PCIexpress in MicroTCA.4 System at XFEL
150 NAT-MCH-PHYS and 50 NAT-MCH-PHYS80
PCIexpress in MicroTCA Systems

Agenda

- PCIe: Topology, device tree, root complex
- Sub-dividing a system with N root complexes
- Hot-swap
lspci –s bb:dev.func

Configuration Space

CPU

Root Complex

System Memory (DRAM)

P2P

Switch

Type 1 Headers (Virtual PCI-PCI Bridges)

Type 0 Headers

PCle Endpoint

PCle Endpoint

Endpoint
PCIexpress devices & endpoints

root@nat: #lspci

```
00:00.0 Host bridge: Intel Corporation 2nd Generation Core Processor Family DRAM Controller (rev 09)
00:01.0 PCI bridge: Intel Corporation Xeon E3-1200/2nd Generation Core Processor Family PCI Express Root Port (rev 09)
00:01.1 PCI bridge: Intel Corporation Xeon E3-1200/2nd Generation Core Processor Family PCI Express Root Port (rev 09)
00:02.0 VGA compatible controller: Intel Corporation 2nd Generation Core Processor Family Integrated Graphics Controller (rev 09)
00:16.0 Communication controller: Intel Corporation 6 Series Chipset Family MEI Controller #1 (rev 04)
00:19.0 Ethernet controller: Intel Corporation 82579LM Gigabit Network Connection (rev 04)
00:1a.0 USB Controller: Intel Corporation 6 Series Chipset Family USB Enhanced Host Controller #2 (rev 04)
00:1c.0 PCI bridge: Intel Corporation 6 Series Chipset Family PCI Express Root Port 1 (rev b4)
00:1d.0 USB Controller: Intel Corporation 6 Series Chipset Family USB Enhanced Host Controller #1 (rev 04)
00:1f.0 ISA bridge: Intel Corporation QM67 Express Chipset Family LPC Controller (rev 04)
00:1f.2 IDE interface: Intel Corporation 6 Series Chipset Family 4 port SATA IDE Controller (rev 04)
00:1f.3 SMBus: Intel Corporation 6 Series Chipset Family SMBus Controller (rev 04)
00:1f.5 IDE interface: Intel Corporation 6 Series Chipset Family 2 port SATA IDE Controller (rev 04)
01:00.0 PCI bridge: Integrated Device Technology Inc. Device 808f
01:00.2 System peripheral: Integrated Device Technology Inc. Device 808f
02:08.0 PCI bridge: Integrated Device Technology Inc. Device 808f
03:00.0 PCI bridge: PLX Technology Inc. Device 8748 (rev ba)
04:00.0 PCI bridge: PLX Technology Inc. Device 8748 (rev ba)
```
Star Topology
Any Slot can be Root Complex
Application Example: MPI Greifswald

12 * 4 PCIe lanes 2 * 16 PCIe lanes
Welcome to the HTML based NAT-MCH configuration tool.

Setup Functions:

Base Configuration:
- Changes Base Configuration.

Age Time:
- MAC Table setup: set the aging of the MAC Table Entries.

Port VLAN:
- Port based VLAN setup and port enable/disable.

802.1Q VLAN:
- 802.1Q VLAN setup.

802.1X:
- 802.1X security setup.

802.1p:
- 802.1p Quality of Service setup.

Port Mirroring:
- Mirroring of the inbound and outbound traffic on a port

Jumbo frames:
- Support of the Jumbo frames on a port

Link Aggregation:
- Support of up to four Link Aggregation groups

Rapid Spanning:
- Support of the Rapid Spanning Tree by 10GbE-Switch

IGMP Snooping:
- Support of the IGMP Snooping by 10GbE-Switch (FM4000 only)

Link Status:
- Show the current status of the Ethernet links

Counter Statistic:
- Show the counter statistic of the Ethernet switch

Maintenance Functions:

Script Management:
- Backup/Restore settings to/from flash memory or file.

Board Information:
- Provides hardware information of this NAT-MCH.

System Information:
- Collects hardware information of this system.

Reboot NAT-MCH:
- Allows rebooting of the NAT-MCH over the Web-Interface.

Update MCH:
- Allows updating several components over the Web-Interface.

Change/Reset Password:
- Allows changing or resetting of the MCH Password over the Web-Interface.

N.A.T. Webpage:
- Opens the N.A.T. webpage in a new browser window.

Home:
- Shows this page.
PEX8748
Multi-Host Configuration: up to 6 Cluster
NAT-MCH by N.A.T.

**PCIe Virtual Switch configuration**

Select Host AMC (Upstream) for each virtual switch that shall be enabled first.
Select Host AMC (Non-Transparent Upstream) for each virtual switch that shall be enabled afterwards.
Select which AMCs shall be connected to each virtual switch as downstream in the end.

<table>
<thead>
<tr>
<th>Virtual Switch</th>
<th>Upstream AMC</th>
<th>NT-Upstream AMC</th>
<th>A1</th>
<th>A2</th>
<th>A3</th>
<th>A4</th>
<th>A5</th>
<th>A6</th>
<th>A7</th>
</tr>
</thead>
<tbody>
<tr>
<td>none</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Virtual Switch 0</td>
<td>PIM</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Virtual Switch 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Virtual Switch 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Virtual Switch 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Virtual Switch 4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Virtual Switch 5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: You need to click apply before you can save your changes to EEPROM.
PCIexpress Requirements
Hot Plug and not Hotswap Solution

- Hardware Requirements
- MCH requirements
- Operating system ie Linux requirements

![Diagram showing the relationship between Operating System, Device Driver, PCIexpress Switch of MCH, HotPlug Controller, AMC with PCIe, MMC, and IPMI.]

User Interface
Hot Removal of AMC with PCIe

Hotswap handle pulled out: standard procedure

- AMC IO
- MMC
- MCMC
- MCH
- HotPlug Controller
- Linux on AMC Root Complex
- Device Driver
- HotPlug Driver
- Udev event

- IPMI HotSwap Open
- IPMI Blue LED Blink
- Attention Button Pressed
- Wait for PCIe PwrEN
- PCIe PwrEN
- IRQ Attention Button
- Check Slot status
- Set Slot Control Registers
- Wait for 5 sec
- Call driver remove
- Back from remove
- HotPlug Service
- Remove Payload
- Red LED on Green LED off
- Blue LED on

Attention Button

Check Slot status

Set Slot Control Registers

Wait for 5 sec
Testing HotPlug

Prepare **MCH** and OS

- The PCI Express Hot-Swap in MTCA depends on:
  - Linux OS Hotplug Driver (pciehp)
  - PCI Express Switch with the Hot-Plug controller on the MCH

- We have to ...

- Configure the MCH (PCI Express Switch)
  - Enable Hot-Plug Controller on the PCI Express Switch

![Configuration table](image)
Thank you very much!

Questions?

Vollrath Dirksen
vollrath@nateurope.com

N.A.T. GmbH
Konrad-Zuse-Platz 9
53227 Bonn, Germany

www.nateurope.com

MTCA.4 Training:

techlab.desy.de/support/training

www.nateurope.com/services/support.html