

Micron AI Challenge

Registration Due Date: Apr 30, 2025

Submission Due Date: Jun 15, 2025

Signals to Quality: Unlocking Insights from Process Data

Introduction

In semiconductor manufacturing, complex process steps and toolings are involved. In this data challenge, we are particularly interested in one of the critical process steps. Various tools are involved in this step, and multiple sensors capture the signals during the process. At the end of each run, a wafer level measurement is collected which records the output value at each location of the product (indexed by the x,y coordinates). Due to equipment variations and process dynamics, the resulting measurements can exhibit non-uniformity across different locations. Accurately predicting both the **average measurement** and its **variability** is essential for optimizing manufacturing processes.

On top of the complex relations between tool setup, process signals, and the output, the tools degrade over time, and have varying lifetimes. As a result, the performance of the tools fluctuates over time. It is of interest to account for the tool degradation in the output prediction. It is noted that certain hardware inside the tools can be replaced, and their lifetime will be reset after replacement. An illustration of the data structure and relation is shown in Figure 1.

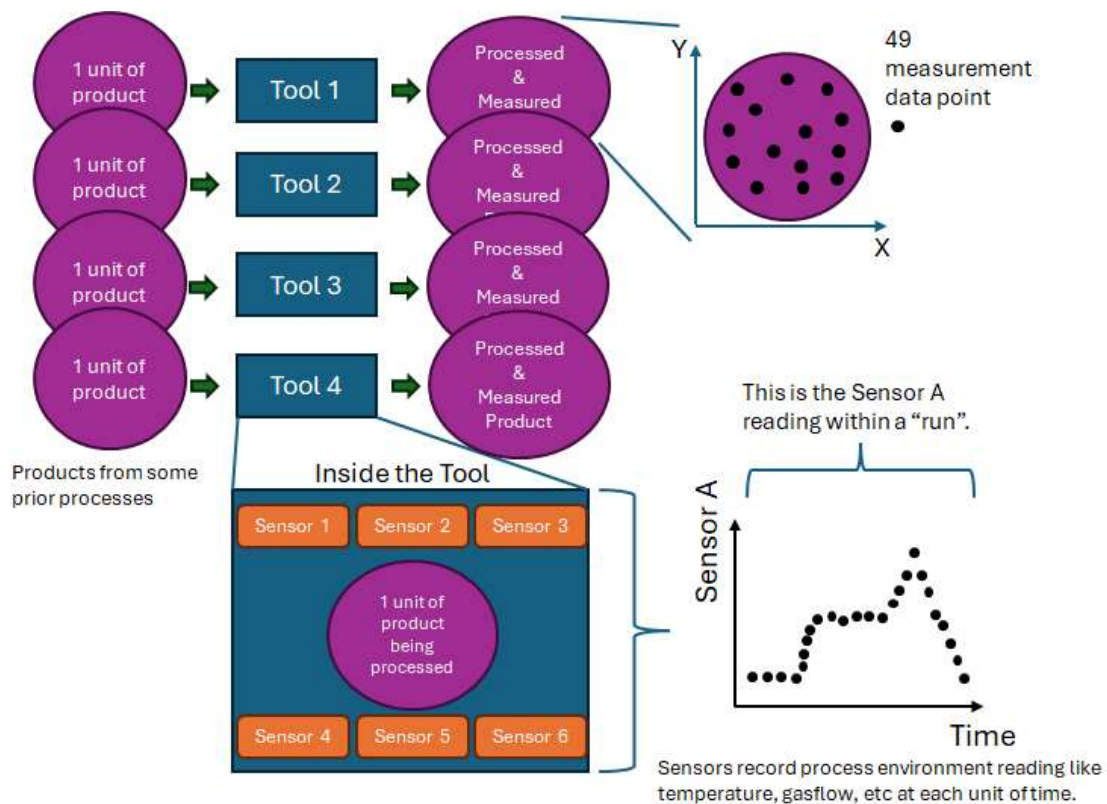


Figure 1. Illustration of the process data and quality data

Data

Data comes from a processing step from a semiconductor manufacturer. Participants will receive two datasets capturing process conditions and measured values:

1. Process Data (run_data.parquet)

This dataset contains **time-series in-situ sensor readings** recorded during each manufacturing run. The input data are organized in the following format. Essentially, the inputs record time series of various sensors associated with tools during the process runs. The age of the tools is recorded as “ConsumableLife” to consider the impact of tool degradation over the performance.

Input data schema

Column Name	Data Format	Description
ToolId	String	UUID for the tool where the unit of product has been processed in.
ConsumableLife	Numeric	The consumable life of the tool since last count reset.
RunId	String	Each unit of product is processed in a tool for a specific time. This span of activity of process is called a "run". RunId is the UUID for the specific run.
RunStartTime	DateTime	The start date time of the specific run
RunEndTime	DateTime	The end date time of the specific run
Recipe	String	Recipe UUID. The specific way that a unit of product processed is called a "recipe".
ProcessStep	String	A unit of product passes through a few processes. This column is to indicate which process the run and sensor data belong to.
TimeStamp	Numeric	Accumulative time stamp when the sensor data is collected, counting from start of the run.
StepId	Numeric	Within a recipe or a process, there are multiple steps. The number is chronologically accumulative from one step to another.
SensorName	String	Name of the sensor
SensorValue	Numeric	Value of the sensor at the point of data collection.

A sample data is shown below.

Input data sample

ToolId	ConsumableLife	RunId	RunStartTime	RunEndTime	Recipe	ProcessStep	TimeStamp	StepId	SensorName	SensorValue
Tool A	23.5	EIWOFHWOFU	3/7/2025 15:34	3/7/2025 16:34	REYUIHEWUHFE	Process Step A	0	1	Sensor A	324.5
Tool A	23.5	EIWOFHWOFU	3/7/2025 15:34	3/7/2025 16:34	REYUIHEWUHFE	Process Step A	0.016689847	1	Sensor B	0.2

Tool		EIWOFHWOFU	3/7/2025	3/7/2025	REYUIHEWUHFE	Process	0.0333796			
A	23.5	EW	15:34	16:34	OW	Step A	94	1	Sensor A	325.5
Tool		EIWOFHWOFU	3/8/2025	3/8/2025	REYUIHEWUHFE	Process				
B	203.5	ER	14:34	15:34	OW	Step A	0	1	Sensor A	314.5
Tool		EIWOFHWOFU	3/8/2025	3/8/2025	REYUIHEWUHFE	Process	0.0166898			
B	203.5	ER	14:34	15:34	OW	Step A	47	1	Sensor B	1.2
Tool		EIWOFHWOFU	3/8/2025	3/8/2025	REYUIHEWUHFE	Process	0.0333796			
B	203.5	ER	14:34	15:34	OW	Step A	94	1	Sensor A	315.5

2. Measurement Data (measurement_data.parquet)

This dataset provides **point-level measurements** across the manufactured product surface. Each Run ID record corresponds to 49 measurement points located at consistent locations of the product surface, whose Run ID can be used to link back to the process data. The data format and examples are shown below.

Output data schema

Column Name	Data Format	Description
X	Numeric	X-direction coordinate of the measurement on the product
Y	Numeric	Y-direction coordinate of the measurement on the product
Measurement	Numeric	Measurement value
RunId	String	Each unit of product is processed in a tool for a specific time. This span of activity of process is called a "run". RunId is the UUID for the specific run.

Output data sample

X	Y	Measurement	run_id
104.7738693	47.48743719	10.09654695	EIWOFHWOFUEW
-139.4472362	-55.02512563	10.13458545	EIWOFHWOFUEW
41.45728643	-143.9698492	10.13446887	EIWOFHWOFUEW

Objective

The primary objective of the data challenge is to develop a predictive model that utilizes the given signal data and tool usage data in each run to predict the output quality. The model should account for the variations in tool performance across different lifetimes and the impact of hardware changes.

A successful model will be evaluated based on its accuracy in predicting the point-level measurement outcome of the product. The accuracy will be evaluated on a separate set of test data provided by the competition committee based on the RMSE of all the point-level data in the test dataset. The model should demonstrate high precision and robustness in handling the time series data and the inherent variability in tool performance.