Journal Name

Crossmark

RECEIVED
dd Month yyyy

REVISED
dd Month yyyy

RESEARCH ARTICLE

First international comparison of three different tabletop Kibble balances from 1 mg to 10 g

Johannes Konrad[1], Kumar Arumugam[2], Norbert Rogge[3], Christian Rothleitner[1], Hugo Gasca-Aragon[1], Leon Chao[2], Stephan Schlamminger[2], Suren Vasilyan[3], and Thomas Fröhlich[3]

[1] Physikalisch-Technische Bundesanstalt, 38116 Braunschweig, Germany
[2] National Institute of Standards and Technology, 20899, Gaithersburg, MD, USA
[3] Institute of Process Measurement and Sensor Technology, Technische Universität Ilmenau, 98684 Ilmenau, Germany

E-mail: johannes.konrad@ptb.de

Keywords: tabletop Kibble balance, small mass metrology, kilogram realization, mass dissemination, digital calibration certificate (DCC), digital metrological expert (DME), interlaboratory comparison

Abstract

We compared results from three different tabletop Kibble balances operated at TU Ilmenau, PTB, and NIST using two circulated mass sets. Each set contained five OIML R 111 nominal values from 1 mg to 10 g and one deliberately chosen non-OIML value of 6.626 070 15 g. The manufacturers calibrated the sets independently against OIML E_1 target specifications. Data exchange used Digital Calibration Certificates (DCCs), and all calibration and participant results were withheld until the final unblinding and evaluation.

For each mass, we calculated an inverse-variance weighted reference value from the three Kibble balance results and compared it with the manufacturer calibration value. The weighted reference values differed from the manufacturer values by at most 15 µg. For the gram-level masses, the corresponding relative deviations were below 2.3×10^{-6}. Eleven of the 12 cases satisfied $|E_n| \leq 1$. The remaining case had $E_n = 1.3$. The individual Kibble balance results were less uniform: reported expanded uncertainties ranged from about 0.4 µg to 160 µg and differed by up to approximately a factor of 230 for the same nominal mass. These results indicate compatibility at the level of the weighted reference values, while the individual uncertainty levels and measurement performance are not uniform across implementations.

1 Introduction and Motivation

The redefinition of the kilogram within the revised International System of Units (SI), which took effect on May 20, 2019, marked a paradigm shift for the global mass metrology community. Previously, the SI unit of mass was bound to the embodiment of a single material object, the International Prototype of the Kilogram (IPK), for which definition and realization were identical and conceptually simple. As a direct consequence, the dissemination of the unit followed a rigid hierarchical traceability chain of mass standards based on comparison weighings, with gradually decreasing accuracy. Starting from a 1 kg reference, multiples and submultiples of the unit, the so-called mass scale, were established as further physical standards using factors of $1 \cdot 10^n$, $2 \cdot 10^n$, and $5 \cdot 10^n$, with positive and negative integer values of n, ensuring traceability for masses of any value [1]. Since the redefinition, the unit of mass is expressed in terms of the Planck constant h, and realization is no longer identical to the definition. Although this approach is more conceptually abstract, it enables greater flexibility and independence. The *mise en pratique* [2], published as an appendix to the SI Brochure, describes how the unit of mass can be realized experimentally using primary methods such as the Kibble balance or the XRCD. Currently, most approaches focus on realization at 1 kg, and dissemination at the national level remains anchored at this value, as reflected in most publications [3, 4]. At present, dissemination is based on a consensus value at 1 kg derived from international comparisons of realization experiments [5].

However, the SI Brochure states that, in principle, primary realizations can be established at any point of the mass scale [6]. This suggests that mass calibration could be performed directly at arbitrary values without relying on a mass scale and the associated hierarchical standards. This could be particularly beneficial for masses much smaller than 1 kg [3]. In this range, the high surface-to-volume ratio leads to reduced relative stability of the standards. In addition, subdivision becomes increasingly demanding, and the permitted relative uncertainties in established calibration frameworks increase [7]. Interestingly, it has been argued in the literature [8], that mass measurements at any value could have been made traceable to CODATA-recommended values of h long before the redefinition, reaching relative uncertainties below 1×10^{-6}.

Present-generation Kibble balances, mainly used for realizations at 1 kg with relative uncertainties on the order of 1×10^{-8} [9], are complex and require significant resources. Accordingly, dissemination of the unit is typically performed via established subdivision methods. Kibble balances designed as tabletop systems, with relative uncertainties on the order of 1×10^{-6} in the sub-gram to gram range, represent an alternative approach. Performance levels compatible with requirements for mass standards, for example as specified by the International Organization of Legal Metrology (OIML) [7], could be targeted. The equivalence of these results could be demonstrated through comparisons of submultiples of the kilogram.

Recent developments indicate growing activity in tabletop Kibble balances. Beyond the systems in this comparison, further compact implementations targeting the gram range are under development [10,11]. They could complement subdivision, extend traceability, and make use of the additional degrees of freedom introduced by the redefinition. In this context, an international comparison of realizations of the unit of mass at submultiples of the kilogram has been initiated. Researchers at the National Institute of Standards and Technology (NIST, USA), Technische Universität Ilmenau (TUI, Germany), and Physikalisch-Technische Bundesanstalt (PTB, Germany) have each developed such instruments and participated in this comparison. The comparison marks a transition from separate single-instrument demonstrations toward a shared experimental evaluation of tabletop Kibble balance implementations, using common transfer standards and independent manufacturer calibration values. This represents an initial effort of its kind and an important step towards extending realization-based mass metrology beyond the kilogram level.

An analogy can be drawn to the introduction of conventional cesium beam clocks, which several decades ago marked the transition from large laboratory-based atomic frequency standards to compact tabletop instruments, thereby enabling broader and routine application of primary time standard technology [12].

2 Comparison design

2.1 Conduct of the comparison

Two independently manufactured mass sets, serving as transfer standards, were circulated between participating laboratories. In this work, a participant refers to an institute operating a tabletop Kibble balance and reporting final measurement results. The two mass sets were manufactured by Häfner Gewichte GmbH (Germany) and Rice Lake Weighing Systems (USA)[1], and are referred to as HAF and RL, respectively.

The study was conducted as an informal interlaboratory comparison. The circulation scheme, measurement campaign, and evaluation approach were agreed among the participants. It was not organized as a key comparison under the CIPM Mutual Recognition Arrangement (MRA).

The comparison was initially planned to involve four participants: PTB, TUI, NIST, and the United States Army Test, Measurement, and Diagnostic Equipment Activity – The Army Primary Standards Laboratory (APSL) at Redstone Arsenal (USA). Measurements started in January 2025 at the APSL site but were not further continued there due to technical difficulties. No complete result was obtained,

[1] Certain equipment, instruments, software, or materials are identified in this paper in order to specify the experimental procedure adequately. Such identification is not intended to imply recommendation or endorsement of any product or service by the National Institute of Standards and Technology (NIST), nor is it intended to imply that the materials or equipment identified are necessarily the best available for the purpose.

and APSL is therefore not included in the final evaluation.

Each participant included in the evaluation reported one estimate of the absolute mass value for each individual mass in each mass set, together with an associated expanded uncertainty ($k = 2$), based on internal measurement data and individual uncertainty budgets. In cases where a mass set was measured more than once, the most recent result was used in the evaluation. Repeated measurements were not analyzed for temporal stability in this study.

During the comparison, each mass set crossed the Atlantic Ocean twice. It was transported either by commercial courier or hand-carried by personnel during scheduled travel. The measurement schedule is summarized in Table 1.

Table 1. Schedule of the comparison.

Month (2025)	HAF	RL
January	—	APSL
April	TUI	—
May	PTB	NIST
Masses cross the Atlantic Ocean		
July	—	TUI
August	NIST	—
October	—	PTB
Masses cross the Atlantic Ocean		
October	PTB	—
November	TUI	NIST

2.2 Mass sets and material properties

Photographs of the two mass sets are shown in Figures 1 and 2. Each of them comprised six masses with nominal values of 1 mg, 50 mg, 200 mg, 1 g, 6.626 070 15 g, and 10 g. Five of these values correspond to nominal values defined in OIML R 111, following the 1–2–5 scheme, while the value 6.626 070 15 g does not belong to this scheme and is therefore referred to as a non-OIML nominal value. The HAF set was commissioned by TUI, and the RL set was commissioned by NIST. In both cases, the manufacturers were instructed to manufacture and calibrate the mass sets targeting OIML E_1 specifications [7], while the calibration values were not disclosed to the participants during the measurement campaign. However, density values were extracted from the manufacturers' calibration data and provided separately to enable consistent air buoyancy corrections.

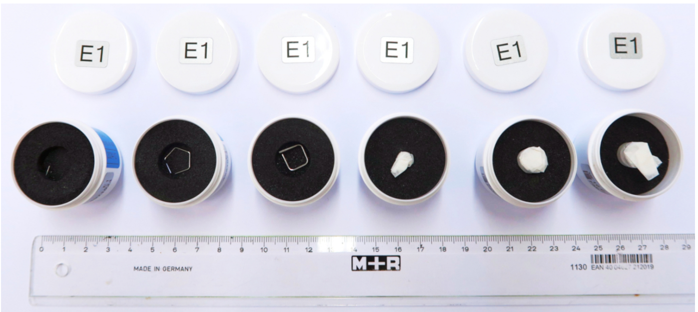

Figure 1. Photograph of the HAF mass set. Some of the weights are wrapped in transport paper. The Planck-mass (6.626 070 15 g) is second from the right.

The value 6.626 070 15 g (for brevity, hereafter denoted as 6.626 g, "Planck-mass") was deliberately included to reduce expectation bias associated with OIML nominal values and tolerance limits. It should not be confused with the fundamental Planck mass $\sqrt{\hbar G/c}$. This result-blind approach, in which nominal values were known while calibration values and inter-participant results were withheld until final evaluation, reduces the risk of unconscious adjustments toward expected results ("intellectual phase locking") [13, 14].

The HAF set consists of wire weights in the milligram range and cylindrical knob weights in the gram range, except for the Planck-mass, which is a cylindrical weight. All masses were manufactured

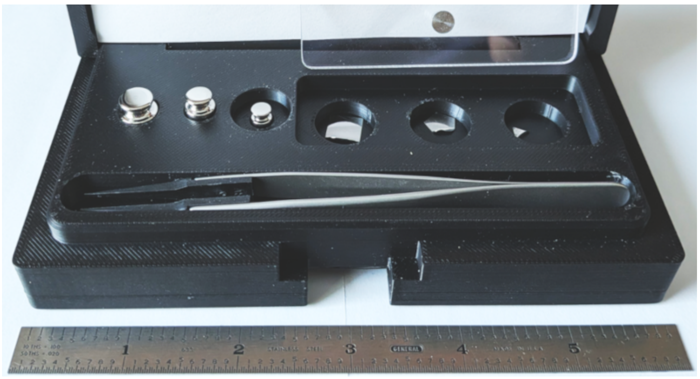

Figure 2. Photograph of the RL mass set. The Planck-mass (6.626 070 15 g) is second from the left.

from stainless steel alloy (HE210) and were stored individually in plastic screw containers. The RL set consists of sheet weights in the milligram range and cylindrical knob weights in the gram range. All masses are made of stainless steel except for the 1 mg mass, which is made of aluminum. The masses were stored together in a custom 3D-printed transport case.

Density values were used consistently by all participants and the manufacturers for air buoyancy corrections, ensuring identical input conditions for the comparison. Independent density determinations would in principle be possible but were outside the scope of this study. The density values for both sets are listed in Table 2.

Table 2. Manufacturer-supplied density values expressed in $\mathrm{g\,cm^{-3}}$, with expanded uncertainties ($k = 2$).

Nominal mass	HAF	RL
1 mg	8.00(6)	2.7(5)
50 mg	8.00(6)	7.95(5)
200 mg	8.00(6)	7.95(5)
1 g	8.007(33)	7.97(29)
6.626 g	8.0064(87)	7.97(5)
10 g	8.0085(70)	7.980(29)

3 Measurement apparatus and procedure

In general, a Kibble balance is an instrument that compares virtual mechanical power with virtual electrical power. A central idea proposed by Bryan Kibble in 1976 [15] was the addition of a velocity mode to the ampere balance, enabling the elimination of the geometric factor. This concept led to the development of the watt balance, later renamed in his honor. A detailed description of the underlying principles and historical context can be found in [16]. In weighing mode, the difference between the coil currents required to maintain equilibrium in the mass-on (load applied) and mass-off (load removed) states is denoted by I and used to determine the test mass as

$$m = Bl \cdot \frac{I}{g}, \tag{1}$$

where I is obtained from the voltage V across a precision resistor R using Ohm's law, B is the magnetic flux density, l is the effective coil wire length, and g is the local acceleration of free fall. The relation above represents the fundamental Kibble balance equation. Condition-dependent corrections may apply, e.g., for air buoyancy. The product Bl, also called geometric factor, is determined in velocity mode, in which a voltage U is induced in the coil moving with velocity v through the magnetic field, such that

$$Bl = \frac{U}{v}. \tag{2}$$

The instruments participating in this comparison are referred to as tabletop implementations of Kibble balances. While they operate according to the same fundamental principle, this term is used here to denote a class of systems aimed at measuring masses well below the kilogram level over several orders of magnitude, with reduced size and cost, simpler operation, and typically higher, yet practically relevant, uncertainties. In this sense, they are intended to function as laboratory instruments comparable to analytical balances.

Several characteristics are common to the tabletop Kibble balances used in this comparison:

- a single degree-of-freedom flexure mechanism for moving the coil, based on the so-called Kibble–Robinson theory (KRT) [17–19];
- a Keysight 3458A digital voltmeter;
- a commercial displacement measuring interferometer with vertically aligned laser beam (parallel to g), referenced by a liquid-pool surface;
- oven-controlled crystal oscillators (OCXOs) as frequency sources, referenced to external standards such as GPS or hydrogen masers;
- determination of the local acceleration of free fall using relative gravimeters referenced to nearby absolute measurements;
- magnet systems based on Sm_2Co_{17} permanent magnets;
- operation in air;
- use of oscillatory coil motion in velocity mode.

The determination of the test mass value is based on repeated force mode and velocity mode measurements. In force mode, automated load handling is used, with the test mass repeatedly applied to and removed from the weighing pan. Velocity mode measurements are performed before and after force mode to determine Bl, which is interpolated to the time of the force mode measurements. Key measurement parameters are summarized in Table 3. The three instruments are described in the following subsections, where system characteristics are outlined to provide technical context for the comparison and to illustrate the present status of the tabletop Kibble balance systems represented in this study. Each system is shown in a photograph, with an apple included for scale.

Table 3. Typical measurement parameters for the determination of a 1 g test mass. The table includes both instrument settings and measurement sequence parameters used during the comparison. R denotes the measurement resistance value in force mode.

Parameter	NIST	PTB	TUI
Weighings per force mode	5	4	10
Number of measurement sets in force mode	5	80	30
Total measurement duration / h	2	6	16
Bl / T m	138	463	180
Coil oscillation amplitude / mm	2.5	0.175	0.02
Coil oscillation frequency / Hz	0.5	1	4 to 8
Coil wire diameter / µm	80	71	50
max(U) / V	1.1	0.5	0.2
max(v) / mm s^{-1}	8	1.1	1.0
R / kΩ	10	10	1

3.1 KIBB-g2 at NIST

The development of tabletop Kibble balances at NIST began with the first-generation device, KIBB-g1, which demonstrated measurement uncertainties at a level comparable to those of accuracy class E_2 when weighing copper masses with nominal values of 1 g and 5 g [20].

Figure 3 shows the second-generation device KIBB-g2 used in this comparison study. The overhung voice coil attached to the platforms follows an open magnet design similar to commercial actuators. The current balance prototype includes an additional mass-handling device on the right side, compared to the previous prototype presented in [21], resulting in two mass pans mounted on opposing flexure platforms.

The test mass is placed on the left pan, and for gram-level measurements, a tare mass of approximately half the test mass is placed on the right pan to mitigate the influence of the coil-current effect [22]. Since the relative influence of this effect decreases with decreasing coil current, no tare mass was used for the milligram-level measurements in the present study.

The interferometer (Attocube IDS3010) is used to measure both displacement during velocity mode

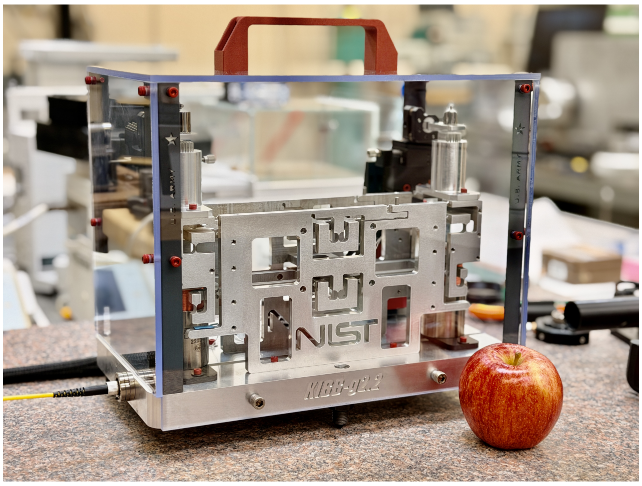

Figure 3. Mechanics of KIBB-g2 housed inside an antistatic draft shield. The test mass is on the left side and a tare mass on the right side.

and as a feedback sensor to maintain equilibrium position of the balance during weighing. As suggested by the Kibble–Robinson theory, its measurement point is located as close as possible to the center of mass of the test mass to reduce sensitivity to eccentric loading errors [18] arising due to the 1 µrad mm^{-1} parasitic rotation of the mass pan during vertical motion.

To characterize the Bl of the weighing actuator on the left side, the flexure is oscillated in open-loop using an auxiliary actuator on the right side. The voltage samples U are integrated over ten power line cycles, $\Delta t = 166.666$ ms, and are sampled at a 2.82 Hz rate. To determine velocity v corresponding to U, a square wave signal output from the voltmeter indicating the start and end of each integration period is used to trigger two interferometer samples to determine the displacement Δz during Δt.

Data acquisition is implemented on a USB-based LabVIEW system using an NI USB-6351 (National Instruments), with data post-processing performed in Python. A least-squares method is used to determine Bl coefficients using a polynomial fit to the voltage samples up to sixth order. The sinc correction is applied to the voltage and velocity signals measured during the determination of Bl and depends on the oscillation frequency f and the integration time Δt used in the velocity mode [23].

The measured values from the voltmeter operating in the 10 V range were corrected using a Zener reference (Fluke 732C) before the mass measurements. The Zener reference was calibrated using a Josephson voltage standard. The 10 kΩ resistor (Fluke 742A) was calibrated against a reference resistor (traceable to a quantum Hall resistor) using a current comparator. The Type A standard uncertainty of the measured mass is calculated as $u = s/\sqrt{n}$ with $n = 4$, corresponding to the minimum of the Allan deviation, and s is the standard deviation of 25 mass determinations (five weighings, each performed in a force mode cycle), see Table 3. For the 1 g mass of the HAF set, the largest contributions to the relative standard uncertainty arise from the Type A component of 3.4×10^{-6} in the force mode, and the 1.5×10^{-6} uncertainty associated with the sinc correction.

3.2 Planck-Balance at PTB

The Planck-Balance concept was jointly initiated by TUI and PTB within a publicly funded research project from 2017 to 2019 [24]. The implementation used and developed at PTB, sometimes referred to as PB1, is based on a modified commercial weighing cell (model WZA26-HC, Sartorius AG). A dedicated asymmetric magnet system with a non-cylindrical inner yoke contour was developed in-house [25]. Measurements are carried out in a former balance testing laboratory without air conditioning, on a weighing stone. The instrument is shown in Figure 4.

Data acquisition and control are implemented on a PXIe-based system using LabVIEW, with an NI PXIe-8861 controller in an NI PXIe-1092 chassis. Data post-processing is performed in MATLAB. Measurements are performed with the voltmeter operated exclusively in the 10 V range. The magnet temperature is continuously monitored and corresponding corrections to Bl are applied during measurements of both modes.

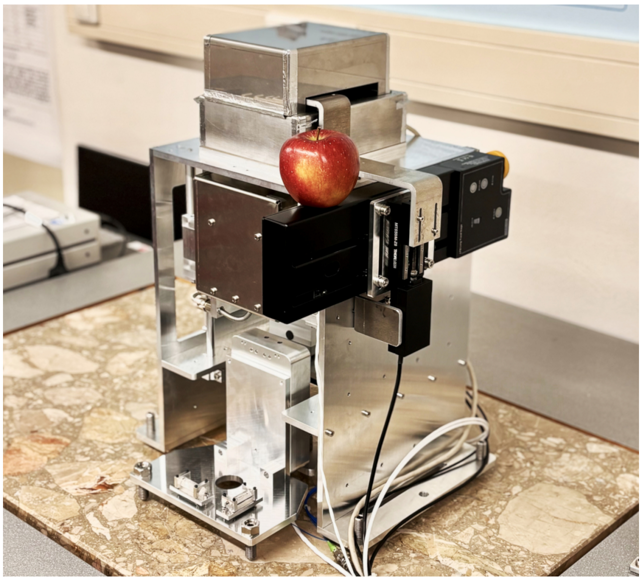

Figure 4. Photograph of the mechanical part of the Planck-Balance at PTB. External instrumentation, electronics, and control systems are not shown. The test mass is located beneath the upper metal hood.

In velocity mode, the coil is driven by a sinusoidal current generated by a function generator. Voltage and displacement signals are sampled simultaneously at 1 kHz, recording complete oscillation periods. The geometric factor is obtained from the Fourier components at the excitation frequency following [26]. The displacement amplitude z_0 is converted to velocity amplitude using $v_0 = 2\pi f z_0$, where f is the oscillation frequency. The ratio of the induced voltage amplitude to v_0 yields Bl. Additional corrections are applied for measured frequency dependence, amplitude distortion, and sinc-filter bias. A commercial interferometer (SIOS, model SP 2000 DI) is used. In contrast to the previous publication, the analog interferometer signals are analyzed rather than the digital data, enabling a global Heydemann correction.

In force mode, feedback uses the original optoelectronic position sensor and associated electronics supplied with the commercial weighing cell, with a feedback loop frequency of about 500 Hz. The voltage across the measurement resistor is measured using an integration time of 0.2 s. The current used in force mode is supplied by an in-house built summing amplifier, rather than the commercial power source described in the latest publication [27], to reduce instrument cost while maintaining sufficient performance for the present setup.

Mass placement and lifting involve a vertical motion of approximately 1.7 mm. For masses larger than 500 mg, symmetric weighing is applied using tare masses, eliminating the need for an additional coil-current correction [22]. For smaller masses, symmetric weighing is not applied and the corresponding coil-current correction is evaluated instead. Potential magnetic interference effects were evaluated by Gauss probe measurements of the vertical field component B_z and its gradient at the weighing pan. The gradient was estimated as $-0.3\,\mathrm{mT/m}$ and B_z as approximately $-5\,\mu\mathrm{T}$.

Traceability of the electrical quantities is established using a Fluke 732C reference for voltage and a commercial Vishay H-series resistor for resistance. Both references are periodically calibrated, typically at intervals of about one year, with a documented calibration history. The voltage reference is calibrated against a Josephson voltage standard, while the resistance value of the Vishay H-series resistor is determined using a cryogenic current comparator referenced to a conventional standard resistor. The Vishay H-series resistor is maintained in an air bath at 23 °C.

In force mode, an offset-force correction of approximately 0.3 µg is applied. It was determined from force mode measurements without an applied test mass and treated as a constant bias during the comparison. Type A standard uncertainties of the measured mass are calculated as the experimental standard deviation of the mean. No significant correlations were observed between successive measurements. For the 1 g mass of the HAF set, the largest contributions to the relative standard uncertainty are the Type A component at about 0.4×10^{-6}, the determination of Bl, eccentric loading, and buoyancy at about 0.3×10^{-6} each, and the force mode voltage and offset-force correction at about 0.2×10^{-6} each. An evaluation of the full uncertainty budget and the most recent measurement results is provided in [27]. At present, the achieved equivalent calibration capability is comparable to that of OIML accuracy class E_1 up to 5 g, E_2 up to 20 g, and F_1 up to 100 g.

3.3 Planck-Balance at TUI

The Planck-Balance at TUI (PB2), as described in [28], was operated in the configuration therein for this comparison. Figure 5 presents the system, which is based on the prototype structure in [29] and uses a commercial weighing cell (model WZA215, Sartorius AG). A compact closed-circuit magnet–coil assembly is integrated below the weighing pan [30], with the closed magnetic path reducing susceptibility to external magnetic fields. The system has a modular mechanical design, allowing replacement of key components such as the load carrier, coil, magnet system, and interferometer.

The balance is operated in a closed-loop control system implemented on a commercial digital signal processor with a loop frequency of 10 kHz. In force mode, feedback is provided by the optoelectronic position sensor integrated in the commercial weighing cell. In velocity mode, the voltage measurement and the data acquisition of the interferometer (SIOS SP 5000 DI) were both sampled at 1 kHz and synchronized using a trigger signal generated by a Keysight 33521B waveform generator.

The PB2 system was operated on a weighing stone in a basement lab with typical temperature stability of less than 200 mK per day but without active air conditioning.

A compact load lifter is integrated into the structure as a modular unit mounted on the housing using a dedicated clamping mechanism. The final mass determination is performed during post-processing.

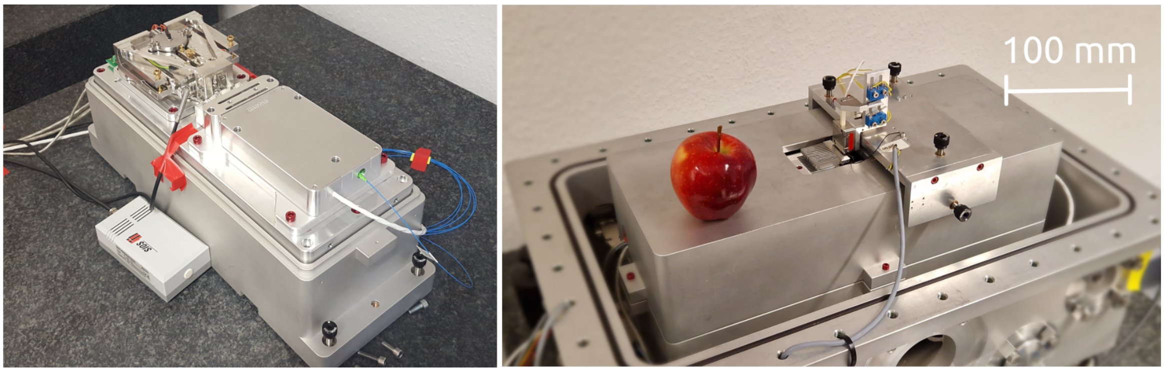

Figure 5. PB2 at TU Ilmenau showing the core mechanical assembly without auxiliary subsystems or external instrumentation. The base structure (left) illustrates the commercial weighing cell and interferometer, while the load lifter is positioned on the housing of the system (right).

The velocity mode data is evaluated using a sinusoidal fitting algorithm [31]. In general, frequencies between 0.5 Hz and 10 Hz are used to characterize dynamic effects and validate Bl. To reduce measurement time in the comparison reported here, the velocity mode was operated at frequencies between 4 Hz and 8 Hz. The corresponding individual Bl values obtained at these frequencies were corrected for frequency dependent effects and their mean value was used to determine the final geometric factor. The small oscillation amplitude is intended to reduce the influence of mechanical and magnetic nonlinearities and support stable long-term operation. Further details on the velocity mode excitation and harmonic contributions to the induced voltage are given in [32].

The system operates without mechanical tare elements. A DC offset current in the velocity mode excitation coil counterbalances the mass pan and other coupling elements, and shifts the operating point such that the electric currents in force mode for the mass-on and mass-off states are equal in magnitude and opposite in sign. In velocity mode, the same principle shifts the oscillation center to coincide with the desired weighing position used in force mode. The measurement infrastructure follows the architecture described in [28]. It uses a high-resolution DAQ system together with precision voltage-to-current converters and control electronics, enabling flexible implementation of control algorithms and signal processing.

The voltmeters were calibrated with a Josephson voltage standard before each new mass determination. A Vishay HZ-series resistor with a nominal value of 1 kΩ was used as a shunt and was calibrated at PTB at the beginning of the measurement campaign.

For the 1 g mass of the HAF set, the main contributions to the relative standard uncertainty were Type A uncertainty 7.9×10^{-6}, residual coil current effects 1.3×10^{-6}, and vertical alignment

1.2×10^{-6}. The Type A uncertainty is a linear combination of the uncertainty determined from a statistical analysis of the weighings and Bl-determinations as well as the Type A uncertainty of the determination of the observed bias effects.

4 Going with the times: DCC and DME

4.1 Digital Calibration Certificates

Digital data exchange was introduced using the Digital Calibration Certificate (DCC), reflecting ongoing efforts toward increased digitalization of calibration workflows. The DCC is an XML-based format that provides structured, machine- and human-readable calibration data together with associated metadata. It contains at least the information provided by a traditional calibration certificate and fulfills the requirements of ISO/IEC 17025 [33]. While not yet widely adopted, the supporting infrastructure is already available, though still evolving. For implementation, existing tools such as the PTB GEMIMEG tool [34] were used. Compared to simple data formats such as text files or spreadsheets, the DCC enables standardized and unambiguous data exchange.

For the comparison, a DCC template was prepared and distributed to all participants and manufacturers. Each mass set and each individual mass within the set was assigned a unique identifier (ID), e.g., `Haf_6.62607015g`. Participants and manufacturers completed the template by providing their measured values and the associated uncertainties for each ID.

In contrast to previous work [35], where a single mass value per certificate was reported, two sets of masses were considered in this comparison. For each mass set, one DCC was provided by the manufacturer and one by each of the three participants. Each certificate contained, in its final form, the complete mass set of six values. The certificates included the measured values, associated uncertainties, coverage factors, and measurement dates. Additional metadata, such as environmental conditions, were included for completeness. This approach enables standardized data storage and facilitates subsequent analysis. The included metadata ensure that the data remain interpretable even beyond the original context of the comparison.

4.2 Digital Metrological Expert for live unblinding

For the evaluation, the Digital Metrological Expert (DME) was used, a software tool developed at PTB for the analysis of interlaboratory comparisons based on DCC data, with the intended capability to generate comparison reports for the Key Comparison Database (KCDB) [36]. The DME extracts the relevant values from the provided DCCs and implements the evaluation scheme used in this study, including the calculation of the Reference Value (RV), Degrees of Equivalence (D), and normalized Errors (E_n). These terms are used here in the descriptive sense implemented in the DME/DCC framework. It then evaluates the results and generates the corresponding plots.

For this comparison, a dedicated evaluation routine was implemented by the development team at PTB. Participants uploaded their completed DCCs individually and received a corresponding identifier for each submission. These identifiers were retained by the participants and not shared prior to the evaluation. The unblinding was performed live during the Kibble Balance Technical Meeting in November 2025 at the BIPM in Sèvres, France. During the session, the participants provided their identifiers through a shared real-time collaborative editor (Etherpad) visible to all attendees [37], which were then used to link the corresponding DCCs within the DME. The evaluation was executed immediately, and the results were generated and presented in real time. The complete evaluation and unblinding procedure can be reproduced using the DCC files and the DME evaluation workflow provided in the supplementary material.

5 Comparison results

5.1 Input data and evaluation

Table 4 summarizes the measurement results reported by the participants and by the manufacturers of the corresponding mass sets. These data form the basis for the subsequent evaluation. The values shown in the table are rounded for presentation, whereas the evaluation of the comparison was carried out using the unrounded reported values.

The comparison reference value (RV) and its expanded uncertainty ($k = 2$) were calculated from the

Table 4. Reported mass values with expanded uncertainties ($k = 2$). Uncertainties are rounded to at most two significant digits, with values rounded accordingly. All values are expressed in the unit of the nominal mass.

Nominal mass	NIST	PTB	TUI	Manufacturer
		HAF mass set		
1 mg	0.9928(67)	1.000 62(46)	0.996(21)	1.0011(6)
50 mg	49.9934(52)	50.000 93(45)	50.003(13)	50.0014(12)
200 mg	199.9934(55)	200.002 18(53)	200.037(32)	200.0019(20)
1 g	0.999 997(8)	1.000 005 6(13)	0.999 965(17)	1.000 006(3)
6.626 g	6.626 082(25)	6.626 092(7)	6.626 017(37)	6.626 088(18)
10 g	10.000 015(38)	10.000 012(11)	10.000 044(53)	10.000 011(6)
		RL mass set		
1 mg	1.0009(29)	1.001 77(42)	1.001(55)	1.001 42(70)
50 mg	50.0019(30)	50.002 41(44)	49.947(68)	50.002 04(71)
200 mg	200.0017(34)	200.001 43(53)	199.98(12)	200.001 14(59)
1 g	1.000 001(10)	0.999 999 8(55)	0.999 95(12)	1.000 001 5(11)
6.626 g	6.625 990(26)	6.625 999 1(93)	6.626 27(16)	6.626 013 6(72)
10 g	10.000 006(38)	10.000 008(12)	10.000 343(73)	10.000 009 6(55)

Kibble balance measurements only using

$$m_{\mathrm{RV}} = \frac{\sum_{i=1}^{3} m_i \cdot U_i^{-2}}{\sum_{i=1}^{3} U_i^{-2}} \tag{3a}$$

$$U_{\mathrm{RV}} = \left(\sum_{i=1}^{3} U_i^{-2} \right)^{-1/2}, \tag{3b}$$

where m_i and U_i denote the reported mass values and associated expanded uncertainties of the three Kibble balances ($i = 1, 2, 3$), neglecting correlations between the individual results. Inverse-variance weighting is used, as commonly applied in interlaboratory comparisons [38]. In the literature this weighting is usually expressed in terms of standard uncertainties. Since all reported uncertainties here are expanded uncertainties with the same coverage factor $k = 2$, the weighting remains unchanged. The degree of equivalence and its expanded uncertainty are defined as

$$D_j = m_j - m_{\mathrm{RV}} \tag{4a}$$

$$U(D_j) = \begin{cases} \sqrt{U_j^2 - U_{\mathrm{RV}}^2}, & j = 1, 2, 3, \\ \sqrt{U_j^2 + U_{\mathrm{RV}}^2}, & j = 4, \end{cases} \tag{4b}$$

where $j = 1, 2, 3$ correspond to the three Kibble balance results and $j = 4$ denotes the value reported by the weight manufacturer. Compatibility was assessed using the normalized error

$$E_{n,j} = \frac{D_j}{U(D_j)}. \tag{5}$$

For the manufacturer results ($j = 4$), E_n indicates compatibility with the reference value, as they are not included in its calculation. Table 5 summarizes the reference values, degrees of equivalence, and normalized errors for both mass sets.

The comparison values were calculated using the unrounded input values provided in the final DCCs. Using the rounded values listed in Table 4 would lead to insignificant deviations in the resulting comparison values.

The reference values were compatible with the corresponding manufacturer calibration values for 11 of the 12 masses according to $|E_n| \leq 1$. The remaining case was the Planck-mass of the RL set, with $E_n = 1.3$. The individual Kibble balance results showed a broader spread, with 22 of 36 values

Table 5. Comparison reference values (RV), degrees of equivalence (D), and normalized errors (E_n) for both mass sets. RV values are derived from the Kibble balance measurements, and are provided with expanded uncertainties ($k = 2$). Uncertainties are generally rounded to two significant digits, with values rounded accordingly. RV values are expressed in the unit of the nominal mass, while D values are given in µg. E_n values are dimensionless.

		NIST		PTB		TUI		Manufacturer	
Nominal mass	RV	D	E_n	D	E_n	D	E_n	D	E_n
HAF mass set									
1 mg	1.000 58(46)	−8(7)	−1.2	0.04(3)	1.2	−5(21)	−0.2	0.5(8)	0.7
50 mg	50.000 88(44)	−7(5)	−1.4	0.05(4)	1.3	2(13)	0.2	1(1)	0.4
200 mg	200.002 11(53)	−9(5)	−1.6	0.07(5)	1.4	34(32)	1.1	0(2)	−0.1
1 g	1.000 005 2(13)	−8(8)	−1.0	0.5(2)	2.0	−41(16)	−2.5	1(3)	0.3
6.626 g	6.626 088 8(66)	−7(24)	−0.3	3(2)	1.4	−72(36)	−2.0	−1(19)	0.0
10 g	10.000 013(10)	2(37)	0.1	−1(3)	−0.4	31(52)	0.6	−2(12)	−0.2
RL mass set									
1 mg	1.001 75(42)	−1(3)	−0.3	0.02(6)	0.3	−1(55)	0.0	−0.3(8)	−0.4
50 mg	50.002 39(44)	0(3)	−0.2	0.01(6)	0.2	−56(68)	−0.8	−0.4(8)	−0.4
200 mg	200.001 43(52)	0(3)	0.1	−0.01(8)	−0.1	−21(116)	−0.2	−0.3(8)	−0.4
1 g	1.000 000 0(48)	1(9)	0.1	0(3)	−0.1	−54(116)	−0.5	1(5)	0.3
6.626 g	6.625 998 9(87)	−9(25)	−0.4	0(3)	0.1	267(162)	1.6	15(11)	1.3
10 g	10.000 015(11)	−9(36)	−0.3	−8(4)	−2.0	328(72)	4.6	−6(12)	−0.5

satisfying $|E_n| \leq 1$ and 34 of 36 satisfying $|E_n| \leq 2$. The maximum absolute value was $|E_n| = 4.6$, corresponding to a largest absolute individual deviation of 328 µg from the reference value.

The normalized errors for the manufacturer values and for the participants have different roles. The manufacturer values were not included in the calculation of the reference values and therefore provide an external comparison with the Kibble-balance-derived reference values. The participant values instead describe the deviations of the individual Kibble balance results from the weighted reference values.

Because inverse-variance weighting is used, the reference values are mainly determined by the results with the smallest reported uncertainties. For this comparison with three participating Kibble balances, all reported results were kept in the calculation. No outlier rejection or consistency-based modification of the reference values was applied.

Across the individual Kibble balance results, the reported expanded uncertainties ranged from about 0.4 µg to 160 µg. The reported expanded uncertainties were also evaluated relative to each other and relative to the corresponding manufacturer calibration uncertainties. For the same nominal mass, the ratio between the largest and smallest participant uncertainty ranged from about 5 to about 230, with a median ratio of about 25. Relative to the corresponding manufacturer calibration uncertainty, the individual Kibble balance uncertainties ranged from about 0.27 to about 203, with a median ratio of about 4.3. These ratios show that the participating balances did not operate at a common uncertainty level during the comparison.

Taken together, the results show two aspects. The weighted Kibble-balance-derived reference values were compatible with the independent manufacturer calibrations according to the stated E_n criterion. At the same time, the individual systems showed clearly different reported uncertainties and normalized deviations under the conditions of this comparison.

A graphical representation of the results, shown as absolute values, is given in Figures 6 and 7.

5.2 Differences in reported uncertainties

The differences in reported uncertainties reflect the experimental configurations and uncertainty budgets used during the comparison. In addition, differences are observed between the two mass sets for the same nominal value and participant. These may partly reflect the different density uncertainties of the mass sets (see Table 2), which affect the air buoyancy correction. As the participating Kibble balances are still under active development, the reported uncertainties depend on the specific implementation and the operating conditions at the time of the measurements.

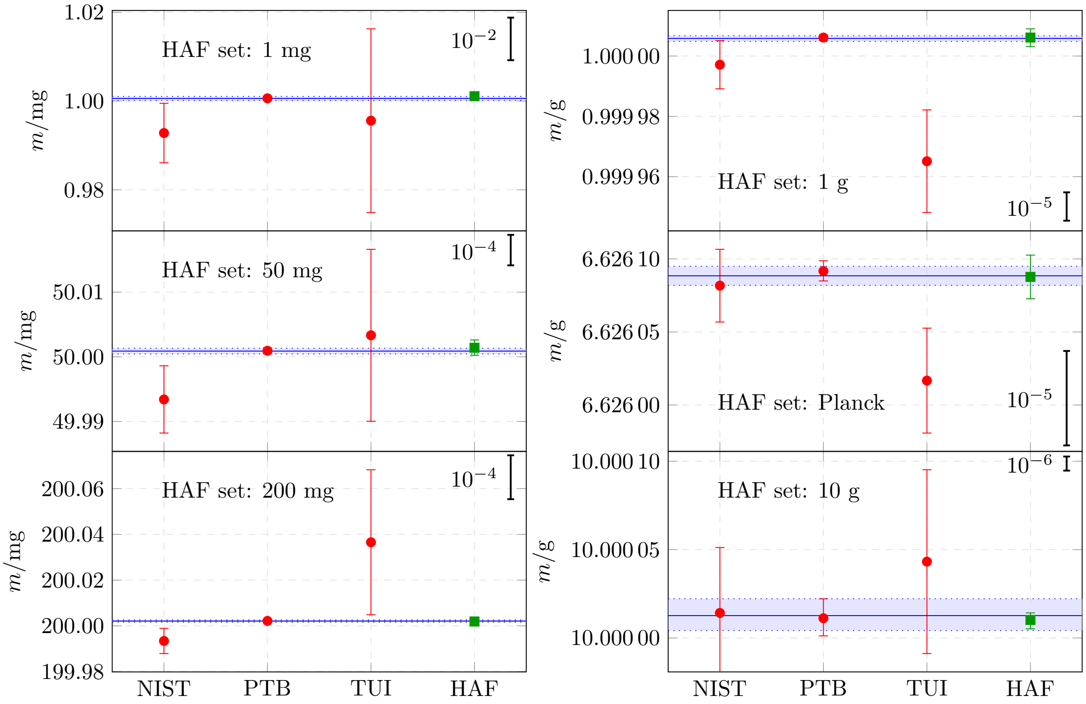

Figure 6. Comparison of measurement results for the HAF mass set. Red markers represent the measurements of the Kibble balances and green markers the values reported by HAF. The blue horizontal line indicates the reference value derived from the three Kibble balance measurements. Expanded uncertainties ($k = 2$) are shown as error bars for individual measurements and as a shaded band for the reference value.

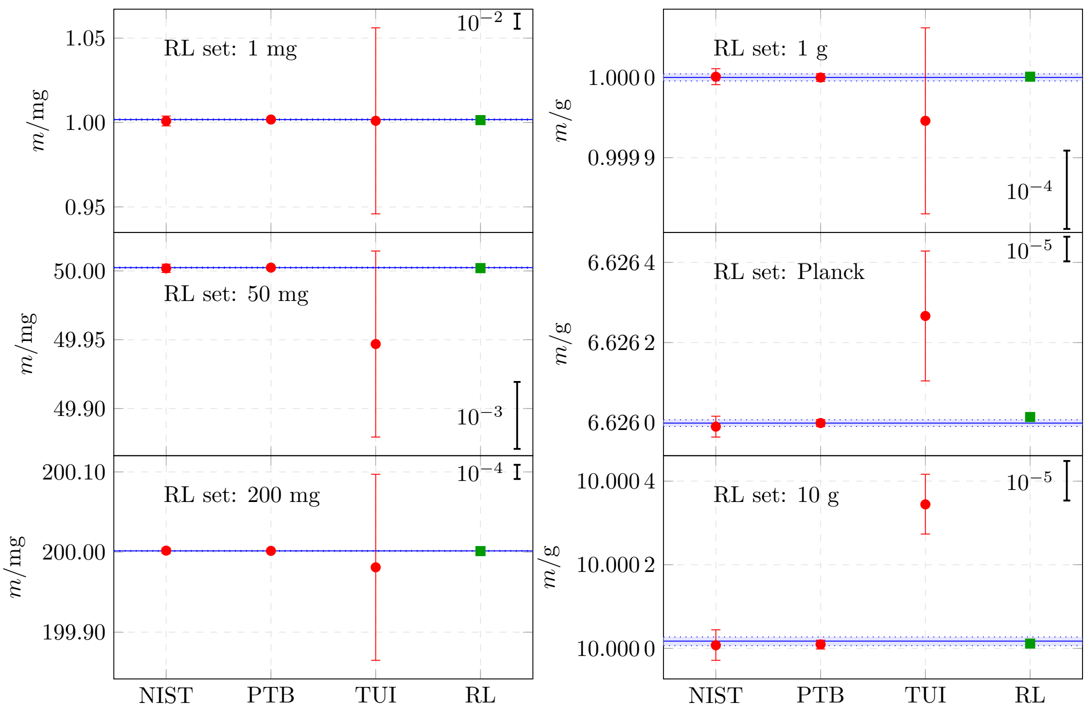

Figure 7. Same representation as in Figure 6, but for the RL mass set. The green markers indicate the values reported by RL.

For the TUI system, the larger uncertainties are mainly caused by an additional bias force effect observed in force mode measurements. The measurements were performed using the system configuration described in [28], for which the uncertainty budget had previously been established. After a modification of the mounting of the weighing pan and the interferometer mirrors, a bias force effect was observed when the system was returned to this configuration. Corrections based on measurements with an empty load lifter were applied. The statistical uncertainty associated with these corrections dominates the uncertainty budget of the reported TUI results. This is reflected in the comparatively large uncertainties shown in the figures. The effect is more pronounced for the RL mass set than for the HAF mass set, for which the measurements were performed later and the corresponding correction was better understood.

5.3 Practical aspects of the comparison campaign

Several practical aspects encountered during the comparison are relevant for future campaigns involving tabletop Kibble balances. First, the international transport of the mass sets caused logistical difficulties. Shipments sent via commercial courier services were delayed during customs clearance. Future comparisons should allow sufficient buffer time for international transport. Alternative ways of transferring the artifacts may be preferable.

Second, the handling of weights across a wide mass range introduces practical challenges. In particular, very small weights require careful placement to ensure stable positioning. The mechanical design of the pan and lifter must ensure reliable operation over several orders of magnitude in mass. In our observations, sheet weights were generally easier to handle than wire weights, although they provide a smaller contact length with the pan. These aspects do not represent fundamental limitations but require appropriate consideration in the design.

Third, the live unblinding using digital calibration certificates (DCCs) highlighted the importance of a fully tested evaluation workflow. Deviations from the agreed DCC template occurred in the manufacturer data. For the HAF mass set, individual DCCs were provided for each weight instead of a single certificate for the complete set. For the RL mass set, uncertainties were reported with a coverage factor of $k = 1.96$ instead of $k = 2$, which could not be processed by the DME at that time.

Rehearsal runs were performed prior to the meeting using test data sets in coordination with the PTB DME/DCC team. The manufacturers were not included in these tests to limit administrative effort and to avoid any risk of premature disclosure of the results. As a consequence, the live unblinding was less smooth than intended. The DME plots initially showed only the Kibble balance results, while the manufacturer results had to be communicated orally. For the final evaluation, the HAF DCCs were combined into a single certificate without modification of the underlying data, and the RL uncertainties were converted to $k = 2$. Future comparisons should enforce standardized DCC schemas and include full rehearsal runs with representative data from all participants.

6 Discussion and outlook

The comparison shows two separate aspects. The Kibble-balance-derived reference values were compatible with the independent manufacturer calibration values for 11 of the 12 investigated masses according to $|E_n| \leq 1$. At the same time, the individual balances showed different uncertainty levels and normalized deviations. This indicates a technically promising but not yet uniform state of tabletop Kibble balances in the investigated range from 1 mg to 10 g.

The inverse-variance weighting is important for interpreting the reference values. With only three participating balances and reported uncertainty ratios of up to about 230 for the same nominal mass, a weighted reference value can be dominated by the lowest-uncertainty result. The uncertainty spread is partly implementation-specific, including the spurious offset forces observed at TUI.

The measurements span nearly one year, and drift of the transfer standards cannot be excluded, particularly for small masses with high surface-to-mass ratios. Stability of transfer standards is a known issue in comparisons of kilogram submultiples [39, 40]. The evaluation is based on one final result per participant, mass value, and mass set. Possible transfer-standard changes are therefore not separated from differences between instruments.

Another observation concerns the dependence of the reported uncertainty on the nominal mass value in conventional calibration. For the non-OIML value of 6.626 g, the manufacturer uncertainties are noticeably larger than for OIML nominal values such as 1 g or 10 g. The magnitude of this increase is not readily explained by a simple scaling with mass. It may be related to the treatment of non-OIML nominal values in conventional calibration, although the specific procedures applied by the manufacturers are not known. In contrast, no comparable increase is observed in the Kibble balance results. The significance of this observation cannot be assessed on the basis of the present data alone.

It is also noteworthy that, for several masses, PTB reported smaller uncertainties than those stated in the manufacturer calibration certificates, particularly in the milligram range. However, these manufacturer uncertainties should not be interpreted as fundamental limits of calibration by mass comparison. They primarily reflect compliance with calibration requirements. Lower uncertainties are achievable within conventional mass comparison and are reflected in published entries for Calibration and Measurement Capabilities (CMCs).

The inclusion of tabletop Kibble balances in formal key comparisons may represent a relevant future step. Such comparisons could also include mass values outside the conventional 1–2–5 mass scale, for example 6.626 g. This would provide a pathway toward establishing CMCs and their publication in the BIPM Key Comparison Database. These aspects are separate from questions concerning the use of such systems in calibration practice, accreditation, or legal-metrology frameworks. This work addresses the use of tabletop Kibble balances under comparison conditions and may be relevant in the context of future standardization efforts.

Further intercomparisons involving additional laboratories operating compact Kibble balances would be of interest. Intercomparisons of this type provide insight into systematic effects and the consistency of results across different implementations. Broader participation may contribute to a more comprehensive assessment of such systems in comparison settings.

Acknowledgment

The authors express their gratitude to Hao Fang for facilitating the unblinding at the BIPM. They also acknowledge the mass manufacturers, Rice Lake Weighing Systems and Häfner Gewichte GmbH, for their steadfast support throughout the project. Furthermore, support from the PTB DME/DCC team was essential in adapting the existing infrastructure to this comparison. In particular, we would like to thank Wafa El Jaoua, David Urban, Moritz Gafert, Daniel Hutzschenreuter, Gisa Foyer and Shanna Schönhals for their dedicated contributions. The authors also thank Michael Borys and Dorothea Knopf for their valuable feedback and insights from mass metrology. The NIST team gratefully acknowledges Brad Suggs and Russell Kauffman from the U.S. Army Primary Standards Laboratory for their contributions to the development of KIBB-g2. Authors from TU Ilmenau were supported by the Deutsche Forschungsgemeinschaft (DFG) under projects 521501354 and the Walter Benjamin Program (546584462). Authors from PTB thank their colleagues for their contributions, in particular, the Scientific Instrumentation Department for fabricating the mechanical parts of the setup.

References

[1] M. Kochsiek and M. Gläser. Comprehensive Mass Metrology. Wiley-VCH Verlag Berlin GmbH, 1999.

[2] Bureau International des Poids et Mesures. The International System of Units (SI). Sèvres, France, 9th edition, 2019. Appendix 2: Mise en pratique for the definition of the kilogram in the SI.

[3] M. Stock, S. Davidson, H. Fang, M. Milton, E. de Mirandés, P. Richard, and C. Sutton. Maintaining and disseminating the kilogram following its redefinition. Metrologia, 54(6):S99–S107, 2017.

[4] M. Gläser, M. Borys, D. Ratschko, and R. Schwartz. Redefinition of the kilogram and the impact on its future dissemination. Metrologia, 47(4):419, 2010.

[5] Consultative Committee for Mass and Related Quantities (CCM). CCM detailed note on the dissemination process after the redefinition of the kilogram. Approved at the 17th CCM meeting, 16–17 May 2019, 2019. BIPM document.

[6] Bureau International des Poids et Mesures (BIPM), Sèvres, France. The International System of Units (SI), 9th edition, 2019. English version.

[7] International Organization of Legal Metrology. OIML R 111-1: Weights of classes E_1, E_2, F_1, F_2, M_1, M_{1-2}, M_2, M_{2-3} and M_3: Part 1: Metrological and technical requirements. OIML, Paris, France, 2004. Edition 2004 (E).

[8] P. Richard, H. Fang, and R. Davis. Foundation for the redefinition of the kilogram. Metrologia, 53(5):A6, 2016.

[9] M. Stock, P. Conceição, H. Fang, F. Bielsa, A. Kiss, Y. H. Wu, S. J. Shen, L. Tsao, S. J. Chen, M. Thomas, et al. CCM.M-K8.2024: Realization of the kilogram. CIPM MRA Comparison Reports, 2026.

[10] S. Li, N. Li, W. Liu, E. E. Qupasie, W. Zhao, and S. Huang. Design of a miniature Kibble balance for kilogram-scale mass calibration—KBmini. Metrologia, 63(2):025001, 2026.

[11] I. A. Robinson. The architecture of the NPL next generation Kibble balance. In Proceedings of the 24th IMEKO TC3 Conference on Measurement of Force, Mass and Torque, pages 198–203, 2023.

[12] J. Vanier and C. Audoin. The classical caesium beam frequency standard: fifty years later. Metrologia, 42(3):S31–S38, 2005.

[13] L. W. Alvarez. Alvarez: Adventures of a Physicist. Basic Books, New York, 1987.

[14] L. M. Branscomb. Views: Integrity in science: Much of the problem of honor—or lack of honor—in science stems not from malice but from self-deception. American Scientist, 73(5):421–423, 1985.

[15] B. P. Kibble. A suggestion for a different way of realizing the ampere. In J. H. Sanders and A. H. Wapstra, editors, Atomic Masses and Fundamental Constants, volume 5, pages 545–551. Plenum, New York, 1976.

[16] R. Steiner. History and progress on accurate measurements of the Planck constant. Reports on Progress in Physics, 76(1):016101, 2013.

[17] B. P. Kibble and I. A. Robinson. Principles of a new generation of simplified and accurate watt balances. Metrologia, 51(2):S132–S139, 2014.

[18] L. Keck, K. Arumugam, L. Chao, Z. Comden, F. Seifert, D. Newell, D. Haddad, and S. Schlamminger. Thoughts on the Kibble–Robinson theory. Metrologia, 62(2):025012, 2025.

[19] J. Konrad, N. Rogge, S. Vasilyan, C. Rothleitner, and T. Fröhlich. Experimental verification of coil motion misalignment insensitivity in a Kibble balance. Measurement: Sensors, 38:101350, 2025.

[20] L. Chao, F. Seifert, D. Haddad, J. Pratt, D. Newell, and S. Schlamminger. The performance of the KIBB-g1 tabletop Kibble balance at NIST. Metrologia, 57(3):035014, 2020.

[21] K. Arumugam, J. E. Draganov, L. Chao, Z. Comden, T. Mametja, S. Schlamminger, F. Seifert, and D. Newell. Honey I Shrank the Kibble Balance: A Second Generation NIST Table Top Balance. In 2024 Conference on Precision Electromagnetic Measurements (CPEM), 2024.

[22] S. Li, F. Bielsa, M. Stock, A. Kiss, and H. Fang. Coil-current effect in Kibble balances: analysis, measurement, and optimization. Metrologia, 55(1):75, 2017.

[23] R. Lapuh and B. Jeckelmann. Sampling with 3458A: Understanding, Programming, Sampling and Signal Processing. Left Right, 2018.

[24] C. Rothleitner, J. Schleichert, N. Rogge, L. Günther, S. Vasilyan, F. Hilbrunner, D. Knopf, T. Fröhlich, and F. Härtig. The Planck-Balance—using a fixed value of the Planck constant to calibrate E1/E2-weights. Measurement Science and Technology, 29(7):074003, 2018.

[25] J. Konrad, C. Rothleitner, and T. Fröhlich. A novel magnet system – which enables uniformity of radial flux density by means of a parabolic yoke – applied in the Planck-Balance. Measurement Science and Technology, 35(10):105004, 2024.

[26] C. M. Sutton. An oscillatory dynamic mode for a watt balance. Metrologia, 46(5):467–476, 2009.

[27] J. Konrad, C. Rothleitner, S. Vasilyan, N. Rogge, R. Behr, M. Götz, M. Kruskopf, and T. Fröhlich. Small mass value realization equivalent to accuracy class E1 using the Planck-Balance at PTB. Metrologia, 62(2):025014, 2025.

[28] S. Vasilyan, N. Rogge, C. Rothleitner, S. Lin, I. Poroskun, D. Knopf, F. Härtig, and T. Fröhlich. The progress in development of the Planck-Balance 2 (PB2): A tabletop Kibble balance for the mass calibration of E2 class weights. tm – Technisches Messen, 88(12):731–756, 2021.

[29] F. Hilbrunner, I. Rahneberg, and T. Fröhlich. Wattwaage mit Hebelübersetzung auf Basis eines kommerziellen EMK-Wägesystems. tm – Technisches Messen, 85(11):658–679, 2018.

[30] J. Schleichert, S. Vasilyan, and L. Günther. Magnet system for the Planck-Balance. In Proceedings of the 59th Ilmenau Scientific Colloquium, Ilmenau, Germany, September 2017.

[31] S. Lin, C. Rothleitner, and N. Rogge. Investigations on the sine fitting algorithm in the Planck-Balance. In Proceedings of the 20th GMA/ITG Conference on Sensors and Measurement Systems, pages 547–553, 2019.

[32] S. Vasilyan, N. Rogge, H. Preißler, M. Starkloff, M. Schubert, and T. Fröhlich. Adaptation of metrology-grade ac current source in velocity mode of Planck-Balance 2: direct referencing induced voltages with ac quantum voltage standard. Measurement Science and Technology, 35(1):015026, 2023.

[33] Physikalisch-Technische Bundesanstalt (PTB). Digital Calibration Certificate (DCC). https://www.ptb.de/cms/en/metrological-services/dkd/dkd-dcc.html. Accessed 2026-03-06.

[34] Physikalisch-Technische Bundesanstalt (PTB). GEMIMEG Tool. https://gemimeg-tool.ptb.de/v1/#/. Version v1, accessed 2026-02-17.

[35] B. Rodiek, G. Álvarez Clara, C. Brown, S. Davidson, M.-A. Demir, S. Eichstädt, W. El Jaoua, Y. H. Fung, J.-L. Hippolyte, T. Hoffmann, D. Hutzschenreuter, M. Jordan, B. Korutlu, N. Kuramoto, S. Cho, Z. Li, A. Malengo, N. Murnaghan, L. M. Peña Pérez, S. Schönhals, and I. Smith. Project on a fully automated evaluation of a virtual comparison of mass using the Digital Calibration Certificate (DCC) schema. Measurement: Sensors, 38:101361, 2025.

[36] C. Brown, W. El Jaoua, T. Hoffmann, D. Hutzschenreuter, B. Rodiek, and S. Schönhals. A digital expert for a digital key comparison report processor. Measurement: Sensors, 38:101463, 2025.

[37] The Etherpad Foundation. Etherpad: A modern really-real-time collaborative editor. https://etherpad.org/. Version 2.6.1, accessed 2026-02-13.

[38] International Organization for Standardization. Statistical methods for use in proficiency testing by interlaboratory comparisons. ISO Standard 13528, 2015.

[39] L. O. Becerra, W. Bich, N. Bignell, G. D. Chapman, J. W. Chung, S. Davidson, M. Gläser, A. Gosset, Z. J. Jabbour, W. G. Lee, et al. Final Report on CIPM key comparison of multiples and submultiples of the kilogram (CCM.M-K2). Metrologia, 40(1A):07004, 2003.

[40] Stuart Davidson. Report on EURAMET.M.M-S2: Supplementary comparison of 100 gram, 10 gram, 500 microgram, 200 microgram and 100 microgram weights. Metrologia, 48(1A):07005, 2011.