Relevance of the weak equivalence principle and experiments to test it: Lessons from the past and improvements expected in space

Anna M. Nobili [a],*, Alberto Anselmi [b]

[a] *Dept. of Physics "E. Fermi", University of Pisa and INFN, Largo B. Pontecorvo 3, 56127 Pisa, Italy*
[b] *Thales Alenia Space Italia, Strada Antica di Collegno 253, 10146 Torino, Italy*

A R T I C L E   I N F O

A B S T R A C T

Tests of the Weak Equivalence Principle (WEP) probe the foundations of physics. Ever since Galileo in the early 1600s, WEP tests have attracted some of the best experimentalists of any time. Progress has come in bursts, each stimulated by the introduction of a new technique: the torsion balance, signal modulation by Earth rotation, the rotating torsion balance. Tests for various materials in the field of the Earth and the Sun have found no violation to the level of about 1 part in 10^{13}. A different technique, Lunar Laser Ranging (LLR), has reached comparable precision. Today, both laboratory tests and LLR have reached a point when improving by a factor of 10 is extremely hard. The promise of another quantum leap in precision rests on experiments performed in low Earth orbit. The Microscope satellite, launched in April 2016 and currently taking data, aims to test WEP in the field of Earth to 10^{-15}, a 100-fold improvement possible thanks to a driving signal in orbit almost 500 times stronger than for torsion balances on ground. The 'Galileo Galilei' (GG) experiment, by combining the advantages of space with those of the rotating torsion balance, aims at a WEP test 100 times more precise than Microscope, to 10^{-17}. A quantitative comparison of the key issues in the two experiments is presented, along with recent experimental measurements relevant for GG. Early results from Microscope, reported at a conference in March 2017, show measurement performance close to the expectations and confirm the key role of rotation with the advantage (unique to space) of rotating the whole spacecraft. Any non-null result from Microscope would be a major discovery and call for urgent confirmation; with 100 times better precision GG could settle the matter and provide a deeper probe of the foundations of physics.

© 2017 Elsevier B.V. All rights reserved.

1. Introduction

The General theory of Relativity (GR) [1] stands on the fundamental assumption that in a gravitational field all bodies fall with the same acceleration regardless of their mass and composition, a 'fact of nature' known as the Universality of Free Fall (UFF) or the Weak Equivalence Principle (WEP). The WEP has been tested for various materials in the field of the Earth and the Sun, and no violation has been found to the level of about $\Delta a/a \simeq 10^{-13}$ [2] (Δa is the difference in acceleration between two test bodies falling in a gravitational field with mean acceleration a, referred to as the 'driving signal').

A WEP experiment is both a test of the foundation stone of GR and a search for a new long range field coupling to matter in a way that depends on composition. A confirmed violation would have the same significance as the discovery of a new force of nature. There is no firm prediction as to the level at which the violation

should occur. However, the WEP is so fundamental a postulate that any experiment that can push limits by many orders of magnitude is highly significant, whether it finds an effect or not.

Substantial progress in WEP test precision has always depended on the introduction of a new technique: the torsion balance at the turn of the 20th century (Eötvös), the Sun providing a daily modulated signal source (Dicke, Braginsky in the 1960s–70s), the rotating torsion balance (Adelberger and collaborators, from the early 1990s to this date). Nowadays, laboratory experiments have run their gamut and any further progress is small and comes at slow pace. A completely different technique, Lunar Laser Ranging (LLR), tests the WEP for the Earth and Moon as bodies of different composition falling in the field of the Sun. Such tests have reached a precision similar to the torsion balance [3,4] and a 10-fold improvement requires not only mm-level laser ranging, but also a matching improvement of the physical model which describes the Earth-Moon system [5].

Today, only space experiments seem capable of a significant step forward. Just because in orbit the driving signal from Earth is stronger by a factor of almost 500 than it is for the torsion balance

---

on ground, a carefully designed orbiting experiment can target a precision improved by a similar factor.

The Microscope mission, launched in April 2016 into a low altitude, sun-synchronous orbit, aims to test WEP in the field of Earth to 10^{-15} [6], a 100-fold improvement over the rotating torsion balances which can be achieved with a lower sensitivity to differential accelerations thanks to the stronger driving signal in orbit.

The guiding principle of the 'Galileo Galilei' (GG) small satellite mission [7] is to fully exploit the advantages of space as well as those of the rotating torsion balance, so as to design a balance optimized for testing the WEP at zero-g. GG aims to reach 10^{-17}: a four order of magnitude improvement over the best ground experiments and a 100-fold improvement over Microscope. For GG to achieve its target, it must be about 20 times more sensitive to differential accelerations than rotating torsion balances, which is possible by exploiting weightless conditions inside an isolated co-rotating laboratory (the spacecraft) passively stabilized by rapid 1 Hz rotation around the symmetry axis.

Today, experimental evidence pointing the way out of the current physics impasse is hard to find, and even a hint of an effect from Microscope would cause excitement and call for confirmation by new measurements with increased precision. GG could provide validation at the level of 1% and settle the matter. Recently, GG was proposed as a candidate in the European Space Agency's M5 competition for a new medium-sized science mission, and is awaiting further inquiry, having passed the first round of selection.

The paper is organized as follows.

Sec. 2 presents the WEP as an 'experimentum crucis' of modern physics. Sec. 3 presents the principles of the torsion balance experiment introduced by Eötvös for testing the equivalence between inertial and gravitational mass, a decisive progress over previous experiments with pendulums by Galileo, Newton, Bessel and others. Sect. 4 elaborates on rotation as the other key element for increased precision. Sec. 5 summarizes the state of the art of current WEP experiments and their limitations. Sec. 6 makes the case for a space experiment in low Earth orbit as the way out of such limitations. Sec. 7 discusses the key features of the Microscope space experiment aiming at 10^{-15}. Sec. 8 shows how, on a similar orbit as Microscope and without requesting cryogenic temperatures, a different experiment design allows GG to aim at a 100 times better precision, to 10^{-17}. Recent experimental results relevant to the GG mission are also reported, along with positive news from the Microscope orbiting experiment which corroborate the choice of exploiting rotation in space. Sec. 9 draws the conclusions.

2. An 'experimentum crucis' of modern physics

The UFF was established experimentally by Galileo at the turn of the 17th century using two pendulums of different composition (see [8] for a general discussion on the universality of free fall and the equivalence principle). In 1687, in the opening paragraph of the 'Principia', Newton stated the equivalence of inertial and gravitational mass and then went on to derive the equations of motion showing that all masses fall with the same acceleration under the gravitational attraction of the Earth. If inertial and gravitational mass are equivalent, UFF holds: this was the 'equivalence principle' until the early 20th century.

In 1907, Einstein made the crucial leap from Newton's principle (now referred to as the weak equivalence principle, WEP), to the strong equivalence principle, SEP (also referred to as the Einstein Equivalence Principle, EEP). In the words of Robert Dicke [9]: "*The strong equivalence principle might be defined as the assumption that in a freely falling, non-rotating, laboratory the local laws of physics take on some standard form, including a standard numerical content, independent of the position of the laboratory in space and time. It is of course implicit in this statement that the effects of gradients in the grav-*itational field strength are negligibly small, i.e. tidal interaction effects are negligible. . . . this interpretation of the equivalence principle, plus the assumption of general covariance is most of what is needed to generate Einstein's general relativity.*" Should experiments invalidate UFF (and the WEP), they would invalidate the SEP as well.

As experimental evidence for UFF, Einstein took the results of Eötvös ([1], p. 773), who had achieved an impressive 1000-fold improvement over previous experiments by suspending test masses of different composition on a torsion balance rather than individual pendulums.

The Standard Model of particle physics and the General theory of Relativity, taken together, form our current view of the physical world. While the former governs the physics of the microcosm, the latter governs physics at the macroscopic level. Gravity couples in the same way to all forms of mass–energy, in all bodies, regardless of composition. Such universal coupling makes gravity different from all known forces of nature described by the Standard Model, and is at the heart of the fact that the two theories have so far resisted all attempts at reconciliation into a single unified picture of the physical world. This is the crossroad physics faces at the present time, which is of vital interest not only to theorists, especially given that the nature of about 95% of the matter–energy in the Universe – the so called dark matter and dark energy – is presently unknown.

An experiment capable of testing UFF to extremely high precision can potentially break this deadlock. The situation is reminiscent of that at the end of the 19th century, when Michelson and Morley tested by very precise light interferometry the propagation of the newly discovered electromagnetic waves through the hypothetical ether [10]. Their null experimental result showed beyond question that although its existence was generally assumed, there was in fact no ether; which led to the special theory of relativity. While Michelson and Morley knew which precision their interferometer had to achieve in order to detect the relative velocity between the Earth and the ether, we do not know which precision a test of UFF-WEP should reach to detect a violation, if any. Nonetheless, the issue is so important and the potential reward so huge that many prominent experimental physicists have spent long years in such tests, renewing the effort whenever the possibility for an improvement has arisen.

A WEP experiment is both a test of the foundation stone of GR and a search for a new long-range field coupling to matter in a way that depends on composition (phenomenologically, on powers of the atomic number Z and nucleon number A). The mass–energy content (A/Z ratio; electromagnetic effects in the proton mass, in the neutron mass and in the binding energy of the nucleus; etc.) varies greatly in different atoms and the validity of WEP at very high precision – implying that all forms of mass–energy fall with the same acceleration – is a very strong constraint for all physical theories to comply with [9]. In the years leading to GR, Einstein realized that the theory could stand or fall depending on the results of a single experiment, and even went so far as proposing one himself, calling it a 'simple experiment which would have the significance of an experimentum crucis' [11].

Gravitational self-energy, neutrinos and photons, matter and antimatter, in the purely geometrical treatment of General Relativity, all obey the WEP. Conversely, a confirmed violation could provide the so far missing clue to a more comprehensive physical theory. The higher the precision of the test, the higher the chances to find new physics.

WEP tests are null experiments, by their nature among the most precise types of experiments in physics. They are conceptually simple, can rely on well proven techniques of experimental physics and do not require large apparatus or resources, which makes it easier to detect and control systematic errors. Their very high probing power has already been demonstrated in the lab to
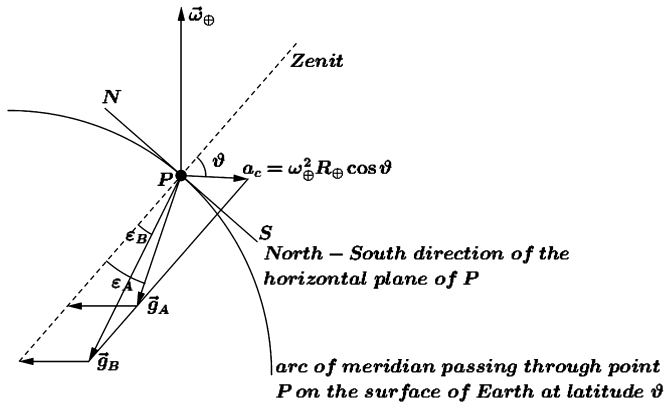
Fig. 1. Two plumb lines in P, at latitude ϑ on the surface of Earth rotating with diurnal angular velocity $\vec{\omega}_\oplus$, have masses of different composition, A and B. We assume them to have the same inertial mass m_i but different gravitational masses: $m_A^g = m_i$, $m_B^g = m_A^g(1 + \eta)$, with $\eta \neq 0$ the Eötvös parameter quantifying the violation of equivalence. The figure shows their deflections towards South (if $\eta > 0$, then $\varepsilon_A > \varepsilon_B$). The figure is obviously not to scale. The deflection angle of a plumb line is very small: $\varepsilon \simeq \frac{\omega_\oplus^2 R_\oplus}{2g} \sin 2\vartheta$, with $R_\oplus$ the radius of Earth and g the local gravitational acceleration, and a maximum deflection of $\simeq 1.7 \cdot 10^{-3}$ rad at exactly 45° latitude.

10^{-13} (Sec. 5); however a quantum leap in precision can only be achieved by performing the experiment in space (Sec. 6).

3. From Galileo to Eötvös: replacing pendulums with the torsion balance

Galileo tested the UFF using masses of different composition suspended from wires of the same length and checking how long the two pendulums would keep on step with each other [12]. From Galileo's own description (see [13], pp. 128–129) and based on a modern analysis of his experiments [14], it is apparent that he achieved a test of the universality of free fall to about 10^{-3}. Newton reports his own pendulum experiments in the 'Principia' and concludes that inertial and gravitational mass are equivalent with a similar precision. Pendulum tests were later improved to reach a few 10^{-5}.

A breakthrough occurred in 1890 [15] when Eötvös published the first results of his tests of the equivalence between inertial and gravitational mass to 5×10^{-8} obtained with test masses suspended on a torsion balance rather than as simple pendulums. In 1909 [16] he won a prize of the University of Göttingen reporting an improved precision of 10^{-8} whose details were published more than a decade later, after his death [17]. In the words of Einstein [18], who referred specifically to the fact that Eötvös had confirmed the equivalence to 10^{-8}, the torsion balance experiment is described as follows:

Eötvös' experimental method is based on the following. A body on the surface of the Earth is acted upon by the terrestrial gravitation and the centrifugal force. The gravitational mass is the determining factor for the first force, and the inertial mass for the second one. If the two did not coincide, then the direction of the resultant of the two (apparent gravitation) would depend on the material of which this body consists.

Fig. 1 is a graphical representation of Einstein's words. If the two masses in the figure, for which violation is assumed, are placed at the two ends of a uniform balance arm, suspended at its center, with radius vectors $\vec{r}_A$, $\vec{r}_B$ from the center of mass of the Earth ($\vec{r} = \vec{r}_B - \vec{r}_A$ is the balance arm), they are subjected to the forces:

$$\vec{F}_A = m_i \vec{g}_A \quad , \quad \vec{F}_B = m_i \vec{g}_B \tag{1}$$

which give rise to the torque:

$$\vec{T} = \vec{r}_A \times \vec{F}_A + \vec{r}_B \times \vec{F}_B \quad . \tag{2}$$

The total gravitational force from Earth $\vec{F}_A + \vec{F}_B$ is applied to the center of mass of the balance, and is counteracted by the tension of the suspension wire. Thus, at 1-g the unit vector along the direction of the wire is:

$$\hat{w} = -\frac{\vec{F}_A + \vec{F}_B}{|\vec{F}_A + \vec{F}_B|} \tag{3}$$

and the component of the torque $\vec{T}$ along $\hat{w}$, which twists the balance until it is counteracted exactly by the restoring torque of the suspension fiber, is:

$$T_w = \vec{T} \cdot \hat{w} = \frac{(\vec{F}_A \times \vec{F}_B) \cdot \vec{r}}{|\vec{F}_A + \vec{F}_B|} \quad . \tag{4}$$

Since $\vec{F}_A \times \vec{F}_B$ lies in the East–West direction of the horizontal plane, the effect of violation on the balance is maximum if its arm is in the same direction.

This formula shows that only forces on the two masses which are not parallel to each other do twist the balance, while parallel forces – even if they have different magnitude – do not affect it. Thus, the torsion balance is an intrinsically differential instrument capable of rejecting common mode forces even of different magnitude – as long as they are parallel. Moreover, by using a very thin suspension fiber, its elastic torsional constant is very small (being inversely proportional to the 4th power of the radius of the fiber), which makes the balance sensitive to extremely tiny torques.

Test masses on the balance are not concentric, therefore they are subjected to different forces because the gravitational field is not uniform (gravity gradient, or tidal effects). In the simple case of a balance with a single arm and one test mass at each end, even if the arm is exactly horizontal, i.e. perpendicular to the local vertical defined by the sum of gravitational and centrifugal force (and assuming no violation), the gravitational accelerations $\vec{g}_A$ and $\vec{g}_B$ on the two masses are not parallel to each other because the masses are separated by the non-zero relative vector $\vec{r}$. Therefore, according to (4), there is a spurious tidal effect which mimics a violation signal of order $r/R_\oplus \simeq 2.3 \cdot 10^{-8}$ for a balance arm $r \simeq 15$ cm.

However, if the balance arm is flipped, rotating it by 180° around $\hat{w}$ (or else, the balance is kept fixed and the test masses are exchanged), the torque due to gravity gradient does not change sign. Instead, in case of violation, the torque T_w caused by the forces (1) would change sign, because $\vec{r} \rightarrow -\vec{r}$, while the forces (1) remain the same. As shown in Fig. 1, the mass which was deflected toward South more than the other because of its composition, will still be deflected more than the other after flipping. Thus, by inverting the masses (or by flipping the balance) the deflection angle caused by a violation of equivalence changes sign while the one due to gravity gradient does not. By rotating the balance around an axis directed along $\hat{w}$ with frequency ν_{spin} (and $\vec{r} \perp \hat{w}$), the violation signal from Earth appears at ν_{spin} while the tidal effect is at $2\nu_{spin}$, and they can be separated. Eötvös' balance was static and could only be flipped manually. It is not surprising that he was able to reach about 10^{-8}.

4. From Eötvös to Eöt-Wash: up-converting the frequency of the signal by rotation

Rotation of the torsion balance around the suspension fiber up-converts a violation signal in the field of Earth from zero (DC) to the rotation frequency, making it possible to detect a deflection angle at a known frequency and to distinguish it from systematics. Moreover, it has long been known [19] that in mechanical

experiments losses due to internal damping are lower at higher frequencies. Therefore, by up-convertion the signal finds itself in a region of reduced thermal noise, a well exploited property in experiments to test the weak equivalence principle [2,20] (see Sec. 5 and Sec. 8).

Rotation of a very sensitive instrument such as the torsion balance has been regarded for a long time as likely to give rise to an unacceptable additional noise. In the mid 1960s Dicke and collaborators [21] used the Sun as source body of a possible violation, by comparing the gravitational and inertial mass of the test bodies which enter, respectively, in the gravitational force from the Sun and in the centrifugal force along the orbit around it. The 'passive' rotation of the balance together with the Earth at the diurnal frequency makes a violation signal from the Sun to appear at this frequency, with no need to 'actively' rotate the balance.

A disadvantage with the Sun as source is a slightly weaker driving signal ($n^2 d_{\oplus\odot} \simeq 6 \times 10^{-3}$ ms^{-2}, with n the annual angular orbital velocity of the Earth and $d_{\oplus\odot}$ the Earth–Sun distance) as compared to that in the field of the Earth ($\omega_{\oplus}^2 R_{\oplus}/2 \simeq 1.7 \times 10^{-2}$ ms^{-2}). More importantly, at the Earth diurnal frequency at which a violation signal would appear, other disturbances are known to affect the torsion balance, such as diurnal thermal effects, diurnal microseismicity, diurnal local mass motions. They are caused by the Sun not through its gravitational force but through its illumination and heating of the Earth's surface and atmosphere (possibly with some time lag), in the day/night, hot/cold cycle. Earth tidal effects are less relevant because they are DC. The main tidal effect from the Sun is at twice the diurnal frequency. More importantly, solar tides are smaller than Earth gradients because the Earth–Sun distance $d_{\oplus\odot}$ is much larger than the Earth's radius. For a balance arm of 15 cm the effect of gravity gradients from the Sun results in a spurious 'violation' of order $r/d_{\oplus\odot} \simeq 10^{-12}$.

Dicke's group reached a precision of 10^{-11}, three orders of magnitude better than Eötvös had done in the field of Earth. A few years later Braginsky [22,23] in Moscow improved the test in the field of the Sun by another order of magnitude, to 10^{-12}.

The improvement achieved in Moscow was possible for various reasons. The ability in suspending masses with tungsten fibers of very low torsional constant and very high mechanical quality made the balance extremely sensitive to tiny torques, and thermal noise very low. The balance carried 8 test masses (4 in Al and 4 in Pt) at the vertices of a regular octagon (the Al masses on one side and the Pt ones on the other). This configuration makes the balance sensitive only to the 5th derivative of the gravitational potential, thus reducing the spurious torques resulting from the gravitational coupling between the mass distribution of the torsion balance and the nonstationary local masses, having a prominent diurnal component for the reasons mentioned above. Last but not least, as Braginsky recalled during a visit by one of us (AMN), the Institute of Physics in Moscow was built on a very deeply rooted rock so that its basement could be used for experiments requiring a very quiet environment. The torsion balance was still in the basement of the Institute at the time of the visit, and records of the balance oscillations impressed on a uniformly rotating film (by means of a laser and a small mirror placed on the balance) were still visible in Braginsky's office.

In 1986 Eötvös' tests in the field of Earth became of interest again because of the possible existence of a new composition-dependent long range force which would act on distance scales accessible by these experiments [24]. Such a force could not be ruled out by the much more precise tests in the field of the Sun, hence pointing out the need to improve the old Eötvös' results.

After Dicke's and Braginsky's experiments it was apparent that the challenge was a rotating torsion balance, in order to achieve in the field of the Earth the same precision demonstrated in the field of the Sun. The Eöt-Wash group led by Eric Adelberger at the University of Washington in Seattle has successfully completed the task, finding no violation of equivalence to a precision of about 10^{-13} in the field of the Earth and to a few parts in 10^{13} in the field of the Sun with experiments that reached the level of thermal noise expected at the rotation frequency of the balance (see [2] and references therein).

The improvement over Eötvös' results in the field of Earth is outstanding, by almost 5 orders of magnitude.

In addition to achieving an extremely smooth and quiet rotation (at about 1 mHz), the authors faced the challenge of spurious torques caused by the gravitational coupling of the Earth and the nearby masses with the mass distribution of the suspended balance.

The torques caused by such couplings may be too large when aiming at high level precision. More importantly, since external masses do not rotate with the balance, some couplings appear at the rotation frequency and therefore compete directly with the violation signal. For instance, the monopole of Earth coupling with a non-zero quadrupole mass moment of the balance results in a torque with the same frequency and phase as the signal.

In order to deal with this issue the suspended balance was carefully designed (and manufactured) in such a way as to minimize its most relevant mass moments. The effects of external masses were measured using various different mass distributions of the balance. By amplifying on purpose, one at a time, its mass moments (starting from the most relevant ones), the measurement provided the torque resulting by its coupling with the external masses, which was then compensated by setting up an appropriate distribution of lead blocks in the vicinity of the balance around it. A similar technique has been used recently by Chinese scientists [25]. The procedure required re-checking and various iterations, and was a complex combination of numerical calculations and experimental measurements carried out with extreme care, until the balance was almost insensitive to mass couplings and sensitive only to a violation of equivalence from Earth or to a new composition-dependent force. Which in the end made possible an improvement by almost 5 orders of magnitude.

In the field of the Sun the improvement by the Eöt-Wash group over Braginsky's experiments has been by less than 1 order of magnitude, to a few parts in 10^{13}. Rotation of the balance moves the signal away from the diurnal frequency to a higher frequency at which thermal noise is lower. However, since the environment does not rotate with the balance, all local disturbances which are induced by the Sun at 1 day period, give rise to torques with the same frequency as the signal. As a result, the improvement over the results obtained by exploiting only the 'passive' rotation provided by the Earth of a stationary balance relative to the Sun could not be as spectacular as in field of Earth.

5. State of the art and limitations

At present the best tests of the weak equivalence principle in the gravitational field of Earth are those by the Eöt-Wash group performed with rotating torsion balances to a precision of 10^{-13}. In the field of the Sun the precision is slightly worse, of a few parts in 10^{13} [2].

The gravitational effect of the Sun on bodies with different mass–energy content has been tested also with LLR experiments. Using more than 45 years of laser ranging data to retroreflectors on the surface of the Moon scientists have shown no violation for Earth and Moon towards the Sun to about 10^{-13} [3,4]. (The composition of the Moon is similar to that of the Earth's mantle and different from the Ni–Fe rich core.)

Rotating torsion balance experiments are limited by gravity gradients, particularly those changing with time (e.g. because of water flow, which is quite relevant in Seattle), and by thermal noise

(losses in the suspension fiber at the rotation frequency of the balance).

Further reduction of gravity gradient effects requires still more care in the balance design as well as in the measurement and compensation of these effects.

Thermal noise can be reduced by lowering the temperature, by increasing the rotation frequency or by manufacturing a suspension fiber with lower internal losses. The Eöt-Wash group has tested a cryogenic balance, but its performance has never matched that of their room temperature balances. The current mHz rotation frequency appears to be the best choice for this mechanical oscillator in 1D (see [2,20]). A fiber fabricated in fused silica, whose losses are expected to be much lower than in tungsten wires, is under consideration. Should such lower losses be confirmed also at the rotation frequency of the balance, and provided that electric charging of the non-conductive fiber is sufficiently small, thermal noise will be reduced.

In addition, the Eöt-Wash group plans to use neutron-rich beryllium and proton-rich polyethylene test bodies because they have a higher composition contrast which would result in a more precise WEP test for the same experiment sensitivity (see [26–28]).

Altogether the Eöt-Wash group is aiming at 1 order of magnitude improvement both in the Earth's and in the Sun's field [2].

LLR tests are affected by gravity gradients between the Earth and the Moon in the field of the Sun. An error Δa in the measurement of the semimajor axis of the lunar orbit yields a spurious 'violation' signal of about $3\Delta a/d_{\oplus\odot}$ [29]. If laser ranging to the Moon has errors at cm level, the limit from gravity gradients is of the order of 10^{-13}.

Considerable efforts have allowed laser ranging to the Moon to be improved by about a factor of 10, with errors at mm level [5], making it possible a similar improvement of LLR tests of WEP. However, for this remarkable hardware achievement to yield a comparable improvement of the WEP test it is necessary that the physical model of the lunar orbit from which the effect of violation is obtained (a polarization of the orbit towards the Sun, absent if WEP holds) be improved too by 1 order of magnitude.

Satellite Laser Ranging (SLR) to Laser Geodynamics Satellites (LAGEOS) orbiting the Earth at a distance of about two Earth radii has been suggested as another possibility for testing WEP. In this case, because of the smaller distance from the source body (Earth) laser ranging errors at cm level yield a larger spurious signal of about $3\Delta a/(2R_{\oplus}) \simeq 2.4 \cdot 10^{-9}$ [29].

It is in fact the large distance from the Sun which allows LLR tests of WEP to reach 10^{-13}, not so much the fact that in the Earth–Moon system 'The heavy masses make it insensitive to any disturbances other than celestial perturbations', as stated in [30]. Non-gravitational forces are a matter of concern for artificial satellites because of their large area-to-mass ratio [31], a parameter which is very small in the case of celestial bodies – such as the Earth and the Moon – due to their large size. In spite of that, the physical model needed to predict the motion of the Moon on the basis of laser ranging data from Earth-based stations to the laser reflectors on the lunar surface, is very complex and hard to improve. From the hardware side, gravity gradient effects at the distance of the Moon resulting from laser ranging errors are indeed the key limitation to WEP tests by LLR, and this is why scientists have worked very hard to reduce them from cm to mm level.

Similarly to the WEP test for the Earth–Moon system around the Sun, one might consider a similar test for the Earth–LAGEOS system in the field of the Sun, since in this case the limitation by gravity gradients would be at 10^{-13} level, not 10^{-9}. However, this test would be a factor 300 less sensitive than in the case of the Moon [29]. In essence, this is because LAGEOS is closer to the Earth than the Moon, therefore its orbit is less affected than the

lunar orbit by the Sun, which is also the source mass of a possible violation.

LLR tests of WEP are therefore not going to be superseded by similar SLR tests, and expectations are for one order of magnitude improvement.

Since Earth and Moon have different non-negligible self-gravitational binding energies LLR can test the property of gravity itself to obey WEP, a test which is obviously beyond the reach of all experiments with artificial test bodies. GR requires self gravitation to obey WEP, while other metric theories of gravity do not, hence these tests can discriminate. Since self gravitation is a very small fraction of the total mass–energy even for celestial bodies, tests of whether it obeys WEP are much less precise. LLR has reached a few parts in 10^4 [4]. Tests of UFF for gravity are performed also in strong field with pulsar-white dwarfs binary [32] and recently with the triple pulsar data for which a result to $3 \cdot 10^{-8}$ is reported [33].

Although WEP tests to the highest possible precision are the deepest probe of GR and the existence of a composition-dependent new force, testing WEP for various forms of mass–energy and as many different compositions as possible, even with low precision, is very important.

With the explosion of supernova 1987A in the Large Magellanic Cloud, observation of a neutrino burst within a few hours of the associated optical burst provided a test of the weak equivalence principle. The observed delay was used to conclude that different particles (neutrinos and photons) undergo the same effect (the Shapiro time delay) from the gravitational field of our galaxy to a few parts in 10^3 [34,35].

At CERN scientists of the AEḡIS collaboration plan to measure the local gravitational acceleration of antihydrogen to 1%, thus performing for the first time a direct test of the weak equivalence principle with antimatter [36].

Another class of WEP tests are the so-called mass drop tests, which have not been used until the late 1980s. Galileo himself, despite the legend, tested UFF with pendulums and not by dropping masses from a height. In 1986 the need emerged to improve Eötvös' tests of WEP in the field of Earth [24] and various groups attempted to do so by dropping masses instead of using a torsion balance. It was clear from Dicke's and Braginsky's experiments that rotation was the key to improve Eötvös' results in the field of Earth, but rotating a very sensitive balance in the lab was regarded as too challenging. On the other hand, the much larger driving acceleration, by almost 600 times ($g \simeq 9.8$ ms^{-2} in mass drop tests versus $g\varepsilon \lesssim 1.69 \cdot 10^{-2}$ ms^{-2} on the torsion balance), and the availability of high resolution, low noise readout based on laser interferometry made mass drop tests very attractive. Non-zero gravity gradients due to initial condition errors (test masses starting from different heights and with different velocities) were a known competing effect which scientists hoped to minimize by devising clever arrangements of the test masses.

In spite of all efforts, and regardless of a factor of about 600 in their favor, drop tests achieved precisions in $\Delta g/g$ of several parts in 10^{10} [37–39]. They would not compete with the Eöt-Wash rotating balance and even the improvement over the torsion balance tests of Eötvös was quite modest, a factor of about 14 despite the sophisticated technologies employed. In terms of sensitivity to differential accelerations, the Eötvös's balance was still more sensitive (its driving signal being about 600 times weaker). As discussed by [39,40] the limitation was due to initial condition errors at release which, because of the gravity gradient of the Earth, give a differential acceleration error which mimics a violation signal.

A longer free fall time would help, since the effect of a violation increases quadratically with time. Drop tests using balloons and sounding rockets have been proposed [41,42]. In [42] the plan was to make drops with reversed axis in order to distinguish the effect of gravity gradient. In [41] the test masses are coupled to form

a sensitive differential accelerometer rotating around a horizontal axis to be dropped inside a capsule carried by a balloon with 30 s free fall time. In this case rotation would make gravity gradient to appear at twice the rotation frequency while a violation signal would be at the rotation frequency, so that they can be separated.

Drop tests with cold atoms have recently reached 10^{-8} [43]. These experiments use known techniques of quantum mechanics to test the universality of free fall with two atom clouds made of atoms of different species (or just different isotopes). What is being tested is whether different atom species (no matter, in principle, how many are contained in each cloud) couple in the same way, or not, with the gravitational field of Earth. As mass drop tests they face the same problems as drop tests with bulk masses, though the way such problems can be solved must take into account the actual features of the experiment.

At the 10^{-8} level the systematic effect of relative initial condition errors in the presence of the gravity gradient of Earth does not matter, but it does when aiming at higher precision.

A cold-atom drop test is underway inside a 10-m-tall vacuum chamber, with a longer free fall time than other ground tests [44, 45]. The precision of the test will depend also on the ability to deal with the gravity gradient/initial condition issue.

6. From ground to space

More than 400 years since Galileo's pendulum experiments in Pisa at the start of 1600, the equivalence of inertial and gravitational mass is confirmed to a precision 10 orders of magnitude better, to 10^{-13}. As shown in Sec. 5, from now on we can expect at most a 10-fold improvement.

The torsion balance is extremely sensitive and very effective in rejecting common mode forces; being manufactured with tolerances at the 10^{-5} level [2] its measurements of differential effects reach a relative precision of 10^{-13}. The problem is that the driving signal relative to which the balance measures extremely small differences, is small. As shown in Fig. 1, the gravitational and inertial mass under scrutiny involve only the small component $g\varepsilon$ of the gravitational acceleration in the horizontal plane arising to counteract, with a tiny deflection ε, the horizontal component of the centrifugal acceleration due to the diurnal rotation of Earth.

Instead, for a mass in orbit around the Earth the centrifugal force (proportional to its inertial mass) equals in magnitude the total gravitational force from Earth (proportional to its gravitational mass), which at low altitude is not much lower than on ground.

An instrument carried by a dedicated spacecraft in orbit around the Earth at altitude h, and capable of detecting a differential acceleration Δa between two masses of different composition, will test WEP to $\eta = \frac{\Delta a}{g(h)}$, with $g(h) = \frac{GM_{\oplus}}{(R_{\oplus}+h)^2}$. With the same sensitivity as the Eöt-Wash balance and $h \simeq 630$ km it would test WEP almost 500 times better, because the driving signal in orbit $g(h)$ is almost 500 times stronger than $g\varepsilon \lesssim 1.69 \cdot 10^{-2}$ ms^{-2} on ground. A further improvement in sensitivity to differential accelerations by a factor of 20 would yield a WEP test in the field of Earth to 10^{-17}. With the Sun as source the improvement would come only from a better sensitivity to differential accelerations, because in this case the driving signal in low Earth orbit is not stronger than on ground.

In principle a WEP experiment in orbit has considerable advantages. At (almost) zero-g masses can be suspended with a very weak coupling constant, which means high sensitivity. As torsion balances have shown, rotation is a key feature of WEP experiments. Unlike on ground, in space the entire laboratory (the spacecraft) is an (almost) isolated system, hence local 'terrain' tilts and microseismicity are much reduced. In addition, the lab co-rotates with the instrument (which would be impossible on ground) needing no stator and no bearings (which eliminates a major source of

noise typical of rotating experiments on ground); if the spacecraft is one-axis stabilized, as it is the case with GG, rotation occurs by angular momentum conservation after initial spin up, and not even thrusters are needed ('passive' rotation, similarly to Earth's diurnal rotation). Co-rotation also eliminates altogether the effects of local mass anomalies (as long as there are no moving parts) because they are DC.

A major issue to deal with in space, which is not present on ground, are non-gravitational forces acting on the outer surface of the free falling spacecraft, such as solar radiation pressure and drag due to residual air along the orbit, which dominates at low altitudes. For WEP experiments in space this is not a small perturbation, but an effect many orders of magnitude larger than the target signal. Microscope deals with it by drag-free control (see Sec. 7); in order to reach higher precision, GG exploits, in addition to drag-free control, the possibility to reject common mode effects (such as the inertial forces resulting from drag) by designing the test masses as a variant in space of the torsion balance (see Sec. 8).

How would a torsion balance work in space?

A practical problem is that in absence of weight it needs to be suspended from both ends. On ground the fiber defines the local vertical for a balance with a given distribution of different composition test masses; its direction is therefore defined by (3) leading to the fact, expressed by (4), that only forces with different directions twist the balance. For the torsion balance the property of high common mode rejection is thus provided mostly by physics. At zero-g this property is lost, and one must rely on precision manufacturing, possibly with some adjustments in flight. In space, with the balance on an orbit that cannot be exactly circular, gravity gradient effects at the same frequency as the signal on test masses a few cm apart would be far too large to be acceptable.

The only way to deal with gravity gradient (tidal) disturbances in space is to have (nominally) concentric masses, in practice two cylinders – of different composition – one inside the other (see Sec. 7 and Sec. 8).

On ground drop tests have been unable to compete with the rotating torsion balance by far. In space they have no factor to gain in strength of the driving signal, which is in fact slightly weaker than it is for drop tests on ground. Some gain would come from a longer time of fall in absence of weight. However, Earth tidal effects due to systematic initial condition errors which limit drop tests on ground, are an issue to be faced also in space. Nonetheless, cold atoms drop tests in space have been investigated [46,47] with the goal of improving the results of similar ground experiments [43,48,49] by 50 million and 5 million times respectively, to reach $2 \cdot 10^{-15}$. At this level the small number of atoms in the clouds compared to Avogadro's number, combined with the gravity gradient/initial conditions errors issue, result in a fundamental limitation [50]. A possible way out is to reduce and/or separate the effect of the gravity gradient of Earth. To this end, an attempt at using methods similar to those employed in experiments with bulk masses has recently been proposed [51]. However, substantial differences between the two kinds of experiments show that the proposed solution would not be as effective as hoped [52].

7. The Microscope experiment in orbit aiming at 10^{-15}

The Microscope satellite was launched on 25 April 2016 with the goal of testing the weak equivalence principle in the field of Earth to 1 part in 10^{15} [6]. It has successfully completed the commissioning phase and since December 2016 is taking scientific data [53]. Preliminary results have been reported at the 2017 conference on Gravitation in La Thuile, Italy.

The principle of Microscope experiment is shown in Fig. 2. It was originally proposed by Paul Worden at Stanford in the
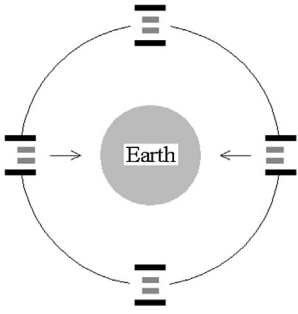
Fig. 2. Principle of Microscope test of the weak equivalence principle. Two concentric, co-axial test cylinders of different composition move on a circular Earth orbit with the symmetry (and sensitive) axis lying in the orbit plane, inside a spacecraft (not shown) whose attitude is actively kept fixed relative to inertial space. It is the original concept proposed at Stanford [54–56], but in Microscope the cylinders are at room temperature rather than cryogenic at superfluid He temperature. Electrostatic forces suspend each cylinder individually in such a way that it is allowed to move only along the symmetry axis, the electrostatic stiffness being low along the axis and high in the plane perpendicular to it. For each cylinder the displacement relative to the enclosure (rigid with the spacecraft, not shown either) unbalances the capacitance bridge and allows the displacement to be measured [57]. The sketch shows a violation of equivalence whereby the inner test cylinder is affected by a slightly stronger acceleration than the outer one towards the center of mass of Earth. The violation signal is at the orbital frequency $\nu_{orb} \simeq 1.7 \cdot 10^{-4}$ Hz. (Figure not to scale; the satellite orbits at $h \simeq 700$ km altitude.)

Fig. 3. Sketch of the effect of the gravity gradient of Earth if the centers of mass of the inner and outer test cylinders shown in Fig. 2 (depicted here as a green and a blue dot) are not exactly coincident along the sensitive axis because of construction errors. The arrows represent the dominant tidal acceleration of the green test cylinder relative to the blue one due to its different radial distance from Earth. If the spacecraft attitude is fixed in space and the orbit is perfectly circular, this acceleration is the same every half orbital period, i.e. its frequency is $2\nu_{orb}$ while the violation signal would be at ν_{orb}, as shown in Fig. 2. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

1970s [54–56] for STEP, a cryogenic space test of the WEP to be carried out at superfluid He temperature to a precision of 1 part in 10^{17}.

With a driving acceleration at the Microscope altitude $g(h) \simeq 8$ ms^{-2}, a WEP test to 10^{-15} requires to detect a differential acceleration between the test cylinders $\Delta a_{WEP} \simeq 8 \cdot 10^{-15}$ ms^{-2}. The corresponding relative displacement depends on the stiffness of the suspension. As reported by [58] (see also [59]), the frequency induced by electrostatic stiffness along the sensitive axis is $1.45 \cdot 10^{-3}$ Hz for the outer cylinder and $9.76 \cdot 10^{-4}$ Hz for the inner one. As a result, a violation of equivalence to 10^{-15} would displace one test cylinder relative to the other by about 100 pm.

Because of construction errors – mostly due to the complex assembly of its electrodes – each cylinder is not exactly centered along the sensitive axis and there is an offset between the two, giving rise to tidal effects from Earth. As shown in Fig. 3, the largest tidal effect is at twice the orbital (and signal) frequency, amounting to $\Delta a_{tide} \simeq \frac{2GM_\oplus}{R_\oplus^3} \Delta x_{off}$. Microscope requires $\Delta x_{off} \simeq$ 20 μm [6], in which case it is $\Delta a_{tide} \simeq 4.5 \cdot 10^{-11}$ ms^{-2}. The offset measured in orbit is only slightly larger than required [60].

With the electrostatic stiffness reported above, the corresponding tidal displacement is $\Delta x_{tide} \simeq 0.6$ μm. If the orbit is not perfectly circular, a fraction of the main tidal effect proportional to the eccentricity will be at the orbital frequency, hence competing with the signal. Luckily, the eccentricity of Microscope's orbit has turned out to be lower than required, more than compensating for the larger offset between the centers of mass of the test cylinders.

The way how Microscope deals with the tidal effect at the signal frequency is, similarly to STEP, to use the main tidal effect at $2\nu_{orb}$ predicted by celestial mechanics to estimate in flight the offset which has generated it to 1/200 of its value ($\simeq 0.1$ μm) and then *a posteriori*, during data analysis, separate from the signal the tidal effect at the same frequency (see [6] for the physical parameters whose knowledge is needed for this procedure to work). Although gravity gradients inside a small lab isolated in space away from Earth (the spacecraft) are not so much a concern as they are for torsion balances on ground, they are still a key issue to deal with when aiming at a high precision test of the WEP.

WEP experiments in space are affected by non-gravitational forces acting on the outer surface of the spacecraft and not on the
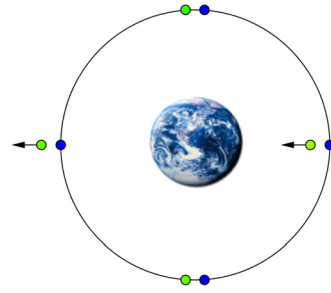
test cylinders weakly suspended inside it. In low Earth orbit the largest one is due to residual air drag, giving rise to an inertial acceleration on each test cylinder equal and opposite to the air drag acceleration of the spacecraft; the largest drag component is at the orbital frequency – like the signal – and many orders of magnitude larger.

Although in principle each test mass should acquire the same inertial acceleration, individually suspended cylinders as in Microscope are very hard to match, thus leaving large differential residuals (poor common mode rejection). In such a case, almost the entire effect of air drag must be compensated by means of thrusters – whose disturbances must be taken care of – which force the spacecraft to follow the common mode motion of the two test cylinders (drag-free control).

Each cylinder sensitive along the symmetry axis is subjected also to a direct non-gravitational acceleration known as 'radiometer effect'. A non-zero residual pressure inside the enclosure, combined with a non-zero temperature gradient along the axis originated by radiation from Earth, results in a spurious acceleration at the orbital frequency, like the signal [61,62].

In Microscope each cylinder is actively controlled using as actuators capacitors similar to those which detect its motion, otherwise it will hit the enclosure and therefore end the experiment.

The electrostatic forces which control the two cylinders so that they are forced to follow the same orbit, ultimately yield the accelerations of the two cylinders relative to each other, including the one due to a violation of equivalence, if present. The control force of each cylinder contains (primarily): a DC term, known as back action [57], due to the misbalance of the capacitors because of construction errors; a component at $2\nu_{orb}$ necessary to counteract the main gravitational tidal effect; a component at ν_{orb} which counteracts the residual inertial acceleration after drag-free control, the tidal effect due to a non-zero orbital eccentricity, the violation signal itself and also – if large enough to compete – the non-gravitational radiometer acceleration.

As reported in [58], the back action acceleration has been estimated by Microscope scientists to amount to $1.85 \cdot 10^{-9}$ ms^{-2} for the outer cylinder (with a misbalance error of 22.3 μm) and to $7.64 \cdot 10^{-10}$ ms^{-2} for the inner one (with an error of 20.3 μm). Its fluctuations have a component at the signal frequency that must be taken into account.

Fig. 2 shows that by maintaining the spacecraft attitude fixed relative to inertial space a WEP violation signal appears at the orbital frequency ν_{orb}. If the spacecraft rotates relative to inertial space at frequency ν_{spin} around an axis perpendicular to the orbit plane, the sensitive axis of the accelerometer rotates with it and

the signal appears at the sum (or difference) of the two frequencies. Being perpendicular to the symmetry axis of the test cylinders, the rotation axis is unstable for small perturbations, hence rotation must be ensured actively, and it is slow in order to add as little noise as possible.

Microscope was originally planned to be operated partly in inertial mode (no rotation) and partly in rotation mode, with a spin frequency up to a few times faster than the orbital one (and in opposite direction) so as to up-convert the signal to $\nu_{WEP} = \nu_{orb} + \nu_{spin}$ reaching almost 10^{-3} Hz [6], a value which is close to the rotation frequency of the Eöt-Wash torsion balance.

As reported recently, it has been found that a spin rate a few times faster than the maximum planned (by a factor 3.9) results in a better sensitivity. Thus, the faster rotation rate has been adopted, resulting in a signal frequency at $3.1 \cdot 10^{-3}$ Hz. Despite the higher amount of propellant required (and consequent shorter duration of the mission) Microscope scientists have concluded that overall 'the balance is in favor of spinning faster'. The inertial mode is no longer mentioned and preliminary results at the faster spin rate shall be announced soon.

Though this result came as a surprise to people who regard rapid rotation as incompatible with precision experiments, in fact it is not. Unlike on ground, rotation in space needs neither stator nor bearings because the whole spacecraft spins together. Thus, even in the case in which rotation requires thrusters and propellant because the spacecraft is not stabilized by conservation of angular momentum around a stable axis, nonetheless, the absence of stator and bearings makes rotation much less noisy than in any rotating experiment on ground.

Furthermore, a faster rotation helps in reducing thermal disturbances, which is what Microscope scientists also report, including a smaller radiometer effect.

Other effects such as gravity gradients, residual drag, radiometer etc., are also up-converted along with the signal. Thermal noise due to internal damping is lower at higher frequencies [19]. In Microscope it is dominated by losses in the thin gold wire connecting each cylinder to the enclosure, with a quality factor of about 100 at the frequency of interest [59]. For a signal to noise ratio SNR = 2 the integration time to reach the target of 10^{-15} is about 1.4 d (20 orbits).

Microscope carries two pairs of test cylinders. In the second pair the two cylinders are made of the same material. An effect with the known frequency and phase of the signal which were detected by the different composition accelerometer (as shown in Fig. 2) and not by the equal composition one, would be a violation signal. This is correct in principle, but other differences between the two accelerometers (such as the fact that they are several cm apart in the nonuniform gravitational field of Earth, or a different response to non-gravitational forces such as the radiometer effect [62]) might yield somewhat different results even in absence of a violation. The equal composition accelerometer is an additional tool to be used together with systematic error checks in order to distinguish spurious effects from a violation signal.

8. The 'Galileo Galilei' (GG) candidate space experiment to test the WEP to 10^{-17}

'Galileo Galilei' (GG), to be flown in low Earth orbit as Microscope, aims at testing the weak equivalence principle to 1 part in 10^{17} at room temperature [7].

GG can target a precision 100 times better than Microscope because of its design. Table 1 shows in summary a quantitative comparison between the two experiments. Two are the key features which make GG different from Microscope: *i*) it spins 1320 times faster (340 times faster after the recently adopted higher spin rate), up-converting the signal from the orbital frequency to
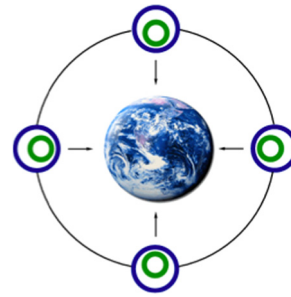

Fig. 4. Principle of the GG test of the weak equivalence principle. Two concentric, co-axial test cylinders of different composition move on a circular Earth orbit with the symmetry axis perpendicular to the orbit plane, inside a spacecraft (not shown) whose attitude is stabilized by passive 1-axis rotation at $\nu_{spin} \simeq 1$ Hz around the symmetry axis. The cylinders form a mechanical oscillator sensitive to differential forces in the plane perpendicular to it, and much stiffer along the axis (Fig. 6). The sketch shows a violation of equivalence whereby the inner (green) cylinder is affected by a slightly stronger acceleration than the outer (blue) one. The violation signal is a vector of constant size (for a perfectly circular orbit) pointing to the Earth's center of mass while orbiting around it at $\nu_{orb} \simeq 1.7 \cdot 10^{-4}$ Hz. The experiment is designed in the framework of Newtonian dynamics because GR effects taking into account rotation of both the test cylinders and the Earth have been shown to be negligible by far [63]. (Figure not to scale; the satellite orbits at $h \simeq 630$ km altitude.) (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

$\nu_{spin} \simeq 1$ Hz ($\nu_{spin} \gg \nu_{orb}$) where thermal noise is much lower; *ii*) the test cylinders (10 kg each, for lower thermal noise and reduced non-gravitational effects) are arranged to form a balance whose arms can be adjusted in flight in order to reject common mode effects.

Rapid spin is possible by turning the symmetry axis of the test cylinders by 90° with respect to Microscope (compare Fig. 4 with Fig. 2), making it the (stable) rotation axis of the whole spacecraft which stabilizes its attitude, while the plane perpendicular to it is the sensitive plane of the test cylinders. The spacecraft conforms to the cylindrical symmetry imposed by the test bodies and the whole system co-rotates passively (by conservation of angular momentum after initial spin-up) up-converting the signal from the orbital to the spin frequency by a factor of about 5800 (see Fig. 4 and Fig. 5).

The way how two concentric cylinders are coupled to form a beam balance is shown in Fig. 6. By balancing its arms in flight against the common mode effect of drag using piezo actuators, the balance allows all common mode effects (primarily the effect of drag) to be rejected. Since this is 50 million times weaker than local gravity for ordinary balances on ground, it is apparent that balancing in space is much easier than on ground, where it has been known since a long time that balances can detect differences of weight by 10^{-7}–10^{-8} [65]. Thus, unlike in Microscope where drag can only be compensated actively by the drag-free control system, in GG drag is partially compensated and partially rejected by the balance, leaving a residual differential effect half the size of the target signal of GG (Table 1, entry III).

The issue of a balance in space, and the original GG setup, were the subject of lively discussions during a few days visit to Professor Braginsky in Moscow by one of us (AMN). He believed that a high precision test of WEP needs a balance, in space as well as on ground, and strongly advocated that the test masses be kept as passive and undisturbed as possible (Table 1, entry X).

Thermal noise due to internal damping decreases as the square root of the frequency [19], as confirmed by rotating torsion balances [2]. The GG balance of Fig. 6 is in essence a 2D harmonic oscillator rapidly rotating at supercritical speed (i.e. above the natural normal mode frequency of oscillation). Its thermal noise is calculated in [20], where the equations of motion and their solution show that: *i*) the signal determines the equilibrium position and
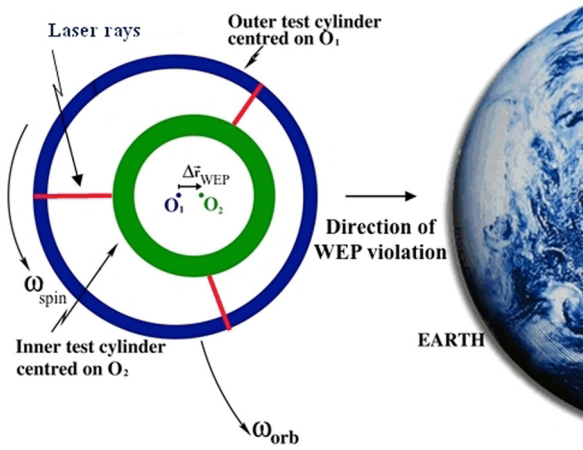
Fig. 5. The sketch shows a relative displacement vector $\Delta\vec{r}_{WEP}$ between the centers of mass of the two cylinders O_1 and O_2 caused by a violating differential acceleration driven by Earth $\Delta\vec{a}_{WEP}$ ($\Delta\vec{r}_{WEP} \simeq \frac{\Delta\vec{a}_{WEP}}{4\pi^2\omega_{dm}^2}$). The red lines depict the laser rays (at 120° from each other) of the interferometry gauge designed to read the relative displacements of the test cylinders. The whole system co-rotates at $\nu_{spin} \simeq 1$ Hz, therefore the signal $\Delta\vec{r}_{WEP}$ is read at $\nu_{spin} - \nu_{orb} \simeq \nu_{spin}$. The cylinders are weakly coupled as a 2D harmonic oscillator with normal mode differential frequency ω_{dm}, hence equilibrium is reached at the displaced position $\Delta\vec{r}_{WEP}$. This is the physical observable measured by the laser gauge, yielding the Eötvös parameter $\eta = \frac{\Delta a_{WEP}}{g(h)}$ that quantifies the level of WEP violation. For the GG target $\eta = 10^{-17}$ the displacement signal to be measured is 0.6 pm. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

in 2D it is up-converted above the critical speed without attenuation; *ii*) the offset errors by construction are reduced as the ratio spin-to-natural frequency squared (a well known phenomenon called self-centering); *iii*) a weak instability ('whirl motion') occurs around the equilibrium position with time constant $\tau = \frac{Q}{\pi}P_n$, where P_n is the natural period of oscillation and Q the quality factor of the oscillator at the (high) spin frequency not at the (low) natural one [66].

Measurements of the quality factor of CuBe suspensions by Virgo scientists report $Q = 20000$ close to 1 Hz ([67], Fig. 2). At a few Hz, for larger oscillation amplitudes at which higher losses are expected, we have measured $Q = 19000$ [68]. With $Q = 20000$, for a 1 Hz signal, thermal noise due to internal damping in GG is only slightly larger than other sources of thermal noise, yielding altogether an integration time of $\simeq 3.5$ h for a signal-to-noise ratio SNR = 2 [69]. Targeting a 100 times smaller signal requires an integration time 10^4 times longer, hence it is quite remarkable that, for the same SNR, GG can reduce thermal noise below the signal even quicker than Microscope (Table 1, entries IV, V and VI).

Because of the limited duration of space missions – especially when they rely on a finite amount of propellant to compensate solar radiation pressure and air drag – a very short integration time is crucial in order to ensure a rigorous campaign of systematic error checks. It needs a readout with low enough noise to match it.

In GG the differential displacements of the test cylinders are read by a heterodyne laser interferometry gauge. It was originally proposed by Mike Shao, based on the laser gauge he developed at JPL for the SIM mission of NASA [70]. It is now under development at the Italian national metrology institute in Torino (INRIM) [71, 72], where it has recently demonstrated a displacement noise of $0.6\frac{pm}{\sqrt{Hz}}$ at 1 Hz (Fig. 7), which means 1 s integration time for the target displacement of 0.6 pm, up-converted to 1 Hz. The high frequency of the signal makes the GG interferometer much less complex than the one which has successfully flown in LISA-PF (also a heterodyne laser interferometer) because in this case low noise is required at frequencies well below 1 Hz [73]. For the GG laser gauge lab measurements have checked error sources such as laser frequency noise, cross talk and temperature sensitivity. A spurious
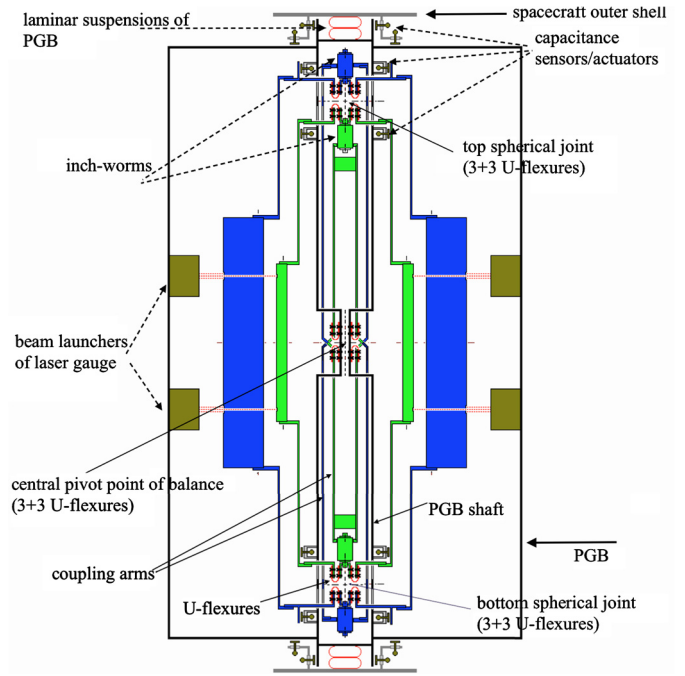

Fig. 6. Sketch of the GG balance made of 2 coaxial concentric test cylinders of different composition (inner in green, outer in blue) spinning around the symmetry axis and weakly coupled in the plane perpendicular to it: section along the spin axis. An animation of this balance, oscillating with 1 s spin period under the effect of a violation signal is available on the GG webpage [64]. In the animation the oscillation amplitude is largely exaggerated: a violation to 10^{-17} would cause an oscillation of 0.6 pm, which is only 3×10^{-11} the size of the 2 cm gap between the outer surface of the inner cylinder and the inner surface of the outer one. The s/c is passively stabilized by one-axis rotation at $\nu_{spin} \simeq 1$ Hz, nutation damping being provided by the weak coupling between the spacecraft and an intermediate stage (the PGB-Pico Gravity Box, and its shaft) to which the balance is connected at the center by means of U-flexures (in red). This is the pivot of the balance. The motion of PGB relative to the s/c is read by capacitance sensors and drives the drag-free control system. At its two ends each test cylinder is connected (through a light rigid interface and 3 U-flexures 120° apart; 2 shown in planar section) to the 2 ends of the coupling arms; the short ones for the inner green cylinder; the long ones for the outer blue cylinder. The result are two spherical joints (top and bottom) such that any differential force acting between the test cylinders in the plane perpendicular to the symmetry axis will displace their centers of mass by tilting the coupling arms (pivoted at the center). The laser gauge boxes are fixed on the PGB and shown in brown. At each end of the coupling arms (in blue and green) are shown the inch-worms which allow the balance to be balanced in order to reject accelerations acting in common mode on both cylinders. The two shorter parts of each coupling arm (pertaining to the inner cylinder and shown in green) have a small additional mass each (in green) so that the pivot center is at their center of mass. In the balance the mass of the test cylinders dominates over the mass of the coupling arms and interfaces. Note the symmetry of the balance both in azimuth and top/down. This clever design of a beam balance with concentric test masses and perfect symmetry, as needed in space for testing the weak equivalence principle, is due to Donato Bramanti. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

effect which might arise because of rotation (the Sagnac effect) has been investigated, finding that it is much smaller than the signal [72].

Comparison with the capacitive readout of Microscope is shown in Table 1, entry VII (a cap readout is used also in the GG demonstrator on ground – GGG). By comparison with capacitive sensors the laser gauge is intrinsically differential, it does not require calibration, its sensitivity does not decrease with increasing gaps, it is less noisy.

Bringing the radiometer effect 100 times below the Microscope target is impossible at room temperature (in the Stanford proposal cryogenics would ensure an extremely low pressure which makes radiometer negligible). Radiometer acts along the symmetry axis. GG can deal with it at room temperature because it is sensitive in

Table 1
Microscope and GG: a quantitative comparison.

Microscope	GG
I. Target of WEP violation $\eta = \Delta a / g(h)$, $g(h) \simeq 8\,\mathrm{ms}^{-2}$	
$\eta_{\mu scope} = 10^{-15}$	$\eta_{GG} = 10^{-17}$
II. Signal: differential acceleration & differential displacement	
Signal along symmetry axis of test cylinders (1D sensitivity)	*Signal in plane $\perp$ to symmetry/spin axis (2D sensitivity)*
$\Delta a_{\mu scope} \simeq 8 \times 10^{-15}\,\mathrm{ms}^{-2}$	$\Delta a_{GG} \simeq 8 \times 10^{-17}\,\mathrm{ms}^{-2}$
$\Delta x_{\mu scope} \simeq 10^{-10}\,\mathrm{m}$	$\Delta r_{GG} \simeq 6 \times 10^{-13}\,\mathrm{m}$
III. Test cylinders coupling, common mode rejection & drag-free control	
– Each test cylinder individually suspended by electrostatic forces along symmetry/sensitive axis, connected to cage by Au wire	– Test cylinders coupled by CuBe flexures to form a beam balance with beam along symmetry/spin axis sensitive in plane $\perp$ to it
– Common mode accelerations of individual cylinders to be reduced by matching their sensitivities with inflight calibrations: CMR = 18	– Beam balance to be balanced in flight (against common mode inertial acceleration due to drag) by adjusting balance arms using differential displacement as driver and piezo as actuators: CMR = 10^5
– Drag-free control & CMR = 18 will bring the residual differential acceleration due to drag (same frequency as signal) to $5.6 \times 10^{-15}\,\mathrm{ms}^2 \simeq 0.7 \Delta a_{\mu scope}$	– Drag-free control & CMR = 10^5 will bring the residual differential acceleration due to drag (same frequency as signal) to $4 \times 10^{-17}\,\mathrm{ms}^2 \simeq 0.5 \Delta a_{GG}$
IV. Signal readout frequency	
– Inertial mode: displacement signal of each test cylinder detected at $\nu_{orb} \simeq 1.7 \times 10^{-4}\,\mathrm{Hz}$ – Rotation mode ($\nu_{spin1} = 3\nu_{orb}$): displacement signal of each test cylinder detected at $\nu_1 = 4\nu_{orb} \simeq 5 \times 10^{-4}\,\mathrm{Hz}$ – Rotation mode ($\nu_{spin2} = 5\nu_{orb}$): displacement signal of each test cylinder detected at $\nu_2 = 6\nu_{orb} \simeq 8.4 \times 10^{-4}\,\mathrm{Hz}$	– Differential displacement signal detected at: $\nu_{spin} \simeq 1\,\mathrm{Hz}$
V. Internal damping at signal readout frequency	
$Q \simeq 100$	$Q \simeq 20000$
VI. Integration time	
SNR=2 in 1.4 d integration time Typical duration of science session: 8.3 d $\rightarrow$ SNR = 4.9	SNR=2 in 3.5 h integration time Typical duration of science session: 1 d $\rightarrow$ SNR = 5.2
VII. Readout and readout noise	
– Capacitance readout for each test cylinder (600 μm gap) – At $\nu \geqslant 10^{-3}\,\mathrm{Hz}$ [59]: $38.5\,\frac{\mathrm{pm}}{\sqrt{\mathrm{Hz}}}$ for outer cylinder displacement $40.4\,\frac{\mathrm{pm}}{\sqrt{\mathrm{Hz}}}$ for inner cylinder displacement ($\simeq 200\,\frac{\mathrm{pm}}{\sqrt{\mathrm{Hz}}}$ reported at $10^{-3}\,\mathrm{Hz}$ [57])	– Differential heterodyne laser interferometer (2 cm gap) – At $\nu_{spin} \simeq 1\,\mathrm{Hz}$: $0.6\,\frac{\mathrm{pm}}{\sqrt{\mathrm{Hz}}}$ for relative displacements of test cylinders
VIII. Radiometer non-gravitational effect	
– Along sensitive axis – Solved with: residual pressure of $10^{-5}\,\mathrm{Pa}$ & passive temperature stabilization	– Along axis $\perp$ to sensitive plane – Causes indirect tidal effect in sensitive plane. Solved with: residual pressure of $10^{-5}\,\mathrm{Pa}$ & passive temperature stabilization
IX. Test cylinders centers of mass offsets	
– 20 μm construction error for each test cylinder – 0.1 μm offset by inflight estimates & *a posteriori* data analysis	– 10 μm construction error for each test cylinder – 0.6 nm self-centering of each test cylinder by physics laws – 1.7 nm test cylinders offset by capacitance control of whirl motion [66]
X. Test cylinders instabilities & control forces	
• Each test cylinder (otherwise unstable) controlled along sensitive axis by capacitance sensors & actuators; control always on [58] • back action (DC) for outer cylinder (22.3 μm offset): $1.85 \times 10^{-9}\,\mathrm{ms}^{-2}$ • back action (DC) for inner cylinder (20.3 μm offset): $7.64 \times 10^{-10}\,\mathrm{ms}^{-2}$	• Each test cylinder weakly unstable by whirl motion with time constant $\tau = \frac{Q}{\pi} P_n \simeq 9.6\,\mathrm{d}$ ($Q \simeq 20000$, $P_n = \frac{2\pi}{\omega_n} \simeq 130\,\mathrm{s}$); controlled by capacitance sensors & actuators; control off during science data taking • control at whirl frequency to keep offset $\leqslant 1.7\,\mathrm{nm}$: $\geqslant 4 \times 10^{-12}\,\mathrm{ms}^{-2}$

the plane perpendicular to it. Nonetheless, the largest tidal effect between the test cylinders of GG in the sensitive plane is indirectly due to radiometer along the much stiffer symmetry axis (Table 1, entry VIII).

Direct tidal effects come from offsets between the centers of mass of the test cylinders in the sensitive plane. Self-centering as predicted theoretically ensures, for the test cylinders of GG, a reduction of the offsets achieved by fabrication by more than
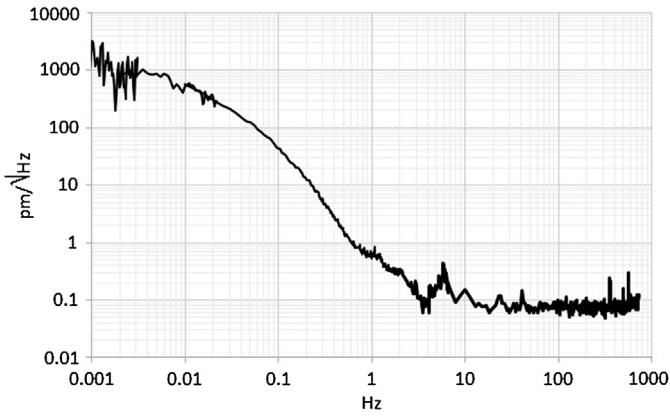
Fig. 7. Displacement noise measured at INRIM with a laser gauge suitable for GG [72]. At the GG signal frequency of 1 Hz the noise is 0.6 $\frac{pm}{\sqrt{Hz}}$. At high frequencies the noise measured is about 0.1 $\frac{pm}{\sqrt{Hz}}$ (electronics noise, not interferometer noise). Noise at low frequencies is related to the optical fibers and can be reduced if needed. The measurement is not in vacuum and the frequency of the laser is not stabilized.

4 orders of magnitude (Table 1, entry IX; see [74], Fig. 5 for experimental evidence of self-centering with the GGG demonstrator).

As shown in [20], losses generate a whirl motion around the equilibrium position at the natural frequency. It is known that, except in the presence of large dissipation, the frequency of whirl is the same as the natural frequency at zero spin. However, there was no experimental demonstration so far that the relevant quality factor is that at the spin frequency, not at the natural one. Since losses are lower at higher frequencies, the issue is very important.

The GGG demonstrator, with complex CuBe cardanic joints to sustain weight, a spin frequency $\nu_{spin} = 0.16$ Hz and a differential mode frequency of $\nu_{diff} = 0.074$ Hz ($P_{diff} = 13.5$ s), shows that whirl grows with $Q_{\nu_{spin}} = 2126$ and time constant $\tau = \frac{P_{diff}}{\pi} Q_{\nu_{spin}} \simeq 2.5$ h (Fig. 8), while at zero spin, the amplitude of oscillation at the differential mode frequency decays faster, with $Q_{\nu_{diff}} = 948$ (Fig. 9).

The theoretical predictions are demonstrated: the frequency of whirl is the same as the natural differential frequency, whirl growth occurs as in the analytical solution and the quality factor improves when the system spins faster than this frequency. Nonetheless, at 1-g it would be hard to have a time constant of whirl growth longer than 2.5 h (a long differential period requires weak suspensions, which is impossible with 10 kg test cylinders at 1-g), and this means that in GGG whirl control, performed by means of capacitance sensors/actuators, must be on all the time (see [7] for the results of one month measurements).

In space with lower stiffness, hence lower natural frequencies, higher spin rate hence higher quality factor, whirl growth is much slower, with a much longer time constant. In GG the time constant for the whirl growth of each test mass is almost 10 days, whirl damping (performed with capacitors as on ground) requires very weak forces and it is off during science data taking, thus leaving the test masses totally passive, as advocated by Braginsky (Table 1, entry X). On ground, only rotors with high rotor/stator and bearings noise, or very high dissipation have reported chaotic behavior [75–77].

In such rotors the frequency of whirl when the system rotates is no longer equal to the corresponding natural normal mode frequency at zero spin, which is 'the smoking gun' for the onset of whirl chaos. GG has no motor, no bearings, no stator and high Q, hence whirl chaos is ruled out. Experimental evidence is provided by the GGG demonstrator on ground, which has motor, stator and bearings noise and higher dissipation than in space (due to more complex suspensions and a lower spin rate), and yet the whirl and
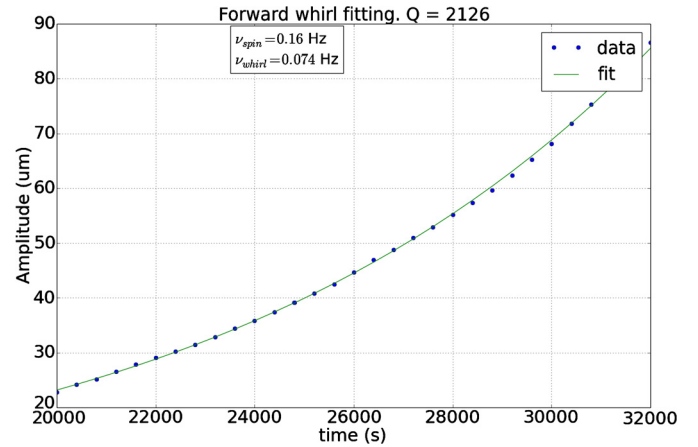

Fig. 8. Exponential growth of whirl motion of the test cylinders relative to each other in the GGG laboratory demonstrator with time constant $\tau = \frac{Q}{\pi} P_{whirl}$. The system spins at $\nu_{spin} = 0.16$ Hz and the whirl period is $P_{whirl} = \frac{2\pi}{\nu_{whirl}} = 13.5$ s. We measure: $Q_{whirl} = 2126$, hence $\tau \simeq 2.5$ h.
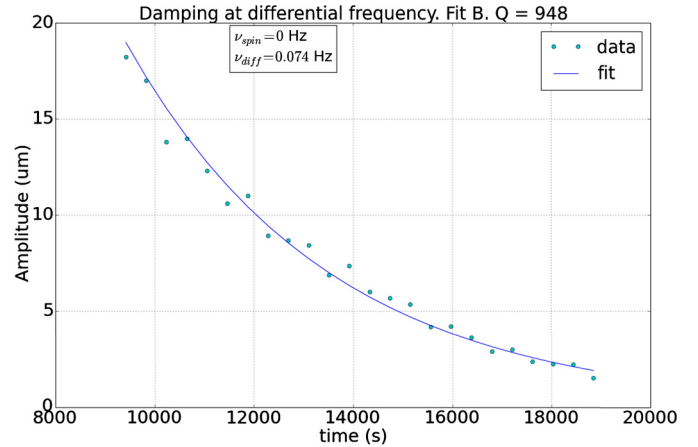

Fig. 9. Exponential decay of the relative oscillation amplitude of the GGG test cylinders not rotating and oscillating at the natural differential mode frequency $\nu_{diff} = 0.074$ Hz. The system is the same as in Fig. 8 after stopping rotation and before breaking vacuum. We measure $Q_{\nu_{diff}} = 948$ (with a decay time constant of 1.1 h).

the normal mode frequencies have always been found to be the same.

The spin axis of GG (perpendicular to the sensitive plane) is essentially undisturbed and remains fixed in space due to its high rotation energy. Instead, being in a sun-synchronous orbit, the normal to the orbit plane precesses around the north pole by $\simeq 1°/$d. Over an orbital arc of 90 d, in a totally passive and deterministic manner, the angle between the two varies from $-45°$ to $+45°$, thus making the violation signal vary in a perfectly known way, different from all systematic effects – even at the same frequency – that we are aware of. With 90 measurements during the cycle (Table 1, entry VI), all to the target precision, the variation of the measured effect with the changing geometry is mapped and the signal, if present, is separated from errors with certainty (there will be at least three 90-d cycles during the mission). A second balance with equal composition test cylinders, co-axial and concentric with the one with different composition cylinders, can also be accommodated for further checking [78].

The test cylinders of GG spin around their symmetry axis. Manufacturing imperfections, being fixed with the cylinders, produce DC effects on their relative displacements which do not compete with the much higher frequency of the signal. This unique feature, together with large test masses (10 kg each), self-centering and

Table 2
Budget of the major systematic errors for GG targeting a WEP test to $\eta = 10^{-17}$.

Acceleration $(x, y$ sensitive plane) due to:	Frequency	Differential acceleration $(\mathbf{m/s^2})$
WEP violation signal	ν_{orb}	$8.1 \cdot 10^{-17}$
External non-gravitational forces after drag compensation <u>and</u> common mode rejection	ν_{orb}	$4 \cdot 10^{-17}$
Earth's monopole coupling with test masses quadrupole moments	ν_{orb}	$6 \cdot 10^{-18}$
Earth's tide coupled to radiometer effect along spin axis z	ν_{orb}	$4 \cdot 10^{-18}$
Earth's tide coupled to thermal emission along spin axis z	ν_{orb}	$8.1 \cdot 10^{-19}$
Earth's tide coupled to non-gravitational accelerations along spin axis z	ν_{orb}	$1.7 \cdot 10^{-19}$
Earth's tide coupled to radiometer effect along spin axis z	$2\nu_{orb}$	$8.5 \cdot 10^{-16}$
Earth's tide coupled to non-gravitational accelerations along spin axis z	$2\nu_{orb}$	$3.4 \cdot 10^{-17}$
Magnetic dipole moment of one test mass coupled to magnetization induced on the other by Earth's magnetic field $B_\oplus$	$2\nu_{orb}$	$1.8 \cdot 10^{-16}$

balancing, allows requirements on their fabrication to be relaxed. It is therefore possible to test 'unusual' materials which are known to probe WEP violation more deeply but have never been implemented because of fabrication issues. The choice of Pb and C_2H_4 proposed for GG in 2009 in [27] has been found to yield – with the same driving signal and the same sensitivity to differential accelerations – a better test of WEP by a factor of 12 [28], which would mean the possibility of reaching 10^{-18}.

We have discussed the key facts whereby GG can aim at a WEP test to 10^{-17}. Although a full report on the error budget (and consequent experiment and mission requirements) is beyond the scope of this paper, it is worth listing the major systematic errors which compete with a WEP violation signal at this level. As shown in Table 2, these are, in the first place, the effects that have the same frequency as the signal (in the non rotating frame of the spacecraft this is the orbital frequency $\nu_{orb} \simeq 1.7 \times 10^{-4}$ Hz, which the rotation of the s/c up-converts to 1 Hz as shown in Fig. 5) and secondly, those not too far from the signal frequency, at twice it. As expected, the most relevant error is due to non-gravitational forces on the outer surface of the spacecraft (although at a large phase difference from the signal) which GG can reduce below the target only thanks to the unique feature of rejecting common modes in flight. Frequency separation allows effects at $2\nu_{orb}$ which are larger than the signal to be clearly distinguished from it. Effects at frequencies much farther away from ν_{orb}, such as the natural/whirl frequencies, can be much bigger than the signal and still be clearly distinguished from it.

GG has been under investigation for over 20 years (including the construction and testing of a full scale ground demonstrator). The most comprehensive industrial study funded by ASI (Agenzia Spaziale Italiana), reported in about 30 documents [79], features a full 3D simulation of the space experiment based on the simulator developed at Thales Alenia Space in Torino, Italy for the successful GOCE mission of ESA [80].

9. Conclusions

In this paper we have shown how the progress in WEP tests has depended on innovations in experimental techniques, driven by theoretical insight.

The torsion balance, first used to test the equivalence of inertial and gravitational mass at the end of the 19th century, was the first high accuracy instrument: intrinsically differential to reject common mode forces, coupled to the lab by a thin fiber with very low torsional stiffness to enhance sensitivity and high mechanical quality for low thermal noise, it was limited by the lack of modulation of the signal in the field of Earth. Later on it was realized that, by taking the Sun as source, the diurnal rotation of Earth would provide the desired modulation, with no need to flip or rotate the balance. Next, the ability to smoothly rotate the torsion balance itself, and to reduce and/or compensate gravity gradients, along with the understanding that losses due to internal damping decrease at higher frequency, whereby a rotating balance moves the signal to a region where thermal noise is lower, have led to about five orders of magnitude improvement in the field of the Earth and almost one in the field of the Sun.

Eventually, rotating torsion balance experiments are limited by gravity gradients, particularly those changing with time, and by thermal noise driven by losses in the suspension fiber at the rotation frequency of the balance. Current laboratory experiments have reached a point where significant improvement is hard to achieve, and the same applies to LLR, for the reasons discussed in Sec. 5.

Advancing by orders of magnitude requires moving the experiment to space.

Microscope is the first WEP experiment to have been implemented in space, realizing a concept originally proposed (in a cryogenic version) at Stanford University in the 1970s. The Microscope sensor is made of two concentric cylinders – their common symmetry/sensitive axis lying in the plane of the orbit – whose displacements along the axis are individually read and controlled in order to minimize the separation between their centers of mass. In low Earth orbit the driving acceleration is almost 500 times larger than for a torsion balance in the lab, and by this fact alone a corresponding increase in sensitivity can be targeted. The main design parameters of Microscope are discussed in Sec. 7. It appears that the 10^{-15} target would be very hard to improve with this design because of such features as tidal accelerations resulting from relative position errors by construction and mounting, residual non-gravitational accelerations due to poor common mode rejection, large control forces to be applied in the direction of the signal, thermal noise due to large losses at low frequency in a loose gold wire. Even cryogenics, with all the attendant practical limitations and problems, would not solve all the open issues.

The GG sensor, too, is a pair of concentric cylinders but their symmetry axis is placed orthogonal to, rather than in, the orbit plane, and is sensitive in 2D rather than along the axis. By this simple choice, the rotation rate is no longer limited by stability issues and can be chosen sufficiently high where losses due to internal damping are low and thermal noise is small. With carefully designed U-flexures and some ingenuity, the two concentric cylinders are coupled as in a beam balance with the beam along the symmetry/rotation axis, sensitive in the plane perpendicular to it. In flight, against the common mode effect of drag which is many orders of magnitude weaker than 1-g, the beam balance can be precisely balanced by adjusting its arms, more easily than a balance on ground against the very large effect of local gravity, thus

achieving good common mode rejection. As a result, air drag is partially compensated by drag-free control and partially rejected by the balance, making its residual differential effect much smaller than it could possibly be by active compensation only, with no need to push the requirement for low noise trusters to unrealistic limits. When rotating above a critical rate, physics itself takes care of aligning the rotation axis to the symmetry axis much better than it can possibly be achieved by construction (self-centering), the remaining offset being very small, calculable and measurable with the ground demonstrator. Because of losses in the flexures, motion around the equilibrium position is weakly unstable (whirl motion). The relevant losses which determine the growth rate are predicted to occur at the (high) spin frequency (at which losses are known to be small). This fact, predicted theoretically and demonstrated experimentally (see Sec. 8) ensures a very slow growth, such that in GG whirl control can be switched off during science data taking, leaving the test cylinders totally passive.

GG was conceived to address and solve the shortcomings of alternative experiment designs, as they have emerged in an almost 45-year long history of proposals for testing the WEP in space. A comparison of the key issues, as carried out in Secs. 7 and 8 and summarized in Table 2, shows how the GG approach makes it possible to target a WEP test 100 times more precise than Microscope, to 10^{-17}. Should Microscope detect a sign of new physics, another mission capable of greater precision would be called for, and urgently. As we show in Sec. 8, a design based on different principles is needed.

In recent years, three space missions have shown that very high precision physics experiments are feasible in space, and that the relevant technological challenges can be met. Most technological ingredients of GG have received confirmation as part of successful satellite missions. GOCE (2009–2013) demonstrated drag-free control, high passive thermal stability and high precision capacitive measurement and control technology. LISA-PF (2015–2017) confirmed these findings and demonstrated very low noise heterodyne laser metrology on cm-length scales. Microscope (2016-ongoing) showed that precise tests of WEP in space are feasible and can take advantage of rotation because – contrary to the widespread concern, originating from lab experience, whereby rotation impairs precision – the space environment is favorable to rotation for the simple reason that in space, unlike on ground, the whole 'lab' (the spacecraft) rotates relative to inertial space. As to the GG instrument, its high quality suspensions have heritage from gravitational wave detectors and the Eöt-Wash balance, and it has been tested as far as feasible for an instrument designed to fly with a dedicated demonstrator at 1-g.

Currently, GG is a candidate in the ESA M5 competition, the final phase of which will take place in the same time frame as the disclosure of the early results of Microscope. Should the need arise, peer review of all proposed experiments would be a prerequisite before a confirmation mission is launched. The arguments reported here indicate that GG has good chances to emerge as the most viable and cost-effective option.

Acknowledgements

References

[1] A. Einstein, Grundlage der allgemeinen Relativitätstheorie, Ann. Phys., Ser. 4 49 (1916) 769–822;
see A. Einstein, The foundation of the general theory of relativity, in: The Principle of Relativity, Dover Pub. Inc., N.Y., USA, 1952.

[2] T.A. Wagner, S. Schlamminger, J.H. Gundlach, E.G. Adelberger, Torsion-balance tests of the weak equivalence principle, Class. Quantum Gravity 29 (2012) 184002.

[3] J.G. Williams, S.G. Turyshev, D.H. Boggs, Lunar laser ranging tests of the equivalence principle, Class. Quantum Gravity 29 (2012) 184004.

[4] J. Mueller, F. Hoffman, L. Biskupek, Testing various facets of the equivalence principle using lunar laser ranging, Class. Quantum Gravity 29 (2012) 184006.

[5] T.W. Murphy Jr., E.G. Adelberger, J.B.R. Battat, C.D. Hoyle, N.H. Johnson, R.J. McMillan, C.W. Stubbs, H.E. Swanson, APOLLO: millimeter lunar laser ranging, Class. Quantum Gravity 29 (2012) 184005.

[6] P. Touboul, G. Metris, V. Lebat, A. Robert, The MICROSCOPE experiment, ready for the in-orbit test of the equivalence principle, Class. Quantum Gravity 29 (2012) 184010.

[7] A.M. Nobili, M. Shao, R. Pegna, G. Zavattini, S.G. Turyshev, D.M. Lucchesi, A. De Michele, S. Doravari, G.L. Comandi, T.R. Saravanan, F. Palmonari, G. Catastini, A. Anselmi, "Galileo Galilei" (GG): space test of the weak equivalence principle to 10^{-17} and laboratory demonstrations, Class. Quantum Gravity 29 (2012) 184011.

[8] A.M. Nobili, D.M. Lucchesi, M.T. Crosta, M. Shao, S.G. Turyshev, R. Peron, G. Catatstini, A. Anselmi, G. Zavattini, On the universality of free fall, the equivalence principle, and the gravitational redshift, Am. J. Phys. 81 (2013) 527.

[9] R.H. Dicke, The Theoretical Significance of Experimental Relativity, Blackie and Son Ltd., London and Glasgow, 1964.

[10] A.A. Michelson, E.W. Morley, On the relative motion of the Earth and the luminiferous ether, Am. J. Appl. Sci. 34 (1887) 333.

[11] J. Illy, Einstein und der Eötvös-Versuch: Ein Brief Albert Einsteins an Willy Wien, Ann. Sci. 46 (1989) 417.

[12] D. Bramanti, G. Catastini, A.M. Nobili, E. Polacco, E. Rossi, R. Vergara Caffarelli, Galileo and the universality of free fall, in: STEP Symposium, ESA WPP-115, 1993, p. 139.

[13] G. Galilei, Discorsi e dimostrazioni matematiche intorno a due nuove scienze, in: Edizione Nazionale delle Opere, vol. VIII, Barbera, Firenze, 1968.

[14] F. Fuligni, V. Iafolla, Galileo and the equivalence principle, in: STEP Symposium, ESA WPP-115, 1993, p. 104.

[15] R.V. Eötvös, Über die Anziehung der Erde auf verschiedene Substanzen, Math. Nat.wiss. Ber. Ung. 8 (1890) 65;
see also R.V. Eötvös, On the gravitation produced by the Earth on different substances, Abraham Zelmanov J. 1 (2008) 6.

[16] C. Runge, Nachrichten der der Königlichen Gesellschaft der Wissenschaften zu Göttingen, vol. 1, Weidmann, Berlin, 1909, pp. 37–41.

[17] R.V. Eötvös, D. Pekár, E. Fekete, Beiträge zum Gesetze der Proportionalität von Trägheit und Gravität, Ann. Phys. 373 (1922) 11–66.

[18] A. Einstein, Zum Relativitätsproblem, Scientia 15 (1914) 337–348; in The Collected Papers of Albert Einstein vol. 4, Doc. 31, p. 306.

[19] P.R. Saulson, Thermal noise in mechanical experiments, Phys. Rev. D 42 (1990) 2437–2445.

[20] R. Pegna, A.M. Nobili, M. Shao, S.G. Turyshev, G. Catastini, A. Anselmi, R. Spero, S. Doravari, G.L. Comandi, A. De Michele, Abatement of thermal noise due to internal damping in 2D oscillators with rapidly rotating test masses, Phys. Rev. Lett. 107 (2011) 200801.

[21] P.G. Roll, R. Krotkov, R.H. Dicke, The equivalence of passive and gravitational mass, Ann. Phys. 26 (1964) 442–517.

[22] V.B. Braginsky, V.I. Panov, Verification of the equivalence of inertial and gravitational mass, Sov. Phys. JEPT 34 (1972) 463–466.

[23] V.B. Braginsky, A.B. Manukin, Measurement of Weak Forces in Physics Experiments, University of Chicago Press, Chicago, 1977.

[24] E. Fischbach, D. Sudarsky, A. Szafer, C. Talmadge, S.H. Aronson, Reanalysis of the Eötvös experiment, Phys. Rev. Lett. 56 (1986) 1427.

[25] Jia-Hao Xu, Cheng-Gang Shao, Jie Luo, Qi Liu, Lin Zhu, Hui-Hui Zhao, Effect of gravity gradient in weak equivalence principle test, Chin. Phys. B 26 (2017) 080401.

[26] E.G. Adelberger, et al., Torsion balance experiments: a low-energy frontier of particle physics, Prog. Part. Nucl. Phys. 62 (2009) 102.

[27] A.M. Nobili, et al., "Galileo Galilei" (GG) Phase A2, Study Report, Agenzia Spaziale Italiana (ASI), 2009.

[28] M.A. Hohensee, H. Mueller, R.B. Wiringa, Equivalence principle and bound kinetic energy, Phys. Rev. Lett. 111 (2013) 151102.

[29] A.M. Nobili, G.L. Comandi, D. Bramanti, Suresh Doravari, D.M. Lucchesi, F. Maccarrone, Limitations to testing the equivalence principle with satellite laser ranging, Gen. Relativ. Gravit. 40 (2008) 1533.

[30] J.P. Blaser, Can the equivalence principle be tested with freely orbiting masses?, Class. Quantum Gravity 18 (2001) 2509.

[31] A. Milani, A.M. Nobili, P. Farinella, Non-Gravitational Perturbations and Satellite Geodesy, Adam Hilger Ltd., Bristol, 1987.

[32] P.C.C. Freire, M. Kramer, N. Wex, Tests of the universality of free fall for strongly self-gravitating bodies with radio pulsars, Class. Quantum Gravity 29 (2012) 184007.

[33] Lijing Shao, Testing the strong equivalence principle with the triple pulsar PSRJ0337 + 1715, Phys. Rev. D 93 (2016) 084023.

[34] M.J. Longo, New precision tests of the Einstein equivalence principle from SN1987A, Phys. Rev. Lett. 60 (1988) 173.

[35] L.M. Krauss, S. Tremaine, Test of the weak equivalence principle for neutrinos and photons, Phys. Rev. Lett. 60 (1988) 176.

[36] D. Krasnický, et al., Measuring $\bar{g}$ with AE$\bar{g}$IS, progress and perspectives, Int. J. Mod. Phys. Conf. Ser. 30 (2014) 1460262.

[37] T.M. Niebauer, M.P. McHugh, J.E. Faller, Galilean test for the fifth force, Phys. Rev. Lett. 59 (1987) 609.

[38] K. Kuroda, N. Mio, Test of a composition-dependent force by a free-fall interferometer, Phys. Rev. Lett. 62 (1989) 1941.

[39] S. Carusotto, V. Cavasinni, A. Mordacci, F. Perrone, E. Polacco, E. Iacopini, G. Stefanini, Test of g universality with Galileo type experiment, Phys. Rev. Lett. 69 (1992) 1722.

[40] A.M. Nobili, Testing the weak equivalence principle with macroscopic proof masses on ground and in space: a brief review, Int. J. Mod. Phys. Conf. Ser. 30 (2014) 1460254.

[41] V. Iafolla, S. Nozzoli, E.C. Lorenzini, V. Milyukov, Methodology and instrumentation for testing the weak equivalence principle in stratospheric free fall, Rev. Sci. Instrum. 69 (1998) 4146.

[42] R.D. Reasenberg, B.R. Patla, J.D. Phillips, R. Thapa, Design and characteristics of a WEP test in a sounding-rocket payload, Class. Quantum Gravity 29 (2012) 184013.

[43] Lin Zhou, Shitong Long, Biao Tang, Xi Chen, Fen Gao, Wencui Peng, Weitao Duan, Jiaqi Zhong, Zongyuan Xiong, Jin Wang, Yuanzhong Zhang, Mingsheng Zhan, Test of equivalence principle at 10^{-8} level by a dual-species double-diffraction Raman atom interferometer, Phys. Rev. Lett. 115 (2015) 013004.

[44] S. Dimopoulos, P.W. Graham, J.M. Hogan, M.A. Kasevich, Testing general relativity with atom interferometry, Phys. Rev. Lett. 98 (2007) 111102.

[45] S.M. Dickerson, J.M. Hogan, A. Sugarbaker, D.M.S. Johnson, M.A. Kasevich, Multiaxis inertial sensing with long-time point source atom interferometry, Phys. Rev. Lett. 111 (2013) 083001.

[46] STE-QUEST space mission proposal: http://sci.esa.int/ste-quest/.

[47] D.N. Aguilera, et al., STE-QUEST-test of the universality of free fall using cold atom interferometry, Class. Quantum Gravity 31 (2014) 115010.

[48] A. Bonnin, N. Zahzam, Y. Bidel, A. Bresson, Simultaneous dual-species matter-wave accelerometer, Phys. Rev. A 88 (2013) 043615.

[49] D. Schlippert, J. Hartwig, H. Albers, L.L. Richardosn, C. Schubert, A. Roura, W. Schleich, W. Ertmer, E.M. Rasel, Quantum test of the universality of free fall, Phys. Rev. Lett. 112 (2014) 203002.

[50] A.M. Nobili, Fundamental limitations to high-precision tests of the universality of free fall by dropping atoms, Phys. Rev. A 93 (2016) 023617.

[51] S. Chiow, J. Williams, N. Yu, H. Müller, Gravity-gradient suppression in space-borne atomic tests of the equivalence principle, Phys. Rev. A 95 (2017), 021603(R).

[52] A.M. Nobili, A. Anselmi, Tidal effects in experiments to test the universality of free fall and the weak equivalence principle, 2017, in preparation.

[53] Onera & CNES press release 7 December 2016.

[54] P.W. Worden Jr., C.W.F. Everitt, Test of the equivalence of gravitational and inertial mass based on cryogenic techniques, in: Proc. Int. School of Physics E. Fermi, Course LVI: Experimental Gravitation, Academic Press, New York, 1973, p. 381.

[55] P.W. Worden Jr., A Cryogenic Test of the Equivalence Principle, PhD thesis, Stanford University, Stanford, California, 1976.

[56] P.W. Worden Jr., Equivalence principle tests in Earth orbit, Acta Astronaut. 5 (1978) 27.

[57] V. Josselin, P. Touboul, R. Kielbasa, Capacitive detection scheme for space accelerometers applications, Sens. Actuators 78 (1998) 92.

[58] M. Rodrigues, Development Status of the Microscope Payload, Talk given at "Gravitation and Fundamental Physics in Space Colloquium" Les Houches, 2009.

[59] P. Touboul, The microscope mission and its uncertainty analysis, Space Sci. Rev. 148 (2009) 455.

[60] Onera & CNES press release 27 September 2016.

[61] A.M. Nobili, D. Bramanti, G.L. Comandi, R. Toncelli, E. Polacco, G. Catastini, Radiometer effect in space missions to test the equivalence principle, Phys. Rev. D 63 (2001), 101101(R).

[62] A.M. Nobili, D. Bramanti, G.L. Comandi, R. Toncelli, E. Polacco, Radiometer effect in the Microscope space mission, New Astron. 7 (2002) 521.

[63] A.M. Nobili, Do general relativistic effects limit experiments to test the universality of free fall and the weak equivalence principle?, Phys. Rev. D 94 (2016) 124047.

[64] 'Galileo Galilei' (GG) webpage, 2017. http://eotvos.dm.unipi.it.

[65] T.J. Quinn, Private communication, 1993.

[66] A.M. Nobili, D. Bramanti, E. Polacco, G. Catastini, A. Anselmi, S. Portigliotti, A. Lenti, P. di Giamberardino, S. Monaco, R. Ronchini, Evaluation of a proposed test of the weak equivalence principle using Earth-orbiting bodies in high-speed co-rotation: re-establishing the physical bases, Class. Quantum Gravity 16 (1999) 1463.

[67] G. Cagnoli, L. Gammaitoni, J. Kovalik, F. Marchesoni, M. Punturo, Low-frequency internal friction in clamped-free thin wires, Phys. Lett. A 255 (1999) 230.

[68] A.M. Nobili, D. Bramanti, E. Polacco, I.W. Roxburgh, G.L. Comandi, A. Anselmi, G. Catastini, A. Lenti, A. Severi, 'Galileo Galilei' (GG). Proposed Space Experiment to Test the Equivalence Principle and Preliminary Results from the Prototype on the Ground, NASA Document D-18925, 2000, pp. 309–327.

[69] A.M. Nobili, R. Pegna, M. Shao, S.G. Turyshev, G. Catastini, A. Anselmi, R. Spero, S. Doravari, G.L. Comandi, D.M. Lucchesi, A. De Michele, Integration time in space experiments to test the equivalence principle, Phys. Rev. D 89 (2014) 042005.

[70] M. Shao, SIM-Like Laser Gauge and Adaptation to GG, Talk given at GG Workshop, Pisa 2010.

[71] M. Pisani, G. Mana, A.M. Nobili, Design of an Interferometric Displacement Sensor with Picometer Resolution for the Galileo Galilei Mission, IEEE 978-1-4799-7569-3/15, 2015, pp. 591–595.

[72] M. Pisani, M. Zucco, A.M. Nobili, Laboratory Tests of a High-Precision Laser Interferometry Readout for the GG Experiment in Space, IEEE 978-1-4673-8292-2/16, 2016, pp. 260–265.

[73] M. Armano, et al., Sub-femto-g free fall for space-based gravitational wave observatories: LISA pathfinder results, Phys. Rev. Lett. 116 (2016) 231101.

[74] A.M. Nobili, G.L. Comandi, S. Doravari, F. Maccarrone, D. Bramanti, E. Polacco, Experimental validation of a high accuracy test of the equivalence principle with the small satellite 'Galileo Galilei', Int. J. Mod. Phys. D 16 (2007) 2259.

[75] P. Goldman, A. Muszynska, Chaotic behavior of rotor/stator systems with rubs, J. Eng. Gas Turbines Power 116 (3) (1994) 692–701.

[76] U. Ehehalt, R. Markert, Instability of unbalance excited synchronous forward whirl at rotor-stator-contact, PAMM 2 (2003) 60–61.

[77] G. Adiletta, A.R. Guido, C. Rossi, Chaotic motions of a rigid rotor in short journal bearings, Nonlinear Dyn. 10 (1996) 251–269.

[78] A.M. Nobili, D. Bramanti, G.L. Comandi, R. Toncelli, E. Polacco, M.L. Chiofalo, 'Galileo Galilei-GG': design, requirements, error budget and significance of the ground prototype, Phys. Lett. A 318 (2003) 172.

[79] GG Preliminary Requirements Review Data Package (GG-PRR-Data-Pack), ASI-Thales Alenia Space Contract No. I/039/08/0 Studio di Fase A2 della Missione Galileo Galilei-GG, 2009.

[80] G. Catastini, S. Cesare, S. De Sanctis, M. Dumontel, M. Parisch, G. Sechi, Predictions of the GOCE in-flight performance with the end-to-end system simulator, in: ESA SP-627, 2007.