



Flash Memory Summit

SNIA STORAGE
SECURITY SUMMIT
Wednesday, May 11, 2022 • Virtual

Insights from building storage arrays out of raw NAND

Hari Kannan

Pure Storage



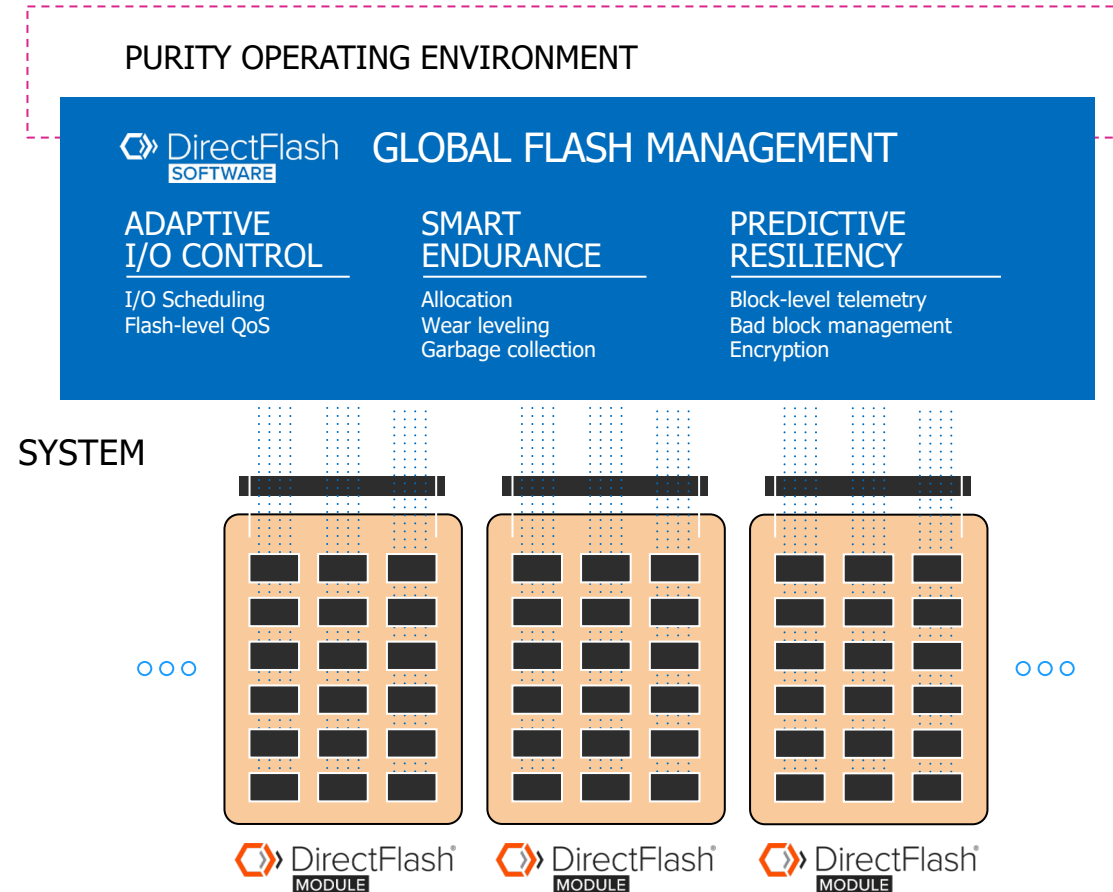
The Challenge with QLC

HIGHER DENSITY ALSO MEANS LOWER PERFORMANCE AND ENDURANCE

Media Performance (lower write latency = higher performance)			Media endurance (# P/E cycles)	
~0.5 ms	Best latency	SLC	>100k	Best endurance
1-2 ms		MLC	10k	
2-3 ms		TLC	3k	
10-20 ms	Worst latency	QLC	<1k	Worst endurance



“DirectFlash”: talk to NAND directly



1

RELIABILITY

GLOBALLY OPTIMIZED,
WITH FEWER POINTS OF FAILURE

2

EFFICIENCY

MORE USABLE FLASH WITHOUT
DOUBLE OVER-PROVISIONING

3

DENSITY

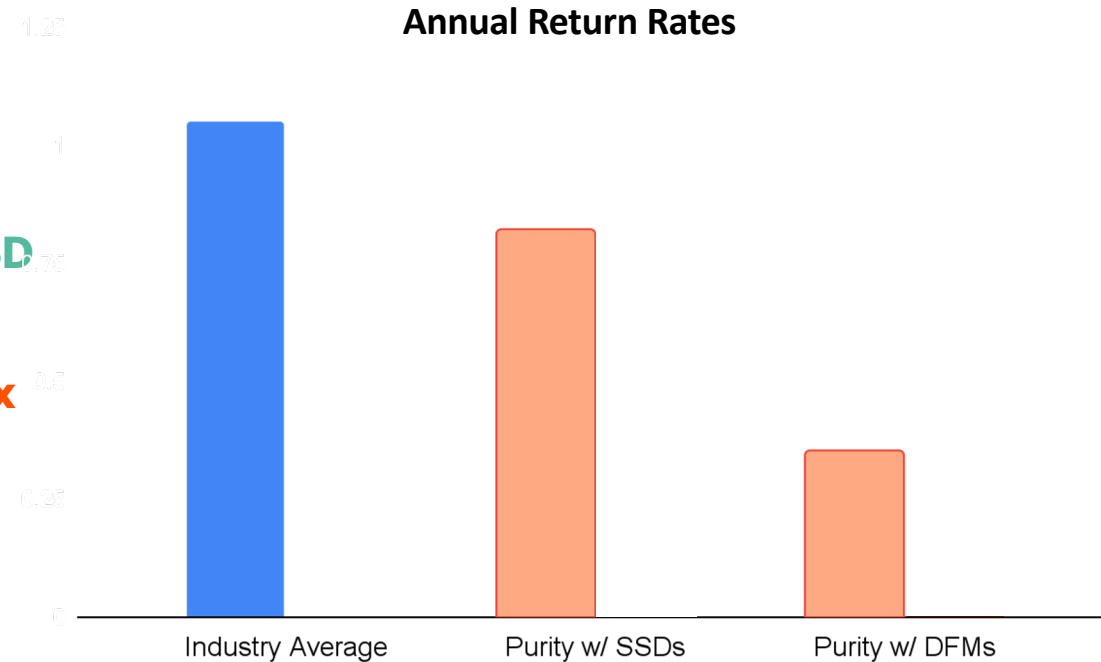
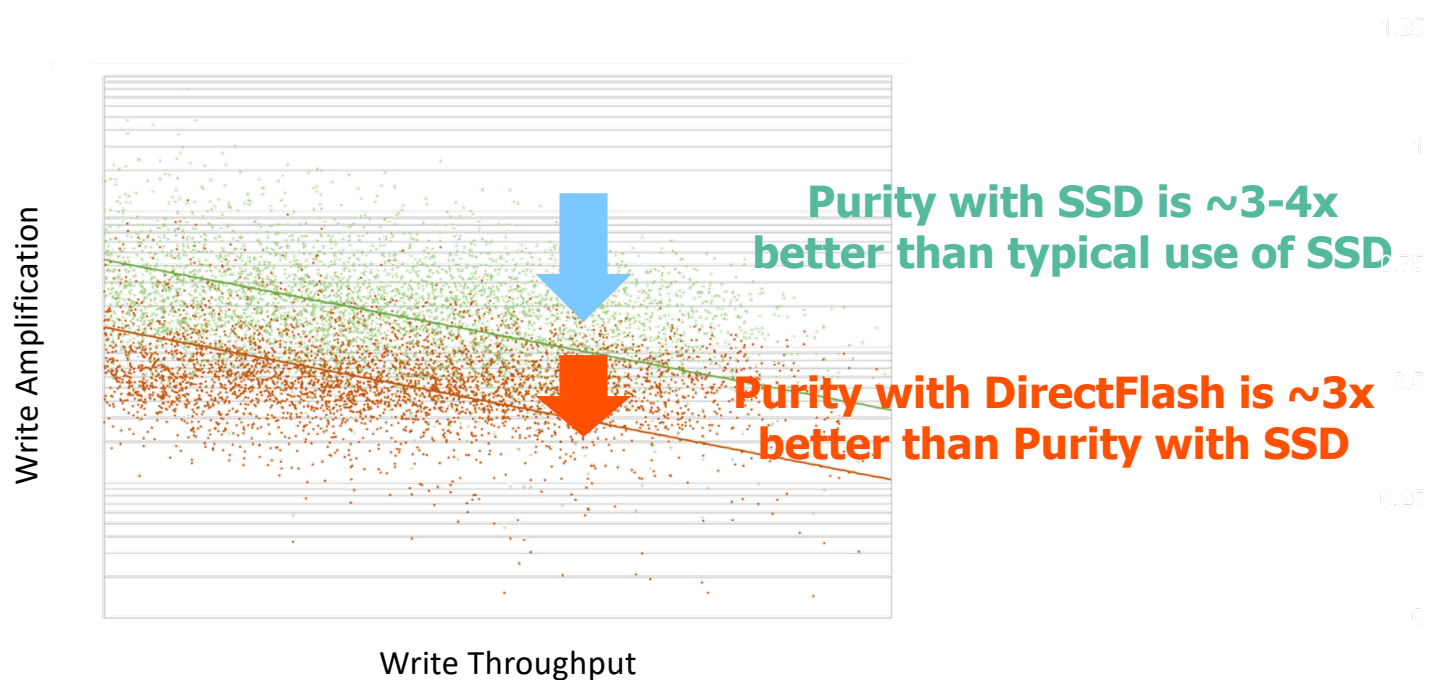
INDUSTRY-LEADING CAPACITIES,
WITHOUT TRADEOFFS

4

PERFORMANCE

HIGHER BANDWIDTH AND
PREDICTABLE LATENCY

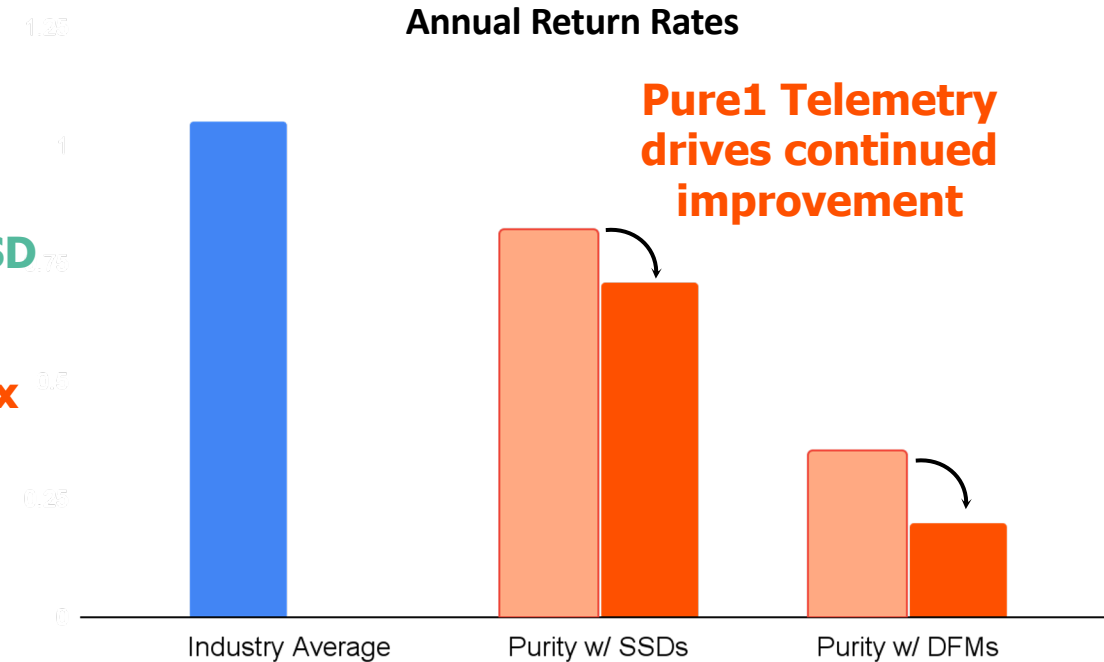
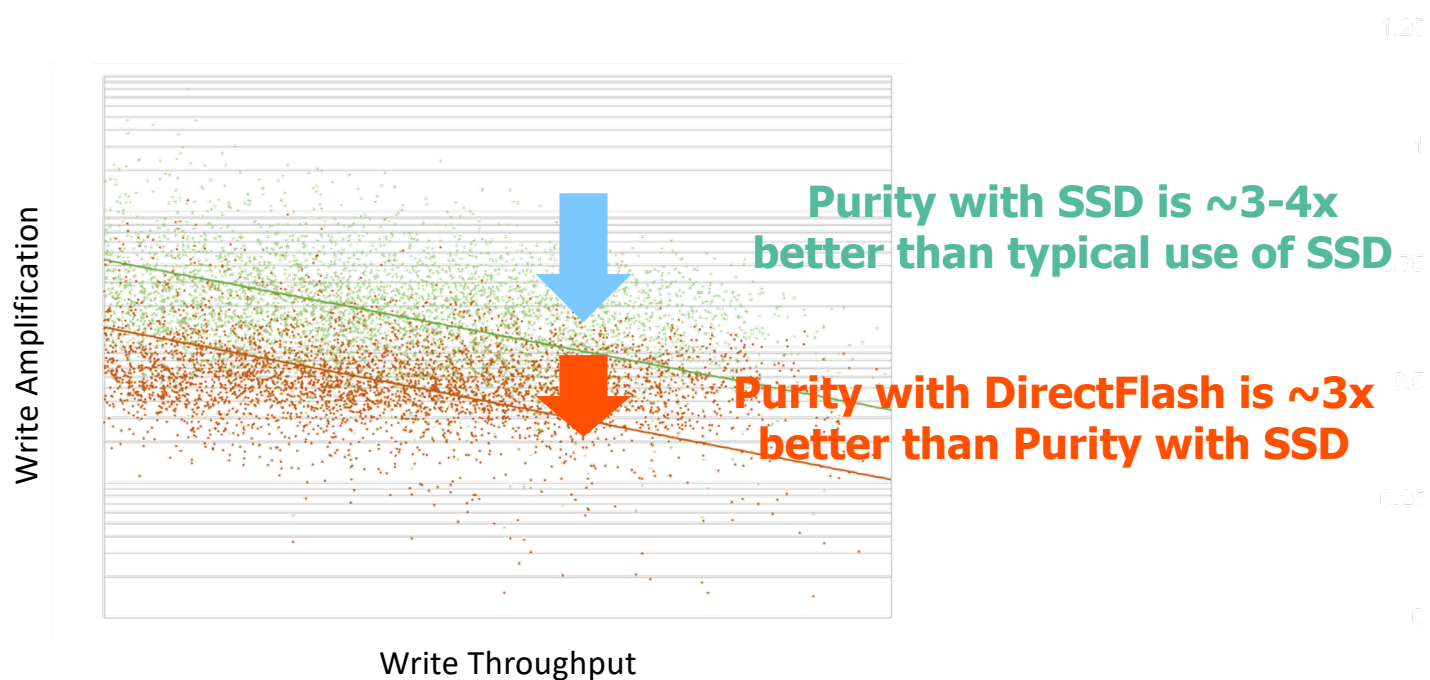
~3-4x over conventional SSDs



Moving GC & wear leveling to system-level with broader context
Reduces write-amplification and increases reliability

1 DirectFlash Improves Reliability

~3-4x over conventional SSDs



Moving GC & wear leveling to system-level with broader context
Reduces write-amplification and increases reliability

2 DirectFlash Drives Efficiency

DirectFlash eliminates unnecessary drive-level mappings, enabling

- DRAM to be sized proportional to performance
- Removing flash over-provisioning from the DFM



40x less DRAM



20% more capacity

This improvement in efficiency allows for efficiently building large flash modules

3 DirectFlash Drives Density

... **Efficiency** enables **Density**

DirectFlash eliminates unnecessary drive-level mappings, enabling

- DRAM to be sized proportional to performance
- Removing flash over-provisioning from the DFM

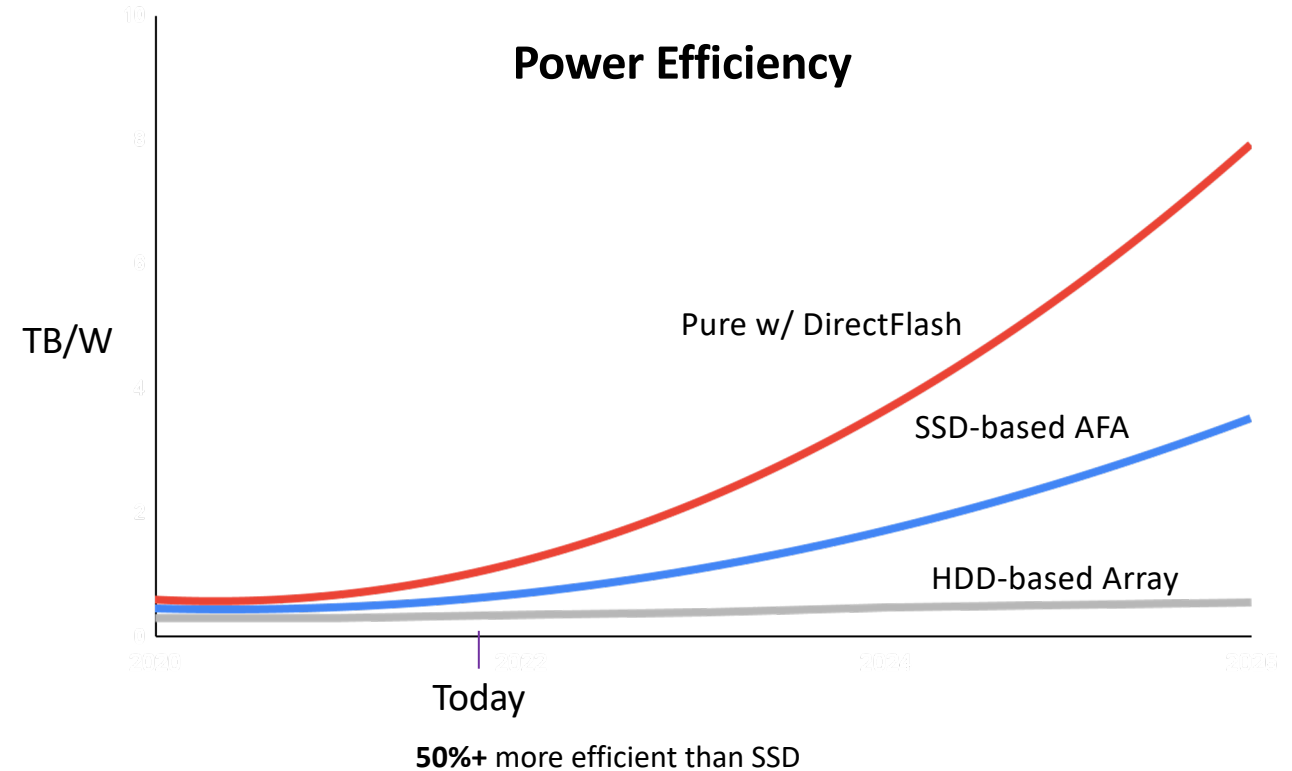


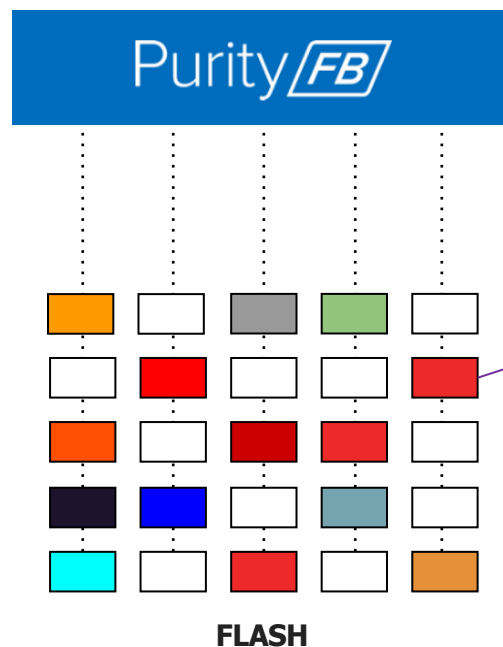
40x less DRAM



20% more capacity

This improvement in efficiency allows for building large flash modules





Slow commands like erases can block higher priority operations causing **high tail latencies**

DirectFlash provides **granular scheduling** and **reduces background operations by 3x** to provide consistent, high performance.

Scheduling GC at the system-level with broader context
Reduces write-amplification and improves performance

Dealing with Failures at Scale

NAND fails in many ways beyond bitflips

- Bad planes/die
- Failures in SRAM (cache-registers)
- Failures to initialize die state machines
- Read disturb issues
- Byzantine failures ...

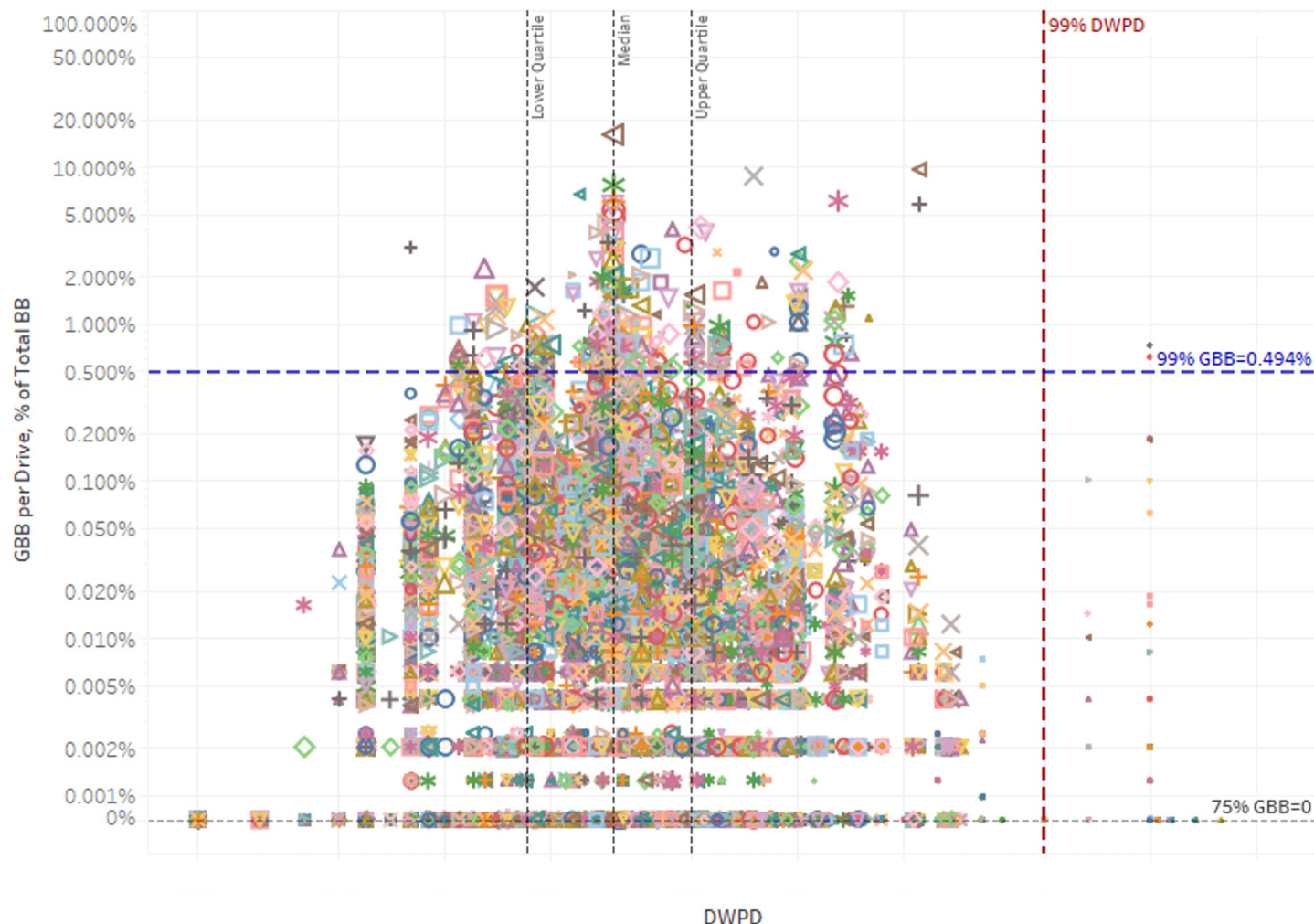
Handling these require **system-wide solutions** and **sophisticated telemetry**



Real-time, fleet-wide statistics & analytics

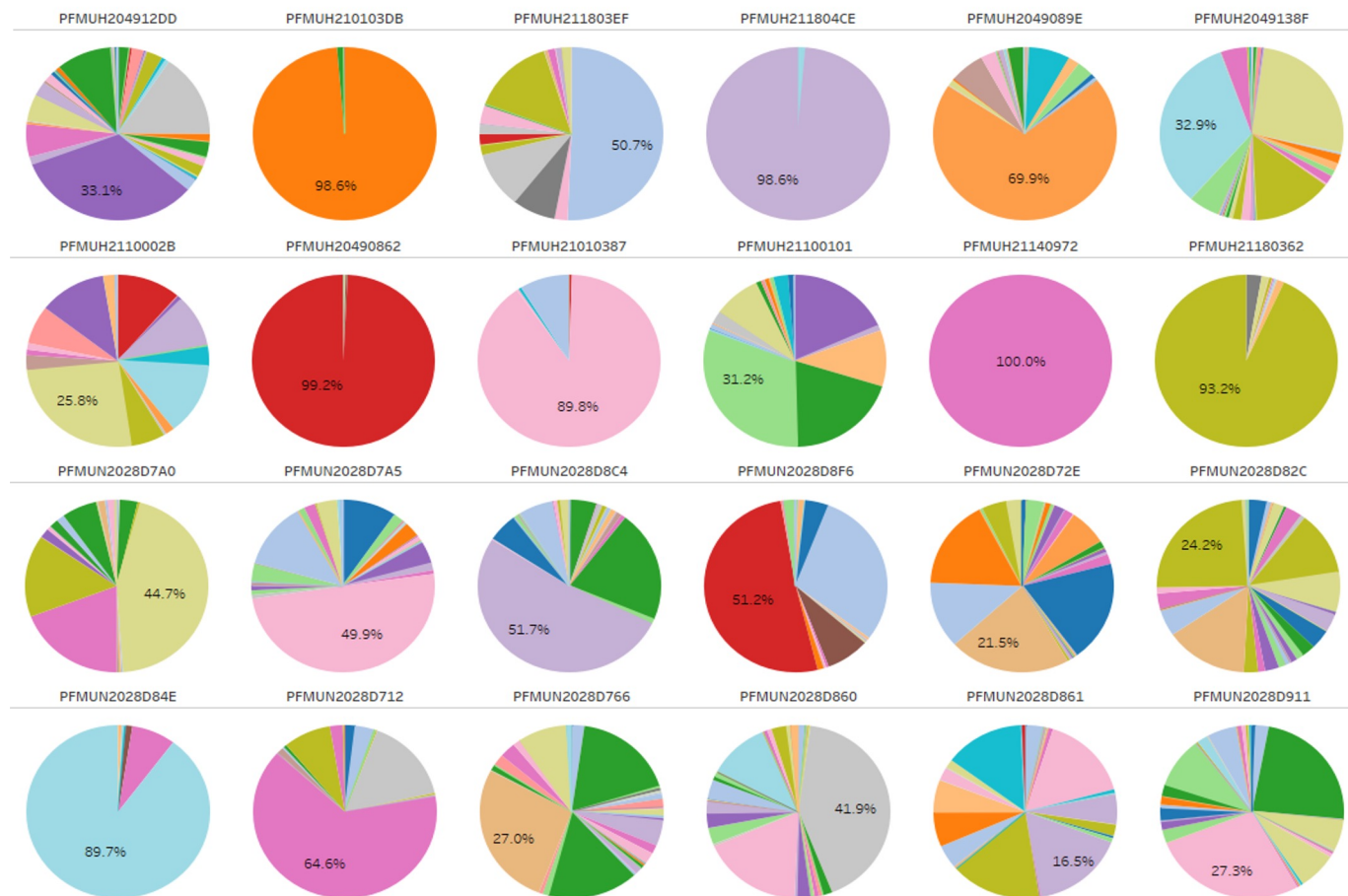
- Millions of die
- Billions of blocks
- Trillions of data points

Example: Analyzing trends for grown bad blocks



- One marker per drive
- For 99% of drives:
 - Grown Bad Blocks (GBB) are < 0.5% of total allowed bad blocks.
 - There are outliers with high GBBs – what's driving them?

Outlier analysis



Pie charts show breakup of bad blocks across die for some of the worst drives

Majority of GBB stem from 1 or 2 bad die

Tailor system solution to handle **“bad die”**

Targeted solutions to vintage-specific issues possible thanks to sophisticated telemetry