Report on CCM Pilot Study CCM.R-kg-P1

Comparison of future realizations of the kilogram

Final Report

27 June 2017

M. Stock[1], P. Barat[1], F. Beaudoux[2], P. Espel[2], P. Pinot[2], F. Piquemal[2], M. Thomas[2], D. Ziane[2], P. Abbott[3], D. Haddad[3], Z. Kubarych[3], J.R. Pratt[3], S. Schlamminger[3], K. Fujii[4], K. Fujita[4], N. Kuramoto[4], S. Mizushima[4], L. Zhang[4], S. Davidson[5], R.G. Green[6], J. Liard[6], C. Sanchez[6], B. Wood[6], H. Bettin[7], M. Borys[7], I. Busch[7], M. Krumrey[7], A. Nicolaus[7]

[1] BIPM, Pilot laboratory
[2] LNE, France
[3] NIST, USA
[4] NMIJ, Japan
[5] NPL, UK
[6] NRC, Canada
[7] PTB, Germany

Bureau
International des
Poids et
Mesures

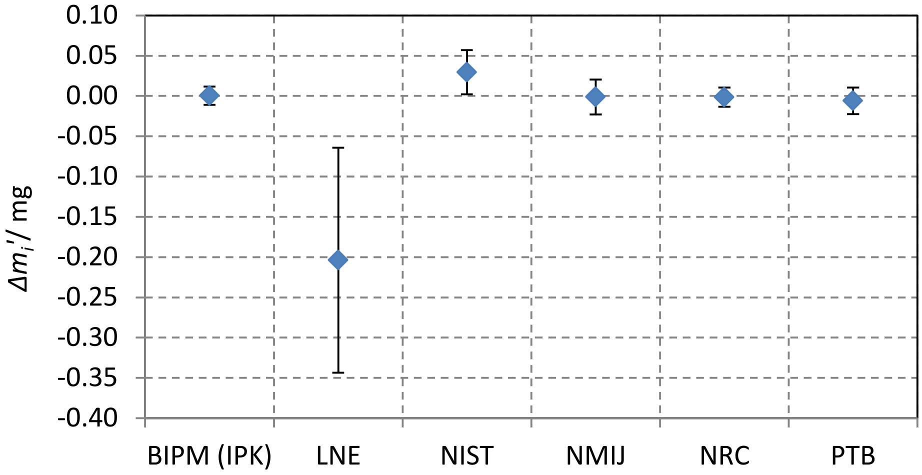
Table of Contents

Executive Summary

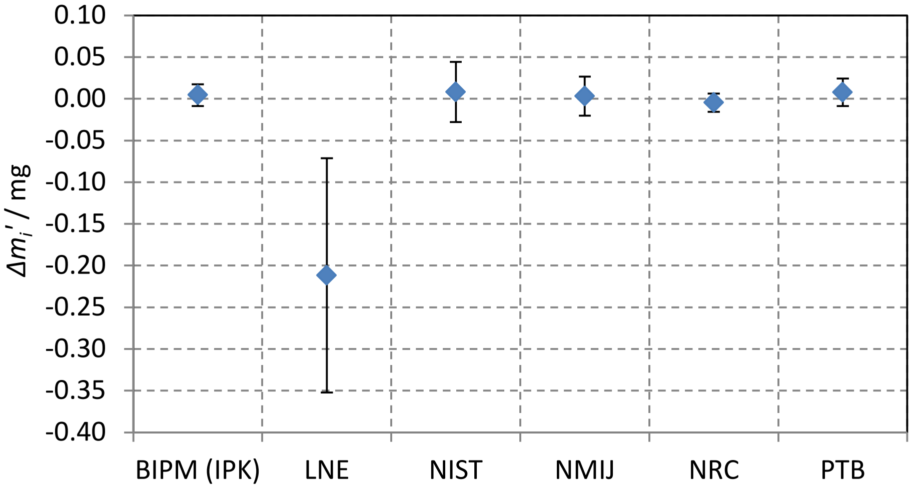
The CCM Pilot Study of future realizations of the kilogram has been carried out on the request of the Consultative Committee for Mass and Related Quantities (CCM) to test (1) the consistency of future realizations of the kilogram carried out by different National Metrology Institutes (NMIs), (2) the continuity with the present definition of the kilogram and (3) to evaluate and optimize the technical protocol for future key comparisons.

The comparison was organized by the BIPM. Participants were the LNE, the NIST, the NMIJ, the NRC and the PTB. The LNE, NIST and NRC used watt/Kibble balances, the NMIJ and the PTB used Avogadro spheres.

The results of calibrations of 1 kg standards under vacuum, using the primary methods of the participants, and traceable to the IPK at the BIPM, are compared in the following figure. The zero line corresponds to the weighted mean of the results of the five NMI participants. The uncertainty bars correspond to standard uncertainties (*k* = 1).



The results of calibrations of 1 kg stainless steel standards in air, based on the primary methods of the participants and traceable to the IPK at the BIPM are compared in the second figure. The zero line corresponds to the weighted mean of the results of the five NMI participants. The uncertainty bars correspond to standard uncertainties (*k* = 1).



Four of the kilogram realizations agree within the standard uncertainties, and also with the present kilogram realization. The LNE result is offset, but in agreement with the others at the level of *k* = 2.

Abstract

This report describes a comparison of realizations of the kilogram using methods which will become primary methods after the planned redefinition of the kilogram. This CCM Pilot Study is one of the essential activities on the joint CCM-CCU roadmap towards the redefinition of the kilogram. The main objectives are to determine the level of agreement of future realizations of the kilogram, carried out by different National Metrology Institutes (NMIs), and to assess the consistency with the present kilogram realization based on the International Prototype of the Kilogram (IPK).

The comparison was organized by the BIPM and had five participants. The LNE, NIST and NRC operated Kibble balances, the NMIJ and the PTB used ^{28}Si-spheres from the International Avogadro Coordination project. The realization methods were used to calibrate two sets of 1 kg travelling standards, with uncertainties ranging from 0.015 mg to 0.140 mg. One set had to be calibrated under vacuum, the other in air. The standards were sent to the BIPM where they were compared with each other and with standards calibrated traceable to the IPK.

The results of the comparison demonstrate a high level of uniformity between the calibrations of 1 kg mass standards with different realization experiments and a high level of consistency between realizations of the future kilogram definition and the as-maintained mass unit of the BIPM, traceable to the IPK.

1 Introduction

It is planned that in the near future four of the seven base units of the International System of Units (SI), the kilogram, the ampere, the kelvin and the mole will be redefined [1]. It is expected that the decision will be taken at the General Conference on Weights and Measures (CGPM) in November 2018. The present definition of the kilogram, the mass of the International Prototype of the Kilogram (IPK), would then be replaced by a definition based on the fixed numerical value of the Planck constant (and on the definitions of the second and the metre).

The future definition of the kilogram does not prescribe any particular experiment for the practical realization of the mass unit. Any experiment or method for determining a mass, of any value, in terms of the Planck constant without the use of a mass standard can be used. Such experiments are called primary methods or realization experiments. An artefact whose mass has been calibrated by a primary method becomes a primary mass standard and can be used to disseminate the mass unit to secondary mass standards.

The new definition will in principle allow any National Metrology Institute (NMI) to realize the kilogram, by developing a realization experiment. At present the draft *mise en pratique* of the definition of the kilogram describes the Kibble[1] balance [2] and the X-ray crystal density (XRCD) technique using silicon spheres [3]. This makes the new definition universal but leads to the question of the consistency of the independent realizations. Another important aspect is the continuity between realization based on the present definition and those based on the future definition. The size of the kilogram should not change noticeably as a consequence of the redefinition.

The Consultative Committee for Mass and Related Quantities (CCM) has developed a *Roadmap towards the redefinition of the kilogram* [4], which identifies the main steps towards the ratification of the redefinition. An important task on the roadmap is the Pilot Study of kilogram realizations, as a crucial test of the future mass realization and dissemination schemes. In October 2013 the BIPM was selected as pilot laboratory for this study.

---

[1] The CCU decided in its meeting in June 2016 to refer to the watt balance as the “Kibble balance” in homage to Bryan Kibble, who originally conceived the idea of this experiment.

2 Objectives of the Pilot Study

The purpose of the Pilot Study is to check the realization and dissemination of the new kilogram as described in the proposed *mise en pratique* for the future definition of the kilogram [5]. The primary aim is to test the consistency of kilogram realizations based on different realization experiments. The Pilot Study will also test the continuity between the kilogram realizations according to the future definition and according to the present definition. Since measurements of the Planck constant are traceable to the IPK and the recent extraordinary calibration campaign using the IPK in 2014 [6, 7] has improved its disseminated uncertainty, one can expect that this continuity should indeed exist within the uncertainties. The Pilot Study will allow verification of this assumption directly. Additionally the experience gained from the Pilot Study will optimize the technical protocol for future comparisons of kilogram realizations.

To achieve the lowest possible uncertainty, the realization experiments are typically operated under vacuum, and the primary mass standards are calibrated under vacuum. The protocol for the Pilot Study has been designed such that realizations of the kilogram obtained by different realization experiments may be compared as directly as possible. This part of the Pilot Study requires a comparison of mass standards under vacuum.

Although for highest accuracy, the future realization of the mass unit will be carried out under vacuum, it is expected that practical mass dissemination will continue to be done with stainless steel standards kept in air. Making the mass unit available in air requires the application of sorption corrections for mass standards used under vacuum and in air, and the application of buoyancy corrections. Therefore, the Pilot Study will compare the dissemination of the mass unit with stainless steel standards kept in air.

3 Organization of the Pilot Study

3.1 General principle

The Pilot Study was carried out by using two sets of 1 kg travelling standards per participant, which were provided by each of the participants from its own standards:

Set 1: One 1 kg Pt-Ir standard and optionally one 1 kg standard of the participant’s choice, to be calibrated as directly as possible with the realization experiment (under vacuum if possible);

Set 2: Two 1 kg stainless steel standards, to be calibrated in air, traceable to the realization experiment.

The standards of Set 1 had to be calibrated as directly as possible with respect to the realization experiments. This could be a direct calibration, under vacuum, in a Kibble balance or in a mass comparator against an Avogadro sphere. Another possibility would be the use of an intermediate transfer standard, which was calibrated using the realization experiment and which was then used in a second step to calibrate the travelling standards in a vacuum mass comparator. The mass values of the standards of Set 1 were calculated by all participants from the same value of the Planck constant. The comparison protocol stated that the value recommended by the 2014 CODATA fundamental constants adjustment [8],

$$h = 6.626\,070\,040 \times 10^{-34}\ \text{Js}$$

should be used. The particular choice of this value is not relevant for the investigation of the uniformity between different realizations, as long as all NMIs use the same value. However, the choice of the value impacts the continuity between the present and the future realizations. It is for this reason that the CODATA value has been selected. It should be noted that the numerical value for the Planck constant which will ultimately be used in the new definition of the kilogram, could be different from the 2014 CODATA value.

Set 2 consists of two 1 kg stainless steel standards which had to be calibrated in air and traceable to the realization experiment. This required making a correction for surface sorption and applying a buoyancy correction for the weighing in air.

Both sets of calibrated travelling standards were sent to the pilot laboratory, BIPM, where the standards from all participants were compared with each other, the standards of Set 1 under vacuum, and those of Set 2 in air. By using the results of these comparisons and the calibration results provided by the participants, it is possible to investigate the consistency of the NMIs’ kilogram realization results. Since the BIPM working standards are traceable to the IPK, it is also possible to investigate the agreement between the new realizations and the IPK.

The primary objective of the Pilot Study is to compare the mass units as realized and disseminated in practice by using Kibble balances or XRCD experiments. The participants were therefore asked to use those calibration procedures which they plan to apply after the redefinition for the calibration of both sets of standards in the Pilot Study.

3.2 Measurement sequence

After the initial calibration at the NMIs, the standards were hand-carried to the BIPM. Although the standards of Set 1 had to be calibrated under vacuum, they were stored in containers filled with air during transport, since no practical solution for shipping under vacuum exists. The unavoidable air-to-vacuum transfers will have some impact on the mass stability, but recent investigations indicated that these effects should be small with respect to the uncertainties of the realization experiments [9]. Mass comparisons were carried out at the BIPM among all standards of Set 1 under vacuum (except No. 13 sent by the LNE) and among all standards of Set 2 in air. The travelling standards of both sets were also compared with BIPM working standards, the masses of which are calibrated traceable to the IPK from comparisons made in 2014. For the weighings under vacuum of the standards of Set 1, a set of two 1 kg Pt-Ir sorption artefacts was used to link with the Pt-Ir BIPM working standards that are kept in air.

To verify the stability of the standards of Set 1 during transportation, the participants also determined the mass of these standards in air, traceable to their national prototype (and ultimately to the IPK), before sending them to the BIPM and after receiving them back. The BIPM determined the mass of the standards in air upon receipt and before return. Since the participants’ and the BIPM’s measurements were all traceable to the IPK, this allowed the detection of any significant mass changes during transportation.

The full series of measurements for the standards of Set 1 was as follows:

- Mass of travelling standards (under vacuum, except at LNE) calibrated traceable to realization experiment (Kibble balance, Avogadro sphere), at each NMI
- Mass of travelling standards (in air) calibrated with respect to Pt-Ir reference standard (traceable to the IPK), for a check of the stability during transport, at each NMI
- Standards transferred to BIPM (in air)
- All travelling standards compared at BIPM in air with BIPM working standards in Pt-Ir, for a check of the stability during transport
- All travelling standards compared under vacuum (except No. 13 of LNE), including BIPM sorption artefacts to establish traceability to the IPK held in air

- All travelling standards compared at BIPM in air with BIPM working standards in Pt-Ir, to enable a check of transport stability upon receipt by the NMI
- Standards returned to NMIs (in air)
- Mass of travelling standards (in air) calibrated with respect to Pt-Ir reference standard (traceable to the IPK) at each NMI, for a check of the stability during transportation.

For the standards of Set 2, the sequence of measurements was the following:

- Travelling standards calibrated in air with NMI stainless steel standards, traceable to the realization experiment
- Standards transported to BIPM
- All travelling standards compared directly to one another and against BIPM working standards in stainless steel
- Standards returned to NMIs
- Travelling standards compared in air with NMI stainless steel standards traceable to the realization experiment

The travelling standards arrived at the BIPM between 25 March and 26 April 2016. Measurements were made at the BIPM from May to July 2016. The travelling standards were returned to the NMIs in the period from 8 August to 29 September 2016.

4 Participants and travelling standards

All NMIs working on primary methods were invited to participate in the Pilot Study, under the condition that they would be able to realize the kilogram according to its future definition with a relative standard uncertainty below 2 parts in 10^7, that is 200 µg at the 1 kg level. Five institutes participated (table 1): LNE (France), NIST (USA), NMIJ (Japan), NRC (Canada) and PTB (Germany). The LNE, NIST and NRC used Kibble balances, the NMIJ and the PTB used the ^{28}Si-spheres AVO28-S5c and AVO28-S8c, respectively, from the International Avogadro Coordination [10] as the basis for their calibrations.

Table 1: Participants of the CCM Pilot Study, and their realization method.

Institute	Contact person(s)	Realization method
LNE	François Piquemal Patrick Pinot	Kibble balance
NIST	Jon Pratt Patrick Abbott	NIST-4 Kibble balance
NMIJ	Shigeki Mizushima	XRCD method, AVO28-S5c
NRC	Carlos Sanchez	Kibble balance
PTB	Horst Bettin Michael Borys	XRCD method, AVO28-S8c

Table 2 lists the travelling standards sent by the participants for Set 1 (for calibration under vacuum as directly as possible with realization experiment). The LNE requested that prototype No. 13 not be weighed under vacuum. It was compared with the other travelling standards by an air-to-vacuum transfer with BIPM sorption standards.

Table 2: Travelling standards of Set 1 and their properties as communicated by the participants prior to the start of the Pilot Study.

Institute	Identification of standard	Type	Approx. deviation from 1 kg / mg	Volume at 20 °C / cm^3	Unc. of volume / cm^3
LNE	No. 13	Pt-Ir prototype	-0.1068	46.4181 (at 0 °C)	0.0001
NIST	K104	Pt-Ir prototype	0.386	46.41000	0.0003
	141714	Mettler-Toledo stainless steel	0.4	124.79757	0.0011
NMIJ	No. 94	Pt-Ir prototype	0.3	46.4334	0.0003
	E59	Pt-Ir from Stanton Instrum.	4.9	46.4095	0.0004
NRC	K50	Pt-Ir prototype	Not comm.	46.53041	0.0003
PTB	Pt109	Pt-Ir prototype	<1	46.41460	0.0003
	Si14-02	nat. Si-sphere	<10	429.351730	0.000025

Table 3 shows the travelling standards sent by the participants for Set 2 (for calibration in air, traceable to realization experiment).

Table 3: Travelling standards of Set 2 and their properties as communicated by the participants prior to the Pilot Study. All standards are made of stainless steel.

Institute	Identification of standard	Manufacturer and Type	Approx. deviation from 1 kg / mg	Volume at 20 °C / cm^3	Unc. of volume / cm^3
LNE	E	Mettler Toledo, OIML E0	0.144	125.612	0.004
	INM	Mettler Toledo, OIML E2	0.942	125.6327	0.0042
NIST	Zwiebel 7	Zwiebel, OIML one piece	0.02	125.1553	0.0050
	Zwiebel 8	Zwiebel, OIML one piece	0.275	125.1167	0.0050
NMIJ	S1_2	Chyo balance, right cylinder	-1.3	126.9007	0.0004
	S2_1	Chyo balance, right cylinder	0.4	126.8905	0.0004
NRC	HSA2	Häfner, right cylinder	Not comm.	124.81504	0.00012
	HSA3	Häfner, right cylinder	Not comm.	124.81542	0.00012
PTB	D1	PTB, cylinder	<3	126.2680	0.0010
	D2	PTB, cylinder	<3	126.2682	0.0010

5 Traceability schemes of the participants

In the following we describe briefly the traceability schemes chosen by the participants for the realization of the kilogram. Full details can be found in the NMIs' measurement reports, which are reproduced in annex 1.

5.1 LNE

At the time when the measurements for the Pilot Study were made, the LNE Kibble balance was not yet operating under vacuum. In addition, the prototype No. 13 showed a very regular linear drift with time over more than a century and should not be used under vacuum.

A 500 g Pt-Ir mass standard W1 was calibrated in air on the LNE Kibble balance. W1 served for the calibration of a similar 500 g Pt-Ir mass standard W2 using an M_one mass comparator. W1 and W2 were then used together to determine the mass in air of the travelling standard No. 13.

W1 and W2 were also used together to calibrate the 1 kg Pt-Ir standard PtIV. This served to determine the mass of the stainless steel standard MET7, which in turn was used to calibrate the two stainless steel transfer standards of Set 2. The buoyancy correction was made using artefacts.

5.2 NIST

The travelling standards of Set 1 were measured under vacuum directly in the NIST-4 Kibble balance. The air-to-vacuum correction of K104 was determined using sorption artefacts K105 (cylinder) and C18 (stack). Both have the same surface characteristics as K104. The travelling standards of Set 2 were compared directly with K104 in air. The buoyancy correction was made using the CIPM-2007 formula.

5.3 NMIJ

The NMIJ used the sphere AVO28-S5c from the International Avogadro Coordination. A new determination of the core volume was carried out by optical interferometry. The mass of the surface layers was determined by new XPS and ellipsometry investigations. For the lattice constant, the relative atomic mass, and the influence of point defects, the results from the previous determination of the Avogadro constant were used [10]. The travelling standards were compared with AVO28-S5c in a vacuum mass comparator.

For the calibration of the standards of Set 2, the mass of AVO28-S5c in air was determined using sorption artefacts. The sphere was then used to calibrate the stainless steel standard S2_2. The buoyancy correction was carried out using a set of buoyancy artefacts. S2_2 served to calibrate the two travelling standards S1_2 and S2_1. For these measurements, the buoyancy correction was made with the CIPM-2007 equation.

5.4 NRC

Vacuum cycling experiments were performed to both stabilize and determine the stability of masses during air-vacuum cycling. Six 500 g masses were cycled six times. Masses were stable within ± 0.4 µg after three cycles. Standard N13SiB4 (500 g, boron doped single crystal Si, diamond turned) was then extracted from the vacuum balance and calibrated under vacuum in the Kibble balance. The standard was then reinserted in the vacuum balance via the load-lock and compared again with the five other standards of the first measurement series. A correction of – 4.4 µg (with the same uncertainty) was applied to correct for the observed mass change. During this series, N13SiB4 was used to calibrate the 500 g standards H11SB1 and H11SB2, both made of stainless steel. Finally, H11SB1 and H11SB2 were used together to calibrate the 1 kg travelling standard K50 under vacuum in an M_one mass comparator.

For the calibration of the standards of Set 2, a set of stainless steel sorption artefacts were compared under vacuum to K50 and the H11SB1-H11SB2 stack traceable to the realization experiment. The stainless steel sorption artefacts were then corrected for their sorption and used to calibrate the stainless steel travelling standards in air. The air buoyancy correction was carried out using the CIPM-2007 formula.

5.5 PTB

The PTB used the sphere AVO28-S8c from the International Avogadro Coordination. Only the volume of the sphere and the surface layers were measured anew. For the other parameters, the results from the previous determination of the Avogadro constant were used [10]. The travelling standards of Set 1 were compared with AVO28-S8c under vacuum in an M_one mass comparator.

The Pt-Ir cylinder PtSk-Z was calibrated against AVO28-S8c in vacuum. This cylinder is one of the two standards used as sorption artefacts for the link between the mass of AVO28-S8c under vacuum and the mass of the two travelling standards of Set 2 in air. The buoyancy correction was made using the artefact method.

Since the PTB used the sphere AVO28-S8c and the NMIJ the sphere AVO28-S5c, their results are partly correlated. The core volumes and surface layers have been independently re-determined in both institutes. The point defect corrections are different and only partly correlated. All other correlated uncertainty contributions have been estimated by the PTB as 8 µg and are probably the same as in the measurements of NMIJ, resulting in a correlation coefficient of about 13 % for the masses of the spheres.

6 Measurements at the BIPM

6.1 Preparatory measurements

All measurements were carried out with the BIPM M_one mass comparator (Mettler Toledo), equipped with a six-place mass exchanger. In March 2016 a complete study on the influence of the mass handler position was carried out. Its contribution was not significant, the uncertainty due to the eccentric loading has been estimated as 0.8 µg. This study was repeated in December 2016, with the same result.

Typically at the BIPM, only platinum-iridium and stainless steel standards are involved in comparisons. The sensitivity for the M_one mass comparator is checked during each comparison and the nonlinearity is checked using a sensitivity weight of 95 mg. The latter corresponds approximatively to the air buoyancy correction to be applied to 1 kg standards made of stainless steel when compared to 1 kg standards made of platinum-iridium. For the present work the nonlinearity has been checked using a sum of weights of 460 mg which corresponds approximatively to the air buoyancy correction to be applied to 1 kg standards made of silicon when compared to 1 kg standards made of platinum-iridium. The uncertainty associated with the error of nonlinearity for calibrations in air of a silicon sphere has been estimated as 0.9 µg.

6.2 Comparison measurements

The measurement series carried out at the BIPM are shown in table 4. The series shown in grey are the principal mass comparisons for the Pilot Study. The other series served to calibrate BIPM reference standards and to verify their stability during the comparison.

All measurements were made with the M_one mass comparator. All weighing data of a series were grouped together and mass values were determined using a least-squares adjustment.

The principal reference standards for the study were the two BIPM Pt-Ir working standards Nos. 63 and 88. They were calibrated against the set of BIPM working standards for limited use (three Pt-Ir standards) in April 2016. The latter was calibrated with respect to the IPK in 2014. Nos. 63 and 88 were compared at the beginning, in the middle and at the end of the Pilot Study (series 1, 7, 11) with four other Pt-Ir standards: Nos. 42', 77, 97 and 103. These comparisons showed mass variations of Nos. 63 and 88 of about ± 2 µg. Their mass was interpolated linearly between the dates of the verifications.

Nos. 63 and 88 were directly used as references for the comparison of the travelling standards of Set 1 in air (series 2, 10). A set of air density artefacts, Cc2 and Cp2, was used for buoyancy corrections. The measurements on Set 1 in air, at the beginning and the end of the study, were treated in one single adjustment, but the masses of the standards at the beginning and the end were assumed to be independent. In total 96 mass differences served as input data to determine the masses of the 8 travelling standards, at the beginning and the end. The masses of Nos. 63 and 88, interpolated as explained above, served as constraints for the adjustment. The typical statistical uncertainty of the fit was 1.3 µg.

For the vacuum comparisons the BIPM sorption artefact A0 served as reference (series 4). The pair of sorption artefacts A0 (cylinder) and A18 (stack) was compared before and after the vacuum measurements on 19-20 May and 30 May-1 June (series 3, 5) with standards Nos. 63 and 88. The mass of A0 was stable to better than 1 µg during the vacuum comparisons. A linear interpolation was applied. One-hundred twenty mass differences were available to determine the masses of the 7 transfer standards of Set 1 (excluding No. 13 from LNE). The mass of A0 served as the constraint. The statistical uncertainty of the fit was 0.7 µg.

During all measurements under vacuum, the M_one mass comparator stayed under vacuum; the loading of travelling standards was done through the Artefact Transfer Device. The pair of sorption artefacts A0 and A18 stayed inside the chamber of the M_one mass comparator and was not exposed to air during loadings.

The stainless steel standards of Set 2 were compared with the BIPM stainless steel references C1 and C2 (series 8). They were compared with Nos. 63 and 88 on 7-8 June and 30 June-4 July 2016 (series 6, 9). The buoyancy correction was made using the buoyancy artefacts Cc2 and Cp2. Between these dates a mass increase of about 2 µg was observed for both stainless steel standards. A linear interpolation was applied to calculate their mass between the dates of verification. The buoyancy corrections for the comparison of the Set 2 standards (all in stainless steel) with C1 and C2 were made using the CIPM-2007 formula.

For the measurements on the standards of Set 2, the masses of the two BIPM standards C1 and C2, linearly interpolated, were used as constraints. A set of 150 mass differences served as input data for the determination of the masses of the 10 travelling standards. The statistical uncertainty of the fit was 0.7 µg.

The fact that the fitting uncertainty was largest for the weighings of Set 1 in air can be explained by the fact that these weighings required the largest buoyancy corrections because Set 1 contains masses in Pt-Ir, stainless steel, and Si.

Table 4: Measurement series carried out at the BIPM. Grey fields indicate the principal comparison measurements, the others serve to verify the stability of BIPM reference standards. The BIPM standards Nos. 841 and 691 were used to fill all six positions in the mass comparator but were not used for the Pilot Study.

Measurement series	Start date	End date	BIPM standards	NMI travelling standards
1 - Evolution of Pt-Ir references	5/5/16	6/5/16	42', 63, 77, 88, 97, 103	
2- Set 1, comparison in air, at arrival	9/5/16	10/5/16	63, 88, Cc2, Cp2	E59, 94
	10/5/16	11/5/16	63, 88, Cc2, Cp2	K50, 13
	11/5/16	12/5/16	63, 88, Cc2, Cp2	K104, 141714
	12/5/16	13/5/16	63, 88, Cc2, Cp2	Pt109, Si14-02
3- Evolution of A0, A18 in air	19/5/16	20/5/16	63, 88, A0, A18, 841	13
4 - Set 1, comp. under vacuum	20/5/16	23/5/16	A0, A18	E59, 94, K104, 141714
	23/5/16	25/5/16	A0, A18, 691	K50, K104, 141714
	25/5/16	27/5/16	A0, A18, 691	K50, Pt109, Si14-02
	27/5/16	29/5/16	A0, A18	E59, 94, Pt109, Si14-02
5 - Evolution of A0, A18 in air	30/5/16	1/6/16	63, 88, A0, A18, 691	13
6 - Evolution of C1, C2 in air	7/6/16	8/6/16	63, 88, C1, C2, Cc2, Cp2	
7 - Evolution of Pt-Ir references	12/6/16	13/6/16	42', 63, 77, 88, 97, 103	
8 - Set 2, comparison in air	16/6/16	17/6/16	C1, C2	S1_2, S2_1, E, INM
	17/6/16	20/6/16	C1, C2	Zwiebel7, Zwiebel8, E, INM
	20/6/16	21/6/16	C1, C2	Zwiebel7, Zwiebel8, HSA2, HSA3
	21/6/16	24/6/16	C1, C2	D1, D2, HSA2, HSA3
	24/6/16	27/6/16	C1, C2	D1, D2, S1_2, S2_1
9 - Evolution of C1, C2 in air	30/6/16	4/7/16	63, 88, C1, C2, Cc2, Cp2	
10 - Set 1, comparison in air, before departure	5/7/16	6/7/16	63, 88, Cc2, Cp2	E59, 94
	6/7/16	7/7/16	63, 88, Cc2, Cp2	K50, 13
	8/7/16	11/7/16	63, 88, Cc2, Cp2	K104, 141714
	11/7/16	12/7/16	63, 88, Cc2, Cp2	Pt109, Si14-02
11 - Evolution of Pt-Ir references	20/7/16	21/7/16	42', 63, 77, 88, 97, 103	

7 Comparison in terms of the Planck constant

The measurements made by the participants for the calibration of the standards of Set 1 with their realization experiments also enabled the determination of a value for the Planck constant (assuming the mass of the standards is also established with respect to the IPK). The results communicated by the NMIs are shown in table 5, together with the latest published results. Figures 1 and 2 show the results in graphical form, as relative deviations from the CODATA 2014 recommended value, $h_{2014} = 6.626\ 070\ 040 \times 10^{-34}$ Js.

Table 5: Latest published values of the Planck constant and values determined during the Pilot Study.

Institute	Latest published value of *h* / 10^{-34} Js	Reference	Value of *h* from Pilot Study / 10^{-34} Js	Reference
LNE	6.626 068 8(20)	[11]	6.626 071 33 (93)	Report
NIST	6.626 069 83 (22)	[12]	6.626 069 86 (24)	Report
NMIJ	6.626 070 15 (13)	[10]	6.626 070 06 (16)	Priv. comm.
NRC	6.626 070 11 (12)	[13]	6.626 070 03 (10)	Priv. comm.
PTB	6.626 070 15 (13)	[10]	6.626 070 07 (13)	Report

The results of the NMIJ, NRC and the PTB are in very good agreement. The NIST result is slightly lower but the difference is comparable with the standard uncertainties. The LNE result obtained during the Pilot Study shows a deviation larger than the standard uncertainty, but still consistent at the level of $k = 2$.

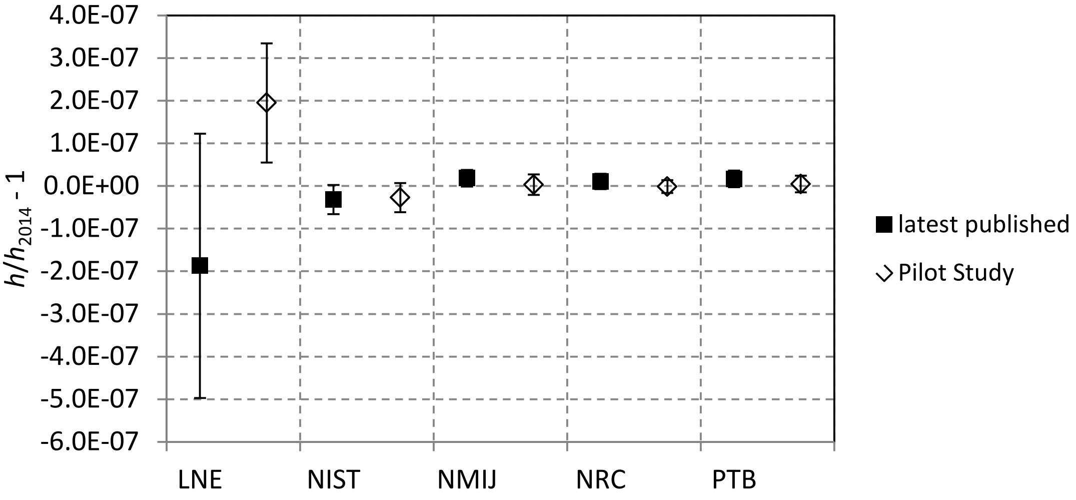


Fig. 1: Results for the Planck constant obtained during the Pilot Study, and latest published results. Uncertainty bars represent standard uncertainties (k = 1).

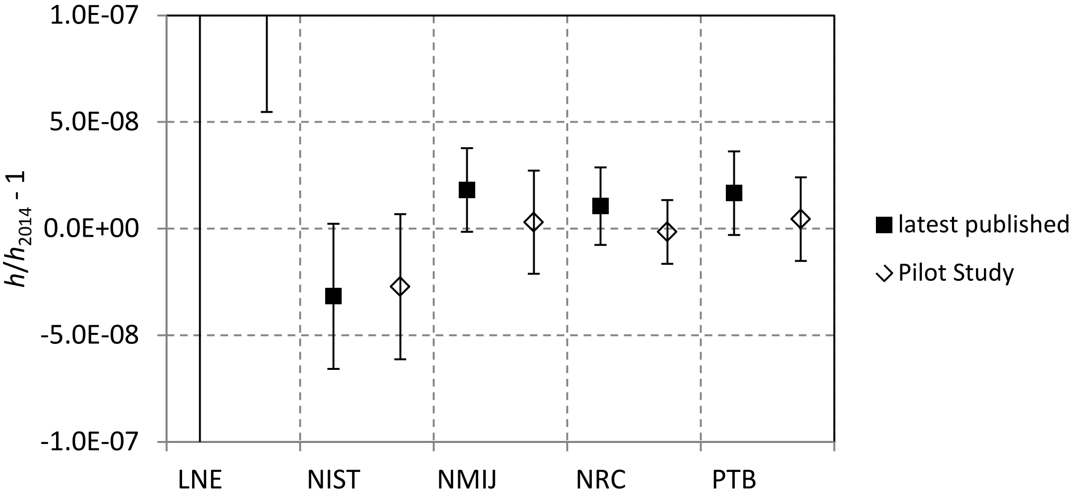


Fig. 2: Same as figure 1, but with a five-times expanded vertical scale. Uncertainty bars represent standard uncertainties (k = 1).

8 Results for the travelling standards of Set 1

8.1 Stability of the travelling standards of Set 1

The results of the Pilot Study are influenced by the mass stability of the travelling standards during transportation from the NMIs to the BIPM. The stability during the transport back to the NMIs is of less importance, because the calibrations against the primary realization experiments were only made at the NMIs *before* sending the standards to the BIPM.

To assess the stability of the standards during transportation, they were calibrated at the NMIs before sending them to the BIPM and after receiving them back. At the BIPM, the travelling standards were calibrated after the receipt of all standards and before return. These measurements were made in air and in terms of the present definition of the kilogram.

Table 6 shows the results of these measurements. Figures 3 and 4 show them in graphical form. In these graphs the numerical values are normalized for each standard to the value obtained by the NMI before sending it to the BIPM.

According to the measurement report from NIST, standard 141714 showed some instability during measurements in air following its calibration in the NIST-4 Kibble balance. The mass value just after receiving the mass from the Kibble balance was 1000.000 424 8 g. Over the next month the mass gradually decreased to 1000.000 398 g, which is a change of -0.0268 mg. The value in table 6 for the mass of standard 141714 obtained at the NIST before sending the standard to the BIPM is the last value obtained. The corresponding points in figures 3 and 4 represent the same mass value.

As requested by the PTB, the sphere Si14-02 was washed three times by the BIPM by applying the NMIA method before it was weighed under vacuum, that is, between the initial and final BIPM measurement in air.

The differences between the initial NMI and BIPM measurements depend on a possible difference in the mass scale and on a possible mass change during transportation. The differences range from +9 µg to -10 µg. Although standard 141714, which shows the largest positive change, has already shown some instability at the NIST, this spread seems to be too large to be explained by differences in the mass scale between the BIPM and the participants, in particular because all of them had standards recalibrated with respect to the IPK at the end of 2014. It might be that the mass stability of the standards has been negatively affected by their use under vacuum. This does not explain the difference of 9 µg of No. 13 which has never been used under vacuum.

One can also not exclude the possibility of an anomaly in the “BIPM arrival” results, which for several standards are below the initial NMI results. In contrast to this the “BIPM departure” results agree in general quite well with the final NMI results. However, the data files from the initial BIPM weighings do not indicate any anomaly. Since the NMIs’ measurement reports arrived much later, this effect could not be noticed at the time of the BIPM measurements.

Table 6: Results of the mass determinations of the standards of Set 1 in air, with respect to the IPK, at the NMIs and the BIPM.

		Calibration in air with respect to IPK at NMI, before BIPM		**Calibration in air with respect to IPK at BIPM, at arrival**		**Calibration in air with respect to IPK at BIPM, at departure**		**Calibration in air with respect to IPK at NMI, after BIPM**	
Institute	**Identification of standard**	**Deviation from 1 kg / mg**	**Std. unc. / mg**	**Deviation from 1 kg / mg**	**Std. unc. / mg**	**Deviation from 1 kg / mg**	**Std. unc. / mg**	**Deviation from 1 kg / mg**	**Std. unc. / mg**
LNE	n° 13	-0.107	0.005	-0.1157	0.004	-0.1167	0.004	-0.118	0.005
NIST	K104	0.3953	0.0037	0.3881	0.004	0.3903	0.004	0.3910	0.0037
	141714	0.3980[2]	0.0071	0.4074	0.005	0.4152	0.005	0.3818	0.0046
NMIJ	n° 94	0.3356	0.0038	0.3291	0.004	0.3331	0.004	0.3330	0.0038
	E59	4.9281	0.0038	4.9205	0.004	4.9233	0.004	4.9232	0.0038
NRC	K50	-0.072	0.005	-0.0726	0.004	-0.0705	0.004	-0.075	0.005
PTB	Pt109	0.1558	0.0049	0.1514	0.004	0.1560	0.004	0.1565	0.0051
	Si14-02	-4.205	0.012	-4.2145	0.010	-4.2048	0.010	-4.205	0.012

---

[2] The NIST calibration report mentions a significant mass change of this standard before sending it to the BIPM. The value in the table is the last value obtained at NIST before sending the standard.

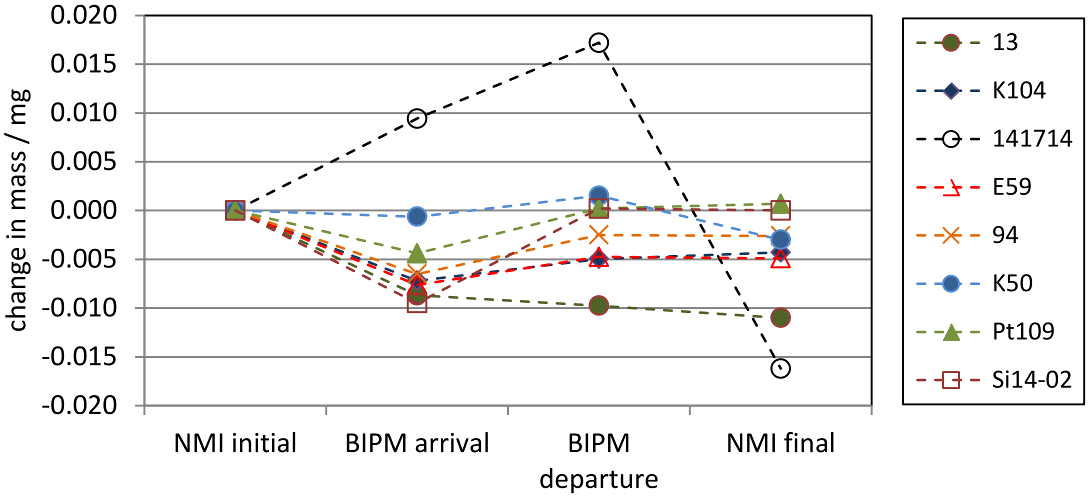


Fig. 3: Mass of standards of Set 1 in air, obtained by the NMIs and the BIPM, to verify the stability of the standards during transportation. All masses traceable to the IPK and normalized to the initial NMI measurement. No uncertainty bars are shown for clarity.

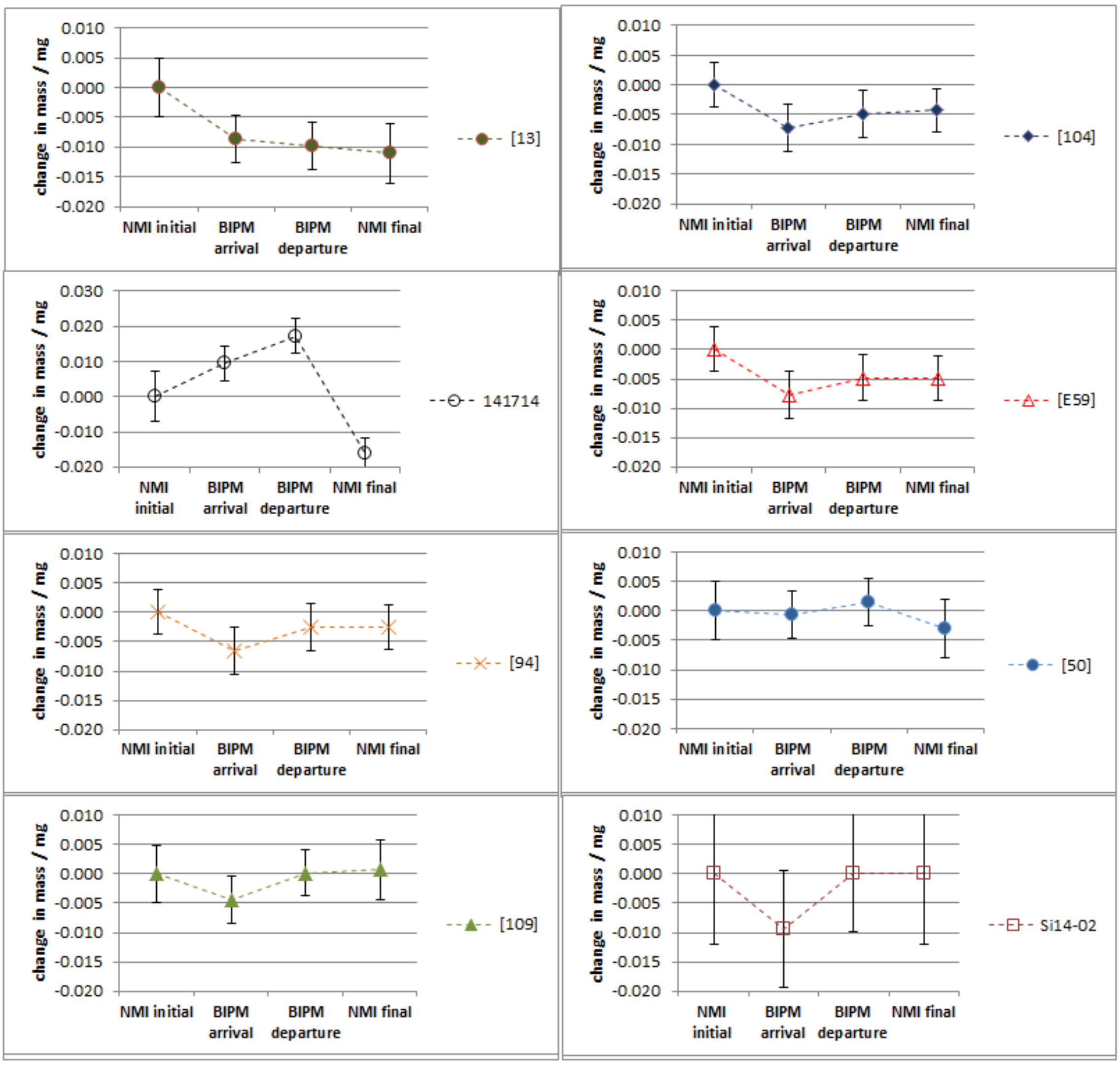


Fig. 4: Same as figure 3, but separate graphs for each travelling standard. Standard 141714 had shown some instability before sending it to the BIPM. The first point on the graph for this standard is the last value obtained at the NIST, before sending the standard to the BIPM.

Following discussion with the participants, the uncertainty related to the transport to the BIPM was estimated using the equation

$$u_{\text{transport}} = \frac{1}{2}\left(\left|m_{\text{BIPM}}^{\text{arrival}} - m_{\text{NMI}}^{\text{initial}}\right| + \left|m_{\text{NMI}}^{\text{final}} - m_{\text{BIPM}}^{\text{departure}}\right|\right) \qquad (1)$$

For all travelling standards, except for 141714, the equation leads to very similar uncertainties, on average 0.004 mg with a standard deviation of 0.001 mg. Therefore a common transport uncertainty of 0.004 mg is used for all standards, except for 141714, where the use of equation (1) leads to 0.0214 mg. These values are shown in the third column of table 7.

The uncertainty of the Pilot Study is also influenced by the stability of the travelling standards under repeated air-vacuum transfers. The relevant quantity is the change of the vacuum mass between the NMI and the BIPM due to the storage and transport in air in between. The data from the study does not allow assessment of this quantity directly. In principle the comparison of the in-air calibrations at the BIPM at the time of receipt and at the time of departure allows the effect of the time spent under vacuum to be quantified. It can be supposed that both effects are of comparable size.

$$u_{\text{air-vac}} = \left|m_{\text{BIPM}}^{\text{departure}} - m_{\text{BIPM}}^{\text{arrival}}\right| \qquad (2)$$

However, as has been discussed above, for some of the travelling standards the mass determined after arrival at the BIPM appears to be too small. This would lead to an overestimation of the effect of cycling of the standards between vacuum and air. It has therefore been agreed with the participants to estimate this uncertainty based on previous experience. For Pt-Ir standard Pt109 PTB reported a mass stability of the order of 0.001 mg. NRC stated in its report a cycling/short term stability of 0.0025 mg. The BIPM made the experience that after a small number of cycles the vacuum mass becomes stable to within 0.001 mg. For the purpose of the Pilot Study the uncertainty related to cycling between vacuum and air has been estimated as 0.002 mg for Pt-Ir standards. Since the stainless steel standard 141714 has about twice the surface of a Pt-Ir standard an uncertainty of 0.004 mg was assumed. For the Si sphere, the PTB reported a mass stability of 0.005 mg, including the effect of repeated cleaning. These uncertainties are indicated in the fourth column of table 7.

Table 7: Uncertainty contributions due to transportation of the standards and to transfer between air and vacuum.

Institute	**Identification of standard**	**Transport uncertainty $u_{\text{transport}}$ / mg**	**Air-vac-transfer unc. $u_{\text{air-vac}}$ / mg**
LNE	n° 13	0.004	0.000 (no vacuum use)
NIST	K104	0.004	0.002
	141714	0.0214	0.004
NMIJ	n° 94	0.004	0.002
	E59	0.004	0.002
NRC	K50	0.004	0.002
PTB	Pt109	0.004	0.002
	Si14-02	0.004	0.005

8.2 Comparison results for Set 1

Table 8 shows in columns 3 and 4 the results of the NMIs' calibrations of the standards of Set 1 under vacuum. An exception is the standard No. 13 of the LNE, which was calibrated in air. The results are expressed as deviations from 1 kg, in milligrams.

The following approach was chosen to determine the degrees of equivalence of the NMIs' calibrations: In a first step for each travelling standard the difference between the calibration at the NMI and the BIPM was calculated. In this way the mass unit maintained at the BIPM served as a common reference to compare the NMI calibrations. For NMIs which sent two standards, the mean of both differences was been calculated. The weighted mean of these differences was calculated to obtain the comparison reference value (CRV). Finally, the differences between the NMIs' results and the CRV were determined. This procedure is described in detail in the following.

As shown in table 4, at the BIPM the travelling standards of Set 1 were compared under vacuum with the pair of BIPM Pt-Ir sorption standards A0 and A18 during four comparisons in the M_one mass comparator. Each travelling standard was included in two groups of standards. The mass of A0 was known with respect to the IPK. A least-squares adjustment was carried out to determine the masses $m_{i,j}^{\mathrm{BIPM}}$ (*i*=1..5, index for NMI; *j*=1,2 index for travelling standards of NMI *i*) of the travelling standards against the known mass of A0, with a statistical uncertainty of 0.7 µg. The results are shown in the two rightmost columns of table 8. Because standard No. 13 could not be brought under vacuum, its mass value was determined in air.

In the first step we calculate the difference in mass between the calibration results provided by the NMIs $m_{i,j}^{\mathrm{NMI}}$ and the calibration at the BIPM:

$$\Delta m_{i,j} = m_{i,j}^{\mathrm{NMI}} - m_{i,j}^{\mathrm{BIPM}} \tag{3}$$

These differences are shown in the third column of table 9. The table also shows the combined uncertainty $u_{i,j}$, obtained by adding the transport and air-vacuum transfer uncertainties from table 7 to the calibration uncertainties stated by the NMIs, from table 8. These additional uncertainties increase the NMIs' calibration uncertainties by less than 6 %, except for the standard 141714, where the increase is 27 %. The results are shown in figure 5.

Table 8: Results of the mass determinations of the standards of Set 1 under vacuum, using the realization experiments at the NMIs and with respect to the IPK at the BIPM. The measurements at the BIPM were made in May 2016.

		Calibration under vacuum using real. exp. at the NMIs			**Calibration under vacuum at the BIPM, traceable to IPK**	
Institute	**Identification of standard**	**Deviation from 1 kg / mg**	**Std. unc. / mg**	**Date**	**Deviation from 1 kg / mg**	**Std. unc. / mg**
LNE	n° 13	-0.320 (in air)	0.140	Jan – Mar 2016	-0.1157 (in air)	0.004
NIST	K104	0.4081	0.0359	Feb – Mar 2016	0.3872	0.005
	141714	0.4366	0.0279	Jan – Feb 2016	0.3995	0.005
NMIJ	n° 94	0.3266	0.0238	Mar 2016	0.3286	0.005
	E59	4.9150	0.0238	Mar 2016	4.9164	0.005
NRC	K50	-0.086	0.015	Feb 2016	-0.0839	0.005
PTB	Pt109	0.150	0.019	Feb – Mar 2016	0.1520	0.005
	Si14-02	-4.223	0.019	Feb – Mar 2016	-4.2118	0.005

Table 9: Mass differences $\Delta m_{i,j}$ of the transfer standards, between calibration at the NMI and at the BIPM. The combined uncertainty $u_{i,j}$ includes the transport and air-vacuum transfer contributions. The rightmost columns show the difference Δm_i for each NMI, averaged over its travelling standards, and the associated uncertainty u_i.

Institute	Identification of standard	$\Delta m_{i,j}$ / mg	$u_{i,j}$ / mg	Δm_i/ mg	u_i / mg
LNE	n° 13	-0.2043	0.14	-0.2043	0.14
NIST	K104	0.0209	0.0362	0.0290	0.0292
	141714	0.0371	0.0354		
NMIJ	n° 94	-0.0020	0.0242	-0.0017	0.0240
	E59	-0.0014	0.0242		
NRC	K50	-0.0021	0.0157	-0.0021	0.0157
PTB	Pt109	-0.0020	0.0195	-0.0066	0.0194
	Si14-02	-0.0112	0.0201		

In the two rightmost columns of the table, for NMIs which sent two travelling standards, the mean of the two results is shown:

$$\Delta m_i = \frac{1}{2}(\Delta m_{i,1} + \Delta m_{i,2}) \tag{4}$$

The uncertainty contributions provided by NMIJ and PTB for both standards have been assumed as being completely correlated because they are dominated by the realization experiment. The NIST has provided a detailed uncertainty budget which identifies the correlation coefficient, 42.9 %, between the uncertainties for their standards. The transport and air-vacuum contributions have been assumed as uncorrelated. The combined uncertainty u_i of the mean values was calculated based on these assumptions:

$$u_i^2 = u(\Delta m_i)^2 = u_{i,\mathrm{corr}}^2 + \frac{1}{4}(u_{i,1,\mathrm{uncorr}}^2 + u_{i,2,\mathrm{uncorr}}^2) \tag{5}$$

The averaged results for all NMIs are shown in figure 6.

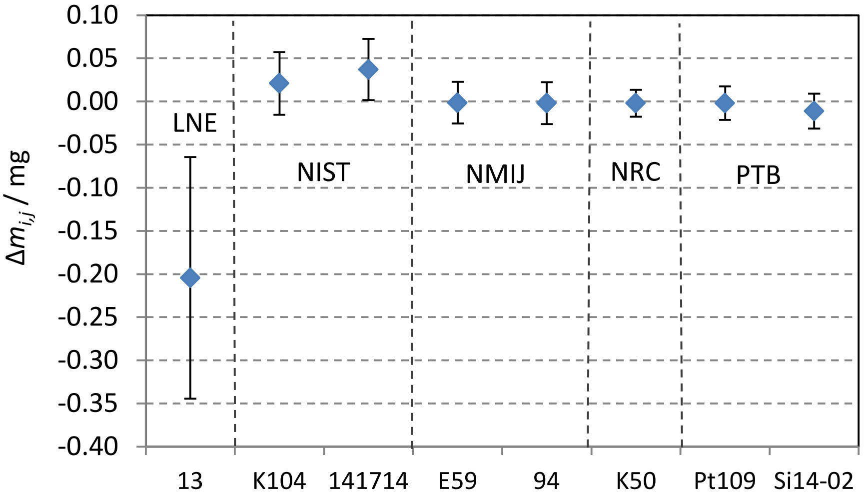


Fig. 5: Differences $\Delta m_{i,j}$ between the calibrations of each travelling standard using its NMI's realization experiment and at the BIPM and associated combined standard uncertainties.

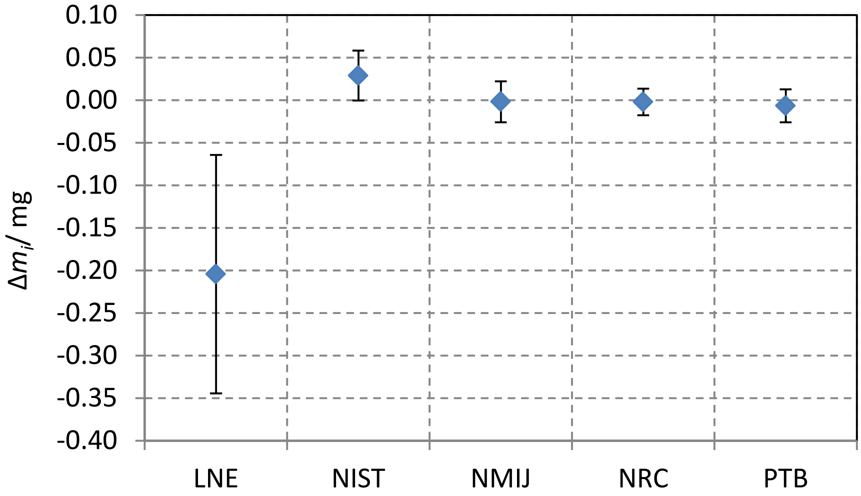


Fig. 6: Differences Δm_i between calibrations made with the NMIs' realization experiment and at the BIPM and associated combined standard uncertainties. For NMIs which sent two standards the average result is shown.

The next step is the calculation of the comparison reference value in the form of a weighted mean of the results Δm_i of the five NMIs. It was calculated in the usual way, using the inverse squared combined uncertainties as weights:

$$\overline{\Delta m} = \sum_{i=1}^{5} w_i \, \Delta m_i \, ; \quad w_i = \frac{u_i^{-2}}{\sum_j u_j^{-2}} \tag{6}$$

$$u(\overline{\Delta m}) = \sqrt{\frac{1}{\sum u_j^{-2}}} \tag{7}$$

It follows from the data of table 9 that $\overline{\Delta m}$ = -0.0006 mg with a standard uncertainty of 0.0102 mg. This means that the weighted mean of the calibration results for 1 kg standards of the participants is only 0.0006 mg different from a calibration based on the IPK.

This result was calculated without taking into account the correlation between the uncertainties of the PTB and the NMIJ, the correlation coefficient of which has been estimated by the PTB as 13 %. A generalized least-squares adjustment including this correlation leads to the very similar reference value of -0.0004 mg. For the purposes of this study, the results of all NMIs can be assumed as uncorrelated.

The differences of the participants' results from the reference value are calculated as

$$\Delta m_i' = \Delta m_i - \overline{\Delta m} \tag{8}$$

$$u(\Delta m_i') = \sqrt{u_i^2 - u(\overline{\Delta m})^2} \tag{9}$$

The minus sign is a consequence of the correlation between the participant's uncertainty and that of the weighted mean. We also calculate the difference between the result based on the present mass unit and the reference value. Since the BIPM results have not been used for the calculation of the reference value, in this case

$$u(\Delta m_{\mathrm{BIPM}}') = \sqrt{u_{BIPM}^2 + u(\overline{\Delta m})^2} \tag{10}$$

The BIPM calibration uncertainty in terms of the IPK, u_{BIPM}, is estimated as 0.005 mg. The differences from the reference value and the related standard uncertainties are shown in table 10 and on figure 7.

This comparison has the status of a Pilot Study. Had it been a key comparison, the report would have to state the degrees of equivalence, which are expressed by two quantities: the deviation from the key comparison reference value and the expanded uncertainty of this deviation (for 95 % confidence, often replaced by k = 2). Since the Pilot Study can be seen as the starting point for a future series of key comparisons, the expanded uncertainty is also shown in table 10.

Table 10: Deviations of the NMIs' results from the reference value and related standard uncertainties. The difference between the BIPM calibration based on the IPK and the reference value is also indicated.

Institute	**Deviation from reference value** $\Delta m_i'$ **/ mg**	**Standard uncertainty of deviation** $u(\Delta m_i')$ **/ mg**	**Expanded uncertainty of deviation (*k* = 2)** $U(\Delta m_i')$ **/ mg**
BIPM (IPK)	0.0006	0.0113	0.0226
LNE	-0.2038	0.1396	0.2792
NIST	0.0296	0.0274	0.0548
NMIJ	-0.0012	0.0218	0.0436
NRC	-0.0015	0.0119	0.0238
PTB	-0.0061	0.0165	0.0330

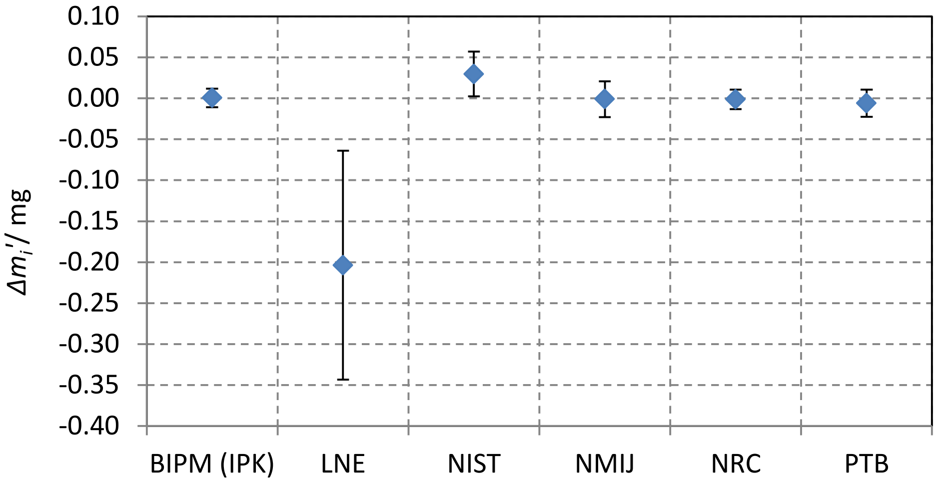


Fig 7: Deviations of the NMIs' results from the reference value and related standard uncertainties. The difference between the BIPM calibration based on the IPK and the reference value is also indicated. The standard uncertainty of the reference value is 0.010 mg.

To check the consistency of the data the Birge ratio R_{B} can be calculated according to

$$R_{\mathrm{B}}^2 = \frac{1}{n-1}\sum\frac{(\Delta m_i - \overline{\Delta m})^2}{u_i^2} \quad (11)$$

with $n = 5$. The results from table 10 lead to $R_B = 0.90$. A Birge ratio below one indicates consistent data for the realization part of the Pilot Study. This is confirmed by the fact that for four participants the uncertainty bars overlap, which is also the case for the LNE at the level of $k = 2$.

It is interesting to compare the comparison results in terms of the Planck constant (figures 1 and 2) with those in terms of mass (figure 7), by taking into account that a higher value for the Planck constant corresponds to a smaller value for mass. The dominating uncertainty contributions are in both cases those related to the realization experiments and they affect both results in the same way. The comparison in terms of the Planck constant should therefore show a high degree of consistency with that in terms of mass. Any difference can only be due to:

- the behavior of the travelling standards (which only affects the comparison in terms of mass),
- the comparison of the travelling standards at the BIPM (which only affects the comparison in terms of mass),
- differences in the maintained mass unit between the NMIs and the BIPM (which only affects the comparison in terms of the Planck constant).

Comparison of figures 1, 2 and 7 shows that the results for the Planck constant and the mass calibration results are indeed consistent at the level of a few micrograms, that is several parts in 10^9. This confirms that the outcome of the Pilot Study is not dominated by the behavior of the travelling standards or the comparison measurements at the BIPM, but that the observed mass differences reflect differences between the realization experiments.

9 Results for the travelling standards of Set 2

The comparison procedure for the stainless steel standards of Set 2 was simplified with respect to that for Set 1, because it did not require calibrations in terms of the IPK by the NMIs and the BIPM. The stability of the travelling standards could therefore only be deduced from the measurements made by the NMIs before and after sending the standards to the BIPM, which is a slight disadvantage because it includes potential mass changes during the return which in principle do not affect the prior comparison measurements done at the BIPM. This might be optimized if such a comparison would be repeated in the future.

Table 11 shows the mass changes reported by the participants and figure 8 shows them in graphical form. In some cases the mass changes and the related uncertainty could not be deduced from the measurement reports. In these cases the BIPM contacted the NMI and asked it to provide this information.

The LNE compared the standards E and INM with a stainless steel standard (MET7) before and after the measurements at the BIPM. The uncertainty of these mass determinations was 0.015 mg. The observed mass changes were -0.009 mg and -0.054 mg, respectively. The mass loss of standard INM of 0.054 mg appears to be anomalous. After discussion with the LNE, the origin remains unclear. We have noticed that the density determination in a hydrostatic balance was done just one week before the initial mass determination. It hypothesized that the mass might still have been unstable after the immersion of the standard in water. The result of the final mass determination of INM is much more consistent with the weighing results of the BIPM for the second LNE standard, E, than the initial mass. The participants agreed to use in the case of standard INM the LNE calibration result from the initial measurement report, but corrected for a mass change of -0.054 mg.

The NMIJ had compared the standards S1_2 and S2_1 with a stainless steel standard (S1_1) before and after the measurements at the BIPM. The uncertainty of the individual weighings was 0.019 mg, but due to the correlations, the uncertainty of the mass changes was only 0.0024 mg. S1_2 has gained 0.0001 mg. No change was detectable on S2_1.

At the NIST the two stainless steel standards Zwiebel 7 and Zwiebel 8 were determined against the Pt-Ir standard K79, before and after the measurements at the BIPM. The observed mass changes were -0.0228 mg for Zwiebel 7 and -0.0093 mg for Zwiebel 8. The uncertainty is estimated as 0.004 mg.

The NRC reported mass changes of -0.001 mg for standard H2A2 and of +0.003 mg for H2A3, with uncertainties of 0.006 mg.

The PTB reported mass changes of -0.009 mg and -0.0086 mg for the standards D1 and D2. The uncertainty of both changes is 0.004 mg.

Table 11: Mass changes of the stainless steel standards determined by the NMIs: mass after the comparison minus mass before the comparison.

Institute	Identification of standard	Mass change / mg	Uncertainty of mass change / mg
LNE	E	-0.009	0.005
	INM	-0.054	0.005
NIST	Zwiebel 7	-0.0228	0.004
	Zwiebel 8	-0.0093	0.004
NMIJ	S1_2	+0.0001	0.0024
	S2_1	0.0000	0.0024
NRC	HSA2	-0.001	0.006
	HSA3	+0.003	0.006
PTB	D1	-0.0090	0.004
	D2	-0.0086	0.004

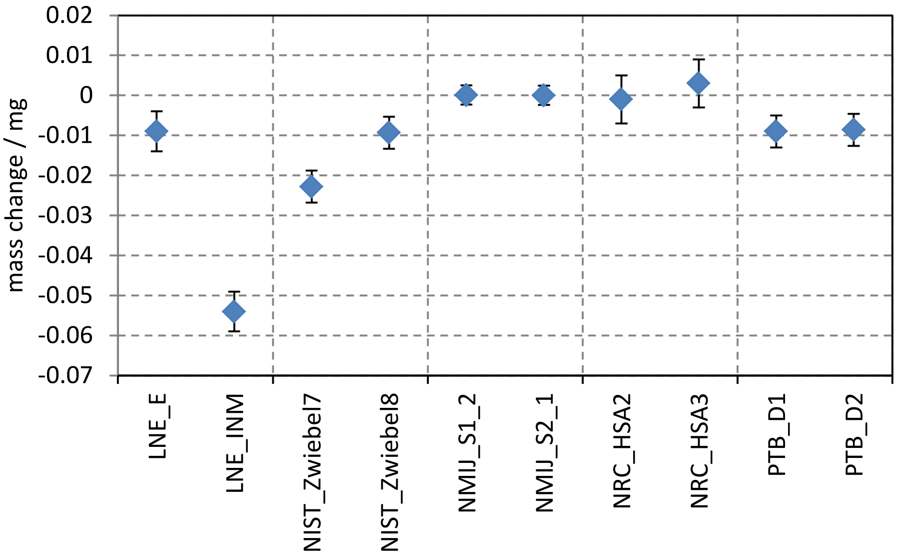


Fig. 8: Mass changes of the stainless steel standards determined by the NMIs: mass after the comparison minus mass before the comparison.

With the exception of the standards of the LNE, the changes seem to be correlated with the NMI to which the standards belong: the two standards of the NMIJ and of the NRC did not change significantly and the two standards of the PTB changed by about -0.009 mg. The PTB did not indicate any significant changes on the visual inspection forms. The standards of the NIST changed by -0.009 mg and -0.023 mg; we did not receive visual inspection forms from NIST. For the moment, the reason for this observation is unexplained. It is worth noting that for the two standards C1 and C2 of the BIPM, which served as references in the comparison and which were used more often than those of the participants, a mass increase of 0.002 mg was observed, as reported in section 6.

Similar to the standards of Set 1, an estimate for the transport uncertainty was derived from these results. Since the observed changes are the cumulative effect of the trip to the BIPM and back to the NMI, we take them as an upper limit and derive the transport uncertainty by dividing their absolute value by $\sqrt{3}$. The uncertainty for the LNE standard INM was calculated in the same way.

As shown in table 4, the standards of Set 2 were compared with the BIPM stainless steel standards C1 and C2 in air during five comparisons in the M_one mass comparator. As for Set 1, each standard was measured twice. The masses of C1 and C2 were known with respect to the IPK, which allowed the masses of the travelling standards to be expressed with respect to the IPK, too. Table 12 summarizes the results from the NMI measurement reports and from the BIPM.

Table 12: Results of the mass determinations for the travelling standards of Set 2. They were calibrated at the NMIs in air, traceable to the realization experiments, and at the BIPM with respect to the IPK. The measurements at the BIPM were made in June 2016.

		Calibration in air traceable to primary realization experiment at NMI			**Calibration in air traceable to IPK at BIPM**	
Institute	**Identification of standard**	**Deviation from 1 kg / mg**	**Std. unc. / mg**	**Date**	**Deviation from 1 kg / mg**	**Std. unc. / mg**
LNE	E	-0.080	0.140	Mar 2016	0.1300	0.008
	INM	0.656[3]	0.140	Mar 2016	0.8786	0.008
NIST	Zwiebel 7	0.0389	0.0368	Mar 2016	0.0333	0.009
	Zwiebel 8	0.2856	0.0368	Mar 2016	0.2841	0.009
NMIJ	S1_2	-1.3095	0.0255	Mar 2016	-1.3088	0.006
	S2_1	0.3928	0.0255	Mar 2016	0.3949	0.006
NRC	HSA2	-0.162	0.015	Mar – Apr 2016	-0.1522	0.006
	HSA3	-0.161	0.015	Mar –April 2016	-0.1526	0.006
PTB	D1	2.033	0.019	Mar 2016	2.0294	0.006
	D2	2.275	0.019	Mar 2016	2.2720	0.006

---

[3] The result obtained by the LNE before sending the standard to the BIPM, 0.710 mg, was corrected for the mass change of -0.054 mg, which we believe occurred before arrival at the BIPM.

The determination of the degrees of equivalence of the NMIs' calibrations for the Set 2 standards follows the same approach as applied for Set 1.

As the first step, the differences between the NMI calibrations and the BIPM calibrations, $\Delta m_{i,j}$ (equation 3), were determined, see table 13 and figure 9. The combined uncertainties $u_{i,j}$ include the transport uncertainty. This additional uncertainty component increases the NMIs' calibration uncertainties by 6 % in the worst case. Since each NMI sent two travelling standards, the results for both were averaged to obtain Δm_i (equation 4). The mean values for each NMI are shown on the right side of table 13 and on figure 10. The calibration uncertainties of the NMIs were again considered as being completely correlated between both standards, the transport uncertainties as uncorrelated (eq. 5).

Table 13: Mass difference between calibration using primary realization experiment and against the IPK. The rightmost columns show the results for each NMI, by averaging over its travelling standards.

Institute	Identification of Standard	$\Delta m_{i,j}$ / mg	$u_{i,j}$ / mg	Δm_i / mg	u_i / mg
LNE	E	-0.2100	0.140	-0.2163	0.141
	INM	-0.2226	0.143		
NIST	Zwiebel 7	0.0056	0.0391	0.0036	0.0375
	Zwiebel 8	0.0015	0.0372		
NMIJ	S1_2	-0.0007	0.0255	-0.0014	0.0255
	S2_1	-0.0021	0.0255		
NRC	HSA2	-0.0098	0.0150	-0.0091	0.0150
	HSA3	-0.0084	0.0151		
PTB	D1	0.0036	0.0197	0.0033	0.0193
	D2	0.0030	0.0196		

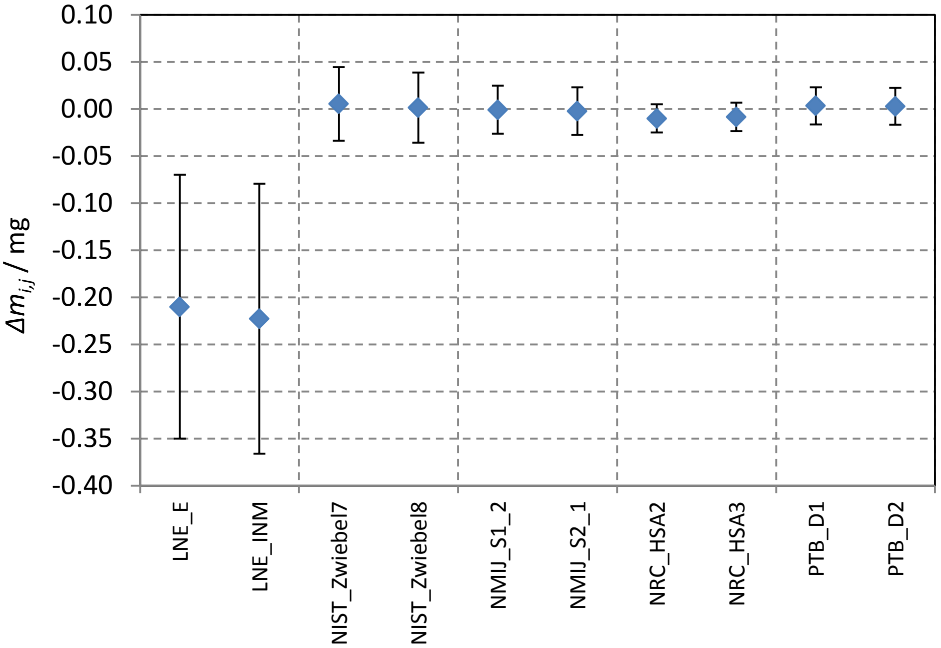


Fig. 9: Differences between the calibration of each travelling standard with its NMI's realization experiment and at the BIPM, $\Delta m_{i,j}$ (eq. 3), and associated combined standard uncertainties.

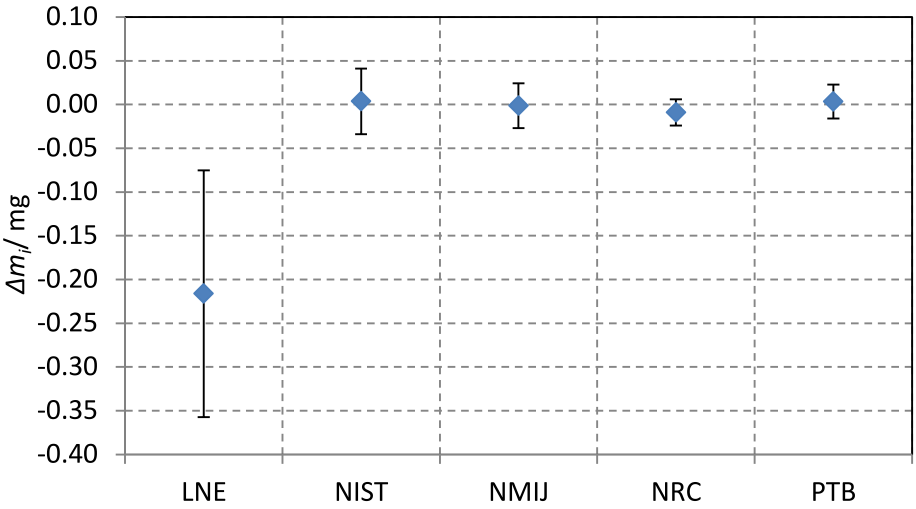


Fig. 10: Differences between the calibrations made using the NMIs primary realization experiment and at the BIPM and associated combined standard uncertainties. Shown are the mean values of the results for the two standards of each NMI from figure 9.

Following the same approach as for Set 1, the comparison reference value was calculated as the weighted mean of the five participants' results (equations 6 and 7).

The reference value is -0.0045 mg with a standard uncertainty of 0.0103 mg. The differences of the NMIs' results from the reference value are shown in table 14 and in figure 11. Also shown is the position of the BIPM calibrations based on the IPK with respect to the reference value. The Birge ratio, calculated according to equation (11) is 0.80.

Table 14: Deviations of the NMIs' results from the reference value and related standard uncertainties.

Institute	**Deviation from reference value** $\Delta m_i'$ **/ mg**	**Standard uncertainty of deviation / mg**	**Expanded uncertainty of deviation (*k* = 2) / mg**
BIPM (IPK)	0.0045	0.0131	0.0262
LNE	-0.2118	0.1405	0.2810
NIST	0.0080	0.0360	0.0720
NMIJ	0.0031	0.0233	0.0466
NRC	-0.0046	0.0109	0.0218
PTB	0.0077	0.0164	0.0328

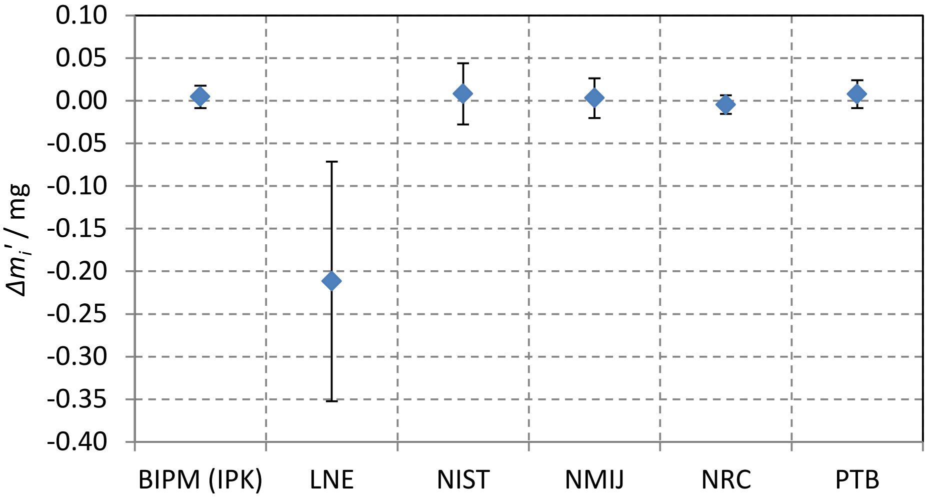


Fig. 11: Deviations of the NMIs results from the reference value and related standard uncertainties. The difference between the BIPM calibration based on the IPK and the reference value is also indicated. The standard uncertainty of the reference value is 0.010 mg.

The outcome of the comparison with the standards of Set 2 looks very similar as that obtained with Set 1 (figure 7). The main difference is that for Set 2, the NIST result is closer to the Comparison Reference Value than for Set 1. The similarity of the results for both sets demonstrates that the dissemination of the mass unit, realized under vacuum, does not present any particular problems.

10 Summary and conclusions

- This CCM Pilot Study had the objectives to (1) test the consistency of kilogram realizations based on the Kibble balances of LNE, NIST and NRC and on the Avogadro spheres from the PTB and the NMIJ, (2) to test the continuity between the future realizations and the present one and, (3) to evaluate and optimize the technical protocol for future key comparisons.
- To compare the realizations as closely as possible, the travelling standards of Set 1 were calibrated and compared under vacuum (except for the LNE).
- To compare the dissemination of the mass unit, the stainless steel travelling standards of Set 2 were calibrated and compared in air.
- The calibrations of 1 kg standards of Set 1 from the NIST, NMIJ, NRC and the PTB agree with each other within their standard uncertainties; the LNE is in agreement at the level of k = 2. The standard uncertainty of the weighted mean of the five laboratories is 0.010 mg.
- The weighted mean agrees with the calibration based on the IPK to within 0.001 mg.
- The calibrations of 1 kg standards of Set 2 from the NIST, NMIJ, NRC and the PTB agree to within their standard uncertainties, the LNE is in agreement at the level of k = 2. The standard uncertainty of the weighted mean of the five laboratories is 0.010 mg.
- The weighted mean of the calibrations of the standards of Set 2 agrees with the calibration based on the IPK to within 0.0045 mg.
- The result of the comparison of mass calibrations is highly consistent with the comparison of the determinations of the Planck constant carried out during this study.

- The results of this comparison are dominated by the performance of the primary realization experiments and for the stainless steel standards, to a lower degree, by the behavior of the travelling standards.
- Transport appeared to cause significantly higher mass changes for the stainless steel masses than the silicon and Pt -Ir masses.
- Condition 1 of the CCM Recommendation G1 (2013) [14] requests that “at least three independent experiments, including work from watt balance and XRCD experiments, yield consistent values of the Planck constant with relative standard uncertainties not larger than 5 parts in 10^8”. This condition is fulfilled by the results of NIST, NMIJ, NRC, and PTB.
- Condition 2 of the CCM Recommendation G1 (2013) [14] requests that “at least one of these results should have a relative standard uncertainty not larger than 2 parts in 10^8”. This condition is fulfilled by the results of NRC and PTB, which both have uncertainties below 2 parts in 10^8.
- Condition 4 of the CCM Recommendation G1 (2013) [14] requests that “the procedures for the future realization and dissemination of the kilogram, as described in the *mise en pratique*, have been validated in accordance with the principles of the CIPM MRA”. The Pilot Study was carried out in a form similar to a future key comparison. Calibrations of 1 kg mass standards using future realization methods agreed for four participants within the estimated standard uncertainties, and for the fifth participant within the expanded (k=2) uncertainty. Although the CCM has to make the final judgement, it can be concluded that the procedures for the future realization and dissemination of the kilogram, as described in the *mise en pratique*, have been validated in accordance with the principles of the CIPM MRA.

11 Recommendations for future key comparisons of primary mass standards

One of the uncertainty contributions in a comparison of primary mass standards is the stability of the mass standards under transportation. To allow an estimate of possible mass changes, the protocol for the Pilot Study requested that the mass standards of Set 1 had to be calibrated in air, traceable to the IPK, at the NMIs before sending them to the BIPM, at the BIPM upon receipt and before return, and again at the NMI after return. For the stainless steel standards of Set 2 the protocol asked only for the calibrations using the realization experiment at the NMIs and for the comparison at the BIPM. The uncertainties of the initial and final calibrations of the NMIs include the contribution from the realization experiment, which are correlated between both measurements. The information collected on the reporting forms does not allow the estimation of the mass changes of the stainless steel standards with an uncertainty independent of the contribution of the realization experiment. This information had to be requested by the pilot laboratory during the analysis of the data.

- It is recommended that in a future comparison, the protocol shall explicitly ask for a comparison of the stainless steel standards at the NMIs against a stable stainless steel reference standard, before and after the comparison measurements, so that the mass change during the comparison can be determined with a small uncertainty.

In the present comparison the dominant uncertainty component was related to the realization experiments, but in future comparisons the contribution related to the stability of the masses will become more important because of the improvement of the realization experiments.

- It is recommended that participants shall carry out a number of vacuum-air cycles to stabilize the travelling standards and provide an estimate for the related uncertainty.

12 References

[1] Milton M.J.T., Davis R., Fletcher N., Towards a new SI: a review of progress made since 2011, *Metrologia*, 2014, **51(3)**, R21-R30.

[2] Robinson I.A., Schlamminger S., The watt or Kibble balance: a technique for implementing the new SI definition of the unit of mass, *Metrologia*, 2016, **53(5)**, A46.

[3] Fujii K. et al., Realization of the kilogram by the XRCD method, *Metrologia*, 2016, **53(5)**, A19.

[4] Joint CCM and CCU roadmap towards the redefinition of the SI in 2018, http://www.bipm.org/utils/common/pdf/SI-roadmap.pdf.

[5] *Mise en pratique* for the definition of the kilogram, version 9.0, working document CCM/15-02A, http://www.bipm.org/cc/CCM/Allowed/15/02A_MeP_kg_141022_v-9.0_clean.pdf.

[6] Stock M., Barat P., Davis R., Picard A., Milton M., Calibration campaign against the International Prototype of the kilogram, part I: comparison of the international prototype with its official copies, *Metrologia*, 2015, **52**, 310-316.

[7] de Mirandés E., Barat P., Stock M., Milton M., Calibration campaign against the international prototype of the kilogram in anticipation of the redefinition of the kilogram, part II : evolution of the BIPM as-maintained mass unit from the 3rd periodic verification to 2014, *Metrologia*, 2016, **53**, 1204-1214.

[8] Mohr P.J., Newell D.B., Taylor B.N., CODATA Recommended Values of the Fundamental Physical Constants: 2014, *Rev. Mod. Phys*., 2016, Vol. 88, 035009-1 – 035009-73.

[9] Berry J., et al., A comparison between seven National Measurement Institutes to evaluate inert gas transfer of mass standards against current best practice, NPL Report ENG 61, December 2015. http://www.npl.co.uk/content/ConPublication/6872

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

[11] Thomas M. et al., First determination of the Planck constant using the LNE watt balance, *Metrologia*, 2015, **52**, 433-443.

[12] Haddad D. et al., A precise instrument to determine the Planck constant, and the future kilogram, *Rev. Sci. Instrum.* **87** 061301 (2016).

[13] Sanchez C. et al., Corrigendum to the 2014 NRC determination of Planck's constant, *Metrologia*, **52** (2015), L23.

[14] Recommendation G1 (2013), Report of the 14th meeting of the Consultative Committee for Mass and Related Quantities (CCM), http://www.bipm.org/utils/common/pdf/CC/CCM/CCM14.pdf .