PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A
MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

The Constancy of G and Other Gravitational Experiments

R. D. Reasenberg

Phil. Trans. R. Soc. Lond. A 1983 **310**, 227-238
doi: 10.1098/rsta.1983.0083

Phil. Trans. R. Soc. Lond. A **310**, 227–238 (1983)

Printed in Great Britain

The constancy of G and other gravitational experiments

By R. D. Reasenberg

Radio and Geoastronomy Division, Smithsonian Astrophysical Observatory, Harvard-Smithsonian Center for Astrophysics, Cambridge, Massachusetts 02138, *U.S.A.*

Traditionally, theories of gravitation have received their most demanding tests in the solar-system laboratory. Today, electronic observing technology makes possible solar-system tests of substantially increased accuracy. We consider how these technologies are being used to study gravitation with an emphasis on two questions:

(i) Dirac and others have investigated theories in which the constant of gravitation, G, appears to change with time. Recent analyses using the Viking data yield $|\dot{G}/G| < 3\times10^{-11}$ per year. With further analysis, the currently available ensemble of data should permit an estimate of $\dot{G}/G$ with an uncertainty of 10^{-11} per year. At this level it will become possible to distinguish among competitive theories.

(ii) Shapiro's time-delay effect has provided the most stringent solar-system test of general relativity. The effect has been measured to be consistent with the predictions of general relativity with a fractional uncertainty of 0.1%. An improved analysis of an enhanced data set should soon permit an even more stringent test.

Technology now permits new kinds of tests to be performed. Among these are some that measure relativistic effects due to the square of the (solar) potential and others that detect the Earth's 'gravitomagnetic' field (the Lense–Thirring effect). These experiments, and the use of astrophysical systems are among the experimental challenges for the coming decades.

1. Introduction

The energetic interplay of theory and experiment has led to the rapid advance of our understanding of the physical universe. In no other field has our knowledge of reality evolved so rapidly. For many decades it appeared that general relativity was to be an exception to this rule. Now, thanks to the application of modern technology, gravitational physics is again an experimental subject. Most tests today are limited to the first-order post-Newtonian effects and have, at best, an accuracy of a part in 10^3. Of all the reputable experiments done, there has not been one that yielded results inconsistent with the predictions of general relativity. Over the next decades, the evolving level of technology applied to these tests will lead to order-of-magnitude increases in the accuracy of the results. But we must ask ourselves whether we should not be looking at alternative régimes for testing this theory of gravitation. The identification of new physically realizable experiments is a significant frontier for relativity today.

The classical laboratory for gravity research is the solar system, where the theories of Newton and Einstein have been tested in turn. Although certain observations are particularly important to a particular 'test', the entire ensemble of data is advantageously used for each of the possible tests. In the traditional laboratory environment, an apparatus is assembled and operated in such a manner as to obtain an estimate of a single or, at most, a few parameters of interest. These are generally described as physical quantities. The contrivances of a terrestrial laboratory are precluded in a solar-system study. Here an ensemble of diverse data is collected and a multi-parameter model is fitted to the data. The vast majority of the parameters estimated are of no interest to the relativist. These 'nuisance' parameters, however, are essential to the reliable determination of those few quantities that are of relativistic interest.

The study of post-Newtonian effects in the solar system is difficult. They are scaled by the solar potential which reaches 2×10^{-6} at the limb of the Sun, a region accessible to the experimenter only by means of photons. As will be seen below, the solar-system laboratory is useful only because of the availability of modern technologies of electronics and space travel. The natural alternative, terrestrial laboratory experiments, is not yet practical because of the small strength of the gravitational interaction between laboratory bodies and the high level of ambient noise, both natural and man made. Finally, there is the possibility of using astrophysical systems as serendipitous laboratories. For such systems, the gravitational potential, ϕ, may be large. However, these systems are remote, unfamiliar, and generally observed by a single means. Thus, degeneracy and model ambiguity are major problems. The pulsar in a binary system, PSR 1913+16, is a prime example of an astrophysical system that yields information on the nature of the laws of gravitation.

TABLE 1. SOME RELATIVISTIC EFFECTS

effect	parametric dependence	state of determination: present	expected: σ	proposed: σ
deflexion of light	$\frac{1}{2}(1+\gamma)$	1.007 ± 0.009	0.001	10^{-6}
perihelion advance	$\frac{1}{3}(2+2\gamma-\beta)$	0.98 ± 0.04	0.003	—
gravitational redshift	$\frac{\text{observed}}{\text{expected}}$	1.0000025 ± 0.00014	—	10^{-8}
Shapiro time delay	$\frac{1}{2}(1+\gamma)$	1.000 ± 0.001	0.0005	—
Nordtvedt's principle of equivalence violation	$\eta = 4\beta-\gamma-3$	0.0 ± 0.015	0.005	—
secular variation of G	$\dot{G}/G$	$(0 \pm 3) \times 10^{-11}/\text{a}$	$10^{-11}/\text{a}$	—
gravitational radiation damping of orbit	$\frac{\text{observed}}{\text{expected}}$	0.957 ± 0.09	—	—
geodetic precession	$\frac{1}{3}(1+2\gamma)$	—	—	0.001
Lense–Thirring effect*	$\frac{1}{2}(1+\gamma)$	—	—	0.1

* The significance of a test of this effect transcends the importance of determining the p.p.n. coefficient. The detection by the Stanford gyroscope experiment of the Lense–Thirring effect would be the first experimental verification of the relativistic effect of the (rotational) motion of the source.

Tests of general relativity, like those of any theory, are often described in terms of 'effects'. These may be thought of as predictions or observations that differ from those expected according to a previous description of the laws of nature. Table 1 lists some effects that have been or will be important in tests. As a list of known relativistic effects, it is far from complete. For example, it excludes variations of the gravitational constant with distance on a laboratory scale, gravitational waves, and quantum gravity effects.

In the following sections, we will consider first the ensemble of data under analysis at the Center for Astrophysics (C.f.A.). This work has been in collaboration with R. W. Babcock, J. F. Chandler and I. I. Shapiro at C.f.A. and with R. W. King at M.I.T. We will then discuss areas of relativistic interest, the secular variation of G and the Shapiro time-delay test. Finally, § 5 will be devoted to a discussion of experiments currently under consideration. Some of these are well developed and ready for implementation; others are in an early phase of a long process and have an uncertain future.

2. SOLAR SYSTEM DATA SET

Before considering specific tests, we review the current state of solar-system analysis at C.f.A. Table 2 shows our present data set. For current purposes, the most powerful data are the Viking

Lander observations, of which we use about 1060, spanning 6 years. In addition, we have 1075 Doppler observations of the Viking Landers and 4060 Viking Orbiter normal points (n.p.). Although these last two kinds of data have little direct value for the relativity tests, the former are useful for determining the rotation of Mars, and the latter establish the mean distance of the Landers from the Mars equator. Since these are part of the solar-system modelling problem, the Doppler and n.p. data are indirectly useful for the relativity analysis. The Mariner-9 mission yielded spacecraft orbiter normal points for 1971 and 1972. Although there are fewer than 200 of these, they are important in determining the Mars ephemeris because they have epochs long before the Viking era. The table shows no Mars radar data, though we do have 2.5×10^4 of these, all of which we have previously processed. The combination of Viking and Mariner-9 data is so much more powerful that we temporarily discarded the Mars radar data.

Table 2. Combined sets of data

source†	no. of data	approximate range of error assumed in estimator: min	max	unit
Viking				
Lander delay (plasma corrected)	798	20	60	ns
Lander delay (not plasma corrected)	263	50	300	ns
Orbiter n.p.‡	4060	100	900	ns
Lander Doppler	1075	20	40	mHz
l.l.r.				
Observing session n.p.§	2613	6	14	ns
Mariner 9				
Orbiter n.p.‡	185	0.1	10	µs
radar				
Mercury	642	1	15	µs
Venus	784	1	15	µs
meridian circle‖				
Sun	1023	≈ 2		″
Moon	212	≈ 0.5		″
inner planets (M, V, M)	1518	≈ 1		″
outer planets (J, S, U, N)	1643	≈ 1		″
outer planet n.p.	6	25	500	µs

† All observables are time delays except for the Viking Lander Doppler and meridian circle data.
‡ The orbiter normal point (n.p.) is a compressed datum: the equivalent Earth–Mars time delay measured between the centres of mass of the planets.
§ The lunar laser ranging (l.l.r.) normal point (n.p.) is a single estimate of the round trip propagation time between a tracking station and a single lunar retroreflector. The estimate is an average based on all photons received during an observing sequence.
‖ The data are a mixture of right ascension and declination measurements.
¶ The outer planet normal point (n.p.) is a compressed datum from a spacecraft encounter with either Jupiter or Saturn. The n.p. is the equivalent Earth–planet time delay measured between the centres of mass of the planets.

Lunar laser ranging has given us 2600 so-called laser ranging normal points. Traditionally, the lunar data were analysed separately from the planetary data, in part because the scientific basis for the work is different and in part because different subgroups within the research group had been doing the analyses.

In principle, the combination of the lunar and planetary data sets should have been a simple matter as each had been previously reduced to normal equations. Even though disparate nominal values were used for the adjustable parameters in the two reductions, a simple linear operation, which our software will perform, should properly combine the data. In practice, the problem is more complicated. The estimator shows large (0.9999) correlations that tend to enhance the undesirable effects of both estimator nonlinearity and (economically necessary) approximations that cause errors in the variational equations. The common (and economically justified) practice of re-using previously calculated variational equations exacerbates this problem. It therefore is necessary to iterate the estimator in order to combine properly the data. At the same time, the combined data set may require an enhanced model. Thus there are two competing lines of activity that must be pursued in balance to achieve a properly combined data set.

It now appears that we have solved the most serious of the problems and successfully have combined the data sets. To remove remnants of the separate histories of the two data groups required four iterations of the estimator, the last of which was distinguished in that we re-computed the variational equations.

3. Secular variation of G

Dirac (1937, 1938) has investigated the cosmological consequences of the large numbers hypothesis. It was noted even earlier that it is possible to combine physical constants to create dimensionless numbers that generally differ from unity by at most a few orders of magnitude. Yet there are some such numbers that are extremely large. One of these is the ratio of the electric to the gravitational force between an electron and a proton; it is close to 10^{40}. Another is the age of the universe expressed in units of atomic time; it is also close to 10^{40}. Finally there is the mass of the visible universe expressed in proton masses; it is close to $10^{(40\times 2)}$. The hypothesis is that this coincidence is a message, not an accident; perhaps these quantities are related by some small time-invariant constants. If that is correct, then one or more of the 'constants' used to make each of the large numbers must be time varying. The original description was that the gravitational 'constant' was the most likely candidate for the time variable. A discussion of some alternatives is given by Dyson (1972).

How would one detect such a variable G in the solar system? We should have a complete theory of dynamics with a variable G, yet none exists. As an experiment, based on Dirac's early papers, we use an *ad hoc* model (Shapiro 1971)

$$G = G_0 + \dot{G}(t - t_0), \tag{1}$$

where t_0 is a convenient recent epoch. This is not an entirely satisfactory way to describe the dynamics, but it allows us to determine whether or not there may be a secular variation of G.

With any of the schemes by which one has a time-varying gravitational constant, there are two kinds of times. One is the time determined by any system dominated by gravity, so-called gravitational time, such as the independent argument of an ephemeris. The other is the time determined by any system that is not dominated by gravity, so-called atomic time, such as the time kept by a hydrogen maser atomic clock. Of course, there may be more kinds of time defined for example by the strong or the weak interaction. These need not be considered here.

The use of (1), when applied to a planetary orbit, gives rise to a relation among the gravitational constant, the semi-major axis a, the period p, and the mean motion n (Counselman & Shapiro 1968),

$$2\dot{G}/G = -2\dot{a}/a = -\dot{p}/p = \dot{n}/n. \tag{2}$$

As suggested by (2), equation (1) leads to two observable effects in the solar system (Shapiro

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

1964*a*). First, the scale of the system appears to change. Second, the linearly varying period gives rise to a quadratically growing increment in the mean longitude of each body. In principle, both effects can be measured. For observations that span many orbital periods, the latter effect dominates. For Mercury, over a period of 10 years, the quadratic effect is two orders of magnitude larger than the linear.

An alternate approach is to consider explicitly the two time scales. A comparison of this with the approach represented by (1) is given by Canuto *et al.* (1983). They find that either approach will suffice for the purpose of detecting the existence of a secular variation of G. Should it be found by either approach that $\dot{G} \neq 0$, a more refined analysis would be required to distinguish the alternate formulations.

A later version of Dirac's hypothesis included two forms of matter creation (Dirac 1974). In the first, multiplicative creation, new matter appeared where old matter was present. In the second, additive creation, new matter appeared uniformly throughout the universe. These two postulated modes of matter creation give rise in turn to a pair of modifications of (2):

$$\dot{G}/G = \dot{n}/n, \tag{3}$$

$$\dot{G}/G = -\dot{n}/n. \tag{3a}$$

Again, there are the linear and quadratic observable effects. Recently, both Dirac (1979) and Canuto *et al.* (1977) have advanced theories that feature the two kinds of time. The observable consequences of these new descriptions are similar to those of (1). However, the coefficients of the linear and quadratic terms have a different ratio in the new theories.

Theoretical predictions yield $\dot{G}/G = \alpha H$ where α is a small constant and H is the Hubble constant. We will consider Mercury after nine years of observation. The accumulated effect of $\dot{G}/G = 10^{-11}$ is a mean-longitude shift such that a radar time-delay echo will show a peak increment of 8 μs. If we compare that with the uncertainty of modern radar measurement, as little as 0.2 μs, it appears that it should be very easy to determine whether or not G varies at $10^{-11}\,G$ per year. In fact, a simple order-of-magnitude calculation, which assumes a thousand observations and propagates the errors, yields an uncertainty of $10^{-14}\,G$ per year.

The problem is not so simple. Aside from the fractional solar mass loss of about 10^{-14} per year, there are problems in the analysis of the data. First, we do not know *a priori* either the initial phase or period of the Mercury orbit. For uniformly spaced and weighted data, the change from estimating one parameter to three causes the uncertainty in the estimate of $\dot{G}$ to increase by a 'masking factor' of six. But the solar-system analysis involves a very large number of parameters, not three. In a typical study of $\dot{G}$ with radar data, we find that the masking factor is about 80. Above, we used an uncertainty of measurement of 0.2 μs, about the best now available. It is not typical historically and it is almost irrelevant because of the unknown planetary topography. If we define the radar delay as a measure of the distance to the mean target-body surface, then the accuracy must include the unmodelled part of the topography; the ratio of measurement accuracy to peak effect is reduced from 40 to more like 3. Not all observations are made when the line of sight is tangent to the orbit of Mercury. In addition, there are model errors that interfere with the accuracy of the interpretation of the measurements. Our standard procedure is to perform a series of numerical experiments to uncover the effects of these errors and thus to provide realistic estimates of parameter uncertainties. When all of these factors are taken into consideration, the determination of $\dot{G}/G$ that initially appeared to have an uncertainty of 10^{-14} per year

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

is found to have an uncertainty of 10^{-10} per year. Our published result (Reasenberg & Shapiro 1976) was a bound on $\dot{G}/G$ of $(5 \pm 10) \times 10^{-11}$ per year. Surprisingly, this result has been taken by some as evidence for a non-zero $\dot{G}$. This interpretation is, of course, not correct; we have found no significant evidence for a non-zero $\dot{G}$.

What is the prognosis for improving the accuracy of the $\dot{G}$ determination? Our present data set is listed in table 1. In some recent numerical studies we have investigated the accuracy to which we will be able to determine $\dot{G}/G$. A preliminary result is $|\dot{G}/G| < 3 \times 10^{-11}$ per year, where the bound is one standard deviation. However, by the time we finish the present series of studies, we expect to see a further reduction in the uncertainty by a factor of at least three. What are the limits to the accuracy with which $\dot{G}$ can be estimated using solar-system data? As previously noted, the uncertainty of the solar mass loss rate, about $10^{-14}\,M_{\odot}$ per year, sets a bound on the accuracy of a $\dot{G}$ estimate from planetary observations. The asteroids probably represent the most serious present limit. We now estimate the masses of eight large asteroids. Unfortunately, there are many asteroids excluded from our solutions that are nearly as large as some that are included. These and the thousands of smaller asteroids generate a 'gravitational noise' that is nearly impossible to model correctly, although various approximations are possible. This gravitational noise probably sets a bound of 0.5×10^{-11} per year to $\sigma(\dot{G}/G)$ determined from observations of the motion of Mars (J. Williams, Jet Propulsion Laboratory, personal communication 1983).

4. The fourth test of relativity

The Shapiro (1964*b*) time-delay effect provides the basis for the fourth test, the most stringent solar-system test of relativity (compare with tests of the underlying principle of equivalence by means of the gravitational redshift (Vessot *et al.* 1980) and references therein). For a signal that passes close to the Sun during a round trip from Earth to a spacecraft or planet, the round-trip propagation time is that expected from Euclidean geometry plus an additional term,

$$\Delta\tau = \frac{2r_0}{c} S \ln\left[\frac{r_e + r_p + R}{r_e + r_p - R}\right], \tag{4}$$

where $r_0 = 3\,\text{km}$, $2r_0/c = 20\,\mu\text{s}$, and $S = \frac{1}{2}(1+\gamma)$. In (4), r_e, r_p, and R are, respectively, the distances from the Sun to the Earth, from the Sun to the target, and from the Earth to the target; γ is one of the parameters of the PPN formalism. (See Will 1981, and references therein.) For an impact parameter d substantially less than r_p or r_e:

$$\frac{r_e + r_p + R}{r_e + r_p - R} \approx \frac{4r_e r_p}{d^2}. \tag{5}$$

In estimating γ from time-delay data, as in the other uses of the time-delay data, there are three principal problems:

(1) determination of the location of the end points of the observations;

(2) measurement of the time delay to high accuracy, including the calibration of the instrumentation; and

(3) correction of the measured time delay to the equivalent vacuum delay.

For the time delay experiment, the solar corona has traditionally been the most serious obstacle. The corona is highly variable: fluctuations are of the order of the mean. However, a simple approximate expression for the average effect of the corona on signal delay under quiet-Sun conditions can be given:

$$\tau_p = 300/f^2 d \tag{6}$$

where τ_p is in μs, f is in GHz, and d, the impact parameter of the signal, is in solar radii.

There are several important results from pre-Viking analyses of time-delay experiments. Based on radar data through 1971, both X-band (8 GHz) and U.H.F. (400 MHz), the coefficient of the Shapiro effect, $\frac{1}{2}(1+\gamma)$, was estimated to be 1.01 with an uncertainty of 5 %. As a result of adding radar data taken in 1972, the estimate became 1.00 ± 4 %. An analysis of tracking data from the Sun-orbiting Mariners 6 and 7 yielded an estimate of 1.00 ± 3 %. Finally, there was the Mariner-9 mission: the spacecraft was placed in a 12 h orbit around Mars. Because of the short period of the orbit (cf. Mariners 6 and 7) the effects of unmodelled accelerations could not accumulate pathologically. This advantage was in part offset by the effect of the poorly known irregularities of the Mars gravitational potential. That mission yielded a measurement of the Shapiro effect of 1.00 ± 2 %.

The Viking mission provided a dramatic improvement. There were four Vikings, two orbiters and two landers, each equipped with a high-gain antenna. For the Vikings, the time delay observable far from superior conjunction has a precision of about 10 ns, that is, the distance measurements are uncertain by roughly the height of a person. Near superior conjunction, the measurement uncertainty is dwarfed by the plasma problem. Each Orbiter was equipped with a transponder that would receive an S-band signal from the ground, and send back coherent S-band and X-band signals. From the difference in the delays of these returning signals, one can determine the plasma columnar content of the path between Earth and Mars. Because they were connected to a massive body, the Landers, unlike Orbiters, were not buffeted by non-gravitational low-thrust forces. The Landers' orbits with respect to the centre of mass of Mars are relatively easy to model to high accuracy; the Landers do not execute a substantial random walk. True, there are small geophysical effects. The spin-rate of Mars is uneven and that causes the Lander to move a few metres with respect to the prediction of a simple rotation model. Mars may wobble with about a 190 day period, corresponding to the Earth's 400 day wobble. But these effects are very small compared with the kinds of effects which have plagued the analysis of spacecraft data.

Figure 1 shows the observing situation. A signal is sent from Earth, up to the Lander, and returns to the Earth which has moved since the signal was sent. A second tracking station is used to send an S-band signal to an Orbiter which returns signals at S-band and at X-band. Thus, it is the path from the Orbiter to Earth that provides the only measure of the plasma in the vicinity of the Sun between the Earth and Mars. We make the approximation that all the plasma is in a thin screen perpendicular to that line of sight and containing the Sun. In that approximation, we can determine the plasma contribution to the Lander signal if we are willing to ignore the fact that the four paths do not pierce the thin screen in the same place. The Lander plasma delay is calculated by assuming that the down-link contribution is the same as it is for the Orbiter and the up-link contribution is the same as it was for the Orbiter down-link at a time earlier by one Sun–Mars–Sun propagation time.

A preliminary analysis was done based on observations made during 40 days surrounding the first superior conjunction on 25 November 1976. This first analysis was done very crudely in comparison with our present work. Nonetheless, it yielded $\frac{1}{2}(1+\gamma) = 1 \pm 0.005$ (Shapiro *et al.* 1977), a factor of 4 improvement over the results of the Mariner-9 (relativity) experiment.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

When we did a more formal analysis based on 14 months of observation, we obtained a much more accurate result. By then, we had developed a systematic procedure for correcting the effect of the plasma. All aspects of the analysis were under computer control, with no potentially prejudicial hand intervention. Most of the residuals are less than 100 ns; they get worse near superior conjunction as is to be expected. From this analysis we have determined that the data are consistent with relativity: $\frac{1}{2}(1+\gamma) = 1.000 \pm 0.001$ (Reasenberg *et al.* 1979). In terms of the Brans–Dicke theory, this result implies that ω is greater than 500. A histogram of the distribution of the residuals shows that, as is often the case with real data, the distribution resembles a Gaussian except that the tails are too high. This and other evidence of systematic error was taken into consideration in determining the uncertainty.

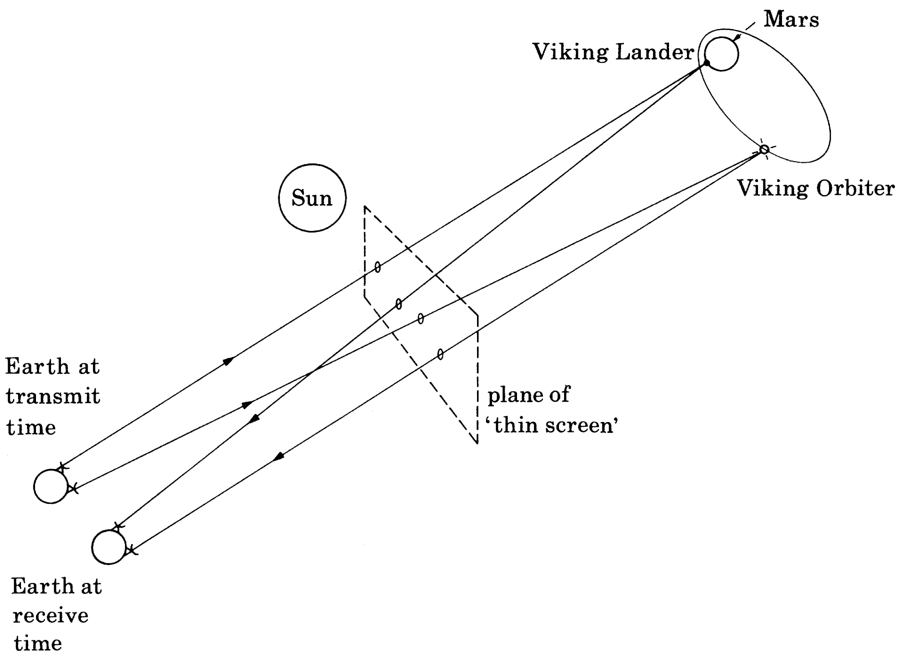

FIGURE 1. The observing geometry, not to scale.

We are not yet finished with the analysis of this experiment. What are the prognoses for further improvement? There have been two superior conjunctions for which the equipment was working. The above results are based on the first of those. We had four objects, all of which have now failed after functioning far past the end of their design life. The second Orbiter failed on 5 August 1980, after which there was no way of measuring the contribution to the delay from the plasma between Earth and Mars. Lander One, which was the last Viking to fail, was working until late 1982; during the last several months, useful data were being obtained about twice per month. Although there is no way to correct these data for the effect of the plasma, far from superior conjunction the plasma effects are small and relatively stable; these data are quite useful. They are directly applicable to improving the planetary ephemerides and thus indirectly useful for the relativity tests.

We have done some numerical experiments to determine what sort of accuracies we ought to expect from a complete analysis of all of the data now available. These suggest that we can decrease the uncertainty in the estimate of $\frac{1}{2}(1+\gamma)$ by a factor of two or three below our best published results. It will not be until we finish the present series of studies, that we will be able to make a new and reliable determination of the time-delay coefficient.

5. POINTS

The results of all accepted experiments are consistent with the predictions of general relativity. However, with the exception of the binary pulsar study (Taylor & Weisberg 1982), all the experiments depend on first-order post-Newtonian effects. One must wonder whether it is particularly fruitful at this time to invest in further improvement of first-order tests. Does it seem reasonable that general relativity fails deep in first order or should we be looking elsewhere? I believe that it is no longer reasonable to make substantial investments in small improvements to first-order tests of the kinds already performed. New kinds of first-order tests, such as those based on motion of the source (e.g. the Stanford gyroscope experiment), remain interesting.

How could one perform a second-order test of relativity? The problem is difficult because the solar potential at the limb of the Sun is 2×10^{-6}. A second-order experiment must be at least three orders of magnitude more sensitive than the present solar-system tests: a considerable challenge.

A second-order test, if it is to be done within the solar system, will likely utilize the potential close to the limb of the Sun. Putting an instrument package close to the Sun causes substantial engineering problems. Such a mission, STARPROBE (Reasenberg *et al.* 1982), is being considered by NASA but is now on hold and its future is uncertain. Thus we must find a way of doing an experiment in which we use photons that nearly graze the Sun. The two reasonable possibilities, the light deflexion experiment and the time-delay experiment, depend on essentially the same phenomenon. For the foreseeable future, the time-delay experiment ineluctably involves spacecraft tracking and the NASA deep space network which, for a combination of technical and political reasons, seems unlikely in the foreseeable future to be able to yield data with the accuracy necessary to measure the second-order time delay. Thus we are limited to considering a light deflexion experiment.

At the solar limb, the second-order light deflexion is 10.9 μs (Epstein & Shapiro 1980; Fischbach & Freeman 1980; Richter & Matzner 1983), five orders smaller than the 1.75 s first-order deflexion. The plasma-related problems, which are a limiting factor in first-order experiments using microwave signals, demand the use of substantially higher frequency signals for a second-order test. A reasonable solution is to use signals in the optical region. Since the resolution of any angle-determining instrument is closely related to the ratio of the wavelength to the size of the instrument, the optical region offers an additional advantage. Thus we are led to the possibility of using optical interferometry as a means of determining the light deflexion to second order in the solar potential. Any device that would perform such an experiment would have to be above Earth's atmosphere because of the large corrupting effect of atmospheric refraction.

Figure 2 shows a conceptual design of an instrument to perform the second-order deflexion experiment by means of Precision Optical INTerferometry in Space (POINTS). It is a dual optical interferometer with nominal baselines of 10 m. There are four one-metre telescopes that bring together light at the two central sections, where fringes are formed and analysed. A beam splitter, lens, and prism yield a spectrum of the interference (i.e. a channelled spectrum) which is allowed to fall on a linear array of photon-counting detectors. The complementary channelled spectra from the two exit ports of the beam splitter contain all the information necessary to determine the displacement of the source from the optical axis of the interferometer. As the principal axis of such an instrument is moved away from the source, the number of channels in the spectrum increases. The ability of the linear detector array to separate the undulations in the channelled spectrum sets the limit to the allowed mispointing of the instrument. In the nominal design the linear array has 1000 detectors; the pointing error limit is $\pm 2.5''$.

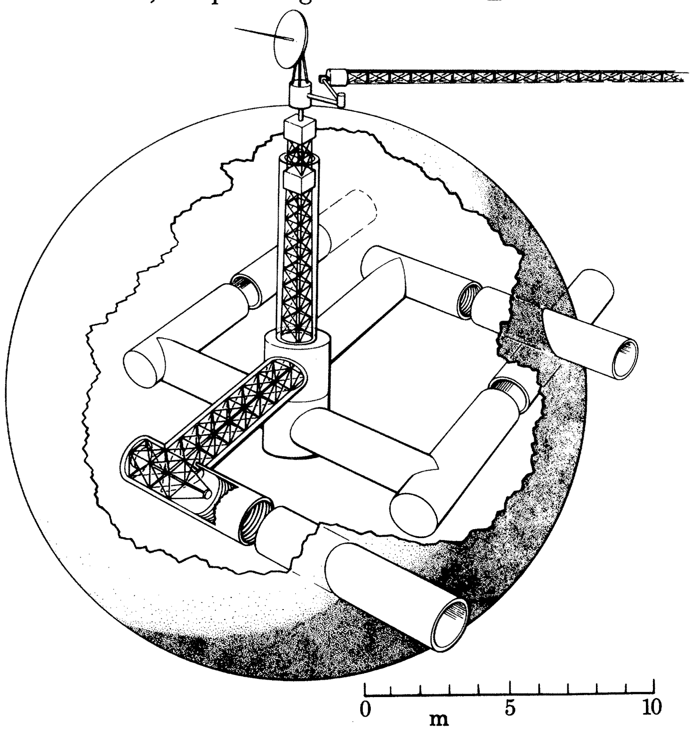

FIGURE 2. An artist's rendition of the proposed astrometric interferometer, POINTS. The satellite diameter is 20 m; each of the two interferometers has a pair of 1 m telescopes at the ends of a 10 m baseline. This design is the result of a collaboration with the C. S. Draper Laboratory.

A simple photon-statistics analysis of the system indicates that a single interferometer observing a 10th magnitude source at 7000 K has a measurement uncertainty of 1 μs with an integration of time of 150 s. Thus when POINTS is used to observe bright stars, it is a fast measuring device. There would be time to conduct the relativity experiment along with several other kinds of research: looking for planetary systems, determining the lower rungs of the cosmic distance ladder by measuring the distance to Cepheid variables, and numerous other astrophysically interesting studies. The diversity of application is important, not only to the astrophysicist directly, but also indirectly. It is the speed of measurement and the potential diversity of application that will make possible a large constituency from the astronomical community to back the development of POINTS.

The spherical enclosure is intended to reduce severe thermal problems such an instrument would experience if exposed directly to the Sun. Several of the technical aspects of this cartoon are now obsolete, as we have come to understand the instrument better, but the concept is still considered valid. The two interferometers shown are at right angles. This configuration maximizes the number of reference stars available for a given target star and simplifies the measurement of the angle between the optical axes of the two interferometers.

To achieve the required accuracy with POINTS, the locations of the optical components must be known with an uncertainty of no more than about 0.1 Å,† 10^{-12} of the baseline length of 10 m. This cannot be done entirely by the use of stable materials. In principle, the required measurements can be performed with onboard laser metrology. However, consider what it means to know the location of an extended object like a telescope mirror to 0.1 Å. Such an object is a chunk of glass, quartz, or some composite material which has a non-zero thermal expansion coefficient, distorts under mechanical stress, and may creep. Optical components will not only change temperature, but they will have varying temperature distributions and the corresponding distortions. Thus, it will not be sufficient to keep track of the locations of a few reference points on a large optical element; the entire surface of the element must be monitored. Fortunately, we know in principle how to construct a system to provide the required 'full aperture' metrology.

The problem now is one of detailed engineering and forming a constituency. It seems unlikely that we can start construction of POINTS during this decade. Perhaps some advanced design work can be started by the end of the decade, and the device can be constructed in the next decade for launch either in the late 1990s or the early 21st century.

6. Conclusion

There is now a small, international industry dedicated to testing relativistic theories of gravitation. In the decades since the emergence of general relativity, and particularly in the last 20 years, considerable experimental progress has been made. There are several successful first-order tests and the strong evidence from PSR 1916+13, all supporting the validity of general relativity. We appear now to be on the threshold of a new era. It is time to investigate additional phenomena including the Lense–Thirring effect, second-order effects, and the interaction of gravitation with quantum mechanical systems. These are among the experimental challenges for the coming decades.

I thank N. L. Murphy for her careful preparation of the text. I am grateful to I. I. Shapiro for his comments. This work was supported in part by the Smithsonian Astrophysical Observatory, the Secretary's Fluid Research Fund of the Smithsonian Institution, the National Aeronautics and Space Administration (NASA) through the Mars Data Analysis Program Contract NSG-7556, and the National Science Foundation through grant PHY-8106036.

References

Canuto, V., Adams, P. J., Hsieh, S.-H. & Tsiang, E. 1977 *Phys. Rev.* D **16**, 1643–1663.
Canuto, V. M., Goldman, I. & Shapiro, I. I. Testing the strong equivalence principle by radio ranging. (In preparation.)
Counselman, C. C., Kent, S. M., Knight, C. A., Shapiro, I. I., Clark, T. A., Hinteregger, H. F., Rogers, A. E. E. & Whitney A. R. 1974 *Phys. Rev. Lett.* **33**, 1621.
Counselman, C. C. & Shapiro, I. I. 1968 *Science, N.Y.* **162**, 352–355.
Dirac, P. A. M. 1937 *Nature, Lond.* **139**, 323.
Dirac, P. A. M. 1938 *Proc. R. Soc. Lond.* A **165**, 199.
Dirac, P. A. M. 1974 *Proc. R. Soc. Lond.* A **338**, 439–446.
Dirac, P. A. M. 1979 *Proc. R. Soc. Lond.* A **365**, 19–30.
Dyson, F. J. 1972 In *Aspects of quantum theory* (ed. A. Salam & E. P. Wigner), pp. 213–236. Cambridge University Press.

† 1 Å = 10^{-10} m.

Epstein, R. & Shapiro, I. I. 1980 *Phys. Rev.* D **22**, 2947–2949.
Fischbach, E. & Freeman, B. S. 1980 *Phys. Rev.* D **22**, 2950–2952.
Fomalont, E. B. & Sramek, R. A. 1977 *Comments Astrophys.* **7**, 19.
Reasenberg, R. D. 1980 In *Cosmology and Gravitation* (ed. P. G. Bergmann and V. DeSabbata). New York and London: Plenum Press.
Reasenberg, R. D. 1983 In *Proceedings of the Third Marcel Grossmann Meeting on the Recent Developments of General Relativity* (In the press.)
Reasenberg, R. D., Anderson, J. D., DeBra, D. B., Shapiro, I. I., Ulrich, R. K. & Vessot, R. F. C. 1982 In *STARPROBE scientific rationale, a report of the ad hoc working groups* (ed. J. H. Underwood & J. E. Randolph). JPL Publication 82–49, Pasadena.
Reasenberg, R. D. & Shapiro, I. I. 1976 In *Atomic masses and fundamental constants*, vol. 5 (ed. J. H. Sanders & A. H. Wapstra), pp. 643–646. New York and London: Plenum Press.
Reasenberg, R. D., Shapiro, I. I., MacNeil, P. E., Goldstein, R. B., Breidenthal, J. C., Brenkle, J. P., Cain, D. L., Kaufman, T. M., Komarek, T. A. & Zygielbaum, A. I. 1979 *Astrophys. J. Lett.* **234**, 219.
Richter, G. W. & Matzner, R. A. 1983 *Phys. Rev.* D. (In the press.)
Shapiro, I. I. 1964*a* *Effects of general relativity on interplanetary time-delay measurements.* M.I.T. Lincoln Laboratory technical report 368, Lexington, Massachusetts.
Shapiro, I. I. 1964*b* *Phys. Rev. Lett.* **13**, 789.
Shapiro, I. I., Counselman, C. C. & King, R. W. 1976 *Phys. Rev. Lett.* **36**, 555.
Shapiro, I. I., Reasenberg, R. D., MacNeil, P. E., Goldstein, R. B., Brenkle, J. P., Cain, D. L., Komarek, T., Zygielbaum, A. I., Cuddihy, W. F. & Michael, Jr, W. H. 1977 *J. geophys. Res.* **82**, 4329.
Shapiro, I. I., Smith, W. B., Ash, M. E., Ingalls, R. P. & Pettengill, G. H. 1971 *Phys. Rev. Lett.* **26**, 27–30.
Taylor, J. H. & Weisberg, J. M. 1982 *Astrophys. J.* **253**, 908.
Texas Mauritanian Eclipse Team 1976 *Astr. J.* **81**, 452.
Vessot, R. F. C., Levine, M. W., Mattison, E. M., Blomberg, E. L., Hoffman, T. E., Nystrom, G. U., Farrel, B. F., Decher, R., Eby, P. B., Baugher, C. R., Watts, J. W., Teuber, D. L. & Wills, F. D. 1980 *Phys. Rev. Lett.* **45**, 2081.
Will, C. M. 1981 *Theory and experiment in gravitational physics.* Cambridge: Cambridge University Press.