Deep Science Publishing
https://doi.org/10.70593/978-93-49910-47-8_8

Chapter 8: Harnessing big data analytics to improve semiconductor yield, reliability, and predictive maintenance

8.1. Introduction

As the pace of technological advancement increases, the demand for novel solutions across fields is also burgeoning. Industries that rely on intricate and sophisticated systems of manual assembly lines are tasked with eradicating downtime, increasing output, and reducing production costs—all while ensuring the highest product quality. In tandem, upcoming technologies are driving change and improvement in industry. In this landscape, existing production processes are developed and enhanced through the use of cloud systems, data analytics, and connectivity. The goal is to gain insights and implement preventive measures. However, the increased reliance on data results in new challenges and possibilities that were not present before, thus creating a need for organizations to rethink strategy.

The semiconductor manufacturing sector is at the forefront of complexity and requires machinery that produces unique products of intricate design, for which only techniques not previously presented in the manufacturing world are used. Thus, existing approaches from industries that rely on machinery with tight and repeating cycles may not be feasible and might produce results that fail to achieve the expected outcome. Still, the market demands for more solutions from production lines that are increasingly complex themselves. A semiconductor fabrication plant is a building, commonly several stories high and covering more than 500.000 m2 in its entirety, comprising several hundreds of machines, dozens of modules, and expected to produce several thousands of wafers per day. Alternatively, a semiconductor back-end factory is smaller but with equal complexity (Parmar, 2021; Kalusivalingam et al., 2022; Anang & Chukwunweike, 2024).

As the broadest area of industry, the semiconductor industry is also highly autonomous, whereby each machine is tasked with carrying out a subsection of the overall process, e.g., a lithographic process that deposits a layer of photoresist and enables designing at the nanoscale level. The back-end assembly used to produce and prepare the chips in the package is equally technological and demanding. Here, more than seventy individual machines handle photonic chips, usually more than 300 times per chip. Still, these machines are not perfect, and complexity emerges in the form of machinery breakdowns, production invalidation, error messaging, or human decision-making. In the highly automated world of serviced machinery, especially in semiconductor manufacturing, Big Data Analytics enables plant-wide data utilization turned into actionable insights (Siddiqui et al., 2024; Wright et al., 2024; Rath et al., 2025).

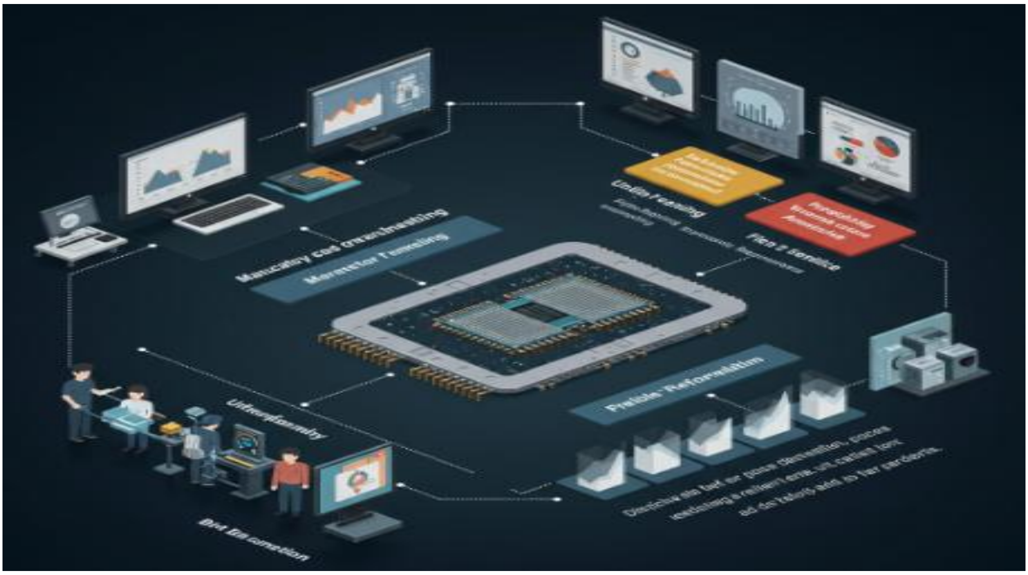

Fig 8.1: Big Data Analytics to Improve Semiconductor Yield, Reliability, and Predictive Maintenance

8.1.1. Background and Significance

Intelligent manufacturing focuses on time-based operation, cost control, quality assurance, and responsive production. Complex manufacturing processes generate big data. The exploration of process big data can enhance operational efficiency and reduce cycle time in semiconductor manufacturing processes. Big data in semiconductor manufacturing includes wafer characteristics, metrology, sensor data, and raw test data from manufacturing process execution. The challenge of big data is to extract useful information and knowledge stored in the data to provide new understanding and insights into current operations. Big data analytics is essential to develop intelligent manufacturing strategies and solutions in semiconductor manufacturing.

Rapid technology evolution in logic processing, memory, and advanced packaging integration led to fundamental changes in the semiconductor industry. Important technological innovations include advanced lithography, deposition, fusion bonding, and etching processes. These technologies result in the biggest chips with enormous capability and integration. However, rapid evolution leads to more complex manufacturing processes, which reduce the process yield rate and intrinsic quality of products. Further R&D investment in new materials costs a great deal. Big data analytics is developed to better extract useful information from big data collected in the production and explore knowledge to enhance operational efficiency in the manufacturing processes.

Since the launch of the smart factory and the Industry 4.0 initiative, big data analytics using the Internet of Things has been widely investigated in manufacturing engineering. Overall data analytics solutions are designed to develop intelligent manufacturing strategies in manufacturing scheduling, equipment maintenance, and production planning. There exist some papers focusing on semiconductor manufacturing big data analytics approaches to extract useful knowledge from process data for better understanding and enhancement of operational efficiency in semiconductor manufacturing processes. The challenges and opportunities of big data analytics in semiconductor manufacturing processes remain a new, open research area.

8.2. Understanding Big Data in Semiconductor Industry

Big data concept in the semiconductor manufacturing industry is viewed differently from various perspectives. The big data and analytics framework for smart factory defines big data as the data that exceed the storage, access, and analytics capacities of traditional hardware and software, including data that are large in volume, complex in structure, and high velocity. Examples of big data in semiconductor fabs are in Wafer-Level Reliability (WLR). With the faster and smarter devices, the test time for 4G devices has been shortened to 20 seconds to avoid yield loss due to the thermal runaway. Battery leakage failures become prominent among the WLR typical failures which require a longer test time of 60 seconds to guarantee device reliability. The new ATEs with a truly parallel testing architecture can support 200 devices per second testing, 16,000 I/O high speed pins, and memory capacity of 32 GB per W/LNTA. In smart manufacturing, data is the core competency. In semiconductor processes, different types of sensors are deployed which generate data in terabytes every day. Data generated in equipment, process chambers, and FDC systems are heterogeneous in structured and unstructured format. Data generated may not be suitable for immediate analysis. Moreover, it is important to select profitable data features for result prediction while data is big. Therefore, a systematic approach matching big data analytics technologies is desired for task-orienting semiconductor big data analytics.

The manufacturing industry has been revolutionized by the widespread adoption of big data technologies for extracting and leveraging manufacturing big data. This focus area emphasizes the significance of adopting and utilizing big data in the manufacturing industry and explores the opportunities and challenges brought by big data to the manufacturing industry. Specifically, it summarizes recent advances and research efforts in facilitating big data collection, storage, query, processing, analysis, visualization, and knowledge discovery. Also, key enabling technologies such as big data analytics architecture, big data processing, big data analytic models, algorithms/techniques, and big data services in WLCSP assembly are reviewed. In addition to an overview of big data analytics in the manufacturing industry, research challenges and directions to exploit big data for the intelligence of the manufacturing system and the manufacturing process are discussed.

8.2.1. Research design

Knowledge generated in the BA processes cannot be reused. Thus, separating data processing from knowledge generation to enable reusability of the big data analytics processes is essential.

An exploratory research process was employed to gain deeper knowledge of key capabilities regarding expert knowledge communities in semiconductor manufacturing. The research was contextualized in the case company's context of mature products, wherein some big data analytics processes had already been implemented prior to the data-analytics program. Furthermore, existing knowledge infrastructures based on the company’s digital twin concept were utilized to mitigate the inherent challenges of representing the complex domain knowledge of semiconductor manufacturing accurately and comprehensively. The research design entailed multiple case studies focusing on how the BA processes were separated from the modeling and analysis. Moreover, the semi-structured interview-based data collection process was designed collaboratively with the company to enhance the relevance and validity of the data. Knowledge communities were identified to receive structured data from the building blocks of the company’s digital twin concept, and the BA processes applicable to the context were derived from research on data preparation and processing. This led to the production of a design proposal in the form of a matrix depicting the linkages between key components of the company’s digital twin concept, the knowledge communities, and the BA processes.

8.3. Data Collection Techniques

Data logging has become a standard procedure in almost all industries. In semiconductor manufacturing, production data is one of the crucial elements allowing the process feedback loop to function effectively. Ideally, it is also a service to customers who can be provided with trace data like production log and electrical test results are the lagging indicators affecting product quality. Manufacturing data logging and collection rates are increased by advances in high-speed communication and data storage technologies. At the same time, data mining and statistical analysis programs are also developed to explore and analyze this abundant collection of data. The knowledge banks are continually enriched with new data being included and compiled on a 24-hour basis. It has also been recognized that only a small part of the knowledge in the banks is extracted and analyzed with commonly used data mining techniques such as statistical process control and regression analysis. It is estimated that about 90% of the process-related knowledge is still buried in the database. This ratio is even worse for large data banks where such as one hundred queries return more than one hundred thousand records and 0.1% of the data can be analyzed with reasonable effort and time using currently available tools.

To target manufacturing quality improvement, it is necessary to divert some resources to analyze the accumulated knowledge banks. A quality index is required to assess the process capability of the manufacturing equipment. Performance monitoring charts are required to ensure that equipment performance is within a set limit over a period and warnings are necessary regarding degradation before a process failure actually occurs. There are, however, some unique characteristics of the semiconductor manufacturing processes which impose constraints on the development of above-mentioned tools. Regardless of how sophisticated the analysis technique is, it is still inevitable that during this process, the knowledge about previously unrecognized issues is buried. In addition, the objective is to improve product quality and discard defective devices in the end. Thus the health indication derived should reflect the machine's influences on wafer and product quality. In addition, during the wafer fabrication, the mask process also introduces some capability in differentiating and identifying this process-related failure/reliability issue. Thus in turn, the tool is required to detect and identify systematically defects that have been flagged during the electrical failing test and customer returns.

8.3.1. Sensors and IoT Devices

Big Data Analytics, big data, internet of things, IoT As semiconductor manufacturing becomes more complex, the volume, variety, and velocity of data generated from manufacturing equipment are rapidly increasing. Smart manufacturing with big data analytics is regarded as the best practice for deriving business values from the high-volume, high-speed data generated in manufacturing factories. With the emergence of the IoT, a network of smart connected devices, manufacturing equipment have been equipped with an array of sensors, which measure a variety of variables including temperatures, pressures, voltages, currents, flow rates, and so forth. In addition, there are many other devices such as OPC servers, aligners, probes, manipulators, stockers, shipping carts, etc., that become available with sensors and other types of smart modules. These machines are used in manufacturing processes, quality inspection, factory automation, and material handling. IoT devices measure variables associated with their operations, including temperatures, pressures, voltages, currents, flow rates, state indicators, etc. As a first step toward smart data-driven manufacturing, it has become crucial to establish an IoT-enabled smart manufacturing environment in any modern manufacturing industry, where production equipment, inspection systems, and logistics systems are integrated to monitor the state of the factory with detailed information.

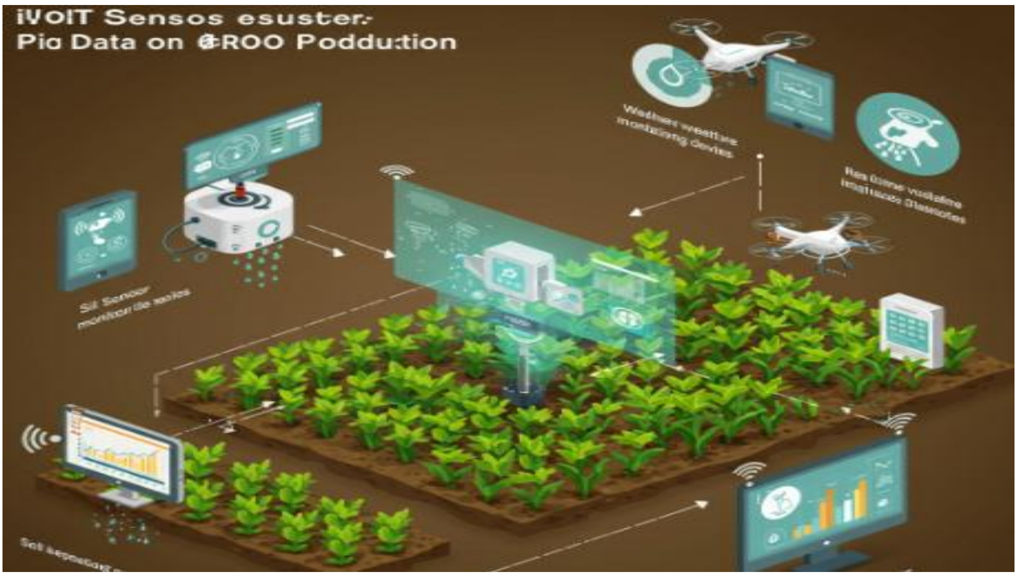

Fig 8.2: IoT Sensors and Big Data to Improve Precision Crop Production

8.3.2. Data Acquisition Systems

Data acquisition systems (DASs) can classify as a type of computer-aided, data-driven system employing automated means for collection of process data from production, manufacturing, and assembly systems. Process data is generated as a by-product of production, but such data must be stored into systematic databases before they can be utilized for managerial and engineering purposes. In principle, the major tasks performed by a DAS includes data acquisition (softwares acquiring data from distributed intelligent sensors), data storage (historically functional data warehouse), data pre-processing (cleaning, filtering, and transformation of raw data into specified identified data formats), data aggregation (historically functional MSc fields), data backup systems (ensuring historical local and remote backup of database contents without data loss), secured access and usage (ensuring security by administration of data usage privileges), and on-line/offline status monitoring (ensuring health monitoring of critical units in runtime mode).

Data collection from machine tools and processes has been widely studied and commercialized. In semiconductor equipment, some data acquisition systems have been deployed around the globe. However, improvements in terms of smarter algorithms, efficient data analysis methods, and investigation of smarter collection mean-ings in high-volume data generation areas would still be needed. Meanwhile, data processing, data warehousing, and applied data mining fields are very mature in both research and industrial applications. Plenty of software tools that focus on general emphasis and/or specific areas have been developed and utilized. Most semiconductor factories also own server farms or cloud storage systems for large amounts of data storage and processing power requirements. However, since human-machine-enforced intelligent processes have not been widely applied yet, improvement opportunities still exist around the globe.

8.4. Data Processing and Management

Data processing and management in semiconductor manufacturing are labor intensive. Traditionally, data processing and preparation involve steps including importing raw data from multiple measurement tools, quality checking by human annotators, feature extraction, and writing model training data. This process is time consuming, as inspecting a whole wafer lot containing 100 slices may take days. Each slice must fulfill quality checks on all its range of measurements to see whether they meet the pass and fail criteria. A half-inch-wide data science tool has to be developed for feature handling, as data formatting varies across tools, making unified formatting a challenge. Semiconductor manufacturing relies on customized measurement tools and processes. Each supplier is responsible for its tools' design and controlling measurements. This arrangement leads to data standardization issues, as the same kind of feature is computed in different ways in different tools. Because of an environment requirement that involves air humidity and temperature control, equipment malfunctions in the production environment cannot be preemptively flagged. Problems are communicated through a textbook-sized human inspection report, and solutions are implemented via a lengthier revisited-measurement process. External data rich in time series of stalking processes and environmental fluctuation, such as air humidity and temperature, chemical concentrations, and maintenance record, have put pressure on computational units. Attempts to speed up the analysis path have not yielded the expected result due to the linearity of the steps involved. If the explanation generation takes ten days for a worst-case scenario, model training on new data will take at least another ten days. It is heavily reliant not only on prior step processing but also on the vendor side who holds the data. Moreover, unsolicited outliers are returned, as representative outliers will trigger engineering trials to analyze and calculate the production cost. A novel data processing method is needed to integrate on-device preprocessing into the current workflow.

With greater scrutiny of the air and battery quality and increasing customer demands for battery lifetime, rapid-pulse current cycling analysis is becoming increasingly important for batteries in daily monitoring. In a pulse current measurement, pulse ing is done through rapid switches between two voltage ranges; the lower range has a greater resolution but a lesser current range, while the higher range has a wider area but makes it harder to detect faults. With a battery pack, in which each cell is different, a better solution must be tried on-line to accommodate the variability in these aspects. Any responses like a flat current but a distorted voltage signal shift in phase from the reference data will trigger a data science tool to judge the health state. A combination of an expert-designed heuristic algorithm for wholesale data screening and an anomaly detection algorithm for scanty anomalies has revealed potential struggles in regards to the time complexity and handling method. Proposed real-time feedback on pulse smoothness failure has been modeled using hidden markov chains.

8.4.1. Data Cleaning and Preprocessing

Despite advances in streaming analytics algorithms, a significant portion of the effort involved in advanced process control, process analytics, and machine learning involves acquiring and preparing data. However, when industrial case studies are published they often lack important details on data acquisition and preparation. This is unfortunate as while data pre-processing is unfairly maligned as trivial, in practice it has an out-sized influence on the success of real-world artificial intelligence applications. This text describes best practices for acquiring and preparing operating data to pursue data-driven modelling and control opportunities in industrial processes. Practical considerations for pre-processing industrial time series data to inform the efficient development of reliable soft sensors that provide valuable process insights are presented. The analytics landscape is rapidly evolving with an increasing interest in extracting actionable insights from process data. Advances in communication, sensor, and storage technologies enable the collection of vast amounts of monitoring and operational data, but sophisticated computing infrastructure to extract insights from this big data remains scarce. Automation facilities are attempting to capitalize on the opportunities offered by big data analytics but face numerous challenges managing and analysing the wealth of process data they hold. The extremely high dimensionality of process data presents a significant barrier in pursuing data-driven modelling and control methodologies. Because of the historically exclusive reliance on mechanistic modelling, modelling needs are often highly vague and ill-defined. Inadequate understanding of the available analytics tools often results in inappropriate use of tools, limiting the efficacy of data-driven analyses. Avoiding these pitfalls necessitates a clear and well-defined understanding of the desired outcome of the analysis, including clear metrics to assess quality and performance, as well as understanding the limitations of analytical methods used. Sewage treatment is a complex nonlinear biological process with uncertainties in the form of stochastic disturbances and outliers. It falls under the category of big data due to high volume, frequency, and dimensionality. Reinforcement learning-based data-driven control, under the framework of the Markov decision process, is a promising approach to address challenges thought to be impractical. Simulation-based metrics are proposed to discover stable control policies that are no longer degenerated and sample efficient, tremendously speeding up the learning process. Further performance improvement techniques to improve data efficiency, control performance, and interpretability, respectively, are combined to augment the application of the framework.

8.4.2. Data Storage Solutions

When deciding how to store data, several aspects need to be addressed. First, the various data that need to be stored in the system should be identified. Then the right types of data storage solutions should be chosen, such as relational or non-relational solutions. Finally, solutions to transfer data to the data storage should be established. This section starts by addressing the first two aspects, focusing on types of data and storage solutions. For the last aspect, architectures for transferring data to data storage transferred and transformed during an example of a disruptive event are described.

Various types of data are collected, analyzed, processed, reported, and stored in the use case systems. These include recipe, correlation, signature, parametric, capabilities, parameter value history, tool, probe, lot, update, updated, zone, constraint and sensitivity basis data. Several data entities, such as alarms, analysis results, and notifications of results being updated, are also used but are more reactive and ephemeral in nature. For some data types, e.g., alarm, there are numerous attributes, but most are not of interest in this section. There are essentially three levels of detail for most data types, which consist of summaries, single data point instances, and a history of all such instances.

For storing on relational systems, a star schema is identified for each type with various tables. Some of the tables are used to store general attributes common for all data type instances, whereas others are used to store type-specific attributes. In both types, several tables have surrogate primary keys and various foreign keys. A set of simple queries to report on contents of the star schemas is also produced. Further, level-1 reports that use the aggregate functions of the business intelligence tools are defined. They are aimed at displaying either the outcome of traversing through the result table or a desired version of a hexagonal plot.

8.5. Big Data Analytics Techniques

Supported by the growth of Information and Communication Technology (ICT) and the Internet of Things (IoT), big data analytics has advanced rapidly in many sectors while generating emerging opportunities, particularly in semiconductor manufacturing. With the construction of the fifth generation (5G) of cellular networks, a large quantity of sensing data is collected from nodes such as the terminal equipment and base station antennas. In these circuits, many parameters can be collected to model, evaluate, and quantify the performance. Semiconductor manufacturing and IoT products not only encompass manufacturing environments (factories, cleanrooms) equipped with tools for lithography, etching, deposition, and so forth) but also include foundries and design stages of the devices. A comprehensive review of previous works on big data analytics for key opportunities in semiconductor manufacturing is presented, including challenges in the process of converting raw data to highly valuable data-driven solutions.

Big data refers to large volumes of data, both structured and unstructured. Social media, sensors, and machine-to-machine data are a few examples of the growing volume of data that can be analyzed for insights. It is easy to define what is big data for an individual case, as it varies according to the data source industry, processes, and analytics requirements. However, generating valuable analytics from big data requires a robust analytics framework to include the relevant analytics sources. With the advent of the Third Platform Computing Systems (cloud, big data, mobile, or social), advanced production machinery and equipment, inventory control, scheduling, quality control, obsolescence forecasting, and supply chain management production activities require Data-Volume-Metrics (data acquisition, storage, processing, mining, presentation, and decision-making) based data analytics activities.

Descriptive analytics can be performed to analyze the reasons for these incidents and classify them by severity. It provides insights into which modules most frequently raise incidents and should consequently be prioritized for fixing. Prescriptive analytics could help identify the root causes behind severe and frequent ones and generate concrete recommendations. Support logs can be aggregated over the entire product lifecycle to build a top view of usage and cold facts about where to focus on.Finally, a novel and effective visual display viewed the predicted behaviour of the pins in several time horizons. The collaborative configuration between AI and human operators is also an important aspect that has not been widely explored. AI approach methodologies have so far been developed in a completely unilateral configuration with humans as passive and secluded recipients of AI productivity without space for interaction, feedback, and partnership. In this critical aspect, the proposed system bridges the gap between human and AI.

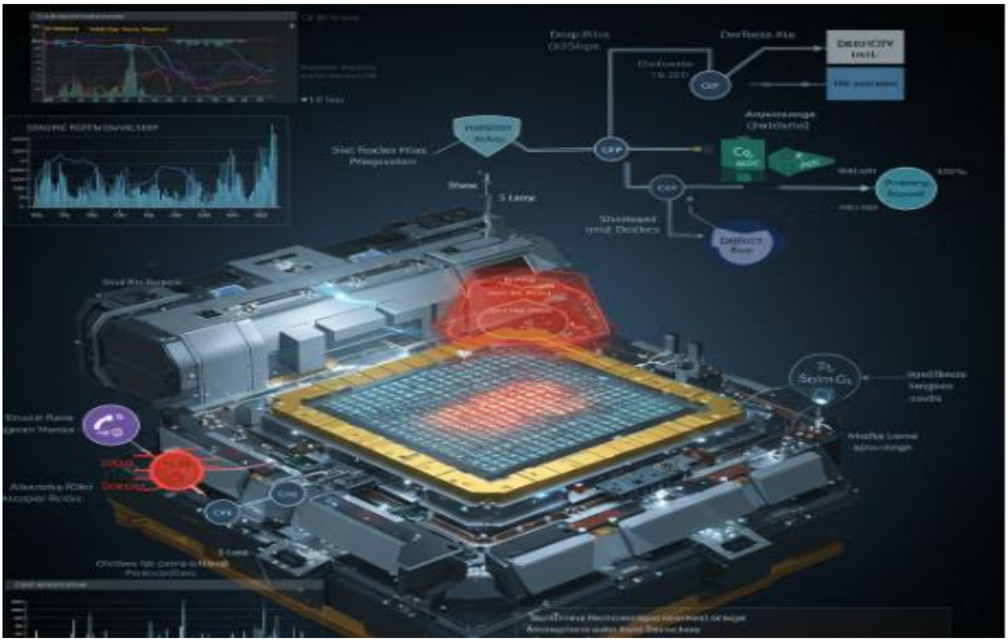

Fig : Semiconductor Process Anomaly Detection and Prediction

8.5.1. Descriptive Analytics

Big data analytics can be categorized into four general categories as shown below: descriptive analytics, diagnostic analytics, predictive analytics, and prescriptive analytics. Descriptive, diagnostic, and predictive analytics combine to solve the ‘what is/lived’ problem. The outputs of the analysis can then feed into prescriptive analysis to help answer the ‘what is/foreseen’ problem. Prescriptive analysis can also combine with scenarios built in the earlier stages of product management to provide input for strategic decision making.

Descriptive analytics aims to describe past data and events more clearly, which could also help identify trends and patterns that otherwise could go unnoticed and provide insights to fuel the other levels of analytics. It typically entails extracting, processing, and visualizing data of interest in dashboards and reports. In a software-intensive product, the core events of interest for the management can be usage-related (e.g., number and length of incidents in specific product modules), support-related (e.g., number and length of issues raised regarding a module), performance-related (e.g., product crashes), or demographic (e.g., geographical distribution of instance installations). A combination of usage and demographic analytics is used to prioritize the product modules to be changed in the next major release. Collaborative visualization tools can track product usage with respect to business processes.

8.5.2. Predictive Analytics

In the semiconductor industry, simulation processes can be affected by multiple uncertainties that directly affect all obtained results. For this reason, predictive analytics systems are used in semiconductor manufacturing factories to forecast next and future results from several input parameters and conditions. The predictive approach can be applied directly on one result or is modularized to allow explanation of the predicted behavior on the basis of the physical model. Predictive maintenance is mainly utilized in many maintenance operations in semiconductor manufacturing factories. As milling and laser repair operations are currently performed manually, knowledge-based systems have been built to analyze, in real-time, the history and status of head pins in equipment. Process data, manufacturing data, and machine health data are gathered and analyzed in a cloud-based architecture. The analytics phase detects potential problematic pins and generates a visual explanation view for the operator. The effectiveness of the proposed system has been demonstrated through its application to a real data scenario of a machine in a semiconductor manufacturing factory. A similar approach can be used in advance for other operations in the semiconductor industry based on their existing processes and available data. The proposed predictive maintenance system has proven its capability in assisting maintenance operations of semiconductor manufacturing equipment by gathering and analyzing real conditions of the equipment. A knowledge-based module adopted an AI approach using libraries in the data analytics. Business analytics were carried out in a cloud-based architecture, while data layers supported the deployment of the process.

8.6. Improving Semiconductor Yield

In the semiconductor industry, yield enhancement is a pivotal concern, with direct consequences for cost efficiency and market competitiveness. Wafer fabrication is the stage that incurs the most direct cost in the whole semiconductor manufacturing process, while enhancing yield has emerged as a pivotal lever to curtail costs and amplify financial returns. In advanced logic wafer fabrication facilities (fabs), a mere 1% increase in yield can translate to a substantial $150 million in additional estimated net profit, including both the on-wafer profit and the overall fab operation profit including material, equipment, and labor cost. Moreover, a yield decrease of 1% from a 90% yield may even result in an estimated $370 million in 1-year profit loss. Recognizing the importance of yield enhancement, the semiconductor industry is building and expanding robust and comprehensive yield enhancement strategies involving statistical analysis, data analysis, hardware debugging, and pattern yield learning, all of which are mostly manual processes. Consequently, much effort is dedicated to gathering and curating process and inspection data from a plethora of industrial data sources to unveil underlying defects in silicon wafers and facilitate yield enhancement measures.

Machine learning (ML) has been increasingly employed to augment these established yield enhancement strategies. ML techniques hold the potential to release engineers from tedious and repetitive daily routine analysis, provide insight into the new yield enhancement measures, and enhance the overall yield enhancement paradigm. However, in terms of statistical data modeling, ML techniques are externally opaque such that they are typically not interpretable or explainable even though the comprehension of critical data is imperative for process optimization in semiconductor manufacturing. On the other hand, despite the potential of these ML techniques, their development and deployment typically require extensive expertise in both semiconductor manufacturing and algorithm development, presenting a considerable barrier for rapid and agile integration and responsiveness in semiconductor smart manufacturing (SSM). Consequently, there is a perennial quest in SSM for easily implementable and interpretable data modeling measures to quickly adapt to the changing of things, improve the yield of semiconductor products, and optimize the utilization efficiency of precious resources.

8.6.1. Yield Analysis Techniques

The yield analysis can be classified into In-Line Yield Analysis and End-of-Line Yield Analysis. In-Line Yield Analysis investigates the root causes of yield loss by taking various pre-dispensed data collected from the semiconductor manufacturing process, to identify and determine the possible defect types which degrade or destroy the yield. End-of-Line Yield Analysis, on the other hand, refers to the post-silicon activities after wafer fabrication and test. The performance data collected during the wafer test sometimes contains a large number of defects so that it is infeasible to analyze them one by one. Instead, grouping the defects according to their characteristics would help understanding and isolating the yield loss.

The defect cluster recognition systems previously worked on the wafer test result of packaged devices and semi-finished chips. Results were aggregated on mold lines in packaged devices and functional test results were tabled on the probe station for semiconductor wafer testing. They are much more detailed than current study and should illustrate higher resolution for dependability tests, however in semiconductor manufacturing, the activities are far more complicated. There are a huge number of different data such as alarming logs, process wafer in-transit data and inspection results. Therefore, yield analysis is more complex for semiconductor manufacturing due to diversified defects and therefore higher uncertainties.

The defects are diverse and introduced in any part of the semiconductor manufacturing process. However, all embedded defects were masked by patterns and shunted as functionality off-chip IC's after wafer fabrication. This requires statistical process capability monitoring of both the fab and the ODT to identify problems. Masking defects concentrate in a certain area on the wafer to a pattern and there exists a bitmap to describe them. The bitmap can be used to simplify the recognition of a group of defects. Pattern matching and image processing methods are currently used for recognizing defects. However, it is time-consuming if internecine mask shapes are used. The aim to extract the depicted bitmap on the wafer is to compute this from the test data.

8.6.2. Root Cause Analysis

In semiconductor wafer fabrication, the production data is singled out as one of the crucial elements in aiding the process feedback loop. The tightly coupled interdependence and multidimensionality of the fabrication processes require continuous monitoring, analysis, and modification. Data logging has become a commercially feasible task and is now a standard procedure in virtually all high-volume manufacturing-based industries. The data logging ensures that virtually all industrial process-related data, be it raw, modified, or newly derived data, are always ready for extraction and analysis. There is a lot of knowledge buried in the huge data collection, which is waiting to be discovered. The knowledge banks are continually enriched with new data. Hence applying new appropriate analysis tools can further improve semiconductor manufacturing.

The production of microelectronic devices has an important feature that makes it significantly different from other manufacturing processes. The semiconductor wafer fabrication is a batch manufacturing of a multitude of Integrated Circuits (ICs) produced simultaneously in a multitude of sequential fabrication steps on a single piece of silicon substrate. The pieces of silicon substrate are known as wafers. In addition to the traditional analog processing unit, the addition of more and more digital and radio frequency processing units makes the ICs more elaborate in design and fabrication. Since the fabrication process is quite complicated, there are many possible sources of yield loss, including faulty equipment, faulty process parameters, and human errors. Cleaning up the small isolated defects before they can propagate into major faults is the goal of the current defect disposition process. Many faults in a fabricated wafer are automatically detected. These defects are classified based on their image patterns (shape, size, and texture). The probability that the a priori judgment of a defect being bad given it is classified as a certain type is calculated and used as the order of magnitude to execute the fault detection and diagnosis process.

Defect clusters are defined as defects that are identified as belonging to the same defect type, and that are located near each other on the wafer. In other words, if two defects are classified as belonging to the same defect type, and if the difference in positions of the two defects on the wafer is smaller than a predefined threshold, then the two defects are said to form a defect cluster. It is a known phenomenon that under normal and stable working conditions, out-of-specification defects tend to cluster due to a common process-related fault. The more closely a defect is grouped with those that already possess a defect cluster label, the higher is the chance that the defect label can propagate through the defect clustering process.

8.7. Conclusion

Recent rapid technological development, active promotion, and extensive application of AI technologies on chip methodology, logic, process and EDA tools, an upgrade framework has formed and been followed by adoption across semiconductor design, manufacturing and supply chain, demonstrating applications of chip deep learning yield prediction, die yield and quality prediction, chip cost prediction, device lifetime prediction, hot spot prediction, pattern discovery, lithography simulation, circuit parasitic parameters extraction, yield ramp-up evaluation, and yield improvement, etc. However, there still exist deep challenges and major difficulties to be overcome across semiconductor chip manufacture, assembly and test before large-scale deployment in chip factories, including: pruning and efficient inference of deep network, graph convolutional deep model for heterogeneous chip capacity prediction, failure classification and root cause analysis of die FA-test, active learning for hot spot discovery, cross-domain pattern mosaic and interpretation, rapid training of simulation-supported computation model, knowledge-embedded explainable AI approaches for AIOL, etc. To address different challenges and extend attention to sophisticated AI techniques, extensive development and deep collaboration across academia and industry are anticipated to be played.

Real-time big data analytics with AI-enhanced deep learning and edge computing are key enabling technologies to smart manufacturing and smart factory 4.0, with efficient modeling and mining of peer-to-peer spatiotemporal high-dimensional data of high throughput and data velocity in 5G time-sensitive semiconductor trade with ubiquitous manufacturing nodes. The structure and components of intelligent edge-cloud big data analytics and mining architecture with novel AI time series models for machine-level big data mining are introduced first. Then, a comprehensive framework of AI-enabled big data analytics systems for cross-location multiple users with diverse techniques, tools, engine and domain knowledge exploration is elaborated for scenario-based predictive modeling and actionable intelligence, supported with different example cases, along with research discussion on scientific challenges and trends.

8.7.1. Future Trends

Detailed analyses of data generated by manufacturing companies in the semiconductor industry have grown enormously in recent years. Several factors are impacting this increase and are spurring a demand for real-time, high-dimensional data and processes. As a result, the semiconductor industry is experiencing drastic technological changes and rapid advances in the tools and platforms to accommodate all the new data. Within 10 to 20 years, chipmakers and OEMs may not architect chips and their corresponding equipment but leave it all to machines working in harmony. The equipment deciding the capability of products will be selected and designed by themselves. It is a world where deterministic chips are fabricated based on a set of parameters similar to the way more than a dozen airlines now operate scheduling of thousands of aircrafts with no human intervention. Every company may start in-house manufacture but end up a pure IDM, fabless, or supplier of a single piece of equipment if they are unable to keep up the pace. Only future mega-cap tech companies branded under Big chips will remain if the balance is tipped towards integration. Others may evolve into net cash and then disappear altogether. Intelligent manufacturability offers a more complex system and architecture that demands a renewed understanding of physical, chemical, biological processes, and new quantum, molecular, atomic, etc. As manufacturing companies aggressively develop their capability intelligence, they face a discontinuity due to poor physics controls at the atomic and molecular levels. In addition to ultra-broadband optics, new ways may include shear-thickening fluid as self-sensing and adaptive pressurizing media; robotic arms under ultra-high vacuum that can move millions of gallons of heavy liquid per second; multi-layer assembly that can print nanostructured coatings that change optical properties in real time according to external stimulus; etc. As an addendum, while underestimating feedback and loop delay, the majority of today's equipment also cannot account for multiple types of correlations among signals and therefore may miss the boat entirely.

References

Parmar, T. (2021). Data-centric Approach to Decision Making in Semiconductor Manufacturing: Best Practices and Future Directions.

Kalusivalingam, A. K., Sharma, A., Patel, N., & Singh, V. (2022). Leveraging Reinforcement Learning and Predictive Analytics for Continuous Improvement in Smart Manufacturing. International Journal of AI and ML, 3(9).

Anang, A. N., & Chukwunweike, J. N. (2024). Leveraging Topological Data Analysis and AI for Advanced Manufacturing: Integrating Machine Learning and Automation for Predictive Maintenance and Process Optimization. Int. J. Comput. Appl. Technol. Res, 13, 27-39.

Rath, K. C., Mishra, D., Tripathy, S. K. T., Mishra, B. K., & Muduli, K. (2025). Potential of AI, Quantum Computing, and Semiconductor Technology Adoption in Future Industries: Scope, Challenges, and Opportunities. Integration of AI, Quantum Computing, and Semiconductor Technology, 415-440.

Siddiqui, A. H., VP, S., Chowdhary, H., Krishna, R. V. V., Muniyandy, E., & Maguluri, L. P. (2024). Harnessing Big Data: Strategic Insights for IT Management. International Journal of Advanced Computer Science & Applications, 15(7).

Wright, T., Tse, B., Nsiye, E., Azinord, T., Medina, D., & Mondesire, S. (2024, May). Probabilistic modeling and machine learning for preventative maintenance prediction in semiconductor manufacturing. In 2024 35th Annual SEMI Advanced Semiconductor Manufacturing Conference (ASMC) (pp. 1-6). IEEE.