yms Magazine

Process Watch: Having Confidence in Your Confidence Level

By Douglas G. Sutherland and David W. Price

Author's Note: The Process Watch series explores key concepts about process control—defect inspection and metrology—for the semiconductor industry. Following the previous installments, which examined the 10 fundamental truths of process control, this new series of articles highlights additional trends in process control, including successful implementation strategies and the benefits for IC manufacturing.

While working at the Guinness® brewing company in Dublin, Ireland in the early-1900s, William Sealy Gosset developed a statistical algorithm called the T-test.[1] Gosset used this algorithm to determine the best-yielding varieties of barley to minimize costs for his employer, but to help protect Guinness' intellectual property he published his work under the pen name "Student." The version of the T-test that we use today is a refinement made by Sir Ronald Fisher, a colleague of Gosset's at Oxford University, but it is still commonly referred to as Student's T-test. This paper does not address the mathematical nature of the T-test itself but rather looks at the amount of data required to *consistently achieve* the ninety-five percent confidence level in the T-test result.

A T-test is a statistical algorithm used to determine if two samples are part of the same parent population. It does not resolve the question unequivocally but rather calculates the probability that the two samples are part of the same parent population. As an example, if we developed a new methodology for cleaning an etch chamber, we would want to show that it resulted in fewer fall-on particles. Using a wafer inspection system, we could measure the particle count on wafers in the chamber following the old cleaning process and then measure the particle count again following the new cleaning process. We could then use a T-test to tell if the difference was statistically significant or just the result of random fluctuations. The T-test answers the question: what is the probability that two samples are part of the same population?

However, as shown in Figure 1, there are two ways that a T-Test can give a false result: a false positive or a false negative. To confirm that the experimental data is actually different from the baseline, the T-test usually has to score less than 5% (i.e. less than 5% probability of a false positive). However, if the T-test scores greater than 5% (a negative result), it doesn't tell you anything about the probability of that result being false. The probability of false negatives is governed by the number of measurements. So there are always two criteria: (1) Did my experiment pass or fail the T-test? (2) Did I take enough measurements to be confident in the result? It is that last question that we try to address in this paper.

T-Test Conclusion	Actual: No Change	Actual: Change
Change	False Positive	
No Change		False Negative

Figure 1. A “Truth Table” highlights the two ways that a T-Test can give the wrong result.

Changes to the semiconductor manufacturing process are expensive propositions. Implementing a change that doesn’t do anything (false positive) is not only a waste of time but potentially harmful. Not implementing a change that could have been beneficial (false negative) could cost tens of millions of dollars in lost opportunity. It is important to have the appropriate degree of confidence in your results and to do so requires that you use a sample size that is appropriate for the size of the change you are trying to affect. In the example of the etch cleaning procedure, this means that inspection data from a sufficient number of wafers needs to be collected in order to determine whether or not the new clean procedure truly reduces particle count.

In general, the bigger the difference between two things, the easier it is to tell them apart. It is easier to tell red from blue than it is to distinguish between two different shades of red or between two different shades of blue. Similarly, the less variability there is in a sample, the easier it is to see a change.[2] In statistics the variability (sometimes referred to as noise) is usually measured in units of standard deviation (σ). It is often convenient to also express the difference in the means of two samples in units of σ (e.g., the mean of the experimental results was 1σ below the mean of the baseline). The advantage of this is that it normalizes the results to a common unit of measure (σ). Simply stating that two means are separated by some absolute value is not very informative (e.g., the average of A is greater than the average of B by 42). However, if we can express that absolute number in units of standard deviations, then it immediately puts the problem in context and instantly provides an understanding of how far apart these two values are in relative terms (e.g., the average of A is greater than the average of B by one standard deviation).

Figure 2 shows two examples of data sets, before and after a change. These can be thought of in terms of the etch chamber cleaning experiment we discussed earlier. The baseline data is the particle count

per wafer before the new clean process and the results data is the particle count per wafer after the new clean procedure. Figure 2A shows the results of a small change in the mean of a data set with high standard deviation and figure 2B shows the results of the same sized change in the mean but with less noisy data (lower standard deviation). You will require more data (e.g., more wafers inspected) to confirm the change in figure 2A than in figure 2B simply because the signal-to-noise ratio is lower in 2A even though the *absolute change* is the same in both cases.

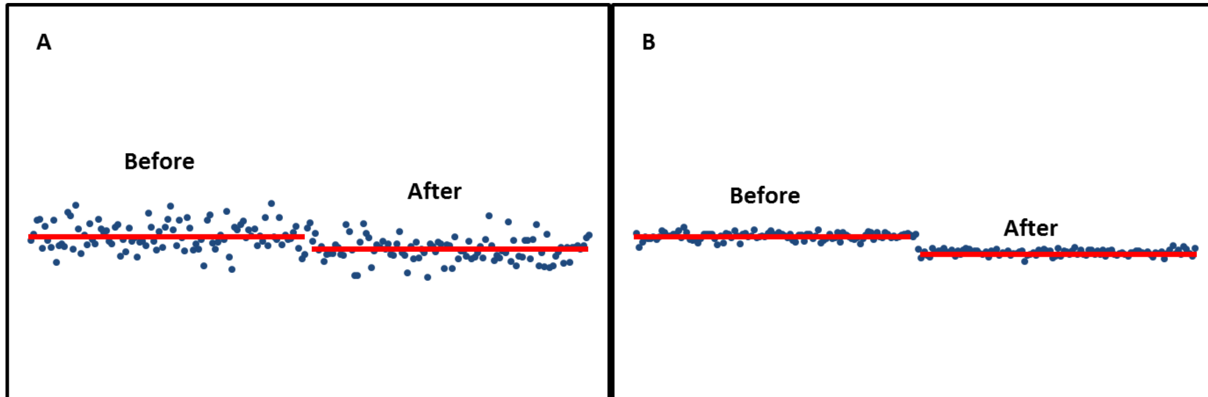

Figure 2. Both charts show the same absolute change, before and after, but 2B (right) has much lower standard deviation. When the change is small relative to the standard deviation as in 2A (left) it will require more data to confirm it.

The question is: how much data do we need to confidently tell the difference? Visually, we can see this when we plot the data in terms of the Standard Error (SE). The SE can be thought of as the error in calculating the average (e.g., the average was X +/- SE). The SE is proportional to $\sigma/\sqrt{n}$ where n is the sample size. Figure 3 shows the SE for two different samples as a function of the number of measurements, n.

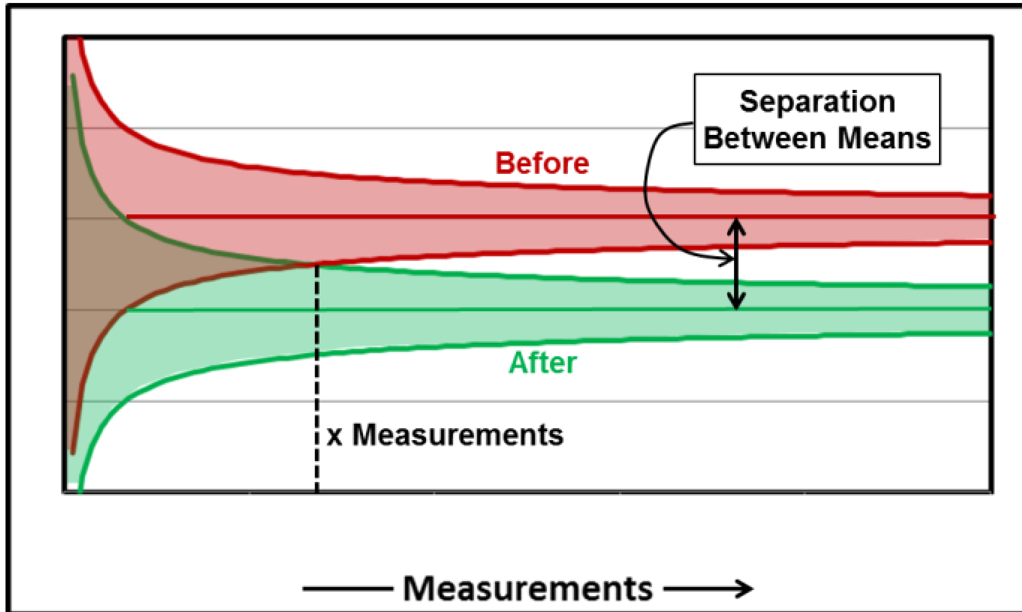

Figure 3. The Standard Error (SE) in the average of two samples with different means. In this case the standard deviation is the same in both data sets but that need not be the case. With greater than x measurements the error bars no longer overlap and one can state with 95% confidence that the two populations are distinct.

For a given difference in the means and a given standard deviation we can calculate the number of measurements, x, required to eliminate the overlap in the Standard Errors of these two measurements (at a given confidence level).

The actual equation to determine the correct sample size in the T-test is given by,

$$n = \frac{2(Z_{\beta} + Z_{1-\alpha/2})^2}{\text{Delta}^2} \quad \text{Eq 1}$$

where n is the required sample size, "Delta" is the difference between the two means measured in units of standard deviation (σ) and Z_x is the area under the T distribution at probability x. For α=0.05 (5% chance of a false positive) and β=0.95 (5% chance of a false negative), $Z_{1-\alpha/2}$ and Z_{β} are equal to 1.960 and 1.645 respectively (Z values for other values of α and β are available in most statistics textbooks, Microsoft® Excel® or on the web). As seen in Figure 3 and shown mathematically in Eq 1, as the difference between the two populations (Delta) becomes smaller, the number of measurements required to tell them apart will become exponentially larger. Figure 4 shows the required sample size as a function of the Delta between the means expressed in units of σ. As expected, for large changes, greater than 3 σ, one can confirm the T-test 95% of the time with very little data. As Delta gets smaller, more measurements are required to consistently confirm the change. A change of only one standard deviation requires 26 measurements before and after, but a change of 0.5 σ requires over 100 measurements.

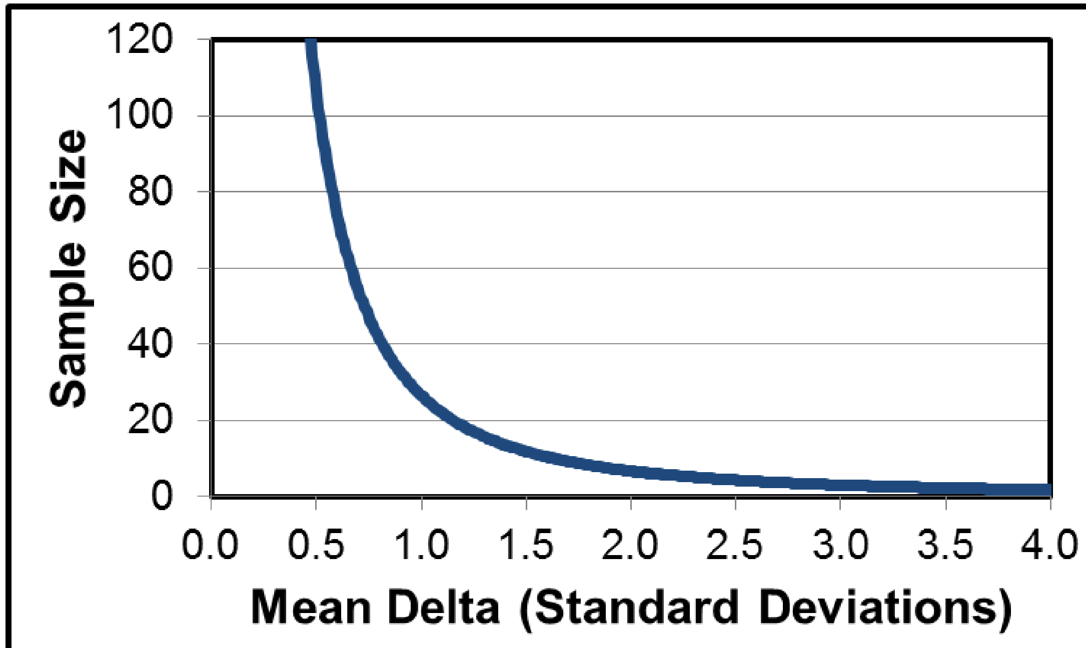

Figure 4. Sample size required to confirm a given change in the mean of two populations with 5% false positives and 5% false negatives

The relationship between the size of the change and the minimum number of measurements required to detect it has ramifications for the type of metrology or inspection tool that can be employed to confirm a given change. Figure 5 uses the results from figure 4 to show the time it would take to confirm a given change with different tool types. In this example the sample size is measured in number of wafers. For fast tools (high throughput, such as laser scanning wafer inspection systems) it is feasible to confirm relatively small improvements (<0.5σ) in the process because they can make the 200 required measurements (100 before and 100 after) in a relatively short time. Slower tools such as e-beam inspection systems are limited to detecting only gross changes in the process, where the improvement is greater than 2σ. Even here the measurement time alone means that it can be weeks before one can confirm a positive result. For the etch chamber cleaning example, it would be necessary to quickly determine the results of the change in clean procedure so that the etch tool could be put back into production. Thus, the best inspection system to determine the change in particle counts would be a high throughput system that can detect the particles of interest with low wafer-to-wafer variability.

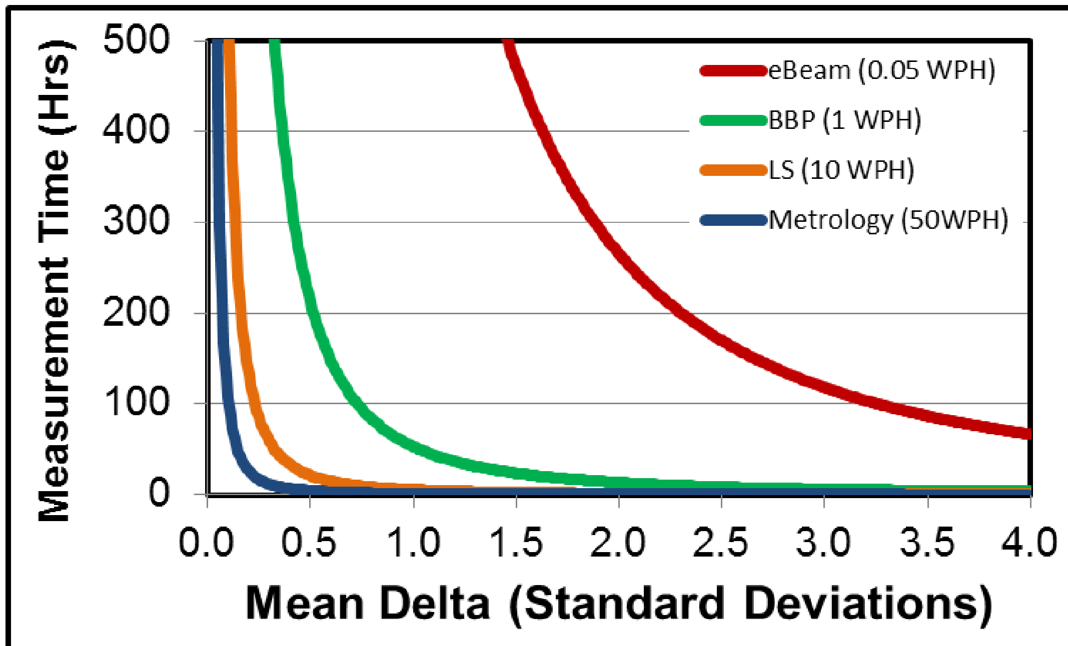

Figure 5. The measurement time required to determine a given change for process control tools with four different throughputs (e-Beam, Broadband Plasma, Laser Scattering and Metrology)

Experiments are expensive to run. They can be a waste of time and resources if they result in a false positive and can result in millions of dollars of unrealized opportunity if they result in a false negative. To have the appropriate degree of confidence in your results you must use the correct sample size (and thus the appropriate tools) that correspond to the size of the change you are trying to affect.

References:

1) https://en.wikipedia.org/wiki/William_Sealy_Gosset

2) Process Watch: Know Your Enemy, *Solid State Technology*, March 2015

***About the Authors*:**

Dr. David W. Price is a Senior Director at KLA-Tencor Corp. Dr. Douglas Sutherland is a Principal Scientist at KLA-Tencor Corp. Over the last 10 years, Dr. Price and Dr. Sutherland have worked directly with more than 50 semiconductor IC manufacturers to help them optimize their overall inspection strategy to achieve the lowest total cost. This series of articles attempts to summarize some of the universal lessons they have observed through these engagements.