Accelerating RAID Rebuild and Reducing Write Amplification on High Density Solidigm QLC Drives With xiRAID

Dmitry Livshits, Chief Executive Officer, Xinnor

Sarika Mehta, Sr. Solutions Architect, Solidigm

Agenda

- xiRAID Introduction
- Solidigm High-density QLC Total Cost of Ownership
- Recovery Time and Large Capacity Rebuild Concern
- Background Terms
- Configuration and Results
- Summary
- Conclusion

The Xinnor idea

The world demands ever-faster access to data, without compromises on data integrity

Xinnor's know-how:

1. Fastest data encoding

2. Unique vector calculations approach

3. Optimal system resource allocation

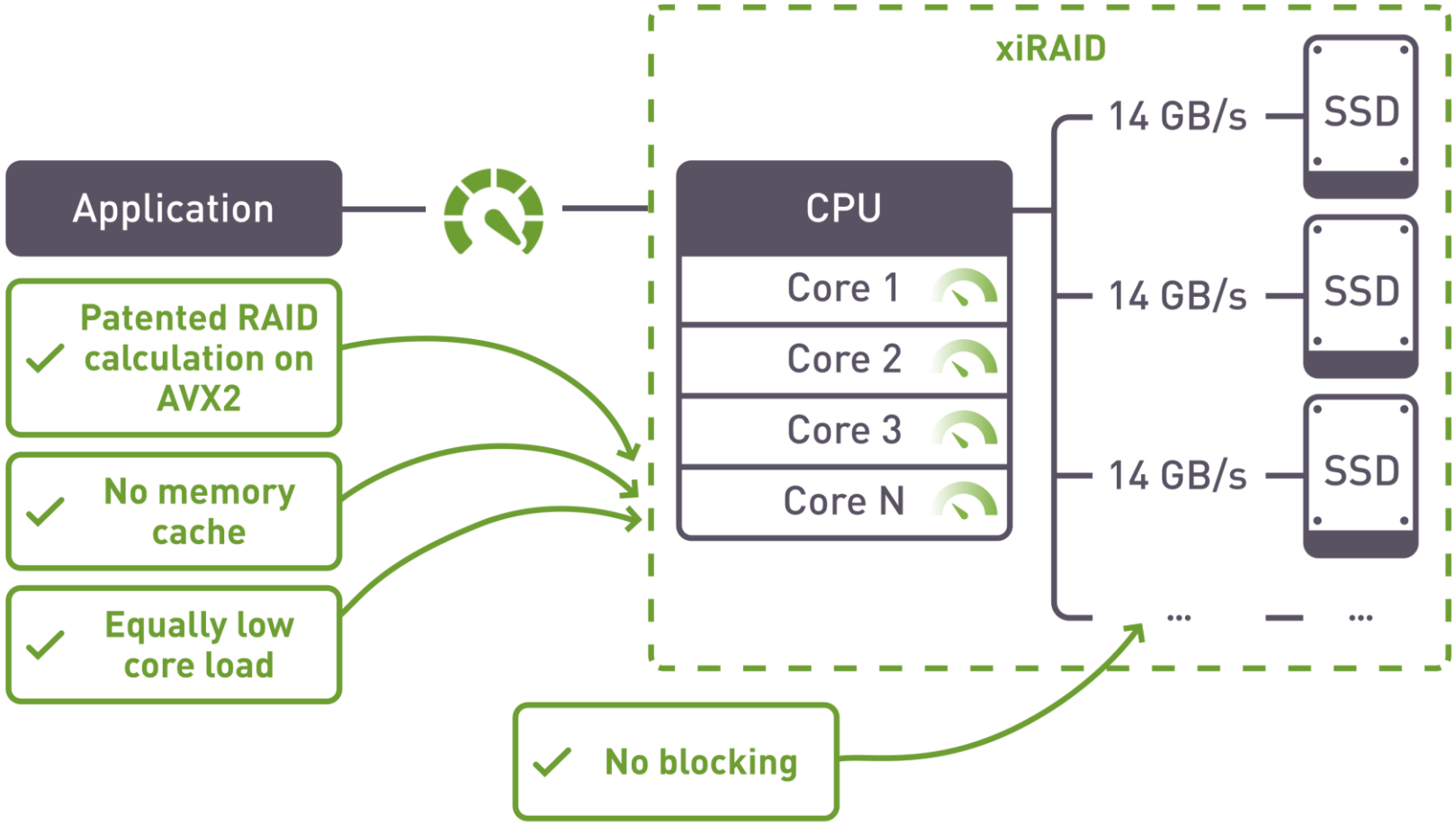
Xinnor delivers extreme performance in accessing data, while ensuring protection against hardware and software failures and their consequences:

- **xiRAID** – the fastest data protection engine.
  Already in production
- **xiSTORE** – high-performance and reliable storage, specifically designed for AI workloads offering the best TCO.
  In development

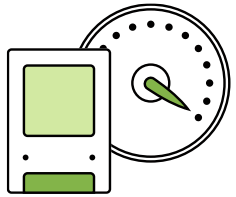
Xinnor's xiRAID Classic

CPU-assisted RAID (AVX)

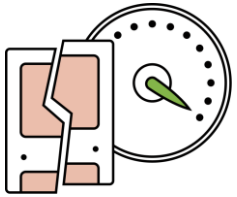
Lockless data path



xiRAID: High performance RAID engine
Competitive advantages

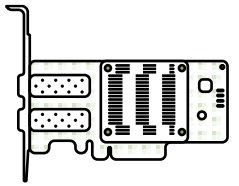


Superior performance in normal operation



High performance in degraded mode

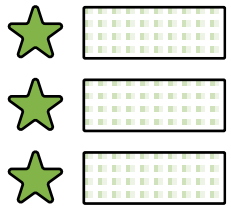
>10X performance boost vs competitive options



No PCIe taxation

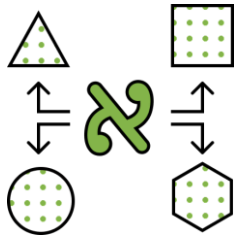
Minimal CPU load for checksum calculation

No need for dedicated hardware



Features rich

RAID Migration, Restriping, Hot Spare, Notification, Merge, Write-hole protection, Variable strip-size, CPU affinity, High Availability



Flexibility

Any drive, any capacity, RAID level 0,1,5,6,7.3,10, 50,60,70 and N+M supported, up to 64 drives in a single RAID, unlimited number of RAID

"By 2028, more than 65% of on-premises storage capacity will be based on QLC NVMe SSD, compared to less than 28% in 2025."

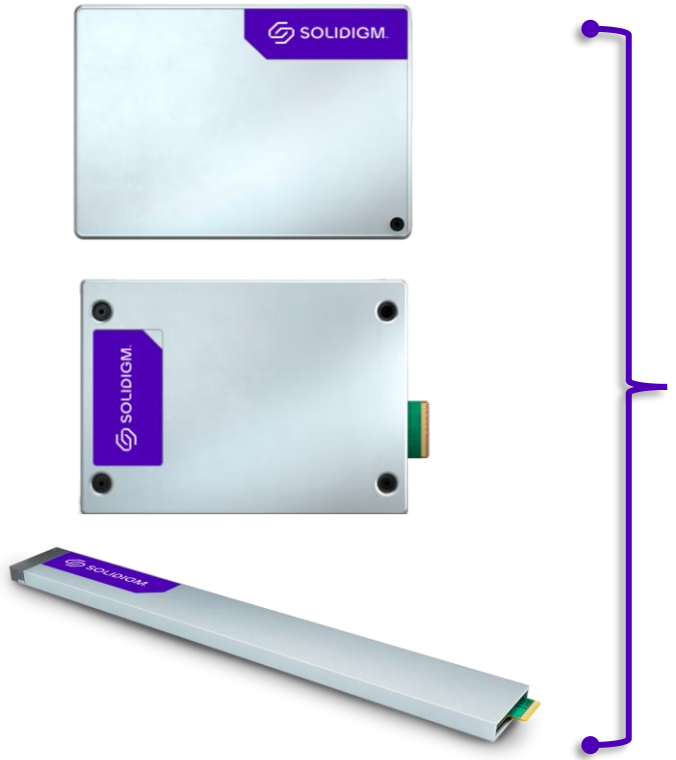
Gartner, "Hype Cycle for Storage Technologies, 2025"

Solidigm™ D5-P5336
122TB in Full Volume Production Enabling AI Deployments

U.2 15mm
7.68TB – 122.88TB

E3.S 7.5mm
7.68TB – 30.72TB

E1.L 9.5mm
15.36TB – 122.88TB



- Accelerate storage in **AI and data-intensive** workloads
- **Massive scalability** for high-density storage environments
- **Substantially improve TCO and sustainability** in hyper-scale environments

Solidigm™ D5-P5336 61.44TB **value vs. all-HDD**

Total Cost of Ownership Value When Solving for **10PB** Object Storage Solution	**30TB HDD** 3.5″	**Solidigm D5-P5336** U.2 61.44TB	
# of capacity drives	1,429	**343**	**~4.2x fewer drives**
Total # of servers	120	**15**	**~8x fewer servers**
Total racks	7	**1**	**7x smaller rack footprint**
TBe/W	4.04	**11.53**	**2.9x better power density**
5-year energy cost	$2.07M	**$266K**	**~7.8x lower energy cost**
5-year total cost	$3.92M	**$2.06M**	**47% BETTER TCO**[1]

1 Source – Solidigm. See Appendix 2: 30TB HDD vs QLC 61.44TB TCO Parameters for configuration details and assumptions.

Food for Thought

In enterprise environments, the key question is:
how fast can we recover when hardware needs replacement?

Drive capacities are exploding, but rebuild times scale linearly

Traditional RAID solutions create unacceptable business risks

Why this matters:

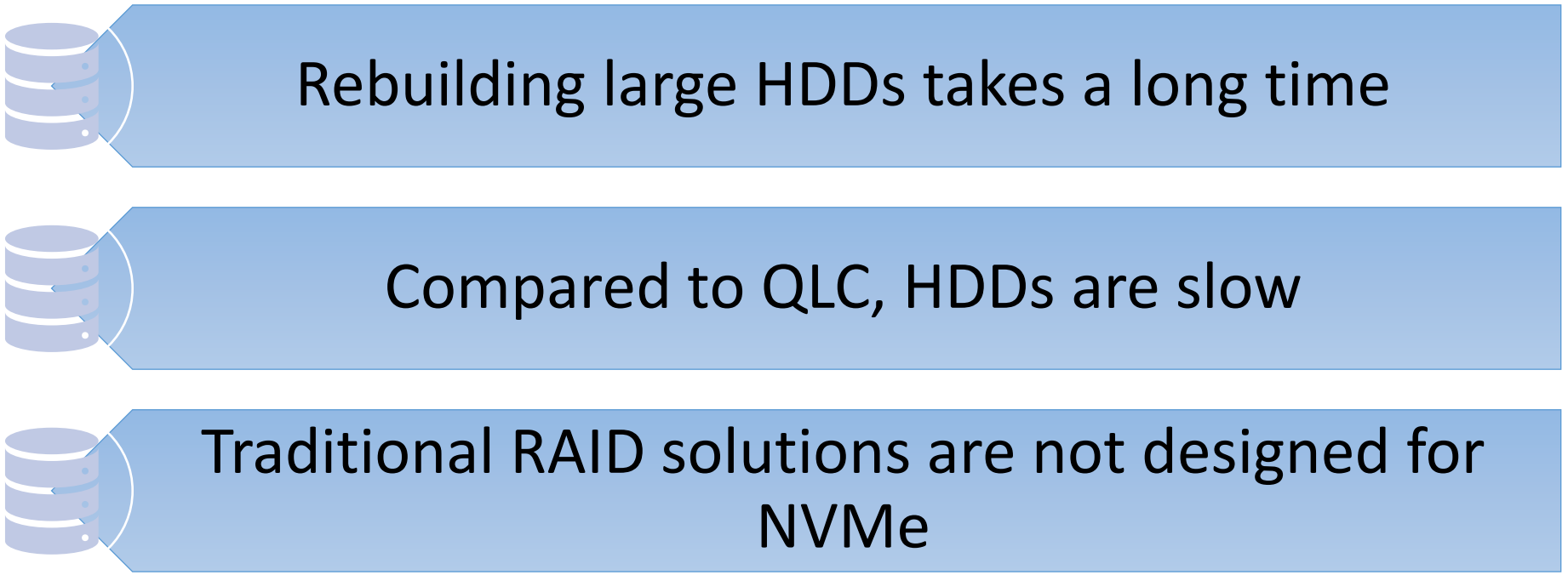
Business Continuity Risk

- Degraded mode operation = reduced performance for all applications
- Extended vulnerability window = exponentially higher data loss risk

Economic Impact

- Massive capacity investments sitting in compromised state
- Performance bottlenecks affecting entire infrastructure

Rebuild Concern for High Density QLC



How long does it take to rebuild a Solidigm™ D5-P5336 61.44TB QLC SSD?

Background:
Write Amplification Factor (WAF)

- Write amplification factor is a well-known measure of NAND SSD endurance.
- It is also a measure of storage stack efficiency.
- Lower WAF is better.

$$\text{WAF} = \frac{\text{NAND writes for the workload}}{\text{Host writes for the workload}}$$

Background: Parity RAID Rebuild

- Parity RAIDs (e.g., RAID5, RAID6) use mathematical syndrome formulas for parity calculations.
- These calculations can be time-consuming, depending on RAID engine efficiency.
- Rebuilds involve RAID recalculating data for a new drive to replace the failed one.

Rebuild speed depends on RAID checksum and recalculation performance.

Measuring RAID Rebuild Performance – RAID configuration

RAID engines:

- mdadm 4.3
- xiRAID Classic 4.3

RAID configuration:

- RAID5 of 9 x Solidigm™ D5-P5336 61.44TB: **~0.5PB in total**
- mdadm set `speed_limit_max` to 9GB/s
- xiRAID Classic additional setting: reconstruction priority set to 80%

Host writes and NAND writes are captured via SSD SMART logs before and after the reconstruction to calculate WAF.

Results – Rebuild Without Workload

Rebuilding 1x Solidigm D5-P5336 61.44TB QLC

RAID engine	**Rebuild time**	**Rebuild speed**	**WAF** (lower is better)
mdraid	53h 40m	322MB/s	1.2
xiRAID Classic 4.3	5h 22m 10x faster rebuild	3.18GB/s 10x higher throughput	1.02 15% lower WAF

See Appendix 1: Configuration for system and workload configuration

Results – Rebuild With Workload

Rebuilding 1x Solidigm D5-P5336 61.44TB QLC

RAID Engine	**Rebuild time**	**Rebuild speed**	**WAF** (lower is better)	**Workload speed under rebuild**
mdraid	>67 days	10.5 MB/s	1.58	Read: ~100MB/s Write: ~45MB/s
xiRAID Classic 4.3	53h 53m 25x faster rebuild	316 MB/s 30x higher throughput	1.21 23% lower WAF	Read: 44GB/s Write: 13GB/s 290-440x higher

See Appendix 1: Configuration for system and workload configuration

Summary

Solidigm™ D5-P5336 QLC

- High-density QLC capacities
- Storage infrastructure compaction with power and cost advantage
- Delivers on demands of RAID rebuild and host workload simultaneously

xiRAID Classic

- Leverages the raw SSD performance to deliver fast drive rebuild.
- Provides excellent workload performance during RAID rebuild.
- Generates low WAF, enhancing the endurance of the SSDs.

Conclusion

Accelerated RAID Rebuild

Enhanced Storage Endurance

High Host Workload Performance During Rebuild

xiRAID Classic + Solidigm™ D5-P5336 Deliver an Impeccable High-density RAID Solution

Thank you

Appendix 1: Configuration

Test System Configuration:

System: Dell PowerEdge R760

BIOS: Vendor: Dell Inc.; Version: 2.3.5

CPU: Intel(R) Xeon(R) Gold 6430 x 2; 2 x sockets @3.4GHz, 32 cores/per socket

NUMA Nodes: 2

DRAM: Total 512G DDR4@3200 MHz

OS: Rocky Linux 9.5

Kernel: 5.14.0-503.22.1el9_5.x86_64

SSD: 10 x Solidigm D5-P5336 61.44TB, FW Rev: 5CV10302, PCIe Gen4x4

Fio: Version: 3.35

xiRAID: Version: 4.3

Mdraid: Version: 4.3 2024-02-15 - 3

Host Workload:

```
[global]
iodepth=64
direct=1
ioengine=libaio
group_reporting
runtime=604800 # a week

[write]
rw=write
bs=1MB
numjobs=38
offset_increment=15G
filename=/dev/<raid_device_name>

[read]
rw=read
bs=1MB
numjobs=90
offset_increment=10G
filename=/dev/<raid_device_name>
```

Appendix 2: 30TB HDD vs QLC 61.44TB TCO Parameters

	Seagate Exos Mosaic 30TB	Solidigm D5-P5336 122.88TB
Drive Cost ($/GB)	1x $/GB	6x $/GB
Deployment Term (years)	5	5
Max Read B/W (MB/s)	OP'd to 250	7,000
AFR	1.57%	0.44%
Average Active Read Power (W)	9.5	16
Idle Power (W)	6.9	5
Power Cost ($/KWHr)	$0.25	
PUE	1.35	
Drives Per 2U Server	12	24
Storage U Available per 42U Rack*	36	
Non-Storage Power per Server (W)	1080	
Empty Rack Purchase Cost	$1,100	
Storage Server Purchase Cost	$10,335	$14,222
Drive Duty Cycle	20%	3%
Capacity Utilization Factor	70%	95%
Error Encoding Factor	3x (Hadoop Triplication)	2x (RAID 1)
Data Reduction/Compression Factor	1.0:1 (none)	

* Assumes 6U per rack reserved for networking/other

Solidigm Disclaimers

All information provided here is subject to change without notice.

- The products described in this document may contain design defects or errors known as errata, which may cause the product to deviate from published specifications. Current characterized errata are available on request.
- Solidigm technologies' features and benefits depend on system configuration and may require enabled hardware, software or service activation. Performance varies depending on system configuration. No computer system can be absolutely secure. Check with your system manufacturer or retailer.
- Solidigm disclaims all express and implied warranties, including without limitation, the implied warranties of merchantability, fitness for a particular purpose, and non-infringement, as well as any warranty arising from course of performance, course of dealing, or usage in trade.
- Tests document performance of components on a particular test, in specific systems. Differences in hardware, software, or configuration will affect actual performance. Consult other sources of information to evaluate performance as you consider your purchase.
- Cost reduction scenarios described are intended as examples of how a given Solidigm-based product, in the specified circumstances and configurations, may affect future costs and provide cost -savings. Circumstances will vary. Solidigm does not guarantee any costs or cost reduction.
- Solidigm does not control or audit the design or implementation of third-party benchmark data or Web sites referenced in this document. Solidigm encourages all of its customers to visit the referenced Web sites or others where similar performance benchmark data are reported and confirm whether the referenced benchmark data are accurate and reflect performance of systems available for purchase.
-