TEST AND CALIBRATION OF DISPLACEMENT MEASURING LASER INTERFEROMETERS

Jack Stone
Semiconductor and Dimensional Metrology Division, National Institute of Standards and Technology, 100 Bureau Dr. Stop 8211, Gaithersburg, MD, USA. jack.stone@nist.gov

ABSTRACT: At the National Institute of Standards and Technology (NIST), we have a capability to calibrate and test laser interferometer systems used to measure displacements. Recently our calibration protocol has been modified so as to bring it into accord with a new standard, ANSI/ASME B89.1.8-2011. In addition to quantifying length-proportional errors, our measurements quantify a zero-drift parameter that was not previously measured but can be a significant source of uncertainty for some types of interferometers. The essence of the calibration is a comparison of a customer interferometer to a master interferometer maintained by NIST. The two interferometers are arranged in a back-to-back geometry that is compensated so as to reduce typical sources of drift. The primary advantage of this geometry is its flexibility to measure any customer displacement interferometer.

Key words: laser interferometry, displacement measurement, dimensional metrology.

1. INTRODUCTION

Laser interferometers are an indispensable part of length measurement at the highest levels of accuracy. In particular, displacement measuring interferometers are commonly used either directly as a scale for a measuring system or indirectly to verify measurement accuracy (e.g., error mapping of a coordinate measuring machine). In one manner or another, the traceability chain for a large number of length measurements relies on laser displacement interferometry, and it must be verified that the laser interferometer is itself traceable to national and international standards.

Laser interferometers all measure distance or displacement in terms of the wavelength of light, almost always in air, and consequently the metric of the measurement is set by the air wavelength \( \lambda_{\text{air}} \), which must be tied back to national standards and the SI units. This quantity in turn depends on the vacuum wavelength \( \lambda_{\text{vac}} \) and the air refractive index \( n \), where \( \lambda_{\text{air}} = \lambda_{\text{vac}}/n \). The vacuum wavelength can be determined with negligible uncertainty through a frequency comparison to an iodine stabilized laser, which is an internationally accepted frequency standard [1]. The refractive index is usually determined from air temperature, pressure, and humidity measurements, via a calculation employing one of several versions of the Edlén equation or Ciddor equation [2]. Thus, a clear chain of traceability of the interferometer system to national standards might be established through calibrations of atmospheric sensors in combination with measurements of the laser vacuum wavelength. For many situations this is an entirely reasonable approach to achieving traceability, but there are some situations where additional testing would be desirable. The calibrations described above cannot fully verify that an arbitrary interferometer will perform according to its specifications, because an interferometer system requires a complex integration of sensors, electronics for counting and interpolating fringes, and software for data capture and processing. This is not always implemented correctly, and consequently it may be important to carry out a test of overall system performance to verify that all sub-systems operate correctly as a unified whole. Such a test has been described in a new standard, “Performance Evaluation of Displacement-Measuring Laser Interferometers.” (ANSI/ASME B89.1.8 – 2011) [3]. (This standard will subsequently be referred to as B89.1.8.) Typically, an overall system test at ambient laboratory conditions is supplemented by calibration of individual sensors over a wide operating range to determine errors and uncertainty of the interferometer system over its entire operating range. The NIST Dimensional Metrology Group performs a system test but does not calibrate individual sensors; the purpose of this article is to describe how NIST implements the system test. The B89.1.8 standard allows for combining results from a system test at one laboratory with sensor calibrations done by other laboratories to give the final results.

2. GEOMETRY OF THE COMPARISON SET-UP

At NIST, the B89.1.8 overall system test employs a set-up as shown in Fig. 1, referred to as a compensated back-to-back comparison. Essentially, a master interferometer is set up back-to-back with the test interferometer, with retroreflectors for the two systems mounted on a moving carriage. When the carriage is moved, the two interferometers ideally register displacements of opposite sign but equal magnitude, so that the sum of the two readings is constant; any observed deviation from this expectation is interpreted as an error in the test interferometer’s measurement. The reference arm of the master interferometer has been modified so that the retroreflector is rigidly mounted as close as possible to the reference arm of the device under test [4]. This compensates for possible changes in the total optical path between the two interferometers due to uniform changes in air refractive index or uniform thermal expansion of the machine base. This particular arrangement has the primary

![Fig. 1: Compensated back-to-back comparison.](image-url)
advantage that it can in principle test any displacement measuring interferometer, independent of any details of the system under test, because the test system optics is entirely independent of the master interferometer system. The test system is exactly the same as it is for normal operation, with no additional optics required within the beam path (as is required for some other comparison testing schemes). For our immediate purposes an important consideration is that we can compare the master to commercial systems that have the reference path built into the laser head. This capability is important because, if the commercial interferometer is not designed well, it could be subject to much drift during its warm-up period, and it is important to quantify this drift.

Other possible comparison techniques do not fully share this flexibility but can more easily achieve high accuracy. Several alternative geometries are possible with a tradeoff between flexibility and accuracy [3-6]. The very highest accuracy can be obtained in an arrangement such as shown in Fig. 2, where the master and test laser are set up so as to share the same optics and follow the same air paths, eliminating errors associated with Abbe offset, temperature gradients, and mechanical instabilities [5]. This set-up can achieve very high accuracy with minimal effort and minimal demands on equipment; it is an attractive approach if all interferometer systems to be calibrated are of similar design, compatible with the optical layout. We have employed common optics comparisons to help verify the accuracy of our master interferometer but do not use this technique for normal customer calibrations, where it is important to maintain a flexibility to measure any kind of system. The flexibility of our compensated back-to-back geometry is achieved at the expense of sensitivity to sources of error that are easily eliminated using the common-optics geometry, but these error sources nevertheless can be kept small enough that they do not greatly increase the overall uncertainty of a calibration.

3. MEASUREMENT PROTOCOLS

The essence of a measurement is simply to move the carriage through a series of displacements up to a maximum value (approximately 4 m for our apparatus), stopping periodically to read the two interferometers statically so as to avoid a need for careful synchronization of the readings from the two interferometers. The measured errors (sum of master and test interferometer readings or difference in magnitudes) are fit to give a slope and intercept. Another set of measurements is made returning to the original zero position. The out-and back “runs” are repeated 5 times and the resulting 10 runs are analyzed separately to give a slope (length-proportional error) and intercept for each run. The timing of runs is arranged so that measurements are uniformly distributed in time and the full set of measurements is spread over 1 hour. A sample set of 10 runs is shown in Fig. 3. This is a comparison of two NIST interferometers, where both are compensated to the same refractive index. One is our usual master interferometer and the second plays the role of a customer interferometer under test. The B89.1.8 test would normally call for comparing a customer interferometer using its own atmospheric compensation to the NIST system compensated using NIST sensors and software, so that the accuracy of compensation can be evaluated. A test with both interferometers compensated to the same conditions, as in Fig.3, may also be carried out as a recommended diagnostic, but it does not include the effects of sensor measurement errors on compensation.

The average slope of the 10 best-fit lines is reported as the length-proportional error of the test interferometer at ambient conditions, denoted as $LD_{Ec}$ (Length Dependent Error measured by the comparison test). Here the average slope gives $LD_{Ec} = 8.8 \times 10^{-8}$, or equivalently, 0.088 μm/m. This error mostly arises from the fact that the actual laser vacuum wavelength differs from the value used by the interferometer firmware. After correcting for this effect, the slope is reduced to $1.7 \times 10^{-8}$, where this residual slope is consistent with what could be expected due to uncertainty of the measurement.

Normally, when comparing to a customer interferometer that uses independent sensors for determining refractive index, $LD_{Ec}$ includes compensation errors of the interferometer at ambient conditions. Additional errors may
occur if the interferometer is used under atmospheric conditions that differ greatly from what occurred at the time of testing, and this can be determined by calibration of the atmospheric sensors over the full operating range, as mentioned previously.

For each of the ten runs, an intercept is also calculated. A constant intercept has no significance, but variations in the intercept over time are indicative of the zero drift mentioned previously. The full range of the 10 intercept values is reported as the “zero drift.” It can be expected that this would be greatest during the first hour of operation, and most dimensional measurements can be completed within 1 hour, which is the reason why the B89.1.8 standard specifies that measurements are to be spaced over a 1 hour interval. For the interferometers tested here, which do not have an internal reference path, the only zero drift expected would be associated with noise of the measurement, but a systematic drift in the data is in fact observed, giving rise to an apparent zero drift of 50 nm. Mounting instability of optical components is probably responsible for this drift, which limits our ability to meaningfully measure the true interferometer zero drift below 100 nm.

4. UNCERTAINTY EVALUATION

When performing the system test, there are two basic elements to an uncertainty budget: uncertainty of the master interferometer and uncertainty associated with the comparison process. In addition, the B89.1.8 standard calls for including an additional Type A uncertainty evaluated from the standard deviation of the slopes (LDEc values) of the ten runs.

4.1 Uncertainty of the master interferometer.

The largest source of uncertainty in the master interferometer is atmospheric compensation, which depends primarily on the calibration uncertainty for the atmospheric sensors (pressure, temperature, and humidity) and on the stability of the sensors following calibration. This uncertainty may be as large as several parts in 10^6 for commercial systems. Smaller uncertainties (< 5 parts in 10^8) are associated with the Edlén equation and with uncertainties in atmospheric composition, such as variable CO₂ concentration (which is usually not measured). For comparisons such as shown in Fig. 3, which do not include effects of atmospheric compensation, the primary source of uncertainty in the master interferometer is the calibration of vacuum wavelength (typically below 1 part in 10^8 fractional uncertainty) and its stability between calibrations (typically a few parts in 10^9). Additional small uncertainties are fringe interpolation errors (typically below 10 nm) and effects of beam collimation (typically a few parts in 10^9).

If a testing laboratory is to offer calibrations in accord with the B89.1.8 standard, it is required that their master interferometer system’s uncertainty should be verified by comparison to another interferometer system whose uncertainty has been previously verified, where verification requires a comparison of the type described in this paper.

4.2 Uncertainty of the comparison process.

The primary uncertainties of the comparison process include mechanical and thermal-induced distortions or drift, alignment errors, Abbe errors, and imperfect compensation of variations in refractive index. An uncertainty budget for the comparison can be developed based on estimates of the magnitudes of these errors. Some of the individual errors are difficult to quantify with confidence, but it is also possible to directly test the overall accuracy of the comparison procedure and thus verify that the uncertainty budget is plausible. This can be done by comparing two interferometers of good quality, as was shown in Fig. 3. After correcting for the vacuum wavelength of the lasers, the residual slope was 1.7×10^-9, and this length-proportional error can be attributed to the comparison process. Actually, the error of the comparison process could be determined even if the vacuum wavelengths of the lasers were unknown. Errors that are associated with the comparison process can be distinguished from differences intrinsic to the interferometers by switching the positions of the two interferometers while keeping all mounting hardware and other conditions as similar as possible. In fact, it usually suffices simply to leave the optics in place and exchange the positions of the master and test laser. If one interferometer indicates a larger displacement than the second in both positions, then the difference is intrinsic to the interferometers, such as an error in vacuum wavelength. If the interferometer that reads a larger displacement in one position indicates a smaller displacement in the second position, then the error is intrinsic to the comparison process, such as mechanical or thermally induced changes in optical mounts that are correlated with carriage position. In performing this switch when using heterodyne interferometers, it is necessary to keep in mind that, depending on the orientation of the polarizing beamsplitters, the polarization component travelling the measurement arm might be different following the switch, and the difference in frequency/vacuum wavelength must be taken into account if the highest accuracy is desired.

If such comparisons are carried out multiple times, exchanging the positions of the master and test interferometer, then most important errors in the comparison process can be sampled and must be consistent with the claimed uncertainty of the comparison.

For our apparatus, the agreement in LDEc between results with the two interferometers exchanged can be kept below 5×10^-9 when great care is taken with the set-up and thermal management. Under more typical conditions, the agreement is not so good. In particular, there appear to be thermal related errors as large as 1.5×10^-8 if we do not take great care with thermal management and do not employ a fan to help reduce temperature gradients along the length of the waybed. This can be done only at the expense of
increased turbulence in the beam paths (in spite of shielding) and is not used for routine work.

It may be noted that any calibration laboratory can verify the uncertainty of their comparison procedure by carrying out such a test, and this is required in the B89.1.8 standard. With this uncertainty verified, the calibration results of the laboratory will be reliable if the laboratory’s master interferometer is verified by comparison to another interferometer at a national laboratory. However, note that these tests only set a floor on the claimed uncertainty of the calibration lab for routine testing; a judicious uncertainty estimate will likely be larger to account for factors that may be somewhat larger during routine calibrations than under the special conditions of testing.

5. CONTROLLING SOURCES OF UNCERTAINTY

This section briefly discusses sources of error in the comparison procedure and how they can be minimized.

5.1 Mechanical and thermal variations of the total optical length.

The comparison will be in error if there is drift in the optical length of the test interferometer that is not compensated by a corresponding change in the master interferometer. Mechanical instability or thermal expansion is mostly compensated in the arrangement of Fig. 1. Uniform thermal expansion of the waybed is compensated; any increased length in the sum of the two measurement arms is compensated by a corresponding increase in the length of the reference arm of the master interferometer. This compensation is exact if the length of this reference arm is just equal to the sum of the two measurement arms. This statement is true in regards to both changes in the physical length resulting from thermal expansion and changes in optical length as a consequence of changes in refractive index with temperature (or pressure). In fact, both spatial and temporal variations in the air refractive index will not affect the comparison results except when there are temporal variations in the spatial gradient of refractive index between the master interferometer reference arm and the parallel path of the measurement arms of the two interferometers (about 3 cm separation). One manifestation of this problem is a sensitivity to air turbulence. We reduce air turbulence (and other effects associated with changing temperature) by covering the air path and optics with insulating shielding. Compensation is never exact. The air paths in our apparatus are not perfectly compensated; the lengths may differ by a few centimeters depending on details of the set-up. This is of only secondary importance, but might require a small deadpath correction needed when large variations in atmospheric pressure occur.

The master interferometer has the geometry of an angle-measuring interferometer, and any thermally or mechanically induced distortion of the mounts that causes a variation in angle between the two ends of the set-up will result in error. In fact our laboratory has good temperature control and variations in the temperature of hardware are small, below 0.02 °C at fixed locations. Spatial variations along the length of the machine can be several tenths of a degree under our usual operating conditions. If the mounting of the retroreflectors on the carriage has time to respond to these variations as the carriage moves from one end to the other, significant errors might occur. As stated previously, there is evidence suggesting that thermal problems can cause errors in LDEc at least as large as 1.5×10⁻⁵, and the uncertainty must be increased accordingly when thermal gradients are present.

Another concern is mechanical creep in the mounting of optics. This creep is set-up dependent and it is necessary to monitor such drifts and correct the set-up if persistent drifts are present. We commonly see drifts on the order of 50 nm/h, but the drift could be arbitrarily large if care is not taken in setting up. The drift over a single “run” would be only 1/10 of this value, not a great concern, but the drift over the full 10 runs is large enough that it limits our ability to measure zero drift.

5.2 Abbe Error

Abbe offset is not a great concern in our apparatus, but a few comments may be useful. Abbe offset is illustrated in Fig. 4, where it is assumed that the laser beams in the two interferometers are precisely aligned horizontally with the direction of motion. The Abbe offset is the vertical component of the distance between the nodal points (or “optical centers”) of the two back-to-back cube-corner reflectors. (For a hollow cube-corner, the nodal point is the physical tip, but for a solid glass reflector the nodal point is located at a point within the glass.) More precisely, this is the vertical component of the Abbe offset, and there is an analogous horizontal component. It may be difficult to set the Abbe offset precisely to zero if the retroreflector tips are not visible within the housing and the accuracy of their alignment relative to the housing is unknown. If the Abbe offset were as large as 1 mm, then pitch and yaw errors in our apparatus (< 70 μrad) could generate unacceptable errors as large as 70 nm.

In Fig. 4, the nodal points of the two cube-corners (shown as located at the tips) are separated by distance $L$ tilted at angle $\theta$ relative to the horizontal. The Abbe offset is the distance $\delta = L \sin \theta$. The distance between the tips along the direction of the laser beams is $L \cos \theta$. When the angle $\theta$ varies by a small amount $d\theta$, the resulting change in
the horizontal separation of the tips is given by the change in
the cosine function, with magnitude \( L \sin \theta d \theta = \delta d \theta \). This
is the familiar expression for the Abbe error. Note that the
interferometers measure distance along the direction of
propagation (horizontal), so the sum of the two
interferometer readings changes by \( \delta d \theta \) and produces this
error in the results.

If the retroreflectors are mounted on a rotary stage, then
the sum of the two interferometer readings will change as
the stage is rotated. We have observed the expected cosine
dependence of the error on the rotation angle. Abbe offset
can be eliminated by tracing out this cosine curve and
rotating the stage to the position where the sum of the
interferometer readings passes through a minimum. This
point—where the first derivative of the cosine curve
vanishes—corresponds to zero Abbe offset. The procedure
must be carried out twice, using rotations about orthogonal
axes, to eliminate Abbe offsets along both axes perpendicular to the beam. Without much difficulty the
Abbe offset can be reliably reduced to below 0.1 mm,
reducing the expected errors to less than 7 nm. The
uncertainty could be reduced quite a bit further if more
accurate measurements of the angle \( \theta \) were employed.

5.3 Non-simultaneous reading

For customer interferometer systems, it may be
impractical to fully automate data capture simultaneously
with the master interferometer; readings of the two
interferometers captured manually may differ in time by
a few tenths of a second. Different filtering time constants
in the two systems will have a similar effect. To avoid
problems with non-simultaneous readings, data are taken
under static conditions with the cart stopped. Turbulence and
vibration can cause errors that are at least partially
compensated when readings are simultaneous but the
compensation is less effective for non-simultaneous
measurements. Covering the air paths can significantly
reduce effects of turbulence. In our apparatus the
fluctuations of 1-second averages are typically about 10 nm
(1 standard deviation), where this number is a measure of
the combined effects of non-simultaneous measurements
along with noise and resolution of the phase interpolation.

When using atmospheric compensation, a further
complication in obtaining truly simultaneous measurements
is that thermistors in the master and test system may have
significantly different time constants, and in some cases the
update rate of the test system sensors may be slow. A slow
update of the pressure sensor, if coupled with pressure
variations that occur on windy days, can add significant
noise to the comparison.

5.4 Alignment errors

Cosine errors occur if the measurement beams of the
two interferometers are not perfectly aligned. In principle,
cosine errors are eliminated by making the two
measurement beams exactly parallel to each other, even if
they are not well aligned with the direction of motion of the
carriage, but in practice it is preferable to align both beams
as precisely as possible with the direction of motion. Alignment can be done most exactly by using a position
sensitive detector or quadrant detector to measure lateral
changes in the position of the beam from the retroreflector
as the carriage is displaced. When this is done, the primary
limitation on alignment accuracy lies not with the quadrant
detector but with the stability of the alignment. Drift of the
alignment is not expected to exceed 60 \( \mu \)rad, giving an
expanded uncertainty in the length-proportional error of
2x10^{-8}. If a quadrant detector is not used, it is necessary to
visually observe the reflected laser beam. If we assume that
a variation of spot position by 1.5 mm can be reliably
detected when the carriage is displaced through 4 m (8 m
round trip), this would give rise to an error \( \Delta L/L = 1.7 \times 10^{-8} \).
We take this number as the expanded uncertainty due to
alignment errors when the quadrant detector is not
employed. The problem would be much more severe (16x
greater) if the displacement of the carriage were limited to
only 1 m.

6. COMBINED UNCERTAINTY

Uncertainty sources when calibrating an interferometer
using our apparatus are summarized in this section. These
are all Type B expanded uncertainties estimated assuming
typical conditions of a normal customer calibration, which is
not the ultimate achievable uncertainty. First, consider a test
exclusive of atmospheric compensation; in effect, this
measures the errors of the customer interferometer if it were
to measure displacement in vacuum. Length-proportional
and length independent errors are summarized below.

Type-B expanded uncertainties for error sources independent
of length include:
(a) Master laser fringe interpolation:       10 nm
(b) Vibration/turbulence/ non-simultaneous reading: 30 nm
(c) Abbe offset:                          7 nm
(d) Bending of the apparatus with carriage motion: 15 nm
(e) uncompensated drift in mechanical or optical length
between interferometers (during 1 run):       20 nm

The estimates above are conservative according to our
current knowledge. The total length-independent uncertainty
of a given displacement measurement, found by adding the
above uncertainties as a root sum of squares, is 40 nm \( (k = 2) \).
All of the items above may give rise to deviations from a best-
fit line, and the final result is consistent with any deviations
from a straight line that we have seen in our data when
compensation is not employed.

For a test such as shown in Fig. 3 that does not test
atmospheric compensation, fractional expanded uncertainties
for length-proportional errors include:
(a) Uncertainty of master interferometer: 1.5x10^{-8}
(b) Alignment errors: 1.7x10^{-8}
(c) Thermal-related errors of comparison process: 2x10^{-8}
The total expanded uncertainty for measuring a displacement \( l \),
found by combining the length-dependent and length-
independent uncertainties in quadrature, is \( [3.0 \times 10^{-8} l]^{2} + [40 \times 10^{-8}] \).
nm)\(^{1/2}\). At 4 m, this uncertainty is 129 nm, or, expressed as a relative expanded uncertainty, 3.2×10\(^{-8}\). All of these uncertainties could be reduced if needed. The uncertainty of the master interferometer could be reduced as needed via more frequent calibration. Alignment errors assume that alignment is done by “eye” without the aid of a quadrant detector. As discussed previously, the thermal-related errors could be reduced with careful control of air flow and attention to heat sources within the room. Initial, very careful testing suggested that the uncertainty in LDE\(_C\) could be reduced below 1×10\(^{-8}\) with sufficient care, but this is not practical or needed for most calibrations. There is little motivation to reduce the uncertainty further because the primary measurement of interest is LDE\(_C\) when using atmospheric compensation, and the uncertainty in this measurement is dominated by uncertainties in atmospheric sensors. When atmospheric compensation is employed, additional uncertainties include the following expanded Type-B uncertainties. The expanded uncertainties are given both in terms of the measured parameter (pressure, temperature, etc.) and in terms of the corresponding fractional uncertainty in refractive index, or equivalently the uncertainty of LDE\(_C\).

(a) Edlén Equation: \(2\times10^{-8}\)
(b) Air Pressure: \(15\) Pa or \(4\times10^{-8}\)
(c) Air Temperature: \(0.03\) °C or \(3\times10^{-8}\)
(d) Humidity: \(3\%\) RH or \(3\times10^{-8}\)
(e) Atmospheric contaminants and unknown CO\(_2\) concentration: \(3\times10^{-8}\)

The expanded uncertainty given for the Edlén equation above is appropriate near standard atmospheric conditions. A larger expanded uncertainty (5×10\(^{-8}\)) is associated with the Edlén equation when environmental parameters and laser wavelength vary over a greater range. (This uncertainty also applies to the Ciddor equation.)

These uncertainties, combined (using root sum of squares) with the previous estimate of uncertainty for the case where compensation is not used, yield an expanded relative uncertainty of 7.6×10\(^{-8}\) for a comparison of the interferometers with atmospheric compensation in use.

7. CONCLUDING REMARKS

We have demonstrated that the comparison process itself has low uncertainty even in a system that is optimized for flexibility of operation at the expense of some loss of accuracy. To achieve high accuracy using this method requires some care, but any laboratory can demonstrate their ability to do comparisons of known accuracy by performing tests such as described here. The B89.1.8 standard sets out in detail how a calibration laboratory can thus establish a verified uncertainty for their calibrations. We may expect that some calibration laboratories or manufacturers of interferometer systems will set up testing facilities to verify interferometer operation for customers.

It is still an open question as to when this type of test is necessary. The classic approach of calibrating laser vacuum wavelength and sensors provides a transparent traceability path without the additional complication of an overall system test. From a strictly practical standpoint, it might be unnecessary to test the overall system integration more than once in the lifetime of the device, and one might go further to argue that it is not even necessary to test every individual interferometer if the overall system has been verified for some interferometers of the same model and software version. The added value of additional testing must be weighed against economic costs. The final decision, of course, will be in the hands of an assessor or auditor.

REFERENCES