

# Recalibration of the U.S. National Prototype Kilogram

R. S. Davis

National Bureau of Standards, Gaithersburg, MD 20899

Accepted: June 14, 1985

The U.S. national prototype kilogram, K20, and its check standard, K4, were recalibrated at the Bureau International des Poids et Mesures (BIPM). Both these kilograms are made of platinum-iridium alloy. Two additional kilograms, made of different alloys of stainless steel, were also included in the calibrations. The mass of K20 in 1889 was certified as being 1 kg-0.039 mg. Prior to the work reported below, K20 was most recently recalibrated at the BIPM in 1948 and certified as having a mass of 1 kg-0.019 mg. K4 had never been recalibrated. Its initial certification in 1889 stated its mass as 1 kg-0.075 mg. The work reported below establishes the new mass value of K20 as 1 kg-0.022 mg and that of K4 as 1 kg-0.106 mg. The new results are discussed in detail and an attempt is made to assess the long-term stability of the standards involved with a view toward assigning a realistic uncertainty to the measurements.

**Key words:** International System of Units; kilogram unit; mass; national prototype kilograms; platinum-iridium kilograms; stainless steel kilograms; standards of mass; Système International d'Unités.

## 1. Introduction

The International Prototype Kilogram (IPK), made of an alloy of 90% platinum and 10% iridium, is kept at the International Bureau of Weights and Measures, or Bureau International des Poids et Mesures (BIPM), in Sèvres, France. The mass of this single artifact defines the kilogram unit in the International System of Units (SI): "The kilogram is the unit of mass; it is equal to the mass of the international prototype of the kilogram" [1]<sup>1</sup>. Thus, calibration of national standards in SI units implies traceability to the IPK.

The IPK is actually one of three nearly identical kilo-

gram artifacts ordered in 1878 from Johnson, Matthey and Company of London. Four years later, 40 more replicas of the IPK were ordered; these eventually becoming the first national prototypes. The United States was allotted artifacts No. 4 and No. 20 from these 40. (It is customary to refer to the *i*<sup>th</sup> national prototype as *K<sub>i</sub>*. Thus NBS is in possession of national prototypes K4 and K20.) Since their receipt by the U.S. government, K20 has been considered the national standard and K4 the "check standard" by which the constancy of K20 is monitored.

The BIPM now stores six replica or check kilograms in the same vault along with the IPK. All seven are kept under the strict control of the International Committee for Weights and Measures (CIPM). A brief history of the IPK and its check standards through 1975 is contained in [2].

The establishment of long-term stability limits for the national prototypes is important—if only because of the

---

**About the Author:** R. S. Davis is a physicist in the Length and Mass Division of NBS' Center for Basic Standards.

---

<sup>1</sup> Numbers in brackets indicate literature references.

infrequency of international comparisons. For this reason, we have included a fairly detailed review of the calibration history of the U.S. prototypes as well as a complete report of their recalibration by the BIPM in 1984.

The mass calibration service of NBS uses nichrome working standards at the kilogram level. These have a density of about  $8.4 \text{ g/cm}^3$ , considerably lower than the  $21.55 \text{ g/cm}^3$  of the prototypes. Calibration of a nickel-chrome or a stainless steel kilogram by comparison to a platinum-iridium prototype presents many additional problems. Because of the metrological importance of these problems, two stainless steel kilograms, designated CH-1 and D2, also were sent to BIPM for calibration. These results are also examined.

## 2. BIPM Working Standards

The IPK has remained locked in a strongbox at the BIPM since its last use in 1946. The use of it or of any of its six check standards, which reside in the same enclosure, requires special permission of the CIPM. For routine calibrations, the BIPM uses two prototype kilograms, K9 and K31. These are now compared, at about

5-year intervals, with K25—a prototype which is reserved for exceptional usage. The use of K25 can be authorized by the director of the BIPM—a procedure that is far simpler than the protocol for use of either the IPK or any of its six check standards. The last measurements made with any of the six check standards occurred in 1965. Since then, BIPM calibrations of mass have been based on the assumption of long-term stability of K25. At 5-year periods, K25 has been cleaned and then used to recalibrate, K9 and K31.

The calibration history of K9 and K31 is shown in figure 1. Working standards K9 and K31 are each assigned an uncertainty at a level of one standard deviation of  $8 \mu\text{g}$  [3]. Their present mass values and uncertainties result from the following chain:

Measurements	Year
K31 and K9 against K25	1983
K25 against two IPK check standards	1965
IPK check standards against the IPK	1946

The 1983 measurements were performed using the NBS-2 balance (see sec. 5.1) and thus are assigned a

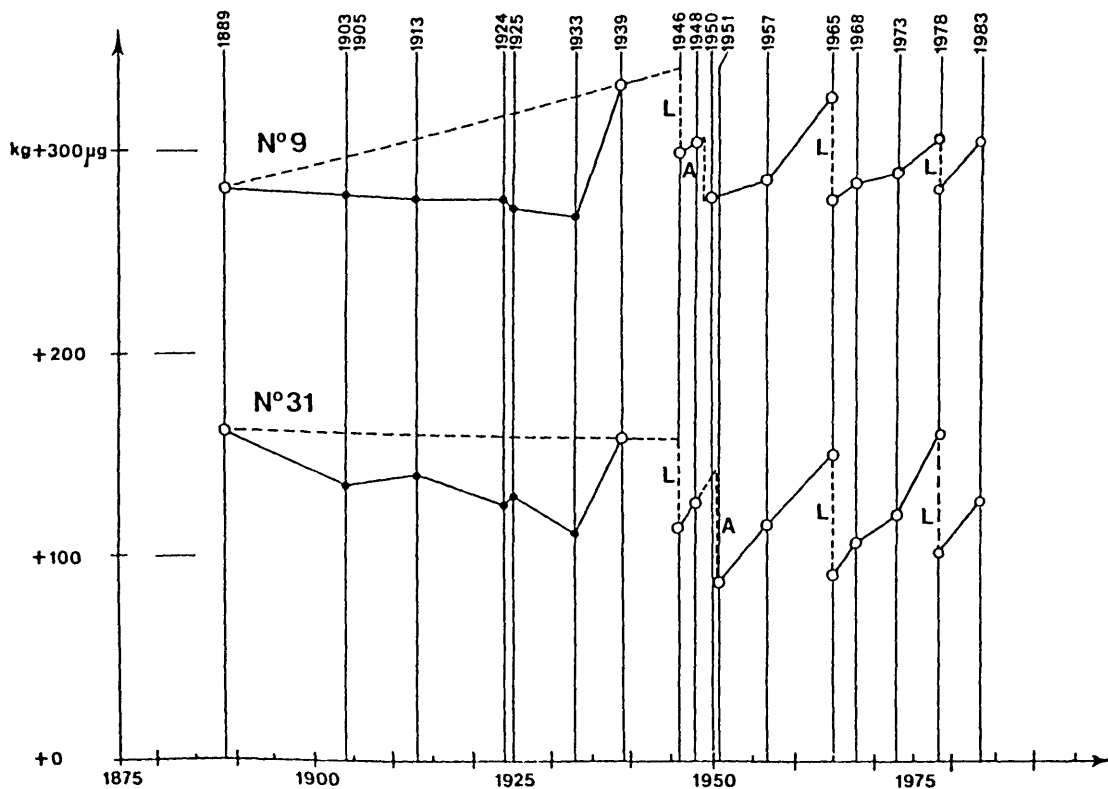


Figure 1—Calibration history of prototypes 9 and 31. The filled circles represent measurements which are tied less directly to the IPK than the open circles. Cleanings according to the method of Bonhoure [4] are indicated by "L." Two accidents are labeled "A." This graph, which extends the graph published in [2], was kindly supplied to us by the BIPM.

standard deviation of 1  $\mu\text{g}$ . Measurements prior to 1970 were made primarily on a Ruedrecht balance with a standard deviation for a single measurement of about 10 micrograms. Since the adoption of the cleaning method described by Bonhoure in 1946 [4], the long-term behavior of K9 and K31 has been easily interpreted: 1) the two accidents involved an appreciable loss in mass; 2) both prototypes show a mass increase with time if they are not cleaned (relative to K25 which was cleaned prior to the comparisons); 3) K31 has a tendency to pick up pollutants more quickly than K9 (the scratched surfaces of K31 make it a better trap for airborne particles [5]); and 4) cleaning the prototypes restores their mass to within 10  $\mu\text{g}$  of their prior "clean" values.

### 3. Plan for BIPM Calibration of NBS Artifacts

Four one-kilogram artifacts were hand-carried from NBS, Gaithersburg to BIPM. The transatlantic crossings were made on commercial airlines. The artifacts were stored in the passenger section of the aircraft. Two of the artifacts, K4 and K20, are the national prototypes belonging to the United States. The other two, CH-1 and D2, are made of stainless steel. A detailed description of the artifacts used is given in section 4.

The carrying container for K20 was essentially that described in [6]. In this type of enclosure, the kilogram is held firmly on the top and bottom and is also clamped gently at three places around the side. All clamp faces conform to the contour of the adjacent kilogram surface. The clamped areas of the kilogram are protected by low-abrasive tissue paper which is backed by chamois skin, the latter having been previously degreased through successive soakings in benzene and ethanol. The outer case of the container is metal. The seal is not airtight.

The carrying case for K4 was of simpler design. The artifact was completely wrapped in tissue, then wrapped in chamois skin, and finally placed in a snugly-fitting brass container. Again, the container seal was not airtight.

Both stainless steel kilograms were wrapped in tissue paper and were then padded with successive layers of cotton batting and soft polyethylene foam. The outer container was a stiff cardboard tube. The padding held the kilogram fast within the tube.

After their arrival at BIPM at the beginning of December, 1983, all the artifacts were unpacked and stored in the mass laboratory. In February 1984 they were all calibrated twice against the BIPM working standards, K9 and K31. The first calibration was made without cleaning any of the artifacts. Before the second calibration, K20 and K4 were cleaned in the prescribed

manner—wiped with benzene-soaked chamois skin, wiped with ethanol-soaked chamois skin, and then cleaned with steam from bi-distilled tap-water [4]. Except for steaming, the stainless steel artifacts underwent an identical cleaning procedure prior to the second calibration.

## 4. Description and History of Mass Standards Used

### 4.1 National Prototype K20 and K4

The manufacture of K20 and K4 has been well documented [6]. Such prototypes are replicas of the IPK. They are made of an alloy of 90% platinum and 10% iridium. At the time of manufacture, the ingots were swaged until their density reached a maximum, then formed into right-circular cylinders (height=diameter = 39 mm), and polished to a "rather handsome, but not specular" finish [7]. The edges of each prototype are slightly chamfered and each unit is identified with a number which is delineated by a slight change in the surface finish. The number is located on the side of the cylinder.

The surfaces of all the prototypes were carefully examined at the time of manufacture. The final step of polishing was done with emery which left the surface with a regular pattern of shallow marks. (Measurements on K22 reported in 1960 give a mean depth of 2  $\mu\text{m}$  to these marks [8]). Occasionally small pits (*piqûres*) with a maximum diameter of 0.1 mm were noticed on the finished surfaces. K20 was found to have two such pits on its top and some striation marks in a small area on its base. The center of K4's top surface was found to have a dull finish and striations were noted in another small area of the top.

The surface of K20 has remained nearly as initially described. However, an NBS notebook entry of December 5, 1947 remarks that "K4 has numerous scratches on top and bottom and a black spot 15/32" [11.9 mm] from the bottom almost opposite the '4'." There is no indication of how or when K4 suffered this change in appearance, which persists.

Inspection of K4 after a cleaning in December 1971 revealed "two scratches on the top that appeared to be new." (See Appendix.)

The 1889 certificate listed the following properties for the U.S. national prototypes [6].

Mass	Volume <sup>2</sup> at 0 °C
K20 1 kg—0.039 mg	46.4030 cm <sup>3</sup>
K4 1 kg—0.075 mg	46.4296 cm <sup>3</sup>

<sup>2</sup> The volume in the original certificate was specified in milliliters. These numbers were converted to cubic centimeters by using the relation 1 ml = 1.000028 cm<sup>3</sup>.

With a coefficient of volume expansion to a temperature  $t(^{\circ}\text{C})$  of <sup>3</sup>:

$$\alpha = (25.859 + 0.0065t) \times 10^{-6}/^{\circ}\text{C}$$

The expansion is assumed the same for all prototypes.

The standard deviation assigned to the measured masses was 3.3  $\mu\text{g}$  (we have inferred this from the reported probable error of 2.2  $\mu\text{g}$ ). Measurements were made on balances having a standard deviation of 10  $\mu\text{g}$  and results were solved by the method of least squares. No additional sources of error are mentioned. The fact that all artifacts were virtually identical in their physical properties undoubtedly minimizes or eliminates most errors of a nonstatistical nature (but see 5.1.3 below). The standard deviations for the volume assignments to K4 and K20 may again be inferred from the probable errors listed, i.e., 0.00010  $\text{cm}^3$  and 0.00028  $\text{cm}^3$ . It was, however, recognized that in the case of hydrostatic volume measurements, many nonstatistical sources of error intervene. These sources are given in [9] along with a description of the pains taken to eliminate their effects.

After their initial calibration, artifacts K4 and K20 were shipped to the U.S. on separate vessels. K20 arrived first and was received with sufficient bureaucratic pomp to secure its status as the primary standard of mass in the U.S., K4 being henceforth treated as a check standard [10].

The U.S. artifacts were transferred to NBS soon after its establishment.

Prior to the work reported below, K20 was returned twice to BIPM—in 1937 and 1948. There was to have been the second reverification of all national prototypes in 1939 (the first occurred in 1899–1900; the U.S. did not participate). It is possible that K20 was sent to BIPM in 1937 as part of this planned exercise. At any rate, the deteriorating international situation put an end to the reverification plans. K20 was returned to NBS in 1938 with only a preliminary calibration certificate. After World War II, the results of this measurement were stated to be: mass of K20 = 1 kg–0.021 mg. The standard deviation of the result was thought to be an insignificant contribution to the total uncertainty of about 20  $\mu\text{g}$ . The total uncertainty was based on experience with other prototypes which had been recalibrated.

The second periodic reverification was carried out in 1948–54, and K20 was again sent to BIPM for a portion of 1948. Measurements were made using K32 and K43, check standards of the IPK, as working standards. These check standards had been compared directly to the IPK in 1946 [11]. The result of this work showed the mass of K20 to be 1 kg–0.019 mg.

<sup>3</sup> This number was modified to  $(25.863 + 0.00562t) \times 10^{-6}/^{\circ}\text{C}$  in the 1947 certificate.

Although K4 had never been returned to BIPM, it was used several times at NBS in its role as check standard to K20. The results are summarized in table 1, which also includes the results of the newest comparisons. The implications of this table for long-term stability of K4 and K20 are discussed in the Appendix.

## 4.2 K650

In order to ascertain whether the artifacts sent to BIPM for calibration were in some way altered by their travel, it was essential to have at least one platinum-iridium artifact at NBS which could remain as a check standard. This was graciously provided by BIPM in the form of K650. This is one of a new series of prototypes

Table 1

Year	K20	K20-K4	Comments
1889	1 kg–0.039 mg	0.036 mg	BIPM, cleaned with alcohol and water vapors
1922		0.036	NBS, not cleaned
1937		0.05	NBS, prior to shipment to BIPM
1937	1 kg–0.021 mg		BIPM, wiped clean
1938		0.039	NBS, subsequent to return from BIPM
1947		0.045	NBS, prior to shipment to BIPM
1948	1 kg–0.019 mg		BIPM
1949		0.011	NBS, after return from BIPM
1970		0.058	NBS, measurements made on NBS-2
1971		0.038	NBS
1972		0.066	NBS
1983		0.072	NBS, prior to shipment to BIPM
1984	1 kg–0.001 mg	0.074	BIPM, as received
1984	1 kg–0.022 mg	0.084	BIPM, after cleaning
1984		0.075	NBS, after return from BIPM (as received)
1984		0.081	NBS, after cleaning

which is being manufactured by diamond turning [12], a machining process that leaves the surface of the kilogram in no need of polishing. As K650 was an early attempt, and the difficult task of machining to mass tolerance was not yet perfected, the finished mass of K650 was some 2 mg too small. This rendered it unsuitable for service as a national prototype but ideal for our purposes.

Its characteristics were described as follows:

Mass	Volume at 0 °C
1 kg—2.256 mg	46.4352 cm <sup>3</sup>

with a coefficient of volume expansion,

$$\alpha = (25.863 + 0.00562t) \times 10^{-6}/^{\circ}\text{C}.$$

The artifact was received by NBS in good condition. Prior to its first use, however, a wide, shallow scratch was noticed on its basal surface. This damage was thought to invalidate its mass assignment; however, later recalibration at BIPM showed no significant decrease in mass (see results below). The artifact was returned to BIPM in October 1984.

### 4.3 KA

A second kilogram was also used as a check standard at NBS during the recent measurements. This artifact has belonged to the U.S. since 1821. It is designated KA and has been referred to as the "Arago kilogram" because its original certificate bore the surname of Dominique-Francois Arago, the 19th century scientist and liberal statesman. Its shape is a cylinder of minimum surface (i.e., height equal to diameter).

This artifact was manufactured from platinum sponge. Its density (20.8863 g/cm<sup>3</sup> at 20 °C) is about 2.6% below theoretical and its surface shows a great deal of pitting. The mass of this artifact is demonstrably unstable, decreasing monotonically with time, as can be seen in table 2. In this respect, its behavior is identical to other old platinum weights of similar manufacture—the British pound standards and the so-called Kilogram of the Archives [11], for example. But, while KA is unsuitable as a long-term standard, its short-term stability is sufficient for its use as a second check standard.

Table 2

Year	Mass of KA
1888	1 kg—4.63 mg
1922	1 kg—4.753 mg
1970	1 kg—4.803 mg
1984	1 kg—4.845 mg

### 4.4 D2, CH-1

Two stainless steel artifacts were also used in the comparisons with BIPM. These are designated D2 and CH-1. They are single-piece weights which are roughly cylindrical in shape except for a lifting knob on the upper surface. The bases of both weights are relieved so that each weight rests on a narrow annular area whose outer diameter is that of the weight itself. Their important physical properties are:

Weight	Nominal Mass	Volume at 20 °C	Coefficient of Expansion
D2	1 kg+13.49 mg	127.6250 cm <sup>3</sup>	45×10 <sup>-6</sup> /°C
CH-1	1 kg—0.36 mg	124.9681 cm <sup>3</sup>	45×10 <sup>-6</sup> /°C

D2 was manufactured by the Troemner Company of Philadelphia from an austenitic alloy similar to 18-8 stainless steel. D2 has been used extensively in calibration and research work at NBS for the last 15 years.

CH-1 was acquired from the Chyo Company of Kyoto, Japan in 1983. It was manufactured from an austenitic stainless steel alloy having the following composition: 25.1% Ni, 29.9% Cr, 2.2% Mo, 1.45% Mn, 0.53% Si, 0.2% Cu, 0.07% C, 0.019% P. The metal was vacuum melted before being machined.

The volumes of both weights were determined at NBS at a temperature of 23 °C using distilled water as the standard. The uncertainty in the volume determination is calculated to be 5 ppm.

The volumetric thermal expansion of D2 near room temperature was supplied by the manufacturer. The volume expansion of CH-1 was determined by linear expansion measurements at NBS made on a sample rod supplied by the manufacturer.

## 5. Principles of Mass Comparisons

### 5.1 Description and Function of an Ideal Balance

The balances used in the recent mass comparisons were: NBS-2, a single-pan balance designed and built at NBS and then permanently transferred to BIPM in 1970 [13]; and V-1, the primary kilogram comparator of NBS, manufactured by the Volland Corporation of Hawthorne, NY. Both balances are similar in design and are based on design principles established by NBS researchers during the 1960's [14,15].

The major design features of these balances are well known [13]. Weights are manipulated on or off the pan remotely by the proper combination of raising, lowering, and rotation of a weight table. Comparisons of two weights are done by substitution on the single pan,

often referred to as Borda's method. Both balances are free-swinging so that the equilibrium position of the beam must be inferred from brief observations of the beam oscillations. At both BIPM and NBS, these observations consist of recording successive turning points of the swinging beam. These points are at present determined by reading an optical scale, the image of which is projected from a mirror fixed to the balance beam. At NBS, five successive turning points are read. The beam amplitude is always adjusted to  $2 \text{ mg} \pm 0.04 \text{ mg}$  (peak to peak). Since the beam has a period of about 30 s, this means that the beam swings freely for 1-1/2 min for each determination of an equilibrium position. The beam is first stopped using a braking mechanism at an angle corresponding to 2 mg below equilibrium. The brake is then released and readings commence after one complete oscillation. The first turning point is taken on the same side of the equilibrium position for every measurement. If the five turning points read are  $l_1, l_2, \dots, l_5$ , we infer the rest point  $\lambda$  from the relation

$$\lambda = (1/12)(2l_1 + 3l_2 + 2l_3 + 3l_4 + 2l_5). \quad (1)$$

Bignell has recently discussed the advantages of eq (1) over other possible formulas [16]. For our measurements, we may use Bignell's results to show that eq (1) leads to a bias in  $\lambda$  of approximately  $0.01 \mu\text{g}$ . This bias will be constant to 2%. It is thus negligible.

A very important feature of these balances is the inclusion of a constant-load stop which maintains all but about 20 mg of the full 1-kg load on the beam, its knives and its bearings, through the exchange of weights. This feature, which causes the balance behavior to better approximate ideality, has been discussed in great detail elsewhere [15].

Differences in balance equilibrium position ideally correspond to different forces on the pan. In order to relate observed differences in beam angle to differences in force (or mass) it is necessary to determine the balance sensitivity. This is accomplished by the addition of a small sensitivity weight  $s$  (2 mg at NBS). The mass and volume of this weight must be known sufficiently well so that the apparent mass which it adds to the balance pan is determined to  $\sim 0.5 \mu\text{g}$ . In determining the mass of the sensitivity weight from a calibration chain starting from 1-kg standards, the kilogram standards need only be known to  $\pm 0.25 \text{ g}$  (i.e., 1 part in 4000), a tolerance which is never in doubt.

When one considers that the ratio of the beam arms of the balance ideally should be constant to  $1 \times 10^{-9}$  (i.e., a differential change in length of less than  $0.075 \text{ nm}$  for V-1), it is not surprising that successive equilibrium readings of a single weight often increase or decrease monotonically for long periods of time. This could be

due, for instance, to a very slight differential heating of the two ends of the brass beam ( $\Delta T = 56 \mu\text{K}$  would be sufficient). A change in balance equilibrium for successive observations of the same weight will be referred to as balance drift.

In order to take account of linear drifts, as well as to measure the balance sensitivity, the following double substitution involving five equilibrium observations is used to determine the difference in apparent mass<sup>4</sup> between an object A and an object B:

Observation	Balance Equilibrium Value	Object on Balance Pan
1	$\lambda_1$	A
2	$\lambda_2$	B
3	$\lambda_3$	B + s
4	$\lambda_4$	A + s
5	$\lambda_5$	A

The apparent mass difference between A and B,  $P(A-B)$  is then estimated by

$$P(A-B) = \frac{\lambda_1 - \lambda_2 - \lambda_3 - \lambda_4}{-\lambda_2 + \lambda_3 + \lambda_4 - \lambda_5} P(s), \quad (2)$$

where  $P(s) = m_s - \rho v_s$ ,  $\rho$  is the average air density during the double substitution and  $v_s$  is the volume of  $s$ .

Using eq (2) to estimate  $P(A-B)$  serves to eliminate the effect of a linear drift in time of the equilibrium value of the balance. In fact, it has been pointed out [53] that the right-hand side (r.h.s.) of (2) provides an unbiased estimate for  $P(A-B)$  with no statistical degrees of freedom, provided only that the drift between observations 1 and 2 equals that between observations 3 and 4, and the drift between observations 2 and 3 equals that between observations 4 and 5. This may be a better model of reality than a simple linear drift because the drift which occurs when kilogram artifacts are exchanged on the balance pan may differ from that which occurs when only  $s$  is added or subtracted (although even in the latter case, the kilogram weight is removed from the pan and then replaced).

The standard deviation of measurement of  $P(A-B)$  on V-1 is about  $4 \mu\text{g}$  and about  $1 \mu\text{g}$  on NBS-2.

In order to estimate the mass difference between A and B,  $m_A - m_B$ , from  $P(A-B)$ , one must recognize that

$$m_A - m_B = P(A-B) + \rho(V_A - V_B) + f_1(\Delta s) + f_2(\Delta h) \quad (3)$$

where  $V_A$  and  $V_B$  are the volumes of A and B,  $f_1$  is a function of the difference in surface of A and B and  $f_2$  is

<sup>4</sup> By "apparent mass" we mean the force introduced by adding an object to the balance pan divided by the gravitational acceleration at the center of mass of the added object.

a function of the difference in centers of mass of A and B. The functions  $f_1$  and  $f_2$  are often neglected although, for many conditions, are not really negligible. Their effects will be mentioned below. For the moment, we will estimate  $m_A - m_B$  by the relation

$$m_A - m_B = \frac{\lambda_1 - \lambda_2 - \lambda_3 + \lambda_4}{-\lambda_2 + \lambda_3 + \lambda_4 - \lambda_5} (m_s - \rho V_s) + \rho(V_A - V_B). \quad (4)$$

The determination of  $m_A - m_B$  by eq (4) has no statistical degrees of freedom. In order to provide a good statistical check, four weights, A, B, C, and D, are usually intercompared in all six possible pairings: A-B, A-C, A-D, B-C, B-D, C-D. This can be thought of as a weighing "design" composed of six subweighings. If the mass of one of these weights, for example A, is known, the mass of the others can be determined by least squares fitting to the data [17]. If  $m_B$ , for example, is also known, the least squares estimate of  $m_B$  can be compared to the accepted value, thus providing the means for a t-test [18].

## 5.2 Non-Ideal Behavior

The V-1 and NBS-2 balances do not behave in the ideal way described above. Thus the method of calibrating unknown masses must be altered from the simple algorithm just outlined. We now discuss the necessary modifications.

**5.2.1 Pressure Drift:** The NBS-2 balance is sealed against changes in the ambient barometric pressure. The V-1 balance, however, is not sealed. Thus changes in barometric pressure will affect that balance equilibrium during a double substitution if the volume of the weight on the pan differs significantly from the volume of the balance counter-weight. (Changes in temperature and relative humidity are insignificant for the times involved.) In order to eliminate this unwanted effect, the average barometric pressure during the 1-1/4 min needed to determine each  $\lambda_i$  is recorded. Then,  $\lambda_1, \lambda_3, \lambda_4$  and  $\lambda_5$  are corrected to the value they would have had at  $P_2$ , the pressure recorded for  $\lambda_2$ . The corrected value  $\lambda'_i$  is given by

$$\lambda'_i = \lambda_i + \frac{(P_i - P_2)(V_i - V_{cw})\rho}{P_2 S}$$

where  $V_{cw}$  is the volume of the balance counter-weight,  $V_i$  is the volume of the weight on the balance pan, and  $\rho$  is now the air density at the time of the second observation, and  $S$  is the balance sensitivity (see sec. 5.2.2).

The ratio of the balance arms is assumed equal to 1 to within 4%. When stainless steel and platinum weights are used in the same weighing design, the magnitude of  $\lambda'_i - \lambda_i$  can reach 15  $\mu\text{g}$ .

**5.2.2 Sensitivity:** Every subweighing includes an estimate of the current balance sensitivity,

$$S = \frac{m_s - \rho V_s}{-\lambda_2 + \lambda_3 + \lambda_4 - \lambda_5}.$$

In the course of a complete weighing design, which requires some 150 min, the sensitivity is seen to decrease by about 0.2%. This is true of both NBS-2 and V-1. The sensitivity recovers when the balance is fully arrested for several hours (i.e., knives and flats separated). The decrease in sensitivity was noted in the paper reporting the construction of NBS-2 [13]. At that time, the author speculated that elastic deformation of the metal knives could account for the observed behavior. In our view, this remains a viable hypothesis. The effective radii of the knives are  $\sim 20 \mu\text{m}$  which is the same order of magnitude as the distance between the center of curvature of the main knife and the center of mass of the balance beam. The latter distance largely determines balance sensitivity [19].

Since the decrease in balance sensitivity during a series is slow and monotonic, it has no serious consequences. The observed decrease in sensitivity simply underlines the necessity of using an algorithm which measures balance sensitivity for each subweighing.

**5.2.3 Bias:** In order to use a least squares approach to the solution of a weighing design, it is necessary that the r.h.s. of eq (4) be an unbiased estimator of  $m_A - m_B$ . For both NBS-2 and V-1, this assumption has been found to be violated. A detailed account of the problem as regards NBS-2 is found in [53].

For V-1 we have found that the usual subweighing involving any two weights A and B gives a biased estimate of  $m_A - m_B$ . That is, instead of eq (4), we find that

$$\delta_{A,B} \equiv \epsilon + m_A - m_B = \frac{\lambda_1 - \lambda_2 - \lambda_3 + \lambda_4}{-\lambda_2 + \lambda_3 + \lambda_4 - \lambda_5}. \quad (4)$$

The additional term,  $\epsilon$ , seems independent of which subweighing of the design is involved, although we cannot be definitive on this point. This is a somewhat different outcome from that described for NBS-2 in [53]. We have found that  $\epsilon$  is about +2.4  $\mu\text{g}$  with a mean standard deviation of 0.4  $\mu\text{g}$  (22 degrees of freedom).<sup>5</sup>

We have essentially adopted the approach taken by BIPM for dealing with this bias (see [53]). As the authors

<sup>5</sup> It may be of historical interest to point out that unexplained systematic behavior has been reported before for a two-pan balance [30]. In that case, a change in procedure removed the problem.

of this reference point out, although the bias cannot yet be satisfactorily explained, its effect can, nevertheless, be eliminated. This is done by duplicating all the subweighing in a design on the succeeding day, the only change being that all weighings are reversed:

Design for	Design for
Day 1	Day 2
A-B	B-A
A-C	C-A
A-D	D-A
B-C	C-B
B-D	D-B
C-D	D-C

Then for each pair of weights  $x$  and  $z$ ,

$$(\delta_{x,z} - \delta_{z,x})/2$$

is now an unbiased estimator for  $m_x - m_z$ . The quantity  $(\delta_{x,z} + \delta_{z,x})/2$  gives an estimate for  $\epsilon$ .<sup>6</sup>

Six weights were involved in the measurements at NBS—K20, K4, KA, K650, CH-1, and D2. Six weights were also involved in the measurements at BIPM—K9, K31, K20, K4, CH-1, and D2. These were measured in three separate groupings. In the case of NBS, the scheme shown in table 3 was carried out over a period of 8 days. Unlike NBS-2, balance V-1 can only accommodate four weights on its remote-controlled weight changer. Therefore, in the case of V-1, the balance needed to be opened after Days 2 and 5. After opening, the balance required a full day to re-equilibrate.

The design of table 4 provides 18 unbiased estimates of mass differences among pairings of the six weights. The least squares solution as well as the variance-covariance matrix are readily found for the restraint that the mass of K20 is known. These matrices are also shown in table 4.

<sup>6</sup> In [53], a design is advocated which requires rotation of the weight table between successive subweighings. We have not, for technical reasons, used the BIPM design but have, instead, rotated the weight table  $\pm 1/8$  revolution between consecutive subweighings involving the same weight for the first observations.

## 6. Necessary Corrections

### 6.1 Added Weights

Kilogram balances such as NBS-2 and V-1 have a very narrow on-scale range—typically 40 mg. There are reasons, however, why it is imprudent to operate the balance over its full range. First, the scale may be non-linear. This could occur, for instance, if one of the two knives had an effective radius which was a function of beam angle [21]. Second, the balance reading is more immune to vibrational noise if the beam is maintained nearly horizontal [19].

In order to minimize unwanted effects, the following steps are taken: 1) The amplitude of the beam swing is always adjusted to be the same for each measurement—in our case, 4 mg peak-to-peak, and 2) small weights are added to the kilogram artifacts as needed so that the four artifacts agree in apparent mass to within 1 mg. These precautions ensure that the five observations of a subweighing are all derived from roughly the same place on the knife-edge profiles.

In the case of comparing platinum-iridium prototypes to D2 and CH-1, as much as 95 mg must be added to the stainless steel artifacts to bring their apparent masses to within the 1-mg tolerance. Obviously the masses and volumes of the added weights must be known with sufficient accuracy (i.e., about  $1 \mu\text{g}$  and  $1 \text{mm}^3$ ) so that the final mass assignments are not compromised.

Calibration of the added weights is similar to the problem of calibration of the sensitivity weight and is similarly resolved. We must work down in a calibration chain from 1 kg to 1 mg and we must have a resulting inaccuracy to no worse than  $1 \mu\text{g}$  in 95 mg ( $11 \times 10^{-6}$ ). Thus at the outset we must know the value of the 1-kg standards to 0.011 g. This is no problem. As in the case of the sensitivity weight, the calibration accuracy is limited primarily by the ratio of precision-to-load of the balances used in the calibration chain. Typical results are shown in table 5. The uncertainties are correlated because of the weighing design which was used [17].

Table 3

Day 1	Day 2	Day 4	Day 5	Day 7	Day 8
K20,K4	K4,K20	KA,K650	K650,KA	CH-1,D2	D2,CH-1
K20,KA	KA,K20	KA,CH-1	CH-1,KA	CH-1,K20	K20,CH-1
K20,K650	K650,K20	KA,D2	D2,KA	CH-1,K4	K4,CH-1
K4,KA	KA,K4	K650,CH-1	CH-1,K650	D2,K20	K20,D2
K4,K650	K650,K4	K650,D2	D2,K650	K2,K4	K4,D2
KA,K650	K650,KA	CH-1,D2	CH-1,D2	K20,K4	K4,K20



**Table 4.** The matrix of observations, variance-covariance matrix, and transpose of the solution matrix [17]. The value for each observed difference is derived from a weighing and its opposed weighing as explained in section 5.1.3 and [53]. It is assumed that the mass of K20 is known.

Obs #	Matrix of observations					
	K20	K4	KA	K650	CH-1	D2
1	1	-1	0	0	0	0
2	1	0	-1	0	0	0
3	1	0	0	-1	0	0
4	0	1	-1	0	0	0
5	0	1	0	-1	0	0
6	0	0	1	-1	0	0
7	0	0	1	1	0	0
8	0	0	1	0	-1	0
9	0	0	1	0	0	-1
10	0	0	0	1	-1	0
11	0	0	0	1	0	-1
12	0	0	0	0	1	-1
13	0	0	0	0	1	-1
14	-1	0	0	0	1	0
15	0	-1	0	0	1	0
16	-1	0	0	0	0	1
17	0	-1	0	0	0	1
18	1	-1	0	0	0	0
Restraint	1	0	0	0	0	0
<i>Variance-covariance matrix:</i>						
	0	0	0	0	0	0
(1/10000)x	0	2500	1250	1250	1250	1250
	0	1250	2917	1667	1458	1458
	0	1250	1667	2917	1458	1458
	0	1250	1458	1458	2917	1667
	0	1250	1458	1458	2667	2917
<i>Transpose of solution matrix:</i>						
	0	-2500	-1250	-1250	-1250	-1250
	0	-1250	-2917	-1667	-1458	-1458
	0	-1250	-1667	-2917	-1458	-1458
	0	1250	-1667	-417	-208	-208
	0	1250	-417	-1167	-208	-208
	0	0	1250	-1250	0	0
	0	0	1458	208	-1458	-208
(1/10000)x	0	0	1458	208	-208	-1458
	0	0	208	1458	-1458	-208
	0	0	208	1458	-208	-1458
	0	0	0	0	1250	-1250
	0	0	0	1250	1250	-1250
	0	1250	1458	1458	2917	1667
	0	-1250	208	208	1667	417
	0	1250	1458	1458	1667	2917
	0	1250	208	208	417	1667
	0	-2500	-1250	-1250	-1250	-1250

## 6.2 Air Buoyancy

Additional weights were necessary to bring the apparent mass of a stainless steel kilogram standard into near equality with a platinum-iridium prototype, i.e.,

$$|P(K - [R + \Sigma W])| < 1 \text{ mg}$$

where K is made of platinum-iridium, R is made of

stainless steel and  $\Sigma W$  represents a summation of small weights whose mass is nominally 95 mg.

However, the quantity of interest is  $m_K - m_R$ :

$$m_K - m_R = P(K - [R + \Sigma W]) + m_{\Sigma W} - \rho(V_R - V_K) - \rho V_{\Sigma W} + f_1 + f_2. \quad (5)$$

The term  $\rho(V_R - V_K)$  is approximately equal to  $m_{\Sigma W}$ .

**Table 5.** Typical calibration values for a set of fractional weights. The uncertainties are given at a level of 1 standard deviation according to [31]. The uncertainty is virtually all statistical in nature (Type A). The results are correlated so that, for example, the uncertainty in the combination of (50 mg + 30 mg + 10 mg) is 0.27  $\mu\text{g}$ .

Nominal Mass (mg)	Correction (mg)	Uncertainty ( $\mu\text{g}$ )	Vol. at 20 °C ( $\text{cm}^3$ )	Coef. of Exp. ( $10^{-6}/^\circ\text{C}$ )
50	-0.02936	0.23	0.00301	20
30	-0.01361	0.19	0.00181	20
20	-0.02715	0.15	0.00740	69
10	-0.02037	0.18	0.00370	69
5	-0.01514	0.15	0.00185	69
3	-0.02223	0.16	0.00110	69
2	+0.00971	0.14	0.00074	69
1	+0.03106	0.18	0.00038	69

Clearly, if  $\rho$  were zero, neither the added mass  $m_{\text{sw}}$  nor the buoyancy correction  $-\rho(V_{\text{R}} - V_{\text{K}})$  would be necessary. Performing the required mass comparisons under vacuum conditions presents a number of as yet unstudied problems, however; mass stability and balance performance are chief among them. As a related problem, one would have to re-open the question of whether the SI definition of the kilogram ought to be the mass of the IPK in equilibrium with laboratory air or in vacuum. These concerns have potential interest but are beyond the scope of this report.

A brief word should be said about the reliance of the necessary buoyancy corrections on SI units of mass. The density  $D_x$  of a weight  $x$  is defined as  $D_x \equiv m_x/V_x$ . Quantities such as  $m_x - \rho V_x$  which occur in weighing equations can equally be written  $m_x(1 - \rho/D_x)$ . Thus, for instance, instead of estimating  $m_{\text{R}}$  by

$$m_{\text{R}} = -P(\text{K} - \text{R}) + m_{\text{K}} + \rho(V_{\text{R}} - V_{\text{K}})$$

we can, equally, write

$$m_{\text{R}} = \frac{-P(\text{K} - \text{R}) + m_{\text{K}}(1 - \rho/D_{\text{K}})}{1 - \rho/D_{\text{R}}}$$

In the latter formulation, it becomes clear that buoyancy corrections only involve density ratios ( $\rho/D_{\text{K}}$  and  $\rho/D_{\text{R}}$ ) and therefore are fundamentally independent of which consistent system of units is employed. One can, in fact, easily conceive of experiments to measure the densities of  $\rho$  and  $D$  relative to the density of some natural standard—distilled water, for instance. Such experiments usually involve weighing but the instrument used need only be a linear force transducer—the propor-

tionality constant between transducer reading and SI units need not be known. This approach relies on the suitability of water as a density reference for laboratories separated by large distances. Recently, renewed interest in water suggests that distilled tap water at a given temperature is probably uniform in major metrology laboratories to at least  $10 \times 10^{-6}$  [22].

Although improvements in establishing reference densities in SI units through solid objects have lowered uncertainties by an order of magnitude [23], this has been an iterative procedure which relied in part on the soundness of the mass unit. The mass unit was able to be established and disseminated through the national prototypes to the required precision because distilled water, for all its shortcomings, was a sufficiently good density reference.

**6.2.1 Densities of Mass Artifacts:** Since the density of artifact standards is stable, an error in the density (or volume) assignment to a mass standard propagates as an error which is systematic to all future measurements. The magnitude of the error incurred is  $(\rho_1 - \rho_2)\delta V_x$ , where  $\delta V_x$  is the error in the volume assigned to a mass artifact  $x$ ,  $\rho_1$  is the air density at the time of calibration of  $x$  and  $\rho_2$  is the air density of the time of subsequent use of  $x$ . This means that although the error in assignment of an SI mass value to  $x$  may be large (i.e.,  $\rho_1\delta V_x$ ),  $x$  can nevertheless be used without significant error to propagate SI mass values to other weights provided that the subsequent calibrations are carried out at an air density  $\rho_2$  sufficiently close to  $\rho_1$ .

**6.2.2 Density of Air:** Historically, the density of laboratory air has been determined from an equation of state involving temperature, pressure, and relative humidity. The equation was derived by making small corrections to the density of air under specified reference conditions near those normally encountered [24]. The reference density was determined relative to distilled water [25].

Recently, the equation of state has been rederived on a firmer theoretical basis. Also, refinements have been added—such as a correction for changes to the ambient levels of carbon dioxide [26]. This approach has won endorsement by the CIPM, which has recommended a specific formulation, “Formula for the Determination of the Density of Moist Air (1981),” for international work [27].

In the 1981 formula, the mass unit appears in the measurement of ambient laboratory pressure, i.e., pressure must be determined in SI units. Atmospheric pressure calibrations at NBS are, in turn, currently based on a controlled clearance piston gage [28]. Its use requires an air buoyancy correction for stainless steel weights. This correction is about 95 ppm. (Ignoring the air density correction would produce pressure measurements

systematically in error by 95 ppm. One iteration of the measurements, now including a buoyancy correction, would be sufficient to reduce pressure errors due to buoyancy to acceptable levels.)

The devices used to determine ambient temperature, pressure, humidity, and carbon dioxide conditions at NBS will now be described.

**6.2.2.1 Temperature:** Temperature was read using a quartz digital thermometer (Hewlett-Packard 2804A) with sensor mounted in the balance.<sup>7</sup> This device can resolve 0.1 mK and dissipates negligible power in the sensor, which is located in the balance at the approximate level of the weights being calibrated.

There are several drawbacks to this device which necessitated using it in the differential mode: First, since the device measures temperature by looking at changes in the resonant frequency of a crystal oscillator with respect to a reference oscillator, the stability of the reference is of primary concern. By using the reference oscillator supplied with the thermometer, one can achieve millikelvin accuracies only for differential measurements of two sensors which differ by less than 10 °C. In addition, the sensors are altered by mechanical and thermal shock and also are subject to hysteresis if cycled between temperature extremes.

Based on these concerns, we decided that all air measurements should be made differentially and that the sensors should always remain within a few degrees of the balance temperature. Thus one of the sensors must be held in a bath near room temperature. The bath temperature can then be determined by a calibrated platinum resistance thermometer. As a routine surveillance, both sensors are placed in the bath to make sure that their properties have not changed. They are also checked at room temperature against the platinum thermometer.

Since the temperature bath used plays a vital role in our knowledge of the temperature in the balance case, we will briefly describe its design and operating characteristics. We have used a simplified version of the air bath described by Cutkosky and Field [29]. The bath is constructed of aluminum alloy. The bath temperature is servo-controlled at about 25.25 °C using a circuit which has already been described [30]. There are two wells 20-cm deep which can receive either quartz thermometric sensors or a platinum resistance thermometer. The lowest 12-cm of the wells is actively controlled. The wells are filled with enough liquid to surround the volume of the inserted sensor. The sensor stems are sealed to the top of the well by a thermal short-circuit and insulated to 10 cm above the well entrance.

The liquid used is a fluorinated hydrocarbon which has poor but adequate thermal conductivity

<sup>7</sup> Brand names are given to specify experimental conditions. Such mention does not imply endorsement by the National Bureau of Standards.

( $0.7 \times 10^{-3} \text{ W/cm} \cdot ^\circ\text{C}$ ), a low vapor pressure (40 Pa), but is not viscous or oily. Since the room temperature is never more than 3 K below the bath temperature, stem corrections for the thermometers are negligible. By permuting temperature sensors between the two wells, it can be verified that any thermal gradient between wells is less than 0.5 mK.

The diurnal variations in room temperature affect the bath. A change in room temperature of 1 K produces a corresponding change in bath temperature of about 1 mK. This is considerably worse than the control reported in [29], the reason probably being the presence of the wells in the modified bath design. At any rate, the bath stability is adequate for our purposes.

The bath temperature is determined via a Myers-type platinum resistance thermometer placed in one well. This thermometer has been calibrated on the IPTS-68. From time to time its resistance at the triple-point of water is redetermined. This number is well behaved. The thermometer was read twice daily using a Leeds and Northrup G-2 Bridge. The error in balance temperature inferred by these means is thought to be about 5 mK.

A final problem with the quartz thermometer is its time constant. We have determined that, in still air, the time constant,  $\tau$ , of our quartz sensor is about 6 min. This means that if the sensor were placed in an environment in which the temperature changed linearly with time, the thermometer at time  $t$  would read the temperature which existed at time  $t - \tau$ . The thermal lag time imposes a limit on the acceptable rise in temperature within the balance case during the course of our measurements; i.e., the rate of change in temperature must be less than  $E/\tau$  where  $E$  is the acceptable temperature error. If  $E$  is limited to 3 mK, then the rate of temperature increase within the balance must be less than 0.5 mK/min. This criterion has been met throughout our measurements.

**6.2.2.2 Pressure:** Pressure was measured using a model DDR-6000 pressure-to-voltage transducer manufactured by Ruska Instruments, Houston, TX. A 5 1/2-digit voltmeter (Hewlett-Packard Model 3556A) was used to read the voltage. The pressure transducer was configured as a differential gage. When atmospheric pressure was required, the reference chamber of the gage was evacuated and a small correction was applied for any residual pressure on the reference side. The zero drift of the gage was monitored twice daily by allowing the reference pressure to come to the laboratory ambient.

A calibration history of the gage indicated that the calibration near 1 atmosphere was constant for months at a time. The zero drift required correction on a daily basis, however. Calibration of the device required car-

rying it about 30 m to facilities of the NBS Temperature and Pressure Division. Here the transducer was calibrated against a piston gage throughout the normal range of local atmospheric pressures. The effective area of the piston gage used was determined by calibration against a controlled-clearance piston gage which serves as primary standard near atmospheric pressure.

The primary standard has a calibration uncertainty of 3 Pa at atmospheric pressure. Therefore, the pressure-to-voltage transducer can, in principle, be calibrated to an accuracy of about 5 Pa ( $5 \times 10^{-5}$ ). This is a calibration uncertainty. The uncertainty in a single measurement, based on recommended BIPM practices for combining error [31] is approximately 3 Pa. (Unless otherwise stated, all uncertainties in this paper are given in accordance with [31] at an estimated level of one standard deviation.)

Because of the critical importance of pressure measurements and the possibility of the transducer changing its characteristics between calibrations, a check standard was used daily. This standard is a remarkable aneroid barometer—one of a handful of which were manufactured 40 years ago by the Wallace and Tiernan Company. The stability of this device is such that the long-term standard deviation of a measurement is 3.3 Pa. Thus the total uncertainty of a single measurement is about 5 Pa.

In the second series of measurements made at NBS (1984), the Wallace and Tiernan gage was used to calibrate our pressure-to-voltage transducer. The reason for this is that a lamp within the transducer failed and was replaced; so that the previous calibration—based on a piston-gage—was no longer valid. The pressure-to-voltage transducer developed further problems subsequent to all the 1984 mass measurements. We were therefore forced to rely on the Wallace and Tiernan gage as our pressure standard. The latter was completely recalibrated after its use by us and was found not to have changed throughout the pressure range of our measurements.

We should also mention that the test port of the pressure transducer was directly connected to the interior of the weighing chamber. The transducer must be placed at the same level as the balance pan or a head correction of  $1.4 \times 10^{-6}$ /cm must be applied.

**6.2.2.3 Relative Humidity:** Relative humidity,  $U$ , was measured using Hygrodynamics wide-range humidity sensors manufactured by the American Instrument Company. The sensors were calibrated by the manufacturer at the time of purchase. The calibration at time of purchase had been verified periodically by placing the sensors in the atmosphere above a standard salt solution. Agreement of two points along the calibration curve with fixed points determined by standard salt

solution was thought to be sufficient evidence that the humidity probes were not in need of recalibration.

However, when the probes were recalibrated in 1984 by NBS' Chemical Process Metrology Division, it became clear that our assumption had led to serious error—the probe used for our most critical mass measurements prior to sending our standards to the BIPM had changed its calibration by about 0.035 relative humidity units ( $\Delta U \sim 3.5\%$ ) in the ambient region. Its calibration at the relative humidity produced by the salt-solution fixed points had not changed, however.<sup>8</sup> We have, subsequent to this episode, required a 7-point calibration of each probe at approximately 6-month intervals when critical work is involved. We also compare two probes at ambient humidity to help ensure that important changes have not occurred between calibrations. In this way, we now believe the relative humidity in our balance case is known to 0.01 relative humidity units.

**6.2.2.4 Carbon Dioxide:** Carbon dioxide fraction in the balance ambient,  $X_{\text{CO}_2}$  is checked each time the case is opened. Measurements are made using an infrared detector (MIRAN 101, manufactured by the Foxboro Company). Calibration of the instrument is accomplished by using two cylinders of air. One cylinder contains  $490 \pm 5$  parts per million carbon dioxide; the second cylinder contains  $400 \pm 5$  parts per million carbon dioxide. A manifold allows the user to connect the detector to either gas cylinder or to the balance ambient. The first cylinder serves to calibrate the span of the detector while the second cylinder determines its sensitivity. The detector has a known nonlinearity so that the sensitivity measurement must be applied with caution. Nevertheless, it is apparent that the  $\text{CO}_2$  content of ambient air can be determined, using the above procedure, to 10%; i.e., the background ratio of  $\text{CO}_2$  in air can be known to within 50 ppm.

Typical ambient levels and uncertainties for the air-density data are summarized in table 6.

The equation-of-state itself has a component of uncertainty independent of  $T$ ,  $P$ ,  $U$  and  $X_{\text{CO}_2}$ . This is, however, difficult to assess. From estimates of Jones [26], one may infer that the uncertainty in the equation-of-state itself is about  $6 \times 10^{-5}$ . Recently, Balhorn has compared results obtained using the CIPM equation and direct measurements based on Archimedes' principle (the latter measurements involve vacuum weighing to find mass differences independent of a buoyancy correction). His expected experimental uncertainty is at the level of that estimated by Jones for uncertainties in the equation-of-state. Balhorn finds no unexpected results

<sup>8</sup> The author has since become aware of an identical experience with this type of sensor [32].

**Table 6.** The effect of uncertainties in ambient levels of temperature, pressure, relative humidity and CO<sub>2</sub> fraction at NBS on the calibration of mass of a stainless steel standard from a national prototype.

	Parameter	Ambient Air at NBS (typical values)		Mass Uncertainty
		Value	Uncertainty	
July 1984	<i>t</i>	22.170 °C	0.005 °C	2 μg
	<i>P</i>	752.02 mm Hg	0.038 mm Hg	5 μg
	<i>U</i>	41%	1.0%	9 μg
	<i>X</i> <sub>CO<sub>2</sub></sub>	440 ppm	50 ppm	1 μg
				RSS 11 μg
March 1983	<i>t</i>	22.253 °C	0.005 °C	2 μg
	<i>P</i>	751.929 mm Hg	0.030 mm Hg	4 μg
	<i>U</i>	25.2%	2%	17 μg
	<i>X</i> <sub>CO<sub>2</sub></sub>	430 ppm	150 ppm	6 μg
				RSS 19 μg

[33]. This is an important confirmation because Balhorn has measured air density by a buoyancy method. Similar results, but at somewhat increased uncertainty, have been obtained in other laboratories as well [34,35].

**6.2.3 Surface Area:** Recall that eq [3] contains a term  $f_1$  which is a function of the difference of surface between A and B. In general  $f_1$  will depend on relative humidity, surface composition and finish. Modern studies of the functional relationship for a variety of metals and surface finish have been carried out at The Physikalisch Technische Bundesanstalt (PTB) in the Federal Republic of Germany and The National Research Laboratory of Metrology (NRLM) in Japan [36,34]. The NRLM proposes the following empirical formula for the correction to be applied between a platinum-iridium kilogram and a stainless steel kilogram:

$$f_1 = -(0.01)S_K + (0.0092U - 0.20)(S_K - S_R) \quad (6)$$

where  $f_1$  is in micrograms and

$S_K$  = geometrical area of prototype in cm<sup>2</sup>

$S_R$  = geometrical area of stainless-steel weight in cm<sup>2</sup>

$U$  = relative humidity in percent.

The linear functional dependence of  $U$  is claimed valid for  $20\% < U < 70\%$ . Equation (6) corrects measurements done at atmospheric pressures and at relative humidity  $U$  to vacuum conditions. It would be useful to obtain confirmation of eq (6) for different prototypes and for stainless weights of different alloys and finishes. Measurements reported by BIPM, for instance, appear to be inconsistent with eq (6) [32]. The PTB results suggest that the magnitude of  $df_1/dU$  may be overestimated by eq (6) in some instances—perhaps by a factor of three [36].

To correct results to a reference humidity  $U_0$  within the range of validity of eq (6),

$$f_1 = 0.0092(U - U_0)(S_K - S_R).$$

For D2 and CH-1,  $S_K - S_R$  is about  $-62 \text{ cm}^2$  and  $-72 \text{ cm}^2$ , respectively. Thus,  $df_1/dU = -0.57 \mu\text{g}/\%RH$  for D2 and  $f_1 = -0.66 \mu\text{g}/\%RH$  for CH-1.

**6.2.4 Difference in Centers of Mass:** Equation (3) contains a function  $f_2$  to correct for the difference in gravitational acceleration between the centers of mass of weights K and R. If R is either stainless steel weight, D2 or CH-1, the elevation differences of the centers of mass of R and K is about 1 cm.

A naive calculation which assumes the laboratory is on the surface of a homogeneous and spherical earth leads to a gravitational gradient of  $-3.14 \times 10^{-9}/\text{cm}$  or a correction of  $3.1 \mu\text{g}$  [37]. Most mass laboratories are located below ground level, however. For such conditions, the magnitude of the gravitational gradient might be reduced by as much as 50% [38].

The correction for gravitational gradient was studied very early in the history of BIPM [39]. In a series of experiments remarkable for their attention to systematic sources of uncertainty, the gradient at BIPM was inferred from weighing results of obtained on a two-pan balance whose pans differed in elevation by 11 m. One pan was above ground level while the other was below. The results were in good agreement with theoretical calculations which took into account local inhomogeneities. The experimental result was:  $g^{-1}(dg/dz) = -2.5 \times 10^{-9}/\text{cm}$ .

While no such experiments have been done in our mass laboratory due to the relative unimportance of this correction, we believe that the results of [39] suggest that an assumed gradient value of  $-2.5 \times 10^{-9}/\text{cm}$  is reasonable in our laboratory as well. Should this correction prove in the future to be important relative to the overall uncertainty, a measurement of the gradient would, of course, become essential. (As a matter of completeness, we point out that the gravitational gradient may also be a necessary correction even in less precise weighing experiments. An example is hydrostatic weighing where

the object in the hydrostatic bath may be 50 cm below the balance weights [40]. This, in fact, was the motivation for the early BIPM experiments.)

## 7. Results at BIPM

G. Girard, who performed the calibrations at BIPM, has summarized his results as follows:

In practice the comparisons were carried out in the following manner:

### 3 weighings:

K32	D2	CH-1	K4				
		CH-1	K4	K20	K9		
				K20	K9	K31	D2

Design of Weighing #1		Design of the Opposed Weighing	
K31	D2	D2	K31
D2	CH-1	CH-1	D2
CH-1	K4	K4	CH-1
K4	K31	K31	K4
K31	CH-1	CH-1	K31
D2	K4	K4	D2

The mass value of the sensitivity weight used was last determined in September–October, 1983.

The four NBS standards were compared to the two platinum-iridium standards of BIPM, first in the state in which they arrived at BIPM, then after a cleaning in benzene and alcohol; prototypes K4 and K20 of platinum-iridium were, in addition, washed under a steam jet of doubly-distilled water.

These comparisons took place from January 30 to February 7, 1984 (group a, before cleaning-washing) and from February 23 to March 1, 1984 (group b, after cleaning-washing).

**Correction for Air Buoyancy:** The density of moist air, in the course of each weighing, was calculated with the aid of the "Formula for the determination of the density of moist air (1981)" [24].

The different ambient parameters and the instruments used for their determination in the course of the two groups of measurements are the following:

**Temperature:** Rosemount platinum resistance thermometer and Automatic Systems Laboratories F17A alternating current bridge. The readings of this thermometer were verified, before the start of the measurements, by comparison to a Tinsley platinum resistance thermometer calibrated in the IPTS-68.

Uncertainty  $\approx 2$  mK

Mean temperatures (a) 20.27 °C  
(b) 20.24 °C

**Pressure:** Ruska DDR-6000 calibrated by comparison with the BIPM interferential manobarometer before and after each group of measurements.

Uncertainty  $\approx 2$  Pa

Average pressure (a) 100,922 Pa  
(b) 101,015 Pa

**Relative Humidity:** Hygrodynamics transducer calibrated against an EG&G dewpoint hygrometer before and after the comparisons, at the ambient relative humidity.

Uncertainty  $\approx 1\%$  to  $2\%$

Mean relative humidities (a) 41%  
(b) 38%

**CO<sub>2</sub>:** Ambient air near the balance is brought to an MSA analyser. This device is based on infrared absorption; the mole fraction of CO<sub>2</sub> in the air is measured before closing up the balance.

Uncertainty  $\approx 0.0001$

Mean mole fractions (a) 0.00041  
(b) 0.00062

The mean values of air density in the two groups are, therefore,

(a) 1.194 40 mg/cm<sup>3</sup>  
(b) 1.195 91 mg/cm<sup>3</sup>

After making corrections for gravitational gradient, the following results were obtained:

	Before Cleaning		After Cleaning
K4	1 kg - 0.075 mg	K4	1 kg - 0.106 mg
K20	-0.001 mg	K20	-0.022 mg
CH-1	-0.377 mg	CH-1	-0.384 mg
D2	+13.453 mg	D2	+13.447 mg

The estimated standard deviation of each result is 1.2  $\mu$ g.

The estimated standard deviation of each result is 1.3  $\mu$ g.

Kilogram K650 was remeasured at BIPM in February 1985. The results are: mass = 1 kg - 2.264 mg;  $\sigma = 0.5$   $\mu$ g [52].

## 8. Results at NBS

The 1984 results at NBS, carried out after the BIPM recalibration, will be presented in detail. The 1982 measurements will be summarized.

### 8.1 1984 NBS Measurements

Upon return to NBS, K20 and K4 were used in some preliminary measurements with KA and K650. K20 and K4 were then cleaned in benzene and ethanol after which they were washed in a vapor jet of distilled water. After cleaning, they were again compared with KA and K650. The results clearly show that K20 was un-

changed by the cleaning whereas K4 lost about 4  $\mu\text{g}$ . The difference, if significant, could be attributed to the traveling container of K4 which required the prototype to be entirely wrapped in paper backed with chamois. KA and K650 were not cleaned for these measurements.

A set of 18 symmetrized observations was then made using the six weights as described in section 4 above. The results are shown in table 7.

Observation 12 appears to be an outlier. Its deviation from the least squares fit and its deviation from the average value of  $\epsilon$  are the maxima of the set. The data were recomputed eliminating observation 12 with the following results: The standard deviation of the fit was reduced from 3.9  $\mu\text{g}$  to 3.2  $\mu\text{g}$ ; the fitted values for the masses of K4, KA and K650 were unchanged (this could have been predicted from examining the solution matrix); the values of CH-1 and D2 were changed by only 1.2  $\mu\text{g}$ .

We have chosen not to eliminate observation 12. This decision has a very small impact on the inferences we

will draw from the calibration exercise.

At this point, CH-1 and D2 were cleaned by vapor degreasing. Observations 13 through 18 were then repeated after which the new results were compared with the original observations.

Assuming K20 and K4 to be invariant during these weighings, the results may be interpreted as CH-1 having lost 16.5  $\mu\text{g}$  and D2 having lost 19.3  $\mu\text{g}$  as a result of the cleaning. The relative humidity throughout the measurements was essentially unchanged.

Thus the "as-cleaned" values for CH-1 and D2 are:

		Estimated SD
CH-1	1 kg -0.3887	4.8 $\mu\text{g}$
D2	+13.4516	4.8 $\mu\text{g}$

## 8.2 1983 NBS Measurements

We should also make use of the measurements taken at NBS in 1983—before the artifacts were sent to BIPM.

Table 7. Results of measurements made using the design of table 3. The ambient conditions given are the average found in two successive days as explained in section 5.1.3. The standard deviations are derived exclusively from the least squares fitting. Results for CH-1 and D2 shown in parentheses were found after cleaning as described in the text.

Obs. #	Mass Diff. (Not Fitted)	Deviation from LS fitted value	$\epsilon$	Average Ambient Conditions
1	0.0846 mg	4.0 $\mu\text{g}$	2.7 $\mu\text{g}$	$P = 753.17$ mm Hg
2	4.8467	2.0	2.0	$t = 22.16$ °C
3	2.2468	-0.1	1.5	$U = 40.9\%$
4	4.7615	-2.6	1.0	$X_{\text{CO}_2} = 430$ ppm
5	2.1691	3.0	4.2	$\rho = 1.18008$ mg/cm <sup>3</sup>
6	-2.5967	1.3	0.3	
7	-2.6009	-2.9	0.4	$P = 749.82$ mm Hg
8	-4.4882	3.8	1.9	$t = 22.30$ °C
9	-18.3376	-2.5	2.8	$U = 40.8\%$
10	-1.8955	-1.5	2.5	$X_{\text{CO}_2} = 400$ ppm
11	-15.7344	2.7	-3.0	$\rho = 1.17418$ mg/cm <sup>3</sup>
12	-13.8355	7.6	10.2	
13	-13.8444	-1.3	1.9	$P = 754.76$ mm Hg
14	-0.3559	-3.2	2.9	$t = 22.37$ °C
15	-0.2731	-1.0	-0.4	$U = 41.2\%$
16	13.4963	5.9	3.7	$X_{\text{CO}_2} = 440$ ppm
17	13.5618	0.8	6.2	$\rho = 1.18164$ mg/cm <sup>3</sup>
18	0.0772	-3.4	4.5	
Sum Squares:		193.8 $\mu\text{g}^2$		
Std. Dev. of Fit:		3.9 $\mu\text{g}$ (13 degrees of freedom)		
Artifact	LS Fitted Values	Estimated Std. Dev.		
K20	1 kg -0.022 mg	restraint		
K4	-0.1026	2.0 $\mu\text{g}$		
KA	-4.8667	2.1		
K650	-2.2687	2.1		
CH-1	-0.3722	2.1	( -0.3887 mg	4.8 $\mu\text{g}$ )
D2	+13.4709	2.1	(+13.4516	4.8)

These did not use the symmetrized algorithm shown in section 5.1.3. There are other differences as well. The most important of these is that, in 1983, the balance pan was loaded and all knife-flat bearings were engaged for 18 h prior to the start of measurements. In the BIPM measurements and in 1984 NBS measurements duplicating the BIPM algorithm, the balance is fully arrested until the time of use. At BIPM, the balance is "warmed up" by making three observations prior to the start of data taking. At NBS, warm-up consists of a practice observation of each of the four weights in the design.

Four designs were run using K20, K4, KA and K650. All the weights were rinsed in benzene, rinsed in ethanol, and washed with steam prior to the first and third measurements. The pooled standard deviation for these four runs was  $4.5 \mu\text{g}$  (12 degrees of freedom). The mass of K20 given in the 1948 certificate was used to restrain the least squares solution. We therefore expect that the standard deviation of the derived masses of K4, KA and K650 will be  $\sqrt{1/2} \cdot 4.5 \mu\text{g} = 3.2 \mu\text{g}$ . The pooled standard deviation of the six duplicate measurements made without cleaning is  $2.9 \mu\text{g}$ . This agrees well with the least squares value and therefore does not indicate a "between times" component to the standard deviation [18].

One can also examine the results for changes which may have occurred upon cleaning. This is done by looking at the average values for K4, KA and K650 for measurements 1 and 2 as compared with measurements 3 and 4:

	K4	KA	K650
Average Difference	$-8.7 \mu\text{g}$	$-1.9 \mu\text{g}$	$-9.4 \mu\text{g}$
Expected SD of Difference	$4.5 \mu\text{g}$		

These results suggest that the cleaning procedures used may produce perceptible changes in the relative masses of platinum weights. Two rounds of measurements were carried using K20, K4 and various lower-density weights. K20 and K4 were recleaned before each round and used in five designs. The average differences for the two additional rounds as compared to the average of the first two measurements (1 and 2) is:

	(K4)	(K4)'
Average Differences	$-3.8 \mu\text{g}$	$-4.0 \mu\text{g}$
Expected SD of Difference	$2.7 \mu\text{g}$	

These last results are consistent with the *average* value of the first four measurements. Since the author had no experience cleaning kilogram prototypes, the

progression of values shown above may represent a learning curve. The metrologist with the most experience in cleaning platinum-iridium prototypes is undoubtedly G. Girard of BIPM. Using NBS-2, he has demonstrated a cleaning stability of approximately  $2 \mu\text{g}$ . (The data presented are for K63—measurements involved three complete cleaning cycles during a period of 8 months [32].)

The 1984 results at NBS may be compared with the 1983 results as shown in table 8a. We have assumed that KA and K650 have not changed in mass over the 16-month period. The difference in mass seen for K20 and K4 is consistent with what was observed at BIPM upon cleaning. We can, therefore, conclude that the cleaning administered to K4 and K20 at BIPM in 1984 was more thorough than that performed at NBS in 1983. The  $6\text{-}\mu\text{g}$  discrepancy between the cleaning loss measured at BIPM and that inferred from measurements at NBS can be explained in at least three ways:

1) The  $6\text{-}\mu\text{g}$  difference is barely significant given the combined standard deviations of the NBS ( $4.2 \mu\text{g}$ ) and BIPM ( $1.4 \mu\text{g}$ ) results.

2) KA and K650 may have picked up surface dirt during the intervening 16 months—about  $2 \mu\text{g}/\text{year}$  seems typical for weights stored in a similar fashion [5].

3) The measurement scheme used in 1983 may give systematic differences from that used in 1984. To check this possibility, we performed several additional measurements involving K20, K4, KA and K650 using the 1983 measurement scheme. The results are shown in table 8b.

The comparisons of mass values found for CH-1 and D2 at NBS before and after the BIPM calibrations show closure to an average of  $11.5 \mu\text{g}$ . This number does not include a correction to the same value of relative humidity for both sets of measurements—such a "correction" would worsen agreement by an additional  $10 \mu\text{g}$ . The 1983 measurements at NBS were corrected for the recalibration of our humidity sensors in 1984 (see above). Also, in 1983 there were no background measurements of  $\text{CO}_2$  fraction in the balance air. We estimate that these shortcomings increase the uncertainty of buoyancy corrections for the 1983 measurements by  $8 \mu\text{g}$  over the 1984 NBS uncertainties.

We should add that D2 was measured against K20 in 1970. The measurements were made on NBS-2 at NBS. The 1948 calibration value of K20 was assumed. The results were:

$$\text{Mass of D2 (1970)} = 1 \text{ kg} + 13.521 \text{ mg} (6 \mu\text{g})$$

where the  $6 \mu\text{g}$  is the standard deviation of the mean of three measurements. The measurements were made at a relative humidity of about 21%. The 1970 data have



**Table 8.** In A, one sees that KA and K650 give self-consistent results. One may infer that K20 and K4 have lost mass after leaving NBS in 1983. In B, one may note the effect of a change in measurement algorithm. The conclusions found in A are still intact, but the quantitative mass losses now agree better with cleaning losses observed at BIPM. The tabulated numbers are all corrections, in milligrams, to 1 kg. Least squares fitted values of the 1984 measurements have been used.

Values found at NBS in 1983				
K20	1 kg - 0.019 mg (1948 BIPM certificate)			
K4	1 kg - 0.090 mg			
KA	1 kg - 4.891 mg			
K650	1 kg - 2.292 mg			

**A. 1984 NBS values of K20, K4, KA, and K650 based on 1983 NBS values for KA and K650 (corrections in milligrams to 1 kilogram)**

	K20	K4	KA	K650
KA	-0.045	-0.128	-4.892	-2.293
K650	<u>-0.045</u>	<u>-0.127</u>	<u>-4.890</u>	<u>-2.292</u>
Ave.	<u>-0.045</u>	<u>-0.128</u>	<u>-4.890</u>	<u>-2.292</u>
1983 value:	<u>-0.019</u>	<u>-0.090</u>	<u>-4.891</u>	<u>-2.292</u>
Δ :	-0.026	-0.038	0.001	-0.000

**B. Same as A except 1983 measurement algorithm used in 1984**

KA	-0.040	-0.124	-4.891	-2.298
K650	<u>-0.035</u>	<u>-0.118</u>	<u>-4.886</u>	<u>-2.292</u>
Ave.	<u>-0.038</u>	<u>-0.121</u>	<u>-4.888</u>	<u>-2.295</u>
1983 value:	<u>-0.029</u>	<u>-0.090</u>	<u>-4.891</u>	<u>-2.292</u>
Δ :	-0.029	-0.031	0.003	-0.003

been recalibrated using the 1981 equation of state for moist air [27]. We estimate that the calibrations of the barometer, thermometer, and humidity sensors used at the time, as well as the assumption of a background level of CO<sub>2</sub>, introduce an additional uncertainty of about 16 μg.

Comparison of the 1970 and 1984 values suggest that D2 has lost a considerable amount of mass in the last 15 years. The loss could be due to wear because D2 has been used extensively.

A comparison of the 1984 NBS results with the 1984 BIPM calibration is shown in table 9a. Since all measurements were made at nearly the same relative humidity, no correction is necessary. There is reasonable self-consistency among K20, K4, and CH-1, but D2, appears to have gained mass. Recall, however, that we are looking at values of the stainless steel weights after they were simply dusted upon removal from their packaging. The results inferred from measurements made after the stainless steel weights were cleaned by vapor degreasing (see Appendix) are shown in table 9b. These now appear consistent with BIPM measurements (but are now less consistent with the 1983 measurements at NBS).

**Table 9.** In A, one sees that K20 and K4 give self-consistent results but the results are inconsistent with results based on CH-1 and D2. After cleaning, B shows better self-consistency and also good consistency with BIPM results. The tabulated numbers are all corrections, in milligrams, to 1 kg. Least squares fitted values of the 1984 measurements have been used.

Values found at BIPM in 1984				
K20	1 kg - 0.022 mg			
K4	1 kg - 0.106 mg			
CH-1	1 kg - 0.384 mg			
D2	1 kg + 13.447 mg			

**A. 1984 NBS values of K20, K4, CH-1, and D2 based on 1984 BIPM values**

	K20	K4	CH-1	D2
K20	-0.022	-0.103	-0.372	13.471
K4	-0.025	-0.106	-0.376	13.467
CH-1	-0.034	-0.114	-0.384	13.459
D2	-0.046	-0.126	-0.396	13.447

**B. Same as A except CH-1 and D2 cleaned prior to measurements**

	K20	K4	CH-1	D2
K20	-0.022	-0.103	-0.389	13.452
K4	-0.025	-0.106	-0.392	13.448
CH-1	-0.017	-0.098	-0.384	13.456
D2	<u>-0.027</u>	<u>-0.107</u>	<u>-0.393</u>	<u>13.447</u>
Ave.	<u>-0.023</u>	<u>-0.104</u>	<u>-0.390</u>	<u>13.451</u>
BIPM values:	<u>-0.022</u>	<u>-0.106</u>	<u>-0.384</u>	<u>13.447</u>
Δ :	-0.001	-0.002	0.006	-0.004

These results can be interpreted in so many different ways that we are forced to expand our uncertainty accordingly. Thus it appears that long-term measurements of platinum-iridium artifacts based on K20 can be stable to 10 μg provided that the artifact is vigorously cleaned before use, according to the BIPM method. Mass values can be supplied to stainless steel weights with an uncertainty of about 30 μg. This includes all known sources of uncertainty as well as an additional "between times" component.

We should point out, however, that other laboratories have reported somewhat better results for comparisons of platinum prototypes with stainless steel weights [41]. These results, although based on a data sample too small to be definitive, strongly suggest that the 30-μg uncertainty on stainless steel weights can be significantly reduced once the influencing effects are better understood.

Because of the difficulty of assigning an SI mass value to a stainless steel weight, it has been suggested that a practical mass system based on stainless steel or nickel-chrome artifacts be adopted. Indeed, this is essentially the approach that has been taken by NBS [42]. The danger in this approach is that the long-term stability of

the artifacts is not assured to the same extent as is the stability of the national prototypes.

## 9. Recommendations

Several steps must be taken in order to improve both our ability to make reproducible mass measurements and our prospects for understanding the effects of influencing parameters:

1) Since virtually all secondary mass standards are now made of stainless steel, it is desirable for NBS to use stainless steel working standards for routine calibrations. The stability of these standards however, must be determined—not assumed. This will necessitate periodic measurements against K20 which, in turn, leads to recommendations 2) and 3).

2) A balance (preferably automated) must be made available which has a standard deviation of 1  $\mu\text{g}$  or better. In all probability, the time required to complete a weighing design will always be measured in hours. This has more to do with transporting weights to the pan without creating air turbulence than with the speed of operation of the balance itself. By having a device which measures to a standard deviation of 1  $\mu\text{g}$  instead of the present 4.5  $\mu\text{g}$ , a single set of measurements would yield mass values with the same standard deviation as the average of 20 sets obtained using the V-1 balance. Since the mass artifacts are mutable, a balance with low standard deviation is the only practical method of acquiring necessary data with an acceptable number of weighing operations. The best mass standards seem stable enough to warrant this level of precision.

3) The balance should be enclosed in a hermetically sealed chamber. This is probably prerequisite to achieving the desired standard deviation because changes in room air pressure would otherwise lead to air currents in the balance. In any event, it is very desirable that the density of air remain constant during a complete weighing design. Such a chamber would also permit research on other possible weighing atmospheres—including reduced air densities (i.e., partial vacuum). Thus the potentially large correction for relative humidity suggested by NRLM could be verified.

4) A cleaner environment for storing and using the weights should be considered. Cleaning is a major source of instability in mass standards. Technology exists to maintain weights in environments which have greatly reduced levels of dust and hydrocarbon vapors from those of our present facility.

The author is greatly indebted to G. Girard of BIPM for performing the calibrations at his laboratory. Useful discussions at BIPM with T. Quinn, G. Girard, and P. Carré are also acknowledged. He is also indebted to R. D. Cutkosky for several air bath design suggestions and to Lai Lee for technical assistance.

## References

- [1] Goldman, D. T., and R. J. Bell, eds., The International System of Units (SI): Natl. Bur. Stand. (U.S.) Spec. Publ. 330 (1981 December).
- [2] Page, C. H., and P. Vigoureux, eds. The International Bureau of Weights and Measures 1875–1975 (translation of the BIPM centennial volume). Natl. Bur. Stand. (U.S.) Spec. Publ. 420: Chapter 2, 43–63 (1975 May).
- [3] Girard, G. Masses. CIPM Procès-Verbaux 42: 45–48 (1973).
- [4] Bonhoure, A. Note sur l'efficacité de quelques procédés de nettoyage des poids en platine iridié, Appendix 3. CIPM Procès-Verbaux 20: 171–178 (1947).
- [5] Page, C. H., and P. Vigoureux, eds. Natl. Bur. Stand. (U.S.) Spec. Publ. 420; *op. cit.*; caption to figure 2; 49.
- [6] Benoit, J.-R.; Rapport à la Conférence sur les nouveaux prototypes métriques, 2me partie. Kilogrammes prototypes. Travaux et Mémoires du BIPM 7: 5–112 (1890).
- [7] Thiesen, M. Kilogrammes prototypes-1re partie. Travaux et Mémoires du BIPM 8: 3–71 (1893).
- [8] Laporte, H. Etude sur l'état de surface du kilogramme prototype no. 22, Appendix 5. CIPM Procès-Verbaux 28: 141–146 (1960).
- [9] Thiesen, M. Kilogrammes prototypes-2me partie. Travaux et Mémoires du BIPM 9: 3–51 (1898).
- [10] Cochrane, R. C. Measures for Progress, a history of the National Bureau of Standards. Natl. Bur. Stand. (U.S.) MP 275; Appendix A, 515–526 (1966).
- [11] Bonhoure, A. Kilogrammes prototypes. Travaux et Mémoires du BIPM 22 (3): C1–C82 (1966).
- [12] Girard, G. Masses. CIPM Procès-Verbaux 49: 38–43 (1981).
- [13] Almer, H. E. One kilogram balance. NBS No. 2. J. Res. Natl. Bur. Stand. (U.S.) 76C (1 and 2): 1–10 (1972 January–June).
- [14] Bowman, H. A., and L. B. Macurdy. Gimbal device to minimize the effects of off-center loading on balance pans. J. Res. Natl. Bur. Stand. (U.S.) 64C (4): 277–279 (1960 October–December).
- [15] Bowman, H. A., and H. E. Almer. Minimization of the arrestment error in one-pan, two-knife systems. J. Res. Natl. Bur. Stand. (U.S.) 64C (3): 227–235 (1963 July–September).
- [16] Bignell, N. Estimation of the centre of swing of an underdamped balance. Metrologia 19: 127–128 (1983).
- [17] Cameron, J. M.; M. C. Croarkin and R. C. Raybold. Designs for the calibration of standards of mass. Natl. Bur. Stand. (U.S.)

- [18] Croarkin, M. C. Measurement Assurance Programs Part II: Development and implementation. *Natl. Bur. Stand. (U.S.) Spec. Publ. 676-II* (1984 April).
- [19] Cage, M. E., and R. S. Davis. An analysis of read-out perturbations seen on an analytical balance with a swinging pan. *J. Res. Natl. Bur. Stand. (U.S.)* **87** (1): 23-45 (1982 January-February).
- [20] Thiesen, M. *Travaux et Mémoires du BIPM* **8**: 24 *op. cit.*
- [21] Corwin, A. H. Weighing, Chapter in *Technique of Organic Chemistry, Part 1, Physical methods of organic chemistry, Vol. 1*, 71-129; A. Weissberger, ed. New York: Interscience (1957).
- [22] Menaché, M.; C. Beauverger and G. Girard. Etude de la variation de la masse volumique de l'eau (eau pure et eau de mer) en fonction de la composition isotopique et de la teneur en gaz atmosphériques dissous. *Annales Hydrographiques, 5th series* **6** (3): 37-76 (1978).
- [23] Bowman, H. A.; R. M. Schoonover and C. L. Carroll. A density scale based on solid objects. *J. Res. Natl. Bur. Stand. (U.S.)* **78A** (1): 13-40 (1974 January-February).
- [24] Broch, O. J. Poids du litre d'air atmosphérique. *Travaux et Mémoires du BIPM* **1**: A52-A57 (1881).
- [25] e.g., Lord Rayleigh. On the densities of the principal gases. *Proc. Roy. Soc.* **53**: 134-139 (1883).
- [26] Jones, F. E. The air density equation and the transfer of the mass unit. *J. Res. Natl. Bur. Stand. (U.S.)* **83** (5): 419-428 (1978 September-October).
- [27] Giacomo, P. Formula for the determination of the density of moist air (1981). *Metrologia* **18**: 33-40 (1982).
- [28] Heydemann, P. L. M., and B. E. Welch. Piston Gages, chapter 4, 147-202 in *Experimental Thermodynamics, Vol. II, Experimental Thermodynamics on Non-reacting Fluids*, B. LeNeindre and B. Vodar, eds., London: Butterworths (1975).
- [29] Cutkosky, R. D., and B. F. Field. Standard cell enclosure with 20- $\mu$ K stability. *IEEE Trans. Instrum. and Meas.* **IM-23**: 295-298 (1974).
- [30] Cutkosky, R. D., and R. S. Davis. Simple control circuit for temperature regulation and other bridge applications. *Rev. Sci. Instrum.* **52**(9): 1403-1405 (1981 September).
- [31] Giacomo, P. News from the BIPM-Meeting of experts on the subject of expression of uncertainties. *Metrologia* **17**: 73-74 (1981).
- [32] Girard, G. *Masses. CIPM Procès-Verbaux* **44**: 40-43 (1976).
- [33] Balhorn, R. Berücksichtigung der Luftdichte durch Wägung beim Massevergleich. Kochsiek, M., ed. *Massebestimmung hoher Genauigkeit. PTB(F.R.G.) Bericht Me-60*; 65-77 (1984 June).
- [34] Kobayashi, Y. On a more precise correction for buoyancy and gas adsorption in mass measurement. Taylor, B. N.; Phillips, W. D., eds. *Precision measurement and fundamental constants II*; 1981, June 8-12; Gaithersburg, MD. *Natl. Bur. Stand. (U.S.) Spec. Publ. 617*; 441-443 (1984 August).
- [35] Koch, W. F., R. S. Davis and V. E. Bower. Direct determination of air density in a balance through artifacts characterized in an evacuated weighing chamber. *J. Res. Natl. Bur. Stand. (U.S.)* **83**(5): 407-413 (1978 September-October).
- [36] Kochsiek, M. Measurement of water adsorption layers on metal surfaces. *Metrologia* **18**: 153-159 (1982).
- [37] Almer, H. E., and H. F. Swift. Gravitational configuration effect upon precision mass measurements. *Rev. Sci. Instrum.* **46** (9): 1174-1176 (1975 September).
- [38] J. Faller, *Natl. Bur. Stand. (U.S.)/JILA*. private communication.
- [39] Thiesen, M. Détermination de la variation de la pesanteur avec la hauteur au Pavillon de Breteuil. *Travaux et Mémoires du BIPM* **7**: 1-32 (1890).
- [40] Bowman, H. A., and R. M. Schoonover, with appendix by M. W. Jones. Procedures for high precision density determinations by hydrostatic weighing. *J. Res. Natl. Bur. Stand. (U.S.)* **71C** (3): 179-198 (1967 July-August).
- [41] Girard, G.; Y. Kobayashi, K. Uchikawa, G. A. Bell, and J. W. Humphries. Rapport sur la comparaison d'étalons de masse de 1 kg en platine iridié et en acier inoxydable entre le Bureau International des Poids et Mesures, le National Research Laboratory of Metrology (Japon) et le CSIRO, Division of Applied Physics (Australie). *Rapport BIPM-82/3. Recueil de Travaux du BIPM* **8**: paper No. 19 (1981-1982).
- [42] Standard spec. for lab. weights and precision mass standards. ASTM designation E617-81. *1984 Annual book of ASTM standards, Vol. 14.02*. 458-477 Philadelphia, ASTM (1984).
- [43] Sears, M. J. E. in *Comptes rendus de la 9me CGPM. Travaux et Mémoires du BIPM* **21**: 32 (1952).
- [44] e.g., Benoit, M. in *Comptes rendus de la 5me CGPM. Travaux et Mémoires du BIPM* **16**: 31 (1917).
- [45] Kochsiek, M., rapporteur. Comité Consultatif pour la Masse et les Grandeurs Apparentées, session de 1981. Sèvres: BIPM (1983).
- [46] Pérard, A. *Masses. CIPM Procès-Verbaux* **21**: 27-31 (1948).
- [47] Bonhoure, A. Secondary standards of mass. *Microtecnica* **6** (3): 151-154 (1952).
- [48] Bowman, H.A.; R. M. Schoonover and C. L. Carroll. Reevaluation of the densities of the four NBS silicon crystal standards. *Natl. Bur. Stand. (U.S.) NBSIR* **75-768** (1975 August).
- [49] Takehashi, T.; Y. Kobayashi and T. Mizushima. On the calibration of the set of the secondary standard weights No. 1. *Bull. of the NRLM (Japan)* **9**: 6-10 (1964).
- [50] IMGC, L'Activité de l'IMGC dans le domaine des masses. Document CCM/81-32, Comité Consultatif pour la Masse et les Grandeurs Apparentées.
- [51] Plassa, M.; M. Tolomelli and A. Torino. Chemical properties of alloys for mass standards. Preprints of the 10th Conf. of IMEKO TC-3 on measurement of force and mass; 1984, September 11-14; Kobe, Japan. preprint 5-1.
- [52] G. Girard, Private communication.
- [53] Carré, P. and R. S. Davis, Note on weighings made with the NBS-2 balance, submitted to *J. Res. Natl. Bur. Stand. (U.S.)*.

# APPENDIX

## Cleaning and Long-Term Stability of Prototypes and Secondary Standards

### Platinum-Iridium Prototypes

Since the calibration of K4 and K20 is infrequent, it is important to know as much as possible about the long-term stability of these prototypes. The question inevitably involves cleaning, because airborne pollutants inevitably increase the mass of the prototypes at an unknown rate. This rate would, presumably, be reduced if storage in a cleaner environment could be arranged. The cleaning method must return the surface of the prototype to an original condition.

After manufacture, the prototypes were cleaned with alcohol and water vapor before calibration by BIPM metrologists [6]. In the years after, prototypes returned to BIPM for calibration were wiped clean with solvent-soaked chamois skin [11]. In 1946, A. Bonhoure systematically studied various cleaning procedures and reached the following important conclusions:

- Steam cleaning produces a more reproducible mass than any other method.
- Steam cleaning alone is insufficient in some instances—especially when the weight surface is visibly discolored.

For these cases one must remove the dirt by rubbing with chamois skin or soft cloth soaked in a convenient solvent (benzene, alcohol, and ether were solvents actually used). This operation must be followed by steam cleaning.

- Vigorous rubbing with chamois skin as described above, if done with care, does not damage the weight involved.

The cleaning technique has been further perfected by Girard at BIPM who profited from the low standard deviation ( $\sim 1 \mu\text{g}$ ) of the NBS-2 balance for his studies. Girard used successive rubbings with chamois skins soaked in benzene and ethanol followed by steam cleaning. He showed that prototype K63 was stable to about  $2 \mu\text{g}$  using such a procedure. A second platinum weight whose surface was badly scratched, however, required more than one cleaning procedure to reach a reproducible minimum mass [32].

In light of what is now known, it has been suggested that it is “not impossible” that the variability observed in measurements of prototypes between 1890 and 1946 is due to inferior methods of cleaning [43]. Indeed, it was the custom of BIPM to assume that observed changes in mass of less than  $50 \mu\text{g}$  were to be expected [44].

Thus, when examining the variability of prototypes after 1890, it is reasonable to dismiss results prior to 1946. Of the remaining results, one must also exclude cases where the prototype is known to have suffered an accident. The remaining ensemble of measurements reported by BIPM has a standard deviation of  $10 \mu\text{g}$  based on pooling 18 mass differences involving 13 prototypes. Data between 1891 and 1945 have been excluded.

Besides computing a pooled standard deviation, it is also interesting to see if the magnitude of the deviations has a time dependence—that is, if two calibrations of a prototype separated by several decades tend to be more discrepant than calibrations separated by a shorter period of time. We again examine only those data for weights which were steam cleaned just prior to calibration. Although the data are limited, the observed discrepancies appear flat between 10-year periods to the maximum 95-year period. The data of [32] suggest that, for periods of months, the scatter seen in repeated cleaning and calibrations is less than  $2 \mu\text{g}$ . These data, however, involved a single prototype which was measured on NBS-2. Also, the period of time involved was short enough that recleaning of the reference standard was unnecessary. These two factors undoubtedly contribute to the lower scatter. In particular, from the limited data presented by Bonhoure, one can see that a significant portion of the instability of assigned mass values must have been due to the balance that was used. Now that NBS-2 has been in service for over 10 years, it will be interesting to see if the measured long-term stability of the mass values assigned to the prototypes is improved.

We might note, in addition, anomalous behavior has been observed in at least one prototype. The short-term stability of K18 (belonging to the U.K.) is rather poor [45], although the long-term stability is representative of the ensemble of prototypes. The source of the problem has not yet been discovered. This behavior indicates that one must be cautious in drawing inferences for a

single prototype based on the behavior of an ensemble.

Since the U.S. has two prototypes, one would think that measurements of the mass difference between the two would provide valuable information. Such measurements are difficult to interpret, however, as can be shown by referring to table 1. The values for the mass difference of K20 and K4 presented range from 84  $\mu\text{g}$  to 11  $\mu\text{g}$ . This confusing picture becomes somewhat clearer upon close examination of the laboratory notebooks involved.

In particular, let us examine the data for 1947–1984. It appears that K4 was not cleaned between 1889 and 1970. In 1947, just prior to returning K20 to BIPM for recalibration, K20 was cleaned at NBS. The cleaning procedure used differed in many ways from that adopted at BIPM in 1948. It was clear from a series of measurements made at NBS that K20 actually *gained* about 20  $\mu\text{g}$  as a result of the “cleaning” used. This excess, as well as additionally accumulated surface dirt was, apparently, removed in the cleaning at BIPM in 1948 so that the mass value of K20-K4 decreased by 34  $\mu\text{g}$  between 1947 and 1949.

In 1970, both K20 and K4 were steam cleaned. The difference in value between 1970 and 1889 could be due to the observed surface damage on K4, noted in 1947 (see sec. 4.1).

In 1971 the kilograms were again washed with steam. The measured value for the mass of K20-K4 was low enough, however, for the metrologists involved to suspect that K4 had been insufficiently cleaned. After a vigorous recleaning, new surface damage was noticed on K4. This damage coincided with a new value of 66  $\mu\text{g}$  for the mass difference.

In 1983, both K20 and K4 were cleaned several times at NBS. The procedure used was rinsing in baths of benzene and ethanol, followed by steam cleaning. The use of rinsing baths is apparently inferior to actual rubbing with solvent-soaked chamois skin. This can be seen in the BIPM data which show that the cleaning at BIPM removed an additional 21  $\mu\text{g}$  from K20 and an additional 31  $\mu\text{g}$  from K4.

The lesson of these data is that strict adherence to the cleaning method used by BIPM is prerequisite to obtaining both stable results and results which would obtain were the measurements conducted at BIPM.

## Stainless Steel Standards

The research into the cleaning of stainless steel weights is less definitive than that concerning the prototypes. Bonhoure tried to apply the same cleaning procedure to stainless steel weights that he had found so

effective for platinum-iridium. The masses of the artifacts used were measured after each step of the cleaning procedure. The final steam cleaning caused a loss in mass of about 100  $\mu\text{g}$  although successive steam cleanings had no further effect [46,47]. Bonhoure also found that single-piece stainless steel weights which had been used in hydrostatic measurements might change their mass value by an appreciable amount. These changes were not permanent but recovery could take months [47]. For these reasons, it is the practice of BIPM to clean stainless steel weights exactly as platinum-iridium prototypes with the important omission of steam cleaning.

At NBS, we have used vapor degreasing in inhibited 1,1,1-trichloroethane as the final step in cleaning weights of stainless steel. Both this method and the BIPM method were used on steel spheres whose diameters were then measured optically. Dimensional measurements on the spheres which had been vapor degreased had the lower standard deviation [48]. (The exact reason for this, as well as the explanation for a systematic difference in the dimensional measurements for the two cleaning methods, is unknown.)

Vapor degreasing has proved to be an acceptable method of cleaning. A stackable set of stainless steel weights with nominal value of 1 kg was found to be stable under numerous vapor degreasing operations over a period of more than one year. The balance used had a standard deviation of 12  $\mu\text{g}$  and the kilogram weight set had double the surface area of D2.

Three vapor degreasings of CH-1 did not result in any noticeable changes in mass. The measurements were carried out on the V-1 balance.

More research is required to establish both an agreed-upon method for the cleaning of stainless steel weights and an alloy with optimal qualities. Such research is being carried out under the aegis of Working Group 3 of the Consultative Committee for Mass and Related Quantities.

The long-term stability of stainless steel standards is not well established. There is no doubt that some older weights have lost mass monotonically with time [48,50]. This is probably due both to wear<sup>1</sup> and to outgasing. The latter component would, presumably, not be significant in vacuum melted weights. Other possible mechanisms for instabilities in stainless steel weights have recently been studied at the Istituto di Metrologia “G. Colonnetti” (IMGC) in Turin, Italy [50,51].

<sup>1</sup> The hardness of annealed 90/10 platinum-iridium alloy is 130 HB. The alloy of which CH-1 is made has the same hardness. Type 18/8 stainless steel is somewhat less hard.