



# NVMe<sup>™</sup> 1.4 Features and Compliance: Everything You Need to Know

Sponsored by NVM Express<sup>™</sup>, Inc.

October 2, 2019

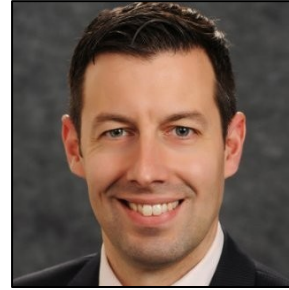


# Speakers



**Nick Adams**

**Platform Storage  
Architect**



**David Woolf**

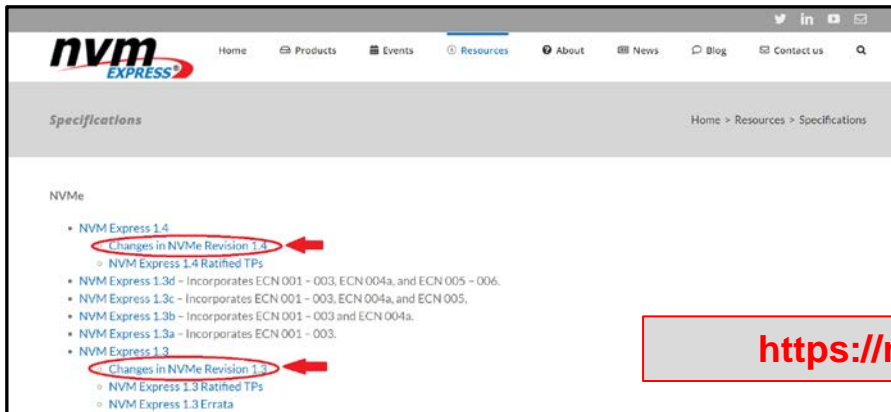
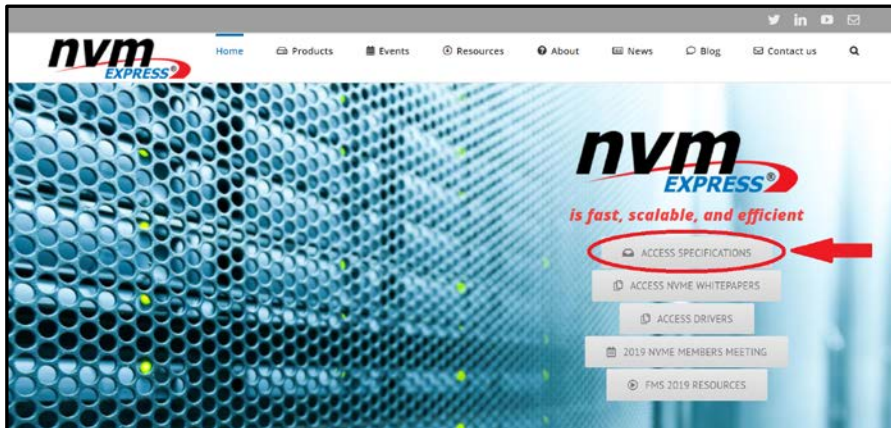
**Senior Engineer  
Data Center  
Technologies**



# Agenda

- NVMe™ Base Specification 1.4 Changes
  - Overview of New Features
  - Scope of Mandatory Changes
- Compliance Program and Tools
  - Overview and Deliverables
  - Dive into individual compliance test cases for some new NVMe features

# Where do I start?



The NVM Express™ website, of course!

- <https://nvmexpress.org>
- Spec details at link: “Access Specification”

Great resources

- Current Spec
- Current ECNs & TPs
- Historical Specs
- Detailed change documents

<https://nvmexpress.org/changes-in-nvme-revision-1-4/>

# NVMe™ 1.4 Specification

## New Feature & Enhancement Overview

# NVMe™ 1.4 Specification New Features & Enhancements\*

## *For today's overview*

- IO Determinism
- IO Performance & Endurance Hints
- Persistent Event Log
- Namespace Write Protect
- Verify Command
- Rebuild Assist

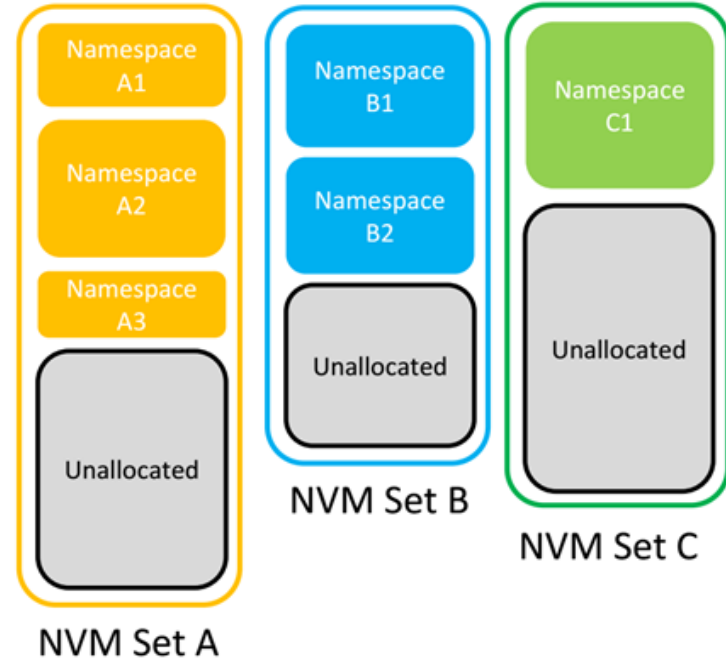
## *Additional New Features for NVMe 1.4*

- Persistent Memory Region
- Asymmetric Namespace Access
- NVM Sets
- Read Recovery Levels
- Endurance Groups
- Traffic Based Keep Alive
- UUIDs for Vendor Specific Information
- Administrative Controller
- Submission Queue Association

*Many additional features were enhanced with New Capabilities!*

# IO Determinism – NVM Sets

- NVM Sets are QoS Isolated
  - Write to namespace A1 does not impact QoS associated with namespace B2
- NVM Subsystem may support one or more NVM Sets
- One or more Namespaces may be allocated to an NVM Set



# IO Determinism – Predictable Latency Mode

Deterministic Window

Non-Deterministic Window

- Service isolation region
- Increase Read IOPs and reduce tail latency
- Provides strict QoS profile
- Significantly improves P99 and P9999 for a well-behaved host

NVM Set #1

Deterministic Window

Non-Deterministic Window

Deterministic Window

Non-Deterministic Window

NVM Set #2

Non-Deterministic Window

Deterministic Window

Non-Deterministic Window

Deterministic Window

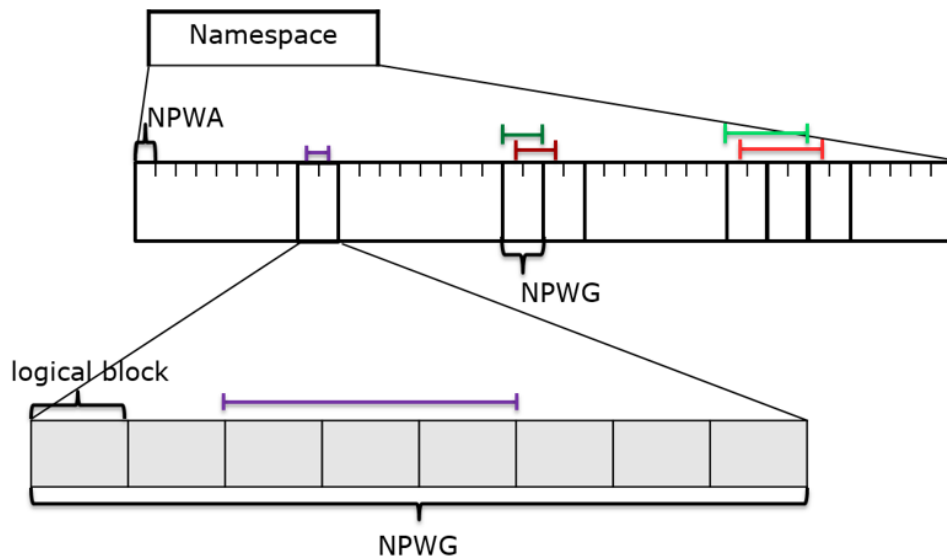




# IO Performance and Endurance Hints

Created new **mechanisms** for Hosts to **optimize** their use of NVMe™ devices



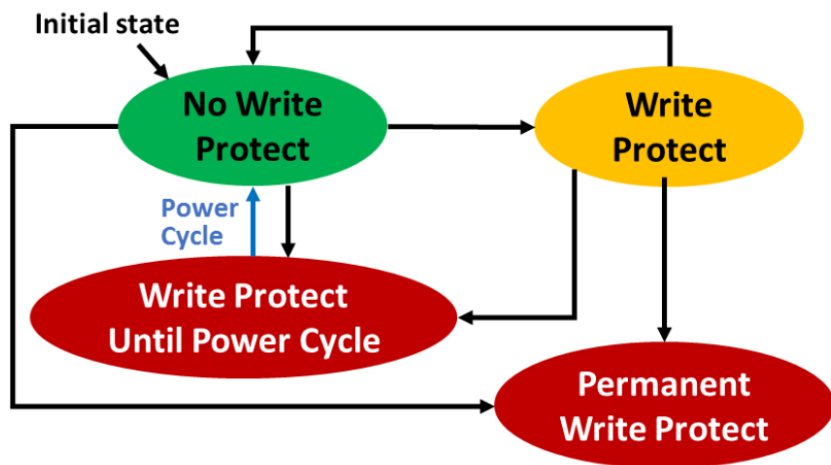
- IO Performance & Endurance Hints
  - Exposes preferred Size, Granularity and Alignment for both Write and Deallocate to the Host



Legend	
	Conformant I/O
	Non-Conformant I/O
Namespace Preferred Write Alignment (NPWA)	
Namespace Preferred Write Granularity (NPWG)	

# Namespace Write Protect

- New Feature allowing a Host to set the Write Protection Status of a Namespace
- Two supported protection states:
  - Write Protect until Power Cycle
  - Permanent Write Protect

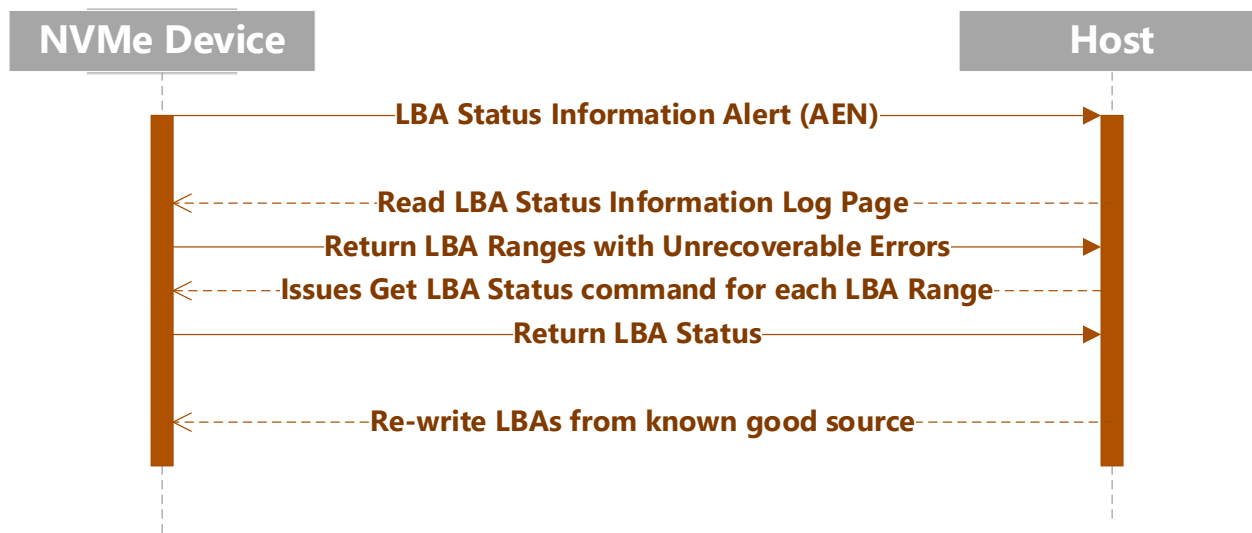


## Allowed Commands under WP

Admin Command Set	NVM Command Set
Device Self-test	Compare
Directive Send	Dataset Management
Directive Receive	Read
Get Features	Reservation Register
Get Log Page	Reservation Report
Identify	Reservation Acquire
Namespace Attachment	Reservation Release
Security Receive	Vendor Specific
Security Send	Flush
Set Features	Verify
Vendor Specific	

# Rebuild Assist

- Host is able to configure NVMe™ Device for notifications about Unrecoverable Errors
- Establishes mechanism for early communication of two types of errors:
  - ‘Tracked LBA’ list – Blocks discovered to now be bad by Device
  - ‘Untracked LBA’ list – LBA ranges associated with a component failure



# Verify Command

- New Command to check the integrity of stored data
  - Effectively acts as Read without transferring data to the Host
  - Controller reads & discards the data – while performing equivalent Protection Information checks
  - Errors are generated if data cannot be read correctly

*Drive diagnostics & data scrubbing during drive operation require integrity verification, but don't require access to the actual data.*

*Verify significantly increases the efficiency of this type of operation!*

# Enhanced Telemetry Capabilities

- The **Persistent** Event Log defines the features necessary to build a scaffolding that enables extensible debug infrastructure that is usable at scale
- Comprehensive set of events defined
  - Health Snapshot
  - Firmware Commits
  - Timestamp Changes
  - Power-on or Resets
  - Thermal Excursions
  - Vendor Specific
  - TCG-defined Events
  - Hardware Errors
  - Changed Namespace
  - Set Feature Events
  - Format NVM Start & Complete
  - Sanitize Start & Complete



***Allows SSD customers to get consistent debug capabilities across vendors!***

***Allows SSD vendors an extensible framework for custom debug content!***

# NVMe™ 1.4 Specification Required, Incompatible Changes

# NVMe™ 1.4 Specification Required Changes\*

- New NSID value usages
- New errors and reporting requirements
- Temperature threshold clarifications
- Controller Memory Buffer & Persistent Memory Region Enhancements
- New Sanitize requirements
- Reservation Notification Log usage
- Clarified LBA Range feature behavior
- Reservation Report command conflicts resolved
- New Abort command behavior

\* Not to scale. These are *categories* of changes, not the full list of changes themselves

# Example: Mandatory Change Controller Memory Buffer (CMB)

## Overview

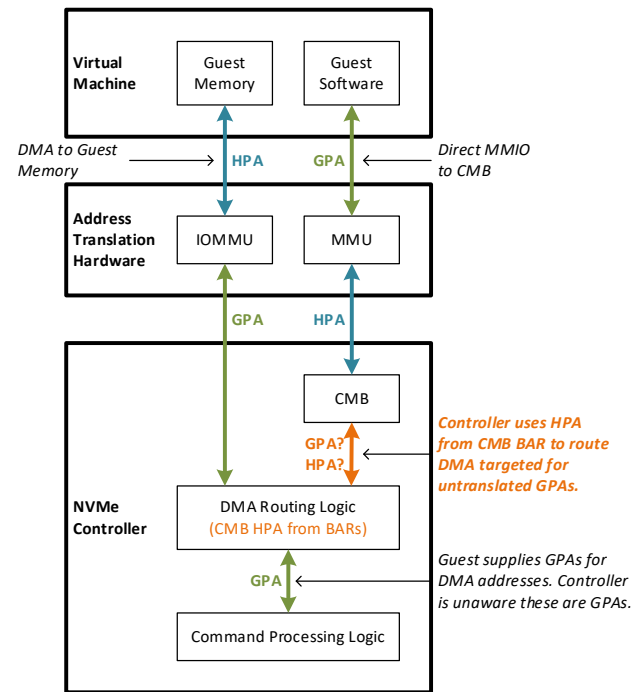
- Controller Memory Buffer now requires Support (CMBS) and Enable (CRE) bit usage
- Removed restrictions on the usages of the CMB – SQ, CQ & Data

## Why the changes?

- Requires explicit configuration of the feature by the driver
- Hardens the Controller Memory Buffer implementation
- Relaxes the restrictions on host usage of the CMB

## Impacts of inaction...

- Leaves the potential for DMA misrouting with CMB implementations



## References

NVMe™ revision 1.4 section 3.1, 4.7, 4.8 & 7.3  
Technical Proposal 4054



# Example: Mandatory Change NSID value - FFFFFFFFh

## Overview – Namespace Identifiers

- All usages of NSID value FFFFFFFF are now well-defined
- Generally used to mean a broadcast action against all Namespaces

## What are the changes?

- Clarifications in many sections: I/O Commands, Set/Get Features, Admin Commands, and Reservations
- Explicitly defines when NSID of FFFFFFFF can be used and how to use it

## Why the changes?

- The specification was quiet on a number of use cases
- Need to provide consistency across Device and OS implementations
- Improve the end-user experience and ease of NVMe™ device consumption

## Impacts of inaction

- Inconsistent results when using devices from various hardware vendors

# Compliance

# Compliance Program Overview



## Coverage

- NVMe™ Base Spec, NVMe-MI™, NVMe-oF™

## Timeline

- Test Specifications usually lag specification by 1-2 quarters. Test Specifications are updated twice a year and try to address any ratified TPs and ECNs since previous update.



## Mandatory vs. FYI

- New tests are introduced as FYI. After the test implementation is vetted, it can be transitioned to being Mandatory. Test spec and tool call out Mandatory vs. FYI tests.



## Optional Features

- Optional features are skipped if not supported. (You've don't have to do it, but if you do it, you have to do it right). Tests check for feature support first.



## Test Tools

- Tests available through UNH-IOL test tools. Tools can be run in-house to check compliance on an ongoing basis. SSD vendors, Controller Vendors, Integrators, IP Houses, Datacenter companies regularly run these tools (some weekly and nightly) to ensure continued compliance.



# Compliance Program Deliverables

Test Specs

Test Tools

Plugfests and Private Testing

Integrators List

A screenshot of the 'NVM™ Integrator's List' website. The page title is 'NVM™ Integrator's List'. It features a search bar and a table of integrators. The table has columns for 'Product', 'Product Type', 'Firmware Version', 'Integrator Program Revision', 'Date Listed', and 'Further Info'.

Product	Product Type	Firmware Version	Integrator Program Revision	Date Listed	Further Info
Intel Optane™ P5800	SSD	1.00.01.0	1.00.01.0	08/01/2016	Intel Optane™ P5800
Intel Optane™ P5600	SSD	1.00.01.0	1.00.01.0	08/01/2016	Intel Optane™ P5600
Intel Optane™ P5400	SSD	1.00.01.0	1.00.01.0	08/01/2016	Intel Optane™ P5400
Intel Optane™ P5200	SSD	1.00.01.0	1.00.01.0	08/01/2016	Intel Optane™ P5200
Intel Optane™ P5000	SSD	1.00.01.0	1.00.01.0	08/01/2016	Intel Optane™ P5000
Intel Optane™ P4800	SSD	1.00.01.0	1.00.01.0	08/01/2016	Intel Optane™ P4800
Intel Optane™ P4600	SSD	1.00.01.0	1.00.01.0	08/01/2016	Intel Optane™ P4600
Intel Optane™ P4400	SSD	1.00.01.0	1.00.01.0	08/01/2016	Intel Optane™ P4400
Intel Optane™ P4200	SSD	1.00.01.0	1.00.01.0	08/01/2016	Intel Optane™ P4200
Intel Optane™ P4000	SSD	1.00.01.0	1.00.01.0	08/01/2016	Intel Optane™ P4000
Intel Optane™ P3800	SSD	1.00.01.0	1.00.01.0	08/01/2016	Intel Optane™ P3800
Intel Optane™ P3600	SSD	1.00.01.0	1.00.01.0	08/01/2016	Intel Optane™ P3600
Intel Optane™ P3400	SSD	1.00.01.0	1.00.01.0	08/01/2016	Intel Optane™ P3400
Intel Optane™ P3200	SSD	1.00.01.0	1.00.01.0	08/01/2016	Intel Optane™ P3200
Intel Optane™ P3000	SSD	1.00.01.0	1.00.01.0	08/01/2016	Intel Optane™ P3000
Intel Optane™ P2800	SSD	1.00.01.0	1.00.01.0	08/01/2016	Intel Optane™ P2800
Intel Optane™ P2600	SSD	1.00.01.0	1.00.01.0	08/01/2016	Intel Optane™ P2600
Intel Optane™ P2400	SSD	1.00.01.0	1.00.01.0	08/01/2016	Intel Optane™ P2400
Intel Optane™ P2200	SSD	1.00.01.0	1.00.01.0	08/01/2016	Intel Optane™ P2200
Intel Optane™ P2000	SSD	1.00.01.0	1.00.01.0	08/01/2016	Intel Optane™ P2000
Intel Optane™ P1800	SSD	1.00.01.0	1.00.01.0	08/01/2016	Intel Optane™ P1800
Intel Optane™ P1600	SSD	1.00.01.0	1.00.01.0	08/01/2016	Intel Optane™ P1600
Intel Optane™ P1400	SSD	1.00.01.0	1.00.01.0	08/01/2016	Intel Optane™ P1400
Intel Optane™ P1200	SSD	1.00.01.0	1.00.01.0	08/01/2016	Intel Optane™ P1200
Intel Optane™ P1000	SSD	1.00.01.0	1.00.01.0	08/01/2016	Intel Optane™ P1000

# Compliance Test Cases for NVMe™ 1.4 Specification

- IO Determinism
- Namespace Write Protect
- Persistent Event Log
- Verify Command

# IO Determinism

Test 3.6 Case 1 : Predictable Latency Mode Supported (FYI, OF-FYI)



Testing Station  
NVMe™ Host

NVMe  
Controller

Identify CNS=01h

Identify Controller Data Structure  
CTRATT Bit 5 Predictable Latency Mode = 0

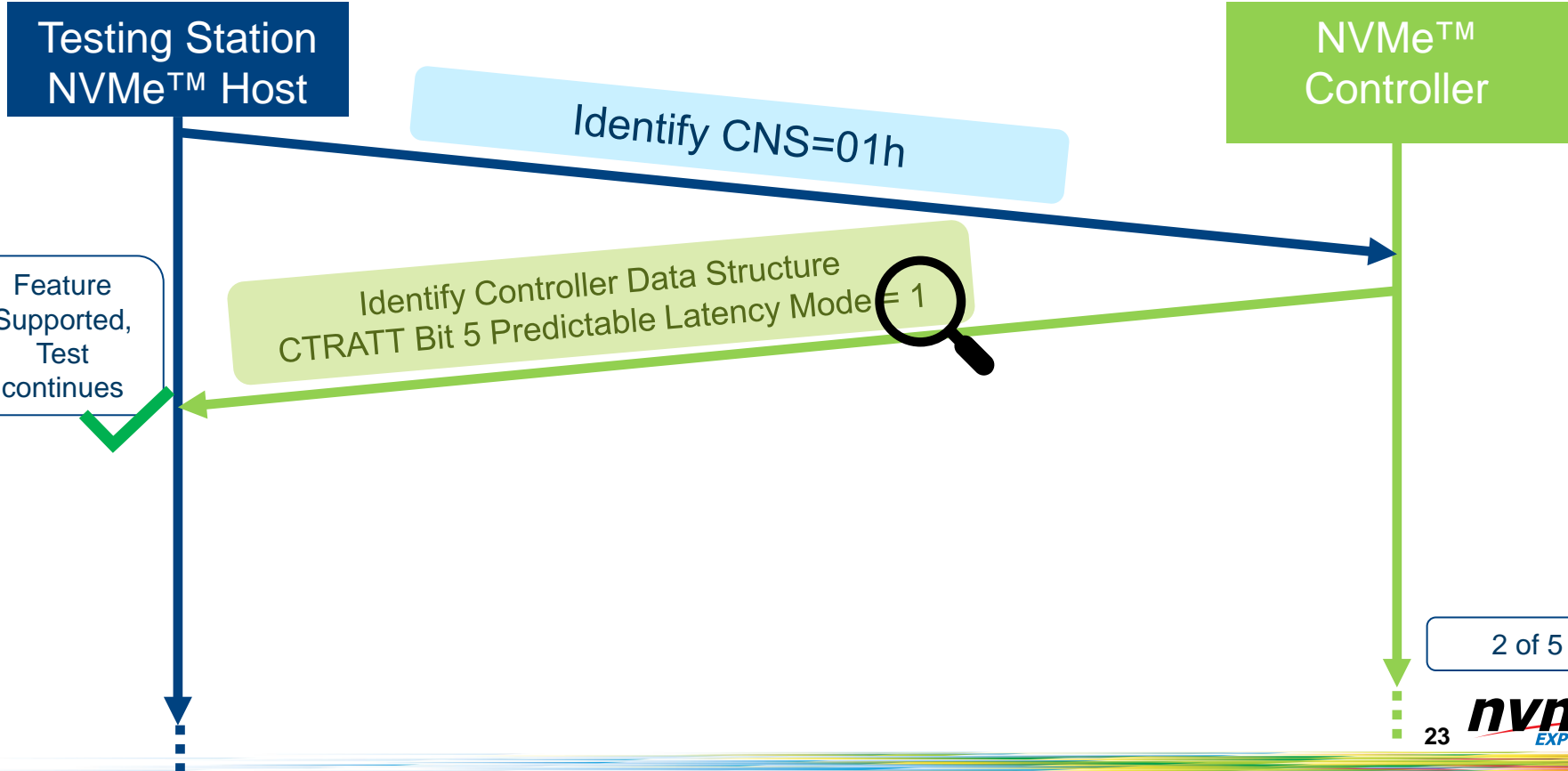
Feature not  
Supported;  
Test not  
applicable



1 of 5

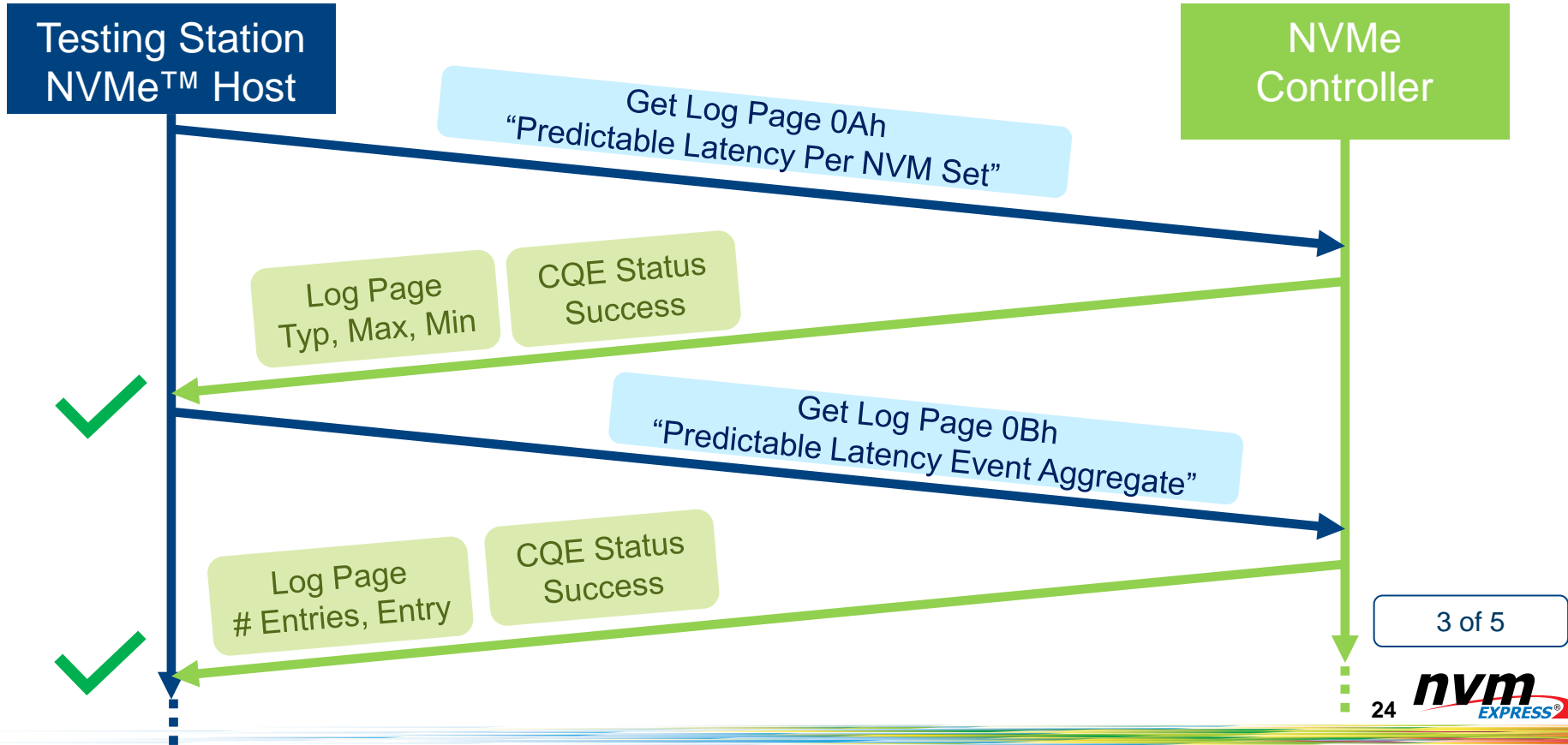
# IO Determinism

Test 3.6 Case 1 : Predictable Latency Mode Supported (FYI, OF-FYI)



# IO Determinism

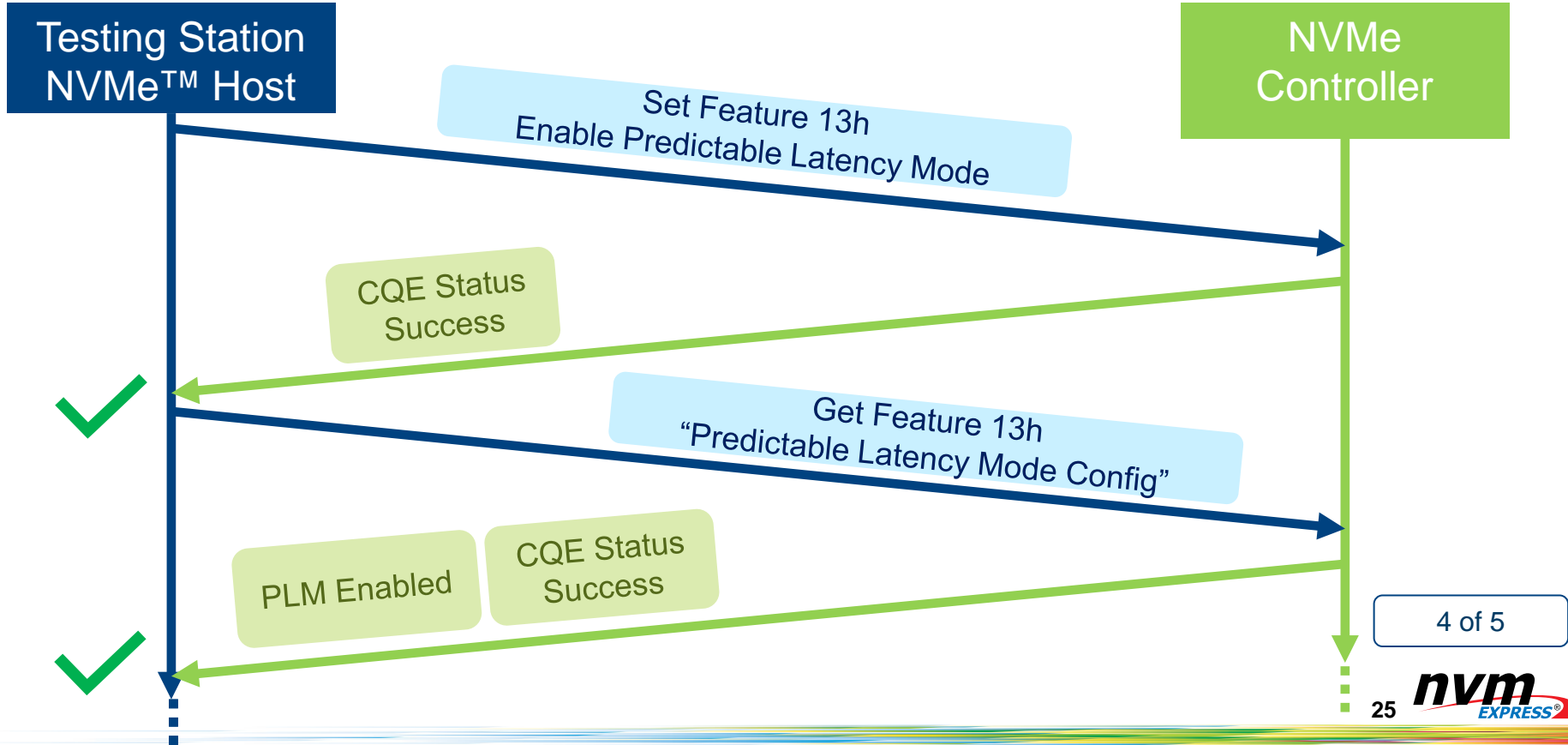
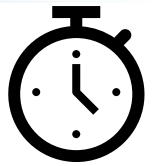
Test 3.6 Case 1 : Predictable Latency Mode Supported (FYI, OF-FYI)





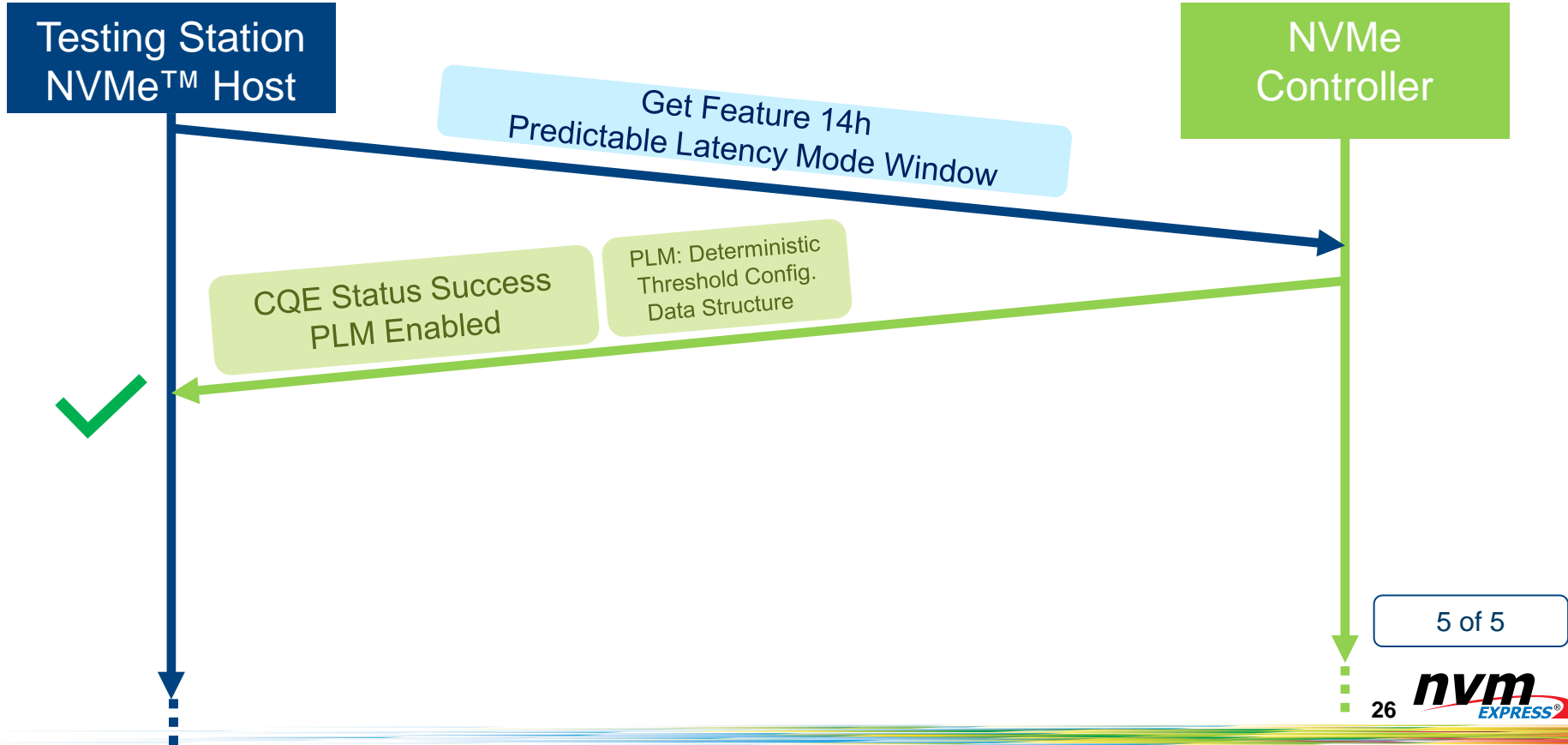
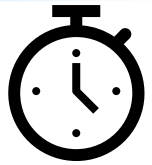
# IO Determinism

Test 3.6 Case 1 : Predictable Latency Mode Supported (FYI, OF-FYI)

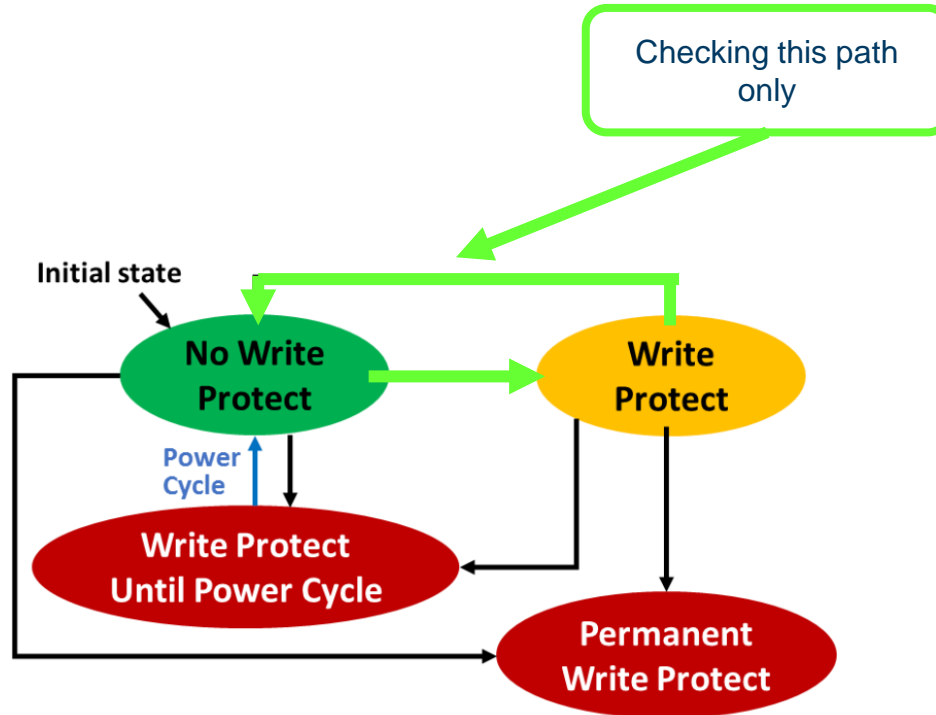


# IO Determinism

Test 3.6 Case 1 : Predictable Latency Mode Supported (FYI, OF-FYI)



# Namespace Write Protect



# Namespace Write Protect



Test 3.7 Case 1 : Enable and Disable Write Protection (FYI, OF-FYI)

Testing Station  
NVMe™ Host

NVMe  
Controller

Identify Controller Data  
Structure CNS = 01h

Identify Controller Data Structure  
NWPC = 000b

Feature not  
Supported;  
Test not  
applicable

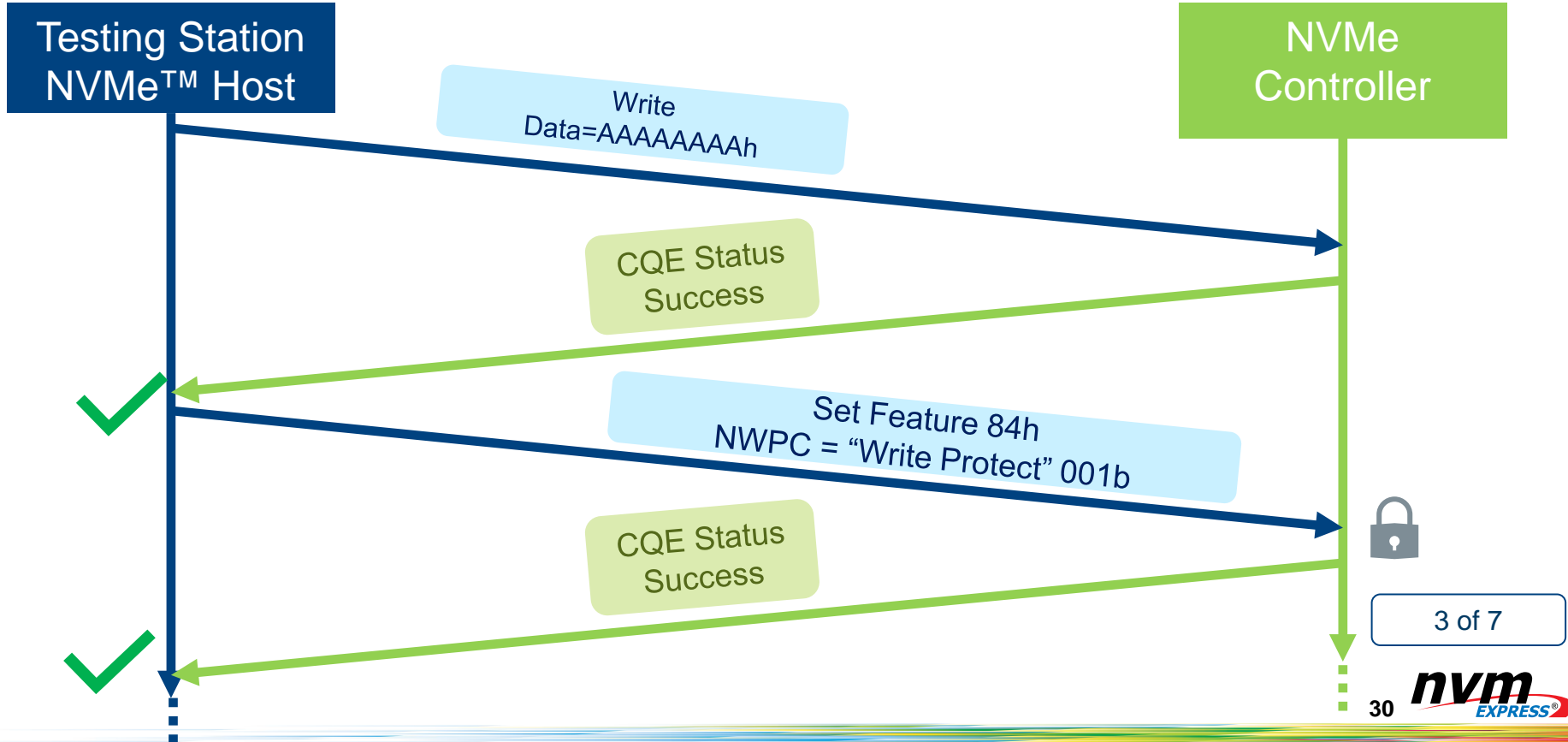


1 of 7



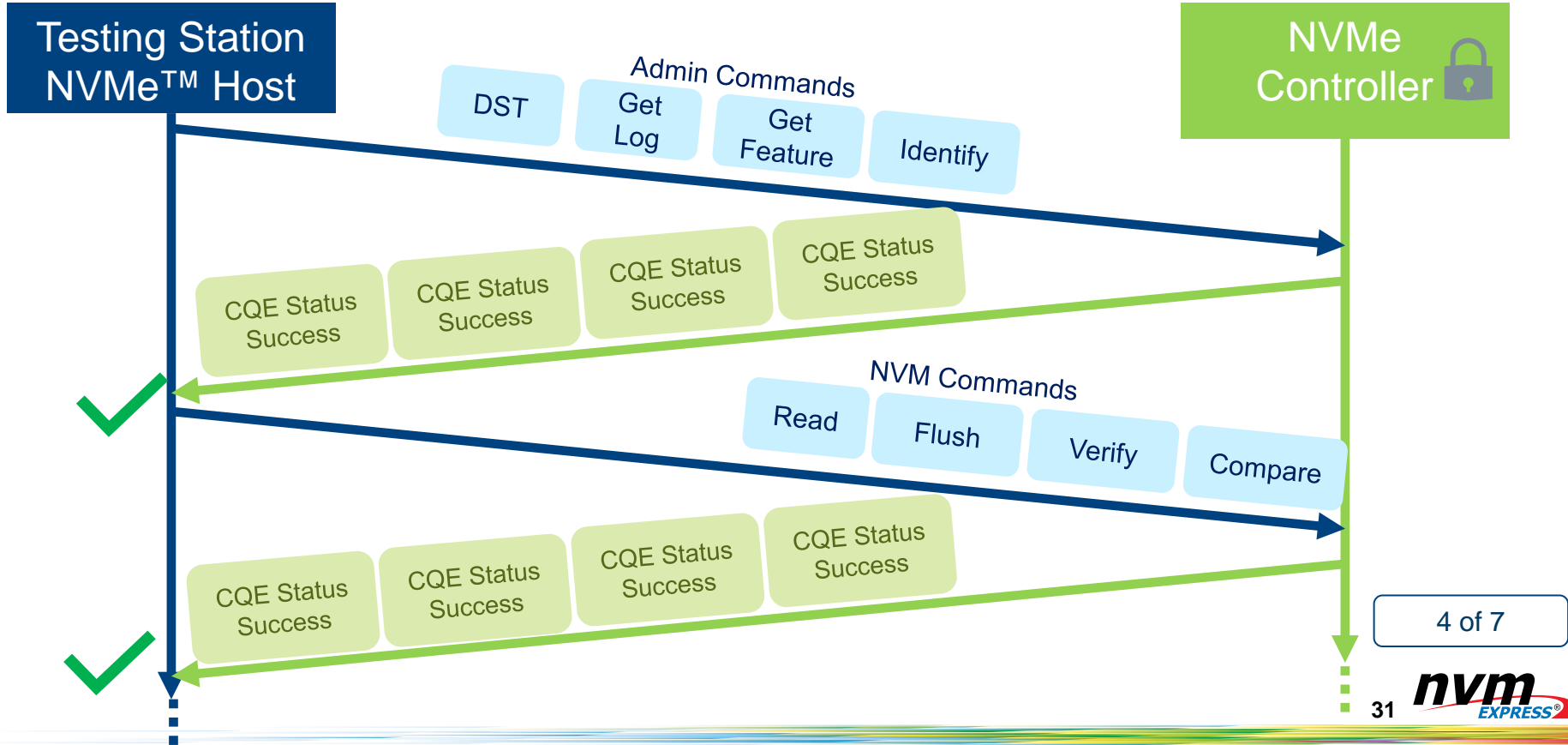
# Namespace Write Protect

Test 3.7 Case 1 : Enable and Disable Write Protection (FYI, OF-FYI)



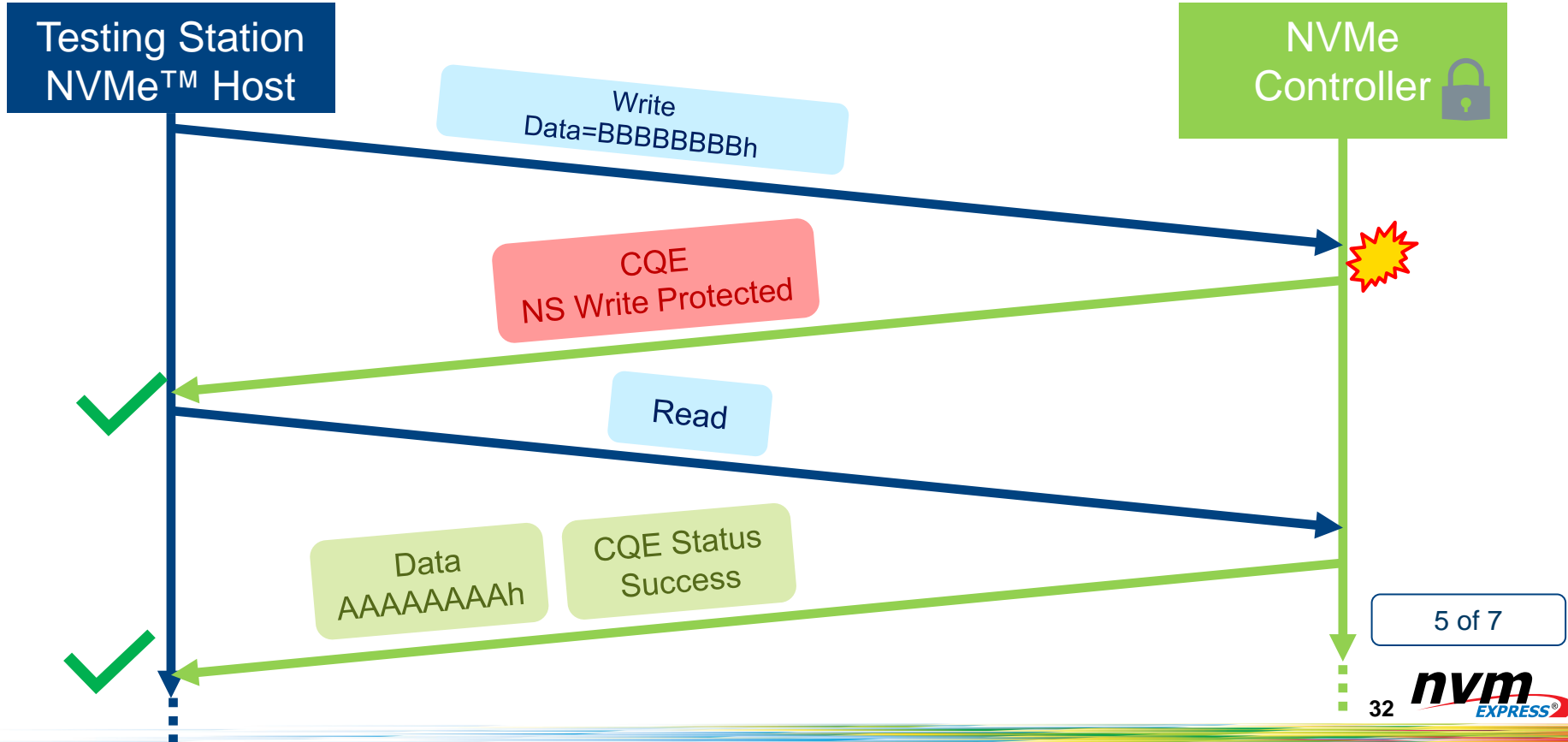
# Namespace Write Protect

Test 3.7 Case 1 : Enable and Disable Write Protection (FYI, OF-FYI)



# Namespace Write Protect

Test 3.7 Case 1 : Enable and Disable Write Protection (FYI, OF-FYI)





# Namespace Write Protect

Test 3.7 Case 1 : Enable and Disable Write Protection (FYI, OF-FYI)



Testing Station  
NVMe™ Host

NVMe  
Controller

Set Feature 84h  
NWPC = "No Write Protect" 000b

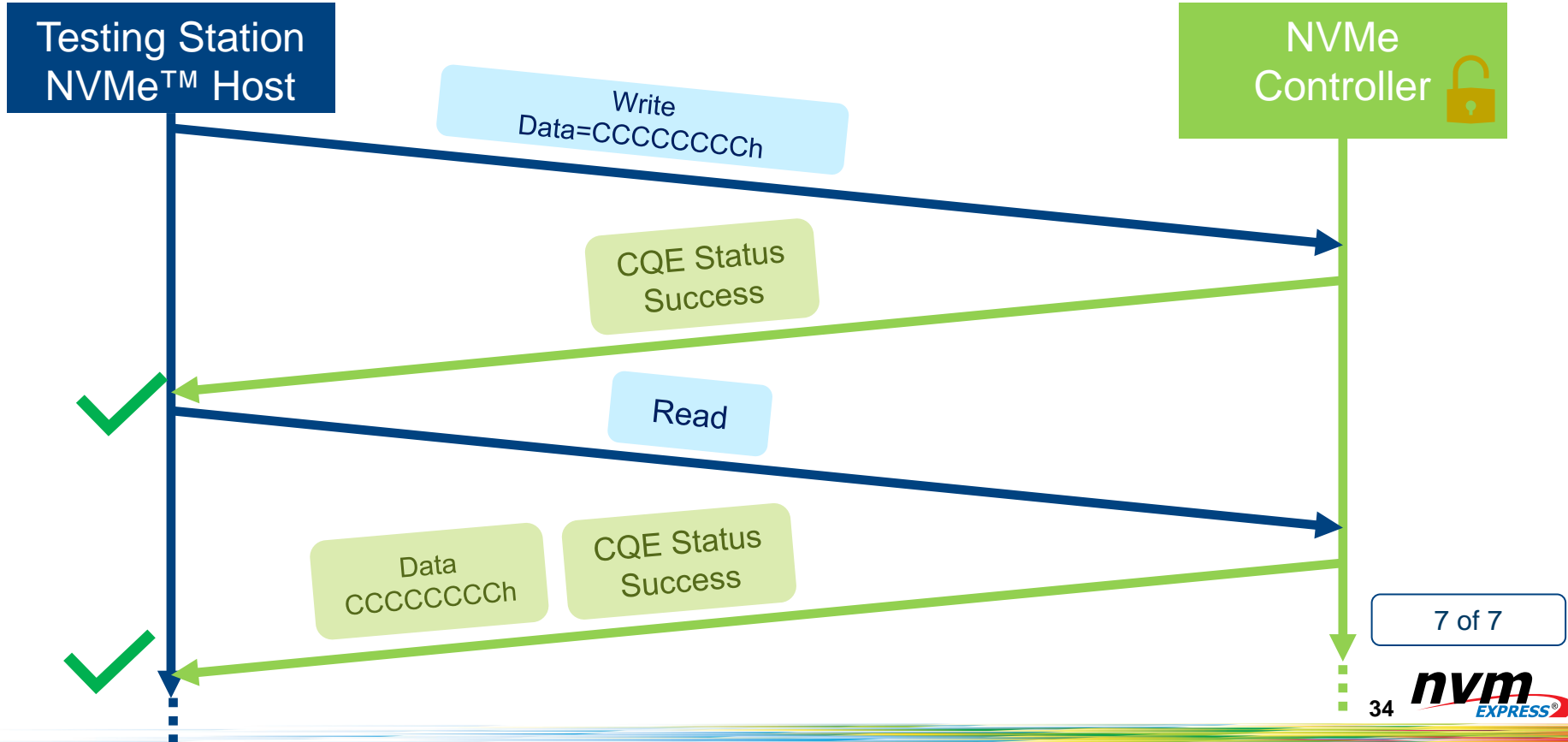
CQE Status  
Success



6 of 7

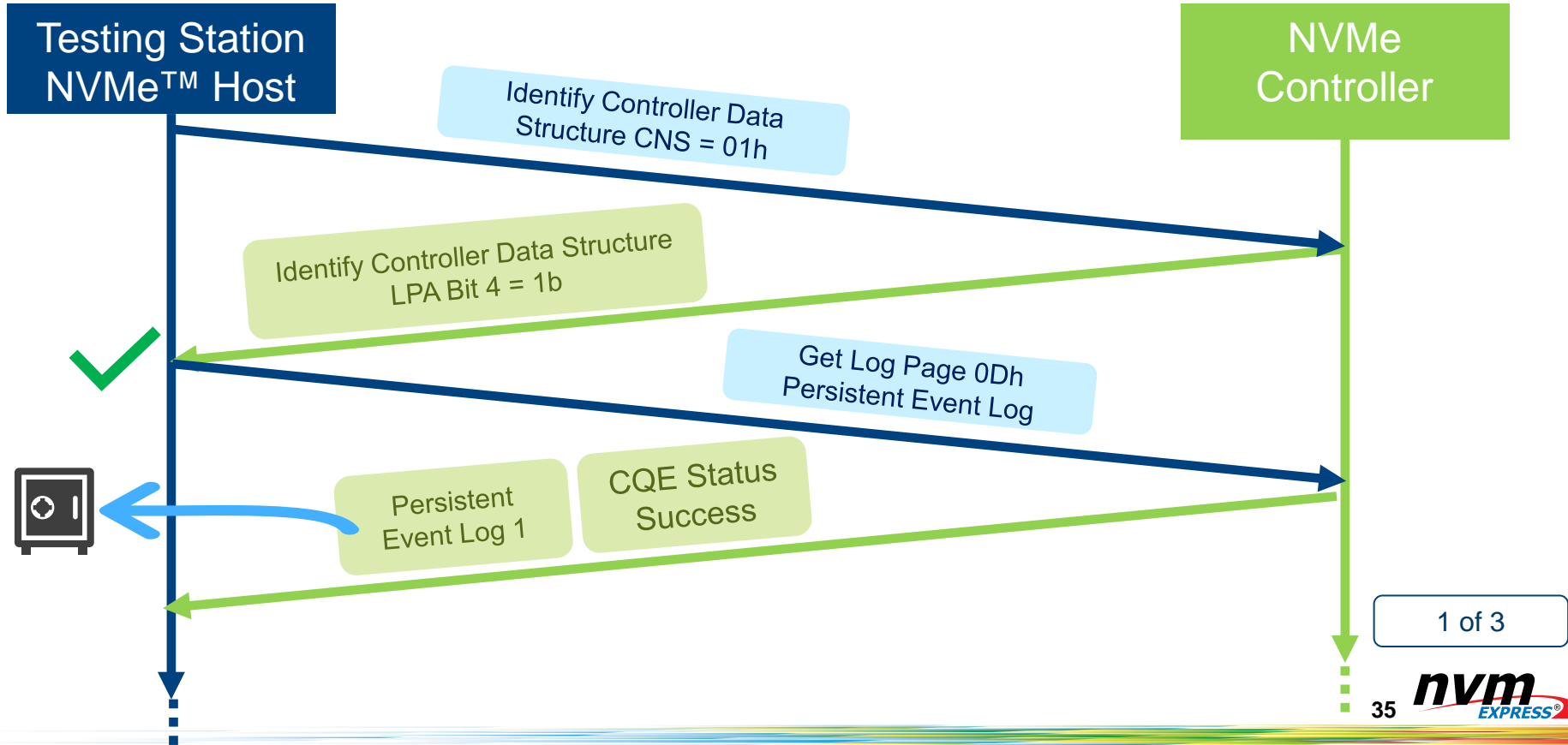
# Namespace Write Protect

Test 3.7 Case 1 : Enable and Disable Write Protection (FYI, OF-FYI)



# Persistent Event Log

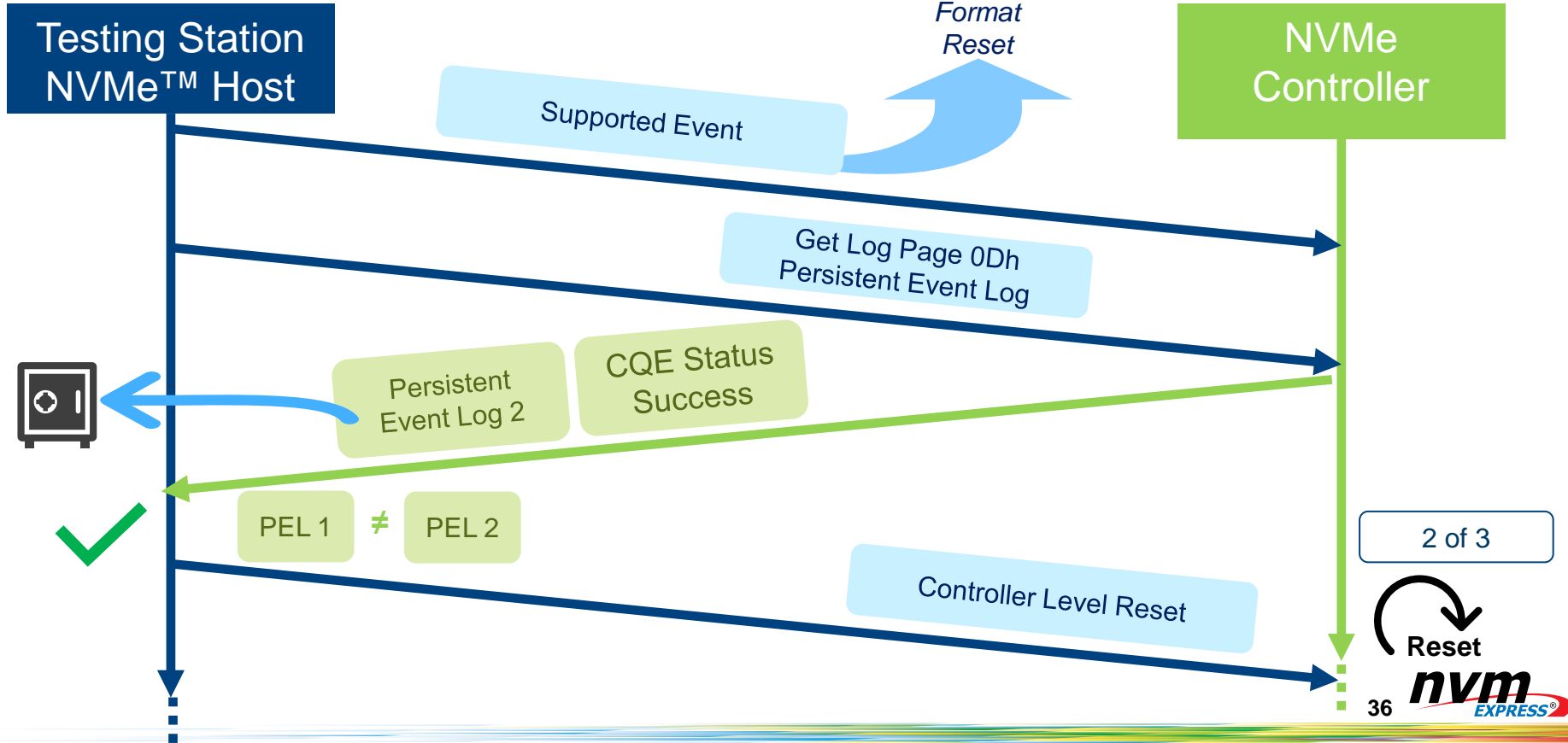
Test 1.3 Case 18 Persistent Event Log (FYI, OF-FYI)



1 of 3

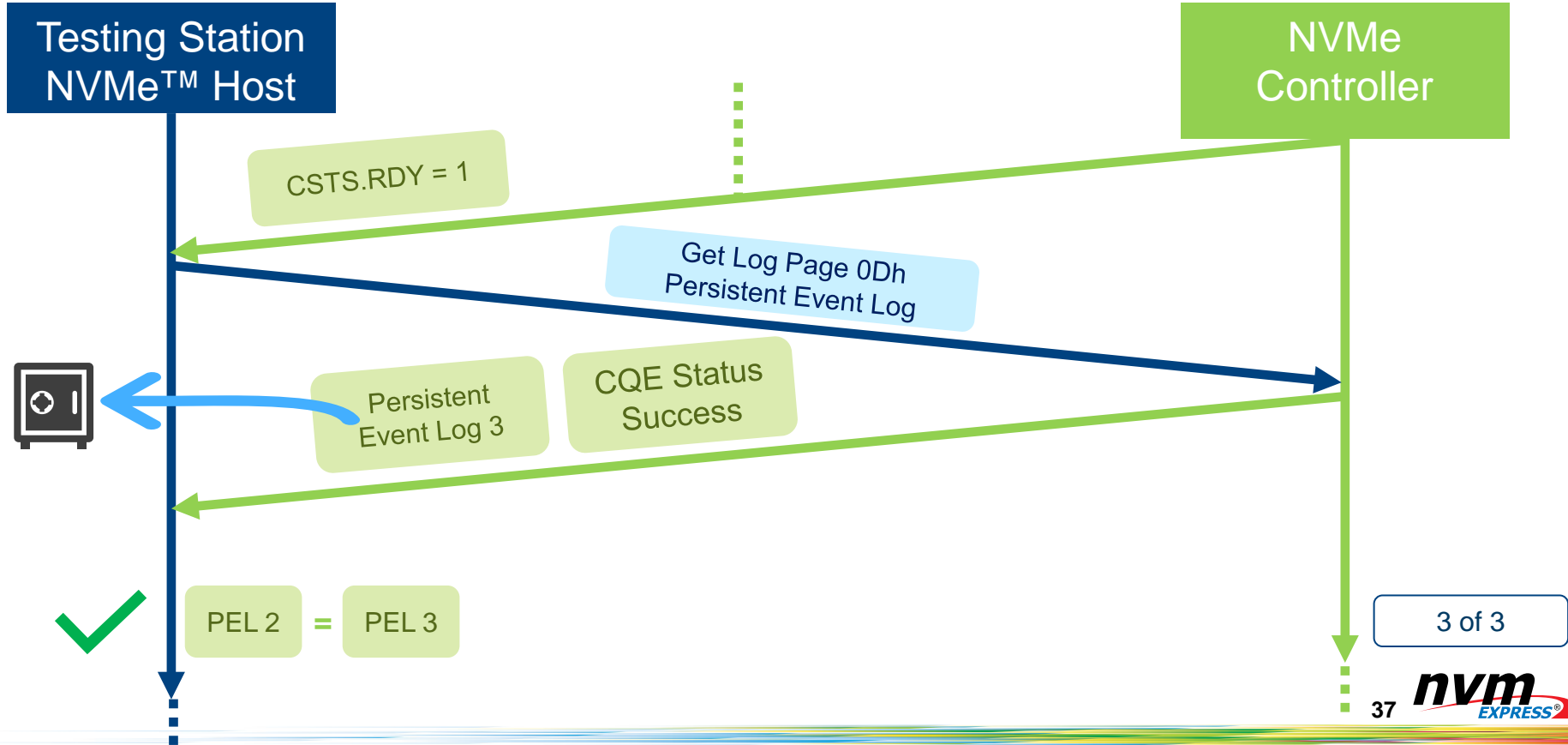
# Persistent Event Log

Test 1.3 Case 18 Persistent Event Log (FYI, OF-FYI)



# Persistent Event Log

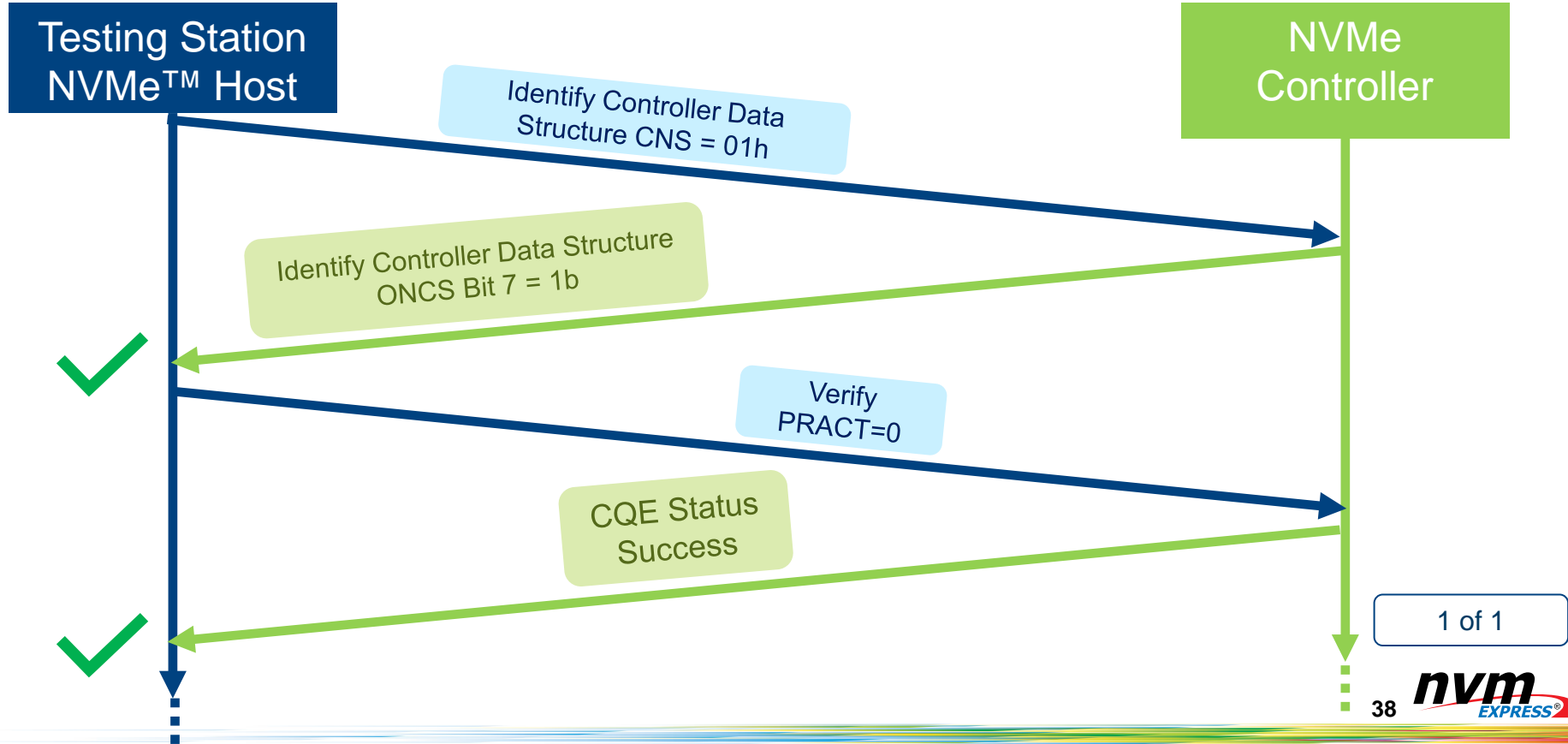
Test 1.3 Case 18 Persistent Event Log (FYI, OF-FYI)





# Verify Command Basic Operation

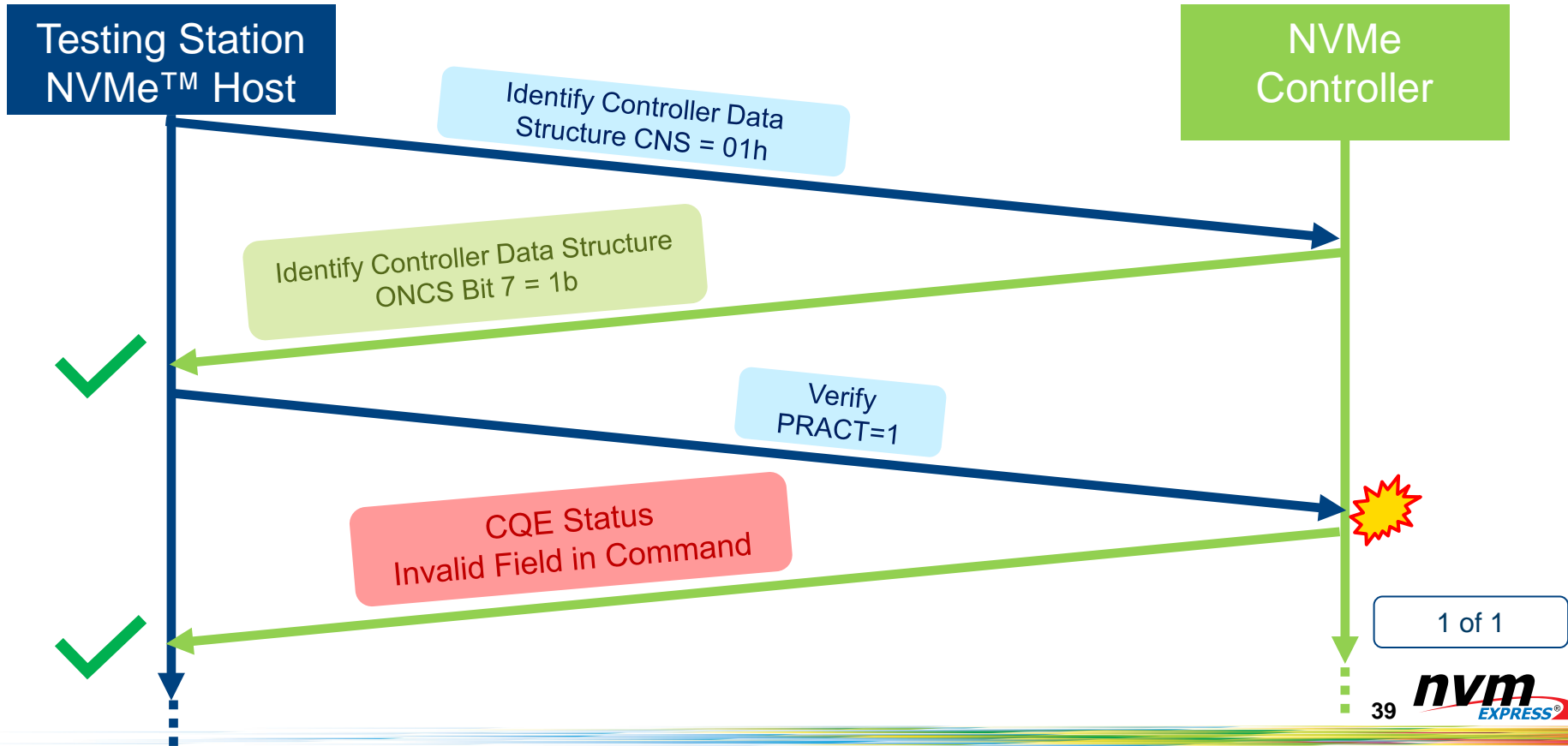
Test 2.11 Case 1 Valid Command (FYI, OF-FYI)





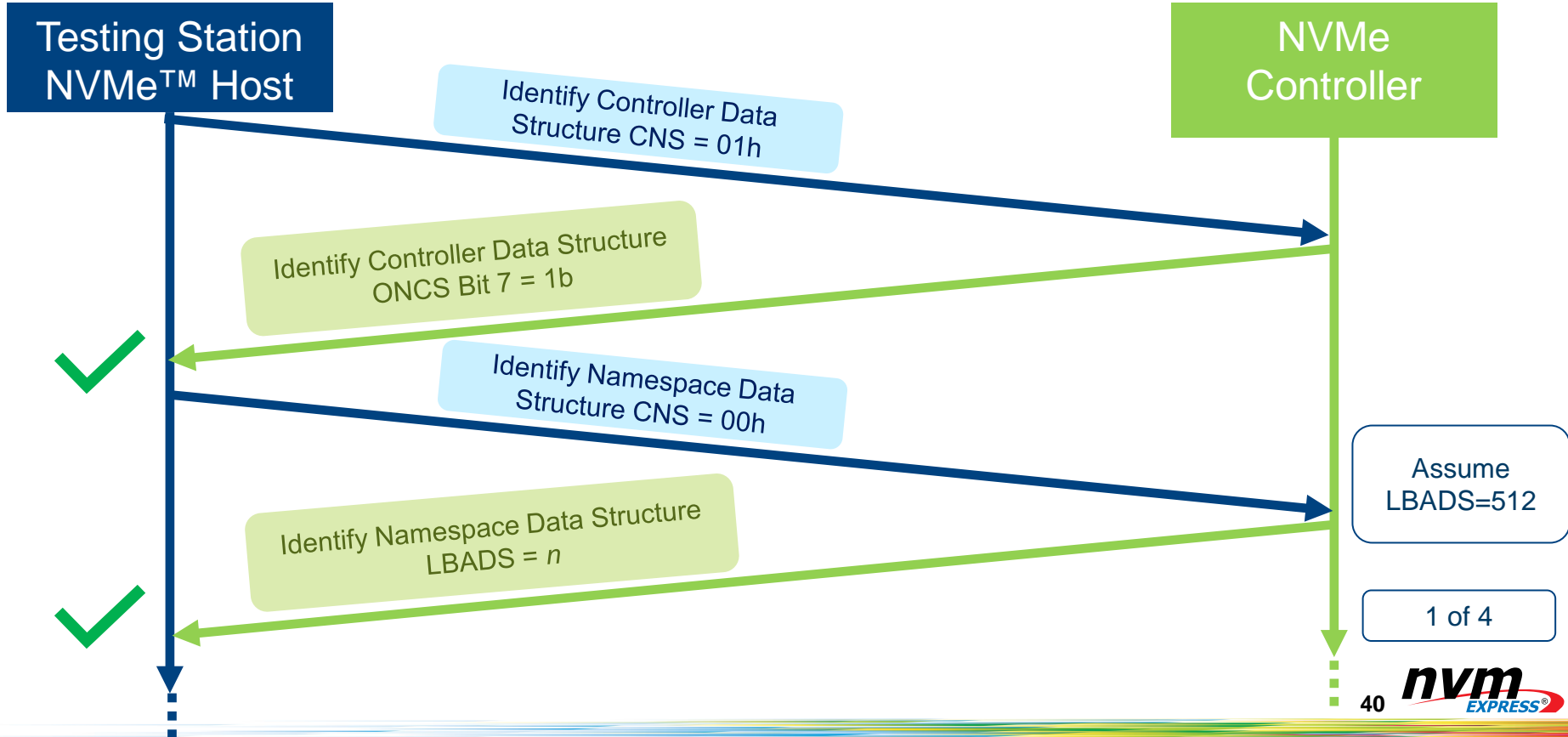
# Verify Command PRACT=1

Test 2.11 Case 2 PRACT=1 (FYI, OF-FYI)



# Verify Command – SMART/Health Log

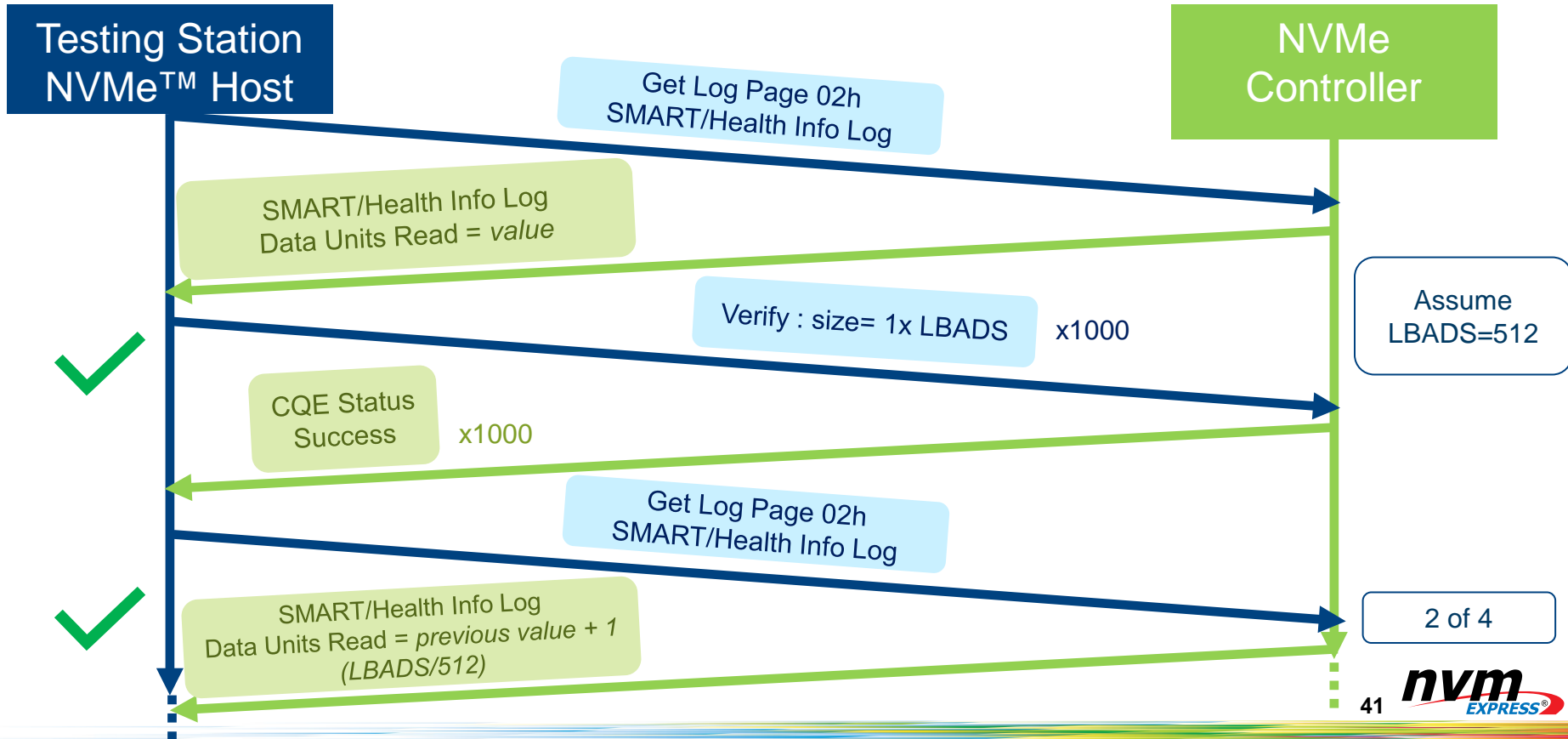
Test 1.3 Case 19 Data Units Read Count – Verify (FYI, OF-FYI)





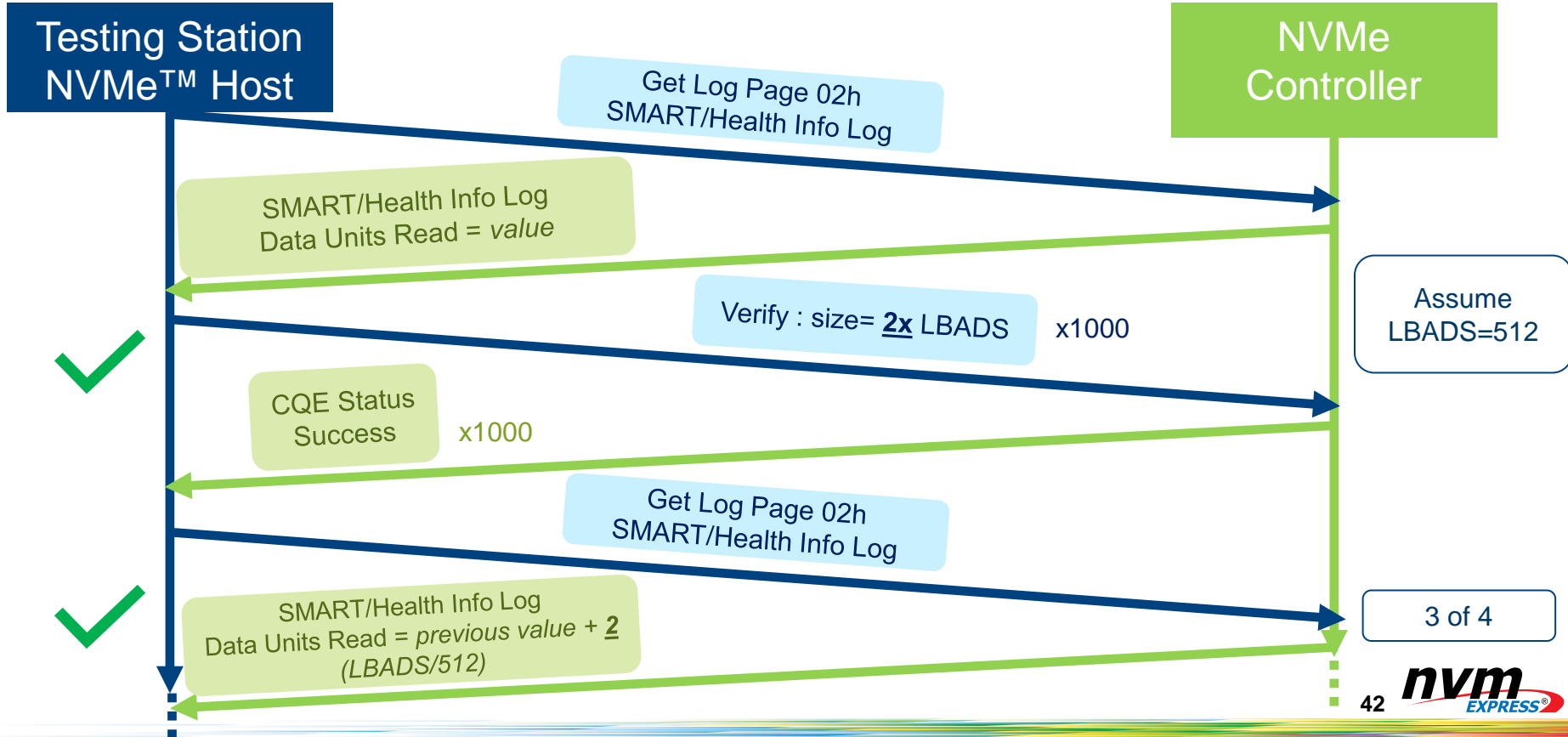
# Verify Command – SMART/Health Log

Test 1.3 Case 19 Data Units Read Count – Verify (FYI, OF-FYI)



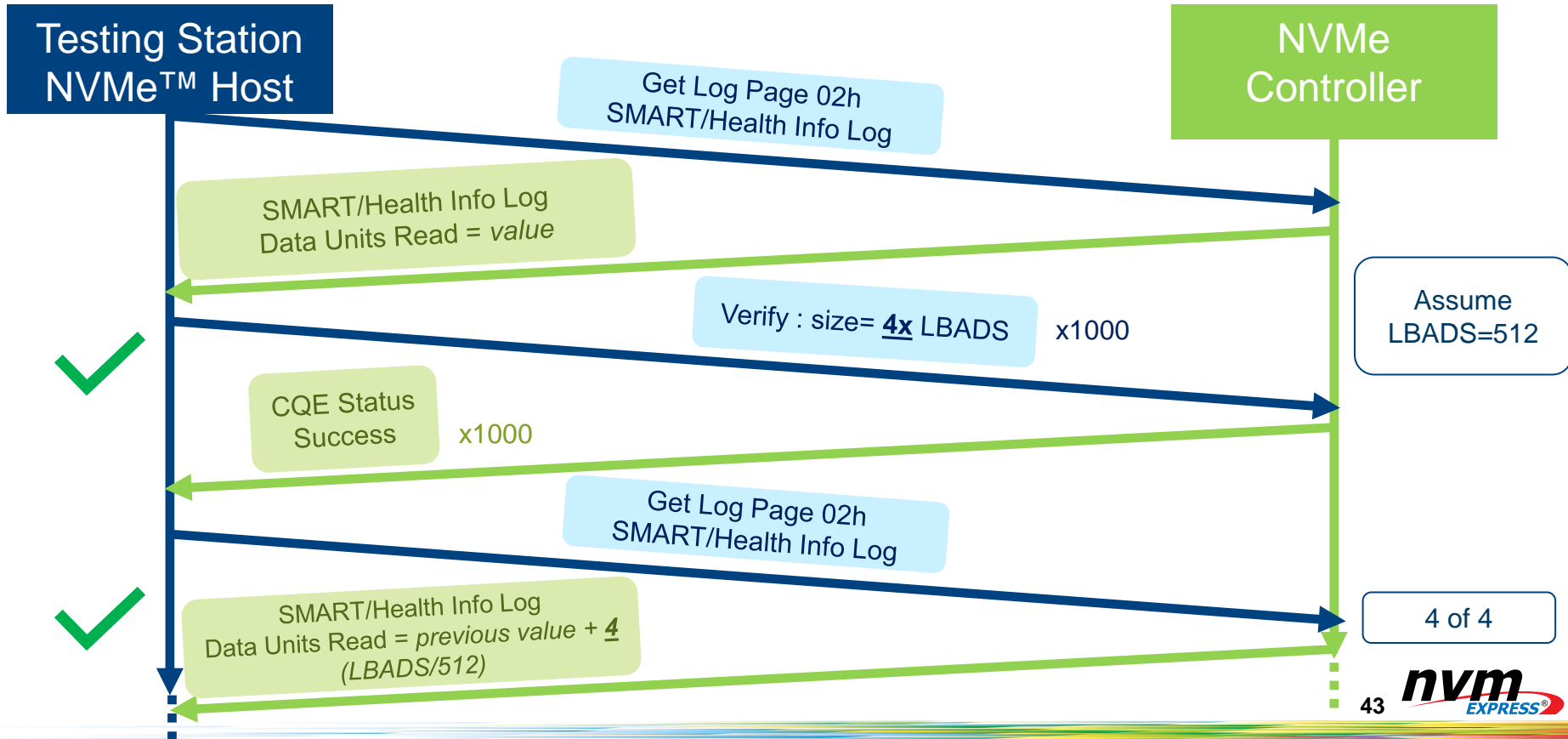
# Verify Command – SMART/Health Log

Test 1.3 Case 19 Data Units Read Count – Verify (FYI, OF-FYI)



# Verify Command – SMART/Health Log

Test 1.3 Case 19 Data Units Read Count – Verify (FYI, OF-FYI)



# Compliance for NVMe™ 1.4 Specification

- Test Tools currently supporting initial set of NVMe 1.4 features
- Download and run tools to prove compliance
- Feedback welcome

# Q&A

