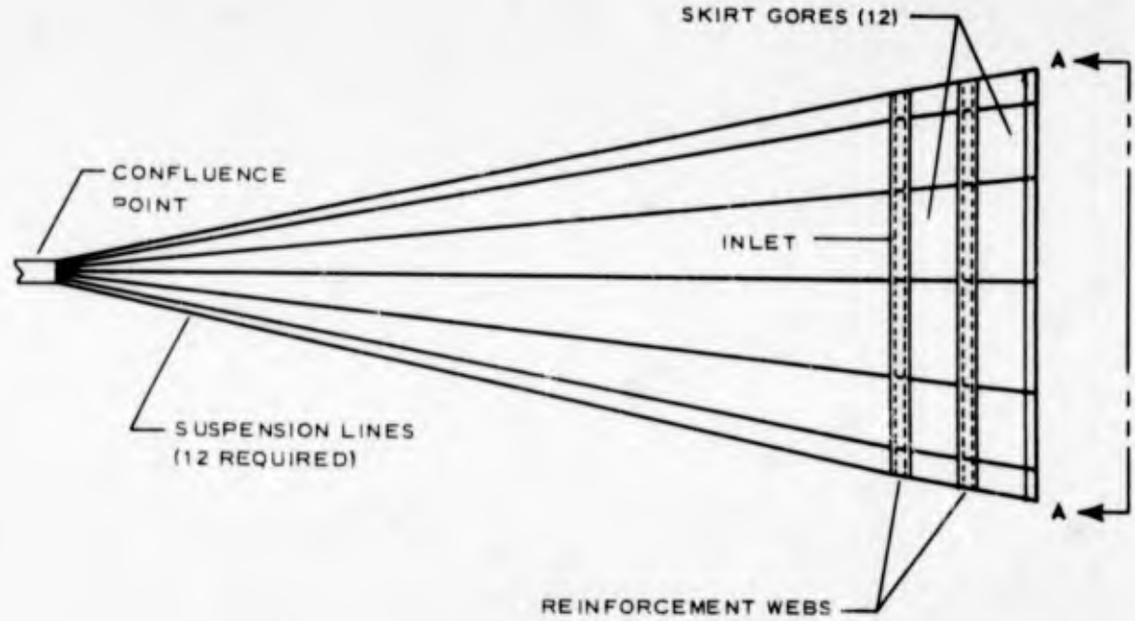
10. Bartz, D. R.: "A Simple Equation for Rapid Estimation of Rocket-Nozzle Convective Heat Transfer Coefficients." Jet Propulsion, Vol. 27, No. 1, pp 49-51, January 1957.

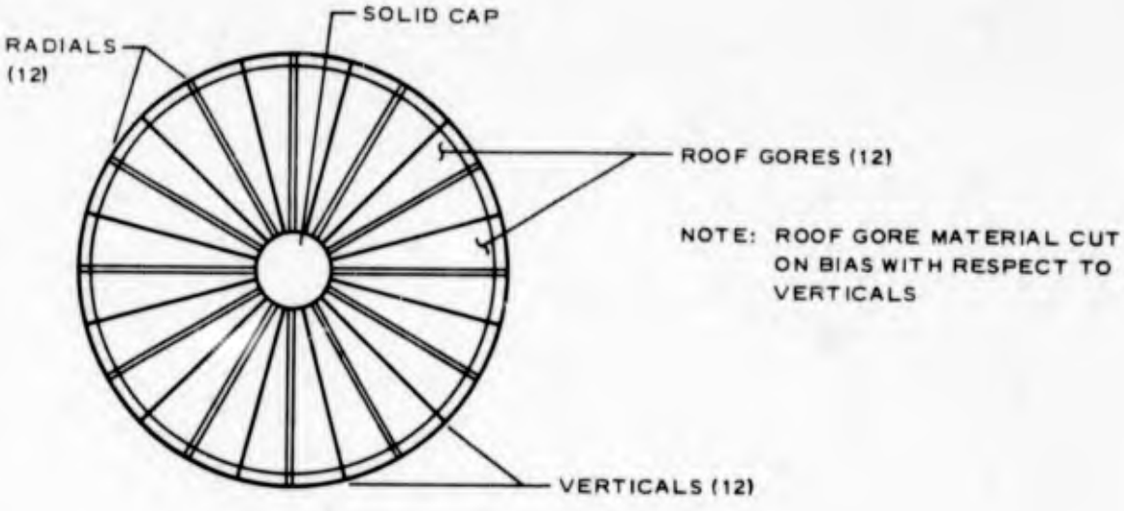
11. GER-11317: Hyperflo Test Item SP-3. Akron, Ohio, Goodyear Aerospace Corporation, November 1963.

12. GER-11415: Hyperflo Test Item SP-5. Akron, Ohio, Goodyear Aerospace Corporation, January 1964.

13. GER-12323: Aerodynamic Deployable Decelerator Performance Program - Large Test Items. Akron, Ohio, Goodyear Aerospace Corporation, December 1965.

BLANK PAGE





VIEW A–A, ROOF

Figure 1A - Small Hyperflo Parachute Configurations

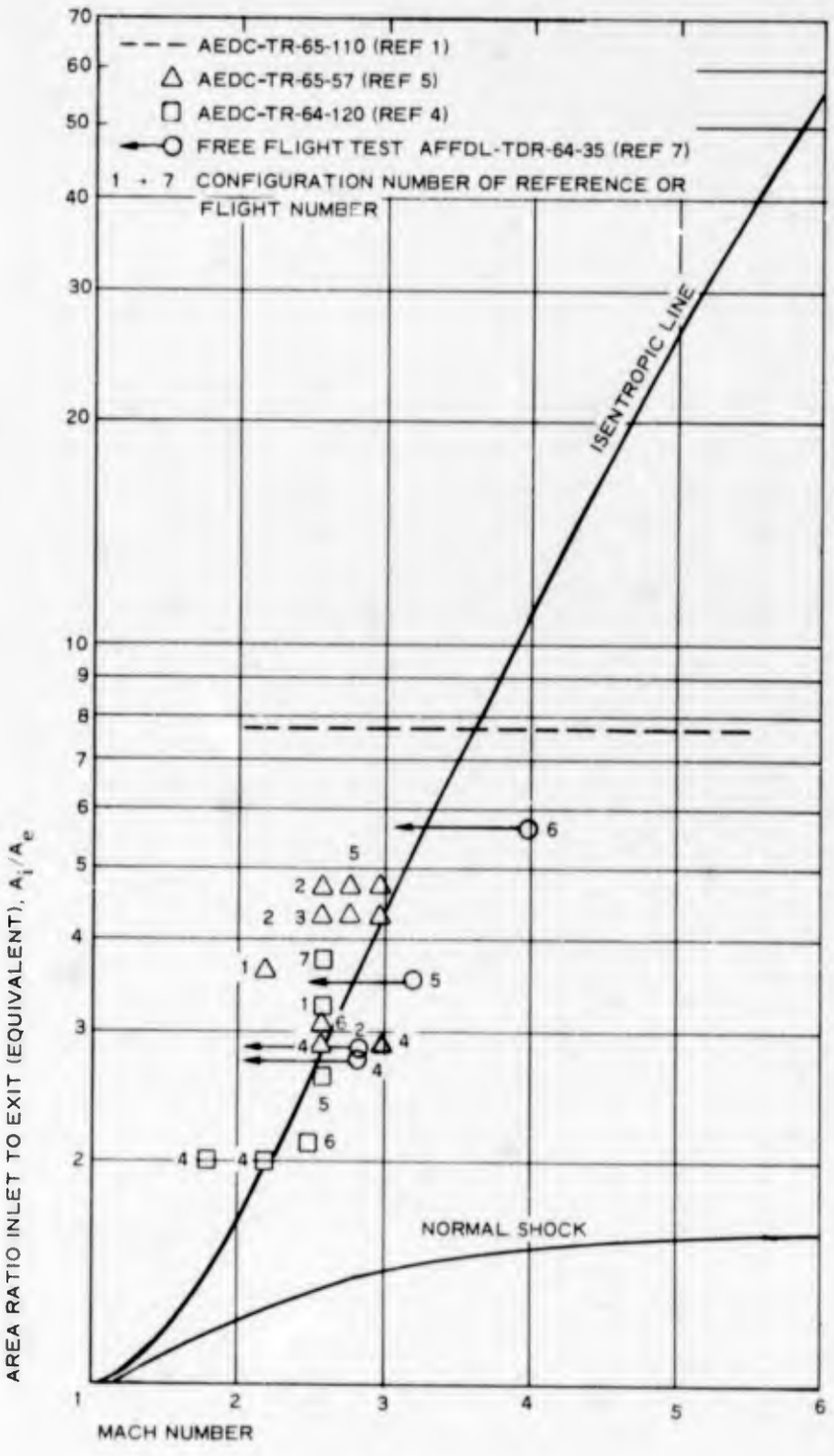


Figure 2 - Inlet to Exit Area Ratio versus Mach Number

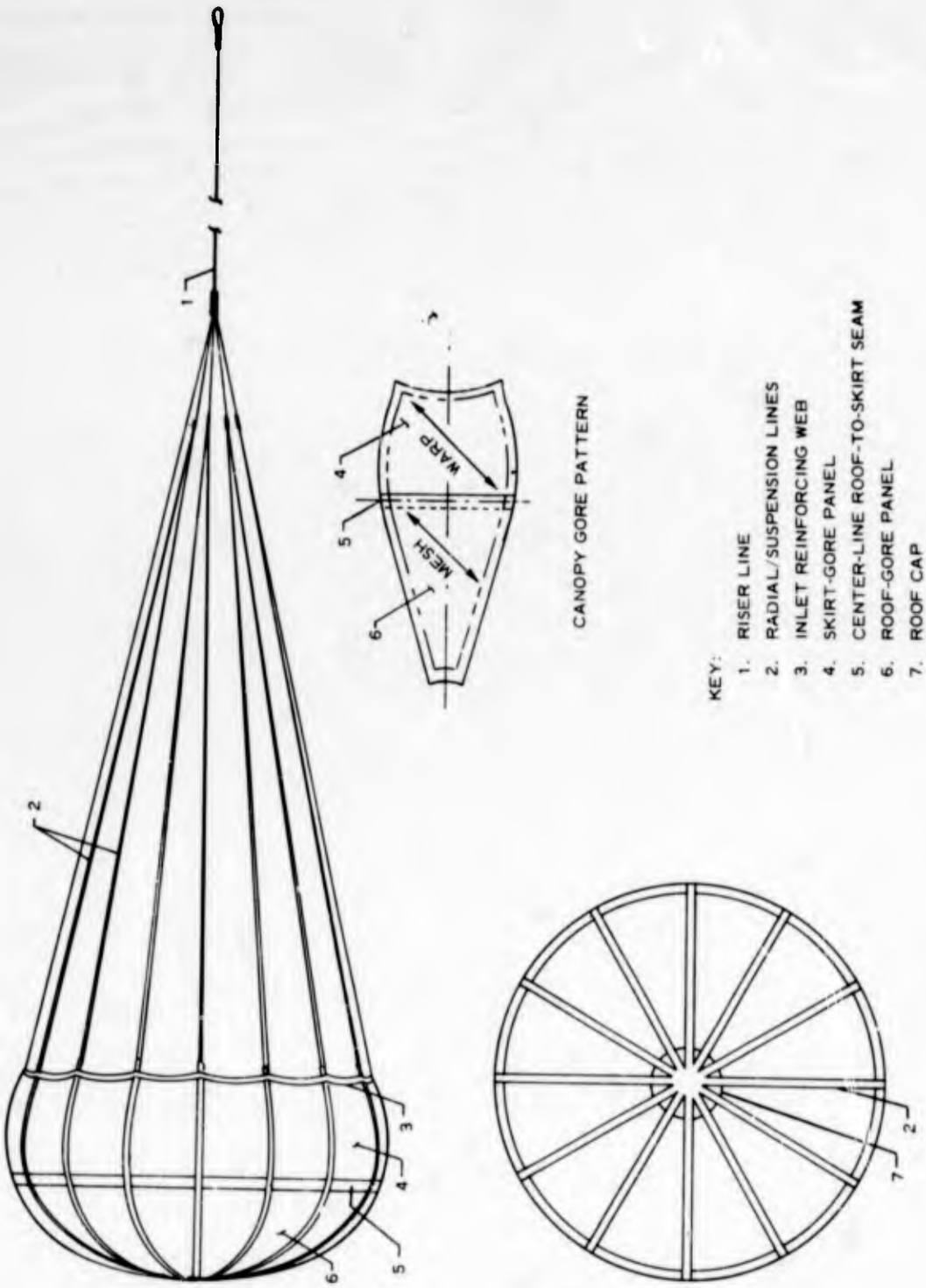


Figure 1B - Small Parasonic Parachute Configurations

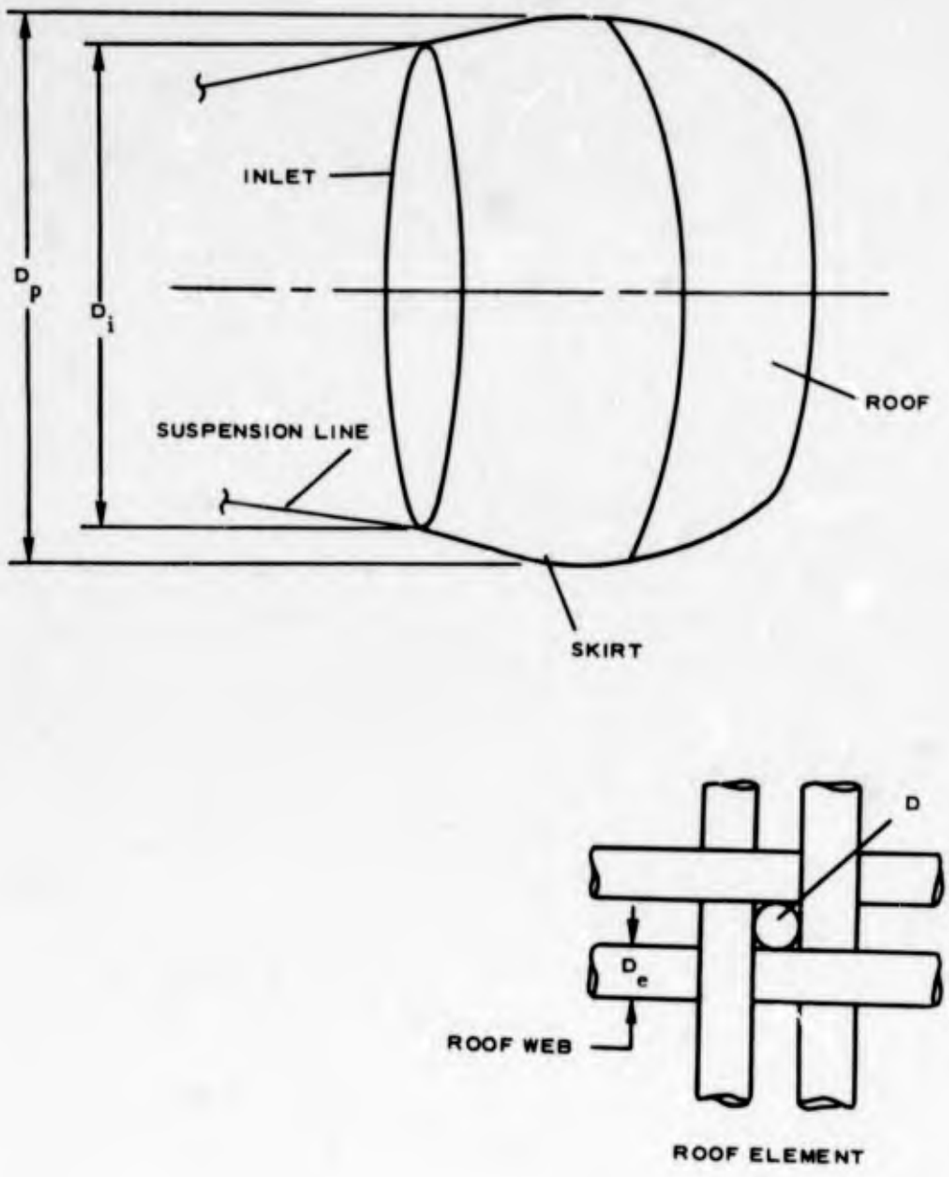


Figure 3 - PARASONIC Parachute Configuration and Roof Element

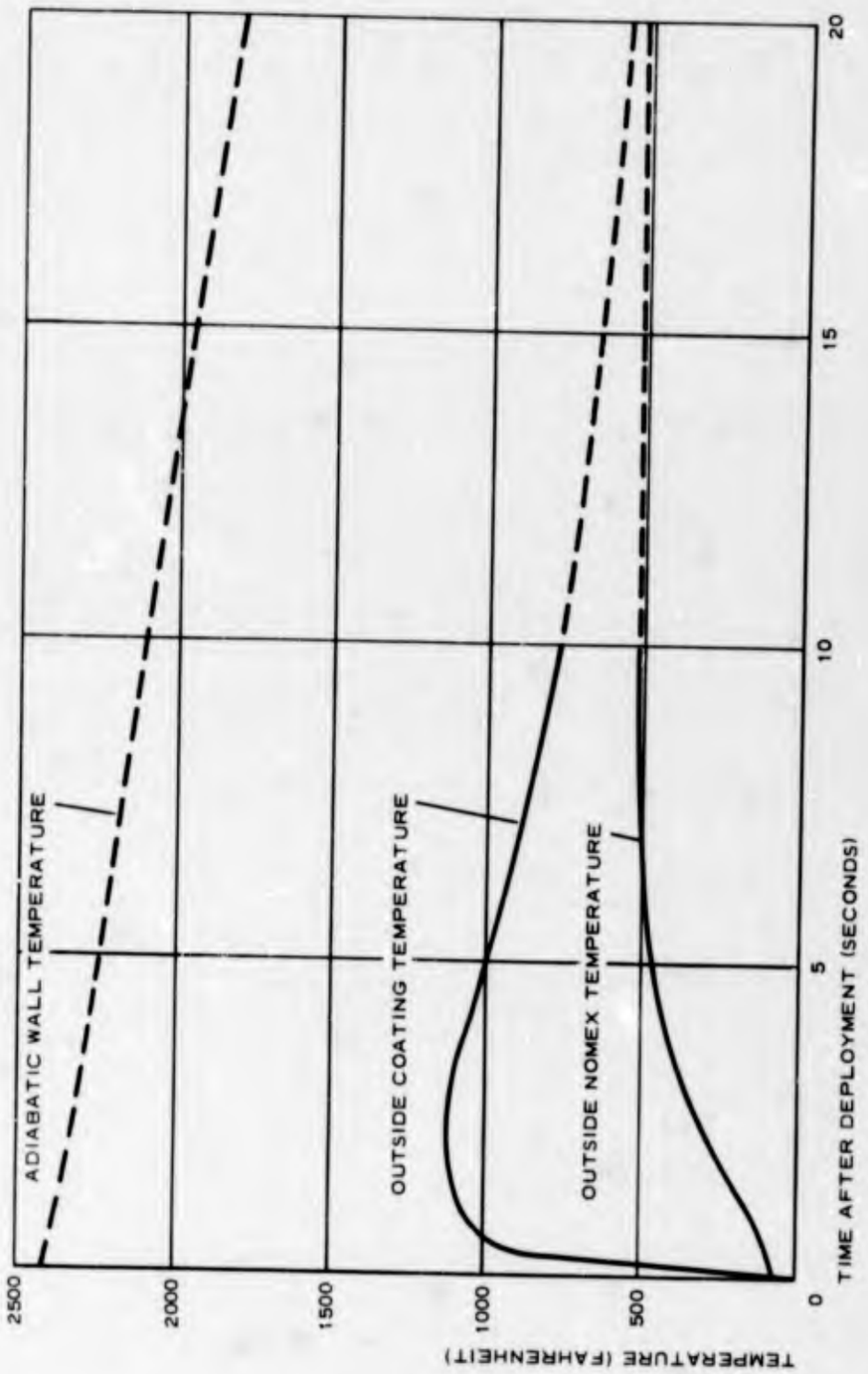


Figure 4 - Temperature Response of Coated Nomex Roof Element Parasonic Design for Mach 5.6 at 121,000 Ft

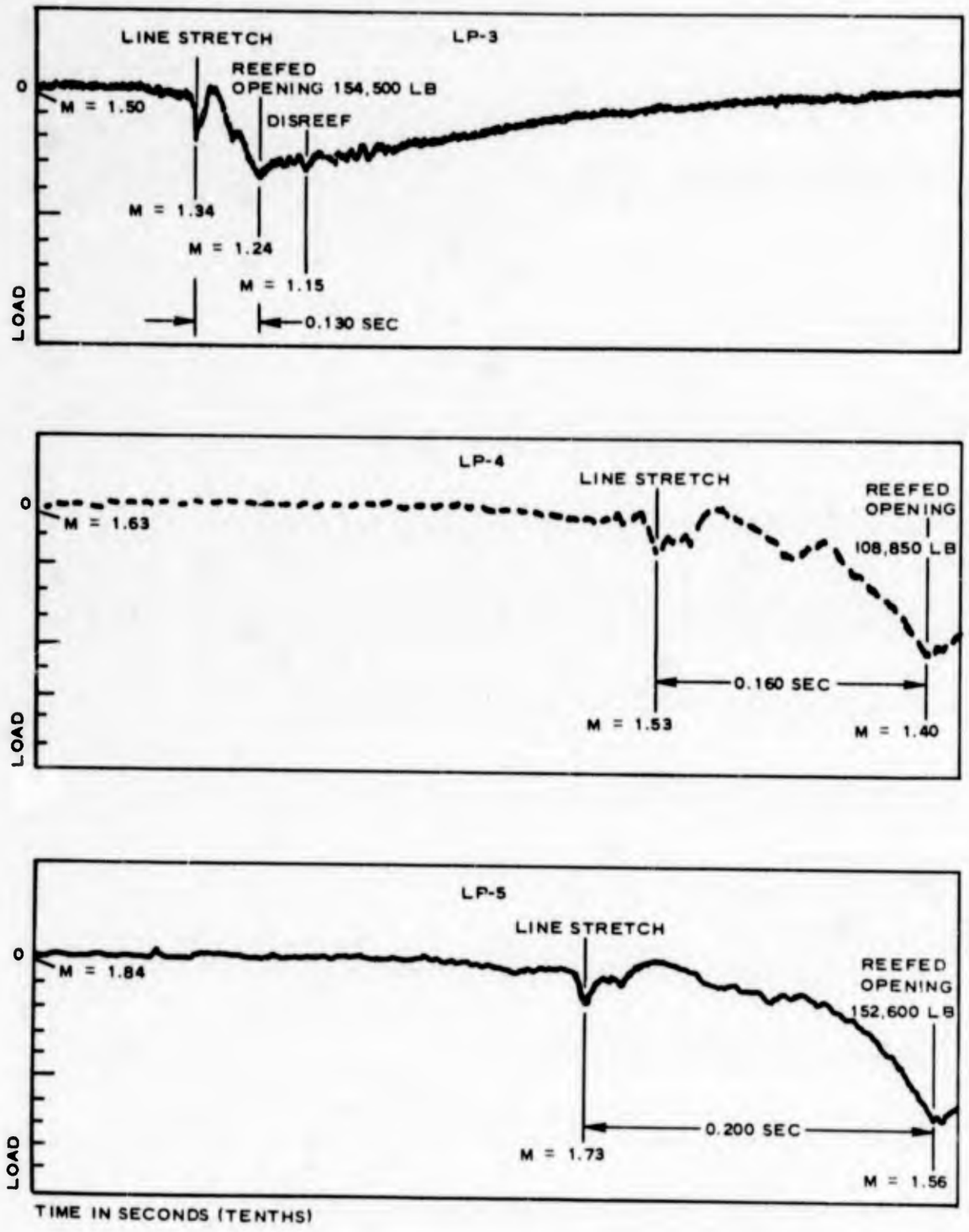


Figure 5 - Load Traces

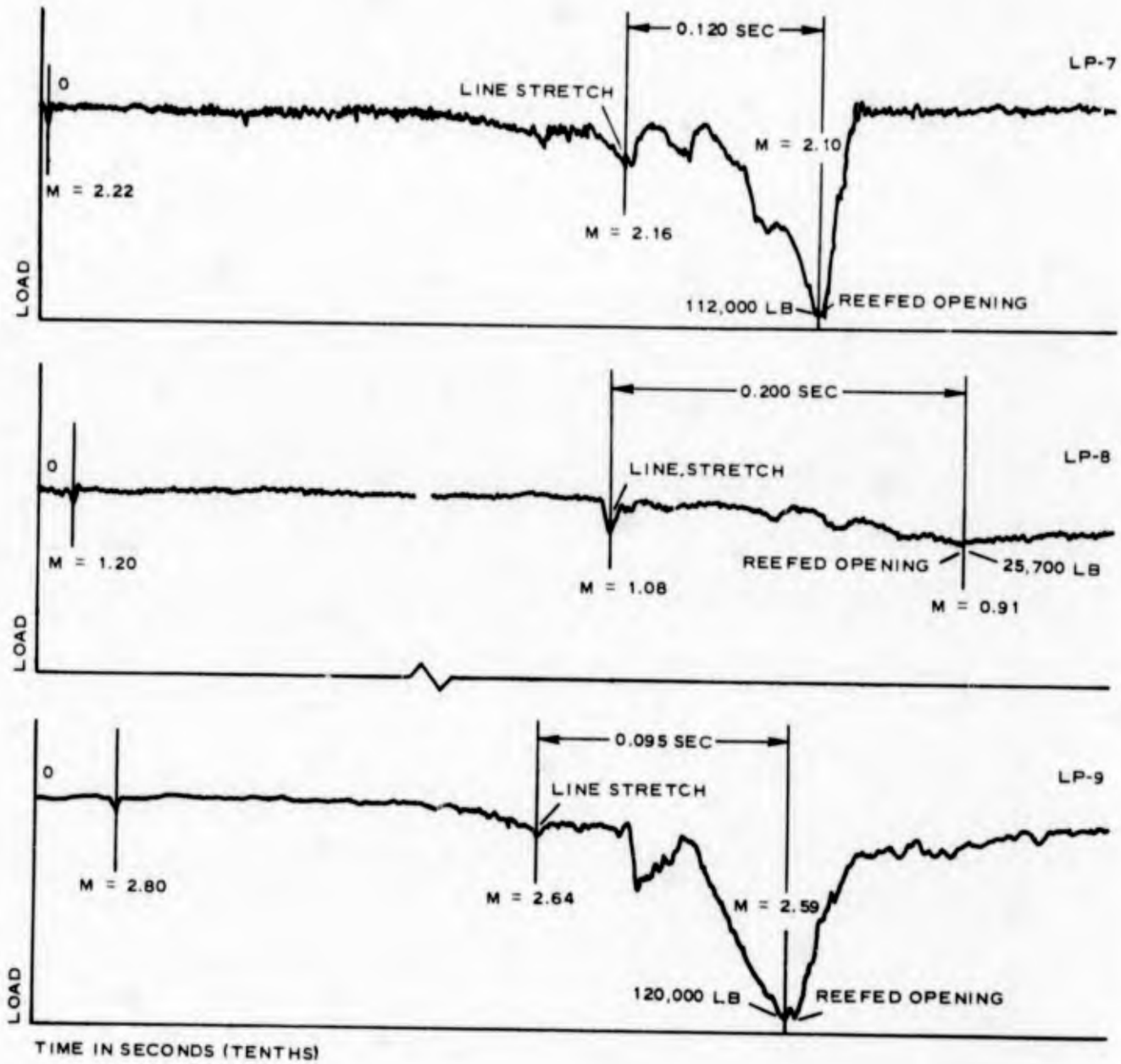


Figure 6 - Load Traces

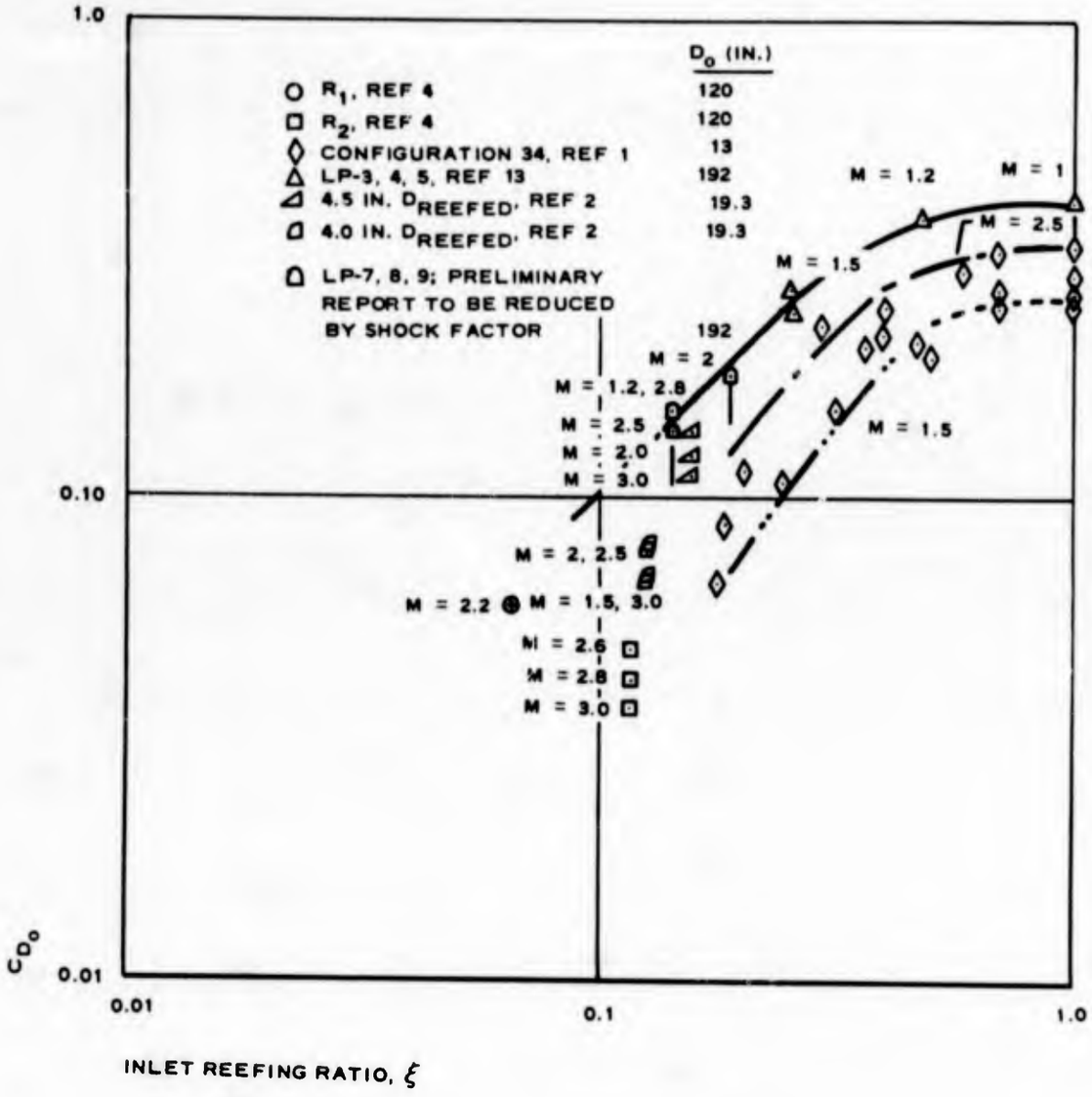


Figure 7 - Comparison of Small-Scale and Free-Flight Data

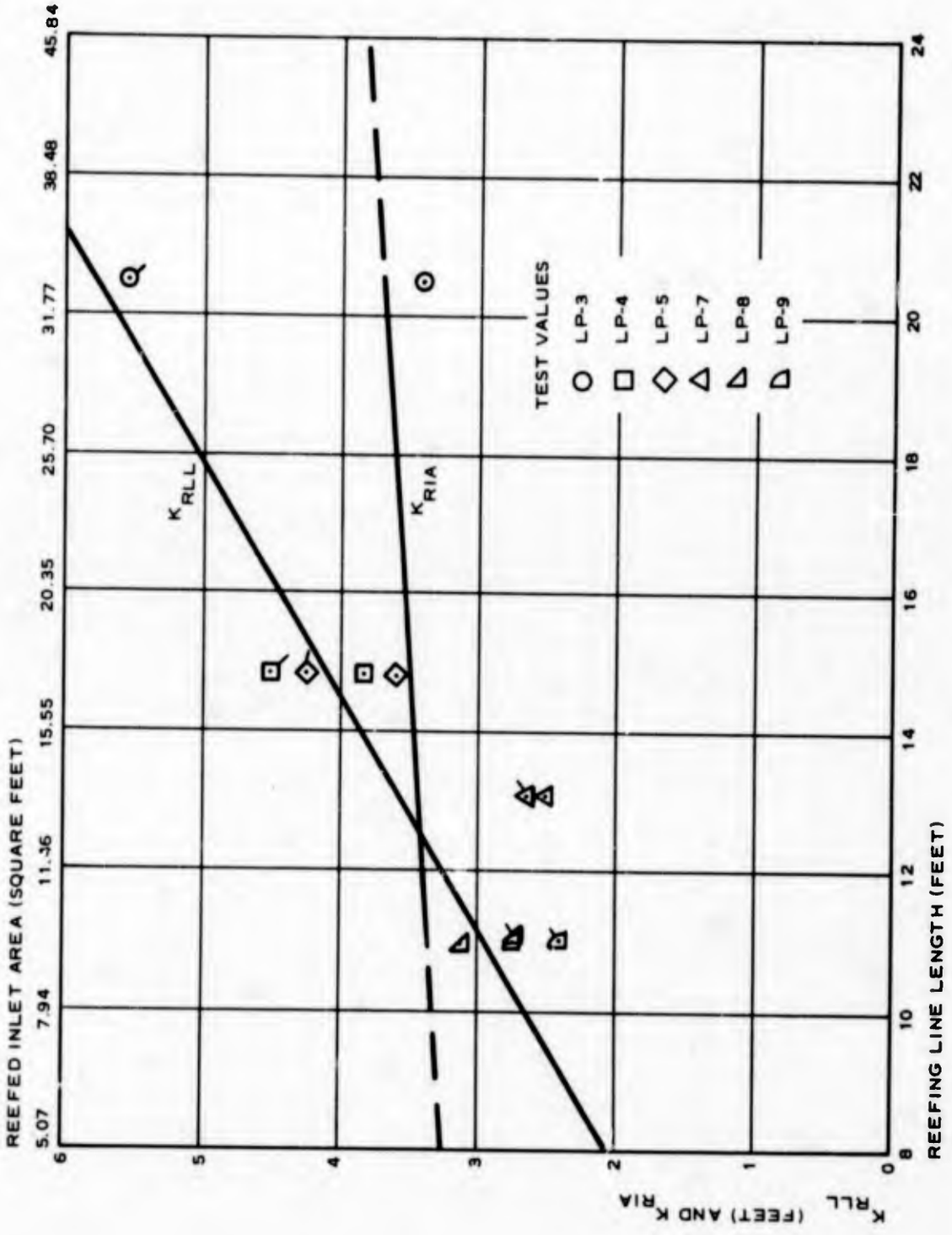


Figure 8 - K_{RLL} and K_{RIA} versus Reefing Line Length and Reefed Inlet Area

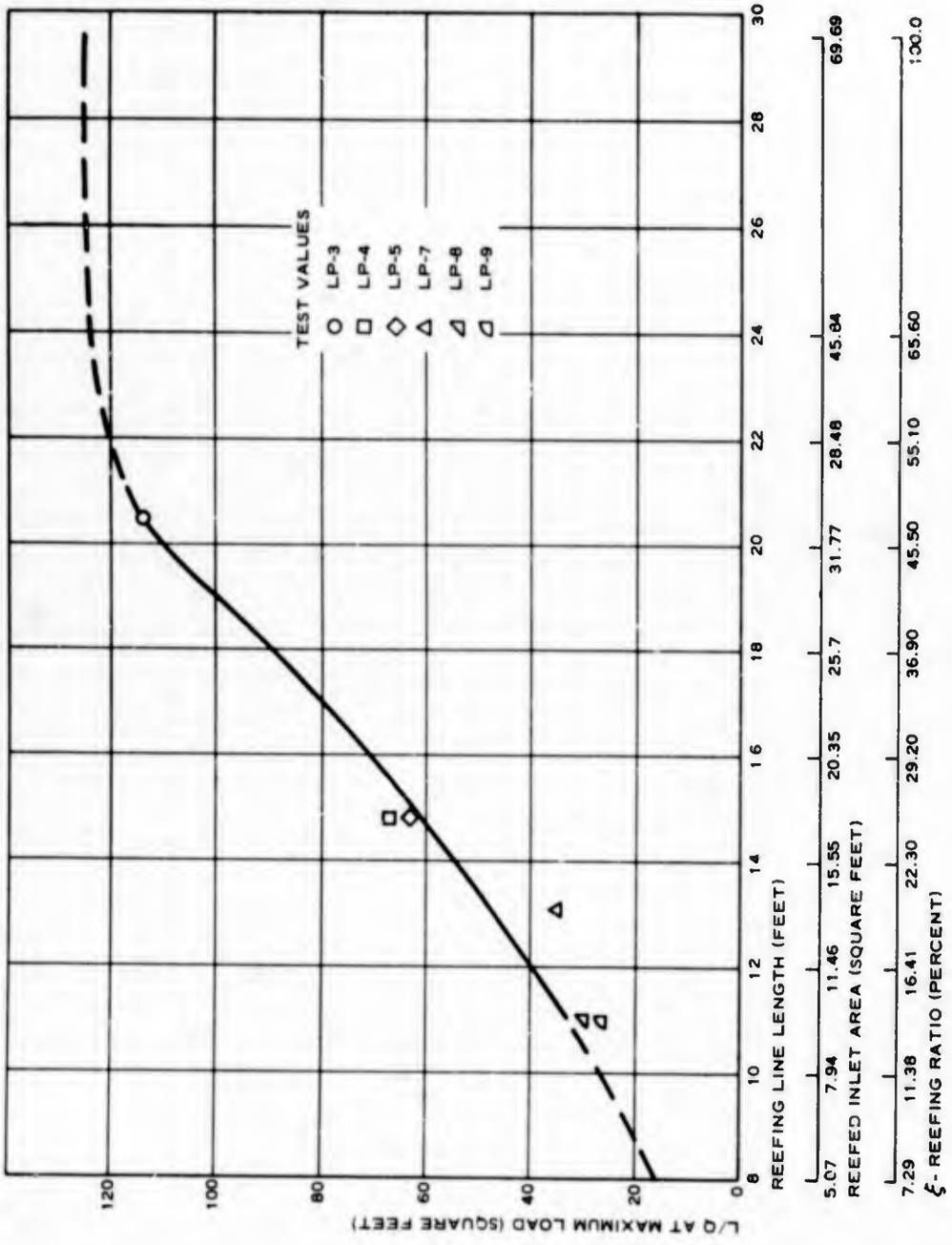


Figure 9 - L/q at Maximum Load versus Reefing Parameters Based on K_{RLL} and K_{RIA}

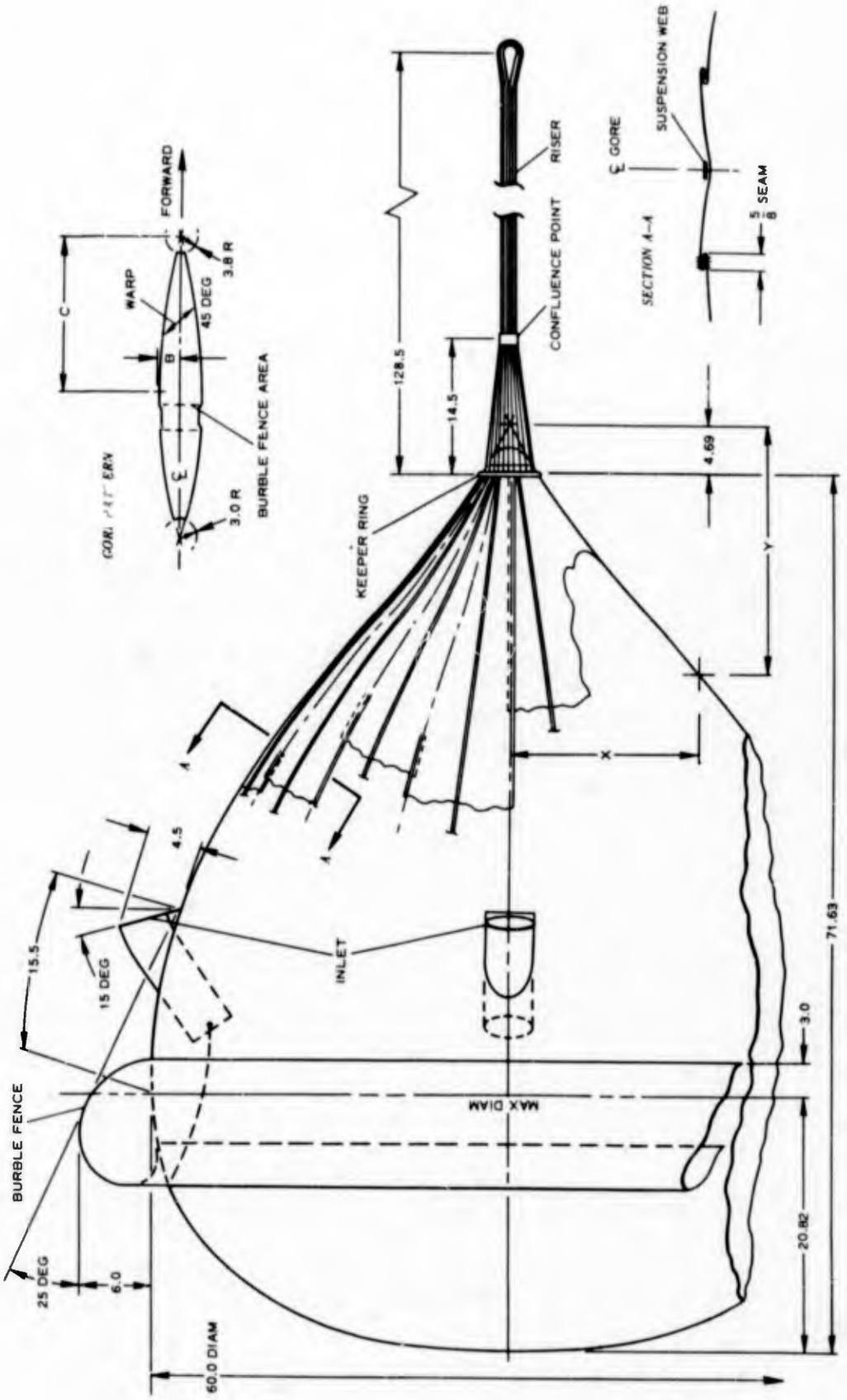


Figure 10 - Example of BALLUTE Configuration

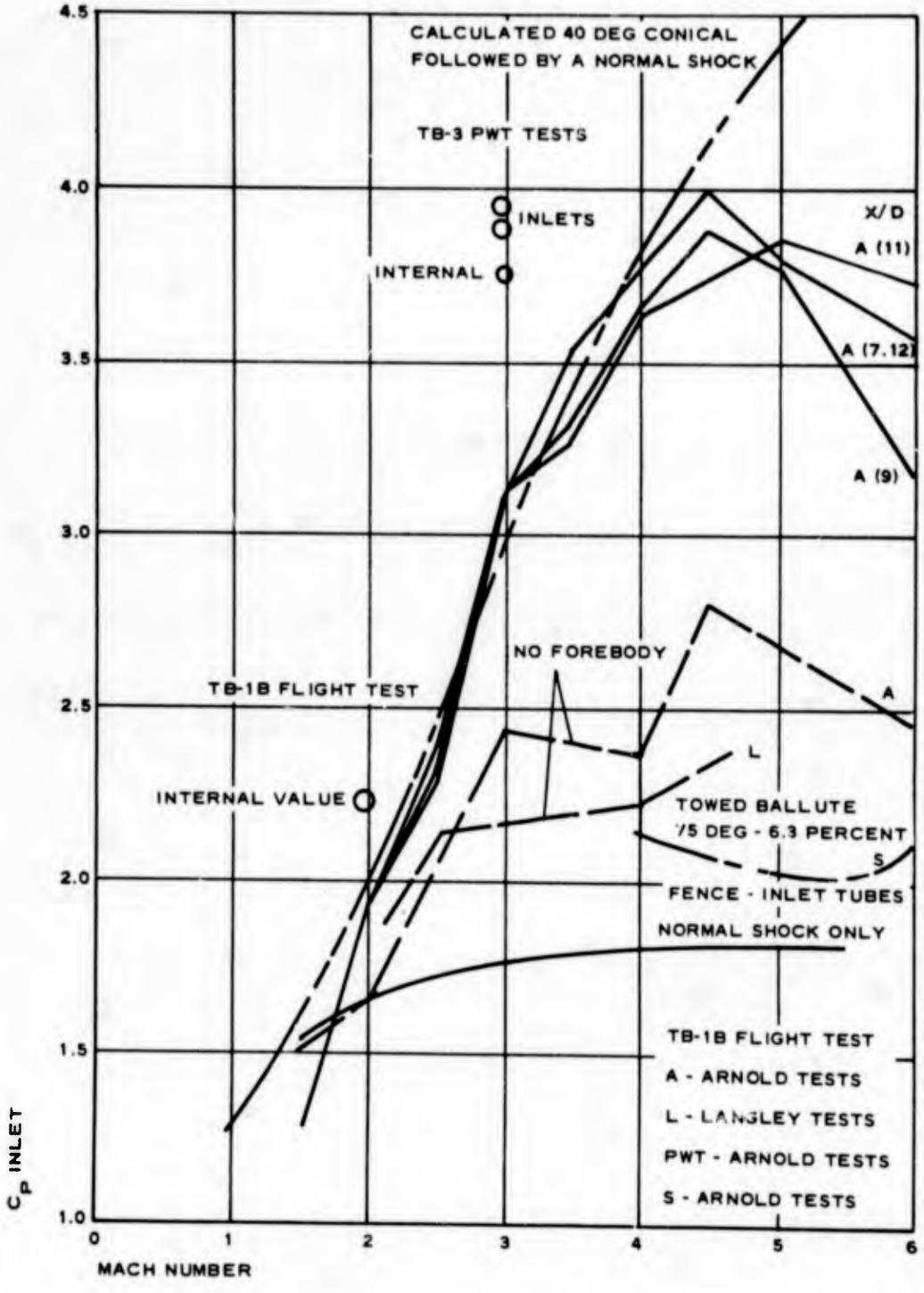


Figure 11 - Inlet and Internal Pressure Coefficient (C_p) Data from Tests of Small Rigid Wind-Tunnel Models and Full-Scale Items

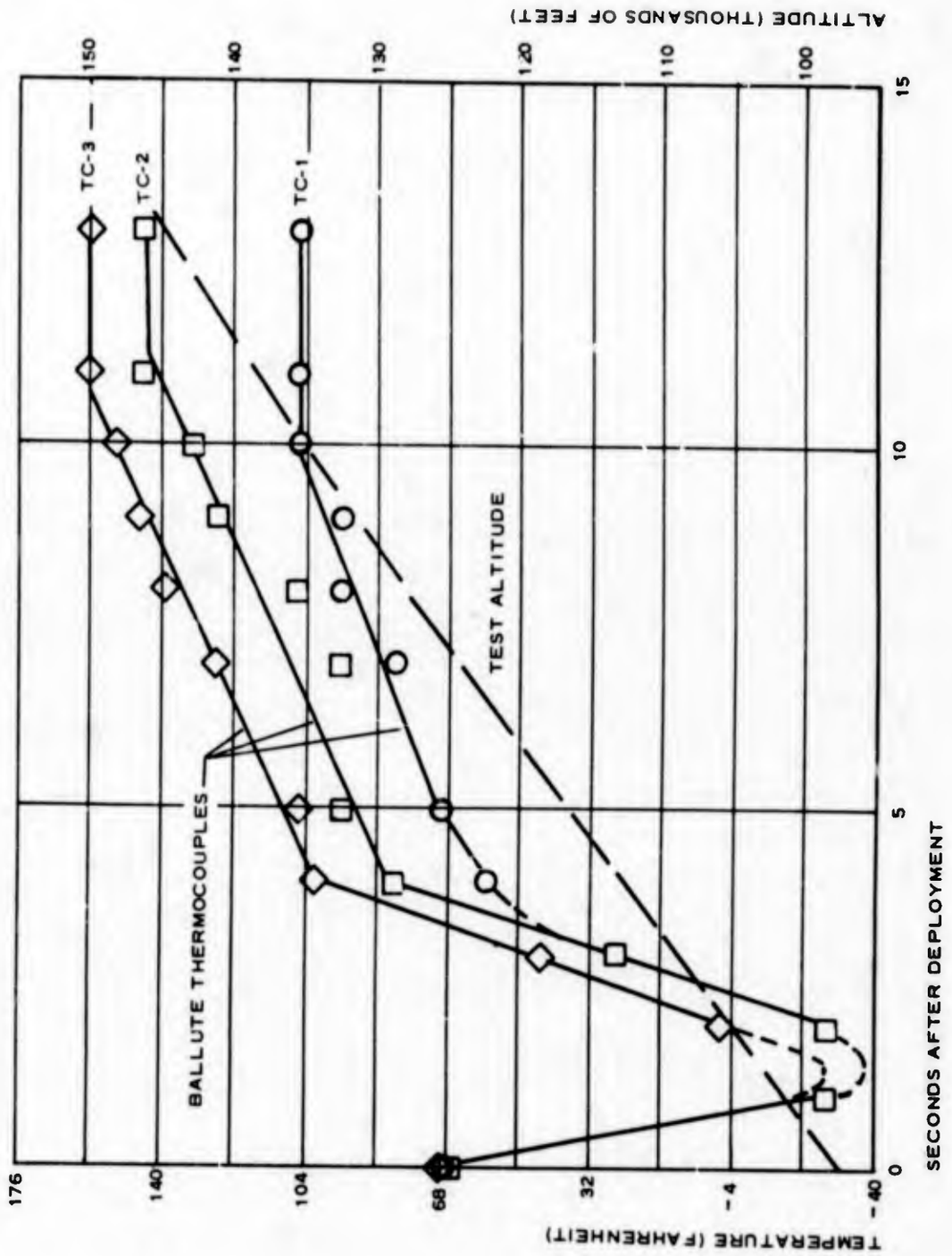


Figure 12 - Measured Thermocouple Values - Degrees Fahrenheit (TB-3 Flight) and Test Altitude versus Time

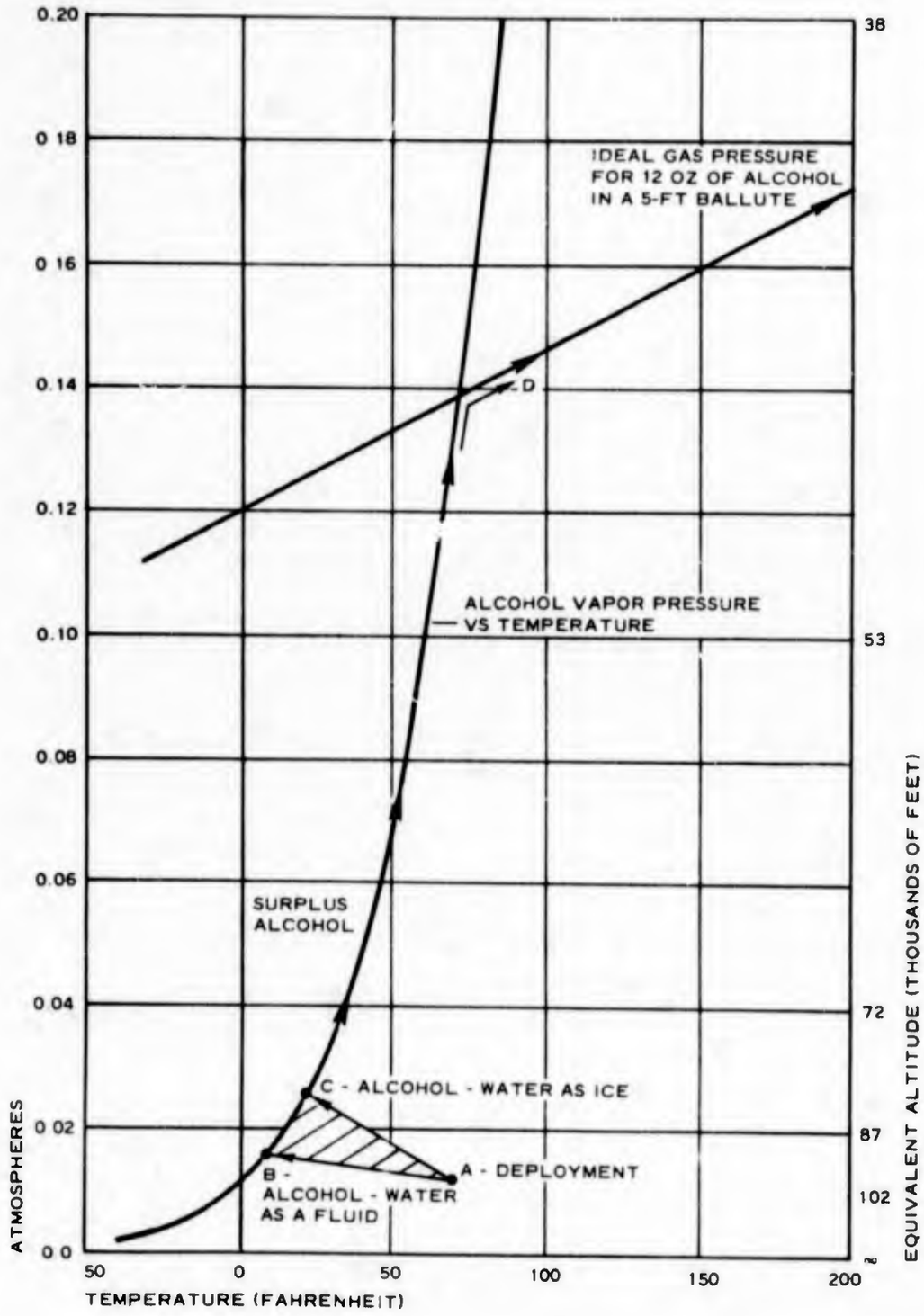


Figure 13 - Calculated BALLUTE Internal Pressures and Temperatures Based on TB-3 Flight Test Conditions

(U) BLAST WAVE EFFECTS ON THE PITCHING OF BLUNT CONES

by

Brian P. Quinn, 1st Lt, USAF

Hypersonic Research Laboratory
Aerospace Research Laboratories
Office of Aerospace Research
Wright-Patterson Air Force Base, Ohio

Lt Brian P. Quinn

BIOGRAPHY

1/Lt Brian P. Quinn

1/Lt Brian P. Quinn was assigned to the Aerospace Research Laboratories on 10 June 1964 as an Aerospace Research Engineer in the Hypersonic Research Laboratory. This was his first active duty assignment.

From September 1962 - June 1964 Lt Quinn held the position of "Chercheur" at the Laboratoire de M'ecanique des Fluides, Centre Orsay while a doctoral candidate at the University of Paris, France.

During the summer of 1962 Lt Quinn was a Research Engineer in the Research Department of Grumman Aircraft Engineering Corporation, Bethpage, Long Island, New York. His principal duties at that time involved force measurements on models of re-entry configurations in Grumman's shock tunnel.

From September 1960 - June 1962 Lt Quinn was a graduate student at the University of Notre Dame. In the capacity of assistant in the Aero Space Engineering Department he taught undergraduate aerodynamics initially and later researched a hydrodynamic problem under contract with the David Taylor Model Basin.

Prior to 1960 Lt Quinn's work was centered in academic pursuits with the exception of small summer jobs.

Lt Quinn was born in Flushing, New York, attended St. Andrew Auellino School and All Hallows Institute and earned a B.S. in engineering from the University of Notre Dame.

He and the former Mary Josephine Shipp of Tulsa, Oklahoma have two children, Brian II and Mary Jo.

BLANK PAGE

ABSTRACT

The flow around a blunted slender cone performing pitching oscillations is theoretically analyzed by means of Chernyi's blast-wave piston analogy. Stationary and transient shock shapes and pressure distributions are presented in power series form. Within the region of applicability, the stationary pressure agrees well with Chernyi's numerical integration. Normal force and pitching moment coefficient derivatives are presented and the results are plotted for a 10° cone. The theoretical static and dynamic stability derivatives agree satisfactorily with recent experiments in a Mach 14 wind tunnel. It is concluded that the experimentally observed reduction in static and dynamic stability due to increased nose bluntness may be attributed to blast wave effects.

INTRODUCTION

The delivery of a given payload to a prescribed location is the final mission objective of any missile system. The attainment of this objective requires that the system be capable of correcting any deviation from the planned trajectory during all phases of its flight. With particular reference to the vehicle's performance in an atmosphere, completion of the mission demands that the configuration be aerodynamically stable both statically and dynamically. Because this is a fundamental requirement which will persist as long as trajectories include atmospheric flight, research into the stability of re-entry vehicles continues to play a timely role in the Air Force's missile program.

Recognizing this need, a program of continuing research on the basic fluid mechanics of stability at hypersonic speeds was begun by the Hypersonic Research Laboratory at the Aerospace Research Laboratories several years ago when it became clear that dynamic stability was indeed a problem during re-entry. The study of reference (1) represents one of the major results of this effort. Walchner and Clay therein presented conclusive experimental evidence of the destabilization provoked by blunting the nose of circular cones, a configuration of obvious practical importance. Their semiempirical analysis was based on the method of tangent cones applied to Griffith and Lewis' [2] correlation of experimental pressure distributions along blunted cones at zero lift. The results of their analytical study agreed with their measured static pitching moment derivatives over the entire range of nose bluntnesses investigated. However, the apparent limitations of the tangent cone technique in such a complicated flow field were illustrated by the divergence of the trends of their analytical and experimental dynamic stability curves at rather large values of the nose bluntness ratio.

This paper is a theoretical study concerned with the influence of marked nose bluntness on the hypersonic stability derivatives of slender cones. As a complement to the work of reference (1), the explicit objective is to determine if blast-wave effects can explain the trend of experimental dynamic stability data.

CONTENTS

NOMENCLATURE

C_m = $\dfrac{\text{Pitching Moment}}{\frac{1}{2}\rho V^2 S \ell}$, pitching moment coefficient

C_N = $\dfrac{\text{Normal Force}}{\frac{1}{2}\rho V^2 S}$, normal force coefficient

$( )_\alpha$ = $\lim_{\alpha \to 0} \dfrac{\partial ( )}{\partial \alpha}$

$\left(\dfrac{q \ell}{V}\right) = 0$

$( )_q$ = $\lim_{\substack{\alpha = 0 \\ \partial \left(\frac{q\ell}{V}\right)}} \dfrac{\partial}{}$

$\left(\dfrac{q \ell}{V}\right) \to 0$

C_{D_n} = $\dfrac{\text{Nose Drag}}{\frac{1}{2}\rho V^2 \pi \ r_N^{\ 2}}$, nose drag coefficient

C_p = $\dfrac{P_{surface} - P_\infty}{\frac{1}{2}\rho V^2}$, pressure coefficient

C_p' = $\dfrac{P_{surface\ perturbation}}{\frac{1}{2}\rho V^2}$, perturbation pressure coefficient

E = energy released by explosion, related to nose drag of blunt cone

I = impulse imparted by explosion

K = $M_\infty \tan \Theta$, hypersonic similarity parameter

L = $\left(\dfrac{E}{\pi \rho U^2}\right)^{1/2}$, characteristic length

ℓ = body length

M_∞ = free stream Mach number

P = $\dfrac{C_p}{2\Theta^2}$, dimensionless pressure

NOMENCLATURE (CONT'D)

$p = \frac{C_p'}{2\Theta^2}$, dimensionless perturbation pressure

$R(\xi)$ = dimensionless shock shapes or shock radius

$R_p(\xi)$ = dimensionless piston radius

$r(\xi)$ = dimensionless shock perturbation

$r_p(\xi)$ = dimensionless piston perturbation

r_N = nose radius of body

r_B = base radius of body

$S = \pi r_B^2$

$T = \frac{L}{U}$, characteristic time

t = time

$U = V \tan \Theta$

V = free stream speed

x = distance from nose in flow direction

x_{cg} = distance of pitch axis from nose

y = body fixed ordinate

Greek symbols

$\alpha = |\alpha| \sin \omega t$, pitch angle

$|\alpha|$ = amplitude of α

γ = ratio of specific heats

$\epsilon = -\frac{|\alpha|}{\Theta} \cos \phi$

$\eta = \frac{r_N}{r_B}$, nose bluntness ratio

Θ = cone half angle

$\lambda = \left(\frac{\Theta(1-\eta)}{2\eta}\right)^{1/2}$, a correlation parameter

NOMENCLATURE (CONT'D)

ξ = $\frac{t}{T} = \sqrt{\frac{2}{C_D}}\ \Theta^2\ \frac{x}{r_N}$, dimensionless time

ρ = free stream density

ϕ = peripheral angle (see fig. 3)

ω = circular pitching frequency

subscript $(\ )_0$ = { stationary shock shape and pressure or pitch axis through nose

$(\dot{})$ = $\frac{d(\ )}{d\xi}$

ANALYSIS

1. Equations of Motion

Blunted leading edges on hypersonic vehicles are responsible for a host of fluid dynamic complications including detached and curved shock waves, rotational flow, chemical effects, existence of subsonic and supersonic speed regimes, and many others. Although it is sometimes necessary to account for all or most of these complexities in a theoretical study, expediency and efficiency frequently suggest that a simplified model, retaining only the global aspects of the flow, serve as the foundation of the analysis. Chernyi[3] has proposed such a model for the hypersonic flow over blunted cones. By virtue of the equivalence principle, he treats the blunted cone problem indirectly by considering an analogous problem in piston theory. The analogy is represented schematically in figure 1. In the piston plane, a cylindrical charge of unit depth explodes at time t = 0 and releases a quantity of energy into fluid which is initially at rest. At the same instant, a cylindrical piston is formed at the point of the explosion and begins to move outwards toward the shock wave produced by the explosion. The analogy is related to the hypersonic flow over a blunted cone by equating the energy released by the explosion with the work done by the nose drag in moving a unit distance.

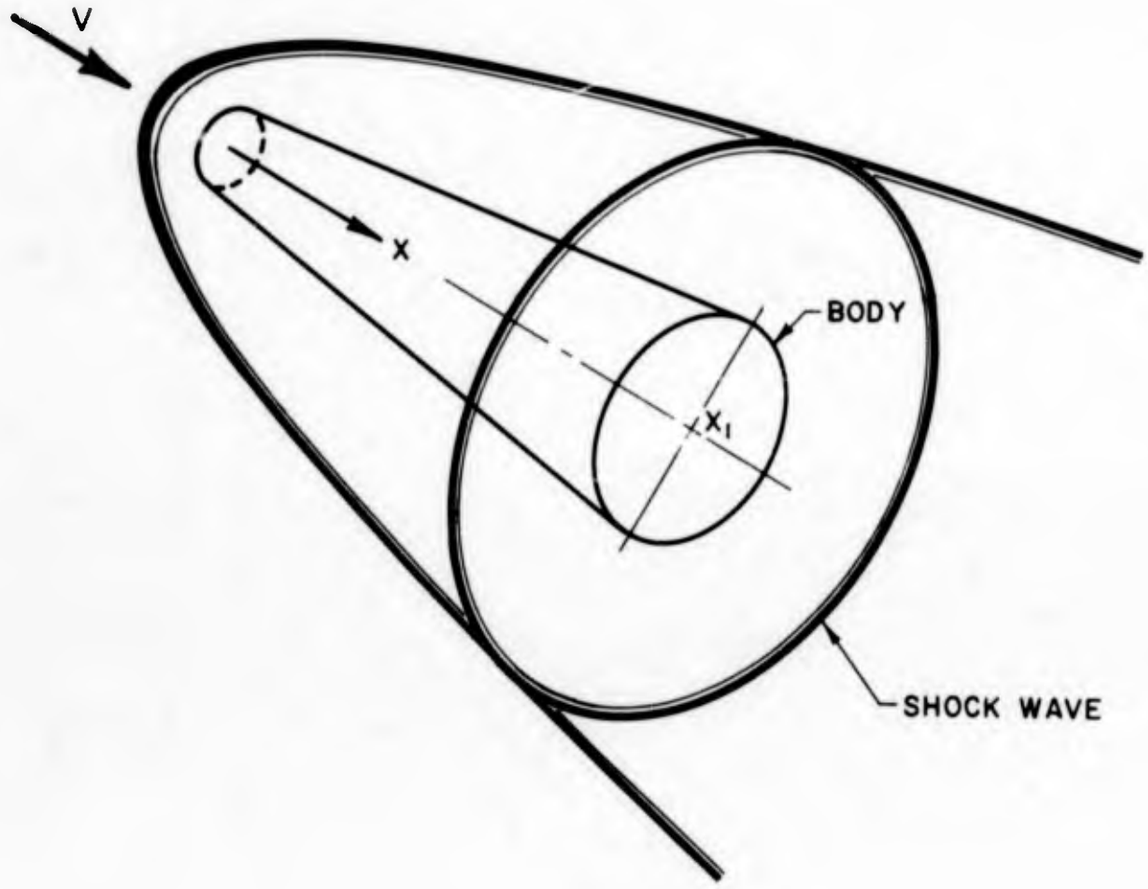
Conservation of energy requires that the fluid's energy at any instant be equal to the sum of the energy released by the explosion, the initial energy of the gas and the work done by the piston. In like manner, the momentum of the fluid at any instant is equal to the sum of the impulses imparted by the explosion and by the piston. In view of these considerations, Chernyi has deduced that the equations expressing conservation of energy and momentum may be approximated by the following dimensionless equations:

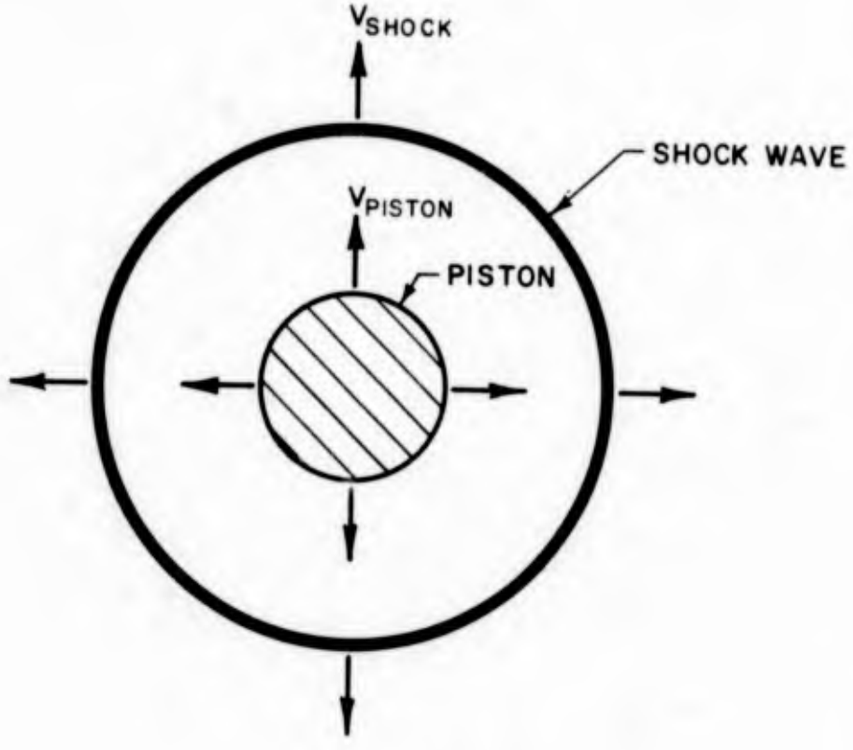
$$\frac{1}{2}\left(\frac{2}{\gamma+1}\right)^2\left(\dot{R}R\right)^2 + \frac{1}{\gamma-1}\left(R^2 - R_p^2\right)P = 1 + 2\int_0^{\xi} P R_p \, dR_p(\xi) \qquad (1)$$

and

$$\left(\frac{2}{\gamma+1}\right)\dot{R}R^2 = \frac{IU}{E} + 2\int_0^{\xi} P R \, d\xi \qquad (2)$$

where characteristic length and time have been chosen as $L = (E/\pi\rho_\infty U^2)^{1/2}$ and $T = L/U$, respectively. The term IU/E in equation (2) is generally considered to be negligibly small for slender cones.





$$0 < t_1 = \frac{X_1}{V}$$

FIGURE 1

The above equations differ somewhat from those set down by Chernyi in that the initial pressure of the fluid has been neglected. This implies an infinite value for the hypersonic similarity parameter, $K = M_\infty \tan \Theta$. The deletion of the initial pressure brings to question the applicability of equations (1) and (2) to physical problems wherein K is finite. This simplification is, however, consistent with the fact that equations (1) and (2) were not derived from consideration of the physical flow about a blunted cone but from a very simplified model. More than qualitative information should not be expected from their solution. Moreover, we note that Chernyi's pressure expression does not approach true cone surface pressures for large ξ, but it does tend toward the Newtonian value $P = 1$ for all K. This is a consequence of certain approximations used in the formulation of the model, such as the constancy of the speed of the gas in the region bounded by the piston and the shock wave, and the concentration of mass near the shock wave. These approximations become realistic in the double limit $K \to \infty$, $\gamma \to 1$, and for this reason Chernyi's equations may be considered as limiting equations for large K. In this sense, deletion of the initial pressure from the energy and momentum equations, as has been done in equations (1) and (2), should not influence their qualitative validity.

2. Solution of the Steady Problem

The approximate equations (1) and (2) serve as the point of departure for this study. Before proceeding to the case of a pitching cone, it is first necessary to solve the governing equations for the case of hypersonic flow over a stationary blunt cone.

Eliminating the pressure terms from equations (1) and (2), and noting that $R_{p_0} = \Theta x / L = \xi$, results in the shock shape being governed by the following equation:

$$\frac{1}{2}\left(\frac{2}{\gamma+1}\right)^2 \left(\dot{R}_0 R_0\right)^2 + \frac{R_0^2 - \xi^2}{\gamma^2 - 1} \frac{1}{R_0} \frac{d}{d\xi}\left(R_0^2 \dot{R}_0\right)$$

$$= 1 + \frac{2}{\gamma+1} \int_0^\xi \frac{\xi}{R_0} \frac{d}{d\xi}\left(R_0^2 \dot{R}_0\right) d\xi \qquad (3)$$

where subscript zero indicates a nonpitching condition.

For the bluntness ratios of interest, that is for $\eta = r_N / r_B$ more than around 10 per cent and for small cone angles Θ, the entire body will generally be contained in the interval

$$0 \leqq \xi \leqq 1 \qquad (4)$$

For such small values of ξ and in view of the asymptotic form of R_0 established by Chernyi, the solution of equation (3) has been approximated by a power series of the form

$$R_0 = A\xi^{1/2} + B\xi^{3/2} + C\xi^{5/2} + \ldots \tag{5}$$

Substitution into equation (3) results in the following expressions for the coefficients:

$$A = \left[\frac{4(\gamma+1)^2(\gamma-1)}{3\gamma-1}\right]^{1/4}$$

$$B = \frac{(2\gamma-1)(\gamma+1)}{32\gamma A} \tag{6}$$

$$C = \frac{\gamma+1}{A^2(15\gamma+3)}\left[\frac{7}{2}\gamma B - \frac{AB^2(47\gamma+3)}{2(\gamma+1)}\right]$$

Representative numerical values are given in Table I.

	$\gamma = 1.4$	$\gamma = 1.6667$
A	1.3027	1.4756
B	.0741	.0791
C	.0153	.0140

TABLE I

The pressure distribution over slender blunted cones whose shock shape is given by equation (5) is found from equation (2) to be

$$P_0 = \frac{1}{4(\gamma+1)}\left[A^2\xi^{-1} + 14AB + (21B^2 + 34AC)\xi + \ldots\right] \tag{7}$$

Figure (2) compares Chernyi's numerical solution of the system of equations (1) and (2) with the above result.

3. The Pitching Problem

The approach used herein to formulate the problem of hypersonic flow around a blunted cone performing small pitching oscillations about

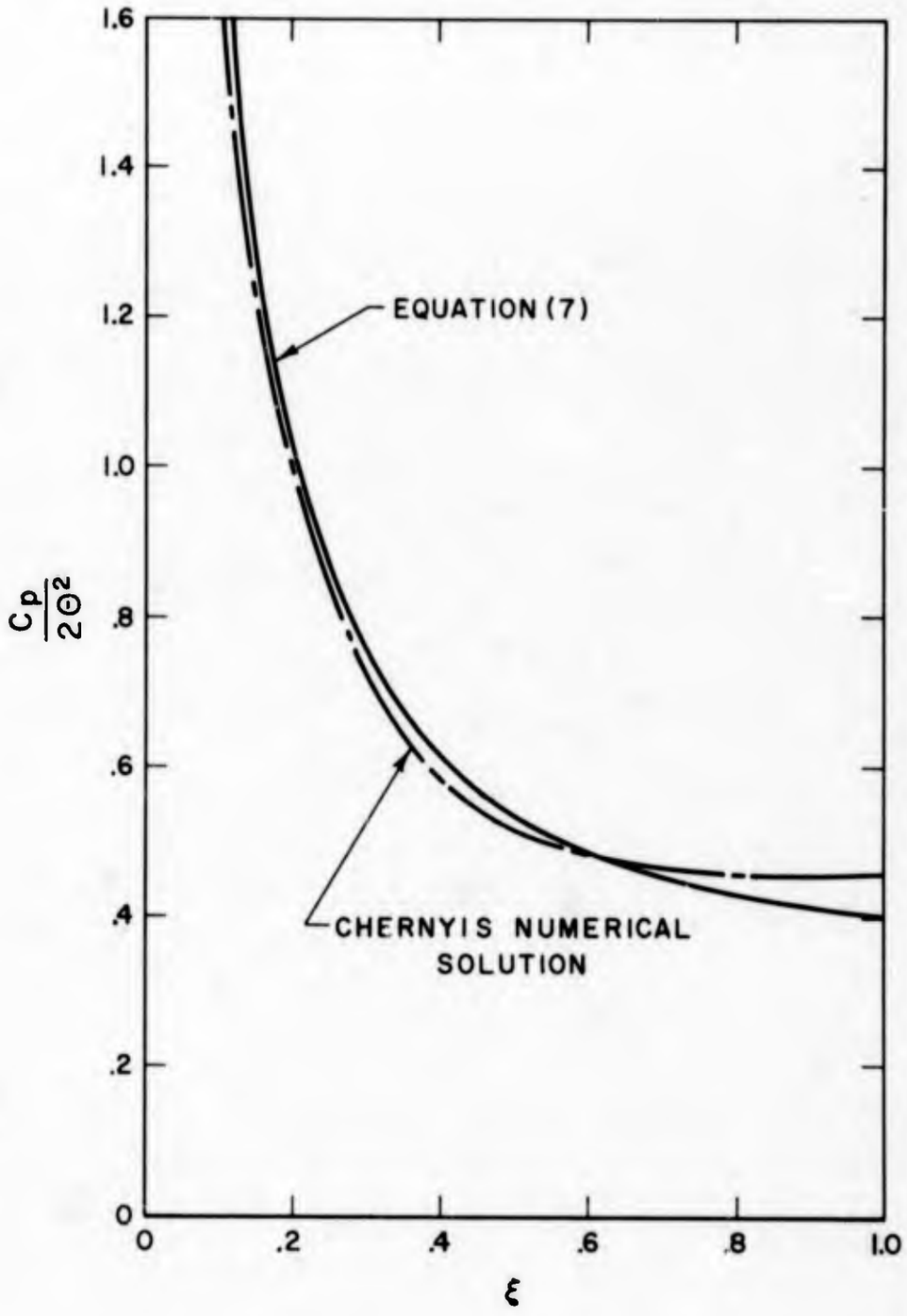

1.6
1.4
1.2
1.0
.8
.6
.4
.2
0
0
.2
.4
.6
.8
1.0
$\frac{C_p}{2\Theta^2}$
ξ
EQUATION (7)
CHERNYIS NUMERICAL SOLUTION

FIGURE 2

zero angle of attack consists of a perturbation technique. Thus, corresponding to a piston whose radius is perturbed an amount $r_p(\xi)$ such that the piston radius may be expressed as

$$R_p(\xi) = R_{p_0}(\xi) + r_p(\xi) \tag{8}$$

the shock radius and pressure may be represented by

$$R = R_0(\xi) + r(\xi) \tag{9}$$

$$P = P_0(\xi) + p(\xi) \tag{10}$$

Assuming that R_{p_0}, R_0 and P_0 are themselves solutions to the equation of motion, and assuming that products of the perturbations and their derivatives are negligibly small, the linearized equations governing the perturbations are found from substitution into equations (1) and (2).

Thus,

$$p = \frac{1}{\gamma + 1}\left[R_0 \ddot{r} + 4\dot{R}_0 \dot{r} + \ddot{R}_0 r \right] \tag{11}$$

and

$$\left(\frac{2}{\gamma+1}\right)^2 R_0 \dot{R}_0 \frac{d}{d\xi}\left(R_0 r\right) + \frac{1}{\gamma - 1}\left[2P_0\left(R_0 r - r_p\right) + p\left(R_0^2 - \xi^2\right) \right]$$

$$= 2\int_0^{\xi} \left[P_0\left(\xi \dot{r}_p + r_p\right) + \xi p \right] d\xi \tag{12}$$

With reference to figure (3),

$$R_p = \frac{\Delta y}{L} \tag{13}$$

so that at zero incidence $\Delta y = \Theta x$ and $R_{p_0} = \xi$. At incidence, however, the cosine law indicates that

$$(\Delta y + r_N)^2 = (r_N + \Theta x)^2 + (\alpha x)^2 - 2(r_N + \Theta x)\alpha x \cos\phi \tag{14}$$

For small incidences we may assume that Δy can be expressed by an expansion of the form

$$\Delta y = \Delta y_{\alpha=0} + x\alpha f_1(\alpha, \phi) + x\alpha^2 f_2(\alpha, \phi) + \ldots \tag{15}$$

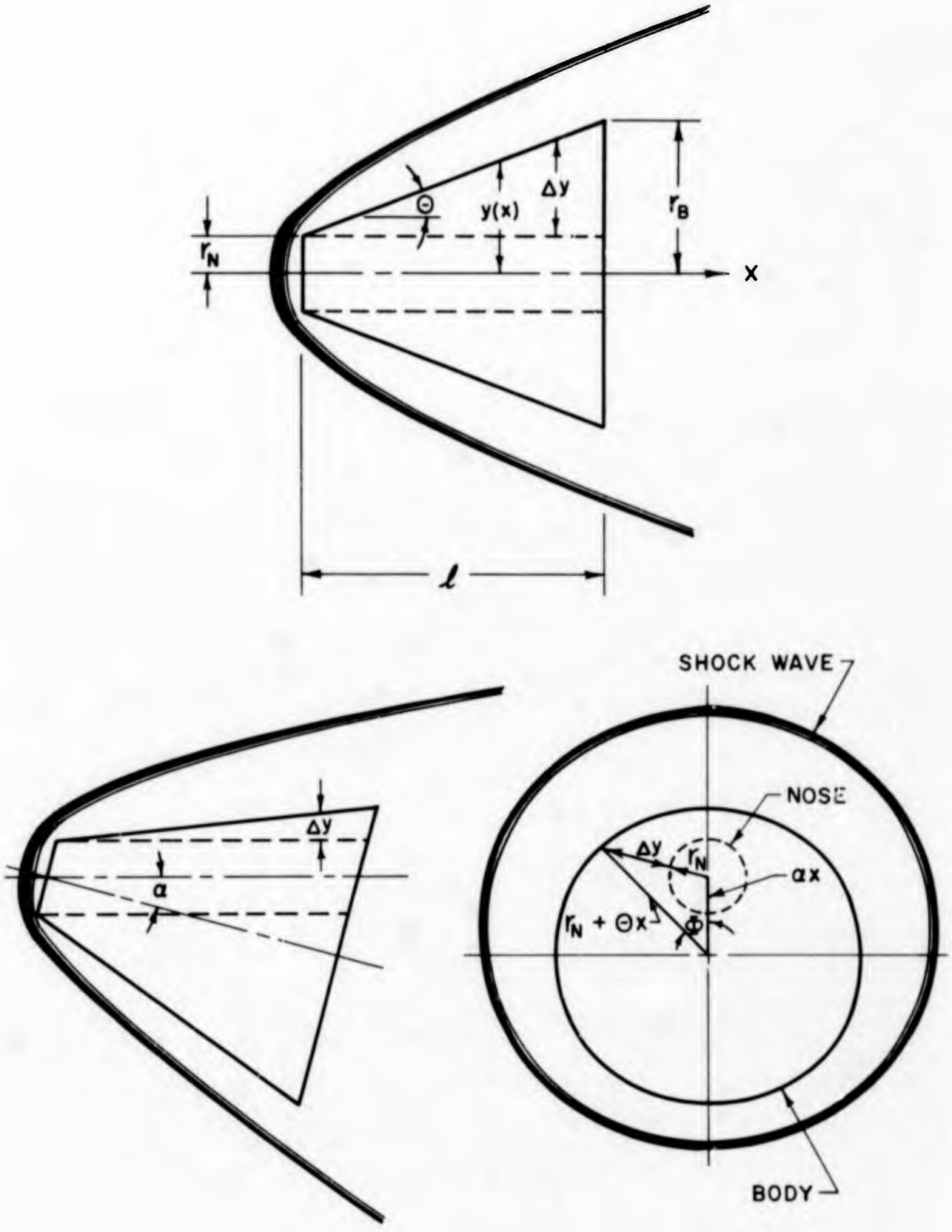


FIGURE 3

and by comparing like terms of equations (14) and (15) Δy may be written

$$\Delta y = \Theta x \left(1 - \frac{a}{\Theta} \cos \phi + \mathcal{O}(a^2) + \ldots\right) \tag{16}$$

whereupon

$$R_p = \xi \left(1 - \frac{a}{\Theta} \cos \phi\right) \tag{17}$$

if higher order terms are neglected so that comparison with equation (8) yields

$$r_p(\xi) = - \frac{a}{\Theta} \cos \phi \, \xi \tag{18}$$

Strictly speaking, equations (11) and (12) are not compatible with the piston perturbation described above because of its dependence on the peripheral angle ϕ. The $\cos \phi$ term will introduce into the model crossflows for which no account has been taken. However, since the stability derivatives are to be evaluated under the limiting conditions of vanishing incidence and vanishing pitch rate, it is assumed that cross-flow components may be neglected.

According to equation (18), the piston perturbation which is analogous to the disturbances set up by a blunted cone performing pitching oscillations about an axis through its nose may be written as

$$r_p(\xi) = - \frac{\xi \cos \phi}{\Theta} \, |a| \, \sin \omega t \tag{19}$$

where $|a|$ and ω are the amplitude and circular frequency of the oscillation, respectively. The argument of the sine function may be rewritten as follows:

$$\omega t = \omega T \, \frac{t}{T} = \frac{\omega L}{U} \, \xi \tag{20}$$

so that for small arguments the piston perturbation becomes

$$r_p(\xi) = - \frac{|a| \cos \phi}{\Theta} \, \frac{\omega L}{U} \, \xi^2$$

$$= \epsilon \, \frac{\omega L}{U} \, \xi^2 \tag{21}$$

thereby defining ϵ . When this piston perturbation is introduced into equations (11) and (12) together with the appropriate boundary conditions

$$r = \dot{r} = 0 \qquad \text{for} \quad \xi = 0 \tag{22}$$

there results a well-defined system of equations from which the shock and pressure perturbations may be determined. If the shock perturbation is expanded as

$$r(\xi) = (a\xi^{3/2} + b\xi^{5/2} + c\xi^{7/2} + \ldots) \sin \omega t \tag{23}$$

substitution into the equations (11) and (12) results in the coefficients being given by

$$a = \frac{\epsilon}{4A} \frac{(\gamma + 1)^2}{(15\gamma + 3)}$$

$$b = \frac{17}{48} \frac{a}{A^2} (2\gamma + 1)(\gamma + 1) + \frac{7}{6} \epsilon \frac{B}{A^2} \gamma(\gamma + 1) - \frac{Ba}{A} \left(\frac{97\gamma + 33}{8} \right) \tag{24}$$

$$c = \frac{1}{F_1 A^3} \left[F_2 - F_3 + F_4 + F_5 \right]$$

where A, B and C are defined by equations (6) and

$$F_1 = \frac{5}{\gamma + 1} \left(\frac{3\gamma - 1}{\gamma^2 - 1} \right)$$

$$F_2 = \frac{1}{\gamma + 1} \left[\frac{31}{4} Ab + \frac{39}{4} Ba + \frac{\epsilon}{8} \left(21B^2 + 34AC \right) \right]$$

$$F_3 = \left(\frac{2}{\gamma + 1} \right)^2 \left[\frac{21}{2} A^2 Bb + \frac{23}{2} A^2 Ca + \frac{27}{2} AB^2 a + 8AB^2 a \right]$$

$$F_4 = \frac{\epsilon}{\gamma^2 - 1} \left[\frac{21}{2} B^2 + 17AC \right]$$

$$F_5 = \frac{1}{\gamma^2 - 1} \left[\frac{31}{2} Ab + \frac{39}{2} Ba - 70A^2 Bb - 26A^2 Ca - 86AB^2 a \right]$$

In table II values of a, b and c are given for representative values of γ .

	$\gamma = 1.4$	$\gamma = 1.6667$
$\bar{a}$	.04607	.04304
$\bar{b}$	.20355	.21320
$\bar{c}$	.22705	.16355

$$a = \frac{\bar{a}}{\epsilon} \qquad b = \frac{\bar{b}}{\epsilon} \qquad c = \frac{\bar{c}}{\epsilon}$$

TABLE II

Direct substitution of the shock perturbation into equation (11) results in the perturbation pressure being given by

$$p = \frac{1}{\gamma+1}\left\{\left[\frac{7}{2}Aa + \left(\frac{17}{2}Ab + \frac{21}{2}Ba\right)\xi + \left(\frac{31}{2}Ac + \frac{39}{2}Bb + \frac{39}{2}Ca\right)\xi^2\right]\sin\omega t + \frac{\omega L}{U}\left[5Aa\,\xi + \left(7Ab + 9Ba\right)\xi^2 + \left(9Ac + 11Bb + 13Ca\right)\xi^3\right]\cos\omega t\right\} \quad (25)$$

Notice that the perturbation pressure is the sum of two components: the first, which is in phase with the piston displacement, is associated with the oscillation frequency, while the second, which is ninety degrees out of phase, is associated with the damping of the oscillation. Each component is plotted in figure (4).

4. Forces and Moments

Using the notation of figure (3), it has been shown[4] that the normal force and pitching moment coefficients can be determined from the following relations:

$$C_N = -\frac{2\ell^2}{S}\int_0^1 \frac{y(x)}{\ell}\, d\left(\frac{x}{\ell}\right)\int_\pi^0 C_p \cos\phi\, d\phi \quad (26)$$

$$C_m = \frac{2\ell^2}{S}\int_0^1 \left(\frac{xy(x)}{\ell^2} + \frac{\Theta y^2(x)}{\ell^2}\right) d\left(\frac{x}{\ell}\right)\int_\pi^0 C_p \cos\phi\, d\phi \quad (27)$$

where the pitching axis has been taken through the nose of the body and the small angle approximations have been assumed.

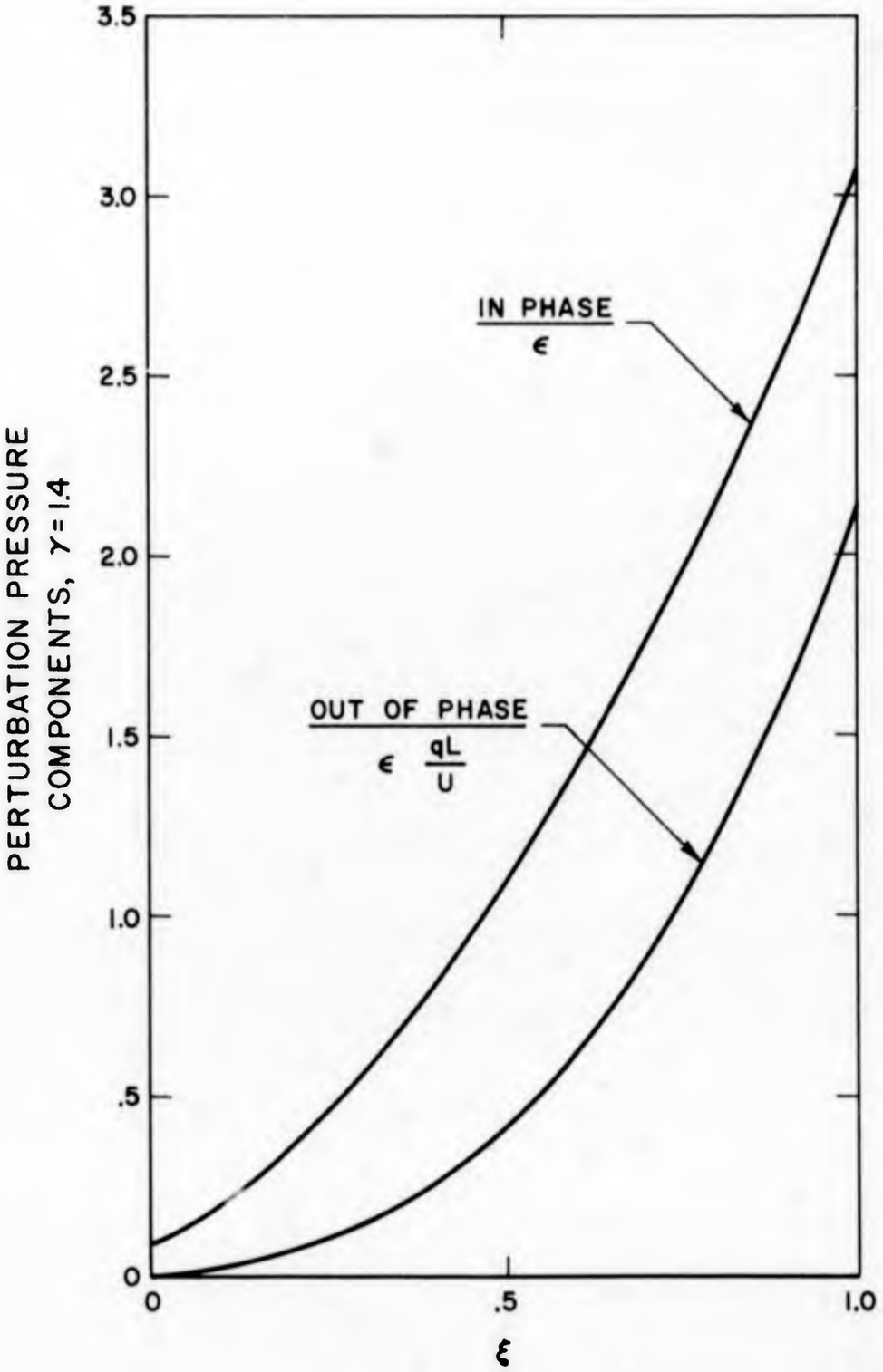

PERTURBATION PRESSURE COMPONENTS, γ=1.4
3.5
3.0
2.5
2.0
1.5
1.0
.5
0
0
.5
1.0
ξ
IN PHASE
ε
OUT OF PHASE
ε qL/U

FIGURE 4

The pressure integrand which is common to both equations (26) and (27) is, of course, the sum of two terms; the stationary pressure at zero angle of attack and the perturbation pressure. Due to symmetry, the integrated contribution of the former is null and the pressure integral becomes simply

$$\int_{\pi}^{0} C_p' \cos\phi \, d\phi \tag{28}$$

Recalling that

$$p = \frac{C_p'}{2\Theta^2}, \quad \epsilon = \frac{-|a| \cos\phi}{\Theta}, \quad a = \epsilon \bar{a}, \quad b = \epsilon \bar{b}, \quad c = \epsilon \bar{c}$$

and using equation (25), the pressure integral becomes

$$\int_{\pi}^{0} C_p \cos\phi \, d\phi = -\frac{\pi \Theta^2 a_0}{\gamma+1} \left\{ \left[\frac{7}{2} A\bar{a} + \left(\frac{17}{2} A\bar{b} + \frac{21}{2} B\bar{a} \right) \xi + \left(\frac{31}{2} A\bar{c} + \frac{39}{2} B\bar{b} + \frac{39}{2} C\bar{a} \right) \xi^2 \right] \sin \omega t + \frac{\omega L}{U} \left[5\bar{a}\xi + \left(7A\bar{b} + 9B\bar{a} \right) \xi^2 + \left(9A\bar{d} + 11B\bar{b} + 13C\bar{a} \right) \xi^3 \right] \cos \omega t \right\} \tag{29}$$

From the geometry of figure (3) we can deduce

$$\frac{x}{r_N} = \frac{x}{\ell} \frac{r_B}{r_N} \frac{\ell}{r_B} = \frac{x}{\ell} \frac{1-\eta}{\Theta\eta}$$

so that

$$\xi = \sqrt{\frac{2}{C_D}} \Theta^2 \frac{x}{r_N} = \sqrt{\frac{2}{C_D}} \Theta \frac{1-\eta}{\eta} \frac{x}{\ell} .$$

Moreover, we note that

$$\frac{\omega L}{U} = \frac{\omega \ell}{V} \frac{x}{\ell} \frac{1}{\xi}$$

so that the integral (29) can be expressed in terms of the body abscissa x, the angle of attack $\alpha = |\alpha| \sin \omega t$, and the pitching speed $q = d\alpha/dt = |\alpha| \omega \cos \omega t$ as follows:

$$\int_{\pi}^{0} C_p \cos \phi \, d\phi = \frac{-\pi \Theta^2}{\gamma + 1} \left\{ \left[D_1 + D_2 \frac{x}{\ell} + D_3 \left(\frac{x}{\ell} \right)^2 \right] \alpha + \left[D_4 \left(\frac{x}{\ell} \right) + D_5 \left(\frac{x}{\ell} \right)^2 + D_6 \left(\frac{x}{\ell} \right)^3 \right] \frac{q\ell}{V} \right\} \tag{30}$$

where $D_0 = \sqrt{\frac{2}{C_D}} \; \Theta \; \frac{1-\eta}{\eta}$

$$D_1 = \frac{7}{2} A \bar{a}$$

$$D_2 = \left(\frac{17}{2} A \bar{b} + \frac{21}{2} B \bar{a} \right) D_0$$

$$D_3 = \left(\frac{31}{2} A \bar{c} + \frac{39}{2} B \bar{b} + \frac{39}{2} C \bar{a} \right) D_0^2$$

$$D_4 = 5 A \bar{a}$$

$$D_5 = \left(7 A \bar{b} + 9 B \bar{a} \right) D_0$$

$$D_6 = \left(9 A \bar{c} + 11 B \bar{b} + 13 C \bar{a} \right) D_0^2$$

After substitution of the pressure integral (30), equations (26) and (27) may be integrated along the length of the body to give the normal force and pitching moment coefficients. Differentiating will then result in the required derivatives:

$$C_{N_\alpha} = \frac{2\Theta}{\gamma + 1} \left\{ \eta \left(1-\eta \right) \left[D_1 + \frac{1}{2} D_2 + \frac{1}{3} D_3 \right] + \left(1-\eta \right)^2 \left[\frac{1}{2} D_1 + \frac{1}{3} D_2 + \frac{1}{4} D_3 \right] \right\} \tag{32}$$

$$C_{N_{q_0}} = \frac{2\Theta}{\gamma + 1} \left\{ \eta \left(1-\eta \right) \left[\frac{1}{2} D_4 + \frac{1}{3} D_5 + \frac{1}{4} D_6 \right] + \left(1-\eta \right)^2 \left[\frac{1}{3} D_4 + \frac{1}{4} D_5 + \frac{1}{5} D_6 \right] \right\} \tag{33}$$

$$C_{m_\alpha} = \frac{-2\Theta}{\gamma+1}\left\{\eta\left(1-\eta\right)\left[\frac{1}{2}D_1 + \frac{1}{3}D_2 + \frac{1}{4}D_3\right] + \left(1-\eta\right)^2\left[\frac{1}{3}D_1 + \frac{1}{4}D_2 + \frac{1}{5}D_3\right]\right\} \tag{34}$$

$$C_{m_{q_0}} = \frac{-2\Theta}{\gamma+1}\left\{\eta\left(-\eta\right)\left[\frac{1}{3}D_4 + \frac{1}{4}D_5 + \frac{1}{5}D_6\right] + \left(1-\eta\right)^2\left[\frac{1}{4}D_4 + \frac{1}{5}D_5 + \frac{1}{6}D_6\right]\right\} \tag{35}$$

where the subscript zero indicates that the pitching axis has been taken through the nose. The above stability derivatives have been plotted in figure (5) as a function of the bluntness parameter η for a 10° cone in an atmosphere where $\gamma = 1.4$.

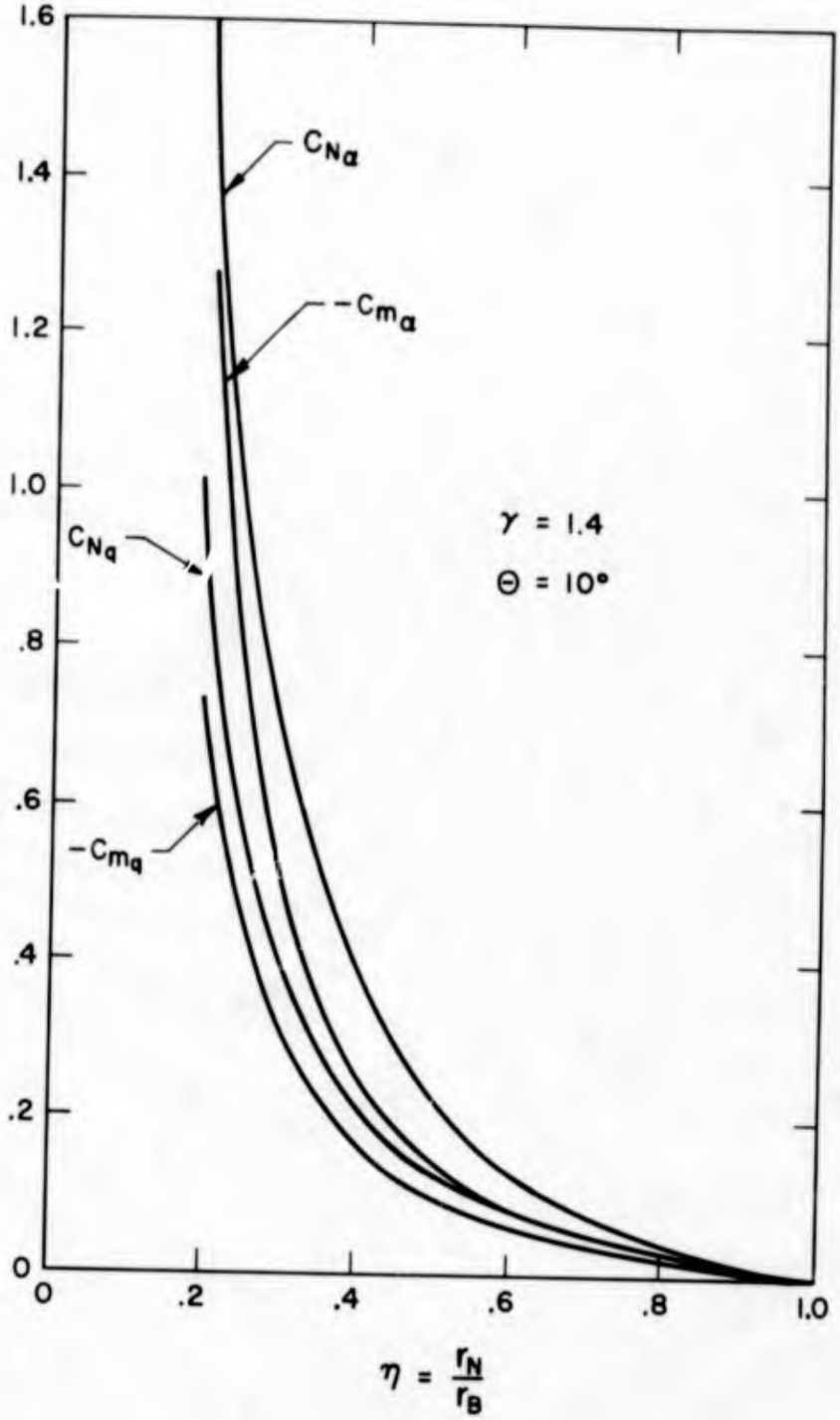

1.6
1.4
1.2
1.0
.8
.6
.4
.2
0
0
.2
.4
.6
.8
1.0
$C_{N\alpha}$
$-C_{m\alpha}$
C_{Nq}
$-C_{mq}$
$\gamma = 1.4$
$\Theta = 10°$
$\eta = \frac{r_N}{r_B}$

FIGURE 5

COMPARISON WITH EXPERIMENT

The experimental values for C_{m_α} and C_{m_q} which were established over a broad range of nose bluntness ratios and pitching frequencies in reference (1) serve as the basis of comparison with the preceding analysis. For the configurations with larger nose bluntness ratios (greater than approximately 10 per cent) these experiments found a slight and linear increase in the dynamic derivatives $C_{m_q} + C_{m_{\dot\alpha}}$, with increasing pitch frequencies, although the static derivatives were insensitive to changes in the frequency. Insofar as C_{m_q} is defined under the limiting process of frequency tending to zero, the experimental values of $C_{m_q} + C_{m_{\dot\alpha}}$ extrapolated to zero frequency have been plotted in figure (7). Frequency effects, if they exist in C_{m_q}, are of a higher order and would require a greater degree of precision in the power series expansions (5) and (23) in order to be delineated theoretically.

The stability derivatives given above in equations (32) through (35) are only valid for blunted cones performing pitching oscillations about an axis through the nose of the body. This unlikely circumstance, however, presents no serious limitation to the applicability of these results to bodies pitching about another axis location, in view of the axis transfer relations established by Tobak and Wehrand[4]. In brief, if the pitch axis is located a distance x_{cg} from the nose, the normal force and pitching moment derivatives become:

$$C_{N_q} = C_{N_{q_0}} - \frac{x_{cg}}{\ell} C_{N_\alpha} \tag{36}$$

$$C_{m_\alpha} = C_{m_{\alpha_0}} + \frac{x_{cg}}{\ell} C_{N_\alpha} \tag{37}$$

$$C_{m_q} = C_{m_{q_0}} + \frac{x_{cg}}{\ell} C_{N_{q_0}} - \frac{x_{cg}}{\ell} C_{m_{\alpha_0}} - \left(\frac{x_{cg}}{\ell}\right)^2 C_{N_\alpha} \tag{38}$$

while C_{N_α} is, of course, unchanged.

With the values of η and x_{cg}/ℓ corresponding to the experimental model of reference 1, C_{m_α} and C_{m_q} have been computed from the above expression and plotted in figure (6).

In view of the restriction (4) on ξ, the region of valid comparison in figure (6) extends up to values of the correlation parameter of reference 1, $\lambda = [\Theta(1-\eta)/2\eta]^{1/2}$ equal to around one-half. Within this

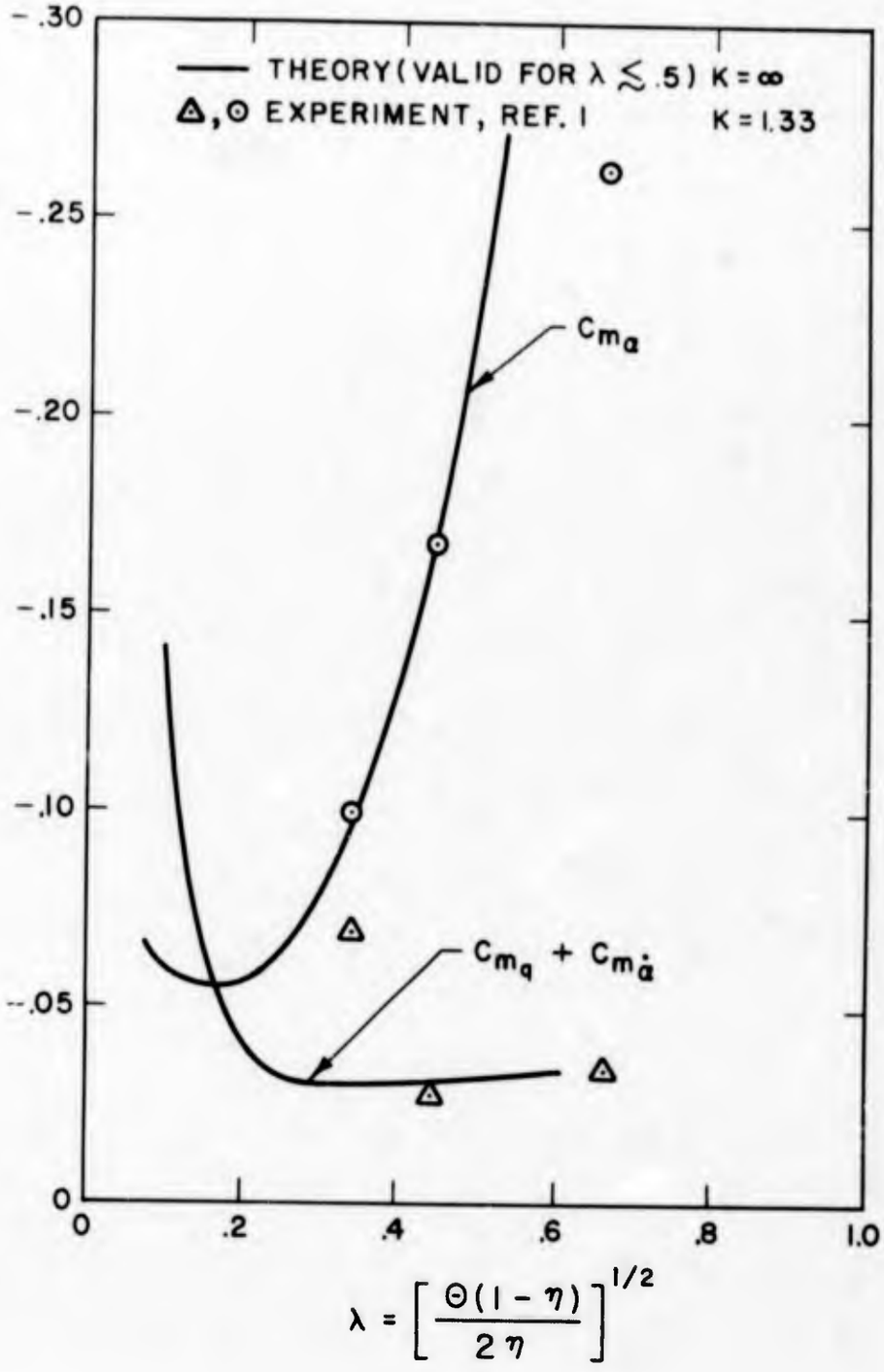

-.30
-.25
-.20
-.15
-.10
-.05
0
0
.2
.4
.6
.8
1.0
THEORY (VALID FOR λ ≲ .5) K = ∞
Δ, ⊙ EXPERIMENT, REF. 1 K = 1.33
C_{m_α}
$C_{m_q} + C_{m_{\dot\alpha}}$
$\lambda = \left[\frac{\Theta(1-\eta)}{2\eta}\right]^{1/2}$

FIGURE 6

range, the experimental and analytical values of C_{m_α} agree very well although there is an obvious discrepancy in the dynamic derivative $C_{m_q} + C_{m_{\dot{\alpha}}}$. This quantitative discordance might be attributable to the analysis' tacit assumption that the shock wave responds immediately to piston perturbations. With such a model, $C_{m_{\dot{\alpha}}}$ is zero since phase lags in the perturbations do not exist. On the other hand, the inability to separate the experimental C_{m_q} and $C_{m_{\dot{\alpha}}}$ unfortunately necessitates comparing the C_{m_q} of the theory with the $C_{m_q} + C_{m_{\dot{\alpha}}}$ of the experiment. It is also noted that the experiments were run at a Mach number of fourteen which corresponds to a value of 1.37 for the hypersonic similarity parameter K, while the theory has assumed that K is infinite. In view of the approximate nature of the model, the overall agreement is considered qualitatively satisfactory.

SUMMARY

An analysis of the effects of bluntness on the static and dynamic pitching moment derivatives of a slender cone has been presented. The theory was founded on Chernyi's approximate model of the flow around a blunted cone and analyzed the unsteady flow around the pitching body by perturbing its analogous piston.

The solution to the integrodifferential equation which describes the stationary flow field was approximated by a power series, as was the solution to the linearized perturbation equation. Equations for the stationary and perturbation shock shapes and pressure distributions have been presented while a comparison, over a defined range, between the present stationary pressure distribution and Chernyi's numerical solution confirmed the accuracy of the power series technique.

It was found that the expression for the perturbation pressure distribution contained a component which was in phase with the displacement and a component which lagged the displacement by ninety degrees. From this perturbation pressure, static and dynamic normal force and pitching moment derivatives were calculated with graphic results presented for a 10° half-angle cone.

With the aid of already existing axis transfer equations, the theoretical stability derivatives were compared with recent experimental results; and within the framework of such an approximate analysis, the agreement was satisfactory. In view of this, it can be concluded that the experimentally observed reduction in dynamic stability may be attributed to so-called blast wave effects.

ACKNOWLEDGEMENT

Grateful appreciation is extended to Mr. Otto Walchner for his continuing encouragement and advice during the course of this work.

REFERENCES

1. Walchner, O. and Clay, J. T., "Transactions of the Second Technical Workshop on Dynamic Stability Testing" (Sponsored by AEDC and ARO, Inc.), Volume I, Paper 8, (1965).

2. Griffith, B. J. and Lewis, C. H., "A Study of Laminar Heat Transfer to Spherically Blunted Cones and Hemisphere Cylinders at Hypersonic Conditions," AEDC TDR 63-102 (June 1963).

3. Chernyi, G. G., "Introduction to Hypersonic Flow," Translated by R. F. Probstein, Academic Press, New York (1961).

4. Tobak, M. and Wehrand, W. R., "Stability Derivatives of Cones at Supersonic Speeds," NACA TN 3788, (September 1956).

DEVELOPMENT OF A UNIQUE MOBILE INTEGRATED SUPPORT SYSTEM FOR TACTICAL ELECTRONICS (U)

CO-AUTHORS: MAURICE N. SCHEIDERICH and ANTHONY N. CIANCIO
Mechanical Engineering Section
Rome Air Development Center
Griffiss Air Force Base, New York

---

8 June 1966

Maurice N. Scheiderich

BIOGRAPHY

MAURICE N. SCHEIDERICH
Supervisory Mechanical Engineer
Mechanical Engineering Section

Born: March 13, 1921, at New Hartford, New York

Education: Elementary and High School - Utica, N. Y.

College - Mohawk College, Utica, N. Y., 1946 - 1948;
Clarkson College of Technology, Potsdam, N. Y., 1948-1950,
B.M.E.; Graduate Courses, Syracuse University, Syracuse, N.Y.

After graduation, Mr. Scheiderich worked as a design engineer on pneumatic tools for Ingersoll-Rand Corporation, Athens, Pa., until 1951. From 1951 to the present he has worked in the Mechanical Engineering Section, RADC, where he has held positions involving the design of mobile installations, shock and vibration, mechanical systems support, and project engineer for special structures and equipment.

From 1959-1963 Mr. Scheiderich served as lead engineer for the ancillary subsystem of Air Weapons Control System. In 1964 he was made supervisor of the Mobility Engineering Group, the position he now holds.

Mr. Scheiderich has acted as consultant to SPO's in the area of mobility, and to RADC in programs involving mobility and mechanical engineering.

Mr. Scheiderich is presently the chief investigator in the in-the-house program for the design and fabrication of the Mobile Integrated Support System.

He is a member of the American Society of Mechanical Engineers.

Anthony N. Ciancio

BIOGRAPHY

ANTHONY N. CIANCIO
Mechanical Engineer
Mechanical Engineering Section

Born: December 24, 1937, at Dobbs Ferry, New York

Education: Elementary and High School - Dobbs Ferry, N. Y.

College - New York State Maritime College, Bronx, N.Y., 1955-1959, Bachelor of Marine Engineering, Third Assistant Engineering Officer, Steam and Motor Vessels.

Following graduation Mr. Ciancio was employed at the Light Military Electronics Department of the General Electric Company, involved in the design of packaging electronic test equipment. In 1964, Mr. Ciancio accepted an appointment with the Rome Air Development Center where his present duties include development of mobilizers and packaging techniques for electronic systems.

Mr. Ciancio is an Associate Member of the American Society of Mechanical Engineers.

BLANK PAGE

ABSTRACT

DEVELOPMENT OF A UNIQUE MOBILE INTEGRATED SUPPORT SYSTEM FOR TACTICAL ELECTRONICS (U)

CO-AUTHORS: MAURICE N. SCHEIDERICH and ANTHONY N. CIANCIO
Mechanical Engineering Section
Rome Air Development Center
Griffiss Air Force Base, New York

---

8 June 1966

At present, tactical electronic equipments are transported on highways or cross-country using conventional means, such as trailers and standard "M" series trucks. These equipments are housed in hut-type shelters 8 to 12 feet long. This standard method of providing mobility has a number of serious disadvantages. One 12 ft package weighing 4000 lbs requires one standard military M-35 truck, weighing 12,600 lbs, for transport; one power generating package, of at least 500 lbs, either towed or carried on another vehicle; and a loading-unloading device, estimated to weigh 200 lbs, for removing the package from the back of the truck. The combined weight of these support items is 13,250 lbs. The ratio of support weight to equipment weight is 3.3:1. One C-130 aircraft is required to transport this one-package system; two helicopters are required to transport the electronic package and power. In addition, it will require at least 45 minutes for four men to unload the equipment from the truck and hook up power.

Our in-the-house design of a mobile integrated support system, which will be described in this paper, has the following goals: (1) a support to equipment weight ratio of 1:1; (2) volume reduction such that two complete systems will fit in one C-130 aircraft; (3) helicopter transport of the complete system by two helicopters; (4) transport set-up time of 10 minutes by two men.

Our mobile integrated support system is being designed to meet the specific mobility requirements of tactical electronic equipment. Road transport life of an electronic package is estimated at less than 10,000 miles. A top road speed of 45 mph is sufficient for convoy operation. Cross-country operation with the capability to climb a 60% slope is a necessity. By designing specifically to these requirements, a decrease in weight and complexity can be achieved over conventional design.

This mobile integrated support system is basically a two-unit transporting system. The front wheeled unit consists of the prime power plant, steerable front end, and cab. This unit attaches to the front of the electronic package. The rear unit is an axle and wheel assembly which attaches to the rear of the shelter. All four wheels of the system are driven.

The main power plant is used in a dual role of providing <u>both</u> electrical power and motive power. The engine, a 100 HP diesel, is used to drive a 32 kw alternator to provide electrical power for operation of the electronic equipment. The diesel is also used to drive a pump which supplies high pressure hydraulic fluid to hydraulic motors located at each wheel. The support system is capable of raising the electronic package from the ground to the transport position or lowering it to the ground from the transport position. When the front and rear units are

detached from the package they can be joined together for movement without the electronic package.

The total support system weight is 4000 lbs and can be lifted by helicopter as a unit. This capability will allow full mobility of the system after helicopter transport.

The mobile integrated support system will give full tactical mobility to any equipment or system such as radar, communications, navigation, etc., that will fit into an 8 to 12 foot long package, weigh up to 4000 lbs, and require up to 32 kw of power.

INDEX

DEVELOPMENT OF A UNIQUE MOBILE INTEGRATED SUPPORT SYSTEM FOR TACTICAL ELECTRONICS

I. SUMMARY

Mobility, for tactical electronic equipment, can be improved significantly if supporting equipment is designed specifically for electronic systems. Mobility cannot be added to a system as an afterthought, but must be considered with and be part of the over-all system design. Designing for the specific tactical requirements and integrating the functions of the supporting equipment, where possible, will reduce system weight, support aircraft requirements and set-up time. This paper describes a support system designed for transporting an S-141 shelter type package and compares this system with the conventional means of transporting this electronic package.

II. INTRODUCTION

The need for mobility of our tactical ground electronic equipment has been expressed many times in requirements, presentations, and operational plans, but little advancement has been made in this field. Continual advancement is being made in decreasing the weight and volume of the electronic equipment, but often this weight and size reduction is accompanied by a need for increased electronic capability resulting in no reduction in package size or weight. Our studies have shown that the support system, required to transport and to power an electronic package, is usually two to three times the weight of the electronic package. The ratio of the volume of the support system to the electronic package is at least two to one.

This severe penalty we are paying because of the support equipment is detrimental in several ways: (1) the over-all mobility of a system is poor because of limited cross-country capabilities, long set-up time, and reduced handling ability; (2) parts of the support system, such as the truck, cannot be helicopter lifted, thereby restricting or eliminating movement of the electronic package after a helicopter movement; (3) excessive numbers of cargo aircraft are required because of the bulk and weight of the support items. It becomes apparent, then, that reduction of support equipment weight and volume is of prime importance if we are to improve the mobility of our electronic systems.

The term, "support system", as used in this paper, includes all the ancillary equipment necessary to transport, lift, and power the electronic package. The term, "electronic package", includes the shelter and all equipment mounted or stored within the shelter. The term, "electronic system", includes both the electronic package and the support system.

III. REQUIREMENTS

A tactical system must meet the following general operational and technical requirements:

1. Mobility - By definition, quickly and easily movable. The equipment must be capable of being moved from place to place within a theater of operations over highways, secondary roads, and cross-country. Transportation over long distances, within the theater of operations and to the theater, must be possible by C-130 aircraft. Air mobility must be obtained with the least possible number of cargo aircraft. Movement within the theater by helicopter must be possible and should include sufficient support equipment for relocation of the basic electronic equipment.

2. Weight and Bulk - To meet the mobility requirements, the weight and bulk of the complete support system must be held to an absolute minimum. Individual package weights should be no more than 4000 lbs for helicopter lift.

3. Set-up Time - The time required to set up an electronic system, from transport condition to operating condition, or from operating condition to transport condition, is an inherent characteristic affecting the over-all mobility of the system. Mobility is improved as much by a reduction in set-up time as by a reduction in road-time or flight-time.

4. Power - Electrical power for:

   a. Operation of the electronic equipment

   b. Operation of the cooling equipment

5. Package Handling - Package positioning, unloading, loading.

IV. ANALYSIS OF A CONVENTIONALLY SUPPORTED SYSTEM

For the purpose of this analysis, systems for supporting one electronic package and two electronic packages were considered. The one package system provided the best mobility and the two package system resulted in the best ratio of payload to support system weight. The electronic package can be any electronic equipment installed in a 12 ft long S-141 shelter with a total package weight of 4000 lbs.

A single electronic package is generally transported in the back of an M-35, 2 1/2 ton truck, as shown in Figure 1. The M-35 truck weighs 12,600 lbs. An unloading-loading mechanism is required to lift the package in and out of the truck. This mechanism consists of a winch, cables, and rollers, and weighs approximately 200 lbs. A power generating set is also required to provide electrical power for operation of the electronic equipment. For the purpose of this comparison, a 30 kw, 400 cycle, turbine driven generator was chosen on the basis of weight and size only. This set weighs 450 lbs. The generator set would be transported inside the electronic package and removed for operation. (It should be noted

that, if longer life and a lower fuel consumption is required, a diesel driven generator would have to be provided in the system. A suitable 30 kw, 400 cycle diesel driven set would have to be trailer mounted and would weigh about 5500 lbs, including the trailer.) The total weight of this system is 17,250 lbs - 4000 lbs for the electronic package and 13,250 lbs for the support system. One C-130 cargo aircraft would be required to transport the system. Two helicopters would be needed just to transport the electronic package and the generator set. It is estimated that four men would require 45 minutes to set up the system (not including electronic set-up). After the electronic package is transported by helicopter, the package will remain in its location. There is no suitable way to move or reposition it.

Two electronic packages can be moved by placing one package in the back of an M-35 truck and mounting the second package on a mobilizer and towing it behind the truck. This configuration is shown in Figure 2. The total two package system weight is 23,350 lbs - 8000 lbs of electronic package and 15,350 lbs of support equipment. Airlifting this system requires the cargo space of one and one-half C-130 aircraft. Three helicopters are required to transport the operating equipment only. It is estimated that four men will require 60 minutes to set up these packages, excluding electronic set-up.

The cross-country mobility would be limited because of the trailed package. The electronic packages would have no ground mobility after being positioned by helicopter.

Two major areas of redundancy are present in the conventional system. An engine is used to drive the truck for surface mobility. A second engine is provided to drive the generator for electrical power. The vehicle engine is idle during electronic equipment operation and the power unit engine is idle during road transport. One of the two engines is idle at all times. The second area of redundancy is the duplication of chassis. The S-141 type shelter, in which the electronic equipment is mounted, is capable of supporting its rated payload without additional chassis support. In the conventional system the shelter is additionally supported by the chassis of the M-35 truck. Elimination of this redundancy will decrease system weight.

V. ANALYSIS OF AN INTEGRATED SYSTEM APPROACH

The operational and technical requirements can better be met with an integrated support system. This system is designed to transport any electronic equipment that is installed in an S-141 type shelter with a maximum package weight of 4000 lbs and a maximum length of 12 feet. The electrical power requirement of the system should not exceed 32 kw. An analysis of such an integrated system indicates that the following criteria should be used as a basis for the system design:

a. Size Requirements:

Sizing a hydrostatic drive involves many of the same parameters as any other drive. In vehicular applications, the maximum and minimum speeds under various conditions of load, grade, and acceleration must be determined. The torque requirement will determine maximum system pressure; the motor speed will determine system fluid flow.

Full horsepower must be developed under two extremes of speed and torque; at low speed to negotiate a 60 per cent grade, and at high speed to overcome rolling and air resistance on level roads. The high torque requirements at low speed determine the displacement of the hydraulic motor and the gear ratio used. The upper speed limit of the motor, and the pump capacity, determine the maximum vehicle speed.

b. Description:

Versatility was a major reason for selecting a hydrostatic drive for the MISS. It offers many advantages over the conventional drive shaft-differential power transmission.

A hydrostatic drive provides stepless control throughout the entire speed range, and high efficiency over a wide range of output torques. This makes it possible to match engine power to the vehicle requirements throughout the range of speed and torque. By restricting return flow from the motor, dynamic braking is provided for the vehicle. Travel direction can be changed smoothly by a single lever which reverses the pump inlet and outlet ports without gear change. At zero pump delivery, the transmission will not creep. A relief valve limits the load on all mechanical members in the power train. Operating pressure and pump delivery can be controlled to prevent engine overloading. The smooth application of torque reduces the tendency to spin the wheels. Hose and flexible joints used in hydraulic lines result in freedom of motion between units. The elimination of a mechanical drive train permits the new and unusual design concept of MISS.

c. Operation:

In a hydrostatic drive, energy is imparted to the fluid by a positive-displacement pump and is transmitted to a positive-displacement motor where it is transformed into useful work.

This particular application consists of a single, variable displacement, axial piston pump, and four motor units, with associated control and fluid conditioning components. The use of axial piston type units allow for higher pressures; this reduces the size of components and lines required for a given power capacity. It is a closed circuit and not adversely affected by environmental conditions.

1. Design life, for surface mobility, should be approximately 10,000 driven miles between major overhauls. This includes highway and cross-country mileage under conditions where first class highways are severely damaged or do not exist. This distance will certainly be adequate for tactical operations since the prime purpose of the package is electronic operation, with less than 10% of its life spent in surface transport.

2. No design restrictions need be placed on the type of vehicle used to obtain surface mobility because:

a. The vehicle need not be suitable for general purpose or cargo use because it will be used solely to support the electronic package either for transport or for electrical power.

b. The advantages of parts standardization becomes insignificant in comparison with the potential systems benefit.

VI. GENERAL DESCRIPTION OF THE MOBILE INTEGRATED SUPPORT SYSTEM (MISS)

The main drive engine of the MISS is a lightweight diesel which is used in a dual role. It is used to drive the alternator to produce electricity for electronic equipment operation. It is also used to drive the hydraulic pump to provide power to the wheels for surface transport.

The Mobile Integrated Support System is a hydraulically driven system made up of two sections - the front section contains the power generating equipment, automotive type controls, steering, package lifting device, two driving wheels with a hydraulic motor in each, and two retractable support wheels. The rear section contains two driving wheels, with a hydraulic motor in each, and a device for lifting the package. The sections are attached to the front and the rear of the electronic package and, when attached, are capable of lifting the package to a road transport height of 14 inches, or lowering it to the ground. The front and rear sections are interconnected hydraulically to provide four-wheel drive for the system. Figure 3 shows an artist's concept of the MISS. When the front and rear sections are removed from the shelter they can be connected together for road movement or helicopter lift. The total weight of this system is 8000 lbs -- 4000 lbs for the Mobile Integrated Support System and 4000 lbs of electronic package. Figure 4 shows the MISS configuration.

VII. DESIGN FEATURES

Significant features of the Mobile Integrated Support System are as follows:

1. Vehicular Drive System:

A hydrostatic drive was selected as the system offering most promise in this application.

In the system chosen for MISS, engine power is transferred to a nine piston, variable displacement, 3500 psi main system pump capable of supplying 78 gpm at 2500 rpm. (See Figure 5) The pump tilt plate can be displaced angularly from 0° to 20° , either side of center, to adjust the pump displacement from zero output to .031 gpm/rev. The pump is nominally supercharged to 120 psi by a hydraulically driven gear pump mounted on the main pump. A crossline relief valve, mounted on the main pump, protects the pump from overload pressures, and through pilot control of this valve, pressures are regulated to control coasting and braking.

Hydraulic power is fed to four axial piston motors, one mounted within each wheel. The hydraulic motors are similar to the main pump. Each front wheel motor has a three position stroke cylinder to obtain 0°, 6°, and 22° tilt plate angles which provide the three speed ranges of the vehicle. Each rear wheel motor has a stroke cylinder which positions the tilt plate at 6° and 22°. In the low speed range, all the motor tilt plates are at the 22° angle; in medium speed, all tilt plates are at the 6° angle. However, in the high speed range the rear motor tilt plates are set at 6° and the front motor tilt plates go to zero displacement.

The wheel motor stroke positions are controlled by a manually operated directional valve, utilizing a 1000 psi supply from a dual volume gear pump mated to the multiple power take-off. The operator selects a high, medium, or low speed range, which in turn shifts the motor tilt plates to the angular position controlling motor displacement.

Operation of the accelerator pedal simultaneously varies the main pump governor and the engine governor. Operation is such that any tendency of the engine to stall results in reducing the stroke of the main pump, thereby decreasing the torque requirements of the pump. The governor is adjustable to set up a variety of power curves to meet specific applications.

d. Hydraulic Wheel Hub:

All of the wheels of the support system are powered. Each hydraulic motor is mounted in an aluminum hub to which the tire and rim are attached. The motor drives the rim of the hub through a 7.35 to 1 gear reduction. The hub has been designed so that the hydraulic motor may be removed for service or repair by simply detaching the hydraulic lines and the mounting bolts. No jacking or other disassembly is necessary for motor removal. (See Figure 6)

The hub also contains a disc brake system. This system serves as a parking brake and as a back-up for the dynamic braking system. An automatic wear compensating device has been incorporated in the disc brake system.

2. Power:

Power will be required to produce electricity for equipment operation as well as drive the system in road and cross-country operation. Thirty-two kw of electrical power are provided for operation of the electronic equipment in the shelter. The ability to negotiate 60% slopes or maintain a speed of 45 mph on highways requires up to 70 horsepower. (Sample calculations are in appendix.)

The prime power source for the Mobile Integrated Support System will be a new type of 100 BHP, two-stroke cycle, diesel engine developed by Dynastar Laboratories. This engine has adequate power, torque and speed relationships to meet the mobilizer and power requirements. It was selected because, in addition to the usual advantages of the diesel, namely good fuel and long service life, the Dynastar engine has a unique cylinder configuration resulting in a lightweight compact design.

a. Description:

Basically, the diesel utilizes four pairs of "U" type twin cylinders arranged with the cylinder bores in a plane perpendicular to the crankshaft axes. Two pistons per cylinder block operate with a common combustion chamber. This dual piston operation provides the advantages of the opposed piston type, in a simpler and more compact package.

The cylinders of each pair are at an angle to each other which allows one piston of a pair to lead or trail the other, (Figure 7), establishing the proper phase relationship between exhaust and inlet port openings. Additional port timing control may be obtained by varying the distance between the knucklepins of each cylinder pair.

The pistons of each pair of cylinders are connected to a "true-motion" frame by individual connecting rods. The true motion frame rides on the crankpins of two single throw crankshafts. The frame links the crankshafts together so that they rotate in phase, and any point on the frame describes a true circle, whose radius is equal to the crankthrow. A common output shaft is geared to the two single throw crankshafts.

This diesel engine is liquid cooled and weighs approximately 400 pounds, with envelope dimensions of 26 x 26 x 21 inches. General specifications are: bore 3 inches; stroke 3 1/2 inches; rated 100 hp at 2700 rpm.

b. General Advantages:

As a two-cycle operation with no conventional valving, the diesel engine has the characteristic of simplicity. This lends itself to

multiple grouping of small cylinders which can stand higher specific outputs (BHP/sq. in. of piston area) because the smaller pistons dissipate excessive heat more efficiently than large ones.

Although this design basically replaces one large cylinder with two small cylinders, these cylinders act as a single combustion unit; therefore, it is not necessary to double up on fuel injection units -- the number of such units remains the same as for the large cylinder configuration. This saves weight and reduces cost. Furthermore, two small cylinders provide 41% more cylinder periphery for ports than would one larger cylinder of the same piston area.

Only four holddown bolts are required for each cylinder pair, and the cylinder heads are integral with the cylinder blocks, eliminating the cylinder head gaskets.

Good dynamic balance in an engine is achieved when all shaking forces are balanced out before they reach the main bearings. In this engine, all the inertia loads of the pistons, rods, and frame are balanced out by the counterweights within the crankshaft structure, and thus they never reach the main bearings on the crankcase. The engine operates almost without vibration.

The basic structural configuration of the diesel is an "X". This has proven to be the most efficient in regard to structural integrity. It requires a minimum of redundant structure and provides the most uniform stress loading throughout. This contributes to the light weight, compactness, and excellent installation characteristics.

3. Multiple Output Gear Box:

The gear box was designed specifically for the Mobile Integrated Support System and provides power outputs for the following components:

1. 30 kw Alternator
2. Main System Hydraulic Pump
3. Dual Volume Control Pump
4. Power Steering Pump
5. Water Pump

Mounting pads for attachment of the engine-transmission assembly to the vehicle frame are located on two sides and rear of the gear box. There are two modes of operation. Mode selection is provided by a single lever which engages either the 32 kw alternator or the main system and control pumps. Neither mode can be engaged or disengaged while the engine

is running, nor can the main system pump and alternator operate simultaneously. The power steering and water pumps are driven during both modes. (See Figure 8)

4. Alternator:

The alternator provides power for the electronic package. This AC generator is a self-excited, air-cooled, brushless type machine, designed for use in aircraft. It is lightweight, weighing only 85 lbs, and delivers 40 kv-a (32 kw) output. It delivers this rated output when driven at 6000 rpm. The output is three phase, four-wire, 120/208 volts, 400 cps. Dimensionally, its mounting flange diameter is 11 inches and its length (from the mounting flange) is approximately 13 1/2 inches. (See Figure 9)

Brushless design means the elimination of brushes, commutators, and slip rings of the conventional design, with a resultant simplification of maintenance and increase in reliability. Such design was made possible by the development of high-temperature silicon rectifiers. The three-phase output of the rotating exciter armature is fed into a three-phase, full-wave, rotating rectifier where it is converted to direct current. This direct current provides the needed excitation of the main rotating field of the generator. The exciter armature, the rectifier circuit, and the generator field are all mounted on the same shaft; therefore there is no relative motion between them, and DC excitation to the generator field is provided without brushes, commutators, or slip rings.

The voltage regulator, control panel (for control and protective functions), control switch, current transformers, and a load contactor (circuit breaker) are mounted external to the generator.

5. Suspension System:

The suspension system, used for MISS, includes independent front wheel suspension using a parallel arm mechanism with air springs. The rear unit uses a floating axle with air springs supporting the load. The air springs are independently adjustable to compensate for uneven weight distribution. The suspension system is designed to give protection to the electronic equipment during cross-country and highway movement. Aluminum has been used extensively to keep the weight to a minimum.

6. Mobility Characteristics:

The MISS is designed to meet the following service requirements:

1. Terrain: Cross-country, unimproved roads, trails, and open rolling, hilly terrain.

2. Gradeability:

a. 60% slope at 2 mph.

b. Highway travel at 45 mph.

3. Life: 10,000 miles before complete overhaul.

4. Payload: 4000 lbs - electronic package.

5. Air Mobility:

a. Transport by C-130 aircraft: Two MISS's with payloads will fit in one C-130 aircraft.

b. Helicopter Transport: Two CH-3 helicopters are required to transport the complete system with payload. Full mobility is available after helicopter lift.

c. Set-up Time: The set-up of the MISS is an extremely simple operation. After an electronic package is driven to the operating location, the package is lowered to the ground and the power cable is connected. With two men, the time required for this set-up is 10 minutes. This does not include the electronic set-up of the package. If the package can be operated without lowering to the ground, set-up is limited to connection of the power cable.

A comparison of weights, mobility, and set-up of the conventional support systems and the MISS is shown in Figure 10.

VIII. APPLICATION

The MISS can be used to transport any electronic equipment package, installed in an S-141 type shelter, with a total package weight of 4000 lbs. This means that most tactical electronic packages, with a helicopter transport requirement, can be handled by this system. No major changes have to be made to the shelter if it has the standard mobilizer attachments. The hydraulic lines for the rear unit can be temporarily attached to the shelter. The artist concept, Figure 11, shows applications of MISS in the field.

This paper has described a prototype of a specially designed transporting system for electronic equipment. This initial design, used to prove the technique, incorporated off-the-shelf components. Future models of this system will take advantage of new or specially designed components to reduce the weight and improve the reliability.

APPENDIX

SAMPLE CALCULATIONS

SYMBOLS

A = Cab frontal area, sq. ft.

Ca = Coefficient of air resistance, lb-sec^2-ft^{-4}

Dm = Motor displacement, in^3 per rev

F = Tractive force, lb

f = Coefficient of rolling resistance

Nm = Motor speed, rpm

P = Power consumption of hydraulic pump, hp

p = Pump discharge pressure, psi

Δpm = Pressure drop across motor, psi

Qp = Pump discharge flow, gpm

r = Rolling radius, in.

Ra = Air resistance, lb

Rg = Grade resistance, lb

Rr = Rolling resistance, lb

Tm = Motor torque output, lb-in

Vr = Vehicle speed relative to air, mph

W = Vehicle weight, lb

ηp = Total pump efficiency

ηv = Over-all volumetric efficiency

γ = Wheel gear ratio

θ = Slope angle, deg

I. Power required for a 60% slope - 2 mph:

At low speeds air resistance is considered negligible and at constant speed tractive force balances only rolling and grade resistance.

$$F = Rr + Rg$$

$$= fW = W \sin \theta$$

$$= (.06)(8000) + (8000) \sin 31^{\circ} = 4600 \text{ lbs}$$

$$Tm = \frac{(F)(r)}{4}$$

$$= \frac{(4600)(14.5)}{(4)(7.35)} = 2261 \text{ lb-in}$$

$$\Delta pm = \frac{2\pi Tm}{Dm} = \frac{2\pi(2261)}{7.24} = 1961 \text{ psi}$$

Qp = pump delivery at 2 mph

$$Qp = \frac{4 \, Nm \, Dm}{231 \eta v} = \frac{(4)(171)(7.24)}{(231)(.90)} = 23.8 \text{ gpm}$$

Power consumption of hydraulic pump

p = 2000 psi

ηp = 80%

$$P = \frac{QpP}{(1714)(\eta p)} = \frac{(23.8)(2000)}{(1714)(.80)} = 34.7 \text{ HP input to pump}$$

II. Power required for 45 mph speed on a 0% grade:

$$Qp = \frac{2 \, Dm \, Nm}{\eta v \, (231)}$$

$$= \frac{2(2.23)(3830)}{(.90)(231)} = 82 \text{ gpm}$$

Tractive force required principally to overcome rolling and air resistance (headwind excluded).

$A = 46.6$ sq. ft.

$$F = Rr + Ra$$

$$= fW + Ca\,A\left(\frac{Vr}{10}\right)^2$$

$$= (.012)(8000) + (.26)(46.6)(20.2)$$

$$= 96 + 244.5 = 340.5 \text{ lbs}$$

$$Tm = \frac{(F)(r)}{2\gamma} = \frac{(340.5)(14.5)}{(2)(7.35)} = 336 \text{ lb-in}$$

$$\Delta pm = \frac{2T_m}{D_m} = \frac{2(336)}{2.23} = 947 \text{ psi}$$

Power consumption of hydraulic pump:

$p = 1025$ psi

$$P = \frac{(Qp)(p)}{(1714)(\eta p)} = \frac{(82)(1025)}{(1714)(.80)} = 61.3 \text{ HP}$$

SYMBOLS

$$H = H_{pa}\,\frac{P}{P_a}\,\sqrt{\frac{T_a}{T}} \quad \text{where}$$

P = absolute pressure

P_a = absolute pressure (at standard atmospheric conditions)

T_a = absolute temperature (at standard atmospheric conditions)

T = absolute temperature

H_{pa} = horsepower (at standard atmospheric conditions)

H_p = horsepower at some specified altitude

III. Engine Power for 32 km:

$$H_p = H_{pa} \frac{P}{P_a} \sqrt{\frac{T_a}{T}}$$

At 10,000 ft and 23.3°F

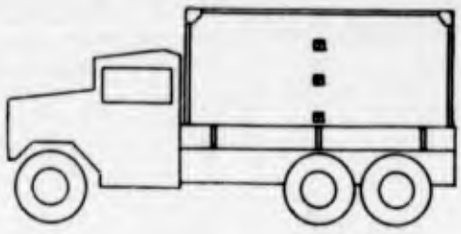
$$H_p = H_{pa} \frac{20.58}{29.92} \sqrt{\frac{518.7}{483.0}}$$

$$H_p = H_{pa} \; .691 \sqrt{1.063} = .714 \; H_{pa}$$

$$H_{pa} = 78$$

$$H_p = (.714)(78) = 56$$

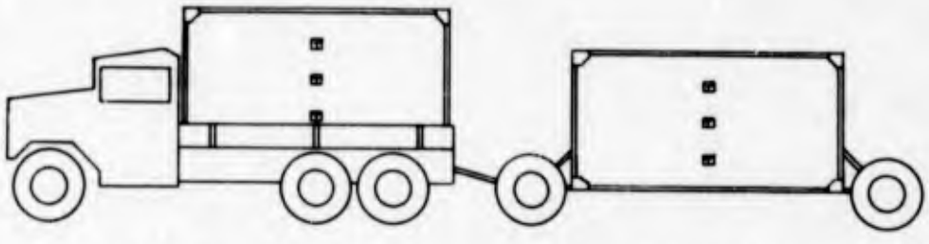
CONVENTIONAL TRANSPORT—ONE PACKAGE



WEIGHT		HANDLING REQ'D	
M-35 TRUCK	12,600 LBS	2—HELICOPTERS	(NO MOB.)
ELECTRONIC PKG	4,000 LBS	1—C-130 AIRCRAFT	
TURBINE GEN	450 LBS	4—MEN SET UP TIME	45 MIN
HANDLING DEVICE	200 LBS		
TOTAL	17,250 LBS		

Figure 1

CONVENTIONAL TRANSPORT—TWO PACKAGES



WEIGHT		HANDLING REQ'D.	
M-35 TRUCK	12,600 LBS	3—HELICOPTERS	(NO MOB.)
ELECTRONIC PKG'S	8,000 LBS	1½—C-130 AIRCRAFT	
HANDLING DEVICE	200 LBS	4—MEN SET-UP TIME	60 MIN.
TURBINE GEN	450 LBS		
UNDERCARRIAGE	2100 LBS		
TOTAL	23,350 LBS		

Figure 2

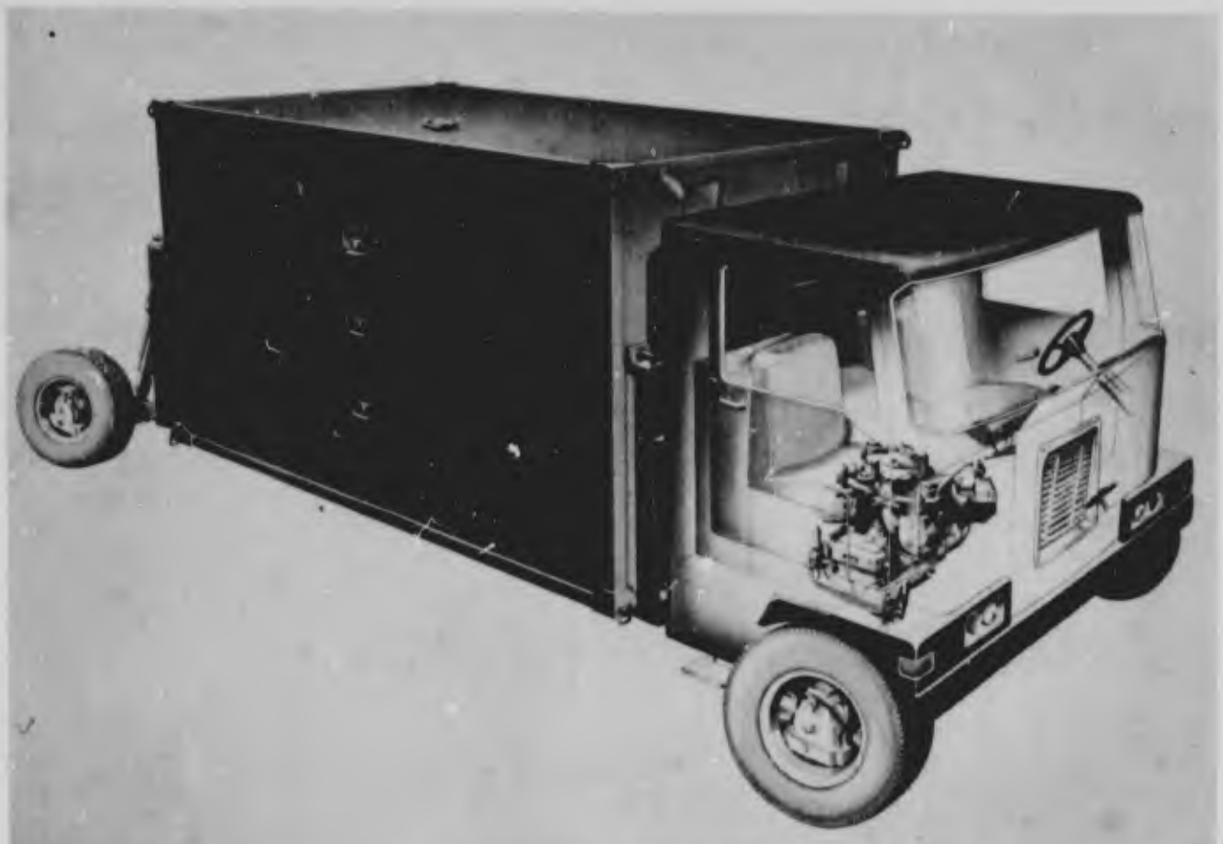
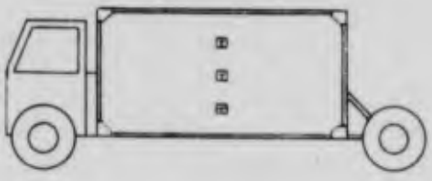


Figure 3 Artist Concept of MIBS

INTEGRATED SYSTEM—ONE PACKAGE



WEIGHT		HANDLING REQ'D.	
ELECTRONIC PKG.	4000 LBS	2—HELICOPTERS	(FULL MOB)
SUPPORT SYSTEM (INCL. POWER)	4000 LBS	½—C-130 AIRCRAFT	
TOTAL	8000 LBS	2—MEN SET-UP TIME	10 MIN
TO TRANSPORT TWO PACKAGES			
DOUBLE WEIGHTS			

Figure 4

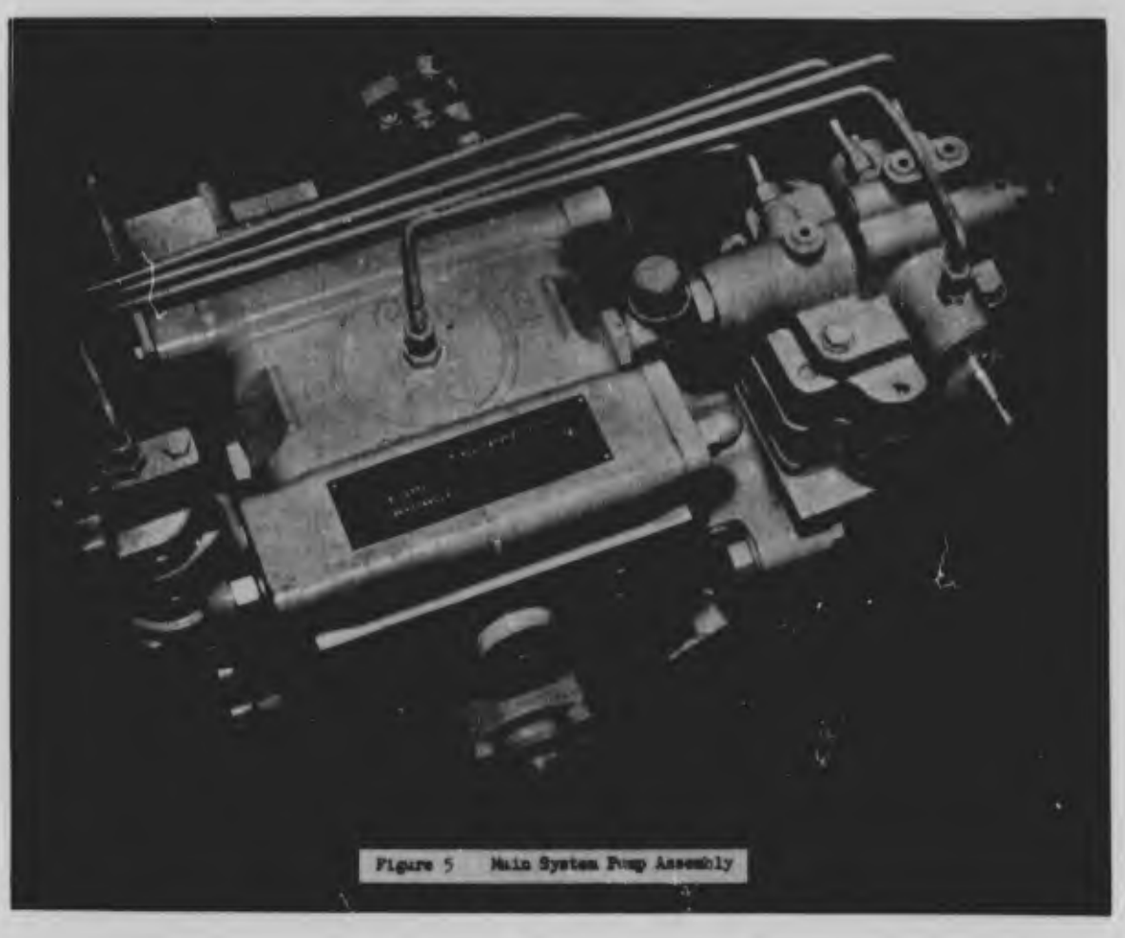


Figure 5 Main System Pump Assembly

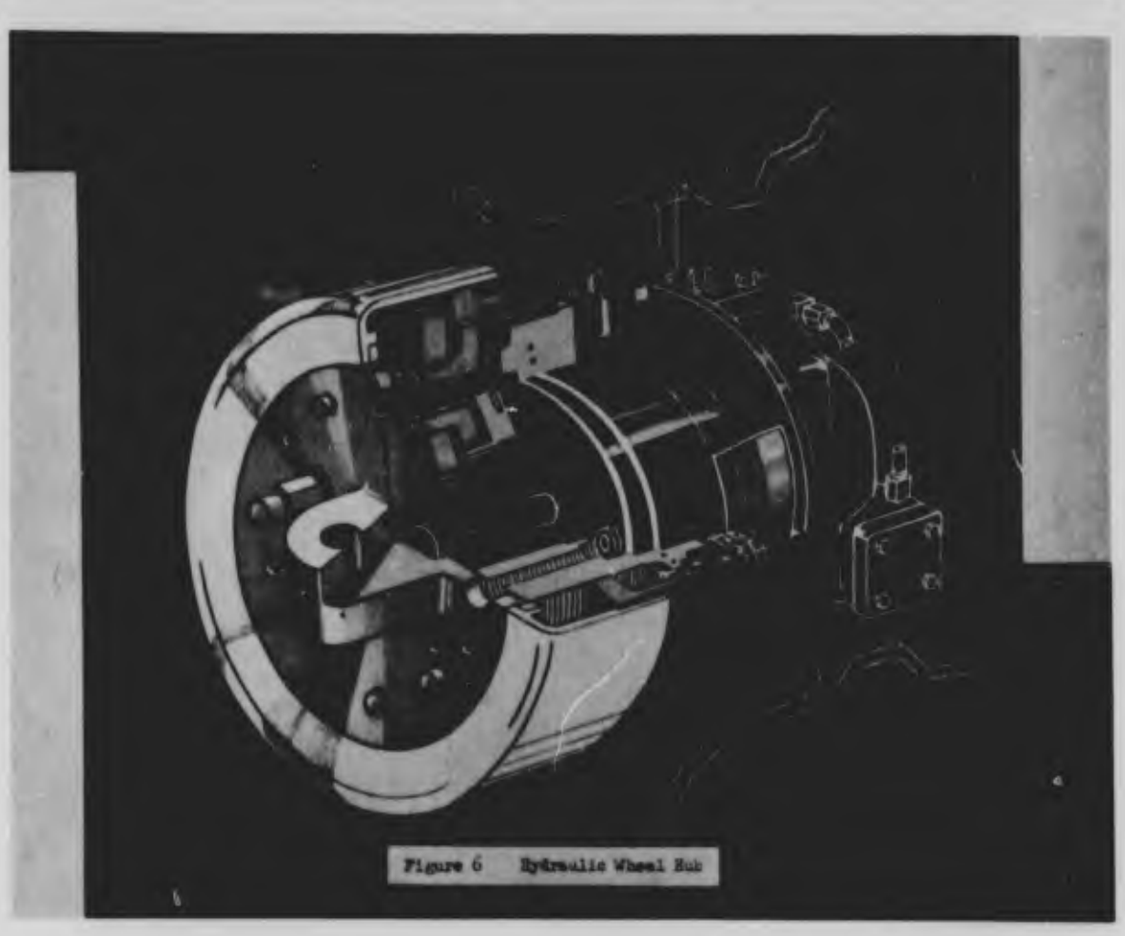


Figure 6 Hydraulic Wheel Hub

Figure 7 Diesel Engine Cross-section

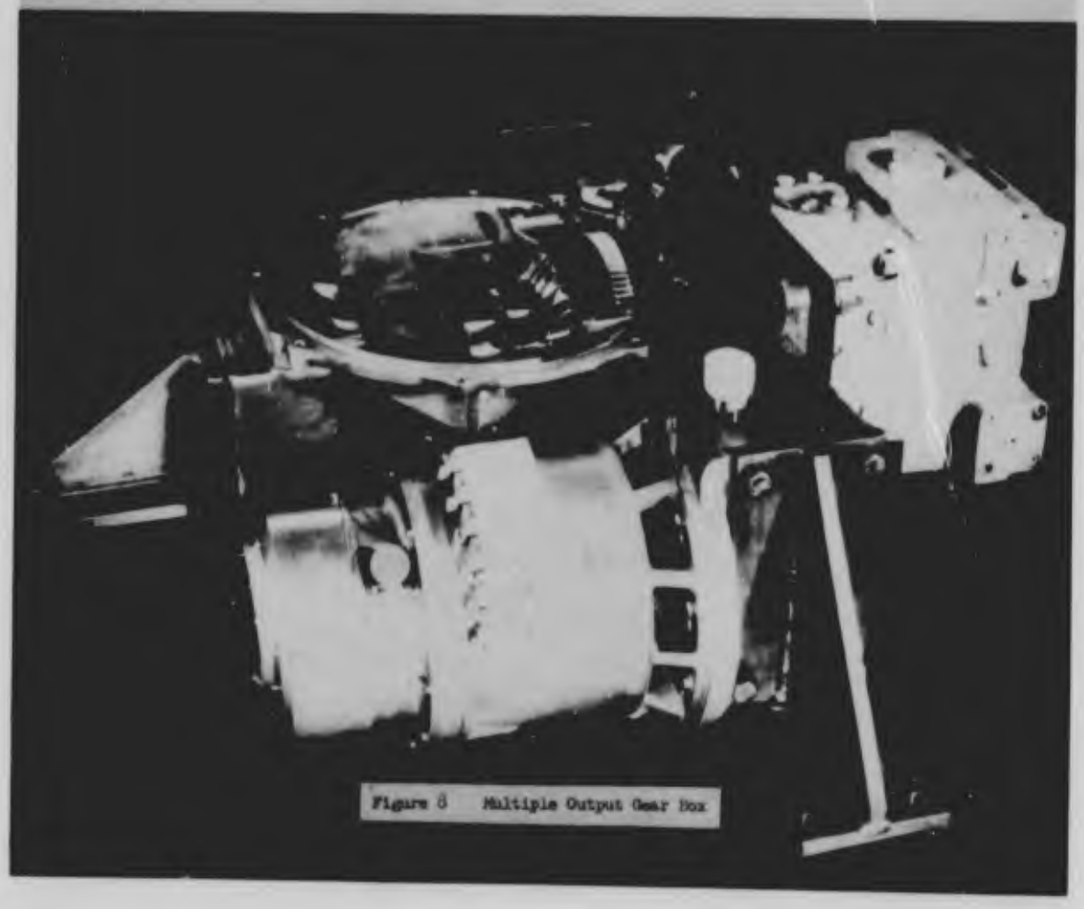


Figure 8 Multiple Output Gear Box

Figure 9 Alternator

COMPARISON OF CONVENTIONAL vs. INTEGRATED SYSTEM

	TRANSPORTING			
	ONE PACKAGE		TWO PACKAGES	
	CONV	INT	CONV	INT
WEIGHT TOTAL	17,250	8,000	23,350 (LM)	16,000
HELICOPTERS REQ'D	2 (NM)	2 (FM)	3 (NM)	3 (LM)
C-130 AIRCRAFT REQ'D	1	½	1½	1
SET-UP TIME — MIN	45 (4 MEN)	10 (2 MEN)	60 (4 MEN)	10 (4 MEN)

NM — NO MOBILITY

LM — LIMITED MOBILITY

FM — FULL MOBILITY

Figure 10

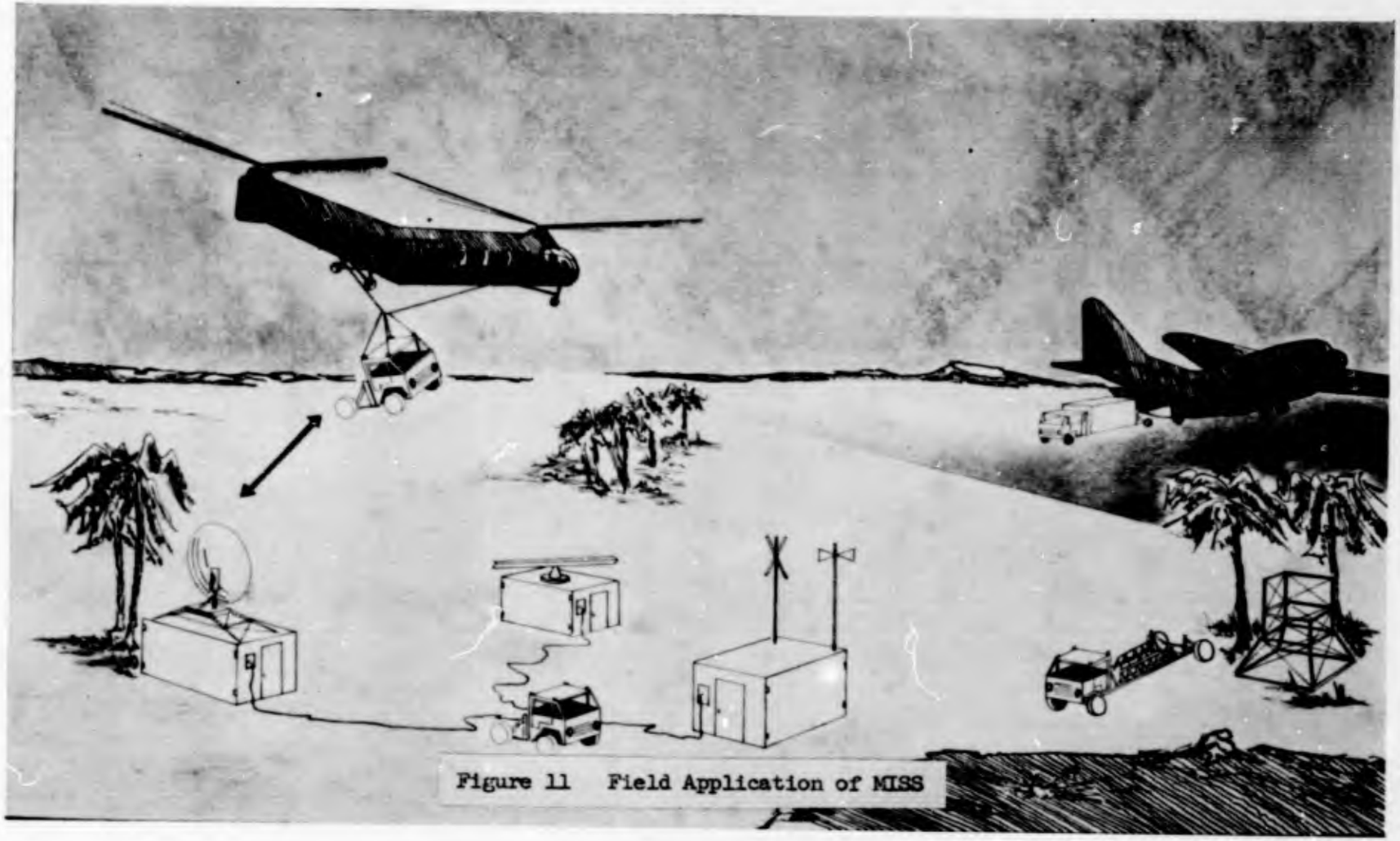


Figure 11 Field Application of MISS

(U) MACHINES THAT COMMUNICATE AND DESIGNERS WHO DON'T

By

J. ALBERT SOUTHERN

Systems Engineering Group
Research and Technology Division
Wright-Patterson Air Force Base, Ohio

J. Albert Southern

BIOGRAPHY (J. ALBERT SOUTHERN)

Mr. Southern is a Human Factors Engineer in the Systems Engineering Group at Wright-Patterson AFB, Ohio. He assisted in the application of the Voice Warning System to the B-58 and is presently engaged in several projects for limited warfare. He has completed graduate courses in human engineering at Ohio State University and U.C.L.A.

A native of North Carolina, he completed an A.B. and M.A. at Duke University before joining the faculty there. A pilot since 1940, he joined the staff at George Washington University (HumRRO) and developed techniques for the evaluation of Army helicopter training methods. A believer in learning at first hand, he completed helicopter pilot training as a part of his own indoctrination.

He joined the staff of the Aerospace Medical Laboratory in 1958 and served as Chief of the Manned Systems Section. Prior to the organization of the R&T Division, he served as Chief of ASD's Bioastronautics Division of the Strategic and Tactical Systems Group.

His published papers have included "Evaluation of a Measurement Project" in the Personnel and Guidance Journal and "Human Factors Evaluation of a Radar Warning System" SEM-TM-64-4.

He is a member of the Society of Engineering Psychologists (APA), Human Factors Society, Soaring Society of America, Mensa, and The International Society for General Semantics.

Hobbies include flying gliders, automobile modification and teaching Process Dynamics.

BLANK PAGE

J. ALBERT SOUTHERN
SEMCP
Extensions 52314/54597

ABSTRACT

(U) MACHINES THAT COMMUNICATE AND DESIGNERS WHO DON'T

The essential difference between a system and an assortment of parts is the capability for communication that links the parts together to form an interconnected total. Whenever hardware is being designed the mechanical or electrical interconnectors are usually carefully developed and effectively tested. However, whenever a human link is inserted in the system a multitude of communication errors become possible and adequate pre-testing of the components of this subsystem is the exception rather than the rule. The aim of this presentation will be to call attention to a large and important area of system design and human factors engineering that is frequently neglected, overlooked or misunderstood -- often to the detriment of the system's effectiveness.

This area consists of the language words or symbols that are frequently made a part of the machines, tools, or components of the system, plus the language elements that make up the charts, blueprints, and manuals prepared for use by the human operators who become a critical link in the functioning of the system. Examples, illustrations, and data are cited to show that changes in the words used in man-machine systems may produce great improvements in system performance.

The viewpoint is presented that this critical area exists because it has been taken for granted for decades and is rarely recognized as an

engineering function in system development. Laughter and jokes over communications failures are common, but "Accident Investigations" of breakdowns in inter-personnel communication are rare. The identification and reduction of these breakdowns in Air Force systems improve system reliability and increase system output. It demands the best efforts of a team composed of members representing the needs of the system designer, the human factors engineer and the user. Critical points of communications links and kinks are discussed and practical techniques for resolving many of the problems are presented.

Picture yourself in your favorite automobile; the sun is shining and a few car lengths ahead of you there is a bridge. When you have this image pictured in your mind raise your hand.

We have just completed one of the most complex and fascinating experiences that is open to a human being. By emitting a pattern of sounds I succeeded in helping you form a picture in your mind and through your non-verbal response you, in turn, gave me a message that you had understood my request. In brief, we communicated.

OR DID WE???

SLIDE

1

Here is the "bridge" I was talking about!!

Did it resemble your "bridge"?? By this exercise I am trying to call your attention to the significant fact that the word "bridge" like most of the words we use does not have a meaning. In fact, <u>words never</u> have meanings! Only <u>people</u> have meanings and though a word may be the most useful communication subsystem component available, generally it must be placed in the system with caution and used with care.

The essential difference between a system and an assortment of parts is the capacity for interaction. This capacity exists only when a capability for communication links the parts together. Designers frequently produce magnificent subsystems of mechanical-electrical components -- carefully tested and highly reliable. However, as a Human Factors Engineer I wish to show you how some of the <u>links</u> become <u>kinks</u> when human communication is added as a part of the system. Three forms of communication will be examined. Note that in these the sender moves progressively further away from the receiver. In this paper the original sender is always assumed to be the designer of the system.

Our three classes of communication are:

I. PERSON-TO-PERSON
(Verbal and Non-verbal)

SLIDE 2

II. MACHINE-MAN INTERFACE
(Labels and Legends)

III. DESIGNER'S PUBLISHED DATA
(Technical Manuals)

The goal of this paper is to focus attention on some of the design hazards that produce communication kinks and what can we do to reduce them. For a glimpse at some of the underlying characteristics of person-to-person communication let's return to our BRIDGES and see if we can explore some of the things that enter into person-to-person communication.

Glance backward at your feelings when I showed you my bridge. How many of you felt a little resentful at being "tricked"? A little amused that I was "different"? A little curious about my unusual choice or why I would use such an example? This time I am trying to explore with you the fact that the meaning of words exist in people and all people have feelings as well as thoughts. Words or symbols can stir those feelings and to fail to recognize this attribute of our communication subsystem's raw material is to ignore the noise you can introduce into your system along with your signal. Effectively used, these are powerful resonances and harmonics in your listener that can reinforce and strengthen your message.

The ability of symbols to evoke feelings has long been known and used by leaders of men. The well-chosen word has helped shape history. Can you listen to these words without feelings?

"Damn the torpedoes -- Full speed ahead!"

"Your income tax will help pay for it."

"Touchdown! Army is leading 20 to 7!"

Words are one variety of symbols and no matter how or by whom they are generated they remain merely sound waves in air or marks on paper until someone responds meaningfully to them. The meaning you attribute to my words will always be a mixture of the logical meanings that some group has agreed they can mean plus the emotional feeling that, through your own previous experience, has become associated with this symbol. If I choose a word inappropriate for you the logical meaning will be lost in the noise generated by your feelings. If your feelings have reached a sufficiently high level of intensity it ceases to matter what words I choose.

During person-to-person communication the human mind generally operates in a manner resembling a three dimensional sound motion picture, to which are added the sensations of taste, smell and touch. Since most of us have somewhat similar capabilities for sensing our environment we tend to experience the same event in a similar manner when we are in the same place at the same time. When we are separated by time or space we still have a storehouse of memory units that may be selected for re-runs. It thus becomes possible to trigger these re-runs within our minds or the minds of others by using visual or auditory symbols. These memory patterns or symbol concepts vary in character from very specific items such as "the Statue of Liberty" to very abstract ideas such as "American Democracy". They vary from individual to individual according to the circumstances under which they were learned. The thought and feelings stirred within you by the word ROCK will be as distinctive as your fingerprint, as unique as your own life experiences.

Communication is made possible by our ability to use visual or auditory symbols to elicit recall within an individual. It is important to note that the would-be communicator who is preparing the message always omits part of the total material available to him and whatever the medium of transmission he chooses, he is forced to further reduce the stimulus transmitted. The choice of the mode of transmission limits the message that can be sent. In a face-to-face situation we frequently are unaware of some of the cues that we are receiving. A friend of mine is a psychiatrist. He tells the story of asking a patient if he ever heard voices without knowing where they came from. "Oh yes," replied the man, "every time I answer the telephone." The receiver of the 'message' habitually elaborates the stimulus material he receives and attempts to arrange it as if he were experiencing the original situation. Thus a strange kind of arithmetic operates here in that a speaker always sends less than the total yet the hearer always 'receives' more than was sent.

Therefore, whenever person-to-person communication is made a part of a system, the chances for error or reduced reliability are multiplied by the number of such links that are created and the volume of data that must be transmitted through each of the links. (3)

SLIDE

3

$$Ep = \frac{f V(TL)}{100}$$

Ep = Error probability

V = Volume of data

T = Number of Transmissions

L = Links (person-to-person)

The designer of a system may be aware of these hazards and sometimes one of his recourses can be to remove one or more of the human links from the system and allow the machine to transmit the message. We are looking now at our second classification of communication: "The Machine-Man Interface." In this instance the machine has been given the task of sending and the human operator or observer is the receiver.

Come with me aboard one of our B-58 bombers as this mighty ship with four powerful engines roaring leaves the runway and climbs into the atmosphere above the earth. The pilot is watching for other traffic, when suddenly a feminine voice, speaking decisively but without panic tells him "FIRE in engine three -- FIRE in engine three." He glances rapidly at the instruments before him and a few moments later he has corrected a condition which if left uncorrected could have led to a fatal crash. A mechanized Voice Warning System has communicated a danger signal rapidly and clearly to the human operator.

Many jokes have been made about the Voice Warming System now in use on some Air Force and Navy planes ....

SLIDE

4

but some of these machines have communicated so effectively with the pilot that some of these pilots will frankly admit they feel they owe their lives to a communicating machine.

Not all communicating machines are new. One that may have been used by your grandmother and that operated quite effectively is pictured here ....

SLIDE

5

This machine was simple in concept yet capable of saying, "The water is hot!" in an almost unmistakable language. Or at least in a language so simple that a child could understand it. Far more familiar to most of us is another household device probably linked in your mind with communicating person-to-person ....

SLIDE

6

but have you considered that it has its own machine language it has taught you? Lift the receiver and listen! Can you hear it saying, "I am connected and waiting for you to tell me whose telephone you wish me to try to ring?" It probably did this in about one tenth the time that it took me to say this in words.

That these are actual communicating machines cannot be denied. They generate and transmit a meaningful message just as surely as any human operator who might be observed practicing his 'art' of communicating.

SLIDE

7

Machines DO communicate!! But sometimes they fail miserably. Let's examine some occasions when the machine stuttered instead of transmitting the message the system designer wished to have someone receive.

The foregoing examples have leaned heavily on the use of the auditory channel and it is widely known that "if you write it down there is less chance for error." Let's look at some examples: Here is a written message that was placed in one of our Air Force vehicles.

SLIDE

8

EMERGENCY OPERATION
RESCUE HOIST

PUSH & HOLD MANUAL OVERRIDE ON 4-WAY VALVE IN DIRECTION DESIRED AND OPERATE MANUAL SPEED CONTROL.

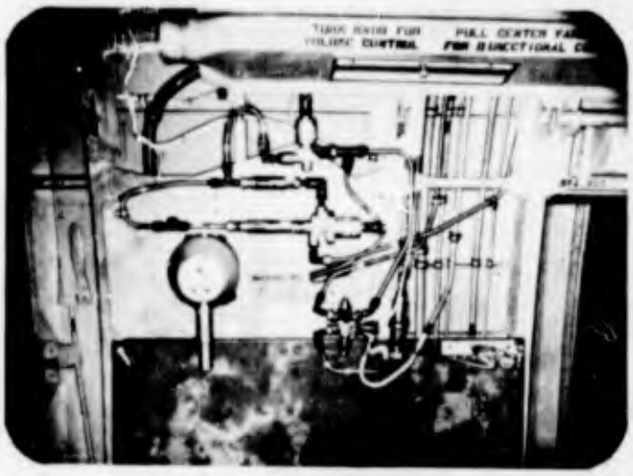
— CAUTION —

DO NOT FULLY EXTEND OR RETRACT CABLE DURING EMERGENCY OPERATION.

If this seems less than completely clear when taken out of context perhaps you would like to see where it was placed.

SLIDE

9



"But," you may say, "these are confusing because they are technical problems." So let's look at an example chosen by Dr. Alphonse Chapanis (2) from the hallway of a public building in a large Eastern city.

SLIDE

10

Here is a closeup of the instructions.

SLIDE 11

Now try to imagine your feelings if you walked up one flight or down two and found the same sign waiting for you on each of the other floors! Eventually you might guess what the system designer intended, but could have said more clearly as follows:

SLIDE 12

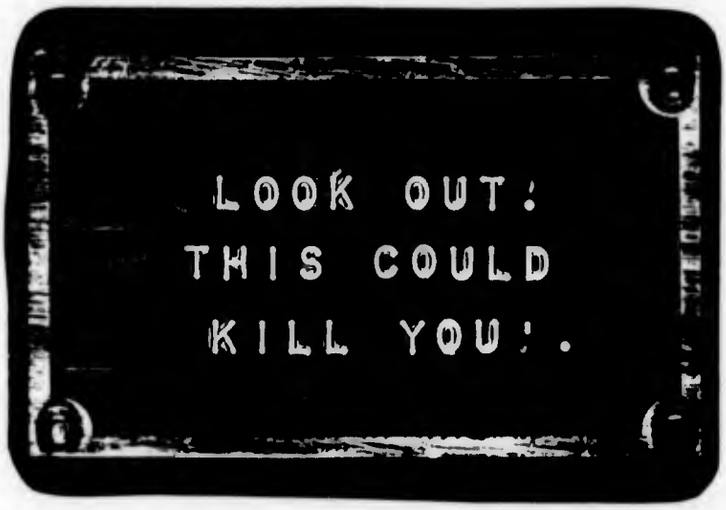
An outstanding advantage of well prepared machine-man communications is that they are at the place they are needed when they are needed. Again I am indebted to Dr. Chapenis for this example where the designer included a warning in the instruction manual which said:

SLIDE 13

"WARNING: The batteries in the AN/MSQ-55 could be a lethal source of electrical power under certain conditions."

But, how does this compare, in your judgment, with the one prepared by the unknown serviceman who translated this and placed a sending unit directly on the equipment?

SLIDE 14



Doesn't this communicate more clearly the basic message?

Our third category of designer-to-operator communication methods is the charts, drawings and printed text that make up the Technical Manuals for the system. These documents are certainly the heart and backbone of the system and its subsystems -- that they sometimes fail to be dancing feet, forever ready to respond to system changes is a weakness that is almost inevitable. Thus the need for the best possible original is apparent. To prepare a document that will be used both as a textbook for beginners and a reference library for a skilled employee is a difficult task. The thoroughness and detail that must be included for the system description frequently makes the locating and understanding of a particular reference item difficult or impossible.

Another problem of any technical document is the jargon that becomes the language of any specialized field. Such a specialized language expedites the communication process within the trained group but poses a problem for the student who has not yet mastered the language. It also handicaps personnel in related areas who may encounter the document or the equipment only on infrequent occasions. The communication process is blocked if both sender and receiver are not familiar with the same jargon and the identical language will not be understood equally well by different classes of readers. Such a misunderstanding by different experience groups lies in the story of the mother who was monitoring the intercom connected to the children's playroom. Her 6-year old daughter was playing with a neighbor's 7-year old son. The young man was heard to ask. "Do you know how to make love?" "Oh yes," replied the daughter, "come here and I'll show you." The puzzled parent reached the playroom in time to see her daughter, chalk in hand, write L-O-V-E on the playroom blackboard.

In addition to problems of jargon, the layout, indexing, diagrams, photographs and drawings may add to or detract from the comprehensibility of the manual. The Aerospace Medical Laboratory selected a list of desirable features and organized a checklist which was applied to a group of representative Air Force Technical Manuals as a means of surveying comprehensibility features of such manuals. (4) In this study, the investigator says: "The results indicate that each manual was prepared with deficiencies in number sufficient to bring about a serious reduction in its comprehensibility. On an overall basis, some differentiation between manuals was shown, but it was only between levels of mediocrity. For the most part, the manuals received correspondingly low scores on the specific aspects of design reflected by different sections of the checklist."

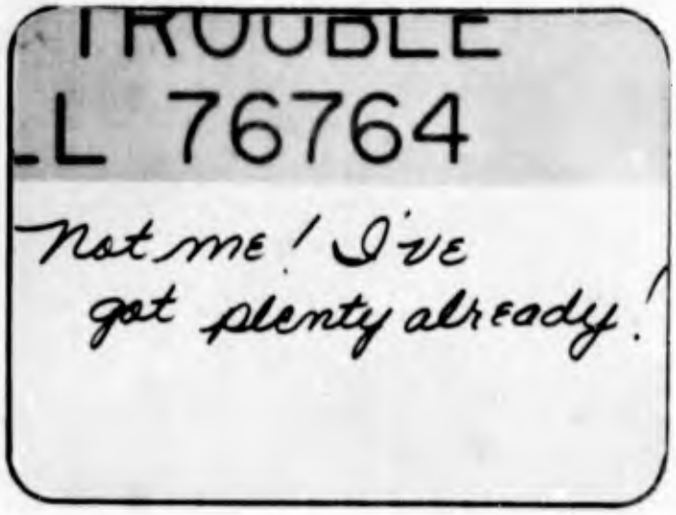
We have examined three categories of designer-to-user communication methods and a few of the ways in which these fail to maintain the reliability that is built into the hardware of the system. The tremendous

A possible disadvantage of the short message is the possibilities of its being misunderstood. In my final example of machine-man messages the trend to brevity was overdone and left the message ambiguous. On an electrical panel at Wright-Patterson AFB the instructions for emergency service read:

SLIDE 15

The misunderstanding was revealed by a handwritten message just below it.

SLIDE 16



importance of this problem is indicated in the findings of a survey of human factors deficiencies of eight missile systems.(1) In this survey almost 1/3 of the items (29%) cited as inadequate were communications deficiencies.

SLIDE

17

Can we afford reliability-reducing errors like these in our systems? If we are unwilling to accept deficiencies of this magnitude what can we do about it?

I recommend a system approach be applied to this problem and that this frequently neglected area can be brought up to a somewhat comparable status with the remainder of the system. No magic is needed -- just hard work and sufficient funds and effort to apply the approach so familiar to hardware designers but often overlooked where human links are made a part of the system.

SLIDE

18

1. ANALYSIS
2. DESIGN
3. TEST
4. REVISE & RETEST

The first step: Analysis should always include both a system analysis and a Communication Procedures Analysis. The first one defines the goal of the system or the subsystem and establishes a context for the communication analysis.(5) Some of the questions to be answered by the Communication Procedures Analysis are: (a) What messages will be generated and by whom? (b) Where should they go? (c) What is the optimum routing for them and what modes of transmission are available? (d) What are the characteristics of the receivers?

The second step is the design of a communication network. What alternatives are available? How much redundancy is needed? Is feedback necessary? How is it achieved?

Step three should be a thorough testing of the design. The test should be realistic and it should give a positive answer to the question: "How well does this subsystem work and where are the areas that need improvement?" This should be completed early enough to permit the next step.

Step four is: Re-design and re-test as necessary. Zero Defects can be a goal for the communication subsystem but it should not be an expectation of even the finest first effort.

The goal of my communication effort has been to focus attention on a problem. Perhaps I have risked oversimplifying it. If so, I suggest that you consider that all systems are created by men and for the use of men. It is the communication capability that makes it a system -- that links it together internally and relates it to the men who profit from its existence. The communication links should be designed with caution and tested with care. We are presently tolerating slop in this area that would not be tolerated in the mechanical portion of our systems. This is a difficult, complex and challenging area. This was what I meant to transmit. What did you receive?

REFERENCES

1. Chaikin, Gerald., Survey of Recommendations by Human Factors Engineering Evaluators During Army Missile System Development Programs, Report RC-S-63-1, Redstone Arsenal, Alabama., Aug 1963.

2. Chapanis, Alphonse., Words, Words, Words, Human Factors Journal, 1965, Volume 7, page 1-17.

3. Potter, Norman R., Unpublished communication

4. Ross, Donald A., Comprehensibility Evaluation of Technical Manuals, WADCTN-59-442, July 1959.

5. Rees, David W., Guide to Design of Air Force Check-List Publications, WADC TR-59-758, December 1959.

DRAWINGS BY:

Robert D. Fletcher and Richard J. Kennedy

(U) PHOTOGRAPHIC DISPLAY SYSTEMS - TYPICAL USAF APPLICATIONS

by

LEON MCDOWELL

Rome Air Development Center
Research and Technology Division
Griffiss AFB, New York

Leon McDowell

BIOGRAPHICAL SKETCH

Mr. Leon McDowell was born in Portsmouth, Virginia on January 30, 1934. He received his B.S. degree in Electrical Engineering in 1960 from Howard University of Washington, D.C. and has also taken courses at Syracuse University.

Mr. McDowell was employed at the Rome Air Development Center as an electronic engineer in 1960. Since that time he has been assigned to the Display Systems area. At present, he is employed in the Projection Techniques Section of the Display Techniques Branch and is engaged in the development of projection display systems and techniques for operational applications.

BLANK PAGE

ABSTRACT

PHOTOGRAPHIC DISPLAY SYSTEMS - Typical USAF Applications

by Leon McDowell

This paper presents a comparative study of the rapid process photographic film equipments that are currently being used in United States Air Force Command and Control Systems, and it includes a review of the technological and operational aspects of photographic film systems. Rapid process film systems are presently utilized at Hq NORAD (425L), Hq SAC (465L), and Hq USAF (473L). These systems provide the using commands with a large screen, computer generated display capability to assist the decision makers in executing their command and control functions, and these specific film systems are included in the scope of this study. The paper will also concern itself with the effectiveness of photographic techniques in providing the display capability in a command and control environment.

Photographic recording and projection may be subdivided into four functional areas: a) digital to image conversion, b) camera optics, c) film and processor, and d) projection optics. A discussion is made with respect to the specified functional areas of the varied design approaches for these equipments.

Additional data are included in the form of a tabulated summary of pertinent performance parameters such as resolution, symbol brightness and contrast, and symbol size and stroke-width, etc.

This paper also addresses itself to the maintenance requirements and the annual operating cost of the systems under study. The primary areas

of concern in this portion of the paper will be the amount of preventive maintenance time required per day to maintain an acceptable level of performance and the amount of chemicals and film consumed per year.

The above mentioned areas are considered to be the most undesirable features of rapid processing film equipments, and much of the present development effort is designed towards correcting and/or minimizing them.

INTRODUCTION

The group display concept of man-machine information transfer within command and control centers has generated the requirement for displays which are larger and have greater capability than the direct view cathode ray tube. These additional requirements have resulted in the development of a variety of projection displays designed to provide bright, large-screen, computer-generated displays to assist the decision makers in executing their command and control functions. The three techniques presently utilized in large screen applications are the oil film projector, the scribe projector, and the photographic film projector. The photographic film projector is the most widely used in command and control applications The major command headquarters which utilize the photographic technique to provide their display capability are: Hq NORAD (425L), Hq SAC (465L), and Hq USAF (473L).

This paper reviews the technological and operational aspects of current photographic display systems. It includes a review of the functional areas of photographic display systems, a discussion of the 465L and 473L display concepts, and a description of the 425L camera processor projector. Maintenance requirements, cost considerations, and performance characteristics are also included to aid in acquiring an appreciation for the effectiveness of the technique in providing the required capability.

The command and control systems in the Air Force may be divided into two general categories: 1) Planning and Management, and 2) Surveillance and Detection. Both categories exercise certain command and control functions in executing the broad concepts of threat evaluation and battle management. However, planning and management systems usually deal with the status of military resources on a world-wide basis and do not handle individual weapon systems in a specific strategic or tactical manner as does surveillance and detection systems (1). A representative number of Air Force systems are categorized in Figure 2. The subdivisions in Category 1 suggest the relative extent to which the command functions are exercised. The systems in Subdivision B are primarily planning tools as opposed to the systems in Subdivision A which are major command and control systems. These general categories will to a large extent determine many of the system requirements and constraints with which the display system must comply, such as the size of the data base, the maximum quantity of data to be presented per display, the multiplicity of display formats, the response time of the display, and the time-rate of changing information. Planning and management systems operate out of a relatively large data base and require large quantities of static-type data to be presented in a variety of formats. Surveillance and detection systems, whose primary functions are track-following and weapons control, usually present fixed-format, situation type displays where certain elements of information in the display field are periodically changing. It is the author's belief that there is no one technique that can be employed efficiently in all system applications, but rather a particular technique is generally best suited in a given application. The photographic technique is

believed to be best suited for the first category. It is capable of generating and presenting in color large quantities of static-type data in a few seconds. In planning and management applications, where no requirement exists to display or otherwise indicate continuous motion of certain display elements, the inherent inability of this technique to present a "dynamic" situation does not represent a serious limitation.

Although photographic display systems are considered effective in providing display capability in command and control centers, they require considerable maintenance and logistic support. The maintenance and logistic requirements, in addition to the annual operating cost, are considered to be the most undesirable features of silver halide film systems. The amount of preventive maintenance time required per day to maintain an acceptable performance level and the amount of expendable materials (chemicals and film) consumed per year are the two primary areas where improvement is desired. Development efforts are underway at RADC to minimize the effects of these undesirable features. The specific areas of development which should increase the over-all effectiveness of the photographic technique are the development of a dry process to ease the maintenance requirements and the development of a reusable recording medium to reduce the logistic support and the annual cost of expendable materials.

Functional Components of Photographic Display Systems

The Photographic Display System is comprised of five functional components which are presented in Figure 3. These areas are: digital-to-image converter, camera optics, processor, projector, and screen. A brief discussion of each component follows:

Digital-to-Image Converter - The digital-to-image converter accepts the coded digital data from the data processor and converts it into visual images suitable for photographic recording. The data provided to the display system are in coded digital form in order to identify which element of information is to be displayed and where in the display field it is to be displayed. This component provides the input/output control, the digital-to-analog conversion, symbol generation, and the electrical-to-image conversion.

The input/output control provides an orderly transfer of data between the data processor and the display system. The information transfer in the majority of applications is asynchronous (variable rate) and normally occurs in either one of the following two modes: "word-demand" and "block-of-words." In the "word-demand" mode, computer words are transferred only at the request of the display system. Thus the rate at which the display can accept a digital word is not exceeded. This mode of data transfer is used with a display system which has no local storage capability. The computer does not necessarily have to be unavailable for

other tasks during the total time it takes to transmit a complete display message. A more complicated program could make efficient use of the time between word transfers. It becomes a matter of the extent of complexity which can be tolerated in the programming effort versus the value of the total time required for display generation. In the "block-of-words" mode, all the computer words are transferred into the display memory before any image generation is begun. Therefore, the entire display message is transferred in comparatively little time, enabling the computer to return to its other tasks. This mode of data transfer is used with a display system with local storage capability.

The digital-to-analog conversion converts the digital code which represents the position of the information element to an analog voltage which will determine where the deflection circuits will position the electron beam of the cathode ray tube (CRT). The position data for a particular element is given in reference to an X-Y coordinate system which in many cases has its origin in the lower left hand corner. This origin is chosen in order to have all positive coordinates which simplifies the hardware design and the programming function. Present systems are designed to receive up to 10 bits in "X" and 10 bits in "Y", and position at random the CRT electron beam at any one of over 1,000,000 points in a rectangular position matrix.

The symbol generation function has the obvious task of receiving the digital code representing the identity of the element (character or vector) to be displayed and providing the proper analog signals required to generate the element on the CRT. The most popular techniques for generating symbols in present equipment are stroke writing, line segment, beam shaping, and the monoscope. The speed of generation is optimized with other recording parameters and the symbol write-time in the present systems extends from 100 to 1000 u sec. This time includes the deflection and settling time for the electron beam (50 u sec).

The CRT is the optical transducer which converts the electrical signals to visual images suitable for photographic recording. CRT's have been designed to produce images with adequate stability, quality, spectral content, and intensity. The types generally used are the 5" or 7" flat face, high resolution CRT's with electromagnetic deflection and focus, and using P-11 phosphor as the light emitter. Display non-linearities contributed by the CRT and the driving equipment are generally less than 0.5% of display width, and electron spot sizes range from 2 to 4 mils.

Camera Optics - The camera optics transfer the image of the CRT to the recording medium with minimum distortion and aberration. The input optics are also designed to minimize the effects of off-axis illuminance and vignetting.

Color is one of the most popular coding techniques used to present different categories of information. The use of color film is prohibited

because of the excessive development time required to produce the final image. As a result, the color additive approach is utilized in many applications to produce a multi-color capability where a rapid response is required. To meet the multi-color requirement, the camera has to produce three precisely positioned images on the film which correspond to the three primary color channels in the projector. By sequential presentation of data on the CRT, in conjunction with the appropriate shuttering, a multi-color image can be recorded without movement of the film. Camera optics in present systems are capable of the following performance characteristics, depending on the specific design:

Resolving Power - greater than 120 optical lines/mm

Speed Range - f/6 to f/12

Processor - The processor converts the latent image on the exposed film to a positive image suitable for projection. The processing performed by contemporary systems is a wet process involving the movement of chemicals. In conventional processors, the most common recording medium used is the silver halide light sensitive emulsion. The sensitivity of silver halide film is such that it can be exposed with a standard cathode ray tube. It can be observed from the spectral response graph of Figure 4 that the spectral response is relatively flat out to 500 millimicrons, making this emulsion ideal for exposure to P-11 phosphor which peaks at approximately 460 millimicrons. The demand for rapid response for computer displays has been met by the development of a rapid film processing technique. This process consists of the use of special films with hardened emulsions which make them suitable for rapid processing at elevated temperatures without appreciable loss in image quality. A general rule of thumb adopted by the film industry allows the development time to be cut in half with each 15° rise in chemical temperature. Most film process data is given at a temperature of 68° F and a development time of four minutes. Therefore, when the temperature is raised to approximately 140° F under rapid process conditions, the development time is reduced to approximately 10 to 15 seconds. Processing of silver halide film is divided into three categories: conventional, partial reversal, and full reversal (2). A negative image is produced in conventional processing and is not suitable for projection in a color additive process because the information areas are dark and the background areas are clear. The steps involved in conventional processing are:

Develop - Reduces the silver halide crystals exposed to the CRT image to metallic silver and leaves the others unreduced. The light energy absorbed by the exposed crystals increases their sensitivity to the developer and consequently they are reduced before the unexposed crystals.

Fix - Changes the unexposed silver halide into a soluble salt in order that it can be washed from the base material.

Clear - Removes the soluble silver halide products of the fixing process and leaves the unexposed parts clear.

Reversal processing produces a direct positive image suitable for projection. The steps involved in reverse processing are:

Develop - Same as in conventional processing.

Bleach - Converts the metallic silver into a water soluble compound in order that it can be washed from the base material.

Clear - Removes the soluble products of the bleach process and leaves the exposed areas of the film clear.

The above three steps constitute a partial reversal process. The final image is the clear information areas against the unexposed silver halide background. To obtain a full reversal process, a fourth step is added:

Rinse - Converts to metallic silver the silver halide crystals in the background that are re-exposed by the radiation from the projection lamp.

Under zero ambient illumination conditions, the partial reversal processes are capable of producing image contrast of better than 50 to 1, and full reversal processes are capable of producing image contrast of better than 100 to 1. In applications where the surface area to be illuminated is greater than 150 sq ft, the increased contrast afforded by the full reversal process can compensate for the reduced incident illumination level and consequently preserves the display legibility.

The sensitometric curve of Fig. 4 indicates the maximum contrast that the fully reversed silver halide emulsion is capable of producing. The indicated density range of 2.2 to 0.2 is equivalent to a light transmission of 0.63% in the dark areas of the film and a light transmission of 63% in clear areas. The spatial frequency response curve, also of Figure 4, indicates that the maximum brightness and subsequent contrast degrades rapidly as the resolution (line pairs/mm) of the image to be recorded and processed exceeds 50 line pairs/mm. Thus, it is significant to note that although the inherent resolving power of silver halide emulsion is rated in excess of 100 line pairs/mm, the maximum response of the emulsion is realized at some lesser value of resolution.

When utilizing a negative film processor, it is necessary to produce a positive image suitable for projection from the negative. This is done by transferring the image to Kalvar film. Image formation in Kalvar photographic materials is based upon light scattering rather than light absorption as is typical of silver halide materials (3). The pertinent characteristics of Kalvar film are shown in Figure 5.

The absolute sensitivity of Kalvar is such that it cannot be exposed with a standard CRT. It is sensitive to near UV (response peak at 3800A) and is developed and fixed by heat alone. For a photographic medium depending upon light scattering, the projection density depends upon the speed of the projection lens. This may also be interpreted to mean the contrast of the projected image is inversely related to the speed of the projector lens. It can also be observed from Figure 5 that an f/3.5 projection aperture will produce excellent contrast (density range of 2.0-0.2) and is an excellent compromise between screen brightness and image contrast. Kalvar film is capable of presenting contrast ranges of 100 to 1, and its inherent resolving power exceeds 140 optical lines per mm. Since Kalvar film does not absorb as much light as does silver halide, air cooling at the projection station is usually sufficient.

An evaluation of some of the pertinent physical characteristics of present day film processors indicates that many of the effects of problems experienced with early processors have been minimized. The problems of fluid leaks, chemical/material compatibility, and contamination were the most serious ones to plague the film processor industry. Characteristics of present day processors are discussed in the following paragraphs:

Chemical Storage and Transfer - The primary consideration for the selection of materials for the construction of chemical tanks is the hot caustic solutions which react with metals producing corrosion. Use of teflon, polyethylene, or any other chemically inert material as liners in stainless steel tanks, or use of plastic containers has proved to be very effective against corrosion (4).

The fluid systems in the referenced "L" systems utilize the negative pressure concept. This concept reduces the possibility of fluids escaping from the process heads into the atmosphere in the event of a leak, and is applicable in both continuous and pulse flow systems. Transfer of caustic liquids by rotary pumps has been a serious problem because of the leakage of fluids around the rotary seals of the pumps. This problem has been solved in the continuous flow system by utilizing a peristaltic pump. The chemicals do not come in physical contact with this pump and, therefore, it is not subject to either leaks or corrosion. In the pulse flow system, atmospheric pressure is the prime fluid mover, working into a partial vacuum in the waste tank. The partial vacuum is produced by pumping air rather than chemicals. This approach is also not subject to either pump corrosion or pump leakage.

Film - To preserve emulsion integrity, the emulsion is allowed to come in contact with a minimum number of mechanical surfaces. "Wetting" agents are used in the processing fluids to moisten the emulsion,

allowing a more uniform chemical action at the emulsion/chemical interface. The emulsion is supplied on a mylar base to give the film adequate mechanical strength.

Projector - The projector provides the required incident illumination on the screen and transfers the image at the film gate to the screen with the required magnification. The primary color fields used in the color additive process are produced by the filtering of white light. The worst case misregistration of colors produced by the color additive process in present systems is approximately 40 to 50 per cent of the stroke width. Based on recent studies at RADC (5), a color misregistration of 67% is considered the upper limit of acceptability. Xenon light sources are used in large scale projection equipment because of their color temperature, high efficiency, and intensity per unit area of the source. Projector efficiency (ratio of total lumens emitted from source to total lumens incident on screen) of current systems is approximately two to fifteen per cent.

Screen - The screen collects the light from the projector and converts it into a visual image. Screens have a wide variety of distribution characteristics and are selected to meet the requirements of a particular screen/audience configuration as well as environmental considerations such as ambient illumination and physical space.

Application of Photographic Techniques

The photographic techniques as applied in the 465L, 473L, and 425L Systems will be reviewed in the following sections. The concepts of 465L and 473L will be discussed and the 425L Camera Processor Projector (CPP) equipment will be described.

465L - The 465L Concept as presented in Figure 6 is a two-step process utilizing the conventional processing of 70mm silver halide film and the contact exposure and heat development of Kalvar film. The character generator is a charactron tube which forms characters via beam shaping. The camera optics provide for the exposure of three image areas per film frame. Each area represents one of the three primary colors. Sequential presentation of data on the charactron tube and the appropriate shuttering determine the particular color of the projected display. The conventional negative is obtained with silver halide film produced by the Ansco Company. After the negative processing is complete, it is then moved to the contact print station. The positive film (Kalvar) is exposed to a one millisecond UV flash lamp and then moved to a heat platen (240°F) for developing for approximately $\frac{1}{2}$ sec. It is then moved to the project station. The projector produces the three primary color beams by dichroic separation of white light from a Xenon source. These three color beams pass through their

respective image areas and are superimposed at the screen. This concept produces a full color 8' X 8' display (six colors plus white). The background geographic reference data is projected from a separate reference projector.

473L Concept - The 473L Multicolor Concept as presented in Figure 7 is a one-step process utilizing the partial reversal processing of 70 mm silver halide film. The 473L System has a small multi-color display within the console with a screen size of 1' X 1'. The system also will have two large scale displays which will be silver halide devices and use film chips rather than roll film. The large screens will be 6' X 6'. The performance data included in this paper refers to the 473L Console Display. The character generation technique is the monoscope which forms characters by the video modulation of a sub-raster. Each character is formed individually with a separate modulated raster. The camera optics provide for the exposure of four image areas per film frame. Each image area represents a specified color. Sequential presentation of data on the monoscope and the appropriate shuttering determine the particular color of the projected display. The partially reversed negative is produced from Kodak So-198. After the processing is complete, the film is moved to the project station where it is projected on to a 1' X 1' screen. The four color channels which supply the required colored light to each image area are produced by the filtering of white light. This concept is a four-color, non-additive approach. The colors are red, yellow, green, and white. Geographic background data is entered from another projector.

425L CPP - The 425L CPP as presented in Figure 8 is essentially a camera, rapid processor, and multi-color film projector in which these elements are combined to achieve a seven color large screen display. The primary colors are achieved by introducing color filters in the projector. The secondary colors are achieved by optical mixing of the primary colors at the screen. The basic display unit (BDU) presents a visual image on the CRT for processing by the CPP.

Theory of Camera Operation - The three-image frames of the film which correspond to the three primary color channels are exposed at the camera station to symbology appearing on the CRT. Exposures are under the control of BDU generated shutter opening signals: The sequence of operation is as follows:

The BDU transmits shutter-opening signals to the CPP. The shutter opening signals correspond to the color to be displayed. The CPP responds by adjusting its shutters to the required configuration. There are three shutters which correspond to the three image areas.

The actuation of single shutters permits the exposure of primary "color" frames; the simultaneous actuation of multiple shutters causes exposures in more than one frame yielding secondary colors at the projection screen.

Camera Optical System - The camera utilized in transferring the CRT image on the film consists of a 3-lens system which positions the CRT image on three adjacent areas on the film. Certain key design parameters are:

Lens	6" focal length f/6.8
Minification	3.125
CRT Image Size	2.25" X 3.0" (4" diagonal)
Film Image Size	0.72" X 0.96" (1.28" diagonal)

Processor - The processor of the CPP is a full reversal type with three continuous flow fluid systems operating under negative pressure. After exposure, the film is moved to the three process stations sequentially (develop, bleach, and rinse). The film (Kodak So-142) remains at each station for approximately 3.5 seconds. To meet the 10-second update requirement, the final exposure and development are accomplished at the project station.

Projector - The three-frame exposure, after processing, is projected as an additive primary color combination to a full color display with minimum color misregistration. The projector consists of a 6.5 KW Xenon light source (operated at 5.0 KW input) and three source condensers (f/1) that illuminate three film gates. Three matched projection lenses (f/3.5) transfer the image at each film gate to the screen with the required magnification (200X). Color filters in each film channel provide the appropriate colors. The 425L CPP produces a projected image of 12' X 16' at a throw distance of approximately 25'.

An artist's view of the NORAD combat operation center and the screen-audience configuration is presented in Figure 9. Figure 10 is a picture of the display equipment (camera-processor-projector and the basic display unit) located behind the screen, and Figure 11 is a picture of the display screen as seen from the operator's position. The actual display is a locally generated test pattern designed to test the functions in the camera-processor-projector (CPP) and the basic display unit (BDU) independent of any computer inputs.

Performance Characteristics

The photographic technique is capable of producing displays of exceptional quality. Figure 12 is a tabulated summary of the pertinent performance parameters of the three systems being reviewed. The values of

the individual parameters do not necessarily represent a maximum, but rather the list of the parameters taken together represents a set of performance characteristics which can exist simultaneously. Although the majority of the parameters are self-explanatory, the fall-off refers to variations in illumination intensity from screen center to edge. In a properly aligned optical system, the center will be brighter as a result of the $\cos^4 \theta$ illumination law and vignetting within the condenser and projector. The resolution figure is classified as system resolution and it includes the degrading effects of high temperature processing. The registration parameter refers to the registration of corresponding points in the computer-generated display and the background reference slide. The color fringing parameter describes the color misregistration in the computer generated display.

Maintenance Requirements

The major portion of the maintenance required for photographic display equipment can be traced to the fluid system which consists of chemical storage and transfer. Although "wet" film processors are becoming more reliable, they still require considerable attention, and many of the daily preventive maintenance procedures are quite lengthy. The daily procedures usually include cleaning the process heads to remove the excess sludge particles that have accumulated, checking for leaks in plumbing and connections, checking for the normal rate of fluid flow through fluid lines, checking all motors and pumps, and checking chemicals for proper operating temperature.

The above procedures, in addition to others, generally require from one to two manhours to perform. The 425L System is presently scheduled for two hours per day (24-hour day) for preventive maintenance (PM). The 465L facility at Westover AFB is presently scheduled for four hours of daily PM with at least four people performing maintenance. The projection display unit in the 465L System is called a quadrajector. There are four quadrajectors at the Westover Facility; therefore, the average amount of preventive maintenance required is one manhour per quadrajector.

The reliability of photographic systems is generally good. The latest reliability and maintainability report from the NORAD site (6) specifies that the inherent MTBF of the 425L Large Scale Display is in excess of 100 hours and the MTTR is 103 minutes. This gives an availability figure of 96%. The latest SACCS analysis summary dated Feb 66 and produced by Hq Second Air Force, Barksdale AFB specifies that the reliability percentage figure for the 465L Quadrajector is in excess of 95% (7). Although no similar figures were available on the 473L System, it is expected that with the improved fluid system, the reliability figures will exceed those stated above.

Cost Considerations

Although the annual operating cost of photographic display equipment is not excessive, any effort designed to reduce that cost even further does warrant consideration. It is determined by the quantity of expendable materials consumed per year. The expendable materials consumed per year is in turn a function of the total number of updates and the characteristics of the processing equipment. The number of updates on which the operating cost of each equipment is based is considered a nominal rate by operations personnel. Figure 13 is a tabulated summary of the cost estimates for the three systems.

425L - The estimated nominal update rate for the CPP is 70 updates/24-hr day. During exercise of stress conditions, the rate is much higher, but 70 is considered on the high side for the average rate over a year. It has a continuous flow processor and the chemicals in the supply tanks are maintained at their operating temperatures. These elevated temperatures reduce the "shelf" life of the chemicals; therefore, costs are more time dependent than update dependent. Based on a 24-hr replacement cycle, the chemical costs are estimated at $10.00 per day. The 35mm silver halide film is used at a rate of 19 inches per update. The film costs about $25.00 per 400-ft. roll. Therefore, the yearly costs are:

Chemicals	- $3650
Film	- $2660
Total Cost	- $6310

465L - The estimated update rate for one of the quadrajectors is 84/24-hr day. This is a SAC estimate based on eventual system operation. It has a continuous flow processor and the chemicals in the supply tanks are maintained at their operating temperatures. These elevated temperatures reduce the "shelf" life of the chemicals; therefore, costs are more time dependent than update dependent. Based on a 24-hr replacement cycle, the chemical costs per quadrajector per day are $15.00. The 70mm silver halide film is used at a rate of 1-ft./update. The image is transferred to Kalvar (1-ft./4 updates). Both films cost about $50 for 400 feet. Therefore, the yearly costs are:

Chemicals	- $5475
Film	
Silver Halide	- $3837
Kalvar	- $ 955

Total Cost per Quadrajector - $10,267

It is significant to note that the 465L Westover Facility has operated successfully with a 5-day chemistry change cycle. This would reduce the cost per update from $0.33 to approximately $0.27.

473L - The estimated average update rate for one of the consoles or the large screen film projector is 50 updates/day. This is an average figure over a year. It has a pulse flow processor, and only that portion of the chemicals required for processing is heated to a high temperature. The chemicals in the supply tanks are not maintained at high temperatures; therefore, they have longer "shelf" life. This means that chemical cost is more update dependent than time dependent. The estimated chemical cost is 2.8¢ per update. Six inches per update of 70mm silver halide film is used. This film costs about $50.00 for 400 feet. Therefore, the yearly costs are:

Chemicals	- $ 510
Film	- $1140
Total Cost	- $1650

Conclusion

The Photographic Display System has proved operationally effective for providing a display capability in command and control systems. It is capable of presenting a multi-color, computer-generated display consisting of alphanumerics and vectors within a 10 to 15-second time period. The resultant image possesses all the characteristics of an excellent quality display. It is bright multi-color and has high resolution and good contrast.

The Photographic Display System is a relatively reliable system. However, the reliability and/or availability figures do not accurately reflect the effort required to maintain an acceptable performance level. This is not to imply that the maintenance requirements for photographic systems far exceed those of other operational techniques, but rather it is to suggest that this is an area where improvement is desirable.

Development Objectives

Current development efforts at RADC have a two-fold objective: 1) to ease the maintenance effort, and 2) to reduce the annual cost of expendable materials with no appreciable sacrifice in performance. To achieve both objectives simultaneously would be an ideal accomplishment. However, a more realistic goal would be the accomplishment of the objectives in the order stated. The "dry" process appears very attractive for the next generation of film processors. It eliminates the fluid system which in turn should result in more simplified maintenance procedures. The development of a reusable recording medium should result in a reduction of the annual cost of expendable materials.

There are two film techniques which could conceivably be used to accomplish objective number one. They are the direct CRT exposure of Kalvar film and the development of silver halide films with a "dry" film which is imbibed with the required chemicals. Kalvar film is a commercially available product; has the resolution and contrast suitable for large scale projection displays, but its photographic sensitivity is not adequate for direct cathode ray tube exposure. CRT's and other near-UV exposure devices are being developed for this application. The dry development of silver halide has been realized (8); however, the process has not been realized for reversal processing and projection. This process, although minimizing the maintenance requirements, could conceivably result in higher operating cost. Both of the above processes are essentially dry processes although they utilize expendable materials.

Several other techniques such as photoplastic recording, thermoplastic recording, and photochromics have the potential not only for "dry" processes but also for reusability. A reuse capability would reduce by a significant amount the cost of expendable materials. Photoplastic and thermoplastic recordings utilize surface deformation as the mechanism for image retention, and form images that remain fixed until erased (9). Photochromics film is a material in which a reversible chemical action occurs when exposed to certain wavelengths. The image is transient and must be periodically regenerated to form a fixed image. These concepts are still in R&D and it is difficult at this time to predict when either or both of the initial objectives will be realized.

References

(1) Bernberg, R.E., "A Look at Future Management Data Display Technology," Information Display, Vol. 2 (Jan/Feb 1965) 14-23

(2) Larmore, L., Introduction to Photographic Principles, Prentice-Hall

(3) Kalvar Corp., The Kalvar Handbook, Technical Bulletin No. 101

(4) Bendick, M. and Scott, W., "Development of A Continuous Film Processor to Overcome Common Deficiencies," SMPTE, (1963), Vol. 72, Pg. 184

(5) Rizy, E.F.; Elias, M.F.; Snadowsky, A.M., "Misregistration in Color Additive Displays," RADC TR 64-488 (Aug 4, 1964)

(6) Burroughs Corporation, NORAD COC Equipment Monthly Reliability and Maintainability Report, dated June - Dec 1965

(7) Hq Second Air Force, Barksdale AFB, SACCS Analysis Summary (Feb 1966)

(8) Tarkington, R., "A New Kodak Processing Method for The Aerospace Age," paper presented at the 1963 conference of the SPIE.

(9) Gaynor, J. and Aftergut, S., "Photoplastic Recording," SPSE Vol 7 (1963) Pg. 209

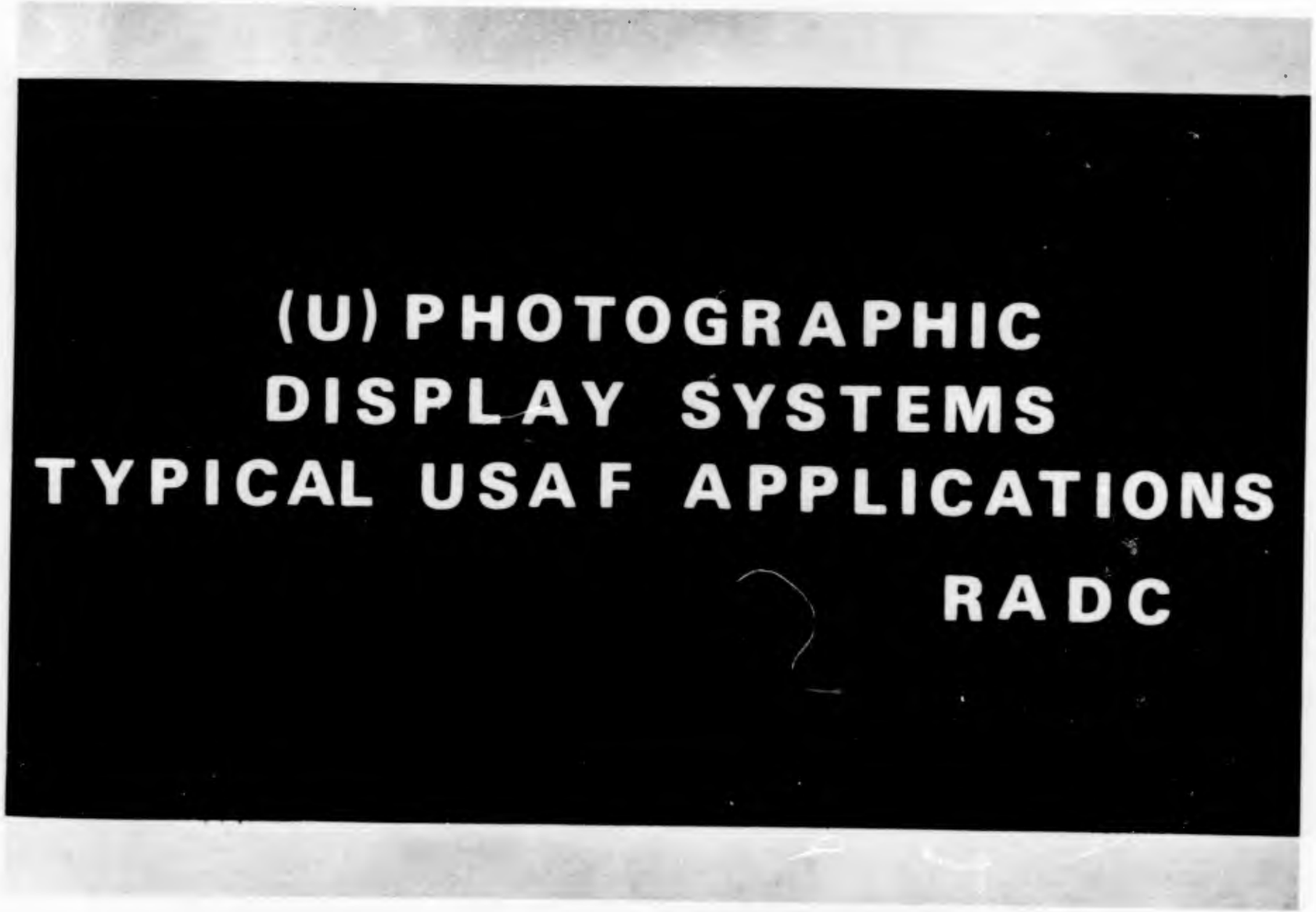


FIGURE 1 TITLE

CATEGORIES OF "L" SYSTEMS

I PLANNING AND MANAGEMENT

A) COMMAND TYPE

1. 425 L NORAD
2. 465 L SAC

B) NON- COMMAND TYPE

1. 473 L USAF
2. SACOPS

II SURVEILLANCE AND DETECTION

A) 412 L ADC

B) 416 L (AAC) ADC

FIGURE 2 CATEGORIES OF "L" SYSTEMS

FUNCTIONAL BLOCK DIAGRAM OF PHOTOGRAPHIC DISPLAY SYSTEMS

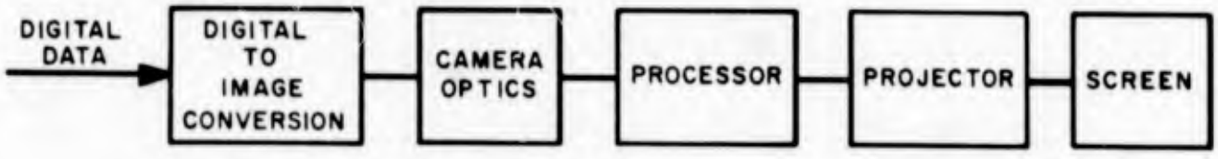


FIGURE 3 FUNCTIONAL BLOCK DIAGRAM OF PHOTOGRAPHIC DISPLAY SYSTEMS

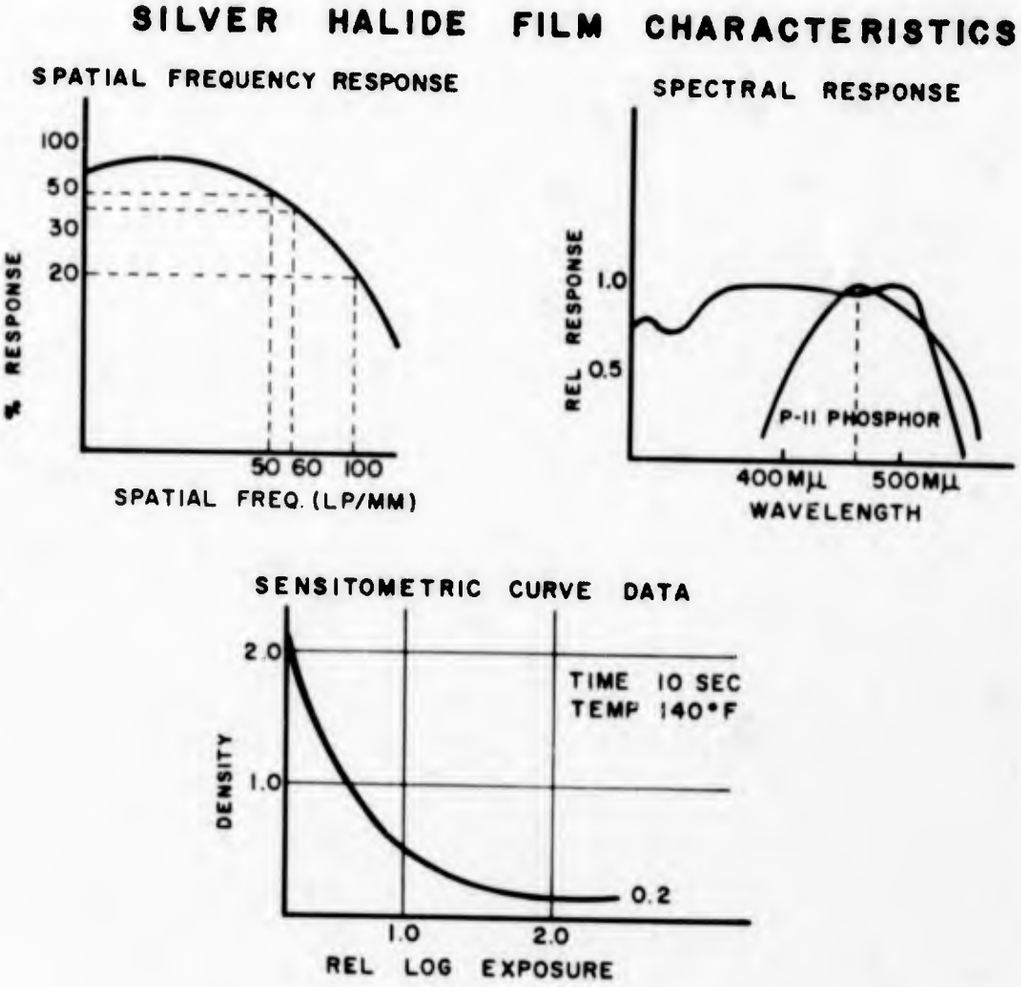


FIGURE 4 SILVER HALIDE FILM CHARACTERISTICS

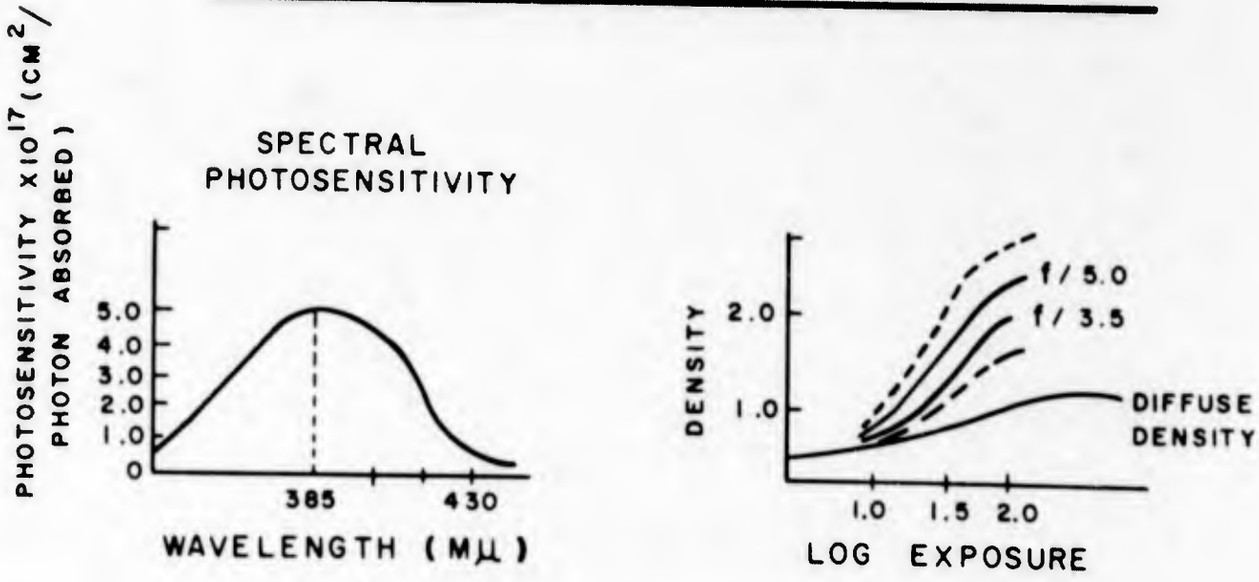


FIGURE 5 KALVAR FILM CHARACTERISTICS

465 L PROCESSOR- PROJECTOR CONCEPT

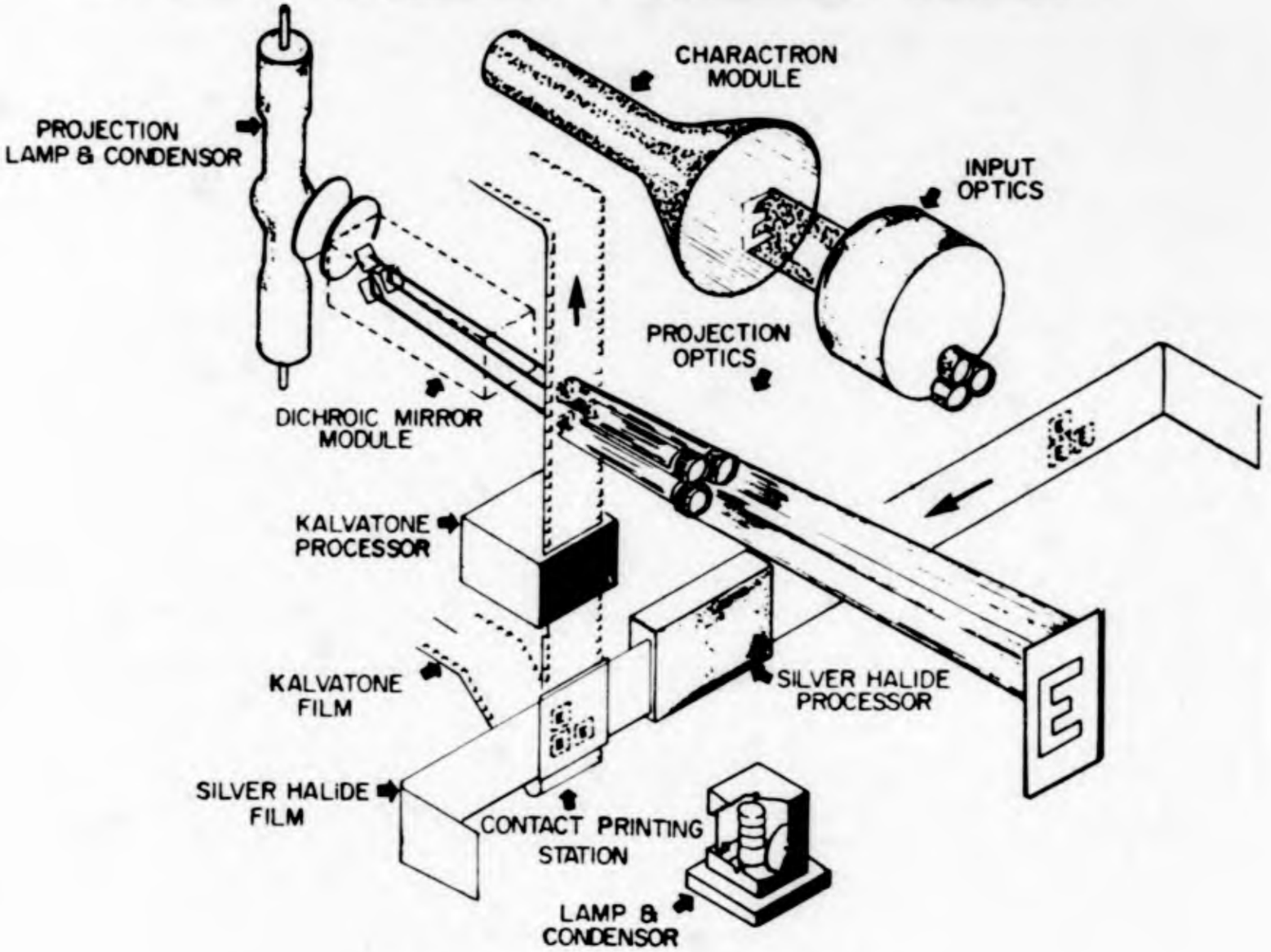


FIGURE 6 465L DISPLAY CONCEPT

473 L PROCESSOR - PROJECTOR CONCEPT

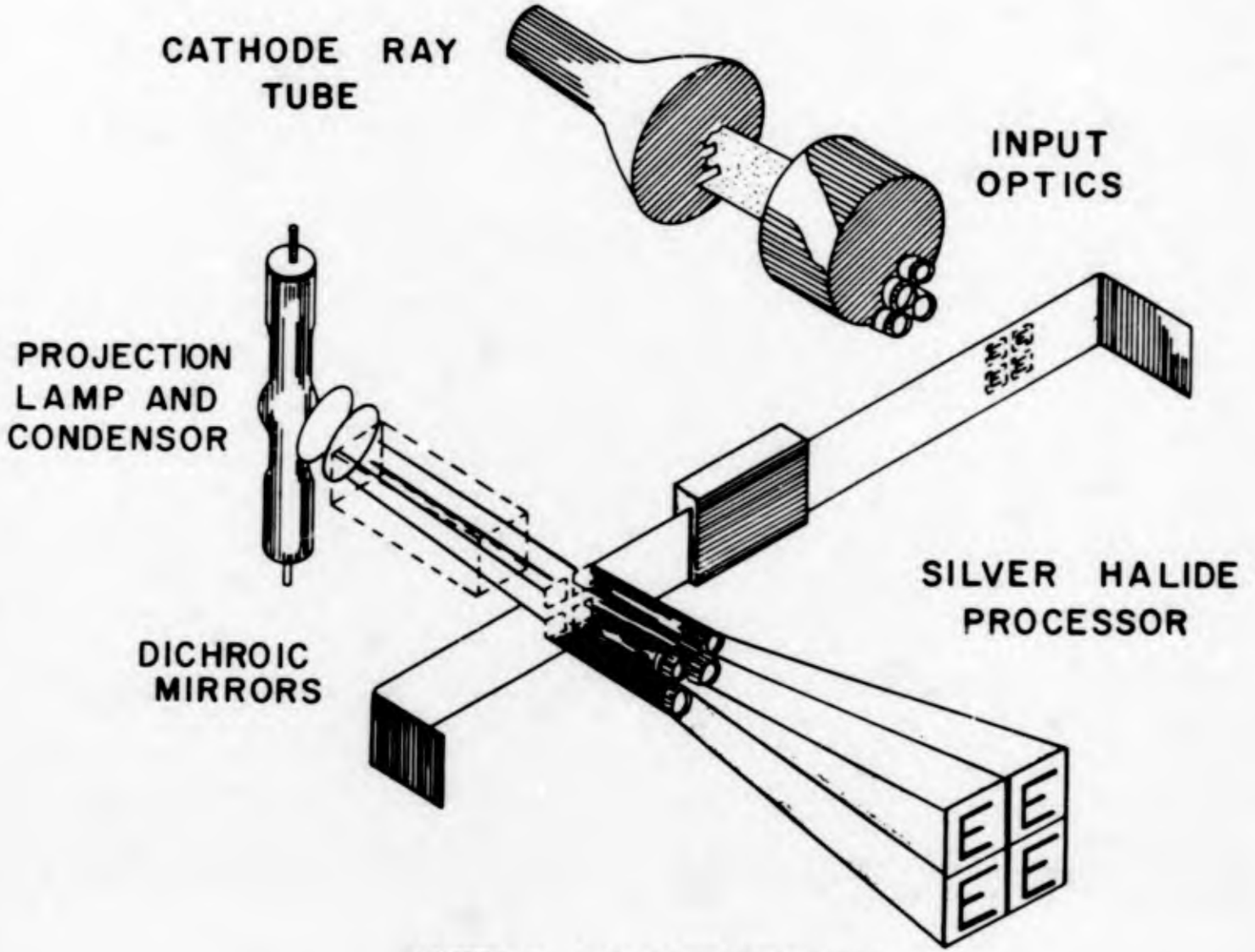


FIGURE 7 473L DISPLAY CONCEPT

SCHEMATIC OF CAMERA PROCESSOR PROJECTOR UNIT 425 L

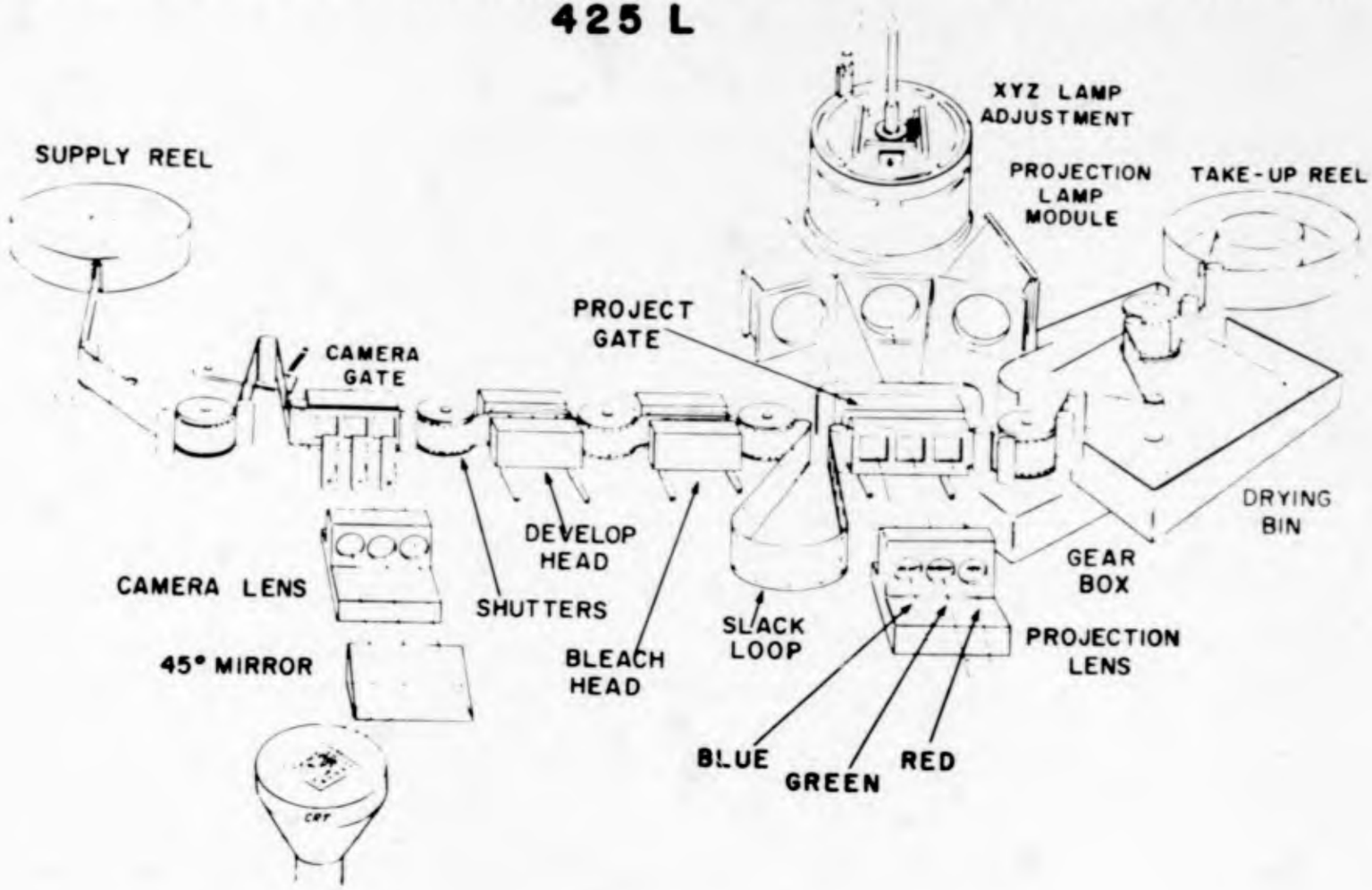


FIGURE 8 425L CAMERA-PROCESSOR-PROJECTOR UNIT

FIGURE 9 425L SYSTEM CONCEPT

FIGURE 10 425L DISPLAY EQUIPMENT

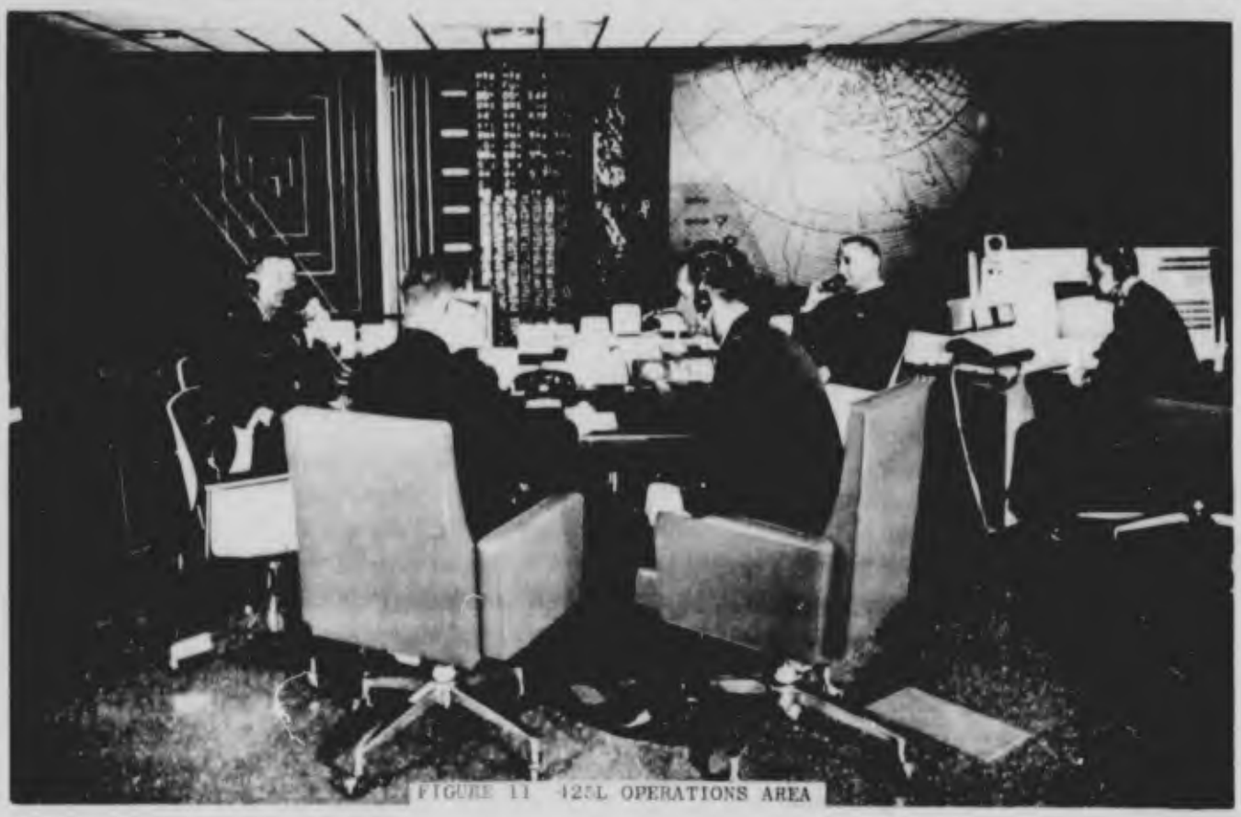


FIGURE 11 425L OPERATIONS AREA

PERFORMANCE CHARACTERISTICS

	425 L	465L	473L
PROJECTION SCHEME	REAR	FRONT	REAR
SCREEN SIZE (FT)	12 X 16	8 X8	1 X 1
THROW DISTANCE (FT)	22 ± 1/4	28 ± 1/2	3
ILLUMINATION (FT CANDLES)	26	30	100
FALL - OFF (%)	22	23	25
SYMBOL BRIGHTNESS (FT. LAMBERTS)	13	15	50
CONTRAST RATIO (ZERO AMB.)	150:1	100:1	30:1
SYMBOL HEIGHT (INCHES)	1.63	1.40	0.210
COLORS	SEVEN	SEVEN	FOUR
COLOR FRINGING (INCHES)	0.13	0.12	N/A
RESCLUTION (LINE PAIRS PER SCREEN HEIGHT)	940	910	1000+
LINEARITY (%)	0.5	0.5	0.5
REGISTRATION (%)	N/A	1.0	0.7
RESPONSE TIME (SEC)	10	15	15

FIGURE 12 PERFORMANCE CHARACTERISTICS

ANNUAL OPERATING COST

	425L	465L	473L
NO. OF UNITS	1	16	10
UPDATES / 24 HR DAY	70	83	50
COST / UPDATE	$ 0.25	$ 0.33	$ 0.09
CHEMISTRY COST/ UNIT	$ 3650	$ 5475	$ 510
FILM COST / UNIT	$ 2660	$ 4792	$ 1140
TOTAL COST / UNIT	$ 6310	$ 10,267	$ 1650
TOTAL COST / YEAR	$ 6310	$ 164,212	$ 16,400

TOTAL COST PER YEAR FOR THE THREE SYSTEMS BASED ON ABOVE → $ 186,972

TOTAL COST PER YEAR FOR THE THREE SYSTEMS BASED ON A (5) MIN UP-DATE RATE FOR THE THREE SYSTEMS → $ 616,456

FIGURE 13 ANNUAL OPERATING COST

BLANK PAGE

(U) ANALYSIS OF DIGITAL COMMUNICATION SYSTEMS ON TROPOSPHERIC SCATTER

by

Robert G. McLaughlin, 1/Lt., USAF

Rome Air Development Center
Research and Technology Division
Griffiss AFB, New York

1/Lt Robert G. McLaughlin

BIOGRAPHY - 1ST LT. ROBERT G. MCLAUGHLIN

Lt. McLaughlin graduated with a BEE degree from Rensselaer Polytechnic Institute, Troy, New York in June 1962 and received an Air Force Reserve Commission as a 2nd Lieutenant. He entered active duty in August 1962 and was sent to RPI by the Air Force Institute of Technology for graduate school. He received his MEE degree in June 1963 and was assigned to the Rome Air Development Center, Air Force Systems Command as a project officer in the Communications Signal Processing Section. He was promoted to 1st Lieutenant in February 1964 and received a Regular Commission in July 1964. He attended Squadron Officer School in January 1966 and was a Distinguished Graduate.

Lt. McLaughlin does research in statistical communication theory. He has published several papers as RADC technical reports. He is a member of the Institute of Electrical and Electronic Engineers and is a member of the IEEE Professional Technical Groups on Circuit Theory, Communication Technology, and Information Theory.

ABSTRACT

In FY-65, the Air Force tested three new time-modulation-multiplex communication systems. The tests were run over a 375 mile tropospheric scatter circuit at the Atlantic Missile Range. At the same time, the Communications Signal Processing Section of the Rome Air Development Center analyzed the performance of these new systems using the same path.

This paper discusses the three systems, the mathematical models and analysis, and presents the theoretical and experimental results. We used the Sunde model of tropo-scatter to compute one performance measure, called the error probability, for each system. The model describes the effects of tropo-scatter on signals in terms of statistics. For each system, our final result is a curve of error probability versus the received average signal-to-noise ratio.

The curve for each system shows how well that system will work on a specific communication circuit. Pulse distortion limits the performance of each system, no matter how large the signal-to-noise ratio is made. We compared our results with the test results for each system and found excellent agreement. The tests also showed performance limits at high signal levels, and the test limits were very close to the predicted limits.

The combined results of the analysis and tests show that the Sunde model may be used to predict system performance. With the modeling techniques we used, the Air Force can now analyze proposed digital systems for tropospheric scatter.

Introduction

For several years, tropospheric scatter has been used for communication between stations 100 to 400 miles apart. Communications specialists have devoted a lot of effort to mathematically describe the effects of tropo-scatter on communication signals. This paper will discuss the performance analysis of three digital tropo-scatter systems. The discussion will include the mathematics used to describe tropo-scatter, each of the systems, and both analytical and experimental results. First, I will give a little history of the problem. Then, I will discuss the mathematics, or model, used to describe tropo-scatter. Finally, I will discuss each of the three systems, describe the analysis, and give the results.

In FY-65 the Air Force tested three new digital communication systems on a 375 mile tropo-scatter circuit at the Atlantic Missile Range. Each of the systems used pulses to transmit up to twenty-four channels of either digital or voice-like information. At the same time the Communications Signal Processing Section at the Rome Air Development Center analyzed the performance of these systems. The purpose of the analysis was to evaluate and verify a recently developed mathematical model of tropo-scatter. If we could show that results using the model were close to experimental results, the Air Force could then analyze new proposals for tropo-scatter communication systems. The analysis cannot replace testing new systems in the field, but it can allow us to eliminate systems that cannot possibly do the job. A full report of the systems analysis and results has been published[1] and is available from the Defense Documentation Center.

Tropo-Scatter Model

In 1964 a new model of tropo-scatter was developed by E. D. Sunde of the Bell Telephone Laboratories[2]. Our section felt that this would be a good model of tropo-scatter and thus, our analyses were based on Sunde's model. I will only discuss Sunde's model as it applies to digital communication systems.

A very common performance measure for digital communication systems is called the error probability. Let us consider a binary system, one that transmits either a "zero" or a "one". Each "zero" or "one" is called a bit. Most present computers also use binary data to represent information. If the transmitter sends a "one", but the receiver decodes the signal as a "zero", or vice versa, then the receiver has made an error. A measure of system performance or reliability, is the probability that the receiver will make a mistake when it decodes a bit. This probability is called the error probability for the system. Experimentally we can measure this number for any system by sending many "zeros" and "ones" and counting how many mistakes the receiver makes. Then the measured error probability is equal to the number of errors divided by the total number of bits sent. If error probability is the performance measure we decide to use, the system with the smallest error probability is the best. Of course there are many other measures of performance, but this is the one we used in our analysis.

The Sunde model assumes that a tropo-scatter circuit distorts signals in certain ways that can only be described statistically. Given the

circuit geometry, transmitter and receiver characteristics, and the receiver average signal-to-noise ratio, the model tells how to compute the error probability for a system. For each system we computed a curve of error probability vs. average receiver signal-to-noise ratio. As with any statistical measurement, the experimental data varied above and below our theoretical curve. Our results did, however, appear reasonable when we compared them with the test results.

The Sunde model categorizes errors into three types. The first type of error is caused by variations in the troposphere which produce a phase or frequency (doppler) shift in the received signal. This type of error becomes important as the transmission rate, or number of transmitted bits per second, becomes small. In the three systems we considered, the transmission rate was high enough that errors from phase or frequency shifts were insignificant. The second type of error is caused by random noise at the receiver and variations in the strength of the received signal. The variations are called "fading" and are important for any communication system. The type of fading found in tropo-scatter is called Rayleigh fading and is completely described by the average signal-to-noise ratio. With this type of error cause, the error probability goes down as the average signal-to-noise ratio goes up. The error probability, however, never goes to zero, although we can make it as small as we like. The third type of error is due to the fading, called frequency-selective fading, being different at different frequencies. This type becomes important as we increase the transmission rate and thus the bandwidth of our system. If the pulses we use to transmit "zeros" and "ones" have a large bandwidth, this type of fading distorts the pulses and actually spreads the pulse energy in time. If the spread is large enough, energy from one pulse will interfere with energy from another pulse and the receiver may make an error. This distortion is important because it limits how small we can make the error probability, no matter how powerful we make the transmitter. It is sometimes called the irreducible error probability.

The Sunde model describes how to compute the error probabilities due to the three causes I have described above. Actually, the system errors are due to a combination of the three causes. If the error probability for each cause is small enough, however, the overall error probability is very close to the sum of the individual error probabilities. For any useful communications system this condition is satisfied so that our analysis does add the individual probabilities to get system performance.

Rayleigh fading describes the variations in received signal strength found on tropo-scatter. The term, Rayleigh fading, means that for a constant transmitted signal power, the received signal power is described by a Rayleigh probability density function. If we let the received signal power at any time be S, the average received signal power be S_i, and the probability density function for S be p(S), then:

$$p(S) = e^{-(S/S_i)}/S_i \qquad S \geq 0$$
$$= 0 \qquad S < 0 \qquad (1)$$

The receiver's internal noise is usually described by a Gaussian probability density function. Let the noise voltage at any time be v, the average noise voltage be zero, the average noise power be N and the probability density function for v be p(v). Then:

$$p(v) = \frac{1}{\sqrt{2\pi N}} \; e^{-(v^2/2N)} \qquad (2)$$

With Equations (1) and (2) plus a knowledge of the receiver structure, the model shows how to compute the error probability due to noise and fading.

The model also gives a method for calculating the error probability due to frequency selective fading. To compute the effects of tropo-scatter on pulse distortion, the model assumes that, at any specific time, the troposphere acts as a linear filter. The characteristics of this filter vary statistically in time, but vary much more slowly than the transmission rate. Thus, we can get the error probability by finding the error probability for fixed filter characteristics and averaging over all possible filter characteristics. The model assumes that at any time, the filter has a gain function and an envelope delay function that both vary linearly with frequency (a graph of either function vs. frequency is a straight line). Sunde has derived the probability functions for the slopes of the two straight lines. To use the model then, we need to find the amount of interference between pulses for the different slopes of gain and envelope delay, and then find the average distortion and error probability.

To get the complete error probability, we just add the two error probabilities I have discussed above. Then we can plot a curve of error probability vs. average signal-to-noise ratio, which is the desired result.

Time Division Multiplex/FSK System

The first system we analyzed was a time division multiplex/FSK system. The bits from each channel were multiplexed in time, the first bit from Channel 1, the second from Channel 2, and so on. Then this stream of bits was transmitted by frequency shift keying (FSK). If a bit is a "zero", a pulse on one frequency is sent; if the bit is a "one", a pulse on a second frequency is sent. The receiver examines the amount of energy on each frequency, and decides which frequency has more energy. Then the receiver puts out the appropriate bit. These bits are unscrambled into the individual channel bits. This system was analyzed, as an example, in Sunde's paper so that we were able to use his equations directly. Table I shows the number of channels and error probability caused by frequency selective fading only (no noise).

TABLE I

CHANNELS	ERROR PROBABILITY
24	0.034
12	0.005
6	0.00065

Figure 1 compares the theoretical results, including noise, with the experimental results. As can be seen, the system operation is marginal. Note that the system cannot be improved by merely increasing the transmitter power. This system, with the required number of multiplexed channels, is apparently not suitable for a tropo-scatter circuit of the length used in the test.

Random Acess/TSK System

The second system we investigated used a very complex multiplexing scheme. Voice information was transmitted by pulse-position modulation. The binary information was transmitted by a set of pulses starting either at one time or another, which I have labeled as time shift keying. Twenty-four channels were designed for information and the twenty-fifth channel was used for synchronizing the system. Now, I will describe how the twenty-five channels are multiplexed and then separated at the receiver.

Each of the twenty-five channels was given its own address, consisting of four pulses centered on four different frequencies (which are the same for all channels) and occurring in four out of six possible time positions. The addresses for the channels are shown in Table II, where the numbers 1, 2, 3, 4 represent the frequency at which a particular pulse is sent and the blanks represent the absence of a pulse.

TABLE II - CHANNEL ADDRESSES

Time Position	1 2 3 4 5 6
	1 2 3 4
	1 2 3 4
	1 2 3 4
	1 4 2 3
	1 3 4 2
	2 4 1 3
	2 4 1 3
	2 4 1 3
	2 3 4 1
	2 1 3 4
	3 1 4 2
	3 1 4 2
	3 1 4 2
	3 2 1 4
	3 4 2 1
	4 3 1 2
	4 3 1 2
	4 3 1 2
	4 2 3 1
	4 1 2 3
	4 3 2 1
	4 3 2 1
	1 4 3 2
	2 1 4 3
	3 2 1 4

The receiver had a filter centered at each of the four frequencies. The output of each filter was passed through a signal detector which is called a square law envelope detector. Each channel had switches which connected the channel to each filter at the proper time, as determined by the channel's address. Each channel had delay circuits for each filter which lined up the filter outputs in time. Then the outputs were added to form the total output. For each channel, this procedure was done for each of the possible pulse positions (two in the binary case). The receiver then decided what bit was sent based on the position with the largest total output

Our analysis first assumed that the channel addresses did not interfere with each other. Then, the problem was one of combining four signals with Rayleigh fading and noise and comparing that with four channels with noise alone. This problem had been solved before in the literature so we just used the existing equation for error probability. This result is shown in Figure 2 and is the curve that varies with average signal-to-noise ratio.

The structure of the addresses for this system was such that the addresses could interfere with each other. For example, the receiver could not discriminate between the Channel 1 address in one position and Channels 2, 14, 10, and 8 in the other position. Another problem was caused by a non-linear power amplifier at the system output, so that the transmitter power for any one pulse depended on the number of pulses that were transmitted at the same time. We were unable to solve either of these problems directly so we resorted to a digital computer simulation of the problem. Using our knowledge of the code structure, we wrote computer programs for the RADC Control Data Corporation 1604B digital computer. The programs randomly put the channels into the various positions and counted how often enough interference occurred to cause an error. With one million trials, we found that the error probability, even with no noise, could not be better than 1.44×10^{-4}. This completed our analysis of the system and the results are shown in Figure 2 along with the experimental results.

Combination TDM-FDM/FSK System

The third system we analyzed used a combination of time division multiplexing (TDM) and frequency division multiplexing (FDM). Then the information was transmitted by frequency shift keying. The difference in FSK between this system and the first system I discussed is that this system does not require all channels to use the same two frequencies for FSK. This system used five pairs of frequencies to transmit by FSK. The channels were broken into groups of five and each group was assigned a pair of frequencies. With more frequency pairs available, this system was able to have large time gaps between pulses on the same frequency, thus allowing the pulse distortion to die away before it interfered with another pulse. Actually, the system still had interference from different frequencies and different times, but the amount of interference was reasonable. The receiver was very much the same as for the other systems, the difference being in the switch circuitry to connect the channels to the various filters.

Sunde's paper has an equation for the error probability for this system, FSK, working with Rayleigh fading and random noise. This equation is graphed in Figure 3. Then we proceeded to calculate the error probability due to

pulse distortion. The method for finding the pulse distortion for a fixed slope of envelope delay is given in Sunde's paper and I will briefly discuss it in the next paragraph. Once we had the pulse distortion for any fixed delay slope we averaged the distortion over all possible delay slopes. Doing this for the various cases of pulse interference gave us the ratio of average signal-to-distortion power ratio. Then, we treated this ratio as though it were a signal-to-noise ratio and thus, we could use the same equation for error probability that I discussed earlier in this paragraph. Note that this signal-to-distortion ratio cannot be reduced by increasing the transmitter power, and thus, we have an irreducible error probability. Using the analysis in the next paragraph, the minimum error probability for this system is 2×10^{-5}. Now I will discuss how to compute the signal-to-distortion ratio.

From our study of the equipment we knew the transmitted pulse shape and the receiver filter transfer function. To this we had to add the tropo-scatter transfer function. Then by the use of Fourier transforms we computed the received distorted pulse shape for any pulse combination. We were unable to compute the transform exactly, so we used the digital computer to numerically approximate the transform. Then, we averaged the pulse interference values for all values of delay slope and computed the signal-to-distortion ratio. The theoretical and experimental results are shown in Figure 3.

Conclusions

In this paper I have described the performance analysis of three communication systems operating on a tropospheric scatter circuit. The analysis used the Sunde model to mathematically describe the effects of tropo-scatter on the signals. The performance of each system is described by its error probability. The theoretical results agree very well with the experimental results for all of the systems. This agreement indicates that the Sunde model is a good model of tropo-scatter, and its use can give a good idea of the performance of tropo-scatter digital communication systems.

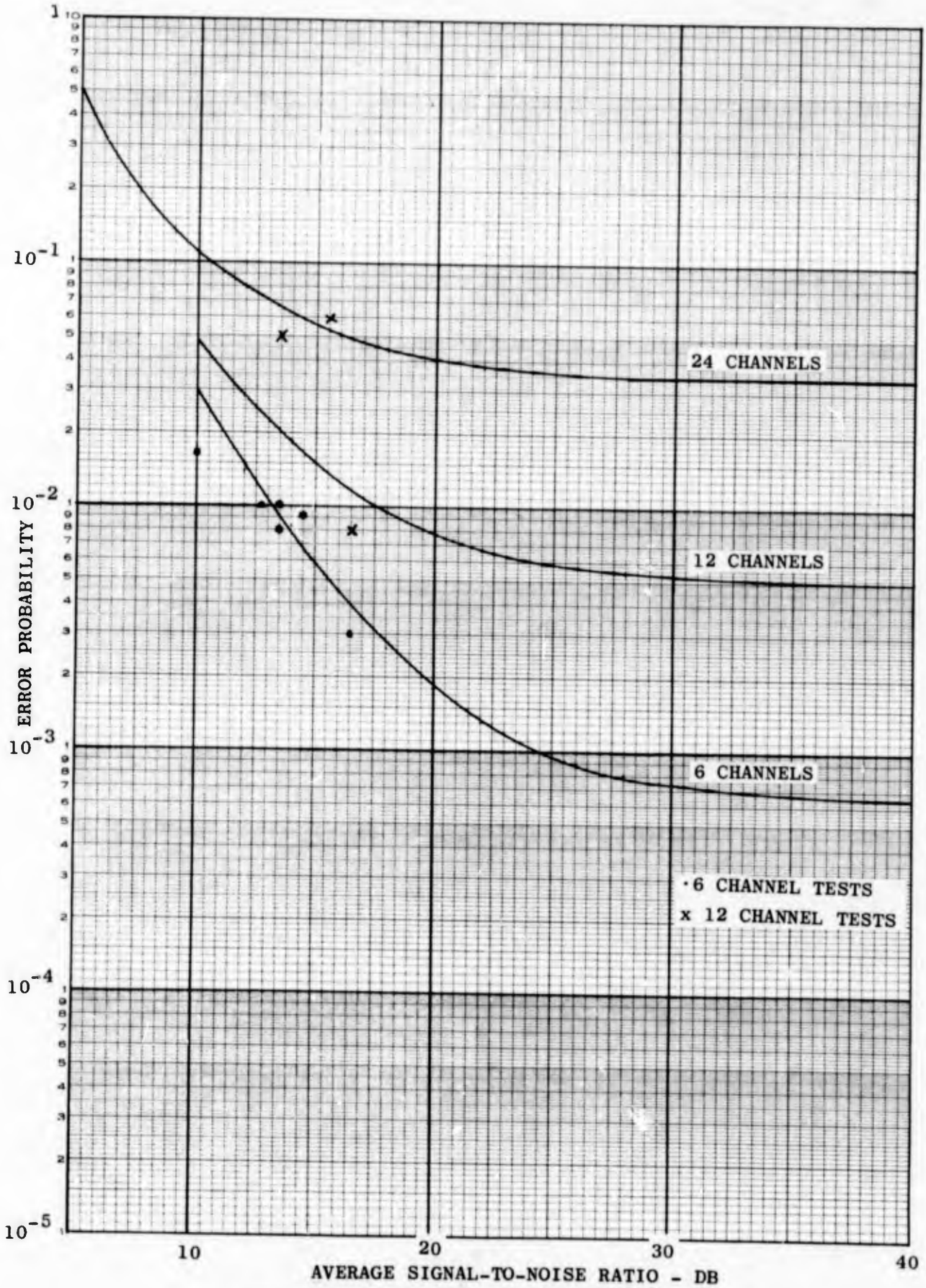


FIGURE 1 TIME DIVISION MULTIPLEX/FSK SYSTEM

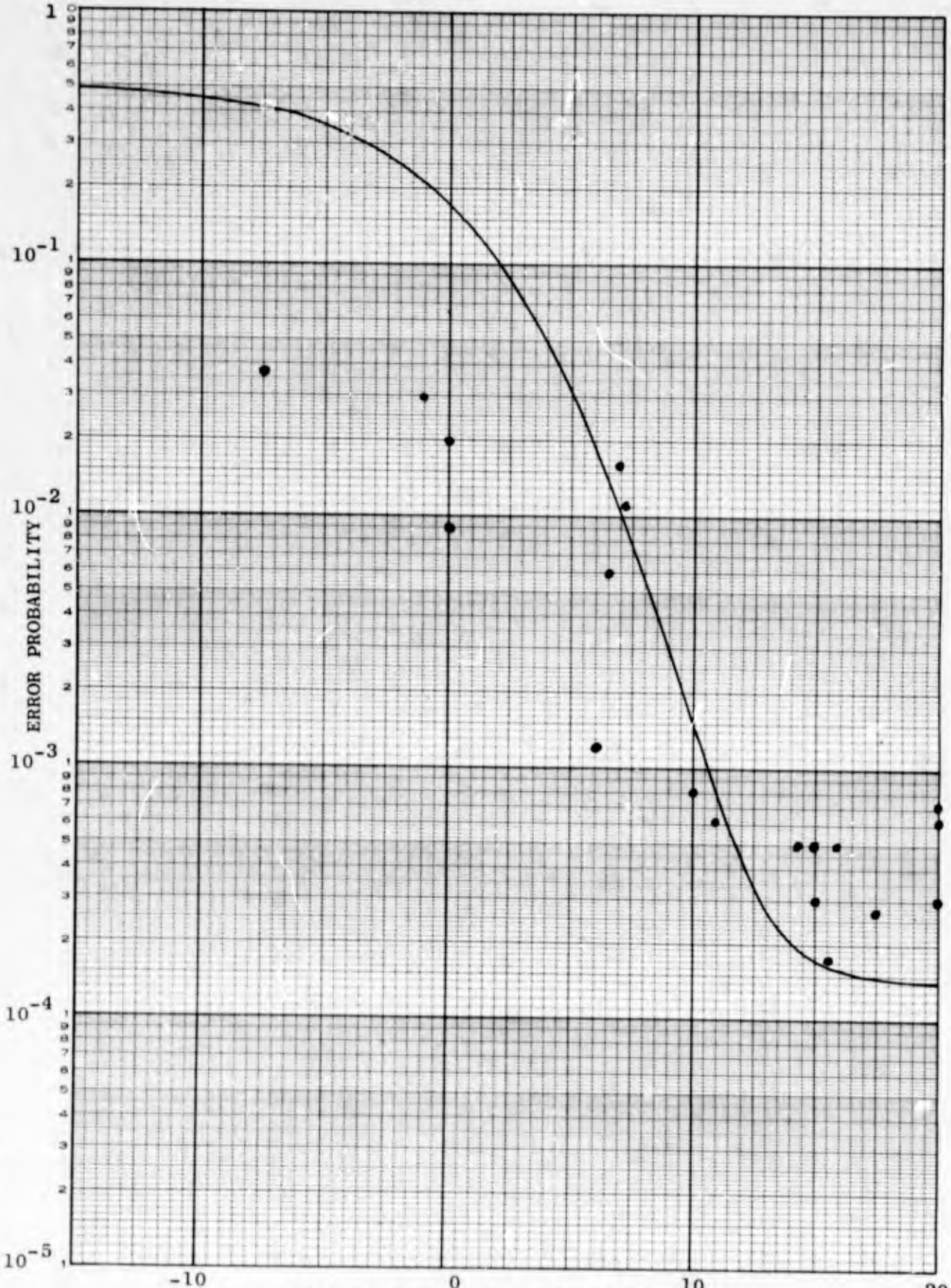


FIGURE 2 RANDOM ACCESS/TSK SYSTEM

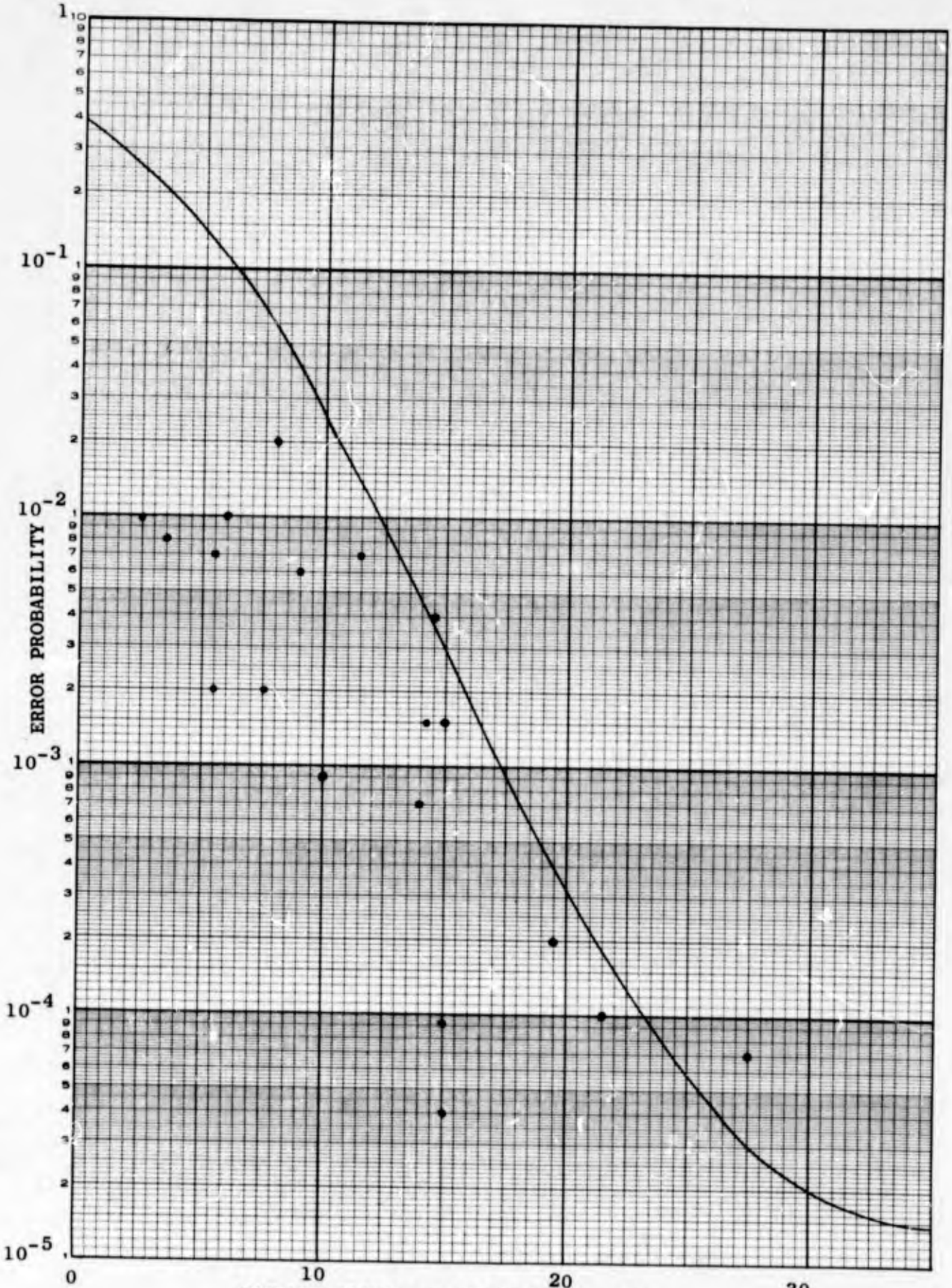


FIGURE 3 COMBINATION TDM/FDM/FSK SYSTEM

(U) EVOLUTION OF COMPUTER SYSTEMS TO PERFORM PARALLEL PROCESSING

by

Morris A. Knapp

Rome Air Development Center
Research & Technology Division
Griffiss Air Force Base, New York

BLANK PAGE

Morris A. Knapp

BIOGRAPHY

Name: Morris A. Knapp

Date and Place of Birth: 11 December 1938; Utica, NY

Elementary School: Schuyler Elementary School, Schuyler, NY, Graduated 1952

High School: Whitesboro Central High School, Whitesboro, NY Graduated 1956

College: Clarkson College of Technology, Potsdam, N.Y. B.S.E.E. 1960

Graduate Work: Syracuse University GAFB Center, Rome, NY

Employment: Rome Air Development Center, GAFB, NY

Employed in the Data Processing Section, Information Processing Branch, Intelligence and Information Processing Division. Mr. Knapp is responsible for advanced developments of computer organization to fulfill future Air Force requirements in data processing. This work is carried out jointly, in-house and contractual support.

BLANK PAGE

Evolution of Computer Systems to Perform Parallel Processing

ABSTRACT

The paper begins with a brief historical development of digital computers, starting with the Roman-type abacus and being carried up to the computer design described by Dr. von Neumann. Some problems, like numerical weather forecasting, handling of data from large phased array radar, mapping and charting, pattern recognition, solution of large matrices, atomic reactor calculations, etc., have data processing requirements which cannot be handled effectively with present day computer systems and require some type of parallel processing.

The Iterative Array Computer is the first and to date the only general parallel network computer which is capable of performing up to one hundred simultaneous calculations. The organizational design of this system, whereby a general purpose computer was used as the central control unit for a ten by ten array of processing elements, will be given in detail. These one hundred processing elements which make up the array are all identical, small, serial, one megacycle per bit computers without instruction decoding capability. The logical and hardware design of the entire system will be given in detail where integrated molecular circuits were used for the implementation of the logical functions.

The software design will be given along with an explanation of how the micro-commands can be used by the programmer to generate his own special instructions for the array. The Iterative Array was designed around the use of the bit patterns which can be generated on the

input/output lines of the CDC 160 computer. Two CDC 160 words are required for one 24 bit instruction word for the array; this instruction word is broken down as follows: 13 bits for the micro-command, 5 bits for routing of data, and 6 bits to address memory. An assembler language has been developed for the system and will be described.

A sample of the target assignment problem using the algorithm developed by James Munkres has been programmed for the Iterative Array Computer system to illustrate how the parallel organization can be used in problem solutions.

I. Introduction to Computer Development

The basic reason for the development of a computer is to have a tool which can be of assistance to humans in the solution of their problems. Humans are very slow and generally unreliable in the performance of detailed numerical operations. Therefore, with this understanding of the requirements for a device to aid humans, the abacus could be considered as the first such device, or the forefather of a computer. The modern abacus consists of a rectangular frame with a number of rods or wires. These rods are divided into two unequal portions by a transverse bar. On the upper, smaller portion of each rod are two beads and on the lower portion, five beads. (The modern Japanese abacus has one bead above and four beads below the bar.) The whole acts as a numerical register, each rod representing a decimal order as in our familiar "arabic" notation.

The principal disadvantage of the abacus, however, is that the operator must perform his calculations mentally. The device merely stores his results step by step. It does not perform the actual calculations as does the modern desk calculator.

The principles of modern computers, albeit in mechanical form, were laid down by Charles Babbage in England. He worked on the design of an analytical engine, which, had it been built, would have incorporated ideas familiar to the modern digital computer.[1] Babbage's idea was to extend the capabilities of the difference engine so that it could not merely add and print, but also multiply, divide and call for new data from its human operator. In order for the analytical engine to perform the tasks expected of it, it was necessary to express all the possible instructions to the machine in the form of stereotyped commands. Such a procedure is known as the logic or the logical design of a computer, and Babbage's engine was strikingly modern in regard to its logical design. Except for the need of a human attendant to read into the machine values from mathematical tables, this engine was logically parallel to most of the recent automatic computers.

The next major development, prior to the advent of the high-speed electronic automatic computer, came in 1944 when Howard Aiken, in cooperation with some IBM engineers and some graduate students at Harvard University, completed the Mark 1 Relay Computer, also called the Automatic Sequence-Controlled Calculator.[2]

This machine performed thousands of calculating steps, one after another, according to a scheme fixed ahead of time. Its individual operations were automatic, once the punched tape fixing the chain of operations had been put on the machine, and it was sequence-controlled, since control over the sequence of its operations was built into the machine.

Dr. John von Neumann is considered the father of digital computers as they are known today (stored program computers). Dr. von Neumann defined an automatic computing system as "a device which can carry out instructions to perform calculations of a considerable order of complexity."[3] He then stated that instructions governing this operation must be given to the

device in absolutely exhaustive detail; that instructions must be given in form the device can sense; and that once these instructions are given to the device it must be able to carry them out completely and without any need for further intelligent human intervention.

The general purpose computers on the market by various manufacturers even today follow the basic layout given by von Neumann. There have been improvements made in many areas of the machine, of course; but the basic organization is very similar to the concepts of von Neumann's computer. Several orders of magnitude increase in processing speeds have been acquired over the years. Most of these increases have resulted from improved hardware, improved programming techniques, and some minor modifications to the organization of the machine. Even the newest series of computers (IBM 360 series, RCA Spectra series, G.E. 600 series, CDC 6000 series, Honeywell series, etc.) have basically the same processor internal organization and perform calculations sequentially. The only minor exception is the CDC 6600 where the main processor can overlap two or three arithmetic operations. Even the computers being designed for time sharing operations perform calculations serially. Each processor operates on one program at a time for a short burst and then sequences to the next program within the queue.

There are many general problems, both military and commercial, that are not satisfied by currently available sequential processing computers. The inabilities of the sequential type computers to meet some real-time processing requirements are imbedded in the limitations of component switching speeds, memory speeds and time delays of signal propagation. Some problems like numerical weather forecasting, handling of data from phased array radar, mapping and charting, pattern recognition, atomic reactor calculations, etc. have data processing requirements which cannot be handled effectively with present day computers. Some type of parallel or multi-processing has to be performed to effectively handle these types of problems within realistic time requirements.

The development of a Parallel Network Computer was initiated by Rome Air Development Center, Griffiss AFB, in September 1961. At that time there were two distinctly different concepts for a new computer organization based upon a parallel network approach. One concept was formulated by Dr. J. Holland[4], University of Michigan and was called the Holland Machine. This machine concept has an array of identical computer modules, interconnected to the four neighbors, in a regular array, each module having arithmetic and logic control capabilities. Therefore, this type of computer has distributed control capability and is extremely difficult to program. The other concept, formulated by Dr. D. Slotnick[5], was called the SOLOMON computer. The SOLOMON computer concept also has an array of identical processors, but differs from the Holland Machine in that its processors are operated in synchronism under the control of a centralized controller. The Parallel Network Computer Organization using Central Control was attacked, both contractually for the design and development of the SOLOMON computer, and in-house at RADC by the development of the Iterative Array Computer. The Iterative Array Computer is the only general computer built to date which is capable of performing up to one hundred operations simultaneously on different data.

II Organization of the Iterative Array Computer

The functional layout and preliminary organization design was started in September 1961 at RADC and in August 1962 it became a Discretionary Fund Project (DW-63-2) with Mr. Morris A. Knapp as Principal Investigator. The basic approach was to use a standard general purpose computer as the central control unit and to design an array of processing units to operate in unison under the control of the central processor. A Control Data Corporation 160 computer was selected, because of availability, as the central processor and the design of the array proceeded around the CDC 160. Modification to the CDC 160 was restricted to minor ones only, the preference being no modifications at all; therefore, the organization and design of the array had to be oriented around the information available from the computer's input/output channels. Also the flow of data and information between the array and the external world was to be via the CDC 160 computer. The system's function organization (as illustrated in Figure 1) is oriented around the use of the CDC 160 as the central controller and the communication to the external world.

The CDC 160 computer having a 12 bit word length has only 12 information lines in the input/output channel. In the design of the Iterative Array, it was impossible to address memory and still have enough control bits within the single 12 bit word; therefore, it was necessary to use a double CDC 160 word as an instruction to the Iterative Array. A 24-bit word was used as the word length in each processing element to obtain more accuracy for fixed point arithmetic operations within the array.

The Iterative Array Computer is composed of one hundred identical processing elements, each of which may be thought of as a small computer, capable of performing logic and storing data. A control buffer between the Iterative Array and the control unit accepts 12-bit instructions from the control unit and translates them into 24-bit commands for the Iterative Array. An input/output buffer converts the serial-by-word parallel-by-bit 12-bit control unit data words to the 24-bit parallel-by-word serial-by-bit format of the array data words and transmits the array data words to the 10 left edge processing elements. There is also a single 24-bit register, called the broadcast register, which can simultaneously transmit the same data word to all 100 processing elements.

The Iterative Array, the heart of the parallel processing system, is a 10 x 10 array of identical processing elements (PE's) interconnected. Each processing element performs the same operation, in parallel, on its own unique data. Each processing element, except some of the Pe's on the edge of the array, is connected so that it may receive data from each of its four nearest neighbors and the broadcast register. Each PE can perform logic on data stored in its own memory and accumulator, its own memory and one of its neighbors accumulators, or its own memory and the broadcast register. There can be direct transfer of information from a neighboring accumulator to the accumulator or memory of a PE. The routing from the above, below, left, or right neighbor's accumulator, from the broadcast register, or from a PE's own accumulator to the module's arithmetic unit is the same for all 100 PEs during any given instruction cycle. By using

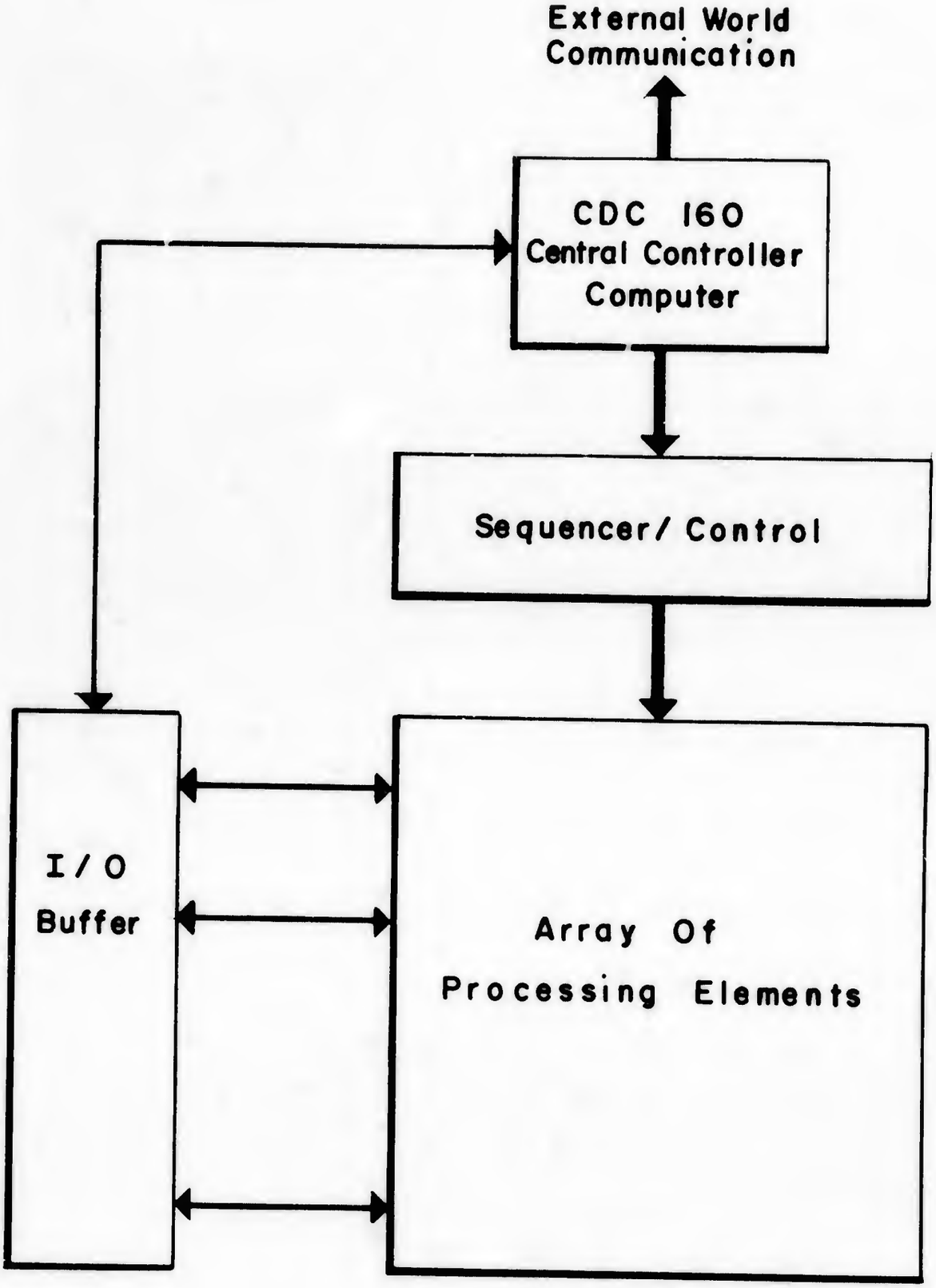


Figure 1. System Configuration

the broadcast register, the array can also simulate an associative memory of 100 parallel search words with storage of up to 6300 words, with complete arithmetic and logic capability at each search word.

Figure 2 illustrates the functional logic diagram of a PE which is duplicated 100 times in the system. The basic information rate of each PE is 1-megacycle serial-by-bit. Both the accumulator and the 63-word memory are serial delay lines which operate at 1 megacycle bit rate. The main reason for using delay lines in place of flip-flops and core memories was to make the PEs as inexpensive as possible to fabricate when they were to be reproduced 100 times within the system. Information for the adder or binary logic unit is provided by the memory, accumulator (its own or one of the neighbors') or broadcast register. The arithmetic unit can add two binary bits of information, transfer information without modification from either memory to accumulator or accumulator to memory, or perform Boolean operations (AND, OR, exclusive OR, etc.) on the 24 pairs of bits in 2 words. The arithmetic operations are performed in twos' complement.

The PE network is designed to simultaneously execute a common instruction. However, control is provided within each PE to give it the capability of executing or ignoring a given instruction. The S flip-flop can block the flow of information into the memory or accumulator; this provides a measure of limited distributed control within the array. The only control distributed in the array is the ability to keep individual or groups of PEs from executing a specific command, depending on the mode of its S flip-flop. The mode of the S flip-flop can be set by information within each PE. The techniques used to control the execution of instructions within the individual PEs is called selective or multi-modal control. Basically, each PE can be in either of two mode states. Certain 24 bit instructions can specify selective control and the control status required for execution. As an example, if a given instruction specifies addition in mode 1, only those PEs in mode 1 will execute the instruction, those in mode 0 do not. A group of mode control instructions which allow for sensing, changing states, and loading and storing can be built from the micro-instructions.

The broadcast register is a special 24 bit delay line which is connected to each of the 100 processing elements. It is connected to the arithmetic unit of each PE (as illustrated in Figure 2) in the same manner as the accumulators. The same data word may be transferred from the broadcast register to the accumulator or memory of each PE. Broadcast register data may also be combined with memory data in each of the PEs, either in an arithmetic or Boolean operation before being stored in the memory or accumulator.

The input/output buffer performs the function of transposing the 12 bit parallel word from the CDC 160 to 24 bit sequential for the processing elements and the reverse conversion. This buffer consists of ten 24-bit delay lines connected to the 10 PEs at the left hand edge of the array and a 12-bit shift register which is connected to the input/output lines of the CDC 160 computer. Data is loaded into the memories of each processing element via this input/output buffer. Twenty CDC 160 words are transferred, sequentially, to the input/output buffer to completely fill it, then ten words (one to each processing element in the column) are transferred to the array sequential by bit.

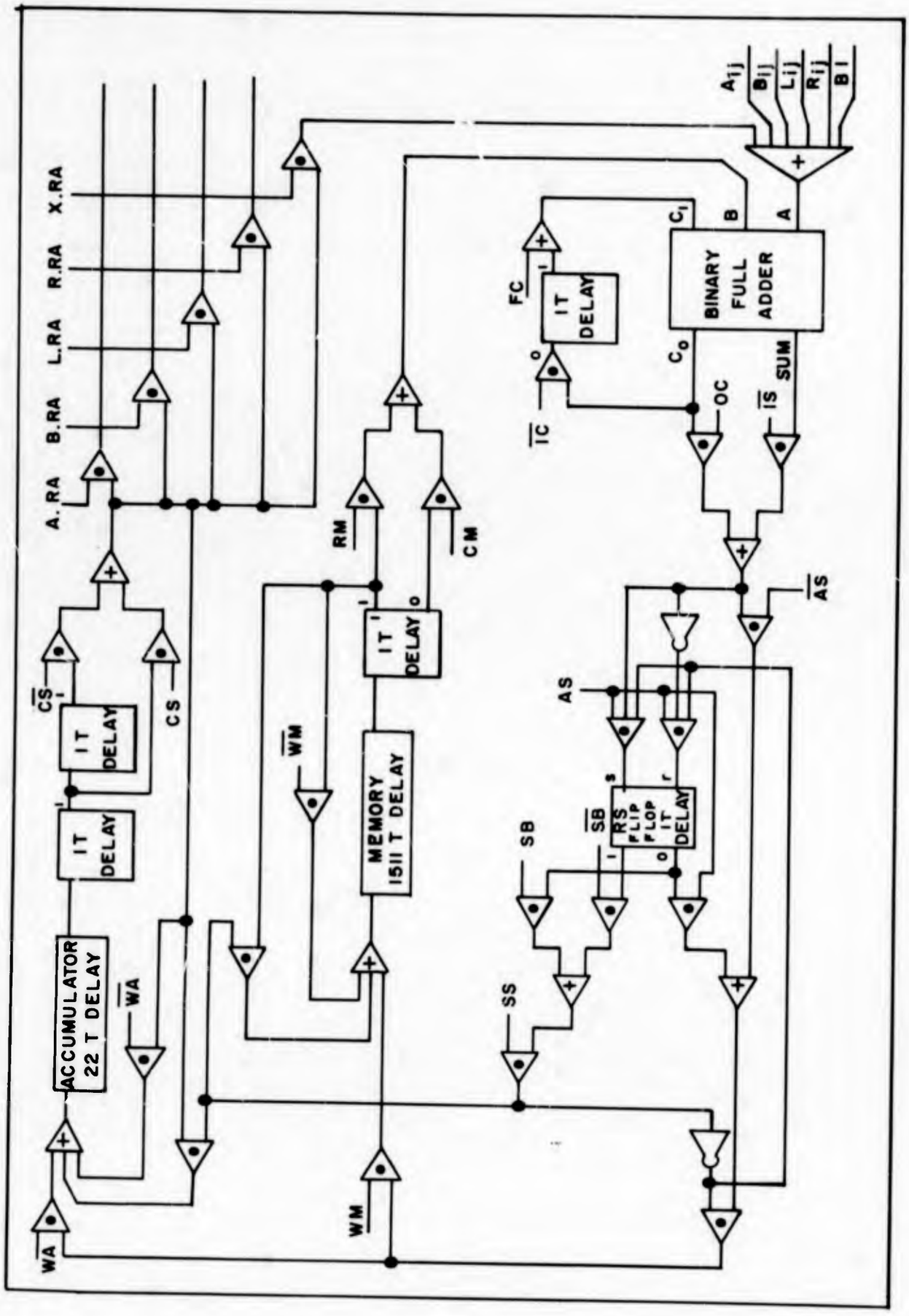


Figure 2. Functional Block Diagram of A PE

The synchronizer has the function of interpreting and routine the control information being sent to the Iterative Array from the CDC 160 computer. This device has a function decoder which interprets the following codes and routes the control signals to the correct location.

Function Code	Mode
7700	Execute Instructions
7720	Sense Mode Status
7740	Input/Output of Data
7760	Broadcast Load

These control signals will be accepted by the different devices to mean that the particular device is to accept from, or supply to, the CDC 160 some data. An OUT command within the CDC 160 program will start the flow of data on the input/output lines of the CDC 160 after it receives an information request signal.

The instructions for the Iterative Array evolved around the Micro-Commands which are used to control the different functions to be performed by the Processing Elements. As stated earlier, two 12-bit words from the CDC 160 are required to make a single instruction for the Iterative Array. The command word format illustrated in Figure 3 shows the bit locations of the individual micro-commands. The 24 bit instruction word is broken down as follows: 13 bits - micro-commands; 5 bits required to route information; and 6 bits required to address any one of the 63 words of memory location which is to be used by each processing element. The thirteen basic micro-commands have been defined as follows:

First Command Word

Microcommand	SS	SB	IC	OC	IS	AS	FC	CM	WM	RM	WA	RA
Bit	11	10	9	8	7	6	5	4	3	2	1	0

Second Command Word

Microcommand	B	A	L	R	BI	CS	ADDRESS					
Bit	11	10	9	8	7	6	5	4	3	2	1	0

Figure 3. Command Word Format

BIT	MNEMONIC	DESCRIPTION
0	RA	READ ACCUMULATOR - The contents of the accumulator not only recirculate in the accumulator, but are gated into the arithmetic logic at the adder input of the processing modules specified in the second half word. Data circulates from one processing module to any of the nearest neighbors through the RA gating.
1	WA	WRITE ACCUMULATOR - The recirculation of data in the accumulator delay line is inhibited, and the output of the arithmetic logic is read into the accumulator.
2	RM	READ MEMORY - The contents of the processing module memories at the address specified in the second half word are gated into the arithmetic logic at the adder input in the processing module. Recirculation of the memory is not affected. Data in the memory of the (i,j)th processing module can only be read into the (i,j)th processing module's arithmetic logic.
3	WM	WRITE MEMORY - The circulation of the data stored in the processing element memory at the address specified in the second half of the command word is inhibited, and the output of the arithmetic logic is gated into the specified memory location.
4	CM	COMPLEMENT MEMORY - This micro-command performs the same function as does the RM micro-command except that the one's complement of the memory location contents is read into the arithmetic logic instead of the memory location contents.
5	FC	FICTITIOUS CARRY - This command causes a one to be presented to the carry input of the adder circuitry at the lowest order (least significant) bit time. This micro-command is the only one THAT DOES NOT AFFECT processing element operation for a whole word time.
6	AS	ACTIVATE S FLIP-FLOP - The output of the arithmetic logic is used to set the S flip-flop, and the normal arithmetic output path to the memory and/or accumulator is inhibited, and the contents of the S flip-flops at each bit time are read into the memories and/or accumulators.
7	IS	INHIBIT SUM - The sum output of the adder is inhibited so that it does not circulate into the S flip-flop accumulator, or memory input logic.

BIT	MNEMONIC	DESCRIPTION
8	OC	OUTPUT CARRY - The carry output of the adder is read into the S flip-flop, accumulator, and memory input logic. If the IS micro-command is not activated with this command, the "or" function of the sum and carry will be read into the S flip-flop, accumulator, and memory input logic.
9	IC	INHIBIT CARRY - The normal circulation of the carry output of the adder through the one timing period delay flip-flop into the carry input of the adder is inhibited.
10	SB	SELECTED BIT - This micro-command works in connection with the SS (Selective) micro-command. If the SS micro-command is activated, only those processing modules whose S flip-flop matches the SB bit at the bit time under consideration will execute the instruction called for.
11	SS	SELECTIVE - This micro-command operates in conjunction with the SB micro-command. If the SS micro-command is activated, only those processing modules whose S flip-flop contents match the SB micro-commands execute the specified array instruction. Instruction execution is inhibited in all other processing modules.
6 (second word)	CS	CONTROL SHIFT - This micro-command causes information coming from the accumulator to be right shifted one place algebraically, by by-passing a flip-flop in the output of the accumulator. The lowest order bit (b_o) of the accumulator is lost and the sign bit (b_{23}) is rewritten in position b_{22} and b_{23}. Information in all other bit positions, n + 1, is shifted into bit position n.

Theoretically, using the 13 micro-commands $2^{13}(8192)$ instructions or macro-commands could be produced. Though all these instructions are possible, many combinations of these micro-commands are useless in instructing the processing elements to perform a function. Many of the micro-commands have been grouped to produce most of the standard computer instructions; however, it is possible for a user of the system to make up new instructions by using different combinations of the micro-commands.

III Implementation of the Iterative Array Computer

After formal solicitation of industrial organizations for the fabrication of the Iterative Array Computer, Westinghouse Aerospace Div. was selected as the group to fabricate the computer. Delay lines (Computer Devices Corp.) were used for the accumulator and memory of each PE, where all 200 delay lines have to be time phased. Integrated circuits make up approximately 80 percent of the electronics of the system. The processing elements which are duplicated one hundred times throughout the system are completely integrated circuits except for the indicated lamps which give the status of the mode flip-flop and the read/write amplifier for the delay lines. The integrated circuit used in the system are special logic circuits which were produced by both Motorola and Westinghouse. The resulting system is depicted in Figure 4.

Integrated circuits were used for the logical devices in this system. Both monolithic and chip molecular circuits were used. The circuits consist of six different types of microelectronic devices called MED circuits. The MED circuits are packaged in 10-pin TO-5 cans and each can is capable of dissipating a maximum of 180 milliwatts of power. These circuits employ the STROKE or NAND system of positive logic. To each of the six different types of circuit is assigned a unique MED number. MED 1 and MED 2 are STROKE gates. MED 6, MED 7 and MED 11 are AND gates while MED 3 is a driver stage. A logical 1 corresponds to a 2.2- to 6.0-volt level and a ground to +1.5-volt level corresponds to a logical 0. The difference between the two techniques, monolithic and chip, in the fabrication of molecular devices is illustrated by Figure 5,6.The monolithic element is where the complete circuit is fabricated by different diffusion steps on a single silicon wafer, while the chip techniques consists of many transistors, diodes, and resistors being diffused separately on different silicon wafers, diced and then interconnected into the different circuits within the TO-5 can.

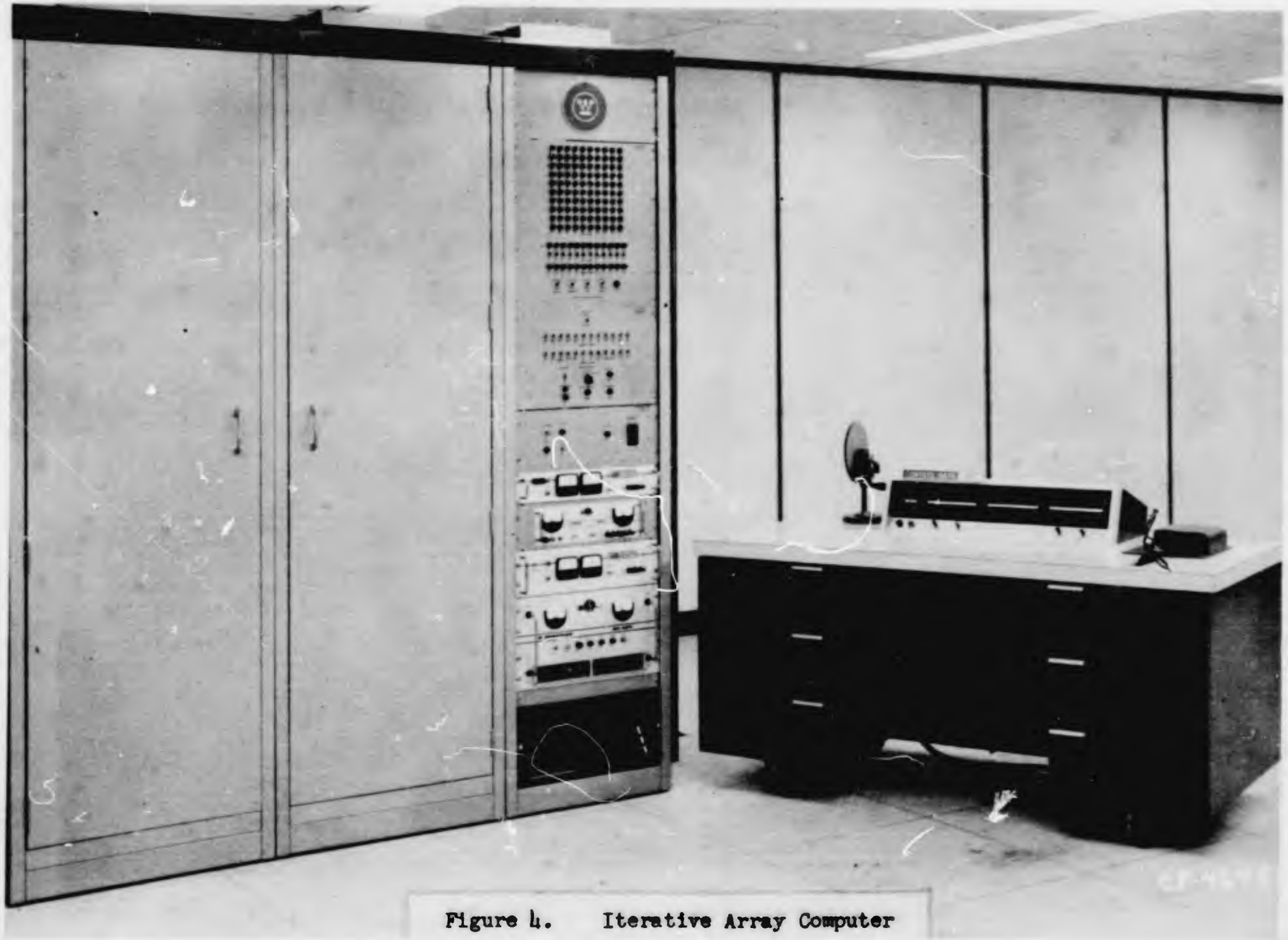


Figure 4. Iterative Array Computer

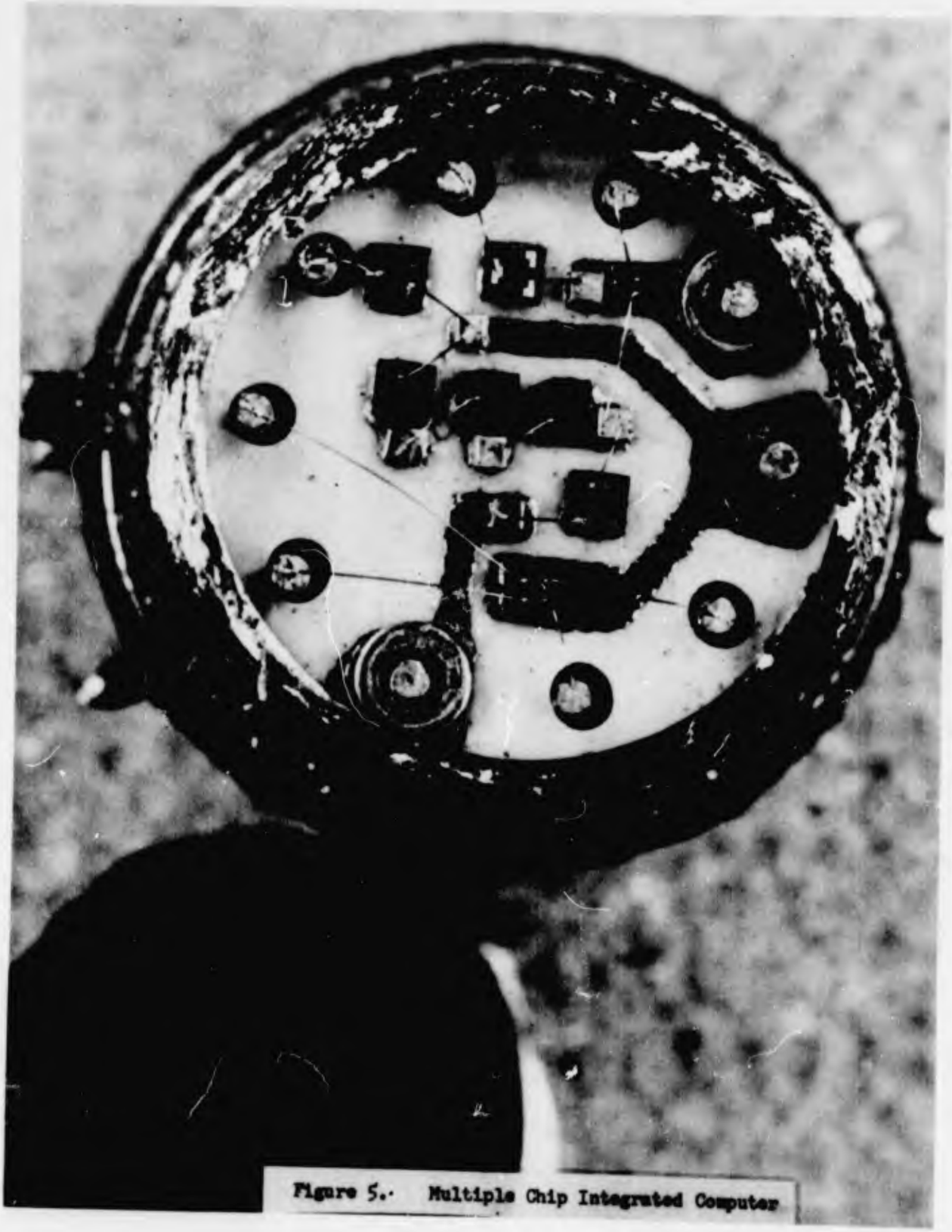


Figure 5.· Multiple Chip Integrated Computer

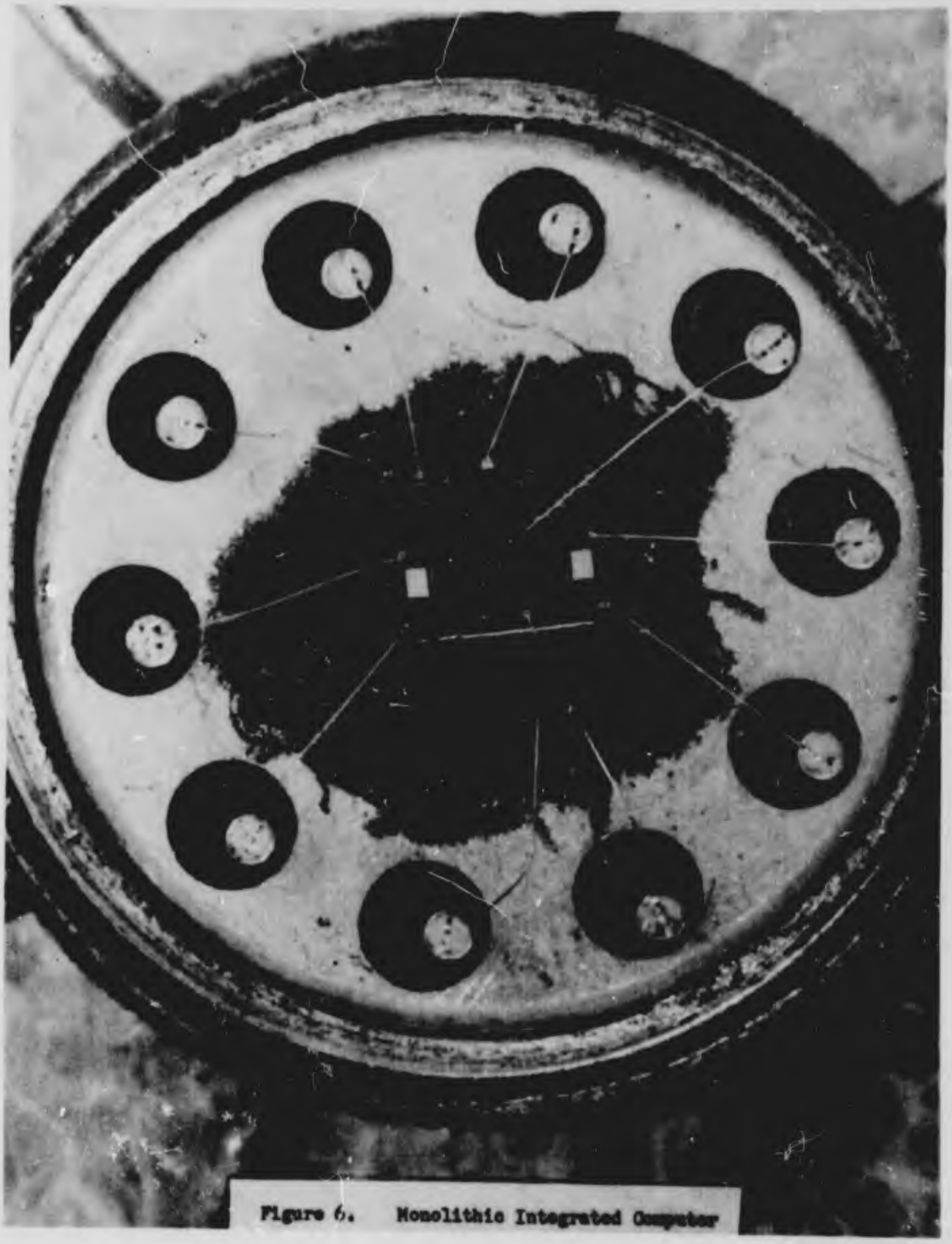


Figure 6. Monolithic Integrated Computer

IV Programming the Iterative Array

The Iterative Array Computer System can be programmed in machine language quite easily if the programmer is familiar with the instruction code and machine language of the CDC 160 computer. This 160 is programmed in machine language exactly the same as any 160 without having the Iterative Array connected. The Iterative Array is selected by the 160 the same way any other peripheral device is selected, with the exception that the special external function code as described earlier is used. Instructions for programming the Iterative Array are described in a report entitled, "A Centrally Controlled Iterative Array Computer (RADCIAC)". Different bit patterns as presented in that report will instruct the Iterative Array to perform different operations.

The input/output channel of the CDC 160 computer has to be set up with an external function code, followed by the steps to output a block of data from the 160. This block of data contains the correct bit patterns to set up instructions within the Iterative Array. Following is an example of how this procedure is programmed within the CDC 160 to set up instructions for the Iterative Array.

Memory Location	Contents	Description
7000	7500	Select
7001	7700	Execute
7002	7303	Output
7003	7---	Last Word Address +1
7004	6502	Jump Back 2
7005	7006	First Word Address
7006	0042	Store a 1 in the accumulators
7007	0000	of all PE's
7010	0500	Store a 1 in memory location
7011	0002	02 of each PE
7012	.".	Any length of PE instruction
.".	.".	can be given

An assembler system called RADICAL has been developed for the system which uses a special symbolic representation for CDC 160 instruction. Any symbol can be defined within the assembler system; therefore, it is still possible for the programmer to make up his own special instruction for the Iterative Array. The programmer has to give a name to the new instruction and define the correct micro-command for the instruction within the symbol tables of the assembler system. The standard instructions for the array do not have to be defined. This assembler system uses the Iterative Array while it is doing the assembling; therefore it would be impossible to use the language on any CDC 160 unless the Iterative Array is connected. The set up of the input/output line of the 160 by the external function code for the Iterative Array is automatically taken care of; therefore, to perform the program given in the previous paragraph using the assembler system, it would appear as follows:

```
Ţ LDO Ţ C
Ţ STO Ţ 02 C
```

After passing through the assemble system the machine code as shown previously would be prepared for the machine to perform the operations.

V Application of The Iterative Array Computer

There are two fundamentally different strategies involved in the assignment of missiles to a target complex in time of national emergency, namely:

1. A sequential assignment of missiles to a target complex,

2. Entire missile battery assignment to the target complex.

In (1) the strategy is to allow time to evaluate the bomb damage before the next missile or groups of missiles are launched. However, if the total expected engagement time is less than the time required to evaluate the destruction of a target, then obviously strategy (2) should be employed. The problem this part of the paper addresses itself to involves strategy (2). However, the same problem essentially exists when the strategy under (1) consists of sending major groups of missiles in waves where the assignment of each wave, after the first, depends upon the damage evaluation for the proceeding wave. For instance, if there were initially 900 missiles, it might be decided to send 300 missiles each in three waves.

This weapons assignment problem can be couched within the problem of the general type commonly known as the transportation problem first formulated by F.L. Hitchcock[7] in 1941. Independently, during World War II, T.C. Koopmans[8] also formulated this same problem. Consequently, it is also often referred to as the Hitchcock-Koopmans transportation problem. Both men suggested a procedure for solving this problem similar to the linear programming simplex method.

There are two variations on the Hitchcock-Koopmans transportation problem which have been termed:

1. The assignment problem.

2. The transportation problem.

As will become evident below, the assignment problem is only a special case of the more general transportation problem. In each case, there is a rating matrix (a_{ij}), as illustrated in Figure 7, where the rows of the matrix correspond to initial surpluses of materials, like missiles for the case of concern, and the columns correspond to initial shortages, like target requirements for the case of concern. The elements, a_{ij}, of the rating matrix represent costs or values associated with sending each of the surpluses to each of the shortages. In Hitchcock's problem as he originally stated it, the row designators represented warehouses with inventory surpluses and the column designators represented stores with inventory requirements. The rating matrix represented costs associated with shipping the surpluses from each of the warehouses to each of the stores. The problem was to find the assignment of the inventory from each of the warehouses to each of the stores, as represented by an assignment matrix (x_{ij}), such that the total shipping cost was minimized, i.e.,

Targets / Missiles	Storage #1	Storage #2	Storage #n
Surplus #1	a_{11}	a_{12}	a_{1n}
Surplus #2	a_{21}	a_{22}	a_{2n}
Surplus #m	a_{m1}	a_{m2}	a_{mn}

The problem statement

find an assignment matrix $\left[x_{ij} \right]$

such that

$$\text{maximum or minimum} \left[\sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij} \, x_{ij} \right]$$

Figure 7. Example of Rating Matrix

$$\text{Minimum} \left[\sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij} x_{ij} \right]$$

In the case of the weapons assignment problem, the rating matrix can be thought of as a value matrix where:

$$a_{ij} = W_j D_{ij} P_{ij}$$

and

W_j = Worth or value assigned for each target "j"

D_{ij} = Destruction capability for each target "j" by each missile "i"

P_{ij} = Probability of each missile "i" reaching each target "j"

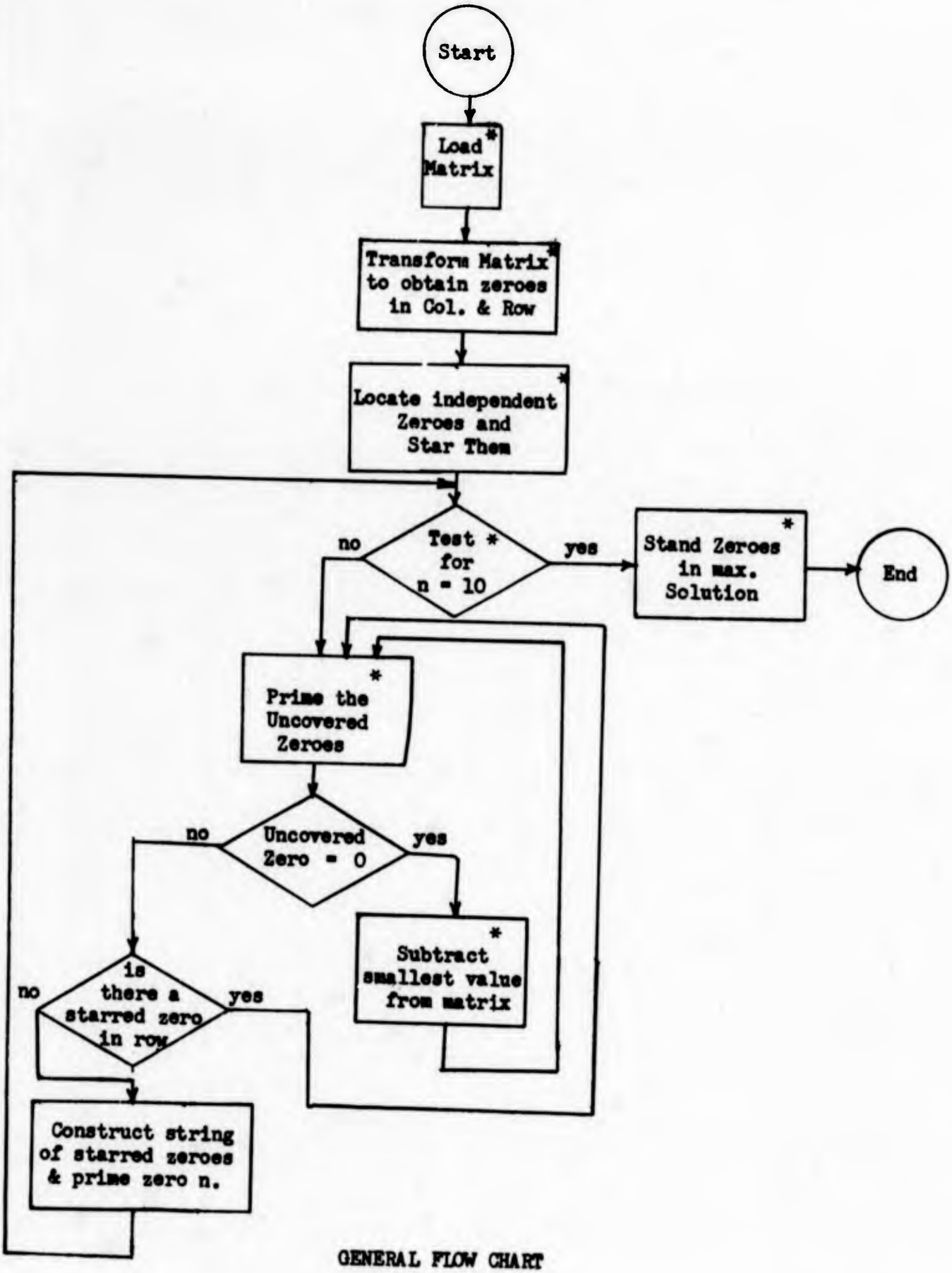
a_{ij} = Expected worth of each missile "i" for each target "j"

Here the problem is to determine the assignment matrix (x_{ij}), so that the total expected value is maximized, i.e.,

$$\text{Maximum} \left[\sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij} \quad x_{ij} \right]$$

This algorithm which was used by the author was developed by James Munkres and presented in a paper entitled "Algorithm for Assignment and Transportation".[9]

This algorithm was used to solve a n = 10 matrix on the Iterative Array Computer. There were many steps which could be done in parallel; however, with the extremely limited directly distributed control within the array more steps were required. The flow charts for the program follow. The blocks marked (*) are at least partially performed within the array and have some parallelism in them.



GENERAL FLOW CHART

VI Conclusions and Recommendations

The Iterative Array was designed to be an experimental tool with which interested researchers could obtain some experience with parallel processing, and to prove that increase in data processing can be obtained through parallel processing. The processing elements which are reproduced one hundred times were designed so that they could be fabricated at the least possible expense and still have an operable general system. Distributed control capability is limited to the "S" mode flip-flop which can assume one of two states. In the application of the system to the assignment problem, this highly restricted distributed control capability was extremely noticeable in programming and in the systems efficiency. If the system had more states for mode control and some external control similar to geometric control in the SOLOMON computer, it would have been easier to program and more efficient in the solution of problems. Also, more parallelism could have been realized because the CDC 160 computer had to handle many of the manipulations which could have been done within the array if it had the extra distributed control described above.

Most of the mathematical solutions for problems which exist today were generated so that the problems could be solved by humans or by general purpose computers which are on the market today. Both humans and general purpose computers perform operations sequentially (only one operation at a time). When these techniques are used on a parallel processor, it is difficult to take full advantage of the parallelism since the data for one part is often dependent on the output of another part. The basic mathematical techniques for the solution of problems should be examined to find which existing techniques are more applicable to parallel processing and where new mathematical techniques should be developed for parallel processing.

The ILLIAC IV (being developed at the University of Illinois under RADC direction and ARPA funding) will be the second generation parallel processor which will have much more extensive distributed control and much faster array of processor units. The speed goals of the system are 10^8 floating point operations per second and 10^8 fixed point character operations per second.

1. "Historical Monograph Electronic Computers Within the Ordnance Corps" Aberdeen Proving Grounds, Md.

2. "Proposed Automatic Calculating Machine", Howard Aiken, IEEE Spectrum August 1964.

3. John von Neumann Collected Work, Vol. V, "Design of Computers, Theory of Automata and Numerical Analysis", A Pergamon Press Book, The Macmillan Co, N.Y. 1963

4. "A Universal Computer Capable of Executing an Arbitrary Number of Sub-Programs Simultaneously", John Holland, 1957, Proceeding of the Eastern Joint Computer Conference, pps 108-113.

5. "The SOLOMON Computer", Workshop on Computer Organization, Proceedings of the 1962 Workshop on Computer Organization, Edited by A.R. Barnum and M.A. Knapp, Spartan Books, Inc., pps 66-92.

6. For more detail on the programming of the Iterative Array refer to RADC-TR-64-251, "A Centrally Controlled Iterative Array Computer", obtainable from DDC, AD 605 370.

7. Hitchcock, F.L., "The Distribution of a Product from Several Sources to Numerous Localities", Journal of Math and Physics, Vol 20, 1941, pps 224-330.

8. Koopmans, T.C., "Optimum Utilization of the Transportation Systems" Proc. of the Intelligence Statistical Conference, 1947, Wash, D.C. (later reprinted as Supplement to Econometrica, Vol 17, 1949)

9. Munkres, James, "Algorithms for the Assignment and Transportation Problems"

VALIDATING VELA SATELLITE ORBIT DETERMINATION PROCEDURES USING OBSERVED ECLIPSE ENTRANCE TIMES

BY

Charles B. Huelsman III

Lieutenant United States Air Force

Air Force Satellite Control Facility

Lt Charles B. Huelsman III

BIOGRAPHICAL HISTORY

Lieutenant Charles B. Huelsman III attended Miami University (Ohio) graduating with an A.B. in 1963, majoring in Mathematics. There, he wrote an article on Infitie Series, "Raabe's Test", which was published in the Spring 1965 issue of the Pi Mu Epsilon Journal. During 1962-3 he was Chairman of the Interreligious Relations Committee of the National Newman Club External Affairs Department. He was Program Chairman of the 1963 Eastern Regional Interreligious Relations Conference. From September of 1963 until February of 1964 he was a graduate assistant in the Department of Mathematics at The Ohio State University. In July of 1964 he was commissioned, and in December of 1964 he received his M.S. in Mathematics from The Ohio State University. The first two chapters of this thesis, "Tauberian Theorems", were published in the Autumn 1965 issue of the Pi Mu Epsilon Journal under the title "Tauberian Theorems for Abel Summability."

Lieutenant Huelsman is presently stationed at the Air Force Satellite Control Facility (SSD) in Sunnyvale, California. There he specializes in orbit determination and has published an enlarged version of his presented paper: SSOTU Technical Memo, "Validating VELA Vectors Using Observed Eclipse Entrance Times." He has served as the President of the Catholic young adult group near Sunnyvale. He has been selected by AFIT for a Ph.D. program in mathematics. Presently he is serving a temporary tour of duty at Andrews Air Force Base, Washington, D. C., with the Fundamental Space Operations Study Group, SCLS/SPAD. The group is studying the future role of the Air Force in space.

ABSTRACT

VALIDATING VELA SATELLITE ORBIT DETERMINATION PROCEDURES USING OBSERVED ECLIPSE ENTRANCE TIMES

The validation of orbit determination calculations has been constantly improved by the VELA program. One suggested validation standard was the observed time the satellite entered the shadow of the earth. This time can be predicted on the computer using the results of an orbit determination procedure; that is using the resulting vector. Theoretically, a vector is valid if $/t/ < c/R/+k$, where t is the time difference, R is the number of revolutions elapsed since the eclipse for the vector in question, and c and k are constants. To test these guidelines and determine the constants c and k several eclipse predictions were made resulting in 248 (R,t) situations. The guidelines were found to be lacking. It was discovered that the data had not been corrected for the effect of the atmosphere on the shadow of the earth. Atmospheric corrections were calculated and the data was corrected. Validation guidelines were calculated and suggestions for improving the eclipse prediction program were made.

CONTENTS

ILLUSTRATIONS

TABLES

LIST OF MAJOR SYMBOLS

a	mean radius of the earth
c	constant used to describe the maximum t'
h	height of the atmosphere refracting sunlight, 100km
k	constant used to describe the maximum t'
m	mean t' over an interval on the R axis
p	the time the vehicle is in the shadow of the earth
R	$R_1 - R_o$, the number of revolutions a vector v is away from an eclipse at revolution R_o
R_o	revolution during which the eclipse took place
R_1	nodal revolution ot a vector v
t	t_1-t_o, eclipse time prediction error initially attached to the vector v
t'	t-x, error t corrected for the atmospheric effect
t_o	observed eclipse time
t_1	predicted eclipse time using vector v
T	maximum allowable eclipse time prediction error
T_m	lower bound of the maximum allowable eclipse time prediction error t'
T_M	upper bound of the maximum allowable eclipse time prediction error t'
u	sigma for t' over an interval on the R axis
v	given vector at rev R_1, predicting an eclipse entry time of t_1
v_o	vector at eclipse time, calculated using vector v
x	atmospheric correction for t

GLOSSARY

KECLIPSE	The computer program used to predict eclipses
KUPDATE	The computer program which calculates a vector of a satellite at a new time, from a previously given vector
MEDIAN	50th percentile
MODE	Interval of greatest population
VECTOR	The satellite's position (x, y, z) and its velocities ($\dot{x}$, $\dot{y}$, $\dot{z}$) in the respective directions at a given time.

SECTION 1

INTRODUCTION

1. The determination of vector accuracy for high altitude satellites has bee.. constantly improved. Initial vector validation procedures had inaccurate foundations. From a given vector the time of nodal crossing was calculated. With a second vector, the same time was approximated. If the nodal crossing times were similar, both vectors were considered accurate. If the nodal crossing times differed by more than five minutes, one of the vectors was considered inaccurate. Which was acceptable was open to argument. This provided an initial method for judging the accuracies of a new vector without knowing the validity of the original vector.

2. A second method of judging a vector was to compare its properties with those of the previous vector; however, some of these properties varied with the positions of the moon and the sun. This method again compared potentially inaccurate vectors with each other. There were no assurances that one vector was better than another. These methods were sufficient for calculating vehicle acquisition information, but required refinement if individualistic interpretations of the data were to be eliminated and/or accurate eclipse times were required.

3. In view of these problems an improved method of vector validation was sought. It was known that as the satellite entered the shadow of the earth the penumbral entry time could be accurately found by observing the battery voltage and solar array output readings. With a given vector, this entry time could also be predicted on a computer. With the observed eclipse time as a standard, a reliable validation procedure was established.

4. A method for comparing the actual and predicted shadow entry times is already in use but requires further theoretical development. It is not known how many minutes of error are allowed before the new vector generates unusable station acquisition information. It is the objective of this study to develop criterion for judging vector reliability.

SECTION II
THEORY

1. If for a given vector v the predicted eclipse time t_1 is close to the observed eclipse time t_o (see Table 1), then v is an acceptable vector. If $t = (t_1 - t_o)$ is large in magnitude, then the vector is inaccurate. The natural question is how large can t be before the vector becomes unusable. It is the thesis of this study that the maximum allowable error T is directly proportional to the difference R, between the nodal revolution R_1 of the vector v and the revolution R_o during which the eclipse took place (Figure 1); i.e.,

$$R = R_1 - R_o \text{ and } |T| = c|R| + k \qquad (1)$$

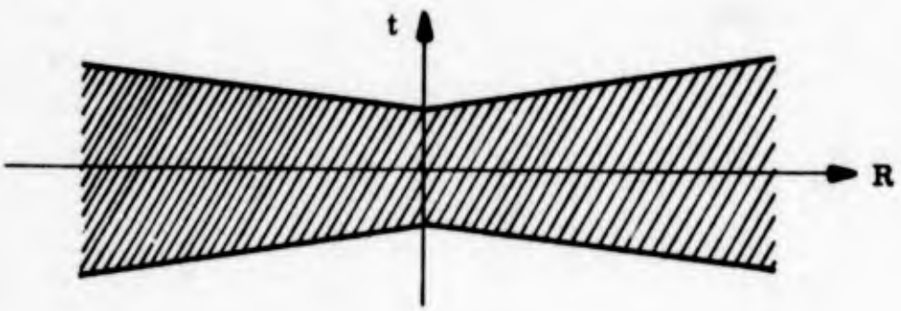


Figure 1 Theoretic Cone of Validity

It would be easier to set the allowable margin of error T at some constant, say five minutes. But this method ignores errors present in the General Purpose Tracking Programs (GPTP) that must be considered. To make an eclipse prediction for a given vector v , v must be integrated with the computer program KUPDATE (Ref. 1) to the eclipse time. KUPDATE yields a natural amount of error in calculating the eclipse entry vector v_o. The greater the number of revolutions R to be integrated over, the more significant this internal error becomes.

2. Admittedly, the relation [Eq. (1)] might actually be a quadratic or a polynomial of higher degree. Thus it must be judged from the data which kind of equation would be justified.

Table 1

OBSERVED VELA EARTH ECLIPSE PENUMBRAL ENTRANCE TIMES t_o

Vehicle No.	Rev	Date (day)	(month)	(yr)	Time (ZULU) (hr)	(min)	(sec)
1801	118	1	Apr	65	1	9	37
	119	5	Apr	65	13	37	30
	160	5	Oct	65	0	5	53
1851	118	31	Mar	65	15	34	38
	119	5	Apr	65	4	22	-
	160	5	Oct	65	10	46	56
3662	62	3	Apr	65	13	5	-
	107	6	Oct	65	14	3	40
	108	10	Oct	65	19	7	-
3674	63	4	Apr	65	8	0	-
	108	6	Oct	65	12	32	3
	109	10	Oct	65	17	24	20
6564	20	17	Oct	65	7	14	56

SECTION III

EXPERIMENTAL METHOD

A. INTRODUCTION

1. All vectors generated from 22 March 1965 to 6 February 1966 for normal VELA Program operations on Launches I and II and for Vehicle 6564 were considered. Using tracking data from normal operations, more vectors were generated and considered together with the above data.

2. For the purposes of this study the differences in the orbits of the five vehicles were considered negligible. The eccentricities varied from 0.006 to 0.114. The maximum altitudes varied from 62,135 to 66,453 nautical miles. The integrated periods varied from 6000 to 6714 minutes. While the inclinations to the equatorial plane varied from 34.4 to 41.0 degrees, and the right ascensions varied between 194.1 and 213.4 degrees. The values R and t were calculated for all the vectors and plotted on a t vs R graph (Figure 2). A search was made for the cone described in Figure 1. From experience it was known that a majority of the VELA vecotrs were accurate. Theoretically, the data points (R, t), should cluster into a cone with the R axis as its center. Noting Figure 2, most of the points were above the R axis, indicating that most of the predicted times came after the observed time. A statistical graph, (Figure 3), was constructed. It possessed a mode (interval of greatest population) of 1 minute 15 seconds. All data within 6 minutes 15 seconds of the mode formed a near normal curve. The data within this interval had a median of 1 minute 52 seconds, indicating that the data was biased. Theoretically the mode and the median should be 0 seconds.

3. Research by Gersten (Ref. 2) indicated that the atmosphere of the earth causes the shadow of the earth to be larger than assumed in the computer eclipse prediction program, KECLIPSE (Ref. 3). This program computes the position of the earth's shadow as pictured in Figure 4. The KECLIPSE program ignores the refraction of light caused by the atmosphere of the earth. The difference is seen in a comparison of Figure 4 with Figure 6. Gersten computed a minimum correction of 52 seconds. Realizing more accurate data could be obtained with atmospheric corrections, the following theory was developed.

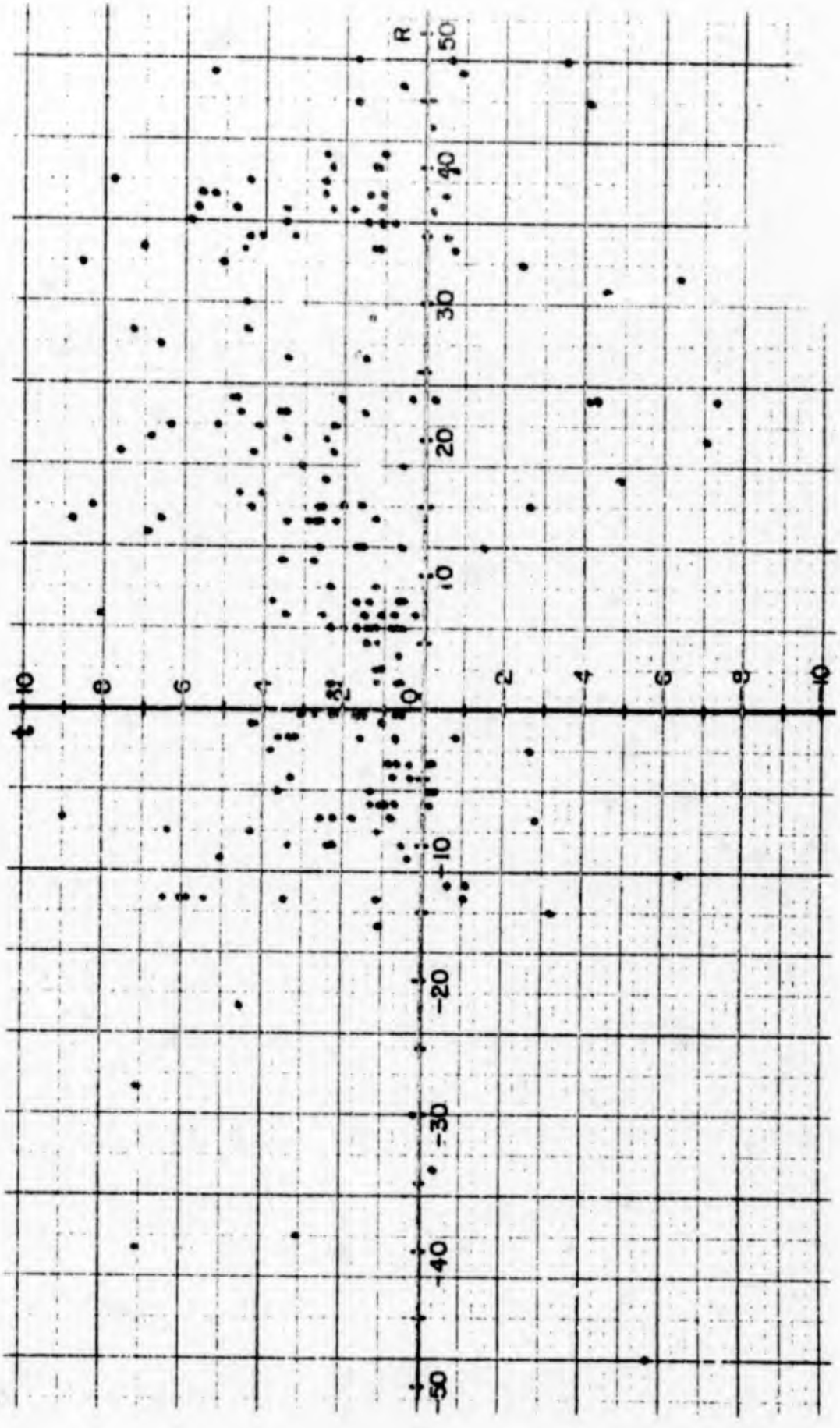


Figure 2. t vs R Graph

Time Interval		Number
7' 31" -	DDDDDDDDDDDDDDDDDDDDD	21
7' 1" - 7' 30"	DDDD	4
6' 31" - 7' 0"	DDDDDD	6
6' 1" - 6' 30"	DDD	3
5' 31" - 6' 0"	DDDDD	5
5' 1" - 5' 30"	DDDDD	5
4' 31" - 5' 0"	DDDDDD	6
4' 1" - 4' 30"	DDDDDDDDDDDDDD	14
3' 31" - 4' 0"	DDDDDDDDDDDDDD	14
3' 1" - 3' 30"	DDDDDDDD	8
2' 31" - 3' 0"	DDDDDDDDDDDDDDD	15
2' 1" - 2' 30"	DDDDDDDDDDDDDDDDDDDDD	21
1' 31" - 2' 0"	DDDDDDDDDDDDDDDDD	17
1' 1" - 1' 30"	DDDDDDDDDDDDDDDDDDDDDDDDDDDDD	30
0' 31" - 1' 0"	DDDDDDDDDDDDDDDDDDDDD	21
0' 0" - 0' 30"	DDDDDDDDDDD	11
-0' 30"- -0' 1"	DDDDDDDDD	9
-1' 0"- -0' 31"	DDDDDDD	7
-1' 30"- -1' 1"	DD	2
-2' 0"- -1' 31"	DD	2
-2' 30"- -2' 1"	D	1
-3' 0"- -2' 31"	DDD	3
-3' 30"- -3' 1"	D	1
-4' 0"- -3' 31"	DD	2
-4' 30"- -4' 1"	DDD	3
-5' 0"- -4' 31"	DD	2
-5' 30"- -5' 1"		0
-6' 0"- -5' 31"	D	1
-6' 30"- -6' 1"	DD	2
-7' 0"- -6' 31"		0
-7' 30"- -7' 1"	DD	2
- -7' 31"	DDDDDDDDDD	10

Number of t's	248
Mode	1' 15"
Number of t's within 6' 15" of Mode	212
Resulting Median	1' 52"

Figure 3. Statistical Graph of All t's

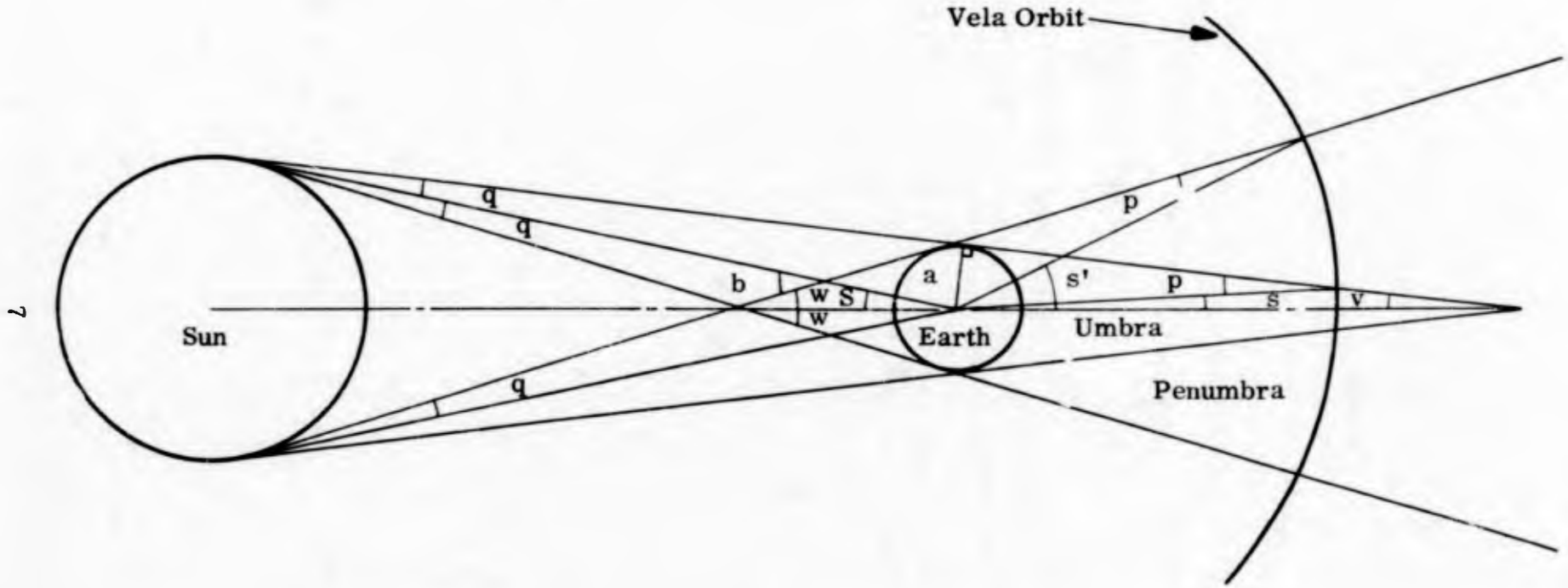


Figure 4. Sun-Earth Eclipse Geometry

B. ATMOSPHERIC CORRECTIONS

1. The computer program KECLIPSE uses the diagram shown in Figure 4 as a model for the sun-earth eclipse geometry. Considering various triangles: $p = s + v$; $S = q + v$; $s' = p + w$; $b = w + S$; and $2w = b + q$. From the last two equations $w = S + q$ and thus

$$s' = p + q + S \tag{2}$$

$$s = p + q - S \tag{3}$$

$$s' - s = 2S \tag{4}$$

2. Letting a equal the average radius of the earth, and R'' equal the distance from the vehicle to the center of the earth, it follows that

$$p = \sin^{-1}(a/R'') \tag{5}$$

By considering the average distance between the earth and the sun, it is easily calculated that $q = 0.15'$ and $S = 16'$. Thus $s' - s$ is always close to $32'$.

3. With reference to Figure 5, arc V is a given satellite path through the shadow, while arc O is a path through the center of the umbra. The amount of time needed for a satellite to traverse each of the indicated distances W, X, Y and Z is represented by w, x, y and z respectively. Since the velocity of the satellite on arc O is arbitrary, it may be assumed that this satellite's velocity is equal to the velocity of the satellite on arc V. For a given eclipse, w was obtained from the eclipse prediction printout. The time z was obtained by dividing $s' - s$ by the instantaneous angular velocity at point A, $\dot{s}$. The time y was found through the more complex methods explained below.

4. By extending previously developed lunar eclipse theory (Ref. 4), Gersten (Ref. 2) developed a method of calculating the s'' shown in Figure 6. By constructing a shadow density function, with an accurate, complex method, the following equation was found:

$$s'' = 74.0' + (p - 57')\left(1 + \frac{h}{a}\right)$$

where h is 100 km, the maximum height of atmosphere affecting the shadow, and a is again the average radius of the earth. Thus y was found by dividing $s'' - s'$ by $\dot{s}$. The angle s' was obtained using Eq. (2). As both s'' and s' were dependent on p, individual p's were calculated for each eclipse.

5. By similar triangles

$$\frac{X}{Y} = \frac{W}{Z}$$

and

$$\frac{x}{y} = \frac{X\dot{s}}{Y\dot{s}} = \frac{W\dot{s}}{Z\dot{s}} = \frac{w}{z}$$

thus

$$x = \frac{wy}{z}$$

If ϕ was sufficiently small, the atmospheric correction calculated above was accurate. If ϕ exceeded 10 degrees, other means were used.

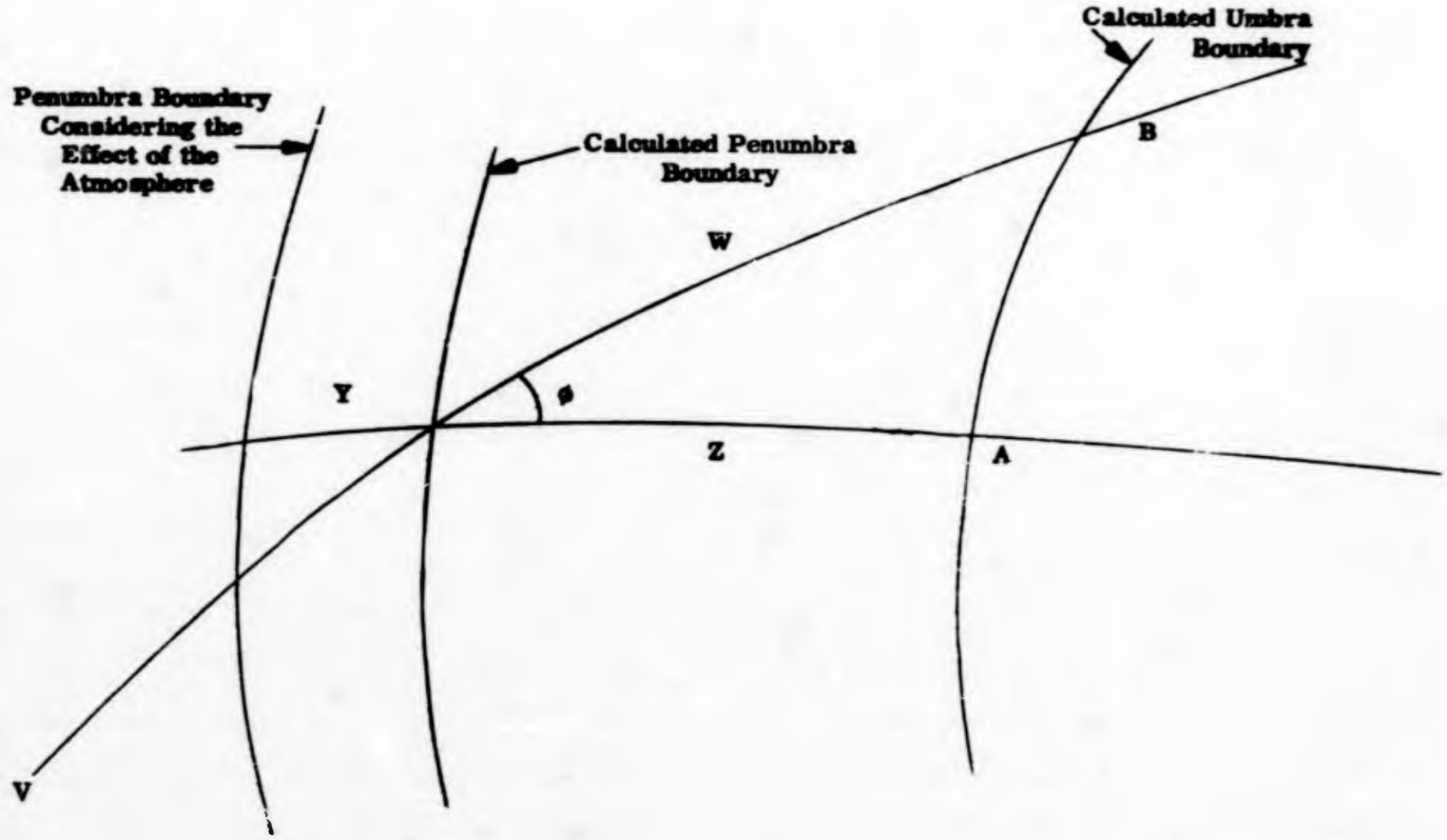


Figure 5. Shadow Cross-Section with the Atmospheric Effect

6. Vehicle 1351 had passed through the earth's shadow on two consecutive revolutions during the March-April eclipse season. The angle ϕ for the 5 April 1965 eclipse was small enough to ensure an accurate atmospheric correction. The 31 March 1965 eclipse traversed only the penumbra. In this case ϕ was nonexistent. The two eclipses, being on consecutive revolutions, were comparable. As the velocities during each eclipse were approximately equal, the distances shown in Figure 7 were assumed to represent time. From Figure 7:

$$\frac{r}{y} = \cos a = \frac{\frac{1}{2}P}{R'}$$

and

$$\frac{r}{x} = \cos b = \frac{\frac{1}{2}p}{R'}$$

$$x = \frac{R'r}{\frac{1}{2}p} = \frac{\frac{1}{2}R'P}{\frac{1}{2}R'p} \quad y = \frac{Py}{p}$$

7. For a vehicle where there was no comparable eclipse, the following procedure was used. Using the two right triangles in Figure 8:

$$B^2 = (R')^2 - A^2 = D^2 - (A + X)^2$$

$$X = \sqrt{D^2 - (R')^2 + A^2} \quad - A$$

Without loss of generality it may be assumed that the velocities along all three of the paths were the same. Letting x, d, r and a represent the respective times needed to traverse X, D, R' and A,

$$x = \sqrt{d^2 - r^2 + a^2} \quad - a$$

8. The data from Section IV was altered and another statistical graph was composed as Figure 9,

$$t' = t - x$$

The mode of this graph was 15 seconds, meaning that to a limited extent the data was statistically healthy. The median however was 36 seconds, meaning that biases were still present.

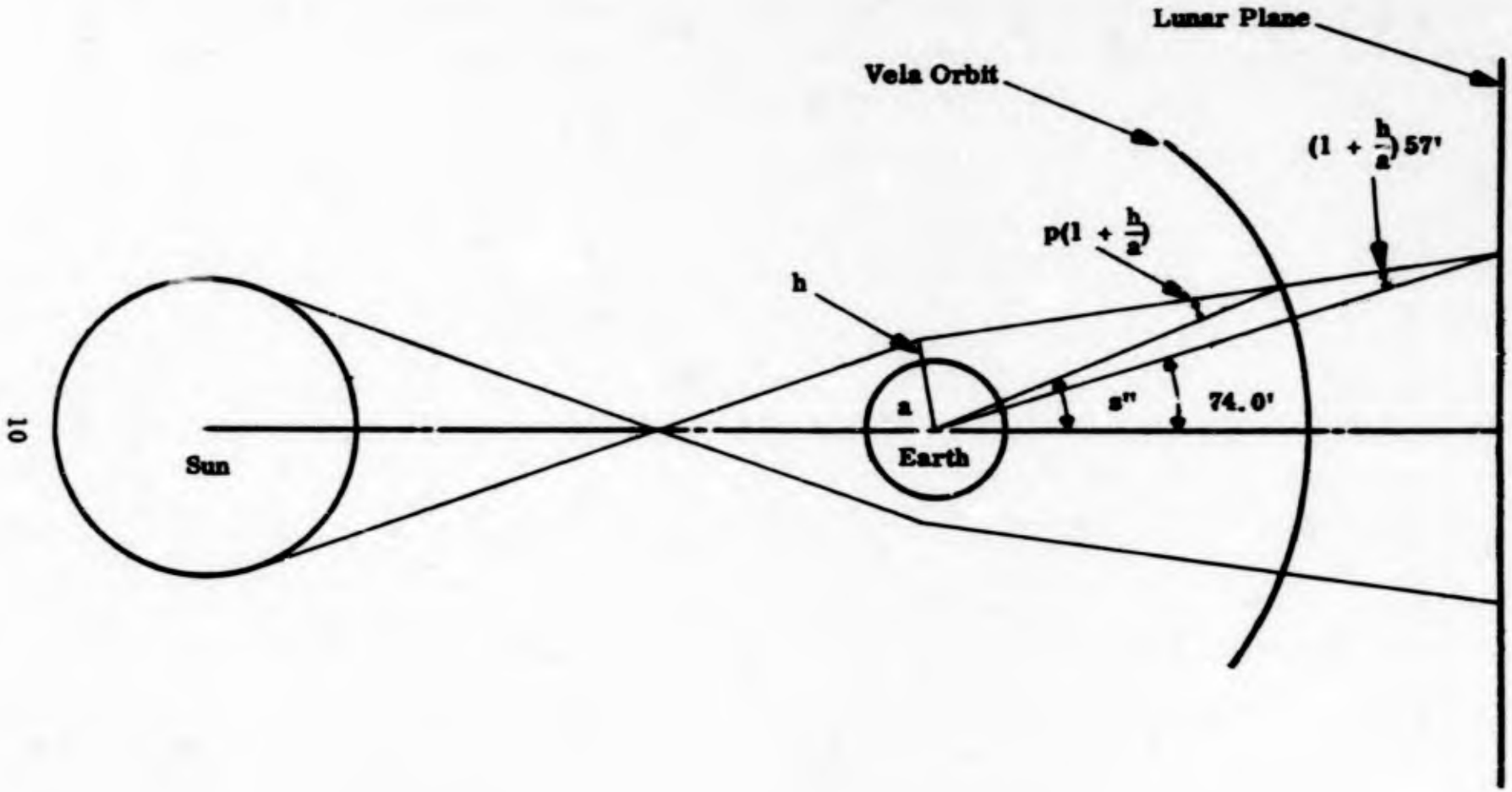


Figure 6. Sun-Earth Eclipse Geometry with the Atmosphere

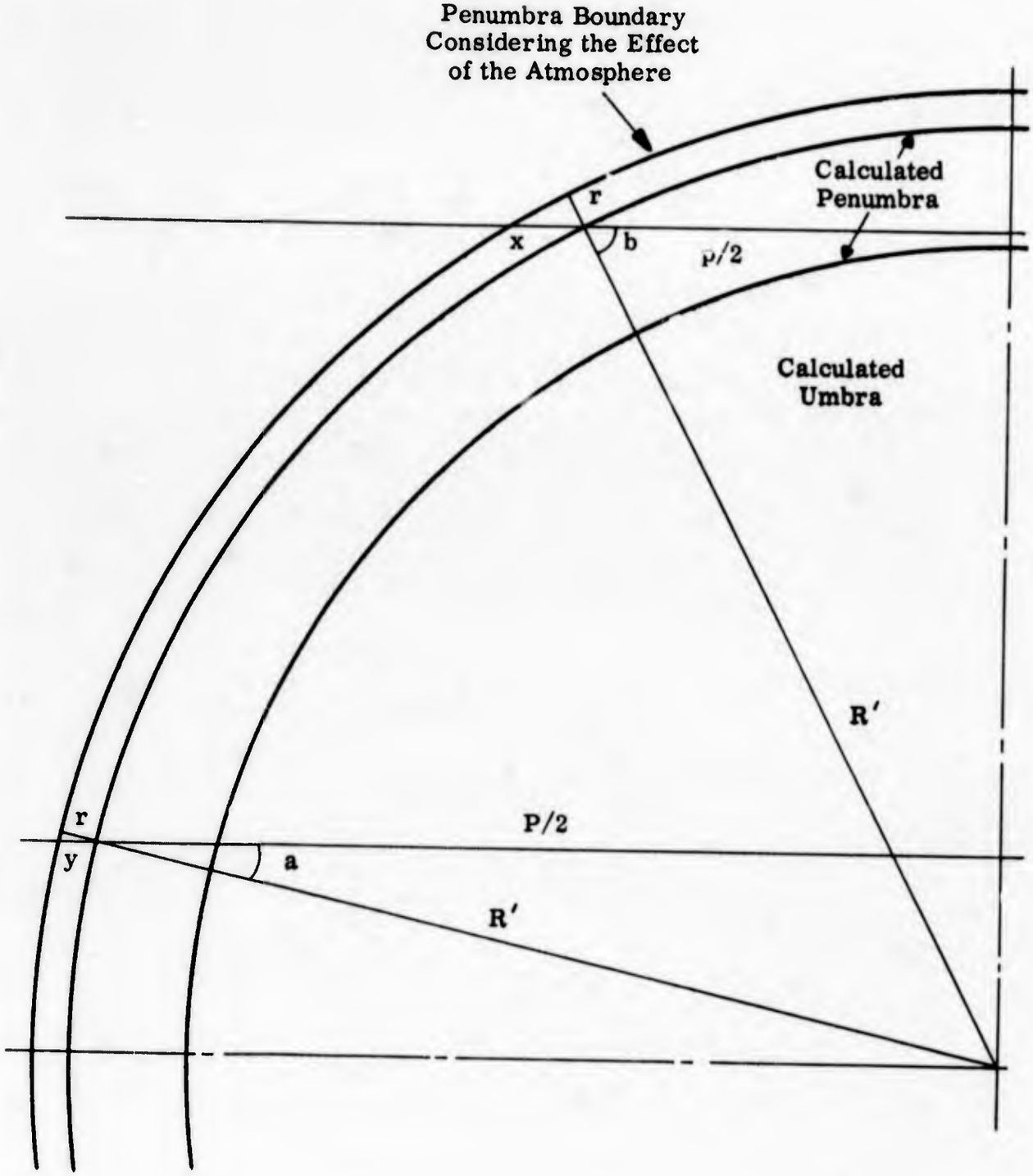


Figure 7. Earth Shadow Geometry I

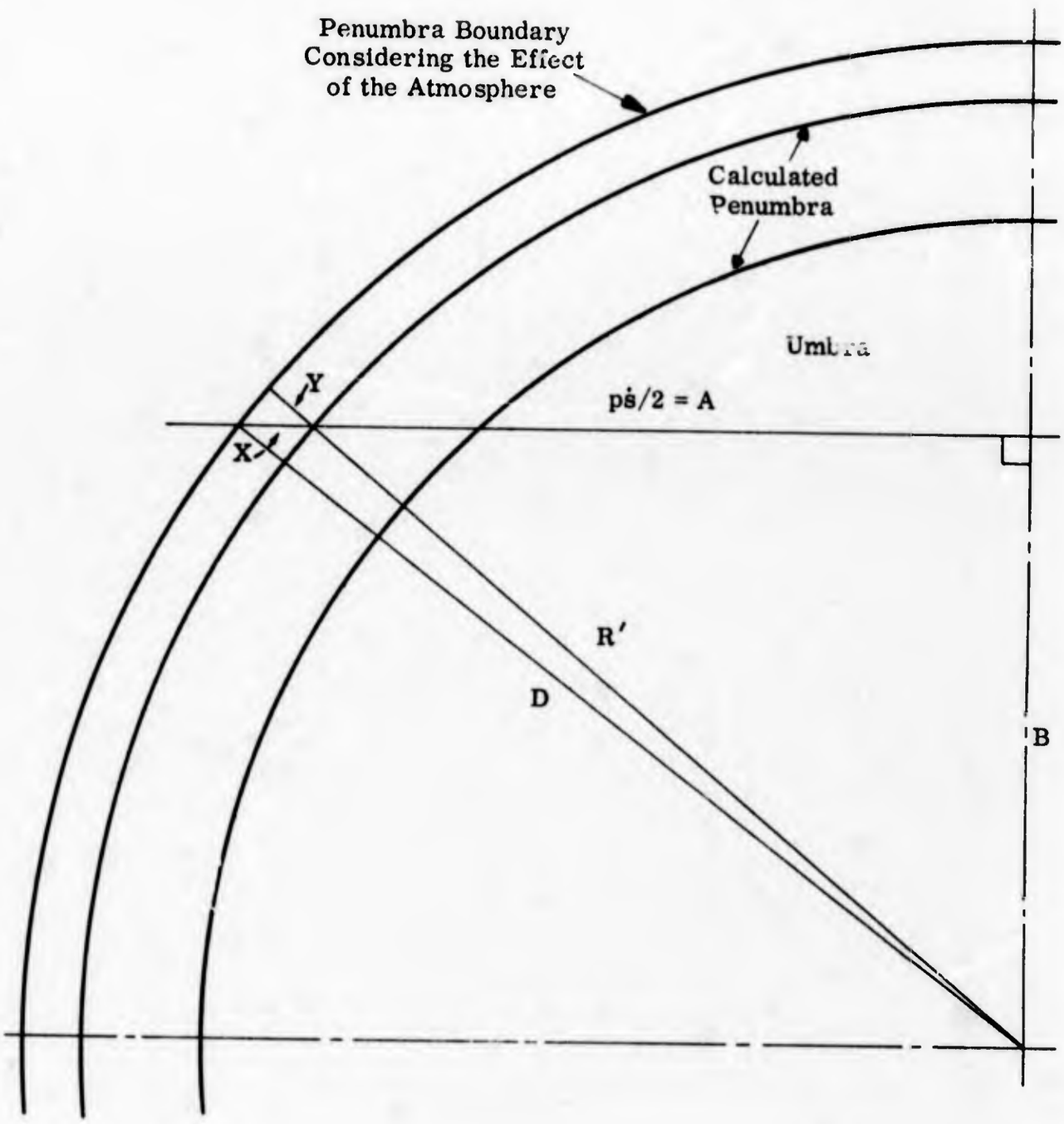


Figure 8. Earth Shadow Geometry II

Time Interval						Number
7'	31"	-			DDDDDDDDDDDDDDD	15
7'	1"	-	7'	30"	DD	2
6'	31"	-	7'	0"	D	1
6'	1"	-	6'	30"	DDDD	4
5'	31"	-	6'	0"	DDD	3
5'	1"	-	5'	30"	DDDDDDD	7
4'	31"	-	5'	0"	DDD	3
4'	1"	-	4'	30"	DDDD	4
3'	31"	-	4'	0"	DDDDDDDDD	9
3'	1"	-	3'	30"	DDDD	4
2'	31"	-	3'	0"	DDDDDDDDDDDDDD	14
2'	1"	-	2'	30"	DDDDDDDDDDDDD	13
1'	31"	-	2'	0"	DDDDDDDDDDDDDDDD	16
1'	1"	-	1'	30"	DDDDDDDDDDDDDDDDDD	18
0'	31"	-	1'	0"	DDDDDDDDDDDDDDD	15
0'	0"	-	0'	30"	DDDDDDDDDDDDDDDDDDDDDDDDDDDDDD	30
-0'	30"	-	-0'	1"	DDDDDDDDDDDDDDDDDD	18
-1'	0"	-	-0'	31"	DDDDDDDDDDDDDDDD	16
-1'	30"	-	-1'	1"	DDDDDDDDDDDDD	13
-2'	0"	-	-1'	31"	DDDDDDD	7
-2'	30"	-	-2'	1"	DDDD	4
-3'	0"	-	-2'	31"	DDDD	4
-3'	30"	-	-3'	1"	D	1
-4'	0"	-	-3'	31"	DD	2
-4'	30"	-	-4'	1"	DD	2
-5'	0"	-	-4'	31"		0
-5'	30"	-	-5'	1"	DDDD	4
-6'	0"	-	-5'	31"	DDD	3
-6'	30"	-	-6'	1"		0
-7'	0"	-	-6'	31"	DD	2
-7'	30"	-	-7'	1"	DD	2
		-	-7'	31"	DDDDDDDDDDDD	12

Number of (t')'s	248
Mode	0' 15"
Number of (t')'s within 6' 15" of Mode	214
Resulting Median	36"

Figure 9. Statistical Graph of all (t')s

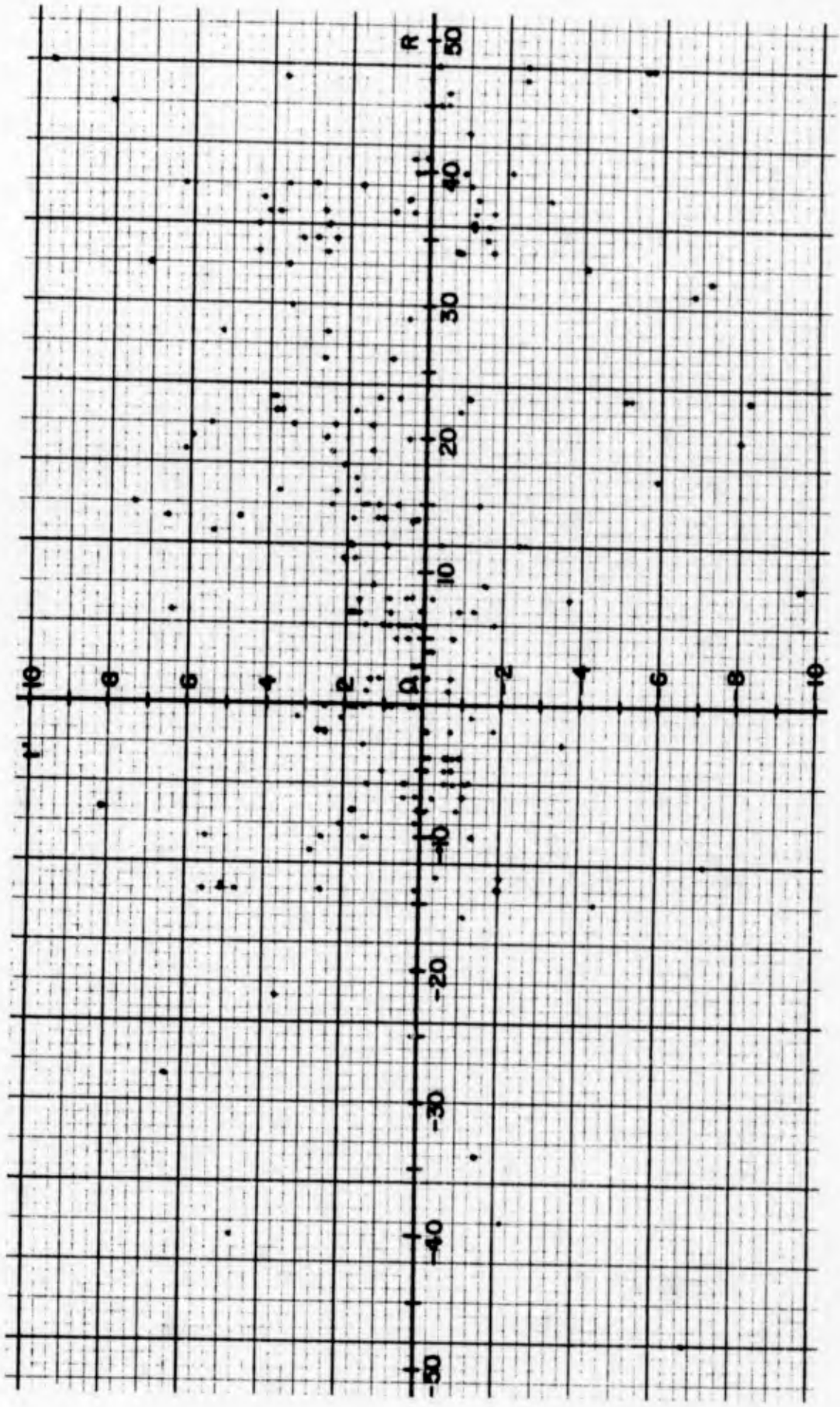


Figure 10. t' vs R Graph

SECTION IV

EXPERIMENTAL ERROR

1. The possibility of errors occurring in the data is not negligible. The error due to KUPDATE has already been noted. It is caused by the approximation errors within the numerical integration process, and by integrating a vector around an inaccurate model of the earth. It was theorized that this error would increase proportionately with R. This is verified by noting the breadth of each cluster of data in Figure 10. From experience, it was known that the launch vectors of Vehicle 6564 were not more than 5 minutes off when used to predict ephemeris within one revolution of their time of epoch; however these vectors were hours off when projected 16 to 18 revolutions to predict an eclipse. Thus, the usually larger errors of launch vectors were magnified irregularly by KUPDATE. So, for these vectors, Eq. (1) does not hold; but the smaller errors of vectors, based on 300 points spread over 10 or more revolutions of data, and an established vector, as shown in Figure 10, appear to satisfy an approximation of Eq. (1). Thus, the maximum error T appears to become linearly related to R when the vectors used are based on 10 revolutions of well spread data and an established vector.

2. The computer program KECLIPSE assumes that the earth's shadow is a perfect cone. The earth, however, is a slightly oblate spheroid. Assuming a conical shadow calculated with an earth of mean radius, the greatest possible error in t' is ± 6 seconds (Ref. 2). Other surface irregularities would account for even less error. It is known that light curves toward mass, but it is not a gross assumption this effect on the size of the earth's shadow would mean less than a few seconds error in t'. It was assumed that the angle S in Figure 4 remained a constant equal to 16'. Because the distance between the earth and the sun varies, S varied between 15.724' and 16.267'; assuming $\dot{s} = 3.0$ minutes of arc/minute yields an error of ±5 seconds in t' if individual S's are not calculated for each eclipse in the KECLIPSE program. This error in S would also mean an error of less than one-half second in the atmospheric correction x.

3. The equations given in Section III are estimates of the atmospheric corrections. They are based on a comparison of triangles whose sides were known to varying degrees of accuracy. Some of the distances were estimated using a constant velocity, but the velocity of each vehicle was constantly changing. In view of these factors an error of no more than ±5 percent of the value of x may be attached to the atmospheric correction.

4. It is estimated that the errors in calculating the penumbral entrance times were ±60 seconds during the March-April eclipse season and ±30 seconds during the October season. The difference is due to improvement in technique. The entrance time is calculated by observing the first detectable change in battery current and solar array output readings. These values are given in only two significant figures. Thus, they must be averaged to determine the entry time. This data must be interpreted in the light of previous uses of the satellite and of what fraction the satellite's reference illumination is of one sun (Ref. 4). The technique is complex.

5. In this study it was assumed that all the data used were good. This was not the case. Some bad data was accepted and inaccurate vectors were generated. These inaccurate vectors were later rejected with the previously mentioned statistical argument. Some of the March-April season shadow entrance times were only calculated to the nearest minute. Thus, errors of up to 29 seconds were possible in these cases. Different reference dates were used to calculate the vectors produced from normal VELA Program Operations. One second of error can be assigned to t' for every month within the difference between the time of epoch and the reference date. This error is due to the rotation of the vernal equinox 360 degrees every 26,000 years. Thus, a December vector produced with a 1 January reference date will be 12 seconds in error predicting eclipses. The possible errors are numerous; only a few are significant. With such a variety of errors, only a statistical interpretation can give the data meaning.

SECTION V
INTERPRETATION OF DATA

1. The data was not perfectly uniform. The vectors were produced using varying methods, different amounts of data, and three different software models. No one earth eclipse was identical to another. To perform a study accounting for all these variables would take more effort and money in computer time than the results would be worth. The data is sufficiently uniform for the intended purpose. Because of the variety of eclipse situations considered, only a statistical interpretation can give the data significance.

2. Noting Table 2, the data is not perfectly distributed over the vehicles per revolution. A closer look at the data reveals several low (t')s around Rev 6 for Vehicle 6564. Several small negative values of t' can be noted between Revs -4 and -6 for Vehicle 1851. While distortions can be expected at these revolutions, the distribution is adequate. Means (m) and sigmas (u) were calculated for any cluster of 30 or more points spanning less than 10 revolutions. These numbers and their associated, weighted, revolutions are given in Table 3. From experience it is known that about 95 percent of all VELA satellite vectors are accurate enough to produce good station acquisition data. For each revolution the values m - 2u and m + 2u were calculated. For a normal curve the area within 2u of the mean is about 95 percent of the total area. The expected distortions at -5 and +6 revs were found in a plot of the double points (R, m - 2u, m + 2u). The density of (R, t') points below the broken line connecting all (R, m) points did not match the (R, t') point density at Revs -5 or +36. This was taken as another distortion.

3. The values (m - 2u, m + 2u) between -5 and +5 revs were averaged yielding the t' intercept points of the cone of validity. With the previously mentioned distortions in mind, a cone of validity was approximated. A vector having an associated t' is valid if t' is between T_M and T_m where:

$$T_M = \begin{cases} 2.72 \text{ min}, & |R| \leq 5 \\ .107\,|R| + 2.16 \text{ min}, & |R| \geq 5 \end{cases}$$

$$T_m = \begin{cases} -1.67 \text{ min}, & |R| \leq 5 \\ -.064\,|R| - 1.35 \text{ min}, & |R| \geq 5 \end{cases} \qquad (6)$$

This method of determining the cone of validity lacks perfection. The cone was not meant to be precise; it was meant to serve as a realistic guideline, based on past experience, and will not account for every possible error.

Table 2

DISTRIBUTION OF DATA USED TO
CALCULATE THE CONE OF VALIDITY
(Number of Points/Rev Span)

	Vehicle					
Rev Span	1801	1851	3662	3674	6564	Interval Totals
Below -10	3	1	1	4	4	13
-10 to -1	7	8	4	14	7	40
0 to 10	4	11	5	6	18	44
11 to 20	4	4	11	7	1	27
21 to 30	3	1	9	5	0	18
31 to 40	14	14	3	5	0	36
Beyond 40	10	0	2	4	0	16
Total	45	39	35	45	30	194

Table 3

INITIAL CONE OF VALIDITY CALCULATIONS

Rev No.	Mean m	Sigma u	Validity Interval Minimum		Maximum	
Revs	(sec)	(sec)	(min)	(sec)	(min)	(sec)
-8.3	16	73	-2	10	2	42
-6.6	14	59	-1	44	2	12
-5.1	13	68	-2	3	2	29
-2.9	29	82	-2	14	3	12
-0.7	44	76	-1	49	3	17
0.3	53	72	-1	31	3	17
0.5	48	70	-1	32	3	8
3.0	35	49	-1	3	2	13
5.3	29	47	-1	4	2	2
8.1	34	73	-1	52	3	0
10.9	49	73	-1	36	3	14
12.8	65	75	-1	25	3	35
15.8	85	73	-1	1	3	51
16.7	87	79	-1	12	4	5
18.4	96	80	-1	4	4	16
35.8	52	149	-4	5	5	49
37.1	73	148	-3	44	6	10
38.3	60	142	-3	43	5	43

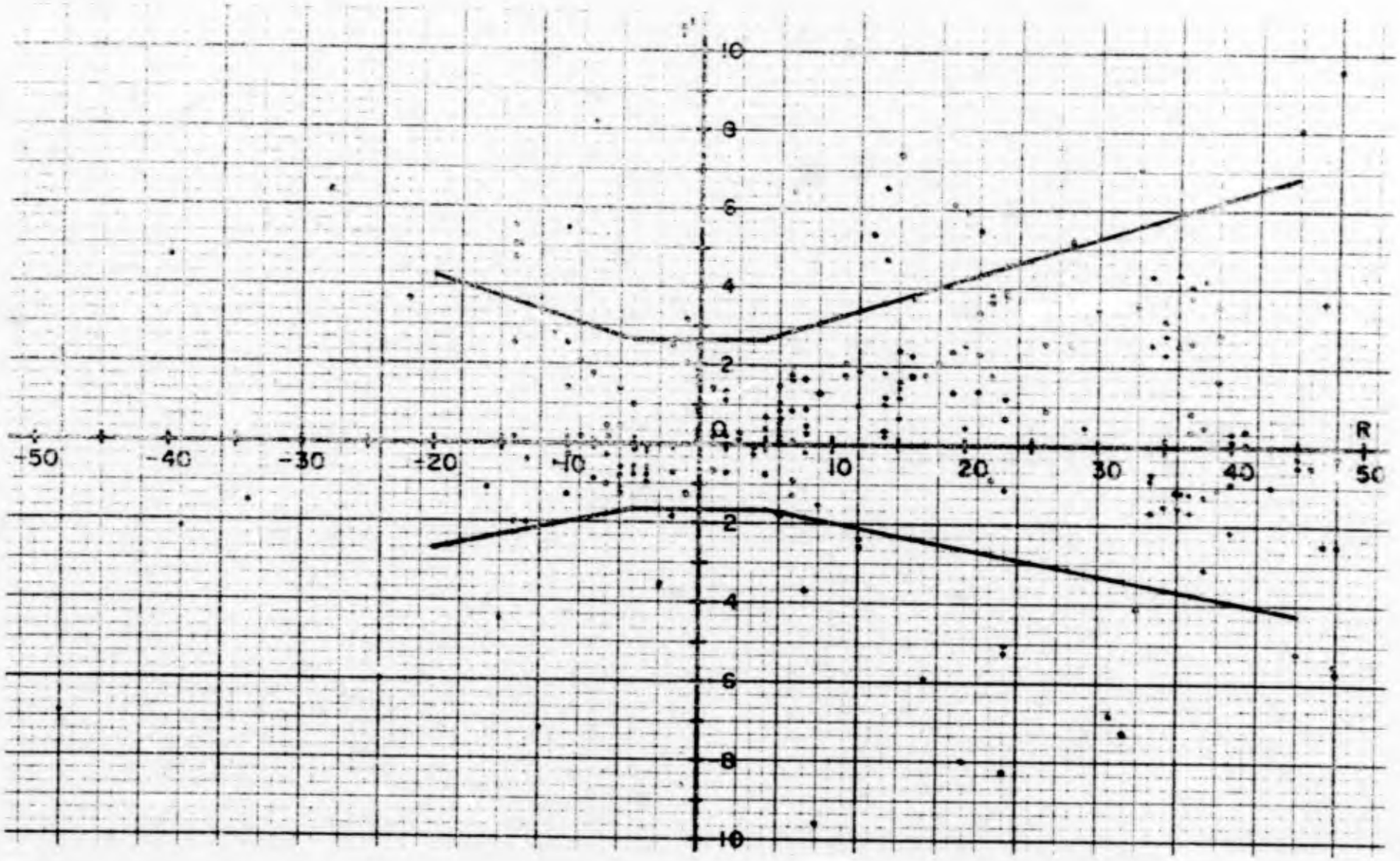


Figure 11. Cone of Validity

SECTION VI

CONCLUSIONS

1. Realizing the uniformity of the data and the purpose of the study, it may be concluded that the guidelines Eq. (6) for judging VELA vectors are valid. The following recommendations are proposed:

   a. That the cone-of-validity method be accepted as the standard method of judging VELA vectors.

   b. That the computer program KECLIPSE be changed to account for the atmospheric effect on the shadow of the earth.

2. Noting Figures 9, 10, and 11, the errors in the program KECLIPSE, the errors within the observed entrance times, plus other noted errors appear to cause a bias of about 30 seconds. The normal error (1 sigma), due to these errors seems to be 65 seconds in either direction. The program KUPDATE appears to be in error, over-approximating by at most 64 seconds and under-approximating by at most 38 seconds for KUPDATING 10 revolutions. These errors are reflected in the cone of validity and are acceptable. The error caused by the absence of an atmospheric correction is between 52 and 170 seconds. This should be corrected.

3. The cone-of-validity method should not be the only method for validating a vector. Orbit determination is a field that is so complex that other means of vector validation should be available. Thus, the methods of comparing vector data with previous orbital characteristics, ephemerides, or station look angles should be retained.

NOTE: A list of the actual data used can be obtained by writing the author.

REFERENCES

1. Data Dynamics, Inc., Orbit/Ephemeris Subsystem Computer Program and Supporting Documentation, Milestone 5, Subroutine KUPDATE - J30DA, Rev. A DDI 1215-26, Contract AF 04(695)-611 Monterey, California, 29 Jan 1965 (U)

2. Aerospace Corporation, Eclipse Predictions and Observations for VELA Spacecraft, TOR-669(6106-01)-1, Contract AF 04(695)-669 El Segundo, California, August 1965

3. Data Dynamics, Inc., Orbit/Ephemeris Subsystem Computer Program Preliminary Design Specifications, Milestone 3, Subroutine KECLIPSE - J61, Programming Document 2820/03/01, Contract AF 04(695)-611 Monterey, California, 5 Feb 1965 (U)

4. F. Link, Lunar Eclipses, Physics and Astronomy of the Moon, New York, Academic Press, 1962

AD641922

ERRATA SHEET

VOLUME II

PROCEEDINGS OF THE 13TH ANNUAL AIR FORCE SCIENCE AND ENGINEERING SYMPOSIUM

27-29 September 1966

Paper #33
Author: Raymond E. Christal
Title: (U) USING THE ELECTRONIC COMPUTER TO DEFINE AND IMPLEMENT POLICY

Add

Page 3, 11th Paragraph, 2nd Sentence: the word "such" between the words "... no _____ assumption ..."

Paper #41
Author: Robert G. McLaughlin, 1/Lt., USAF
Title of Paper: (U) ANALYSIS OF DIGITAL COMMUNICATION SYSTEMS ON TROPOSPHERIC SCATTER

Page 6 Line Bottom Add: References listed below

References

1. Staff of EMCRS, "Use of Sunde Tropo Channel Model to Predict Error Probabilities of AMRT Tests", RADC-TR-65-427, December 1965, AD625868.

2. Sunde, E. D., "Digital Tropo-scatter Transmission and Modulation Theory", Bell System Technical Journal, Part 1, January 1964.

Paper #43
Author: Lt. Charles B. Huelsman III
Title: (U) VALIDATING VELA SATELLITE ORBIT DETERMINATION PROCEDURES USING OBSERVED ECLIPSE ENTRANCE TIMES

Page	Line	Delete	Add
iv	15	22	21
iv	21	20	19
iv	22	21	20
17	7	from	during