White Paper
Manufacturing | Artificial Intelligence

High-Value AI in Intel's Semiconductor Manufacturing Environment

With over two decades of experience in developing and applying artificial intelligence (AI) to specific use cases, Intel has established best practices that enable AI at scale to generate optimal business value

Authors

Rao Desineni
Senior Director, Manufacturing, Supply Chain and Operations

Eugene Tuv
Intel Fellow, Technology Development

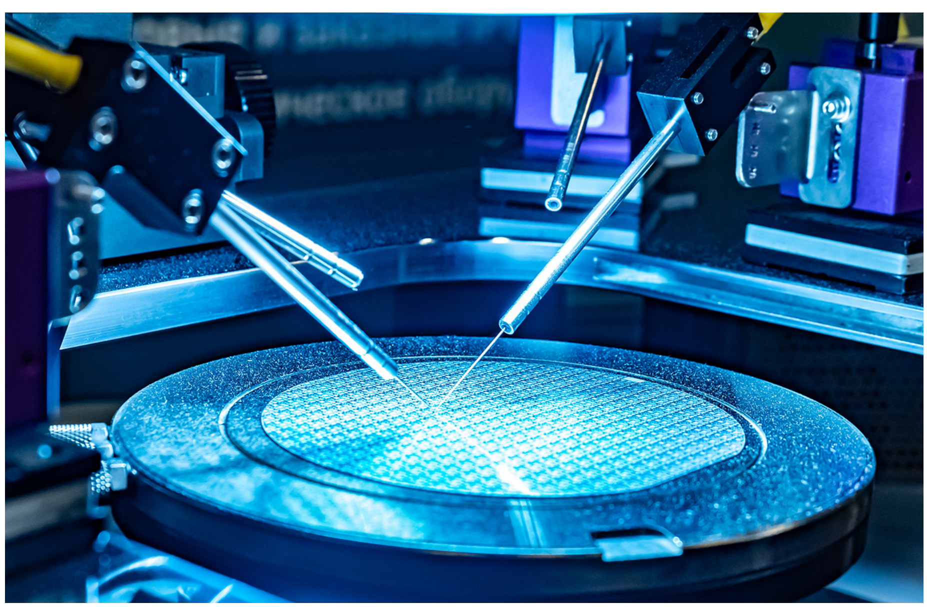
Table of Contents

Executive Summary 1
Introduction 2
Applying AI in Semiconductor Manufacturing: Solving the n-i, n Problem 2
Deploying AI at Scale in Intel's Manufacturing Environment. 2
Prioritizing AI Use Cases to Optimize Business Value. 3
Approaching AI with a Clear Purpose and Goal. 3
An AI Sampler: Three At-Scale AI Solutions Boosting Manufacturing Quality and Productivity 4
ADC with Computer Vision and Machine Learning. 4
Root-Cause Analysis 5
Sort Test Probe Card Inspection. 5
Conclusion 6
Further Reading. 6

Executive Summary

As one of the world's leading silicon manufacturers, Intel continually strives to increase product quality while improving manufacturing efficiency. Artificial intelligence (AI) is a powerful tool to transform vast amounts of manufacturing data into insights that can improve the manufacturing process. To maximize success, Intel has fine-tuned its approach to AI in manufacturing by focusing on applications that demonstrate substantial business value, practical feasibility, and quick time to value. Each solution solves a specific, well-defined problem with measurable success metrics. We've deployed a large range of manufacturing AI solutions involving thousands of AI models at scale over the last 20 years. Each successful AI solution is proliferated across all of Intel's factories for all products.

Applying AI to more use cases, in more places, for more businesses, through deep investments in an open AI platform is one of Intel's top missions. From inline defect detection to end-of-line yield analysis and many analytical steps in between, Intel's at-scale manufacturing AI solutions have generated millions of dollars in business value, accelerated manufacturing processes, and contributed to higher yield and productivity.



Introduction

Semiconductor manufacturing is complex, with tens of mask layers; hundreds of process steps; thousands of pieces of equipment, each with tens to hundreds of individual sensors; and tens to thousands of integrated circuits (ICs) per wafer. Each wafer contains billions of transistors and interconnect lines and hundreds of electrical and physical measurements per wafer, and tens of thousands of wafers are manufactured in multiple fabs every week.

This complexity leads to hundreds of petabytes of data during the manufacturing of products in advanced technologies. Semiconductor companies have always been leaders in generating and analyzing data. In the artificial intelligence (AI) spring that we currently find ourselves in, the obvious question is: Can AI be used in semiconductor manufacturing to gain insights from all this collected data?

Intel has been developing and using AI-enabling methods—machine learning, deep learning, computer vision and image processing, advanced multivariate statistics, operations research, and others—in various aspects of IC technology development (TD) and high-volume manufacturing (HVM) for almost two decades. The complexity of TD and the scale of HVM forced us to steadily replace rule-based systems with learning-based systems (aka AI) for use cases where it made sense. And the journey continues.

Applying AI in Semiconductor Manufacturing: Solving the n-i, n Problem

Semiconductor manufacturing comprises many steps—process, measurement, shipping, accounting, and others. Data is generated at each of the n steps that a silicon wafer or die must traverse before revenue is realized.

AI solutions in semiconductor manufacturing must include at least one of the following:

1. Detect a problem at step n.
2. If a problem is detected at step n, rapidly find its root cause using the data collected at the previous n-i steps, where 0< i ≥ n.
3. Predict the outcome at step n using the data from the previous n-i steps, where 0< i ≥ n, and devise control methods to optimize outcomes at step n.

Deploying AI at Scale in Intel's Manufacturing Environment

Some of the AI applications deployed in production in Intel factories include the following (also see Figure 1):

- Inline defect detection
- Tool/fleet/fab matching
- Multivariate process control
- Automated wafer map pattern detection and classification[1]
- Fast root-cause analysis (RCA)
- Detecting outliers at sort test for both test time reduction and quality improvements in downstream shipped products

Behind the scenes, multiple AI-enabling techniques, such as advanced statistics, machine learning, optimization, and various forms of computer vision are used, depending on the use case.

Once we develop an AI solution for a specific use case and validate its business value, we proliferate it across Intel's entire line of factories—thereby optimizing the return on investment for our efforts and promoting consistency between factories.

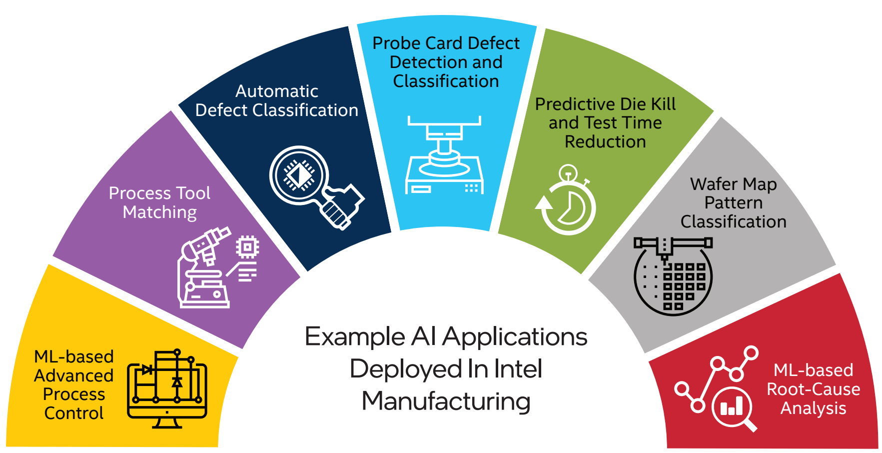


Figure 1. The adoption of AI in Intel's factories includes various applications.

Each application in Figure 1 fits the n-i, n problem statement. Here are some examples:

- **RCA.** Data from n-i steps is used to find the root cause for the abnormality observed at step n.
- **Machine learning for Advanced Process Control (APC).** Data from n-i steps is used to control the process at step n.
- **Predictive Die Kill and Test Reduction.** Machine-learning models are built using data from n-i steps to predict failures at some downstream step n.
- **Automatic Defect Classification (ADC).** Applications are devised to detect anomalies and quantify baseline non-systematic defects at a given step n.

Prioritizing AI Use Cases to Optimize Business Value

Defining a framework for prioritizing and deploying AI is critical for success. It's important because enormous volumes of data are generated from manufacturing operations every day. Equally essential is the burgeoning interest in AI from engineers and executives, as well as the pervasiveness of self-learning courses and the relative ease of creating pilot AI solutions. The prioritization process turns out to be simple, but not trivial, comprising three main ingredients: substantial business value, feasibility, and time to value (see Figure 2). Specifically, it is critical to answer the following questions to assess the need to automate existing business processes and workflows with AI:

- Are there any cost, productivity, or yield benefits?
- Is the targeted application error-tolerant? That is, are occasional false positives acceptable?
- Can the solution be truly automated, at scale? That is, can it be integrated into existing manufacturing automation systems in a way that the AI models can be built, monitored, and updated with minimal manual intervention?
- Will the solution be ready in time to produce the intended business impact?

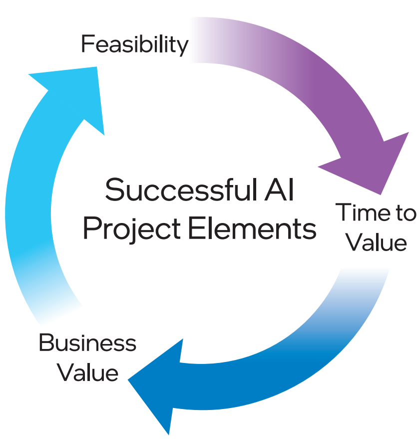


Figure 2. Prioritization matrix for AI projects.

Approaching AI with a Clear Purpose and Goal

The internet abounds with articles highlighting the colossal failure rates of AI and big data initiatives across industries. The AI project failure rate ranges between 60 percent and 85 percent,[2] despite huge investments in AI. One of the primary reasons for failed AI projects is that some are created without a clear use case in mind. In our opinion, AI should not become a solution in search of a problem.

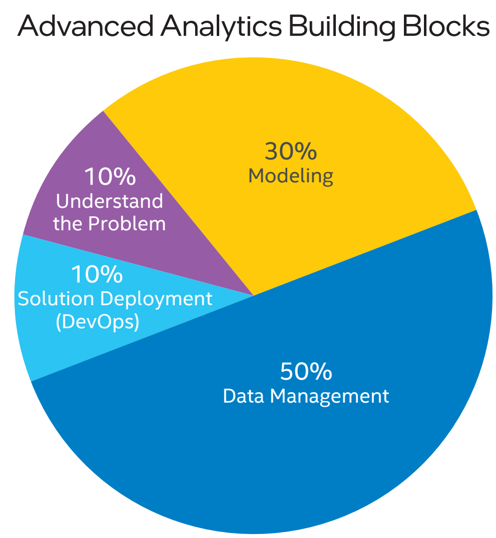


Figure 3. Building blocks of an advanced analytics solution.

> As well-documented on the internet, the failure rate for AI projects ranges between 60 percent and 85 percent across industries, despite huge investments in AI. We believe that one of the primary reasons for failed AI projects is that a majority of them were created without a clear use case in mind.

Besides choosing a specific use case, another aspect of AI project success is to take a holistic approach, covering four essential ingredients: understanding the problem, solution deployment, modeling, and data management. Ignoring one ingredient puts the entire project at risk:

1. Although it represents only 10 percent of the pie-chart, it's crucial to **understand the problem** to be solved. There is no substitute for domain expertise. AI implementation cannot be isolated from the business functions where the solutions will be ultimately deployed. We start by bringing domain experts—process engineers, equipment engineers, and yield engineers—into a working group to first understand the problem statement and associated business value before generating a proof of concept (PoC) AI solution. The PoC is thoroughly validated by the domain experts and iteratively perfected prior to the next step.

2. **Solution deployment (DevOps)** accounts for another 10 percent of the advanced analytics pie. DevOps is the only way to democratize complex algorithms across the organization. Even at the PoC stage, we are already planning for HVM deployment, which includes integrating the solution into our factory automation systems, to ensure adoption.
3. **Modeling** represents 30 percent of the advanced analytics pie. Here, we follow two rules:
   - Start with simple interpretable techniques, such as robust linear models or single decision trees, before employing less transparent AI methods such as ensembles or much heavier neural nets.
   - Use the best possible AI engines (algorithms) that are customized for our data domain. Specifically, we have invested heavily in exceedingly performant engines that can handle the unique characteristics of semiconductor data—highly imbalanced data, missing data, categorical data, and very often, "dirty" data. We constantly benchmark the performance and accuracy of our custom AI engines against the open-source engines used by millions of data scientists around the world.[3] Depending on the use case, we may use our custom AI engines or the open-source ones, or a mix of both.
4. **Data management** is the biggest slice of the advanced analytics pie—about 50 percent. And yet it is often the least exciting facet of advanced analytics. The problem with data is that it lives in different formats—structured and unstructured, video files, text, and images—and is stored in different places with different security and privacy requirements. Our data challenges are similar to those at other companies; we have addressed them and continue to improve upon them, through multiple data initiatives.

An AI Sampler: Three At-Scale AI Solutions Boosting Manufacturing Quality and Productivity

Thousands of AI models are in production at Intel. We used the project prioritization framework to develop and deploy each of the AI solutions shown in Figure 1. Let's examine three AI solutions in our factories, to illustrate how the prioritization framework contributes to project success.

ADC with Computer Vision and Machine Learning

ADC with computer vision and machine learning was among the first AI solutions that we deployed in HVM more than a decade ago.[4] Inline defect metrology helps detect excursions and catch issues in nanometric silicon chip layouts before they become serious yield and quality issues. ADC helps pinpoint issues at their source.

As illustrated in Figure 4, it can take six to nine months to train people to manually classify defects with 90 percent accuracy. Even after training is complete, an expert operator typically maintains only 70–85 percent accuracy over time due to multiple reasons, including:

- The work is highly repetitive.
- Process changes can result in new defect types that require further training.
- Classifying IC defects is inherently difficult. Some defects require design layout cross-referencing for accurate diagnosis, while others simply cannot be perceived by the human eye and brain.

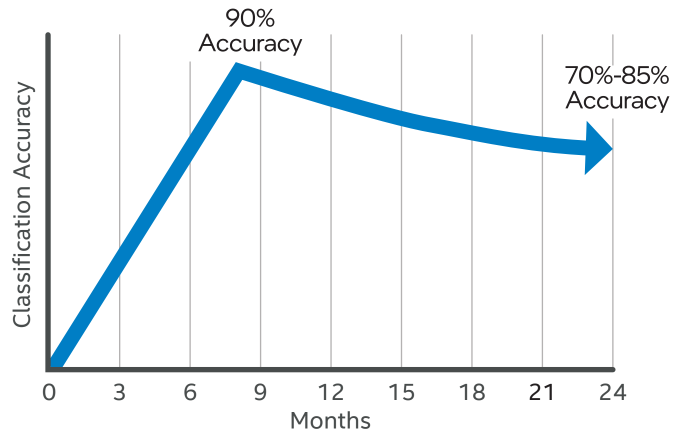


Figure 4. Defect measurements – accuracy vs. time.

With the problem statement clearly understood, we worked with a cross-functional team of process, yield, defect metrology, and equipment engineers to implement a machine-learning (including deep neural nets) ADC solution. This solution has since been deployed in both TD and HVM for every technology node manufactured at Intel—including Intel® Xeon® Scalable processors and Intel® Optane™ technology. The deployment itself required enormous effort and investment to integrate the AI algorithms into factory automation systems. Integration included several levels:

- The input side with the defect inspection systems.
- The user side to allow defect engineers and technicians to label images and configure corresponding target layout information.
- The factory operations side to automatically generate statistical process control (SPC) alerts and place factory lots on hold.

ADC enables us to measure and classify most of the defects on wafers produced by Intel's factories with a required rate of accuracy, and we have not experienced any increase in the total cost of ownership compared to other solutions. We have also been able to use existing imaging equipment in the post-wafer manufacturing process to implement ADC with computer vision and machine learning where it did not previously exist, **helping to prevent errors early and increase yield at no additional cost.**

Applying AI to the Manufacturing of the Latest Innovations from Intel

Two of Intel's flagship products are Intel® Xeon® Scalable processors and Intel® Optane™ technology. Intel Xeon Scalable processors are optimized for many workload types and performance levels, and deliver built-in AI acceleration and advanced security capabilities. Intel Optane technology—including Intel Optane persistent memory (PMem) and Intel Optane SSDs—is transforming the memory and storage hierarchy in the data center by bridging gaps, reducing bottlenecks, and crushing data latency.

Because we deploy our AI solutions at scale across all Intel factories, the resulting productivity and quality enhancements are available for these products as well as all the other silicon products Intel manufactures.

Root-Cause Analysis

Another good example of at-scale machine learning and advanced analytics is the RCA solution, which we have democratized across all technology nodes in Intel factories. In semiconductor manufacturing, rapidly finding the root cause of yield and quality issues is critical for both profitability and customer satisfaction.

The problem: At the scale of Intel's manufacturing, finding the root cause for a yield issue typically requires mining billions of parameters across e-tests, SPC, tools, operations, defects, queue times (QTimes), process times, wafer slot order, equipment logs, and many other data types. It is similar to finding the proverbial needle in a haystack. An analytically savvy engineer, with strong domain knowledge and years of experience, might be able to intelligently mine all available data in a few hours or days; but this knowledge is difficult to share between even two engineers, let alone across all of Intel's factories.

To democratize RCA, we developed interpretable machine-learning engines (including enhanced decision trees, novel committee methods, feature selection, and rule induction techniques) that can handle massive, noisy, heterogeneous, and frequently missing not at random (MNAR) manufacturing data. These engines provide solutions for tasks such as RCA, but a large effort is required to first transform the data into an analyzable form. We applied semiconductor domain expertise to create a custom big data storage infrastructure that provides very fast data access to the multi-dimensional data required for RCA; **engineers can now find potential root causes in a few minutes compared to hours or days.** And by seamlessly integrating machine-learning analytics on top of ready-to-use fast data infrastructure, we have substantially reduced the repetitive tasks of finding, extracting, cleaning, and joining data.

Sort Test Probe Card Inspection

Another practical AI solution deployed in Intel factories happens during sort test (see Figure 5). Sort test is the final step in wafer manufacturing, where the individual dies on a wafer are tested to determine yield (the number of good dies). Technicians use a piece of hardware called a Probe Card to deliver test patterns to the dies on the wafer, wherein tester pins make physical contact with the Probe Card.

The problem is due to the physical nature of the contact. Probe Cards are subject to wear and tear, which in turn can confound the test results. Historically, technicians periodically used a microscope to manually inspect the Probe Cards. This task is laborious, time-consuming, and presents a significant ergonomic risk. Additionally, this method suffers from the other limitations highlighted earlier in the ADC section.

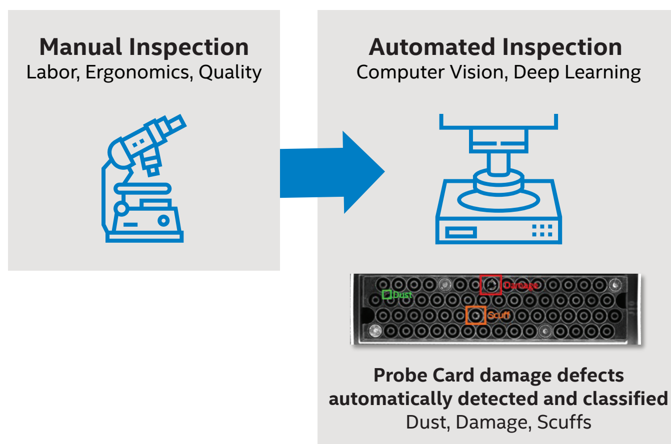


Figure 5. Probe Card defects – from manual to AI inspection.

We took a multi-stage approach to build a fully automated inspection system for Probe Cards. At each stage, we built intermediate applications that reduced the technician's workload. The overall solution now coalesces these applications. For example, one application automatically collects the image data while the Probe Card is on the test equipment. Review tools flag only anomalous areas of the Probe Card. Another application allows the technicians to easily label the data, which in turn allows us to create a labeled dataset to train a deep-learning AI system. By starting with a minimum viable product and augmenting the functionality incrementally, we developed a solution that worked for the technicians even as they helped us attain our goal of full automation. The system is now fully automated and deployed at multiple factories, providing significant productivity enhancements: **A task that previously took as long as 46 hours per week per factory has been reduced to less than 60 seconds.**

Conclusion

AI is poised to bring transformative changes to semiconductor manufacturing. We have been using a variety of AI solutions in Intel factories for almost two decades now and have experienced their value in terms of yield, cost, and productivity gains. The solutions described in this paper are examples of when it made sense to allow machines to do the repetitive tasks that humans would otherwise do painstakingly. In these examples, AI techniques offer more precise results, especially when compared to work performed by a less experienced engineer.

However, AI is not magic. In each case, after the problem statement was clearly understood, the AI algorithms had to be selected, adopted, or developed from scratch by machine-learning experts. PoC solutions had to be extensively validated by users and the perfected algorithms had to be integrated into factory automation systems via DevOps. Also, while we generate hundreds of petabytes of data, it is critical to apply AI to those use cases that can provide substantial business value, practical feasibility, and the quickest time to value. Once use cases have been prioritized, it is crucial to make appropriate investments compute resources, DevOps, and integrating the algorithms into existing workflows and automation systems. Investments are also important to free up bandwidth for domain experts.

Many related trends are contributing to the ongoing AI spring in Intel's factories:

- Storage and compute are becoming more affordable.
- Human talent is increasing by virtue of abundant AI courses and an active open-source community.
- AI momentum is growing, based on incremental success.

We expect the current AI spring to continue at least into the next decade. In semiconductor manufacturing as a whole, we are seeing growing awareness of AI across OEM suppliers, electronic design automation (EDA) vendors, data infrastructure providers, and our competitors. Just as the Industrial Revolution took a while to mature, the vision of "AI everywhere" will take time to become a reality. We are certain it will happen much faster than the former, but it will take some enterprise-wide cultural changes to supplement the technical solutions and make AI ubiquitous.

Intel manufacturing is on a path to implement AI solutions where they provide business value. We continually modernize our automation systems, data infrastructure, and most importantly, organizational culture, knowing that the greatest advantage comes when humans and machines work together.

Further Reading

You may find the following resources helpful:

- Goodwin, Randall et al., November 2004. "Advancements and Applications of Statistical Learning/Data Mining in Semiconductor Manufacturing," Intel Technology Journal, Volume 8, Issue 4.
- "Faster, More Accurate Defect Classification Using Machine Vision," Eugene Tuv et al., IT@Intel white paper, November 2018.
- "Streamline Deep-Learning Integration into Auto Defect Classification," Eugene Tuv et al., IT@Intel brief, July 2020.
- "Spatial Patterns in Sort Wafer Maps and Identifying Fab Tool Commonalities," Eric R. St. Pierre; Eugene Tuv; Alexander Borisov, Proceedings of the 2008 IEEE/SEMI Conference and Workshop on Advanced Semiconductor Manufacturing.
- Isabelle Guyon et al. "Analysis of the AutoML Challenge Series 2015-2018," Automated Machine Learning. The Springer Series on Challenges in Machine Learning, (2019): pp 177-219.

Find the solution that is right for your organization. Visit **Global Manufacturing at Intel** or contact your Intel representative.

The authors would like to acknowledge the contributions of many brilliant Intel employees from the Technology Development; Manufacturing, Supply Chain & Operations; and IT organizations who helped develop and deploy the AI solutions discussed in this publication. Special thanks also to all those users—engineers, technicians, operators, and managers—whose passionate involvement helped bring these solutions to see light of the day.

[1] "Spatial Patterns in Sort Wafer Maps and Identifying Fab Tool Commonalities," Eric R. St. Pierre; Eugene Tuv; Alexander Borisov, Proceedings of the 2008 IEEE/SEMI Conference and Workshop on Advanced Semiconductor Manufacturing.

[2] "85% of big data projects fail, but your developers can help yours succeed," Matt Asay, TechRepublic, November 10, 2017.

[3] Isabelle Guyon et al. "Analysis of the AutoML Challenge Series 2015-2018," Automated Machine Learning. The Springer Series on Challenges in Machine Learning, (2019): pp 177-219.

[4] "Faster, More Accurate Defect Classification Using Machine Vision," Eugene Tuv et al., IT@Intel white paper, November 2018.