PCI Express in QEmu

Isaku Yamahata <yamahata@private.email.ne.jp>
<yamahata@valinux.co.jp>
VA Linux Systems Japan K.K.
KVM-forum 2010: August 10, 2010

Agenda

- Introduction
- Current status and future work
- Summary

Introduction

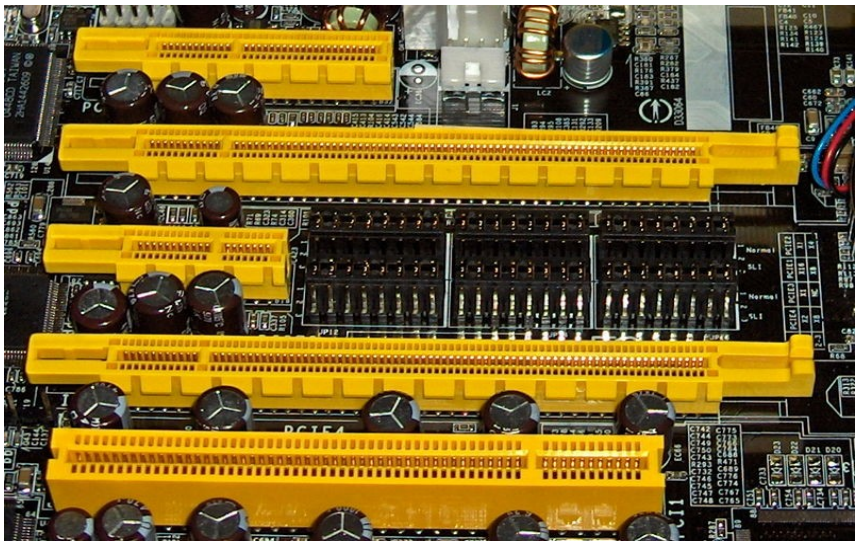
Why PCI Express?

- New features: enhancements as a successor
  - Used as express is widely accepted in the market.
- Some device drivers require express
  - They check if the device is really express
  - Existing PCI device assignment isn't enough
- Hardware certification requires express
- Current PCI support is also limited

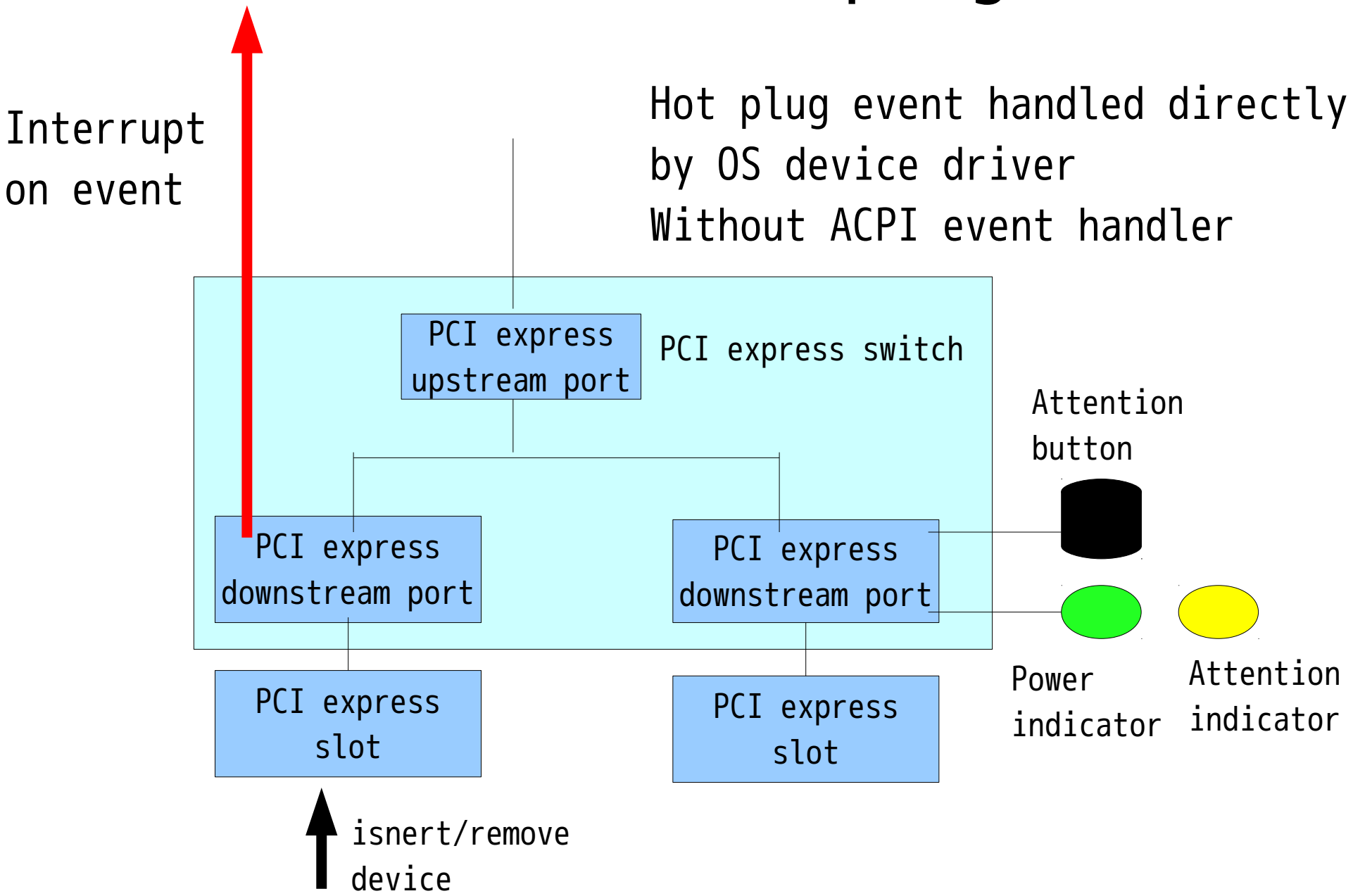
PCI Express features from software point of view

- Many enhancements from PCI, for example
  - MMCONFIG: larger configuration space
  - Native hotplug:not ACPI based
  - Native power management
  - AER(Advanced Error Reporting)
  - ARI(Alternative Routing ID)
  - VC(Virtual Channel)
  - FLR(Function Level Reset)

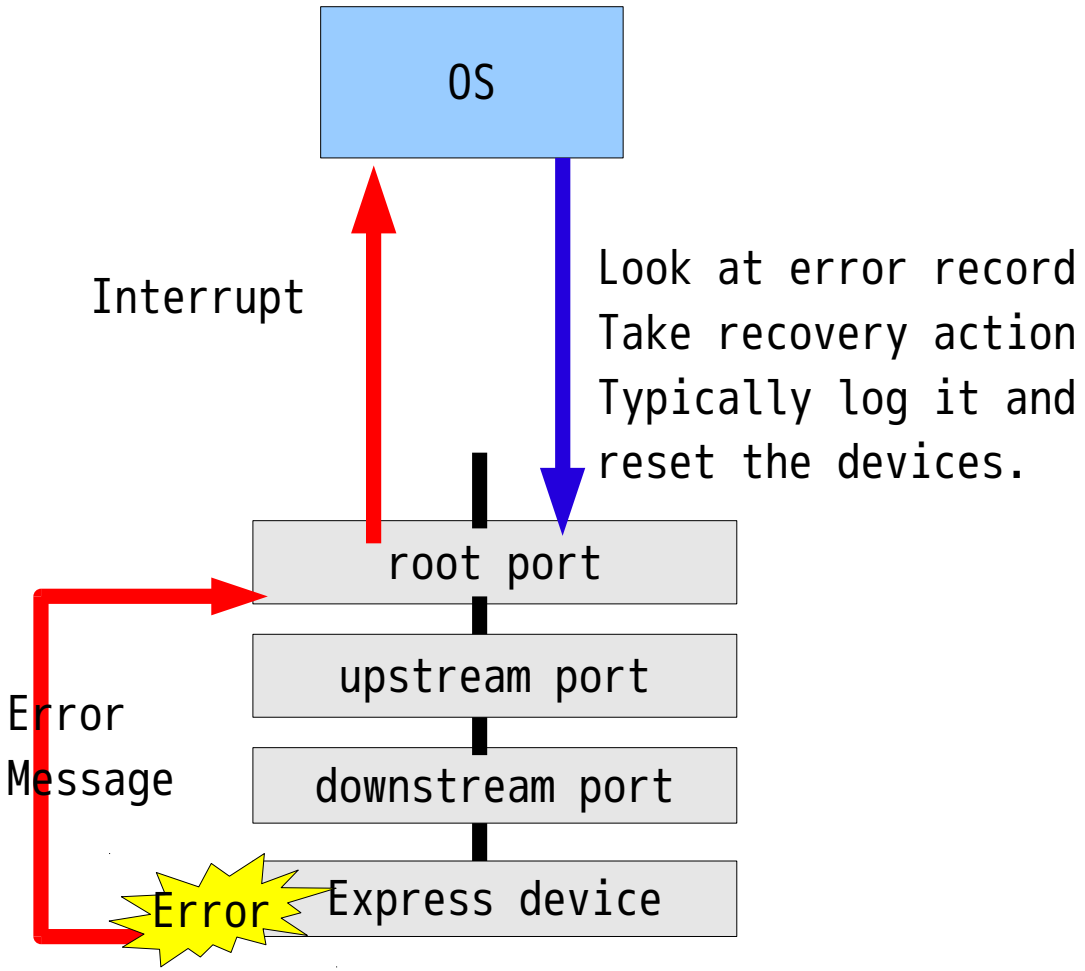
From wikipedia



Native hot plug

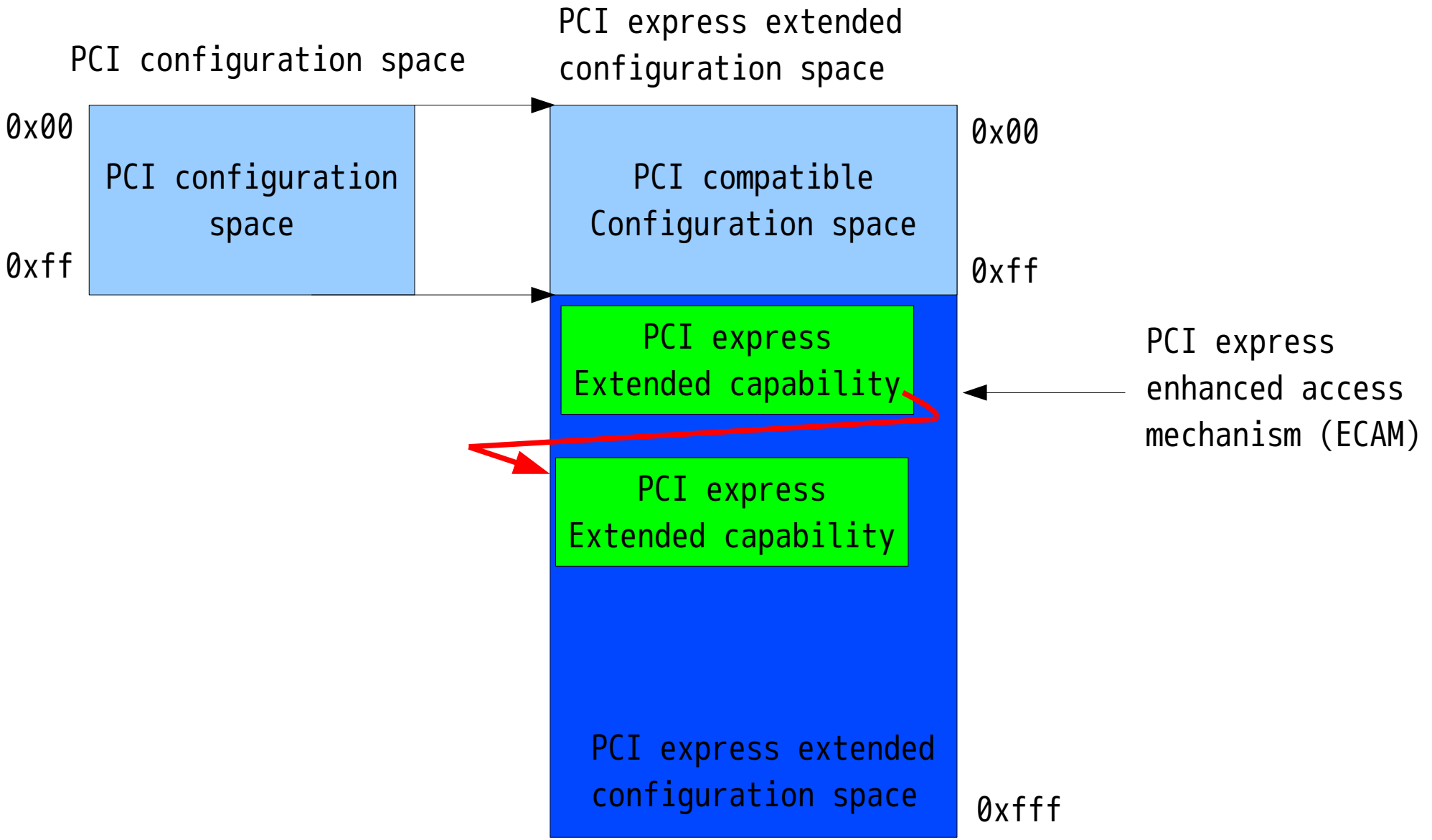


Advanced Error Reporting(AER)

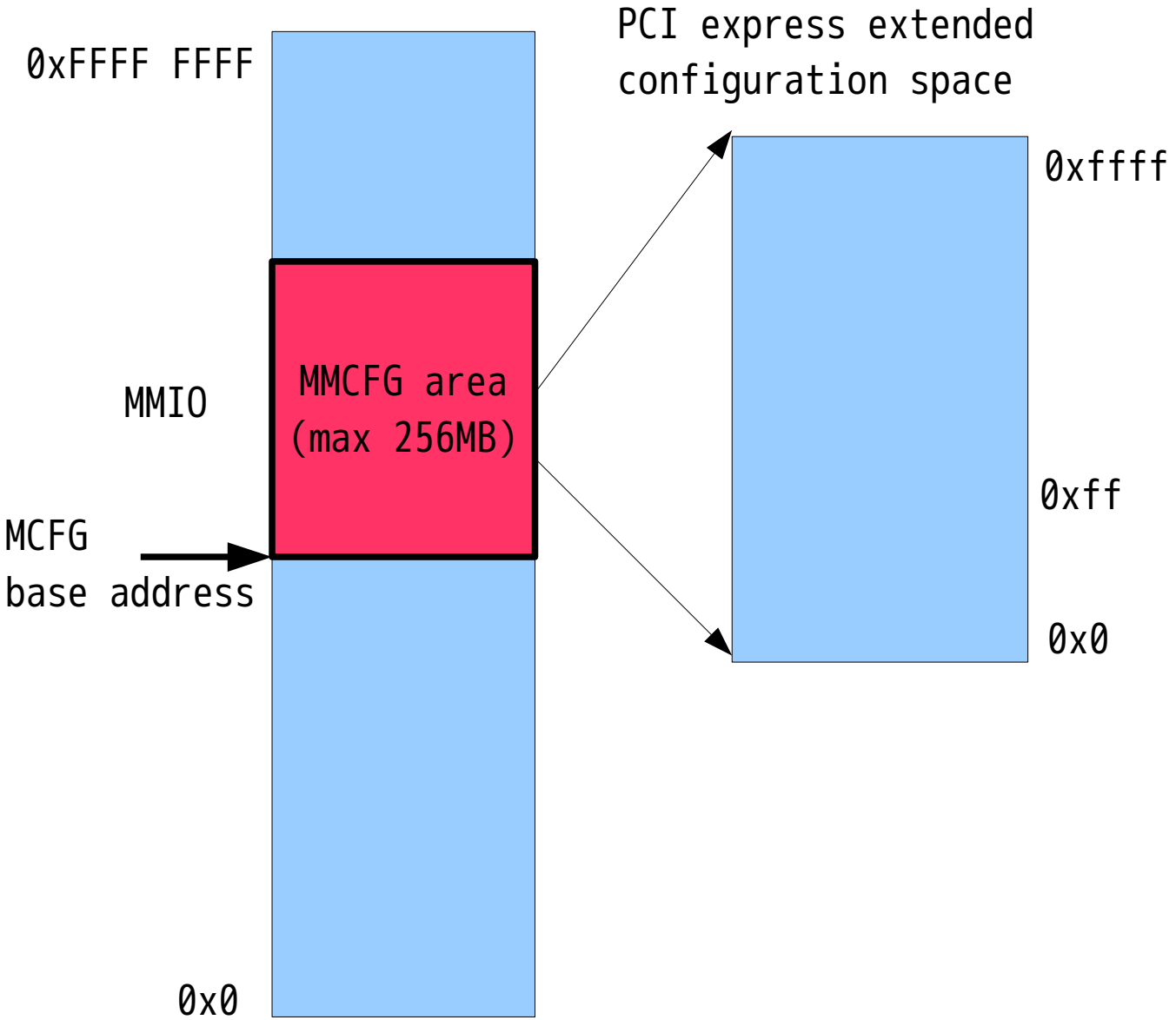


- Standardized error reporting.
- Important for RAS

PCI express extended configuration space

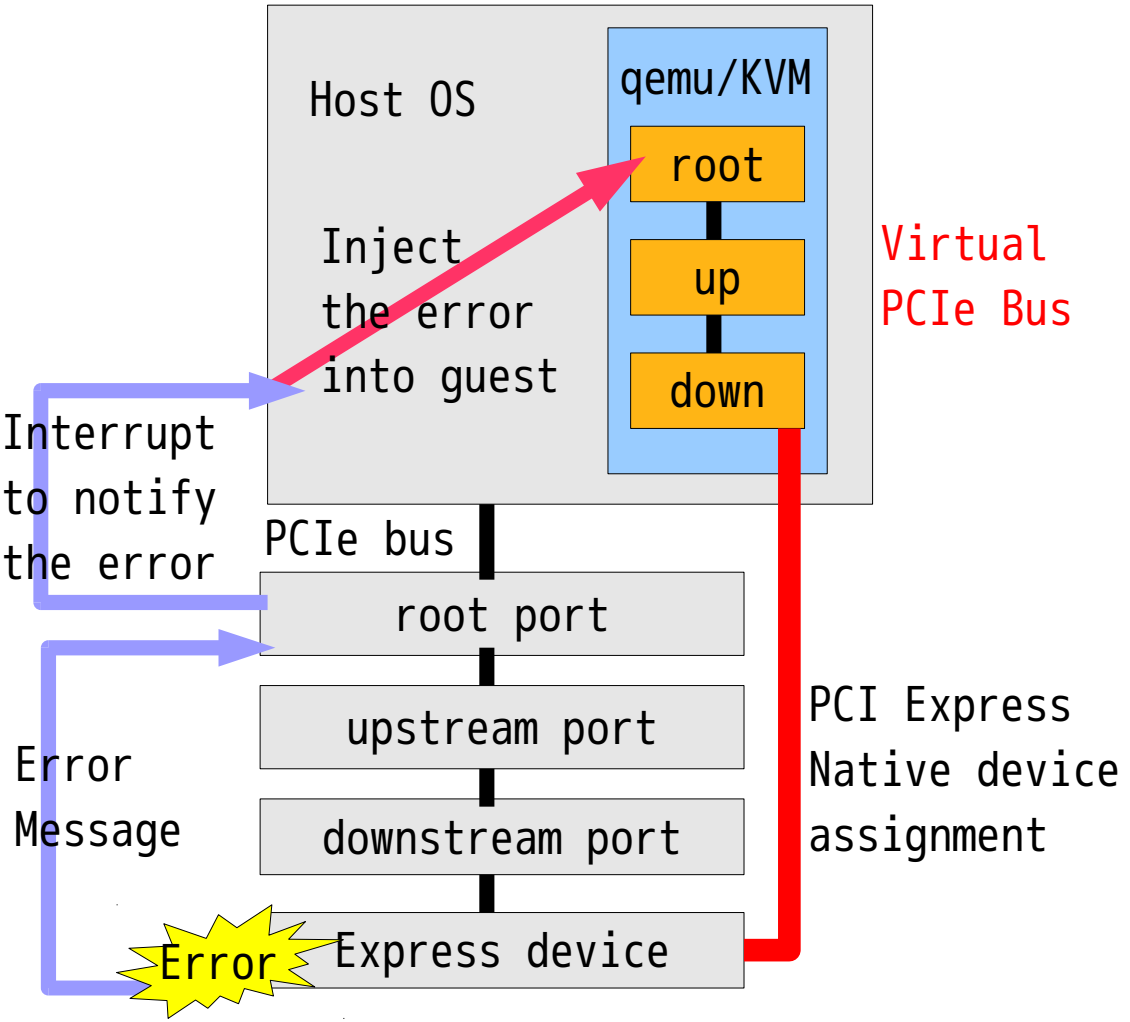


PCIe MMCONFIG



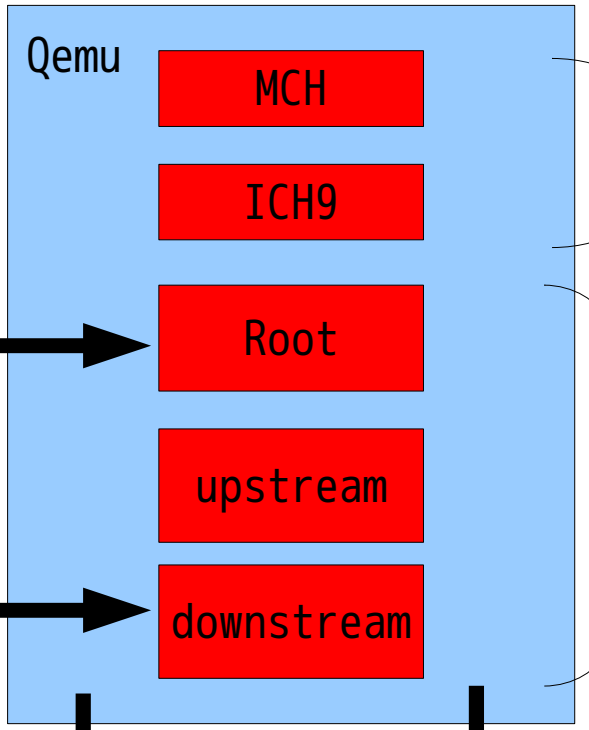
Goal

- Enable QEmu to support PCI Express
- Enable PCI Express native device assignment with
  - Native hot plug
  - RAS
- Then, bring Express support to qemu derivative, ie KVM.



Current implementation and future work

I440fx chipset refactoring
64bit BAR
Extended config space
MMConfig
PCI-to-PCI bridge clean up
PCI bus reset



Q35 chipset
New DSDT

PCI express port switch

AER error injection
pcie_aer_inject_inject

Native hotplug
pcie_abp

Pass DSDT
(avoid rom
size limit)

PV pci bus numbering
Pass hint for pci bus number

chipset abstraction(i440fx)
64bit BAR
Multi pci bus init
DSDT loading
MCFG
Q35 support

Seabios

Hot plug function

Function	Supported?
Attention Button	yes
Power Controller	No
MRL Sensor	No
Attention Indicator	Yes
Power Indicator	Yes
Hot-Plug Surprise	Yes
EMI	Yes

Merged
Under review
To be posted

New chipset emulator(Q35 based)

- Why new chipset?
  - Current QEmu chipset is
    - I440FX/PIIX: 10+ years since its release
    - Flat PCI bus(single PCI bus)
  - Discard legacy compatibility
    - It's very difficult to test various legacy OSes
    - Only for modern OSes
- Introduce new functionality with new chipset
  - Q35(MCH/ICH9) chipset based emulator
  - PCI Express
  - Multi pci bus(PCI-to-PCI bridge, pci express port switch)
  - It lacks iommu/graphics emulation so it should be called P45?

Current status

QEmu

Items	Status
64bit BAR	Merged
PCI Bridge lib	Merged to PCI branch
PCI Bus reset	Under review
MMCONFIG(PCI layer)	Merged
PCIe port switch Including native hotplug AER error injection	To be posted
DSDT overriding	posted(to be resend)
Q35 Chipset	To be posted
PV PCI bus numbering	To be posted

Seabios

Items	Status
64bit BAR	Merged
Multi pci bus	Merged
Chipset abstraction	Merged
DSDT overriding	Under review
MCFG	Under review
Q35	To be posted
Q35 DSDT	To be posted
PV pci bus numbering	To be posted

VGABios

Items	Status
VBE	Waiting Gerd's patch

Future Work

- Upstream merge
- PCI express native device assignment
  - PCI express specific configuration registers should be virtualized
    - Device serial number cap, VSEC...
  - AER(Advanced Error Report)
    - Catch the error in host.
      - Currently Linux AER port driver does only printk(). Poll errors from targeted devices.
    - inject errors from host to guest OS for RAS.
  - Native Power management
  - VC(Virtual channel)
  - Assigning bus hierarchy tree

Future Work(cont.)

- PCI Express device emulator?
  - Currently pcie port switch only
  - Interesting express native device?
    - IGB? IGBVF? IXGB? IXGBVF?
- Possibly, VT-d/IOMMU shadow paging for nested virtualization …
  - Qemu iommu emulation is coming. So device assignment version would be wanted.

Summary

- PCI Express is useful even in virtualized environment
- Q35 new chipset patch enables QEmu to support PCI Express
- It benefits all qemu derivatives, including KVM.

Thank you

Questions?