

Measurement of gravitational coupling between millimetre-sized masses

<https://doi.org/10.1038/s41586-021-03250-7>

Tobias Westphal¹✉, Hans Hepach^{1,4}, Jeremias Pfaff^{2,4} & Markus Aspelmeyer^{1,2,3}✉

Received: 17 September 2020

Accepted: 18 January 2021

Published online: 10 March 2021

 Check for updates

Gravity is the weakest of all known fundamental forces and poses some of the most important open questions to modern physics: it remains resistant to unification within the standard model of physics and its underlying concepts appear to be fundamentally disconnected from quantum theory^{1–4}. Testing gravity at all scales is therefore an important experimental endeavour^{5–7}. So far, these tests have mainly involved macroscopic masses at the kilogram scale and beyond⁸. Here we show gravitational coupling between two gold spheres of 1 millimetre radius, thereby entering the regime of sub-100-milligram sources of gravity. Periodic modulation of the position of the source mass allows us to perform a spatial mapping of the gravitational force. Both linear and quadratic coupling are observed as a consequence of the nonlinearity of the gravitational potential. Our results extend the parameter space of gravity measurements to small, single source masses and low gravitational field strengths. Further improvements to our methodology will enable the isolation of gravity as a coupling force for objects below the Planck mass. This work opens the way to the unexplored frontier of microscopic source masses, which will enable studies of fundamental interactions^{9–11} and provide a path towards exploring the quantum nature of gravity^{12–15}.

The last decades have seen numerous experimental confirmations of Einstein's theory of relativity, our best working theory of gravity, by observing massive astronomical objects and their dynamics^{5,16}. This culminated in the recent direct detection of gravitational waves from the merger of two black holes¹⁷ and the direct imaging of a supermassive black hole¹⁸. Meanwhile, Earth-bound experiments have been continuously increasing their sensitivity to gravity phenomena at laboratory scales, including general relativistic effects^{19,20}, tests of the equivalence principle^{6,21}, precision measurements of Newton's constant^{22–24} and tests of the validity of Newton's law at micrometre-scale distances^{25–27}. Although test masses in such experiments span the whole range from macroscopic objects to individual quantum systems^{19–21,24,28}, the gravitational source is typically either Earth or masses at the kilogram scale and beyond⁸. This is contrasted by an increasing interest to study gravitational phenomena originating from quantum states of source masses, for example, in the form of 'quantum Cavendish experiments'^{1,12–14,29}. Because quantum coherence is easily lost for increasing system size, it is important to isolate gravity as a coupling force for as small objects as possible.

Experiments with smaller source masses are much more difficult—the gravitational force generated at a given distance by a spherical mass of radius R shrinks with R^{-3} —and hence only few experiments have so far observed gravitational signatures of gram-scale mass configurations^{27,30,31}. In one case, a hole pattern in a rotating, 5-cm-diameter attractor disk made from platinum generated a periodic mass modulation of a few hundred milligrams, which was resolved in a torsional balance measurement²⁷. In another case, a single 700-mg tungsten sphere was used to resonantly excite a torsion pendulum³¹. Isolating gravitational interactions generated by even smaller, single source-mass objects is a challenging task because

it requires increasing efforts to shield residual contributions from other sources of acceleration, in particular of seismic and electromagnetic nature³². In addition, resonant detection schemes, which are typically employed to amplify the signal above the readout noise, amplify displacement noise as well, and hence do not yield any gain in terms of separating the signal from other force noise sources. Here we overcome this limitation by combining time-dependent gravitational accelerations with an off-resonant detection scheme of a well balanced differential mechanical mode and independent noise estimation. In this way, we can measure the gravitational field of single source masses smaller than 100 mg.

Experiment

In our experiment, the gravitational source is a nearly spherical gold mass of radius $R = 1.07 \pm 0.04$ mm and mass $m_s = 92.1 \pm 0.1$ mg. A similarly sized gold sphere acts as a test mass of $m_t = 90.7 \pm 0.1$ mg. The idea is that a periodic modulation of the position of the source mass generates a time-dependent gravitational potential at the location of the test mass, the acceleration of which is measured in a miniature torsion pendulum configuration (Fig. 1). The experiment is conducted in high vacuum (6×10^{-7} mbar), which minimizes residual noise from acoustic coupling and momentum transfer of gas molecules³² (see Methods). To prevent nonlinear coupling of high-frequency vibrations into the relevant low-frequency measurement band around the modulation frequency $f_{\text{mod}} = 12.7$ mHz, the support structure of the pendulum rests on soft, vacuum-compatible rubber feet³³.

We optically monitor the angular deflection of the pendulum, which provides a calibrated readout of the motion of the test mass with a

¹Institute for Quantum Optics and Quantum Information (IQOQI) Vienna, Austrian Academy of Sciences, Vienna, Austria. ²Faculty of Physics, University of Vienna, Vienna, Austria. ³Research Platform TURIS, University of Vienna, Vienna, Austria. ⁴These authors contributed equally: Hans Hepach, Jeremias Pfaff. ✉e-mail: tobias.westphal@oeaw.ac.at; markus.aspelmeyer@univie.ac.at

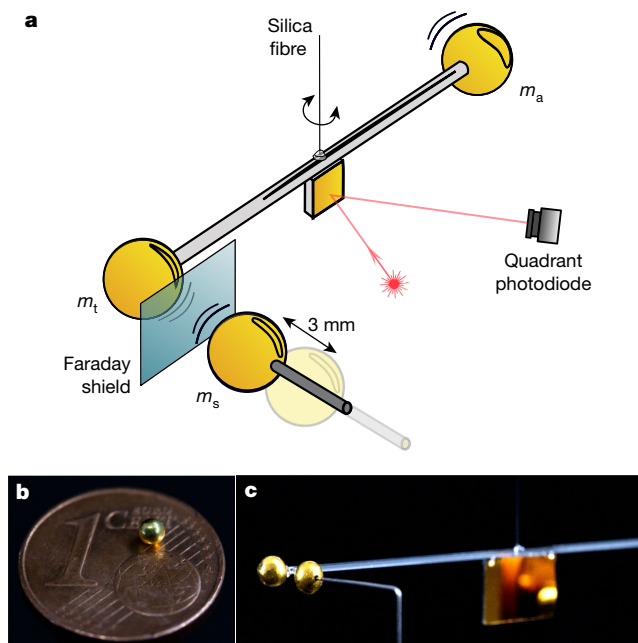


Fig. 1 | Experimental setup. **a**, Schematic of the experiment. The torsion pendulum, which acts as a transducer for gravitational acceleration, consists of two gold spheres of radius $r \approx 1$ mm held at 40 mm centre distance by a glass capillary. One mass serves as a test mass ($m_t = 90.7$ mg) and the other ($m_a = 91.5$ mg) as counterbalance that provides vibrational-noise common-mode rejection. A 4- μ m-diameter silica fibre provides a soft ($f_0 \approx 3.6$ mHz) torsional resonance that is well separated from other degrees of freedom. The torsion angle is read out via an optical lever directed to a quadrant photodiode. The gravitational interaction is modulated by harmonically moving the source mass ($m_s = 92.1$ mg) by about 3 mm at a frequency of $f_{\text{mod}} = 12.7$ mHz, well above the torsion resonance, using a shear piezo. Direct electrostatic coupling is suppressed by discharging and by a 150- μ m-thick Faraday shield. **b**, The source mass on a 1 euro cent coin. **c**, Photograph of the torsion pendulum and the mounted source mass.

detector-noise-limited sensitivity of about $2 \text{ nm Hz}^{-1/2}$. Figure 2 shows an amplitude spectrum of the displacement of the test mass. The torsional mode resembles a damped harmonic oscillator with resonance frequency $f_0 = 3.59$ mHz and mechanical quality factor $Q = 4.9$. It is well decoupled from other oscillation modes of the pendulum, which do not appear below 0.5 Hz. Diurnal variations in the low-frequency noise floor limit the times of highest sensitivity to those hours during the night when local public transport and pedestrian and car traffic are minimized (typically between midnight and 5 a.m.). Then, the oscillation of the test mass is mainly governed by thermal noise (see Methods). The corresponding force spectrum is obtained by deconvolution of the test-mass displacement time series with the inverse of the deduced mechanical susceptibility. The spectrum exhibits a flat noise floor that allows off-resonant detection of test-mass accelerations at frequencies of up to 0.1 Hz at a sensitivity better than $2 \times 10^{-11} \text{ m s}^{-2}$ within half a day. We use this to obtain a live estimate of the pendulum noise conditions, which is combined with inverse-variance weighting of the data to optimize the information obtained per experimental run.

The source mass is mounted on a 300- μ m-diameter titanium rod that is connected to a bending piezo, which provides a travel range of more than 5 mm in the vicinity of the test mass. The geometric separation of the masses is determined with an accuracy of 20 μ m by linking the drive piezo signal to position information from video tracking through a template-matching algorithm (see Methods). During the experiment, the centre distance between source and test mass is varied between 2.5 mm and 5.8 mm, with a minimal surface distance of about 0.4 mm.

To isolate gravity as a coupling force, we need to minimize other influences on the test mass. In addition to seismic and acoustic effects, these

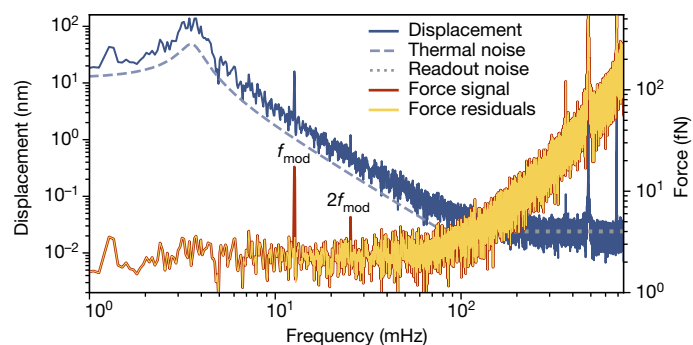


Fig. 2 | Spectrum of a single measurement run. The rotation of the torsional oscillator, measured over a period of 13.5 h, is calibrated via the test-mass displacement (blue solid line) and the corresponding exerted force (red solid line). The measurement is limited by both thermal noise (dashed line) and diurnally varying white force noise below 100 mHz, whereas displacement readout noise (dotted line) dominates above 100 mHz. The torsional motion is well decoupled from other oscillator modes, starting at 0.5 Hz. Force residuals (yellow solid line) represent the difference between measured and expected gravitational force. The spectrum shows gravitational coupling at a modulation frequency of $f_{\text{mod}} = 12.7$ mHz and nonlinearity of the gravitational potential at $2f_{\text{mod}} = 25.4$ mHz, well above the resonance frequency $f_0 = 3.59$ mHz of the oscillator. The flatness of the force residuals indicates that both signals are in agreement with the expected gravitational signal. Other sources of nonlinearity are found to be negligible (see Methods).

are predominantly electromagnetic interactions. We ground the source mass by directly connecting it to the vacuum chamber. The test mass is discharged to less than 8×10^4 elementary charges using ionized nitrogen³⁴, a method that was developed for charge mitigation in interferometric gravitational-wave detectors¹⁷. Further shielding is provided by a 150- μ m-thick conductive Faraday shield of gold-plated aluminium that is mounted between the source and test masses. In that way, electrostatic forces are suppressed to well below 3% of the expected gravitational coupling strength. Other shielding measures include housing the source-mass drive piezo inside a Faraday shield to suppress coupling via the applied electric fields, as well as gold-coating and grounding most surfaces inside the vacuum chamber to exclude excitation from time-varying charge distributions. We also rule out the presence of relevant magnetic forces by independently measuring the magnetic permeabilities of the masses. As expected, permanent magnetic dipoles are negligible in both (diamagnetic) gold and (paramagnetic) titanium. Coupling via induced magnetic moments, either from Earth's magnetic field or from low-frequency magnetic noise originating in nearby urban tram traffic, is also found to be orders of magnitude below the expected gravitational coupling strength. In the current experimental geometry, Casimir forces are negligible, although they probably become a dominant factor for considerably smaller masses³⁵. Another relevant noise source is Newtonian noise, which is caused by non-stationary environmental gravitational sources and cannot be shielded from. For comparison, at a centre separation of 2.5 mm the static gravitational force between our masses is expected to be 9×10^{-14} N. The same gravitational force is exerted on the test mass by an experimenter standing at a distance of 2.5 m, or by a typical Vienna tram at 50 m distance from the laboratory building. Consequently, our experiment experiences a complex low-frequency gravitational noise of urban origin. It is obvious that such gravitational noise sources will pose an increasing challenge for future experiments. At present, our torsion pendulum is sufficiently small compared to the distance of typical environmental sources to be insensitive owing to common mode rejection.

Results

We observe gravitational coupling between the two masses by harmonically modulating the source-mass position $x_s = d_0 + d_m \cos(2\pi f_{\text{mod}} t + \phi)$

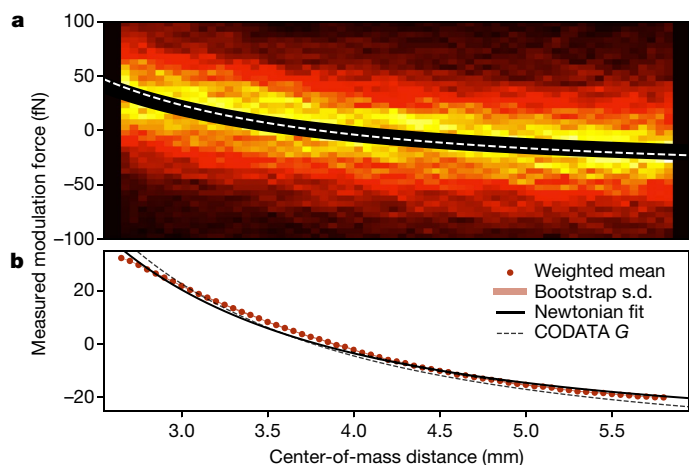


Fig. 3 | Spatial mapping of the gravitational force. **a**, The probability density distribution of the exerted force versus source–test mass separation shows the spatial nonlinearity of the source–mass potential. This method makes full use of our measurement data by taking into account weighting with the current noise conditions. The apparent width of the distribution is mainly given by noise at frequencies other than that of the modulation (a.c. force). **b**, The actual measurement precision becomes apparent by extracting the mean of the force measured at given distances (red points) along with their standard deviation obtained by bootstrapping (pink band). We note that the small statistical error collapses the error band to a narrow pink line. A full-weighted fit of the 13.5-h-long data segment with Newtonian gravity (black line) yields $G_{\text{fit}} = (5.89 \pm 0.20) \times 10^{-11} \text{ m}^3 \text{ kg}^{-1} \text{ s}^{-2}$, which is in agreement with our long-term combined value (see Fig. 4). Comparison with the literature (CODATA) value $G_{\text{CODATA}} = 6.67 \times 10^{-11} \text{ m}^3 \text{ kg}^{-1} \text{ s}^{-2}$ (dashed line) and our known systematic errors shows that influences other than gravity are below 10%.

at a frequency of $f_{\text{mod}} = 12.7 \text{ mHz}$, well above the fundamental torsional pendulum resonance (mean centre-of-mass distance, d_0 ; modulation amplitude, d_m). In this frequency regime, the test-mass response becomes independent of the pendulum properties and behaves essentially as a free mass. According to Newton’s law, the source mass generates a gravitational acceleration of the test mass of $a_G = Gm_s/x_s^2(t)$ (G , Newton’s constant). The $1/r$ dependence of the gravitational potential results in higher-order contributions to the coupling at multiples of the modulation frequency f_{mod} . Figure 2 shows the measured force spectrum of the test mass for a distance modulation of $d_m \approx 1.6 \text{ mm}$ at a centre separation of $d_0 \approx 4.2 \text{ mm}$, where we resolve the linear and quadratic acceleration modulations at f_{mod} and $2f_{\text{mod}}$, respectively. The flat noise residuals clearly indicate that the observed peak heights agree well with the expected force modulation due to Newtonian gravity. This confirms the gravitational origin of the interaction.

To more accurately quantify the strength of the coupling, we correlate the measured separation x_s between source and test mass with the independently inferred force on the test mass. This provides us with a position-dependent mapping of the gravitational force (Fig. 3). All data evaluation is carried out in post-processing and using non-causal zero-phase filtering (see Methods). Each dataset, consisting of up to 13.7-h-long measurements, is fitted using the expression for Newtonian gravity to obtain a value for the measured coupling strength.

Figure 4 summarizes the results of a series of measurements that were taken during the seismically quiet Christmas season. The weighted combination of measurement runs yields a coupling constant of $G_{\text{comb}} = (6.04 \pm 0.06) \times 10^{-11} \text{ m}^3 \text{ kg}^{-1} \text{ s}^{-2}$, with a statistical uncertainty of 1% (all errors reported are 1 s.d.). The observed coupling deviates from the recommended CODATA value for Newton’s constant ($G_{\text{CODATA}} = 6.67430(15) \times 10^{-11} \text{ m}^3 \text{ kg}^{-1} \text{ s}^{-2}$) by around 9%. This offset is fully covered by the known systematic uncertainties in our experiment, which include unwanted electrostatic, magnetic and gravitational

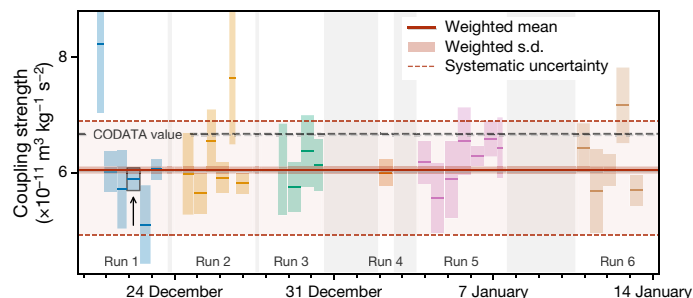


Fig. 4 | Collection of measurement runs. Over a stretch of three weeks during the seismically quiet Christmas season, a total of 29 measurement runs were conducted, with electrostatic coupling being characterized during periods shaded in grey. The dataset presented in Figs. 2, 3 is highlighted in black (run 1). The statistical error (1σ error bar) of the best-fitting Newton-like coupling of each data segment (dark lines) varies because of highly non-stationary noise in the measurement band. Combining individual measurements weighted by their respective data quality yields $G_{\text{comb}} = (6.04 \pm 0.06) \times 10^{-11} \text{ m}^3 \text{ kg}^{-1} \text{ s}^{-2}$. The plotted systematic uncertainty (red dashed line) includes the identified systematic influences from the experiment as summarized in Extended Data Table 2. We explicitly do not claim a significant systematic deviation of the combined coupling value from the CODATA recommended value (blue dashed). The experiment and its data evaluation was mainly designed with the reduction of statistical errors in mind, and the deviation is fully covered by the known experimental systematics.

influences from the masses and supports, as well as geometric uncertainties in the centre-of-mass distance due to the actual shape of the masses (we provide a detailed list in Methods). Our results show that we are able to isolate the gravity of a single, small source mass, with other influences being below the 10% level. The small statistical error underlines the precision character of our measurement and our ability to measure even smaller source masses with our approach in the future.

Discussion and outlook

We have demonstrated gravitational coupling between a test mass and a 90-mg spherical source mass. To our knowledge, this is the smallest single object whose gravitational field has been measured. Our measurements resolve an acceleration modulation of about $3 \times 10^{-10} \text{ m s}^{-2}$ at the drive frequency f_{mod} with high accuracy, quantified by a systematic uncertainty of around $3 \times 10^{-11} \text{ m s}^{-2}$, and high precision, quantified by a statistical uncertainty of about $3 \times 10^{-12} \text{ m s}^{-2}$ over an integration time of 350 h. Contributions of non-gravitational forces were kept below 10% of the observed signal. Our results extend gravity experiments to the regime of small gravitational source masses, with the potential of going considerably smaller. Given that our current sensitivity is limited by the thermal noise floor of the torsion pendulum when operated under optimal ambient conditions, it can be readily improved by increasing the mechanical quality factor Q . Probing the gravity of objects even smaller than the Planck mass ($m_p \approx 22 \mu\text{g} = 2.4 \times 10^{-4} m_s$) should become possible for $Q \geq 20,000$ (see Methods). The thermal-noise-limited force sensitivity of such miniature torsion pendula can be improved even further by dissipation dilution, as recently demonstrated³⁶. Another interesting route is to use levitated test masses, which provide almost perfect charge mitigation and similar acceleration sensitivities^{37–41}. To make full use of improvements in thermal noise, it is important to understand and to mitigate other, predominantly anthropogenic, low-frequency noise sources that currently dominate our measurement performance most of the time. Possible countermeasures include the appropriate choice of a remote laboratory location or pushing the experiment to much higher pendulum frequencies^{32,42}. Going to smaller masses will involve additional challenges, because other noise contributions will play an increasing role. For example, even for the ideal case of electrically neutral source and test

masses, Casimir forces will become relevant at distances below 100 μm . We note, however, that electromagnetic shielding combined with signal modulation can overcome this particular challenge. We also note that our method allowed us to directly measure the position dependence of the gravitational force, which constitutes a one-dimensional mapping of the gravitational field strength. A more complex test-mass geometry may enable full mapping of the metric tensor of such small source masses⁴³.

Our experiment provides a viable path to enter and explore a regime of gravitational physics that involves precision tests of gravity with isolated microscopic source masses at or below the Planck mass. This opens up possibilities such as a different approach to determine Newton's constant³², which so far remains the least well determined of the fundamental constants⁴⁴. In general, miniaturized precision experiments may allow tests of the inverse-square law of gravity at considerably smaller scales than possible today (the best reported value is of the order of 50 μm)²⁷, thereby probing, for example, fundamental features of string theory such as extra dimensions and massless scalar particles⁴⁵. Small source masses also allow us to test the consequences of speculative scalar fields that have been discussed in the context of dark energy^{9–11}. More stringent parameter constraints for so-called chameleon forces may already be possible for source masses at the level investigated here⁹. Another example is experimental access to gravitational accelerations at or even below the level of galaxy rotations. Modification of Newtonian dynamics in this regime has been suggested as an alternative to dark-matter scenarios⁴⁶, and some of these models can therefore be directly tested⁴⁷. Finally, the ongoing discussion on the quantum nature of gravity has revived a gedankenexperiment by Feynman to probe gravitational coupling between quantum systems^{12–14}. The underlying question is whether it is possible to probe gravitational phenomena that cannot be explained by a purely classical source-mass configuration¹. Such experimental tests have so far been impossible, because they require the ability to prepare quantum states of motion of the source mass. Because decoherence phenomena scale markedly with system size, isolating the gravity of microscopic masses is a prerequisite for such future experiments. Our result provides a first step in this direction—although we stress that, with current quantum experiments using sub-micrometre-sized objects of 10⁻¹⁷ kg (refs. 48,49), adding quantum coherence to experiments in the relevant mass regime remains a completely different experimental challenge.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-021-03250-7>.

- Unruh, W. G. In *Quantum Theory of Gravity: Essays in Honor of the 60th Birthday of Bryce S. DeWitt* (ed. Christensen, S. M.) 234–242 (Adam Hilger Limited, 1984).
- Preskill, J. Do black holes destroy information. In *Proc. of the International Symposium on Black Holes, Membranes, Wormholes and Superstrings* (eds Kalara S. & Nanopoulos, D. V.) 1992 (World Scientific, 1993).
- Greenberger, D. M. The disconnect between quantum mechanics and gravity. Preprint at <https://arxiv.org/abs/1011.3719> (2010).
- Penrose, R. On the gravitization of quantum mechanics 1: quantum state reduction. *Found. Phys.* **44**, 557–575 (2014).
- Will, C. M. The confrontation between general relativity and experiment. *Living Rev. Relativ.* **17**, 3 (2014).
- Adelberger, E. New tests of Einstein's equivalence principle and Newton's inverse-square law. *Class. Quantum Gravity* **18**, 2397 (2001).
- Hossenfelder, S. Experimental search for quantum gravity. Preprint at <https://arxiv.org/abs/1010.3420v1> (2010).
- Gillies, G. T. & Unnikrishnan, C. S. The attracting masses in measurements of G: an overview of physical characteristics and performance. *Philos. Trans. R. Soc. A* **372**, 20140022 (2014).
- Feldman, B. & Nelson, A. E. New regions for a chameleon to hide. *J. High Energy Phys.* **2006**, 002 (2006).
- Burrage, C., Copeland, E. J. & Hinds, E. Probing dark energy with atom interferometry. *J. Cosmol. Astropart. Phys.* **2015**, 042 (2015).
- Hamilton, P. et al. Atom-interferometry constraints on dark energy. *Science* **349**, 849–851 (2015).

- DeWitt, C. M. & Rickles, D. (eds) *The Role of Gravitation in Physics: Report from the 1957 Chapel Hill Conference* (Max Planck Research Library for the History and Development of Knowledge, 2011).
- Bose, S. et al. Spin entanglement witness for quantum gravity. *Phys. Rev. Lett.* **119**, 240401 (2017).
- Marletto, C. & Vedral, V. Gravitationally induced entanglement between two massive particles is sufficient evidence of quantum effects in gravity. *Phys. Rev. Lett.* **119**, 240402 (2017).
- Belenchia, A. et al. Quantum superposition of massive objects and the quantization of gravity. *Phys. Rev. D* **98**, 126009 (2018).
- Ransom, S. M. et al. A millisecond pulsar in a stellar triple system. *Nature* **505**, 520–524 (2014).
- Abbott, B. P. et al. Observation of gravitational waves from a binary black hole merger. *Phys. Rev. Lett.* **116**, 061102 (2016).
- Akiyama, K. et al. First M87 event horizon telescope results. VI. The shadow and mass of the central black hole. *Astrophys. J. Lett.* **875**, L6 (2019).
- Chou, C. W., Hume, D. B., Rosenband, T. & Wineland, D. J. Optical clocks and relativity. *Science* **329**, 1630–1633 (2010).
- Asenbaum, P. et al. Phase shift in an atom interferometer due to spacetime curvature across its wave function. *Phys. Rev. Lett.* **118**, 183602 (2017).
- Rosi, G. et al. Quantum test of the equivalence principle for atoms in coherent superposition of internal energy states. *Nat. Commun.* **8**, 15529 (2017).
- Gundlach, J. & Merkowitz, S. Measurement of Newton's constant using a torsion balance with angular acceleration feedback. *Phys. Rev. Lett.* **85**, 2869–2872 (2000).
- Quinn, T., Parks, H., Speake, C. & Davis, R. Improved determination of G using two methods. *Phys. Rev. Lett.* **111**, 101102 (2013).
- Rosi, G., Sorrentino, F., Cacciapuoti, L., Prevedelli, M. & Tino, G. M. Precision measurement of the Newtonian gravitational constant using cold atoms. *Nature* **510**, 518–521 (2014).
- Geraci, A. A., Smullin, S. J., Weld, D. M., Chiaverini, J. & Kapitulnik, A. Improved constraints on non-Newtonian forces at 10 microns. *Phys. Rev. D* **78**, 022002 (2008).
- Tan, W.-h. et al. Improvement for testing the gravitational inverse-square law at the submillimeter range. *Phys. Rev. Lett.* **124**, 051301 (2020).
- Lee, J. G., Adelberger, E. G., Cook, T. S., Fleischer, S. M. & Heckel, B. R. New test of the gravitational 1/r² law at separations down to 52 μm . *Phys. Rev. Lett.* **124**, 101101 (2020).
- Colella, R., Overhauser, A. & Werner, S. Observation of gravitationally induced quantum interference. *Phys. Rev. Lett.* **34**, 1472–1474 (1975).
- Al Balushi, A., Cong, W. & Mann, R. B. Optomechanical quantum Cavendish experiment. *Phys. Rev. A* **98**, 043811 (2018).
- Hoskins, J. K., Newman, R. D., Spero, R. & Schultz, J. Experimental tests of the gravitational inverse-square law for mass separations from 2 to 105 cm. *Phys. Rev. D* **32**, 3084–3095 (1985).
- Mitrofanov, V. P. & Ponomareva, O. I. Experimental test of gravitation at small distances. *Sov. Phys. JETP* **67**, 1963 (1988).
- Schmölle, J., Dragosits, M., Hepach, H. & Aspelmeyer, M. A micromechanical proof-of-principle experiment for measuring the gravitational force of milligram masses. *Class. Quantum Gravity* **33**, 125031 (2016).
- Shimoda, T. & Ando, M. Nonlinear vibration transfer in torsion pendulums. *Class. Quantum Gravity* **36**, 125001 (2019).
- Ugolini, D., Funk, Q. & Amen, T. Discharging fused silica test masses with ionized nitrogen. *Rev. Sci. Instrum.* **82**, 046108 (2011).
- Canaguier-Durang, A. et al. Casimir interaction between a dielectric nanosphere and a metallic plane. *Phys. Rev. A* **83**, 032508 (2011).
- Komori, K. et al. Attonewton-meter torque sensing with a macroscopic optomechanical torsion pendulum. *Phys. Rev. A* **101**, 011802 (2020).
- Prat-Camps, J., Teo, C., Rusconi, C. C., Wiecek, W. & Romero-Isart, O. Ultrasensitive inertial and force sensors with diamagnetically levitated magnets. *Phys. Rev. Appl.* **8**, 034002 (2017).
- Timberlake, C., Gasbarri, G., Vinante, A., Setter, A. & Ulbricht, H. Acceleration sensing with magnetically levitated oscillators above a superconductor. *Appl. Phys. Lett.* **115**, 224101 (2019).
- Monteiro, F. et al. Force and acceleration sensing with optically levitated nanogram masses at microkelvin temperatures. *Phys. Rev. A* **101**, 053835 (2020).
- Lewandowski, C. W., Knowles, T. D., Etienne, Z. B. & D'Urso, B. High sensitivity accelerometry with a feedback-cooled magnetically levitated microsphere. *Phys. Rev. Appl.* **15**, 014050 (2021).
- Kawasaki, A. et al. High sensitivity, levitated microsphere apparatus for short-distance force measurements. *Rev. Sci. Instrum.* **91**, 083201 (2020).
- Liu, Y., Mummery, J. & Sillanpää, M. A. Prospects for observing gravitational forces between nonclassical mechanical oscillators. Preprint at <https://arxiv.org/abs/2008.10477> (2020).
- Obukhov, Y. N. & Puetzfeld, D. In *Fundamental Theories of Physics* 87–130 (Springer, 2019).
- Speake, C. & Quinn, T. The search for Newton's constant. *Phys. Today* **67**, 27–33 (2014).
- Adelberger, E., Heckel, B. & Nelson, A. Tests of the gravitational inverse-square law. *Annu. Rev. Nucl. Part. Sci.* **53**, 77–121 (2003).
- Milgrom, M. A modification of the Newtonian dynamics as a possible alternative to the hidden mass hypothesis. *Astrophys. J.* **270**, 365 (1983).
- Ignatiev, A. Testing MOND on Earth 1. *Can. J. Phys.* **93**, 166–168 (2015).
- Tebbenjohanns, F., Frimmer, M., Jain, V., Windey, D. & Novotny, L. Motional sideband asymmetry of a nanoparticle optically levitated in free space. *Phys. Rev. Lett.* **124**, 013603 (2020).
- Delić, U. et al. Cooling of a levitated nanoparticle to the motional quantum ground state. *Science* **367**, 892–895 (2020).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2021

Methods

Torsion pendulum

Geometry. The torsion pendulum consists of the test mass ($m_t = 90.7$ mg) and a counterbalance mass ($m_a = 91.5$ mg). The masses are rigidly connected by a 37.9-mm-long, square-shaped glass capillary with an outer dimension of 0.4 mm and wall thickness of 0.1 mm (9.6 mg). It is supported by a 35-mm-long quartz fibre pulled down to 3.6 μ m diameter over a 20-mm section at the centre for maximal torsional compliance. The 5–10-mm-long, up to 50- μ m-wide taper region below this torsion section increases the pitch and roll stiffness for better mode separation and provides a bigger cross-section for attachment to the capillary. The attachment point is tuned to minimize static roll to <10 mrad. A 5.85-mm-wide, 0.3-mm-thick, hand-ground and gold-coated mirror (27.0 mg) is attached underneath the capillary for deflection readout of the torsion (yaw) angle. This turns to point downwards by 50 mrad in pitch, potentially coupling the non-torsion motion—in particular roll—into the yaw readout as well. All connections are established by an ultraviolet-curing adhesive. The fundamental mode frequency f_0 is determined by the stiffness of the suspension fibre and the effective mass $m_{\text{eff}} = 183.8$ mg $\approx m_t + m_a = 182.2$ mg. The effective mass is almost minimal, that is, hugely dominated by the two gold masses for the balanced, differential mode used here. It can only be reduced further by using a single mass in a non-differential setup—which would then, however, be subject to the full environmental translational acceleration.

Pendulum characterization. Because we aim to characterize the gravitational force exerted by our source mass, we quantify the test-mass motion by its spatial displacement rather than by the pendulum deflection angle. The amplitude spectrum of the test-mass displacement (shown in Fig. 2) follows the slope of the mechanical susceptibility of a damped harmonic oscillator, which is solely defined by its eigenfrequency f_0 , mechanical quality factor Q and mass m_{eff} . (Owing to the low frequency of the observation and the extremely good frequency separation of the mechanical modes of a torsion pendulum in general, all transfer functions except the mechanical susceptibility of the yaw mode can be neglected.) The eigenfrequency and mechanical quality factor were obtained by a fit to the spectrum as $f_0 = 3.59$ mHz and $Q = 4.9$. Both parameters were verified during measurement runs and independently by a ring-down measurement with relatively small amplitude. For larger amplitudes, we observe a sizable decrease in mechanical losses. Values in excess of $Q = 2 \times 10^4$ were observed for larger amplitudes ($-\pi$ rad) and different rest positions. The effective mass was determined by weighing and modelling all components except for the small amounts of glue that were used.

The drive frequency f_{mod} was chosen to lie well above the torsion resonance (~ 3.6 mHz), but far away from other, highly excited modes (for example, around 500 mHz and 700 mHz) and sufficiently below the readout noise (~ 100 mHz), such that the $2f_{\text{mod}}$ (and eventually the $3f_{\text{mod}}$) signal could be resolved. The reason for choosing 12.7 mHz is to prevent frequency mixing in nonlinear systems, such as discretization and clipping of narrow-band signals, to accidentally match with the gravitational signal.

By choosing the signal frequency in such a way, the test mass effectively behaves like a free mass (m_{eff}) (more precisely, if the signal frequency is well above the mechanical linewidth, $\gamma_0/(2\pi) = f_0/Q$, that is, $f_{\text{mod}} \gg \gamma/(2\pi) \approx 0.7$ mHz). Hence, in this regime the response is mostly independent of the resonant enhancement factor Q and the resonance frequency f_0 .

Prospects towards smaller masses. The experiment presented here verifies our approach to isolating gravitational coupling of 100-mg-sized objects. The achieved signal-to-noise ratio (SNR) of 6.04/0.06 ≈ 100 already allows measuring the gravity exerted by 1-mg-sized objects with SNR ≈ 1 for a measurement duration of ~ 150 h.

We note that our current measurement runs undergo diurnal variations in the noise floor, to which the thermal noise of our $Q = 4.9$ oscillator assembly poses a hard limit. To make better use of our measurement time ($\text{SNR} \propto T_{\text{meas}}$), we must improve our understanding and reduce the coupling of daytime environmental noise. Furthermore, mechanical quality values in excess of $Q = 2 \times 10^4$ have been observed even for this particular pendulum. Because the off-resonant thermal noise amplitude decreases with $1/\sqrt{Q}$, this provides a factor of $\sqrt{20,000/5} \approx 65$ improvement over the current sensitivity. Combined, these aspects show the potential of our approach for sensing the gravitational field of a Planck-mass-sized object ($m_p = 22$ μ g).

Calibration

Optical-lever readout. The angular position of the torsion balance is read out by means of an optical lever using 1.8 mW of laser light (In-nolight Mephisto) with an angle of incidence of ~ 50 mrad. The signal is obtained as the horizontal differential photo-voltage V_x of an off-the-shelf quadrant photodetector (QPD; Thorlabs QPD80A; sum voltage V_{sum}) with a high dynamic range (0.2–0.5 V_x/V_{sum}) and low noise ($\sim 2 \times 10^{-9} V_x/V_{\text{sum}} \text{ Hz}^{-1/2}$) in its approximately linear region of 0.5 mrad. The data are sampled by a Picoscope 4824 oscilloscope at 250 and 286 samples per second (anti-aliasing filter at 132 Hz) and digitally downsampled by a factor of 10. The analogue-to-digital converter (ADC) provides 12-bit resolution (± 10 V full scale, 4.9 mV resolution). Although we oversample the signal, the high-frequency quantization noise is improved when using even higher data rates. Reducing the full range would also help, but we require it to cover large low-frequency amplitudes, in particular during noisy times of data acquisition.

For future experiments, the actual (not quantization-noise-limited) readout sensitivity can be improved by using interferometric methods, although at the cost of limited absolute range. Systematic calibration errors can be substantially reduced by exchanging the current readout with a perfectly linear, self-calibrating auto-collimator at the expense of increased complexity and limited oversampling capability⁵⁰.

Position calibration. Position calibration of the test-mass motion was carried out by measuring the pendulum yaw deflection both via the optical lever and, independently, via video tracking. For the latter measurement, the torsion pendulum was recorded from the top with a resolution of 24 μ m per pixel, with both gold spheres being pattern-tracked to obtain the deflection (compare Methods section ‘Distance measurement’). Each method contributes its own error: for small amplitudes, the video-tracked deflection angle is relatively noisy, whereas at large deflections the optical transfer function of the QPD readout becomes nonlinear. For our calibration, the pendulum was excited to an amplitude of ~ 125 μ m, largely exceeding the linear detection range.

In the ideal case, the rotation angle and the normalized QPD signal are related via the Gauss error function. In our case, clipping and other effects degrade this relation at large amplitudes. Because our experiment is confined to small deflections, we require a good calibration only in the central QPD range. Therefore, a 3rd-order polynomial was fitted to the data surrounding $V_x/V_{\text{sum}} \approx 0.09$ to serve as yaw angle calibration for all our QPD data.

Force calibration and force noise estimate. A major feature of off-resonant detection is in our case the flatness of the force noise floor around the signal. This allows us to obtain a live estimate of the current noise conditions without additional sensors. Known narrow-band signals, such as f_{mod} and harmonics, are removed from the inferred force signal, and the instantaneous amplitude in the 5–75 mHz band is obtained by the absolute value of the signal plus its Hilbert transform. Specifically, evaluating amplitude (real part of complex signal) and change of amplitude (imaginary part) gives a continuous estimate for the noise coupling into the system. The goal of the algorithm is to

Article

identify excess excitation segments of the data that occur over a long time, in particular to minimize the data degradation from the sometimes up-to-tenfold increased noise conditions during daytime. We therefore low-pass-filter (moving average, $1/f_{\text{mod}} \approx 79$ s) the instantaneous amplitude and use the result for inverse-variance weighting of the measurement data. Compared to completely vetoing data during noisy conditions, this allows to make use of the full measurement time under these extremely non-stationary conditions. By employing this method, we could optimally combine a total of 350 h of non-stationary measurement time, which is worth ~150 h of higher-quality data, as presented in Figs. 2, 3, into the single evaluation shown in Fig. 4.

Non-stationary environmental conditions. Each data chunk contains up to 13 h 53 min of data (5×10^4 s). To display the time evolution of our noise, we perform a spectral density estimate on 1-h segments using Welch's method with 20-min-long segments and 50% overlap (Extended Data Fig. 2). (We note that the spectral resolution is inversely proportional to the integration time and therefore worse for the segmented spectra than in Fig. 2. In addition, the gravitational signal is smaller for these short segments because less of the signal is integrated. If we choose the amplitude spectrum representation instead, our signal will remain at the same height but the noise floor will rise, because less noise is averaged out.)

The low-frequency noise (below ~100 mHz) is dominated by the actual motion of our torsion pendulum. Although we reach the theoretical thermal noise limit of the damped harmonic oscillator during quiet times, we observe up to 10 times higher noise levels during other times of the day. The origin of this noise has not been fully identified yet (see also Methods sections 'Seismic noise' and 'Magnetic interaction' for more details). However, it always exhibits the same pattern: within 1-h segments, the noise appears to be white in terms of the force exerted onto the test mass, but its level varies slowly. The noise power is well correlated with the diurnal variability of anthropogenic noise, be it in higher-frequency seismic activity (evidenced by the excitation of the 0.7-Hz mode) or in 10-mHz magnetic fields. Nights are better than daytimes, even with nobody being present in the laboratory. Weekdays present worse conditions compared to weekends and public holidays. Therefore, the Christmas 2019 time was exceptionally well suited for the low-noise measurements presented here.

At higher frequencies, above ~100 mHz, the noise flattens out (except for higher-order modes) at a readout noise level of 2×10^{-9} m Hz^{-1/2}. Because this noise further improves with increased sampling frequencies, it is suspected to be quantization noise.

Data evaluation

Data handling. All data are sampled by a Picoscope 4824 at 250 or 286 samples per second using the internal clock. The ADC provides 12-bit resolution (± 10 V full scale, 4.9 mV resolution). A typical data trace spans 5×10^4 s, just short of 14 h. An analogue anti-aliasing filter at 132 Hz is applied to the horizontal difference of the photodiode quadrants to remove high 50-Hz harmonics (room light) and other narrow-band large-amplitude signals. The QPD signals are digitally de-rotated by ~20 rad, which maximizes the sensitivity to the torsion motion while minimizing the impact of pitch motion (vertical spot motion). The calibration procedure described earlier provides a signal corresponding to test-mass displacement. From this, the force signal is deduced as described below. At this point, the momentary noise estimation signal is deduced, processed with the other channels. Only after that, the signals are digitally down-sampled by a factor of 10 (a 10th-order Chebyshev type 1 low-pass filter with a cut-off frequency 0.8 times the new, reduced, sampling frequency $f_{s,\text{new}}$, with the response at f_{mod} normalized to unity, is applied before the decimation) to reduce the amount of data. The initial oversampling reduces the digitization noise (optimally by a factor of $\sqrt{10}$). Finally, the signals are cropped—that is, usually the first 5% are removed. Besides eliminating general edge

effects of any filters, this mainly reduces the ring-in effect of the force estimation.

Force inference. In general, a damped harmonic oscillator can be represented by a 2nd-order low-pass filter. The mechanical susceptibility $\chi(\omega)$ is defined as the transfer function from the externally applied force $\tilde{F}_{\text{ext}}(\omega)$ to the induced displacement $\tilde{x}(\omega)$,

$$\chi(\omega) = \frac{\tilde{x}(\omega)}{\tilde{F}_{\text{ext}}(\omega)} = \frac{1}{m_{\text{eff}}}(-\omega^2 - i\gamma\omega + \omega_0^2)^{-1}. \quad (1)$$

At a given frequency ω , χ is defined by the (viscous) damping rate $\gamma = \omega_0/Q$, the (un-damped) resonance frequency ω_0 and the effective mass m_{eff} . We map this to a zero-pole gain (zpk) representation

$$\chi \stackrel{\dagger}{=} k(s - p_0)(s - p_0^*) \quad (2)$$

(where the notation $\stackrel{\dagger}{=}$ stands for 'must be mapped to') in the Laplace domain ($s = a + ib$, where a and b are real numbers and we choose $a = 0$) by means of a pair of complex conjugate poles (p_0, p_0^*) and a d.c. response k (which contains the unit conversion). We find

$$k = \frac{1}{m_{\text{eff}}}, \quad p_0 = \frac{\gamma}{2} \pm i \sqrt{\omega_0^2 - \frac{\gamma^2}{4}}. \quad (3)$$

As long as the zpk representation of a system is known, then the time-domain response $x(t)$ can be inferred from the driving force $F_{\text{ext}}(t)$ without solving the equation of motion, by the following recipe; for example, using the MATLAB built-in functions `zp2tf()`, `bilinear()` and `filtfilt()`:

1. convert the zpk model with zeros z and poles p to its transfer function form defined by the numerator `num` and the denominator `den` by `[den, num] = zp2tf(z, p, k)` (appropriate k must be chosen to take into account the re-normalization by placing poles/zeros),

2. convert a continuous time model to filter the coefficients `bd`, ad of a discrete time model sampled at frequency f_s by `[bd, ad] = bilinear(den, num, f_s)`,

3. apply the filter by `x = filtfilt(bd, ad, F)`.

Whereas steps 1 and 2 are more or less technical re-definitions, step 3 convolves the force time series with the impulse response in a very resource-friendly manner.

We use this method in reverse by inverting the mechanical susceptibility, which is equivalent to exchanging poles and zeros ($p_0 \rightarrow z'_0$, $z_0 \rightarrow p'_0$ and $k \rightarrow 1/k'$) to infer the force F_{ext} required to produce the measured displacement $x(t)$. To avoid an unphysical (infinite) response at or above the Nyquist frequency $f_N = f_s/2$ from the two zeros, we must limit the response by a pair of compensation poles. By allowing them to be complex as well, their phase loss around the detection band can be limited. In step 2, we apply pre-warping at 100 mHz to reduce malicious effects onto the transfer function. The compensation poles are chosen to lie at $\omega_{\text{comp}} = 2\pi \times 1$ Hz with $Q_{\text{comp}} = 3$. This choice is verified not to influence the measured gravitational signal considerably (compensation pole in Extended Data Table 2).

Data visualization. The spectrum estimate shown in Fig. 2 was produced by Welch's method (`pwelch()`). Each time the series is divided into five blocks with 50% overlap and tapered by a Hanning window. The spectrum estimates of the blocks are then averaged. Hence, the noise floor in the figure is elevated but smoothed out compared to that achieved by performing a single spectral estimate on the whole trace at once. The residuals shown in the plot are a spectral estimate of the difference of the measured force signal (red trace) and the expected/nominal gravity signal. By performing the subtraction before the spectral (power) estimate, these residuals become sensitive to the relative phase of the two signals.

Because our goal is to quantify a (nonlinear, but stationary) relation between the source–test mass separation and the exerted force, plotting one against the other comes with the implicit assumption of equal significance for each data point. However, because our momentary noise estimation channel enables us to assess the significance (in terms of SNR) of each point, we chose to weight the data points in Fig. 3 with this significance to obtain the most general visual representation of the force versus separation relation. The hidden assumption here is that the momentary noise level estimation procedure does not introduce systematics into the quantitative data evaluation (see next step). We tried to verify this by checking that there is no quantitative dependence of the results on noise-estimation filter parameters.

Only after this most general data representation, the relation is subjected to assumptions using Newton’s law to quantify the relation. Hence, the width of the signal band is determined by all noise within the bandwidth of evaluation (5–75 mHz bandpass; see above) and not only at the signal frequency or its harmonics. Although subjecting data observed in dedicated high-precision experiments to the rigid corset of assumptions (fitting parameters) is a well established method, we believe the least-assumptions data representation approach of Fig. 3 to be extremely valuable in the yet-unexplored regime of gravitational sensing.

Data evaluation. To quantify the measured gravitational force, we assume the validity of Newton’s law and treat the masses as point masses concentrated at their respective centres of mass. Strictly speaking, this assumption is valid only for perfectly rotationally symmetric masses. Possible deviations due to bumpiness are considered as systematic errors. We use the gravitational constant as a fitting parameter resembling the strength of $1/x_s^2$ coupling. We further allow for a distance-independent force resembling the measured force averaged over one signal period, given that the experiment was not designed to measure the d.c. attraction, that is, the force with relation to the source mass being infinitely far away.

The statistical uncertainty of the data can be inferred by resampling. While often bootstrapping is used, it turns out that without further pre-conditioning, high-frequency temporal correlations such as high-frequency vibrational modes spoil the results as the data points cannot be regarded as sufficiently independent. Subsampling, by contrast, holds under weaker assumptions and appears to work for our data because short-term correlations remain within the same segment.

Each time trace is cut into N equally long segments. For each segment, the coupling strength G_i is fitted with the inverse-variance weight factors described in the previous section in non-stationary environmental conditions. Furthermore, a common force offset F_0 is fitted to all segments (allowing each segment to have its own d.c. force $F_{0,i}$ leads to similar results, but is less physical, because this force should not drift). The weighted mean $G = \sum_i G_i / \sigma_i^2$, where $1/\sigma_i^2$ are the normalized inverse-variance weights for the segment, gives the same result as a weighted fit to the whole trace. From the (inverse-variance-weighted) standard deviation $\sigma_w(G_i)$ we infer the statistical uncertainty for each G_i . The uncertainty for the whole trace is then inferred by means of $\sigma(G) = \sigma_w(G_i) / \sqrt{N_{\text{eff}}}$, where N_{eff} is the effective number of segments with independent noise contributing to the weighted standard deviation. The number N of segments per trace is varied between 11 and 30. For small N , drifts within the segments and non-stationarity of noise within the segments may become sizable, whereas for large N the exact resonance frequency and quality factor of the torsion resonance gain importance as the segment length approaches a torsion period. To exclude such effects, for each trace the results are verified to not depend on N , in order to avoid introducing systematic errors by its exact choice, and the fit parameters are averaged over the full range of segmentations.

This evaluation gives the independent estimate of the coupling constant and an estimate of the statistical uncertainty for each trace presented in Fig. 4. All deduced coupling constants agree well within

their statistical error budgets. A coupling constant with even lower statistical uncertainty is obtained by averaging the coupling constants of all traces while using the respective deduced statistical uncertainties as inverse-variance weights. The value of the time-resolved representation of Fig. 4 is that it shows that there are no major (monotonous) time-dependent systematics such as adsorption and continuous electrical charging.

Vacuum

The experiment was conducted in high vacuum (10^{-6} – 10^{-7} hPa) to reduce excitation from residual gas molecules and direct momentum transfer. The vacuum chamber is pumped by a turbo molecular pump (Pfeiffer TMU 071P) running at 25 Hz during measurement runs. These narrow-band vibrations were not observed to produce sensitivity degradation of the gravitational signal. The pressure could not be monitored permanently during measurement runs, because the Pirani pressure gauge (Pfeiffer PKR 251) needed to be detached from the chamber; it was observed to constantly charge the torsion balance when in use, and its relatively strong magnetic field was suspected to cause magnetic interaction between the gold masses.

The effects of residual gas molecules cause a Brownian force noise that is derived as an additional damping rate⁵¹

$$\gamma_{\text{wall}} = \frac{P}{m v_{\text{air}}} r^2 \frac{\pi^{3/2} r^2}{\sqrt{2} d^2 \ln\left(\frac{r^2}{d^2} + 1\right)} \quad (4)$$

(P , pressure; $v_{\text{air}} = k_B T / m_{\text{air}}$, average velocity of ‘air molecules’ of mass m_{air} ; d , separation between the surface of the cylinder and the electromagnetic shield; r , cylinder radius; m , cylinder mass; k_B , Boltzmann constant; T , temperature). The additional force noise damping due to the proximity to the electromagnetic shield is taken into account. Moreover, we assume the worst-case scenario of a cylindrical test mass, which would trap considerably more gas molecules compared to a sphere of similar size close to the wall. Using our experimental parameters ($P = 6 \times 10^{-7}$ mbar, $m_{\text{air}} = 4.8 \times 10^{-26}$ kg, $r = 1$ mm, $m = 91$ mg, $d = 150$ μm , $T = 300$ K), we obtain an additional damping of $\gamma_{\text{wall}} \approx 5.1 \times 10^{-8}$, which is much smaller than the natural damping rate, $\gamma_m = \omega_0 / Q \approx 5.0 \times 10^{-3}$, of our test-mass oscillator. Looking ahead, even for a projected experiment in the Planck-mass regime, which would require a mechanical quality factor of $Q > 20,000$, the effect of the force noise from gas collisions would be at the 5% level in this pressure regime—and can be further reduced by pushing vacuum levels into the ultrahigh-vacuum range.

Vibration isolation

The vacuum chamber is placed on an optical table with deflated air springs. Although inflating the springs did attenuate high-frequency noise, as expected⁵², the pendulum yaw mode, by contrast, was observed to get heavily excited. We suspect a relation to the table tilt, which we independently observed to couple strongly into the experiment (3 rad of yaw motion per 1 rad of tilt motion). In the inflated state, air density fluctuations act differently onto the air springs and can thereby cause excessive table tilt at low frequencies, which could explain the excessive yaw excitation. Operating our experiment with inflated air springs will require an active tilt stabilization system, as described in ref. ⁵³.

High-frequency vibrations are known to couple nonlinearly into lower-frequency bands, where the measurement takes place⁵³. To decouple the pendulum from such vibrations, for example, such as those generated by the turbo pump, the 420-g pendulum support structure rests on soft vacuum-compatible rubber feet (Viton).

Source-mass position drive

The position of the source-mass gold sphere is modulated by a long-range bending piezo (PI PL140.10, PICMA Bender). Applying a

Article

voltage of 0–60 V bends the piezo and provides a travel range of the order of 1 mm at the piezo tip, which is connected to a 0.31-mm-diameter non-magnetic titanium rod carrying the source mass (Extended Data Fig. 3). The rod is inserted only half way into the source mass so that its contact potential that faces the test mass is effectively shielded by the surrounding gold.

Upon actuation, the piezo bending is translated to a mostly linear source-mass motion of up to 5 mm and a rotation around its centre, which can be neglected. The position of the source mass is modulated harmonically to avoid d.c. noise of the torsion pendulum such as slow thermal drifts.

Distance measurement

A major source of uncertainty for the measurement of the gravitational coupling G is the absolute separation between source mass and test mass. We measure the distance with a digital single-lens reflex (DSLR) camera (Canon EOS 60D) that is mounted on top of the vacuum chamber, looking down vertically onto the experiment. This means that only a two-dimensional (2D) projection of the assembly can be monitored. Videos are taken with a 60-mm macro lens and extension tubes. The positions of the masses are tracked using a template-matching algorithm based on the open-source image and video processing library OpenCV.

At the beginning of each distance measurement, templates of the masses are defined in the first frame of a video. Their positions are tracked in subsequent frames by integrating the squared pixel value distance over the dimensions of the template

$$R(x, y) = \sum_{i \in \{r, g, b\}} \left\{ \sum_{x', y'} [T_i(x', y') - I_i(x + x', y + y')]^2 \right\}. \quad (5)$$

Here, $T_i(x, y)$ denotes the 8-bit pixel value of the colour channel i of the template image, and $I_i(x, y)$ the same for the video frame to be evaluated. The minimum $\min(R) = R(\tilde{x}, \tilde{y})$ then corresponds to the position of the best match (\tilde{x}, \tilde{y}) for the template. This method is limited to an accuracy of 1 pixel, which corresponds to about 24 μm . Therefore, for each dimension we perform a Gaussian fit over the five neighbouring pixels centred at the best-match position to improve the tracking resolution.

To limit the effect of changes in laboratory lighting, as well as the influence of reflexes on the source-mass surface, only the outer perimeter of the masses is tracked. We verify the tracking quality by visual inspection of the residuals—that is, by looking at a video with each frame realigned to the tracked position—similar to a digitally stabilized video. The resolution and accuracy of the tracking is limited by the (manual) determination of the edges of the masses in the template, the low contrast inside the vacuum chamber due to its gold-plated interior, and the resolution of the DSLR, and is taken to be 20 μm . We measure the separation of source and test masses via video at the beginning and the end of each measurement run.

Furthermore, the source–test mass separation is defined not only by the momentary source-mass position, but also by the motion of the residual test mass. In particular, the centre-of-mass motion of the pendulum varies the distance by more than 100 μm under bad conditions. Therefore, the gravitational force is averaged over these distances. The separation modulation changing from this test-mass motion cannot be extracted from the optical-lever signal to be calibrated out. In the future, continuous video tracking will provide this information.

Piezo calibration. The piezo is operated without feedback. Because the position cannot be monitored continuously, the piezo's position response to the applied voltage needs to be characterized and modelled. The response of the piezo can be modelled as an ideal mechanical transducer combined with an additional low-pass filter⁵⁴. Fundamentally, this reflects electron mobility/diffusion into the piezo material.

As a consequence, the step response is delayed and the drive shows a slow creep towards its final position, as well as hysteresis, when driven cyclically. The creep (knowledge of the exact position) is of particular interest for the charge measurements, whereas the hysteresis (knowledge of the exact amplitude and phase) is important for the gravity measurements, because it results in the source-mass position lagging behind the applied voltage signal and reduces the amplitude. The stationary creep (the step response of the piezo) depends on the size of the applied voltage step and the starting position (the voltage levels applied to the piezo before and after) and can reach up to 30 μm over a period of 15 min. As the creep depends on parameters over which we have little control, the creep is fitted for every measurement separately rather than using global piezo parameters. The observed time lag depends on the drive frequency, but does not show any dependence on the modulation amplitude. For a modulation frequency of 12.7 mHz we find a time lag of 509 ± 3 ms by correlating the drive voltage with the source-mass position as measured by video tracking.

We also observe a drift over time in both the mean position and the modulation amplitude of the source-mass drive. A possible reason is continuous charging of the piezo material, similar to the observed creep for a step response in quasi-static operation, even in dynamic operation. In contrast to the stationary case, in which creep-induced piezo deflection reaches a final state after about 1 h, under cyclic operation the mean deflection and amplitude continue to drift over a measurement period of 100 h. For a peak-to-peak modulation of about 2 mm, the amplitude is reduced by 13 μm and the mean position drifts by 55 μm . Although this long-term drift resembles the stationary creep as described in ref.⁵⁴ with a much longer time constant, we cannot fit the corresponding function with only the start and end videos. The drift of the mean position, as well as the modulation depth, therefore, are taken to be linear over the period of a measurement run.

Additionally, the source-mass position modulation exhibits higher-harmonics contributions. We measure the relative amplitudes and phases of these higher harmonics with respect to the fundamental mode in the start and end videos. We infer the actual position modulation for the gravitational measurement runs from the drive signal by taking the time lag, the amplitude change and mean position drift, and the higher harmonics into account. For example, the amplitude of the 2nd harmonic of the drive position is fitted and found to be below 0.3% of the fundamental amplitude. Remaining higher-order contributions are estimated to be below <0.1% and are therefore irrelevant for our experiment.

Charge measurements

We discharge the test mass to less than 8×10^4 elementary charges using a nitrogen gas discharge and ion diffusion as developed in ref.⁵⁵. The nitrogen gas is channelled over a set of high-voltage a.c. tips. After passing through an aperture, the nitrogen ions neutralize the surfaces of the vacuum chamber by diffusion. After discharging, the unshielded electrostatic interaction is still approximately three times stronger than gravity (see Extended Data Fig. 6). We note that the interaction is enhanced over a pure Coulomb interaction by induced surface charges of the source and test masses.

Because of that, even if the source mass is perfectly grounded, a charged test mass will induce charge separation resulting in parasitic attraction. Whereas most torsion experiments aiming to measure small forces conductively coat their suspension filament and thereby sacrifice mechanical quality (statistical error) for charge control (systematic error), we explicitly chose not to do so. One reason is that when down-scaling the torsion balance further, the proportion of lossy coating (surface) grows with respect to the fibre material (volume). The other reason is the long-term prospect to contactlessly levitate the test mass, which would prohibit the use of any electrical contacting.

It can be shown by finite-element simulations (Extended Data Fig. 5), however, that electrostatic coupling between source-mass position

and test-mass force can be suppressed well below 3% of the expected gravitational signal by means of a 150- μm -thick conductive Faraday shield inserted between the source and test masses (see Extended Data Fig. 6). The shield comes at the added benefit of reducing other electrodynamic interactions, such as short-range Casimir forces, as well.

The drive piezo is housed in a Faraday shield in order to suppress coupling of the applied electric fields onto the pendulum, for example, via residual charges or polarization effects. During an 8.5-h measurement without a source mass attached but with the piezo modulation active, we did not observe any test-mass actuation. Although this verifies the effectiveness of the shield, a full analysis of this drive piezo contribution to the systematic error of our experiment would require the same amount of effective integration time as in the gravity run, that is, a measurement run would have to last for weeks under quiet conditions as well.

Because long-term drifts from charge diffusion on non-conductive materials are observed to degrade our low-frequency sensitivity, all surfaces inside the vacuum chamber (except for optical viewports) are made from conductive materials and are gold-plated wherever possible to equalize surface potentials and mitigate thermo-electric potentials.

Charge characterization. Static charges on our test mass can be characterized by means of induced mirror charges. They can be quantified by a force versus separation measurement without an electrostatic shield and with the source mass being grounded as shown in Extended Data Fig. 4. This is effectively a macroscopic realization of a Kelvin probe force microscope, which is typically used to contact-free measure (surface) potentials. The theory for the electrostatic interaction, including polarization effects of extended conductive spheres, is presented in ref. ⁵⁶. In addition to the expected signals (electrostatic, gravity), we observe an unidentified coupling that is proportional to the sphere separation.

To overcome this and other experimental challenges, such as unknown contact potentials, we isolate the effect induced by residual test-mass charges. We therefore quantify them by biasing the source-mass potential and observing the test-mass charge versus source-mass potential interaction at a set of discrete separations. Extended Data Table 1 summarizes the results of the intermittent charge measurements over the whole duration of the measurements. A more detailed description of the charge determination will be presented in an upcoming publication. In conjunction with the electrostatic shield between the masses and our finite-element simulation, we thereby verify that electrostatic coupling is suppressed to below 3% of gravity over the whole measurement period.

Seismic noise

Even during the environmentally quiet Christmas season, our fundamentally thermal-noise-limited sensitivity is dominated by a yet-unidentified contribution with a strong diurnal variation (see Extended Data Fig. 2). Reduction of this noise is important to advance towards smaller detectable source masses.

Long-period seismic ground motion in an urban environment is a major concern for extremely sensitive low-frequency experiments⁵⁷⁻⁵⁹. For ideal harmonic motion, the rotational mode of an oscillator will not be directly excited by translational ground motion. Unfortunately, unavoidable imperfections in the construction of the pendulum cause asymmetries that couple horizontal ground motion into rotational motion^{60,61}. Furthermore, down-conversion of noise at higher frequencies, such as cross-coupling from other modes, cannot be neglected³³.

To monitor horizontal and vertical ground displacement down to 100 mHz, a Raspberry Shake 3D seismometer was installed in close proximity to the experiment. Low-frequency seismic displacements at the experiment site were monitored with a RefTek Observer 60 s broadband seismometer. In addition, we have access to the broadband data of the ZAMG seismic station (Streckeisen STS2) at Hohe Warte, Vienna, roughly 3 km away from the laboratory.

Correlating strong seismic events, such as the 6.7-mag earthquake in Turkey on 24 January 2020 in the data of the local seismometer as well as the STS2, with the response of our oscillator provides a rough estimate of seismic coupling, which suggests that the diurnal low-frequency noise observable in our data is not due to linear coupling of horizontal or vertical ground motion.

This claim is supported by directly comparing the non-stationary low-frequency noise in the STS2 data in Vienna (Extended Data Fig. 7a) and the torsion balance (Extended Data Fig. 2). There is a strong variability of the seismic noise floor in the 30-mHz band but a relatively steady primary microseismic peak around 70 mHz. The daytime-dependent change in the 30-mHz noise floor is consistent in amplitude and time with the varying white force noise that we see in our test mass below 100 mHz. Yet, the frequency distribution is very different. In particular, the microseismic peak at 70 mHz is a very strong, stationary seismic feature within the frequency band of the experiment. Because we do not see it in the torsion pendulum data, we conclude that our rotational mode is not dominantly driven by linear coupling of the horizontal seismic noise.

Mechanical coupling

Recoil of the source-mass drive is of major importance for higher modulation frequencies such as 50 Hz (ref. ³²). Here we argue that, owing to its strong frequency dependence, this effect can be neglected for our experiment. A 100-kg freely suspended optical table, for example, would move by 1 nm per millimetre of recoil motion of a 100-mg source mass, which is of the order of our observed signal. The coupling into pendulum yaw is attenuated further by the pendulum's common mode rejection, that is, how well balanced the assembly is. Yet, already a table suspension resonance of 1 Hz, which can roughly be assumed for the air springs, attenuates this by another factor of $(f_{\text{mod}}/f_0^{\text{table}})^2 \approx 1/6,000$. A deflated table, by contrast, typically exhibits internal resonances well beyond the 100-Hz region, yielding even higher resistance to recoil.

Another possible coupling at low frequencies is the cyclic deformation of the environment under the changing load distribution of the drive setup. Again, we consider the 1-m-sized, 100-kg optical table suspended by air springs ($\sim 4 \times 1 \text{ kN m}^{-1}$ stiffness). A 1-mm horizontally shifted 100-mg mass causes 1 μN differential load and thereby induces a tilt of approximately 1 nrad. In conjunction with the abovementioned tilt-to-yaw coupling, this can result to a test mass motion of the order of 1 nm per millimetre of a 100-mg source-mass motion, which is of the order of the observed signal and dominating the recoil effect. In practice, the air springs employ active tilt feedback at low frequencies, stiffening the table tilt motion. When deflated, the table leg stiffness may increase to 1 MN m^{-1} , providing suppression of source-mass-induced table tilt by a factor of 1,000. Internal table flexing modes even in the low 100-Hz region would not allow increased tilt coupling at the modulation frequency. In the future, a more in-depth investigation of tilt-to-yaw coupling and cancellation of both effects by means of a counteracting mass will be investigated.

Therefore, both effects are indeed of importance for massive source masses and high accuracy measurements aimed to determine G very accurately. However, to our current understanding, they can be neglected in this experiment and in similar future low-frequency designs aiming at further reduced source masses.

Magnetic interaction

To minimize influence from magnetic coupling, the experiment was designed without ferromagnetic materials. All masses are made of diamagnetic gold, and the source mass is mounted on a paramagnetic titanium rod. As a consequence, Earth's magnetic field ($\sim 50 \mu\text{T}$ at our location), as well as anthropogenic magnetic noise in an urban environment, induce dipoles in the gold spheres, leading to a repulsive force between them.

Using literature values⁶² for the permeability of commercial-grade gold and titanium, we estimate dipole forces of

$$F_{\text{mag}} = \frac{6\mu_0 m_1 m_2}{4\pi d_c^4} \quad (6)$$

between two magnetic dipoles m_i at centre distance d_c , such as between our two diamagnetic spheres, as well as between the titanium rod and one gold sphere (μ_0 , vacuum permeability). The strength of the dipoles is determined by the strength of the externally applied field, the susceptibility, and the volume and geometry of the material.

Iron impurities in the gold are suggested to dominate the magnetic susceptibility⁶³. Yet gold–iron alloys act diamagnetically up to 1,000 ppm iron, paramagnetically up to 10% iron, and only then start acting as a ferromagnetic material. The gold used in the experiment is specified to have impurities below 1,000 ppm but, according to ref.⁶⁴, commercial-grade gold has iron impurities between 10 and 100 ppm. As a rough experimental verification for the permeability, a 1-cm cubic NeFeB magnet with a remanent magnetization of around 1.4 T was brought within 7 mm of the test and feedback masses. The literature value of the magnetic susceptibility χ_{Au} is confirmed within a factor of -4.

Projecting the dipoles induced by Earth's magnetic field to the exerted forces by means of equation (6) shows that they are about four orders of magnitudes weaker than gravity in this setup.

To estimate magnetic signals induced by anthropogenic fields, we measure the magnetic field close to the experiment with a three-axis magnetic sensor (Stefan Mayer FLC3-70) and two different Samsung phones containing fluxgate magnetometers (see Extended Data Fig. 7b). The 10-mHz frequency comb is correlated with a close-by traffic light period. We verify that this is a magnetic signal by measuring it with different magnetometers using different sampling ADCs. Moreover, the signal strength does not depend on the exact location in the laboratory. Yet, it is not visible in the force signal of the experiment. We therefore conclude that the masses do not couple magnetically via dipoles induced by anthropogenic magnetism.

Casimir force

The steady-state Casimir force for a plane–sphere geometry can be calculated⁶⁵ using

$$F_{\text{Casimir}} = \frac{R_t k_B T \zeta(3)}{8d_s^2}, \quad (7)$$

where ζ is the Riemann zeta function. For our setup, the parameters are $T = 300$ K, test-mass radius $R_t = 1$ mm and plane–sphere distance $d_s \approx 250$ μm . Thus, the Casimir force between the test mass and the Faraday shield is more than four orders of magnitude smaller than Newtonian gravity⁶⁶. For extending the experiment towards smaller source masses, however, the Casimir force will dominate the gravitational attraction of a Planck-mass-sized source object for shield–sphere separations below ~ 100 μm . Because in our approach the gravitational potential is modulated, even in that source-mass regime we will be able to distinguish Casimir forces exerted onto the test mass (which are static in the presence of an electrostatic shield) from gravity. Furthermore, for reduced shield separations, the Casimir force might become relevant as a static nonlinearity or instability of the torsion balance potential. For a 183-mg oscillator with a natural frequency of 1 mHz, the Casimir force will render the oscillator unstable by surpassing the restoring force only for surface distances below ~ 100 nm.

For any setup without a Faraday shield, Casimir forces set a lower bound for test–source mass separation, which in turn sets an ultimate lower limit on the detectable Newtonian force. For our scheme, these limits are approximately 2 μm for the symmetric setup with 90-mg masses and 20 μm for a source object with $m = m_p$.

Systematic deviations

Here we list identified sources of systematic over- or underestimation of the measured gravitational force. Neither of these has been corrected for in the evaluation, given that most of these can only be guessed. In Extended Data Table 2 we assess the most relevant contributions considered. In general, we find that most effects scale with the source mass. Hence, they limit the accuracy of our G estimation but will not pose a limit to measurements of smaller source masses in the future.

All involved masses except for the tiny amounts of glue used were measured at an accuracy of 0.1 mg (Kern ABS120-4). Although in the evaluation we treated our source and test mass as point sources of gravity, they also act gravitationally onto the other components, such as the capillary and glue. Gravity between the source-mass support rod (titanium gravity) and the test mass is numerically simulated by integrating over $(10 \mu\text{m})^3$ elements. Because the mass distribution of the complete source mass assembly deviates from spherical symmetry, it cannot be summarized as an effective point mass co-moving with the source mass. Instead, the modulation caused over a centre-of-mass separation of 2.4–6 mm is put into relation to the gravity modulation caused by the source- or test-mass modulation over the same range. Similarly, the influence of the spatial extent of the hole used for mounting the titanium rod inside the source mass is estimated. The difference between a spherically symmetric source mass with a 300- μm -diameter hole and a spherically symmetric source mass without a hole but with the same mass is evaluated and put into relation with the nominal source- or test-mass gravity. The modulation of gravity between the counterbalance mass of the torsion pendulum and the source mass is quantified over the same 2.4–6 mm range. The calibration of the QPD signal was carried out once on best-effort basis and the results are shown in Extended Data Fig. 1. The quality was checked by varying assumptions, but cannot be tested independently or continuously. The source drive position is calibrated to an accuracy of 1 pixel ($\sim 20 \mu\text{m}$) at the beginning and end of each measurement run. The accuracy of the interpolation in between cannot be quantified, because the camera could not be operated continuously. Instead of only a 2D projection, the three-dimensional (3D) mass separation will be monitored continuously in the future, improving the data for close approaches, such as in Extended Data Fig. 4, by giving height offset and high-frequency pendulum and roll mode excitation information. The 3D mass distribution of our gold 'spheres' is not known, given that we can only observe 2D projections. From microscopic imaging under different angles, the masses used are known to be bumpy (source-mass radii vary between 1.03 and 1.11 mm), whereas in high-accuracy experiments for the determination of G , even the slightest shape and density deviations are known to be relevant.

The masses used in the experiment deviate from perfect spheres. To quantify the influence of these shape deviations, we took photographs of the test sphere under a microscope and at different angles, as shown in Extended Data Fig. 8. We extract the angle-resolved radius from the geometric centre of mass of each projection. We assume that each combination of angle and radius defines a spherical wedge and we calculate the 2D deviation of the combined centre of masses of the wedges from the geometric centre of mass of the projection. We repeat the evaluation for every projection and take the mean deviation from the geometric centre of mass as a correction for the centre-of-mass distance used in the evaluation. Comparing the fits to the uncorrected and corrected centre-of-mass distance results in an estimated systematic uncertainty of 4.2% for both test- and source-mass shapes.

An additional damping rate caused by the impacts of residual gas molecules is estimated according to ref.⁵¹. Using our experimental parameters and taking into account the close proximity of the electromagnetic shield to the test mass, we obtain a damping rate five orders of magnitude smaller than the natural damping rate of our test-mass oscillator.

A worst-case estimate for sorption effects on uncleaned gold surfaces^{67,68} results in a relative mass change of the order of 10^{-7} and can therefore be neglected. We also note that we do not observe time-dependent systematic effects, in agreement with the absence of substantial sorption effects.

Data availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

50. Turner, M. D., Hagedorn, C. A., Schlamminger, S. & Gundlach, J. H. Picoradian deflection measurement with an interferometric quasi-autocollimator using weak value amplification. *Opt. Lett.* **36**, 1479–1481 (2011).
51. Schmoele, J. *Development of a Micromechanical Proof-Of-Principle Experiment for Measuring the Gravitational Force of Milligram Masses*. PhD thesis, Univ. of Vienna (2017).
52. Newport pneumatic optical table performance. Newport <https://www.newport.com/n/compliance-and-transmissibility-curves>
53. Lewandowski, C. W., Knowles, T. D., Etienne, Z. B. & D'Urso, B. Active optical table tilt stabilization. *Rev. Sci. Instrum.* **91**, 076102 (2020).
54. Displacement of open loop piezo actuators. PI <https://www.pi-usa.us/en/products/piezo-motors-stages-actuators/piezo-motion-control-tutorial/tutorial-4-20/>
55. Weiss, R. *Charging of the Test Masses Past, Present and Future*. LIGO Document T1100332 (2011); <https://dcc.ligo.org/public/0115/G1401153/002/charging.pdf>
56. Lekner, J. Electrostatics of two charged conducting spheres. *Proc. R. Soc. A* **468**, 2829–2848 (2012).
57. Díaz, J., Ruiz, M., Sánchez-Pastor, P. S. & Romero, P. Urban seismology: on the origin of earth vibrations within a city. *Sci. Rep.* **7**, 1 (2017).
58. Groos, J. C. & Ritter, J. R. R. Time domain classification and quantification of seismic noise in an urban environment. *Geophys. J. Int.* **179**, 1213–1231 (2009).
59. Bourdillon, A., Ropars, G., Gaffet, S. & Le Floch, A. Opposite sense ground rotations of a pair of Cavendish balances in earthquakes. *Proc. R. Soc. A* **471**, 20140997 (2015).
60. Shimoda, T., Aritomi, N., Shoda, A., Michimura, Y. & Ando, M. Seismic cross-coupling noise in torsion pendulums. *Phys. Rev. D* **97**, 104003 (2018).
61. Gettings, C. & Speake, C. An air suspension to demonstrate the properties of torsion balances with fibers of zero length. *Rev. Sci. Instrum.* **91**, 025108 (2020).
62. Lide, D. R. *CRC Handbook of Chemistry and Physics* Vol. 85 (CRC Press, 2004).
63. Shih, J. W. Magnetic properties of gold-iron alloys. *Phys. Rev.* **38**, 2051–2055 (1931).
64. Henry, W. & Rogers, J. XXI. The magnetic susceptibilities of copper, silver and gold and errors in the Gouy method. *Philos. Mag.* **1**, 223–236 (1956).
65. Sushkov, A., Kim, W., Dalvit, D. & Lamoreaux, S. Observation of the thermal Casimir force. *Nat. Phys.* **7**, 230–233 (2011).
66. Emig, T., Graham, N., Jaffe, R. L. & Kardar, M. Casimir forces between arbitrary compact objects. *Phys. Rev. Lett.* **99**, 170403 (2007).
67. Beer, W. et al. The METAS 1 kg vacuum mass comparator-adsorption layer measurements on gold-coated copper buoyancy artefacts. *Metrologia* **39**, 263 (2002).
68. Gläser, M. & Borys, M. Precision mass measurements. *Rep. Prog. Phys.* **72**, 126101 (2009).

Acknowledgements We thank E. Adelberger, A. Buikema, P. Graham, N. Kiesel, N. Klein, D. Racco and J. Schmöle for discussions. We are grateful for the suspension fibre provided by A. Rauschenbeutel and T. Hoinkes and for the mechanical design assistance by M. Dragosits. This project was supported by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement number 649008, ERC CoG QLev4G), by the Austrian Academy of Sciences through Innovationsfonds Forschung, Wissenschaft und Gesellschaft, by the Alexander von Humboldt Foundation through a Feodor Lynen Fellowship (T.W.) and by the Austrian Federal Ministry of Education, Science and Research (project VCQ HRSM).

Author contributions T.W., J.P. and M.A. conceived the experiment, T.W., H.H. and J.P. constructed and conducted the experiment and analysed the results. All authors wrote and reviewed the manuscript.

Competing interests The authors declare no competing interests.

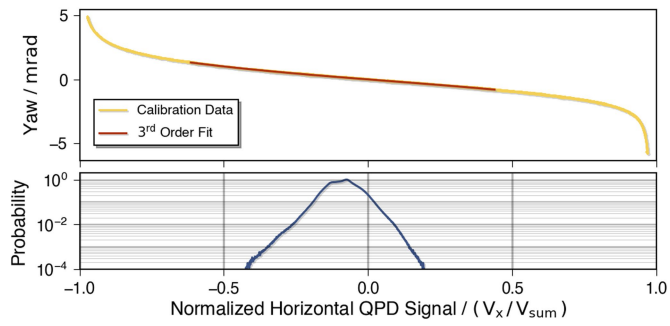
Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41586-021-03250-7>.

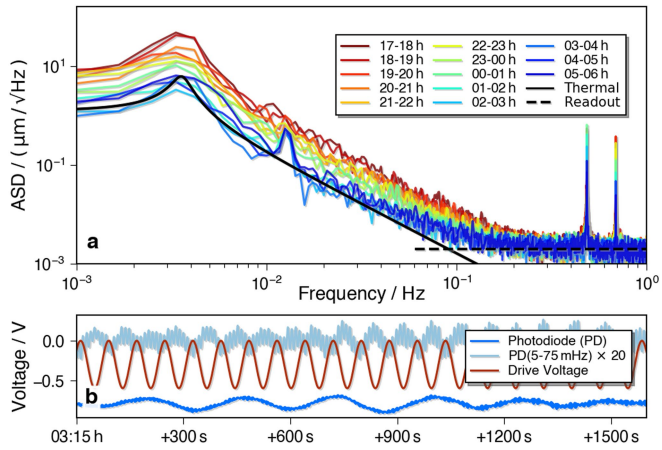
Correspondence and requests for materials should be addressed to T.W. or M.A.

Peer review information *Nature* thanks the anonymous reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

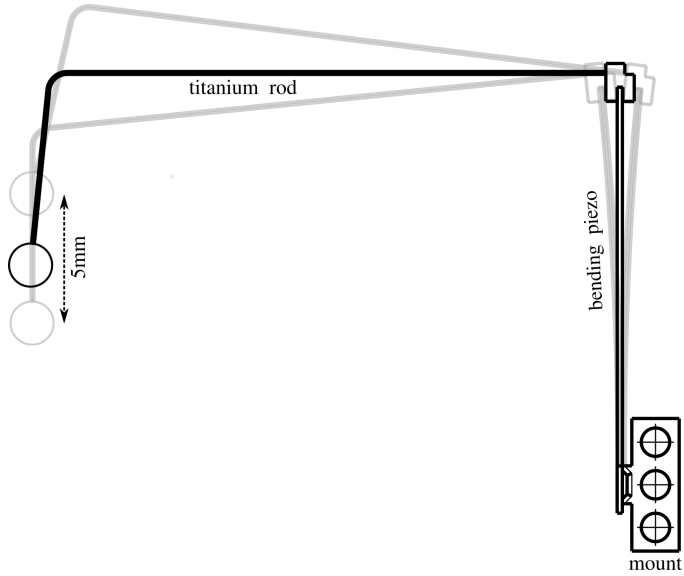
Reprints and permissions information is available at <http://www.nature.com/reprints>.



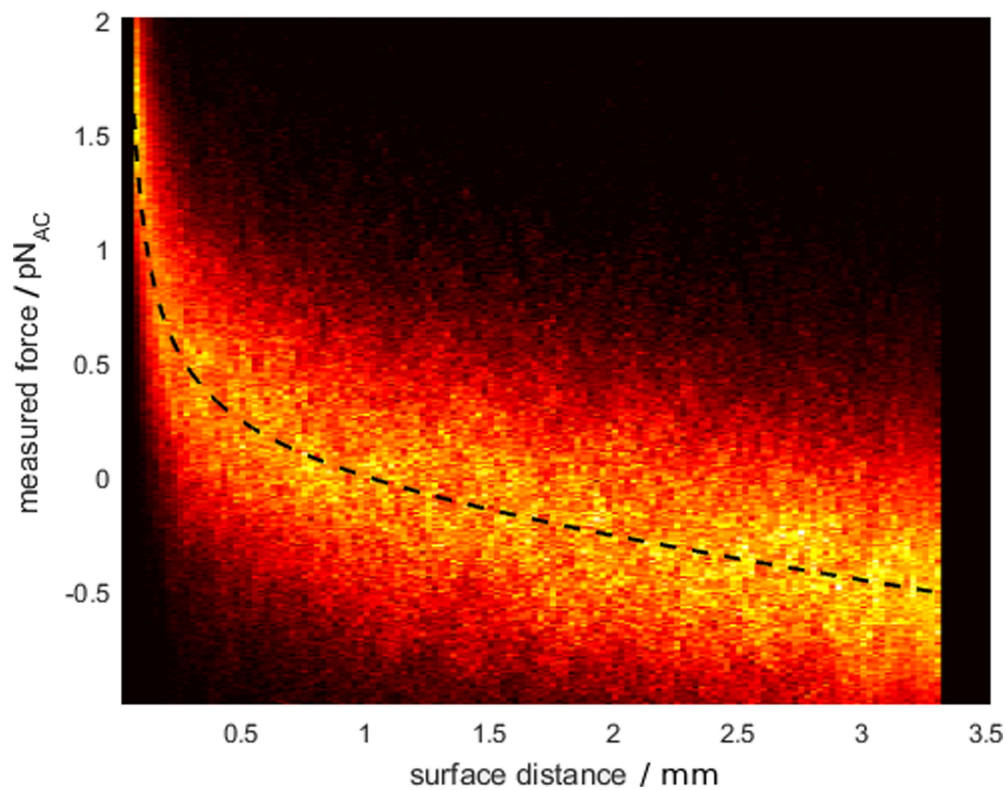
Extended Data Fig. 1 | Quadrant photodiode calibration. The video-tracked yaw angle of the torsion pendulum is compared to the QPD readout (horizontal difference signal normalized by the total sum signal to reduce laser intensity noise coupling). A third-order polynomial fit to these data provides our QPD signal calibration. The occurrence probability accumulated over all our data runs shows the region in which a calibration is required, that is, $V_x/V_{\text{sum}} \in [-0.4, 0.2]$.



Extended Data Fig. 2 | One-hour displacement spectra. a. The displacement amplitude spectral density (ASD) estimate of 1-h data chunks of representative data taken between 26 and 27 December shows stationary readout noise at high frequencies and non-stationary white force noise at low frequencies. The level of the latter varies up to tenfold, consistently being the lowest at 0–5 a.m., particularly in the night after public holidays or Sundays, that is, before normal workdays. During such quiet periods, the thermal noise of the pendulum is reached. **b.** Time-series data spanning half an hour during the 3–4 a.m. segment. The data trace is bandpass-filtered in the frequency range 5–75 mHz with a 6th-order Butterworth filter and magnified by a factor of 20.

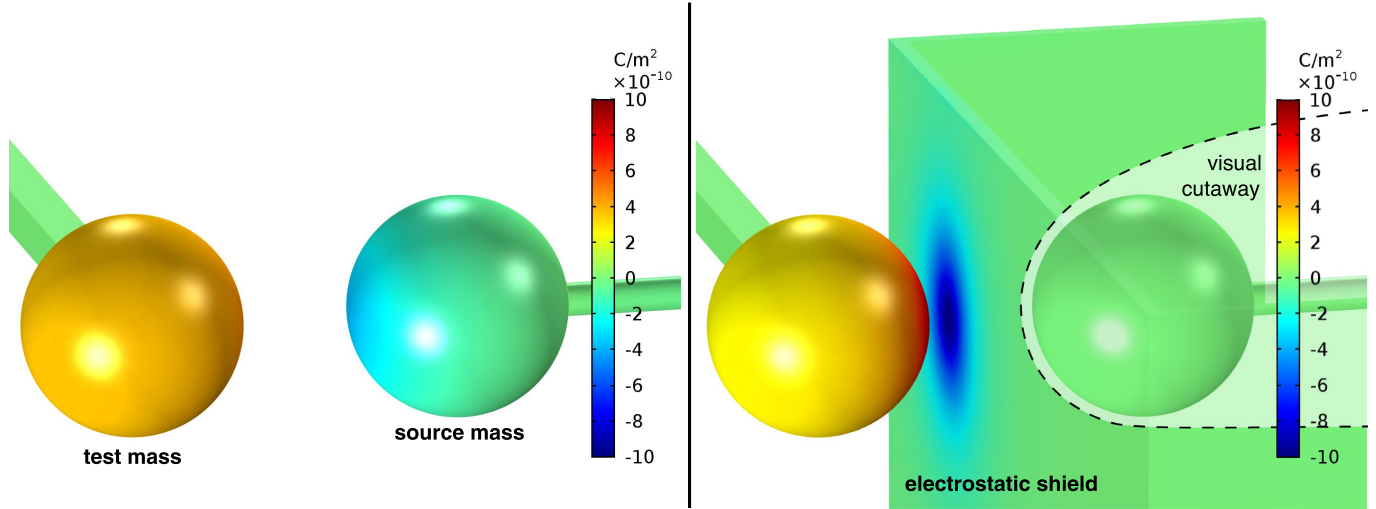


Extended Data Fig. 3 | Schematic of drive mechanism. The source-mass drive, consisting of a bending piezo and a titanium rod, can modulate the source-mass position by more than 5 mm peak to peak at frequencies as high as 1 Hz. The geometry of the titanium rod amplifies and translates the piezo deformation into an approximately linear motion.



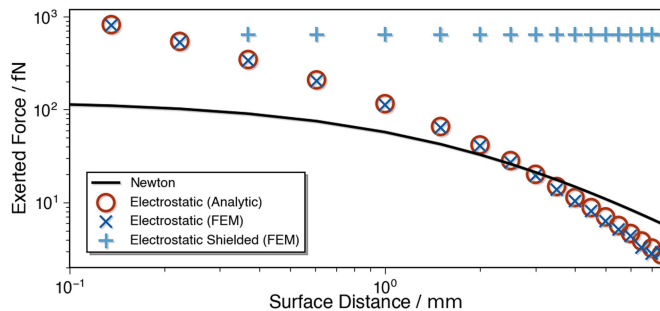
Extended Data Fig. 4 | Probability distribution of $F(d)$ without an electromagnetic shield. Without electrostatic shielding, the probability density distribution of inferred force versus source–test mass separation requires three constituents to describe: Newtonian gravity (see Fig. 3),

electrostatic interaction between charged ($-8 \times 10^4 e^-$; e , elementary charge) test mass and grounded source mass⁵⁶, and an unexplained force proportional to the separation.

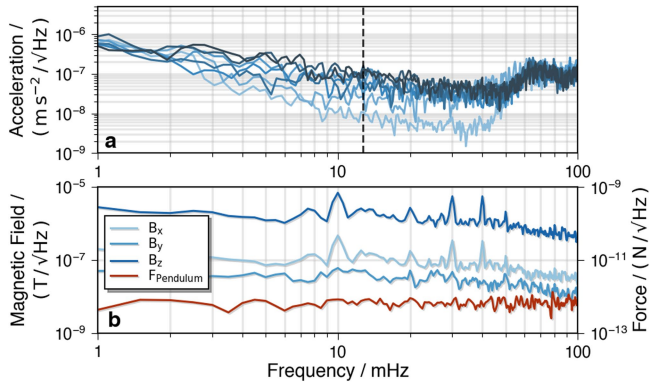


Extended Data Fig. 5 | Finite-element method simulation of charge distribution. The charge distribution induced by a $3 \times 10^4 e^+$ charged source mass on the grounded test mass was determined by a finite-element simulation (COMSOL; shown for 1 mm surface separation as an example). With the grounded, conductive electrostatic shield in place, many mirror charges are

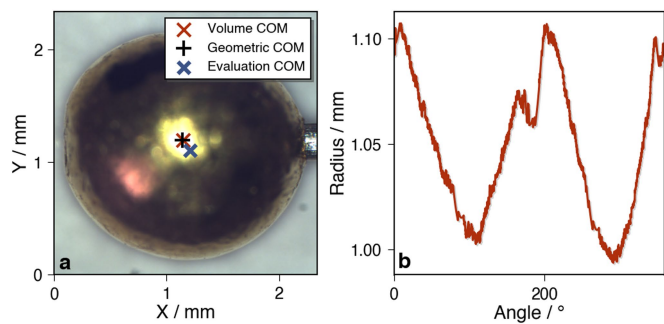
induced in the unmodulated shield, resulting in a d.c. force. The induced surface charge density (colour-coded) on the position-modulated source mass is suppressed by a factor of approximately 100. These charges are screened by the shield, resulting in further suppression of the exerted force. The actual suppression could not be quantified owing to numerical inaccuracies.



Extended Data Fig. 6 | Simulation of electrostatic coupling. Our finite-element simulation of the electrostatic interaction of a charged test mass ($10^5 e^-$) and a grounded source mass was validated using the analytical method described in ref. ⁵⁶. When inserting a 150- μm -thick, conductive electrostatic shield between them, the electrostatic force exerted onto the test mass is dominated by the test mass–shield interaction, that is, it becomes independent of the source-mass position. Residual fluctuations of the force are on the 3–10% level of gravity, without a clear source position dependence, and are expected to stem from numerical errors.



Extended Data Fig. 7 | Magnetic and seismic coupling. **a**, The time-resolved amplitude spectral density of the horizontal ground motion in Vienna, recorded by an STS2 broadband seismometer, shows a strong frequency dependence of the variability. A rise of the noise floor is observed throughout the frequency range, starting from midnight (dark blue) to noon (light blue), but is most prominent at 30–40 mHz. By contrast, the noise floor at the microseismic peak, at around 70 mHz, varies only slightly. The modulation frequency of the drive mass is marked by a vertical dashed line. Each spectrum corresponds to a single 2-h measurement. **b**, The amplitude spectral density of magnetic field variations at the site of the experiment were measured with a three-axis magnetometer (Stefan Mayer FLC3-70; z = vertical, y = along source–test mass axis). The signal at 10 mHz is explained by the period of the traffic light at a nearby crossing, regulating car traffic as well as trams. The inferred force acting on the pendulum (red line), shows no sign of this signal. Each spectrum corresponds to a single 8,000-s measurement.



Extended Data Fig. 8 | Shape of test-mass sphere. a, 2D projection of the test mass. The centre of mass determined by the geometric centre of mass (mean over all image dimensions), as well as the centre of mass calculated by our shape evaluation, are indicated ('Volume COM'). The centre of mass used in the determination of G is indicated by a blue cross. **b**, The radius measured from the geometric COM over all angles is shown.

Article

Extended Data Table 1 | Surplus charges over time

	12/28/2019	12/23/2019	12/27/2019	01/03/2020	01/07/2020	01/15/2020
surplus charges ($\times 10^3 e^+$)	34.7	25.0	44.4	78.8	38.8	52.8
accuracy ($\times 10^3 e^+$)	0.7	2.6	9.4	0.4	3.9	1.1

The surplus charges on the test mass were measured between our measurement runs during the time periods marked by grey bands in Fig. 4. All measured charges stay well below $8 \times 10^6 e^+$, which we take as an upper limit in our estimation of electrostatic coupling between source and test mass.

Extended Data Table 2 | Identified systematics

effect	influence / F_G
hole gravity	+8.4E-3
titanium gravity	+6.1E-3
capillary gravity	+4.5E-3
glue gravity	+3E-3
counterbalance grav.	-1.1E-4
electrostatic	$\pm 3E-2$
magnetic	$\pm 1E-4$
calibration	$\pm 6E-3$
drive calibration	$\pm 1.6E-2$
mass separation	$\pm 1E-3$
m_s accuracy	$\pm 1.1E-3$
m_t accuracy	$\pm 1.1E-3$
m_{glue} accuracy	+3E-3
height offset	$\pm 1.5E-2$
m_s roundness	$\pm 3.2E-2$
m_t roundness	$\pm 1.0E-2$
Q accuracy	$\pm 5E-3$
bandpass $1f_m$	+1.6E-2
bandpass $2f_m$	-7.6E-2
downsampling	$\pm 1E-3$
compensation pole	$\pm 1E-5$
anti aliasing filter	$\pm 2E-6$
upper limit	+15.9E-2
lower limit	-11.8E-2

Identified systematic deviations in units of expected Newtonian gravity between source and test mass.