

DMQA Open Seminar

Active Learning in Semiconductor Manufacturing

2021. 08. 20

Data Mining & Quality Analytics Lab.

발표자 : 황성진

Contents

1. 반도체 산업 데이터

- 반도체 제조 산업
- 반도체 데이터 소개
- 반도체 관련 Task

2. Active Learning

- Active Learning?
- Sampling Strategy
- Active Learning in Semiconductor Manufacturing

3. Conclusion

발표자 소개

❖ 황성진

- 고려대학교 산업경영공학과
- Data Mining & Quality Analytics Lab (김성범 교수님)
- 석사 과정 (2021. 3~)

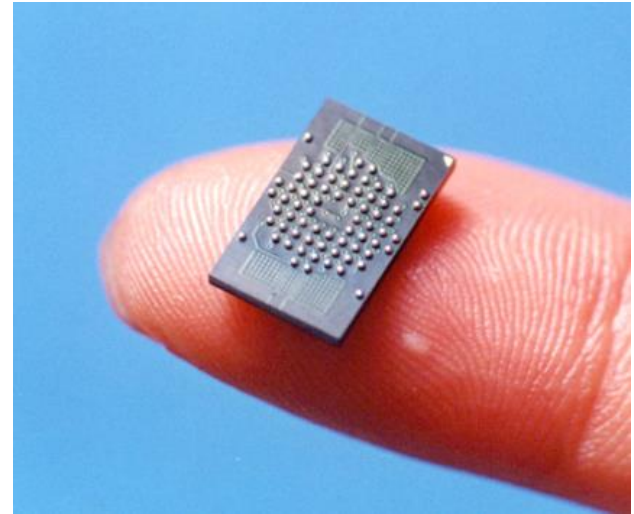
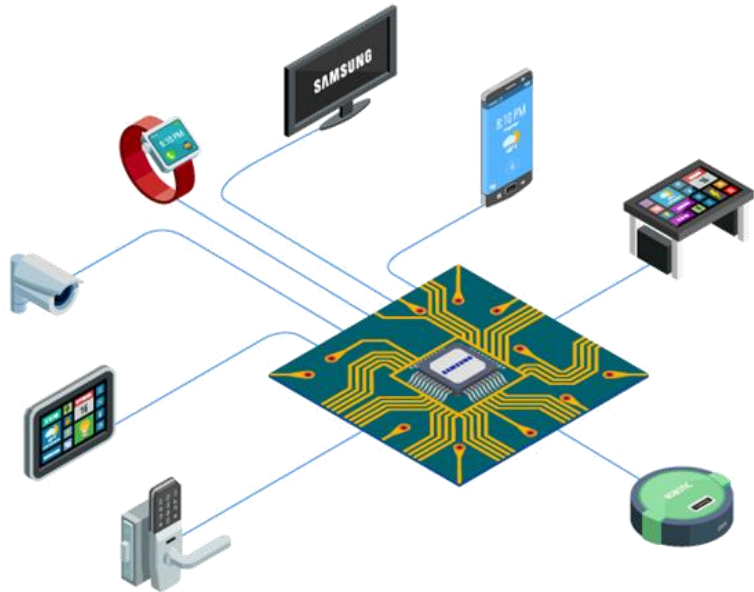


❖ 관심 연구분야

- Machine Learning and Deep Learning
- Active learning in semiconductor manufacturing

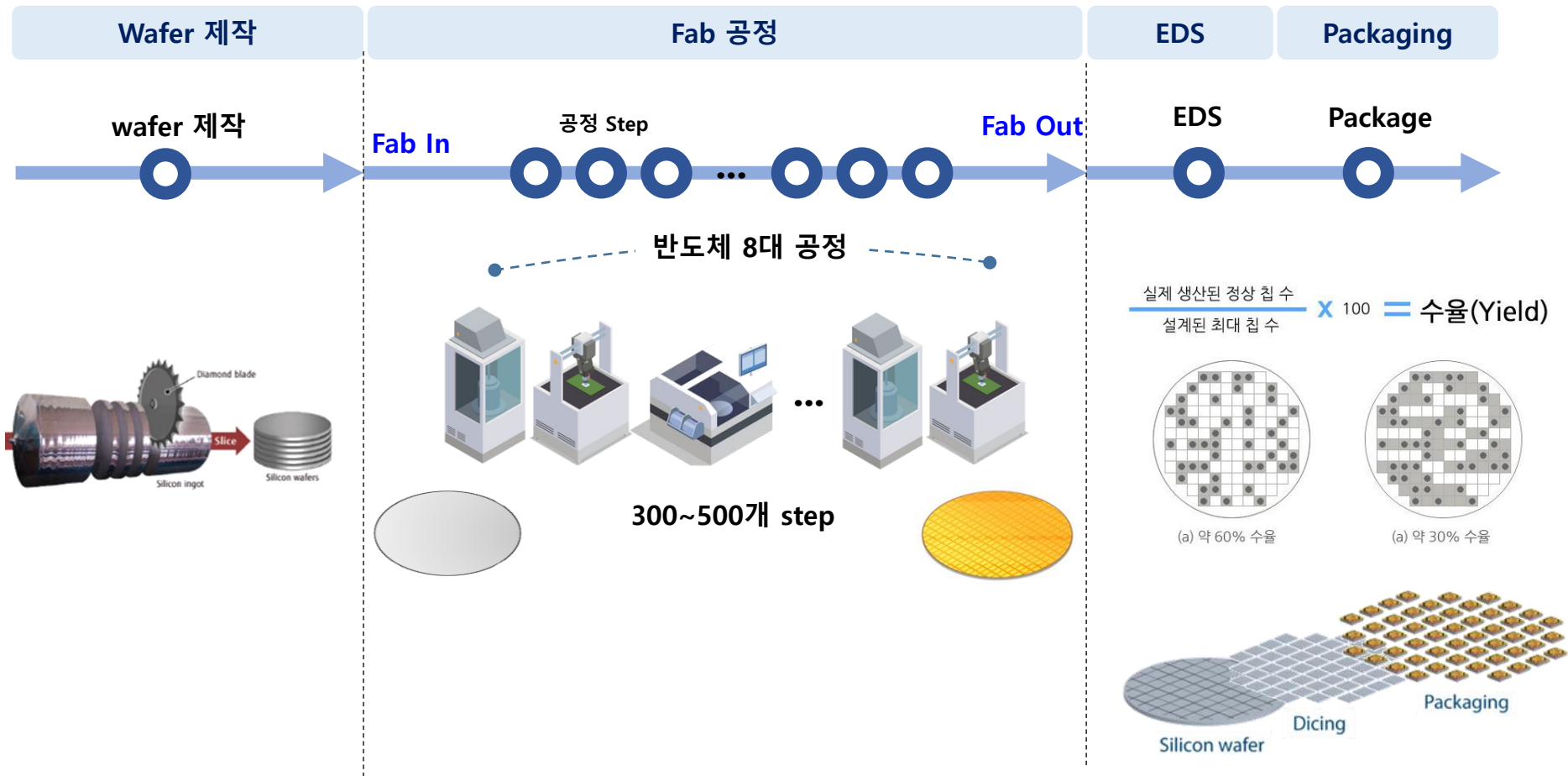
반도체 제조 산업

- ❖ 반도체 : 전기전도도가 도체와 부도체의 중간 정도인 물질을 총칭
- ❖ 집적 회로(Integrated Circuit) : 특정 기능을 수행하는 전자 회로를 하나의 칩에 모아 구현한 것.



반도체 제조 산업

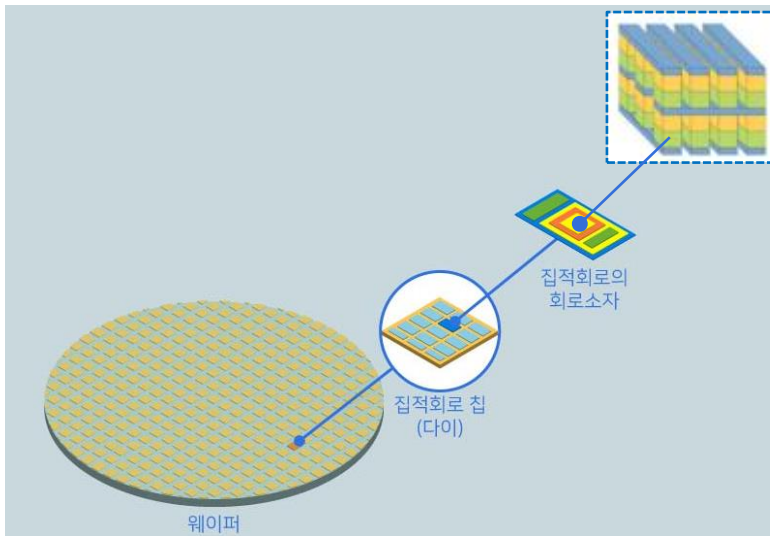
❖ 반도체 제조 Overview



반도체 제조_① Wafer 제작

❖ 웨이퍼(Wafer) : 집적 회로를 만들기 위한 바탕이 되는 얇은 기판

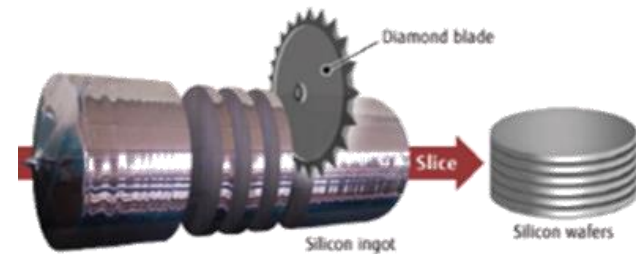
• Wafer와 집적회로



• Si 결정 성장을 통한 Ingot의 성장

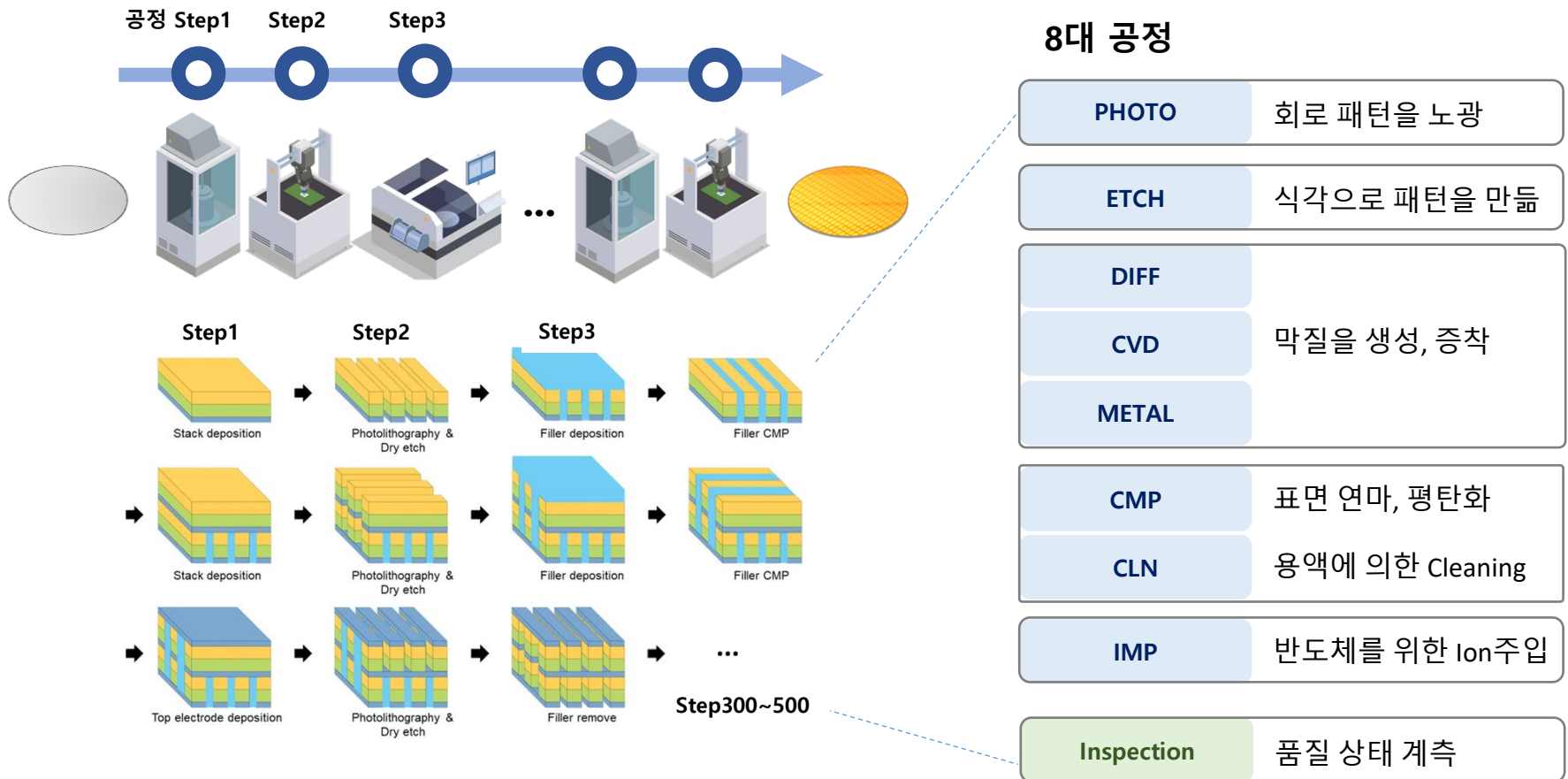


• 절단 및 연마를 통해 Wafer 완성



반도체 제조_② Fab 공정

❖ Fab 공정(8대공정) : 원하는 구조의 반도체를 실질적으로 만드는 단계



반도체 제조_③ EDS

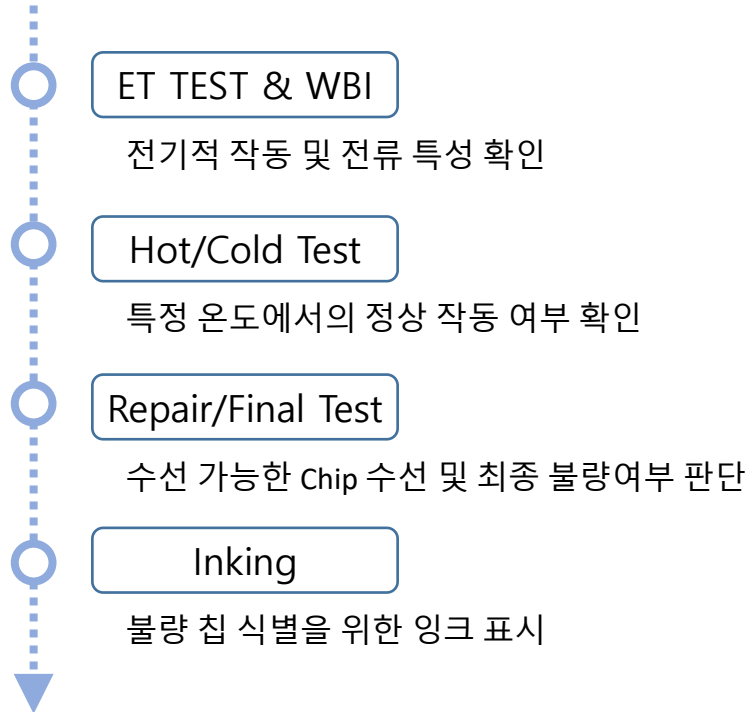
Wafer 제작

Fab 공정

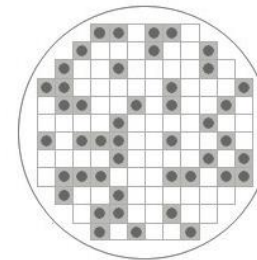
EDS

Packaging

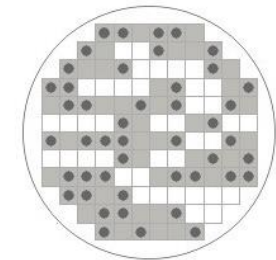
- ❖ **EDS(Electrical Die Sorting)** : 전기적 특성검사를 통해 Wafer 상태의 각 칩들이 원하는 품질 수준에 도달하는지 Test



$$\frac{\text{실제 생산된 정상 칩 수}}{\text{설계된 최대 칩 수}} \times 100 = \text{수율(Yield)}$$



(a) 약 60% 수율



(a) 약 30% 수율

반도체 제조_④ Packaging

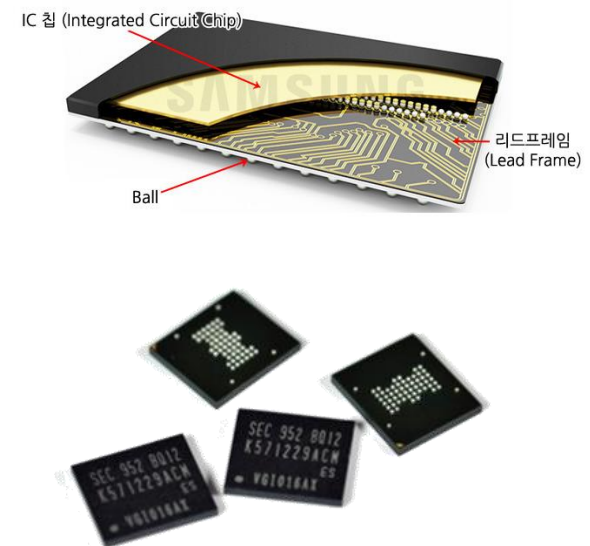
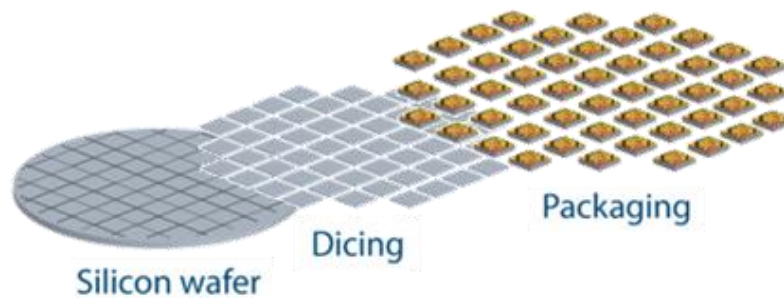
Wafer 제작

Fab 공정

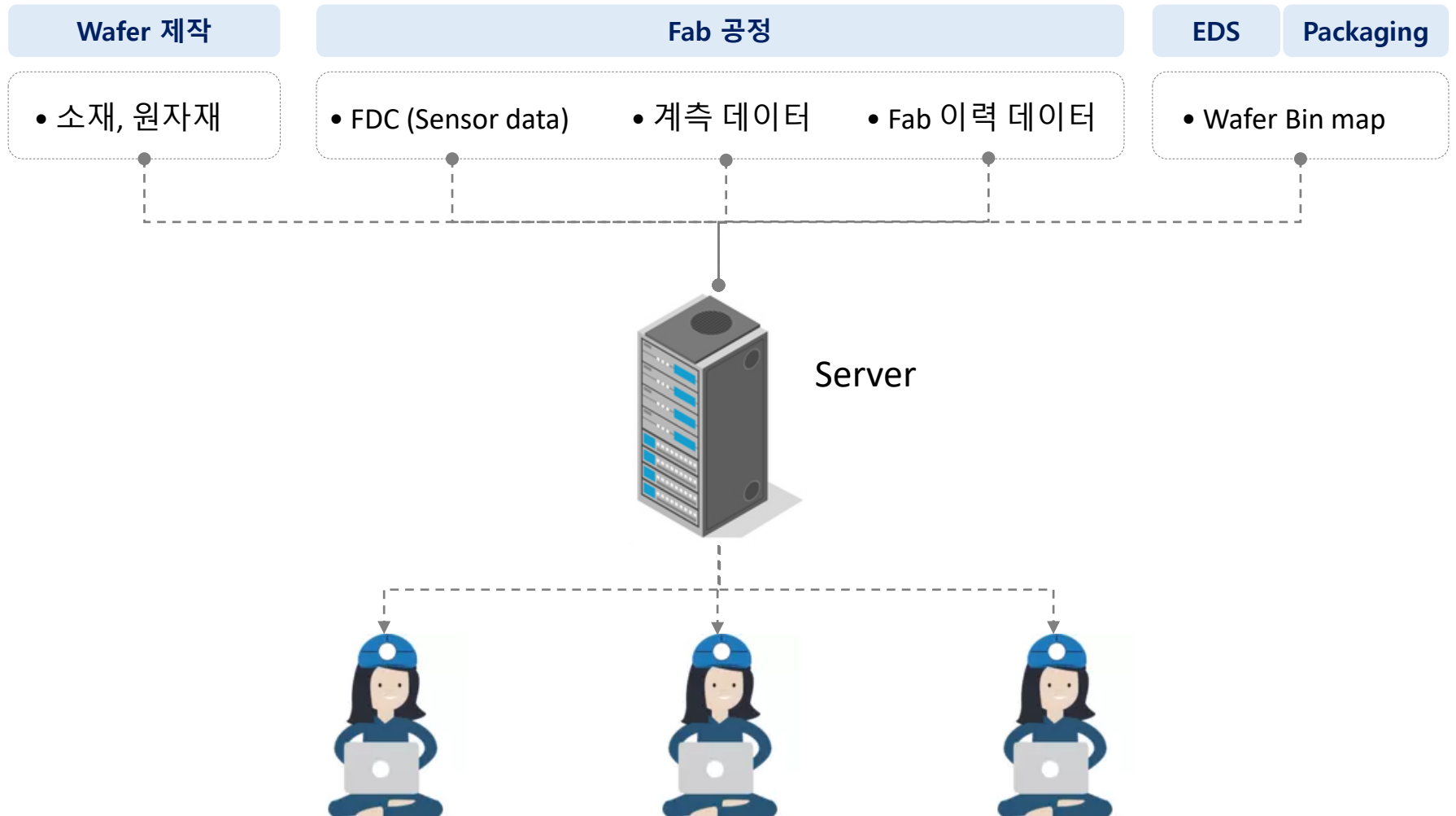
EDS

Packaging

❖ **Packaging** : Wafer를 칩으로 나눈 뒤, 단자 간 연결을 위한 전기적인 포장 진행



반도체 데이터 Domain



반도체 데이터 Domain

Wafer 제작

- 소재, 원자재

Fab 공정

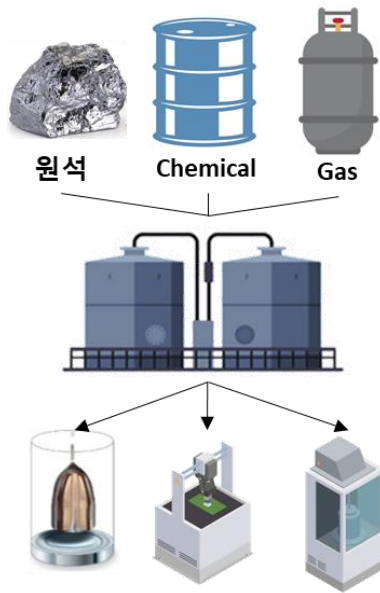
- FDC (Sensor data)
- 계측 데이터
- Fab 이력 데이터

EDS

Packaging

- Wafer Bin map

❖ 소재 : 성분, 불순물의 정보를 제공, Tabular data 형태로 가진 경우가 대표적



입고 시점	성분1	성분2	불순물1	불순물2
...	568	752	0.001	0.00001
2021/08/15	557	125	0.003	0.00002
2021/08/16	562	685	0.005	0.00001
2021/08/17	566	432	0.005	0.00002
2021/08/18	560	452	0.003	0.00001

반도체 데이터 Domain

Wafer 제작

- 소재, 원자재

Fab 공정

- FDC (Sensor data)
- 계측 데이터
- Fab 이력 데이터

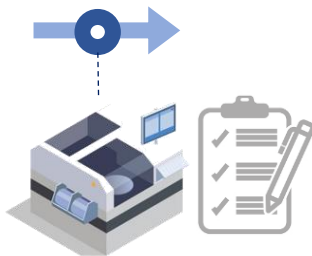
EDS

Packaging

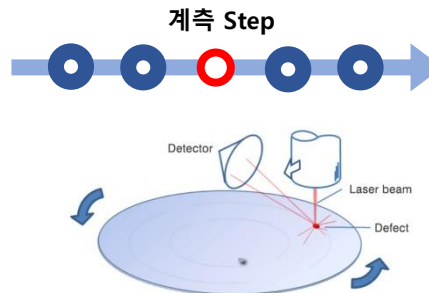
- Wafer Bin map

- ❖ FDC (Fault Detection and Classification) : Fab 공정 설비에서 모니터링되는 Sensor Data
- ❖ 계측 데이터 : 공정 사이사이, Wafer의 상태를 체크하기 위한 별도 검사 Data
- ❖ Fab 이력 데이터 : 생산라인 물류 이동과 같은 이력 데이터

• FDC (Sensor data)



• 계측 데이터



• Fab 이력 데이터



반도체 데이터 Domain

Wafer 제작

- 소재, 원자재

Fab 공정

- FDC (Sensor data)

- 계측 데이터

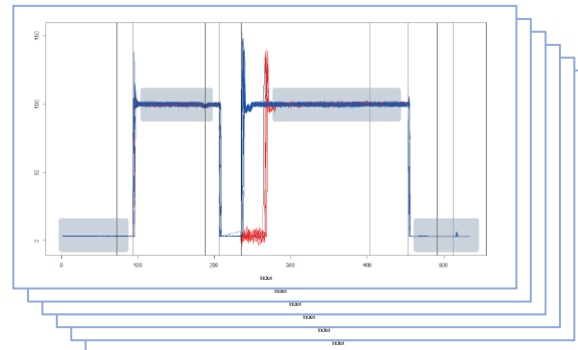
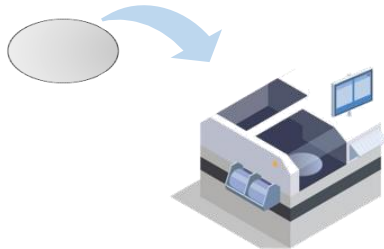
- Fab 이력 데이터

EDS

Packaging

- Wafer Bin map

- ❖ FDC: 공정 설비의 상태를 모니터링하기 위하여 Sensor로부터 받아오는 각종 Parameter로, 일반적으로 진행 시간에 따른 온도, 압력, chemical량 등 Time series 형태의 데이터를 확인.



Sensor 1~n개

Time	parameter1	parameter2	parameter3	...
20210815111255	556	200	0.001	...
20210815111256	557	200	0.003	...
20210815111256	558	200	0.005	...
20210815111256	1002	232	0.005	...
20210815111256	1008	256	0.003	...

반도체 데이터 Domain

Wafer 제작

- 소재, 원자재

Fab 공정

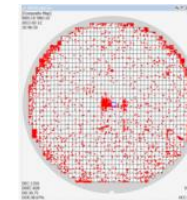
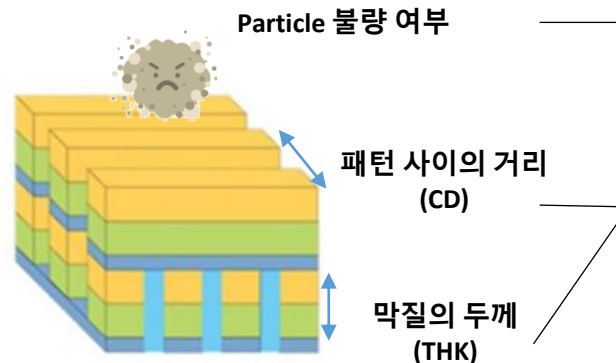
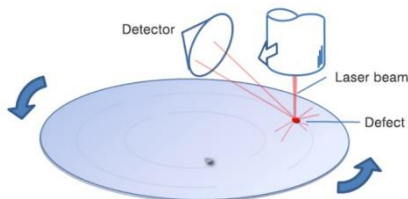
- FDC (Sensor data)
- 계측 데이터
- Fab 이력 데이터

EDS

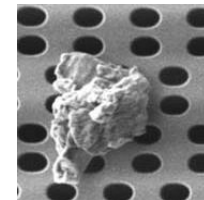
Packaging

- Wafer Bin map

- ❖ 계측 : 검사 목적에 따라 다양한 종류가 있으며, Wafer의 현재 품질 관련 정보를 수집
(Defect map / Defect Image / CD(critical Dimension) / THK(Thickness) 등)

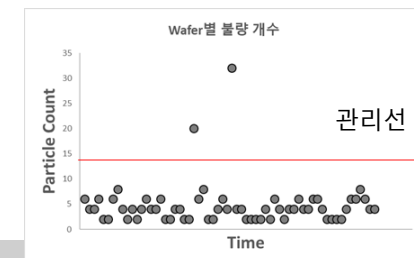


Defect map



Defect image

좌표	cd1	cd2	cd3	...
1	33	23	400	...
2	34	24	399	...
3	33	24	400	...
4	33	24	399	...



반도체 데이터 Domain

Wafer 제작

- 소재, 원자재

Fab 공정

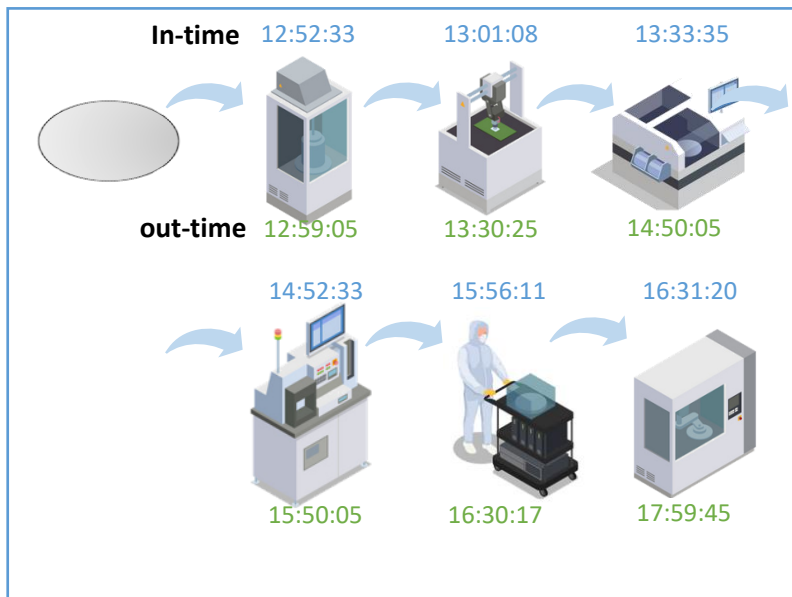
- FDC (Sensor data)
- 계측 데이터
- Fab 이력 데이터

EDS

Packaging

- Wafer Bin map

❖ Fab 이력 데이터 : Wafer가 각 step에서 설비에 in-out된 이력을 전부 저장하는 데이터



In-Time	out-time	Step	설비명	내용
2021.08.15_12:52:33	2021.08.15_12:59:05	step4	DIFF1	...
2021.08.15_13:01:08	2021.08.15_13:30:25	step5	PHOTO1	...
2021.08.15_13:33:35	2021.08.15_14:50:05	step6	ETCH2	...
2021.08.15_14:52:33	2021.08.15_15:50:05	step7	CLN4	...
2021.08.15_15:56:11	2021.08.15_16:30:17		...	...

반도체 데이터 Domain

Wafer 제작

- 소재, 원자재

Fab 공정

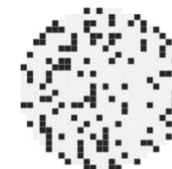
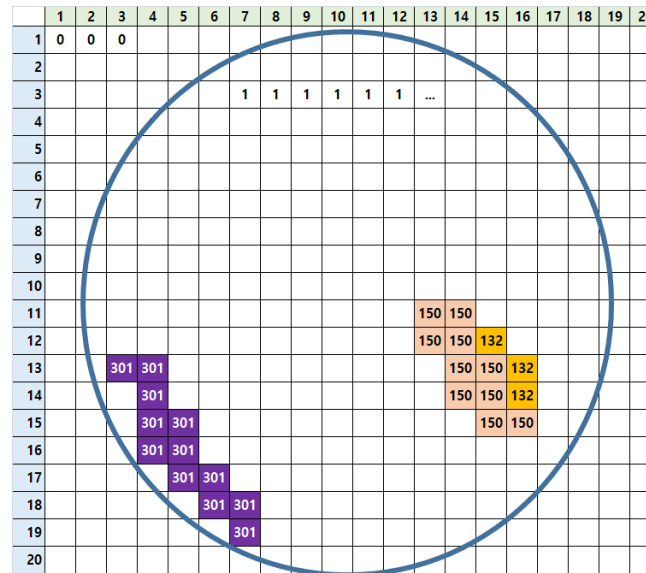
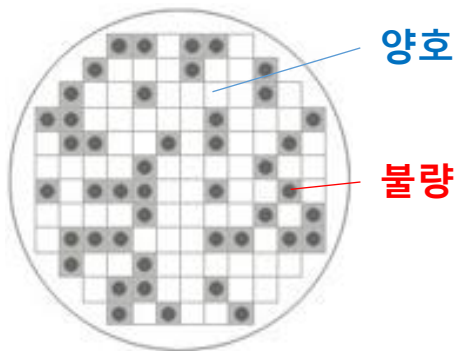
- FDC (Sensor data)
- 계측 데이터
- Fab 이력 데이터

EDS

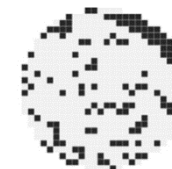
Packaging

- Wafer Bin map

- ❖ WBM(Wafer Bin Map) : EDS 검사에 의한 결과 데이터로,
각 Chip의 불량 유무를 확인하고, 불량 종류에 따라 Bin number를 부여



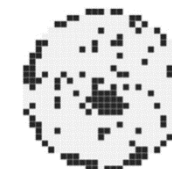
(a) None



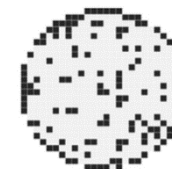
(b) Edge-local



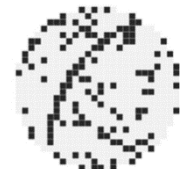
(c) Local



(d) Center



(e) Edge-ring



(f) Scratch

[n x n x1] 형태의 Image

반도체 데이터 Domain

Wafer 제작

Fab 공정

EDS

Packaging

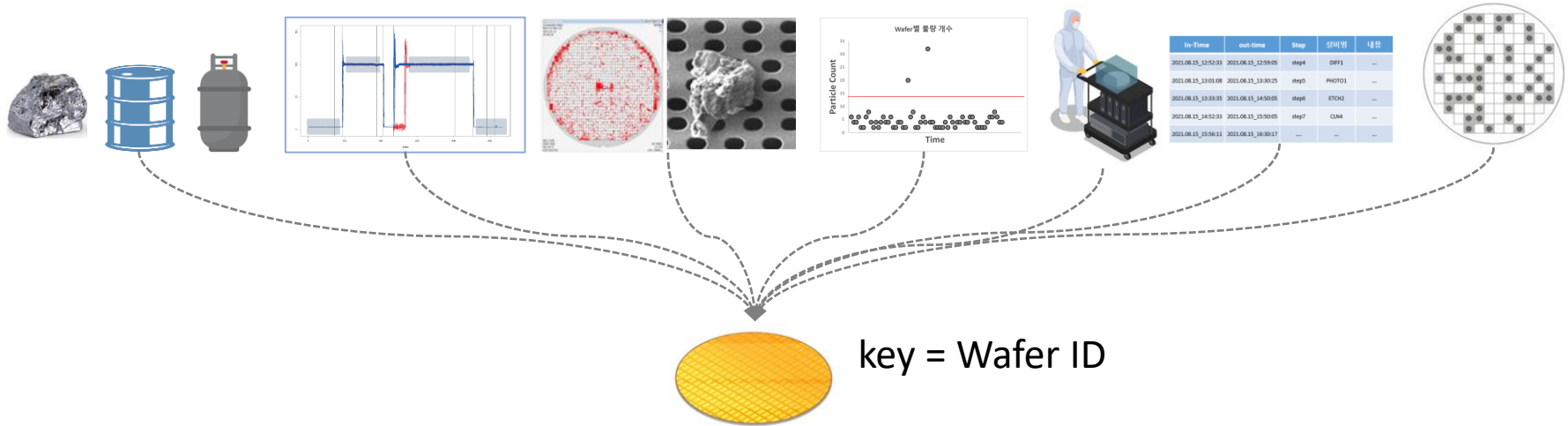
- 소재, 원자재

- FDC (Sensor data)

- 계측 데이터

- Fab 이력 데이터

- Wafer Bin map



❖ Wafer를 key로 모든 Data는 통합되어 서로의 상관관계에 대한 분석 진행.

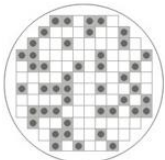
반도체 데이터 분석 예시

1. Regression : 선행하는 데이터로부터, 수율 등의 후측 데이터를 예측하는 문제

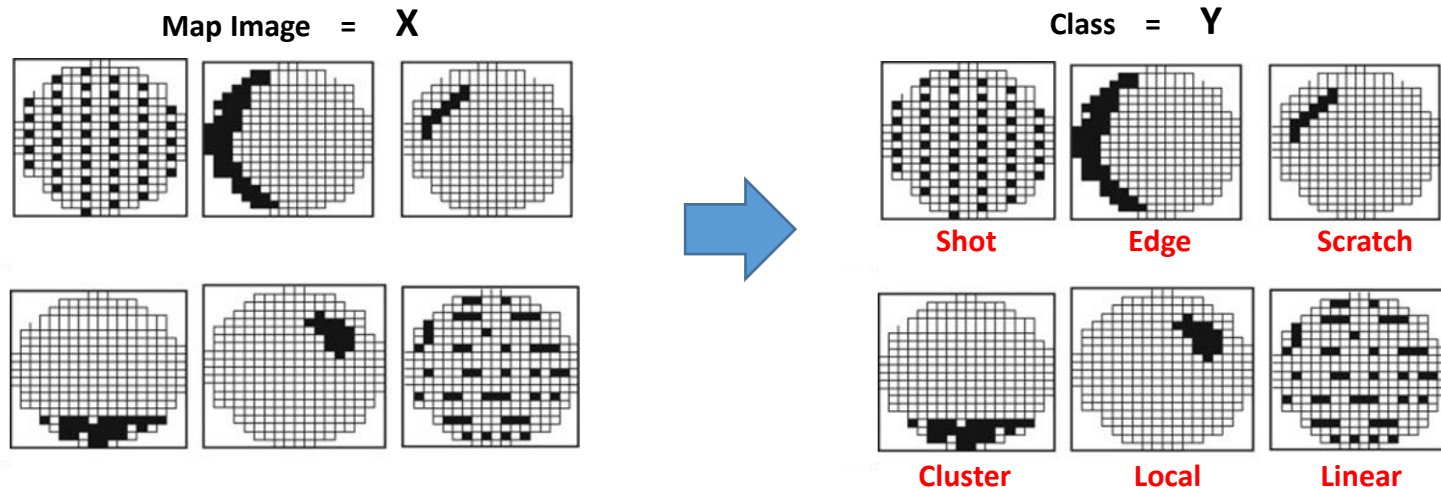
FDC + 계측 데이터 = X

Time	parameter1	parameter2	parameter3	...	작표	cd1	cd2	cd3	...
20210815111255	556	200	0.001	...	1	33	23	400	...
20210815111256	557	200	0.003	...	2	34	24	399	...
20210815111256	558	200	0.005	...	3	33	24	400	...
20210815111256	1002	232	0.005	...	4	33	24	399	...
20210815111256	1008	256	0.003	...					

수율 = Y



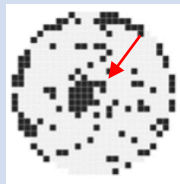
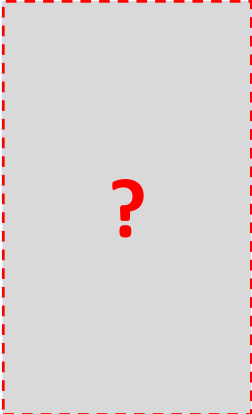
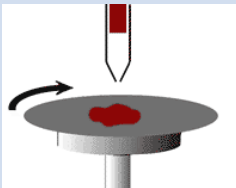
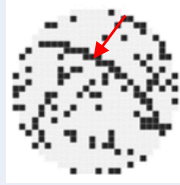
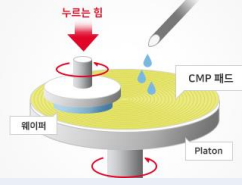
2. Classification : 불량률의 종류등을 분류하는 문제 (이상탐지 등)



Wafer Bin Map 분류 문제

❖ Wafer Bin Map의 분류 : 불량률의 원인을 찾기 위한 시작점

Map의 형태로부터, 어떤 공정에 문제가 있는지 파악

WBM	분류 Label	추정원인	관련부서	비고
			PHOTO, CLN	...
			CMP	...

☞ WBM 데이터 : daily 수천~수만매 대량 생성되지만 Label이 없기 때문에, 엔지니어에 의한 분류



Unlabeled Data

Label	발생량
Normal	
Center	
Edge	
Neck	
Scratch	
...	

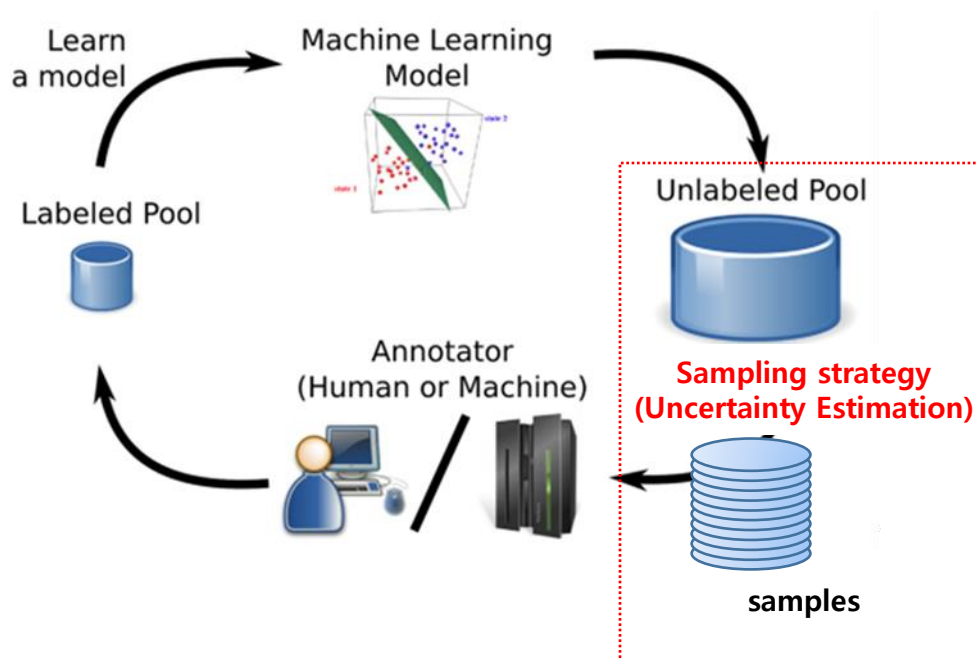
Classification

▶ 학습에 필요한 Labeling 비용을 최소화 필요

Active Learning

❖ Active Learning

- : 소량의 Labeled Data를 통해 모델을 구축하고, 이후 학습에 더 좋은 Sample을 선별해서 update하여 단기간에 일정 수준 이상의 성능을 확보하는 준지도학습의 일종
- : 사람이 관여하여 Update하는 부분이 존재하여, 인공지능과 사람이 협동한다는 특징을 가짐.



[1 Phase]

- ① 소량의 Labeled data로 모델 구축.
- ② 모델에 Unlabeled data를 input으로 넣음.
- ③ 학습에 더 좋은 sample을 계산.

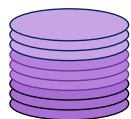
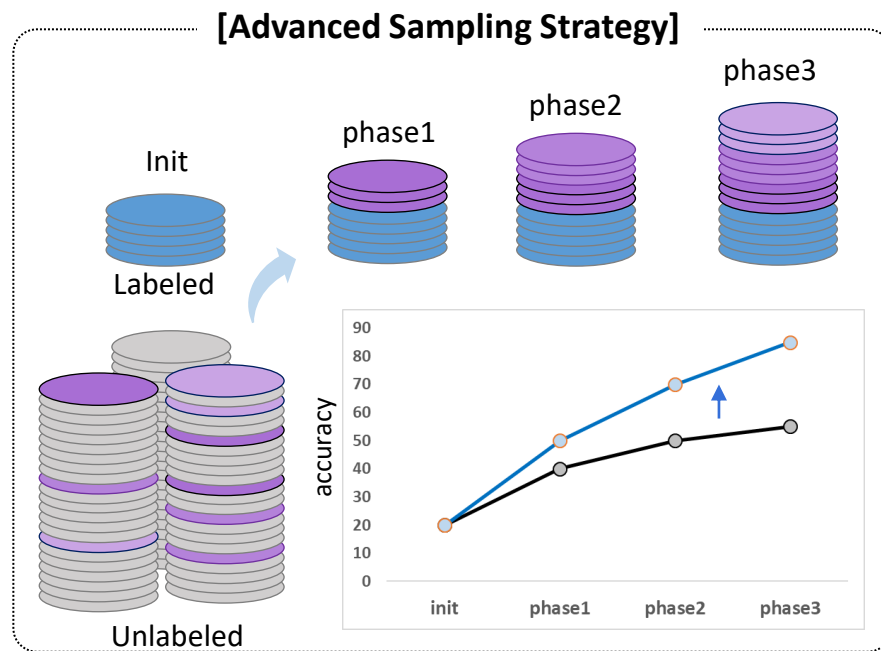
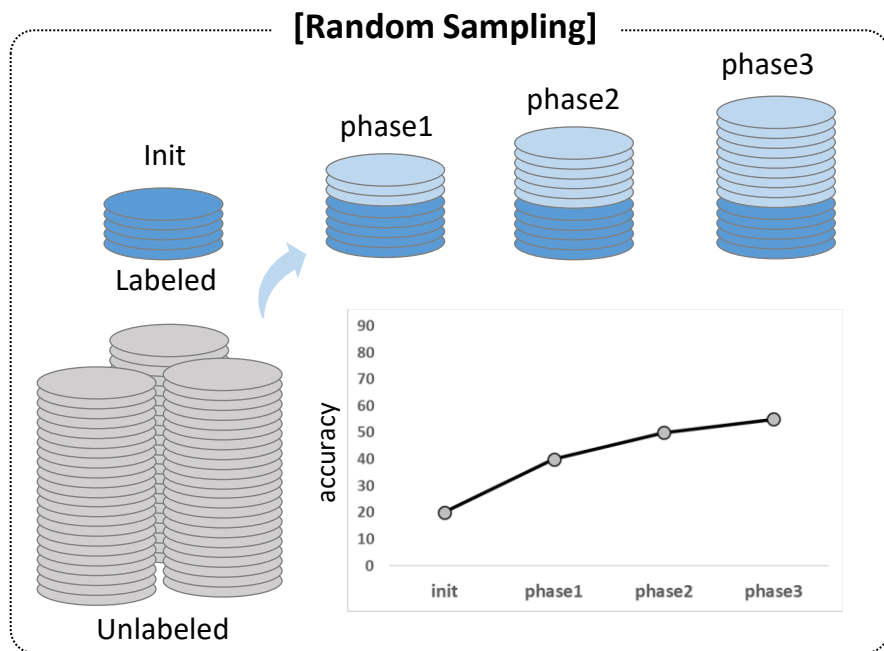
(Sampling strategy or Uncertainty Estimation)

- ④ 해당sample을 엔지니어가 분류하여 Labeled로 update.
- ⑤ Update된 Data로 모델 재학습하여 성능 향상.

▶ Phase를 반복하여 계속해서 성능을 높여감

Active Learning

- ❖ 모델의 성능을 더 많이 향상시킬 수 있는 "Useful Sample"이 존재한다고 보고, 이를 우선적으로 골라 학습시키면, 더 빠르게 성능을 확보할 수 있음.



Useful Sample의 기준?

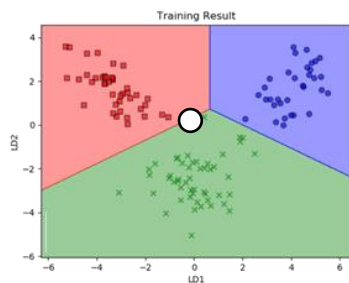
Sampling Strategy

❖ Sampling Strategy :

- Heuristic Approach
- Ensemble Approach
- Distribution Approach
- Bayesian Approach

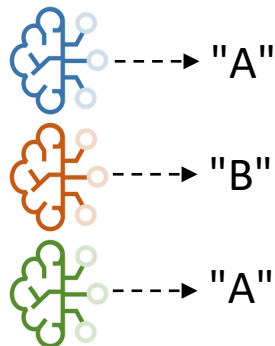
Useful Sample을 정의하는 기준은 접근 방식에 따라 다름

[Heuristic]



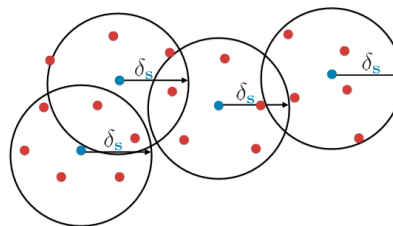
확실하지 않은
예측확률을 갖는 Sample

[Ensemble]



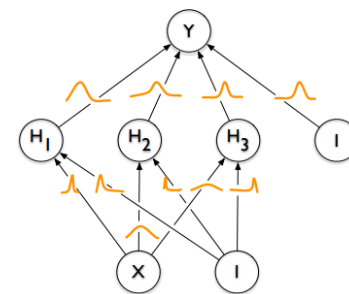
여러 모델이
서로 다르게 예측하는
어려운 Sample

[Distribution]



다른 Sample들을 대표하는
중심 Sample

[Bayesian]



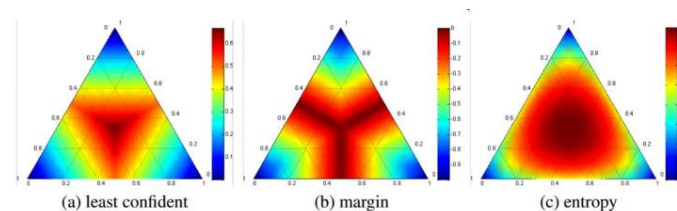
Bayesian Neural Network의
Uncertainty가 큰 Sample

Sampling Strategy

❖ **Heuristic Approach** : 가장 고전적인 형태, Class 분류에 대한 확률값으로부터 판단.



sample	Class A	Class B	Class C
①	0.9	0.05	0.05
②	0.3	0.5	0.2



Least Confidence

$$lc_i = -\max_j p_j(\mathbf{X}_i, y_i).$$

① : 0.9

② : 0.5

→ 0.9 > 0.5 이므로 더 작은 ② 선택

Margin

$$lm_i = -(p_{j_1}(\mathbf{X}_i, y_i) - p_{j_2}(\mathbf{X}_i, y_i)).$$

① : 0.9-0.05

② : 0.5-0.3

→ 0.85 > 0.2 이므로 더 작은 ② 선택

Entropy

$$en_i = -\sum_{j=1}^C p_j(\mathbf{X}_i, y_i) \log p_j(\mathbf{X}_i, y_i)$$

① : $-(0.9 \log 0.9) + (0.05 \log 0.05) + \dots = 0.171$

② : 0.447

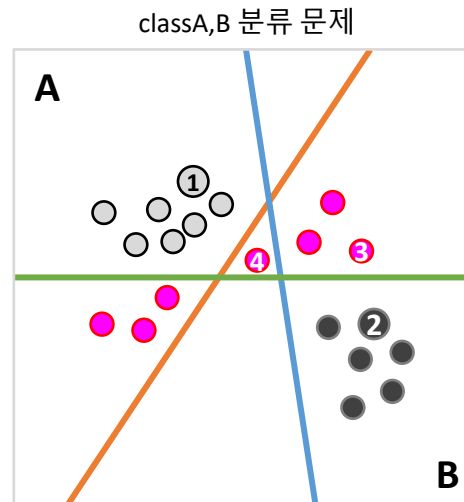
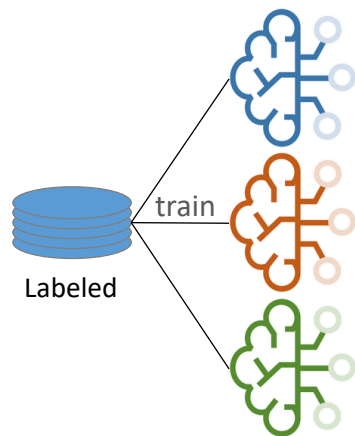
→ 0.171 < 0.447 이므로 더 큰 ② 선택

Sampling Strategy

- ❖ **Ensemble Approach** : 동일 sample에 대한 여러 모델들의 결과variation을 통해 Uncertainty를 정의

Query-by Committee

각각 학습시킨 모델들이 투표를 진행하고, 그 결과가 가장 엇갈린 Sample 선택.



Sample	M_Blue	M_Red	M_Green
①	A	A	A
②	B	B	B
③	B	B	A
④	A	B	A

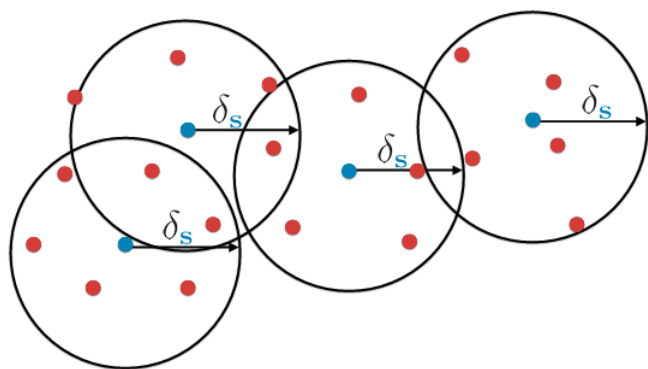
[Vote-entropy]

$$x_{VE}^* = \operatorname{argmax}_x - \sum_i \frac{V(y_i)}{C} \log \frac{V(y_i)}{C}$$

Sampling Strategy

- ❖ **Distribution approach** : Sample의 분포 내에서 다른 Sample을 대표하는 주요한 Sample을 선택

Core-Set



- ① ●을 대표하는 Center ●을 선정.
- ② ●을 중심으로 다른 sample을 커버하는 Distance δ 를 최소화하도록 Optimization 진행.
- ③ 이후, 전체 sample들의 Distribution을 전부 커버하게 된 ●을 선택

$\{\bullet\}$ is δ -cover of $\{\bullet\}$

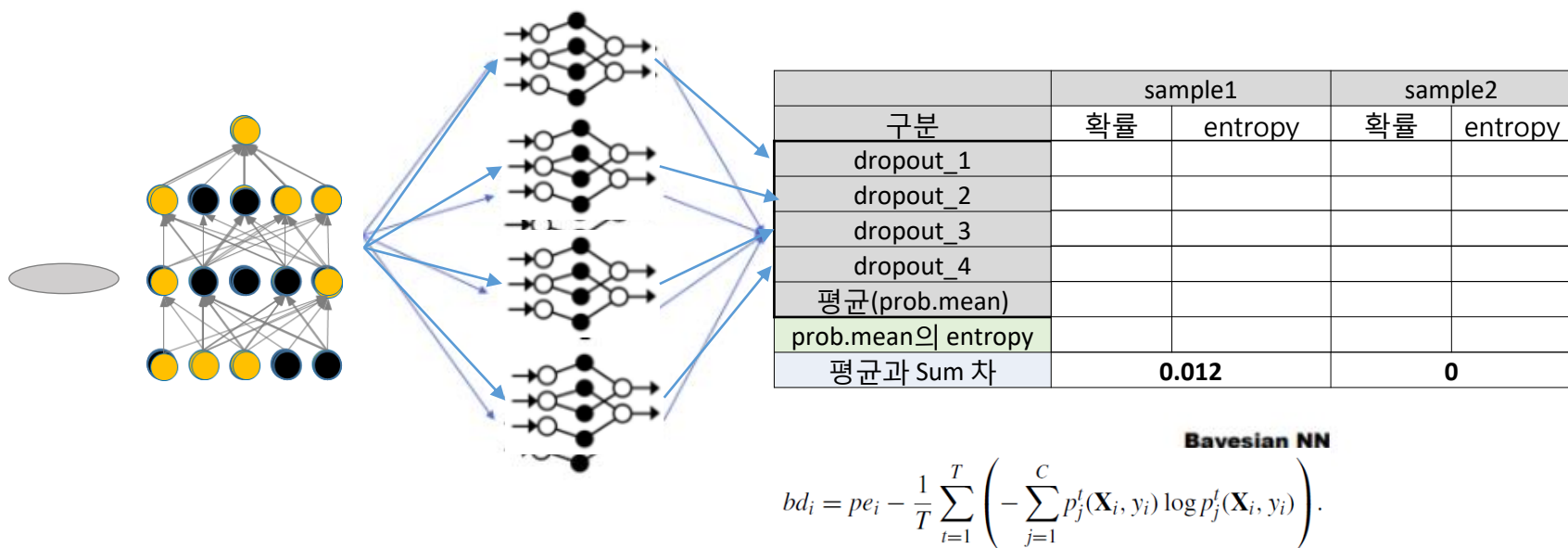
$$\{x\} = \min_{\{x\}} \delta$$

Sampling Strategy

- ❖ **Bayesian approach** : Bayesian Neural Network의 Uncertainty를 기준으로 Sample을 선택

BALD_Dropout (Bayesian Active Learning by Disagreement)

: Drop out을 시도할 때 마다, 확률의 변동이 큰 Sample이 Uncertainty가 높은 Sample으로 선택



Active Learning in Semiconductor Manufacturing

❖ Active Learning of Convolutional Neural Network for Cost-Effective Wafer Map Pattern Classification

- Jaewoong Shim; Seokho Kang; Sungzoon Cho
- IEEE Transactions on Semiconductor Manufacturing, May 2020 , 7회 인용

Active Learning of Convolutional Neural Network for Cost-Effective Wafer Map Pattern Classification

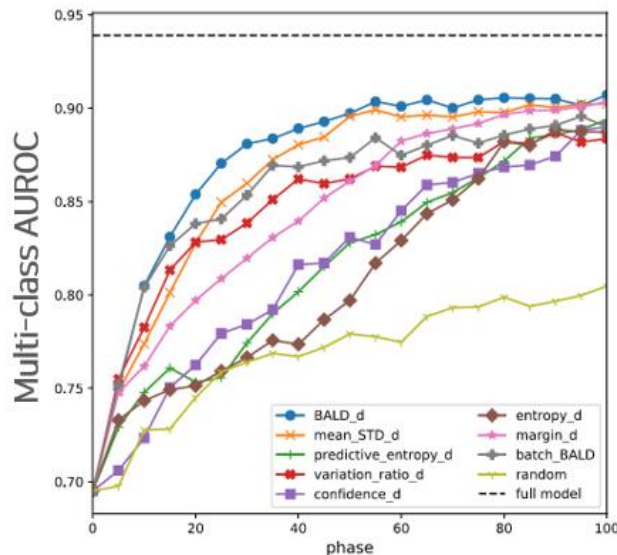
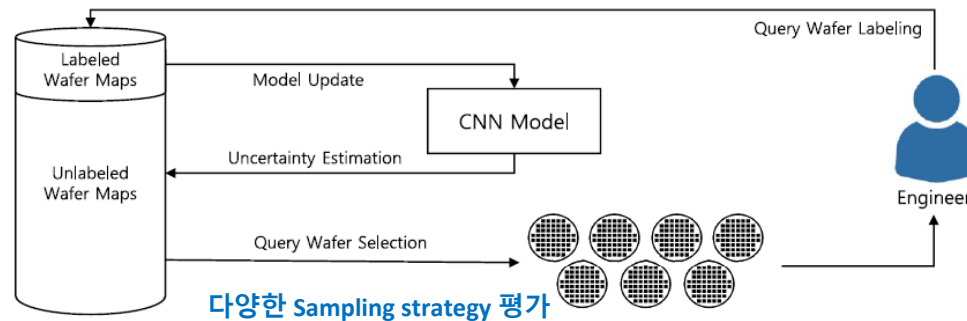
Jaewoong Shim^{id}, Seokho Kang^{id}, and Sungzoon Cho^{id}

Abstract—Wafer maps provide important information for engineers for detecting root causes of failure in a semiconductor manufacturing process. Thus, there has been active research into the automation of wafer map pattern classification. With recent advances in deep learning, a convolutional neural network (CNN) has yielded state-of-the-art performance in wafer map pattern classification. Because a large amount of labeled training data is required, experienced engineers need to annotate large quantities of wafer maps manually which is costly. To construct a well-performing CNN model with a lower labeling cost, we propose a cost-effective wafer map pattern classification system based on the active learning of a CNN. In the system, a CNN

The wafer map provides engineers with very important information for root cause identification. Wafers with similar spatial failure patterns are likely to have the same root cause in the fabrication process. For example, if wafers with the same wafer map pattern appear more frequently than a certain level in a short period of time, engineers can take action, e.g., finding a common fabrication facility or process for those wafers. Therefore, wafer map pattern classification is an important task for finding the cause of a failure, and provides clues to improve the manufacturing quality.

Active Learning in Semiconductor Manufacturing

❖ Wafer Bin map의 분류문제에 다양한 Sample Strategy 평가



BALD(Bayesian Dropout) 방법론이 가장 좋은 성능

Random Sampling 대비 대부분의 방법론이 효과적임

Conclusion

❖ 반도체 제조 공정과 데이터 Domain

- 대량의 Unlabeled Data가 생성되어 엔지니어 분석 업무로 가중됨.
- 이를 머신러닝으로 학습시킬 때, Labeling에 대한 비용 절감 필요.
(Wafer Bin Map Classification 문제)

❖ Active Learning in Semiconductor

- 다양한 Sampling Strategy를 통해 최적의 Sample만을 골라내는 Active learning은 Labeling에 대한 시간과 비용을 절약시켜줄 수 있는 좋은 방안이 될 수 있음.
- 다만 여전히 엔지니어가 직접 관여하는 부분은 개선될 여지가 있음.
- Classification을 돕는 별도 모듈을 추가하거나, 모델의 성능 자체를 올릴 수 있는 다양한 시도들을 병행한다면 더 효율적으로 사용할 수 있을 것으로 보임.

감사합니다.
