Int J Thermophys (2009) 30:360–369
DOI 10.1007/s10765-008-0488-8

# Demonstrating Competency and Equivalency of Two Commercial SPRT Calibration Facilities

**T. J. Wiandt**

**Abstract** The Hart Scientific Division of Fluke Corporation operates two accredited low-uncertainty SPRT calibration facilities; one in the USA and the other in the UK. Competency and equivalency must be demonstrated for both facilities. However, because of the low uncertainties involved, the required experiments are both expensive and challenging. In the USA, a proficiency test (PT) is available through NVLAP based on the long-standing NIST measurement assurance program to accomplish this purpose. Although needed, a PT of this level is not readily available elsewhere in the world. Consequently, an alternative approach is required. This paper describes the approach taken in an effort to show both competency and equivalency of these two facilities and a logical link to the USA NVLAP PT conducted at the USA facility. Additionally, the description of the tests and establishment of performance criteria will disclose the seriousness and rigor to which this activity was held. Finally, the data will demonstrate that not only are such tests possible, but also the degree of equivalence attained can be very high.

**Keywords** Accreditation · Calibration · Competency · Equivalency · Fixed-point cell · Interlaboratory comparison · NVLAP · SPRT · UKAS

## 1 Introduction

In order to achieve and maintain accreditation, one of the requirements of ISO 17025:2005 is that monitoring of the laboratory be a planned and reviewed activity. Interlaboratory comparisons and proficiency tests (PTs) can serve this purpose [1].

---

T. J. Wiandt (✉)
Hart Scientific Division, Fluke Corporation, American Fork, UT 84003, USA
e-mail: tom.wiandt@hartscientific.com

Generally, the PT consists of the laboratory under test calibrating one or more measurement standards for which the characteristics are known. The results are evaluated and pass/fail criteria are established based on the normalized error, denoted $E_{\mathrm{normal}}$ or $E_{\mathrm{n}}$. Because of the expense, difficulty, and expertise required, PTs for SPRT calibrations at the lowest levels of uncertainty are not generally available.

The Hart Scientific Division of the Fluke Corporation operates two accredited SPRT calibration facilities, one at the Hart Scientific factory in Utah, USA, and the other at a service facility in Norwich, UK. The USA facility is accredited through NVLAP, whereas the UK facility is accredited through UKAS. Both provide SPRT calibrations using similar equipment and procedures, and with similar uncertainties. These uncertainties are among the lowest commercially available [2]. Consequently, the PT requirements are extremely demanding. In the USA, a suitable PT is available through NVLAP based on the long-standing NIST SPRT measurement assurance program [3,4]. Thus, a conventional PT was conducted in the USA facility. No such PT is offered as a regular service in the UK or within EUROMET. All attempts by the author to achieve the goals of the PT using conventional methods proved unsuccessful. Consequently, it became clear that these goals had to be met using a different approach.

To further complicate matters, a requirement exists for the fixed-point cells used in the SPRT calibration process to be verified periodically [5]. Due to the delicate nature of these cells and the difficulty of transport, the traditional approach of returning the cells to a central location to accomplish the verification is extremely inconvenient. Again, it was determined that this requirement had to be met using a different approach.

Finally, in addition to demonstrating competency for the purpose of accreditation, the additional challenge faced by these two laboratories is one of the demonstrating equivalencies to interested customers. It is important from the customer's viewpoint that calibrations provided by the two laboratories be equivalent relative to the stated uncertainties in order to satisfy the requirements of the SPRT user while maintaining operational efficiency.

## 2 Measurements

### 2.1 SPRT Calibration Comparison

### *2.1.1 Strategy*

For the SPRT portion, it was determined that an interlaboratory comparison between the US and UK facilities, when taken in conjunction with the NVLAP PT conducted at the US facility, could accomplish the purpose. This section describes the interlaboratory comparison, the results of the comparison, and the results of the NVLAP PT upon which the comparison was based.

### *2.1.2 Measurements*

The NVLAP PT is available for several ranges of temperature. The range selected for this test should cover the range of accreditation. In our case, we needed the widest span available; from the triple point of argon to the freezing point of aluminum, approximately −190 °C to 660 °C. This corresponds to ITS-90 ranges commonly referred to as 4 and 7 [6]. It made sense to apply this same range to the interlaboratory comparison. The NVLAP PT utilizes three SPRTs, calibrated over this entire range. Since SPRTs are available with glass or steel sheaths, and it is reasonable to assume that calibration lab performance might be different as the calibration applies to differing types, it was decided to include both types in the NVLAP PT. These SPRTs belong to NIST, and were not all manufactured by Hart Scientific. The use of multiple SPRTs proved impractical for the interlaboratory comparison; consequently, one artifact was used (as is common with most PTs). We desired the most conclusive of results, so we elected to use the best available SPRT from our product portfolio that covers this range—a Model 5681 glass-sheathed SPRT.

First, to ensure that the SPRTs were stable and arrived without damage, the initial resistance at the triple point of water, $R_{\text{TPW}}$, was measured and compared to the previous value. In the case of the NVLAP PT, the values were provided to NIST, and NIST performed the check. Calibration commenced in the conventional manner once the stability of the SPRTs was demonstrated. The repeatability requirement for the SPRT before calibration can begin is 0.25 mK [2]. The procedures used for the actual calibration were essentially conventional SPRT calibration procedures with the exception that all of the ITS-90 fixed points in the range were included. In the case of the NVLAP PT, additional checks were employed to verify proper application of the corrections and to validate the mathematical operations. These additional steps were not necessary for the interlaboratory comparison because both laboratories employ the same internally written software. After calibration was completed, the SPRTs and calibration results were returned to the reference laboratory. In the case of the NVLAP PT, NIST recalibrated the SPRTs to demonstrate that stability was maintained throughout the process. This was not possible in the case of the interlaboratory comparison because the SPRT was required for another project.

### *2.1.3 Results*

The NVLAP PT report includes the measurement results, details pertaining to the ITS-90 fixed-point cell corrections and mathematics, information related to the redundant fixed points, and data related to the stability of the SPRTs involved [7]. This paper describes only the measurement results, including the redundant fixed points.

PT results are generally evaluated using the normalized error, denoted $E_{\text{normal}}$ or $E_{\text{n}}$ [8,9]. The normalized error is the ratio of the difference in the measurement results relative to the combined measurement uncertainties. $E_{\text{n}}$ is calculated as follows:

$$E_{\text{n}} = \frac{\left(x_i - x_{\text{r}}\right)}{\sqrt{(U_i)^2 + (U_{\text{r}})^2}}, \tag{1}$$

where $E_{\text{n}}$ is the normalized error, $x_i$ is the measurement result from the laboratory under evaluation, $x_{\text{r}}$ is the measurement result from the reference laboratory, $U_i$ is the expanded uncertainty of the measurement under evaluation ($k = 2$), and $U_{\text{r}}$ is the expanded uncertainty of the reference measurement ($k = 2$).

**Table 1** SPRT uncertainty components and correlation assumptions

| Uncertainty component | Type | Correlation assumption |
|---|---|---|
| Process variability as observed by check standard SPRT | A | Uncorrelated |
| Precision of measurement (procedure limit $n = 40$) | A | Uncorrelated |
| Fixed-point value (reference cell certification) | B | Partially correlated |
| SPRT self-heating correction | B | Fully correlated |
| Hydrostatic head correction | B | Fully correlated |
| Non-ideal immersion profile | B | Fully correlate |
| $R_{\mathrm{TPW}}$ propagation | B | Uncorrelated |
| Shunt losses | B | Fully correlated |
| Bridge nonlinearity | B | Uncorrelated |
| Reference resistor instability during process | B | Uncorrelated |

When $|E_{\mathrm{n}}| \leq 1$, the comparison is deemed successful, and when $|E_{\mathrm{n}}| > 1$ the results are considered unacceptable. In most cases, this outcome is considered conclusive even when the two uncertainties are similar in magnitude. However, when $E_{\mathrm{n}}$ is close to 1, ambiguity can arise [10]. When this occurs and the uncertainty in the reference value is small relative to the uncertainty in the unknown value, it can be logically concluded that the unknown value is suspect because the reference uncertainty does not contribute much to the combined uncertainty. However, when the two uncertainties are similar in magnitude, the result may be inconclusive because one may not be able to determine which value is correct without additional evidence [10]. Nevertheless, in the case of a conventional PT, one must assume that the reference laboratory is correct and the participating laboratory is in error. Consequently, the participating laboratory must take all appropriate measures before doubt can be placed upon the reference laboratory or the reference values.

Formally speaking, in the case of the NVLAP PT, the issue with regard to the magnitude of the uncertainties is irrelevant because, as a formal PT, the results will stand on their own. Nevertheless, the NIST uncertainties are significantly smaller than the uncertainties of the laboratory under test; therefore, a conclusive result (successful or unsuccessful) is expected. In the case of the interlaboratory comparison, the uncertainties of the two labs are essentially identical and the nature of the comparison is informal; consequently, an inconclusive result is possible.

Finally, since both Fluke laboratories are traceable through one set of cells and apparatus, it is expected that some of the individual components of uncertainty may be correlated. If this is the case, the $E_{\mathrm{n}}$ calculation may under-represent the actual errors. Consequently, we attempted to identify the components that might be correlated and remove them from the result. The components of uncertainty, along with the correlation assumptions, are shown in Table 1. Since equivalence of the two laboratories is very important to our customers, we decided to take a very conservative position on the value of $E_{\mathrm{n}}$. Although values of $E_{\mathrm{n}}$ between $-1$ and 1 are considered passing results, we decided to evaluate any conditions where $E_{\mathrm{n}}$ is between $-1$ and $-0.5$ or 0.5 and 1 with the intent to improve the equivalence.

The results of the comparison experiments are shown in Tables 2 and 3 and graphically in Fig. 1. The results of the redundant fixed-point measurements at the Ga MP

**Table 2** SPRT NVLAP PT results

| ITS-90 fixed point | $\Delta W_{T90}$ (mK) | $U_{NIST}$ $(k = 2)$ (mK) | $U_{HART}$ $(k = 2)^{a}$ (mK) | $U_{C}$ $(k = 2)$ (mK) | $E^{b}_{normal}$ |
|---|---|---|---|---|---|
| $LN_2$ | −0.28 | 0.14 | 0.60 | 0.62 | −0.5 |
| Hg TP | −0.04 | 0.15 | 0.40 | 0.43 | −0.1 |
| Ga MP | −0.02 | 0.07 | 0.40 | 0.41 | 0.0 |
| In FP | −0.04 | 0.18 | 0.90 | 0.92 | 0.0 |
| Sn FP | 0.00 | 0.28 | 0.90 | 0.94 | 0.0 |
| Zn FP | −0.53 | 0.51 | 1.10 | 1.21 | −0.4 |
| Al FP | −0.88 | 0.79 | 2.10 | 1.24 | −0.5 |

[a] Uncertainties shown are those on the laboratory scope of accreditation (NVLAP lab code 200348)

[b] $E_{normal}$ values shown were calculated with the accredited uncertainties rather than the preliminary uncertainties estimated at the time of the PT. Therefore, $E_{normal}$ values are negligibly different from the values shown on the PT report

**Table 3** SPRT interlaboratory comparison results

| ITS-90 fixed point | $\Delta W_{T90}$ (mK) | $U_{HARTAF}$ $(k = 2)^{a}$ (mK) | $U_{HARTUK}$ $(k = 2)^{a}$ (mK) | $U_{C}$ $(k$=2) $U_{C}(k$=2) (mK) | $E_{normal}$ |
|---|---|---|---|---|---|
| $LN_2$ | 0.23 | 0.55 | 1.55 | 1.64 | 0.1 |
| Hg TP | 0.18 | 0.30 | 0.30 | 0.42 | 0.4 |
| Ga MP | −0.07 | 0.35 | 0.35 | 0.49 | −0.1 |
| In FP | 0.13 | 0.69 | 0.69 | 0.98 | 0.1 |
| Sn FP | −0.25 | 0.83 | 0.83 | 1.17 | −0.2 |
| Zn FP | −0.51 | 1.05 | 1.05 | 1.48 | −0.3 |
| Al FP | −0.52 | 1.83 | 1.93 | 2.66 | −0.2 |

[a] Uncertainties represent those on the respective scopes of accreditation with the correlated components removed (NVLAP lab code 200348, UKAS certificate number 0775)

and In FP are shown in Fig. 2. The comparison experiments demonstrate definitive agreement both between the NIST and the Hart Scientific USA facility and between the Hart Scientific USA and the UK facilities. When taken in conjunction with our rigorous cut-off criterion of ±0.5 for $E_n$, we conclude that the PTs were successful and no further action is required. Additionally, in both cases, the non-uniqueness is consistent with expectations, demonstrating good internal consistency in the calibration process.

## 2.2 Fixed-point Cell Comparison

### *2.2.1 Strategy*

The USA facility maintains three sets of fixed-point cells and the UK facility maintains two sets of fixed-point cells. In the USA facility, these cells functioning as the primary set, working set, and SPRT calibration set. In the UK facility, these cells functioning as an SPRT calibration set and backup and cross-check set. The USA primary set has been tested at NIST. The use of these cells is restricted to the certification of the working cells, the SPRT calibration cells, the cells for the UK facility, and newly purchased cells for selected customers who require the lowest possible uncertainties (primarily NMI customers). The USA working cells are used for the routine certification

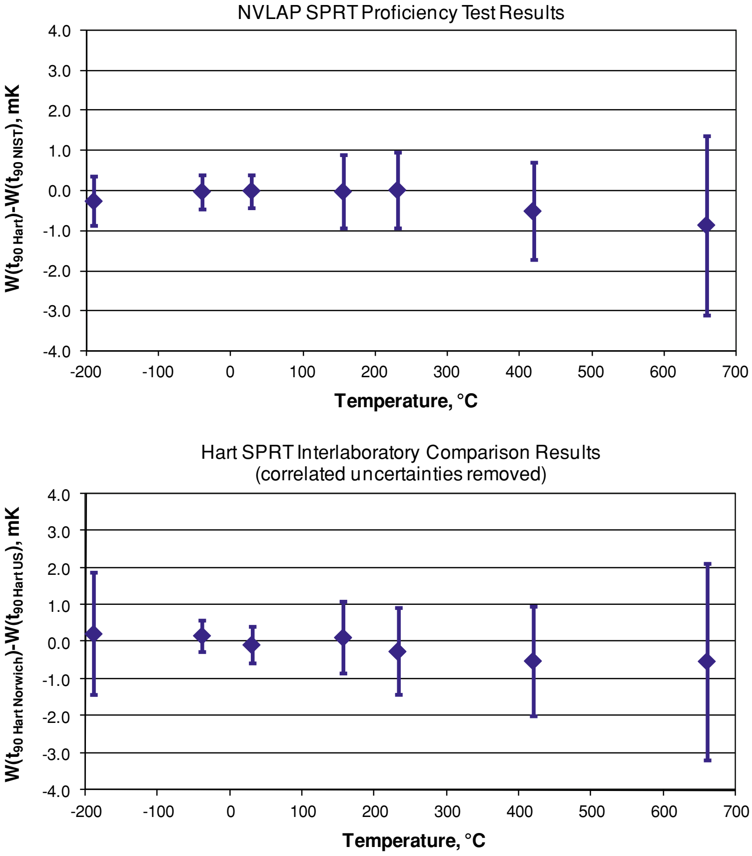

**Fig. 1** NVLAP SPRT PT results and Hart SPRT interlaboratory comparison results, both at 0 mA, covering the range from $-200\,^{\circ}\mathrm{C}$ to $660\,^{\circ}\mathrm{C}$. Data points represent the error in mK at the individual fixed points $(T_{\mathrm{Lab}} - T_{\mathrm{Ref}})$, and error bars denote the combined uncertainty $U_{\mathrm{C}}$

of customer fixed-point cells and as cross-check cells for the SPRT calibration cells. Since the primary set was tested at NIST, the uncertainties assigned to these cells are smaller than the other sets. The uncertainties assigned to the other sets are comprised mainly of the uncertainties attributed to the primary cells and the uncertainties of the comparison process. As a result, the uncertainties assigned to these sets are nominally identical. If the uncertainties of the comparison process are kept as small as possible, equivalency among the cells should be fairly straightforward to demonstrate.

Although the fixed-point cells can be assigned an uncertainty based upon purity, construction, and other characteristics, we find it simpler to treat the cells as calibrated artifacts. This approach is somewhat unconventional but makes the traceability and uncertainty analysis more direct. In line with this approach, the certification includes the temperature difference $(\Delta T)$ observed during the comparison experiment, corrected to $\Delta T$ from the ITS-90 nominal value, along with the uncertainties. The uncertainties propagate from the uncertainties of the original NIST certification and the uncertainties of the various comparison experiments. For most cells, the observed

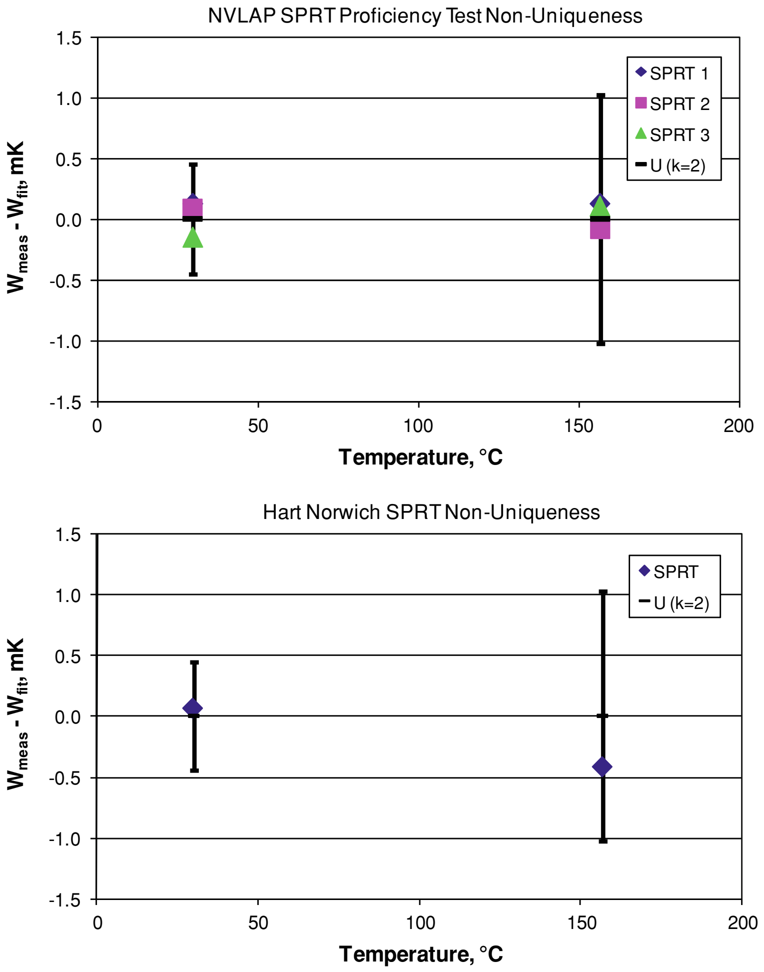

**Fig. 2** Error/non-uniqueness in the SPRT calibrations at the Ga MP and In FP relative to the RSS propagated subrange uncertainty. Data points represent the error in mK ($T_{\text{calculated}} - T_{\text{measured}}$), and error bars denote the RSS combined uncertainty $U_C$

$\Delta T$ is small relative to the uncertainties of the comparison experiment. Traceability to NIST is established through an unbroken chain of comparisons in the conventional manner for calibrated instruments.

Finally, both NVLAP and UKAS require periodic verification that the cells in daily use are stable over time. It was determined that this requirement could be met by alternating semi-annual plateau evaluation with semi-annual comparison of the working cells with new cells or the backup cells.

### *2.2.2 Measurements*

As described previously, the various sets of cells have been certified with the $\Delta T$ from the nominal ITS-90 temperature. Therefore, the difference in observed temperature of any two cells can be calculated and compared. The results can then be evaluated relative to the uncertainties to determine the success of the comparison. When the $\Delta T$ exceeds the expectation, the test can be considered unsuccessful, whereas when the $\Delta T$ is smaller than the expectation, the test can be considered successful.

The measurements were conducted using an ASL F18 or MI 6010T bridge (or both), SPRTs known to be stable at the temperatures of interest, thermally regulated reference resistors, and appropriate realization apparatus. In all cases, multiple SPRTs were used for each cell. For direct comparison, it has been suggested that $R_{\mathrm{T90}}$ is superior to $W_{\mathrm{T90}}$ in detecting small differences [11]. However, to ensure SPRT stability during the comparison process, $R_{\mathrm{TPW}}$ was measured at the opening and closing of each fixed-point cell comparison experiment. The bridges were controlled using software both to reduce the possibility of operator error and to improve the resolution and reproducibility of the results. The software used with the ASL F18 was written in-house. The software used with the MI 6010T is commercially available software purchased with the bridge. In order to ensure that the plateaux were evaluated at the identical percent of sample frozen (or melted), the initiations of the plateaus were offset by the time interval required to complete one measurement sequence. The measurements were executed with two currents and the results extrapolated to zero power. The zero-power values were taken as the values representing $R_{\mathrm{T90}}$. In order to ensure the achievement of thermal equilibrium, the measurement sequence consisted of three elements; nominal power, double power, and nominal power. Thermal equilibrium and thermal stability were verified, to the extent possible, before the measurement was accepted [2]. Once the data were obtained, the measured $\Delta T$ values were compared to the calculated $\Delta T$ values.

### *2.2.3 Results*

The normalized error parameter, $E_{\mathrm{n}}$, will be used to demonstrate equivalence. However, unlike the comparison of SPRTs, the comparison of fixed-point cells within the individual laboratories should not contain significant correlated uncertainties. However, $E_{\mathrm{n}}$ was calculated using both the combined uncertainties, $U_{\mathrm{C}}$, and the uncertainties of only one lab, $U_{\mathrm{Norwich}}$. Nevertheless, the formal decision regarding success must be based on $U_{\mathrm{C}}$ as described in Eq. 1.

The results are shown in Table 4 and in Fig. 3. In both cases, the comparison experiments demonstrate definitive agreement between the two Hart Scientific facilities. When taken in conjunction with our rigorous cut-off criterion of $\pm 0.5$ for $E_{\mathrm{n}}$, we conclude that the comparison tests were successful and, with the exception of including the Al FP and $\mathrm{LN}_2$ comparison later this year, no further action is required.

## 3 Conclusions

Through the experiments described above, it can be concluded that the two Fluke Corporation laboratories show excellent equivalence both in SPRT calibrations and in fixed-point comparison tests. Furthermore, the NVLAP PT results demonstrate excellent equivalence between the USA facility and the NIST.

**Table 4** Fixed-point cell interlaboratory comparison results

| Fixed Point | $\Delta t_{\text{calculated}}$ (mK) | $\Delta t_{\text{measured}}$ (mK) | $\Delta t_{\text{difference}}$ (mK) | $U_{\text{AF}}$ ($k = 2$) (mK) | $U_{\text{UK}}$ ($k = 2$) (mK) | $U_{\text{C}}$ ($k = 2$) (mK) | $E_{\text{n}}^{\text{a}}$ | $E_{\text{n}}^{\text{b}}$ |
|---|---|---|---|---|---|---|---|---|
| Hg TP | −0.14 | −0.20 | −0.06 | 0.20 | 0.20 | 0.28 | −0.2 | −0.3 |
| Ga MP | −0.07 | −0.11 | −0.04 | 0.08 | 0.08 | 0.11 | −0.4 | −0.5 |
| In FP | −0.47 | −0.65 | −0.18 | 0.50 | 0.50 | 0.71 | −0.3 | −0.4 |
| Sn FP | −0.24 | −0.09 | 0.15 | 0.60 | 0.60 | 0.85 | 0.2 | 0.3 |
| Zn FP | −0.65 | −0.44 | 0.21 | 0.80 | 0.80 | 1.13 | 0.2 | 0.3 |

The formal acceptance criteria are based upon $E_{\text{n}}$ calculated using the combined uncertainty ($U_{\text{C}}$) as shown in Eq. 1, and denoted as $E_{\text{n}}^{a}$ in the table rather than $E_{\text{n}}$ calculated using the Norwich-only uncertainties ($U_{\text{UK}}$) and denoted as $E_{\text{n}}^{\text{b}}$ in the table. These values are shown for completeness (as they were included in the official report to UKAS) and to suggest "worst case" values of $E_{\text{n}}$

[a] $E_{\text{n}}$ calculated using combined uncertainty, $U_{\text{C}}$

[b] $E_{\text{n}}$ calculated using individual uncertainty of Norwich laboratory, $U_{\text{UK}}$

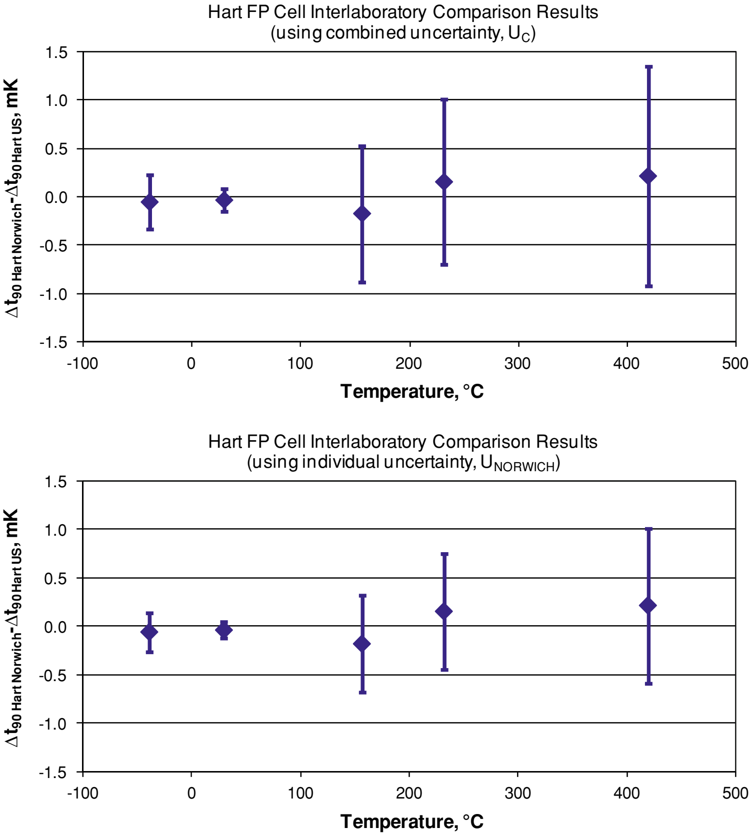

**Fig. 3** Hart fixed-point cell interlaboratory comparison results covering the cells Hg TP through Zn FP, illustrating both the combined uncertainty $U_{\text{C}}$ and the individual uncertainty $U_{\text{Norwich}}$. Data points represent the error in mK ($T_{\text{Lab}} - T_{\text{Ref}}$), and error bars denote the uncertainty $U_{\text{X}}$

**Acknowledgments** The author would like to thank Miss Alison Shrimpling of our Norwich facility for her assistance and participation in conducting the majority of the measurements in the Norwich laboratory, as well as Greg Strouse of NIST for his guidance in designing the experiment and tutelage in interpreting the NVLAP PT results.

## References

1. ISO/IEC 17025:2005, General requirements for the competence of testing and calibration laboratories (2005)
2. T. Wiandt, Int. J. Thermophys. **29**, 890 (2008)
3. NVLAP Lab Bulletin LB-10-2004 Thermometry Proficiency Tests (2007)
4. G.F. Strouse, NIST measurement assurance program for the ITS-90. Presented at Test and Calibration Symposium (1995)
5. C.D. Faison, C.S. Brickenkamp (eds.), *NIST Handbook 150-2H, Calibration Laboratories Technical Guide for Thermodynamic Measurements* (2004)
6. B.W. Mangum, G.T. Furukawa, "Guidelines for Realizing the International Temperature Scale of 1990 (ITS-90)," *NIST Tech. Note 1265* (1990)
7. NVLAP Calibration Laboratories Accreditation Program Proficiency Test Report, Test folder number 836/270208-04 (2004)
8. ISO/IEC Guide 43:1997, Part 1: development and operation of proficiency testing schemes (1997)
9. ILAC-G22:2004, Use of proficiency testing as a tool for accreditation in testing (2004)
10. A.G. Steele, R.J. Douglas, Metrologia **43**, S235 (2006)
11. G.F. Strouse, in *Proceedings of TEMPMEKO 2004,* 9$^{th}$ *International Symposium on Temperature and Thermal Measurements in Industry and Science*, ed. by D. Zvizdić (FSB/LPM, Zagreb, Croatia, 2004), pp. 879–884