
기계 학습을 이용한 셀 문턱 전압 분포 기반의 Soft Decision 센싱 레벨 예측 방법

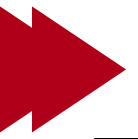
(Prediction Method of Soft Decision Sensing Level

Based on Distribution of Cells Threshold Voltage Using Machine Learning)

노해동, 백준걸 †

고려대학교 산업경영공학과 석사과정 / 고려대학교 산업경영공학과 교수†

nhd0113@korea.ac.kr / jungeol@korea.ac.kr †



Contents

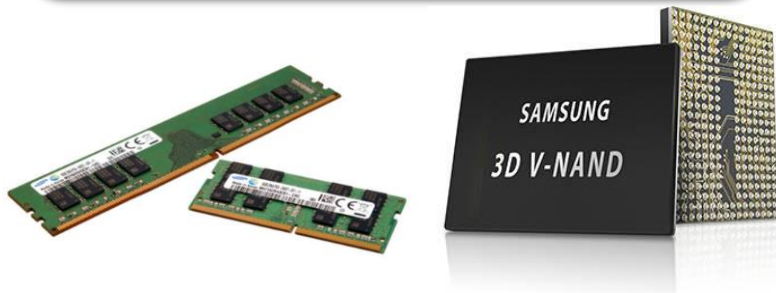
1. 서론
2. 제안 방법
3. 실험 결과
4. 결론

1. 서론

낸드 플래시 메모리 시장

• 제품군

- ✓ SSD(Enterprise , Server 용)
 - ✓ Mobile Storage Device
- 빠른 접근 속도 + 비휘발성 메모리



• 시장 규모

- ✓ '20년 \$732.7억(약 9조원) 규모 예상
- ✓ 지속 증가 전망

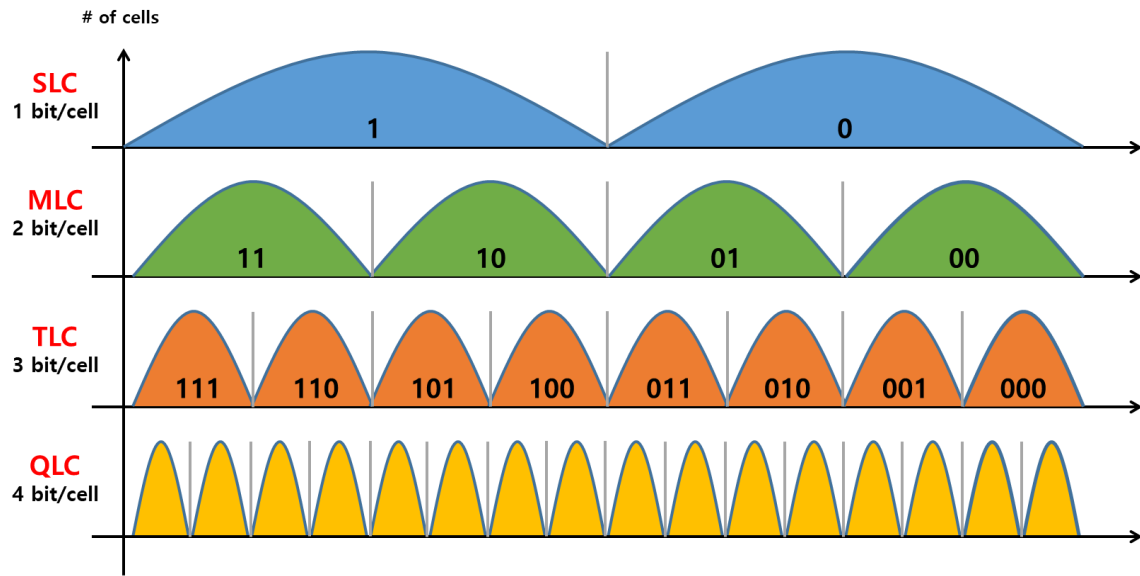


1. 서론

낸드 플래시 Cell의 발전

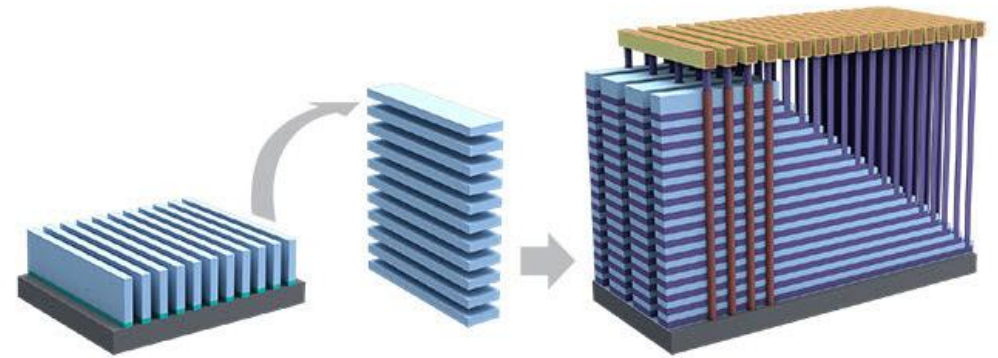
• Cell 당 저장 bit 증가

✓ SLC → MLC → TLC → QLC



• 3D 낸드 플래시 적층 증가

✓ 공정 미세화 동반



2D to 3D

3D NAND

그림 출처 : <https://www.etnews.com/20200217000106>

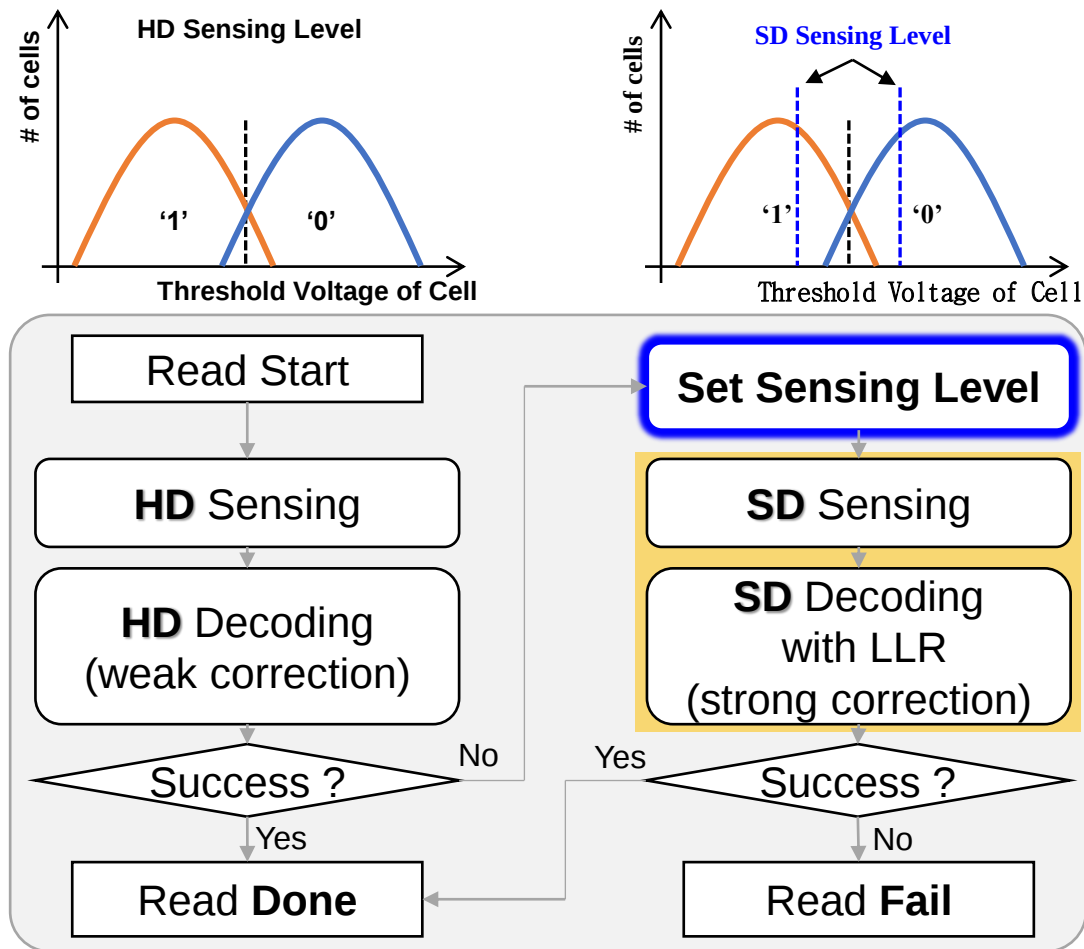
☞ 분포도 겹침과 지속적인 공정 미세화에 따른 **데이터 Error 발생률 급증**인 상황에서
저장 장치의 신뢰성과 수명 유지를 위해서는 **강한 오류 정정 코드 필요**

1. 서론

강한 오류 정정 코드 : LDPC

• Low-Density Parity-Check (LDPC) 특징

- ✓ 지난 10여년간 낸드 플래시 메모리에 널리 사용
- ✓ Error-Correction-Code(ECC)의 한 종류
- ✓ Hard Decision(HD) 체계 디코딩 수행
- ✓ **Soft Decision(SD)** 체계 디코딩 추가 수행



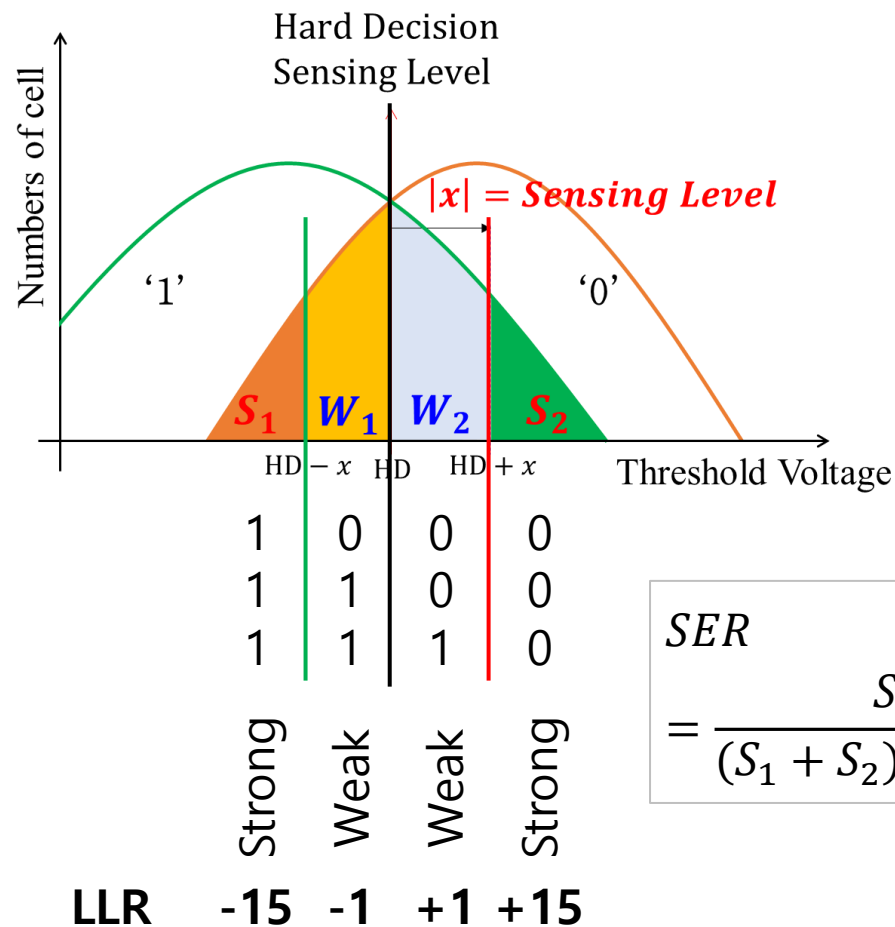
☞ SD 체계 수행으로 **강력한 오류 정정 능력이 보장**되기 위해서는 **최적 Sensing Level 필수적**

1. 서론

Sensing Level과 SER의 관계

• Soft Decision 체계 디코딩

- ✓ Sensing Level(SL)을 기준으로 Cell Sensing을 통해 Log-Likelihood-Ratio(LLR) 정보 획득
 - SL 설정에 따른 SER의 변화
- ✓ 획득한 LLR 정보를 추가하여 디코딩 수행
 - 강력한 오류 정정 능력 발휘 가능
- ✓ SER은 전체 에러 중 Strong 에러 비율 의미
 - SER은 셀 분포 의미



SER

$$= \frac{S_1 + S_2}{(S_1 + S_2) + (W_1 + W_2)}$$

👉 셀 분포가 반영된 Sensing Level 설정이 필수적

1. 서론

낸드 플래시의 다양한 셀 분포

• Physical 특성에 따른 차이

- ✓ 3D 낸드 플래시 적층 수는 지속 증가 추세
- ✓ Word Line(WL)에 따른 비트 오류율 차이 확인

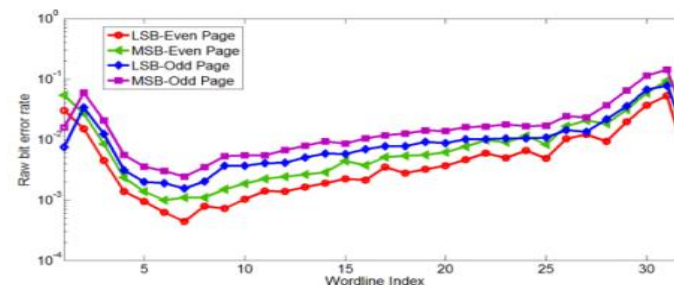


Figure 2. Physical Location and Retention Error (Mizoguchi *et al.* 2017)



• 열화 요인에 따른 차이

- ✓ 쓰고 지우지 동작에 의한 PE Cycle 영향
- ✓ 데이터 방치 시간에 의한 열화
- ✓ 셀 열화의 수준 차이의 증가 추세 확인

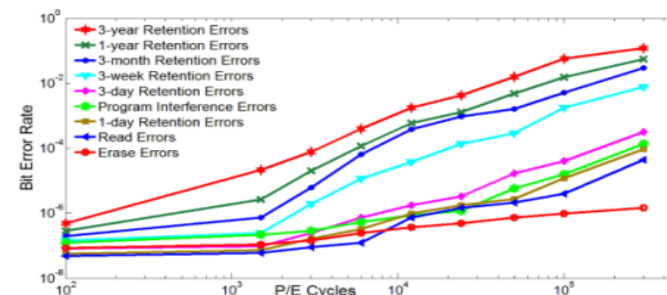


Figure 1. PE Cycle and Retention Error (Mizoguchi *et al.* 2017)



👉 물리적 특성과 열화 요인의 **중첩**으로 **더욱 다양한 셀 분포가 존재**

1. 서론

기존 모델과 한계점

• 임베디드 제품의 제한적인 상황

- ✓ 고성능 요구에 따른 응답시간의 제한 존재
→ SL 가용 개수 제한적 (try 횟수에 제한적 : 3N)

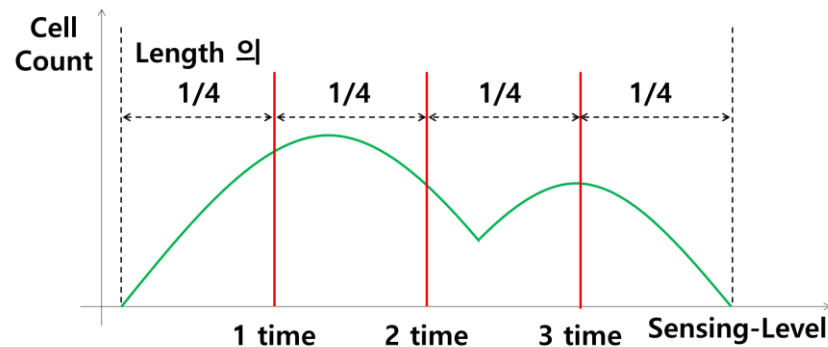
• 균등 분할 모델

- ✓ 테스트를 통해 확보된 다양한 셀 문턱 전압 분포 확보
- ✓ 분포의 Range 구간 확인 → 가용 SL 개수로 균등 분할
- ✓ 제한된 개수로 분포에 대해 낮은 정확도와 큰 오차 발생 문제

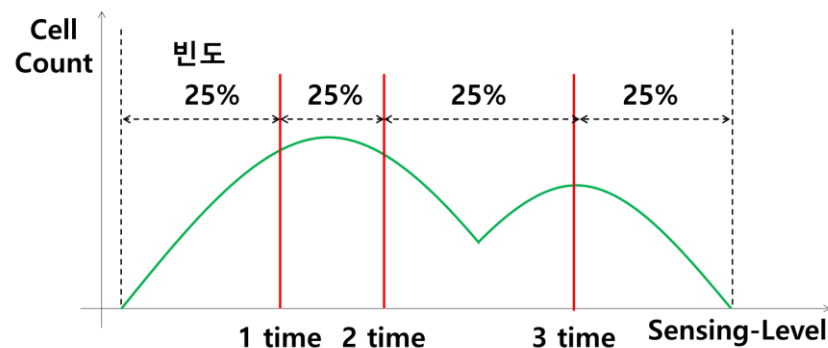
• 확률 기반 모델

- ✓ 제한된 SL 값 개수로 인한 오차 문제 해결을 위해 확률 개념을 도입
- ✓ SL 값의 분포를 고려하여 빈도가 높은 SL 값들을 후보군으로 선정

균등 분할



확률 기반



☞ 제한적인 SL 개수의 상황에서 기존 모델은 **낮은 정확도와 큰 오차**에 대한 **한계 존재**

1. 서론

Motivation

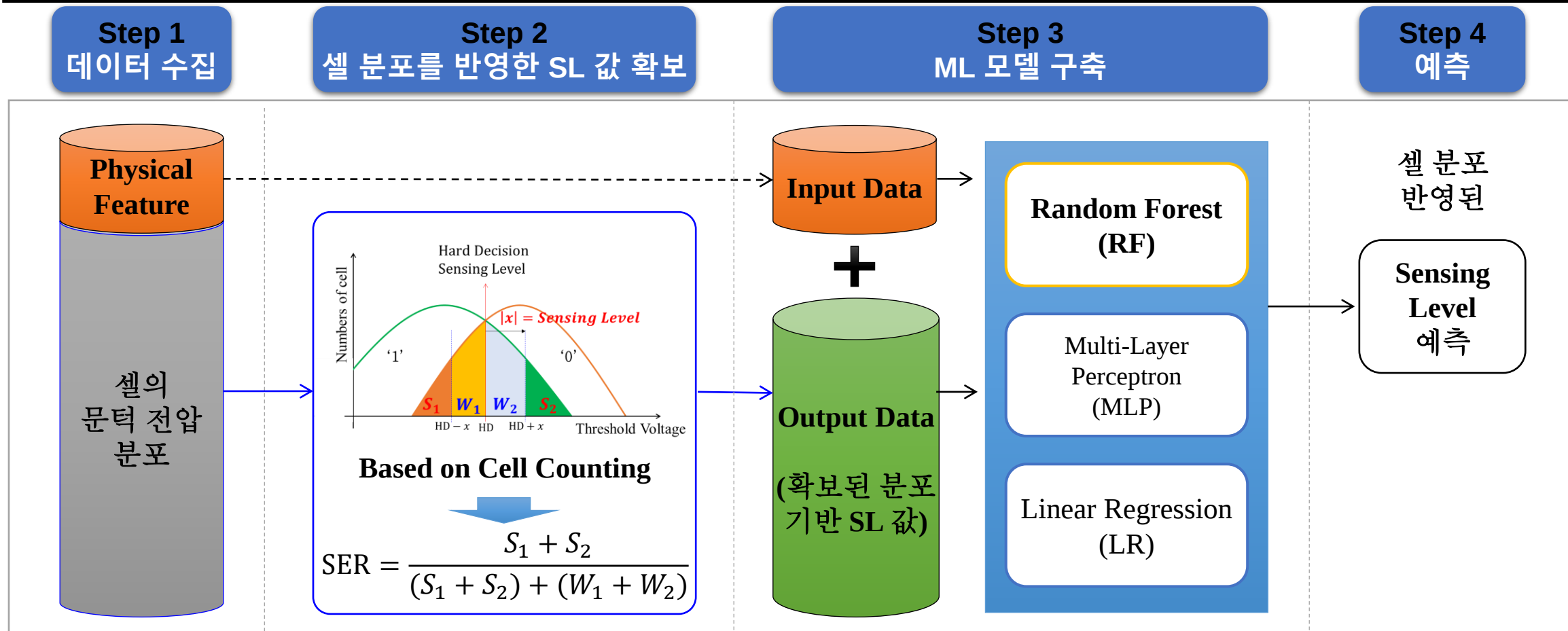
임베디드 제품의 제한적인 상황 속에서

(1) 다양한 분포에 대한 반영 모델

(2) 한번의 예측으로 높은 정확도와 낮은 오차의 모델

→ 셀 분포가 반영된 출력변수와 입력변수로 ML 모델 학습

2. 제안 방법



Step 1. 다양한 조건의 Cell의 Physical 특성에 대해 열화 실험을 통해 **데이터 수집**

Step 2. 셀 개수 산정 방식을 통해 설정 **SER**에 해당하는 셀 문턱 전압 분포가 반영된 **SL 값을 확보**

Step 3. Physical Feature를 독립 변수로, 확보된 SL 값을 결과 변수 사용하여 **ML 모델 구축**

Step 4. 구축된 ML 모델을 통해 셀의 **Physical Feature 값**으로만 **Sensing Level 예측 수행**

2. 제안 방법

방치 열화 실험을 통한 데이터 수집

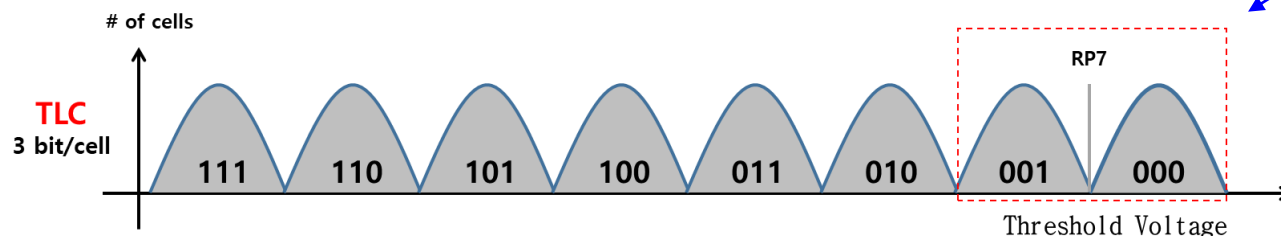
• 데이터 설명

- ✓ 실제 3D 낸드 플래시 메모리 데이터 개수 : 331,190 (학습 : 70% / 테스트 30%)
- ✓ PE Cycle, Retention Time 열화에 따라 24hr 주기로 데이터 수집
- ✓ 다양한 분포와 분산이 존재하는 RP7 데이터 활용
 - Physical Feature : 낸드 읽기 동작 과정에서 확보 되는 Feature
 - 셀의 문턱 전압 분포
- ✓ Min-Max 정규화 수행
- ✓ 전진, 후진법을 이용하여 중요변수 선택

Step 1 데이터 수집

Physical
Feature

셀의
문턱 전압
분포



2. 제안 방법

SER별 셀 분포 기반의 SL 값 확보

• Sensing Level 값 산출 방법

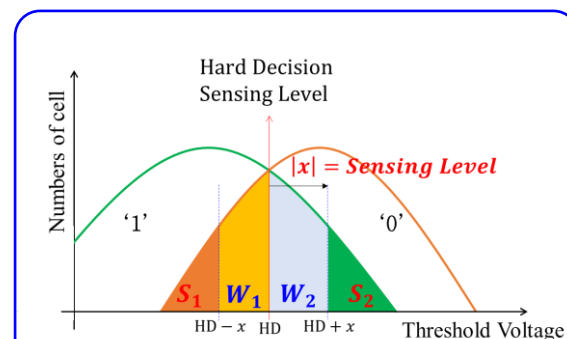
- ✓ S_1, W_1, W_2, S_2 는 HD 레벨을 기준으로 판단되는 Cell 개수
- ✓ $(S_1 + W_1 + W_2 + S_2) =$ 전체 Error Bit 개수
- ✓ Strong Error Ratio(SER)은 전체 Error 중에 Strong Error의 비율 의미
- ✓ SD SL에 따라서 S_1, W_1, W_2, S_2 의 변동이 동반되고, 곧 SER의 변화

- 셀 분포를 반영 의미

- ✓ 설정 SER 값에 맞는 SL 값 확보
- ✓ 해당 실험에서는 SER 30%, 40%, 50%에 대해 진행
 - 설정 SER 변경에도 ML 모델 구축 가능 여부 확인 가능

Step 2

셀 분포를 반영한 SL 값 확보



Based on Cell Counting

$$SER = \frac{S_1 + S_2}{(S_1 + S_2) + (W_1 + W_2)}$$

2. 제안 방법

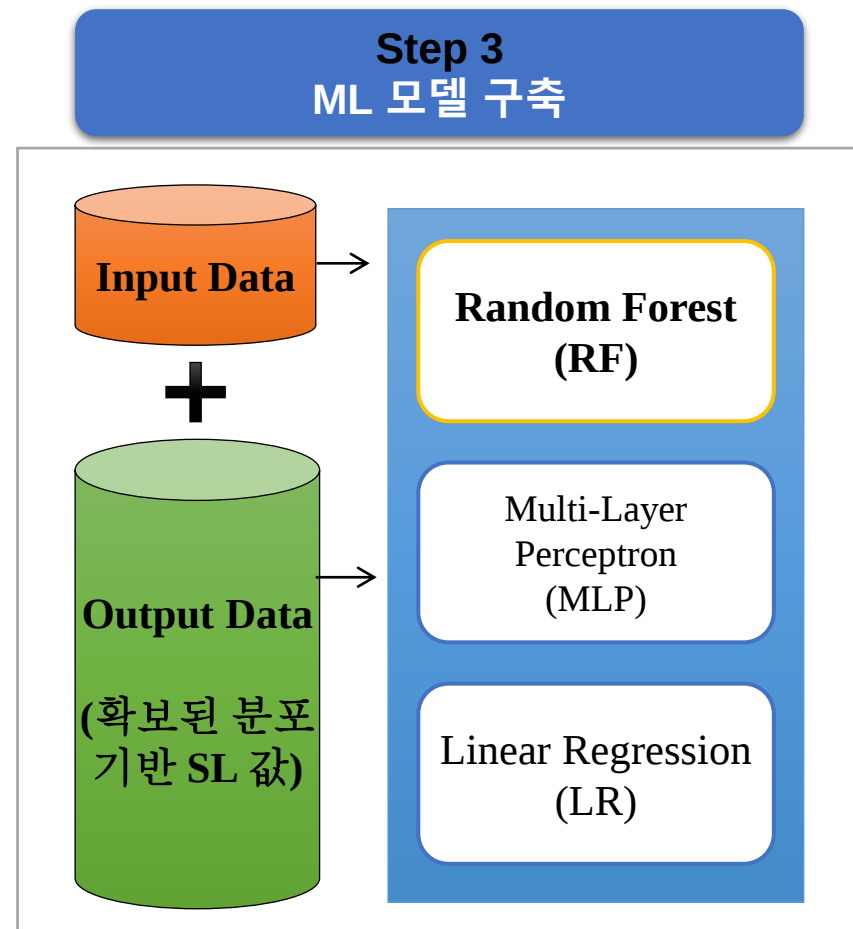
ML 모델 구축 및 예측

• ML 모델 구축

- ✓ ML 예측 회귀 모델은 독립변수와 결과변수의 관계 알고리즘을 개발과 새로운 입력변수에 대한 출력변수의 예측이 가능
- ✓ [Step 1]에서 읽기 동작에서 확보되는 **Feature**를 독립 변수로 사용
- ✓ [Step 2]에서 셀 분포가 반영된 SL 값을 **결과 변수**로 사용
- ✓ Random Forest(RF), Multi-Layer Perceptron(MLP), Linear Regression(LR) 모델을 통해 실험 진행

• ML 모델 예측

- ✓ 추가적인 리소스가 소요되지 않는 Physical Feature를 입력으로 사용하여 구축된 모델을 통한 SD Sensing Level 예측 수행



2. 제안 방법

평가 지표

• 성능 지표 MAE, MSE

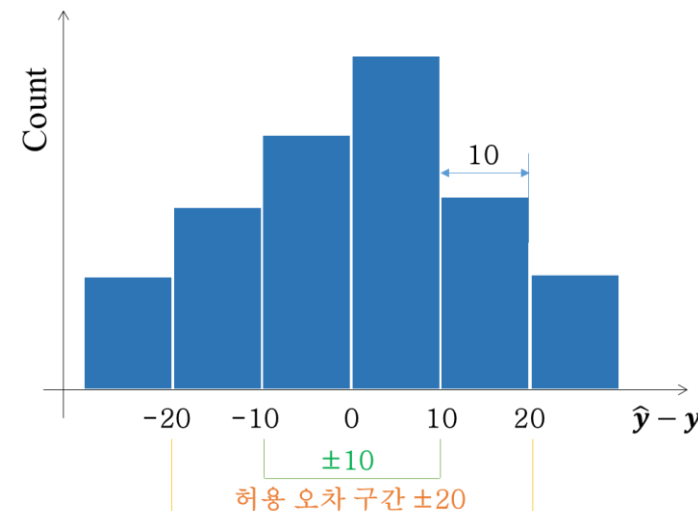
- ✓ y 는 실제 값, $\hat{y}$ 는 예측 값, n 은 전체 개수를 의미한다.
- ✓ 오차 크기의 평균을 의미하는 MAE는 오차에 절대값을 취하기 때문에 직관적으로 오차를 알 수 있는 평가 지표
- ✓ MAE와 MSE는 회귀 모델의 대표적인 평가 지표

$$MAE = \frac{\sum |y - \hat{y}|}{n}$$

$$MSE = \frac{\sum (y - \hat{y})^2}{n}$$

• 오차 허용 구간을 통한 정확도 확인

- ✓ 백분율을 통해 직관적인 성능 수준 확인 용이함
- ✓ 예측 값에서 실제 값을 뺀 오차 값을 0을 기준으로 허용 구간 설정
- ✓ 허용 구간의 개수를 전체 개수로 나누어 %로 확인
- ✓ 해당 실험에서는 6가지의 허용 구간 10, 15, 20, 25, 30, 35mV로 진행



3. 실험 결과

실험 방향

- (1) 모델 간 성능 비교
- (2) 설정 SER에 대한 ML 모델 성능 확인
- (3) 다구간 확률 기반 모델과 성능 비교

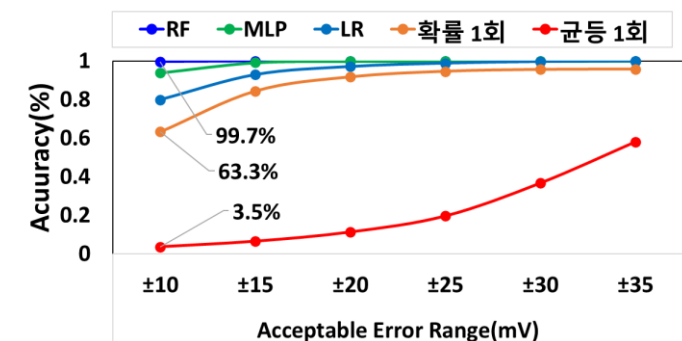
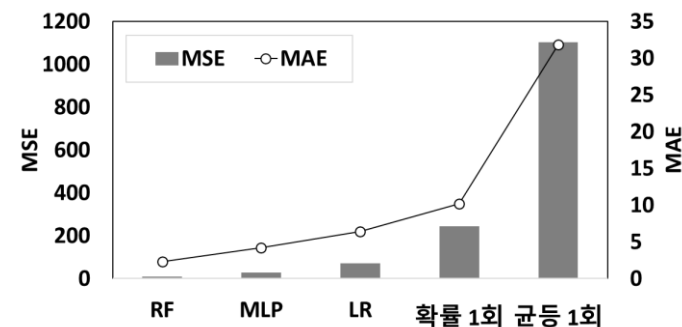
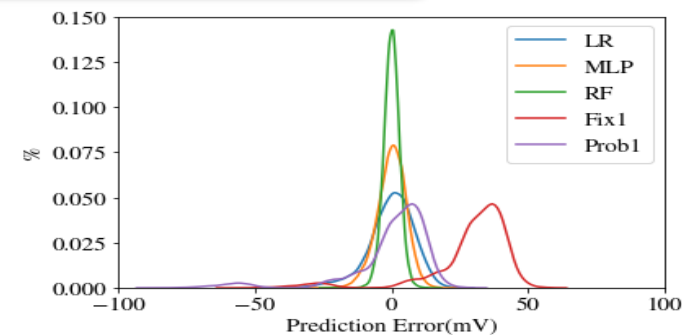
3. 실험 결과

(1) 모델 간 성능 비교

- ✓ 모든 평가 지표에서 **ML 모델 예측 성능의 우수함** 확인
- ✓ RF > MLP > LR > 확률 > 균등 순서로 예측 성능 우수함을 확인
 - RF 성능 : MSE 8.7, MAE 2.3, 정확도 99.7%(±10mV기준)

👉 **ML 예측 모델 구축 가능**

SER	모델	구분	오차 평가 지표		오차 허용 구간별 정확도 ↑ (%)					
			MSE↓	MAE↓	±10mV	±15mV	±20mV	±25mV	±30mV	±35mV
50%	ML	RF	8.7	2.3	99.7%	100.0%	100.0%	100.0%	100.0%	100.0%
		MLP	28.7	4.2	93.8%	99.2%	99.9%	99.9%	100.0%	100.0%
		LR	70.7	6.4	79.9%	92.9%	97.1%	98.8%	99.7%	100.0%
	확률	1 회	244.5	10.2	63.3%	84.4%	91.7%	94.6%	95.6%	95.7%
	균등	1 회	1102.1	31.8	3.5%	6.4%	11.2%	19.6%	36.7%	58.1%



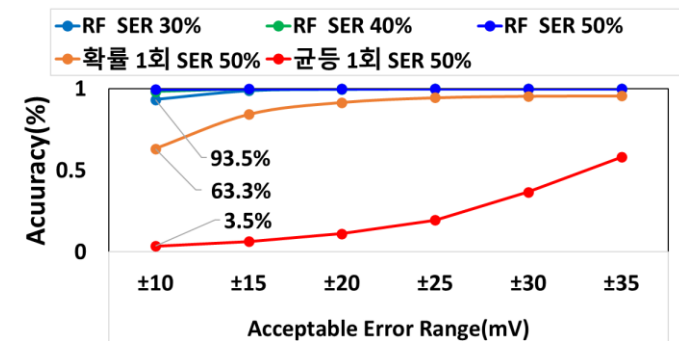
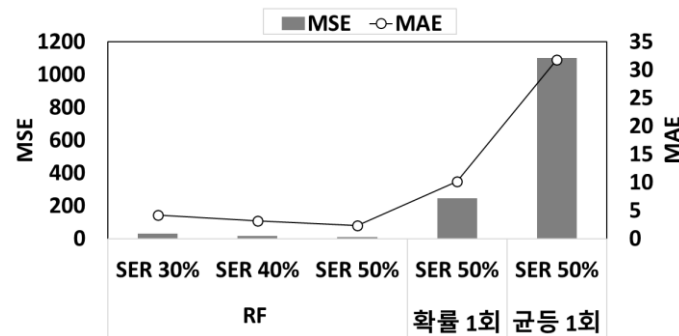
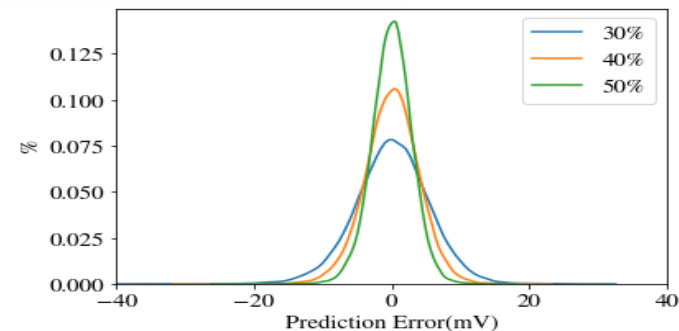
3. 실험 결과

(2) 설정 SER에 대한 ML 모델 성능 확인

- ✓ 가장 우수한 ML 모델인 RF로 SER 30, 40, 50% 성능의 우수함 확인
- ✓ SER별 최소 93.5%이상의 정확도 확보 가능
 - ML 모델 정확도(93.5 ~ 99.7%), 확률 정확도(63.3%) - ($\pm 10\text{mV}$ 기준)

👉 설정 SER별 예측 모델 구축 가능

SER	모델	구분	오차 평가 지표		오차 허용 구간별 정확도 $\uparrow$ (%)					
			MSE $\downarrow$	MAE $\downarrow$	$\pm 10\text{mV}$	$\pm 15\text{mV}$	$\pm 20\text{mV}$	$\pm 25\text{mV}$	$\pm 30\text{mV}$	$\pm 35\text{mV}$
30%	ML	RF	29.7	4.2	93.5%	98.9%	99.8%	99.9%	100.0%	100.0%
40%	ML	RF	15.8	3.1	98.3%	99.8%	100.0%	100.0%	100.0%	100.0%
50%	ML	RF	8.7	2.3	99.7%	100.0%	100.0%	100.0%	100.0%	100.0%
	확률	1회	244.5	10.2	63.3%	84.4%	91.7%	94.6%	95.6%	95.7%
	균등	1회	1102.1	31.8	3.5%	6.4%	11.2%	19.6%	36.7%	58.1%



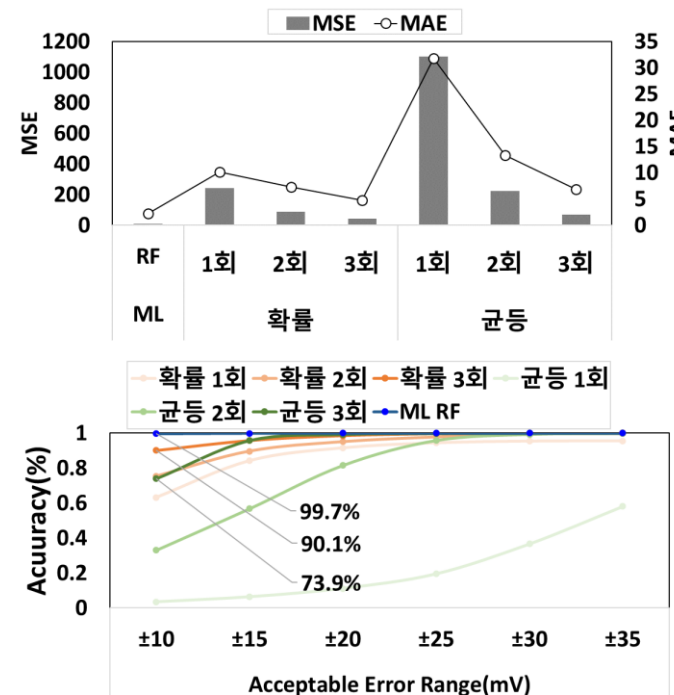
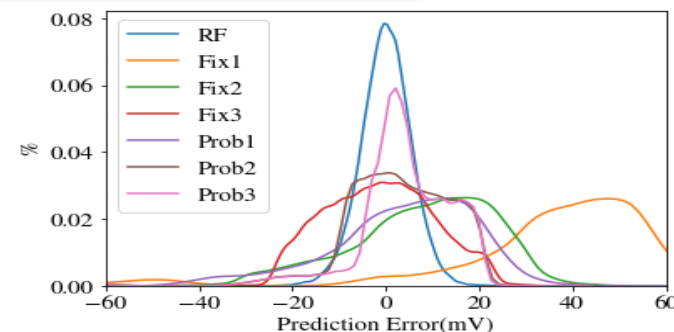
3. 실험 결과

(3) 다구간 확률 기반 모델 성능 확인

- ✓ 모든 평가 지표에서 ML 모델 예측 성능의 우수함을 확인
- ✓ RF > 확률 3회 > 균등 3회 순서로 예측 성능 우수함을 확인
 - RF는 1회 예측을 통한 성능이 확률기반 3회보다 우수함 확인

☞ 다구간 확률 기반보다 우수한 성능의 예측 모델 구축 가능

SER	모델	구분	오차 평가 지표		오차 허용 구간별 정확도 ↑ (%)					
			MSE↓	MAE↓	±10mV	±15mV	±20mV	±25mV	±30mV	±35mV
50%	ML	RF	8.7	2.3	99.7%	100.0%	100.0%	100.0%	100.0%	100.0%
	확률	1회	244.5	10.2	63.3%	84.4%	91.7%	94.6%	95.6%	95.7%
		2회	89.6	7.3	75.4%	89.7%	95.2%	97.9%	99.5%	100.0%
		3회	43.0	4.8	90.1%	95.9%	98.6%	99.8%	100.0%	100.0%
	균등	1회	1102.1	31.8	3.5%	6.4%	11.2%	19.6%	36.7%	58.1%
		2회	225.2	13.4	33.1%	56.9%	81.6%	95.8%	99.3%	99.9%
		3회	67.8	6.8	73.9%	95.8%	99.3%	99.9%	100.0%	100.0%



4. 결론

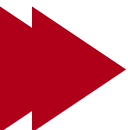
기계 학습을 이용한 셀 문턱 전압 분포 기반의 Soft Decision 센싱 레벨 예측 방법 제안

- ① 여러 종류의 열화 실험을 통해 다양한 셀 분포와 Physical Feature 데이터 수집
- ② 셀 개수 산정을 통해 설정 SER에 대한 셀 분포가 반영된 센싱 레벨 값을 확보
- ③ 독립변수(Physical Feature)와 결과변수(센싱 레벨 값)을 사용하여 ML 모델을 구축
- ④ Physical Feature으로만 ML 예측 모델을 통하여 Soft Decision 센싱 레벨 예측 가능

- ✓ ML 모델 구축 가능함을 확인함 (정확도 36~52% 향상)
- ✓ 다양한 설정 SER에 대해 ML 모델 구축의 가능함을 확인 (최소 정확도 94% 이상)
- ✓ 다구간에 우수한 성능을 실험적으로 확인 (3회 대비 정확도 9~26% 향상)



한번 시도에도 높은 정확도를 가지고 있기에 실용성이 높다



Acknowledgement

이 연구는 2019년도 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원으로 수행된 연구(NRF2019R1A2C2005949)이며, 2019년도 정부(산업통상자원부)의 재원으로 한국산업기술진흥원의 지원을 받아 수행된 연구(P0008691, 2019년 산업전문인력역량강화사업)이며, 삼성전자(Samsung Electronics)의 지원을 통하여 진행되었습니다.

감사합니다

Appendix

• SER별, 모델 별 성능 지표 값

SER	Model	Case	Measure		오차 허용 구간 별 정확도 ↑ (%)					
			MSE↓	MAE↓	±10mV	±15mV	±20mV	±25mV	±30mV	±35mV
30%	ML	RF	29.7	4.2	93.5%	98.9%	99.8%	99.9%	100.0%	100.0%
		MLP	87.2	7.3	73.0%	89.7%	96.7%	99.0%	99.8%	99.9%
		LR	174.4	10.3	57.2%	76.3%	88.1%	94.2%	97.0%	98.4%
	확률 기반	1 회	582.1	16.2	41.9%	59.6%	75.2%	85.5%	90.5%	93.0%
		2 회	234.7	12.0	50.0%	69.9%	83.6%	90.7%	94.3%	96.4%
		3 회	111.3	7.8	76.3%	87.9%	93.5%	96.3%	98.2%	99.4%
	균등 분배	1 회	1819.3	39.9	5.1%	7.7%	11.0%	15.7%	22.8%	33.3%
		2 회	301.9	14.6	36.0%	52.9%	69.7%	84.7%	94.6%	98.2%
		3 회	134.2	9.6	57.0%	77.1%	91.5%	99.4%	99.8%	100.0%
40%	ML	RF	15.8	3.1	98.3%	99.8%	100.0%	100.0%	100.0%	100.0%
		MLP	49.3	5.5	85.6%	96.4%	99.2%	99.9%	99.9%	99.9%
		LR	110.7	8.1	69.1%	86.4%	94.1%	97.2%	98.7%	99.5%
	확률 기반	1 회	381.1	12.9	52.3%	73.2%	86.4%	91.5%	94.0%	95.3%
		2 회	144.8	9.4	62.3%	81.9%	91.1%	95.1%	97.4%	99.1%
		3 회	68.1	6.1	84.3%	93.1%	96.6%	98.7%	99.7%	100.0%
	균등 분배	1 회	1462.9	36.3	4.0%	6.5%	10.1%	15.2%	25.0%	40.1%
		2 회	267.0	14.2	33.7%	53.1%	73.3%	90.6%	97.6%	99.4%
		3 회	86.4	7.7	69.1%	89.3%	99.0%	99.7%	100.0%	100.0%
50%	ML	RF	8.7	2.3	99.7%	100.0%	100.0%	100.0%	100.0%	100.0%
		MLP	28.7	4.2	93.8%	99.2%	99.9%	99.9%	100.0%	100.0%
		LR	70.7	6.4	79.9%	92.9%	97.1%	98.8%	99.7%	100.0%
	확률 기반	1 회	244.5	10.2	63.3%	84.4%	91.7%	94.6%	95.6%	95.7%
		2 회	89.6	7.3	75.4%	89.7%	95.2%	97.9%	99.5%	100.0%
		3 회	43.0	4.8	90.1%	95.9%	98.6%	99.8%	100.0%	100.0%
	균등 분배	1 회	1102.1	31.8	3.5%	6.4%	11.2%	19.6%	36.7%	58.1%
		2 회	225.2	13.4	33.1%	56.9%	81.6%	95.8%	99.3%	99.9%
		3 회	67.8	6.8	73.9%	95.8%	99.3%	99.9%	100.0%	100.0%