Applications for Machine Learning in Semiconductor Manufacturing

Richard Burch, Luke Merrick, Qing Zhu, Tomonori Honda, Jeff David

PDF Solutions

Abstract

This paper presents a machine learning system for comprehensive root cause analysis of low yielding wafers, enabling rapid reaction to correct problems and identify practical actions to reduce the probability and severity of repeated occurrences of the same root cause.

(Keywords: Manufacturing, FDC, Modeling, Machine Learning, Artificial Intelligence)

Introduction

For most non-manufacturing applications, the accuracy of machine learning prediction is more important than a detailed understanding of which variables drive the prediction. This is true because the prediction is sufficient to drive the actions leading to profit improvement through reduced cost or increased revenue. Applying this type of machine learning to semiconductor manufacturing is very challenging, since the extensive actionable labels required are very difficult to obtain.

Nonetheless, the techniques of modern machine learning are incredibly valuable for semiconductor manufacturing. However, the focus is to identify the variables driving accurate prediction and to quantify the impact of each variable on each wafer. This quantification of input variables can be used to quickly identify root cause of yield problems and dramatically accelerate the speed of corrective actions.

Overview

This paper presents a comprehensive, machine learning-based automatic fail mode identification and drilldown system, called ASD (Adaptive Signature Diagnostics), for semiconductor manufacturing. If root cause of the failures can be isolated to specific tools, or event-specific sensor traces, then corrective action can be quick and decisive. ASD can be divided into four distinct capabilities:

1. **Automatic Wafer Classification**: Based on die level bin data, and optionally wafer sort parametric measurements, low yield wafers are classified to separate fail modes and enhance signal detection capability.
2. **Univariate Screening**: For each wafer class, all relevant parameters from all data sources are quickly screened to reduce the hundreds of thousands of potential variables to a few hundred of the most significant.
3. **Multivariate Variable Selection**: For a single wafer class, multivariate variable selection can explain almost all of the predictability with only a few (usually less than 5) parameters.
4. **Signal Review Template**: Although there may only be two or three parameters that can be used to explain most of the wafers for a single wafer class, there are often tens of alternate parameters that are correlated to the selected parameters and may be the actual root cause. Engineering judgement is critical here and ASD uses an interactive, guided analysis drilldown system to help the user identify the true root causes for each wafer class.

Automatic Wafer Classification

During any reasonable time period, there are multiple independent root causes for low yielding wafers. Since they are independent, most low yielding wafers are impacted by a single root cause. Figure 1 gives an example showing a limited section of time with three repeated spatial patterns. The patterns are different bins and/or distinctly different spatial patterns and thus are almost certainly different root causes. Traditionally, wafers with similar patterns were grouped with clustering techniques or simple algorithms, heuristics, or fixed rules. These approaches have proven to be useful but are limited in the patterns they can correctly classify and are difficult to maintain and expand.

Instead of these traditional approaches, ASD uses the same modern machine learning techniques, such as convolutional neural networks, used for image recognition. Figure 2 shows a brief sampling of some of the spatial patterns (both fixed zonal patterns and local clustering patterns) that ASD can identify. Once ASD has completed wafer classification, the huge number of potential variables must be screened to identify probable root causes.

Univariate Variable Screening

2021 Electron Devices Technology and Manufacturing Conference (EDTM)

2021 5th IEEE Electron Devices Technology & Manufacturing Conference (EDTM) | 978-1-7281-8176-9/21/$31.00 ©2021 IEEE | DOI: 10.1109/EDTM50988.2021.9421052

Semiconductor data presents a very challenging problem for machine learning algorithms, since the number of parameters greatly exceeds the observations. In practice, a specific repeated wafer class can be expected to have a few hundred pattern wafers, or even less than 100. Conversely, the number of possible explanatory variables is orders of magnitude greater (Table 1).

Even if FDC variables are excluded, the number of explanatory variables will be 10 to 100 times larger than the number of pattern wafers. Any brute force modeling approach would drastically overfit the available data.

To avoid this, ASD first performs univariate screening using a simplified, cross-validated machine learning pipeline to identify variables with significant prediction power between the variable and the pattern wafers. Figure 3 shows the ROC Curve [1], True Positive Rate (TPR) .vs. False Positive Rate (FPR), with the optimal. For selection purposes, the Matthews Correlation Coefficient [2] or simply use TPR-FPR can be used to estimate the detection power of the variable.

If FDC variables are available, they should be included, but the 100X or more increase in number of variables requires efficient big data architectures with massively parallel processing.

Multivariate Variable Selection

Univariate screening is an important first step to deal with the large number of variables inherent in semiconductor root cause analysis. However, incorporating multiple variables improves the prediction significantly. This often occurs for one of two reasons: the real root cause is an interaction between two or more variables or the spatial signatures of two different root causes were not distinct and there are actually two different root causes.

In the interaction case, the two variables are usually physically related. In the independent root cause case, the variables are often unrelated.

For multivariate variable selection, ASD adapts a standard stepwise regression [3] approach to use a nonlinear machine learning model instead of a simple linear model.

Using the same example as Figure 3, Figure 4 shows the improvement in detection power when the second term is added to the multivariate model. The two steps, Gate Etch and Sidewall Etch, indicate that this is an interaction between two related steps.

Interactive Drilldown and Alternate Selection

The first three steps (Automatic Wafer Classification, Univariate Variable Screening, and Multivariate Variable Screening) are run automatically. The final step is Interactive Drilldown and Alternate Selection using the Signal Review Template. A detailed examination of all of the capabilities of the review template is beyond the scope of this paper. Figure 5 shows the pairwise drilldown page examining the two process steps previously identified and shows the interaction between two bad tools that explain most of the pattern wafers.

Preventative Actions

Although, not the topic of this paper, the identification of specific root causes, especially among metrology, defect, and FDC data can be used to incorporate alarms to identify when the problem reoccurs long before it can be detected at wafer sort.

Conclusion

This paper has presented a modern machine learning approach to identifying explanatory variables responsible for specific failure modes. These explanatory variables are close enough to the true root cause to enable rapid corrective actions and provide information needed for effective prevention.

References

[1] https://en.wikipedia.org/wiki/Receiver_operating_characteristic

[2] https://en.wikipedia.org/wiki/Matthews_correlation_coefficient

[3] https://en.wikipedia.org/wiki/Stepwise_regression

Figure 1: Wafer Yield plotted versus time showing normal wafers (open green circles) and several repeated fail models: Bin A Bottom Edge (Red Circles), Bin B Sparse Donut (Gold Triangles) and Bin 8 Edge (Blue Diamonds).

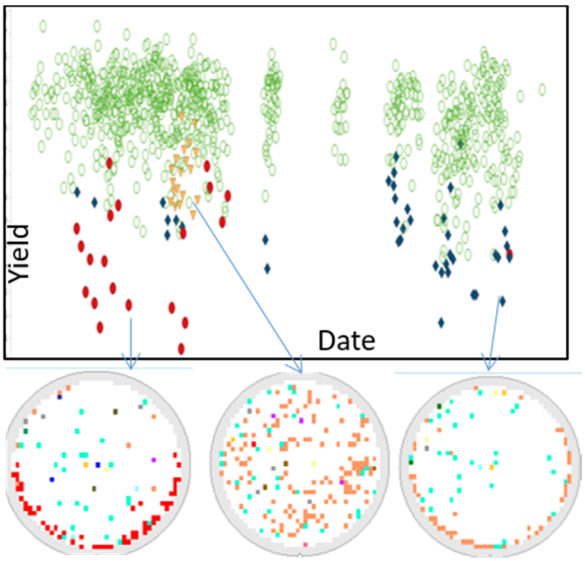

Figure 2: This figure gives a small sampling of spatial patterns (including local clusters) that occur in real data sets.

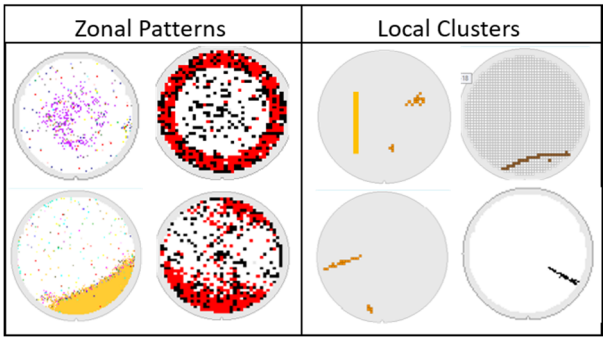

Table 1: Typical data sources used to localize root cause of fail modes identified in wafer classification.

Group	Data Source	Typical Variables
Confirm	Wafer Sort	2,000 Continuous
	PCM	1,000 Continuous
	Metrology	200 Continuous
Action	WEH	1,000 Categorical
	FDC (Sensor Trace Stats)	1,000,000 Continuous
	Defect	1,000 Continuous

Figure 3: ROC cuve and truth table illustrating detection capability of Gate Etch step which has one bad tool explaining over 90% of pattern wafers. Interaction or transient effect is likely since large number of normal wafers also go through that tool.

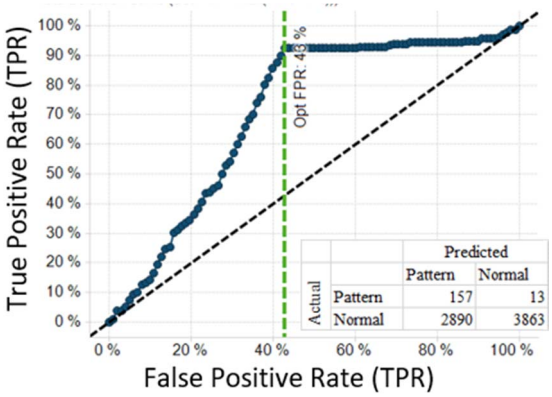

Figure 4: The ROC curve for the cumulative detection capability of Gate Etch and Sidewall Etch still explain nearly 90% of the Pattern wafers but reduce the falsely identified normal wafers by 43%.

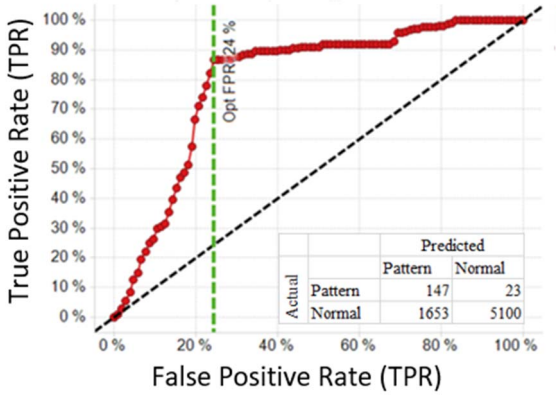

Figure 5: Both Gate Etch and Sidewall Etch have 3 tools with two good tools and one bad tool (GE3 & SE3) with a very large number of pattern wafers. However, when both bad tools process a wafer, the chance of that wafer being bad is nearly 10%. This is strong evidence of an interaction driving the pattern.