Challenges For Use Of Statistical Software Tools In The Semiconductor Industry

Richard Kittler
Yield Dynamics, Inc.
Santa Clara, California

Contact Information:
Richard Kittler, Yield Dynamics, Inc., 2855 Kifer Road, Santa Clara, 95051, USA,
Phone: 408.330.9320 x105, Fax: 408.330.9326, E-mail: rich@ydyn.com

Keywords: **software statistics yield semiconductor manufacturing**

Abstract:

This paper surveys statistical methodologies and associated software tools in use within semiconductor manufacturing to support process characterization, process control, and yield improvement. Current difficulties and challenges are discussed as well as emerging trends that will require the development of new methods and the software to support them.

Introduction

The practice of process development and manufacturing in the semiconductor industry requires knowledge and application of a broad range of statistical methods. Processes are typically so complex that large amounts of data need to be collected to characterize process windows and manage variability during manufacturing. Storage and analysis of this data today requires use of advanced database, data integration, and analysis capabilities.

The industry has thus far evolved a mixture of such capabilities. For lack of commercial alternatives, larger semiconductor manufacturers were forced early on to develop many capabilities in-house. Commercial suppliers then emerged but tended to focus only on particular aspects of the whole. Early efforts at automation evolved within the process tools by the tool manufacturers and at the factory level in the form of shop floor control systems. Following this came basic SPC capabilities, either standalone or connected to the shop floor control system, together with basic tool communication capabilities to enable automated data collection and recipe control. Later, various forms of extracts were set up to feed subsets of the data centralized in the shop floor control system to offline databases coupled to standalone analysis tools such as SAS [1] and RS/1 [2] to provide engineers with improved tools for characterization and improvement of the process.

This paper surveys the methodologies that software tools are being built to support in the areas of process characterization, process control, and yield improvement. Current difficulties and challenges are discussed as well as emerging trends that require the development of new methods and the software to support them.

The remainder of the paper is organized as follows. The next section provides an overview of the data infrastructure currently in place to support semiconductor manufacturing engineers. We next survey typical tasks performed during process characterization, process control, and yield improvement and the methods in practice to accomplish them. The following sections describe the difficulties of the current methods, trends that are emerging in current practice, and the challenges of developing new methods to meet the need for the timely solution to problems of every increasing complexity.

Manufacturing Context

We begin with a brief overview of the semiconductor manufacturing process. Circuit manufacturing begins in a facility called a ‘fab’ that is responsible for processing blank slices, or wafers, of single crystal silicon through until they are patterned with finished devices ready for separation and packaging. Wafers are processed through the line in batches called ‘lots’. Each lot has 25 to 50 wafers and is most often dedicated to making parts for a specific circuit type. During the course of processing they are deposited with various films and patterned through lithographic and etch steps. Today’s leading edge processes typically involve 300 or more processing steps and printed transistor features on the order of 0.18 micron. Detailed data is collected following many of the process steps and monitored using statistical process control to insure that control of the process is maintained. Near the end of processing electrical testing is done on test structures distributed across the wafers. This data is used to maintain monitors of device performance and provide the diagnostics necessary to quickly react to problems if critical device characteristics begin to drift due to unanticipated process changes. When fab processing is completed the circuits, or die, on the wafers are probed electrically to determine functionality. The data from this test step provides critical feedback on the yield of the manufacturing process and the relationship between the transistor characteristics and the performance of the completed circuit. Such information is used to drive yield improvement efforts and improve understanding of process characteristics that could impact new process technologies and devices under development. Following wafer probe, the wafers are sent on to an assembly facility where they are sawn up into individual chips, packaged, and re-tested for final performance in the package. Final testing may also involve some accelerated life testing to insure early life failures are screened out prior to shipment. Following shipment, any customer returns are subjected to detailed analyses to determine the root cause of failure.

Data is collected from many sources during processing to both characterize critical process steps and track their variability. Data collection is more intense during the development and introduction of a new process than later on when the process is in volume production. Data collection is added as new sources of variability are discovered. When sources of variability are removed or reduced, some vestige of the data collection plan often remains as an insurance policy. Since succeeding generation of processes only become more complex, data collection has therefore increased over time. Today there are elements of data collection from the very beginning to the very end of the production process. This includes data on the raw materials used in processing, the history of tools used to process a lot, metrology data, defect inspection data, electrical data, probe yields, assembly processing histories, assembly sub-contracting histories, final test, customer returns, SPC violations, design revisions, dates of process changes, etc.

As data collection has evolved from various sources it has only recently begun to be knitted together into integrated data repositories from which data can be pulled to address problems. The completeness of these repositories is still sorely lacking in many areas and there are many problems with managing the data that is available from them. These problems include regular gathering of data from a diversity of sources, timely loading, storage of the huge volumes of certain data types, e.g. wafer probe, and maintaining acceptable response times for data retrievals to support analyses. In addition, due to the diversity of data types there are not always unambiguous methods of aggregating and merging the data of various types to feed into an analysis. Due to different data collection plans some data may be collected at the lot-level while other is collected on a few random wafers in the lot, or multiple sites on selected wafers, etc. In the extreme measurements are made on every site on the wafer in a mode know as wafer mapping. At some locations in the process the order with which wafers are processed is also recorded in an effort to investigate the variability associated with processing sequence.

As lots progress from one portion of the process to another and enter different manufacturing facilities the lot numbers are often changed, the lot may get split into several sub-lots, and re-merged later on into yet another lot number. This renaming of lots makes traceability difficult and requires considerable effort to setup and maintain the systems necessary to organize and make available this lot genealogy.

In addition to the data associated with a lot some of the data is associated with the tools, e.g. the repair and calibration history of the tools or environmental measurements of air or chemical quality. Recently some of the internal sensor data within process tools, known as trace data, has also become available through extensions of the equipment interface capabilities.

To access all of this data in it’s various forms and locations there exists a combination of general purpose tools for accessing commercial databases, in-house applications built on top of the databases, and commercial software applications that bundle a database together with access and analysis tools. Due to the complexity of the data mentioned above most of these tools are designed to address certain tasks well rather than the whole of what might be needed by a semiconductor manufacturer's engineering group.

Tasks and Methodologies

The types of analyses that are performed on these data are particular to the engineering functions within development and manufacturing. Within each group, if statisticians are present, they will play an important role in defining the methodologies to be used, the selection of software tools to support the methodologies, and in providing expert guidance in the analysis of data.

Of the engineering groups, *product engineers* are responsible for interfacing between the design, development, and manufacturing groups, and often have responsibility for the characterization of new products. Characterization involves running of so-called ‘corner lots’ on a new design to characterize the performance of the product at the extremes of the data sheet specs. Ideally this is based on a well-defined methodology involving designed experiments. The characterization data is then used to model the variability of the process across the spec range at different test conditions and establish guard band limits for product testing. The statistics are not complicated and the difficulties come mainly in organizing and reducing the data prior to basic plots and analyses.

Process development engineers run designed experiments to characterize the capabilities of new tools and develop recipes to fabricate new structures. They also work to characterize the stability of new processes through control charting and process capability studies prior to the handoff to manufacturing. In extreme competitive situations the process will not have reached stability prior to handoff and manufacturing will have the job of completing the task.

Technology development engineers invent new structures and work with the process development engineers to fabricate them for electrical characterization. Very often the electrical characterization of these new structures will use special test patterns with thousands of electrical tests. They may also use process characterization data to better calibrate the simulation tools used to model the performance of new device structures.

Integration engineers assist in the transfer of new processes to manufacturing, provide ongoing review of process change requests, review the design and protocol for experiments run by manufacturing, and act as liaison between manufacturing and the development and product groups.

Manufacturing engineers are responsible for ramping volume on new technologies and tools received from the development groups. Their tools of the trade are control charts, ongoing process capability studies, trouble-shooting guides, and designed experiments to test process improvements. Due to the benefits of automation, the task of setting up the proper charts and staying on top of them is often overwhelming.

Yield engineers monitor probe yields, analyze low yielding lots, and correlate yield data to process histories to understand the important factors influencing yield. They are responsible for profiling yield limiting factors and working with the manufacturing engineers to eliminate yield problems. They are the biggest drivers for the development of comprehensive repositories of engineering data collected during manufacturing and the biggest users of data analysis tools for finding relationships in the data. Their responsibilities are to ramp yield following the introduction of new technologies and products into volume manufacturing and to respond quickly when processes experience yield excursions. In addition to exploratory data analysis through use of SAS JMP, Statistica, RS/1 or commercial yield management software, some of the methods being used include automated :

- ANOVA or Kruskal-Wallis analyses to look for tool-to-tool performance differences
- Regressions to monitor the relationships between electrical parameters
- Commonalty studies to look for what is in common across a group of bad lots
- ANOVA or Kruskal-Wallis differences analyses to contrast high and low yielding lots
- Linear and time-series models to forecast probe yields based on in-line metrology data
- Searches for systematic behavior in process sequence data, e.g. every other wafer effects due to the use of two-chamber process tools
- Routines to correlate SPC violations to prior processing histories

For the most part, methods in common use are based on traditional parametric and non-parametric statistical methods such as regression, ANOVA, chi-square, and Kruskal-Wallis that are found in commercial statistical analysis software. New methods such as data mining are just starting to come into use. Yield engineers must also find ways of integrating other forms of data such as that coming back from failure analysis and in-line defect metrology data, for which there are no clean models, but for which root cause relationships are known to exist.

As common threads to all of the analyses performed by the various engineering groups there are ongoing questions regarding:

- the statistical methods to use,
- the battle between an engineer’s desire to automate or ‘can’ analyses and the statistician’s preference for interactive data exploration,
- confusion over when to base a decision on raw data and when to use summarized statistics, and
- what to show in graphical form and what to tabulate.

Difficulties

Much of the difficulty in the analysis of data in semiconductor manufacturing is due to the diversity of data and complexity of its interrelationships. This has led to point solutions evolving within each of the engineering groups and slow progress toward an integrated data and analysis tool environment. These difficulties pervade both the commercial statistical software tools, which are too general to address semiconductor-specific data access or the analysis tasks, and the commercial semiconductor software applications, such as yield management and SPC packages, which are too specific to cover the breadth demanded by the spectrum of tasks described above. Although general statistical software has an advantage in that almost any task demanded of it by the engineer is available, to use them one must often progress through a lengthy series of manual steps. On the other hand, industry-specific packages are designed to be very efficient to use, but trade this off against a narrower scope of use.

The ease and flexibility of data access is still a major problem for many of the engineering groups. Data is often spread across various systems, the names of many of the collected parameters may follow obscure naming conventions, data often has to be cleaned in some manner prior to use, and there are often times no easy methods for merging the data together from the various sources. All of this results in long lead times to gather the data together to run analyses leading many engineers and statisticians to lament that it takes 90% of the time to get the data and only 10% of the time to do the analysis! As commercial tools directed at the semiconductor engineering market evolve they will obviously need to address this issue.

In much of the characterization work, in looking for root cause on SPC excursions, and in almost all of the yield enhancement efforts, there can be many variables and relatively few observations. This is because with no a priori knowledge of where a problem originates engineers will want to search through all of the known data about recently processed lots to see if the can find any relationships. Compounding this situation is the fact that there are often significant outlier effects. When analyzing an excursion the outliers are important while in other cases it may be necessary to employ some robust methods of screening them out so that the main portion of the distribution can be analyzed. How to properly deal with outliers and avoid false signals remains a difficult task to learn and perform. Automated analyses can be particularly susceptible to generating false signals and sending engineers off on wild goose chases unless careful thinking goes into outlier and sample size handling. Early efforts to automatically to search for relationships were very prone to this type of problem, and many still are.

A related problem comes up in automating SPC charts. There can be thousands of charts if one blindly creates a chart for each of the 300+ process steps and each of the tools at those steps. Given this number of charts, how does one monitor them all, and how does the engineering group avoid becoming numb to the constant barrage of out of control email messages?

With a wealth of data available and an opportunity to automate the generation of graphs and statistical summaries, engineering groups look constantly for new methods of generating and visualizing health metrics for components of the manufacturing process. The most common versions of this are Cpk-related, i.e. Cpk reports, composites of Cpk statistics from different areas of the process, and Cpk trend charts. These metrics, though the result of a need to abstract only the essential elements for decision-making, can be deceiving if one loses sight of the components of the index and their underlying dynamics.

Visualization has other challenges. Some are related to the difficulty of accomplishing them in software, while others are awaiting the development of new methods. For instance, wafer maps show the variation of a measured device characteristic across all of die locations on a wafer. While it has been shown that using a glyph of the wafer map as a plot symbol can significantly extend the informational content of a scatter plot or a trend chart [3], no commercial software is yet capable of this type of graphic. In another case, a yield engineer finds that yield is lowered due to an interaction between two process tools with range of processing dates and is unclear how to represent this relationship visually. There are many visualization problems which yearn for the sort of elegant solution developed by Minard to describe Napoleon’s Russian campaign [4].

Some of the graphical information that is easily perceived visually is more difficult to extract programmatically. The spatial variation that is present on wafer maps may often be the result of process-induced non-uniformities that are systematic in nature. Since it is impossible for an engineer to visually scan the wafer maps associated with all testing, algorithms are needed to extract features from maps in a rigorous fashion. However, not unlike the false signals that are generated by some the automated routines to search for tool-to-tool differences, there is well-founded skepticism regarding the reliability of algorithms to do such feature extractions. If a method can be shown to be reliable, then the next step would be to use this feature strength information as a response and look for correlation back to process histories, such as the order of processing at each of the process steps. The same goes for the ability to extract features from x-y plots of electrical parameters and plots of key device parameters versus processing sequence. Out of all of the possible plots the engineers would like to look at those that contain something interesting, and for those relationships that are known to be serious, the presence of a significant signal needs to trigger follow-on analyses to trace potential problems, or pre-cursors, to root cause.

This need highlights two further difficulties. First is the general lack of power in automating analyses and second the lack of general methods for capturing knowledge and embodying it in a form where it can be both used to train new members of the organization and embedded into automated line monitoring routines. Automation methods need to be capable of easily testing any aspect of a prior result, reporting on it in a flexible fashion, and conditionally triggering follow-on analyses. There should also be the ability to build classes of methods to eventually take on more and more of the routine monitoring and follow-up analyses.

Faced with the complexity of the data and the diversity of statistical methods available to them, engineers often take the simplest rather than the most appropriate path. Analytical tools designed for the semiconductor engineering community need to eventually offer structured analysis methods in the form of guided macros or wizards that will understand the context of the problem from the initial steps that a user has taken and suggest next steps.

Some of the data collected during processing is not directly associated with the processing of a lot. Rather, it is environmental data or data associated with a processing tool such as calibration, preventative maintenance events, or qualification data. Such data can be analyzed today using generic methods. The difficulty comes when engineers would like to associate this data with the processing of lots. There are no easy methods in commercial tools for interleaving data to associate a time-stamped event with lot processing. These are needed in order to answer questions such as whether there is any relationship between lot yield and the amount of material processed since the last preventative maintenance.

Outliers pose a significant problem with many types of data. Among these are electrical test data, where testing problems can cause unusually large values, and defect metrology data, where a die, which normally would have zero to one defect, might record tens of thousands of defects. Unless the underlying distributions are understood outliers may often be erroneously excluded. In some such cases it is the outliers that contain real information about the variability of the process. In other cases a data transformation may be appropriate to bring these values into the realm of other data. The complexity of ‘what to do when’ is non-trivial and needs further work to develop context-sensitive wizards to guide the user.

Emerging Trends

There are some notable trends in systems capabilities that offer hope for addressing these difficulties. Most significant is the availability of semiconductor-industry specific commercial tools with more flexible and powerful integrated data access and analysis capabilities. While general statistical software such as MS/Excel, SAS, S-Plus, Statistica, and others will continue to improve in the depth and breadth of their features, they will still fall short in providing the industry-specific automated routines needed by engineers. In addition to one or more general statistical software tools, every company will need a complementary in-house or commercial yield management tool.

Since many semiconductor companies will develop new data sources over time, analysis applications will need to be flexible enough to interface to new data sources as they evolve. This includes the ability to couple with commercial data warehouse technology. In a similar manner, the analysis framework will need to be able to incorporate new statistical technologies, such as data mining, or provide the customer with the ability to easily link them in through standard programmatic interfaces using a plug-and-play paradigm. This would parallel the manner in which processing tools come with standardized communication interfaces and are the building blocks of the factory.

Challenges

One of the biggest challenges for the use of statistical software tools in semiconductor manufacturing remains to significantly improve the flexibility and ease of data access. These data access methods must be general and yet have variants that are very targeted toward common types of problems. Merging of data must be totally automated and new data sources must be able to be brought online easily with little to no additional user training in understanding what is available. In some cases, additional data should be fetched behind the scenes to supplement or meet the needs of analyses as they progress without the user having to go back and request them. Oh yes, and it needs to be fast.

Methods of automating the access and analysis of data need to be improved so that it does not take a programmer to automate the types of line monitoring performed by today's engineers. Families of watchdog agents need to be developed and evolved to take responsibility as intelligent overseers of regions of the process.

Of course these software systems challenges impact little on the work of the statistician. For the statistician, the challenges are large nonetheless! Statisticians need to guide the development of increasingly powerful analytical tools and graphical techniques to couple with the data access and automation capabilities. After all, it makes no sense to automate nonsense. Among the more significant challenges for statisticians are the needs to develop:

- New forms of visualization together with standardized methods to extract features from wafer maps, sequence data, and x-y plots, etc.

- Methodologies for recognizing and dealing with outliers in different types of data need to be more fully developed and embedded in a context sensitive fashion

- An understanding of alternate statistical methods that could be useful, such as decision trees, how they relate to traditional methods, and when they should be used.

- Methods for correlating lot- and non-lot based data together with new graphical techniques for visualizing these relationships

- Suggestions for sequences of analyses that make sense and can be embedded in the form of wizards which can be selected or will pop up when they detect the user going down a certain road.

- Rules to follow when looking at graphs or the output of methods that alert the user that something unusual is present, such as an odd cluster of outliers on a plot, and will advise on next steps to take, or offer to take them if the user desires.

Some of these challenges may already be solved in some semiconductor companies. None are yet fully embodied in commercial software. It is likely that it will be several years before elements of all of them are realized. Your help is needed and would be appreciated in shortening this timeline!

Conclusion

With the large amount of data being generated to support semiconductor manufacturing new methods are needed to ease access, bring diverse types of data together, and provide analyses. Point solutions need to evolve into global capabilities with applications capable of embedding knowledge in the form analytical agents that assist the user along the way and provide them with the benefits of prior learning. New statistical technologies, such as data mining, and database technologies, such as data warehousing and online analytical processing will act to enable portions of this evolution.

Acknowledgements

Thanks to Jon Buckheit and Weidong Wang of Yield Dynamics for their helpful suggestions on improving the manuscript and to Barry Hembree of Advanced Micro Devices, Inc. for his suggestion of the topic.

References

1. SAS is a product of SAS Institute, Inc.
2. RS/1 is a product of Domain, a division of Brooks Automation.
3. "Process Monitoring and Improvement Through Spatial Analysis of Wafermap Data", Vijay Nair, David A. James, and Mark H. Hansen, *1999 Statistical Methods Group Symposium*, *SEMATECH, Austin, TX*.
4. The Visual Display of Quantitative Information, Edward Tufte, Graphics Press, Cheshire, Connecticut, 1983, p. 40.

Biography:

Dr. Kittler is currently Vice President of Product Development at Yield Dynamics, Inc., a leading supplier of yield management and process control software to the semiconductor industry. Prior to joining Yield Dynamics he spent nearly twenty years at Advanced Micro Devices, Inc. in yield management software development and management roles, most recently as AMD Fellow and Fab Integration and Modeling Manager. He was the principal architect of AMD's own internal yield management tool, holds four patents in semiconductor process technology, and has numerous publications in the area of semiconductor yield improvement and analysis. Dr. Kittler received his Ph.D. in Physics from the University of California, Berkeley.