Advancements in SR-IOV Technology in Cloud Computing SSDs

Douglas Arens, Technical Staff, Applications
Microchip Technology Inc.

Agenda

1. **What is SR-IOV?**
   - Current standard to achieve multi-tenancy
   - Multiple SR-IOV instances on Server attach to NVMe® Objects (Namespaces) on SSD drive
2. **What is SIOV?**
   - How is it the same?
   - How is it different?
     - Footprint
     - Performance
     - Software ecosystem
     - DRs vs VFs
   - Are there any disadvantages?
3. **Market preferences/adoption**
4. **What comes next?**

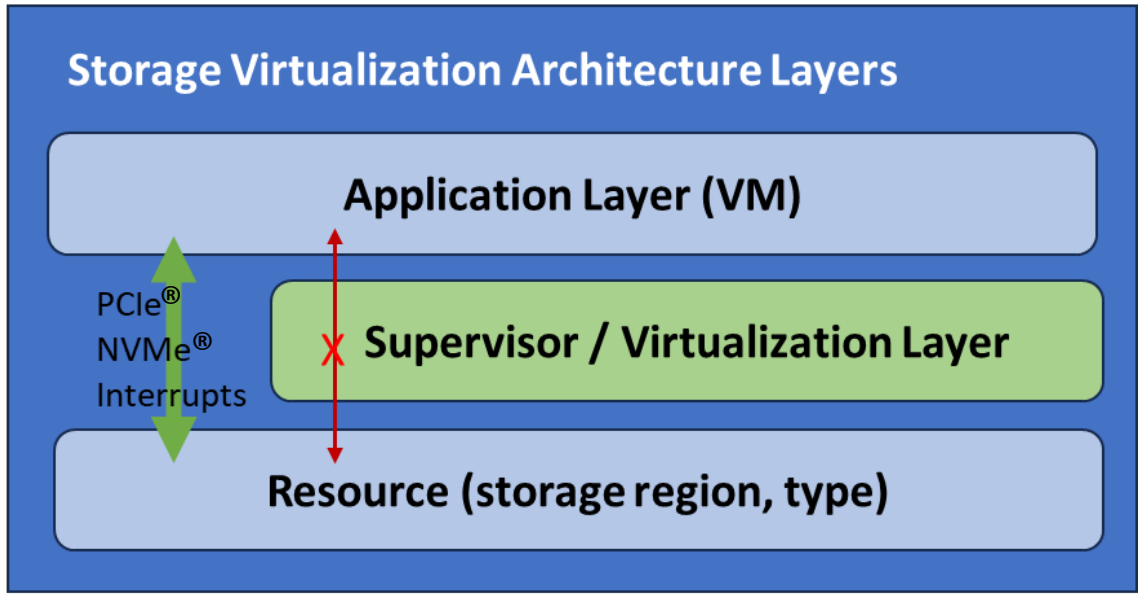
SR-IOV: Virtualization Architecture and Benefits

Architecture Layers

- More direct-access from an Application (VM) to a resource
- Minimal supervisor-layer to initially configure access from Application to the Virtual resource.
- Then supervisor-layer gets out of the way.
- Utilize the direct-access as a superhighway for unencumbered data flow

Benefits of Virtualization

- ✓ Reduced capital and operating costs
- ✓ Minimized or eliminated downtime.
- ✓ Increased IT productivity, efficiency, agility and responsiveness
- ✓ Faster provisioning of applications and resources



SR-IOV: Virtualization Architecture and Benefits

What is the impact at the datacenter?

1. **Hardware Improvements**
   - More tenants per platform (density, compact)
   - Larger SSD drives to utilize existing PCIe® slots
   - More HW automation (performance)
2. **Firmware/Software improvements**
   - Focus on customer application needs
   - FW/Driver optimizations
   - Application improvements

SR-IOV: Challenges of Multi-tenant Virtualized Systems

1. Server Memory
   - More VMs (tenants) require more RAM
2. Server Storage
   - Large boot drive with separate VMs
   - More VMs (tenants) often require more storage
3. Server Performance
   - Each tenant takes
4. Maintain application independence & isolation
5. Maintain Security while Sharing System
6. Migrate Applications and Namespace Data

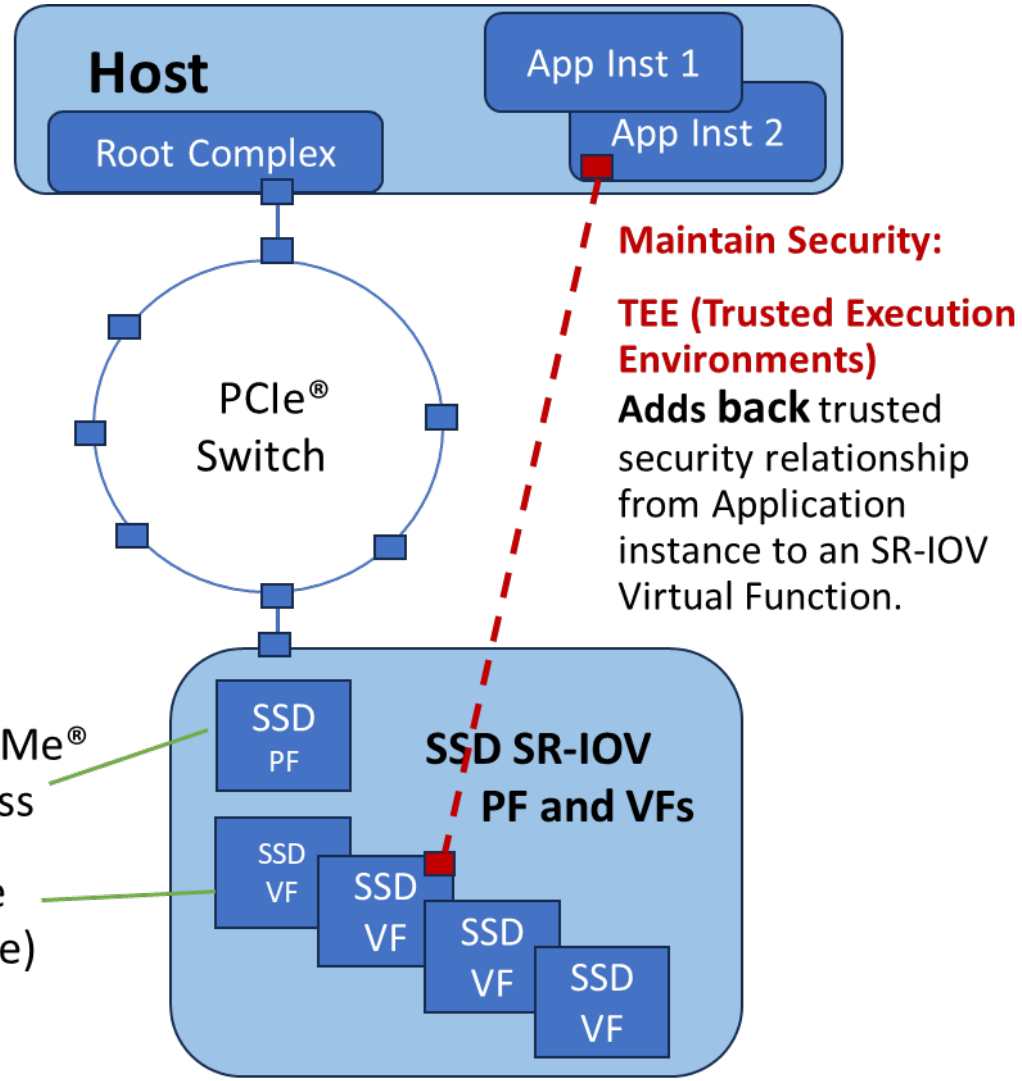
PCIe® SR-IOV (Single-Root I/O Virtualization)

USE CASE: Multifunction device at PCIe layer, better cost/features

SR-IOV allows a PCIe physical device under a single root complex to appear as multiple separate devices to the hypervisor or the guest operating system.

Virtualization Benefits:

- ✓ It makes it possible to run ~16 physical/virtual functions per SSD, which reduces the need for separate hardware and the resultant costs of space and power required by hardware devices



Maintain Security:

TEE (Trusted Execution Environments)
Adds back trusted security relationship from Application instance to an SR-IOV Virtual Function.

Physical Function- Real NVMe® device, configuration, access

Virtual Function- Exchange of data (Vendor defined use)

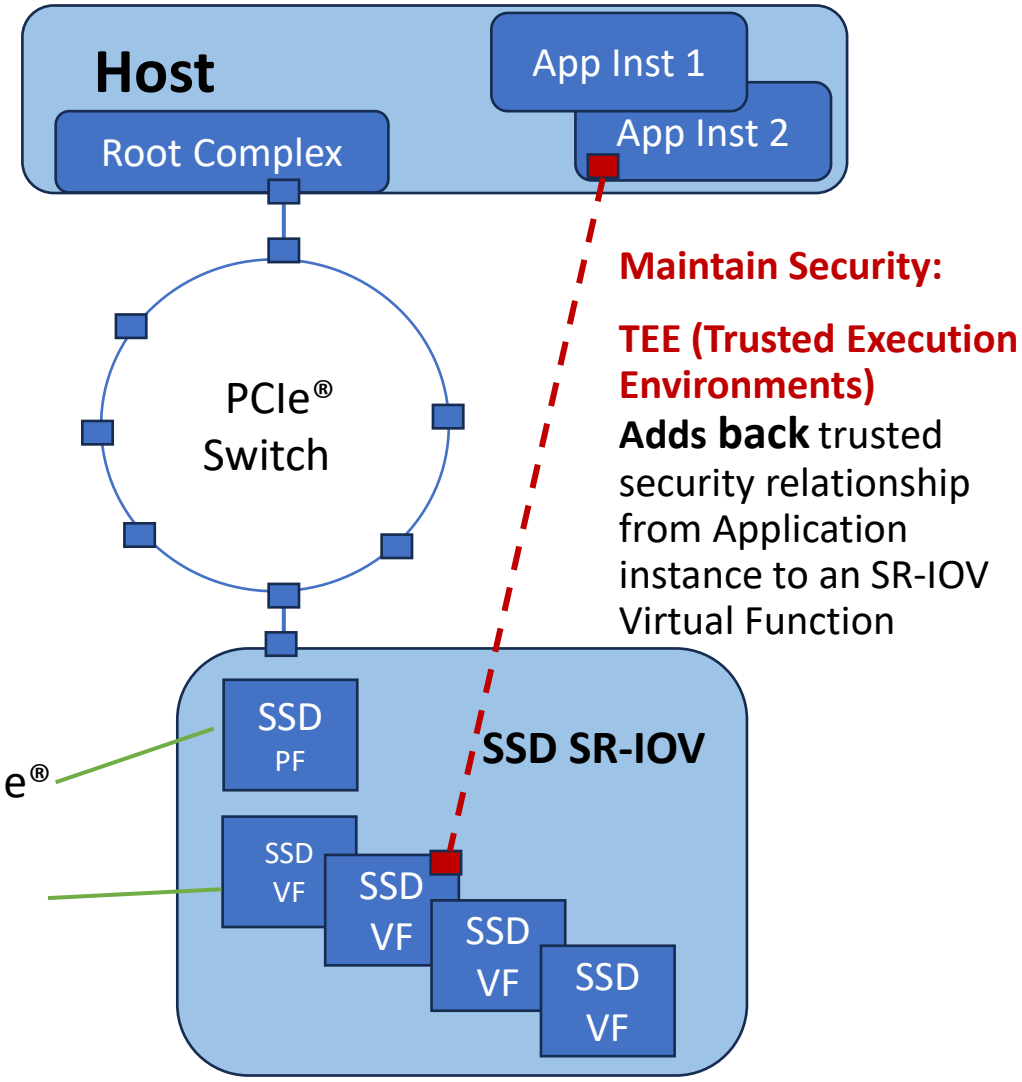
PCIe® SR-IOV (Single-Root I/O Virtualization)

Limitations

- PCIe® endpoint
- Resources for VFs and Apps

Moving beyond those limitations:

- Become PCIe switch, not endpoint
- HW automation
- SW/FW efficiency
- Resource use is unavoidable



Physical Function- Real NVMe® device, configuration, access

Virtual Function- Exchange of data (Vendor defined use)

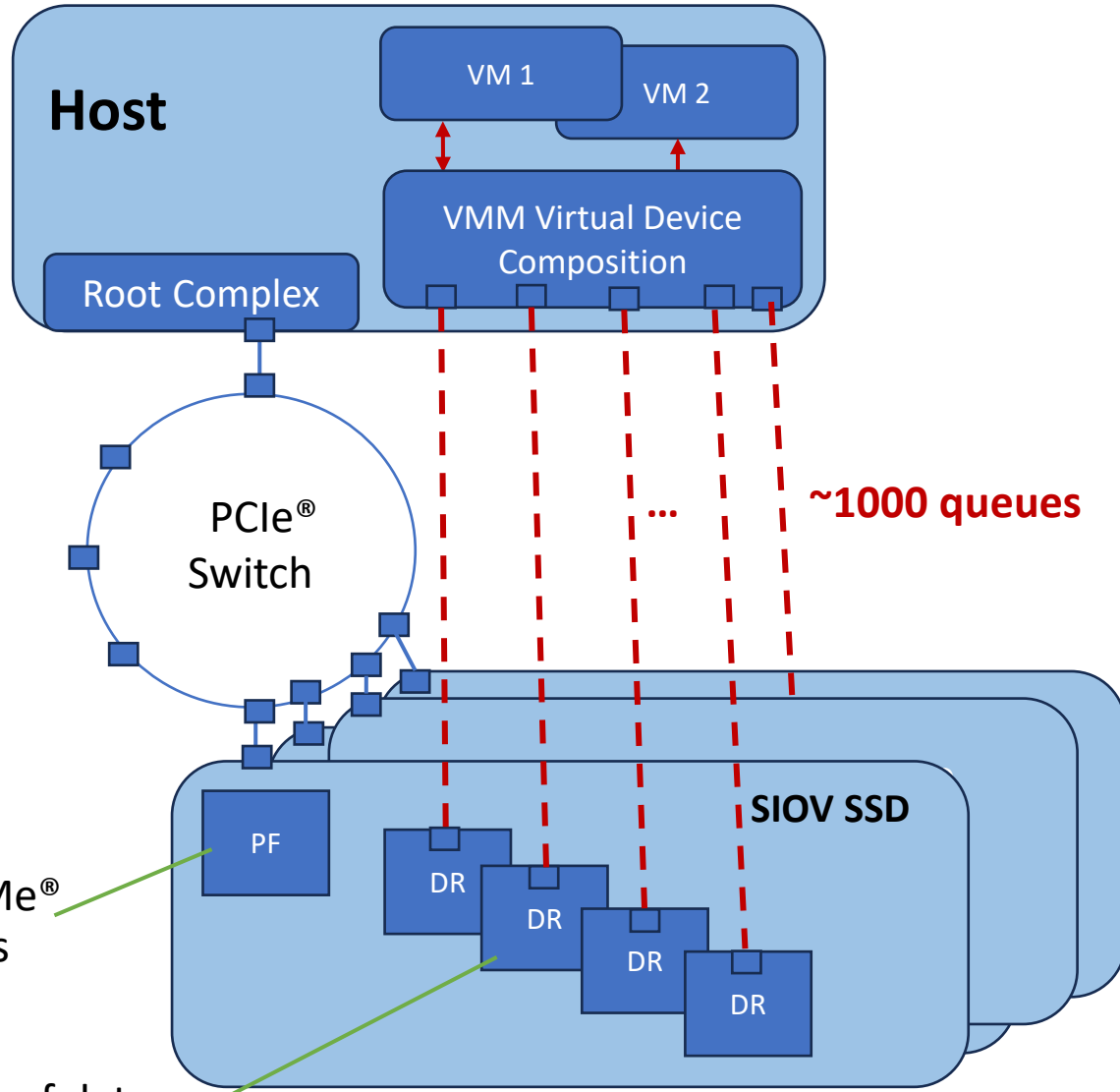
Intel® Scalable I/O Virtualization

USE CASE: Highly Scalable Multifunction device

Intel Scalable IOV improves upon SR-IOV to allow a single SSD under a single root port to virtualize ~1000 separate storage devices for each host in hyper-scale datacenters

Intel Scalable I/O Benefits:

- ✓ Storage is 16X more scalable, hardware-assisted IO paired with software flexibility. Scalable 20-bit ID (over 16-bit ID) between VM and DR.
- ✓ DMA and interrupt remapping
- ✓ VMM Directed IO to NVME® queue level at device
- ✓ Separation of fast path I/O from slow path (configuration, reset)



Physical Function- Real NVMe® device, configuration, access

Device Resources- Exchange of data (direct-addressing to queue level)

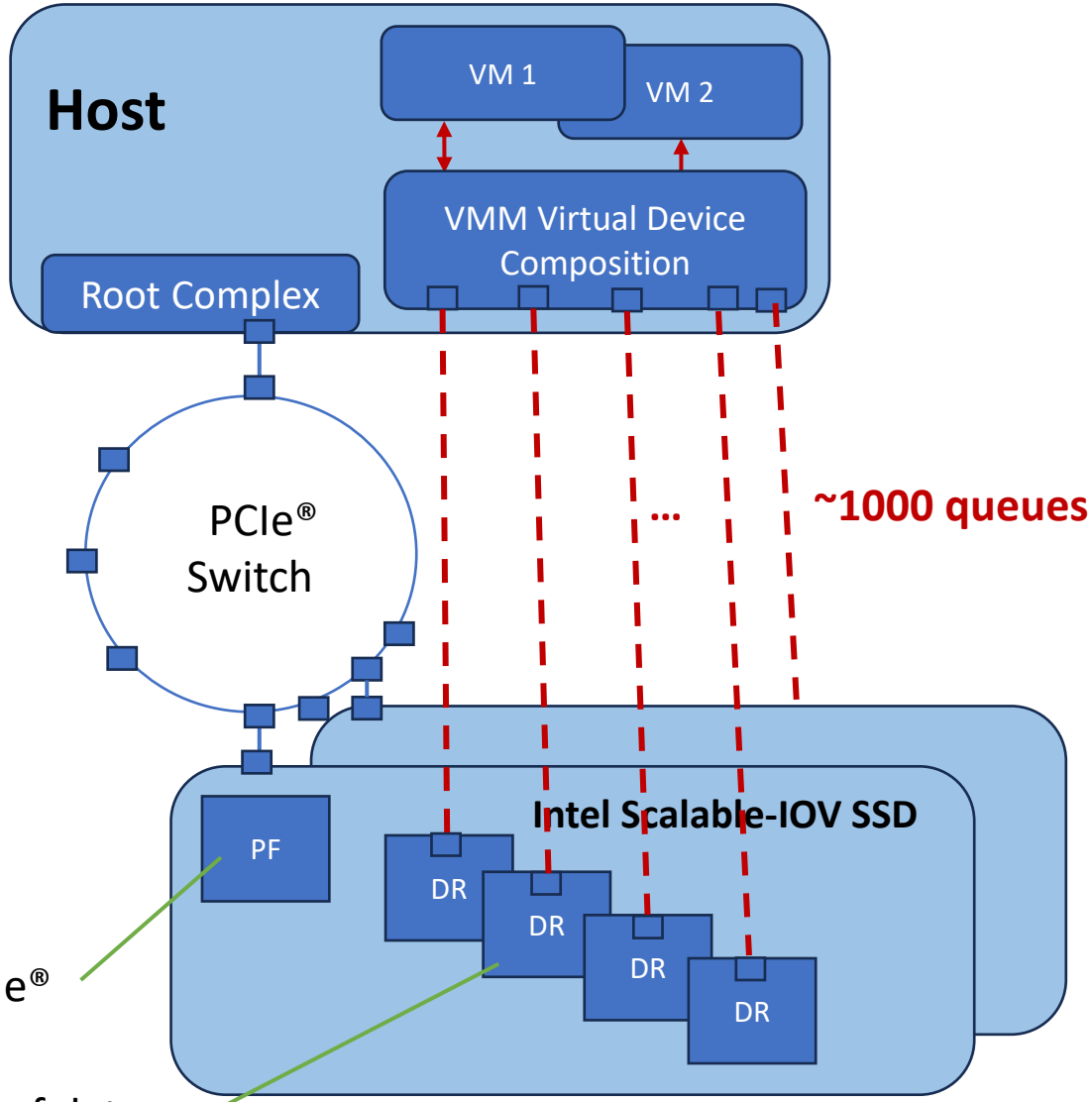
Intel® Scalable I/O Virtualization

Limitations

- PCIe® endpoint
- Resources for VFs and Apps

Moving beyond those limitations:

- Become PCIe switch not endpoint
- HW automation
- SW/FW efficiency
- Resource use is unavoidable



Physical Function- Real NVMe® device, configuration, access

Device Resources- Exchange of data (direct-addressing to queue level)

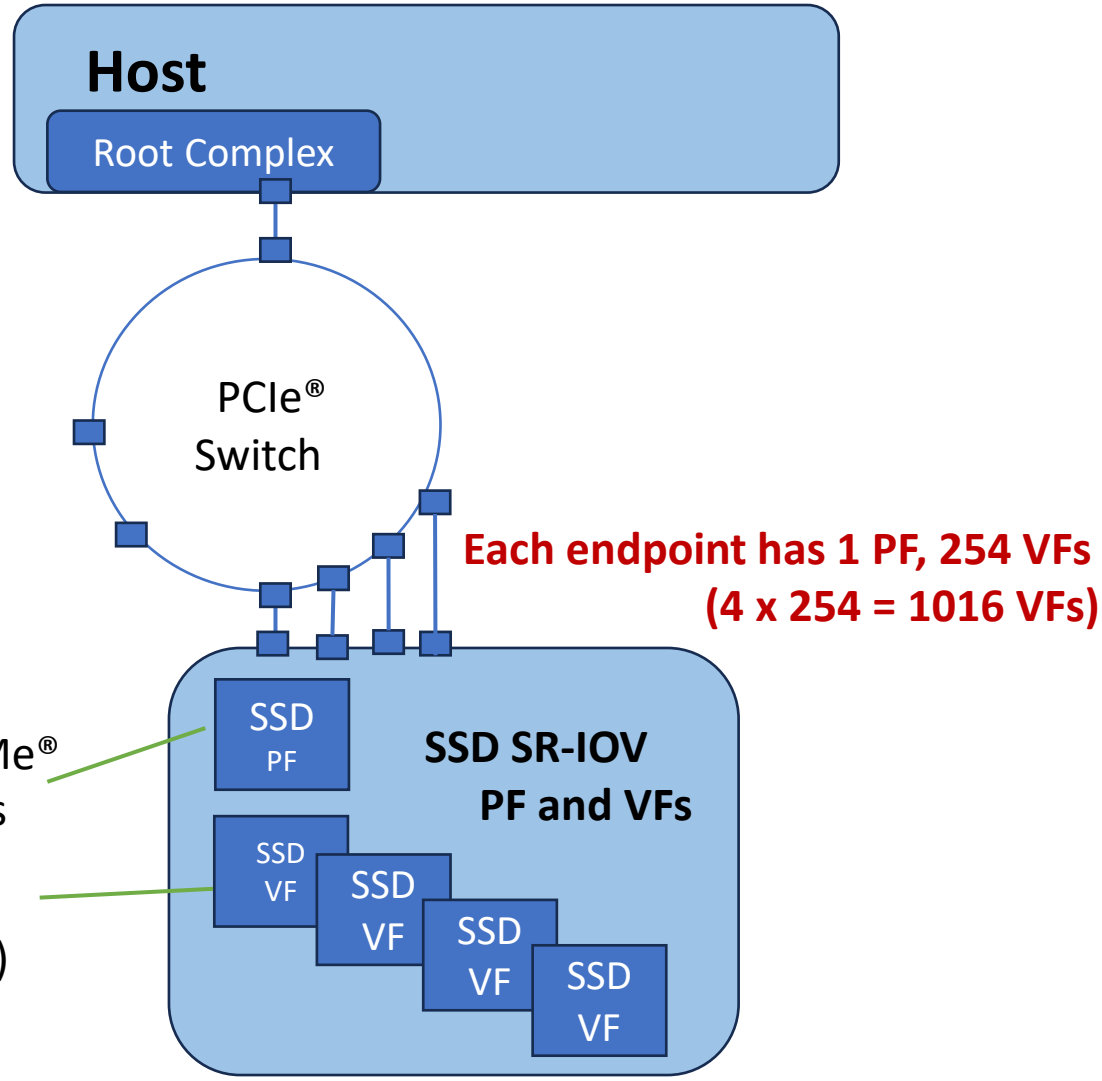
Moving beyond PCIe® limitation of 254 VFs

Gather groups of “Synthetic VFs” under a switch interface instead of just one endpoint

ARI addressing permits 254 VFs per endpoint

(this is the limitation)

If device presents itself as a switch it can have several endpoints inside it. Then add a mapping layer in the device that translates host addressing to the corresponding synthetic VFs.



Physical Function- Real NVMe® device, configuration, access

Virtual Function- Exchange of data (Vendor defined use)

What's next? … Another wave of improvements

1. **Hardware Improvements**
   - More tenants per platform (density, compact)
   - Larger SSD drives to utilize few PCIe slots
   - More HW automation (performance)

2. **Firmware/Software improvements**
   - Focus on customer application needs
   - FW/Driver optimizations
   - More Application features
     - SPDK (OCP support from multiple tech giants)
     - Flexible Data Placement

Thank you !

Visit Microchip Booth # 419