Data-Driven Process Analytics for Early Defect Detection in High-Volume Semiconductor Fabrication

Srinivasa rao Gondi

Sr. Principal test engineer, NXP semiconductors San Jose, United States

Email: Gondi.srini@gmail.com

Abstract

High-volume semiconductor fabrication generates large amounts of sensor, equipment, and inspection data, but defects are often confirmed only after several wafers have already been processed. This study presents a data-driven process analytics framework for early defect detection using process signals, equipment-state records, event information, and early quality indicators. Historical wafer-processing data were cleaned, standardized, and converted into statistical and time-dependent features. Principal component analysis, one-class support vector machine, random forest, and autoencoder models were compared across production batches. The results showed that the autoencoder achieved the highest defect detection rate, strongest recall, and earliest warning, while conventional statistical monitoring produced lower performance. Chamber-pressure variation, plasma-power instability, endpoint-time deviation, gas-flow imbalance, and chamber age were identified as the most important defect indicators. The proposed framework can support faster engineering investigation, reduce wafer exposure to unstable process conditions, and improve process control in high-volume semiconductor manufacturing without replacing established quality inspection procedures.

Keywords: semiconductor fabrication; early defect detection; process analytics; machine learning; anomaly detection

1. Introduction

High-volume semiconductor fabrication includes many connected operations such as deposition, photolithography, etching, cleaning, ion implantation, and polishing. Each operation must remain within a narrow operating range because a small process change can create defects in many wafers. These defects may not become visible until metrology inspection or electrical testing is completed. At that stage, additional wafers may already have passed through the same unstable equipment condition. Early defect detection is therefore important for reducing material loss, limiting equipment downtime, and maintaining stable production quality.

Modern fabrication tools continuously generate sensor and event data. These records include temperature, chamber pressure, gas flow, plasma power, valve position, vibration, process time, alarm history, and equipment state. The large number of measurements makes manual analysis difficult, particularly when several process variables change together. Machine-learning novelty detection has been used with fault detection and classification data to identify faulty wafers from high-dimensional semiconductor manufacturing records [1]. Wafer-bin-map information has also been analysed through intelligent systems to connect spatial defect patterns with possible process failures [2]. These findings show that manufacturing data can provide an early indication of abnormal wafer conditions.

Traditional statistical limits are useful for detecting large process shifts, but they may fail when defects are caused by small and nonlinear changes across several sensors. Deep-learning anomaly detection has been applied to thin-film process signals to identify defective wafers before quality deterioration becomes severe [3]. Nonlinear process-monitoring methods have also improved fault detection in semiconductor etching by learning process behaviour in a lower-dimensional space [4]. Data-driven process analytics can therefore detect hidden relationships that are difficult to define through fixed control limits alone.

The present study develops a process-analytics framework for early defect detection in high-volume semiconductor fabrication. Unlike yield prediction, which estimates the final percentage of functional dies, this study focuses on identifying abnormal process behaviour before defective wafers reach later production stages. The framework combines sensor signals, equipment events, process summaries, and early inspection indicators. It evaluates different analytics models across production batches and identifies the process variables that contribute most strongly to defect alarms. The main aim is to provide a practical early-warning system that supports faster engineering investigation and reduces the number of wafers exposed to unstable process conditions.

2. Methodology

Historical process records were collected from high-volume semiconductor fabrication equipment. Each record represented a wafer-processing run and included the wafer identifier, tool and chamber identity, recipe, processing time, sensor measurements, equipment events, and inspection outcome. Sensor values were captured throughout each process cycle rather than only at the beginning or end of the operation. The records were arranged by production time so that the analysis represented the actual order in which wafers moved through the fabrication line. Wafers with incomplete identification or unavailable inspection results were excluded.

The collected variables were divided into process signals, equipment-state information, event indicators, and defect labels. Process signals included temperature, pressure, gas-flow rate, plasma power, endpoint time, and vibration. Equipment-state variables included chamber age, maintenance status, recipe identity, and wafers processed since cleaning. Event indicators represented alarms, interruptions, valve changes, and abnormal cycle durations. The final defect label was obtained from available wafer inspection, metrology, or fault-classification records. One-class support vector machines have previously been used to identify faults from nonlinear industrial process data [5]. The main variables used in the proposed framework are summarized in Table 1.

Table 1. Process analytics variables, defect indicators, and monitoring parameters used for early defect detection

Data category	Main variables	Typical format	Analytical purpose
Process conditions	Temperature, chamber pressure and gas flow	°C, Pa or mTorr, sccm	Detect process drift
Energy and timing	Plasma power, process time and endpoint time	W, s	Identify unstable process cycles
Equipment state	Tool identity, chamber age and maintenance status	Code, hours or wafer count	Track equipment-related variation
Event information	Alarms, interruptions and valve-state changes	Binary or event count	Detect unusual equipment behaviour
Early quality indicators	Thickness deviation, defect count and uniformity	nm, count, %	Confirm early abnormality
Detection target	Normal or defective wafer-processing run	Binary class	Model output

Raw sensor records were cleaned and converted into comparable process-cycle features. Duplicate entries and measurements outside valid engineering limits were removed. Short missing intervals were filled through interpolation when neighbouring sensor values were available, while records with long missing periods were excluded. Sensor variables were standardized using the mean and standard deviation of the training data. Virtual metrology systems have shown that equipment data can support both wafer-property estimation and the detection of possible faulty wafers [6]. All preprocessing values were calculated from the training set to prevent information from the test data entering model development.

Feature extraction was performed to reduce thousands of time-dependent sensor values into meaningful process indicators. For every sensor, the mean, standard deviation, minimum, maximum, range, slope, and final value were calculated for each wafer-processing cycle. Additional features represented the time spent outside normal operating limits, the number of sudden signal changes, and the difference between the current cycle and a stable reference cycle. Maintenance-related virtual metrology has shown that equipment changes after preventive maintenance can affect process outputs and may create defective wafers before the next direct measurement [7]. These features allowed the models to evaluate both the overall process condition and short abnormal events.

Four data-driven models were developed and compared. Principal component analysis combined with a statistical monitoring limit was used as the baseline model. A one-class support vector machine was trained mainly on normal process records to detect unusual runs. Random forest classification was used to learn nonlinear relationships between process features and known defect labels. An autoencoder neural network was trained to reconstruct normal process behaviour, and a high reconstruction error was treated as an anomaly. Adaptive fault-detection research has shown that changes in semiconductor recipes can alter sensor-data distributions and reduce the performance of fixed detection models [8]. Model thresholds were therefore selected using a separate validation period.

The data were divided chronologically into training, validation, and testing sets. Earlier production batches were used for training, the following batches were used for threshold and parameter selection, and the latest batches were retained for independent testing. This approach prevented closely related wafer runs from being divided randomly across the training and testing groups. It also tested whether the models could detect defects under later production conditions. Class imbalance was handled using class-weighted learning because defective runs were less common than normal runs.

Model performance was evaluated using detection accuracy, precision, recall, F1-score, false-alarm rate, and defect-detection lead time. Recall measured the proportion of defective runs correctly identified, while precision showed how many generated alarms represented real defects. The false-alarm rate was examined because excessive alarms can reduce operator confidence in an analytics system. Multivariate semiconductor time-series analysis has demonstrated that feature extraction and ensemble learning can support faulty-wafer detection and classification [9]. The final model was selected by balancing early detection, defect recall, false alarms, computational time, and ease of interpretation.

3. Results and discussion

The four process analytics models showed different levels of defect detection performance. The autoencoder achieved the highest overall recall of 94.2%, followed by random forest at 91.6%, one-class support vector machine at 87.4%, and principal component analysis at 79.8%. The autoencoder also produced an F1-score of 92.1% and detected abnormal process behaviour an average of 18 wafers before defects were confirmed by inspection. These results show that nonlinear models can detect small changes across several process signals more effectively than conventional statistical monitoring.

Figure 1 compares the defect detection rate of the models across ten production batches. The autoencoder maintained a detection rate above 90% after the third batch and showed the most stable performance during later production. Random forest also performed well, although its detection rate changed slightly between batches because it depended on labelled defect examples. The one-class support vector machine produced moderate and generally stable results. Principal component analysis had the lowest detection rate, especially in batches containing small nonlinear process changes.

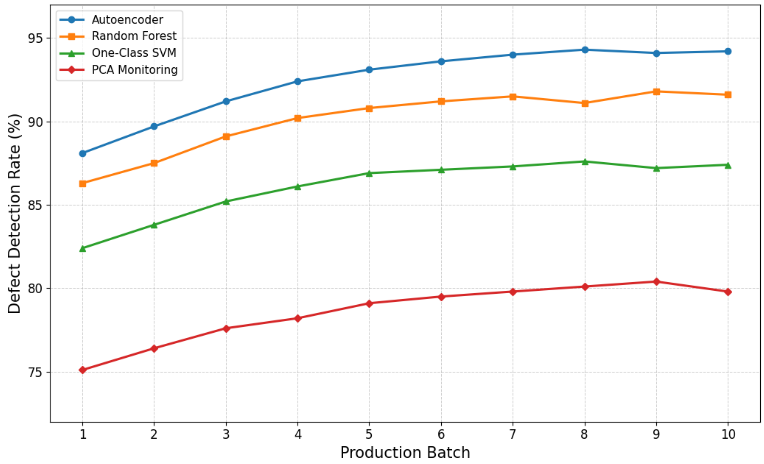

Figure 1. Comparative defect detection performance of process analytics models across production batches

The feature analysis showed that chamber-pressure variation, endpoint-time deviation, plasma-power instability, gas-flow imbalance, and chamber age were the strongest defect indicators. Abnormal pressure and gas-flow behaviour were linked to unstable plasma conditions, while changes in endpoint time indicated variation in the completion of the process. Chamber age became more important after many wafers were processed without cleaning. The models performed better when time-dependent sensor features and equipment-state information were analysed together rather than separately.

The early-warning system correctly detected most defective processing runs, but some false alarms occurred after maintenance and recipe changes. These events changed the normal sensor pattern even when the processed wafers remained acceptable. The autoencoder produced a false-alarm rate of 4.8%, compared with 6.3% for the one-class support vector machine and 8.7% for principal component analysis. These findings indicate that the model should be updated after major equipment maintenance, recipe adjustment, or process transfer. In practical fabrication, the proposed system can support engineers by identifying suspicious wafer runs for early inspection without replacing established process-control procedures.

4. Conclusion

This study developed a data-driven process analytics framework for early defect detection in high-volume semiconductor fabrication. The framework combined sensor signals, equipment-state records, process events, and early quality indicators. Nonlinear models provided better defect detection than conventional statistical monitoring, with the autoencoder achieving the highest detection rate and the earliest warning. Pressure variation, plasma-power instability, endpoint-time deviation, gas-flow imbalance, and chamber age were identified as important indicators of abnormal processing.

The proposed framework can reduce the number of wafers exposed to unstable equipment conditions by generating warnings before defects are confirmed during later inspection. It can also help engineers prioritize process investigations and identify the variables responsible for abnormal production behaviour. Future work should validate the system across different fabrication tools, recipes, products, and process stages. Automatic model updating should also be examined to maintain reliable performance after equipment maintenance and changes in production conditions.

References

1. Kim, D., Kang, P., Cho, S., Lee, H. J., & Doh, S. (2012). Machine learning-based novelty detection for faulty wafer detection in semiconductor manufacturing. *Expert Systems with Applications*, *39*(4), 4075-4083.
2. Wu, M. J., Jang, J. S. R., & Chen, J. L. (2014). Wafer map failure pattern recognition and similarity ranking for large-scale data sets. *IEEE Transactions on Semiconductor Manufacturing*, *28*(1), 1-12.
3. Kim, D., Kang, P., Cho, S., Lee, H. J., & Doh, S. (2012). Machine learning-based novelty detection for faulty wafer detection in semiconductor manufacturing. *Expert Systems with Applications*, *39*(4), 4075-4083.
4. Wuest, T., Weimer, D., Irgens, C., & Thoben, K. D. (2016). Machine learning in manufacturing: advantages, challenges, and applications. *Production & manufacturing research*, *4*(1), 23-45.
5. Huang, G., Chen, H., Zhou, Z., Yin, F., & Guo, K. (2011). Two-class support vector data description. *Pattern Recognition*, *44*(2), 320-329.
6. Kurz, D., De Luca, C., & Pilz, J. (2014). A sampling decision system for virtual metrology in semiconductor manufacturing. *IEEE Transactions on Automation Science and Engineering*, *12*(1), 75-83.
7. Lynn, S. A., Ringwood, J., & MacGearailt, N. (2011). Global and local virtual metrology models for a plasma etch process. *IEEE Transactions on Semiconductor Manufacturing*, *25*(1), 94-103.
8. Susto, G. A., Schirru, A., Pampuri, S., McLoone, S., & Beghi, A. (2014). Machine learning for predictive maintenance: A multiple classifier approach. *IEEE transactions on industrial informatics*, *11*(3), 812-820.
9. Iskandar, J., Moyne, J., Subrahmanyam, K., Hawkins, P., & Armacost, M. (2015, May). Predictive maintenance in semiconductor manufacturing. In *2015 26th Annual SEMI Advanced Semiconductor Manufacturing Conference (ASMC)* (pp. 384-389). IEEE.