ON THE WEIGHT OF PHOTONS

R. V. POUND

Usp. Fiz. Nauk 72, 673-683 (December, 1960)

EINSTEIN[1] showed in 1911 that his "principle of equivalence" predicts that electromagnetic radiation should gain energy in falling through a gravitational potential difference. The observations from Galileo onward of the proportionality of mass and weight for material bodies could be taken as evidence of the validity of the principle of equivalence. The principle was invoked as a step in bringing the highly successful special theory of relativity into contact with gravitation, leading ultimately to the so-called general theory of relativity.

In the past fifty years the direct effect of gravity on electromagnetic energy has defied definitive observation. There have been astronomical measurements of the so-called "gravitational red-shift" but these have been rendered relatively inconclusive by the presence of other effects not fully understood. For example,[2] the shifts determined by observation of spectral lines of the sun depend strongly on the distance from the center of the solar disk of the point under observation, the shift rising asymptotically to near the predicted value in the limb. The physical conditions that account for such dependence must be established unequivocally before the observations could be said to support the prediction of a gravitational shift or not. The inclusion of radioactive material among these for which the principle of equivalence has been shown to hold to a high degree of precision can be argued to leave little room for doubt of the validity of the effect on radiation itself but such indirect proofs naturally overlook the unknown. The intimate connection between the theories of relativity and electromagnetism renders the challenge to find a direct test especially strong.

Einstein's simple computation was based on substituting for a gravitational field in a region, as prescribed by the principle of equivalence, a gravity free coordinate system under constant acceleration g upward, where g is the downward acceleration due to gravity in the actual space. The time of flight of radiation from a source vertically a distance h above the observer is h/c and in the equivalent coordinate system, if gh/c is small compared to c, a fractional increase of frequency $\Delta\nu/\nu_0$ equal to gh/c^2 results from the first order Doppler effect. The observer sees a frequency $\nu = \nu_0(1 + gh/c^2)$ as the consequence of the change of relative speed during the time of flight. For non-uniform gravitational fields one would obtain, similarly, $\nu = \nu_0(1 + \Delta\varphi/c^2)$ where $\Delta\varphi$ is the difference of gravitational potential. An entirely equivalent view, derived by Einstein at the same time, regards time at the two points as proceeding at rates differing by gh/c^2. The translation from the one to the other picture invokes only the well proven transformation theory of special relativity.

Numerically the effect is so small as to have been far beyond hope of detection in wholly terrestrial experiments until very recently, being fractionally about 10^{-16} per meter of height at the surface of the earth. As a consequence, experiments have been planned to compare the time keeping of satellite born atomic clocks with similar ones on the ground, with the effect then being at about the limit of presently achievable stability. A wholly terrestrial experiment offers far greater accessibility to allow changes, such as inversion whereby the effect should be reversed, which can thereby circumvent pitfalls similar to those encountered astronomically.

About two years ago R. T. Mössbauer published[3] results of novel experiments on nuclear resonance fluorescence in a cooled solid. He called attention to the fact that the binding of a nucleus to a lattice very significantly affects the dynamics and, therefore, the spectrum of low energy γ-ray emission and absorption. In effect, the binding results in two rather different things. First, the nucleus is prevented from recoiling freely during emission or absorption but, instead, it partially carries its neighbors with it. Analysis leads to the conclusion that there is a certain probability that the momentum be absorbed by the relatively massive entire lattice and that there be negligible energy lost to the recoiling mass. Thus, contrary to the situation for free nuclei, an emitting and an absorbing nucleus in appropriate states, can resonate with one another even for very narrow linewidths. The second function performed by the lattice is the great reduction of the broadening that might be expected to result from thermal motions. This comes about because the lattice frequencies, for instance in a Debye spectrum, are mostly so high that over the time of transition the velocities, and the resulting Doppler broadening, averages nearly to zero. Classically one may visualize a frequency modulation that preserves some energy in the Fourier component corresponding to the monochromatic carrier. The two effects lead to a component in the γ-ray spectrum undisplaced by recoil and unbroadened beyond the width set by lifetime. The fractional intensity of this component is quantitatively described by a factor that has come to be designated as f closely resembling the

Debye-Waller factor for coherent X-ray scattering. In a simple lattice described by a Debye spectrum of temperature θ_D, f is approximately

$$f \approx \exp\left\{-\frac{3}{2}\left(E^2_\gamma/2Mc^2k\theta_D\right)\left[1+\frac{2}{3}\left(\frac{\pi T}{\theta_D}\right)^2\right]\right\}$$

for temperatures T much less than θ_D. Similarly, a bound nucleus in the ground state has a total peak resonance scattering cross section σ equal to $f\sigma_0$ where σ_0 is the resonant electromagnetic cross section,

$$\sigma_0 = 2\pi\lambda_e^2\,(2I_e+1)/(2I_g+1)\,(1+\alpha),$$

where I_e and I_g are the spins of the excited and ground states respectively and α is the internal conversion coefficient. For scattering γ-rays of width matched to the resonance, with no hyperfine structure splittings, a net cross section equal to one-half the peak value for $f\sigma_0$ results. If there are equal hyperfine splittings in the source and the scatterer, the statistical weight factors are altered and σ is reduced.

After first observing such scattering for Ir^{191} through the decrease in transmission where source and absorber were cooled, Mössbauer introduced a motion of the source toward or away from the absorber to produce a Doppler shift that could destroy the resonance. The extraordinary sharpness of the resonance is made graphic by the fact that a relative velocity of one centimeter per second reduced the absorption appreciably. In effect, the lifetime lead to absolute width comparable with the smallest obtainable in optical spectra but the carrier frequency was about 10^5 times higher. At last a way had been found to use a solid to limit thermal broadening, without suffering severe broadening by solid state interactions. The weakness of the coupling of the nucleus to external fields plays its role here.

When my associate, G. A. Rebka, Jr. and I became aware of this new development of Mössbauer, less than a year ago, we set out to see if it could be extended, by finding other examples, far enough to make possible a terrestrial measurement of the gravitational red shift.[4] Other groups[5,6] independently proposed such an experiment and all appear to have fixed upon the same two examples among the multitude of nuclear decay schemes so far tabulated. These two are the 14.4-kev γ ray of 0.1-μsec Fe^{57} and the 93-kev γ ray of 9.4-μsec Zn^{67}. Both are emitted as the final step in the decay to a stable isotope of an electron capturing parent, 270-day Co^{57} and 78-hour Ca^{67} respectively.

The experiment visualized the detection of the displacement from resonance of a source and absorber at different heights and it is tempting to describe the usefulness of a line by specifying the height that should result in a displacement to half the peak scattering. For Fe^{57} this should be 3 km and for Zn^{67} it should be about 5 meters. There are, however, other factors involved. For example, the low γ-ray energy and the high Debye temperature of metallic iron results for Fe^{57} in the possibility of 0.8 for f even at room temperature, whereas temperatures below about 50° K would required to obtain an f value approaching 0.01 with Zn^{67} in zinc. A disadvantage of Fe^{57} is the large conversion, α equal to about 15, which increases the problems of background from the higher energy γ rays in the spectrum relative to that desired.

Our initial experiments on Fe^{57} yielded a very encouraging result.[7] The source was prepared by electroplating Co^{57} onto an iron disk and annealing for about an hour at 950° C. In this way Co^{57} was diffused a short distance into the iron to give the Fe^{57} a homogeneous iron environment. When the 14.4-kev γ ray was observed with a scintillation spectrometer through an iron foil 0.0015 cm thick, it was found at room temperature that vibration of the source resulted in about 20 percent increase in transmission. The line shape was mapped by moving the source with a magnetic moving-coil transducer driven by a triangular current wave and the result is reproduced in Fig. 1. The Lorentz curve drawn, of half-width 0.017 cm/sec, is to be compared to a theoretical width expected of 0.01 cm/sec. Other laboratories[8] have since obtained widths apparently as small as theoretical, using modified but similar techniques.

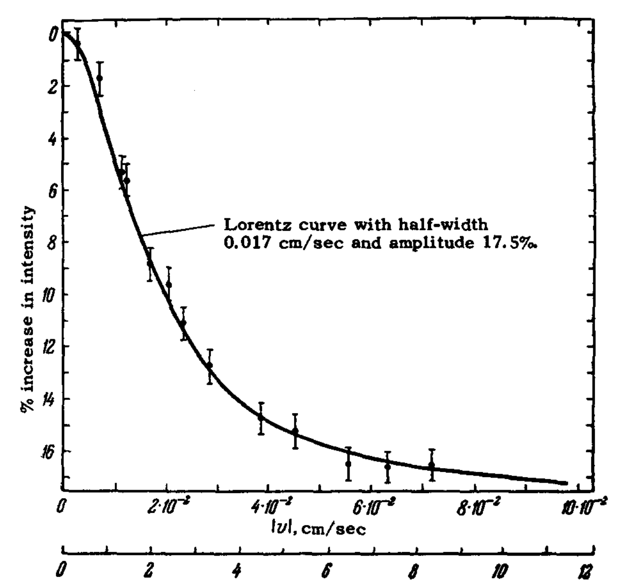

FIG. 1. The resonance absorption vs. source velocity for Fe^{57}. The absorber is about 0.0015 cm thick and is natural iron.

In a limited way, we sought hyperfine structure splitting. The excited level was known to have spin I_e equal to 3/2 and the ground state I_g equal to 1/2. In ferromagnetic iron a very large effective magnetic field would be expected and therefore both levels should be split. The magnetic dipole γ-ray would then have six components of different energies but, with identical fields in source and scatterer or absorber, the lines would be matched one to one. All would be similarly affected by relative motion and the line width could be unchanged, although the absorption would be weakened by the hyperfine splitting. At larger speeds, steady relative motion may result in other partial matching of emission and absorption lines. We explored out to 0.8 cm/sec and found three subsidiary resonances as shown in Fig. 2. In all, if the structure is fully resolved, there should be seven satellites on each side. Others[9] have unravelled the hyperfine spectrum completely, finding two more lines similar to ours but further out than we could reach in our initial run. The remaining two weak lines have been shown to be included, almost unresolved, in these five. Detailed fitting to theory has been aided by the use of transverse polarizing fields on the iron which shows large changes of satellite intensity depending on whether the source and absorber are polarized parallel or perpendicular to one another. The internal field has been found to be 330 kilogauss and the sign and magnetic moment of the excited level evaluated.

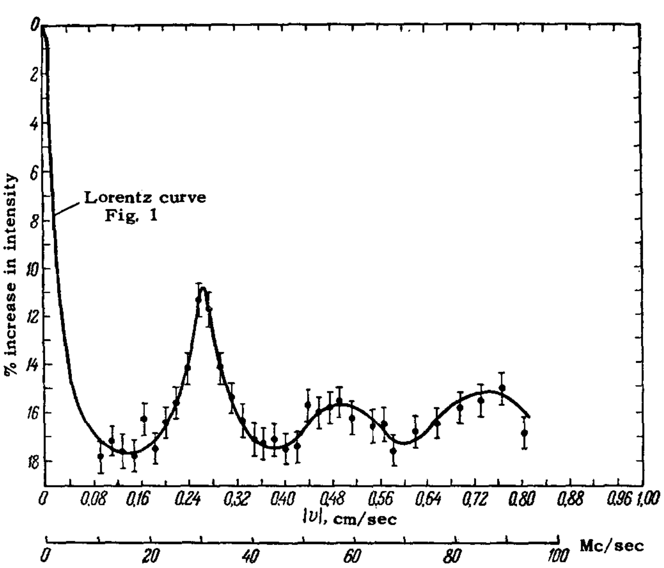

FIG. 2. The inner part of the hyperfine structure spectrum of Fe^{57} vs. Doppler velocity.

Many laboratories have joined the exploration and similar data are being reported for other ferromagnetic and paramagnetic environments.[10,11,12] In addition to internal fields, electric quadrupole splitting and a dependence of energy on chemical structure are being found by these experiments. The existence of the hyperfine structure and its dependence on the environment dictates the care needed in preparation of sources and absorbers to obtain as narrow a line as possible for our application.

Our preliminary experiments[13] with Zn^{67}, mostly conducted at 4° K, failed to find a resonant absorption larger than about 0.1 percent. We tried a number of lattices but have reason to think that small frequency shifts, owing to differences between source and absorbers in such properties as electronic structure, Debye temperature and lattice symmetry, prevented the absorption from occurring at zero relative speed. A group at Los Alamos has recently reported[14] a very small effect in sintered ZnO, but, as yet, the actual line width is not known because they have relied upon a Zeeman effect to destroy the resonance.

In spite of the breadth of the Fe^{57} line, its large intensity, the long parent lifetime and the convenience of operating at room temperature encouraged us to proceed to use it to try to measure the gravitational effect. The description of its width as requiring 3 to 5 km of height to reduce the absorption to one-half is misleading. The important factor is actually the change of the absolute absorption for a given fractional frequency shift of the source relative to the absorber. To obtain highest sensitivity to a frequency shift that is a small part of the line width, one may first introduce motion to displace, by a Doppler effect, the line of the source relative to the absorber to the inflection point where the maximum rate of change of transmission with differential frequency shift occurs. With a perfectly symmetrical experimental line, it is convenient to use alternate displacements in opposite directions produced by oscillating the source upward and downward. The transmission during upward periods and downward ones may then be compared and a small inherent shift is then shown by a difference of the two transmissions. In the event that the source strength is limited and, therefore, that statistical fluctuations dominate the uncertainty of a measurement, it becomes evident that the detectability of a shift that is proportional to the separation h is independent of h so long as the counting rate is proportional to h^{-2}. On the other hand, the possible presence of systematic errors, like the dependence of the frequency on other uncontrolled variables, dictates the choice of an adequate height.

Our experiment has been conducted in an enclosed tower that was a feature of the architectural design of the Jefferson Physical Laboratory at Harvard, built about eighty years ago. It has a separate foundation and a bearing wall separate from the surrounding structure runs from basement level to the roof of the building. The height h used is about 21 meters. By its nature this site provides an admirable isolation from vibration and thermal expansion. An important advantage in the use of a broad line to determine the shift as compared to a narrower one is that vibrations, so long as they do not involve velocities beyond the linear range of the slope, or in other words so long as they are small compared to the line width, average out of the long term data. The small static shifts can be measured even in the presence of larger vibrations. To avoid attenuation by air in the path we have employed a 40 cm. diameter cylindrical plastic bag inflated with helium gas at atmospheric pressure. A slow flow of the helium gas was maintained to sweep out air that diffused into the bag.

The overall system is described by the block diagram of Fig. 3. The upward and downward oscillation of the source was provided by a sinusoidal voltage at about 50 cps applied to a ferroelectric ceramic cylinder on which was mounted the 5 cm diameter iron foil source. This contained about 0.4 curies of carrier free Co^{57}, diffused into the iron in the same manner as for the earlier tests. A detector of large area was obtained by the use of seven separate 7 mm thick NaI scintillation crystals mounted on multiplier phototubes of similar diameter. The absorbers were made by electroplating iron enriched in Fe^{57} to about 32 percent onto seven 1-mm-thick beryllium disks of the same diameter as scintillators. In the later phases of the experiment these disks were mounted over holes in a thick brass sheet and a set of ceramic permanent magnets provided a magnetic field of about 200 oersteds more or less parallel and in the plane of the disks. Similar magnets were applied at the source. The purpose of the magnets was to polarize the source and absorbers parallel to one another with the aim of suppressing some hyperfine lines and increasing the absorption since the absorbers were thin. Another purpose was to reduce any influence of changes of orientation of the system relative to the magnetic field of the earth.

The 50-cycle source was caused to trigger a gate, of duration 5 μsec timed symmetrically about the point of maximum velocity of the transducer. During the dead time between gate pulses switches were thrown by a square-wave generated from the 50-cps source to connect the amplified and discriminated pulses to one or another scaler. In this way one scaler collected data during upward and the other during downward motion of the transducer.

The difference of these counts would depend on the basic asymmetry of the line and the modulation and on the stability of the system. To monitor the stability and to provide a means of eliminating such effects from the data, a duplicate channel of amplifier, discriminator, gate and scalars was provided. This was driven by a scintillation counter with its own small, but similar absorber mounted only about one meter from the source. Normally the counting rate in this channel was about six times that in the main channel so that a difference between the fractional differences observed by it and by the main channel could be taken without much increase in statistical uncertainty as compared to that of the main channel.

The connection between an observed difference in counts and a frequency shift such as was being sought depends on the modulation swing, the line shape, width and intensity, the background correction which in turn depends on amplifier gain and the pulse height analyzer, the helium purity and perhaps other parameters. Although these quantities could be measured the more reliable and direct approach was adopted of using a Doppler effect calibration. For this, the

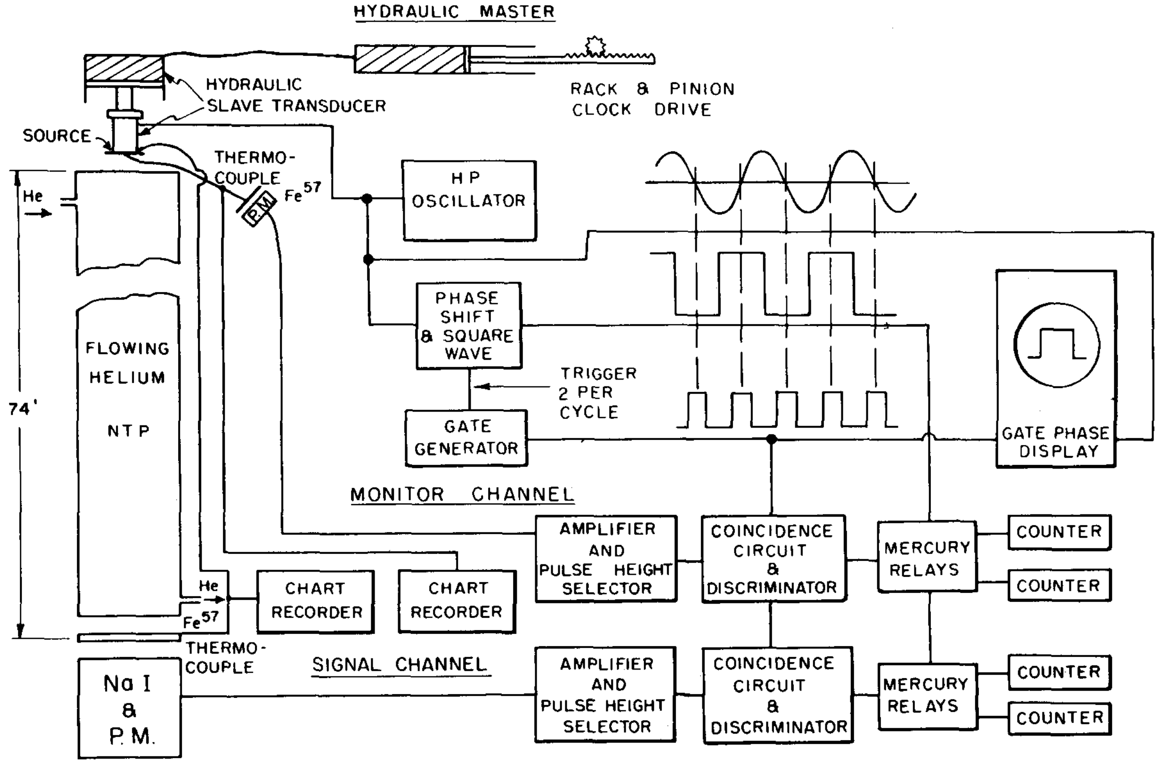

FIG. 3. A block diagram of the arrangement set up in an enclosed tower at Harvard University to measure the gravitational rod-shift.

transducer providing the sinusoidal modulation was itself mounted on a hydraulic piston which, via a piston of smaller bore and a geared down synchronous motor rack and pinion drive, provided a steady motion at about 6×10^{-4} cm/sec, upward or downward. Assuming the apparatus to be stable over periods of upward and of downward steady motion the two sets of data can be combined to yield the sensitivity and the asymmetry that would have been present without the calibrating motion. In early runs the calibration device was operated about one third of the time. Later an automatic timer was added which reversed the direction of motion of the calibrating piston after 300 secs. and switched both channels to a duplicate set of scalers. In this way data was collected with continuous calibration. A frequency displacement averaged over a run of any duration, but usually of about 24 hours, could be calculated directly from the data provided, with no information carried over from other runs. The consistency of sensitivity data provided a useful check on the operation of the system. A run of twenty-four hours measured the apparent fractional frequency shift with an r.m.s. statistical error less than 10^{-15}. The fractional shift expected due to gravity in the tower was about 2.5×10^{-15}.

It was planned from the beginning that the experiment would be performed by comparing the apparent displacement of source and absorber frequencies seen with the source at the top to a similar quantity with the source at the bottom. The effect sought, in the 21 meter height, was only about 2.5×10^{-15} and we did not suppose that the line or instrumental symmetry could be relied upon to the scale of about 2×10^{-3} times the line width, as would be necessary in a one way experiment. The possibility of such a reversal is a principle attraction of the terrestrial experiment, as mentioned earlier. Although a similar quantity can be attained by comparing data at different heights, such as was provided by the monitor and main channels, it is not possible to make them identical geometrically and it is not, therefore, a reliable technique.

Our initial attempts to obtain meaningful results with the apparatus were discouraging. They showed quite severe fluctuations of the apparent displacement of the frequency of the source from that of the absorber which even seemed at times to be diurnal in character. We tried to find malfunctions in the apparatus for some time before we realized, with the help of results of some tests made for other purposes, that we were actually observing frequency shifts caused by changes of the temperature difference between the source and the absorber. Although the average value of the velocity of a bound nucleus vanishes over long times, and so must the linear Doppler shift, the average of the square of the velocity measures the zero point plus the thermal free energy and its variation with temperature is directly related to the lattice specific heat C_L. Over the times of emission or absorption, 10^{-7} sec. in the case of Fe^{57}, the average is well defined, and little broadening of the line would be caused because the spectrum of lattice oscillations is broad and mainly based at frequencies much higher than 10^7 cps. On the other hand the non-zero value of v^2 results in a frequency decrease, as observed in the laboratory, fractionally by $v^2/2c^2$ owing to the time dilation or the relativistic Doppler effect. Thus, the frequency of emission or absorption would vary with temperature as

$$\frac{\partial \nu}{\partial T} = -\frac{\nu_0 C_L}{2Mc^2},$$

where M is the nuclear mass. In the classical limit, where $C_L = 3R$, one would find a decrease of apparent frequency of $2.44 \times 10^{-15}\ \nu_0$ per degree Centigrade. Using a Debye function for C_L one obtains at 300° K about $2.20 \times 10^{-15}\ \nu_0$ per degree or one degree should produce a shift just about equal to the gravitational effect sought in our experiment.

This proposition was tested directly[15] by varying the relative temperatures of the source and absorber and the "red-shift" apparatus was used to measure the frequency shift. The fit to the prediction based on a Debye curve was found to be quite satisfactory. (See Fig. 12 of Mössbauer's article p. 866.)

Following this discovery, the gravitational experiment was modified by the addition of a thermocouple, with one junction at the source and the other embedded in the brass plate carrying the absorbers, connected to a recording microvoltmeter. A similar unit was added to the monitor channel. At once the data became stable and consistent when corrections were added to include frequency shifts caused by the temperature difference averaged over a run.[16] From that point onward useful data has been collected and, except for brief intervals during which modifications were introduced, or checks were being made, the apparatus has been operated essentially continuously.

The data has demonstrated very conclusively the importance of looking only to the difference between rising and falling γ-rays to find the gravitational effect. A displacement of about $-15 \times 10^{-15}\ \nu_0$, about six times the shift sought for a one-way experiment, was found to exist between the source and the absorbers. The seven iron coated beryllium disks were tested separately and found to be different, ranging from -8 to -30×10^{-15}, whereas the one used on the monitor channel showed about $+0.5 \times 10^{-15}$ relative shift. In specifying these shifts a negative sign is used to indicate that the source frequency is lower than the frequency of maximum absorption. The origin of these shifts is still uncertain but it must be remembered that even the largest is only 3×10^{-2} times the line width. The line width itself is probably about $3/2$ times its natural width owing to imperfection of the source or absorbers, or both, and it would hardly be surprising if such imperfection introduced an effective shift or asymmetry in the line. Differences of Debye temperature, through differences in zero point amplitudes and thermal motions, could also produce such effects. The important point at the moment is that over the four months of operation, the shifts have appeared to be constant and, therefore, to have permitted the extraction of the gravitational effect from the data.

Altogether so far about six separate experiments have been carried out, each of two or three weeks duration. The operation is viewed in this way because certain modifications have been introduced between such periods of running. Each such period produced a result with an uncertainty of about 0.1 of the effect predicted and each agreed with the value 4.67×10^{-15} for the fractional difference for the two directions well within that statistical uncertainty. The average of all runs to date yields

$$0.98 \pm 0.04$$

in units of the predicted value of gh/c^2 when the data of the main channel is used alone. A value of

$$0.97 \pm 0.05$$

is found when the shift is taken relative to that of the monitor channel. The difference between the two numbers is not statistically significant, indicating rather good stability in the system. We believe our control over sources of systematic errors places them below the r.m.s. statistical errors quoted. The result demonstrates that the predicted "red-shift" exists and is quantitatively in agreement with expectation at least within about 4 percent. In effect the effective mass and weight of radiation have the same proportionality as do those measures of material bodies.

It is our intention to continue to reduce the uncertainty of the measurement by collecting more data, by improving the resolution, and eventually, by setting it up on a larger scale with a longer path, where systematic errors would contribute less strongly. Thereby we may be able to reduce the uncertainty well below a percent.

[1] A. Einstein, Ann. Physik **35**, 898 (1911).
[2] E. Finlay-Freundlich, Phil. Mag. **45**, 303 (1954).
[3] R. T. Mössbauer, Z. Physik **151**, 124 (1958); Naturwissenschaften **45**, 538 (1958); Z. Naturforsch. **14a**, 211 (1959).
[4] R. V. Pound and G. A. Rebka, Jr., Phys. Rev. Letters **3**, 439 (1959).
[5] I. Ya. Barit, M. I. Podgoretskiĭ, and F. L. Shapiro, JETP **38**, 301 (1960), Soviet Phys. JETP **11**, 218 (1960).
[6] J. P. Schiffer and W. Marshall, Phys. Rev. Letters **3**, 556 (1959).
[7] R. V. Pound and G. A. Rebka, Jr., Phys. Rev. Letters **3**, 554 (1959).
[8] M. A. Clark and J. M. Robson, Bull. Am. Phys. Soc. **5**, 369 (1960).
[9] S. S. Hanna, J. Heberle, C. Littlejohn, G. J. Perlow, R. S. Preston, and D. H. Vincent, Phys. Rev. Letters **4**, 177 (1960).
[10] G. K. Wertheim, Phys. Rev. Letters **4**, 403 (1960).
[11] O. C. Kistner and A. W. Sunyar, Phys. Rev. Letters **4**, 412 (1960).
[12] I. Solomon, C. R. Acad. Sci. (Paris) **250**, 3828 (1960).
[13] R. V. Pound and G. A. Rebka, Jr., Phys. Rev. Letters **4**, 397 (1960).
[14] D. E. Nagle, P. P. Craig, and W. E. Keller, Nature **186**, 707 (1960).
[15] R. V. Pound and G. A. Rebka, Jr., Phys. Rev. Letters **4**, 274 (1960).
[16] R. V. Pound and G. A. Rebka, Jr., Phys. Rev. Letters **4**, 337 (1960).