NIST-Traceable Calibration of CD-SEM Magnification Using a 100 nm Pitch Standard

M. Tortonese, Y. Guan, J. Prochazka
VLSI Standards, Inc., 3087 North First Street, San Jose, CA 95134

ABSTRACT

This paper is a practical guide to the calibration of magnification of a CD-SEM using a pitch standard. It answers two fundamental metrology questions: 1) how many individual pitch measurements should one take in order to estimate the average pitch of the sample with a specified uncertainty and with a specified confidence level?, and 2) when is it appropriate to recalibrate the instrument following the measurement of the standard? In answering these questions, this paper identifies Cost of Ownership (CoO) elements of the calibration process and outlines best engineering practices for the calibration procedure. The discussion is then extended to the case of tool matching and calibration of not just a single measurement tool, but an entire measurement system comprised of several measurement tools all matched to each other. Finally, this paper discusses the problem of hydrocarbon contamination in a CD-SEM, which limits the number of times that a certain location on the standard can be used for calibration, and presents a methodology to determine how often the measurement location should be changed.
Keywords: pitch, 100 nm, CD-SEM, metrology, litho metrology, NIST-traceability.

1. INTRODUCTION

The analysis of a measurement system typically includes five components: bias, repeatability, reproducibility, stability, and linearity [1]. The component that we are interested here is the bias, often called accuracy. Bias is defined as the difference between the observed average of measurements and the reference value of a standard. Grating structures are most commonly used as standards for calibration of magnification of a CD-SEM. In this paper we will refer to a 100 nm NIST-traceable pitch standard [2]. Figure 1 shows a SEM image of the sample, from reference 2.

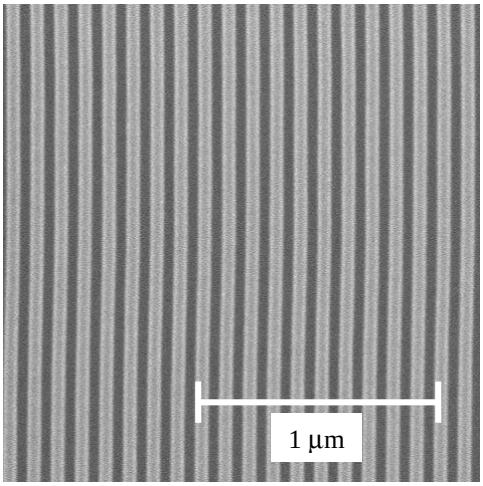


Fig. 1: SEM image of the 100 nm pitch calibration sample

The measurement of the bias of an instrument is a simple task. It requires to measure the pitch a number of times, preferably in a number of different locations, and calculate the average of the measurements. The difference between the average of the measurements and the reference value of the standard is the bias. This procedure is well documented, for example, in the Measurement Systems Analysis Reference Manual of the QS-9000 standard [1]. Measuring the bias is only the first step in the calibration process. The next step is to decide whether the instrument should be recalibrated or

Metrology, Inspection, and Process Control for Microlithography XVII, Daniel J. Herr, Editor,
Proceedings of SPIE Vol. 5038 (2003) © 2003 SPIE · 0277-786X/03/$15.00

not. Because we have not been able to find a reference that teaches when it is appropriate to recalibrate the instrument, we develop in this paper a best practice based on a statistical hypothesis testing on the averages [3].

2. THEORY

The theory presented here is of general applicability and can be found in more detail in reference 3. The only assumption is that the individual measurements be normally distributed. This assumption is generally true.
Suppose that one takes N measurements of pitch (q_1, q_2, …q_N).
The best estimate of the pitch is the average of the measurements:

$$\bar{q} = \frac{1}{N}\sum_{i=1}^{N} q_i \,, \tag{1}$$

and the expanded uncertainty in the measurement is

$$u_{q,\exp} = t_{P,N-1} u_{q,s} \,, \tag{2}$$

where $u_{q,s}$ is the standard uncertainty, or standard deviation of the mean (SDOM)

$$u_{q,s} = \frac{\sigma_{N-1}}{\sqrt{N}}, \tag{3}$$

where σ_{N-1} is the standard deviation of the N measurements

$$\sigma_{N-1} = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}\left(q_i - \bar{q}\right)^2} \,, \tag{4}$$

and $t_{P,N-1}$ is the value of the t-distribution for a level of confidence p, with $N-1$ degrees of freedom. The level of confidence p is expressed in percentage. It is typical, although not a requirement, to use a level of confidence p = 95%. A table, as well as a mathematical formula for calculating the value of the t distribution for any value of degrees of freedom and confidence level can be found in reference 5, Section C.3.8 and Section G.
Equation (2) answers the following question: what is the expanded uncertainty in estimating the value of pitch with the average of N measurement with a specified level of confidence? Notice that the uncertainty decreases with the number of measurements. Therefore, no matter how noisy is the measurement instrument, and how severe are the non-uniformity of the sample and line edge roughness of the sample, the uncertainty in the measurement can be made infinitesimally small if the number of measurements, *N*, is made infinitely large.
Equations (2) and (3) can be rewritten as follows:

$$N = \left(\frac{t_{P,N-1}\sigma_{N-1}}{u_{q,\exp}}\right)^2 . \tag{5}$$

Equation (5) answers the following question: how many individual pitch measurements should one take in order to estimate the pitch of the sample within a specified expanded uncertainty and with a specified confidence level? Note that Equation (5) is not a closed form equation because the standard deviation of the N measurements, σ_{N-1}, is not known until the measurements are taken. However, the standard deviation, σ_{N-1}, is roughly independent of the number of measurements for large *N*, and for normally distributed data, and can therefore be estimated from any prior set of measurements. Likewise, the value of the t distribution, $t_{P,N-1}$, is not known until the actual number of measurements, N, is determined. However, the t distribution is a very slowly varying function of the number of degrees of freedom, for a number of degrees of freedom larger than 8. Therefore it is reasonable to use $t_{P,N-1} = 1.9$ for 90% confidence level, $t_{P,N-1} = 2.3$ for 95% confidence level, and $t_{P,N-1} = 3.4$ for 99% confidence level. $u_{q,\exp}$ is the desired expanded uncertainty, expressed in the same units as the standard deviation σ_{N-1}, for example nm.

3. COST OF OWNERSHIP ELEMENTS OF THE CALIBRATION PROCESS

It can be seen from equation (2) that a smaller measurement uncertainty, $u_{q,\exp}$, can be obtained from the same number of measurements N if the standard deviation, σ_{N-1}, is smaller. Conversely, it can be seen from equation (5) that the smaller the standard deviation, σ_{N-1}, the fewer number of measurement are necessary in order to achieve the same desired expanded uncertainty, $u_{q,\exp}$. Fewer measurements translate directly into a lower Cost of Ownership (CoO) for the calibration process. Fewer measurements also prolong the lifetime of the standard, because the standard gets contaminated with use by electron beam induced deposition of hydrocarbons. A small standard deviation σ_{N-1} is therefore key in obtaining a fast and accurate calibration with minimal sample damage. The standard deviation of a series of measurements depends primarily on the noise in the CD-SEM instrument, on the spatial uniformity of the pitch sample, and on the line edge roughness of the pitch lines[2]. Those are quality attributes of the instrument and of the sample that translate directly into CoO of the calibration process. It is possible to reduce the measurement uncertainty due to line edge roughness and instrument noise by measuring a distance corresponding to several adjacent pitch structures, and dividing the measured distance by the number of pitch structures measured. This is rigorous because what is of interest is the average pitch across the sample. On the other hand, the measurement uncertainty due to pitch non-uniformity across the sample can only be reduced by increasing the number of measurement locations.

4. RECALIBRATION STRATEGY

Another important question is whether or not to change the calibration factor of a CD-SEM following measurement of the standard. There are two practices. The first is to always recalibrate the instrument by applying an appropriate correction factor so that the mean value of the measurements is the same as the certified value of the standard. This method guarantees the best possible measurement accuracy. The drawback of this method is that it complicates the tracking of the performance of the instrument through Statistical Process Control (SPC) charts, as baseline shifts occur every time that the instrument is recalibrated. We advocate a second technique as the engineering best practice. The engineering best practice is to recalibrate the instrument when the measured value of the standard, $\bar{q}$, differs from the certified value, q_{cert}, by more than the combined uncertainties of the following two measurements: 1) the measurement of the standard performed with the instrument to be calibrated and 2) the measurement of the standard performed by an accredited calibration laboratory. The combined expanded uncertainty of the two measurements is given by[1]

$$u_c = t_{p,\nu_{eff}} \sqrt{u_{Master,S}^2 + u_{q,S}^2} \,, \tag{6}$$

where $u_{Master,S}$ is the standard uncertainty reported by the calibration laboratory, $u_{q,s}$ is given in equation (3), and $t_{p,\nu_{eff}}$ is the value of the t-distribution for a desired level of confidence p, with ν_{eff} degrees of freedom, where

$$\nu_{eff} = \frac{\left(u_{Master,S}^2 + u_{q,S}^2\right)^2}{\dfrac{u_{Master,S}^4}{\nu_{Master}} + \dfrac{u_{q,S}^4}{N-1}}, \tag{7}$$

where ν_{Master} is the number of degrees of freedom for the measurement of the standard performed by the calibration laboratory.

According to the definition of NIST-traceability[4], every NIST-traceable measurement must be accompanied by a statement of uncertainty. There are a variety of ways in which a calibration laboratory may report the standard uncertainty and number of degrees of freedom. Some laboratories report uncertainty at the 2-sigma level. In this case the standard uncertainty is obtained by dividing the reported 2-sigma uncertainty by 2. Other laboratories may report the uncertainty at the 95% confidence level. In this case the standard uncertainty is obtained by dividing the 95% confidence level uncertainty by the value of the t-distribution corresponding to the 95% confidence level and to the

number of degrees of freedom for the measurement performed at the calibration laboratory. In any case, the number of degrees of freedom must be reported by the calibration laboratory.
Equation (7) is the Welch-Satterthwaite formula, which is the method prescribed in the ISO Guide to the Expression of Uncertainty in Measurement[5] to calculate the number of effective degrees of freedoms for a measurement that is the combination of several measurements, each with its own associated standard uncertainty and number of degrees of freedom. There are other expressions in the literature for calculating the number of effective degrees of freedom. Reference 3, pg. 103, reports a slightly different expression.
In practice, it is a reasonable approximation that the number of degrees of freedom is high enough, that we can use $t_{p,v_{eff}} \cong 2$ at the 95% confidence level. Therefore we can approximate equation (6) with the following more practical expression:

$$u_c \cong 2\sqrt{u_{Master,S}^2 + u_{q,S}^2} \tag{8}$$

The engineering best practice defined here therefore calls for the following decision:
If

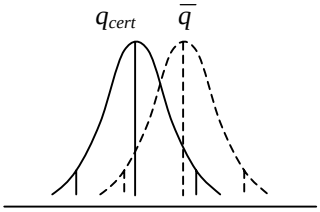
$$\left|\bar{q} - q_{cert}\right| \le u_c \tag{9}$$

then do not alter the calibration of the instrument; otherwise, apply a proportional correction factor k_{corr} to the measurement equal to

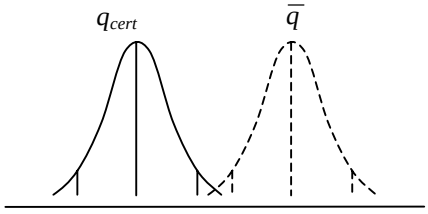
$$k_{corr} = \frac{q_{cert}}{\bar{q}} \tag{10}$$

This best practice is equivalent to what is generally referred to in textbooks as the hypothesis testing on means of normal distributions, with variance unknown[3]. The hypothesis test answers the following question: given two normally distributed populations with unknown mean and variance, are the means of the two populations statistically the same or not, within a specified confidence level? The best practice for instrument recalibration described here requires a recalibration if the answer to that question is no. That is, the instrument needs to be recalibrated if there is a statistically significant difference, within a given confidence level, between the measurement done on the standard at an accredited laboratory and the measurement done with the instrument under investigation.
All cases fall into one of the following three categories:

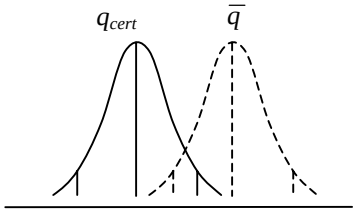
1. The measured value is within the certified confidence interval of the standard. In this case Eqn. (9) is satisfied regardless of the value of $u_{q,s}$. There is no need to recalibrate the instrument.



2. The measured value is outside the certified confidence interval of the standard, and there is no overlap between the certified confidence interval of the standard and the confidence interval of the measurement of the standard with the tool under investigation. In this case Eqn. (9) cannot be satisfied. The instrument needs recalibration.



3. The measured value is outside the certified confidence interval of the standard, and there is overlap between the certified confidence interval of the standard and the confidence interval of the measurement of the standard with the tool under investigation. In this case Eqn (9) must be evaluated to determine whether the instrument needs to be recalibrated.



It is very important to take into account the stability of the instrument before changing the calibration factor on the instrument. Stability is defined as the variation in measurement results over an extended period of time. To take into account the stability of the instrument one needs to take measurements over an extended period of time, or account for tool stability with a separate uncertainty component that needs to be added to the standard deviation of the measurements. The analysis carried out here indicates that the less repeatable, or less stable is the instrument, the less stringent is the requirement to recalibrate the instrument, which is consistent with the fact that a tool with better measurement capability has more stringent calibration requirements.

5. TOOL MATCHING

Pitch samples are an excellent choice for matching the magnification of CD-SEMs. For tool matching it is not necessary to use a NIST-traceable standard. Any sample can be used, provided that the same location on the sample can be measured by both tools. In the case of a CD-SEM, because of hydrocarbon contamination, which builds up on the sample as a result of imaging, it may not be possible to use the same location multiple times or in multiple instruments. Therefore it may be best to use a highly uniform sample, with very low line edge roughness, so that any area on the sample can be measured. This will also save navigation time to find the same location repeatedly. The sample uniformity and line edge roughness will be a limiting factor in the ability to match different tools or, in other words, the tool matching ability cannot be better than the uncertainty in the measurement due to line edge roughness and uniformity. The effect of line edge roughness, however, can be minimized by increasing the number of measurements, or increasing the number of adjacent pitches that are measured.

The analysis and procedure discussed above can be used to match two tools. In a tool matching study the same hypothesis test used above to determine whether a tool needs to be calibrated can be used. If the mean of N_1 measurements from tool 1 is $\bar{q}_1$, with a standard uncertainty $u_{q1,s}$, and the mean of N_2 measurements from tool 2 is $\bar{q}_2$, with a standard uncertainty $u_{q2,s}$, then the two tools match each other if

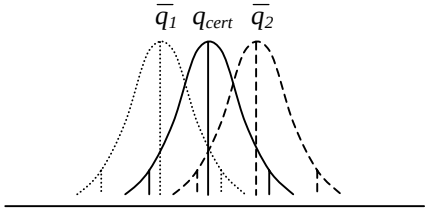
$$\left|\bar{q}_1 - \bar{q}_2\right| \le t_{p,\nu_{eff}} \sqrt{u_{q1,S}^2 + u_{q2,S}^2} \tag{11}$$

where $t_{p,\nu_{eff}}$ is the value of the t distribution for the level of confidence p, and for the number of degrees of freedom of the combined measurement, calculated as in (7), with the obvious substitutions:

$$\nu_{eff} = \frac{\left(u_{q1,S}^2 + u_{q2,S}^2\right)^2}{\frac{u_{q1,S}^4}{N_1 - 1} + \frac{u_{q2,S}^4}{N_2 - 1}} \tag{12}$$

As in the case of calibration, instrument stability must be taken into account by allowing the measurements to span a period of time that encompasses the variability due to instrument instability. If this is impractical, then a term must be added to the standard uncertainty to account for instrument stability.

It is worth noting that two tools that are both calibrated against a NIST-traceable standard, may not match, as illustrated below:



Therefore matching and traceability need to be carried out simultaneously. A possible strategy is the following:

1. Carry out the NIST-traceable calibration of each tool independently as discussed in section 4.
2. Use Equation (11) on each pair of tools to see if they all match each other. If there are n tools, then Equation (11) must be evaluated on $\frac{n(n-1)}{2}$ pairs. The results can be organized in a two-dimensional matrix.
3. If all the pairs match, then no further calibration is required. If any of the pairs do not match, the tool with the highest number of mismatches needs to be calibrated using Equation (10). Then the matrix can be re-evaluated, and the process iterated until there are no more mismatches.

Alternatively, for best matching, all tools can be calibrated using Equation (10). This method assures matching, but requires the largest number of recalibrations and baseline shifts.

A further level of complication arises when matching different Fabs. This can be done using the methodology explained above if using the same standard, or standards that exactly match each other. If different standards are used in different Fabs, then the matching ability is decreased by the uncertainty with which the standards are certified.

NIST-traceable standards, being all matched to the International System of Units (SI) facilitate the matching task when traceability to the SI is also required.

6. ELECTRON-BEAM-INDUCED CONTAMINATION

A practical guide on CD-SEM calibration would not be complete without a discussion of electron beam contamination. Of particular interest is the question of how many times a certain location on the sample can be used for calibration before it is contaminated and a new area needs to be chosen.

We have measured the effect of electron-beam-induced sample degradation quantitatively by measuring 10 consecutive pitches (1 µm nominal distance) 4500 times in the same location using a KLA-Tencor 8100 CD-SEM, a field of view of 1.5 µm, landing energy of 600 eV, and e-beam current of 30 pA. The accumulated e-beam exposure time after 4500 measurements was 27 min. Figures 2a-d show the progression of sample contamination.

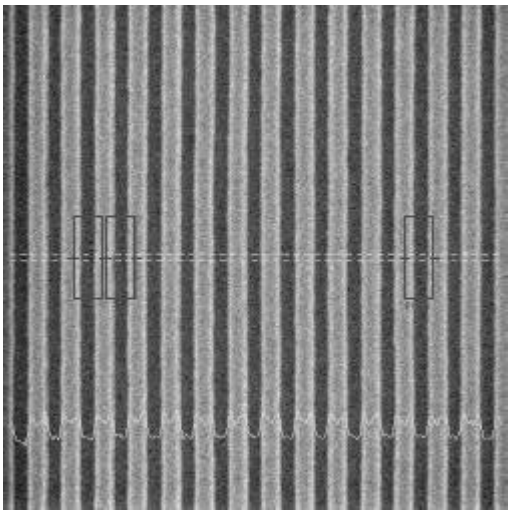


Fig. 2a: Image of the standard after the first measurement.

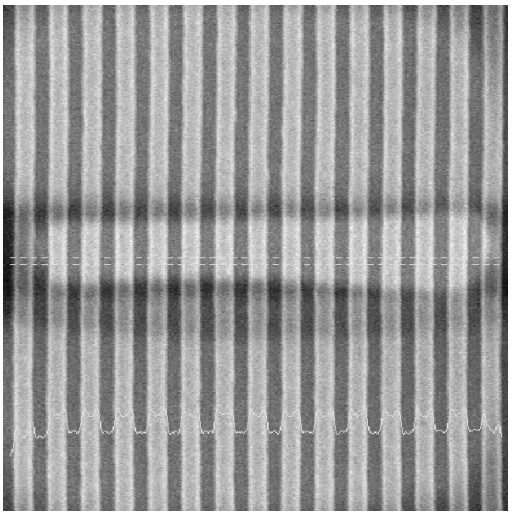


Fig. 2b: Image of the standard after 90 seconds of continuous imaging, or 250 measurements.

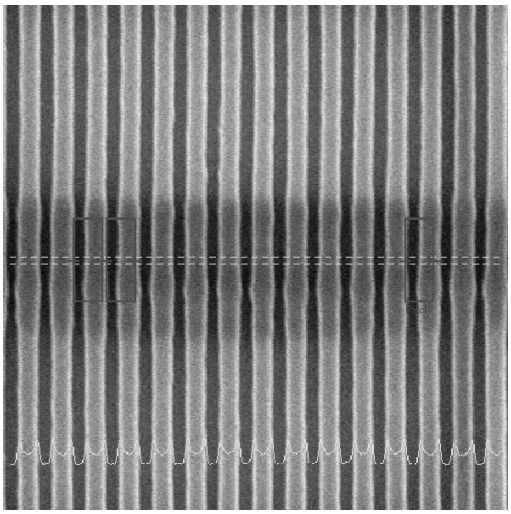


Fig. 2c: Image of the standard after 6 minutes of continuous imaging, or 1000 measurements.

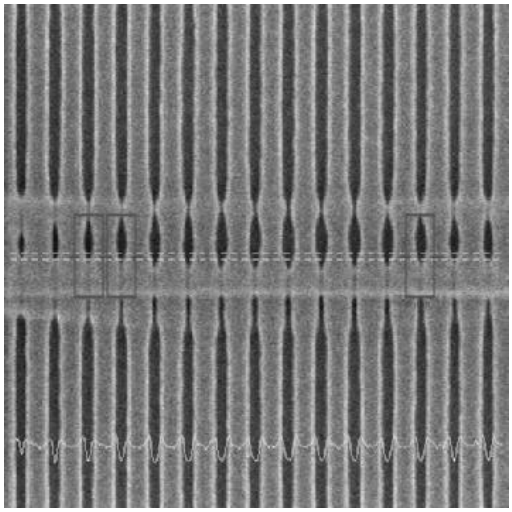


Fig. 2d: Image of the standard after 24 minutes of continuous imaging, or 4000 measurements.

All the measurements took place between the white dotted lines in the middle of the images. The brightness profile between the dotted lines is displayed in the lower portion of the image. The outer two of the three rectangular boxes on the dotted line are used as gates to determine the pitch using an edge detection algorithm with 50% threshold.

As shown in Figure 2, prolonged measuring causes a widening of the lines that constitute the grating. The effect is already noticeable after 250 measurements. After 1000 measurements changes in the critical dimension (CD) and line edge roughness (LER) have become obvious. After 4000 measurements, the CD and LER of the grooves have degraded severely (Fig. 2d). Even at this point, the pitch value determined by the edge detection algorithm described above remains about the same. The reason is that the lines all widen at approximately the same rate, so that CD variations cancel out when measuring pitch. The edge detection algorithm eventually fails when the lines start to merge with each other, and at that point the measured pitch value shows a sudden and dramatic departure from its normal value. In our case, this happened after approximately 4500 measurements, or 27 minutes of continuous measuring.

The pitch variation over time is illustrated more quantitatively in Fig. 3. To obtain the plot in Fig. 3, the 4500 pitch measurements were divided up into 18 blocks of 250 consecutive measurements each. The average of the 250 measurements in each block is plotted in chronological order. Note that, while the pitch measurement in this study was done over a distance of 1 μm, or 10 pitches, we have plotted in Fig. 3 the fundamental pitch value, which is obtained by dividing the measured value by 10. This plots shows quantitatively that the measured pitch value is not affected by the electron-beam-induced sample degradation up to the point where a sudden departure from the expected value occurs. In this case this occurred after approximately 4500 measurements, or 27 min of continuous imaging on the same location.

Based on these results, we reach the following conclusions:

1. Unlike CD, the pitch value is not noticeably affected by the electron-beam contamination.
2. With the CD-SEM system that we tested, the pitch value of the standard remains well within its uncertainty specifications for at least 4000 measurements.
3. Some sample degradation in the measured area can be observed after about 250 measurements. Such degradation does not affect pitch measurement results as long as some contrast is maintained. On the other hand, the electron beam should not be left on at the same location for prolonged time because electron-beam-contamination still takes place even though no measurements are taken.
4. When using automated recipes, the ability of a CD-SEM to reposition the beam to the same location is limited by stage precision. Slight changes in positioning could affect the pitch measurement, if the pitch measurement overlaps areas that have been subjected to different electron beam exposures. Therefore we recommend that the number of times that the same area is measured be conservatively set at no more than 250. We also recommend obtaining measurements as the average of 10 measurements. Therefore, any individual area could be used for 25 calibrations.

5. The above results have only been validated with the particular CD-SEM and beam configuration described here. This study should be used as a methodology to investigate the effect of electron-beam contamination on the measurement of pitch. Similar studies need to be carried out on the specific CD-SEM being used on a case-by-case basis.

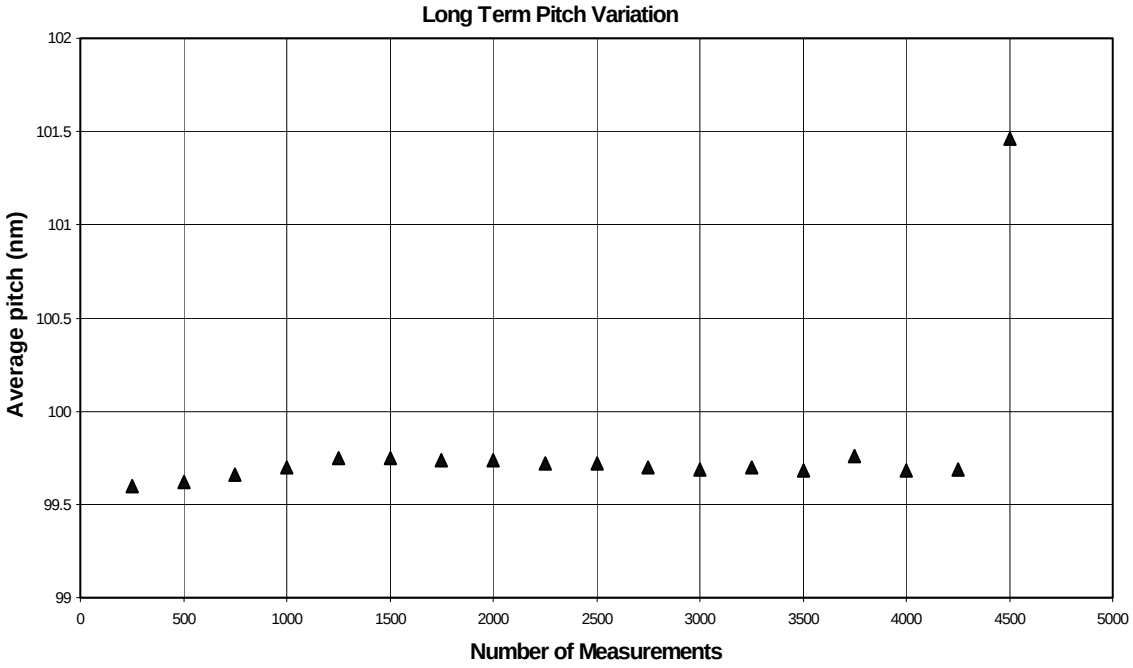


Figure 3: Measured pitch as a function of the number of measurements. 4500 pitch measurements were taken, and divided up into 18 blocks of 250 consecutive measurements each. The average of the 250 measurements in each block is plotted in chronological order

6. CONCLUSIONS

Through a rigorous statistical analysis based on statistical hypothesis testing we have defined a best practice for calibration of a measurement tool. The analysis takes into consideration both the uncertainty in the measurement of the standard by an accredited calibration laboratory, and the uncertainty in the measurement of the standard with the tool to be calibrated. The fundamental result can be summarized saying that a tool needs to be recalibrated if the mean of the two measurements differ by more than the combined expanded uncertainty of the two measurements, calculated at the desired confidence level. The same theory is applied to the case of tool matching. The paper identifies line edge roughness and sample uniformity as Cost of Ownership elements of the calibration process because repeated measurements are required to reduce the effects of both sample uniformity and line edge roughness. Finally, the paper presents a methodology to determine how often the measurement location should be changed due to hydrocarbon contamination in a CD-SEM. On the system that we used we set the limit to 25 measurements, each of them comprised of 10 consecutive measurement for a total imaging time of 90 seconds.

REFERENCES

1. QS-9000 *Measurement Systems Analysis Reference Manual, 2nd edition,* 1995.
2. M. Tortonese, J. Prochazka, P. Konicek, J. Schneir, and I. R. Smith, "100 nm pitch standard characterization for metrology applications", *Proc. SPIE*, **4689**, pp. 558-564, 2002.
3. Montgomery, D. C. *Introduction to Statistical Quality control,* 3rd ed., John Wiley, New York, 1997.
4. http://www.nist.gov/traceability/
5. *Guide to the expression of uncertainty in measurement,* ISO, Geneva, 1995.