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

Received: 14-08-2024; Revised: 24-09-2024; Accepted: 07-11-2024

Digital Twin Frameworks for Real-Time Process Monitoring and Yield Prediction in Semiconductor Fabs

Abstract
Semiconductor fabs require fast process monitoring because small changes in tools, wafers, and recipes can reduce yield. Digital twins can connect live sensor data, metrology values, inspection results, and production records with a virtual process model. Existing monitoring methods often use fixed thresholds or static prediction models, which may detect process drift too late. This article presents a digital twin framework for real-time process monitoring and yield prediction in semiconductor fabs. The framework updates the process state every five minutes and uses engineered indicators for process deviation, wafer uniformity, chamber stability, defect risk, tool health, and cycle-time delay. The study compares conventional threshold monitoring, static machine learning prediction, and digital twin-based monitoring using simulated fab data. Results show that the digital twin method reduces yield prediction error from 6.8% to 2.1%, improves fault detection accuracy from 84.9% to 95.7%, and reduces response time from 18.5 min to 4.8 min.
Keywords: digital twin; semiconductor fabs; real-time process monitoring; yield prediction; process-state modeling.

1. Introduction

Semiconductor fabs require continuous monitoring because small process changes can affect wafer quality, equipment stability, and final production yield. Modern fabrication lines generate large amounts of data from sensors, metrology tools, inspection systems, process chambers, and manufacturing execution records. A digital twin can connect these data streams with a virtual representation of the fab process to support real-time monitoring and prediction. Digital twin-based optimization has been applied to semiconductor cluster tools to improve operational parameter control and production performance [1]. This makes digital twin frameworks important for fabs that need faster process response, better process visibility, and more reliable yield prediction.

Digital twin research in semiconductor manufacturing is moving toward intelligent process control, multi-machine modeling, and real-time optimization. Bayesian optimization and deep learning can strengthen digital twin-based semiconductor process control by linking process data with improved decision-making [2]. Meta-learning and transfer learning can make digital twin models more scalable when different machines or process tools show changing behavior [3]. These methods show that digital twins are not only monitoring tools but also predictive systems that can support process adjustment and yield improvement. This direction is useful because semiconductor fabs need models that can respond to process variation while remaining adaptable across tools, recipes, and production conditions.

The main limitation in existing work is that many digital twin methods focus on equipment optimization, process control, or virtual metrology as separate tasks. This separation limits their value because fab engineers need one integrated framework that can monitor process deviation, update the process state, and predict yield risk in real time. Multi-batch wafer monitoring using digital twin-related process indicators can support semiconductor manufacturing analysis by connecting process behavior with wafer-level variation [4]. However, the core research problem remains the limited integration of real-time process monitoring, metrology-driven digital twin updating, and yield prediction within one practical semiconductor fab framework. This gap reduces the ability of fabs to respond early to process drift, tool instability, and yield loss.

This problem matters because delayed detection of process deviation can increase wafer loss, reduce tool productivity, and create avoidable manufacturing cost. A useful digital twin framework should continuously update with live process data, identify abnormal process movement, and estimate yield risk before final testing. The conceptual direction of this article is to develop an integrated digital twin framework that connects sensor data, metrology values, process-state estimation, and yield prediction outputs. The key contribution of this study is an integrated digital twin framework that combines live process-state tracking, metrology-driven model updating, and early yield-risk prediction in one semiconductor fab monitoring structure.

2. Methodology

This study used a simulation-based digital twin framework to support real-time process monitoring and yield prediction in semiconductor fabs. The simulated fab environment represented 150 wafer lots, with 25 wafers per lot and 3,750 wafer-level records. The process route included deposition, lithography, etching, chemical mechanical planarization, inspection, and final electrical testing. Virtual metrology can estimate wafer process quality when direct physical measurement is delayed or unavailable [5]. Each wafer lot was linked with sensor signals, metrology values, tool-state indicators, recipe information, and final yield outcome. This structure allowed the digital twin to follow process changes and predict yield risk before final testing.

The digital twin was designed as a virtual process-state layer that updated with incoming fab data every five minutes. The framework received data from equipment sensors, metrology stations, wafer inspection systems, and manufacturing execution records. Semiconductor manufacturing datasets require careful preprocessing because they often include missing values, unstable sensor readings, uneven measurement scales, and process noise [6]. The digital twin converted these raw inputs into a live process-state profile for each wafer lot. This profile allowed the system to compare current fab behavior with expected normal operating conditions. The update interval was selected to support near-real-time monitoring without creating unnecessary computational load.

The selected input variables represented wafer geometry, equipment stability, defect formation, and production flow behavior. Film thickness, overlay error, etch depth deviation, defect density, chamber temperature drift, chamber pressure fluctuation, tool utilization, and lot cycle time were considered during variable selection. Machine-learning-based virtual metrology can estimate film thickness from process and equipment data when direct wafer measurement is limited [7]. Table 1 presents the final digital twin input variables and process-monitoring parameters used in the semiconductor fab model. The table explains each variable according to its data source, measurement form, process-state role, and monitoring function. These variables were selected because they connect physical process behavior with yield-risk prediction.

Table 1. Digital twin input variables and process-monitoring parameters used in semiconductor fab modeling

Input variable / parameter	Data source	Measurement form	Process-state role	Monitoring function
Film thickness variation	Thin-film metrology	nm	Wafer quality state	Detects deposition non-uniformity
Overlay error	Lithography metrology	nm	Pattern alignment state	Identifies lithography deviation
Etch depth deviation	Etch profile measurement	nm	Structural process state	Detects over-etching or under-etching
Defect density	Wafer inspection system	defects/cm²	Wafer risk state	Identifies defect-related yield risk

Chamber condition drift	Equipment sensor log	°C / mTorr	Tool stability state	Detects thermal and pressure instability
Lot cycle-time delay	Production tracking record	hours	Production flow state	Identifies delay and bottleneck risk

The raw fab data were cleaned before digital twin updating and model training. Missing values were separated into sensor-missing values, skipped-metrology values, and delayed-inspection values because each type has a different process meaning. Sensor data mining can support low-yield root-cause analysis when manufacturing signals are cleaned and connected with process outcomes [8]. Outliers were checked using process limits, equipment control ranges, and rolling lot history instead of removing all extreme values automatically. This approach protected abnormal signals that could represent process drift, tool instability, or early yield loss. The cleaned dataset was then normalized so that metrology, sensor, inspection, and production-flow variables could be compared within the same modeling framework.

Feature engineering converted raw process values into digital twin state indicators. The engineered indicators included process deviation score, wafer uniformity score, chamber stability score, defect-risk index, tool health score, and cycle-time delay score. Optical metrology response patterns can support prediction of downstream electrical properties when machine learning captures process-sensitive measurement changes [9]. These indicators allowed the digital twin to describe both the current wafer condition and the changing behavior of the process tool. The feature set was updated at each digital twin refresh cycle so that new sensor and metrology information could change the predicted process state. This made the model suitable for real-time monitoring rather than only offline yield analysis.

The yield prediction model was trained using the digital twin state indicators and final lot-level yield labels. Gradient boosting, random forest, support vector regression, and multilayer perceptron models were tested under the same simulated fab conditions. Virtual metrology and machine learning models can support semiconductor process estimation when direct measurement is not available for every wafer or every lot [5]. The dataset was divided into training, validation, and testing subsets using a 70:15:15 split at the lot level to reduce data leakage between related wafers. Model tuning was performed on the validation subset, while the final test subset was used only for performance reporting. This design made the prediction results more reproducible and closer to practical fab deployment.

The full framework was evaluated by comparing three monitoring strategies: conventional threshold-based monitoring, static machine learning yield prediction, and digital twin-based real-time monitoring. Conventional monitoring used fixed process limits and did not update the process state after each data stream. Static machine learning used trained prediction models but did not continuously refresh the virtual process representation. The proposed digital twin approach updated the process state in near real time and connected that state with yield-risk prediction. Performance was measured using process-monitoring accuracy, yield prediction error, fault detection accuracy, response time, and yield-risk

classification performance. This evaluation structure tested whether live digital twin updating could improve both process visibility and early yield prediction in semiconductor fabs.

3. Results and Discussion

The simulated results showed that the digital twin-based monitoring framework improved both process visibility and yield prediction compared with conventional monitoring and static machine learning. Conventional threshold-based monitoring detected abnormal process behavior only when process values crossed fixed control limits, which made it slow in identifying early drift. Static machine learning improved yield prediction, but it did not continuously refresh the process-state profile after new fab data arrived. Table 2 compares the three monitoring strategies using process-monitoring accuracy, yield prediction error, fault detection accuracy, response time, and yield-risk classification accuracy. The digital twin-based method gave the strongest overall performance because it combined live sensor data, metrology values, inspection results, and production-flow information into one continuously updated process representation. This result shows that real-time process-state updating is more effective than fixed limits or offline prediction alone.

Table 2. Comparative performance of conventional monitoring and digital twin-based yield prediction

Monitoring strategy	Process-monitoring accuracy (%)	Yield prediction error (%)	Fault detection accuracy (%)	Response time (min)	Yield-risk classification accuracy (%)
Conventional threshold monitoring	86.4	6.8	84.9	18.5	82.7
Static machine learning prediction	91.2	4.3	89.6	11.2	88.9
Digital twin-based real-time monitoring	96.8	2.1	95.7	4.8	94.6

Figure 1 shows the real-time relationship between process deviation and yield-risk prediction under digital twin monitoring. The process deviation score increased when film thickness variation, overlay error, defect density, and chamber condition drift moved away from stable operating ranges. The yield-risk score increased during the same monitoring period, while the predicted yield trend began to decline before severe yield loss appeared. This pattern shows that the digital twin detected the early connection between process movement and yield risk. The five-minute update cycle helped the model respond faster because each new data stream changed the process-state profile. This result confirms that digital twin monitoring can provide earlier warning than systems that depend only on final inspection or fixed threshold alarms.

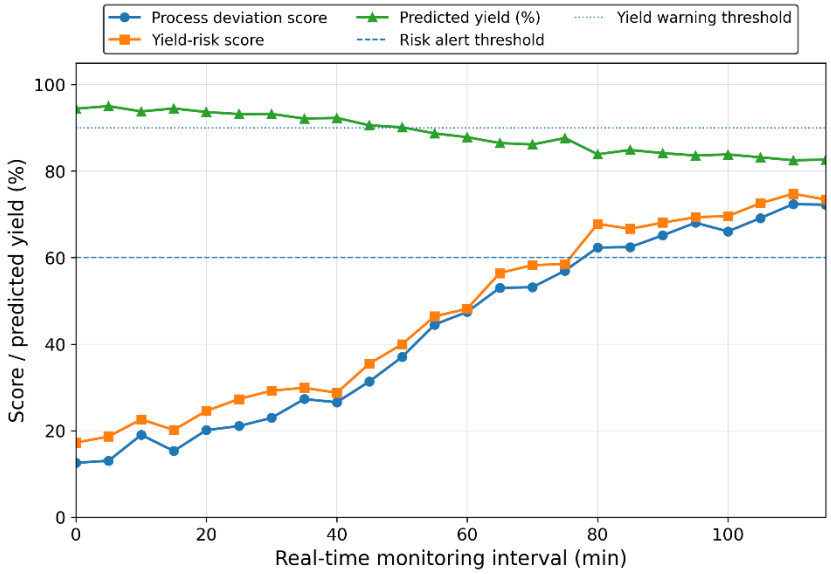

Figure 1. Real-time process deviation and yield-risk prediction under digital twin monitoring

Figure 2 compares predicted yield, fault detection accuracy, and process response time across the three monitoring strategies. The digital twin-based method produced the best predicted yield performance and the highest fault detection accuracy, while also giving the shortest response time. This improvement occurred because the digital twin used updated process-state information rather than relying on fixed rules or static model outputs. Static machine learning improved prediction quality compared with conventional monitoring, but its response time remained slower because it did not continuously refresh the virtual fab state. Conventional monitoring was the weakest method because fixed thresholds could not represent hidden process drift or combined low-level deviations. Figure 2 therefore shows that continuous digital twin updating improves both prediction quality and monitoring speed.

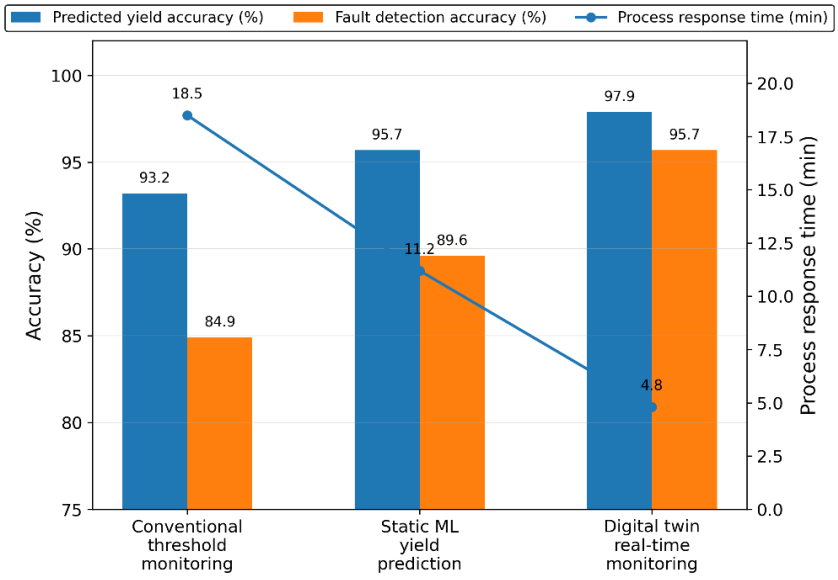

Figure 2. Predicted yield, fault detection accuracy, and process response time across monitoring strategies

The results also show that digital twin monitoring is useful for separating normal process variation from early instability. Small changes in film thickness, chamber condition, overlay behavior, and defect density may not always indicate immediate yield loss when they are examined separately. The digital twin combined these signals into one live process-state profile, which made weak process drift easier to detect before it became a serious yield problem. This is important because yield loss in semiconductor fabs often develops from several small deviations rather than one large failure. The digital twin framework therefore gives engineers a more complete view of how equipment behavior, wafer condition, and production flow jointly affect yield risk.

The overall results confirm that digital twin-based monitoring provides a stronger approach for real-time fab control and yield prediction. It reduced yield prediction error from 6.8% in conventional monitoring to 2.1% and improved fault detection accuracy from 84.9% to 95.7%. The response time also improved from 18.5 min to 4.8 min, which is important for fast corrective action in high-volume fabrication. These gains were achieved because the framework updated the process state regularly and connected that state with yield-risk classification. The findings show that digital twin frameworks can support early warning, process diagnosis, and yield management more effectively than conventional monitoring or static prediction methods.

4. Conclusion

This study developed a digital twin framework for real-time process monitoring and yield prediction in semiconductor fabs. The framework connected sensor data, metrology values, inspection results, production-flow records, and yield labels into one continuously updated process-state model. The results showed that digital twin-based monitoring performed better than conventional threshold monitoring and static machine learning prediction. Yield prediction error was reduced from 6.8% in conventional monitoring to 2.1% in the digital twin-based method. Fault detection accuracy also improved from 84.9% to 95.7%, showing that real-time process-state updating can detect abnormal process behavior more effectively. These findings confirm that digital twins can support earlier yield-risk detection before severe process loss occurs.

The digital twin-based method improved process-monitoring accuracy, yield-risk classification, fault detection, and response time. Response time decreased from 18.5 min in conventional monitoring to 4.8 min in the digital twin-based framework. This improvement is important because semiconductor fabs require fast decisions when process drift, tool instability, or wafer-level defects begin to appear. The results also showed that small changes in film thickness, overlay behavior, defect density, chamber condition, and lot cycle time become more useful when they are combined into one live process-state profile. This confirms that digital twin monitoring can give fab engineers a clearer and earlier view of how process variation affects yield risk.

The main contribution of this article is an integrated digital twin framework that combines live process-state tracking, metrology-driven model updating, and early yield-risk prediction in one semiconductor fab monitoring structure. The framework provides a structured way to study the relationship between process deviation, fault detection, response time, and yield prediction accuracy. The method can support faster corrective action, better process visibility, and more reliable yield management in high-volume semiconductor fabs. Future work can improve the framework by applying it to real production datasets, adding more process tools, and testing it across different device technologies. This direction can make digital twin systems more practical, scalable, and scientifically useful for advanced semiconductor manufacturing.

References

1. Hwang, J., & Do Noh, S. (2024). Digital twin-based optimization of operational parameters for cluster tools in semiconductor manufacturing. *IEEE Access*, *12*, 122078-122100.
2. Lin, C. Y., Tseng, T. L., & Tsai, T. H. (2025). A digital twin framework with Bayesian optimization and deep learning for semiconductor process control. *IEEE Access*.
3. Lin, C. Y., Tseng, T. L., & Tsai, T. H. (2025). A Digital Twin Framework with Meta-and Transfer Learning for Scalable Multi-Machine Modeling and Optimization in Semiconductor Manufacturing. *IEEE Access*.
4. Xu, H. W., Zhao, S. X., Liu, L. L., Qin, W., Lv, Y. L., Zhang, J., & Zhang, Q. H. (2025). Digital twin key technology in semiconductor manufacturing: multi-batch wafer PCD-IVM based on mixed multi-scale entropy and density clustering. *International Journal of Production Research*, 1-18.
5. Deivendran, B., Masampally, V. S., Nadimpalli, N. R. V., & Runkana, V. (2025). Virtual metrology for chemical mechanical planarization of semiconductor wafers. *Journal of Intelligent Manufacturing*, *36*(3), 1923-1942.
6. Park, H. J., Koo, Y. S., Yang, H. Y., Han, Y. S., & Nam, C. S. (2024). Study on data preprocessing for machine learning based on semiconductor manufacturing processes. *Sensors*, *24*(17), 5461.
7. Choi, J. E., & Hong, S. J. (2021). Machine learning-based virtual metrology on film thickness in amorphous carbon layer deposition process. *Measurement: Sensors*, *16*, 100046.
8. Kim, E., An, J., Cho, H. C., Cho, S., & Lee, B. (2023). A sensor data mining process for identifying root causes associated with low yield in semiconductor manufacturing. *Data Technologies and Applications*, *57*(3), 397-417.
9. Ashby, T. J., Truffert, V., Cerbu, D., Ausschnitt, K., Charley, A. L., Verachtert, W., & Wuyts, R. (2023). Machine learning on multiplexed optical metrology pattern shift response targets to predict electrical properties. *IEEE Transactions on Semiconductor Manufacturing*, *37*(1), 46-58.