NVMe SSD data reliability and protection

Yuekuan Li
yuekuan.li@memblaze.com

Reliability NVMe SSD

How to evaluate NVMe SSD reliability

1. JESD218A Enterprise standard

UBER < 1 sector per 10^{16} bits read

2. MTBF-- Manufacturer liked

MTBF: predicted elapsed time between inherent failures of a system during operation.

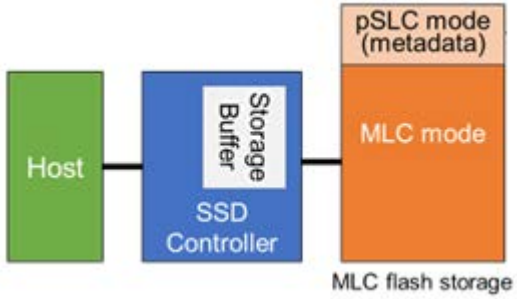
MTBF refers to the failure rate of a drive over its expected lifetime.
This doesn't mean a 1 million hour MTBF drive will last 1 million hours

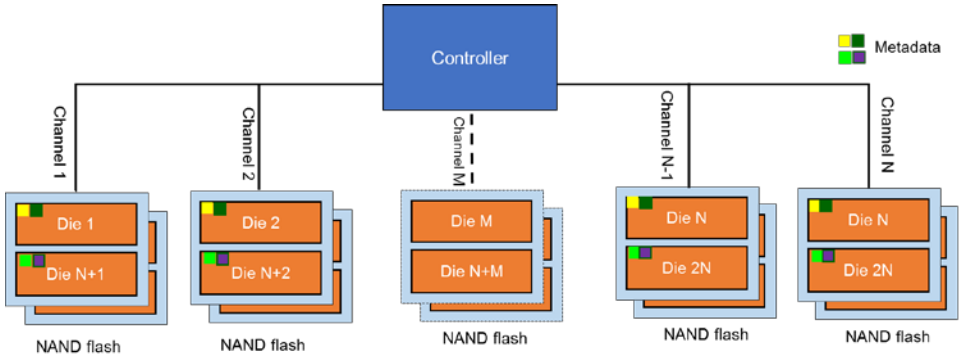
Reliability NVMe SSD

- Metadata Protection (pSLC)
- Power Failure Protection
- High Temperature Protection

Metadata Protection

- Metadata includes firmware, information on wear leveling, error correction, FTL, read/erase counts, free/bad block bitmap, and so on
- pSLC for metadata storage, has SLC's speed and durability
- 4 copies of redundancy
- Distributed on multi-channel

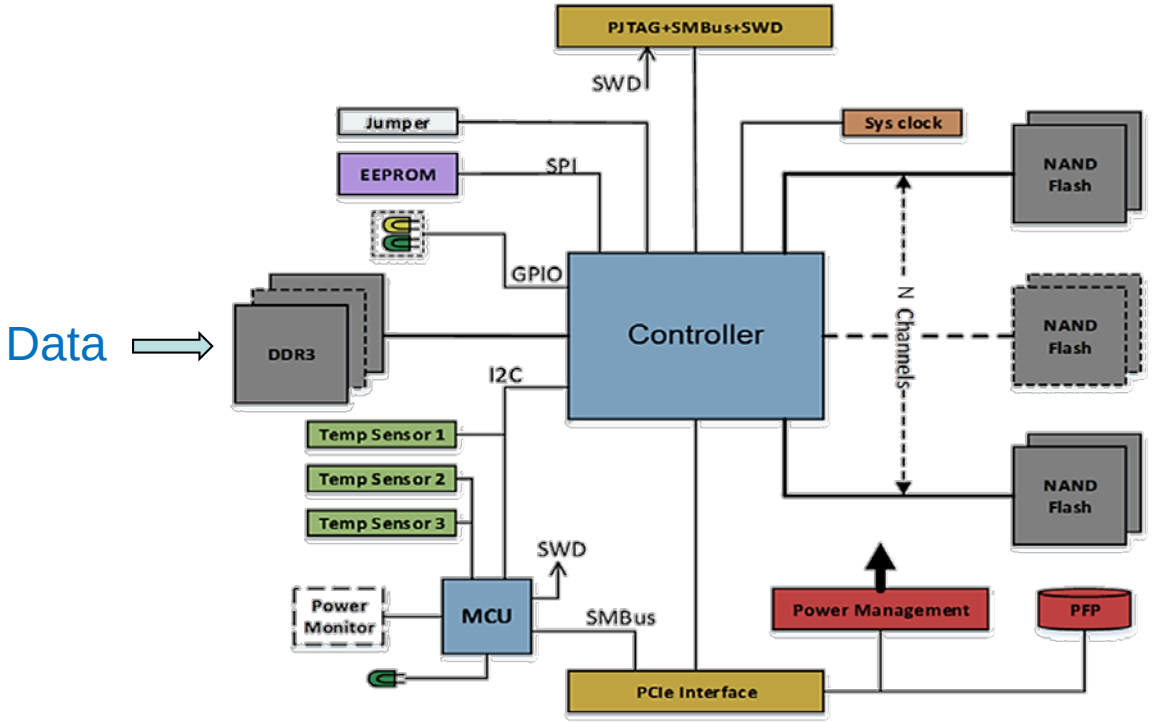




NVMe SSD write

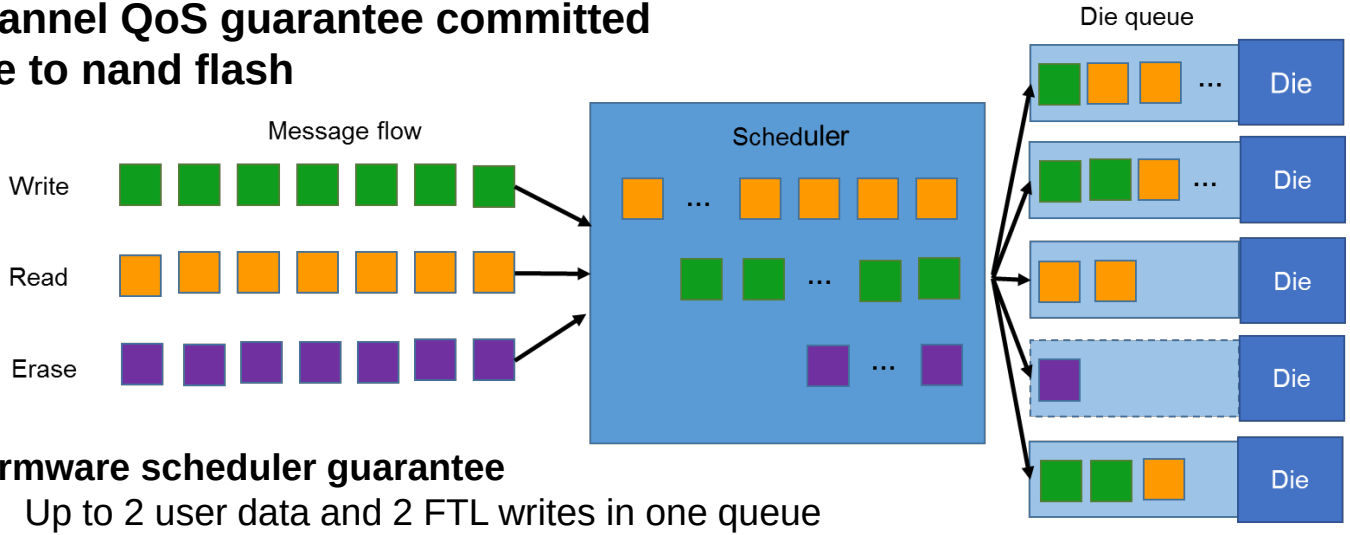
Write back
Power failure
Loss data?

No!



Power Failure Protection

Flash Channel QoS guarantee committed data write to nand flash



Firmware scheduler guarantee

- Up to 2 user data and 2 FTL writes in one queue
- No user data Write and Erase in same queue at any time
- Up to 4 commands in one queue

Capacitors guarantee

- additional 15ms since power failure detected

Power Failure Capacitor solution

$$W=P*t=\frac{1}{2}*C*V^2=\frac{1}{2}*C*(V_1^2-V_2^2)$$

P: write operation
T: time
C: capacity
V: voltage

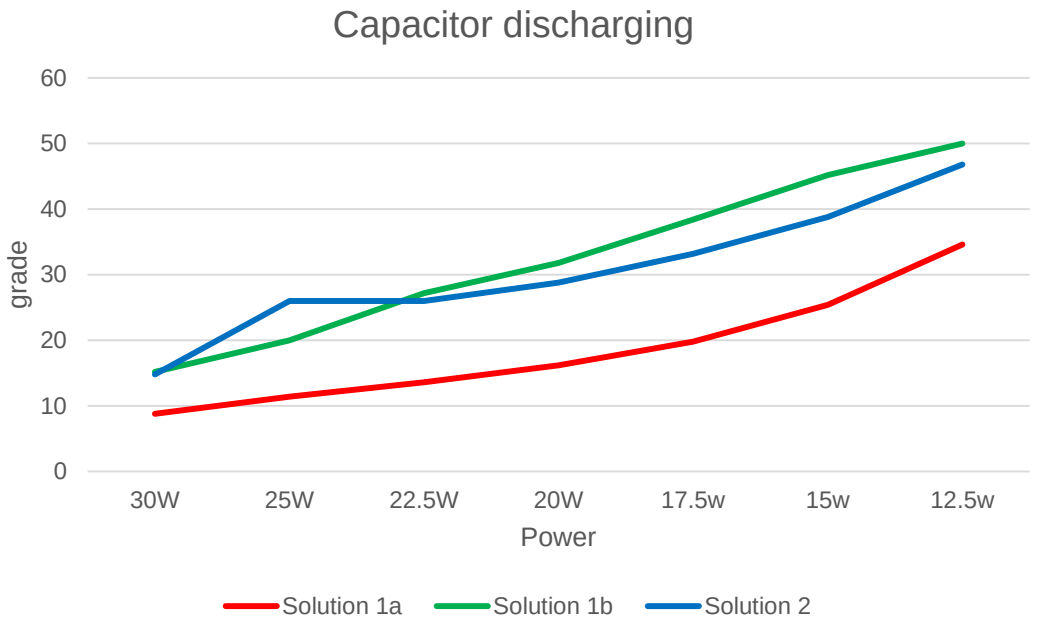
Solution 1: bigger C

Advantage:
Simple, don't need additional charging, add to bus is ok
Disadvantage:
Cost high. Require higher capacity, require super capacitor or several capacitors

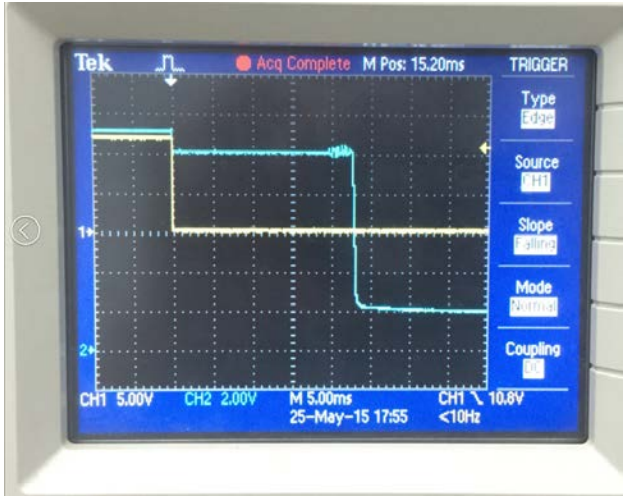
Solution 2: Higher V_1 and Lower V_2

Advantage:
Low cost, commonly used solid-state capacitors or aluminum capacitor for energy storage
Disadvantage:
Circuit complex, require boost converter and buck converter, take up more space for PCB layout

Reliability capacitor protection

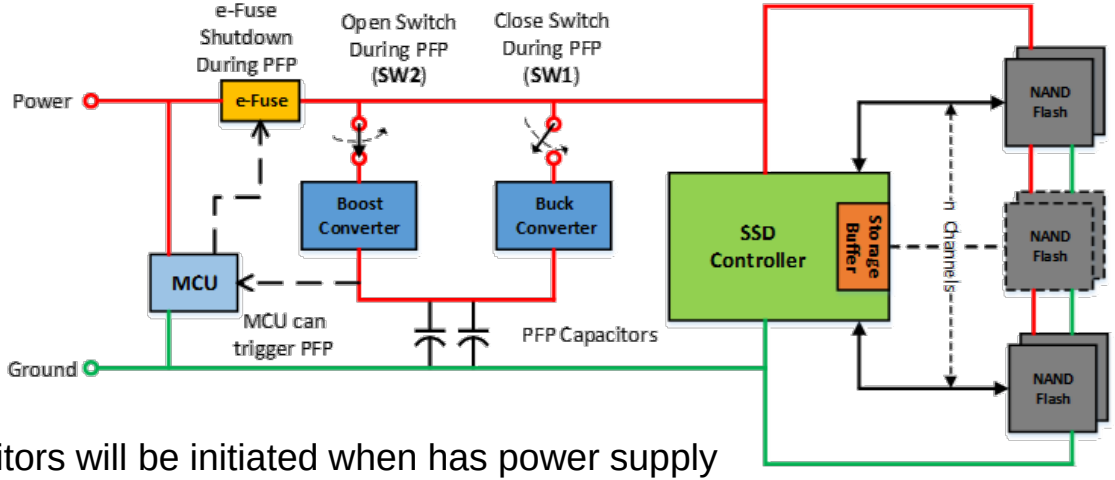


AIC Solution
Power 25W
Discharging 23ms



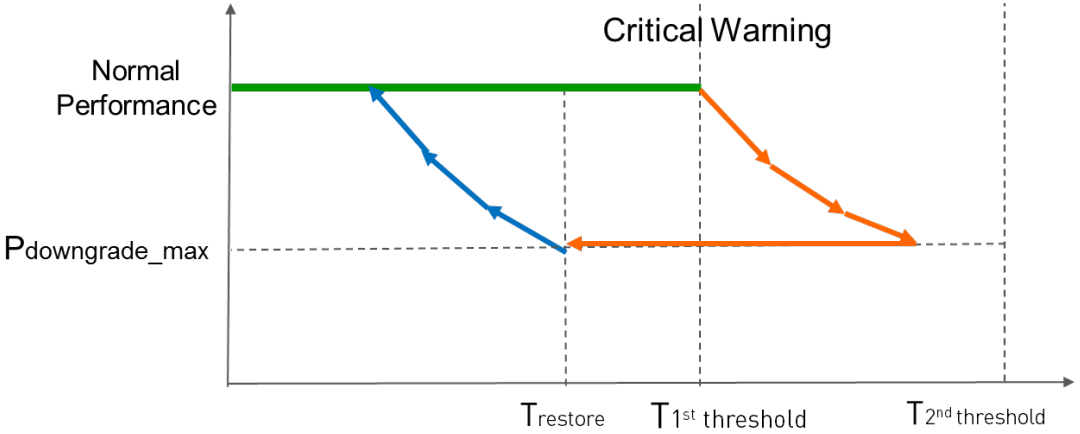
Power Failure Protection

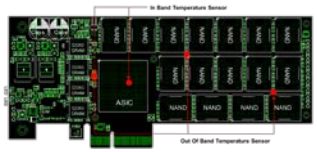
AIC solution



- Charging capacitors will be initiated when has power supply
- Capacitors guarantee additional 15ms since power failure detected
- 12V supply voltage will be boosted via Boost Converter for capacitor charging. Buck converter is to adjust voltage for 3.3Vaux operating and other components working voltage
- e-Fuse shutdown if PFP detected, at the same time, SW2 open and SW1 close for capacitor discharging
- BIST capacitor during manufacture
- MCU keep tracking capacitors health during whole product life

High Temperature Protection





- Monitor core, board and nand flash temperature
- If temperature higher than critical warning, go to downgrade performance
- If temperature lower than $T_{restore}$, go to recovery performance
- If temperature higher than T_2^{nd} threshold, will be freeze

Flash Memory SUMMIT

Continue improve NVMe SSD reliability!

Thanks!