VFIO-based Mediated Pass Through – KVMGT as an example

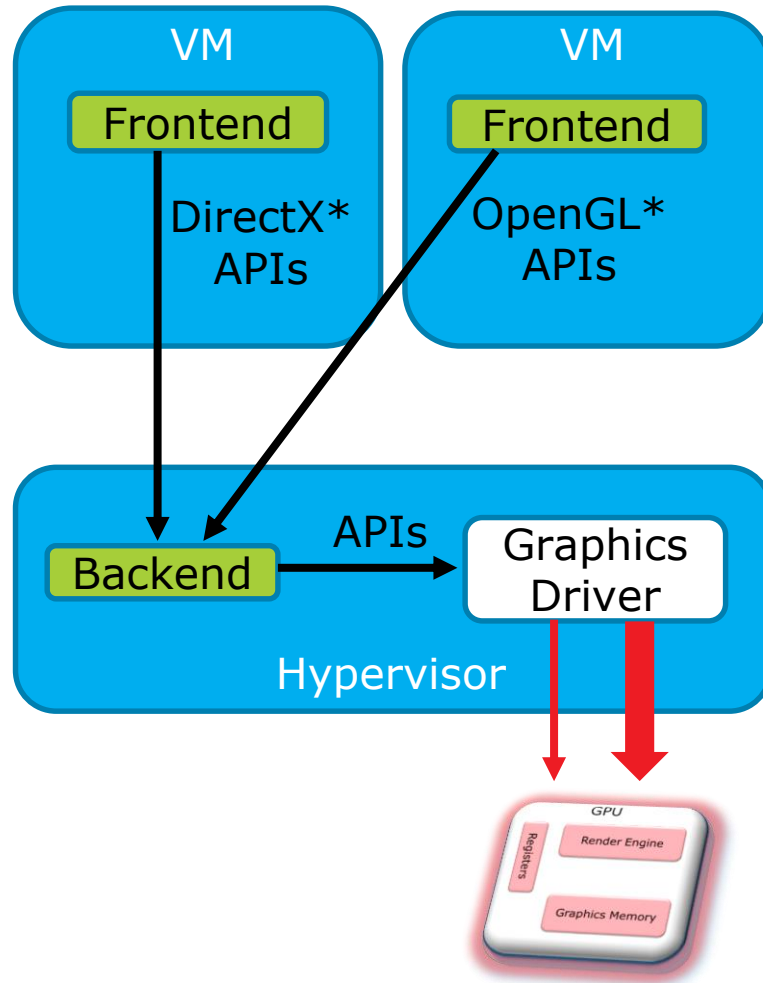Jike Song <jike.song@intel.com>

Agenda

❑ Graphics Virtualization: KVMGT and MPT

❑ VFIO

❑ VFIO-based KVMGT

Intel OTC

Graphics Virtualization – Before MPT: API Forwarding
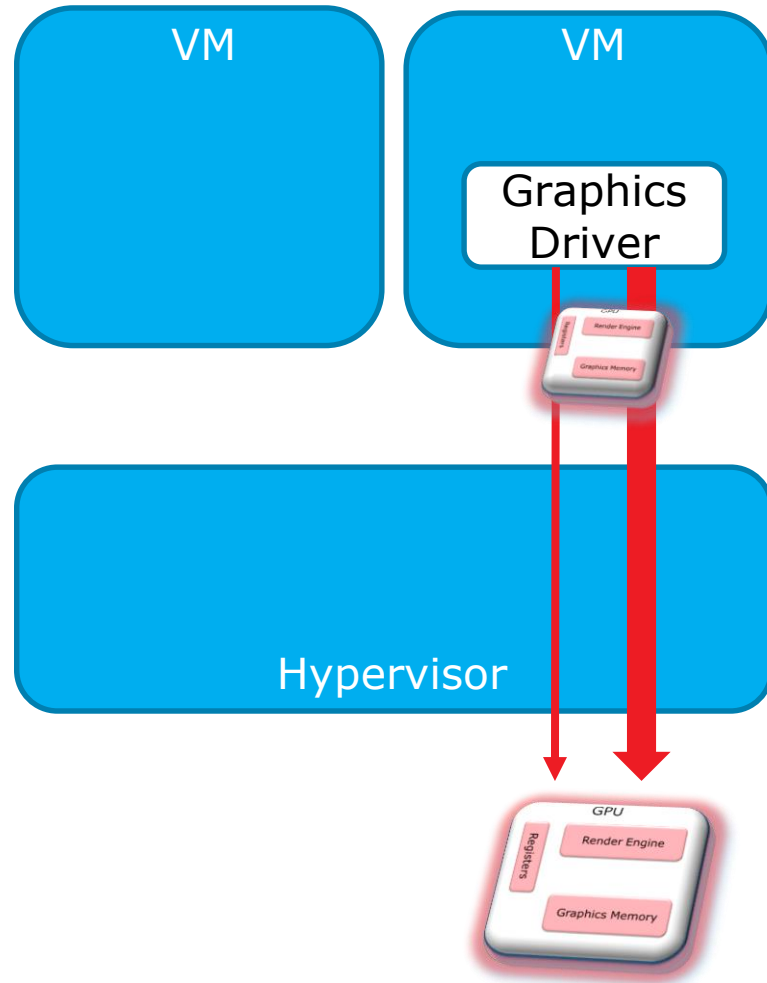

Pros

- Performance
- Scalability

Cons

- Lagging features
- Incompatible APIs
- Maintenance burden

Intel OTC

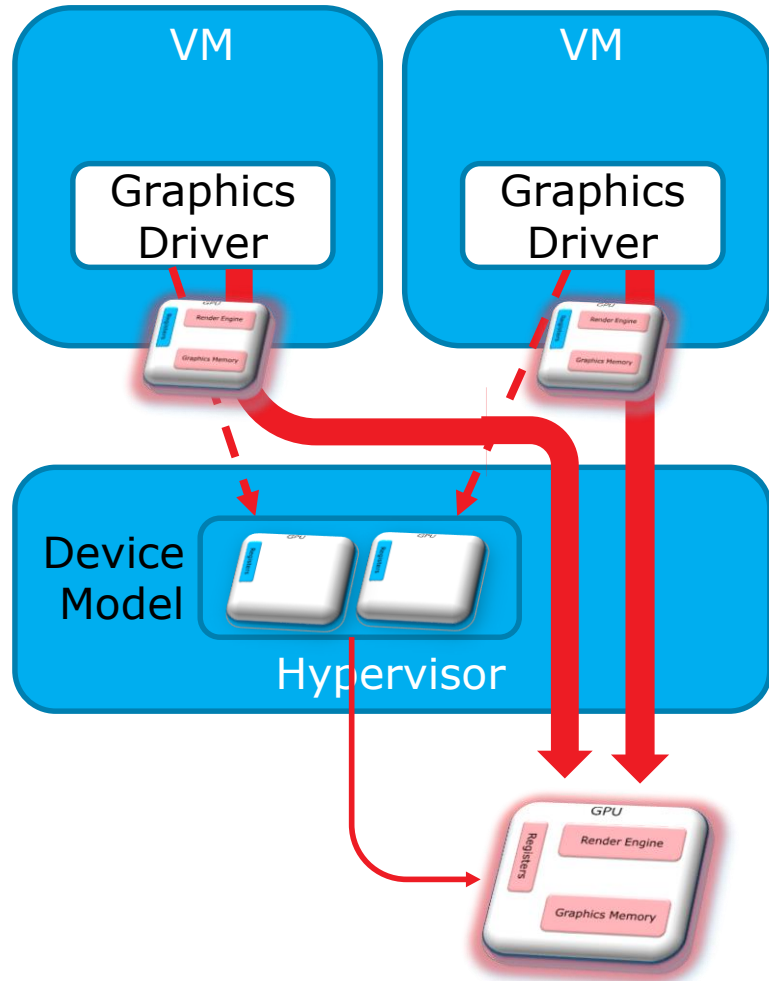# Graphics Virtualization – Before MPT: direct Pass-Through

VM

VM

Graphics Driver

Hypervisor

GPU

Registers

Render Engine

Graphics Memory

Pros

- Performance
- Full features

Cons

- No or limited sharing

Intel OTC

intel®

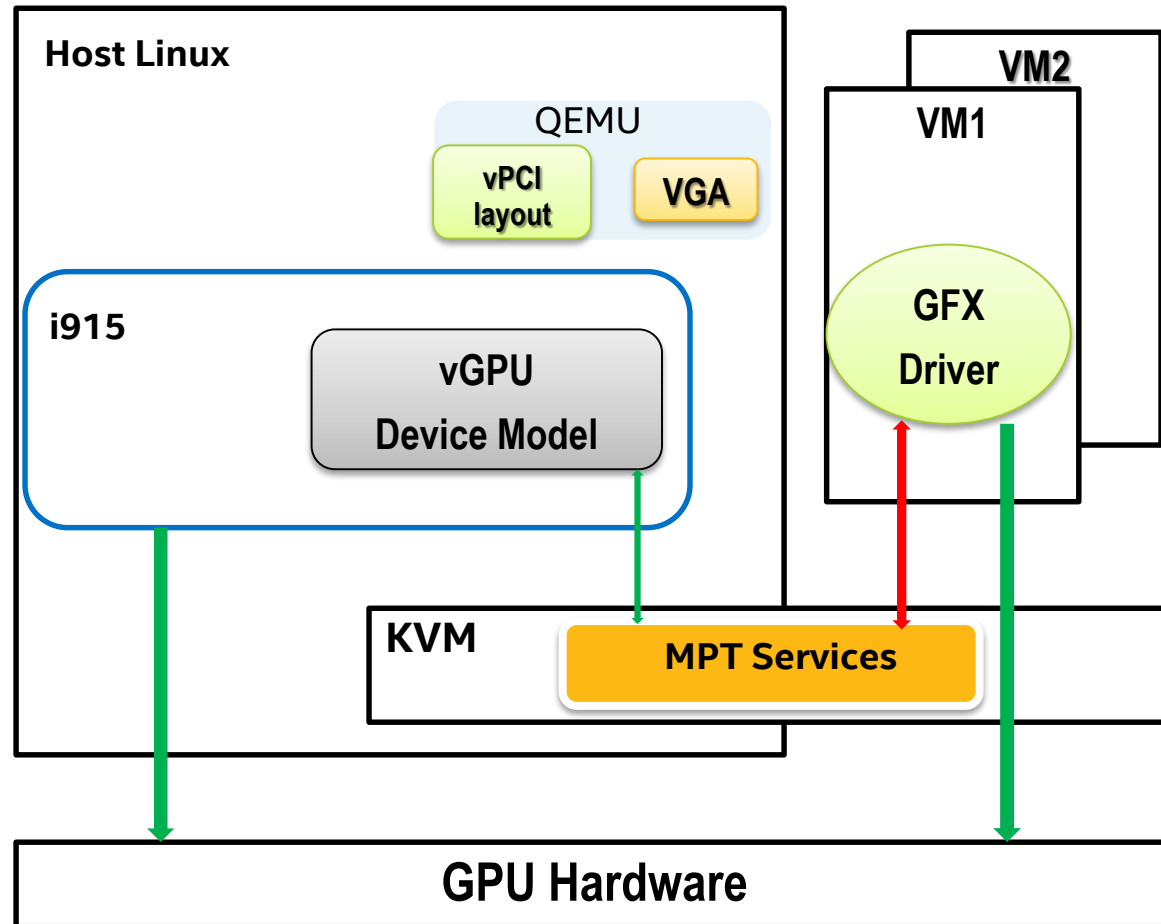# Graphics Virtualization – MPT: Mediated Pass-Through

Pros

- Performance
- Full feature
- Scalability

Cons

- Vendor specific

Intel OTC

Graphics Virtualization – MPT based KVMGT

Pros

- Full Feature
- Performance
- Scalability

Cons

- Touched a lot: Kernel, KVM, i915, QEMU, SeaBIOS ...

Intel OTC

Agenda

❑ Graphics Virtualization: KVMGT and MPT

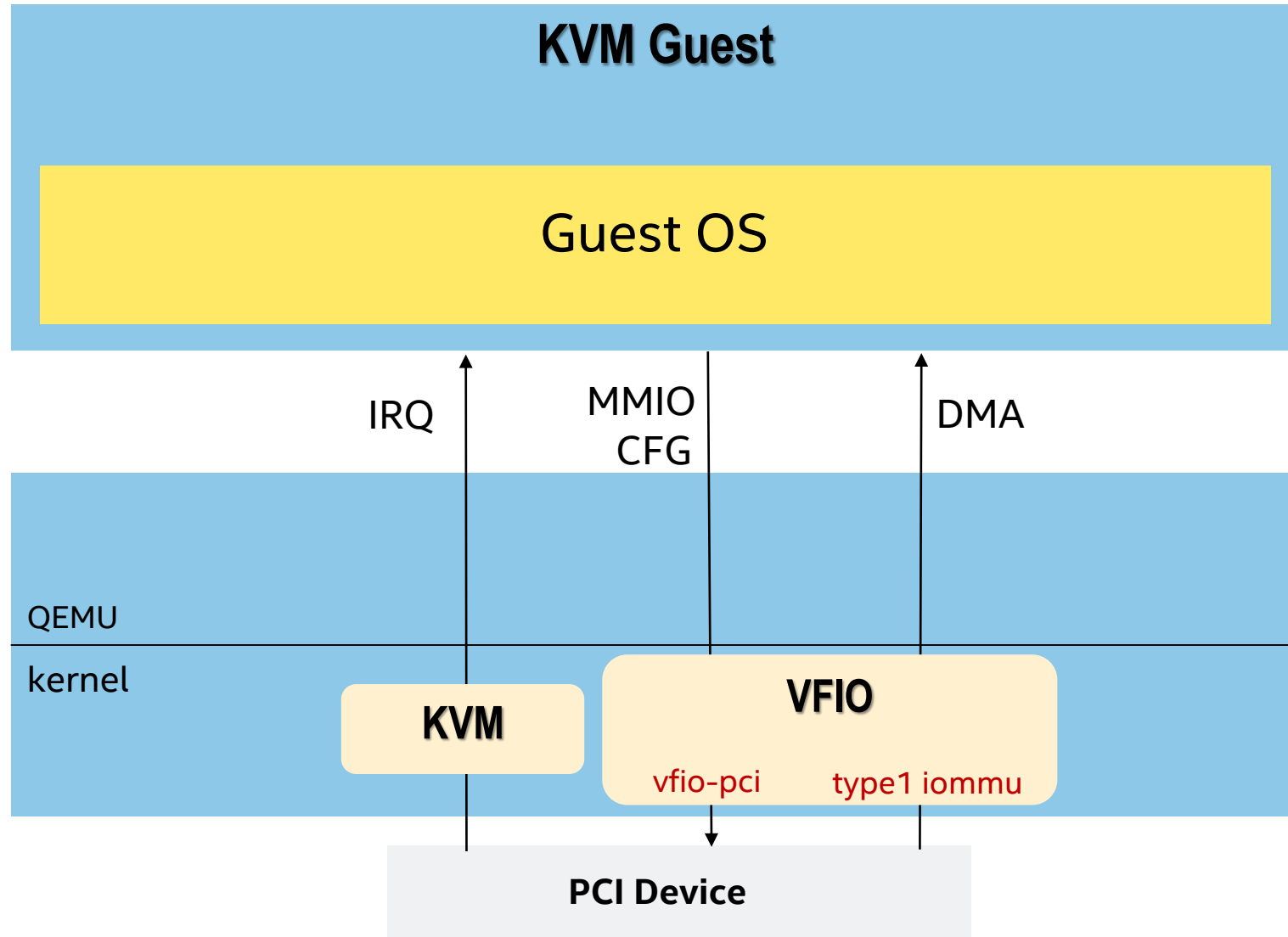❑ VFIO

❑ VFIO-based KVMGT

Intel OTC

VFIO – Virtual Function I/O

❑ By Alex Williamson @ Redhat

❑ Used for: device Pass-Through in KVM

❑ Used for: Userspace Drivers

❑ Modular Bus drivers, Modular IOMMU backends

➢ Available Bus drivers: PCI, platform

➢ Available IOMMU backends: type1, SPARR

VFIO – PCI device Pass-Through to KVM Guest

Intel OTC

VFIO – PCI device Pass-Through to KVM Guest

❑ A PCI Device or VF consists of:
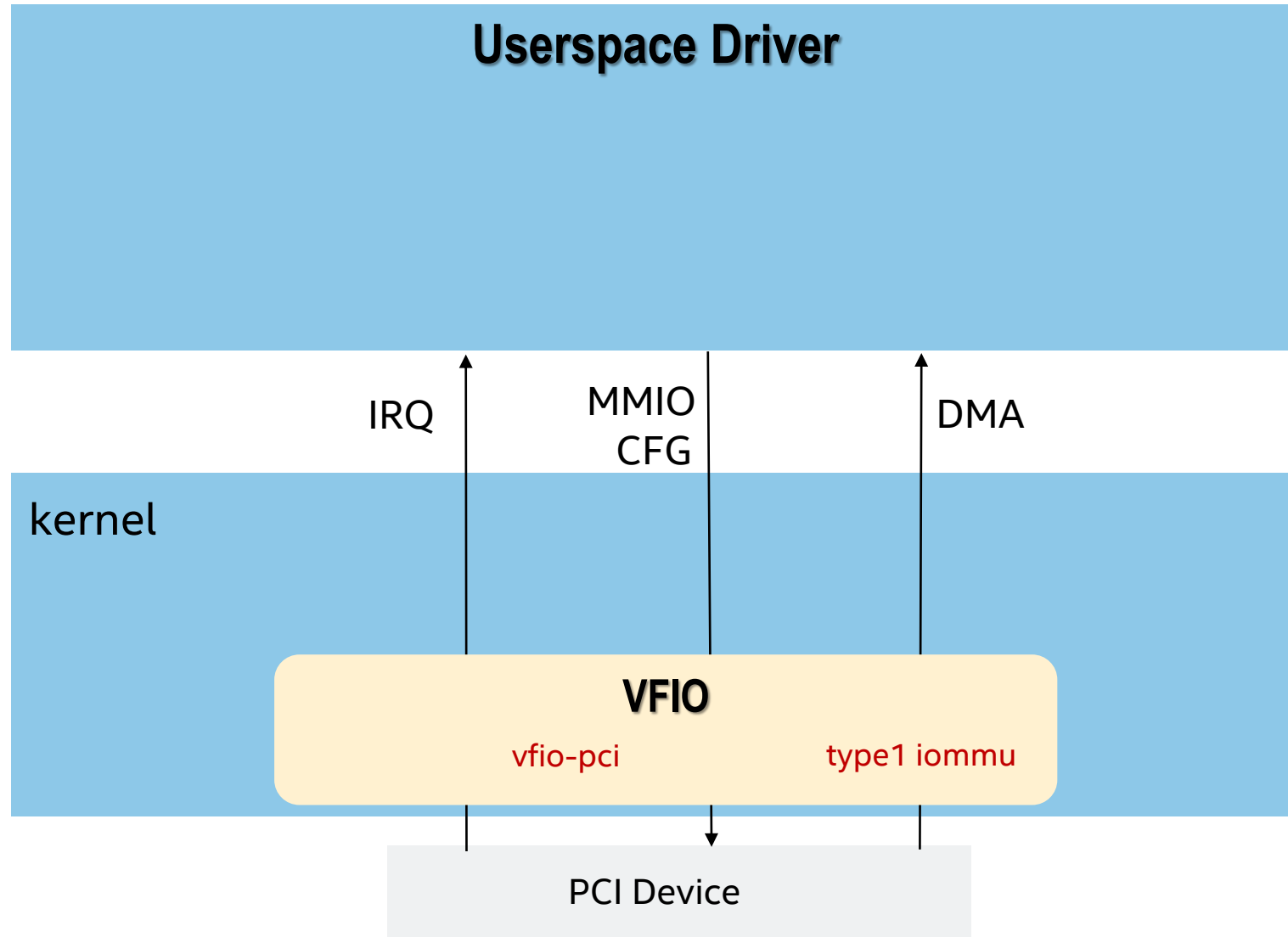
  ➤ PCI Configuration Registers

  ➤ MMIO Registers

  ➤ INTx/MSI/MSI-X IRQ

  ➤ DMA

❑ VFIO passthrough it by:

  ➤ vfio_pci bus driver

    ✓ PCI CFG : proxying the access

    ✓ MMIO : mmap to QEMU, thereby to guest

    ✓ IRQ : irqfd to QEMU, ioctl to KVM & inject to guest

  ➤ Type1 IOMMU backend

    ✓ DMA: pin & map GPA(Guest Physical Address) to HPA(Host Physical Address)

Intel OTC

VFIO – PCI device Userspace Driver

Userspace Driver

IRQ   MMIO
       CFG   DMA

kernel

VFIO

vfio-pci   type1 iommu

PCI Device

Intel OTC

VFIO – PCI device Userspace Driver

❑VFIO enables userspace driver by:

➢vfio_pci bus driver

✓PCI CFG : porxying the access

✓MMIO : mmap to userspace

✓IRQ : irqfd to userspace

➢Type1 IOMMU backend

✓DMA: pin & map userspace virtual addres to physical address

Intel OTC

(intel)

Agenda

❑ Graphics Virtualization: KVMGT and MPT

❑ VFIO

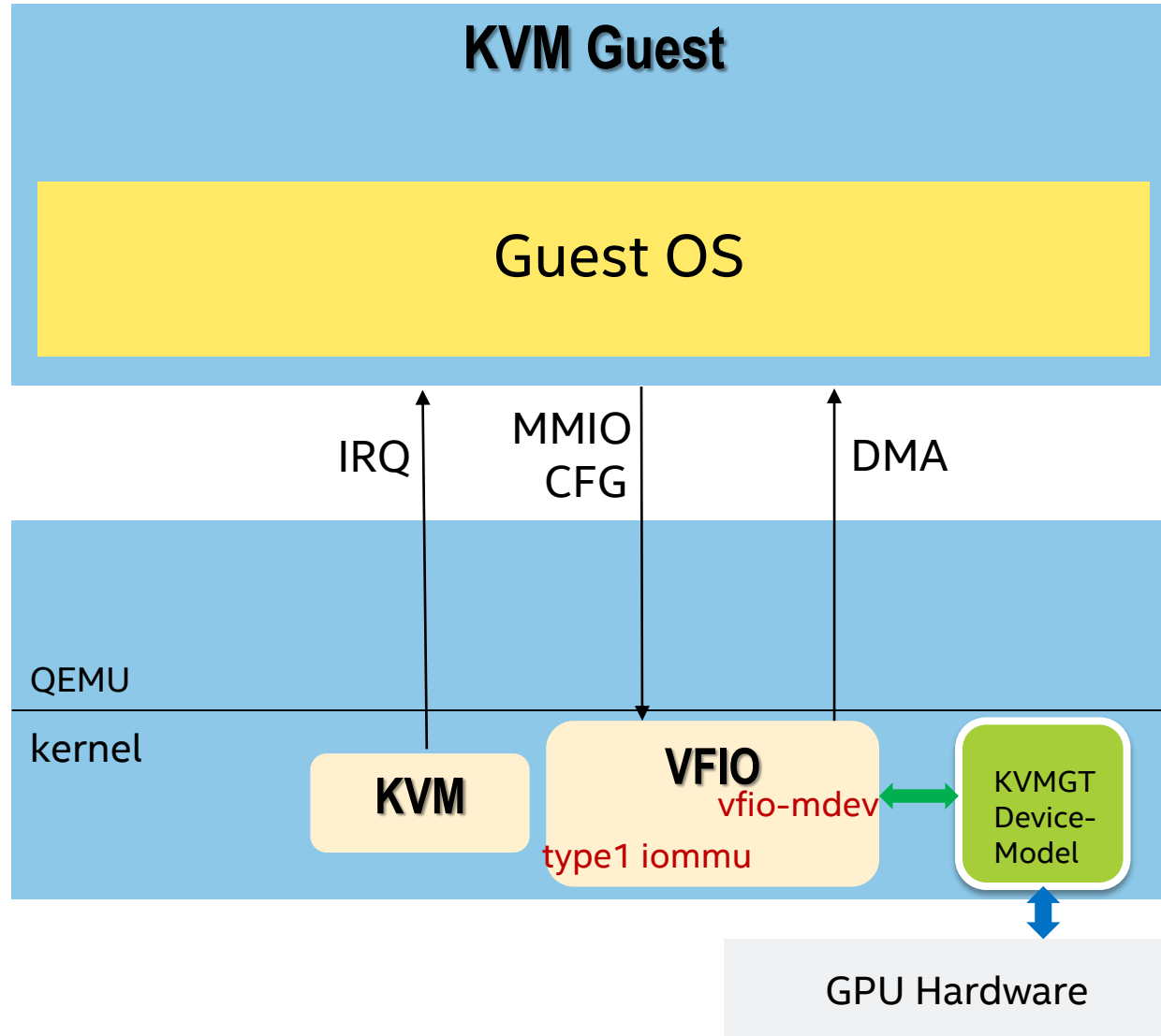❑ VFIO-based KVMGT

Intel OTC

VFIO Based KVMGT – the VFIO MDEV support

❑ Framework first implemented by Nvidia

❑ Upstreaming in progress

❑ New Bus driver for Mediated Device

  ➢ vfio-mdev

  ➢ flexible configuration: trapping or passing-through

  ➢ Capable of being compatible with the existing userspace API for PCIDEV

  ➢ Yet not PCI-specific

❑ Extended type1 IOMMU backend

  ➢ Pin guest pages on-demand

  ➢ Without hardware IOMMU dependency

❑ Multiple Usage

  ➢ vGPU Solution : Nvidia, Intel

  ➢ CCW Pass-Through : IBM

  ➢ Probably other mediated devices in the near future

Intel OTC

(intel)

VFIO Based KVMGT – the new KVMGT

KVM Guest

Guest OS

IRQ

MMIO
CFG

DMA

QEMU

kernel

KVM

VFIO
vfio-mdev
type1 iommu

KVMGT
Device-
Model

GPU Hardware

Pros compared with old KVMGT

- API compatibility with vfio-pci and all vGPU vendors
- No QEMU/SeaBIOS changes

Cons compared with old KVMGT

- More difficult to support primary GPU mode

(intel)

Thank you!

Questions?

Intel OTC