An AI-Enabled Framework for Smart Semiconductor Manufacturing

(Article begins on next page)

An AI-Enabled Framework for Smart Semiconductor Manufacturing

Khaled Sidahmed Sidahmed Alamin*, Davide Appello†, Alessandro Beghi‡, Nicola Dall'Ora§, Fabio Depaoli*, Santa Di Cataldo*, Franco Fummi§, Sebastiano Gaiardelli§, Michele Lora§, Enrico Macii*, Alessio Mascolini*, Daniele Pagano¶, Francesco Ponzio*, Gian Antonio Susto‡, Sara Vinco*

*: Politecnico di Torino, Italy, {name.surname}@polito.it †: Technoprobe Spa, Italy, {name.surname}@technoprobe.com
‡: University of Padova, Italy, {name.surname}@dei.unipd.it ¶: STMicroelectronics, Italy, {name.surname}@st.com
§: University of Verona, Italy, {name.surname}@univr.it

***Abstract*—With the rise of Machine Learning (ML) and Artificial Intelligence (AI), the semiconductor industry is undergoing a revolution in how it approaches manufacturing. The SMART-IC project (*DATE'24 MPP category: initial stage*) works in this direction, by proposing an AI-enabled framework to support the smart monitoring and optimization of the semiconductor manufacturing process. An AI-powered engine examines sensor data recording physical parameters during production (like gas flow, temperature, voltage, etc.) as well as test data, with different goals: (1) the identification of anomalies in the production chain, either offline from collected data-traces or online from a continuous stream of sensed data; (2) the forecasting of new data of the future production; and (3) the automatic generation of synthetic traces, to strengthen the data-based algorithms. All such tasks provide valuable information to an advanced Manufacturing Execution System (MES), which reacts by optimizing the production process and management of the equipment maintenance policies. SMART-IC is a 300k€ academic project funded by the Italian Ministry of University and supported by STMicroelectronics and Technoprobe with industrial expertise and real-world applications. This paper shares the view of SMART-IC on the future of semiconductor manufacturing, the preliminary efforts, and the future results that will be reached by the end of the project, in 2025.**

I. INTRODUCTION

Semiconductor chips form the building blocks of modern technology: as proven by the recent chip shortage crisis, their impact spans across very different markets, ranging from automotive to medical equipment and household appliances [1]. This crucial role, together with the increase in demand and production, makes it essential to ensure the functionality and reliability of semiconductor manufacturing [2]. Typically, a wafer fab with a construction cost of $7 billion would need to recover roughly $4 million per day to amortize the investment. Thus, any breakdown halting production, even for a minute, is highly undesirable. Semiconductor companies are thus in a constant quest to improve quality and reduce the generation of waste and defects to open the way to smaller, faster, and higher-quality devices [3].

Effective chip testing aims at ensuring the quality of semiconductor products, to address any issues before their products reach the market, identifying abnormal processes that cause wafer defects, and ensuring that quality meets specifications. To optimize tests, companies adopt automated test equipment, able to conduct electrical tests on large numbers of chips in parallel. However, such advanced equipment is extremely expensive, ranging from 500k€ for a test handler to 1.2M€ for the most advanced tester infrastructures with 3D support and enhanced tracing capabilities.

Advanced Machine Learning (ML) and Artificial Intelligence (AI) offer new opportunities to optimize the testing processes and achieve greater levels of cost-effectiveness. ML algorithms can indeed be trained to identify patterns and relationships in complex data sets. This makes them well-suited for analyzing the vast amounts of data that are generated during semiconductor manufacturing and testing. ML and AI solutions can thus be highly beneficial when using data collected on the field to monitor the operating conditions of production equipment, predict the correctness of produced wafers, and optimize the operating scheduling in a way that is aware of ongoing processes.

This paper presents SMART-IC, an academic project funded by the Italian Ministry of University and Research and by the Next Generation EU plan. The goal of SMART-IC is to build a complete framework for smart monitoring and production optimization, with the support of STMicroelectronics and Technoprobe, that provide industrial expertise and real-world applications. The main pillars of the project are the development of an *AI-powered engine* and of an *enhanced MES* for achieving *production-aware manufacturing system management*, that exploits relevant information distilled by the AI-powered engine to optimize the scheduling and allocation of production. Many efforts exist in the literature to improve the effectiveness of semiconductor manufacturing. However, they either target specific processes or focus on scheduling and maintenance management without exploiting the availability of large amounts of data with AI-based approaches [4]–[8]. Our project, currently in its initial phase, aims to close the loop with a comprehensive framework where AI and advanced production management cooperate to face the challenges of such a complex scenario.

The paper is organized as follows: Section II details the SMART-IC perspective on semiconductor manufacturing, Section III exemplifies the strategy and shows our preliminary

The work has been partially supported by the PRIN 2022T7YSHJ SMART-IC - Next Generation EU project.

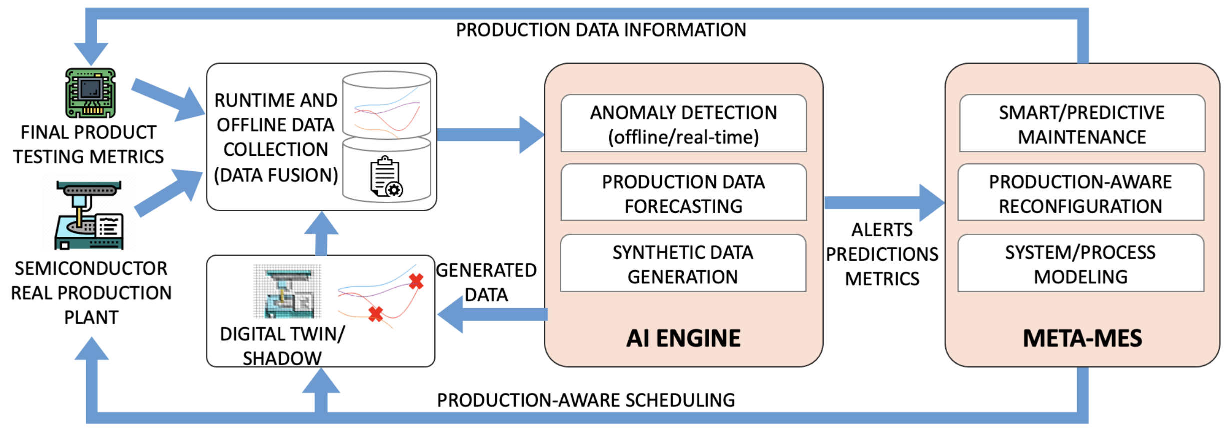

Fig. 1. SMART-IC pillars: AI-powered engine (for anomaly detection, production data forecasting and generation), and meta-MES development for smart semiconductor manufacturing.

study, and Section IV summarizes our final remarks and underlines the impact on the Design, Automation and Test in Europe (DATE) Conference.

II. SMART-IC VIEW ON SEMICONDUCTOR MANUFACTURING

The perspective of SMART-IC on semiconductor manufacturing is to develop and integrate new data-driven approaches for smart monitoring and production optimization, with the goal of reducing defect generation, identifying latent defects on masks and wafers during components manufacturing, and reducing waste products at all levels.

Fig. 1 outlines the framework. The semiconductor manufacturing chain is monitored to *collect relevant data traces relative to physical parameters*, like gas flow, temperature, voltage, and final product testing metrics. Such data is fused in a coherent form and then made available to an AI-powered engine, both in the form of a database of historical values and a continuous stream of real-time data.

Such AI-powered engine is the heart of the SMART-IC framework, as it provides data-based algorithms for:

- *anomaly detection*, applied (i) offline on collected data to assess the quality of a finished product, or (ii) in real-time during a production phase, to send alerts about a possible faulty product/process;
- *production data forecasting*, to predict future data traces of physical quantities and the quality of the corresponding products;
- *synthetic data generation* based on ML algorithms; the generated data can be used to augment the available data to improve the quality of the AI-based algorithms, as well as to construct a digital twin of the manufacturing equipment.

The output of the AI-powered engine is fed to a *Meta-MES, a MES extended with advanced automated features*, such as reconfiguration of the production line, autonomous execution of production orders, resource management, and advanced scheduling [9]. The Meta-MES can react to alerts, predictions, and metrics of the AI-powered engine to:

- apply *smart maintenance*, to estimate mean time to failure/repair and reduce unscheduled equipment downtime (i.e., costs of production time and yield losses as well as maintenance costs);
- obtain a *reconfiguration of the production line* by reallocating the production to apply mitigation strategies based on the conditions of the production equipment or on the quality of the generated products;
- both steps rely on *system and process modeling* to ease the management of the information flows.

Product data information, including testing outputs and production trace-based metrics, are used to annotate the final products and to further enrich the training of the AI-based algorithms.

A. AI-powered engine for anomaly detection, production data forecasting and generation

Driven by the Industry 4.0 revolution, the semiconductor industry is investing heavily in the digitalization of its production chain [10]. As a result of these investments, the chip production process has been equipped with multiple sensors that constantly monitor the evolution of each manufacturing phase, from oxidation to testing and packaging, thus collecting a tremendous amount of heterogeneous data, commonly referred to as *traces*. To fully unveil the potential and hidden knowledge of such data, Artificial Intelligence technologies, and especially Machine Learning, are widely acknowledged to have a fundamental role [11].

Traditionally, data analytics techniques have been applied in semiconductor manufacturing mainly for fault detection (FD) and fault classification (FC) tasks, referring to the monitoring, analysis, and categorization of variations in tool and/or process data to detect faulty parts [12]. Nevertheless, most of the available solutions have limitations in how they approach the problem, as they typically address these tasks as simple problems of univariate statistics, neglecting the inter-dependency between the different variables [13]. Some researchers try to circumvent the problem with “de-correlation approaches”: different time series are de-correlated to formally translate the original multi-dimensional problem into multiple 1-dimensional ones. Nonetheless, the foundations of such decorrelations are generally difficult to demonstrate in real-world scenarios.

On the other hand, modern ML technologies have recently shown great potential in dealing with multivariate time series. Deep learning has become a particularly powerful approach in many applications, including smart manufacturing, offering the capability to capture intricate patterns and dependencies within complex multivariate datasets and across multiple dimensions [14]. This paradigm leverages artificial neural networks with a large number of hidden layers to learn latent representations directly from the temporal sequences of multiple variables, and then exploit the learned representations to solve a number of diverse tasks, that are equally relevant from a smart manufacturing point of view.

Traditional *diagnostic* tasks, such as fault detection and classification, where the algorithm needs to detect and recognize an unwanted deviation of the process *after* it has already occurred, can be solved in a completely unsupervised way by deep learning-based *anomaly detection* techniques: a model is trained on regular data from non-faulty processes and then asked, at run-time, to recognize discrepancies from the learned behavior. Literature provides several successful examples, based on different formulations of Autoencoders or of ensemble methods [15]. By fine-tuning the training and the response window of the model, the diagnosis can happen either *offline* after an entire production step or, working towards *real-time* implementations, at any point within a production step.

Interestingly, state-of-the-art deep learning can go far beyond fault diagnosis and solve way more complex *prognostic* tasks: in this case, a *forecasting* model is trained to predict the future evolution of the data, based on its past characteristics. Promising architectures for this purpose are, again, Autoencoders, as well as recurrent models such as Long Short-Term Memory (LSTM) Networks or auto-regressive deep architectures [15], [16]. From a semiconductor manufacturing viewpoint, this translates into the possibility to predict the future evolution of the sensed variables at multiple levels, to provide metrics related to the future risks of faults in the chip production chain.

Finally, latest deep learning models have shown potential even in the task of *generating* multivariate time series data from scratch. This is a particularly challenging problem, as it requires that the network learns the inherent characteristics of multivariate distributions from the training data, so as to be able to replicate them at run-time. The best architectures for this task can be borrowed from computer vision and natural language processing domains, where deep generative models such as Generative Adversarial Networks (GANs) and Diffusion Models have recently reached impressive results [17]. In a smart manufacturing scenario, the generation of synthetic production data can be exploited in many ways, from solving data-related issues such as missing recordings or lack of training traces, to creating a realistic digital twin of the production process.

Based on these considerations, SMART-IC's strategy is to integrate the aforementioned state-of-the-art deep learning models into a single *AI-powered engine*, which will serve as the core of the decision-making of our framework.

B. Meta-MES for production scheduling and optimization and smart maintenance

Semiconductor manufacturing requires the careful coordination of hundreds of process steps, and even small changes in the production recipes may lead to different production throughput and wear, as a result of both complex processes and of the presence of multiple wafers being processed at the same time, potentially with different processing flows. In this perspective, considerations about anomalies and production forecasts can be used to improve manufacturing efficiency through reconfiguration, with a positive impact on the reduction of unscheduled downtime and the utilization rate of different machinery [18].

To implement this strategy, it is necessary to extend the traditional MES into a Meta-MES, as proposed by [9], i.e., to include advanced automated features such as reconfiguration of the production line, resource management, and advanced scheduling. The Meta-MES acts as the manufacturing operating system, enabling the implementation of custom applications to monitor the production plant, reconfigure production processes, and collect additional key parameters. The Meta-MES takes in input information generated by the AI-powered engine in terms e.g., of punctual anomalies, metrics on the distribution of faulty wafers in lots, or as remaining time to failure estimation. This information can be used, together with models of the production flow or of specific processes, to pursue optimization and management steps.

Among the optimization strategies, the minimization of the machine setup time is one of the most widely used in the manufacturing field [19]. This heuristic tries to schedule production tasks with the same machine setup one after the other. However, manufacturing systems are affected by a set of unexpected events that make the schedule unfeasible, such as machine breakdown, maintenance tasks, and new production orders, which reduce the performance of this heuristic in real production environments. Another known issue in semiconductor manufacturing consists of the quality of the products, which are classified into low-end products and high-end products based on the performance detected in the testing phases [20]. To avoid this problem, other approaches try to increase the resilience of manufacturing systems by forecasting the demand of production orders [21] and integrating in advance maintenance tasks with production tasks [21], [22].

The Meta-MES approach enables advancing a step further by implementing also a "dynamic" schedule, which continuously reconfigures the production plant based on the events notified by the AI-powered engine. Thus, the integration with the digital twin allows testing manifold promising schedules on a digital replica of the production plant. This creates an optimization loop in which the digital twin provides additional information on the computed schedules, allowing to select the best solution among the promising ones.

Finally, all data produced from the machinery during a production batch can be collected and historicized by Meta-MES. These data are fundamental because they allow for comparing testing data, which identifies valid pieces against wastes with data measured during the previous production steps. This allows the generation of new training data that can be used to improve the AI-powered engine to detect anomalies as soon as possible.

III. Preliminary study

Smart-IC will build on top of real semiconductor manufacturing data collected throughout project development and testing. Given the initial stage of the project, in this paper we exploit public data to (1) showcase, even though just in principle and on a limited scale, Smart-IC's strategy; (2) demonstrate the potential impact to the semiconductor industry; and (3) plan the future developments of our project.

A. Scenario

In a typical semiconductor production chain, operations are performed on wafers, and wafers passing through the same operations are aggregated into so-called *lots*. Every step of the production is monitored by a number of sensors recording physical parameters (gas flow, temperature, voltage, etc.). Finally, at the end of the whole production chain, each device is tested to identify faulty wafers.

As a first case-study for Smart-IC, we selected the dataset presented in [23], that well represents the targeted scenario. Sensor data are provided with a 1s timestamp for two different production steps: the deposition process (SPUT, 32 sensors) and the rapid thermal process (RTP, 24 sensors). Each wafer is represented by a trace consisting of a multidimensional time series, with a total duration of 176 s. The data is collected for a total of 54 successive lots, each made of 25 wafers (wafer belonging to the same lot are given the same lot label). The post-production test results of each wafer are also provided, in the form of a binary label (1: faulty, 0: not faulty), which is the only ground truth made available for our analysis.

B. Anomaly detection

The first task of the AI-powered engine is *anomaly detection*, that in our case is the automatic identification of faults in the production chain. For this purpose, a ML model is typically trained offline on non-faulty data, to learn the characteristics of a "normal" behavior, so as to be able at run-time to identify any events that significantly deviate from the normality. This can be done at different time-resolutions, by simply changing our definition of *event*.

1) Offline: Given that the ground truth is available only at the wafer-level, and each wafer is represented by 176 s long traces, the first and most obvious task for *anomaly detection* is to consider a whole recording as an event, and to try to identify the anomalous wafers from the normal ones. This goes by the name of *offline* anomaly detection, implying that an entire data

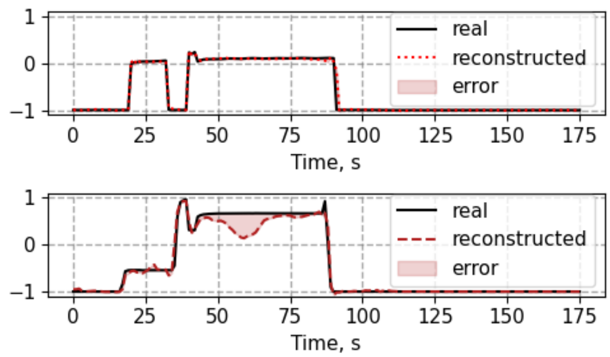

Fig. 2. Example of a real and a reconstructed sensor trace from a normal (top) and faulty (bottom) wafer.

trace needs to be collected before we can understand whether the corresponding wafer is faulty.

A promising approach is, for example, *reconstruction-based*: a ML model can be trained on the task of reconstructing a 176s long trace corresponding to a single wafer. At inference time, it can be expected that a data trace corresponding to a non-faulty wafer will be similar to the ones the model was trained on, and hence will be reconstructed with a low reconstruction error. Vice-versa, anomalous data traces will be reconstructed with high error (see Fig. 2).

Starting from these considerations, we interpret anomaly detection as a binary classification task, where an anomalous trace can be identified by simply imposing a threshold on its reconstruction error at run-time.

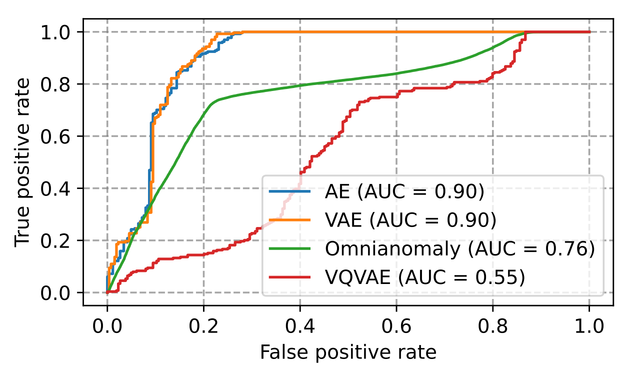

Fig. 3. ROC curves of the tested anomaly detection methodologies.

To assess the feasibility of this task with the available dataset, we made preliminary experiments with different widely used ML models: *Autoencoder (AE)* [24], *Variational Autoencoder (VAE)* [25], *Vector Quantized Variational Autoencoder (VQ-VAE)* [26], as well as *Omnianomaly* [27], that leverages a combination of stochastic recurrent neural networks and VAE. In Fig. 3, we show the results of these experiments in terms of ROC curves. In such curves, the true positive and the false positive rates of each model are plot at varying values of the threshold used to identify the anomaly. Hence curves closer to the top-left corner indicate a better performance, and the area under the curve (AUC) provides a single threshold-less $[0, 1]$ measure of the algorithm's ability to identify the anomaly.

From the obtained results, we can see that most techniques provided a good performance, corresponding to AUC values of 0.76 for Omnianomaly and 0.9 for AE and VAE. This demonstrates our initial hypothesis, that the reconstruction error has a good discriminative power, and can be effectively used to identify the faulty wafers during production.

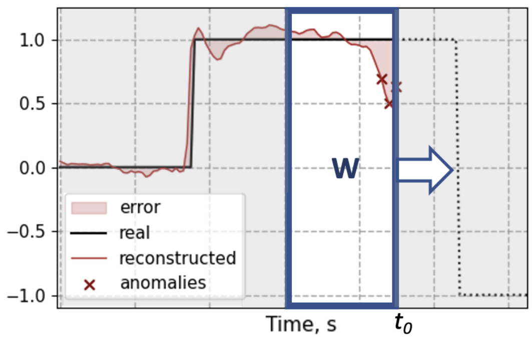

Fig. 4. Example of window-based real-time anomaly detection: the reconstruction error of a sliding window W is computed at each new timestamp t_0. If the error exceeds the imposed threshold, t_0 is identified as anomalous.

2) Real-time: By receiving the sensor data in the form of a continuous stream, the anomaly detection algorithm can be easily adapted to identify anomalous events at a higher time resolution, or even to identify punctual anomalies, corresponding to single data samples. While the currently available dataset cannot be used to provide quantitative results because it does not contain annotations of punctual anomalies, we can use it to exemplify the strategy in principle, as shown in Fig. 4. Instead of the entire recording, the ML model is trained to reconstruct a small time-window W. At inference-time, W moves with a stride of 1s (equal to the given timestamp), so as to reconstruct the windowed signal up-to the current timestamp t_0. At each new t_0, the reconstruction error of the current window is computed to decide whether t_0 is anomalous or not, and send a corresponding alert to the Meta-MES.

C. Production data forecasting

The same ML backbones can be effectively exploited also for the forecasting of future production data, as follows.

- Instead of being trained to reconstruct the input signal as is, the ML model is trained to predict a future signal $X(t_1)$, by taking as input a current signal $X(t_0)$, and/or a past one $X(t_{-1})$.
- Same as for the anomaly detection task, the width of the time windows can be tuned, so as to obtain prediction of single data samples, or of progressively wider time windows. For example: predicting the production data of a future lot, given the data of the past N ones.
- Finally, by applying anomaly detection on the forecast data instead of the past data, it is possible to predict the future evolution of the production process, so as to anticipate faulty events. This is a valuable information for the Meta-MES, as it can be used to optimize future production steps, as well as to implement targeted maintenance policies (details will follow).

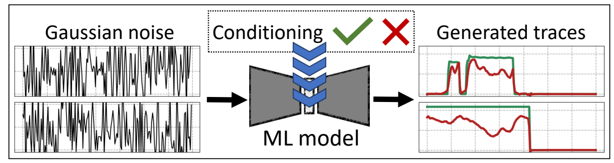

Fig. 5. Example of data generation task. Given random Gaussian white noise traces as input, a generative model can be conditioned to output realistic production data of normal (green) and faulty wafers (red).

D. ML-based synthetic data generation

The last task addressed by the AI engine is the generation of synthetic production data from scratch. In this case, the ML model takes in a random input, (for example, Gaussian noise) and produces as output a completely new data trace, that attempts to mimic the training data distribution. Latest generative algorithms such as GANs [28] and Diffusion Models [29] are very promising for this task, as they can be opportunely conditioned towards a specific type of distribution by simply adding label information during the training. By doing so, the model learns to generate synthetic data of a specific class. This principle is shown in Fig.5, where the output is data traces of either a normal wafer (green) or of a faulty one (red).

A known problem of generative models is that they typically necessitate huge training sets, considerably larger than the one of our preliminary case study. To circumvent this problem, SMART-IC's framework is designed to expose the AI engine to both runtime and offline data collection, thus allowing to continuously train the ML models with newly produced sensor data and test labels, to progressively improve its generative capability.

E. Maintenance and production scheduling/optimization

In the SMART-IC project vision, the Meta-MES is the production supervisor that is responsible of optimizing the production and the maintenance phases in order to ensure high quality standards for obtaining high-end products. The Meta-MES receives from the AI engine different information: (1) production alerts in the presence of anomaly or punctual data traces; (2) predictions of future production conditions, from which it is easy to derive degradation metrics, e.g., Mean Time To Failure (MTTF); (3) synthetic data traces related to nominal production behaviors or faulty synthetic data traces obtained with data generation and augmentation techniques. Based on these Meta-MES inputs, several scheduling/optimization techniques are being tested to improve the prediction capabilities of the Meta-MES and keep each production phase of the manufacturing plant in optimal working conditions. For example, a practical case could arise when, through the continuous monitoring of the different data sources and based on the feedback provided by the AI engine, the MES identifies a specific portion of wafers as subject to physical defects. In this case, the Meta-MES could suggest and apply new strategies to circumvent this problem associated with specific production phases and automatically label the chips made from this wafer portion as low-end products. Consequently, all actions that the Meta-MES could apply to the manufacturing plant are in the view of not wasting finished products and avoiding unexpected maintenance issues through preventive actions on the plant's machinery. Another practical case could happen in the last phases of semiconductor production, when the chip needs to be tested after the packaging. In this case, the AI engine could process data from machinery' sensors and data retrieved from sensors on-chip to access more precisely chip's internal conditions. Based on the AI engine output, the Meta-MES will react consequently to the production schedule to set the parameters of each machinery correctly and maintain the chip yield as high as possible.

IV. Summary and connection with DATE conference's community

As explained in this paper, SMART-IC targets semiconductor manufacturing, a sector that is strategic for the digital transformation, with the aim of improving its overall efficiency, reliability and sustainability. By doing so, it offers significant benefits to the community of the Design, Automation and Test in Europe (DATE) Conference, as well as to all the researchers and practitioners in the field of digital technologies.

To achieve its aims, the project will develop an one-of-its-kind framework, where an advanced MES handles scheduling, maintenance and production reconfiguration based on knowledge extracted by an AI engine, empowered by state-of-the art deep learning. Hence, SMART-IC brings together the latest technologies in IoT, Artificial Intelligence and data analytics, which are all topics of interest for the DATE conference.

Ours is indeed a very ambitious project, that presents many challenges that need to be addressed to fully realize its potential. We are confident that the interaction with the DATE audience, which includes the most experienced specialists of design, test and manufacturing of electronic circuits and systems, will help us identify and understand those challenges in the best way possible, so as to be able to tackle them efficiently.

References

[1] Semiconductor Industry Association, "Global semiconductor sales increase 1.9% month-to-month in september," https://www.semiconductors.org, Nov. 2023.

[2] European Commission, "European Chips Act," https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age/european-chips-act_en, 2022.

[3] T. F. Edgar, S. W. Butler, W. J. Campbell, C. Pfeiffer *et al.*, "Automatic control in microelectronics manufacturing: Practices, challenges, and possibilities," *Automatica*, vol. 36, p. 1567–1603, 2000.

[4] S. Umeda, K. Tamaki, M. Sumiya, and Y. Kamaji, "Planned maintenance schedule update method for predictive maintenance of semiconductor plasma etcher," *IEEE Transactions on Semiconductor Manufacturing*, vol. 34, no. 3, pp. 296–300, 2021.

[5] A. M. Ham and M. Cho, "A practical two-phase approach to scheduling of photolithography production," *IEEE Transactions on Semiconductor Manufacturing*, vol. 28, no. 3, pp. 367–373, 2015.

[6] A. Upasani, R. Uzsoy, and K. Sourirajan, "A problem reduction approach for scheduling semiconductor wafer fabrication facilities," *IEEE Transactions on Semiconductor Manufacturing*, vol. 19, no. 2, pp. 216–225, 2006.

[7] Y. Tang and M. Zhou, "Design of reconfigurable semiconductor manufacturing systems with maintenance and failure," in *Proceedings 2001 ICRA. IEEE International Conference on Robotics and Automation (Cat. No.01CH37164)*, vol. 1, 2001, pp. 559–564 vol.1.

[8] E. Kim, T. Kim, D. Lee, H. Kim *et al.*, "Practical reinforcement learning for adaptive photolithography scheduler in mass production," *IEEE Transactions on Semiconductor Manufacturing*, pp. 1–1, 2023.

[9] S. Gaiardelli, S. Spellini, M. Panato, M. Lora *et al.*, "A software architecture to control service-oriented manufacturing systems," in *Design, Automation & Test in Europe Conference & Exhibition (DATE)*, 2022, pp. 40–43.

[10] D. Kim and K. Cho, "Digital transformation characteristics of the semiconductor industry ecosystem," *Sustainability*, vol. 15, no. 1, p. 483, 2022.

[11] C.-F. Chien, H. Ehm, J. Fowler, L. Mönch *et al.*, "Production-level artificial intelligence applications in semiconductor manufacturing," in *2021 Winter Simulation Conference (WSC)*. IEEE, 2021, pp. 1–8.

[12] J. Moyne and J. Iskandar, "Big data analytics for smart manufacturing: Case studies in semiconductor manufacturing," *Processes*, vol. 5, no. 3, p. 39, 2017.

[13] K. Shaukat, T. M. Alam, S. Luo, S. Shabbir *et al.*, "A review of time-series anomaly detection techniques: A step to future perspectives," in *Future of Information and Communication Conference*. Springer, 2021, pp. 865–877.

[14] G. Li and J. J. Jung, "Deep learning for anomaly detection in multivariate time series: Approaches, applications, and challenges," *Information Fusion*, vol. 91, pp. 93–102, 2023.

[15] A. A. Cook, G. Mısırlı, and Z. Fan, "Anomaly detection for IoT time-series data: A survey," *IEEE Internet of Things Journal*, vol. 7, no. 7, pp. 6481–6494, 2019.

[16] R. Kumar and N. Agrawal, "Analysis of multi-dimensional Industrial IoT (IIoT) data in edge-fog-cloud based architectural frameworks: A survey on current state and research challenges," *Journal of Industrial Information Integration*, p. 100504, 2023.

[17] R. Gozalo-Brizuela and E. C. Garrido-Merchán, "A survey of generative ai applications," *arXiv preprint arXiv:2306.02781*, 2023.

[18] D. Suerich and T. Young, "Reinforcement learning for efficient scheduling in complex semiconductor equipment," in *SEMI Advanced Semiconductor Manufacturing Conference (ASMC)*, 2020, pp. 1–3.

[19] I.-B. Park, J. Huh, J. Kim, and J. Park, "A reinforcement learning approach to robust scheduling of semiconductor manufacturing facilities," *IEEE Transactions on Automation Science and Engineering*, vol. 17, no. 3, pp. 1420–1431, 2020.

[20] H. Chen and D. Boning, "Online and incremental machine learning approaches for ic yield improvement," in *2017 IEEE/ACM International Conference on Computer-Aided Design (ICCAD)*, 2017, pp. 786–793.

[21] W. Fu, C.-F. Chien, and Z.-H. Lin, "A hybrid forecasting framework with neural network and time-series method for intermittent demand in semiconductor supply chain," in *Advances in Production Management Systems. Smart Manufacturing for Industry 4.0*, I. Moon, G. M. Lee, J. Park, D. Kiritsis *et al.*, Eds. Cham: Springer International Publishing, 2018, pp. 65–72.

[22] T.-G. Lee, T.-S. Yu, and T.-E. Lee, "Cleaning plan optimization for dual-armed cluster tools with general chamber cleaning periods," *IEEE Transactions on Automation Science and Engineering*, vol. 20, no. 3, pp. 1890–1906, 2023.

[23] M. Pleschberger, A. Zernig, and A. Kaestner, "Equipment Sensor Data from Semiconductor Frontend Production," Nov. 2020. [Online]. Available: https://doi.org/10.5281/zenodo.4322353

[24] D. H. Ballard, "Modular learning in neural networks," in *Proceedings of the sixth National conference on Artificial intelligence-Volume 1*, 1987, pp. 279–284.

[25] D. P. Kingma and M. Welling, "Auto-encoding variational bayes," *arXiv preprint arXiv:1312.6114*, 2013.

[26] A. Van Den Oord, O. Vinyals *et al.*, "Neural discrete representation learning," *Advances in neural information processing systems*, vol. 30, 2017.

[27] Y. Su, Y. Zhao, C. Niu, R. Liu *et al.*, "Robust anomaly detection for multivariate time series through stochastic recurrent neural network," in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ser. KDD '19. New York, NY, USA: Association for Computing Machinery, 2019, p. 2828–2837.

[28] A. Jabbar, X. Li, and B. Omar, "A survey on generative adversarial networks: Variants, applications, and training," *ACM Computing Surveys (CSUR)*, vol. 54, no. 8, pp. 1–49, 2021.

[29] L. Yang, Z. Zhang, Y. Song, S. Hong *et al.*, "Diffusion models: A comprehensive survey of methods and applications," *ACM Computing Surveys*, 2022.