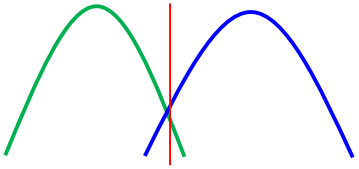
Using Machine Learning Techniques To Reduce SSD Costs

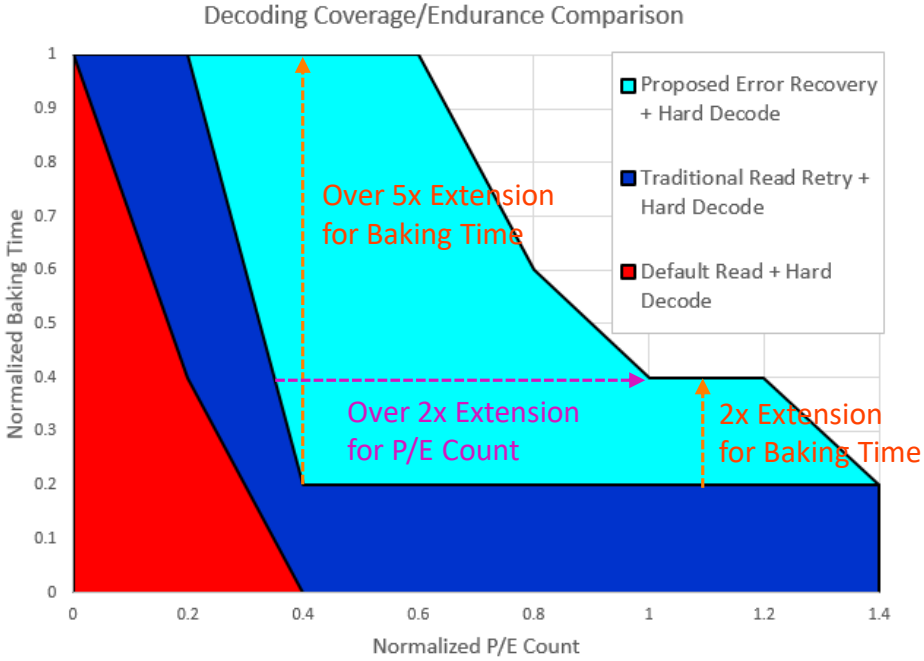
Cloud Zeng

LITEON/Storage/NVM Lab

eTLC Endurance Improvement

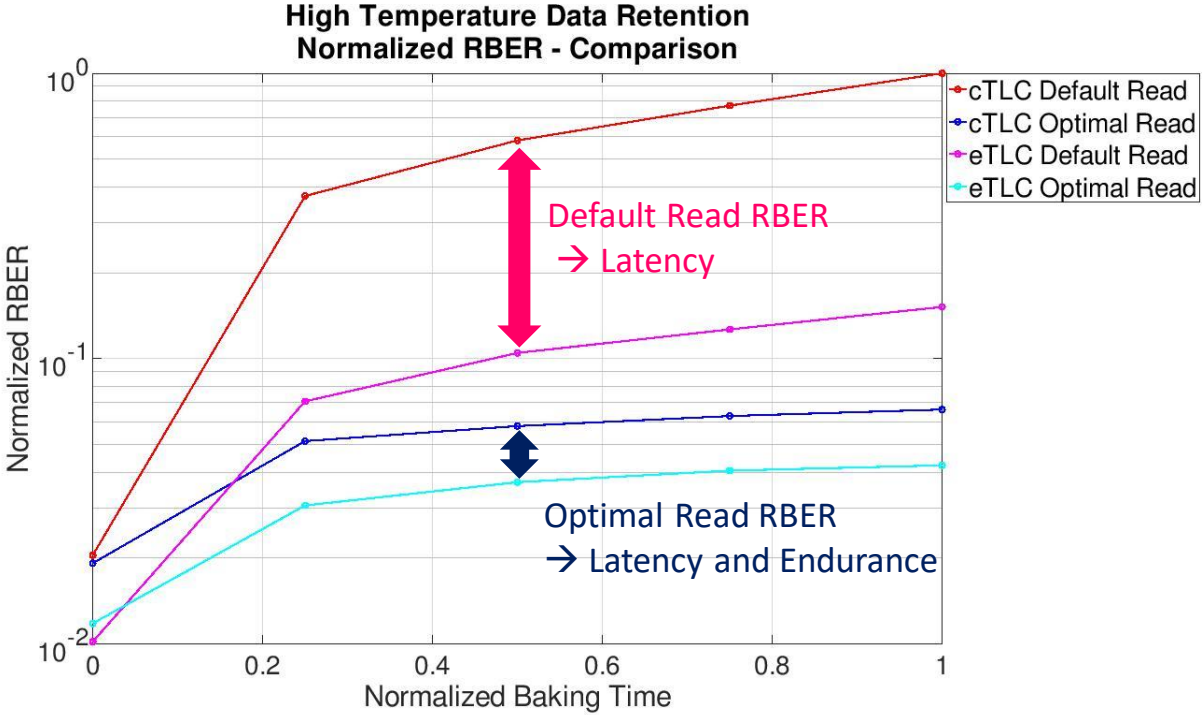


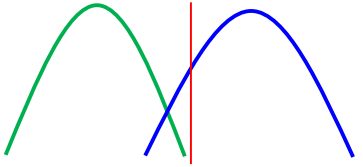
Optimal Read Level with Hard Decoding



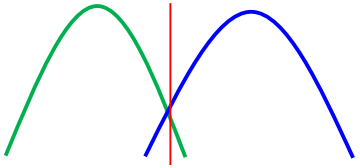
- Our Error Recovery Scheme use ML to find Optimal Parameters for variant operation conditions (**combination of {PE, DR, RD, Temperature}**)
- 5x Extension for Baking Time & 2x Extension for P/E Count

cTLC vs eTLC – RBER – HTDR





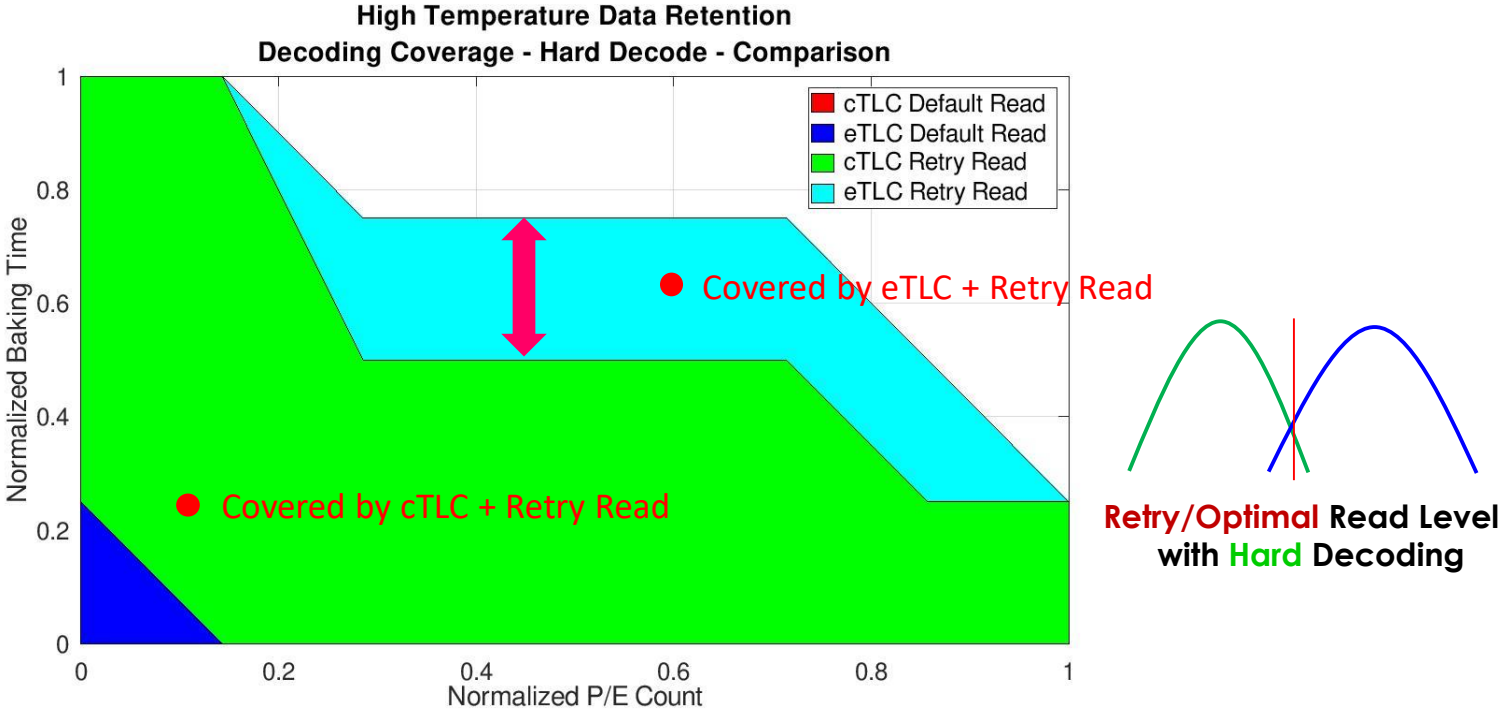
Default Read Level



Optimal Read Level

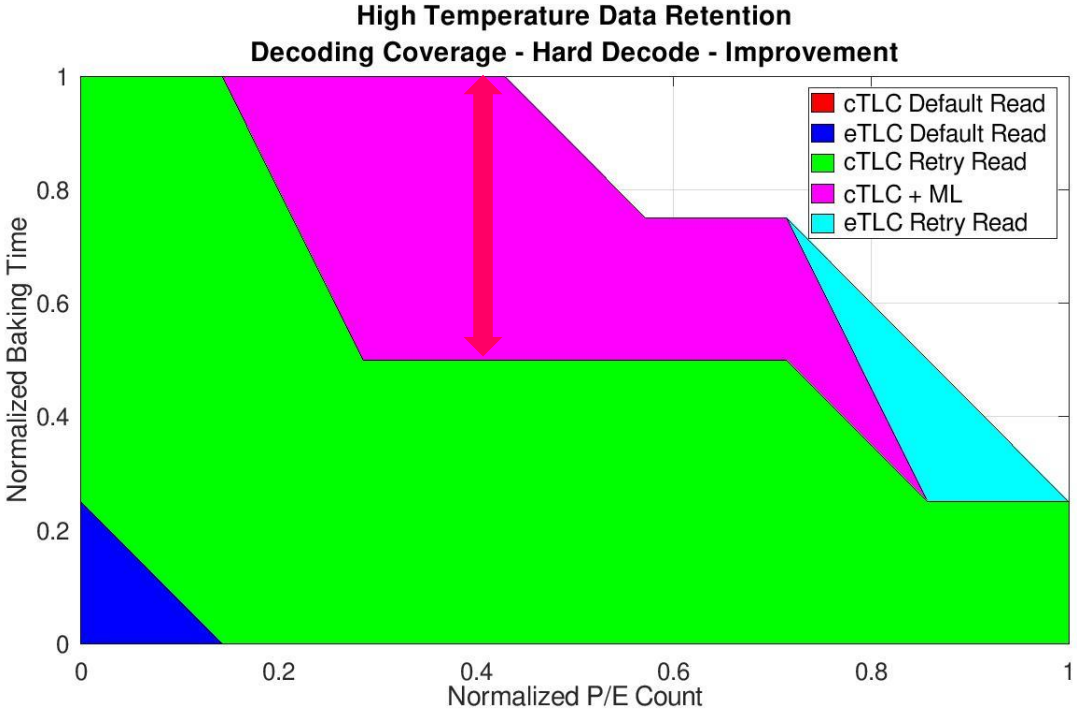
- RBER change with operation condition {P/E, DR, RD, Cross Temp….}
- There is always a RBER gap between cTLC and eTLC

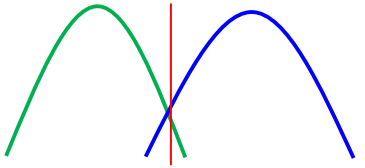
cTLC vs eTLC – Endurance – HTDR



- There is always an endurance gap between cTLC and eTLC
- eTLC with Retry Read has better decoding coverage

cTLC vs eTLC – Endurance – HTDR

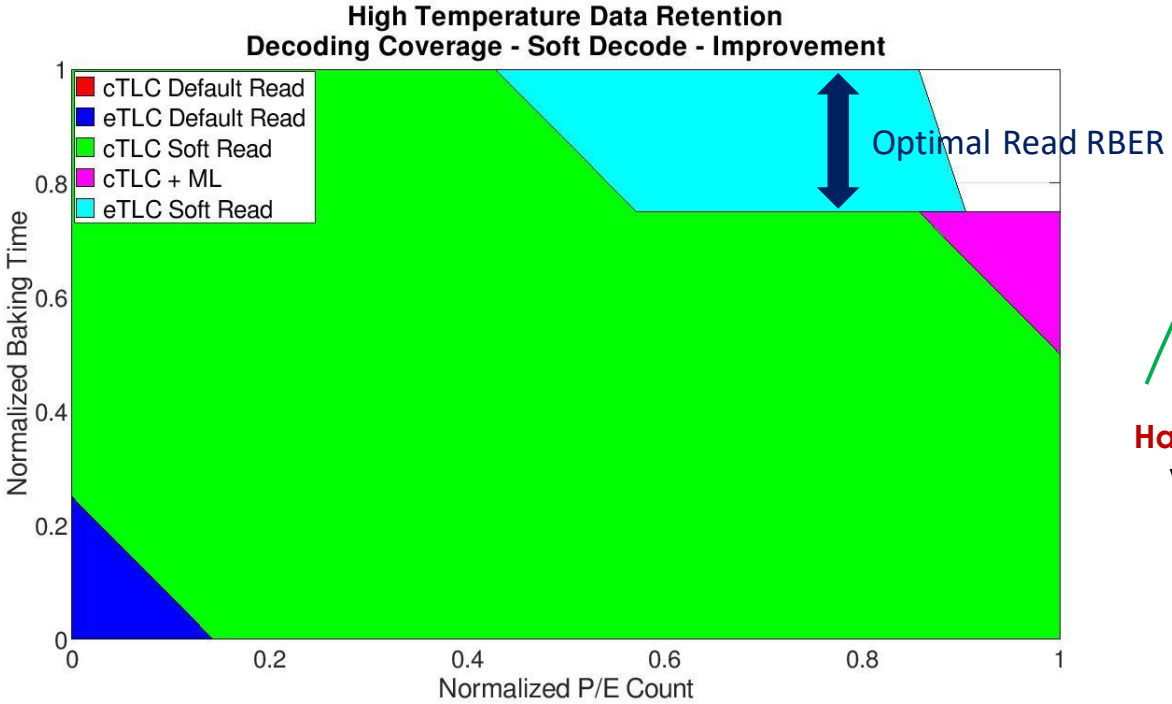


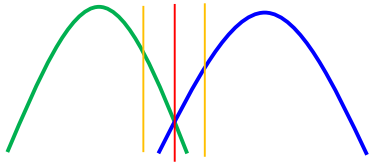


Retry/Optimal Read Level with Hard Decoding

- Our ML techniques can extend decoding coverage for cTLC (Optimized Retry Sequence and Read Level Prediction Model)

cTLC vs eTLC – Endurance – HTDR

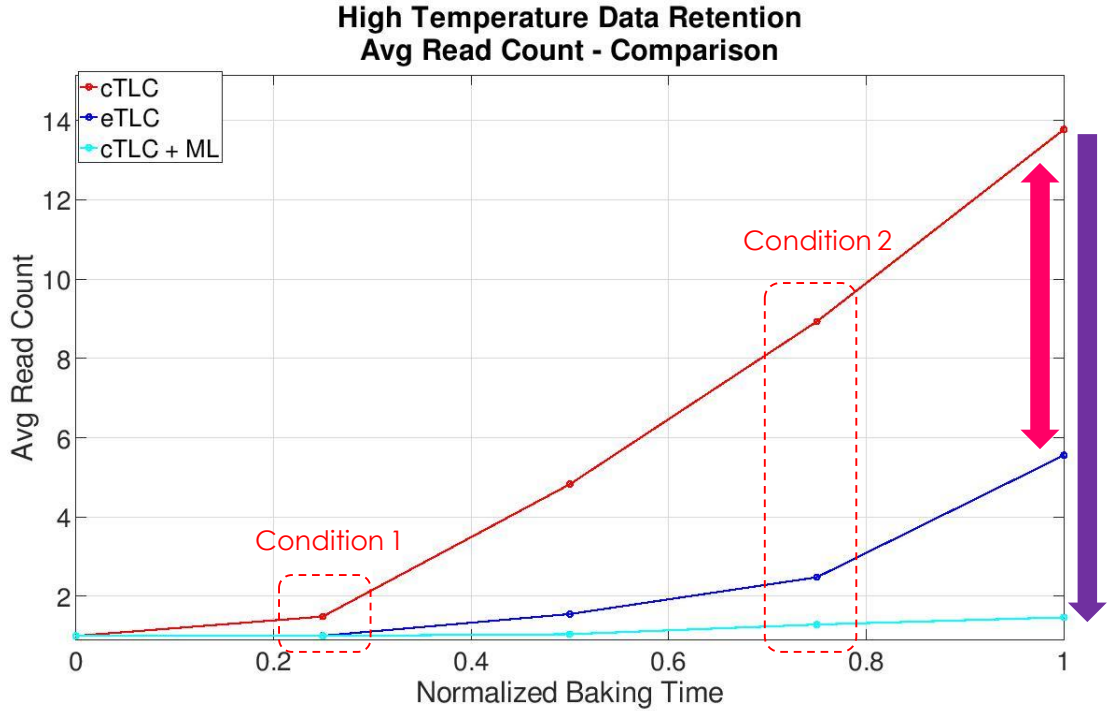


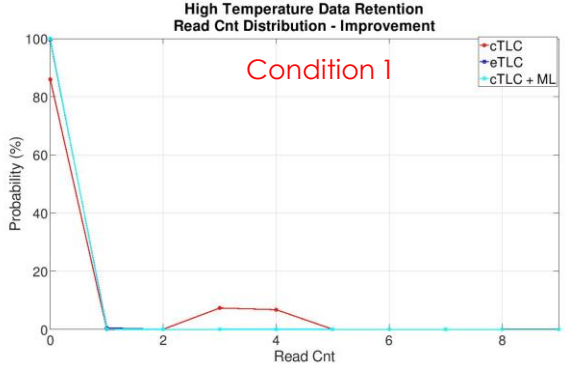


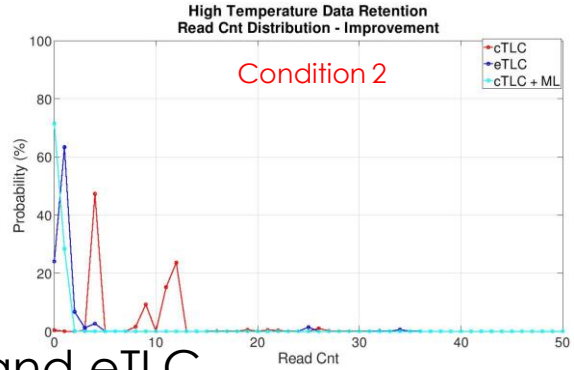
Hard/Soft Read Level with Soft Decoding

- There is an intrinsic gap for soft decoding coverage
- The soft decoding coverage is far beyond the spec

cTLC vs eTLC – Latency – HTDR

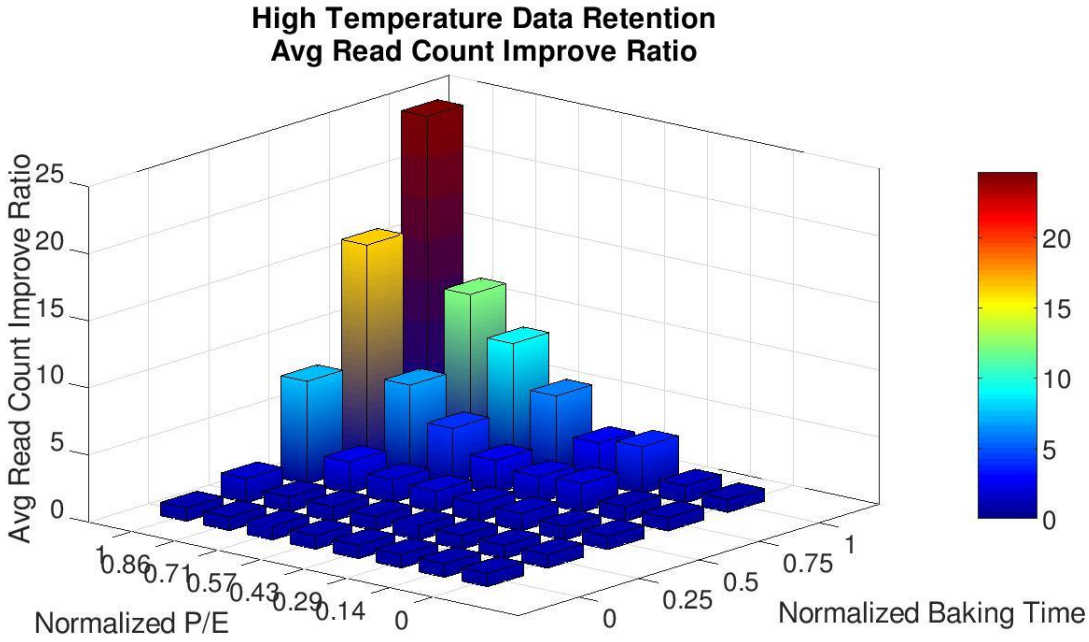






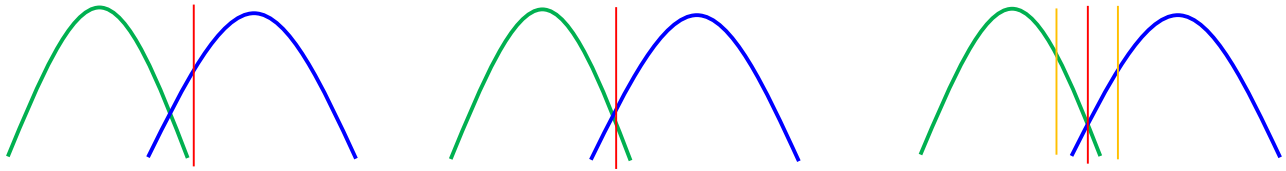
- There is always a latency gap between cTLC and eTLC
- cTLC + ML can greatly reduce the read latency

cTLC vs eTLC – Latency – HTDR

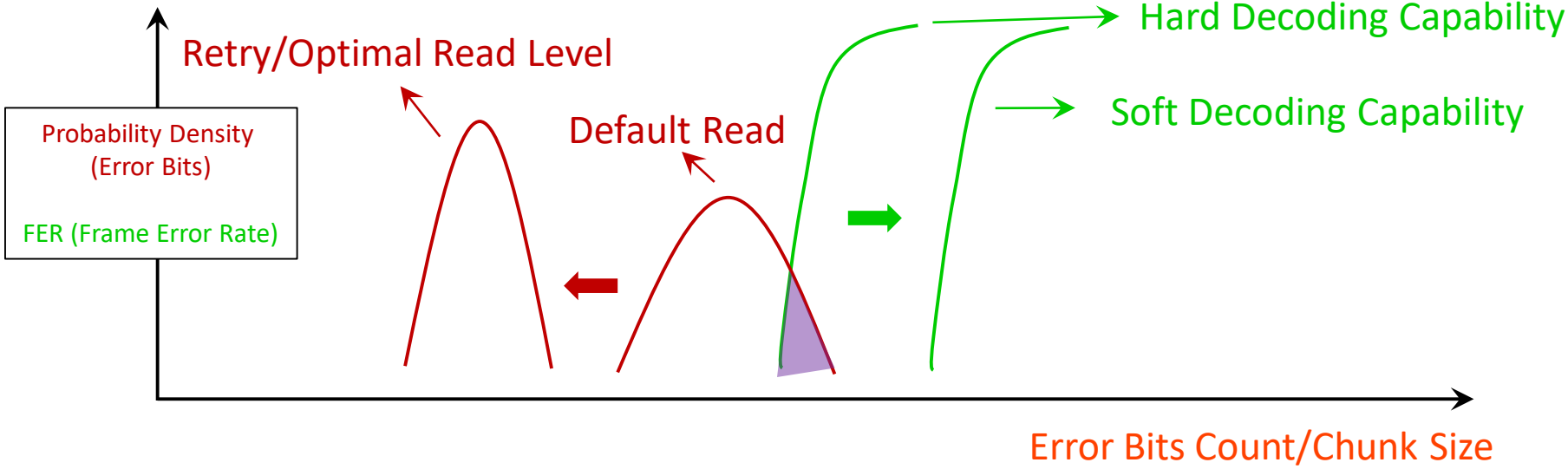


- Improvement Ratio ： Avg Read eTLC / Avg Read (cTLC + ML)
- cTLC + ML always has less read count compared with eTLC

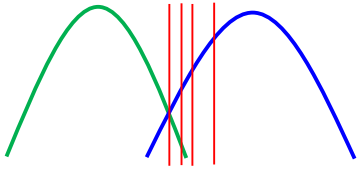
Error Recovery Flow



1. Default Read Level with Hard Decoding
2. Retry/Optimal Read Level with Hard Decoding
3. Retry/Optimal Read Level with Soft Decoding

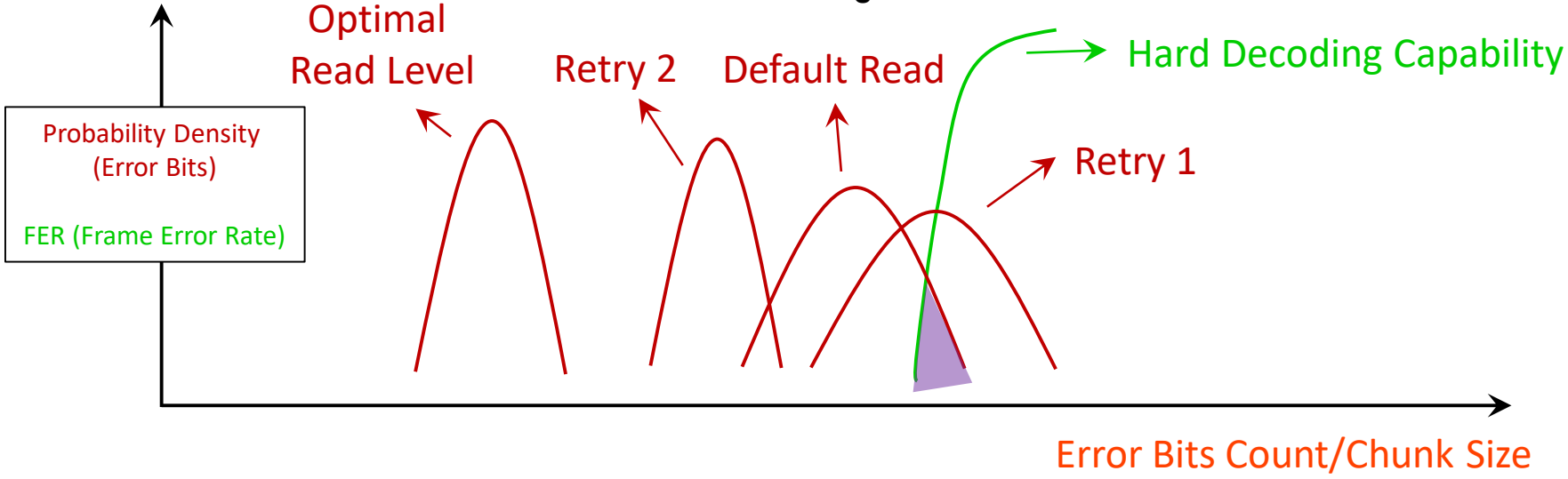


Read Level Management



Read Level
with Hard Decoding

1. Optimized Retry Sequence
2. Adjust Default Read Level
3. Accurate Optimal Read Level

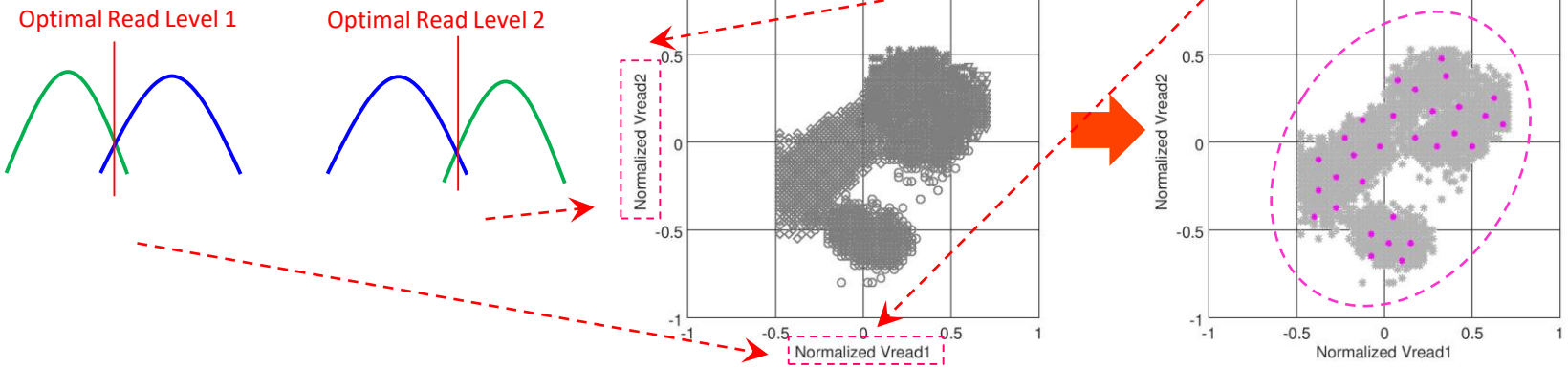


Read Retry Table – Clustering

Billions of ECC Chunks Info were collected over dice under different failure mode

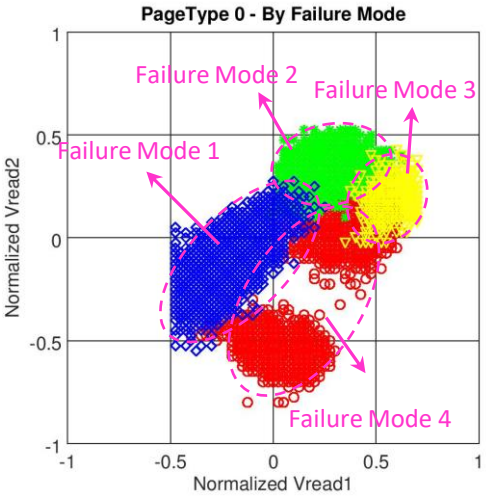
Die	Plane	BLK	WL	PageType	P/E Count	Baking Time	Optimal Read LV1	Optimal Read LV2
2	0	100	64	0	3000	24	+10	-6
2	1	101	78	0	4000	24	+7	+9
3	1	120	31	0	4000	36	+3	-12
…								

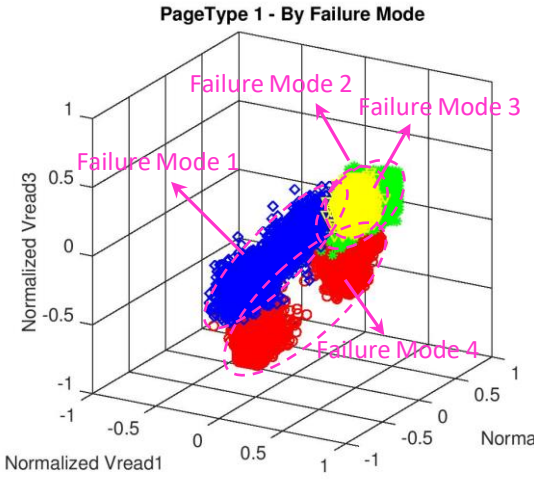
ECC Chunks Info

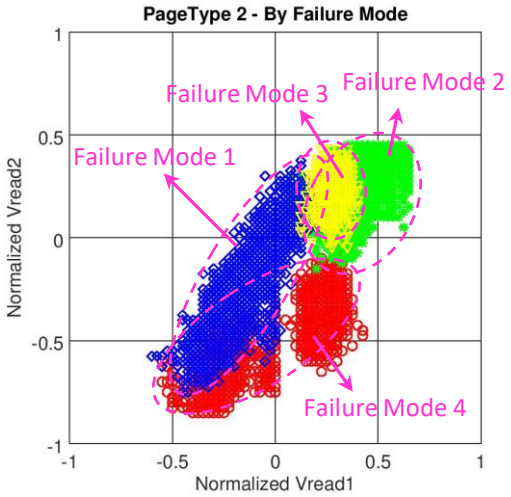


Optimized Retry Sequence

Find some indexes to separate the data, reduce number of retry tables





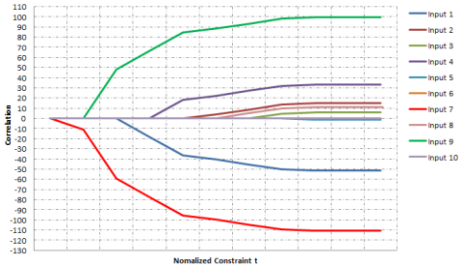


Prediction Model – Optimal Read Level

Example: Data Collection

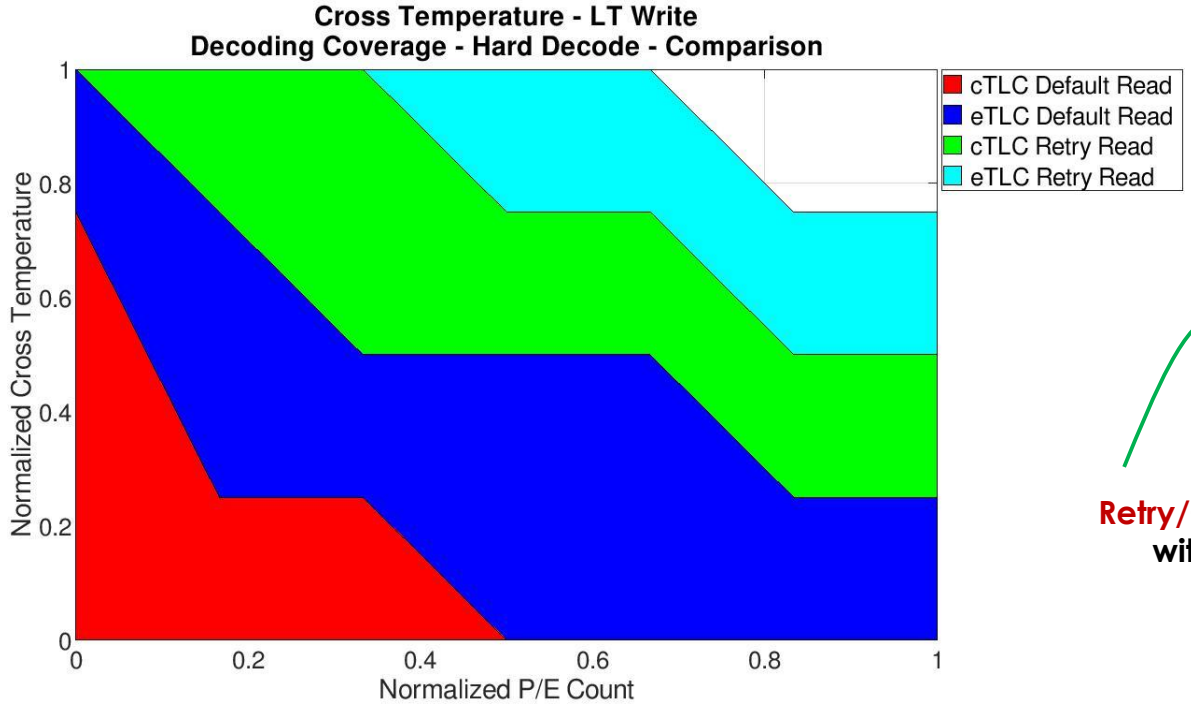
	Input Para 1	Input Para 2	Input Para 3	Input Para 4	Input Para 5	Input Para 6	Optimal HD Read Level
Data 1	1100	589	1794	6322	1000	1000	6
Data 2	932	908	1503	7849	500	500	-5
…	…						
Data N	990	842	1894	5692	300	400	3

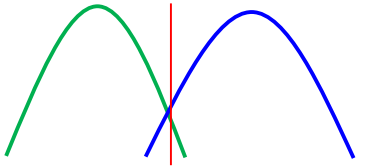
Feature Selection



- **What's the Optimal HD Read Level after n Days/Weeks?**
- **Input Parameters:**
  - P/E Cycle, Retention Time, Read Count, Temperature, Dwell … Program/Erase Time, Histogram ….
- **Regression Problem:**
  - Ordinary Least Square(OLS) Regression
  - Ridge Regression (Hoerl and Kennard, 1970)
  - Other Regression Analysis can be used to solve this problem

cTLC vs eTLC – Endurance – Cross Temp

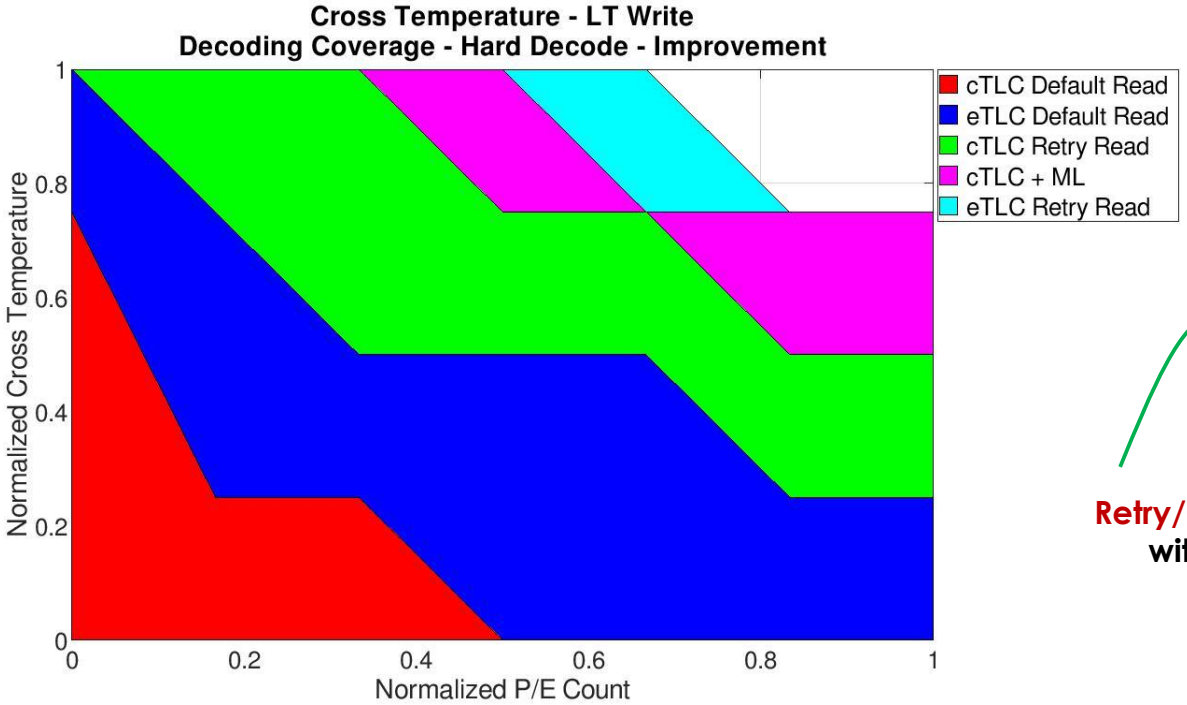


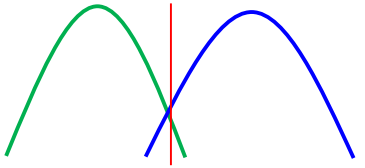


Retry/Optimal Read Level with Hard Decoding

- Typical Result : eTLC > cTLC , Retry Read > Default Read

cTLC vs eTLC – Endurance – Cross Temp



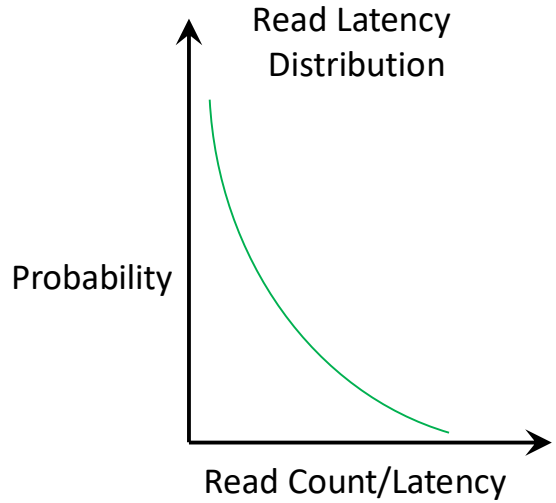


Retry/Optimal Read Level with Hard Decoding

- Our ML techniques can extend decoding coverage for cTLC (Optimized Retry Sequence and Read Level Prediction Model)

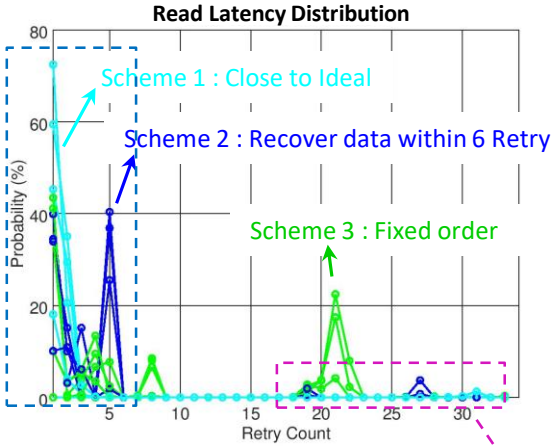
Latency Improvement

Change Read Level and the Priority of Retry Table dynamically

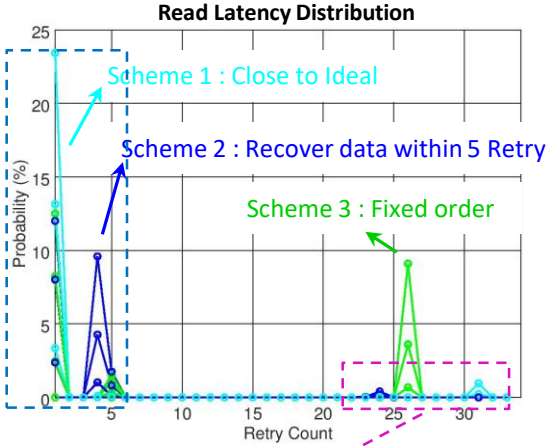


Ideal : monotonically decreasing

Failure Mode 1



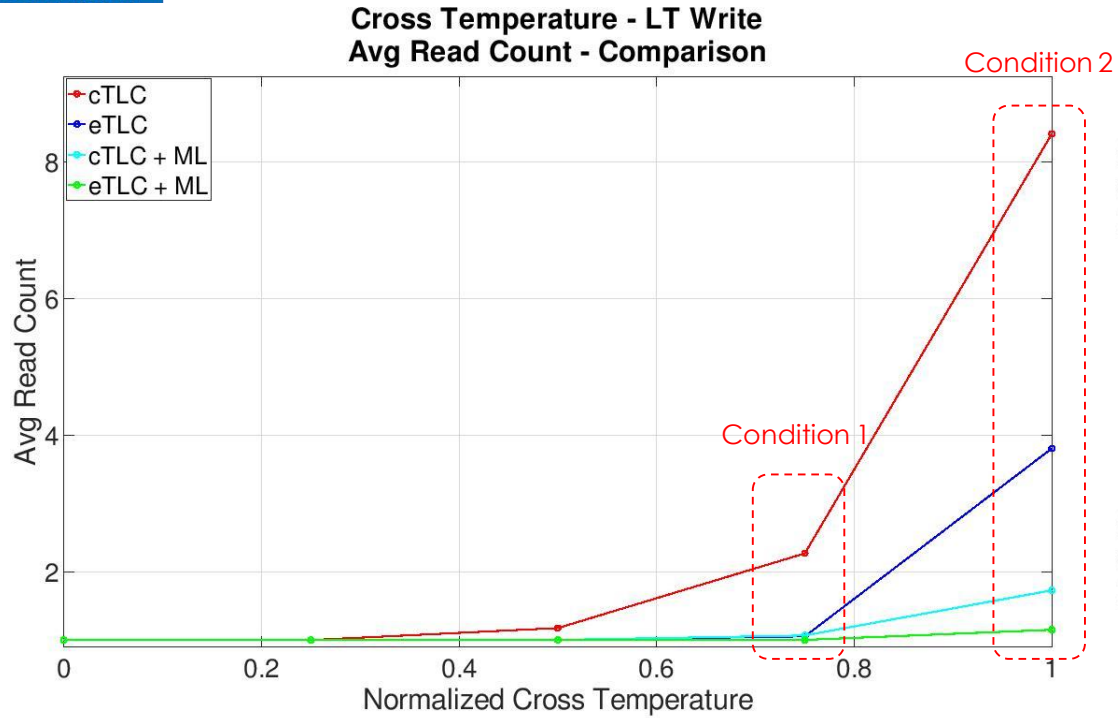
Failure Mode 2

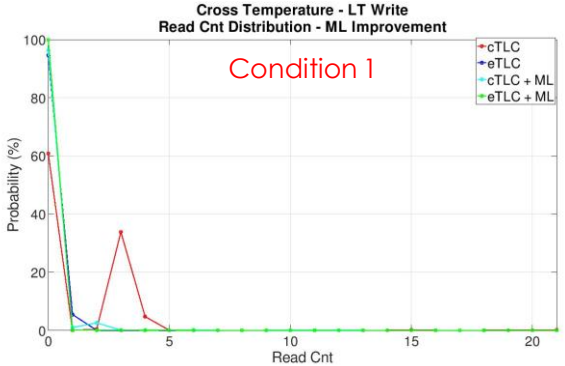


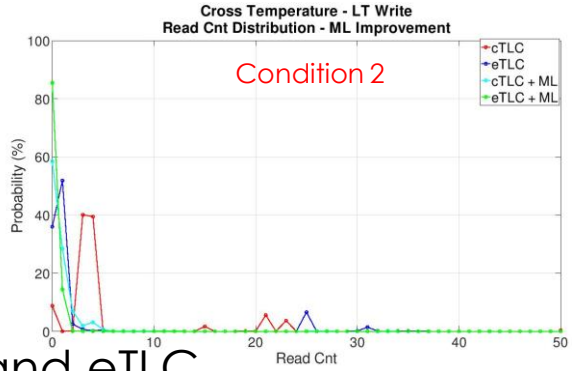
1. Arranged Optimized Retry Sequence + Read Level Adjustment
2. Optimized Retry Sequence (Fixed Order)
3. Traditional Retry Table (Fixed Order)

Last line of defense → Prediction Model : Optimal Read Level/LLR

cTLC vs eTLC – Latency – Cross Temp

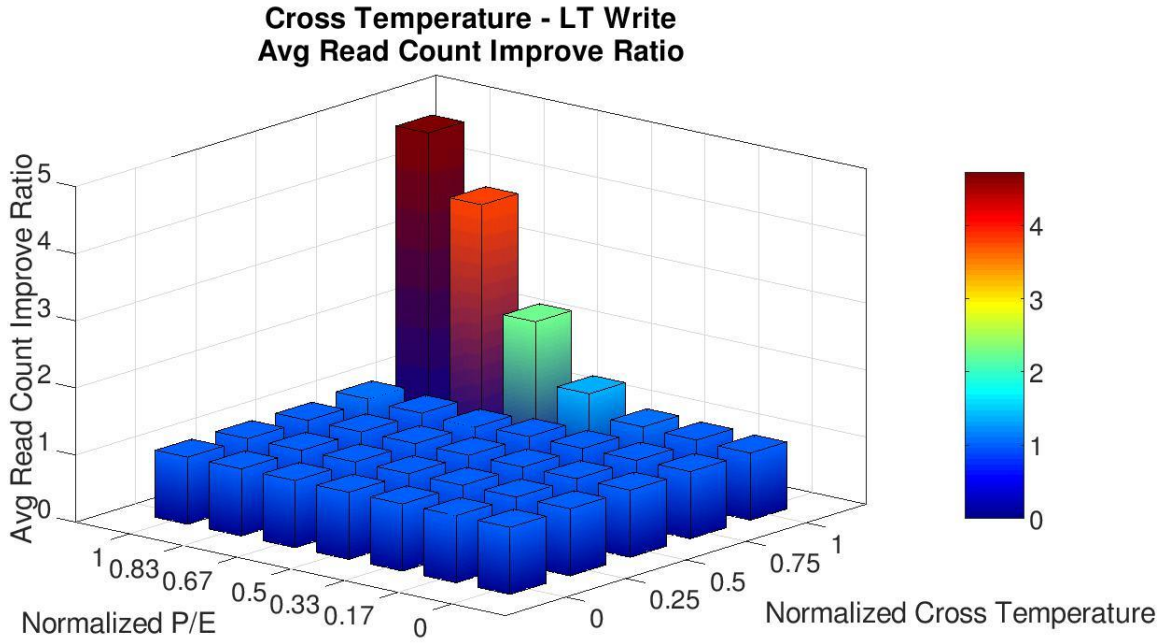






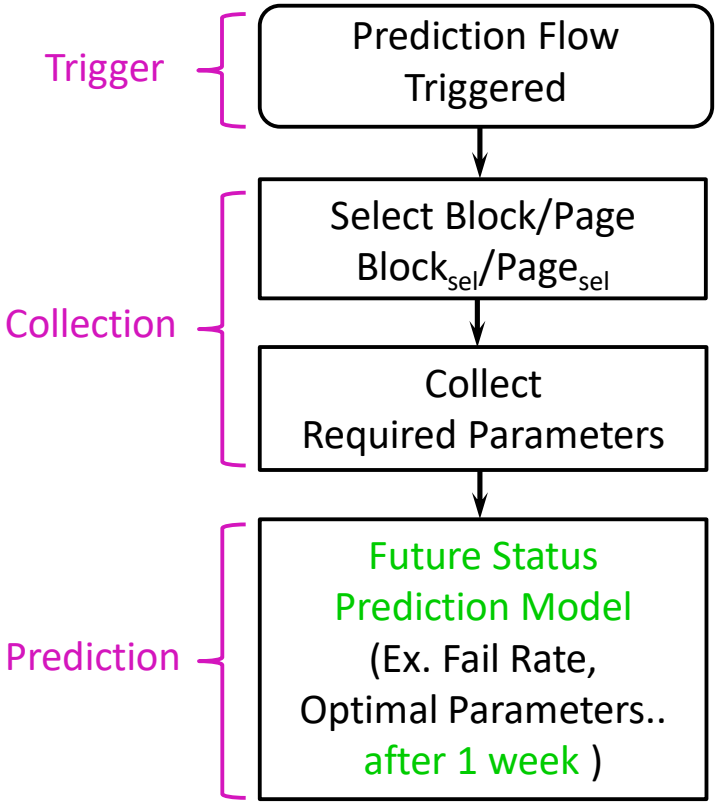
- There is always a latency gap between cTLC and eTLC
- cTLC + ML can greatly reduce the read latency

cTLC vs eTLC – Latency – Cross Temp



- Improvement Ratio : Avg Read eTLC / Avg Read (cTLC + ML)
- cTLC + ML always has less read count compared with eTLC

Implement For Future Status Prediction



P1 : Trigger Condition/Frequency ?

P2 : Block/Page(s) to be selected ?

P3 : Important/Required Parameters ?

P4 : Operation Condition in 1 week !!!

P5 : Power-Off !!!

Summary

- **How to Extend Endurance & Reduce Error Recovery Latency**
  - Optimized Read Retry Sequence
  - Optimal Read Level, LLR, Status Prediction Model
  - Design Error Recovery Scheme based on NAND Flash Characteristic/Controller Architecture
- **Replace Enterprise Level TLC with Client Level TLC**
  - cTLC + ML can achieve eTLC specification (JESD218)
  - cTLC + ML can greatly reduce the latency compared with eTLC
  - Large sample testing follow JESD218 is on-going

Flash Memory SUMMIT

LITEON®

THANK YOU!

Any questions?

Come by LITE-ON **Booth# 621** for Demos!

- Learn about Machine Learning & the latest SSD Technology
- Get a chance to win special prizes!