Application Study: Smart Factory of Kioxia Utilizing AI

1. Smart Factory

Kioxia's flash memory is fabricated in a smart factory, which incorporates cutting-edge technologies. This smart factory has thousands of different manufacturing and inspection/metrology equipment, generating more than 3 billion data points every day (as of February 2025). The data is mainly sensor data, and inspection and measurement results, amassing 50 terabytes (TB) per day. To stabilize yield and output, it is essential to analyze the generated data. However, the sheer volume of data exceeds what can be analyzed by humans. As a result, Kioxia has been utilizing AI technology for data analysis for many years to help identify equipment and processes issues, develop countermeasures for these problems, and control production to maximize output. Our factory continues to improve and evolve on a daily basis, thanks to our long track record, abundant production experience, and incorporation of the latest technologies.

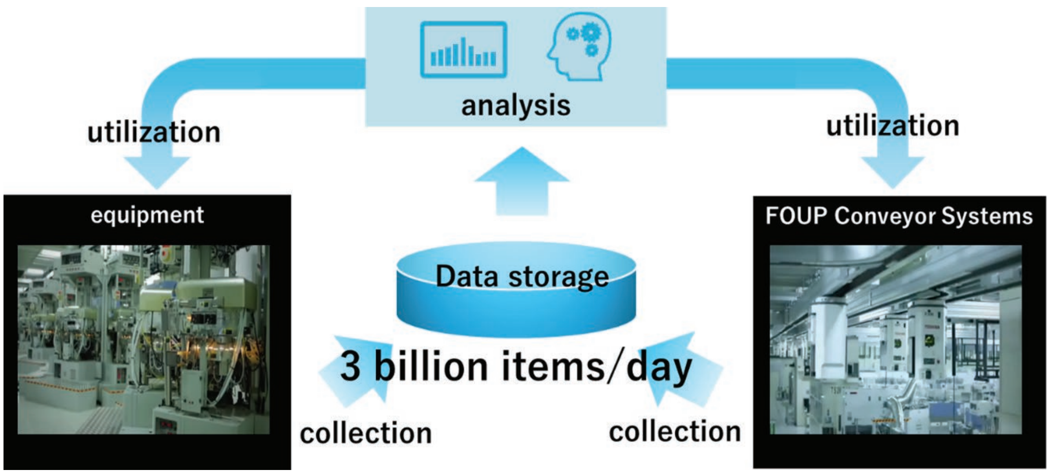

Figure 1. Smart factory of Kioxia utilizing ai

<Smart Factory> Kioxia Corporation Official Site
An autonomous intelligent factory that leverages AI and the Internet of Things (IoT)
https://www.kioxia.com/en-jp/about/yokkaichi/smart-factory.html

1-1. Big Data Infrastructure

The huge amounts of data generated by a smart factory cannot be processed by conventional database (DB) technology, therefore a dedicated big data infrastructure is being built. It consists of many powerful computers, and has distributed DB and distributed processing capabilities. It is based on cutting-edge big data software technologies, such as Apache Hadoop/Apache HBase/Apache Spark, an open-source software framework that supports distributed processing of large-scale data. This infrastructure has enabled the realization of many AI application benefits. In recent years, we have been studying the fusion of structured data, such as sensor data, and unstructured data, such as text and human knowledge, using generative AI and large language models (LLMs).

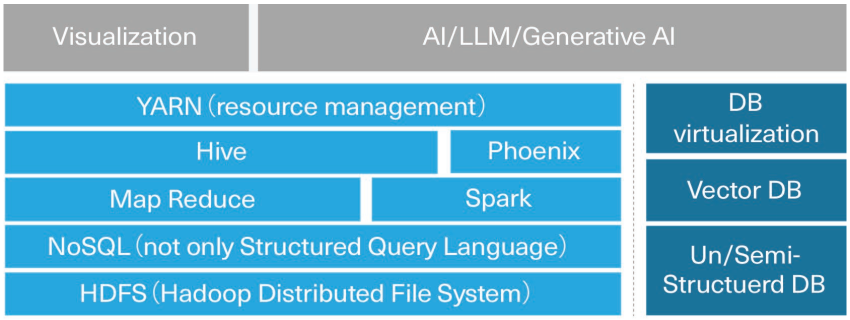

Figure 2. Configuration of big data infrastructure in smart factory

In addition, data from eight fabs, six at our Yokkaichi Plant and two at our Kitakami Plant, are incorporated in Kioxia's Big Data Infrastructure, making all the data appear virtually as one giant factory. Based on this data, we create a digital world, analyze the data using AI, and optimize and predict the future using simulations. The results are fed back into the physical world (real world) to improve productivity and quality.

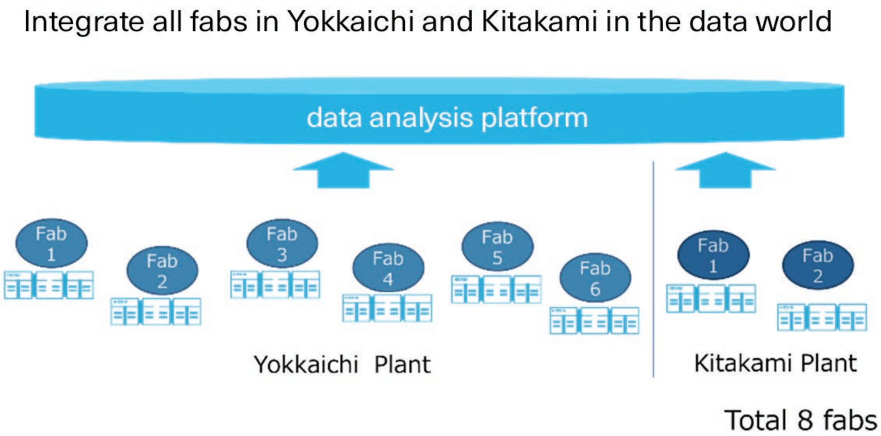

Figure 3. Data integration of 8 fabs

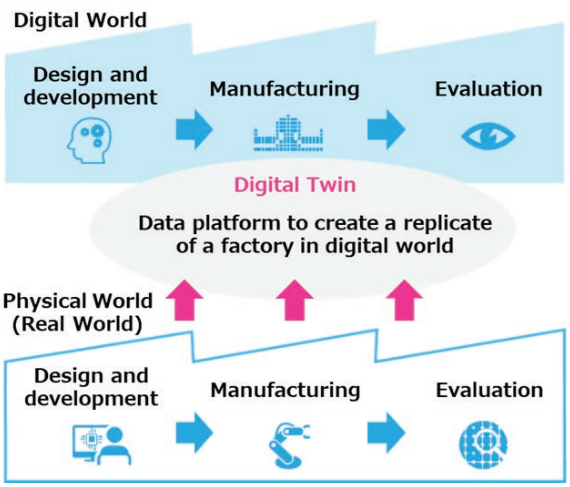

Figure 4. Replication of factories in digital world using big data infrastructure

1-2. Image Classification

In semiconductor manufacturing, many inspections and metrologies are performed to monitor the production process. Our AI system automatically analyzes inspection images and extracts useful information for quality control. In particular, image classification, which classifies inspection images as good or defective, or those deemed defective by defect mode, is one of the most important analyses performed.

For image classification, we employ the optimal method depending on the problem. In supervised classification, labels such as “good” or “defective” are assigned to inspection images in advance, and the AI learns the labels and images to create a model. Although the process of creating the model is labor intensive, it is useful for reliable and stable monitoring. Unsupervised classification, on the other hand, requires no prior labeling of images; instead, the AI identifies the features of the images and classifies them. This makes it possible to handle the frequent model updates that occur during the process development phase. Advances in AI technology have greatly improved the accuracy of unsupervised classification in recent years, expanding the options for problem solving.

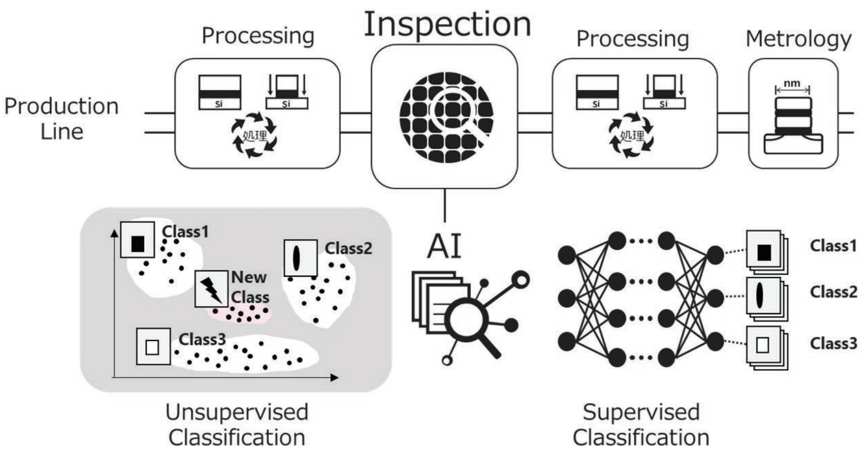

Figure 5. Automatic classification of inspection images utilizing AI

1-3. Failure sign detection

Most manufacturing equipment uses dry pumps to create a vacuum in the processing chamber. Sudden failure of a dry pump not only interrupts the manufacturing process, but also allows gas to flow back into the vacuum chamber from the dry pump, negatively affecting the product. Therefore, it is important to monitor the large number of dry pumps in a factory to predict failures. In collaboration with Mie University, we are working on the development of failure prediction technology.

In order to detect signs of dry pump failure, acceleration sensors are attached to dry pumps to extract features from the data. The feature values are then evaluated using two statistics that reflect the long-term deterioration and short-term abnormality of the pump. As a result, we have confirmed that we can detect signs of failure with high accuracy at the stage immediately before it fails. We are continually working on the development of our failure sign detection technology to enable detection at an even earlier stage.

Figure 6. Failure sign detection of dry pump utilizing AI

2. Process informatics

Process informatics is a technology that uses AI to optimize the manufacturing process. 3D flash memory has been increasing in number of layers and shrinking horizontally as its capacity increases, making semiconductor process development more difficult and extending the time required for development. In semiconductor process development, operating conditions of the manufacturing equipment, process conditions, are optimized to obtain the desired processing dimension. In addition, improvements in the design of the processing chambers for manufacturing equipment may also be necessary. We are promoting the use of AI not only to shorten the time required for these optimizations, but also to achieve optimizations that that are beyond human capability.

2-1. Reduction of number of experiments in process condition optimization

We utilize Bayesian optimization to optimize process conditions with a small number of experiments.

When Bayesian optimization is utilized in process optimization, conditions that are highly likely to produce optimal conditions are presented based on the results of previous trials. By repeating the trials of the proposed conditions, the optimal conditions can be reached quickly. To further reduce the number of trials in Bayesian optimization, a novel initial experimental design method has been developed. In this initial experimental design, a small number of experimental points are uniformly placed in a high-dimensional space with multiple experimental parameters. The number of experimental trials can be significantly reduced by conducting initial experiments under the conditions derived by this initial experimental design method, followed by condition search using Bayesian optimization. This initial experimental design method was developed in collaboration with the University of Tokyo.

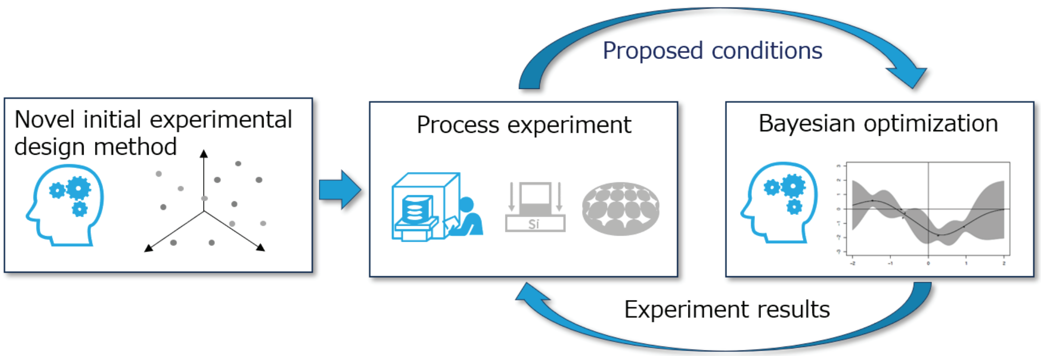

Figure 7. Bayesian optimization for process condition search and introduction of a suitable initial experimental design method

2-2. Optimization of Processing Chamber Design for Manufacturing Equipment

When it is difficult to obtain the desired fabricated shape by process condition optimization alone, optimization of the processing chamber design becomes necessary. In conventional methods, design optimization is often conducted through experiments using actual wafers, which can be time-consuming and costly. Since engineers set and consider the conditions, they are dependent on their experience and intuition, which is inefficient in many cases. To address this issue, we developed a method to automatically optimize the design of processing chambers using simulations of fluid behavior inside the chambers and Bayesian optimization. In this method, Bayesian optimization is performed using not only numerical information obtained from fluid simulations, but also features of contour plots in the fluid simulations to optimize the processing chamber design. This enables us to optimize the design of the processing chamber in a short time and at a low cost, without the need for experiments using actual wafers.

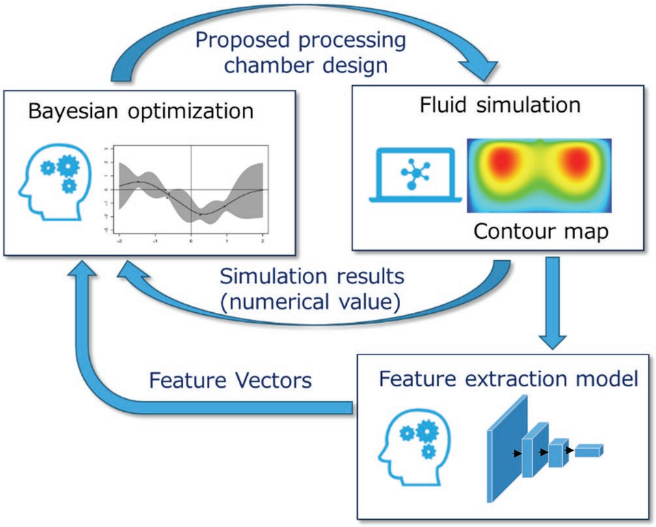

Figure 8. Optimization of processing chamber design through fluid flow simulation and Bayesian optimization

NOTES:

Definition of capacity: KIOXIA Corporation defines a megabyte (MB) as 1,000,000 bytes, a gigabyte (GB) as 1,000,000,000 bytes and a terabyte (TB) as 1,000,000,000,000 bytes. A computer operating system, however, reports storage capacity using powers of 2 for the definition of 1 GB = 2^30 = 1,073,741,824 bytes and therefore shows less storage capacity. Available storage capacity (including examples of various media files) will vary based on file size, formatting, settings, software and operating system, and/or pre-installed software applications, or media content. Actual formatted capacity may vary.

TRADEMARKS:

Apache Hadoop, Apache HBase, and Apache Spark are trademarks or registered trademarks of Apache Software Foundation in the U.S.A and other countries.
Other company names, product names, and service names may be trademarks of third-party companies.