2023 SNU Data Science Invited Seminar - I

Nov 1, 2023

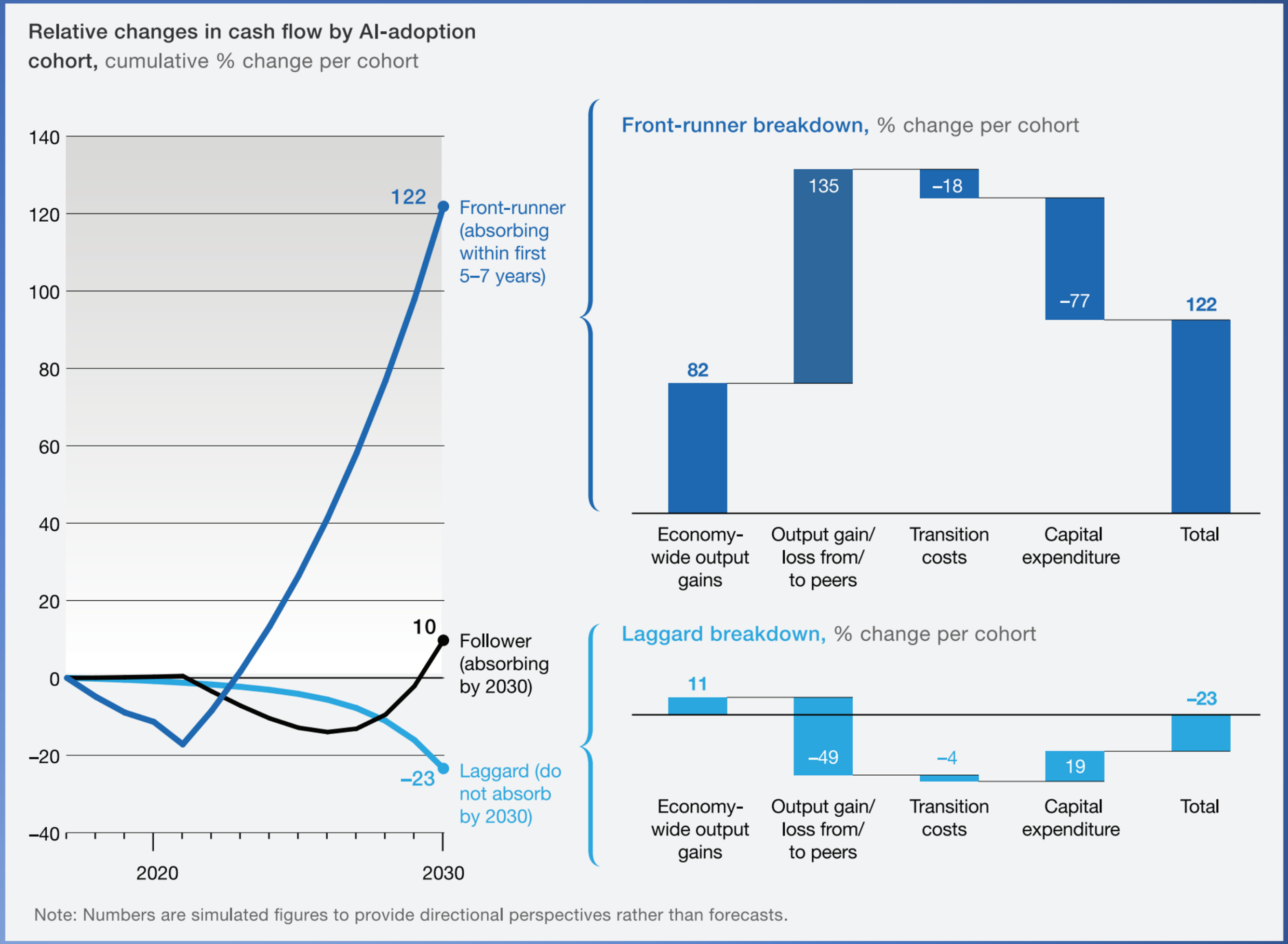
# Industrial AI Technology and Software Platform for Manufacturing

Sunghee Yun
Chief Applied Scientist & Head of Global R&D @ Gauss Labs

# Today

1. Why Manufacturing AI?
2. Computer vision ML for manufacturing
3. Time-series ML for manufacturing
4. Difficulties with time-series ML in manufacturing
5. Gauss Labs success story:  Virtual Metrology

# Fast AI adoption WILL create way larger economic gains




* Source: McKinsey Global Institute Analysis (2019)

**Data Characteristics**

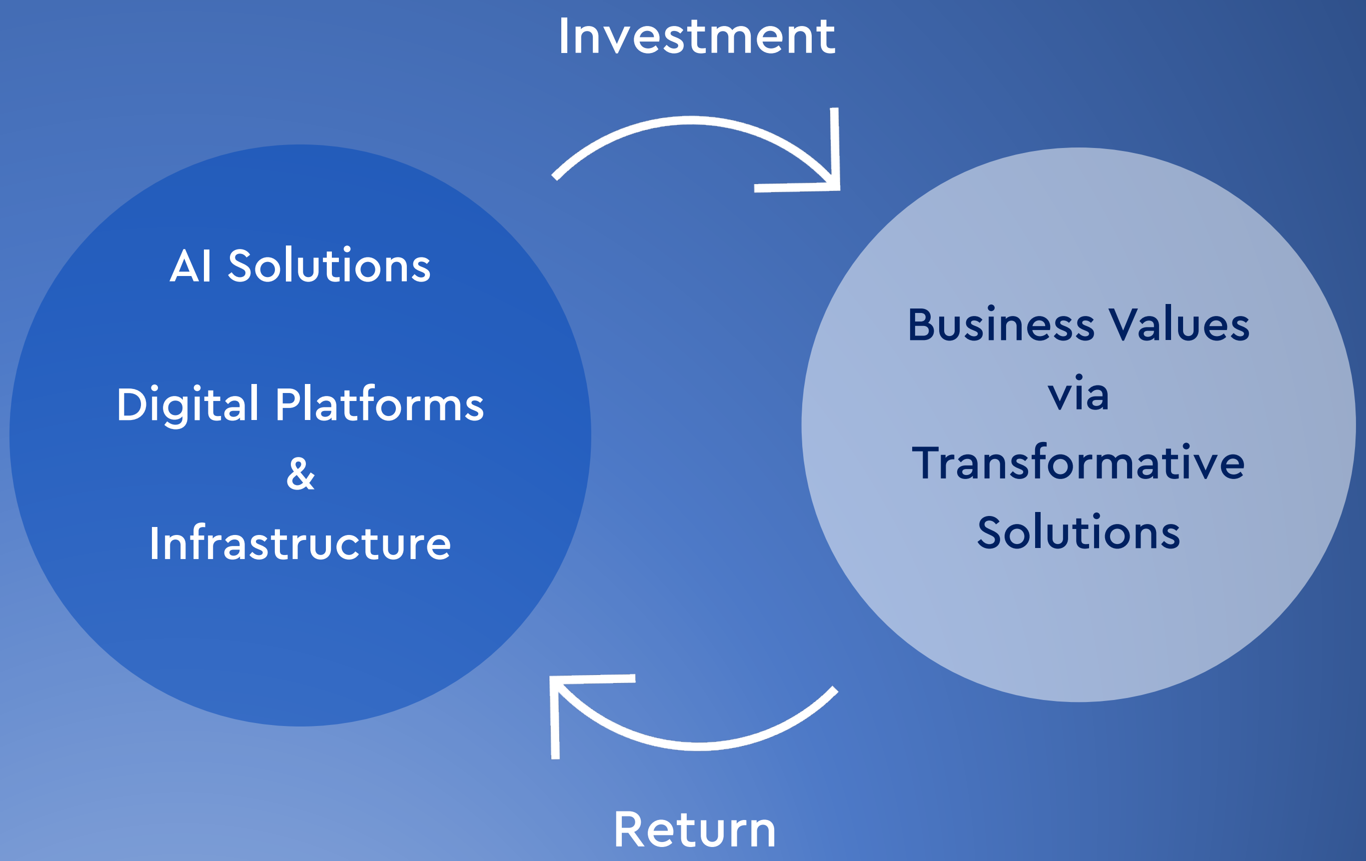
**Virtuous (or Vicious) Cycle**

**Data-centric AI**

## Data Characteristics

## Virtuous (or Vicious) Cycle

## Data-centric AI



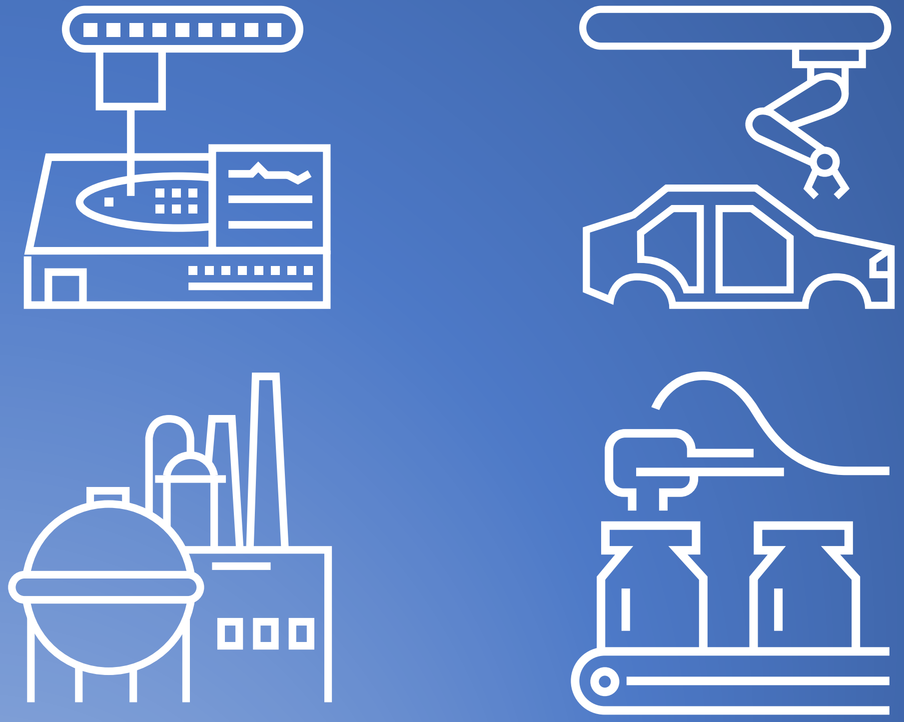
"We need 1,000 models for 1,000 problems" – Andrew Ng

Data-centric AI
Discipline of systematically engineering the data used to build an AI system

Data Characteristics

Virtuous (or Vicious) Cycle

**Data-centric AI**



Every company or sector has its own problems

# Our initial focus for 10x changes

GAUSSIAN MIXTURE 2021

## Semiconductor Fab

**A modern mega fab has ...**

- ~1,000 process equipement
- ~100 metrology equipment
- ~1,000 wafers per day, per equipment
- ~1,000 processes per wafer
- 3-6-month cycle time

## Servers and Systems

**Equipment Sensor Data**
(~1M parameters, ~1Tb/day)

**Metrology Image Data**
(~1M images, ~10 Tb/day)

**Manufacturing Execution Data**
(~10M events, ~10 Gb/day)

## Why Semiconductor?

Data availability from advanced digitalization

Diverse and sophisticated processes, ideal for expanding to new customers & sectors

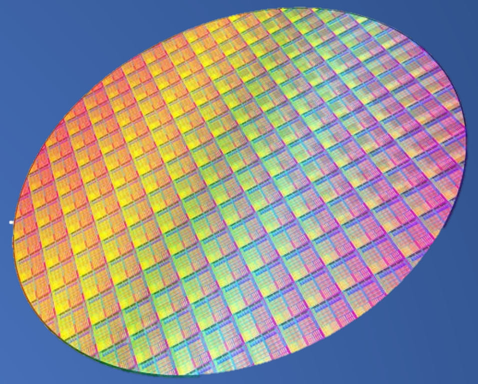
Huge impact even within the sector itself

# Computer vision and time-series ML in Manufacturing

## *lots of image data to measure and inspect*

Scanning electron microscope (SEM) images, transmission electron microscope (TEM) images, etc.

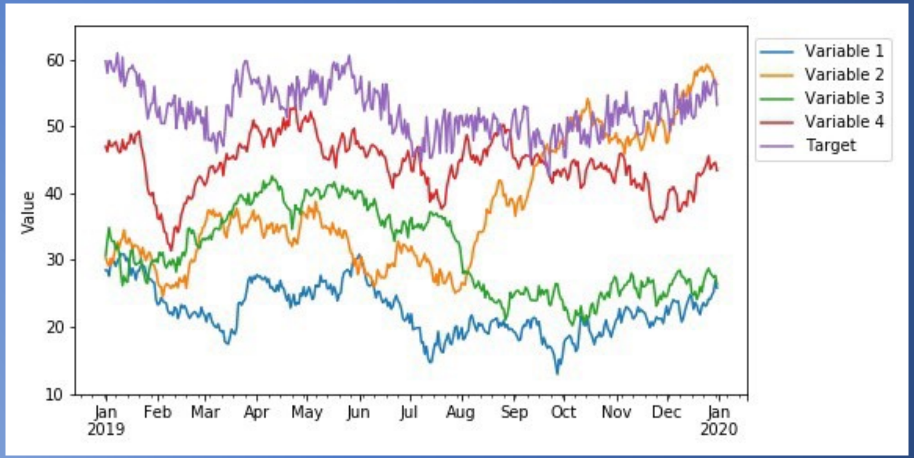
→ Image pattern classification, image anomaly detection, defect inspection, etc.



## *(almost) All the data coming from manufacturing are time-series data*

Equipment sensor data, process times, material measurement, *etc.*

→ time-series (TS) regression / prediction/estimation, TS anomaly detection, etc.



# Computer Vision ML for manufacturing



**Metrology**

*Measurement of critical features*

**Inspection**

*Anomaly detection,*
*localization and classification*

Image courtesy of ASML

# Scanning Electron Microscope

GAUSSIAN MIXTURE 2021

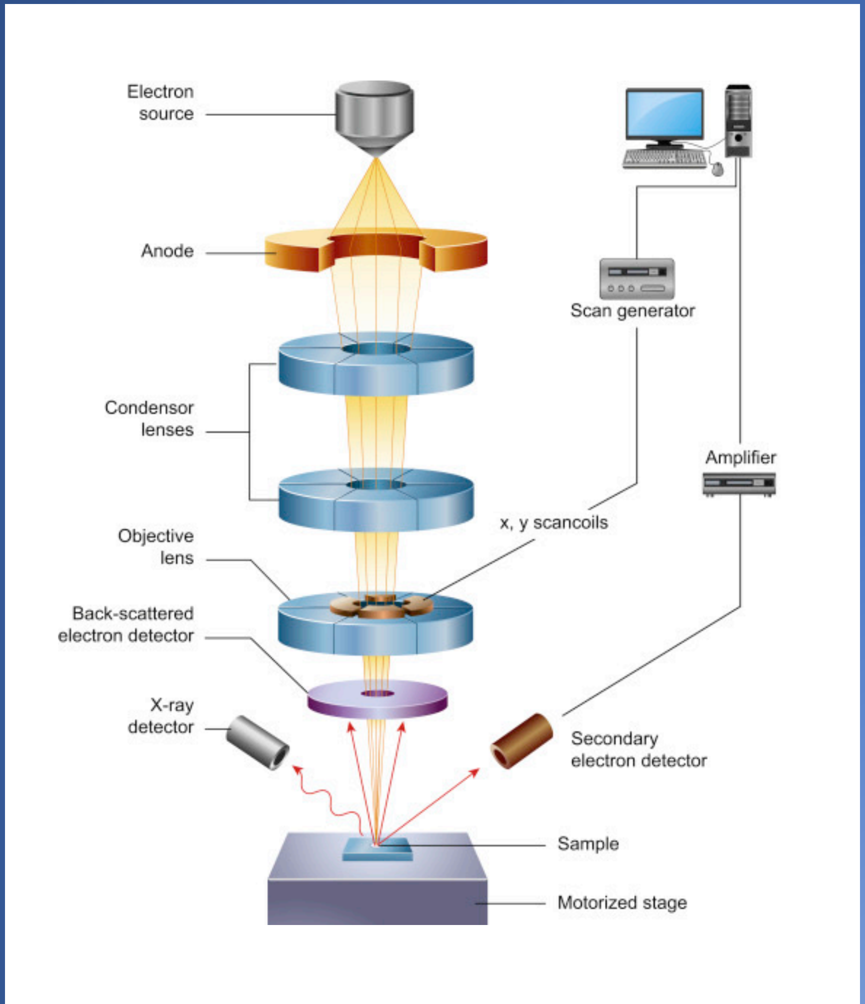
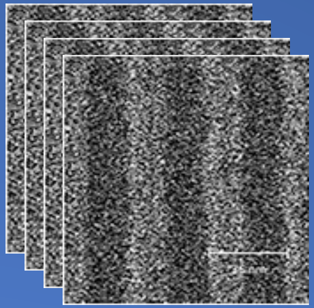
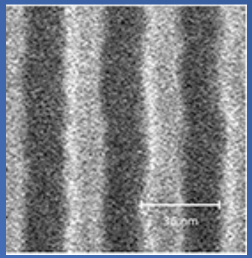



Image courtesy of https://www.sciencedirect.com/science/article/pii/B978008100040300002X



raw images



average of multiple images

Shot Noise Image courtesy of https://en.wikipedia.org/wiki/Shot_noise

# Image restoration

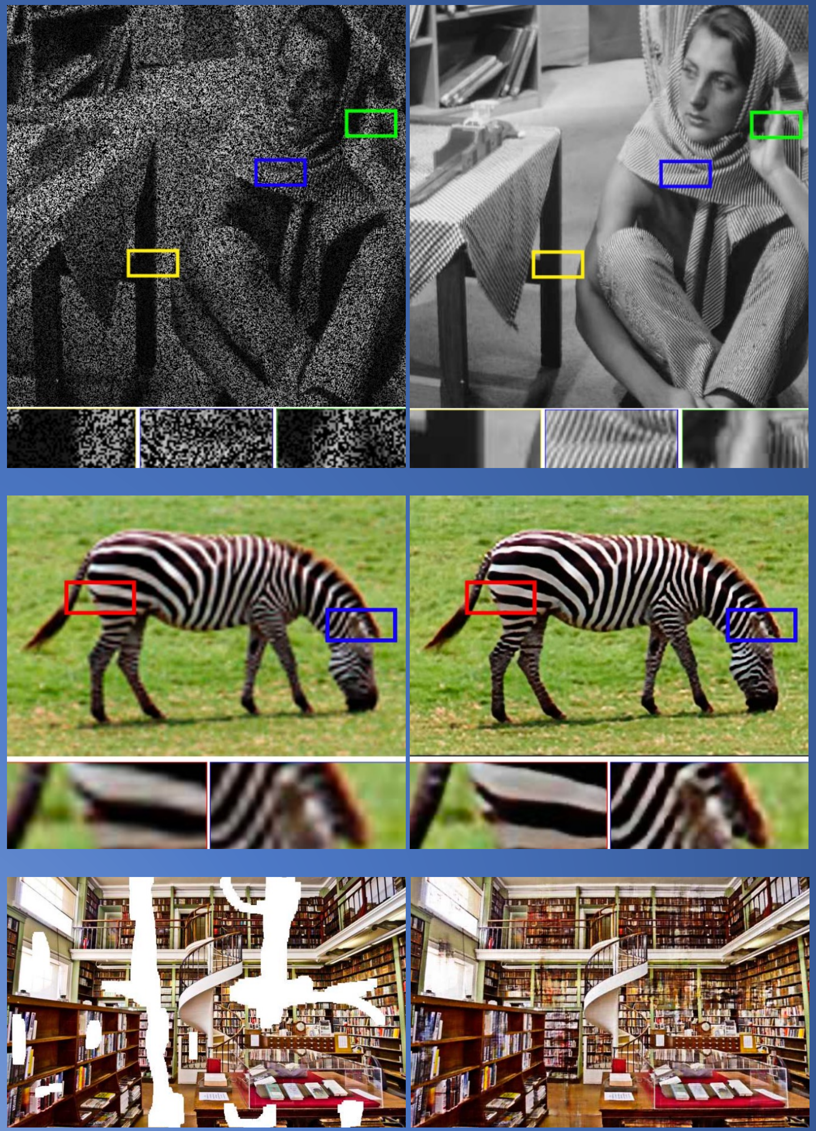
## Inverse problem of image corruption

$x = f(y) + n$

- y: clean image
- x: corrupted image
- n: noise

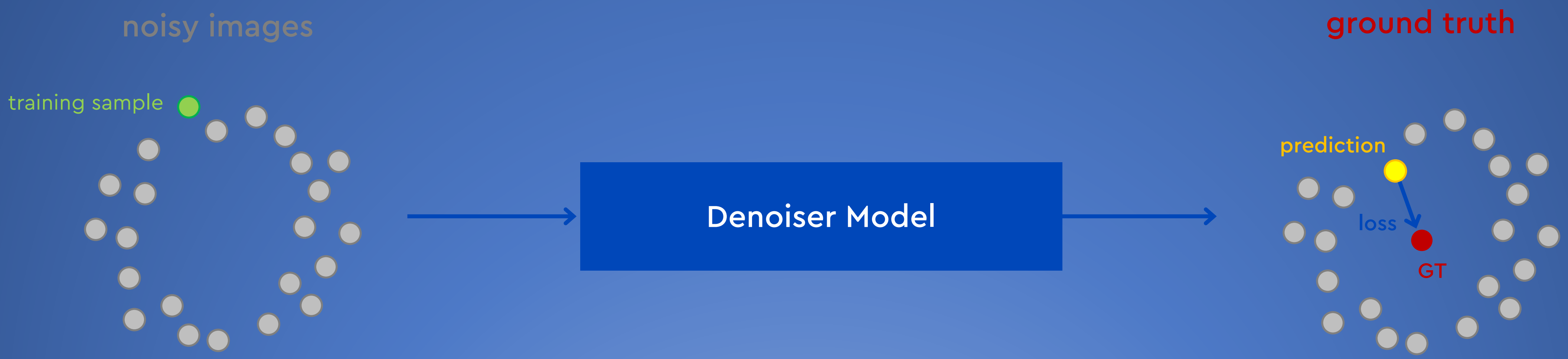
## f(·) and corresponding solutions

- Noising: Identity function → Denoising
- Downsampling → Super-resolution
- Missing pixels → Inpainting



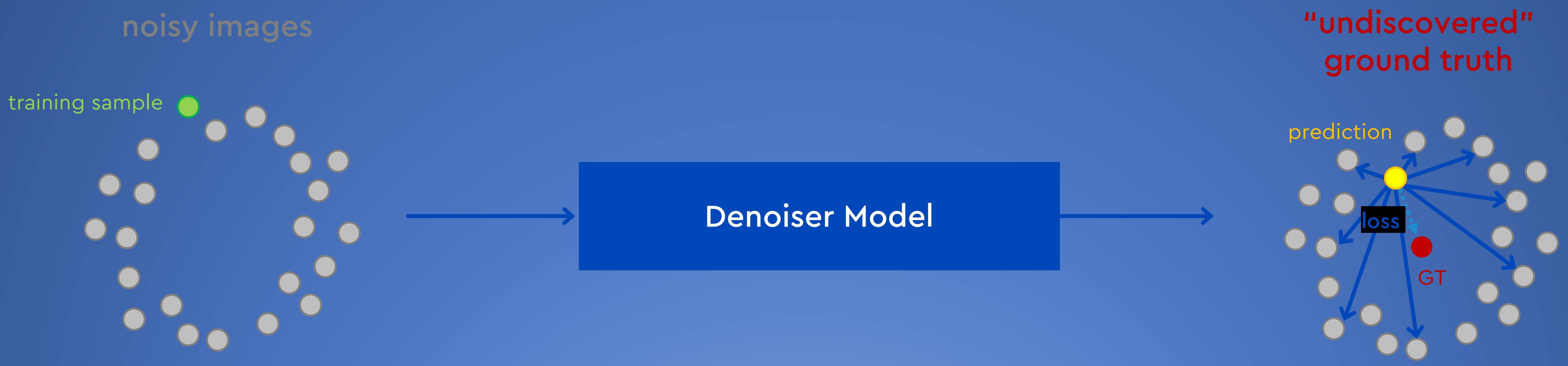
Dmitry Ulyanov et al. Deep Image Prior. CVPR, 2018

# Supervised image denoising




*However, it is not possible to acquire ground-truth images from SEM device, in practice.*

# Blind denoising without ground truth




*If the mean of the noise is zero, the average of the gradients that model takes is same with the gradient to the ground truth*

J. Lehtinen, et al. Noise2Noise: Learning Image Restoration without Clean Data. ICML, 2018.

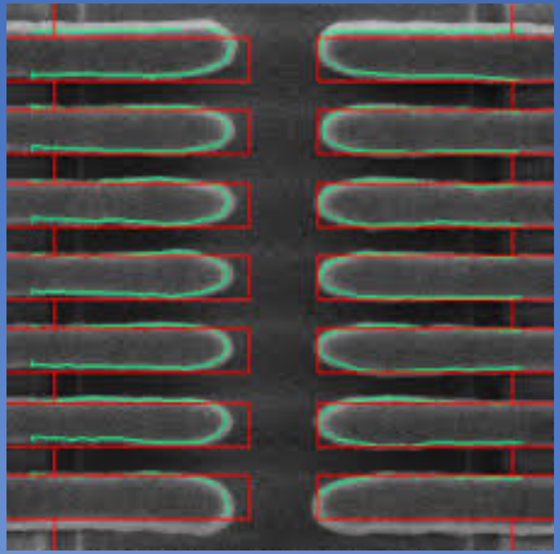
# Metrology based on segmentation and pattern recognition

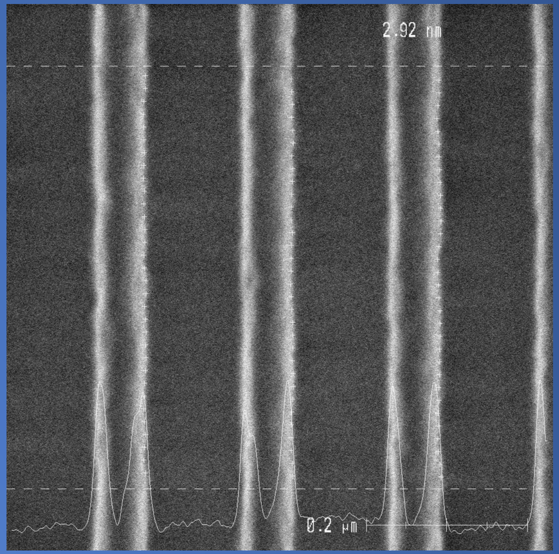
Investment

## Automatic measurement of critical dimensions

## Approaches

- Unsupervised texture segmentation
- Repetitive pattern recognition

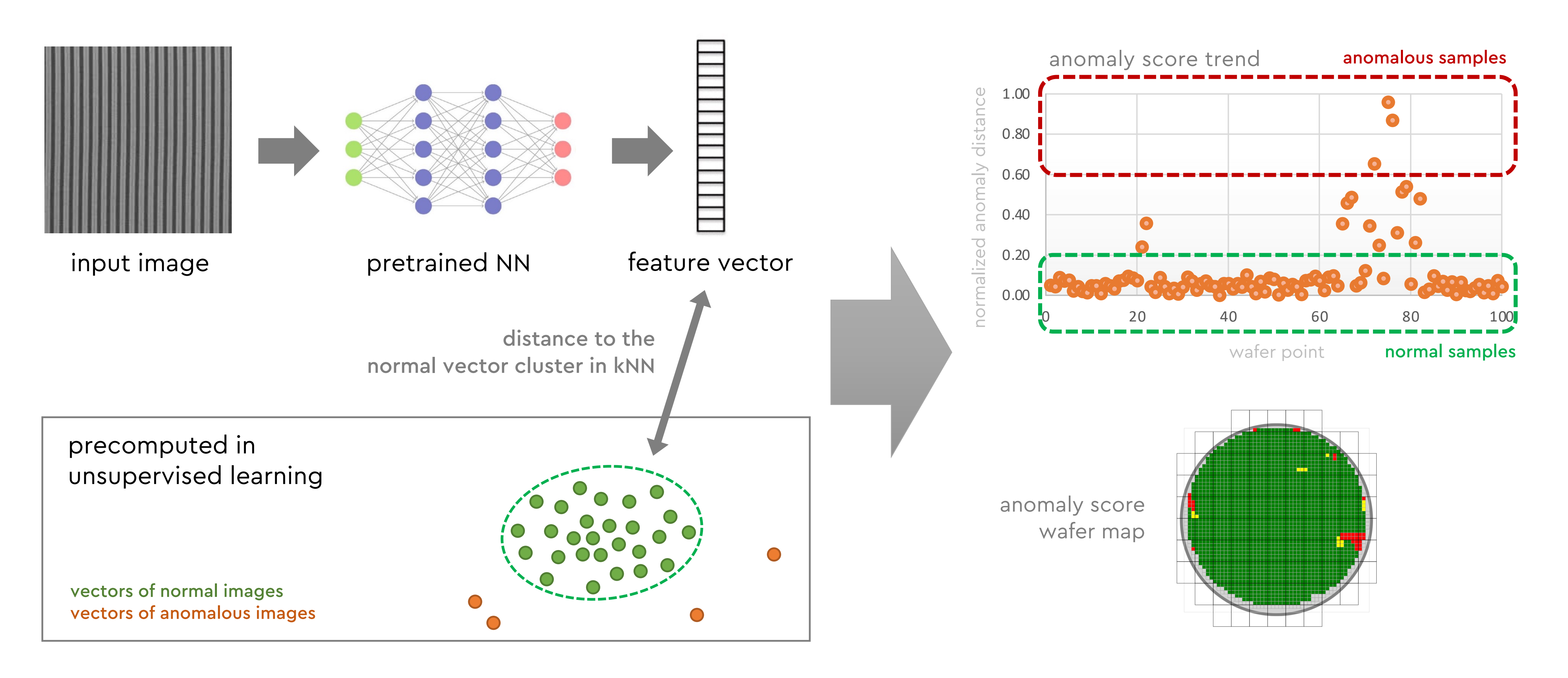




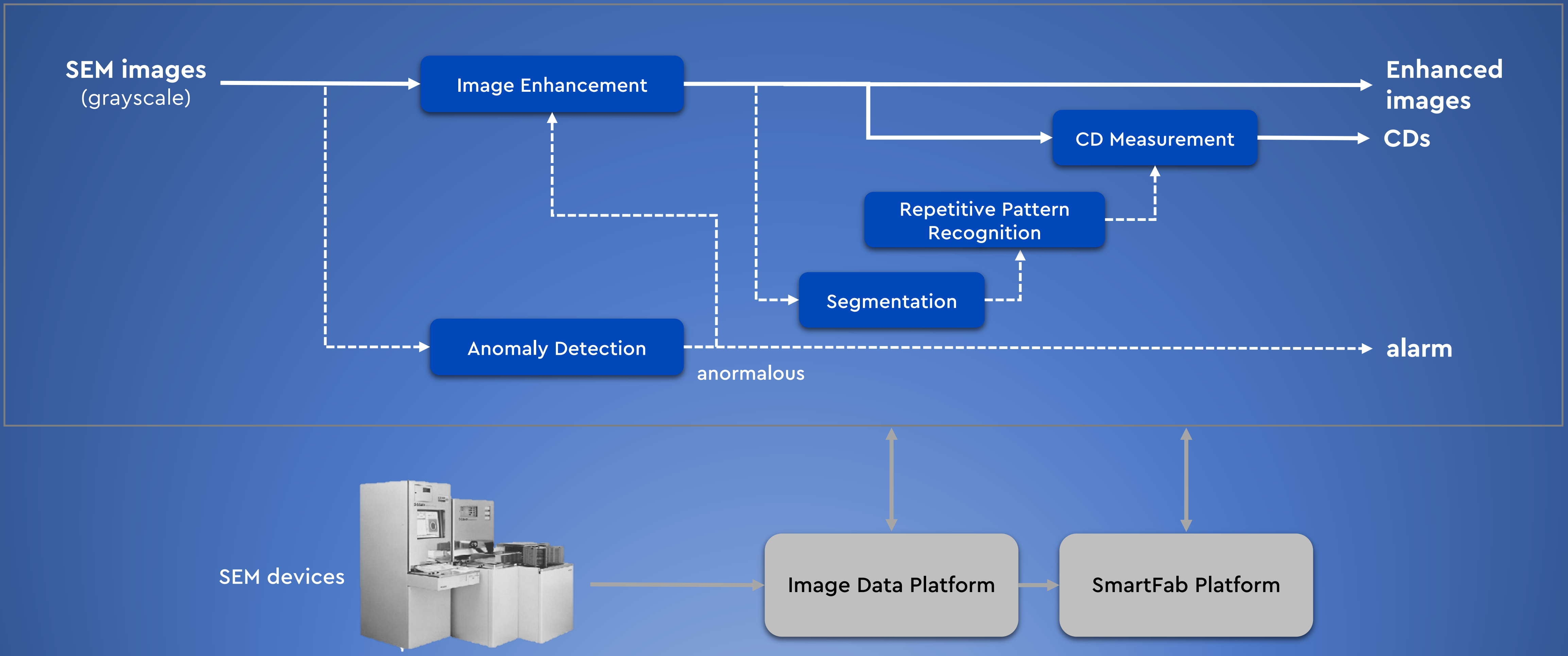

<0.1 nm measurement precision guaranteed

# Anomaly detection in unsupervised learning

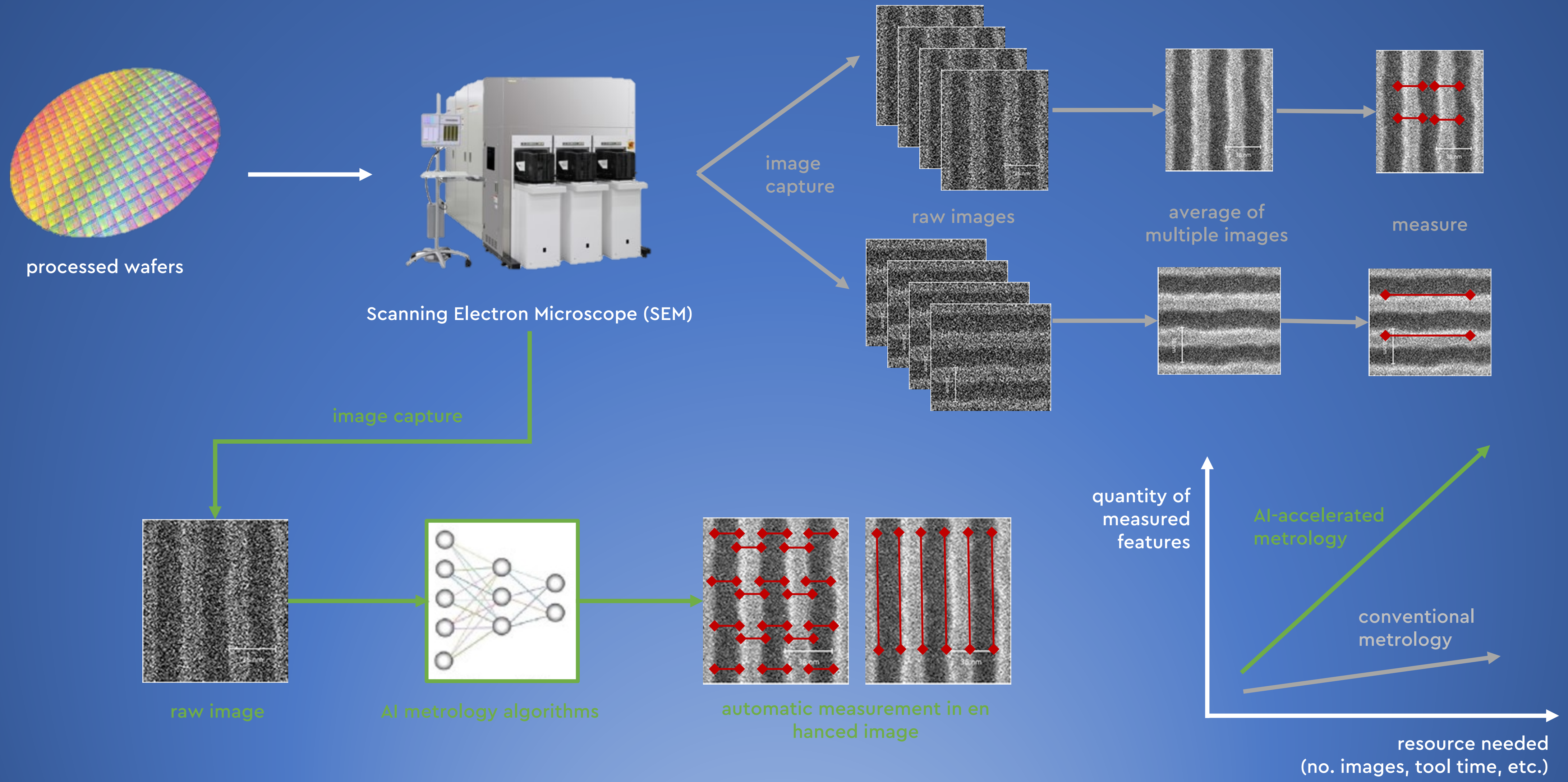




# AI-accelerated metrology system



# Automatic measurement for semiconductor manufacturing

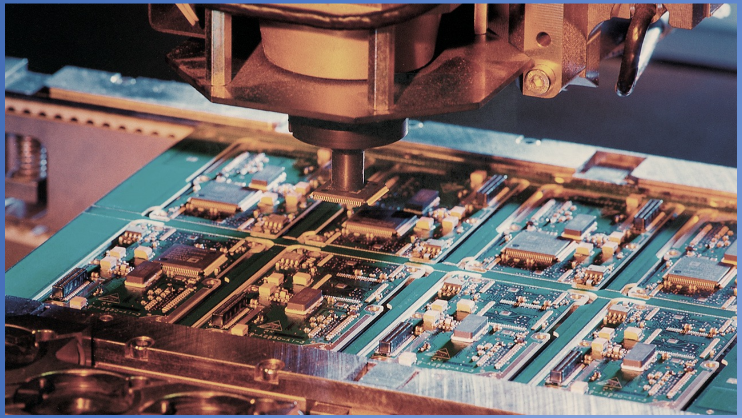


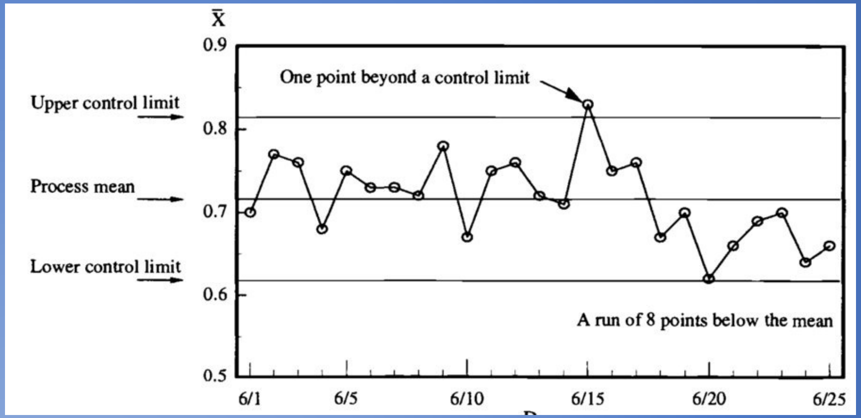
# Time-series ML for manufacturing

# Why time-series ML?

***manufacturing application is about one of the followings:***

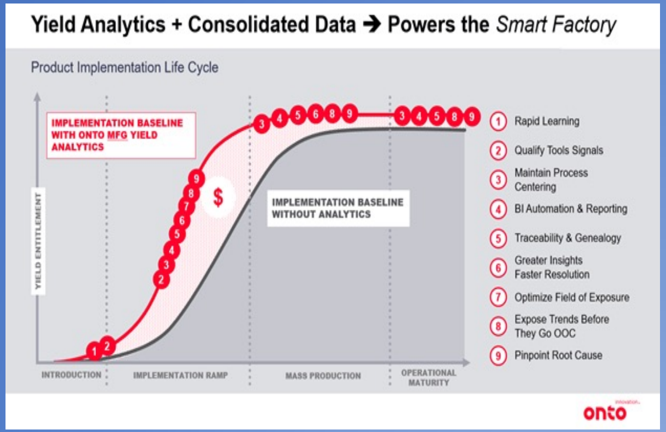
- prediction of time-series values - virtual metrology, yield prediction
- classification of time-series values - equipment anomaly alarm generation
- anomaly detection on time-series data - root cause analysis, yield analysis

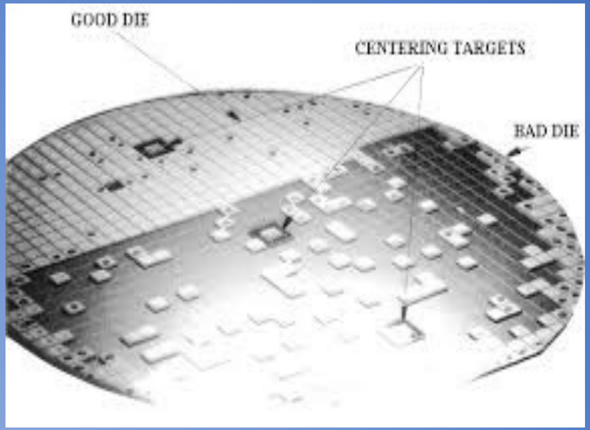




# Time-series regression/prediction/estimation

- virtual metrology
  - *measure unmeasured processed materials using equipment sensor signals*
  - *save investment on measurement equipment, downstream applications such as process control, statistical process control, yield improvement*
- yield prediction
  - *predict yield (# working dies / # total dies*
  - *better product quality and larger profit, business impact*

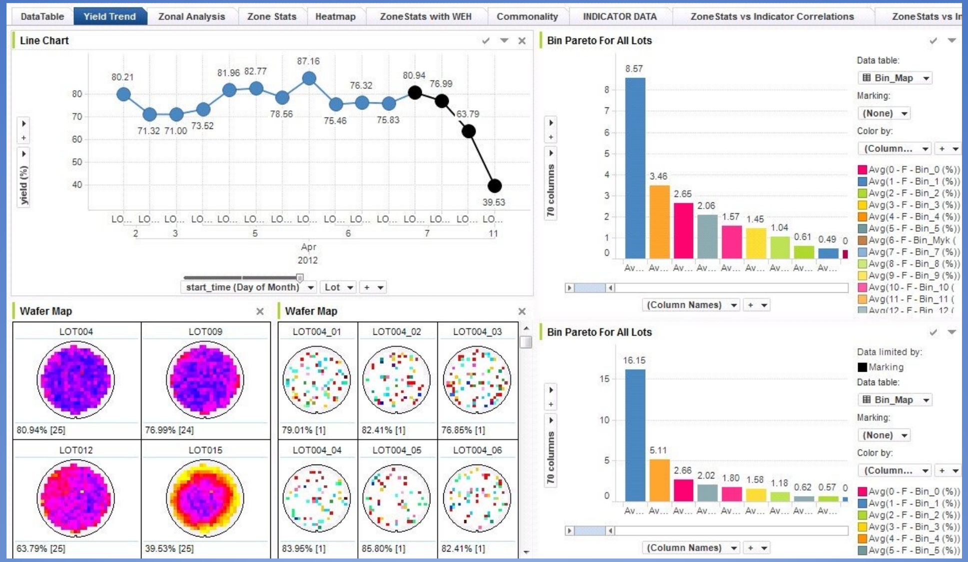


# Root cause analysis using time-series anomaly detection

- **equipment alarm root cause analysis**
  - *when alarm goes off, find responsible equipment and root causes*
  - *reduce equipment downtime, make process engineers' lives easier*

- **yield analysis**
  - *find responsible equipment and root causes for yield drop*
  - *a few % yield improvement brings profit increase of tens of millions of dollars!*



# Difficulties with Time-series ML in manufacturing

# Data challenges

- covariate shift & concept drift

$p\left(x\left(t_k\right), x\left(t_{k-1}\right), \ldots\right)$ *changes over time*

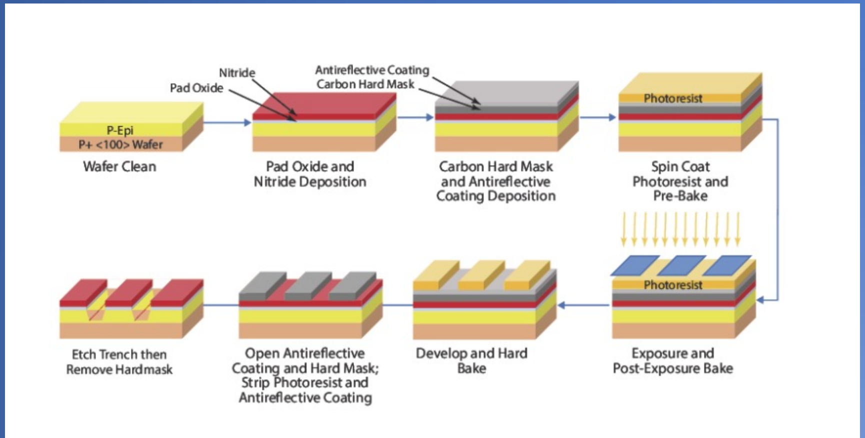
$p\left(y\left(t_k\right) \mid x\left(t_k\right), x\left(t_{k-1}\right), \ldots, y\left(t_{k-1}\right), y\left(t_{k-2}\right), \ldots\right)$ *changes over time*

- fat data, *i.e.*, # features way larger than # data
- poor data quality; missing values, anomalies, wrong formats
- huge volume of data to process
- multi-modality - different types of data

# Domain knowledge and fully home-grown algorithms

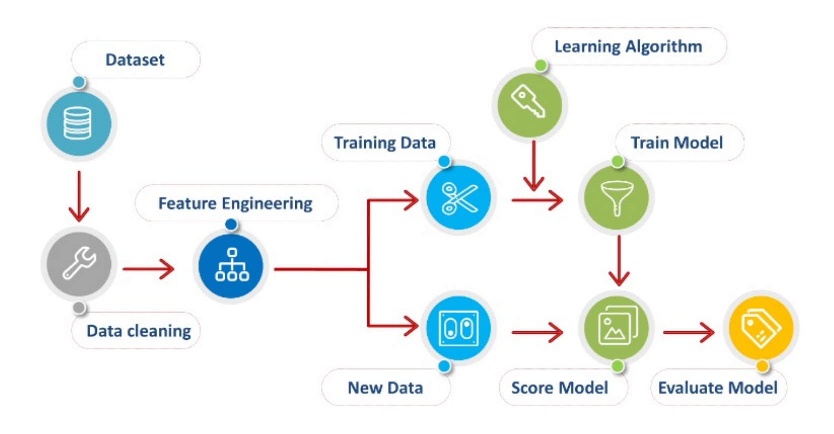
*in most cases,*
*domain knowledge is critical!*

close collaboration with customers required




*off-the-shelf algorithms*
*not working!*

developing fully customized algorithms needed



# Virtual Metrology (VM)

# What is VM?

***in many cases, we cannot measure all processed materials***

- measurement equipment is too expensive
- measuring every materials makes production slow inducing low throughput

***thus, we do sampling (with very low sampling rate)***

- in semiconductor manufacturing line, average sampling rate is less than 5%

***PROBLEM***

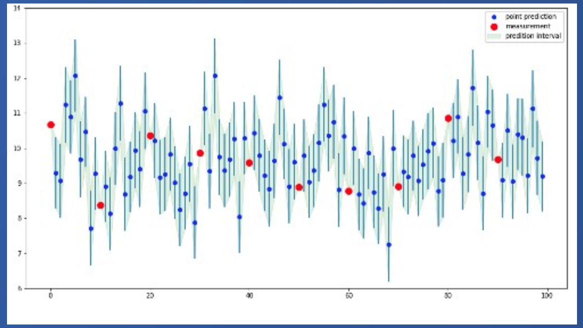
- predict the measurement of unmeasured material using indirect signals
- sensor data, maintenance history, operation data, . . .

process

Sample measurements

Sensor data, maintenance history, operation data, ...

AI Model
(semi-supervised learning)



# Business Impact made by VM

*To the best of our knowledge*

- no organization has even been successful with VM

***Gauss Labs VM***

- uses online learning to cope with data drift/shift
- RMSE comparable to measurement equipment precision
- also predicts uncertainty of predictions - providing prediction reliability information

***VM implications***

- measuring ALL wafers – equivalent to investing on 100x measurement equipment
- enables optimal re-allocation of limited measurement resources

# Conclusion

*supervised and unsupervised ML everywhere in industrial AI applications*

*lots of challenges*

- data challenge, domain knowledge required, need for customizing algorithms

*huge changes potentially made via various applications*

**Impacts**

- Tens of Millions of dollars by 1% yield increase
- 100x measurement equipment save by VM