PROCEEDINGS

of the

CONFERENCE ON EXPERIMENTAL TESTS OF GRAVITATION THEORIES

November 11-13, 1970

California Institute of Technology

Pasadena, California

Sponsored by

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION

EUROPEAN SPACE RESEARCH ORGANIZATION

JET PROPULSION LABORATORY, CALIFORNIA INSTITUTE OF TECHNOLOGY

PROCEEDINGS

of the

CONFERENCE ON EXPERIMENTAL TESTS OF GRAVITATION THEORIES

November 11-13, 1970

California Institute of Technology

Pasadena, California

Edited by R. W. Davies

November 1, 1971

Sponsored by

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION

EUROPEAN SPACE RESEARCH ORGANIZATION

JET PROPULSION LABORATORY, CALIFORNIA INSTITUTE OF TECHNOLOGY

PREFACE

The work described in the eleven papers presented by members of the Jet Propulsion Laboratory represent one phase of research carried out at the Jet Propulsion Laboratory, California Institute of Technology, under Contract NAS 7-100, sponsored by the National Aeronautics and Space Administration. Six of the papers presented were sponsored totally or partially by the European Space Research Organization.

The six sessions held over the three days from November 11-13, 1970, are listed below:

Session I: Astrophysics and Theory

Session II: Gravitation Waves, Redshift, and Gyroscope Experiments: Current and Future

Session III: Deflection, Time Delay, and Planetary Orbit Experiments: Past and Present

Session IV: Inherent Limitations and Future Technology Capabilities

Session V: Spacecraft Mission Analysis

Session VI: Potential of Space Technology for Testing Gravitation Theories

Panel Discussion 1:

Priorities of Theoretical Questions

Moderator: K. S. Thorne

Members: R. H. Dicke, L. Schiff, J. A. Wheeler, D. Sciama, W. M. Fairbank

Panel Discussion 2:

Technical Program Considerations

Moderator: R. H. Dicke

Members: R. Pacault, J. Mitchell, D. B. DeBra, R. Kraemer, R. Juillerat

FOREWORD

The structure and timing of the conference reflected the feelings of a number of people that technology, particularly that spawned by previous space activities, had made it possible to think realistically in terms of a long-range cooperative effort in the testing of General Relativity and other modern theories of gravity. Naturally, the conference was attended primarily by experts and enthusiasts. However, there are a number of competitive philosophies within this group on both the operational and on the theoretical level, and it was a constructive confrontation of these issues that the conference organizers sought to develop.

Not long after the conference ended, one of the most important personalities in this exciting field, Leonard Schiff, died. We will no longer have the pleasure of his gentle company or his sincere counsel, and so we modestly attempt to say we miss him by dedicating these proceedings to his memory.

R. W. Davies
Jet Propulsion Laboratory

CONTENTS

NOTE

The following presentations, not published in this proceedings, were also given at the conference:

Frontiers of Relativity, J. A. Wheeler, Princeton University (published in Physics Today, Jan. 1971)

The Detection of Gravitational Waves, Joseph Weber, University of Maryland (published in Scientific American, 224, No. 5, May 1971)

Gravitational Redshift Experiment, B. Persson, California Institute of Technology

Physical Limitations in Tracking Spacecraft: Earth Rotation and Transmission Media, D. W. Trask, Jet Propulsion Laboratory, California Institute of Technology

The Cosmic Helium Abundance

Leonard Searle
Hale Observatories
Carnegie Institution of Washington, California Institute of Technology

I. Introduction

It would be interesting to know the composition of the material out of which galaxies formed. This would obviously place an important constraint on cosmological theory. As you well know, some such theories make quantitative predictions concerning the helium content of pre-galactic matter. If the bulk of the helium in the world was formed at a pre-galactic epoch, we should expect that the helium content of cosmic material would be the same at all places and at all times since the formation of galaxies began. I shall here, very briefly, summarize the observational evidence relating to this expectation.

Elements other than hydrogen and helium certainly appear to have been formed by events subsequent to the formation of galaxies. Apart from the direct evidence provided by the abundance of radioactive species, the distribution of the elements shows this. Different subsystems within our own galaxy have very different compositions. For example, the abundance ratio of hydrogen to the common metals is more than one hundred times greater in many of the stars of the old spherical subsystem of our galaxy than it is in the common stars of the galactic disk. Moreover, entire galaxies differ one from another in their metal content. Among the near neighbors of our galaxy there is a clear difference in the metal content of the bulk of the stars belonging to the Large Magellanic Cloud on the one hand and to the dwarf elliptical galaxies (e.g., the Sculptor system) on the other. Evidently there is no cosmic metal abundance. Is there a cosmic helium abundance?

II. The Helium Abundance Here and Now

The helium abundance in young stars and in the interstellar gas of the solar neighborhood is by now well determined. If there is a cosmic helium abundance we know what it is! There are three lines of evidence.

A. Interior Structure

Studies of the interior structure of young stars predict relations between mass, radius, and luminosity for unevolved homogeneous hydrogen-burning stars that depend upon composition and in particular on the value of the helium mass fraction Y. From a comparison of these predicted relations and those actually observed, it is found that the helium mass fraction $Y = 0.25 \pm 0.05$. For examples of work of this kind see Percy and Demarque (Ref. 1) and Morton (Ref. 2).

B. Photospheric Spectra

Studies of the profiles of helium absorption lines in the photospheric spectra of hot young stars provide the second line of evidence. Early attempts to derive helium abundances in this way were based on crude theories of line broadening and line formation and the results were quite uncertain. Recently, however, much progress has been made concerning these matters and the new work, based on this theoretical progress, yields $Y = 0.30 \pm 0.04$. Examples of this work are Hyland (Ref. 3), Shipman and Strom (Ref. 4), Norris (Ref. 5), and Poland (Ref. 6). There remains the problem of interpreting the photospheric

abundances. I think that there is little doubt that generally they represent the composition of the material out of which the star formed. But there are, even among young stars, rare cases where the photospheric helium content differs from the normal value. Overabundances of helium can result from nuclear evolution within stars and the mixing of some of the processed material to the stellar surface. Overabundances, then, are not too surprising and need not detain us. More alarming are cases where the photospheric material is underabundant in helium. The best studied case is 3 Cen A (Ref. 7). On the surface of this star Y = 0.04 and most of the small amount of helium present is He^3. There are other abundance abnormalities. Phosphorous is overabundant compared with the normal composition of stars near the sun by more than a factor of a hundred. The processes that have led to this photospheric composition are unexplained but I do not think that anyone believes that the star was formed from helium deficient material. This star is a member of a binary system and its companion has a quite normal surface composition. There are numerous examples of individual helium-poor stars that are found in star clusters containing dozens of other hot stars of normal helium content. These stars are a clear warning that we cannot always take the photospheric composition of a star as indicating the composition of the material from which that star formed.

C. Emission-Line Spectrum of Ionized Gas

Finally, the helium abundance can be determined from the emission-line spectrum of ionized gas surrounding hot young stars. The ratio of the number of He^+ ions to protons in these regions can be obtained directly from the intensity ratio of the recombination lines of HeI and HI. Suitable pairs of recombination lines can be measured in both the optical and in the radio frequency regions of the spectrum. Observation and theory both show first that there is very little He^{++} in the ionized hydrogen region surrounding young stars and second that helium is generally either nearly all He^o or nearly all He^+ throughout the H^+ zone, cases where He^o and He^+ are present in comparable abundance being very rare. This circumstance is, of course, very favorable to abundance determination. For the average of a number of regions from which helium recombination lines are seen, optical studies give $n(He^+)/n(H^+) = 0.090 \pm 0.010$ (Ref. 8). Radio studies (Ref. 9) give 0.084 ± 0.003 for this same ionic number ratio. If the amount of neutral helium in these regions is negligible, $Y = 0.26 \pm 0.01$. If, as seems probable, a small amount of neutral helium is present in the H^+ zone, then this number should be construed as a lower limit.

There is good agreement between the results of the three independent lines of investigation. I think we can safely conclude that the helium mass fraction here and now lies within the limits $0.26 < Y < 0.32$.

III. The Helium Abundance Long Ago

For the helium abundance early in the history of our galaxy, only one of the three possible lines of evidence yields a decisive result. Studies of the interior structure of old stars predict evolutionary changes in luminosity and radius that a star undergoes as it ages. For an assemblage of coeval stars of different mass (such as a star cluster) the theory predicts a distribution in luminosity and radius which changes with the age of the assemblage and whose character depends on the initial helium abundance. Comparison of the observed and predicted distributions yields both the age and initial helium content of the stars in the cluster. All the numerous investigations based on this idea give the same result, namely that the initial helium abundance of the oldest and most metal-poor stars of our galaxy was about thirty percent. Values lower than twenty percent are clearly inconsistent with the observations. For different approaches see Cayrel (Ref. 10), Schwarzschild (Ref. 11) and Hartwick (Ref. 12). In a similar fashion the comparison of the observed and computed pulsational characteristics of old stars yields Y = 0.32 (Ref. 13).

In apparent conflict with these results, analyses of photospheric spectra show that helium is frequently underabundant by large factors in the hot old stars (which are the only old stars where the matter can be investigated spectroscopically). The best studied case is HD205805 (Ref. 14). I believe that the reality of the helium deficiency in the photospheres of these hot old stars is established beyond reasonable doubt. What is in doubt is whether this photospheric composition is the same as that of the material out of which the star originally formed. As I have already mentioned there is also convincing evidence for helium-deficient photospheres among young stars. In the one case where high resolution spectra could be obtained for an old helium-poor star (Ref. 15), that star was found to exhibit large overabundances of phosphorus - just like 3 Cen A (Ref. 7) - surely not a property of the primordial composition.

One might hope that the helium content could be obtained from the emission-line spectrum of ionized gas surrounding hot old stars. In a planetary nebula gas ejected from an evolved old star is ionized by the hot stellar remnant. For one such object in a metal-poor globular cluster a helium abundance Y = 0.42 was found (Ref. 16) i.e., the helium content was higher than that of young stars. There is a large spread in the helium content of planetary nebulae (Ref. 17) and it seems likely that their compositions have, in general, been contaminated by nuclear processes within the evolved stars themselves.

I conclude from this that of the three possible lines of evidence only the one involving observational checks on theories of stellar interior structure and evolution is at present capable of yielding information on the helium content of old stars. Such studies show that the oldest and most metal deficient stars of our galaxy were formed from material with a helium content indistinguishable (in the present state of theory and observation) from the helium content of the young stars in the solar neighborhood now. In any case, where and when the metals-to-hydrogen ratio was one or two orders of magnitude smaller than it is locally now the helium-to-hydrogen ratio was smaller by, at most, a factor of two. It is not yet possible to investigate how helium content depends upon age in any galaxy beyond our own.

IV. The Helium Abundance Far Away

The spatial homogeneity of the helium abundance outside our own galaxy can be studied from

the optical emission-line spectra of ionized gas around hot young stars in the spiral arms of external galaxies. The helium content has been determined in this way for a few galaxies of the local group (M31, M33, LMC, SMC, NGC6822) and for a few galaxies beyond it (M51, M101, NGC2403, 4214, 4449, 7679). A large fraction of the available information comes from the work of Peimbert and Spinrad (Ref. 18). None of these galaxies have helium contents significantly different from the interstellar gas of the solar neighborhood of our own galaxy. Always $Y = 0.28 \pm 0.06$.

Recently I have found, from observations of these emission regions, that there are composition gradients across the disks of spiral galaxies. In particular the ratio N/O decreases from the inner to the outer spiral arms. Furthermore, the N/O ratio is systematically higher in some galaxies than it is in others. In contrast to this observed inhomogeneity of the nitrogen and oxygen abundances the helium-to-hydrogen ratio is the same wherever I have been able to measure it.

It is perhaps worth remarking that these helium surveys refer to only a very small part of the observable universe. No helium abundances have been determined for galaxies further away than about 50 megaparsec.

V. Conclusion

At other places and at other times the composition of the cosmic material was significantly different from the composition in the solar neighborhood now. In contrast, the helium content has not changed by observable amounts over the lifetime of our galaxy and is the same for all galaxies so far surveyed. These results suggest that the bulk of the helium in the world was formed in an event or events that occurred before the galaxies themselves formed.

But, analyses of emission-line spectra of quasars lead to the result that in them the He/H ratio is sometimes an order of magnitude smaller than this "universal" abundance. See, for example, Bahcall and Kozlovsky (Ref. 19). This is, at the moment, the only serious evidence known to me that opposes the notion of a "cosmic" helium abundance. The determination of abundances from quasar spectra is no routine matter, however. Despite the progress that these models undoubtedly represent, they remain in some respects arbitrary. In fact they have internal inconsistencies; for example, they are not in thermal equilibrium. That this is a serious limitation is clear from the consideration that radiation in the observed emission lines (whose intensities are used to put constraints on the model) is precisely the cooling mechanism of the gas.

I think that it is fair to say that the helium abundances so far inferred from quasar spectra are much less certain than the results on which this review has been based.

References

1. Percy, J. R., and Demarque, P., Ap. J., 147, 1200, 1967.
2. Morton, D. C., Ap. J., 151, 285, 1968.
3. Hyland, A. R., Theory and Observation of Normal Stellar Atmospheres, 361, Edited by O. Gingerich, M.I.T. Press, 1969.
4. Shipman, H. S., and Strom, S. E., Ap. J., 159, 183, 1969.
5. Norris, J., Thesis, Australian National University, 1970.
6. Poland, A. I., Ap. J., 160, 609, 1970.
7. Hardorp, J., Zs. f. Ap., 63, 137, 1966.
8. Peimbert, M., and Costero, R., Bol. de los Obs. Ton. y Tac., No. 31, 1969.
9. Palmer, P., et al., Ap. J., 156, 887, 1969.
10. Cayrel, R., Ap. J., 151, 997, 1968.
11. Schwarzschild, M., Q.J.R.A.S., 11, 12, 1970.
12. Hartwick, F. D. A., Ap. J., 161, 845, 1970.
13. Sandage, A. R., Ap. J., 157, 515, 1969.
14. Baschek, B., and Norris, J., Ap. J. Supp., 19, 327, 1970.
15. Sargent, W., and Searle, L., Ap. J., 150, L33, 1967.
16. O'Dell, C. R., Peimbert, M., and Kinman, T. D., Ap. J., 140, 119, 1964.
17. Kaler, J. B., Ap. J., 160, 887, 1970.
18. Peimbert, M., and Spinrad, H., Ap. J., 159, 809, 1970.
19. Bahcall, J. N., and Kozolovsky, B., Ap. J., 155, 1077, 1969.

Stellar and Solar Rotation*

Jeremiah P. Ostriker†
Princeton University Observatory

I. Introduction

The fact that most stars rotate at observable rates is interesting to relativists for at least two reasons. Even if the rotating star is nearly spherical, its rotation causes the inertial frames in its vicinity to rotate with respect to inertial frames at distance. This effect, the "dragging of inertial frames," which, of course, does not exist in Newtonian Physics, provides one of the tests of relativistic theories of gravitation. Another effect due to oblateness is important chiefly for the confusion it can cause. If the matter distribution in a star is aspherical, the gravitational field will in general have a quadrupole moment, which, in turn, can affect the motion of a test particle in a way that can mimic (or mask) relativistic effects (Ref. 1).

Measurements of non-Newtonian effects are normally so delicate that the only suitable experimental laboratory is our solar system. Unfortunately, the sun does not rotate rapidly. This makes the anticipated effects small and difficult to estimate reliably. For example, we must know the state of interior rotation to calculate either of the above-mentioned effects. If the sun had a surface rotation rate 100 times the observed rate, it would be noticeably flattened and its luminosity would be significantly less than that of the rotating star having the same mass and age. Then the reduction would give a quantitative estimate of the state of interior rotation. For our sun, the reduction is less than 0.1% even if the central regions are rotating 15 times as fast as the surface. In addition, the sun is a relatively old star, and so quite inefficient processes like viscous angular-momentum transfer and "spindown" might have altered the initial angular momentum distribution. For these reasons, the direct approach to determining the rotation of most of the sun's mass is very difficult. Direct observation is complicated because the anticipated effects are small, and direct calculation is dangerous because so many very small physical effects may be acting in unanticipated ways. For these reasons, I shall take an indirect approach in which the sun is placed in the context of stars in general, and the probable initial angular momentum and subsequent evolution are estimated on the basis of continuity with stars of other masses and ages.

Taking the theoretician's approach, I will first discuss what we might on an a priori basis expect the angular momentum of stars to be and how we might expect it to be distributed within the stars. Then I will turn to the accumulated observations which show how the observed surface velocities depend on a bewildering variety of factors, including age, mass, spectral type, binarism, etc. Finally, I shall return to the sun.

II. Theoretical Expectations

A. General Remarks

There are two integral properties to consider, the angular momentum $J = \int v_\phi \tilde{\omega}\, dm$ and the rotational kinetic energy $T = 1/2 \int v_\phi^2\, dm$, where v_ϕ is the rotation velocity and $\tilde{\omega}$ the perpendicular distance from the rotation axis. For a given star,

*Work supported by Air Force Contract F44620-70-C-0033.
†Presently an Alfred P. Sloan Fellow.

the natural unit of energy is the gravitational energy $W = -1/2\int V_g\, dm$, where V_g is the gravitational potential, so we can use the ratio $|T/W|$ to characterize the importance of rotation for a given star. The angular momentum J is useful if different stars are to be compared and is particularly valuable since, under many circumstances, it is an invariant of the equations of motion.

B. Distribution of Angular Momentum Within Stars

Our laboratory experience leads us to expect that rotating objects rotate uniformly, the angular velocity of rotation being constant throughout the object. The contrary can be expected for stars. There are three principal ways in which cosmic bodies differ from laboratory objects, and all three differences lead one to expect that nonuniform rotation will be the rule, not the exception.

(1) Stars are very inhomogeneous with respect to density: the calculated central density of the sun is $10^{8.5}$ times larger than the density of its radiating surface. Thus, even if a star contracts from a uniform-density, uniformly rotating cloud, its central regions, having contracted further (to a higher density), will rotate more rapidly than average for the star. Since $\Omega \propto \rho^{2/3}$ (for fixed angular momentum) the center of the sun would be expected to rotate $\sim 10^6$ times more rapidly than its surface if angular momentum transfer could be neglected. Furthermore, high- and low-density regions are not strongly coupled together dynamically; density gradients (and, even more, molecular weight gradients) tend to stabilize a star against shear instabilities which might oppose differential rotation.

(2) Stars are very big. The time t_{vis} required for viscous forces to act is, by dimensional analysis, $t_{vis} = L^2/\nu$, where L is a characteristic length and ν the kinematic viscosity; notice that the velocity itself does not appear in the expression for t_{vis}. In a star's interior, $\nu < 10^{1.5}$; thus, even in 10^{10} years, viscous forces are able to couple together only regions smaller than about 0.1 $R_\odot$. Furthermore, as we shall shortly see, the viscous forces will not necessarily lead to uniform rotation but can work in the opposite direction.

(3) Stars are held together by gravity. The inertial forces associated with rotation are of a repulsive inverse-cube type. Under laboratory conditions, these are balanced by attractive short-range forces having a radial dependence as steep as, or steeper than, inverse-cube dependence. Stars, of course, are held together by the inverse-square long-range attractive force of gravity. As a consequence of this difference, stars, in contrast to flywheels, will not "break up" if their angular momentum is increased beyond some critical point. If rotation is rapid ($|T/W| \gtrsim 0.1$), in a centrally condensed star, the outer parts feel an essentially point gravitational potential, and the angular velocity distribution must become Keplerian. Under these circumstances, the angular velocity Ω will be $\propto \varpi^{-3/2}$ regardless of viscous effects (so long as the viscous time scale is long compared to the few-hour dynamical time scale); then it can be shown that viscous effects will tend to increase rather than decrease the Lagrangian gradient in angular momentum per unit mass.

There are, however, limits on the angular velocity (or momentum) gradients possible in a star. If the angular momentum per unit mass should increase toward the rotation axis [$d(\varpi^2\Omega)/d\varpi < 1$], the flow is unstable to the axisymmetric dynamical Rayleigh instability. In addition, Goldreich and Schubert (Ref. 2; see also Ref. 3) have shown that, if the rotational velocities are not too small, there exist slower instabilities (depending on diffusion) which require the angular momentum to be constant on cylinders, so that the angular velocity cannot increase inward from surface to center along surfaces parallel to the axis.

Thus, in summary, we may expect the central angular velocities of rotating main-sequence stars to be several times larger than the observed surface equatorial angular velocities.

C. Distribution of Angular Momentum Among Stars

The initial angular momentum of stars after they are formed from the interstellar gas and dust depends, of course, on which highly uncertain theory of stellar formation is adopted. But, although the detailed theories vary greatly in their predictions, all of the simpler (nonelectromagnetic) theories lead to values of angular momentum J much greater than those observed in main-sequence stars. This point is easily made in rough quantitative fashion. Consider (1) the angular momentum of a spherical blob of gas, due simply to the fact that the local standard of rest rotates about the galaxy with a period of $\approx 10^8$ years; and (2) the angular momentum expected if the protostar cloud has rotational kinetic energy in equipartition with its translational kinetic energy ($v_{tr} \approx 10$ km/s). In both cases, we find that a protostar with an original density of $\approx 10^{-24}$ g/cm^3 would have

$$J_a \approx 10^{56} (M/M_\odot)^{5/3} \text{ g cm}^2 \text{ s}^{-1} \text{ (galactic rotation)}$$

$$J_b \approx 10^{58} (M/M_\odot)^{4/3} \text{ g cm}^2 \text{ s}^{-1} \text{ (equipartition)} \qquad (1)$$

In comparison to this, even the most rapidly rotating main-sequence stars have $J \approx 10^{50} - 10^{52}$ g cm^{2}s^{-1}. The difference between these two sets of figures indicates that physical processes must exist which are capable of transferring angular momentum from a contracting, spinning star to its surroundings. Two general classes of mechanism have been considered, and it is likely that both types operate. Macroscopic instabilities may occur which transform spin angular momentum to the orbital form; the prevalence of binary stars (more than half of all stars) makes fission

plausible. On a "microscopic" level, angular momentum can be transferred from one mass element to another by a variety of processes, including molecular and radiative viscosity, magnetic stresses, and turbulent interchanges (see Ref. 4 for details). Magnetic braking can be very efficient in slowing the rotation of a star (Refs. 5 and 6), if it is simultaneously suffering mass loss via a stellar wind analogous to the observed solar wind. Stellar winds, in turn, are theoretically expected for the lower-mass main-sequence stars ($M < 1.7\ M_\odot$) having convective envelopes.

The fission process is beginning to be understood (Ref. 7). It is capable of reducing the spin angular momentum to the upper end of the observed range but always leaves the stars in a state of relatively rapid rotation ($|T/W| \approx 0.2$), often more rapid than the observed rate. Thus, it is likely that one or more of the suggested microscopic processes operate during the contraction phase. We may expect the same relatively slow angular momentum loss mechanisms to operate during the much longer main-sequence evolutionary phase if they are significant during the briefer contraction phase. In brief, old stars should rotate more slowly than young ones.

Tidal interactions, which are classically known to be capable of transferring angular momentum from spin to orbital form, can account for the slow rotation of the moon and perhaps even Venus (Ref. 8). Correspondingly, we may expect stars in close binary systems to rotate synchronously with their orbital motions, which, in astronomical terms, would lead to slow rotation.

Finally, since the most efficient of the proposed loss mechanisms require magnetic fields, we might expect magnetic stars to rotate slowly.

Needless to say, the list of "predictions" has been constructed with one eye on the observed facts; however, the theoretical arguments are still useful if only to provide an intellectual framework within which to organize the complex observations of stellar rotation.

III. Observations of Surface Stellar Velocities

Figure 1 summarizes the accumulated data on early-type main-sequence stars adopted from Slettebak (Ref. 9). The sample excludes the relatively common Am and Ap stars, which have very low rotational velocities but which also display spectral peculiarities indicating abundance anomalies and, for some, strong magnetic fields. Be stars (e-showing spectral lines in emission) have been included; they occupy the upper left portion of the cross-hatched area. The sample shown in Fig. 1 has been chosen to represent essentially unevolved stars which, according to prevailing theory, have convective cores and radiative envelopes. In the mass range $1.1 < M/M_\odot < 1.7$, the angular momentum increases very steeply with increasing mass, the approximate dependence being $J \propto M^{4.5}$. These stars have both small convective cores and small convective envelopes, according to current stellar interior theory (Ref. 10). The upper main-sequence stars ($M/M_\odot > 1.7$) do not have convective envelopes. The average angular momentum for stars of given mass follows approximately the law $J \propto M^{5/3}$ (Ref. 11). This dependence is close to what one would expect if the outer parts were in Keplerian orbits, although the observed velocities tend to be somewhat below this limit. The Keplerian limit — sometimes misleadingly called "breakup velocity" — represents the maximum surface velocities consistent with hydrodynamical equilibrium. Actually, a somewhat lower limit exists to the possible angular momentum if stability to fission as well as equilibrium is required. The stability limit — $|T/W| \approx 0.2$ — (Ref. 7) is quite close to the left-hand part of the upper envelope shown in Fig. 1. Young upper main-sequence stars (those in galactic clusters) tend to rotate at velocities ranging from the maximum compatible with a stable equilibrium to a factor of 4 slower (Ref. 12). The field stars comprising Slettebak's sample have similar velocities, except that the less massive stars (A5 - F5) tend to have somewhat lower velocities than their counterparts in clusters. This is probably an evolutionary effect since, for these spectral types, field stars are considerably older than cluster members.

In Fig. 2 (adopted from Ref. 12) the variation from cluster to cluster is shown dramatically. All stars observed in the specified clusters have been used, including Ap, Am, spectroscopic binaries, and Be stars. Much of the variance at a given spectral type can be understood qualitatively in terms of the mechanism described in Section II. The Ap stars are typically magnetic variables and rotate slowly. The Am stars are often close binaries. When close binaries and spectroscopically peculiar stars are removed from the sample, the cluster curves look much more like one another and more like those in Fig. 1. Much of the remaining variance can be ascribed to the age of the cluster.

Kraft (Ref. 13) has demonstrated such an effect among field and cluster stars of different age categories. His results, reproduced in Table 1, show that, for F5 stars, the surface velocities decrease by more than a factor of 6 in a few billion years.

The lower main-sequence stars ($M < 1.1\ M_\odot$) rotate much more slowly than more massive stars and as much as 2 orders of magnitude more slowly than the limits given by stability requirements. Although observation becomes difficult because of both the faintness of the stars and the weakness of the rotational effects, there are indications that these stars also slow down significantly with increasing age. They all have significant convective envelopes, and many show signs of chromospheric activity more dramatic than that of the sun. It seems a reasonable presumption that all have stellar winds. If the chromospheric activity (indicated by flares and emission lines) is a measure of the strength of the winds, then these are much stronger when the stars are young (cf. Ref. 14 for a summary discussion). The angular momentum removed from the sun by the solar wind can be measured by space probes. Taking the mass flux of 1.2×10^{12} g s^{-1} from Ref. 15, and the ratio of mass flux to angular momentum flux from Ref. 16, I obtain an angular momentum loss rate of 6×10^{30} g cm^{2}s^{-1}. This implies that the sun's surface velocity will decrease by a factor e in about 9×10^9 years if it is rotating uniformly, but in only 1.0×10^8 years if the convective envelope alone is being decelerated.

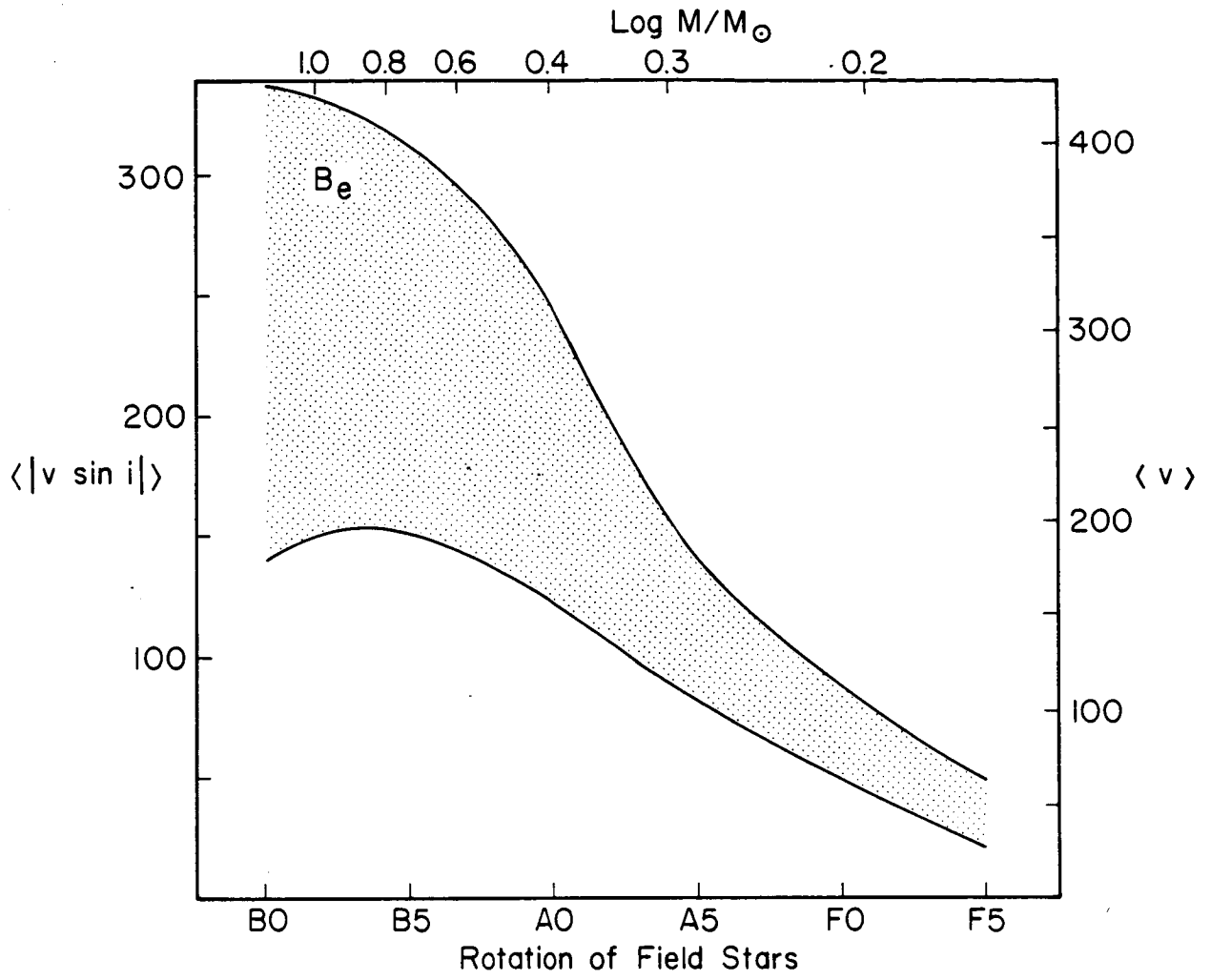


Fig. 1. Rotation of main-sequence field stars adopted from Ref. 9. Left ordinate: product of equatorial velocity (km/s) and sin of angle between rotation axis and line of sight (with assumption made in data reduction that stars are uniformly bright, uniformly rotating spheres); range in the average $|v \sin i|$ at given spectral type due to differences in samples investigated by various observers. Right ordinate: average equatorial velocity assuming random orientation of axes. Lower abscissa: observed spectral type. Upper abscissa: approximate calibration giving the relation between mass (in units of the sun's mass) and spectral type; notice extreme nonlinearity of scale.

On a rough quantitative basis then, the small and secularly decreasing velocities of the lower main-sequence stars can be understood in terms of stellar wind angular momentum transport. *

IV. The Sun

Evidence from the youngest clusters indicates that the sun may have arrived on the main sequence with a surface velocity of ~15 ± 5 km/s (data from Ref. 14). The rotation period at this time was ~3.3 days and applied to the interior as well since, during its pre-main-sequence phase, convection prevailed throughout the star and turbulent viscosity was probably important. During its early evolution, chromospheric activity (as indicated by CaII emission) was intense, the solar wind stronger than it is at present, and the angular momentum loss rate was greater.

Suppose $dJ/dt = QJ^n$. Then (for $n > 1$) the characteristic time $(J/\dot{J})_0$ measured at any instant is $(n - 1)$ times the age of the object if $J_0 \ll J_{initial}$. The observed characteristic time for the sun is about twice its present age (cf. Section III) assuming solid body rotation. The corresponding index, $n = 3$, is consistent with observations of other stars (see Table 1), which, in rough quantitative terms, indicates that solar wind angular momentum transport may have been sufficient to slow the sun, as a whole, to its presently observed rotation rate. These arguments are not definitive; furthermore, our understanding of the dynamical processes coupling the decelerated envelope to the central core is insufficient to permit the assumption that

*We have intentionally neglected the fact that for the sun and at least one other late type dwarf (Ref. 17), most of the angular momentum is probably in a planetary system.

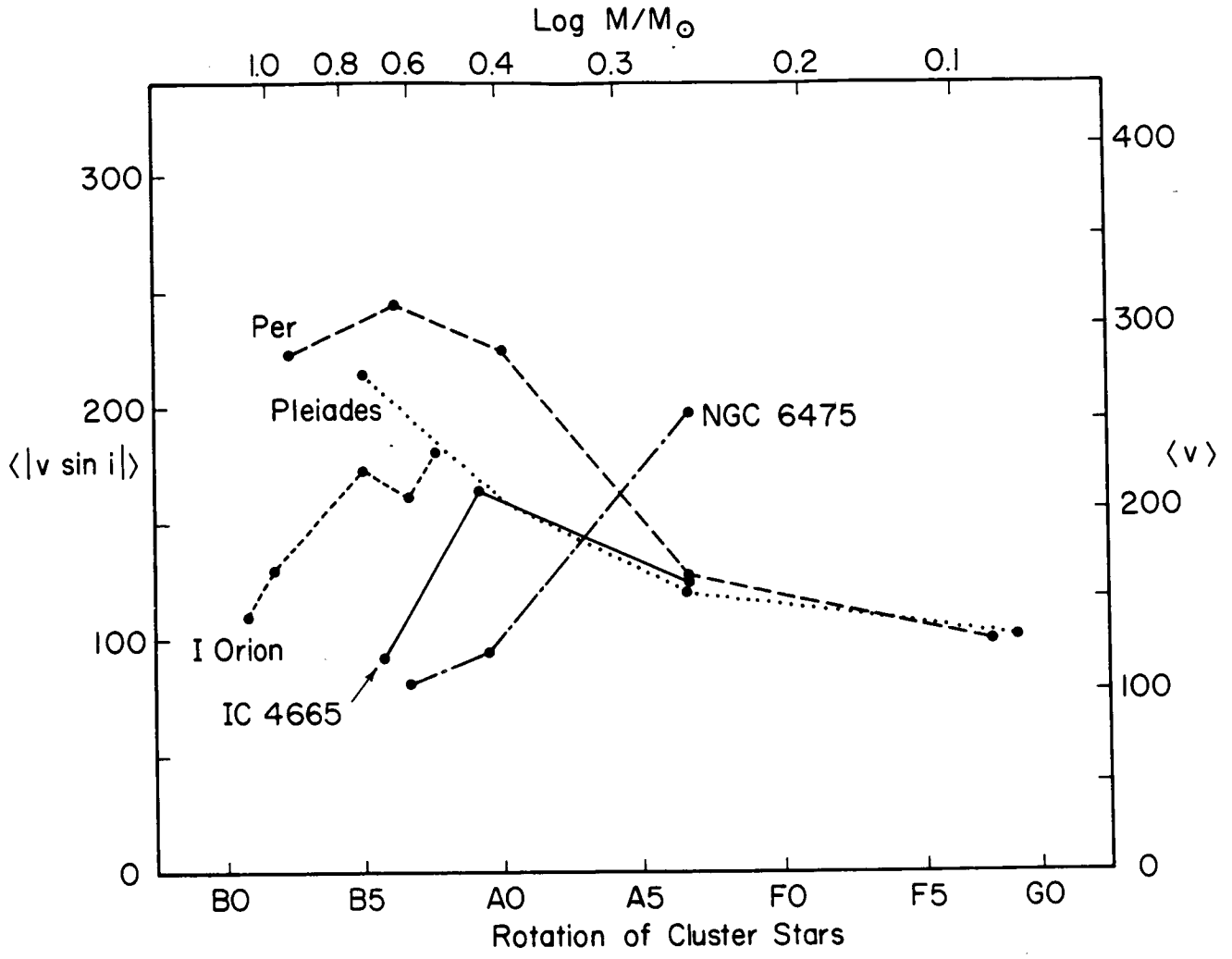


Fig. 2. Average equatorial rotation velocities of stars in various clusters, adopted from Ref. 12. Scales defined as in Fig. 1. When close binaries and stars with spectral peculiarities (e.g., Am, Ap) are removed from the cluster samples, the curves become more like one another and more like those in Fig. 1, for the early spectral types. However, F and G cluster stars do rotate more rapidly than field stars of the same spectral type, probably indicating a decrease in surface velocity with increasing age.

Table 1. <v> as a function of age for stars of mass M = 1.2 $M_\odot$

Kind of star	Approximate average age, years	<v> km/s
Field non-emission	3×10^9	6
Hyades Cluster	4×10^8	18
Pleiades Cluster	3×10^7	39

the core will not maintain its original rotation rate. Allowing for these uncertainties, it seems possible to restrict the present value of the core rotation rate to the range 3 da < P_c < 25 da.

References

1. Dicke, R. H., Nature, 202, 432, 1964.
2. Goldreich, P., and Schubert, G., Ap. J., 150, 571, 1967.
3. Fricke, K., Z. Astrophys., 68, 317, 1968.
4. Spitzer, L., in Stars and Stellar Systems, Vol. VII, Chap. 1, Edited by B. M. Middlehurst and L. H. Aller, University of Chicago Press, Chicago, 1968.
5. Mestel, L., and Selley, C. S., in Stellar Rotation, pp. 269-273, Edited by A. Slettebak, Reidel Publishing Co., Dordrecht, Holland, 1970.

6. Kulsrud, R. M., Ap. J., 163, 567, 1971.

7. Ostriker, J. P., in Stellar Rotation, pp. 147-156, Edited by A. Slettebak, Reidel Publishing Co., Dordrecht, Holland, 1970.

8. Goldreich, P., and Peale, S. J., Ann. Rev. of Ast., 6, 287, 1968.

9. Slettebak, A., in Stellar Rotation, pp. 3-7, Edited by A. Slettebak, Reidel Publishing Co., Dordrecht, Holland, 1970.

10. Cox, J. P., and Giuli, R. T., Principles of Stellar Structure, Chap. 26, Gordon and Breach, Inc., New York, 1968.

11. Kraft, R. P., in Spectroscopic Astrophysics, Edited by G. Herbig, University of California Press, Berkeley, 1970.

12. Abt, H., in Stellar Rotation, pp. 193-203, Edited by A. Slettebak, Reidel Publishing Co., Dordrecht, Holland, 1970.

13. Kraft, R. P., Ap. J., 150, 551, 1967.

14. Kraft, R. P., in Stellar Astronomy, Vol. 1, p. 317, Edited by H-Y. Chiu, R. Warasilo, and J. Remo, Gordon and Breach, Inc., New York, 1969.

15. Neugebauer, M., and Snyder, C., J. Geophys. Res., 71, 4469, 1966.

16. Grzedzielski, S., Astrophys. and Space Science, 3, 139, 1969.

17. van de Kamp, P., A. J., 74, 757, 1969.

Theoretical Frameworks for Testing Relativistic Gravity — A Review*

Kip S. Thorne, Clifford M. Will, and Wei-Tou Ni
California Institute of Technology

I. Introduction

The technology of the 1970s makes possible many experimental tests of general relativity. But their cost, in manpower and money, is high. (The megabuck will be a useful unit of measure for some of the tests.) For this reason, it is crucial that we have as good a theoretical framework as possible for comparing the relative values of the various experiments — and for proposing new ones that might have been overlooked.

The most simple-minded theoretical framework would be a direct comparison of general relativity with Newtonian theory. Indeed, just such a comparison motivated Einstein's original three tests — the gravitational red shift, the deflection of light, and the perihelion shift of Mercury. One might think that we should merely continue to measure these and other non-Newtonian, general relativistic effects to higher and higher accuracy, and only if a discrepancy between experiment and theory is found, should we begin to consider other theories.

This would be a reasonable approach if we had enormous confidence in general relativity; but we do not — at least some of us don't some of the time. So we would prefer to design the experiments to be as unbiased as possible; we would like to see them force us, with very few a priori assumptions about the nature of gravity, toward general relativity or some other theory. And, of course, this can happen only if we first open our minds to a wide variety of theoretical possibilities.

A leading exponent of this viewpoint is Robert H. Dicke.** It has led him and others to perform several high-precision null experiments (Dicke-Eotvös experiment; Hughes-Drever experiments; ether drift experiments) which greatly strengthen our faith in the foundations of general relativity (Ref. 1). Without this viewpoint, some of the null experiments might not have been performed, and we would certainly not understand so well their significance.

Dicke himself has suggested one type of theoretical framework for comparing various theories of gravity and analyzing the significance of various experiments. This "Dicke framework" assumes almost nothing about the nature of gravity. It helps one to design and discuss experiments which test, at a fundamental level, the nature of space-time and gravity. Within it, one asks such questions as: Do all bodies respond to gravity with the same acceleration? Is space locally isotropic in its intrinsic properties? What types of fields, if any, are associated with gravity — scalar fields, vector fields, tensor fields, affine fields, ... ?

We will not discuss the Dicke framework in this paper. Our reasons: (1) Detailed reviews of the Dicke framework are readily available elsewhere

*Supported in part by the National Science Foundation [GP-19887, GP-28027, GP-27304] and the National Aeronautics and Space Administration Caltech/JPL Contract No. NAS 7-100 [188-41-54-02-01].

**See also pp. 100-101 of Ref. 2 for a very convincing discussion of it.

(Ref. 1, especially Appendix 4; Ref. 3). (2) Its forte is the analysis of null experiments, which are deemphasized at this conference; for analyzing "solar system experiments," the Dicke framework is much less useful.

A second theoretical framework, one better suited to solar system experiments, is the Parametrized Post-Newtonian (PPN) formalism of Nordtvedt (Ref. 4) and Will (Ref. 5), which is based on earlier work by Eddington (Ref. 6), Robertson (Ref. 7), Schiff (Ref. 8), and Baierlein (Ref. 9). The PPN framework starts where the Dicke framework leaves off: By analyzing a number of experiments within the Dicke framework, Thorne and Will (Ref. 3) arrive at (among others) two "fair-confidence" conclusions about the nature of gravity: (1) that gravity is associated, at least in part, with a symmetric tensor field, the "metric" and (2) that the response of matter and fields to gravity is described by $\underset{\sim}{\nabla} \cdot \mathbf{T} = 0$, where $\underset{\sim}{\nabla}\cdot$ is the divergence with respect to the metric, and $\mathbf{T}$ is the stress-energy tensor for all matter and nongravitational fields. These two conclusions in the Dicke framework become the postulates upon which the PPN framework is built; and one calls theories of gravity that satisfy these two postulates "metric theories."

To construct the PPN framework, one first takes the slow-motion, post-Newtonian limits of all conceivable metric theories. One then builds a single "post-Newtonian theory of gravity," which possess those limits as special cases. This post-Newtonian theory or framework contains a set of real-valued parameters. The post-Newtonian limit of each metric theory is characterized by a set of particular values for these PPN parameters. The task of solar system gravity experiments in the coming decade can be regarded as one of measuring the values of these PPN parameters and thereby delineating, hopefully, which theory of gravity is correct.

It is important for the future that experimenters concentrate not only on measuring the PPN parameters. They should also perform new experiments within the Dicke framework to strengthen — or destroy — the foundation which it lays for the PPN framework.

Section II of this paper discusses metric theories of gravity; it explains why every currently viable theory is a metric theory, and it lists and classifies a large number of metric theories. Section III reviews the PPN framework and lists the values of the PPN parameters for various metric theories. Section IV interprets past and future experiments in terms of the PPN framework, thereby delineating their significance. Section V compares the predictions of various metric theories with the results of past experiments.

II. Metric Theories of Gravity

A. Definition of Metric Theory

A given theory of gravity is a "metric theory" if and only if it can be given a mathematical representation in which two conditions hold: Condition I. There exists a metric of signature -2, which governs proper length and proper time measurements in the usual manner of special and general relativity:

$$ds^2 = g_{ij}dx^i dx^j \tag{1}$$

Condition II. Stressed matter and fields being acted upon by gravity respond in accordance with the equation

$$\underset{\sim}{\nabla} \cdot \mathbf{T} = 0 \tag{2}$$

where $\mathbf{T}$ is the total stress-energy tensor for all matter and nongravitational fields.

It is significant that these two postulates can be obtained directly from a single, attractive assumption: the existence of local Lorentz frames everywhere, in which all the laws of special relativity take on their usual form (a weakened form of the strong equivalence principle; cf. C below).* However, we prefer to put the PPN formalism on the narrower base of metric plus $\underset{\sim}{\nabla} \cdot \mathbf{T} = 0$, so that its experimental justification can be discussed more clearly.

*In applying this assumption, one must be careful to allow for coupling to the Riemann curvature tensor in certain of the usual special-relativity equations. For example, the usual laws of vacuum electrodynamics in terms of the physical observables $\underset{\sim}{E}$ and $\underset{\sim}{B}$,

$$\underset{\sim}{\nabla} \cdot \underset{\sim}{E} = \underset{\sim}{\nabla} \cdot \underset{\sim}{B} = 0, \qquad \underset{\sim\sim}{\nabla \times B} = \frac{-\partial \underset{\sim}{E}}{\partial t}, \qquad \underset{\sim\sim}{\nabla \times B} = \frac{\partial \underset{\sim}{B}}{\partial t}$$

in curved space-time imply that the vector potential $\underset{\sim}{A}$ in the Lorentz gauge ($\underset{\sim}{\nabla} \cdot \underset{\sim}{A} = 0$) satisfies

$$\Box \underset{\sim}{A} - \mathbf{R} \cdot \underset{\sim}{A} = 0$$

rather than $\Box \underset{\sim}{A} = 0$. Here $\Box$ is the wave operator (d'Alembertian) and $\mathbf{R}$ is the Ricci tensor.

One should keep in mind that any metric theory of gravity can perfectly well be given a mathematical representation that violates conditions I and II. For example, the Brans-Dicke theory, in the mathematical representation of Dicke (Ref. 10) does not satisfy either condition: Dicke's scalar field causes deviations from geodesic motion, and physical rods and clocks do not measure $ds^2 = g_{ij}dx^i dx^j$. However, in the original mathematical representation of Brans and Dicke (Ref. 11), the theory satisfies both conditions, so it is a metric theory.

Notice that, in the "canonical representation" of a metric theory (the representation where conditions I and II are satisfied), the metric is the only gravitational field which enters into the response equation $\nabla \cdot \mathbf{T} = 0$. (The metric determines ∇; and $\mathbf{T}$ contains no gravitational fields.) This does not mean that the metric is the only gravitational field present. On the contrary, as in Brans-Dicke theory, there may be other fields. However, the role of the other fields can only be that of helping to generate the space-time curvature associated with the metric. Matter may create them, and they plus matter may create the curvature, but they cannot act back directly on the matter. The matter responds only to the metric!

In this paper, we shall restrict ourselves entirely to metric theories of gravity. Our justification is two-fold: (1) conditions I and II for metric theories are supported strongly by experimental data (B and C below), and (2) every currently viable theory of gravity is a metric theory (D below).

B. Evidence for the Existence of a Metric

There is strong experimental evidence for the existence of a symmetric "metric" field $\mathbf{g}$, whose orthonormal tetrads are related by Lorentz transformations, and which determines the ticking rates of atomic and nuclear clocks and the lengths of laboratory rods.

This evidence comes largely from elementary particle physics. It is of two types: first, experiments which measure space and time intervals directly — e.g., measurements of the time dilation of the decay rates of unstable particles;* second, experiments which reveal the fundamental role played by the Lorentz group in particle physics,** including every-day, high-precision verifications of four-momentum conservation and of the relativistic laws of kinematics. To cast out the metric tensor would destroy the theoretical backing of such experiments.

Let us notice what particle-physics experiments do and do not tell us about the metric tensor $\mathbf{g}$: First, they do not guarantee that there exist global Lorentz frames — i.e., coordinate systems extending throughout all of space-time in which†

$$g_{ij} = \text{Minkowskii metric } \eta_{ij} \equiv \text{diag}\,(1, -1, -1, -1) \tag{3a}$$

However, they do demand that at each event q there exist local frames, related by Lorentz transformations, in which $g_{ij}(q) = \eta_{ij}$. Moreover, given such a frame, elementary differential geometry guarantees the existence of coordinates in which

$$g_{ij} = \eta_{ij} + 0\left(\sum_k \left|x^k - x^k(q)\right|^2\right); \qquad \frac{\partial g_{ij}}{\partial x^k} = 0 \text{ at } q \tag{3b}$$

Such a coordinate system is called a "local Lorentz frame at q."

Second, particle experiments do not guarantee that freely falling particles move along geodesics of the metric field, i.e., along straight lines in the local Lorentz frames. In particular, we do not know from elementary-particle experiments whether the local Lorentz frames in an earth-bound laboratory are freely falling (so they fly up from the center of the earth and then fall back with Newtonian acceleration $g = 980\ \text{cm/s}^2$), whether they are forever at rest relative to the laboratory walls, or whether they undergo some other type of motion. The strong equivalence principle (Einstein elevator argument) predicts that the local Lorentz frames should fall freely, so that a free particle initially at rest in one would always remain at rest in it. Contrast this with flat-spacetime theories of gravity, in which rods and atomic clocks are governed by the global Minkowskii metric (3a), and gravity, like electromagnetism, is described by a field (scalar, vector, tensor, or combination) which resides in flat space-time. In such theories a Lorentz frame initially at rest in an earth-bound laboratory would remain always at rest (except for accelerations $\ll 980\ \text{cm/s}^2$ due to the earth's rotation and orbital motion). These possibilities and others are permitted by all elementary-particle experiments to date (except the Mossbauer red-shift experiments discussed in C1 below).

Third, elementary particle experiments do tell us that the times measured by atomic clocks depend only on velocity, not upon acceleration.

---

*For a 2% test of time dilation with muons of $(1 - v^2)^{-1/2} \sim 12$ in a storage ring, see Ref. 12. For earlier time dilation experiments, see Refs. 13 through 17. For an experiment which verifies, to one part in 10^4, that the speed of light (γ-rays) is independent of the velocity of its source (decaying π°) for source velocities $v > 0.99975c$, see Ref. 18.

**See Ref. 19 for a discussion of Lorentz invariance, spin and statistics, the TCP theorem, and relevant experiments.

†Here and throughout most of this paper, we use units in which the speed of light is unity.

The measured squared interval is $ds^2 = g_{ab}dx^a dx^b$, independently of acceleration. Equivalently but more physically, the time interval measured by a clock moving with velocity v^α relative to a local Lorentz frame is

$$ds = \left(\eta_{ab}\, dx^a\, dx^b\right)^{1/2} = \left[1 - \left(v^x\right)^2 - \left(v^y\right)^2 - \left(v^z\right)^2\right]^{1/2} \tag{4}$$

independently of the clock's acceleration d^2x^α/dt^2. If this were not so, then particles moving in circular orbits in strong magnetic fields would exhibit different decay rates than freely moving particles, which they do not (Ref. 12),* and ^{57}Fe nuclei would show acceleration dependence in the frequency of their Mossbauer transitions, which they do not (Ref. 20).

C. The Response of Matter to Gravity

1. Test-Body Trajectories and the Gravitational Redshift

According to the Dicke-Eötvös experiment (see, e.g., Ref. 1), the trajectory of a freely falling, neutral, laboratory-sized object ("test body") is independent of its structure and composition — at least to a high degree of accuracy. We shall assume complete independence (Dicke's weak equivalence principle).

This means that space-time is filled with a family of preferred curves, the test-body trajectories. Any initial event in space-time and initial velocity through that event determine a test-body trajectory which is unique except for parametrization. If we knew all the test-body trajectories, we would know a great deal — perhaps everything — about gravity.

There is a second family of preferred curves filling all of space-time: the geodesics of the metric **g**. It is tempting to identify these geodesics with the test-body trajectories (Einstein's equivalence principle). However, we should not do so without rather convincing experimental proof.

In order to see what kinds of experiments are relevant, let us elucidate the physical significance of the goedesics. A geodesic of **g** is most readily identified locally by the fact that it is a straight line in the local Lorentz frames. Put differently, a body's motion is unaccelerated as measured in a local Lorentz frame if and only if the body moves along a geodesic of **g**. Hence, to determine whether test-body trajectories are geodesics, we must compare experimentally the motion of a local Lorentz frame with the motion of a test body.

It is easy to study experimentally the motions of test bodies; relative to an earth-bound laboratory they accelerate downward, with $g = 980\ cm/s^2$; and this acceleration can be measured at a given location on the earth to a precision of one part in 10^6 (Ref. 23).

Unfortunately, it is much more difficult to measure the motion of a local Lorentz frame. It seems to the authors that the only experimental handle one has on this today is gravitational redshift experiments.

The red-shift experiment of highest precision is that of Pound and Rebka (Ref. 24), as improved by Pound and Snider (Ref. 25). It reveals a red shift of $z = \Delta\lambda/\lambda = (gh/c^2)(1 \pm 0.01)$ for photons climbing up through a height h in the earth's locally homogeneous gravitational field — if the emitter and receiver are at rest relative to the earth's surface. This tells us that the local Lorentz frames are not at rest relative to the earth's surface as predicted by flat-space-time theories of gravity; rather, as predicted by the strong equivalence principle, they accelerate downward with the same acceleration g as acts on a free particle (to within 1% precision) (cf. B above). To arrive at this conclusion from the experiment, we argue as follows.**

We wish our argument to be as independent of the special-relativistic laws of physics as possible. The only aspects of special relativity that we shall use are (1) the relationship between the Minkowski metric of the local Lorentz frames and the ticking rates of atomic clocks and (2) the conservation of wave fronts in electromagnetic waves. Let us assume (falsely) that the local Lorentz frames were unaccelerated relative to the walls of the tower used in the Pound-Rebka experiment. We can then perform a calculation in that particular Lorentz frame which was attached to the walls of the tower and was large enough to cover the entire tower. The static nature of the emitter, receiver, gravitational field, and Lorentz coordinate system guaranteed that, although the space-time trajectories of the wave crests might have been bent by gravity, they were certainly the same from one crest to another, except for a translation Δt_L in the Lorentz time coordinate. Thus, the coordinate rates $1/\Delta t_L$ emission and reception of wave crests

*The experiment of Farley et al. (Ref. 12 is a 2% check of acceleration independence of the muon decay rate for energies $E/m = (1 - v^2)^{-1/2} \sim 12$ and for accelerations, as measured in the muon rest frame, of $a = 5 \times 10^{20}\ cm/s^2 = 0.6/cm$. Note that, at accelerations a factor of 10^{13} larger than this ($a \sim 10^{33}\ cm/s^2 \sim 10^{12}/cm$), in one light travel time across the muon, it accelerates up to near the speed of light if it was initially at rest. Such large accelerations will probably affect the decay rates — not because of any breakdown in relativity theory, but because the decay cannot be analyzed within a single comoving local Lorentz frame. The muon ceases to be a valid special relativistic clock. See Refs. 21 and 22.

**We thank Charles W. Misner for helping us to formulate this version of the argument. For a variety of somewhat similar arguments see chapter V of Ref. 2.

were the same. But, by assumption, these Lorentz coordinate rates were also the proper rates measured by the atomic clocks (^{57}Fe nuclei) of the experiment. Hence, theory predicts zero red shift, in contradiction with experiment. Our assumption that the local Lorentz frames were unaccelerated must be wrong!

We must assume, then, that the local Lorentz frames were accelerated relative to the tower. Since gravity pointed vertically and all horizontal directions were equivalent in all respects, the acceleration of the Lorentz frames must have been vertical. Denote by a its value in the downward direction. As in our previous argument, in a static coordinate system (i.e., in coordinates at rest relative to emitter, receiver, and earth's static gravitational field) the wave-crest trajectories must have been identical, except for a time translation Δt_s from one crest to the next. But in this case the static coordinates were not Lorentz coordinates. Rather, they were accelerated upward (in the +z-direction) relative to the Lorentz frames (here we show the speed of light explicitly):*

$$ct_L = \left(\frac{z_s + c^2}{a}\right) \sinh\left(\frac{at_s}{c}\right),$$

$$z_L = \left(\frac{z_s + c^2}{a}\right) \cosh\left(\frac{at_s}{c}\right),$$

$$x_L = x_s, \qquad y_L = y_s$$

Hence, proper time as measured by atomic clocks was given by

$$c^2 d\tau^2 = c^2 dt_L^2 - dx_L^2 - dy_L^2 - dz_L^2$$

$$= \left(\frac{1 + az_s}{c^2}\right)^2 c^2 dt_s^2 - dx_s^2 - dy_s^2 - dz_s^2$$

Since, as before, the wave-crest emission and reception rates were the same ($1/\Delta t_s$) when measured in static coordinate time, they were related by

$$\frac{\Delta\lambda}{\lambda} = \frac{\nu_{em}}{\nu_{rec}} - 1 = \frac{\left[1 + (az_s)_{rec}/c^2\right]\Delta t_s}{\left[1 + (az_s)_{em}/c^2\right]\Delta t_s} - 1$$

$$\frac{a\,(z_s)_{rec} - (z_s)_{em}}{c^2} = \frac{ah}{c^2}$$

when measured in the proper time of the atomic clocks. But the experimentally measured red shift was gh/c^2 to a precision of 1%. Hence, the downward acceleration of the inertial frames was the same as that of a free particle, $g = 980\ \text{cm s}^{-2}$, to a precision of 1%.

The Pound-Rebka-Snider experiment is the easiest red shift experiment to interpret theoretically because it was performed in a uniform gravitational field. Complementary to it is the experiment by Brault (Ref. 27), which measured the red shift of spectral lines emitted on the surface of the sun and received at earth. To a precision of 5% he found a red shift of $GM_\odot/R_\odot c^2$, where $M_\odot$ and $R_\odot$ are the mass and radius of the sun. This is just the red shift to be expected if the local Lorentz frames, at each point along the photon trajectory, are unaccelerated relative to freely falling test bodies. It certainly could not result if there were a single global Lorentz frame, extending throughout the solar system and at rest relative to its center of mass!**

In summary, the red-shift experiments reveal that, to a precision of $\sim 0.01\ GM/R^2$, where M and R are the mass and radius of the earth, the local Lorentz frames at the earth's surface are unaccelerated relative to freely falling test bodies. Equivalently, test bodies move along straight lines in the local Lorentz frames. Equivalently, the test-body trajectories are geodesics of the metric $\mathbf{g}$.

Because this conclusion is crucial to the foundations of metric theories, of the PPN framework, and of general relativity, it is important that the precision of the red-shift experiments be improved as much as possible — both on earth (homogeneous field) and elsewhere in the solar system (inhomogeneous fields). Of particular interest will be experiments in which atomic clocks are flown in space craft (see, e.g., Refs. 28, 29, and 30).

Our view that much effort should be put into future red shift experiments is disputed by Dicke and by Schiff. Dicke (Ref. 1, pp. 5 and 6) believes that "the gravitational red shift is not a very strong test of general relativity" because it can be derived from the weak equivalence principle, plus energy conservation, plus equivalence of inertial mass and conserved energy. The authors do not find Dicke's argument fully compelling. The fact that general relativity has no satisfactory local energy conservation law, except in static external gravitational fields, makes us worry about the a priori assumption of energy conservation. More importantly, we see no convincing a priori arguments why the inertial mass must equal the conserved energy to the precision required by Dicke's argument. In fact, this is not true in some theories with two tensor fields (see Ref. 31; we thank Professor Dicke for pointing this out to us). Finally, there exist a variety of relativistic gravitation theories which have been considered viable and attractive at one time or another but which disagree with the gravitational red-shift experiments (see p. 100 of Ref. 2).

*For an elementary derivation and discussion of this transformation law between Lorentz frames and accelerated frames, see, e.g., chapter 6 of Ref. 26.

**See chapter V of Ref. 2 for further discussion of this point

Schiff (Ref. 32, item 2 in "note added in proof") has suggested that the red shift can be derived from the Dicke-Eötvös experiment plus fundamental concepts of quantum theory. He was working on the details of such a derivation in January 1971, at the time of his tragic death. If others succeed in completing his derivation, then the importance of red-shift experiments will be diminished. (For further discussion see D2 below.)

2. The Response of Stressed Matter to Gravity

To discuss solar system tests of gravity, one must assume something about the response of stressed matter (e.g., the matter inside planets) to gravity. Metric theories all assume that, as in special relativity (gravity absent), so also in the real world where gravity is present,

$$\underset{\sim}{\nabla} \cdot \mathbf{T} = 0 \qquad (5)$$

Here $\mathbf{T}$ is the total stress-energy tensor for all matter and non-gravitational fields, and $\underset{\sim}{\nabla}\cdot$ is the divergence with respect to the metric $\mathbf{g}$ and its affine connection $\{{}^{a}_{bc}\}$.

Unfortunately, we do not have a firm experimental basis for the validity of Eq. (5) in the presence of gravity. However, we can make it seem reasonable — perhaps even compelling — by the following argument.

Geodesic motion for test bodies and $\underset{\sim}{\nabla} \cdot \mathbf{T} = 0$ for stressed matter go hand-in-hand. In particular, from the assumption $\underset{\sim}{\nabla} \cdot \mathbf{T} = 0$ we can derive geodesic motion (see, e.g., Ref. 33). From geodesic motion, i.e., straight-line motion in local Lorentz frames, we can derive $\underset{\sim}{\nabla} \cdot \mathbf{T} = 0$ for the smeared-out stress-energy tensor of a swarm of noninteracting test particles. For test particles that interact only by means of instantaneous collisions, each of which conserves energy and momentum in the local Lorentz frames, geodesic motion again guarantees $\underset{\sim}{\nabla} \cdot \mathbf{T} = 0$.

Unfortunately, one cannot prove that geodesic motion implies $\underset{\sim}{\nabla} \cdot \mathbf{T} = 0$ in all circumstances. The closest one has come is the following: Consider a laboratory-sized object made of stressed material. Geodesic motion and conservation of rest mass mean that the body's 4-momentum is conserved, as seen in any local Lorentz frame:

$$\int_{\Sigma} T^{ab}\, dS_b = P^a \text{ is independent of } \Sigma \qquad (6)$$

Here Σ is any spacelike 3-surface, contained entirely within the local Lorentz frame, which passes all the way through the body.* Using Stokes' theorem in the local Lorentz frame, we can infer from Eq. (6) that

$$\int_{\nu} T^{ab}{}_{,b}\, d\nu = \int_{\nu} (\underset{\sim}{\nabla} \cdot \mathbf{T})^a\, d\nu = 0 \qquad (7)$$

Here ν is any 4-volume contained entirely within the local Lorentz frame, which is intersected by all parts of the body. Equation (7) is equivalent to geodesic motion. The most straightforward way to guarantee the validity of Eq. (7) is by imposing $\underset{\sim}{\nabla} \cdot \mathbf{T} = 0$. But that is not the only way. For example, if $\underset{\sim}{n}$ and $\underset{\sim}{w}$ are vector fields whose variation through the body is completely negligible, if $\underset{\sim}{n}$ is spacelike, and if $T = T^{a}_{a}$ is the trace of the stress-energy tensor, then

$$\underset{\sim}{\nabla} \cdot \mathbf{T} + \underset{\sim}{w}(\underset{\sim}{n} \cdot \underset{\sim}{\nabla} T) = 0 \qquad (8)$$

would imply (7) and thence geodesic motion. However, there is no obvious, satisfactory way to pick out the vectors $\underset{\sim}{n}$ and $\underset{\sim}{w}$.

It is tempting, as another alternative to $\underset{\sim}{\nabla} \cdot \mathbf{T} = 0$, to demand that $\underset{\sim}{D} \cdot \mathbf{T} = 0$, where $\underset{\sim}{D}\cdot$ is the covariant derivative with respect to some affine connection Γ^{a}_{bc} different from $\{{}^{a}_{bc}\}$ — for example, $\{{}^{a}_{bc}\}$ plus a torsion. However, Thorne and Will (Ref. 3, appendix) show that this is untenable.

It is important to seek, in the future, direct experimental proof that $\underset{\sim}{\nabla} \cdot \mathbf{T} = 0$. To the accuracy of all laboratory experiments performed thus far (i.e., measurements of the behavior of stressed bodies in the earth's gravitational field), $\underset{\sim}{\nabla} \cdot \mathbf{T} = 0$ is true. But these experiments are probably not of sufficiently high precision for the purposes of the PPN formalism.**

D. Every Currently Viable Theory of Gravity is a Metric Theory

So far as the authors know, every theory of gravity which is currently viable is a metric theory. Of course, this statement is meaningless without a definition of "viable." By "viable" we mean (1) self-consistent, (2) complete, and (3) compatible with all experiments performed in the past. These three criteria, and theories which violate them, are discussed in the following three sections.

1. Self-Consistency

The classic example of an internally inconsistent theory is the spin-two field theory of gravity, derived with standard field theoretic techniques by Fierz and Pauli (Ref. 34). (For a brief review,

*We ignore small corrections due to the Christoffel symbols $\{{}^{a}_{bc}\}$, which vanish only at the origin of the local Lorentz frame. Clearly those corrections go to zero linearly with L, the size of the space-time region under consideration — i.e., the "size of the local Lorentz frame."

**Note the great difference in spirit between the above discussion and the usual viewpoint. One usually assumes $\underset{\sim}{\nabla} \cdot \mathbf{T} = 0$; and when confronted by any apparent violation of it (e.g., the apparent breakdown in energy-momentum conservation in beta decay), one normally seeks a modification of the stress-energy tensor $\mathbf{T}$ which will then restore the validity of $\underset{\sim}{\nabla} \cdot \mathbf{T} = 0$ (e.g., Pauli's 1930 postulate of the existence of neutrinos). By contrast, we are assuming (without much justification) that all the contributions to $\mathbf{T}$ are known, and that the metric and covariant derivative $\underset{\sim}{\nabla}$ are known; and we are then asking whether $\underset{\sim}{\nabla} \cdot \mathbf{T} = 0$.

see chapter 7 of Ref. 26.) This theory is identical to "Linearized General Relativity" — which one obtains by linearizing about flat space-time. It has a flat background metric η_{ij} and a symmetric second-rank gravitational field h_{ij}. The field equations of the theory demand that, in a Lorentz frame of the background metric, the stress-energy tensor have vanishing coordinate divergence, $T^{ij}{}_{,j} = 0$. From this, one infers that test particles move along straight lines in the background metric, which means they are affected not at all by gravity: $d^2x^\alpha/d\tau^2 = 0$ — a result contradicting the equations of motion of the theory,

$$(\eta_{\mu\nu} + h_{\mu\nu}) \frac{d^2x^\nu}{d\tau^2}$$

$$= -\frac{1}{2}\left(\frac{\partial h_{\nu\alpha}}{\partial x^\beta} + \frac{\partial h_{\nu\beta}}{\partial x^\alpha} - \frac{\partial h_{\alpha\beta}}{\partial x^\nu}\right)\frac{dx^\alpha}{d\tau}\frac{dx^\beta}{d\tau}$$

This internal inconsistency of the theory can be removed by modifying the Lagrangian from which the theory is derived. The resultant modified theory turns out to be identical to general relativity (see papers by Gupta, Thirring, Feynman, Weinberg, and Deser cited in chapter 7 of Ref. 26).

Another example of a theory with internal inconsistencies is that of Kustaanheimo (Ref. 35). It predicts zero gravitational red shift when the wave version of light (Maxwell theory) is used, and non-zero red shift when the particle version (photon) is used. Such an inconsistency is fairly common in "flat-space theories of gravity." By "flat-space theory" we mean a theory with a global Minkowskii metric η_{ij} that governs atomic clock rates and rod lengths. Such theories must predict zero gravitational red shift when the wave version of light is used (see, e.g., C1 above, Ref. 2, and chapter 7 of Ref. 26). But their equations of motion typically predict the standard red shift (Δ energy/energy $= -g\Delta h$) for particles of zero rest mass. No such theories are viable. However, most of them can be (and have been) salvaged by a physical reinterpretation. The Minkowskii metric is replaced, as regulator of clock rates and rod lengths, by a curved-space metric which is constructed from it and from the "gravitational field." For example, Whitehead's original theory of gravity (Ref. 36) had a Minkowskii metric η_{ij} which governed clock rates and rod lengths; and it had a symmetric second-rank tensor h_{ij} as its gravitational field. To make its wave-calculated red shift come out right, Synge (Ref. 37) reinterpreted $g_{ij} \equiv \eta_{ij} + h_{ij}$ as the metric that governs clock rates and rod lengths.

2. Completeness

To be complete, a theory of gravity must be capable of analyzing from "first principles" the outcome of every experiment of interest. Of course, this requires that it mesh with and incorporate a consistent set of electromagnetic laws, quantum mechanical laws, etc.

No theory is complete if it postulates that atomic clocks measure the "interval" $ds = (g_{ij}\,dx^i dx^j)^{1/2}$ constructed from a particular metric. Atomic clocks are complex systems whose behavior must be calculated from the fundamental laws of quantum theory and electromagnetism.

No theory is complete if it postulates that planets move on geodesics. Planets are complex systems whose motion must be calculated from fundamental laws for the response of stressed matter to gravity.

Most theories of gravity proposed in the past were incomplete. Only subsequently were they put into forms which meshed consistently with the other laws of physics, so that the behavior of complex systems could be calculated. So far as the authors know, all cases of successful meshing have succeeded by identically the same approach: (1) the introduction of a curved space-time metric g_{ij} and (2) the demand that all special relativistic laws of physics be valid in the metric's local Lorentz frames (strong equivalence principle). Two consequences of this universal approach are that (1) atomic clocks and laboratory rods measure the interval whose square is $ds^2 = g_{ij}\,dx^i dx^j$ and (2) stressed matter and (non-gravitational) fields respond to gravity in accordance with the law $\underset{\sim}{\nabla} \cdot \mathbf{T} = 0$, where $\underset{\sim}{\nabla}\cdot$ is the metric's covariant derivative. Hence, successful meshing has always yielded a metric theory of gravity.

To construct a theory of gravity which meshes consistently in some other manner with the laws of electromagnetism and quantum theory is an intriguing theoretical challenge. So far as the authors know, nobody has ever succeeded. Since the only known way of meshing produces metric theories, and since metric theories which give the correct Newtonian orbits for freely falling particles near earth must have $g_{00} = 1 - 2GM/r$ and must thus give the correct red shift, red-shift experiments might seem less important than was claimed in C1 above. However, the authors retain a belief in the importance of red-shift experiments, since the failure of theorists to find another meshing scheme is no proof at all that other schemes are impossible.

3. Compatibility with Experiments

The necessity that a theory agree, to within several standard deviations, with the "four standard tests" (gravitational red shift, perihelion shift, electromagnetic-wave deflection, and radar time delay) is obvious. Equally obvious but often forgotten is the need to agree with observations at the more every-day, Newtonian level. Example: Birkhoff's theory (Ref. 38) predicts the same red shift, perihelion shift, deflection, and time delay as general relativity. But it requires that the pressure inside gravitating bodies equal the total density of mass-energy, $p = \rho$, and as a consequence, it demands that sound waves travel with the speed of light. Of course, this disagrees violently with experiment, so Birkhoff's theory is not viable.

Table 1 lists many non-viable theories of gravity and explains at least one difficulty with each. It is only a preliminary list; other theories, which the authors have not yet studied closely, will turn out to be non-viable.

Table 1. Non-viable theories of gravity: a partial list[a]

Theory and references	Comments, including some but not all the reasons why the theory is non-viable
Hoyle's C-field theory (Refs. 39, 40, 41)	Was devised originally as a foundation for the steady-state model of the universe. Is incomplete – lacks an equation governing the rate at which the C-field creates particles.
Milne's kinematical relativity (Refs. 42, 43)	Was devised originally to handle certain cosmological problems. Is incomplete – makes no red-shift prediction; predicts zero deflection of light.
Poincare's theory (Ref. 44) as generalized by Whitrow and Morduch (Ref. 45)	Action-at-a-distance theory in flat space time with an adjustable parameter n. For $n < 2$ predicts zero red shift. For $n \geq 2$ gives internally inconsistent treatment of light propagation.
Whitrow-Morduch vector theory (Refs. 45, 46)	Contains a vector gravitational field that resides in flat space time; possesses a freely specifiable parameter p. For $p = 0$, predicts no red shift. For $p \neq 0$, gives internally inconsistent treatment of light propagation.
Kustaanheimo's various vector theories (Refs. 35, 45, 46, 47, 48)	Contain a vector gravitational field that resides in flat space time; possess several freely specifiable functions, which can be adjusted to give agreement with the "four standard tests." Are incomplete – do not mesh with other laws of physics, except by imposing them in the global Lorentz frame of the flat space-time metric, which then gives internal inconsistencies: different red shifts for light viewed as photons and light viewed as waves; non-zero red shifts for photons between points at the same gravitational potential. Also several scalar and tensor theories which suffer the same problems.
Birkhoff's theory (Ref. 38)	Gives same prediction as general relativity for four standard tests; but predicts crazy results for internal behavior of matter – $p = \rho c^2$, $v_{sound} = v_{light}$.
Rastall's theory (Refs. 49, 50)	Is such a complicated theory mathematically that it is far from obvious whether the field equations have unique solutions – or any solution at all – for given initial data. Until the initial value problem has been treated properly, we (the authors) consider this a non-viable theory.
Yilmaz's theory (Refs. 51, 52)	Scalar-metric theory. Gives same prediction as general relativity for four standard tests, but does not have complete account of non-static gravitational fields. Modified version of the theory is included in Tables 2 and 6.

[a]These theories are non-viable in their present forms. Future modifications or specializations might make some of them viable. If we have misinterpreted any theory here, we apologize to its proponents, and we urge them to demonstrate explicitly its self-consistency, completeness, and correct experimental predictions.

E. A Partial Catalogue of Metric Theories

All metric theories contain a metric g_{ij}, which governs the lengths of laboratory rods and the ticking of atomic clocks in accordance with $ds^2 = g_{ij}dx^i dx^j$, and which influences the motion of stressed matter in accordance with $\nabla \cdot \mathbf{T} = 0$. In only one significant way do various metric theories differ: their law for the generation of the metric. In general relativity theory, the metric is generated directly by the stress-energy of matter and of non-gravitational fields. In the scalar-tensor theories of Brans and Dicke (Ref. 11), and of Bergmann (Ref. 53) (as specialized and made complete by Wagoner, Ref. 54), the matter and non-gravitational fields generate a scalar gravitational field ϕ; then ϕ acts together with the matter and fields to generate the metric. In other theories, e.g., that of Whitehead (Ref. 36), there exists a global Lorentz metric η_{ij}, which does not affect matter directly, in addition to the physical metric g_{ij}. The Lorentz metric acts together with the stress-energy and non-gravitational fields to produce the physical metric. In still other theories, e.g., the generalized Nordström theory (Refs. 45 and 55), a global Lorentz metric η_{ij}, together with stress-energy, generates a scalar field ϕ; then ϕ and η_{ij} work together to generate the physical metric g_{ij}. All these examples and others are spelled out in Table 2.

Table 2. A partial catalogue of metric theories of gravity[a]

Theory and references[b]	Entities present	Generation scheme	Arbitrary parameters and functions
1. General relativity (Ref. 56)	T, g	$T \xrightarrow{W} g$	None.
2. Scalar-tensor theories	T	$\mathrm{tr}(T) \xrightarrow{W} \phi$	General case (Bergmann-Wagoner): two arbitrary functions of ϕ; in post-Newtonian limit (without cosmological function) — two arbitrary parameters, ω and Λ.
a. General case (Refs. 53[c], 54)	ϕ	$(T, \phi) \xrightarrow{W} g$	
	g		
b. Special cases (Refs. 11 [same as Ref. 57, with $\eta = -1$, $b = 0$], 58)			Brans-Dicke-Jordan case: one arbitrary parameter, ω; in P-N limit — $\Lambda = 0$.
			Nordtvedt case: one arbitrary function of ϕ; in P-N limit — ω and Λ.
3. Conformally flat theories of the Nordstrøm type	T	$(T, \eta) \xrightarrow{W} \phi$	General case (Ni): two arbitrary functions of ϕ; in P-N limit — two arbitrary parameters, p and q.
	η	$(\phi, \eta) \xrightarrow{L} g$	
a. General case (Ref. 55)	ϕ		Whitrow-Morduch case: one arbitrary function of ϕ; in P-N limit — q arbitrary, $p = -4$.
b. Special cases (Refs. 45, 46, 59, 60, 61, 62, 63, 64)	g		Littlewood-Bergmann case: none; in P-N limit — $q = 1/2$, $p = -4$.
			Nordstrøm case (Ref. 61): none; in P-N limit — $q = 0$, $p = -4$.
			Nordstrøm-Einstein-Fokker case: none; in P-N limit — $q = 1/2$, $p = -3$.
4. Stratified theories with conformally flat space slices (Ref. 55)	T, η, t	$(T, \eta) \xrightarrow{W} \phi$	
	ϕ, g	$(\phi, \eta, t) \xrightarrow{L} g$	
a. Einstein (Ref. 65)			a. None
b. Whitrow and Morduch (Refs. 45, 46) (They call it "theory with variable velocity of light.")			b. None
c. Modified Yilmaz (Refs. 51, 52, 55)			c. None
d. Ni's Lagrangian-based theory (Ref. 55)			d. None
e. Ni's general stratified theory (Ref. 55)			e. Two arbitrary functions; in P-N limit — two arbitrary parameters, p and q.

Table 2. (contd)

Theory and references[b]	Entities present	Generation scheme	Arbitrary parameters and functions
5. Action-at-a-distance theories with Lorentz invariance	η, T	$(\eta, T) \xrightarrow{AD} h$	
	h, g	$(\eta, h) \xrightarrow{L} g$	
a. Whitehead (Ref. 36); see also Synge (Ref. 37) and Schild (Ref. 66)			a. None.
b. Schild's generalization of Whitehead (Ref. 66)			b. Two arbitrary functions.

[a]Notation:

η = a "background" Lorentz metric whose existence is postulated by the theory.

ϕ and h = scalar and symmetric tensor gravitational fields, generated by stress-energy, which help to generate the physical metric, but which do not act back directly on matter or non-gravitational fields.

t = a scalar field which plays the role of a preferred, universal time coordinate.

g = the physical metric which governs clock rates and rod lengths, and to which matter responds via $\nabla \cdot T = 0$.

T = the total stress-energy tensor for all matter and non-gravitational fields.

$tr(T)$ = the trace of the stress-energy tensor.

$B \xrightarrow{W} C$ means that B generates the field C via a wave-type equation, which has homogeneous wave-type solutions representing a freely propagating B field.

$B \xrightarrow{AD} C$ means that B generates the field C via a Lorentz-invariant action-at-a-distance equation.

$B \xrightarrow{L} C$ means that B generates C in a local, algebraic manner; e.g., $(\phi, \eta) \xrightarrow{L} g$ might represent the generation equation $g_{ij} = e^{\phi} \eta_{ij}$.

[b]Some of these theories were incomplete (cf. IIID) in their original forms. Ni (Ref. 55) has completed them and converted them into metric theories by making minor modifications. A more detailed version of this catalogue will be given in Ref. 55.

[c]Bergmann's theory (Ref. 53) was not complete in its original form. Wagoner's version (Ref. 54) completed it by assuming a particular form for the coupling of matter in the field equations.

(Note that Table 2 is not a complete catalog; other metric theories will be added after the authors have studied them more closely.)

III. Parametrized Post-Newtonian (PPN) Framework

A. Description of the Framework

The solar system, where experiments to distinguish between metric theories are performed, has weak gravity,

$$U = \text{(Newtonian potential)} \lesssim 10^{-6}$$

Moreover, the matter which generates its gravity moves slowly,

$$v^2 = \text{(velocity relative to solar system center of mass)}^2 \lesssim 10^{-7}$$

and has small stress and internal energies:

$$\frac{|T_{\alpha\beta}|}{\rho_o} = \text{(stress divided by density of rest mass-energy)} \lesssim 10^{-6}$$

$$\Pi = \frac{\rho - \rho_o}{\rho_o} = \text{(internal energy density per unit rest mass-energy)} \lesssim 10^{-6}$$

Consequently, the analysis of solar system experiments using any metric theory of gravity can be simplified, without significant loss of accuracy, by a simultaneous expansion in the small parameters U, v^2, $|T_{\alpha\beta}|/\rho_o$, and Π. Such a "weak-field, slow-motion expansion" gives: (1) flat, empty space time at "zero order," (2) the Newtonian treatment of the solar system at "first order," and (3) post-Newtonian corrections to the Newtonian treatment at "second order."

The formalism of Newtonian theory plus post-Newtonian corrections is called the "post-Newtonian limit." Each metric theory has its own post-Newtonian limit. Despite the great differences between the metric theories themselves, their post-Newtonian limits are very similar; so similar, in fact, that one can construct a single post-Newtonian theory of gravity, devoid of any reference to the fields ϕ, $\mathbf{h}$, $\boldsymbol{\eta}$ of Table 2, which contains the post-Newtonian limit of each metric theory as a special case. This all-inclusive post-Newtonian theory is called the "Parametrized Post-Newtonian (PPN) framework." It contains a set of parameters (called PPN parameters), which can be specified arbitrarily. One set of values for these parameters makes the PPN framework identical to the post-Newtonian limit of general relativity; another set of values makes it the post-Newtonian limit of the Brans-Dicke theory; etc.

How many parameters does the PPN framework need in order to reproduce the post-Newtonian limit of every metric theory? Nine parameters suffice, if one is satisfied to treat all theories listed in Table 2, except Whitehead and Generalized Whitehead. The Whitehead theories contain complexities unknown to other catalogued theories, so they require an extension of the PPN framework beyond nine parameters.

The restricted, nine-parameter framework which the authors use was divised by Will (Ref. 5), and is also presented in Part 9 of Ref. 26. A very brief exposition of it is given in Table 3 of this paper. Table 4 contains a list and heuristic description of its parameters.

Table 3. The nine-parameter PPN framework (Ref. 5; see also Part 9 of Ref. 26)

Coordinate system	The framework uses a particular, nearly globally Lorentz coordinate system in which the solar system's center of mass is at rest, or nearly so. The coordinates are (t, x^1, x^2, x^3); and three-dimensional, Euclidean vector notation is used throughout. All coordinate arbitrariness ("gauge freedom") has been removed by specialization of the coordinates.
Units	Geometrized units are used. These are units in which the speed of light and the Newtonian gravitational constant (as measured today far from the solar system in a frame at rest with respect to the distant galaxies) are unity ($c = G = 1$).
Matter variables	1. ρ_o = density of rest mass measured in a local Lorentz frame momentarily comoving with the gravitating matter. 2. $v_\alpha = (dx^\alpha/dt)_{\text{comoving}}$ = coordinate velocity of the matter; to Newtonian precision (fractional errors $\lesssim 10^{-6}$) this can also be interpreted as the "proper velocity" or the "Newtonian velocity" or any other kind of velocity desired. 3. $t_{\alpha\beta}$ = stress tensor as measured in a local Lorentz frame momentarily comoving with the matter. 4. $p = 1/3(\sum_\alpha t_{\alpha\alpha})$ = pressure as measured in a local Lorentz frame momentarily comoving with the matter. This isotropic pressure is the dominant part of the stress. The much smaller anisotropies are important in maintaining the deformations of planets but generate completely negligible post-Newtonian corrections in the solar system. 5. Π = internal energy per unit rest mass. It includes all forms of non-rest-mass, non-gravitational energy — e.g., energy of compression and thermal energy. It obeys the first law of thermodynamics: $\rho_o d\Pi/dt = (p/\rho_o) d\rho_o/dt$

Table 3. (contd)

PPN parameters (see also Table 4)	$\gamma,\ \beta,\ \beta_1,\ \beta_2,\ \beta_3,\ \beta_4,\ \Delta_1,\ \Delta_2,\ \zeta$
Metric	$g_{oo} = 1 - 2U + 2\beta U^2 - 4\Phi + \zeta\mathcal{A}$ $g_{o\alpha} = (7/2)\,\Delta_1 V_\alpha + (1/2)\,\Delta_2 W_\alpha$ $g_{\alpha\beta} = -(1 + 2\gamma U)\,\delta_{\alpha\beta}$ where $U(\underset{\sim}{x}, t) = \int \frac{\rho_o(\underset{\sim}{x}', t)}{\lvert \underset{\sim}{x} - \underset{\sim}{x}' \rvert} d\underset{\sim}{x}'$ = Newtonian potential $\Phi(\underset{\sim}{x}, t) = \int \frac{\rho_o(\underset{\sim}{x}', t)\Phi(\underset{\sim}{x}', t)}{\lvert \underset{\sim}{x} - \underset{\sim}{x}' \rvert} d\underset{\sim}{x}'$ $\phi = \beta_1 v^2 + \beta_2 U + (1/2)\beta_3 \Pi + (3/2)\beta_4 p/\rho_o$ $\mathcal{A}(\underset{\sim}{x}, t) = \int \frac{\rho_o(\underset{\sim}{x}', t)\left[(x_\alpha - x'_\alpha)v_\alpha(\underset{\sim}{x}')\right]^2}{\lvert \underset{\sim}{x} - \underset{\sim}{x}' \rvert^3} d\underset{\sim}{x}'$ $V_\alpha(x, t) = \int \frac{\rho_o(\underset{\sim}{x}', t)v_\alpha(\underset{\sim}{x}')}{\lvert \underset{\sim}{x} - \underset{\sim}{x}' \rvert} d\underset{\sim}{x}'$ $W_\alpha(x, t) = \int \frac{\rho_o(\underset{\sim}{x}', t)v_\beta(\underset{\sim}{x}')\,(x_\beta - x'_\beta)\,(x_\alpha - x'_\alpha)}{\lvert \underset{\sim}{x} - \underset{\sim}{x}' \rvert^3} d\underset{\sim}{x}'$
Stress-energy tensor	$T^{oo} = \rho_o(1 + \Pi + v^2 + 2U)$ $T^{o\alpha} = \rho_o(1 + \Pi + v^2 + 2U)\,v_\alpha + t_{\alpha\beta}v_\beta$ $T^{\alpha\beta} = t_{\alpha\beta}(1 - 2\gamma U) + \rho_o(1 + \Pi + v^2 + 2U)v_\alpha v_\beta + 1/2(v_\alpha t_{\beta\gamma}v_\gamma + v_\beta t_{\alpha\gamma}v_\gamma)$
Equations of motion	$\underset{\sim}{\nabla} \cdot \mathbf{T} = 0$, i.e., $T^{ij}_{;j} = 0$, i.e., $T^{ij}_{,j} + \left\{ {i \atop jk} \right\} T^{kj} + \left\{ {j \atop kj} \right\} T^{ik} = 0$, where $\left\{ {i \atop jk} \right\}$ are Christoffel symbols calculated from the metric.

Table 4. Heuristic description of the nine PPN parameters

Parameter	What it measures, relative to general relativity	Value in general relativity
γ	How much space curvature ($g_{\alpha\beta}$) is produced by unit rest mass?	1
β	How much nonlinearity is there in the superposition law for gravity (g_{oo})?	1
β_1	How much gravity (g_{oo}) is produced by unit kinetic energy $[(1/2)\rho_o v^2]$?	1
β_2	How much gravity (g_{oo}) is produced by unit gravitational potential energy ($\rho_o U$)?	1
β_3	How much gravity (g_{oo}) is produced by unit internal energy ($\rho_o \Pi$)?	1
β_4	How much gravity (g_{oo}) is produced by unit pressure (p)?	1
ζ	How much more gravity (g_{oo}) is produced by radial kinetic energy $[(1/2)\rho_o (\underline{v} \cdot \underline{\hat{r}})^2]$ — i.e., kinetic energy of motion toward observer — than by transverse kinetic energy?	0
Δ_1	How much dragging of inertial frames ($g_{o\alpha}$) is produced by unit momentum ($\rho_o v_\alpha$)?	1
Δ_2	How much easier is it for momentum ($\rho_o v_\alpha$) to drag inertial frames radially (toward the observer) than in a transverse direction?	1

A more specialized PPN framework, which Will used as a guide in constructing his, was devised by Nordtvedt (Ref. 4), and is used by Nordtvedt in his paper in this proceedings. Nordtvedt's framework can be regarded as a "point-particle limit" of Will's framework. Whereas Will treats the sun, planets, and moon as made of stressed, continuous matter, Nordtvedt treats them as made from "gases" of point particles (atoms) which interact gravitationally and electromagnetically. The relationship between Will's parameters and Nordtvedt's parameters is given in Table 5.

B. Accuracy of the Framework

How accurate is the PPN framework? Or, stated more precisely, how accurately does the post-Newtonian limit agree with the metric theory from which it comes? In the solar system, where U, v^2, $|T_{\alpha\beta}|/\rho_o$, and Π are all $\lesssim 10^{-6}$, the post-Newtonian limit makes fractional errors of $\lesssim 10^{-12}$ in post-Newtonian-order quantities, and fractional errors of $\lesssim 10^{-6}$ in Newtonian-order quantities. For example, it misrepresents the deflection of light by $\lesssim 10^{-6} \times$ (post-Newtonian deflection) $\sim 10^{-6}$ arc-s. And it ignores relativistic deformations of the earth's orbit of magnitude $< 10^{-12} \times$ (one astronomical unit) ~ 10 cm. Clearly, there is no need in the 1970s to use higher-order corrections to the post-Newtonian limit, and hence, no need to construct a "parametrized post-post-Newtonian framework."

Actually, there are a few exceptions to the claim that the post-Newtonian limit suffices. They occur where the external universe impinges on and influences the solar system. For example, gravitational waves propagating into the solar system from distant sources are ignored by every post-Newtonian limit and by the PPN framework. They must be treated using a full metric theory or a weak-field, "fast-motion" approximation to one. Similarly, time dependence of the "gravitational constant," induced by expansion of the universe, is beyond the scope of the PPN framework.

C. PPN Parameters for Various Metric Theories

Ni (Ref. 55) and Will (Refs. 5 and 67) have calculated the values of the nine PPN parameters corresponding to each metric theory catalogued in Table 2. Those values are listed in Table 6.

Table 6 shows some surprising regularities in the PPN parameter values. For example, in every theory listed, $\Delta_2 = 1$, $\zeta = 0$, and $\beta_3 = 1$; and in several, $\beta_4 = \gamma$ and $2\beta_1 = \gamma + 1$. Nordvedt (Ref. 68) and Will (Refs. 69 and 70) have shown that one can put theory-independent constraints such as these on the values of the parameters by demanding, for example, that the predicted flat-space speed of propagation of gravity should be the same as that of light for any theory of gravity. This is a special case of a more stringent demand, stated in the following terms: the post-Newtonian metric of any well behaved theory of gravity (when written in a particular gauge) should be invariant under a Lorentz transformation of the coordinate system far from the matter. This demand leads to the following constraints on the PPN parameters:

Table 5. Relationship between Will's PPN parameters (Ref. 5) and those of Nordtvedt (Ref. 4)[a]

Will	Nordtvedt
γ	γ
β	β
β_1	$1/4\ (4\alpha'' + 1 - \chi)$
β_2	$1/2\ (3\gamma - \alpha')$
β_3	Absent (because of point-particle approximation)
β_4	Absent (because of point-particle approximation)
ζ	$\alpha''' - \chi$
Δ_1	$(8/7)\Delta - (1/7)\chi$
Δ_2	$8\Delta' + \chi$

[a] These relations cannot be inverted to give Nordtvedt's parameters in terms of Will's because one of Nordtvedt's parameters is arbitrary, corresponding to an arbitrariness in his coordinate system.

Table 6. Values of PPN parameters for metric theories catalogued in Table 2 [based on calculations by Ni (Ref. 55), Will (Ref. 5), and Nordtvedt (Ref. 58)]

Theory and its parameters	γ	β	β_1	β_2	β_3	β_4	ζ	Δ_1	Δ_2
1. General relativity (none)[a]	1	1	1	1	1	1	0	1	1
2. Scalar-tensor theories (ω, Λ)[a]	$\frac{1+\omega}{2+\omega}$	$1+\Lambda$	$\frac{3+2\omega}{4+2\omega}$	$\frac{1+2\omega}{4+2\omega} - \Lambda$	1	$\frac{1+\omega}{2+\omega}$	0	$\frac{10+7\omega}{14+7\omega}$	1
3. Conformally flat theories of Nordstrom type (p, q)	-1	1 - q	0	$\frac{1}{2}p$	1	0 or -1[b]	0	$-\frac{1}{7}$	1
4. Stratified theories with conformally flat space slices									
a. Einstein	0	0	0	0	1	0	0	$-\frac{1}{7}$	1
b. Whitrow-Murduch	0	-1	0	0	1	0	0	$-\frac{1}{7}$	1
c. Modified Yilmaz[a]	1	1	0	0	1	0	0	$-\frac{1}{7}$	1
d. Ni's Lagrangian-based theory[a]	1	1	1	1	1	1	0	$-\frac{1}{7}$	1
e. Ni's second theory (p, q)[a]	1	1 - q	0	$\frac{1}{2}p$	1	0 or -1[b]	0	$-\frac{1}{7}$	1

Table 6. (contd)

Theory and its parameters	γ	β	β_1	β_2	β_3	β_4	ζ	Δ_1	Δ_2
5. Whitehead and Generalized Whitehead	Require more than nine parameters for their description.								

[a]These theories are currently viable. All others in the table disagree violently with one experiment or another (see Table 7).

[b]The value of β_4 depends on what generates the scalar field of the theory: rest mass ($\beta_4 = 0$) or the trace of the stress-energy tensor ($\beta_4 = -1$).

$$\Delta_2 + \zeta - 1 = 0$$

$$4\beta_1 = 2\gamma + 2 + \zeta \qquad (9)$$

$$7\Delta_1 + \Delta_2 = 4\gamma + 4$$

(see Ref. 70). These constraints are violated by all the scalar-metric theories in Table 6 (theories 4a, b, c, d, e).

One can also insist that any well behaved theory of gravity should have integral conservation laws. If one demands that an isolated gravitating system (in otherwise empty space) should possess a conserved four-momentum P^i and a conserved angular momentum J^{ij} (which contains six conserved quantities — three for the ordinary angular momentum $[J^{\alpha\beta}]$ and three expressing uniform motion of the center of mass $[J^{0\alpha}]$) — then one is led (Ref. 70) to the following constraints on the parameters:

$$\beta_1 = \tfrac{1}{2}(\gamma + 1) \qquad \zeta = 0$$

$$\beta_2 = \tfrac{1}{2}(3\gamma - 2\beta + 1) \qquad \Delta_1 = \tfrac{1}{7}(4\gamma + 3) \qquad (10)$$

$$\beta_3 = 1 \qquad \Delta_2 = 1$$

$$\beta_4 = \gamma$$

These constraints are satisfied only by general relativity, by scalar-tensor theories, and by Nordström-type theories with $p = 2q - 4$ — and perhaps by other theories not treated in this paper. Of course, these are purely theoretical constraints on the values of the parameters, and should themselves be subjected to experimental scrutiny (see, for example, IVC4, and Ref. 69).

IV. Experiments and Their PPN Interpretations

One can regard solar system gravity experiments as attempts to measure the PPN parameters and thereby disprove as many theories as possible. In the following sections we list a variety of experiments performed in the past or performable in the 1970s, and we describe how their outcomes depend upon the values of the PPN parameters.

A. Experiments to Measure γ

[Recall that γ measures how much space curvature is produced by unit rest mass.]

1. Deflection of Electromagnetic Waves

According to general relativity, electromagnetic waves (light) passing near the sun should be deflected through an angle of

$$\alpha_{GR} = \frac{+GM_\odot}{\ell} = (1''.75 \text{ arc})\frac{R_\odot}{\ell} \qquad (11)$$

where $M_\odot$ and $R_\odot$ are the sun's mass and radius, and ℓ is the impact parameter of the photons (Ref. 56). The PPN framework predicts the alternative deflection angle

$$\alpha_{PPN} = \tfrac{1}{2}(1 + \gamma)\alpha_{GR} \qquad (12)$$

(For derivation see, e.g., Part 9 of Ref. 26.) The quasar radio-wave experiments of Muhleman, Ekers, and Fomalont (Ref. 71) and of Seielstad, Sramek, and Weiler (Ref. 72) yield

$$\frac{\alpha_{EXP}}{\alpha_{GR}} = 1.04 \,{}^{+0.15}_{-0.10} \quad [\text{MEF}]$$

$$\frac{\alpha_{EXP}}{\alpha_{GR}} = 1.01 \pm 0.12 \quad [\text{SSW}] \qquad (13)$$

corresponding to

$$\gamma = 1.08 \begin{matrix} +0.30 \\ -0.20 \end{matrix} \quad [\text{MEF}]$$

$$\gamma = 1.02 \pm 0.24 \quad [\text{SSW}] \tag{14}$$

2. Radar Time Delay

A radar beam passing near the sun suffers a relativistic time delay with a complicated dependence on impact parameter, according to general relativity (Ref. 73; see also Refs. 74, 75, and 76). The dominant part of the general relativistic round-trip delay is

$$\Delta\tau_{GR} = 4GM_\odot \ln\left[\frac{\left(a_T + \sqrt{a_T^2 - \ell^2}\right)\left(a_R + \sqrt{a_R^2 - \ell^2}\right)}{\ell^2}\right] \tag{15}$$

where ℓ is the ray's impact parameter and a_T and a_R are the radii (distance from sun) of the transmitter and reflector. The PPN framework predicts the alternative delay

$$\Delta\tau_{PPN} = \frac{1}{2}(1 + \gamma)\Delta\tau_{GR} \tag{16}$$

(For derivation see, e.g., part 9 of Ref. 26.) The passive-radar experiments by Shapiro reported at this conference yield

$$\frac{\Delta\tau_{EXP}}{\Delta\tau_{GR}} = 1.015 \pm 0.05, \qquad \gamma = 1.03 \pm 0.10 \tag{17}$$

The Mariner VI and VII active-radar experiments reported at this conference by Anderson, Muhleman, and Esposito yield

$$\frac{\Delta\tau_{EXP}}{\Delta\tau_{GR}} = 1.02 \pm 0.04, \qquad \gamma = 1.04 \pm 0.08 \tag{18}$$

3. Geodetic Gyroscope Precession

A gyroscope in a satellite orbiting the earth should precess relative to the distant stars, according to general relativity (Ref. 77). The dominant part of the precession is due to the space curvature, which is regulated by γ in the PPN framework. This dominant part, called the "goedetic precession" or "de Sitter precession," has an angular velocity of

$$\underset{\sim}{\Omega}_{\text{geod.,GR}} = \frac{3}{2}\,\underset{\sim}{v} \times \underset{\sim}{\nabla} U \tag{19}$$

where $\underset{\sim}{v}$ is the velocity of the gyroscope and U is the Newtonian gravitational potential. The PPN framework predicts the alternative angular velocity of precession

$$\underset{\sim}{\Omega}_{\text{geod., PPN}} = \frac{1}{3}(1 + 2\gamma)\underset{\sim}{\Omega}_{\text{geod., GR}} \tag{20}$$

(For derivation see, e.g., Part 9 of Ref. 26; for further discussion see the papers of Everitt and O'Connell in this proceedings, and references cited therein.)

B. Experiments to Measure β and γ

1. Perihelion Shifts

The perihelion of a planetary orbit (or drag-free satellite orbit) should shift forward slightly each time the body "circles" its orbit, according to general relativity (Ref. 78). All nonrelativistic contributions to the perihelion shift are well known and can be subtracted out of the data, except the shift produced by the (unknown) solar quadrupole moment. The combined general relativistic shift and quadrupole shift per orbit is given by

$$\delta\phi = \frac{6\pi M_\odot}{a(1 - e^2)}\,\lambda_{P,GR};$$

$$\lambda_{P,GR} = 1 + J_2\,\frac{R_\odot^2/M_\odot}{2a(1 - e^2)} \tag{21}$$

Here a and e are the semimajor axis and eccentricity of the planetary orbit, and J_2 is a standard measure of the solar quadrupole moment defined by

$$U = (\text{Newtonian potential of sun})$$

$$= \frac{GM_\odot}{r}\left[1 - J_2\left(\frac{R_\odot}{r}\right)^2 \frac{3\cos^2\theta - 1}{2}\right] \tag{22}$$

(J_2 would be near 1 if the sun were rotating near break-up speed.) The PPN framework predicts the alternative shift per orbit

$$\delta\phi = \frac{6\pi M_\odot}{a(1 - e^2)}\,\lambda_{P,PPN};$$

$$\lambda_{P,PPN} = \frac{2 - \beta + 2\gamma}{3} + J_2\,\frac{R_\odot^2/M_\odot}{2a(1 - e^2)} \tag{23}$$

(For derivation see, e.g., Part 9 of Ref. 26.) Shapiro's experimental results, obtained by combining all available optical and radar data for several planets, and reported at this conference, are

$$(\lambda_P)_{\text{Mercury}} = 0.98 \pm 0.01 \tag{24}$$

$$J_2 \lesssim 3 \times 10^{-5} \tag{25}$$

(The limit on J_2 is obtained by comparing λ_P for different planets.) By combining these data with formula (23) for $\lambda_{P,\,PPN}$, we obtain

$$\frac{1}{3}(2 - \beta + 2\gamma) = 0.98 \,{}^{+\,0.01}_{-\,0.10} \tag{26}$$

and by combining this with the Mariner time-delay value for γ (Eq. 18), we obtain

$$\beta = 1.14 \,{}^{+\,0.2}_{-\,0.3} \tag{27}$$

2. Periodic Effects in Orbits

General relativity produces periodic, noncumulative perturbations in planetary and drag-free satellite orbits (Ref. 79). The perturbations have amplitudes $\sim GM_\odot \sim 1$ km, independently of how far the body is away from the sun. In the PPN framework these perturbations will depend on the same parameters, β and γ, as the perihelion shift; but the precise nature of the dependence has never been calculated.

C. Experiments to Measure Other Parameters

1. Frame-Dragging Gyroscope Precession

According to general relativity the earth, as it rotates, should "drag along" the inertial frames near it. This (very small) frame dragging should produce Coriolis forces in any reference frame fixed relative to the distant stars, and it should therefore cause a gyroscope to precess relative to the distant stars (see, e.g., Refs. 77 and 80). The angular velocity of this "frame-dragging" or "lens-Thirring" precession should be

$$\underset{\sim}{\Omega}_{\text{f.d., GR}} = \frac{G}{r^3}\left[-\underset{\sim}{J} + \frac{3(\underset{\sim}{J}\cdot\underset{\sim}{r})\underset{\sim}{r}}{r^3}\right] \tag{28}$$

where $\underset{\sim}{J}$ is the earth's angular momentum vector and $\underset{\sim}{r}$ is the vector from the earth's center to the gyroscope's position. The PPN framework predicts the alternative angular velocity.

$$\underset{\sim}{\Omega}_{\text{f.d., PPN}} = \left(\frac{7}{8}\Delta_1 + \frac{1}{8}\Delta_2\right)\underset{\sim}{\Omega}_{\text{f.d., GR}} \tag{29}$$

(For derivation see, e.g., Part 9 of Ref. 26; for further discussion see the papers of Everitt and O'Connell in this proceedings, and references cited therein.)

2. Polarization of the Lunar Orbit by the Sun

Nordtvedt (Ref. 4) has pointed out that in most relativistic theories of gravity, but not in general relativity, massive, self-gravitating bodies should violate the equivalence principle. This "Nordtvedt effect" was discovered independently by Dicke (Ref. 81), using elementary arguments about the response of gravitational energy to gravity. (See also Ref. 3.) The Nordtvedt effect produces a number of phenomena in the solar system which are potentially observable; Nordtvedt enumerates them in his paper in this proceedings. The most promising of these is a "polarization" of the earth-moon orbit due to the fact that the moon should fall toward the sun with a greater acceleration than does the earth. This polarization results in an eccentricity in the orbit which points always toward the sun and has amplitude

$$\delta r = 840\Big[3\gamma + 4\beta - 7\Delta_1 - \tfrac{1}{3}(2\beta + 2\beta_2 - 3\gamma + \Delta_2 - 2)\Big]\ \text{cm} \tag{30}$$

(See e.g., Ref. 82 for derivation of this particular formula.) Bender, in this proceedings, expresses considerable optimism about measuring or placing limits on this effect by means of lunar laser-ranging data.

3. Other Manifestations of the Nordtvedt Effect

Other manifestations of the Nordtvedt effect, listed and discussed in Nordtvedt's paper in these proceedings, look less promising experimentally but might be measurable in the 1970s. They include polarizations of planetary and spacecraft orbits by Jupiter (same parameter dependence as lunar orbit polarization, Eq. 30); also an anomaly in Kepler's third law for the sun-Jupiter system (same parameter dependence); also a displacement of the stable triple point of Lagrange for two orbiting masses (same parameter dependence); also time-dependent perturbations in the "gravitational acceleration" g measured by a gravimeter attached to the earth's surface (different parameter dependence; see end of 4, below).

4. Gravimeter Measurements ("Earth Tides")

The PPN formalism predicts a variety of time-dependent, relativistic perturbations of gravimeter measurements on the earth's surface — perturbations which are all absent in general relativity. The largest of these is caused by an anisotropy in the (Cavendish-measured) gravitational "constant," which in turn is due to motion of the earth through the "ether" (Ref. 69). A number of different "ethers" with different properties are predicted by various gravitation theories. An ether that perturbs gravimeter measurements is present in any theory with parameters $\Delta_2 + \zeta - 1 \neq 0$. (For an example of such a theory see Ref. 69.) This type of ether produces not only

gravimeter perturbations but also a different speed of propagation for gravitational waves than for light (cf. IIIC). The gravimeter perturbations, as calculated by Will (Ref. 69), are

$$\frac{\Delta g}{g} = 1 - \frac{1}{2}(\Delta_2 + \zeta - 1)(\underset{\sim}{v} \cdot \underset{\sim}{e}_r)^2 \qquad (31)$$

where $\underset{\sim}{v}$ is the earth's velocity relative to the ether, and $\underset{\sim}{e}_r$ is a unit vector pointing from the earth's center toward the gravimeter's location. As the earth rotates, $\underset{\sim}{e}_r$ rotates, and $\Delta g/g$ fluctuates with a sidereal 12-h period. Taking as a reasonable guess

$$\underset{\sim}{v} \equiv (\text{velocity of earth through ether}) - (\text{velocity of earth through galaxy})$$

one finds

$$(\text{amplitude of } \frac{\Delta g}{g}) \sim (\Delta_2 + \zeta - 1)(3 \times 10^{-8}) \qquad (32)$$

Actual gravimeter measurements show fluctuations with amplitude $\sim 5 \times 10^{-8}$ and with dominant periods of 12 solar hours (due to solar tidal effects), 12 lunar hours (due to lunar tidal effects), and 12 sidereal hours (due to the declination of the lunar and solar orbits). All these effects are well accounted for by standard Newtonian theory, and in particular, there is no sign of any anomalous 12-sidereal-hour effect down to amplitude $\sim 10^{-9}$ (see Ref. 69). Hence, gravimeter experiments place a very tight limit on $\Delta_2 + \zeta - 1$:

$$|\Delta_2 + \zeta - 1| \lesssim 3 \times 10^{-2} \qquad (33)$$

Other gravimeter perturbations, too small to show up in current gravimeter data, are spelled out at the end of Nordtvedt's paper in this proceedings (see also Ref. 83). The translation from Nordtvedt's notation (N) to Will's (W) for the parameters involved is

$$(2\beta + \alpha' - \gamma - 2)_N = (2\beta + 2\gamma - 2\beta_2 - 2)_W \qquad (34a)$$

$$(4\Delta' + 4\Delta' - 2 - 2\gamma)_N = \frac{1}{2}(7\Delta_1 + \Delta_2 - 4\gamma - 4)_W \qquad (34b)$$

$$\left(\alpha' - 4\Delta' - \frac{\gamma + \chi}{2}\right)_N = \frac{1}{2}(5\gamma - 4\beta_2 - \Delta_2)_W \qquad (34c)$$

$$\eta_N \equiv (8\Delta) - 4\beta - 3\gamma - \chi)_N = \frac{1}{2}(7\Delta_1 - 4\beta - 3\gamma)_W \qquad (34d)$$

The post-Newtonian limit of Whitehead's theory of gravity is too complex to fit into Will's nine-parameter version of the PPN formalism. However, Will (Ref. 69) has calculated its gravimeter predictions directly. He finds that the galaxy should produce an anisotropy in the (Cavendish-measured) gravitational "constant," and thence in gravimeter readings, given by

$$\Delta g/g \approx \left(\begin{matrix}\text{Newtonian potential due} \\ \text{to galaxy at earth}\end{matrix}\right) \times \cos^2\left(\begin{matrix}\text{angle made at gravimeter by} \\ \text{earth's center and galaxy's} \\ \text{center}\end{matrix}\right)$$

The amplitude of these 12-sidereal-hour perturbations is $\sim 2 \times 10^{-7}$ — a factor of 200 greater than the experimental limit on them. Hence, Whitehead's theory of gravity cannot possibly be correct.

5. Cavendish Experiments

According to general relativity, Cavendish experiments to measure the gravitational "constant" G_{local} must yield a single, universal number, independent of orientation of the test masses or location in space time. No so in most other theories of gravity. They predict (1) a change of G_{local} with time due to expansion of the universe (not analyzable in the PPN framework); (2) variations in space due to the proximity of matter [given in the PPN framework by

$$\Delta G_{local} = -G_\infty(2\beta + 2\gamma - 2\beta_2 - 2)U \qquad (35)$$

where G_∞ is the value of G_{local} in isolation (at "infinity"), and $U > 0$ is the Newtonian potential due to nearly matter; see Refs. 58, 69 and 83]; and (3) variations with velocity and orientation of the Cavendish apparatus, due to motion through the "ether" (given in the PPN framework by

$$\Delta G_{local} = G_\infty\left[\frac{1}{2}(4\beta_1 + 2\gamma + 1 - 7\Delta_1)\underset{\sim}{v}^2 - \frac{1}{2}(\Delta_2 + \zeta - 1)(\underset{\sim}{v} \cdot \underset{\sim}{e}_r)^2\right] \qquad (36)$$

where $\underset{\sim}{v}$ is the velocity of the apparatus relative to the ether, and $\underset{\sim}{e}_r$ is the unit vector between the two masses of the apparatus; see Ref. 69).

These effects all show up in gravimeter readings (4 above) as well as in Cavendish experiments. Cavendish experiments performed thus far — by contrast with gravimeter measurements — have been too insensitive to detect any of these effects or place significant limits on their parameters. A Cavendish experiment of sensitivity $\sim 10^{-11}$ would be required to detect with 10% precision the typical predicted variations in G_{local} due to the sun's changing Newtonian potential along the earth's eccentric orbit. In an

interplanetary spacecraft with fairly eccentric orbit a sensitivity of $\sim 10^{-8}$ would be adequate.

6. Active Versus Passive Gravitational Mass

The Cavendish experiments of the last section can be discussed from several different viewpoints. The viewpoint used above attributes variations in the attraction between Cavendish masses to variations in the local gravitational "constant." Another viewpoint, used below, declares that the gravitational "constant" is a true, universal constant, and that variations in the attraction must be due to variations in the "active gravitational masses" of the attracting bodies.

This viewpoint requires one to distinguish carefully between the four different masses that can be attributed to a body.* The inertial mass m_I is the mass which resists acceleration:

$$\underset{\sim}{F} = m_I \underset{\sim}{a}, \text{ i.e., (force)} = \text{(inertial mass)} \times \text{(acceleration)} \quad \text{for a body momentarily at rest} \quad \text{(no special relativistic effects)} \tag{37}$$

The conserved mass m_C is the "rest mass" which enters into the energy-momentum conservation laws of special relativity and classical mechanics. When one measures an atom's "mass" by mass spectroscopy (bending a beam in a magnetic field), the resulting number is m_I; when one measures its mass by applying energy-momentum conservation laws to nuclear reactions, the resulting number is m_C. These two types of measurements yield the same results to a precision of

$$\left| \frac{m_I - m_C}{m_C} \right| < 10^{-6} \tag{38}$$

for typical atoms. The passive gravitational mass of a body m_P tells how much force must be applied to prevent the body from falling in a static, external Newtonian gravitational field

$$\underset{\sim}{F} = -m_P \underset{\sim}{\nabla} U_{ext} \tag{39}$$

The Dicke-Eötvös experiment reveals that laboratory bodies of widely differing chemical composition have the same ratio of m_P/m_I to a precision of $\lesssim 3 \times 10^{-11}$. The active gravitational mass m_A tells how much "1/r" Newtonian gravitational field is produced by a body

$$U = \frac{-G_\infty m_A}{r} + \text{terms of higher order in } \frac{1}{r} \tag{40}$$

Here G_∞ is a "universal" gravitational constant which one might choose, for example, by demanding that $m_A = m_C$ for a small sphere of some standard substance (e.g., carbon) far from all other matter ("at infinity").

In the language of "inertial," "conserved," "active," and "passive" mass, the PPN prediction (Eqs. 35, 36) for the dependence of a Cavendish measurement on location, velocity, and direction reads

$$\frac{\Delta m_A}{m_C} = \frac{\Delta G_{local}}{G_\infty} = -(2\beta + 2\gamma - 2\beta_2 - 2)U_{ext} + \tfrac{1}{2}(4\beta_1 + 2\gamma + 1 - 7\Delta_1)\underset{\sim}{v}^2 - \tfrac{1}{2}(\Delta_2 + \zeta - 1)(\underset{\sim}{v} \cdot \underset{\sim}{e}_r)^2 \tag{41}$$

i.e., it says that the active mass of a body depends on the external Newtonian potential it resides in, on its velocity through the ether, and on the angle $\theta = \cos^{-1}(\underset{\sim}{v} \cdot \underset{\sim}{e}_r / v)$ between the field point and the velocity direction. (Note that dependence on θ means the active mass is anisotropic!) These variations in active mass, as emphasized in 4 and 5 above, have shown up most sensitively thus far in their influence on gravimeter measurements (measurements of earth's active mass).

The PPN formalism predicts yet another type of variation in m_A/m_C: a variation with chemical composition and with self-gravitational energy

$$\frac{m_A}{m_C} - 1 = 2(2\beta_1 - \beta_4 - 1)\frac{E_{kin}}{m_C} + (\beta_3 - 1)\frac{E_{int}}{m_C} - \zeta \frac{E_{ket}}{m_C} + (6\gamma - 4\beta_2 - \beta_4 - 1)\frac{E_{grav}}{m_C} \tag{42}$$

(For derivation see Part 9 of Ref. 26, and Refs. 68 and 70.) Here E_{kin} is the kinetic energy of all the nucleons in all the atoms of the gravitating body; E_{int} is the internal energy (electrostatic, nuclear binding, etc.) of all the atoms; E_{ket} is a projection of the body's "kinetic-energy tensor" on the direction toward the observer:

$$E_{ket} = \tfrac{1}{2} \int \rho_0 (\underset{\sim}{v} \cdot \underset{\sim}{e}_r)^2 \, d^3x \tag{43}$$

and E_{grav} 0 is the self-gravitational energy of the body.

*Bondi (Ref. 84) distinguishes three types of mass: "inertial," "active," and "passive." To these we have added "conserved" mass.

There is no hope of measuring the dependence on self-gravitational energy. There is also no hope of measuring the anisotropy in m_A produced by E_{ket}. The contributions to such an anisotropy from each nucleus (due to deviations from spherical symmetry) would average to zero over all the (randomly oriented) nuclei in a laboratory-sized body, and E_{ket} would take the form

$$\left\langle E_{ket} \right\rangle_{\text{av. over nuclei}} = \frac{1}{3} E_{kin}$$

Thus the variation in m_A/m_C takes the form, for laboratory-sized bodies,

$$\frac{m_A}{m_C} - 1 = 2(2\beta_1 - \beta_4 - 1 - \frac{1}{6}\zeta) \frac{E_{kin}}{m_C} + (\beta_3 - 1) \frac{E_{int}}{m_C}$$

However, the "active-versus-passive-mass" experiment of Kreuzer (Ref. 85) would have revealed the dependences on E_{kin} and E_{int} if the relevant parameter combinations $(2\beta_1 - \beta_4 - 1 - 1/6\zeta;\ \beta_3 - 1)$ were near unity. Kreuzer finds that bromine and fluorine have the same ratio of m_A/m_P to a precision $\lesssim 5 \times 10^{-5}$. Since m_P/m_I and m_P/m_C are independent of chemical composition to higher precision than this (see above), one concludes that

$$\left| \frac{(m_A/m_C)_{\text{Fluorine}} - (m_A/m_C)_{\text{Bromine}}}{(m_A/m_C)_{\text{Bromine}}} \right| \lesssim 5 \times 10^{-5} \qquad (44)$$

The kinetic energies of bromine and fluorine, as evaluated using the "Uniform Model of Wigner" (Ref. 86, pp. 266-270), are

$$\left(\frac{E_{kin}}{M_C}\right)_F = 1.004 \frac{T_o}{M_{AMU}}, \qquad \left(\frac{E_{kin}}{M_C}\right)_{Br} = 1.009 \frac{T_o}{M_{AMU}} \qquad (45a)$$

where T_o is a constant energy defined by Blatt and Weisskopf (Ref. 86), M_{AMU} is the energy of one atomic mass unit, and

$$\frac{T_o}{M_{AMU}} \approx 0.0140 \qquad (45b)$$

The internal energies, $E_{int} = [-(\text{measured binding energy}) - E_{kin}]$, are

$$\left(\frac{E_{int}}{M_C}\right)_F = -0.0224, \qquad \left(\frac{E_{int}}{M_C}\right)_{Br} = -0.0234 \qquad (45c)$$

Consequently, unless there is a fortuitous cancellation of kinetic and internal contributions, Kreuzer's experiment yields the limits

$$\left|2\beta_1 - \beta_4 - 1 - \frac{1}{6}\zeta\right| \lesssim 0.04, \quad \left|\beta_3 - 1\right| \lesssim 0.05 \qquad (46)$$

Of course, the PPN formalism is rather cavalier in its lumping of all types of internal energy into a single entity, Π. A more cautious treatment would separate the electromagnetic energy, the nuclear binding energy, etc., and would write

$$\Pi = \Pi_{EM} + \Pi_{NB} + \cdots$$

It would also separate out the stresses t_{jk} due to the electromagnetic field, the nuclear pion fields, etc.; and it would include a provision for gravitation produced by anisotropic stresses. Correspondingly it would introduce separate PPN parameters $\beta_{3,EM}$, $\beta_{3,NB}$, $\beta_{4,EM}$, $\beta_{4,PION}$, ... for all possible sources of gravity; and it would use the Kreuzer experiment to put limits on as many of these parameters as possible. Such an analysis is underway in the authors' group.

V. Conclusion

Experiments to date place the following limits on the PPN parameters:

$\gamma = 1.04 \pm 0.08$	(time delay and light deflection)
$\beta = 1.14 \,^{+0.2}_{-0.3}$	(perihelion shift plus time delay)
$\left\|\Delta_2 + \zeta - 1\right\| \lesssim 0.03$	(gravimeter measurements)
$\left\|2\beta_1 - \beta_4 - 1 - \frac{1}{6}\zeta\right\| \lesssim 0.04$	(Kreuzer measurement of m_A/m_P)
$\left\|\beta_3 - 1\right\| \lesssim 0.05$	(Kreuzer measurement of m_A/m_P)

(47)

Comparison of these experimental results with the parameter values of various theories (Table 6) enables one to "pass judgment" on the theories (see Table 7). All theories are thereby disproved except general relativity, the scalar-tensor theories (Brans-Dicke-Jordan; Bergmann-Wagoner), Ni's theories, and possibly the modified Yilmaz theory. However, other theories not yet studied by the authors might also survive the comparison with experiment.

As indicated in Table 7, higher-precision measurements of time delay, light deflection, and perihelion shift are needed to test the scalar-tensor theories, while successful completion of the Stanford gyroscope experiment and of the lunar laser ranging experiment (Nordtvedt effect) are needed to test the theories of Yilmaz and Ni.

Table 7. Comparison of theory and experiment for the metric theories of Table 2

Theory and its parameters	Status
*1. General Relativity	Agrees with all experiments to date.
*2. Scalar-tensor theories (ω, Λ)	Agree with time-delay and deflection experiments: to 1σ accuracy if $\omega > 23$, to 2σ accuracy if $\omega > 6$. Agree with perihelion shift measurements: to 1σ accuracy if $-0.16 < \Lambda < 0.34$, to 2σ accuracy if $-0.46 < \Lambda < 0.64$. Agree completely with gravimeter and Kreuzer experiments. Future experiments should concentrate on pushing ω toward $+\infty$ and Λ toward 0 (general relativity limit). This is best done by experiments of highest precision — time delay, light deflection, and perihelion shift.
3. Conformally-flat theories of Nordstrom type (p, q)	Disagree violently with time-delay and light-deflection experiments.
4. Stratified theories with conformally flat space slices	
a. Einstein	Disagrees violently with time delay, light deflection, and perihelion shifts.
b. Whitrow-Murduch	Diagrees violently with time delay and light deflection.
c. Modified Yilmaz[a]	Agrees completely with time delay, light deflection, and perihelion shifts. Disagrees mildly but not conclusively with Kreuzer experiment (dependence of active mass on chemical composition). More careful theoretical analysis of Kreuzer experiment might rule it out conclusively. Predicts zero dragging of inertial frames by earth's rotation; would thus be tested by Stanford gyroscope experiment. Predicts a polarization amplitude of 71 m for the lunar orbit (Nordtvedt effect); would thus be tested by lunar laser ranging.
d. Ni's Lagragian-based theory[a]	Agrees with all experiments to date. Predicts zero dragging of inertial frames by earth's rotation; would thus be tested by Stanford gyroscope experiment. Predicts a polarization amplitude of 67 m for the lunar orbit (Nordtvedt effect); would thus be tested by lunar laser ranging.
e. Ni's general stratified theory (p. q)[a]	Agrees with all experiments to date. Predicts zero dragging of inertial frames by earth's rotation; would thus be tested by Stanford gyroscope experiment. Predicts a polarization amplitude of 280 (26 - 10q - p) cm for the lunar orbit; would thus be tested by lunar laser ranging.
5. Whitehead's theory	Agrees with time delay, light deflection, and perihelion shifts. Disagrees violently with gravimeter measurements (see IVC4 of text).
Generalized Whitehead	Has not yet been studied adequately to compare with experiment.

*These theories are currently viable.

Davies, in his contribution to this proceedings, argues that experimenters should not take the PPN framework too seriously because theorists are generally less clever than nature. We agree. The PPN framework might indeed be too narrow to encompass the correct theory of gravity; so each experimenter should state the results of his experiment not in the PPN language, but instead in highly empirical language. Nevertheless, experimenters will forgive theorists like us, we hope, if we use the PPN framework to get a rough measure of the meaning of their experiments. And they will applaud the PPN framework, we hope, whenever it proposes new experiments to distinguish various gravitation theories from each other. Indeed, new experiments it has been proposing from time to time during the last few years: experiments to measure the "polarization" of lunar and planetary orbits (IVC2 above); other experiments to detect other consequences of the Nordtvedt effect (IVC3); and experiments to measure anisotropies in the gravitational "constant" (IVC4 and IVC5).

C

Tests of the Equivalence Principle and Gravitation Theory Using Solar System Bodies

Kenneth Nordtvedt, Jr.
Montana State University

I. Introduction

The Equivalence Principle of Einstein is the assertion that uniform gravitational fields (realistic gravitational fields neglecting their gradients) and accelerated coordinate systems are locally indistinguishable; they yield equivalent experimental results (Ref. 1). By invoking this principle, one expects all bodies to be accelerated at the same rate in a given external gravitational field.

In Newtonian language the ratio of the acceleration of a body to the gravitational field in which it is located is given by the passive gravitational to inertial mass ratio;

$$\vec{a} = \left(\frac{M_g}{M_i}\right)\vec{g}_{ex} \qquad (1)$$

I will describe possible experiments which can measure with precision the M_g/M_i ratio of celestial bodies. Also, I will briefly indicate in what manner such measurements are deep probes of the post-Newtonian structure of gravitation theories.

II. Discussion

Bodies for which the M_g/M_i ratio are measured can be divided into two classes: (1) those bodies containing negligible internal gravitational potential energy; and (2) those bodies containing a substantial amount of the same. A typical laboratory object is of the former class, its gravitational potential energy being of order 10^{-24} its mass energy.

A wide class of theories can be constructed so that all laboratory objects, independent of their internal nature, fall at the same rate in a gravitational field. Given a body which: (1) is in internal equilibrium, and (2) is not experiencing external forces, then the volume integral of the body's energy density equals the volume integral of the body's contracted scalar energy density;

$$\int T_{oo}\, d^3x = \int T\, d^3x, \qquad (2)$$

$$T = T_{oo} - T_{xx} - T_{yy} - T_{zz}$$

Any gravitational theory which in the Newtonian limit couples the gravitational potentials to a linear combination of T_{oo} and T will fulfill the Equivalence Principle for laboratory objects. In particular the scalar-tensor theories fulfill this (Refs. 2 and 3). Laboratory bodies supported on the earth's surface by solid-state contact forces have their T_{oo} integral differ from the T integral by order 10^{-18} the body's mass energy, so non-null experimental results might be found at that level. The Eötvös (Ref. 4) and Princeton (Ref. 5) experiments have verified the constancy of M_g/M_i for such bodies to order 10^{-11}.

Celestial bodies on the other hand have internal gravitational potential energies of order 10^{-8} to 10^{-5} their mass energy. We inquire therefore whether the M_g/M_i ratio of these bodies is altered at this level due to the presence of the internal gravitational energy.

Phenomenologically the acceleration of a rotating massive fluid body in an external gravitational field $\vec{g}_{ex}$ can be expressed as follows:

$$\vec{a} = \vec{g}_{ex}\left[1 + \eta\frac{U}{Mc^2}\right] + \eta'\left[\frac{J^2}{3}\vec{g}_{ex} - \vec{J}\cdot\vec{g}_{ex}\vec{J}\right]\Big/2\mathcal{I}Mc^2 \qquad (3)$$

where

$$U = -\frac{1}{2}\int\frac{\rho(x)\,\rho(x')d^3xd^3x'}{|\vec{x} - \vec{x}'|}$$

is the gravitational potential energy in the body, $\vec{J}$ is the body's angular momentum, and $\mathcal{I}$ its moment of inertia. Experiments are sought which will measure non-zero values of the dimensionless parameters η and η'.

For celestial bodies which are not on the verge of centrifugal breakup, rotational energy is much smaller than the body's gravitational potential energy, so I will concentrate on detection of possible η effects.

First consider Kepler's Third Law for the sun-Jupiter system:

$$\omega^2 R^3 = G\left[M_s\left(\frac{M_g}{M_i}\right)_J + M_J\left(\frac{M_g}{M_i}\right)_s\right] \qquad (4)$$

If it is assumed that

$$\left(\frac{M_g}{M_i}\right)_s \cong 1 + 10^{-5}, \qquad \left(\frac{M_g}{M_i}\right)_J \cong 1 + 10^{-8}$$

and

$$\frac{M_J}{M_s} \cong 10^{-3}$$

then a part in 10^8 correction to the law will result.

To detect this will require that Jupiter's orbit frequency and orbit radius be measured to better than a part in 10^8 accuracy, while the mass of Jupiter must be known to better than a part in 10^5.

Consider now the gravitational three body problem; a third body M_x perturbs a two-body system M_1 and M_2. The relative acceleration of the two bodies produced by the third body is given by (Ref. 6)

$$\delta\vec{a}_{12} = GM_x\left[\frac{\vec{R} - \vec{r}}{|R - r|^3}\left(\frac{M_g}{M_i}\right)_2 - \frac{\vec{R}}{R^3}\left(\frac{M_g}{M_i}\right)_1\right]$$

$$= \frac{GM_x\vec{R}}{R^3}\left[\left(\frac{M_g}{M_i}\right)_2 - \left(\frac{M_g}{M_i}\right)_1\right]$$

$$+ \frac{GM_x}{R^3}\left[\vec{r} - 3\frac{\vec{r}\cdot\vec{R}\vec{R}}{R^2}\right]$$

$$+ \text{octupolar term}$$

$$+ \cdots \qquad (5)$$

If bodies M_1 and M_2 have different M_g/M_i ratios, their orbit is polarized by M_x in a dipole manner, in addition to the dominant classical Newtonian perturbations which are quadropolar and octupolar polarizations.

Applying these considerations to the sun's effect on the earth-moon system leads to a perturbation on the earth-moon distance:

$$\delta r_{EM}(t) \cong \frac{1}{6}\,10^{13}\,\delta\cos(\omega - \Omega) + \text{cm} \qquad (6)$$

with

$$\delta = \left(\frac{M_g}{M_i}\right)_{\text{earth}} - \left(\frac{M_g}{M_i}\right)_{\text{moon}}$$

ω = the moon's angular frequency around the earth

Ω = the earth's angular frequency around the sun

If the lunar laser ranging experiment can ultimately detect a range oscillation of this frequency to accuracy 3 cm, then a δ of size 2×10^{-12} can be detected, bettering the fractional accuracy of the laboratory Equivalence Principle experiments. The magnitude of δ calculated in the Brans-Dicke scalar-tensor gravitational theory is about 1.5×10^{-10} while General Relativity predicts $\delta = 0$ (Ref. 7).

Are there other sources which produce perturbations of the same frequency on the

earth-moon distance, possibly making detection of this effect difficult or impossible?

(1) The classical octupolar perturbation contributes at this frequency. However the orbital parameters of the lunar orbit are known to sufficient accuracy to calculate this classical effect and look for residuals.

(2) The sun's illumination of the lunar surface and an accompanying thermal expansion wave travels around the moon at this frequency. The effect of crustal movement on the distance from the earth to the laser reflector can be minimized by using lunar polar locations or ±90° longitude locations. Also, if the reflector is at a lunar longitude other than directly toward the earth, then out of phase, $\sin(\omega - \Omega)t$, range signals will be generated which will be solely thermal in origin, allowing direct measurement of the thermal effect. Workers who have theoretically estimated the thermal expansion effect conclude that it will be negligible.

Laser ranging to an earth orbiting satellite could also be used to search for an Equivalence Principle violation; however the orbit's range perturbation diminishes with orbit radius as $r^{3/2}$, so large satellite orbits must be used.

If we consider a possible dipole polarization by Jupiter on the earth orbit around the sun, this will amount to about 1 km assuming

$$\delta_s \simeq 10^{-5}$$

There are some consequences of $M_g/M_i \neq 1$ in the solar system which exploit resonance conditions to amplify the size of the perturbations. Let the galaxy's mass M_g at distance R_g perturb the earth orbit. Assuming a solar δ_s of order 10^{-5}, the earth's orbit is polarized toward the galactic center by an amount (Ref. 6)

$$\delta r(t) = \frac{3}{2}\frac{GM_g}{R_g^2}\frac{\delta_s}{\omega_E \Omega_E}\cos\omega_E t \qquad (7)$$

where

ω_E = the earth's orbital frequency

and

Ω_E = the earth's total perihelion precession frequency

This orbital perturbation amounts to about 13 km, but I know of no efficient way to observationally disentangle this fixed polarization from the very slowly advancing natural eccentricity of the earth's orbit.

The stable triple point of LaGrange for two orbiting masses is also effected by Equivalence Principle violations. For the earth-sun system the stable point moves toward the earth by (Ref. 8)

$$2\delta R_E \simeq 10^3 \text{ km} \cong 3 \times 10^{-3} \text{ s} \qquad (8)$$

if

$$\delta_s = 10^{-5}$$

This suggests placing a satellite at the earth-sun triple point and ranging on that satellite as it wanders about the equilibrium point.

The Trojan asteroids which wander near the Jupiter-sun triple points will have their equilibrium positions shifted also. As viewed from the earth this movement would amount to about 1" of arc (Ref. 6).

What can experimental results on M_g/M_i for massive bodies tell us about gravitation theories? Below is an outline of the calculation of M_g/M_i for massive bodies in geometrical gravitational theories.

Starting from the action integral for an assembly of particles which possess both mass and electromagnetic charge (Refs. 3, 7, and 9),

$$A = -\sum_i \int dt \left[m_i \sqrt{g_{\mu\nu}(x,t)\frac{dx^\mu}{dt}\frac{dx^\nu}{dt}} + e_i A_\mu(x,t)\frac{dx^\mu}{dt}\right] - \frac{1}{16\pi}\int \sqrt{-g}\, d^4x F_{\mu\nu}F^{\mu\nu} \qquad (9)$$

the equation of motion of each particle can be obtained. The space-time metric $g_{\mu\nu}(x,t)$ which appears in Eq. (9) is assumed to be given by a parameterized expansion in a set of gravitational potentials produced by all the matter in the world, including that matter external and internal to the assembly of particles under consideration.

From the equation of motion of each particle, the equation of motion of the entire assembly of particles is obtained by proper averaging; this collection of particles is assumed to be in the Newtonian gravitational field of an external mass. We then identify the M_g/M_i ratio for the massive body. Will and Thorne (Ref. 10) have obtained the same results using a hydrodynamical model for a massive body plus the curved space conservation equations

$$T^{ij}{}_{ij} = 0$$

For a collection of masses m_i moving at velocities $\vec{v}_i$ and accelerations $\vec{a}_i$, the

parameterized space-time metric including the potentials needed to solve this problem is (Ref. 7)

$$g_{oo} = 1 - 2\psi + 2\beta\psi^2$$

$$+ 2\alpha' \sum_{ij} \frac{m_i m_j}{r_{ij}} \left(\frac{1}{|\vec{r} - \vec{r}_i|} + \frac{1}{|\vec{r} - \vec{r}_j|} \right)$$

$$+ \chi \sum_i \frac{m_i}{|\vec{r} - \vec{r}_i|} \overrightarrow{r - r_i} \cdot \vec{a}_i + \cdots$$

$$g_{ok} = 4D \sum_i \frac{m_i}{|\vec{r} - \vec{r}_i|} (v_i)^k$$

$$+ 4D' \sum_i \frac{m_i \vec{v}_i \cdot \overrightarrow{r - r_i}}{|\vec{r} - \vec{r}_i|^3} (r - r_i)^k + \cdots \quad (10)$$

$$g_{kk'} = - (1 + 2\gamma\psi)\, \delta_{kk'} + \cdots$$

with

$$\psi = \sum \frac{m_i}{|\vec{r} - \vec{r}_i|}$$

In terms of the metric parameters the gravitational to inertial mass ratio tensor of a massive rotating fluid body is then found to be given by the expression (Ref. 11)

$$\left(\frac{M_g}{M_i} \right)_{\alpha\beta} = \left[1 - \eta \frac{U}{Mc^2} \right] \delta_{\alpha\beta}$$

$$+ \eta' \left[\frac{J^2}{2\mathcal{I}}\, \delta_{\alpha\beta} - \frac{3 J_\alpha J_\beta}{2\mathcal{I}} \right] \Big/ Mc^2 \quad (11)$$

with

$$U = -\frac{1}{2} \sum \frac{G m_i m_j}{r_{ij}}$$

$$\eta = (8\Delta - 4\beta - 3\gamma - \chi) + \eta'/3$$

$$= 0 \text{ in General Relativity} \quad (11a)$$

$$\eta' = (8\Delta' + 2\beta + \chi - \alpha' - 2)$$

$$= 0 \text{ in General Relativity} \quad (11b)$$

If the metrical gravitation theory is in accord with Special Relativity, then all the parameters in (11a and b) are not independent of each other. Applying a Lorentz transformation and retardation corrections to the static linearized metric,

$$g_{oo} = 1 - 2\psi$$

$$g_{ok} = 0$$

$$g_{kk'} = - [1 + 2\gamma\psi] \delta_{kk'}$$

gives us some parameter values determined by Special Relativity alone:

$$\chi = 1$$

$$\Delta' = 0$$

$$\Delta = (1 + \gamma)/2$$

Then (11a and b) read

$$\eta = (3 + \gamma - 4\beta) + \eta'/3$$

$$\eta' = (2\beta - \alpha' - 1)$$

The expressions (11 a and b) illustrate what an extensive probe of the post-Newtonian structure of gravitation theory M_g/M_i measurements will be. γ and β are parameters of the Schwarzschild metric which are being measured by the electromagnetic signal deflection and time delay experiments and planetary orbit perihelion advance measurements. The parameters Δ and Δ' label the Lense-Thirring potentials which will be sought in the Stanford orbiting gyroscope experiment. χ, which appears in (11a), is a parameter that measures the Special Relativistic retardation correction to the Newtonian potential and is not measurable in any other present experiment (but is inferred in the detection of gravity waves). α' is a parameter labeling a second class of non-linear potentials which also is only presently measurable in M_g/M_i experiments.

In a gauge where $\Delta' = 0$ all the parameters in General Relativity have the value one:

$$\gamma = \beta = \Delta = \chi = \alpha' = 1$$

but this is not true in most gravitational theories.

If the lunar laser ranging experiment can put an upper limit of 3 cm on the oscillatory earth-moon range term given in Eq. (6), that will result in a part in 500 ($|\eta| \leq 2 \times 10^{-3}$) confirmation of much of General Relativity's post-Newtonian structure.

Dicke (Ref. 12) and also Thorne and Will (Ref. 10) have presented a heuristic argument which obtains the M_g/M_i value in the scalar-tensor theories as a consequence of the change of Newton's gravitational constant when mass is nearby. It can be shown that their argument only holds in those gravitational theories whose non-linear structure parameters β and α' obey the relation

$$2\beta = \alpha' + 1$$

It is conjectured (but not yet proven) that this relationship between β and α' is fulfilled in all LaGrangian-based gravitational theories.

The discussion so far can be said to have investigated the global aspects of the Equivalence Principle and massive bodies. How does an entire body move in an external gravitational field? I want to mention some local or differential aspects of the Equivalence Principle and massive bodies.

If a massive body is in a uniform external gravitational field, does every part of the body experience the same external acceleration? Specifically, when the earth and a laboratory on the earth fall toward the sun, is there a differential external acceleration (neglecting gravitational gradient effects) which is compensated by the laboratory's support forces?

Theoretical analysis shows that a differential acceleration which must be provided by the solid-state supports is given by (Ref. 9)

$$\delta a_r = (2\beta + \alpha' - \gamma - 2)\left(\frac{GM_s}{c^2 R_s}\right)\vec{g}_E$$

$$+ \left[\eta \frac{U}{Mc^2} - (4\Delta + 4\Delta' - 2 - 2\gamma)\frac{GM_E}{c^2 r_E}\right]\vec{g}_s \cdot \hat{r}_E$$

$$- \frac{2}{3}\left(\alpha' - 4\Delta' - \frac{\gamma + \chi}{2}\right) I_2 \vec{g}_s \cdot \frac{\hat{r}_E}{r_E^3}$$

$$\delta a_t = \left[\eta \frac{U}{Mc^2} - (8\Delta - 4\beta - 3\gamma - \chi)\frac{GM_E}{2c^2 r_E}\right]\vec{g}_s \cdot \hat{t}$$

$$+ \frac{1}{3}\left(\alpha' - 4\Delta' - \frac{\gamma + \chi}{2}\right) I_2 \hat{g}_s \cdot \frac{\hat{t}}{r_E^3} \qquad (12)$$

with

$$I_2 = \frac{G}{c^2}\int \rho_E(x) x^2 d^3x$$

$\vec{g}_E$ = the earth's central gravitational field

$\vec{g}_s$ = the sun's gravitational field

r_E = radius of the earth

$\hat{r}_E$ = unit radial vector

R_s = distance to the sun

δa_r and δa_t are the radial and tangential components of the Equivalence Principle violating differential acceleration. All the effects in Eq. (12) vanish in General Relativity but not in other theories. But again the post-Newtonian metric structure is needed to predict the General Relativity null result.

The term in Eq. (12) proportional to $(2\beta + \alpha' - \gamma - 2)$ is interpreted as a change in Newton's gravitational constant due to the proximity of the sun's mass. As the earth goes from orbital apogee to perigee, this term varies by order $10^{-10}\vec{g}_E$.

The other terms are seen as 24-hour period, a part in 10^{12}, variations in the magnitude and direction of the earth's gravitational field. Gravimeter and horizontal pendulum-type experiments can search for such effects. However, the solar 24-hour tidal force drastically overwhelms these relativistic effects.

III. Conclusions

If one or more of the experiments I have discussed can be performed and the results are in accord with the Equivalence Principle -- massive celestial bodies and their various parts fall at the same rate as test particles in gravitational fields -- we will still have learned a great deal new about gravitation. The presence of several post-Newtonian gravitational potentials which are not yet observable in other experiments but are predicted by Einstein's General Relativity will be seen. A detection of $M_g/M_i \neq 1$ for a celestial body would signal a breakdown of General Relativity.

References

1. Einstein, A., The Meaning of Relativity, 5th Edition, Chapter III, Princeton University Press, 1955.
2. Brans, C., and Dicke, R. H., Phys. Rev., 124, 925, 1961.
3. Nordtvedt, K., Jr., Int. J. Theor. Physics, 3, 133, 1970.
4. Eötvös, R. V., Ann. Physik, 68, 11, 1922.
5. Roll, P. G., Krotkow, R., and Dicke, R. H., Ann. Phys. (N. Y.), 25, 442, 1964.
6. Nordtvedt, K., Jr., Icarus, 12, 91, 1970.
7. Nordtvedt, K., Jr., Phys. Rev., 169, 1017, 1968.
8. Nordtvedt, K., Jr., Phys. Rev., 169, 1014, 1968.

9. Nordtvedt, K., Jr., Phys. Rev. (in press).

10. Thorne, K., and Will, C., Astrophys. J. (in press).

11. Nordtvedt, K., Jr., Phys. Rev., 180, 1293, 1969.

12. Dicke, R. H., Gravitation and the Universe, Jayne Lectures for 1969, American Philosophical Society, Philadelphia (in press).

A Low-Temperature Gravitational Radiation Detector

William O. Hamilton
Louisiana State University

I. Introduction

There are few times in the history of science when one has the opportunity to participate in the discovery and investigation of a completely new phenomenon. It is even more rare that the phenomenon be of such importance that the results of the experiments are sure to profoundly modify our concept of the origin and history of the universe. Weber's report of the discovery of gravitational radiation signals the discovery of such a phenomenon. We will discuss the beginning design of an experiment to investigate gravitational radiation by the use of massive detectors which are cooled to ultralow temperatures in order to improve the signal-to-noise ratio and the effective range and stability of the detectors. It may also be possible to measure the speed of propagation of the radiation.

II. The Gravitational Detector

Weber has shown that a gravitational wave propagating through a material will induce a time-varying strain which is proportional to the magnitude of the Riemann curvature tensor in the material. If the body is chosen to be a mechanical oscillator, and if the gravitational wave has appreciable harmonic content at the resonant frequency of the oscillator, then the time-varying strain will force the body into oscillation. The amplitude of oscillation will depend on the length of the body, the wavelength of the gravitational radiation, the Q of the body, and the amplitude of the time-varying curvature. Weber has reported the detection of gravitational radiation by monitoring the strain in a 5-ft-long aluminum bar. He is sensitive only to the frequency of the strain which appears in the lowest longitudinal mode of oscillation of the bar. Most of the signal is due to the equilibrium thermal motions of the bar. Occasionally a large pulse is detected, indicating an increase in the strain associated with the lowest mode. Weber has verified that these pulses are not due to thermal fluctuations by examining coincidences between two separated bars. He claims that the probability of seeing the number of coincidences he has detected is extremely small. His detection probabilities are statistical, however, since his signal-to-noise ratio in a single bar is 5:1 for the largest signals. Thus, if a gravitational disturbance tends to oscillate the bar out of phase with the thermal motion of the bar at the time of arrival of the disturbance, the signal will appear much smaller than it would have if the disturbance had arrived in phase with the thermal motion; as a result, the disturbance might be overlooked.

III. A Low-Temperature Detection System

We are preparing an experiment to check Weber's discovery and to greatly increase the sensitivity of the antenna, and hence the detection probability for gravity waves. We intend to cool a bar to below helium temperatures in order to reduce the thermal contributions to the motion of the bar. The experiment is to be performed cooperatively with Stanford. One bar will be located at Palo Alto and the other at Baton Rouge. Each bar will be absolutely independent from the other, the data being stored at each site on magnetic tape, along with an accurate time signal. This allows the data to be analyzed easily by computer, using the latest techniques of optimum filtering and stochastic signal detection. The long baseline between the two detectors offers the possibility of determining the speed of propagation of the radiation. A new superconducting accelerometer will be used to detect the motion of the bar. This new

instrument promises to provide the sensitivity and low noise necessary to detect the 10^{-16}-cm motions of the end of the bar when it is cooled to less than 0.1 K.

IV. The Cooled Detector and Magnetic Support

The most obvious advantage of the low-temperature environment is the reduced thermal noise in the antenna bar, the average kinetic energy of an oscillator in equilibrium being 1/2kT. The reduced noise will allow much weaker signals to be detected and will eliminate statistical uncertainties in those signals which are large enough to be detected on room-temperature apparatus.

The bar will be supported within the helium dewar by a magnetic field. The bottom of the aluminum bar will be coated with a superconductor, either niobium or niobium-titanium foil. Niobium-titanium wires will be wound on a cradle in which the bar rests. After the apparatus is cooled to liquid helium temperatures, a current will be passed through the wires of the support cradle and the bar will be raised by the magnetic pressure exerted by the field of the wires at the surface of the superconducting foil. The current can be made persistent to insure a constant supporting flux, with no heat dissipation in the helium or the bar. A ballast inductance can make the current practically constant, thus assuring a long support time constant and maximum vibration isolation.

There are two principal advantages to this magnetic support. The first and most obvious is that the supporting force is almost uniform and can be made practically independent of the position of the supported bar, thus resulting in a long-time constant support and good vibration isolation. Vibration isolation is further improved by the uniformity of the support, which makes it difficult to couple any vibrational energy into the longitudinal modes of oscillation. The second advantage to the magnetic support and its uniform support pressure is the absence of mode selection in the bar. Most methods of support will work well only for modes which have nodes at the support points. The magnetic support is uniform and will not damp any mode more than another. As a result, we can look for gravitational stimulation of any of the normal modes of the bar.

This technique then allows us to infer details of the frequency spectrum of the gravitational event. It also allows us to discriminate against some types of vibrational excitation. Since a gravitational wave affects an oscillator like a time-varying force gradient, a gravitational wave should not excite any of the even harmonics of the bar. Thus, an event which is coincidental in two bars tuned at, say, 1600 Hz but which does not excite the second harmonic (also at 1600 Hz) of a third bar is most probably a gravitational event. The first and third harmonics of the third bar should be excited if there is appreciable spectral density at these frequencies. This then gives us a completely new tool with which to analyze the nature of the incoming radiation.

The magnetic suspension also makes possible another exciting development. If thermal noise is the factor which limits the sensitivity of the gravitational detector, then greatly increased sensitivity can be obtained by cooling the bar to as low a temperature as possible. In this way, a much smaller gravitational signal can be detected above the noise. The magnetic suspension eliminates the otherwise unavoidable conduction of heat to the bar by mechanical supports and thus makes it not unreasonable to consider cooling the bar to below 1 K. At 1 K, the enthalpy of an aluminum bar weighing 1300 kg (approximately the size used by Weber) is only 40 J. This amount of heat can be removed rather easily by He^3 or He^3-He^4 pumping or demagnetization of a large salt. The vibration isolation and lack of thermal conductivity afforded by the magnetic support make the use of ultralow temperatures feasible and make it possible to consider keeping the gravitational detector at these low temperatures for extended times. There will be no eddy current heating since the normal metal in the bar is shielded from the support field by the superconducting foil. Other sources of heating will be conduction from the surrounding foil and from the surrounding walls via thermal radiation, and residual gas conduction and heat generation by flux motion in the superconducting foil. The first two sources can be made arbitrarily small by proper design, and the third can be minimized by choosing the proper support foil. The foil can be either a type I material operated below Hc_1 or a type II material operated in the region where the trapped flux is tightly pinned. This second choice is probably preferable since a larger support pressure can be sustained by type II material, and hence a heavier bar can be supported.

V. Superconducting Shielding

The use of a low-temperature environment also allows the complete elimination of electromagnetic noise sources by the use of superconducting shielding. Experiments performed by Cabrera and Hamilton indicate that a superconducting lead shield will offer a shielding factor of better than 10^9 against electric and magnetic fields of any frequency from dc up to hundreds of megacycles. Thus, the entire antenna and, if necessary, some of the low-level electronics can be placed inside a closed superconducting shield. We have previously built electronic circuits using field-effect transistor circuitry which worked well at the temperature of liquid helium. The use of superconducting shielding eliminates not only the obvious electromagnetic pickup at the signal frequencies of the gravitational detector but also the possibility of detection due to the magneto-acoustic effect, which can parametrically pump energy into a bar through the interaction between eddy currents in the bar and a slowly varying external field. We have shown that the magnetic field inside a properly constructed superconducting shield is absolutely constant, or at least changes by less than 10^{-8} G/month.

VI. The Superconducting Accelerometer Detector

The motion of the end of the vibrating bar will be monitored by measuring the acceleration of the end of the bar. This detection scheme has the advantage that it is sensitive to the motions of any longitudinal mode of the bar and hence allows the use of the bar for a spectral analysis of the gravitational wave. In addition, by using entirely superconducting circuits, the accelerometer gains stability and freedom from noise not obtainable by any other means.

The accelerometer consists of a superconducting mass which is supported on a persistent-current magnetic field. The current which creates the supporting magnetic field is rigidly attached to the accelerometer case. If the case is accelerated, the distance between the supported mass and the support circuit will change, and hence the inductance of the support circuit will change. This inductance change can be monitored either by monitoring the current in the persistent-current support using a superconducting magnetometer, or by rf measurement of the inductance. The measurement of current by a superconducting magnetometer is straightforward and consists of monitoring the magnetic flux in an inductance in series with the support circuit. Sensitivities of 0.01 flux unit are routinely obtained and are adequate to detect the motion of the bar at helium temperatures and below.

Another detection scheme, which will offer more rapid response time, consists of tuning the support circuit at a high resonant frequency, e.g., 30 MHz. The voltage across the tuned circuit is measured when the circuit is driven by a constant-current generator, and the out-of-phase component is extracted by synchronous detection. This signal can be fed back to hold the mass at a constant position with respect to the detection inductance. The readout of the acceleration is proportional to the feedback current. This detection scheme offers a sensitivity to motions of 10^{-17} cm at 1000 Hz. The use of a persistent current support for the test mass guarantees the absolute stability of the support and also assures that no noise other than noise of thermal origin will be introduced in this stage of the detector. There is very little thermal noise because of the low temperature. A superconducting magnetic support is also friction-free, insuring that the accelerometer will have a minimum dead zone and no hysteresis.

VII. Summary

It is foolhardy to talk of much greater sensitivity for very difficult measurements unless a radically new experimental procedure is envisioned. The use of low temperatures is such a procedure and offers much more than the necessary reduction of thermal noise of the massive detector. It enables the use of an accelerometer detector of absolute stability and great sensitivity. The gravitational detector can be supported magnetically, thus decreasing vibrational coupling to the very minimum and allowing ultra low-temperature techniques to be used. The procedure permits the entire experiment to be surrounded by a perfect superconducting shield, thus making it impossible for the bar to be excited by electromagnetic radiation at any frequency. In short, the construction of a low-temperature gravitational radiation detector promises the greater sensitivity necessary to detect radiation from other galaxies and also provides solutions to the problems which limit the magnitude of the signal that can be observed with room-temperature apparatus.

Properties of "Hollow Square" Gravitational Wave Detectors

D. H. Douglass, Jr.
University of Rochester

I. Introduction

There is a consensus among theorists that various astrophysical events should produce gravitational radiation. A supernova in its early stages should produce a pulse of gravitational radiation with a characteristic frequency in the kHz band; two gravitationally bound masses spiraling to collapse will produce a "chirp" of gravitational radiation; the radiation power may reach sizeable magnitudes also in the kHz band; when a mass, m, falls into a "black hole," a gravitational radiation pulse followed by "ringing" should occur. It is obvious that conformation (or lack of conformation) of the prediction of gravitational radiation coming from these or any other astrophysical sources can only come from experiment. Weber (Ref. 1) has set up an antenna system to look for radiation at 1661 Hz and has recently reported events which may be interpreted as pulses of gravitational radiation coming from the center of our galaxy (or anywhere in the plane defined by the earth's axis and the galactic center). Measurements at other frequencies would be of great importance in establishing the nature of these events.

Other properties of gravitational radiation that would be of interest are the directivity, the polarization, and the scalar and tensor character of the wave. I will describe an experimental program to develop antennas which have many of these properties.

These devices are mechanically resonant systems in the shape of "hollow squares" with a fundamental frequency in the 100-Hz range. The cross section, for absorption of quadrupole gravitation radiation at this frequency, can be made comparable to that achieved by Weber (Ref. 1); in addition, this mode is excited only by one polarization. The next highest mode of vibration is sensitive to scalar gravitational radiation (if it exists). A most remarkable property of these antennas is that the absorption cross section for quadrupole radiation at the higher normal frequencies

$$\omega_n = n^2 \omega_1$$

of vibration is the same as that of the fundamental if the damping factors are the same. This means that these devices can be operated at any of these higher frequencies and under suitable conditions can have a cross section at high frequencies larger than can be achieved by a cylinder-type detector.

The basic idea behind these new antennas is that mechanical structures vibrate under flexure at frequencies lower than the frequencies induced by longitudinal sound waves. To be more specific, for rods and cylinders of length, ℓ, the longitudinal vibration frequency, ω, is related to ℓ by $\omega \propto \ell^{-1}$; whereas these same rods and cylinders vibrate under flexure at a frequency which is proportional to ℓ^{-2}. We have previously classified gravitational antennas into these two groups (Ref. 2).

Class I: $\omega \propto \ell^{-1}$

Class II: $\omega \propto \ell^{-2}$

The particular structures that we are developing are in the form of "hollow squares" (see Fig. 1). These Class-II systems typically vibrate in their fundamental mode at a frequency two orders of magnitude lower than a Class-I system of the same size and mass.

We now proceed to calculate the cross section for absorption of quadrupole gravitational radiation for a Class-I detector of cylindrical cross section and for a Class-II detector shaped like a "hollow square."

II. Class-I Detectors

The equation of motion of a mass element moving parallel to the cylinder axis due to interaction of a (sinusoidal) gravitational wave $Be^{i\omega t}$ propagating parallel to the X-axis (see Fig. 1) is

$$E^* \frac{\partial^2 \xi}{\partial y^2} - \rho \frac{\partial^2 \xi}{\partial t^2} = \rho c^2 y B e^{i\omega t} \tag{1}$$

where ξ is the longitudinal displacement of the mass element from equilibrium and ρ is the density. Damping is represented through the complex part of Young's modulus

$$E^* = E(1 + j/Q) \tag{2}$$

and Q is the familiar quality factor.

The sinusoidal solution $\xi e^{i\omega t}$ is obtained by requiring that $\partial\xi/\partial y = 0$ at $y = \pm \ell/2$ and is

$$\xi = \frac{\rho c^2 B}{E^* k^3}\left[-\frac{\sin ky}{\cos (k\ell/2)} + ky\right] \tag{3}$$

where

$$k^2 = \frac{\omega^2 \rho}{E^*} \tag{4}$$

The amplitude becomes large at the resonant frequencies

$$\omega_n = \sqrt{\frac{E}{\rho}}\frac{\Pi}{\ell} n, \qquad n = 1, 3, 5, \cdots \tag{5}$$

$\omega_1 n$

and is

$$\xi_n \simeq -\frac{i4Q\ell^3 \rho c^2 B}{(n\Pi)^4 E} \sin k_n y \tag{6}$$

where n is the index corresponding to the higher modes of oscillation. The power can be transferred to an external recording instrument is (Ref. 3)

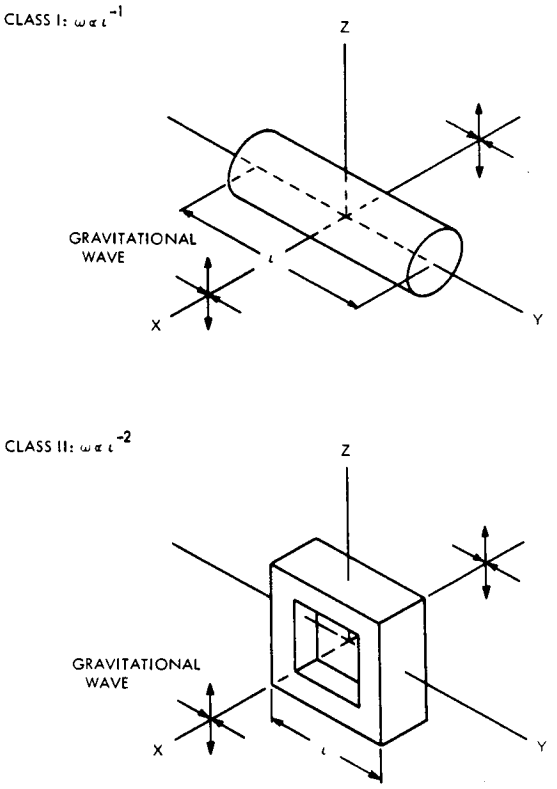


Fig. 1. Gravitational wave detectors

$$P_n = \frac{\omega_n}{Q_{ex}} \int \frac{dm\, \overline{\dot{\xi}^2}}{2} \tag{7}$$

which becomes for our case

$$P_n = \frac{\omega_n^3}{4Q_{ex}} \int dm\, \xi_n^2 \tag{8}$$

$$= \sqrt{\frac{\rho}{E}} \frac{c^4 B^4 Q_{in} M\ell^3}{4(n\Pi)^5} \tag{9}$$

where M is the total mass; Q_{in} and Q_{ex} are the internal and external quality factors and the condition (Ref. 3) for maximum power $Q_{in} = Q_{ex}$ has been used along with the relation

$$Q_{in}^{-1} + Q_{ext}^{-1} = Q^{-1}$$

At this point it is convenient to use a relationship given by Weber (Ref. 3) between the energy flux, τ, for one polarization and the amplitude, B, of the gravitational wave (under condition of negligible relative velocity between the source and antenna)

$$B^2 = \frac{4\pi G\omega^2}{c^7}\tau \qquad (10)$$

Equations (9) and (10) can be used to compute the cross section $\sigma_n(=P_n/\tau)$ at the frequency $\omega_n (n = 1,3,5, \cdots)$

$$\sigma_n = \frac{2GQ_{in}A}{c^3\pi^3 n^3} \qquad (11)$$

where

$$A = \omega_1 M\ell^2 \qquad (12)$$

The quantity A, previously defined (Ref. 2) is the only parameter that can be varied substantially in the design of a detector. It is noted that the cross section of these Class-I detectors, when operated in their higher modes, decreases as ω^{-3}; for example, the cross section of Weber's detector, $f_1 = 1661$ Hz, when operated at 4983 Hz will be down by a factor of 27.

III. Class-II Detectors

The vibrational frequencies of the "hollow square" detector, shown in Fig. 1, closely approximate those of four transversely vibrating beams with pinned ends. The lowest frequency is the quadrupole mode and corresponds to two opposite sides moving in and out in phase with each other, but opposite to the other two sides. In the next highest mode, the scalar mode, all four sides move in and out together in phase; thus scalar gravitational radiation if it exists will excite this mode. The cross section for absorption of quadrupole radiation will now be calculated.

The equation of motion for a mass element in one of the beams (the one on the +Z axis say) is

$$\rho S\frac{\partial^2\xi}{\partial t^2} + E^*I\frac{\partial^4\xi}{\partial y^4} = \frac{Mc^2Be^{i\omega t}}{8} \qquad (13)$$

where ξ is the transverse displacement from equilibrium, S is the cross sectional area, I is the moment of inertia of the area, M/4 is the mass of the beam; the gravitational wave $Be^{i\omega t}$ is assumed to be propagating along the X-axis. The boundary conditions are

$$\xi = \frac{\partial^2\xi}{\partial y^2} = 0 \quad \text{at} \quad y = \pm\ell/2$$

The sinusoidal solution $\xi e^{i\omega t}$ is found to be

$$\xi = \frac{Mc^2B}{16E^*Ik^4}\left[\frac{\cos ky}{\cos\frac{k\ell}{2}} + \frac{\cosh ky}{\cosh\frac{k\ell}{2}} - 2\right] \qquad (14)$$

where

$$k^4 = \frac{\rho S}{E^*I}\omega^2 \qquad (15)$$

The amplitude becomes very large at the resonant frequencies

$$\omega_n = \sqrt{\frac{EI}{\rho S}}\frac{\pi^2}{\ell^2}n^2, \qquad n = 1,3,5, \cdots \qquad (16)$$

$$= \omega_1 n^2$$

and is

$$\xi_n \cong -\frac{iQ\ell^4Mc^2B}{2EI(n\pi)^5}\cos k_n y \qquad (17)$$

The power deliverable to an external instrument is evaluated by using Eq. (7); the equation corresponding to Eq. (9) is

$$P_n = \sqrt{\frac{\rho S}{EI}}\frac{c^4B^2Q_{in}M\ell^4}{8(n\pi)^4} \qquad (18)$$

where a factor of 4 has been included because there are four sides.

Using Eq. (10) the cross section is

$$\sigma_n = \frac{GQ_{in}A}{2\pi c^3} \qquad (19)$$

where

$$A = \omega_1 M\ell^2 \qquad (20)$$

as before.

One sees the remarkable fact that the cross section for absorption at the higher normal frequencies $\omega_n = n^2\omega_1$ $(n = 1,3,5, \cdots)$ is the same as the cross section of the fundamental; this is in contrast to the Class I detector which fell off as n^{-3} (see Eq. 11). If the Q's of the higher modes are comparable to that of the fundamental, then these "hollow square" detectors will have a larger cross section at some frequency than can be achieved by a Class-I detector operating in its fundamental. Other properties of the "hollow square" detector are that for the directions considered in this paper the detector does not interact with the other polarization of quadrupole radiation. Also, quadrupole radiation does not excite the scalar modes and scalar radiation will not excite the quadrupole modes.

IV. Experimental Program

Prototype aluminum "squares" have been constructed with these parameters: $\ell \sim 10^2$ cm, $M \sim 10^5$ g, $f_1 \sim 10^2$ Hz. Quality factors of the order 10^5 have been achieved.

A detector is being designed to operate at $f_1 = 66.4$ Hz with $\ell \sim 5 \times 10^2$ cm, $M \sim 10^7$; this would yield a cross section for quadrupole radiation comparable to that achieved by Weber. As noted above the cross section in the higher modes $\omega_n = n^2 \omega_1$ is the same as for the fundamental (assuming the Q's are the same). Thus the quadrupole gravitational power spectrum can be measured at many discrete frequencies simultaneously; for example, $f_5 = 5^2 f_1 = 1661$ Hz which is the frequency of Weber's detector and reveals the reason for the particular choice of f_1. In addition one can also monitor the scalar modes of vibration to observe if there is any scalar radiation component.

V. Conclusions

In concluding this paper, I would like to mention that f_1 can be lowered by mass loading of the square without serious degradation of the Q. Thus it will be possible to change the f_1 of this "square" to 60.4 Hz, the frequency at which the crab pulsar should be emitting gravitational radiation if it is a rotating neutron star with a transverse quadrupole moment. Estimates (Ref. 4) for the amount of gravitation radiation coming from the crab pulsar do not cause one to be overly optimistic about obtaining a positive result. Since this experiment can be "done for nothing," very little is lost if no radiation is observed.* One should also keep in mind that theorists are sometimes wrong.

Acknowledgment

I wish to acknowledge many discussions with J. A. Tyson and the general assistance of L. Cooley and B. Smith in the experimental measurements.

This research is supported in part by ARO(D) and the National Science Foundation.

References

1. Weber, J., Phys. Rev. Letters, 25, 180, 1970.
2. Douglass, D. H., Gravity Essay, 1970. Douglass, D. H., and Tyson, J. A., Nature (to be published).
3. Weber, J., General Relativity and Gravitational Waves, Interscience Publishers, Inc., New York, 1961. Also appears in Phys. Rev. Letters, 17, 1228, 1966; Phys. Rev. Letters, 18, 498, 1967.
4. Ostiker, J. P., and Gunn, J. E., Astrophys. J., 157, 1395, 1969.

*Not observing radiation allows upper limits on its value to be set.

On the Heterodyne Method of Detecting Gravitational Waves

V. B. Braginsky and V. S. Nazarenko
Moscow State University

The idea of the proposed method (Ref. 1) can be explained as follows. Suppose that a dumbbell consisting of two masses M is situated in a plane P, and that a plane-fronted gravitational wave is traveling perpendicular to P. In this case, a pair of equal and opposite forces F^{μ} acts on the masses (Ref. 2):

$$F^{\mu} = -Mc^2 R^{\mu}_{o\alpha o} \ell^{\alpha} \qquad (1)$$

where $R^{\mu}_{o\alpha o}$ is the Riemann curvature tensor and ℓ^{α} is the vector connecting the two masses. Taking into account the polarization of the gravitational wave (Ref. 2), one can see that this pair of forces produces a torque on the dumbbell. The magnitude of this torque depends on the angle ϕ between ℓ^{α} and one of the two polarization axes of the wave, and is given by

$$\mathcal{T}_g F_g = 2F_o r \cdot \sin 2\phi \cdot \sin \omega_o \tau \qquad (2)$$

Here ω_o is the angular frequency of the wave, r is the radius of the dumbbell (r = 1/2 l); F_o is the amplitude of F^{μ} and is equal to $M\omega_o l(8\pi Gt/c^3)^{1/2}$ (Ref. 3), G is the gravitational constant, and t is the energy flux (ergs/cm^2/s) in the wave. Using this formula for F_o, we have

$$\mathcal{T}_g F_g = 4M\omega_o r^2 \left(\frac{8\pi Gt}{c^3}\right)^{1/2} \cdot \sin 2\phi \cdot \sin \omega_o \tau \qquad (3)$$

Suppose now that the dumbbell is rotating freely around the axis perpendicular to P with some angular velocity ω_1 near 1/2 ω_o. In this case, $\mathcal{T}_g$ contains a component that oscillates very slowly (angular velocity $\omega_o - 2\omega_1$):

$$\mathcal{T}_g F_g = 2M\omega_o r^2 \left(\frac{8\pi Gt}{c^3}\right)^{1/2} \times \left[\cos (\omega_o - 2\omega_1)\tau - \cos (\omega_o + 2\omega_o)\tau\right] \qquad (4)$$

It is this change in frequency from ω_o to $\omega_o - 2\omega_1$ which is the main point of the proposed method. It should be noted that such a change can also be obtained by modulating r instead of rotating the dumbbell.*

A small torsional rigidity can be used to couple the dumbbell to a platform that rotates with the same velocity ω_1 and serves as a reference frame. This will turn the detector into a high-Q oscillator with a resonant frequency $\Omega = \omega_o - 2\omega_1$. The result is a substantial in the amplitude of the angular displacements. This amplitude, $\Delta\phi$, produced by a wave train of

*This suggestion is due to A. D. Sacharov.

duration $\hat{\tau}$, can be derived as a solution of the familiar equation for forced oscillations:

$$\Delta\phi\Big|_{\hat{\tau}} = \frac{\omega_o \hat{\tau}}{2\Omega}\left(\frac{8\pi G t}{c^3}\right)^{1/2} \qquad (5)$$

A capacitor type of transducer can be used for recording these oscillations.

The obvious advantage of this detector is its capability of being "tuned" over a rather wide range of wave frequencies by changing the angular velocity of rotation ω_1.

The other advantage is its very high sensitivity. The sensitivity is restricted by the Nyquist force (thermal fluctuations), and therefore by the friction between the dumbbell and the laboratory. The condition for detectability of a torque $\mathcal{T}_g$ that lasts for a time $\hat{\tau}$ (length of wave train) can be derived from the following general equation:

$$\mathcal{T}_g \geq (4kTH\phi\Delta f)^{1/2}; \qquad \Delta f \approx \frac{1}{\hat{\tau}} \qquad (6)$$

Here T is the temperature of the thermal fluctuations and $H\phi$ is the coefficient of friction.

The magnitude of $H\phi$ depends on the particular type of torsional rigidity chosen. If the dumbbell is suspended on a thin fused quartz thread and the vacuum around it is high enough, $H\phi$ is equal to $Mr^2\Omega^2\eta N^{-1}$. Then, combining Eq. (4) and Eq. (6) gives the following estimate of the weakest detectable gravitational wave flux $[t]_{min}$:

$$[t]_{min} \geq \frac{c^3}{8\pi G} \cdot \frac{kT}{\hat{\tau}} \cdot \frac{\Omega^2}{\omega_o^2} \cdot \frac{\eta}{NMr^2} \qquad (7)$$

where η is the quartz viscosity and N is the modulus of rigidity. For example, if $T = 300°K$; $M = 10^3$ g, $\Omega = 10^{-2}$ s^{-1}, $\hat{\tau} = 10^5$ s, $\omega_o = 10^3$ s^{-1}, and $r = 60$ cm, we obtain $[t]_{min} = 3 \times 10^{-4}$ ergs/s cm^2.

The sensitivity can be improved by using a form of rigidity with less dissipation. The ultimate limit of sensitivity is determined by the random forces, with which the recording device acts on the dumbbell. As shown in Ref. 4, if a recording device is set up in an optimal way and a mechanical oscillator is ideal, then the weakest force whose effect on the oscillator can be detected is given by

$$[F]_{min} \geq \frac{2}{\hat{\tau}} \sqrt{\hbar\Omega M}, \qquad (8)$$

where $\hbar$ is Planck's constant.

Using Eqs. (4) and (8), the ultimate sensitivity limit of the heterodine detector can be derived:

$$[t]_{min} = \frac{c^3}{8\pi G} \cdot \frac{\hbar\Omega}{\omega_o^2 \hat{\tau}^2 M r^2}. \qquad (9)$$

For values of Ω, M, ω_o, $\hat{\tau}$, and r the same as above, we have $[t]_{min} = 1.5 \times 10^{-14}$ ergs/s cm^2.

It should be noted that, according to Weber's report (Ref. 2), the total output noise in his experiments has its origin in thermal acoustical waves within the body of the detector. That means that the sensitivity of Weber's detectors and other detectors of the same type, estimated (Ref. 5) as 1×10^6 ergs/s cm^2 for $T = 300°K$, is not likely to be improved substantially without employing very complicated and expensive superlow-temperature techniques.

References

1. Braginsky, V. B., Zeldovitch, Ja. B., and Rudenko, V. N., JETP Letters, 10, 437, 1969.

2. Weber, J., Phys. Rev. Letters, 22, 1320, 1969.

3. Braginsky, V. B., Uspekchi (USSR), 83, 433, 1965.

4. Braginsky, V. B., JETP, 53, 4 (10), 1434, 1967.

5. Braginsky, V. B., and Rudenko, V. N., Uspekchi, 100, 3, 1970.

Observations of Earth Eigen Vibrations Possibly Excited by Low-Frequency Gravity Waves

V. S. Tuman
Stanislaus State College, California

I. Introduction

During the last 6 months, over two-hundred eigen vibrations of the earth have been identified from three different records of a cryogenic gravity meter obtained during February 18-20, February 27-March 2, and on July 31, 1970. The interpretation is made on a Fourier transform of the digitized record.

The cryogenic gravity meter is made of two parts -- a magnetic suspension unit and a detection module. An upward push is applied to a hollow niobium sphere from a specially designed magnetic field with an appropriate field gradient. The hollow niobium sphere, which is superconducting at a liquid helium temperature of 4.2°F, has a mass of 2.4 g, a diameter of 2.54 cm, and a thickness of about 100 μm. By adjusting the push to be equal to the gravitational pull, the hollow niobium sphere is made to float freely at the geometric center of the instrument.

The necessary magnetic field and field gradient are generated by energizing a set of superconducting coils made of niobium-zirconium alloy wire. Once the sphere is made to float at a particular datum, the superconducting magnets are made persistent. The detection module is a double Josephson junction magnetometer, also known as Superconducting Quantum Interference Device (SQUID) (Refs. 1 and 2). With the aid of the detection module, a vertical motion of less than an angstrom can be sensed by this unit. The hollow niobium sphere floats freely in liquid helium, and consequently the system forms a relatively low Q oscillator. The instrument, which is immersed in a bath of liquid helium, is protected from the earth magnetic perturbations by a niobium canister and a mu-metal shield. The detailed design of the individual module of the gravity meter and the calibration of the SQUID are discussed in Ref. 3. The output signal depicts the change in the gravitational field at the observatory. The theoretical sensitivity of the instrument is about one part in 10^{11} of the earth field.

II. Discussion

Some of the interesting features of the earth eigen vibrations observed by this instrument have already been discussed elsewhere (Refs. 4, 5 and 6). Because of new evidence and some numerical estimations of the energies involved in the low-level earth modes, I would like in this article to discuss further the possibility of earth spheroidal oscillations excited by the tensor gravity waves.

On two out of three records, I had observed that the energy content of the fundamental even eigen vibrations of the earth denoted by ${}_0S_2$, ${}_0S_4$, ${}_0S_6 \cdots {}_0S_{2n}$ with quadrupole moments was higher than the energy content of the odd harmonics ${}_0S_3$, ${}_0S_5$, $\cdots$ ${}_0S_{2n+1}$.

Although it is feasible that an unknown earthquake mechanism within the earth is capable under certain conditions of giving rise to such amplitude modulations (Ref. 5), for the present the author is not aware of such a mechanism. For this reason, it was speculated in Ref. 5 that the tensor gravity waves interacting with the earth may give rise to the observed anomalous effect. Such are the predictions using general relativity. Formerly, because of lack of theoretical data of a model of the earth, I had identified only the fundamental ${}_0S_{2n}$ modes, the first and second overtones ${}_1S_{2n}$ and ${}_2S_{2n}$ modes, all of which had higher energy

content compared to the non-symmetric eigen vibrations ${}_0S_{2n+1}$, ${}_1S_{sn+1}$, and ${}_2S_{2n+1}$.

After receiving the theoretical periods of higher overtones based on a model of the earth (Ref. 7), we have now observed the same trend in third and fourth overtones--namely, the energy content of ${}_3S_{2n}$ and ${}_4S_{2n}$ is also higher than the energy content of ${}_3S_{2n+1}$ and ${}_4S_{2n+1}$. Furthermore, we have divided the 60-h record of February 27 - March 2 into two sections. The Fourier analysis of the first 30 h indicates the amplitude modulations, with even eigen vibrations having higher energy content. The trend does not appear regularly among fundamentals and the overtones of the second half of the record. This fact indicates that necessary energy was coupled to the earth prior and just during the first 30 h of recording. All these observations have strengthened the position of the anomalous effect within our records. The source of this effect is either a mysterious type of earthquake, or the low-frequency tensor gravity waves coming from a binary star system. The energy difference between the even and odd harmonics is estimated to be about 10^{14} - 10^{15} erg.

III. Results

The position of the hollow niobium sphere at any time is a function of the earth gravitational field at that moment. As the observatory is displaced upward away from the center of the earth, the earth gravitational field decreases by a small amount, and consequently the niobium hollow sphere also is pushed up by the unbalanced magnetic force. This upward motion of the sphere is detected by the magnetic flux change produced at the pickup coil of the detection module. The instrument output is a voltage depicting the motion of the hollow niobium sphere. The output is recorded on a chart recorder and later digitized, and with the aid of a digital computer, a Fourier transform is performed on the data. The results for the first 30 h of the second recording for fundamental eigen vibrations and first to fifth overtones are given in Tables 1 through 6.

Table 1. Fundamental eigen vibrations

${}_0S_n$ mode	Period, s	Relative amplitude	Relative energy	Cumulative energy	${}_0S_n$ mode	Period, s	Relative amplitude	Relative energy	Cumulative energy
2	3233				3	2133.3			
4	1551	2.037	4.149		5	1177.0	1.413	2.000	
6	966.0	0.497	0.248		7	812.0	0.510	0.260	
8	706.2	1.948	3.800		9	632.0	1.447	2.160	
10	578.5	1.103	1.220		11	536.1	1.092	1.190	
			9.417	9.417				5.610	5.610
12	502.0	0.455	0.207		13	469.7	1.245	1.550	
14	449.1	0.784	0.616		15	426.7	0.477	0.228	
16	406.3	0.720	0.518		17	389.5	0.655	0.428	
18	373.7	0.995	0.990		19	359.3	0.584	0.342	
20	348.3	0.628	0.396		21	336.8	0.610	0.372	
			2.727	12.144				2.920	8.530
22	324.1	0.585	0.342		23	316.0	0.621	0.386	
24	305.7	1.071	1.145		25	296.8	0.629	0.396	
26	290.1	1.016	1.032		27	282.8	0.853	0.728	
28	275.3	1.157	1.335		29	268.1	0.578	0.334	
30	261.9	1.171	1.371		31	256.0	0.658	0.434	
			5.225	17.369				2.278	10.808
32	249.8	0.795	0.632		33	244.4	0.634	0.403	
34	239.8	0.588	0.346		35	235.9	0.943	0.890	
36	230.6	0.683	0.467		37	224.6	1.041	1.082	
38	220.7	0.543	0.295		39	216.0	0.539	0.291	
40	212.0	0.421	0.178		41	209.0	0.874	0.764	
			1.918	19.287				3.430	14.238

Table 2. First overtones of eigen vibrations

$1S_n$ mode	Period, s	Relative amplitude	Relative energy	Cumulative energy	$1S_n$ mode	Period, s	Relative amplitude	Relative energy	Cumulative energy
2	1462.9	1.433	2.05		3	1077.7	1.664	2.77	
4	846.3	1.771	3.14		5	731.4	1.422	2.02	
6	656.4	1.503	2.26		7	609.5	1.100	1.21	
8	556.5	1.155	1.33		9	512.0	0.417	0.17	
10	467.6	1.808	3.27		11	426.7	0.477	0.23	
			12.05	12.05				6.40	6.40
12	396.9	0.867	0.75		13	364.4	0.777	0.60	
14	337.9	0.610	0.37		15	319.0	0.787	0.62	
16	303.9	0.640	0.41		17	287.6	0.752	0.57	
18	278.3	0.719	0.52		19	266.0	0.430	0.18	
20	256.0	0.658	0.43		21	246.7	0.570	0.32	
			2.48	14.53				2.29	8.69
22	237.6	0.640	0.41		23	230.1	0.683	0.47	
24	224.0	0.679	0.46		25	213.8	0.714	0.51	
26	207.7	0.770	0.59		27	202.0	0.578	0.33	
28	196.2	0.390	0.15		29	191.4	0.354	0.13	
30	186.2	0.710	0.50		31	181.9	0.529	0.28	
			2.11	16.64				1.72	10.41

Table 3. Second overtones of eigen vibrations

$2S_n$ mode	Period, s	Relative amplitude	Relative energy	Cumulative energy	$2S_n$ mode	Period, s	Relative amplitude	Relative energy	Cumulative energy
4	721.1	1.061	1.13		5	656.4	1.503	2.26	
6	591.9	1.306	1.71		7	536.1	1.092	1.19	
8	487.6	1.403	1.97		9	449.1	0.784	0.61	
10	416.3	0.824	0.68		11	389.4	0.655	0.43	
			5.49	5.49				4.49	4.49
12	368.3	1.548	2.40		13	346.0	0.119	0.01	
14	327.2	0.696	0.48		15	308.4	0.446	0.20	
16	291.7	0.919	0.84		17	273.8	0.455	0.21	
18	258.6	0.817	0.67		19	244.9	0.634	0.40	
20	231.7	0.952	0.91		21	220.7	0.542	0.29	
			5.30	10.79				1.11	5.60
22	210.7	0.668	0.45		23	200.8	0.648	0.42	
24	193.6	0.764	0.58		25	186.9	0.558	0.31	
26	181.2	0.885	0.78		27	175.6	0.408	0.17	
28	171.5	0.456	0.21		29	166.5	0.679	0.44	
30	162.3	0.574	0.33		31	158.3	0.334	0.11	
			2.35	13.14				1.45	7.05

Table 4. Third overtones of eigen vibrations

$_3S_n$ mode	Period, s	Relative amplitude	Relative energy	Cumulative energy	$_3S_n$ mode	Period, s	Relative amplitude	Relative energy	Cumulative energy
2	585.1	1.115	1.33		3	489.9	0.840	0.71	
4	439.5	1.227	1.51		5	416.3	0.824	0.68	
6	393.8	0.867	0.75		7	374.4	0.835	0.70	
8	354.3	1.243	1.55		9	341.3	0.485	0.24	
10	322.0	0.791	0.63		11	312.2	0.427	0.18	
			5.77	5.77				2.51	2.51
12	299.4	0.606	0.37		13	286.0	0.354	0.13	
14	273.8	0.455	0.21		15	263.2	0.380	0.14	
16	251.0	0.607	0.37		17	242.1	0.972	0.94	
18	233.8	1.080	1.17		19	224.1	0.607	0.37	
20	217.4	0.404	0.16		21	210.7	0.668	0.42	
			2.28	8.05				2.00	4.51
22	203.6	0.324			23	1969.9	0.463		

Table 5. Fourth overtones of eigen vibrations

$_4S_n$ mode	Period, s	Relative amplitude	Relative energy	Cumulative energy	$_4S_n$ mode	Period, s	Relative amplitude	Relative energy	Cumulative energy
2	478.5	1.301	1.69		3	463.3	0.521	0.27	
4	419.7	1.238	1.53		5	371.0	1.021	1.04	
6	330.3	0.739	0.55		7	303.9	0.490	0.24	
8	282.9	0.853	0.73		9	270.2	0.552	0.30	
10	258.6	0.818	0.67		11	249.8	0.796	0.63	
			5.17	5.17				2.48	2.48
12	242.0	0.973	0.95		13	233.2	0.699	0.49	
14	226.5	0.807	0.65		15	218.3	0.839	0.70	
16	212.0	0.422	0.18		17	205.2	0.422	0.18	
18	199.2	0.701	0.49		19	192.8	0.523	0.27	
20	187.5	0.509	0.26		21	179.6	0.766	0.59	
			2.27	7.44				2.23	4.71

Table 6. Fifth overtones of eigen vibrations

$_5S_n$ mode	Period, s	Relative amplitude	Relative energy	Cumulative energy	$_5S_n$ mode	Period, s	Relative amplitude	Relative energy	Cumulative energy
2	401.6	1.690	2.86		3	356.8	0.817	0.67	
4	325.1	0.411	0.17		5	307.5	0.120	0.01	
6	294.3	0.840	0.71		7	279.8	0.953	0.91	
8	268.1	0.578	0.33		9	253.5	0.427	0.18	
10	237.6	0.640	0.41		11	224.7	1.041	1.08	
			4.48	4.48				2.85	2.85
12	212.9	0.743	0.55		13	202.8	0.563	0.32	
14	195.0	0.696	0.48		15	189.6	0.655	0.43	
16	182.5	0.747	0.56		17	177.8	0.743	0.55	
18	172.7	0.349	0.12		19	167.0	0.415	0.17	
20	163.0	0.359	0.13		21	159.0	0.904	0.82	
			1.84	6.32				2.29	5.14

Figure 1 shows the cumulative energy of fundamental first and second overtones, while Fig. 2 shows the cumulative energy for third, fourth, and fifth overtones. The energy modulation for individual eigen vibrations is given in Figs. 3, 4, and 5.

When we compare the first 30 h of the record with the second 30 h, we note a very distinct shift in the period. In fact, when a plot of data is made, three sets of curves emerge. In one group, there is no shift in the frequency; the second and third groups indicate a pronounced change in period (shift in the frequency). The shift of the second group is larger than that of the first group. These results are given in Fig. 6. One is tempted to interpret Fig. 6 as an indication that the earth was being stimulated by a driving mechanism during the first 30 h, while it was oscillating freely during the second half of the record.

Acknowledgment

I would like to express my thanks to Drs. William Fairbank, Francis Everitt, and Arthur Hebard of Stanford University, Dr. Bruce Bolt of Berkeley, and to Mr. Robert Uhrhammer, my assistant, for many stimulating discussions. I am grateful to Drs. F. Gilbert and John Derr for providing me with theoretical data of overtones of earth eigen vibrations. This research is supported by the Air Force Cambridge Laboratories.

References

1. Jacklevic, et al., Phys. Rev. Letters, 12, 274, 1964.

2. Hebard, A., Quark Meter, personal communication, 1969, and Ph. D. Thesis, Stanford, 1970.

3. Tuman, V. S., Cryogenic Gravity Meter, Report No. AFCRL-70-0449.

4. Tuman, V. S., Cryogenic Gravity Meter, Proceedings of L. T-12, Kyoto, Japan, 1970 (to be published).

5. Tuman, V. S., "Terrestrial Spectroscopy by a Cryogenic Gravity Meter," Nature (to be published).

6. Tuman, V. S., "Observation of Earth Eigen Vibrations Possibly Excited by Gravity Waves," (to be published).

7. Derr, J., Ph. D. Thesis, U. C. Berkeley, 1968.

2

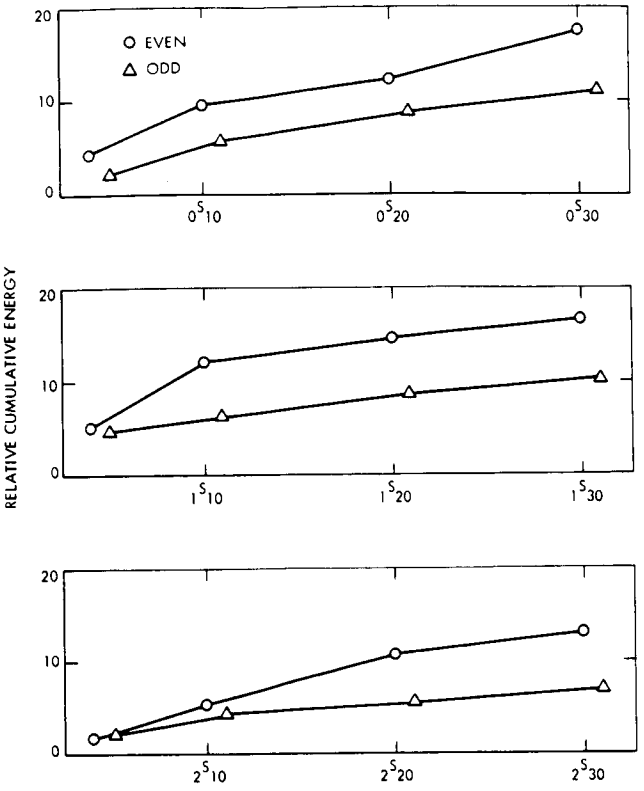


Fig. 1. Cumulative energy of fundamental first and second overtones of the earth

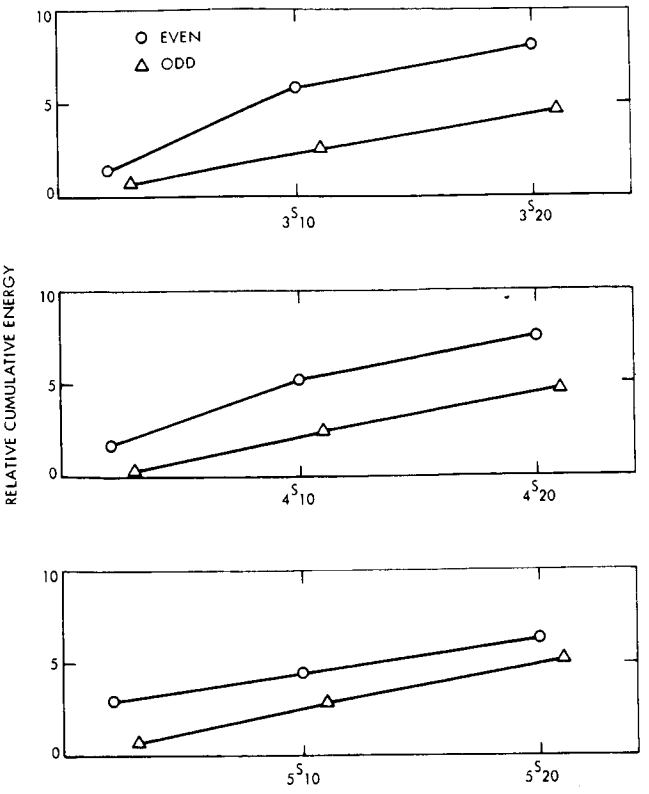


Fig. 2. Cumulative energy of third, fourth, and fifth overtones of the earth

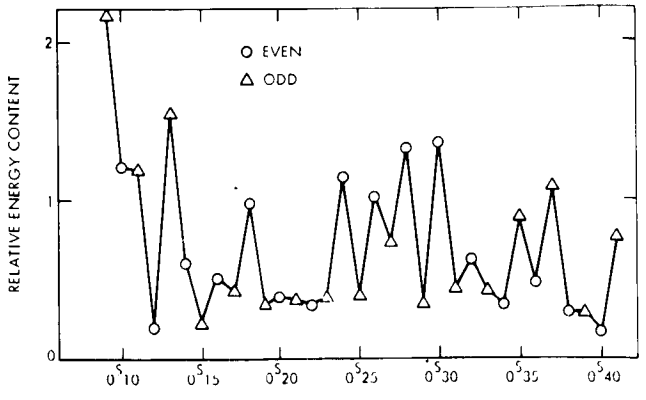


Fig. 3. Energy content of the fundamental eigen vibrations of the earth during the first 30 h of experiment 1 (Feb. 27 - Mar. 2, 1970)

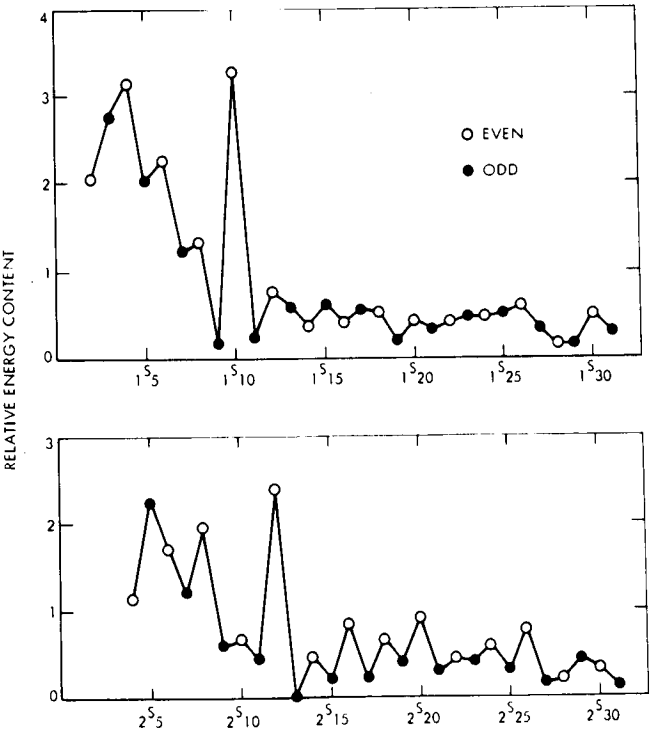


Fig. 4. Energy content of first and second overtones of the earth

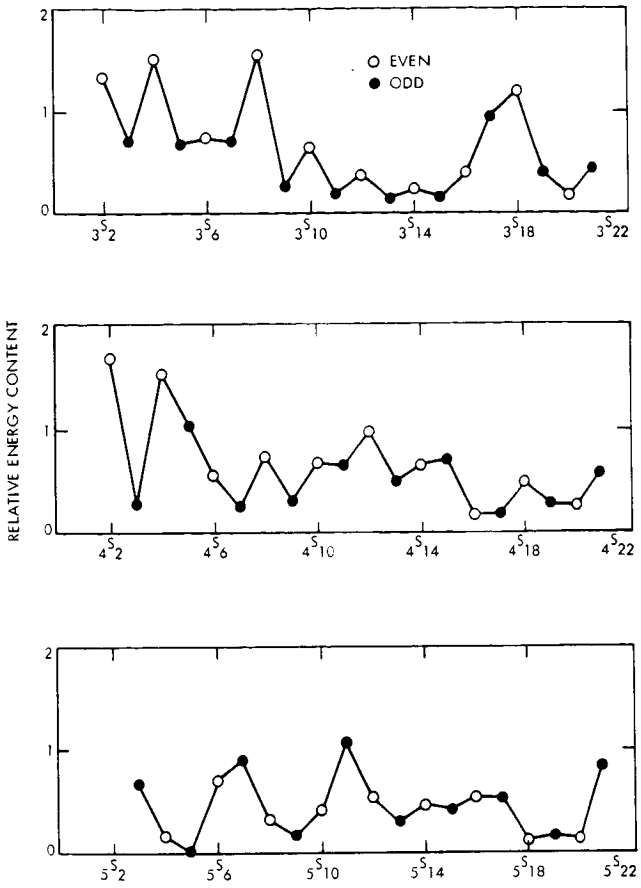


Fig. 5. Energy content of the third, fourth, and fifth overtones of the earth

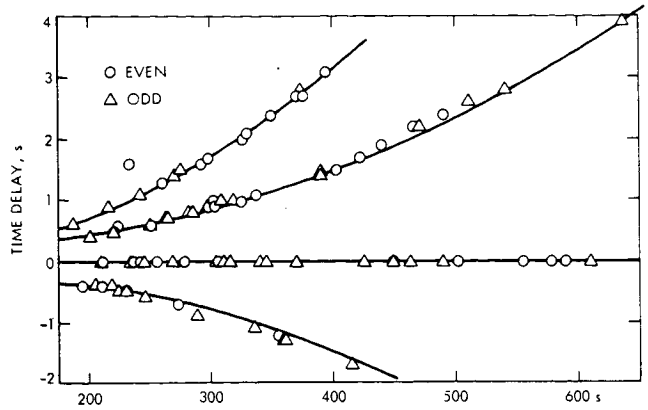


Fig. 6. Consistent time delay between first and second 30 h of experiment 2

Measurement of the Gravitational Redshift Using a Clock in an Orbiting Satellite*

R. F. C. Vessot and M. W. Levine
Smithsonian Astrophysical Observatory
and
Harvard College Observatory

I. Introduction

The proposed experiment uses a hydrogen-maser clock in a satellite to measure the gravitational effect on time scales with an accuracy substantially higher than has ever been used before. This is a test of the principle of equivalence, which asserts that there is no way of distinguishing locally between a gravitational acceleration and an oppositely directed mechanical acceleration. This principle, first stated by Einstein (Ref. 1) as a generalization of the observed proportionality between gravitational mass and inertial mass, can be justified only by experiments. Experiments were carried out by Isaac Newton using pendula of various materials; more recently, the proportionality has been tested to a few parts in 10^{11} by Roll, et al. (Ref. 2) using a highly refined Eötvös balance.

The principle of equivalence affects radiation and manifests itself in the gravitational redshift, where a source radiating in a gravitational field will appear to be shifted in frequency by a fractional amount:

$$\frac{\Delta f}{f} = \frac{\Delta \phi}{c^2}$$

where $\Delta\phi$ is the gravitational potential difference between the source and the observer.

A number of proposals have been made to test the equivalence principle for clocks (Refs. 3-7). Pound and Rebka (Ref. 8) and Pound and Snider (Ref. 9) used the extremely narrow linewidth of Fe^{57} radiation and absorption due to the Mössbauer effect in a series of experiments over a 75-ft (approximately 23-m) vertical distance. Their results confirmed the prediction of the equivalence principle to 1 part in 100.

The advances in space-flight technology over the last decade and the availability of atomic oscillators with frequency stabilities better than 1 part in 10^{14} make possible a much more sensitive test of the principle of equivalence applied to clocks. The experiment we propose is a direct test of the relation with an accuracy of 1 part in 10^5 between the rates of proper clocks located at substantially different gravitational potentials.

A discussion of this experiment has been reported recently by Kleppner, Vessot, and Ramsey (Ref. 10).

The expression describing the fractional shift in frequency of a satellite-borne oscillator observed from the earth is given by

*This work was supported in part by contract NSR 09-015-098 from the National Aeronautics and Space Administration.

$$S = \frac{f_s - f_e}{f_c} = \frac{1}{c^2}(\phi_s - \phi_e)$$

$$- \frac{1}{2c^2}\left(v_s^2 - v_e^2\right) \qquad (1)$$

where the velocity is measured in terms of an inertial frame whose origin is at the center of the earth and whose axes are aimed at the "fixed" stars, and where $(\phi_s - \phi_e)$ is the gravitational potential difference between the satellite and the earth's surface. If we neglect the effect of the earth's rotation (or locate our ground station at the north or south pole) and if we consider the earth to be spherical, we obtain the time average of the redshift:

$$\langle S_0 \rangle = \left\langle \frac{GM_e}{c^2 R_e}\left(1 - \frac{3R_e}{2a}\right)\right\rangle \qquad (2)$$

where a is the semimajor axis of the orbit, and $GM_e/c^2 R_e = 6.94 \times 10^{-10}$ is the total redshift that would result if the satellite were at rest at a very large distance from the earth. The value of $\langle S_0 \rangle$ is shown plotted in terms of orbital radius in Fig. 1. It is seen that with a 24-h orbit, we obtain 77% of the total effect due to the earth's gravity. The value of $\langle S_0 \rangle$ for these orbits is 5.37×10^{-10}.

If the orbit is eccentric, S will vary periodically, and we can describe the variations as a gravity-induced frequency modulation of the oscillator in the satellite. We can use two separate properties of the oscillator – the precision* and its stability† – by means of an eccentric orbit.

Assuming a spherical earth and no other perturbations, the time average value of S is independent of eccentricity and depends only on the orbital period. The average frequency of the satellite oscillator can be related to its previously determined value on the ground with an accuracy that is limited by the precision of the oscillator. On the other hand, the extent of modulation of the oscillator frequency due to excursions back and forth in the earth's gravitational potential can be determined with an accuracy that depends on the oscillator's stability. This stability, as will be seen later, depends on the averaging time. The applicable averaging time will depend on the period of the modulation. As in other modulated systems, the accuracy of the determination can be improved by repeating the measurement over many cycles.

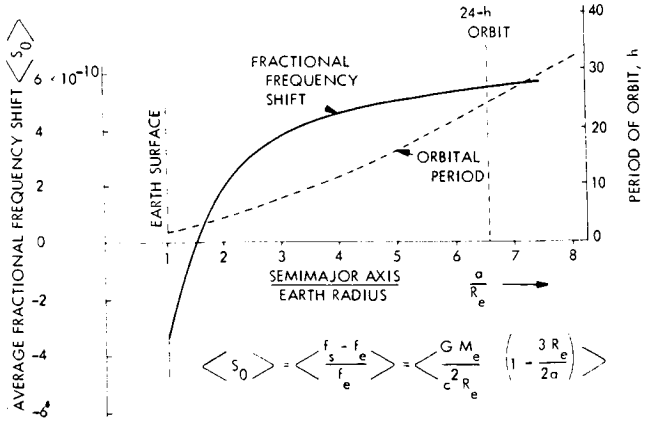


Fig. 1. Redshift and orbital period plotted versus semimajor axis in units of the earth radii

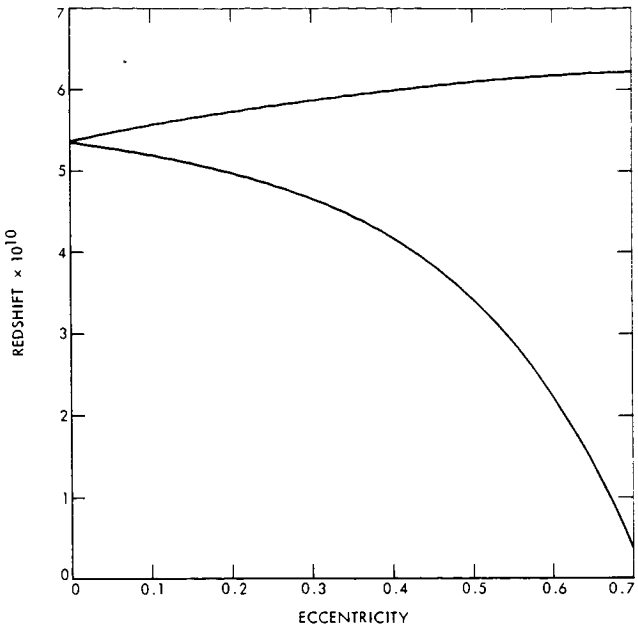


Fig. 2. Extremes of redshift versus eccentricity (The upper and lower branches give the redshift at apogee and perigee, respectively.)

For a 24-h orbit, the extremes in the value of S are plotted in Fig. 2 as a function of orbital eccentricity.

There are limits imposed on the eccentricity of the orbit by the desirability of keeping the satellite constantly in view of the ground station so that the telemetry system will operate with no interruption of carrier phase. In Fig. 3 (from Ref. 10), we show the minimum angle of elevation h_{min} as a function of eccentricity of the orbit for a 24-h orbit with inclination 28.5°. The value of h_{min} observed from earth stations at the equator and at latitude 20° are shown. On the same figure, the value of the diurnal variation is plotted versus eccentricity. Since it is desirable to keep the minimum elevation angle more than 15° above the horizon, we will look into the characteristics of

*By precision, we mean the ability of the oscillator to be independently restarted to oscillate at a previously determined frequency.

†The stability is defined here as the two-sample variance or the Allan variance (see Ref. 17).

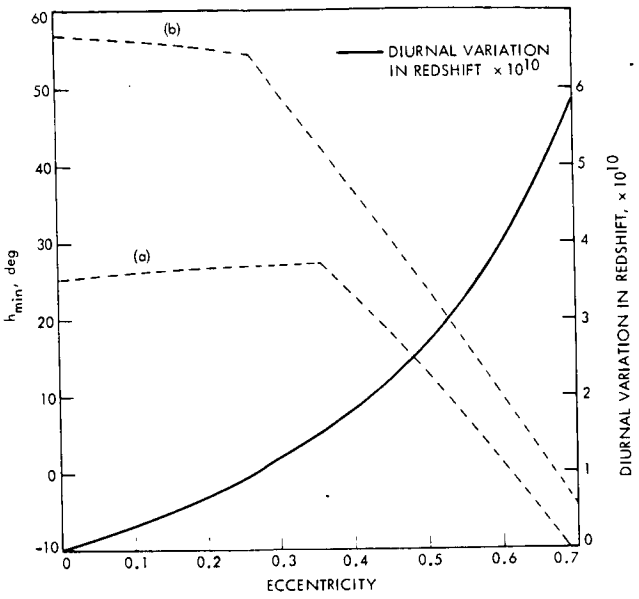


Fig. 3. Variation in redshift and minimum angle of elevation for 24-h orbit inclined at 28.5° as seen from an earth station: (a) 20° (b) 0°

an experiment performed with a 24-h orbit with eccentricity 0.52 and a ground station at latitude 20°. Under these conditions, we can expect the shift to vary from 3.09×10^{-10} to 6.08×10^{-10} with a 24-h period. The behavior with time of the redshift is plotted in Fig. 4 (from Ref. 10). Shown in the same figure as a first-order doppler effect for a carrier frequency of 2.5 GHz.

II. The Telemetry System

From Fig. 4, we see that the frequency shift we want to measure is very small compared to the doppler shift of the telemetry carrier. However, by use of a phase-coherent system, it is possible

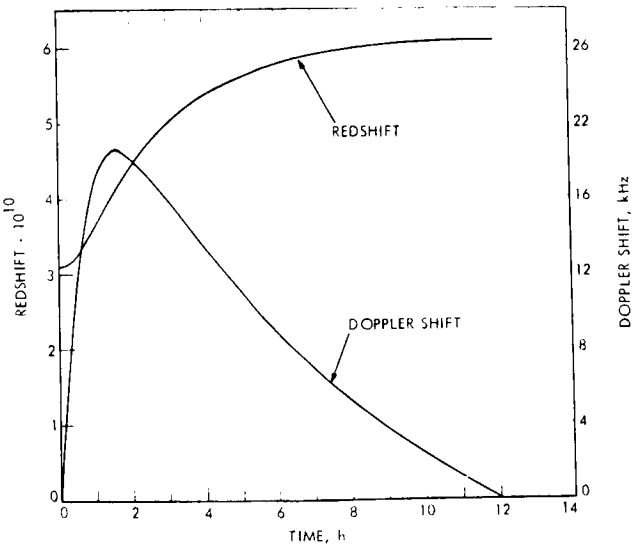
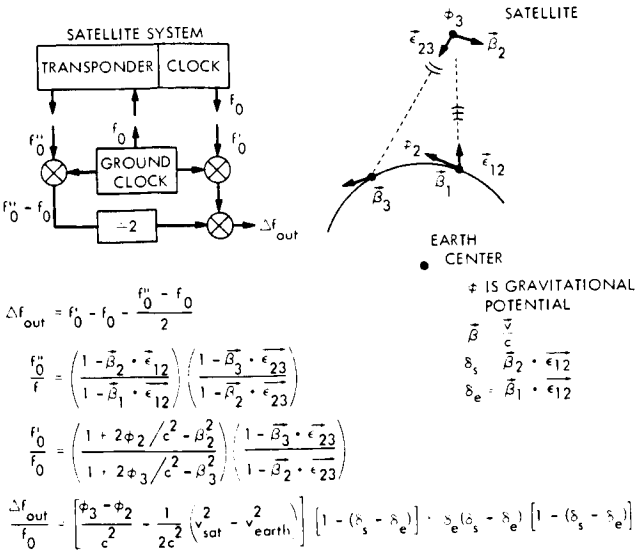


Fig. 4. Redshift and doppler shift versus time for a 24-h orbit with eccentricity 0.52 and inclination 28.5° (obtained from an earth station at 20° latitude using a 2.5-GHz carrier)



$$\Delta f_{out} = f'_0 - f_0 - \frac{f''_0 - f_0}{2}$$

$$\frac{f''_0}{f} = \left(\frac{1 - \vec{\beta}_2 \cdot \vec{\epsilon}_{12}}{1 - \vec{\beta}_1 \cdot \vec{\epsilon}_{12}}\right)\left(\frac{1 - \vec{\beta}_3 \cdot \vec{\epsilon}_{23}}{1 - \vec{\beta}_2 \cdot \vec{\epsilon}_{23}}\right)$$

$$\frac{f'_0}{f_0} = \left(\frac{1 + 2\Phi_2/c^2 - \beta_2^2}{1 + 2\Phi_3/c^2 - \beta_3^2}\right)^{1/2}\left(\frac{1 - \vec{\beta}_3 \cdot \vec{\epsilon}_{23}}{1 - \vec{\beta}_2 \cdot \vec{\epsilon}_{23}}\right)$$

$$\frac{\Delta f_{out}}{f_0} = \left[\frac{\Phi_3 - \Phi_2}{c^2} - \frac{1}{2c^2}\left(v_{sat}^2 - v_{earth}^2\right)\right]\left[1 - (\delta_s - \delta_e)\right] \cdot \delta_e(\delta_s - \delta_e)\left[1 - (\delta_s - \delta_e)\right]$$

$\vec{\beta} \equiv \frac{\vec{v}}{c}$

$\delta_s \equiv \vec{\beta}_2 \cdot \vec{\epsilon}_{12}$

$\delta_e \equiv \vec{\beta}_1 \cdot \vec{\epsilon}_{12}$

Fig. 5. Concept of doppler-canceling telemetry system

to remove the doppler shift and to extract the desired information. This system is shown in its conceptual form in Fig. 5, and its operation is described below.

Two signals are used, a clock signal from the satellite and a tracking signal that is transmitted from the ground, received at the satellite, and coherently transmitted back to the ground station. The ratio of the received signal frequency f'' to the transmitted frequency f_0 of the tracking signal is given by

$$\frac{f''_0}{f_0} = \left(\frac{1 - \vec{\beta}_2 \cdot \vec{\epsilon}_{12}}{1 - \vec{\beta}_1 \cdot \vec{\epsilon}_{12}}\right)\left(\frac{1 - \vec{\beta}_3 \cdot \vec{\epsilon}_{23}}{1 - \vec{\beta}_2 \cdot \vec{\epsilon}_{23}}\right) \qquad (3)$$

As discussed in Ref. 10, this expression results from a direct application of the special theory of relativity, and its accuracy is sufficient for the purpose of this experiment. The symbols are explained by Fig. 5.

The clock signal is transmitted from the satellite at a frequency f_0 that is proper to the clock in the satellite. This signal is received by the ground station at a frequency f'_0. The relationship between f_0 and f'_0 is given by

$$\frac{f'_0}{f_0} = \left(\frac{1 + (2\Phi_2/c^2) - \beta_2^2}{1 + (2\Phi_3/c^2) - \beta_3^2}\right)^{1/2}\left(\frac{1 - \vec{\beta}_3 \cdot \vec{\epsilon}_{23}}{1 - \vec{\beta}_2 \cdot \vec{\epsilon}_{23}}\right) \qquad (4)$$

This expression results directly from the principle of equivalence. It also results from the general theory of relativity when terms in v^2/c^2 and ϕ/c^2 are taken to first order.

When applied to the simple case where the radial accelerations are small and when the

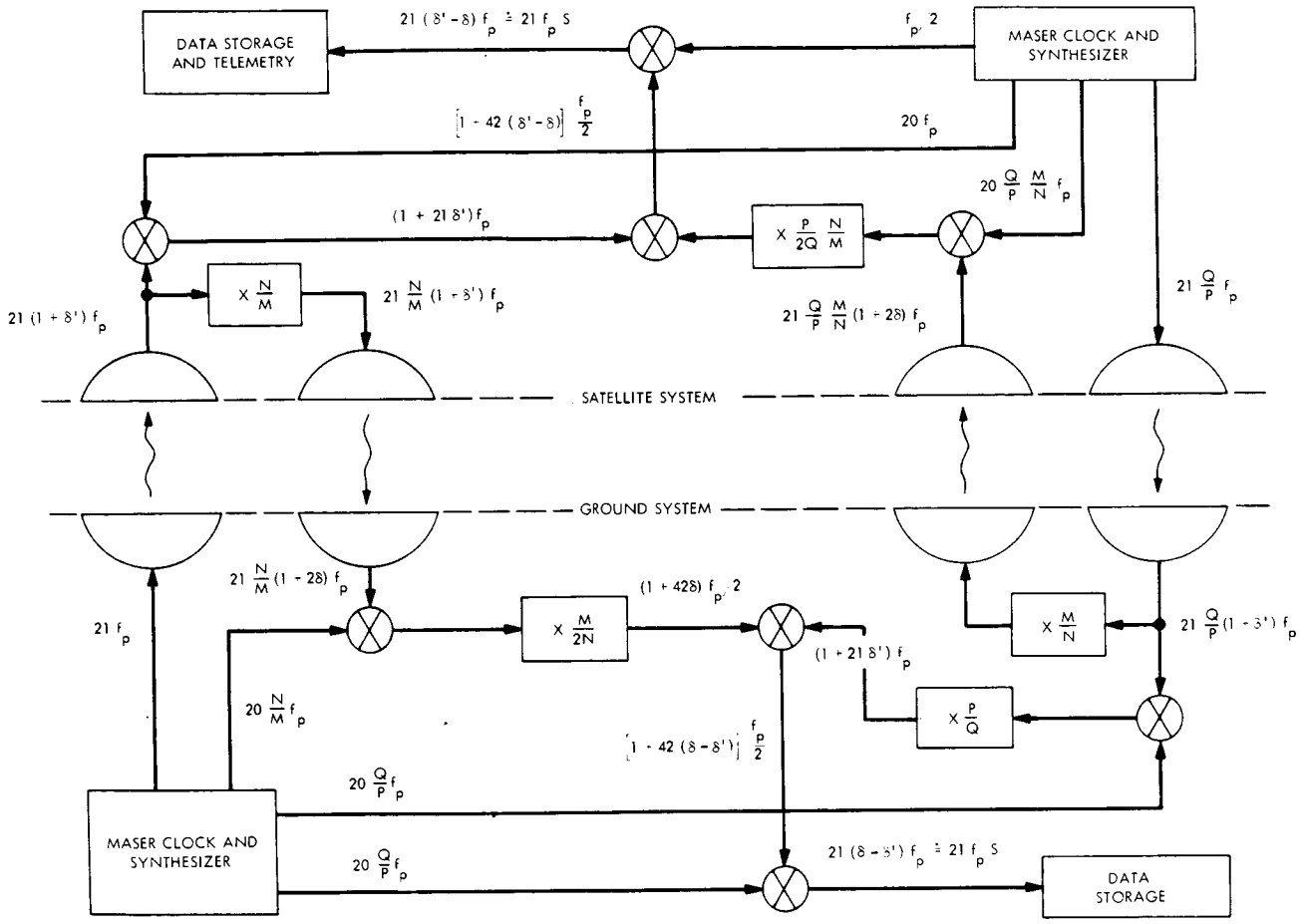


Fig. 6. Schematic block diagram of a symmetrical doppler-canceling telemetry system (Data are taken both in the satellite and on the ground.)

propagation path to and from the satellite is the same, Eqs. (3) and (4) will give us an expression for S. For this case, where $\overrightarrow{\epsilon_{23}} = -\overrightarrow{\epsilon_{12}}$ and $\overrightarrow{\beta_1} = \overrightarrow{\beta_3}$ and δ_s and δ_e are defined as $\overrightarrow{\beta_2} \cdot \overrightarrow{\epsilon_{12}}$ and $\overrightarrow{\beta_1} \cdot \overrightarrow{\epsilon_{12}}$, respectively, we obtain

$$\frac{\Delta f_{out}}{f_0} = \frac{f_0'}{f_0} - 1 - \frac{1}{2}\left(\frac{f_0''}{f_0} - 1\right) = S\left[1 - (\delta_s - \delta_e)\right]$$

$$+ \delta_e(\delta_s - \delta_e) \times \left[1 - (\delta_s - \delta_e)\right] \qquad (5)$$

The first term is the shift S multiplied by a small doppler correction; the second is a residual doppler effect resulting from the ground-station velocity.

The effects of changes in radial velocity of the satellite and the change in position of the ground station due to the earth's rotation depend on the choice of orbit and the location of the ground station. The correction terms that result from these effects can be computed from knowledge of the orbit and the station location.

The telemetry system described above cancels the first-order doppler by dividing the frequency shift in the tracking (go-return) signal by two and subtracts this average shift from the frequency of the one-way clock signal. Since asymmetrical propagation can occur in the up-down paths, the doppler correction of the frequency of the clock signal may be in error. To avoid this problem, it is very desirable to operate the telemetry system in a symmetrical manner, taking the redshift data both on the ground and in the satellite. This is accomplished by the addition of a transponder on the ground that coherently retransmits the clock signal from the satellite. In this way, the satellite can obtain information for doppler cancellation and correct the received frequency of the ground-station signal. This signal is controlled by the ground-based clock. The symmetrical system is shown schematically in Fig. 6. All frequencies are derived from the "proper" frequency f_p, which in this illustration is taken as 100 MHz. Doppler shifts are denoted by δ and relativistic-plus-doppler shifts by δ'.

The up-link to the satellite transponder is at 2100 MHz, and the down-link is at 2100 × N/M = 2280.5 MHz (N = 240, M = 221). The satellite clock signal down-link is at 2100 Q/P = 2290 MHz (Q = 241, P = 221), and the ground-station transponder frequency is at 2100 (Q/P) (M/N) = 2108.7 MHz. These carrier frequencies should be sufficiently close together to avoid dispersion problems in the atmosphere and ionosphere.

This system will allow continuous tracking of the maser clocks by a continuous monitoring of the phase of the two signals transmitted from each of the ground and satellite clocks. Data will be recorded both in the satellite and on the ground and will be compared to the known orbital parameters of the satellite. To make a redshift comparison compatible with the clock stability of 7 parts in 10^{15}, the following requirements on orbit information must be met:

Radial distance	$\Delta r \sim 240$ m
GM	known to 1 part in 10^5
Vector velocity of satellite	$\Delta v \sim 60$ cm/s
Radial acceleration	$\ddot{r} \sim 10^{-4}$ m/s^2

Note that much of the tracking information is available from the doppler measurements made by the system itself and that the limitations to the accuracy of the redshift determination are not likely to result from the tracking system but will depend on the stability of the clock.

III. Corrections to the Redshift

The data obtained will contain contributions from other gravitational effects. These have been estimated in Ref. 10. The significant ones reported here apply to the elliptic orbit described earlier, with the ground station at the equator:

(1)	Quadrupole moment of the earth	
	(a) Correction to the ground-station redshift	$= 3.76 \times 10^{-13}$
	(b) Correction to the satellite redshift at the perigee of the $\epsilon = 0.52$ orbit	$= 9.0 \times 10^{-14}$
(2)	Gravitational effect of the sun	
	(a) Correction to the ground-station redshift	$= 1.81 \times 10^{-17}$
	(b) Correction to the satellite redshift	$= 1.83 \times 10^{-15}$
(3)	Gravitational effect of the moon	
	(a) Correction to the ground-station redshift	$= 3.91 \times 10^{-17}$
	(b) Correction to the satellite redshift	$= 3.28 \times 10^{-15}$
(4)	Effect of earth tides	
	Correction to the ground-station redshift	$= 1 \times 10^{-15}$

Corrections for these effects can be computed from information relating to the geopotential of the earth and earth-body tides. This information is available to far greater accuracy than we need in Gaposchkin and Lambeck (Ref. 11) and Köhnlein (Ref. 12).

IV. Hydrogen-Maser Clock System

The atomic hydrogen maser has been under continuous development since its invention in 1960 by Goldenberg et al. (Ref. 13) and Kleppner et al. (Ref. 14). At present, when used to control the phase of a crystal oscillator, it provides the highest stability of any known system. The characteristics of this system are compared to typical characteristics of other oscillators in Fig. 7 (Vessot, Ref. 15). Note that statistical descriptions of the Rb-gas-cell-controlled crystal oscillators and of the crystal oscillator have been obtained by removing linear drift of frequency.

The high stability of the maser results from the storage-bulb technique that is used to confine ground-state hydrogen atoms in the upper hyperfine level, allowing them to interact with RF resonance radiation within a cavity for intervals of 1 s or longer. During this time, the atoms are coherently stimulated and deliver their energy to the cavity, sustaining the level of RF field if a sufficient flux of atoms enters the bulb.

The device is a self-oscillator at 1.4 GHz with a linewidth given by πT_T^{-1}, where T_T is total effective storage time of the bulb. Normally, the oscillator Q is in the order of 10^9.

Figure 8 is a schematic diagram of the maser and shows the energy levels of atomic hydrogen as a function of magnetic field strength. A cutaway view of the compact maser for space applications is shown in Fig. 9. In brief, the maser operates as follows. Molecular hydrogen is fed to an RF discharge dissociator, and the dissociated molecules are collimated into a beam of hydrogen atoms directed into a vacuum system along the axis of a hexapole magnet. Atoms in the $F = 1$, $M_F = 0$, and 1 states are focused into the storage bulb through a small, well-collimated hole. Atoms

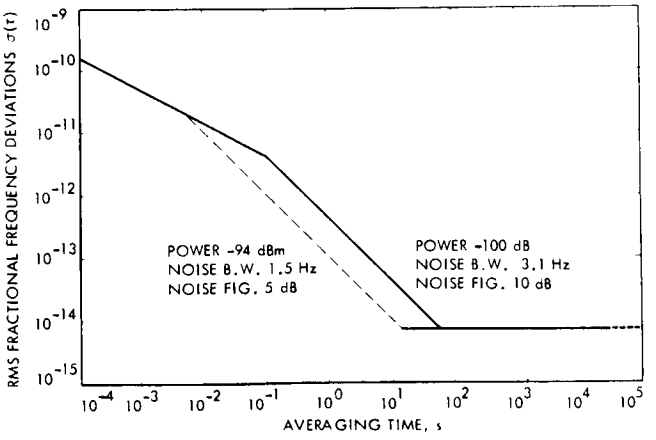


Fig. 7. Stability of hydrogen maser as a function of averaging time

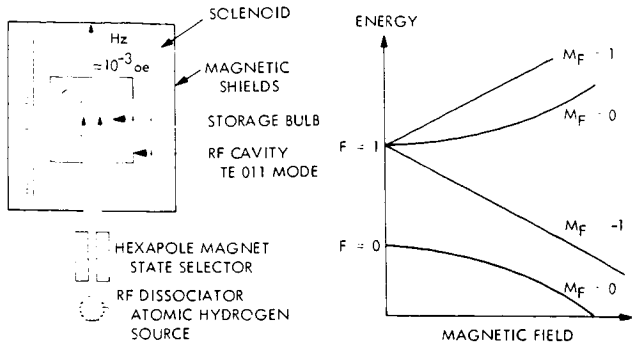


Fig. 8. Schematic diagram of maser and ground-state energy levels of atomic hydrogen

in the other two states are deflected away from the beam axis. The storage bulb confines the atom in an unperturbed way in the in-phase region of the RF magnetic field of a circular E-mode resonant cavity. Upper state, $M_F = 0$, atoms are stimulated to radiate their energy during their stay in the bulb and leave the bulb eventually to be scavenged by the ion pump. When the power available from the atoms in the bulb exceeds the cavity losses and the power coupled out, the maser oscillates.

Normally the output power is about -97 dBm and the signal is used to control the phase of a crystal oscillator. In this way, signal outputs at useful frequencies can be generated. More detailed descriptions of the maser and its associated electronics are available in the listed references.

The stability of the maser is limited by thermal noise whose frequency components lie within the linewidth of the oscillator and by additive noise that competes with the output signal in the bandwidth of the phase-lock system. The expression relating the maser stability to the noise within the linewidth of the oscillator and within the bandwidth of the receiver is given by Cutler and Searle (Ref. 16):

$$\sigma(\tau) = \left[\frac{kT}{2P}\left(\frac{F\omega_B Q_e}{\omega_0^2 \tau^2 Q_c} + \frac{1}{Q_c^2 \tau}\right)\right]^{1/2} \qquad (6)$$

Here, $\sigma(\tau)$ is the fractional rms frequency deviation averaged over a time interval τ, kT is the thermal noise power per unit bandwidth, ω_B is the receiver half-bandwidth (single tuned bandpass), P is the power delivered to the cavity by the atoms, Q_e is the atomic line Q, and Q_c and Q_e are the loaded cavity and external cavity Q, respectively.

The ultimate limit to the maser stability for long time intervals ($\tau > 10^4$ s) is imposed by variations in the resonance frequency of the cavity that are usually of thermal origin. The cavity mistuning "pulls" the output frequency by the ratio of the cavity Q to the line Q times the amount of the mistuning. Normally this ratio is about 10^{-5}, so a 10 Hz change in the cavity causes a 10^{-4} Hz change in the output (or 7 parts in 10^{14}). The

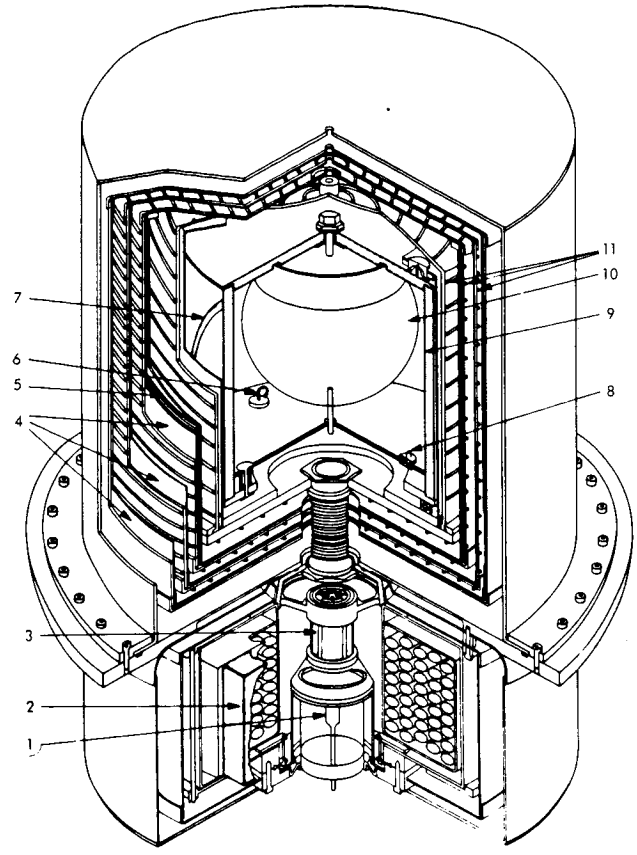


Fig. 9. Cutaway view of NASA maser

cavity frequency variations have been found to have a f^{-1} spectral density, evident from the flattening out of the $\sigma(\tau)$ plot of Fig. 7 (Ref. 17).

It is possible to improve long-term stability at the expense of short-term stability by increasing the storage time of the bulb. This is accompanied by a reduction in the output power from the maser and a corresponding worsening in the short-term stability that goes as $P^{-1/2}$. The tradeoff of short- versus long-term stability is shown in Fig. 10 for several values of the total bulb relaxation rate γ, which is the reciprocal of the effective bulb-storage time.

Several processes limit the storage time in the maser bulb (see References). Recent studies at Harvard and at the Smithsonian Astrophysical Observatory (SAO) on specially prepared polytetrafluorethylene (PTFE-Teflon) surfaces inside the storage bulbs have shown that 10^5 or more collisions with the bulb surface can occur before the atom loses phase with the RF field in the cavity. With bulb sizes now in use, this means that storage times approaching 10 s should be possible; and with the usual type of cavity and thermal controls, values of the rms frequency stability will approach 1 part in 10^{15} for averaging times of 10^3 s.

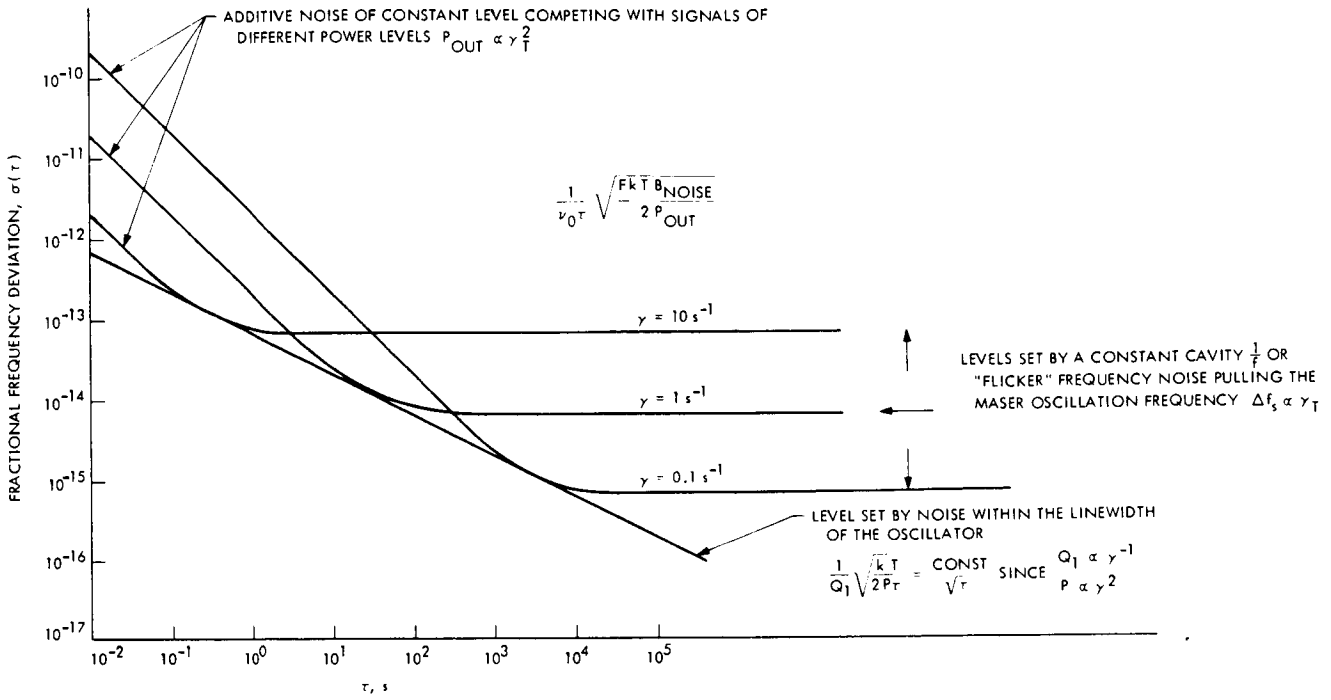


Fig. 10. Maser stability showing influence of total bulb relaxation rate at long and short averaging times

Experiments to verify this expectation are in progress at SAO.

The accuracy of the output frequency from the hydrogen maser is limited by the ability to determine (1) the magnetic field in the bulb, (2) the second-order doppler shift of the atoms in the bulb (which is related to the bulb temperature), (3) the amount of systematic residual mistuning, and (4) the systematic average phase shift per collision of the stored atoms with the walls of the bulb (wall shift). These effects are discussed in detail in Refs. 18, 19, and 20 and will be reviewed only briefly here.

A. Magnetic Field

The average field in the bulb is determined by observing the quenching of the maser output when a resonance transition is caused between the F = 1 magnetic sublevels. These transitions have a frequency dependence of 1.4 MHz/gauss, and using this resonance, the field can be measured to better than $\pm 3 \times 10^{-6}$ gauss. Since the field dependence of the $0 \rightarrow 0$ transition is described by $\Delta\nu_M = 2750 \langle H^2 \rangle_b$, the mean square value of the field over the bulb volume is required. Because gradients in the magnetic field can be observed through the effect of magnetic quenching of the output signal, we can trim the magnetic field, using separate coils in the solenoid so that $\langle H \rangle_b^2 - \langle H^2 \rangle_b$ is very small. An upper bond on the estimate of inaccuracy due to this is a few parts in 10^{14}.

B. Second-Order Doppler

The second-order doppler frequency shift is due to the thermal motion of the atoms in the storage bulb. Since each atom in the bulb makes a large number of collisions, the atoms are in thermal equilibrium with the bulb. The expression describing the shift is

$$\Delta\nu_T = -\nu_0 \frac{3kT}{2mc^2} = -1.9557\ T$$

where ν_0 = 1.4204 GHz, k is Boltzmann's constant, T is in kelvins, c is the velocity of light, and m is the mass of the hydrogen atom. Measurement of the bulb temperature to an accuracy of ±0.25°K at 320°K allows a determination of $\Delta\nu T/\nu_0$ to about 3 parts in 10^{14}.

C. Cavity Mistuning

Cavity mistuning, as described earlier, will "pull" the output frequency, and a systematically mistuned cavity can be a source of inaccuracy in the output frequency of the maser. A part of the NASA contract that lead to the development of the maser shown in Fig. 9 consisted of the invention and development of an automatic cavity tuning servo system.

The exact expression for the effect of cavity mistuning (Ref. 21) is

$$\Delta\nu_s = \left(\frac{\Delta\nu_{cavity}}{\nu_0} Q_c - BT^{1/2}\right)\frac{1}{\pi}\nu_T \qquad (7)$$

where B is a constant that depends on the bulb and the cavity and on the cavity-bulb geometry, T is

in kelvins, and γ_T is the total relaxation rate of the atoms in the bulb.

The $BT^{1/2}$ term results from shifts in frequency due to hydrogen-hydrogen collisions. This effect is described by Bender (Ref. 22). Crampton et al. (Ref. 21) have shown that the shift is proportional to the linewidth, and by use of the spin-exchange quenching tuning method described below, the shift is eliminated.

Spin-exchange collisions occur among the atoms in the storage bulb and produce pressure-dependent relaxation among the energy levels in the hydrogen atom. This relaxation process, negligible under normal operating conditions, can be used to modulate the total relaxation rate γ_T. If the quantity in parentheses in Eq. (7) is not zero, there will be a modulation in $\Delta\nu_s$ whose magnitude and phase are in proportion to the magnitude and sign of the bracketed quantity. The automatic cavity tuning servo consists of a synchronous detection system to observe the variations in the output frequency due to $\Delta\nu_s$ and to correct the cavity resonance frequency so that $\Delta\nu_s$ is nulled.

The cavity itself is made of CER-VIT and is compensated, by use of re-entrant posts of an alloy with a high-expansion coefficient, to remove the variations in dielectric loading of the fused silica bulb due to the temperature dependence of its dielectric constant. The temperature coefficient of the cavity resonance frequency has been reduced to 100 Hz/°C. With thermal control within limits of $\pm 1 \times 10^{-2}$°C, the resultant output-frequency variations due to the cavity can be maintained below 1 part in 10^{14}.

The fractional frequency stability between two masers operating with automatic tuning systems has been measured to be 2×10^{-14} for averaging times of 4×10^4 (1/2 day) (Ref. 23). Thus, for a single maser, the stability can be estimated to be $2/\sqrt{2} \times 10^{-14}$, or 7 parts in 10^{15} for the above averaging time.

The accuracy of the automatic-tuning technique depends on the type of tuning reference oscillator used and on the time allowed for the servo to operate. When a second hydrogen maser is employed as a tuning reference, less than 1 h is necessary for an accuracy of about 1 part in 10^{13} to be accomplished. Since no systematic effects on the accuracy of the maser output frequency have thus far been attributable to the tuning procedure, the accuracy of the tuning should approach the stability of the tuning, although this remains to be proved experimentally.

D. Wallshift

The systematic average phase shift per wall collision has been the chief limitation in the intrinsic accuracy of the hydrogen maser. Variations in the wallshift determined at several laboratories have led to differences of some 5 parts in 10^{12} in the maser output frequency. Since the wallshift depends on the collision rate of the atoms in the bulb, the texture of the Teflon wall surface is an important consideration in a determination of the surface-to-volume ratio of the bulb. It is difficult to reproduce the same texture of the Teflon from one bulb to another. During the last year, there has been considerable activity investigating the properties of Teflon surfaces. Zitzewitz at Harvard (Ref. 24) found that the phase shift per collision for FEP Teflon* could be made to cross through zero from negative to positive at about 110°C, a function of increasing wall-coating temperature. Vessot and Levine (Ref. 25) at SAO found that for high-molecular-weight PTFE surfaces sintered and rapidly quenched, a zero wallshift temperature of 83°C could be obtained that was independent of the surface-to-volume ratio of the bulb. Brenner (Ref. 26) suggested using a deformable bulb to vary, by a known amount, the surface-to-volume ratio of the bulb, thus allowing a determination of the wall shift of the coating in a particular bulb. This method was successfully demonstrated by Debely (Ref. 27), using a cylindrical bulb whose conical end, made of thin flexible Teflon and stretched inside or outside the cylinder, would determine two easily calculable bulb volumes with the same surface.

Since the zero wallshift temperature is independent of the collision rate, knowledge of the collision rate is no longer required if the correct temperature can be determined. Recently, we suggested that the flexible bulb also be used to vary the collision rate in order to determine the temperature at which there is no collision-rate dependence on the wallshift. We have described a servosystem that accomplishes this automatically (Vessot and Levine, Ref. 28).

Recent measurements by SAO and the National Bureau of Standards of PTFE surfaces applied in a reasonably standardized way have given agreement to 5 parts in 10^{13} in the maser frequency (Ref. 29).

E. Summary of Present Maser Capability

The intrinsic accuracy of the maser can be estimated from the rms of the several accuracy limitations:

$$\Delta\nu_{temp}/\nu_0 \simeq 3 \times 10^{-14}$$

$$\Delta\nu_{mag}/\nu_0 \simeq 3 \times 10^{-14}$$

$$\Delta\nu_{cav}/\nu_0 \simeq 1 \times 10^{-13}$$

$$\Delta\nu_{wall}/\nu_0 \simeq 5 \times 10^{-13}$$

$$\text{Accuracy } \Delta\nu_{rms}/\nu_0 \simeq 5.1 \times 10^{-13}$$

The precision of resettability can be estimated from the rms of the parameters that vary with time:

$$\Delta\nu_{temp}/\nu_0 \simeq 3 \times 10^{-14}$$

$$\Delta\nu_{mag}/\nu_0 \simeq 3 \times 10^{-14}$$

$$\Delta\nu_{cav}/\nu_0 \simeq 1 \times 10^{-13}$$

$$\text{Resettability precision } \Delta\nu_{rms}/\nu_0 \simeq 1.1 \times 10^{-13}$$

*Hexafluoropropylene-polytetrafluoroethylene copolymer.

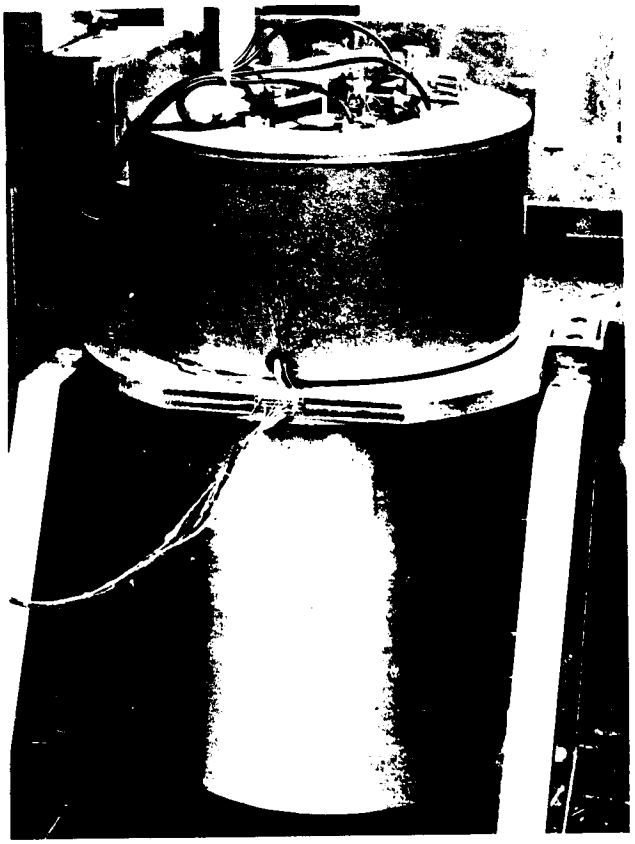
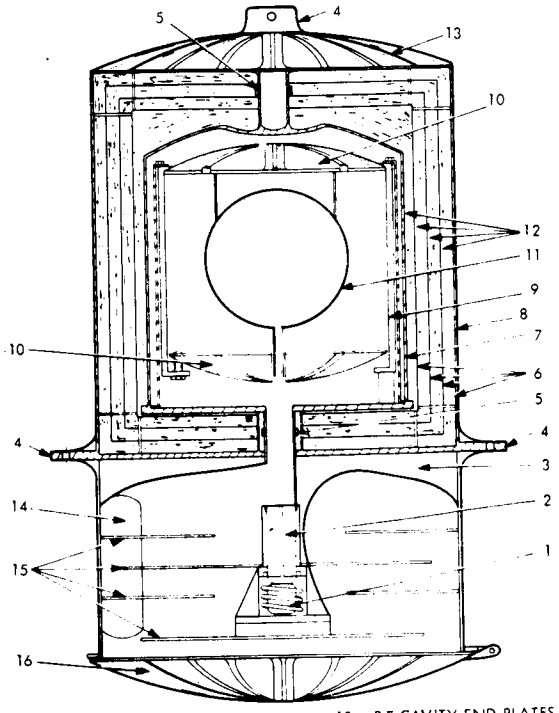


Fig. 11. Photograph of NASA maser developed for satellite use



1 RF DISCHARGE H_2 DISSOCIATOR
2 HEXAPOLE STATE SELECTOR MAGNET
3 SUPPORTING WEB (ONE OF THREE)
4 MOUNTING LUGS
5 SUPPORTING TUBE WITH HEAT STATION
6 MAGNETIC SHIELDS
7 VACUUM ENCLOSURE
8 OUTER CAN
9 RF CAVITY CYLINDER
10 RF CAVITY END PLATES
11 MASER BULB
12 MULTILAYER REFLECTIVE INSULATION
13 OUTER CAN REINFORCEMENT
14 HYDROGEN CYLINDER AND PRESSURE CONTROL
15 BAFFLES
16 REMOVABLE COVER OUTSIDE OF SATELLITE

Fig. 12. Lightweight maser based on existing design concepts

The fractional frequency stability of the maser is 7×10^{-15} for averaging time intervals from 100 to 4×10^4 s.

We see that for an unmodulated gravitation experiment, for example a circular orbit, the accuracy of the measurement will depend on the magnitude of the shift (see Fig. 1) and the precision of resettability of the clock. Prior to launch, the satellite clock and ground clock would have been compared. For those clocks not included in this preflight calibration the experiment accuracy is limited by the accuracy of the clocks.

The advantage of using an eccentric orbit results from the excellent stability of the maser and the fact that the measurement can be repeated many times, as long as the satellite is operating. We must, however, be certain that there are no systematic frequency shifts that are synchronized with the orbital period.

V. Maser Configurations, Weight, and Power Estimates

Since 1960, a continuous development of hydrogen masers has been under way with NASA's support. In 1966, NASA supported the development of a small, relatively lightweight maser system for use in spacecraft. The maser oscillator of this system is shown in Fig. 11; the automatic tuner system, phase-lock synthesizer, thermal controls, pressure controls, and RF dissociator have been successfully operated in breadboard form. The cavity and bulb assembly was vibration tested in the sine sweep mode; it successfully survived such tests in all three axes. The complete structure shown in Fig. 11 weighs about 112 kg and requires about 30 W to operate; it is about 81 cm high and 48 cm in diameter.

The vacuum of space can profitably be used in place of the ion pump to scavenge the expended hydrogen. A considerable reduction in weight and power will result from use of multilayer, reflecting insulation under vacuum to replace the microballoon foam insulation currently used to surround the bell jar and ovens. The total weight of the maser can be reduced to about 34 kg, and its power to about 10 W. A sketch of this lightweight maser is shown in Fig. 12. The design concepts of this maser are similar to those of the self-contained unit. Some changes in operation are required, however. Before launch, the maser will be operated by use of a titanium sublimation pump located inside the hatch cover. This system will enable ground comparisons and preflight testing to be conducted.

During launch, the maser will be shut down; later, in orbit, it will be restarted and the cavity

retuned. Several hours of operation will be possible before the titanium evaporated just before launch is used up, and this will allow preliminary system checks. The hatch will be opened to the void of space as soon as the satellite outgassing in the direction of the pump reaches an acceptably low level.

Contamination streaming back into the maser will be kept to a low level by baffles, and the contamination entering the bulb, if it has not been adsorbed on the metal baffles, is very unlikely to attach itself to the Teflon-coated interior of the bulb, which will be maintained at 83°C. In view of the even slight possibility of wallshift changes due to contamination, a test of the continued accuracy of the maser can be made by comparing the average frequency over several orbits early in the experiment with the average frequency at a later time. The stability of the maser, over 4×10^4 s averaging time intervals, would be negligibly affected by bulb-contamination processes that slowly and monotonically could shift the maser output frequency. The data from redshift modulation will continue to be available.

Other, more compact, maser configurations are possible with the use of dielectric or capacitive loading of the RF cavity to reduce its size. Experiments are in progress at SAO* on a spherical, fused-silica, dielectrically loaded cavity 20.3 cm in diameter with an inside diameter of 12.7 cm. The features of spherical geometry and the integral bulb make a very rugged, compact unit. The original proposal to NASA for a satellite-borne maser (Ref. 30) included this cavity. Development and procurement of this cavity proceeded until 1966, when the emphasis on light-weight and low-power consumption was changed to that of pressing for the greatest possible precision of resettability and stability. The cavity was taken over by the USAECOM and successfully used in a maser (Ref. 31). The recent improvements on wall coatings and the desire for a compact lightweight device suitable for a wide range of vehicles bring this design back into prominence. Figure 13 shows the maser with a spherical cavity.

VI. Conclusion

The proposed gravitational redshift experiment described in this paper makes use of technology currently available. Since 1964, when the original proposal was submitted to NASA, the hydrogen-maser clock has been under continuous development for its use in spacecraft, and the feasibility of a lightweight, low-power device has been demonstrated.

Plans for the experiment have evolved considerably. In 1968, the George C. Marshall Space Flight Center completed a Preliminary Program-Development Plan for a Hydrogen-Maser Relativity Satellite, which envisioned a 2700-lb (approximately 1000 kg) satellite containing two masers requiring a total of 1200 W. This satellite would be launched by a Titan IIIC booster into a synchronous, circular orbit. Its objective would be the measurement of the redshift with an accuracy of about 500 parts per million.

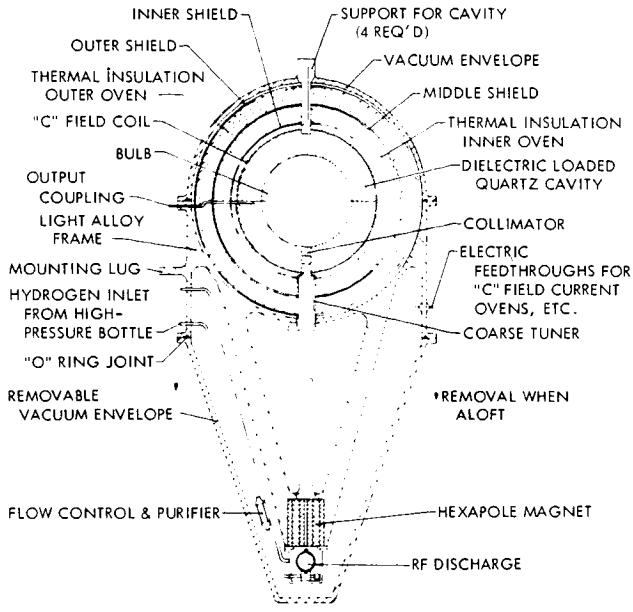


Fig. 13. Hydrogen maser using a dielectrically loaded spherical cavity

Since 1968, the precision and stability of the maser clock have been substantially improved, largely as a result of continuing NASA support for maser development. Also, the use of an eccentric orbit, where the satellite scans back and forth through large differences in the earth's gravitational potential, has made it possible to utilize the maser's excellent frequency stability to good advantage. The present concept of the satellite clock experiment will give much better accuracy at substantially lower cost. A total satellite weight of 500 lb (187 kg), consisting of the maser, its electronics, a transmitter-transponder, batteries, and solar panels, is now possible. The thrust-augmented Thor-Delta vehicle can launch this payload into an eccentric orbit.

The previously mentioned limits to eccentricity, imposed by the desire to keep the satellite constantly in sight of the ground station, can be relaxed, thus giving a larger gravitational modulation. Since a satellite with a 24-h period in a highly eccentric orbit cannot be observed both at perigee and at apogee from a single ground station, a different period should be used, allowing a ground station to track all phases of the orbit over a period of time, though not continuously. A further advantage to this highly eccentric orbit is that more than one ground station can independently participate in the experiment. The limit on eccentricity is probably imposed by the requirement that the satellite be in sight of a particular station long enough to allow 1 h or longer of phase tracking of the doppler-corrected carrier. At the 2.5-GHz carrier frequency, assuming about 10° of phase resolution, the system should allow each measurement to be made to the limits of the clock stability. At present, there are several tracking stations already equipped with hydrogen masers that, with a modest amount of additional equipment, can be used to track the satellite maser signals. Studies

*Supported in part by contract DAAB07-70-C-A-A108 from the U.S. Army Electronics Command, Ft. Monmouth, New Jersey.

on weight, power, orbits, and booster capability are continuing.

We believe that an accurate test of the equivalence principle for clocks in space is a necessary phase of the continuing space program and that this test, successfully performed, will not only advance our knowledge of science but also open new fields of technology that will be of value to our exploration of the universe.

References

1. Einstein, A., Jahrb. Radioakt. u. Elektronik, 4, 411, 1907.
2. Roll, P. G., Krotkov, R., and Dicke, R. H., Ann. Phys., 26, 442, 1964.
3. Ramsey, N. F., Molecular Beams, Oxford Univ. Press, 1956.
4. Winterberg, F., Astronautica Acta, 2, 25, 1956.
5. Singer, S. F., Phys. Rev., 104, 11, 1956.
6. Hoffman, B., Phys. Rev., 106, 358, 1957.
7. Refsdal, S., Phys. Rev., 124, 996, 1961.
8. Pound, R. V., and Rebka, G. A., Jr., Phys. Rev. Letters, 4, 337, 1960.
9. Pound, R. V., and Snider, J. L., Phys. Rev., 140, B788, 1965.
10. Kleppner, D., Vessot, R. F. C., and Ramsey, N. F., Astrophys. Space Sci., 6, 13, 1970.
11. Gaposchkin, E. M., and Lambeck, K., Smithsonian Astrophys. Obs. Spec. Rep. No. 315, 1970.
12. Köhnlein, W., Smithsonian Astrophys. Obs. Spec. Rep. No. 198, 1966.
13. Kleppner, D., Goldenberg, H. M., and Ramsey, N. F., Phys. Rev., 126, 603, 1962.
14. Goldenberg, H. M., Kleppner, D., and Ramsey, N. F., Phys. Rev. Letters, 5, 361, 1960.
15. Vessot, R. F. C., Hewlett-Packard Journal, p. 15, Oct. 1968.
16. Cutler, L. S., and Searle, C. L., Proc. IEEE, 54, 136, 1966.
17. Barnes, J. A., et al., NBS Tech. Note 394, Washington, D. C., 1970.
18. Vessot, R. F. C., and Levine, M. W., "Proceedings of 24th Annual Symposium on Frequency Control," U.S. Army Electronics Command, Ft. Monmouth, N.J., Apr. 27-29, 1970.
19. Vessot, R. F. C., et al., IEEE Trans. Instr. Meas., IM-15, 165, 1966.
20. Vessot, R. F. C., et al., "Proceedings of International Conference on Precision Measurement and Fundamental Constants," National Bureau of Standards, Gaithersberg, Md., Aug. 3-7, 1970.
21. Crampton, S. B., Kleppner, D., and Ramsey, N. F., Phys. Rev. Letters, 11, 338, 1963.
22. Bender, P. L., Phys. Rev., 132, 2154, 1964.
23. Risley, A. S., et al., "Conference on Precision Electromagnetic Measurements," National Bureau of Standards, Boulder, Colo., June 1970.
24. Zitzewitz, P. W., Uzgiris, E. E., and Ramsey, N. F., Rev. Sci. Instr., 41, 81, 1970.
25. Vessot, R. F. C., and Levine, M. W., Meteorologia, (in press).
26. Brenner, D., Bull. Amer. Phys. Soc., 14, 943, 1969.
27. Debely, P., Rev. Sci. Instr., 41, 1290, 1970.
28. Vessot, R. F. C., and Levine, M. W., NASA Disclosure HQN-10654, 1970.
29. Hellwig, H., Vessot, R., Levine, M., Zitzewitz, P., Allan, D., and Glaze, D., IEEE Trans. Instr. Meas., IM-19, Nov. 1970.
30. Vessot, R. F. C., Kleppner, D., and Ramsey, N. F., "Atomic Hydrogen Maser for Space Vehicle Applications," Varian Associates Proposal BP 812-64-09, Oct. 30, 1964.
31. Hellwig, H., and Pannaci, E., J. Appl. Phys., 39, No. 12, 5496, 1968.

Tests of General Relativity Using Pulsars

Paul E. Reichley
Jet Propulsion Laboratory
California Institute of Technology

I. Introduction

Several pulsars are being observed on a regular basis at the Jet Propulsion Laboratory. The arrival times of the pulses from each pulsar are measured by a cesium clock, which in turn is compared with clock 8 of the National Bureau of Standards. The observations are all made at a frequency of 2388 MHz (12.5 cm wavelength) on a 26-meter dish antenna at the Goldstone Tracking Station. The method used to measure the pulse arrival times (Ref. 1) and the antenna and receiver used in the measurements (Ref. 2) have been described elsewhere.

The arrival times of the pulses in the inertial frame of the solar system are influenced by several effects. The motion of the clock about the barycenter of the solar system causes the phase of the received pulse train to vary. The absolute velocity of the clock in the barycentric inertial frame and the clock's distance from the gravitational masses in the solar system both vary, causing the clock's rate to vary with respect to the coordinate time rate (Refs. 3 and 4). This difference in rate appears as a variation in the phase of the measured pulse train. The sun's gravitational mass causes a delay of the pulses as they pass through the solar system (Refs. 5 and 6). This delay causes the phase of the received pulse train to vary. Because of the high frequency of observation, the effect of charged particles within the solar system can be neglected for the pulsars observed. The effect of interstellar charged particles seems to be a random one that simply increases the noise level on the arrival time measurements.

We are concerned here with the variation in rate of the clock and the delay of the pulses. The variation in clock rate consists of two effects: the time dilation effect of special relativity and the red-shift effect of general relativity. Both of these effects have been verified using earth-based experiments (Refs. 7 and 8). Because of the mechanics of the experiment, as will be explained in Section II, the two effects are alike and inseparable, and appear as a red-shift effect only. We hope to improve on the present 1% measurement of the red shift (Ref. 9). The delay effect is at present beyond our measurement capability for the pulsars under observation. We hope to make a measurement of the delay effect eventually, using techniques to be discussed later, but we do not expect to be able to achieve the accuracy obtainable with radar ranging.

II. Experimental Analysis

In order to measure changes in the rate of our cesium clock due to solar system gravitational effects, we need a clock free from these effects for comparison. The "ticks" or pulses from the pulsars under observation furnish just such a set of clocks. In order to make use of these pulsar clocks, we must be able to determine the rate at which they tick. Hence, we must eliminate the non-relativistic effects and determine the rate at which the pulsar clocks pulse from our measurements of the arrival times, which furnish the desired comparison.

The non-relativistic effects on the arrival times of the pulses are caused by the cesium clock's movement about the solar system

barycenter. We eliminate these effects by reducing the arrival times to the center of an inertial reference frame centered at the barycenter of the solar system. We use the right ascension-declination system of 1950.0 as the inertial reference frame. The position of the earth with respect to the barycenter is computed by means of Jet Propulsion Laboratory Ephemeris DE-69 (Ref. 10). The position of the cesium clock with respect to the geocenter, the reduction to the epoch 1950.0, and the conversion from UTC to ET are computed by means of standard equations using parameters developed at JPL (Ref. 11).

The reduction of the pulse arrival times to the barycenter requires knowledge of the pulsars' positions. Since the positions are not known accurately enough a priori, we must solve for them. The arrival times of the pulses at the barycenter depend on the periods (pulse repetition intervals) of the pulses. We must also solve for the period and its time derivatives, called the period characteristics. Hence, from the arrival time measurements, we solve for the pulsars' positions and period characteristics as well as the relativistic effects we seek. The solutions are accomplished by least-squares techniques and differential correction, and have been described (Ref. 1) for all unknowns except the relativistic effects.

The metric used in our calculations is the Robertson line element (Ref. 12):

$$ds^2 = \left(1 - \frac{2\alpha m}{r} + \frac{2\beta m^2}{r^2} + \ldots\right) dt^2$$

$$- \frac{1}{c^2}\left(1 + \frac{2\gamma m}{r} + \frac{3\delta m^2}{2r^2} + \ldots\right) d\sigma^2$$

where

$d\sigma^2 = dx^2 + dy^2 + dz^2$

c = speed of light

m = gravitational radius of the sun

$\alpha, \beta, \delta, \gamma$ = free parameters (= 1 for Einstein's theory)

$r = (x^2 + y^2 + z^2)^{1/2}$

We neglect all terms of order m^2/r^2 or higher as the effects they contribute are well below the accuracy of our arrival time measurements. The relativistic effects of the planets and the moon are neglected for the same reason. Since the earth moves around the sun, the sun's motion in the barycentric inertial frame causes a time dilation effect on the cesium clock. However, since the sun's velocity is so low, this effect can also be neglected.

As seen from the above discussion, the relativistic effects we can measure are due to the cesium clock's motion about the sun. Under this motion, within our measurement accuracy, the time dilation effect and the red-shift effect are inseparable. Because of the way in which the two effects combine, we are measuring the red shift effect only (or the time dilation only). The variation in the cesium clock's rate integrates over the span of the measurements and appears as a sinusoidal effect with an amplitude of 1.7α ms and a period of 1 year.

The delay effect appears logarithmically with a coefficient of $\alpha + \gamma$ and reaches a maximum once a year. The maximal effect of $(\alpha + \gamma)50$ μs occurs for a pulsar passing behind the sun, as seen from the earth.

III. Results

We have used the arrival time measurements from several pulsars to solve for the red-shift and delay effects. The delay effect was buried in the noise of the arrival time measurements in all cases, and we were unable to make a measurement of $\alpha + \gamma$. The red-shift effect was correlated almost 100% with position variations in all cases, but we were able to measure α.

The failure to measure $\alpha + \gamma$ was due to an unfortunate set of circumstances. The pulsars which passed close enough to the sun to have a significant effect were too weak to make sufficiently accurate arrival time measurements possible. The pulsars that were strong enough to yield accurate arrival time measurements did not pass close enough to the sun to have a significant effect.

We were able to make measurements of α in all cases, but because of the high correlations with position, we do not put much faith into the results. We obtained a mean value of 0.99 ± 0.03, where the error is one standard deviation, which agrees well with the Einstein value of 1. Our major reason for putting little faith into this result is that varying the value of α about the value of 1 and solving for the other parameters of the model did not change the sum of squares of the residuals in a statistically significant manner. On the positive side, however, the positions obtained by this technique with $\alpha = 1$ agreed well with the best interferometric results.

IV. Conclusions

We are presently writing a computer program that will combine data from several pulsars. Since the relativistic effects are common to the data, this technique will reduce the correlations of positions with the red-shift effect and should raise the delay effect above the noise level.

If the arrival time measurements could be made with uncertainties on the order of a microsecond, the door would be opened to a host of relativistic effects. Some of the more interesting effects would be the red shift of Jupiter, the advance of perihelion of the earth, and periodic variations in the earth-pulsar distance.

Arrival time uncertainties on the order of a microsecond (Ref. 13) have been obtained with the Crab Nebula pulsar at optical wavelengths. This was accomplished using fast sampling techniques and averaging. Unfortunately, the pulsar's period is not well behaved and creates noise of its own.

Our major obstacle at present is the inability to sample the pulses fast enough. Even with a sampling period of 50 μs, we are able to obtain 10-μs uncertainties on the stronger pulsars by averaging only 20 or 30 pulses. We will have the capability to sample at a period of 1 μs within a year. This could yield a 0.05% measurement of the red shift. The largest error source with fast sampling is the random noise on the pulse arrival times caused by interstellar charged particles.

Acknowledgment

We would like to thank Professor Kip Thorne for helpful discussions; Dr. George Downs and Mr. George Morris for help with the observations and data reduction; and the staff of the Venus Station, Goldstone Deep Space Communications Complex, for their help with the observations.

References

1. Reichley, P. E., Downs, G. S., and Morris, G. A., Ap. J. (Lett.), 159, L35, 1970.

2. Downs, G. S., Morris, G. A., and Reichley, P. E., Nature, 222, 1257, 1969.

3. Aoki, S., Astron. J., 69, 221, 1964.

4. Clemence, G. M., and Szebehely, V., Astron. J., 72, 1324, 1967.

5. Shapiro, I. I., Phys. Rev. Lett., 13, 789, 1964.

6. Muhleman, D. O., and Reichley, P. E., Space Programs Summary 37-29, Vol. IV, 239, Jet Propulsion Laboratory, Pasadena, Calif., 1964.

7. Pound, R. V., and Rebka, G. A., Phys. Rev. Lett., 4, 337, 1960.

8. Frisch, D. H., and Smith, J. H., Am. J. Phys., 31, 342, 1963.

9. Pound, R. V., and Snider, J. L., Phys. Rev., 140, B788, 1965.

10. O'Handley, D. A., Holdridge, D. B., Melbourne, W. G., and Mulholland, J. D., Technical Report 32-1465, Jet Propulsion Laboratory, Pasadena, Calif., 1969.

11. Melbourne, W. G., Mulholland, J. D., Sjogren, W. L., and Sturms, F. M., Technical Report 32-1306, Jet Propulsion Laboratory, Pasadena, Calif., 1968.

12. Robertson, H. P., Space Age Astronomy, 228, Edited by A. J. Deutsch and W. E. Klemperer, Academic Press, New York, 1962.

13. Nelson, J., Hills, R., Cudaback, D., and Wampler, J., Ap. J. (Lett.), 161, L235, 1970.

The Stanford Gyroscope Experiment*

C. W. F. Everitt
Stanford University

I. Description of Experiment

In 1960, L. I. Schiff pointed out that a gyroscope in motion about a massive body such as the earth might be expected to undergo a relativistic precession in the framework of the fixed stars (Ref. 1). Two main effects were predicted. The geodetic effect is due solely to the motion of the gyroscope about the earth and yields a drift rate Ω_G given by

$$\Omega_G = \frac{3GM}{2c^2R^3}(\underline{R} \wedge \underline{v}) \qquad (1)$$

The motional effect is due to the rotation of the earth itself and is

$$\Omega_M = \frac{GI}{c^2R^3}\left\{\frac{3\underline{R}}{R^2}(\underline{\omega} \cdot \underline{R}) - \underline{\omega}\right\} \qquad (2)$$

where $\underline{m}$, $\underline{R}$, and $\underline{v}$ are, respectively, the mass, coordinate, and velocity of the gyroscope, and M, I, and $\underline{\omega}$ are the mass, moment of inertia, and angular velocity of the earth.

An experiment to detect either effect would consist in measuring the precession rate of one or more gyroscopes, mounted either on earth or in a satellite, with respect to the axis of a telescope pointing at a suitable fixed star. In a satellite moving in a 500-mile circular polar orbit, the two effects are at right angles, the integrated geodetic rate being 6.9 arc-s/year and the motional rate 0.05 arc-s/year. There are also a number of smaller terms in the range 0.0001 to 0.01 arc-s/year, arising from the quadrupole mass moment of the earth, the rotation of the sun, and the orbital motion of the earth about the sun (Ref. 2).

Expressing the results in terms of the principal post-Newtonian formalism developed by Eddington, Schiff, and most recently, by Thorne and Will (Ref. 3), a measurement of Ω_G determines the geodetic parameter $(1 + 2\gamma)$ and of Ω_M the parameter Δ associated with the effects of rotating matter. In the Brans-Dicke scalar-tensor theory, the predicted value of Ω_G is about 6.3 arc-s/year. An experiment with accuracy of about 0.1 arc-s/year would thus provide a clear test between general relativity and the scalar-tensor theory, and would also check certain other theories which fall outside the PPN formalism (for example, the Birkhoff, Belifante-Swihart and Whitrow-Morduch theories), each of which gives a different prediction for Ω_G (Ref. 4).

However the most interesting goal for the experiment is a measurement of the motional rate Ω_M. To obtain such a measurement within 1 or $2^{0}/_{0}$, a gyroscope with residual errors from extraneous sources of less than 0.001 arc-s/year is required. This desired accuracy, which is equivalent to a drift rate of 1.6×10^{-16} rad/s, will be called Ω_0. It appears to be just within the capabilities of existing technology.

Analysis and experimental work on the gyroscope program have been supported at Stanford by the National Aeronautics and Space Administration since October 1963. The work is being carried on jointly by members of the Department of Physics and the Department of Aeronautics and Astronautics.

*Work supported in part by NASA Grant 05-020-019.

After prolonged thinking, we have concluded that almost the only way of reaching the the extremes of accuracy needed is by extensive combination of three areas of technology which have opened up during the past 15 years: (1) space research, (2) large-scale low-temperature physics research, and (3) certain advanced techniques in electronics instrumentation and in control theory. The work on the gyroscope experiment has both stimulated and benefited from other activities in all three fields. Elsewhere in this volume, W. M. Fairbank describes several possible future applications of low-temperature technology which interest us at Stanford, while D. B. DeBra discusses the drag-free satellite technology which forms part of the program of research on guidance and control currently being pursued in the Department of Aeronautics and Astronautics.

A complete laboratory prototype of the gyroscope experiment has been designed, and parts of it are now under test. We expect to make the first spinup and gyro drift measurements on earth during 1971. Plans for an engineering test flight as soon as possible thereafter are being formed in conjunction with a group at NASA Marshall Space Flight Center under the leadership of Dr. R. Decher.

A general view of the space experiment as presently conceived is shown in Fig. 1. It comprises a superinsulated dewar vessel filled with about 300 l of liquid helium, containing a single star-tracking telescope and four gyroscopes grouped in two pairs, spinning in opposite directions, to give a double check on each of the two relativity effects. Mechanical stability was ensured by making the entire gyro-telescope structure of fused quartz, optically contacted together and all at helium temperatures to eliminate distortion due to thermal expansion. The reference axis is a bright star near the equator, possibly Procyon. The telescope is a folded Schmidt-Cassegrain system of 150-in. focal length and 5.5-in. aperture, in which the light is divided by a beam splitter to give two star images. Each image falls on the sharp edge of a roof prism, where it is again subdivided and passed to a light chopper and photodetector at ambient temperature. In this way, a reference accuracy better than 0.001 arc-s is obtained even though the determination is far beyond the diffraction limit, since the location of the center of the star image is limited only by the sharpness of the knife edge and by photon counting statistics. Attitude control of the satellite is accomplished by gas jets supplied from the helium gas evaporating from the dewar. The pointing accuracy so obtained is about ±0.5 arc-s; fine pointing for the telescope is provided by an inner servo driven by cryogenic actuators. An internal proof mass, acting as a zero-g reference, is used to apply translation control to the gas jets. The use of a drag-free system helps in two ways. It improves the averaging of the residual acceleration acting on the gyroscopes and it reduces errors in the orbit determinations needed in analyzing the relativity data.

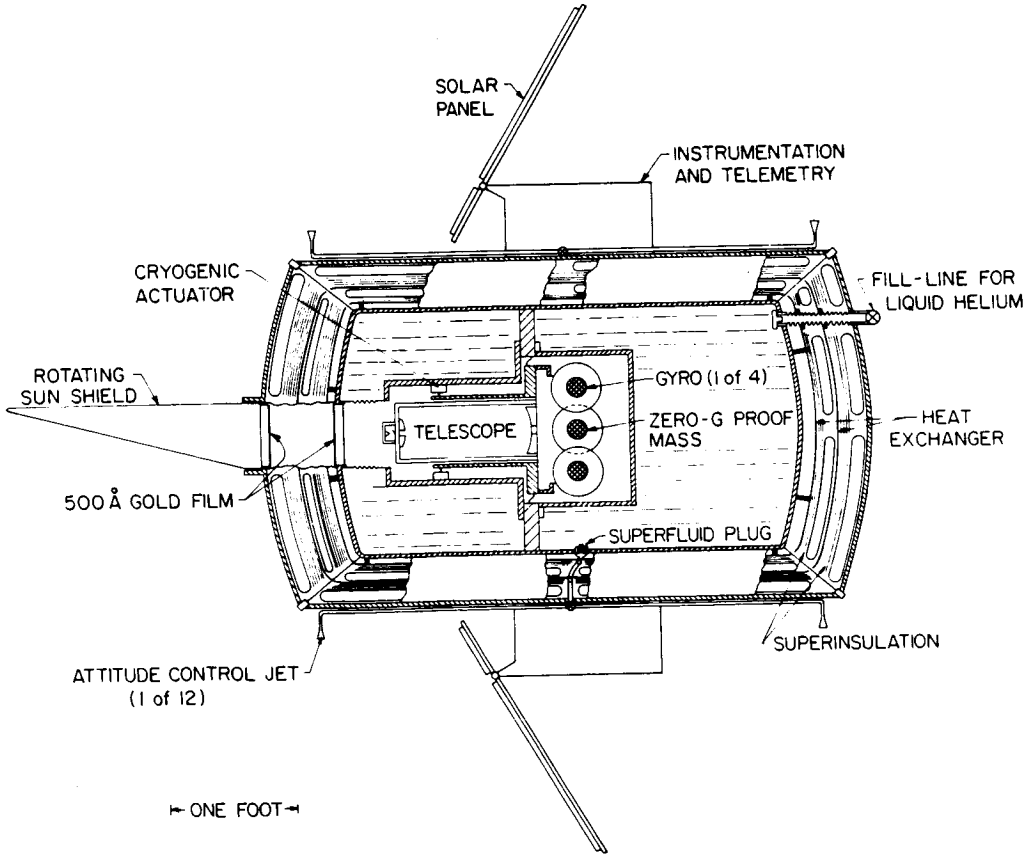


Fig. 1. General view of relativity satellite

The helium dewar has several novel features. The most interesting of these is a new solution to the problem of containing a liquid in space (Ref. 5). The dewar is filled with superfluid helium and sealed with a fine porous plug of high thermal conductivity. With an appropriate choice of design parameters, the flow is self-regulating; the helium emerges from the plug, evaporates on the outside, and refrigerates the dewar by conduction through the plug materials. Inside the vessel, thermal equilibrium is maintained by the creeping helium film.

Great attention has also been paid to thermal design. Conductive heat loads are minimized by suspending the inner vessel from fine wires, which may be erected in space after the temporary supports used in the launch phase have been withdrawn. The dimensions of the wires determine the resonant frequency between the inner and outer parts of the dewar and therefore influence the design of the attitude control servos. Radiative heat loads are substantially reduced by coating the outside of the satellite with paint which reflects sunlight strongly but has high emissivity in the infrared to reduce its operating temperature to about 240 K. Conversely, radiation into the telescope from the neck of the dewar is reduced by using quartz windows coated with thin gold films which transmit starlight but have low emissivity in the infrared.

Special attention is also paid to the design of the superinsulation. The overall performance is increased by a factor of approximately 20 by applying the sensible heat available from warming the helium vapor to ambient temperatures in heat exchangers located at optimized positions in the superinsulation. Present estimates suggest that a 300-l dewar may be expected to hold helium for more than a year. Many of the design features of the space dewar are currently under test at Stanford in the laboratory model.*

The gyroscope designed for the experiment is illustrated in Fig. 2. It consists of a ball of quartz 4 cm in diameter, coated with a thin film of superconductor, electrically supported by three mutually perpendicular sets of condenser plates, and surrounded by a superconducting magnetic shield. The ball is spun up initially to a speed of about 200 rev/s by a gas-jet system designed by Bracken and Everitt (Ref. 6), after which the gas is pumped out and the ball is allowed to run freely in a vacuum. The basic idea of the electrically suspended gyroscope is not new. The present design is a modification of the one invented by Nordsieck and coworkers and subsequently developed by Honeywell, Inc.

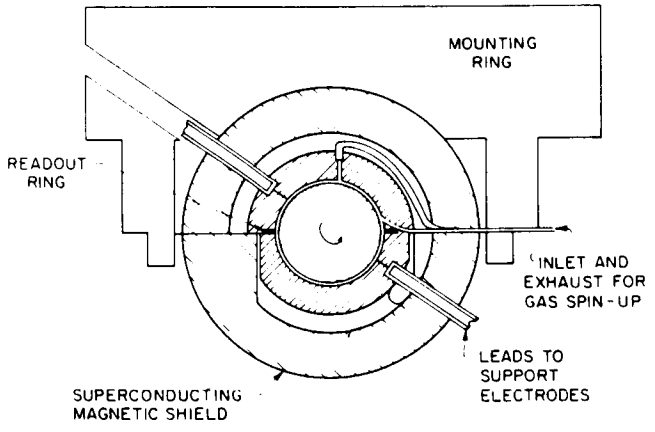


Fig. 2. Gyroscope for laboratory experiment

When such a gyroscope is operated on earth, by far the largest external torques acting on it are those arising from the electrical suspension and from mass unbalance of the rotor. In space, the suspension voltage may be turned down almost to zero, and hence such torques may be reduced below the level needed. There are, however, a number of torques which are not reduced in space, and it is in eliminating these that low-temperature techniques prove specially helpful. One of the most significant is the torque due to magnetic fields. In ordinary electrically supported gyroscopes, the interaction of eddy currents in the spinning metal rotor with magnetic fields obtained by conventional shielding techniques leads to residual drift rates several orders of magnitude larger than Ω_0. The rotor for the present gyroscope, being made of non-magnetic quartz coated with superconductor is not susceptible to any magnetic torques except those associated with trapped flux and with the small magnetic moment generated in a spinning superconductor known as the London moment. These turn out to be negligible provided the superconducting shield is cooled initially in a field below 10^{-6} G. Shielding techniques developed at Stanford and described in the paper by Fairbank already surpass that limit by 2 orders of magnitude.

The calculations summarized in section 2 show that, with appropriate design, all known torques acting on the gyro-rotor may be reduced below the level needed. To reach the desired limits, however, the rotor has to be made spherical to 1 part in 10^6 and homogeneous in density to 1 part in 10^6. This poses a readout problem. Conventional gyro readouts require knowledge of the position of the principal axes of the ball, which is unobtainable when the moments of inertia are so nearly identical. The London moment provides a method of locating the axis without reference marks on the ball. Figure 3 shows the proposed readout. The spinning sphere generates a magnetic dipole moment M_L, parallel to the instantaneous axis of spin, of magnitude

$$M_L = \frac{mc}{3e} sr^3\omega_s = 3 \times 10^{-8} r^3 \omega_s \text{ G cm}^3 \quad (3)$$

where r is the radius of the gyro rotor and ω_s its spin angular velocity.

A superconducting loop surrounds the sphere. Since the resistance is zero, any change in flux due to change of orientation of M_L will generate a canceling current in the loop, which may be used to measure the direction of the spin axis. The circuit used to measure the current is shown in Fig. 4. A second superconducting loop (modulator) is placed

*Special credit for this work is due to J. Lipa.

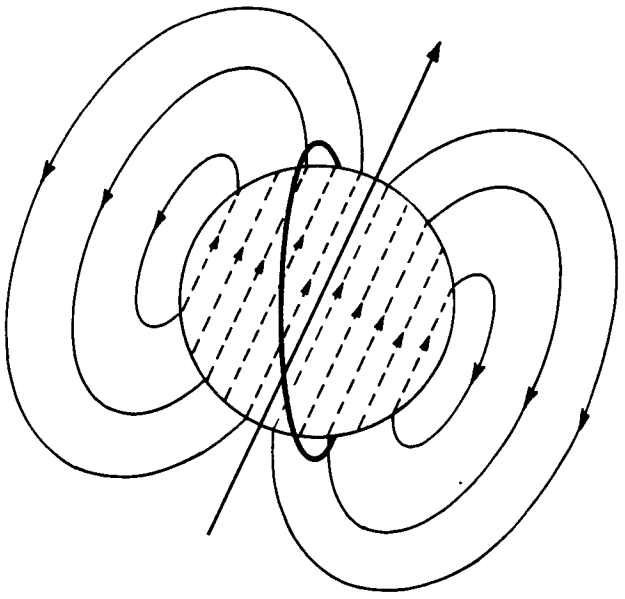


Fig. 3. London moment readout

in series with the first loop (detector). The canceling current is now distributed in the two loops proportional to the ratio of their inductances. By modulating the inductance of the second loop 10^5 times per second, a 100-kHz alternating signal is produced, proportional to the readout angle. The signal is picked up and amplified in another circuit connected in parallel with the two loops. Its magnitude is increased by connecting a tuning condenser across the output to make the system resonant. With suitable design, the readout sensitivity is 0.1 arc-s in 0.1 s of time; averaging over longer periods increases the resolutions to the desired value.

Practical development of the magnetometer is due chiefly to J. E. Opfer (Ref. 7). The inductance is modulated by a vibrating superconducting ground plane facing the second loop. driven resonantly be a quartz crystal or a magnetic forcer. Pickup is eliminated by enclosing the modulator in a superconducting shielded box. Further details of magnetometer design are given in the paper by Fairbank.

II. Summary of Error Analysis

A detailed account of the error analysis made during the past few years will be published shortly elsewhere. The present summary is intended to indicate briefly some of the principal results.

The drift rate Ω_r of a gyroscope due to some extraneous non-relativistic torque may be written

$$\Omega_r = \frac{\Gamma_r}{I\omega_s} \qquad (4)$$

where I is the moment of inertia, ω_s the spin angular velocity, and Γ_r the resolved component of the r'th torque perpendicular to the spin axis. For a gyroscope to be capable of performing the experiment, either the Ω_r must all be made much smaller than Ω_G and Ω_M, or their magnitude must be exactly known. With one partial exception to be discussed later, it is not practicable to determine the Ω_r in orbit, so the gyroscope has to be designed to make the upper limit on $\sum_r \Omega_r$ less than the desired value of 1.6×10^{-16} rad/s for Ω_0.

The moment of inertia of a sphere of density ρ and radius r is $(8\pi/15)\,\rho r^5$. Replacing ω_s in Eq. (4) by v_s/r, where v_s is the peripheral velocity of the ball, one has

$$\Omega_r = \frac{15\Gamma_r}{8\pi\rho r^4 v_s} \qquad (5)$$

Now there is an upper limit to the peripheral velocity of any gyro rotor, set ultimately by the bursting speed of the material and more immediately by the elastic distortion due to centrifugal forces. For a spinning elastic sphere, the difference Δr between polar and equatorial radii is given very nearly by

$$\frac{\Delta r}{r} = \frac{2\rho v_s^2}{7E}\left(1 + \frac{11}{14}\sigma\right) \qquad (6)$$

where E is Young's modulus and σ Poisson's ratio. Hence, if a certain limit is assigned to $\Delta r/r$, the maximum peripheral velocity is essentially proportional to $\sqrt{E/\rho}(1 - 11\sigma/28)$.

The extraneous torques may be divided into two broad categories: those related to the surface area of the ball and those related to its volume. Each surface-dependent torque Γ_σ will be proportional to (area) × (radius) × $\sigma(r)$, where $\sigma(r)$ is a function which in some instances is constant and in others depends on deviations from perfect sphericity in the shape of the ball. Over a fair range of radii, $\sigma(r)$ may be taken as proportional to r^s, where $0 < s < 1$. Thus, Γ_σ varies as $r^{(3+s)}$. Substituting into (5) and replacing v_s by the term derived from (6), one finds that the drift

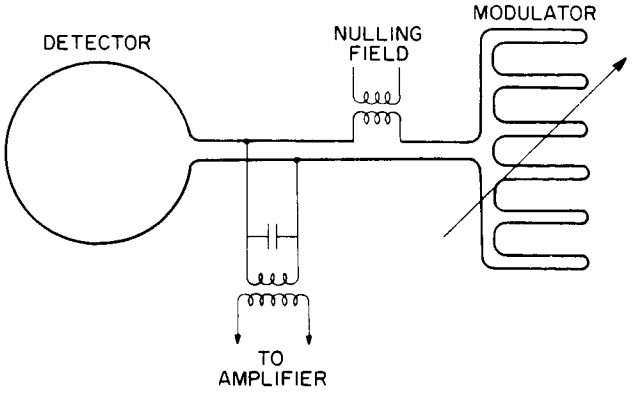


Fig. 4. Superconducting readout

rate Ω_σ arising from any surface-dependent torque of this kind has the functional form

$$\Omega_\sigma \alpha \frac{r^{(s-1)}}{\sqrt{E\rho}} \left(1 + \frac{11}{28}\sigma\right) \tag{7}$$

For a rotor of given material and radius, Ω_σ is proportional to ω_s^{-1}.

For volume-dependent torques associated with variations in density of the ball, the torque is proportional to (density) × (volume) × (radius) × $\phi(r)$, where again $\phi(r)$ may be taken as proportional to r^v and $0 < v < 1$. Assuming for the moment that different materials have identical density variations, one finds that the drift rate Ω_ϕ due to volume-dependent torques of this kind has the functional form

$$\Omega_\phi \alpha r^v \sqrt{\frac{\rho}{E}} \left(1 + \frac{11}{28}\sigma\right) \tag{8}$$

Again, Ω_ϕ varies as ω_s^{-1} for a given rotor.

There are exceptions to this general torque classification. Thus, from Eq. (3), the action of a magnetic field on the London moment is proportional to $\omega_s r^3$, and the resultant error ${}_1\Omega_m$ is proportional to $r^{-2}\rho^{-1}$. Again, errors due to differential damping of the gyroscope about different axes, caused by a gas drag or rotating trapped flux in the ball, have the form $\Omega_\delta \alpha r^{-1}\rho^{-1}$. Both ${}_1\Omega_m$ and the Ω_δ are independent of ω_s. Finally, there are the electric and gravity gradient torques associated with centrifugal distortion, which are proportional to $\Delta r/r$ in Eq. (6). The electric torques are surface-dependent and have the form $\Omega'_\sigma \alpha r^0/\sqrt{E\rho}(1 + 11\sigma/28)$; the gravity gradient torque is volume dependent and has the form $\Omega'_\phi \alpha r^{+1}\sqrt{\rho/E}(1 + 11\sigma/28)$. Both Ω'_σ and Ω'_ϕ vary as ω_s^{+1} for a given rotor. In summary, the errors due to extraneous torques depend on radius, spin speed, and mechanical properties according to the scheme set out in Table 1. In comparing the results with the explicit formulas given below, it should be observed that the expressions in the last column are all specified for fixed $\Delta r/r$.

Thus, some errors increase and some decrease with increasing rotor diameter, but except for the primary London moment torque, which may be made negligible in other ways, the advantage gained by varying the size of the rotor is meager. In fact, the diameter of 4 cm chosen for the experiment was determined by the design of the readout rather than by drift performance. The ratio $\sqrt{\rho/E}(1 + 11\sigma/28)$ is essentially constant for all materials, but $1/\sqrt{\rho E}(1 + 11\sigma/28)$ tends to decrease with increasing density. The choice of material for the rotor therefore depends on two distinct questions. To minimize the Ω_σ, Ω_δ, and ${}_1\Omega_m$, the material should be selected for lowest $1/\rho$ and $1/\sqrt{\rho E}$. To minimize the Ω_ϕ, it should be selected for extreme homogeneity. The correct choice depends upon which category of torque is dominant. Fused quartz was chosen for the present experiment chiefly for homogeneity, since its density variations could be determined optically by studying the variations in refractive index. The existing rotors are homogeneous to 1 or 2 parts/10^6. Other materials may be considered after the laboratory experiments have been completed. Once the material and dimensions of the rotor have been decided on, the optimum spin speed is determined by minimizing the sum of the drift rates with respect to ω_s. The condition is

$$\sum_{\sigma'} \Omega'_\sigma + \sum_{\phi'} \Omega'_\phi = \sum_\sigma \Omega_\sigma + \sum_\phi \Omega_\phi \tag{9}$$

A second method for categorizing the sources of drift error is to divide them into three groups as follows: (1) support-dependent torques; (2) support-independent torques; (3) support-independent disturbances of random walk type. The support-dependent torques are those associated with the action of the electrical suspension and mass unbalance of the rotor. These are greatly reduced by performing the experiment in space. Their effect depends on the time history of the acceleration environment, and influences the arrangement of the gyroscopes in the experimental package as well as the design of the rotor and suspension system. The support-independent torques include those due to the gravity gradient, the magnetic field, electric charges, and the drag of residual gas in the gyro housing. They impose conditions on the density, field strengths, and gas pressure. The disturbances in the third class arise from the random impacts of gas molecules, photons, and cosmic rays on the rotor. Their effect is not represented by Eq. (4) but by an effective drift rate:

$$\Omega_\rho = \frac{\kappa_\rho}{I\omega_s} \frac{1}{\sqrt{t_o}} \tag{10}$$

where t_o is the time of observation. Thus, in addition to constraints on the environment, the random walk disturbances establish in principle a minimum time for performing the experiment.

A typical support-dependent error is the one due to the mass-unbalance torque, when the center of geometry and center of mass of the rotor do not quite coincide. Evaluating the error for a rotor with a uniform density gradient parallel to the spin axis, and then introducing a factor $c_1 (0 < c_1 < 1)$ representing the symmetry of the actual density distribution, one finds

$$\Omega_u < \frac{c_1}{2} \frac{\Delta\rho}{\rho} \frac{f}{v_s} \tag{11}$$

where f is the residual acceleration on the gyroscope. Taking $\Delta\rho/\rho$ as 10^{-6}, c_1 as 0.2, and v_s as 2500 cm/s, the condition on $\bar{f}$ – the average value of f needed to make Ω_u less than 10^{-16} rad/s – is $\bar{f} < 2.5 \times 10^{-9}$ g. Conversely, if $\bar{f}$ were made equal to g, the variations in density would have to be less than 1 part/10^{15}. Thus, consideration of

Table 1. Relation of errors to size, speed, and mechanical properties of rotor

Error symbol	Description	Power of r (v_s fixed)	Power of ω_s (r fixed)	Mechanical properties ($\Delta r/r$ fixed)
Ω_σ	Surface-dependent due to polishing errors	$s - 1$ $0 < s < 1$	-1	$\frac{1}{\sqrt{\rho E}}\left(1 + \frac{11}{28}\sigma\right)$
Ω_ϕ	Volume-dependent due to density variations	v $0 < v < 1$	-1	$\sqrt{\frac{\rho}{E}}\left(1 + \frac{11}{28}\sigma\right)$
${}_1\Omega_m$	London moment	-2	0	$\frac{1}{\rho}$
Ω_s	Differential damping	-1	0	$\frac{1}{\rho}$
Ω'_σ	Surface-dependent due to centrifugal distortion	0	+1	$\frac{1}{\sqrt{\rho E}}\left(1 + \frac{11}{28}\sigma\right)$
Ω'_ϕ	Volume-dependent due to centrifugal distortion	+1	+1	$\sqrt{\frac{\rho}{E}}\left(1 + \frac{11}{28}\sigma\right)$

the mass-unbalance torque alone immediately reveals the advantage of performing the experiment in space.

The electrical suspension system utilizes three mutually perpendicular pairs of condenser plates deposited on the spherical inner surface of the gyro housing. The plates are circular, with a diameter of about 2 cm, and are about 5×10^{-3} cm from the rotor. Twenty-kilocycle alternating voltages are applied to each plate; when the gyroscope experiences an acceleration parallel to an electrode axis, the voltage is raised on one plate and lowered on the other to keep the ball centered. Such a system does not generate any torques on a perfect sphere of high electrical conductivity, since changes in the orientation of the sphere do not alter the energy in the plates. Thus, suspension torques depend on deviations from sphericity in the shape of the gyro rotor. They are evaluated by expanding the shape in a series of spherical harmonics and determining the effect for each term.

Since the gyroscope spins at high speed, it may be treated as an axially symmetric body. Upper bounds on the drift errors are obtained by associating the known limits of distortion and polishing precision with the first significant even and odd harmonics, as the effects of higher-order terms are necessarily smaller. For a suspension system in which the electrodes are perfectly spherical and symmetrical, three terms only need to be considered: (1) an ellipsoid misaligned with the electrode axes, (2) an ellipsoid with its axis parallel to the electrodes but displaced from the center of the cavity, (3) a misaligned pear-shaped body. There are also terms arising from asymmetries in the electrodes, which are small compared with the worst cases of the misaligned bodies, but become significant if advantage is taken of the reduction in the main terms due to symmetry.

The torques are first evaluated in terms of the potentials applied to the condenser plates, and then most conveniently expressed in terms of the pre-load acceleration h, where h is a measure of the stiffness of the suspension, defined as the acceleration parallel to an electrode axis required to send the voltage on one plate to zero. For $h = 10^{-6}$ g, which is a reasonable stiffness for the space experiment, the normal operating voltage is found, on substituting numerical values, to be about 0.7 V. On earth, h must be greater than g, and the operating voltage is about 1000 V.

The torques on an elliptical rotor are significantly reduced by designing the suspension system to maintain the preload accelerations nearly equal on all three axes and by making the gyro spin axis nearly coincide with an axis of symmetry among the electrodes. Defining a preload compensation factor ζ equal to $(h_x - h_y)/h_z$ etc., and a misalignment from perfect symmetry of θ_0, one finds the following upper limit on the drift rate ${}_1\Omega_e$ due to a misaligned elliptical rotor:

$$ {}_1\Omega_e < 1.35 \frac{\Delta r}{r} \frac{1}{\rho v_s} \left[\frac{f^2}{h} + \zeta h\right] \theta_0 \qquad (12) $$

The torque may also be reduced by introducing "sum-of-the-squares" control, that is, by mechanizing the suspension system in such a way as to hold constant the sum of the squares of the voltages on the six condenser plates. In the present state of technology, sum-of-the-squares control serves chiefly as an alternative to electrode symmetry, since either method alone is capable of reducing the torque on an ellipsoidal rotor to a value comparable with the higher-order electric torques. If it is used, θ_0 in Eq. (12) must be replaced by a sum-of-the-squares control factor ξ, defined in a similar way to the preload compensation factor ζ. In a typical suspension, ζ may be about 2×10^{-3} and θ_0 or ξ about 5×10^{-3}. In (12), Δr is the combined deviation from sphericity due to polishing errors and centrifugal distortion. In practice, the centrifugal term is dominant and $\Delta r/r$ may be evaluated from (6), giving

$$ {}_1\Omega_e' < 0.39 \frac{v_s}{E}\left(1 + \frac{11\sigma}{14}\right)\left[\frac{f^2}{h} + \zeta h\right]\Bigg\{^{\theta_0}_{\xi} \qquad (13) $$

where the curly bracket signifies that either θ_0 or ξ must be substituted, depending on the choice of mechanization.

The torques due to a miscentered ellipsoid and a pear-shaped rotor are obtained in a similar way. The upper limit on drift rate for a miscentered ellipsoid is

$$ {}_2\Omega_e < 3.28 \frac{\Delta r}{r} \frac{t}{d} \frac{f}{v_s} \qquad (14) $$

where t is the displacement, d the rotor-electrode gap, and again, $\Delta r/r$ is determined primarily by centrifugal distortion. The upper limit for a pear-shaped rotor, assuming the use of symmetry, is

$$ {}_2\Omega_e < 2.58 \frac{\Delta r'}{r} \frac{f}{v_s} \theta_0 \qquad (15) $$

where, in this instance, $\Delta r'/r$ is due entirely to polishing errors. If no advantage is taken of symmetry, θ_0 in Eq. (15) should be replaced by 0.67.

Equations (11) to (15) give upper limits on the drift rates associated with the instantaneous acceleration acting on the gyroscope. Some advantage may also be gained from the time-averaging of the accelerations. Here the most prudent course is to take averages over each satellite orbit, any additional averaging over longer periods being treated as insurance rather than as a reliable design parameter. The discussion is restricted to nearly circular polar orbits; comments about non-polar orbits are given in section 3. The residual acceleration on the gyroscope may be written

$$ \underline{f} = \underline{f_1} + \underline{f_2} + \underline{f_3} \qquad (16) $$

where $\underline{f_1}$ is the acceleration arising from air drag, solar radiation, and other external forces on the satellite, $\underline{f_2}$ is the action of the self-gravitation of the satellite on the gyroscope, $\underline{f_3}$ is the gravity gradient acceleration arising from the earth's field because the gyroscope is some distance from the mass center of the satellite. For a typical satellite in a 500-mile orbit,

$$ f_1 \sim 3 \times 10^{-8}\ \text{g} $$

$$ f_2 \sim 10^{-9}\ \text{g} $$

$$ f_3 \sim 1.5\, \ell(\underline{a} \sin\alpha + 2\underline{b} \cos\alpha) \times 10^{-9}\ \text{g} $$

where ℓ is the distance between gyroscopes and mass center, α is the angle between $\underline{\ell}$ and the orbit plane, and $\underline{a}$ and $\underline{b}$ are unit vectors defining the plane through $\underline{\ell}$ perpendicular to the orbit.

To some extent, f_1 averages over an orbit, but since the atmospheric density varies from night to day and since the departures from circularity in the orbit may be significant in comparison with the scale height of the atmosphere, the improvement so gained is probably not more than 20% per orbit at best. The average $\bar{f}_1$ is therefore about ten times greater than f_2. Because the limitation on zero-g control is determined by the action of the self-gravitation of the satellite on the control body, which is comparable with the self-gravitation on individual gyroscopes, the average acceleration $\bar{f}$ in a drag-free system may be made comparable with f_2 or 10^{-9} g. Thus, the drift performance obtained with a drag-free satellite is about one order of magnitude better than with an uncompensated satellite. The effectiveness in averaging f_3 depends on the position of the gyroscope in the satellite. If the gyroscope is displaced from the mass center in a direction perpendicular to the orbit plane, there is no averaging in any one orbit, although the effects may be averaged over long periods by rotating the satellite at intervals through 180 deg about the telescope axis. On the other hand, averaging of f_3 does take place when the gyroscope lies in the orbit plane. For the terms Ω_u, ${}_2\Omega_e$, and ${}_3\Omega_e$, which are linear in f, the errors vary sinusoidally at orbital frequency and the averaging may be assumed good to about 1%, the residue being due to imperfect alignment and departures from circularity in the orbit. Taking ℓ as 15 cm, $\bar{f}_3$ is about 5×10^{-10} g, that is, somewhat less than $\bar{f}$. The contributions to f_3 from the gravity gradients of the sun and moon are negligible. For the term ${}_1\Omega_e'$, however, which is quadratic in f, the effect of f_3 rectifies, and the errors average only to an extent proportional to $[f_0^2/2h + \zeta h]$, where f_0 is the peak value of f_3. Given f_0 and ζ, the error is minimized by making the preload h equal to $f_0/\sqrt{2\zeta}$, that is, about $10 f_0$.

The suspension is made automatically adjustable to higher preloads for brief emergencies such as the impact of micrometeorites on the satellite. With reasonable numerical values as set out in Table 2, the worst-case sum of the four support-dependent errors averaged over each orbit, assuming the use of symmetry or sum-of-squares control, is 1.2×10^{-16} rad/s. This figure corresponds to an error of 0.0007 arc-s/year. Some additional improvement may be derived from long-term averaging, and from occasional rotations of the satellite about the telescope axis, but the advantage so gained is small. Thus, the support-dependent errors in a polar orbit are essentially independent of mission lifetime.

The support-independent errors include the effects of magnetic fields, residual electric charge on the ball, residual gas in the cavity, and the action of the gradient of the earth's gravitational field on the quadrupole mass moment of the gyroscope. The secular drift rate Ω_g due to gravity gradients may be written

$$ \Omega_g = \frac{3}{2} \frac{\Delta I}{I} \frac{r}{v_s} \frac{GM}{R^3} \sin 2\beta \qquad (17) $$

Table 2. Design requirements for accuracy of 10^{-16} rad/s

Gyro rotor	Homogeneity	$\Delta\rho/\rho \sim 10^{-6}$
	Sphericity at rest	$\Delta r'/r \sim 10^{-6}$
Gyro suspension	Sphericity of cavity	$\Delta r''/r \sim 10^{-5}$
	Centering accuracy	$t/d \sim 10^{-2}$; $t \sim 4 \times 10^{-5}$ cm
	Preload: acceleration	$h \sim 10^{-6}$ g
	Preload: voltage	$V \sim 0.7$ V
	Preload compensation factor	$\zeta \sim 3 \times 10^{-3}$
	Symmetry or sum-of-squares control factor	θ_0, ξ $\sim 5 \times 10^{-3}$
Environment	Acceleration due to self-gravitation	$f \sim 10^{-9}$ g
	Acceleration due to earth's gradient	$f_o \sim 2 \times 10^{-8}$ g
	Residual magnetic fields: best	$H < 10^{-7}$ G
	Residual magnetic fields: using symmetry	$H < 4 \times 10^{-6}$ G
	Residual electrification on ball: charge	$Q < 10^{8}$ electrons
	Residual electrification on ball: voltage	$V < 0.03$ V
	Residual gas pressure: magnitude	$p \sim 10^{-9}$ cm Hg
	Residual gas pressure: gradients	$\Delta p/p \sim 5 \times 10^{-2}$
Spinup system	Torque switching ratio	$\Gamma_r/\Gamma_s \sim 2 \times 10^{-13}$

where ΔI is the difference between principal moments of inertia, β is the angle between the gyroscope axis and the orbit plane, and G, M, and R have the same meanings as in Eq. (1). The difference in moments of inertia may arise from inhomogeneities in the ball, in which case $\Delta I/I$ may be replaced by $c_2\,\Delta\rho/\rho$, where c_2 is a symmetry factor different from and smaller than c_1 in Eq. (11); alternatively, the difference may arise from centrifugal distortion of the ball, in which case $\Delta I/I$ may be replaced by $\Delta r/r$ from Eq. (6). According to (17), the gravity gradient error vanishes when the gyroscope is parallel or perpendicular to the orbit plane, as is to be expected from symmetry. Actually, slight corrections must be applied to this condition when the oblateness of the earth is taken into account (Ref. 8). The equations may be transformed into a condition on $\Delta I/I$ for the required performance. For a gyroscope aligned to within 1 deg of the orbit plane, which is reasonable for a near-polar orbit, $\Delta I/I$ must be less than 1.5×10^{-6}. The error Ω'_g associated with centrifugal distortion of the ball differs from every other disturbance acting on the gyroscope in that it depends on quantities whose magnitude can be calculated exactly. Furthermore, in non-polar orbits, it may be independently checked in space by the methods described in section 3. Thus, even if the choice of orbit is such as to make Ω'_g appreciable, it may legitimately be corrected for in data analysis.

The chief magnetic disturbance comes from the action of any residual trapped flux in the magnetic shield on the London moment. The magnitude is

$$ {}_1\Omega_m = 2 \times 10^{-8} \frac{H_s}{\rho r^2} \qquad (18) $$

where H_s is the component of trapped flux perpendicular to the spin axis, and the numerical constant has been evaluated from the quantities in Eqs. (3) and (5). Two other magnetic disturbances, ${}_2\Omega_m$ and ${}_3\Omega_m$, have been evaluated due, respectively, to the reaction of the gyro readout loop on the London moment and on the residual trapped field H_t in the ball. In general, H_t and H_s are roughly equal. If the gyro-shield is maintained in a fixed orientation, Eq. (18) sets an upper limit

on the H_s of 10^{-7} G. This requirement may be relaxed slightly if the gyroscope is aligned parallel to the telescope axis, since the sign of ${}_1\Omega_m$ may then be periodically reversed by rolling the satellite about the telescope axis. In that case, the upper limit on the H_t set by ${}_3\Omega_m$ is about 5×10^{-6} G. The London moment reaction error ${}_2\Omega_m$ is negligible unless the gyroscope spin axis is inclined at an angle greater than 5 deg to the readout loop.

Residual electric charge on the ball may cause a torque through the reaction of induced charge on the inner surface of the gyro housing. This term evidently vanishes if the ball is exactly centered and either it or the cavity is perfectly spherical. It may be evaluated using a spherical harmonic expansion given originally by Maxwell for the problem of "a nearly spherical conductor enclosed in a nearly spherical and nearly concentric conducting vessel" (Ref. 9). Assuming that the deviations of the ball from sphericity are dominated by centrifugal distortion, the error is

$$ {}_4\Omega_e < \frac{Q^2}{14E}\left[1 + \frac{11}{14}\sigma\right]\left[\frac{v_s}{r^4}\right]\left[\frac{5t}{d'} + \frac{3\Delta r'}{d'}\right]\left[\frac{A'd}{A'd + Ad'}\right] \tag{19} $$

where Q is the total charge in electrostatic units, t the displacement of the ball from the center of the cavity, A the electrode area, d the rotor-electrode distance, and A', d', and $\Delta r'$ are the area, distance, and polishing error of the elevated conducting regions which surround the gas spinup channels.

Numerical substitution gives an upper limit of 10^8 electrons for the charge on a ball of 2-cm radius to keep the drift rate below 10^{-16} rad/s. The corresponding static potential difference between the ball and the cavity is 0.03 V, which may be contrasted with the potential of 0.7 V applied to the support. The difference comes about through the absence of symmetry and compensation factors in the static torque. Since $Q = 4\pi r^2\sigma$, the error ${}_4\Omega_e$ is independent of the radius of the ball for constant charge density σ.

Residual gas in the cavity gradually slows down the ball. The exponential spindown time τ is inversely proportional to the pressure p, and at a pressure of 10^{-9} mm is about 1000 years. A drift error will occur if the value of τ varies with the orientation of the spin axis. In practice, τ is independent of orientation so long as p is uniform, even though the ball and cavity may be nonspherical, but significant errors may occur if there are any pressure gradients in the cavity. In a typical situation, the error is

$$ \Omega_p = \frac{5}{6}\sqrt{\frac{m}{2\pi k}}\frac{\Delta p}{\rho r T^{1/2}} \sin 2\gamma \tag{20} $$

where m is the mass of a helium atom, k Boltzmann's constant, T temperature, Δp the total pressure variation in the cavity, and γ the angle between the spin axis and the axis of minimum pressure. For a pressure of 10^{-9} mm Hg, Ω_p may be reduced to the required value by making γ about 2 arc-min and $\Delta p/p$ about 5%.

There remain the random walk disturbances. A gyroscope at a given temperature T may be expected, on statistical mechanical grounds, to experience fluctuations in the transverse components of its angular momentum ΔJ of order $\sqrt{kTI}$. The root-mean-square fluctuation in direction $\bar{\theta}$ is then given by

$$ \bar{\theta} \sim \frac{1}{\omega_s}\sqrt{\frac{kT}{I}} \tag{21} $$

Equation (21) gives the jitter in direction and sets a limit on the precision of measurement. If Ω is the desired drift rate (10^{-16} rad/s), the minimum time t_0 for a significant observation is $\bar{\theta}/\Omega_0$, which yields, on numerical substitution, a value for t_0 of about 6 h. This limitation of measurement must be distinguished from the random walk drifts of the spin axis, which are evaluated by considering the transfer of angular momentum to the ball through each impact of a gas molecule, photons, or cosmic rays. Much the largest of these is due to gas molecules, which give an effective drift rate

$$ {}_1\Omega_\rho \sim \frac{3}{4\pi}(N_0^{1/2} m^{1/2} k^{3/4}\frac{p^{1/2}T^{3/4}}{\rho r^2 v_s} t^{-1/2} \tag{22} $$

where N_0 is Loschmidt's number and the remaining quantities have the meanings assigned in Eq. (21).

In contrast to Ω_p, the error ${}_1\Omega_p$ decreases with decreasing temperature. The random walk motion ${}_2\Omega_\rho$ due to photons from the walls of the cavity is a much stronger function of temperature, scaling as $T^{5/2}$. Numerical substitution makes ${}_1\Omega_\rho$ about $1.5 \times 10^{-16} t^{-1/2}$ at a temperature of 2 K and a pressure of 10^{-9} mm Hg. The effects of photons and cosmic rays are at least two orders of magnitude smaller. Thus, the random walk drifts are negligible after a few seconds, and the only significant effect of thermal fluctuations is to set a minimum time for observation through jitter. The reason for the drifts being so much smaller than the jitter, as first explained to me by L. I. Schiff, is that the relaxation times of the processes are very much longer than the characteristic observation time t_0.

One more important problem bearing on the gyro performance is spinup. To spin a gyro rotor initially at rest, a torque Γ_s of some kind must be applied for a time t_s, after which Γ_s must be reduced to a level where the residual component Γ_r perpendicular to the spin axis does not cause significant drift errors. Neglecting drag torques, $\Gamma_s = I\dot{\omega} = I\omega_s/t_s$. Substituting into Eq. (4), one finds

$$ \frac{\Gamma_r}{\Gamma_s} < \Omega_0 t_s \tag{23} $$

Since it is operationally convenient for t_s to be shorter than the half period of the satellite orbit (2700 s), its upper limit is about 2000 s; hence, with the upper limit on Ω_0 of 10^{-16} rad/s, Γ_r/Γ_s must be less than 2×10^{-13}. If drag torques during spinup are taken into consideration, the torque switching ratio required is even more extreme. There are also two other general restrictions on any spinup mechanism which will be discussed elsewhere. The design of the gas spinup system now planned for the experiment is described in Ref. 6.

The error analysis provides design restrictions on the rotor, housing, suspension system, spinup system, and environment, which must be fulfilled to attain a drift performance of 10^{-16} rad/s. Results for a typical gyroscope of 4-cm diameter, spinning at 200 Hz, and contained in a satellite moving in a 500-mile near-polar orbit are summarized in Table 2. The various restrictions are not independent: an improvement in performance in one area would allow relaxation of restrictions elsewhere. The immense challenge that exists in achieving all the requirements simultaneously needs no emphasis; nevertheless, each one does seem within the bounds of possibility. The specifications on the gyro rotor, low magnetic fields, and drag-free satellite design have already been met in the laboratory. The three areas in which most progress is needed are: (1) mechanizing the suspension electronics for space, (2) reaching satisfactory operating pressures, (3) maintaining low residual charge on the ball. Work on each problem is in progress at Stanford.

III. Readout, Instrumentation, and Attitude Control

Given adequate drift performance, the experimental problem reduces itself to designing gyro and telescope readouts, electronics instrumentation, and an attitude control system precise enough to handle the relativity data and transmit them to earth without error. Before these are discussed, it is useful to establish some general requirements determined by the form of the data. The final output consists of difference signals between the gyroscope and telescope readouts, whose content may be divided into six parts:

(1) Relativity drift, the measurement of which is the goal of the experiment. In a near-polar orbit, this will consist of linearly increasing signals with predicted values on Einstein's theory of 6.9 arc-s/year for Ω_G and 0.05 arc-s/year for Ω_M. In non-polar orbits, the signals will be modulated through the apsidal rotation of the orbit plane, and will normally consist of superposed linear and sinusoidal terms, whose magnitudes and period will depend on the orbit and on the orientations of the gyroscopes.

(2) Proper motion of the reference star.

(3) Annual aberration of starlight, which gives a sinusoidal displacement between gyroscopes and telescope outputs in the plane of the ecliptic, with amplitude of ±20.116 arc-s and period of 1 year.

(4) Orbital aberration of starlight, with amplitude in the plane of the orbit of about ±5 arc-s and period of 90 min. When the telescope axis lies in the orbit plane (as it must for measurement of Ω_G in a polar orbit), the star is occulted by the earth for half of each orbit, and only one side of the sinewave is observed.

(5) Disturbances due to attitude motions of the satellite.

(6) Intrinsic noise and errors in the gyroscope and telescope readouts.

For a satellite in a polar orbit, then, the total range of motion between the gyroscope and telescope axes may be as much as 20 arc-s in the plane of the orbit and 40 to 50 arc-s in the plane of the ecliptic. In principle, the displacement might be allowed to appear in either the gyroscope readout or the telescope readout, or even in both. In practice, it is a far less forbidding problem to obtain wide linear range from the gyroscope, which has a readout derived from linear changes in a magnetic field, than from the telescope, which depends on the motions of a star image. Thus, the telescope is maintained as close to null as possible, and the displacements due to relativity drift and aberration are taken up in the gyro readouts.

Consideration must next be given to the linearity, zero stability, scaling accuracy, and noise performance of the readouts, and to the bearing of these parameters on attitude control. Thereafter, the various merits of different satellite orbits will be briefly examined.

The ultimate limit to sensitivity of the gyroscope readout circuit illustrated in Fig. 4 is set by the input noise of the amplifier. An analysis by J. M. Pierce* gives the mean-square error of the measurement flux in electromagnetic units as

$$\overline{\phi^2} = \frac{(1+\beta)^3}{\beta} \frac{L}{\alpha^2 Q} k T_{eff} \frac{\Delta\nu}{\nu} \qquad (24)$$

where L is the average total inductance, α the modulation factor, β the ratio of inductances in the modulator and detector, Q the quality factor of the resonant circuit, T_{eff} the effective noise temperature of the amplifier, ν the modulation frequency, and $\Delta\nu$ the bandwidth. Optimum performance is obtained with β equal to 1/2. In the current design, $\alpha \sim 0.2$, $Q \sim 10^4$, $T_{eff} \sim 30$ K $\nu \sim 100$ kHz. The value of L is determined by the dimensions of the readout loop. Combination of (24) with (3) and with the equation for inductance of a single loop of wire yields a formula for the noise-equivalent angle of a gyroscope with optimum β:

$$\overline{\theta_g} > \frac{2.4}{\nu_s \sqrt{r}} \sqrt{\frac{T_{eff} \Delta\nu}{\alpha^2 Q \nu}} \qquad (25)$$

*In Proc. Symposium on Superconducting Devices, University of Virginia, April 28-29, 1967.

with a bandwidth of 10 Hz, the resolution is 0.1 arc-s. Greater resolution is of course obtained by integrating for longer periods.

Equation (25) may be transformed to give t_0', the minimum time needed to determine Ω_0, by replacing $\bar{\theta}_g$ by $\Omega_0 t_0'$ and $\Delta\nu$ by $1/t_0'$. The result is that t_0' must exceed 10^6 s or about 20 days. This may be compared with the value for t_0 obtained from (23) of about 6 h. At the present stage of experimental development, therefore, resolution is limited by gyro readout accuracy rather than by statistical mechanical jitter of the ball.

Since the readout circuit depends on measurement of the current in a superconducting circuit, it has, in principle, absolute null stability. However, errors may arise from changes in the external field acting on the gyroscope even though much of the external flux is excluded from the loop by the presence of the superconducting ball. If the gap between the ball and the loop is s' (about 0.05 cm), the maximum change in field allowed over t_0' is $10^{-7} v_s \Omega_0 t_0'/2s'$ or about 10^{-12} G. Field stabilities of this level may be assured by using a superconducting shield more than 10^{-4}-cm thick and by attaching a long superconducting tube to each hole in the shield. It is possible but probably not necessary to relax the limit by counterwinding a second loop of area approximately $2\pi r s'$ around the outside of the detector. Similar considerations apply to the modulator.

Another possible source of zero error is trapped flux. This may affect the system in three ways. Rotating trapped flux in the ball may, if large enough, saturate the readout amplifiers. It also causes a rectification signal, associated with phase shifts in the readout circuit, which may change and produce a zero error as the ball slows down. Finally, trapped flux in the vibrating ground plane of the modulator introduces an error if the modulation amplitude changes. With proper design, all these terms may be made negligible so long as the trapped fields are kept below 10^{-6} G.

The last important source of zero error is displacement of the ball from the center of the readout loop. Defining the displacement axes with x in the direction of spin and z perpendicular to the plane of the loop, the error angle θ_d is

$$\theta_d = \frac{\pi}{3}\frac{xz}{r^2}\left(1 - \frac{2\sqrt{x^2 + y^2}}{\pi r}\right) \qquad (26)$$

From (26), the maximum displacement allowed during the observation time of 20 days is 2×10^{-5} cm, which may be compared with the limit of 8×10^{-6} cm determined by the suspension torques. After deriving Eq. (26), I was interested to discover that Maxwell had obtained the analogous formula for a magnet with axis parallel to z in 1863 in his beautiful analysis of the errors of the British Association determination of the ohm (Ref. 10).

Linearity of the gyroscope readout is ensured by a feedback circuit, using a digital signal with 17-bit accuracy. The scale factor is calibrated to 1 part in 10^5 by a precision dc reference. Thus, provided the restrictions set out below are fulfilled, the gyroscope readout gives the desired performance.

Field constancy	$< 10^{-12}$ G
Trapped field	$< 10^{-6}$ G
Centering stability	$< 2 \times 10^{-5}$ cm
Scaling accuracy	1 part in 10^5
Minimum time of observation	20 days

The ultimate limit to sensitivity of the telescope is set by photon noise from starlight. The noise-equivalent angle $\bar{\theta}_t$ for each readout is calculated from the fluctuations in the intensity of light falling on the photocell during each signal period. For a star of magnitude M and color temperature $\bar{\lambda}$, the result is approximately

$$\bar{\theta}_t = 2 \times 10^{-6}\frac{\delta}{D}\sqrt{\frac{2.51^{-M}\Delta\nu}{\lambda\epsilon\eta}} \qquad (27)$$

where δ is the image diameter, D the telescope aperture, ϵ the effective light loss in each channel, η the quantum efficiency of the detectors, and $\Delta\nu$ the bandwidth. For diffraction-limited optics, δ may be taken as $1.22\gamma/D$; and with a telescope of 5.5-in. aperture having $\epsilon\eta$ about 0.01, the direction of a first-magnitude star is resolved to 0.01 arc-s in 0.1 s of time. Resolutions approaching this limit have been obtained on an experimental star simulator and detector in the laboratory. Significant nonlinearities occur if the range of motion Δ of the image exceeds

$$\Delta < 1.15\delta^{2/3}\Omega_0^{1/3}t_0'^{1/3} \qquad (28)$$

or, for diffraction-limited optics with $t_0' = 20$ days, about ±0.03 arc-s. With defocussing, a wider range is allowed, but then $\bar{\theta}_t$ increases and other errors also occur.

More precise computer studies by W. L. Pondrom, Jr., confirm the essential features of the results based on (27) and (28). Their significance is twofold. First there is a lower limit on the brightness of the reference star that can be used without introducing excessive noise into the final output and increasing the observation time t_0'. The limit is obtained by setting $\bar{\theta}_t$ equal to $\bar{\theta}_g$. For the present telescope design, the dimmest acceptable star is of third magnitude. Second, the two equations supply design criteria for the telescope pointing system. The present two-loop system was conceived by me in collaboration with D. B. DeBra, D. P. Chandler, and R. A. Van Patten. Analysis by J. Bull establishes that adequate pointing accuracy can be achieved in the presence of telescope noise and normal satellite disturbances. Occasional emergencies such as telescope flares or meteorite impacts are handled by interrupting data acquisition until the disturbance has settled out.

Errors occur in the telescope if the roof prism has curvature, or small nicks on the dividing edge. An analysis by R. A. Nidey gives the maximum diameter ∂ relative to the objective as

$$\partial < 3.6\Delta^{3/4}\Omega_0^{1/4}t_0'^{1/4} \qquad (29)$$

from which the maximum ∂ for a telescope of 150-in. focal length is 25×10^{-6} cm. Prisms have already been made with no nicks greater than 8×10^{-6} cm in diameter.

Other optical and electronic errors in the telescope readout have been studied in some detail. The "seeing" due to atmospheric disturbances, which would be disastrous on earth, vanishes completely so long as observations are confined to the portion of the orbit in which the telescope is well above the horizon. Formulas have also been obtained for mechanical distortions of the gyro-telescope structure. It is essential for the entire structure to be optically contacted together and maintained at low temperatures. At ambient temperature, the required stability could be reached only if the transverse heat flux falling on the satellite were balanced at the telescope to within 1 part in 10^6. At helium temperatures, the permissible heat flux is several times the total input into the dewar. On earth, the telescope sags under its own weight by about 0.1 arc-s, and also undergoes creep. In space, both effects are negligible. There is, however, a significant possibility of error from the delayed elastic effect associated with the relaxation of inherent strains in the material. These may be eliminated by proper manufacturing procedures; in particular, by annealing the telescope parts in a vertical position for several days at 1100°C so that the strains acquired under gravity have axial symmetry. The last important telescope error to be considered arises if there is a long-term pointing error Δ' combined with differences in scaling between the gyro and telescope readouts. If the nominal scale factor of each readout is k, but there is a difference Δk between them, then the limit on $\Delta k/k$ is

$$\frac{\Delta k}{k} < \frac{\Omega_0 t_0'}{\Delta'} \qquad (30)$$

with $t_0' = 20$ days and $\Delta' = 0.03$ arc-s, the maximum acceptable scaling difference is 1 part in 10^3.

An instrumentation system has been conceived with processes relativity data and, at the same time, provides signals for use in attitude control and in calibrating the telescope scale factor. The proposed design, which is due largely to R. A. Van Patten, is illustrated in Fig. 5. Portions of the system and gyro electronics are now being developed by J. Nikirk. The heavy lines represent an integrating data loop which supplies continually updated relativity information in digital form after subtracting and summing the gyroscope and telescope readouts, with the final signal in the precision summing amplifier Σ_1. The amplifier output consists of an amplitude-modulated, suppressed-carrier, alternating current signal. This signal is processed in a sampling demodulator and filter to obtain a direct current output with extremely low zero offset. It is then integrated by means of a bipolar voltage-to-frequency converter driving a 17-bit up-down binary counter, which contains the readout signal for storage and telemetry. The integrating loop is closed via a 17-bit digital-to-analog converter summed into Σ_1. Its operation may be understood as follows. Call the gyro output G, the telescope output T, and the signal in the up-down counter R. The summing amplifier provides the function (T - G + R), which is maintained at null, making the final signal R equal to (G - T). This is the quantity of interest in the experiment.

The pointing servos are controlled by signals derived from the second summing amplifier Σ_2. In normal operation, Σ_2 gives the direct telescope error, but an interesting feature of the design is that interruptions of data do not impede the action of the control loop. When the telescope signal is interrupted, for example by meteorite impact or occultation of the star by the earth, the output will continue to represent attitude motions but will refer them to the function (G - R), where R is held static and T inhibited by the action of control logic. Thus, the telescope remains pointing in the correct line of sight by reference to the gyroscope. When data are resumed, the reversible counter will run until the instrumentation loop has been driven to null, and the output will come to the value it would have reached with continuous recording. The only loss is the immediate loss of data for the final statistical analysis. The telescope scale factor is calibrated against the gyroscope by introducing a low-frequency dithering signal into the inner servo loop, which makes the entire gyro-telescope structure swing back and forth through a small angle across the line of sight at about 0.1 Hz. If the scale factors of the two readouts are not equal, a signal appears at the output of the instrumentation loop, where it is synchronously detected; and automatic gain control for the telescope is obtained by digital sampling, subtraction, and difference accumulation.

The instrumentation system may be made with solid-state integrated circuitry for maximum reliability. It provides enough filtering to process the data and store them in a memory of reasonable size, without the complexities and reliability problems of an on-board computer. The choice of time constant for the integrating loop depends on several factors. The optimum figure appears to be about 20 s. This is long enough to provide substantial noise filtering while allowing the system to recover from interruptions within 1 or 2 min. The output of the instrumentation loop is sampled every few seconds and stored in a core memory containing 10^4 words of 17 bits, from which it is transmitted about once per day for further processing on the ground. With a time constant of 20 s the output of the data loop consists of (1) the full relativity signal, (2) the full annual aberration in ecliptic plane, (3) 99.98% orbital aberration in the orbit plane, and (4) residual noise. An important feature of the experiment is that the aberration signal, being known with great precision, provides an inherent calibration of the data. Thus, the telescope scale factor (which might be expected to change through effects such as aging of the photomultiplier) is automatically driven to the same value as the gyroscope by the dithering technique, while the gyroscope scale factor (which should remain constant) is automatically checked and calibrated in space by the aberration signals.

In the actual experiment, certain error sources must be checked by maneuvers of the satellite. The mission lifetime should therefore be not less than 6 t_0', or about 4 months. An experiment lasting 1 year has the advantage of following

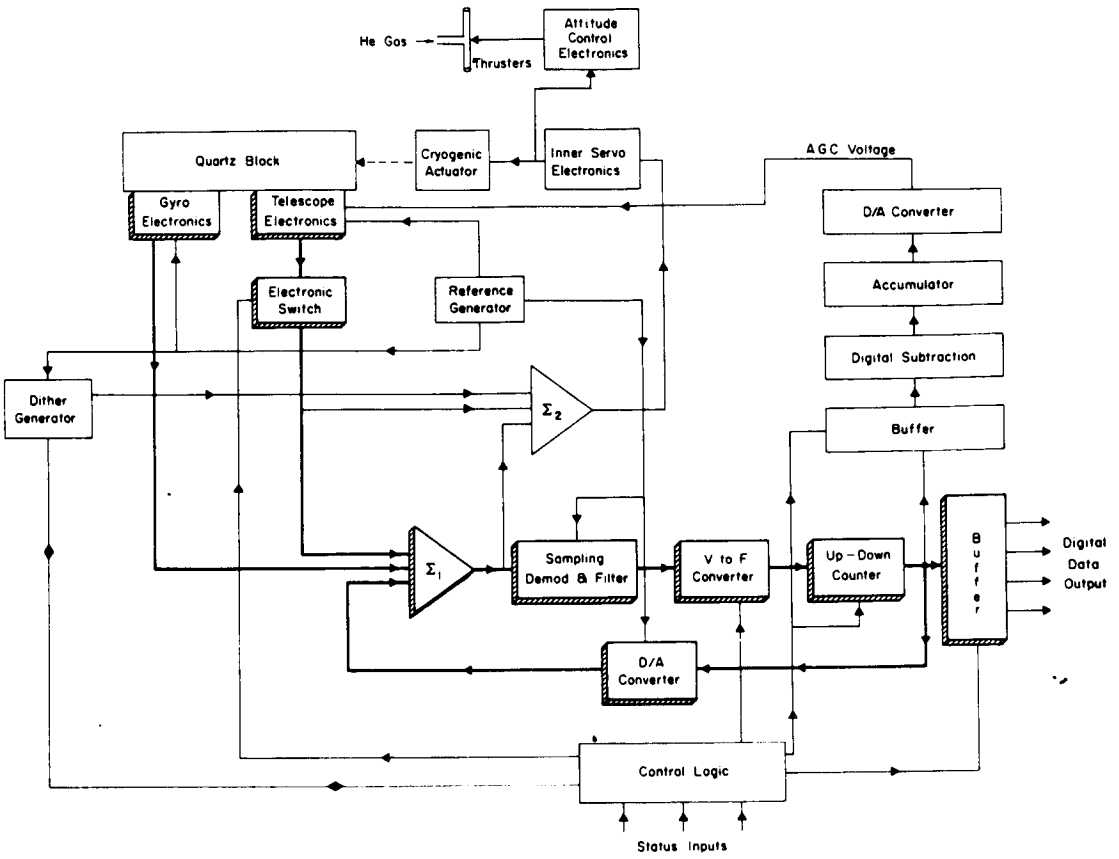


Fig. 5. Instrumentation system

the complete cycle of annual aberration besides permitting slight additional averaging for particular error torques. The simplest experimental configuration consists of following an exact polar orbit, and having one gyroscope pair spinning parallel to the earth's axis and the other pair parallel to the orbit axis, with the telescope pointing at a star in the plane of the earth's equator. The terms Ω_G and Ω_M are then completely separated. An alternative is to make the spin axes parallel to the telescope axis, in which case Ω_G and Ω_M appear concurrently in the two readout planes of each gyroscope. The second arrangement has the advantage of allowing the signs of some of the errors to be reversed periodically by rolling the satellite about the telescope axis. On the other hand, it requires two full accuracy readouts for each gyroscope and does not give such direct separation of Ω_G and Ω_M if the orbit is slightly non-polar.

If the orbit is highly non-polar the apsidal precession becomes relatively rapid, being about 5 deg per day, or one revolution in 10 weeks, for a 45-deg orbit. A gyroscope spinning parallel to the earth's axis will see a geodetic precession of sinusoidal form with an amplitude of about ±0.5 arc-s and a period of 10 weeks, together with orbital aberration signals amplitude-modulated at the same frequency, and a component of annual aberration. A gyroscope perpendicular to the earth's axis will see superimposed sinusoidal and linear geodetic terms, a linear motional term, annual aberration, and an amplitude-modulated orbital aberration 90 deg out of phase with the first term. Both gyroscopes will also experience gravity gradient and suspension errors larger than those in polar orbits, which are also modulated at apsidal period but with different phase and scale relations from Ω_G and Ω_M. The form of the data is therefore much more complex for non-polar orbits, but since the problem of resolving a sinusoidal signal of known period and phase in the presence of noise is essentially equivalent to the problem of resolving a linearly increasing signal, in principle, the data may be recovered with precision comparable to that for a polar orbit. More detailed studies are necessary before a final decision is reached on the choice of orbit.

Corrections to Ω_G and Ω_M for non-circular orbits, including the effects of higher-order mass moments from the earth, have been derived by R. F. O'Connell (Ref. 11) and by D. C. Wilkins (Ref. 12). The higher mass moments affect both the form of the orbit and the structure of the relativity terms themselves. Wilkins has established limits on the knowledge of orbit parameters needed to recover data with the desired accuracy. A precision better than 1 part in 10,000 is obtained by integrating the equations explicitly for a single orbit and then piecewise fitting of successive orbits in data analysis.

There remains the problem of proper motion. In a polar orbit, the linear component of the proper motion will cause displacement in each axis indistinguishable from relativity data. In non-polar orbits, the error disappears from the

sinusoidally modulated terms that measure Ω_G but not from Ω_M. The magnitude of the motion varies from star to star, but in many instances, the uncertainties in its value definitely exceed the design goal of determining Ω_M to 1 part in 50. Ultimately, it can be eliminated by flying two satellites in orbits of different radii to give different values for Ω_M. Meanwhile, studies are still under way to determine which star will give the least uncertainty in reference and, at the same time, be bright enough to fulfill the requirements on the attitude control system. On balance, the experimenters will probably count themselves fortunate if proper motion proves to be the greatest difficulty of the gyroscope program.

Acknowledgments

All aspects of the experiment have been discussed with W. M. Fairbank, to whom I am indebted in many ways. In addition to colleagues mentioned in the text, special thanks are due to Dr. Robert H. Cannon.

References

1. Schiff, L. I., Proc. Nat. Acad. Sci., 46, 871, 1960.

2. Eddington, A. S., Mathematical Theory of Relativity, p. 99, Cambridge University Press, 1962; O'Connell, R. F., Phys. Rev. Letters, 20, 69, 1968; Barker, B. M., and O'Connell, R. F., ibid., 25, 1511, 1970; Wilkins, D. C., Ann. of Phys., 61, 2, 277, 1970.

3. Thorne, K. S., and Will, C. M., Astrophys. J., 163, 595, 1971.

4. Pustovoit, V. I., and Bautin, A. V., JETP, 46, 1386, 1964; Wilkins, D. C., private communication. The numerical predictions for a 500-mile orbit are: Birkhoff +4.3, Belifante-Swihart -(2.3 to 3.8), Whitrow-Morduch, +1.1 arc-s/y.

5. Selzer, P., Fairbank, W. M., and Everitt, C. W. F., Adv. Cry. Eng., 16, 1971.

6. Bracken, T. D., and Everitt, C. W. F., Adv. Cry. Eng., 13, 168, 1968.

7. Opfer, J. E., Rev. de Phys. Appliqué, 5, 37, 1970.

8. DeBra, D. B., "Vectors and Dyadics: The Natural Tools of Space-Vehicle Mechanics," American Astronautical Society Meeting, San Francisco, Calif., Aug. 1, 1961.

9. Maxwell, J. C., Treatise on Electricity and Magnetism, Second Edition, Vol. 1, Sec. 141c, 1879.

10. Maxwell, J. C., Reports of British Association, 33, 172, correction E, 1864.

11. O'Connell, R. F., this proceedings.

12. Wilkins, personal communication.

Gyroscope Test of Gravitation: An Analysis of the Important Perturbations

R. F. O'Connell
Louisiana State University

We discuss two important perturbations due to (1) the earth's quadrupole moment (Refs. 1 and 2) (say Q) and (2) the earth's revolution around the sun (Ref. 3), which must be taken into account in an analysis of Schiff's (Ref. 4) proposed gyroscope test of gravitation. The rate of change of the spin axis $\vec{S}$ of a gyroscope may be written (Ref. 4)

$$\frac{d\vec{S}}{dt} = \vec{\Omega} \times \vec{S} \tag{1}$$

where $\vec{\Omega}$ is the angular velocity of precession.

Explicitly, in Einstein theory ($\vec{\Omega} \equiv \vec{\Omega}_E$), we have (Refs. 1-4)

$$\vec{\Omega}_E = \vec{\Omega}_T + \vec{\Omega}_{DS} + \vec{\Omega}_{LT} + \vec{\Omega}_Q + \vec{\Omega}^S_{DS} \tag{2}$$

where $\vec{\Omega}_T$, $\vec{\Omega}_{DS}$, $\vec{\Omega}_{LT}$, $\vec{\Omega}_Q$, and $\vec{\Omega}^S_{DS}$ are the so-called Thomas, de Sitter, Lense-Thirring, quadrupole-moment (Refs. 1 and 2) and sun (Ref. 3) contributions, respectively. The corresponding quantity in the Brans-Dicke (BD) scalar-tensor theory (Ref. 5), say $\vec{\Omega}_{BD}$, is given by (Ref. 6)

$$\vec{\Omega}_{BD} = \vec{\Omega}_T + \frac{4 + 3\omega}{6 + 3\omega}\left(\vec{\Omega}_{DS} + \vec{\Omega}_Q + \vec{\Omega}^S_{DS}\right) + \frac{3 + 2\omega}{4 + 2\omega}\vec{\Omega}_{LT} \tag{3}$$

where ω is the dimensionless coupling constant of the BD theory. It is possible to have $\vec{\Omega}_T$ essentially zero (Ref. 4) by putting the gyroscope in a satellite. Henceforth, we will regard the $\vec{\Omega}$ as being averaged over a period of the motion and confine our discussion to orbits which would be circular in the absence of the earth's quadrupole moment Q (although elliptic orbits have also been treated: Ref. 2). The term $\vec{\Omega}_Q$ is referred to as the direct quadrupole moment effect, but there is also an indirect effect which manifests itself only when the principal term $\vec{\Omega}_{DS}$ is averaged over a period of the motion (Ref. 1).

We define a length "a" such that Kepler's law holds in its normal form for the distorted (due to Q) circular orbit; i.e.,

$$\omega = \frac{2\pi}{T} = \left(\frac{GM}{a^3}\right)^{1/2} \tag{3a}$$

where T is the period and ω is the average angular velocity (Ref. 2) of the gyroscope in the field of the earth of mass M, spin angular momentum $\vec{S}$(2), and quadrupole moment Q, given by (Ref. 7)

$$2Q = (1082.64 \pm 0.08) \times 10^{-6} R^2 \tag{4}$$

where R is the earth's equatorial radius. It follows (Refs. 1-4) that, to the accuracy required,

$$\vec{\Omega}_{DS} = \frac{3GM\omega}{2c^2 a}\left(1 + k\frac{Q}{a^2}\right)\vec{n} \qquad (5)$$

$$\vec{\Omega}_{LT} = \frac{GS^{(2)}}{2c^2 a^3}\left[\vec{n}^{(2)} - 3\cos\theta\,\vec{n}\right] \qquad (6)$$

$$\vec{\Omega}_{Q} = \frac{3GM\omega}{2c^2 a}\left(\frac{3Q}{a^2}\right)\left[\frac{1}{2}(5\cos^2\theta - 1)\vec{n} - \cos\theta\,\vec{n}^{(2)}\right] \qquad (7)$$

$$\vec{\Omega}^{S}_{DS} = \frac{3GM_{\odot}\,\omega_E}{2c^2 r}\,\vec{n}^{(S)} \qquad (8)$$

where k is a number of order unity which depends on the inclination of the orbit (Ref. 2) (it equals 0.5 for a polar orbit and -1.0 for an equatorial orbit), $\vec{n}$ and $\vec{n}^{(2)}$ are unit vectors along the orbital angular momentum of the gyro and $\vec{S}^{(2)}$ directions, respectively, and θ is the angle between the $\vec{n}$ and $\vec{n}^{(2)}$ directions. In addition, $M_{\odot}$, ω_E, and $\vec{n}^{(S)}$ denote the mass of the sun, the average angular velocity of the earth around the sun, and the unit vector along the earth's orbital angular momentum; r is the earth-sun distance.

Perhaps the most unique feature of Schiff's gyro test is that it is the only experiment thus far proposed which is likely to measure the off-diagonal Lense-Thirring terms in the metric tensor. Everitt and Fairbank (Ref. 8) and Fairbank (Ref. 9) expect to carry out this experiment in the near future by launching a satellite containing two pairs of superconducting gyroscopes into a polar orbit around the earth: the spin of one pair (gyro 1) will be parallel to the earth's axis and the spin of the other pair (gyro 2) will be perpendicular to the plane of the orbit.

A polar orbit ($\theta = \pi/2$) was selected because (from Eqs. 5 and 6) $\vec{\Omega}_{DS}$ and $\vec{\Omega}_{LT}$ are at right angles for such an orbit (and, in addition, precession of the gyro due to a possible gyro quadrupole moment is zero; Ref. 10). For definiteness, consider the earth's angular velocity to be in the z-direction and the polar orbit to be in the xz-plane so that the orbital angular momentum of the satellite points in the y-direction. Then $\vec{\Omega}_{DS}$ lies along y and $\vec{\Omega}_{LT}$ along z. Thus, gyro 1 (with spin along z) will not be affected by $\vec{\Omega}_{LT}$, and gyro 2 (with spin along y) will not be affected by $\vec{\Omega}_{DS}$. Therefore, it was thought that gyro 2 would provide a "clean" test of the Lense-Thirring terms -- clean in the sense of being sensitive to $\vec{\Omega}_{LT}$ only. However, as we now make clear, this possibility is ruled out because of the sensitivity of gyro 2 to the sun's perturbation.

The magnitude* of $\vec{\Omega}_{LT}$ for a satellite in a circular polar orbit 300 miles above the earth is 43.8×10^{-3} sec/yr (at this altitude the magnitude of $\vec{\Omega}_{DS}$ is the oft-quoted value of 7.0 sec/yr). Using BD theory, this value is reduced (Ref. 6) by a factor of 1/16, i.e., by 2.7×10^{-3} sec/yr. As before (Ref. 6), we take $\omega = 6$. Thus, to distinguish between the Einstein and BD theories, the experiment should be capable of measuring such small precession angles. In fact, measurement accurate to 10^{-3} sec/yr will be possible (Ref. 9).

The question we wish to consider is whether there are any perturbations of magnitude greater than 10^{-3} sec/yr along the z-axis, in addition to $\vec{\Omega}_{LT}$. With regard to $\vec{\Omega}_{Q}$, as is clear from Eq. (7), this contribution turns out to be in the same direction as $\vec{\Omega}_{DS}$ for a polar orbit (though this is not true in general), and thus it has no component along z. However, the magnitude** of $\vec{\Omega}^{S}_{DS}$ is 19.2×10^{-3} sec/yr, and since the earth's equator is inclined at an angle ϕ of 23.44 deg to the ecliptic, the z-component is 0.917 $\vec{\Omega}^{S}_{DS}$, i.e., 17.6×10^{-3} sec/yr. It is more than 6 times as large as the difference between the Lense-Thirring contributions arising from the Einstein and BD theories and 17.6 times larger than what can be measured! Actually, the best orbit to select for the purpose of obtaining the "cleanest" test of $\vec{\Omega}_{LT}$ is one inclined at an angle θ to the equator so that $\vec{n}$ and $\vec{n}^{(S)}$ point in the same direction (i.e., $\theta = \phi$). Gyro 2 is again placed in the $\vec{n}$ direction so that it will now be insensitive to both $\vec{\Omega}_{DS}$ and $\vec{\Omega}^{S}_{DS}$. Unfortunately, a price must be paid; that is, the Lense-Thirring contribution to the precession of gyro 2 is reduced by a factor $\sin\phi = 0.398$ to a value 17.4×10^{-3} sec/yr, and the difference between the Einstein and BD contributions in this case is 1.1×10^{-3}, very close to the limits of the expected experimental accuracy. In addition, there is now a contribution from $\vec{\Omega}_{Q}$ amounting to 3.6×10^{-3} sec/yr. For this configuration, similar to the polar orbit case, a possible gyro quadrupole moment does not contribute to the precession.

To summarize, we wish to emphasize that, although the perturbations discussed will certainly make the analysis of the observations more complex, they have all been calculated precisely to the accuracy desired, with the result that the experiment should be capable of deciphering each separate contribution to the angular velocity of precession of the spin.

References

1. O'Connell, R. F., Astrophys. Space Sci., 4, 199, 1969; O'Connell, R. F., Lett. Nuovo Cimento, 1, 933, 1969; Barker, B. M., and O'Connell, R. F., Nuovo Cimento Lett., 4, 561, 1970.

*The value of $\vec{\Omega}_{LT}$ is usually calculated by assuming that the earth is homogeneous in density. However, for our present purposes, it is necessary to use a value for the angular momentum of the earth which is based on a more realistic mass distribution for the earth (Ref. 11). This is a significant modification because it reduces $\vec{\Omega}_{LT}$ by about 17%, which is considerably larger than the approximately 6% difference between the Einstein and Brans-Dicke predictions, the quantity which the experiment proposes to measure.

**From Eq. (76a), Ref. 2, it follows that $\vec{\Omega}^{S}_{DS}$ is simply one-half of the perihelion precession of the earth around the sun.

2. Barker, B. M., and O'Connell, R. F., Phys. Rev., D2, 1428, 1970.

3. Barker, B. M., and O'Connell, R. F., Phys. Rev. Lett., 25, 1511, 1970.

4. Schiff, L. I., Proc. Nat. Acad. Sci. U.S., 46, 871, 1960.

5. Brans, C., and Dicke, R. H., Phys. Rev., 124, 925, 1961.

6. O'Connell, R. F., Phys. Rev. Lett., 20, 69, 1968.

7. Smith, D. W., Planet. Spac. Sci., 13, 1151, 1965.

8. Everitt, C. W. F., and Fairbank, W. M., in Proceedings of the Tenth International Conference on Low Temperature Physics, Moscow, U.S.S.R., 1966, Vol. 2B, p. 337, Viniti, Moscow, 1967.

9. Fairbank, W. M., London Award Lecture, in Proceedings of the Eleventh International Conference on Low Temperature Physics, St. Andrews, Scotland, 1968, Vol. 1, pp. 14 and 15, Edited by J. F. Allen, D. M. Finlayson, and D. M. McCall, St. Andrews University, St. Andrews, Scotland, 1969.

10. Everitt, C. W. F., Fairbank, M., and Schiff, L. I., "Theoretical Background and Present Status of the Standard Relativity-Gyroscope Experiment," preprint.

11. Allen, C. W., Astrophysical Quantities, 2nd Edition, p. 108, Athlone Press, London, 1963.

4

Some Future Applications of Low-Temperature Technology in Space

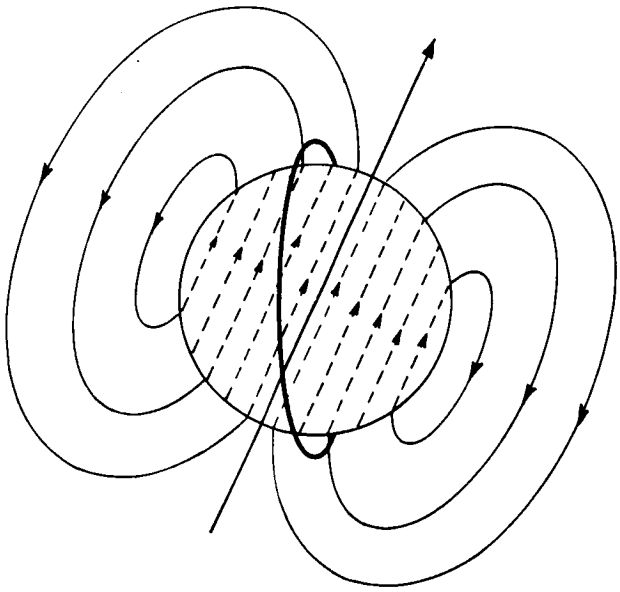
William M. Fairbank
Stanford University

The gyro-relativity experiment suggested by Leonard Schiff (Ref. 1), and the gravitational radiation experiment being developed jointly at Louisiana State University and Stanford using radiation detectors cooled to a few millidegrees in temperature (Ref. 2) are discussed elsewhere in this proceedings. These experiments are part of a low-temperature program begun at Stanford to perform fundamental experiments in physics, many of which require new and sometimes large-scale low temperature techniques. In this paper, I wish to discuss a number of other possible applications of low-temperature technology in space which we at Stanford, and other scientific colleagues elsewhere, have been thinking about.

Let me begin by reviewing some particular ways in which low temperatures play a crucial role in the relativity experiment described by Francis Everitt (Ref. 1). It appears desirable to perform the experiment with an accuracy approaching 10^{-3} arc-s per year. For this purpose, the drift error of the gyroscope due to applied torques, the error of the gyro readout, and mechanical alignment between the gyro housing and the reference telescope must all be reduced below the 10^{-3} arc-s level. Low-temperature techniques provide an enormous advantage in achieving this accuracy. Operation of the telescope at liquid helium temperatures eliminates errors due to thermal distortion, the use of a superconductor provides a perfect magnetic shield for the gyroscope, and the London moment in the spinning superconducting sphere supplies an entirely new kind of gyro readout. The detailed calculations summarized by Everitt show that a gyroscope with the required accuracy can be obtained by the use of a highly homogeneous sphere spinning in a vacuum under conditions such that electric and magnetic torques have been reduced to extremely low levels. The requirement of nearly perfect sphericity poses a readout problem. Conventional readouts require knowing the position of the axis of rotation with respect to the ball. If the moments of inertia of all the axes of the ball are the same, it is not possible to anticipate about which axis the gyroscope will spin. Furthermore, if the ball is to be kept at 1.2°K in a vacuum, it can be cooled by black-body radiation, and effectively no heat can be allowed to fall on the ball in the process of reading it out. These factors eliminate conventional readouts.

Superconductivity leads to a unique solution of the readout problem. A superconductor spun up in a zero field develops along its axis of spin a uniform magnetization of $10^{-7}\ \omega$ G (Refs. 3, 4) as a result of macroscopic flux quantization. With a gyroscope spinning at an ω of 2×10^{3} rad/s, this would give 2×10^{-4} G along the axis of spin. The question then arises how one can detect to 0.001 arc-s the orientation of such a gyroscope by use of this very small magnetic field.

Figure 1 illustrates the proposed readout, which makes use of a superconducting loop. Shown in the figure is a spinning superconductor with a magnetic field as indicated along the axis of the spin. Around the spinning sphere is placed, as a method of readout, a superconducting loop. Since the resistance of the superconducting loop is zero, any change in the flux through the loop caused by a change in orientation of the gyro sphere will cause a current to flow in the loop which exactly cancels this change in flux. If one could read out this current, one could determine the change in orientation of the direction of the ball.



LONDON MOMENT FIELD $H = 10^{-7}\omega$ GAUSS

Fig. 1. Principle of London moment readout for relativity gyroscope

Figure 2 shows the method we have developed to read out this current. In series with the first loop is placed a second superconducting loop indicated with an arrow through the loop. The current that flows in the two superconducting loops produces a canceling flux, which is distributed in the two loops instead of being confined to one. The ratio of the flux in the two loops is equal to the ratio of the inductances of the loops. Thus, the change in flux through the first loop caused by the reorientation of the ball produces a canceling flux distribution in the two loops. If the inductance of the second loop is changed, the current flowing in the two loops changes and the distribution of the canceling flux in them changes. If the inductance is changed 10^5 times/s, then a 10^5-Hz ac signal is produced which can be detected by a readout coil. Development of the practical design is due to J. E. Opfer (Ref. 5). The modulator consists of a long superconducting lead evaporated on a flat surface. Adjacent to this long superconducting wire is a superconducting ground plane evaporated onto a quartz crystal. The crystal and surface are placed about 2000Å apart, and the crystal is driven such that the ground plane periodically approaches and recedes from the superconducting circuit. This modulates the inductance of the circuit and causes the flux to be pumped back and forth between the two loops. The oscillating current in the two loops flows through the coil as indicated and is read out through a transformer by an amplifier. It is possible to increase the sensitivity of this circuit by placing a condenser in the circuit, as indicated on the diagram. The modulating current flows in and out of the condenser plates in such a way as to provide additional parametric amplification.

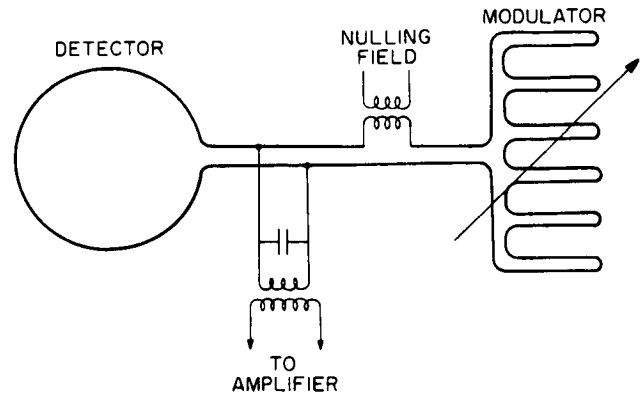


Fig. 2. Vibration plane magnetometer

John Pierce (Ref. 6) has worked out in detail the sensitivity of such a circuit compared with the theoretical Johnson noise in an amplifie :

$$\Phi^2 \geq \frac{27\pi}{16}\frac{L}{Q}\left(\frac{L}{\Delta L}\right)^2 kT\,\frac{\Delta\nu}{\nu}$$

where L is the total inductance of the circuit, ΔL the inductance change, T the noise temperature of the circuit, Q the quality factor of the circuit, ν the frequency of modulation, and $\Delta\nu$ the bandwidth. We have verified the validity of this equation both with experiments and by model circuits on an analog computer. With the modulating crystal operating at 10^5 Hz and a Q of 1000 with room temperature amplifier noise, the sensitivity is predicted to be 10^{-10} G, which would allow an accuracy of readout of 0.1 arc-s in 0.1 s of time. The desired sensitivity of 0.001 arc-s is then obtained by integrating over longer periods of time. In the above circuit, feedback is used to keep the signal as near as possible to zero. Thus, if all the pickup is eliminated from the vibrating crystal because of the quantized nature of the zero in the superconducting loop, there should be no drift in this circuit over the course of a year. A Josephson junction magnetometer might also be used for the gyro readout providing the zero drift can be kept sufficiently small. Tests are being made to determine zero drift in both the vibrating plane and Josephson junction magnetometers.

In order to eliminate magnetic errors, the gyroscope must be kept in a magnetic field region of less than 10^{-7} G. Blas Cabrera and William Hamilton have been experimenting at Stanford and L. S. U. with the possibility of eliminating the last quantum of trapped flux from a superconducting shield and have discovered that, if a lead shield is cooled from one end very slowly by heat flushing technique in a low magnetic field, the trapped flux can be reduced to less than one quantum of flux every 20 cm^2 to give a magnetic field of less than 10^{-8} G. It appears possible that the last quantum of flux may ultimately be eliminated. Another technique which helps in reducing magnetic fields is expanding a folded superconducting balloon, since the constant flux condition makes the field inversely proportional to the cross-section of the balloon. Figure 3 illustrates a large annular dewar being constructed at Stanford, which may contain either fixed or expansible superconducting balloons.

The gyro-relativity experiment will not only play a role in checking Einstein's General Theory of Relativity but we believe that it can be the prototype of other astronomical experiments. A

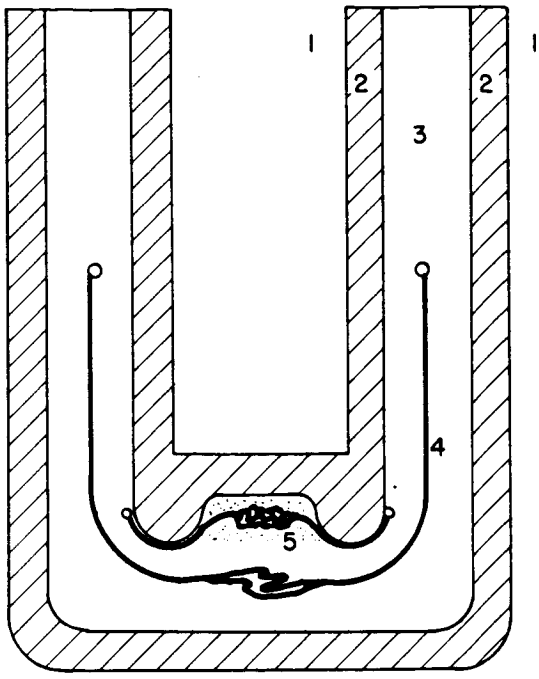


Fig. 3. Large-scale low-magnetic-field facility; (1) room-temperature region, (2) insulation, (3) helium-temperature region, (4) partially expanded bladder, (5) second bladder before expansion

gyroscope which can be read out to the order of a 0.001 arc-s per year and whose drift at this level is determined only by the predictable effects of general relativity can serve as an absolute reference of astronomical telescopes. At the present time, one star is referenced with respect to a second star. When the relativity experiment was first discussed, Schiff started with the consideration of whether a gyroscope could be built perfectly enough to determine the position of fixed stars. This appeared hopeless at the time, but the discussion at this meeting shows that one of the limitations of the relativity experiment is the proper motion of the fixed stars. So we have come full swing. The experiment has arrived at the potential accuracy with which the gyroscope can in fact check the proper motion of fixed stars. Indeed, we can speculate that the experiment might be used to investigate whether a planet the size of the earth exists around any of the nearer stars. If we take a star the size of our sun 4 light years away, a planet of terrestrial size would cause a proper motion of about 5×10^{-6} arc-s over a year. It is conceivable that the accuracy of such a gyroscope could be improved to the point that a periodic proper motion of this kind could be detected.

With large-scale helium environments in space, telescopes of extraordinary mechanical stability become possible, since cooling to helium temperatures eliminates thermal distortion, while the reduction in gravitational acceleration eliminates mechanical distortion. A further beautiful advantage of having a telescope at low temperatures is that noise in many kinds of optical detectors may be greatly reduced, with a consequent improvement in sensitivity. This is particularly true of infrared detectors, and several experimenters have expressed interest in using a liquid helium dewar in space to observe the 3°K black-body radiation. In addition to the germanium detectors commonly used, there are some intriguing possibilities which we are studying at Stanford of developing an infrared detector capable of observing single photons by means of Josephson junction devices.

Other potential applications for helium in space include the use of superconducting magnets to produce large magnetic fields. One such application already planned is in the cosmic ray experiment designed by Alvarez, Buffington, Smith and their collaborators (Ref. 7) for the second High Energy Astronomical Observatory in Space (HEAO). Yet another intriguing future application which we have been considering for a satellite with a cryogenic reference telescope identical to that used for the gyro relativity experiment, consists in a development of an experiment started at Stanford by Hamilton to measure the electric dipole moment of He^3 (Ref. 8). This experiment involves polarizing He^3 nuclei and placing them at the center of a very spherical superconducting shield from which the last quantum of flux has been eliminated. The He^3 nuclei can be thought of as electrically supported gyroscopes with a magnetic readout analogous to that of the relativity gyroscope. If time reversal invariance were valid, the nuclei would be symmetrical with respect to displacement of positive and negative charges. Hence, any precession arising when an electric field is applied at right angles to the axis of magnetization implies the presence of an electric dipole moment and a violation of time reversal invariance. On earth, the accuracy of the experiment is limited by the residual magnetic fields in the shield and by the precision of the inertial reference. Assuming that the magnetic field can in fact be made vanishingly small by the quantized flux condition, the accuracy is improved by about three orders of magnitude by placing the experiment in a satellite, in which case it can be used to check time reversal invariance in the superweak interaction.

All such experiments require a method for controlling the flow of helium from a dewar in the low residual acceleration environment of space. During the course of research on the gyro relativity experiment, Selzer, Fairbank, and Everitt (Ref. 9) have developed a promising technique for doing so which makes use of a porous plug of tightly rolled aluminum foil. Helium flows through and evaporates at the outer surface, from which the refrigeration is carried into the dewar by the high thermal conductivity of the plug. With appropriate choice of design parameters, the plug may be made to operate either in the normal or superfluid condition as an automatic regulator, removing the necessary liquid to overcome the heat leak without allowing excess liquid to flow out of the dewar. The plug has been tested in the laboratory at Stanford in a dewar which can be turned upside down with no change in the rate of helium boiloff. The first flight test planned in the gyro ralativity experiment program is a test of this superfluid plug in space.

As a further illustration of the kind of improvements in signal to noise performance obtained by low temperature techniques, I should like to refer once again to the Stanford-Louisiana State

University gravity wave experiment described by W. O. Hamilton. The limit on the experiment is the thermal motion in the detecting massive bar. Since the bar must be resonant at a particular frequency, for example 1600 cycles, it is not possible to increase the mass beyond the present size of Weber's bar. Therefore, the sensitivity must be increased by decreasing the background noise in the bar. If the temperature is reduced to 0.003°K, the background noise is reduced by a factor of 10^5. It seems possible to further increase the sensitivity by a better coupling to the detector. If a net increase in sensitivity of 10^6 is obtained, then it would be possible to see gravitational radiation signals with the same detection amplitude from sources 10^3 times farther in distance, since the gravitational radiation falls off as $1/r^2$. This would mean that gravitational radiation signals from a thousand galaxies could be seen with the same sensitivity as the room-temperature detector sees from the center of our galaxy. It seems very important in the ultimate understanding of gravitational radiation to be able to look at distant galaxies. A possibility of such large-scale low-temperature cooling of a massive bar is provided by the extensive large-scale low-temperature engineering that has gone into the development of a superconducting accelerator at Stanford.

In summary, space has made it possible for us to escape from the disturbing effects of our atmosphere and the further disturbing effects of the gravitational field we live in. The zero-g environment and the vacuum of space offer an unprecedented reduction in the errors of both telescopes and gyroscopes. When these advantages are combined with low-noise devices and the ideal magnetic shielding that can be obtained from the cryogenic environment, a new dimension has been added to the excitement of space physics.

Acknowledgments

This paper was prepared in collaboration with C. W. F. Everitt. I am grateful to colleagues at Stanford and elsewhere for many discussions, especially to W. O. Hamilton, J. A. Lipa, J. M. J. Madey, J. E. Opfer, H. A. Schwettman, F. C. Witteborn. Above all, I remember the extraordinary, unselfish help and interest supplied on so many occasions by the late Leonard Schiff.

References

1. Everitt, C. W. F., this proceedings.

2. Hamilton, W. O., this proceedings.

3. London, F., Superfluids, Vol. 1, p. 78f, John Wiley and Sons, New York, 1950; and Dover Publications, New York, 1961.

4. Hildebrandt, A. F., Phys. Rev. Letters. 12, 190, 1964; Bol, M., and Fairbank, W. M., Proc. IX Int. Conf. on Low Temp. Phys., p. 471, Plenum Press, 1965; King, A., Jr., Hendricks, J. B., and Rorschach, H. E., Jr., ibid., p. 471.

5. Opfer, J. E., Rev. de Phys. Appliqué, 5, 37, 1970.

6. Pierce, J. M., Proc. of Symposium on Superconducting Devices, University of Virginia, April 28-29, 1967.

7. Alvarez, L., Buffington, A., and Smith, L. H., SSL, 374, University of California Space Sciences Laboratory, University of California, Berkeley, May 29, 1970.

8. Fairbank, W. M., Hamilton, W. O., and Everitt, C. W. F., Proc. OAR Research App. Conf., p. 153, Washington, 1966.

9. Selzer, P., Fairbank, W. M., and Everitt, C. W. F., Adv. Cryo. Eng., 16, 1970.

Light Deflection*

Henry A. Hill
Wesleyan University
University of Arizona

I. Introduction

A major effort has been concentrated on the design, fabrication, and development of a telescope which constitutes the primary piece of instrumentation in an experiment to obtain an improved measurement of the gravitational deflection of light. The method employed is that of using a highly accurate measurement of the diameter of the sun to establish a scale of distances on the sky, and measuring the separation between two stars as a function of time while the sun appears to move on the celestial sphere; the experiment is to be carried out in full sunlight, obviating the necessity of waiting for a total eclipse of the sun. Positions of stars and the sun are to be determined photoelectrically, and compensation for first order differential atmospheric refraction, atmospheric dispersion, and changes in scale in the focal plane is an integral part of the apparatus. An analysis of the characteristics of this instrument, based in part on experience gained with a telescope designed to measure the solar oblateness, indicates that an improvement in accuracy in excess of a factor of 10 over that of the earlier photographic deflection measurements can be achieved by the use of this instrument. In addition to yielding a value for the deflection of light at the sun's limb, these measurements would provide the form of the deflection curve as a function of distance from the center of the sun. Furthermore, it affords an alternate way of carrying out a solar oblateness experiment to check the Princeton results (Ref. 1).

By virtue of the method to be employed in this measurement, the data acquired on star positions relative to the sun would contain much information in addition to the deflection of light. In particular, since the instrument is designed to include an atomic time standard, the measurements would yield an accurate value for the instantaneous right ascension and declination of the sun, that is, the projection of the earth-sun line on the celestial sphere. Since the accuracy of these instantaneous coordinates can be expected to exceed that accessible by classical astrometric methods, three categories of results can be obtained:

(1) A comparison of atomic and astronomical time would be a direct consequence of comparing the apparent motion of the sun to the signals from the atomic clock. Using these data to obtain the earth's period of rotation around the sun to an improved accuracy for a period of 10 years would provide evidence related to possible changes in the universal constant of gravitation.

(2) The coordinates of the plane of the earth's motion, the fundamental inertial system of astronomy, would be established.

(3) The precession of the perihelion of the earth's orbit in space would be determined to sufficient accuracy that the relativistic contribution to this effect may

*This work supported in part by National Science Foundation grant GP9295.

be separated from that produced by a solar oblateness, since the radial dependence of the perihelion motion due to a solar oblateness differs from that of purely relativistic origin (Ref. 2).

These measurements are of fundamental importance to both physics and astronomy, in that they offer a more accurate test of the predictions of a fundamental physical theory as well as providing basic astronomical data. Applications of this instrument to work in the night sky would increase the accuracy of our knowledge of the positions of stars relative to one another, as well as their parallaxes and proper motions. Improved parallaxes of nearby stars and new parallaxes for more distant stars are crucial to astrophysical theory, in that the stars of known trigonometric parallax are the sole means for calibrating the absolute luminosities and bolometric corrections of the stars (Ref. 3); the extension of parallax measurements to a sphere of greater radius than is now available extends not only the number of calibration stars, but also the range of spectral type and luminosity class included in the calibration. More accurate proper motions, on the other hand, would improve the knowledge of the true space motions of stars enabling more accurate determination of galactic rotation, velocity dispersion as a function of spectral type, and dynamical characteristics of stars in clusters.

The operation which has developed since the beginning of the program in August of 1963 is presently known as SCLERA, an acronym for the Santa Catalina Laboratory for Experimental Relativity by Astrometrics. SCLERA is an off-campus research facility located in the Santa Catalina Mountains north of Tucson, Arizona.

II. Telescope

The telescope is a modified Schupmann telescope (Ref. 4), maintained vertical, fed by an elevation-azimuth mirror system and evacuated. Basically, the instrument is a long focal length refractor inside a tube which can be evacuated. The telescope is maintained in a tower with the top of the tower, i.e., the dome, rotatable to follow the sun's diurnal rotation. Figure 1 illustrates a cut-away view of the telescope tube, the tower and instrument room.

The objective is a f-100 singlet and has a 1200 cm focal length. The site elevation is 8500 ft (2586 m) and the site latitude is 32°25'N.

In an attempt to obtain a value for the light deflection to better than 1% a careful analysis has been made of the problems: optical aberrations, apodization of the objective, atmospheric refraction and dispersion, longitudinal chromatic aberration, scattered light from mirrors and lenses, star tracking near the sun, and the solar edge definition. The results of these analyses can be found in the following theses and/or papers:

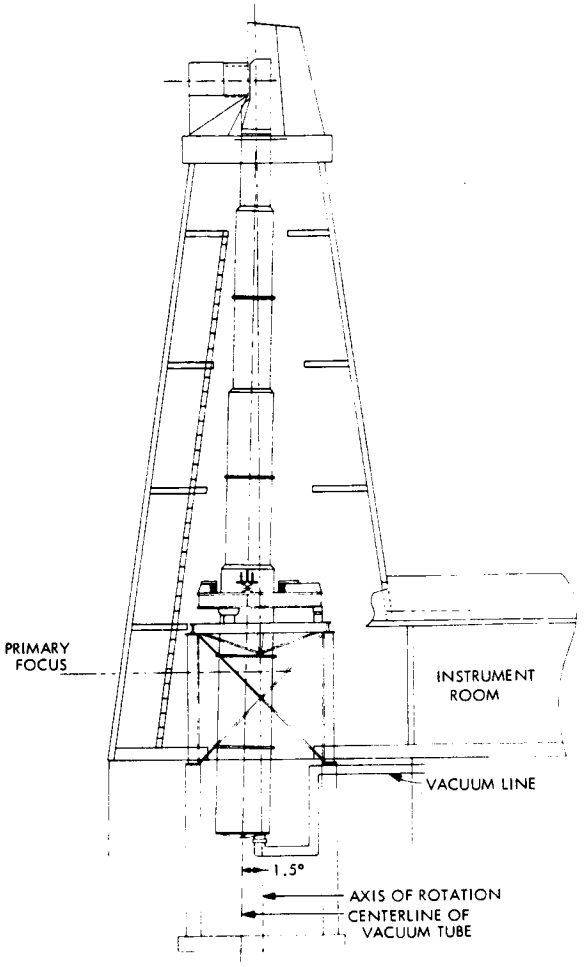


Fig. 1. Cut-away view of instrument

(1) Development of Daytime Astrometry to Measure the Gravitational Deflection of Light, C. A. Zanoni, Ph. D. Thesis, Princeton University (1966).

(2) Reduction of Diffracted Light for Astrometry Near the Sun, C. A. Zanoni and H. A. Hill, JOSA 55, 1608, 1965.

(3) Correction of Lateral Color Aberration Produced by the Atmosphere, H. A. Hill and C. A. Zanoni, JOSA 56, 1655, 1966.

(4) Measurements of Scattered Light from Mirrors and Lenses, G. R. Hostetter, D. L. Patz, H. A. Hill, and C. A. Zanoni, Applied Optics 7, 1383, 1968.

(5) The Fourier Coefficient Technique for Defining the Edge of the Sun, R. T. Stebbins, Senior Thesis, Wesleyan University, 1970.

(6) SCLERA Progress Report, The Mangin Design Section, University of Arizona 1970.

III. Current Projects

The efforts at SCLERA are presently directed towards five experiments with several near completion and one in its early stage of development.

These experiments are:

(1) Measurement of the Solar Oblateness, P. Clayton, Ph. D. Thesis, University of Arizona.

(2) Anisotropy of the Solar Limb Darkening, D. Patz, Ph. D. Thesis, University of Arizona.

(3) Gravitational Deflection of Starlight, B. Cardon, Ph. D. Thesis, Wesleyan University.

(4) G/G, A. Healy, Ph. D. Thesis, University of Arizona.

(5) Parallax of Spectroscopic Binary, preliminary work by Ya Yue Dunker, Wesleyan University.

The supervision of these theses in my absence, as I spend only 30% of my time at SCLERA, is by Dr. James Oleson, a Wesleyan University Research Associate.

The first measurements made at SCLERA were on the solar oblateness in the fall and winter of 1968-69. The reason that these were chosen as the first measurements rather than the light deflection itself is because the ability to measure a solar diameter is a fundamental part of the light deflection experiment. Also, the oblateness experiment is important in its own right.

One immediate observation that could be drawn from the 1968-69 data was that the signal-to-noise was much better than that of Princeton. However, it soon became apparent that there was a systematic error in the measurement of different solar diameters of the order of $\sim 1/10$ sec of arc, giving a measured oblateness about two times larger than the Princeton result. After several months of trying to ascertain the origin of this error, it was discovered to depend directly on anisotropy in the earth's atmospheric seeing. Further study showed that this systematic error also existed in the Princeton experiment but not as effective by a factor of 2. These results precipitated a complete restudy of methods to determine the edge of the sun and led to the development of the Fourier Coefficient Technique (FCT) by Mr. Stebbins (Ref. 5). This technique is presently being implemented.

There are no new results to report on the solar oblateness because of the seriousness of the systematic error mentioned above.

The edge of the sun as located by FCT is defined as that point r_o on the disc for which the Fourier coefficient of the intrinsic limb darkening curve from $r_o - a$ to $r_o + a$ is zero. Experimentally, one must work with the intrinsic limb darkening curve convoluted with the transfer function of the earth's atmosphere and the instrument. If r is the point at which the Fourier coefficient in FCT is zero for the convoluted limb darkening curve, then r_o is obtained by taking limit of r as the width of the transfer function goes to zero.

A Mark II Solar Oblateness Detector has been designed, built, and is presently being installed to take advantage of this new technique. This work has been done by P. Clayton and should yield results on the solar oblateness in the near future.

A second program on the solar oblateness is a look at the anisotropy of the solar limb darkening by D. Patz using FCT. It appears that this technique affords an extremely sensitive way of doing this.

A measure of this technique can be obtained by using the empirical limb darkening function in Allen (Ref. 6) and adding the temperature dependent term recently suggested by Ingersoll and Spiegel (Ref. 7). This is a brightness variation that goes as $x^{-1/2}$ when x is measured in from the edge. This leads to an observed oblateness which depends on a of FCT and is equal to $0.016a^{1/2}$ arc sec where a is in units of arc sec. For a change in $a^{1/2}$ of 3 arc $sec^{1/2}$, there is a change in oblateness of 0.048 arc sec, the size of the total result reported by Dicke and Goldenberg (Ref. 1).

The star detectors from the light deflection experiment are near completion by B. Cardon. Both this experiment and the $\dot{G}/G$ are expected to be operating this year. The work on $\dot{G}/G$ by A. Healy is expected to give considerable information relating to the earth's motion as indicated in the introduction.

A preliminary study has been made of the problems associated with the measurement of the parallax of spectroscopic binaries and it indicates that for these objects distances up to 10^5 pc might be measured. Work in this area is continuing.

References

1. Dicke, R. H., and Goldenberg, H. M., Phys. Rev. Letters, 18, 313, 1967.

2. Dicke, R. H., Nature, 202, 432, 1964.

3. Schwarzchild, M., Structure and Evolution of Stars, Princeton University, 1958.

4. Schupmann, L., Die Medial Feinrohre, Teubner, Leipzig, 1899.

5. Stebbins, R. T., "The Fourier Coefficient Technique for Defining the Edge of the Sun," Senior Thesis, Wesleyan University, 1970.

6. Allen, C. W., Astrophysical Quantities, second edition, Athlone Press, London, p. 170, 1963.

7. Ingersoll, A. P., and Spiegel, E. A., Temperature Variation and the Solar Oblateness (submitted for publication).

Radio Propagation Measurements of the Solar Corona and Gravitational Field: Applications to Mariner 6 and 7

Duane O. Muhleman, John D. Anderson,
Pasquale B. Esposito, and Warren L. Martin
Jet Propulsion Laboratory
California Institute of Technology

I. Introduction

The great advancement of the techniques of radio astronomy and the precise radio tracking of space probes in solar orbits have made it possible to perform significant experiments on electromagnetic wave propagation in the solar gravitational field. These experiments include the measurement of the apparent angular deflection of radio sources (natural and artificial), the propagation time delay to space probes and planets, the relativistic doppler effect essentially caused by the time variation of the delay, and effects on pulsar apparent periods caused by a combination of the above phenomena.

However, the measurement of these effects must be carried out through a medium containing a significant electron density which strongly varies with distance from the sun and in a random fashion in both space and time. The electron plasma causes both systematic effects which resemble the general relativistic phenomena and, by virtue of the plasma's stochastic nature, difficulties in signal processing due to a loss of coherence. The systematic effects can be minimized (but never eliminated), or they can be studied as separate phenomena by using multiple-frequency systems (the plasma differs from the gravitational field by being dispersive) or by modeling the electron density profiles, the parameters of which may be estimated from the observations.

At this time, the stochastic effects appear to be more serious, particularly for propagation near the sun. These effects are best minimized by employing clever signal detection techniques and by using the highest possible frequency. However, experimenters are severely limited as to frequency by practical matters such as the fixed frequency utilized by the deep space probes of the USA (~2300 MHz), the paucity of strong natural radio sources at frequencies beyond, say, 8 GHz, and the small number of discovered pulsars with significant power in their spectra beyond 2 GHz. Clearly, there is little value in a discussion of these experiments without regard for the ever-present plasma effects or in dismissing them on the basis that they can be eliminated by multifrequency or dispersion techniques.

An attempt is made to face these questions squarely in this paper. The necessary theory for propagation in the gravitational field and plasma is developed in some detail in section II. In section III, we will discuss the available results from interferometric measurements of the apparent angular deflection of QSO (quasi stellar object). Some preliminary results from Mariner 6 and 7 experiments, particularly from the standpoint of the plasma effects, are discussed in section IV. Future experiments and plasma scintillation phenomena are described in section V.

II. Theoretical Observables

In all cases, the effects of general relativity on the observables such as time delay, doppler

shift, or angular deflection are very small relative to those of other parameters in the problem, e.g., six constants representing the position and motion of the earth. Consequently, the presence and magnitudes of relativity terms must be estimated simultaneously with the other constants of the problem, usually by a least-squares, differential corrections scheme. In this procedure, the observed quantities, e.g., time delay, are compared to quantities computed from numerical integrations of the equations of motion of all the solar system bodies.

For our purposes, we can consider the computed observables as a linear combination of geometrical terms which include all classical effects of the planetary positions and motions plus the small propagation effects of relativistic and plasma terms. The computer programs developed by the Jet Propulsion Laboratory numerically integrate fully relativistic equations of motion written in isotropic coordinates. Consequently, in the discussion here the relativistic dynamic effects are included in the "classical" motions, and the observables are computed as coordinate quantities. The relativistic dynamic terms cause only a second-order effect on the propagation observables and will not be considered further in this paper. Clearly, the particular coordinate system chosen is unimportant. The observables are independent of the coordinate system.

Maxwell's equations and, consequently, the electromagnetic wave equation in a gravitational field can be rigorously interpreted as the usual equations of special relativity written in a flat space with a particular permittivity due to the field (Ref. 1). In isotropic coordinates, this leads to the concept that the gravitational field can be replaced by a medium with a spatially isotropic and unique index of refraction $n(\bar{r})$. An eikonal equation can be written and observables can be computed using the techniques of geometrical optics. As pointed out by Muhleman and Johnson (Ref. 2), the rigorous treatment of propagation characteristics in the solar system requires that the total refractive index be written at each point because of the combination of the gravitational field and the electron plasma. The refractive index for the spherically symmetric gravitational field can be written

$$n_R^2(r) = 1 + \frac{4m}{r}; \qquad R_\odot \le r \tag{1}$$

to be first order in $m/R_\odot$, where $m = GM_\odot/c^2$ and $R_\odot$ is the radius of the sun. It should be realized that n_R is the ratio of the vacuum speed of light to the coordinate speed of light at each point in space. As shown in Ref. 2, for the relatively weak field of the sun, the total refractive index at a point $\bar{r}$ in the solar system can be written

$$n^2(\bar{r}) = 1 + \frac{4m}{r} - \frac{4\pi e^2}{m_e w^2} N_e(\bar{r}) \tag{2}$$

where w is the signal frequency and $N_e(r)$ is the electron density at r. It can be successfully argued that the effects of magnetic fields in the corona and particle collisions can be neglected at frequencies higher than about 300 MHz.

The electron density profile in the solar corona and the interplanetary medium have been extensively investigated with photometric and spectroscopic methods during solar eclipses (see Ref. 3 for a review), mostly in the region from the chromosphere out to about 10 solar radii. Recently, Blackwell et al. (Ref. 4) have extended these data to about 40 solar radii and obtained a reasonable extrapolation to the orbit of the earth. The plasma has been intensively studied near 1 AU by space probe measurements. In general, the electron density profile can be described as being very nearly spherically symmetrical, or at least cylindrically symmetrical, about the rotational axis of the sun. Variations of the mean density with solar latitude are seen, particularly very near the solar disk. Superimposed on this mean structure are numerous streamers which apparently co-rotate with the sun. The effects of this random-like structure tend to average out for propagation between two distant points in the solar system.

Since these effects, when integrated over the ray paths, are expected to be small, we will assume for now that the electron profile is spherically symmetrical to first order. Experimental deviations from this model based on Mariner 6 time delay measurements are discussed in section IV. Assuming that the total index of refraction profile is smooth in the mathematical sense, the ray path through the medium connecting the two end points (the radio source and the earth) may be found using Fermat's principle. The mathematical procedures are straightforward and are described in detail for both the nondispersive (gravitational field) and the dispersive case (plasma) in Refs. 5-8. The results we need from these references are the time delay between two bodies at radial distances r_1 and r_2 from the sun

$$\rho = \int_{r_0}^{r_1} \frac{r n_g(r) n(r) dr}{\sqrt{r^2 n^2(r) - b^2}} + \int_{r_0}^{r_2} \frac{r n_g(r) n(r) dr}{\sqrt{r^2 n^2(r) - b^2}} \tag{3}$$

and the total ray bending

$$\psi = -\int_{r_0}^{r_1} \frac{b\left(\frac{dn}{dr}\right) dr}{n(r)\sqrt{r^2 n^2(r) - b^2}} - \int_{r_0}^{r_2} \frac{b\left(\frac{dn}{dr}\right) dr}{n(r)\sqrt{r^2 n^2(r) - b^2}} \tag{4}$$

where n(r) is given by Eq. (2), r_0 is the radial distance of closest approach to the sun, b is the impact parameter of the ray asymptote, and $n_g(r)$ is a symbolic representation of the "group index of refraction" given by the usual dispersion relationship

$$n_g(r) = n(r) + w\frac{dn}{dw} \tag{5}$$

Equations (3) and (4) are exact equations (for a spherically symmetric region) which take fully into account the curvature of the ray path at each spatial point, e. g., the electron density profile causes the ray path between two given end points to be curved in toward the sun, where the gravitational field is slightly more intense.

Fortunately, for all practical experiments at frequencies higher than about 300 MHz the medium is "thin" and the equations can be linearized without a significant loss in accuracy. For example, experiments at 2000 MHz are possible to within about 3 solar radii (see section V). At this point, the refractivities ($|n(r)-1|$) are 1.4×10^{-6} for the gravitational field and less than 5×10^{-6} for the coronal plasma. Clearly, the squares of these terms are negligible for all currently proposed experiments. The linearization procedure is carried out for Eqs. (3) and (4) in the appendix and expressions for the three observables, time delay, total bending, and doppler velocity are derived. It is shown there that for each case of interest, the integrals over the curved path can be replaced by integrals over the straight line connecting the end points and, consequently, the contributions to the observables from the gravitational field and the plasma simply add linearly. The results for the time delay in terms of the impact parameter of the geometrical line of sight p are (Eq. A-12)

$$\rho = \rho_{geom} + 2m \ln\left[\frac{r_1 + r_2 + \rho_{geom}}{r_1 + r_2 - \rho_{geom}}\right]$$

where the geometrical distance ρ_{geom} is, of course, computed from the numerical orbital integration in isotropic coordinates.

The first-order expression for the total ray bending is (Eq. A-9)

$$\psi = \frac{2m}{p}\left[\frac{1}{r_1}\sqrt{r_1^2 - p^2} + \frac{1}{r_2}\sqrt{r_2^2 - p^2}\right]$$

If the source and the earth are removed to infinity, ψ becomes $\psi = 4m/p = 1''.75/p$ (p measured in solar radii), which is the usual expression for the Einstein effect. The terms in brackets arise from the refractivities at the end points, which are exactly equivalent to the finite space curvature at the end points in the usual curved-space formulation.

If we introduce the critical electron density N_w,

$$N_w = \frac{4\pi e^2}{m_e w^2}$$

the time delay due to the plasma is given by Eq. (A-16):

$$\rho_{p_1} = \frac{1}{2N_w}\left[\int_p^{r_1} \frac{Ne(z)zdz}{\sqrt{z^2 - p^2}} + \int_p^{r_2} \frac{Ne(z)zdz}{\sqrt{z^2 - p^2}}\right]$$

and the bending due to the plasma is

$$\psi_{p_1} = \frac{p}{2Nw}\left[\int_p^{r_1} \frac{\frac{dNe}{dz}dz}{\sqrt{z^2 - p^2}} + \int_p^{r_2} \frac{\frac{dNe}{dz}dz}{\sqrt{z^2 - p^2}}\right] \tag{6}$$

Finally, the first-order expression for the doppler velocity can be written from Eq. (A-19) in the form

$$\frac{d\rho}{dt} = \frac{d\rho}{dt}_{geom} + \psi(p)\frac{dp}{dt} + F \tag{7}$$

where F is a small end-point term arising from the finite refractivity at the two end points:

$$F = \frac{\eta(r_1)}{\sqrt{r_1^2 - p^2}}\left(r_1\frac{dr_1}{dt} - p\frac{dp}{dt}\right)$$

$$+ \frac{\eta(r_1)}{\sqrt{r_2^2 - p^2}}\left(r_2\frac{dr_2}{dt} - p\frac{dp}{dt}\right) \tag{8}$$

For solar orbits with semi-major axes > 0.3 and eccentricities < 0.3, the F terms are on the order of 10-4 m/s and are below the accuracy of current doppler tracking systems. It can be seen from Eq. (6) that the doppler velocity due to either the gravitational field or the plasma is essentially given by the product of the bending angle and the rate of change of the geometrical impact parameter. Since the bending angles are of opposite sign for the two effects, the doppler shifts are also of opposite sign. A phase tracking doppler system will actually measure a decrease in the optical path length due to the plasma as the ray path is swept toward the sun.

III. Ray Bending Experiments

At the time of this writing, the only reported successful measurements of ray bending in the solar gravitational field are those of Muhleman et al. (Ref. 9) and Seielstad et al. (Ref. 10). The first of these interferometer experiments was performed at 2388 MHz on a baseline of 21 km, and the second at 9600 MHz on a baseline of 1 km. The final results for the bending angle referred to the solar limb, were $\psi = 1''.82 \pm 0.20$ and $\psi = 1''.74 \pm 0.20$ ($1''.77$ is reported in the abstract of Ref. 10; the $1''.74$ figure is also reported in the paper using a method of analysis which includes the very small effect of the plasma at this frequency). In both cases, the angular separation between two radio point sources, 3C273 and 3C279, was measured as a function of the angular distance from the sun. The published results were obtained from observations to within about 4.0 solar radii from the sun. The difficulty in precisely removing the plasma bending at 2388 MHz is responsible for most of the $\pm 0''.20$ error. The quoted error on the 9600 MHz result is caused by the observational errors due in part to the relatively short baseline; plasma bending at this frequency at distances greater than 4 solar radii is nearly negligible.

In order to analyze data taken at a single frequency, it is necessary to model the electron density profile $N_e(r)$. Considerable work has been done on this problem during this century, the most definitive study being that of van de Hulst (Ref. 11; see also Ref. 3). The run of the scattered light intensity with distance from the sun has been measured during many total solar eclipses, usually to distances of about 10 solar radii. Recently, Blackwell et al. (Ref. 12) have extended the observations to about 40 solar radii. The observations are well represented by an interpolation function

$$N_e(r) = \frac{A}{r^6} + \frac{B}{r^{2.3 \pm 0.3}} \qquad (9)$$

for $r > 2$ solar radii. Space probe in situ measurements near the orbit of the earth approximately confirm the inverse-square dependence in that region (the r^{-6} term is negligible beyond about 5 solar radii). These results are in good agreement with radio scintillation measurements and solar wind theory as reviewed by Newkirk (Ref. 3). The ray bending experiments utilized Eq. (8) as a plasma model and estimated values of A and B simultaneously with the relativistic bending angle ψ. Muhleman et al. (Ref. 9) obtained estimates of A and B which result in an electron density profile about a factor of 2 smaller than that of Blackwell et al. (Ref. 12).

Recently, Muhleman et al. (Ref. 13) have combined their two sets of observations (which were made in the same time period) to obtain a better estimate of the A and B parameters, in particular. Two sets of observations are available: the data used in the analysis of Refs. 9 and 10 and an extended set which includes 9600-MHz observations taken within 1.8-2.9 solar radii. This latter group were not used by Seielstad et al. (Ref. 10), since the data are strongly affected by the corona even at their high frequency and would considerably degrade the analysis of their relativistic bending.

According to Ref. 13, the combined data for distances greater than 4 solar radii yield the same corona as reported in Ref. 9 and a relativistic bending of

$$\psi = 1''.78 \pm 0''.14$$

i.e., the mean of the two published results. This occurs because the 9600 MHz data contain essentially no information about the corona.

The combined data set, including the data very near the sun, cannot be used to improve the estimate of the relativistic bending due to the strong influence of the corona. Instead, they fixed the relativistic bending at $1''.75$ and estimated A and B only. The resulting electron density profile was found to be

$$N_e(r) = \frac{(0.80 \pm 0.27) \times 10^8}{r^6} + \frac{(0.51 \pm 0.30) \times 10^6}{r^{2.3}}$$

electrons/cm^3 near the solar equator during October 1969. This expression gives a value of the electron density at 1 AU of 1.9 ± 1.1 cm^{-3}, which is smaller than the well established mean value from space probe measurements of 5 ± 3 cm^{-3}.

Although it is possible that the electron densities were low during the observation period (the sun remained relatively quiet), Muhleman et al. (Ref. 13) believe that the ray bending method systematically underestimates the electron density relative to that obtained by the eclipse methods. This can be seen from Eq. (6). The bending depends on the integrated gradient of the electron density, whereas the light scattering during an eclipse (and also the total time delay) depends on the integrated electron density. If the wavefront encountered an ideal slab of electrons, the total bending through the slab would be zero. However, these electrons would fully contribute to light scattering and time delay. A similar agreement can be derived for propagation through an ensemble of spherical globs of electrons imbedded on the mean profile. The contribution to the total bending would average to a zero mean. An adequate treatment of these configurations cannot be performed with geometrical optics. It is shown in section IV that the electron density profile obtained from time delay measurements agrees very well with eclipse results.

IV. Plasma-Relativity Effects Separation for Mariners 6 and 7

Since the Mariner spacecraft employ a single tracking frequency (wavelength ~13.1 cm), the electron density profile for the solar corona (beyond 4 solar radii, $R_\odot$) must be modeled. We use the form of Eq. (9), with three parameters, A, B, and ϵ, to be determined from the observations:

$$N_e(r) = \frac{A}{r^6} + \frac{B}{r^{2+\epsilon}} \qquad (10)$$

The degree of success in estimating these parameters separately from the other parameters of the problem is determined by the signatures (i.e., partial derivatives) for each parameter in the observables. In an ideal case, the signature of each parameter would be entirely different or mutually "orthogonal," and each parameter could be computed free of correlated errors from the others. For example, if a significant component in $N_e(r)$ existed with a distance dependence of $1/r$, this component could not be separated from the relativity effect since it has the same structure. Fortunately, there is no evidence to suggest such a component. The signatures of A, B, and ϵ are sufficiently different from the relativity effect for the Mariner data to ensure an adequate parameter separation. However, considerable experience with the data indicates that these effects limit the accuracy of the Mariner experiment to 1%.

A great deal of information exists on the numerical values of A, B and ϵ from eclipse and radio-astronomical scintillation observations. The latter experiments bound ϵ between 0 and 0.4 (Refs. 14 and 15). A priori estimates of the parameters and values for their standard deviations were obtained from an exhaustive study of the literature of solar eclipse investigations. The a priori values that we have adopted are

$$A = 1.3 \times 10^8 \pm 0.9 \times 10^8$$

$$B = 1.15 \times 10^6 \pm 0.7 \times 10^6$$

$$\epsilon = 0.3 \pm 0.3$$

where the units are such that Eq. (10) yields the electron density at a point r in electrons per cubic centimeter when r is expressed in solar radii. The values of the parameters are of no importance in the analysis of the data, since these parameters are estimated in the process. However, the standard deviations are very important, since they are used in the a priori weighting matrix in the least-squares analysis. These standard deviations statistically constrain the final solution for all the parameters in such a manner that the final estimates of the plasma parameters are statistically consistent with the other independent information on the plasma from, e.g., eclipse investigations. The a priori standard deviations used here are actually significantly larger than the eclipse data suggest, and their use in the weighting matrix makes our weighting procedure conservative.

Another important independent result which we must use is that the average electron density at 1 AU is 5 $\pm$3 cm^{-3} (again with a conservative standard deviation). Consideration of the B and ϵ parameters shows that errors in these parameters must be highly correlated to be consistent with the standard deviation of the density being ± 3 cm^{-3} at 1 AU. Consequently, we have adopted a correlation coefficient of B,ϵ of 0.9999 in the a priori covariance weight matrix. Obviously, in estimating the relativity parameter, this is nearly equivalent to fixing either B or ϵ (with a zero standard deviation) and estimating the other. The real importance in these procedures lies in computing a realistic estimate of the final error on the relativity parameter which will accurately include the available independent information on the medium.

The parameterization of the relativistic propagation effect is achieved with a parameter γ^* which multiplies the relativity expression (essentially the log term in Eq. A-12). Clearly, Einstein's formulation requires γ^* to be unity. (Actually, the analysis program estimates the usual metric parameters γ and β, where $\gamma^* = (1 + \gamma)/2$; see Ref. 16.) We have performed an extensive error analysis on the separation of the data of γ^* from the remaining parameters of the problem, which include the orbital parameters of the spacecraft orbit, those for the earth's orbit, certain parameters of the spacecraft, plasma parameters, etc. These matters are extensively discussed in Ref. 16. The highest correlations occur between γ^* and the plasma parameters, since their signatures are "similar" in that they peak near solar occultation. However, the plasma delays are small compared with the relativity effect.

The expected time delays for the Mariner 7 trajectory due to the A and B terms, using the a priori values, are shown in Fig. 1. The symbols are drawn at the times for which we actually have time delay observations. It can be seen from the figure that, for the Mariner 7 orbit, the A term was sampled at only five points and the maximum delay is only 100 m. For this spacecraft (but not for Mariner 6), there is no need to estimate A. The a priori time delays for the total plasma and the relativity effects ($\gamma^* = 1$) are shown in Fig. 2. Although the plasma delay at the peak point is 6% of the relativity effect, the important factor is a weighted integral of the two

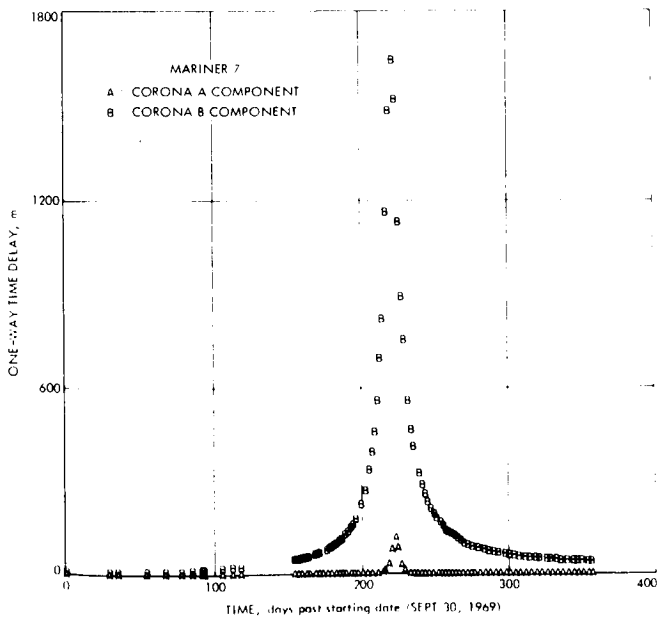


Fig. 1. Predicted plasma delays for the A and B terms for Mariner 7 (The symbols appear for days on which measurements are available.)

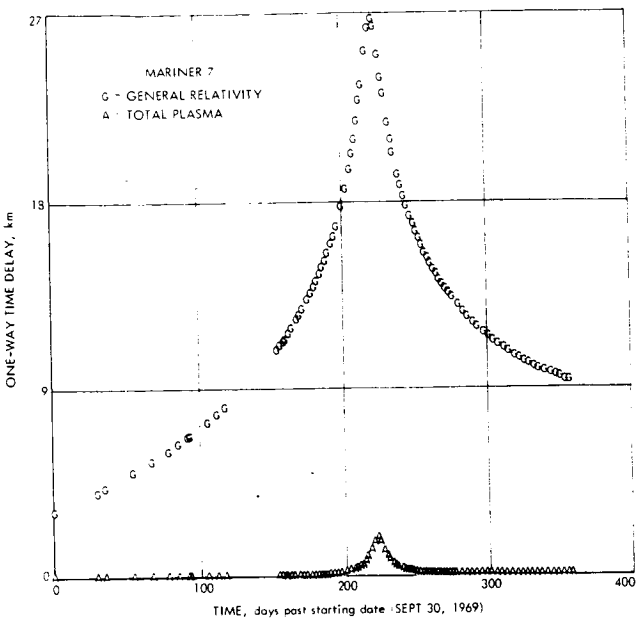


Fig. 2. Predicted relativity delays (G) and total plasma delays (A) for Mariner 7

curves. In fact, if we simply ignored the existence of the plasma in the data analysis, we would obtain a bias error on our estimate of γ^* of about 1%.

The Mariner 6 a priori curves are very similar to those for Mariner 7. However, Mariner 6 passed about 1 solar radius closer to the sun, and the A term is about a factor of 10 larger. The effect of this is to approximately double the total plasma delay within the time period of ±5 days around solar occultation, with no effect outside of this interval. The data obtained for Mariner 6 are sufficient for the estimation of all of the parameters, as is shown below.

The effect on the final error of γ^* due to correlated errors in A and B is well illustrated in a series of numerical simulations of the actual Mariner 6 experiment. Observations were simulated over an 87-day portion of the trajectory centered on solar occultation, with range errors added to the theoretical observables, which were drawn from a table of gaussian random numbers with appropriate variances. We used one observation per day, except for 6 days around solar occultation. The parameters γ^*, A, and B were then estimated from the pseudo-data. The exact parameters used and the results of the numerical experiments are shown in Table 1. In each case, two values of range-measurement standard deviations were used, the larger value of σ_{SO} data being selected for the 10 observable days nearest to the sun to model the possible degradation of the measurement accuracy due to plasma turbulence. The data errors chosen are certainly larger than those of the actual measurements. Table 1 shows that in each case, the estimated value of γ^* agrees with the input value to a fraction of 1%. The A parameter is rather poorly determined because very few of the observations are near enough to the sun to be sensitive to the A term. The B parameter is determined to better than 10%. Although these numerical experiments hardly provide a definitive error analysis, the results are completely consistent with the computations employing the actual data.

Deviations from our static and spherically symmetric plasma model are certainly to be expected in the actual medium. The motion of the line of sight to Mariner 6 with respect to the sun in heliographic coordinates is shown in Fig. 3. The probe ray path passed north of the sun on April 30, 1970, and a complete range of North-solar latitudes were covered over a 3-month period.

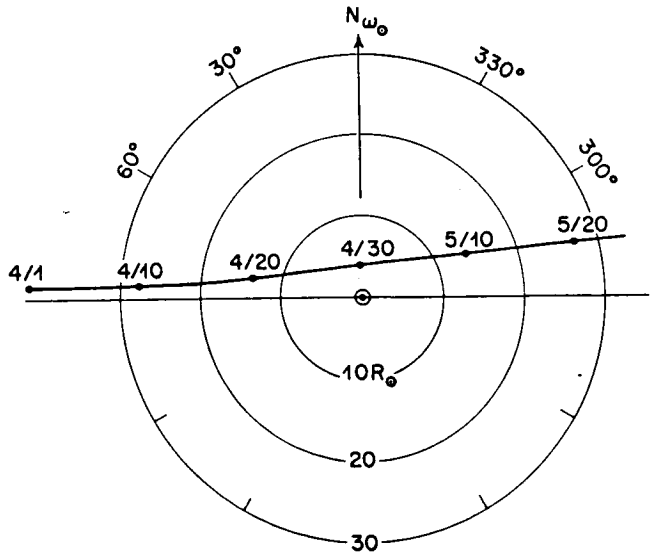


Fig. 3. Projected motion of Mariner 6 with respect to the solar equator

Table 1. Numerical simulations of Mariner 6 (87-day arc, 6-day gap at spacecraft)

Experiment	Input parameters			σ_{data}, m	σ_{SO} data, m	Estimates of parameters		
	γ^*	A	B			γ^*	A	B
1	1.0000000	1.38×10^8	1.15×10^6	±20	±200	0.99955	1.103×10^8	1.189×10^6
2	1.0000000	1.38×10^8	1.15×10^6	±40	±600	1.00005	0.823×10^8	1.168×10^6
3	1.0000000	1.38×10^8	1.15×10^6	±75	±1000	0.99995	0.619×10^8	1.154×10^6
4	1.0000000	1.38×10^8	1.15×10^6	±75	±1000	1.00076	2.416×10^8	1.043×10^6

As a further test of our model, we performed a special solution for all of the parameters, with γ^* fixed at unity except for A, B, and ϵ. We assumed that no plasma existed. The time delay residuals from such a solution essentially exhibit the plasma delays (for Einsteinian relativity). These residuals are shown in Fig. 4, along with the theoretical plasma delays computed from our model with the a priori parameter values. The agreement with the a priori model is remarkable. Significant short-period deviations do exist, there being times when the delays are less than predicted, or equivalently, when the integrated electron density along the ray path is less than expected. No significant correlations between these deviations and solar activity can be explained. These results are shown in terms of the fraction of the a priori model in Fig. 5.

The observed short-period variations never exceeded 50% of the expected plasma delays. The overall structure in the data is probably due to the solar-latitude effect. The point marked with (?) in Fig. 4 was measured at the closest approach to the sun in a region of extreme turbulence. The detailed behavior of the radio receiver strongly suggests that this point is not valid, and it will not be used in the further analysis.

Finally, numerous solutions to the complete problem for Mariner 6 have been obtained. The various solutions are discussed in Ref. 16. In

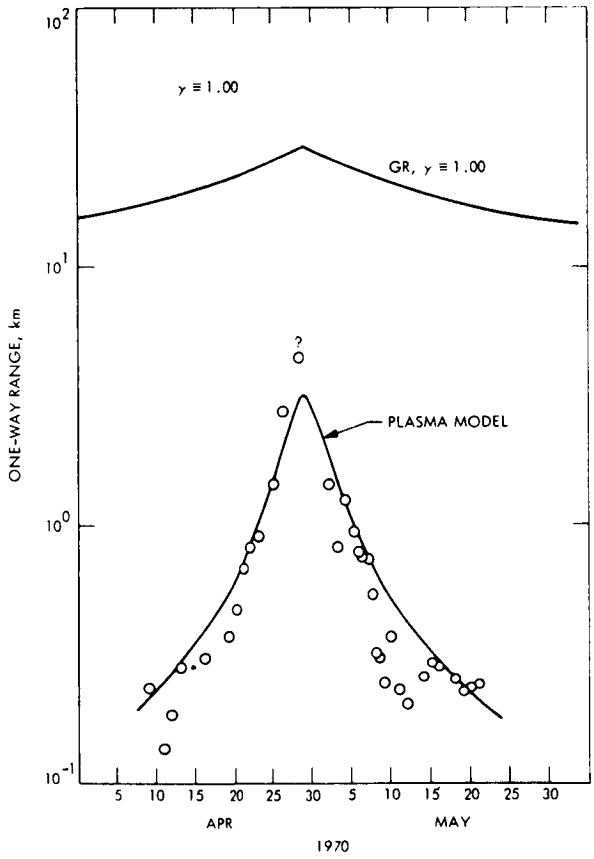


Fig. 4. Mariner 7 range residuals for a solution, with γ^* fixed at unity and the plasma effect assumed to be zero (The full line represents the delays for the a priori plasma model.)

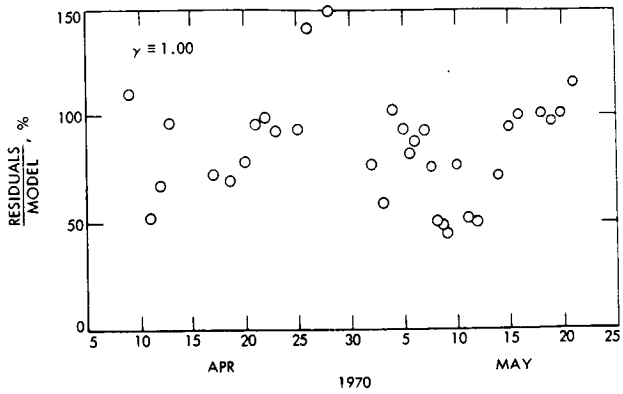


Fig. 5. Ratio of the range residuals from Fig. 4 to the predicted delay from the plasma model

this paper, we will consider only that solution which we regard as the most reliable at this time. The observational set consisted of time delay and doppler velocity data over a 3-month period around occultation. The pertinent results are:

$$A = (1.92 \pm 0.70) \times 10^8, \text{ cm}^{-3}$$

$$B = (1.41 \pm 0.48) \times 10^6, \text{ cm}^{-3}$$

$$\epsilon = 0.41 \pm 0.21$$

$$\gamma^* = 0.9973 \pm 0.0144 \text{ (formal error)}$$

We must again emphasize that in this paper we are primarily interested in the plasma effects; a full discussion of the γ^* determinations is given in Ref. 16. In particular, the quoted errors are statistical formal errors which must be slightly increased, as shown in Ref. 16. The above parameters yield a value of the electron density at 1 AU, averaged over the 3-month period, of 3.4 ± 1.1 electrons/cm^3.

The resulting electron density profile is indicated by the dashed line in Fig. 6, along with the eclipse results of Blackwell et al. (Ref. 12), and the profile obtained by the interferometric measurements of ray-path bending using discrete radio sources (Ref. 13). As pointed out above, the ray bending measurements apparently yield a low estimate of the electron densities due to the effects of plasma irregularities, although the possibility that the medium was underdense during October 1969 cannot be ruled out.

The determination of estimates of A, B, and ϵ separately, while of considerable interest, is not vital to the relativity problem. The important factor is the integrated electron density along the ray path, which is not affected by the correlations between the three plasma parameters. The integrated density is, of course, directly proportional to the plasma time delay and is measured directly in our procedure. The integrated electron density was computed by integrating Eq. (10) using the determined values of A, B, and ϵ along a radial direction from the sun between a point at distance r to 1 AU. The effects of the correlated errors in the three parameters essentially vanish in the integration. The results are shown by the solid

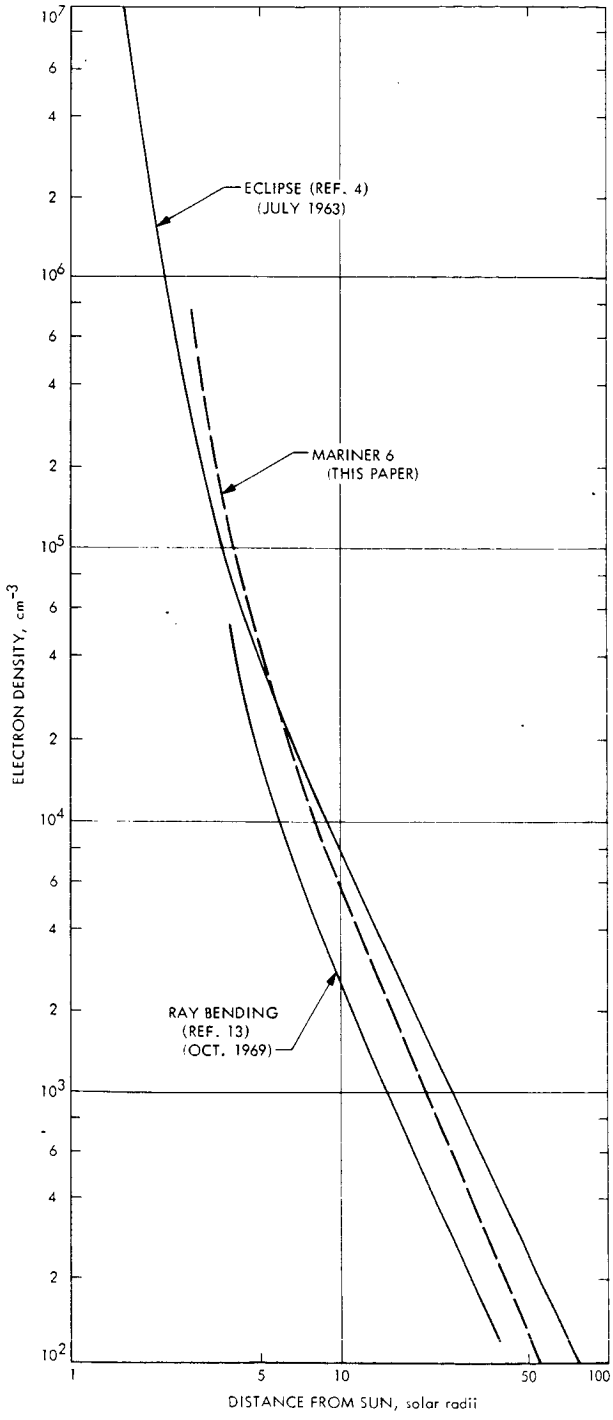


Fig. 6. Electron density profiles from the Mariner 6 experiment, 1963 solar eclipse measurements, and interferometric ray bending measurements

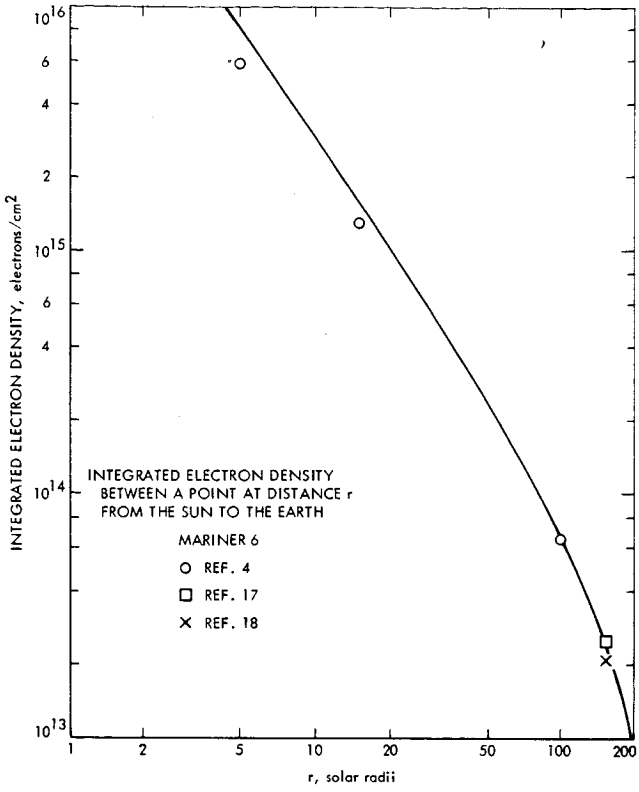


Fig. 7. Integrated electron densities from Mariner 6 and eclipse measurements (Independent measurements for the path between the earth and Venus are indicated by × and □.)

line in Fig. 7. Also shown are three representative points of the Blackwell et al. profile (circles) and two independent determinations of the integrated density between the earth and Venus. The result obtained from a two-frequency measurement to the Mariner 4 spacecraft when it was near Venus (Ref. 17) is indicated by a cross, and that for two-frequency time delay radar measurements to the surface of Venus (Ref. 18) is indicated by a square. These results are all in excellent agreement. The difference between the Mariner 6 curve and the eclipse curve is probably real, indicating the difference between the integrated density near the solar cycle maximum (Mariner) and minimum (eclipse during 1963). For example, a literal interpretation of Fig. 7 yields a total electron content between 6 $R_\odot$ and the earth of 8×10^{16} electrons/cm^2 in 1970 and 6×10^{16} electrons/cm^2 in 1963, a change of about 30%.

We believe that our results concerning the average plasma are consistent with all other investigations of the coronal and interplanetary electron plasmas. Furthermore, the evidence is strong that our method for estimating the relativity effect has removed the plasma effect to an accuracy of better than 1% in γ^*.

V. Future Radio Propagation Tests of General Relativity

Any test of general relativity which uses radio propagation near the sun will be strongly affected by plasma irregularities. This is particularly true for at least the next 5 years, since such experiments will be limited to wavelengths longer than 1 cm for interferometric ray bending experiments and longer than about 4 cm for all spacecraft experiments. The plasma irregularities moving with the solar wind cause a random phase modulation and amplitude scintillations on

radio signals. These effects have been extensively studied from about 5 $R_\odot$ to 1 AU during the last decade.

The root-mean-square phase jitter can be computed from amplitude scintillation observations and scintillation theory. The phase jitter essentially depends on the electron density fluctuations, and the observing wavelength and the density fluctuations are approximately proportional to the mean density. Assuming that these ideas can be extended into the unobserved region from 5 $R_\odot$ to the solar limb, we can roughly estimate the rms phase jitter to be expected at any frequency. The results of these calculations are shown in Fig. 8 for a one-way passage through the medium.

Quite generally, spacecraft radio tracking systems which measure time delays and doppler shifts require a phase jitter in their pass bands of less than $\pi/2$ rad, i.e., a measure of coherence over time scale on the order of 1 s. Figure 8 indicates the rough distance from the sun at which this incoherence sets in for wavelengths of 2, 3.7, and 13.1 cm. These curves were computed from scintillation theory and associated empirical parameters and are strongly supported by our experience with Mariner and Pioneer space probes at 13.1 cm; i.e., we were unable to make range measurements closer than 6 $R_\odot$ and doppler measurements closer than about 10 $R_\odot$. An increase in frequency to a wavelength of 3.7 cm (currently in progress) only slightly improves the situation. All of this certainly does not mean that improved coding and signal detection techniques cannot be developed, but such developments do not seem to be forthcoming within the next 5 years.

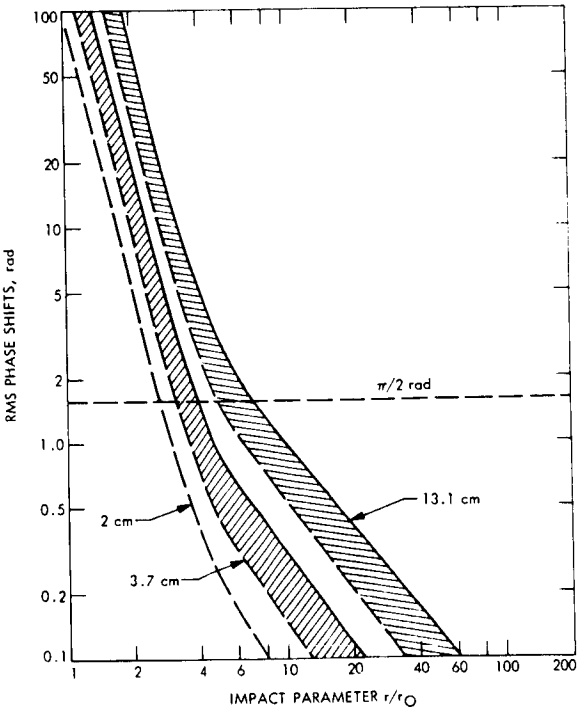


Fig. 8. Estimated values of the rms phase jitter due to plasma scintillations as a function of the ray impact parameter

The situation for interferometric measurement of ray bending in the gravitational field is much more favorable for at least three reasons: (1) it is not necessary to work very near to the sun, since the relativity effects do not have to be separated for the orbital effects as is the case for space probe experiments; (2) the signal detection process is less vulnerable to the scintillations; and (3) there is no limit on frequencies that can be employed; e.g., interferometry at 1.3 cm of natural radio sources has been demonstrated.

The effect of scintillations on the measurement of precise radio source positions is to cause the apparent position in the sky to jump about rapidly, forming a "tremor disk" with an intensity pattern which is approximately gaussian. The standard deviation of this pattern is roughly

$$\phi \sim \frac{160''}{f_G^2 r^{2.5}}, \qquad 5R_\odot \le r \tag{11}$$

where ϕ is measured in arc seconds, f_G is the frequency in gigahertz, and r is expressed in solar radii. If the angular spacing between the interferometer fringes is larger than ϕ, then long records can be taken to find the centroid of the pattern which is a good estimate of the apparent source position. At higher resolution (larger spacing of the interferometer), where the fringe spacing is small compared with ϕ, the source will be resolved out and no measurement is possible; for example, at 8 GHz and r = 5 $R_\odot$ ϕ ~ 0".045, the source would become overresolved at spacings greater than about 160 km. Thus, very long baseline interferometry (VLBI) techniques apparently cannot be used so near the sun. At about 40 $R_\odot$ (~10°) from the sun, the scintillation broadening should be negligible at 8 GHz. However, the relativity bending is reduced to 0".045, and source positions would have had to be measured to an accuracy of 4".5 × 10^{-4} to achieve a 1% relativity test. This accuracy may or may not be feasible in the near future.

Finally, we will consider tests that utilize the doppler velocity shift arising from the motion of the ray path relative to the sun for spacecraft transmitters and pulsars. The doppler effect has not as yet been exploited for general relativity experiments. The theoretical one-way doppler shifts for probes moving in circular orbits with various semi-major axes are shown in Fig. 9. The pulsar case is the curve labeled a = ∞; i.e., the variation arises from the motion of the earth about the sun for a fixed source at infinity. The figures on the right of each curve give the times in days required for the ray path to cover the range of impact parameters. The present capability for measuring the doppler velocity for deep space probes is about ±5 × 10^{-3} m/s in regions removed from the sun. All of the space probes launched thus far have a range of semi-major axes from 0.7 to 1.5 AU and doppler systems operating at a wavelength of 13.1 cm. It can be seen from Fig. 9 that the doppler effects are very small in this range, and, according to the above arguments, it is not possible to measure the doppler shift at this wavelength closer than, say, 6-8 solar radii. Mercury- and Jupiter-type orbits are more favorable.

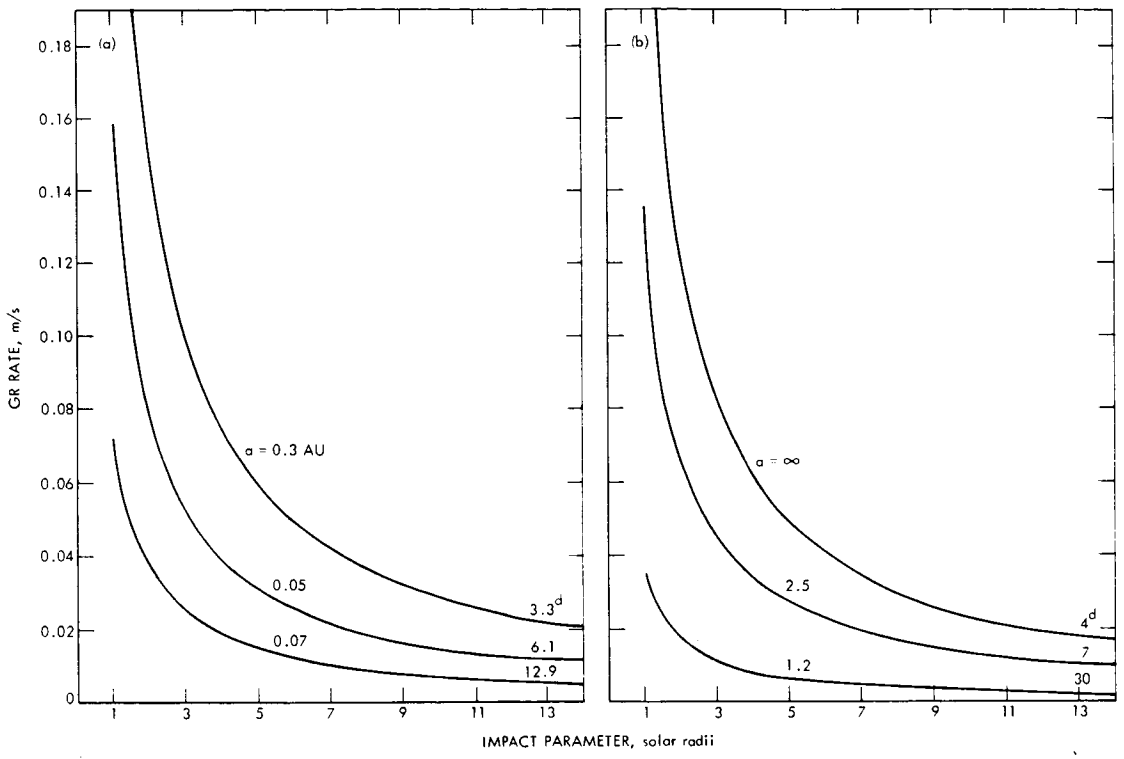


Fig. 9. Time rate of change of the relativistic time delays for circular orbits: (a) interior to the orbit of the earth; and (b) exterior to the orbit of the earth (The semi-major axes of the orbits are given by a.)

Doppler experiments have several advantages over time delay experiments. The systems are very simple, since only a continuous wave is required. Doppler systems operating at wavelengths as short as 1 cm could easily be developed for spacecraft applications. Furthermore, since the relativity signature goes through its major variations in just a few days (with a change in sign at occultation), the experiment would be less vulnerable to the effects of non-gravitational forces on the spacecraft.

Perhaps the most nearly optimum experiment could be achieved by flying an oscillator which is stable to 1 part in 10^{12} (or better) over a period of a few weeks. In this case, only one passage through the medium is required, and a wavelength in the millimeter region could be employed, since ground-transmitter facilities would not be required. The signal could be tracked nearly to the solar limb at millimeter wavelengths.

The pulsar case is, of course, different. With the present ensemble of known pulsars, the experiments must be performed at low frequencies, i.e., less than 1 GHz, where the plasma delays and scintillations are severe. The delays due to steady-state plasma can, in principle, be measured from the dispersion effects at several frequencies. An educated guess at this time, however, is that these measurements will not rival current space probe or interferometry experiments in the near future.

Appendix

I. First-Order Approximations

The exact expressions for propagation in a spherically symmetric medium for an arbitrary index of refraction n(r) are given by Eqs. (2), (3), and (4). However, in all of the applications of interest here, the refractive index differs from unity by $\sim 10^{-5}$. Clearly, the quantities of interest can be accurately computed from approximations of the exact equations to first order in the refractivities. Furthermore, our expressions for the refractivities themselves are accurate only to first order; i.e., we neglect terms in the gravitational field of order higher than GM/c^2 and magnetic field effects in the plasma terms.

We will fully develop the analysis for the relativistic term; the plasma terms can be treated identically. In order to save space, we will write only the expressions for one portion of the ray path, from body 1 to the perihelion point of the ray. Identical terms for body 2 must be added to all the answers. Starting with the differential equation for a ray in a spherical medium, the total heliocentric angle between the earth and the perihelion point of the ray is

$$\theta_1 = \int_{r_0}^{r_1} \frac{b\,dr}{r\sqrt{r^2n^2 - b^2}} \qquad (A\text{-}1)$$

We make the change of variables

$$z = rn = r(1 + \eta(r)) \qquad (A\text{-}2)$$

which can be written to first order in $\eta(r)$ as

$$r = z(1 - \eta(z))$$

where η is the sum of all the refractivity terms. Then, Eq. (A-1) becomes

$$\theta_1 = \int_b^{r_1 n_1} \frac{b\left(1 - z\frac{d\eta}{dz}\right)dz}{z\sqrt{z^2 - b^2}} \qquad (A\text{-}3)$$

where $n_1 = n(r_1)$. We cannot assume that the index of refraction is unity at body 1. Equation (A-3) becomes

$$\theta_1 = \cos^{-1}\frac{b}{r_1 n_1} - b\int_b^{r_1 n_1} \frac{\left(\frac{d\eta}{dz}\right)dz}{\sqrt{z^2 - b^2}} \qquad (A\text{-}4)$$

It can be shown that in all integrals whose integrands contain $\eta(z)$ or $d\eta/dz$, the b should be replaced by p in the first-order analysis. Obviously, θ_1 is also equal to

$$\theta_1 = \cos^{-1}\frac{p}{r_1} \qquad (A\text{-}5)$$

Upon equating Eqs. (A-4) and (A-5), we find that the relationship of the impact parameter of the ray asymptote b with that for the geometrical straight line p is

$$b = p(1 + \eta(r_1)) - p\sqrt{r_1^2 - p^2}\int_p^{r_1} \frac{\left(\frac{d\eta}{dz}\right)dz}{\sqrt{z^2 - p^2}} \qquad (A\text{-}6)$$

The relationship for the gravitational field alone, i.e., $\eta_R = 2m/r$, is

$$b_R = p + \frac{2mr_1}{p} \qquad (A\text{-}7)$$

II. Ray Bending

The first-order expression for the ray bending can be written immediately from Eq. (4), since it is an expression in terms of $d\eta/dr$ directly. Thus,

$$\psi_1 = -p\int_p^{r_1} \frac{\frac{d\eta}{dz}dz}{\sqrt{z^2 - p^2}} \qquad (A\text{-}8)$$

This equation illustrates the important result that the total ray bending is the linear sum of the bending angles from each refractivity term to first order. For the gravitational field alone, we get

$$\psi_{1R} = +\frac{2m}{p}\left(\frac{\sqrt{r_1^2 - p^2}}{r_1}\right) \qquad (A\text{-}9)$$

and the observed bending would be $\psi = \psi_{1R} + \psi_{2R}$.

III. Time Delay

The exact expression for the time delay is given by Eq. (3) (in length units):

$$\rho_1 = \int_{r_0}^{r_1} \frac{n_g n r dr}{\sqrt{r^2 n^2 - b^2}}$$

The two cases with n(r) greater or less than unity are best treated separately. For $n \geq 1$, $n_g = n$. Upon making the change of variables,

$$z^2 - p^2 = r^2 n^2 - b^2 \qquad (A\text{-}10)$$

the time delay becomes

$$\rho_1 = \int_p^{z_u} \frac{z dz}{\sqrt{z^2 - p^2}} - \int_p^{z_u} \frac{z^2\frac{d\eta}{dz}dz}{\sqrt{z^2 - p^2}}$$

where the upper limit is given by

$$z_u = \sqrt{r_1^2 n(r_1^2) - b^2 + p^2} \qquad (A\text{-}11)$$

from which b can be eliminated by use of Eq. (A-6) for the general case. The time delay for the gravitational field can be computed using Eq. (A-7) in Eq. (A-11) for which $z_u = r_1$. We then get

$$\rho_1 = \sqrt{r_1^2 - p^2} + 2m \ln\frac{r_1 + \sqrt{r_1^2 - p^2}}{p}$$

The total (one-way) time delay is $\rho = \rho_1 + \rho_2$, which, after some elementary algebra, yields

$$\rho = \rho_{geom} = 2m \ln\frac{r_1 + r_2 + \rho_{geom}}{r_1 + r_2 - \rho_{geom}} \qquad (A\text{-}12)$$

where

$$\rho_{geom} = \sqrt{r_1^2 - p^2} + \sqrt{r_2^2 - p^2}$$

Thus, in isotropic coordinates, the relativistic effect appears in a single term when the geometrical time delay is also computed in these coordinates.

We now consider the group time delay due to the electron plasma. To first order in the

refractivity, $n_{gn} = 1$ in Eq. (3), since the medium is dispersive and $\eta(z)$ is negative. Then, the time delay becomes

$$\rho_{p_1} = \int_{r_0}^{r_1} \frac{r\,dr}{\sqrt{r^2n^2 - b^2}} \qquad \text{(A-13)}$$

Using the change of variables, Eq. (A-10), in this expression and writing $\beta(r)$ for $\eta(r)$,

$$\rho_{p_1} = \int_{p}^{z_u} \frac{z\,dz}{\sqrt{z^2 - p^2}} - \int_{p}^{z_u} \frac{2z\beta(z)\,dz}{\sqrt{z^2 - p^2}} - \int_{p}^{z_u} \frac{z^2\left(\frac{d\beta}{dz}\right)dz}{\sqrt{z^2 - p^2}} \qquad \text{(A-14)}$$

where z_u is computed from Eq. (A-11) and b from Eq. (A-6). In general, $z_u \neq r_1$, and the first integral in Eq. (A-14) becomes

$$\sqrt{z_1^2 - p^2} + \beta(r_1)\sqrt{r_1^2 - p^2} + p^2 \int_{p}^{r_1} \frac{\left(\frac{d\beta}{dr}\right)dz}{\sqrt{z^2 - p}}$$

The integral in this expression is added to the third integral in Eq. (A-14), yielding an expression which can be integrated by parts. This results in

$$\rho_{p_1} = \sqrt{r_1^2 - p^2} - \int_{p}^{r_1} \frac{z\beta(z)\,dz}{\sqrt{z^2 - p^2}} \qquad \text{(A-15)}$$

or

$$\rho_{p_1} = \int_{p}^{r_1} \frac{z n_g(z)\,dz}{\sqrt{z^2 - p^2}} \qquad \text{(A-16)}$$

The total plasma delay is then $\rho_{p_1} + \rho_{p_2}$, where the delays from each refractivity term are combined linearly.

IV. Doppler Velocity Effects

As a result of the motion of the ray path with respect to sun and, consequently, the refracting medium, doppler effects appear in both the gravitational and plasma fields. These effects are usually measured by radio-tracking the signal phase. In the case of pulsars, a doppler-like effect occurs in the form of an apparent change in the pulse period which is clearly a group-like phenomenon. Again, two cases must be distinguished: a phase retardation occurs for $n > 1$ and a phase advance for $n < 1$. We will treat only the former case; the results for $n < 1$ are identical after a change in sign.

Starting with Eq. (3), with $n_g = n$, and utilizing the same change of variables and approximation techniques as above, the time delay becomes

$$\rho_1 = [1 + \eta(r_1)]\rho_{geom} - \int_{p}^{r_1} \sqrt{z^2 - p^2}\left(\frac{d\eta}{dz}\right)dz \qquad \text{(A-17)}$$

Using the same trick as after Eq. (A-14) and differentiating with time, we get

$$\frac{dp_1}{dt} = [1 + \eta(r_1)]\left(\frac{dp}{dt}\right)_{geom} - \frac{dp}{dt}\int_{p}^{r_1} \frac{p\left(\frac{d\eta}{dz}\right)dz}{\sqrt{z^2 - p^2}} \qquad \text{(A-18)}$$

Substitution of Eq. (A-8) in (A-18) yields the desired result:

$$\frac{dp_1}{dt} = [1 + \eta(r_1)]\left(\frac{dp}{dt}\right)_{geom} + \psi_1(p)\frac{dp}{dt} \qquad \text{(A-19)}$$

That is, except for a small end-point correction, the doppler effect results from the projection of the impact parameter rate-of-change by the bending angle. The total (one-way) doppler velocity is given by

$$\frac{dp}{dt} = \frac{dp_1}{dt} + \frac{dp_2}{dt}$$

References

1. Eddington, A. S., Mathematical Theory of Relativity, Cambridge University Press, 1923.
2. Muhleman, D. O., and Johnston, I. D., Phys. Rev. Lett., 17, 455, 1966.
3. Newkirk, G., Ann. Rev. Astron. Astrophys., 5, 213, 1967.
4. Blackwell, D. E., and Petford, A. D., Monthly Notices Roy. Astron. Soc., 131, 383, 1966.
5. Muhleman, D. O., and Reichley, P., Space Programs Summary 37-29, Vol. IV, 239-241, Jet Propulsion Laboratory, Pasadena, Calif., Oct. 31, 1964.
6. Muhleman, D. O., and Reichley, P., Space Programs Summary 37-31, Vol. IV, 342-345, Jet Propulsion Laboratory, Pasadena, Calif., Feb. 28, 1965.

7. Reichley, P., and Muhleman, D. O., Space Programs Summary 37-29, Vol. IV, 229-239, Jet Propulsion Laboratory, Pasadena, Calif., Oct. 31, 1964.

8. Reichley, P., and Muhleman, D. O., Space Programs Summary 37-32, Vol. IV, 273-276, Jet Propulsion Laboratory, Pasadena, Calif., Apr. 30, 1965.

9. Muhleman, D. O., Ekers, R. D., and Fomalont, E. B., Phys. Rev. Lett., 24, 1377, 1970.

10. Seielstad, G. A., Sramek, R. A., and Weiler, K. W., Phys. Rev. Lett., 24, 1373, 1970.

11. van de Hulst, H. C., Bull. Astron. Inst. Neth., 11, 135, 1950.

12. Blackwell, D. E., Dewhirst, D. W., and Ingham, M. F., Advances in Astronomy and Astrophysics, 5, 25, Academic Press, New York, 1967.

13. Muhleman, D. O., Weiler, K. W., Fomalont, E. B., and Ekers, R. D., 1971 (in preparation).

14. Okoye, S. E., and Hewish, A., Monthly Notices Roy. Astron. Soc., 137, 287, 1967.

15. Little, L. T., private communication.

16. Anderson, J., this proceedings.

17. Fjeldbo, G., and Eshleman, V. R., Radio Sci., 4, 879, 1969.

18. Campbell, D. B., and Muhleman, D. O., J. Geophys. Res., 74, 1138, 1969.

Probing the Solar Plasma With Mariner Radio Tracking Data

P. F. MacDoran, P. S. Callahan, and A. I. Zygielbaum
Jet Propulsion Laboratory
California Institute of Technology

Shortly after their encounters with Mars in the summer of 1969, Mariners VI and VII entered an extended mission phase to take advantage of the continuing spacecraft life. With the advent of the Mariner 1969 extended mission came a new ranging system (Fast Acquisition Ranging), using a sequential binary code (Ref. 1). This new system possesses 40 times the sensitivity of the ranging system used during the primary Mariner 1969 mission. The increased sensitivity made it possible for the Mariners VI and VII spacecraft to be ranged around their entire orbit, particularly at 2.6 AU at their solar superior conjunctions. The range and doppler radio tracking close to the sun made it possible to measure the solar plasma dynamics about to be discussed.

The plasma dynamics were measured by means of a method known as differenced range versus integrated doppler (DRVID), which exploits the opposite change of group and phase velocity as the plasma density changes along the radio raypath. The method is sensitive only to changes in the columnar electron content, not to the total content. The original purpose of DRVID (Ref. 2) was to provide charged-particle calibrations for doppler tracking data; however, it has proved useful in probing the solar plasma as well.

Approximately 150 h of DRVID data suitable for calibration of doppler tracking data and for studying the solar plasma have been collected during the extended mission, covering 4 months after the Mariners' superior conjunction. From a preliminary study of these data we have found

1. Solar plasma clouds with typical sizes of 6×10^4 to 2×10^6 km at heliocentric distances of 27 solar radii.
2. Local electron density variations more than a factor of 4 above steady-state density predictions.
3. A correspondence between plasma fluctuations in the raypath and McMath sunspot regions on the solar surface.

The DRVID method is based on the apparent path differences as measured by group and phase methods in a plasma. In a dynamic plasma, the group and phase velocities are not constant but vary such that, for an increasing columnar electron content, the phase velocity increases by the same amount that the group velocity decreases. Range code modulation is propagated at the group velocity, while doppler information propagates at the phase velocity.

By comparing the path changes indicated by range differences against those found by integrating the velocity inferred from the doppler data, a remainder results that is proportional to the number of electrons which have entered or left the raypath during the interval of observation. It should be noted that such a technique cancels out all common effects in the doppler and range such as the earth's troposphere and, more importantly, the tracking station/spacecraft relative motions. Even the less common effects of general relativity and possible gravity waves are removed by this differencing technique.

As shown in Ref. 2, the DRVID function is given in MKS units by

$$\mathrm{DRVID} = \frac{40.3}{f^2} \int_{t_1}^{t_2} \int_{\text{raypath}} \frac{d}{dt} n(s,t)\, ds\, dt$$

where

DRVID = two-way range change, m

f = radio frequency, Hz

n(s,t) = space and time variable electron density, electrons/m^3

t_1, t_2 = time limits of observation span

The columnar electron content is given by $I(t) = \int_{raypath} n(s,t)\, ds$, so that the change in the columnar electron content is given by

$$\Delta I(t) = 6.007 \times 10^{16}\ DRVID(t),\ electrons/m^2$$

for the case of S-band range and doppler tracking.

The validity of the DRVID technique, using the Fast Acquisition Ranging System, was established in a limiting case in late 1969. In these tests, with the Mariner spacecraft at an angular distance of 60 deg from the sun, the earth's ionosphere was measured by DRVID and independently verified by a VHF Faraday rotation method (Ref. 3).

For observations within a few degrees of the sun, independently determined solar plasma dynamics are unavailable to check the DRVID method to the required precision. Thus, an alternate, although less rigorous approach had to be devised to test the internal consistency of the observations. The method involved eliminating the plasma contribution from the doppler data and examining the residuals for systematic effects. The internal consistency test is valid since the solar plasma excursions are assumed uncorrelated with the inherently diurnal variations that normally occur in doppler tracking data. However, if the plasma dynamics were introduced by the earth's ionosphere, the effect would be diurnal in character and, therefore, would invalidate any doppler residual analysis to check the DRVID data independently.

The simultaneous presence of an uplink and downlink in the medium between the earth and spacecraft offers an opportunity to measure the position of plasma intersections with the radio raypath. As plasma irregularities transit the raypath, they will, in general, cause a particular signature in the DRVID data. The plasma signature will arrive at the earth twice, first on the downlink and then again at a time τ^* later when the uplink signature is received. Given a plasma stream which crosses the raypath at a particular point and persists for a time comparable to the length of the observations (several hours), an autocorrelation of the DRVID data should exhibit a correlation maximum at a time shift τ^*. The domain of physically significant autocorrelation time shifts obviously extends from zero (at the spacecraft) to a round-trip light time (at the earth).

Although independently conceived by the authors, the autocorrelation technique applied to bistatic tracking was first proposed by Thiede and Lusignan (Ref. 4), in the context of phase excursions introduced into doppler tracking data by refractive index variations. The DRVID data type differs somewhat from the proposed uses of Ref. 4 since DRVID has sensitivity to only the charged particle effects and no dependence on relative spacecraft motion and neutral atmospheric effects.

Figure 1 shows columnar electron changes and the range change at S-band for an event observed in Mariner the VII DRVID data of 29 May 1970. The spacecraft range was 2.5 AU; the sun-earth-probe angle 6 deg. The received ranging power was -190 dBm, with a total uplink power of 200 kW from the 64-m antenna at Goldstone, California. The curve fit to the data points is the result of a least-squares 15th-order power series. The event itself is rare because it apparently shows plasma entering and leaving the raypath. Most excursions observed were of plasma either entering or leaving but not both. The shape of the excursion suggests the transit of two connected clouds. The change in the columnar electron content is approximately 65×10^{17} electrons/m^2. Assuming that the excursion results from two similar plasma clouds, the duration of each is about 2-1/2 h.

Approximately 25 similar events have been analyzed for data taken within a few degrees of the sun. Columnar content rates of change lasting several hours have been found up to 3×10^{15} electrons/m^2/s. It is likely that there were even larger rates of change; however, the internal consistency test showed that there was a failure to maintain doppler phase-lock in such cases. Phase-lock failures were few, only about 5% for a month around superior conjunction. For comparison, columnar content rates in the earth's ionosphere rarely exceed 10^{14} electrons/m^2/s.

Figure 2 shows the result of autocorrelating the residuals of the individual DRVID samples from the 15th-order least-squares power series fit to the data shown in Fig. 1. Autocorrelation relative maxima occur at time shifts of 18, 24, and 36 min. The absolute significance of the autocorrelation function has been distorted by the inability of the power series fit to adequately remove the long-period (150 min) oscillation in the data.

The solar plasma propagation model is quite simple. The sun rotates with an angular rate of approximately 13.4 deg per day. If heliographic latitude dependence is ignored, disturbances originating at the surface propagate radially at a speed between 300 and 350 km/s. These assumptions result in a simple spiral structure for the propagated disturbances. A specific example, the data

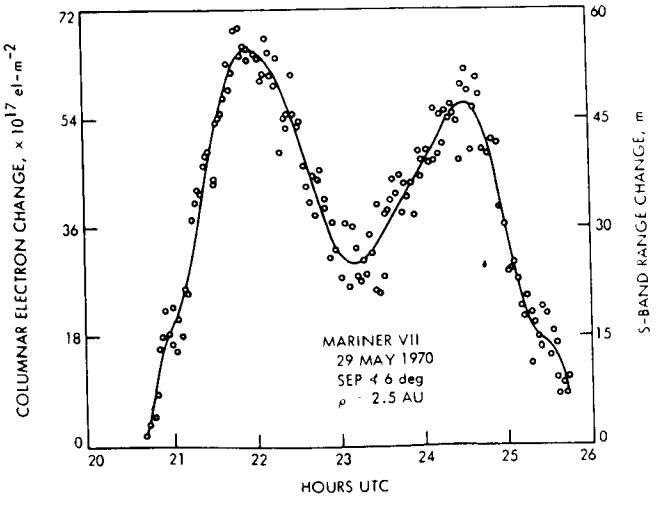


Fig. 1. DRVID measured solar plasma dynamics

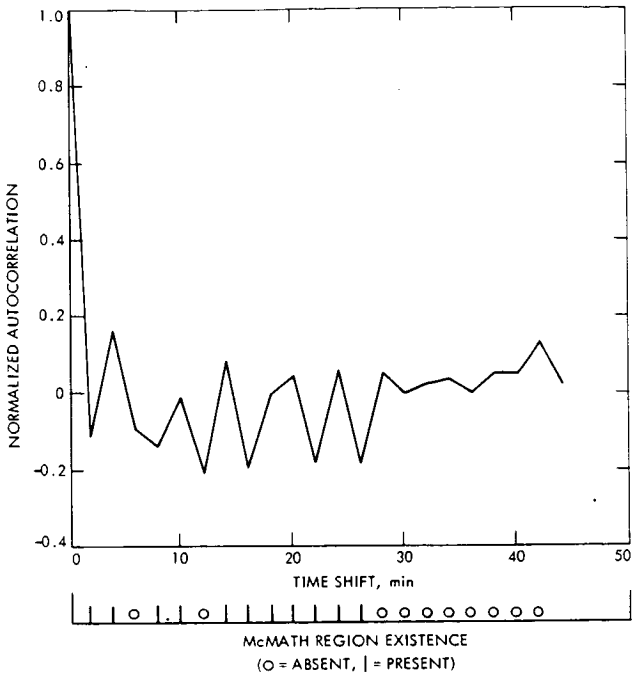


Fig. 2. Normalized autocorrelation, Mariner VII, 29 May 1970

in Fig. 2 for Mariner VII on 29 May 1970, will be helpful in illustrating the use of the model.

The correlation $\tau^* = 24$ min implies that the plasma crossed the raypath 24 round-trip min from the spacecraft, or 18 min from the earth, since the earth/spacecraft round-trip light time was 42.0 min. The radial path from the sun to the raypath plasma intersection is 0.12 AU (0.18×10^8 km) and is observed on day 149.9. If an average velocity of 320 km/s is assumed, the disturbance had to leave the solar surface on day 149.3 (29.3 May). The assumed radial motion of the plasma requires it to have departed the sun's surface at a relative earth/sun longitude of 120°W. The question is now whether or not a solar surface feature is located at such a longitude at the required time.

The NOAA* Research Laboratories, Solar-Geophysical Data (Prompt Reports) contain Hα spectroheliograms taken on a daily basis. In the spectroheliogram for 11 June 1970 at 2325 UT, McMath region 789A, B extends over a range of longitudes from 45 to 80°E. If a rotation rate of 13.4 deg/day is assumed, and if the 789 complex remains substantially unchanged, on 29.3 May the complex will be located between longitudes 96 and 131°W. Therefore, the calculated longitude of 120° is in the center of region 789. McMath region 789 is not observed to make a west limb transit before it is detected in the raypath. However, projecting ahead to the expected time of east limb transit for the plasma emitter, region 789 appears at the predicted time. Thus, it appears possible to observe the development of active regions on the back of the sun with the autocorrelation technique.

The other autocorrelation peaks at 18 and 36 min, by an analysis similar to that above, correspond to McMath regions 740 and 759, respectively. Figure 3 illustrates the plasma crossing, and its mapping to the sun.

Three important correlations were found in Fig. 4 (Mariner VI, 2 June 1970), at 4, 14, and 24 min, and a conspicuous lack of correlations is seen between 30 and 40 min. Using a velocity of 300-350 km/s, the 4-, 14-, and 24-min correlations are found to correspond to McMath regions 774, 781A, B, and 792, respectively. Region 774 is a return of region 740. It should be noted that region 740 appeared in the analysis of Mariner VII on 29 May, 3 days prior to these Mariner VI observations.

The simplest possible model to relate McMath regions to autocorrelations in the DRVID data would presume that there is a constant stream of material emitted from each region. Because of

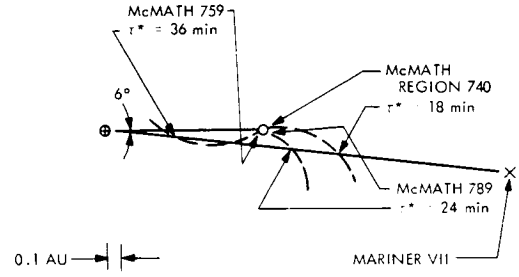


Fig. 3. Plasma transit/McMath region correspondence, Mariner VII, 29 May 1970

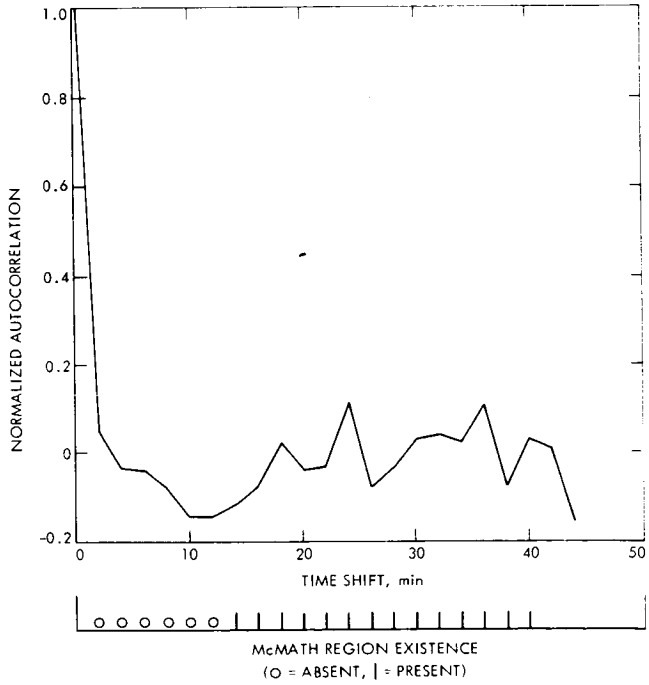


Fig. 4. Normalized autocorrelation, Mariner VI, 2 June 1970

*National Oceanic and Atmospheric Administration, Space Disturbances Laboratory, Boulder Observatory, Boulder, Colo.

the peculiar mapping involving the earth-spacecraft geometry, the plasma velocity, and the solar rotation, the correspondence between the correlation time shifts and the sun's surface is not immediately obvious. Tables 1 and 2 contain the relative longitudes and dates of ejection from the sun if material were to cause correlations at the specified time shifts. In the fourth column of the tables is an indication of which, if any, McMath region is located at the longitude-time pair on Hα spectroheliograms. These results are plotted at the bottom of Figs. 2 and 4 so that the correspondence between correlation features and McMath regions can be seen more clearly.

Figures 2 and 4 show that (1) for regions of the correlation that are particularly featureless, there is a paucity of McMath regions, and (2) if there is a local correlation maximum, there is a corresponding McMath region. The other time shifts for which there are McMath regions but no notable correlation maxima indicate that this simple model is not completely satisfactory; time variations in the emission from the active regions are a virtual certainty. Thus, having a McMath region at the appropriate longitude and time is necessary, but, because of temporal variations, not sufficient to observe a correlation maximum in the DRVID data.

Table 1. Correlation times τ^* mapped to longitude-time pairs on the sun's surface (Mariner VII, 29 May 1970)

τ^*, min	λ ejection (relative to earth-sun line), °W	T ejection, day number, 1970	Identification (McMath region number)
2	170.2	143.1	None found
4	169.8	143.7	None found
6	169.3	144.3	None found
8	168.8	144.7	None found
10	168.0	145.3	None found
12	167.1	145.9	None found
14	165.8	146.4	740A
16	164.0	147.0	774-740A, B[a]
18	161.2	147.7	774-740B
20	156.5	148.2	774/741B[b]
22	146.7	148.8	781A, B-743
24	119.9	149.3	789A, B
26	58.6	149.4	753
28	25.1	148.9	Unnumbered
30	13.5	148.3	760A, B
32	8.1	147.7	Unnumbered
34	5.1	147.1	Unnumbered
36	3.1	146.5	759
38	1.7	145.9	759/753
40	0.8	145.3	753

[a](-) returning McMath region.
[b](/) other possible region.

Table 2. Correlation times τ^* mapped to longitude-time pairs on the sun's surface (Mariner VI, 2 June 1970)

τ^*, min	λ ejection (relative to earth-sun line), °W	T ejection, day number, 1970	Identification (McMath region number)
2	161.6	146.3	774-740A, B[a]
4	160.9	146.9	774-740A, B
6	160.2	147.5	None found
8	159.2	147.9	780A, B/740B[b]
10	158.1	148.5	780C/741B
12	156.6	149.1	None found
14	154.7	149.6	781A, B-743
16	152.2	150.2	743
18	148.6	150.8	789A
20	143.1	151.3	789A, B
22	134.2	151.8	789B, C/748
24	118.2	152.3	792
26	89.9	152.6	796/759
28	55.8	152.5	None found
30	32.6	152.2	None found
32	19.9	151.6	None found
34	12.6	151.1	None found
36	7.9	150.5	None found
38	4.8	149.9	None found
40	2.5	149.3	None found
42	0.8	148.9	None found

[a](-) returning McMath region.
[b](/) other possible region.

For the case of Mariner VII on 29 May 1970, the presence of three solar surface disturbances contributing to the columnar electron dynamics complicates the interpretation; however, an order of magnitude estimate of the parameters for the clouds causing the large changes is possible. For purposes of estimation, let us assume that the large-scale dynamics of the double cloud were contributed by the region with the smallest heliocentric distance to the raypath, McMath region 789. The velocity transverse to the raypath is smaller than the average velocity of 320 km/s by the dot product of the radial plasma velocity vector and the impact vector. Thus, the transverse velocity is v_T = 260 km/s, implying that the clouds have a transverse dimension of approximately 2.3×10^6 km. Assuming the longitudinal dimensions of the cloud to be comparable, an estimate can be made for the change in the average electron density. The columnar content is observed to increase by 65×10^{17} electrons/m^2. Given a cloud of size 2×10^6 km, it follows that the average density must have changed by 3×10^9 electrons/m^3 (3000 electrons/cm^3). If one assumes that these plasma dynamics occurred at the position indicated by the 24-min correlation peak, then 3000 electrons/cm^3 variations can occur at 0.12 AU (27 solar radii).

It is of interest to compare this variation with the steady-state electron density predictions for the solar corona. Various models for the solar corona (Refs. 5, 6, 7) predict substantially the same electron density at 0.12 AU, approximately 700 electrons/cm^3. Thus, a variation in density

of 3000 electrons/cm^3 is rather unexpected. There is, of course, the possibility that the plasma dynamics resulted from one of the other two McMath regions, 740 or 759. The plasma from both those regions intersects the raypath at heliocentric distances of 0.4 and 0.6 AU, respectively. The implied cloud dimensions would be 0.6×10^6 km, with an average electron density variation of 10^4 electrons/cm^3. Since the steady-state electron density at 0.5 AU is expected to be between 50 and 100 electrons/cm^3, a variation of $\geq 10^4$ electrons/cm^3 seems hard to accept.

The DRVID autocorrelation peaks often appear to be preceded and/or followed by relative negative correlations, as can be seen in Figs. 2 and 4. These relative minima accompanying maxima could be physically interpreted as indicating that the plasma irregularities causing the correlations are of a compression/rarefaction nature. Determining the detailed structure of the irregularities is not possible because the data frequency is limited to 2 min per sample. The indications are, however, that the correlation width of the irregularities is approximately 4 min. If the irregularities have a simple rectangular electron density compression/rarefaction structure, their autocorrelation would have triangular shape. Plasma compressions associated with velocity waves have been observed by Neugebauer and Synder (Ref. 8) in data from Mariner II. The deviations from the triangular pattern are possibly caused by data noise and the oversimplified model. The width of the correlation triangle measures the irregularity size divided by the transverse velocity. Using v_T = 260 km/s and a correlation time scale of 240 s, it is found that the irregularities have a typical size of 6×10^4 km. Because of the data sampling rate, the correlation time scale will always be approximately 240 s. If the apparent transverse velocity changes because of geometry, the correlation will be sensitive to a different size of plasma irregularity. The maximum size observable from the correlation is approximately 10^5 km, while the minimum is about 2×10^4 km.

The plasma propagation speed also appears to increase with the increasing heliocentric distance. For heliocentric distances to plasma intersections with the raypath less than 0.5 AU, an average propagation speed of 320 km/s appears to fit the model best. However, for heliocentric distances between 0.5 and 1 AU, the best speed seems to be 350 km/s.

The data obtained from the superior conjunctions of Mariners VI and VII have provided a previously unavailable opportunity to investigate plasma effects near the sun. No special equipment aboard the spacecraft was required to make the measurements. The necessary doppler and ranging transponders are likely to be a part of future planetary missions, making further investigations possible.

The observed plasma effects were larger than expected. In one special case, the size of the plasma cloud was estimated to be 2×10^6 km, and a density change was calculated to be at least 3000 electrons/cm^3, more than a factor of 4 times the predicted steady-state density.

The size of plasma clouds observed are in the range of 6×10^4 to 2×10^6 km. The latter is a deduction dependent only on the plasma's transverse velocity. The smaller size is inferred from the width of autocorrelations in the data. Because the correlation technique is sensitive to different cloud sizes at different points along the raypath, it appears that, with sufficient data, the spectrum of sizes in the range 10^4 to 10^6 km could be deduced. The correlations have a structure that suggests the electron density variations may be of a compression/rarefaction type.

With a simple model of a rotating sun and radial disturbance propagation at average velocities of 300 to 350 km/s, the observed correlations can be mapped to active McMath regions on the sun's surface. Under 0.5 AU, the average plasma velocity appears to be 320 km/s, while between 0.5 and 1 AU, the velocity appears to increase to 350 km/s. McMath regions that develop on the backside of the sun are observed via the autocorrelation and their times of east limb transit predicted to within less than one-half day. Furthermore, if the autocorrelation is observed to be particularly featureless, it is found that there are no active regions in the appropriate area of the sun's surface. From the evidence presented, there appears to be a probable correspondence between McMath sunspot regions on the solar surface and large solar wind variations.

References

1. Martin, W. L., Space Programs Summary 37-57, Vol. II, 72-81, Jet Propulsion Laboratory, Pasadena, Calif., May 31, 1969.

2. MacDoran, P. F., Space Program Summary 37-62, Vol. II, 28-34, Jet Propulsion Laboratory, Pasadena, Calif., Mar. 31, 1970.

3. MacDoran, P. F., and Martin, W. L., Space Programs Summary 37-62, Vol. II, 34-41, Jet Propulsion Laboratory, Pasadena, Calif., Mar. 31, 1970.

4. Thiede, E. C., and Lusignan, B. V., "A Technique to Study Randomly Varying Media," IEEE Trans. on Antennas and Propagation, AP-18, 1, Jan. 1970.

5. Whang, T. C., and Chang, C. C., J. Geophys. Res., 70, 4175, 1965.

6. Hartle, R. E., and Sturrock, P. A., Astrophys. J., 151, 1155, 1968.

7. Hundhausen, A. J., and Gentry, R. A., J. Geophys. Res., 74, 2908, 1969.

8. Neugebauer, M., and Snyder, C. W., J. Geophys. Res., 71, 4469, 1966.

Acknowledgments

The authors thank J. H. Bieging and R. N. Wimberly for early developmental work on the DRVID technique; J. G. Williams for stimulating discussions; W. L. Martin for development of the Fast Acquisition Ranging System; and the Mariner 1969 Extended Mission operations team. The authors also gratefully acknowledge the sponsorship by the NASA Office of Tracking and Data Acquisition of the DRVID technique.

3

A Measurement of the General Relativistic Time Delay With Data From Mariners 6 and 7

John D. Anderson, Pasquale B. Esposito,
Warren L. Martin, and Duane O. Muhleman
Jet Propulsion Laboratory
California Institute of Technology

I. Introduction

On June 13, 1966 the National Aeronautics and Space Administration (NASA) approved a Celestial Mechanics Experiment for the Mariner Mars 1969 mission. The objectives of that experiment were to use the range and doppler tracking data obtained by the NASA/Jet Propulsion Laboratory (JPL) Deep Space Network (DSN) to obtain new information on the ratio of the mass of the earth to that of the moon, the gravity field of Mars and the ephemerides of Mars and the earth. At the same time it was realized that it should be possible to carry out a relativity propagation and solar coronal plasma experiment with the Mariner 1969 range data during solar conjunction. However, in the summer of 1966 there were some serious doubts as to whether ranging to the Mariner spacecraft at distances of 2 astronomical units (AU) or more would be possible.

These doubts, as it turned out, were not justified. New developments in the ground-based ranging equipment made it possible to obtain ranging data within 1 deg of the sun. In addition, even though the regular Mariner Mars 1969 mission ended 3 months after encounter with Mars (i.e., about November 1, 1969) there was every reason to expect that the Mariner spacecraft could be tracked for a far longer period, probably at least until the end of 1970. Realizing that this was the case, on December 8, 1969, NASA approved an extension to the Celestial Mechanics Experiment for the purpose of performing a general relativistic time delay experiment. Tracking passes were scheduled by JPL to provide essentially continuous coverage of both Mariners 6 and 7, at least from a celestial mechanics point of view, until the end of 1970. This extended coverage permitted an accurate determination to be made of the orbits of the Mariners, and consequently, by comparing the actual distance to the spacecraft with the apparent distance, it was possible to measure the relativistic time delay during the time of solar conjunction (April 29, 1970, for Mariner 6 and May 10, 1970, for Mariner 7).

With the Celestial Mechanics Experiment now extended to a period of almost 2 years (i.e., from launch on February 24, 1969, to the end of 1970) it made sense to segment the experiment into three orbital arcs, each with its own objectives. In addition, there was a real criterion for the separation of arcs. During the encounters of Mariners 6 and 7 with Mars, an infrared spectrometer (IRS) was operated in a cryogenic environment. This environment was produced by expelling nitrogen and hydrogen gas through the IRS into space. As a result, significant nongravitational forces on the order of 100 dyn or more acted on the spacecraft during the encounter period. The total capability of the IRS gas jetting system to produce a velocity change in the trajectory of the spacecraft was on the order of 0.1 m/s. Consequently, it was practically impossible to fit one orbital arc all the way from launch

through encounter and beyond. It was primarily for this reason that the analysis of the tracking data was partitioned into three orbital arcs.*

The first arc included the data taken during the orbital transfer between the earth and Mars, the second arc encompassed the encounter with Mars, and the third arc, which was the one of interest to the relativity experiment, extended from a time 5 days after the closest approach to Mars (July 31, 1969, for Mariner 6 and August 5, 1969, for Mariner 7) to the end of the usable data in the middle of December 1970.

A tracking schedule which would assure good determinations of the orbits of the Mariners was established shortly after encounter with Mars. It specified that horizon-to-horizon range and doppler passes would be obtained on the basis of one per week from the start of the post-encounter orbit to December 20, 1969; two horizon-to-horizon passes per week from December 20, 1969, to February 10, 1970; one pass every other day from February 10, 1970, to April 15, 1970; and one pass every day from April 15, 1970, to May 25, 1970, during the critical period about solar conjunction. After conjunction, the tracking frequency would revert to one pass every day from May 25, 1970, to July 1, 1970, and then to two passes per week from July 1, 1970, to January 1, 1971, when tracking was scheduled to terminate.

During the tracking periods, there were two types of data of importance to the relativity experiment. The first was phase-coherent integrated doppler data which were transmitted at an S-band frequency of about 2.2×10^9 Hz. The second type of data was provided by a new ranging system conceived by R. Goldstein and implemented by W. Martin of JPL for the relativity experiment (see Section II). The ranging system went into operation on September 30, 1969, and provided accurate range measurements through solar conjunction at a distance of 400 million km. A new high-power transmitter was put into operation just a few days before the solar conjunction of Mariner 6. As a result, range data from both Mariners 6 and 7 were obtained within 1 deg of the sun through the solar corona with a transmitted power of 200 kW. These measurements could not have been made with the older 20-kW transmitter. However, in order to obtain the greater power for the solar conjunction, it was necessary to abandon the collection of range data in the month of February while modifications were made to the 210-ft antenna at Goldstone. It was possible, however, to obtain doppler data from an 85-ft antenna during this period.

It is difficult to assess the accuracy of the two types of data available for this experiment, because there is a severe noise contribution from the solar corona as the two spacecraft near conjunction with the sun. At reasonably large separation angles between the sun and the spacecraft, the residuals on the doppler data, after the best-fit orbit has been obtained, indicate that they are accurate to 0.002 Hz at S-band over a 10-min integration time. In terms of an averaged range-rate error, this is better than 0.2 mm/s. With the 200-kW transmitter, the ranging data indicate RMS residuals on the order of 12 m at reasonably large separation angles between the spacecraft and the sun. It should be noted, however, that the range machine was operated only at distances at which there was a severe signal-to-noise problem. An assessment of how well it can perform in regions where the signal-to-noise ratio is significantly greater than for Mariners 6 and 7 must await tracking of future spacecraft, or perhaps the tracking of Mariner 6 in 1971 when it is near opposition. Mariner 7 cannot be tracked at this time because the attitude control system will not operate. The attitude control gas was depleted in December 1970. However, Mariner 6 should be operable until May 1971.

II. Technique for Obtaining Range Measurements of Mariners 6 and 7

The extended mission relativity experiment imposed two stringent requirements on the range measurement equipment. Foremost was a capability for detecting the very low signal levels which resulted from the use of the spacecraft's omnidirectional antenna. At the critical period, near superior conjunction, the ranging signal declined to -203 dBm. The second requirement was a high internal stability for the measurement of charged-particle and solar-plasma dynamics to aid in their separation from relativistic effects.

Range is determined by measuring elapsed time between the transmission of an encoded signal from the ground station and its return from the spacecraft. The spacecraft carries a turnaround transponder which retransmits the received ranging code on a slightly different carrier frequency. The ultimate precision is dependent upon the accuracy and stability of a frequency standard at the ground station which quantizes time into measurable increments.

A rubidium vapor oscillator serves as the primary reference. Its stability of 1 part in 10^{11} makes practical the synchronous detection of signals returned from distant spacecraft. Additionally, the standard is capable of quantifying the round-trip time of flight with uncertainties of only a few nanoseconds at sufficiently strong signal levels.

Figure 1 is a simplified block diagram of a binary-coded sequential-acquisition ranging system. A frequency synthesizer, including the rubidium reference oscillator, generates f_s (nominally 22 MHz), which is multiplied by 3 and phase-modulated by the transmitter coder. The code is generated by dividing the × 3 multiplier output by 64 and applying the result to an 18-stage binary counter. Each of the 18 binary counter outputs is individually selectable for modulating the transmitter. The period t_n of the nth component is given by

$$t_n = \frac{64 \times 2^n}{3f_s}$$

and the resultant code is coherent with the

*For a complete discussion of the pre-encounter and encounter orbit determination see JPL TM 33-469, The Mariner VI and VII Flight Paths and Their Determination from Tracking Data, December 1, 1970.

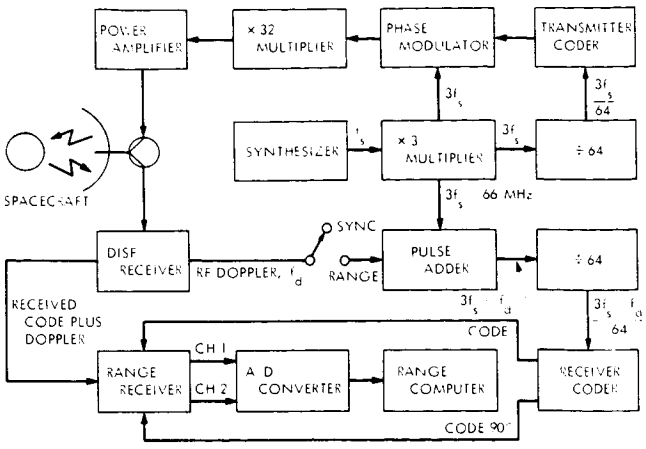


Fig. 1. Block diagram of the binary-coded sequential-acquisition ranging system

transmitted carrier. Sequential transmission of the binary-code components maximizes available power in each component and reduces the time required for a range measurement. Since the orbit determination process is sufficiently accurate to predict the round-trip time of flight to within a few hundred microseconds, no difficulty results from the serial transmission. Periods of the individual code components vary from 2 μs to 0.25 s, as specified by the equation for t_n. The shortest code component (2 μs) determines the system's resolution, while the others serve only to resolve ambiguities associated with the highest-frequency code.

Returning to Fig. 1, we note that a virtually identical set of hardware exists in the receiver. However, the +64 stage is preceded by a doppler pulse adder circuit. The pulse adder accepts the 66-MHz output of the × 3 multiplier from the transmitter chain. The second input is the RF doppler, properly scaled from the ground receiver. In the ranging mode the pulse adder's output frequency is the algebraic sum of the two input frequencies.

The receiver coder is a duplicate of its counterpart in the transmitter, except for a second code output which is phase delayed by $\pi/2$. These two receiver coder outputs are then combined to establish the amplitude of the returning signal prior to measuring its phase.

When the Range-Sync switch is in the Sync position, the two coders operate synchronously. Because of the topocentric range rate of the spacecraft, the received code slips in phase with respect to the transmitted code. If at time t_0 the switch is changed to the Range position, the receiver coder frequency is modified by RF doppler and becomes coherent with the signal being received from the spacecraft. Assuming that the two coders were synchronized prior to the changeover, the phase difference between the receiver coder and the incoming signal is a measure of the range. This phase difference will remain essentially constant by virtue of the doppler rate aiding, allowing the range measurements to be made at leisure.

Thus, a coherent model of the received range code can be generated by modifying the frequency of the transmitted code by the spacecraft's doppler signal. Note that the tracking operation is open-loop; this eliminates any settling time and greatly simplifies the hardware.

Square-wave coding produces a triangular correlation function, as shown in Fig. 2. Two channels are required to measure the phase difference between the received code and that generated by the receiver coder. The method is analogous to the optimum phase estimator for sine waves. Both channels are sampled every 10 ms and the results summed over some integration time t_i. After sufficient samples have been accumulated, the phase displacement can be computed using the relationship

$$\tau = \frac{\sum_{t=0}^{t_i} V_{CH1}}{\sum_{t=0}^{t_i} V_{CH1} + \sum_{t=0}^{t_i} V_{CH2}} + \frac{T}{4}, \quad 0 < \tau < \frac{\pi}{2}$$

The above equation holds only for the first quadrant, and similar expressions are necessary for the others.

Figure 3 shows the relative code correlation characteristics of the first three components with the quadrature channel deleted. Suppose that the spacecraft's range results in $\tau = \tau_R$. A measurement of the first component's position on its correlation function determines the phase angle between the received code and the local coder. The latter is now shifted to bring it into phase with the code received from the spacecraft. This is equivalent to shifting to a positive peak (point A) on the correlation function.

A number corresponding to the magnitude of that shift is entered into the range tally. Note that whenever the system is tracking at a positive peak, the next lower frequency code component will also be at a peak, either positive or negative. A negative peak implies that the received and local codes are out of phase, and a shift of one half of that component's period will be required to align them. Conversely, a positive peak indicates code synchronization, and no shifting is required. Whenever a shift takes place, the range tally is incremented appropriately.

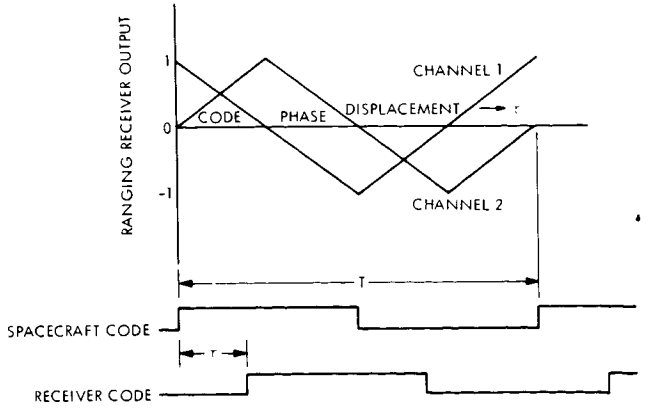


Fig. 2. Ranging receiver output characteristics

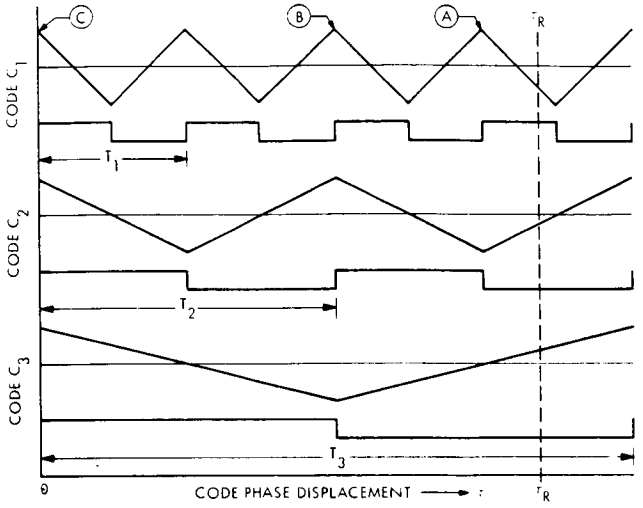


Fig. 3. Relative code correlation characteristics

Returning to the example of Fig. 2, we observe that the process causes the system to move from point A to point B to point C. At that time, the first three components of the receiver coder will be in phase with those received from the spacecraft. If this procedure is continued until all components have been synchronized, the range tally will contain a measure of the original phase displacement between transmitted and received codes. Since the phase shift was caused by the finite round-trip time to the spacecraft, it is obvious that a phase measurement is equivalent to a range measurement.

An earlier observation that the orbit determination process can predict the range to within a few hundred microseconds makes it unnecessary to constrain the code's length to more than one round-trip time. Practically, the relativity experiment operated with a ten-component code whose total period was just under 1 ms. It is irrelevant that the code repeats millions of times in the 45 min required for the signal to reach the spacecraft and return; the number of repetitions is predictable. The important data are the remaining fractional period which can not be forecast. By utilizing available information, ranges of up to 400 million km have been measured to an accuracy on the order of 12 m in only a few minutes of time on the 210-ft antenna.

Figure 4 specifies the time required to complete a ranging measurement as a function of the received signal power. Curves representing three different system noise temperatures are shown. The times indicated are based upon the acquisition of 18 code components. By reducing the number of components, it is possible to proportionally diminish the time necessary to complete the measurement. For example, by using ten components, a ranging measurement at -200 dBm can be completed in only 33 min for a typical system noise temperature N_T of 30 K and with a probability P_E of an erroneous range acquisition of 0.01.

In summary, this section has described a system designed to extend the ranging threshold to weaker signals than has been possible previously without sacrificing accuracy. Improved performance results from square wave coding,

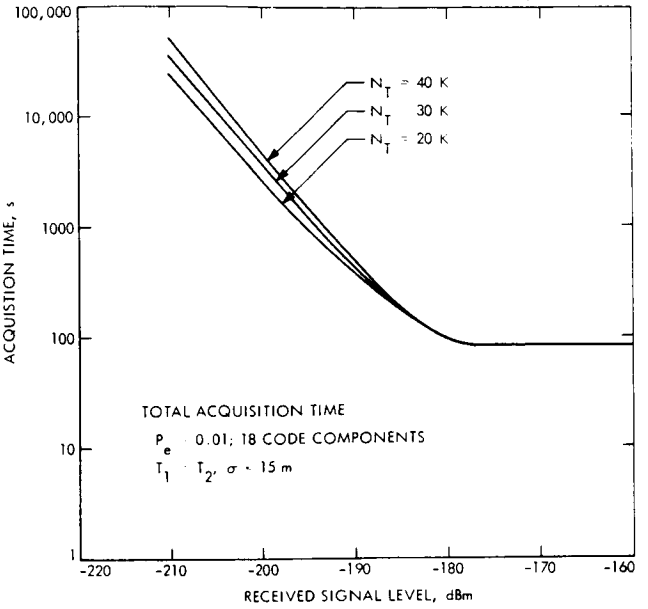


Fig. 4. Acquisition times for an 18-component ranging code

sequential transmission of code components, and open-loop operation. To date the system has collected several hundred range points on two spacecraft. Analysis of the data indicates that the measurement uncertainties from day to day are on the order of 12 m when 200 kW of transmitted power are used. This error is almost certainly caused by long-term drifts over several hours in the spacecraft transponder. Calibrations of future transponders to account for variations in temperature and signal level could improve the accuracy in range by at least an order of magnitude over the Mariner 6 and 7 data.

III. Theoretical Background for the Mariner Mars 1969 Relativity Test

The testing of general relativity by astronomical techniques is, of course, not new. The three classical tests are well known. They depend first on orbital motions through the precession of the perihelion of Mercury, second on the bending of starlight as it passes the sun during a total eclipse, and third on the red shift of lines in the spectrum of the companion of Sirius, a white dwarf. Of these three tests, only the precession of the perihelion of Mercury is capable of providing relatively high accuracies on the order of 1%. However, recently it has been shown (Ref. 1) that the sun's disk is sufficiently oblate to suggest that a good portion (about 10%) of the precession of the perihelion, which was formerly attributed to general relativity alone, could be explained equally well by a quadrupole moment in the sun's gravitational field. There is disagreement on the magnitude of the quadrupole moment which can be inferred from the oblateness measurements (Ref. 2), and Ingersoll and Spiegel (Ref. 3) have recently shown that the observed oblateness could be the result of temperature effects alone. In any case, the current difficulties with the solar oblateness make the testing of Einstein's theory by the classical methods uncertain to the level of about 10%. The more recent methods provided by planetary radar measurements have refined this figure to the level of about 5% (Ref. 4), and the Mariner experiment

discussed here is currently providing a 4% test of the theory.

According to the Newtonian theory of gravitation, the propagation of electromagnetic waves near massive bodies is unaffected by the gravitational field. However, in the theory of general relativity, the waves appear to propagate at speeds less than the velocity of light. The exact nature of the effect of the solar gravitational field on such signals has been computed by Muhleman and Reichly (Ref. 5) and by Shapiro (Ref. 6) for the case of the Schwarzschild solution of the Einstein field equations. The possibility of a measurement of the general relativistic delay in the propagation of radio signals was first suggested several years ago by Shapiro (Ref. 7). The observational reality of the delay was subsequently analyzed very carefully by Ross and Schiff (Ref. 8) and by Shapiro (Ref. 9).

The magnitude of the propagation delay in range can be expressed in several different coordinate systems. The simplest equation has been obtained in isotropic coordinates by Holdridge (Ref. 10) and Tausner (Ref. 11). In order to parameterize the magnitude of the relativity effect, we have used a general isotropic, static, spherically symmetric metric with two free parameters, β and γ, as suggested by Eddington (Ref. 12) and Robertson (Ref. 13). Orbital calculations, time transformations, and propagation effects have been computed within the framework of this generalized metric. It is given by

$$ds^2 = \left[1 - 2\frac{m}{r} + 2\beta\left(\frac{m}{r}\right)^2 + \cdots\right] c^2 dt^2 - \left[1 + 2\gamma\frac{m}{r} + \cdots\right](dx^2 + dy^2 + dz^2) \tag{1}$$

In terms of this metric, the total one-way light time is written in the following form:

$$c\Delta t = \rho + (1 + \gamma)m \ln \frac{r_0 + r_1 + \rho}{r_0 + r_1 - \rho} \tag{2}$$

where ρ, r_0, and r_1 are coordinate distances representing, respectively, the observer-spacecraft distance, the sun-observer distance, and the sun-spacecraft distance. All distances are computed from orbits (geodesics) defined by the generalized metric. As such, they depend implicitly on both β and γ.

The parameter m is equal to one-half the gravitational radius of the sun and is equal to

$$m = \frac{GM_\odot}{c^2} = 1.4766 \text{ km} \tag{3}$$

All distances in this paper are expressed in laboratory metric units, but it should be realized that the basic unit of length is the light second. Metric distances are obtained by multiplying the distances in light-seconds by the IAU adopted value of the speed of light (299792.5 km/s). Questions of coordinate systems and the conversions of clock time measurements to proper times can be treated in a rather straightforward way. It is necessary to relate clock times and the coordinate time t used in the gravitational theory (Ref. 14). What is important is to make certain that all calculations are performed in the same coordinate system, in this case the isotropic coordinates of Eddington and Robertson.

A. Computation of Orbits

The Lagrangian L for the orbital motion is related to the metric given in Eq. (1) by

$$L = -c^2 \frac{ds}{dt} \tag{4}$$

which is obtained by expanding the square root of $(ds/dt)^2$ in powers of $1/c^2$. The result, to powers of $1/c^4$ in ds/dt, is

$$L = -c^2 + \frac{GS}{r} + \left(\frac{1}{2} - \beta\right)\frac{GSm}{r^2} + \frac{1}{2}v^2\left[1 + 2\left(\gamma + \frac{1}{2}\right)\frac{m}{r}\right] + \frac{1}{8}\frac{v^4}{c^2} \tag{5}$$

The speed v of the spacecraft can be written in cartesian components, as can the heliocentric distance r.

$$v^2 = \dot{x}^2 + \dot{y}^2 + \dot{z}^2 \tag{6}$$

$$r^2 = x^2 + y^2 + z^2 \tag{7}$$

Therefore, the Lagrangian L can be expressed as a function of the cartesian position vector $\vec{r} = (x, y, z)$ and velocity vector $\dot{\vec{r}} = (\dot{x}, \dot{y}, \dot{z})$, and the Euler-Lagrange equations for the motion are

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{\vec{r}}}\right) - \frac{\partial L}{\partial \vec{r}} = 0 \tag{8}$$

where

$$\frac{\partial L}{\partial \dot{\vec{r}}} = \left[1 + 2\left(\gamma + \frac{1}{2}\right)\frac{m}{r} + \frac{1}{2}\frac{v^2}{c^2}\right]\dot{\vec{r}} \tag{9}$$

$$\frac{\partial L}{\partial \vec{r}} = -\frac{GS\vec{r}}{r^3}\left[1 + (1 - 2\beta)\frac{m}{r} + \left(\gamma + \frac{1}{2}\right)\frac{v^2}{c^2}\right] \tag{10}$$

The equations of motion for the Lagrangian under consideration are then

$$\ddot{\vec{r}} = -\frac{GS\vec{r}}{r^3} - \left[(1 - 2\beta)\frac{m}{r} + \left(\gamma + \frac{1}{2}\right)\frac{v^2}{c^2}\right]\frac{GS\vec{r}}{r^3}$$

$$+ \left[(1 + 2\gamma)\frac{m\dot{r}}{r^2} - \frac{v\dot{v}}{c^2}\right]\dot{\vec{r}}$$

$$- \left[(1 + 2\gamma)\frac{m}{r} + \frac{1}{2}\frac{v^2}{c^2}\right]\ddot{\vec{r}} \qquad (11)$$

The Newtonian acceleration $\ddot{\vec{r}}_N$ is simply

$$\ddot{\vec{r}}_N = -\frac{GS\vec{r}}{r^3} \qquad (12)$$

and the relativistic equations can be equated to the Newtonian acceleration plus a perturbative acceleration $\ddot{\vec{r}}_{Rel}$. The total relativistic acceleration is

$$\ddot{\vec{r}} = \ddot{\vec{r}}_N + \ddot{\vec{r}}_{Rel} \qquad (13)$$

The perturbative acceleration is of order v^2/c^2, and this will still be so if the Newtonian values of the acceleration are substituted for $\dot{v}$ and $\ddot{\vec{r}}$. The vector acceleration $\ddot{\vec{r}}$ is given by Eq. (12) and the scalar $\dot{v}$ by

$$v\dot{v} = -\frac{GS\dot{r}}{r^2} \qquad (14)$$

Thus, $\ddot{\vec{r}}_{Rel}$ can be written

$$\ddot{\vec{r}}_{Rel} = \frac{m}{r^3}\left\{\left[2(\beta + \gamma)\frac{GS}{r} - \gamma v^2\right]\vec{r} + 2(1 + \gamma)(r\dot{r})\dot{\vec{r}}\right\} \qquad (15)$$

This is the general form of the perturbative acceleration. It is added to the Newtonian acceleration to produce relativistic motion for the spherically symmetric, isotropic metric. For Einstein's theory of gravitation, $\beta = \gamma = 1$, and the acceleration is given by

$$\ddot{\vec{r}}_{Rel}(\text{Einstein}) = \frac{m}{r^3}\left[\left(4\frac{GS}{r} - v^2\right)\vec{r} + 4(r\dot{r})\dot{\vec{r}}\right] \qquad (16)$$

These equations have been derived by Moyer (Ref. 15). He also obtains relativistic equations of motion for n-bodies for an isotropic metric which reduce to the two-body equation of Eq. (16) if the relativistic effects of all bodies but the central one are neglected. In later work, Moyer (Ref. 16) has shown that the two-body equation can be used for all planets except the earth. The exceptional motion of the earth about the sun is complicated by the moon. Moyer shows that the heliocentric acceleration of the earth by the sun is still given by the two-body equation if the potential GS/r in Eq. (16) is augmented by the potential for the moon. Thus, for the relativistic perturbation in the earth's motion caused by the sun, the acceleration of Eq. (15) is written

$$\ddot{\vec{r}}_{Rel}(\oplus) = \frac{m}{r_\oplus^3}\left\{\left[2(\beta + \gamma)\left(\frac{GS}{r_\oplus} + \frac{GM}{r_{\oplus\mathbb{C}}}\right) - \gamma v_\oplus^2\right]\vec{r}_\oplus + 2(1 + \gamma)(r_\oplus\dot{r}_\oplus)\dot{\vec{r}}_\oplus\right\} \qquad (17)$$

where $r_{\oplus\mathbb{C}}$ is the distance between earth and moon.

Similarly, the relativistic solar perturbation on the moon is given by

$$\ddot{\vec{r}}_{Rel}(\mathbb{C}) = \frac{m}{r_\mathbb{C}^3}\left\{\left[2(\beta + \gamma)\left(\frac{GS}{r_\mathbb{C}} + \frac{GE}{r_{\oplus\mathbb{C}}}\right) - \gamma v_\mathbb{C}^2\right]r_\mathbb{C} + 2(1 + \gamma)(r_\mathbb{C}\dot{r}_\mathbb{C})\dot{\vec{r}}_\mathbb{C}\right\} \qquad (18)$$

The mass of the earth is E in Eq. (35), and M is the mass of the moon in Eq. (34). By combining $\ddot{\vec{r}}_{Rel}(\oplus)$ and $\ddot{\vec{r}}_{Rel}(\mathbb{C})$ with the moon-to-earth mass ratio ($\mu = M/E$), the heliocentric relativistic perturbative acceleration $\ddot{\vec{r}}_{Rel}(\oplus + \mathbb{C})$ of the earth-moon barycenter is obtained.

$$\ddot{\vec{r}}_{Rel}(\oplus + \mathbb{C}) = \frac{1}{1 + \mu}\ddot{\vec{r}}_{Rel}(\oplus) + \frac{\mu}{1 + \mu}\ddot{\vec{r}}_{Rel}(\mathbb{C}) \qquad (19)$$

In a numerical integration of the planets, Eq. (19) is used for the earth-moon system, and Eq. (15) is used for the other eight planets. The heliocentric relativistic motion of the earth and moon as separate bodies involves more than the solar relativistic perturbations. Other terms representing the relativistic acceleration of the earth caused by the moon and the relativistic acceleration of the moon caused by the earth are also important. However, unless the relative motion of the earth and moon is required to a very high accuracy, the integration of the earth-moon barycenter and the eight other planets can be accomplished using only the relativistic accelerations caused by the sun. Thus, the isotropic, static, spherically symmetric metric with arbitrary β and γ can be used for the integration of all the planets. In a higher approximation to the n-body problem, the metric is neither static nor spherically symmetric because it involves the positions and velocities of the n-bodies. What Moyer has shown with respect to the solar system is that the sun is so dominant that the relativistic interactions of the planets are negligible. Consequently, the heliocentric relativistic perturbations in planetary orbits can be described by a static

spherically symmetric metric like Eq. (1). A precise description of the integration of the planets whose ephemerides are used in the reduction of Mariner data has been given by Lieske (Ref. 17)

B. Propagation of Electromagnetic Signals

According to general relativity, electromagnetic signals will propagate along null geodesics with ds = 0, and from Eq. (1), the coordinate speed of light v is a variable given by

$$\frac{v^2}{c^2} = 1 - \frac{2m}{r} - \frac{2\gamma m}{r} - \cdots \qquad (20)$$

Now, if dρ represents an infinitesimal arc length in the three-dimensional space represented by the coordinates x, y, and z, so that

$$d\rho^2 = dx^2 + dy^2 + dz^2 \qquad (21)$$

then the coordinate speed of light can be represented by

$$v = \frac{d\rho}{dt} \qquad (22)$$

The coordinate range given as a finite time interval $\Delta t = t_1 - t_0$ can be obtained as an integral of the form

$$c\Delta t = c\int_{t_0}^{t_1} dt = \int_0^{\rho_1} \frac{c}{v}\, d\rho \qquad (23)$$

where the signal originates at a transmitter at time t_0. The ratio c/v is given by Eq. (20) as

$$\frac{c}{v} = 1 + (1+\gamma)\frac{m}{r} \qquad (24)$$

and the integral of Eq. (23) becomes

$$c\Delta t = \rho_1 + (1+\gamma)m\int_0^{\rho_1} \frac{d\rho}{r} \qquad (25)$$

The fact that the signal does not follow a straight path in the space (x, y, z) is neglected in the following as a higher-order effect, and the geometry of the propagation between a transmitter at S_0 and receiver S_1 is given by Fig. 5. From this figure, it can be seen that the heliocentric distance r of the signal, after it has propagated through a distance ρ, can be obtained from the law of cosines in the form

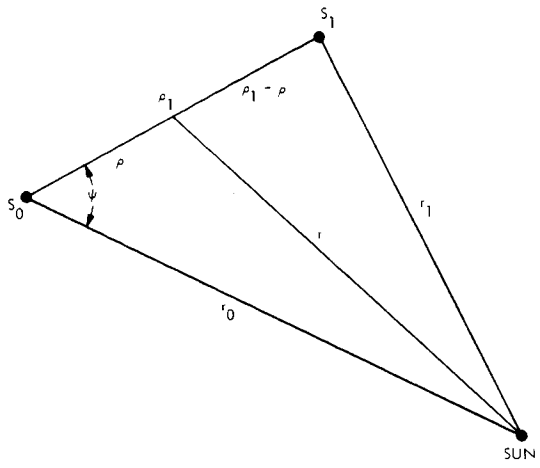


Fig. 5. Electromagnetic signal propagation

$$r^2 = r_0^2 + \rho^2 - 2r_0\rho\cos\psi \qquad (26)$$

and the integral in Eq. (25) can be written as

$$\int_0^{\rho_1} \frac{d\rho}{r} = \int_0^{\rho_1} \frac{d\rho}{\left(r_0^2 + \rho^2 - 2r_0\rho\cos\psi\right)^{1/2}} \qquad (27)$$

which, upon integration, becomes

$$\int_0^{\rho_1} \frac{d\rho}{r} = \ln\frac{r_1 + \rho_1 - r_0\cos\psi}{r_0 - r_0\cos\psi} \qquad (28)$$

The cosine of the angle ψ is obtained from Fig. 5 in terms of r_0, r_1, and ρ_1.

$$2r_0\rho_1\cos\psi = \rho_1^2 + r_0^2 - r_1^2 \qquad (29)$$

When Eq. (29) is substituted into Eq. (28), the resulting expression can be simplified to yield

$$\int_0^{\rho_1} \frac{d\rho}{r} = \ln\frac{r_0 + r_1 + \rho_1}{r_0 + r_1 - \rho_1} \qquad (30)$$

and the range cΔt can be written as

$$c\Delta t = \rho_1 + (1+\gamma)m\ln\frac{r_0 + r_1 + \rho_1}{r_0 + r_1 - \rho_1} \qquad (31)$$

This expression for range has been derived in the Einstein form ($\alpha = \gamma = 1$) by Tausner

(Ref. 11) and Holdridge (Ref. 10). Ross and Schiff (Ref. 8) give the equivalent of Eq. (30) in a different form. Again, it is important to interpret the coordinate lengths in Eq. (30) in terms of orbits computed with the perturbative acceleration of Eq. (15).

C. Time Transformation

In Sections A and B, the coordinate time has been used in the orbital calculations and the propagation equations. In order to represent data, it is necessary to introduce the transformation to the observer's proper time τ. Time at a radar station is recorded by an atomic clock, whose frequency ν_τ is related to a coordinate frequency ν_t by the ratio of intervals of proper time and coordinate time. In the limit, as these intervals become arbitrarily small, the frequency ratio is given by

$$\frac{\nu_t}{\nu_\tau} = \frac{d\tau}{dt} \tag{32}$$

Thus, if we define uniformity of time in some universal sense, it follows that, with respect to a constant coordinate frequency ν_t, the rates of various observers' clocks will vary in accordance with their coordinate velocity v and their local gravitational potential function Φ. From Eq. (1), this variation is given by

$$\left(\frac{d\tau}{dt}\right)^2 = 1 - \frac{2\alpha}{c^2}\Phi - \frac{v^2}{c^2} + \ldots \tag{33}$$

The higher-order β and γ terms are neglected because they are several orders of magnitude smaller than the uncertainties in the best atomic frequency standards. For an observer on the earth the potential function Φ includes contributions from the gravitational fields of the sun, earth, and moon. Planetary contributions are negligible. Also, we are interested only in variations in Φ, since any constant terms are absorbed in the definition of the second. Therefore, contributions to Φ from other matter in the universe can be neglected. Variations produced by galactic rotation and the proper motions of stars are of such long period that they can be considered constant.

In the expression for $d\tau/dt$, there is a general relativistic term in the gravitational potential Φ and a special relativistic term in v^2/c^2. The coefficient α in the general relativistic term is unity, but we will carry it throughout this development to make clear how the two terms enter in the time transformation. The coordinate velocity v represents the motion of the observer in the x, y, z coordinate system used to compute orbits. It involves the revolution of the earth-moon system about the sun, the revolution of the earth-moon barycenter, and the rotation of the earth on its axis. The precession of the earth's axis is another long-period term, which can be neglected. The nutation is insignificant.

The several components of the observer's motion are summarized in Table 1, and their numerical contributions to $d\tau/dt$ are summarized in parts in 10^{11}. At present, the stability of atomic clocks is about 1 part in 10^{11}, and terms smaller than this can be neglected. The largest term in Table 1, term 1, can be neglected because it is constant. This leaves terms 5, 7, and 11 in the expression for $d\tau/dt$ to achieve an accuracy of 1 part in 10^{11}. At this level, simple formulas for the time transformation can be derived which involve the diurnal and annual motions of the observer. The annual motion has been discussed by Aoki (Ref. 18) and Clemence and Szebehely (Ref. 14). For purposes of keeping time, the annual term is all that is required, but the transformation between t and τ for computing range and doppler data requires the diurnal term as well. If accuracies greater than 1 part in 10^{11} are needed in the future, all terms in Table 1 become important, with the exception of 2 and 6. It is probably not advisable to compute $d\tau/dt$ by approximate formulas like those developed here if accuracies on the order of 1 part in 10^{12} or better are required. Then, the velocity v can be computed rigorously by adding the various vector velocities for the motion.

We will not repeat the derivation of the annual term in the expression for $d\tau/dt$ because it is given in Refs. 14 and 18. It is sufficient to note that the observer's velocity v can be approximated to 1 part in 10^{11} by

$$v^2 = v_\oplus^2 + 2\vec{r}_\oplus \cdot \dot{\vec{R}} \tag{34}$$

where $\vec{r}_\oplus$ is the heliocentric vector velocity of the earth, $v_\oplus$ is the magnitude of $\vec{r}_\oplus$, and $\dot{\vec{R}}$ is the geocentric vector velocity of the observer. This expression for v can be substituted into Eq. (33). The other term in Eq. (33), the general relativistic term, is significant only in the annual motion. Other contributions from the diurnal and monthly motion are negligibly small in their periodic contributions to Φ. Also, the variations in the earth's gravitational potential over the surface of the earth are negligible with respect to atomic clocks. Thus, the potential Φ in Eq. (33) can be approximated by

$$\Phi = \frac{GS}{r_\oplus} \tag{35}$$

where $r_\oplus$ is the distance of the earth from the sun. The heliocentric velocity of the earth in Eq. (34) can be approximated by

$$v_\oplus^2 = \frac{2GS}{r_\oplus} + c_1 \tag{36}$$

where c_1 is a constant and is not important to the derivation of $d\tau/dt$. When Eqs. (34), (35), and (36) are substituted into Eq. (33), the resulting expression for $d\tau/dt$ is

$$\frac{d\tau}{dt} = 1 - \frac{(1+\alpha)m}{r_\oplus} - \frac{1}{c^2}(\dot{\vec{r}}_\oplus \cdot \dot{\vec{R}}) \tag{37}$$

Table 1. Special relativistic terms

Term	Maximum effect on time transformation, parts in 10^{11}
1. Mean annual motion about sun	493
2. Mean monthly motion about earth-moon barycenter	9×10^{-5}
3. Diurnal motion about earth's rotation axis	0.12
4. Product of annual and monthly motions	0.42
5. Product of annual and diurnal motions	15.4
6. Product of monthly and diurnal motions	6×10^{-3}
7. Periodic component in earth's orbital velocity (first order in eccentricity)	16.6
8. Periodic component in earth's orbital velocity (second order in eccentricity)	0.28
9. Effect of earth's eccentricity in term 5	0.52
10. Planetary perturbations in earth's orbital speed	0.10
11. Effect of the obliquity of the ecliptic in term 5	1.27

The diurnal term in Eq. (37) can be developed by expressing the velocity vectors $\dot{\vec{r}}_\oplus$ and $\vec{R}$ in ecliptic coordinates. The expressions for the components of the velocity of the earth are

$$\left.\begin{aligned} \dot{x}_\oplus &= v_\oplus \sin \psi_\odot \\ \dot{y}_\oplus &= -v_\oplus \cos \psi_\odot \\ \dot{z}_\oplus &= 0 \end{aligned}\right\} \tag{38}$$

where $\psi_\odot$ is the longitude of the sun measured along the ecliptic from the vernal equinox.

The geocentric velocity of the observer in ecliptic coordinates is given by

$$\left.\begin{aligned} \dot{X} &= -v_s \sin \theta \\ \dot{Y} &= v_s \cos \theta \cos \epsilon \\ \dot{Z} &= -v_s \cos \theta \sin \epsilon \end{aligned}\right\} \tag{39}$$

where v_s is the velocity of the observing station about the earth's polar axis, θ is the local sidereal time, and ϵ is the obliquity of the ecliptic.

The scalar product of the two velocity vectors results in the expression

$$\dot{\vec{r}}_\oplus \cdot \vec{R} = -\frac{1}{2} v_s v_\oplus [(1 + \cos \epsilon) \cos (\psi_\odot - \theta) - (1 - \cos \epsilon) \cos (\psi_\odot + \theta)] \tag{40}$$

The numerical values of the quantities in Eq. (40) can be approximated by

$$v_\oplus = 29.80 \text{ km/s}$$

$$\cos \epsilon = 0.9174$$

$$v_s = 0.4638 \cos \phi' \text{ km/s}$$

with the geocentric latitude given by ϕ'. Now, divide Eq. (40) by c^2 to obtain

$$\frac{1}{c^2}(\dot{\vec{r}}_\oplus \cdot \vec{R}) = -7.689 \times 10^{-11} \cos \phi' \cdot [1.9174 \cos (\psi_\odot - \theta) - 0.0826 \cos (\psi_\odot + \theta)] \tag{41}$$

The second term in $\cos (\psi_\odot + \theta)$ is smaller than 1 part in 10^{11} and can be neglected. The first term can be approximated by setting $\psi_\odot - \theta$ equal to 12:00 h minus the local mean solar time measured

from midnight. Thus, the final expression for $d\tau/dt$ is obtained by substituting the first term of Eq. (41) into Eq. (37) and by taking the annual term from Aoki (Ref. 18). The result is

$$\frac{d\tau}{dt} = 1 - 1.654(1+\alpha) \times 10^{-10} \cos M_{\odot} - 1.474 \times 10^{-10} \cos \phi' \cos (UT + \lambda) \quad (42)$$

where $M_{\odot}$ is the mean anomaly of the sun, UT is the universal time, and λ is the longitude of the station measured eastward from Greenwich.

In converting intervals of coordinate time to observed intervals of atomic time, the diurnal term is significant for the accurate computation of spacecraft range data. At 1 AU, the diurnal term in round-trip range is equal to $1.474 \times 10^{-10} \times 2 \times 1.496 \times 10^{11}$ m, or about 44 m.

For purposes of time keeping, it is necessary to integrate Eq. (42), as done by Clemence and Szebehely (Ref. 14) for the annual term. The result for both the annual and diurnal terms is

$$\tau - \tau_0 = t - t_0 - 0.^{s}8307(1+\alpha) \times 10^{-3} \left[\sin M_{\odot} - \sin M_{0\odot}\right] - 2.^{s}03 \times 10^{-6} \cos \phi' \cdot \left[\sin (UT + \lambda) - \sin (UT_0 + \lambda)\right] \quad (43)$$

The initial epochs τ_0 of atomic time and t_0 of coordinate time are arbitrary, but to conform with astronomical definitions of ephemeris time and atomic time, we set $t_0 - \tau_0$ equal to the difference of ephemeris time and UT2 time on January 1, 1958, at 0 h

$$\Delta T(1958) = t_0 - \tau_0 = 32.^{s}15$$

On January 1, 1958, the mean anomaly of the sun is about $357.^{\circ}925$, so that the sine of $M_{0\odot}$ is -0.0362. Substituting these constants into Eq. (38), the transformation between atomic time τ and coordinate time t is

$$t = \tau + 32.^{s}15 + 0.^{s}8307(1+\alpha) \times 10^{-3} \left[\sin M_{\odot} + 0.0362\right] + 0.^{s}0020 \times 10^{-3} \cos \phi' \left[\sin (UT + \lambda) - \sin \lambda\right] \quad (44)$$

This completes the specification of the equations needed to represent spacecraft tracking data in general relativity theory. They form the basis for the subsequent solutions for γ from the Mariner data.

IV. Interplanetary and Coronal Plasma Effects

Range and doppler observations at S-band are affected significantly by the electron density in the interplanetary medium and the outer solar corona. However, the associated range delays are small compared to the relativistic delay. Since the closest approach of the radio ray path to the sun is greater than 3 solar radii ($R_{\odot}$) throughout the experiment, the medium may be regarded as collisionless, and the index of refraction is given to sufficient accuracy by

$$n^2(r) = 1 - \frac{N_e(r)}{N_c} \quad (45)$$

where $N_e(r)$ is the electron density per cubic centimeter at distance r from the sun, and N_c is the critical density at the Mariner S-band frequency ($N_c \sim 6 \times 10^{10}$). The maximum possible deviation of the refractive index from unity for the Mariner experiment is on the order of 10^{-5}, and the medium can be regarded as thin. It can be shown that the group propagation time from the earth to the spacecraft is

$$\rho = \int_{r_m}^{r_{\oplus}} \frac{r\,dr}{\left[r^2 n^2(r) - b^2\right]^{1/2}} + \int_{r_m}^{r_{s/c}} \frac{r\,dr}{\left[r^2 n^2(r) - b^2\right]^{1/2}} \quad (46)$$

where r_m is the distance of closest approach to the sun of the curved ray, b is the impact parameter of the ray asymptote, and $r_{\oplus}$ and $r_{s/c}$ are the radial distances of the earth and spacecraft from the sun. Since the departure of n from unity is very small, it is permissible to work to first order in this term. After considerable reduction, the following equation for the group propagation time is obtained.

$$(\rho - \rho_g) = \frac{1}{N_c} \int_p^{r_{\oplus}} \frac{N_e(r)\,r\,dr}{\sqrt{r^2 - p^2}} + \frac{1}{N_c} \int_p^{r_{s/c}} \frac{N_e(r)\,r\,dr}{\sqrt{r^2 - p^2}} \quad (47)$$

where ρ_g is the geometrical distance and p is the distance of closest approach of the geometrical path to the sun. Strictly speaking, the actual path

followed by the radio signal results from the combined effects of the ray being displaced toward the sun due to relativistic ray bending and away from the sun due to plasma bending, the latter effect being frequency-dependent. The fact that the medium is "thin" means that the bending can be completely neglected to high accuracy in computing range-delay observables, and the components of delay due to the several terms can be combined linearly. Thus, the formula for $\rho - \rho_g$ gives the group delay due to the medium. The additional doppler shift caused by the medium (phase delay) is obtained by differentiating the negative of the group delay with respect to time. In general, the doppler velocity not only involves motions of the earth and spacecraft but is also sensitive to the rate of change of the electron density along the ray path.

Considerable information concerning the electron density profiles is available from solar eclipse measurements made during the last few decades. The intensity of the scattered light seen around the sun's limb during a total eclipse has been used to determine electron density profiles out to distances of about 40 solar radii from the sun (e.g., Ref. 19). Recent work reported by Blackwell, Dewhirst, and Ingham (Ref. 20) yields the profile

$$N_e(r) = \frac{A}{r^6} + \frac{B}{r^{2+\epsilon}} \; ; \qquad 2.5 \leq r \qquad (48)$$

where A and B are electron densities, $\epsilon \sim 0.3$, and r is expressed in solar radii (r = 215 at earth's distance). The value of ϵ in this equation is not well known but appears to be greater than zero based on spacecraft in situ measurements and solar wind theory. The coefficient A apparently varies by a factor of 2 or 3, depending on the solar latitude and the solar activity; i.e., it is a function of the sun-spot cycle. The B coefficient is much less variable and seems to be only a weak function of solar activity. A compilation of values of A and B taken from the literature is shown in Table 2. All of the values are from eclipse measurements except line 6. The numerical values of this line were obtained from measurements of ray bending near the sun with radio interferometry (Ref. 21).

The plan of the Mariner experiment was to estimate A, B, and ϵ simultaneously with the relativity parameters from the range and doppler observables. It was recognized that it might be necessary to estimate a series of these parameter sets sequentially over the total tracking period of the experiment. However, their values within a month of the solar conjunction were of primary significance to the relativity experiment. In any case, it was concluded that the plasma effects would not be the major source of error in the relativity determination. As it turned out, the major source of error arose from the nongravitational forces acting on the spacecraft (see Section V).

A difficulty with using the parameterized electron density of Eq. (48) in the least-squares solutions for the relativity parameter γ is that the three parameters A, B, and ϵ are not independent, although for lack of anything better, it is assumed that A is not correlated with either B or ϵ. With respect to the term in Eq. (48), which contains B and ϵ, the a priori uncertainty in the interplanetary electron density arises from (1) time variations in B and ϵ because of solar activity and (2) experimental error in the determination of the electron density as reflected in the discordant values in Table 2. With these two points taken into account, the following values and 1 σ uncertainties are adopted for the Mariner experiment:

$$A = (1.3 \pm 0.9) \times 10^8 \text{ electrons/cm}^3$$

$$B = (5 \pm 3) \times (215)^{2.3} \text{ electrons/cm}^3$$

$$\epsilon = 0.3 \pm 0.3$$

In order to obtain the correlation between B and ϵ, it is reasonable to constrain the uncertainty in the electron density at the distance of the earth ($215 \times R_\odot$). At this distance, the electron density is given by

$$N_e(r_\oplus) = \frac{B}{(215)^{2+\epsilon}} \qquad (49)$$

The uncertainty in $N_e(r_\oplus)$ is related to the uncertainties in B and ϵ by

$$\frac{\Delta N_e(r_\oplus)}{N_e(r_\oplus)} = \frac{\Delta B}{B} - (\ln 215)\Delta\epsilon \qquad (50)$$

and the variance on $N_e(r_\oplus)$ is related to the variances on B and ϵ and the correlation coefficient ρ between them by

$$\frac{\sigma^2_{N_e}}{N_e^2(r_\oplus)} = \frac{\sigma_B^2}{B^2} - 2\rho\,(\ln 215)\frac{\sigma_B}{B}\sigma_\epsilon + (\ln 215)^2\sigma_\epsilon^2 \qquad (51)$$

The numerical form of Eq. (51) for $N_e = 5$ electrons/cm^3 is

$$\left(\frac{\sigma_{N_e}}{5}\right)^2 = \left(\frac{3}{5}\right)^2 - 2\rho(5.37)\left(\frac{3}{5}\right)(0.3) + (5.37)^2(0.3)^2 \qquad (52)$$

and with $\rho = 1$, the error on N_e is constrained to ± 5 electrons/cm^3. For values of ρ smaller than unity, the uncertainty on N_e will be larger. Because a value of ± 5 electrons/cm^3 is conservative enough, we adopt a correlation coefficient of 1.0 between B and ϵ for the Mariner experiment. This implies that the a priori value and uncertainty at the earth's distance N_e is 5 ± 5 electrons/cm^3.

Table 2. Coefficients for solar corona model

A, cm^{-3}	B, cm^{-3}	ϵ	N_e at 1 AU	State	Ref.
1. 0.96×10^8	-	-	-	Solar minimum, equatorial	19
2. 1.71×10^8	-	-	-	Solar maximum, equatorial	19
3. 1.18×10^8	1.46×10^6	0.3	6.4	Solar minimum, equatorial (July 1963)	22
4. 1.01×10^8	2.01×10^6	0.33	7.4	Average solar minimum, equatorial	20
5. 2.62×10^8	2.07×10^6	0.33	7.6	Strong solar maximum, equatorial and polar	20
6. 0.60×10^8	0.8×10^6	0.33	2.9	Measured Oct. 2-14, 1969, 13 deg solar latitude	21

V. Measurement of the General Relativistic Propagation Delay

The measurement of the general relativistic propagation delay is accomplished by estimating the relativity parameter γ along with the corona parameters (A, B, ϵ) and the elements of the spacecraft orbit in a least-squares differential correction process. In this way the "best" determination of the relativistic delay is defined as that which makes the weighted sum of squares of the range and doppler data a minimum. In order to separate the parameter γ from other parameters in the least-squares solution, it is necessary in principle to determine the heliocentric orbit of the spacecraft from data outside of superior conjunction, and then to derive the relativistic delay from the characteristic increase in the round-trip travel time of ranging signal beyond that predicted by the position on the heliocentric orbit at superior conjunction. In practice this is accomplished by least-squares fits to many weeks or months of doppler and range data centered about superior conjunction.

The least-squares algorithm used in the analysis of the Mariner data is

$$\left(A^T W A + \tilde{\Gamma}_x^{-1}\right)\Delta x = A^T W\left(\Delta z + \tilde{\Gamma}_x^{-1}\Delta\tilde{x}\right) \tag{53}$$

where Δz represents the column vector of residuals in the range and doppler data, Δx represents the column vector of corrections to the current set of parameters, $\Delta\tilde{x}$ represents the corrections to the a priori values of the parameters, $\tilde{\Gamma}_x$ is the a priori covariance matrix on the parameters, W is a diagonal weighting matrix for the data, and the matrix A contains the differential coefficients that relate variations in the data vector to variations in the parameter vector ($dz = A dx$).

The method of solution is that of weighted least squares with a modification to allow the introduction of a priori information into the process. As in any least-squares solution it is necessary to compute residuals Δz in the data, and by convention the sense of the residuals is the observed minus the computed (O - C) values. The adopted procedure is simply to represent the range and doppler measurements as accurately as necessary by mathematical formulae and then to form the O - C subtraction. The actual measurement O is stored on magnetic tape. An accurate representation of the data will involve considerations of light-time, atmospheric refraction corrections, and an interpretation of the station procedure used to record the time of an observation.

The equations of motion for the Mariner spacecraft are expressed in mean equatorial coordinates of 1950.0. They represent a sixth-order system of differential equations where only the coordinates of the spacecraft are obtained by numerical integration. Coordinates of other bodies in the solar system are stored on magnetic tape and are provided by JPL Development Ephemeris 69 (DE69). The equations of motion are expressed in the relative motion form and, in vector notation, are given by

$$\frac{d^2\vec{r}_{12}}{dt^2} = -k^2(m_1 + m_2)\frac{\vec{r}_{12}}{r_{12}^3} + k^2\sum_{j=3}^{n} m_j\left(\frac{\vec{r}_{2j}}{r_{2j}^3} - \frac{\vec{r}_{1j}}{r_{1j}^3}\right) + \vec{P}_2 \tag{54}$$

By convention, a position vector $\vec{r}_{ij}$ represents the coordinates of the j^{th} body of mass m_j with respect to the i^{th} body of mass m_i. Thus, the first term in Eq. (54) represents the two-body acceleration of the Mariner spacecraft with respect to the primary body of mass m_1. The mass of the probe is negligible with respect to the primary mass m_1. Therefore, m_2 can be set equal to zero.

The second term in the equations represents the contribution to the relative acceleration from other bodies in the solar system. For the Mariner orbit the primary body is the sun, and the other bodies in the n-body system are the remaining planets and the moon. The third term, $\bar{P}_2$, represents perturbative accelerations on the spacecraft which arise from forces aside from the gravitational attraction of the sun, moon, and planets. In particular, $\bar{P}_2$ includes effects from solar radiation pressure on the spacecraft and low-thrust forces from the spacecraft attitude control system, which operates by releasing cold nitrogen gas through a number of jets. Because neither of these non-gravitational forces has a significant effect on the primary body, the form of $\bar{P}_2$ can be equated to the inertial acceleration from solar pressure and low-thrust forces.

If k^2 in Eq. (54) is set equal to the Gaussian gravitational constant, the units are astronomical units, solar masses, and ephemeris days in the equations. However, the units used in the integration of Eq. (54) are km and sec, and $k^2 m_i$ is combined into a single factor GM_i (km^3/sec^2). The formula for GS in the case of the sun is given by

$$GS = (86,400)^{-2} k^2 A^3 \qquad (55)$$

where A is the astronomical unit in kilometers and, by definition (Ref. 23), the Gaussian constant is

$$k = 0.01720209895 \text{ AU}^{3/2}/\text{day}$$

The value of GM_i for any planet whose mass M_i^s is given in solar mass units is

$$GM_i = (GS)\frac{M_i}{S} = (86,400)^{-2} k^2 A^3 M_i^s \qquad (56)$$

It is understood that the value A of the astronomical unit in kilometers is based on the adopted value of c because, as mentioned earlier, the standard meter is of no consequence in the mathematical representation of the tracking data. The adopted values of the constants for all the solutions of this paper are:

Astronomical unit	A = 149,597,893 km
Velocity of light	c = 299,792.5 km/s
Geocentric gravitational constant	GE = 398,601.2
Earth/moon mass ratio	μ = 81.3011

The adopted values for the masses of the planets are given in Table 3.

The matrix of differential coefficients given by A and the a priori covariance matrix Γ_x are determined by the parameter set x used in the differential correction. The variations of the position and velocity of the spacecraft with respect to variations in the initial conditions, as well as the other parameters necessary to represent the orbit, are obtained by numerically integrating a set of variational equations.

Table 3. Adopted planetary masses

Planet	Reciprocal mass
Mercury	5,983,000
Venus	408,522
Mars	3,098,700
Jupiter	1,047.391
Saturn	3499.2
Uranus	22930
Neptune	19260
Pluto	1,812,000

In constructing the weighting matrix W, the important consideration is the relative weighting of the range and doppler data. In all the solutions of this paper we have weighted the doppler data with a standard error of 0.05 Hz (3.4 mm/s) when sampled at a rate of one point per minute. For any other sample interval T_s, the standard error is modified by the rule $\sigma = 0.05\,60/T_s$, where T_s is given in seconds. Of course the weight w for the matrix W is the inverse of the square of the standard deviation ($w = 1/\sigma^2$). The range data are weighted with a standard error of 90 m in one-way range. For both range and doppler the weighting is conservative in that the standard errors are about a factor of 7 high in both cases. However, the relative weighting is representative of the RMS residuals obtained from the data included in the least-squares fits.

A. Post-Encounter Trajectories of Mariners 6 and 7

Trajectories have been computed for Mariners 6 and 7 which show the relative geometry of the earth and the spacecraft with respect to the sun. The post-encounter trajectory of Mariner 6 projected onto the plane of the ecliptic is shown in Fig. 6; that for Mariner 7 is shown in Fig. 7. The heliocentric trajectories of Mariners 6 and 7 are quite similar.

Soon after Mariners 6 and 7 flew by Mars on July 30, 1969, and August 5, 1969, respectively, the spacecraft went into heliocentric orbits, and the general relativity time delay experiment commenced. A summary of the orbital parameters of the spacecraft at the initial epoch and superior conjunction is given in Table 4. Note that during superior conjunction neither spacecraft underwent a solar occultation. However, the radio signal penetrated deep into the solar corona and, at

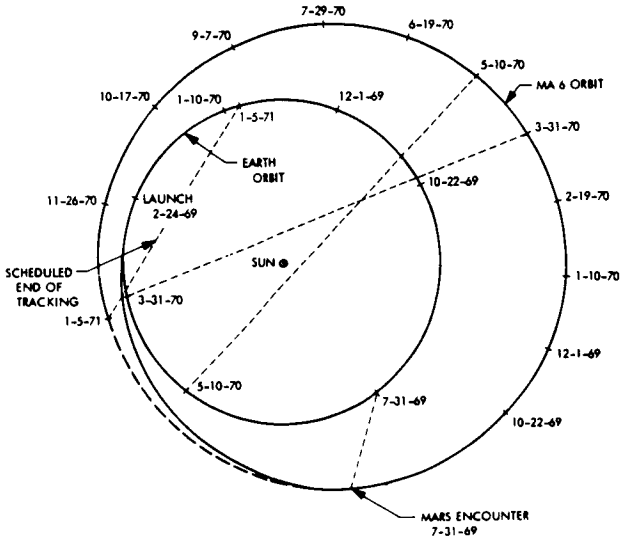


Fig. 6. Heliocentric orbit for Mariner 6

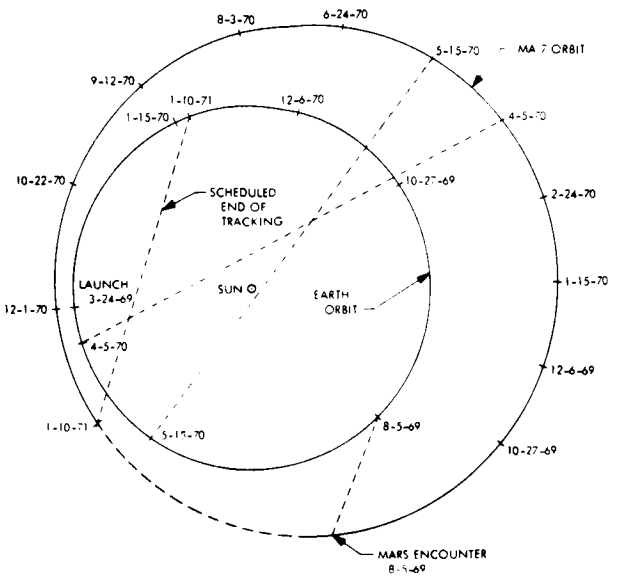


Fig. 7. Heliocentric orbit for Mariner 7

closest approach, passed within 5 solar radii of the solar surface for Mariner 7 and 3 solar radii for Mariner 6.

It is interesting to note that the heliocentric trajectories after encounter with Mars are significantly different from the earth-to-Mars transfer orbits. After successfully completing close encounters with Mars, the two Mariner spacecraft gained energy so that they were in solar orbits, with perihelion distances significantly greater than the aphelion distance for the earth. Though not designed as Mars swing-by missions, the Mariner trajectories were fairly efficient in this respect. If Mars had been a massless planet, and if an impulsive maneuver had been performed to change the pre-encounter orbits into the resulting post-encounter orbits, the velocity requirements would have been 1.584 km/s for Mariner 6 and 1.446 km/s for Mariner 7. However, no maneuvers were performed at the time of encounter.

There is another important implication of the Mariner trajectories. The spacecraft were attitude-stabilized so that the solar panels always pointed towards the sun. For the same reason, the low-gain antenna on the spacecraft was always directed toward the sun and always pointed toward the earth during the period of superior conjunction. However, the high-gain antenna axis made an angle of 41.6 deg to the low-gain antenna axis. Therefore, in order to send a ranging signal to the high-gain antenna, it would have been necessary to reorient the spacecraft so that the high-gain antenna pointed at both the sun and earth. This mode of operation, although somewhat risky, was seriously considered prior to conjunction. Fortunately, however, the availability of the high-power transmitter during solar conjunction made the orientation of the spacecraft unnecessary; all ranging was accomplished by means of the spacecraft's low-gain antenna.

A scale drawing of the superior conjunction of Mariner 6 is shown in Fig. 8. The spacecraft approached the sun from the west at a rate of about 1/3 deg per day, and on April 29, 1970, the ray path to the spacecraft passed over the north polar region of the sun. The geometry of the Mariner 7 conjunction was similar, but the closest approach was at about 5 solar radii instead of 3.

B. Data Summary

The experimental data include time-integrated doppler and active radar or round-trip range. When the experiment began, seven Deep Space Network (DSN) stations were receiving doppler data from both spacecraft. The first range data were received near the end of September 1969, when the range equipment became operational. At this time, the round-trip travel time of the radio signal was approximately 20 min. The frequency of the range data was about one measurement per week.

During the critical period of superior conjunction, high-quality range measurements were obtained daily. Within several days of superior conjunction, the received range code could not be compared in an unambiguous manner with that transmitted. This situation was caused by the interference introduced to the ranging code as the signal propagated through the solar corona. Thus, no range data exist within this relatively short interval of time. The doppler data were even more seriously affected. The solutions of this paper do not utilize doppler data taken within a 1-month interval centered on superior conjunction. An effort is being made to recover some of these doppler data.

At present, thousands of doppler and several hundred range measurements exist for each spacecraft. While the doppler data supply important orbital information, it is the range data that provide a direct measurement of the relativistic time delay. Thus, the two types of data complement each other in yielding the most reliable and accurate measurement of the relativistic time delay that is possible with the Mariner spacecraft. Doppler and range measurements were obtained at

Table 4. Mariner Orbital elements and superior conjunction information

Data	Mariner 6	Mariner 7
Orbit		
Epoch, ET	6 Aug 69 $0^h0^m0^s$	10 Aug 69 $0^h0^m0^s$
Semi-major axis, km	216.62×10^6	210.32×10^6
Eccentricity	0.211	0.206
Longitude of ascending node,* deg	-18.495	-35.311
Argument of Periapsis,* deg	-155.740	-142.620
Inclination,* deg	1.799	2.571
Period, days	636.460	608.865
Conjunction		
Superior conjunction, ET	29 Apr 70 17^h	10 May 70 1^h
Earth-spacecraft distance, km	402.25×10^6	385.556×10^6
Round-trip travel time, min	44.72	42.87
Sun-earth-spacecraft angle, deg	0.946	1.555
Closest approach of signal to solar surface ($R_\odot$)	2.58	4.90

*Measured with respect to the earth mean orbit of 1950.0.

regular intervals until the end of 1970, when the data collection came to an end.

C. Data Analysis

The basis of the time delay test is the generalized metric of Eq. (1) and the propagation equation (Eq. 31) that follows from it. In addition, the solar corona also introduces a delay in the ranging signal because of interactions of the S-band signal with free electrons.

Expressed schematically, the equation for the round-trip range observable is

$$\Delta t(\text{observed}) = \frac{\rho}{c}(\text{uplink}) + \frac{\rho}{c}(\text{downlink})$$

$$+ \frac{2m}{c}(1+\gamma)\ln\left(\frac{r_p + R + \rho}{r_p + R - \rho}\right)$$

$$+ \Delta t\,(\text{solar corona})$$

$$+ \Delta t\,(\text{earth's atmosphere}) \quad (56)$$

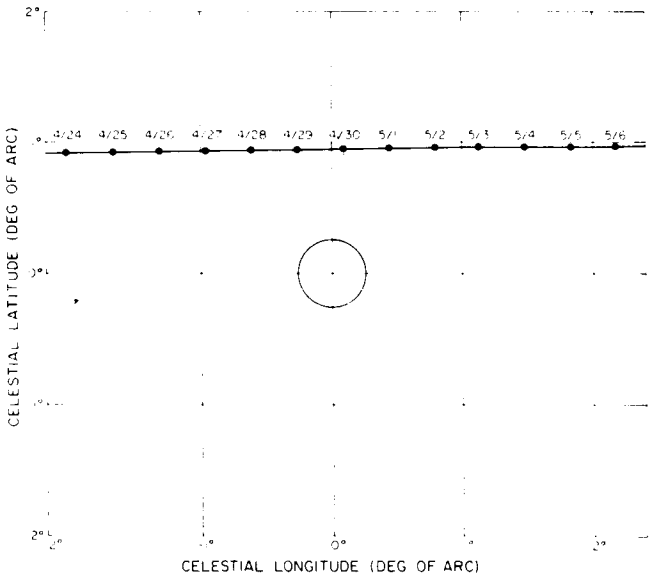


Fig. 8. Geometry of the Mariner 6 superior conjunction

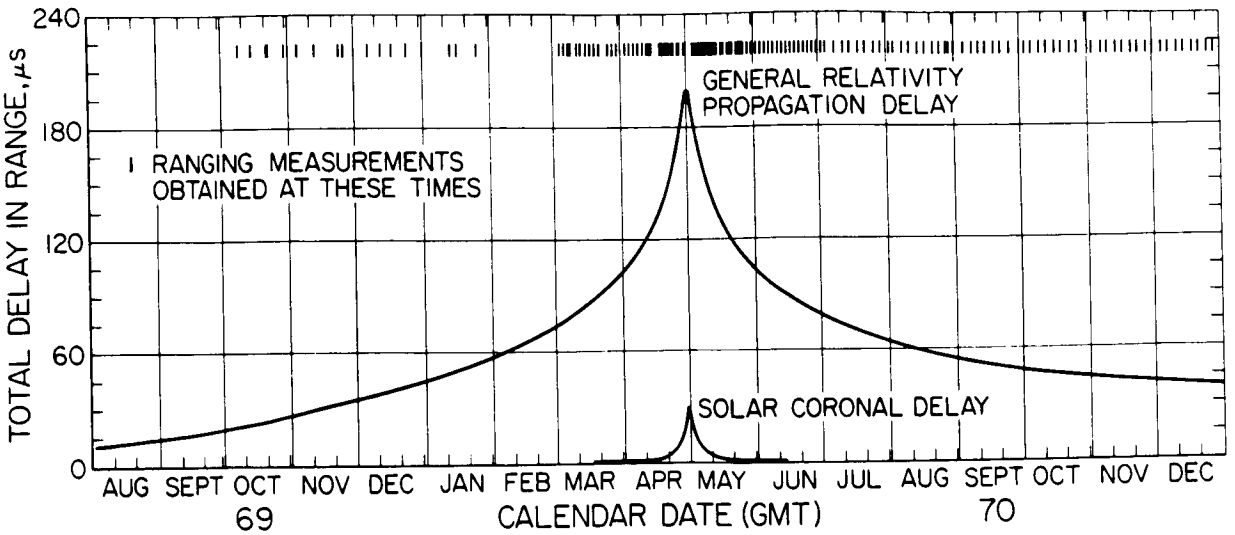


Fig. 9. Mariner 6 time delay due to general relativity and the solar corona

where r_p is the sun-spacecraft distance, R the sun-tracking station distance and ρ the station-spacecraft distance. As this equation indicates, the range observable consists of the finite travel time of the range signal from transmission at the 210-ft antenna at Goldstone, California, to reception at the spacecraft (uplink), and the retransmission from the spacecraft to the same antenna (downlink). In addition, Eq. (56) contains the expression for the relativistic time delay, which also consists of an uplink and downlink portion, as well as further delays introduced by the solar corona and the atmosphere of the earth.

Figures 9 and 10 show the relativistic time delay, with γ equal to unity, for Mariners 6 and 7 as a function of calendar date. For Mariner 6, the maximum relativistic time delay is approximately 200 μs, and the width at half amplitude is about 60 days. In addition, the maximum time delay introduced by the solar corona is 30 μs, with a width at half amplitude of about 6 days. At superior conjunction, the round-trip travel time was roughly 45 min. Across the top of Figs. 9 and 10, vertical bars show the times of the range measurements. While the signatures of the relativity and coronal time delay appear similar, their differences are sufficiently great to allow for a good separation of these two distinctly different effects.

Figure 10 shows a maximum relativistic delay of about 180 μs and a width at half amplitude of 95 days for Mariner 7. With respect to the solar corona, the maximum delay is approximately 15 μs, with a width at half amplitude of 12 days.

The absence of ranging points at the peak of the relativity curve can be explained primarily by multipath effects in the solar corona which prevented an acquisition of the ranging code, and to a lesser extent by operational problems with the high-power transmitter operating at 200 kW. Also, the relativity experimenters were required to share the 210-ft antenna with other groups, in particular with a group of JPL experimenters taking doppler measurements of Mercury during its inferior conjunction in April of 1970. Thus, horizon-to-horizon attempts to acquire the Mariners were not possible during the superior conjunctions.

Although it is the parameter γ that is estimated in the least-squares differential correction process, it is preferable to express results in terms of a parameter γ^* which represents the proportionality constant in the total relativistic delay. This proportionality constant is defined by

$$\gamma^* = \frac{1}{2}(1 + \gamma) \tag{57}$$

and it is equal to unity in the Einstein theory. In the Brans-Dicke theory it is related to the coupling constant ω between the scalar and tensor fields by

$$\gamma^* = \frac{3 + 2\omega}{4 + 2\omega} \tag{58}$$

Thus, for a value of 5 for ω, $\gamma^* = 0.93$, and the Brans-Dicke theory would predict a 7% decrease in the relativistic delay from that predicted by the Einstein theory. For Mariner 6, the predicted delay would be decreased from 200 to 186 μs in the Brans-Dicke theory. Of course, if ω should be taken larger than 5, then the predicted decrease from the Einstein delay would be less.

The parameter β in the metric of Eq. (1) is unity in both the Einstein and Brans-Dicke theories. There is no sensitivity of the Mariner data to a determination of β, nor is there any significant sensitivity to the sun's quadrupole moment.

All of the data for the relativity experiment have been processed with the JPL Double Precision Orbit Determination Program (DPODP). In most of the analysis, 18 parameters are included in the least-squares fits. The estimated parameters are:

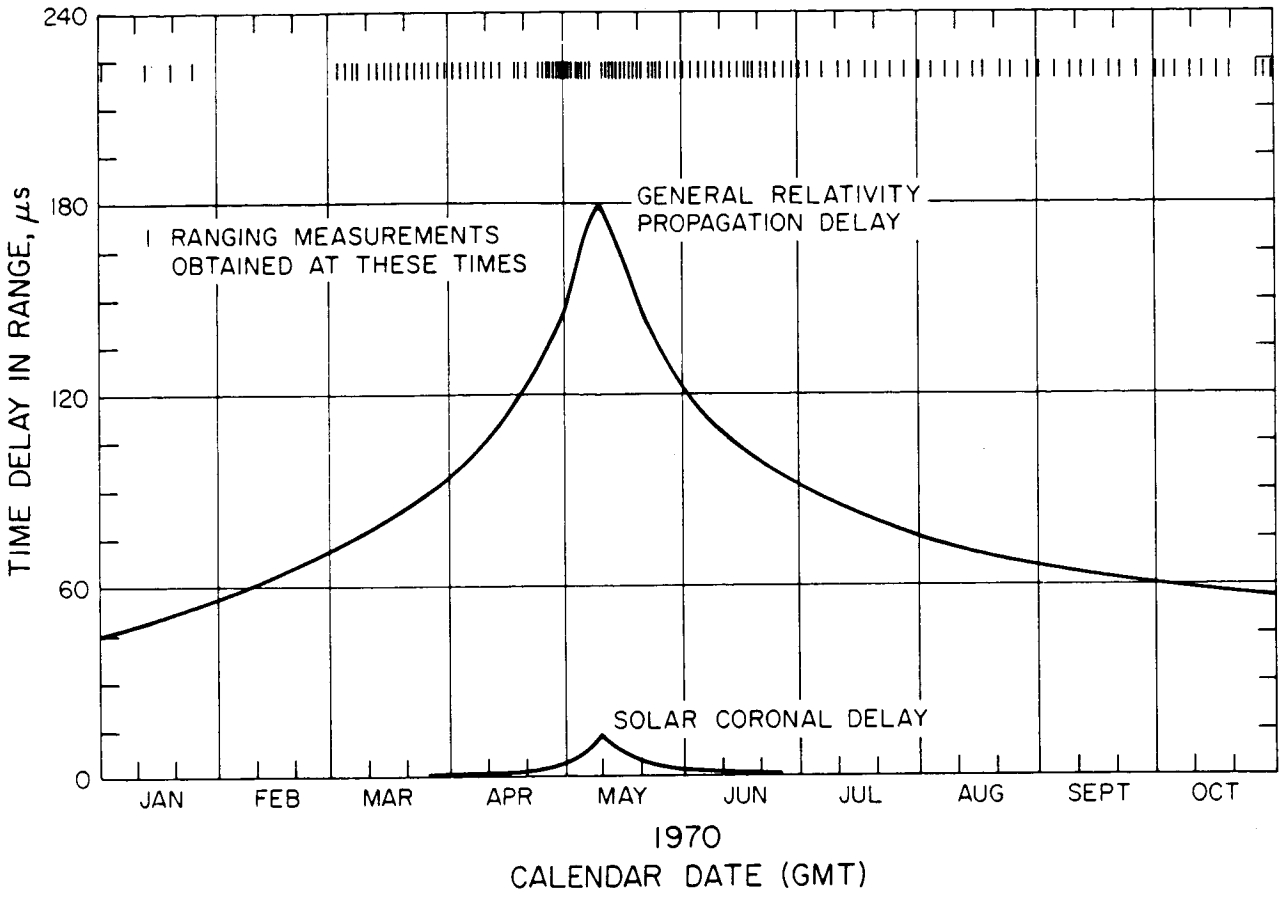


Fig. 10. Mariner 7 time delay due to general relativity and the solar corona

Six initial conditions for the spacecraft orbit

Solar radiation pressure model (3 components)

Attitude control force model (3 components)

Earth-moon mass ratio

Astronomical unit

Station locations

Solar coronal electron density model (see Section IV)

General relativity parameter (γ)

It would make little sense to attempt a solution for all of these parameters from the Mariner data alone, but this is precisely the advantage of incorporating an a priori option into the least-squares algorithm. A priori values of well known parameters, such as the astronomical unit, can be used along with their standard deviations to condition the solution for the parameters of interest. In this way, the uncertainties in parameters which are peripheral to the Mariner problem can be accounted for in the estimate of the uncertainties on parameters such as γ. The most straightforward method of determining the best value of γ from the Mariner data would be to process all the available data in one least-squares solution for the parameters listed above. Unfortunately, as is often the case, the most straightforward approach leads to erroneous results. Fourteen months of Mariner 6 data from August 5, 1969 to October 10, 1970 have been fit in the least-squares sense with the result that the residuals in range exhibit large systematic trends (see Fig. 11). There are two possible explanations for this.

One explanation for the large residuals shown in Fig. 11 is that some important dynamical effect on the spacecraft orbit has been neglected. We have investigated a number of possible causes for this, among them an error in the mass of Mercury or a contribution to the spacecraft motion from the total of mass in the asteroid belt, and have found nothing that could reasonably produce residuals of the size shown in Fig. 11 or that would affect the uncertainty in the relativity parameter γ* at the level of 1%.

Although we remain open to the possibility that some unexpected dynamical effect might be present in the residuals of Fig. 11, there is another explanation for the residuals that is much more acceptable at this time. Random nongravitational forces are acting on the spacecraft which could arise from variations in the solar pressure acting on the spacecraft (Plamondon has reported variations in the solar flux at the level of 0.3% with a correlation width of a few hours; Ref. 24), from variations in the solar wind, or from random outgassing from the spacecraft's attitude control system. All of these effects are roughly of the

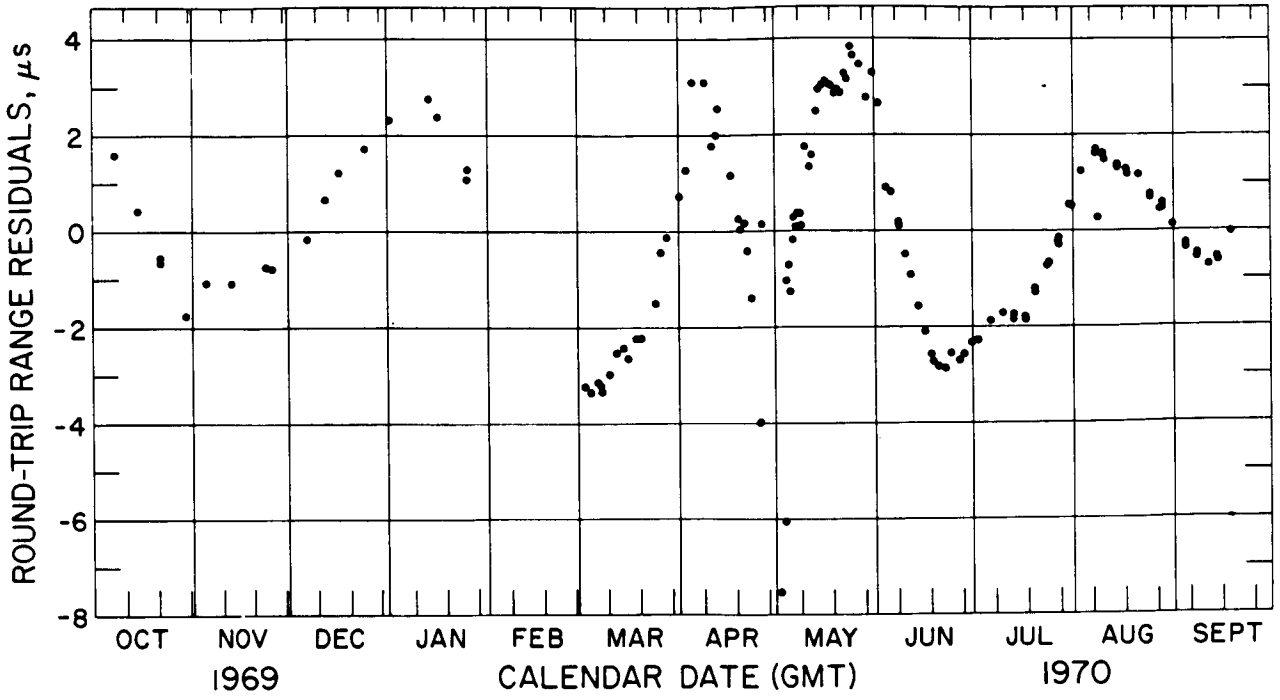


Fig. 11. Range residuals from a least-squares fit of 14 months (long arc) of Mariner 6 data

same order of magnitude and cause random accelerations on the spacecraft at the level of 10^{-10} to 10^{-9} m/s^2. By following an idea of Brouwer's (Ref. 25) for studying errors in numerical integration, it can be shown that random accelerations will produce a cumulative error in position that increases with time to the 3/2 power. Thus, for a standard error of σ_a on the spacecraft's random acceleration, the error σ_p in its position will increase roughly according to the formula

$$\sigma_p \sim t^{3/2} \Delta t^{1/2} \sigma_a \qquad (59)$$

where Δt is the correlation width of the random forces. For $\sigma_a = 5 \times 10^{-10}$ m/s^2 and $\Delta t = 4$ h, the position error would build up to the order of 13 km in 14 months. It is not surprising, therefore, to see residuals on the order of 4 μs (600 m) when fits over 14 months of data are attempted.

The implications of accepting the fact that random accelerations on the order of 10^{-10} to 10^{-9} m/s^2 are acting on the spacecraft are that (1) the random accelerations are the dominant error source for the relativity experiment and (2) the least-squares fits to the data must be accomplished within this constraint.

Perhaps the easiest method of removing the effect of the random forces is to process arcs of data which have a sufficiently short time span such that the random forces exert a negligible influence on the least-squares fits. The difficulty with this procedure is that it is self-defeating. If the total time interval of the data is too short, then it is impossible to separate the determination of the relativity parameter γ^* from the determination of the heliocentric orbit of the spacecraft. A good compromise is to process data over a time period of about 3 months centered about superior conjunction. In this amount of time the cumulative error in the position of the spacecraft is held to an acceptable level of the order of 1.3 km, and yet the relativity parameter can be determined to about 4%. Note that the cumulative error of 1.3 km is not mapped directly into an error on the relativity parameter because of the smoothing nature of the least-squares process.

Another means of accounting for the random forces is to resort to methods of sequential or recursive estimation of the parameters. These methods were originally explored by Kalman (Ref. 26) and by Schmidt and his colleagues at the Ames Research Center (Ref. 27). Our preliminary experimentation with these methods (see Section V-C1) indicates that the relativity parameter can be determined to about 4%, even when all the Mariner data are processed. With this knowledge in hand, it is possible to prepare a table of the important error sources which are outside the parameters of the least-squares solutions and, in addition, to list estimates of their contribution to the error in γ^*. Table 5 is the result of such a compilation. It can be seen at once that the random forces are definitely the major source of error in the Mariner relativity experiment.

1. Sequential Estimation

The basis of sequential estimation is to write the least-squares algorithm in a recursive form. For the differential correction of the parameters from the ith batch of data, where a batch can be anything from one measurement to all of the data, the recursive form of the algorithm is

Table 5. Error sources

Source	Assumed uncertainty	Error contribution to γ^*, %
Random variations in nongravitational forces	$\pm 10^{-9}$ m/s^2	3
Random columnar variations in solar coronal electron density over 30 days	±10%	<1.5 (Mariner 6) <0.8 (Mariner 7)
Earth's ephemeris	±0.''02	1
Earth effects	(see Ref. 28)	<1
Mass of Mercury	±1%	<1
Relativity parameter β	±20%	<0.02
Solar oblateness J_2	$\pm 10^{-5}$	<0.2

$$\left(A_i^T W_i A_i + \tilde{\Gamma}_i^{-1}\right)\Delta x_i = A_i^T W_i \Delta z_i + \tilde{\Gamma}_i^{-1}\Delta\tilde{x}_i \tag{60}$$

In the absence of random accelerations,

$$\Delta\tilde{x}_i = \Delta x_{i-1} \tag{61}$$

and

$$\tilde{\Gamma}_i^{-1} = A_{i-1}^T W_{i-1} A_{i-1} + \tilde{\Gamma}_{i-1}^{-1} \tag{62}$$

An application of Eqs. (60), (61), and (62) for any number of batches will result in the same final correction to the parameters that would result from a processing of all the data in a single batch. Of course, this is true only if the corrections are within a linear region of convergence and if enough numerical significance is carried in the computations.

When random accelerations are present, as they are in the Mariner experiment, then modifications to Eqs. (61) and (62) are in order. We have not altered Eq. (61) for the six position (x, y, z) and velocity ($\dot{x}$, $\dot{y}$, $\dot{z}$) components of the spacecraft, or for the relativity parameter γ or the corona parameter B. However, for the other parameters of the problem, they have been restored to their original values before the application of Eq. (60). The rationale for this is that such parameters as the earth-moon mass ratio, and the station locations are actual constants, and they are known a priori with more confidence than can be established with a single batch of Mariner data. The restoration of the parameters representing the deterministic part of the solar radiation pressure and the attitude control forces to their original values is done to permit a degree of independence between solutions for these parameters with different batches of data.

The a priori covariance matrix $\tilde{\Gamma}_i$ is modified significantly from what would result by a straightforward application of Eq. (62). First of all, for each batch of data the covariances on the position and velocity of the spacecraft are degraded appreciably from their values after the processing of the previous batch. This is to allow for the unknown action of the random accelerations on the orbit. Further, the uncertainties on the six parameters of the solar radiation pressure and attitude control forces are restored in their original values at each batch. In this way a set of nongravitational parameters is associated with each batch of data, and any variations in the nongravitational forces are effectively smoothed out somewhat by the least-squares process; the relaxation of the uncertainty on position and velocity at each batch need take account of only the residual noise in the nongravitational forces.

The sequential estimation has been applied to Mariner 6, with the 14 months of data divided into 6 batches. Table 6 shows the solutions for each batch, along with the standard error and the a priori error (the figure in parentheses) which was assumed before the introduction of the data for the batch under consideration. The six batches were processed chronologically in the order given. In all six solutions the nominal values of the parameters upon which the differential correction is based are the same. The six position and velocity components of the spacecraft are referred to the same epoch on August 5, 1969. The parameter GM in the table is the gravitational constant for the moon. The two corrections to the station location refer to the 210-ft antenna and are in a direction perpendicular to the earth's axis of rotation for Δr_s and along a direction of increasing longitude for $r_s\Delta\lambda$. The third direction parallel to the earth's axis cannot be determined from the data. The nominal values for γ^*, B, and GM are 1.0, 1.15×10^6, and 4902.80, respectively. The nominal values for the parameters of the solar radiation pressure (GR, GX, GY) and the attitude control forces (AR, AX, AY) are all zero, except for GR, which is 1.31.

The residuals for the six individual batches of data are shown in Fig. 12 on one continuous plot. Note that the large systematic trends evident in Fig. 11 are now gone. Figure 12 is a much

Table 6. Least-squares solutions for sequential estimation for Mariner 6 (batches 1-6)

Parameter	8/5/69-11/1/69	11/1/69-2/1/70	2/1/70-4/1/70	4/1/70-6/1/70	6/1/70-8/1/70	8/1/70-10/1/70
Δx, km	0 ± 86 (10^6)	164 ± 391 (500)	639 ± 1583 (2000)	4052 ± 2370 (5000)	4818 ± 2553 (7000)	3486 ± 3260 (8000)
Δy, km	252 ± 53 (10^6)	218 ± 241 (500)	-1069 ± 856 (2000)	768 ± 1008 (5000)	-768 ± 1005 (7000)	204 ± 1836 (8000)
Δz, km	-503 ± 70 (10^6)	-541 ± 411 (500)	1587 ± 1662 (2000)	-4673 ± 2333 (5000)	-625 ± 2241 (7000)	198 ± 2277 (8000)
$\Delta \dot{x}$, mm/s	-67 ± 20 (10^6)	-60 ± 32 (300)	-21 ± 105 (900)	-268 ± 148 (700)	-314 ± 154 (700)	-219 ± 212 (700)
$\Delta \dot{y}$, mm/s	38 ± 28 (10^6)	18 ± 63 (300)	-16 ± 140 (900)	453 ± 246 (700)	402 ± 277 (700)	173 ± 357 (700)
$\Delta \dot{z}$, mm/s	-31 ± 32 (10^6)	29 ± 99 (300)	288 ± 225 (900)	261 ± 212 (700)	450 ± 132 (700)	375 ± 180 (700)
GR	1.31 ± 0.09 (0.1)	1.32 ± 0.09 (0.1)	1.31 ± 0.09 (0.1)	1.32 ± 0.09 (0.1)	1.32 ± 0.09 (0.1)	1.34 ± 0.09 (0.1)
GX	0.05 ± 0.09 (0.1)	-0.03 ± 0.09 (0.1)	-0.05 ± 0.09 (0.1)	-0.02 ± 0.09 (0.1)	-0.02 ± 0.09 (0.1)	-0.03 ± 0.09 (0.1)
GY	-0.06 ± 0.09 (0.1)	-0.05 ± 0.09 (0.1)	-0.05 ± 0.09 (0.1)	-0.04 ± 0.09 (0.1)	-0.04 ± 0.08 (0.1)	-0.04 ± 0.08 (0.1)
AR, $m/s^2 \times 10^{-9}$	6.5 ± 5.1 (10)	1.1 ± 6.0 (10)	-10.4 ± 8.9 (10)	-1.0 ± 7.9 (10)	3.6 ± 7.2 (10)	6.5 ± 9.4 (10)
AX, $m/s^2 \times 10^{-9}$	8.7 ± 4.6 (10)	6.3 ± 5.1 (10)	1.6 ± 6.1 (10)	10.9 ± 6.3 (10)	5.2 ± 5.9 (10)	2.4 ± 7.2 (10)
AY, $m/s^2 \times 10^{-9}$	-2.1 ± 6.1 (10)	-1.0 ± 8.5 (10)	8.0 ± 9.3 (10)	-3.0 ± 9.5 (10)	2.6 ± 9.6 (10)	14.8 ± 9.9 (10)
GM, km^3/s^2, - 4902	0.74 ± 0.03 (0.1)	0.79 ± 0.03 (0.1)	0.87 ± 0.03 (0.1)	0.73 ± 0.02 (0.1)	0.81 ± 0.02 (0.1)	0.76 ± 0.03 (0.1)
$r_s \Delta\lambda$, m	-6.1 ± 1.5 (10)	2.2 ± 6.8 (10)	-5.0 ± 7.1 (10)	-8.8 ± 8.0 (10)	-0.5 ± 7.4 (10)	-9.2 ± 7.7 (10)
Δr_s, m	2.8 ± 2.2 (10)	-4.8 ± 3.1 (10)	-5.5 ± 2.3 (10)	2.5 ± 5.6 (10)	1.3 ± 4.9 (10)	-1.0 ± 4.4 (10)
B, electrons/$cm^3 \times 10^6$	--	--	--	1.16 ± 0.20 (0.70)	1.13 ± 0.19 (0.20)	1.13 ± 0.19 (0.19)
Y*	--	--	0.999 ± 0.1 (0.1)	1.032 ± 0.047 (0.5)	1.042 ± 0.042 (0.047)	1.043 ± 0.042 (0.042)

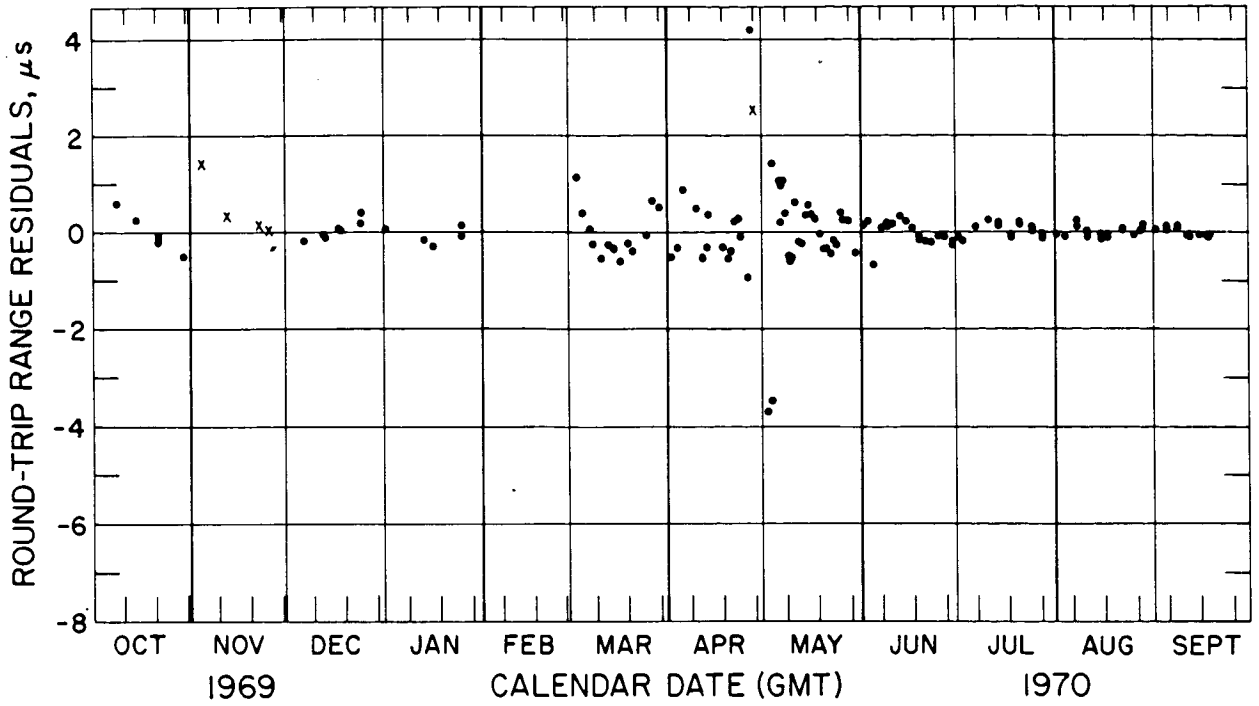


Fig. 12. Range residuals from a sequential fit of 14 months of Mariner 6 data

more representative indicator of the noise on the range measurements than is Fig. 11. At superior conjunction the residuals are more noisy because of the presence of the solar corona. The residuals after conjunction are better behaved than those before conjunction because of the use of 200 kW of transmitter power for post-conjunction measurements. A power of 20 kW was used for the measurements before conjunction.

The value of γ^* (1.043 ± 0.042) obtained from the sequential estimation should be reasonably reliable, although the fairly arbitrary selection of batch sizes and a priori variances in each batch suggests that more solutions should be performed as a check. The next two sections consider fits to just 3 months of data for Mariners 6 and 7, respectively, with the data centered about superior conjunction.

2. Determination of γ^* for 3 Months of Data from Mariner 6

Instead of analyzing all the data, we now confine our attention to the interval of data surrounding superior conjunction. As shown in Figs. 9 and 10, the relativistic time delay changes rapidly in this region, and extensive range coverage of the curve is available. In order to minimize the effect of the random accelerations and to consider enough data such that the relativity and solar coronal parameters uncouple, a 3-month data arc (short arc) centered on conjunction is analyzed. Three procedures are adopted. First, γ^* is estimated along with the parameters previously mentioned, with a nominal solar corona model; second, γ^* and B are estimated assuming the nominal A component of the solar corona; and finally, γ^*, A, and B are all estimated simultaneously. The results of this analysis are tabulated as follows:

	γ^*	$A \times 10^8$	$B \times 10^6$
Short arc (γ^*)	1.017 ± 0.010	Nominal	Nominal
Short arc (γ^*, B)	0.992 ± 0.032	Nominal	1.28 ± 0.16
Short arc (γ^*, A, B)	1.017 ± 0.038	2.23 ± 0.73	1.06 ± 0.24

The range residuals for the short arc solution in which γ^* and B are estimated together are given in Fig. 13. The nominal values of A, B and ϵ are given in Section IV.

In addition to the above three solutions, a solution in which all three corona parameters are estimated is given. For this solutions an a priori correlation of 0.999 is assumed between B and ϵ. The epoch for the position and velocity of the spacecraft is the same (August 5, 1969) as for the sequential estimation in Section 5-C1. Again, the a priori standard error is given in parentheses.

Parameter	Solution
Δx, km	2923 ± 2300 (10^6)
Δy, km	-1383 ± 936 (10^6)
Δz, km	1594 ± 1920 (10^6)
$\Delta \dot{x}$, mm/s	-155 ± 136 (10^6)
$\Delta \dot{y}$, mm/s	246 ± 263 (10^6)
$\Delta \dot{z}$, mm/s	259 ± 138 (10^6)

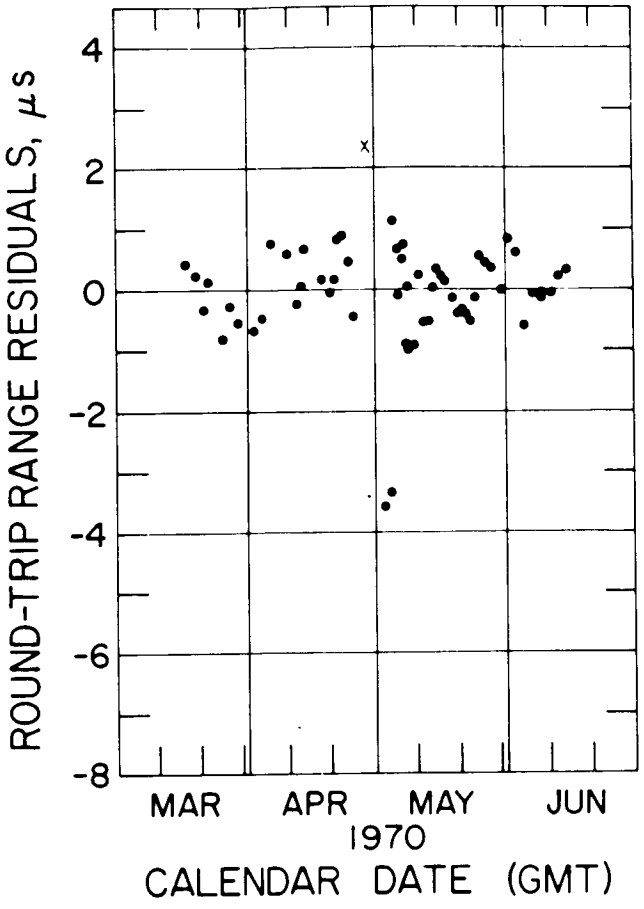


Fig. 13. Range residuals from a least-squares fit of the Mariner 6 conjunction data (short arc) (Residuals indicated by x correspond to data not used in the analysis.)

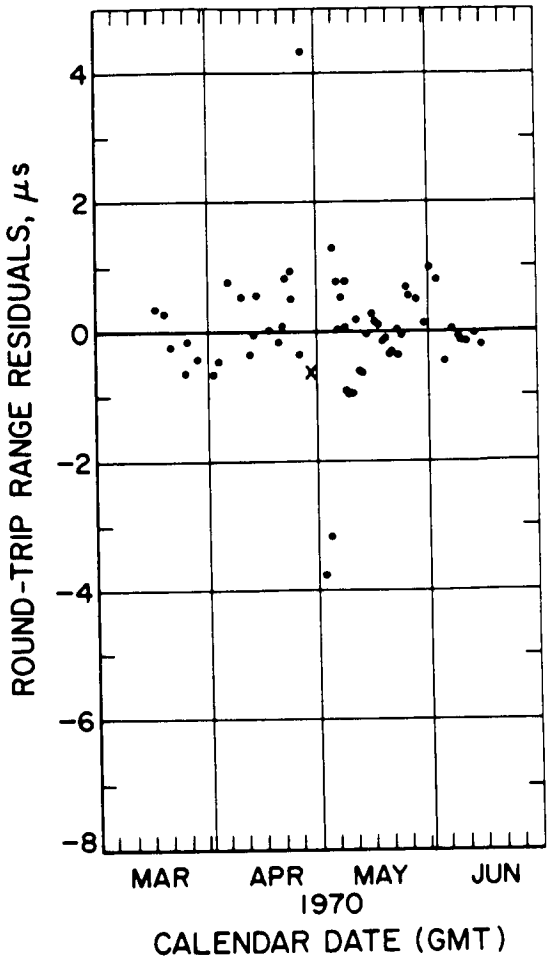


Fig. 14. Range residuals from a least-squares fit of the Mariner 6 conjunction data

Parameter	Solution
GR	1.32 ± 0.01 (0.01)
GX	-0.03 ± 0.01 (0.01)
GY	-0.05 ± 0.01 (0.01)
AR, $m/s^2 \times 10^{-9}$	-9.05 ± 3.27 (10)
AX, $m/s^2 \times 10^{-9}$	-2.87 ± 5.10 (10)
AY, $m/s^2 \times 10^{-9}$	7.58 ± 8.02 (10)
$r_s \Delta\lambda$, m	11 ± 9 (10)
Δr_s, m	15 ± 3 (10)
A, electrons/$cm^3 \times 10^8$	1.92 ± 0.70 (0.9)
B, electrons/$cm^3 \times 10^6$	1.41 ± 0.48 (0.7)
ϵ	0.41 ± 0.21 (0.3)
γ^*	0.997 ± 0.014 (0.5)

The range residuals for the short arc solution, where A, B and ϵ are all differentially corrected, are shown in Fig. 14. Doppler residuals for this same solution are shown in Fig. 15. Note that a 1-1/2-month arc of doppler data centered about superior conjunction is excluded from the fit because of a severe noise contribution to these data from the solar corona. The significance of the solution for the corona parameters is discussed separately in Ref. 29.

3. Determination of γ^* for 3 Months of Data from Mariner 7

While the orbit of Mariner 7 is basically the same as that of Mariner 6, an important difference occurs at superior conjunction. At closest approach, the ray path passes approximately twice as far from the solar surface for Mariner 7 as for Mariner 6. Consequently, the maximum relativistic time delay is reduced by 10% (180 versus 200 μs), while the maximum solar coronal time delay is reduced by 50% (15 versus 30 μs). The conclusion is that the solar corona has a significantly smaller effect on the Mariner 7 data analysis than on the Mariner 6 analysis. However, our experience with the Mariner 7 data leads us to believe that Mariner 7 is much more seriously affected by random accelerations than is Mariner 6. Therefore, we do not place as much confidence in the value of γ^* obtained from the Mariner 7 data.

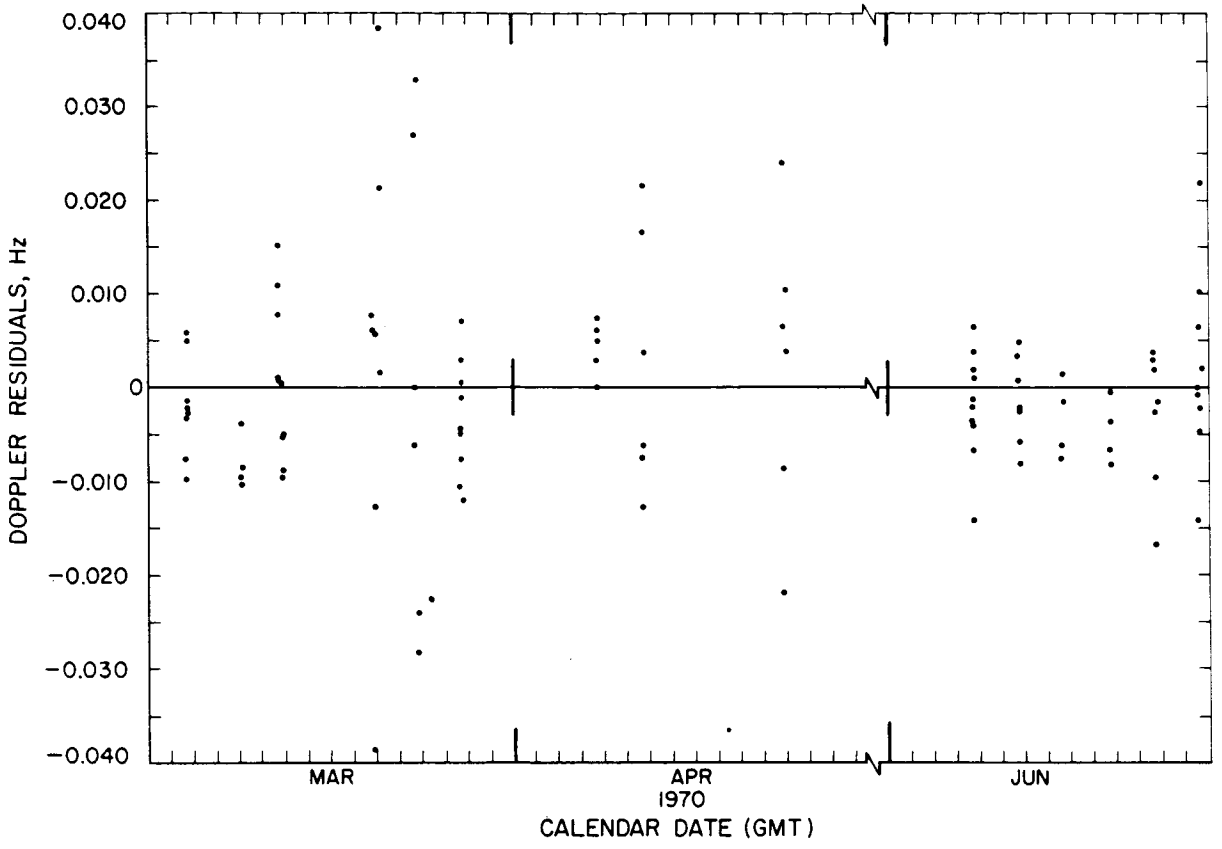


Fig. 15. Integrated doppler residuals from a least-squares fit of the Mariner 6 conjunction data (1 Hz ≈ 67 mm/s) (Doppler data within the interval April 15 - June 1, 1970, were not used in the analysis.)

Because of the larger random accelerations acting on Mariner 7, it is proving more difficult to make fits to the data over long orbital arcs. The only possibility for obtaining something worthwhile from Mariner 7, at least at this stage of the analysis, is to process as short an arc of data as possible about superior conjunction. Figure 16 shows a fit to data from April 1, 1970, to July 1, 1970. Again, as in the case of Mariner 6, the doppler data over a period of 1-1/2 months centered about superior conjunction are not included in the fit. The results of the differential correction are as follows:

Parameter	Solution
Δx, km	$-3393 \pm 1994\ (10^6)$
Δy, km	$362 \pm 751\ (10^6)$
Δz, km	$2004 \pm 1754\ (10^6)$
$\Delta \dot{x}$, mm/s	$202 \pm 134\ (10^6)$
$\Delta \dot{y}$, mm/s	$-584 \pm 226\ (10^6)$
$\Delta \dot{z}$, mm/s	$189 \pm 95\ (10^6)$
GR	$1.35 \pm 0.01\ (0.01)$
GX	$0.04 \pm 0.01\ (0.01)$
GY	$0.01 \pm 0.01\ (0.01)$
AR, $m/s^2 \times 10^{-9}$	$-3.6 \pm 4.4\ (10)$
AX, $m/s^2 \times 10^{-9}$	$0.5 \pm 4.6\ (10)$
AY, $m/s^2 \times 10^{-9}$	$1.6 \pm 7.6\ (10)$
$r_s \Delta\gamma$, m	$-4 \pm 7.5\ (10)$
Δr_s, m	$16 \pm 2.7\ (10)$

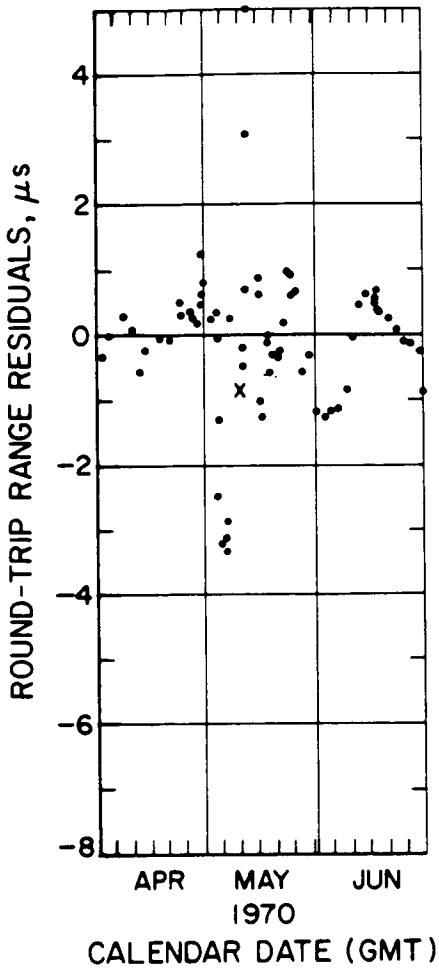


Fig. 16. Range residuals from a least-squares fit of the Mariner 7 conjunction data

Parameter	Solution
B, electrons/cm^3 × 10^6	1.73 ± 0.44 (0.7)
ϵ	0.55 ± 0.19 (0.3)
γ^*	1.025 ± 0.011 (0.5)

The corona parameter A is not included in the Mariner 7 solution because the data are not sensitive to it.

4. Summary of Results

The values of the relativity parameter γ^* from the various Mariner 6 solutions range from a low of 0.992 for one of the solutions over 3 months of data to a high of 1.043 for the sequential estimation. Our general attitude toward these solutions is that (1) the sequential estimation is a very preliminary effort, which needs considerable refinement before a fit to 14 months of data can be considered more reliable than the fits to 3 months of data, (2) the 3-month solutions without the corona parameter ϵ should be representative of the time delay measurement from Mariner 6, and (3) the 3-month solution with a full estimation of the three corona parameters, including the perfect positive correlation between B and ϵ, should provide the most reliable estimate of γ^* of any of the solutions.

An unexpected result of the solutions is that the formal error on γ^* is significantly less when the perfect correlation between B and ϵ is taken into account than it is for the case where γ^*, A, and B are estimated, with ϵ left out of the solution. The formal error in the former case is 1.4%, while in the latter it is 3.8%. The inclusion of ϵ in the solution presumably provides a better separation of the coronal time delay and the relativistic time delay than does a solution for B alone, where the shape of the coronal delay curve cannot be adjusted.

The solution for γ^* from the Mariner 7 data is not as reliable as that from Mariner 6 because, as mentioned earlier, Mariner 7 is subjected to larger and more erratic random forces. However, the value for γ^* (1.025 ± 0.011) tends to corroborate the Mariner 6 results.

In all cases, with the possible exception of the sequential estimation, the formal standard errors on γ^* can not be taken too seriously because the effects of the random forces are not included directly in the error estimates. Short-term fluctuations in the solar corona are not included either, but as indicated in Table 5 their influence is relatively small. The dominant error source is definitely the random forces, and based on the solutions of this paper it is reasonable to bound the error on γ^* because of these forces and the other sources of error to 4%. Also, a detailed theoretical discussion of the contribution of the random forces (Ref. 30) seems to indicate that the expected error in γ^* for 3 months of data should be in the neighborhood of 4%, although the results of the theoretical investigation are very dependent on the assumptions about the nature of the random forces and on how the corona is included in the simulated solutions.

Based on the foregoing discussion, we are inclined to accept the value of $\gamma^* = 0.997$ as the most representative yet obtained from the Mariner data and to attach a "realistic" standard error to it of ±0.04.

References

1. Dicke, R. H., and Goldenberg, H. M., Phys. Rev. Lett., 18, 313, 1967.

2. Dicke, R. H., Ap. J., 159, 1, and Ann. Rev. Astr. and Ap., 8, 30, 1970.

3. Ingersoll, A. P., and Spiegel, E. A., Ap. J., 163, 375, 1971.

4. Shapiro, I. I., this proceedings.

5. Muhleman, D. O., and Reichley, P., Space Programs Summary 37-29, Vol. IV, 239-241, Jet Propulsion Laboratory, Pasadena, Calif., Oct. 31, 1964.

6. Shapiro, I. I., Phys. Rev. Lett., 13, 789, 1964.

7. Shapiro, I. I., Bulletin Astronomique, 25, 1965.

8. Ross, D. K., and Schiff, L. I., Phys. Rev., 141, 1215, 1966.

9. Shapiro, I. I., Phys. Rev., 141, 1219, 1966.

10. Holdridge, D., Space Programs Summary 37-48, Vol. III, 2-4, Jet Propulsion Laboratory, Pasadena, Calif., Dec. 31, 1967.

11. Tausner, M. J., MIT Technical Report 425, 1966.

12. Eddington, A. S., The Mathematical Theory of Relativity, Harvard University, Cambridge, Mass., 1923.

13. Robertson, H. P., in Space Age Astronomy, p. 228. Edited by A. J. Deutsch and W. E. Klemperer. Academic Press, New York, 1962.

14. Clemence, G. M., and Szebehely, V., Astron. J., 72, 1324, 1967.

15. Moyer, T. D., Relativistic Equations of Motion, MS Thesis, UCLA, Los Angeles, Calif., 1965.

16. Moyer, T. D., Space Programs Summary 37-49, Vol. III, 40-54, Feb. 29, 1968, and 37-50, Vol. III, 1-7, Apr. 30, 1968, Jet Propulsion Laboratory, Pasadena, Calif.

17. Lieske, J., Ap. J. (in press).

18. Aoki, S., Astron. J., 69, 221, 1964.

19. Van de Hulst, Bull. Astron. Inst., Netherlands, 11, 135, 1950.

20. Blackwell, D. E., Dewhirst, D. W., and Ingham, M. F., Advances in Astronomy and Astrophysics, 5, 25, Academic Press, New York, 1967.

21. Muhleman, D. O., Ekers, R. D., and Fomalont, E. B., Phys. Rev. Lett., 24, 1377, 1970.

22. Blackwell, D. E., and Petford, A. D., Monthly Notices Roy. Astron. Soc., 131, 383, 1966.

23. Clemence, G. M., Ann. Rev. Astron. and Ap., 3, 93, 1965.

24. Plamondon, J. A., Space Programs Summary 37-59, Vol. III, 162-168, Jet Propulsion Laboratory, Pasadena, Calif., Oct. 31, 1969.

25. Brouwer, D., Astron. J., 46, 149, 1937.

26. Kalman, R. D., J. Basic Eng., 35-45, Mar. 1960.

27. McLean, J. D., Schmidt, A. N., and McGee, L. A., NASA Technical Manual D-1208, 1962, and Smith, G. L., Schmidt, A. N., and McGee, L. A., NASA Technical Report R-135, 1962.

28. Trask, D., talk presented at Conference on Experimental Tests of Gravitation Theories, Pasadena, Calif., Nov. 11-13, 1970.

29. Muhleman, et al., this proceedings.

30. Curkendall, et al., this proceedings.

Testing General Relativity: Progress, Problems, and Prospects

Irwin I. Shapiro
Massachusetts Institute of Technology

I. Introduction

Having remained in a state of almost complete lethargy for nearly half a century, experimental research in gravitation is now perhaps the fastest growing of the subdisciplines of physics. The reasons are simple: Until about a decade ago, physicists in the field constituted essentially a null set; empirical support for the fundamental theory of gravity rested almost completely on the work of a dedicated, unbroken retinue of astrometers. With such initial conditions, not many physicists need be converted to yield a rapid growth condition. The conversion was prompted by the truly fantastic increase in instrumental capability that has accompanied the space age. The Golden (Parkinsonian) Rule states that whenever more accurate measurements can be made they will be made – provided only that a major fraction of the world's GNP is not required. The results are sometimes even important, although perhaps in unsuspected ways as, for example, in the discovery of pulsars.

The remainder of this paper is divided into three main sections: in the first, we consider the results from present ground-based experiments; in the second, prospects for improvement in these experiments; and in the final section, we compare ground-based and spacecraft techniques on an accuracy vs cost basis. The ratio of content to rhetoric increases – at least on average – in these latter sections.

II. Recent Results from Radar and Radio Experiments

Here we shall give a progress report on the various radar and radio tests of general relativity that have been undertaken at MIT. But first, we describe briefly our philosophical approach to these tests.

A. Parameterization Philosophy

One could adopt and defend a variety of approaches for the characterization of tests of a class of physical theories. There is no unique procedure. The choice basically boils down to a matter of taste, especially when, as here, even the class to be considered is by no means universally agreed upon. For example, one could try to formulate a mathematical model for gravitation, as fully parameterized as possible, constrained by only the most basic theoretical concepts that we consider inviolable. This approach has been followed to greater or lesser extents by Eddington (Ref. 1), Robertson (Ref. 2), Dicke (Ref. 3), Nordtvedt (Ref. 4), Thorne (Ref. 5), Will (Ref. 6), and others. Near the opposite end of the spectrum, one can simply assume that general relativity, as originally formulated, is the standard for experimental comparisons. From a set of measurements, one can then estimate the inevitably unknown parameters* and examine the post-fit residuals. If statistical analyses of these show no deviations of significance, one concludes that

*Not even Einstein succeeded in formulating a theory which predicted planetary initial conditions and masses.

general relativity is an adequate theory – at least for that set of measurements. Most physicists will probably agree that this latter approach is unsatisfactory: One is given essentially no quantitative inkling as to how stringently the various aspects of the theory are being tested by the data, except insofar as the fractional measurement accuracies themselves supply clues. Nonetheless, some – no doubt a small minority – could argue that so long as the theory is in agreement with the measurements in a statistically satisfactory manner, nothing more need be asked of it.

The next level up in parameterization is Eddington's (Ref. 1), in which the metric concept of general relativity is accepted but the form is parameterized. Thus, the "known" coefficients in the metric for the Schwarzschild solution are replaced by multipole expansions with unknown coefficients; the latter are to be determined from the data. To some, unsatisfying aspects of this parameterization include the lack of direct identification of an expansion coefficient with a "physical" effect. Without further analysis, one cannot tell, for example, whether from a collection of solar system radar observations, γ is being determined primarily from the planetary perihelion advances, primarily from the predicted retardation of the radar signals by solar gravity, or from a more or less equal combination; one cannot tell whether the data are testing small-amplitude, short-period orbital effects or only the perihelion advances, etc.

The approach my colleagues and I adopted when first getting started in gravitation experiments was designed to test as directly as feasible each specific physical effect. We were mindful, of course, of the ad hoc nature of our approach, of the openness to varied interpretations of the words "physical effect," and of some of the theoretical objections that could be made to our approach. What, in fact, is this approach? We start with general relativity as the theoretical "yardstick" and introduce separate ad hoc parameterizations for each of the various different physical effects that are predicted – and, in some cases, not predicted – by Einstein's Theory. More specifically, we developed the general relativistic formulas for particle motion and for light-ray propagation within the solar system. We then considered the anticipated sensitivity of our measurements and the approximate time base over which they would extend (or, more properly, over which our interest would extend). On this basis, we introduced a separate parameter for each of the effects we could conceivably discern: the increase in radar echo time delays attributable directly to solar gravity, the gravitational deflection of radio waves, the various non-Newtonian contributions to particle motion (secular and short-period), a possible time variation of the gravitational constant, a possible violation of the principle of equivalence in regard to the relative contribution of gravitational binding energy to gravitational and inertial mass, etc. The full range of parameters has not yet been incorporated in the enormous digital computer program that has been constructed to analyze the relevant measurements, but the parameterization is safely ahead of our measurement capability. A more explicit characterization of the programmed parameters is included in the discussion of experimental results.

As an illustration of the moot nature of the categorization of phenomena into distinct physical effects, we may consider the radar echo time-delay test (Ref. 7). There are some physicists who claim that this test is "completely equivalent" (Ref. 8) to the light-deflection experiment. The justifications usually given for this classification are: (1) The effect of a gravitational field on light propagation is analogous to the classical electromagnetic effect of an index of refraction: The direction of propagation is bent toward the region of higher index (or gravitational potential) as the speed of propagation decreases. In other words, just as in electromagnetic interactions, so in gravitational ones deflection and propagation speed are both manifestations of the same physical effect. (2) In terms of the Eddington generalized-metric formalism, the first-order expressions for the delay and the deflection both depend on γ only.

Since I first suggested the delay test, it is not too surprising that I take a somewhat different view of the comparison. I feel that one should not try too hard to prejudge nature. Although often useful, if argument by analogy as in (1) above were always accepted as valid in physics, we could never believe that we understand the stability of the atom. Similarly, in regard to (2), there is insufficient reason to assume a priori that the "true" theory of gravitation must be in accord with the generalized metric. It is certainly possible to imagine a theory that would not be consistent even qualitatively with predictions based on the generalized metric. A simple example of this type, which can of course be criticized, involves the assumption that in a gravitational field, photons behave as (slowly moving) mass particles. For such photons, the "extra" delay would be negative; the photons would, in effect, be speeded up rather than slowed down by solar gravity when passing near the sun. The deflection, although again different in magnitude from Einstein's prediction, is toward the sun in both theories.

B. Radar-Echo Time Delays

Although first suggested as a feasible test of general relativity in 1964 (Ref. 7), results from the radar-echo time-delay experiment were not obtained until 1967 (Ref. 9). At that time the 120-ft-diameter Haystack antenna was able to detect radar echoes from both Mercury and Venus at superior conjunction, where the path of the radar waves passes close to the sun's limb. The radar frequency employed was 7.840 GHz. The observations, each of which depended on the (incoherent) integration of the echo signal for a period of several hours, had individual delay uncertainties in the 10- to 15-μs range out of a total round-trip time of about 1500 s. At this level, the effects of the solar corona and planetary topography are unimportant (Refs. 9, 10). On the other hand, the orbits of the planets directly involved, as deduced from classical optical observations, are in some important respects of wholly insufficient accuracy for the purposes of the delay test. Thus, certain characteristics of these orbits had to be estimated from the same type of radar data as were used for the delay test. The other planets, indirectly involved through their perturbative effects, cause no difficulty despite their being unobservable by radar. Because their effect

is only indirect, the classical orbits are sufficiently accurate. Similarly, the classical orbit can be relied upon for the moon, whose main effect is only to cause the observation platform – the earth – to orbit about the earth-moon barycenter.

Other parameters for the theoretical model are also taken from optical results. For example, the radar observations are not very sensitive to the earth's spin and do not involve star observations at all; consequently, the radar data cannot be used to infer the orientation of the solar system with respect to the conventional stellar approximation to an inertial reference frame. We therefore take this orientation and the earth's rotational motion as obtained from classical determinations. Similarly, the mutual orbital inclinations for the inner planets are so small that the radar delay data are relatively insensitive to their effects. So these orbital initial conditions, as well as the corresponding longitudes of the ascending nodes, are also fixed in accord with the conventional wisdom in optical astrometry. But once these orbital conditions (two for each planet) are fixed, the radar data do turn out to be sufficiently sensitive – through the mutual orbital inclinations – to enable the "absolute" longitudes of each of the observed planets to be usefully determined. In summary, the radar data must be used to estimate the in-plane orbital parameters for each of the observed planets, with one exception: The absolute longitude in the adopted system can either be estimated or taken from optical results.

To actually use these radar data to distinguish the predicted delay effects of solar gravity from the remaining orbital effects, we introduced parameters into the equations for planetary motion and for signal propagation. One, which we shall denote by λ_p, multiplies the set of non-Newtonian terms in the harmonic-coordinate formulation of the planetary equations of motion that we have used throughout. The second, λ_s, multiplies the extra, logarithmic term in the coordinate-time expression for the round-trip travel time of the radar pulses. In general relativity, $\lambda_p = \lambda_s = 1$; these parameters being zero would correspond, respectively, to Newtonian equations of planetary motion and rectilinear light propagation at a constant speed.

The number of parameters in addition to the λ's that must be estimated simultaneously from the data becomes rather large: the four orbital initial conditions for each observed planet and for the earth-moon barycenter, one average radius for each target planet, the mass of Mercury, the earth-moon mass ratio, and the speed of light far from the source in astronomical units. With the possible exception of the earth-moon mass ratio, values for the last four parameter types can be determined better from the radar data than from any other source. The speed-of-light parameter enters because of our choice of units: We take the sun's mass to be unity and the day (defined as 86,400 A.1 s) to be the unit of time; the further stipulation that $(GM_\odot)^{1/2} = 0.01720209895\ AU^{3/2}/day$ defines the astronomical unit of length, leaving the speed of light in these units as a parameter to be estimated.

The result obtained from the first experiment for λ_s was 0.9 ± 0.1 (formal standard error), with the estimate of the "true" 1σ uncertainty having been raised to 0.2, partly as a result of the disagreement that then existed between the Haystack and Arecibo radar data (see below). Inclusion of U.S. Naval Observatory meridian-circle observations of the sun and planets from 1950-1967 allowed all the orbital elements of each of the relevant planets to be estimated; the result for λ_s was, as expected, unaffected.

What have data from the past 3 years added? The accuracy of individual measurements has improved somewhat and the time-base over which they extend has tripled, thus yielding far better orbit determinations. (The improved Haystack radar system went into operation only at the end of 1966; the first relativity experiment was based on data to September 1967.) Further, radar data obtained at the Arecibo Observatory, which initially appeared to disagree systematically with Haystack measurements made at nearly the same time, have been reanalyzed and now are in excellent agreement with the Haystack data (Ref. 11). We used both data sets in an attempt to refine our estimate of λ_s. The Arecibo data, with some of the measurement uncertainties at the 3-μs level, contribute importantly to strengthening the orbit determinations and the earth-Venus "excess-delay" measurements. The operating frequency at Arecibo is 430 MHz, and so the interplanetary plasma can produce noticeable effects. Therefore, another parameter was added to estimate the time-averaged electron density at the distance of 1 AU from the sun on the assumption that the density falls off with the inverse square of the distance – an adequate approximation for the region covered by, and the accuracy of, the Arecibo observations. Very few Arecibo observations were made with ray paths passing within about 15 deg or so from the sun. Only the Arecibo Venus observations were included with the Haystack Venus and Mercury data, the re-analysis of the considerable body of Arecibo Mercury data not having been completed in time for inclusion. The currently available high-quality JPL Goldstone radar observations of Venus are confined to the region near inferior conjunction and are too few in number to affect the solution; there are no Goldstone earth-Mercury data. (Earth-Mars radar time-delay observations were also made at Arecibo in 1965, at Haystack in 1967 and 1969, and at Goldstone in 1969; these data are of interest here only in that they serve to further refine the earth's orbit. The special problems and opportunities offered by the large topographic variations – up to 12 km – on Mars constitute too long a story to warrant relating in the present context.)

The result that we have obtained (Ref. 11) from analyzing these data is $\lambda_s = 1.01_5 \pm 0.02$ (formal standard error). A typical sample of the earth-Venus time-delay residuals derived from this solution are displayed in Fig. 1 relative to the "excess delays" given by the logarithmic term in the expression for signal propagation. Based on our considerable experience with these experiments and the possible sources of systematic error, we feel that a realistic estimate of the uncertainty is about 0.05. Thus, our result $\lambda_s = 1.01_5 \pm 0.05$ (i.e., $\gamma \sim 1.03 \pm 0.10$) can be compared with the prediction from general relativity, $\lambda_s = 1$, and with the currently expected

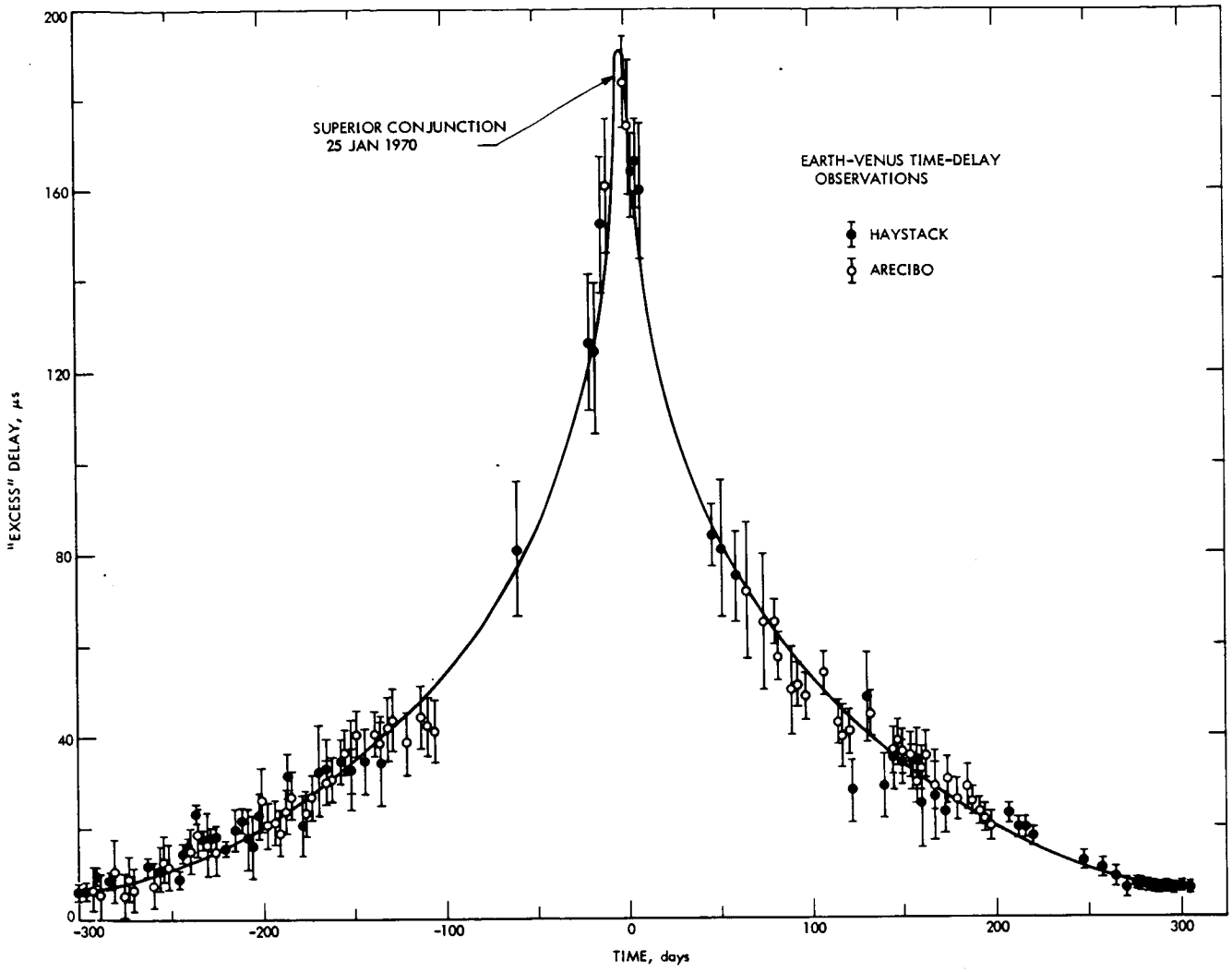


Fig. 1. Typical sample of post-fit residuals for earth-Venus time-delay measurements, displayed relative to the "excess" delays predicted by general relativity. Corrections were made for known topographic trends on Venus. Note the dramatic increase in accuracy that was obtained with the radar system improvements incorporated at Haystack just prior to the inferior conjunction of November 1970.

value, $\lambda_s = 0.93$, based on the Brans-Dicke scalar-tensor theory.* Although far from a conclusive discriminant, our result does not seem to support the scalar-tensor theory with a value of 0.07 for the fractional contributions of the scalar field (see, e.g., Ref. 8). Of course, if general relativity is correct, the existence of a scalar-field interaction can never be disproved by this type of experiment; only successively more stringent upper limits can be placed on the parameters.

The time-averaged value obtained for the interplanetary electron density (Ref. 11) is 7 ± 2 el/cm^3 at the earth's orbit – in excellent agreement with results from other techniques, which are in the range of 5 to 7 el/cm^3. In fact, the agreement is much better than could be expected, considering the uncertainty accompanying our estimate.

C. Perihelion Advance and Solar Oblateness

The accuracy of the radar data is such that λ_p serves to measure with significance only the non-Newtonian perihelion advance of Mercury. If we assume that the solar gravitational quadrupole moment, characterized by J_2, is identically zero, we obtain (Ref. 12) $\lambda_p = 0.98 \pm 0.01$ (formal standard error). Various tests and analyses performed with the data lead us to conclude that, for this secular effect, 0.03 is a more reliable estimate of the true uncertainty in λ_p. It is not possible from these data alone to estimate accurately both λ_p and J_2; the correlation between these parameters is too high to allow a meaningful separation.

How well do the accumulated optical observations of solar system bodies test the predicted

*This value was given by R. H. Dicke at the Third Cambridge Conference on Relativity (New York, June 1970).

non-Newtonian perihelion advances? The definitive, classical analysis indicated that (for $J_2 = 0$) the optical observations are in agreement with the general relativistic prediction of 43"/century excess perihel on advance for Mercury to well within the estimated uncertainty of 1% (Ref. 13 and 14).

To obtain an independent result from the optical data, we (Ref. 15) undertook 6 years ago the massive task of converting almost the entire world's collection of meridian-circle and relevant photographic observations of the sun, moon, planets, and selected asteroids into a common, machine-readable format.* The number of separate measurements involved is about 400,000, spanning the time period from 1750 to 1970. We completed the conversion task this past spring. The analysis of these data is quite complicated and requires discussion of the possible corrections required for the disparate sets of star catalogs that were used by different observatories at the same time and by the same observatory at different times, for declination biases, for biases in the phase corrections used to reduce inner-planet data, for different constants used in different time periods for aberration, nutation, precession, etc. Since it is not possible in this paper to describe our analysis procedures adequately, I shall instead be brief.

We obtained several hundred different least-squares solutions involving the optical data. The number of parameters estimated in any one solution ranged up to 300. The different solutions were required to test the sensitivity of the results to a wide variety of changes in assumptions. Because we obtained good "bench-mark" estimates for almost all of the relevant parameters from initial experimentation, we needed to apply in most cases only one differential correction to find the least-squares, or maximum likelihood, parameter estimates for each case. This process can be carried out very efficiently; on the average, a single solution required about 30 s on our IBM 360/67 computer.

What results did we obtain? Although incomplete, our studies did establish that the formal standard error — considering only the data fluctuations with no allowances for possible, uncompensated biases — for λ_p was 0.01, indicating that, at very best, Mercury's non-Newtonian perihelion advance is determined to 1% by the optical data alone. Our sensitivity studies indicate that 3% is a more reliable percent value for this standard error, with the estimate in agreement with general relativity within that range. We are thus led to conclude that past estimates of the accuracy of the determination of Mercury's perihelion advance were somewhat optimistic.

The combination of the optical and radar data provides a powerful tool for the estimate of λ_p. Our preliminary solutions yield a formal standard error of 0.004; the "true" 1σ uncertainty is probably at least 0.01. Considerably more analysis of this combined data set is required before any firm conclusions can be drawn.

Will these combined data allow us to estimate λ_p and J_2 simultaneously and to separate their contributions at a significant level of accuracy? We made some computer studies to answer this question; a "typical" solution, in which 150 parameters were estimated, yielded $\lambda_p = 0.96 \pm 0.03$ and $J_2 = (0.8 \pm 0.8) \times 10^{-5}$ (formal standard errors; J_2, defined in the usual manner, is dimensionless). Further analysis is definitely required to test the sensitivity of these results to changes in assumptions. Enough work has already been done, however, to convince us that the true uncertainty in J_2 is no less, and perhaps even somewhat larger, than 3×10^{-5}. Upon comparing such a value with Dicke's deduction of $J_2 = (2.7 \pm 0.5) \times 10^{-5}$ from measurements of the sun's visual oblateness (Ref. 16), we conclude that presently existing optical and radar data cannot provide a useful estimate of J_2.

D. Time Variation of the Gravitational Constant

Many conjectures** have been made during the past few decades concerning a possible variation with time of the gravitational "constant" G. Several years ago, I pointed out (Ref. 7) that the development of planetary radar systems and atomic clocks has made possible the placement of a fairly stringent experimental limit on the magnitude of $\dot{G}$. Gravitational time can be compared with atomic time by making, in effect, repeated measurements on an atomic-time scale of the orbital period of a planet. The Mercury radar data are more significant than the Venus data for this purpose in virtue of Mercury's five times higher orbital angular velocity with respect to the earth.

The clocks used for the measurements may be assumed to have kept track of atomic time without error, since this source of uncertainty is far too small to be of significance for our experiment. The set of intercompared cesium beam atomic standards used by the U.S. Naval Observatory to determine A.1 time have long-term errors of only about 1 part in 10^{12}, corresponding to less than 200 μs error in epoch after 6 years. The effect on the interpretation of a delay measurement between the earth and Mercury is therefore always less than 0.04 μs — far less than the delay measurement uncertainties themselves, which were never less than 5 μs. Similarly, the clock errors contributed insignificantly to the measurement of delay, since the former, over the round-trip times, were always accurate to within 2 parts in 10^{12}, whereas the latter were never more accurate than 2 parts in 10^9. Thus, for both important functions — the

*The optical observations were culled from modern and ancient observatory reports and transformed into machine-readable form in a common format. An exception was Eros, the observations of which were kindly provided to us in machine-readable form by J. H. Lieske.

**See, for an early example, P. A. M. Dirac, Proc. Roy. Soc. A165, 199 (1938).

determinations of epochs and intervals – the atomic clocks can be considered errorless.

The somewhat ad hoc model that we used to analyze the data for a possible variation in G can be described briefly. Each planet was assumed to obey the usual equations of motion that follow from the Schwarzschild metric for the sun and from the Newtonian perturbations attributable to the moon and other planets, except that the gravitational constant was replaced by $G_\odot + \dot{G}_o (t - t_o)$, where the coordinate time t_o is some (arbitrary) epoch at which G and $\dot{G}$ are evaluated. This formulation appears adequate to test most cosmological theories, especially since the time span of our data is relatively so short.

The use of the usual Schwarzschild metric to determine the gravitational effect of the sun instead of, say, the corresponding metric for the Brans-Dicke theory, has no substantial effect on our estimate of $\dot{G}$, nor do possible small spatial variations of G within the solar system. A similar conclusion follows for our assumption of a zero solar gravitational quadrupole moment and a zero solar mass loss.

Our result shows no evidence for a time variation of the gravitational constant, the magnitude of the estimate of $\dot{G}/G$ being only a small fraction of the formal standard error, 1.5×10^{-10}/year. To make a reasonable allowance for unknown but possibly important vitiating effects on our estimate, we take 4×10^{-10}/year as a more reliable indicator of the actual uncertainty (Ref. 17). We are not aware of any other experimental limit on $\dot{G}/G$ of comparable stringency.

The use of optical data may not improve this result, since prior to 1956, atomic time was not kept continuously. The earth's rotation time is all that would be available for comparison with orbital time, but, unfortunately, the former is affected by relatively large variations that are not well understood.

E. Deflection of Radio Waves

The first suggestion to use radio interferometry to detect the predicted deflection of electromagnetic waves by solar gravity was made in 1967 (Ref. 18). Since then, my colleagues and I* have been trying to determine accurately the relative deflection of radio waves from the quasars 3C279 and 3C273. The former is occulted by the sun on October 8. The most promising data were obtained last month using the "Goldstack" very-long-baseline interferometer composed of the 210-ft-diameter antenna at Goldstone and Haystack's 120-ft-diameter dish. Observations were made at 7.840 GHz to optimize signal-to-noise ratios (both systems were equipped with low-noise receivers operable at this frequency) and to minimize solar corona problems. Two NRAO Mark I tape recording systems were available at each site so that data could be obtained almost continuously. The procedure followed was to observe 3C279 for about 2 min, then to move the antenna to observe 3C273 for about 1 min. Since 1 min was allowed to swing the antennas between sources, one "cycle" of observations occupied 5 min. These cycles were repeated without a break throughout the 4 to 5 h of observation that were scheduled for a given day. Observations were taken each day from October 1 to 6 and from October 10 to 15. Except for the observations of 3C279 on the 6th and the 10th, when solar corona problems were apparently severe, fringes have already been obtained on over 90% of the total of over 1200 3-min tape recordings that were made. We are currently trying to connect the fringe phases from successive cycles unambiguously; if successful, the result should yield a value for the deflection with an error no greater than a few percent of the value predicted by general relativity. The fringe-rate data can also be used to detect the deflection, but with far less accuracy.

While we have not yet obtained any results, two groups at Caltech have already reported deflection values found from short-baseline radio interferometry 3C279 and 3C273 experiments (Ref. 19). These results, which agree with general relativistic predictions to well within the 10% formal standard errors quoted, compare favorably with the best values so far obtainable from optical photographs of stars during total solar eclipses.

III. Future Prospects for Ground-Based Tests of General Relativity

Here we shall describe with extreme brevity the possible improvements in some of the above-described tests that might be achievable in the next few years.

A. Goldstack Radar System

Only relatively small improvements can be expected in the time-delay measurement accuracy that can be achieved near the superior conjunctions of Mercury and Venus with Haystack. However, a bistatic radar combination with the 210-ft antenna at Goldstone being the receiver and Haystack the transmitter, affords a potential gain in signal-to-noise of about 10 dB over Haystack alone. This Goldstack configuration was tested out for the first time last summer when Mercury was near superior conjunction.** Although analysis of these data is incomplete, it seems clear that the system worked and yielded about an 8-dB improvement over Haystack. Certain anomalies in the data indicate that Mercury's scattering law, especially in its effect on 10-μs-baud-length phase-coded signals, may be highly variable with aspect. This interpretation is consistent with recent radar observations of Mercury made at Arecibo which, at an almost 20 times lower frequency, exhibited striking variations in detectability when these short baud-length codes were used (Ref. 20).

*Members of the group working on these experiments include T. A. Clark (GSFC), R. M. Goldstein (JPL), and H. F. Hinteregger, C. A. Knight, D. S. Robertson, A. E. E. Rogers, I. I. Shapiro, and A. R. Whitney (MIT).

**The Goldstack experiment was carried out jointly by R. M. Goldstein, J. H. Lieske, and W. G. Melbourne (JPL) and R. P. Ingalls and I. I. Shapiro (M.I.T.).

A very substantial improvement over the prior echo-time-delay tests could be achieved with the Goldstack radar next summer near the superior conjunction of Venus. This predicted improvement is based on a number of factors: (1) A new 400-kW S-band transmitter will be available at Goldstone, thus allowing alternate S-band monostatic and X-band bistatic observations to eliminate solar corona effects almost completely. Because changeover between frequencies may be time-consuming, a better procedure may be to defer the S-band observations until the end of each day's bistatic observations, i.e., until after Venus has "set" at Haystack. Less observing time would be required at S-band to achieve the same accuracy as at X-band, since the two systems (monostatic and bistatic) are about comparable in sensitivity but the Venus atmosphere attenuates the S-band signal about 6 dB less. Radar observations made at Arecibo would allow a further check on the coronal effects. (2) Venus presents the same face to the earth at both inferior and superior conjunctions; hence, the scattering law is well known at the short baud-lengths that would be employed. Furthermore, from past and present observations at inferior conjunctions, the topography (i.e., the surface height variations) will soon be determined with sufficient precision so as not to degrade the results of the bistatic radar time-delay test (see also Section III. B). (3) Superior conjunction occurs in mid-summer and therefore at reasonably high declinations, which extend the time of mutual visibility between Goldstone and Haystack. (4) The radar observations already accumulated are very well distributed around Venus' and the earth's orbits and are more accurate on the average than the observations of Mercury, and thus provide better relative a priori orbits. Many of these observations, especially those taken surrounding the past two inferior conjunctions, have uncertainties of no more than 3 μs; quite a few have even smaller errors. In combination with an extended series of Goldstack earth-Venus observations, the "orbital problem" should be solvable at least at the 2- to 3-μs level and should therefore not degrade significantly the propsects for the time-delay test. The good agreement — at the several-microsecond level — obtained between the most precise Haystack and Arecibo delay measurements that were made nearly simultaneously (see Fig. 1) indicates that there will probably be no insuperable problems in relating the results of Goldstack Venus observations to the other Venus data. In any event, to the extent that the scattering law of Venus is independent of aspect, one additional bias parameter would suffice to remove any constant difference in delay estimates to the "leading edge" of the target. Of course, Venus, too, is known to have aspect dependences in its scattering law, but these have already been quite well studied at S- and at X-band. The 8-year cycle in the relative orbits of the earth and Venus insures almost exact repetitions of aspect. Radar data have already spanned one complete cycle.

B. Next-Generation Radar Systems

Aside from languishing documents such as the NEROC proposal for a 440-ft-diameter radome-enclosed, fully steerable antenna usable efficiently at frequencies at least up to X-band, the only dramatically improved radar system on the horizon is the proposed upgrading of the Arecibo system. Under this plan, the accuracy of the inner 600-ft-diameter part of the antenna will be improved to make it usable efficiently at S-band frequencies. A powerful S-band radar transmitter system and low-noise receiver system have been proposed for use with the improved antenna. This project is apparently to be started in the near future and hopefully will result by the mid-1970's in a radar system with about 20 dB greater sensitivity than any presently available.

With the increased measurement precision that will become possible with this instrument (errors <1 μs at all points in the relative orbits of the earth, Mercury, Venus, and Mars that are visible at Arecibo), the topography and plasma problems loom more important in the use of such measurements to improve the radar tests of gravitational theories discussed in Section II. But the problems posed are not insuperable — at least in principle, if not in practice. For ray paths that don't pass closer than 1 to 2 deg from the solar limb, two-frequency operation (say, at 2500 and 3500 MHz) should enable the solar corona to be calibrated with a loss in precision of only a factor of about 2 introduced by the errors inherent in the two-frequency calibration procedure. The limitations on the usefulness of the close approaches (≲1 deg) of the ray path to the sun depend critically on what might be termed the turbulence of the corona, which is rather time-variable. These limitations are set primarily by the loss of coherence in the echoes, which prevent precision delay measurements. A far more serious problem would undoubtedly be the increased funding required for two-frequency operation. The coronal problems, in any event, will be a serious limitation only for the test of the direct effect of solar gravity on echo delays; the other tests involve secular orbital effects for which measurements near superior conjunction aren't needed, and the plasma need not begin to degrade individual measurement accuracies unless the latter penetrate the 0.1-μs level.

Topography must certainly be calibrated to make full use of these precision radar measurements. Luckily, the time scale for secular changes in surface structure is likely to be long compared even to the average astronomer's lifetime. We may therefore count on topography calibration remaining valid over all time scales of concern here. Various methods are being developed for this calibration. One technique that yields good resolution both in altitude and in the surface area region to which the altitude refers makes use of the power-vs-delay profiles for each small frequency band in the radar echoes from the target planet. In essence, the topographic highs and lows are determined from the first arrival of echo power in these profiles, as compared to the first arrival time to be expected for a spherical target. The technique is applicable only along the Doppler equator since, for the energy returned in a narrow-frequency band, the echo power received first is reflected from the region of the Doppler equator (Ref. 10). Of course, for purposes of the measurement of echo time delay referred to the subradar point, or "leading edge," on the target planet, the important topography lies on the Doppler equator. The reliability of the topographic corrections can be accurately tested by comparison of delays measured at the same subradar point for each of the four rotations made by Venus as seen by

the earth-observer between successive inferior conjunctions.

Another technique (Ref. 21) will enable full, three-dimensional maps of the planets to be obtained. It builds on the by now well-known radar delay-Doppler technique: By adding the fringe-phase information obtainable from radar interferometry to the delay-Doppler data, one can obtain the three-dimensional coordinates of each reflecting area. Calibration of the fringe phase can be accomplished by comparison with the phase for the subradar point reflections.* The most useful antenna spacings will probably lie in the several-hundred-kilometer range. There is no doubt that the successful application of this technique to yield surface height resolution of about 50 to 100 m on Venus is fraught with technological difficulties, but it is not clear that any are insurmountable and it is clear that the technique is workable. Good preliminary results have already been obtained for the moon using this technique with the Haystack-Westford interferometer (Ref. 21). (The second element is a 60-ft-diameter antenna system located about 1 km from Haystack.)

In summary, although much more careful study is required to assess the achievable accuracy reliably, it appears that presently planned advanced radar systems can be used to measure interplanetary echo delays at some relative orientations with errors of the order of only 0.1 μs, and that almost full use can be made of the accuracy attainable for almost all of the inner planet configurations in the performance of tests of gravitational theories such as those described in Section II.** The resultant accuracies of such tests will be discussed in Section IV.

C. Very-Long-Baseline Interferometry (VLBI)

The quasar pair 3C279 and 3C273 at present appears to be the most suitable for the deflection-of-light experiment. Other, better ones might be uncovered in the near future. For example, if point-source water-vapor emissions were found near the OH sources W28S and W28N, one would have a pair that passes on opposite sides of the sun each December 22. Water vapor, if also present near the OH source W24, would provide a nearby calibrator. The advantage of such sources, aside from their proximity to the apparent solar path in the sky, lies in the high frequency (≈23 GHz) of the emissions, which would markedly decrease the solar corona effects. At the rate at which spectral-line emissions are being discovered in the galaxy, it certainly seems reasonable to expect that some high-frequency sources will be found near the ecliptic.

Aside from these possibilities for new sources, what are the prospects for improvement of the deflection experiment? I suspect them to be quite good, but will defer until Section IV any explicit estimates of achievable accuracy and devote myself here only to a brief description of the manner in which such improvements might be realized. The main limitations on accuracy are: (1) the complicated structure of 3C279 (and perhaps 3C273); (2) the solar corona; (3) the earth's atmosphere and ionosphere; (4) the available baseline projections; and (5) the phase stability of the receivers. I shall discuss each in turn.

Preliminary examination of the time-dependence of the fringe amplitude of 3C279 indicates that the source structure at the level of interest is quite complicated but should not present insurmountable problems. The beauty of the relativistic deflection effect is that it changes sign on October 8; it is hard to believe that anything in the source would appear to vary in a correlated manner.

The solar corona and the earth's ionosphere can be handled in principle by observing at a very high radio frequency or simultaneously at two widely separated frequencies — say, 7500 and 10,000 MHz. A large fractional separation is needed for accurate plasma calibration; high absolute frequencies are desirable to allow fringe-phase information to be obtained closer to the sun and to attempt to ensure that the paths followed by the "rays" at the two frequencies pass through nearly the same plasma environment.

The differential earth's atmospheric effects can be calibrated in a variety of ways. One promising method involves monitoring radiometrically, along the line of sight to the source, the antenna temperature of two narrow bands — one at and one near the 23-GHz water vapor absorption line.+ The accuracy of the calibration is uncertain; preliminary indications are that, on an absolute scale, the electrical path length along the line of sight can be estimated with errors equivalent to only about 1 to 2 cm. For the deflection experiment, the atmospheric effect enters, in essence, only as a triple difference: the difference on different days of the difference in fringe phases from 3C279 and 3C273, each phase being determined from the difference in the phase delays from the source to the separate elements of the interferometer. Thus, a high order of cancellation of the residual, uncalibrated contribution of the atmospheric effects can be anticipated.

Aside from the constraint imposed by possible source resolution, the optimum baseline to use in the deflection experiment is a compromise between achieving the longest one possible (along the right direction — primarily east-west for the 3C279-3C273 pair) and having a long enough daily period of common visibility to measure the variations in fringe phase from which the deflection can be deduced. The "absolute" fringe phase is ambiguous by multiples of 2π unless a very wide effective bandwidth is employed. Such a wide-bandwidth technique has been developed at M.I.T. and used successfully in VLBI experiments, but has not yet evolved to the level where the fringe-phase ambiguity can be completely removed. This removal is, however, well within state-of-the-art capabilities. Thus, unless the fringe-phase ambiguity

*Venus atmospheric effects tend to cancel, since the beams to each antenna of the interferometer from a given reflecting element on the planet's surface pass through almost identical paths in its atmosphere.

**The deflection-of-light experiment can also be done by radar but would probably not be competitive with the quasar approach.

+This technique is being developed by D. Staelin and J. Waters in collaboration with the National Radio Astronomy Observatory.

is removed, reliance in the deflection experiment must be placed on monitoring continuously the variations in the relative 3C279-3C273 fringe phases during each day's observations.

The phase stability of the receiver systems clearly enters importantly into the realization of the highest possible accuracy in this experiment. With current hydrogen-maser frequency standards and with carefully constructed local oscillator chains driven directly from a 100-MHz, or higher, frequency output from the maser, the short-term (≈30-s) system phase noise ought to be contained at the several-degree level at X-band observing frequencies. The main limitation on accuracy might then be the unknown component of the earth's polar motion during the course of observations. But other VLBI observations, made during the same period, could be used to reduce this "noise" contribution sharply.

The "ideal" setup, then, to conduct the deflection experiment would involve a continuous monitoring of the relative 3C279-3C273 fringe phase simultaneously at two widely separated frequencies over an optimum baseline. (The requirement of continuous observations could be relaxed if the fringe ambiguity were removed via wide-bandwidth observations.) These goals could be realized by using a pair of antennas at each end of the baseline, with one antenna of each pair observing 3C279 and the other 3C273. The video signals from both sources could be recorded simultaneously, with only a small penalty in signal-to-noise ratio.

Of course, improvements on this "ideal" setup are not hard to envision; for example, more sites could be added to the interferometer to allow truly continuous monitoring and to improve the "uv plane" coverage. The corona would be easiest to calibrate during solar minimum (the mid-1970's).

IV. Comparison of Spacecraft and Ground-Based Techniques: Accuracy vs Cost

By the expression "ground-based techniques" I imply experiments in which all the man-made instrumentation resides on the earth. I examine below very briefly the relative merits and demerits of the ground-based and spacecraft techniques in the context of experiments designed to test various aspects of gravitational theories. Both accuracy and cost are considered in the comparison.

A. "Red-Shift" Experiment

The classical red-shift experiment has already been performed in an earth-based laboratory to an accuracy of 1% (Ref. 22) at an estimated cost for time and labor of several tens of kilodollars. The experimenters estimate (Ref. 23) that at an expenditure level near 100 kilodollars, the accuracy could be improved by about one order of magnitude (mainly through use, I believe, of a longer shaft). It has been proposed (Ref. 24) to place a pair of hydrogen-maser frequency standards in an eccentric orbit about the earth at a synchronous mean altitude. This experiment might yield an uncertainty as low as 10^{-3}%, i.e., two orders of magnitude more accurate than a possible improved Mossbauer laboratory-based experiment. But the earth-orbiting satellite experiment, if charged solely to the hydrogen-maser red-shift measurements, would have a cost reckoned in the tens of megadollars. A proposal that I made to include an "absolute" frequency standard — such as a cesium clock — aboard the Mariner Venus/Mercury spacecraft could yield a test over a "nonlinear" region of the solar gravitational potential with an accuracy approaching 0.1%. The incremental cost of this experiment, however, would probably be several megadollars.

A possibly competitive ground-based experiment (at the 0.1% level) might be accomplished by monitoring arrival times of pulsar pulses for a period of several years.* The pulsar (in principle) could be expected to transmit at equally spaced "coordinate-time" intervals, whereas atomic clocks on earth, used to measure the arrival times, would vary relative to these because of the varying solar gravitational potential through which the earth-based clock passes (and because of other effects such as Doppler shifts). The amplitude of the atomic-clock variation is about 1-1/2 ms. The difficulty with accurate measurement is that, for the pulsars so far studied, stable periods are accompanied by long pulse lengths (many milliseconds), making difficult a really precise measurement of the arrival time, say at the microsecond level. Only for the Crab pulsar — the one with the shortest pulse length known — have the arrival times been measured with errors at the several-microsecond level. But the Crab pulsar undergoes irregular, and so far unpredictable, changes in pulse rate which make a useful red-shift experiment difficult to carry out. (There is, of course, also the problem of estimating the pulsar position relative to the earth's orbital plane, which, for the Crab pulsar, will be highly correlated with the effect sought.) It is possible, of course, that new pulsars with more desirable properties — at least from this specialized point of view — might be discovered in the near future. Costs here for use of existing facilities would be in the 100 kilodollar range.

All in all the spacecraft approach appears appreciably more promising. But the costs are very high, especially since the outcome of this experiment is very unlikely to be different from predictions.

B. Principle of Equivalence

The principle of equivalence has been tested in the laboratory (Ref. 25) to about 3 parts in 10^{11}. But such tests cannot hope to reach sensitivity levels at which the relative contributions of the gravitational binding energy to inertial and gravitational mass could be determined. Much larger masses are required before the gravitational binding energy becomes a sensible fraction of the total rest energy. It is well known and easily demonstrated that a simple violation of this so-called weak principle of equivalence cannot be discerned from studies of two-body orbital motion. At least three mutually orbiting bodies are required.

We have been examining the observable consequences of a possible equivalence principle

*This possibility was first suggested by B. Hoffman.

violation in the sense of the gravitational binding energy contributing to only one of the two mass types.* This investigation has centered on earth-Mars and earth-moon measurements and, for the former case, has included reasonably sophisticated models of the "noise" introduced by the asteroids. Although final conclusions have not yet been reached, ground-based earth-Mars experiments appear inadequate with presently foreseeable measurement accuracies. A combination of results from the Mariner Mars 1971 and the Viking 1975 missions might just barely yield a useful result. But the cost — if all allocated to this experiment — would be in the kilomegadollar range!

Preliminary results indicate that lunar laser observations might be more fruitful in this regard. With measurement errors at the 0.1 ns level, several years of data might provide interesting results. (Much more analysis is required before this tentative statement can be given any credence.) The cost for such lunar laser experiments would probably be in the several megadollar range — provided one doesn't include any costs to place the corner reflectors on the moon.

C. Deflection of Light

Spacecraft, when nearly occulted by the sun, can be used in the same manner as quasars to measure the gravitational deflection of light. The advantages of spacecraft are (1) the true point-source nature of their radio (or optical) emissions and (2) the potentially greater signal strength. Spacecraft would best be required to transmit on two well separated frequency bands if radio signals are used and on only one if the transmissions are in the optical range. Knowledge of the spacecraft orbit could be obtained with the requisite accuracy by having the spacecraft gravitationally anchored to, or actually emplaced on, a planet. "Drag-free" techniques could be used to advantage for non-anchored craft. Requirements aren't too severe: A 1-km orbital position uncertainty for a spacecraft at 1.5 AU from the sun causes an angular position uncertainty, as viewed from the earth at superior conjunction, of only about 0.0005 arc-s. Without a comparison companion reasonably nearby, however, atmospheric effects may seriously limit the achievable accuracy. Differencing, as described in Section IIIC, is a powerful technique for eliminating systematic errors. Depending on the orientation conditions, it may well be possible to use a quasar as the comparison for the spacecraft. The cost of the experiment — if the spacecraft is allotted totally to this test — would be in the 100-megadollar region. Because suitable spacecraft were planned for other purposes, I proposed this deflection experiment for the Mariner and Viking missions scheduled for the next half decade. However, the accuracy achievable may not be appreciably different from the accuracy achievable with a full exploitation of the ground-based technique. This latter might entail expenditures at the several-hundred-kilodollar level (given that no new radio telescopes would be built for this purpose). What accuracy can, in fact, be achieved? Although predictions of this sort are notoriously unreliable, still, I feel that the error in the estimation of the deflection ought to be reducible to the 0.5% level and perhaps even somewhat lower. A detailed discussion of the reasons for, and possible faults in, these predictions is not possible here.

D. Echo Time Delays

Ground-based observations may place bounds on the uncertainty of the predicted solar gravitational effect on radar-echo time delays as low as 0.3% with the next generation radar system planned for Arecibo — if the solar corona can be calibrated with sufficient accuracy. The cost would be at the 5-megadollar level if the capital investment for the new radar system is included, and at the 1-megadollar level if it is not.

Spacecraft of the Mariner 1969 type can probably not be used to determine these "excess" echo delays with uncertainties below several percent. The main limitations are (1) the solar corona, which is dynamic, not static, and therefore difficult to model and (2) the "process noise" effects on the spacecraft orbit, which inhibit the interpretation of the measurements further from the sun.**

Neither of these limitations is uncorrectable. In the Mercury/Venus flyby mission, dual-frequency (S-band and X-band) downlink coherent ranging and Doppler capability will exist. Although only available on the spacecraft-ground link, the plasma calibration should be accurate to about the 10-ns level for ray paths that don't pass closer than a few degrees to the sun. The superior conjunction should take place in the "quiet sun" period, a further advantage. The orbital uncertainties will still be severe, but somewhat less so than for Mariner 1969, because the Mercury/Venus spacecraft will "pass through" the superior conjunction region several times faster, with a consequent smaller period of time available for a buildup of the corrupting effects of process noise. Offsetting this advantage to some extent is the higher level of unmodeled sunlight-pressure accelerations.

Orbital uncertainties can be greatly reduced by planetary "anchoring." Thus, the two Mariner 1971 spacecraft are to be placed in orbit about Mars. The spacecraft orbits can therefore be calibrated with respect to Mars' orbit over time intervals that are short compared to the time required for the effect of the process noise to build up to intolerable levels. This calibration is made possible by the relatively short orbital periods of the spacecraft about Mars. The unknown gravity anomalies in Mars are also a source of concern, but since the drift (or precession) of the spacecraft orbits with respect to Mars is slow, the Mars gravity field (except for rotational effects on the one non-resonant orbit) will remain essentially fixed for the "duration" of the superior conjunction phase.

Unfortunately, the Mariner 1971 spacecraft, like Mariner 1969, will have only S-band tracking

*This work is being carried out primarily by G. Sherman at M.I.T.

**For a complete discussion of the Mariner 1969 results, which are almost identical to the radar determinations given in Section IIB, see J. D. Anderson, P. B. Esposito, W. L. Martin, and D. O. Muhleman, this proceedings.

equipment. Viking, on the other hand, will have dual S/X-band frequency capability on the two Mars orbiters and S-band capability on the two landers.

The accuracy achievable from these missions in the time-delay test might be at the 0.1% level with the Vikings and perhaps the 0.3% level with the Mercury/Venus mission; Mariner 1971, because of the single-frequency capability, may be limited to about the 1% level. More study would be required to see whether the benefits of planetary anchoring will allow this limit to be lowered. The costs of the latter two are in the 100-megadollar class, whereas the Viking mission is nearly at the kilomegadollar level. Of course, if these missions were to be performed solely to test general relativity, their costs would be drastically lowered, perhaps to the 30 megadollar range.

E. Orbital Motion of Massive Particles

There are a number of small but interesting non-Newtonian effects that could be examined, but I shall restrict the discussion solely to the problem of distinguishing J_2 and λ_p. The present radar time-delay data set, as mentioned, is incapable of estimating J_2 and λ_p simultaneously at a meaningful level of accuracy. However, if the radar measurements are continued and the expected accuracy improvements attained (see above), then a useful separation could be obtained by the middle to late 1970s. Thus, a covariance analysis shows that with 3 years of data from the upgraded Arecibo radar added, the uncertainty of J_2 would be reduced to about 3×10^{-6} and that of λ_p to 0.3%.

Spacecraft orbits that pass as close to the sun as about 0.2 AU offer the best possibility for an achievement of comparable accuracy in the simultaneous determination of J_2 and λ_p (Ref. 26). But the combination of the optical, radar, and spacecraft data probably offers the most powerful tool for the reduction of the uncertainty in estimates of these quantities.

Thus, we may anticipate that in the 1970s the gravitational oblateness of the sun will be estimated from its dynamical effects directly, at a level of accuracy sufficient to either confirm or refute the interpretation of the Princeton solar oblateness experiment in terms of J_2.

F. Gyroscope Precession

The relativistic contribution to the precession of the moon's orbit may be distinguishable from ground-based observations, but certainly not accurately. The corresponding contribution to the precession of the earth's pole of rotation may be inferred from long-term VLBI observations of quasars. These latter data must be used to determine both the earth's precession and nutation so as to be able to distinguish the "geodesic precession" from present uncertainties in the fractional difference between the earth's polar and equatorial moments of inertia. Aside from the long time involved – at least as long as the 18-year period of the nutation – there are also important difficulties in accounting properly for the non-rigid-body aspects of the earth's rotation. In any event, these ground-based techniques do not seem even remotely capable of discerning the Lense-Thirring effect, which is due to rotation of the primary body. The only hope here lies at present with earth-orbiting gyroscopes which would be able as well to measure the geodesic precession accurately (Ref. 27). The cost of such a spacecraft experiment would probably reach nearly 50 megadollars if it alone bore the launching expense.

G. Variation of Gravitational Constant

Radar observations of Mercury will continue to provide the best means for reducing the upper limit on possible time variations of the gravitational constant until such time, for example, as long-lived transponders are placed on Mercury's surface near its pole of rotation to facilitate continuous tracking from the earth. Lunar laser observations, partly because of problems with earth-moon tidal interactions, seem far less suitable for this task. By the middle to late 1970s, if regular radar observations of the expected accuracy are continued through this period, the error in the estimate of $\dot{G}/G$ would be reducible reliably to 1×10^{-11} per year. The cost, if all but the capital investments were charged to this experiment, would reach the several-megadollar level. The placement of a transponder on Mercury would involve at least two orders of magnitude greater expenditure.

H. Conclusions

The above discussion can be summarized as follows: (1) Testing gravitational theories at a meaningful level is expensive. (2) Spacecraft experiments are about two orders of magnitude more expensive than ground-based ones. (3) Spacecraft experiments have the potential to "outperform" their ground-based counterparts. (4) In three of the most important presently feasible experiments – the deflection of light, the echo time delay, and the relativistic perihelion advance-solar oblateness separation – ground-based techniques appear capable of doing the job at a 1% or better level of accuracy.

These statements lead me to conclude that, at present, a spacecraft mission devoted solely to gravity exploration will probably not provide sufficiently important scientific results to warrant the cost. On the other hand, a spacecraft mission justified on other grounds should be modified to optimize its usefulness for testing gravitational theory if the incremental cost does not exceed more than a few percent of the total. Although many qualifications are needed, this approach would seem to strike a better balance between the costs of similar ground-based and spacecraft experiments. We presuppose, of course, that ground-based gravity experiments will also be supported.

References

1. Eddington, A. S., The Mathematical Theory of Relativity, p. 105, Cambridge University Press, New York, 1957.

2. Robertson, H. P., in Space Age Astronomy, Edited by A. J. Deutsch and W. E. Klemperer, p. 228, Academic Press, New York, 1962.

3. Dicke, R. H., The Theoretical Significance of Experimental Relativity, Gordon and Breach, Inc., New York, 1964.

4. Nordtvedt, K., Jr., Phys. Rev. 169, 1017, 1968.

5. Thorne, K. S., and Will, C. M., this proceedings.

6. Will, C. M., Astrophys. J., 163, 611, 1971.

7. Shapiro, I. I., Phys. Rev. Letters, 13, 789, 1964.

8. Dicke, R. H., and Peebles, P. J. E., Space Sci. Rev., 4, 419, 1965.

9. Shapiro, I. I., Pettengill, G. H., Ash, M. E., Stone, M. L., Smith, W. B., Ingalls, R. P., and Brockelman, R. A., Phys. Rev. Letters, 20, 1265, 1968.

10. Smith, W. B., Ingalls, R. P., Shapiro, I. I., and Ash, M. E., Radio Science, 5, 411, 1970. See also, Ingalls, R. P., and Rainville, L. P., in preparation.

11. Shapiro, I. I., Ash, M. E., Ingalls, R. P., Smith, W. B., Campbell, D. B., Dyce, R. B., Jurgens, R. F., and Pettengill, G. H., Phys. Rev. Letters, 26, 1132, 1971.

12. Shapiro, I. I., et al., in preparation.

13. Clemence, G. M., Astron. Papers Am. Ephemeris 11, Part 1, 1943.

14. Duncombe, R. L., Astron. Papers Am. Ephemeris 16, Part 1, 1958.

15. Shapiro, I. I., Ash, M. E., and Smith, W. B., in preparation.

16. Dicke, R. H., and Goldenberg, H. M., Phys. Rev. Letters, 18, 313, 1967.

17. Shapiro, I. I., Smith, W. B., Ash, M. E., Ingalls, R. P., and Pettengill, G. H., Phys. Rev. Letters, 26, 27, 1971.

18. Shapiro, I. I., Science, 157, 806, 1967.

19. Seielstad, G. A., Sramek, R. A., and Weiler, K. W., Phys. Rev. Letters, 24, 1373, 1970; Muhleman, D. O., Ekers, R. D., and Fomalont, E. B., Phys. Rev. Letters, 24, 1377, 1970.

20. Pettengill, G. H., private communication.

21. Shapiro, I. I., Slade, M. A., Rogers, A. E. E., Zisk, S. H., and Thompson, T. W., in preparation.

22. Pound, R. V., and Snider, J. L., Phys. Rev., 140, B788, 1965.

23. Pound, R. V., private communication.

24. Kleppner, D., Vessot, R. F. C., and Ramsey, N. F., Astrophys. Space Sci., 6, 13, 1970.

25. Roll, P. G., Krotkov, R., and Dicke, R. H., Annals Phys. (New York), 26, 442, 1964.

26. Anderson, J. D., and Esposito, P. B., AIAA Preprint No. 70-1317, prepared for AIAA 7th Annual Meeting and Technical Display, Houston, Texas, October 19-22, 1970.

27. Everitt, C. W. F., this proceedings.

The Effects of Random Accelerations on Estimation Accuracy With Applications to the Mariner 1969 Relativity Experiment

D. W. Curkendall, S. G. Finley, M. W. Nead,
V. J. Ondrasik, and C. L. Thornton
Jet Propulsion Laboratory, California Institute of Technology

I. Introduction

In a parameter estimation problem, a problem which characterizes most of orbit determination in general and determination of relativity parameters in particular, the calculation of the accuracy of the estimates is a difficult task. Insofar as the physical model used in establishing the estimation procedure is a faithful replica of the real world, the error covariance is calculated as an adjunct to that estimation procedure and is easily determined. It is the determination of the impact of the likely discrepancies between the estimation filter's view of reality and reality itself that poses the most serious difficulties.

In this paper, we develop a simple point of view toward establishing the relationship between a general modeling error and the estimation accuracy. In particular, we explore and provide tentative models for the random, non-gravitational forces thought likely to be affecting the Mariner VI and VII spacecraft. (The navigation data from these spacecraft were used to determine the relativity parameter γ, as reported by Anderson et al., this proceedings.) The effect that these accelerations have on the classical least-squares filter, which pretends that they do not exist, is calculated, and it is demonstrated that their presence seriously affects the estimates, particularly when an extended data arc is employed. Finally, some preliminary exploration using sequential-type filters, which attempt to model and thus ameliorate the effects of these same accelerations, is reported.

II. On the Effect of Modeling Errors

In an estimation problem, the data z will be related to the unknown parameters of the system X through the relation

$$z(t) = h(t, X) + \epsilon(t) \tag{1}$$

where ϵ represents the data error. The problem is: given a set of observations $z(t_i)$ at discrete time points t_i, estimate X in the presence of ϵ. We transform to a linear problem by making an initial guess of X, X_o, constructing

$$H \triangleq \begin{bmatrix} \dfrac{\partial h(t_1, X_o)}{\partial X} \\ \dfrac{\partial h(t_2, X_o)}{\partial X} \\ \cdot \\ \cdot \\ \dfrac{\partial h(t_m, X_o)}{\partial X} \end{bmatrix}$$

$$\delta z(t_i) = z(t_i) - h(t_i, X_o)$$

and finally, by defining

$$\delta z \triangleq \begin{bmatrix} \delta z(t_1) \\ \delta z(t_2) \\ \vdots \\ \delta z(t_m) \end{bmatrix}, \quad \epsilon \triangleq \begin{bmatrix} \epsilon(t_1) \\ \epsilon(t_2) \\ \vdots \\ \epsilon(t_m) \end{bmatrix}$$

$$\delta X = X - X_o$$

arrive at

$$\delta z = H\delta X + \epsilon \tag{2}$$

The solution of (2) for δX becomes the "problem." If ϵ is mean zero with covariance

$$E\left[\epsilon\epsilon^T\right] = \Lambda_\epsilon^*$$

the familiar weighted least-squares estimate of δX, $\delta\hat{X}$, is given by

$$\delta\hat{X} = \left(H^T\Lambda_\epsilon^{-1}H\right)^{-1}H^T\Lambda_\epsilon^{-1}\delta z \tag{3}$$

Substituting (2) into (3), we obtain

$$\delta\hat{X} - \delta X = \left(H^T\Lambda_\epsilon^{-1}H\right)^{-1}H^T\Lambda_\epsilon^{-1}\epsilon \tag{4}$$

The covariance of the error in $\delta\hat{X}$ is readily seen as

$$E\left[(\delta\hat{X} - \delta X)(\delta\hat{X} - \delta X)^T\right] = \left(H^T\Lambda_\epsilon^{-1}H\right)^{-1} \tag{5}$$

This brief analysis illustrates that, when the statistical structure of ϵ is known and the estimation treats this structure properly, the description of the likely estimation errors is easily constructed. This, however, is a rare circumstance.

In the treatment of earth-based, spacecraft tracking data, there are three principal causes for the departure of the behavior of ϵ from that assumed when specifying Λ_ϵ in the filter** design:

(1) In order to make the computations tractable, Λ_ϵ is usually specified as having a diagonal form; i.e., the data errors are assumed uncorrelated, or as having other similar structure. In actual practice, the data, which for our purposes are measurements of station-spacecraft range and range rate, have complex error structures arising from the effects of the ionosphere, the troposphere, and space plasma, station and spacecraft electrical path variations, oscillator instability, etc., none of which lend themselves to simple statistical modeling.

(2) Since the data are functions of the relative station-spacecraft geometry, any mismodeling of the station dynamics contributes to the ϵ of (1). Estimating the latitude, longitude, and radius of the station, i.e. including these parameters in X, can shift some of these effects away from ϵ, but we are still left with variations in the effective station location. Major contributors here are uncalibrated polar motions and undetected variations in Universal Time (the spin rate of the earth).

(3) Any phenomenon affecting the motion of the probe will change the earth-probe geometry and thus contribute indirectly to a change in the data. If these effects are not modeled, they will appear and contribute to ϵ. To be more precise, partition X as follows:

$$X \triangleq \begin{bmatrix} x(o) \\ p \end{bmatrix}$$

where x(o) is the six-dimensional state of the probe (position and velocity) at some initial epoch, and p are the additional parameters being estimated.†

Rewrite (1) as

$$z(t) = h(t;\, x(t, x(o), p);\, p) + \epsilon(t)$$

That is, the data are a function of the instantaneous probe state x(t), which in turn is influenced by the additional parameters p. Some portions of p may affect the data directly, and this is denoted by the final argument of h. To the extent that x(t) is determined by x(o) and p, we have the problem modeled correctly. But suppose there is, in addition, a random forcing function acting on the

*Small letters, either Greek or Roman, are column vectors, matrices are denoted by capital letters, superscript T denotes the transpose, and E is the expected value operator.

**Filter refers to the estimator, or estimation procedure.

†For the Mariner relativity experiment, the parameter γ would be included in p. In the estimation of X, only the value γ would be of direct interest — all the remaining parameters, including x(o), are estimated in an attempt to move data signatures from ϵ to the modeled portion $H\delta X$ of Eq. (2), where their presence does less damage.

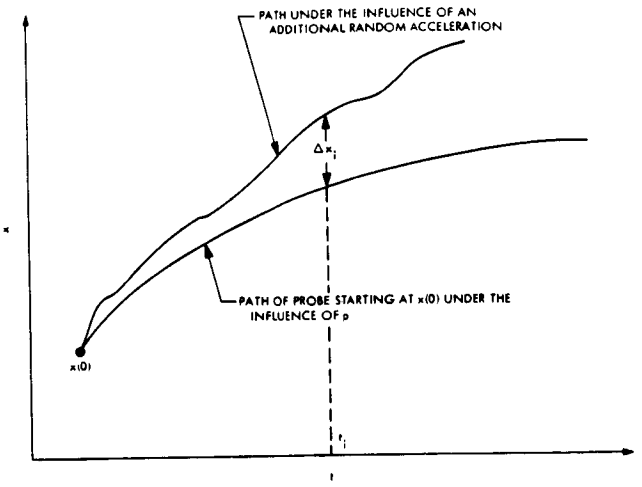


Fig. 1. Comparison of a nominal trajectory and one under the influence of a random forcing function

probe and influencing its state, as illustrated in Fig. 1.

According to this figure, at time t_i, the actual trajectory has departed from the nominal case by an amount Δx. If this random acceleration is not modeled, the contribution to ϵ will be

$$\delta\epsilon(t_i) = \left(\frac{\partial h(t_i)}{\partial x}\right)^T \Delta x_i$$

A. Relation Between ϵ and Estimation Error

We return now to Eq. (3) and view Λ_ϵ as an assumed statistical description of ϵ; its actual behavior will be modified by the three classes of errors just discussed. Defining

$$\left(H^T \Lambda_\epsilon^{-1} H\right)^{-1} H^T \Lambda_\epsilon^{-1} \triangleq F$$

F becomes the matrix operator defining the filter. Since

$$FH = I \tag{6}$$

$$\delta\hat{X} - \delta X = F\epsilon \tag{7}$$

If we are interested in the error induced by a single component of δX, say δX_i, then

$$\delta\hat{X}_i - \delta X_i = f_i^T \epsilon \tag{8}$$

where f_i^T is the i^{th} row of F. To facilitate the visualization, think of f not as an m-dimensional vector but as a scalar function of the single parameter t. (This is a rather nonrigorous transformation, but it facilitates the discussion and presentation of later figures.) Thus, if tracking data are processed on the interval [0, T], the error becomes

$$\delta\hat{X}_i - \delta X_i = \int_0^T f_i(t)\epsilon(t)dt \tag{9}$$

The structure of f_i, then, is the key in relating data errors, from whatever source, to errors in the final estimate. For example, if f(t) were a constant, a slowly varying or constant ϵ function would produce a rather large estimation error, whereas an ϵ function of complex structure or high frequency (of equivalent power) would be much less serious. The reverse would be true if f(t) had the complex character.

The structure of f in a practical estimation problem depends on the data partials (the H matrix of Eq. 2 and the Λ_ϵ assumed when constructing the filter F. In single-dimension problems, where Λ_ϵ is diagonal, f simply copies the form of the original partial ($f^T = (H^TH)^{-1}H^T$ = (scalar)H^T when H is a m × 1 column vector). In multidimensional problems, this simple form does not appear because of the additional requirement that

$$f_i^T h_j = \delta_{ij} \tag{10}$$

where h_j is the j^{th} column of H and δ_{ij} is the Kronecker delta (see Eq. 6).

By way of examples, Fig. 2 displays the original partial of the ranging data with respect to γ for the full data arc of Mariner VI. For reference, superior conjunction occurs some 268 days

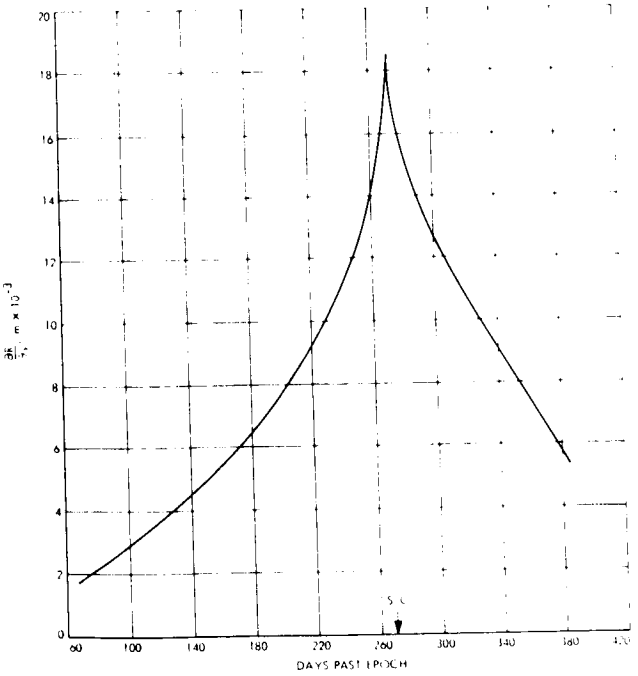


Fig. 2. Partial of range with respect to γ

past epoch, as shown in the figure. As already pointed out, if γ were the only quantity estimated, the f_γ function would replicate this signature. Figure 3 shows f_γ for the more realistic case, where spacecraft state and three orthogonal but constant non-gravitational forces are estimated as well. Note that while the basic character is still preserved, unlike the original partial, both positive and negative excursions occur and the function rings somewhat on either side of superior conjunction.

Figure 4 demonstrates the dependence on data arc and the inclusion of additional parameters. The solid line gives f_γ for the same estimation list but includes only the data within ±1-1/2 months of superior conjunction. Note that the magnitude has increased (there is a factor of 10 scale change) and that the ringing is more pronounced. More dramatically, the function reverses near superior conjunction and becomes a mirror of the original partial. Finally, the dotted f_γ is obtained when B, the major parameter of the solar corona is included in the estimate list. (B is defined in Section IV.) The high-frequency content becomes even more pronounced, but adding parameters tends to diminish the amplitudes of the excursions.

B. Specializing to Random Forces

The preceding f_γ functions could be used to determine the effects of any error source on the estimation of γ. It is thought that the qualitative structure of the functions gives an important clue as to what errors might be important. (For example, a low-level data bias would be of little consequence in any of the circumstances shown because of the nearly equal positive and negative excursions.) However, we wish to apply this point of view to the specific problem of determining the effect of random non-gravitational forces.

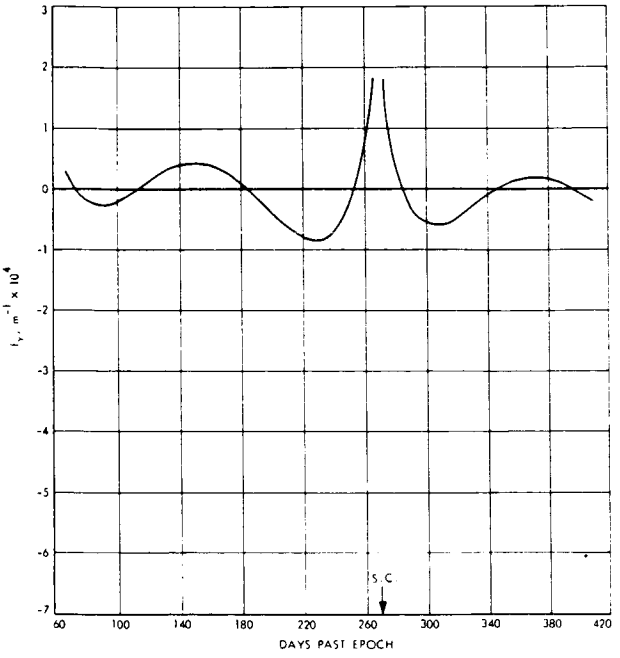


Fig. 3. Range portion of the f function of γ when state, three constant forces, and γ are estimated

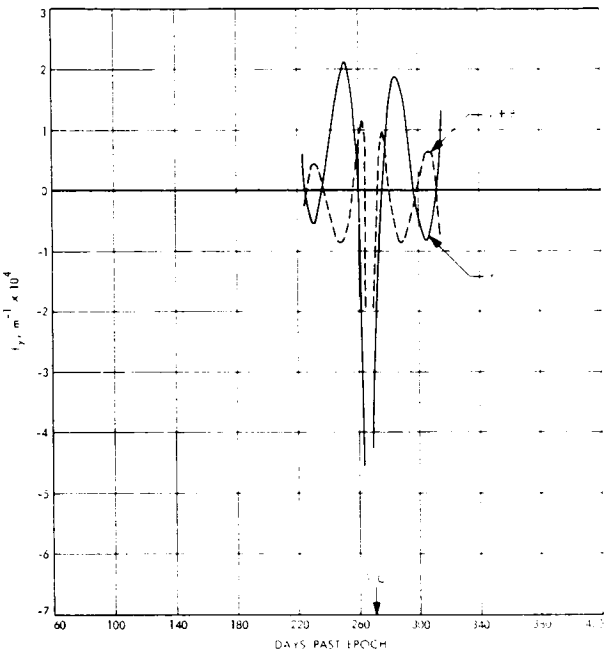


Fig. 4. Range portion of the f function of γ for the short data arc

To this end, Fig. 5 displays the range residuals created by the application of two types of forces on the probe. The solid curve results from the application of constant accelerations of magnitude 10^{-9} m/s^2 in each of three orthogonal directions. The dashed curve arises from a single realization of an exponentially correlated random process having a standard deviation of 10^{-9} m/s^2 and a correlation time of 25 days. Both curves have the same basic structure. The latter is somewhat diminished in amplitude because of the averaging which occurs when the process is random. The scale does not permit showing the additional difference, i.e., that the dashed curve is not as smooth – again due to the randomness.

Both these curves were applied to the f_γ shown in Fig. 3; the results are presented in Table 1. The difference between the effects of these phenomena is dramatic. Whereas the Case I acceleration produces almost no error in the estimation of γ, the Case II acceleration produces an error large enough to invalidate the experiment.

The reason for this disparate behavior is clear. The filter contains provisions for estimating the constant forces; hence f_γ will, by design, be orthogonal to the partials of those parameters (Eq. 10) and, by extension, orthogonal to any data residuals caused by their presence. On the other hand, no provisions are included to make f_γ orthogonal to the residuals caused by the random process; as a result, a significant response occurs.

This last remark implies that we could make f_γ orthogonal to Case II if we wished, and indeed we could do so. But Case II is only a single realization; it is, of course, not possible to construct a fixed non-zero function which is orthogonal to every member of a random process. Recourse must be taken to the concepts of minimizing the mean-squared error in these circumstances, concepts that are beyond the domain of parameter

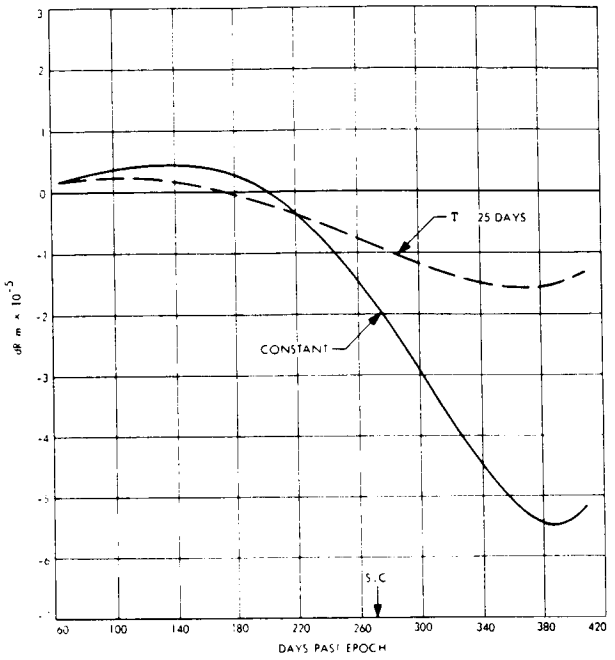


Fig. 5. Range residuals due to accelerations of 10^{-9} m/s^2

estimation. In order to lessen the response to random processes, one must resort to the principles of sequential or Kalman filtering. This will be discussed in the concluding section.

III. A Measure of the Uncertainty in γ^* Produced by Random Accelerations

The preceding example clearly demonstrates the large impact seemingly small random accelerations of the spacecraft can have on the orbit determination solution in general. Of particular concern here is to obtain some idea of how random accelerations, of the type likely to be experienced by Mariners VI and VII, degrade the solutions for γ^*. A study which will yield this type of information divides rather naturally into the following three areas:

(1) An investigation to determine what type of random accelerations (both in magnitude and character) may be expected to influence the Mariner-class spacecraft.

(2) Obtaining a measure of how the solution for γ^* is degraded by using a batch filter, which is ignorant of random spacecraft accelerations, in the presence of these same accelerations.

(3) Obtaining some idea of the degree of superiority that sequential filter solutions for γ^* can be expected to exhibit over the batch filter solutions.

A. Possible Random Accelerations

The random accelerations experienced by Mariner-class spacecraft arise primarily from three sources: (1) leakage and imbalance of the attitude control jets, (2) variations in the solar radiation pressure, and (3) variations in the solar wind.

Table 1. Illustration of filter behavior on estimates of γ^*

Case	Acceleration	Perturbation in γ^*
I	Constant 10^{-9}-m/s^2	$<10^{-4}$%
II	Random 10^{-9}-m/s^2 with a correlation time of τ = 25 days	18%

The time history of the accelerations of various Mariner spacecraft produced by these sources may be obtained by examining data collected by these spacecraft. Table 2 contains the standard deviations and the associated exponential correlation times describing such accelerations when they are modeled as exponentially correlated phenomena.

The parameters characterizing the random accelerations due to attitude control jets were computed from an autocorrelation analysis made of the accelerations observed along the Mariner V spacecraft's pitch and yaw axes. (Ref. 1). The acceleration time history was obtained by examining telemetered limit cycle data (Ref. 2). The examination of such data is quite involved and, as yet, has not been carried out for Mariners VI or VII. However, a cursory examination of the data has shown that, like Mariner V, Mariners VI and VII are both generally clean spacecraft. Thus, the values quoted for Mariner V (the first and second numbers, respectively, in Table 2), are probably representative of the accelerations experienced by Mariners VI and VII.

The random accelerations due to the solar radiation pressure are produced by 0.1 to 0.2% variations in the solar constant (Ref. 3). These accelerations will occur primarily in the sun-spacecraft direction.

The first value given in Table 2 describing the magnitude of the acceleration due to the solar wind was calculated using a model describing a quiet sun (Ref. 4). Examination of solar wind data taken by Mariner II (1962) over a 4-month period indicates a maximum inelastic collision acceleration of 0.15×10^{-9} m/s^2, the second value in Table 2 (Ref. 5). Although these data were taken during a relatively low period of the solar cycle and the Mariner VI and VII experiment took place during a relatively high period, this is not expected to substantially affect the magnitude of the maximum acceleration but only the frequency of larger accelerations. Hence, the parameters given in Table 2 describing the solar-wind-induced accelerations are probably representative of those experienced by Mariner VII. Once more the acceleration will take place primarily in the sun-spacecraft direction.

Table 2. Description of random accelerations experienced by various Mariner-class spacecraft

Cause of random acceleration	Source of data	Standard deviation, 10^{-9} m/s^2	Correlation time, days
Attitude control jets	Mariner V (1967)	0.2 and 0.6	22 and 45
Fluctuations in			
Solar pressure	Mariner VI, VII	0.1 (at Mars distance)	< 1 day
Solar wind	Mariner II (1962)	0.03 and 0.15 (at Mars distance)	< 1 day

B. Degradation of the Solution for γ^* in the Presence of Random Forces

To obtain some idea of the effect random accelerations can have on solutions for γ^*, such solutions made in the presence of a particular random acceleration environment will be examined. This environment will be composed of random accelerations that are constant for a period of T days and whose magnitude is selected from a normal distribution with variance σ^2 (ran), and which is exponentially correlated from period to period with a time constant τ. A judicious choice of the parameters of this environment allows it to be a good representation of the attitude control, the solar pressure, and, to a lesser degree, the solar wind accelerations mentioned in the previous section.

In a classical least-squares solution of an estimation problem, a measure of the uncertainty in a particular parameter is given by its variance. This quantity, which will be called the computed variance, defined in Eq. (5), is a function only of the data noise, the a priori values of the "solve for" parameter set, and the partial derivatives of the data with respect to these parameters. The computed variance of γ^* for Mariner VI data arcs, which start at the initial epoch of August 5, 1969, and are extended for 386 days, is shown in Fig. 6 in terms of its square root, the standard deviation. Unfortunately, this is a much too optimistic picture because, as mentioned previously, the computed variance does not take into account the effect of unmodeled phenomena such as random accelerations which can be expected to disturb the data and/or the state of the spacecraft.

A much more reliable measure of the uncertainty in γ^* is the consider variance. This quantity is composed of the computed variance and a generally much larger term that reflects the effects error sources (which must be described in some statistical sense) have on the solution when the employed solution filter is ignorant of these error sources. For the types of data arcs discussed in the preceding paragraph, the consider variance, which results from a batch filter (estimating state, constant accelerations, and γ^*) operating in a random acceleration environment of

$$\sigma^{(ran)}_{R,x,y} = 10^{-9}\ m/s^2$$

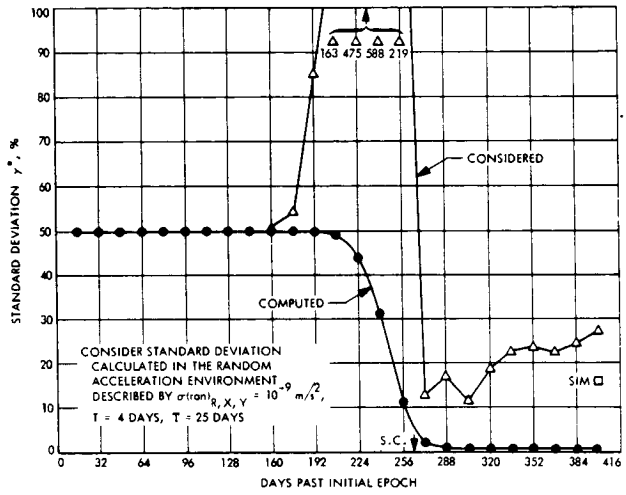


Fig. 6. Computed and consider standard deviations for data arcs starting at the initial epoch

(where R, x, y represent the spacecraft roll, pitch, and yaw axes), τ = 25 days, and T = 4 days, is also shown in Fig. 6. The point denoted by SIM in the figure gives the error in γ^* which would result if the 384 days' worth of the range residuals, shown in Fig. 5, and the associated doppler residuals were operated on by a batch filter.

The consider variance curve presented in Fig. 6 is very instructive and contains the following gross features:

(1) The consider variance for γ^* does not substantially change from its a priori value until more than 160 days' worth of data are included in the solution.

(2) As the data arc is extended toward superior conjunction, the consider variance, instead of diminishing as does the computed variance, actually exhibits a rapid increase and finally reaches a maximum of about 600%.

(3) As the data arc passes through the superior conjunction period, the consider variance is rapidly reduced to a minimum

(which is still much larger than the corresponding computed variance).

(4) After this minimum is reached, the inclusion of more data in the solution causes a generally continuous rise in the consider variance.

The explanation of these gross features is quite simple:

(1) The partial derivatives of the data with respect to γ^* during the first 160 days are so small that they contain very little information which can be used to increase the a priori knowledge regarding γ^*. However, during this time period, the partials of the data with respect to the remaining nine parameters of the solution are not small, and the batch filter assumes that after 160 days it has a very good knowledge of these parameters.

(2) As the data arc is extended toward superior conjunction, the partials with respect to γ^* continue to increase in magnitude until their information content far exceeds the a priori information. In this situation, the filter will change γ^* to try to reduce the residuals which have been produced by the random accelerations. The filter attempts to perform this reduction almost entirely with γ^* because, as mentioned above, it assumes that the data used during the earlier portion of the data arc have yielded a very good knowledge of the remaining parameters in the solution set.

(3) As the data arc passes through superior conjunction, the data partials with respect to γ^* exhibit a rather distinctive shape, as shown in Fig. 2, and this drastically increases the information content. Since it is unlikely that the data residuals produced by the random accelerations will resemble the distinct shape of the partials, the filter will probably not be able to reduce the residuals by making a large and spurious change in γ^*. Consequently, the consider variance reduces as shown.

(4) As additional data are included in the solution, the distinctive nature of the partials around superior conjunction becomes relatively less important, and the filter may reduce the random acceleration residuals by making larger changes to γ^*.

As just mentioned, the large errors in the solution for γ^* in the preceding example are the results of the filter's attempt to absorb the random acceleration effects in γ^*. It does this because, by the time the γ^* partial becomes large enough to start supplying information to the solution, the filter thinks it has the remaining parameters very well determined. Clearly, a partial remedy to this problem would involve forming the solution so as to diminish the extent to which the filter assumes it knows the other parameters, so that they may also be used to help reduce the random acceleration residuals. One way to do this is to design a filter that will operate in a sequential mode, such as discussed in the next section. Another possible way of improving the solution is still to use a batch filter but to start the data arc at a later time, so that the filter has not firmly established the values of all the parameters except γ^* before the superior conjunction data are reached.

The results of performing a consider variance analysis for data arcs starting at 1-1/2 and 2-1/2 months before superior conjunction and progressively extended to 1-1/2 and 2-1/2 months after superior conjunction are given in Fig. 7. These results clearly demonstrate the superiority of the short arc versus the long arc for solutions performed in the presence of the type of random accelerations under consideration.

Unfortunately, the above consider variance analysis, even for the ±1-1/2-month data arc, presents too optimistic a measure of the error in γ^* if it is necessary to include the solar corona parameter B in the solution. This is clearly demonstrated in Fig. 8, which shows the results of performing a consider variance analysis of a solution containing both B and γ^* in the solution parameter set and covering the data arcs that start at 1-1/2 months before superior conjunction and may continue for up to 3 months. The "γ^* only" curve in Fig. 8 has been taken from Fig. 7 and is included for comparison. The degradation of the solution for γ^* produced by including B in the solution parameter set arises because the data partials for γ^* and B exhibit somewhat similar behavior around superior conjunction. This similar behavior diminishes the distinctive nature of the γ^* partial derivatives around conjunction, and hence degrades the solution for γ^* in the presence of random accelerations.

Also included in Fig. 7 is a consider variance analysis of these same data arcs for a batch filter operating in the random acceleration environment described by $\sigma_{R,x,y} = 10^{-9}$ m/s^2, T = 1 day, and τ = 2 days. This type of random acceleration environment is more representative

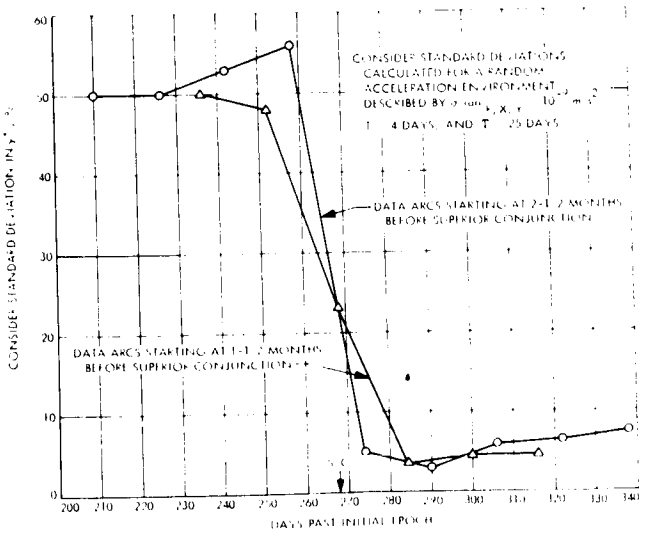


Fig. 7. Consider standard deviations for data arcs starting at 1-1/2 and 2-1/2 months before superior conjunction

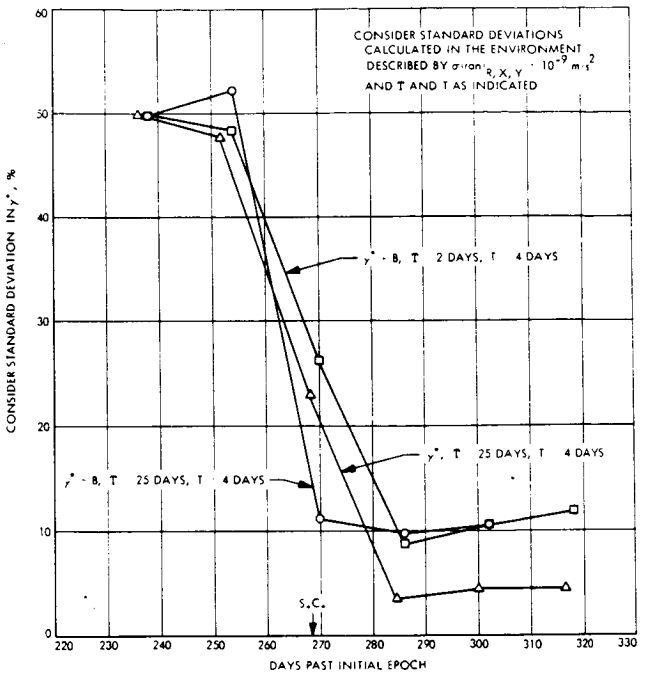


Fig. 8. Consider standard deviations with and without B in the solution

of the solar pressure and solar wind random accelerations, described in Table 2, than the random acceleration environment considered until now.

From an examination of the two "γ + B" curves in Fig. 9, it does not appear that the errors in the solution of γ^* are particularly sensitive to the time constant of the random acceleration environment.

Before attempting to use Figs. 6, 7, and 8 to estimate the errors which random accelerations can be expected to introduce in the Mariner VI and VII solutions for γ^*, one should scale these errors by the values given in Table 2.

IV. Preliminary Sequential Estimation Results

Thus far, we have established that, because of substantial random processes, γ^* cannot be accurately estimated from the long arc of data (410 days from Mariner VI) with the familiar batch filter. Although the shorter arc of data near superior conjunction does give reasonable results with the batch filter, it is desirable to develop a sequential filter which can perhaps improve the ability to solve for γ^* by processing the full data arc. The following paragraphs present some preliminary results from one such sequential filter.

The particular filter described here does not pretend to be optimal with respect to the likely models of random forces given in the preceding section. It is an early experimental design. Work is in progress to produce a more nearly optimal filter, but in the meantime, we shall content ourselves with results using this rather primitive sequential filter. There is one advantage in that this filter can be applied to records of random forces drawn from the formal statistical models we have suggested. This approach gives an important clue as to how sequential filters respond to random processes that are modeled somewhat incorrectly by the filter. In a sense, we are pursuing the same type of analysis as in the preceding section: propose a filter, and then determine its performance when the actual environment is other than that assumed during the filter design.

The term sequential filter, here, refers to a "batch sequential," weighted, least-squares filter. That is, the data are processed in blocks or batches — sequentially in time. The size of the data blocks is selected by the analyst. For each block of data processed, the filter employs an a priori estimate of the parameters at the initial epoch and the associated covariances which reflect the information obtained from all the preceding data. This sequential filter departs from the typical batch filter, however, by allowing each a priori covariance to account for uncertainties in the filter's model of the physical world. That is, for each estimate obtained, a covariance matrix Λ_c is computed. This covariance matrix represents the uncertainties in the estimate due to errors in the data and errors in a priori values of the parameters. The sequential filter then computes the a priori covariance for the next batch to be

$$\tilde{\Lambda} = \Lambda_c + Q$$

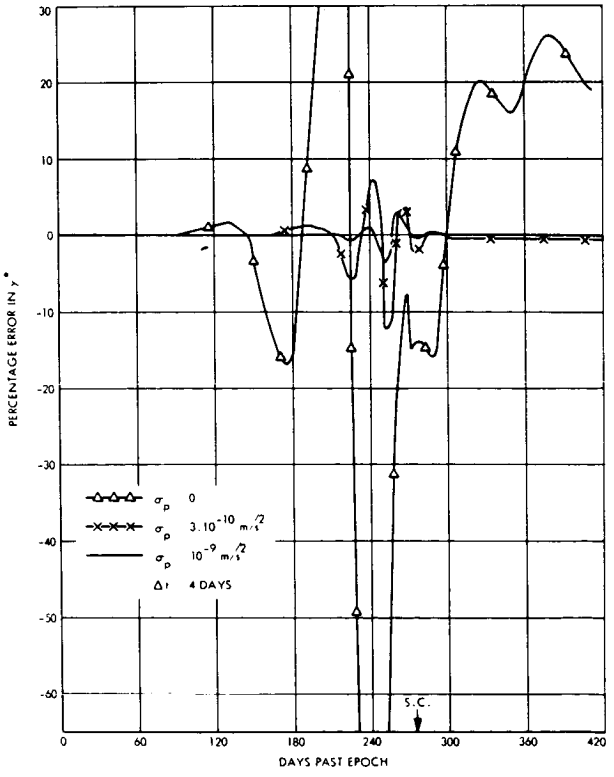


Fig. 9. Sequential filter results — percentage error in γ^* for different assumed levels of random accelerations

where Q represents the uncertainties in the latest estimate of the parameters due to the effects of improperly modeled accelerations on the last block of data processed. The matrix Q prevents the filter from "knowing the wrong answer too well." That is, the a priori information on the trajectory and the estimated accelerations remains at a level at which new data can be used by the filter to re-evaluate these parameters.

In order to compute Q for each batch, the filter assumes that piecewise constant accelerations affect the spacecraft. These accelerations are in three orthogonal directions and have the following statistical properties. Let $\underline{p}_{i-1}$ = accelerations experienced during the i^{th} block of data.

$$E\left[\underline{p}_{i-1}\right] = 0$$

$$E\left[\underline{p}_{i-1}\underline{p}_{i-1}^{T}\right] = \begin{pmatrix} \sigma_p^2 & & 0 \\ & \sigma_p^2 & \\ 0 & & \sigma_p^2 \end{pmatrix} \quad \text{for } i = 1, 2, \ldots$$

Each matrix Q is a function of σ_p and the partials $V = \partial \underline{X}_i / \partial \underline{p}_{i-1}$, where $\underline{X}_i$ = spacecraft state at the beginning of the (i+1) data block. The standard deviation of the random forces σ_p is an input parameter to the program. For the case $\sigma_p = 0$, the matrix Q is zero, and the filter performs like the familiar batch processor.

Currently, the filter processes range and doppler data to obtain sequential estimates of the following twelve parameters:

$\underline{X}_o$ = six state parameters at the initial epoch

γ^* = relativity parameter

B, ϵ = solar corona parameters*

$\underline{w}_o$ = three orthogonal forces, constant over the entire trajectory.

Several cases have been processed with this sequential filter using all or part of the long arc of data from Mariner VI, where the data residuals have been generated to reflect only random accelerations on the spacecraft. These random accelerations have a correlation time of 25 days and a magnitude of 10^{-9} m/s^2. That is, for the bulk of the studies, we employed the same single realization of the process that was discussed in Section II.

For each of the cases studied, the initial standard deviations of the estimated parameters were the following:

σ_x = 10^7 m in each position component at epoch

$\sigma_{\dot{x}}$ = 3.3×10^3 m/s in each velocity component at epoch

σ_{γ^*} = 50%

σ_B = 0.7×10^6 cm^{-3}

σ_ϵ = 0.2

σ_{w_o} = 10^{-8} m/s^2 in each direction

All cross correlations were zero, except $\sigma_{B_\epsilon} = 0.99999$.

Table 3 shows the final percentage error in the estimate of γ^* and the computed standard deviation (σ_{γ^*}) for the cases where $\sigma_p = 0$, 3×10^{-10}, and 10^{-9} m/s^2. The Δt indicates the length (in days) of the sequential blocks of data.

From Table 3, we can conclude that, of the cases studied, the sequential filter with $\sigma_p = 10^{-9}$ m/s^2 and $\Delta t = 4$ days yields the smallest percentage error in γ^* (-0.008%), with a standard deviation of 6.9%. However, for $\sigma_p = 3 \times 10^{-10}$ m/s^2 and $\Delta t = 10$ days, the error is only -0.17%, with $\sigma_{\gamma^*} = 3\%$. These results are significant when they are compared to the 19% error in γ^* for $\sigma_p = 0$, i.e., the batch filter result.

The sequential results are shown in Figs. 9, 10, and 11. Figure 9 gives the percentage error in γ^* for the cases in which the data were processed in blocks of 4 days and $\sigma_p = 0$, 3×10^{-10}, and 10^{-9} m/s^2. Figure 10 gives the computed standard deviations of these estimates as a function of time from the initial epoch. In Fig. 11, the percentage error in γ^* is plotted for the short arc of data (226-318 days from epoch), with $\sigma_p = 0$, and for the long arc, with $\sigma_p = 10^{-9}$ m/s^2. (This latter curve is repeated from Fig. 9.) The short-arc results are representative of the batch filter errors. For this particular model of random accelerations and choice of σ_p, the long-arc sequential solution of γ^* is preferable to the short-arc batch solution.

V. Conclusions

We have shown that random, non-gravitational forces of the magnitude and character believed to be affecting the Mariner VI and VII spacecraft have a significant effect on the navigation data

*These solar corona parameters are defined by the following electron density profile, determined by Blackwell, Dewhirst, and Ingham (Ref. 6):

$$N_e(r) = \frac{A}{r^6} + \frac{B}{r^{2+\epsilon}} \quad \text{for } r \geq 2.5$$

where A and B are electron densities, r is the distance from the sun measured in solar radii, and $\epsilon \cong 0.3$.

Table 3. Percentage error in estimate of γ^* for long arc of data calculated for the random process described by $\sigma_{R,x,y} = 10^{-9}\,m/s^2$, $\tau = 25$ days

Δt, days	σ_p, m/s^2		
	0	3×10^{-10}	10^{-9}
4	$\Delta\gamma^* = 19\%$	-0.55	-0.008
	($\sigma_{\gamma^*} = 0.45\%$)[a]	(3.5)	(6.9)
10	$\Delta\gamma^* = 19\%$	-0.17	0.1
	($\sigma_{\gamma^*} = 0.45\%$)	(3.0)	(4.9)
25	$\Delta\gamma^* = 19\%$	2.8	-0.45
	($\sigma_{\gamma^*} = 0.45\%$)	(2.5)	(3.4)

[a]Numbers in parentheses denote the filter's own computation of its performance at the end of the data arc.

from these spacecraft, and become a limiting error source for the relativity experiment.

The effect of the random forces on the classical least-squares filter, which assumes they do

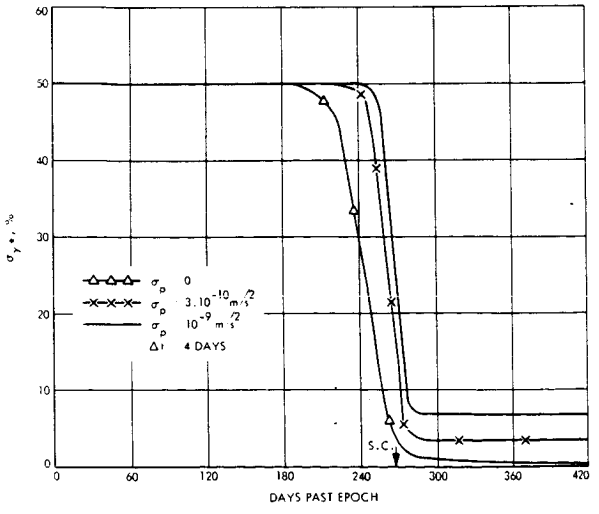


Fig. 10. Sequential filter results – standard deviation of γ^* for different assumed levels of random accelerations

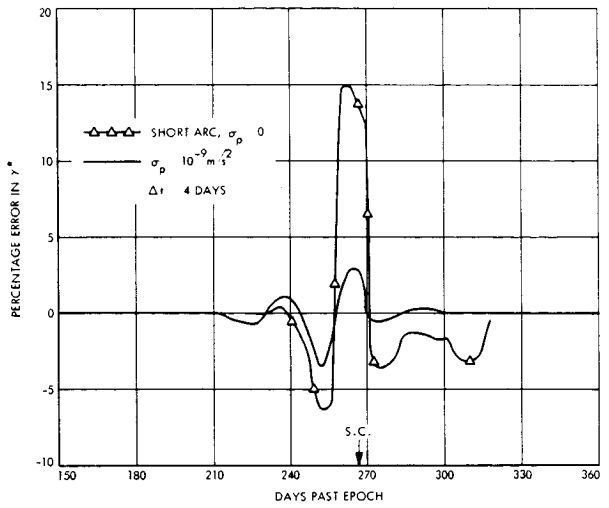


Fig. 11. Percentage error in γ^* for short-arc batch and long-arc sequential

not exist, produces large errors in the estimate of γ^* from the long arc of data. Although shorter arcs of data near superior conjunction can yield reasonable estimates of γ^* with this same filter, it is desirable to develop an estimator which more faithfully represents the actual physical environment of the spacecraft. We have suggested that a judiciously designed sequential filter can be of significant benefit; and we believe the results presented promise that the employment of the sequential filter can permit use of the entire data arc, producing results superior to any strategy employed with the classical least-squares filter.

References

1. Bourke, R. D., McReynolds, S. R., and Thuleen, K. L., J. Spacecraft Rockets, 6, No. 9, 1063-1066, Sept. 1969.

2. Mohan, S. N., Modeling of Translational Forces Due to Gas Leakage in the Mariner V Attitude Control System, Jet Propulsion Laboratory Internal Technical Memorandum 391-113, Sept. 1970.

3. Plamondon, J. A., Space Programs Summary 37-59, Vol. III, 162-168, Jet Propulsion Laboratory, Pasadena, Calif., Oct. 31, 1969.

4. Hundhausen, A. J., Rev. Geophys. Space Phys., B, No. 4, 779, Nov. 1970.

5. Neugebauer, M., and Snyder, C. W., J. Geophys., Rev. 71, 4469, 1966.

6. Blackwell, D. E., Dewhirst, D. W., and Ingham, M. F., Advances in Astronomy and Astrophysics, 5, p. 25, Academic Press, New York, 1967.

Radio Tracking With the Deep Space Network

Warren L. Martin
Jet Propulsion Laboratory
California Institute of Technology

I. Introduction

For many years precision doppler existed as the major tool available to those working in the field of celestial mechanics. Until recently, ranging never achieved widespread popularity despite the fact that it is potentially more powerful in certain situations. The first application of ranging to a major project came in 1966 with the Lunar Orbiter series. The following year, a new planetary ranging system capable of operating with deep space probes was used on Mariner V at Venusian encounter. This same equipment provided valuable data on Mariners VI and VII at Martian encounter in 1969. Recognition of ranging as an important tool came in August of 1969 when a battery failure on Mariner VII caused the spacecraft to veer slightly from its intended course. Range and doppler data collected following the event provided the information needed by controllers to carry out the experiment objectives. The Mariner VI and VII extended mission relativity experiment marked the first time that range information became more important than doppler data in establishing orbital parameters. During a few days around solar superior conjunction, ranging was the only usable data type. Thus, ranging has emerged as an important tool and a large portion of this paper will be devoted to a discussion of its capabilities. The current doppler extraction system will also be considered together with the rather unique S-X band radio tracking system planned for Mariner Venus-Mercury 1973.

II. Range and Doppler Measurements

Before considering the performance capabilities of the present radio tracking systems it would be worthwhile to distinguish between the measurement of range and doppler. Range is obtained by counting the cycles of a precise frequency standard which occur between the transmission and reception of an encoded signal. Thus, range in this context is a measurement of time rather than of distance, and any conversion to the latter must involve assumptions regarding the propagation of radio waves.

The topocentric range rate is measured by a frequency shift on the RF carrier produced by the doppler effect. Mechanization of the doppler extractor dictates that the sample be taken over some non-zero interval. Therefore, the measurement is equivalent to a range change which occurred while the sample was being taken. Obviously any attempt to convert the data to a velocity (meters/second) encounters difficulties identical to the ranging situation. Therefore, doppler is collected as a cumulative cycle count and is treated solely as a variation in range.

In the following discussion of the Deep Space Network's (DSN) range and doppler tracking system the Mars site (DSS 14) at the Goldstone Tracking Station in California will be quoted as typical. This can be justified by the prototype nature of DSS 14 and the fact that identical 210-ft

(approximately 64 m) antennas are under construction or are planned for the major overseas DSN sites.

III. Doppler System

Figure 1 is a simplified block diagram of the DSIF doppler tracking system. The most important elements are the frequency standard and its distribution system since they ultimately limit the measurement's accuracy. Following frequency multiplication and amplification, the carrier is transmitted to the spacecraft where a phase-locked receiver/transmitter system filters the signal and returns it to the earth. Doppler is measured by comparing the received frequency with that currently being transmitted. The doppler extractor's output v_e is described by (Ref. 1)

$$v_e = \cos\left[\frac{2\pi bf}{4}\int_{t_0}^{t}\left(\frac{2v}{c}+\frac{v^2}{c^2}+\alpha\right)dt\right] \qquad (1)$$

where

b = transponder multiplication ratio (240/221)

f = transmitted S-band carrier frequency

v = spacecraft topocentric range rate

c = velocity of light in a vacuum

α = columnar electron dynamics

A value can be computed for α using the relationship

$$\alpha = \frac{B}{cf^2}\left(\dot{I}_{up}+\frac{1}{b^2}\dot{I}_{dn}\right) \qquad (2)$$

where

B = 40.3 in mks units

$\dot{I}_{up}$ = uplink columnar charged-particle dynamics (electrons/m^2-s)

$\dot{I}_{dn}$ = downlink columnar charged-particle dynamics

Since the frequency standard and distribution system serve as a common reference to both transmitter and receiver, any drift will affect the doppler data. When discussing errors in the standard it is important to indicate the measurement interval. Typically, three separate times are important. First, the short-term stability ($t < 1$ s) specifies the amount of high-frequency noise contributed by the reference. Second, the medium-term stability (1 s $< t <$ 10,000 s) describes the error accumulated in one round trip. Third, the long-term stability ($t >$ 10,000 s) provides information about how well day-to-day tracking data can be fitted over a long arc.

Figure 2 shows the stability characteristics for typical rubidium vapor and hydrogen maser oscillators. Note that the frequency deviation becomes constant for all sample times in excess of 100 s. The decrease in stability at shorter time intervals is indicative of the high frequency noise characteristics of the standard.

Figure 2 also demonstrates the substantial improvement which can be realized by employing a hydrogen maser as the timing source. For averaging periods in excess of 1 min the variations should be only a few parts in 10^{14}. Observe also that the high-frequency noise characteristics are substantially better than those of the rubidium oscillator. The advantages of a hydrogen maser are most apparent when considering very deep space missions where integration over long round-trip times makes the rubidium maser unsuitable.

The discussion thus far has disregarded any degradation due to frequency synthesis or distribution, and clearly errors from these sources are equally important to those present in the frequency standard. Present equipment has stabilities about equivalent to the rubidium oscillator (5 parts in 10^{-12}), and the distribution system planned for the hydrogen maser will have characteristics compatible with that instrument.

A pair of hydrogen maser standards have been installed at the Goldstone Mars site (210-ft antenna) and will become available on an experimental basis for Mariner Mars 1971. Assuming satisfactory operations, these masers will serve as prototypes for similar installations at other DSN stations. Aside from the potential improvement in range and doppler data for deep space missions, the new frequency references

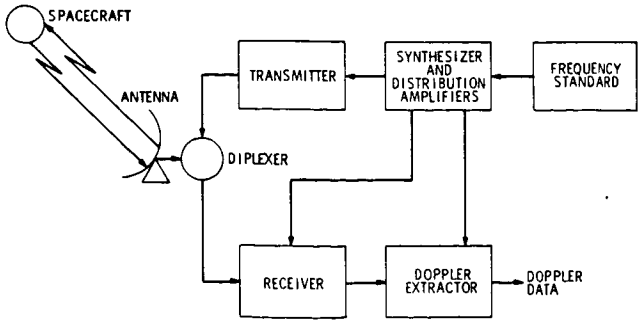


Fig. 1. Doppler data system

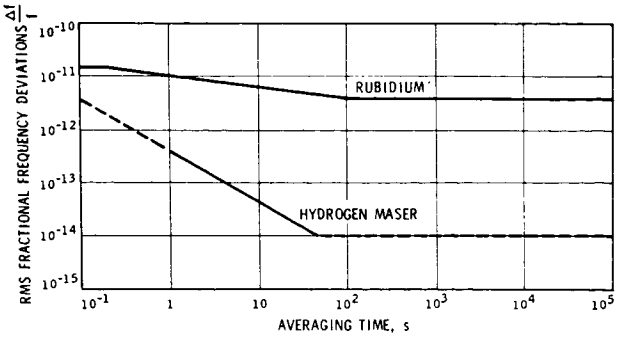


Fig. 2. Typical frequency standard stability

are the first step in providing a very long baseline interferometric capability within the DSN.

IV. Ranging

Recent experimental activity in the areas of relativity and charged-particle dynamics has resulted in an increased interest in ranging as a data type. To better understand its capabilities and limitations, a brief description of system mechanization will be presented.

Two different spacecraft ranging machines presently exist within the DSN. These are the planetary system used for Mariner Venus 1967 and Mariner Mars 1969, and the sequential system used for the Mariner Mars 1969 extended mission relativity experiment. Present plans include using both systems to support Mariner Mars 1971, and the feasibility of locating one of them at an overseas DSN site, 85-ft antenna is currently being explored. Conceptually, the two equipments are similar. The basic differences lie in the coding and in the method used to generate a replica of the signal returned from the spacecraft. Due to the space limitations and availability of data, the specific performance data contained herein pertain only to the sequential ranging equipment.

Figure 3 is a simplified block diagram of the binary coded sequential acquisition ranging system (Ref. 2). A frequency standard-synthesizer generates f_s (nominally 22 MHz), which is multiplied by 3 and phase-modulated by the transmitter coder. The code is generated by dividing the $\times 3$ multiplier's output by 64 and applying the result to an 18-stage binary counter. Each of the 18 binary counter outputs is individually selectable for modulating the transmitter. The period T_n of the n^{th} squarewave component is given by

$$t_n = \frac{64 \times 2^n}{3f_s} \qquad (3)$$

From the above relationship it can be seen that the code's period is irrevocably tied to the transmitter frequency. Changing the transmitter frequency, as is done from time to time to assure optimum reception at the spacecraft, also changes the coder's frequency in direct proportion.

A virtually identical set of hardware exists in the receiver; however, the $\div 64$ stage is preceded by a frequency adder circuit. The adder accepts as one of its inputs 66 MHz from the $\times 3$ multiplier in the transmitter chain. A second input connects RF doppler f_d, properly scaled, from the DSIF receiver. In the ranging mode, the output from the frequency adder circuit is the algebraic sum of the two inputs:

$$f_{out} = 3f_s + f_d \qquad (4)$$

The receiver coder is a duplicate of its counterpart in the transmitter except for the inclusion of a second output providing code delayed by $\pi/2$. This second channel is combined with the first to establish the amplitude of the returning signal — a necessary step in measuring its phase.

When the range-sync switch is in the sync position, the two coders will be operating synchronously. Because of the topocentric range rate of the spacecraft, the received code slips with respect to the transmitted code. If at time t_0 the switch is changed to the range position, the receiver coder's frequency is modified by the RF doppler and becomes coherent with the signal being received from the spacecraft. Assuming that the two coders were synchronized prior to the changeover, the phase difference between the receiver coder and the incoming signal is a measure of range. This phase difference will remain essentially constant by virtue of the doppler rate aiding, allowing the range measurement to be made at leisure.

Thus, a coherent model of the received range code can be generated by modifying the frequency of the transmitted code by the spacecraft's doppler. Besides a drastic simplification in the hardware design, this method of rate aiding provides a number of subsidiary benefits, specifically, greater reliability, lower cost, and a means for calibrating the columnar charged-particle dynamics in the

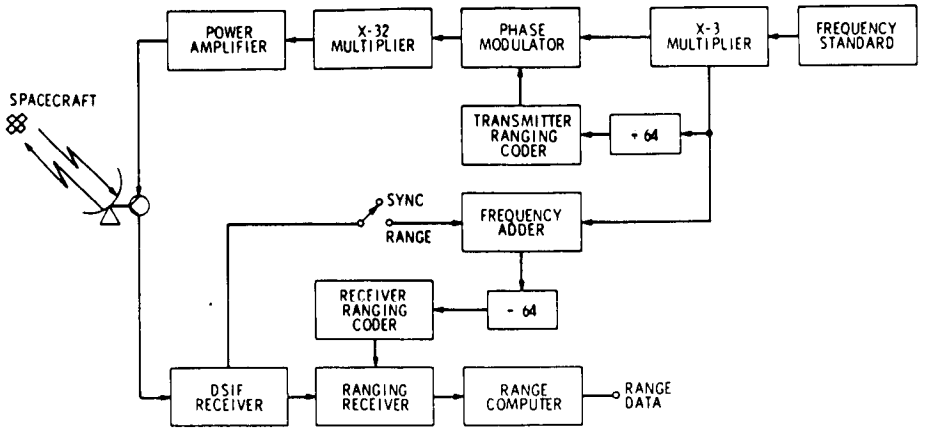


Fig. 3. Binary coded sequential acquisition ranging system

ray path. It has been shown (Ref. 1) that the received ranging signal is of the form

$$v_r = \cos\left[\omega_i t + \phi'_m \cos(\gamma - \theta_0)\right] \quad (5)$$

where

ω_i = $2\pi \cdot 10^7$ (receiver IF carrier frequency)

ϕ'_m = spacecraft modulation index

cos ( ) = squarewave approximation of cos ( ); +1 if cos ( ) > 0; -1 if cos ( ) < 0

θ_0 = phase shift in modulation due to spacecraft range

γ = perturbation of the range code due to spacecraft velocity and particle activity

In the above expression

$$\dot{\gamma} = \omega_m \int_{t_0}^{t} \left[\left(1 + \frac{v}{c}\right)^2 - \alpha\right] dt \quad (6)$$

where $\omega_m = 2\pi f_m$ (angular range code frequency) and α is defined in Eq. (2).

A comparison of Eqs. (1) and (6) illustrates the effect of phase and group propagation velocities upon the respective signals. This difference forms the basis of the differenced range versus integrated doppler (DRVID) calibration technique (Ref. 1).

V. Charged-Particle Calibration

In a dynamic charged-particle environment, the group and phase velocities of a radio wave are not constant. Consider a situation where charged particles are entering the radio path to a receding spacecraft. The group velocity would then be decreasing while the phase velocity would be increasing. Since the range code is propagated at the group velocity, the range data would make the spacecraft appear to have moved further than its actual physical displacement. If doppler cycles were added over the same interval, the spacecraft's apparent displacement would be smaller than the true distance. Thus, a dynamic phase error results when an RF doppler synthesized range code model is compared with an actual received range code. In the supposed situation of an increasing columnar content, the modeled code will lag further and further behind the received range code. The resulting "drift" in the calculated range is actually a direct measure of the DRVID function, which is

$$\text{DRVID} = \int_{t_0}^{t} \frac{B}{f^2}\left(\dot{I}_{up} + \frac{1}{b^2}\dot{I}_{dn}\right) dt \quad (7)$$

where

B = a constant (40.3 in mks units)

f = transmitted carrier frequency, $96 f_s$

$\dot{I}_{up}$ = uplink columnar charged-particle dynamics (electrons/m^2-s)

b = transponder ratio, 240/221

$\dot{I}_{dn}$ = downlink columnar charged-particle dynamics

Comparing the DRVID function with the output from the ranging equipment, which is

$$\Delta\tau(t) = \int_{t_0}^{t} \frac{2B}{cf^2}\left(\dot{I}_{up} + \frac{1}{b^2}\dot{I}_{dn}\right) dt \quad (8)$$

results in

$$\text{DRVID} = \frac{c}{2}\Delta\gamma(t) \quad (9)$$

Thus, the apparent drift in the range is a direct measure of the columnar charged-particle dynamics.

Two charged-particle calibrations were conducted on November 24 and December 11, 1969, with the Mariner VI spacecraft and the results are plotted in Fig. 4a and b, respectively. Also shown in Fig. 4 are the range changes attributable to the earth's ionosphere as determined by VHF Faraday rotation monitoring (Ref. 3).

Note the excellent agreement (within $1/2\sigma$) of the DRVID measurement with the ionospheric data. Despite the spacecraft's great distance (approximately 2 AU) virtually all of the charged-particle activity is indicated to be within the

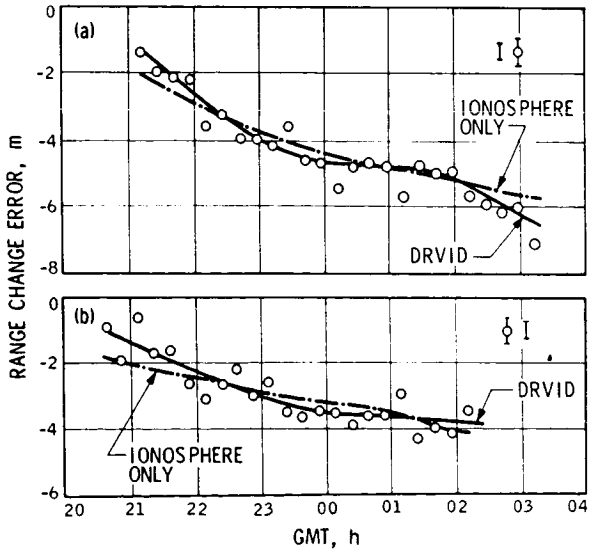


Fig. 4. Differenced range versus integrated doppler

ionosphere. The figure also demonstrates that the ranging equipment is capable of measuring the charged-particle activity, although the effect may be very small, amounting to only a few meters over several hours.

The ionospheric calibration was conducted when the sun-earth-probe angle was relatively large (60 deg). Additional data were collected when the signal ray path passed closer to the sun. The results appear in Ref. 4.

VI. Experimental Results

During the past year the tracking equipment has supported the Mariner Mars 1969 extended mission relativity experiment. Interpretive results obtained from the information which has been gathered will be found in Ref. 5. This discussion is concerned only with the data type and its quality.

Both range and doppler information have been used to generate a reference orbit. Range (or time delay) data gathered near solar superior conjunction are compared with the expected round-trip time, and the disparity is used to compute gamma. Obviously the experimental solutions will be directly affected by the tracking data quality.

To place the findings in their proper perspective it is necessary to examine the ranging signal's characteristics upon return to the earth tracking station. Figure 5 shows the received signal strength for several months around superior conjunction. For 2 months prior to superior conjunction the received ranging power was below -200 dBm (10^{-23} W). This low value resulted from the probe's distance, which was nearly 2.5 AU., and a battery problem aboard the spacecraft which made it impossible to use the high-gain antenna. Toward the end of April a new high-power transmitter (200-kW) became available resulting in a 10-dB (10 ×) increase in the received signal's power. The remainder of the mission has been conducted using this new transmitter.

Figure 6 is a plot of the range - doppler residuals over a 1-year period as computed by the JPL Double Precision Orbit Determination Program (DPODP, Ref. 5). The data are clearly divisible into three distinct categories — before, during, and after superior conjunction. Excluding the period around superior conjunction (March through May) it is evident that the post-conjunction data are considerably less noisy than the information gathered earlier. The difference is due largely to the new 200-kW transmitter. Data collected over the 4-month post-conjunction interval exhibit a 1-σ deviation of less than 15 m despite the spacecraft's distance of more than 2 AU. The true errors may even be somewhat smaller than those shown because of systematic trends and biases not included in the processing (Ref. 5).

Around superior conjunction, two phenomena combined to degrade the data quality and produce an increase in the 1-σ deviation to 60 m. First, as the sun-earth-probe angle became small, the antenna began to receive solar radiation, increasing the effective system noise temperature. This is caused by imperfections in the antenna pattern as well as by reflections from the quadripod structure. Second, significant spectral broadening resulted from signal scintillations as the ray path passed through the solar corona.

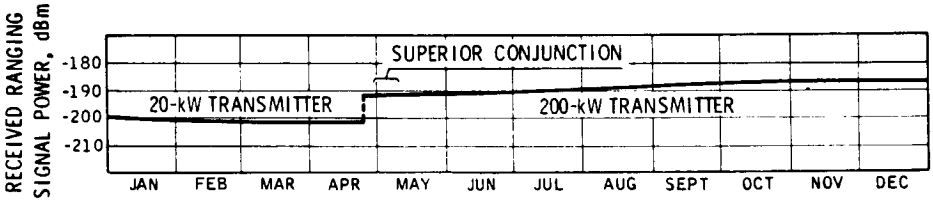


Fig. 5. Mariners VI and VII received ranging signal power

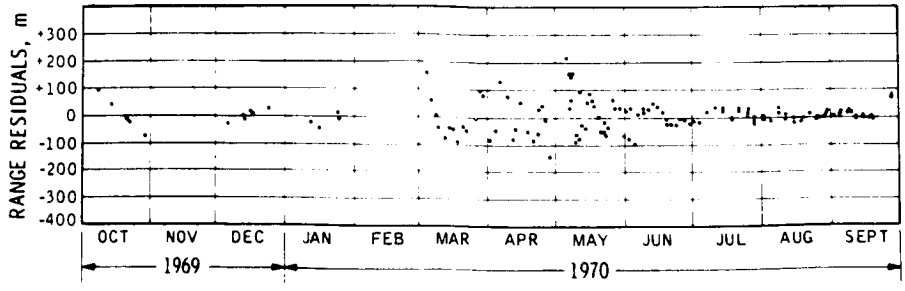


Fig. 6. Mariner VI range residuals

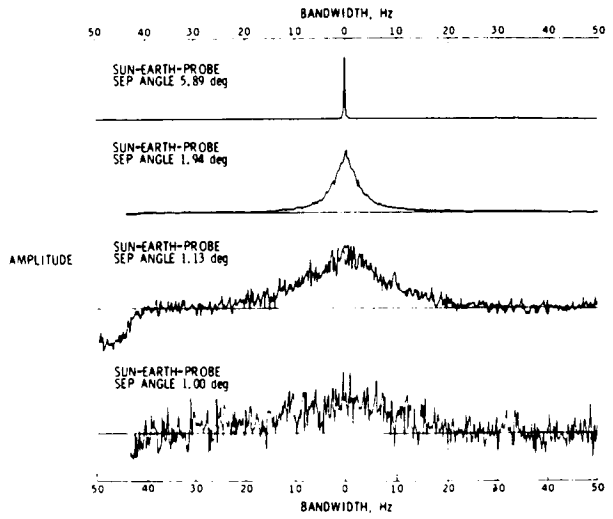


Fig. 7. Spectral distribution of received radio signal

Figure 7 shows the spectral spreading of the S-band radio signal as the ray path neared the sun. These measurements were made by R. Goldstein, P. Reichley, and G. Downs of JPL using the Mariner VI spacecraft and the 210-ft antenna at DSS 14. The lower plot was obtained on April 30, 1970 when Mariner VI was at superior conjunction. It graphically demonstrates the reason for the gap in the range residual data of Fig. 6. The signal degradation depicted in Fig. 7 is the result of both scintillations and increased noise temperature, which rose to well over 100°K. The figure also illustrates the rapid deterioration which occurs when the sun-earth-probe angle falls below 3 deg.

VII. Equipment Characteristics

To make the most effective use of the ranging equipment the user must specify an acceptable uncertainty in the data. Typically, a trade must be made between data noise and tracking time. During studies of charged-particle dynamics, an additional factor should also be considered, namely, the required resolution of the medium's dynamical characteristics. For example, referring to Fig. 4, a slow variation over several hours is clearly discernible. However, deviations on the order of 5 min would not be visible since each point represents data averaged for 15 min. Thus, in some instances a larger data undertainty may be preferable to the masking of short-term variations occurring at long integration times.

Figure 8 provides a means for selecting the proper averaging time and is a plot of the relationship derived by Goldstein (Ref. 6).

$$\sigma = \frac{1}{4}\sqrt{\frac{No}{Pt}} \qquad (10)$$

where

No = noise spectral density

P = received signal power

t = integration (averaging) time

Consider the situation that existed during the recent segment of the Mariner Mars 1969 relativity experiment. The received ranging signal level was approximately -192 dBm using the 200-kW transmitter, and a typical integration time of 120 s was employed for each point. Ten samples were collected and averaged to get one range number for each acquisition. Therefore, the effective integration time was 1200 s. Figure 8 indicates that a signal level of -192 dBm and an integration time of 1200 s should yield an uncertainty of approximately 20 ns. Returning to Fig. 6 and considering the interval from June through September, the period when the above conditions were applicable, the 1-σ uncertainty is approximately 40 ns. Two conditions combined to increase the data noise. First, no effort has been made to correct the range data for plasma induced errors. Second, there is the experimental process noise described by Anderson (Ref. 5). The effects of process noise are clearly discernible in the early range data where a strong sinusoidal character is evident.

Drift or instabilities in the equipment must also be considered as a potential source of error. The frequency standard has already been discussed in relation to the doppler data, and any instabilities produce similar effects upon the range information. Specifically, the greater the round-trip time the greater the potential uncertainty from this component. For example, the round-trip time of 45 min that existed near superior

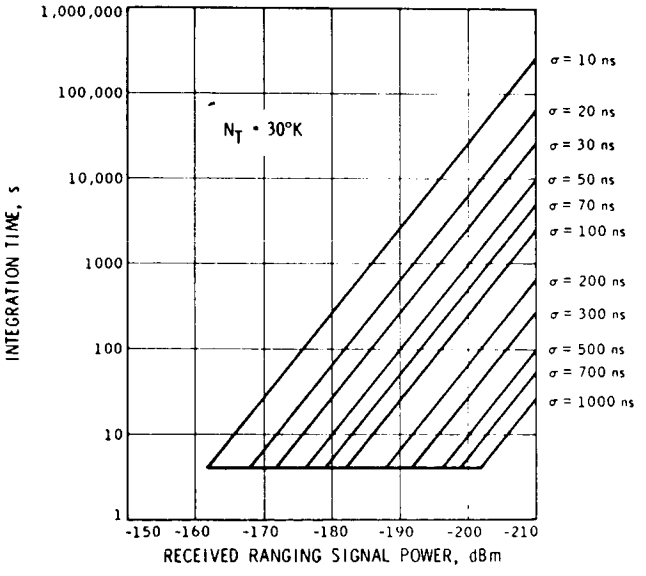


Fig. 8. Constant phase jitter contours, μ ranging

conjunction could result in range errors of several nanoseconds when using the rubidium vapor oscillator.

Stability measurements were made on the ground equipment which was configured to measure its own delay (Ref. 7). All of the subsystems normally employed in a ranging measurement were used in these calibrations and in an identical manner. The single addition was a zero-delay device affixed to the antenna's surface which replaced the spacecraft's transponder. Its purpose was to convert a portion of the transmitted signal to the received frequency without introducing a measurable delay. The results appear in Fig. 9, which shows the change in the measured station delay over an 8-h period. A least-squares fit to the data indicated a maximum change over the test interval of approximately 8 ns with worst case slopes approaching 2 ns/h. It is reasonable to conclude from these data and other similar tests that ground equipment drift is not likely to be a major problem to most experimenters.

Comparing the data of Fig. 9 with the expected uncertainty predicted by the constant phase jitter contours (Fig. 8) produces an interesting confirmation of the theoretical calculations. The test conditions were: received signal level -185 dBm; integration time 300 s. The 1-σ error predicted by Fig. 8 is about 18 ns while the value computed from the drift data of Fig. 9 is 20.4 ns.

VIII. Transponder

The discussion thus far has neglected the spacecraft transponder as an error source. This is largely due to the paucity of information regarding the transponder's performance in a varying environment. Understandably, the problems associated with calibration are enormous and it simply may not be economically feasible to collect sufficient data. The major problem is accessibility. All measurements must be made prior to launch, since no practical method has been devised for in-flight calibrations. Long-term delay stability measurements are impractical since the transponder is not normally available until just prior to launch. In the past, calibrations have been conducted over a period of a few days at three different temperatures and five signal levels, with the spacecraft enclosed in a thermal vacuum chamber. Because of thermal gradients, the chamber with its simulated solar radiation is the only method for providing the correct test environment. Obviously, the procedure is very expensive, making transponder calibration one of the first things to be curtailed when cost cuts occur.

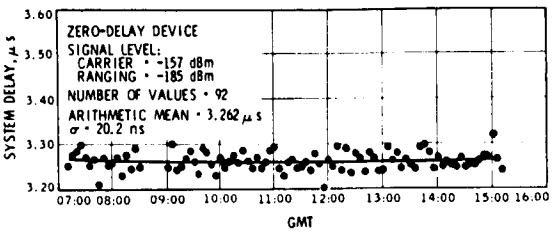


Fig. 9. Ranging system drift (Nov. 4, 1969)

Testing of identical, non-flight, transponders outside of the vacuum chamber has been considered, but even if it were possible to simulate the thermal environment, other major difficulties exist. Figure 10 accurately depicts the problem. The figure shows the range code delay through the transponder at various signal levels for two "identical" units. These devices were flown on the Mariner Mars 1969 spacecraft indicated on each curve. Two major problem areas can be identified. First, test equipment limitations precluded measurements at carrier signal levels lower than -120 dBm, although a major portion of the extended mission was conducted below this point. Second, note the rather large difference in delay through the two transponders at strong signal levels. For example, the disparity is 44 ns at -120 dBm and becomes larger with stronger signals. The former problem can be solved by procuring better test equipment, but the latter one with its implications on long-term stability may not be so simple. It must be emphasized that a predictable variation of range delay with signal power need not be important since signal strength information is telemetered back from the spacecraft. Rather it is the long-term changes of those characteristics described in Fig. 10 that are of concern. The ultimate solution to the problem may be one of transponder redesign to reduce both signal strength dependency and unit to unit variation.

While examining future transponders, it would be useful to consider a universal unit containing an encoder identical to the one at the transmitting station. The system acts as a very narrow-band tracking filter which locks to, and reconstructs, the range code received by the

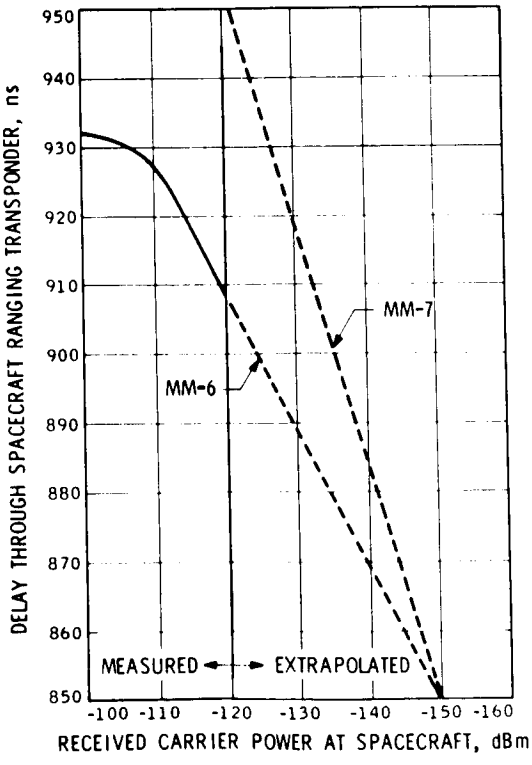


Fig. 10. Mariner Mars 1969 spacecraft ranging transponder delay

spacecraft. Although the received code may be very noisy, the signal which is retransmitted from the spacecraft is virtually noise-free. Restated, the potential improvement approximates the reciprocal of the signal-to-noise ratio. The implications for deep space missions are enormous and the technique may offer considerable savings in terms of ground equipment requirements.

IX. Mariner Venus-Mercury 1973 S/X-Band Radio Tracking System

Figure 11 is a simplified block diagram of the radio tracking system planned for the Mariner Venus-Mercury 1973 Project. Beginning with the ground equipment, an S-band carrier at 2116 MHz is transmitted from the 210-ft antenna. Upon arrival at the spacecraft, the received frequency is multiplied by 240/221 and retransmitted with a power of 20 W at 2298 MHz. This part of the spacecraft's transponder is identical to earlier units and will provide an S-band range and doppler capability.

The X-band section of the transponder coherently multiplies the received carrier frequency by 2640/663 to 8427 MHz. A 200-mW amplifier connects through a separate feed system to a 48-in. (122-cm) partially steerable antenna. A separate range channel is included in order to provide a simultaneous S/X-band ranging capability.

Upon return to the ground station the signal enters the S- and X-band receivers via two separate feeds. Each receiver provides independent doppler extractors which are referenced to the same frequency standard. Present plans include an S-band ranging capability, and it is expected that a separate X-band system will also be available.

The advantages of X-band over S-band are greater potential antenna gain and less interference from the medium. Accordingly, the existence of the dual doppler and ranging subsystems will provide some unique opportunities for radio propagation studies.

X. Conclusions

This paper has considered those aspects of the radio tracking equipment affecting data quality. An effort was made to include a description of planned expansion and development and to assess their meanings in terms of enhanced performance. A potential trouble area was identified and possible solutions were suggested. Finally, an attempt was made to provide experimenters with sufficient information to accurately estimate the capability of the equipment and, hence, to judge the probability of the experiment's success.

References

1. MacDoran, P. F., and Martin, W. L., Space Programs Summary 37-62, Vol. II, Jet Propulsion Laboratory, March 31, 1970.

2. Martin, W. L., Space Programs Summary 37-57, Vol. II, Jet Propulsion Laboratory, May 31, 1969.

3. Mulhall, B. D., Space Programs Summary 37-58, Vol. II, Jet Propulsion Laboratory, July 31, 1969.

4. MacDoran, P. F., this proceedings.

5. Anderson, J. D., et al., this proceedings.

6. Goldstein, R. M., Space Programs Summary 37-52, Vol. II, Jet Propulsion Laboratory, July 31, 1968.

7. Martin, W. L., Space Programs Summary 37-62, Vol. II, Jet Propulsion Laboratory, March 31, 1970.

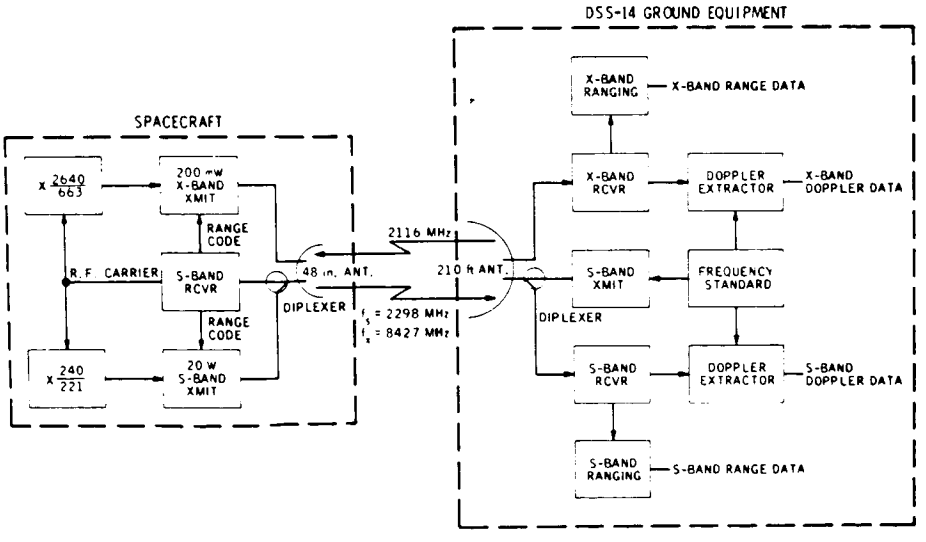


Fig. 11. Mariner Venus-Mercury 1973 X/S-band radio tracking system

Future Possibilities in Spacecraft Ranging and Doppler Systems

Mahlon Easterling
Jet Propulsion Laboratory
California Institute of Technology

I. Introduction

There is always a need for knowledge about the orbit of any spacecraft. The knowledge may be needed to properly evaluate the data collected by the spacecraft, to guide the spacecraft into a desired future orbit, or just to plan future operations and assure contact with the spacecraft. In certain cases, as when the spacecraft is used to probe a gravity field, the orbit itself is the prime scientific datum.

Knowledge about the orbit of a spacecraft is obtained by processing tracking data. While a variety of tracking instruments producing different kinds of data are used for near-earth orbits, the tracking of spacecraft at lunar distances and beyond is universally done by means of an active radar that is also used as the basis of the communication system for transmitting information to and from the spacecraft. The communications aspects are not of concern here, and we will concentrate on the tracking functions of the radar. It is well to review the structure of the radar as a starting point for a discussion of what the future possibilities of ranging and doppler systems may be. In this discussion, we will restrict our attention to systems that would be useful at lunar distances and beyond.

The structure of a spacecraft tracking radar is shown in Fig. 1. A master oscillator provides a stable frequency signal, which the transmitter multiplies up to the carrier frequency, amplifies to a high power level, and supplies to the antenna for radiation to the spacecraft. The spacecraft is shown as having separate receive and transmit antennas, which is typical of the major mode of operation of Mariner-class spacecraft; but only one antenna is sometimes used. The basic system structure is the same in either case. The carrier signal is received by a phase tracking receiver one of whose outputs is locked in phase to the received carrier. The term locked in phase means that the phase of the output signal is always proportional to the phase of the input signal except for momentary transient disturbances due primarily to noise. The average of these disturbances approaches zero with increasing time.

In general, the output signal is at a lower frequency than the input signal. This output signal is the input to the spacecraft transmitter. The transmitter multiplies the input up to the downlink carrier frequency, which is different from the uplink carrier frequency, amplifies it to a high power level, and delivers it to the antenna for radiation to the earth. The ground receiver is similar to the spacecraft receiver in that one of its outputs is phase-locked to the input carrier. This output is fed to the doppler extractor, where it is compared with a signal from the master oscillator to determine the two-way doppler shift on the carrier caused by the relative motions of the spacecraft and the ground tracking station. The doppler data are the primary tracking data produced by the radar.

The radar also measures range by means of a second signal, which is carried as a modulation on the up- and downlink carriers. It is produced in the ranging equipment from a signal provided by the master oscillator. In general, the ranging signal is an aggregate of signals, which may be transmitted together or sequentially, and which together have a very long period. In addition, they

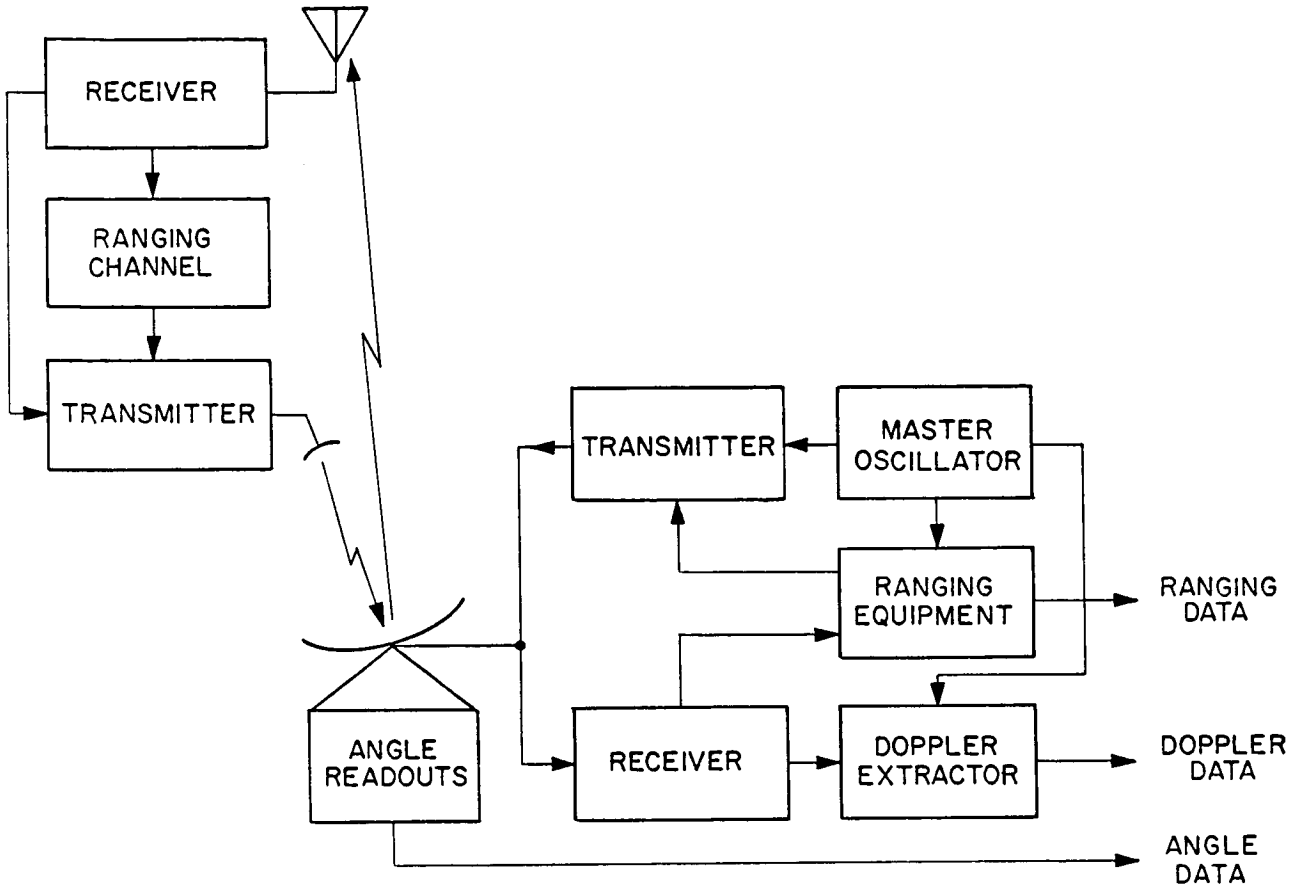


Fig. 1. Spacecraft tracking radar

have a structure that permits the phase of the aggregate to be measured to a very small fraction of a period. This signal is modulated onto the uplink carrier by the ground transmitter and demodulated from the uplink carrier by the spacecraft receiver. The signal is amplified and filtered by the ranging channel and modulated onto the downlink carrier by the spacecraft transmitter. It is demodulated from the downlink carrier by the ground receiver, which also amplifies and filters it. The signal from the ground receiver is compared with the originally transmitted signal to determine the phase shift or time displacement of the returned signal. This time displacement is a measure of the time required for the signal to travel to the spacecraft and return, and hence, of the range to the spacecraft.

It is possible to use the radar to determine the angular direction to the spacecraft by measuring the direction in which the ground antenna points. The ground antennas are always equipped with angle readouts which may be used for this purpose. However, once the spacecraft is more than a few tens of thousands of kilometers from the earth, the direction can usually be computed from the orbit much more accurately than it can be measured by the pointing of the antenna. In fact, the usual procedure is to use the angle readouts to point the antenna from computed angles rather than using measured angles to help determine the orbit. It should be noted that this is not the case for near-earth orbits. For our purposes, angle measurements need be considered no further.

One final comment on the overall structure is in order. All signals used in the radar are derived from one master oscillator. This means that all frequencies, phases, and time scales are referred to one basis. This permits the frequencies of the signals at various parts of the system to be selected so as to avoid interference and to solve mechanization problems without destroying the phase relationships between these signals. As an example, the radar measures the doppler shift essentially as though the uplink and downlink frequencies were identical to each other and to the frequency of the oscillator itself, even though there are many deliberate shifts in frequency in the system. Of course, the frequency manipulations must all be very carefully done, but as long as there is only one oscillator, the fundamental character of the data is as though all of the frequencies were the same. Moreover, since the ranging signal is also derived from the master oscillator, it is possible to use the fixed relationship between phases of the carrier and the modulation to assist in measuring the phase of the returned modulation. As we discuss below, this phase relation is disturbed by the interaction of the radio signal with charged particles, and this disturbance can be used to determine something about the number of charged particles in the propagation path.

Now that we have looked at the structure of the radar, it is appropriate to examine more carefully just what the radar is measuring. When radars were first applied, the distances were such that the time required for the radio signal to

propagate to the target and back was very short on the time scale by which movements of the target were measured. This is still the case in most applications, and the terminology applied to the data reflects this fact. Doppler measurements are usually understood to be measurements of the frequency shift of the returned carrier caused by the radial velocity of the target relative to the tracking station. The time at which a measurement is made can be considered to be the time of transmission, the time of reception, or any time in between without appreciable error.

In the case of radars for tracking spacecraft at planetary distances, the round-trip propagation time for the radio signals is no longer short. Indeed, it is many minutes or even tens of minutes long. While this is still short in terms of the orbital period of the spacecraft, it is long enough so that the radio path length to the spacecraft can change appreciably during a round-trip propagation time. That is, it can change appreciably in terms of the accuracy to which it can and needs to be measured. So also can the doppler frequency shift change appreciably during a round-trip time. This points up the need for considering the measurements in a somewhat different light than the terms, doppler and range, would indicate.

To see what the measurements really are in the case of spacecraft tracking, it is helpful to consider that the radar establishes a continuous flow of radiation to and from the spacecraft. It is continuous in time in that it is emitted by both transmitters as a continuous carrier and continuous in space in that it exists everywhere along the path between the tracking stations and the spacecraft in both directions. It is also helpful to consider the doppler extractor as a continuous phase comparitor. If there were a situation in which nothing was moving, the doppler extractor would show some constant phase between the carrier transmitted by the ground transmitter and the carrier transmitted by the ground receiver. If things were then allowed to move, the phase comparitor would show a different phase at some later time.

In practice, the output of the phase comparitor is sampled periodically, say every few seconds, for example. The phase comparitor is fitted with an accumulating counter arranged so that the sampled output is shown in terms of wavelengths and fractions of wavelengths. The difference between the outputs at two samples is the difference between the path length traversed by the radio signal that was arriving when the first sample was read and the path length traversed by the radio signal that was arriving when the second sample was read. This difference is given in terms of wavelengths and fractions of wavelengths of the radio carrier frequency. There are complications due to the fact that the uplink and downlink carriers are of somewhat different frequencies and that the accumulating counter is fed a fixed bias of one million counts per second so that it never has to count negative changes, but these do not invalidate our conceptual picture.

Each arriving segment of the radio signal has a phase relative to the constant phase transmitted signal, which is determined by the total path length traversed by that segment. Since the path is changing continuously, the phase changes continuously. The total change in phase from the start of tracking is accumulated and sampled out periodically. The counter does, of course, have finite capacity, and when the capacity is reached, it merely starts over again. This causes no difficulty, since if one sample reads 9968 and the next reads 0083, we can supply the missing fifth digit from the previous sampled counts that have been obtained. It should be stressed that the counter is not reset when the count is sampled out but is allowed to continue counting undisturbed. Even the technique for obtaining fractions of wavelengths does not interfere with the accumulation process.

Since each cycle of the carrier is just like every other cycle, it is not in general possible to tell how many cycles are stretched out along the path to the spacecraft and back. That is, when the phase of the received signal changes by one cycle (the path length changes by one wavelength), we cannot tell whether the total phase has changed from 100 cycles to 101 cycles or from 101 to 102 or, more realistically, from 687,354,693,872 to 687,354,693,873 or from 687,354,693,873 to 687,354,693,874. In practice, the accumulating counter is started at an arbitrary count at the beginning of a tracking pass. Thus, each phase measurement is in error by an unknown amount, but all measurements made during a pass are in error by the same unknown amount, so that the difference between two measurements made during the same pass is correct. Thus, the doppler measurements taken during a pass together show a time history of how the path length varied over the pass.

Obviously, the basic shortcoming of the doppler measurement is that it gives only the change in path length rather than the total path length. This shortcoming is removed if a ranging system is added to the tracking system. The long wavelength of the ranging signal permits the initial phase of the returned ranging signal to be determined from a priori knowledge. The measurement itself is made in the same way as the doppler measurement in terms of the phase of the returned signal relative to the transmitted signal. The measurement is made continuously, and changes of phase are accumulated in a counter. In practice, only fractions of a cycle are measured; the integer number of wavelengths is supplied when the data is processed. Since the wavelength is usually about 300,000 km, there is no uncertainty about how many need be added to obtain the total path length.

The fundamental differences in the range and doppler measurements arise from the following points. First, of course, the one measures the total radio path length while the other measures only changes over a tracking pass. Another difference, which becomes important when extreme accuracy is required, is that one measurement is based on the phase of the carrier while the other is based on the phase of the modulation. Thus, one propagates at the phase velocity of the signal, the other at the group velocity. In the presence of charged particles, these are not quite the same. The other points are concerned with the difference in the character of the signals. Because the ranging signal has such a long period and it is necessary to measure its phase to a very small fraction of a cycle, the signal has a complex structure and the phase measuring process is quite complicated. Even so, it is not possible to measure the phase of the ranging signal to nearly as great precision

as the carrier signal in terms of absolute time displacement. This is primarily because of the frequency limitations on the ranging signal. Since it is carried as modulation, the signal must fit within the bandwidth allocated. Thus, the highest frequency component is less than one one-thousandth of the carrier frequency. Even more complex measurement schemes cannot begin to compensate for this difference.

In summary, then, both the ranging and the doppler measurements are measurements of the phase of a continuous signal that propagates from the ground tracking station to the spacecraft and back. In both cases, the measurement is made continuously, and all changes in phase are accumulated in a counter. The counter is sampled periodically, and these samples constitute the data. The doppler data are a very fine-grained measurement made on the short-period carrier signal and show the time variation of the path length over a tracking pass. The ranging measurement is a much coarser-grained measurement made on a long-period complex modulation signal and shows the total path length as well as changes over a tracking pass. The measurement of the changes is, of course, much less precise than that made by the doppler. Because of the significant differences, the two data types are truly complementary, even though both are related directly to the path traveled by the radio signal and both would respond to changes in that path in the same way if the path were entirely in truly empty space.

It has always been recognized that the ranging and doppler measurements were inherently complementary and could be useful in different ways. The doppler measurements have been the primary measurements used for orbit determination. In fact, because of a combination of circumstances, the first ranging system on a U.S. spacecraft that went to the moon or beyond was that on the Lunar Orbiter series. This was essentially a system under development for, and later used on, the Apollo spacecraft. It was intended to assist in rapidly redetermining an orbit after a maneuver. Its aim was to achieve an overall accuracy at lunar distances of 15 m, but because of limitations to transponder development, its accuracy was probably only two or three times that. This seemed adequate, since it was much more than good enough to assist with the orbit determination process. The Lunar Orbiters did show, however, that range measurements had other uses. Once the orbit about the moon was determined, unaccounted for variations in range measurements were used to refine the earth-moon ephemeris.

The first ranging system used on a planetary spacecraft was that on Mariner V to Venus. The transponder was the same as that flown on the Lunar Orbiter and which had been developed for the Mariner missions to Mars in 1964. In fact, this same basic transponder was also used in Mariners VI and VII and will be carried on the Mariners intended to orbit Mars in 1971. The planetary ranging system differed from the lunar ranging system only in having a much more sensitive ground equipment capable of making the measurements on the vastly weaker ranging signals returned from planetary distances. The overall accuracy was the same and was realistically quoted at 50 m.

The ranging data from Mariners V, VI, and VII were very useful in the celestial mechanics experiments associated with those missions, as well as being helpful in orbit determination; there appeared to be no real need for greater accuracy. Thus, all ranging systems used on U.S. lunar and planetary spacecraft were developed as an adjunct to the doppler system intended for quickly redetermining an orbit after a maneuver. The accuracies achieved are estimated to be about 50 m. Although more sensitive ground equipment has been developed to cope with the Mariner V, VI, and VII flights and the Mariner VI and VII extended mission, the development was on an ad hoc basis, with only one tracking station equipped for ranging at planetary distances. Moreover, all the transponders used were of essentially the same kind developed in the early 1960's.

II. Probable Future Uses of Ranging and Doppler

Recently, two new uses have arisen for range measurements that have somewhat different requirements than heretofore. The one is to use ranging in conjunction with doppler to attempt to determine the total number of charged particles in the radio path. The other employs ranging as the primary data type in determining certain kinds of orbits. These uses may be different enough from past ones to require significant further development of ranging systems. The doppler system would seem to be adequate except for needing possible increases in stability as noted below. For this reason, the discussion of future developments and a detailed consideration of the limitations of these developments will emphasize ranging rather than doppler systems.

It was noted above that when there are charged particles in the radio path, the propagation velocity for the carrier and for the modulation is different or, more precisely, the total phase shift in the returned carrier and in the returned ranging signal is different. This implies that the change in path length over a tracking pass as determined by the doppler is different from that determined by the ranging. If both systems have the requisite precision, stability, and accuracy, the effect of the charged particles can be determined and the tracking data corrected accordingly. This is necessary to achieve higher navigational accuracy, and the information about the charged particles is also of intrinsic interest.

The requirements on a ranging system for such use are quite different than for the present uses. Most important is that the system must be very stable over a time scale of a tracking pass. To make the notion of "very stable" more precise, we note that it is customary to construct a ranging waveform as a binary combination of elementary binary waveforms, each with a duration of 1 μs. The most fundamental combination is an alternating sequence of positive and negative elements, i.e., a 0.5-MHz square wave. In measuring the phase of the received ranging signal, it is natural to work in terms of an integer number of elementary waveforms, i.e., microseconds, plus some fraction of a cycle of the 0.5-MHz component. The use of the 1-μs element instead of a 2-μs element is a consequence of the way in which the various components of the signal need to be combined and is not important to this discussion.

In most phase measurements, 1 deg of phase accuracy is considered quite good. We now seem to see a requirement for a resolution of three to ten times that much, corresponding to a range resolution of 0.3 to 0.1 m. More to the point, we wish to measure changes in range over a pass to this accuracy, so the system must be effectively stable to such a level over the pass. We discuss the fundamental limitations on stability, resolution, and accuracy below. It should be noted that there are similar requirements on the doppler, but here the only difficult aspect of the problem is the stability because the required accuracy and resolution are comparable to a whole cycle of the carrier. In the more distant future, it may be that the required range change accuracy will be even greater, perhaps as little as 5 cm.

The use of range data for orbit determination imposes somewhat different requirements, namely in overall accuracy. There is some reason to believe that overall accuracies of the order of 1 m may be useful. Here the problem is not resolution but absolute accuracy and the long-term stability to assure the accuracy even after the spacecraft has spent many months in space.

III. Basic Configuration of Ranging and Doppler System

The probable new uses of ranging and doppler data discussed in the preceding section do not in themselves impose any changes on the basic structure of the tracking radar. They are based on exactly the same kind of data as is now produced. The changes are in requirements for new standards of accuracy, stability, and precision. All three can be improved by further development of the basic system. However, it appears that it may be desirable to change the structure of the ground tracking station somewhat to permit continuous calibration as an alternative to inherent stability. This is discussed further in the next section.

Another factor which may lead to a change in the structure is the use of an X-band downlink for high-data-rate telemetry. It was mentioned above that the communication system shares the basic radar with the tracking system. Actually, the communication system dominates the radar and, in a very real sense, dictates the direction of its development. Thus, although it is well known that the number of charged particles in the radio path can be determined by comparing the change in path length of a tracking pass as measured by doppler at two different frequencies, such a method could not be incorporated into the tracking system unless there were a communications need for a second frequency. There is now such a need, but only on the downlink. Given the X-band downlink, it is quite possible to use it for tracking. A probable structure of such a radar is shown in Fig. 2. The important thing to observe is that there is an S-band uplink and both an S- and X-band downlink. Moreover, both downlinks are coherent with the uplink, i.e., both downlink carriers are derived from the uplink carrier that is received at the spacecraft. This permits the ground tracking station to measure not only the S-band doppler but also a hybrid doppler, which is an S-band up and X-band down, allowing the two-frequency method of determining charged particles to be used.

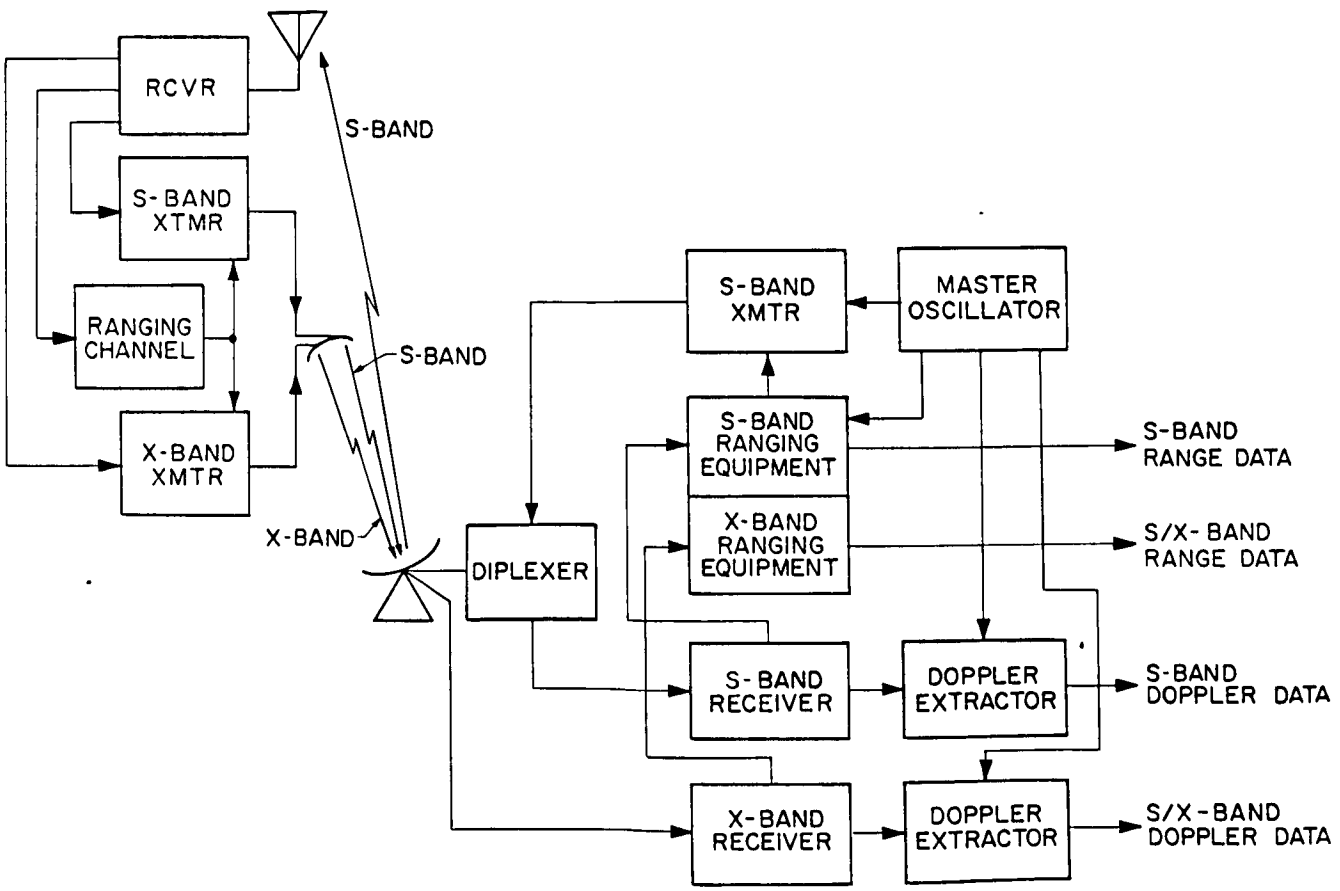


Fig. 2. S/X-band radar

The uplink ranging signal is also sent down on both carriers. Exactly the same ranging signal is modulated onto both carriers. The measurement of the two ranges uses essentially identical equipment. The two ranging equipments are shown joined together, since they share that part which generates the signal used to modulate the uplink. In the absence of charged particles, the two measurements would be identical just as would the two doppler measurements. In the presence of charged particles, the two measurements are different from each other, as well as giving a change in path length over a pass length which is different from that given by either doppler measurement. Obviously, having four kinds of data to compare yields more accurate information about the effects of the charged particles than just having two.

It is also obvious that having separate uplinks, too — essentially two independent simultaneously operating radars — would produce even better information. However, in the opinion of the writer, there is little likelihood of this happening in the foreseeable future. Should this opinion prove wrong, the limitations discussed below would still generally apply. The single major exception would be the possible use of a wider-band ranging signal on the X-band radar. This would permit both greater accuracy and greater resolution. The results derived below would scale by the bandwidth increase.

The hybrid S/X-band radar would not in itself yield better data for orbit determination than the S-band radar, but improved determination of the charged particle effects would, of course, improve the data, both range and doppler. An X-band radar would yield more accurate range data for orbit determination by permitting the use of a broader-bandwidth ranging signal, but, as remarked above, this does not seem probable.

IV. Inherent Limitations on Accuracy

The discussion above has been essentially qualitative and was intended to explain how the tracking system operates and why it is likely to evolve in a certain direction. In this section, we consider the accuracy that might be achieved in the future in terms of some of the inherent limitations on accuracy. How closely these limitations are actually approached will, of course, be determined by the needs for increased accuracy and by the resources available to continue the development.

It is convenient to consider the limitations on accuracy due to each of several parts of the system separately. These limitations will be those of the tracking instrumentation itself and will not include any that may be associated with the external physical world in which the system must operate.

A. Oscillator

The oscillator affects the accuracy of the tracking system in several ways. Perhaps the most fundamental is in determining the wavelength of the signals. Any error in frequency is translated directly into error in wavelength and hence in path length. The oscillators used at present are accurate to a few parts in 10^{11}, and oscillators under development appear to be accurate to one part in 10^{14}. This corresponds to range accuracies of 5 m and 1.5 cm, respectively, at a distance of 1 AU.

The error in oscillator frequency also affects the change in range over a pass as measured by either the doppler or the ranging systems. However, an extreme change in range over a 10-h pass might be 10^6 km. The corresponding errors for the oscillator accuracies quoted above would be 3 cm and 0.01 mm.

The stability of the oscillator also affects the tracking accuracy. Oscillator stability is a large subject and incompletely worked out, but for our purposes, a simple analysis will provide a bound on the limitation imposed by oscillator stability. A worst-case form of instability is an abrupt change in frequency by the full maximum frequency error. If the change were instantaneous, the effect, at a distance of 1 AU, would be a uniform apparent change in path length that would grow to 1.5 cm over one round-trip propagation time. After that, there would be no further change due to the oscillator. Experience shows that the frequency rate of change of the oscillators under development is very slow, so the calculation above is an absurd worst case. Thus, for the oscillators under development, instabilities would have an effect of only a fraction of a centimeter on the change in path length measured over a tracking pass.

B. Transponder

The transponder has an effect on the accuracy of ranging and doppler because there are finite propagation times of the signals through the transponder. We will not specifically treat the case of the carrier signal except to note that the hypothetical systems discussed here are approaching the point at which the propagation time of the carrier signal through the transponder will become of concern.

The propagation time of the ranging signal through the transponder, or more properly, through the spacecraft, is a part of the total propagation time measured by the ranging system. To obtain an accurate measure of the true propagation time to and from the spacecraft, the time through the spacecraft must be subtracted from the total measurement. This requires the careful measurement of the time through the spacecraft before launch. Since this is usually done with equipment very much like actual ranging equipment, the measurement can be made with both an accuracy and a precision appropriate to the system with which the spacecraft is to be used. The real problem is one of stability. This can be expressed in terms of the basic squarewave on which a ranging signal is built. In the S/X-band system hypothecated for the foreseeable future, the basic squarewave would have a wavelength of 600 m. An absolute range accuracy of 1 m corresponds to just over 1 deg in phase of this signal. While it is certainly possible to measure to this accuracy, it is difficult to achieve such stability in the spacecraft over a period of many months or years. It certainly could not be done using present design techniques.

An approach to spacecraft design which might achieve a long-term stability corresponding to a

meter or less is based on the fact that the path through the spacecraft has three parts. There is the passive RF circuitry, which includes the antennas and feed lines. This circuitry is very wideband, and its propagation stability is determined essentially by its mechanical stability. It should be possible to maintain that stability to an electrical equivalent of a few centimeters. The second part is the more conventional electronic RF circuitry in the transmitter and receiver. With care, and by maintaining broad bandwidths, it should also be possible to maintain a long-term stability of a few centimeters in this part. The third part is the ranging channel itself, which is intended to include all critical circuits and especially the filters that limit the bandwidth of the ranging signal. While it is customary to refer to the ranging signal as binary, it is really a binary signal that is filtered to fit within the band allocated. The filtering circuitry is the part that is most important to the stability, since small changes in a filter that cuts off at about 1.5 MHz will shift the phase of a 0.5-MHz squarewave. As will be noted below, there are some performance as well as frequency allocation reasons for not widening this bandwidth.

One effective method for controlling phase shifts in sensitive circuitry is to carefully design the circuitry and then enclose it in a controlled-temperature enclosure. By this means, it should be possible to achieve a long-term stability of less than 1 m, perhaps as little as 25 cm. It should be noted that this is the technique used to stabilize crystal oscillators. In an extreme case, it would be possible to borrow a further technique from the oscillator art and put in three channels. Range measurements would be made using the channels in sequence, say, on successive days, and compared to verify that no channel had developed an unexpected shift in propagation time.

Short-term instabilities in the ranging channel have an effect similar to that of short-term instabilities in the oscillator. However, a channel with a very-long-term stability of less than 1 m would be expected to have a stability over a few seconds, minutes, or hours of about 1 cm, so short-term stability would not be a problem.

C. Ground Equipment

The ground equipment under discussion here includes all parts of the tracking station that are involved in the ranging system. It is convenient to consider the effect of the ground equipment in two ways. The one has to do with the technique of making the measurements, the other with the techniques of calibrating the equipment.

The measurement of range is accomplished in two parts. The time displacement of the received signal relative to the transmitted signal is given in terms of periods of the signal, periods of the basic 1-μs element, and fractions of a period of the 1-μs element. The first part is supplied from a priori knowledge and is not a part of the measurement process. The second part consists of a kind of counting, and except for a certain very small probability of making a wrong count, is accurate and precise. The final part is a measurement in the conventional sense. It is already customary to make this measurement with a resolution of one part in a thousand or 0.15 m. There seems to be no reason why that resolution could not be extended to one part in ten thousand or 1.5 cm.

The precision of the measurement is probably several times as great as the resolution. Experiments have indicated that successive measurements under static laboratory conditions can differ by 1 m or more. This seems to be due to the counting process, in that the digital equipment which generates the local model of the received signal for comparison with the actual signal may end up in one of a number of different internal states, depending on the particular way in which the count was obtained. This situation is subject to control by more careful design, and it should be possible to make the precision approach the resolution.

The question of accuracy of the range measurement is related to calibration. The ranging equipment in a tracking station is actually a very small part of all the equipment involved in making a measurement. Most of the tracking station is really involved. One is attempting to make a measurement of the propagation time from the tracking station to the spacecraft and back. First, of course, there is the question of defining the positions of the tracking station and of the spacecraft. There are some interesting problems in deciding just where in a tracking station that has nearly an acre of reflector in the antenna and some hundreds of feet of cable between the antenna and the ranging equipment one places the point from which range is actually measured. Even the spacecraft is physically large relative to the precision of range measurements we are considering. One usually chooses a point in the tracking station related to the axes of rotation of the antenna. For some antennas, the axes do not intersect, but a suitable, although variable, point can be chosen. In the spacecraft, the center of mass would probably be preferable, but the line of sight from the earth to the antenna might not extend to pass through the center of mass. Also, there is likely to be more than one antenna, and one may be movable. Even the center of mass may move as the attitude control gas or rocket fuel is expended. Still, a suitable point can be chosen.

Conceptually, one would like to calibrate the entire ranging system by placing the spacecraft at a known distance from the tracking station and making a measurement. The amount by which the measurement exceeded the known distance would be the calibration constant and would be subtracted from all future measurements made at unknown distances. The two major obstacles to the application of this concept are that the spacecraft may not operate with the tracking station until after launch, and the tracking station is not stable enough to make do with one calibration. Accordingly, techniques have been developed for calibrating the two ends of the system separately.

In the case of the spacecraft, a procedure involving the disconnecting of antennas, measuring the rest, and calculating the propagation time through the antennas has been adequate but will require some refinement for the future. The ground equipment was originally calibrated by providing a sort of dummy spacecraft in the collimation tower at each tracking station. However, the dummy device itself required calibration, so a simpler

approach was devised. A small pickup probe is mounted in the main reflector of the antenna and feeds directly into a crystal mixer. The other input to the mixer is a signal whose frequency is the difference between the uplink carrier frequency and the downlink carrier frequency. One of the signals produced in the mixer is a replica of the uplink signal but at the downlink frequency. This is reradiated by the pickup probe and received by the ground receiver. The path length from the probe to a reference plane in front of the antenna is calculated; the path length from the probe to the subreflector and into the feed is the same as for signals to and from a spacecraft, as is the path inside the tracking station. Since the pickup probe and crystal mixer are both very small and wideband, the propagation time through them is negligible. This approach permits the calibration of a tracking station to an accuracy limited only by the precision of the measuring equipment and the accuracy of the master oscillator. If both of these were increased without limit, one would eventually run into the limits imposed by the mechanical stability of the antenna structure as a whole – perhaps 2 or 3 cm. Thus, there appears to be no difficulty with the calibration of the tracking station per se.

The stability of the tracking station is a problem quite like that of the spacecraft, but since the tracking station is electrically more complex and physically more extensive, the problem tends to be more severe. Fortunately, the tracking station can be calibrated often, before and after each pass, if required. Recent experience shows that the calibration may change by as much as a meter or two over a few hours' time. No doubt this could be reduced by careful redesign of parts of the equipment and by special cables less sensitive to changes in temperature. However, it should be recalled that the design of a tracking system, including the tracking station, is likely to be dominated by the communications needs rather than by the tracking needs, so that it might not be possible to control the station to the extent required. An alternative to increased stability is continuous calibration.

The signal normally transmitted to a spacecraft is exactly the same as the signal used in calibration. The transmitter is even run at the same power to be sure that is has the same propagation time. Thus, it would be possible to calibrate the tracking station while transmitting to a spacecraft. However, the receiver and the ranging equipment (as well as the doppler extractor) would all be operating on the calibration signal rather than on the signal being received from the spacecraft. A second receiver and ranging equipment (plus doppler extractor) would have to be provided to receive the signal from the spacecraft. If this were done, then the configuration of the tracking station consisting of the transmitter and the one set of receiving equipment could be calibrated, while the other configuration consisting of the same transmitter and the other set of receiving equipment was tracking the spacecraft. Of course, there would be only one of that portion of the ranging equipment which generated the transmitted ranging signal but two sets of the portion which measured the phase of the received ranging signal. A block diagram of a tracking station with two sets of receiving equipment is shown in Fig. 3. The diagram represents only the S-band portion of the station. If a hybrid S/X-band system were in use, then the X-band equipment would have to be extended, too.

There are two parts to a normal ranging operation. First, the ranging signal is acquired, i.e., the phase is determined by a search, and the counter that keeps track of the phase is set to the correct value. For the signal levels encountered at planetary distances, this may require several minutes. Once the signal is acquired, it is tracked, i.e., the phase is measured continuously, and the counter that keeps track of the phase may be sampled as required. However, if the phase of the received ranging signal were known, the signal could be acquired very quickly and the acquisition checked in a few seconds at most. Use can be made of this property to devise a strategy for continuous calibration.

Before the beginning of a pass, both configurations of the tracking station would be calibrated. When the pass began, one configuration would be used to acquire the signal from the spacecraft. Both the carrier and the ranging signal must be acquired; the ranging signal cannot be acquired until after the carrier has been acquired, but carrier acquisition is usually much easier at planetary distances than ranging acquisition. Once the ranging signal had been acquired, range (and doppler) data could be read out as required. Also, the second set of receiving equipment could acquire the carrier and ranging signals very quickly by using information about carrier frequency and ranging phase provided by the first. Then data could be read out from the second set of receiving equipment as required and the first set switched over to calibrate mode. Since both the carrier frequency and ranging phase are essentially known for the calibrate situation, acquisition is very fast. The two sets of receiving equipment could be switched back and forth as often as need be, even every minute or two if desired. Thus, a new calibration could be obtained as often as the stability of the two configurations was likely to be different, and the data produced would have to be tagged with the correct calibration number. It might also be

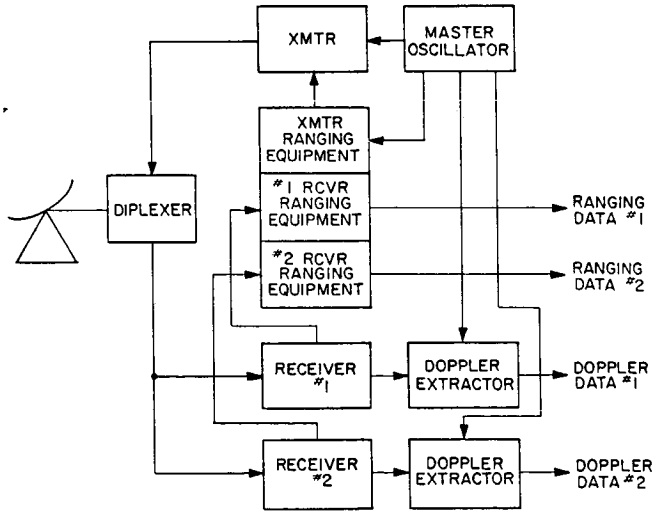


Fig. 3. Tracking station setup for continuous calibration

necessary to take the two calibration numbers into account when performing the fast acquisitions. Experience with present tracking stations has shown that they are very stable on a time scale of a few minutes, so a continuous calibration scheme of the sort described should eliminate the stability of the tracking station as a factor limiting the accuracy of a ranging system.

Although a detailed discussion of the limitations of the doppler system will not be presented in this paper, it should be evident from the above discussion that a similar calibration strategy could be devised for the doppler ground equipment. Of course, the total number of wavelengths in the path through the ground equipment would not be known, but changes during a pass would be detected. The ranging calibration is of no help to the doppler calibration, and vice versa, because significant parts of the paths taken by the two signals are different.

Finally, there is the question of whether a change over a time of the order of a round-trip time is due to a change in the transmitter or a change in the receiver. If it is the latter, it should be applied immediately to correct the signal being received, if the former, it should be used to correct a signal arriving one round-trip time later. What is most likely is that a change involves both the transmitter and receiver and that both receivers would change in much the same way, so that the changes in the parts of the system could not be separated. Fortunately, experience suggests that changes are likely to be slow enough for this not to be a problem; i.e., the changes of a meter or two over a pass occur rather uniformly in time, so that changes over a round-trip time are only a few centimeters.

D. Signal-to-Noise Ratio

The signal-to-noise ratio of the received ranging signal affects the accuracy obtainable. In principle, it is possible to average together as many observations as necessary to obtain any desired certainty in the measurement. The averaging is actually a least-squares fit of a curve through the observed data points, and the more scatter to the points, i.e., the more points involved in a fit, the more any fine structure in the data is obscured. Thus, in practice, a high signal-to-noise ratio would be desirable.

A high signal-to-noise ratio also facilitates the acquisition process. Again in principle, it is possible to take up to several hours for the acquisition. Experience shows, however, that it is generally undesirable to take more than about 20 min for an acquisition. This is a fairly arbitrary period, and it might be quite acceptable under some circumstances to take an hour or two. More than that would certainly interfere with the usefulness of ranging to measure charged particles.

Both of the above points require a more quantitative treatment. If an optimum estimation procedure is used, a certain averaging time is required to achieve a certain standard deviation in the estimate of the fraction of the basic 1-μs element of the ranging signal. This averaging time also depends on the signal-to-noise ratio or, more precisely, on the ratio of signal power to noise spectral density. A brief analysis is made in Appendix A, with the results presented in Fig. 4. In the figure, the required signal-to-noise ratio is given as a function of averaging time, with the standard deviation of the estimate as a parameter. A signal-to-noise ratio and averaging time can be chosen so as to achieve a desired standard deviation. The choice would be made depending on the accuracy required, the fine structure to be eliminated, and constraints imposed by the particular system in use.

The averaging time developed in Appendix A is, strictly speaking, applicable only to a static situation. However, it is customary to use information from the doppler extractor to determine the rate of change of path length. This permits the range measuring equipment to measure the phase of the received signal relative to a signal of the same frequency rather than relative to a signal of a different frequency. In effect, it converts a dynamic situation into a static one. The analysis given is thus applicable as a limiting case. The real meaning is that if the data are fitted to a curve over a certain time period, the rms deviation would be expected to approach that indicated by Fig. 4.

It is interesting to note that if high accuracies, say, better than 1 m or so, and short averaging times, say, less than a few hundred seconds, are required, a signal-to-noise spectral density considerably greater than 1 is necessary. It turns out that the corresponding acquisition time for any one of several acquisition schemes that have been used or proposed is well within the 20 min arbitrarily allowed above.

Of course, Fig. 4 and the preceding paragraph do not entirely dispose of the problem, because a user of a ranging system thinks in terms of range rather than signal-to-noise ratio. The translation of signal-to-noise ratio into range requires making several assumptions about ranging systems. The

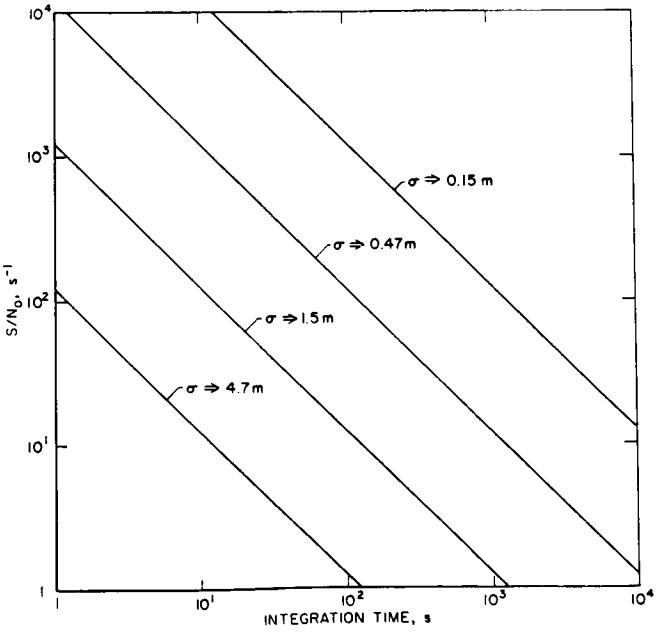


Fig. 4. Ranging signal-to-noise ratio as a function of integration time

nature of the assumptions that should be made about future systems is the subject of considerable debate among people working in this area. The reader should be aware that the assumptions made here are not concurred in by all of the people concerned. Moreover, the assumptions made fall into two categories. The one is the assignment of specific values to certain system design parameters. The other includes the form of the ranging channel assumed and the method of analysis used. The first, while there might not be general agreement, is not too important since the results of the analysis can easily be scaled for other values of parameters. The second is very fundamental and requires some justification.

The form of ranging channel assumed is the one in use now, e.g., on Mariners VI and VII, consisting of a wideband, high-gain, hard limiting amplifier. It converts the ranging signal plus the noise in the bandwidth of the amplifier to a slightly filtered binary waveform of constant amplitude. It is very simple and susceptible to being made quite stable by using the techniques discussed above. This form of channel is also adequate for missions approximately out to the orbit of Neptune, as shown by the analysis in Appendix B.

It is often argued that a different and more narrowband channel is needed to accommodate missions like the Mariner VI and VII extended mission. This argument overlooks the fact that such extended missions are essentially opportunistic and make use of spacecraft that were designed and launched for other purposes. For the other purposes, namely missions for obtaining and sending back scientific data, a significant telemetry capability is required. This capability will permit the use of the simple ranging channel assumed. Thus, there is no justification for a more complex channel to fulfill a purpose outside the scope of the mission no matter how much an extended mission user may desire it. On the other hand, if the ranging is required for the mission itself, there will certainly be adequate telemetry capability, so that the simple ranging channel will suffice.

Bandwidths other than the conventional 1.5 MHz are sometimes considered also. For accuracy and precision, a bandwidth as wide as possible would be desirable, together with a basic element length short enough to make use of it. The 1-μs basic element for the ranging signal is just about optimum for the 1.5-MHz bandwidth that is compatible with the RF frequency band allocation of 3.3 MHz. Large changes in bandwidth would be necessary to affect the performance significantly. For example, a reduction in bandwidth by a factor of 2 would change the received signal strength by a factor of 2 also, which would increase the accuracy by a factor of 2 (see Fig. 4), but the corresponding factor of 2 increase in element length would reduce the accuracy by a factor of 2, for a net loss by a factor of 2. Thus, it is desirable to keep the bandwidth as wide as possible and the element length correspondingly short.

The method of analysis employed is based on the use of a simple "turnaround" ranging channel and assumes that, whatever the range, there will be a prescribed downlink telemetry capability. This permits the uplink and downlink to be separated. The uplink analysis is based on a set of parameters which, the writer feels, will be about the most favorable achieved over perhaps the next 10 years. This places the upper limit on the range at which the uplink per se will operate at about 17 AU, as determined by the signal level required to operate the spacecraft carrier tracking loop. The value is derived in Appendix B.

The downlink effective received ranging power relative to the ground receiver noise spectral density is derived from the uplink analysis by assuming a certain downlink telemetry capability. Since the telemetry capability is different for different missions, it is included as a parameter. The results of the analysis given in Appendix B are shown in Fig. 5, which shows the ranging signal-to-noise ratio versus range for various telemetry capabilities. Together with Fig. 4, it illustrates the accuracy achievable as a function of range and integration time. Since the downlink analysis is based on a prescribed telemetry capability, it applies to either the S-band or X-band downlink. Of course, in a particular situation, the two links may have different capabilities and, therefore, different ranging capabilities.

E. Operational Limitations

One major operational limitation has already been referred to, namely, the time required for acquisition. If this time exceeds more than a few tens of minutes, it begins to be burdensome, although it might be possible to take as long as an hour or more in special cases. Fortunately, it seems that such long times will not be required for regular missions.

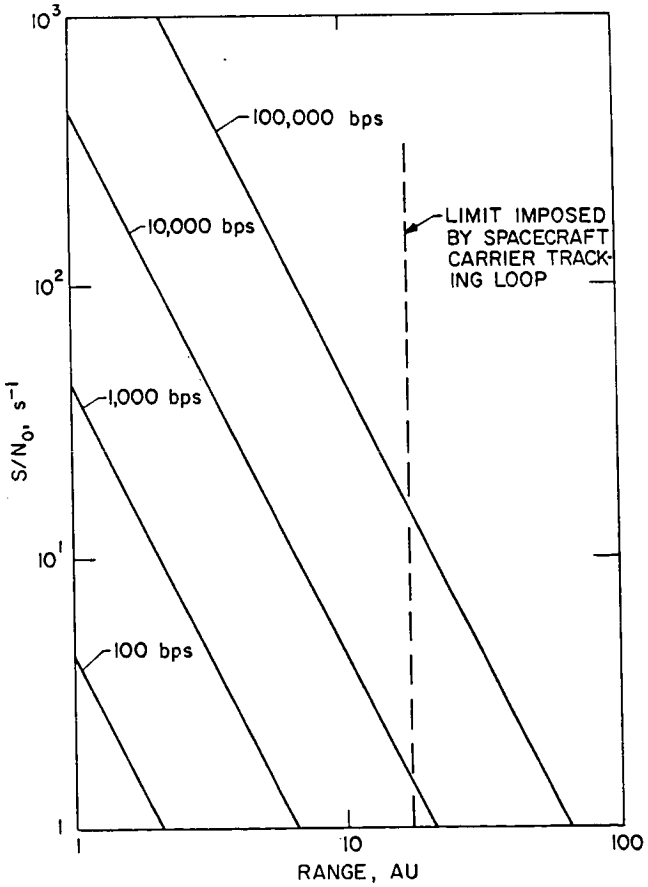


Fig. 5. Ranging signal-to-noise ratio vs range for various telemetry capabilities

Another possible operational limitation arises at extreme ranges. The round-trip time is approximately 16 min per astronomical unit of range, or 2 h 40 min at 10 AU. This reduces the time during a pass for which ranging (and doppler) data can be obtained and may impose some limitations on the use of tracking to determine the charged particles in the path.

V. Summary and Conclusions

We have considered the probable future uses of ranging and concluded that there will be three major ones. The first is the present use in orbit determination to complement doppler. The second is as the primary data in the determination of certain types of orbits. The third is to determine, in conjunction with doppler, the charged particles in the radio path. It seems that the kind of ranging data now obtained are suitable for all three of these uses but that improvements in accuracy and precision are required. Also, the use of a second carrier frequency would greatly aid in the charged particle determination. The need for an X-band telemetry link from the spacecraft makes possible the use of hybrid S/X-band tracking radar, with an S-band uplink and both S- and X-band downlinks. The structure of this radar was described.

The limitations on accuracy were explored, and it was concluded that the several parts of the radar would be capable of an accuracy of 0.1 m or less, except for the transponder, which, because of the need for long-term stability, might be capable of only about 0.25 m. This would permit absolute range measurements to an accuracy of less than 1 m and measurements of changes in range over a pass to less than 0.1 m. The very high accuracy might require continuous calibration of the tracking station.

An analysis of the effects of range on accuracy and acquisition time was performed, and it was shown that accuracy, not acquisition time, was limited by range. This is especially to the point for the new uses foreseen for ranging. With the kinds of high-data-rate telemetry proposed for future spacecraft, it should be possible to achieve high-accuracy ranging at distances up to 10 AU, with an absolute limit to ranging set by the uplink at about 17 AU using the equipment in prospect for the next 10 years. The question of a different form of ranging channel was considered, and it was concluded that, because of the downlink capability needed for telemetry, there was no need to abandon the present simple form of channel.

The absolute limit to ranging is set by the uplink, but if one speculates about further improvements, for example 420-ft ground antennas and 5-MW ground transmitters, the absolute limit on distance moves out toward the edge of the solar system. This, of course, would require a corresponding change in the spacecraft to increase the downlink capability. Finally, it would be necessary to find new locations for tracking stations that could keep the spacecraft in sight for longer continuous periods of time; the round-trip propagation time to the orbit of Pluto, for example, is about 11 h.

Appendix A

Analysis of the Effect of Signal-to-Noise Ratio on Averaging Time

If optimum independent estimates of the phase of the ranging signal are made, the variance on each estimate is

$$\sigma^2 = T^2 N_o / 8St \text{ s}^*$$

where

T = duration of the basic element

N_o = one-sided noise spectral density

S = signal power

t = integration time

In a typical ranging system, $T = 1\ \mu s$. Also, since the speed of light is 3×10^8 m/s, a standard deviation of 10^{-9} s in the estimate of the fraction of the basic element corresponds to a standard deviation of 0.15 m in the estimate of range. Thus, the above equation can be used to relate S/N_o to t with σ as a parameter, i.e.,

$$\frac{S}{N_o} = \frac{10^{-12}}{8\sigma^2 t}$$

This relationship is plotted in Fig. 4.

Appendix B

Analysis of Ranging Signal-to-Noise Ratio as a Function of Range

Consider the signal power on the uplink as a function of range. Assume:

210-ft ground antenna
500-kW ground transmitter
10% of total power in carrier
90% of total power in ranging signal
Omnidirectional receiving antenna on spacecraft
300 K spacecraft receiver temperature
1.5-MHz ranging channel bandwidth

This leads to the following calculation of the received ranging signal-to-noise ratio (SNR) at a distance of 1 AU:

Transmitter power	+87	dBm
90% of power in ranging	-0.5	dB
Transmitter antenna gain	+60	dB
Space loss	-263	dB
Receiving antenna gain	0	dB
Receiving ranging power	-116.5	dBm
Receiver noise spectral density	-173.8	dBm/Hz
Ranging channel bandwidth	61.8	dB/Hz
Noise power	-112.0	dBm

*Goldstein, R. M., "Ranging With Sequential Components," JPL Space Programs Summary 37-52, Vol. II, May 1 to June 3, 1968, pp. 46-49.

Thus, at 1 AU, the signal-to-noise ratio in the ranging channel bandwidth is -4.5 dB. Since the output SNR is -2 dB relative to the input because of the limiter, the output has a signal-to-noise ratio of -6.5 dB. Thus, the reduction in actual downlink transmitted ranging power relative to allocated ranging power is

$$\frac{P_{R_{effective}}}{P_{R_{allocated}}} = 0.224 \times \left(\frac{1\ AU}{range}\right)^2$$

There is a limiting range beyond which the uplink does not work because the spacecraft carrier tracking loop will not operate. The usual loop bandwidth is 20 Hz, and the loop requires a 10-dB SNR in this bandwidth to operate properly. The limiting range can be derived by calculating the SNR at 1 AU and then scaling by $1/R^2$.

Transmitter power	+87	dBm
10% of power in carrier	-10	dB
Transmitter antenna gain	+60	dB
Space loss	-263	dB
Receiving antenna gain	0	dB
Received carrier power	-126	dBm
Receiver noise spectral density	-173.8	dBm/Hz
Carrier loop bandwidth	+13	dB/Hz
Noise power	-160.8	dBm

$$SNR = +34.8\ dB$$

excess SNR = +24.8 dB - factor of 300 in R^2. Therefore, the limitation in range due to the carrier loop is 17.3 AU.

The analysis of the downlink requires a different approach. The basic assumption on which the approach is based is that a certain telemetry capability will be provided on each spacecraft. This capability requires some combination of transmitter power, antenna gain, and other system parameters to guarantee a certain received ratio of signal energy per bit to noise spectral density ratio no matter what the distance is. For reasonably high data rates and coding, this ratio is about 2, i.e.,

$$\frac{P_D T_B}{N_o} = 2$$

where

P_D = data signal power

T_B = duration of each bit

N_o = one-sided noise spectral density

If the downlink ranging power is allocated 10% as much power as the data signal, then

$$\frac{P_{R_{received}}}{N_o} = \frac{0.2}{T_B}$$

As shown earlier in this appendix, the effective ranging power is less than the allocated ranging power, so that the ratio of the effective received ranging power to the noise spectral density is

$$\frac{S}{N_o} = 0.0448\ \frac{data\ rate}{(range)^2}$$

where the data rate is in bits per second and the range is in astronomical units. This relation between S/N_o and range is plotted in Fig. 5, with downlink data rate as a parameter. Since the downlink analysis is based on the assumption of a certain telemetry capability, it applies to either S- or X-band, although in a particular case, they may have different rates and, therefore, different capabilities.

The Lunar Laser Ranging Experiment

P. L. Bender, Joint Institute for Laboratory Astrophysics; R. H. Dicke and D. T. Wilkinson, Princeton University; C. O. Alley and D. G. Currie, University of Maryland; J. E. Faller, Wesleyan University; J. D. Mulholland, Jet Propulsion Laboratory, California Institute of Technology; E. C. Silverberg, McDonald Observatory; H. E. Plotkin, Goddard Space Flight Center; W. M. Kaula, University of California at Los Angeles; and G. J. F. MacDonald, University of California at Santa Barbara

I. Introduction

The emplacement of the Apollo 11 retroreflector package on the lunar surface has made possible very accurate measurements of the lunar distance (Refs. 1-5). A continuing program of range measurements to the package at nearly all phases of the moon is being carried out by the McDonald Observatory under NASA support. Returned signals have also been obtained by the Pic du Midi Observatory in France and by the Air Force Cambridge Research Laboratories (AFCRL) Lunar Laser Observatory near Tucson, Arizona. It is hoped that several other lunar ranging stations will be in operation within the next year or two, including ones in Japan, Hawaii, Russia, and the southern hemisphere.

The present observing program at the McDonald Observatory consists of three observing periods on most nights when the weather permits, except for a period of 5 days around the new moon. One observing period is near the time of meridian transit for the moon, and the others are 3 or 4 h earlier and later. Several runs of about 50 shots each are normally fired during each observing period.

The ruby laser system being used at present gives 3-J pulses with a repetition rate of one every 3 s. The total pulse length between the 10% intensity points is 4 ns. The rms variation in the observed transit time due to the laser pulse length and the jitter in the photomultiplier receiving the returned signal is 2 ns. The present overall accuracy of the measured transit time is 2 ns, and improvement to less than 1 ns requires only refinement in the calibration procedures.

The uncertainty in the range correction for the effect of the atmosphere was shown some time ago (Ref. 6) to be less than 6 cm out to zenith angles of 70 deg. This result was based on using the surface value of the atmospheric refractive index as a predictor for the correction, as is often done in radio work. Recently, H. S. Hopfield (Ref. 7) pointed out that very much better corrections for the optical case can be obtained by using the surface pressure as the predictor. It now seems clear that the total error in the range correction for zenith angles of up to 70 deg will be less than 1 cm except under rare atmospheric conditions.

With data from two or more well located observing stations, the lunar range can be corrected accurately for the effects of polar motion and fluctuations in the earth's rotation rate. Very accurate corrections can be made for the earth tides at each station. It appears that the use of lasers giving roughly 0.1-ns pulse lengths is highly desirable. With them, single-shot ranging accuracies of about 3 cm are expected.

II. Model for Discussing Accuracy of Lunar Results

The actual lunar range results presumably will be analyzed by fitting a numerical integration for the lunar motion such as that of Garthwaite,

Holdridge, and Mulholland (Ref. 8) to the data. All the planetary effects are then included, and the lunar libration parameters and the location of the retro-reflector package with respect to the lunar center of mass can also be determined. However, in order to analyze the accuracy to be expected from the lunar ranging program, it is desirable to use a somewhat simplified model. For this model, we neglect the planetary perturbations and use the Hill-Brown solution (Ref. 9) to the main problem for the motion of the moon. The planetary masses and orbits appear to be well enough known so that uncertainties in them will not affect the results. Terms higher than second degree in the lunar gravitational field are neglected, and the theory of Eckhardt (Ref. 10) for the lunar librations is used. Effects of the earth's bulge, including precession and nutation, are neglected. Finally, we assume that the polar motion and angular position of the earth are well enough known so that changes in the errors in these quantities over a few days are small.

The parameters in the model are as follows:

a, e, e', i, a_1 -- parameters in Brown's theory corresponding to the mean lunar distance, lunar eccentricity, earth's eccentricity, inclination of the lunar orbit, and ratio of the lunar to solar distance. The a_1 also includes a factor $(E - M)/(E + M)$, where E and M are the masses of the earth and moon, and this factor will give the main uncertainty in a_1 when a has been determined fairly well.

ℓ, D, F, ℓ'; L -- Brown's four-angle variable and the mean longitude L of the moon; ℓ, D, and F are, respectively, L minus the mean longitude of perigee, of the sun, and of the node, and ℓ' is the mean longitude of the sun minus that of its perigee.

x_1, x_2, x_3 -- rectangular coordinates of the retro-reflector package with respect to the lunar center of mass measured along the moon's principal axes (x_1 is in the mean direction of the earth, and x_3 is toward the north lunar pole).

β, γ -- lunar libration parameters; $\beta \equiv (C - A)/B$, $\gamma \equiv (B - A)/C$, where A, B, and C are the lunar moments of inertia about the three principal axes.

σ, λ, z -- cylindrical coordinates of the observing station; ρ is the distance from the axis of rotation, λ the longitude, and z the distance above the equatorial plane.

ϵ -- inclination of the equatorial plane to the ecliptic.

As is well known from spacecraft tracking results, the differences between measurements before, during, and after meridian passage determine $\sigma \cdot \cos\delta$ and λ, where δ is the declination of the moon. The period of the terms from which $\sigma \cdot \cos\delta$ and λ are determined is 25 h, and it turns out that the shortest period terms which have to be determined accurately in the lunar motion and librations have periods of about 14 days. The accuracy of determining these two quantities thus will not be appreciably degraded by uncertainty in the lunar part of the problem once the lunar parameters have been moderately well determined. If data from only one station are available, some loss in accuracy will result from having to determine $\sigma \cdot \cos\delta$ and λ, but with several well located stations, little accuracy will be lost. There are thus in effect only 17 parameters instead of 19 which have to be determined from the lunar range terms with periods of longer than 1 day. Three of these, e', ℓ', and a_1 (actually a_1/a) are probably better determined from planetary radar or spacecraft tracking.

We take a right-handed earth-centered coordinate system for the observing station location, with the x-axis pointed from the center of mass of the earth to the center of mass of the moon. The y-axis is in the ecliptic plane, and the z-axis in the northern hemisphere. For the retro-reflector location we take a similar coordinate system, except that it is centered on the moon and rotated 180 deg about the z-axis. The topocentric distance to the reflector is then given by

$$\rho^2 = (r - x_E - x_M)^2 + (y_E + y_M)^2 + (z_E - z_M)^2$$

where r is the center-to-center distance, x_E, y_E, z_E are the earth station coordinates in the new system, and x_M, y_M, z_M are the reflector coordinates. Expanding this gives

$$\rho = \left[r - x_E - x_M\right] + \frac{1}{2}\left[\frac{\left(r_E^2 - x_E^2 + r_M^2 - x_M^2\right)}{(r - x_E - x_M)}\right] + \left[\frac{(y_E y_M - z_E z_M)}{(r - x_E - x_M)}\right] + \theta\left[\frac{\left(y_E^4 + \ldots\right)}{(r - x_E - x_M)^3}\right]$$

where r_E and r_M are the radii of the earth and moon at the observing station and reflector locations. If we write D as

$$\rho = \left[r - x_E - x_M\right] + \left[\frac{(y_E y_M - z_E z_M)}{a}\right] + C$$

we can show that the uncertainty in C affects only the constant term in ρ. The second bracketed term is always less than 3×10^4 m in amplitude, so that only the largest terms in it need to be considered. Thus, with some relatively small corrections, the conclusions concerning the accuracy achievable from the lunar range measurements are almost the same as if only the first bracketed term in ρ were present.

Using any of the above forms for ρ, we can calculate the partial derivatives of ρ with respect to each of the parameters. Each partial derivative consists of a sequence of terms of different frequencies. How well the corrections needed to the different parameters can be separated depends on whether the partial derivatives involve the different frequencies in substantially different ways. A slightly conservative estimate of the accuracy of the results can be obtained by taking 17 of the frequency terms in the partial derivatives which

appear to affect the results most and considering that only the coefficients of these terms in the range residuals are determined by the observations. The problem can then be inverted to find the accuracy of each parameter, given the accuracy with which the residuals at the different frequencies are determined.

In the above model, it may seem that the mean motions for the five angle variables in the problem should be included as parameters. However, if the corresponding phase angles are determined separately from the earlier and the later parts of the data, then the changes in the phase angles give a measure of the errors in the mean motions. Similarly, differences in phase angles from three sections of the data can be used to determine the secular acceleration. While additional parameters for the corrections to the mean motions and the secular accelerations would be included in a final analysis, they do not seem necessary at present.

III. Checks on Gravitational Theory

For several centuries the motion of the moon provided one of the most stringent tests of gravitational theory. With the very accurate range measurements now being made, it appears that this will again be the case. The most general test, of course, is to see which if any of the existing theories can be made to fit the observations by adjusting the known parameters in the problem. Whether or not a given deviation from an existing theory can be detected depends on whether the residuals it causes have the same frequencies as do the residuals caused by errors in some of the parameters to be determined.

One general relativistic correction term which enters in the motion of the moon and should be measurable has been pointed out by Baierlein (Refs. 11 and 12). It has an expected amplitude of about 1 m and a time dependence of the form cos 2D, where D is the difference in the mean longitudes of the sun and the moon. The only non-Newtonian part of the metric which plays a role in determining the coefficient of this term depends on Baierlein's parameter η'. The cos 2D term would have an amplitude of 1.5 m instead of about 1 m if η' were zero instead of one. The Einstein theory and the Brans-Dicke theory with $\omega = 5$ give a difference in amplitude for this term of about a centimeter.

Isolating the roughly 1-m term with cos 2D time dependence requires some care, since the main term of this form due to the strong perturbation of the moon by the sun has an amplitude of about 3000 km. The values determined for a, e, and x_1 are all strongly affected by the observed amplitude of the cos 2D term. In effect, the constant term in the range depends mainly on $a - x_1$, and the cos ℓ term depends mainly on $e \cdot a$. The cos 2D term depends mainly on a and is the most efficient term to use in decoupling the three parameters. However, with about a factor 5 loss in accuracy, one can regard the cos 2ℓ term as being used with the constant and cos ℓ terms to determine the three parameters. The resulting expected value of the cos 2D term can then be compared with the observed value to look for the relativistic term.

A second effect to look for has been suggested by Nordtvedt (Refs. 13-16). This is a possible difference between the gravitational mass and the inertial mass of a large body. Under Einstein's theory, the ratio is exactly unity, but in other theories it can have a slightly different value. The question raised by Nordtvedt is whether the gravitational self energy of the body enters in the same way for both types of mass. If it does not, then he predicts that the anomalous acceleration of the earth toward the sun will lead to an extra term in the earth-moon distance, with time dependence cos D. For the maximum plausible effect, where the gravitational self energy does not contribute at all to the gravitational mass, Nordtvedt gives a magnitude of roughly 24 m for this term. The scalar-tensor theory predicts an effect which is smaller than this by a factor $(\omega + 2)$. With $\omega = 5$, this would yield a term of roughly 3-m amplitude.

The experimental separation of Nordtvedt's term from other terms of the same frequency seems quite feasible. The largest such term present in the lunar motion has an amplitude of 110 km and is proportional to $a \cdot a_1$. With a small increase in the accuracy of a, the main limitation will come from uncertainty in the moon-earth mass ratio, which enters in the definition of a_1. For an uncertainty of 3×10^{-5} in this ratio, as obtained from the tracking of space probes, the corresponding cos D term can be calculated to 7 cm. The effect of uncertainties in the other parameters in the problem is even smaller. Thus, the equality of gravitational and inertial mass for the earth apparently can be checked to high accuracy by the lunar ranging experiment. The cos D term is fortunately well separated in frequency from the other terms in the problem which have to be determined.

A third way in which a departure from Einstein's gravitational theory would affect the motion of the moon has been discussed by Dicke (Ref. 17). The scalar-tensor theory with $\omega = 5$ predicts a decrease of roughly 3×10^{-13} to 3×10^{-11} per year in the gravitational constant, depending on the present mass density of the universe. This would lead to a secular deceleration of the lunar motion and an increase in the lunar distance. The total secular deceleration of the moon could be determined very accurately from the lunar range measurement in a period of 5 to 10 years (Ref. 18), but there now appears to be doubt about whether the tidal part of the deceleration can be found well enough to permit separation of the scalar-tensor theory effect. Van Flandern (Ref. 19) has recently obtained an apparent secular acceleration of the moon which is quite different from the previously accepted value derived from the work of Spencer Jones. In discussing possible sources of systematic errors in the earlier work, Van Flandern states: "In particular, corrections applied to the observations to smooth out the effects of changes in observing procedure are sufficient by themselves to alter the derived acceleration by 100%." Whether future re-analysis of the material used by Spencer Jones could give the tidal deceleration with the necessary accuracy is thus doubtful.

References

1. Alley, C. O., et al., J. Geophys. Res., 70, 2267-2269, 1965.

2. Faller, J., et al., Science, 166, 99-102, 1969.

3. Alley, C. O., et al., Science, 167, 368-370, 1970.

4. Alley, C. O., et al., Science, 167, 458-460, 1970.

5. Faller, J. E., and Wampler, E. J., Scientific American, 38-49, Mar. 1970.

6. Alley, C. O., and Bender, P. L., in Continental Drift, Secular Motion of the Pole, and Rotation of the Earth, Symposium No. 32 of the International Astronomical Union, 86-90, Edited by W. Markowitz and B. Guinot, Reidel, Dordrecht, Holland, 1968.

7. Hopfield, H. S., Radio Science, 6, 357-367, 1971.

8. Garthwaite, K., Holdridge, D. B., and Mulholland, J. D., Astron. J., 75, 1133-1139, 1970.

9. Brown, E. W., Mem. Roy. Astron. Soc., 57, 1-145, 1908.

10. Eckhardt, D. H., The Moon, 1, 264-275, 1970.

11. Baierlein, R., Phys. Rev., 162, 1275-1288, 1967.

12. Krogh, C., and Baierlein, R., Phys. Rev., 175, 1576-1579, 1968.

13. Nordtvedt, Jr., K., Phys. Rev., 169, 1014-1016, 1968.

14. Nordtvedt, Jr., K., Phys. Rev., 169, 1017-1025, 1968.

15. Nordtvedt, Jr., K., Phys. Rev., 170, 1186-1187, 1968.

16. Nordtvedt, Jr., K., Phys. Rev., 180, 1293-1298, 1969.

17. Dicke, R. H., in The Earth-Moon System, 98-164, Edited by B. G. Marsden and A. G. W. Cameron, Plenum Press, New York, 1966.

18. Alley, C. O., et al., The Application of Modern Physics to the Earth and Planetary Interiors, Proceedings of the NATO Advanced Study Institute, 523-530, Edited by S. K. Runcorn, John Wiley and Sons, Inc., London, 1969.

19. Van Flandern, T. C., Astron. J., 75, 657-658, 1970.

The Accuracy of Laser Tracking

J.-E. Blamont
University of Paris

I. Introduction

In 1968 it was proposed to ESRO to measure the coefficients of the space-time metric around the sun by tracking a heliocentric probe, with the nongravitational effects such as the solar radiation pressure and the solar wind pressure carefully taken into account (Ref. 1). The tracking would be accomplished with the use of a laser transmitter placed on the earth whose pulses would be observed from the probe.

Such a system, possibly using a CO_2 TEA laser, could provide an accuracy of better than 1 m for the distance measurement. With an accuracy of 10^{-13} for the measurement of the time, it would be possible to measure γ with an accuracy of $2 - 3 \times 10^{-4}$.

In this presentation, a brief discussion of the actual laser tracking performance of geodetic satellites is followed by a discussion of the errors inherent in the laser system, and an estimate is made of the capability expected to be available in the next few years for the tracking of a heliocentric probe.

II. The Systems of Today

A. Description

Retro-reflectors are carried on six satellites now in orbit (see Table 1), including Explorer 22, which was launched by NASA on October 10, 1964, at a 79.7-deg inclination, 1100-km apogee, and 939-km perigee.

Table 1. Satellites with retroreflectors

Satellite	Apogee, 10^6 m	Perigee, 10^6 m	Inclination, deg	Period, min	Average visual magnitude	Number of cube corners	Effective area, cm^2	Stabilization
BE-B (1964 64A)	1.09	0.89	80	105	8	360	80	Magnetic
BE-C (1965 32A)	1.32	0.94	41	108	8	360	80	Magnetic
Geos 1 (1965 89A)	2.27	1.12	59	120	8	334	940	Gravitational
D1-C (1967 11A)	1.35	0.53	40	104	8	144	20-100	Magnetic
D1-D (1967 14A)	1.85	0.58	39	110	8	144	20-100	Magnetic
Geos 2 (1968 2A)	1.61	1.08	74	112	8	400	1100	Gravitational

When the laser pulse energy E (in joules) is fired, the number of reflected photons S at 6943 Å can be estimated as in the following range equation:

$$\frac{S}{E} = \frac{1}{R^4} \frac{A_S A_R}{\Omega_T \Omega_S} T^2 \frac{10^{19}}{2.86} \frac{\text{photons}}{\text{J}} \qquad (1)$$

where

A_S = effective area of the satellite's retroreflector, m^2

A_R = effective area of the receiver, m^2

Ω_T = solid angle of the transmitted laser beam, sr

Ω_S = solid angle of the beam reflected from the satellite, sr

R = range of the satellite, m

T = atmospheric transmission (one way)

The maximum range R_{max} will be automatically derived from Eq. (1) by giving the minimum detectable number of photons, taking into account the background noise (which is determined by probability) and the possible peak output of the laser light, because all other parameters A_S, A_R, Ω_T, and Ω_S are dependent on the tracking system.

With the typical values found in Table 2 combined with $A_S = 9.35 \times 10^{-2}$ and $\Omega_S = 7.85 \times 10^{-9}$ for the Geodetic Earth-Orbiting Satellite (GEOS), the received photons at a 2000-km range and 1-J transmitted energy will be obtained as 3×10^4 photons. Atmospheric transmission, which varies widely with meteorological conditions, is taken as T = 0.7, a value for clear weather. Actually, the received photons are considerably dependent on range, irregularity of laser pulse energy, incidence angle of the light ray falling on the satellite, transverse energy distribution of the laser beam, atmospheric sway, and velocity aberration of light.

Table 2. Typical values

Laser		
Oscillation mode		"Q"-switched mode with rotating prism
Wave length		6943 Å
Peak laser power output		20 MW nominal
Pulse length		50 ns
Pulse energy		1 J
Pulse repetition rate		1-30 s, variable
Beam divergence		4 mr (without optics) 0.6-5.5 mr (with 4" optics)
Transmitting telescope		
Type		Galileo telescope
Aperture		100 mm
Focal length	objective lens	700 mm
	eyepiece lens	70 mm

B. The Errors in the System

1. GEOS-A

NASA has conducted a Geodetic Earth Orbiting Satellite-A (GEOS-A) Observation Systems Intercomparison Investigation (Refs. 2, 3). Part of this investigation consisted of side-by-side tracking tests of the GEOS-A spacecraft by the Goddard Range and Range Rate (GRARR) system and the Goddard Laser tracking system. Seventeen passes were observed from July to November 1966 by the Rosman, North Carolina, GRARR station, and ten of these were evaluated. In the investigation, the laser system tracks of the spacecraft were used as a reference trajectory for the GRARR system.

The types of tracking stations which participated in GEOS-A are listed in Table 3.

a. Laser System. The laser tracking data consist of one measurement per second of range in nanoseconds and elevation and azimuth in degrees, punched on paper tape. Using a preprocessing program, the following corrections are added to the whole-second time for each measurement to give the time at the spacecraft:

WWV correction (3.6 ms for the Rosman station).

Delay time between the 1-s pulse and actual laser firing.

One-half the round-trip interval of the laser beam.

An internal delay correction due to photomultiplier, cables, and receiving telescope (90-ns round trip) is made to each range measurement when it is converted from a time interval into meters.

b. GRARR System. The GRARR system is a high-precision spacecraft tracking system that determines range using the sidetone ranging technique, and range rate applying the principles of coherent doppler. Angular data are obtained from X/Y-mounted antennas but are not used for orbit determination. Each GRARR station uses an S-band system and a VHF system in conjunction with a multichannel transponder on the spacecraft being tracked. Only the S-band system was used for this evaluation.

Data, at one measurement per second, consisting of the range in meters, range rate in meters per second, and X and Y angles in degrees, were used for this evaluation. In the operational preprocessing, the times to the spacecraft are corrected and a constant transponder bias

Table 3. GEOS-A tracking systems

	Intercomp. net		World net		Frequency, Mc/s		Data measures	Estimated accuracy
	Min.	Max.	Min.	Max.	Up	Down		
Radio systems								
Army/SECOR	1	4	1	9	420.9	244.5, 449.0	R	10 m
Navy/TRANET	4	7	17	21	None	162, 324, 972	$\dot{R}$	5 cm/s
NASA/Minitrack	3	6	9	12	None	136.8	A	20 arc-s
NASA/GRARR	1	1	3	3	2271	1705	R and $\dot{R}$	10 m and 5 cm/s
Total	9	18	30	45				
Lasers								
NASA/laser	0	1	0	1	Opt. freq.		R	2 m and 1 cm/s
SAO/laser	0	1	0	1	Opt. freq.		R	2 m
Total	0	2	0	2				
Cameras								
Air Force/PC 1000	0	14	1	14	Opt. freq.		A	1 arc-s
NASA/STADAN MOTS	5	7	11	14	Opt. freq.		A	1 arc-s
NASA/SPEOPT MOTS	8	12	8	12	Opt. freq.		A	1 arc-s
SAO/Baker Nunn	4	7	11	14	Opt. freq.		A	2 arc-s
ESSA (C&GS)/BC-4	1	8	1	8	Opt. freq.		A	2 arc-s
International/Opt.	0	0	12	15	Opt. freq.		A	1 arc-s
Total	18	48	44	77				

correction equivalent to 3677 ns is made for each range measurement before submission to the Data Center at Goddard Space Flight Center (GSFC).

c. Summary of Results. Data were obtained from colocated laser and GRARR systems at Rosman for 17 passes. Tables 4 and 5 give a summary of the laser and GRARR results, respectively. They are presented as an example of data actually obtained. The average bias error for the Rosman GRARR relative to the laser was found to be -5.3 ±2.5 m. The random noise after removal of the bias was 6.8 m. The average difference between the laser and the GRARR range timing was -2.1 ±1.2 ms. After an appropriate error analysis, it was shown that the average difference between the laser and the GRARR range rate timing was only -0.20 ±0.2 ms.

d. Analysis of Errors. In order to investigate further the sources of error in the GRARR system, three passes were chosen for closer examination. The average values for these three passes were:

Range rms, m	Range bias, m	Range rate rms, m/s	Range time difference, ms
6.2	-4.2 ±2.0	2.6	-2.31 ±1.50

e. Laser Errors. For the three laser passes used for GRARR error modeling, the laser range residuals had a mean value of -0.002 m and a standard deviation of 1.5 m. A chi-square test of normality was run on the range residuals, and none of the three sets were found to be significantly abnormal, although a slight skewness was noticed in each of the three data sets. However, the residuals appeared to be random and if any systematic effects were present, they were quite small.

Tests have been made for serial correlation in the laser data. The results showed the serial correlation to be insignificant. Using the assumption of independence, the eigenvalues of the covariance matrix of the velocity parameters X, Y,

Table 4. Summary of laser data for GEOS-A intercomparison study

Run. No.	Date	Pass duration, s	Range rms, m	Azimuth rms, mrad	Elevation rms, mrad
1	July 27	392	27.1[a]	0.34	1.91
2	July 28	252	17.1[a]	0.17	0.11
3	July 29	No overlap	2.1	0.20	0.15
4	Aug. 10	96	4.3	0.14	0.09
5	Sept. 9	345	2.1	0.27	0.16
6	Sept. 10	304	1.5	0.20	0.07
8	Oct. 6	205	1.7	1.07	0.14
9	Oct. 7	377	1.6	1.11	0.24
10	Oct. 8	344	1.2	0.26	0.13
11	Nov. 15	174	1.6	0.18	0.13
12	Nov. 18	305	1.0	0.31	0.07
13	Nov. 19	153	2.3	0.47	0.14
14	Nov. 19	No overlap	1.4	0.32	0.12
15	Nov. 20	416	1.6	0.69	0.30
16	Nov. 20	336	1.1	0.49	0.54
17	Nov. 21	423	1.6	0.51	0.24
Average			1.8	0.42	0.28

[a]Not included in average

and Z can be used to get an estimate of the accuracy with which range rate can be determined using orbits defined by the laser. It was calculated that range rate would be no more accurate than 1 cm/s using the laser as a standard.

f. Conclusions of GEOS-A Analysis. The GRARR range bias of -5.3 ±2.5 m and the range timing error of -2.1 ±1.2 ms could be explained in part by inaccuracies in the transponder delay curve.

From this intercomparison experiment, it appears that laser orbits can be used to detect systematic errors in both the range and the range rate to about 2 m and 1 cm/s, respectively. After the laser data were edited, systematic effects therein seemed to be at a minimum.

2. GEOS-2

As part of the GEOS Observation Systems Intercomparison Investigation, several of the geodetic satellite tracking systems used with GEOS-2, including a NASA Laser, an Army Secor, and a Navy Tranet, were moved to the NASA Wallops Island station and located near the FPQ-6 and FPS-16 C-band radars there (Ref. 4). GEOS-2 was simultaneously tracked by all of these systems during April, May, and June, 1968, to compare the tracking data without the effects of uncertainties in survey, gravity field, and system time synchronization.

Six separate short-arc laser reference orbits were derived from the laser range, azimuth, and elevation data for the first six simultaneous passes. In deriving these reference orbits, the laser data were assumed to be unbiased, and were combined so as to minimize the weighted residuals in a least-squares sense. The a priori estimates and relative weighting factors for the laser range and angle residuals were 2 m and 1 mrad (206 s of arc), respectively.

Range, range rate, azimuth, and elevation values were computed from the laser reference orbits and compared with the observations from each of the other tracking systems. Comparison of tracking data from the radio tracking systems with the laser reference orbits yielded residuals from which zero-set and timing biases were derived for each system. The preliminary results for the six passes reported indicate that the Secor had consistent, unexplained, negative zero-set biases, averaging about 12 m. The two C-band radars generally agreed with the laser to within ±5 m.

Table 5. Summary of GRARR data for GEOS-A intercomparison study

Run No.	Date	Range rms, m	Range bias, m	Range rate rms, cm/s	Range time difference, ms
4	Aug. 10	12.9	-5.6 ±1.0	No range data used	-1.40 ±0.77
8	Oct. 6	6.4	8.7 ±1.3	21.6	-3.75 ±0.52
9	Oct. 7	6.5	10.6 ±0.8	5.4	-3.28 ±0.23
10	Oct. 8	6.1	-3.6 ±0.8	3.3	-4.04 ±0.28
11	Nov. 15	5.6	-4.1 ±2.2	3.3	-1.73 ±1.27
12	Nov. 18	6.1	-35.2 ±1.0	8.6	-0.82 ±0.42
13	Nov. 19	5.7	-4.7 ±1.5	9.9	-0.77 ±0.11
15	Nov. 20	6.2	-2.6 ±0.6	2.0	-1.47 ±0.17
16	Nov. 20	6.2	-10.1 ±1.1	5.8	-2.02 ±0.44
17	Nov. 21	6.2	-6.5 ±0.7	2.4	-1.41 ±0.17
Average		6.8	-5.3 ±12.4	6.9	-2.07 ±1.19

3. Improvement of the SAO Laser Ranging Accuracy

The returns from the ruby-laser system at the Smithsonian Astrophysical Observatory's (SAO) Mt. Hopkins Observatory exhibit a variation in signal strength that is typical of laser systems (Refs. 5, 6). The variation is due in part to the fact that the signal varies inversely with the fourth power of the satellite range, and also to an observed "scintillation," or random variation in returns from the same satellite range, as illustrated in Fig. 1. The range measurements are affected by this variation in signal strength when they are obtained directly from a time-interval counter that is started by the transmitted pulse and stopped by the received pulse. The resolution of the counter is 1 ns, which corresponds to 15 cm, but the duration of the pulse is 18 ns (full width between half-power points). Consequently, counter readings corresponding to a given range can vary significantly if the counter stops at different points on the pulse's leading edge. A correction, which is the time between the intersection of the counter's "stop" threshold and the centroid of the pulse, is obtained from oscilloscope photographs of the return pulses. By applying such corrections to the readings of the time-interval counters, the instrumental errors of the ruby-laser systems of SAO can be reduced from 1 - 2 m to 20 - 50 cm.

The French group, using the GEOS and DIADEME satellites, has reached the same conclusions as the SAO. This group has observed errors of 1.5 m and estimates that an accuracy of 20 cm can be achieved by refining the technique.

It is concluded, therefore, that the accuracy of laser tracking, currently 1 to 2 m, will be improved to 20 cm in a period of 2 to 3 years; it is already better than any other tracking technique and is used for calibration of other tracking systems.

III. Errors in the Tracking of a Heliocentric Probe

A. Errors Due to the Telluric Atmosphere

1. Variation of Refraction Index

Since the refraction index of the atmosphere n is a function of the altitude h, $n = n_0 e^{-d(h-h_0)}$ (d is a constant ≈ 0.1385/km), the difference between the optical path through the atmosphere and the optical path through vacuum ΔR is for an elevation angle β:

$$\Delta R \simeq \frac{10^{-6} n_0}{\alpha \sin \beta} \sim 2.1/\sin \beta \text{ (in meters)}$$

Then, for useful elevation angles ($\beta > 60$ deg), this correction is about 2.5 m.

2. Bending of the Beam

The path length X is increased by the value $\Delta R \simeq X/\cos \theta \, (\Delta\theta/2)^2$

$$\Delta\theta = (n - 1) \operatorname{tg} \theta$$

where θ is the angle between the ray and the normal to the earth's surface, $\theta = 90 - \beta$. $\Delta\theta$ is of the order of 10^{-5} to 10^{-4} rad and has to be taken into account in order to strike the target but introduces no error in the tracking data.

3. Atmospheric Absorption

The laser beam, passing through the atmosphere, undergoes a certain amount of extinction due to scattering by water droplets. The following

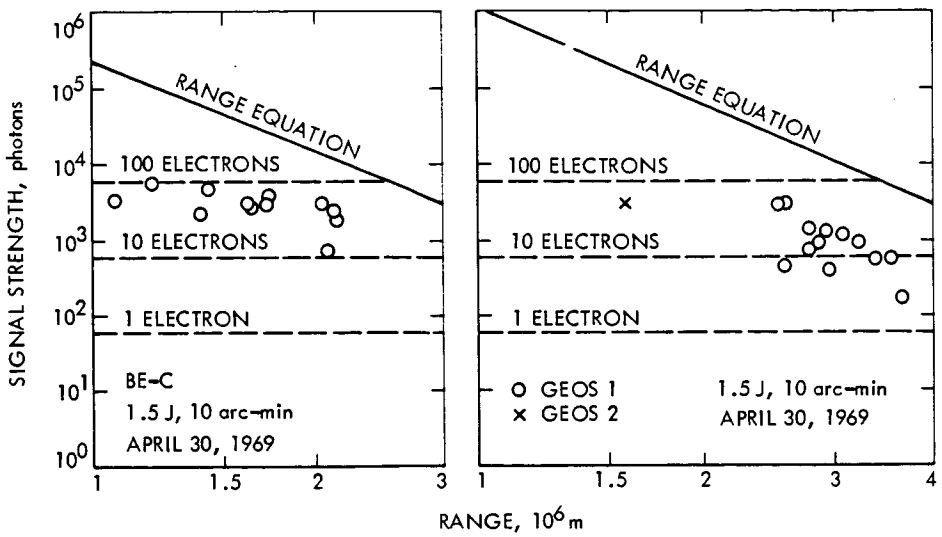


Fig. 1. Laser returns

expression for the atmospheric transmission Δ is given by

$$\Delta = \exp(-r/\cos\theta)$$

where θ is the angle the beam makes with the normal and r is given by

$$r = 0.009\,\lambda^{-4} + 0.223$$

where λ is the wavelength in microns, and the first term represents the scattering by gas molecules, the second one the scattering by water droplets.

The transmission Δ as a function of the elevation angle β_0 is shown below for the ruby laser transmitting at 6963 Å:

β_0, deg	Δ
10	0.275
20	0.515
30	0.635
40	0.705
50	0.745
60	0.77
70	0.79
80	0.795
90	0.8

If we consider a useful angle for the measurements of the order of 60 deg (i.e., ±30 deg around the normal to the ground, which allows about 4 h per day measuring time), we see that the transmission is larger than 0.75.

4. Atmospheric Turbulence

Air temperature varies from point to point in the atmosphere in a random way and is a function of altitude and wind speed. The index of refraction of the air depends on its temperature. Therefore, when a light beam passes through an air region in which there is a temperature change, the beam is partially or totally deviated, depending on the relative dimensions of the beam and of the inhomogeneous region. The main effects of turbulence are:

(a) Beam steering: The random angular deviation of the beam (as a whole) from the line-of-sight path, which can cause the beam to miss the receiver.

(b) Beam spreading: Small angle scattering, which increases the beam divergence and causes a decrease in spatial power density at the receiver.

(c) Beam scintillation: Small-scale destructive interference within the beam cross section, causing variations in the spatial power density at the receiver.

To see which of these effects is predominant, we must look at the ratio between the beam diameter b and the inhomogeneity dimension l. If $b/l \ll 1$, the major effect of the turbulence is to deflect the beam as a whole (beam steering). This is the case of communications from an earth transmitter and a deep-space receiver.

The theory of Tatarski gives the following formula for the corresponding rms angle of deviation:

$$\overline{(\Delta\alpha)^2} = 2.91\, b^{-1/3} \sec\theta \int_0^\infty C_n^2(Z)\, dZ$$

where b is the beam diameter, θ is the angle with the zenith, and C_n is the structure constant of the turbulent medium.

The integral

$$\int_0^\infty C_n^2(Z)\, dZ$$

has been evaluated in the case of medium turbulence. It was found to have a value of 1.3×10^{-11} $m^{1/3}$. The value of $\Delta\alpha$ as a function of the elevation angle β_0 is shown below:

β_0, deg	$\Delta\alpha$, rad
10	19.4×10^{-6}
20	13.8×10^{-6}
30	11.4×10^{-6}
40	10.1×10^{-6}
50	9.2×10^{-6}
60	8.7×10^{-6}
70	8.35×10^{-6}
80	8.15×10^{-6}
90	8.1×10^{-6}

The calculations have been performed for a beam diameter of 0.2 m, corresponding to a beam aperture of 5×10^{-5} rad at a 4000-m height (where more or less maximum turbulence occurs).

In the case of strong turbulence, the value obtained above must be multiplied by a factor of 2, whereas in the case of weak turbulence they must be divided by a factor of 6. We see that for useful elevation measuring angles ($\beta_0 \geq 60$ deg) and for a laser beam with an initial divergence of 5×10^{-5} rad, the steering effect is negligible, while strong turbulence can cause a deviation of the beam from the line of sight of the order of 10% of the beam divergence.

5. Atmospheric Distortion of Short Laser Pulses

Only very short laser pulses (~30 ps) may experience noticeable time dispersion when passing through the atmosphere, so we can neglect this effect in our case.

B. Errors Due to Coronal Effects

When a laser beam passes near the sun (for instance at 1 or 2 sun radii), it can undergo a certain amount of dispersion and scattering due to the presence of plasmas in the corona region. Some calculations have been performed to take into account these effects. It is found that a gaussian pulse of 10-ns duration passing at a distance of 1 sun radius from the surface of the sun undergoes a delay of 6×10^{-4} ns and a broadening of $\sim 10^{-10}$ ns due to the dispersion introduced by the coronal plasmas. These effects of the dispersion can be completely neglected, and the effect of scattering is also negligible. Indeed, the RMS angle deviation ϕ_0 is of the order of 2×10^{-4} rad, and the intensity fluctuations

$$\left\langle \left(\frac{\Delta I}{I}\right)^2 \right\rangle^{1/2} = \sqrt{2}\,\phi_0$$

are of the order of 3×10^{-4}.

IV. Possible System

The CO_2-N_2-He laser transmits at the wavelength $\lambda = 10.6\ \mu$; it can be operated either as a CW or as a pulsed transmission. In the continuum mode, it has achieved a power of 10 kW. Recently, with the use of short pulses at low repetition rates, it has been found that excessive gas heating can be avoided. A transversally excited CO_2 atmospheric pressure gas laser (TEA) has achieved the following performances (Ref. 7): efficiency 17%, 2 J per pulse, repetition rates 1000 pulses/s, half-power pulse 300 ns. Furthermore, Dumanchin, et al. (Ref. 8), have achieved a peak energy (for one giant pulse) of 650 MW for a 0.2-μs pulse. It is estimated that a CO_2 TEA laser should be able to achieve the following performances within 2 to 3 years: power 200 MW, repetition rates 1 to 10 kHz, pulse duration (by mode locking) 1 ns.

The advantage of the CO_2 laser for long-distance transmission lies in the fact that it is diffraction-limited; therefore, it can be easily matched to any diffraction-limited system. Placed at the Coudé focus of a 1-m telescope, which can easily be built diffraction-limited, it would have a beam aperture of 3 arc-s. The transmission factor of such a beam through the terrestrial atmosphere is 70%.

Such a system would provide a signal of about 10^{-6} W at a distance of 2 AU, as compared with a signal given by the sun of 10^{-8} W. This signal could easily be detected by a state-of-the-art (Hg Cd Te) nitrogen-cooled photovoltaic detector with a detectivity of 2×10^{10} cm $Hz^{1/2}/W$ dual response time of 2 ns.

However, it can be remarked that even with a ruby laser of today, the tracking of the probe would be possible with a small error. With a duration of 10 ns, a repetitive frequency of 1 pulse/min, an energy of 1 J (peak power of 100 MW), and a beam divergence of 5×10^{-5} rad, the signal would be received with a 30-cm-diameter Cassegrain telescope (f/2), a 10-Å-bandwidth filter, and a 2- to 4- ns rise time of the receiver.

For both systems, the major errors would be:

(a) Uncertainty of the knowledge of the speed of light: We have to limit ourselves to deviations from an average orbit smaller than 1000 km, which would introduce errors of ±1 m.

(b) Refraction index of the atmosphere: This introduces an error of 2.5 m which can be known at 10%; then the resultant error is 0.25 m.

(c) Indetermination of the measurement of the time of arrival of the pulse due to turbulence: In 95% of the cases the error would be ±8 ns or 2.4 m for the ruby laser or 0.2 m for the CO_2 laser. The total error would be 2.6 m for the ruby laser or 0.2 m for the CO_2 laser.

This is obtained without any integration method; thus, ordinary electronic circuits may be used.

Therefore, it is concluded that a system using a laser situated on the earth and a small telescope on a probe would provide an accuracy at least as good as any tracking system in the S- and X-bands.

The major advantage introduced by the use of a laser tracking system would be the possibility of obtaining measurements down to a few minutes of the solar limb, over the part of the orbit of an heliocentric probe where the effect of the gamma term is large.

References

1. Blamont, J., Proposition pour la détermination des coefficients du tenseur métrique de l'espace-temps à partir d'une sonde spatiale héliocentrique, Internal Report, Service d'Aéronomie No. 224, June 15, 1968.
2. Berbert, J. H., Intercomparison of GEOS-A Observation Systems, G.S.F.C. Technical Note X-514-67-315, 1967.
3. Berbert, J. H., Maresca, P., Norris, P., and Reich, R., Intercomparison of Collocated Laser Optical and GRARR Radio Ranging System Tracks on GEOS-A, G.S.F.C. Technical Note X-514-67-447, 1967.
4. Berbert, J. H., and Parker, H. C., Some Comparisons of Secor, Tranet, FPQ6 and FPS16 With Laser Tracking, G.S.F.C. Internal Note, 1968.
5. Lehr, C. G., Pearlman, M.R., and Scott, J. L., A Photographic Technique for Improved Laser-Ranging Accuracy, S.A.O. report to the GEOS-2 Review Meeting, G.S.F.C., June 1970.
6. Lehr, C. G., Pearlman, M. R., and Scott, J. L., Range Corrections From Oscilloscopic Displays of Laser Returns, S.A.O. Laser Report No. 4, August 1970.
7. Beaulieu, A. J., Appl. Phys. Letters, 16 (12), 504, 1970.
8. Dumanchin, R., Forcy, J. C., Michon, M., and Rocca Serva, J., Paper No. 18-6, 6th International Quantum Electronics Conference, Japan, Sept. 1970.

Drag-Free Satellite Control System Technology

Daniel B. DeBra
Stanford University

I. The Drag-Free Principle

Satellites are disturbed by external forces such as radiation pressure, atmospheric drag, micrometeorite impacts, etc. A proof mass located inside the satellite is shielded from these external forces and will have a purely gravitational orbit if the satellite does not bump it or disturb it. If the satellite can be propelled in response to an error signal generated when it approaches the proof mass, the satellite will follow the same orbit as the proof mass. The orbit of the proof mass is drag-free; hence, the satellite is called a "drag-free satellite."

Although the idea had been proposed previously (Ref. 1), it was independently conceived in 1961 by Lange (Ref. 2), who performed the first detailed analysis of drag-free satellite performance and control mechanization (Ref. 3). As a result of his studies and as part of a program to perform a gyro test of relativity (Ref. 4), the development of drag-free technology was started in 1964. Since that time, a number of applications for drag-free operation have been proposed and studied in detail. One of these, the Navigation Mission, has been funded, and Stanford University has designed, and is building, the drag-free control system for the Johns Hopkins Applied Physics Laboratory as part of the Transit Navigation Satellite Improvement Program. The most ambitious drag-free experiment to date is designed for a heliocentric relativity experiment (Ref. 5), and Juillerat presents a thorough error analysis in this proceedings.

In this paper, some of the technology developed in building a drag-free satellite simulator for laboratory use will be described, the design decisions made in order to achieve a 10^{-11} g perturbation level for the Navigation Satellite will be discussed, and the control system development that will make possible drag-free operation of spinning satellites to reduce perturbation levels by averaging will be described.

II. Laboratory Simulator

In 1964, the first version of a laboratory drag-free satellite simulator was completed. There have been many modifications of this vehicle since that time, and it continues to serve as a test bed for new drag-free satellite technology development. The simulator is supported on an air film, with very low friction in the two lateral directions. For typical simulator velocities, the friction level is much smaller than the uncertainties in the levelness of the table on which it rides.

An automatic table leveling system was designed and built to permit simulation corresponding to altitudes of up to 275 km (Ref. 6). This table holds a level reference with an accuracy of 0.1" for periods of 12 h during which experimental work can be performed. The limitation in simulating an orbital environment in the two horizontal directions is primarily the asymmetry of the gas flow in the bearing (this is different from the drag due to relative motion between the vehicle and table and is quite sensitive to the location of the mass center of the vehicle with respect to the bearing), and the air currents in the room.

In the simulator, the proof mass is supported over the table. The simulator moves with respect to it much as the satellite will move with respect

to the reference proof mass that will be in free fall in orbit. The simulator is perturbed by tilting the table, and the control system must provide the thrust to hold the simulator centered, just as the orbiting satellite must react. The degree of rotational freedom available is used to simulate spinning satellites.

There are two types of problems associated with drag-free satellite design. The first are the perturbing forces that act on the proof mass due to the presence of the satellite. The second type are the design problems of controlling the satellite so it will not touch the proof mass. For illustration, the Navigation Satellite proof mass perturbations and recent control system developments in the mechanization of spinning vehicles will be discussed.

III. Navigation Satellite Design

The TRIAD I configuration is shown in Fig. 1. The satellite is composed of three main bodies: the power section at the top, the main electronics for the navigation function at the bottom, and the disturbance compensation system (DISCOS), which is the drag-free control system, at the center. This three-part design not only provides gravity stabilization for the satellite, but keeps the major portion of the satellite mass sufficiently far away from the proof mass so that the accuracy of calculating and compensating for mass attraction is considerably relaxed.

The principal proof mass perturbations are the interactions of the DISCOS with the proof mass. The largest of these is mass attraction. Figure 2 shows the layout of the DISCOS. The

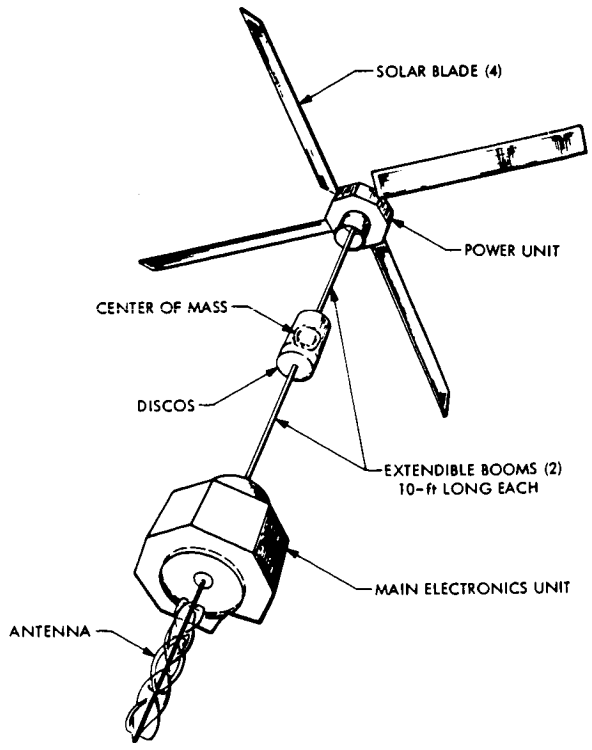


Fig. 1. TRIAD I configuration

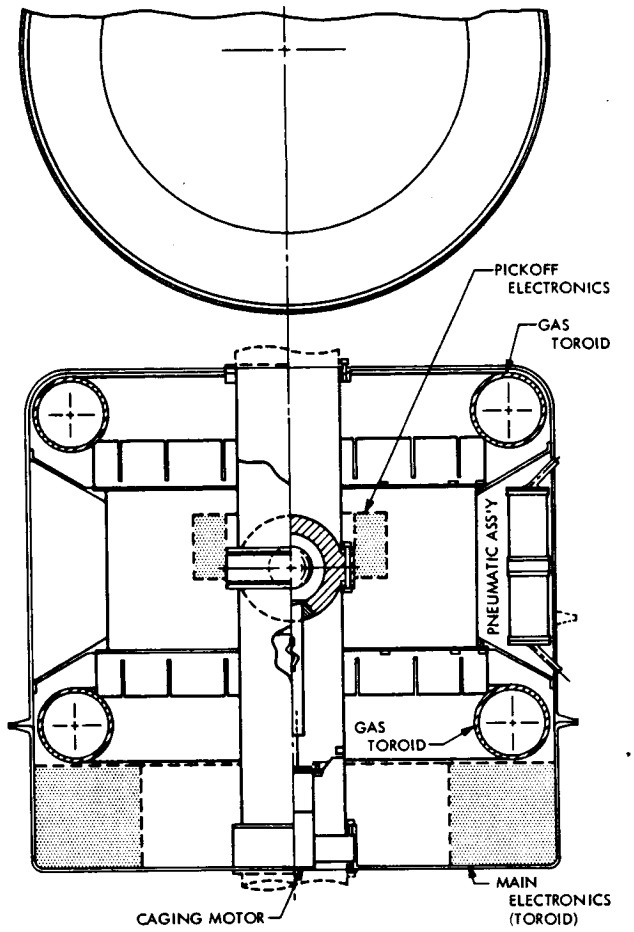


Fig. 2. Layout of DISCOS

positioning of individual masses is critical, and each mass must be weighed accurately and located precisely. The masses closest to the proof mass have a much greater effect and are more sensitive to errors in weighing measurement and location than those farther away. The error budget for mass attraction was divided, with a weighing factor of m/r^2 for each part of the system, where m was the mass of the part and r was an effective average distance. For example, a ring around the proof mass would use the radius of the ring even though the net force is 0 by symmetry because the sensitivity to errors is related to the uncertainty in the individual particles of the ring, and each of these has the ring radius.

Fixed masses can be measured and later compensated. The mass of the propellant, however, is time-varying, and its effect must be compensated at all times. Therefore, a cold-gas propellant was used which distributes itself uniformly through two toroidal tanks. The symmetry of these tanks with respect to the proof mass ensures that the net force from the propellant is always 0. Because there is relative motion of the satellite with respect to the proof mass within a deadband which is introduced to ensure efficient use of the propellant, the gradient in the mass attraction field is also important. The attractive force along the axis of symmetry of the toroid increases at first from zero at the center of the toroid to a maximum, and then decreases approximately

r^{-2} as one gets far enough away from the toroid to treat it as a particle. At the point of maximum force, the slope of the force curve with respect to distance is zero. The derivatives of the force with respect to lateral displacements must be equal to each other by symmetry, and the trace of the gradient must be zero by Laplace's equation. Hence, independent of the amount of propellant in a toroidal tank, the gradient in the mass attraction field is zero at this point. The dual toroidal tank also provides zero mass attraction, but a weight penalty results because cold gas is most efficiently stored in a spherical propellant tank. In this design, the principal perturbing force on the proof mass is the mass attraction of the satellite. Using state-of-the-art machining tolerances, the knowledge and measuring techniques available for obtaining mass properties including density homogeneity, the mass attraction perturbation can be made as small as 10^{-11} g and the gradient in that mass attraction kept smaller than 10^{-11} g.

A number of methods for sensing the relative position of the satellite with respect to the proof mass were considered. A capacitive pickoff was chosen, and with careful design, the maximum electrical field forces — due to the imbalance in the bridge when the satellite moves relative to the pickoff — can be kept smaller than 10^{-12} g, with a ball-to-wall gap of 9 mm.

The proof mass material was selected to minimize the effects of surface forces and perturbations produced by gradients in the magnetic field. The acceleration due to surface forces is minimized by using the highest density possible. The effect of a gradient in the magnetic field is minimized if the magnetic susceptance is made small. An alloy of 30% platinum (which is paramagnetic) and 70% gold (which is diamagnetic) gives a density greater than 20 g/cm^3 and a susceptance of less than 5×10^{-8} cgs units. An etched sample of this alloy measured by the National Bureau of Standards had a magnetic susceptance of 10^{-8} cgs units. (The specification level of five times this amount was based on the assumption that a magnetic moment of 100 pole cm might be located approximately 12 cm from the proof mass.) With this material, the magnetic gradient forces are considered to be negligible for the Navigation Mission.

Temperature differences within the DISCOS produce perturbations in several ways. First, the mass of the satellite is redistributed as a consequence of thermal distortions, and a change in the mass attraction of the proof mass results. The gaseous propellant is most sensitive, requiring a temperature gradient of less than 2°C across the propellant tanks to keep perturbations to 10^{-12} g. A temperature differential across the pickoff housing produces both a radiation pressure and a differential gas pressure in the chamber. To minimize this effect, beryllium oxide was selected as the insulating material for the housing to maximize the thermal conductivity. While this has been an excellent choice in regard to ensuring good mechanical stability and stable electrical properties, the density inhomogeneity of ceramic materials is undesirably large. We have been unable to obtain material with better than 0.3% homogeneity.

IV. Spinning Vehicle Translation Control

The perturbation levels described above are, in general, body-fixed and relatively slow-changing. In drag-free satellite experiments, where the sensitivity to perturbations is not isotropic, it is possible to spin the satellite and significantly reduce the average perturbing forces in the plane of spin. The spin axis must be perpendicular to the sensitive direction(s); for example, in navigation and geodesy, the sensitive direction is along the orbit, and the vehicle may be spun perpendicular to the orbit plane for a circular orbit about the vertical. The averaging obtained is effective for body-fixed forces. However, a systematic satellite displacement to produce an error signal, coupled with the force gradient in the satellite, may produce a perturbing force in a space-fixed direction. Furthermore, some forces — for example, temperature gradient effects — may be created in a way which is systematic with respect to the environment producing them. The following discussions describe the translation control system development that has been done to permit operating a spinning satellite drag-free and simultaneously to cancel the forces produced by systematic error signals that couple with the gradient. These techniques do not correct for systematic perturbations due to temperature differentials or other effects which act directly on the proof mass.

A. Trapping

Powell has discovered a nonlinear phenomenon in spinning vehicle control systems which he refers to as "trapping" (Ref. 7). Figure 3 shows the conditions for trapping in a controller with deadband. When the thrust due to the actuating error signal (which is the sum of position error and velocity) is parallel to the position error of the vehicle with respect to the mass center, an equilibrium can exist in which the thrusters continually provide the force to produce the centripetal

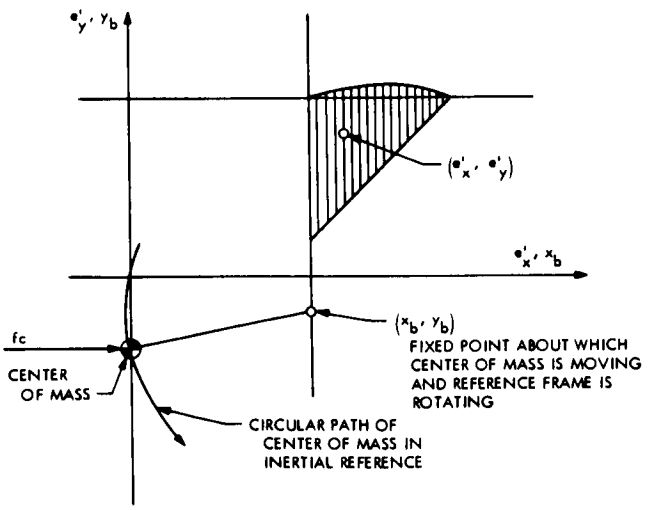


Fig. 3. Control direction for point in trapping region (from Ref. 7)

acceleration required to maintain the mass center in a circular trajectory. Figure 4 shows the increase in propellant expenditure due to trapping compared with normal propellant use due to limit cycling in a disturbance-free environment.

The trapping effect depends on the shape of the deadband of the controller and on the location of the mass center of the satellite with respect to the pickoff null. Powell developed two techniques for minimizing the trapping effect. First, he devised a round deadband which decreases the sensitivity to trapping significantly (Fig. 4). Secondly, he developed a mass center estimator, which determines the location of the mass center from the dynamic behavior of the satellite. The null of the sensor is then electrically shifted to the observed mass center.

The experimental results using a mass center estimator are in good agreement with the predicted results shown in Fig. 5. The figure also shows the effectiveness of a mass center estimator in maintaining the pickoff null at the mass center within the minimum displacement before trapping occurs. The estimator accuracy was 25 μm (0.001 in.), with a gap of 2.0 mm and a 5-cm-diameter ball.

The shape of the deadzone in the plane of the spin determines the sensitivity of trapping. With a circular deadband, the displacement of the mass

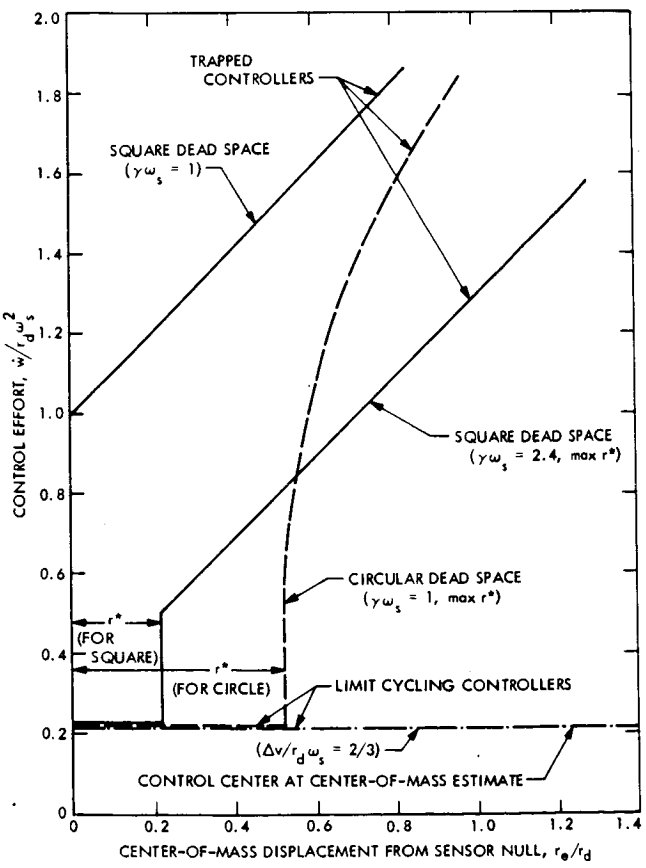


Fig. 4. Control effort vs center of mass ($\Delta v/r_d\omega_s = 2/3$; from Ref. 7)

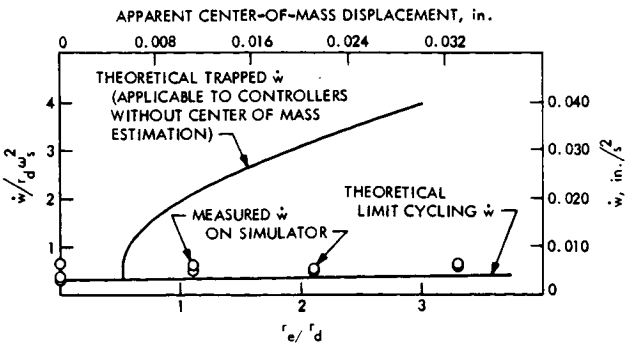


Fig. 5. Control effort vs apparent center of mass with center-of-mass estimation (from Ref. 7)

center with respect to the null point must not be greater than one-half of the deadband, or trapping will occur. This ratio decreases as the number of sides of a polygonal shape approaches a square. For a square, it is theoretically possible to get trapping even though the mass center is exactly aligned with the pickoff null point. For a three-sided deadzone, trapping is ensured in spite of perfect alignment of the null and center of mass.

Special deadbands and mass center estimation are needed only when the spin rate must be "high." The rotation rate ω is high enough to influence the behavior; i.e., when $\Gamma_d\omega^2$ represents a significant amount of total propellant expenditure, where Γ_d is the deadspace radius.

For values of spin rate that are sufficiently small, the control system may be considered to operate as it would if the orientation were fixed with respect to inertial space.

Figure 6 shows a typical response of the mass center estimator of the simulator following a

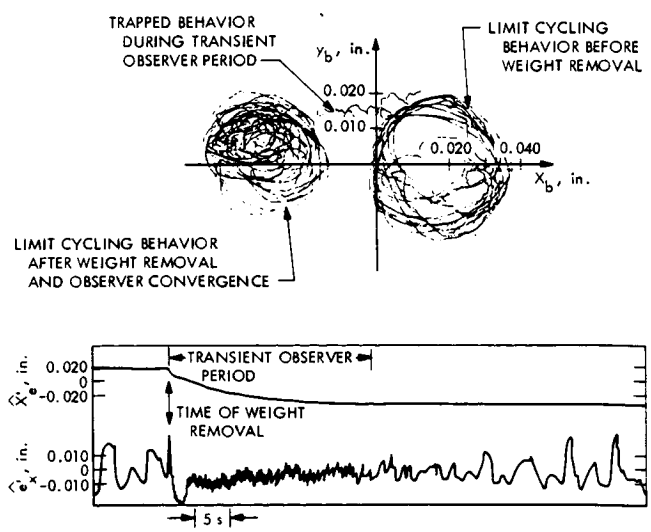


Fig. 6. System response to center-of-mass shift (from Ref. 7)

change in the mass center location. The upper curves are the relative motion of the vehicle with respect to the proof mass as shown in body axes. The circular nature of these trajectories is typical in the presence of an external disturbing force which requires an error signal biased on one side of the deadzone with respect to inertial space. The relative motion therefore appears to be a circle in body axes. The axes shown have an arbitrary origin that is fixed in the vehicle. The null point of the controller is located at the estimated mass center, which is approximately at the center of the circular trajectories. The upper figures show the behavior before the mass center is relocated, with the relative motion trajectories typically nearly circular. The movement of the mass center due to the addition of a weight causes trapping to occur, which continues during the 20 to 30 s required for the mass center estimator to determine the proper location for the pickoff null. After this transient period, the vehicle resumes its limit cycle behavior, with circular trajectories centered at the new center-of-mass location. The lower part of the figure shows the time history at the mass center estimator and the error signal which activates the propulsion system. Figure 7a presents a typical scale factor calibration for the mass center estimator, showing a stability and resolution of approximately 0.001 in. (25 μm).

The practical realization of a mass center estimator must be accomplished carefully and requires relatively high-quality electronics. Figure 7b shows a comparison of an estimator mechanized using an analog computer, with the successful circuit design that was developed for the laboratory simulator.

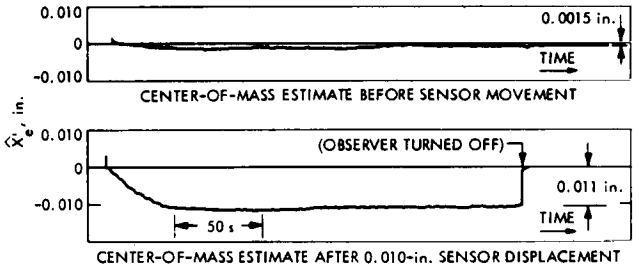


Fig. 7a. Verification of analog-observer center-of-mass estimation accuracy (from Ref. 7)

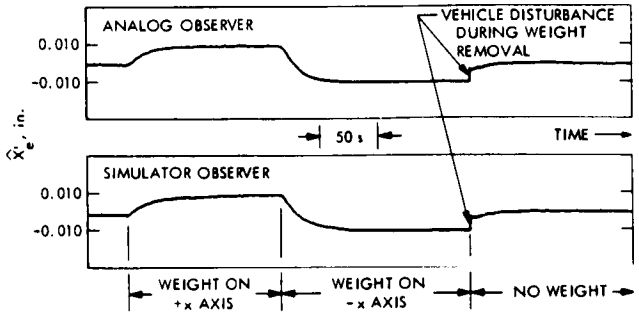


Fig. 7b. Comparison between analog and simulator center-of-mass estimates (from Ref. 7)

B. Integral Control

Spinning vehicle control is capable of averaging any body-fixed forces in the plane of spin. There are two other ways in which disturbances reach the proof mass. One is the coupling of the relative motion of the vehicle with respect to the proof mass and the gradient to produce systematic disturbing forces in the same direction as the external forces that are perturbing the vehicle. The second is a class of disturbances which act directly on the proof mass. The thrust to compensate an external force requires an actuating signal which is consistently in a direction to cancel the force. This signal does not have to be produced by an error signal from the pickoff. By the use of integral control, one can produce an error signal which is maintained as long as the external force does not change. To do this on a spinning vehicle, it is necessary to transform the error signal into a space which is relatively fixed with respect to the source of the disturbance, perform the integral control (or estimation of the disturbance force), and transform the results back into the error signal space. If no integral control is used, compensating an inertially fixed disturbing force requires an error signal which varies sinusoidally at spin frequency in body axes with fixed amplitude. With integral control, a stepped input in external disturbance results in an initial sinusoidal response, but the amplitude decays as the integral reaches a steady-state value which is capable of compensating the external forces. This effect is shown in Figs. 8a and 8b (from Ref. 6).

V. Conclusions

Some perturbations act directly on the proof mass without depending on the mechanization of the control system or on whether the vehicle is spinning or not. For example, a gradient in the

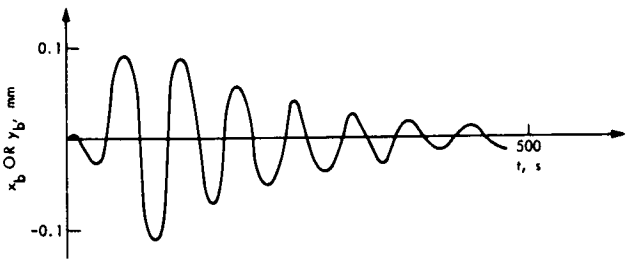


Fig. 8a. Plot of x_b/y_b vs t; low K_c; not enough damping (from Ref. 6)

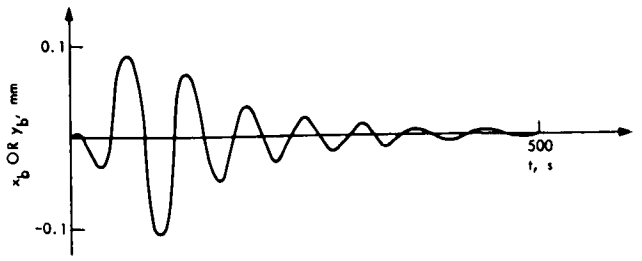


Fig. 8b. Plot of x_b or y_b vs t; optimal K_c (from Ref. 6)

magnetic field of the earth would act directly on the proof mass in proportion to its susceptance. Similarly, a temperature gradient induced across a spinning vehicle remains fixed in the vehicle with respect to the heat source rather than being fixed in the vehicle axes which are rotating. As a result, radiation pressure and thermally induced gas pressure differences are not influenced by the spin averaging or integral control except insofar as the spin speed may be fast compared to the thermal relaxation time of the vehicle, which would cause a reduced gradient at higher spin speeds.

Though mass attraction is the limiting factor for nonrotating vehicles, the smaller effects which act directly on the proof mass in a rotating vehicle become the limiting factors in how successfully a vehicle may be made drag-free. Even if the effects were comparable to the state of the art in predicting mass attraction compensation, there is no practical way of testing to see if the predicted level has been achieved prior to flight; hence, spinning may be desirable as insurance against errors. Furthermore, the mass attraction levels may be relaxed considerably (say, by a factor of 30) without loss of ultimate drag-free performance if full advantage is taken of the averaging due to spin. Table 1 (from Ref. 5) shows what can be achieved for an earth orbit satellite using spin and an integral controller. It can be seen that the cost of relaxing the mass attraction requirements is the need for very accurate attitude control, so that the component of the unaveraged perturbation along the spin axis that is contributed in the sensitive direction is kept small.

References

1. Lange, B. O., AIAA J., 2, No. 9, 1590-1606, Sept. 1964.
2. Conference on Experimental Tests of Theories of Relativity, Stanford University, July 20-21, 1961, sponsored by the National Aeronautics and Space Administration.
3. Lange, B. O., "The Control and Use of Drag-Free Satellites," SUDAER No. 194, Ph. D. Dissertation, Dept of Electrical Engineering, Stanford University, Stanford, Calif., June 1964.
4. "Semiannual Status Report(s) on the Engineering Portion of a Research Program to Develop a Zero-g, Drag-Free Satellite and To Perform a Gyro Test of General Relativity in a Satellite," submitted by Stanford University to NASA. Semiannual reports are from 1964 through May 1969, inclusive.
5. Blamont, J. E., "Proposition pour la détermination des coefficients du tenseur métrique de l'espace-temps à partir d'une sonde spatiale héliocentrique," submitted to CNES, June 15, 1968.
6. DeBra, D. B., et al., J. Spacecraft and Rockets, 5, No. 9, 1040-1045, Sept. 1968.
7. Powell, J. D., "Control of a Spinning Drag-Free Satellite With an Application of Estimation Theory," SUDAAR No. 402, Ph. D. Dissertation, Dept of Aeronautics and Astronautics, Stanford University, Stanford, Calif., May 1970.
8. Jhin, P., "Control of a Spinning Drag-Free Satellite to Reduce Trajectory Errors Due to Mass Attraction," SUDAAR No. 408, Engineer's Thesis, Dept. of Aeronautics and Astronautics, Stanford University, Stanford, Calif., Aug. 1970.

Table 1. Integral-controller errors

Error source	Relation	Key magnitudes	$\tilde{f}_{eh}$, g	In-track trajectory error after 1 yr, m
Spin rate $\omega_h - \bar{\omega}_h$	$\tilde{f}_{eh} = \frac{C\rho}{2}\left(\frac{\omega_h - \bar{\omega}_h}{k_c}\right)^2$	$C = 10^{-10}$ g/mm $\rho = 0.1$ mm $k_c = 0.02$ 1/s $\omega_h - \bar{\omega}_h = 10^{-3}$ rad/s	10^{-14}	150
Yaw attitude error	$\tilde{f}_{eh} = f_{ez_o}\psi$	$f_{ez_o} = 3 \times 10^{-10}$ g $\psi = 10''$	1.5×10^{-14}	225
Integrator bias	$\tilde{f}_{eh} = C\frac{b_c}{k_c}S_c$	$C = 10^{-10}$ g/mm $b_c = 50\ \mu$V/s $k_c = 0.02$ 1/s $S_c = 0.1$ mm/V	2.5×10^{-14}	400
	$\tilde{f}_{eh} = \frac{1}{2}C(10^{-i})S_c$	$i = 4$ decimal digits $C = 10^{-10}$ g/mm $S_c = 1$ mm/full scale	5×10^{-15}	75
Mass-attraction model	$\tilde{f}_{eh} = 2.5 \times 10^{-3}\left(\frac{\partial^3 f_{ex}}{\partial x^3} + \frac{\partial^3 f_{ey}}{\partial y^3}\right)\frac{d_h^3}{2}$	$\frac{1}{2}\left(\frac{\partial^3 f_{ex}}{\partial x^3} + \frac{\partial^3 f_{ey}}{\partial y^3}\right) = 10^{-12}$ g/mm^3 $d_h = 0.2$ mm	2×10^{-17}	0.4
	$\tilde{f}_{eh} = \frac{1}{8\pi}\Delta C\frac{\Delta v}{\omega_h}$ if $\omega_{Lc} = 2\omega_h$	$\Delta C = 10^{-10}$ g/mm $\Delta v = 10^{-3}$ mm/s $\omega_h = 0.1$ rad/s	5×10^{-14} at resonance	50 after 3 months of resonance
Position sensor	$\tilde{f}_{eh} = 0.1\left(k_{3x}\frac{\partial f_{ex}}{\partial x} + k_{3y}\frac{\partial f_{ey}}{\partial y}\right)\frac{1}{2}\left(\frac{r_d}{k_{1x}}\right)^3$	$C = 10^{-10}$ g/mm $k_{3x} \simeq k_{3y} = 0.3$ 1/mm^2 $k_{1x} \simeq k_{1y} = 1$ $r_d = 0.1$ mm	3×10^{-15}	50

Physical Limitations in Sensors for a Drag-Free Deep-Space Probe

Rémy Juillerat
Office National d'Etudes et de Recherches Aérospatiales (ONERA), Paris

I. Introduction

The relativistic effects derived from an experiment performed with a deep-space probe can be calculated within the necessary accuracy only if the perturbing acceleration exerted on the probe falls below 10^{-12} m/s^2 (Refs. 1 and 2). Since the action of external perturbing forces, with particular reference to solar radiation pressure, exceeds that value by several orders of magnitude, the need arises to device a drag-free spacecraft.

The performance of a drag-free spacecraft was carefully studied by B. Lange (Ref. 3). A vehicle of this type consists of a free-falling, ball-shaped proof mass shielded by the spacecraft itself against external perturbing nongravitational forces. A sensor measures the space between the respective centers of proof mass and cage, and acts on the controls of the craft in such a way as to nullify this distance, forcing the vehicle to follow the proof mass trajectory. Thus, only external perturbation affecting the spacecraft is automatically balanced out by a propulsive counteraction (Fig. 1).

However the spacecraft's perfect shielding of the proof mass against external perturbing forces does not come equally into play against any residual internal forces that may exist between proof mass and craft. The guidance control system tends to compensate for them by giving rise to an acceleration of the vehicle which produces an antagonistic force of inertia. The integration of this acceleration is responsible for the difference between the actual and the ideal, purely gravitational trajectories.

Research conducted at Stanford University (Ref. 4) led to the evolution of a drag-free earth satellite stabilized by a gravity gradient, with an internal disturbance of less than 10^{-11} g (approximately 10^{-10} m/s^2).

Investigations recently made at ONERA were concerned with the feasibility of a drag-free deep-space spacecraft set spinning around an axis normal to its orbital plane, with a spurious acceleration in that plane required to fall below 10^{-12} m/s^2. This research was able to draw on the experience gained in developing the CACTUS-type ONERA accelerometer having a sensitivity of 10^{-8} m/s^2 (Refs. 5 and 6).

This paper analyzes the inner perturbing forces, while taking into account the technological limitations imposed on the proof mass position pick-up and on the proof mass acquisition system. The resulting perturbing accelerations are evaluated as a function of the drag-free sensor parameters. These data are to be used to study the laws of guidance, thus making for an optimization of the probe as a whole.

II. Internal Perturbing Forces

A. Nature of Perturbations

The origins of the particular forces, shown in Fig. 2, comprise

(1) Gravitational attraction, exerted by the masses of the spacecraft upon the proof mass.

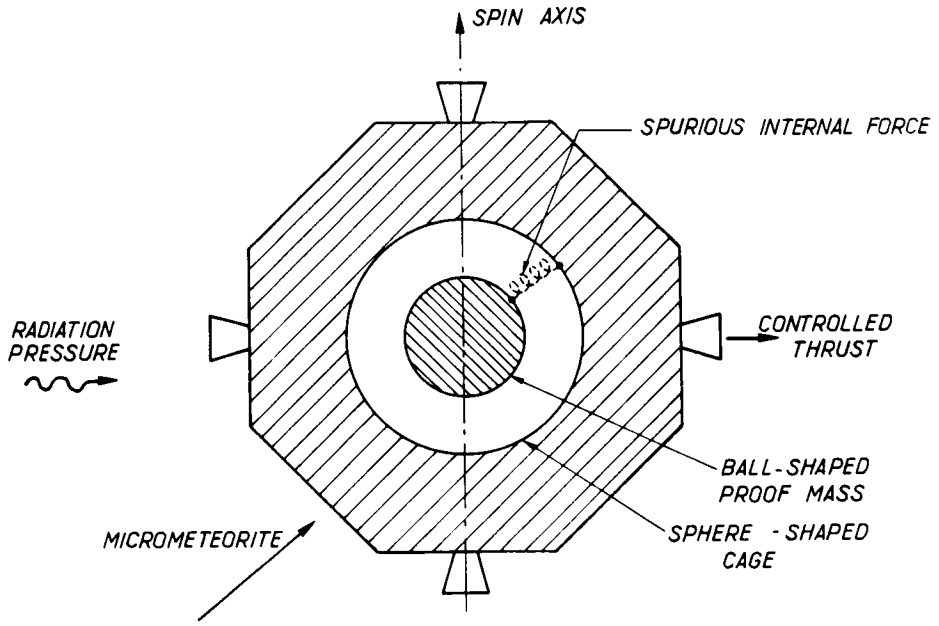


Fig. 1. Drag-free deep-space probe control system

(2) Electrical action, taking two different forms: (a) attraction between the proof mass carrying an electric charge Q and the cage wall carrying the image charge -Q, and (b) a force applied to the proof mass (carrying charge Q) by any electric leakage field generated by the spacecraft inside the cage or by elements of the cage (spurious charges from dielectrics).

(3) Magnetic action, also of two kinds: (a) the effect of the magnetic field gradient of the craft on the proof mass, whose magnetic susceptibility is not strictly zero, and (b) a force due to the motion of the proof mass (bearing electric charge Q) in the interplanetary magnetic field.

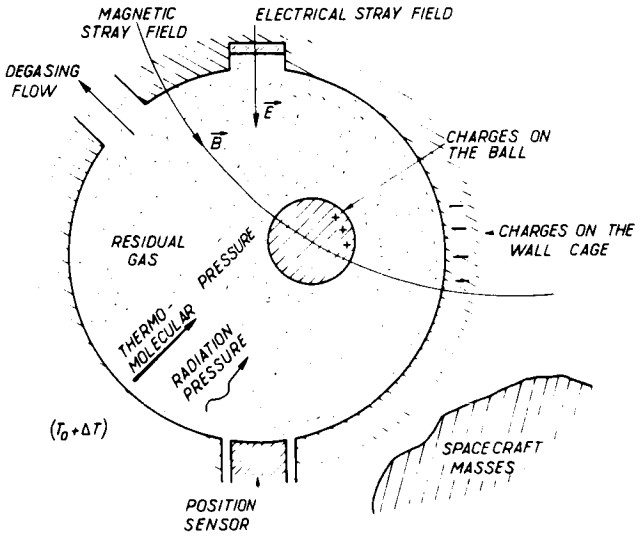


Fig. 2. Sources of perturbing forces

(4) Pressure effects due to residual gases in the cage. The proof mass may be subject to anisotropic pressure in the following cases: (a) dissymmetry of the outgassing molecular flow of elements within the cage, and (b) momentum transfer between the proof mass and thermalized gas molecules on the cage wall, which differs in temperature at various points (thermomolecular pressure). This effect is due to a temperature gradient on the cage wall.

(5) Radiation pressure effects within the cage. The same temperature gradient on the cage wall has a further anisotropic effect on the proof mass through thermal radiation from the cage wall toward the proof mass.

(6) Action of position pickup (capacitive or optical).

All these forces are brought to bear on the same mass m (mass of proof mass). They will therefore be expressed in terms of acceleration directly representing the perturbation of motion of the spacecraft without the neutralizing effect of spin.

B. Classification of Perturbing Accelerations

Perturbing accelerations fall into three categories (Fig. 3), viz:

Type I: Acceleration of constant direction within spacecraft-related axes and of an amplitude independent of the proof mass position. It is a nonrotating (in craft axes) acceleration in a spinning craft, denoted by the subscript 0; i.e., Γ_0.

Type II: Acceleration dependent on the distance $\Delta\xi$ between the proof mass center and the zero point of the position pickup. The

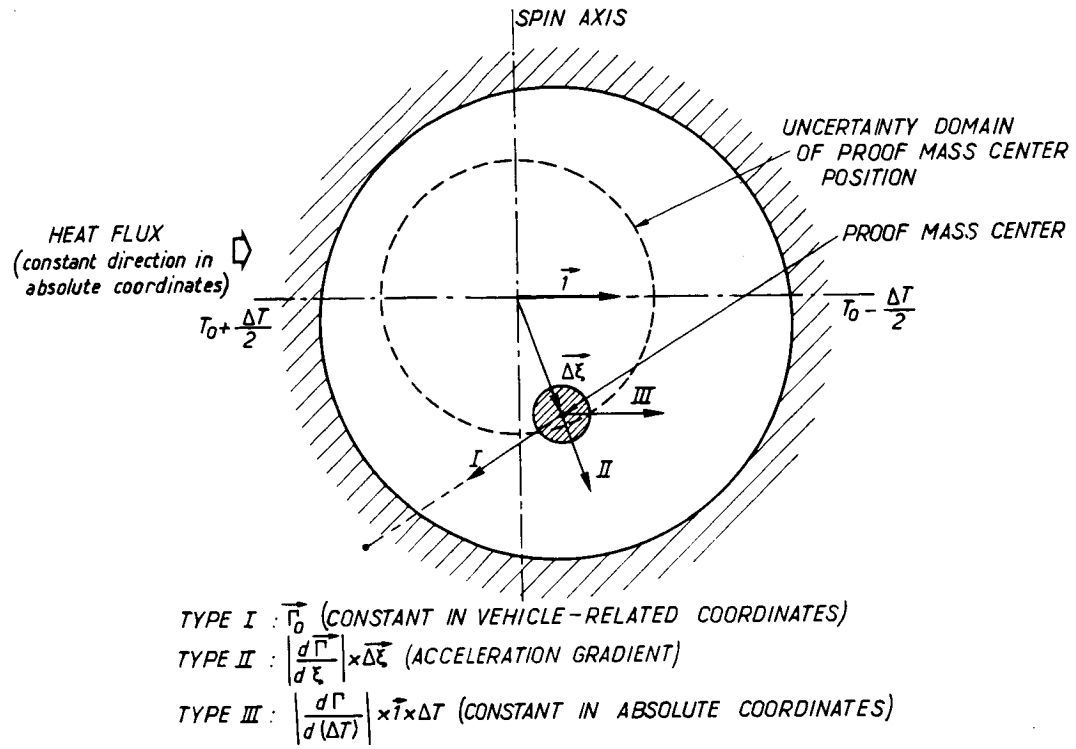


Fig. 3. Types of perturbation

relationship between force and displacement is generally a tensorial one. The term $|d\Gamma/d\xi|$ stands for the value of the highest tensor coefficient.

Type III: Constant acceleration within absolute coordinates. It is a rotating (in craft axes) acceleration in a spinning craft. Such an acceleration may be caused, e.g., by a thermal gradient ΔT whose direction is determined by the position of the sun. This temperature gradient may originate either in the heat dissipation of equipment switched to a frequency that equals the angular frequency of rotation of the spacecraft or in solar flux through the craft (spinning or not). In the case of a spinning craft, only the amplitude of the first harmonic of temperature variation can give rise to a constant acceleration along absolute axes.

Without going into detail over the piloting control of the craft, it should be noted that

(1) Using a three-axis stabilized spacecraft, the three types of acceleration will disturb the craft motion.

(2) Using a spinning craft, with the spin axis normal to the orbital plane, (a) the orbital-plane components of Type I acceleration are averaged with a residual effect diminishing with increasing accuracy of attitude control; (b) the less the uncertainty of the proof mass center position, which depends on the position pickup noise, the slighter will be the untoward effect of Type II acceleration; and (c) Type III acceleration gives rise to the same disturbance as in the case of a non-spinning craft.

C. Forces of Gravity Incident to the Vehicle

Keeping these forces down to manageable levels is largely a problem of technology and, more specifically, dimensional design for, in theory, there is always a way of so distributing the masses of the spacecraft as to ensure that the action of gravity on the proof mass is nil.

In practice, however, the proof mass will remain subject to some residual gravity effects through

(1) Errors in allocating the masses.

(2) Errors in evaluating, on the strength of inertial measurements, the gravity effects of each individual item of equipment.

(3) Consumption of mass of the propulsion units.

(4) Mass displacements due to backlash, looseness, creep, etc.

(5) Mass displacements due to temperature gradients.

Gravity disturbances can be assessed by several different methods, as illustrated by M. Bismut (Ref. 7). One approach he suggests, as applied to a ball-shaped, homogeneous vehicle, rests on the following reasoning: The body is conceived of as split into two equal hemispheres, and the gravity effect is estimated for either half, i.e., $\Gamma_{1/2}$. With position uncertainty and wrong appraisal of the gravity effects of each element taken as represented by an error δ, the value of $\Gamma_{1/2}$ can be estimated with an uncertainty of $\Delta\Gamma_{1/2}/\Gamma_{1/2} = \delta$.

With the same error referred to both hemispheres, the residual action of gravity of Type I will accordingly be on the order of

$$\Gamma_{G_0} = \delta\sqrt{2} \cdot \Gamma_{1/2}$$

Applying this calculation to a spacecraft 2 m in diameter and 350 kg in mass, and separating the near masses (those of the proof mass cage and of the associated electronic equipment--2 kg) from the far masses (all spacecraft equipment located beyond a radius of 15 cm), we have

$$\Gamma_{1/2} \text{ of near masses} = 1.4 \cdot 10^{-8} \text{ m/s}^2$$

$$\Gamma_{1/2} \text{ of far masses} = 1.5 \cdot 10^{-8} \text{ m/s}^2$$

On the assumption that error δ can be kept to below 0.5%, the value of perturbation Type I will be

$$\Gamma_{G_0} = \text{approx. } 10^{-10} \text{ m/s}^2 \qquad (1)$$

This reasoning offers the advantage of a simple way to estimate both the gravity gradient tensor coefficients and the higher-order terms. For each of these coefficients, the sphere is divided into two sections so that by suppressing one of them, the particular coefficient is given its maximum value. The shapes of the sections will depend on the order of magnitude of the coefficient concerned. Thereafter, the same computation is made.

Working with the same data, this leads to

$$\left|\frac{d\Gamma_G}{d\xi}\right| = \text{approx. } 3 \cdot 10^{-9} \text{ m/s}^2/\text{m} \qquad (2)$$

for perturbation Type II.

The near masses account for only a minor part of the total spacecraft mass, and their geometry is much better known than that of the far masses. Their share in the above result is accordingly slight. In first approximation, then, gravity effects will be treated as being independent of the size of the position pickup.

The effect of a temperature gradient inside the craft will depend largely on its structure. An attempt can be made to assess it by considering a cylindrical, homogeneous spacecraft model having an expansion coefficient close to that of aluminum ($2 \cdot 10^{-5}$/deg) (Fig. 4). Choosing a law of constant temperature along a generatrix of the cylinder, dependent on the polar coordinates R_S and ϕ of any point of the spacecraft through the relation

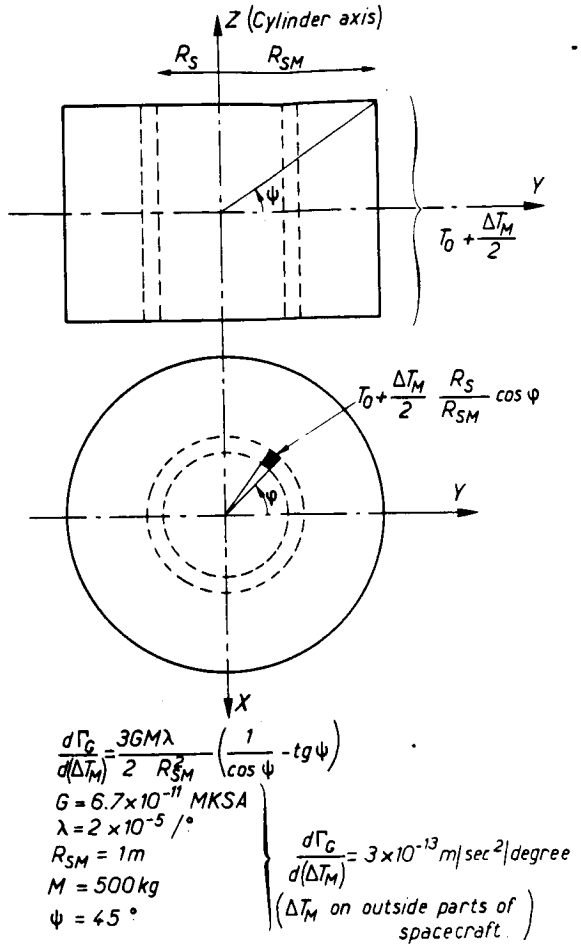


Fig. 4. Effect of temperature gradient on mass attraction

$$T = T_0 + \frac{\Delta T_M}{2} \times \frac{R_S}{R_{SM}} \cos\phi \qquad (3)$$

the following relation is obtained for the Type III perturbation:

$$\frac{d\Gamma_G}{d(\Delta T_M)} = \frac{3GM\lambda}{2R_{SM}^2} \left(\frac{1}{\cos\phi} - tg\psi\right) \qquad (4)$$

For a craft weighing 350 kg and having a diameter equal to its height, namely, 2 m, this Type III perturbation works out as

$$\frac{d\Gamma_G}{d(\Delta T_M)} = 3 \cdot 10^{-13} \text{ m/s}^2/\text{deg} \qquad (5)$$

This value merely represents an order of magnitude, for if the temperature gradient is due to heat

dissipation within the spacecraft; whether cyclic or not, the above law (3) no longer applies.

All the same, it is a useful result in that it points up the impact of temperature gradients on internal gravity effects. Thus, for a three-axis stabilized spacecraft, a temperature difference of the order of 100°C between two diametrically opposite points will induce a perturbing acceleration of $3 \cdot 10^{-11}$ m/s^2. In a spinning spacecraft, on the contrary, ΔT_M is merely the amplitude of the first harmonic of the temperature change, and the higher the rotational speed of the craft, the lower its value will be.

D. Electrical Forces

1. Acceleration Γ_Q Resulting From Attraction Between Charge Q on Proof Mass and Image Charge on Cage

Contrary to the case of gravity effects, this type of perturbation is related to factors which are difficult to control by means of present technology. There are at least two good reasons why the electric charge on the proof mass should not be zero:

(a) When contact is broken between the proof mass and the cage, the proof mass retains an electric charge due to the difference between the work functions of the two materials. Even if the two bodies are plated with the same metal, residual surface pollutions through gas adsorption and surface treatment maintain a difference of approximately 0.05 to 0.1 V between the work functions (Ref. 8). Again, in non-contact conductive bodies whose nearest elements are spaced less than some 10 Å apart, a charge exchange takes place through tunneling (Ref. 9). Yet, the resistance equivalent to this effect shows an extremely rapid variation as a function of the interbody spacing; so much so that the proof mass will retain, even for a very slow separation speed of some 10^{-8} m/s, a charge Q defined by the work function difference (in terms of V) multiplied by the capacitance present between the two bodies at the vanishing point of the tunneling.

The ONERA investigations have shown that the residual charge between two gold-plated bodies given the same surface treatment is on the order of from 10^{-11} to 10^{-12} C. For a proof mass of a few centimeters radius, this charge corresponds to a potential of something like 1 V.

(b) Under free-fall conditions inside the spacecraft, the charge varies as a result of the capture of cosmic particles with sufficient energy to pass through the vehicle. A straightforward assessment (Ref. 10), which is confined to the primary effect (no secondary emission) and neglects nuclear reactions induced by very-high-energy particles, leads to the conclusion that this 1-V potential is reached in a few days under the influence of galactic cosmic particles and can be attained within a few minutes when a high solar flare occurs.

It seems practical to control this charge either by purely physical or by technological means.

The physical method was suggested by W. M. Fairbank of Stanford University (Ref. 11). It consists of inserting between proof mass and cage a luminous flux previously filtered on the short wavelength side so as to limit the energy of the photoelectrons emitted from cage and proof mass to a value of v volts. In this way, the proof mass potential will be stabilized to a value between -v and +v volts.

For physical reasons, such as the fluctuations of the work functions between the different surface points of the cage and the proof mass elements, and effects of the variation of these work functions on the outgassing of these surfaces, no fine stabilization can reasonably be expected. It is most likely a realistic assumption, then, that this potential can be stabilized to a value of no more than 1 V.

This can be achieved, for example, by means of a gold layer over all proof mass and cage surfaces, gold having a work function of something like 4.8 eV, or by an ultraviolet light flux filtered to a minimum wavelength of 2100 Å, which corresponds to energy photons of 5.8 eV. This light can be drawn from the solar flux. The photoelectric stabilizing current yielded by the flux passing through an aperture of about 1 cm^2 considerably exceeds that induced by cosmic particles. The radiation pressure of this luminous flux is of a negligible value compared with other perturbations.

Even should laboratory experiments bear out the feasibility of this procedure, it would seem wise to locate on the cage an electrometric sensing device for remote measurement of the proof mass potential. A vibrating reed electrometer should be sufficiently sensitive, particularly as the integration time of the instrument need not be short.

Such metering provides a means of checking the proper functioning of the equipment, and it will also enable the potential to be servo-controlled by a value closely approaching zero (or at any rate, lower than 1 V) through operating charge injection devices (such as photoemissive apparatus). In any event, this potential will be assumed to have a maximum value of 1 V.

The law defining the acceleration, Γ_Q, imparted to a proof mass of mass m, radius r, and potential V (taken as constant), located in a cage of radius R = r + e, with its center offset with respect to the cage center by the value ξ, is written (Fig. 5):

$$\Gamma_Q = \frac{4\pi\epsilon_0 V^2 r^2}{3me^2}\left[\frac{\xi}{e} + \alpha\left(\frac{\xi}{e}\right)^3 + \cdots\right] \qquad (6)$$

if $e/r \ll 1$ (slight ball-cage gap).

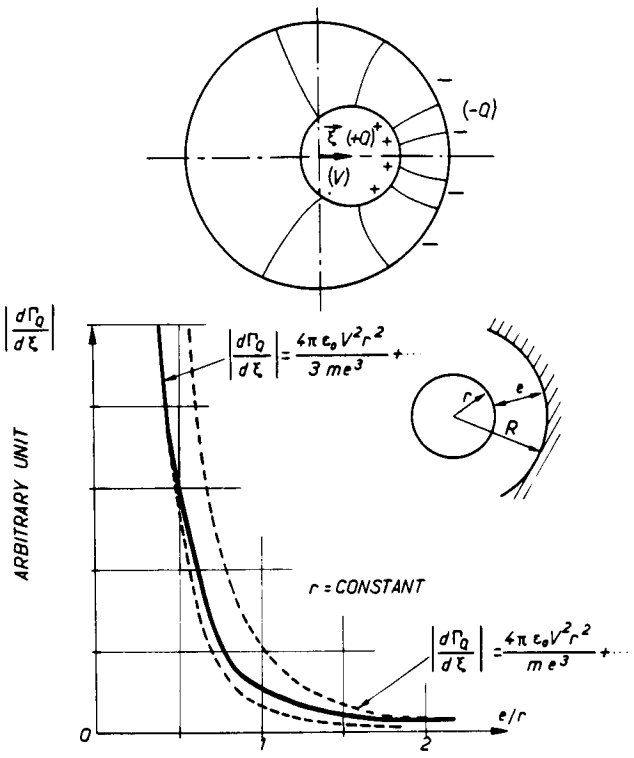


Fig. 5. Action of electric charge on ball

$$\Gamma_Q = \frac{4\pi\epsilon_0 V^2 r^2}{m e^2}\left[\frac{\xi}{e} + \alpha\left(\frac{\xi}{e}\right)^3 + \cdots\right] \qquad (7)$$

if $e/r \gg 1$ (point charge, small ball). The expressions of these two extreme cases differ by only a factor of 3.

For $e/r < 1$, the first expression comes closer to the actual law. The expansion to the third order gives α the value 1/2:

$$\Gamma_Q = \frac{4\pi\epsilon_0 V^2 r^2}{3m e^3}\left[1 + \frac{1}{2}\left(\frac{\xi}{e}\right)^2 + \cdots\right] \qquad (8)$$

or, for $\xi/e < 0.1$:

$$\Gamma_Q \simeq \frac{4\pi\epsilon_0 V^2 r^2}{3m e^3}\xi \qquad (9)$$

With the error made in setting the position pickup to zero on the geometric center of the cage taken as 1% of the proof mass/cage gap (i.e., 10^{-2} e), the following values can be derived from expression (8):

Constant term (Type I perturbation)

$$\Gamma_{Q_0} = 3.7 \cdot 10^{-13}\frac{r^2}{m e^2} \qquad (10)$$

Coefficient of first-order term (Type II perturbation)

$$\left|\frac{d\Gamma_Q}{d\xi}\right| = 3.7 \cdot 10^{-11}\frac{r^2}{m e^3} \qquad (11)$$

2. Acceleration Γ_E Through the Action on Proof Mass Charge Q of a Stray Electric Field E Inside the Cage

Electric field E is largely due to electric charges attaching to the surface of the cage dielectrics, which may be pickup insulators or lenses of the optical devices.

Given a small relative gap e/r (the kind to be chosen later), this leakage field has a local action exerted only on part of the proof mass. If, for instance, a 1-kg proof mass of 150-cm^2 overall surface area, brought to a potential of 1 V, is subjected to a field of 1 V/m operating locally on an area of 1 cm^2, the disturbance will be a mere $2 \cdot 10^{-13}$ m/s^2.

These stray fields seem amenable to being limited to values below 10^{-12} m/s^2 by the use of a suitable staggered shielding. In optical detection, however, shielding tends to become an altogether more delicate operation, requiring the optical system to be placed inside a deep enclosure, or the use of metal grids at the exit of the optical beams.

E. Magnetic Forces

1. Acceleration Γ_B Through the Action on the Proof Mass of Magnetic Susceptibility κ of a Stray Magnetic Field Gradient Inside the Cage

In a proof mass of non-ferromagnetic material, the only magnetic moment is one induced by the stray magnetic field B within the cage. This perturbing acceleration is therefore expressed by

$$\Gamma_B = \frac{\kappa}{2\mu_0\rho}\nabla(B^2) \qquad (12)$$

Given a copper proof mass of $\kappa = 10^{-6}$ magnetic susceptibility and $\rho = 8.9 \cdot 10^3$ kg/m^3 density, a perturbing acceleration of $0.5 \cdot 10^{-12}$ m/s^2 is produced by a dipole located 25 cm off the proof mass center, and having a magnetic moment of 1 A · m^2. The induction field generated at the proof mass center is $B = 3 \cdot 10^{-5}$ Wb/m^2. Thus, the perturbation appears limitable to below 10^{-12} m/s^2 by an expedient arrangement of the spacecraft parts and effective shielding of the detection system.

A word of warning is appropriate here about the need for stringent precautions against the variable magnetic induction fields caused by equipment recurrently switched to a frequency that equals the craft's speed of rotation. The resulting effect would induce a constant acceleration (Type III) in the absolute axes.

Equally careful safety measures must be taken to protect the proof mass from ferromagnetic impurities. A ferromagnetic particle of 1 μg (dust adhering to surface or trapped inside mass) imparts to the proof mass located in an induction field of $3 \cdot 10^{-5}$ Wb/m^2 an acceleration of the order of 10^{-12} m/s^2.

2. Acceleration Γ_S Due to Spacecraft's Movement in the Interplanetary Magnetic Field

With the spacecraft moving at velocity v orthogonally to the interplanetary magnetic field H_S, the acceleration acting on the proof mass carrying electric charge Q and unshielded against that field will be

$$\Gamma_S = \frac{\mu_0 Q v H_S}{m} \qquad (13)$$

The mean value of the interplanetary magnetic field is on the order of 5Y (Ref. 12). By Mariner's measurements, the field may reach values of 20Y (Ref. 13), or $2 \cdot 10^{-8}$ Wb/m^2. At a velocity of 35 km/s, the spurious acceleration is, accordingly, something like 10^{-15} m/s^2. Since such an effect is identically zero in a perfectly conductive cage, and as the actual cage will be made from a good conductor, this kind of perturbation may safely be regarded as totally negligible.

F. Effects Due to Residual Gases

1. Pressure of Residual Gases Between Ball and Cage

To ensure the lowest possible pressure, it is proposed to create a vacuum between ball and cage by having this gap issue into outer space. This is easy to effect by means of cylindrical ducting some 10 cm in diameter (Fig. 6), complete with baffles against micrometeorites, and with dust filters.

With no part of the spacecraft protruding beyond the plane of the duct outlets, the molecules emitted from the vehicle by the outgassing and sublimation of some constituents are left with little, if any, possibility of access to the ducts owing to the very high value of the mean-free path of the gases evacuated. In other words, the outer vacuum will act as an environment of negligible residual pressure in relation to that prevailing inside the ducting. Residual pressure between ball and cage is accordingly computed from the outgassing rate of the constituents and from the area of the internal surfaces.

The outgassing rates of non-outgassed high-grade metals (copper, stainless steel, gold) are at present put at 10^{-6} Pa · m/s (approx. 10^{-9} torr · ℓ · s^{-1} · cm^{-2}) after 10 h under vacuum, and of the same metals outgassed at 400°C at 10^{-9} Pa · m/s (approximately 10^{-12} torr · ℓ · s^{-1} · cm^{-2}) (Refs. 14 and 15).

As a function of time, these rates obey a 1/t rule. This rule has been checked for periods of several hundred hours. Over longer periods, the rate necessarily decreases more rapidly, as the calculation of the mass of gas emitted would lead to a divergent integral. The 1/t rule thus gives a maximum value for the evaluation of the rate on a mission of several months' duration.

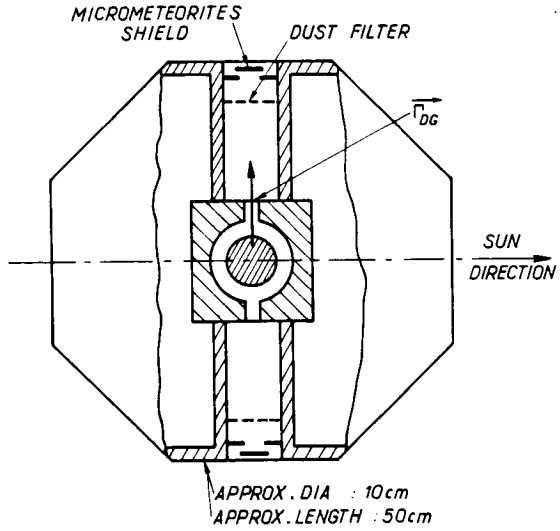


Fig. 6. Proposed sensor vacuum system, acceleration due to sensor outgassing

Accordingly, the outgassing rate of the constituent cage and duct materials is taken to be 10^{-8} Pa · m/s, or approximately 10^{-11} torr · ℓ · s^{-1} cm^{-2}. This rate can be achieved (a) after an outgassing time in orbit of approximately 1000 h (1-1/2 months), with the system not previously subjected to hot outgassing, or (b) by giving the drag-free sensor a hot outgassing and protecting it from the earth atmosphere with a cover to be removed in orbit. In this case, the rate will be reached in a shorter time, probably within some tens of hours.

The residual pressure between ball and cage will be largely due to the low conductance of the dust filters. That of a perforated-foil filter of 30% transparency, with 10-μm holes, is approximately $3 \cdot 10^{-2}$ m^3 s^{-1} (30 ℓ/s).

Given a 0.5-m^2 overall surface area of the ball, cage and set of accessories located within the filter-bounded space, the internal pressure will have the value

$$P_0 = 0.8 \cdot 10^{-7} \text{ Pa, or approximately } 10^{-9} \text{ torr} \qquad (14)$$

2. Perturbing Effect Due to Outgassing

On the strength of the above values, with the ball-to-cage gap giving into the open via a single

outlet, the outgassing flux leads to an acceleration imparted to the ball and having the computed value (with $e/r < 1$) .

$$\Gamma^{+}_{DG} = \frac{4 \cdot 10^{-10}(r + e/2)^4}{m\, e^2} \qquad (15)$$

The use of a system of symmetric ducting will obviously reduce this thrust. A 20% unbalance of the conductances of the two ducts has been shown to give rise to a ten times lower acceleration, a value taken as the basis for this particular effect:

$$\Gamma_{DG} = \frac{4 \cdot 10^{-11}(r + e/2)^4}{m\, e^2} \qquad (16)$$

To minimize the perturbation, the ducting axis may be set perpendicular to the orbital plane.

3. Thermomolecular Pressure Effect Due to a Temperature Gradient

With a temperature gradient present on the cage wall, the neutral particles of the residual gas may, by thermal accommodation on the wall, take up an energy varying with the temperature of the wall section concerned. It follows that the whole series of collisions between these molecules and the proof mass will impart to the latter a certain momentum of a direction dependent upon that of the cage temperature gradient. To estimate this force, the following assumptions are made:

(a) Steady state: there is no source and no sink, i. e., neither outgassing of walls nor condensation on them.

(b) Full thermal accommodation: after a collision, the energy of the particle is determined by the wall temperature at the point of impact.

(c) The molecules are re-emitted in compliance with Lambert's law (spherical emission).

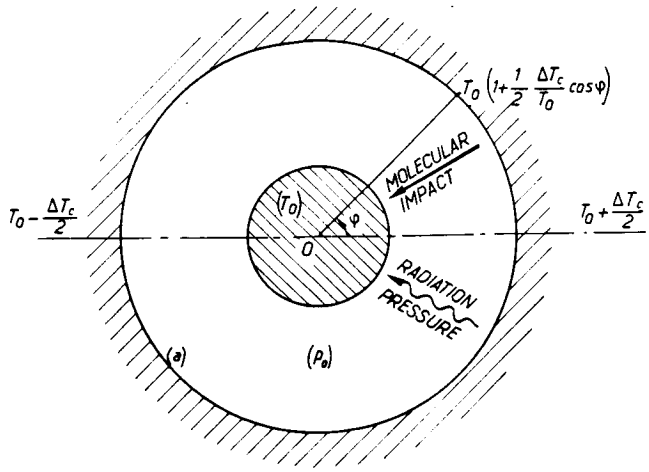
(d) There is no temperature gradient on the proof mass skin.

(e) The temperature on each cage wall point is defined by the relationship (Fig. 7).

$$T = T_0\left(1 + \frac{1}{2} \cdot \frac{\Delta T_C}{T_0} \cos \phi\right) \qquad (17)$$

Assumption (a) implies that the flux emitted per unit area of wall surface is equal at each of its points.

Assumption (d) leads to the conclusion that the flux of the molecules re-emitted by the proof mass gives a zero resultant force, thus leaving only the



	THERMO-MOLECULAR PRESSURE	RADIATION PRESSURE
$\frac{e}{r} \ll 1$	$p_{TM} = \frac{p_0 \Delta T}{3T_0}$	$p_R = \frac{8a\sigma T_0^3 \Delta T}{3c}$
$\frac{e}{r} \gg 1$	$p_{TM} = \frac{p_0 \Delta T}{2T_0}$	$p_R < \frac{4a\sigma T_0^3 \Delta T}{3c}$

$p_0 = 10^{-7} P_a = \text{app. } 10^{-9}$ Hgmm
$T_0 = 300K$
$a = 0.1$
$\sigma = 5.7 \times 10^{-8}$ MKSA
$c = 3.10^8$ m/s
$\frac{e}{r} \leqslant 1$

$$\frac{d\Gamma_{TM}}{d(\Delta T_c)} = \frac{5 \times 10^{-10} r^2}{m}$$

$$\frac{d\Gamma_R}{d(\Delta T_c)} = \frac{5 \times 10^{-9} r^2}{m}$$

Fig. 7. Action of cage thermal gradient through differential pressures

effect of the cage re-emission molecules to be considered.

For the 10^{-7} Pa pressure value computed in Section 2, the mean-free path is on the order of 50 km. Hence, there are practically no collisions between molecules inside the cage. The result of this calculation, expressed in terms of differential pressure P_{TM} brought to bear on the cross section of the proof mass, is

$$\frac{P_{TM}}{P_0} = A\frac{\Delta T_C}{T_0} \qquad (18)$$

Coefficient A is a form factor having a value of 1/2 for a point proof mass ($e/r \gg 1$), and 1/3 for an infinitely small relative gap ($e/r \ll 1$).

From this is derived the expression for the Type III perturbation due to the temperature gradient of the cage, assuming a residual pressure of 10^{-7} Pa (10^{-9} torr) and a temperature of 300°K:

$$\frac{d\Gamma_{TM}}{d(\Delta T_C)} = \frac{5 \cdot 10^{-10} r^2}{m} \qquad (19)$$

G. Effect Due to Thermal Radiation of Cage Wall

The same temperature gradient ΔT_C present on the cage wall will give rise to a flux of anisotropic radiation on the proof mass (Fig. 7).

Retaining assumptions (d) and (e) in Section 3, i.e., the same temperature laws for cage and proof mass, and treating both as black bodies, the relationship below is derived, giving expression to this effect in terms of differential pressure applied to the cross section of the proof mass:

$$P_R = \frac{A'}{3} \cdot \frac{\sigma T_0^3 \Delta T_C}{C} \qquad (20)$$

Coefficient A' is a form factor of the values A' = 4 for $e/r \gg 1$ and A' = 8 for $e/r \ll 1$.

In practice, however, if the two bodies have an emission coefficient, a, of less than 1 (bright bodies), the multiple reflections from this radiation will have the effect of reducing the influence of the gradient. It can be shown that, in the first approximation, and for a small gap ($e/r < 1$), coefficient a will occur as a factor in the above expression:

$$P_R = \frac{8a\sigma T_0^3 \Delta T_C}{C} \qquad (21)$$

For a larger gap, the averaging effect is more pronounced, and the expression below gives a maximum value

$$P_R < \frac{4a\sigma T_0^3 \Delta T_C}{C} \qquad (22)$$

This leads to the expression for the Type III perturbation due to the temperature gradient of the cage and its own thermal radiation, on the assumption of an emission coefficient of 0.1 (highly reflecting gold coating):

$$\frac{d\Gamma_R}{d(\Delta T_C)} = \frac{5 \cdot 10^{-9} r^2}{m} \qquad (23)$$

This effect is predominant over that due to residual gases at a pressure of 10^{-7} Pa.

H. Disturbance Due to Position Pickup

1. Capacitive Pickup

The characteristics of capacitive pickup are described in a later section. What should be said here is that, if the photoelectric control of the proof mass potential described in Section IV A is not to be upset, the voltage applied to the pickup electrode must not exceed 1 V.

As this pickup is essentially made up of capacitance bridges (one bridge per axis) (Fig. 8), the electrostatic force of the sensing electrodes is differential, and the resulting acceleration is

$$\Gamma_D = 2\epsilon_0 \mathcal{V}_d^2 \times \frac{\pi D^2}{4} \cdot \frac{\xi}{m\, e^3} \qquad (24)$$

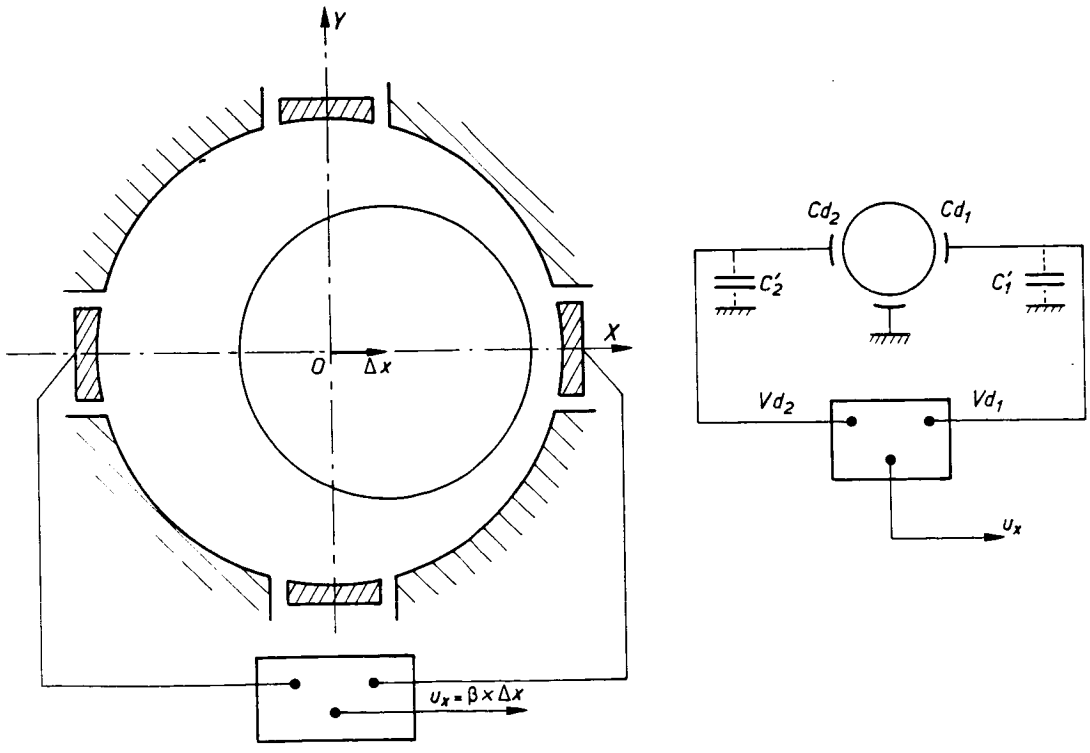


Fig. 8. Capacitive pickup

Where the six electrodes do not cover the entire cage surface, a comparison of expression (24) with (6) shows that the perturbation will invariably be less than that caused by the electric charge on the proof mass.

2. Optical Pickup

Whatever kind of optical device is used, the luminous flux needed to detect the proof mass position is invariably very slight, and a simple calculation can demonstrate that the incident perturbation falls several orders of magnitude below those discussed above.

I. Inferences From the Analysis of Perturbing Forces

The foregoing analysis has shown the main sources of perturbing acceleration to be

(1) Gravity action due to the spacecraft.

(2) Effect of the electric charge on the proof mass.

(3) Effects of the temperature gradient of the craft and the cage of the drag-free sensor.

These accelerations were expressed as a function of

(1) The proof mass radius r.

(2) The gap between proof mass and cage e.

(3) The mass of the proof mass m.

Obviously, the last parameter must be of the highest possible value. This prompts the choice of a high-density material. Copper seems to meet the case fairly well, as it has a relatively high density ($8.9 \cdot 10^3$ kg/m^3), it is a diamagnetic material of low susceptibility (10^{-6}), and, in the O. F. H. C. quality, it is a stable material well suited to vacuum techniques.

On the basis of this choice, Figs. 9, 10, and 11 illustrate the main perturbations as referred to a proof mass radius of between 1 and 10 cm, i. e., a mass of between 37 g and 37 kg, for the two relative gap values $e/r = 0.2$ and $e/r = 1$. Figure 9 shows acceleration Type I, and Fig. 10 represents the acceleration gradients (Type II perturbation). The acceleration induced by the temperature gradients of the spacecraft, ΔT_M, and of the cage wall, ΔT_C, is shown in Fig. 11.

III. Position Pickup

A. Capacitive Pickup

In actuality, the capacitive pickup comprises three identical pickups, one for each of the three axes defining the proof mass position. The pickup on each axis is essentially made up of a capacitance bridge (Fig. 8), whose response on its own axis is given by

$$\frac{U_x}{V_d} = 2\,\frac{\Delta C_d - \Delta C'}{\Sigma C_i} \qquad (25)$$

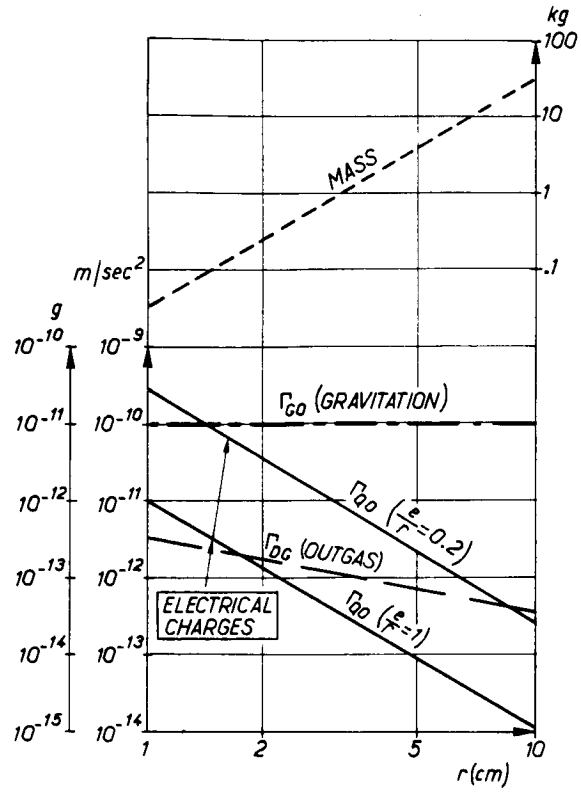


Fig. 9. Main perturbing accelerations (Type I)

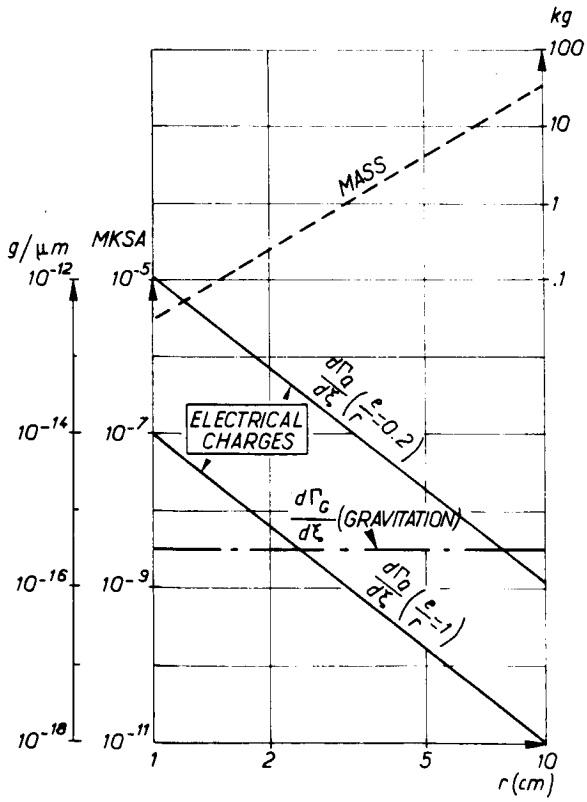


Fig. 10. Main perturbing acceleration gradients (Type II)

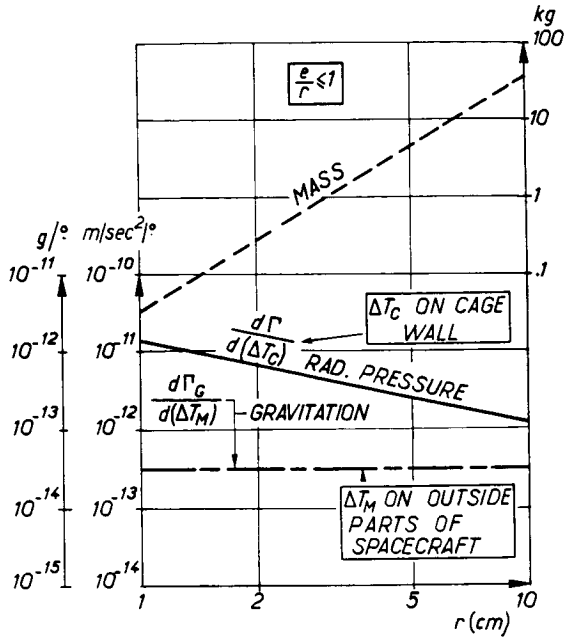


Fig. 11. Perturbing accelerations due to thermal gradients (Type III)

where

$$\Delta C_d = C_{d1} - C_{d2}$$

and

$$\Delta C' = C'_1 - C'_2$$

The pickup's sensitivity is a result of the relationship existing between ΔC_d and the component Δx along the axis of the sensor detecting the distance between the proof mass center and the point corresponding to the position pickup zero. It was assumed that this point could be set in a sphere of 0.01 e radius (Section IV-A). This relationship can be written, for $\Delta x/e \ll 1$, as

$$\frac{\Delta C_d}{C_d} = S\frac{\Delta x}{e} \qquad (26)$$

The sensitivity coefficient S is dependent on the surface area of each capacitance electrode (Fig. 12). Since the cage carries a large number of devices (six capacitance electrodes, the acquisition device, the input device for the luminous flux, the electrometric pickup, etc.), each electrode has a relatively small surface area. Through the selection of a 1% ratio of pickup area to proof mass area, the sensitivity coefficient can be expressed as a function of the relative gap e/r. For an e/r between 0 and 0.2, the coefficient lies quite close on 1 (Fig. 12). For e/r = 0.4 it drops to 0.5, and decreases at a rapid rate from then on. This justifies the choice of e/r = 0.2 in the earlier sections.

The domain of uncertainty on the proof mass center position is, accordingly, a cube with its edges lying in the direction of the position pickup axes and having the value η. This value is given by the thermodynamic noise of the electronic circuitry associated with the capacitance bridge, and also by the variations $\Delta C'$ of the stray capacitances C' caused by local temperature fluctuations. Any variation in stray capacitance $\Delta C'$ will be interpreted as an equivalent deviation Δx_C, defined by the relationship

$$\Delta x_C = \frac{1}{S} \cdot \frac{\Delta C'}{C'} \cdot \frac{C'}{C_d} \cdot e \qquad (27)$$

By suitably disposing the circuits, the stray capacitances can be reduced to very low values, arising exclusively from the presence of electric leakage fields at the ends of some conductors. With the leakages occurring through the coats of varnish and embedding resins, such variations $\Delta C'/C'$ are caused by variations in the dielectric constants of these materials. Their value, on the strength of experience gained with the CACTUS accelerometer (Ref. 5), is given as

$$\frac{\Delta C'}{C'} = 10^{-4}/\text{deg} \qquad (28)$$

Again, the value of C_d is written in first approximation:

$$C_d = 10^{-2}\frac{4\pi\epsilon_0 r^2}{e} = \text{approximately } 10^{-12}\frac{r^2}{e} \qquad (29)$$

Stray capacitance C' is only slightly dependent on dimensions e and r, and can be estimated at less than 0.5 pF. This gives

$$\Delta x_C = \frac{5 \cdot 10^{-5}}{S}\left(\frac{e}{r}\right)^2/\text{deg} \qquad (30)$$

Thus, this first source of position pickup noise is seen to depend solely on the relative gap e/r.

Putting the variance of local temperature fluctuation at less than $(0.5°)^2$, relation (30) provides the following estimate for the noise at zero mean value of the position pickup, due to local temperature variations:

$$\text{for } e/r = 0.2, \quad \eta = \left[\overline{(\Delta x_C)^2}\right]^{1/2} < 1\ \mu\text{m} \qquad (31)$$

$$\text{for } e/r = 0.4, \quad \eta < 8\ \mu\text{m}$$

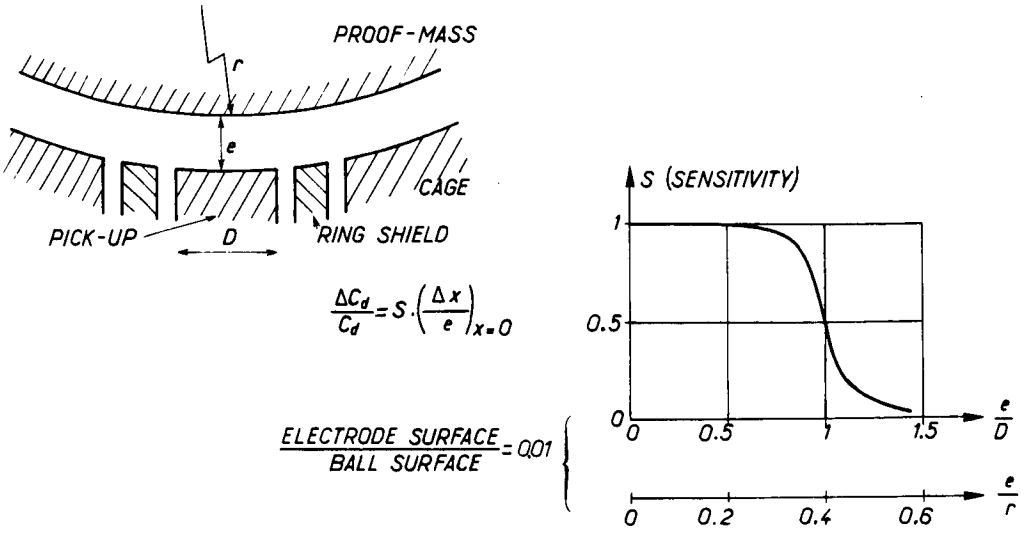


Fig. 12. Capacitive pickup sensitivity

It is easy to ascertain that, with V_d chosen to be 1 V, the thermodynamic noise of the electronic circuitry is negligible by comparison with the above value. An inquiry into the thermodynamic noise also shows that, for one particular amplifier, it, too, depends on the relative gap e/r alone.

In conclusion, then, it can be said that the uncertainty of the proof mass position is given by relation (31) for each axis under measurement.

B. Optical Pickup

The systems applicable include light beam intercepting devices and scanning photomultiplier detectors. Unlike capacitance pickups, these devices give better service with larger relative gaps (e/r) in use.

On the debit side, optical equipment takes up more cage space than do capacitance electrodes and its dynamic range falls below the corresponding capability of a capacitance pickup. These considerations make for the conclusion that, unless there is need, for any reason, to use a relative gap of over 0.2 or even 0.4, the capacitance pickup seems to be the better choice as a position-detecting device.

IV. Acquisition and Reset System

A. General Considerations

In the launching phase, the proof mass needs to be mechanically secured by retractable arms. Moreover, the impact of a micrometeorite or some other incident occurring on a mission may bring the proof mass into contact with the cage wall, to which it may stay attached by short-distance forces of attraction (Van de Waals forces). Such attachments through contact between ball and cage can be minimized by an appropriate control logic so long as provision is made for the proof mass to be returnable to the cage center from any point of adherence. If there are, say, eight retractable arms, this operation is feasible for a ball diameter of roughly half the cage diameter.

This is what brings us to the view that the e/r ratio must not exceed 1 whatever the type of position pickup in use. An examination of the perturbing forces has shown, incidentally, that there is no need for e/r > 1.

B. Principle of Acquisition and Reset With $e/r \leq 0.4$

The retractable arms can be mounted as shown in Figs. 13 and 14. The ends of the ball supporting arms are insulated from the cage and connected to a voltage generator. Once the spacecraft is in orbit, the arms are brought back to the level of the cage surface and can act as electrodes, exerting an electrostatic action on the ball. The electrode voltages can be controlled by the proof mass position pickup so as to keep the ball centered in the cage. This form of suspension works on the same principle as in electrostatic accelerometers.

When the relative speed of the ball in the cage reaches a sufficiently low rate, the operating voltages are cut off, and the ball is kept centered by the guidance control of the spacecraft. This device

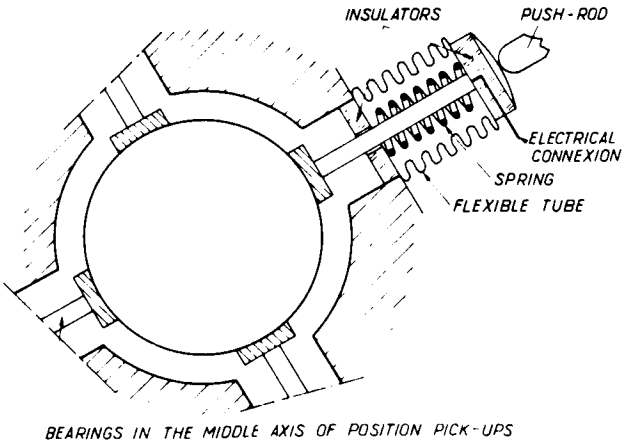


Fig. 13. Proof mass bearing for launch phase

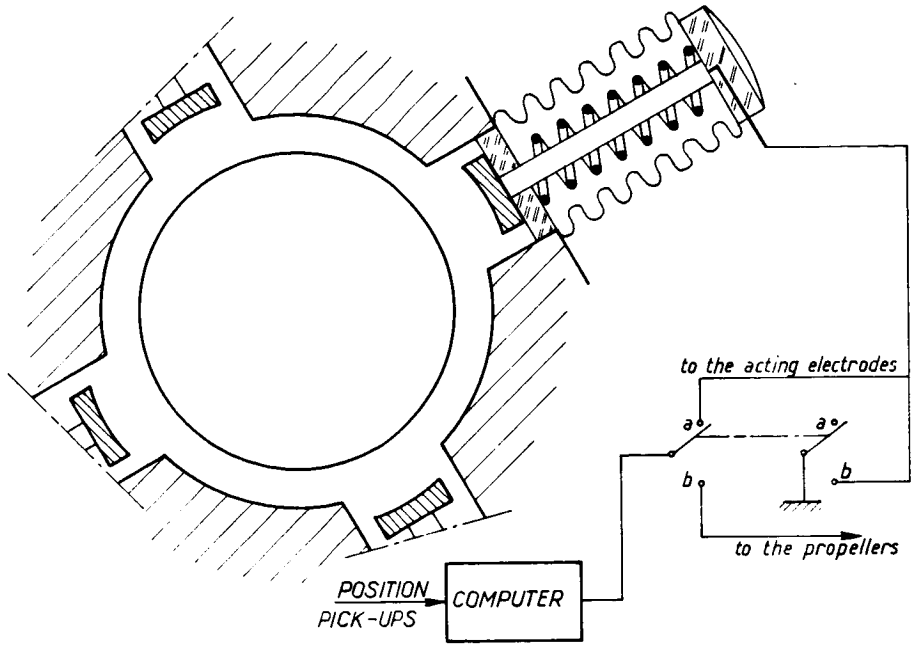


Fig. 14. Acquisition system

offers a dual advantage: (1) In the course of the initial acquisition, the backing movement of the support arms can barely be effected without imparting speed to the ball; the electrostatic suspension remedies this effect. (2) Should the proof mass happen to work back on to the cage wall, it can be brought back into position by exclusive reliance on the play of electrostatic forces. The mechanism will therefore have to be used only once for securing the proof mass on launch.

C. Assessment of the Electric Forces Required for Reset With $e/r = 0.2$

The electric forces must of necessity overcome the short-distance forces of attraction. The latter were estimated by P. Bertrand (Ref. 16), following up work by Casimir (Ref. 17) and Fierz (Ref. 18). He gives the following expression:

$$F_C = \frac{\pi h C}{1440} \cdot \frac{2\pi (r + e/2)^2}{e\, d^3} \qquad (32)$$

This theoretical expression has been experimentally confirmed on the CACTUS accelerometer. It will take further testing in reference to the system parameters to make certain that the law is applicable to all drag-free sensors. In any case, expression (32) can be regarded as the first approximation to the relevant orders of magnitude. It is shown in Fig. 15 for a proof mass radius between 1 and 10 cm, and a ball surface roughness estimated at 0.1 μm.

The maximum electric recovery force is dependent on the maximum admissible voltage on the work electrodes. Granting that, for $r = 10$ cm, these electrodes and the cage may be spaced no more than 1 mm apart, and with a maximum electric field of 200 kV/cm (admissible under vacuum), the operating voltage can be expressed by

$$V_a = 2 \cdot 10^5\, r$$

With the proof mass spaced at a distance of 2 e from the electrode, the electrostatic force of attraction is written as

$$F_a = \frac{\pi \epsilon b V_a^2 r^2}{4 e^2} \qquad (33)$$

where b = electrode surface area/ball surface area.

Figure 15 gives the value of F_a versus r, with $b = 10^{-2}$. This force exceeds the theoretical

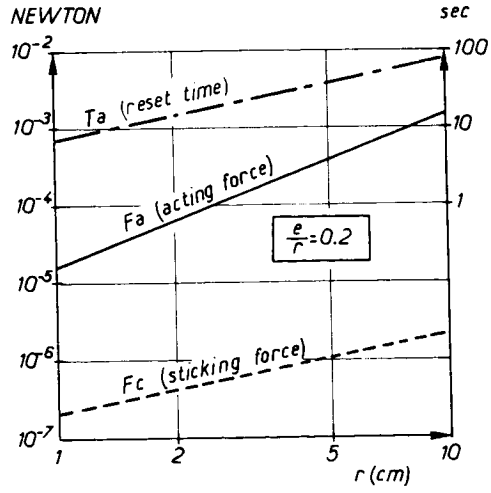


Fig. 15. Reset system performance

value of the force of adhesion by 2 or 3 orders of magnitude. The figure also shows the time taken to effect the reset by this procedure referred to r; this time ranges from 8 to 80 s.

D. Principle of Acquisition and Reset With $e/r = 1$

With $e/r = 1$, the value of the maximum electric forces attainable approaches that of the forces of adhesion; thus the reliability of the device seems to become inadequate, and the operation is best carried out in two phases:

Phase 1: Partial withdrawal of support arms and suspension by electrostatic means, as before.

Phase 2: With operating voltage cut off and spacecraft guidance control switched into circuit, complete withdrawal of support arms.

In the event of accidental adherence, reset is to be effected accordingly by bringing the electrodes closer to the ball, i. e., by concurrent operation of both the electrical and mechanical facilities. This is a serious drawback to the use of a large gap system, in which an optical pickup is required.

V. Conclusions

For a drag-free deep-space spacecraft used in experiments on the theories of gravitation to have a perturbing acceleration of less than 10^{-12} m/s^2 in its orbital plane, the inevitable choice is a spinning craft, since its constant acceleration, in the craft axes, is on the order of 10^{-10} m/s^2. The acceleration gradients and the uncertainty of the position of the proof mass play a vital part in such spacecraft, and induce the residual perturbing acceleration, which is dependent on the guidance control pattern chosen.

The present investigations were conducted at ONERA by M. Bismut (Ref. 7), who showed that, with the use of a control system including a good estimator of the integral term, the equivalent residual acceleration due to position pickup noise may be reduced to $2 \cdot 10^{-13}$ m/s^2; the relevant data were as follows:

Proof mass radius	3 cm
Proof mass mass	1 kg
Relative gap e/r	0.2
Pickup noise	2 μ
Limit cycle	0.1 mm
Computer accuracy	10^{-3}

Because of the spin, Type I accelerations are not involved except for their components along the spin axis. This kind of acceleration, on the order of 10^{-10} m/s^2 and mainly due to gravitational attraction, calls for an attitude control accuracy of 0.1 deg.

A study of Type III acceleration (constant in absolute axes) has, moreover, pointed up the importance of temperature gradients. They need, however, to be clearly separated into the constant term and the term related to the first harmonic at the spacecraft rotational frequency; the constant term gives rise to a Type I and the cyclic term to a Type III acceleration. Thus, for a 1-kg proof mass, the cage wall may show a 20° temperature difference between two opposite points, but the variation of this difference at the spacecraft rotational frequency must not exceed 0.2°. The same applies to the spacecraft structure, for which the order of magnitude of the temperature variations in phase with the rotating motion must be below 1°. Stringent as these requirements may be, they are well within the reach of present-day space technology.

In light of the investigations described, the achievement of a perturbation level below 10^{-12} m/s^2 seems to be within the range of technological possibility. This has been shown by the present feasibility study, and may be confirmed by laboratory experiments, in which the validity of the assumptions made would be checked. But the imaginative powers of scientists will be severely strained to devise the comprehensive, yet low-cost, experiment in which the actual operating conditions are closely approached, so that the true level of the residual perturbations may be ascertained. In this respect, experiments performed in an orbital station are bound to be of supreme interest.

Nomenclature

A, A' = numerical coefficients; shape factors

a = emissivity of proof mass and cage

B = magnetic induction vector

b = numerical coefficient

C_{d1}, C_{d2} = pickup electrode capacitance

C'_1, C'_2 = stray capacitance

c_i = bridge capacitance picked off of a voltage amplifier

C = speed of light in vacuum

D = diameter of pickup electrode roughness

d = proof mass surface unevenness

E = stray electric field

e = gap between proof mass and cage, or charge of electron

F_a = acting force utilized for acquisition of proof mass

F_c = sticking force between proof mass and cage

G = gravitation constant = $6.7 \cdot 10^{-11}$ MKSA

H_S = interplanetary magnetic field

h = Planck's constant = $6.6 \cdot 10^{-34}$ MKSA

M = satellite mass

m = mass of proof mass

P_0 = pressure in cage

P_{TM} = thermomolecular pressure acting on proof mass

P_R = radiation pressure acting on proof mass

R = radius of cage

R_S = modulus in cylindrical coordinates of any point of the craft

R_{SM} = outer radius of a cylindrical spacecraft

r = proof mass radius

S = sensitivity factor of capacitance pickup

T = temperature

T_0 = mean temperature of satellite

ΔT = temperature gradient

ΔT_C = temperature difference at two opposite points of cage

ΔT_M = temperature difference at two opposite points of satellite

Δx = component on the X-axis of the distance of proof mass center from zero point of position pickup

U_x = output voltage of capacitance bridge on the X-axis

V = proof mass potential

V_a = voltage of acting force electrodes

V_{d1}, V_{d2} = voltage of capacitance pickup electrodes

$\mathcal{V}$ = satellite velocity

α, α' = numerical coefficients

β = overall sensitivity of capacitance bridge

Γ_B = perturbing acceleration due to stray magnetic field gradient

Γ_{DG} = perturbing acceleration due to outgassing flow

Γ_E = perturbing acceleration due to stray electric field

Γ_G = perturbing acceleration due to internal gravitational forces

Γ_Q = perturbing acceleration due to electric charge on proof mass

Γ_R = perturbing acceleration due to radiation pressure

Γ_S = perturbing acceleration due to interplanetary magnetic field

Γ_{TM} = perturbing acceleration due to thermomolecular pressure

$\Gamma_{1/2}$ = acceleration due to gravitational forces of spacecraft half

δ = relative uncertainty of gravitational forces in all parts of satellite

ϵ_0 = permittivity of vacuum = $8.85 \cdot 10^{-12}$ F/m

κ = magnetic susceptibility of proof mass

λ = expansion coefficient of satellite

μ_0 = permeability of vacuum = $1.25 \cdot 10^{-6}$ H/m

η = uncertainty on proof mass center position due to position pickup noise

ξ = distance of proof mass center from cage center

$\Delta\xi$ = distance of proof mass center from zero point of position pickup

ρ = proof mass density

σ = Stefan's constant = $5.7 \cdot 10^{-8}$ MKSA

ϕ = polar angle of any point of spacecraft or cage as measured in reference to axis set in direction of temperature gradient

ψ = angle, in cylindrical coordinates of any point of spacecraft

References

1. Israel, G. M., this proceedings.
2. Marchal, C., this proceedings.
3. Lange, B. O., AIAAJ., 2, 9, 1590-1606, Sept. 1964.
4. DeBra, D. B., this proceedings.
5. Delattre, M., "L'accéléromètre ONERA a grande sensibilité," AGARD - Guidance and Control Symposium, Braunschweig, May 7-9, 1968, AGARD Conference Proceedings No. 43.
6. Bouttes, J., Delattre, M., and Juillerat, R., "Résultats des essais en vol de l'accéléromètre à haute sensibilité CACTUS," 21st I. A. F. Congress, Constanz, October 4-10, 1970.
7. Bismut, M., "Détecteur et pilotage d'une sonde héliocentrique à trainée compensée," unpublished ONERA document, ESTEC Contract 1032/70/HP.

8. Palevsky, H., Swank, R. K., and Grenschik, R., The Review of Scientific Instruments, 18, 5, May 1957.

9. Harper, W. R., Contact and Frictional Electrification, Clarendon Press, Oxford, London, 1967.

10. Tiffon, J., "Evaluation de l'évolution du potentiel de la masse d'epreuve sous l'influence des particules cosmiques à haute énergie," ONERA, ESTEC Contract 1032/70/HP.

11. Fairbank, W. M., quoted by B. Lange in Ref. 3, p. 1602.

12. Elbaz, C., "Introduction de la physique du milieu spatial," N. T. 05-4 - O. N. E. R. A. / C. E. R. T. S., p. 87, Toulouse, 1970.

13. Neugebauer, M., and Snyder, C. W., quoted in Atomes, No. 270, p. 680, Nov. 1969.

14. Zhilnin, W. S., Zhilnina, L. P., and Kuzmin, A. A., "Gas Emission from Structural Materials in Vacuum at Room Temperature and after Degassing at 150-450°C," Ministry for the Electronic Industry, Moscow, USSR, Proceedings of the Fourth International Vacuum Congress, Part 2, Manchester, Apr. 1968.

15. Henry, R. P., "Measure des taux de dégazage," These Faculté des Sciences de Paris, Oct. 30, 1968.

16. Bertrand, P., "Forces d'attraction moléculaires entre corps solides électriquement neutres," ESTEC Contract 1032/70/HP.

17. Casimir, H. G. B., Proc. Kon. Ned. Akad. Wetenschap, 51, 793, 1948.

18. Fierz, M., Helvetica Physica Acta, 33, 8, 1960.

Applications of Presently Planned Interplanetary Missions to Testing Gravitational Theories

L. D. Friedman
Jet Propulsion Laboratory
California Institute of Technology

I. Introduction

This paper will present a summary of the probable interplanetary missions for the 1970s that might prove useful in testing the general theory of relativity. Although it is to be understood that relativity testing is not (yet) a major scientific objective of any of these missions, we will try to show that significant testing can be done on them.

Anderson (Ref. 1), Curkendall (Ref. 2), and Trask (Ref. 3) have discussed the non-gravitational limitations of interplanetary spacecraft modeling and possible means of circumventing these limitations. The dispersive media model limitation on knowledge of the ray path is not severe with the introduction of dual-frequency tracking. The unmodelled non-gravitational forces, however, do constitute the important limitations on using interplanetary spacecraft. Estimates in the following sections and covariance studies for both the Helios and the Venus-Mercury projects suggest that to make significant estimates of β and $J_{2\odot}$, the unmodelled accelerations must be less than 0.1% of the modelable part, e.g., at approximately 10^{-10} m/s^2, in the inner planet region of the solar system. For longer arc tracking on outer planet missions, we feel that 10^{-11} m/s^2 will be nearer the minimum upper bound tolerable for unmodeled accelerations. This value could be achieved by (1) designing a drag-free spacecraft, (2) using very accurate accelerometers (see section V), or (3) seeking very tight engineering specifications on gas leakages during the mission and on material reflective property changes. We are also quite hopeful that improved filtering and modeling will reduce sensitivity to these forces. Anderson (Ref. 1) has shown how filtering improvements can allow the tolerance to the unmodeled non-gravitational forces to increase. If the filter is fortuitously chosen, or if several filters are permitted to operate on the data, improvement in the estimates may be forthcoming (see also Ref. 5 in which the improved treatment of process noise — the unmodeled accelerations driving the state — is described).

Non-gravitational forces can also be eliminated by anchoring the spacecraft to the planet; indeed, such missions offer exciting possibilities, for they provide the opportunity of tracking massive bodies with highly accurate doppler and time-delay radio (active radar) data. Possibilities for experiments with such spacecraft are just beginning to be investigated (Ref. 4).

II. Interplanetary Space Program

It is impossible to present a definite plan for the interplanetary space program for the next decade, not because none exists (many do) but because of the vagaries of politics and technical developments. The following summary is taken from recommendations of the Planetary Exploration Planning Panel at NASA (Ref. 5) and from information from the present NASA 5-year plan, which is being used in the budget preparations. Cislunar and earth orbital missions are not included. Inclusion of the activities of other nations has not been attempted, with the exception of Helios — a joint West German-U.S. project. The

Soviet Union has not announced any future plans for interplanetary missions, although they have indicated that at least a vigorous Venus and Mars exploration program will continue.

In Fig. 1, the space program of the 1970s is depicted as a plot of the missions' heliocentric distance vs date. By reading horizontally across the plot, we see the missions planned over the decade to various parts of the solar system. By reading vertically, we can obtain a picture of the range through the solar system of scientific measurements at any time. The dashed lines on the plot depict possible extensions to the basic missions.

We briefly summarize the missions and some of their characteristics in Tables 1 and 2. Not mentioned in the tables are the solar-electric propulsion mission to an asteroid, listed in the 5-year plan for a 1978 launch, or the Jupiter Atmospheric Entry Probe, similarly listed for a 1980 launch, since they are more problematical at this time and their functions are quite obvious.

III. The Mariner Venus-Mercury 1973 Mission

The Mariner mission planned by the United States for flyby of Venus and Mercury in early 1974 is the first announced planned use of a planetary swingby to provide additional energy to a spacecraft's orbit in order to reach a more distant planet (Ref. 6). The trajectory for the basic mission is depicted in Fig. 2.

The scientific objectives for the Mariner Venus-Mercury (MVM) mission are primarily to obtain close flyby observations of Mercury and its environs, and secondarily to obtain Venus encounter and interplanetary measurements. To achieve these objectives, NASA formulated a Science Steering Group (SSG) to advise and coordinate the preliminary science instrumentation and recommendations. The SSG met for about 6 months preceding the payload (i.e., experimenters) selection. Celestial mechanics, including gravitational theory testing, was represented on the Radio Science and Celestial Mechanics Team. Reports of this team and the SSG report have been published as MVM project documents (Refs. 7 and 8).

One very important outgrowth of the SSG-Project early studies was the planned use of both S- and X-band tracking data (ranging and doppler) throughout the mission. As seen in Fig. 2, the MVM mission is characterized by relatively short cruise arcs between two planetary encounters. Since there will be three or four midcourse corrections, no very long arc of tracking data will be available for reduction. Thus, the emphasis on this mission for the celestial mechanics experimenters is on using the tracking data of the planetary encounter to deduce information about the target planet's gravitational field. If only the basic mission were flown (ending at Mercury encounter plus 20 days), no definitive measure of γ, β, or $J_{2\odot}$ could result.

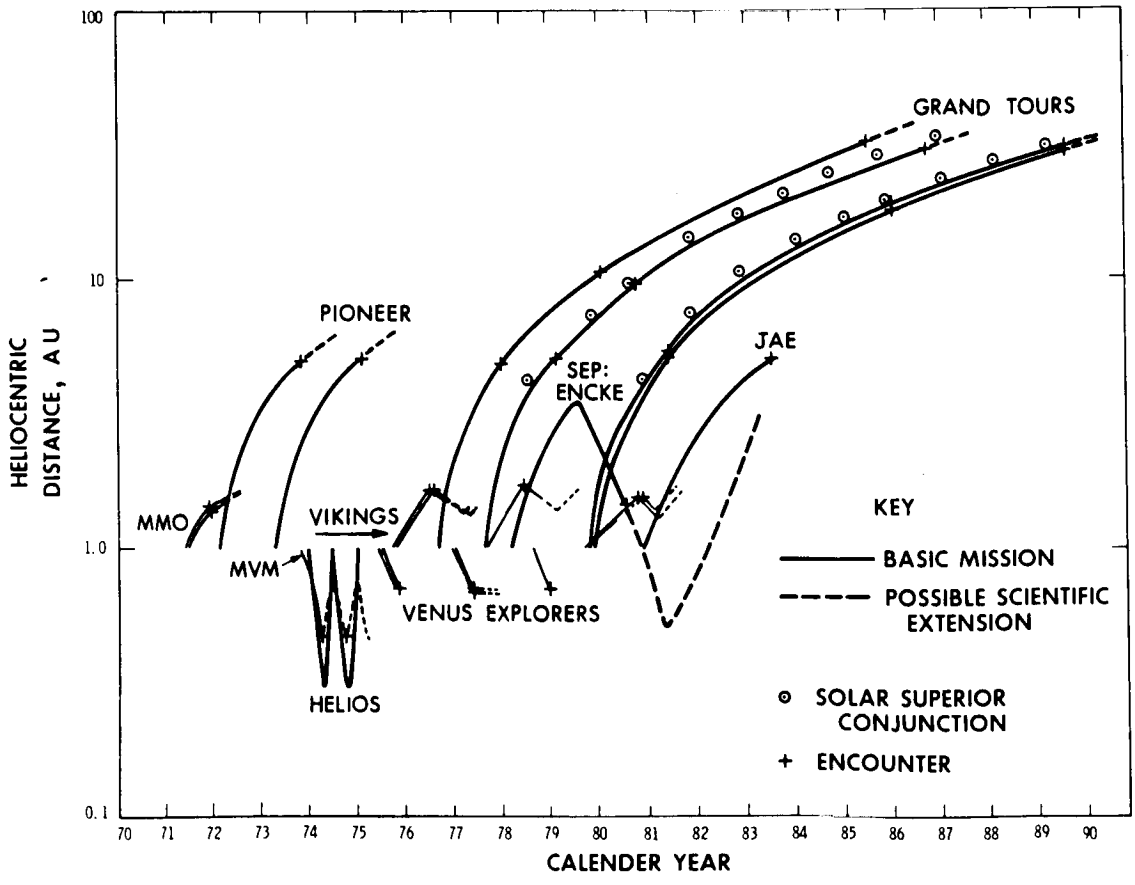


Fig. 1. Possible interplanetary mission schedule in the 1970s

Table 1. Tentative interplanetary mission set (approved)

Mission	Spacecraft	Primary objectives	Science payload	Celestial mechanics investigators	Relevance to relativity testing
MM 1971 (Mars orbiter)	2 Mariner orbiters (JPL)	Mars mapping, investigation of atmosphere, topography, internal activity, and mass distribution. Fields, particles	60 kg: TV, UVS, IRIS, IRR	J. Lorell (JPL) I. I. Shapiro (MIT)	Anchored spacecraft Ephemeris improvement Precise orbit about central body
Pioneer F, G 1972 (Jupiter flybys)	Pioneer flyby (NASA-ARC, TRW)	Exploration of interplanetary medium, Asteroid belt, and Jupiter environment	25 kg: Magnetometer fields and particles, radiation, IRR, UV, meteoroids, imaging, polarimeter	J. Anderson (JPL) G. Null (JPL)	Long-arc tracking (only doppler) Improvement of Jupiter mass
MVM 1973 (Venus, Mercury flybys)	Mariner flyby (JPL, industry)	Observations and measurements of Mercury and its environment	60 kg: Imaging, IRR, UVS, plasma science, magnetometer, charged particle telescope	J. Anderson (JPL) I. I. Shapiro (MIT)	Long-arc tracking, possible Possible superior conjunction communication, dual-frequency Possible two Mercury encounters (anchors) Ephemeris improvements
Helios 1973 (solar probe)	Solar probe (W. Germany – MBB, USA – NASA GSFC)	Measurement of solar wind, charged particles and particularite matter, particle-field interactions	52 kg: Plasma, magnetometer, micrometeoroids, cosmic ray, zodiacal light, photometer	W. Kundt (Hamburg) J. Anderson (JPL) W. Melbourne (JPL)	Long-arc tracking Eccentric solar probe, low periapsis Superior conjunction
Viking 1975, 1977?, 1979? (Mars orbiter, lander)	2 orbiters (NASA LaRC, JPL) 2 landers (NASA LaRC, Martin)	Acquisition of visual, thermal, and water vapor data; radio science Life and organic systems search, TV and atmospheric observations; radio and in situ measurements	70 kg: TV, IR 30 kg: TV; life, soil, and water samplers; atmospheric sensors	W. Michael (NASA LaRC) D. L. Cain (JPL) I. I. Shapiro (MIT)	Both kinds of anchored spacecraft Orbiter/lander geometry Precise orbit about central body Precise planetary motion

In February 1970, a Mercury science symposium was held at Caltech. At this meeting, Prof. Columbo queried as to whether the Mariner spacecraft, after encountering Mercury, returned to the planet again. Subsequent investigation found that the swingby of Mercury indeed provided wide control of the post-encounter period of the spacecraft, and that there existed a family of ballistic trajectories which returned to Mercury 176 days later (two Mercury orbital periods). What was most remarkable is that there were trajectories among this family which satisfied all the other trajectory requirements of the baseline mission. A heliocentric plan view of a Mercury-to-Mercury return trajectory is shown in Fig. 3.

With the astrodynamic feasibility of a Mercury re-encounter established, the Project made a study of a so-called expanded MVM mission. This study provided conclusions which stated that the expanded mission was scientifically extremely desirable and, with some engineering modifications to the spacecraft (connected with making additional midcourse maneuvers, and providing continuous communication with the high-gain antenna), could be provided at a relatively small incremental additional cost. It also showed that multiple Mercury returns could be made repeatedly as long as sufficient midcourse fuel was available to perform trajectory guidance.

Table 2. Tentative interplanetary mission set (not yet approved)

Mission	Spacecraft	Primary objectives	Science payload candidates	Celestial mechanics investigators	Relevance to relativity testing
Venus probes 1975, 1978 (Venus entries)	2 Explorers (NASA GSFC)	Composition and circulation pattern of atmosphere, distribution and composition of clouds	12-25 kg: UV, IR, and imaging; fields and particles	Not yet selected	Incorporation of tracking data Improved Venus ephemeris
Venus orbiter 1976 (Venus orbiter)	Explorer (NASA GSFC)	Venus fields and particles environment and atmospheric observations	12-25 kg: Fields and particles, UV, IR, and visual imaging (camera)	Not yet selected	Anchored spacecraft may improve Venus ephemeris, but limited by orbit trim maneuvers and atmospheric perturbations
Grand Tour 1976, 1977, 1979 (outer planet flyby)	TOPS (JPL, industry)	Observation of all outer planets and several of their natural satellites	100 kg: imaging TV, IR, and UV; radio emission; occultation; fields and particles; micrometeorites	Not yet selected	Long-arc tracking Superior conjunctions Ephemerides of all planets Satellite ephemerides Solar system escape

With an expanded mission, relativity tests become feasible, and four specific tests have been suggested:

(1) The S/X-band ranging and doppler tracking through superior conjunction should provide a confirmation, and at least one order of magnitude improvement, to the Mariner 1969 experiment because the dual-frequency ranging and doppler data will both eliminate the limitation due to the dispersive media models and reduce the dependence on a previously established accurate orbit (hence relieving the low-thrust force limitation). The solar corona model in particular will not be critical as with the Mariner 1969 experiment. Furthermore, the spacecraft is on a much faster orbit and the superior conjunction passage occurs much more quickly than on the Mariner Mars missions, so that the buildup in error due to unmodeled accelerations does not occur to as large an extent.

The Radio Science and Celestial Mechanics Team estimated that γ could be measured to 0.002 and β to 0.1 with the dual-frequency S/X-band system in use through solar superior conjunction. This experiment does not depend on reencounter, just on extension of the mission, and on the continuous S/X-band operation through superior conjunction.

(2) The long-arc tracking of the spacecraft, depending on the midcourse correction strategy, in the Mercury-to-Mercury return phase may allow estimation of β, γ, and $J_{2\odot}$ in the full least-squares orbit determination solution. The limit here will be with the non-gravitational forces and will thus depend on the "cleanliness" of the spacecraft. Preliminary studies have shown that a 2% estimate of β and an estimate of $J_{2\odot}/R_{\odot}$ with a standard deviation of less than 10^{-5} are possible with a long arc of tracking, when the low-thrust forces can be modeled, but also have indicated significant correlation between the β and $J_{2\odot}$ estimates and those parameters of the low-thrust forces. The author has performed computer covariance analyses for this mission (Ref. 9) and found that even if the tracking period is nearly a year, the correlation coefficient between β and $J_{2\odot}$ estimates is about 0.9, and the coefficients between these parameters and the modelable nongravitational force parameters are greater than 0.96.* The high correlations show both the high sensitivity to the nongravitational force model and the need for independent data to separate the β and

*The parameter set included probe initial conditions, low-thrust force parameters, relativity parameters, $J_{2\odot}$, observing site locations, the astronomical unit and interplanetary medium constant. Including $\dot{G}$ in the solution led to unity correlations between it and $J_{2\odot}$ and β.

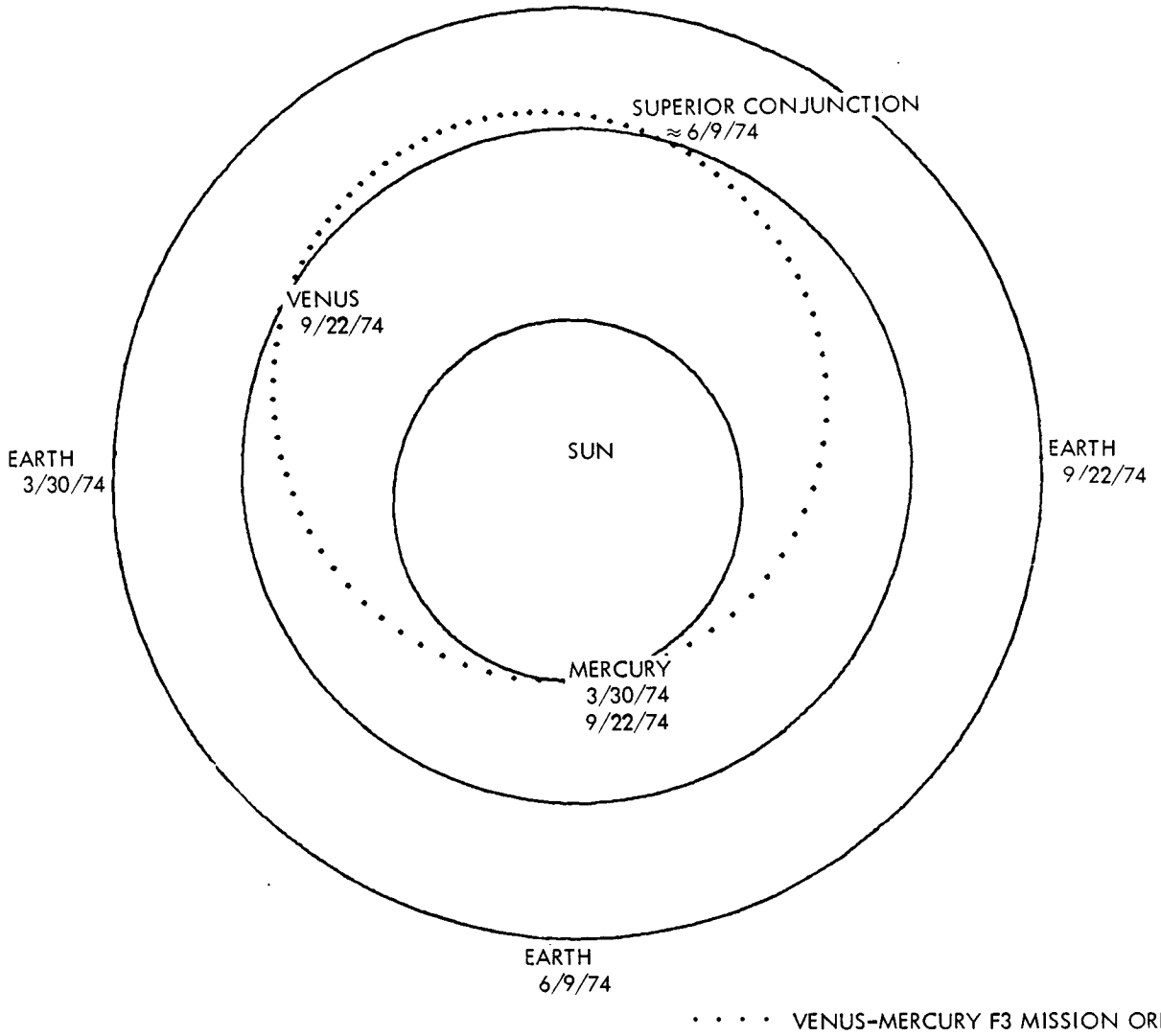


Fig. 2. Venus-Mercury 1973 mission orbit

$J_{2\odot}$ estimates. The results and previous experience indicate that error estimates will increase by at least a factor of 3 to 10 if the unmodeled accelerations are 1% of the modeled part. The long-arc data reduction does not require the Mercury re-encounter.

(3) The two ties to Mercury's orbit may provide extremely accurate indications of Mercury's motion, from which we would like to deduce the solar oblateness coefficient. This possibility is still being investigated, but it appears that if a $J_{2\odot}/R_{\odot}^2$ of 10^{-5} is to be detected, the Mercury position (relative to earth) will have to be resolved to better than 1 km. A resolution of such magnitude has not been predicted but is not far from being feasible. The celestial mechanics investigators on the mission are currently investigating this point.

(4) Finally, we cite the continuing work of Lieske et al. (Ref. 10) at JPL and Shapiro et al. at MIT, who seek to combine the interplanetary spacecraft data with the tens of thousands of planetary observations to estimate all solar system parameters and to provide very accurate planetary ephemerides from which relativistic gravitational effects can be observed. The MVM experimenters include members from these groups, and they will actively pursue studies needed to properly evaluate the potential from concatenating the different data sets into the estimation process. Detailed computer simulations estimating the many parameters are necessary as part of these studies. The two encounters to Mercury should be most beneficial in the combined data set, for they will provide independent benchmarks for the radar data which can resolve ambiguities and allow the filter to give accurate parameter estimates. It seems probable that this combination of radio, radar, and optical data of the inner planets will provide the best gravitational theory experiment from the mission, and indeed perhaps the best of the 1970s.

The basic mission ending at Mercury encounter plus 20 days is now the Project plan. However,

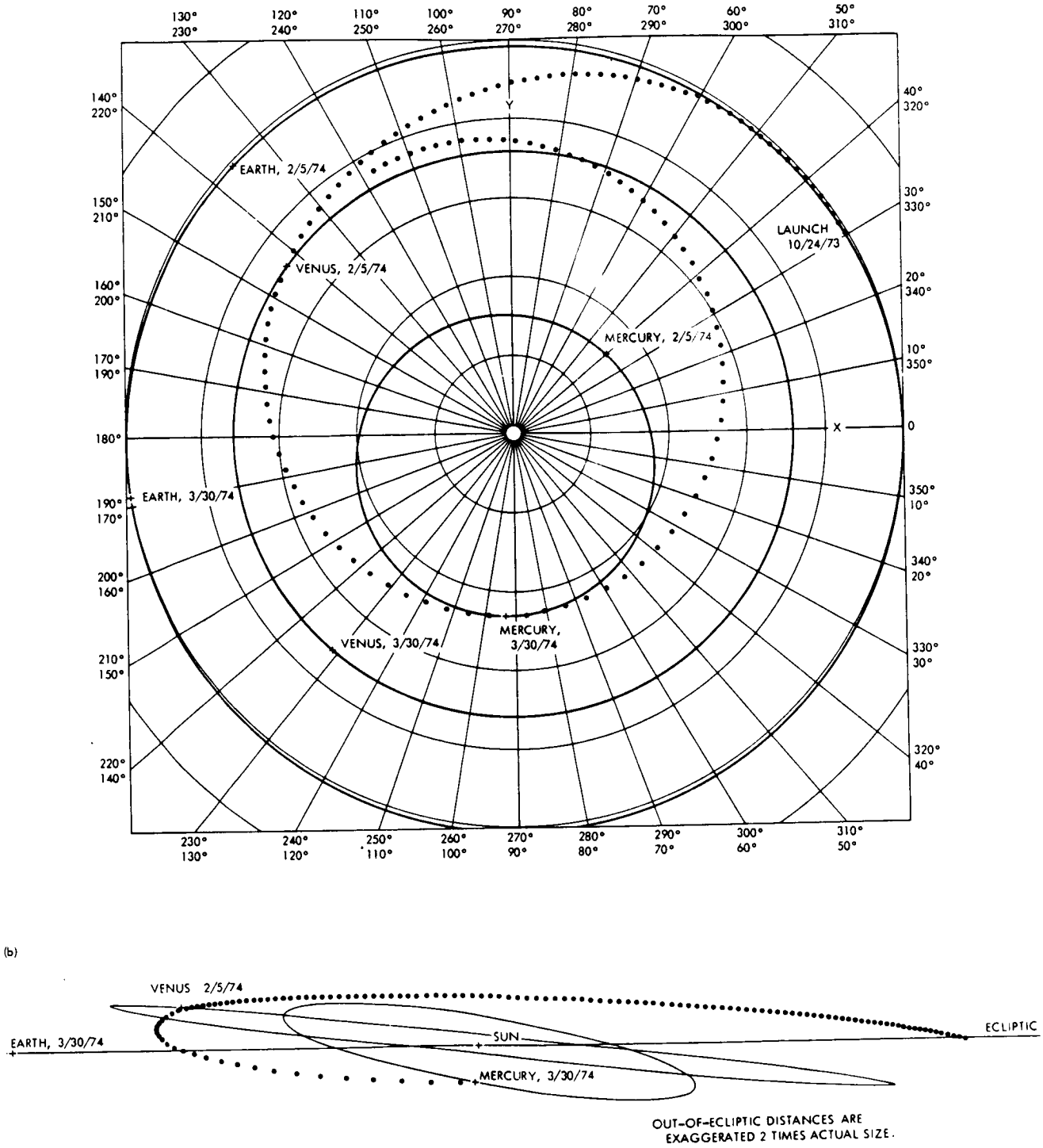


Fig. 3. Heliocentric views of transfer trajectory, October 24, 1973, launch date

an attempt is being made to design the spacecraft and the mission to allow expansion of the mission to accommodate the superior conjunction experiment and Mercury return(s).

IV. Mission Characteristics

In seeking to "hop on board" already existing missions to do relativity testing, we are concerned with both the characteristics of the trajectory and mission plan for obtaining relativistic effects from the tracking, and the characteristics of the spacecraft to suggest instrumentation for measuring relativistic effects directly. In the next section, we discuss additional spacecraft instrumentation for relativity testing; here we consider aspects of the missions. We will continue to confine our attention here to the other flyby missions of the 1970s.

In 1974, a German spin-stabilized solar probe (Helios) will be launched into an orbit with a

perihelion distance of from 0.2 to 0.3 AU.* A representative trajectory for a perihelion distance of 0.2 AU is shown in Fig. 4 in a synodic coordinate system (which rotates at the angular rate of the earth about the sun). As can be seen from the figure, the spacecraft spends a great deal of time within a few degrees of the sun, and a general relativistic time-delay test is certainly feasible.

The Helios project will probably equip the spacecraft with an S-band ranging transponder; they have no plans to include an X-band capability. Nevertheless, it should be possible to obtain a 1% test of the general relativistic time delay, and hence of γ, by ranging to Helios. Also, the relatively small perihelion distance of Helios suggests the possibility that a significant measurement of the parameter β and the dynamical solar oblateness parameter $J_{2\odot}$ might be possible. In order to investigate this possibility, JPL investigators have performed covariance analyses for two orbits with perihelion distances of 0.2 and 0.3 AU. A drag-free situation has been assumed. Although this is not the case for Helios, still the results are useful for comparison purposes between the two trajectories.

The results also indicate what can be achieved with drag-free satellites which make relatively close approaches to the sun. Time delay and doppler measurements with random noise values of 277 ns and 1.37 mHz, respectively, were assumed. Significant β and $J_{2\odot}$ estimates require 100 days of tracking (assuming 1 pass per week), but beyond that, time improvement in the errors of the estimates is very small. With a perihelion distance q of 0.3 AU, estimates of β to 4 - 5% and of $J_{2\odot}/R_{\odot}^2$ to $0.6 - 0.7 \times 10^{-5}$ are all that is possible. From the closer orbit (q = 0.2 AU), a much more favorable determination of β and $J_{2\odot}$ can be achieved. The determinations of β to 0.01% and $J_{2\odot}$ to $0.002 \times 10^{-5}\ R_{\odot}^2$ for q = 0.2 AU are far more significant values. The disparities in the precision between the 0.2 AU and the 0.3 AU case are so great that the investigators have recommended a smaller perihelion distance, if possible, for Helios.

Unfortunately, Helios is not a drag-free satellite, and the cited results cannot be taken seriously for that spacecraft. At this time we do not have any precise estimates of what the non-gravitational forces will do to these curves; results of this sort should soon be available. Preliminary results indicate that the uncertainty in β for the perihelion distance of 0.3 AU increases from an accuracy of about 3% to an accuracy of 5% when the solar radiation and solar wind forces are assumed uncertain by 0.3% of the total radiation pressure force. This is a reasonable assumption when one considers the random fluctuations in the solar flux, which was actually measured on Mariners VI and VII.** Under this assumption, the estimate of the solar oblateness coefficient $J_{2\odot}$ becomes indeterminate.

The results here are consistent with our discussion of the Venus-Mercury mission and with

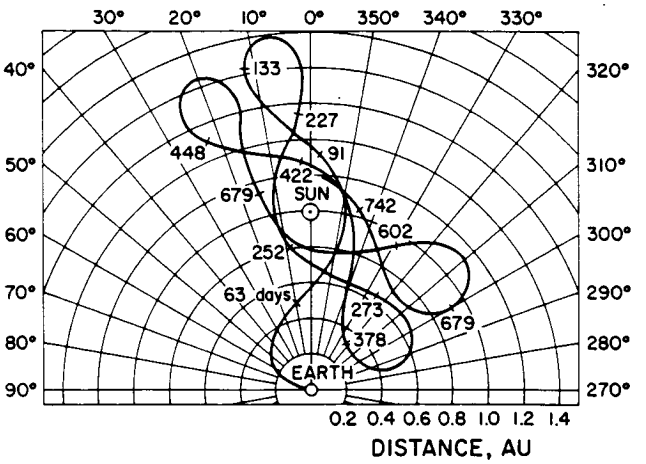


Fig. 4. Helios trajectory; launch date March 13, 1974, perihelion distance 0.2 AU

the comments in Section I about overcoming the unmodeled non-gravitational forces.

The Pioneer F and G missions will not contribute much to the parameter estimates, except insofar as the mass of Jupiter estimate is improved because there will be no range data and also because the tracking will be at a single frequency subject to charged particle limitations. The Venus survey missions may or may not have these limitations. In any case, they will represent only incremental improvements over the Mariner II and V data results.

The Grand Tour missions, with their 9 - 12 year lifetimes and penetration to 40 AU, will add considerably to our solar system data set, especially if S/X-band ranging and doppler tracking continuously operates on these missions.

The encounter data from the outer planet missions will be used to improve ephemerides and mass estimates of the planets. Such improvements are necessary in seeking the measurement of relativistic terms in the motion of the planets. The combining of these data with the two-century arc of optical data should yield an excellent description of motion in the solar system to an accuracy sufficient for gravitational theory testing. In this way, a measured time variation of the gravitational constant may also be sought. This analysis is essentially a determination and fit to a parameterized model of planetary ephemerides. The beginning of this effort has already been reported in Ref. 10.

To be investigated are the motions of the major planet satellites. Over the past few years, scientific interest in observations of the planetary satellites has begun to rank equally in importance to observations of the planets. We can thus expect that mission plans will require close passage to several bodies in the Jovian system. This characteristic makes our probe into a sort of mini-comet relative to the Jovian system. But we shall be

*The discussion of the Helios mission is due to J. Anderson and P. Esposito (see also Ref. 11).

**It is to be realized that solar flux measurements are not measurements of the forces on the spacecraft and do not take into account changing reflective areas and properties nor leaks from the gas tanks.

"on" the comet and receiving accurate tracking information from it, both relative to earth (radio) and relative to the satellite-planet system (optical). Since observations by TV pictures of the satellites against a star background are currently planned, "inertial" references will also be established. Being on such a comet with inertial, geocentric, and central body references should suggest new possibilities for gravitational theory tests. In particular, spacecraft navigation will use observations of these satellites, coupled with radio measurements for reducing spacecraft orbit determination errors. From the solution will come precise orbits of the satellites.

The Jovian and Saturnian systems have many interesting satellite orbital resonances. It seems reasonable that precise orbital knowledge of deviations from perfect resonances should permit estimation of parameters causing the perturbation to the Newtonian gravitational system. The magnitudes or relativistic terms, from General Relativity and from other hypotheses, should be evaluated so that orbit tracking tests can be devised. Nordtvedt's suggested effect from a breakdown in the Equivalence Principle might be observed in the Jovian system and post-Newtonian terms may be measured from precise satellite observations. Further calculations will have to examine the feasibility of detecting such effects and making relativistic parameter estimates from motion and observation in the zeocentric system. Orbiters, especially in the Jovian or Saturnian systems, might offer particularly exciting dynamics detecting capabilities. As mentioned above, using the parametrized version of the post-Newtonian metric may be particularly powerful in this study, for the many interacting effects may yield estimates on many of the parameters.

The outer planet missions will go through a near-solar superior conjunction approximately every 300 days. Certainly, the determination of γ can be confirmed and reconfirmed from this feature. The solar pressure force will become progressively smaller, thus reducing our limitation due to the orbit determination. The latitudinal and radial dependence of the relativistic time delay can be studied from these superior conjunctions, possibly to yield information connected with solar rotation.

The Grand Tour program is not yet rigidly defined. Options in the program plans allow for the missions to include adaptive trajectory design, entry probes, satellite flybys, and possibly even orbiters. Investigation of orbital motion effects on the satellites could be enhanced with close flybys of several satellites or on an orbiter mission.

V. Instrumentation

A. Low-Level Accelerometer

A three-axis, inertially referenced (either through gyros or with independent celestial referencing) accelerometer package capable of accuracies in the 10^{-11}-m/s^2 (10^{-12}-g) range could provide a means of overcoming the problem induced by unmodeled accelerations on the spacecraft. Since the main component of low-thrust forces is (not counting the rotational forces of attitude corrections) from the solar pressure acceleration (of order 10^{-7} m/s^2 in the inner planet region), a dynamic range of 10^{-8} to 10^{-12} g is necessary. It is also necessary to avoid the measuring of the centrifugal forces of the attitude corrections.

The subject of low-level accelerometers has been studied erratically over the past 15 years. The main hindrance has been the lack of incentive or application for such an instrument. One accelerometer with specifications near the 10^{-8} to 10^{-12} g level (Ref. 12) has, however, been constructed and used operationally on several different space programs. Although such accelerations have not yet been measured because of lack of application, the users feel that there are no fundamental limitations impeding their detection. Recently, Neiberling, Lesco and Berkopec (Ref. 13) reported use of that accelerometer on a spacecraft with measurements of about 2 μg to 1% accuracy.

Reinel of West Germany, has presented a method for laboratory and in-flight calibration of this accelerometer to the 10^{-7} g range on earth, and 10^{-9} g in space (Ref. 14), which suggests that the testing of low-level accelerometer instrumentation for space flights is feasible.* We do not intend to neglect or minimize the obvious engineering problems of continuous use of the instruments, calibrated and accurate to the pico-g level, over a mission lifetime of perhaps 10 years, or even for missions of 1 year. Further study and testing do, however, seem to be warranted if a reasonable priority on having such data (for tracking, navigation, or science purposes) is established.

C. S/X-Band Transponders and Antennas

Martin (Ref. 18) and Easterling (Ref. 19) have discussed the S/X-band ranging system, and Trask (Ref. 3) has referred to the advantages of dual-frequency tracking in removing charged-particle and dispersive media effects from the tracking data. The Mariner Venus-Mercury 1973 spacecraft, the Viking orbiter spacecraft and the outer planet (Grand Tour) spacecraft will have such capability. For the Venus-Mercury mission, the main benefit to gravitational theory testing of the S/X-band system will come at solar superior conjunction (if the mission is extended). Since the spacecraft passes relatively close to the sun, the engineering problems of keeping the antenna pointed at the sun near superior conjunction are not trivial, nor is the problem of designing the antenna mount so as to satisfy all the mission objectives.

References

1. Anderson, J. D., this proceedings.
2. Curkendall, D. W., this proceedings.

*A private communication from R. Cannon suggested that tests at Stanford University in the pico-g region were possible. Another such communication by V. Braginsky discussed construction of accelerometers capable of detecting 10^{-18} g.

3. Trask, D., "Earth Rotation and Transmission Media," talk presented at Conference on Experimental Tests of Gravitational Theories (CETGT), Pasadena, Calif., Nov. 11-13, 1970.

4. Cain, D. L., this proceedings.

5. Planetary Exploration Panel (NASA), Final Report, Vol. I, Aug. 1, 1969.

6. Beerer, J. G., and Burke, R. D., "Mariner Venus-Mercury 1973 Preliminary Mission Design," AIAA Paper 70-1049, AAS/AISS Astrodynamics Conference, Santa Barbara, Calif., Aug. 19-21, 1970.

7. Mariner Venus/Mercury 1973 Science Steering Group, Final Report, Internal Document 615-2, Jet Propulsion Laboratory, Pasadena, Calif., March 30, 1970.

8. Mariner Venus/Mercury 1973 Celestial Mechanics and Radio Science Team, Final Report, Jet Propulsion Laboratory, Pasadena, Calif., Apr. 7, 1970.

9. Friedman, L. D., "Potential of Interplanetary Spacecraft Data for Testing Gravitational Theories," Proceedings of the Conference on Scientific Applications of Radio and Radar Tracking in the Space Program, Apr. 9, 10, 1969; Technical Report 32-1475, Jet Propulsion Laboratory, Pasadena, Calif., July 1, 1970.

10. Lieske, J., et al., "Simultaneous Solution for the Masses of the Principal Planets from Analysis of Optical, Radar and Radio Tracking Data," IAU Colloquium No. 9, Heidelberg, Germany, Aug. 12-15, 1970.

11. Anderson, J. D., and Esposito, P. B., "Application of Spacecraft Tracking Data to Experimental General Relativity," AIAA Paper No. 70-1317, AIAA 7th Annual Meeting, Houston, Texas, Oct. 19-22, 1970.

12. Meldrum, N. A., Harrison, E. J., and Milburn, Z., "Development of a Miniature Electrostatic Accelerometer (MESA) for Low g Applications," Report No. BAC-60009-509, NASA CR-54137, Bell Aerosystems Co., Apr. 30, 1965.

13. Neiberling, W. C., Lesco, D. J., and Bercopec, F. D., "Comparative In-Flight Thrust Measurements of the SERT II Ion Thruster," AIAA Paper No. 70-1126, AIAA 8th Electric Propulsion Conference, Stanford, Calif., Aug. 31-Sept. 2, 1970.

14. Reinel, K., "Accelerometer Calibration in the Low g Range by Means of Mass Attraction," AIAA Paper No. 70-1030, AIAA Guidance, Control and Flight Mechanics Conference, Santa Barbara, Calif., Aug. 17-19, 1970.

15. Thorne, K. S., this proceedings.

16. Thorne, K. S., and Will, C. M., "Theoretical Frameworks for Testing Relativistic Gravity, I and II," Orange Aid Preprint Series OAP-217, in Nuclear, Atomic and Relativistic Astrophysics, Aug. 1970.

17. Martin, W., this proceedings.

18. Easterling, M., this proceedings.

Study of the Determination of the Coefficients β and γ of the Generalized Metric of Robertson and of the Dynamical Oblateness of the Sun

Christian Marchal
Office National d'Etudes et de Recherches Aérospatiales (ONERA), Paris

I. Motion of a Probe Around a Spherical Center of Attraction (in the General Relativity Theory)

Let us take the velocity of the light as unity. The most general static and spherical metric may be expressed in the form of Robertson (Ref. 1) as

$$ds^2 = f(\rho)\,dt^2 - g(\rho)\cdot(dx^2 + dy^2 + dz^2)$$

with

$$\rho = \text{radial distance} = \sqrt{x^2 + y^2 + z^2}$$

and f and g being two functions of ρ equal to one at infinity:

$$f(\rho) = 1 - \frac{2\alpha m}{\rho} + \frac{2\beta m^2}{\rho^2} + \cdots$$

$$g(\rho) = 1 + \frac{2\gamma m}{\rho} + \cdots$$

where m is a length equal to GM/c^2 (≈1.48 km in the case of the sun) and $\alpha = 1$ to have a Newtonian motion to the zero order.

The Einstein theory leads to $\beta = \gamma = 1$, but some other theories (Jordan, Thiry, Brans-Dicke) disagree. To test these theories, let us first study the motion according to the metric of Robertson.

We shall use the form of Schwarzschild, in which the coefficient of the circumferential differentials is one. Let us put

$$r = \text{new radial distance} = \rho\sqrt{g(\rho)}$$

$$= \rho + \gamma m + \cdots$$

ϕ = latitude

L = longitude

The differential element ds^2 becomes

$$ds^2 = F(r)\,dt^2 - G(r)\,dr^2 - r^2(d\phi^2 + \cos^2\phi\cdot dL^2)$$

with

$$\left.\begin{aligned} F(r) = f(\rho) &= 1 - \frac{2m}{r} + \frac{2m^2}{r^2}(\beta - \gamma) + \cdots \\ G(r) &= \left(1 + \frac{\rho}{2g}\frac{dg}{d\rho}\right)^{-2} = 1 + \frac{2\gamma m}{r} + \cdots \end{aligned}\right\} \text{according to } \alpha = 1$$

In the Einsteinian case, $F(r) = 1/G(r) = 1 - (2m/r)$ exactly.

The free motion is plane. If we assume that it is in the plane $\phi = 0$, the equations of motion are:

$$r^2 \frac{dL}{ds} = h = \text{integral of the angular momentum}$$

$$F(r)\frac{dt}{ds} = k = \text{constant}$$

$$G(r)\left(\frac{dr}{ds}\right)^2 + r^2\left(\frac{dL}{ds}\right)^2 + 1 - F(r)\left(\frac{dt}{ds}\right)^2 = 0$$

$$= \text{expression of } ds^2$$

The parameter t is a regular time (at a great distance and velocity zero), and the parameter s is the time of the probe itself. These equations may be integrated easily, to the first order, with the help of an auxiliary Keplerian motion. Let us put

$$p = \frac{h^2}{m} + m\left(2\beta - 4 - \frac{5\gamma}{2}\right) + \frac{h^2}{m}(1 - k^2)\left(1 + \frac{\gamma}{2}\right)$$

$$e^2 = k^4 + \frac{h^2}{m^2}(k^2 - 1) - \frac{\gamma m^2}{h^2} + (1 - k^2)(6 + 4\gamma - 2\beta) - \frac{h^2}{m^2}\left(1 - k^2\right)^2(2 + \gamma)$$

and let us consider the Keplerian orbit of eccentricity e and semi-latus rectum p around the same center of attraction. (We shall use the usual parameters n, a, b, p, e, v, E, M, and the gravitational constant $\mu\ (=n^2a^3)$ is here equal to m.)

We then obtain, in terms of p and e,

$$\left.\begin{aligned} h^2 &= mp + m^2\left[3 + 2\gamma - 2\beta + e^2 + \frac{\gamma e^2}{2}\right] + \text{order } \frac{m^3}{p} \\ k^2 &= 1 - \frac{m}{a} + \frac{m^2}{a^2} + \frac{\gamma m^2 e^2(1 + e^2)}{2p^2} + \text{order } \frac{m^3}{p^3} \\ a &= \frac{p}{1 - e^2} \end{aligned}\right\}.$$

And, in terms of the auxiliary true anomaly v, the motion is integrated (to the first order) into

$$\frac{1}{r} = \frac{1 + e\cos v}{p} - \frac{\gamma m e^2}{2p^2}\cos 2v$$

$$L = L_0 + v\left[1 + \frac{m}{p}(2 + 2\gamma - \beta)\right]$$

$$s = s_0 + \frac{M}{n}\left[1 + \frac{m}{2a} + \frac{\gamma m}{p}\left(1 - \frac{e^2}{4} + \frac{3e^4}{2(1 - e^2)}\right)\right] + \frac{\gamma m e^2}{np(1 - e^2)}\left[\left(\frac{3e^3}{2} - 2e\right)\sin E + \frac{2 - e^2}{4}\sin 2E\right]$$

$$t = t_0 + (s - s_0)\left(1 + \frac{3m}{2a}\right) + \frac{2me\sin E}{na}$$

i.e.,

$$t = t_0 + \frac{M}{n}\left[1 + \frac{2m}{a} + \frac{\gamma m}{p}\left(1 - \frac{e^2}{4} + \frac{3e^4}{2(1 - e^2)}\right)\right] + \frac{2me\sin E}{na} + \frac{\gamma m e^2}{np(1 - e^2)}\left[\left(\frac{3e^3}{2} - 2e\right)\sin E + \frac{2 - e^2}{4}\sin 2E\right]$$

with, as usual, $\tan E/2 = \sqrt{(1 - e)/(1 + e)}\tan v/2$ and $M = E - e\sin E$; M/n is the proper time of the Keplerian orbit and L_0, s_0, t_0 are three constants of integration.

These equations give the position, the proper time, and the regular time, that is to say all the parameters of motion, in terms of v; they are valid (to the first order) for any eccentricity and, with spherical symmetry, for any inclination.

It should be noted that

(1) β appears only in the expression of L.

(2) The only visible long-period effect is the secular advance of perihelion [angular velocity $= (nm/p)(2 + 2\gamma - \beta)$].

(3) The second order gives lengths of less than 1 cm (order m^2/p). We shall see that they are useless for our purpose.

II. Motion of a Probe Under the Influence of the Oblateness of the Sun

Let us consider the ordinary potential of an oblate body

$$U = -\frac{m}{r}\left[1 + J_{20}\frac{R^2}{r^2}\frac{3\sin^2\phi - 1}{2} + \cdots\right]$$

and let us neglect the terms after J_{20} (R being the radius of the sun). As usual, $J_{20} = -J_2$ (by definition); J_{20} is negative for an oblate body.

The motion of a free probe in this potential may be expressed, to the first order, with the help of an auxiliary rotating Keplerian orbit O. The elements of O are, as usual, n; a, b, p, e, i, Ω, ω, v, E, M, with

$$b = a\sqrt{1 - e^2}; \qquad p = a(1 - e^2);$$

$$n = \sqrt{\frac{m}{a^3}}; \qquad \tan\frac{E}{2} = \sqrt{\frac{1 - e}{1 + e}} \cdot \tan\frac{v}{2};$$

$$M = E - e\sin E$$

Let us rotate O around the polar axis and around its normal axis, with angular velocities proportional to that of v:

$$\Omega = \Omega_0 + v \cdot \frac{3J_{20}R^2}{2p^2}\cos i$$

$$\omega = \omega_0 + v \cdot \frac{3J_{20}R^2}{4p^2}(1 - 5\cos^2 i)$$

Ω_0 and ω_0 being two constants of integration.

Let us call r_1, ϕ_1, L_1 the coordinates of the point moving along the rotating orbit O:

$$\left.\begin{aligned} r_1 &= \frac{p}{1 + e\cos v} = a(1 - e\cos E) \\ \phi_1 &= \arcsin\left[\sin i \cdot \sin(v + \omega)\right] \\ L_1 &= \Omega + \arctan\left[\cos i \cdot \tan(v + \omega)\right] \end{aligned}\right\}$$

Hence,

$$\tan\phi_1 = \tan i \cdot \sin(L_1 - \Omega)$$

and

$$\cos(v + \omega) = \cos\phi_1 \cdot \cos(L_1 - \Omega)$$

Then the motion in the potential U of the oblate body may be expressed to the first order (Ref. 2), with six appropriate constants of integration a, e, i, Ω_0, ω_0, t_0, in terms of the description parameter v, by

$$r = r_1 - J_{20}R^2\left[\frac{r_1^2}{p^3}\left(\frac{e^2}{2}\cos 2i\cos 2v - \frac{e^3}{16}\sin^2 i\cos 3v\right) + \frac{\cos^2\phi_1}{2r_1}\right]$$

$$\phi = \phi_1 + J_{20}R^2\left[\frac{\sin 2\phi_1}{4r_1^2} - \frac{\sin^3 i \cdot \sin 2(v + \omega) \cdot \cos(L_1 - \Omega)}{8b^2}\right]$$

$$L = L_1 + \frac{J_{20}R^2}{p^2}\left[2e\cos i\sin v\left(1 + \frac{e}{4}\cos v\right) - \frac{1 - e^2}{8}\sin^2 i\sin 2(L_1 - \Omega)\right]$$

t = time

$$= t_0 + \frac{M}{n}\left[1 - \frac{3J_{20}R^2}{p^2}\left(\frac{\cos^2 i}{2} - \frac{e^2}{4} + \frac{e^4}{1 - e^2}\left(\frac{1}{2} - \frac{15\sin^2 i}{16}\right)\right)\right] + \frac{J_{20}R^2}{4nab}\sin^2 i \cdot \sin 2(v + \omega)$$

$$- \frac{J_{20}R^2 e}{np^2(1 - e^2)}\left[\begin{aligned} &\sin E\left(\frac{3e^4}{2} - 2e^2\right) + \sin 2E\left(\frac{e}{2} - \frac{e^3}{4}\right) \\ &+ \sin^2 i\sin E\left[1 - \frac{19}{8}e^2 + \frac{21}{16}e^4\right] \\ &+ \sin^2 i\sin 2E\left[\frac{e}{8} - \frac{3}{32}e^3\right] \end{aligned}\right]$$

These equations give the position and the time in terms of the true anomaly v. They are valid (to the first order) for any eccentricity and any inclination.

III. Estimation of β, γ, and J_{20} by the Study of the Motion of a Solar Probe

A. State-of-the-Art

(1) The deflection of light beams when they pass near to the sun is proportional to $1 + \gamma$ and is known with an accuracy of 20%:

$$\gamma = 1 \pm 0.40$$

(Today's improvements have even yielded $\gamma = 1 \pm 0.10$.)

(2) The advance of the perihelion of Mercury is related to β, γ, and J_{20}, as given by

$$2\gamma - \beta - 9 \cdot 10^3 J_{20} = 1.01 \pm 0.03 \quad (1)$$

(3) The motion of the node of Mercury on the solar equator (when planetary perturbations are removed) is not very well known. There is an upper limit of about $5 \cdot 10^{-5}$ on $|J_{20}|$.

(4) The geometrical oblateness of the sun, as measured by Dicke and Goldenberg, is $\epsilon = 5 \cdot 10^{-5}$.

Since the ratio of the centrifugal to the gravitational force at the equator is $\phi = \omega^2 R^3/GM = 2.1 \cdot 10^{-5}$, the hypothesis of a sun in equilibrium leads to:

$$J_{20} = \frac{\phi - 2\epsilon}{3} = -2.6 \cdot 10^{-5}$$

But if the mean rate of rotation of the sun is near the equatorial one, the same hypothesis leads to $0 \geq J_{20} \geq -5I\phi/4MR^2$, I being the moment of inertia of the sun (hence, $I \leq 2MR^2/5$). Then we must choose between two hypotheses:

(a) $J_{20} \simeq -$ some 10^{-6} and the geometrical oblateness of the sun is due to some phenomenon other than the Newtonian attraction and the centrifugal forces, for instance, the magnetic field or the radiation pressure (inside the sun).

(b) $J_{20} \simeq -2$ or $-3 \cdot 10^{-5}$. This case requires that the inside of the sun rotate with a mean rate of rotation equal to about twice that of the surface. (A uniform rate of rotation leads to $2\phi \leq 4\epsilon \leq 5\phi$, which is not the case.)

B. Measure of γ

Let us call C the angle between the radial vector of the earth and the radial vector of a probe (Fig. 1), and let us construct the Euclidian triangle sun-earth-probe either with the distances r or with the distances ρ defined in Section I.

The radio-electric and the laser measurements of the distance earth-probe give neither the distance d_r nor the distance d_ρ of Fig. 1, but, due to the curvature of space, a length $\mathcal{L}$ is obtained (for $ds \equiv 0$):

$$\mathcal{L} = d_r - m(1+\gamma) \log\left[\tan\frac{A}{2}\tan\frac{B}{2}\right]$$

$$- \gamma m (\cos A + \cos B) + \text{order } \frac{m^2}{r_E \sin A}$$

$$= d_\rho - m(1+\gamma) \log\left[\tan\frac{A'}{2}\tan\frac{B'}{2}\right]$$

$$+ \text{order } \frac{m^2}{\rho_E \sin A'}$$

One can verify that, since $r = \rho + \gamma m +$ order m^2/ρ, the two formulas give concordant results.

This provides an easy way to measure γ accurately when A and B vary near zero; for instance, with a drag-free probe on a 6- or 8-month orbit tangent to that of the earth, γ can be measured with an accuracy of 10^{-3} if one uses laser measurements (and if there is an accurate clock

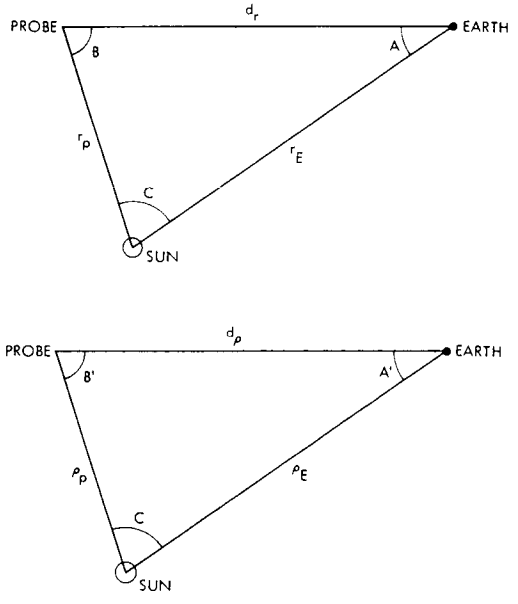


Fig. 1. Measure of γ

in the probe). The accuracy is reduced somewhat if only radio-electric measurements are used because (1) the accuracy of the measure of $\mathcal{L}$ becomes 10 to 15 m instead of 3 m, and (2) $\mathcal{L}$ becomes very difficult to measure through the solar corona.

C. Measure of β and J_{20}

The parameters β and J_{20} can be measured by their action on the motion of a solar probe. For instance, if the orbital plane of the probe is near the solar equator and if planetary perturbations are removed, the mean rate of rotation of the perihelion (n being the mean angular motion of the probe) is

$$n\left[\frac{m}{p}(2 + 2\gamma - \beta) - \frac{3J_{20}R^2}{2p^2}\right]$$

Practically, the first characteristic length $\mathcal{L}_1$--i.e., the greatest measurable length related to γ, β, and J_{20}--is, for one revolution, the length shown in Fig. 2 (the Keplerian motion being closed, this measure avoids the errors upon the orbital elements). This length is about

$$\mathcal{L}_1 = \pi e\left[2m(2\gamma - \beta) - 3J_{20}\frac{R^2}{p}\right]\left(1 + \frac{3e^2}{2} + \cdots\right) \quad (2)$$

Since $\mathcal{L}_1$ is a length of only a few kilometers, the measure of β and J_{20} requires an accurately compensated, drag-free probe: the remaining unknown long-period perturbations must be of the order of 10^{-12} m/s^2 at most; this can probably be achieved by a drag-free probe in the ecliptic plane

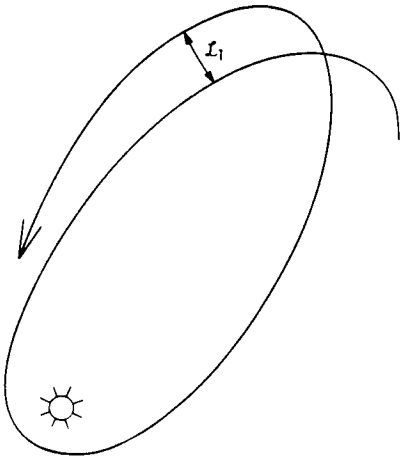


Fig. 2. Measure of β and J_{20}

and spinning in that plane. The out-of-plane component of the perturbing accelerations remains important, of course (10^{-10} or 10^{-9} m/s^2), but this has practically no effect upon the measurement of $\mathcal{L}$ and $\mathcal{L}_1$. (This is the reason why out-of-ecliptic orbits must be avoided.)

Note that since the second characteristic length is generally much smaller than $\mathcal{L}_1$ (Ref. 3), the best way to obtain β and J_{20} is also to use Eq. (1) resulting from the advance of the perihelion of Mercury. However, the 6-month orbit tangent to that of the earth must then be avoided because this orbit has almost the same semi-latus rectum as that of Mercury (0.413 and 0.371 AU, respectively, compared with 0.69 AU for an 8-month orbit), and the ratio of the coefficients of β and J_{20} is proportional to this semi-latus rectum (see Eq. 2).

Finally, three types of experiments with a drag-free probe can be envisioned to measure γ, β, and J_{20}, as shown in Table 1.

Table 1. Experiments to measure γ, β, and J_{20}

Mission	Spacecraft composition	Remaining perturbing acceleration, m/s^2	Error in			Orbit
			γ	β	J_{20}	
I	Laser + clock + dual frequency for the radio-electric measures	10^{-11}	10^{-3}	15%	$3\cdot10^{-5}$	Preferably 6 or 8 months
II	Dual-frequency + rotation of the probe	10^{-12}	2%	3%	$3\cdot10^{-6}$	8 months
III	Mission I + Mission II	10^{-12}	10^{-3}	1%	$3\cdot10^{-6}$	8 months

It should also be noted that, actually, the major source of errors in the measurement of γ by a laser and a clock came from the drift of the clock -- about 10^{-13} at best, or 2.6 m per day in the measurement of $\mathcal{L}$.

IV. Conclusion

It is possible to measure the coefficients β and γ of the metric of Robertson and J20 of the sun with a drag-free probe on an heliocentric orbit of great eccentricity if the remaining unknown long-period perturbing acceleration is of the order of 10^{-11} or 10^{-12} m/s^2.

The second-order effects are too small to be useful.

References

1. Tolman, R. C., Relativity, Thermo-Dynamics, and Cosmology, Clarendon Press, Oxford, London, p. 205.

2. Marchal, C., "Calcul du mouvement des satellites artificiels à partir des solutions exactes du problème des deux centres fixes." Bulletin Astronomique, 1, Part 3, 1966.

3. Marchal, C., "Sur la détermination de l'aplatissement dynamique du soleil et des coefficients β et γ du tenseur de Robertson." La Recherche Aérospatiale, No. 2, p. 57-62, 1970.

4

An Eötvös Experiment in Earth Orbit

Philip K. Chapman
NASA Manned Spacecraft Center and M. I. T. Measurement Systems Laboratory
Andrew J. Hanson
M. I. T. Measurement Systems Laboratory

I. Introduction

It is an axiom in general relativity that the ratio of passive gravitational mass to inert mass

$$k = \frac{m_p}{m_i} \qquad (1)$$

is a constant for all bodies, independent of their composition. This condition is also required in virtually all other theories of gravitation, although in some (for example, the Brans-Dicke cosmology, Ref. 1), the ratio is allowed to vary from point to point in space-time. The assumption of composition-independence for k is a statement of the Weak Principle of Equivalence.

The evidence for this hypothesis consists primarily of experiments of the Eötvös type. Given two bodies, of materials A and B, these experiments may be regarded as measurements of the Eötvös ratio

$$\eta(A, B) = 2\frac{k(A) - k(B)}{k(A) + k(B)} \qquad (2)$$

The most precise experiment of this type to date is that of Roll, Krotkov and Dicke (Ref. 2), which showed that, for gold and aluminum, the ratio is not larger than 10^{-11}. In order to understand the motivation for improving this remarkable accuracy, it is necessary to consider the various forms of energy which together make up the mass of any body or system of bodies. These are:

(1) The elementary particles which make up atoms, particularly protons, neutrons, and electrons.

(2) Energy stored in the strong interaction, which binds protons and neutrons together to make nuclei.

(3) Energy stored in the electromagnetic interaction, which binds electrons and nuclei together to make atoms, atoms together to make molecules and crystal lattices, and which causes protons in a nucleus to repel each other.

(4) Energy stored in the weak interaction, which is responsible for β-decay processes.

(5) Energy stored in the gravitational interaction, which binds planets and stars into solar systems and solar systems into galaxies, but which is very weak indeed on the laboratory scale.

With the exception of the gravitational self-energy, the fractional contribution of each of these energy forms to the overall mass of a body depends strongly on its chemical composition. Eötvös experiments using bodies of widely different composition therefore allow conclusions to be drawn concerning the passive gravitational mass of some of these fundamental forms of energy. Schiff

(Ref. 3) has shown that experiments to date indicate with reasonable accuracy that protons, neutrons, electrons (and their antiparticles), strong interaction energy, and electromagnetic energy do not have anomalous gravitational behavior.

The present experiment is intended to investigate the gravitational interactions of mass energy stored in the weak interaction. If successful, this will leave only the gravitational self-interaction untested. As has been pointed out recently by Nordtvedt (Ref. 4), advances in radar and laser tracking of celestial bodies may soon allow this aspect to be investigated as well. Null results in both these tests would provide strong support for present conceptions of gravitation and increase the impetus to find a satisfactory interpretation of inertial mass in Machian terms. On the other hand, anomalous results might open the door to a totally new understanding of the field. These experiments therefore have an importance far transcending mere improvements in accuracy.

The weak interaction is of particular interest because it is known to violate well-established physical principles, those of parity conservation and charge-conjugation symmetry (Refs. 5 and 6). While no theoretical arguments are available on this point, it is tempting to guess that the weak interaction might violate the Equivalence Principle as well.

II. The Magnitude of Possible Eötvös Anomalies

A. The Weak Interaction

In order to estimate the accuracy required in the experiment, let us assume that energy stored in the weak interaction has inert mass but no passive gravitational mass. If $\alpha(A)$ is the fractional contribution of weak interaction energy to the overall mass energy of body A, and if all other forms of energy do not exhibit anomalous gravitational behavior, then Eq. (2) yields

$$\eta(A, B) = \alpha(B) - \alpha(A) \tag{3}$$

to first order in these small quantities.

After the discovery of non-conservation of parity in weak interactions, the theory of these interactions was substantially improved (Ref. 6), but unfortunately it is still not possible to calculate accurately the contribution of the binding energy to the mass density of matter. However, Blin-Stoyle (Ref. 7) has estimated that weak-interaction energy is approximately one part in 10^7 of the strong-interaction binding energy in a nucleus, a result which appears to have been experimentally confirmed by Lobashov et al. (Ref. 8). On this basis, if A and B are gold and aluminum, Eq. (3) gives

$$\eta(Au, Al) \simeq 10^{-11} \tag{4}$$

In view of the uncertainties in this calculation, it is evident that the accuracy of the experiment of Roll, Krotkov, and Dicke must be improved by several orders of magnitude in order to give any reliable information on the gravitational behavior of weak-interaction energy. The design goal in the present experiment is one part in 10^{14}.

It should be noted that, if an anomaly in the comparison of gold and aluminum were detected at this accuracy, it could be construed as indicating a minor violation of the Equivalence Principle by one of the other forms of mass energy. This possibility could be investigated by carrying out the experiment using several different pairs of materials, except that a weak-interaction effect will not be separable from a small strong-interaction effect until the theory of these interactions has been improved further.

B. The Gravitational Self-Energy

For a spherical body, the fractional contribution of the energy in the gravitational field to the overall inert mass is (Ref. 4)

$$\alpha_g \simeq \frac{Gm}{rc^2} \tag{5}$$

where r is the radius of the body, G is the gravitational constant, and c is the speed of light. This number is of order 10^{-25} for laboratory bodies, much too small for any possible anomaly to be detected. However, it is about 10^{-9} for the earth and 10^{-5} for the sun, so that solar-system experiments are much more promising.

C. Spin-Orbit Interactions

It is predicted by general relativity (Refs. 9 and 10) that a spinning body does not follow exactly a geodesic in the Riemannian space-time determined by neighboring massive bodies, when it is acted on by no forces. In other words, there is an anomalous gravitational force acting on a spinning body: if an Eötvös experiment were carried out which compared a gyroscope with a non-spinning body, a positive result would be obtained. The magnitude of the effect has been calculated by Schiff (Ref. 11) as

$$\eta_{spin} \simeq \frac{H \cdot \Omega}{mc^2} \tag{6}$$

for the case of an apparatus in earth orbit, with orbital angular velocity $\underline{\Omega}$. Here $\underline{H}$ is the angular momentum of the gyro and m its mass. For a large gyro in low orbit, with a wheel radius at 1 m, spinning at 10,000 rpm (close to the value at which centrifugal stresses would tear it apart), the Eötvös ratio would be of order 10^{-17}.

Careful design and a sufficiently large apparatus might make this effect experimentally detectable, thereby allowing another test of general relativity. Although such a device would, in principle, be capable of detecting the difference between inertial acceleration and a gravitational field, this would not necessarily violate the Strong Principle of Equivalence, which is limited to infinitesimal regions (because, in a finite region, it is always possible to differentiate between gravitation and acceleration by measurement of gravity gradients). A spinning body cannot be

infinitesimal, even in principle, essentially because the periphery must move at less than the speed of light. Thus, this experiment would constitute a test of the consequences, and not the foundations, of general relativity.

D. The Electromagnetic Radiation Reaction

At first sight, radiation from an accelerated charged particle seems to violate the Equivalence Principle. How does a charged body, at rest in a terrestrial laboratory, know that it is experiencing a gravitational field and not an acceleration, so that it does not radiate? The usual answer, of somewhat dubious validity, is that radiation phenomena must occur over distances of at least a wavelength, and hence cannot be considered as an experiment in an infinitesimal region. In any case, it is of interest to calculate the anomalous Eötvös ratio which would be expected if a charged body were compared with an uncharged one in earth orbit, because of the radiation reaction.

The power radiated by an accelerated charged particle is given by the Larmor formula (Ref. 12)

$$P = -m\tau\dot{v}^2 = \underline{F}_{rad} \cdot \underline{v} \qquad (7)$$

where

$$\tau = \frac{2}{3}\frac{e^2}{mc^3} \qquad (8)$$

e being the charge and m the mass. In a circular orbit,

$$\dot{\underline{v}}^2 = (\underline{v} \times \underline{\Omega})^2 = v^2\Omega^2 \qquad (9)$$

so that Eq. (7) may be written

$$[\underline{F}_{rad} + m\tau\Omega^2\underline{v}] \cdot \underline{v} = 0 \qquad (10)$$

or

$$\underline{F}_{rad} = m\tau\Omega^2(\underline{R} \times \underline{\Omega}) = -m\underline{g} \times \tau\underline{\Omega} \qquad (11)$$

The radiation reaction force may thus be considered as a drag, opposite to the orbital velocity, whose magnitude is less than the weight of the particle by the factor $\tau\Omega$. The characteristic time τ has its maximum value, 6.26×10^{-24} s, in the case of an electron, which, in low earth orbit, would thus exhibit an anomalous Eötvös ratio of magnitude about 10^{-26}. The effect is therefore negligible.

III. Design of a Sensitive Eötvös Experiment

If it is true that weak interaction energy violates the Equivalence Principle, it appears that the phenomenon may be detectable after a relatively modest improvement in the accuracy of the Eötvös experiment. The other possible phenomena discussed are too small to be detected, at least in a terrestrial or orbiting laboratory environment, without major advances in technology. They will therefore not be considered further here.

The characteristic of Eötvös experiments which distinguishes them from earlier investigations of the equivalence of inert and gravitational mass [such as Galileo's famous and perhaps apocryphal experiment at the Leaning Tower of Pisa, or early attempts to determine whether the period of a pendulum depended on the material of the bob] is that they are null experiments, taking advantage of situations in which, if the Weak Principle of Equivalence is rigorously true, there exist a balance between inert and gravitational forces. Eötvös himself (Ref. 13) used the fact that a plumb bob suspended in the laboratory does not generally lie exactly along the local direction of gravitation, because of the rotation of the earth, so that, in the horizontal plane so determined, there is a small component of centrifugal force (about 1.7 cm/s^2 at 45° latitude) directed toward the equator, and an equal but opposite component of gravitation. Roll, Krotkov, and Dicke (Ref. 2) used the balance between the gravitational pull of the sun (about 0.6 cm/s^2) and the centrifugal force due to the orbital motion of the earth.

The advent of space technology has made possible a substantial improvement in the experimental conditions. In low earth orbit, a balance exists between gravitational and centrifugal accelerations whose magnitude is of order 1 g, about a thousand times greater than those available in terrestrial experiments. Furthermore, the free-fall environment might allow great reduction in the problems associated with suspension of the apparatus, such as coupling to ambient noise. In view of the importance of the experiment to the foundations of general relativity, it is an obvious candidate for performance in space.

The most obvious technique for carrying out this experiment in orbit is to use an adaptation of the rotational balances which have been successful in terrestrial experiments. However, gravity-gradient torques may swamp those due to Eötvös forces unless great care is taken to make the system inertially symmetrical.

Without going into the details of the design of such a balance, some general conclusions may be drawn. For simplicity, consider a system with one of its principal axes along the orbit normal. The gravity gradient torque is then along the orbit normal and of magnitude (Ref. 14)

$$T_g = -\frac{3}{2}\Omega^2\Delta I \sin 2\theta \qquad (12)$$

where ΔI is the difference in the moments of inertia about the principal axes lying in the orbit plane and θ is the angle between one of these axes and the local vertical.

Let us suppose that the system is constructed of two different materials, A and B, for which the ratios of passive to inert mass are k(A) and k(B), respectively. The Eötvös torque about the center of mass is then

$$\underline{T}_e = -\underline{g} \times \left[k(A) \int_A \underline{r}\, dm + k(B) \int_B \underline{r}\, dm \right]$$

$$\simeq -\eta \underline{g} \times \left[\int_B \underline{r}\, dm \right]$$

$$= -\eta m \underline{g} \times \underline{a} \qquad (13)$$

where g is the local gravitational field, m is the total inertial mass of the system, and

$$\underline{a} = \frac{1}{m} \int_B \underline{r}\, dm \qquad (14)$$

With appropriate symmetry, $\underline{a}$ will lie in the orbit plane, but generally not along one of the principal axes. Then, $\underline{T}_e$ is along the orbit normal and has the value

$$T_e = -\eta m g a \sin \phi \qquad (15)$$

where ϕ is the angle between $\underline{a}$ and the local vertical.

In order to estimate the difficulty of inertially balancing the system, a reasonable condition to impose is that the frequency of gravity gradient oscillations be less than that due to Eötvös torques. If I is the moment of inertia about the orbit normal, this may be written

$$\frac{\Delta I}{I} < \frac{\eta R a}{3 b^2} \qquad (16)$$

where R is the radius of the orbit and b the radius of gyration about the orbit normal. For $\eta = 10^{-14}$, low earth orbit, and $a \simeq b \simeq 10$ cm, we obtain

$$\frac{\Delta I}{I} < 2 \times 10^{-7} \qquad (17)$$

Inertial balancing to this accuracy is difficult but not impossible, especially as it may be possible to exploit the double-angle dependence of Eq. (12) and a design difference in the directions of $\underline{a}$ and the principal axis to effect balancing on orbit.

Even if this condition can be met, however, it must be recognized that the torques under consideration are extraordinarily weak. From Eq. (15), the angular frequency of oscillation due to the Eötvös forces is

$$\omega_e^2 = \frac{\eta R a}{b^2} \Omega^2 \qquad (18)$$

which, under the conditions assumed in (17), gives a period of some 2000 h! It is clear that very great care indeed would have to be taken to protect the balance from non-gravitational disturbing torques. Furthermore, it is highly desirable to design an experiment from which the data can be extracted more quickly.

For these reasons, an alternative technique is under study, in which an attempt is made to measure Eötvös forces directly, without converting them into torques. The design was motivated by the following considerations:

(1) Modulation of the Eötvös forces, to minimize the duration of the experiment by operation in a higher frequency region, to allow resonance to be used to enhance the Eötvös effect, and to allow frequency discrimination of many important noise sources.

(2) Minimization of gravity gradient forces by placing the centers of mass of the bodies of different composition under test as nearly as possible at the same point in space.

(3) Exploitation of the tensor properties of residual gravity gradients to distinguish them from the phenomenon under study.

As shown in Fig. 1, the proposed apparatus consists basically of an aluminum wheel, spinning with inertial angular velocity $\underline{\omega}$, which lies along the normal to the plane of a low, circular orbit about the earth. This orientation is, of course, stable under gravity gradient effects. A sensitive accelerometer is mounted radially in the plane of the wheel; it contains two proof masses, one of gold and one of aluminum, suspended coaxially and independently by electrostatic forces, as shown in Fig. 2. The null position of the proof masses is close to the overall center of mass.

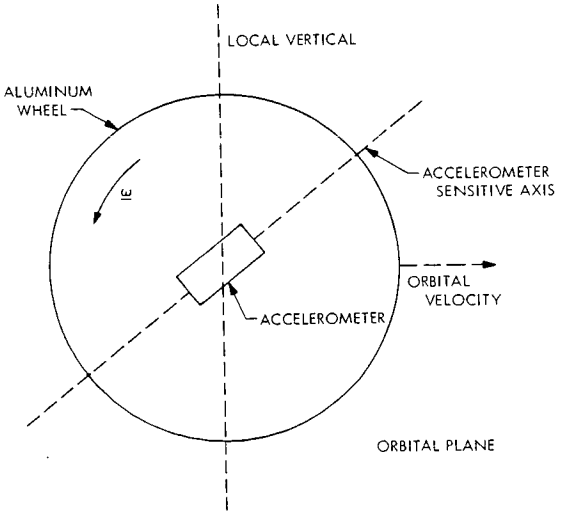


Fig. 1. The orbital Eötvös apparatus

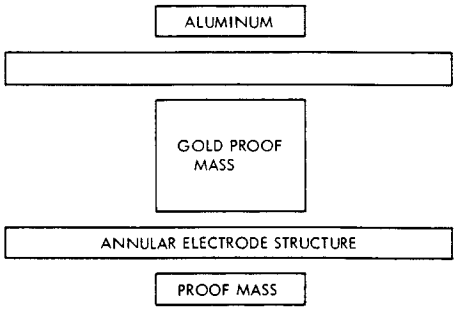


Fig. 2. Accelerometer structure

Let m, M be the inert and gravitational masses of the gold proof mass and $\underline{r}$ its position vector relative to the center of mass of the system. It is assumed that the mass of the wheel is so large that motion of the proof mass does not appreciably shift the position of the CM within the system. Since the axis of the accelerometer may not, in general, pass through the CM, write

$$\underline{r} = \underline{x} + \underline{d} \tag{19}$$

where $\underline{d}$ is the position vector of the null of the accelerometer. The equation of motion along the accelerometer axis is then

$$m\ddot{x} = -mq + m\omega^2 \underline{r} \cdot \underline{i} + (M\underline{g} - m\underline{g}_o) \cdot \underline{i} - m\underline{f} \cdot \underline{i} \tag{20}$$

where mq represents a servo restraint force applied to the proof mass, $\underline{i}$ is a unit vector along the sensitive axis, $\underline{g}$ is the gravitational field of the earth at the proof mass, $\underline{g}_o$ is the inertial acceleration of the CM due to gravitation, and $\underline{f}$ is any disturbing acceleration applied at the CM. In this equation, terms which are definitely perpendicular to the accelerometer axis, such as that due to the Coriolis acceleration, have been dropped.

Without loss of generality, the ratio of passive to inertial mass for aluminum may be taken as unity; this is equivalent to defining the gravitational constant as that measured in a Cavendish experiment using aluminum masses. If there is too little gold or other non-aluminum material in the system to affect the overall ratio of passive to inert mass significantly, then

$$\underline{g}_o = -\Omega^2 \underline{R} \tag{21}$$

where $\underline{R}$ is the geocentric position vector of the CM. Expanding the gravitational field in a vector Taylor series about the CM,

$$M\underline{g} - m\underline{g}_o = m[k\underline{g} - \underline{g}_o] = m[\eta \underline{g}_o + k(\underline{r} \cdot \nabla)\underline{g}|_o + \ldots] \tag{22}$$

where $k = M/m$ and $\eta = k-1$ is the Eötvös ratio for gold and aluminum. In tensor notation, the quantity $(\underline{r} \cdot \nabla)\underline{g}|_o$ is equivalent to

$$r_j \left[\frac{\partial g_i}{\partial R_j}\right]_o = r_j \frac{\partial}{\partial R_j}\left[-GM_\oplus R^{-3} R_i\right]_o = -\Omega^2\left[\delta_{ij} - 3R^{-2}R_i R_j\right] r_j \tag{23}$$

where $M_\oplus$ is the mass of the earth and δ_{ij} is the unit tensor. Substituting in Eq. (20),

$$\ddot{x} = -q + \omega^2 x + \eta g_o \cdot \underline{i} - k\Omega^2\left[1 - 3R^{-2}(\underline{R} \cdot \underline{i})^2\right] x - \underline{f} \cdot \underline{i} + (\omega^2 - k\Omega^2)\underline{d} \cdot \underline{i} + 3R^{-2}(\underline{R} \cdot \underline{d})(\underline{R} \cdot \underline{i})\, k\Omega^2 \tag{24}$$

The term in $\underline{d} \cdot \underline{i}$ is a constant, which may be minimized by choosing $\underline{d}$ as nearly as possible along the spin axis of the wheel. Since this axis is nominally along the orbit normal, such a choice also minimizes $\underline{R} \cdot \underline{d}$. Then, since $\underline{R}$ and $\underline{g}_o$ are rotating at the orbital angular velocity $\underline{\Omega}$, and $\underline{i}$ is rotating at $\underline{\omega}$ about the same axis, the equation becomes

$$\ddot{x} + q - \left[\omega^2 + \tfrac{1}{2}\Omega^2 + \tfrac{3}{2}\Omega^2 \cos 2(\omega - \Omega)t\right] x = \eta g \cos(\omega - \Omega)t - f' \tag{25}$$

where, to an entirely sufficient accuracy, k has been taken as unity. Neglecting constant terms,

$$f' = f \cos\left[(\omega - \Omega)t + \beta_1\right] + \tfrac{3}{2}\Omega^2 d \cos\left[2(\omega - \Omega)t + \beta_2\right] \tag{26}$$

where f and d are now the components of these vectors in the plane of the wheel, β_1 is the angle between f and the local vertical, and β_2 is the angle between d and the accelerometer axis.

If q is a simple spring restraint, $q = Kx$, then Eq. (25) is recognized as a standard Mathieu equation.

For the aluminum proof mass (which has $\eta \equiv 0$), an identical calculation yields the equation of motion.

$$\ddot{x}' + q' - \left[\omega^2 + \frac{1}{2}\Omega^2 + \frac{3}{2}\Omega^2 \times \cos 2(\omega - \Omega)t\right]x' = -f' \qquad (27)$$

On the right-hand side, it is assumed that the null positions of the gold and aluminum proof masses are identical, a condition which may be maintained to within a fraction of the wavelength of light because of the coaxial design.

The quantities which can be measured directly in this system are x, x', and y = x - x'. By subtraction, the equation of relative motion of the two proof masses is

$$\ddot{y} + q - q' - \left[\omega^2 + \frac{1}{2}\Omega^2 + \frac{3}{2}\Omega^2 \cos 2(\omega - \Omega)t\right]y = \eta g \cos(\omega - \Omega)t \qquad (28)$$

The orbital angular velocity is of order 0.01 rpm, whereas the wheel angular velocity may be of order 100 rpm. The time-varying coefficient in (27) and (28) is then some 8 orders of magnitude smaller than ω^2, and may be neglected in the preliminary servo design (although it must be taken into account in the final analysis, since it may cause instability). The Laplace transform of (27) is then

$$\left[s^2 + A(s) - \omega^2\right]x' = -f' \qquad (29)$$

where the transform of q' has been taken as A(s)x'. The transform of q is chosen to be

$$q(s) = B(s)y + (1 + \delta)A(s)x' \qquad (30)$$

so that, from (28),

$$y = \left[s^2 + B(s) - \omega^2\right]^{-1}[\epsilon - H(s)f'] \qquad (31)$$

where

$$\epsilon = \eta g s\left[s^2 + (\omega - \Omega)^2\right]^{-1} \qquad (32)$$

is the Laplace transform of the very small Eötvös acceleration which it is desired to detect, and

$$H(s) = \left[s^2 + A(s) - \omega^2\right]^{-1}\delta A(s) \qquad (33)$$

The block diagram of this system is shown in Fig. 3. Using the differential displacement y as the output, the function of the aluminum proof mass system is to filter disturbing accelerations. Common electronics can be used in the x'- and y-loops for A(s), so that the value of δ depends only on the differences in sensitivity of the displacement detectors and the force transducers used to command the two proof masses: a value as low

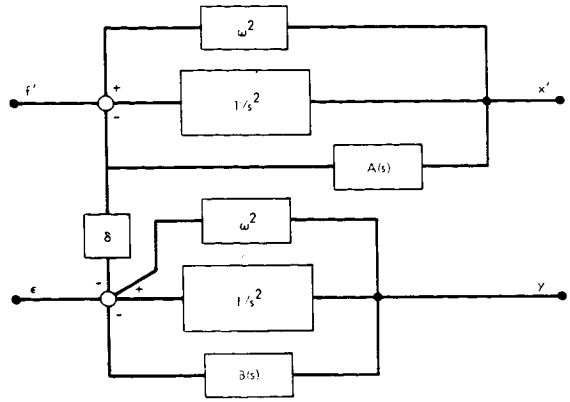


Fig. 3. Block diagram of the accelerometer servo

as 0.001 may be obtainable. The optimum choice for A(s), consistent with stability in the x'-loop, must await specification of the statistics of the disturbing acceleration and is thus dependent on the overall vehicle design.

It is a simple matter to assure stability of the x'-loop, but stability in the y-loop may be affected by the time-varying term in the equation, which can also cause spurious responses that might be mistaken for the Eötvös effect. To investigate this, it is assumed that the disturbing accelerations have been reduced to an acceptable value, and Eq. (28) is written as

$$\ddot{y} + \left[w^2 - \Lambda \cos 2(\omega - \Omega)t\right]y = \eta g \cos(\omega - \Omega)t \qquad (34)$$

where w^2 has the Laplace transform

$$w^2(s) = B(s) - \omega^2 - \frac{1}{2}\Omega^2 \qquad (35)$$

and Λ is used in the time-varying coefficient to allow for the possibility of computing, from external measurements, a compensation term in the servo, if this should prove necessary.

For maximum resonance, the first choice for $w^2(s)$ is clearly a real quantity

$$w^2 = (\omega - \Omega)^2 \qquad (36)$$

but such an ideal, infinite-Q case results in one of the unstable regions for solution of the Mathieu equation (Ref. 15). Examination of the stability boundaries shows, however, that stability can be achieved by operating off the resonant frequency by a fractional amount

$$\frac{w - (\omega - \Omega)}{(\omega - \Omega)} > \frac{1}{4}\frac{\Lambda}{(\omega - \Omega)^2} \qquad (37)$$

The overall Q of the system must then be limited to

$$\frac{Q < 4(\omega - \Omega)^2}{\Lambda} \simeq 10^8 \qquad (38)$$

where the numerical estimate is based on a wheel spin of 100 rpm and Λ has its natural value of $3\Omega^2/2$. On the other hand, a reasonable value of Q, within the limits of mechanical feasibility, is 10^4-10^5; this would allow starting transients to die away in a period of a few hours.

In the steady state, the amplitude of the displacement due to the Eötvös acceleration is

$$\xi = \frac{\eta g Q}{(\omega - \Omega)^2} \qquad (39)$$

which, for $\eta = 10^{-14}$ and $Q = 10^5$, amounts to 10^{-8} cm, well within the capability of electrostatic or interferometric displacement detectors.

More sophisticated servo designs to ensure stability and maximize the sensitivity are currently under study. There are many other error sources which have not been considered here, such as nutation and imbalance of the wheel, deviation of the spin axis from the orbit normal, eccentricity of the orbit, and thermal noise in the accelerometer. Current estimates, however, lead to the expectation that the system should be capable of sufficient accuracy to determine whether or not energy stored in the weak interaction has gravitational mass, if external disturbing accelerations are reduced to a level where they can be effectively filtered.

IV. Spacecraft Design Considerations

It is clear from Eqs. (25) and (26) that disturbing accelerations which are constant in a local-vertical orbital reference frame masquerade directly as Eötvös accelerations, especially any component along the local vertical. An example is residual atmospheric drag on the system: while it may be possible to distinguish the major component along the orbital velocity because of its phase difference from an Eötvös effect, the vertical component can be compensated only by means of the filtering capabilities of the aluminum proof mass system. In order to achieve this, H(s) in Eq. (33) should be made as small as possible, by choice of δ, and in addition be designed for strong rejection at the frequency $(\omega - \Omega)$. Since, in a typical spacecraft at, say, 300-nmi altitude, the vertical component of atmospheric drag is expected to be below 10^{-9} g, it does not appear that this disturbance should upset the measurement at the accuracy under consideration. Similar considerations apply to the other known constant disturbances (e.g., the constant component of solar radiation pressure). At the present time, then, it does not appear to be necessary to take elaborate precautions, such as flying the apparatus unsupported in a servo-controlled windshield (i.e., in a "pure gravity orbit").

Accelerations which appear at a frequency ν in the orbital frame will produce sidebands at frequencies $(\omega - \Omega) + \nu$ in the frame of the accelerometer. These can be filtered by choice of H(s) in the x'-loop as well as by the resonant characteristics of the y-loop itself. With a reasonable choice of Q, only those disturbances with frequencies longer than several hours will appear in the passband of the y-loop. In particular, accelerations which are fixed in inertial space or which, as seen in the orbital frame, appear at orbital frequency, are well outside the passband.

It therefore appears possible to mount the apparatus in gimbals attached to a suitable spacecraft. Indeed, it may be possible to fly the experiment, mounted to the wall in this way, inside a manned space station. In this case, the peak accelerations are expected to be of order 10^{-5} g, with periods measured in seconds. Suitable shock mounts for the gimbals and appropriate servo design may reduce these to effective levels low enough not to interfere with the measurement.

Even discounting the professional bias of one of the authors of this paper, a manned vehicle offers great advantages for this experiment. If flown unmanned, the very sensitive accelerometers would require caging to withstand the boost environment, and some form of orbital gyrocompassing, with attendant horizon trackers, etc., would be necessary to erect the wheel spin axis to the orbit normal. Furthermore, if it is desired to measure the Eötvös ratio of several materials, relative to aluminum, a mechanism would be required for replacing the gold proof mass. Finally, experience with low-level accelerometers in orbit indicates that problems may arise which are very difficult to solve in an inflexible, automatic system. For example, an engineer on the spot could optimize the servo design for the actual disturbing accelerations encountered, instead of relying on statistical estimates obtained from other space missions. It should also be noted that the performance required of this system is of an order which simply cannot be tested in the terrestrial environment. It is not possible to operate accelerometers at levels much below 10^{-6} g on earth, because of seismic noise, leveling difficulties, and cross-coupling from the strong suspension forces required for the proof mass. Similarly, it is probably impossible to balance the wheel sufficiently well in a terrestrial facility. This experiment can make excellent use of the capabilities of man, allowing him to make final design choices as a result of on-orbit experience, tune up and deploy the system in its optimum configuration, monitor the performance and investigate any anomalies observed, and modify the experiment to extend the investigation; at the same time, the equipment may be considerably simpler than in the unmanned case.

Even if it proves impossible to hard-mount this experiment in a manned vehicle, because of the ambient noise levels, much of the advantage of man can be obtained by allowing him to set up the apparatus and then deploy it into a free orbit through an airlock. Even better, it may be possible to provide a rudimentary "pure gravity orbit" capability in some of the manned vehicles which will be orbited during this decade. If the system is free-floating in an evacuated chamber of

reasonable size within the spacecraft, it is only necessary to drive the spacecraft so as to cancel very low-frequency or constant accelerations. For example, an oscillating acceleration of magnitude 10^{-5} g and period as long as 1 min results in a displacement amplitude of only 1 cm. The spacecraft thrusters will have to be fired only when the system approaches one wall of the chamber, either automatically or as a result of direct observation.

The fuel required to provide this capability is very modest. For a typical space station, weighing several hundred thousand pounds, using thrusters with a specific impulse of, say, 250 s, an average fuel flow below 1 lb/h is sufficient to compensate rms accelerations of order 10^{-7} g.* The capability is therefore not difficult to provide; in view of its possible utility in a number of experiments other than the present one, it is suggested that it be considered in the early design phase of any fairly small space station.

V. Conclusion

The apparatus described here is still in a relatively early stage of conceptual design, but it appears capable of improving the accuracy of the Eötvös experiment by several orders of magnitude, so as to allow investigation of the gravitational interactions of energy stored in the weak interaction. If this experiment and that of Nordtvedt (Ref. 4) are successful, the application of the Equivalence Principle to all known forms of energy will have been checked.

The principal advantages of the system over a conventional rotational balance are that it is relatively insensitive to gravity gradients and that it is possible to use resonance to enhance the Eötvös effect. Since techniques are available to isolate the device from disturbing accelerations, or to filter them out, its inherent sensitivity to these is not considered an insuperable obstacle.

Apart from the intrinsic importance of the Eötvös experiment itself, the technology that must be developed for the proposed apparatus is of a kind which may be applicable to other relativistic experiments (Ref. 16). General relativity no longer must rely solely on the "three famous tests" (which were all astronomical observations rather than controlled experiments). Recent advances in aerospace technology have made it possible to consider a broad variety of new experiments in this fundamental field of physics. However, most of those carried out to date – for example, laser ranging to the moon and radar ranging to planets or circumsolar spacecraft near superior conjunction – have used the solar system itself as a laboratory. There is another class, of which the present experiment is an example, that can take advantage of the local, controllable environment in a manned laboratory in earth orbit, and these have received relatively little attention as yet. Conventional laboratory facilities in free fall offer unique opportunities for significant experiments in gravitation, and it is to be hoped that growing awareness of the capabilities under development will lead to imaginative utilization of this potential by experimental relativists.

References

1. Dicke, R. H., The Theoretical Significance of Experimental Relativity, p. 77, Gordon and Breach, Inc., 1964.
2. Dicke, R. H., loc. cit., p. 9.
3. Schiff, L. I., Proc. Nat. Acad. Sci., 45, 69, 1959.
4. Nordtvedt, K., Jr., Icarus, 12, 91, 1970.
5. Lee, T. D., and Yang, C. N., Phys. Rev., 104, 254, 1956.
6. Feynman, R. P., and Gell-Mann, M., Phys. Rev., 109, 193, 1958.
7. Blin-Stoyle, R. J., Phys. Rev., 118, 1605, 1960.
8. Lobashov, V. M., et al., JETP, Letters, 5, 59, 1967. For a review of the theoretical and experimental situation, see Hamilton, W. D., in Brink, P. M., and Mulvey, J. H., Progress in Nuclear Physics, Vol. 10, p. 1, Pergamon Press, Inc., New York, 1969.
9. Mathisson, M., Acta Phys. Polonica, 6, 163, 1937.
10. Sciama, D. W., Nuovo Cim., 8, 417, 1958.
11. Schiff, L. I., ITP-291, Stanford University, 1968 (unpublished).
12. Jackson, J. D., Classical Electrodynamics, p. 581, John Wiley & Sons, Inc., New York, 1962.
13. Eötvös, R. V., Pekar, D., and Fekete, E., Ann. Physik, 68, 11, 1922.
14. Klemperer, W. B., and Baker, R. M., Jr., Astronaut. Acta, III, Fasc. 1, 16, 1957.
15. Graham, D., and McRuer, D., Analysis of Non-Linear Control Systems, p. 356, John Wiley & Sons, Inc., New York, 1961.
16. Weiss, R., and Forward, R. L., private communication.

*Reducing the effective steady acceleration to this level requires that the experiment be close to the center of mass of the vehicle, because of terrestrial gravity gradients.

ESRO Study Program for a Space Experiment on Gravitation Theories

G. M. Israël
European Space Research Organization (ESRO) Headquarters, Paris

I. Introduction

ESRO is currently considering a space experiment on gravitation theories. The study, which is in the definition phase, was initiated when ESRO received a proposal from Blamont in May 1969 for an improved method of conducting the time delay experiment already described by Shapiro, Ross, and Shiff. It was considered that a more complete utilization of the state of the art in space techniques, leading to highly accurate acceleration measurements in a heliocentric spacecraft, together with an improved laser signal propagation method (using a space-borne atomic clock), could substantially increase the validity of the gravitational time delay test during solar conjunction. Subsequently, ESRO decided to set up a group of scientific experts* to estimate the real gain one could expect from a complex space experiment entirely and exclusively devoted to tests of gravitation theories. Preliminary investigations of the primary required techniques were carried out in industry or in institutes under ESRO contract. These studies included an orbit analysis by the Space Division of MBB (Munich, Germany), investigation of the drag-free techniques by ONERA (Chatillon, France), and studies of the time measurement instrument by LSRH (Neuchâtel, Switzerland).

All these studies were completed by the end of February 1970 and used by the group to define the framework of a space experiment on gravitation theories. A preliminary feasibility study of the mission is currently being undertaken, with ERNO (Bremen, Germany) as the prime contractor. The completion of this work is expected by June 1971 and will be reviewed by ESRO before a decision is made to enter a new phase. Also, it is expected that in 1971 an important effort will be made in some advanced technique developments to ascertain the feasibility of critical subsystems.

The broad lines of the project, as viewed at the present time by the ESRO Mission Definition Group, are given below.

The objectives proposed for the mission were to use a spacecraft in heliocentric orbit to measure with the highest accuracy possible the three parameters γ, β, and J_2. Two orbits with an aphelion at approximately 1 AU were selected:

(1) An orbit with a period of 6 months and perihelion of 0.29 AU, which provides a very long superior conjunction after 6 months, and a second one after 18 months (Fig. 1).

(2) An orbit with a period of 8 months and perihelion of about 0.50 AU, the first superior conjunction occurring after 1 year (Fig. 2).

The performance analysis of available American launch vehicles showed that a Titan III D/Centaur can inject a total mass of 420 kg into

*Professor Blamont, University of Paris; Professor Colombo, University of Padua; Dr. Kundt, University of Hamburg; Mr. Lago, Centre Spatial de Brétigny; Mr. Marchal, ONERA, Paris; and Professor Roxburgh, Queen Mary College, London.

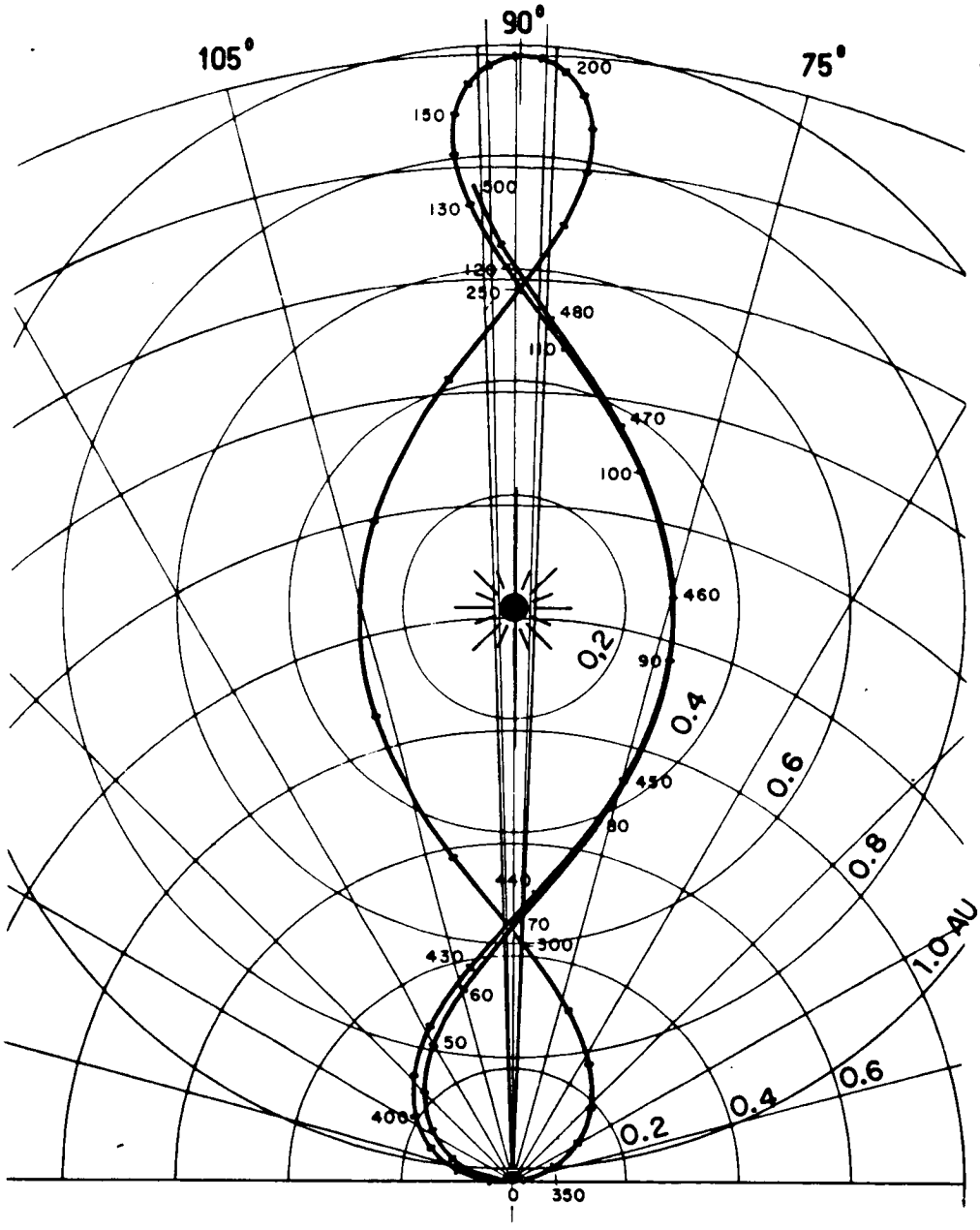


Fig. 1. 0.29-AU orbit in the co-rotating coordinate system. Launch in January, period 184 days (from MBB, Munich)

a 6-month orbit. The same mass can be put into an 8-month orbit by a smaller Atlas/Centaur vehicle. Also considered were orbits internal to earth orbit (achieved by means of velocity impulse at perihelion) and orbits out of the ecliptic plane. However, the spacecraft mass would be reduced to 220 kg, even with a Titan III D/Centaur + Burner II.

The preliminary studies have shown that the experimental requirements, and consequently uncertainties, are different for the determination of β and J_2 than for the measurement of γ.

II. The Orbital Motion Test

The measurement of β and J_2 reduces to the tracking of a heliocentric spacecraft moving over a pure gravitational orbit. The experimental problem is the drag-free system for which the required accuracy is 10^{-12} m/s^2. The physical limitations of the drag-free system have been worked out by ONERA, which conducted two studies:

(1) Definition of a drag-free system with a level of perturbations lower than 10^{-12} m/s^2 with consideration of two possible detectors:

(a) Electrostatic detector (capacity pickup, small cavity; Fig. 3), with the limitation that the electrostatic gradient of acceleration is too high because of the small gap (500 μm).

(b) Optical detector OPCS (large cavity; Fig. 4), whose limitations are

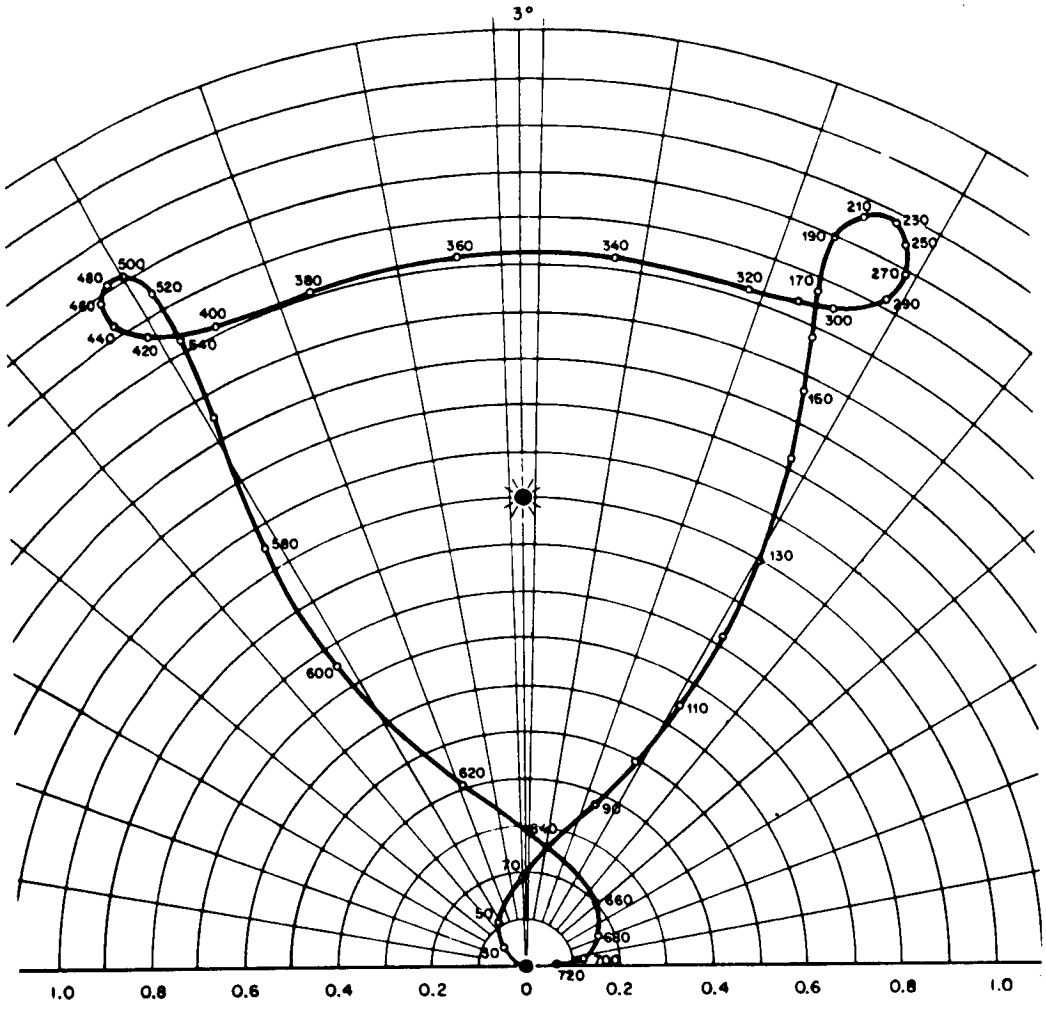


Fig. 2. 8-month orbit for launch in April, perihelion 0.52 AU (from MBB, Munich)

gradients of gravitational forces, uncertainty in the position of the proof mass, time of reacquisition.

(2) New concept of a medium-size cavity (a few millimeters gap), again with an electrostatic detector (1 V for the ball potential). An estimation of the perturbations on the proof mass due to the relative probe/proof mass movement was also made.

Control by cold-gas jets instead of ionic propulsers was also investigated by ONERA. The principal results of these studies, as presented by Juillerat in this proceedings, are the following:

(1) Perturbing accelerations on the proof mass due to internal gravity effects in the spacecraft, electric charge of the proof mass, and thermal effects (thermomolecular flow and radiation pressure) inside the cavity.

The required accuracy for the experiment is 10^{-12} m/s^2 in the ecliptic plane and 10^{-9} m/s^2 in the normal plane. The latter is achievable by proper distribution of masses. However, for the first constraint it is necessary to give the probe a constant rotation.

(2) Perturbations on the proof mass due to the relative motion. Average cancellation of the perturbing accelerations in the plane of rotation is obtained if two conditions are fulfilled: (a) the axis of rotation is perpendicular to the ecliptic plane, and (b) the acceleration on the proof mass is strictly constant in the frame of reference of the probe. These conditions are not completely fulfilled, and there is a "gradient" effect which is not neutralized by rotation if the mean value of ρ_0 is not very small (ρ_0 is the offset in the absolute axis of the position of the proof mass; see Juillerat's article, this proceedings). However, by an appropriate control with a measurement accuracy of 1 μm, the gradient effect can be diminished. This implies a computation of the integral term by the logic control of 1%.

Preliminary results from the present ERNO study confirmed that a rotating satellite compensates all radial gravitation forces from surrounding mass to the level required for the experiment.

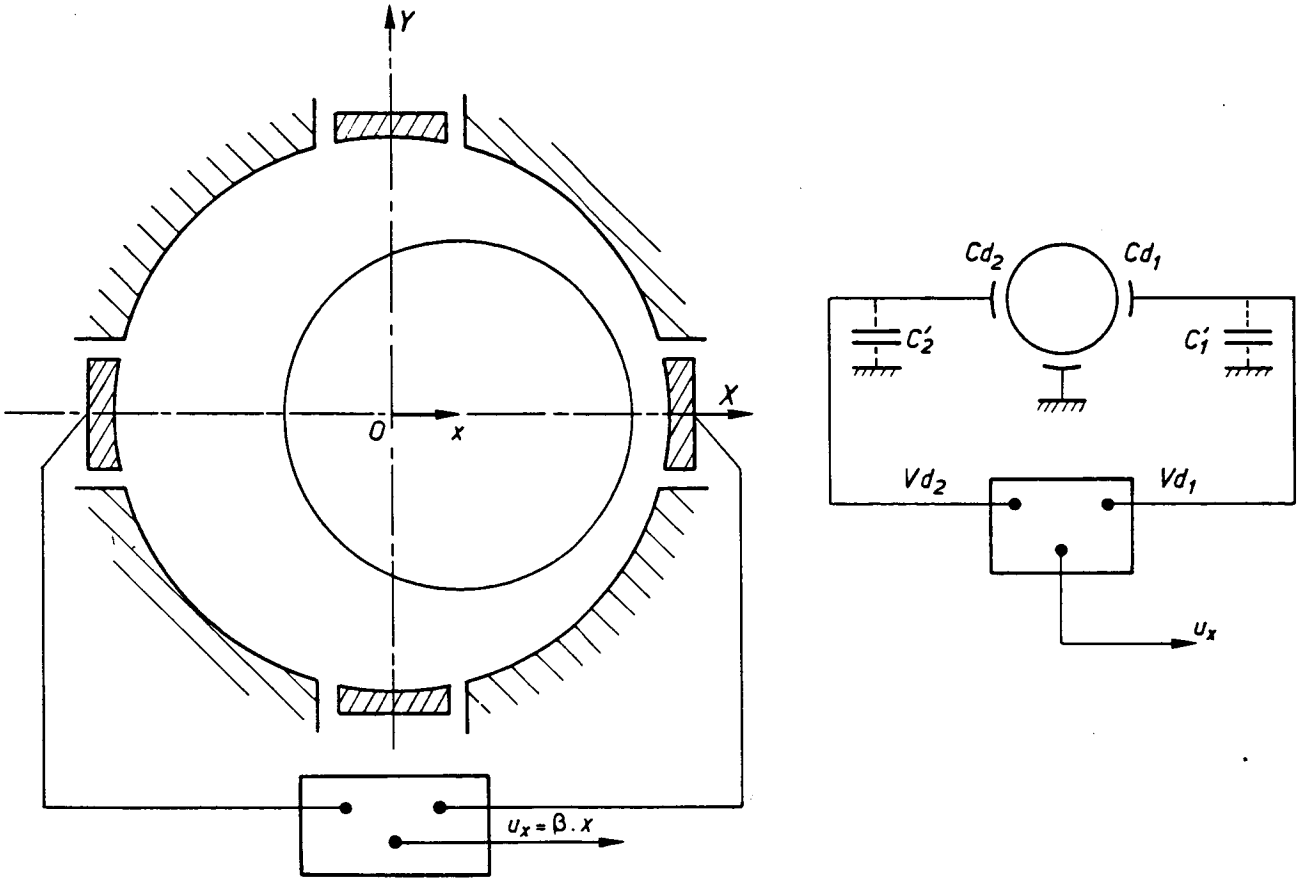


Fig. 3. Capacity position sensor for small cavity (a few millimeters gap)

The remaining effects from components in the axial direction and the perturbing effects of acceleration forces are investigated utilizing the theory of coaxial homogeneous mass rings, leading to preferable mass configuration criteria. Computations were made in the case of the large cavity, which seemed to be the more promising. Statistical investigations of the perturbations have shown that, through proper mass distribution around the large cavity model, it is possible to reduce the gradient effects to the desired level. It is not yet clear, however, whether the large cavity system will still require a computation of the integral term.

Another important conclusion of the preliminary ERNO study is that the control restitution system for the drag-free spacecraft operation can be achieved with cold gas with a small limit cycle of ±0.5 mm and a normal pulse duration of 7.10^{-3} s. This system certainly increases the reliability of the mission as compared to an ionic thruster restitution system.

III. The Time Delay Experiment

The determination of γ consists of the measurement of the transit length of an electromagnetic signal between the earth and the spacecraft. The experimental problems are:

(1) The process noise. A drag-free system is required but with less constraint (a 10^{-10} m/s^2 drag-free system operating for 2 months would provide enough accuracy for this measurement).

(2) The radio ranging (particularly the transponder's performance), which is required anyway with the best accuracy for the β and J_2 measurements, but which appears to be very much compromised by the decrease of the sun-earth-probe angle (antenna noise).

On the other hand, the physical limitations of the propagation method for the two-way experiment are mainly due to the signal scintillations through the solar corona. This cannot be improved much, even with the dual-frequency method (S- and X-bands).

For the one-way experiment, involving a light signal emitted by a laser from earth and recorded on the spacecraft, which implies the use of a clock on board and a clock on earth, the effects of the solar corona are negligible. However, the experimental problems are of another order of magnitude. They are:

(1) The requirement of an on-board laser detector-telescope, especially if one wishes to use a spinning spacecraft and if infrared techniques are employed.

(2) The utilization of a stable clock.

At the present time it seems that a CO_2-TEA laser transmitter (10.6 μm) can be used on the ground with 300 MW peak power, and a pulse duration of 1 ns and a repetition frequency of 1000 pulses/s (pulse-code modulation).

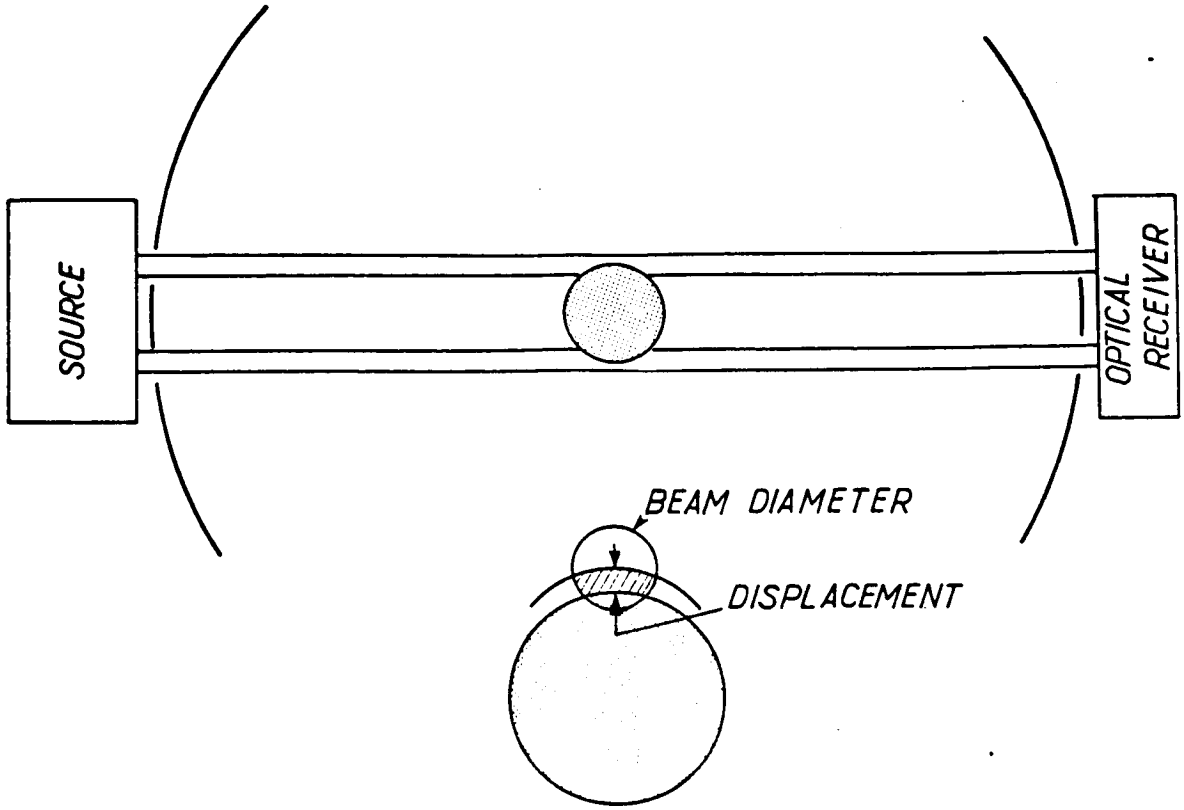


Fig. 4. Optical sensor parallel light beam system for large cavity (diameter = 1 m)

For the time measurement, LSRH strongly recommends the choice of an atomic cesium tube. This was shown to be sufficient for the long-term stability specified to be 5.10^{-12} during the 2-years' duration required for the mission. However, a uniformity of the time scale of ±10 ns for 10^5 s was considered as marginal though feasible. The required chronometry to 1 ns has already been developed for other applications and seems applicable to space missions.

These general considerations on the techniques involved in the mission which ESRO is currently defining will certainly indicate what the difficulties are in designing a spacecraft to test the gravitation theories. During the next phases of the project, many aspects will be reviewed. For instance, ERNO is considering different versions of the system:

(1) A spinning satellite of the Helios type (Fig. 5).

(2) A dual version, where the drag-free spacecraft is linked to a mother spacecraft (used for ranging, communications, and experiment, and forced to follow the daughter spacecraft by means of an appropriate propulsion unit).

Applying the same philosophy, ESRO is also studying an alternative to the drag-free spacecraft concept. In this new version, proposed by Colombo and Bertotti, two probes having identical physical properties, except for a difference of mass, are injected into a heliocentric orbit. The technical constraints related, for instance, to the thermal and optical properties of the two probes appear, however, to be at least as severe as those on the drag-free system.

To conclude, mention should be made of the analytical study at present being undertaken at ERNO and also at ESOC (ESRO's Space Operations Centre) to assess more precisely what are the possible accuracies achievable on the parameters γ, β, and J_2 as determined by the data on ranging and time measurements collected during the mission. Mr. Marchal (ONERA) and Dr. Roth (ESOC) present preliminary results of this aspect of the mission study in this proceedings.

- CYLINDRICAL CENTRAL BODY CARRYING THE CAVITY AND BOTH LASER TELESCOPES
- SIX VERTICAL EQUIPMENT PLATFORMS CARRYING TWO ATOMIC CLOCKS AND THE COMPONENTS OF THE TM/TC, POWER SUPPLY, ATTITUDE CONTROL, AND RESTITUTION SYSTEMS
- TWO CONICAL SOLAR ARRAYS
- SPACECRAFT LAUNCHER ADAPTER

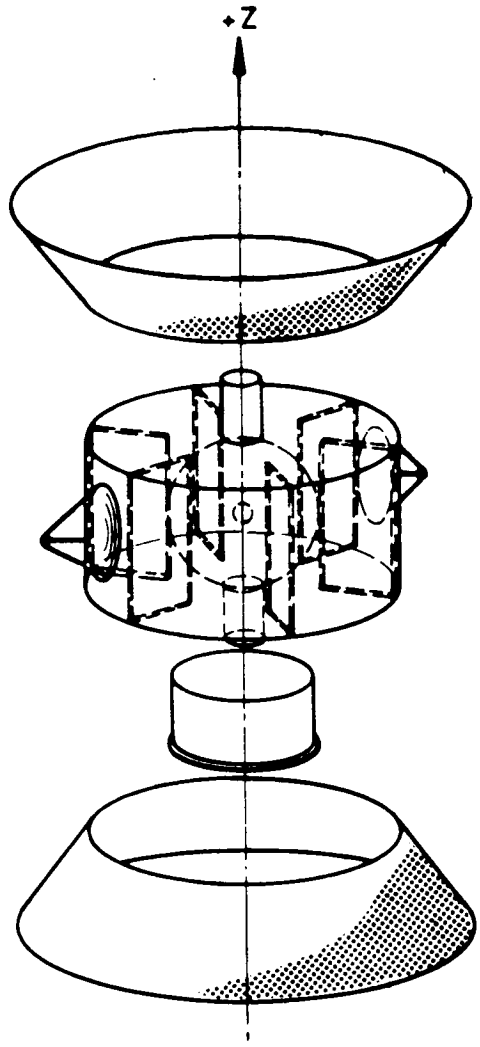


Fig. 5. Spacecraft principal characteristics (from ERNO, Bremen)

On the Accuracy of the Relativistic Parameters β, γ, and the Solar Oblateness Coefficient J_2, As Deduced From Ranging Data of a Drag-Free Space Probe

E. A. Roth
European Space Operations Centre, Darmstadt, Germany

I. Introduction

A space probe will be called drag-free in this study if all non-gravitational perturbations larger than 10^{-12} ms^{-2} are compensated. Such a probe allows for the determination of the coefficients of the second-order terms of the generalized Schwarzschild metric. For this purpose, it is necessary to take into account the short periodic perturbations produced by the deviation of the general gravity field from the Newtonian field. The important result is that the coefficient β can be obtained with a similar accuracy as γ. Up to now, β has been deduced with very poor accuracy from the secular advance of the Mercury perihelion, assuming the solar oblateness to be negligible.

The covariance analysis is based on simple models, including only a limited number of parameters to be estimated. The results are therefore optimistic and show what could be obtained if all the other parameters were known with sufficient accuracy; nevertheless, they give the relative merits of various heliocentric orbits.

II. Motion in the General Gravity Field

The line element of the generalized Schwarzschild metric has, after appropriate expansion, the form

$$ds^2 = \left(1 - 2\alpha\frac{m}{r} + 2\beta\frac{m^2}{r^2} + \cdots\right)dt^2 - \frac{1}{c^2}\left(1 + 2\gamma\frac{m}{r} + \cdots\right)d\sigma^2 \qquad (1)$$

with $m = fM_\odot/c^2$.

In the case of Einstein's general gravity theory, we have $\alpha = \beta = \gamma = 1$, whereas for the Brans-Dicke theory, $\alpha = \beta = 1$, $\gamma = (1 + \omega)/(2 + \omega)$.

It is assumed that a particle moves in this field along a geodetic. Considering the deviation of the line element (Eq. 1) from the Newtonian line element as a perturbation, the acceleration of a particle can be written

$$\ddot{\underline{r}} = \ddot{\underline{r}}_{\text{Newton}} + \ddot{\underline{r}}_{\text{rel}} \qquad (2)$$

The perturbing acceleration $\ddot{\underline{r}}_{\text{rel}}$ has been derived by Anderson (Ref. 1) for the metric (Eq. 1). This perturbation is easily separated into a radial and a transverse component:

$$S = \frac{m}{r^2}\mu\left(2\beta\frac{1}{r} + \gamma\frac{1}{a} + 2(1+\gamma)\frac{1}{p}e^2\sin^2\theta\right)$$

$$T = \frac{m}{r^3}\mu \cdot 2(1+\gamma)\, e \sin\theta \qquad (3)$$

where $\mu = fM_\odot$ and $\alpha = 1$ has been assumed.

Using the planetary equations of Lagrange for the variation of the orbital elements, it is easy to deduce the short-periodic perturbations of a, e, ω, and t. (Full details are given in Ref. 2.)

III. Covariance Analysis

The covariance matrix is given by

$$A = X^T N^{-1} X \qquad (4)$$

where N^{-1} is the covariance matrix of the measurements and X the matrix of the partial derivatives of the earth-space probe distance with respect to the parameters (Ref. 3). The expected values of the errors α_i of the parameters p_i can be derived from A:

$$E(\alpha_i, \alpha_k) = \left\{A^{-1}\right\}_{ik} \qquad (5)$$

In a first model, the primitive model, only the relativistic parameters β and γ, and in addition the solar oblateness coefficient J_2 have been considered. Using the results of the first-order perturbation theory for β, γ, and J_2, the elements of the matrix X follow easily. In this case, the partial derivatives become independent of the parameters. This model shows what could be obtained in the most favorable case, in which all other parameters are already known with sufficient accuracy.

In a more sophisticated model, the simplified model, the orbital elements of the space probe and the earth are included as quantities to be improved. It is assumed that the probe moves in the ecliptic plane, so that the motions are restricted to two dimensions. This model provides a first indication of how the inclusion of more parameters reduces the accuracy with which the parameters are obtained.

IV. Numerical Results

As a first instance, two drag-free space probes with an orbital period of 6 and 8 months, respectively, have been considered. The main reason for this choice is that these probes allow the addition of a sun-occultation experiment from which an independent measurement of γ is obtained from the time delay of an electromagnetic signal. The orbital elements of the two probes are given below:

Element	Probe I	Probe II
a, AU	0.64054	0.77530
e	0.5873	0.3114
ω, deg	0	0
θ_0, deg	180	180
P, months	6	8

For the covariance analysis it was assumed that the range experiment starts 0.1 year after launch and that every day one uncorrelated measurement can be obtained. For the standard deviation of the range measurement, the value α_r = 15m (= 10^{-9} AU) has been adopted.

Figures 1 and 2 show the standard deviations calculated for the drag-free probes I and II in the

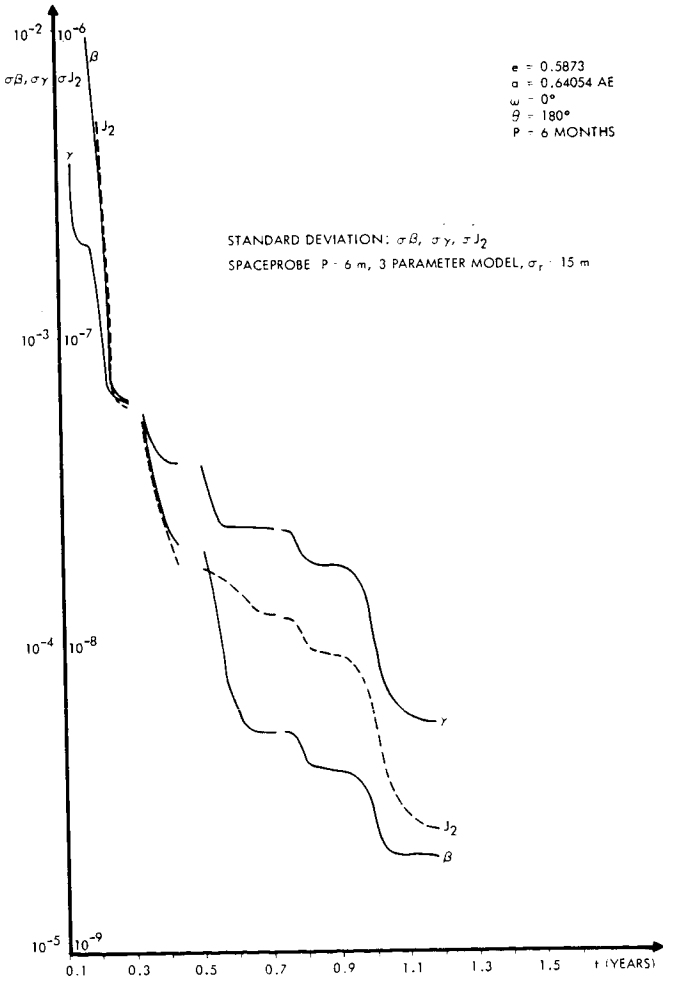


Fig. 1. Standard deviations of parameters β, γ, J_2 vs tracking time for drag-free Probe I; three-parameter model

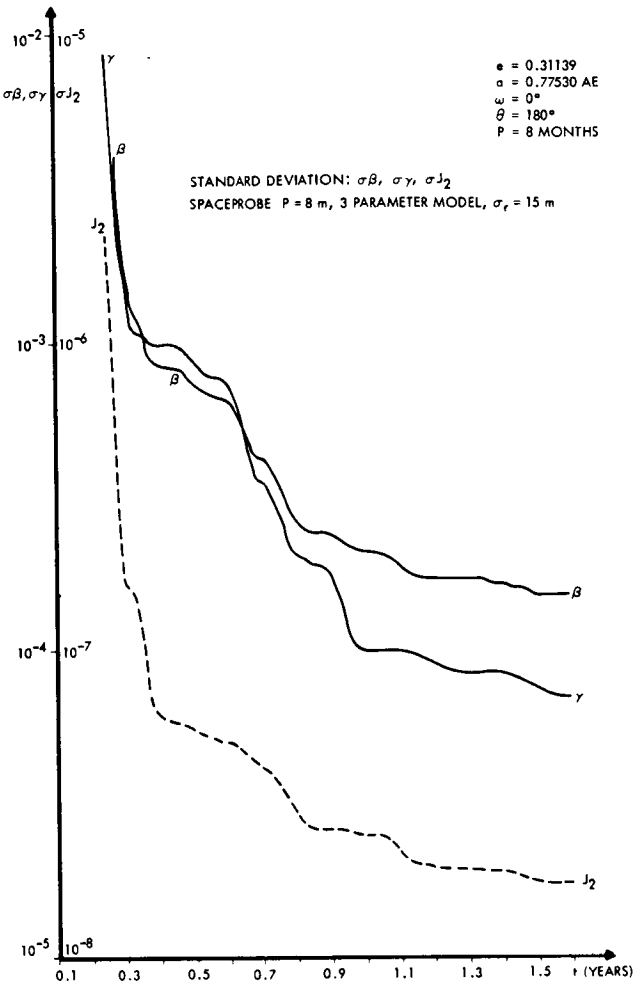


Fig. 2. Standard deviations of parameters β, γ, J_2 vs tracking time for drag-free Probe II; three-parameter model

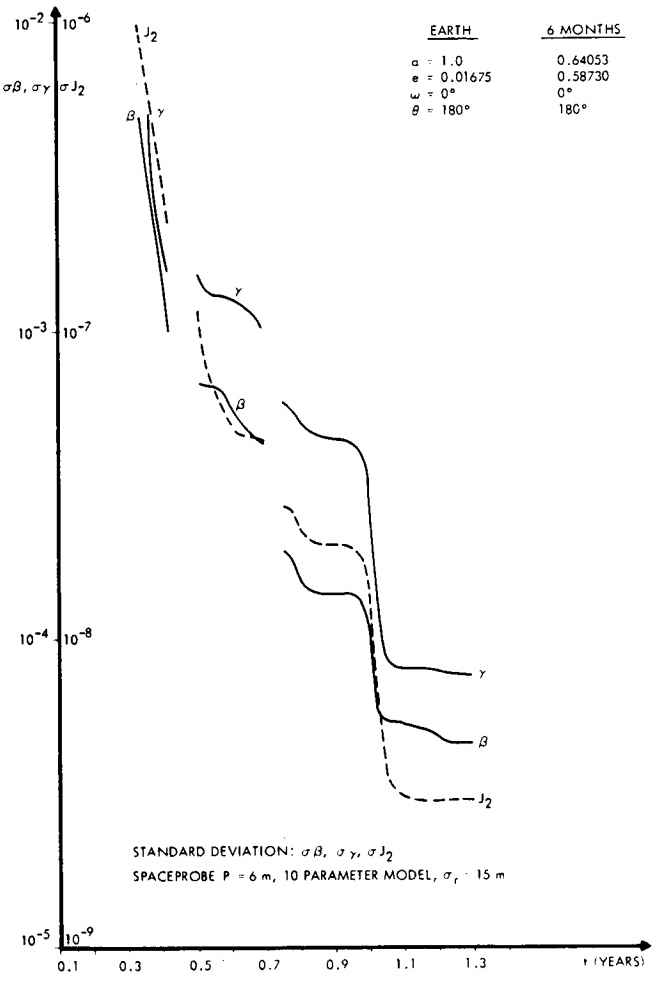


Fig. 3. Standard deviations of parameters β, γ, J_2 vs tracking time for drag-free Probe I; ten-parameter model

case of the primitive model. In order to obtain a reasonable result, tracking has to be performed during at least one revolution, and the second revolution improves the result by a factor of 10. It is interesting to note that probe I gives the coefficient β with a slightly higher accuracy than γ. From Fig. 3 it follows that the inclusion of more parameters reduces the accuracy by a factor of 2 to 3. Parameters β and γ are still obtained to an accuracy of 10^{-4} and J_2 slightly less, but the correlation of γ with β and J_2 is rather high (-0.86, 0.78, respectively, after 1.2 y). The correlation between β and J_2 is -0.46. It seems, therefore, that two space probes with very different orbits are necessary in order to determine the individual parameters.

Note: A further calculation has been performed, assuming that a transponder has been placed on Mars or Icarus. (In this case, the "drag-free" problem is nonexistent.) The result is that a transponder on Mars produces about the same accuracy as probe II, whereas a transponder on Icarus is equivalent to probe I. For a Mars transponder, the correlation of γ with β and J_2 is considerably reduced (-0.68 and -0.42), but the correlation between β and J_2 is large (0.95).

References

1. Anderson, J. D., Space Programs Summary 37-50, Vol. III, 39-47, Jet Propulsion Laboratory, Pasadena, Calif., Apr. 30, 1968.

2. Roth, E. A., "On the Motion in the Field of General Relativity," presented at the XXIst IAF Congress, 1970 (submitted for publication).

3. Hechler, F., Roth, E. A., and Bertotti, B., "Assessment of the Accuracy of the Relativistic Parameters β, γ and the Solar Oblateness J_2, Determined by Range Measurements of an Interplanetary Probe," ESOC Internal Note 73, 1970.

Anchoring Spacecraft to Planets

D. L. Cain
Jet Propulsion Laboratory
California Institute of Technology

I. Introduction

Emphasis is currently on the use of time delay and bending tests of radio and light rays traveling near the sun to verify theories of general relativity. Different relativistic theories differ in their predictions of the motions of the planets (or spacecraft) in the solar system (e.g., perihelion advance), and these differences can be used to select "best" theories.

Proposals have been made to study the motion of spacecraft in orbit around the sun using an accurate accelerometer to calibrate or remove the non-gravitational accelerations.* The limit of experimental accuracy is the accelerometer performance. Without analyzing the latter, it can be said simply that for planet tracking, the immense mass relative to applied non-gravitational forces (for even the smaller asteroids) removes this limitation on our ability to model the motion mathematically. A lander on a natural satellite (other than our moon) would not suffer from non-gravitational effects, and would further have a large mean motion -- important for observing gravitational effects. Of course, the proximity of the moon to the earth contributes to its usefulness in relativity testing as reported elsewhere.**

Before discussing the relative merits of landers and orbiters, we must select candidate planets. Although practical considerations not associated with relativity usually dominate the decision procedure, some comments about the inner planets are summarized in Table 1.

Radio tracking, either radar bounce or active transponder tracking, can be used to infer positional information about planets. The bounce data are less precise because the echo emanates from many points on the surface of the planet; therefore, the topography must be taken into account, which somewhat weakens the solutions. However, the long time span over which this type of data has been and will be taken recommends its use in combination with the more accurate data from transponders when such data become available.

When a transponder on an orbiter is used for the study of planet motion, the most striking effect is the increased precision of the range (and range rate) data as contrasted to radar bounce data. We see the topography and surface reflectivity removed as an error source, but we must contend with the gravity field undulations. Since a planet orbiter experiment has yet to be accomplished, we can only guess at the magnitude of this problem on the basis of experience with the earth and the moon. Based upon exploratory simulations, using reasonable similarity laws to estimate gravity-field uncertainties, and plausible assumptions about the spacecraft non-gravitational forces, we estimate that the center of mass, say of Mars, can be determined by the orbit reduction process to about 50 m, using the ranging data of today's technology. Thus we can see that long-term

*D. DeBra, "Stanford State-of-the Art on Drag-Free Systems" (Conference paper).

**P. L. Bender "Laser Ranging to the Moon" (Conference paper).

Table 1. Considerations in the selection of candidate planets

Planet	Advantage	Disadvantage
Mercury	High eccentricity Large propagation delay Large mean motion	Tracking by day (earth ionosphere) Near solar corona For lander: high energy requirements and high day-time surface temperature
Venus	Large mean motion Good propagation delay	For lander: high surface temperatures, winds For orbiter: potential of the atmosphere difficult to model
Mars	Night-time tracking Less corona	Lower mean motion
Jupiter	Observable effects of relativity on orbiter motion – high Q	Distance Energy requirements Landing difficulties

tracking of an orbiter, yielding a sequence of (in essence) geocentric range data points, can be used in relativity theory tests.

But a lander would yield even better data, because whatever difficulties are involved in the uncertainties of the theory of rotation of a planet, the orbiter, too, would be subject to these, in addition to the velocity, gravity field, and non-gravitational force uncertainties. Table 2 presents a guess at the ranking of the "equivalent range data" capability of three methods.

The theory of the motion of a lander about the center of gravity of a planet such as Mars is subject to much less speculation than that of an orbiter, because rigid-body motion may be safely assumed. For this reason a detailed analysis was made of such a lander.

II. Analysis of a Mars Lander

Preliminary analysis showed that even after tracking periods of several months, the uncertainties in Mars' and earth's orbits predominated and obscured relativity results. Therefore, the analysis incorporated data for a whole synodic period (i.e., a return of Mars to its original position on the celestial sphere as seen from earth). A variance-covariance analysis* was made using ranging data between an earth tracking station and the lander, assumed to be located on Mars' equator. Range standard deviation was assumed to be 3 m with no bias. A tracking pass was taken every 40 days for the 780-day total period, and a range point was taken over 0.1 day during the pass day while the earth was visible to the lander

Table 2. Estimated equivalent range data capabilities

Method	Representative geocentric distances, m
Planet bounce	±500
Orbiters	±50
Landers	±5

(6 data per pass). Data were not taken while the sun-earth-Mars angle was less than 15 deg to avoid the necessity of considering the solar corona.

The solution parameter set consisted of six orbital elements for Mars, three earth orbital elements, the sun's mass, and seven elements of Mars rotation. Keplerian motion (conics) was used to approximate the motion of both Mars and earth. Brouwer and Clemence (Ref. 1) "Set III" elements were used for the partial derivatives.

Three of the earth's six elements (constants of integration of the equations of translation) were eliminated to account for the three degrees of unobservability (e.g., three Eulerian rotations) when only distance measurements were used. The rotational elements of Mars were three lander coordinates, rotation rate, two spin axis orientations, and precession (motion of spin axis).

*The computer program used for this study was a conglomeration of available programs: A. S. Liu furnished the conic routines and did the main program structuring, M. J. Sykes supplied the β and γ partial routines, and the author provided the Mars rotation and associated partials.

A. Solution Characteristics

The solutions were viewed in two separate ways: (1) assume that the relativistic form of the equations of translation is known (Einstein Theory) and (2) assume that the two parameters β and γ (Refs. 2 and 3) in the relativistic "perturbations" of the Newtonian motion are not known but must be estimated (both separately and together).

B. Relativity

When β is estimated, and to a lesser extent, γ, the determination of the in-plane elements of Mars and earth suffers, as shown in Table 3.

In spite of the high correlations* between the in-plane parameters and β, γ, the result produced a 16% test on β (or 7% assuming γ was known by then), a very encouraging test of general relativity.

C. Orbits

When $\beta = \gamma = 1$ is assumed, the in-plane parameters of Mars (relative to the earth reference) are well determined. The periapsis of Mars (Δr in the Set III elements) is determined to 0.8×10^{-3}", showing the degree to which an inertial framework is defined. The other two axes (i.e., small rotations $\Delta\rho$ and Δq) are less well determined by a factor of 10, but this is still impressive relative to today's determinations.

D. Rotational Elements

The parameter of greatest interest here is the precession. In this case, the motion reflects the action of the sun on Mars' equatorial bulge, resisted by the planet's angular momentum. With the mass already known to five figures, the precession yields the polar moment of inertia -- a stringent test of any theory of the internal distribution of mass.

E. Lander Location

The study again confirmed prior simulations regarding the difficulty in determining equatorial distance. The inclusion of β, γ does worsen this determination, changing the 170 m in Table 4 to 1 km. The other two coordinates are well determined.

III. Summary

Much useful information can be obtained from the tracking of a Mars lander over a long period. If the lifetime is sufficiently long, conclusions of fundamental significance can be drawn in the areas of

(1) Relativity – determination of β.

(2) Precise inertial framework determination.

(3) Internal structure – polar moment.

(4) Planetary ephemerides, Mars physical ephemeris.

Table 3. Determination of in-plane elements of Mars and earth*

Estimate	Orbit solution errors (1 - σ)				Relativity parameter determination	
	Mars $\times 10^9$		Earth $\times 10^9$		β	γ
	$\delta a/a$	δe	$\delta a/a$	δe		
β and γ	1.63	2.0	3.9	3.2	0.16	0.03
β only	0.48	0.74	1.4	1.3	0.07	--
γ only	0.21	0.1	0.2	0.05	--	0.01
Neither	0.05	0.03	0.03	0.05	--	--

*No a priori information assumed on any parameters – 3-m ranging standard deviation.

*The out-of-plane and Mars orientation were almost uncoupled with β, γ, and in-plane parameters.

Table 4. Solution standard deviations for 3-m ranging; $\beta = \gamma = 1$ (not in solution)

Parameter	Mars $\times 10^9$	Earth $\times 10^9$
$\Delta \ell_o + \Delta r$	0.18	--
Δp	1.5	--
Δq	1.5	--
$e\Delta r$	0.42	0.028
$\Delta a/a$	0.05	0.030
$\Delta \ell$	0.03	0.049
	AU	20 m
Lander: r		0.52 m
Lander: λ*		0.61 m
Lander: z		170 m
Pole: ω^r/day		0.13×10^{-8}
Pole: i_o*		1.0 m
Pole: Ω_o*		1.3 m
Precession = $d\Omega_o/dt = 0.61 \times 10^{-4}$ r/century		

*Multiplied by Mars radius.

References

1. Brouwer, D. and Clemence, G., Methods of Celestial Mechanics, Academic Press, New York, 1961.
2. Georgevic, R. M., Space Programs Summary 37-56, Vol. III, 8-15, Jet Propulsion Laboratory, Pasadena, Calif.
3. Anderson, J. D., Space Programs Summary 37-50, Vol. III, 39-47, Jet Propulsion Laboratory, Pasadena, Calif.

The Use of the Earth-Moon System as a Radio Tracking Facility

J. Derral Mulholland
Jet Propulsion Laboratory
California Institute of Technology

I. Introduction

Earlier in this meeting, Mahlon Easterling remarked that, for distant missions, tracking stations may have to be put somewhere other than the surface of the earth. He was referring to the fact that the two-way time delay to an object 30 AU distant is about 8 h. Actually, we already have nonterrestrial tracking stations in a sense, in the Orbiting Astronomical Observatories and various orbiting data links such as the communications satellites.

There are several potential applications that one could envision for a tracking station established on a satellite with high-velocity parallax and which would be large enough to accommodate a large variety of tasks. Leaving economic and political considerations aside for the moment, it seems a logical extension of the orbiting data link concept to recognize that our natural satellite, the moon, satisfies this description. Perhaps we should consider, then, the establishment of a radio-tracking facility on the lunar surface. Indeed, Jurkevich (Ref. 1) has suggested such a station for the surface of Mars, but that is an impracticability of an entirely different order of magnitude than that of the present discussion. A much better case can be made for the earth-moon system.

II. Discussion

The suitability of the moon for such a purpose suggests itself because one of the requirements of a radio tracking network is maximum accuracy in the knowledge of relative locations of the stations in the net. In conformance with the discussion by my colleague in the Lunar Ranging Experiment, Peter Bender, there seems good reason to believe that, within a few years, the motion of the moon about the earth and the motion of the moon about its own center of mass will each be determined to something approximating the absolute accuracy with which terrestrial distances are known, perhaps the order of 1 m. Instantaneous topocentric distances to lunar surface sites will be capable of being monitored at the centimeter level, using laser systems and relatively low-cost optical receivers now under development. When this state of affairs is realized, there will be no more problem with the definition of a Goldstone - Sinus Medii baseline than with a Goldstone - Woomera baseline.

What are the reasons for suggesting such a facility? There must be some important advantages if the suggestion is even to be considered, some tasks or researches that either cannot be done as well or not at all from the surface of the earth, such as the following:

(1) Extension of baseline by two orders of magnitude.

(2) Absence of lunar atmosphere.

(3) Higher-velocity parallax with longer period.

(4) The moon's more "modellable" rotation.

The most immediately obvious difference introduced by a tracking station on the lunar surface is the greater distance between it and its sister stations, compared to the present situation. The

baseline between two stations of an earth-bound very long baseline interferometer (VLBI) system is customarily on the order of 2000-9000 km; the ultra-long baseline interferometer (ULBI) will have its endpoints separated by about 360,000 - 410,000 km. Ideally, this could yield directional measures of radio sources to precisions on the order of 10^{-5} arc sec at 1 cm wavelength. Such precision would greatly enhance the execution of radio signal deflection studies during superior conjunctions. Time delay interferometry of this precision could be used in conjunction with more conventional spacecraft tracking techniques to determine the instantaneous three-dimensional geocentric position of the object to the order of 10 m, thus enhancing both the orbit determination process and the fertility of relativistic tests with spacecraft. The application of ULBI to the study of stellar sources would permit the resolution of features as small as 1 light year in extent at cosmological distances, or smaller than the orbit of Jupiter at the galactic center.

One of the current limitation factors in radio tracking, whether in simple tracking or interferometry, is the effects of the earth's atmosphere. This limiting action takes on two forms: (1) Uncertainities in ray-path refraction degrade the accuracy of the data, and (2) atmospheric absorption of large regions of the electromagnetic spectrum imposes a severe limit on the observable frequencies. The moon, of course, has essentially no atmosphere. A tracking station that simply makes observations for subsequent relay to the earth is hampered by neither problem and can thus make higher quality observations, both in the sense of avoidance of the refractive uncertainties and in the sense of being capable of continuous surveys through the frequency spectrum. Such a facility could observe the x-ray emissions expected from black holes. The continuous frequency spectrum capability could be used to investigate such potentially fundamental questions as quasar luminosity distribution over frequency and the wideband characteristics of pulsars.

In conjunction with earth-bound stations, the lunar base possesses other advantages in the matter of atmospheric degradation. For example, one might expect that a ULBI installation would experience refractive degradation that was lower by a factor of $\sqrt{2}$ compared with completely earth-bound interferometers. Furthermore, the influence should be much less in terms of degrading the resolution, because of the increased baseline. The lunar base would provide the possibility of regular multi-frequency monitoring of the earth atmosphere and ionosphere by means of range and range-rate measures between the earth and the moon, much as is now done with artificial satellites. This could provide a calibration of the tropospheric and ionospheric effects and reduce the influence of the atmosphere as an error source for the terrestrial stations.

The cyclic doppler shift imposed on a signal by the revolution of a tracking station about the geocenter can be used in a sense as a single-station non-simultaneous interferometer. The higher-velocity parallax of the lunar station (1000 m/s) compared to an earth-bound station (400 m/s) and the longer period of revolution can be significant advantages in spacecraft tracking.

Finally, a real and present barrier to increased precision from earth-based stations is that posed by unpredictable and unmodellable motions of the earth due to its non-rigidity. The polar motion, variable rotation rate and possible continental drift do corrupt tracking data in a bothersome and interesting way. In principle, and eventually in fact, the motions of a lunar surface point relative to the selenocenter can be removed as an error source in tracking data from a lunar station, because the effects of non-rigidity are much smaller.

Having specified the potential advantages of a lunar tracking facility, it is necessary to observe that these can be utilized in a variety of ways, ranging from the very simple and relatively economical to the very complex and costly. In fact, we have already had the use of some of the simpler ones, which I will designate as "reflective data links" (Fig. 1). These are devices that receive a signal from the earth and retransmit it in some manner back to the earth. The simplest possible device of this kind is the totally passive data link, typified by the laser ranging retroreflector. The entire facility consists of a device that is carried to the lunar surface and implanted, where it just sits being bombarded by photons. It costs a small fraction of the total mission and has no operating parts. It can provide very high-precision data, but relating only to the earth-moon system. One such device is already on the lunar surface, with two more scheduled. The information provided by these devices is expected to render the more sophisticated systems feasible, by tying down station locations relative to the geocenter.

Another simple reflective facility is the active transponder, which receives signals from an earth station, transforms them in some specified fashion, and retransmits the transformer signal back to the earth; the Surveyors carried such devices. They are more complex than the passive reflectors, requiring power supplies. They involve operating equipment, which is to say failure-prone components. With such a station of suitable endurance, one could investigate the earth-moon system and, more importantly, the atmospheric effects on various frequencies of transmission.

Next in complexity in the hierarchy of possible facilities is the three-way transponder tracking

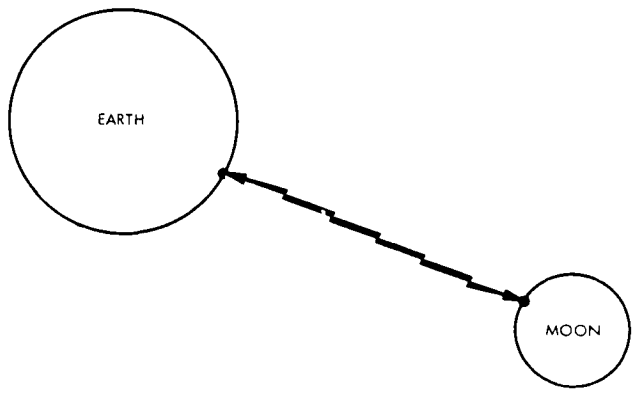


Fig. 1. Reflective data link

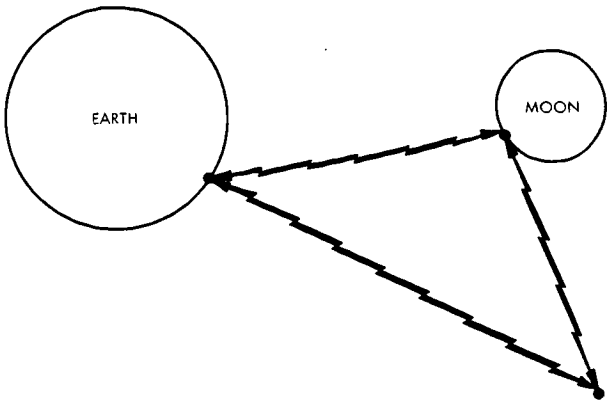


Fig. 2. Transponder tracking station

station, with the capability of tracking something other than the moon itself (Fig. 2). This would be primarily an interferometric station, slaved to an earth-bound station in a way that depends on the objects being observed. For spacecraft tracking, one could maintain a phase-coherent system by transmitting and receiving the basic signal at an earth station that shared a master clock with the station receiving the signal retransmitted from the lunar station. For interferometric tracking of natural objects, it might be necessary to have a reliable time and frequency standard at the lunar site. Either arrangement requires a sizable facility, at the very minimum something like a giant Surveyor, possibly even an astronaut-constructed unit, but it would provide the ULBI capability of extending the angular measures 2 orders of magnitude, of potential interest in tests of relativistic models and the examination of the structure of distant radio sources.

Finally, one may mention the monostatic tracking station and data link (Fig. 3), which provides the additional capability of wideband frequency surveys of radio sources. Again, this would involve a major facility, providing essentially complete radio tracking capability, which might be accomplished more effectively and permanently if constructed by an astronaut team, although there seems no need that it be manned in operation.

In addition to the tracking functions that would be the primary raison d'être for the lunar station, other interesting studies could be pursued at the same site, taking advantage of the data link capability. For example, one could perform the equivalent of a clock-in-a-satellite experiment, if there were a station time standard. Another possibility is presented by the suggestion of de Sabbata (Ref. 2) that the mass concentrations of Muller and Sjogren might be suitable resonating devices for the detection of gravitational waves from pulsars, in the 1-Hz regime. If this is indeed a reasonable suggestion, then detecting equipment could be situated at the tracking station (located on a suitable mare) and could be monitored from the earth-bound DSN. Going further,

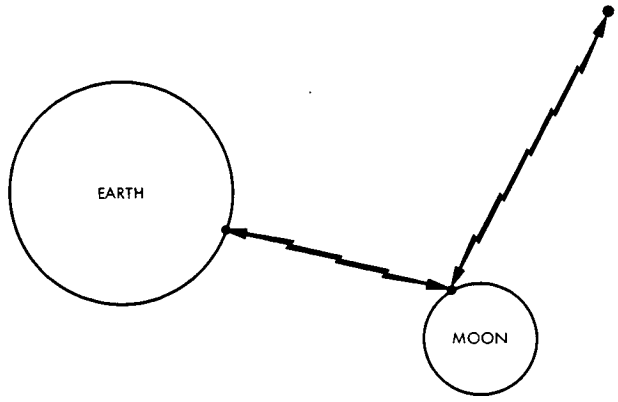


Fig. 3. Monostatic tracking station and data link

Marchal (Ref. 3) proposed to determine the solar oblateness and the coefficients of the metric tensor by means of observations of a laser signal fired from a space probe during the interval in the vicinity of superior conjunction with the sun. Familiarity with the operations of the LURE facility at the McDonald Observatory is sufficient to make me suspect that one does not want to try to observe such a transient and costly phenomenon from beneath the earth's security blanket of air; again, the lunar tracking site provides an "all-weather" location for the detector, with the means for data transmission back to the earth.

III. Summary

The use of the earth-moon system for radio tracking has several potential points of interest from a scientific standpoint, and the facility would be a convenient base station for certain other interesting researches not directly involving the tracking capability but utilizing the facilities as an orbiting data link. The idea seems to have much to recommend it and covers a wide spectrum of possible modes of realization, presenting an equally wide range of possible applications. It is my purpose here only to stimulate your imaginations with these visions, and I hope that I have been successful.

Acknowledgment

I have profited in this discussion from the advice and helpful comments of Mahlon Esterling, Peter F. MacDoran, P. S. Callahan, and T. W. Hamilton.

References

1. Jurkevich, I., Astronautics and Aeronautics, 8, 6, 72-73, 1970.

2. deSabbata, V., Memorie Societa Astronomica Italiana, 41, 1, 65-68, 1970.

3. Marchal, C., La Recherche Aerospatiale, No. 2 (Mar.-Apr.), 57-62, 1970.

Summation and Critique

R. W. Davies
Jet Propulsion Laboratory
California Institute of Technology

Unfortunately, the wide differences of opinion that all of us are aware of emerged more frequently during the cocktail and dinner hours than on the conference floor itself. Consequently, the confrontations of the issues which will have to be faced some day were not dealt with as clearly as some of us had hoped. The individual papers do not faithfully reflect the argumentative excitement of the subject nor the ambience of the three days.

My own attitude on experimental gravitation is very heavily conditioned by the historical record. Not until the late eighteenth century was the Newtonian Theory of Gravity generally accepted because Newton's great continental contemporaries were wedded to Cartesian philosophy. Despite the fact that it was the best theory at the time, the astronomical observations appeared to fit with it to only the first order. Later, Lagrange and Euler did much to develop the mathematical analysis of Newtonian Gravity (leading up to a field theory), but even they were still concerned because of the apparent secular accelerations of Jupiter and the moon. Finally, in 1787, Laplace put the problem in proper perspective by demonstrating that the accelerations were actually of very long period, and Newtonian Gravity was generally accepted.

A half century later, the Frenchman LeVerrier undertook the great task of fitting the observations of the motions of the planets to the developed theory. He found a discrepancy in the precession of the perihelion of Mercury of 35 arc-s per century. The seriousness of this discrepancy became progressively more evident when the search for other planets did not turn up anything.

Dicke once speculated on what the history of the theory of gravity might have been had Lorentz come forward. We can also speculate on what might have happened in the twentieth century if Simon Newcomb had not, a half century after LeVerrier, refined the work of LeVerrier and found he had to correct the ephemeris of Mercury by 43 arc-s per century.

In my opinion, the way is clear. In the last several years, we have seen the development of new operational methods for observing the motions of the planets and the moon. We should utilize them as extensively as possible. Furthermore, we should employ the best theory of gravitation available, the General Theory of Relativity, and make every effort to demonstrate its consistency with the observations.

When the equations

$$G_{ij} = 0$$

seem to be in difficulty, we should follow up LeVerrier style and invoke

$$G_{ij} = T_{ij}$$

before we abandon the structure of G_{ij}. Perhaps I am taking some of the glamour out of the subject by being reluctant to take alternative theories very seriously at this time. Nature has its own standards of drama, and it is not yet time to guess at the next act.

Of course, it is essential that we push for practical new tests of the General Theory of

Relativity. It is even more important, however, to remember that we owe something to the future besides a collection of contesting theories. Scientists have often lamented the destruction of scientific records or the omission of observations. We wish that more would have been observed and recorded about the three supernoval events. We certainly wish that only a century ago observers had had more precise clocks and telescopes. I am hoping that with these new observational techniques at our disposal, we can continue in the tradition of leaving a legacy of fine observations for another generation.

The value of the PPN type of formalism is that it helps us understand to what level of accuracy we have tested General Relativity and to what level we have to strive to detect some discrepancies from the theory's predictions. I believe we are better off not to take the alternative theoretical descriptions too seriously and to search for other physical causes for these discrepancies. Experimenters should be careful not to bias the interpretations of their observations by stating them in PPN parameters only or in any theoretical framework; rather, the results should be presented phenomenologically, giving the theorist not only more freedom to interpret them but also an opportunity to evaluate inconsistencies. For example, I like the fact that Muhleman accepted General Relativity and calculated the electron density of the solar corona. The result is consistent with those obtained from eclipse observations. If the result had been inconsistent, I would have given up on the eclipse observations or on Muhleman's methods before giving up on General Relativity.

Dicke stated that the singularly most important problem was measurement of the gravitational quadrapole moment of the sun. I believe that most of the people at the conference agree with him. Unless Hill and his collaborators find an experimental inconsistency in the Princeton solar oblateness experiment, we will continue to have an interesting problem one way or another. Either General Relativity gives the right answer for the wrong reasons, or there is an astrophysical problem in need of explanation.

What then is the most practical method of measuring the solar quadrapole? Shapiro states unequivocally that it can be accomplished with the proposed upgraded Arecibo radar antenna. JPL personnel feel that it might also be possible to use the Deep Space Network for that purpose. These possibilities should be carefully considered. We can visualize that such experiments might take longer or be less accurate than their proponents now believe. On the other hand, if they are feasible, we must re-evaluate the role of a space mission such as the one being studied by ESRO. The ground-based radar experiments might prove to be inconsistent, or they might be consistent but not in agreement with General Relativity, whereupon a space mission acquires more importance.

Another possibility is the one mentioned by Friedman, where an extension of the Mariner Venus-Mercury mission might provide a measure of the solar quadrapole to a part in 10^5. This level of accuracy, according to Dicke, would settle the gravitational question but not the astrophysical solar oblateness problem.

We have witnessed the presentations of results from three different operational experiments for determining the parameter γ. We may see the fourth experiment at Tucson become operational soon. Because these experiments are now operational and capable of repetition, we can expect to see the realistic error uncertainties decrease with time. If we take the results at face value, then the General Theory of Relativity continues to be the odds-on favorite even though we can expect to see some contradictions once in a while that will continue to keep us uneasy. However, I believe that unless these operational experiments show us that γ is definitely not unity, it will not interest us as much five years hence as will β or the solar quadrapole.

While I have made my preference for General Relativity clear, it would be unwise to accept the results of these recent experiments on γ^* uncritically. On the other hand, much of the criticism expressed outside the conference corridors was completely unfounded. Muhleman, Anderson, and Shapiro had much to say, and they did so with respectable candor.

In a technical sense, Muhleman's presentation was the most instructive, yet the least understood and the least appreciated. I had expected L. T. Little to be rather critical, but it appears that there is no disagreement with Muhleman at a level of accuracy of about 1%. Muhleman's approach was that the plasma delay on the radio signal was not an error term in itself but one of the physical effects to be measured. He uses a well known solar corona model that is static. The variability of the plasma is not represented in the model, and he regards it as an error. It was clear that MacDoran did not seem to agree with Muhleman on the character of the error. MacDoran's principal interest was in relating the fluctuations in the solar corona with the McMath regions on the sun. He demonstrates that these fluctuations occur as far out as 40 solar radii. What we need to know to understand this problem is how much these fluctuations change the integrated electron content in the ray path of the radio signal. How long do they persist and how frequently do they occur?

Unfortunately, the Muhleman-Anderson team needed the ranging capability at the Goldstone radio antenna at the same time MacDoran would have liked to use it. Consequently, MacDoran had only a limited time to gather his information. It would have been useful if he had provided us with a kind of histogram so that the solar event frequency could have been more easily extrapolated to regions near the sun.

For the well documented event of May 29, 1970, the solar corona caused a fluctuation in range of about 55 m in the course of a few hours. The time delay effect at this date is approximately 30 km. At the time of the conference, all the range points on Mariner VI and VII were evenly weighted with an error of 90 m. MacDoran may be correct that the solar fluctuations would be larger and more frequent nearer the sun; nonetheless, 90 m seems pretty conservative. If the fluctuations were too large, then the system would have lost radio lock. Consequently, Muhleman's contention that error in γ^* due to the solar corona is within 1% seems credible.

The conference came a little too soon for a completed presentation of the time delay results from Mariner 6 and 7. Anderson and Esposito had not had time to properly coordinate with Muhleman or with Curkendall. They had an unrefined result which they were still attempting to digest. The task of choosing a least-squares fit of 18 parameters to tracking data takes both skill and fortitude. They obtained fine results when processing sequences of ranging points of several weeks' duration, but when attempting to fit 14 months of tracking data, they found semiperiodic residuals which they could not explain. They systematically eliminated various causes such as the effect of the motion of Mercury on the motion of the earth, and demonstrated to their own satisfaction that the true cause was the random nongravitational forces on the spacecraft. The constant nongravitational forces can be estimated very accurately, but the stochastic accelerations (of the order of 10^{-10} m/s^2) can build up in error as time to the three-halves power. (Anderson has not published a derivation of this formula.)

It is easy for me to accept the thoroughness and the conscientiousness of the JPL and MIT groups. Their results are dependent upon the proper design of their computer programs, and it appears that part of the business is accepting that all this is in order also. I believe that I am reflecting the concerns of others more than my own when I say that it leaves one uneasy.

Shapiro's group at MIT performs the time-delay experiment by reflecting X-band radar waves from the planets. The principal sources of error are due to the roughnesses of the planets. Their data-processing program is not the same as JPL's, so the fact that their results are also consistent with General Relativity lends more credibility to both groups.

The MIT group presents a formal standard error of 0.03 and a realistic estimate of uncertainty of 0.05. The JPL group gives a formal error of 0.14 and a realistic estimate of uncertainty of 0.04. Anderson went to a little more effort to explain the character of his worries, and he seems to be influenced by the perplexing analytical findings of Curkendall, et al., with respect to sequential estimation of the parameter γ^* and the level of the nongravitational forces acting on the Mariner 6 and 7 spacecraft.

The details are numerous and not easy to follow, but both groups vary a variety of parameters systematically and see what happens to their solutions for the parameter γ^*. From this experience and from previous experience, they choose a "realistic" estimate of the error. Somewhere, I feel, the human personality enters into this choice of conservatism at the gut level.

Friedman attempted to evaluate the potential of future planetary missions for the determination of the parameters we have been discussing. The Mars 1971 orbiter looks promising for determining a more refined value of $\gamma^* - 1$% or better. The most interesting possibility lies in extending the Mariner Venus-Mercury 1973 mission. Friedman has performed a covariance analysis and simulated the type of data problems that Anderson, Esposito, and Muhleman have been encountering. He also attempted to introduce some conservatism into his model dealing with the nongravitational forces, and he tries to account for correlations between the various parameters. There is an element of judgment involved in this, but since Friedman has worked with both the JPL and MIT groups, we can assume that his conservatism has a precedent. Presumably ESRO will be undertaking a similar error analysis very soon. With the results of the Mariners in a couple of years, we will be able to test these projective assumptions and better understand how to analyze the possibilities of future space missions.

This conference once again gave us the opportunity to feel the enormous gap that exists between the theoreticians and the technicians. The informational background and personality orientations are sufficient reasons to create a chasm, but each group has its own political condensations which create gaps within the gaps. One man wants to preserve his viewpoint, and he seizes upon the technological information that seems to support him. Meanwhile, the engineer worries about where the resources will come from to allow him to continue his own love, and he hopes that the holy water of pure science will either help him or at least not obstruct him.

Those of us primarily concerned with erecting a solid bridge across the gap must first look at the viewpoints of Muhleman and Shapiro. They were projecting in the right direction several years ago and during the interim have seen their optimism both shattered and reawakened. Shapiro has devoted the most time in his very challenging paper to how he feels we should proceed in the future, and he leaves no doubt that he believes ground-based operations come first. Some conference participants did not hide their skepticism and suspicion. Even if we admit that he is optmistic in his technical projections, his argument cannot be avoided indefinitely because at the present time it is a lower cost option.

There is overlap in the viewpoints of Muhleman and Shapiro on the potential uses of very long baseline interferometry. Muhleman also points out that centimeter and subcentimeter wavelengths might possibly be employed in making accurate doppler measurements on interplanetary spacecraft and thus provide an independent check on the time-delay experiment.

Ranging with spacecraft is comparatively new, and its accuracy appears to improve by approximately a factor of two every couple of years. For very greatly improved ranging accuracy, increased bandwidths are required, and it is likely that political constraints will be encountered before inherent physical problems are the limiting factors. At distances of an astronomical unit or greater, radio ranging and doppler tracking of spacecraft will continue to be fundamental to all interplanetary missions for some time. These comments seem to be the main thrust of the papers by Martin and Easterling.

Laser ranging with the moon, however, is of itself very exciting. It is encouraging to learn that the ranging accuracy is continuously improving. Despite Bender's careful and sensitive presentation, one cannot help but be staggered by the size and representation of the laser ranging team.

On the face of it, the possibilities for making checks on gravitational theories are most impressive. It would be presumptuous of me to question Brown's theory of the motion of the moon, but there are celestial mechanicians who doubt that the Newtonian description of the motion of the moon has ever been handled adequately. On the other hand, I am presumptuous enough to question the likelihood that the scalar tensor parameter ω is only about five, and I doubt if detecting the amplitudes of some special frequencies will give us a reliable method of estimating new parameters in the PPN approach. However, these mild reservations are not meant to be criticisms of the operations experiment itself, because we all know that these range observations will be important.

Of all the gravitational experiments now on the operational level, the one by Weber is the most suspenseful and the most significant to our scientific futures. Much has been said in criticism and defense of his experiment, and I have nothing original to add. If his findings are verified in the coming year by Braginsky and others, the implications will be profound. Under these circumstances, I find I am obliged to be a reactionary and assume that Weber has detected something we cannot explain as yet, rather than believe that the universe is falling apart. If it turns out he has been detecting gravitational waves all this time when he says he has, then, like so many others, I will suffer less the pain of embarrassment than I will enjoy the pleasure of being a new convert. We are all on his side, but is Nature?

Inasmuch as this conference was sponsored by two space agencies, I will confine my remaining criticisms to three very specific formal proposals presented to them to test gravitational theories. They are the gyroscope experiment at Stanford presented by Everitt, the satellite clock experiment at the Smithsonian presented by Vessot, and the ESRO solar satellite mission described by Israël. These proposed missions are potentially expensive by ordinary laboratory standards, so we have to ask ourselves two questions: (1) Are the scientific answers worth the cost? (2) Will the technological developments which run up the cost be sufficiently useful to justify the expenditure? To the first question, the American scientific community would give either a negative answer or a divided reaction at this time. For example, we have heard Thorne and Dicke debate the value of future gravitational red-shift experiments at this conference, so that the satellite clock experiment certainly does not enjoy a cross-the-board endorsement with the community of relativists. Perhaps there would be a different outlook if the experiment would test the red-shift formula to one part in 10^{12}.

The Stanford experiment has two parts, a measurement of the geodetic precession and a test of the precession of a gyroscope due to the Lense-Thirring terms in the space-time metric. The geodetic part of the experiment is an independent way of measuring the parameter γ. Unless the geodetic precession experiment gives results consistent with the time-delay and ray-bending experiments, we cannot have any confidence in the more sensitive Lense-Thirring precession experiment.

It is at this point that I would like to digress and partially contradict myself by mentioning one of the virtues of the PPN development. If the laws of conservation are not arbitrarily discarded, then, as Will has shown, all the PPN parameters can be expressed in terms of β and γ. Consequently, it is not necessary for the Stanford experiment to be so precise as to distinguish between the factors $(3 + 2\omega)/(4 + 2\omega)$ and one, but only to be capable of determining the rotation's direction. Distinction between the General Theory of Relativity and the Brans-Dicke Theory can be determined elsewhere. This would allow relaxing the experimental precision by an order of magnitude, more in line with the known positions of the fixed stars, and would ease some of the difficulties posed by O'Connell in his paper. (A glance at Table 6 in the Thorne, Will, Ni paper is instructive.) Even with these suggestions for lessening the precision requirements of the experiment, it is still a tour de force.

As indicated earlier, the potential scientific value of the ESRO mission concept is, in my opinion, contingent on the results of experiments which will be performed in the next few years.

Schiff, Fairbank, and Everitt had obviously given considerable thought to the second question I posed above. Schiff told me at the conference that he did not know whether the gyroscope experiment could be successfully performed in space but that he was certain that the research efforts of Fairbank and Everitt in low-temperature space technology would prove to be valuable, and Fairbank's paper does a creditable job of reflecting this attitude.

It was a little disappointing, therefore, that Vessot and the ESRO groups presented their concepts as single missions and had nothing to say on the programmatic potentialities of the technical developments required for these flights. Muhleman, in his paper, suggests one possible use of a very accurate and stable clock in a solar satellite. The ESRO concept also requires an atomic clock, but the ESRO advisors prefer a cesium beam clock because the hydrogen maser is both too heavy and does not, in their view, have a sufficiently long-term stability record. NASA, on the other hand, has been skeptical of the cesium clock for space flight purposes.

The drag-free concept was discussed thoroughly and very competently by both Juillerat and DeBra. It is an important development and will appear on the space scene soon. The question is how difficult and expensive it will be to manufacture and test a spacecraft system in which nongravitational accelerations are removed to a level of 10^{-2} m/s^2. For deep space missions, an accelerometer that was integrated twice would do just as well because the fundamental physical measurements are range displacements. The detailed accelerations themselves are not required. However, such a highly accurate system operating over several orders of dynamic range has not been advanced. We have already discussed the modeling on nongravitational forces, and it is clear that attitude-controlled spacecraft that are unattached to a planet in some way cannot compete in theory with a 10^{-12} m/s^2 drag-free system. The

47. Kustaanheimo, P., Phys. Letters, 23, 75, 1966.

48. Kustaanheimo, P., and Nuotio, V. S., Relativistic Theories of Gravitation. I. One-Body Problem, Dept. of Applied Mathematics, University of Helsinki, 1967.

49. Rastall, P., Can. J. Phys., 46, 2155, 1968.

50. Rastall, P., in press, 1970.

51. Yilmaz, H., Phys. Rev., 111, 1417, 1958.

52. Yilmaz, H., in Evidence for Gravitational Theories, Edited by C. Møller, Academic Press, New York, 1962.

53. Bergmann, P. G., Int. J. Theor. Phys., 1, 25, 1968.

54. Wagoner, R. V., Phys. Rev. D., 1, 3209, 1970.

55. Ni, W.-T., paper in preparation, 1971.

56. Einstein, A., Ann. Physik, 49, 769, 1916.

57. Jordan, P., 1955, Schwerkraft und Weltall, Friedrich Vieweg and Son, Braunschweig.

58. Nordtvedt, K., Jr., Ap. J., 161, 1059, 1970.

59. Littlewood, D. E., Proc. Camb. Phil. Soc., 49, 90, 1953.

60. Bergmann, O., Am. J. Phys., 24, 39, 1956.

61. Nordstrøm, G., Phys. Zeit., 13, 1126, 1912.

62. Nordstrøm, G., Ann. Physik, 42, 533, 1913.

63. Nordstrøm, G., Ann. Physik, 43, 1101, 1914.

64. Einstein, A., and Fokker, A. D., Ann. Physik, 44, 321, 1914.

65. Einstein, A., Ann. Physik, 38, 433, 1912.

66. Schild, A., Proc. Roy. Soc., 235A, 202, 1956.

67. Will, C. M., unpublished calculations, 1971.

68. Nordtvedt, K., Jr., Phys. Rev., 180, 1293, 1969.

69. Will, C. M., "Relativistic Gravity in the Solar System. II. Anisotropy in the Newtonian Gravitational Constant," in preparation, 1971.

70. Will, C. M., "Theoretical Frameworks for Testing Relativistic Gravity. III. Conservation Laws, Lorentz Invariance and Values of the Parametrized Post-Newtonian Parameters," in preparation, 1971.

71. Muhleman, D. O., and Reichley, P., Space Programs Summary 37-29, Vol. IV, 342, Jet Propulsion Laboratory, Pasadena, Calif., 1964.

72. Seielstad, G. A., Sramek, R. A., and Weiler, K. W., Phys. Rev. Letters, 24, 1373, 1970.

73. Shapiro, I. I., Phys. Rev. Letters, 13, 789, 1964.

74. Muhleman, D. O., and Reichley, P., Space Programs Summary 37-29, Vol. IV, 342, Jet Propulsion Laboratory, Pasadena, Calif., 1964.

75. Ross, D. K., and Schiff, L. I., Phys. Rev., 141, 1215, 1966.

76. Shapiro, I. I., Phys. Rev., 141, 1219, 1966.

77. Schiff, L. I., Proc. Nat. Acad. Sci. U.S.A., 46, 871, 1960.

78. Einstein, A., Preuss. Akad. Wiss. Sitz., 831, 1915.

79. Anderson, J. D., and Lorell, J., AIAA J., 1, 1372, 1963.

80. Schiff, L. I., Phys. Rev. Letters, 4, 215, 1960.

81. Dicke, R. H., Gravitation and the Universe, Jayne Lectures for 1969, American Philosophical Society, Philadelphia, 1969.

82. Will, C. M., Ap. J., 165, 409, 1971.

83. Nordtvedt, K., Jr., Phys. Rev., in press, 1971.

84. Bondi, H., Rev. Mod. Phys., 29, 423, 1957.

85. Kreuzer, L. B., Phys. Rev., 169, 1007, 1968.

86. Blatt, J. M., and Weisskopf, V. F., Theoretical Nuclear Physics, John Wiley and Sons, New York, 1952.

References

1. Dicke, R. H., The Theoretical Significance of Experimental Relativity, Gordon and Breach, Inc., New York, 1964.

2. Schild, A., in Evidence for Gravitational Theories, Edited by C. Møller, Academic Press, New York, 1962.

3. Thorne, K. S., and Will, C. M., Ap. J., 163, 595, 1971.

4. Nordtvedt, K., Jr., Phys. Rev., 169, 1017, 1968.

5. Will, C. M., Ap. J., 163, 611, 1971.

6. Eddington, A. S., The Mathematical Theory of Relativity, Cambridge University Press, London, 1922.

7. Robertson, H. P., in Space Age Astronomy, Edited by A. J. Deutsch and W. B. Klemperer, Academic Press, New York, 1962, p. 228.

8. Schiff, L. I., in Relativity Theory and Astrophysics. I. Relativity and Cosmology, Edited by J. Ehlers, American Mathematical Society, Providence, Rhode Island, 1967.

9. Baierlein, R., Phys. Rev., 162, 1275, 1967.

10. Dicke, R. H., Phys. Rev., 125, 2163, 1962.

11. Brans, C., and Dicke, R. H., Phys. Rev., 124, 925, 1961.

12. Farley, F. J. M., Bailey, J., Brown, R. C. A., Giesch, M., Jöstlein, H., van der Meer, S., Picasso, E., and Tannenbaum, M., Nuovo Cimento, 45, 281, 1966.

13. Frisch, D. H., and Smith, J. H., Am. J. Phys., 31, 342, 1963.

14. Durbin, R. P., Loar, H. H., and Havens, W. W., Jr., Phys. Rev., 88, 179, 1952.

15. Rössi, B., and Hall, D. B., Phys. Rev., 59, 223, 1941.

16. Ives, H. E., and Stilwell, G. R., J. Opt. Soc. Am., 28, 215, 1938.

17. Ives, H. E., and Stilwell, G. R., J. Opt. Soc. Am., 31, 369, 1941.

18. Alväger, T., Farley, F. J. M., Kjellman, J., and Wallin, I., Phys. Letters, 12, 260, 1964.

19. Lichtenberg, D. B., Meson and Baryon Spectroscopy, Springer-Verlag, New York, 1965, p. 18.

20. Sherwin, C. W., Phys. Rev., 120, 17, 1960.

21. Ageno, M., and Amaldi, E., Accademia Nazionale dei Lincei, 8 (Series VIII), 1, 1966.

22. Bailey, J., and Picasso, E., Prog. in Nucl. Phys., 12, 43, 1970.

23. Cook, A. H., Metrologia, 1, 84, 1965.

24. Pound, R. V., and Rebka, G. A., Phys. Rev. Letters, 4, 337, 1960.

25. Pound, R. V., and Snider, J. L., Phys. Rev., 140, B788, 1965.

26. Misner, C. W., Thorne, K. S., and Wheeler, J. A., Gravitation, W. H. Freeman and Company, San Francisco, in press, 1972.

27. Brault, J., Bull. Am. Phys. Soc., 8, 28, 1963.

28. Kleppner, D., Vessot, R. F., and Ramsey, N. F., Ap. and Space Sci., 6, 13, 1970.

29. Havas, P., paper in press, 1970.

30. Geisler, P., and MacVittie, G. C., Astron. J., 70, 14, 1965.

31. Peebles, P. J., and Dicke, R. H., Phys. Rev., 127, 629, 1962.

32. Schiff, L. I., Am. J. Phys., 28, 340, 1960.

33. Fock, V., The Theory of Space, Time and Gravitation, Pergamon Press, New York, 1964, p. 238.

34. Fierz, M., and Pauli, W., Proc. Roy. Soc., 173A, 211, 1939.

35. Kustaanheimo, P., Ann. Akad. Sci. Fenn., 228, 3, 1957.

36. Whitehead, A. N., The Principle of Relativity, Cambridge University Press, Cambridge, 1922.

37. Synge, J. L., Proc. Roy. Soc., 211A, 303, 1952.

38. Birkhoff, G. D., Proc. Nat. Acad. Sci., 29, 231, 1943.

39. Hoyle, F., M.N.R.A.S., 120, 266, 1960.

40. Hoyle, F., and Narlikar, J. V., Proc. Roy. Soc., 273A, 1, 1963.

41. Hoyle, F., and Narlikar, J. V., Proc. Roy. Soc., 278A, 465, 1964.

42. Milne, E. A., Proc. Roy. Soc., 160A, 7, 1937.

43. Milne, E. A., Kinematical Relativity, Oxford University Press, London, 1948.

44. Poincaré, H., Rend. Circ. Mat. di Palermo, 21, 166, 1906.

45. Whitrow, G. J., and Morduch, G. E., in Vistas in Astronomy, Vol. 6, Edited by A. Beer, Pergamon Press, Oxford, 1965.

46. Whitrow, G. J., and Morduch, G. E., Nature, 188, 790, 1960.

47. Kustaanheimo, P., Phys. Letters, 23, 75, 1966.

48. Kustaanheimo, P., and Nuotio, V. S., Relativistic Theories of Gravitation. I. One-Body Problem, Dept. of Applied Mathematics, University of Helsinki, 1967.

49. Rastall, P., Can. J. Phys., 46, 2155, 1968.

50. Rastall, P., in press, 1970.

51. Yilmaz, H., Phys. Rev., 111, 1417, 1958.

52. Yilmaz, H., in Evidence for Gravitational Theories, Edited by C. Møller, Academic Press, New York, 1962.

53. Bergmann, P. G., Int. J. Theor. Phys., 1, 25, 1968.

54. Wagoner, R. V., Phys. Rev. D., 1, 3209, 1970.

55. Ni, W.-T., paper in preparation, 1971.

56. Einstein, A., Ann. Physik, 49, 769, 1916.

57. Jordan, P., 1955, Schwerkraft und Weltall, Friedrich Vieweg and Son, Braunschweig.

58. Nordtvedt, K., Jr., Ap. J., 161, 1059, 1970.

59. Littlewood, D. E., Proc. Camb. Phil. Soc., 49, 90, 1953.

60. Bergmann, O., Am. J. Phys., 24, 39, 1956.

61. Nordstrøm, G., Phys. Zeit., 13, 1126, 1912.

62. Nordstrøm, G., Ann. Physik, 42, 533, 1913.

63. Nordstrøm, G., Ann. Physik, 43, 1101, 1914.

64. Einstein, A., and Fokker, A. D., Ann. Physik, 44, 321, 1914.

65. Einstein, A., Ann. Physik, 38, 433, 1912.

66. Schild, A., Proc. Roy. Soc., 235A, 202, 1956.

67. Will, C. M., unpublished calculations, 1971.

68. Nordtvedt, K., Jr., Phys. Rev., 180, 1293, 1969.

69. Will, C. M., "Relativistic Gravity in the Solar System. II. Anisotropy in the Newtonian Gravitational Constant," in preparation, 1971.

70. Will, C. M., "Theoretical Frameworks for Testing Relativistic Gravity. III. Conservation Laws, Lorentz Invariance and Values of the Parametrized Post-Newtonian Parameters," in preparation, 1971.

71. Muhleman, D. O., and Reichley, P., Space Programs Summary 37-29, Vol. IV, 342, Jet Propulsion Laboratory, Pasadena, Calif., 1964.

72. Seielstad, G. A., Sramek, R. A., and Weiler, K. W., Phys. Rev. Letters, 24, 1373, 1970.

73. Shapiro, I. I., Phys. Rev. Letters, 13, 789, 1964.

74. Muhleman, D. O., and Reichley, P., Space Programs Summary 37-29, Vol. IV, 342, Jet Propulsion Laboratory, Pasadena, Calif., 1964.

75. Ross, D. K., and Schiff, L. I., Phys. Rev., 141, 1215, 1966.

76. Shapiro, I. I., Phys. Rev., 141, 1219, 1966.

77. Schiff, L. I., Proc. Nat. Acad. Sci. U.S.A., 46, 871, 1960.

78. Einstein, A., Preuss. Akad. Wiss. Sitz., 831, 1915.

79. Anderson, J. D., and Lorell, J., AIAA J., 1, 1372, 1963.

80. Schiff, L. I., Phys. Rev. Letters, 4, 215, 1960.

81. Dicke, R. H., Gravitation and the Universe, Jayne Lectures for 1969, American Philosophical Society, Philadelphia, 1969.

82. Will, C. M., Ap. J., 165, 409, 1971.

83. Nordtvedt, K., Jr., Phys. Rev., in press, 1971.

84. Bondi, H., Rev. Mod. Phys., 29, 423, 1957.

85. Kreuzer, L. B., Phys. Rev., 169, 1007, 1968.

86. Blatt, J. M., and Weisskopf, V. F., Theoretical Nuclear Physics, John Wiley and Sons, New York, 1952.

N72-13776

TECHNICAL REPORT STANDARD TITLE PAGE

1. Report No. 33-499	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle PROCEEDINGS OF THE CONFERENCE ON EXPERIMENTAL TESTS OF GRAVITATION THEORIES NOVEMBER 11-13, 1970, CALIFORNIA INSTITUTE OF TECHNOLOGY, PASADENA, CALIFORNIA		5. Report Date November 1, 1971 6. Performing Organization Code
7. Author(s) JPL Staff		8. Performing Organization Report No.
9. Performing Organization Name and Address JET PROPULSION LABORATORY California Institute of Technology 4800 Oak Grove Drive Pasadena, California 91103		10. Work Unit No. 11. Contract or Grant No. NAS 7-100 13. Type of Report and Period Covered Technical Memorandum
12. Sponsoring Agency Name and Address NATIONAL AERONAUTICS AND SPACE ADMINISTRATION Washington, D.C. 20546		14. Sponsoring Agency Code

15. Supplementary Notes

16. Abstract

This document is a compilation of papers presented by various organizations at the Conference on Experimental Tests of Gravitation Theories, which was sponsored by NASA/European Space Research Organization/JPL. Thirty-two formal presentations of the conference and a summation and critique of the proceedings are presented.

17. Key Words (Selected by Author(s))	18. Distribution Statement
Radio Astronomy Relativity Scientific Instruments Tracking	Unclassified -- Unlimited

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	256	

HOW TO FILL OUT THE TECHNICAL REPORT STANDARD TITLE PAGE

Make items 1, 4, 5, 9, 12, and 13 agree with the corresponding information on the report cover. Use all capital letters for title (item 4). Leave items 2, 6, and 14 blank. Complete the remaining items as follows:

3. Recipient's Catalog No. Reserved for use by report recipients.

7. Author(s). Include corresponding information from the report cover. In addition, list the affiliation of an author if it differs from that of the performing organization.

8. Performing Organization Report No. Insert if performing organization wishes to assign this number.

10. Work Unit No. Use the agency-wide code (for example, 923-50-10-06-72), which uniquely identifies the work unit under which the work was authorized. Non-NASA performing organizations will leave this blank.

11. Insert the number of the contract or grant under which the report was prepared.

15. Supplementary Notes. Enter information not included elsewhere but useful, such as: Prepared in cooperation with... Translation of (or by)... Presented at conference of... To be published in...

16. Abstract. Include a brief (not to exceed 200 words) factual summary of the most significant information contained in the report. If possible, the abstract of a classified report should be unclassified. If the report contains a significant bibliography or literature survey, mention it here.

17. Key Words. Insert terms or short phrases selected by the author that identify the principal subjects covered in the report, and that are sufficiently specific and precise to be used for cataloging.

18. Distribution Statement. Enter one of the authorized statements used to denote releasability to the public or a limitation on dissemination for reasons other than security of defense information. Authorized statements are "Unclassified—Unlimited," "U. S. Government and Contractors only," "U. S. Government Agencies only," and "NASA and NASA Contractors only."

19. Security Classification (of report). NOTE: Reports carrying a security classification will require additional markings giving security and downgrading information as specified by the Security Requirements Checklist and the DoD Industrial Security Manual (DoD 5220.22-M).

20. Security Classification (of this page). NOTE: Because this page may be used in preparing announcements, bibliographies, and data banks, it should be unclassified if possible. If a classification is required, indicate separately the classification of the title and the abstract by following these items with either "(U)" for unclassified, or "(C)" or "(S)" as applicable for classified items.

21. No. of Pages. Insert the number of pages.

22. Price. Insert the price set by the Clearinghouse for Federal Scientific and Technical Information or the Government Printing Office, if known.