Testing the Equivalence Principle in space after MICROSCOPE

Anna M. Nobili[1, 2] and Alberto Anselmi[3]
[1] *Dept. of Physics "E. Fermi", University of Pisa, Largo B. Pontecorvo 3, 56127 Pisa, Italy*
[2] *INFN-Istituto Nazionale di Fisica Nucleare, Sezione di Pisa, Largo B. Pontecorvo 3, 56127 Pisa, Italy*
[3] *Thales Alenia Space Italia, Strada Antica di Collegno 253, 10146 Torino, Italy*
(Dated: March 12, 2018)

Tests of the Weak Equivalence Principle (WEP) can reveal a new, composition dependent, force of nature or disprove many models of new physics. For the first time this test is being successfully carried out in space by the MICROSCOPE satellite. Early results show no violation of the WEP sourced by the Earth for Pt and Ti test masses with random errors (after 8.26 d of integration time) of about 1 part in 10^{14}, and systematic errors of the same magnitude. This result improves by about 10 times over the best ground tests with rotating torsion balances despite 70 times less sensitivity to differential accelerations, thanks to the much stronger driving signal in orbit. The measurement is limited by a thermal noise higher than expected due to the poor quality factor of the gold wires used for electrical grounding. This noise was shown to decrease when the spacecraft was set to rotate faster than planned. The result will improve by the end of the mission, as thermal noise decreases with more data. Not so systematic errors. We investigate major non-gravitational effects and find that MICROSCOPE's "zero-check" sensor, with test masses both made of Pt, does not, in reality, allow their separation from the signal. The early test itself reports systematics in the Pt-Ti sensor which are not detected in the Pt-Pt one, hence would not be distinguished from a violation. The improved test will need more measurements to check systematics, but there is not enough time left. MICROSCOPE demonstrates the huge potential of space for WEP tests of very high precision and indicates how to reach it. To realize the potential, a new experiment needs the spacecraft to be in rapid, stable rotation around the symmetry axis (by conservation of angular momentum), needs high quality state-of-the-art mechanical suspensions as in the most precise gravitational experiments on ground, and must allow multiple checks to discriminate a violation signal from systematic errors. The design of the "Galileo Galilei" (GG) experiment, aiming to test the WEP to 1 part in 10^{17} and currently a candidate for a medium-sized mission of the European Space Agency (ESA), unites all the needed features.

I. INTRODUCTION

The General theory of Relativity (GR) 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 is at the crossroads of the open problems of fundamental physics: the relation of quantum fields and gravitation; the nature of dark matter and dark energy; the absolute character of the fundamental constants of physics. Tests of the WEP provide severe constraints to "new physics" attempting to cross the gap between GR and the Standard Model of particle physics, or make sense of dark matter and dark energy.

The dimensionless Eötvös [1] parameter

$$\eta_{_{Eötvös}} = \frac{\Delta a}{a} \tag{1}$$

quantifies the level of violation. Δa is the differential acceleration measured between two test masses of different composition as they fall in the field of a source body with average acceleration a (the so called "driving signal") .

A reliable measurement of $\eta_{_{Eötvös}} \neq 0$ would amount to the discovery of a new long-range composition dependent force of nature and make a revolution in physics; the higher the precision of the test, the higher the chances to find new physics. Conversely, the more sensitive the test yielding $\eta_{_{Eötvös}} = 0$ the greater the fine tuning required for many physical models and theories to survive.

The best experiments, carried out by the Eöt-Wash group with slowly Rotating Torsion Balances (RTB), have established that there is no violation to about 1 part in 10^{13} [2, 3]. While some improvement is still possible, gaining orders of magnitude requires moving the experiment to a laboratory in space (see [4] and references therein). An experiment to test the WEP in orbit, named STEP, has been studied since the 1970s [5–7]; in the 1990s, following the interest raised by a re-analysis of the Eötvös experiment [8], ESA and NASA have investigated the mission in considerable detail with the goal of testing the equivalence principle to 10^{-17} [9].

For the first time an equivalence principle test is carried out with test masses in low Earth orbit, weakly suspended inside the MICROSCOPE spacecraft aiming to reach 10^{-15} [10]. With the signal at a few mHz, MICROSCOPE scientists report for a Pt-Ti composition dipole a null result relative to the Earth with a random noise on the Eötvös parameter $\eta_{\oplus}$ of about 10^{-14} (after an integration time of 8.26 d), and systematic errors at the same level [11].

In the field of the Earth MICROSCOPE's early result is a 10-fold improvement over rotating torsion balances. The improvement occurs despite a sensitivity to differential accelerations about 70 times worse than RTB at similar frequency. In favour of the space test is the much larger driving signal from Earth –the average free fall acceleration at the denominator of (1)– by almost 500 times at low low altitude as compared to RTB on ground ($\sim 8\,\mathrm{ms^{-2}}$ versus $0.0169\,\mathrm{ms^{-2}}$ at most) [4]. Having RTB superseded mass dropping tests by several orders of magnitude, despite a driving signal almost 600 times weaker ($\lesssim 0.0169\,\mathrm{ms^{-2}}$ versus $9.8\,\mathrm{ms^{-2}}$) the very large factor yet to gain in low Earth orbit and the success of MICROSCOPE strongly indicate that the next leaps in precision tests of the WEP shall occur in space.

MICROSCOPE's measurement is limited by thermal noise due to internal damping in the gold wires, one for each test cylinder, used to connect it to the enclosure in order to ensure electric grounding, which is a serious issue for small force gravitational experiments, especially in orbit. Each cylinder is actively controlled by electrostatic forces (electrostatic suspensions act as a negative spring, hence un-controlled cylinders would be unstable [12]) and the readout is capacitive too, while the connection to the cage by means of the gold wire is meant to have only an ancillary role. Losses in the gold wires turn out to be the limiting factor. The quality factor of the wire, as measured in the lab [13] is 100 times worse, at similar $\sim$ mHz frequency, than that of the torsion fiber of the Eöt-Wash balances [2, 3].

Thermal noise due to internal damping in the suspensions of gravitational wave detectors has been demonstrated to decrease with the frequency as $1/\sqrt{\nu}$ [14], which is particularly important for these detectors whose target signals give rise to extremely small displacements of $\sim 10^{-19}$ m at frequencies above several tens of Hz. Instead, a signal of WEP violation with the Earth as source would be DC on ground and at orbital frequency in space. A way to increase this frequency is by rotating the sensor relative to the Earth, the faster the better. Limitations to the spin rate of RTB come from concerns about rotation noise (on ground it includes motor and bearings noise) and the attenuation of the signal strength at frequencies above the natural oscillation mode (the system being in essence a forced oscillator [15]). With a natural torsional frequency $\nu_{tor} = \frac{1}{798}$ Hz, the highest spin rate so far is $\nu_{spinRTB} = \frac{2}{3}\nu_{tor} \simeq 0.84$ mHz [3].

In MICROSCOPE each test cylinder is sensitive in 1D, along its symmetry axis. Hence, rotation relative to the Earth must occur around an axis perpendicular to the symmetry axis ([11], Fig. 1), which is the only stable axis against small perturbations. Therefore rotation was planned to be slow, below $5\nu_{orb}$ [10], but it has been raised in order to reduce thermal noise since it turned out to be higher than expected. The result reported in [11] has been obtained at $\nu_{spin} = 17.5\nu_{orb} \simeq 2.94$ mHz, and this is the current baseline.

It is the first demonstration of a high precision rotating experiment in space. Rotation of the whole spacecraft relative to inertial space, with no stator and no bearings, has very low noise and is the key to the mission success.

MICROSCOPE scientists are confident that by the end of the mission, with more data available, thermal noise –being random– will be reduced and allow a WEP test closer to the 10^{-15} original target of the mission. Currently reported systematic errors are about 10 times larger than that, and they will not disappear or decrease simply with more data. Hence, all systematics that were to emerge above random noise shall require very careful checking in order to be separated with certainty from a possible violation signal.

For this purpose MICROSCOPE carries a second, "zero-check" equal composition accelerometer (named SUREF) with the test cylinders both made of Pt. Ideally, a violation signal should appear in the Pt-Ti sensor and not in the Pt-Pt one, while systematic effects due to known physics should be detected by both sensors.

In this work we compare random noise and non-gravitational systematic effects in the two sensors showing that –with the given geometry and masses of the test bodies, and even in the simplified assumption of identical physical conditions of the two sensors– the expected separation of the violation signal from systematic errors does not occur. The lower sensitivity of the Pt-Pt zero-check sensor to non-gravitational effects proportional to the area-to-mass ratio of each test body (by 3.6 times) could explain why the systematic effects reported in the early test do not show up in this sensor, despite its lower random noise ([11], Fig. 3). Not being sensitive enough to a wide class of systematic effects, the zero-check sensor cannot, in fact, discriminate a violation signal from disturbances due to known physics.

We conclude that a thorough check of systematic errors cannot be avoided for the experimental result to be reliable. It requires a sufficient number of measurements, all at the same precision, to be carried out in different physical conditions such that the different physical parameters involved allow the signal to be distinguished from systematics on the basis of their different signature, hence different dependence on these parameters.

MICROSCOPE scientists planned to reach the mission target $\eta_{\oplus} = 10^{-15}$ with an integration time corresponding to 120 orbits (roughly 8 days), so that in a 2-yr mission duration there would be many such measurements in different experimental conditions, making it possible to check the result and possibly even improve it. In their own words [10]: "The adopted trade-off remains on different sessions of 120 orbital periods. This is long enough to obtain the Eötvös parameter target exactitude of 10^{-15} in inertial mode and even better in rotating mode, by reducing the stochastic error with respect to the systematic evaluated one. This is also short enough to have time for many sessions with different experimental conditions"

It now turns out that in 120 orbits the measurement is about a factor of 10 short of the target; and the mission will last less than planned because of the higher rotation rate required, which –in the MICROSCOPE design– means a higher consumption of propellant.

Being at present more precise than any ground test, the only way to resolve the ambiguity between systematic errors and a possible violation signal is by flying another experiment with higher precision and a shorter integration time so that systematic errors can be thoroughly checked and a violation signal (if any) identified beyond question. A new mission will benefit from many lessons that can be learned from MICROSCOPE already at this point.

On ground RTB have achieved precisions orders of magnitude better than mass dropping tests (with bulk masses or cold atoms). The gain over ground balances by an additional factor of 500 with suspended masses in low Earth orbit, exploited by MICROSCOPE for the first time, confirms that suspended masses are the right choice (versus drop tests) to reach very high precision. If care is taken in flying an experiment somewhat more sensitive than ground balances, the improvement achievable in space can be impressive.

We have long been involved in the development of the "Galileo Galilei" (GG) project for an experiment in space to test the equivalence principle to 10^{-17}. Rotation of the spacecraft and high quality factors, which MICROSCOPE finds to be pivotal in reducing thermal noise and integration time, are the main drivers of the GG experiment to reach very high precision without invoking cryogenics ([16–18], [4]). GG is the shortlist of candidates to the next medium size mission of ESA, awaiting the decision of the Agency.

The paper is organized as follows.

In Sec. II we present the early MICROSCOPE results, compare them with RTB, show how faster rotation has reduced random noise due to the poor quality factor of the gold wires, discuss the reported systematic errors and their current limited understanding.

In Sec. III we quantitatively compare the effects of thermal noise from internal damping and of major systematic errors in the two sensors, to find that at a precision closer to the 10^{-15} target of MICROSCOPE the Pt-Pt sensor will not allow a violation signal to be separated out. With not enough time left to confirm or rule out a violation, another experiment in space is needed; in Sec. IV we argue that MICROSCOPE itself, through its success and limitations, shows that a much more precise test of the WEP in orbit is possible and points out the key changes to be made in order to achieve it.

In Sec. V we draw the conclusions.

II. MICROSCOPE FIRST TEST OF THE EQUIVALENCE PRINCIPLE IN SPACE: SUCCESS, LIMITATIONS AND OPEN ISSUES

While MICROSCOPE is still in orbit taking science data, early results with an integration time of 8.26 d yield, for Earth as the source body and test masses made of Pt (with 10% of Rh) and Ti (with 10% of Al), a null result at 1σ level of [11]:

$$\eta_{\oplus}(\mathrm{Pt},\mathrm{Ti}) = [-1 \pm 9(\mathrm{stat}) \pm 9(\mathrm{syst})] \times 10^{-15} \quad . \tag{2}$$

It has been obtained with the spacecraft in low Earth orbit at frequency $\nu_{orb} = 0.16818\,\mathrm{mHz}$ ($P_{orb} = 5946\,\mathrm{s}$) rotating at $\nu_{spin} = 2.9432\,\mathrm{mHz}$ ($P_{spin} = 339.8\,\mathrm{s}$), whereby a WEP violation signal would occur at $\nu_{_{EP}} = \nu_{spin} + \nu_{orb} = 3.1113\,\mathrm{mHz}$ ($P_{_{EP}} = 321.4\,\mathrm{s}$).

By comparison with the best tests of WEP achieved on ground by RTB [2, 3] this is a 10-fold improvement. The improvement occurs with a sensitivity to differential accelerations, at the signal frequency, of:

$$\Delta a_{Pt-Ti} \simeq 9 \times 10^{-15} g(h) \simeq 7.1 \times 10^{-14}\,\mathrm{ms}^{-2} \ , \tag{3}$$

where $g(h) \simeq 7.9\,\mathrm{ms}^{-2}$ is the average gravitational acceleration from Earth at 710 km altitude. Instead, RTB are sensitive to $\Delta a = 10^{-15}\,\mathrm{ms}^{-2}$ at a signal frequency of 0.84 mHz for both the Be-Ti and Be-Al composition dipoles tested ([2, 3], Table 3). Despite 70 times less sensitivity, MICROSCOPE's early test is 10 times better thanks to the larger driving signal from Earth in orbit versus RTB on ground [4].

There is no such gain over RTB if data are analyzed taking the Sun, or the dark matter in our galaxy, as source bodies of a possible WEP violation (instead of the Earth). In this case the gravitational and inertial forces which are being compared are the gravitational attraction from the source body (either the Sun or dark matter at the center of the galaxy) and the centrifugal force that keeps the test masses in orbit around them, and there is no larger driving signal in low Earth orbit. The best null results in the field of the Sun and of dark matter in our galaxy have been established by RTB at $\eta_{\odot}$ and $\eta_{_{DM}}$ of a few times 10^{-13} and a few times 10^{-5} respectively ([3], Table 3). No such analysis has been published for MICROSCOPE, but there can be no improvement with a sensitivity to differential accelerations worse than RTB.

A WEP experiment in space must face a huge effect which is absent on ground, due to air drag and solar radiation pressure acting on the outer surface of the spacecraft. It gives rise to an equal and opposite inertial acceleration on every test mass weakly suspended inside the spacecraft, which is many orders of magnitude smaller than 1-g, but also many orders of magnitude bigger than the target violation signal. The effect is common mode in principle, hence, ideally, it gives no differential acceleration competing with the signal. In real experiments there is, in fact, a differential residual which may be relevant because the effect is huge; moreover, its main component is at the same frequency as the signal. For MICROSCOPE the inertial acceleration resulting from drag is roughly 7 orders of magnitude bigger than the signal.

At 1-g the torsion balance can reach a relative precision of 1 part in 10^{13} in the differential effect of WEP violation having been built with tolerances of only 1 part in 10^5, thanks to its capability to reject common mode effects. In MICROSCOPE the test cylinders are suspended individually (they do not form a balance) and their configuration is frozen after assembling. Any difference resulting from construction and mounting errors can only be mitigated (rejected) by in-flight calibrations of their individual responses. Matching by 8.5×10^{-3} (rejection factor of about 118) is reported in [11]. Such a low level of rejection means that in order to achieve the result (2) most of the drag acceleration has been successfully compensated by drag-free control of the spacecraft with propellant and thrusters. If combined with a comparable rejection level by means of an appropriate design of the sensor itself, it would allow a much higher precision to be reached without stringent requirements on thruster noise which are hard to meet [19].

In MICROSCOPE each test cylinder is weakly suspended along its symmetry axis by means of electrostatic forces, with $600\,\mu$m gap. Even after compensation of the large effect of drag, a restoring force is needed in response to small residual forces in order to prevent each cylinder from hitting the cage. Since electrostatic suspensions are unstable (they act as a "negative" spring [12]), a restoring force –including the response to the violation signal, if any– is provided by active electrostatic control.

It is a tradition for this type of accelerometers developed in France at ONERA to add a physical connection between the test mass and the cage by means of a loose, thin, conducting wire made of gold ($7\,\mu$m width and 2.5 cm length in MICROSCOPE [20]), primarily to avoid electric charging.

Concern about thermal noise from internal damping in the wire at the low frequencies of interest has led to extensive measurements of its quality factor Q under realistic flight conditions. An *ad hoc*, electrostatically suspended torsion pendulum has been built in order to avoid the suspension wire and thus achieve a very weak torsional constant [21, 22]. The electrostatic pendulum is in fact 10 times stiffer than the mechanical torsion pendulum of the Eöt-Wash group [2]. It has measured the Q of the gold wire ($7.5\,\mu$m width and 1.7 cm length in this case) at frequencies ranging from about 10^{-4} Hz to several 10^{-2} Hz, showing the presence of large losses, with better Q occurring at higher frequencies. The values measured range from $Q = 36$ slightly above 10^{-4} Hz (the orbital frequency) to 59 at 10^{-3} Hz, while Q values around 110 are measured only at frequencies of 10^{-2} Hz and a few times 10^{-2} Hz([13], Fig. 5).

These losses are much higher than in the suspensions of small force ground experiments. At a signal frequency slightly less than a mHz, the Eöt-Wash group reports a quality factor 100 times better than measured by ONERA at 1 mHz, of about 6000, with a $20\,\mu$m W wire suspending a 70 gram balance at 1-g [2, 3]. Ways are known to avoid large losses also at low frequencies, especially for suspensions to be used in weightlessness conditions in which even large masses need very low stiffness. Monolithic suspensions, manufactured from a single block (to avoid relaxation of bending energy) with enlarged ends (to ensure that clamping is located far from where the flexure undergoes deformation during motion), and with appropriate heat treatments, are commonplace in small force gravitational experiments and have low losses. Instead, a gold wire clamped with droplets of glue at its ends, where most dissipation occurs, is bound to yield large losses. Moreover, losses will be unequal even if all wires are taken from the same coil and assumed to be perfectly identical, because the procedure used for clamping is hardly repeatable.

It has been known since 1990 [14] that thermal noise from internal damping in the suspensions of the test masses in gravitational wave detectors decreases with the frequency as $1/\sqrt{\nu}$. This is how the Virgo/LIGO detectors around 100 Hz can be sensitive to displacements of the mechanically suspended mirrors as small as about 10^{-19} m [23]. A signal of WEP violation is at a much lower frequency: with Earth as the source body, the violation signal would be DC on ground and at orbital frequency in space. For this reason rotation of the apparatus is used, as for torsion balances, to up-convert the signal to higher frequencies, the higher the better.

In MICROSCOPE rotation occurs perpendicularly to the orbit plane and to the symmetry axis ([11], Fig. 1), which however is the only stable axis against small perturbations. For this reason a slow rotation mode was planned with $3\nu_{orb} < \nu_{spin} < 5\nu_{orb}$ [10]. Once in orbit, thermal noise turned out to be higher than expected; being dominated by internal damping, the cure was to reduce it by increasing the rotation rate of the spacecraft. The result (2) was obtained while spinning at $\nu_{spin} = 17.5\nu_{ord} \simeq 2.94$ mHz, which is 3.5 times faster than the maximum spin rate planned before launch. The possibility –unique to space– of rotating the whole spacecraft, with no stator and no bearings, has proved to be much less noisy than rotating experiments in ground laboratories. The original caution was abandoned and the faster rotation rate has been adopted as the new baseline, despite the higher consumption of propellant and consequent shorter duration of the mission.

At a recent conference on "Fundamental Physics in Space" in Bremen MICROSCOPE scientists have reported that when up-converting the signal frequency by increasing the rotation rate of the spacecraft, thermal noise in the Pt-Ti accelerometer improves more than expected according to the $1/\sqrt{\nu}$ dependence [20]. The measurements show that while the signal frequency increases by a factor 3.53, hence thermal noise from internal damping is expected to decrease by $\sqrt{3.53} = 1.88$, it is instead found to decrease by 3.61 times, with an unexplained (favourable) factor of 1.9. We notice that the acceleration due to thermal noise from internal damping depends on the quality factor as $1/\sqrt{Q}$, and the quality factor too has been found to depend on the frequency, usually being higher (lower losses) at higher frequencies. If so, the unexplained improvement by a factor 1.9 might be due to an increase of the quality factor of the gold wires by $1.9^2 = 3.6$ times for the same system in the same conditions, except for the fact that the frequency of the signal has increased (with faster rotation) by 3.53 times. It is quite interesting that a similar improvement has been observed with the ground demonstrator of the proposed GG experment in space: with a frequency increase by

2.16 times, the quality factor (for the same system, except for rotation), was found to increase by 2.24 times ([4], Sec. 8).

The systematic errors shown in (2) have been found to be mostly of thermal origin. The electronics unit and the base plate of the Pt-Ti sensor were subjected to artificially produced thermal variations at the signal frequency, and the resulting differential acceleration between the two cylinders was measured, thus mapping the sensitivity of the instrument to thermal effects. Effects due to thermal variations of the base plate turned out to dominate over those of the electronics unit. But they were larger than expected by more than two orders of magnitude per degree of temperature variation, and the reasons for such behaviour are not known yet. At the same time, by measuring the actual thermal variations (aside from those induced for the sensitivity test), it turned out that they were smaller than expected also by about two orders of magnitude [11]! Both findings call for a convincing physical explanation which may require more information, for instance on the residual pressure.

We note that a rotation rate faster than planned may be responsible for a somewhat better thermal stability, because of a better averaging and also because the signal frequency is farther away from the orbital frequency, at which most of the thermal stress obviously occurs.

III. THERMAL NOISE, SYSTEMATIC ERRORS AND THE “ZERO-CHECK” SENSOR

In addition to the composition dipole SUEP, whose inner (denser) cylinder is made of Pt (with 10% of Rh) and the outer one of Ti (with 10% of Al), MICROSCOPE carries a second sensor, named SUREF, with the same geometry (and the same 600 μm gaps) as SUEP but cylinders made of the same material. The inner cylinder is “identical” to the inner one in SUEP, being made of the same Pt-Rh alloy (the two masses differ only by a few parts in 10^4); the outer cylinder has the same size and volume as the outer one in SUEP, but is made of Pt-Rh alloy like the inner one. All densities are uniform.

The Pt-Pt SUREF sensor should allow systematic errors to be distinguished from a violation signal (“zero-check” sensor): a violation signal –being composition dependent– should appear in SUEP but not in SUREF, while systematic errors should appear in both sensors.

The two sensors are located 17.5 cm away from each other, and none of them is centered on the center of mass of the spacecraft. The drag-free control loop is closed either on one sensor or on the other, with the task of partially compensating the common mode motion, relative to the spacecraft, of the two selected cylinders due to the inertial acceleration (nominally the same for all test bodies) resulting from the effect of drag acting on the outer surface of the spacecraft. Depending on which sensor drives the drag-free control, science data are collected for that sensor only. Hence, SUEP and SUREF never take data simultaneously, which weakens the role of SUREF as zero-check sensor and reduces the integration time available for SUEP to test the WEP.

MICROSCOPE scientists expect, by the end of the mission, to accumulate sufficient data for the random noise of SUEP shown in (2) to reduce to a value close to the 10^{-15} target of the mission. Then, currently reported systematic errors would emerge above noise and call for a full understanding of their physical origin, as they might contain a violation signal.

The Pt-Pt SUREF sensor should detect only systematic effects due to known physics, not a violation signal, and thus solve the problem. In reality, things are not as clear-cut as they appear at a first glance, because SUREF detection of systematic errors depends on its thermal noise (only systematics larger than thermal noise will be detected) and on its own sensitivity to systematics. If it turns out to be less sensitive than SUEP to some systematics, it cannot rule them out completely as possible violation of the WEP.

The measurements are limited by thermal noise from internal damping in the gold wires. At the frequency of the violation signal ν_{EP} the spectral density of the acceleration noise of each cylinder (expressed in $\mathrm{ms}^{-2}/\sqrt{\mathrm{Hz}}$ in SI units) reads [14]:

$$\hat{a}_{thID} = \frac{1}{\mathcal{M}}\sqrt{\frac{4K_B T k_w}{Q_w \omega_{EP}}} \tag{4}$$

where K_B is the Boltzmann constant, T is the equilibrium temperature, $\mathcal{M}$ is the mass of the test cylinder, k_w and Q_w are the stiffness and quality factor of the gold wire connecting it to the cage, and $\omega_{EP} = 2\pi\nu_{EP}$ is the frequency of the signal. Being random noise, and most probably uncorrelated between the inner and outer cylinder in each sensor, the resulting differential acceleration noise competing with the signal is:

$$\hat{\Delta} a_{thID} = \sqrt{\hat{a}^2_{thIDinner} + \hat{a}^2_{thIDouter}} \quad . \tag{5}$$

Assuming the same ambient temperature T in the two sensors and the same k_w and Q_w for all gold wires (even at different frequencies), the ratio of the differential acceleration noise between SUEP and SUREF is:

$$\frac{\hat{\Delta} a_{thID-SUEP}}{\hat{\Delta} a_{thID-SUREF}} = \\ = \sqrt{\frac{\omega_{EP-SUREF}}{\omega_{EP-SUEP}}} \cdot \frac{\sqrt{\frac{1}{\mathcal{M}^2_{inner-SUEP}} + \frac{1}{\mathcal{M}^2_{outer-SUEP}}}}{\sqrt{\frac{1}{\mathcal{M}^2_{inner-SUREF}} + \frac{1}{\mathcal{M}^2_{outer-SUREF}}}} \tag{6}$$

showing that it depends only on the different masses of the individual cylinders and on the different ratio between the frequencies of the signal, which in turn depends on the different rotation frequency during the respective measurements. The masses cannot be changed once in orbit, and their contribution to the noise ratio (6) is 1.5989 (they are measured very precisely). We must therefore expect a higher thermal noise in SUEP than in SUREF by about 1.6, only because of the different values of the masses. The recent measurements for SUEP and SUREF have been made at different rotation rates of the spacecraft, hence they refer to different signal frequencies whose ratio is $3.1113\,\mathrm{mHz}/0.9250\,\mathrm{mHz} = 3.3636$ [11], contributing to the noise ratio (6) by $1/\sqrt{3.3636} = 1/1.8340$. Overall we get:

$$\frac{\hat{\Delta a}_{thID-SUEP}}{\hat{\Delta a}_{thID-SUREF}} = \frac{1.5989}{1.8340} \simeq 0.87 \tag{7}$$

which means that, at the selected rotation frequencies and with the assumptions made we should expect, in the differential acceleration noise competing with the signal, a slightly lower noise in SUEP than in SUREF.

To the contrary, the measured values reported in [11] are $5.6 \times 10^{-11}\,\mathrm{ms}^{-2}/\sqrt{\mathrm{Hz}}$ for SUEP and $1.8 \times 10^{-11}\,\mathrm{ms}^{-2}/\sqrt{\mathrm{Hz}}$ for SUREF, showing that SUEP is in fact 3.11 times more noisy. This means that there is an unexplained factor of about 3.57, SUEP being 3.57 times more noisy than SUREF than one would expect. For one thing this is good news, because the fact that SUREF is less noisy than SUEP helps in discriminating systematics, since it improves the signal-to-noise ratio with which they are detected in SUREF. However, not understanding the physics of the limiting noise in the sensor of a high precision experiment is a problem.

Since the temperature is well measured, only differences in the ratio k_w/Q_w for the test cylinders can be invoked, at least as long as random noise is due to internal damping as in (4). Since all four wires have the same length and are cut from the same coil, while Q_w depends mostly on the glue clamping at the two ends of each wire –which are unpredictable and hardly repeatable– Q_w is more likely to be responsible for the observed discrepancy. It appears in (4) under the square root, hence the Q_w which dominates thermal noise in SUEP should be almost $3.57^2 \simeq 13$ times smaller than the one which dominates in SUREF.

On this issue it is worth recalling that accelerometers similar, in their key features, to those of MICROSCOPE, also built by ONERA, have successfully flown onboard the GOCE geodesy mission of the European Space Agency and a noise level about two times larger than expected has been reported in that case [24–26].

Concerning systematic effects at the frequency of the violation signal, and the respective sensitivities of SUEP and SUREF, we notice the following. The systematic errors which limit the Eötvös parameter (2) as measured with SUEP are depicted in [11], Fig. 3 left plot (and listed in Table III of the paper) at the level of about $7 \times 10^{-14}\,\mathrm{ms}^{-2}$ as function of time with the number N of orbits (120 in total). The same plot shows the random acceleration noise, which instead decrease as $1/\sqrt{N}$ to meet, towards the end of the run, the horizontal line of systematic errors. The same Figure (right plot) shows the (lower) random noise in SUREF, also decreasing as $1/\sqrt{N}$ over a total of 62 orbits, to reach slightly below $3\times 10^{-14}\,\mathrm{ms}^{-2}$ at the end of the run. Should there be systematic effects at the same level as in SUEP, they would clearly appear above random noise in SUREF, but no such systematics are detected. They don't appear either in the spectral density of the acceleration noise of SUREF shown in [11], Fig. 2 right plot. The question as to why it is so is obviously a very relevant one, because systematic errors should be detected by both SUEP and SUREF is order to be distinguished from a violation signal.

The systematic errors reported in SUEP and not detected in SUREF are non-gravitational. We therefore compare the sensitivity of SUEP and SUREF to non-gravitational perturbations. As shown in (1), in WEP tests the physical observable is the differential acceleration of the test masses relative to the source body, hence, the relevant quantities are accelerations, not forces. The accelerations of a number of non-gravitational perturbations are known to be proportional to the area-to-mass ratio of the affected body [27], the area being in this case that of the cross section of the test cylinder perpendicular to its sensitive/symmetry axis. The differential acceleration between the test cylinders in SUEP caused by such non-gravitational perturbation would be:

$$\begin{aligned} &\Delta a_{ng\mathcal{A}/\mathcal{M}-SUEP} = \\ &= a_{ng\mathcal{A}/\mathcal{M}outer-SUEP} - a_{ng\mathcal{A}/\mathcal{M}inner-SUEP} \propto \\ &\propto (\mathcal{A}/\mathcal{M})_{outerSUEP} - (\mathcal{A}/\mathcal{M})_{innerSUEP} \end{aligned} \tag{8}$$

where $(\mathcal{A}/\mathcal{M})$ is the area-to-mass ratio for the test cylinder referred to in the subscript; and similarly for SUREF. If the physical parameters which determine the non-gravitational perturbation under consideration are the same in both sensors, the ratio of the differential accelerations it gives rise to depends only on the ratios of $(\mathcal{A}/\mathcal{M})$ for the test cylinders in the two sensors:

$$\begin{aligned} &\frac{\Delta a_{ng\mathcal{A}/\mathcal{M}-SUEP}}{\Delta a_{ng\mathcal{A}/\mathcal{M}-SUREF}} = \\ &= \frac{(\mathcal{A}/\mathcal{M})_{innerSUEP}}{(\mathcal{A}/\mathcal{M})_{innerSUREF}} \cdot \\ &\cdot \frac{(\mathcal{A}/\mathcal{M})_{outerSUEP}/(\mathcal{A}/\mathcal{M})_{innerSUEP} - 1}{(\mathcal{A}/\mathcal{M})_{outerSUREF}/(\mathcal{A}/\mathcal{M})_{innerSUREF} - 1} \quad . \end{aligned} \tag{9}$$

With data available on the masses and the geometry of the test cylinders [11, 28, 29] we get for this ratio (in modulus) about 3.3. Thus, non-gravitational perturbations whose accelerations are proportional to the area-to-mass ratio of the test cylinders, give rise to differential accelerations 3.3 times larger in SUEP than in SUREF, simply because of the way they have been designed, all other physical parameters being the same.

Since the systematic effects reported in SUEP at the frequency of the signal amount to $7 \times 10^{-14}\,\mathrm{ms}^{-2}$, we should expect systematics in SUREF to be a factor 3.3

smaller, at about $2 \times 10^{-14}\,\mathrm{ms}^{-2}$. Being below the thermal noise measured in SUREF ([11], Fig. 3 right plot) it is not surprising that they are not detected.

This fact questions the use of the zero-check Pt-Pt SUREF sensor to discriminate a violation signal from spurious effects. With a longer integration time and a lower thermal noise SUREF may allow these systematics to be detected. However, as long as their value in SUEP is several times larger than in SUREF the open issue remains that they may contain a violation signal as well.

It is worth considering also non-gravitational perturbations, such as electric charge effects, whose acceleration on a test cylinder does not involve its cross section but only its mass, being inversely proportional to it. In this case the ratio of the differential accelerations between the test cylinders in the two sensors (all other physical parameters being the same) reads:

$$\begin{aligned} \frac{\Delta a_{ng\mathcal{M}-SUEP}}{\Delta a_{ng\mathcal{M}-SUREF}} &= \\ = \frac{(\mathcal{M})_{innerSUREF}}{(\mathcal{M})_{innerSUEP}} &\cdot \\ \cdot \frac{(\mathcal{M})_{innerSUEP}/(\mathcal{M})_{outerSUEP} - 1}{(\mathcal{M})_{innerSUREF}/(\mathcal{M})_{outerSUREF} - 1} \end{aligned} \tag{10}$$

which yields 0.47 (modulus), meaning that this kind of systematic errors would be about 2 times larger in SUREF than in SUEP. This tells us that electric charge effects, if any, are below the level of thermal noise reported in SUREF. Were such spurious effect, at some point, detected above thermal noise in SUEP, the ratio (10) in favour of SUREF would not in principle prevent its separation from the signal.

An important fact to be mentioned, that questions the use of SUREF as a zero-check sensor, is related to the radiometer effect, a well known disturbance competing with the signal which is proportional to the residual pressure around the test cylinders and to the temperature gradient between the two ends of its sensitive/symmetry axis [30].

MICROSCOPE scientists exclude radiometer as the origin of the systematic effect in SUEP because of the extremely good thermal stability and uniformity observed (100 times better than expected!). A residual pressure of 10^{-5} Pa was assumed before launch [10] but no value is given for the residual pressure once in orbit. By comparison, LISA Pathfinder (LPF) finds a value of 2.2×10^{-5} Pa at the initial phase of the mission [32], and a lower value of 10^{-6} Pa at a later time, because of venting to outer space and more time available for degassing [33]. To our knowledge there is no venting to outer space in MICROSCOPE, and the getter pumps it relies upon after final assembling in order to take care of pressure increase due to outgassing surfaces, have a limited lifetime. Thus, a reliable estimate of the residual pressure both in SUEP and in SUREF is needed, especially because different values may be expected due to the fact that the outer cylinder in SUEP is the only coated one of the four, and this may result in a different outgassing as compared to the outer cylinder in SUREF, despite the same geometry.

In any case, the radiometer effect should be carefully investigated because it is known to give no differential acceleration if the two cylinders have the same density, as in the case of SUREF [31]. The acceleration due to radiometer is proportional to the area-to-mass ratio, which is different for the inner and outer cylinder, and to the temperature gradient at its ends, which is different too. However, if the cylinders are made of the same material and have the same density, and the temperature gradient per unit length is the same (as it is reasonable, because the material is the same and they are located in the same enclosure), the acceleration of each cylinder is simply inversely proportional to its density, and therefore it is the same, yielding no differential acceleration. We have checked this fact with the numbers available for the test cylinders [11, 28], and the result for the ratio of the radiometer acceleration on the outer and inner cylinder in SUREF (of length L_{outer} and L_{inner}, as in SUEP) is almost exactly 1:

$$\frac{a_{rad-outerSUREF}}{a_{rad-innerSUREF}} = \frac{(\mathcal{A}/\mathcal{M})_{outerSUREF} \cdot L_{outer}}{(\mathcal{A}/\mathcal{M})_{innerSUREF} \cdot L_{inner}} = 1.009 \tag{11}$$

while the same ratio for SUEP is 4.562. Thus, the radiometer effect gives rise to a non-zero differential acceleration in SUEP and no differential acceleration in SUREF, just as one expects for the violation signal. As suggested in [31], a way out of this impasse might have been to fabricate the Pt-Pt cylinders in SUREF with a different average density, e.g. with some appropriate empty volume in one of them, e.g. the outer one.

Although some additional information may be available (by investigating the accelerations as measured individually by each test cylinder) that could help to mitigate the difficulties outlined here, it is apparent that the main goal of the second equal composition sensor to provide a clear-cut, unquestionable, check of the violation signal versus systematic errors is not met.

The need remains for carefully designed checks of systematic errors in the different composition sensor. This requires many measurements, all at the same sensitivity (possibly the target sensitivity) in different physical conditions, such that systematic errors and violation signal can be distinguished on the basis of their respective signature and consequent different dependence on the physical parameters involved in these measurements, as it is done in ground tests of the WEP with RTB.

The additional complexity and cost of carrying a second sensor should rather be faced for flying two different composition dipoles instead of one. As argued in [2, 3], their measurements can both be analyzed not only in the field of the Earth but also of the Sun and of dark matter at the center of our galaxy, thus avoiding an accidental cancellation of the charges of the test-body dipole or the attractor. This would increase the chance of finding a non-null result and strengthen its physical significance.

IV. LESSONS FROM MICROSCOPE AND ROOM FOR MAJOR IMPROVEMENTS

With the zero-check sensor unable to firmly discriminate systematic errors, the final MICROSCOPE test of the WEP will have to rely on standard procedures of systematic error checks in the different composition sensor. These checks have been envisaged before launch [10] when the expectations were to achieve the target precision in 120 orbits, even in inertial mode, i.e. at zero spin rate, to be possibly improved in rotation mode. During the mission lifetime many such 120-orbit runs would be available in "different experimental conditions", as the authors rightly stress, allowing separation of systematic errors from a violation signal.

It turns out that even in rotation mode, and at a faster rate than the maximum planned, the level of noise over 120 orbits is higher than expected, and the entire mission duration is necessary to reduce it and bring the precision of the WEP test closer to the 10^{-15} target of the mission. This means that there will be no time left to check the systematic errors reported in the early test (2), which are not expected to disappear with a longer integration time.

We shall therefore be left with the most sensitive test of the WEP ever, but no firm conclusion as to whether the equivalence principle is violated or not. Only another experiment in space, with higher precision, shorter integration time and consequent reliable systematic checks, could give the answer. The success of MICROSCOPE, together with its limiting factors, tell us that in orbit it is possible to reach a much higher precision, and give clear indications as to how to proceed in order to reach it.

High precision requires low thermal noise, which results in a short integration time. This needs high Q and high frequency of the signal to be obtained by rotating the spacecraft, the faster the better. As discussed in Sec. II, quality factors better than those of the gold wires of MICROSCOPE can be obtained, also at mHz frequencies. However, this improvement alone would not be enough to reach a sensitivity to differential accelerations better than that already achieved by RTB.

A solution often advocated for MICROSCOPE is to eliminate any physical connection by replacing the gold wires with an active system of electric discharging, as in LPF. It would increase complexity and cost, but it is feasible. Once thermal noise from internal damping were eliminated by eliminating the suspensions, the next relevant one would be thermal noise from gas damping [34, 35]. For the lowest possible residual pressure, a way to reduce gas damping noise is by increasing the gap, up to 4 mm in the case of LPF, as compared with 600 μm in MICROSCOPE.

However, the capacitance is inversely proportional to the gap, so larger gaps mean a less precise readout, which in fact in LPF is at nanometer level [36], while MICROSCOPE reports $3 \times 10^{-11}\,\mathrm{m}/\sqrt{\mathrm{Hz}}$ between 2×10^{-4} Hz and 1 Hz [11]. For LPF this is not an issue because its main readout is based on laser interferometry [32]. Instead, MICROSCOPE relies on the capacitance readout; moreover, the capacitors control the test cylinders, which would otherwise be unstable because of the negative spring of the electrostatic suspensions. Note that gaps in MICROSCOPE are already a factor of two larger than they were in GOCE.

Thus, eliminating the gold wires is not going to improve the MICROSCOPE experiment. Moreover, the problem is not with the mechanical suspensions *per se*, since the most precise and most successful gravitational experiments (RTB and gravitational wave detectors) are all based on mechanical suspensions.

Unlike electrostatic suspensions, mechanical suspensions act as positive springs and naturally provide the restoring force needed by the test masses. For instance, the deflection of the torsion balance under the effect of a torque with a non-zero component along the suspension fiber, including that of a WEP violation, is counteracted by its torsional elastic constant. In MICROSCOPE the restoring force must be provided actively, while the gold wire acts as an ancillary dummy spring with the sole purpose of ensuring electric grounding. Instead, mechanical suspensions can provide both the restoring force and electric grounding. The torsion balance of the Eöt-Wash WEP experiments weighs only 70 grams in total, and can be suspended with a very thin W wire of 20 μm diameter whose torsional elastic constant is very low (being inversely proportional to the 4^{th} power of the thickness). In gravitational wave detectors the mirrors to be suspended are much heavier (about 40 kg) and the fibers much thicker (about 350 μm), but at frequencies around 100 Hz thermal noise from internal damping is very low. Suspension fibers are metallic in Virgo, soon to be replaced with fibers in fused silica (as in LIGO) with even better quality factor. In orbit weight is no longer a limitation, and large masses can be used, which reduces the effects of non-gravitational forces, including those due to thermal noise. Thus, in space tests of the WEP mechanical suspensions are preferable.

The solution we are led to is twofold. In the first place, we must use mechanical suspensions with state-of-the-art fabrication and clamping procedures, so as to ensure high Q. Secondly, we must rotate the spacecraft much faster than MICROSCOPE, so as to up-convert the signal to a much higher frequency, where thermal noise from internal damping is significantly reduced.

MICROSCOPE has demonstrated the advantage of space for high precision rotating experiments. The possibility, unique to space, to spin the entire "laboratory", that is the spacecraft, along with the test cylinders makes rotation noise much lower in orbit than on ground. However, the spin rate of MICROSCOPE is limited by the need to rotate around an axis which is not the symmetry axis of the test cylinders ([11], Fig. 1). So far the highest reported spin rate, achieved during SUEP 120-orbit run which has given the result (2), has been of 2.94×10^{-3} Hz. Spacecraft can spin much faster than that, and be passively stabilized by rotation around the symmetry axis, as in the case of the METEOSAT which spin at 1.7 Hz.

Mechanical suspensions are very versatile and allow the concentric test cylinders to be arranged in such a way that they co-rotate with the spacecraft around the symmetry axis, being sensitive in the plane perpendicular to it. In this plane the relative displacements caused by tiny low frequency differential accelerations between the test cylinders –such as a violation signal– can be detected, upconverted by rotation to a much higher frequency where thermal noise is much lower [15]. After initial spin up, spacecraft stabilitazion is maintained passively by conservation of angular momentum, which ensures extremely low rotation noise and does not need propellant –to be left for drag-free control and occasional manoeuvres [18]. At Hz rather than mHz frequency thermal noise from internal damping is very low, and gaps can be increased so as to reduce also gas damping noise and make the integration time short even for a very high precision target [35]. With very low thermal noise a readout of comparable low noise is needed. With cm level gaps a capacitance readout is not sensitive enough; a laser gauge with very low noise at 1 Hz is well feasible [37, 38].

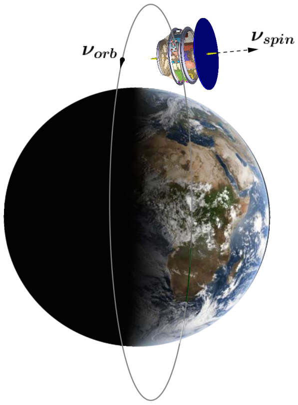

FIG. 1. Sketch of the GG satellite, with cylindrical symmetry and the dish of solar cells facing the Sun, in a high inclination sun-synchronous orbit ($\nu_{orb} \simeq 1.7 \times 10^{-4}$ Hz) similar to that of MICROSCOPE. The spacecraft is passively stabilized by one-axis rotation around the symmetry axis at $\nu_{spin} = 1$ Hz. After initial spin-up relative to inertial space rotation is maintained by conservation of angular momentum. (The Figure is not to scale; the bulk of the spacecraft body has a diameter of about 1.5 m).

The proposed GG space experiment incorporates all the features suggested by the MICROSCOPE experience and aims to test the WEP to 1 part in 10^{17} ([16–18], [4]). GG is currently in the shortlist of candidates to the next medium-sized mission of ESA awaiting the decision of the Agency.

GG will check MICROSCOPE's final result with at least 2 orders of magnitude better precision and improve by 4 orders of magnitude over RTB tests, thanks to the stronger signal in orbit (by about 500 times) and to a less extent by fully exploiting all the advantages of space in order to reach a sensitivity to differential accelerations better than RTB (by 20 times). The latter factor would be an improvement also of RTB tests of the WEP relative to the Sun and to dark matter in our galaxy.

The WEP test relative to dark matter deserves attention. Candidate dark matter particles are typically new particles, not included in the Standard Model of particle physics, which would generate a long-range composition-dependent scalar interaction, hence violate the WEP. RTB tests rule out such a new composition dependent interaction between dark matter and ordinary matter to a few parts in 10^5 [3] stating that, to this level of precision, dark matter in our galaxy interacts with ordinary matter via the gravitational interaction only. Our current understanding of the cosmos is based on the assumption that the required non luminous dark matter interacts with ordinary matter only gravitationally and there is no new long-range interaction. Although this assumption is very often taken for granted, we should be reminded that it is only an assumption and as such it should be tested by the most sensitive possible experiments.

GG, like MICROSCOPE, will be in a low altitude, high inclination, sun-synchronous orbit (with orbital frequency $\nu_{orb} \simeq 1.7 \times 10^{-4}$ Hz). However, thanks to its cylindrical symmetry (built around the concentric test cylinders), it can be passively stabilized by one-axis rotation around the symmetry axis at the rotation rate $\nu_{spin} = 1$ Hz (Fig. 1). The coaxial, concentric test cylinders are located at the center of mass of the spacecraft, co-rotate with it and are sensitive to differential forces acting in the plane perpendicular to the spin/symmetry axis. The relative displacements caused by any such force, like a violation signal, are read by a laser interferometry gauge, also co-rotating with the whole system. In the non-rotating frame of the spacecraft the violation signal is at the orbital frequency; the laser gauge rotating at $\nu_{spin} \gg \nu_{orb}$ reads it at ν_{spin}, where thermal noise from internal damping is much lower than it would be at ν_{orb} making the integration time much shorter [15, 35]. This is the key to reaching a very high precision. It suffices to notice that improving by a factor of 10 in sensitivity requires –with a given level of thermal noise– an integration time 100 times longer. As an example, MICROSCOPE early test (2) has required 8.26 days of integration time; were it aiming at 10^{-17}, the same experiment would need about 6.7 million days for one single measurement.

The expectations for GG, based on theoretical analysis, numerical simulations and laboratory tests, are for a signal-to-noise ratio of 2 in a few hours [35]. This allows a WEP test to 10^{-17} to be completed in 1 d (about 15 orbits). Then, since the spin axis (and the sensitive plane perpendicular to it) are fixed in inertial space, while the nodal line of the sun-synchronous orbit moves by about 1°/d (for the solar panel to follow the annual motion of the Sun), a large number of 1-d runs shall be available, in different physical conditions provided by the dynamical evolution, to allow a violation signal at 10^{-17} level to be separated with certainty from systematic errors [40, 41].

As discussed in Sec. II, for a test of the WEP in orbit to reach a very high precision it must deal with the huge effect of drag which requires, in addition to partial compensation by drag-free control, also a high level of Common Mode Rejection (CMR). CMR is the ability to reject effects which are, by their nature, the same on both test masses and therefore –if the sensor were perfectly differential, i.e. with infinite CMR– would not compete at all with the violation signal which is, by its nature, differential. The largest common mode effect in orbit is the inertial acceleration equal and opposite to the acceleration of the spacecraft caused by non-gravitational forces such as air drag and solar radiation pressure. A high level of CMR can be achieved if the test masses are arranged as a balance. This is how torsion balances have defeated mass dropping tests by many orders of magnitude. A torsion balance as such is not suitable for space [4]. However, space is favourbale to the realization of a very sensitive balance because in orbit the largest common mode force against which the balance is balanced is many orders of magnitude weaker than 1-g on ground (a space version of the Watt balance has been considered at PTB, the German national metrology institute [39]).

In order to achieve a high level of CMR the coaxial test cylinders in GG are arranged to form a beam balance, with the beam along the spin/symmetry axis, hence sensitive to differential forces in the plane perpendicular to it. The peculiarity of the GG balance is that, unlike ordinary beam balances, the two masses are concentric, which is a crucial requirement in space tests of the WEP to minimize classical differential tidal effects. The way such a beam balance with concentric test masses can be realized is based on an ingenious combination of weak, high Q flexures made in CuBe and coupling arms (with special attention to symmetry considerations) whose lengths can be finely adjusted to reach a very good balancing against the common mode inertial acceleration caused by drag. An animation of this balance is available on the front page of the GG website [17]. A 1-g version of it has been realized and tested in the lab with the “GG on Ground” (GGG) demonstrator [16], [4].

V. CONCLUSIONS

For the first time the equivalence principle of Galileo, Newton and Einstein, still at the crossroads of most open problems in fundamental physics, is under test in space, with the test masses weakly suspended inside the MICROSCOPE satellite in low Earth orbit. The mission is performing well and the experiment is successful. Early results based on a 120-orbit run (8.26 d) show no violation for Pt and Ti test masses relative to the Earth to about 10^{-14} [11] and improve by one order of magnitude over the best ground tests with rotating torsion balances [2, 3]. The improvement occurs despite a sensitivity to differential accelerations between the test cylinders about 70 times less than torsion balances, thanks to the much higher driving signal in orbit.

MICROSCOPE demonstrates the potential of space for further orders of magnitude improvement in testing the WEP.

The test turned out to be limited by thermal noise higher than expected due to the poor quality factor of the gold wires used for electrical grounding. The successful result was obtained by rotating the spacecraft faster than planned; it establishes spacecraft rotation as the most effective way of improving the precision of WEP tests.

The result will improve by the end of the mission because random noise decreases with more data. Not so systematic errors. For this purpose MICROSCOPE carries a second “zero-check” sensor with the test masses both made of Pt: all systematic errors should appear in both sensors while the violation signal, being composition dependent, must appear only in the Pt-Ti one and can therefore be identified. In the current situation in which a WEP test with a precision closer to the original target of 10^{-15} is going to require an integration time lasting almost as the entire mission, multiple measurements in different experimental conditions to check systematics will not be available, making the zero-check sensor crucial to separate a violation signal from tiny perturbations to be ascribed to known physics.

In order to establish if the Pt-Pt zero-check sensor will actually allow such separation we have compared the differential acceleration due to major non-gravitational disturbances in the two sensors. Using published data for the masses and the geometry of all test cylinders we have shown that the zero-check sensor is less sensitive to a wide class of systematics effects than the Pt-Ti one, and cannot therefore rule out that such errors (or a fraction of them) might be due to a violation signal. The early test itself reports systematics in the Pt-Ti sensor which are not detected in the Pt-Pt one, despite its lower noise ([11], Fig. 3).

Our work shows that there is no alternative to a rigorous campaign of systematic checks, which needs many measurements in different experimental conditions and therefore requires a short integration time, hence low thermal noise. Ultimately, this means: *i*) high quality state-of-the-art mechanical suspensions, as demonstrated by the most precise gravitational experiments on ground such as torsion balances and gravitational wave experiments, and *ii*) up-conversion of the signal to much higher frequency by faster rotation of the spacecraft, which can be easily achieved provided that it respects the cylindrical symmetry of the test bodies.

Getting rid of the gold wires and using larger gaps for lower damping noise, as in LISA Pathfinder, is shown not to be a solution for testing the WEP. In experiments to test the WEP in space mechanical suspensions are compatible with very low thermal noise and very short integration time at room temperature [15, 35] for the same 10^{-17} target as the cryogenic STEP experiment investigated by ESA and NASA [9] and by ESA alone [19].

Eliminating the mechanical suspensions in a WEP experiment in space because of the poor mechanical quality of the gold wires in MICROSCOPE would be like throwing the baby out with the dirty clothes.

MICROSCOPE demonstrates the huge potential of space for testing the weak equivalence principle. It teaches us that in order to realize the potential a new experiment must rotate much faster than at mHz frequency, and use mechanical suspensions with much better quality factor than a few tens. For both issues viable solutions exist based on proven technology.

We add that in order to deal with the huge effect of drag in orbit it is not enough to rely on drag-free control (which has worked perfectly in MICROSCOPE); it is also necessary to learn from torsion balances that common mode effects must (and can) be rejected very effectively [3]

We also add that it is time to replace capacitance readout with laser interferometry, learning from the successful experience of LPF [32, 33], especially if the high frequency of interest makes its realization by far less demanding, as demonstrated by lab tests carried out at INRIM, the Italian national metrology institute [37, 38].

The success of MICROSCOPE and of LPF demonstrates that space is favourable for extremely small force experiments. In addition, MICROSCOPE demonstrates that rotation of the spacecraft can significantly reduce noise. On ground, the success of rotating torsion balances and of gravitational wave detectors demonstrates that high precision gravitational experiments are made possible by high quality mechanical suspensions. In addition, torsion balances demonstrate that mechanical suspensions make it possible to achieve a high level of common mode rejection.

The proposed space experiment GG is based on all this experience and know-how with the goal of testing the weak equivalence principle to 10^{-17} at room temperature ([16–18],[4]). GG is in the shortlist of candidates to the next medium size mission of ESA awaiting the decision of the Agency.

Acknowledgements. Thanks are due to the colleagues of the GG collaboration for their contributions. The input for this work came during the 656th Wilhelm und Else Heraeus Seminar on “Fundamental Physics in Space” held in Bremen on October 2017. A.M.N wishes to thank the organizers for their invitation.

[1] R. V. Eötvös, Über die Anziehung der Erde auf verschiedene Substanzen, Mathematische und Naturwissenschaftliche Berichte aus Ungarn, Bd. 8, S. 65 (1890); see also On the Gravitation Produced by the Earth on Different Substances, The Abraham Zelmanov Journal 1, 6 (2008)

[2] S. Schlamminger, K.-Y. Choi, T. A. Wagner, J. H. Gundlach, E. G. Adelberger, Test of the equivalence principle using a rotating torsion balance, Phys. Rev. Lett. 100, 041101(4p) (2008)

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

[4] A. M. Nobili and A. Anselmi, Relevance of the weak equivalence principle and experiments to test it: lessons from the past and improvements expected in space, Phys. Lett. A, Special Issue in memory of V. A. Braginsky (2018)

[5] P. W. Worden, Jr. and 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, p. 381 (1973)

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

[7] P. W. Worden, Jr., Equivalence principle tests in earth orbit, Acta Astronautica 5, 27 (1978)

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

[9] J. P. Blaser *et al. (including A. M. Nobili),* Satellite Test of the Equivalence Principle, Report on the Phase A Study, ESA/NASA SCI (93)4 (1993)

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

[11] P. Touboul *et al.*, MICROSCOPE Mission: First Results of a Space Test of the Equivalence Principle, Phys. Rev. Lett. 119, 231101 (2017)

[12] V. Josselin, P. Touboul, and R. Kielbasa, Capacitive detection scheme for space accelerometers applications, Sensors and Actuators 78, 92 (1999)

[13] E. Willemenot, and P. Touboul, On-ground investigation of space accelerometers noise with an electrostatic torsion pendulum, Rev. Sci. Instrum. 71, 302 (2000)

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

[15] R. Pegna, A. M. Nobili, M. Shao, S. G. Turyshev, G. Catastini, A. Anselmi, R. Spero, S. Doravari, G. L. Comandi, and A. De Michele, Abatement of Thermal Noise due to Internal Damping in 2D Oscillators with Rapidly Rotating Test Masses, Phys. Rev. Lett. 107, 200801 (2011)

[16] 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, and A. Anselmi, “Galileo Galilei” (GG): space test of the weak equivalence principle to 10^{-17} and laboratory demonstrations, Class. Quantum Grav. 29, 184011 (2012)

[17] “Galileo Galilei” (GG) Website (2018)

[18] A. Anselmi and G. Catastini, Design of the GG satellite, Phys. Lett. A 318, 205 (2003)

[19] J. P. Blaser *et al.* Satellite Test of the Equivalence Principle, Report on the Phase A Study, ESA SCI (96)5 (1996)

[20] M. Rodrigues, MICROSCOPE mission: a test of the

equivalence principle in space, Invited talk given at 656th WE-Heraeus-Seminar “Fundamental Physics in Space”, Bremen 23-27 October (2017)

[21] E. Willemenot, Pendule de torsion. . . suspension électrostatique, PhD Thesis, Université Orsay-Paris XI (1997)

[22] E. Willemenot, and P. Touboul, Electrostatically suspended torsion pendulum, Rev. Sci. Instrum. 71, 310 (2000)

[23] LIGO Facts, Webpage at Caltech (2018)

[24] B. Christophe, GRACE and GOCE experience, Talk given at MICROSCOPE Colloquium II, ONERA, Paris (2013)

[25] P. Touboul, MICROSCOPE mission overview, Talk given at MICROSCOPE Colloquium II, ONERA, Paris (2013)

[26] B. Frommknecht, D. Lamarre, M. Meloni, A. Bigazzi, and R. Floberghagen, GOCE level 1b data processing, J. Geod. 85, 759 (2011)

[27] A. Milani, A. M. Nobili, and P. Farinella, *Non-gravitational perturbations and satellite geodesy* (Adam Hilger Ltd., Bristol 1987)

[28] P. Touboul, Among space fundamental physics missions, MICROSCOPE, a simple free-fall test, Talk given at Moriond and GPhys Colloquium on gravittaional waves and experimental gravity, La Thuile (2011)

[29] D. Hagedorn *et al.,* MICROSCOPE - fabricating test masses for an in-orbit test of the equivalence principle, Ann. Phys. 525, 720 (2013)

[30] A. M. Nobili *et al.*, Radiometer effect in space missions to test the equivalence principle, Phys. Rev. D Rapid Commun. 63, 101101(R) (2001)

[31] A. M. Nobili *et al.*, Radiometer effect in the Microscope space mission, New Astronomy 7, 521-529 (2002)

[32] M. Armano *et al.*, Sub-Femto-g Free Fall for Space-Based Gravitational Wave Observatories: LISA Pathfinder Results, Phys. Rev. Lett. 116, 231101 (2016)

[33] M. Armano *et al.*, Beyond the Required LISA Free-Fall Performance: New LISA Pathfinder Results down to 20 μHz, Phys. Rev. Lett. 120, 0611011 (2018)

[34] A. Cavalleri, G. Ciani, R. Dolesi, A. Heptonstall, M. Hueller, D. Nicolodi, S. Rowan, D. Tombolato, S. Vitale, P. J. Wass, and W. J. Weber ,Increased Brownian Force Noise from Molecular Impacts in a Constrained Volume, Phys. Rev. Lett. 103, 140601 (2009)

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

[36] M. Armano *et al.*, Capacitive sensing of test mass motion with nanometer precision over millimeter-wide sensing gaps for space-borne gravitational reference sensors, Phys. Rev. D 96, 062004 (2017)

[37] M. Pisani, G. Mana, and A. M. Nobili, Design of an interferometric displacement sensor with picometer resolution for the Galileo Galilei mission, IEEE 978-1-4799-7569-3/15, 591-595 (2015)

[38] M. Pisani, M. Zucco, and 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, 260-265 (2016)

[39] H. Dittus, *recollection during the Bremen Conference on “Fundamental Physics in Space”* (2017)

[40] A. M. Nobili *et al.*, Major challenges of a high precision test of the equivalence principle in space, in Proceedings of XLVIth Moriond Conference on “Gravitational Waves and Experimental Gravity, Eds. E. Augè, J. Dumarchez and J. Tran Thanh Van, The Gioi Publishers, p. 341 (2011)

[41] A. M. Nobili, GG: Abatement of thermal noise by rotation and null checks of an equivalence principle violation signal to 10^{-17}, Talk given at XLVIth Moriond Conference on “Gravitational Waves and Experimental Gravity” (2011)