



Zone-Based Automotive Backbones Tunneling PCIe®

*Dr. Endric Schubert
Chief Technology Officer
Missing Link Electronics, Inc.*

PCI-SIG Virtual Developers Conference 2021

Disclaimer

- Presentation Disclaimer: All opinions, judgments, recommendations, etc. that are presented herein are the opinions of the presenter of the material and do not necessarily reflect the opinions of the PCI-SIG®.

Acknowledgements

- Ulrich Langenbach
Dir. Eng. Missing Link Electronics GmbH (MLE)
- Marcus Pietzsch
Group Mgr, Fraunhofer Institute for Photonic Microsystems (IPMS), Dresden, Germany

Presentation Outline

WHY?

- More safe and eco-friendly vehicles drive automotive connectivity towards so-called Zone-Based Architectures. Inside those Zone Gateways PCIe[®] connects multiple SoC for scalable performance. Zone Gateways connect with each other via the emerging IEEE standards “Time Sensitive Networking” (TSN).

WHAT?

- A solution that fulfills the need for “tunneling” PCIe over TSN (working title PCIe-over-TSN), supporting CPU-to-CPU communication (PCIe NTB) and NVMe storage.

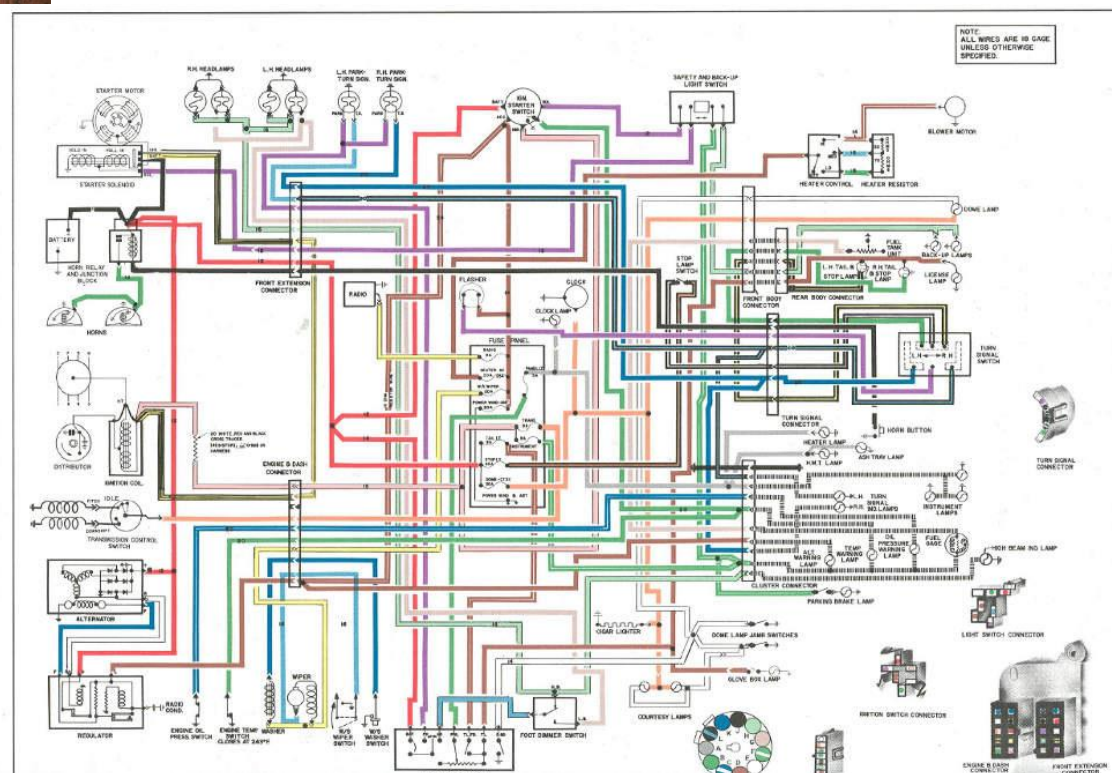
HOW?

- A digital circuit & system stack to encapsulate and to decapsulate PCIe TLPs (along with other protocols) over real-time automotive TSN 10G/25G Ethernet

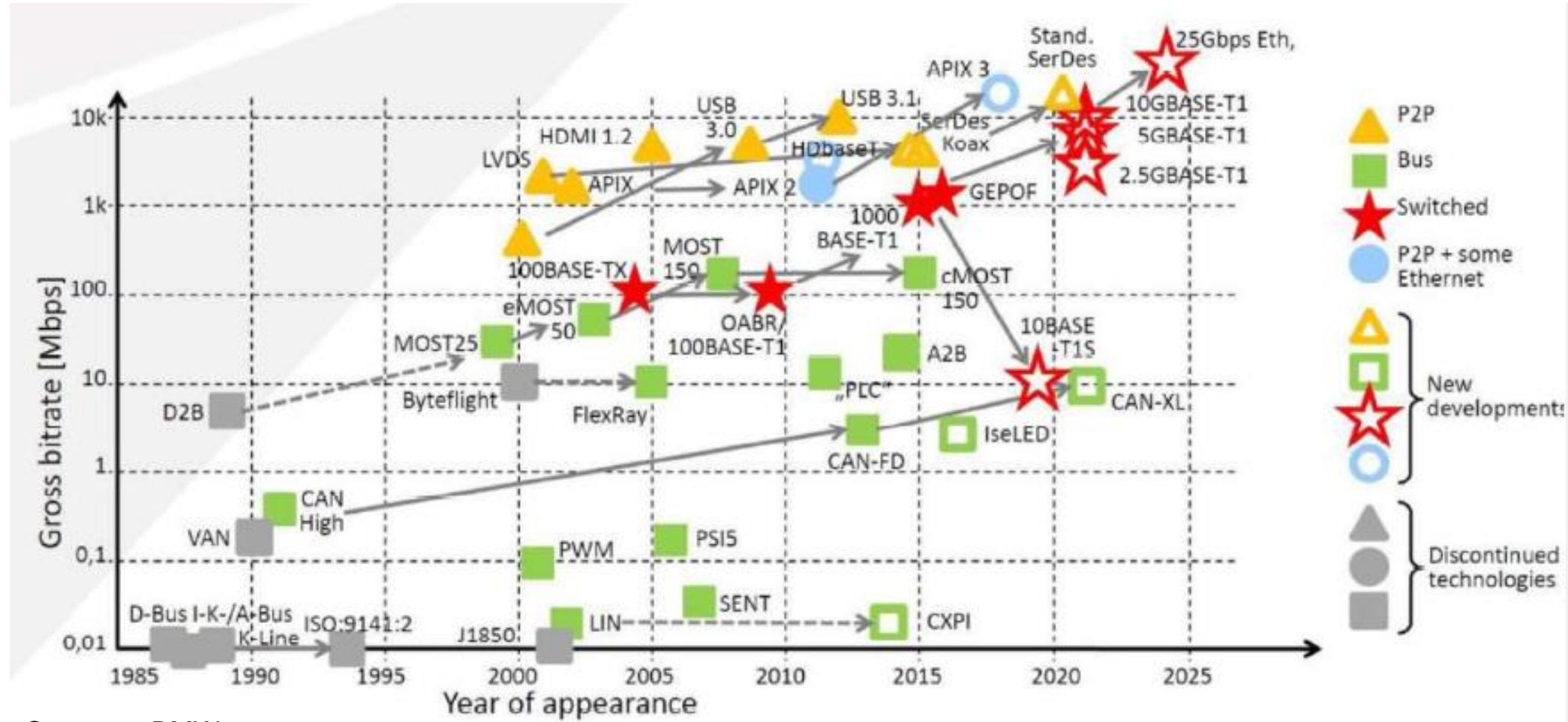
Automotive Network History (1967)



Fig. 12-1—Wiring Diagram (34, 36 and 38 Series, V-8)



Evolution of Automotive Networks



Courtesy: BMW

Today's Expensive Wiring Nightmare

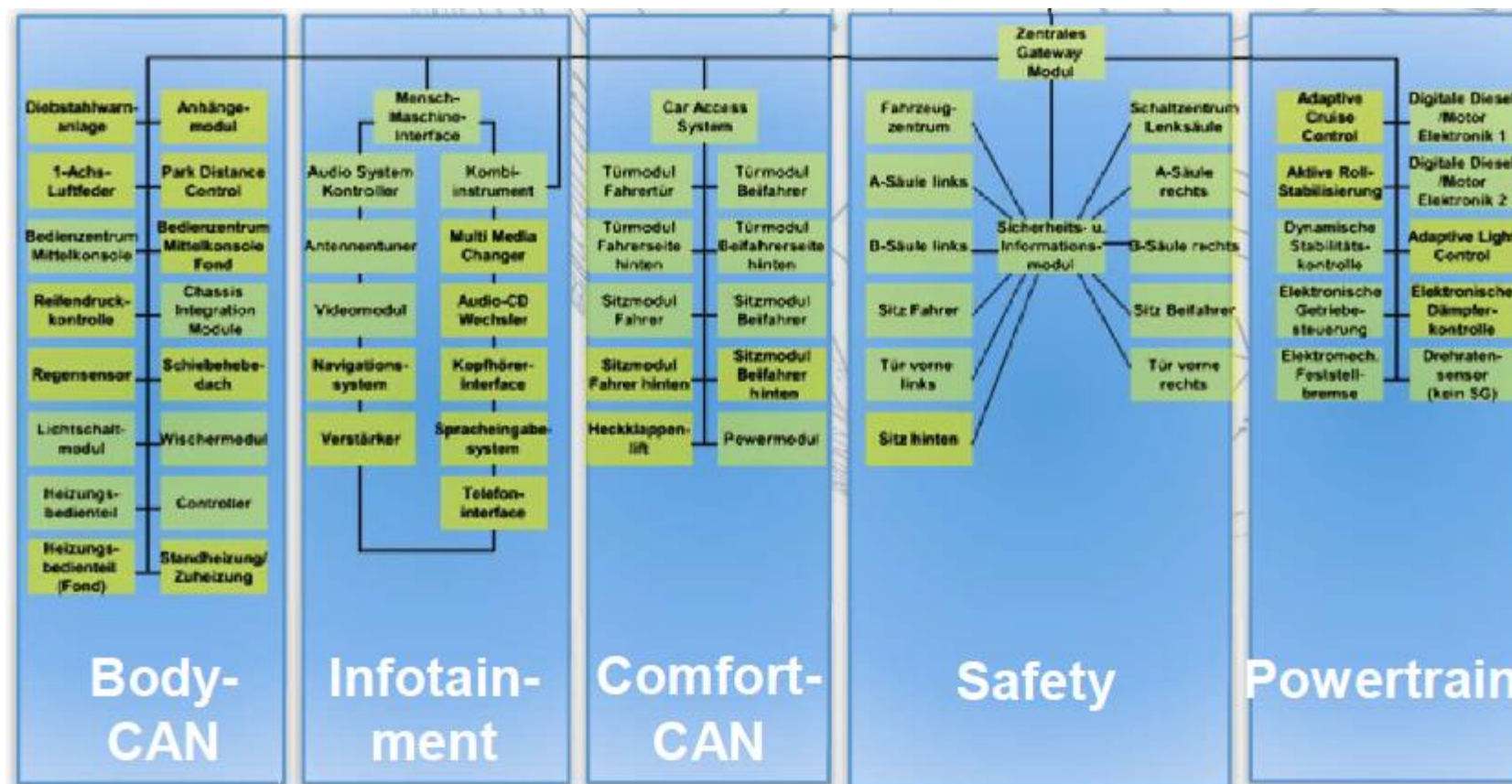


Automotive Networks Today (Domain-Based)



Organized in Functional Domains

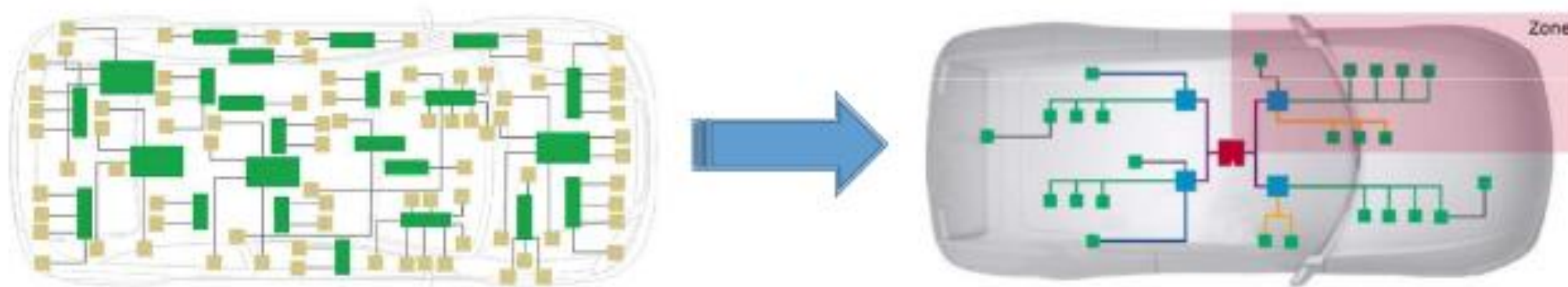
- Powertrain, Body Control, Comfort, Infotainment, etc
- Bus Gateways for Cross-Domain Communication

100 ECUs, or more via many different busses and protocols



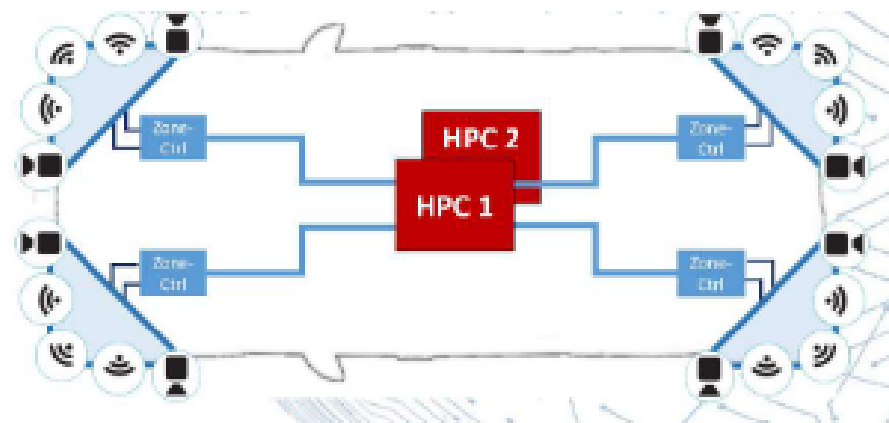
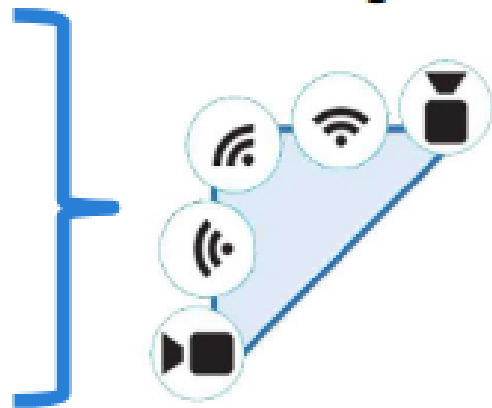
Next: Zone-Based Architectures

- Network connectivity is not based on functional domains, but on *physical location and proximity* inside the vehicle, i.e. “Zones”
- Data aggregation and preprocessing in Zone Controllers: 
- High-bandwidth connectivity towards a central “High Performance Computer” HPC 



Application Example: “Smart Corner”

- Smart Corner: integrates all data sources and sinks located at one corner of a vehicle
- Smart Corner Node contains e.g.:
 - 2 Lidars
 - 2 Radars
 - 2 Cameras
 - 2 Ultrasonics
 - 1 Lighting Unit



Traffic Type	Characteristics
Scheduled Traffic	High-priority real-time traffic transmitted according to a time schedule (time-driven), no interference from other traffic
Stream Reservation	Periodic, guaranteed
Event-driven Traffic	Aperiodic bursts, generated by sporadic events, with real-time constraints
Best-effort Traffic	No guarantees; statistical performance

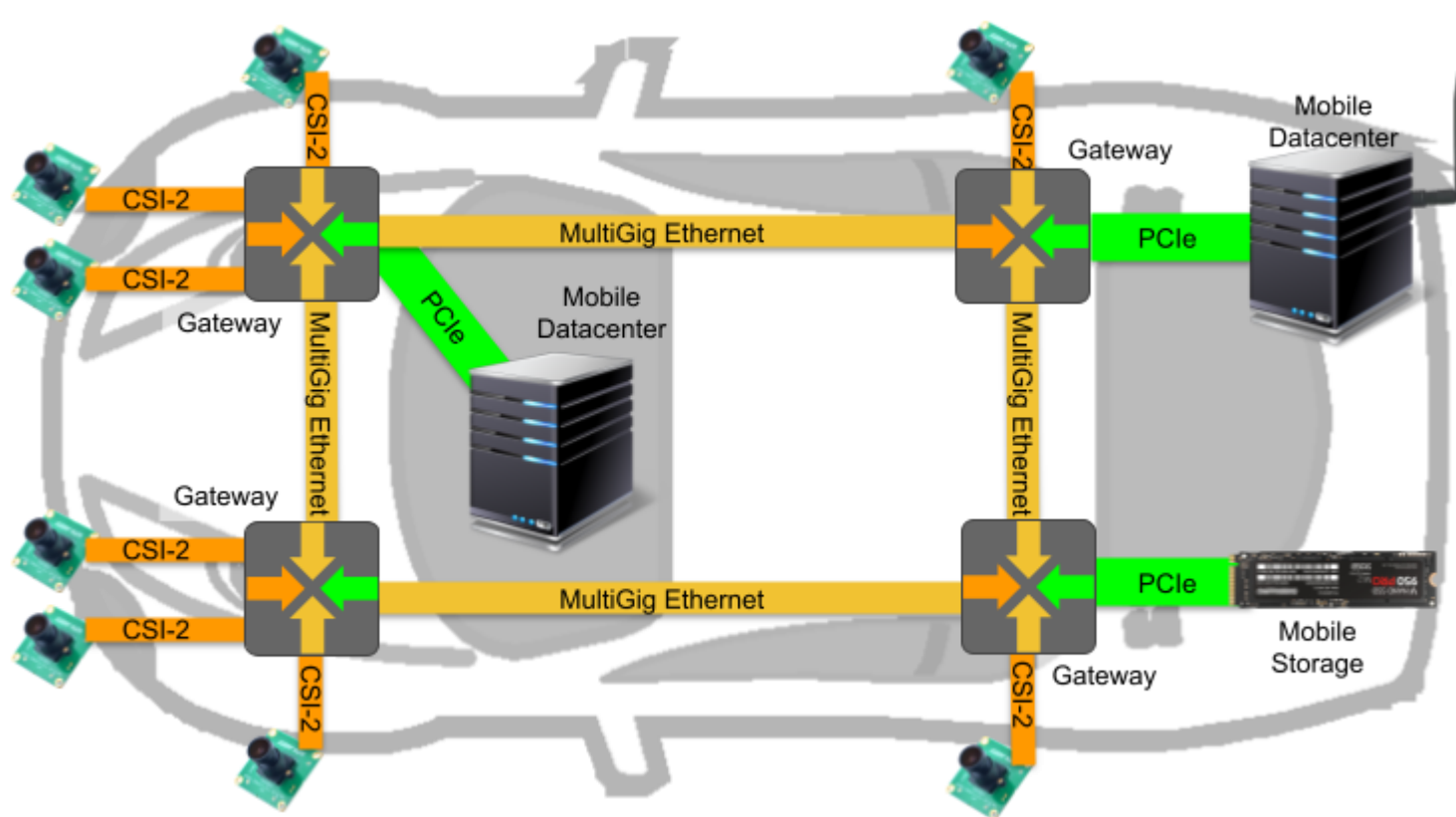
Zone-Based Automotive Network Needs to Transport PCIe

Driven by Cost/Performance, i.e. Centralized Compute & Storage:

- PCIe (for Embedded CPUs, GPUs, FPGAs and SoCs)
- NVMe (for SSDs)

Driven by Compliance

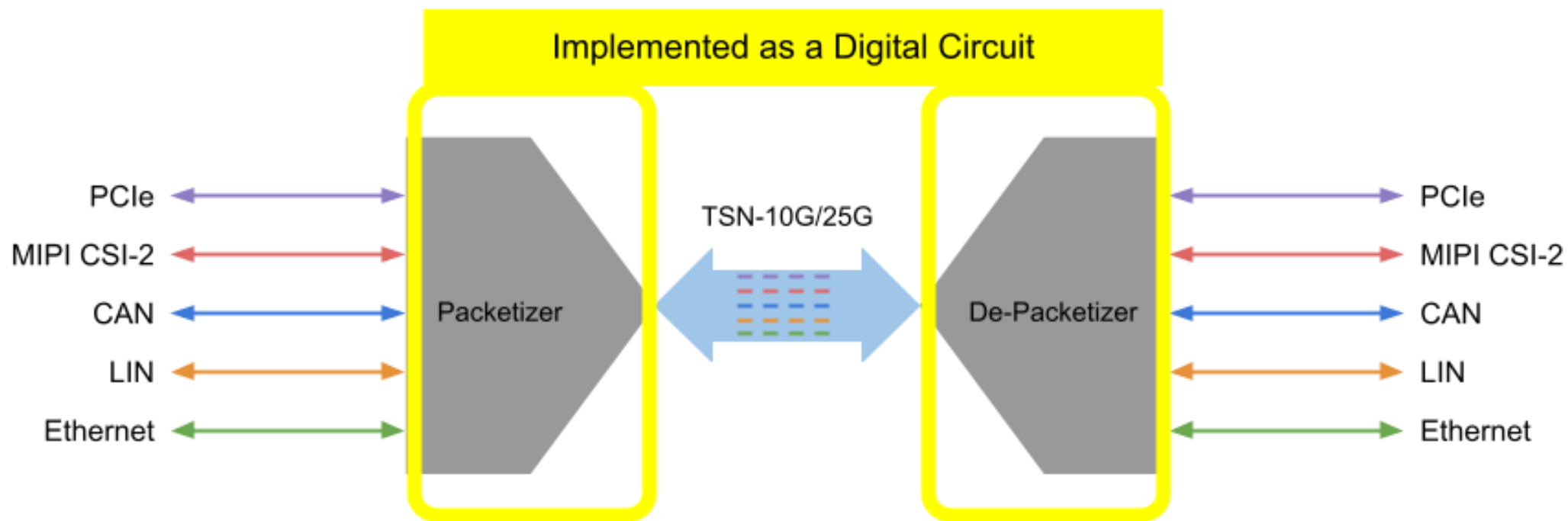
- FuSa ISO 26262
- Security ISO/SAE 21434
- SOTIF ISO 21448
- etc



PCIe-over-TSN is Based on Open Standards

PCIe from PCI-SIG, TSN from IEEE

Symmetric for CPU-to-CPU (e.g. PCIe NTB) or Asymmetric Sensor-to-CPU



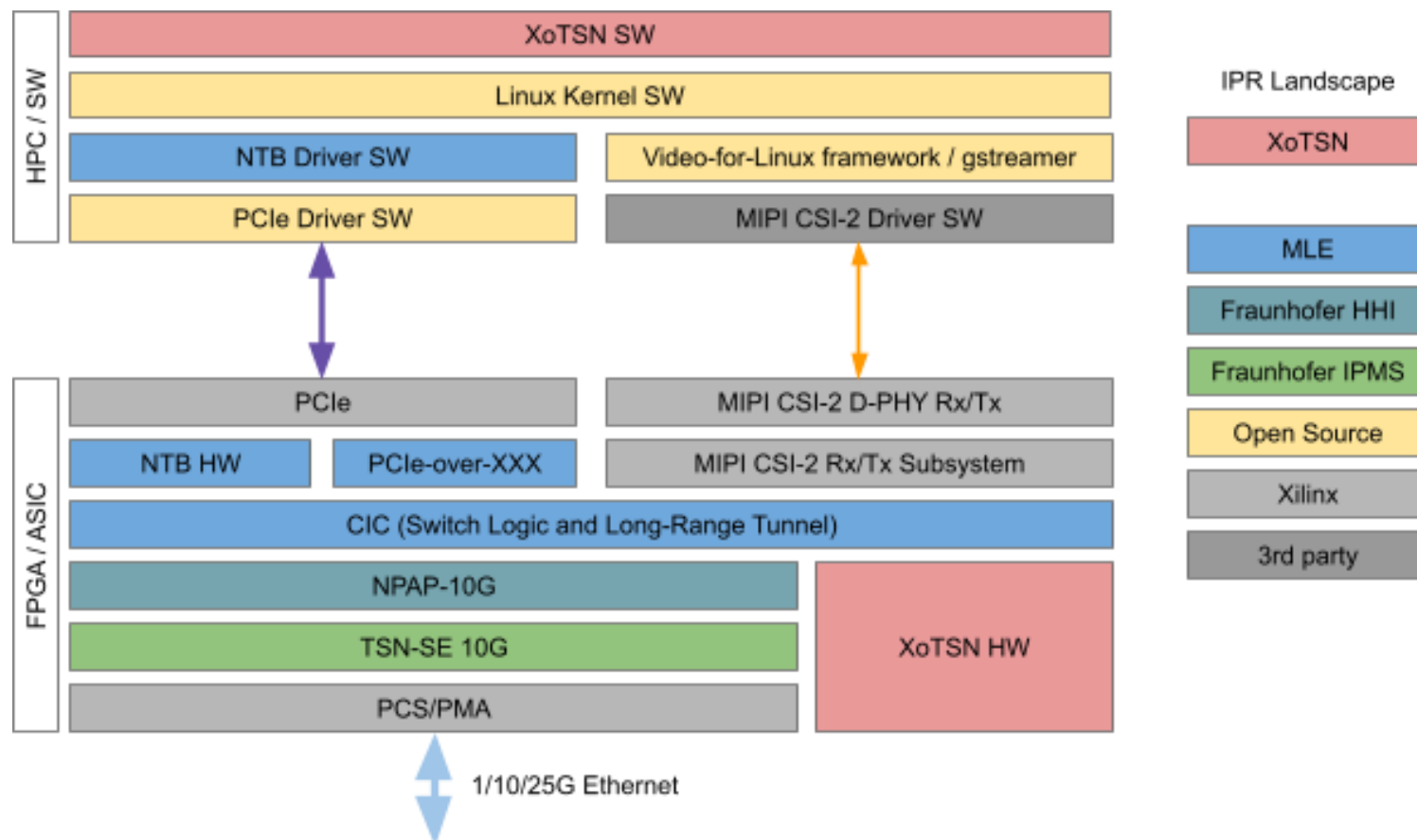
PCle-over-TSN System Stack Example

System Stack is

- Hardware (Digital Circuit)
- Software (Drivers)

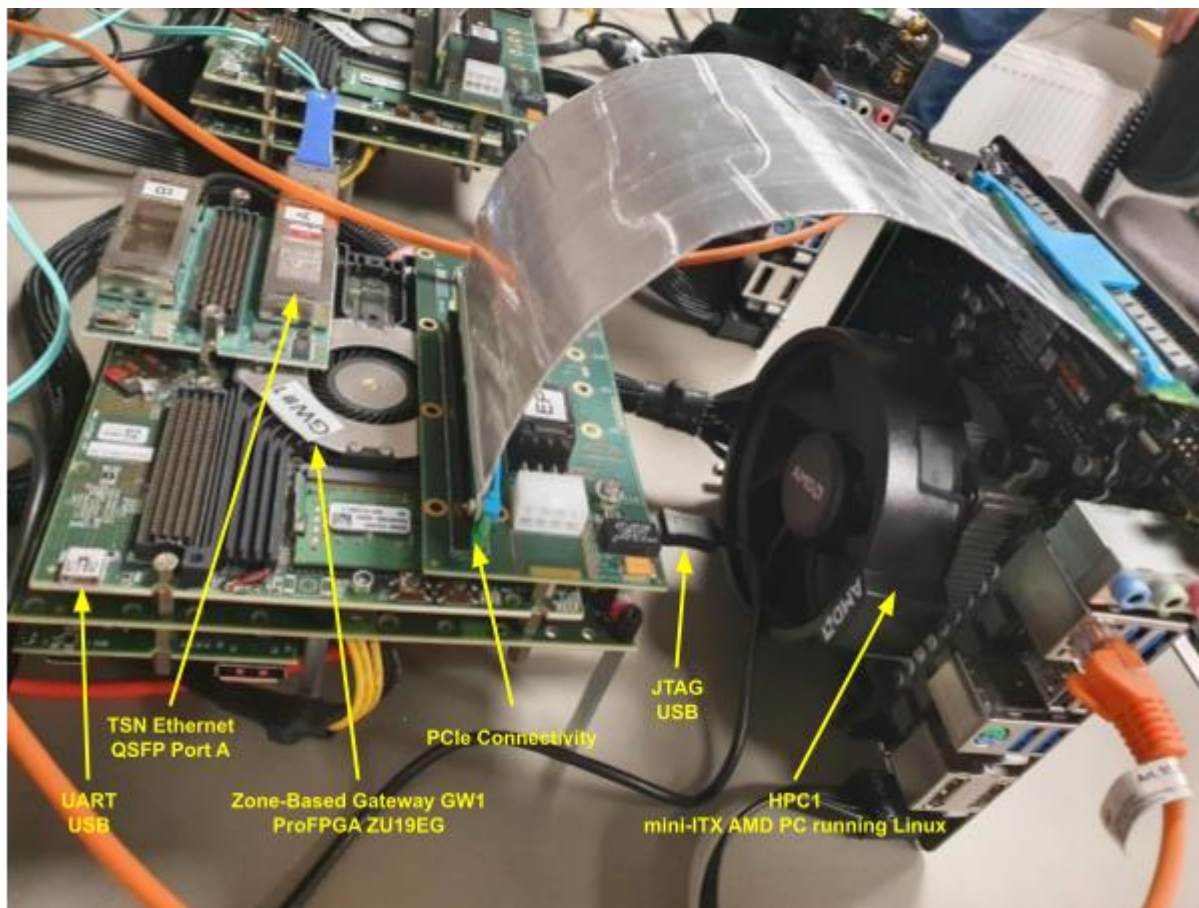
3rd Party (IP Cores)

- Fraunhofer IPMS
- Fraunhofer HHI
- MLE
- Open Source



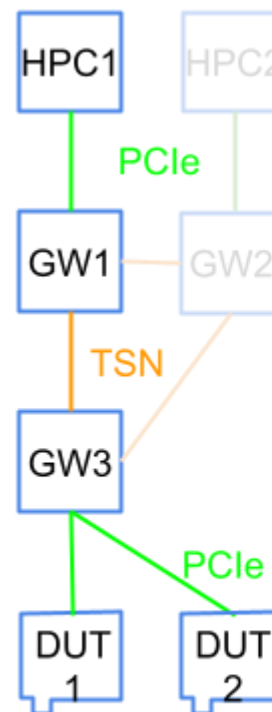
PCIe-over-TSN System Stack Implemented

Labcar Setup w/ PCIe Connect to HPC



Labcar Setup for Experiments

PCIe/NVMe



PCIE NTB



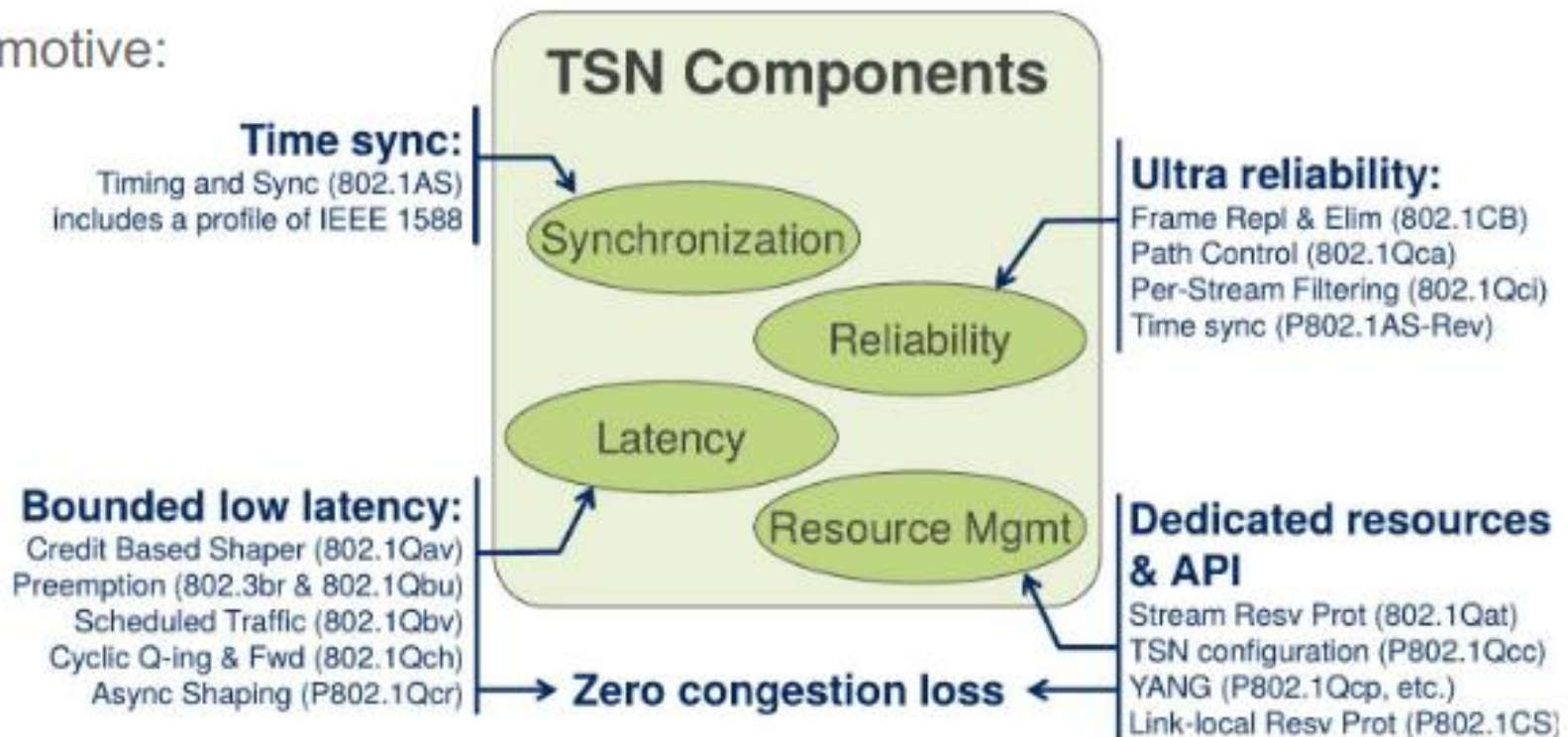
PCIe = 8GT/s x4 TSN = 1GE and 10GE

Backgrounder TSN

TSN is Time Sensitive Networking which is a collection of IEEE Standards for (Layer 2) Ethernet to deliver reliability and bounded latency

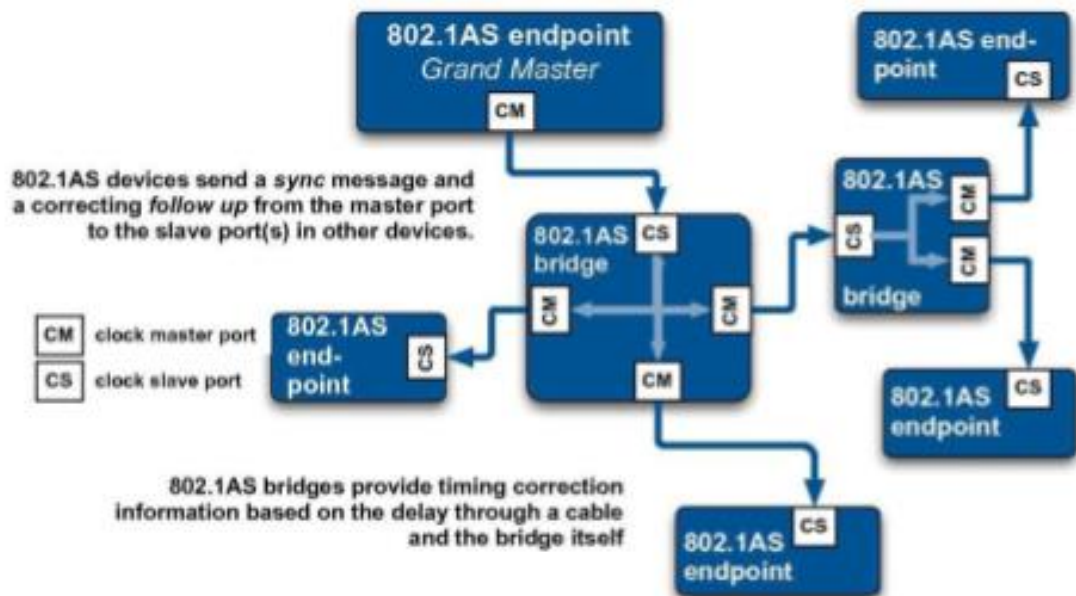
Important for automotive:

- 802.1AS
- 802.1Qbv
- 802.1CB

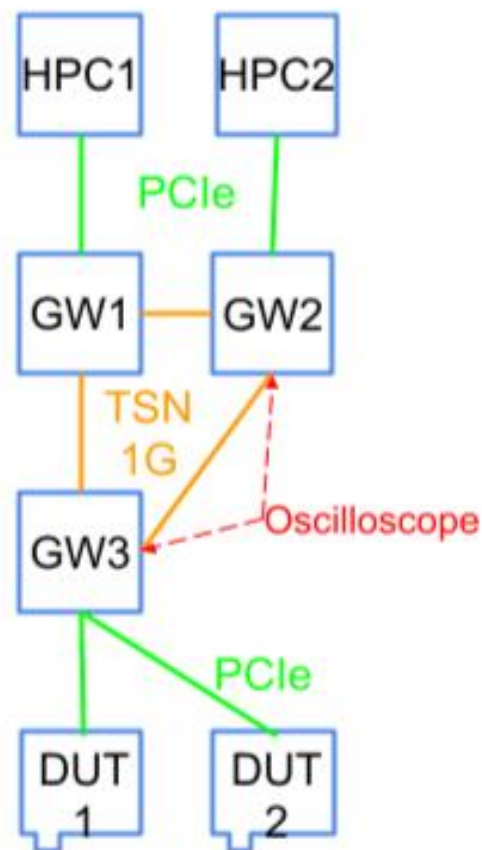


IEEE 802.1AS Precision Timestamping

Concept of Time Synchronization

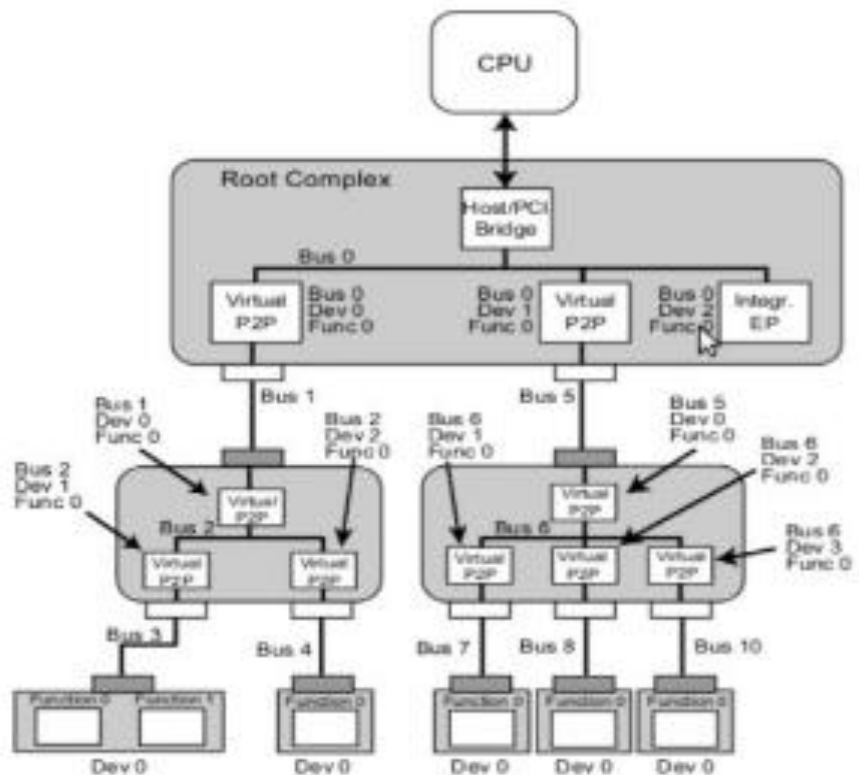


Outcome measured with Oscilloscope is
 ~22 nanoseconds in FPGA (better in ASIC)



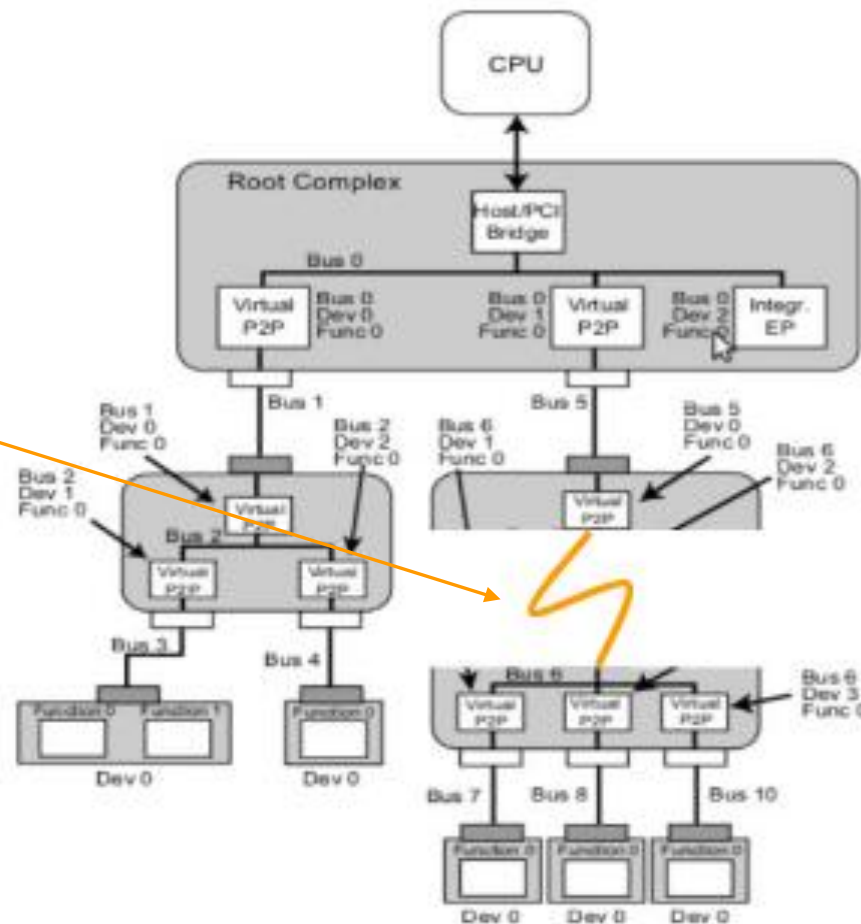
PCIe-over-TSN = “Distributed PCIe Switch”

PCIe Hierarchy with PCIe Switches



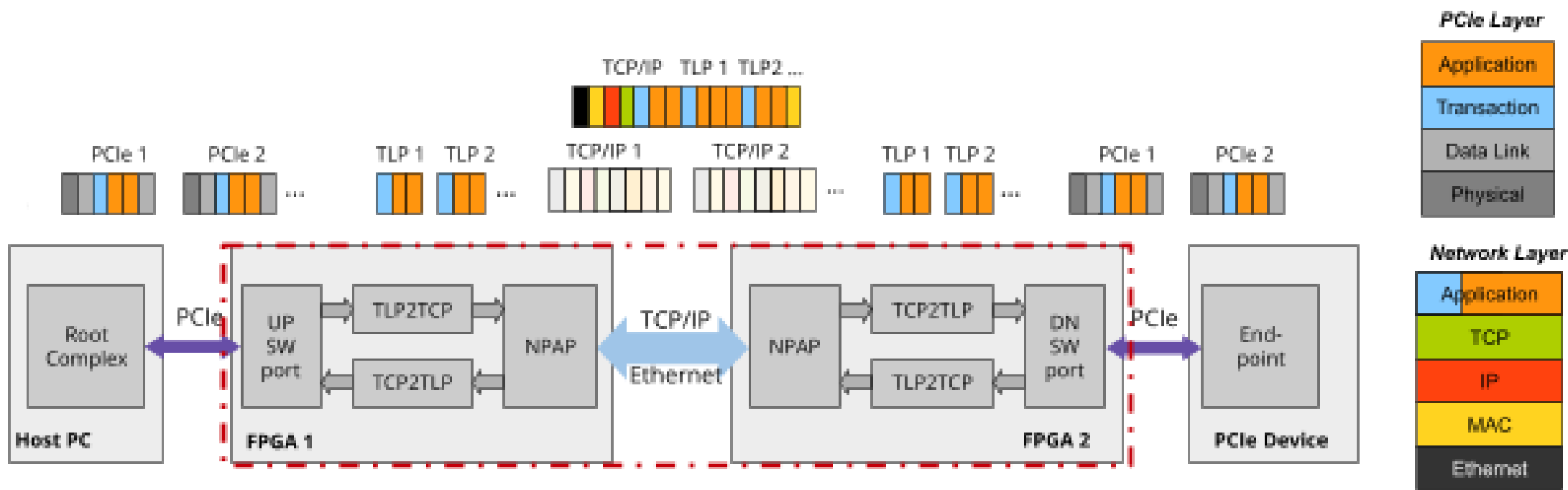
PCIe Long Range Tunnel “cuts open” PCIe Switch

TSN



PCIe-over-TSN Concept

Encapsulate and Decapsulate PCIe TLPs. PCIe demands reliability, therefore we transport TLPs over TCP/IP over TSN over Ethernet.



PCIe-over-TSN Smart TLP Aggregation

Aggregate multiple TLPs to optimize bandwidth dominated communication

Do not aggregate (i.e. send immediately) for latency oriented communication

Obey PCIe ordering rules by handling TLPs “as a stream”

```

00000000 02 00 00 00 00 11 02 00 00 00 00 55 08 00 85 30
00000010 05 c0 9a fa 40 00 ff 06 56 e6 c0 a0 01 09 c0 a8
00000020 01 05 ca 05 ca 06 fc 5a b4 d2 b5 45 28 44 50 18
00000030 01 e0 d3 79 00 00 60 00 00 40 03 00 00 ff 00 00
00000040 00 02 11 c7 4a 00 4b 05 c1 18 50 32 37 45 a9 a0
00000050 72 80 e9 d9 cb 1d 15 04 b9 df 03 b0 23 05 82 ba
<...>
00000130 67 07 92 e3 c1 10 b7 7b 0d 52 be 38 c8 1c 76 2f
00000140 66 0c 11 de 7a 07 00 00 00 00 00 00 00 00 00 00
00000150 00 00 00 00 00 00 60 00 00 40 03 00 00 ff 00 00
00000160 00 02 11 c7 4b 00 ee c5 d1 09 8f d5 a0 18 bd 38
00000170 43 fe c8 95 4e 1e 17 e7 69 03 97 53 d0 1a e2 bc
<...>
00000250 2d 05 34 2c 53 11 09 af 80 c4 af 62 96 8b e1 95
00000260 0b a1 1d ea f9 0a 00 00 00 00 00 00 00 00 00 00
00000270 00 00 00 00 00 00 60 00 00 40 03 00 00 ff 00 00
00000280 00 02 11 c7 4c 00 c0 4c 91 a1 94 95 06 0c 98 29
00000290 02 ed e9 01 f6 0b 33 45 ee 28 54 09 b1 1d a6 48
<...>
00000370 42 21 a5 50 70 00 36 47 4d 67 73 79 35 16 e6 a8
00000380 4e cd 87 eb 06 03 00 00 00 00 00 00 00 00 00 00
00000390 00 00 00 00 00 00 60 00 00 40 03 00 00 ff 00 00
000003a0 00 02 11 c7 4d 00 1c d5 36 a5 c9 f6 66 07 a3 da
000003b0 18 ca 3d 0c 4c 02 54 9b f1 4b 7b 9c df 07 6a 33
<...>
00000490 e2 60 fc 7e 71 15 e6 4e 7d 50 7e ff 29 10 dc a9
000004a0 9c 22 b1 17 10 09 00 00 00 00 00 00 00 00 00 00
000004b0 00 00 00 00 00 00 60 00 00 40 03 00 00 ff 00 00
000004c0 00 02 11 c7 4e 00 06 9c 59 4e c3 d3 45 4c 94 33
000004d0 78 9b 4b 13 b1 16 72 06 a5 59 ad 54 3c 0d ce a0
<...>
000005b0 25 09 84 6a 3f 17 02 a2 1b f9 bd e8 a3 81 48 74
000005c0 22 ee 89 80 63 01 00 00 00 00 00 00 00 00 00 00
000005d0 00 00 00 00 00 00
  
```

Legende:

- Ethernet II Header
[0000-0005] Dst. MAC: (02 00 00 00 00 11) → 02:00:00:00:00:11
[0006-0011] Src. MAC: (02 00 00 00 00 55) → 02:00:00:00:00:55
- Internet Protocol Header
[0017] Protocol: (06) → TCP
[001a-001d] Src. IP: (c0 a8 01 69) → 192.168.1.105
[001e-0021] Dst. IP: (c0 a8 01 65) → 192.168.1.101
- Transmission Control Protocol Header
[0022-0023] Src. Port: (ca 05) → 51717
[0024-0025] Dst. Port: (ca 06) → 51718
- PCIe TLP Header
[0036] FMT/Type: (60) → 64-bit Memory Write Request
[0038-0039] Length: (00 40) → 64 Doublewords (32-bit) → 256 Byte
[003a-003b] Requester ID: (03 00) → 03:0.0
[003e-0045] Address1: (00 00 00 02 11 c7 4a 00)
[015e-0165] Address2: (00 00 00 02 11 c7 4b 00)
[027e-0285] Address3: (00 00 00 02 11 c7 4c 00)
[039e-03a5] Address4: (00 00 00 02 11 c7 4d 00)
[04be-04c5] Address4: (00 00 00 02 11 c7 4e 00)
- Data
- Padding

PCIe-over-TSN Results with NVMe SSD

Linux lspci

```
00:00.0 Host bridge [0600]: Advanced Micro Devices, Inc. [AMD] Device [1022:15d0]
00:01.0 Host bridge [0600]: Advanced Micro Devices, Inc. [AMD] Device [1022:1452]
```

01:00.0 PCI bridge [0604]: Xilinx Corporation Device [10ee:9034] (prog-if 00 [Normal decode])

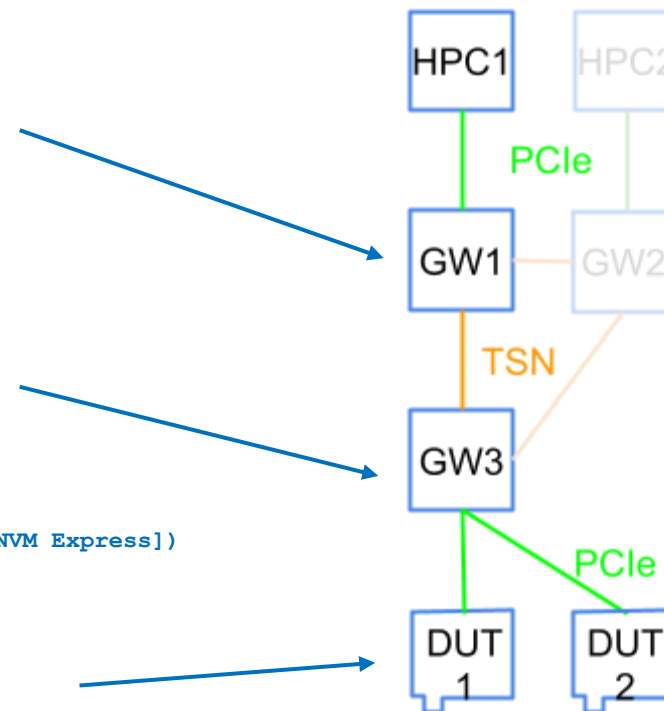
```
Control: I/O+ Mem+ BusMaster+ SpecCycle- MemWINV- VGASnoop- ParErr- Stepping- SERR- FastB2B- DisINTx-
Status: Cap+ 66MHz- UDF- FastB2B- ParErr- DEVSEL=fast >TAbort- <TAbort- <MAbort- >SERR- <PERR- INTx-
Latency: 0, Cache Line Size: 64 bytes
Bus: primary=01, secondary=02, subordinate=03, sec-latency=0
Capabilities: [70] Express (v2) Upstream Port, MSI 00
Capabilities: [1c0 v1] #19
Kernel driver in use: pcieport
Kernel modules: shpchp
```

02:00.0 PCI bridge [0604]: Xilinx Corporation Device [10ee:9134] (prog-if 00 [Normal decode])

```
Control: I/O+ Mem+ BusMaster+ SpecCycle- MemWINV- VGASnoop- ParErr- Stepping- SERR- FastB2B- DisINTx-
Status: Cap+ 66MHz- UDF- FastB2B- ParErr- DEVSEL=fast >TAbort- <TAbort- <MAbort- >SERR- <PERR- INTx-
Latency: 0, Cache Line Size: 64 bytes
Bus: primary=02, secondary=03, subordinate=03, sec-latency=0
Capabilities: [70] Express (v2) Downstream Port (Slot+), MSI 00
Capabilities: [1c0 v1] #19
Kernel driver in use: pcieport
Kernel modules: shpchp
```

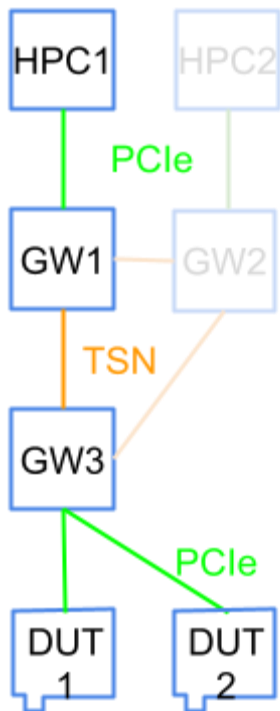
03:00.0 Non-Volatile memory controller [0108]: Samsung Electronics Co Ltd Device [144d:a808] (prog-if 02 [NVM Express])

```
Subsystem: Samsung Electronics Co Ltd Device [144d:a801]
Control: I/O- Mem+ BusMaster+ SpecCycle- MemWINV- VGASnoop- ParErr- Stepping- SERR- FastB2B- DisINTx+
Status: Cap+ 66MHz- UDF- FastB2B- ParErr- DEVSEL=fast >TAbort- <TAbort- <MAbort- >SERR- <PERR- INTx-
Latency: 0, Cache Line Size: 64 bytes
Interrupt: pin A routed to IRQ 30
NUMA node: 0
Region 0: Memory at fcf00000 (64-bit, non-prefetchable) [size=16K]
Kernel driver in use: nvme
Kernel modules: nvme
```

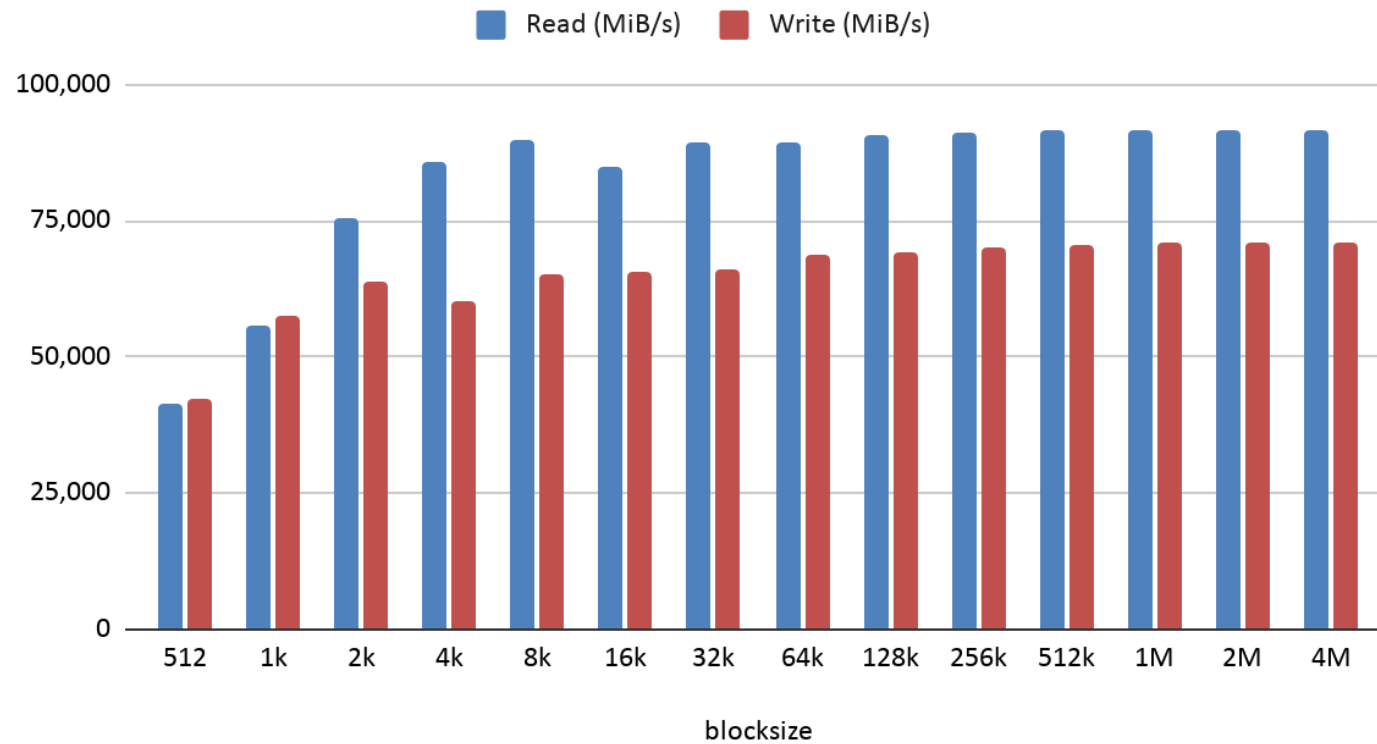


PCIe-over-TSN Robustness

Robust transport, even if TSN “throttles” bandwidth – test results for 1GE

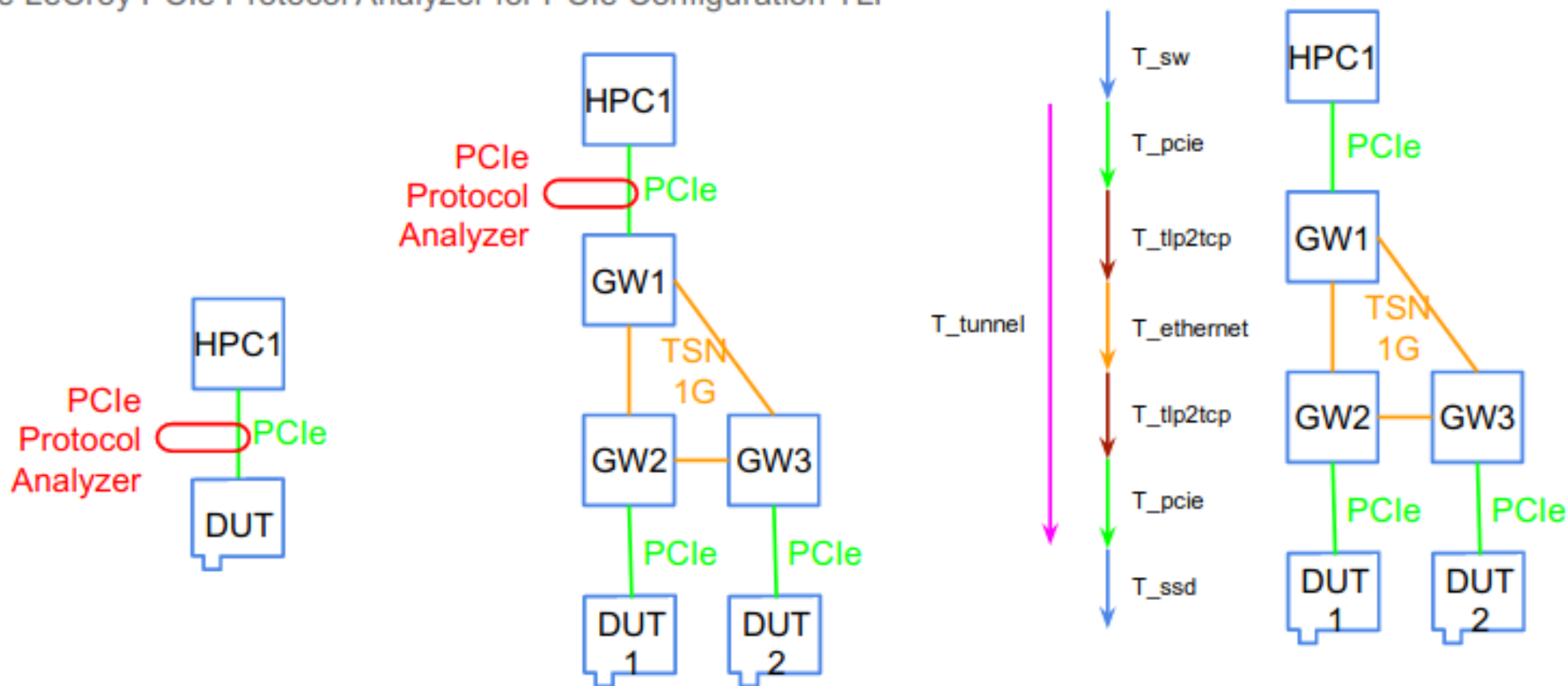


NVMe-over-TSN-1G Read/Write Performance



PCIe-over-TSN Latency Analysis

Compare PCIe Latency for a directly connected SSD against the "tunneled"
 Use LeCroy PCIe Protocol Analyzer for PCIe Configuration TLP



PCIe-over-TSN Latency – Results for 1 GigE

Direct-attached SSD: Baseline one PCIe Config TLP is 646,737,325 - 646,737,045 = 280 ns

0006 . 646 737 045 000 s	200.000 ns	Pkt 4	R→	5.0 / x4	TLP: 303	CfgRd0
0006 . 646 737 245 000 s	76.000 ns	Pkt 5	R←	5.0 / x4	DLLP	ACK
0006 . 646 737 325 000 s	180.000 ns	Pkt 6	R←	5.0 / x4	TLP: 3718	CplD

PCIe-over-TSN: Round-trip time for one PCIe Config TLP is 079,497,529 - 079,490,441 = 7,088 ns

0006 . 079 490 441 000 s	248.000 ns	Pkt 4	R→	5.0 / x4	TLP: 4056	CfgRd1
0006 . 079 490 689 000 s	6.836 us	Pkt 5	R←	5.0 / x4	DLLP	ACK
0006 . 079 497 529 000 s	160.000 ns	Pkt 6	R←	5.0 / x4	TLP: 1452	CplD

$$\Rightarrow T_{\text{tunnel}} = (7,088 - 280) / 2 = 3,404 \text{ ns}$$

10 Gig Ethernet likely shortens this to < 3,000 ns

PCIe-over-TSN for Non-Transparent Bridging (NTB)

```
> lspci -vt
```

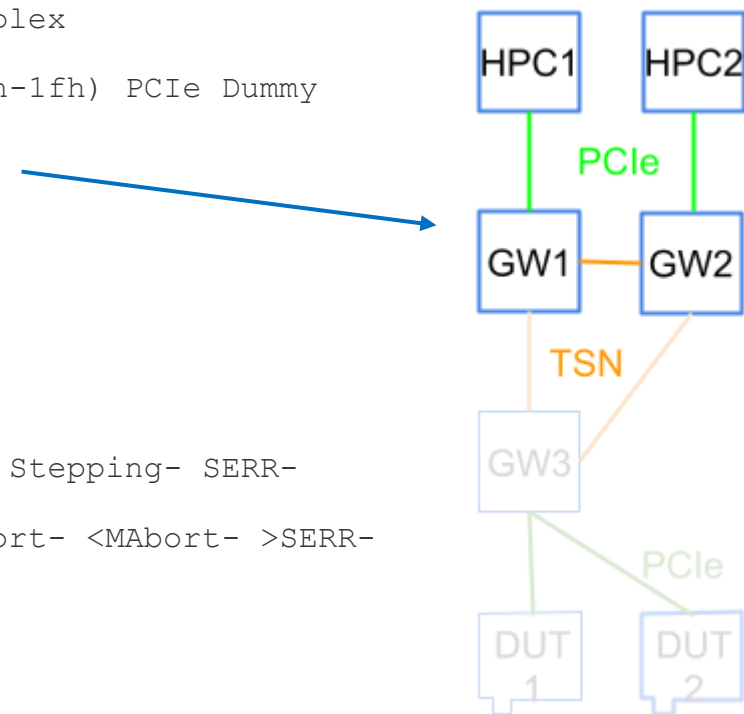
```
-[0000:00]--+-00.0 Advanced Micro Devices, Inc. [AMD] Raven/Raven2 Root Complex
  +-00.2 Advanced Micro Devices, Inc. [AMD] Raven/Raven2 IOMMU
  +-01.0 Advanced Micro Devices, Inc. [AMD] Family 17h (Models 00h-1fh) PCIe Dummy
    Host Bridge
  +-01.1-[01]----00.0 Missing Link Electronics Device 22fb
```

```
...
```

```
> lspci -vv -s 01:00.0
```

01:00.0 Memory controller: Missing Link Electronics Device 22fb

```
Subsystem: Xilinx Corporation Device 0007
Control: I/O- Mem+ BusMaster+ SpecCycle- MemWINV- VGASnoop- ParErr- Stepping- SERR-
  FastB2B- DisINTx+
Status: Cap+ 66MHz- UDF- FastB2B- ParErr- DEVSEL=fast >TAbort- <TAbort- <MAbort- >SERR-
  <PERR- INTx-
Latency: 0, Cache Line Size: 64 bytes
Region 0: Memory at d0000000 (64-bit, prefetchable) [size=256M]
Region 2: Memory at c0000000 (64-bit, prefetchable) [size=256M]
Region 4: Memory at fcf10000 (32-bit, non-prefetchable) [size=64K]
Region 5: Memory at fcf00000 (32-bit, non-prefetchable) [size=64K]
Capabilities: [1c0 v1] #19
Kernel driver in use: ntb_hw_mle
Kernel modules: ntb_hw_mle
```



PCle-over-TSN “Labcar”



Conclusion

- In-vehicle connectivity is quickly evolving, driven by cost/performance and compliance
 - Migration towards centralized so-called Zone-Based Architectures
- PCIe key choice for short-distance local connect between CPUs, GPUs, FPGAs, SoCs
- Our working proof-of-concept underlines feasibility
 - A layered stack of open-standard protocols extends reach of PCIe over real-time Ethernet / Automotive Ethernet – working title PCIe-over-TSN
- Implementation as a digital circuit delivers robust behavior at deterministic, low latencies

Thank you for attending the PCI-SIG Virtual Developers Conference 2021!

For more information about PCIe technology, please go to www.pcisig.com