POLITECNICO DI TORINO
Repository ISTITUZIONALE

SMART-IC: Smart Monitoring and Production Optimization for Zero-waste Semiconductor Manufacturing

Original
SMART-IC: Smart Monitoring and Production Optimization for Zero-waste Semiconductor Manufacturing / Alamin, KHALED SIDAHMED SIDAHMED; Chen, Yukai; Gaiardelli, Sebastiano; Spellini, Stefano; Calimera, Andrea; Beghi, Alessandro; Susto, Antonio; Fummi, Franco; Macii, Enrico; Vinco, Sara. - (2022), pp. 1-6. (Intervento presentato al convegno IEEE Latin-American Test Symposium tenutosi a Montevideo (Uruguay) nel 5th - 8th September 2022) [10.1109/LATS57337.2022.9937011].

Availability:
This version is available at: 11583/2970802 since: 2022-08-29T15:22:34Z

Publisher:
IEEE

Published
DOI:10.1109/LATS57337.2022.9937011

Terms of use:

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

Publisher copyright
IEEE postprint/Author's Accepted Manuscript

©2022 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collecting works, for resale or lists, or reuse of any copyrighted component of this work in other works.

(Article begins on next page)

17 April 2024

SMART-IC: Smart Monitoring and Production Optimization for Zero-waste Semiconductor Manufacturing

Khaled Sidahmed Sidahmed Alamin[1], Yukai Chen[1], Sebastiano Gaiardelli[3], Stefano Spellini[3], Andrea Calimera[1], Alessandro Beghi[2], Antonio Susto[2], Franco Fummi[3], Enrico Macii[1], Sara Vinco[1]
1: Politecnico di Torino, Italy ({name.surname}@polito.it) 2: University of Padova, Italy ({name.surname}@unipd.it)
3: University of Verona, Italy ({name.surname}@univr.it)

***Abstract*—The Industry 4.0 revolution introduced decentralized, self-organizing, and self-learning systems for production control. New machine learning algorithms are getting increasingly powerful to solve real-world problems, like predictive maintenance and anomaly detection. However, many data-driven applications are still far from being optimized to cover many aspects and the complexity of modern industries; correlations between smart monitoring, production scheduling, and anomaly detection/predictive maintenance have only been partially exploited. This paper proposes to develop *new data-driven approaches for smart monitoring and production optimization, targeting semiconductor manufacturing*, one of the most technologically advanced and data-intensive industrial sectors, where process quality, control, and simulation tools are critical for decreasing costs and increasing yield. The goal is to reduce defect generation at the electronic component level and its propagation to the system- and system-of-systems- level by working on (1) *enhanced anomaly detection*, based on the human-in-the-loop concept and on advanced treatment of multiple time-series and of domain adaptation, (2) *smart and predictive maintenance* based on both objective data traces and simulated ones, to mitigate the risk of degrading product quality, and (3) the construction of an *extended manufacturing software stack* that allows anomaly- and maintenance-aware policies to enhance production line scheduling and optimization.**

I. INTRODUCTION

In 2021, semiconductor companies substantially ramped up production to unprecedented levels to address persistently high demand. Furthermore, demand for semiconductor production is projected to rise significantly in the years ahead, as chips become even more heavily embedded in essential technologies in a broad spectrum of target systems and production domains [1]. This increase in demand and production occurred amid the ongoing global chip shortage, which is unlikely to be resolved in the near future, partly because of the complexities of the semiconductor production process and geopolitical and pandemic-related issues [2]. In this scenario, manufacturing production effectiveness is thus even more critical. 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 which brings production to a halt, even for a minute, is highly undesirable. For all the reasons above, semiconductor companies are in a constant quest to improve quality and reduce the generation of waste and defects to open the way to smaller, faster, higher-quality devices [3].

In this context, data-driven solutions can be highly beneficial when intended as techniques that exploit data collected on the field (both with acquisition campaigns and in real-time) to monitor the operating conditions of production equipment, predict its state of health and failure probability, and optimize its operating scheduling in a way that is aware of ongoing processes. This is allowed by modern massive inspection systems, interconnected sensors, and real-time monitoring modules that enable the collection of vast amounts of sensed data. Such a huge amount of data can be managed by the Industry 4.0 technologies, particularly by Machine Learning (ML) algorithms to enhance the complex semiconductor manufacturing processes. Even if such technologies are well established, the integration of all aspects in the context of semiconductor manufacturing is not: correlations between smart monitoring, production scheduling, and anomaly detection/predictive maintenance have been only partially exploited, while they could be beneficial to improve effectiveness and reduce production machinery wear.

This paper proposes SMART-IC*, a complete framework for smart monitoring and production optimization for zero-waste semiconductor manufacturing*, that brings together:

- *Anomaly detection* to identify outliers in the data collected from sensors installed in production equipment, to detect anomalies and correct production parameters with the goal of reducing faulty products and waste;
- *Functional safety*, to boost detection effectiveness by enhancing inspection-driven diagnostic with the identification of their root causes;
- *Smart maintenance* supporting the manufacturing process with metrics on wear and alerts on adverse trends;
- *Production-aware manufacturing system management* with the development of a Meta Manufacturing Execution System (Meta-MES) that exploits relevant information distilled by anomaly detection and smart maintenance to optimize the scheduling and allocation of production;

The paper is organized as follows: Section II details the SMART-IC perspective on semiconductor manufacturing. Section III exemplifies the strategy on some preliminary results, and Section IV draws our perspective and conclusions.

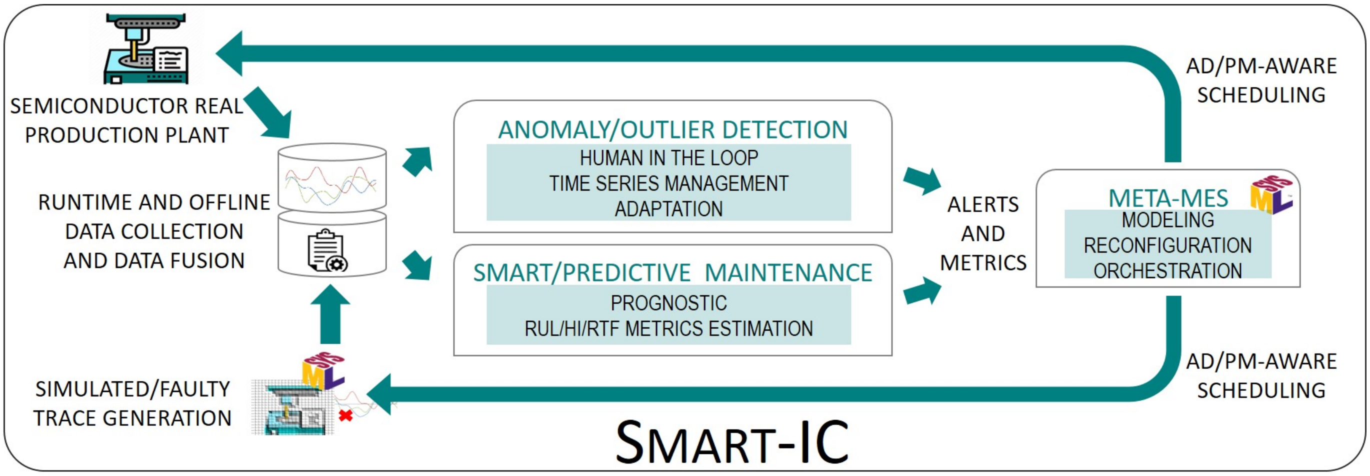

Fig. 1. SMART-IC flow for semiconductor manufacturing optimization based on anomaly detection, smart maintenance, and meta-MES development.

II. SMART-IC VIEW ON SEMICONDUCTOR MANUFACTURING

The perspective of SMART-IC on semiconductor manufacturing is to develop 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, as indicated by the Chips Act promulgated by the European Commission, that underlines the need for *"devices [...] designed for energy efficiency and durability, repairability, upgradability, maintenance, reuse and recycling"* [4].

In the SMART-IC framework, outlined in Figure 1, the semiconductor manufacturing process is monitored to collect relevant data traces that are managed by diagnostics (A) and prognostics (B) algorithms to identify anomalies and outliers and predict future equipment effectiveness and mean time to failure/repair. The outcome of this analysis is used as input by the Meta-MES (C), an advanced manufacturing execution system carrying out diagnostic- and prognostic-aware advanced scheduling and reconfiguration of the manufacturing process (upper-side arrow). The Meta-MES also generates faulty traces via simulation or fault injection on existing traces to overcome the well-known limitations created by limited data availability that typically affect data-centric approaches (lower-side arrow).

A. Anomaly/outlier detection

Anomaly/Outlier Detection (AD) is an important task in unsupervised learning as it aims at detecting anomalous behaviors when dealing with complex and multivariate data. AD approaches summarize the status of a system/phenomenon with a single indicator, the Anomaly Score (AS). While black-box AD approaches have proven to be effective in many real-world scenarios, several factors still limit the applicability in semiconductor manufacturing; examples of limitations are:

a) Lack of interpretability: Two issues are affecting Machine Learning-based technologies: (i) lack of confidence/trust from the users in AD algorithm outcomes and (ii) no immediate association between AD algorithm outcomes and root causes. The first issue arises from the lack of labelled data points (which, on the other hand, is one of the main reasons AD algorithms are appealing in the first place), making it impossible to set up adequate testing procedures. This leads either to trust the algorithm blindly or not to use it at all, both the cases being undesirable. The second issue investigates the possibility of gaining additional knowledge about the task at hand, which may translate into actionable insights for troubleshooting or root cause analysis. The aforementioned issues can be addressed following the principles of eXplainable Artificial Intelligence (XAI) [5].

b) Dealing with time-series data: Semiconductor Manufacturing AD problems are typically associated with multi-dimensional time-series data for batch production (i.e., Multivariate time-series with a finite duration associated with a single processed wafer). In the semiconductor manufacturing literature, such data are called 'trace data'. Nevertheless, most of the time-series AD approaches proposed in the literature are tailored for 1-dimensional continuous time-series [6]. Some researchers deal with multi-dimensionality with 'de-correlation approaches': different time-series signals are de-correlated, and the multi-dimensional problem is formalized as multiple 1-dimensional problems. Once the signals live in a space where they are de-correlated, the anomaly scores can be computed on single time-series (by using any 1-dimensional AD approach) and then composed with some heuristics to provide the final score. Unfortunately, it is quite hard to find a principled and robust model for de-correlating signals in real-world scenarios. For this reason, many researchers tackle the AD task by modeling the phenomena at hand, by resorting to the 'prediction & residual monitoring'. However, such an approach is often unsatisfactory since it requires solving a more complex task (the forecast) than the one we are interested in (the AD), typically leading to overcomplicated or ineffective solutions.

The SMART-IC approach integrates the AS into production optimization systems and exploits AD methodologies that tackles the two aforementioned limitations. While the proposed approach is agnostic to the choice of the AD algorithm, tree-based approaches algorithms [7] are promising candidates in SMART-IC, given their popularity, the availability for interpretability [8] and time-series handling [9].

B. Smart and predictive maintenance

Smart and Predictive Maintenance (PM) continuously monitors equipment conditions to anticipate failures, initiate maintenance activities, and optimize equipment utilization [10]. Its application is, in particular, beneficial in the semiconductor domain, where the demand for suitable maintenance strategies arises from high costs of unscheduled equipment downtime, which comprises the costs of production time and yield losses as well as maintenance costs itself while scheduling maintenance activities in a demand-oriented manner allows minimizing frequency and duration of equipment unavailability [11]. Semiconductor capital equipment suffers indeed at least 8% unscheduled downtime and loses another 7% to scheduled maintenance, and each hour of downtime for a critical unit of process equipment can translate into $100,000 of lost revenue in today's chip-hungry market [12].

Smart maintenance solutions are typically data-based methodologies that use statistical or ML algorithms to calculate equipment health based on sensor data by computing reference metrics, such as Remaining Useful Life (RUL), Overall Equipment Effectiveness (OEE), or Mean Time To Repair (MTTR) [10], [11]. The main drawback is the lack of data, as data may frequently be incomplete or not sufficient: in the real world, failure history data is frequently not enough, as real-time data is rarely stored as a whole, and it is unlikely that data will be present when the actual failure occurred (as an effect of fixing equipment before failure).

The solution proposed by SMART-IC is thus to make the most of available data and to derive faulty traces to compensate for the drawbacks of traditional data-based smart maintenance approaches. Data will be beneficial for building metrics and algorithms that reflect the actual condition of the equipment. On the other hand, *fault injection, transfer learning and data augmentation are used to create faulty (anomalous) traces* and to propagate metrics and findings from one machine to another.

Leveraging a mix of sensed and generated data allows for building an initial metric (such as an RUL model) from generated data to allow the early state of health estimation and to elaborately update the smart prediction model with sensed data traces if failure history is accumulated later (for the equipment under analysis or similar equipment). Faulty data is generated starting from sensed data by exploiting models of failure taken from the state of the threat of every physical domain, allowing a generation of time series of faulty data with a broad temporal horizon. They can be calibrated with respect to the evolution of the production line. For example, model parameters are calibrated once programmed maintenance is completed to align to the new configuration.

Based on the proposed smart maintenance techniques (Figure 1), it is thus possible to obtain prognostic metrics with a two-folding impact on the evolution of the manufacturing process: allow the application of demand-driven maintenance, and provide relevant input to the Meta-MES for subsequent scheduling and reconfiguration operations.

C. A Meta-MES for production scheduling and optimization

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 prognostics and diagnostics 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 [13].

The first ingredient to enable reconfiguration in manufacturing systems is considering the *production scheduling perspective*. Typically, a scheduling technique is constructed to arrange manufacturing processes according to one (or many) optimization objective. To name a few examples, a scheduler may pursue meeting delivery dates, maximizing resource utilization, reducing the makespan and balancing the allocation of tasks to the different workstations. Nonetheless, given the highly complex, non-deterministic challenges implied by scheduling, and the lack of generalized solutions, creating optimal scheduling of manufacturing operations is a labour-intensive process based on heuristic solutions [13]. In the literature, there is a plethora of reconfiguration solutions (i.e., scheduling) based on static and dynamic techniques, but practical problems are proved to be NP-hard [14]. On the one hand, most "dynamic" techniques are based on Artificial Intelligence methodologies [15], [16]. On the other hand, solutions falling into the "static" category are typically based on state-space search algorithms [17], [18]. Scheduling techniques are implemented using a particular optimization model that is strictly related to both the parameter(s) to optimize and the system's constraints. As an example, Resource-Tasks Networks (RTN) [19] and State-Tasks Networks (STN) [20] focus on formalizing production recipes as a directed graph (the former expressing the sequence of material states associated with tasks, the latter including also the allocation of tasks and resources to physical machines).

A critical factor in achieving an efficient reconfiguration is to adopt a lightweight software architecture controlling the production line. In this regard, an emerging paradigm is Service-Oriented Manufacturing (SOM) [21], where systems distribute the responsibility of reconfiguration across the available manufacturing components. This is realized as a Meta-MES [22], that assists the integration of the concept of *production service* within an existing system by acting as an intermediate software control layer. More specifically, the Meta-MES is connected to the traditional MES installed on top of the actual system, which handles the production life-cycle and the definition of production recipes (where one recipe is a set of *tasks* that can be executed by a single, specific machine or a class of components providing the required functionality). The Meta-MES expands such a concept by refining tasks into services, more concretely defining the behaviour of the machine executing the process. In other words, a task is composed of a set of services. The linking between recipe, tasks and services is defined in a model, represented using the Systems Modeling Language (SysML) as a set of behavioural diagrams, one for each level of detail (e.g., tasks and services) [23]. Such a process hierarchy represented in the model is fundamental to carrying out advanced scheduling techniques: the control granularity overproduction is a crucial enabler of precise preemption, interruption and, consequently, interleaving of a set of processes (i.e., orders defined by a production recipe).

To combine the production schedule with prognostic and diagnostic data, the scheduler must be aware of key information such as wafer defects ratios and remaining time before maintenance. In this regard, the Meta-MES handles the production of faulty wafers as *events* (i.e., alarms), constructing a reaction routine to reallocate and reconfigure the production. Such a reaction routine must be defined a priori within the model, alongside the recipe. This step is fundamental for the scheduler to react and correctly schedule the mitigation strategy. In fact, when alarms are raised from monitoring algorithms, the production schedule will be re-planned to guarantee that no production stop will happen. This is possible in a production line with redundancy, by reallocating production to other machines capable of executing the same tasks, thus avoiding a global production stop and allowing to carry out maintenance tasks while meeting the production objectives and deadlines.

III. Exemplification

To prove the impact of the SMART-IC strategy, we applied it to an open dataset, including sensor data for the deposition process (SPUT) and the rapid thermal process (RTP) [24]. The hardware used in this experiment is a workstation with 8GB of RAM and a processor of Intel core i7 of 2GHz.

A. Dataset description and preprocessing

The dataset in [24] includes sensor data (like gas flow, temperature, voltage, and so on.) and testing results (normal/fraud) for 54 lots. Each lot is made of 25 wafers, with a time step of 1s, focusing only on the deposition process (SPUT, 32 sensors) and the rapid thermal process (RTP, 24 sensors). The dataset is organized into three files: the first two files represent sensors data-trace for SPUT and RTP, and the third contains the wafer's final testing result, labelled as 1 or 0 to represent normal and fraud wafer, respectively. Lot number and wafer number are used as unique identifiers.

Data preparation and cleaning have been applied to the dataset. Missing values have been filled with 0s to satisfy the required data dimension (4D for smart maintenance, 3D for anomaly detection). Additionally, 9 sensors data have been removed as considered non informative (75% of the samples were 0s and the maximum value was equal to 1). The final cleaned data contains 237,600 rows with 51 features each.

B. Anomaly detection

To apply anomaly detection techniques, the dataset is analysed with a *convolutional autoencoder*, a connected neural network with fewer neurons in the hidden layers than in the input and output layers, with the goal of capturing the most representative features of the original input data and disregarding the individually specific details of some input data, such as outliers. The autoencoder is then used reconstruct the input data, and typical data is considerably easier to reconstruct than outliers [25].

a) CAE architecture: The proposed anomaly detection model is based on the Convolutional Deep Autoencoder (CAE) in [26], extended with extra convolution neural network layers to increase the model complexity (CAE_D), the subscript D indicates detection. It is composed of the encoder and decoder parts. The encoder part is made up of multiple convolutional layers with a kernel of different sizes, pooling operation, and an activation layer. On the other hand, the decoder part is composed of the same convolution operation number with kernels, an activation function, and an up-sampling operation representing the reverse of a pooling operation.

This work's CAE_D is composed of 10 convolution layers. The encoder comprises five one-dimensional convolutional layers, each connecting with one activation function (ReLU), followed by an average pooling layer, which receives the input data-trace from the 47 sensors. The convolutional and average pooling layers are merged five times in a row, with the output of the last average pooling layer reflecting the final code in the latent space. Simultaneously, dimension reduction decreases data size to around 1/8 of that of the incoming data. Finally, the decoder part reconstructs the input sensor data-trace. For training of CAE_D, 150 epochs are employed using an Adam optimizer with a learning rate of 0.001. The CAE_D uses only normal sensor data-traces for the training phase, while both normal and fraud sensor data-traces are applied in the testing phase. Examples of reconstruction are shown in Figure 2, where the blue line is the data loaded from the dataset, and the red line is the output of the CAE reconstruction.

b) Classification: After reconstructing the input sensor data-trace using the trained CAE_D, a threshold is set to classify the predicted wafer's quality into normal or fraud type depending on the root mean square error (RMSE) between input and output data-trace. Figure 3 shows the output of this classification process for all wafers in the dataset: there is a robust clustering of normal wafers vs. fraud ones, as our trained CAE_D can reconstruct the sensor data-trace of the normal wafer with RMSE of around 0.17, while the fraud wafer reconstruction RMSE is increased to about 0.9. Therefore, we set a reconstruction RMSE threshold of 0.5 to distinguish between the normal and fraud wafer.

c) Classification performance: Classification output has been compared against the label provided by the dataset. The proposed CAE_D solution could achieve an F1-score = 1, MCC = 1, and accuracy = 100% since the threshold of 0.5 can perfectly distinguish the fraud wafer from normal ones. Notice that even with a lower threshold of 0.1 reconstruction error, the trained CAE_D still performs remarkably well, with F1-score = 0.98, MCC = 0.97, and accuracy = 98%. This proves that a classification algorithm can detect faulty wafers even if trained only on normal ones, e.g., on data collected at the

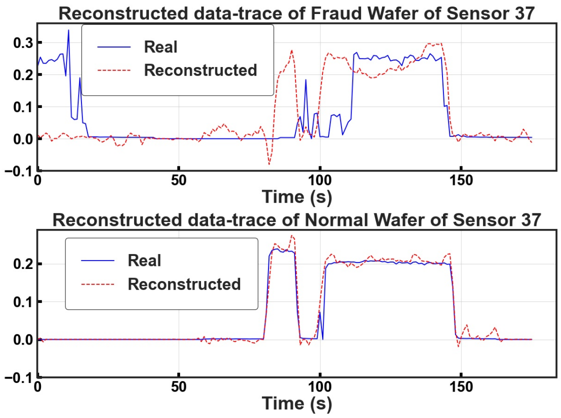

Fig. 2. The real and reconstructed sensor data of normal and fraud wafers.

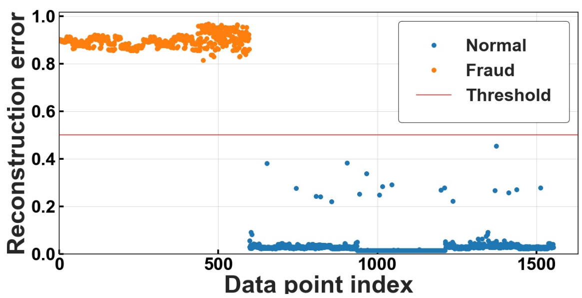

Fig. 3. Reconstruction error (RMSE) for the anomaly detection approach.

beginning of the lifetime of production equipment, when faults do not occur and without time-consuming training on faulty products.

d) Comparison of reconstructed data-trace: To further verify the accuracy of the trained CAE_D, we analyzed the data-trace of the same sensor corresponding to a normal wafer and a fraud wafer for the CAE_D reconstruction (Figure 2). Our CAE_D can reconstruct a normal wafer reasonably well, as indicated in the subplot at the bottom of Figure 2; however, for the reconstructed data-trace of a fraud wafer, our CAE_D performs poorly, as shown at the beginning of the top subplot of Figure 2. The difference in the ability to reconstruct the sensor data-trace of a normal wafer and a fraud wafer allows us to distinguish between the two types of the wafer.

C. Smart maintenance

The strategy adopted for smart maintenance is divided into two phases: (i) prediction of future sensor evolution with a CAE, namely CAE_P, and (ii) subsequent detection of potential future faults with the anomaly detector CAE_D.

a) CAE_P architecture and performance: In order to train a CAE_P with predictive capability, the dataset has been reshaped to a 4-dimensional axis with number of training data, number of wafers in each lot (25), timestamp of each wafer (176), and number of traces used (47). One difference between CAE_P and CAE_D is that CAE_P is based on 3-dimensional instead of 2-dimensional. It has the same neural architecture as the CAE_D in terms of convolutional, pooling, and activation layers. An Adam optimizer with a learning rate of 0.001 is used to train the CAE_P across 100 epochs. The input of CAE_P are the sensors data traces for the current lot, and the output are the estimated sensors data traces of the following lot. The whole dataset has been divided equally into training and testing sets due to its limited size.

The CAE_P is trained until it reaches a satisfactory error rate and then tested on the testing set using regression metrics. When compared with the corresponding sensor data, CAE_P gets a RMSE=0.0495, MAE = 0.1228 and R2=-0.3010.

b) Future frauds detection with CAE_D: The output of CAE_P is then reshaped to 3-dimensional by associating each wafer in the following lot for anomaly detection using the trained CAE_D described in the previous section. Then, the detective result of the CAE_D is compared against the actual label of the wafers in the following lot. As seen from the Figure 4, the CAE_D can distinguish the normal and faulty wafer in the following lot based on the results predicted by the CAE_P, with an F1 score=1.

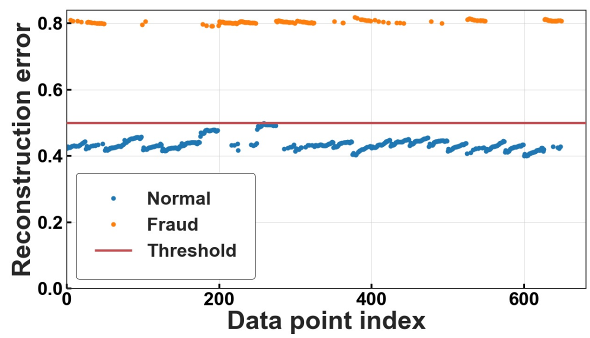

Fig. 4. Reconstruction error (RMSE) for the smart maintenance approach.

Predicting the characteristics of the following lot may be considered limited, even if informative. However, this limitation is due to the limited available data (only 54 lots), which does not allow extensive training. The availability of additional sensed data, and the generation of data with fault injection, would thus allow the training of models that forecast further in the future, thus providing more information to the anomaly-aware production management infrastructure.

D. Anomaly-aware production management

To carry out anomaly-aware production management, the Meta-MES proposes a specific scheduling procedure, that continuously re-schedules and optimizes the production plan based on two kinds of events:

- *anomaly detection* on a task that is being executed by a machine: the scheduler restarts the entire lot from scratch, and allocates it to another machine that is capable of carrying out the same production;
- *prediction* of the failure of a machine: the scheduler plans a maintenance task to fix the faulty equipment and, therefore, to avoid damaging subsequent lots.

The proposed scheduling algorithm is based on a stochastic local search with cost function subject to the total *makespan*:

1) initialize a starting solution and evaluate the makespan;
2) explore the neighborhood, evaluating the makespan for each neighbor;

3) select the neighbor with the minimum makespan;
4) go back to step 2 and repeat until the selected solution cannot be optimized further.

To demonstrate the impact of preventive diagnostic and prognostic events on the reference dataset, it is important to note that the dataset contains 54 unique production lots (Table I), each containing two tasks on two different types of machines (SPUT and RTP), where RTP can be executed only after SPUT [27]. The experiments consider a minimum configuration that enables flexibility for the production schedule (4 equipment, 2 of each type). This allows planning maintenance tasks while carrying on the production by reallocating tasks to other equipment of the same type.

TABLE I
RESULTS COMPARISON OF THE SCHEDULING ALGORITHM WITH AND WITHOUT DIAGNOSTIC AND PROGNOSTIC EVENTS.

Scheduling Type	# Total Lots	# Total Tasks	# Total Maint. Tasks	# Total Failed Tasks	Makespan (hh:mm:ss)
Anomaly Unaware	54	136	0	28	01:48:32
Anomaly Aware	54	130	6	18	01:39:54

Table I summarizes the results that the algorithm can obtain with or without diagnostic and prognostic events. In particular, the first row considers the makespan without considering and, consequently, managing AD and PM events. The second row, on the other hand, reports the makespan of the proposed anomaly-aware scheduling strategy. Even if the dataset comprehends a small production window on a portion of the entire production line, the scheduler manages to reduce the total makespan by 8%. Furthermore, by exploiting the PM data presented in this paper, it also reduces the number of failed tasks by 46%. This is a more significant improvement because it highlights that the same production quantity can be achieved with fewer resources and, therefore, fewer production wastes.

IV. CONCLUSIONS

The paper presented SMART-IC, a data-driven approach to semiconductor manufacturing that can enhance production and reduce waste through three main ingredients: (1) anomaly detection, (2) smart and predictive maintenance, and (3) defect-aware production optimization. The proposed solution has been discussed and applied to an open dataset that, despite the limited data available, allowed us to appreciate the positive impact of SMART-IC on production quality. Future work will apply these techniques to industrial data to further develop the technical solutions to the specific characteristics of a production line.

REFERENCES

[1] Semiconductor Industry Association, "Global semiconductor sales increase 24% year-to-year in October; annual sales projected to increase 26% in 2021, exceed $600 billion in 2022," https://www.semiconductors.org.

[2] D. Takahashi, "KPMG: 56% of chip leaders expect shortage to last in 2023," https://venturebeat.com/2022/03/07/kpmg-56-of-chip-leaders-expect-shortage-to-last-in-2023/, 2022.

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

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

[5] F. K. Došilović, M. Brčić, and N. Hlupić, "Explainable artificial intelligence: A survey," in *International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, 2018, pp. 210–215.

[6] K. Shaukat, T. M. Alam, S. Luo, S. Shabbir, I. A. Hameed, J. Li, S. K. Abbas, and U. Javed, "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.

[7] F. T. Liu, K. M. Ting, and Z.-H. Zhou, "Isolation forest," in *International conference on data mining*. IEEE, 2008, pp. 413–422.

[8] M. Carletti, M. Terzi, and G. A. Susto, "Interpretable anomaly detection with diffi: Depth-based feature importance for the isolation forest," 2020.

[9] S. Guha, N. Mishra, G. Roy, and O. Schrijvers, "Robust random cut forest based anomaly detection on streams," in *International conference on machine learning*. PMLR, 2016, pp. 2712–2721.

[10] D. Fischer, P. Moder, and H. Ehm, "Investigation of predictive maintenance for semiconductor manufacturing and its impacts on the supply chain," in *International Conference on Industrial Technology (ICIT)*, vol. 1, 2021, pp. 1409–1416.

[11] J. Iskandar, J. Moyne, K. Subrahmanyam, P. Hawkins, and M. Armacost, "Predictive maintenance in semiconductor manufacturing," in *2015 26th Annual SEMI Advanced Semiconductor Manufacturing Conference (ASMC)*, 2015, pp. 384–389.

[12] C. Saso, "Remote e-diagnostics: The future of semiconductor manufacturing, part 1," https://www.questteam.com/resources/article.html, 2022.

[13] 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.

[14] J. Blaewicz, W. Domschke, and E. Pesch, "The job shop scheduling problem: Conventional and new solution techniques," *European Journal of Operational Research*, vol. 93, no. 1, pp. 1–33, 1996.

[15] M. Dorigo and C. Blum, "Ant colony optimization theory: A survey," *Theoretical computer science*, vol. 344, pp. 243–278, 2005.

[16] M. Affenzeller, S. Wagner, and S. W. A. Beham, *Genetic algorithms and genetic programming: modern concepts and practical applications*. Chapman and Hall/CRC, 2009.

[17] H. Askari-Nasab, Y. Pourrahimian, E. Ben-Awuah, and S. Kalantari, "Mixed integer linear programming formulations for open pit production scheduling," *Journal of Mining Science*, vol. 47, pp. 338–359, 2011.

[18] M. S. Fox and N. M. Sadeh, "Why is scheduling difficult? a CSP perspective," in *European Conference on Artificial Intelligence (ECAI)*. USA: Pitman Publishing, Inc., 1990, p. 754–767.

[19] A. Barbosa-Povoa and C. Pantelides, "Design of multipurpose plants using the resource-task network unified framework," *Computers & Chemical Engineering*, vol. 21, pp. S703–S708, 1997.

[20] E. Kondili, C. Pantelides, and R. Sargent, "A general algorithm for short-term scheduling of batch operations - milp formulation," *Computers & Chemical Engineering*, vol. 17, no. 2, pp. 211–227, 1993, an International Journal of Computer Applications in Chemical Enginnering.

[21] T. Lojka, M. Bundzel, and I. Zolotová, "Service-oriented architecture and cloud manufacturing," *Acta polytechnica hungarica*, vol. 13, no. 6, pp. 25–44, 2016.

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

[23] S. Gaiardelli, S. Spellini, M. Lora, and F. Fummi, "A hierarchical modeling approach to improve scheduling of manufacturing processes," in *Intern. Symposium on Industrial Electronics (ISIE)*, 2022, pp. 1–7.

[24] 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

[25] Y. Xia, X. Cao, F. Wen, G. Hua, and J. Sun, "Learning discriminative reconstructions for unsupervised outlier removal," in *International Conference on Computer Vision*, 2015, pp. 1511–1519.

[26] H. T. Jebril, M. Pleschberger, and G. A. Susto, "An autoencoder-based approach for fault detection in multi-stage manufacturing: a sputter deposition and rapid thermal processing case study," *IEEE Transactions on Semiconductor Manufacturing*, 2022.

[27] SCREEN Semiconductor Solutions, "Semiconductor manufacturing processes," https://www.screen.co.jp/spe/en/process.