# OTN support for Beyond 400 Gb/s Ethernet

Steve Trowbridge (Nokia)

Tom Huber (Nokia)

# Supporters

- Pete Anslow (Independent)
- Ralf-Peter Braun (Deutsche Telekom)
- Paul Brooks (Viavi Solutions)
- Chris Cole (II-VI Photonics)
- Steve Gorshe (Microchip)
- Kishore Kota (Inphi)
- David Lewis (Lumentum)
- Ted Sprague (Infinera)
- Peter Stassar (Huawei)

# Previous and current projects with OTN support objectives

- IEEE Std 802.3ba-2010 40 Gb/s and 100 Gb/s Ethernet
- IEEE Std 802.3bg-2011 40Gb/s Ethernet Single-mode Fibre PMD*
- IEEE Std 802.3bm-2015 40 Gb/s and 100 Gb/s Fibre Optic
- IEEE Std 802.3by-2016 25 Gb/s Ethernet
- IEEE Std 802.3bs-2017 200 Gb/s and 400 Gb/s Ethernet
- IEEE Std 802.3cc-2017 25 Gb/s Ethernet over Single-Mode Fiber
- IEEE Std 802.3cd-2018 50 Gb/s, 100 Gb/s, and 200 Gb/s Ethernet
- IEEE Std 802.3cm-2020 400 Gb/s over Multimode Fiber
- IEEE Std 802.3cn-2019 50 Gb/s, 200 Gb/s, and 400 Gb/s over greater than 10 km of SMF
- P802.3ct 100 Gb/s over DWDM systems Task Force
- IEEE Std 802.3cu-2021 100 Gb/s per lane Optical PHYs
- P802.3cw 400 Gb/s over DWDM systems Task Force
- P802.3db 100 Gb/s, 200 Gb/s, and 400 Gb/s Short Reach Fiber Task Force



* Variant on the theme: Provide optical compatibility with existing carrier 40Gb/s client interfaces (OTU3/STM-256/OC-768/40G POS)

# Issues to be considered regarding OTN support

Major OTN support issues considered in previous projects:

1. A new rate of Ethernet should “fit” to be carried over a corresponding established transport network signal
2. There should be a single PCS (or other canonical format) common to all PHY types for a given rate of Ethernet that can be used as the basis for the mapping of Ethernet over OTN
3. Ability to reuse Ethernet modules for OTN client interfaces at the corresponding OTN rate (typically slightly higher than the Ethernet signaling rate)

New OTN support issues to be considered:

4. Minimize signaling rate differences between Ethernet and Transport
5. Coding and error marking considerations for OTN transport of Ethernet

# Issue #1 (Original Issue): A new rate of Ethernet should “fit” to be carried over a corresponding established transport network signal

- Less of an issue for newer high signaling rates since Ethernet has become the major client of OTN (replacing SONET/SDH of 20 years ago), and new rates of Ethernet and OTN are normally selected together, where the fact that Ethernet fits into OTN is by design.
- Original 10G OTN (OTU2) container optimized to carry SONET OC-192 or SDH STM-64 signals
- IEEE Std 802.3ae-2002 defined 10GBASE-R at 10.3125 Gb/s (66B coded) after initial OTN standardization. The alternate format 10GBASE-W (rate-reduced to match OC-192/STM-64 signaling rate) was not widely adopted
- Numerous, non-interoperable 10GBASE-R mapping solutions emerged in the market initially on a proprietary basis. Some over-clocked alternatives interfered with the normal OTN signal multiplexing hierarchy
- Most OTN equipment vendors were required to support many different mappings for the 10GBASE-R signal
- Lessons learned: there should be one (canonical) Ethernet format that needs to be transported for any given Ethernet MAC rate, and that signal format should be possible to carry over the corresponding transport rate with at least PCS codeword transparency

# Issue #1 - continued

- The first project to adopt an OTN support objective was P802.3ba, which had the potential for generating a similar problem at 40G as at 10G. The likely coded rate for a 40GBASE-R signal (41.25 Gb/s) was greater than the OTU3 transport payload rate optimized for carrying SONET OC-768 or SDH STM-256 signals
- The problem was addressed by strictly constraining the PCS codeword set to allow ITU-T to transcode 66B to 1024B/1027B to fit, without risk that a 40GBASE-R PHY would generate non-transcodable bit sequences:
  - Clause 82.2.3.3 Block structure: All unused values of block type field are invalid; they shall not be transmitted and shall be considered an error if received.
  - Below Figure 82-5: WARNING-The mapping of 40GBASE-R PCS blocks into OPU3 specified in ITU-T G.709 [B48] depends on the set of control block types shown in Figure 82–5. Any deviation from the coding specified in Figure 82–5 will break the mapping and may prevent 40GBASE-R PCS blocks from being mapped into OPU3 (see ITU-T G.709 [B48] for more details).
  - Clause 82.2.3.4 Control codes: All XLGMII/CGMII, 40GBASE-R, and 100GBASE-R control code values that do not appear in the table shall not be transmitted and shall be considered an error if received.
- This allowed for a single, standardized mapping of 40GBASE-R into OTU3. The explosion of proprietary variants seen at 10G was avoided.

# Issue #2: There should be a single PCS (or other canonical format) common to all PHY types for a given rate of Ethernet that can be used as the basis for the mapping of Ethernet over OTN

- Two important goals:
  - Only one (generally PCS codeword transparent) mapping for each rate of Ethernet into OTN needs to be defined that doesn't depend on PHY type (or force development of new mappings whenever 802.3 specifies a new PHY type)
  - Different Ethernet PHY types at the same signaling rate can be used at the OTN ingress and egress
- Originally this was a natural "fallout" from the tightly-constrained single PCS format for 40GBASE-R and 100GBASE-R as defined by P802.3ba in clause 82
- Mappings of 40GBASE-R into OPU3 and 100GBASE-R into OPU4 are based on serialized and deskewed PCS lanes (66B encoded for 100GBASE-R, 1024B/1027B transcoded for 40GBASE-R).
- 25GBASE-R PCS (clause 107) and 50GBASE-R PCS (IEEE Std 802.3cd-2018 clause 133) are without FEC, and are mapped uniformly in 66B format
- As the 200GBASE-R and 400GBASE-R PCS includes RS(544) FEC which isn't carried over OTN (and is only capable of correcting errors for a single Ethernet ingress or egress link and not over the OTN-carried part of the link), the OTN mapping corrects any errors, removes the Ethernet FEC, and trans-decodes to 66B for the universal OTN mapping at these rates, and adds the OTN overhead and FEC.

# Issue #3: Ability to reuse Ethernet modules for OTN client interfaces at the corresponding rate

- For 10G, commonality of client optics between Ethernet, SONET/SDH and OTN was simple: 10GBASE-R, OC-192/STM-64, and OTU2 client signals were all serial 10G at slightly different signaling rates, but could all use the same SFP form factor pluggable optics
- For 40G, transport interfaces came first, and the initial VSR2000-3R2 client optics built for the low-volume transport market for OC-768/STM-256 and OTU3 client optics were quite costly. Initial 40GBASE-R interfaces were parallel 4×10G interfaces that couldn't share volume with transport. The IEEE Std 802.3bg-2011 later added serial 40GBASE-R using the same VSR2000-3R2 optics, but this came late in the 40G lifecycle and had limited market volumes.
- From 40G and 100G onward, most high-speed Ethernet interfaces were parallel, requiring specification of a parallel version of the transport frame formats to use the modules. It was important to avoid dependency of the logic in pluggable modules that depended on the exact signal format. The blind bit multiplexing used in PMAs for 40 Gb/s, 100 Gb/s, and other high-rate signals was ideal for the dual use of these components for Ethernet and Transport

# Issue #3 – continued
## Examples of dual-use pluggable modules for Ethernet and Transport applications

| Ethernet Spec (optical and logic) | ITU-T Optical | ITU-T Frame Format |
|---|---|---|
| 40GBASE-LR4 (clause 87) | G.695 C4S1-2D1 | G.707 STL256.4 or G.709 OTL3.4 |
| 40GBASE-ER4 (clause 87) | G.695 C4L1-2D1 | |
| 100GBASE-LR4 (clause 88) | G.959.1 4I1-9D1F | G.709 OTL4.4 or G.709.1 FOIC1.4 |
| 100GBASE-ER4 (clause 88) | G.959.1 4L1-9C1F | |
| CWDM4 MSA | G.695 C4S1-9D1F | |
| 4WDM 40km “ER4-lite” | G.959.1 4L1-9D1F | |
| 200GBASE-FR4 (clause 122) | G.695 C4S1-4D1F | G.709.1 FOIC2.4 |
| 200GBASE-LR4 (clause 122) | G.959.1 4I1-4D1F | |
| 400GBASE-FR8 (clause 122) | G.959.1 8R1-4D1F | G.709.1 FOIC4.8 |
| 400GBASE-LR8 (clause 122) | G.959.1 8I1-4D1F | |

IEEE Std 802.3cu-2021 equivalent specifications still to come following recent approval

# New OTN support issues to be considered
## Issue #4: Minimize signaling rate differences between Ethernet and OTN for comparable applications

- While module reuse across Ethernet and OTN applications (Issue #3) has been facilitated for prior generations (400 Gb/s and below), generally there has been a signaling rate difference (in the 6-10% range) across the two applications
- At higher signaling rates, the penalty for operating at the higher speed can be significant.
- The next few slides look at a few of the reasons the rates have tended to be different, and some things that can be explored to minimize the differences

# Issue #4 – FEC as a cause of rate difference

- At low rates, Ethernet interfaces were generally without FEC, and OTN interfaces generally with FEC.
- Even when both Ethernet and OTN used FEC, sometimes the FEC code was different
- Increasingly, Ethernet and OTN choose the same FEC for the same general application space:
  - For 400G Ethernet and OTN client interfaces, 400GBASE-R and OTN FOIC4.x interfaces use the same RS(544,514) FEC (~5.83% overhead)
  - For 100G coherent DCI links, P802.3ct 100GBASE-ZR and G.709.2 OTU4-SC interfaces use the same staircase FEC (~6.7% overhead)
  - For 400G coherent DCI links, P802.3cw 400GBASE-ZR and G.709.3 FlexO4-DSH interfaces use the same CFEC (HD staircase outer, SD Hamming inner) code (~14.8% overhead)
- IEEE 802.3 and ITU-T SG15 should continue to look for opportunities to use the same FEC for the same application space. Increasingly, FEC should not be the factor that creates a rate difference between Ethernet and OTN

# Issue #4: Networking overhead (the “wrapper”) as a cause of rate difference

- OTN signals historically have added a digital wrapper to client signals, adding 16 columns of overhead to 3808 columns of client payload.
- The same ratio of overhead to payload has been maintained for two decades, resulting in an unnecessarily large 160× the amount of overhead for a 400G signal as for a 2.5G signal.
- At a minimum, a “skinnier” wrapper could be adopted by ITU-T with the scale adjusted for the higher signaling rates
- But also to consider – is there some place inband in the Ethernet signal where networking overhead could be added without reducing the packet carrying capability? As an example, 400GBASE-R has a 133-bit pad in the AM field mapped into FEC codewords that adds up to ~1.26 Mb/s. This might scale to 2.52 Mb/s on an 800G interface, or 5.04 Mb/s on a 1.6T interface. Not to unnecessarily increase the signaling rate for Ethernet, but could the framing be designed with an allocation for adding this kind of overhead without increasing the bit-rate?

# Issue #4 – Service multiplexing and interworking with legacy network equipment as a driver for bit-rate differences

- OTN includes a multiplexing hierarchy that enables an OTN signal to carry multiple lower-rate OTN client signals (including their OAM overhead) within the payload of the higher-rate signal. Ethernet doesn't (i.e., PHY layer overhead is removed/added before and after the bridging layer)
- It may be an expectation for OTN equipment that a 1.6T signal could carry four individually wrapped and managed 400GBASE-R signals (or 16 individually wrapped 100GBASE-R signals, etc.)
- Besides the 2$^{nd}$ level of signal wrapping, the coding of previous generation signals (e.g., 100G or 400G mapped in 66B format) is likely to be less efficient than 800GbE or 1.6TbE that will likely be mapped in a format at least as compact as 257B
- Probably little that 802.3 can do that would help ITU-T with this problem

# Issue #5 – Coding and error marking considerations

- As a segue from the previous issue, the fact that recent Ethernet interfaces are 257B coded, and beyond 400 Gb/s Ethernet is expected to use a coding at least this compact (perhaps not even using 66B as a starting point), and the fact that prior-rate OTN mappings are 66B based is a driver for bit-rate differences
- 257B coding sacrifices the coding resilience of 66B (which doesn't help anyway on an interface with FEC), and trades instead for an error marking mechanism normally based on uncorrectable FEC codewords.
- When the FEC is soft-decision (likely for at least some FECs beyond 400 Gb/s), or when the FEC doesn't have high confidence in knowing a FEC codeword is uncorrectable, the method normally used is an inner, block-based CRC (not aligned with the MAC FCS) to detect when there are uncorrected errors as a basis for error marking: e.g., P802.3cw has a CRC32 inside the CFEC codeword, and P802.3bn used a CRC40 inside the LDPC FEC as the way to detect when the FEC decoder was unable to correct errors and the MAC frames need to be error marked.
- Since at least some FECs used for beyond 400 Gb/s Ethernet are likely to be soft-decision, it may make sense to develop a single, universal CRC-based method for detection of uncorrected errors that can be used for all PHYs. Such a mechanism would also serve the purpose of error marking for an Ethernet signal carried over OTN, even if multiplexed inside of a larger signal with no direct access to line-side FEC statistics

# Proposal

As there are many OTN-support issues that deserve further study, the Beyond 400 Gb/s Ethernet Study Group should adopt the following objective:

- Provide appropriate support for OTN

# THANKS!