

July 16th, 2024

Buffer in TDD PHYs

Presenter: Max Turner (max.turner@ieee.org)

Project: IEEE P802.2dm - ISAAC

Event: 2024 July 802 Plenary in Montreal, CAN

Contribution to:  IEEE

IEEE P802.3dm Task Force - ISAAC

ETHERNOVIA

IEEE Std 802.3ch-2020: 44.1.3 Relationship of 10 Gigabit Ethernet to the ISO/OSI reference model

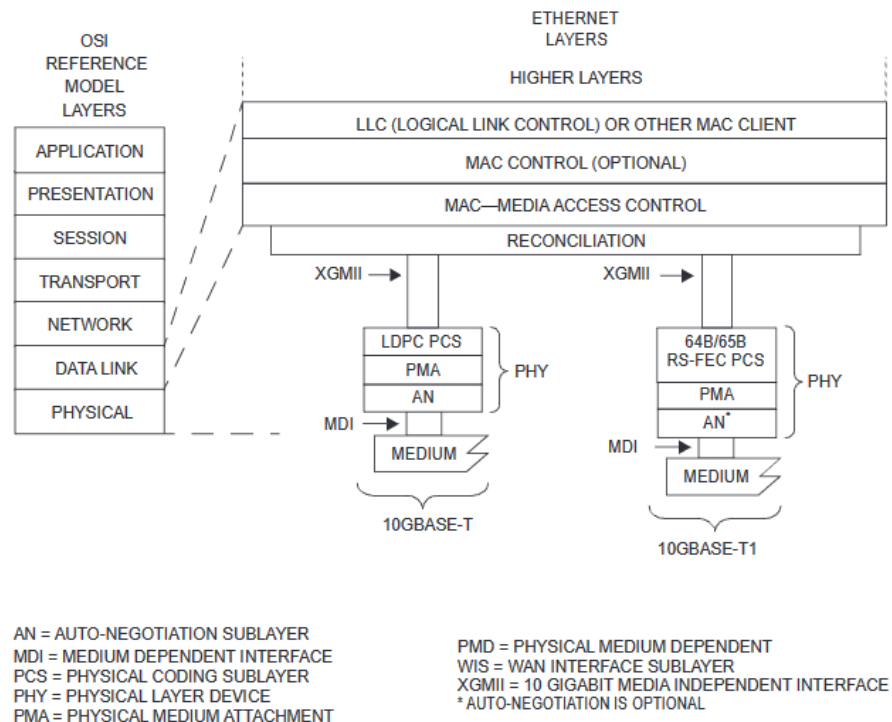
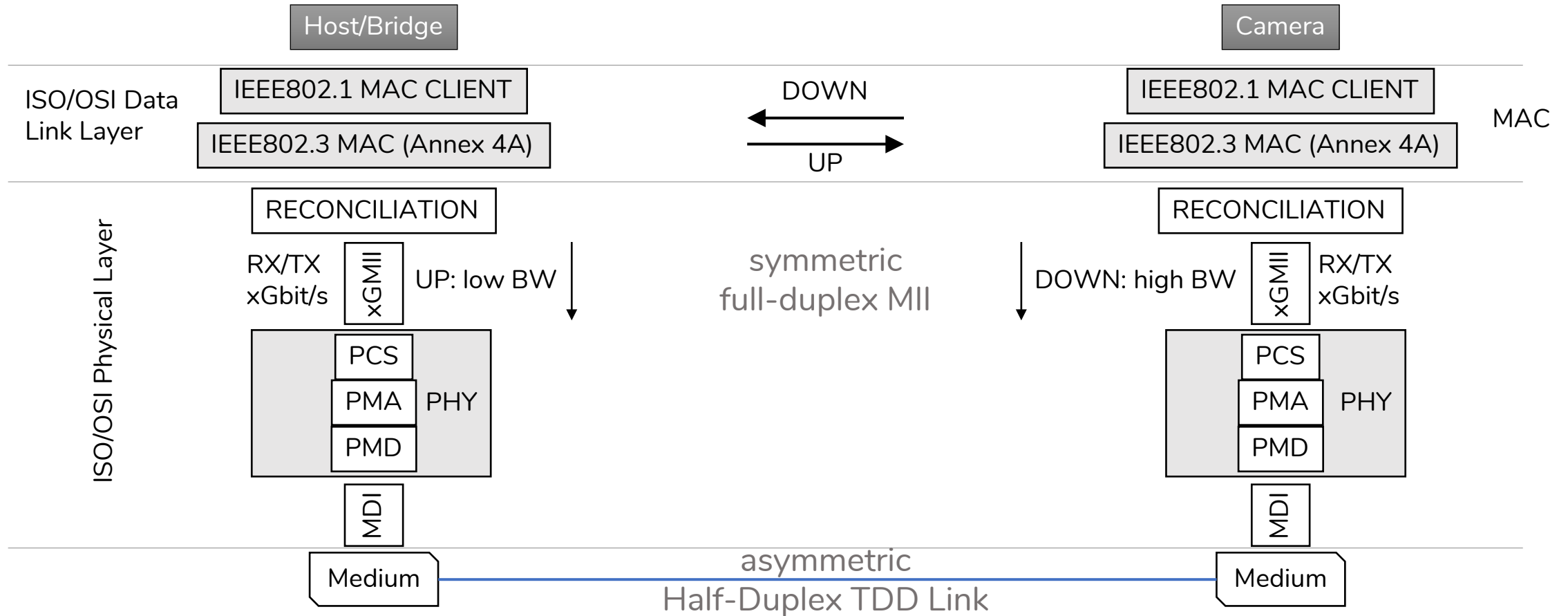
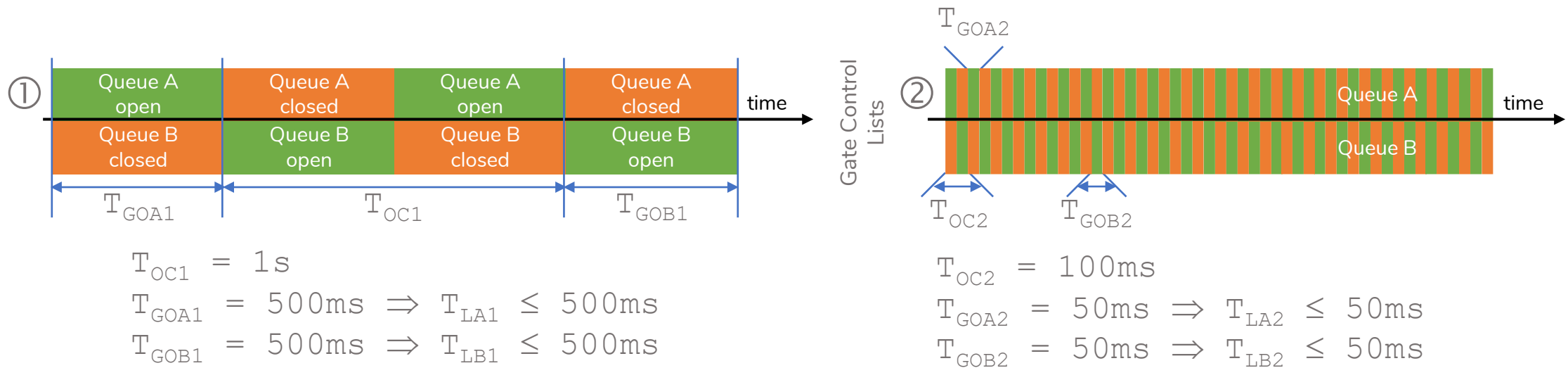


Figure 44-1—Architectural positioning of 10 Gigabit Ethernet

TDD with symmetric MII



TAS¹⁾: Delay vs. Bandwidth

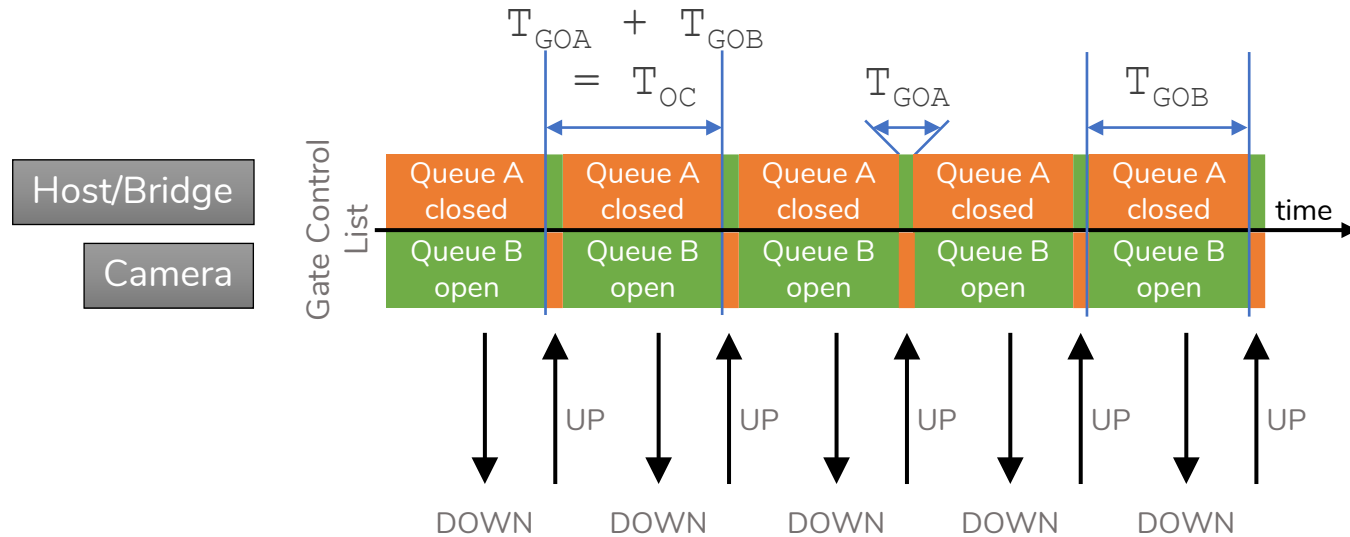


Maximum available TAS bandwidth depends on the Ratio T_{GO}/T_{OC} , but TAS Delay (T_{LA}) depends on the absolute difference of $T_{OC}-T_{GO}$!

T_{OC} ... OperCycleTime
 T_{GO} ... GateOpenTime
 $T_{OC} = gaps + \sum T_{GOx}$

¹⁾ IEEE Std 802.1Qbv™-2015 Enhancements for scheduled traffic

Regular TAS queuing



- The link runs symmetrically at line-rate R
- Queue A is served by the Host/Bridge to transmit the control data (low BW r_A) UP to the Camera
- Queue B is served by the Camera to transmit the image data (high BW r_B) DOWN to the Host/Bridge
- Since $r_B \gg r_A \Rightarrow T_{GOB} \gg T_{GOA}$ and $T_{OC} \approx T_{GOB}$

Max. Delay of Flow A (for any $T_{LA} > T_{GOB}$):

$$R \times (T_{LA} - T_{GOB}) = B_A + r_A \times T_{LA}$$

$$T_{LA} \times (R - r_A) = B_A + R \times T_{GOB}$$

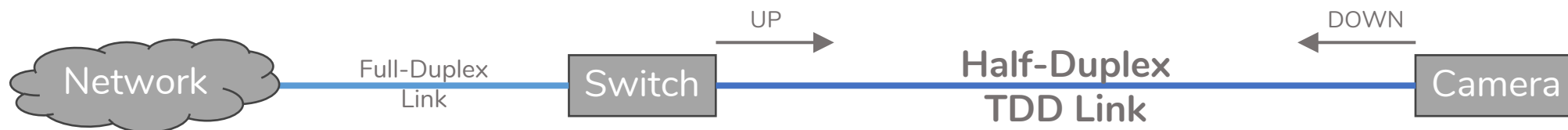
$$T_{LA} = \frac{B_A + R(T_{OC} - T_{GOA})}{R - r_A}$$

- A Burst of data (B_A) is waiting in Queue A as the Gate for A closes
- While the Gate for A is closed, more data is accumulated at (low BW) rate r_A
- When the Gate for A opens (after T_{GOB}), data is transmitted at line-rate R , while more data is accumulated at rate $r_A \ll R$
- Before the Gate for A closes again at the end of the cycle ($T_{OC} = T_{GOA} + T_{GOB}$), all that data must have been transmitted to have an empty buffer

A simplified TDD example

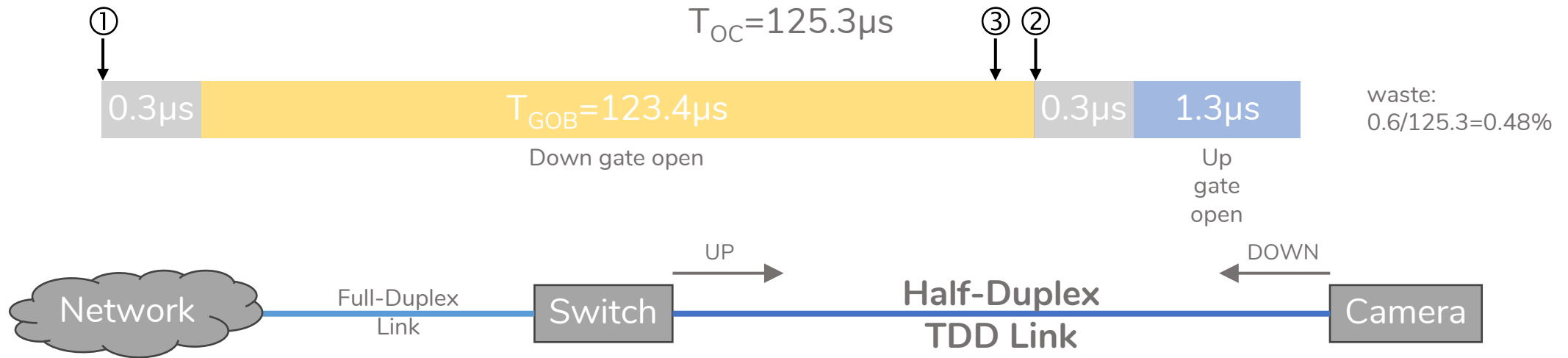
- Run the TDD-Link at 10Gbit/s in both directions.
- We want to allocate a 100× greater BW to the Down direction than to the Up direction.
- In order to not violate Ethernet fundamentals, the 1 time unit Up needs to support at least a 1542Byte Frame; so the Up gate needs to be open for at least $T_{GOA} = 1.2336\mu\text{s}$.
- Therefore: $100 \times 1.2336\mu\text{s} = 123.36\mu\text{s} = T_{GOB}$
- Adding re-sync gaps of $0.3\mu\text{s}$, one can round this to the following simplified TDD cycle:

$$\text{Up: } 123.4\mu\text{s} + \text{gap: } 0.3\mu\text{s} + \text{Down: } 1.3\mu\text{s} + \text{gap: } 0.3\mu\text{s} = 125.3\mu\text{s} = T_{OC}$$



Transmission delays in the simplified example

NOT to scale!

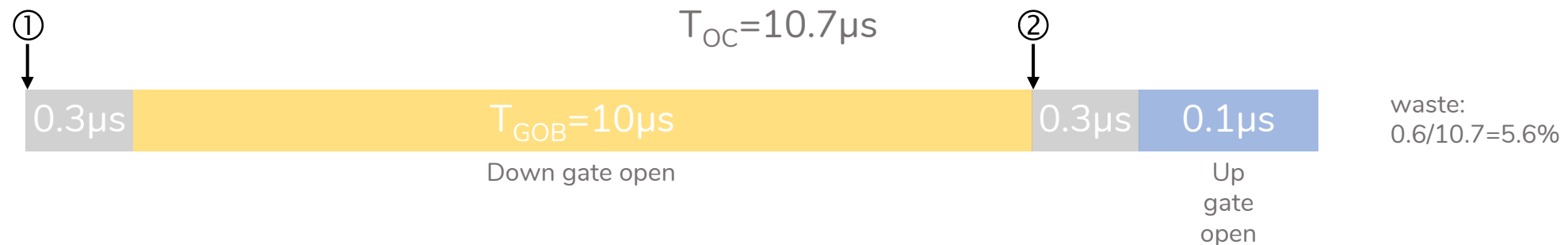


- An upstream Frame eligible for transmission on the Switch's TDD Link at instance ① of the TDD cycle must wait for $124\mu\text{s}$
- A downstream Frame eligible for transmission on the Camera's TDD Link at instance ② of the TDD cycle must wait for $1.9\mu\text{s}$
- Instance ③ of the TDD cycle lies the frame length before instance ②. I.e. the Frame is too long to be transmitted during the gate open time for Down, adding a maximum of $1.23\mu\text{s}$ to the delay at ②.

Does the Gate have to be open for a full Frame?

- If most upstream Frames are small, could one reduce the gate open time for Up to e.g. $0.1\mu\text{s}$?
- Down: $100 \times 0.1\mu\text{s} = 10\mu\text{s}$
- Unfortunately the gaps will likely not change, therefore the wasted time goes up
- And what to do if there is a larger Frame to be transmitted?

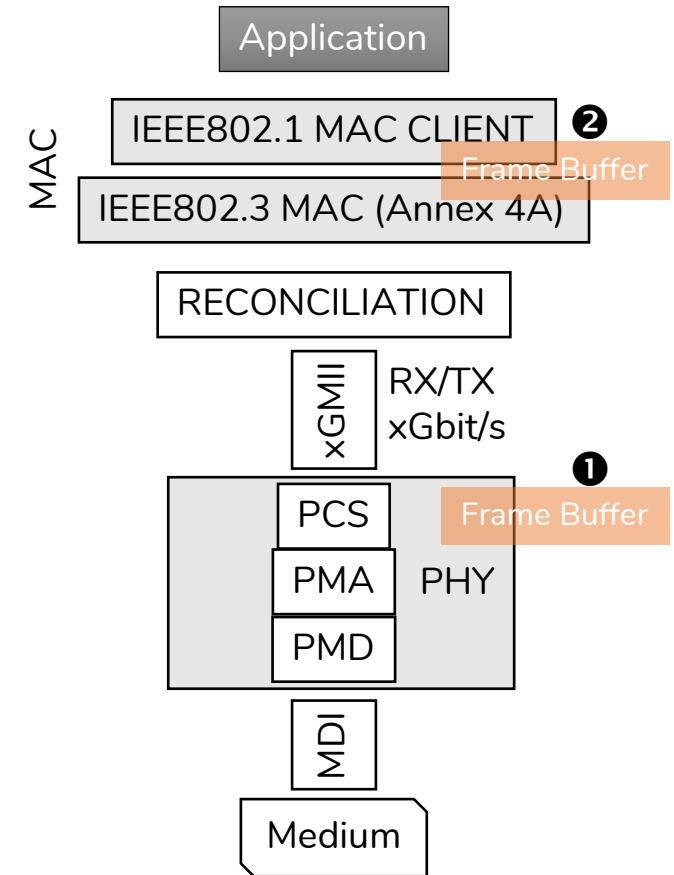
NOT to scale!



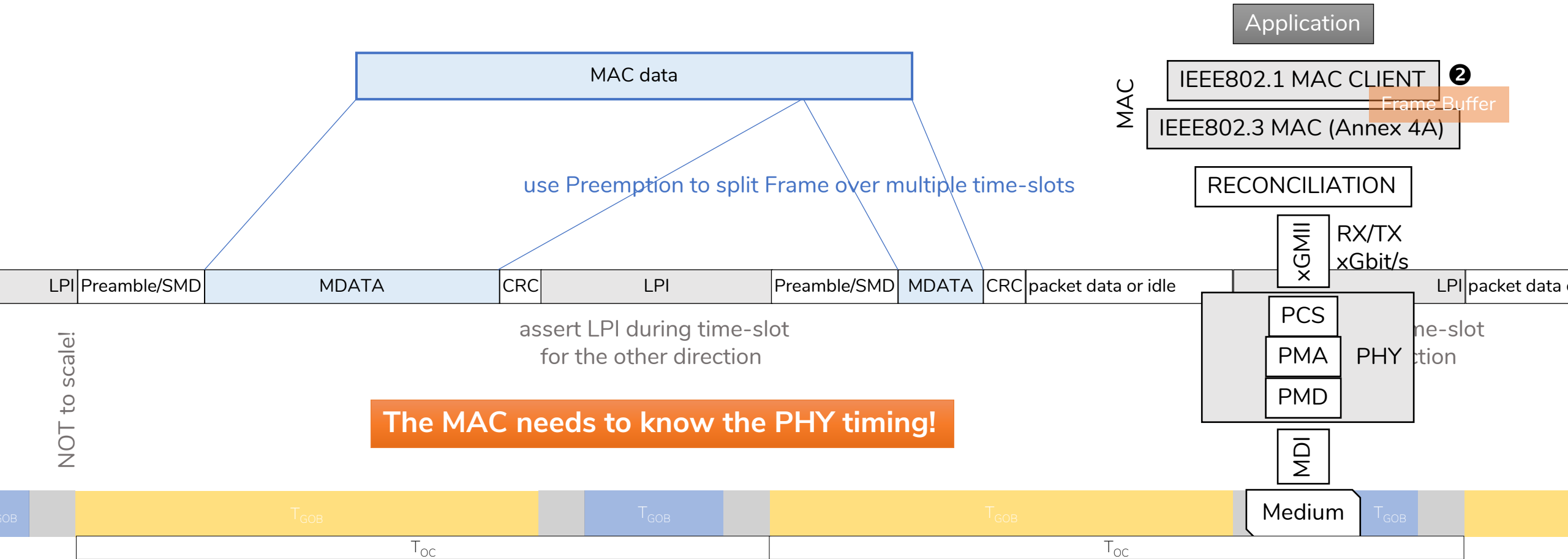
- Now a small upstream Frame eligible for transmission on the Switch's TDD Link at instance ① of the TDD cycle must wait for only $10.6\mu\text{s}$

Need a Frame Buffer

- If the Medium is not available for transmission of a full Frame, it must be buffered somewhere
- **Option 1:** Buffer in the PHY and use CSMA/CA half-duplex signalling from the PHY up to prevent new Frames to be transmitted across the MII
 - The PHY usually does not know about Frames
- **Option 2:** Buffer in the MAC and ...
 - Make sure only full Frames traverse the MII
 - No Idle-Symbols traverse the MII, when the Medium is not available



Using Preemption and EEE's LPI



Conclusion


- Full-Duplex operation would be way nicer to handle ...
 - No Buffering
 - No wait times to access the half-duplex TDD controlled Medium



Max Turner

Utrechtseweg 75
NL-3702AA Zeist
The Netherlands
+49 177 863 7804

max.turner@ethernovia.com


Contribution to:  **IEEE**

IEEE P802.3dm Task Force - ISAAC



ETHERNOVIA

Thank You
for your attention!

Contribution to:  **IEEE**

IEEE P802.3dm Task Force - ISAAC