



Key Attributes and Tools for Testing PCIe Based Solid State Storage

Presenter: Bob Weisickle
CEO/Founder OakGate Technology, Inc



Flash Memory Summit Agenda

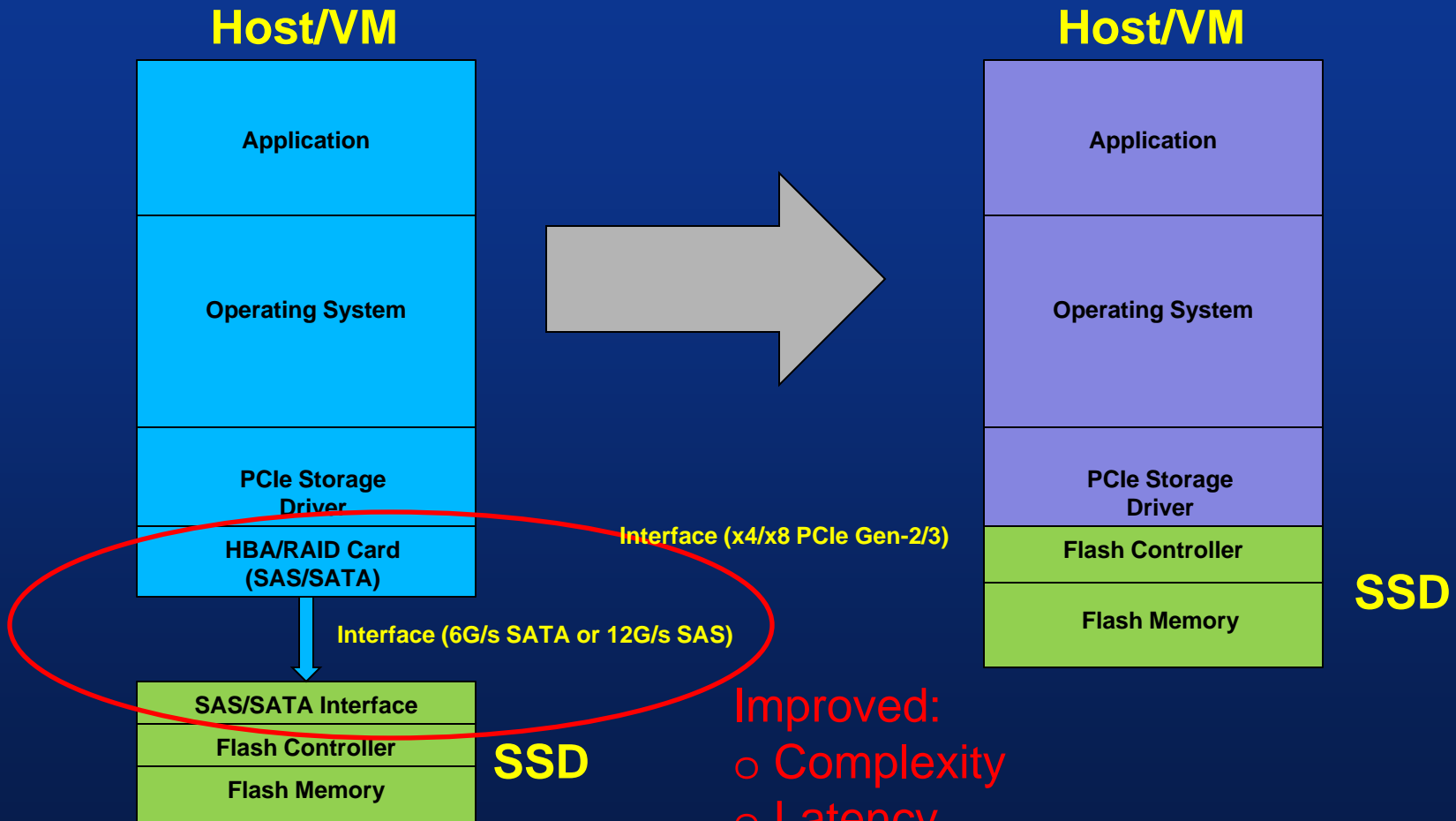
- Premise and Challenge of PCIe Solid State Storage
- Test Environment for Storage Providers and End Users
- What are the unique attributes that distinguishes PCIe SS Storage from SAS/SATA
- Tools and Testing approaches to address some key attributes

PCIe Premise/Challenge

- ◆ Better Performance
 - ◆ Faster Interface
 - ◆ Closer to CPU
- ◆ More CONSISTENT performance (esp. Latency)
- ◆ Simpler Architecture
- ◆ Remove bottlenecks to the Flash

The Premise

PCIe Premise/Challenge



- Improved:
- Complexity
 - Latency
 - Performance
 - Reliability

PCIe Premise/Challenge

◆ Multiple/Emerging PCIe Protocols

- ◆ Vendor Proprietary
- ◆ NVMe
- ◆ SCSIe (SCSI over PCIe)
- ◆ SATAe (AHCI)

◆ New Storage Architecture/Topology

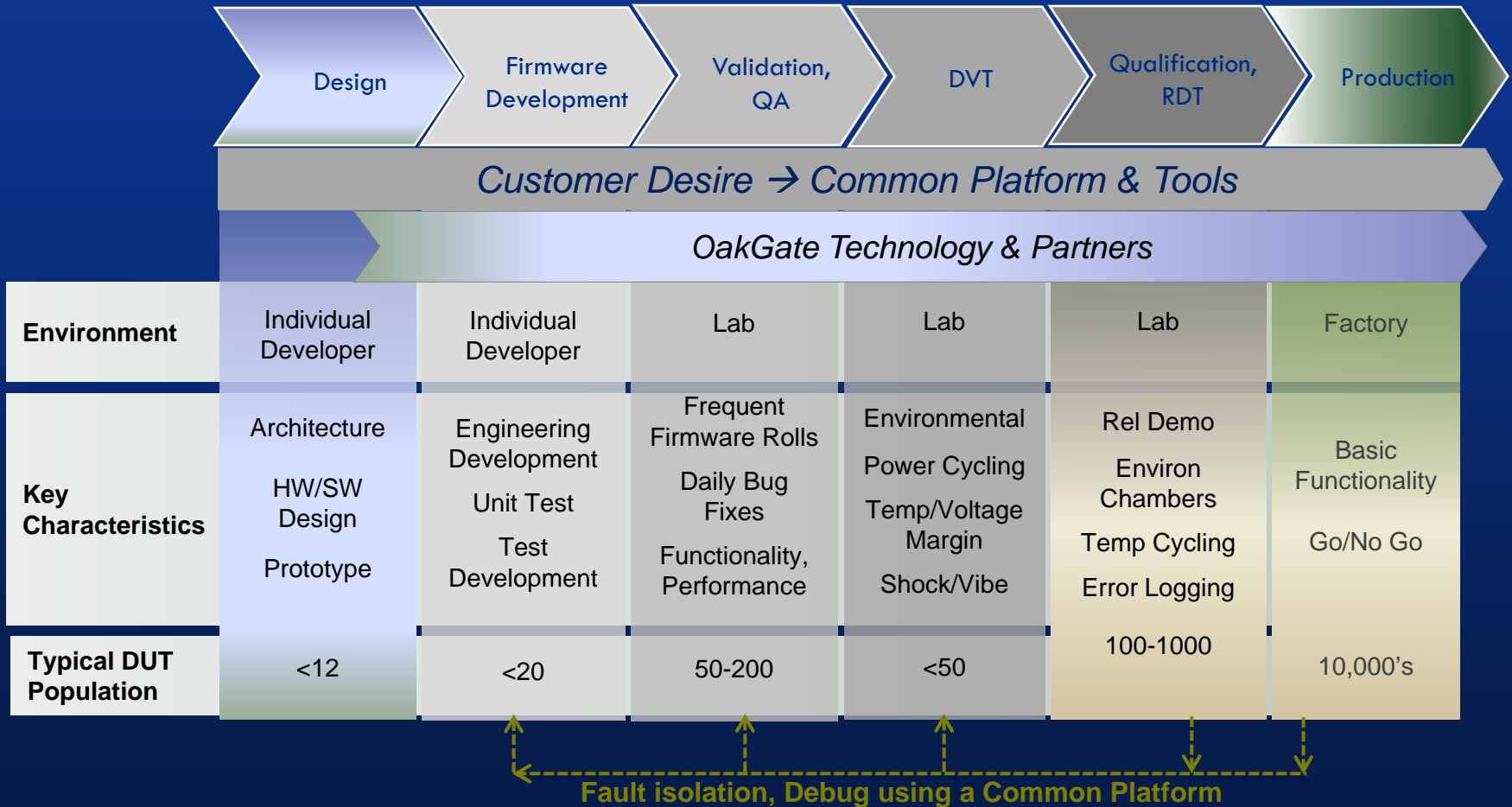
- ◆ PCIe Switching/Bridging
- ◆ New Interfaces/Cables/Enclosures
- ◆ Multi-Host (Physical or Virtual)
- ◆ Multiple Queues
- ◆ Dual Port SFF Architectures

◆ Higher Performance Levels

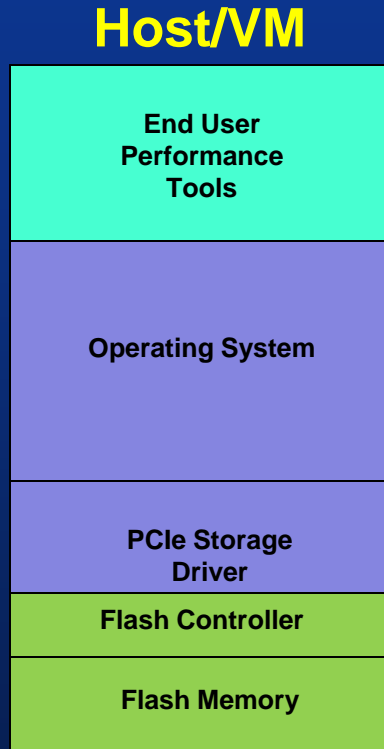
- ◆ Tools to generate/measure xMillion IOPS
- ◆ Measure very short IO Latency
- ◆ Scale with Virtual Machines

The Challenge

Environment --- Storage Providers



Environment --- End Users



SSD

Performance Measurement

- User space Applications like IOmeter
- Limited ability to tune based on 'real' application workload

Dependent on Supplier

- BIOS compatibility
- IO Failure Analysis
- Performance Issues
- Limited means to provide information back to supplier (logs, traces, etc)
- Suppliers don't 'share' internal tools

Extend the Test Model to encompass the User

Distinguishing Attributes of PCIe SSD Storage

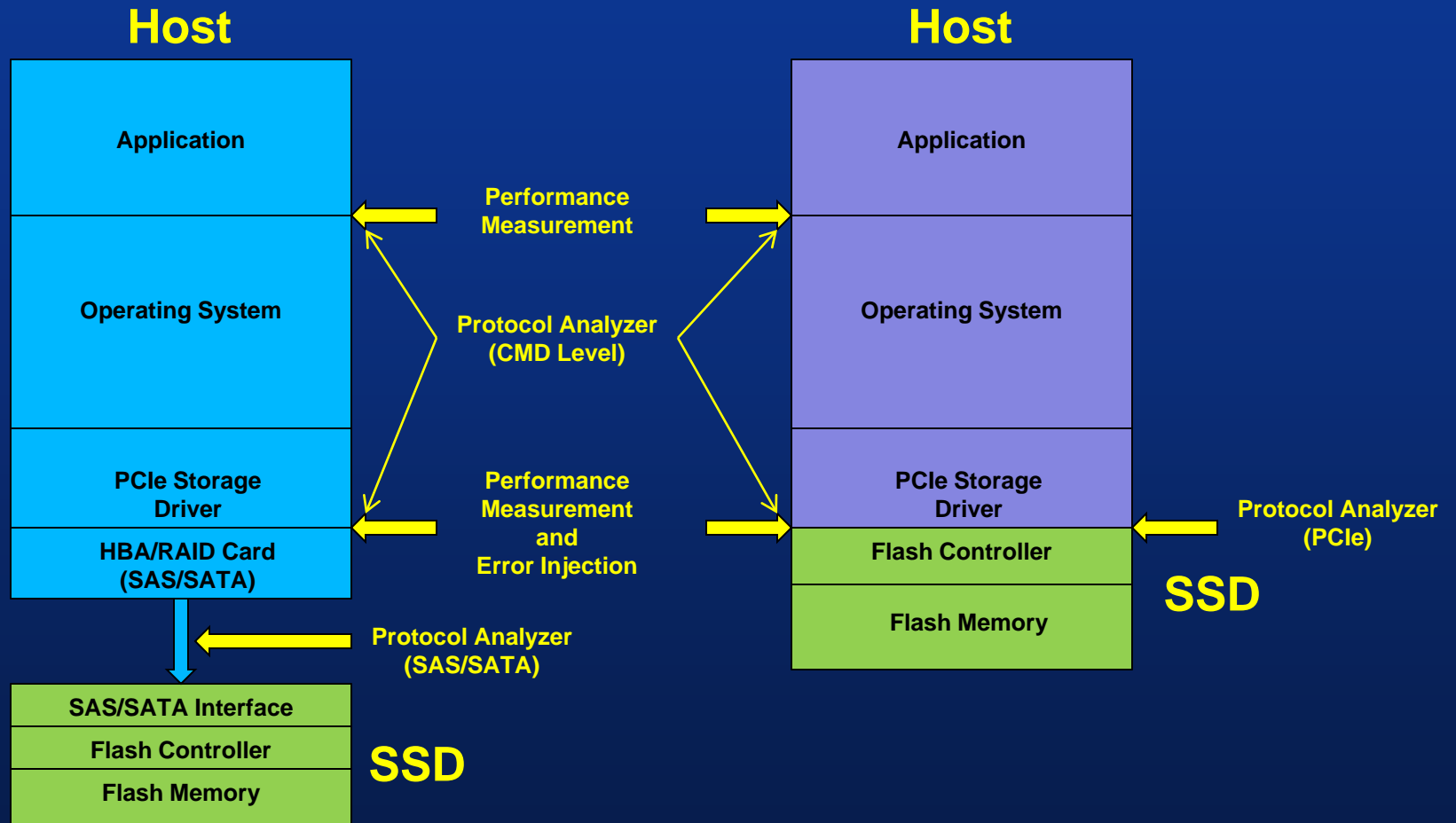
- ◆ What are the unique attributes that distinguishes PCIe SS Storage from SAS/SATA
 - ◆ Emerging protocol(s)
 - ◆ Maturity (Most are 1st Generation releases)
 - ◆ Standards and Proprietary
 - ◆ Architecture implementation differences
 - ◆ Simplicity of Protocols (opportunity)
- ◆ Complexity of Queue'ing model
- ◆ Requirements of Tools and Testing approaches
 - ◆ Evaluation of HW versus HW + Driver
 - ◆ Queue strategy versus Performance
 - ◆ Performance Characterization
 - ◆ Error Injection Methods



Tools and Testing Approaches

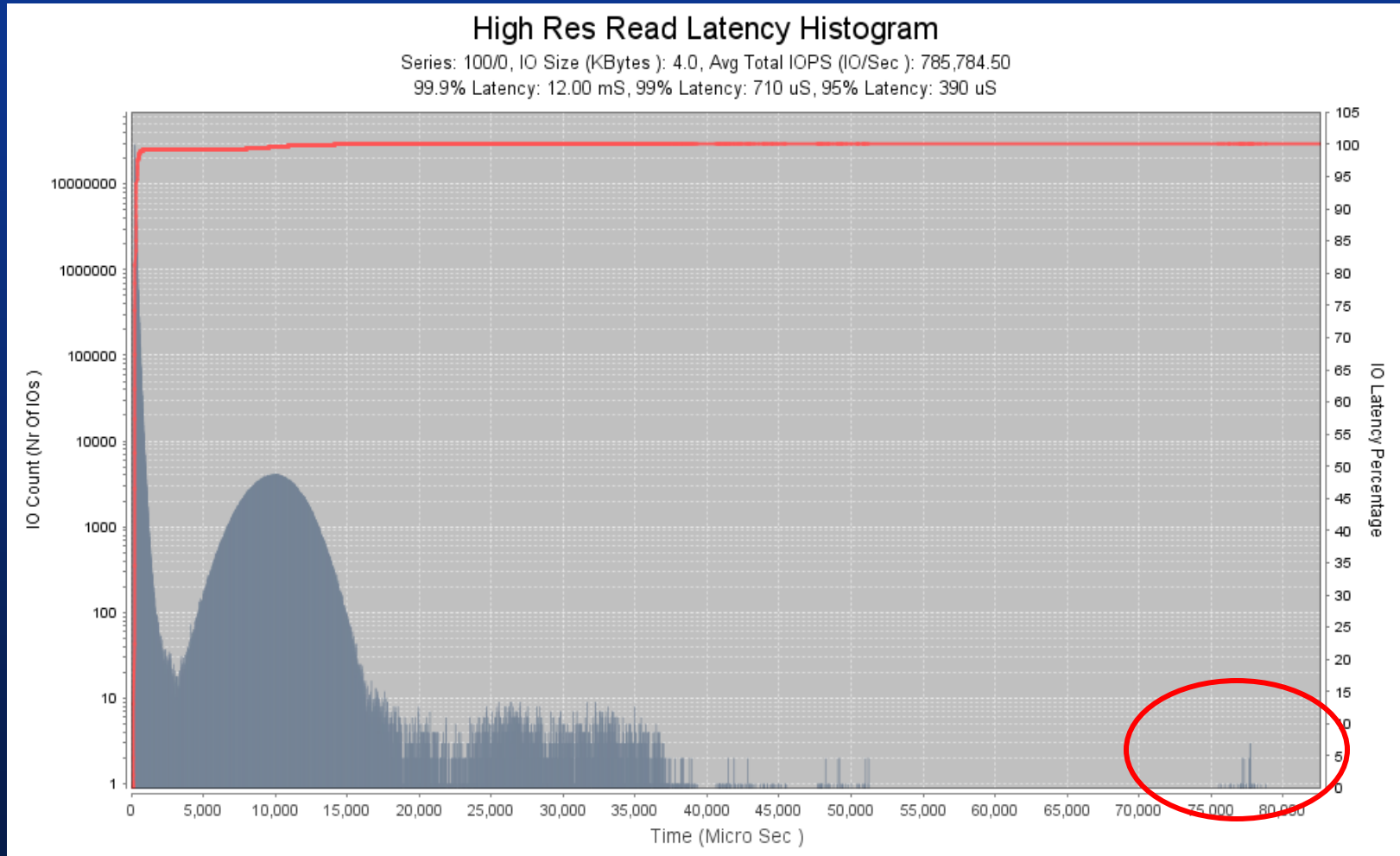
- ◆ Performance Measurement
- ◆ Impact of Multiple Queues
- ◆ Role of Error Injection

Performance Measurement Points



Performance Measurement Examples

Benchmarking

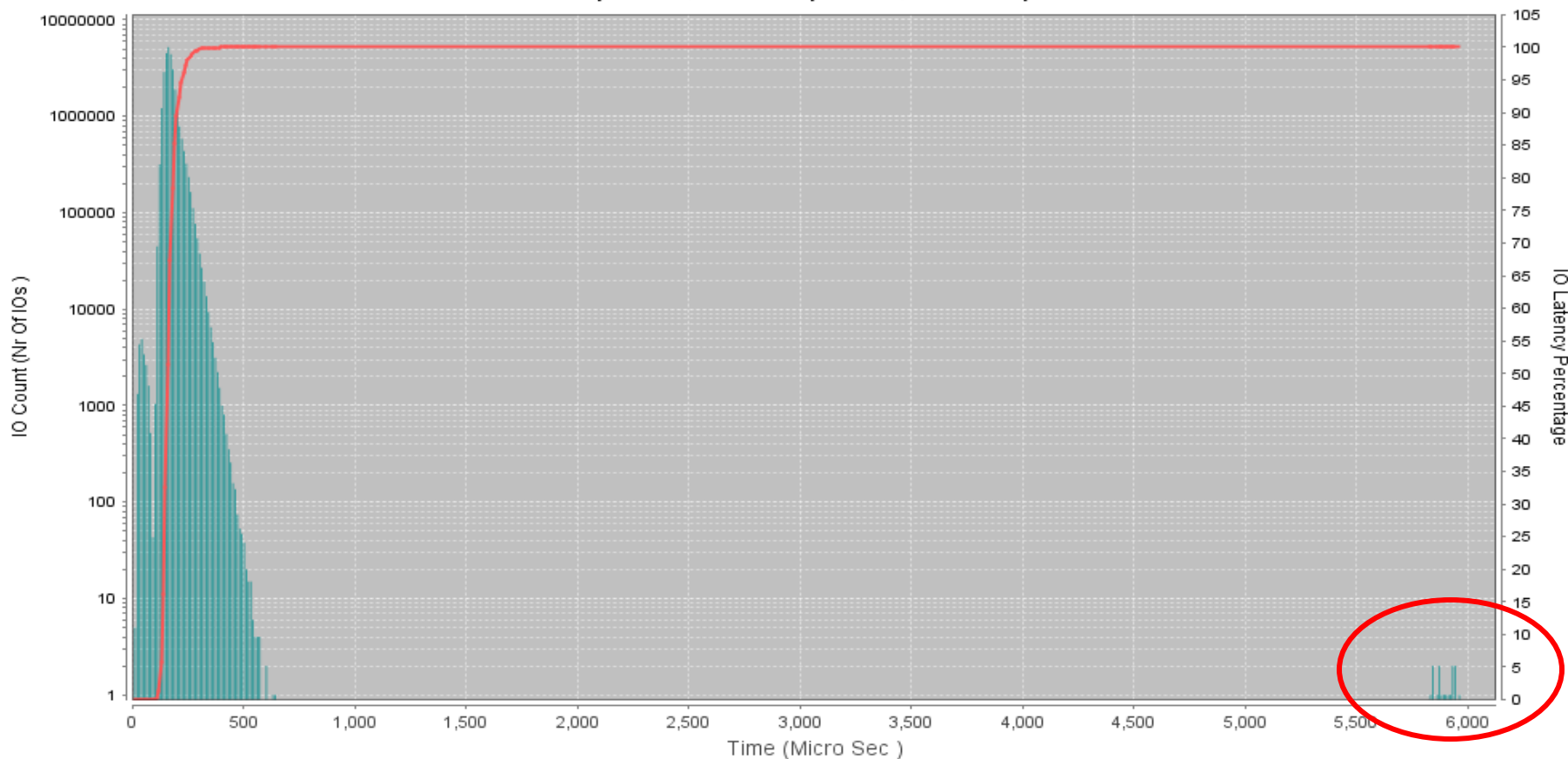


Performance Measurement Examples

Load Based Performance

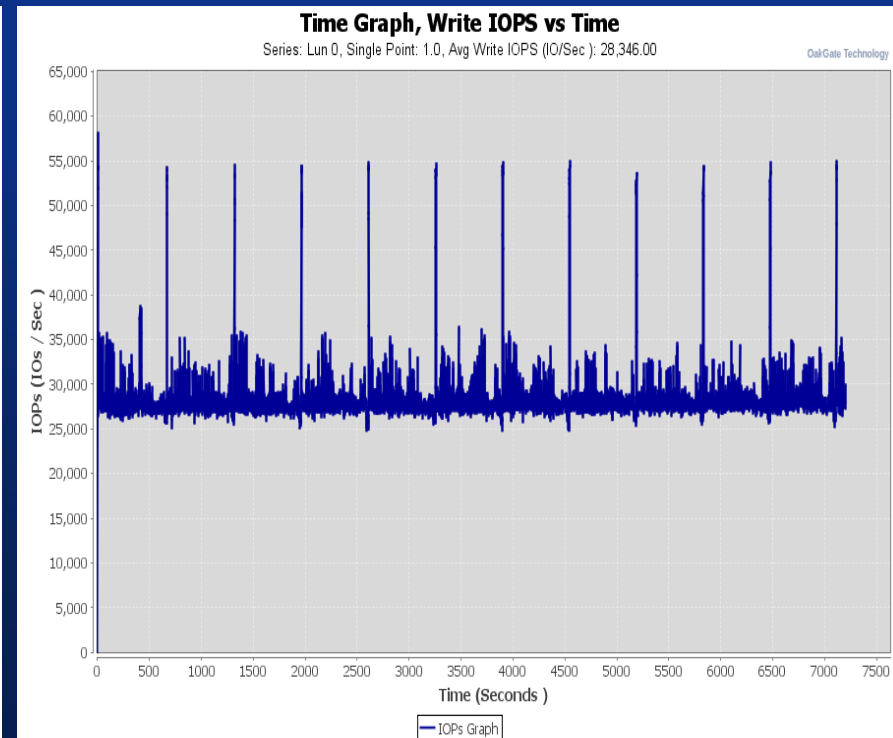
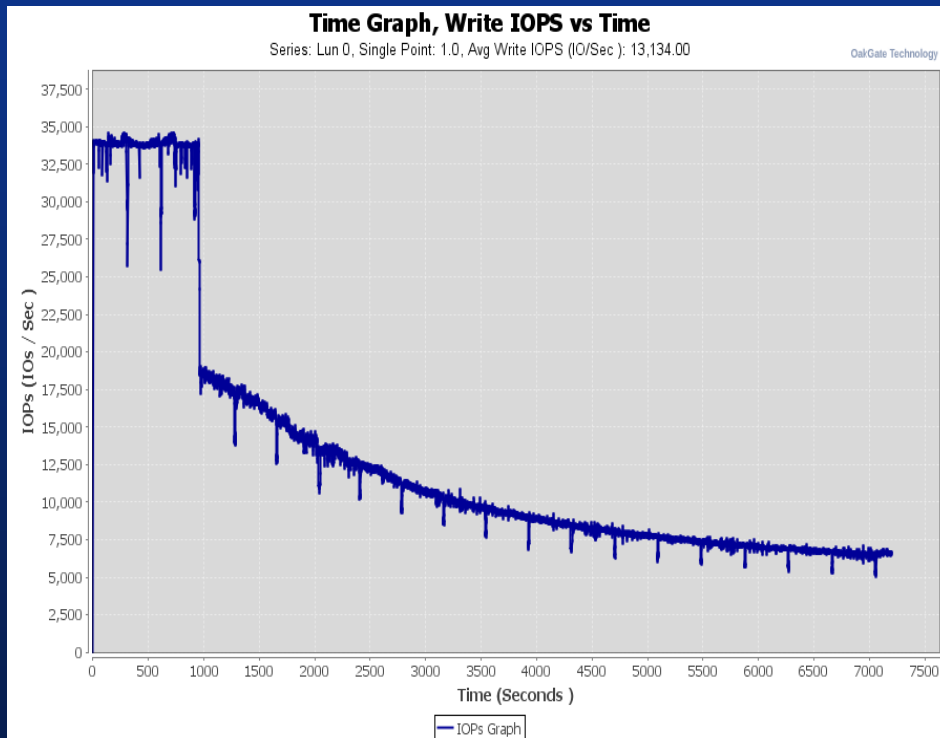
High Res Read Latency Histogram

Series: 16, Loop on: IO Size (IO Size): 4.0, Avg Read IOPS (IO/Sec): 92,345.00
99.9% Latency: 340 uS, 99% Latency: 270 uS, 95% Latency: 230 uS



Performance Measurement Examples

Performance vs Time

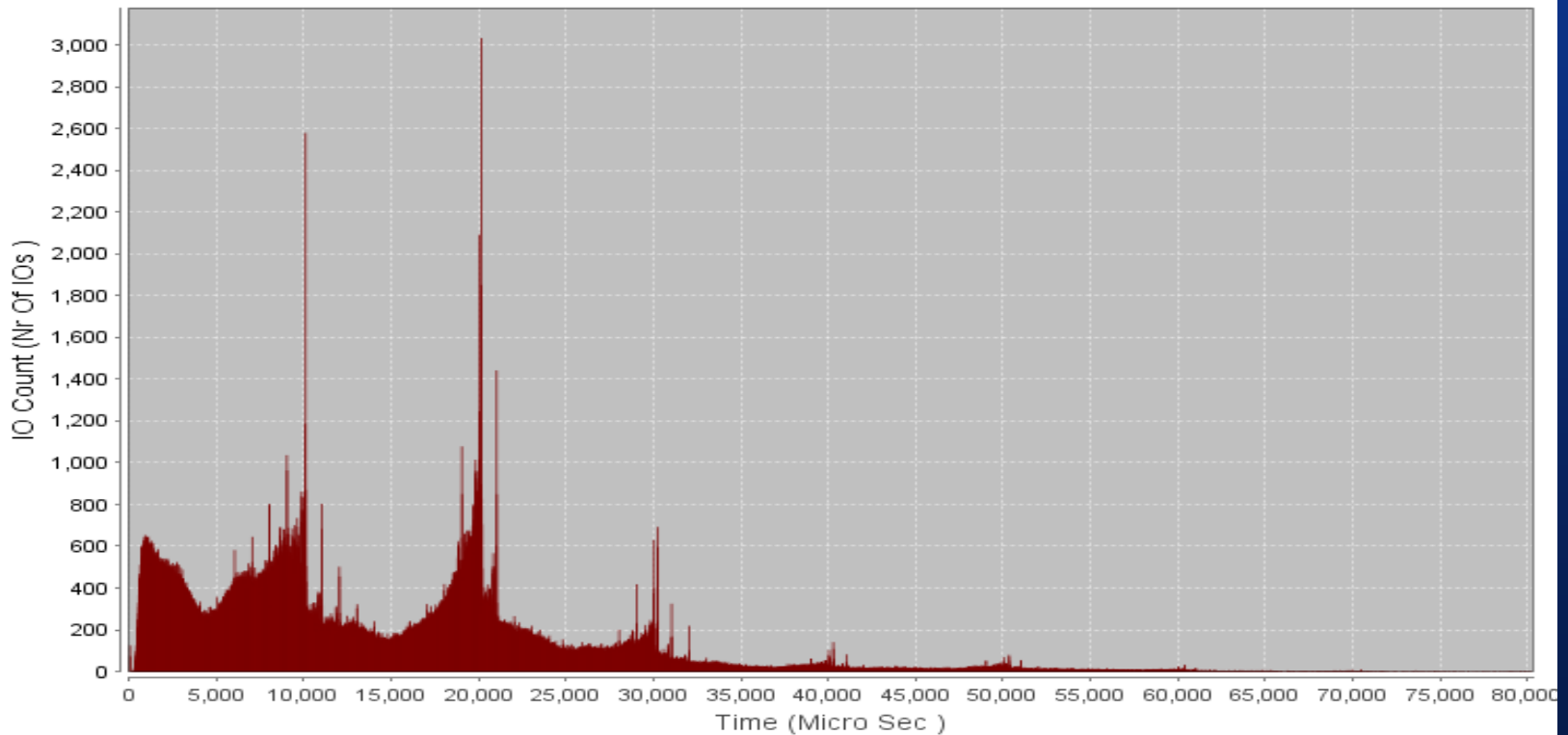


Performance Measurement Examples

VM Environment

Read Latency:Read Latency

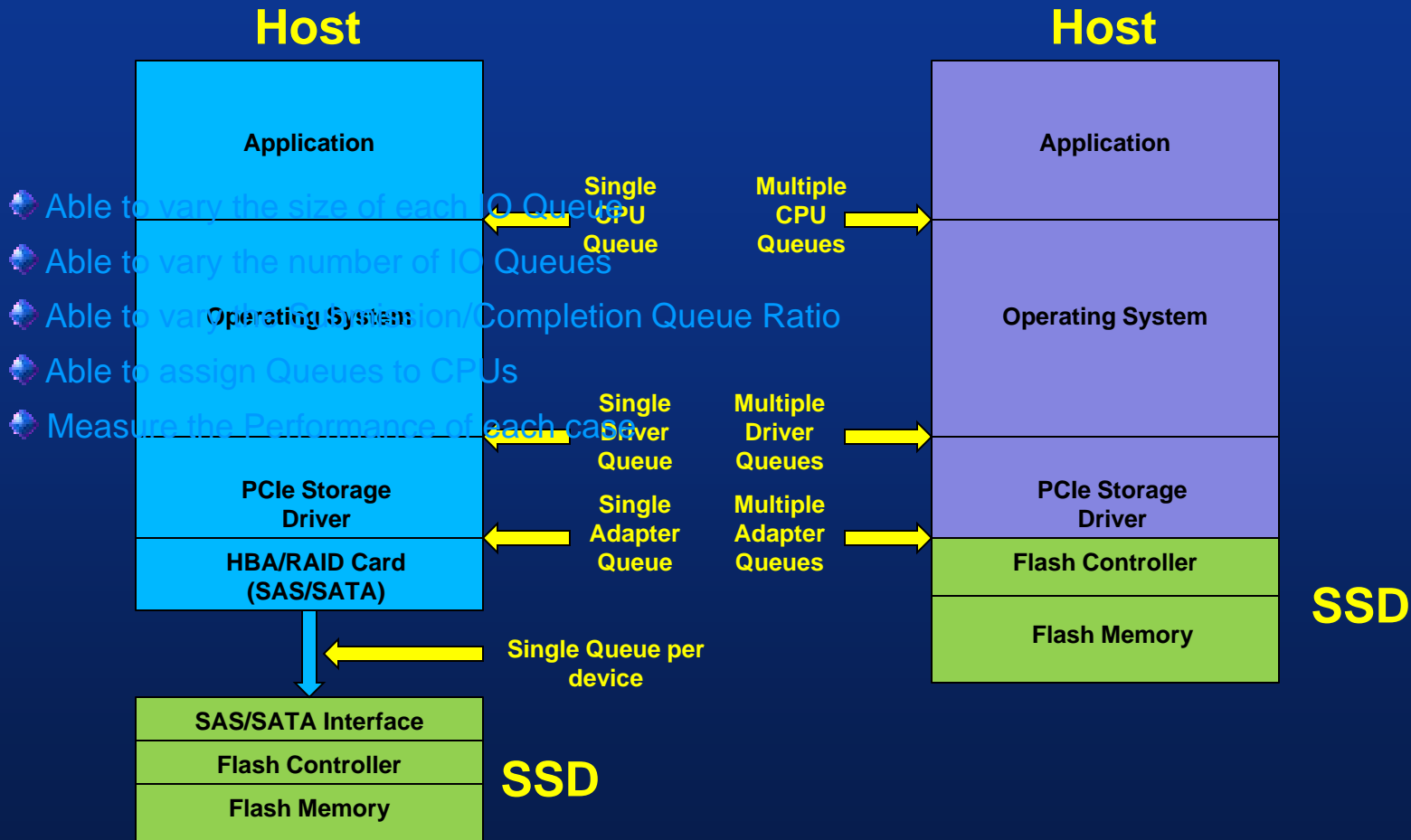
Lun: BK4 /dev/sdd, VMware, Virtual / VMware, Virtual, SN: Unknown SN, LunID: 0
Set Init Cond: DATA_PATTERN = Entropy 100
Test: Eight Systems
Loop On Tests: 8.0 Number Of Systems



Multiple Queues

- ◆ SAS/SATA used a single Queue to manage IO and Management Commands
- ◆ PCIe Proprietary drivers also follow a single Queue model
- ◆ Queuing Characteristics of new PCIe Storage Protocol Standards is different
 - ◆ Each Storage Device will support multiple Queue sets (IO Submission/Completion)
 - ◆ Each Host CPU will have there own Unique Queue sets
 - ◆ Management commands have their own unique Queue set
 - ◆ Performance may vary with the size and number of Queues used
 - ◆ Addressing the needs of Virtual Machines

Multiple Queue Measurement



NVMe/SCSIe Queue Configuration

Device Control: Gemu NVMe Driver 0xabcd

Control | Port Info | PHY Settings | Queue Config

Queue Init

- NVMe Queue Layout
 - Comp Queue 1
 - Comp Queue 2
 - Comp Queue 3
 - Comp Queue 4
 - Comp Queue 5
 - Comp Queue 6
 - Comp Queue 7

Queue Configuration | IO Distribution

Queue ID	IO Distribution	Weight (%)	Lock
1 - 1		17.00 %	<input type="checkbox"/>
1 - 2		3.88 %	<input type="checkbox"/>
1 - 3		3.89 %	<input type="checkbox"/>
1 - 4		3.89 %	<input type="checkbox"/>
2 - 1		3.89 %	<input type="checkbox"/>
3 - 1		3.89 %	<input type="checkbox"/>
3 - 2		3.89 %	<input type="checkbox"/>
3 - 3		3.89 %	<input type="checkbox"/>
3 - 4		3.89 %	<input type="checkbox"/>
3 - 5		3.89 %	<input type="checkbox"/>
3 - 6		3.89 %	<input type="checkbox"/>
4 - 1		13.00 %	<input type="checkbox"/>
5 - 1		3.89 %	<input type="checkbox"/>
6 - 1		3.89 %	<input type="checkbox"/>
6 - 2		3.89 %	<input type="checkbox"/>
6 - 3		3.89 %	<input type="checkbox"/>
6 - 4		3.89 %	<input type="checkbox"/>
6 - 5		3.89 %	<input type="checkbox"/>
6 - 6		3.89 %	<input type="checkbox"/>
7 - 1		3.88 %	<input type="checkbox"/>

Queue Selection Mode: **Weighted Queue Selection**

- Random Queue Selection
- Round-Robin Queue Selection
- Weighted Queue Selection

Nr Of Admin Q Entries: Max Nr Of Queues:



Analyzer Command Verification

Gemu NVMe Driver 0xabcd

Analyzer

State: NOT CAPTURING Save/Load Trace Progress Bar: [] [Start Capture] [Stop Capture]

View Context: All IT Nexus ITL Nexus IO Context

Frame Navigation: [CMD] [◀] [▶] [RSP] [H]

Index	Direction	Time	Frame Type	Decoded Frame	LBA	Sector Size	IO Len (Blks)	IO Thread	IO Tag	Queue ID	Name Space	IO Duration
NVME 15	➡ To Tgt	0.132072s	Command	Create I/O Submission Queue				0	IO Tag 0xc	Sub Q: A...		
NVME 16	➡ To Tgt	0.132079s	Command	Create I/O Submission Queue				0	IO Tag 0xd	Sub Q: A...		
NVME 17	➡ To Tgt	0.132087s	Command	Create I/O Submission Queue				0	IO Tag 0xe	Sub Q: A...		
NVME 18	➡ To Tgt	0.132096s	Command	Create I/O Submission Queue				0	IO Tag 0xf	Sub Q: A...		
NVME 19	➡ To Tgt	0.132104s	Command	Create I/O Submission Queue				0	IO Tag 0x10	Sub Q: A...		
NVME 20	➡ To Tgt	0.132112s	Command	Create I/O Submission Queue				0	IO Tag 0x11	Sub Q: A...		
NVME 21	⬅ To Ini	0.132119s	Response	Status: OK				0	IO Tag 0x3	Comp Q I...		.010267722
NVME 22	➡ To Tgt	0.132127s	Command	Create I/O Submission Queue				0	IO Tag 0x12	Sub Q: A...		
NVME 23	⬅ To Ini	0.132136s	Response	Status: OK				0	IO Tag 0x4	Comp Q I...		.010276626
NVME 24	➡ To Tgt	0.132143s	Command	Create I/O Submission Queue				0	IO Tag 0x13	Sub Q: A...		
NVME 25	⬅ To Ini	0.132152s	Response	Status: OK				0	IO Tag 0x5	Comp Q I...		.010282528
NVME 26	➡ To Tgt	0.132158s	Command	Create I/O Submission Queue				0	IO Tag 0x14	Sub Q: A...		
NVME 27	➡ To Tgt	0.132167s	Command	Create I/O Submission Queue				0	IO Tag 0x15	Sub Q: A...		
NVME 28	➡ To Tgt	0.132174s	Command	Create I/O Submission Queue				0	IO Tag 0x16	Sub Q: A...		
NVME 29	➡ To Tgt	0.132182s	Command	Create I/O Submission Queue				0	IO Tag 0x17	Sub Q: A...		
NVME 30	➡ To Tgt	0.132190s	Command	Create I/O Submission Queue				0	IO Tag 0x18	Sub Q: A...		
NVME 31	➡ To Tgt	0.132198s	Command	Create I/O Submission Queue				0	IO Tag 0x19	Sub Q: A...		
NVME 32	⬅ To Ini	0.132283s	Response	Status: OK				0	IO Tag 0x6	Comp Q I...		.010405718
NVME 33	➡ To Tgt	0.132292s	Command	Create I/O Submission Queue				0	IO Tag 0x1a	Sub Q: A...		
NVME 34	⬅ To Ini	0.132302s	Response	Status: OK				0	IO Tag 0x7	Comp Q I...		.000292691

Buf Fill [] 0% Bw 1 [] 0 MB/Sec Bw 2 [] 0 MB/Sec IOPS [] 0 IOs/Sec

◆ NVMe Submission

- ◆ Command: Create I/O Completion Queue
- ◆ Op Code: 0x5
- ◆ Fused Operation: 0x0
- ◆ Cmd Identifier: 0x0

Error Type	Buffered Errors	Seen Errors

Decoded Frame Raw Frame Settings

Event Info LL Stats Commands Warnings View Search

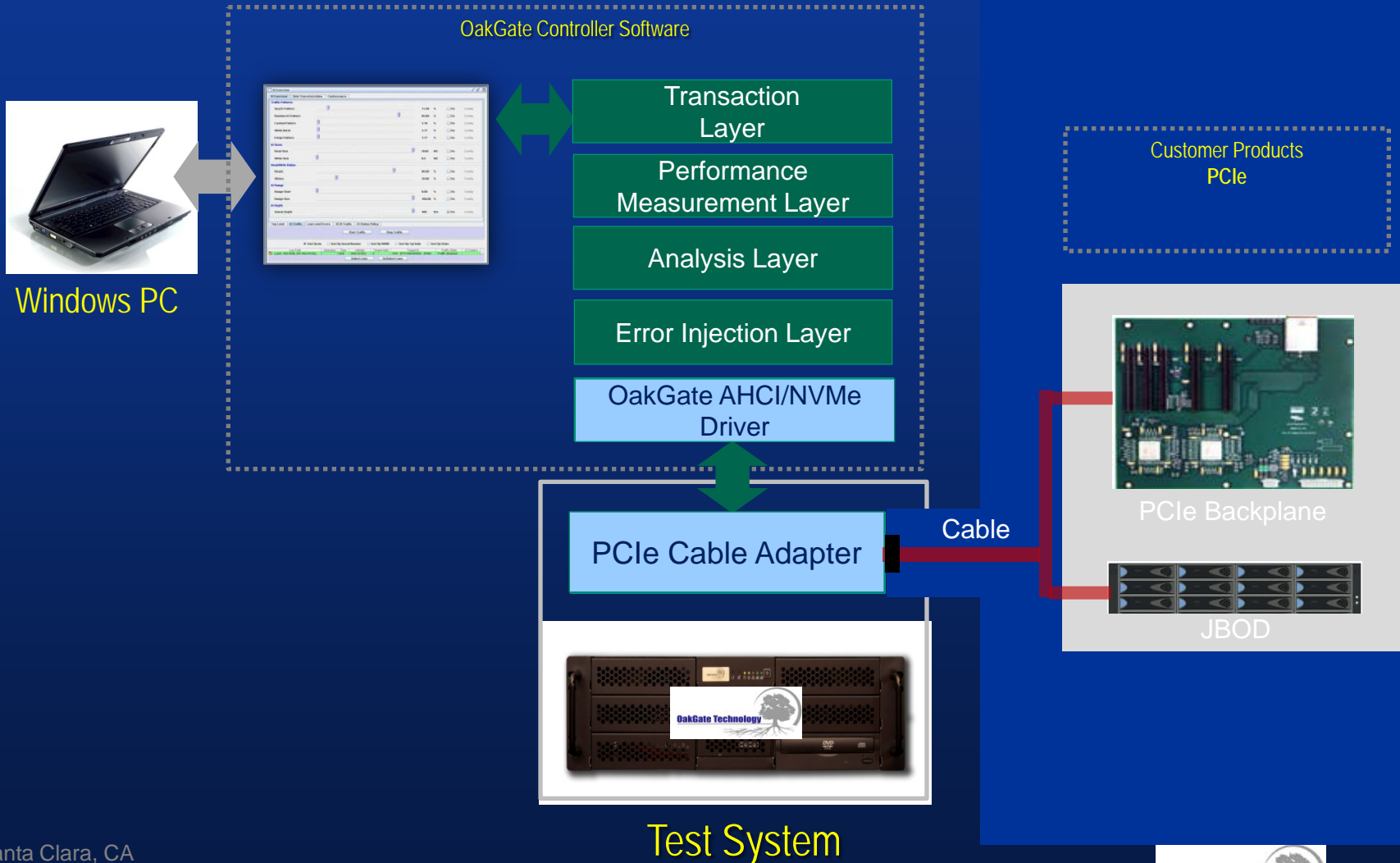
Role of Error Injection

- ◆ Focus is to Improve Product Robustness and Stability
 - ◆ Increased Code Coverage
 - ◆ Verify Error Recovery Paths
 - ◆ Exercise full breadth of Product Features
- ◆ Mixture of IO traffic with:
 - ◆ Management commands
 - ◆ Illegal commands
 - ◆ Queue creations/deletions
 - ◆ Vendor Unique commands
- ◆ Data Validation
 - ◆ Power Fail testing
 - ◆ Write Atomicity
 - ◆ Hot Spot Writes
 - ◆ Stress Wear Leveling

Directed Error Injection/Validation

IO Exerciser										
IO Exerciser Conformance Target Config Editor Global Power Control										
Enabled	Index	Type	Revision	Test Name	Excl. Lun Access	Sub-Tests	Passed Cases	Failed Cases	Status	% Complete
<input type="checkbox"/>	0	nvme	1.1.1	Create CQ Bad ID Tests	×	4	0	0	Idle	0
<input type="checkbox"/>	1	nvme	1.1.1	Create SQ Bad ID Tests	×	6	0	0	Idle	0
<input type="checkbox"/>	2	nvme	1.1.1	Create CQ Bad Contiguous Address Tests	×	3	0	0	Idle	0
<input type="checkbox"/>	3	nvme	1.1.1	Create CQ Bad Discontiguous Address Tests	×	4	0	0	Idle	0
<input type="checkbox"/>	4	nvme	1.1.1	Identify Test	×	256	0	0	Idle	0
<input type="checkbox"/>	5	nvme	1.1.1	Identify Namespace Test	×	1	0	0	Idle	0
<input type="checkbox"/>	6	nvme	1.1.1	Get Features FID Test	×	256	0	0	Idle	0
<input type="checkbox"/>	7	nvme	1.1.1	Set Volatile Write Cache 1->0 Test	×	1	0	0	Idle	0
<input type="checkbox"/>	8	nvme	1.1.1	Aggregation Threshold Test	×	1	0	0	Idle	0
<input type="checkbox"/>	9	nvme	1.1.1	Interrupt Configuration Test	×	96	0	0	Idle	0
<input type="checkbox"/>	10	nvme	1.1.1	Format Test	×	48	0	0	Idle	0
<input type="checkbox"/>	11	nvme	1.1.1	Format All Test	×	0	0	0	Idle	N/A
<input type="checkbox"/>	12	nvme	1.1.1	Format Bad Namespace Test	×	1	0	0	Idle	0
<input type="checkbox"/>	13	nvme	1.1.1	Flush Test	×	1	0	0	Idle	0
<input type="checkbox"/>	14	nvme	1.1.1	Firmware Activate Test	×	24	0	0	Idle	0
<input type="checkbox"/>	15	nvme	1.1.1	Write PI Test	×	16	0	0	Idle	0
<input type="checkbox"/>	16	nvme	1.1.1	Write Max Namespace Size Test	×	32	0	0	Idle	0
<input type="checkbox"/>	17	nvme	1.1.1	Write Max Namespace Capacity Test	×	32	0	0	Idle	0
<input type="checkbox"/>	18	nvme	1.1.1	Write Large LBA Test	×	32	0	0	Idle	0
<input type="checkbox"/>	19	nvme	1.1.1	Write Forced Unit Access Test	×	1	0	0	Idle	0
<input type="checkbox"/>	20	nvme	1.1.1	Read PI Test	×	16	0	0	Idle	0
<input type="checkbox"/>	21	nvme	1.1.1	Read Max Namespace Size Test	×	32	0	0	Idle	0
<input type="checkbox"/>	22	nvme	1.1.1	Read Max Namespace Capacity Test	×	32	0	0	Idle	0
<input type="checkbox"/>	23	nvme	1.1.1	Read Large LBA Test	×	32	0	0	Idle	0
<input type="checkbox"/>	24	nvme	1.1.1	Read Forced Unit Access Test	×	1	0	0	Idle	0
<input type="checkbox"/>	25	nvme	1.1.1	Log Page ID Test	×	256	0	0	Idle	0
<input type="checkbox"/>	26	nvme	1.1.1	Log Page SMART Namespace Test	×	256	0	0	Idle	0
<input type="checkbox"/>	27	nvme	1.1.1	Log Page SMART Global Namespace Test	×	1	0	0	Idle	0
<input type="checkbox"/>	28	nvme	1.1.1	Log Page Vendor Specific Test	×	64	0	0	Idle	0
<input type="checkbox"/>	29	nvme	1.1.1	Log Page SMART Temperature Threshold Test	×	1	0	0	Idle	0

OakGate Technology Tool Set





Thank You

Bob Weisickle
bob.weisickle@oakgatetech.com