An approximate analytical method for the performance evaluation of semiconductor front-end fabrication integrating photolithography inspection strategies

Maria Chiara MAGNANINI
Department of Mechanical Engineering, Politecnico di Milano, Milano, Italy, email: mariachiara.magnanini@polimi.it

Riccardo POMI
Department of Mechanical Engineering, Politecnico di Milano, Milan, Italy, riccardo.pomi@polimi.it

Dragan DJURDJANOVIC
Cockrell School of Engineering, University of Texas at Austin, Austin TX, USA, dragan.djurdjanovic@utaustin.edu

Tullio TOLIO
Department of Mechanical Engineering, Politecnico di Milano, Milano, Italy, email: tullio.tolio@polimi.it

In semiconductor manufacturing, lithography represents the core process of frontend fabrication as the quality outcome in terms of overlay errors depends entirely on it. Hence, particular attention is devoted to the inspection of each wafer layer, having 100% measurements of markers distributed across a wafer with subsequent long inspection times. At the same time, process control is based on each layer's overall measurements, discouraging companies from improving productivity by reducing inspection time. As a consequence, in this context product, process and system are extremely inter-related. Recent developments in joint product-process modelling show that robust model-based control coupled with optimal down-selection of measurement markers enables improved process control without increasing the defects. However, when considering the system level, new dynamics should be accounted for in order to take decisions about production system configuration and operations. This paper proposes a novel analytical model for the evaluation of quality and productivity performance in manufacturing systems characterized by propagation of quality errors, process adaptation and alternative inspection policies. The proposed model is general but particularly useful for the semiconductor sector. Application of this method to an industrial-scale semiconductor manufacturing system shows that when product-process-system are considered together, global optimal solutions can be achieved.

Keywords: Manufacturing systems; Decomposition; Quality control; Semiconductors.

1 Motivation and objectives

Performance evaluation is an essential step towards improvement of manufacturing systems and involves estimating the effect of various factors on performance indicators such as throughput, work in process level, efficiency, among others. Quality is considered one of the most critical factors that can influence system behavior and performance. Therefore, efforts for the integration in performance evaluation models start with including production resources that may produce parts with imperfect quality. Contributions from the community include the integration of SPC methods in analytical models, the evaluation and improvement of systems including quality-quantity coupled operations, quality deterioration, caused by machine stoppages and long waiting times, scrapping policies in response to machine failures, considering buffer-less systems with both presence and absence of deterioration memory, or scrapping policies resulting from stoppages of production systems, and differentiating between long lasting failures, causing deterioration, and rapid failures. The semiconductor manufacturing system is recognized as a highly intricate production process. The initial phase, known as wafer manufacturing or the front-end, incurs significant costs. During this phase, circuits are methodically layered onto the wafer using a series of sequential procedures. Numerous processing steps are involved in this phase. Consequently, the dynamics, performance, and characteristics of both the process and the end product are determined by an extensive range of factors. The propagation of multi-stage dynamics has a clear impact on the responsiveness of the quality strategy. At product level, each stage operates a transformation on the product

and may add product deviations, in form of overlays. Particular attention is devoted to the inspection of each wafer layer, having 100% measurements of markers distributed across a wafer with subsequent long inspection times. Despite technological advancements and research & development efforts in this direction, inspection time could not be reduced yet, and still represents a limit for the system productivity (Chien 2020). To ensure the required quality, process control based on models such as stream of variation (SoV) was introduced and studied (Graff 2023). Errors propagation is analytically described and incorporated in process control models robust to inaccuracies between process parameters and error generation. As a consequence, here product, process and system are extremely inter-related, and a clear understanding of the interaction between quality and production is envied to improve the capacity planning of fabs (Ghasemi 2020). This work integrates state-of-the-art quality models for inspection optimization in lithography combined with model-based process control into an approximate analytical model. Given the unique characteristics, this novel approach is particularly suitable for semiconductor fabrication, but it has general validity. The novel contributions of this work can be summarized as follows:

1. State-based Markovian synthetic representation of semiconductor fabrication stages in lithography threads embedding selective inspection strategies and model-based process control;
2. System-level stochastic model with close-to-reality assumptions, as split and scrap, integrating propagation of product quality and inspection errors along stages;
3. Joint optimization of productivity and quality at system-level in semiconductor fabrication for the downsizing of measurement selection.

2 Methodology and results

The system includes K stages, where each stage S_k is composed of one photolithography machine and one inspection station. Stages are decoupled by buffers B_k characterized by buffer capacity N_k. Both photolithography machines and inspection stations are fully reliable, and no failures occur in either stage. If the downstream carrier has enough capacity, no blocking phenomenon affects the lithography machines. Similarly, starvation may occur if the lithography thread is not adequately balanced. Whenever an inspection station performs a conformity check on the wafer and finds that the patterned layer is defective, a parameter tuning of the photolithography machine is immediately performed, as usual in run-to-run control. At the same time, the defective wafer is promptly unloaded from the inspection machine and rejected from the line. This preserves downstream station capacity from being wasted on processing defective wafers. Given the high automation level and the type of process, both photolithography and inspection have constant cycle time. The target performance measures include the average total throughput th_K at the end of the lithography thread, the average total throughput of good wafers th_K^G and the average throughput of expected defective wafer th_K^{NG}. If the inspection station inspects 100% of the markers on the patterned layer of the wafer, the throughput of expected defective wafers th_K^{NG} includes only the wafers for which the last patterned layer is out of specification. On the other hand, if the inspection station inspects a reduced number of markers on the patterned layer, the throughput of expected defective wafers th_K^{NG} includes also those wafers with non-compliant layers that have not been detected by inspection stations $k = 1, \dots, K-1$. From the perspective of the production flow, there is no conservation of flow, as the production flow decreases along the lithography thread, because at each stage in-process scrap may occur, when non-compliant patterned layers are detected. The product and process model utilizes a robust run-to-run control that considers not only overlay errors but also stack-up overlay errors. These errors are described by the summation of the overlay of non-adjacent layers, using Zernike polynomial-based models (Zhang 2022). The product-process model selects the best combination of a given percentage of available measurement points $m, m \in [0,1]$, known as markers, for the robust control of lithography errors. This ensures that the measure of overlay errors at all the candidate measurement points is minimized. However, it is important to note that the down-selection of markers for the wafer inspection may not detect bad wafers, allowing them to continue along the manufacturing line. The definition of the system-level Markovian model extends the work introduced by Magnanini (2023). Hence, the state space for the stage S_k at system level is $\Omega_k = \{G, BD, BND, S_G, S_{BD}, S_{BND}, B_G, B_{BND}, NQ_1, NQ_j, \dots, NQ_{k-1}\}$, where:

- Local states: state G represents the production condition of good wafers at the lithography machine that afterwards are correctly identified as compliant by the inspection station; state BD represents the production condition of bad wafers at the lithography machine that afterwards are correctly identified as not-compliant by the inspection station; state BND represents the production condition of bad wafers

at the lithography machine that afterwards are not detected as not-compliant by the inspection station, due to the down-selection of markers m.

- Remote states $\{S_G, S_{BD}, S_{BND}\}$ in which the stage is upstream limited from the perspective of production flow, i.e. the considered stage is starved or slowed down.
- Remote states $\{B_G, B_{BND}\}$ in which the stage is downstream limited from the perspective of production flow, i.e. the considered stage is blocked or slowed down.
- Remote states $\{NQ_1, NQ_j, \dots, NQ_{k-1}\}$ in which the stage is processing defective layers from previous stages that were not detected by the down-selection of markers in upstream inspection stations.

The decomposition method obtains the probability density function of each Building Block, then the two-level decomposition approach obtains the transition rate matrix for each Integrated Machine by lumping of the Building Block state-space into the Integrated Machine state-space. In this way, the number of states do not explode. Results in Fig. 1 shows that reducing the number of inspected markers clearly improves the productivity with respect to the baseline case (100%). On the other hand, when the yield is analyzed, it is possible to notice that using a good uncertainty model of the propagation of deviations coupled with the process control guarantees the identification of the optimal set of measurement markers, not only to the optimal number of measurement markers to be inspected. In particular, the highest productivity in terms of good wafers is obtained when the measurement reduction is set according to the proposed novel product-process-system model, as it exploits the knowledge from the product and process level, together with the knowledge about the dynamics at system level.

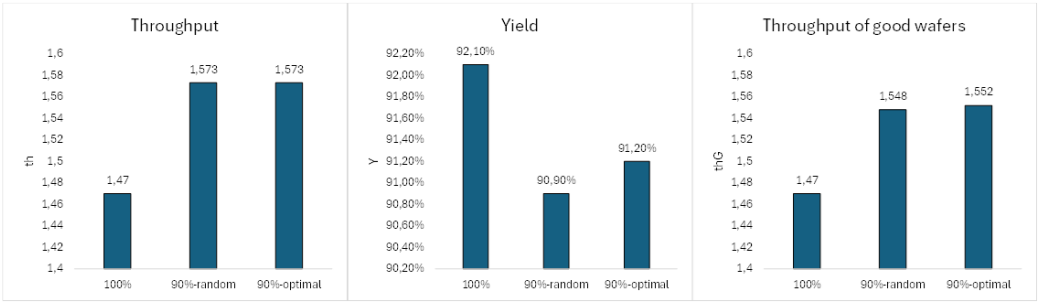

Figure 1. Comparison of performance measures with alternative quality control strategies.

References

Arinez, J., Meerkov, S. M., Zhang, L. 2010. Quality/Quantity Improvement in an Automotive Paint Shop: A Case Study. IEEE Transactions on Automation Science and Engineering 7(4), 755-761.

Chien, C. F., Chen, Y. H., Lo, M. F. 2020. Advanced quality control (AQC) of silicon wafer specifications for yield enhancement for smart manufacturing. IEEE Transactions on Sem. Manuf,, 33(4), 569-577.

Colledani, M., Tolio, T. 2011. Integrated analysis of quality and production logistics performance in manufacturing lines. International Journal of Production Research 49(2), 485-518.

Ghasemi, A., Azzouz, R., Laipple, G., Kabak, K. E., Heavey, C. 2020. Optimizing capacity allocation in semiconductor manufacturing photolithography area–Case study: Robert Bosch. Journal of Manufacturing Systems, 54, 123-137.

Graff, N., Hanasusanto, G. A., Djurdjanović, D. 2023. Robust control of maximum photolithography overlay error in a pattern layer. CIRP Annals, 72(1), 429-432.

Kim, J., Gershwin S. B. 2005. Integrated quality and quantity modeling of a production line. OR Spectrum 27, 287-314.

Liberopoulos, G., Kozanidis, G., Tsarouhas, P. 2007. Performance Evaluation of an Automatic Transfer Line with WIP Scrapping During Long Failures. Manufacturing & Service Operations Management 9(1), 62-83.

Magnanini, M. C., Tolio, T. 2023. A Markovian model of asynchronous multi-stage manufacturing lines fabricating discrete parts. Journal of Manufacturing Systems, 68, 325-337.

Tempelmeier, H., Bürger, M. 2001. Performance evaluation of unbalanced flow lines with general distributed processing times, failures and imperfect production. IIE Transactions 33(4), 293-302.

Zhang, H., Feng, T., Djurdjanovic, D. 2022. Dynamic down-selection of measurement markers for optimized robust control of overlay errors in photolithography processes. IEEE Transactions on Semic. Manuf., 35(2), 241-255.