On the long-term stability of the triple-point-of-water cells
Dedyulin, Sergey; Emms, Rhys; Yang, Lu; Todd, Andrew D. W.; Mester, Zoltan; Peruzzi, Andrea; Van Geel, J. L. W. A.; Dobre, Miruna; Van Den Berghe, Debby; Senn, Remo; De groot, Martin

National Research Council Canada   Conseil national de recherches Canada

Canada

On the long-term stability of the triple-point-of-water cells

S N Dedyulin[1], R Emms[1], L Yang[1], A D W Todd[1], Z Mester[1], A Peruzzi[2], J van Geel[2], M Dobre[3], D Van Den Berghe[3], R Senn[4] and M J de Groot[5]

[1] National Research Council, Ottawa, ON, K1A 0R6, Canada
[2] VSL, Delft, the Netherlands
[3] SMD, FPS Economy, Brussels, Belgium
[4] METAS, Federal Institute of Metrology, Bern-Wabern, Switzerland
[5] Kelvin Training, Vught, the Netherlands

E-mail: Sergey.Dedyulin@nrc-cnrc.gc.ca

Abstract. Slow dissolution of the borosilicate container of triple-point-of-water (TPW) cell is widely recognized as the main cause of long-term drift in observed triple point temperature. We add to the available experimental data a comparison of two large batches of TPW cells (67 cells in total) of various ages (from 1 year to 64 years), manufacturers (NRC, VSL, Fluke, Isotech, etc.), and materials (borosilicate glass and fused-silica) which was undertaken in 2018. After measuring the TPW temperatures realized by all 67 cells, 12 borosilicate cells were opened and their water was analyzed by Inductively Coupled Plasma-Mass Spectrometry (ICP-MS) in order to correlate their impurity content with their respective age and their realized TPW temperature. No direct correlation was observed between the TPW cells age/impurity content and their measured triple-point temperature for neither borosilicate cells nor fused silica cells (Pearson's correlation coefficient r_{xy} is within the range $-0.60 \leq r_{xy} \leq +0.40$ for all the pairs considered). For fused-silica cells, the results indicate that after the isotopic variation in the water source is taken into account, the long-term drift due to the dissolution of glass envelope, if any, is negligibly small: $(+0.4(\pm0.6)\ \mu\text{K·yr}^{-1}$ reported herein). Given that all the fused-silica cells realize the TPW temperature within 100 μK of NRC and VSL national reference cells and since the analyzed time period of 15 years is equal to the average lifespan of a TPW cell, we conclude that fused-silica TPW cells are superior to those made from borosilicate glass.

Keywords: triple point of water, impurities, dissolution of glass, long-term stability

1. Introduction

Although the new definition of the kelvin [1], which came into effect May 2019, no longer relies on the triple point of water (TPW), the TPW continues to be the cornerstone of the International Temperature Scale of 1990 (ITS-90) [2]. All present-day practical temperature measurements are related to this reference temperature. As a consequence, TPW cells, which are the means for the practical realization of the TPW, will retain a fundamental role for many years to come.

The temperature realized by a TPW cell is exactly 273.16 K only for ideally pure Vienna Standard Mean Ocean Water (VSMOW) water. Although the TPW cells are manufactured with extreme care, the water enclosed within a TPW cell is never completely identical to VSMOW water. In particular, it is never free of impurities and it might have an isotopic composition different from VSMOW water.

The isotopic variations arise from the natural fractionation effects occurring in the the continental surface (fresh) water, which depletes the water of the heavy isotopes. In addition, in the distillation and degassing processes during the manufacturing of the cells, the water may be further depleted or enriched of heavy isotopes (see [3] for additional information and references). The combination of these effects leads to TPW realization temperatures typically ranging between $-110 \ \mu$K and $+10 \ \mu$K compared to the temperature realized with VSMOW water [3].

Impurities in the water of TPW cells are the most significant source of uncertainty and arise primarily from [3]:

- chemicals used in cleaning and pre-conditioning (chemical and thermal treatment prior to filling of the cell) of the glass envelope typically made of borosilicate glass or fused silica (also referred to, improperly, as quartz),

- slow dissolution of the glass envelope which over a lifetime of TPW cell releases nonvolatile impurities in otherwise high-purity water,

- low-volatility compounds in the source water not removed by distillation (e.g. light hydrocarbons),

- residual gases sealed in the cell or not fully removed from the water at the time of filling, or the small leak that may introduce air into the cell.

Of these four sources of contamination, the slow dissolution of the TPW glass envelope is the most extensively studied [4–8]. Two distinct mechanisms of the glass dissolution are considered in glass technology: leaching and etching [8, 9]. Leaching is an acidic ion exchange process in which metallic ions near the surface of the glass are swapped with hydrogen ions from the water. The reaction proceeds fastest for the monovalent ions, including sodium. Etching is an alkali attack process in which OH$^-$ ions cut chemical bonds causing dissolution of all components of the glass. The leaching is a diffusion process so decreases as the square root of time and is sensitive to the initial state of the glass surface. In contrast, etching is an exponential process that accelerates as the pH of the water increases with the dissolution of the metallic ions. Historically,

laboratory glass was treated with nitric acid or permanganate and sulphuric acid to pre-leach the glass surface before use, substantially reducing the initial leaching rate and greatly slowing the increase in pH that leads to significant etching. Long exposures to hot water and steam cleaning have the same effect, although are not as effective. Thus, the preparation of the glass prior to cell manufacture and the type of glass will have a major influence on the long-term stability of TPW cells.

The dissolution of the borosilicate glass envelope was first suggested by Hill in 1999 [5] when he opened ten of the most discrepant cells from the previous comparison of 27 borosilicate TPW cells [4] and analyzed the water content using Inductively Coupled Plasma-Mass Spectrometry (ICP-MS). The analyzed cells were sealed between 1954 and 1978; five cells were produced at NRC and the other five were commercial TPW cells produced by Jarrett Instrument Company or its predecessor, J&J; five cells contained excessive air and the remaining five cells exceeded the ± 100 μK deviation from the mean. The results of ICP-MS showed that the dominant impurities were Si, B, Na, and Al in relative abundances by weight that were very close to their relative abundances by weight in the borosilicate glass itself. However, no clear correlation was found between the measured freezing-point depression and the calculated TPW temperature depression from applying Raoult's law of dilute solutions (when using the total impurity content from ICP-MS results and the cryoscopic constant for water) possibly due to presence of excess air identified in some of the cells or due to the presence of additional unidentified impurities or due to differences in the isotopic composition of water [5]. Moreover, some of the results reported by Hill in the 1997 comparison [4] suggested that there exists a great variability in the dissolution rate of the borosilicate glass — some cells appear to remain uncontaminated for many years while others suffer from the gradual dissolution of the glass — and that effects such as the source of the glass and pre-conditioning of the glass envelope "may render predictions of solubility difficult or impossible" [5]. Put another way, as pointed out in [10], the results in [4, 5] (and, by the same token, the current study) are obtained by "simultaneously" measuring at a given time t a batch of N different cells that had been sealed at different antecedent times $\{t_1, t_2, ..., t_N\}$. This was assumed to be the equivalent of focusing on one single cell and following its time evolution by measuring it at times $\{t - t_1, t - t_2, ..., t - t_N\}$. In other words, since no detailed preparation and storage history was available for all the cells in the comparison, this approach assumed out of necessity the same time evolution for all the cells of the batch. Keeping this caveat in mind, an average drift of about -14 μK·yr^{-1} for the ten most discrepant TPW cells reported in [5] and the later revised number of -4 μK·yr^{-1} [6] for all the TPW cells measured in [4] minus "outliers" should be treated with caution as pointed out by Hill himself [6].

Adopting a different approach, in 2007 Strouse and Zhao [11] manufactured one borosilicate and one fused-quartz TPW cell in such a way that the water samples could be periodically (every 6 months) removed and tested for impurity content with ICP-MS to monitor the water composition changes with time, and correlate the changes in composition with the changes in realized TPW temperature. After 2 years of

measurements (4 data points), they confirmed Hill's conclusion about the dissolution of the borosilicate envelope with time and attributed the observed drift in the borosilicate cell of -13 μK·yr^{-1} primarily to the increasing amount of Si in the water, which they estimated having a constant dissolution rate of 94 nmol·mol^{-1}·yr^{-1}. Contrary to Hill's observation of no apparent correlation between measured TPW depression and prediction based on Raoult's law [5], the authors found that applying Raoult's law to Si impurity alone predicts a -10 μK·yr^{-1} drift rate, close to observed experimentally drift rate of -13 μK·yr^{-1}. Unlike the borosilicate ampule samples, there was no increased dissolution of impurities detected from the fused quartz into the water sample [11] with an average drift rate of -2 μK·yr^{-1} (and standard deviation residual of 4 μK) for fused-quartz cells. This observation is in perfect agreement with the earlier prediction of improved stability of high-silica glasses [8] based on the absence of leaching of metallic ions and, consequently, reduced catalytic OH$^-$ attack on the silica.

Given the slow dissolution of the borosilicate glass with time as described above, each laboratory that performs the measurements of the TPW for their ITS-90 realization is left with two main choices. The first option is to switch to using exclusively the fused-silica TPW cells which proved to be more stable compared to their borosilicate counterparts. The second option is to use the most recently manufactured or the most stable borosilicate cell available in order to minimize the impact of the envelope's dissolution. In both cases, the periodic measurement of the ensemble of TPW cells combined with the regular addition of newly manufactured cells (borosilicate or fused-silica) is required to assess the presence of long-term drift. Fused-silica cells are relatively new and there is not enough data on their long-term stability and the use of stable borosilicate cells cannot guarantee that they will not drift at some point in time. The third option, correction for impurities in each individual TPW cell, is still beyond the practical capabilities of any national metrology institute (NMI), as pointed out in [12], due to the lack of: a) a reliable, continuous and representative sampling of the cell water, b) an appropriate analysis technique that can identify all relevant impurity species and their corresponding concentrations (with the required accuracy), and c) a knowledge of the sensitivity coefficient for each impurity.

In this paper, we report on a comparison of a total of 67 TPW cells with the three-fold purpose: 1) periodically assess the long-term stability of various ensembles of borosilicate TPW cells as maintained at each participating NMI and identify the potential "outliers", 2) assess the variability in the fused-silica TPW cells from different manufacturers by checking the temperature depression vs cell's age similarly to what was done in the past for borosilicate TPW cells [4,7], since no such cells were available until 20 years ago, and 3) improve our understanding of the causes of the depression of the TPW cell temperature with time. The cells were manufactured commercially (Isotech/Jarrett/J&J, Fluke/Hart Scientific and Trans-Sonics) or in NMIs (NPL, UK; NIM, China; NRC, Canada and VSL, Netherlands) between 1954 and 2017 from borosilicate glass (48 cells) and from fused silica (19 cells) and were measured at three different laboratories (NRC, Canada; SMD, Belgium and VSL, Netherlands). In section

2 we provide the available information on the investigated cells (identification, year of manufacture, envelope material, design and the laboratory that measured it) and show how the measurements performed at the three different laboratories (NRC, SMD and VSL) can be linked together. In section 3 we report the TPW measurement procedures. In section 4 we report the results of the chemical analysis performed on the water of twelve TPW cells that were opened. Finally, in section 5, we analyze and discuss our results with reference to the existing literature.

2. The investigated cells

Two large batches of TPW cells were investigated: one batch of 36 cells, property of NRC (with a few exceptions) and measured at NRC, and another batch of 31 cells, property of VSL (with a few exceptions), of which 13 were measured at SMD and 18 at VSL. Tables 1 and 2 summarize the identification, manufacturer, year of manufacture, envelope material, design and measuring laboratory for all of the investigated cells.

The equivalence between NRC and VSL results can be obtained in a simple way from the fact that one of the cells measured by NRC, cell VSL16T013, was measured at VSL before shipping it to NRC. The results at NRC and VSL were, respectively:

$$T_{16T013} - T_{NRC}^{Nat.Ref.} = 12 \ \mu K \text{ with expanded uncertainty } U = 46 \ \mu K,$$

$$T_{16T013} - T_{VSL}^{Nat.Ref.} = \text{-6} \ \mu K \text{ with expanded uncertainty } U = 36 \ \mu K,$$

which translates into:

$$T_{NRC}^{Nat.Ref.} - T_{VSL}^{Nat.Ref.} = \text{-18} \ \mu K \text{ with expanded uncertainty } U = 58 \ \mu K$$

The equivalence between SMD and VSL results is obtained from the most recent (2014) direct comparison between SMD and VSL national references [12]. In such comparison, cell 08T004 was used as traveling artifact with the following results:

$$T_{08T004} - T_{SMD}^{Nat.Ref.} = 15 \ \mu K \text{ with expanded uncertainty } U = 20 \ \mu K$$

$$T_{08T004} - T_{VSL}^{Nat.Ref.} = 14 \ \mu K \text{ with expanded uncertainty } U = 36 \ \mu K$$

which translates into:

$$T_{SMD}^{Nat.Ref.} - T_{VSL}^{Nat.Ref.} = \text{-1} \ \mu K \text{ with expanded uncertainty } U = 41 \ \mu K$$

At NRC any "hottest" cell as determined from cell comparison could potentially serve as a national reference since the presence of impurities had been found to depress the TPW temperature [3]. For the purpose of this study, the NRC national reference was defined as the temperature realized by a fused-silica cell Q1049 of most recent manufacture, increased by 8 μK to account for the isotopic composition. All the temperature differences reported in table 1, including the measurements of reference cells, were calculated with respect to that value.

At VSL and SMD the national reference is the average realization temperature of a set of national reference cells and the difference of each cell of the set from the national reference is known from periodic internal comparisons. For the purpose of this

Table 1: The identification, year of manufacture, envelope material and cell design for NRC batch of TPW cells (measured at NRC). In gray are shown the cells that were opened and the water content was analyzed by ICP-MS. In bold is shown the cell that was used to define the NRC national reference for this comparison. In italic are shown the reference cells used for daily measurements in this comparison.

Maker	Model&Serial No.	Envelope material	Age (years)	Relative to Nat.Ref. (μK)	Std. Dev. (μK)	Isotopic Corr. (μK)
NRC	21	BS	64	−206	6	
NRC	24	BS	64	−178	13	
NRC	25	BS	64	−127	15	
J&J	A-11 113	BS	62	−370	18	
J&J	B-11 306	BS	59	−128	15	
Trans-Sonics	T 224	BS	59	−189	8	
Jarrett	B-11 387	BS	55	−166	4	
Jarrett	B-11 539	BS	52	−227	11	
Jarrett	B-11 744	BS	43	−196	36	
Jarrett	B-11 755	BS	43	−149	26	
Jarrett	B-11 806	BS	40	−252	13	
Jarrett	B-11 809	BS	40	−250	10	
Jarrett/Isotech	B-11 2053	BS	21	−23	28	
Isotech	*B-11 2063*	*BS*	17	−84	25	−6
Isotech	B-11 171	BS	17	−1244*	39	
Hart	5901 1256	BS	17	−209	15	
Hart	5901 1257	BS	17	−290	19	
Hart	5901 1308	BS	16	−232	15	
Isotech	B-11 267	BS	15	−101	11	
Isotech	*B-11 269*	*BS*	15	−101	27	
Isotech	B-11 270	BS	15	−134	12	
Isotech	B-11 271	BS	15	−89	9	
Isotech	B-11 273	BS	15	−87	14	
Isotech	B-11 Q321	FS	15	−34	31	+11
Isotech	B-11 Q322	FS	15	−32	6	+11
Isotech	B-11 Q324	FS	15	+25	6	+22
Isotech	*B-11 Q325*	*FS*	15	−39	21	+22
Isotech	*B-11 Q326*	*FS*	15	−25	20	+22
NIM	19[a]	BS	10	−120	9	
Isotech	B11 50400[b]	BS	7	−21	13	
VSL	16T013[c]	FS	2	+12	8	+5
Fluke	**5901C Q1049**	**FS**	1	−8	21	+8
Fluke	5901D Q1150	FS	1	+28	11	+9
Hart	5901C 1031	BS	n/a	−443	14	
Hart	5901C 1032	BS	n/a	−467	17	
Fluke	5901D Q1020[d]	FS	n/a	+10	9	

* Contains air as mentioned in [7].
[a] Courtesy of PTB, Germany.
[b] Courtesy of NPL, UK.
[c] Courtesy of VSL, Netherlands.
[d] Courtesy of LNE-CNAM, France.

Table 2: The identification, year of manufacture, envelope material, design and measuring laboratory for VSL batch of TPW cells. In gray are shown the cells that were opened and the water content was analyzed by ICP-MS. In italic are shown the reference cells used for daily measurements in this comparison.

Maker	Serial No.	Envelope material	Age (years)	Relative to Nat.Ref. (μK)	Std. Dev. (μK)	Measuring laboratory	Isotopic Corr. (μK)
PTB	4-78	BS	40	−631	144	VSL	
VSL	89T023	BS	29	−570	327	VSL	
VSL	89T025	BS	29	−487	90	SMD	+73
VSL	89T026	BS	29	−346	122	SMD	
VSL	89T220	BS	29	no plateau		VSL	
VSL	91T228	BS	27	−468	151	VSL	+49
VSL	92T053	BS	26	−117	20	VSL	+31
VSL	92T057	BS	26	−67	19	VSL	
VSL	93T368	BS	25	−690	169	VSL	+39
NPL	735	BS	24	−471	106	SMD	+56
VSL	94T096	BS	24	−249	48	VSL	+51
VSL	94T223	BS	24	−3112*	212	VSL	
VSL	95T374	BS	23	−474	86	SMD	+53
VSL	95T375	BS	23	−109	34	SMD	
VSL	98T087	BS	20	−229	76	SMD	
VSL	98T089	BS	20	−231	32	SMD	
VSL	98T094	BS	20	−226	39	SMD	
VSL	00T013	BS	18	−596	51	VSL	+4
VSL	03T039	BS	15	−222	136	SMD	+35
VSL	03T030	BS	14	−156	25	SMD	+49
VSL	03T032	BS	13	−88	30	SMD	+15
VSL	*06T003*	*FS*	12	−13	14	*VSL*	+17
VSL	06T006	FS	12	−12	14	VSL	+47
VSL	*07T120*	*FS*	11	+25	14	*SMD*	
VSL	08T001	FS	10	−38	14	VSL	+26
VSL	*08T004*	*FS*	10	−18	14	*SMD*	+3
VSL	*08T005*	*FS*	10	+14	14	*VSL*	+2
VSL	15T009	FS	3	−41	14	VSL	+50
VSL	15T010	FS	3	−20	14	VSL	+28
VSL	17T042	FS	1	+95	14	VSL	−46
VSL	17T048	FS	1	+31	14	VSL	+24

* Possibly contains air.

study, the known differences of reference cells VSL07T120 and VSL08T004 at SMD and VSL06T003 and VSL08T005 at VSL from VSL group of national reference cells were used to calculate the temperature differences reported in table 2, since the difference between the SMD and VSL national references is negligible.

3. TPW measurement procedures

Each laboratory applied its local method for preparing the ice mantle: NRC used crushed solid CO_2, SMD used a heat pipe immersion cooler and VSL used liquid/vapour nitrogen counter-flowing through a multi-tube cooler (methods 1, 2 and 4a respectively described in [3]). In all laboratories, the ice mantle was aged for at least 7 days before starting the measurements. In some cases at NRC, particularly for older borosilicate TPW cells, time required for the measured temperature to stabilize exceeded 3 weeks. One possible explanation of this negative "medium-term" drift — to distinguish it from the short-term drifts due to release of strains in the ice mantle [13] and long-term changes due to dissolution of the glass — would be a slow release of impurities trapped inside the ice in small liquid veins at ice grain boundaries into the pure water of inner melt as suggested in [12]. These medium-term drifts can explain large standard deviations reported in table 1 and 2 and, certainly, can lead to erroneous results and hinder observation of any trends in the data, since there is indication in the literature [14] that they can possibly continue past a one-month mark. Thus any apparent TPW temperature stabilization for a given ice mantle could be only temporary. In this paper, we did not try to investigate these drifts further or mitigate them, in a sense, adopting a "common user" approach to TPW measurement. Note: historical NRC TPW measurements [4] perused in this work were relying on 2-3 measurements performed at a fixed 10-day mark which would lead to higher measured resistance ratios and smaller temperature depressions for impure TPW cells [14].

Prior to each measurement, an inner melt layer was first initiated by inserting a room-temperature metal rod into the thermometer well and then by gently rotating the cell to free the mantle around the well. At VSL and SMD the inner and outer layers of water were then intermixed by gently reversing the cell three times as described in [12], while no such mixing was performed at NRC. Measuring the temperature realized by the cell without mixing and after the cell has been gently inverted several times could result in two different values for measured temperature depression. During the freezing of the mantle, ice forms with a very low impurity concentration, due to the low distribution coefficient of the impurities. When the cell is first used, the water melted around the well has a relatively high purity which can be altered slowly, due to mixing via the grain boundaries or contamination via the top of the mantle, or faster, due to forced mixing via reversal of the cell, causing a gradual depression of the temperature in the first case (labelled medium-term drift above) and more abrupt change in the second.

Two standard platinum resistance thermometers (SPRT) were used by NRC and SMD, while VSL used only one SPRT. All laboratories measured the resistance ratio of

an SPRT to a known reference resistor (100 Ω at NRC and 25 Ω at VSL and SMD) with a resistance ratio bridge (ASL F900 and F18 at NRC, ASL F900 at SMD and MI 6015T at VSL). An SPRT was allowed to thermalize for at least 30 minutes before starting the data acquisition.

NRC used one reference cell per day out of a batch of 4 reference cells (Q325, Q326, 2063 and 269) and measured all 4 reference cells once a week. SMD and VSL used at least one reference cell per day out of a batch of two reference cells (VSL07T120 and VSL08T004 at SMD and VSL06T003 and VSL08T005 at VSL).

Since the batch of cells measured by NRC essentially represents NRC cells and the batch of cells measured at SMD and VSL essentially represents VSL cells, in reporting the measurements results in tables 1 and 2 we have kept them separately. The temperature shift of NRC cells was calculated with respect to the NRC national reference (as realized by cell Q1049 in this comparison as described in section 2), while the shift of VSL cells was calculated with respect to the VSL national reference (using known differences for daily reference cells as described in section 2). Note: the historical NRC data in [5] were calculated with respect to the hottest cell 2053 that hasn't changed much since then (from table 1: -23 μK with a standard deviation of 28 μK with respect to the national reference), so in the following we simply shifted all the measured TPW deviations in [5] by -23 μK for the purpose of the comparison.

4. Chemical analysis

A total of 12 cells (3 from the NRC batch and 9 from the VSL batch highlighted gray in tables 1 and 2) were opened and the nonvolatile impurity content in the water was analyzed by ICP-MS by the Chemical Metrology Group at NRC.

Two of the water samples (from cell 113 (NRC) and cell 03T032 (VSL)) were preliminarily scanned for 70 different chemical elements. Such a preliminary scan allowed to identify the relevant impurities present in the cells: B, Na, Mg, Al, Si, K, Ca and Fe. It was found that these 8 impurity species alone account for 99% of the total impurity content. As a consequence, for the remaining cell water samples, the ICP-MS analysis was limited to only 8 most abundant elements identified above. The results of ICP-MS are reported in table 3 with a stated relative uncertainty of 5% ($k = 1$). Note that VSL cells have significantly higher content of Mg and Ca compared to NRC cells. This is probably due to the different borosilicate glass and, possibly, a different pre-conditioning procedure of the cell envelope prior to filling of the cell.

5. Analysis and discussion

As discussed in the introduction, this study had a three-fold purpose. First, we wanted to periodically assess the long-term stability of an ensemble of borosilicate TPW cells at each participating NMI as part of their respective periodic reassessment programs (the last such comparison at NRC was done in 2014 [7]) and identify the potential "outlier".

Table 3: Mole fraction (nmol·mol^{-1}) of major impurities present in the TPW cells as determined by ICP-MS. Oxides of the elements shown in gray represent the principal constituents of borosilicate glass. The last two columns compare the measured TPW temperature depression and the expected depression as calculated by multiplying the total amount of impurities by the cryoscopic constant of water.

Cell	B	Na	Mg	Al	Si	K	Ca	Fe	Total	Expected Depression (μK)	Measured Depression (μK)
113	651	129	8	6	1547	2	2	0.2	2346	242	370
171	3687	549	0.4	33	1622	76	2	0.5	5971	615	1244*
1031	774	104	0.4	5	1020	2	2	0.2	1908	197	443
89T025	324	444	469	33	1679	26	181	6	3162	326	414
91T228	691	475	250	44	1341	35	63	2	2900	299	419
92T053	157	562	601	30	1833	34	229	7	3453	356	86
93T368	833	457	254	34	1878	24	63	1	3544	365	651
94T096	432	403	217	33	1502	19	75	1	2681	276	198
NPL375	1721	715	271	27	2040	28	104	0.7	4907	505	415
95T374	1715	675	297	20	2769	26	124	0.6	5627	580	421
00T013	517	466	351	27	1788	22	143	2	3317	342	592
03T032	115	1125	1847	1	687	226	646	2	4647	479	73

* Contains excess air as mentioned in [7].

Second, we wanted to assess the variability in the fused-silica TPW cells from different manufacturers by checking the temperature depression vs cell's age, since these cells were relatively new in the past comparisons. Finally, we wanted to further understand the causes of the depression of the TPW cell temperature with time. For performing the TPW measurements, the participating laboratories were left free to apply their own internal procedures as described in detail in section 3. Although this complicates the interpretation of the data, it simulates the "common user" approach and also allows one to perform measurements on a large ensemble of TPW cells relatively quickly. In the following, we divided our discussion according to the goals outlined above.

5.1. Temperature depression vs age for borosilicate TPW cells

All the measured deviations from the national reference at NRC and VSL are listed in tables 1 and 2 with their respective standard deviations and plotted in figure 1. Some of the TPW cells that were found to deviate by more than 100 μK from the national reference, were later opened for the impurity analysis as described in detail further below. It should be noted here that at NRC only fused-silica cells are used for regular measurements with borosilicate cell 2053 being the only exception, while the rest of the
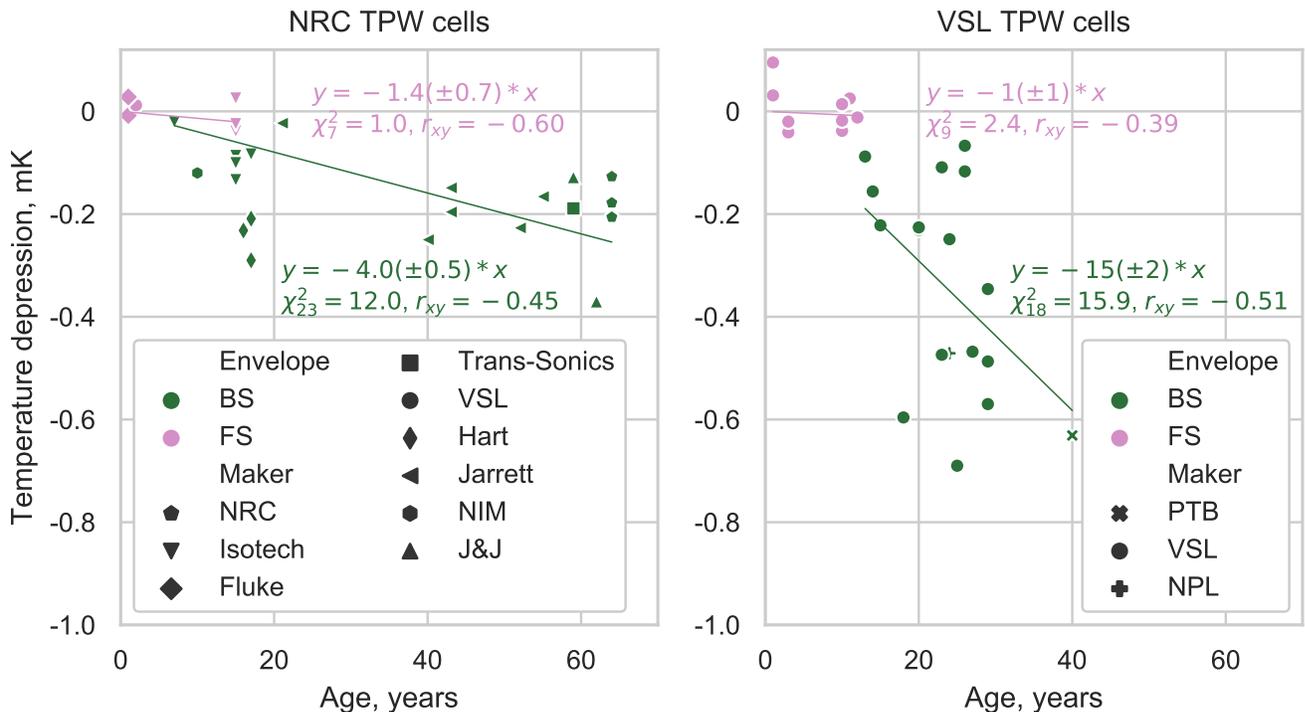
Figure 1: Temperature depression of the NRC and VSL cells with respect to the national reference. Cells 171 (NRC) and 94T223 (VSL) are excluded from consideration due to possible air contamination as mentioned in the text.

cells are measured only in the comparisons such as the one reported here. Similarly, at VSL only 2-3 best cells are used on a regular basis and newly manufactured cells (exclusively fused silica since 2006) are compared with them.

We also looked at the correlation between the age of the cells and their temperature depression specific to the batch and envelope's material as it was done historically [5, 6, 11] by fitting straight line with zero intercept‡ to the data. In addition to traditionally reported linear slope numbers, we list where appropriate the following:

(i) Pearson correlation coefficient, r_{xy}§, which is a statistical measure of the linear relationship between two samples x and y, and which can take on any real value in the range -1 (perfect negative linear relationship) $\leq r_{xy} \leq 1$ (perfect positive linear relationship);

(ii) the standard deviation for the fitted slope;

(iii) the reduced chi-squared $(\chi_\nu^2)\|$, which is a statistical measure of goodness of fit with $\chi_\nu^2 \gg 1$ indicating a poor model fit;

‡ The reasons for choosing a zero intercept are discussed in detail in [6].

§ $r_{xy} = \dfrac{\sum_{i=1}^{N}(x_i-\overline{x})(y_i-\overline{y})}{\sqrt{\sum_{i=1}^{N}(x_i-\overline{x})^2}\sqrt{\sum_{i=1}^{N}(y_i-\overline{y})^2}}$, where N is sample size and $\overline{x} = \frac{1}{N}\sum_{i=1}^{N}x_i$.

$\|$ $\chi_\nu^2 = \chi^2/\nu$, where the chi-squared is a weighted sum of squared deviations, $\chi^2 = \sum\limits_{i=1}^{N}(y_i - f_i)^2/\sigma_i^2$,

(iv) root mean square error (RMSE)¶, which is a statistical measure of how concentrated the data is around the line of best fit.

Looking for linear trends (with fixed zero intercept) in the measured deviations vs. age of TPW cells has its major shortcoming as described in detail in section 1. It presupposes that: 1) all different cells evolve in time in the same way starting from common value (zero), and 2) nuances of the measurement method (e.g. length of time before the first data point is taken, the same melted fraction F for all the cells measured, etc.) have no effect on the outcome. We find that in our own data presented in figure 1 and in the data presented in the literature, the simplest linear approximation (with or without zero intercept, with or without taking measurement standard deviations into account during the fitting) does not describe well the scatter of data points. Thus reporting a single number for the drift slope could be misleading, in particular, if this number is then used to predict the "health" of the triple point of water ensemble at some point in time.

For example, when all the NRC borosilicate cells are considered — excluding sufficiently discrepant cell 171 that was observed to contain excess air in the previous comparison [7] — then the measured average drift of NRC borosilicate cells amounts to $-4.0(\pm0.5)$ μK·yr^{-1} (see figure 1). Comparing this number to the average drift in the VSL borosilicate cells of $-15(\pm2)$ μK·yr^{-1} reported herein (see figure 1) and with the numbers reported in the literature: $-6(\pm2)$ μK·yr^{-1} in [7]; $-14(\pm2)$ μK·yr^{-1} in [5] for the ten most discrepant cells and $-4(\pm1)$ μK·yr^{-1} in [6] for selected eighteen out of twenty seven cells measured in 1997; and $-13(\pm3)$ μK·yr^{-1} in [11], a large fitting residuals and the bimodal distribution of the reported slope values are apparent. When we consider also the very high reduced chi-squared values for the linear fits in figure 1 (and, most likely, all the references above), we have to conclude that one (or both) of the two implicit hypotheses listed above is not supported by the data. Furthermore, the NRC data for borosilicate cells in figure 1 appear to be grouped closely together for the same manufacturer. In some cases, the apparent groups are obviously distinct from the fitted line, thus contributing to the poor statistical goodness, as indicated by large RMSE values: Hart - 181 μK, J&J - 115 μK, NRC - 90 μK, Jarrett - 60 μK, Isotech - 42 μK. This observation suggests that it is the first assumption - that all different cells evolve in time in the same way starting from common value (zero) - that is not supported by the data, possibly, due to the unique nature of each TPW cell, such as water source, glass source, pre-conditioning of the glass envelope, etc. In other words, in the case of borosilicate TPW cells, one cannot simply rely on the prediction from one particular value for the linear drift rate from the literature to estimate the temperature

---

y_i is the observations with variance σ_i^2, related to measurement uncertainty for y_i, f_i is the fitted data, and ν is the number of degrees of freedom and equals the number of observations N minus the number of fitted parameters m, $\nu = N - m$.

¶ Using the notation above, RMSE, or the standard deviation of the residuals, is equal to $\sqrt{\sum\limits_{i=1}^{N} (y_i - f_i)^2/N}$.

deviation of any particular cell from the national reference, unless perchance the time evolution data are obtained on the same cell following the same protocol (and ideally by the same person) as it was done in [11]. Instead, a regular comparison of the borosilicate cells with the newly produced cell (or the cell shown to be stable previously) is required in order to detect the potential drift in the regularly used TPW cell.

5.2. Temperature depression vs age for fused-silica TPW cells

Manufacturing of the TPW cells from high-silica glasses was recommended in [8], which sought to explain the chemical kinetics of the dissolution of borosilicate glass. TPW manufacturers took notice of this recommendation, and fused-silica TPW cells — potentially offering much better long-term stability than their borosilicate counterparts [7, 11] — are now widely available on the market. Since 2004, all the TPW cells purchased by NRC are made from fused-silica. Since 2006, all the TPW cells at VSL are manufactured from fused-silica. However, given the relatively short history of use of fused-silica TPW cells compared to the borosilicate ones and without an absolute reference, it is difficult to demonstrate that fused-silica cells remain stable over long time. With these two limitations in mind, one can only resolve to performing regular comparisons of fused-silica TPW cells between each other and with newly manufactured cells (preferably from different manufacturers) in order to assess their long-term stability. Herein we add to the available information on the long-term stability of fused-silica TPW cells with the comparison of 19 cells from different manufacturers (VSL, Fluke, Isotech) and of different ages (from 1 to 15 years old) to facilitate and benefit future studies.

Once again, all the measured deviations from the respective national reference at NRC and VSL are listed in tables 1 and 2 and plotted in figure 1 with respect to their age. All the fused-silica cells lie consistently higher than their borosilicate counterparts - a good sign as "the hotter is the better" (assuming the same isotopic composition) in the case of TPW cells, since the presence of impurities had been found to depress the TPW temperature [3]. As it was done historically for borosilicate TPW cells, we draw linear fits to the data. However, following the discussion in the previous section, not much weight is given to them in trying to decide on the long-term stability of fused-silica TPW cells until it will be shown that manufacturing and/or measuring method do not contribute to the scatter of data points. Instead, we use the slope from the linear fits exclusively to compare the cell ensembles between each other and with the literature data.

We found that the average drift of NRC commercial fused-silica cells of $-1.4(\pm 0.7)$ μK·yr^{-1} is very similar to that of VSL-manufactured fused-silica cells of $-1(\pm 1)$ μK·yr^{-1} (see figure 1), and both agree with Strouse's number of $-2(\pm 4)$ μK·yr^{-1} measured on the NIST-manufactured single cell [11]. Very small drift rate numbers with large uncertainties suggest that that the dissolution of the fused-silica envelope is very small if any and that the scatter of data we see in figure 1 is potentially coming from the impurities and isotopic variations in the original water source. Also, since all three

aforementioned ensembles were composed of TPW cells from different manufacturers (NRC: Fluke, Isotech; VSL: VSL; NIST: NIST) and yet resulted in similar drift rates, this could imply that fused-silica TPW cells potentially do not show large variability due to the envelope preparation. Another observation is that the temperatures realized by NRC fused-silica cells are higher than NRC borosilicate cells of the same age (approximately 15 years) and from the same manufacturer (Isotech). If one assumes the same water source for both subsets, this would be another indication of a better long-term stability of the fused-silica TPW cells.

In order to test whether the isotopic variation in the original water source influences the scatter of data points we applied a known isotopic correction to 16 out of 19 fused-silica TPW cells. For example, an "outlier" cell 17T042 in figure 1 ($+95\,\mu$K with respect to the national reference) showed an anomalously "enriched" isotopic composition: $+65.5$ ‰ of ^{18}O, $+6.2$ ‰ of ^{2}H and $+3.3$ ‰ of ^{17}O, which will result in isotopic correction of $-46\,\mu$K. After applying the correction, we combined the two sets of fused-silica TPW cells from NRC and VSL in figure 2 in order to increase the representative size of an ensemble and since we have shown that VSL and NRC ensembles of fused-silica TPW cells are very similar when analyzed individually. Analyzed this way, the scatter in the data with respect to zero line should represent the impurity contribution due to glass dissolution assuming that manufacturing/measuring method has no additional effect. In an ideal case scenario, one would expect all the data to lie at the zero line within measurement uncertainties. With NRC cell Q325 being the only exception, all the measured deviations do indeed scatter around zero within expanded measurement uncertainty (in figure 2 we only show k=1 case for clarity) after the isotopic correction is applied. In fact, the entire span of y-axis in figure 2 is equal to the $100\,\mu$K uncertainty of TPW realization at NRC while the age span of the cells covers 15 years - the average lifespan of TPW cell in the lab. The traditional linear fit to all the data points yields an average drift rate of $+0.4(\pm0.6)\,\mu$K$\cdot$yr^{-1} with very large residuals (RMSE is $28\,\mu$K) and a reduced chi-squared value very close to 1 ($\chi^2_{17} = 1.1$). As a matter of fact, the scatter of points in the 1-year-old cells amounts to almost $60\,\mu$K and is nearly identical to that of 15-year-old cells from different manufacturer. These observations combined indicate that the dissolution of fused-silica envelope could not be detected over 15 years time span.

The main conclusion of this part of the study is — using the words of Hill [7] — that "fused-silica-encapsulated TPW cells appear superior to those made from borosilicate glass". So much so that 15 years - an average lifetime of TPW cell - is not enough to confirm the presence of long-term drift in fused-silica cells.

5.3. Temperature depression vs impurity content

Most comparisons of TPW cells end up with the determination of measured differences without elucidating the reasons for those differences. Hill in 1999 [5] has shown that a chemical analysis of the water yields valuable information regarding the type and
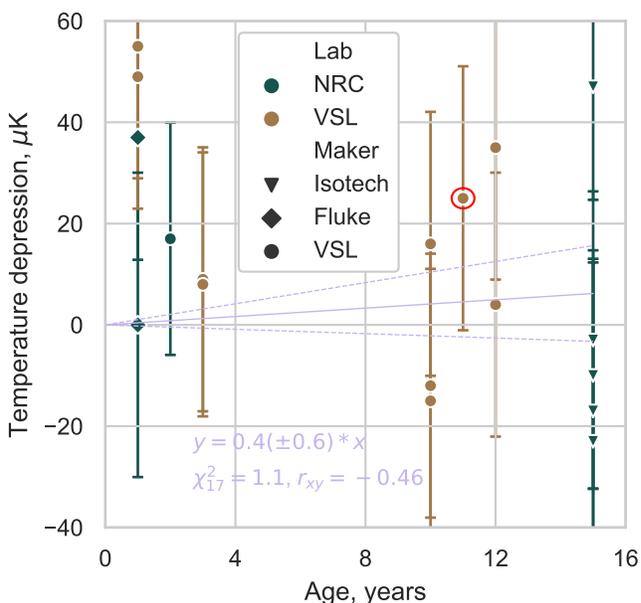
Figure 2: Temperature depression of the fused-silica cells shown as a function of cells' age after the isotopic correction was applied. Measurement uncertainty ($k = 1$) is shown with error bars and the red circle indicates the cell for which there is no isotopic correction available. The linear fit (solid line) to all the data is shown with one standard deviation in the slope parameter (dashed line).

quantity of dissolved impurities. Indeed, after analyzing ten of the most-discrepant cells with ICP-MS, Hill concluded that the envelope of some borosilicate cells dissolves with time leading to the measured temperature depression of the TPW. In this study, we added 12 more TPW cells to the set of 10 cells in [5] — analyzed by the same person in the Chemical Metrology Group at NRC 20 years ago — with the hope that we would be able to discern some additional trends in the measured TPW temperature depression as a function of the individual impurity content. The combined plots of the TPW cells analyzed in 1999 and 2018 are shown in figure 3 together with the attempted linear fits of combined data and the corresponding reduced chi-squared values (as described in section 5.1), while the numerical values used for plotting can be found in table 3 (this work) and table 1 (from [5]).

As follows from a close look at the table 3, in agreement with the previous conclusion by Hill, we found that the dominant impurities are boron, sodium, aluminium and silicon, while the oxides of these elements represent the principal constituents of borosilicate glass [5]. However, no additional trends could be discerned when the temperature depression for the total of 22 cells is plotted in figure 3 as a function of their age, total impurity content (estimated by adding up the measured concentrations of the 8 major impurities) or individual impurity content (only boron, sodium, aluminium and silicon are shown). Moreover, similarly to [5], we found no correlation (Pearson's $r_{xy} = -0.07$) between the expected freezing point depression, calculated by multiplying
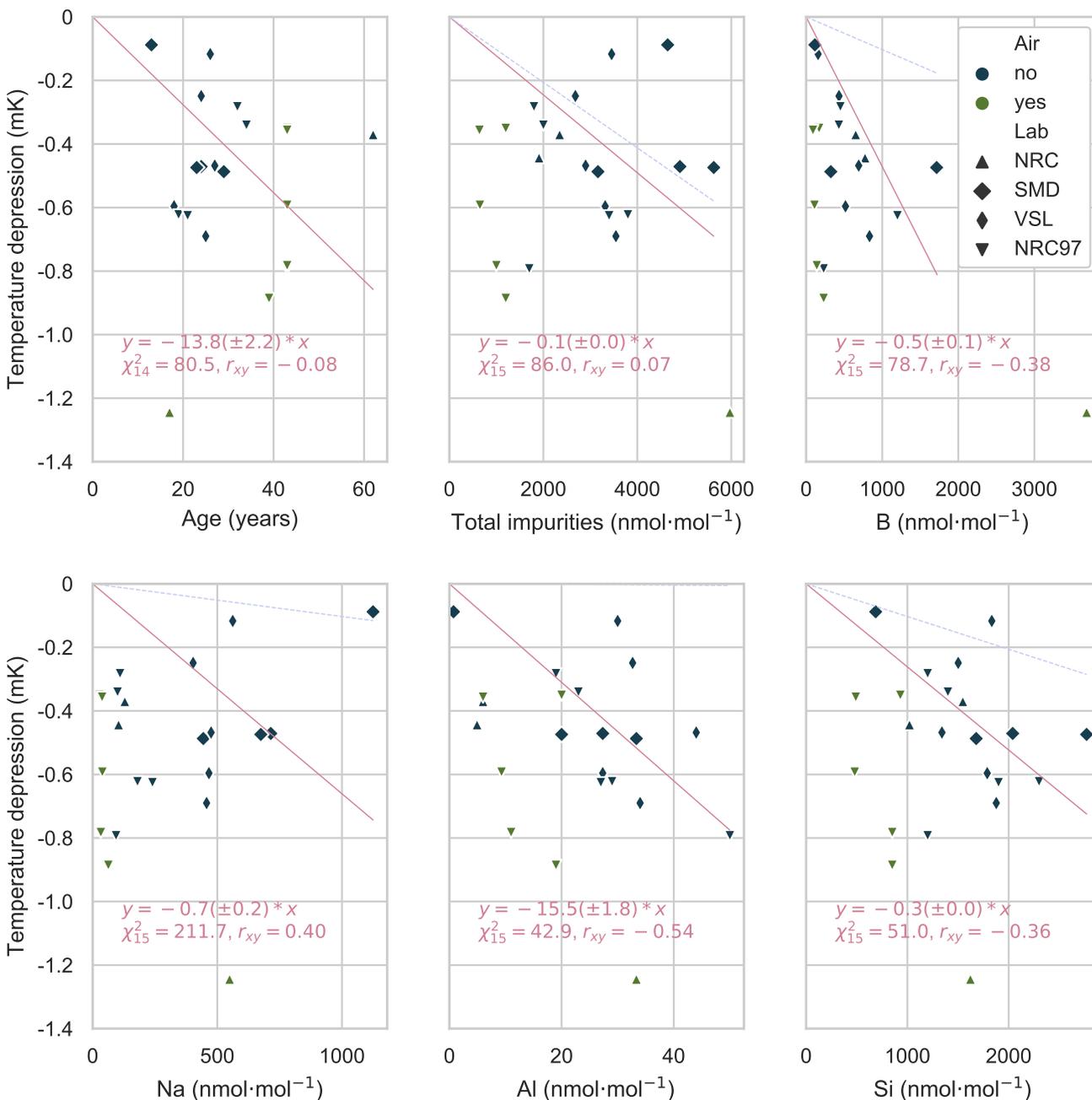
Figure 3: Measured temperature depression for 22 borosilicate TPW cells (including the historical NRC data of 1997 [5]) as a function of their age, total and individual impurity content. Cell 1031 was not included in the first plot since no information on the age of the cell was available. The cells with the confirmed air contamination were excluded from the linear fits of available data (solid line). Dashed blue line shows the Raoult's law prediction.

the total impurity content by the cryoscopic constant for water (103 K per unit mole fraction) and the measured TPW temperature depression for 16 TPW cells without air contamination. In some extreme cases, the calculated freezing point depression underestimates the measured value by a factor of 2 (NRC cell 1031) or overestimates it by a factor of 4 (VSL cell 92T053). This is contrary to the observation by Strouse and Zhao [11] where they were able to explain most of the measured TPW temperature depression with Raoult's law applied to silicon impurity alone — while the concentration of the latter increased linearly with time — as discussed in the introduction. The argument of a varying production method for different manufacturers of TPW cells measured in this work compared to the single cell in [11] cannot alone explain the absence of any linear trends in the data specific to one manufacturer, VSL, in figure 3. Some other factors must be at play, such as: different TPW measurement method used by each lab, impurities introduced during sample preparation for ICP-MS, presence of an organic impurity that cannot be detected by ICP-MS, presence of air in most of the 22 TPW cells that obscures the linear trend or some other, yet unknown factor. The influence of TPW measurement method could probably be ruled out based on the absence of any systematic trends (with respect to the linear trend line) in the data specific to one lab in figure 3. Unfortunately, "common user" approach adapted in this work did not allow us to rule out the influence of any other factors.

For now, we have to conclude that despite an additional set of TPW cells analyzed with ICP-MS in this work, the difficulty to correlate the measured temperature depression in commercial borosilicate cells with impurity content still remains. It is entirely possible that the missing link will remain undiscovered as fused-silica cells will further prove themselves stable long-term and all the labs will eliminate the borosilicate TPW cells from use altogether.

6. Conclusions

Another "snapshot" of the NRC and VSL TPW cells has been completed, with more data now available on the long-term stability of borosilicate and fused-silica cells and the impurities present in the borosilicate "outliers". This comparison is yet another touchpoint for comparisons that may take place in the future at NRC, SMD and VSL with the added benefit that the measurements in three labs can now be linked together going forward. Despite the plethora of available experimental data, the challenges to understanding the behavior of impurities in borosilicate TPW cells still remain and the perspective of predicting the onset of long-term drift in any given TPW cell is dim. The two clear recommendations to come from this work are: 1) to emphasize the importance of periodic comparisons of TPW cells required to assess the presence of long-term drift, and 2) to encourage more measurements of "old" fused-silica cells in the future.

Overall, the results indicate that after the isotopic variation in the water source is taken into account, the long-term drift in the fused-silica TPW cells due to the dissolution of glass envelope, if any, is negligibly small: $(+0.4(\pm 0.6)$ μK·yr^{-1} reported

herein). Given that all the fused-silica cell realize the TPW temperature within 100 μK of NRC and VSL national reference and since the analyzed time period of 15 years is equal to the average lifespan of TPW cell, we conclude that fused-silica TPW cells are superior to those made from borosilicate glass. Future regular comparisons of fused-silica TPW cells combined with the ICP-MS analysis of impurities are required if we are to retain confidence in their ability to provide a reliable reference temperature. We thus encourage all the labs to assess the long-term stability of fused-silica cells in their future comparisons.

While certain old borosilicate TPW cells (e.g., 224, 2053) appear to remain stable over the course of years, we neither can attribute it to a specific cause, nor can we guarantee that the long-term drift will not commence at some point in time. This situation may remain unchanged for the foreseeable future as fused-silica cells will further prove themselves stable long-term and all the labs will eliminate the borosilicate TPW cells from further use altogether.

Acknowledgments

References

[1] Bureau International des Poids et Mesures (BIPM) 2019 *SI brochure, The International system of units (SI)* 9th Ed (Paris: BIPM) (available at https://www.bipm.org/utils/common/pdf/si-brochure/SI-Brochure-9.pdf)

[2] Preston-Thomas H 1990 The International Temperature Scale of 1990 (ITS-90) *Metrologia* **27** 3-10

[3] Consultative Committee for Thermometry under the auspices of the International Committee for Weights and Measures 2015 *Guide to the realization of the ITS-90; Triple point of water* (Paris: BIPM) (available at https://www.bipm.org/utils/common/pdf/ITS-90/Guide_ITS-90_2_2_TPW-2018.pdf)

[4] Hill K D 1997 A comparison of triple point of water cells *Proc. TEMPBEIJING '97: Int. Conf. on Temperature and Thermal Measurements* (Beijing: Standards Press of China) 10-5

[5] Hill K D 1999 Triple point of water cells and the solubility of borosilicate glass *Proc. TEMPMEKO '99: 7th Int. Symp. on Temperature and Thermal Measurements in Industry and Science* ed J F Dubbledam and M J de Groot (Delft: NMi-Van Swinden Laboratorium) 68-73

[6] Hill K D 2001 Is there a long-term drift in triple point of water cells? *Metrologia* **38** 79

[7] Hill K D 2014 An experimental investigation of the long-term stability of triple-point-of-water cells *Int. J. Thermophys.* **35** 611–19

[8] White D R, Downes C J, Dransfield T D, Mason R S 2004 Dissolved glass in triple-point-of-water cells *Proc. TEMPMEKO '04: 9th Int. Symp. on Temperature and Thermal Measurements in Industry and Science* ed D Zvizdic (Zagreb: Laboratory for Process Measurement) 251-6

[9] Adams P B Glass containers for ultrapure solutions *Ultrapurity-methods and techniques* ed M Zief and R Speights (New Yourk: Marcel Decker Inc.) 293-351

[10] Peruzzi A, Dobre M, Strouse G and Van Geel J 2011 Doping experiments with water triple-point cells *Int. J. Thermophys.* **32** 1553-62

[11] Strouse G F, Zhao M 2007 The impact of isotopic concentration, impurities, and cell aging on the water triple-point temperature *Int. J. Thermophys.* **28** 1913-22

[12] Peruzzi A, Dobre M, van Geel J, Uytun A, Kalemci M, Uysal E, Strouse G, Nuevo Ordonez Y and Davis C 2014 Effect of Impurities on Water Triple-Point Cells *Int. J. Thermophys.* **35** 1084-96

[13] Berry R J 1959 The temperature-time dependence of the triple point of water *Can. J. Phys.* **37** 1230-48

[14] Steur P P M and Dematteis R 2014 Water triple-point comparisons: Plateau averaging or peak value? *Int. J. Thermophys.* **35** 620–7