White Paper
Manufacturing
December 2023

IT@Intel: Smart Manufacturing Using Computer Vision and AI for Inline Inspection

To quickly detect micron-level defects introduced during additive processes, Intel IT is deploying an integrated Intel® hardware- and software-based solution that can significantly reduce scrap and business risk

Intel IT Authors

Wu Jia
Senior Solution Integrator

Jun J. Liu
Software Product Manager

Qiangbin Wang
Senior Staff Engineer

Can Wei
Senior Equipment Engineer

Table of Contents

Executive Summary

Keeping Intel's global network of factories running at top efficiency is one of Intel IT's primary roles. Over the last decade, we have introduced numerous automated, artificial intelligence (AI)-based solutions that significantly reduce waste and improve product quality. Recently, we added a new capability using computer vision (CV) and machine learning that enables us to perform inline inspection during assembly and test processes. Inline inspection can catch defects and tool issues much earlier than offline inspection. Additionally, our solution can easily detect micrometer-sized defects that are hard to see even with a microscope, as well as defects that are impossible to detect after the process is complete.

High-resolution cameras take multiple images per second while a grinding tool thins the wafer and a protective polyester film is installed. The images are analyzed by a machine-learning model at the edge. If defects are detected, the solution can sound an alarm or even stop the tool.

The solution, which includes Intel® Core™ i9 processors, Intel® Xeon® Scalable processors and Intel® ARC A770 discrete GPUs, is being deployed to all Intel assembly and test factories. We anticipate that the solution will eventually extend to additional processes, and we are excited to share our success with other manufacturers, who may be able to adapt the solution to additional use cases.

Contributors

Cindy Jiang, IT Director, Intel IT
Jennifer Yuk Ling Lo, Industry Engagement Manager, Intel IT
Zhijing Wu, Account Executive, Intel Sales
Jin Xia, Engineering Manager, Intel Manufacturing
Zhen Xiao, IT Manager, Intel IT
Ying Yang, Engineering Director, Intel Manufacturing
Xiang Zhou, Technical Sales Specialist, Intel Sales

Acronyms

AI	artificial intelligence
CV	computer vision
IWVI	Intelligent Wafer Vision Inspection
PoC	Proof of Concept

Business Challenge

Automation driven by artificial intelligence (AI) and computer vision (CV) has been a critical component of quality control in Intel's fabrication factories (fabs) for over a decade. We collect images from multiple channels during fabrication and compute thousands of features to automate defect detection and classification. We also use AI and CV in our assembly and test factories to repurpose existing images and identify defects. However, there are ample opportunities in our assembly and test factories to extend the use of AI and CV. Early detection of defects can increase confidence in product quality and substantially reduce scrap, which can help lower costs. While Intel IT has always pursued solutions that increase factory efficiency and reliability of products, these goals are increasingly important as Intel ramps up its foundry services.

Assembling semiconductor products is a complex process. Several hundred of the steps are "additive processes" that add or remove material as the product progresses through the assembly line. When a silicon wafer first enters the assembly and test factory, it is relatively thick. The thickness of the wafer must be reduced by delicately grinding away material—especially for newer, smaller products that require the wafer to be extremely thin. After the wafer reaches the desired thinness, a polyester film is applied to the backside of the wafer for protection and stability.

Traditionally, quality inspections have been performed "offline" at certain intervals—usually after two or more additive processes are complete. For example, a certain percentage of wafers are inspected after wafer thinning and film application occurs. As a result, offline inspection (sometimes called offline metrology) presents several challenges:

- **High risk of scrap and escaped defects.** By the time inspection occurs for one wafer lot, up to nine more lots may have already been processed. If a machine or process error has introduced defects, it's likely that many more wafers will also be damaged, leading to high scrap and risk of lower-quality products.
- **Impeded inspections.** Because multiple additive processes occur before offline inspection, directly inspecting a wafer surface might be impossible. For example, the film application to the backside of the wafer—before inspection—prevents direct inspection for grinding errors.
- **Manual processes.** Inspectors use a microscope to inspect the wafers manually. But as products get smaller, it's harder for humans to find micron-level defects. For example, a grinding defect may be only 5 micrometers (µm) long. Comparatively, the entire wafer is 300 millimeters (mm) across—finding the defect is like noticing a grain of rice on a football field.

Guaranteeing quality as production ramps up would require a substantial increase in engineering resources; even so, they might not be able to keep up with the pace of production. In addition, because all new generations of Intel® products are transitioning to advanced packaging,[1] a single defect can cause significant scrap. Not only that, but on tiny products, circuit space is minimal; an escaped defect can potentially cause a critical failure at a customer site, potentially causing harm to the customer's business and to Intel's reputation for quality and reliability.

Intel IT needed a way to find defects quickly, efficiently and reliably. The wafer-thinning process seemed like a good candidate for an automated, CV-driven solution because the grinding tool and the application of the film can cause a variety of defects, some of which include the following:

- The robotic arm of the grinding tool could drop oil on the backside of the wafer.
- The grinding tool could chip or scratch the wafer.
- There could be a bubble in the film.
- There could be a multi-micron indentation on the backside of the wafer.

While these may sound inconsequential, in reality, each of these defects (and others) can potentially negatively impact the performance of the completed product.

Most Manufacturers Face Similar Challenges

Intel IT's computer vision solution for defect detection during assembly and testing can be easily adapted to different processes to reduce scrap and costs and enhance overall quality.

All modern manufactured products, ranging from space exploration rockets to tiny medical robots, have many components. It is common quality assurance practice to inspect a certain percentage of these components before assembling them into the final product. Artificial intelligence (AI) and computer vision (CV) represent a powerful combination that can increase the speed and accuracy of defect detection—freeing up engineers to work on higher-value projects as well as improving the overall quality assurance process.

Intel IT encourages manufacturers to start a conversation about how our CV solution works and how it can be incorporated into their assembly and test facilities.

[1] To learn more about advanced chip manufacturing, visit https://www.mckinsey.com/industries/semiconductors/our-insights/advanced-chip-packaging-how-manufacturers-can-play-to-win

Solution Overview

Intel IT conducted a nine-month proof of concept (PoC) for Intelligent Wafer Vision Inspection (IWVI) using AI and CV in two assembly and test facilities. The solution, based on Intel® hardware and software, represents an affordable approach to digitalization without significant impact on equipment and production pace. The PoC tested cameras and machine-learning models for enabling automated, inline inspection of wafers during the wafer-thinning process. The PoC was so successful that IWVI is now being deployed at Intel's factories worldwide as part of the plan of record. The solution significantly reduces the impact scope of a process problem, such as a bent arm on the grinding tool, and can even detect defects that are undetectable using offline inspection.

System Components

As illustrated in Figure 1, the solution consists of the following components:

- Several high-resolution cameras take images of the wafer backside and a top view of the placement of the polyester film.
- A machine-learning model analyzes images and detects defects.
- If a serious defect is detected, the model sounds an alarm and can automatically stop the grinding tool.
- Edge and high-performance computing (HPC) resources support high-speed image classification.

Because the automated solution can stop the grinding tool much sooner than is possible with offline inspection, we estimate that the solution can help Intel avoid at least USD 2 million worth of wafer scrap.

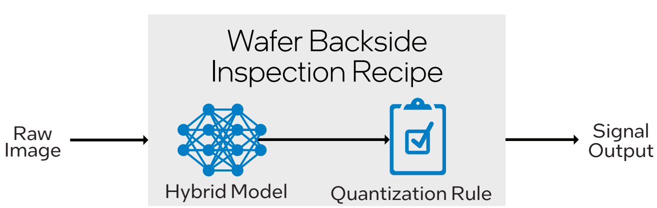

Figure 1. The machine-learning model takes thousands of images from the cameras and uses quantization methods to categorize defects and sound an alarm and/or stop the grinding tool.

Maintainability Considerations

Intel IT carefully evaluates technology deployments for maintainability, including what the sustained costs and effort will be. AI-related projects may have greater risk regarding maintainability due to the dependencies of data sources and changes in processes that could impact the AI model. We found that training and configuring the AI model that was used would enable us to factor in a reasonable degree of change that may occur in data or processes. The result is that this implementation is expected to provide value at minimal operational cost.

Solution Architecture

When exploring how to implement the IWVI solution, we faced several challenges that affected the solution's design. In particular, the solution must:

- Operate in a limited space without interfering with the grinding tool's operation.
- Require zero modification to the grinding tool.
- Be able to communicate with the grinding tool (such as to stop its operation).
- Process up to 20 GB of data per wafer, then submit the data to the machine-learning model for defect analysis and decision—all within the 90-second total wafer processing period.

Meeting these criteria required us to choose solution components carefully; our solution design is described in more detail in the subsequent sections.

Cameras

The vision box containing the cameras is mounted over the film application area to acquire images (see Figure 2). Three 16K time delay integration (TDI) line-scan cameras take images of the wafer backside (see Figure 3 on the following page for a more detailed diagram) and a single, 8K camera captures images of the film mount. The TDI cameras focus on micron-level defects; the defects detected by the mount view camera show larger "gross" defects. Every second, each camera acquires 16 images. The detectability range is 6 μm to 12 μm, and the scanning speed is 250 mm/second.

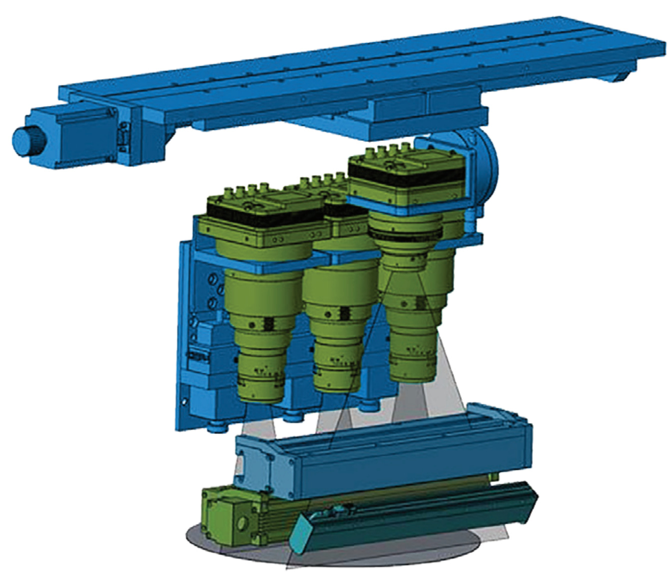

Figure 2. We use three 16K cameras for imaging the wafer backside and one 8K camera for imaging the film mount.

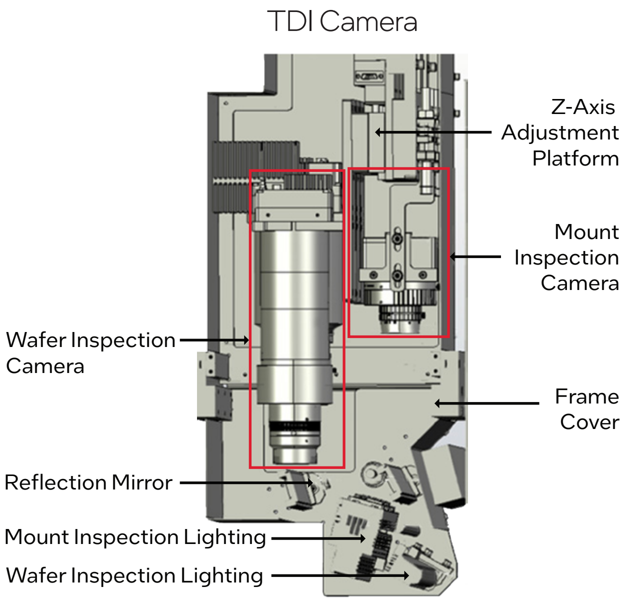

Figure 3. Detailed view of a time delay integration (TDI) camera.

The different resolutions of the backside and mount cameras and the type of light field used by each camera type are shown in Figure 4.

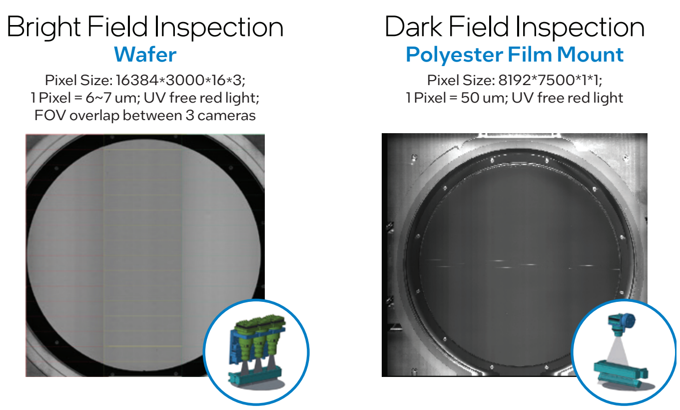

Figure 4. The wafer-backside cameras have a finer pixel size and use a bright field, whereas the film mount camera's pixel size is larger and uses a dark field.

Edge Compute

Every camera has an edge controller with a 12th Gen Intel® Core™ i9 processor and an Intel® ARC A770 discrete GPU for data processing acceleration. The data transfer speed is about 200 Gbit/second. The system can store about 40 TB (three weeks' worth) of raw image and inspection results.

HPC Resources

The machine-learning model resides on an HPC server equipped with an Intel® Xeon® Scalable processor and is located in Intel's private cloud. The server is used as a machine-learning training platform. We train our model on hundreds of thousands of labeled wafer images. The trained model is deployed to multiple edge locations in the production line. The model inference workloads at the edge (on the camera controllers) can be offloaded to the Intel ARC A770 discrete GPU for accelerated image analysis.

The Intel Xeon Scalable processor provides powerful computation to handle the heavy workload, expediting the training process. In our experience, the model trains faster than using other processors, which significantly improves the model iteration speed when we need to retrain the model. We retrained the model weekly during the development stage; now that it is deployed, we retrain quarterly. The model training process is illustrated in Figure 5.

Figure 5. Our solution spans the entire machine-learning lifecycle, from raw data acquisition and labeling to training and validation to deployment.

Software

The images are analyzed by the machine-learning model, which is a hybrid model consisting of more than 50 underlying core algorithms that include various functions such as inspection (building on ResNet50[2]), semantic clip segmentation (to enable defect identification at the pixel level), and identification and gauge (to auto-identify types of edges in the images and auto-measure distances). We optimized the model's efficiency on the Intel hardware platform using the Intel® Distribution of OpenVINO™ toolkit.

We also use an industrial vision control software cluster that includes camera imaging and equipment communication. The software provides an intuitive user interface for image review and offline simulation.

Scalability

All the software functions and algorithms are packaged as an integrated module that can extend the solution to new use cases. This integrated module offers low-code tooling, drag-and-drop programming and a flow-based wizard to enable easy scalability.

[2] ResNet50 (Residual Network, 50 layers) is a convolutional neural network (CNN). More information is available at https://huggingface.co/microsoft/resnet-50.

Results

The IWVI solution has been identified as a critical solution to help ensure the quality and reliability of new generations of products in Intel's factories of the future. Its standardized design results in fast deployment and easy extendibility to new use cases. The PoC demonstrated that the solution could accurately identify many types of defects in the wafer-thinning process, including indentations, scratches of various sizes, grinding marks, stains, cracks, bubbles, wafer shift and mount shift.

The PoC demonstrated that up to 50% of wafer-thinning issues can be detected earlier with inline inspection compared to offline inspection. When the IWVI alarm sounds, the wafer can be reworked, which avoids whole-wafer delamination during downstream processes. More importantly, the solution goes beyond what was possible with offline metrology:

- Detects excursions as they occur.
- Detects process defects and shuts down the tool quickly.
- Enables a new capability to inspect frame cleanness and inner ring.

Using the IWVI solution, we achieve several business benefits:

- Scrap avoidance, saving Intel up to USD 2 million each year.
- Reduced business risk.
- Higher product quality.
- Freeing engineers from tedious manual offline inspection.

Conclusion

The opportunities for continuing to increase factory efficiency and product quality using AI and CV are nearly endless. We can adapt the IWVI solution to other manufacturing processes in Intel's assembly and test factories, and we will continue to train the model for the wafer-thinning process to find other already-known defect types, such as wrinkles in the film mount. Such training could also reveal defect types of which we are unaware.

We anticipate sharing the IWVI solution as a reference design with other manufacturers because the solution is simple, affordable and flexible. For example, a similar solution could be used to inspect automotive welds, airplane paint jobs, ship hull rivets and more.

Related Content

If you liked this paper, you may also be interested in these related stories:

- Expanding Low-Cost Industrial Internet of Things (IIoT) Manufacturing Use Cases
- Intel Cuts Downtime and Costs with Fault Detection Systems for Factory Equipment
- Streamline Deep-Learning Integration into Auto Defect Classification
- Transforming Manufacturing Yield Analysis with AI
- Accelerated Analytics Drives Breakthroughs in Factory Equipment Availability
- Reliability Engineering Helps Intel Cut Manufacturing Downtime in Half
- Minimizing Manufacturing Data Management Costs
- Transforming Industrial Manufacturing with Software-Defined Networking

For more information on Intel IT best practices, visit www.intel.com/IT.

IT@Intel

We connect IT professionals with their IT peers inside Intel. Our IT department solves some of today's most demanding and complex technology issues, and we want to share these lessons directly with our fellow IT professionals in an open peer-to-peer forum.

Our goal is simple: improve efficiency throughout the organization and enhance the business value of IT investments.

Follow us and join the conversation on **X** or **LinkedIn**. Visit us today at **intel.com/IT** if you would like to learn more.