PCI Express Basics

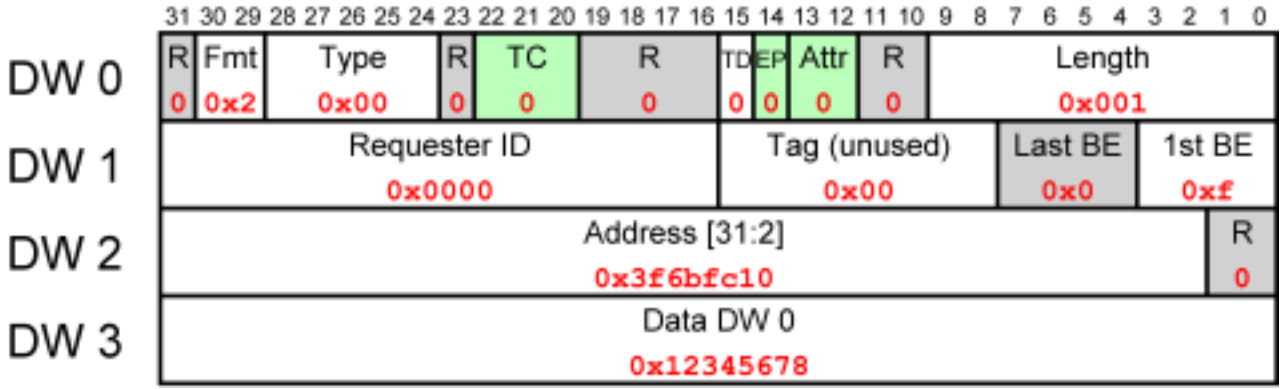
Dolphin Interconnect Solutions
Roy Nordstrøm

PCI Express

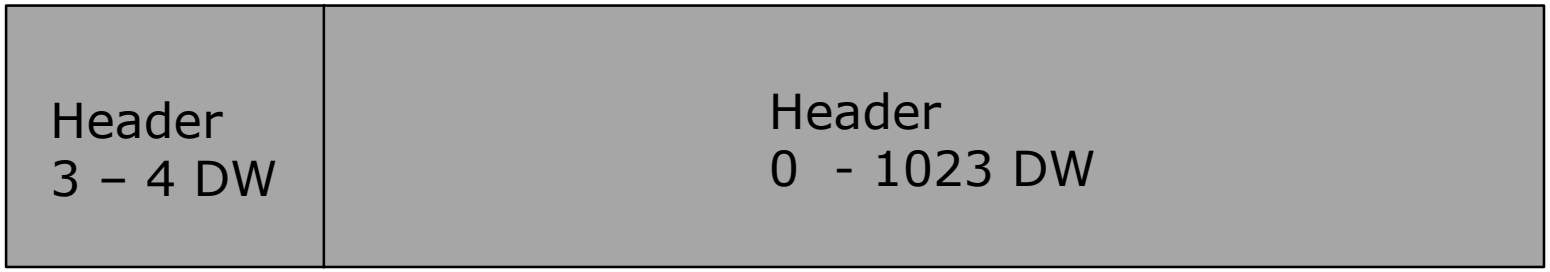
- PCI Express is a packet based protocol
- A high-speed hardware interface for connecting peripheral devices.
- Provides a high-bandwidth scalable solution for reliable data transport
- PCI Express is a serial point-to-point interconnect between two devices
- Scalable performance based on number of signal lanes implemented on the PCI Express interconnect

PCI Express Packets

PCIe Header



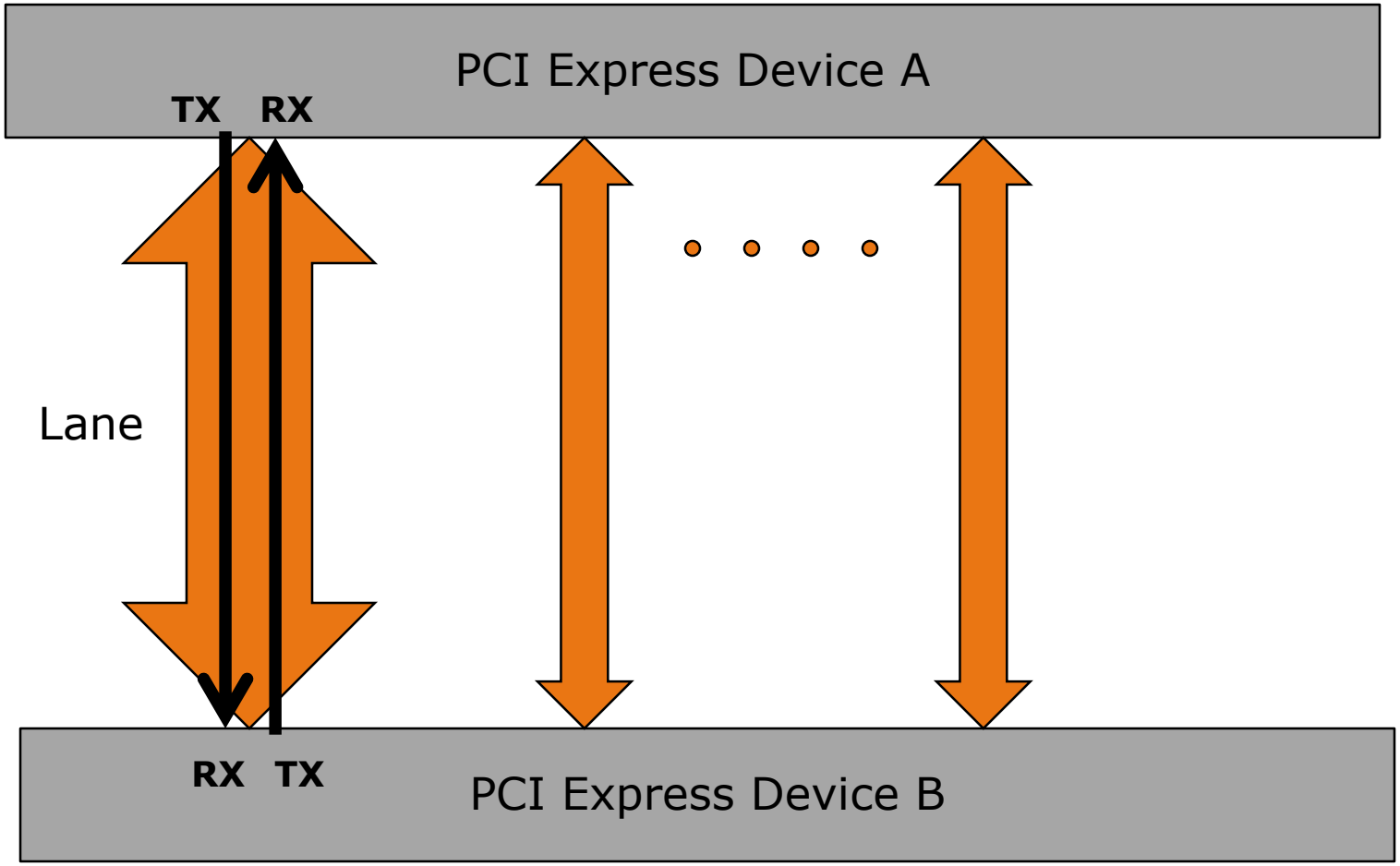
PCIe Packet



PCI Express

- PCI vs PCIe
  - Peripheral Component Interconnect (PCI)
  - PCI is original bus based interconnect
  - PCI Express is high-speed serial connection
- PCIe Link
  - Point to point communication channel between two PCIe ports
- Link width
  - Each lane of a PCIe connection contains two pairs of wires one to send and one to receive.
  - Lanes can be x1, x2, x4, x8, x16 (link width)

PCI Express



PCI Express - Transfer rates

- The specified transfer rate of Gen 1 PCI Express systems is 2.5 GT/s
- Gen 2 PCI Express 5.0 GT/s
- Gen 3 PCI Express systems, 8.0 GT/s
- Gen 4 PCI Express systems, 16.0 GT/s
  - Hardware available in 2018

PCI Express

- Link speed
  - Link speed can be
    - Gen1, 2.5 GT/s
    - Gen2, 5 GT/s
    - Gen3, 8 GT/s
    - Gen4, 16 GT/s (Next generation)
- Ports
  - A combination of lanes into a physical connection
  - Ports can be
    - transparent (**TB**)
    - Non-transparent (**NTB**)
- Link
  - A single port or multiple ports combined into one link
    - E.g., x8 or x16

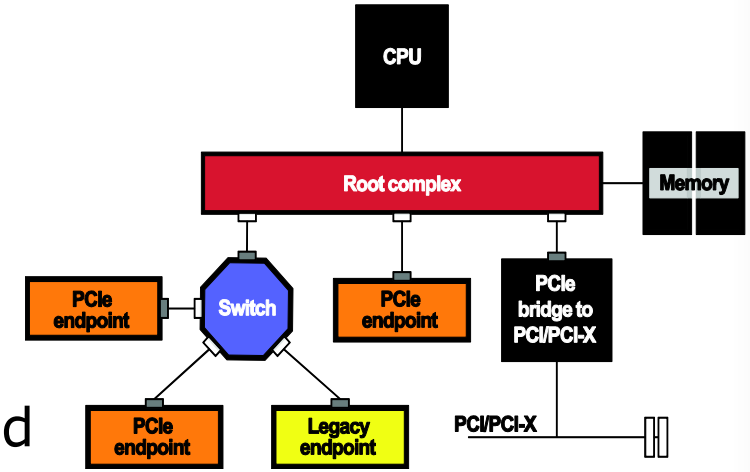
PCI Express Roadmap and performance

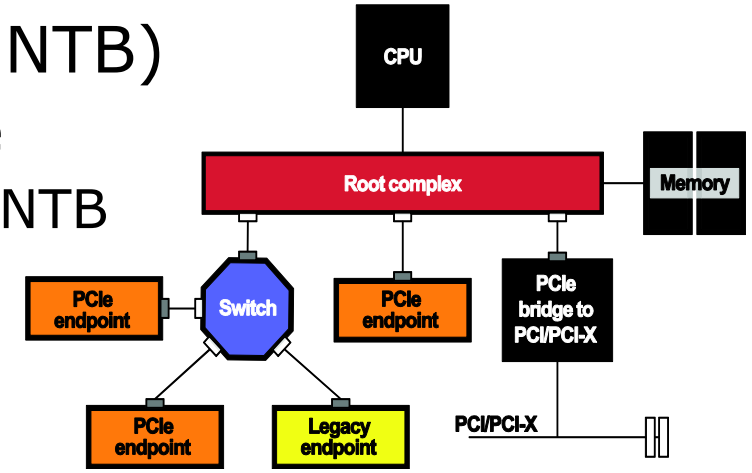
- PCI Express performance chart

	Gen1	Gen2	Gen3	Gen4
Transfer rate				
x1	2.5 GT/s	5 GT/s	8 GT/s	16 GT/s
x2	5 GT/s	10 GT/s	16 GT/s	32 GT/s
x4	10 GT/s	20 GT/s	32 GT/s	64 GT/s
x8	20 GT/s	40 GT/s	64 GT/s	128 GT/s
x16	40 GT/s	80 GT/s	128 GT/s	256 GT/s

PCI Express

- Switching
  - Switch based topologies
- Transparent Bridging
  - Root complex able to see all end point in system
- Non-Transparent Bridging (NTB)
  - PCIe endpoint is a NTB bridge
  - Software used to see beyond NTB Bridge

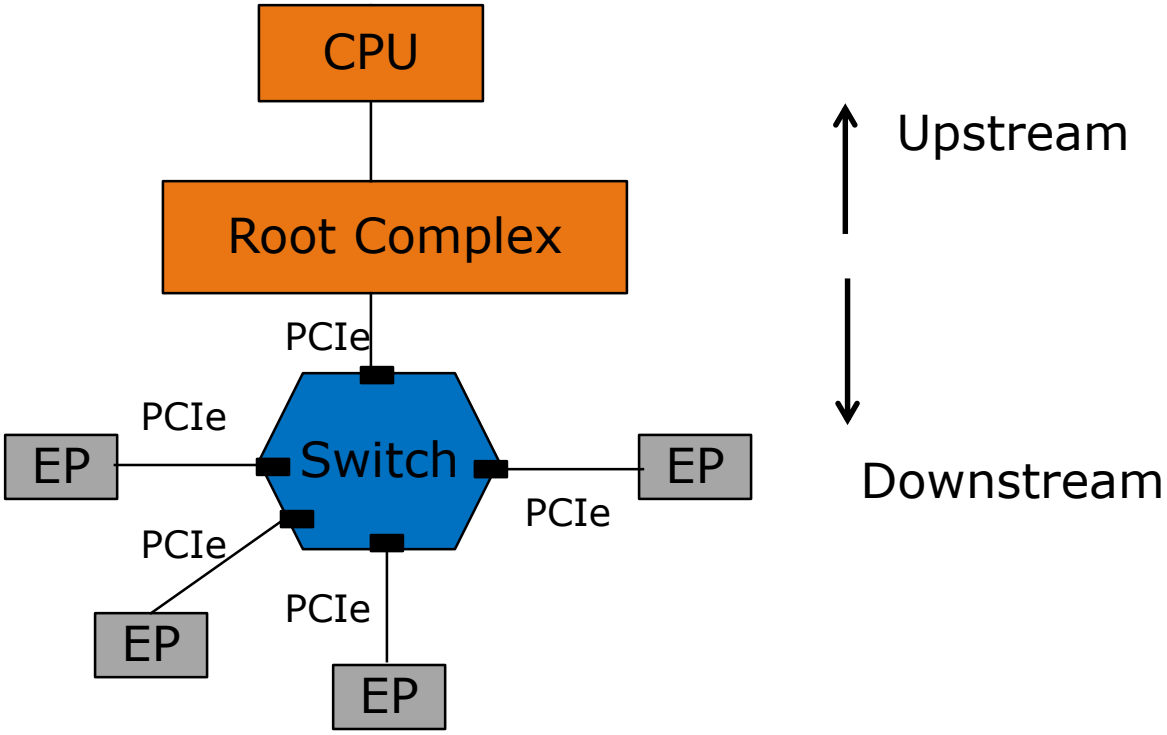




PCI Express – Upstream/Downstream

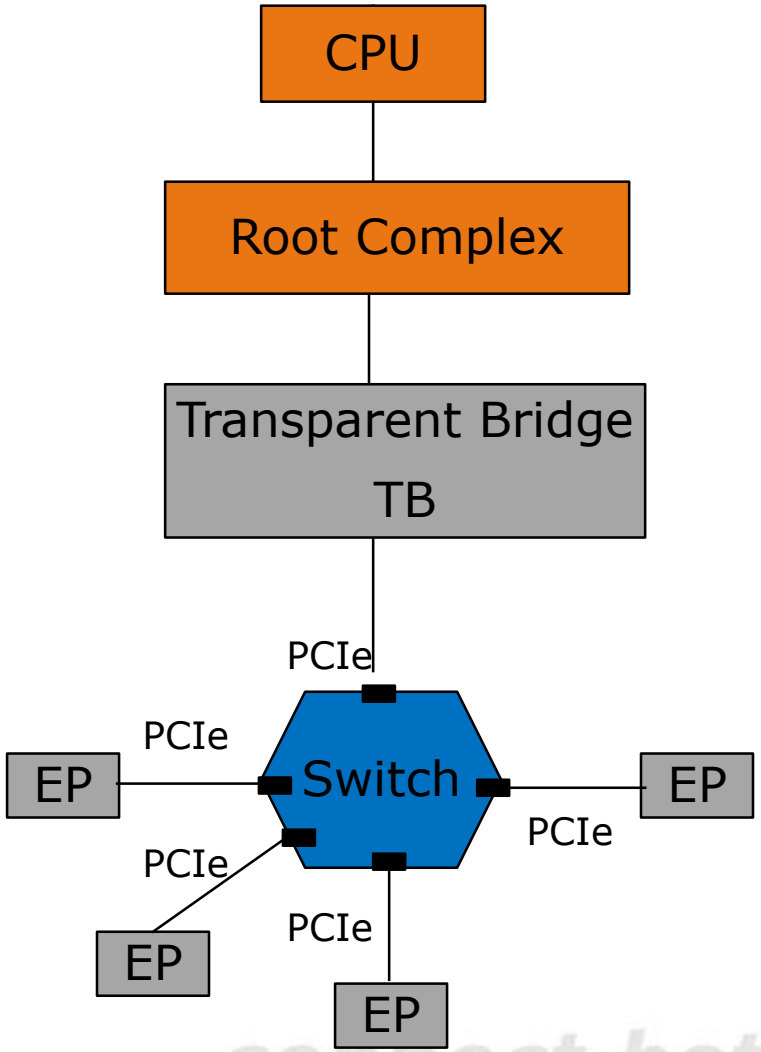
- An Upstream Port is a port that points in the direction of the root complex.
- A Downstream Port is a port that points away from the root complex.

PCI Express – Upstream/Downstream



PCI Express – Transparent Bridge (TB)

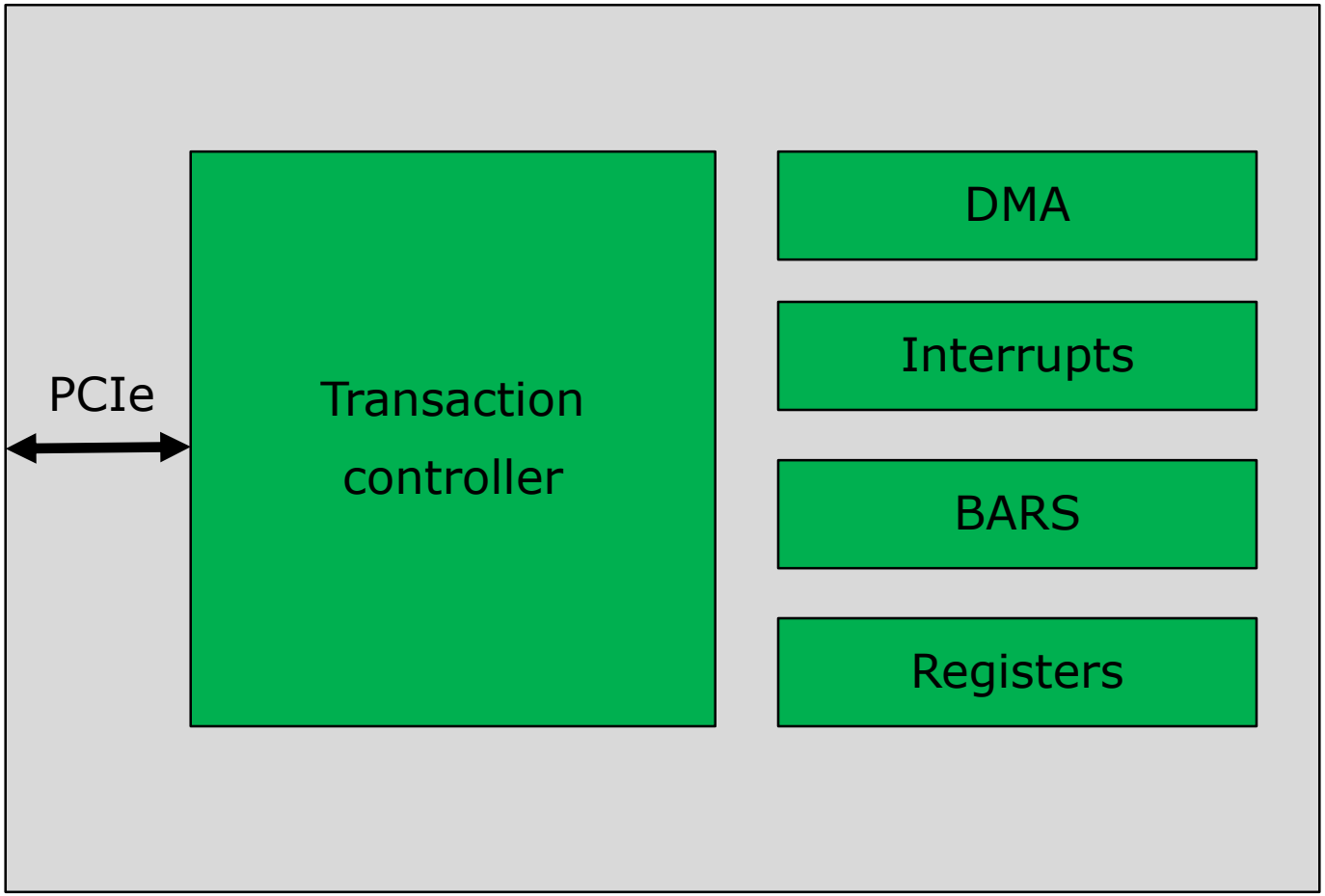
- No software setup



Endpoint (EP)

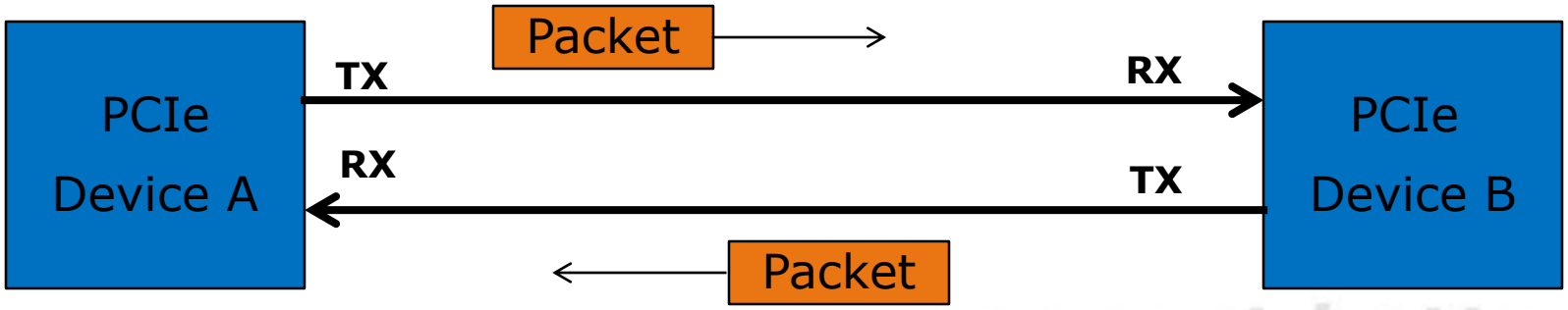
- Supported transaction
  - Configuration or memory mapped transaction
- Requester or completer of PCI Express transaction
- Endpoint initiate transactions as a requester or respond to a transaction as a completer
- Each endpoint is initialized with a requester ID which consist of a bus number, device number and function number

Endpoint (EP)



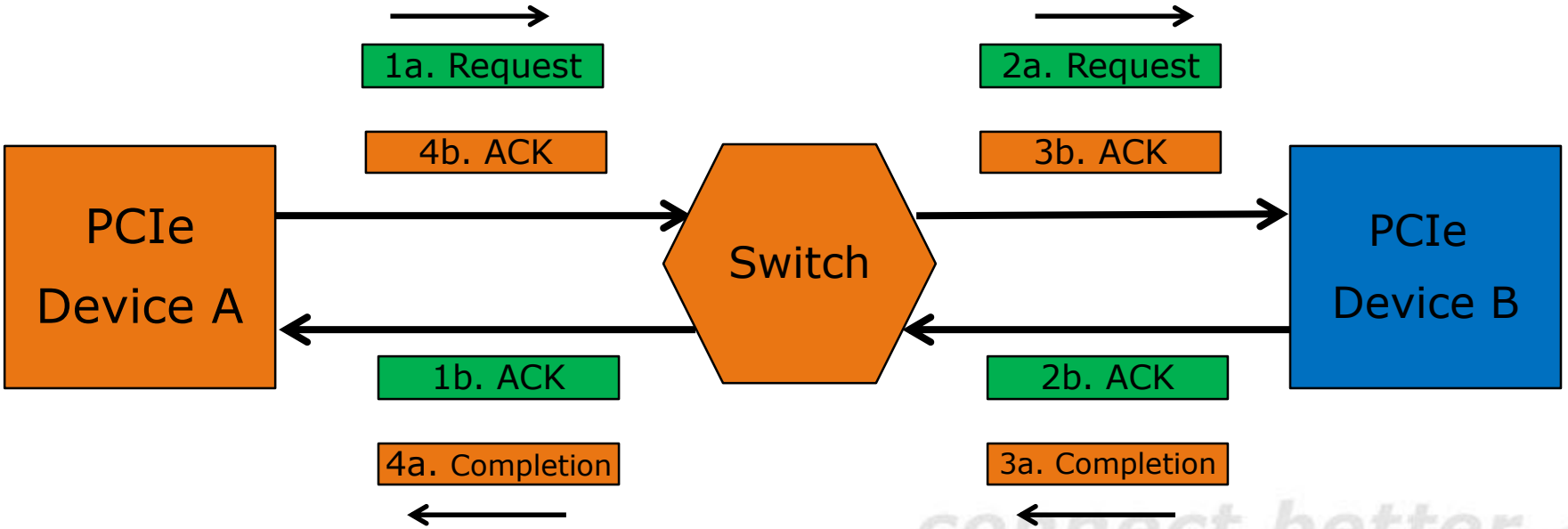
PCI Express features

- Point to point connection
- Scalable: x1, x2, x4, x8, x16, x32
- Packet based transaction protocol
- Differential Signaling



ACK/NAK Protocol: Point-to-Point

- Point to point connection
- ACK returned for good reception of Request or Completion
- NACK returned for error reception of Request or Completion



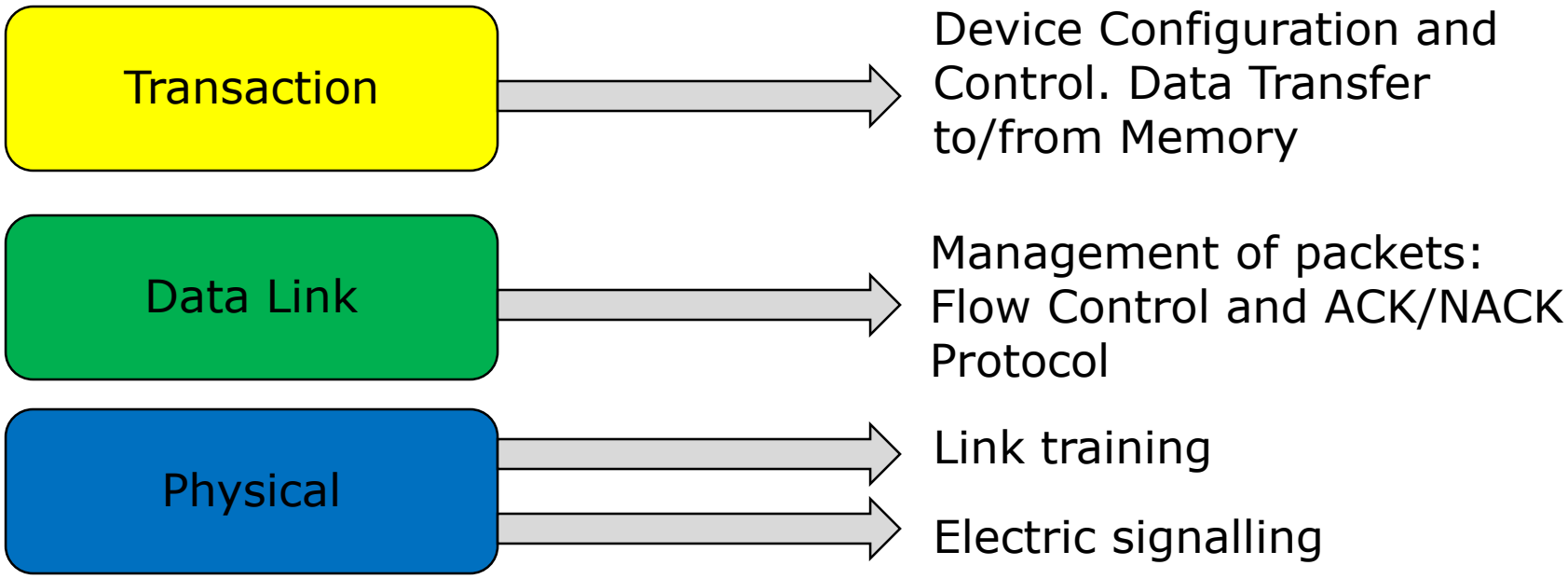
Link training

- The link negotiate to find the appropriate link speed
- The devices send known, ordered sets of symbols to each other and the hardware works its way up from 2.5GT/s.
- Commands to change the link speed is sent to each other
- If the link is unstable, the link can train down again

Transaction Layer Packet (TLP)

- A PCI Express system transfers data in the payload of TLPs.
- Memory data is transferred in Memory Write/Read TLPs
- Completion with Data TLPs are responses to memory read operations.

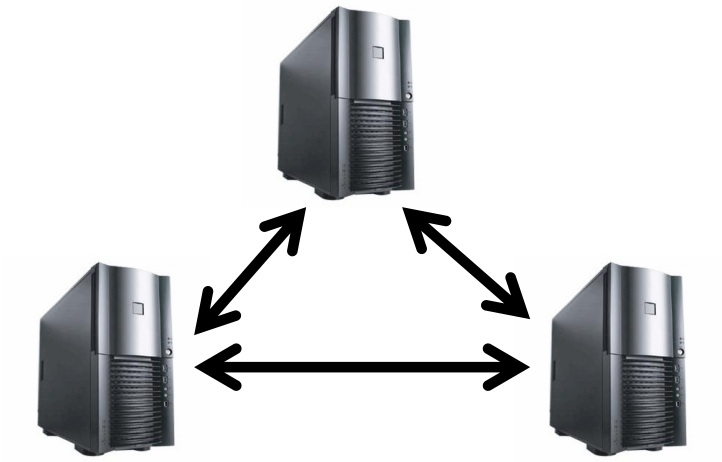
PCI Express Layered Model



Topologies

- DIRECT TOPOLOGY
  - Point to point connection between two nodes

- MULTI DIRECT TOPOLOGY
  - Point to point connection between muliple nodes



Correctable and uncorrectable link errors

- Link errors can occur on the PCIe link
- These errors are reported by **dis_diag** utility
- There a three different error classifications
  - Correctable errors
  - Uncorrectable Non-fatal errors
  - Uncorrectable Fatal errors

Correctable errors

- Correctable errors are handled by the hardware
- Such errors may have inpact on performance
- No information is lost as a result of the error
- Retransmitted by the hardware

Uncorrectable Non-Fatal Errors

- The information has been lost
- Non-fatal errors means that transaction has been corrupted
- PCI Express hardware cannot correct the errors
- The PCI Express fabric continues to function correctly and other transactions are unaffected
- Is handled by device software

Uncorrectable Fatal Errors

- Indicates that a link in a PCI Express fabric is no longer reliable
- Data has been lost and every attempt to recover data under software control can fail
- Fatal errors may be resolved by resetting the link
- On the port link, the software will retransmit the lost data

Uncorrectable error counters – dis_diag –V 9

```
********  PCIe CABLE LINK ERROR INFORMATION FOR ADAPTER NO 0 - Link 1  ********
  NACK error cnt - PCIe Cable Link 1 ..... : 0
     NACK DLLP transmitted .....: 0
     NACK DLLP received .....: 0

  Uncorrectable error cnt - PCIe Cable Link 1 ... : 0
     dlperr cnt ..... : 0
     sdoenerr cnt ..... : 0
     poisoned cnt ..... : 0
     fcperr cnt ..... : 0
     compto cnt ..... : 0
     cabort cnt ..... : 0
     uecomp cnt ..... : 0
     rcvovr cnt ..... : 0
     malformed cnt ..... : 0
     ecrc_cnt ..... : 0
     ur_cnt ..... : 0
     acsv_cnt ..... : 0
     uie_cnt ..... : 0
     mcblktlp cnt ..... : 0
     atopeb cnt ..... : 0
     tlppbe cnt ..... : 0
```

Correctable error counters – dis_diag

Correctable error cnt - PCIe Cable Link 1 ..... : 0

rcverr cnt ..... : 0

badtlp cnt ..... : 0

baddllp cnt ..... : 0

rplyrovr cnt ..... : 0

rplyto cnt ..... : 0

advisorynf cnt ..... : 0

cie cnt ..... : 0

hlo cnt ..... : 0