Multi-Root Stateless VF and PF Migration April 2006

Douglas Freimuth (IBM)

Contents

- **Migration Specification Goal and Usage Scenarios**
- Review of Single-Root and Multi-Root Terms
- Multi-Root Stateless VF Migration
- Multi-Root Stateless PF Migration

Migration Specification Goal

- *Define algorithms for basic mechanisms*
  - ✓ *Migrate a virtual function (VF) from a source VF to a destination VF*
  - ✓ *Migrate a physical function (PF) from a source PF to a destination PF*
  - ✓ *Multi-Root topology*
  - ✓ *Stateless*
- Leverage basic IOV functions
  - ✓ SR & MR PCIM
  - ✓ Hot-plug
  - ✓ Function Level Reset
  - ✓ VF and PF creation
  - ✓ Minimize any migration-specific specification.
- Document usage models, algorithms, etc. as implementation notes
- Discuss requirements and problem statement
  - ✓ Underlying IOV specification isn't complete
  - ✓ Create methods when necessary components of the IOV spec is agreed upon

Usage Scenarios

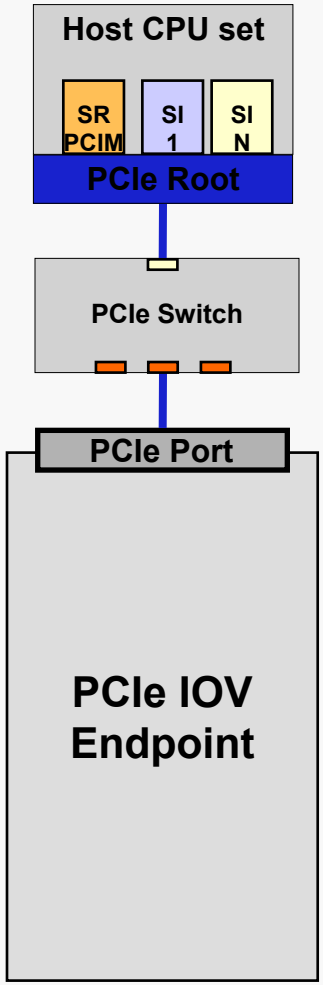
- System Image migration (including I/O)
- Load balancing
  - ✓ System A is overloaded move to system B
  - ✓ Optimization boundary (e.g. performance, utilization)
- Maintenance of hardware
  - ✓ System A needs maintenance; move for maintenance to system B
- Power and cooling management
  - ✓ Move work around for optimal power/cooling usage
- Maintenance of OS
  - ✓ Debug system
  - ✓ System cloning
  - ✓ Move to system with upgrade (e.g. firmware)

Contents

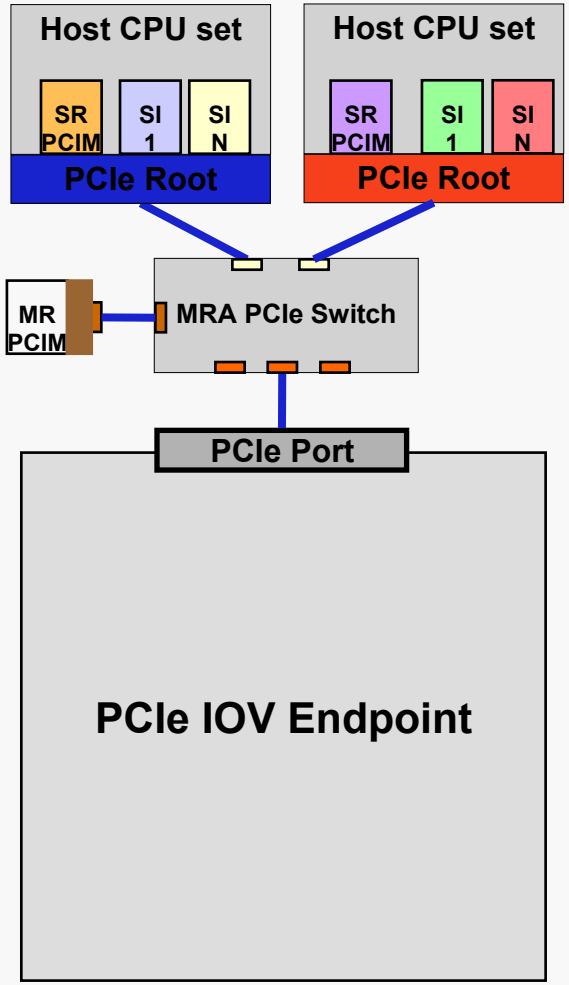
- Migration Specification Goal and Usage Scenarios
- **Review of Single-Root and Multi-Root Terms**
- Multi-Root Stateless VF Migration
- Multi-Root Stateless PF Migration

Topology Overview and Terms

SR Topology



Multi-Root Topology



Terms

Single Root (SR) IOV Overview

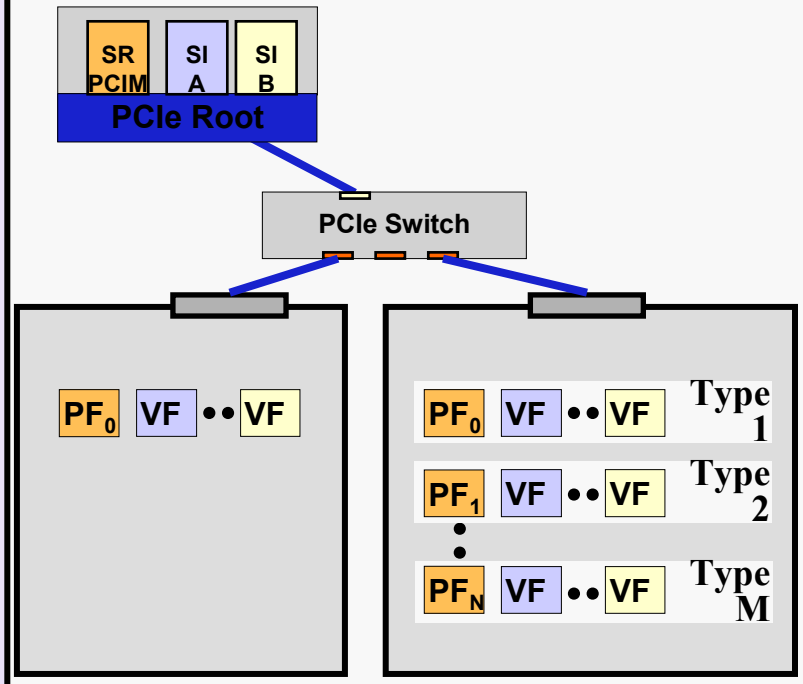
- Only has one Root.
- Switches only need to support PCIe base functionality.
- To make full use of IOV, EP must support SR-IOV capabilities.
- SR-PCIM configures the EP.

Multi-Root (MR) IOV Overview,

- One or more Roots.
- Switches with Multi-Root Aware (MRA) functionality are needed.
- To make full use of IOV, EP must support SR & MR-IOV capabilities.
- MR-PCIM assigns Virtual Endpoints (VEs) to RCs and manages PCIe components.
- SR-PCIM configures its VEs.

Single-Root IOV Function Types and Terms

SR Topology



Terms

For SR topologies,

Physical Function is used by SR-PCIM to manage a set of Virtual Functions.

Physical Function 0 (PF_0) is also used to manage EP functions, such as physical errors and events.

Virtual Function is used by SIs to access resources on the EP.

More details on function assignment, capabilities, etc… later.

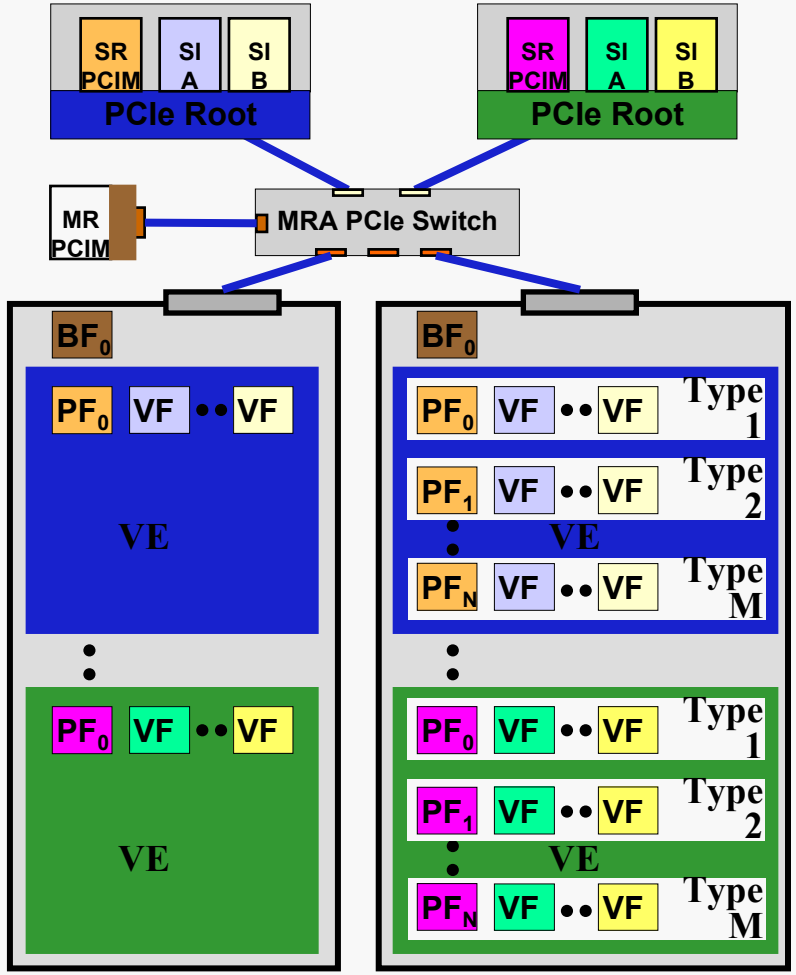
Overview of Function Types used in Single Root Topologies

Function type	Owned by
Physical function	SR-PCIM
Virtual function	System Image

- Physical Function
  - ✓ Used by SR-PCIM to discover and set-up a function's IOV capabilities, such as configuring VFs.
  - ✓ Physical Function 0 is also used by SR-PCIM to manage the physical EP, such as physical errors and events.
- The intention is to enable isolation of Physical Functions to SR-PCIM.
- System Image sees VFs assigned to.

Multi-Root IOV Function Types and Terms

MR Topology



MR Topology Terms

Virtual Endpoint (VE) is the set of physical and virtual functions assigned to an RC.

Each VE is assigned to a **Virtual Hierarchy** (VH).

Virtual Hierarchy is a fully functional PCIe hierarchy that is assigned to an RC or MR-PCIM. Note, all PFs and VFs in a VE are assigned the same VH.

Base Function (BF) only 1 per EP and is used by MR-PCIM to manage an MR aware EP (e.g. assigning functions to Virtual Endpoints).

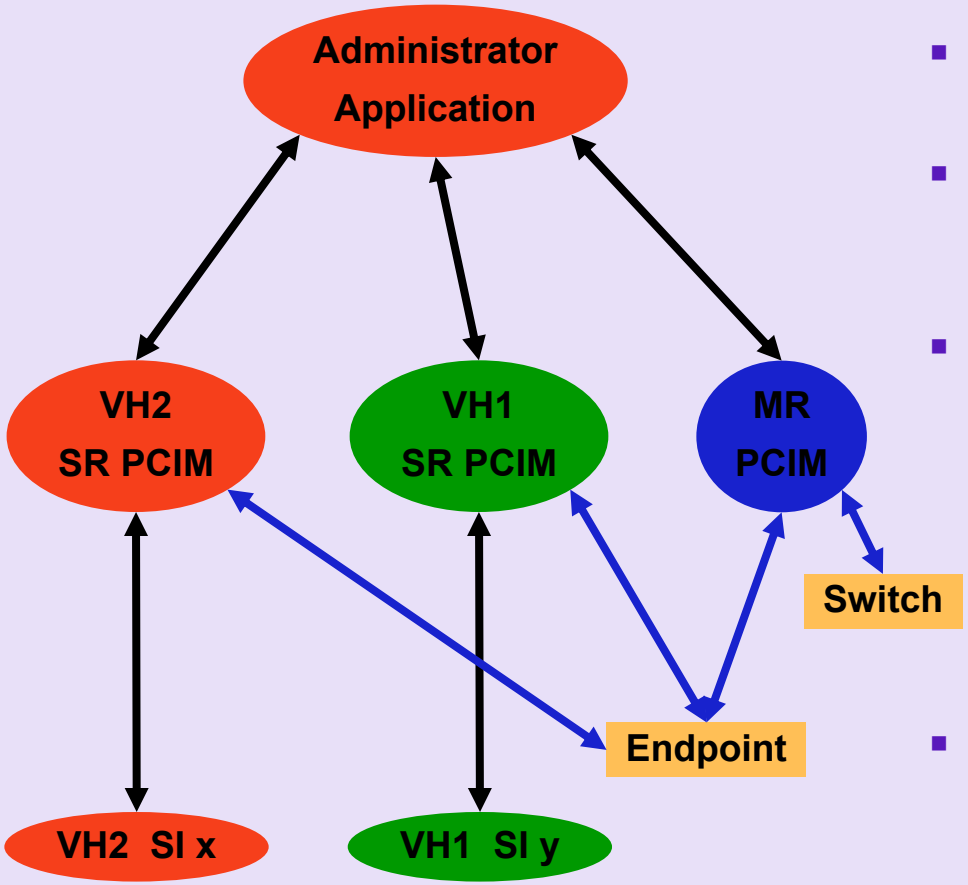
More details later …

Overview of Function Types

Function type	Who owns it for Single Root	Who owns it for Multiple Root
Base function		MR-PCIM
Physical function	SR-PCIM	SR-PCIM
Virtual function	System Image	System Image

- MR-PCIM uses **Base Function** to discover and set-up MR-IOV and SR-IOV capabilities, such as:
  - ✓ Assign PFs to VHs
  - ✓ Assign VFs to PFs
- SR-PCIM uses **Physical Function** to discover and set-up SR-IOV capabilities, such as:
  - ✓ Assign VFs to SIs
  - ✓ Configure BARs for VFs

Management Interfaces (example)



- Administrator Application initiates and manages the transition
- Administrator Application interacts with PCIM(s)
- Administrator Application may interact with SI(s) and/or application being migrated
- All these Interaction Details are NOT Specified (and are out of PCISIG scope)
  - ✓ Only one possible scheme shown here and that is only to provide context
- Blue Interfaces will be standardized
  - ✓ PCIM ←→ Switch
  - ✓ PCIM ←→ Endpoint

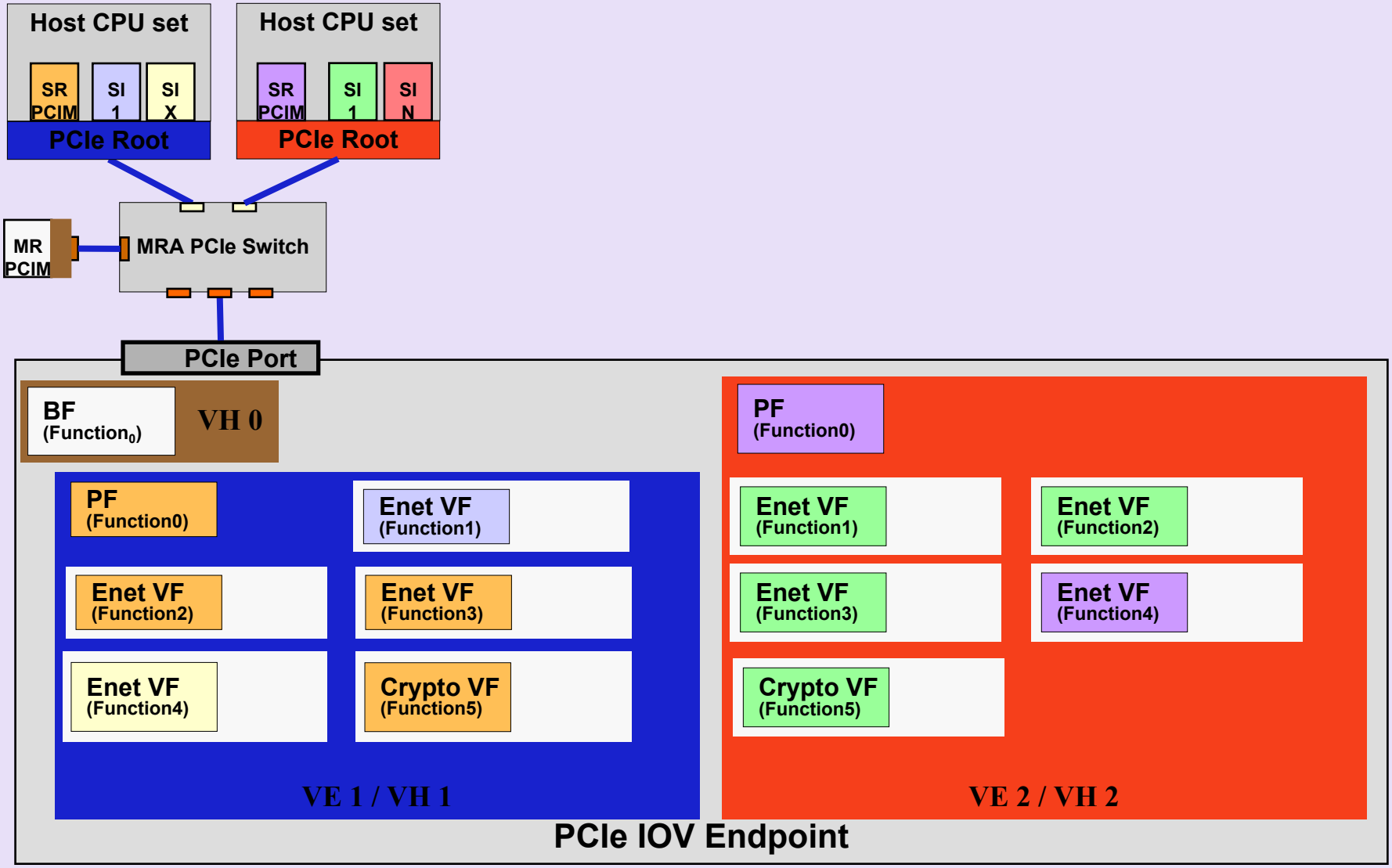
Contents

- Migration Specification Goal and Usage Scenarios
- Review of Single-Root and Multi-Root Terms
- **Multi-Root Stateless VF Migration**
- Multi-Root Stateless PF Migration

Multi-Root Stateless VF Migration Usage Scenario

- Application uses multi-function dual ported Ethernet adapter
- Virtual Function map
  - ✓ VF1..4 Ethernet functions
  - ✓ VF5 Crypto function
- Application using VF5 migrates to new system image
  - ✓ Source system image in VH2
  - ✓ Destination system image in VH1
- Destination system image in VH1 requires Crypto Service
- Migration policy
  - ✓ Source VF5 in VH2
  - ✓ Destination VF5 in VH1

Initial Resource Allocation At Boot

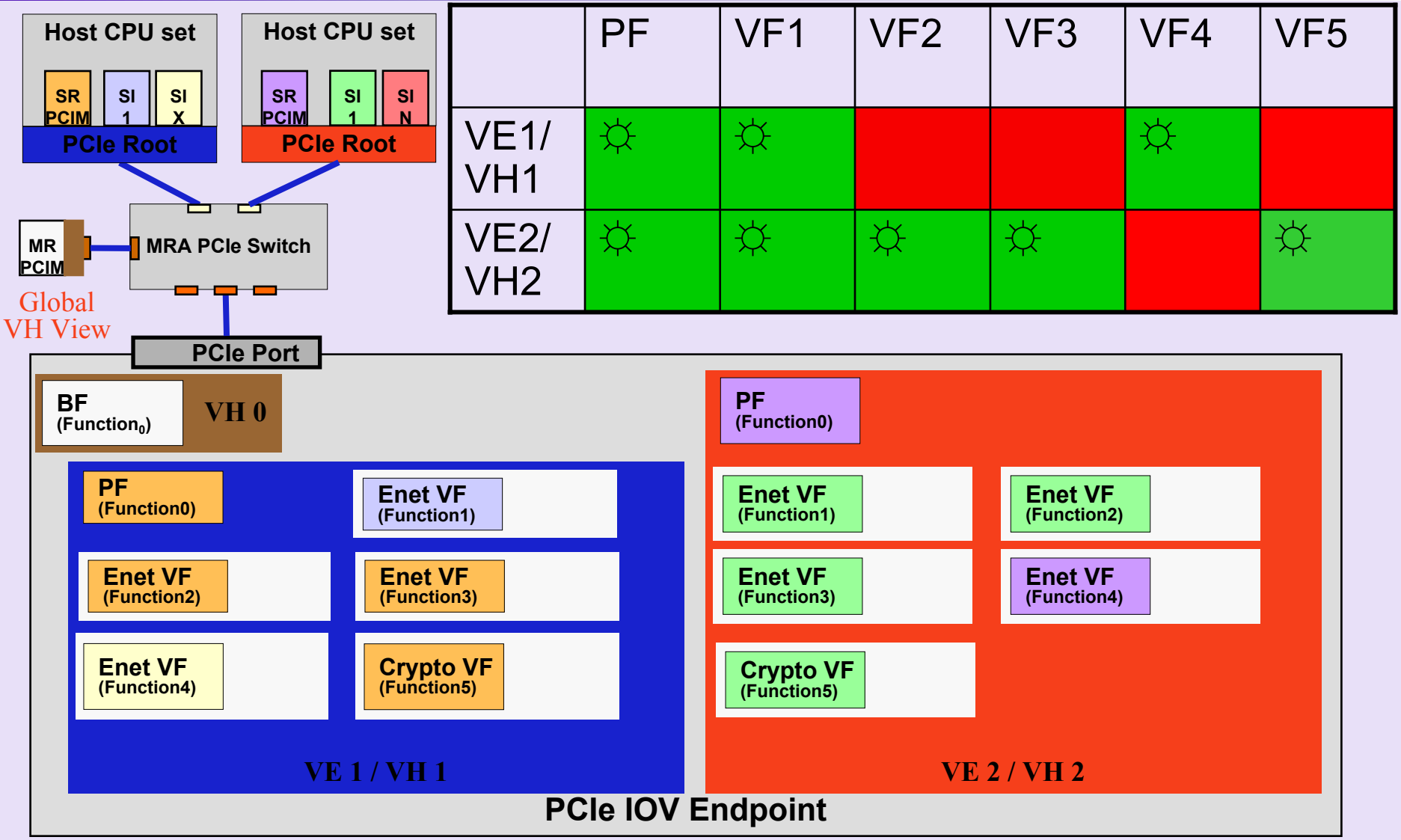


Multi-Root Stateless VF Migration High Level Steps*

1. Administrator asks VH2 SR-PCIM to Hot Remove VF5 from SI x
   a) VH2 SR-PCIM initiates Hot Remove to SI x
   b) SI x indicates Hot Remove done
   c) VH2 SR-PCIM indicates Hot Remove done to Administrator
2. Administrator asks MR-PCIM to move VF5 from VH2 into VH1
   a) MR-PCIM updates EP and Switch tables as appropriate (delete original, add new)
   b) MR-PCIM requests Function Level Reset of VF5
   c) MR-PCIM indicates done to Administrator
3. Administrator asks VH1 SR-PCIM to Hot Add VF5 into SI y
   a) VH1 SR-PCIM initiates Hot Add to SI y
   b) SI y indicates Hot Add done
   c) VH1 SR-PCIM indicates done to Administrator
4. VF5 Migration complete

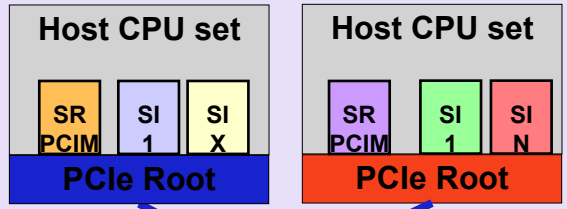
*Note: High level steps represent a possible method for VF migration but don't represent all of the steps and could change with evolving discussion of IOV specification.

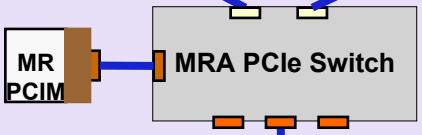
Administrator Application Depicts System State



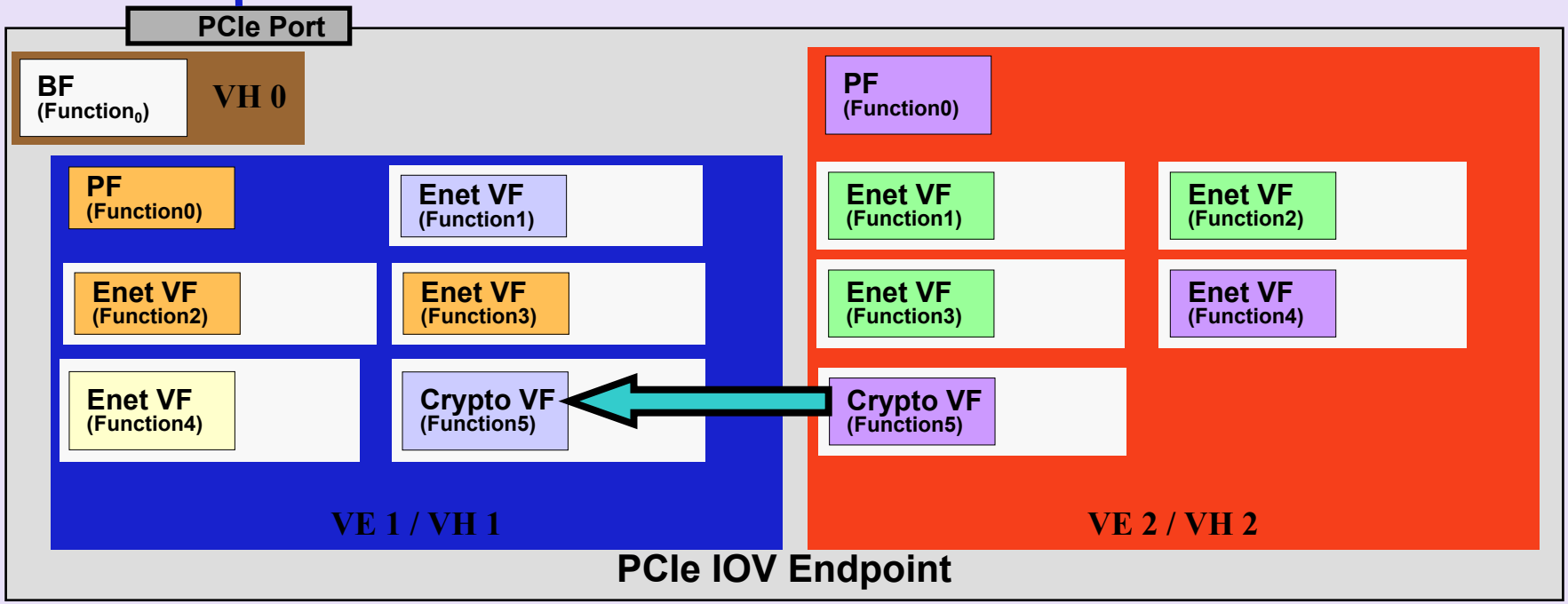
	PF	VF1	VF2	VF3	VF4	VF5
VE1/ VH1	☼	☼			☼	
VE2/ VH2	☼	☼	☼	☼		☼

Multi-Root Stateless VF Migration Complete





	PF	VF1	VF2	VF3	VF4	VF5
VE1/ VH1	☼	☼			☼	☼
VE2/ VH2	☼	☼	☼	☼		



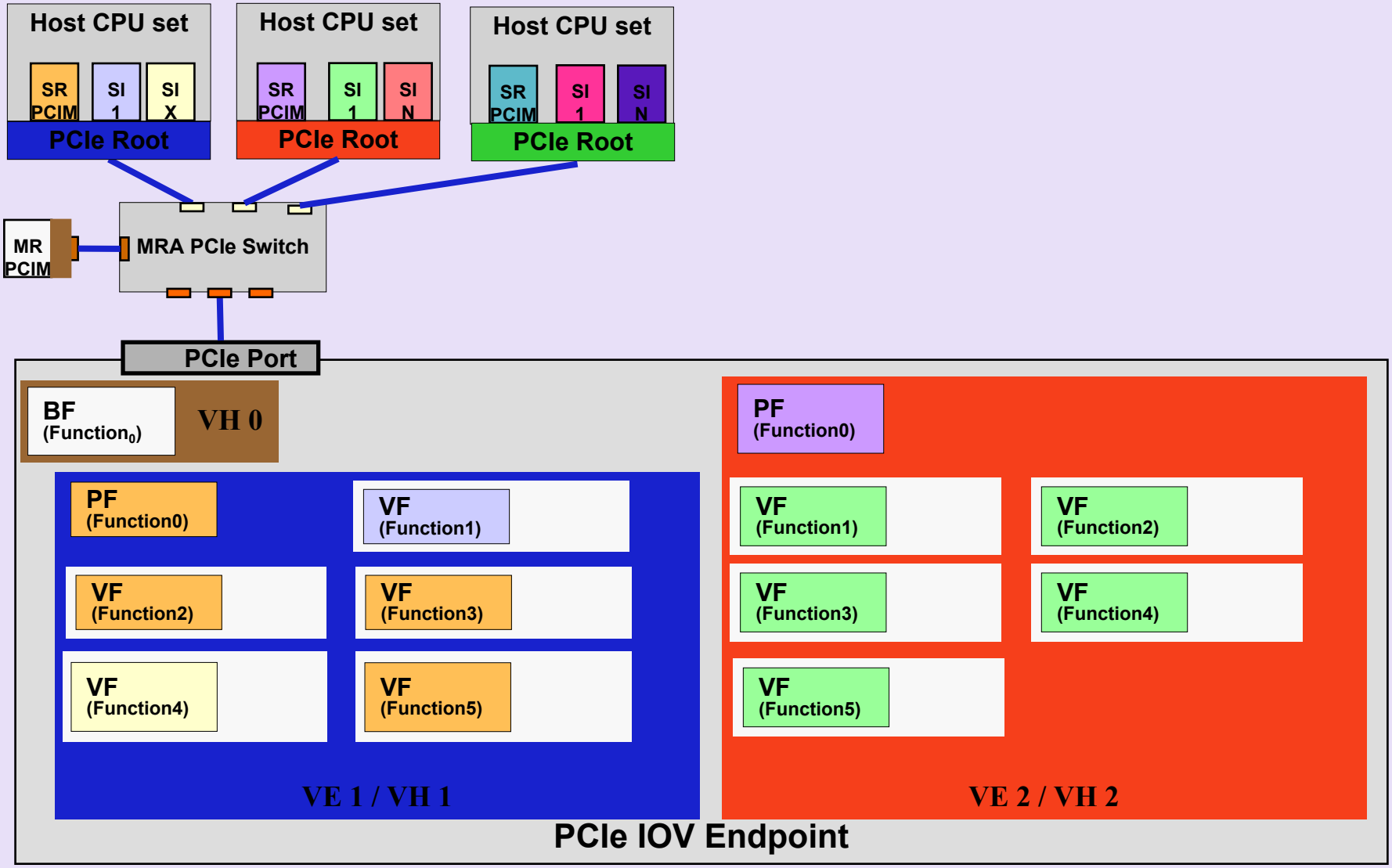
Contents

- Migration Specification Goal and Usage Scenarios
- Review of Single-Root and Multi-Root Terms
- Multi-Root Stateless VF Migration
- **Multi-Root Stateless PF Migration**

Multi-Root Stateless PF Migration Usage Scenario

- Application uses network adapter
- Function map
  - ✓ PF0 in VE2 is the network adapter
- Application using PF0 in VE2 migrates to new system image
  - ✓ Source system image in VH2
  - ✓ Destination system image in VH3
- Destination system image in VH3 requires network adapter
- Migration policy
  - ✓ Source PF0 in VH2
  - ✓ Destination PF0 in VH3

Initial Resource Allocation at Boot

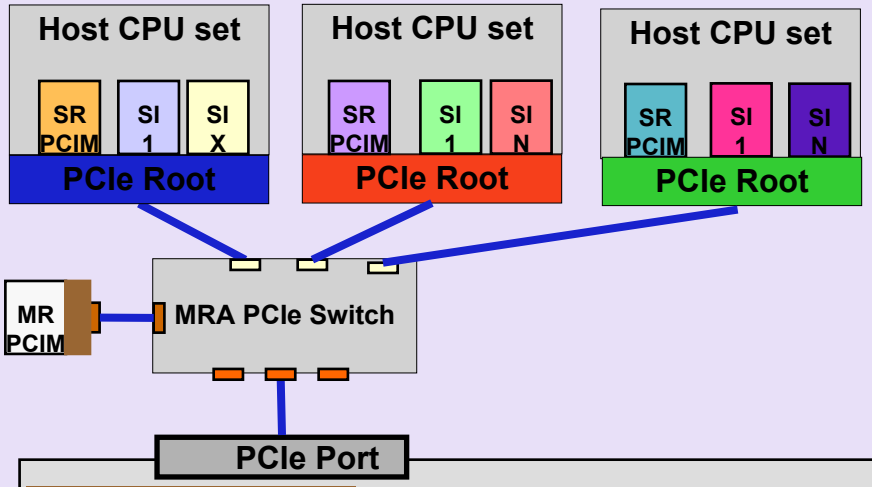


Multi-Root Stateless PF Migration High Level Steps*

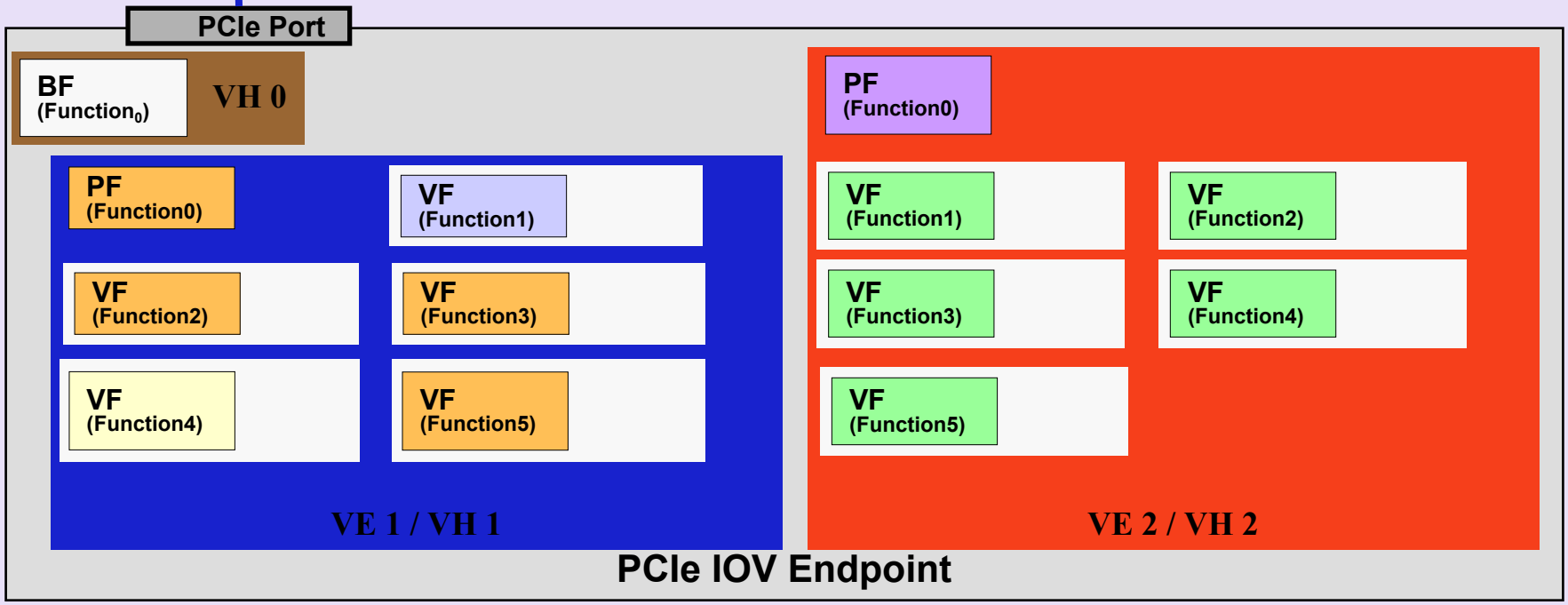
1. Administrator asks MR-PCIM to move VE2 from VH2 to VH3
   a) MR-PCIM "pushes" attention button at Virtual Hot Plug Controller associated with VH2 and port where VE2 is attached
2. VH2 Hot Remove processing
   a) VH2 SR-PCIM gets Hot Remove Request
   b) VH2 SR-PCIM Initiates Emulated Hot Removes from all interested SIs
   c) VH2 SR-PCIM waits for all SIs to finish Emulated Hot Remove
   d) VH2 SR-PCIM initiates DLLP Reset of VE2 to ensure no data leakage out
   e) VH2 SR-PCIM indicates Hot Remove complete using the Virtual HPC
3. continuing in MR-PCIM
   a) MR-PCIM notices VH2 Hot Remove complete
   b) MR-PCIM updates EP and Switch tables as appropriate (delete old, add new)
   c) MR-PCIM initiates Hot Add of VE2 into VH3
4. VH3 Hot Add Processing
   a) VH3 SR-PCIM gets PF Hot Add request
   b) VH3 SR-PCIM configures PF, assigns VFs to SIs
   c) VH3 SR-PCIM initiates Emulated VF Hot Add to SIs
5. MR-PCIM indicates done to Administrator
6. PF Migration Complete

*Note: High level steps represent a possible method for VF migration but don't represent all of the steps and could change with evolving discussion of IOV specification.

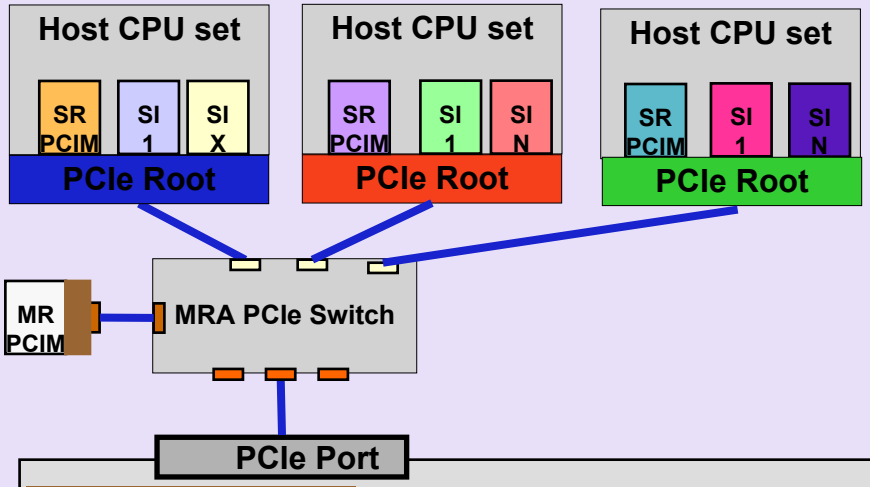
Administrator Application Depicts System State



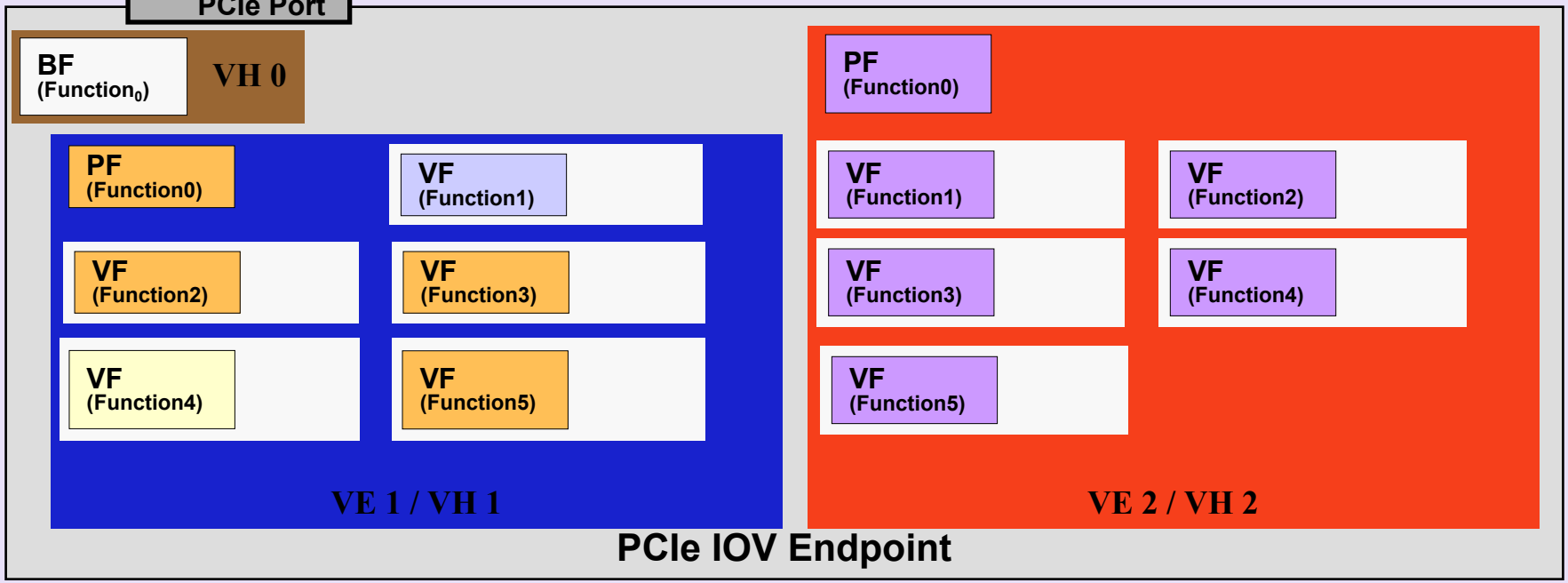
	PF	VF1	VF2	VF3	VF4	VF5
VE1/ VH1	☼	☼			☼	
VE2/ VH2	☼	☼	☼	☼	☼	☼



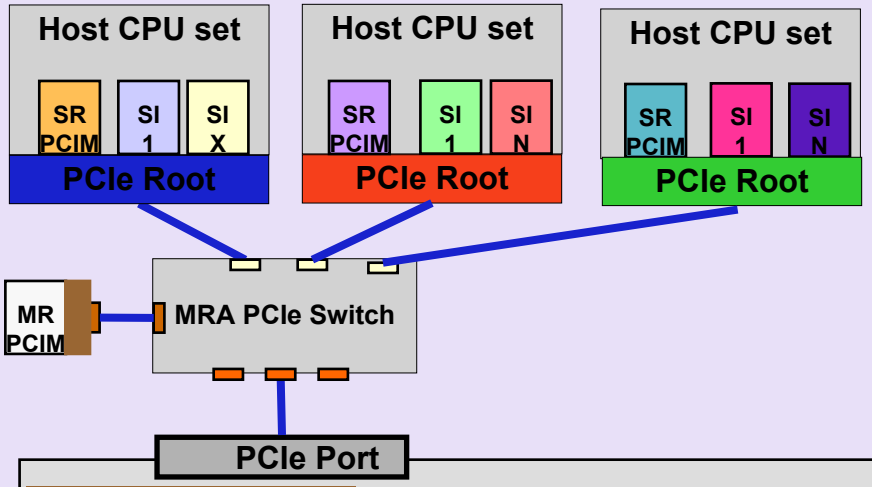
MR PCIM uses switch Hot Plug Controller to Hot Remove in VH2



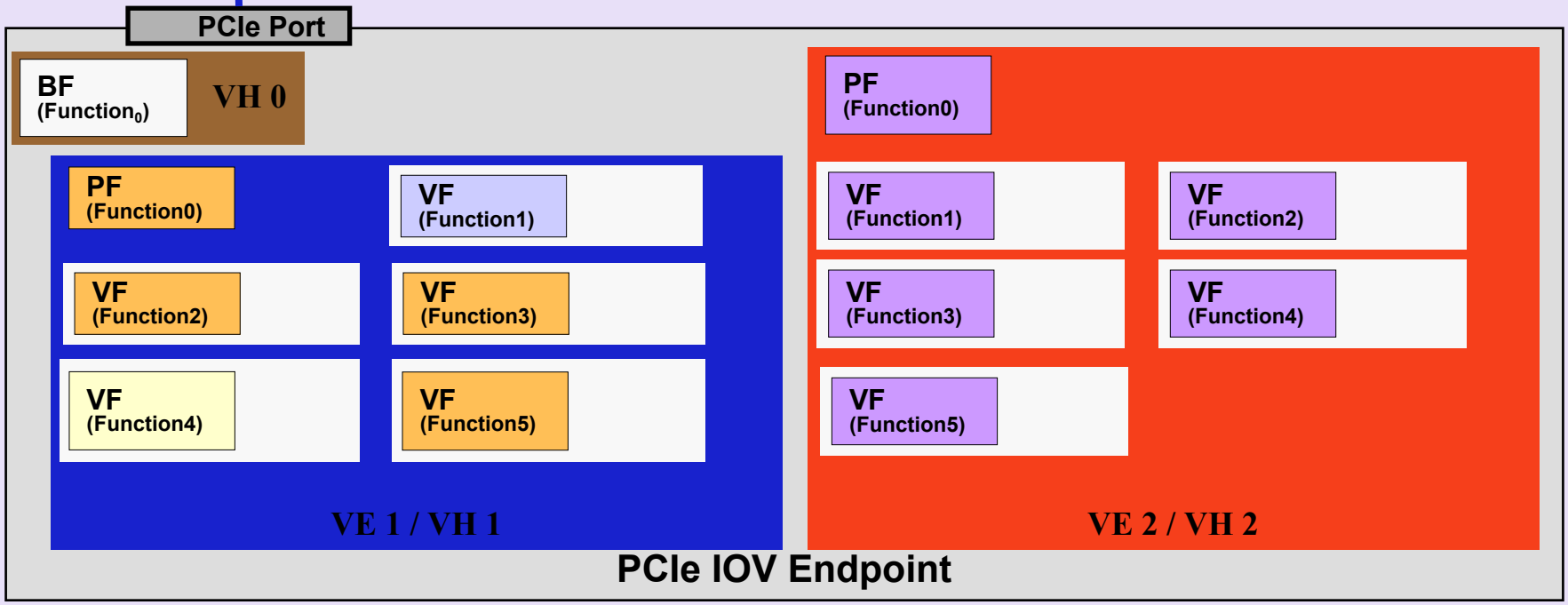
	PF	VF1	VF2	VF3	VF4	VF5
VE1/ VH1	☼	☼			☼	
VE2/ VH2						



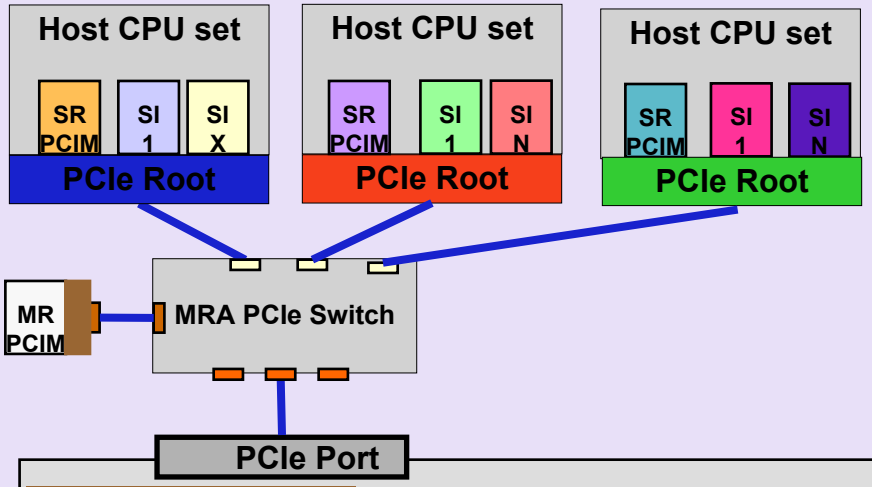
DLLP based Hot Reset of VE2 Functions



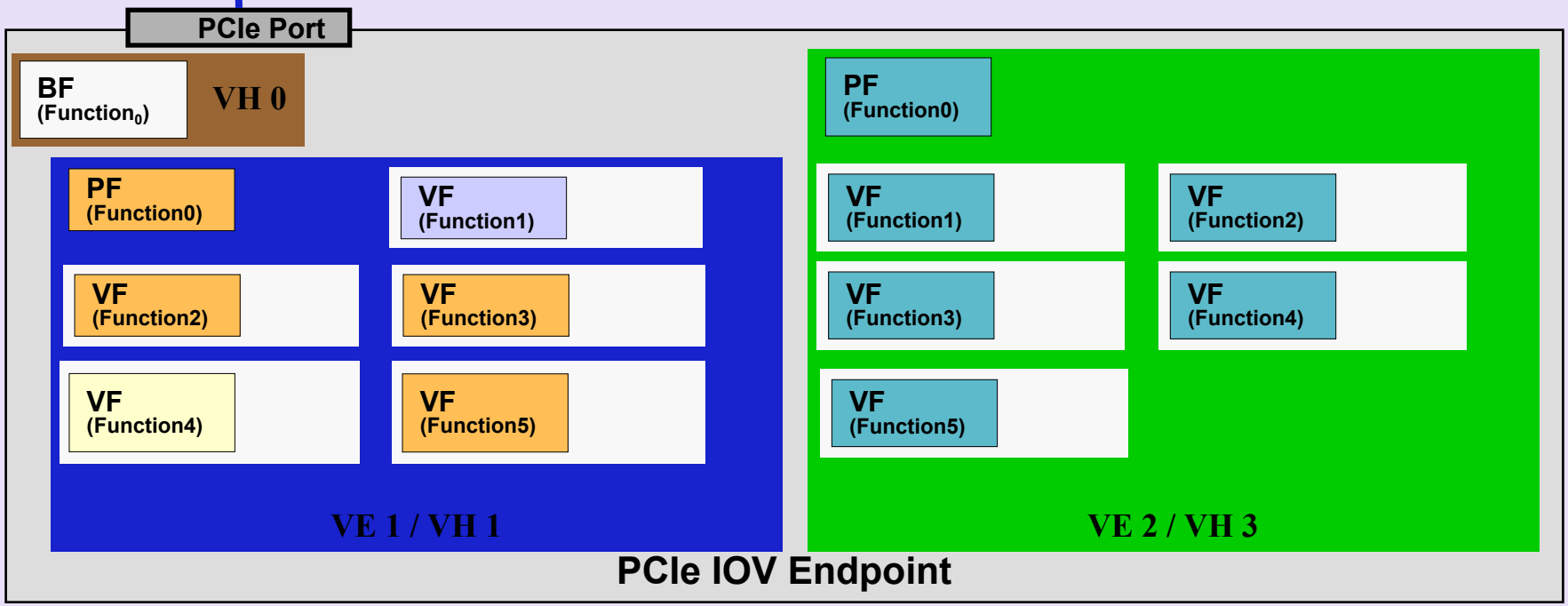
	PF	VF1	VF2	VF3	VF4	VF5
VE1/ VH1	☼	☼			☼	
VE2/ VH2						



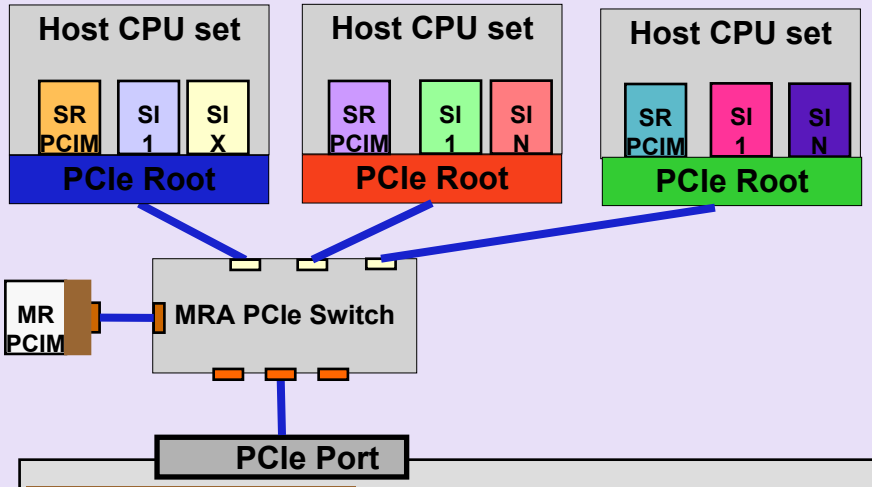
MR PCIM maps VE2 from VH2 to VH3 and tells VH3 Hot Add



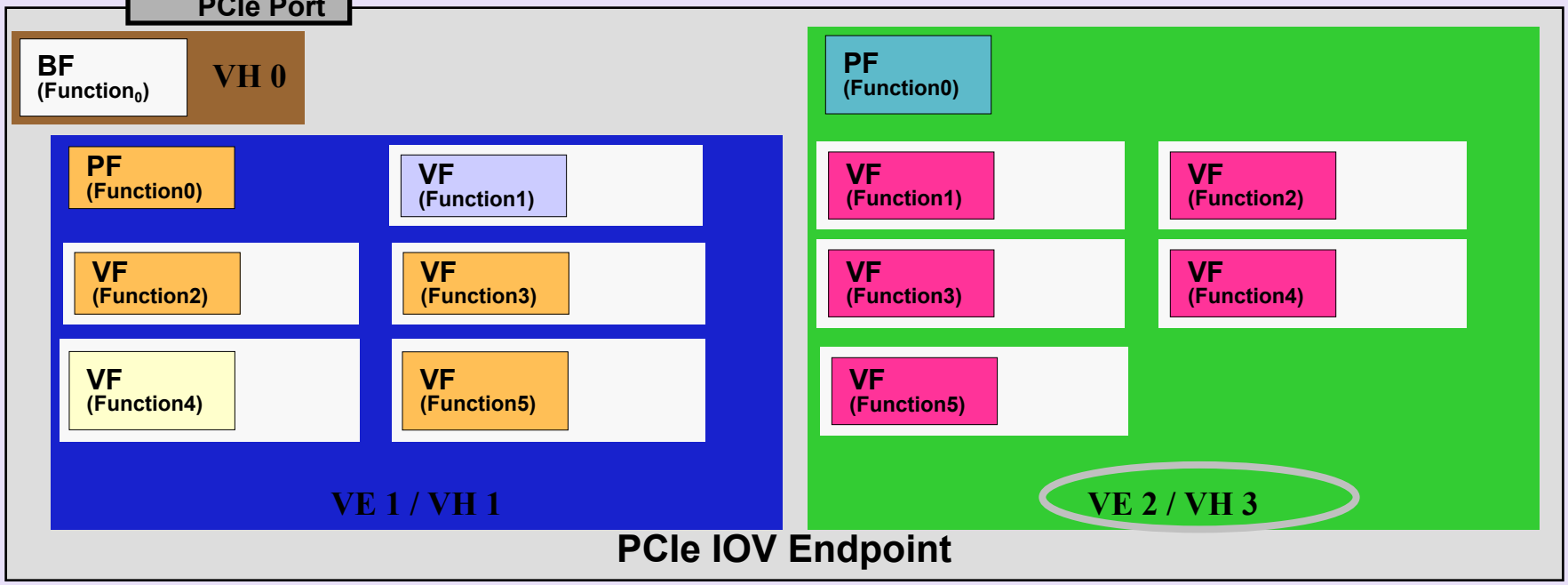
	PF	VF1	VF2	VF3	VF4	VF5
VE1/ VH1	☼	☼			☼	
VE2/ VH3	☼	☼	☼	☼	☼	☼



Multi-Root Stateless PF Migration Complete



	PF	VF1	VF2	VF3	VF4	VF5
VE1/ VH1	☼	☼			☼	
VE2/ VH3	☼	☼	☼	☼	☼	☼



Summary

- Designing mechanisms to migrate VF and PF
- We will continue to consider usage scenarios
- Build on functions created in IOV specification
- Refine usage models, algorithms etc…
- Document usage models as implementation notes

Thank you for attending the
PCI-SIG IO Virtualization Training 2006.

For more information please go to
www.pcisig.com