SOLID STATE TECHNOLOGY

Article Date	November, 1999
Magazine Volume	42
Issue	11
Title	The Emerging Role For Data Mining

Richard Kittler, Weidong Wang, *Yield Dynamics, Inc., Santa Clara, California*

The complexity of semiconductor manufacturing and its wealth of data provide an ideal environment for the application of data mining to discover patterns of behavior leading to knowledge. The demands are, however, such that data-mining technology needs to be extended further to achieve its full potential. Companies successful in developing and harnessing this technology will be rewarded with powerful new capabilities for diagnosing process control, yield, and equipment problems.

Each generation of semiconductor process technology demands more detailed data to support process monitoring and improvement. This is being driven by additional process steps and by new forms of metrology and flexibility in sampling plans. In turn, the wealth of data drives storage, integration, analysis, and automation trends in wafer processing.

Data storage. Islands of engineering data, generated and stored at the tool or cell level, are beginning to be integrated into repositories for analysis packages. These repositories make possible new types of queries that request data from across regions of the process. Thus, new sources of process variation can be traced to their root cause, leading to quicker yield ramps.

Data integration. To access data repositories, integration must allow analysis tools to pull

data from across the network. This can be done via a configurable data integration language or forced consolidation of all data into a proprietary in-house or third party database. Both techniques will remain in use until specific standards are developed to allow plug-and-play among commercial data storage, access, and analysis components.

Table 1. Comparison of Common Data Mining Methods

	Robust to Outliers	Physical Interpretation	Uses Historical Data	Easily Updated with New Data	Continuous Outputs	Discrete Outputs
ANOVA / Regression		X	X		X	X
Neural Nets			X	X	X	X
Bayesian Networks		X	X			X
Decision Trees	X	X	X			X

Data analysis. Analysis tools range from those provided on process tools to domain-specific applications in statistical process control (SPC), yield improvement, and equipment maintenance to ad hoc applications with third-party software packages (e.g., MS Excel, SAS JMP, and SPLUS). Many large IC manufacturers have built in-house domain-specific applications because comprehensive commercial products were limited.

Data analysis automation. Data volume and the number of relationships being monitored dictate a need for improved automation of line and yield monitoring. It is no longer sufficient to generate volumes of trend charts; charts must be analyzed automatically for exceptions and action must be taken if anomalies are found.

Users and suppliers of manufacturing execution systems (MES) tend to look at these problems from the top down; equipment suppliers from the bottom up. MES suppliers must create and maintain such capabilities at the factory level. Tool suppliers see market potential in increasing the communications capabilities of tools and packaging localized automation solutions within a work cell. Automated cells will slowly become more prevalent, similar to the evolution of multichamber tools. This will drive the need for cell management systems that treat the cell as a virtual tool. SPC used within the cell will need to be comprehensive, and the cell will need to be capable of tracing local process and defect related SPC failures to root causes. Eventually, the cell will need to communicate with neighboring cells to elicit information needed to maintain local process targets and anticipate tool wear and breakdown before they have an impact on the process.

Such automated systems will require infrastructure and behavioral models for their various components. The infrastructure is emerging through Sematech's work on a CIM Framework [1]. Behavioral models are already in use at the cell-level for such applications as run-to-run process control of overlay, critical dimensions (CDs), and chemical mechanical planarization (CMP) uniformity [2].

Developing control models involves tedious manual steps: gathering data, cleaning it, and repeatedly exploring it until an analytical model can be established. Techniques — specifically involving data-mining technologies (see "Data mining in brief," page 48) — are evolving, however, that allow some such modeling to be derived, or at least revised, automatically. The promise of such methods is that automated systems guiding tools, cells, and eventually perhaps even factories, will be able to learn and adapt their own control schemes to best achieve high-level directives. We believe that data mining is a key component of the technologies needed to achieve quicker yield ramps, fewer yield busts, higher capacities, and higher productivity metrics.

Current uses of data mining

Data mining is being used extensively in the retail, finance, security, medicine, and insurance industries. Many of these applications were built originally as expert systems. Today, a subset of such systems can be built automatically using data-mining technology to derive rules from contextual historical data. This has opened up new vistas of modeling heretofore deemed impractical because of the volumes of data involved and the transient nature of the models, especially in retail.

Some applications in these industries have analogies in manufacturing. As in the medical industry, for example, use of data mining in manufacturing requires that the derived models have physical understanding associated with their components. It is not enough to predict system behavior, since the underlying root cause of the phenomena usually needs to be identified to drive corrective actions. Also, as with medicine, most applications in manufacturing are for diagnostic systems. Manufacturing groups would like to know the circumstances under which they will encounter certain types of process control excursions, equipment events, and final quality levels. Having a system that can crunch historical data and establish related models improves response time when such events occur and follow-up is needed.

Caveats for manufacturing

Statistical significance does not always imply causality. Given enough variables, one or more will show up as significant regardless of the question asked. In other cases, the answer to a problem may be in the data, but the analytical methods are too narrow to uncover them. Both types of errors are more easily tolerated in off-line than in real-time

systems.

Avoidance of false signals requires ongoing work to improve models by better "cleaning" of the data, linking to previous knowledge, or including certain variables only if certain conditions are true. Today, statisticians or engineers do this weeding out process through the benefit of considerable domain knowledge beyond that contained in the dataset being analyzed.

When data mining misses the correct model, it is usually the result of the narrowness of the algorithms used. For instance, the data-mining algorithm may not be robust to "outliers" and therefore be thrown off-track by "dirty" data. Or, in the decision tree method, models at each node may not be complete enough to catch a behavior (e.g., cyclic vs. linear behaviors).

Data mining can also encounter problems when there are correlations among the variables used in the modeling. Such is often the case when the model involves variables related to the same region of the process. In these cases, the data-mining routine might choose to use one of the correlated variables based on its analytical criteria, even though a statistician or engineer would deem it to have no causal relationship and wonder why the proper variable was not selected. Data-mining tools that offer features for engineers to interact with the analysis and override variable selections help alleviate such black-box limitations.

Given the variety of caveats, the use of data mining in manufacturing is just beginning. The current interactive use will need to reach a level of maturity before being followed by embedding in real-time applications. To be viable for real-time use, data-mining engines must be applied to well-understood phenomena so that output is meaningful and trusted. To do so may require new forms of boundary conditions to define the limits for the solution space. Such rules define what is required of the process outside of the immediate context and what knowledge has been gained outside of the immediate dataset. Today such knowledge is referenced and used through participation of real engineers in the engineering change process.

Applications to wafer processing

In semiconductor manufacturing, possible uses for data mining include process and tool control, yield improvement, and equipment maintenance — analyzing historical data for patterns to derive physically meaningful models for predicting future behavior.

Process control is ripe for use of data mining because of the complexities of problem-solving processes and the richness of the data sources that need to be brought together. SPC is

usually supplemented with troubleshooting guides — out of control action plans (OCAPs). An OCAP is derived from protocol and engineering experience in solving previous problems. Data mining can assist in solving new problems by looking for commonalties in the processing history of previous occurrences — for example, diagnosing excursions of in-line defect metrology and wafer electrical test (WET) data.

In-line defect metrology data is fraught with analytical difficulties. Sophisticated kill-ratio schemes have been derived by correlating with sort data to assist in prioritizing the many out-of-control signals obtained from application of standard SPC methods to random defect data. Once the decision is made to work on a given defect type, analytical information from defect metrology, context, and history of the affected product, and other data, are assimilated by knowledgeable engineers to suggest and eliminate possible causes. This proceeds until the defect goes away or is traced to a root cause.

Given access to this same data, it should be possible to develop a data-mining application to assist in this complex task. Previous results would be embodied in an expert system against which new results would be compared. With a new failure, the system would interrogate other automation components and metrology tools to gather the data needed to suggest a likely cause.

Another possible application is diagnosis of excursion failures at WET. When lots fail at WET, the engineer responsible for disposition must assimilate the pattern of failing tests and either recognize the problem from previous experience or begin an analysis of process history to find the root cause. In the latter case, knowledge of the nature of the failure narrows the process steps to be investigated. Here, a data-mining application could encapsulate relationships between failing tests and the process flow, together with a methodology for gathering data on failing lots and analyzing it for commonalties.

Advanced process control applications close the loop. Process drifts are not only detected, but corrected by feeding process corrections either forward or backward in the process flow. Run-to-run feedback control applications have become critical to achieving production-worthy processes for CD, overlay, and CMP uniformity [2]. Such model-based process control applications require well-founded models for factors that cause drift in the response, tight integration with the factory control system, and an appropriate control algorithm.

The Sematech APC Framework [3] is the foundation for an infrastructure to build such applications. Once this infrastructure is in place, the process model is crucial for achieving success. Modeling needs to take into account the primary factors influencing the response and secondary factors that determine under what sets of conditions the model shall remain valid. For the most part, statisticians and highly knowledgeable engineers have derived these factors through exploratory data analysis. Data mining offers the opportunity to increase the productivity of the tasks leading to the initial model as well as the means to

improve it once in place. For instance, the residual errors from a run-to-run overlay control system could be analyzed to look for patterns that would allow the model to be improved further.

Equipment predictions

On-board tool control systems manage the tool as a self-contained electromechanical system. With the challenge of each succeeding generation of wafer processing technology, these systems have become increasingly complex. This complexity demands sophisticated control systems that monitor and react to problems. During the course of development, extensive data is collected on the failure modes of each subsystem as well as that of the integrated tool.

Data mining has potential application in building models based on historical failure data to detect the precursors of failures. This would permit systems to increase their capabilities for self-repair or graceful shutdown before compromising the process.

Yield management

Yield management covers a diverse set of tasks in both front-end and back-end processing. Yield ramps on new technologies and yield busts on existing technologies demand powerful routines to profile yield loss and trace it to root causes. These types of correlation activities can be tedious and time-consuming because of the time it takes to consolidate data for analysis. Sometimes, it is still a matter of trial and error before a signal is found.

Although various companies have developed automated search routines based on traditional methods like ANOVA and regression, these are of limited use when an interaction exists between two tools or when time is a factor. In such cases, data mining has potential to supplement existing techniques. This is especially true when the data-mining tool is embedded within an exploratory data-analysis environment and can be integrated with existing automation methods.

Macros could be developed to search for certain types of phenomena on a daily or weekly basis and if found, to run data mining to pinpoint the source of the process variation. Similar to an in-line process control application, an example of this might be a monitor that looks at the level of a certain failing bin. When the level exceeds a threshold, the monitor would trigger a data-mining analysis of the last 50 lots to determine the cause of the high counts. This type of system could be extended to trigger data-mining analysis when certain spatial patterns are detected on the wafers.

Fault detection, classification

Fault-detection systems [4] supplement those built within process tools to monitor trace data from tool sensors and look for anomalies. They allow IC manufacturers to customize patterns being monitored and associated action plans.

Data mining is useful in correlating the presence and severity of such faults to downstream quality metrics. This provides an ability to prioritize the response to detected faults. As with excursions detected in defect metrology data, some signals from fault detection on trace data are cosmetic, without a detrimental impact on the wafer; others are indicators of trouble.

Equipment maintenance

Equipment maintenance involves diagnosis and repair of failures, and preventative maintenance (PM). Diagnosis of a tool failure can be a difficult task requiring several pieces of information to be assimilated as clues until a consistent picture can be drawn and the faults isolated. Here again, data mining can make use of historical data to derive expert system-like diagnostic rules to suggest next steps and allow faults to be isolated more efficiently.

PM anticipates the effects of wear and tear to replace parts before they fail and compromise the process. PM frequencies and special PMs are often dependent on processes and recipe changes being performed. For instance, there may be interactions between recipe changes and the process performance immediately after a change. These types of effects are difficult for a semiconductor equipment supplier to anticipate and test. They can have major effects on a manufacturer's success with a tool, however.

Data mining would provide the opportunity to look for such interactions between recipes, as well as establish correlation between the number of lots processed since the last PM and downstream quality measures. Once such models are established they can be used to customize PM procedures and frequencies to improve process performance under given load conditions.

Conclusion and future trends

Although the need for data-mining tools is great, and their value has been proven in other industries, data-mining capabilities needed for semiconductor manufacturing are just beginning to be developed. As with applications in medicine, potential manufacturing applications are hard to picture without involving humans. Current practice necessarily involves an engineer or statistician to guide, interpret, and evaluate the results of data mining.

Today, data mining remains a complementary tool to more traditional statistical and graphical methods for exploratory data analysis and model building. Automated results still need to be reviewed and interpreted prior to direct application in the form of a process change or a new control algorithm for a tool.

Increasingly, however, methods will be developed to reduce these limitations by increasing the breadth of models that can be developed and by reducing the frequency of false signals. Eventually it will be possible to couple data mining to repositories of knowledge as well as data. This will lead to self-optimizing systems of increasing complexity as witnessed by the emerging trend of new offerings for work cell management from the major tool suppliers. This trend will continue and lead to larger centers of automation that will rely on data mining to anticipate and recover from potential problems.

The rate at which this vision is achieved will hinge on the development and adoption of standards for data storage and access between and within tools. Such standards will also need to be extended to the more difficult topic of knowledge storage. Scripting languages will need to be developed or enhanced to manage the complex interactions among tool, cell, and factory control systems and data-mining engines.

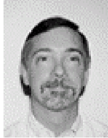
Data-mining technology presents an opportunity to increase significantly the rate at which the volumes of data generated on the manufacturing process can be turned into information. Truly, its time has come!

Acknowledgments

Special thanks to Li-Sue Chen, Bob Anderson, and Jon Buckheit of Yield Dynamics for their help in preparing this article.

References

1. CIM Framework Specification, Vol. 2.0, Sematech document #93061697J-ENG, 1998.
2. Semi/Sematech AEC/APC Symposium X Proceedings, October 11-16, 1998.
3. APCFI proposal and summary, Sematech document #96093181A-ENG, 1996.
4. P.J. O'Sullivan, "Using UPM for Real-Time Multivariate Modeling of Semiconductor Manufacturing Equipment," Semi/Sematech AEC/APC Workshop VII, November 5-8, 1995.

**Richard Kittler** received his PhD in solid-state physics from University of California at Berkeley. He later joined AMD, where he led the development of internal yield management systems. Kittler is currently VP of product development at Yield Dynamics Inc., 2855 Kifer Rd., Santa Clara, CA 95051; ph 408/330-9320, fax 408/330- 9326, e-mail rich@ydyn.com.

Weidong Wang received his PhD in statistics from Stanford University. He is currently a member of the technical staff at Yield Dynamics Inc.; e-mail weidong@ydyn.com.

Data mining in brief

Data-mining methodologies find hidden patterns in large sets of data to help explain the behavior of one or more response variables. Unlike other methods, such as traditional statistics, there is no preconceived model to test; a model is sought using a pre-set range of explanatory variables that may have a variety of different data-types and include "outliers" and missing data. Some variables may be highly correlated and the underlying relationships may be nonlinear and include interactions.

Some data-mining techniques involve the use of traditional statistics, others more exotic techniques such as neural nets, association rules, Bayesian networks, and decision trees.

Neural nets do modeling similar to the learning patterns of the human brain. This model is a network of nodes on input and output layers separated by one or more hidden layers. Each input and output layer node is associated with a variable in the dataset and has connections to all nodes in adjacent layers. Response functions on each hidden layer node determine how a signal from the input direction is propagated to nodes in the output direction. Adjusting the weights on the hidden-layer nodes trains the network to minimize error in the outputs across a set of training data. The trained network can then be used to make predictions for new data — called supervised learning.

Under certain conditions a model with a single hidden layer is equivalent to multiple linear regression. Neural net models are useful when large amounts of data need to be modeled and a physical model is not known well enough to use statistical methods. But with this approach it is difficult to make a physical interpretation of model parameters. Also, predicted outcomes of the model are limited to the scope of the training set used. Because they are not able to discover new relationships in the data, neural nets are not true data mining.

Bayesian networks build a network of relationships using a training dataset where the weights on the links between nodes are constructed from conditional probability distributions. The networks can be built interactively or by searching a database. Though more physical than neural networks because the nodes in the network are measured variables, it is still difficult to extract elements of a physical model from the network or effectively visualize the relationships embodied in it.

Association rules look for patterns of coincidence in data (e.g., how often do failing lots go through various deposition and etch combinations). Its simplest form is the same as a contingency table; advanced forms account for event-occurrence order (e.g., how often do failing lots go through two reworks at first metal and are then etched on a certain etcher). Association rule analysis discovers patterns of behavior, but does not produce a predictive model.

Decision trees are tools for developing hierarchical models of behavior. The tree is built by iteratively evaluating which variable explains the most variability of the response based on the best rule involving the variable. Classes of rules include linear models, binary partitions, and classification groups. For example, the binary partition "if e-test variable $T_{1234} < 1.23 \times 10^{-5}$" would partition data into two groups — true and false. The root node of a tree is a rule that explains the most variability of the response. Child nodes find other variables that explain the most variability of the data subsetted by the first node. The process stops at a point of diminishing returns, similar to automated step-wise regression. Decision trees are useful when relationships are not known and broad categorical classifications or predictions are needed. They are less useful for precise predictions for a continuous variable.

Bayesian networks and decision trees offer the most power in detecting hidden trends in data, in being most physical, and in offering the predictive capability needed to understand patterns of behavior. Both can discover new relationships in data, hence, both are capable of data mining. Of the two, decision trees are easier to interpret physically and to visualize behaviors embodied in their models.

Knowledge-based systems

Knowledge-based systems, such as expert systems, signature analysis, and classification

trees, encapsulate knowledge capable of being derived either whole or in part by data mining.

Expert systems create hierarchical knowledge systems given a set of rules. They guide a user through a decision making or diagnostic process. Many have been built following in-depth interviews with experts. Data mining provides another method of deriving rules for expert systems.

Signature analysis is specifically designed to assimilate clues associated with diagnostic data to fingerprint a process failure. Data mining can be used to discover patterns that associate a given failure with a set of process conditions. Once associations are known, they can be applied to new data through signature analysis to implicate likely process conditions that led to the failure. Classification trees are a special case of data mining when the response variable is categorical. They can be built with or without use of data-mining technology if the knowledge can be obtained through other means.

Requirements for data mining

Data mining requires data availability, efficient access methods, robustness to data problems, efficient algorithms, a high performance application server, and flexibility in delivering results.

The sensitivity of various data-mining methods to data problems must be considered when choosing the product or method for an application. Data problems of concern are those due to gaps and measurements beyond the range of normal observation — so-called outliers. Though data-mining methods have robustness to missing data, results will be improved when gaps can be avoided; but outliers are present in most real-world data sets. Data-mining algorithms that use nonparametric methods (i.e., those that do not rely on normality of the underlying distribution) are less sensitive to outliers. "Cleaning" data prior to analysis can also avoid false signals due to outliers.

Performance of hardware and software in a data-mining system is important when large amounts of data are being mined or response time is critical. Simple data-mining algorithms will by nature be more efficient, but less capable of finding patterns outside a narrow range of behaviors. As the algorithms are made more complex (e.g., evaluate more forms of models at the nodes of a decision tree), efficiency decreases. Once algorithms are chosen, hardware must be sized to deliver the analyses within the required time. For interactive use, it may be sufficient to deliver output to a diagram together with the ability to interact with the diagram to visualize underlying graphical relationships. In other cases, it may be necessary to deliver output to the web. In embedded applications, the output needs to be fed to other systems that can act on the results or message subsystems to perform corrective actions.