Report on the CCM key comparison of kilogram realizations

CCM.M-K8.2019

Final Report

16 October 2020

M. Stock[1], P. Conceição[1], H. Fang[1], F. Bielsa[1], A. Kiss[1], L. Nielsen[2], D. Kim[3], M. Kim[3], K.-C. Lee[3], S. Lee[3], M. Seo[3], B.-C. Woo[3], Z. Li[4], J. Wang[4], Y. Bai[4], J. Xu[4], D. Wu[4], Y. Lu[4], Z. Zhang[4], Q. He[4], D. Haddad[5], S. Schlamminger[5], D. Newell[5], E. Mulhern[5], P. Abbott[5], Z. Kubarych[5], N. Kuramoto[6], S. Mizushima[6], L. Zhang[6], K. Fujita[6], S. Davidson[7], R.G. Green[8], J.O. Liard[8], N.F. Murnaghan[8], C.A. Sanchez[8], B.M. Wood[8], H. Bettin[9], M. Borys[9], M. Mecke[9], A. Nicolaus[9], A. Peter[9], M. Müller[9], F. Scholz[9], A. Schofeld[9]

[1] BIPM
[2] DFM, Denmark
[3] KRISS, Rep. of Korea
[4] NIM, China
[5] NIST, USA
[6] NMIJ, Japan
[7] NPL, UK
[8] NRC, Canada
[9] PTB, Germany

Bureau International des Poids et Mesures

Table of Contents

Abstract

This report describes the first CCM key comparison of realizations of the kilogram definition based on the fixed numerical value of the Planck constant, which came into force on 20 May 2019. The objectives were to determine the level of agreement between realizations of the kilogram using Kibble and joule balances and the X-ray crystal density (XRCD) method and to provide input for the calculation of the first "consensus value" of the kilogram. The consensus value will serve as the basis for an internationally coordinated dissemination of the kilogram which will continue until sufficient agreement between realization experiments has been achieved.

The comparison was organized by the BIPM and had seven participants. The BIPM, KRISS, NIST and NRC operated Kibble balances, the NIM used a joule balance and the NMIJ and the PTB participated using ^{28}Si spheres, the masses of which were determined with the XRCD method. These realization methods were used to calibrate 1 kg mass standards under vacuum. The standards were sent (in air) to the BIPM where they were compared under vacuum with each other and with BIPM Pt-Ir working standards. The latter were calibrated (in air) traceable to the International Prototype of the Kilogram (IPK), the mass of which served as the definition of the kilogram until 20 May 2019.

The results of the weighings at the BIPM together with the measurement results communicated by the participants allowed comparison of the values attributed to mass standards of 1 kg using the participating realization experiments. The level of agreement between mass determinations with the realization experiments and the BIPM as-maintained mass unit, traceable to the Planck constant through the mass of the International Prototype of the Kilogram can also be deduced.

1 Introduction

On 20 May 2019, a revision of the International System of Units, the SI, came into force [1]. It brought new definitions for four of the seven SI base units: the kilogram, the ampere, the kelvin and the mole. Since then, the kilogram has been defined based on a fixed numerical value of the Planck constant [2]. This leads to the fundamentally new situation in mass metrology whereby the mass unit can, in principle, be realized individually by any National Metrology Institute (NMI) which is capable of developing an experiment allowing the realization of the new definition. At present, the *mise en pratique* of the kilogram [3] recognizes two independent primary methods that are capable of realizing the kilogram with relative uncertainties of a few parts in 10^8, corresponding to a few tens of micrograms. The first relies on an electromechanical experiment comparing electrical and mechanical power using electrical quantum standards, which is known as a Kibble balance [4]. Experiments of this type were developed independently by a number of NMIs. A special realization of this principle has been described as a joule balance [5]. The second method determines the mass of a nearly perfect sphere made of ^{28}Si by determining the number of atoms in the sphere, where the mass of the atom is well known in terms of the Planck constant [6]. This technique is known as the X-ray crystal density (XRCD) method and was developed by an international collaboration, the International Avogadro Coordination.

In 2017, the numerical values of the defining constants for the four new SI definitions were determined by a least squares adjustment of all available data by the CODATA Task Group on Fundamental Constants [7]. The set of eight data for the Planck constant was not statistically consistent. In particular, the results of the NIST Kibble balance and the International Avogadro Coordination both of which were published in 2017 and which both had small uncertainties, differed by four times their combined standard uncertainty. The Consultative Committee for Mass and Related Quantities (CCM) discussed this situation at its meeting in 2017. It decided that until the dispersion between values became compatible with the individual realization uncertainties, NMIs should base their dissemination on an agreed "consensus value". The details of this international coordination of the kilogram dissemination are described in the "CCM detailed note on the dissemination process after the redefinition of the kilogram" [8]. The consensus value will be calculated for the first time following the completion of the present key comparison. It will be calculated as the arithmetic mean of three sets of data:

- data directly traceable to the IPK (used for the last time in 2014 and maintained by the BIPM working standards);
- data from the CCM Pilot Study of realization experiments, carried out in 2016 [9], (as maintained on the BIPM working standards) and corrected for the shift of 17 parts in 10^9 in h introduced by the CODATA 2017 adjustment, with respect to the CODATA 2014 value, which was used as the reference in the Pilot Study;
- the key comparison reference value (KCRV) of the present CCM key comparison.

These three data sets from 2014, 2016 and 2020 can be tied together through the controlled stability of the BIPM working standards.

It is planned to repeat the key comparison every two years. After each comparison, the consensus value will be calculated as the arithmetic mean of the three most recent data sets, to reduce temporal changes in the consensus value.

This comparison has the following two objectives: (1) to study the present level of agreement between realization experiments and (2) to provide input for the calculation of the first consensus value.

The consensus value will be determined by the CCM *Task Group on the Phases for the Dissemination of the kilogram following redefinition* (TGPfD-kg). The result will be published in a CCM document.

2 Organization of the key comparison

The comparison followed a similar scheme as the CCM Pilot Study [9]. It was organized in the form of a star-comparison, in which each participating institute was requested to send one or two of its own 1 kg mass standards to the BIPM, which acted as the pilot laboratory. One of the standards should be a Pt-Ir prototype, the other could be chosen by the participant: Pt-Ir, stainless steel or a Si-sphere. The standards should be well characterized with regard to their mass stability, in particular under repeated air-to-vacuum transfers.

The participants determined the mass of their travelling standards under vacuum, using their realization experiment, before sending them to the BIPM. The mass values attributed to the standards were calculated using the numerical value of the Planck constant that is now fixed in the SI,

$$h = 6.626\,070\,15 \times 10^{-34}\ \mathrm{J\ s}.$$

At the BIPM, the travelling standards were all compared under vacuum, over a period of three months, to BIPM working standards which served as reference mass standards. The comparison of the mass standards at the BIPM together with the mass values attributed by the participants allowed a determination of the differences between the participants' realizations of the kilogram. The BIPM reference mass standards are traceable to the International Prototype of the Kilogram (IPK) but also maintain the reference value of the CCM Pilot Study in 2016. This allows comparison of the KCRV of the present comparison with the reference value of the Pilot Study and the IPK and the calculation of the consensus value.

The participants were asked to verify the mass stability of their travelling standards over the period of the comparison by comparing them before and after the measurements at the BIPM with another, stable mass standard, either in air or in vacuum. These measurements should be made with the lowest possible uncertainty, in particular by avoiding large buoyancy corrections.

3 Participants and travelling standards

All NMIs working on realization experiments had been invited to participate in this comparison, under the following conditions:

- standard uncertainty of the mass of the 1 kg travelling standards under vacuum based on the realization experiment < 200 µg, that is 200 parts in 10^9;
- a peer reviewed publication of the realization experiment, including a detailed uncertainty budget (either for a determination of the Planck constant before the redefinition, or for the calibration of a mass standard after the redefinition) and some evidence of the reproducibility of the results over time;
- availability to perform the required measurements within the schedule of the comparison.

Seven institutes participated (table 1): The BIPM, KRISS (Rep. of Korea), NIST (USA) and NRC (Canada) used Kibble balances, the NIM (China) used a joule balance and the NMIJ (Japan) and the PTB (Germany) used the XRCD method as the basis for their mass determinations.

Table 1: Comparison coordinator, support group members and participants of CCM.M-K8.2019.

Comparison coordinator		
Institute	**Contact person**	
BIPM	Michael Stock	
Support group		
Institute	**Contact person**	
DFM	Lars Nielsen	
NPL	Stuart Davidson	
Participants		
Institute	**Contact person(s)**	**Realization method**
BIPM	Hao Fang	Kibble balance
KRISS	Kwang-Cheol Lee	Kibble balance
NIM	Zhengkun Li Jian Wang	Joule balance
NIST	Patrick Abbott	NIST-4 Kibble balance
NMIJ	Naoki Kuramoto	XRCD method
NRC	Richard Green	Kibble balance
PTB	Horst Bettin Michael Borys	XRCD method

Table 2 lists the travelling standards sent by the participants. The BIPM used only one mass standard because no transportation was involved which could compromise the mass stability. Since mass standard 691 is not within the tolerance for a prototype, it is normally used with a small Pt-wire of about 70 mg and is then identified as 691s.

Table 2: Travelling standards and their properties

Institute	Identification of standard	Manufacturer	Type	Estimated air-vacuum surface sorption / µg	Magnetic susceptibility	Magnetic polarization / µT
BIPM	691s	BIPM	Pt-Ir standard with small Pt wire	4	24×10^{-5}	<0.02
KRISS	111	BIPM	Pt-Ir prototype		24×10^{-5}	< 0.02
	SN17	Mettler Toledo	stainless steel cylinder	6.3	<0.002	<0.2
NIM	110	BIPM	Pt-Ir prototype		24×10^{-5}	0
	6601	Changzhou Accurate Weight Co., China	stainless steel cylindrical	5.5	5.16×10^{-4}	0.01
NIST	85	BIPM	Pt-Ir prototype	7.2	24×10^{-5}	<0.02
	104	BIPM	Pt-Ir prototype	3.5	24×10^{-5}	<0.02
NMIJ	94	BIPM	Pt-Ir prototype	5.7(3.3)	3×10^{-4}	< 0.02
	E59	Stanton Instruments	Pt-Ir standard	11.6(3.3)	2×10^{-4}	< 0.02
NRC	106	BIPM	Pt-Ir prototype	3.4	24×10^{-5}	<0.02
	F18	BIPM	Stack of 8 Pt-Ir disks	12.2	24×10^{-5}	<0.02
PTB	109	BIPM	Pt-Ir prototype	2	< 0.001	< 0.1
	Si14-02	PTB	Si sphere	20	-2.6×10^{-7}	0

4 Measurements made by the participants

The following are short summaries of the measurement reports provided by the participants. The reports, including the detailed uncertainty budgets, are available as an annex to this report.

4.1 BIPM

The mass of standard 691s under vacuum (10^{-2} Pa) was determined directly with the Kibble balance during the period 22 August to 10 September 2019.

The BIPM Kibble balance is operated in the one-mode two-phase scheme, in which the same current is present during both measurement phases across one of the windings of a bifilar coil. A 3-axis heterodyne interferometer measures the vertical coil velocity. The laser source is a commercial single frequency 532 nm Nd:YAG-laser stabilized on a hyperfine transition of iodine. The voltages are measured using a system based on a NIST-fabricated PJVS producing up to ±2 V with a resolution of 77 µV. The gravitational acceleration is deduced from measurements made with FG-5 gravimeters during the 2009 International Comparison of Absolute Gravimeters and with the METAS FG5-X in September 2019. The value is corrected for the spatial gradient and the self-attraction of the experiment. The tidal variations are calculated using Tsoft software based on the experimentally determined local parameters in the laboratory. The 100 Ω resistor was calibrated with reference to the quantum Hall resistance standard.

The uncertainty of the mass of standard 691s is 0.049 mg. An article describing the details of the Kibble balance and the uncertainty budget has been published in Metrologia [11].

4.2 KRISS

The masses of two travelling standards were measured using the KRISS Kibble balance at a vacuum pressure between 3×10^{-3} Pa and 4×10^{-3} Pa. The Pt-Ir mass was measured between 20 December 2019 and 5 January 2020. The stainless steel mass was measured twice, from 2 December to 15 December 2019 and from 6 January to 18 January 2020.

A 100 Ω resistor was used to obtain the coil current. The resistor was calibrated before and after the kilogram realization experiment using the KRISS quantum Hall resistance standard.

The voltage was measured against the Josephson voltage standard. The Josephson standard uses a 2 V NIST chip. The voltage difference against the Josephson reference is within a few µV in weighing mode and a few mV in moving mode.

The gravity value was obtained using the result of the gravity survey carried out in 2014. An absolute and a relative gravimeter were used to obtain the gravity at the position of the test mass. The gravity perturbation by the attractive force between the apparatus and the test mass was calculated by finite element analysis. The time varying gravity (tidal signal) was calculated using the commercial software QuickTide Pro.

The coil velocity was measured using a homodyne interferometer. An iodine stabilized helium-neon laser was used for the interferometer.

A 10 MHz signal from the KRISS time and frequency laboratory was used as the time reference for voltage and velocity measurements.

The uncertainty of the mass of the travelling standards is 0.120 mg. An article describing the experiment and the uncertainty budget has been accepted for publication in Metrologia [12].

4.3 NIM

The masses of the travelling standards under vacuum (6×10^{-4} Pa) were determined directly with the joule balance during October and November 2019. A PJVS system is used in the joule balance system for the electrical measurement. An SR102 type standard resistor was used for the current measurement and this resistor has been calibrated by the QHR standard system periodically. The length measurement is traceable to the length primary standard of NIM. The absolute gravity is measured by the Micro-g LaCoste FG5X gravimeter (FG5x-249 of NIM).

The uncertainty of the mass of the travelling standards is 0.052 mg for the Pt-Ir prototype and 0.065 mg for the stainless steel standard. An article describing the details of the joule balance and the uncertainty budget has been accepted for publication in Metrologia [13].

4.4 NIST

NIST measured the two Pt-Ir travelling standards directly on the fourth generation Kibble balance, NIST-4. Each mass was brought under vacuum to NIST-4 (4.5 to 6.1×10^{-4} Pa). The mass was determined on NIST-4, and then transferred under vacuum to a vacuum balance to transfer the mass value to secondary masses to check the mass stability. Two series of realization were done for each mass during September and October 2019. Between the two realizations, NIST-4 was brought to air to check and adjust multiple alignments. Following the measurements under vacuum, both travelling standards were brought to air and measured in a mass comparator against another Pt-Ir standard as part of the check of mass stability.

The uncertainty of the mass values of the travelling standards is 0.027 mg. The latest publication from 2017 includes a detailed uncertainty budget for a combined standard uncertainty of 13.5 parts in 10^9 for the determination of the Planck constant [14].

4.5 NMIJ

The NMIJ used the ^{28}Si sphere AVO28-S5c which was manufactured by the International Avogadro Coordination (IAC). The following is a very succinct summary of a very comprehensive measurement report. The values for the lattice constant, the relative atomic mass of silicon and the influence of point defects, which are not expected to change over time, were taken from the previous work of the IAC [15, 18]. For the new realization of the kilogram only the core volume of the sphere and the mass of the surface layers were re-determined from September 2019 to January 2020. The core volume was determined by an optical interferometer, corrected for the phase shift introduced by the surface layers. The mass of the surface layers was determined from studies with XPS and spectroscopic ellipsometry.

The masses of the travelling standards in vacuum (1.1×10^{-3} Pa) were compared with that of the sphere using a mass comparator in January 2020. The uncertainty of the mass values is 0.0213 mg. This uncertainty is supported by a publication from 2017 including an uncertainty budget for a combined uncertainty of 2.4×10^{-8} for the determination of the Avogadro constant [16].

4.6 NRC

In the first step, a 500 g mass artefact AuCuB was measured in an M_one vacuum mass comparator against a set of 5 other mass standards of various sizes and materials. The artefacts in this set have been found to be stable in vacuum over the time scale of the Kibble balance experiment. AuCuB was then removed from vacuum via the loadlock into air and transferred to the Kibble balance. Calibration of AuCuB was performed under vacuum in the Kibble balance over several days. After which it was removed and transferred through air back to the vacuum balance via the load lock and compared with masses used in the first step. The difference in comparison values observed allowed transfer mass stability to be evaluated. AuCuB was then stacked on top of one of the 5 masses in the comparison set, a second 500 g mass, AuCuA. Both of these 500 g masses together then served to determine the mass of the two travelling standards measured under vacuum (4.7×10^{-3} Pa) in the M_One mass comparator. All measurements were carried out in October 2019. The uncertainty of the mass values is about 0.012 mg. A series of publications has been published on the NRC Kibble balance, the most recent dating from 2017 [17]. It includes an uncertainty budget for a combined uncertainty of 10.3 parts in 10^9 for the determination of the Planck constant. The traveling standards were hand-carried to and from the BIPM housed in custom designed shipping vessels that were sealed to the environment during transport.

4.7 PTB

The PTB used two ^{28}Si spheres: sphere AVO28-S8c which was manufactured by the International Avogadro Coordination (IAC) and the more recently made sphere Si28kg01a, which has a higher enrichment of ^{28}Si and was manufactured by PTB. Only the volume of the spheres and the surface layers were measured anew. In contrast to the previous work [18], the surface layers were measured only by means of XRF/XPS-methods. The point defect corrections are different for both spheres and their uncertainties are only partly correlated. The correlation coefficient between the masses of both spheres is estimated as 0.53, mainly caused by the measurements of the mass of the surface layer, the lattice parameter and the volume measurement.

The travelling standards, a Pt-Ir prototype and a Si-sphere, were compared during October to November 2019 with both ^{28}Si spheres in a CCL 1007 mass comparator under vacuum (4.4×10^{-3} to 4.8×10^{-4} Pa). The PTB provided the masses of the travelling standards with respect to both ^{28}Si spheres, but did not combine the results. Instead, a correlation matrix for the four mass determinations was given (see Section 6). The uncertainty of the mass values of the travelling standards with respect to AVO28-S8c are 0.0155 mg and 0.0156 mg, respectively, and with respect to Si28kg01a 0.0136 mg and 0.0137 mg, respectively. These uncertainties are supported by the most recent publications [18] and [19].

5 Measurements at the BIPM

The objective of the weighings at the BIPM was to determine the mass differences between the participants' travelling standards and to compare them with BIPM working standards. The masses of the latter are traceable to the mass of the IPK from measurements made in 2014 [10] and maintain the reference value of the CCM Pilot Study of 2016 [9].

All measurements were made with the M_one mass comparator, equipped with a six-place mass exchanger. The influence of the mass handler position is determined once a year and has always been found to be within the weighing uncertainty. The sensitivity of the mass comparator is determined at the beginning and at the end of each weighing set using a sensitivity weight of 95 mg, which was recalibrated in September 2019. Since the masses of all travelling standards are within a range of 10 mg, the uncertainty contribution of the sensitivity is negligible. The residual pressure during the vacuum weighings was between 1×10^{-4} Pa and 6×10^{-4} Pa.

The mass comparator stayed under vacuum during the whole duration of the comparison. The cylindrical 1 kg Pt-Ir standard A0 of the BIPM stayed inside the comparator during the whole duration and served as the reference with which all travelling standards were compared. Before and after all the vacuum weighings of the comparison, A0 was compared in air against a set of BIPM Pt-Ir working standards. The BIPM stack A18, made of eight Pt-Ir disks, separated from one another by three small bent Pt-wires, also stayed under vacuum during the whole comparison. Together with A0 it forms a pair of sorption standards, having very similar mass and volume, but very different surface areas. From the measured mass differences between A18 and A0 in air and in vacuum and the calculated surface areas, the mass of the adsorbed (water) layers in air can be calculated. This allows the calculation of the mass of A0 under vacuum, based on its calibration in air. Since it has been observed at the BIPM in the past that the initial transfer from air to vacuum does not always give a reliable estimate of the sorption effect, the transfer at the end of the comparison from vacuum to air was used to calculate the surface sorption value. The mass of the adsorbed water layer on A0 in air was determined as 0.004 mg, corresponding to 55 ng/cm^2. This is consistent with previous determinations.

The mass of A0 was determined in air directly after the period of vacuum weighings with reference to two Pt-Ir working standards for regular use, Nos. 63 and 77. Both of these were then compared with the working standards for limited use, Nos. 9, 31 and 650. These working standards are used only once per year. During March 2019 they were re-calibrated against the working standards for exceptional use, Nos. 25, 73 and 91. These standards are used only every five years and this was the first use after their comparison with the IPK in 2014. The standards for exceptional use were cleaned and washed before use to bring them back to their mass in 2014. This measurement resulted in a correction of +0.008 mg for the working standards, with respect to the predicted mass values. This is related to the contamination of the working standards since 2014 (1.6 µg/year), for which no correction had been applied during this five-year period[1].

Six weighing sets were carried out under vacuum. Each of the sets contained the BIPM standards A0 and A18 and four of the travelling standards. The sets were formed as follows:

Set 1 (20-21/11/2019):	A0	A18	106 (NRC)	F18 (NRC)	85 (NIST)	691s (BIPM)
Set 2 (24-26/11/2019):	A0	A18	106 (NRC)	F18 (NRC)	85 (NIST)	104 (NIST)
Set 3 (29/11-2/12/2019):	A0	A18	109 (PTB)	Si14-02 (PTB)	85 (NIST)	104 (NIST)
Set 4 (10-12/12/2019):	A0	A18	109 (PTB)	Si14-02 (PTB)	6601 (NIM)	110 (NIM)
Set 5 (14-16/2/2020):	A0	A18	SN17 (KRISS)	111 (KRISS)	6601 (NIM)	110 (NIM)
Set 6 (21-23/2/2020):	A0	A18	SN17 (KRISS)	111 (KRISS)	E59 (NMIJ)	94 (NMIJ)

The sphere Si14-02 was cleaned as requested by the PTB, by applying the recommended procedure, before the weighings. Due to the late arrival of the standards from KRISS and NMIJ, there was a delay of two months between the sets 4 and 5. In addition to A0 and A18, the two standards from NIM stayed in the

[1] This first five-year period following the use of the IPK in 2014 was a trial period to determine the contamination rate of the working standards used in the new hierarchical structure introduced in 2014. From now on the contamination rate will be used in the prediction of the mass values.

mass comparator under vacuum during this period to provide a solid link between the first group (sets 1 to 4) and the second group (sets 5 and 6) of weighings.

For each set four full weighing schemes were carried out. In each weighing scheme all 15 pairwise mass-differences between the six standards were determined. This led to 60 mass differences for each set and to 360 mass differences for the whole comparison. Each individual mass difference was obtained from an A-B-A-B-A-B-A scheme. The uncertainty budget for the weighings is shown in table 3 below. The table shows the combined standard uncertainties of the masses of the travelling standards with respect to two different references:

- with respect to the mass of A0 under vacuum
- with respect to the Planck constant (via the IPK)

The mass stability of A0 during the vacuum weighings has been estimated from the observation that during the vacuum weighings the mass of the stack A18 has increased by 0.006 mg with respect to A0. If we assume that the origin is a contamination of the surface from the residual gas (about 0.1 mPa), the ratio of the surface areas leads to a mass increase of 0.002 mg for A0. We did not interpolate the mass of A0 over time but only treat this as an uncertainty. As explained above, the mass value of A0 was obtained from the vacuum-to-air transfer at the end of the vacuum weighings and the calibration in air following it.

All 360 mass differences were used to carry out a generalized least-squares adjustment to obtain the masses of the travelling standards. The vacuum mass of A0 (which is traceable to the IPK) served as the constraint in the adjustment. The mass drift of A18 was included as an adjusted parameter. The statistical uncertainty of the adjusted masses was 0.0005 mg. The results are presented in the following section.

Table 3: Uncertainty budget for the mass of the travelling standards with respect to the vacuum mass of the reference A0 and with respect to the Planck constant (via the IPK). The uncertainties of the second part of the table are totally correlated for all travelling standards.

Source of uncertainty	**Uncertainty**	**Sensitivity coefficient**	**Unc. contribution / mg**
position error of M_one	0.0005 mg	1	0.0005
centre of gravity of standards	0.5 mm	0.0003 mg/mm	0.0002
statistical uncertainty	0.0005 mg	1	0.0005
mass stability of A0 during vacuum weighings	0.002 mg	1	0.002
Standard uncertainty of mass of travelling standards with respect to mass of A0 in vacuum			**0.0021**
air to vacuum transfer of A0	0.002 mg	1	0.002
mass of A0 in air with respect to IPK	0.005 mg	1	0.005
mass of IPK with respect to Planck constant	0.010 mg	1	0.010
Standard uncertainty of mass of travelling standards with respect to Planck constant			**0.0116**

6 Results

The results of the mass determinations by the participants, i = 1,...,7, of the travelling standards, j = 1, 2, $m_{i,j}^{\mathrm{NMI}}$, using their realization experiments are shown on the left side of table 4, as deviations from 1 kg, as derived from the Planck constant. The correlation coefficients r_i between the uncertainties of the two standards of each institute were calculated from the detailed uncertainty budgets, which show the correlations between standards for each uncertainty component. The PTB used two ^{28}Si spheres, AVO28-S8c and Si28kg01a, for the determination of the masses of the travelling standards and reported the results separately. The correlation matrix for the four individual results of the PTB was provided in the measurement report and is reproduced in table 5. The correlations are needed to calculate for each NMI the mean value of the results obtained with its two travelling standards and its uncertainty.

The mass stability of the travelling standards during the comparison was determined by the participants by comparing them in air against a stable reference mass before and after the comparison at the BIPM (table 6). The largest observed mass change was -0.0085 mg with an uncertainty of 0.0027 mg, for n° 85 from NIST. This mass standard shows many scratches on its bottom surface and several indentations on its circumference. The mass standards of PTB and NRC have travelled very well, with mass changes within 0.002 mg. The mass increase of 0.002 mg was observed on a stack with a large surface area. The mass standard of the BIPM, which did not travel, was stable within the uncertainty of the weighings. Due to the travel restrictions during the covid-19 crisis, the travelling standards from KRISS, NIM and NMIJ could not be returned within the duration of the comparison. For the standards of BIPM, NIST, NRC and PTB, the uncertainty related to the mass (in)stability is taken either as the absolute value of the observed mass change or its uncertainty, whatever is larger. For the standards for which the stability could not be verified, we have assumed the average mass stability observed for the other standards (excluding 691s of BIPM, which did not travel), 0.004 mg.

The masses attributed by the participants are compared with those obtained by the BIPM, $m_{i,j}^{\mathrm{BIPM}}$, using its working standards. The latter fulfil a double role: they serve as a common reference to compare the realization experiments and they maintain the BIPM mass unit, which is traceable to the IPK, and also to the Planck constant through its known relationship with the IPK. The latter is important because it allows to link the comparison reference value of the present comparison with that of the Pilot Study of 2016 and with the IPK, used in 2014, in the calculation of the CCM consensus value.

Table 4: Masses (deviations from 1 kg) and standard uncertainties of the travelling standards determined by the NMIs using their realization experiments and by the BIPM using its working standards. On the right, the mass differences $\Delta m_{i,j} = m_{i,j}^{\mathrm{NMI}} - m_{i,j}^{\mathrm{BIPM}}$ are shown.

		NMI realization results				BIPM results		Mass difference NMI-BIPM	
Institute	**Identification of standard**	$m_{i,j}^{\mathrm{NMI}}$ / mg	**Std. unc. / mg**	**Corr. coeff.** r_i	**unc. (mass stability) / mg**	$m_{i,j}^{\mathrm{BIPM}}$ / mg	**Std. unc.[2] / mg**	$\Delta m_{i,j}$ / mg	**Std. unc. / mg**
BIPM[3]	691s	-0.022	0.049	-	0.0015	-0.0284	0.0021	0.0064	0.049
KRISS	111	0.27	0.120	0.600	0.004[4]	0.2737	0.0021	-0.0037	0.120
	SN17	0.20	0.120		0.004[4]	0.0892	0.0021	0.1108	0.120
NIM	110	0.021	0.052	0.264	0.004[4]	0.0407	0.0021	-0.0197	0.052
	6601	0.710	0.065		0.004[4]	0.7603	0.0021	-0.0503	0.065
NIST	85	-0.7679	0.027	0.903	0.0085	-0.7498	0.0021	-0.0181	0.028
	104	0.398	0.0273		0.0056	0.4169	0.0021	-0.0189	0.0279
NMIJ	94	0.3228	0.0213	0.975	0.004[4]	0.3395	0.0021	-0.0167	0.0218
	E59	4.9097	0.0213		0.004[4]	4.9263	0.0021	-0.0166	0.0218
NRC	106	0.4401	0.0116	0.975	0.0012	0.4429	0.0021	-0.0028	0.0118
	F18	0.4688	0.0118		0.0023	0.4738	0.0021	-0.0050	0.0122
PTB (reference AVO28-S8c)	109	0.135	0.0155	See table 5	0.0019	0.1773	0.0021	-0.0423	0.0158
	Si14-02	-4.257	0.0156		0.0048	-4.1912	0.0021	-0.0658	0.0169
PTB (reference Si28kg01a)	109	0.144	0.0136		0.0019	0.1773	0.0021	-0.0333	0.0139
	Si14-02	-4.249	0.0137		0.0048	-4.1912	0.0021	-0.0578	0.0152

[2] Standard uncertainty with respect to mass of A0 under vacuum (see page 14).

[3] For the BIPM, the “NMI result” is the mass of standard 691s obtained with the BIPM Kibble balance, and the “BIPM result” is its mass based on the BIPM working standards.

[4] The mass stability of the travelling standards of KRISS, NIM and NMIJ could not be determined because due to the travel restrictions during the covid-19 crisis, the standards could not be returned to the institutes within the time frame of this comparison. A similar stability as for the other travelling standards is assumed (average mass stability observed).

Table 5: Correlation matrix for the uncertainties of the masses of the PTB travelling standards 109 and Si14-02 using two different reference spheres: AVO28-S8c and Si28kg01a.

		(AVO28-S8c)		(Si28kg01a)	
		109	**Si14-02**	**109**	**Si14-02**
(AVO28-S8c)	**109**	1.00	0.98	0.54	0.53
	Si14-02	0.98	1.00	0.54	0. 55
(Si28kg01a)	**109**	0.54	0.54	1.00	0.98
	Si14-02	0.53	0.55	0.98	1.00

At the present stage of the data analysis, the BIPM working standards serve only as a common reference for the comparison. Therefore the BIPM comparison uncertainty of 0.0021 mg shown in table 4 includes only the uncertainty components of the first part of the uncertainty budget in table 3, which are uncorrelated between the travelling standards. It does not include the last three components of table 3 which are related to the uncertainty of the vacuum mass of A0 with respect to the Planck constant. These components are totally correlated for the travelling standards and can be omitted here, because for the purpose of the comparison of the realization experiments with each other, the BIPM mass unit has to be stable, but not accurate.

The right side of Table 4 shows the differences $\Delta m_{i,j} = m_{i,j}^{\mathrm{NMI}} - m_{i,j}^{\mathrm{BIPM}}$ between the mass values attributed by the participants and the BIPM. The uncertainty includes the NMI realization uncertainty, the mass stability and the (uncorrelated) BIPM weighing uncertainty. For the sphere Si14-02 sent by the PTB an additional uncertainty component of 0.004 mg was included, for the reproducibility of the cleaning of the sphere at the BIPM, as compared to a cleaning at the PTB. This is twice the uncertainty for the repeatability of the cleaning at the PTB. Figure 1 shows these differences and their uncertainties.

Table 6: Mass changes of the travelling standards during the comparison. The standards of the KRISS, NIM and NMIJ could not be returned as scheduled due to the travel restrictions, and their mass stability not be determined experimentally.

Institute	Identification of standard	Mass change / µg	Uncertainty of mass change / µg
BIPM	691s	+0.2	1.5
NIST	85	-8.5	2.7
	104	-5.6	2.7
NRC	106	-1.1	1.2
	F18	+2.1	2.3
PTB	109	-1.0	1.9
	Si14-02	+0.4	4.8

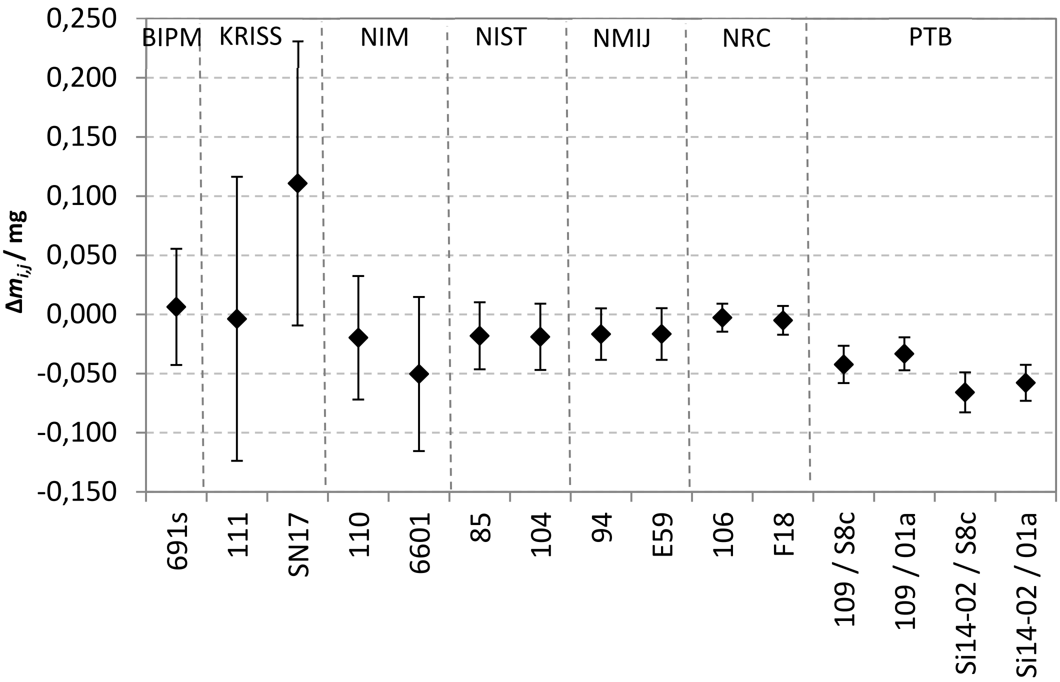

Fig. 1: Differences $\Delta m_{i,j} = m_{i,j}^{\mathrm{NMI}} - m_{i,j}^{\mathrm{BIPM}}$ between the mass determinations of each travelling standard with the NMI's realization experiment and the BIPM working standards and associated combined standard uncertainties

To obtain one single result, representative for each NMI, the results for the two travelling standards were averaged. The correlation coefficients r_i between the uncertainties of the two standards, shown in table 4, were taken into account (the treatment of the four PTB results is described below). The weighted mean Δm_i of the two results $\Delta m_{i,1}$ and $\Delta m_{i,2}$ with uncertainties $u_{i,1}$ and $u_{i,2}$ and covariance $u_{i,12} = r_i\, u_{i,1}\, u_{i,2}$ and its variance were calculated as [20]

$$\Delta m_i = \frac{(u_{i,1}^2 - u_{i,12})(u_{i,2}^2 - u_{i,12})}{u_{i,1}^2 + u_{i,2}^2 - 2u_{i,12}} \left(\frac{\Delta m_{i,1}}{u_{i,1}^2 - u_{i,12}} + \frac{\Delta m_{i,2}}{u_{i,2}^2 - u_{i,12}} \right) \quad \text{(eq. 1)}$$

$$var(\Delta m_i) = \frac{u_{i,1}^2 u_{i,2}^2 - u_{i,12}^2}{u_{i,1}^2 + u_{i,2}^2 - 2u_{i,12}} \quad \text{(eq. 2)}$$

In the case of the PTB results, the correlation matrix, table 5, needs to be taken into account. For each of the two reference spheres, the uncertainties for both travelling standards are very similar and they are nearly completely correlated. Taking into account the additional uncertainties related to the comparison, the correlation coefficient between both travelling standards is reduced from 0.98 to 0.90 for the reference standard AVO28-S8c, and from 0.98 to 0.88 for the reference standard Si28kg01a. In a first step, we calculate for each of the two reference spheres the weighted mean of the $\Delta m_{i,j}$ for both travelling standards from table 4. Considering the correlations (eqs. 1 and 2) the mean values are:

Reference AVO28-S8c	$\langle \Delta m_{\mathrm{PTB,109}}, \Delta m_{\mathrm{PTB,Si14-02}} \rangle$ = -0.0457 mg	u = 0.0157 mg
Reference Si28kg01a	$\langle \Delta m_{\mathrm{PTB,109}}, \Delta m_{\mathrm{PTB,Si14-02}} \rangle$ = -0.0365 mg	u = 0.0139 mg

The diagonal blocks of the correlation matrix show that the results for both reference spheres have a correlation coefficient of about 0.54. The results based on the use of the two references are then combined in the same way. The result is Δm_{PTB} = -0.0399 mg with an uncertainty of 0.0128 mg.

Considering the data reported by the BIPM and PTB (table 4) and the correlation matrix (table 5), there are two notable findings. The difference between mass values assigned to one travelling standard (Pt109 or Si14-02) by the two realizations (with the AVO28-S8c and the Si28kg01a sphere) amounts to 0.009 mg for both travelling standards, which is well within the uncertainty of that difference. But independent of the realisation, the mass difference between the travelling standards measured at the BIPM differs from the mass difference measured at PTB by 0.024 mg, more than three times the uncertainty of this difference.

Following discussion with the colleagues from PTB, it appears that the problem is related to the vacuum weighing of the sphere Si14-02, however the origin of the problem remains unknown. The PTB confirmed their reported results after check measurements with both travelling standards and the AVO28-S8c sphere in vacuum with the CCL1007 and the M_one mass comparator. It should be noted that withdrawal of the sphere result would change the average PTB result by only +0.003 mg as a consequence of the averaging process of the PTB results described above, and the higher uncertainty for the result obtained with the sphere.

Table 7 shows the averaged results Δm_i for all participants and figure 2 shows the results in graphical form. The figure also shows the key comparison reference value (KCRV), calculated as the inverse-variance-weighted mean of the Δm_i $(i = 1, \dots, 7)$ of the realization experiments:

$$\overline{\Delta m} = \frac{\sum_i w_i \Delta m_i}{\sum_i w_i} \quad \text{with} \quad w_i = \frac{1}{u(\Delta m_i)^2} \quad \text{and} \qquad \text{(eq. 3)}$$

$$u(\overline{\Delta m}) = \sqrt{\left(\sum w_i\right)^{-1}} \qquad \text{(eq. 4)}$$

The result is $\overline{\Delta m} =$ -0.0188 mg with a standard uncertainty of 0.0075 mg. The largest statistical weight of 41 % is attributed to the NRC, followed by the PTB with 34 %. This value of the KCRV (with respect to the mass unit maintained on the BIPM working standards) will be the input of this comparison to the calculation of the consensus value, as explained in the introduction.

Table 7: Differences Δm_i between mass values attributed to 1 kg mass standards using the realization experiment of the participants and by the working standards of the BIPM, and associated standard uncertainty.

Institute	Δm_i / mg	$u(\Delta m_i)$ / mg
BIPM	0.0064	0.0491
KRISS	0.0536	0.1072
NIM	-0.0305	0.0456
NIST	-0.0185	0.0270
NMIJ	-0.0166	0.0214
NRC	-0.0034	0.0118
PTB	-0.0399	0.0128

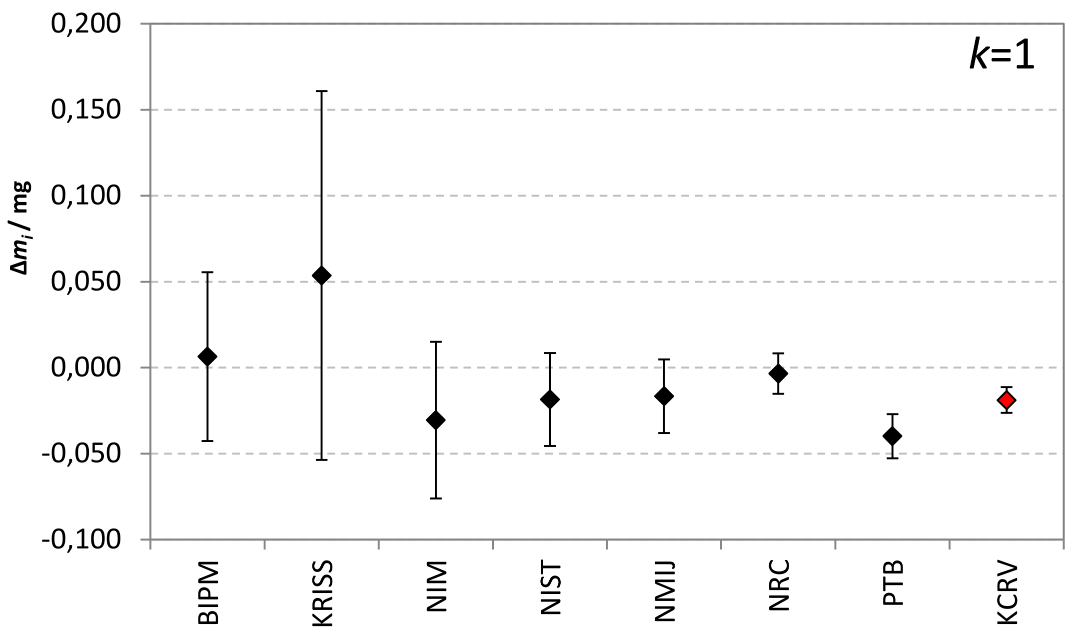

Fig. 2: Differences Δm_i between mass values attributed to 1 kg mass standards using the realization experiment of the participants and by the working standards of the BIPM, and associated standard uncertainty.

In table 7 and figure 2 the BIPM as-maintained mass unit was used as the reference. In the last step of the analysis we calculate the deviation of the participants' results from the key comparison reference value $\overline{\Delta m}$:

$$\Delta m_i' = \Delta m_i - \overline{\Delta m} \qquad \text{(eq.5)}$$

For the participants which contributed to the calculation of the KCRV, the uncertainty is calculated as

$$u(\Delta m_i') = \sqrt{u(\Delta m_i)^2 - u(\overline{\Delta m})^2} \qquad \text{(eq.6)}$$

These deviations and their uncertainties are shown in table 8 and figures 3 and 4. Since the BIPM working standards are traceable to the Planck constant through its known relationship with the International Prototype of the Kilogram (IPK), the difference between mass values based on the BIPM as-maintained mass unit and the KCRV can also be determined:

$$\Delta m_{h(IPK)}' = \Delta m_{h(IPK)} - \overline{\Delta m} = -\overline{\Delta m} \qquad \text{(eq.7)}$$

$\Delta m_{h(IPK)} = 0$ because the BIPM working standards (traceable to the IPK) served as the reference for the comparison. Since the BIPM as-maintained mass unit was not used in the calculation of the KCRV, the uncertainty is calculated as

$$u\left(\Delta m_{h(IPK)}'\right) = \sqrt{u\left(\Delta m_{h(IPK)}\right)^2 + u(\overline{\Delta m})^2} \qquad \text{(eq.8)}$$

The uncertainty $u\left(\Delta m_{h(IPK)}\right)$ includes all uncertainty components of table 3 and is dominated by the uncertainty of the mass of the IPK in terms of the Planck constant, 0.010 mg.

Table 8: Deviations $\Delta m_i'$ of the NMIs' results from the KCRV, related standard uncertainties $u(\Delta m_i')$ and expanded uncertainties for *k* = 2, $U(\Delta m_i')$. The difference between mass values based on the BIPM working standards, traceable to the Planck constant through the IPK, and the reference value is also shown.

Institute	**Deviation from KCRV $\Delta m_i'$ / mg**	$u(\Delta m_i')$ **/ mg**	$U(\Delta m_i')$ **/ mg**
BIPM	0.0252	0.0485	0.0970
KRISS	0.0724	0.1070	0.2140
NIM	-0.0117	0.0449	0.0899
NIST	0.0003	0.0259	0.0519
NMIJ	0.0022	0.0201	0.0401
NRC	0.0154	0.0091	0.0181
PTB	-0.0210	0.0104	0.0209
BIPM (*h*(IPK))	0.0188	0.0138	0.0276

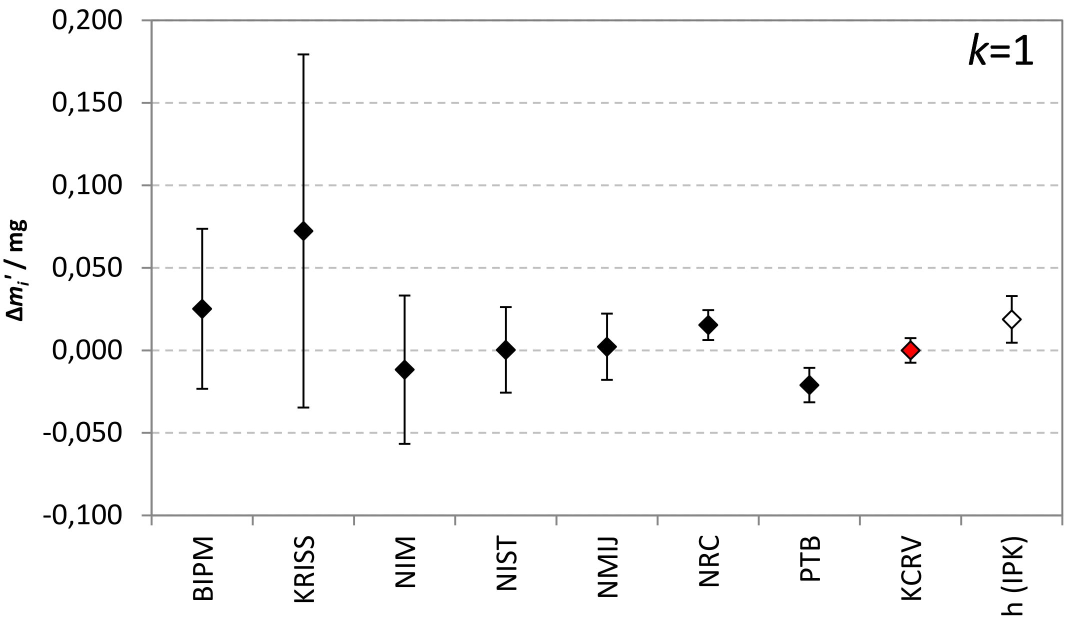

Fig. 3: Differences Δm_i between mass values attributed to 1 kg mass standards using the realization experiment of the participants and the KCRV, calculated as the weighted mean. The difference between the realization based on the BIPM working standards, traceable to the Planck constant through the IPK, and the reference value is also indicated.

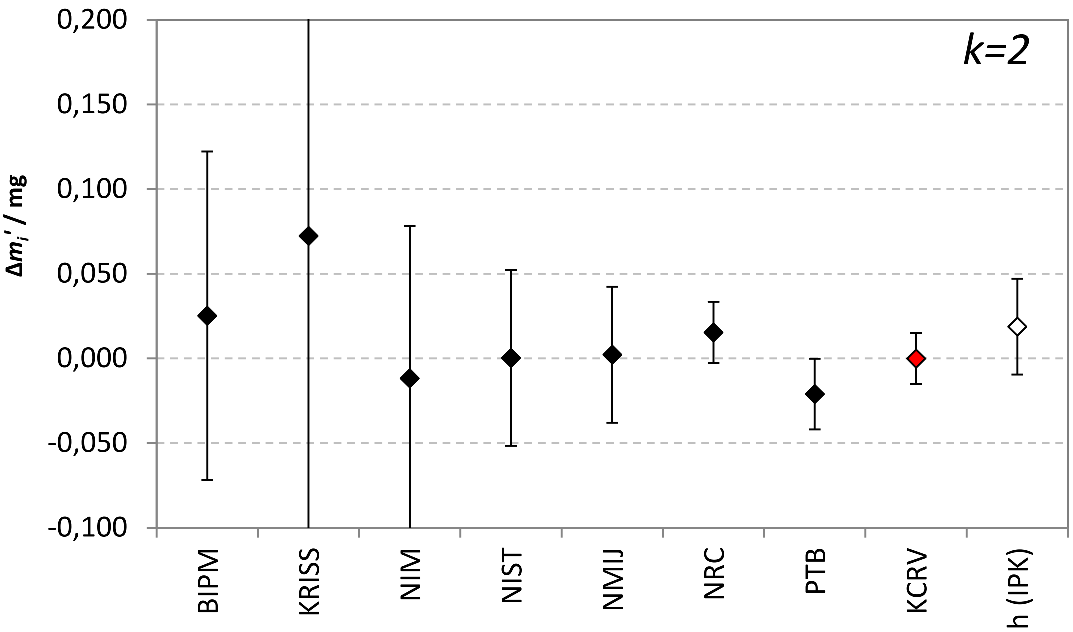

Fig. 4: Same as figure 3, with uncertainty bars showing the expanded standard uncertainty for k=2.

To test the consistency of the data set, the chi-squared value of the set was calculated from the data in table 8 according to

$$\chi^2_{obs} = \sum_{i=1}^{7} \frac{{\Delta m'_i}^2}{u^2(\Delta m'_i)} \qquad \text{(eq. 9)}$$

The experimental chi-squared value is 7.8. The chi-squared value for six degrees of freedom at the 95 % cut-off of the distribution is 12.6. This test is passed although the two results with the smallest uncertainty are not in agreement with each other. A more conservative criterion is to require that the observed chi-squared lies within the expectation value of the chi-squared distribution plus its standard deviation. For ν degrees of freedom this value is $\nu + \sqrt{2\,\nu}$, for 6 degrees of freedom one obtains 9.5. This test is also passed.

It is worth noting that the difference between the two participants with the smallest uncertainties, NRC and PTB, is very close to what was observed in the data set used by the CODATA Task Group on Fundamental Constants in 2017 [7]: 0.036 mg in the present comparison, 4.1×10^{-8} in terms of h (corresponding to 0.041 mg) in 2017.

7 Summary

This key comparison had the objective of comparing realizations of the kilogram based on four Kibble balances, a joule balance and two applications of the XRCD method. The participants determined the mass of one or two 1 kg mass standards traceable to their primary methods and sent them to the BIPM for comparison. At the BIPM all these travelling standards were compared with each other in a vacuum mass comparator using a BIPM mass standard as the reference. The chi-squared test for consistency using the 95 % cut-off criterion was passed, although the two results with the smallest uncertainty are not in agreement with each other.

The key comparison reference value is calculated as the weighted mean of the results. It has a deviation of -0.0188 mg with respect to the mass unit maintained by the BIPM working standards, with a standard uncertainty of 0.0075 mg. This mass value will be the contribution of the present comparison to the calculation of the CCM consensus value.

The stability of the travelling standards could not be verified for all participants due to the travel restrictions during the time of the comparison. The travelling standards of two participants were stable to within 0.001 mg to 0.002 mg. For another participant the changes were at the level of 0.005 mg to 0.008 mg. The largest change was observed on a Pt-Ir prototype which showed many defects due to intense use. In general, the outcome of this comparison is not compromised by the behavior of the travelling standards, and does reflect the capabilities of the participants' realization experiments.

References

[1] M. Stock, R. Davis, E. de Mirandés, M. Milton, "The revision of the SI – the result of three decades of progress in metrology", *Metrologia* **56** (2019) 022001 and Corrigendum in *Metrologia* **56** (2019) 049502

[2] 9th edition of the SI Brochure, available on the BIPM web site: www.bipm.org

[3] *Mise en pratique* for the definition of the kilogram in the SI, available in Appendix 2 of the 9th edition of the SI brochure on the BIPM web site: www.bipm.org

[4] I. A. Robinson, S. Schlamminger, "The watt or Kibble balance: a technique for implementing the new SI definition of the unit of mass", *Metrologia* **53** (2016) A46-A74

[5] Z. Li, *et al*., "The first determination of the Planck constant with the joule balance NIM-2", *Metrologia* **54** (2017) 763-774

[6] K. Fujii *et al*., "Realization of the kilogram by the XRCD method", *Metrologia* **53** (2016) A19-A45

[7] D. B. Newell *et al*., "The CODATA 2017 values for h, e, k, and N_A for the revision of the SI", *Metrologia* **55** (2018) L13-L16

[8] CCM, 2019, "CCM detailed note on the dissemination process after the redefinition of the kilogram", available on the BIPM web site: www.bipm.org

[9] M. Stock *et al*., "Comparison of future realizations of the kilogram", *Metrologia* **55** (2018) T1-T7

[10] M. Stock *et al*., "Calibration campaign against the international prototype of the kilogram in anticipation of the redefinition of the kilogram part I: comparison of the international prototype with its official copies", *Metrologia* **52** (2015) 310-316

[11] H. Fang, F. Bielsa, S. Li, A. Kiss, M. Stock, "The BIPM Kibble balance for realizing the kilogram definition", *Metrologia* **57** (2020) 045009

[12] D. Kim *et al*., "Realization of the kilogram using the KRISS Kibble balance", *Metrologia* **57** (2020) 055006

[13] Z. Li *et al*., "The upgrade of NIM-2 joule balance since 2017", *Metrologia* **57** (2020) 055007

[14] D. Haddad *et al*., "Measurement of the Planck constant at the National Institute of Standards and Technology from 2015 to 2017", *Metrologia* **54** (2017) 633–641

[15] Y. Azuma *et al*., "Improved measurement results for the Avogadro constant using a ^{28}Si-enriched crystal" *Metrologia* **52** (2015) 360−375

[16] N. Kuramoto *et al*., "Determination of the Avogadro constant by the XRCD method using a ^{28}Si-enriched sphere", *Metrologia* **54** (2017) 716−729

[17] B. Wood, C. Sanchez, R. Green, J. Liard, "A summary of the Planck constant determinations using the NRC Kibble balance", *Metrologia* **54** (2017) 399−409

[18] G. Bartl *et al*., "A new ^{28}Si single crystal: counting the atoms for the new kilogram definition", *Metrologia* **54** (2017) 693-715

[19] K. *Fujii*, E. Massa, H. Bettin, N. Kuramoto and G. Mana, "Avogadro constant measurements using enriched ^{28}Si monocrystals", *Metrologia* **55** (2018) L1 – L4

[20] M. Cox, C. Eiø, G. Mana, F. Pennecchi, "The generalized weighted mean of correlated quantities" *Metrologia* **43** (2006) S268-S275