A New Chipset For Qemu - Intel's Q35

Jason Baron
jbaron@redhat.com
November 7th, 2012
http://people.redhat.com/~jbaron/q35/

Agenda

- Introduction to Q35
- PCIe
- Differences between I440FX/PIIX4 and Q35
- Current status
- How to try this at home
- Todo
- Discussion

Why do we need a new chipset?

- We want PCIe, PCIe, PCIe...
- Currently we use I440FX/PIIX4 (sort of)
- 'sort of' goes a long way...10+ years
- So we don't get backed into a corner
- Better experience
- Easier to add chipsets in the future
- I440FX/PIIX4 has been hardened for quite some time

Why Q35?

- Isaku Yamahata said 'I have a real machine'
- Lots of devices already – USB, sound, bridges, AHCI
- Lots of docs (Trust me)
- Has PCIe
- Emulation can be hard – lots of work already on it (Isaku Yamaha, Jan Kiszka)

What is Q35?

- Intel chipset released September 2007
- North Bridge: MCH
- South Bridge: ICH9

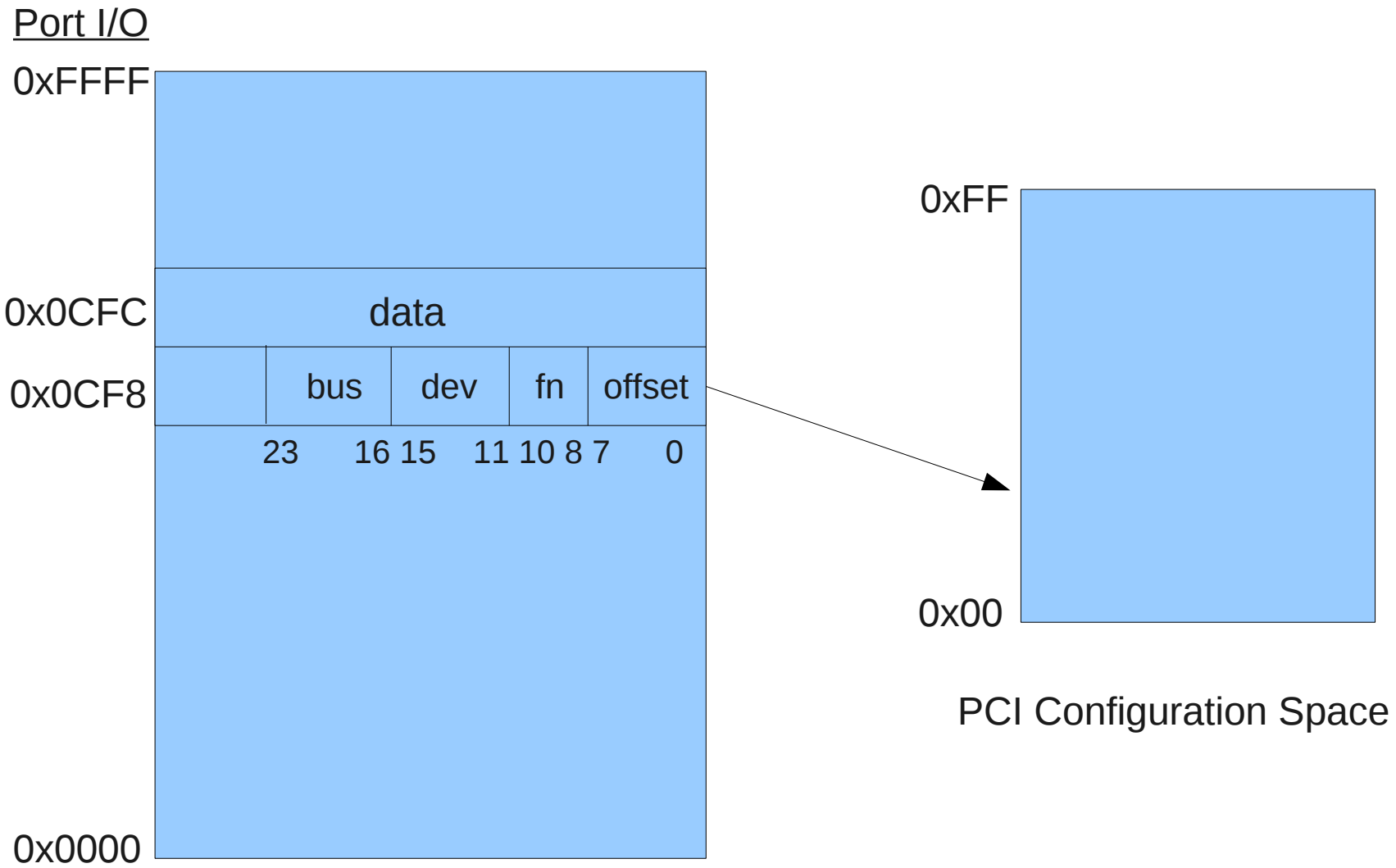
What isn't Q35?

- Isn't the latest Intel chipset (Yes, I know)

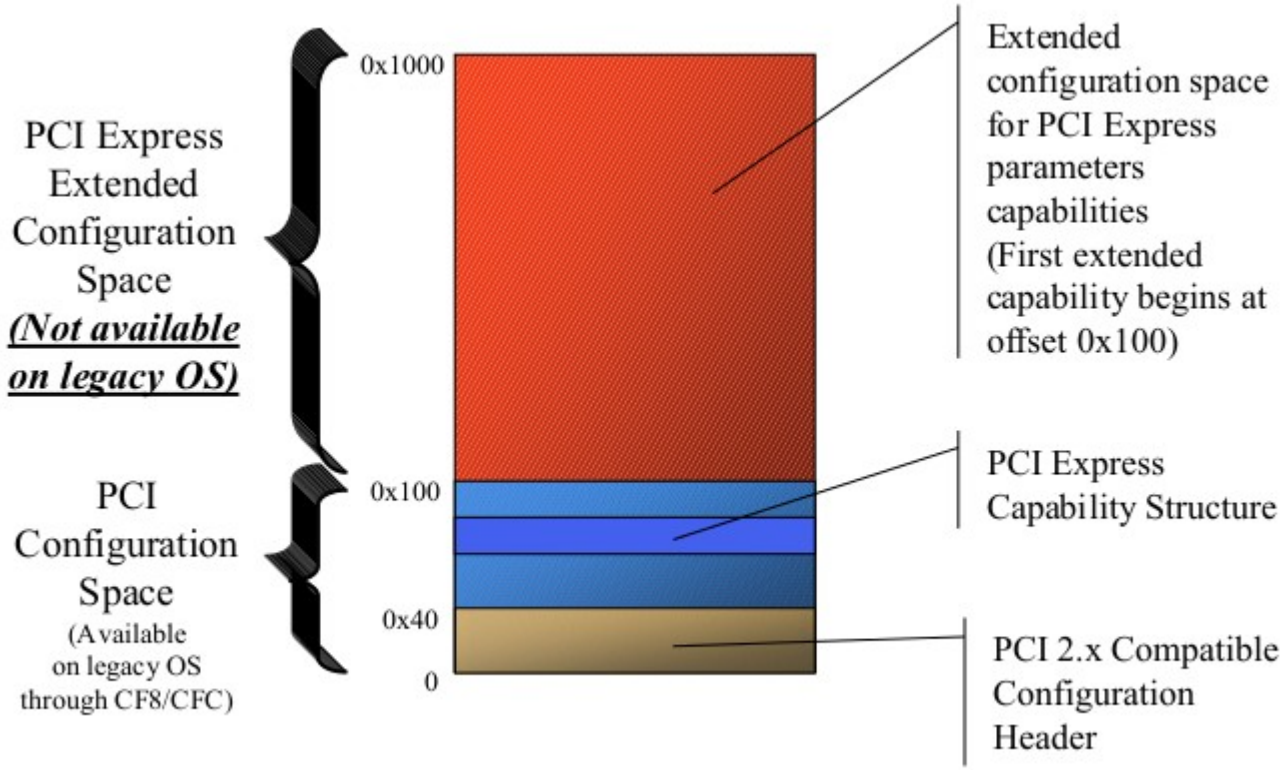
What is PCIe?

- Introduced by Intel/Dell/HP/IBM 2004 designed to replace PCI, PCI-X, and AGP
- Point-to-point topology
- PCIe AER
- PCIe Hotplug
- Backwards compatible with PCI
- Some drivers are PCIe specific
- Extended configuration space

PCI Configuration Space Access



PCIe Configuration Space

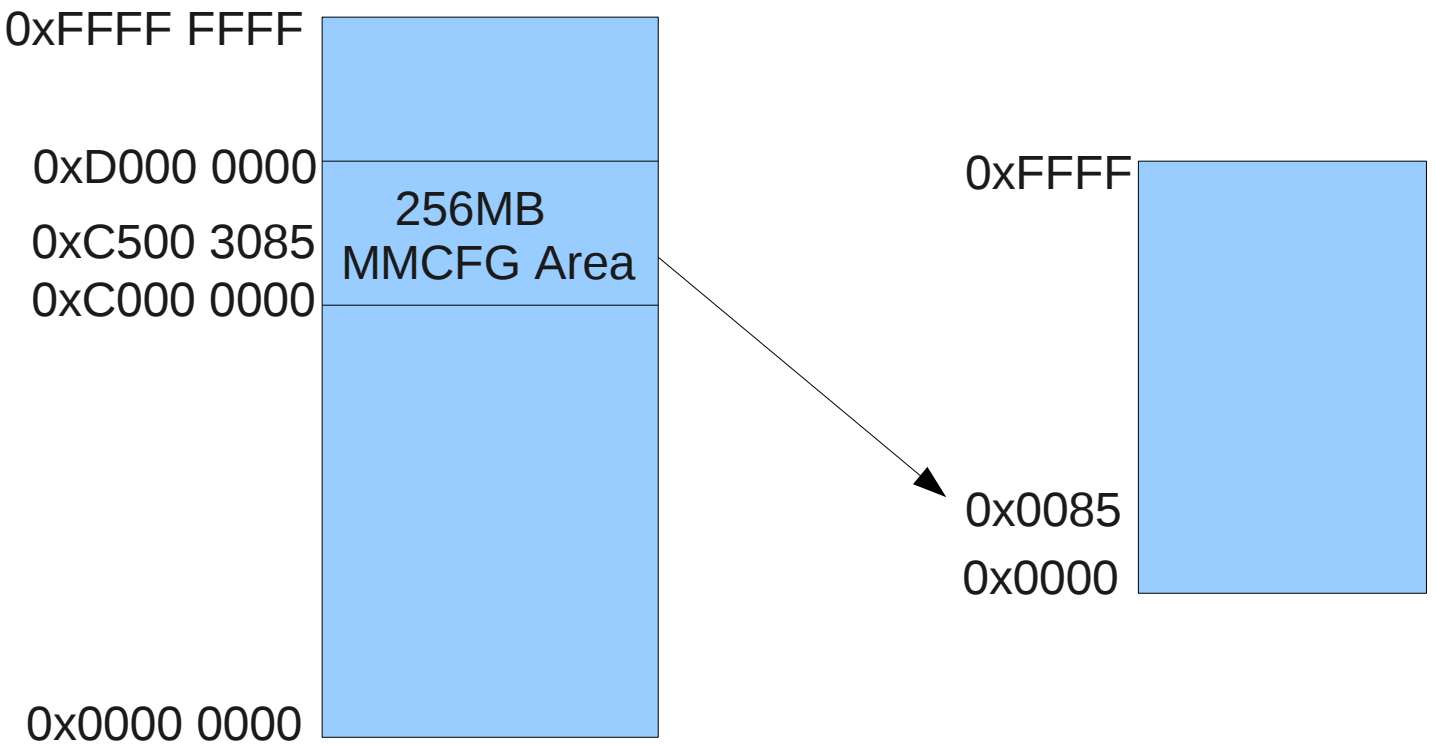


PCIe MMCONFIG

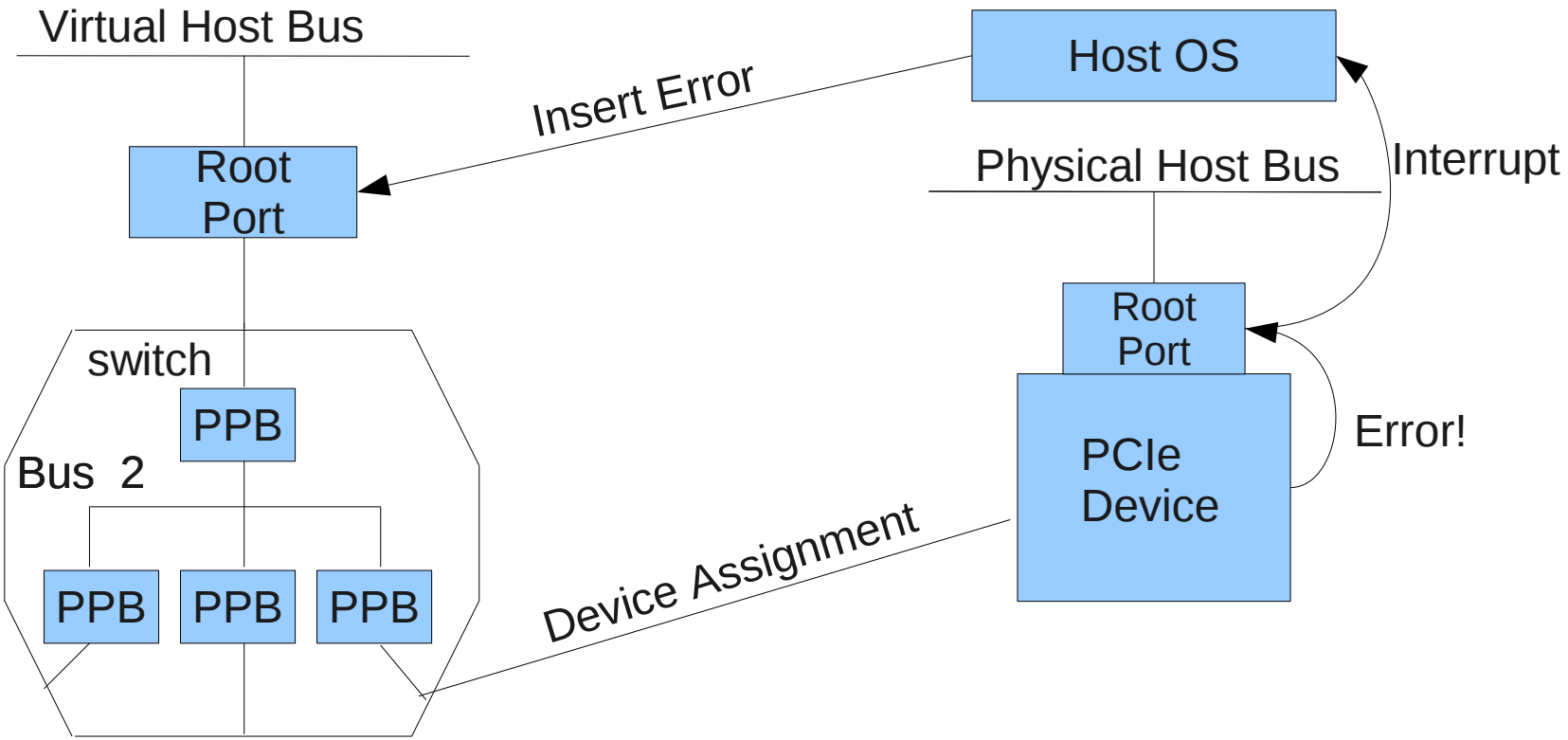
256 buses * 32 devices * 8 functions * 4096 registers = 256MB required

Let's access: Bus 5, Device 0, Function 3, Register offset 85

5 * 0x100000 + 0 * 0x8000 + 3 * 0x1000 + 0x85 = 5003085



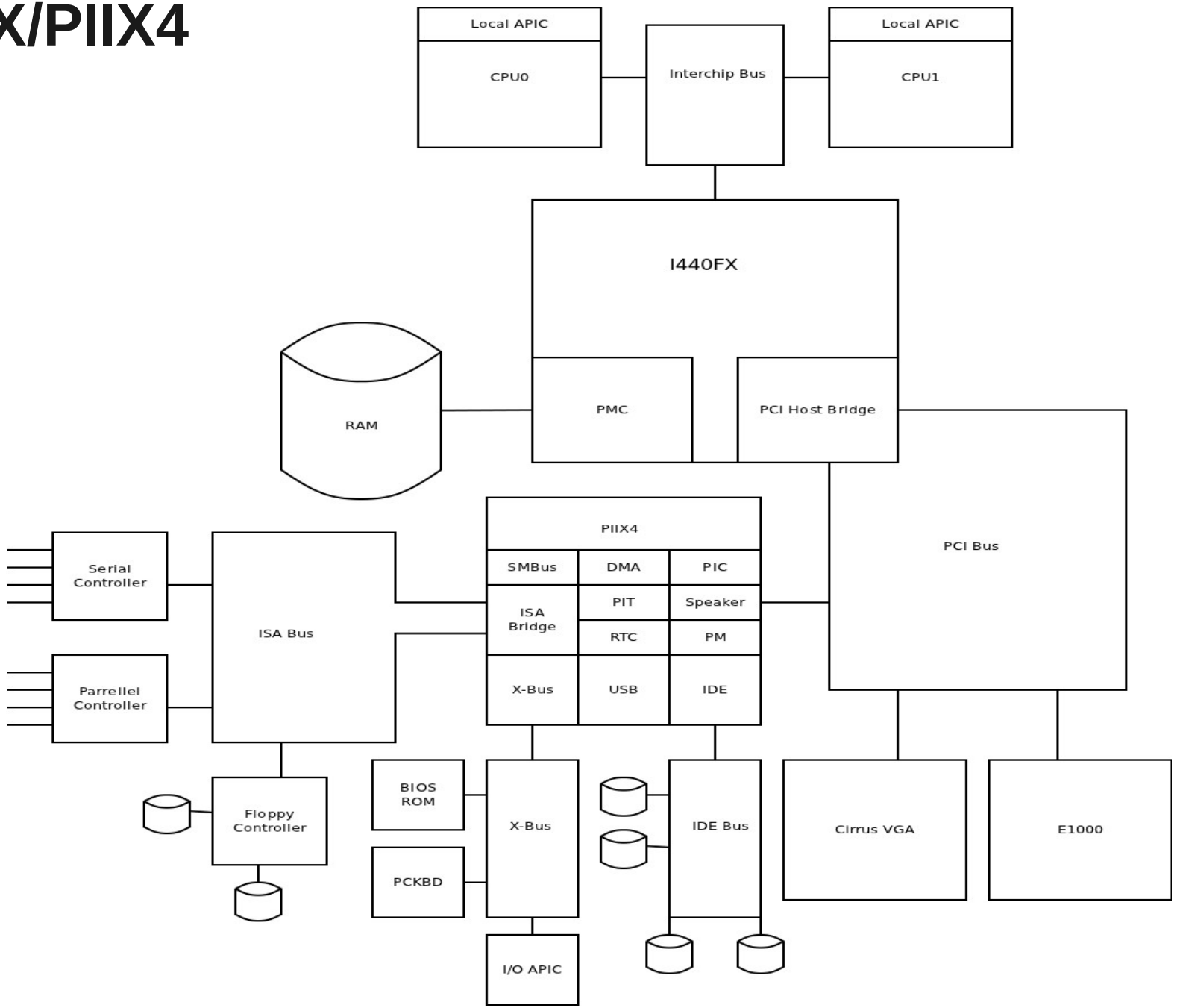
PCIe AER


Virtual Host Bus
Root Port
switch
PPB
Bus 2
PPB
PPB
PPB
Insert Error
Device Assignment
Host OS
Physical Host Bus
Interrupt
Root Port
PCIe Device
Error!

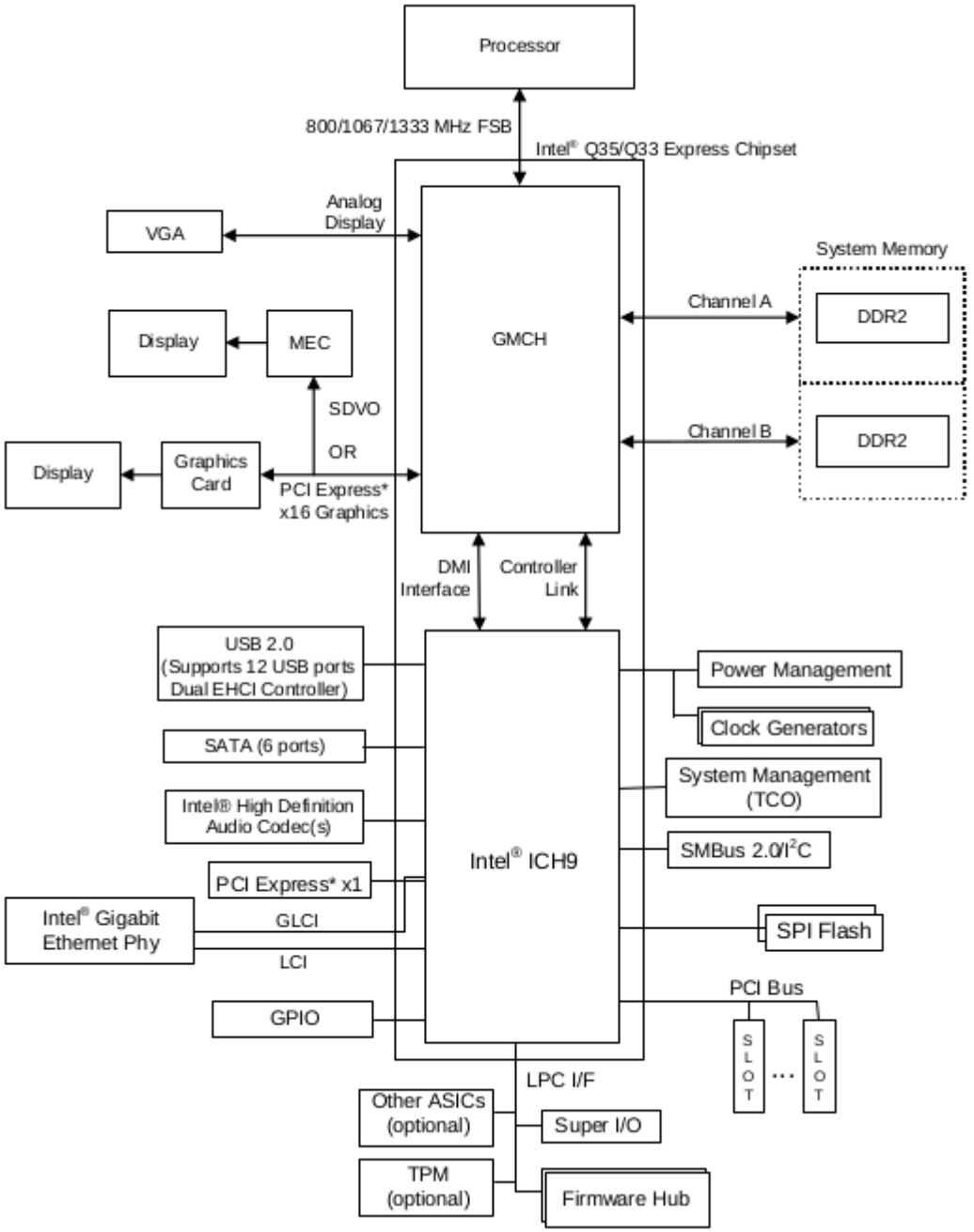
Topology of I440FX/PIIX4 Vs. Q35

- Q35 has IOMMU
- Q35 has PCIe
- Q35 has Super I/O chip with LPC interconnect
- Q35 has 12 USB ports
- Q35 SATA vs. PATA

I440FX/PIIX4



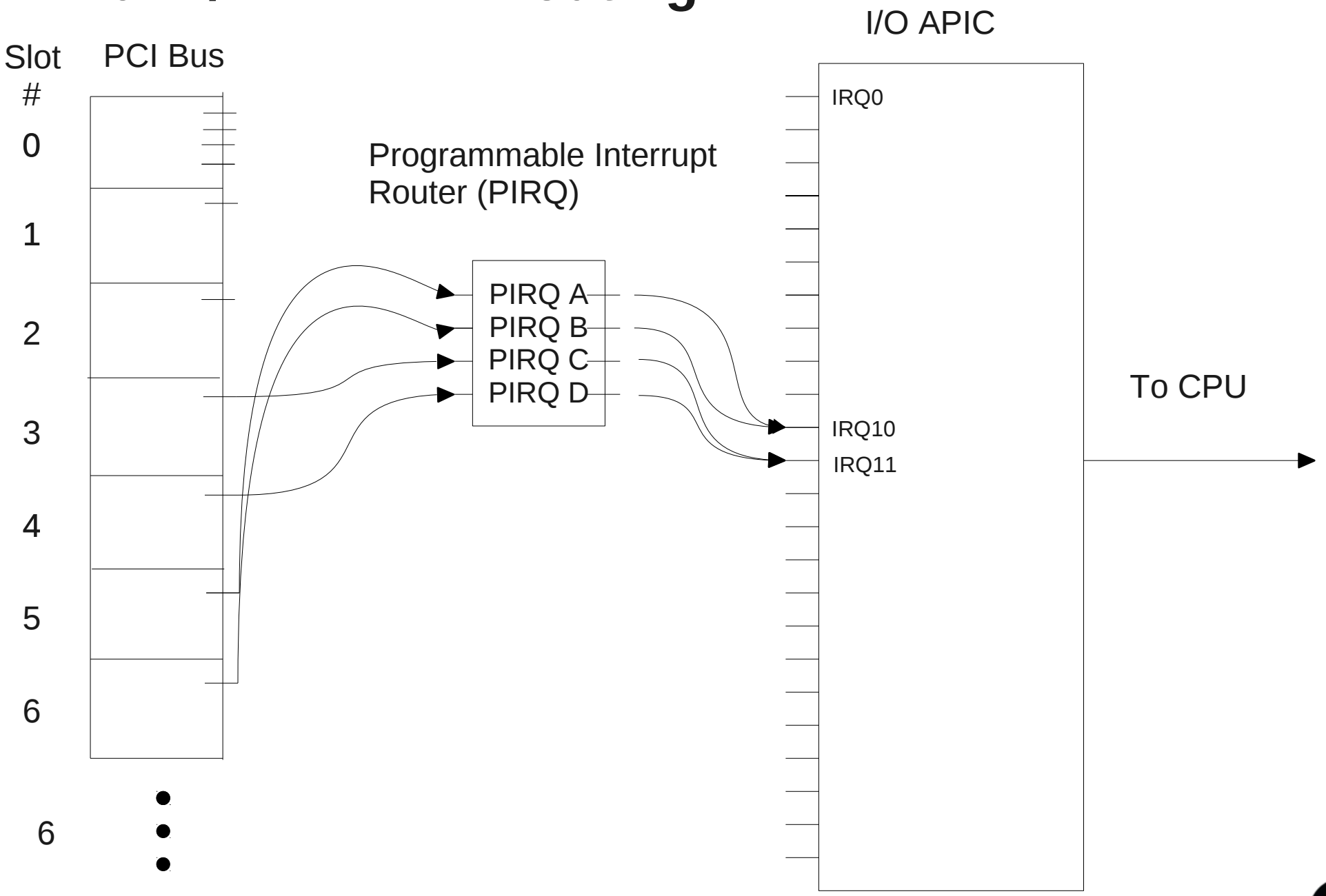
Q35 Topology



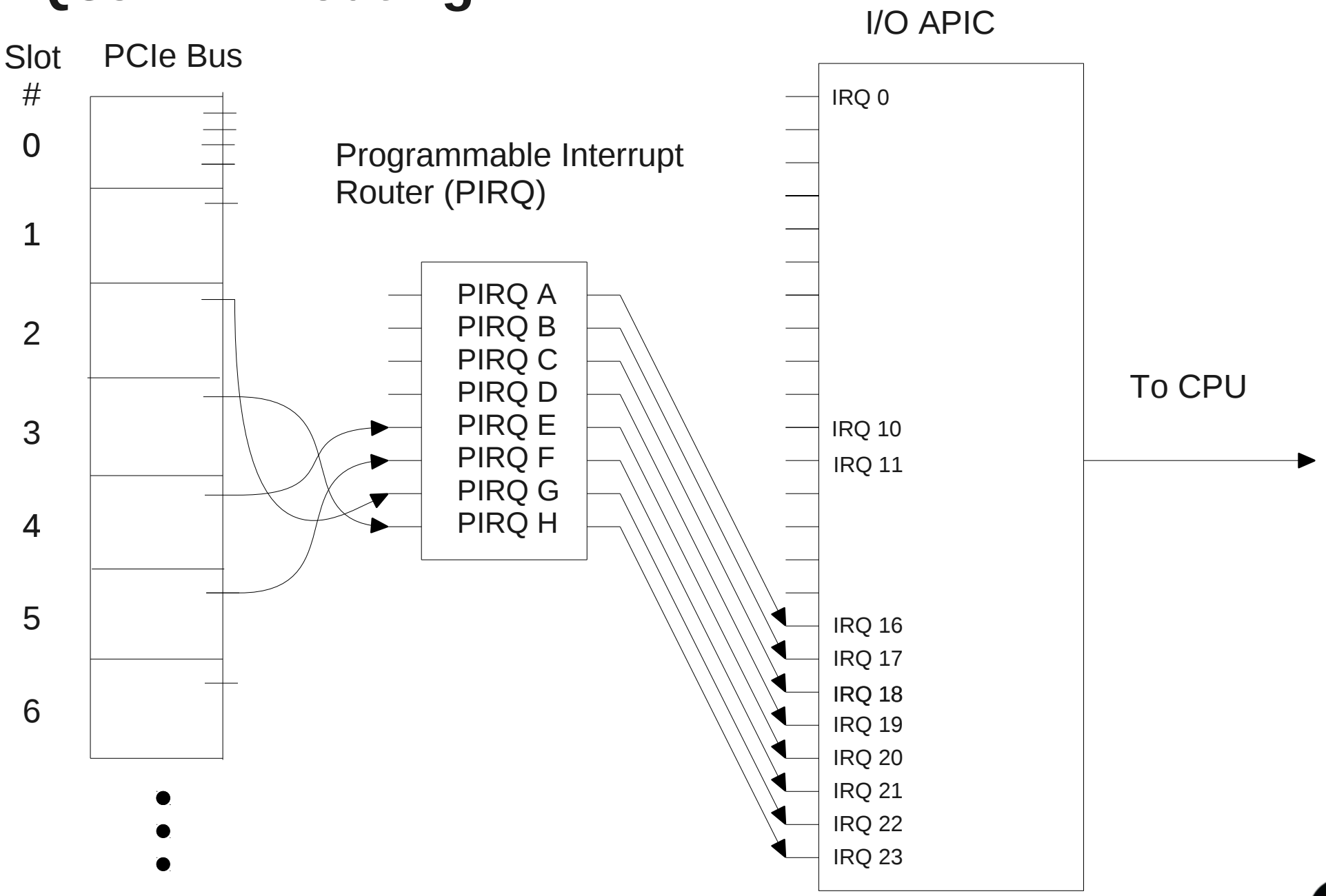
IRQ Routing I440FX/PIIX4 Vs. Q35

- Q35 PIRQ has 8 pins - PIRQA-H
- Q35 has two modes – legacy PIC vs I/O APIC
- Q35 runs in I/O APIC mode
- Slots 0-24 are mapped to PIRQE-H round robin
- PCIe Bus to PIRQ mappings can be programmed
  - Slots 25-31
- Q35 has 8 PCI IRQ vectors available, I440FX/PIIX4 only 2

I440FX/PIIX4 INTx routing



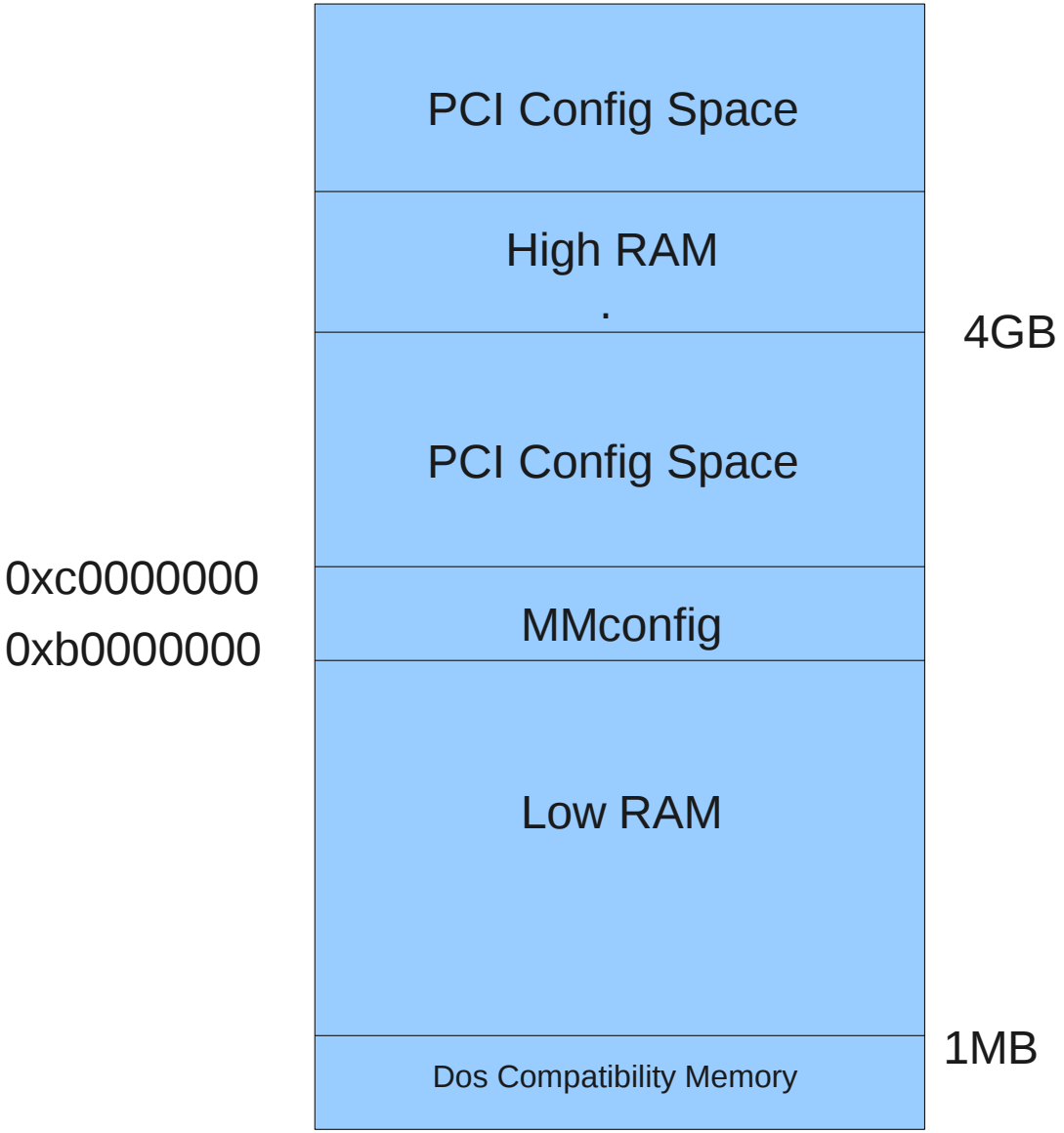
Q35 INTx routing



Physical Memory Layout

- MMConfig window
- 32-bit pci space now fixed doesn't float down
- MMConfig window a problem for a 32-bit OS

Physical Memory Layout



Physical Memory Layout – '(qemu) info mtree'

```
0000000000000000-7ffffffffffffffe (prio 0, RW): system
  0000000000000000-00000000afffffff (prio 0, RW): alias ram-below-4g @pc.ram 0000000000000000-00000000afffffff
  00000000000a0000-00000000000bffff (prio 1, RW): alias smram-region @pci 00000000000a0000-00000000000bffff
  00000000000c0000-00000000000c3fff (prio 1, R-): alias pam-rom @pc.ram 00000000000c0000-00000000000c3fff
  00000000000c4000-00000000000c7fff (prio 1, R-): alias pam-rom @pc.ram 00000000000c4000-00000000000c7fff
  00000000000c8000-00000000000cbfff (prio 1, R-): alias pam-rom @pc.ram 00000000000c8000-00000000000cbfff
  00000000000ca000-00000000000ccfff (prio 1000, RW): alias kvmvapic-rom @pc.ram 00000000000ca000-00000000000ccfff
  00000000000cc000-00000000000cffff (prio 1, R-): alias pam-rom @pc.ram 00000000000cc000-00000000000cffff
  00000000000d0000-00000000000d3fff (prio 1, RW): alias pam-ram @pc.ram 00000000000d0000-00000000000d3fff
  00000000000d4000-00000000000d7fff (prio 1, RW): alias pam-ram @pc.ram 00000000000d4000-00000000000d7fff
  00000000000d8000-00000000000dbfff (prio 1, RW): alias pam-ram @pc.ram 00000000000d8000-00000000000dbfff
  00000000000dc000-00000000000dffff (prio 1, RW): alias pam-ram @pc.ram 00000000000dc000-00000000000dffff
  00000000000e0000-00000000000e3fff (prio 1, RW): alias pam-ram @pc.ram 00000000000e0000-00000000000e3fff
  00000000000e4000-00000000000e7fff (prio 1, RW): alias pam-ram @pc.ram 00000000000e4000-00000000000e7fff
  00000000000e8000-00000000000ebfff (prio 1, RW): alias pam-ram @pc.ram 00000000000e8000-00000000000ebfff
  00000000000ec000-00000000000effff (prio 1, RW): alias pam-ram @pc.ram 00000000000ec000-00000000000effff
  00000000000f0000-00000000000fffff (prio 1, R-): alias pam-rom @pc.ram 00000000000f0000-00000000000fffff
  00000000b0000000-00000000bfffffff (prio 0, RW): pcie-mmcfg
  00000000b0000000-00000000ffffffff (prio 0, RW): alias pci-hole @pci 00000000b0000000-00000000ffffffff
  00000000fec00000-00000000fec00fff (prio 0, RW): kvm-ioapic
  00000000fed00000-00000000fed003ff (prio 0, RW): hpet
  00000000fee00000-00000000feefffff (prio 0, RW): kvm-apic-msi
  0000000100000000-000000014fffffff (prio 0, RW): alias ram-above-4g @pc.ram 00000000b0000000-00000000ffffffff
  0000000150000000-400000014fffffff (prio 0, RW): alias pci-hole64 @pci 0000000150000000-400000014fffffff
```


I440FX/PIIX4 vs. Q35 devices

- AHCI vs. Legacy IDE
- PCI addresses
- Populate slots using flags
- Default slots

Q35 Vs. I440FX/PIIX4 – 'lspci'

Q35:

```
00:00.0 Host bridge: Intel Corporation 82G33/G31/P35/P31 Express DRAM Controller
00:01.0 VGA compatible controller: Cirrus Logic GD 5446
00:02.0 Ethernet controller: Intel Corporation 82540EM Gigabit Ethernet Controller (rev 03)
00:1d.0 USB Controller: Intel Corporation 82801I (ICH9 Family) USB UHCI Controller #1 (rev 03)
00:1d.1 USB Controller: Intel Corporation 82801I (ICH9 Family) USB UHCI Controller #2 (rev 03)
00:1d.2 USB Controller: Intel Corporation 82801I (ICH9 Family) USB UHCI Controller #3 (rev 03)
00:1d.7 USB Controller: Intel Corporation 82801I (ICH9 Family) USB2 EHCI Controller #1 (rev 03)
00:1f.0 ISA bridge: Intel Corporation 82801IB (ICH9) LPC Interface Controller (rev 02)
00:1f.2 SATA controller: Intel Corporation 82801IR/IO/IH (ICH9R/DO/DH) 6 port SATA AHCI Controller (rev 02)
00:1f.3 SMBus: Intel Corporation 82801I (ICH9 Family) SMBus Controller (rev 02)
```

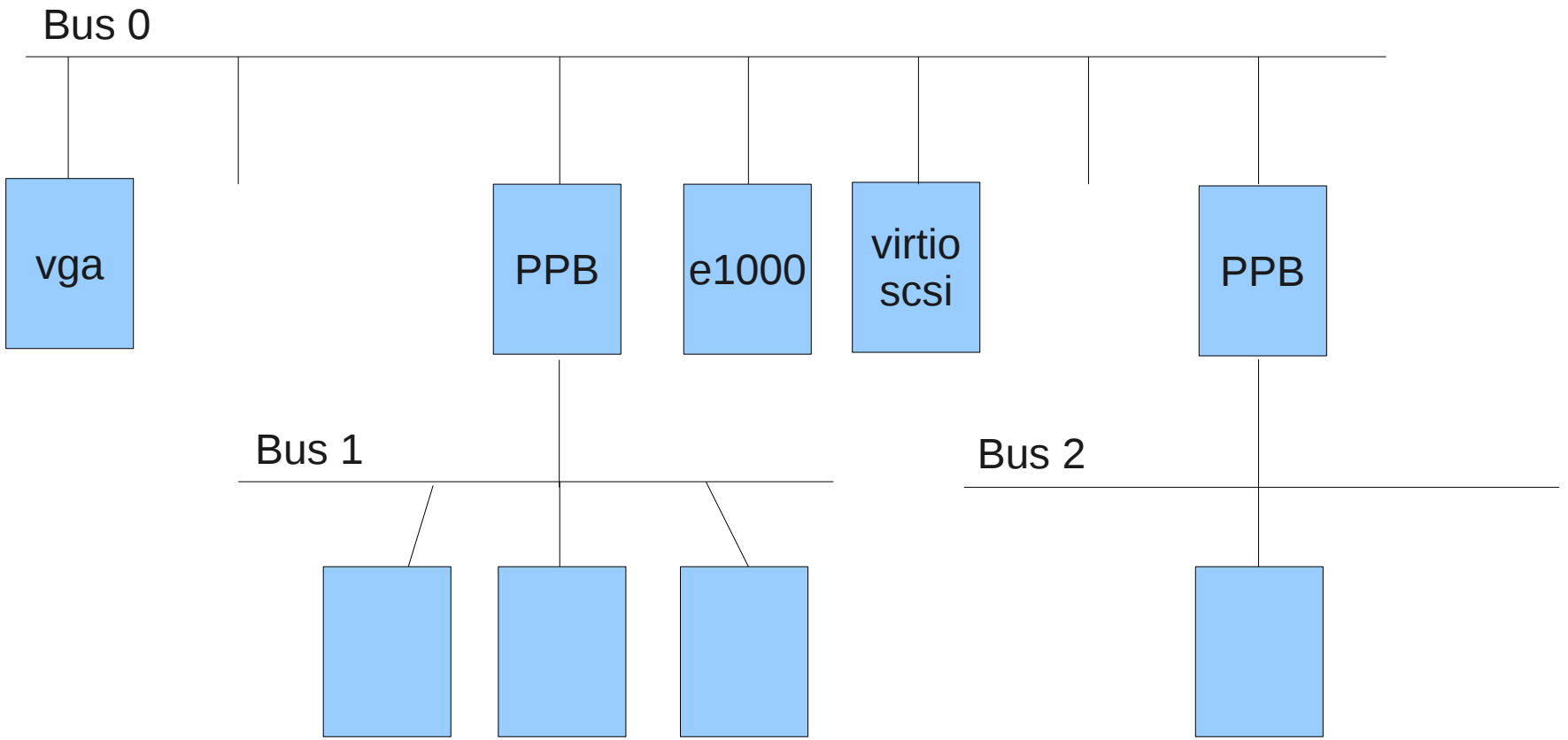
I440FX/PIIX4:

```
00:00.0 Host bridge: Intel Corporation 440FX - 82441FX PMC [Natoma] (rev 02)
00:01.0 ISA bridge: Intel Corporation 82371SB PIIX3 ISA [Natoma/Triton II]
00:01.1 IDE interface: Intel Corporation 82371SB PIIX3 IDE [Natoma/Triton II]
00:01.2 USB Controller: Intel Corporation 82371SB PIIX3 USB [Natoma/Triton II] (rev 01)
00:01.3 Bridge: Intel Corporation 82371AB/EB/MB PIIX4 ACPI (rev 03)
00:02.0 VGA compatible controller: Cirrus Logic GD 5446
00:03.0 Ethernet controller: Intel Corporation 82540EM Gigabit Ethernet Controller (rev 03)
```

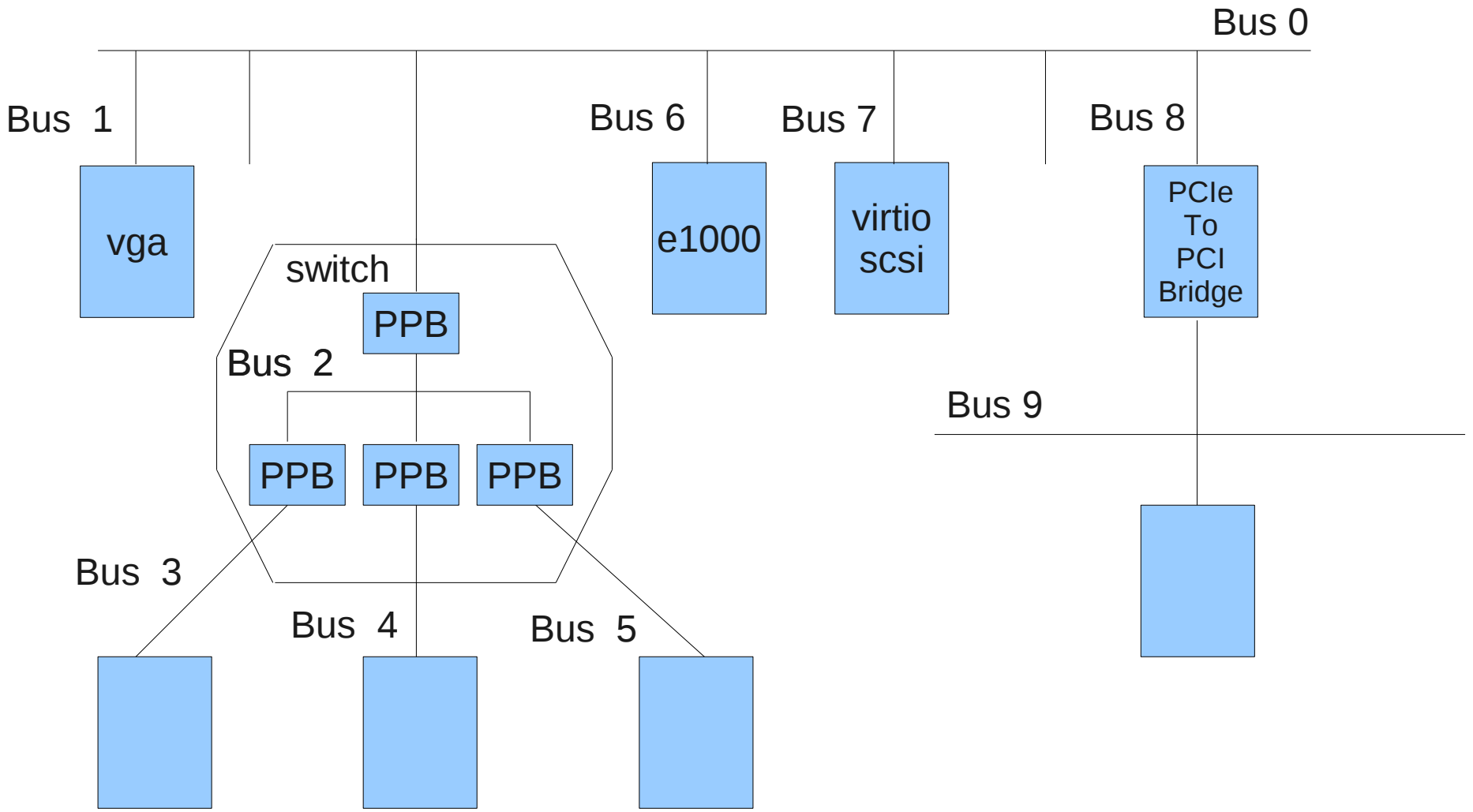

PCI Bridges (I440FX/PIIX4) Vs. PCIe Switches (Q35)

- PCI Bridge
- Root, upstream, downstream ports

PCI Topology



PCIe Topology



Current Status

- Merge Plans
- What I can do on Q35 that I can't on I440FX/PIIX4, right now?
- OS support – Fedora 16, 17, Windows XP, 7, 8, BSD, OSX (Gabriel Somlo)
- Passthrough/VFIO
- Migration testing
- Hotplug

Don't Try This At Home

- git clone git://github.com/jibaron/q35-qemu.git
- git clone git://github.com/jibaron/q35-seabios.git

```
/usr/local/bin/qemu-system-x86_64 -drive if=none,file=/home/jibaron/images/f17.img,id=disk \
       -device ide-drive,drive=disk,bus=ide.0 \
       -drive if=none,file=/home/jibaron/images/isos/Fedora-17-x86_64-DVD.iso,id=cdrom \
       -device ide-cd,drive=cdrom,bus=ide.1 -net nic -net user --enable-kvm \
       -L /home/jibaron/trees/q35-seabios/out -monitor stdio -M q35 -smp 2 -m 2G
```


Todo - AHCI

- Cleanups
- Migration
- Optimizations
- Testing! Testing! Testing!
- Command line interface
  - hda-hdd continues to create IDE disks
  - if=ide continues to create IDE disks
  - AHCI drives specified using if=none and -device (see previous slide for details)

Todo - Hotplug

- Multiple levels of PCI hotplug. Scale?
- Hotplug PCIe switches?
- Goal: no advance setup

Todo

- PCIe passthrough, VFIO
- Add AER capture and pass to guest – topologies could differ
- Windows 8 boots every other time :)
- Add OSX support
- Fill-out options -usb, etc.
- Libvirt changes?
- Create smarter INTx allocations?
- Performance testing – AHCI, INTx devices

Questions?

Thanks!