SEMI® Technical Programs Present

Advanced SPC Techniques for Process Yield Improvement

July 14, 1998
SEMICON® West 98
San Francisco Marriott Hotel
San Francisco, California

DISCLAIMER

SEMI makes no warranties or representations as to the viewpoints expressed in the materials set forth herein, either as to general scientific validity or for any particular application by a potential user. The viewpoints expressed are those of the lecturer(s) at *Advanced SPC Techniques for Process Yield Improvement* as delivered on July 14, 1998 and as reprinted herein.

SEMI takes no responsibility for any use or misuse of any information contained in any of the lectures reprinted herein and cautions readers that an independent evaluation of the viewpoints expressed in each lecture is entirely the responsibility of the reader.

ADVANCED SPC TECHNIQUES FOR PROCESS YIELD IMPROVEMENT

Zigmund Bluvband, Ph.D.
President

Pavel Grabov, Ph.D.
Vice-President

A. L. D. Ltd.,
P.O. Box 679, Rishon Lezion, 75106, Israel

Biography

Zigmund Bluvband is the founder and the President of A.L.D. Ltd. (since 1984) Z. Bluvband holds an M.A. degree in Electronics , M.Sc. in Mathematics and a Ph.D. in Reliability, Metrology and Control. Z. Bluvband was the Manager of Reliability Tools Group, Israel Aircraft Industries until 1984. Later he was the Quality Director at Tadiran, Electronic Systems Division. Z. Bluvband has accrued 29 years of industrial and academic experience, consulting and teaching in Quality Assurance (QA), Reliability, Availability, and Maintainability (RAM) working for both. As a QA professional and an open minded researcher, Z. Bluvband developed advanced methods and designed A.L.D. software tools which are now widely accepted among the best professional RAM analysis tools in the world. Z. Bluvband's significant contribution to RAMS and QA areas is reflected in 3 patents more than 60 papers, books, and tutorials. In 1984 Z. Bluvband was the first to suggest using the Back Order Probability (BOP) as a criteria for spare parts planning (1984, San Francisco) and the new "Z-Charting SPC Technique" together with Dr. P. Grabov (Tokyo, 1994). Z. Bluvband has published a book entitled Software Quality Assurance (Opus, Tel-Aviv, 1996). Z. Bluvband has been successfully conducting CQE, CQA, CRE and Lead Assessors courses for hundreds of Quality Professionals. Z. Bluvband was the President of the Israel Society for Quality from 1989 to 1994.

Pavel Grabov is a Vice-President of A.L.D. Ltd. He holds M.Sc. degree in Physics and Ph.D. degree in Industrial Engineering with an emphasis on quality control using low-energy gamma- radiation. He has over 20 years of academic and industrial experience in quality and process improvement. His current interests are statistical process control, applied statistics and design of experiments. He is the author of over 20 patents and 50 articles on manufacturing engineering, statistical process control and quality improvement, gauge studies, process measurements and experimental design. Prior to joining A.L.D., P. Grabov was a Senior Lecturer at the Quality Assurance & Reliability Department of the Technion (Israel Institute of Technology). Since joining to A.L.D. in 1992, P. Grabov has been responsible for consulting and training services in the field of Industrial & Manufacturing Engineering. He serves as a consultant for the Semiconductor & Hi-Tech Industries providing for quality improvement & yield enhancement techniques. He taught over 30 courses on Practical Statistics for Engineers, Statistical Process Control and Design of Experiments. P. Grabov is a member of American Society for Quality, Israel Society for Quality and Israeli Statistical Association. He is a Certified Quality Engineer (CQE).

Abstract

Procedures and tools of Statistical Process Control (SPC) can increase product yield in the semiconductor industry. However, conventional SPC techniques do not fit the unique processes of the semiconductor industry. A.L.D. has developed a customized, advanced SPC methodology for wafer fabrication process control, and has successfully implemented it in several fab areas. The scope of the methodology includes data collection and treatment as well as analysis interpretation, and decision making. The concept focuses on data variability as a basis for analysis of a process's current performance. Furthermore, the total process variation is

decomposed into its underlying components. The methodology's strongest outcomes are: 1) reduction of the number of wafers for data collection during the process/product design stages; 2) evaluation of relationship between yield and ETs (Electrical Testing results); 3) corrected procedures of GR&R study; 4) 3-level (Sites-Wafers-Lots) 'semicharting' technique; 5) low-*p* (proportion of defectives) charts; 6) 'pure' capability study and computation of relevant capability/performance indices; 7) the knowledge-based process improvement strategy of reactive and proactive activities.

Introduction

Modern concepts of effective process management necessitate advanced Quality Management and Statistical Quality Control (SQC). This paper is primarily focused on one subfamily of SQC—SPC, which pertains to in-process activities. Other SQC methods representing the pre- and post-process activities are beyond the scope of this paper. Nevertheless, those aspects of DOE, Sampling and Final Inspection which are closely connected to SPC will be tangentially mentioned.

Application of "canned" SPC schemes and methodologies has demonstrated that some SPC techniques are not appropriate to the semiconductor manufacturing uniquely complex processes. The reasons for this are as follows:

1. Numerous two-type design rules (geometric and electric) require application of multivariate analysis.
2. Highly restricted amount of data at development stages, and many similar data from production, result in high uncertainty of any statistical inferences based on these data.
3. Insufficient resolution and relatively low precision of measurement/test systems, as well as their periodic faults, result in noisy databases and many outliers.
4. High fab flexibility and frequent process changeovers due to short product lifetime require universal solutions at the technology level.

This paper presents a proven effective SPC methodology customized to the wafer fabrication process. Development of an appropriate SPC technique for the pursuit of an 'increasing yield' policy in the semiconductor industry was our main goal. The methodology includes advanced methods for data processing, analysis, interpretation, evaluation and decision making. Some techniques presented here are described in detail; others are just mentioned due to the limited space.

This paper is primarily intended for practitioners who are responsible for improving quality and increasing productivity. Therefore, we have avoided overusing dull statistics and formulas. The interested reader can refer to the formulas and illustrations in the visuals attached to this tutorial. Due to the proprietary nature of the data, neither names nor values of some analyzed electrical parameters and critical dimensions can be disclosed.

Data Collection and Processing

Data Collection—Statistical methods can be very effective for process improvement, provided that the information underlying all statistical inferences is valid, reliable and timely. It is rarely possible to test an entire population due to high costs or destructive testing methods. Therefore, a sample is used to make an inference about the population. The results of this inductive analysis depend on the degree to which the sample represents the entire population. A fundamental approach to correct sampling for SPC was developed by Shewhart (Shewhart, 1931) known as the rational subgroup concept. This concept is based on two main principles:

Principle 1. Sample units (subgroups) should be selected so as to yield minimum within-sample variation in order to evaluate inherent process variation due to common causes. This means that the time required to select subgroups should be short enough to maintain homogeneous production conditions (no changing tools, no machine adjustments, no operators changes, no material lot changes, etc.).

Principle 2. Sampling frequency should be high enough to provide a maximum chance of capturing between-sample variation due to assignable causes. However, as the frequency of sampling rises, so do the associated costs - often making frequent sampling unfeasible. An optimal sampling frequency should be related to the production rate and rate of occurrence of various types of process shifts. OPTIMAL FREQUENCY SHOULD ALSO ACCOUNT FOR COSTS!

A good empirical 'rule of thumb' for determining sampling frequency is dividing the average time period between two subsequent process adjustments by 5–6 for rather well-known processes, and by 10–12 for almost unknown processes. If much greater protection against continuing an out-of-control process is required, the denominator could be increased to 20–24. This is important when there are significant losses associated with a Type II error, and a process may continue out of control until the next sampling.

We propose supplementing Shewhart's principles with one of our own:

Principle 3. Time is not always an optimal basis for determining sampling frequency. If prior processing has a significant influence on present production results, the frequency should be measured not in units of time but in the number of items produced between consecutive samples.

The choice of the appropriate units of sampling frequency is especially important for fabs wherein a pollution depends on the number of produced wafers. The resulting optimal sampling frequency measured in wafers can help calculate the minimal number of wafers necessary to be produced in order to keep the 'sleeping' technology in the standby state.

Process Distribution Analysis—The next step in the analysis is processing and analyzing the data. Some descriptive statistics are derived from the sample data. Since there are three basic properties characterizing any distribution—location, spread and shape—analysis usually focuses on measures associated with these properties.

One of the most serious problems associated with real data is that they usually contain outliers. This is especially true for transistor leakage, drive currents, threshold voltages, and other phenomena measured on a test structure. Due to the high sensitivity of parametric statistics to extreme values, our inferences could be erroneous without preliminary 'purification' of raw data. Purification could be performed, for example, using *a priori* knowledge of the involved processes and their parameter ranges. Measurement is classified as acceptable if it is within known range limits. Unfortunately, due to the absence of 'hard' *a priori* knowledge, this approach is rarely applicable in real-world situations.

Usually, the practitioner must use intuition to perform some rough purification before data analysis. This could involve a trimmed statistics calculation, wherein some of the ordered observations in a sample are considered outliers and therefore are trimmed from each end (why?). Another technique is the Q-Q plot in which the empirical quantiles are plotted *versus* the corresponding normal quantiles, implicitly supposing the underlying distribution is normal (why?); points deviating from a linear pattern on a plot are outliers. There are many other intuitive techniques. As a rule, these simple methods are characterized by significant errors of both Type I and II, especially for large sets of data.

There are other more sophisticated and much more effective statistical procedures for outlier detection, but they are not used to any extent because they are not particularly well known to practitioners. We recommend Grubbs' test for the univariate case, and Jackknife distance and cross-validation for the multivariate case. In the Grubbs' test, the distances of largest and smallest values from the sample mean measured in sample standard deviation units are compared with some critical values. The Jackknife test calculates the distance of each point from the multivariate mean (centroid), with subsequent comparison of the largest values with some upper limit. Cross-validation employs *a priori* knowledge of strong cross-correlations between some measurable characteristics—if an extreme rise or fall in the value of one characteristic is associated with a corresponding rise of fall of the others.

Suppose the data have been checked for outliers, refined, and some statistics of interest have been calculated. The next step is to draw conclusions about the population's parameters. Validity of the transition from statistics to parameter estimates could be significantly improved using the bootstrap method (Efron, 1981). Bootstrap is an extremely powerful simulation technique whereby raw data are treated as if they constitute the population under study. By replicating those data an infinite number of times, a large number of samples (each the same size as the original one) could be drawn at random from that population. For every bootstrap sample, a statistical estimator of interest is computed. This repeated resampling eventually generates an empirical sampling distribution of sample mean or variance or other estimator of location, spread and shape. The main advantage of bootstrapping is that the distribution could be empirically reconstructed based on the original data characteristics without any *a priori* assumptions about the distribution.

We have successfully applied bootstrapping to both the early stages of new technology development and process improvement effort verification when due to cost/time constraints small sets of data are used for predicting expected results. The procedure's convergence has been evaluated using 532 wafers each tested at 5 sites (some electrical parameter). Subsequent random sampling with increasing size from sample to sample has shown that this population's central tendency can be estimated using 2–3 wafers, and the spread can be estimated using 7–8 wafers, both at the 95% confidence level.

The final stage of any distribution analysis is a comparison of parameter estimate with its desired value. The rules can be defined as follows: 'the Nominal - the Best' for central tendency; 'the Smaller - the Better' for spread; 'In accordance with Specification' for shape (peakedness, symmetry, etc.).

Correlation & Regression Analysis of 'Yield-ETs' Relationship—Until now the discussion has focused on distribution analysis, and has presented various enumerative studies the purpose of which is to determine a process's *current state*. However, the primary objective of any statistical analysis is usually process *improvement*. This can be achieved, for example, by using correlation and regression analysis, which is an analytic method focused on determining cause-effect relationships related to process performance. We'll present our approach to this analysis considering one of most serious problems of wafer fabrication process: yield prediction using Electrical Testing (ET) results.

Since the exact equation relating a chip's electrical parameters to the output of the process (God's Equation) is not known, both the yield prediction and determination of optimal parameter values can be done only by studying past performance, i.e., analyzing the Yield-ET relationship. Experience has shown that any direct attack or straightforward analysis of the Yield-ET's database is doomed to failure due to considerable obstacles.

A.L.D. has developed a proven strategy of root cause search (Fig.1). First, it is necessary to eliminate the yield failure data due to different yield killers. This is because such data do not relate to degradation, and only characterize 'catastrophic' process failures. Second, we contend with errors arising in all measurements using the above methods of data 'purification.' Next, engineering expertise is required for reasonable database reduction, because hundreds of tested parameters are involved. The database should include families of critical parameters that are potentially responsible for process yield (VIA, contact chains, transistors, capacitors, etc.). At least one key parameter of each ET structure should be kept for further analysis. The correct choice of possible yield-related predictors requires extensive engineering knowledge, and could be supported by cross-correlation analysis of ET data.

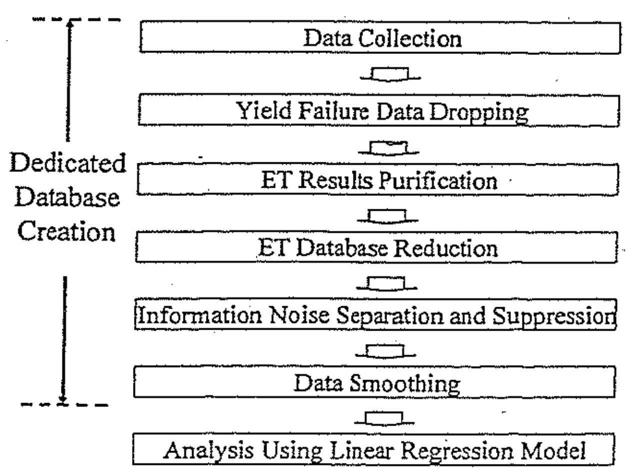

Figure 1. Yield Prediction Using ET's Results

Further database 'purification' requires the separation and suppression of 'information noise.' Any process with normal behavior generates random yield variation representing information noise not related to yield degradation, and should be separated using control chart techniques. The data characterizing yield variations within control limits (inherent process variability) on a *p*-chart are dropped from the database. The remaining points above the upper control limit on the chart constitute 'pure' data for analysis performance. The last stage of data processing involves grouping poor wafers coming from the same lot and ignoring individual poor wafers having no counterparts in the 'mother' lot. This procedure results in data 'smoothing,' and eliminates the necessity of describing every salient point by constructing a yield model.

Experience has shown that the resulting database could be readily analyzed using common techniques of multiple linear regression. More complicated models with second-order terms and two-way interactions between parameters have not resulted in any significant advantages.

This procedure's efficiency has been demonstrated over and over again for different technologies, and has isolated the critical parameters responsible for process yield. As a rule, the coefficient of determination for constructed two-term or three-term models was rather high (0.5–0.8).

Measurement System Analysis

Since measurement systems are used in SPC for making decisions about processes, a conclusion about these systems themselves is necessary. The Gage Repeatability and Reproducibility (GR&R) study is used to estimate the ability of a system to produce precise results. The study is performed by making repeated measurements on the same measurand. The result is evaluated for both repeatability, characterizing variation under identical conditions of measurement (variation due to the gage itself), and reproducibility, characterizing variation under changed conditions of measurement (variation due to operator, time, reloading, reset, environment etc.).

We'll present our approach to GR&R as applied to the precision of Scanning Electron Microscopes (SEM) used for measuring Critical Dimensions (CD). The main CD-SEM metrology problems stem from the fact that repeated imaging cannot produce the same result because the repeated irradiation by the electron beam leads to charge accumulation on photoresist structures surrounding the measured features. The irradiation may also result in contamination, when molecules and atoms in the vacuum and wafer are activated by the electron beam and subsequently deposited on the wafer. This process forms a carbonaceous material decreasing the electron emission from the surface, and results in the features widening. Although contamination tends to build up slower than charging, both act together to distort the repeated imaging results.

Proposed solutions for these problems usually take the form of recommendations concerning operating conditions: reduce beam current, decrease irradiation time, operate at lower magnification, etc. These solutions are rather weak palliatives: they inevitably decrease the signal-to-noise ratio, and the charging and contamination effects will always remain.

Although there is a consensus among metrologists that charging and contamination prevent accurate metrology (AMAG, 1997), conventional GR&R studies of CD-SEM measurements practically reproduce the primary procedure developed for mechanical measurements. The latter does not imply any testing of possible hysteresis: since there is no influence by a micrometer on an object of control, the repeated variation in measurements characterizes only the random errors. By contrast, repeated CD-SEM measurements are always effectively accompanied by hysteresis due to charging and/or contamination. Thus, results of the SEM evaluation without distinction between these effects' influence and measurement error have little in common with true precision estimates.

Therefore, we propose to supplement the conventional GR&R procedure with the testing data obtained *before* the final stage of precision evaluation. The objective of this step is to test if there is only random difference between repeated measurements. Testing could be performed using either the Student's *t*-test for paired readings, or the Wilcoxon's signed ranks test. The latter is preferable, because it is distribution-free. Both tests imply comparison of the average difference between matched pairs (measured in units of difference standard deviation for the *t*-test) with some critical value. If the difference is insignificant, the obtained data could be used 'as is' to gage precision evaluation. A significant difference represents obvious evidence of repeated irradiation influence on the measurement results. In this case, true repeatability and reproducibility estimates could be obtained only after removal of induced trending. This can be achieved by simple subtraction of a linear term from the obtained data. Experimental data support the assumption of a linear physical model for both charging and contamination effects on the repeated measurements (Monahan and Khalessi, 1992).

A.L.D.'s original plan (Fig. 2) to achieve the GR&R study's goals implies the following five-step procedure ('short' form). This procedure has been validated for application on different SEMs and wafer structures.

<u>Step1. First Round Measurements.</u> Some sites on the wafer are measured under specified measurement conditions. Every site is measured twice on those layers (targets) used for production routine measurements.

<u>Step 2. Second Round Measurements.</u> After a specified period of time, the tested wafer is reloaded and the machine is reset. Two repeated measurements are performed as in previous step and on the same sites and layers.

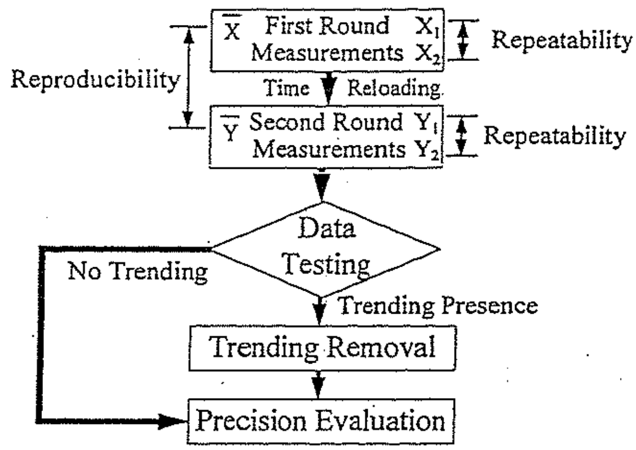

Figure 2. GR&R Study (Short Form)

<u>Step 3. Testing Obtained Data.</u> The Wilcoxon's signed ranks test is applied to every pair of adjacent repeated measurements performed on the same target, and placed in chronological order of the time they were measured. If the variation in these readings characterizes only the random error occurring in every measurement, the precision could be evaluated using the obtained data (Step 5). Otherwise, the data should be preliminary processed using the next step.

<u>Step 4. Trending Evaluation and Removal.</u> Once a data trend is identified, its slope is evaluated using the Least Square Method for Linear Regression and taking into account that the least square line passes through the result of the first measurement on a 'fresh' target. Compensation for the net effect of induced trending on the obtained data is performed by simple subtraction of the linear term as follows:

$$Y_i^c = Y_i^m - b(k_i - 1)$$

where Y_i^c and Y_i^m are corrected and measured values of CD, respectively, b is the computed slope, and k_i is a measurement. number

<u>Step 5. Gage Precision Estimation.</u> Values for both repeatability and reproducibility could be evaluated on a 'purified' database using the standard GR&R procedure. (A detailed description of GR&R can be found in several textbooks, such as Barrentine, 1991.) Repeatability evaluation is based on the average spread of paired measurements on the same targets; reproducibility is characterized by the difference between pooled readings for all targets at the 1st and 2nd steps minus the repeatability

contribution. Finally, the gage precision is derived as square root of sum of variances due to repeatability and reproducibility. Obviously, the obtained pooled estimate of reproducibility reflects differences in all processing conditions, including loading and positioning settings as well as stability—the changes of measurements over time.

This pooled estimate could be decomposed using the 'long' version of the same plan:

Step 1. First Round, First Set Measurements - Identical to Step 1 of the 'Short' form.

Step 2. First Round, Second Set Measurements - The tested wafer is reloaded, the machine is reset, and the second set of measurements is taken on the same targets.

Step 3. Second Round Measurements; Step 4. Obtained Data Testing; Step 5. Trending Evaluation and Removal - Identical to Steps 2, 3, and 4 of the 'Short' form, respectively.

Step 6. Gage Precision Estimation. The 'purified' data obtained in Steps 1, 2 and 3 are used for repeatability evaluation. 'Pure' reproducibility due to reloading, repositioning and reset is characterized by the difference between pooled readings for all targets in the 1st and 2nd steps minus the repeatability contribution. Overall reproducibility is characterized by the difference between pooled readings for all targets in the 1st and 3rd steps minus the repeatability contribution. Finally, the gage precision is computed as the square root of the sum of variances due to repeatability and overall reproducibility.

Note: Examination of the CD-SEM repeated measurements reveals, as a rule, biased estimates dependent on object of control properties. For example, it was a great surprise for us to obtain for the same SEM completely different values of the precision-to-tolerance ratio for three products with different structures but the same CD specification requirements: 8.9%, 22.3% and 48.7%. Thus the precision of the same gage should be considered as excellent, marginally acceptable and absolutely unacceptable, respectively! After applying the proposed procedure for removal of effects associated with feature transformations due to repeated measurements, the core estimates for two last products were computed as 10.3% and 9.7%, respectively. The first result did not change as no trending was detected for this product. One can see that the SEM's true precision is the same for the tested products.

Another problem associated with the metrology of CD-SEM measurements is accuracy evaluation. It can be overcome using either line width standards relevant to the kinds of features encountered in VLSI fabrication or with a traceable instrument. Unfortunately, neither is available to the semiconductor industry (AMAG, 1997), so any discussion concerning accurate CD-SEM measurements seems to be a Catch 22 situation. Achieving accuracy (which is extremely important for contact and VIA CDs) as well as obtaining more precise results for routine production measurements will be presented in further research in progress by the authors.

Customized Charting Methodology

Charts for Variables—The basic SPC concept implies the study of the sources of total process variation and its decomposition into controlled (inherent) and uncontrolled variation due to common and assignable causes, respectively. Obviously, reduction of output variability can be achieved only through identification, separation, and assessment of different sources of variation. Control charts proposed by Shewhart in the1920s, represent a simple but powerful tool used to track process variations and distinguish between inherent and excessive variations. Most readers are familiar with these charts, therefore we will not dwell on their principles or details.

A review of typical SPC schemes used in the semiconductor industry has shown that they usually use some suitable program for generating two charts for variables at different stages of the process: the $\bar{x}$-chart for central tendency monitoring, and the R-chart or S-chart for variability monitoring. The process engineer usually fully relies upon the background processing software, implicitly supposing that if charting is computerized, it is correct.

Unfortunately, experience shows that this null-hypothesis should sometimes be rejected.

To illustrate, let's consider the problem of monitoring a photolithographic process. Suppose that the CD value is measured at several different sites (from 3 to 9 as a rule) on a wafer, and some wafers (usually 1 to 3) are selected at random for testing from each lot. Modern quality control equipment directly transfers the data to a computerized system. The software interprets the obtained measurements as a 'rational subgroup' used for setting control charts, performs the calculations, and displays the charts on a monitor. Aghast, the process engineer observes 15–20% out-of-control points on the $\bar{x}$- chart, and intuitively feels that something is wrong because the process should be shut down every half an hour. The engineer concludes that SPC is merely a game that management plays; SPC is reduced to a buzz word on the shop floor, and the result is understandable frustration. We have often observed this phenomenon, and those who doubt its veracity are referred to a very interesting case study presented by Joshi and Sprague (Joshi and Sprague, 1997).

The problem is associated neither with an incorrect sampling plan nor with the erroneous procedure of calculating control limits on the $\bar{x}$-chart due to non-normal data, site-to-site correlations, or another factor. The data collection plan is perfect. It has been proven that the chart is quite robust, even under conditions of non-normal data and cross-correlation. The problem is that the $\bar{x}$-chart *should not be set at all.* This statement is explained using Fig. 3, showing the sources of inherent variation for the wafer fabrication process. Actually, the measurements performed on the same wafer at some fixed locations characterize the within-wafer variation (its non-homogeneity) only; they cannot be used for evaluation of between-wafer or lot-to-lot variation. (Roughly speaking, doing so is similar to evaluating differences between bolts using diameter measurements performed on the same bolt but in different places.)

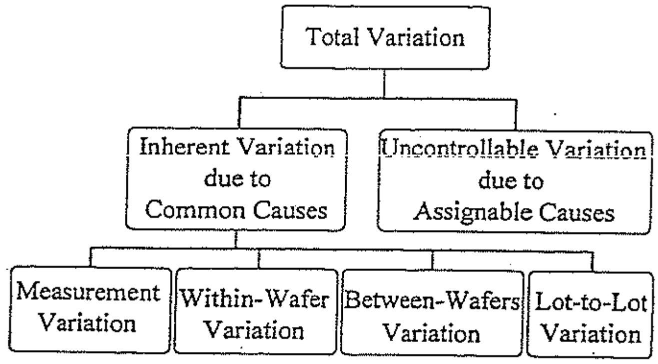

Figure 3. *Total Process Variation*

Generally, within-wafer variation is very consistent, because chips (the true 'production units') on the same wafer are simultaneously built and processed. The wafer is characterized by the pooled results of the measurements on the site level. While during some steps in the production process, such as diffusion, the wafers from the same lot 'run' together, there are some operations (for example, polishing) wherein each wafer is processed separately. Therefore, the between-wafer variation is usually larger than its within-wafer counterpart. By pooling the results of testing some wafers belonging to the same lot, we can obtain a process estimate at the top, lot level. The method of sequentially pooling the results obtained on the site level inherently implies that any attempt at using variation on the lowest level (site) for spread prediction on the higher ones (wafer or lot) will inevitably result in numerous process overadjustments. By contrast, any reverse transition will cause underadjustments when an appreciable process change due to an assignable cause is not detected.

To understand how individual sources of variation contribute to the overall variation, and to better identify where efforts should be concentrated to reduce overall variation, A.L.D. proposes a charting methodology based on the 3-level hierarchy of semiconductor manufacturing. The recommended scheme is intended to simultaneously correct variability evaluation at all levels.

Step 1. Set the R- or S-chart based on the site-level measurements for the monitoring of within-wafer variation. The number of points on the chart is equal to the number of tested wafers.

Step 2. Set the R- or S-chart based on the wafer means treated as individual measurements for the monitoring of between-wafer variation. The number of points on the chart is equal to the number of tested lots.

Step 3. Set the x-chart (not the $\bar{x}$-chart) based on the lot means treated as individual measurements for the monitoring of lot-to-lot variation. The control limits are calculated using the moving range of two successive lots. The moving range chart itself is not constructed because it does not provide any additional information about the process's variability. The number of points on the chart is equal to the number of tested lots.

Note that for the situation when one wafer from each lot is tested, the between-wafer and lot-to-lot variations are indistinguishable, and only two charts (at the 1st and 3rd levels) can be used.

Fig. 4 illustrates the problem of the conventional $\bar{x}$-chart, and shows the 'family' of recommended charts set on the CD data. One can see more than 30% out-of-control points beyond the control limits on the $\bar{x}$-chart, whereas the 'family' exhibits 'hard' evidence of an absolutely stable process.

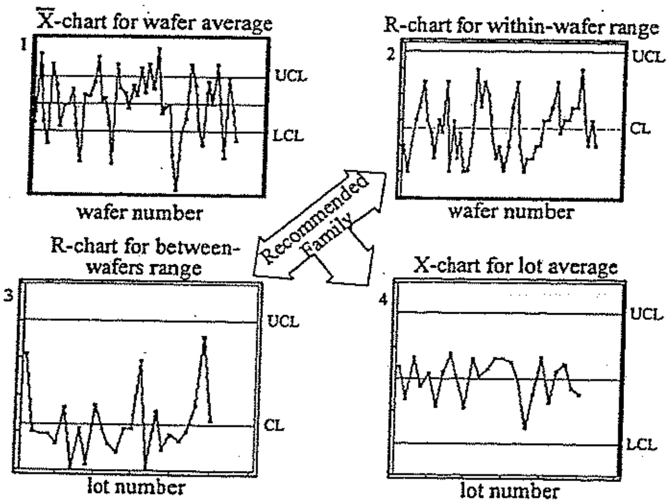

Figure 4. *Commonly used* $\bar{x}$ *- chart, 2; 3; 4 – recommended family*

Chart for Attributes—A chart for fraction nonconforming (p-chart) is usually used in the semiconductor industry for estimating current yields. There are some problems associated with this chart. First, if the fraction nonconforming p is very small, the sample size n should be large enough to provide a high probability of finding at least one nonconforming unit in the sample; otherwise even one defective will result in a salient point outside the upper control limit. Besides, n should be large enough to achieve a lower control limit, indicating the boundary between random variation of the process and its improvement. Obviously, extremely large samples are needed for today's very small p-values. Second, the conventional p-chart is practically unable to recognize small process level shifts due to its rather low sensitivity. An additional problem is associated with the binomial probability model underlying the setting and analysis of a p-chart. This model is not very adequate for wafer fabrication processes wherein nonconforming units (chips) are usually clustered on the wafer, because one of the model's basic assumptions is that successive units of production are independent.

To overcome these obstacles, we propose the Time-Between-Defectives chart, which reverses the situation and displays the number of defective-free samples. The higher the proportion of defectives - the shorter the time between them. Applying the usual procedure of data transformation to the exponentially distributed time between defectives, one can obtain a standard normal distribution and set a control chart with two control limits. Points above the upper limit indicate improvement of the process, whereas those below the lower limit indicate process drift.

Another alternative to the conventional p-chart is its moving counterpart. This method is successful for tracking sample averages. In its unweighted version, the technique is rather simple: after combining some successive samples, the average value is calculated for this pool, then the 'oldest' value in the set is dropped and the newest one is added. The weighted version requires assigning different weights to the current and preceding samples.

These charts preserve sensitivity to small process drifts while simultaneously limiting the overadjustments by combining more than one sample. They are practically insensitive to the interdependence of successive units of production. The lower control limit can be set even for small sample sizes, providing the opportunity for process improvement and troubleshooting problems associated with the upper control limit.

'Pure' Capability Study

The conventional SPC strategy involves judgment about the stability and uniformity of a process, i.e., the ability to yield products with specified properties. Uniformity is evaluated by means of a capability study that describes process performance in terms of customer requirements. Potential process capability is characterized by the capability index C_p, equal to the ratio of the distance between the specification limits and six process standard deviations. Real capability is evaluated by means of the index C_{pk} that accounts for the real process location. This index is computed by taking the minimum range between the process mean and a specification limit divided by three process standard deviations. Obviously, $C_{pk} \leq C_p$ (C_{pk} is equal to C_p when the process is centered). A C_{pk} value of 1.00 is considered to be a *de facto* standard, indicating that the process enables producing a product that conforms to specifications.

Transition from Quality Control to Quality Assurance unavoidably involves the translation of formal quality requirements from the language of inspection (percent defectives) to that of SPC—capability indices. These indices are used for self-evaluation and as a kind of statistical report for customers. In fact, indices have become part and parcel of quality requirements that many customers impose on their suppliers. To illustrate, Section 4.9.3 (Ongoing Process Performance Requirements) of QS 9000 reads as follows: 'For stable processes and normally distributed data, a C_{pk} value ≥ 1.33 should be achieved.' (SPC Reference Manual, 1995). The same requirement is presented in Section 4.20 of the Semiconductor Supplement to QS 9000.

Experience shows that the capability index values are seldom calculated correctly. This is usually due to not understanding the difference between process performance indices (P_p, P_{pk}) and capability indices (C_p, C_{pk}), and substituting one for the other. The only but very important difference between them is that performance indices characterize the total process variation due to both common and assignable causes, whereas capability indices are associated with the inherent variation due to only common causes. The total variation estimate is given by the process standard deviation σ_{total} computed by a single pass through all sampling data. The inherent variation could be evaluated using a stable chart for spread, i.e., a chart without any salient points due to some assignable causes. The $\sigma_{inherent}$ value is given by the centerline on the R- or S-chart divided by a tabular factor d_2 or c_4, respectively, whose values depend on sample size.

Even the most powerful and commonly used statistical software packages sometimes confuse P_{pk} with C_{pk}. As already shown, C_{pk} calculations should be based on the results of the process stability analysis, i.e., directly connected to the control chart setting. However, some packages (such as JMP from SAS Institute, Inc.) suggest that capability evaluation is a part of the static distribution analysis, actually computing P_{pk} while calling it C_{pk}.

Taking into account the 3-level data hierarchy, the ANOVA procedure for SPC in semiconductor manufacturing could be written as follows:

$$\sigma^2_{inherent} = \sigma^2_{within\text{-}wafer} + \sigma^2_{between\text{-}wafer} + \sigma^2_{lot\text{-}to\text{-}lot}$$

If, for example, the R- and S-charts are used for monitoring within-wafer and between-wafers variation, respectively, then

$$\hat{\sigma}^2_{Within-Wafer} = \left(\frac{\overline{R}}{d_2}\right)^2$$

$$\hat{\sigma}^2_{Between-Wafers} = \left(\frac{\overline{S}}{c_4}\right)^2 - \frac{\sigma^2_{Within-Wafer}}{n}$$

$$\hat{\sigma}^2_{Lot-to-Lot} = \left(\frac{\overline{MR}}{1.128}\right)^2 - \frac{\sigma^2_{Within-Wafer}}{nk}$$

where: $\overline{R}$ is the average range, $\overline{S}$ is the average standard deviation, $\overline{MR}$ is the average moving range, n is a number of tested sites on a wafer, and k is the number of wafers drawn from the same lot.

Combining the above equations, we obtain

$$\sigma^2_{Inherent} = \left(\frac{\bar{R}}{d_2}\right)^2\left(1-\frac{1}{n}-\frac{1}{nk}\right)+\left(\frac{\bar{S}}{c_4}\right)^2+\left(\frac{\overline{MR}}{1.128}\right)^2$$

A comparative analysis of the process capability estimates performed on numerous actual cases showed that the erroneous calculations tend to underestimate the capability index by as much as 30–40% of its true value. This serious problem of underestimating indices could result in costly process overadjustments, hyperactive product/process redesign, excessive quality costs, and misleading both supplier and customer.

Strategy of Process Improvement

Quality improvement can be achieved by reducing either the process's output variability, or its failure rate, or both. SPC is focused on variability in the data used as a basis for effective process management, whereas the failure rate is a major concern of reliability programs, and will be considered later.

SPC Strategy—The SPC objectives could be formulated as the detection, recognition and removal of any assignable causes, and also the correct evaluation and elimination of every root source of inherent process variation. As a rule, a process is neither stable nor capable at starting position 1 as shown in Fig. 5. This means the process is not operating in the manner in which it was designed to run, and continued operation under such conditions is merely wasteful. The inherent variation cannot be evaluated yet due to continuous disruptions within the core process. Abnormal variation is dominant, therefore a process set-up starts with a necessary empirical effort to bring about and maintain stability.

Identifying the main sources of an 'out-of-control' state by linking the subgroup identifier to information about process performance conditions encourages the process's galloping toward 'purification' and quick removal of obvious assignable causes. These local actions are usually within the ability of the operators or local supervision and mandate, as a rule, tool replacement, machine or gage recalibration, etc. Management is less involved in resolution at this level. Management involvement is required, for example, when personnel retraining is necessary, standardization can be used as a countermeasure against assignable causes, raw materials or parts are unacceptable and dealing with vendors is on the agenda, etc.

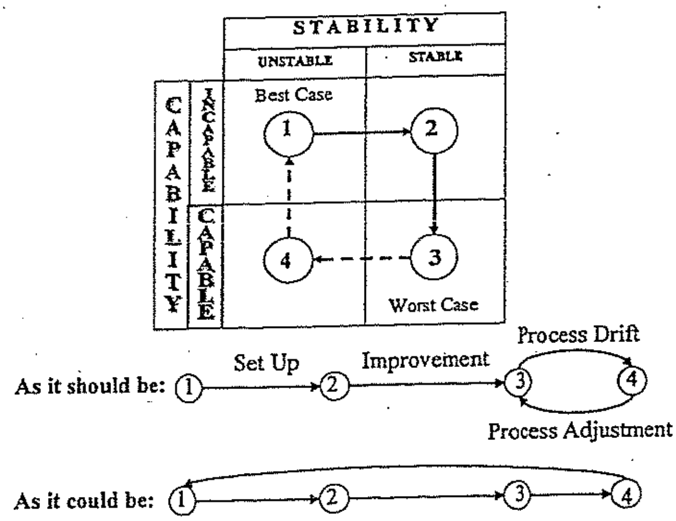

Figure 5. SPC Over Process Life Cycle

Suppose that the process was brought into a state of statistical control (position 2 of Fig. 5), but still fails to meet the customer's requirements. The short-term period of fast SPC achievements by means of 'local' actions is over, and now the process's inherent variability is of interest. Both Juran and Deming warn against confusing between removing assignable causes and process improvement. They emphasize that the former only brings the system back to where it should have been in the first place, while the latter represents the long-term program of never ending gradual reduction of inherent process variability.

Usually, the system-related common causes (such as raw materials, machinery, control equipment, technology, working conditions) are beyond the abilities of operators and local supervision to correct. Instead, they require management involvement to make basic changes in the process. At the same time, machine performance depends on operators (inaccurate settings, sloppy workstation, etc.). Thus SPC requires the participation of the entire company—from top management to shop floor personnel. The only difference is the degree of responsibility for eliminating assignable causes and reducing inherent variability.

We are now at position 3 of Fig. 5, where the process does meet customer expectations. Think of what could happen if a process remains at this point over time: no bad lots, no machine failures, no untrained operators, no measurement errors...now stop dreaming. Processes do vary; they wander over time. The main goal of process monitoring at position 3 is to *prevent* the appearance of defectives as the result of stability loss. The common notion of process drift implies two subsequent stages in deterioration. First, the process loses stability (transition from position 3 to 4 of Fig. 5). Left alone, an unstable process can 'explode' in any direction. What is remarkable is that any such explosion usually leads back to position 1 in Fig. 5, i.e., to poor quality product yields. Therefore, the SPC adjustment policy applies immediate process correction at the boundary 'stable/unstable' (Fig. 5), in contrast to the 'Out-of-Tolerance' adjustment policy based on FI and associated with the 'capable/incapable' boundary. The latter, having nothing to do with any process trends or lack of statistical control, provides feedback signals when a process is already in 'syncope.'

Reactive and Proactive Actions Choice - In the full improvement cycle "**Detect – Isolate – Fix – Verify**" , the first and the last belong to SPC. And what about the tools for problems "isolation" and "fixing"? All of them are well known: Pareto diagram, Flow-Chart, Fishbone diagram, etc., complimented by different experimental practices. We demonstrate the Process FMEA as a powerful tool for fault isolation, better control and corrective actions. Being actually a "corporate memory" Process FMEA is a Knowledge base and Expert System for Control and Decision Making. Proactive QA directed at problem prevention implies elimination, or at least reduction, of the impact of possible product/process malfunctions and failures. Applying FMEA in the early stages of the product life cycle, when changes in products/process are relatively simple and easily implemented, helps to expunge painful syndromes such as late change crises. One of FMEA's most powerful advantages is the ability to take verbal ideas for process improvement and system behavior rules, and transform them into concrete numerical analyses and implementation plans. These ideas and rules comprise a Knowledge Base (KB) which should be incorporated into artificial intelligence systems for expediting and enhancing the improvement efforts. KB is a combination of 'Declarative' and 'Procedural' knowledge (DK and PK, respectively). It is a collection of FMEA libraries, and serves as an organization's 'collective memory.' DK includes databases of failure modes and failure causes on the component level, corrective and preventive actions libraries; end effects and associated severities library; and tests library (detectability, methods, levels, types, etc.). PK includes descriptions of 'next-higher-effect' chains and 'end-effect' allocations. By exploiting the entire knowledge base, advanced engineering methods, and the expertise of a company's personnel at all levels, FMEA allows for analyzing all possible malfunctions, identifying points at which control should take place, and ensuring prevention of potential failures.

Conclusion

SPC is now accepted as the operating procedure for wide range of industries. It provides a common denominator for transitions between all phases of a process, and thereby reduces operational barriers. Applying the right SPC tools at the right time is as important as having a database full of statistics. Therefore, we have tried to develop a statistical methodology adopted and customized for the semiconductor industry. The proposed methods do not require any additional data other than the information used for conventional SPC. The enhancement is that advanced SPC applies engineering expertise and novel statistical methods to raw data, combining them all into a 'clever' database. The motto for advanced SPC is 'Better *versus* Current'—improving processes in an evolutionary manner through minor changes and low investments. Its strong yield orientation restores SPC's image of being a positive, important, and value-adding activity.

References

1. Advanced Metrology Advisory Group (AMAG). *Unified Advanced CD-SEM Specification for Sub-0.18 μm Technology*, Revision 1. 1997.

2. B. Barrentine. *Concepts for R&R Study*, ASQC Quality Press, Milwaukee, Wisconsin. 1991.

3. Z. Bluvband, S. Shahaf. "Sparing Criteria - Clear Management Approach, *Reliability and Maintainability Symposium*, San Francisco, California, 1984, pp. 446

4. B. Efron. "Nonparametric Estimates of Standard Error: the Jackknife, the Bootstrap and Other Resampling Methods". *Biometrika*, 1981, 68, pp. 589-599.

5. P. Grabov, Z. Bluvband. "Statistical Process Control : New Charting Technique". *Statistics in Industry, Science and Technology, Proceedings of International Conference*, Tokyo, 1994, pp. 161-166.

6. M. Joshi, and K. Sprague. "Obtaining and Using Statistical Process Control Limits in the Semiconductor Industry". In *Statistical Case Studies for Industrial Process Improvement*, edited by V. Czitrom and P. D. Spagon, SIAM&ASA, Philadelphia, 1997, pp. 337-356.

7. M. Monahan, and S. Khalessi "Application of Statistical Models to Decomposition of Systematic and Random Error in Low-Voltage SEM Metrology," SPIE Vol. 1673 *Integrated Circuit Metrology, Inspection, and Process Control* VI, 1992, pp. 36-39.

8. W. Shewhart. "Economic Control of Quality of Manufactured Product." Reprinted by ASQC, Milwaukee, WI, 1981.

9. *SPC Reference Manual*. Chrysler Corp., Ford Motor Company, GM Corp. 1995.

Acknowledgments

The authors wish to acknowledge A. Burger, Y. Cohen, M. Gutman , M. Menaker and especially M. Leibovich (Israel Semiconductor Industry) for their helpful interactions and support.