Pre-Migration Verification for NVMe® SSDs: Ensuring Seamless Live Migration

Sponsored by NVM Express organization, the owner of the NVMe® Specifications

Speakers

Prashant Dixit

SIEMENS

Agenda

- Live Migration: The Next Leap in Storage Flexibility
  - What's Driving the Shift to Live Migration?
  - Host Managed Live Migration Model
  - Legacy vs. Live Migration Model (Strongly Suggested)
  - What's New?
- Functional Verification
  - Challenges
  - Solution

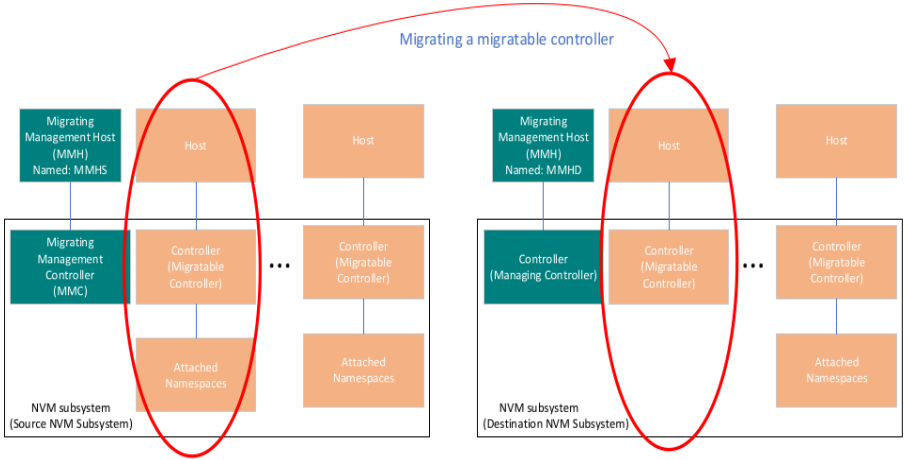
Live Migration: The Next Leap in Storage Flexibility

What's Driving the Shift to Live Migration?

- There is a rising demand for 24/7 uptime due to continuous AI model training and data processing
- Frequent performance disruptions during hardware or software maintenance
- Limited flexibility in moving workloads across systems
- Live Migration addresses these challenges by:
  - Supporting real-time data mobility across infrastructure
  - Maintaining high availability and improving resource utilization

Host Managed Live Migration Model

An Migration Management Controller (MMC) is an I/O controller or an Administrative controller that supports the Host Managed Live Migration capability and provides the ability for the Migration Management Host (MMH) to use privileged actions

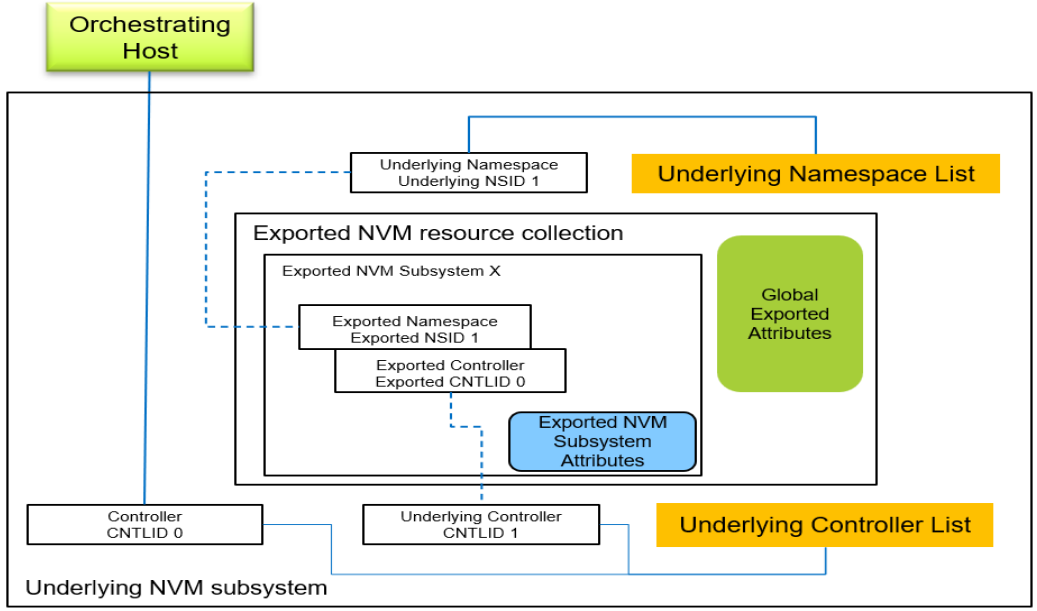


Benefits of Host Managed Live Migration

- Downtime is significantly reduced or eliminated
- Workloads run seamlessly during migration

What's New?

- PCIe® technology exported NVM subsystem Migration
  - Enables VM migration by hiding changes in underlying NVM subsystems, so VMs see consistent storage entities during and after migration



Functional Verification – Challenges & Solution

Pre-Migration Verification Scenarios

- Transfer of Data
  - Ensuring that the user data (namespace changes) and host memory changes are successfully transferred from source NVM subsystem to destination NVM subsystem
- Suspension of Migratable Controller
  - Verifying whether MC when has stopped processing commands when suspended by the MMH
- Configuration of Queues
  - Before the MC controller is resumed at the destination NVM subsystem, it is verified whether the IO queues are successfully configured according to controller state data structure

Pre-Migration Verification Scenarios

- Setting the Controller State
  - Migration Send command with set controller state operation should only be done when the Migratable Controller (MC) for which controller state is being set is suspended
- Checks Before Resuming MC
  - Before resuming the controller, it is checked whether the controller state for that controller has been successfully verified and committed
- Tracking User Data Changes
  - When Track Send command with log user data changes operation is sent, existence of Controller Data Queue (CDQ) and the MC associated with Controller Data Queue Identifier is checked whether it is suspended
- Effects of Controller Level Reset
  - CDQ are deleted, host memory changes are not tracked further, and controller is removed from suspended state

End-to-End Testing of the Migration Flow

- Full Test Coverage
  - Entire Live Migration flow is validated by sending all the related live migration commands in sequence
  - Ensures system behaves as expected during and after migration
- Status Monitoring
  - Successful status codes are checked for all commands in the migration flow
  - Any errors encountered are logged and reported for analysis

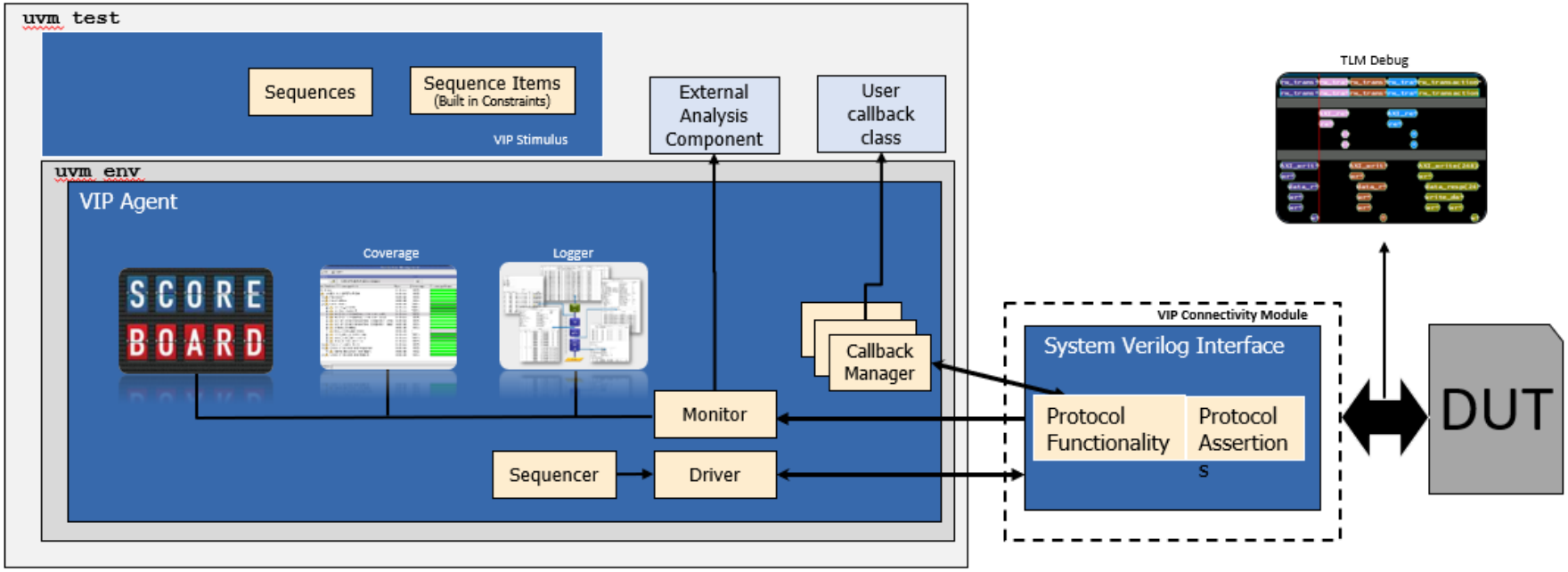
End-to-End Testing of the Migration Flow

- Queue Configuration
  - Queues are created after controller state is set at the destination NVM subsystem
  - Head and tail doorbells for queues are configured
- Seamless NVMe® Command Flow
  - NVMe commands are issued post-migration
  - Validates seamless transition with no loss of functionality

Protocol Compliance

- Verifying that Live Migration commands do not introduce protocol violations or unexpected behavior
- Embedded Monitor
  - Decodes all transport packets
  - Watches complete address space
  - Checks for any unnecessary/ unrelated transport packets
  - Shadow NVM storage models inside Host Software Bus Functional Model (BFM) for data score boarding

Protocol Compliance – Embedded Monitor



Protocol Compliance – Protocol Suite

- Exhaustive Protocol Suite
  - 1800+ checklist items built into BFM and Test Suite
  - 100+ checklist items for HMLMS command set
  - Checklist derived based on spec and UNH test plan

```
NVM21_5_1_4_1_1n7    If the memory specified by this field is not host memory (e.g., CMB or PMR), then the controller shall abort the command with a status code of Invalid Field in Command.
NVM21_5_1_16_1_2n1   Migration Receive: The specified controller is not in a condition to set the controller state.
NVM21_5_1_26_1_1n2   More To Report (MTR): If the NTMCD field is cleared to 0h, then this bit shall be cleared to 0.
NVM21_5_1_17n1       Migration Send : Command Dword 10, 3h to FFh is reserved to select field.
NVM21_5_1_17_1_2n1   If the specified controller is not suspended, then the controller shall abort the command with a status code of Controller Not Suspended
NVM21_5_1_17_1_3n11  Figure 358: The Version field of Controller State Data Structure shall be cleared to 0h.
```

Stimuli / Testing

- Directed Testing – Creating exhaustive test plans
- Stress Testing – Assessing system behavior under high-load conditions
- Handling concurrent operations between live migration tasks and standard NVMe® technology operations

Stimuli / Testing – Compliance Suite

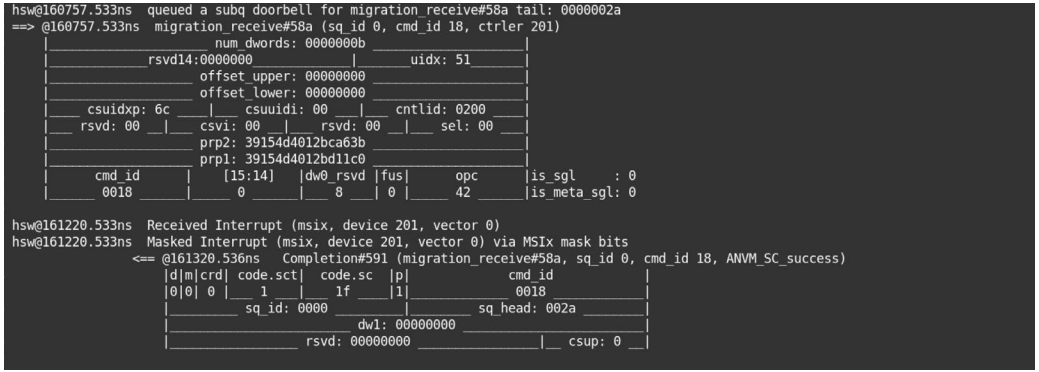
- Transport Independent Stimulus Library
  - 600+ Off-the-shelf compliance tests
- Highly Configurable Command Structure
  - Specification defined fields are directly accessible
- Randomization of Stimulus
  - Corner cases and unexpected scenarios
- Automated Command Creation
  - Constraints, APIs
  - Minimized user input for stress-testing
- Error Injection
  - Can be easily achieve through callbacks and APIs

```
anvmt_sgl_last_seg_dspt.sv
anvmt_sgl_null_data_dspt.sv
anvmt_sgl_seg_dspt.sv
anvmt_sgl_seg_err.sv
anvmt_sgl_use_bit_bucket.sv
anvmt_storage_tag_check.sv
anvmt_subsystem_reset.sv
anvmt_subsystem_shutdown.sv
anvmt_thermal_mng.sv
anvmt_timestamp.sv
anvmt_update_phase.sv
anvmt_virtualization_ctrler_reset.sv
anvmt_virtualization_func_level_reset.sv
anvmt_virtualization_vf_enable.sv
anvmt_zone_append_seq_wr.sv
```

Stimuli / Testing – Transaction Modes

➢ Transaction Mode

- Blocking and Non-Blocking
- Simultaneous or sequential simulation of Live Migration commands along with NVM, Zoned Namespace (ZNS) and Key Value (KV) commands
- Verification IP (VIP) auto memory management for easy of usage
- VIP auto schedules parallel commands among different Submission Q and different controllers



Coverage

- Validate that Device Under Test (DUT) functions correctly under all possible scenarios – functional coverage, code coverage and checklist coverage
- Comprehensive Coverage Plan
  - All fields of Host Managed Live Migration Support Admin, I/O commands
  - Crosses with possible status code types
  - Each cover point has a corresponding test in compliance test suite
- Verification IQ
  - Reduced coverage closure time – hole analysis, heatmaps, bin distribution
  - Debugging tool – failure signature detection

End-to-End Test Suite for Live Migration

- **Pre-existing UNH-IOL Live Migration Testcases**
  - Comprehensive coverage of all checks and validations for various live migration scenarios.
  - Ensures robust testing across different live migration workflows and edge cases.
- **Flexible Test Execution using VIQ Testsuite Configurator**

  Users can choose to execute:
  - All live migration testcases for exhaustive coverage.
  - Specific testcases targeting particular live migration commands as per focused testing needs.

Transaction Recording

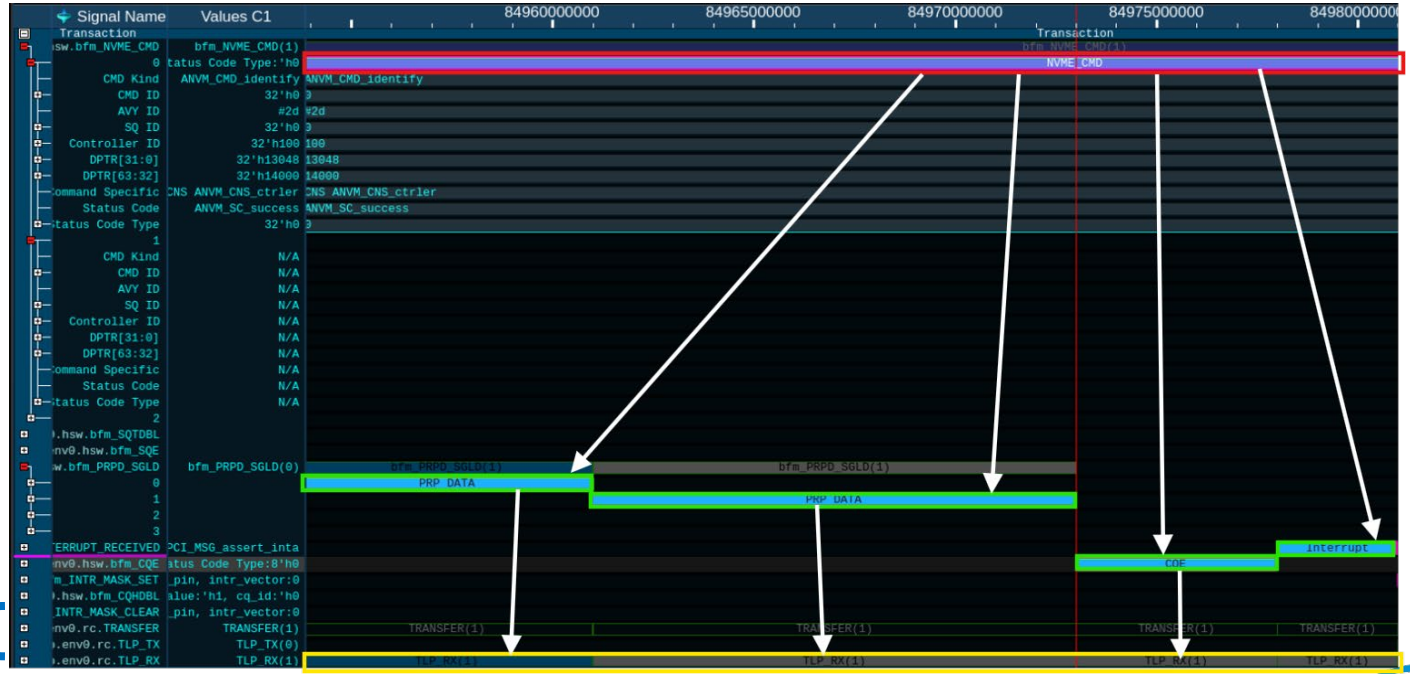
- Comprehensive Command Coverage
  - Each NVMe transaction in the waveform captures the full lifecycle of an NVMe command—from initiation to completion, ending when the interrupt mask is cleared
- NVMe Transactions Mapping
  - Parent NVMe transactions are linked to related child transactions such as PRP/SGL data transfers, interrupts, and completion queue entries (CQEs)
- PCIe Correlation
  - All NVMe transactions are traceable and can be mapped to their underlying PCIe transactions for deeper protocol analysis

Transaction Recording

Parent Transaction

Child Transactions
(SQE, PRP/SGL Data, CQE, Interrupt)

Underlying transport packet
(PCIe® TLPs)



Questions?

Visit us at Siemens EDA booth #


nvm
EXPRESS®
Architected for Performance