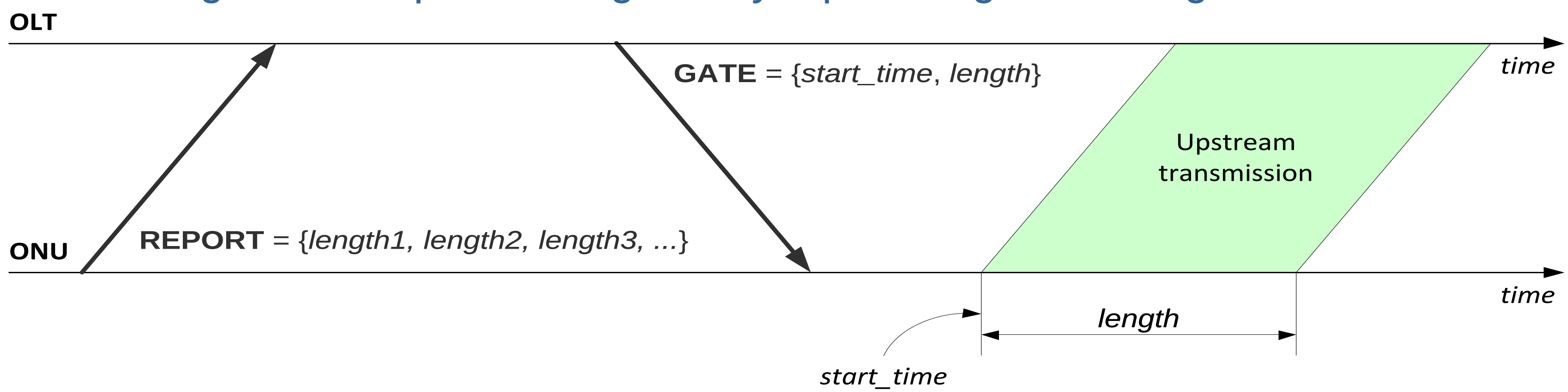
100G-EPON:
Channel Bonding Placement Issues

Glen Kramer, gkramer@broadcom.com

Main MPCP function

- Provide centrally-coordinated access to P2MP medium
  - Downstream is trivial. Main challenge is upstream
- So far, MPCP relied on Reporting/Granting scheme
  - **ONU Reporting process:** inform the central arbiter of the ONU's queue status and provide several "suggested" locally-optimal grant lengths (thresholds)
  - **OLT Granting process:** based on provisioned service levels, state of all the ONUs (queues), and ONUs' locally-optimal grant lengths, compute the globally-optimal grant assignments.

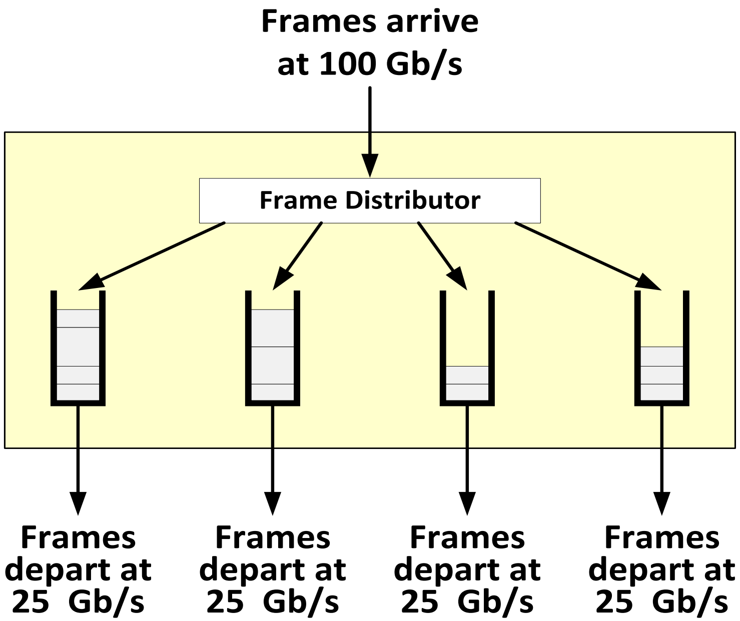


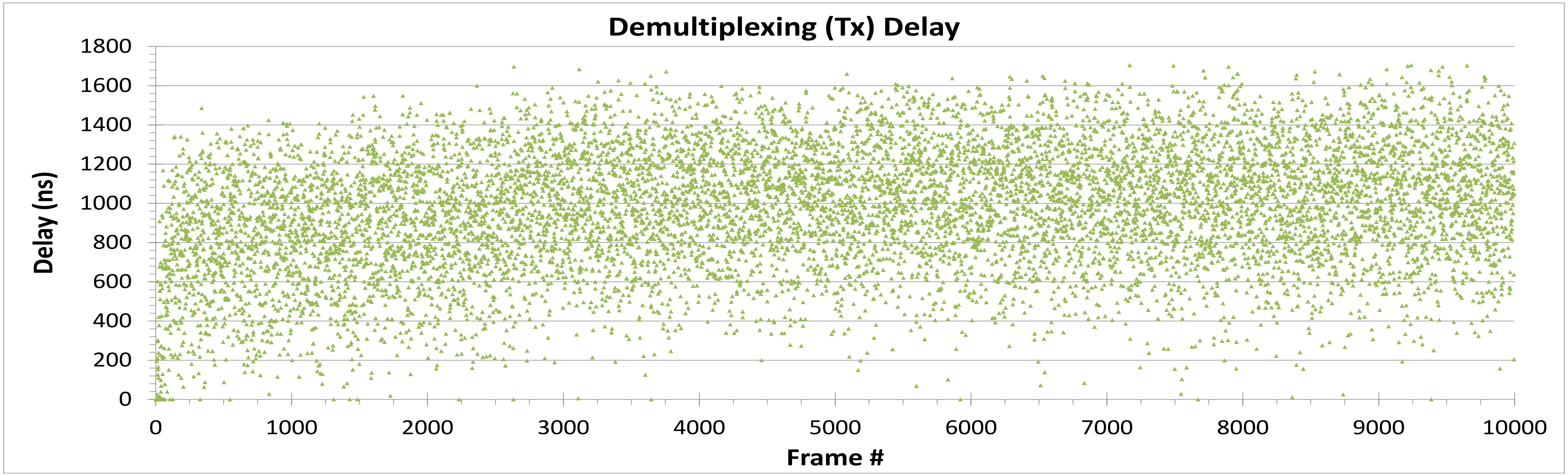
MPCP Requirements

- MAC provides services to MPCP. The layers below MPCP exist to accommodate the MPCP requirements
- Key MPCP requirements to lower layers:
  - Constant propagation delay from Tx MAC Service interface (first bit in) to Rx MAC Service interface (last bit out) for timestamped MPCPDUs.
  - For each pair of MAC Service Interfaces, the lower layers do not change frame order
  - Frame loss ratio equivalent to BER $\leq 10^{-12}$
- MPCP does not know and does not care whether a frame is transmitted over a single lambda or striped across multiple lambdas.

Key Takeaways from kramer_3ca_1a_0116.pdf

- **Frame Demultiplexor adds variable delay**
  - 2000-byte max frame: delay **0 ~ 400ns**
  - 9000-byte max frame: delay **0 ~ 1.7μs**
- **Unknown and hard to predict which lane will be used for a given frame**





Focus of this presentation

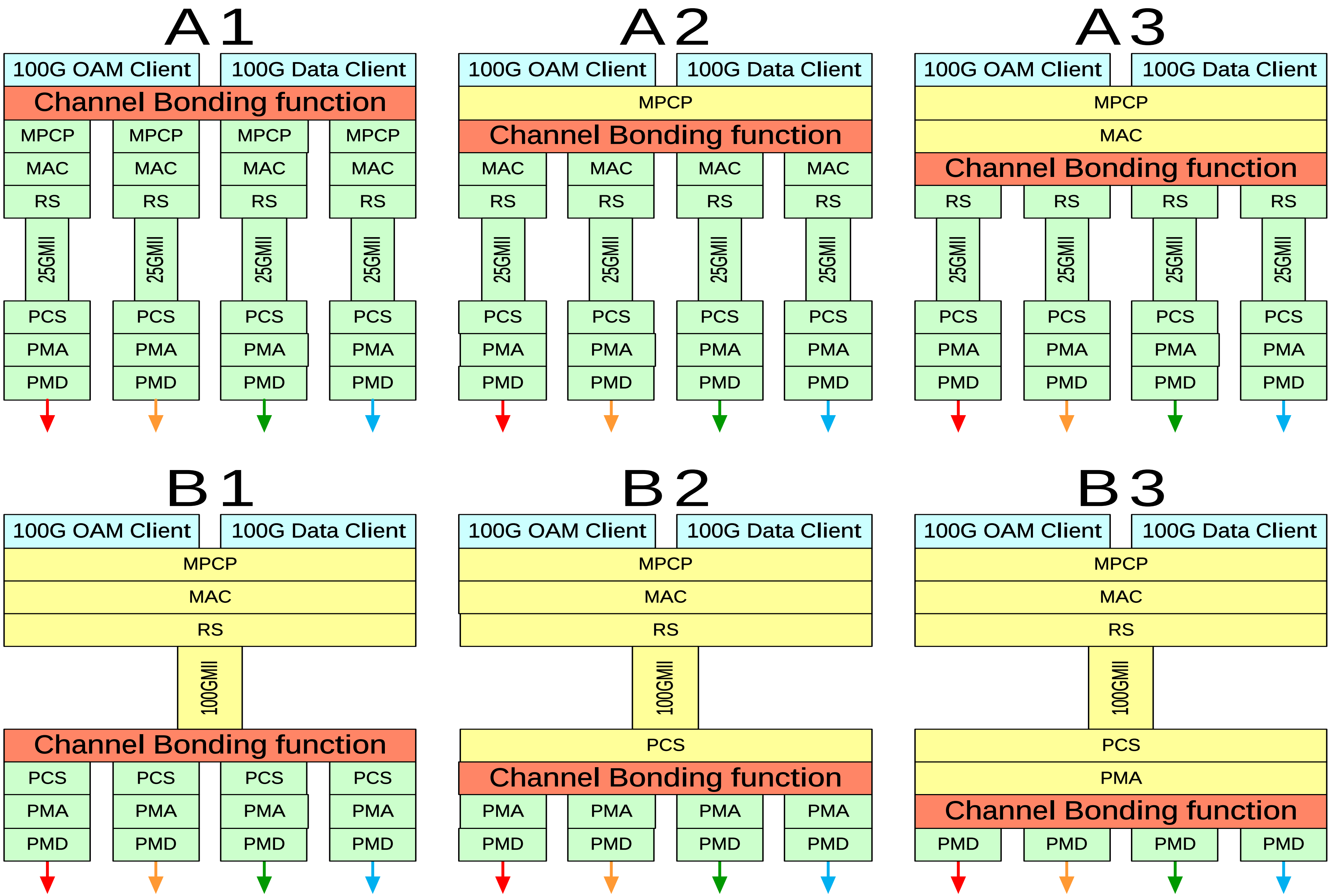
- This presentation focuses on the most challenging part – 100G upstream operation
  - 50G and 25G are easier (subset of 100G upstream)
  - Downstream is easier (like upstream, but without gating)

- In Dallas meeting, it seemed the following consensus started to emerge:

> - **The ONU can transmit on any or all of its equipped channels**
> - **To achieve the required peak rate, multi-channel ONUs must be able to transmit on multiple channels at the same time**
> - **However, simultaneous transmission is not needed always**
>   - **Use cases: power saving, uneven channel loading, unequal start-time from the last transmissions on each channel**
> - **To best utilize the multiple channels, the MPCP gating process should be capable of independent control and transmission on each channel**

See ngepon_1511_effenberger_1.pdf

Where to put channel bonding?



Issue #1 – MAC Speed

- With Channel Bonding located above MAC sublayer, MAC runs at 25 Gb/s
- Objectives require "*supporting symmetric and/or asymmetric MAC data rates of <…> 100 Gb/s in downstream and less than or equal to 100 Gb/s in upstream.*"
- To have MAC operate at full data rate, the Channel Bonding should be located below MAC.
- Maybe we can have another virtual 100Gb/s MAC above the Channel Bonding sublayer, but this becomes 802.1 realm rather than 802.3 realm

Issue #2 – MPCP Synchronization

- MPCP *localTime* counters in the OLT and ONUs re-synchronize upon the reception of each MPCPDU.
- The synchronization mechanism relies on near-constant latency from transmitting MPCP sublayer to receiving MPCP sublayer
  - ONU deregisters if the propagation delay changes by >8 TQ (128 ns).
- Channel Bonding introduces jumps in end-to-end delay, some as high as 100+ TQs.
- MPCPDU timestamping should take place below Channel Bonding

Issue #3 –Tx control

NG-EPON

- MPCP has a very limited signaling capability to control PHY transmission.
- In 10G-EPON, ONU's laser is controlled by presence or absence of data on the XGMII
  - Last N blocks are idles → Turn (or keep) the laser off
  - Else → Turn (or keep) the laser on.
- But for the multi-lane ONU, this approach requires a separate (x)MII for each lane.
- Channel Bonding should be located above (x)MII

Issue #4 – Laser On/Off Timing

- In current EPON architecture, MPCP controls the timing of each burst by starting and stopping the data flow at specific times.

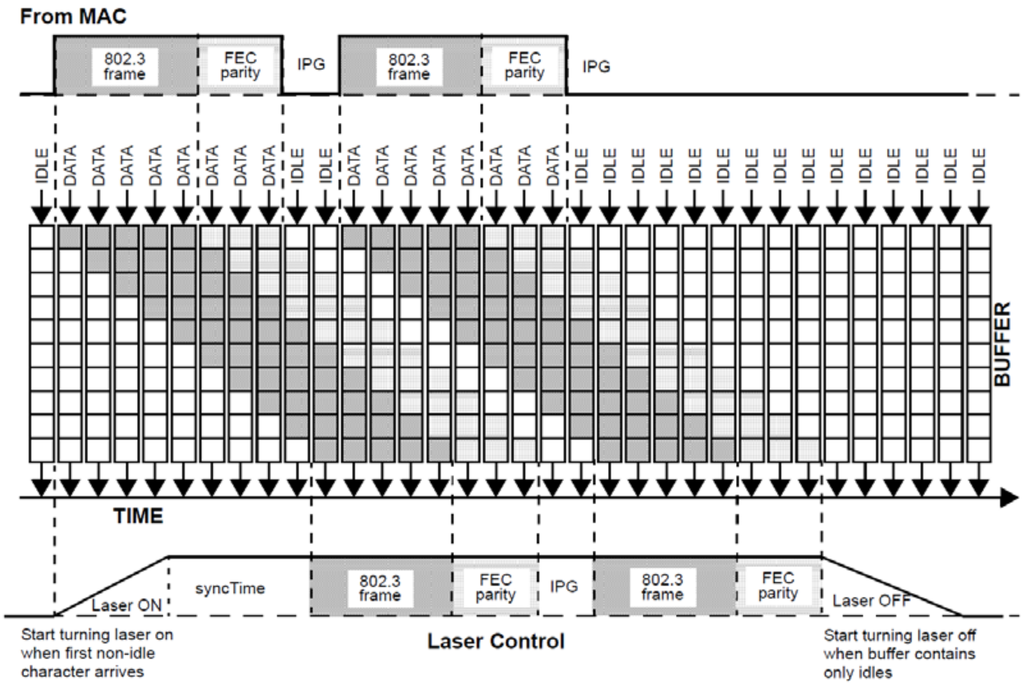


Figure 65–5—Laser control as a function of buffer fill

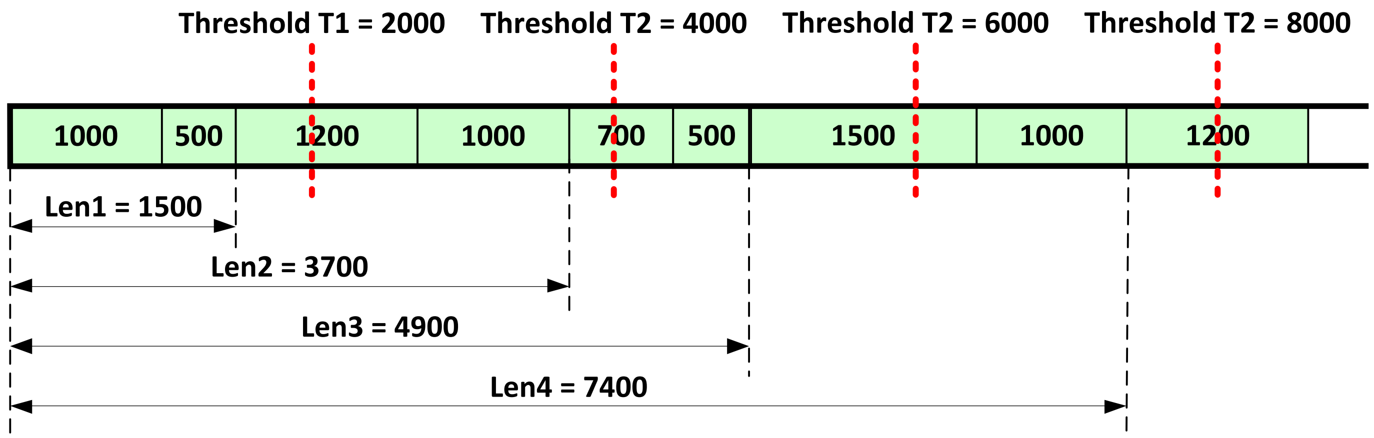
- Channel Bonding (Demultiplexor) introduces variable delay of 0~400ns (2KB max. frame size) or 0~1.7μs (9KB max. frame size)
- Channel Bonding must be located above MPCP

Issue #5 – Frame order

- Channel Bonding may introduce frame reordering.
- For the Multiplexor to be able to restore the original frame order, frames need to include a sequence number.
- If we don't want to mess with the content of user frames, then the sequence number should be carried in the preamble.
- Preamble is not accessible above MAC. In EPON, Reconciliation sublayer overwrites preamble.
- The Channel Bonding should be located below MAC, but above (x)MII.

Existing Reporting / Granting scheme

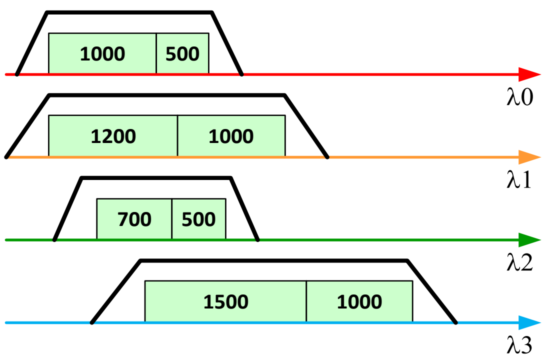
- Once the ONU places a frame in a queue, the frame's order of transmission is set.
- The ONU now can report multiple thresholds based on the fixed order of frames in the queue.



- The OLT can issue a single grant of size 1500 or 3700 or 4900 or 7400
- Or it can issue multiple sequential grants based on deltas: 1500 + 2200 + 1200 + 2500 (in this order)
- All these grants will be packet boundary aligned.

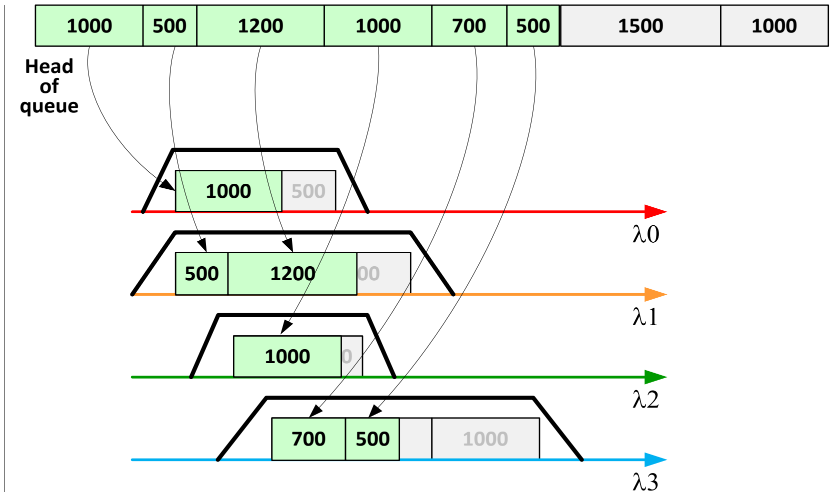
Issue #6 – Grant Packing

- Channel bonding requires ONU to transmit on multiple lanes at the same time. If the ONU keeps frames in FIFO order while packing overlapping grants, the previously-reported frame boundaries become meaningless.



Overlapping grants – desired frame distribution

- OLT has granted 4 grants expecting ONU to send 8 frames with a total of 7400 bytes.



Overlapping grants – actual frame distribution

- ONU had to dequeue frames sequentially. In this example, the ONU could send only 6 frames with a total of 4900 bytes (34% bandwidth wastage)

- Channel Bonding (Demultiplexor) in the ONU must be located above the MPCP (Reporting and transmitting functions).

No Place for Channel Bonding

Technical Issue		Bonding is above xMII			Bonding is below xMII		
		A1	A2	A3	B1	B2	B3
Issue #1	MAC runs at 100 Gb/s	No	No	Yes	Yes	Yes	Yes
Issue #2	MPCP time can be synchronized between the OLT and ONUs	Yes	No	No	No	No	No
Issue #3	ONU is able to control lasers independently for each lane	Yes	Yes	Yes	No	No	No
Issue #4	ONU is able to turn each laser on/off at correct times	Yes	No	No	No	No	No
Issue #5	ONU is able to insert frame sequence number in preamble	No	No	Yes	No	No	No
Issue #6	ONU is able to pack grants based on previously-reported packet boundaries	Yes	No	No	No	No	No

Other questions

- **Synchronization**
  - **Existing requirement:** "ONUs shall operate at the same time basis as the OLT, i.e., the ONU TX clock tracks the ONU RX clock, which in turn locks to OLT TX clock."
  - In multi-wavelength case – which signal to lock on? Does skew matter for MPCP?
  - Should the ONU be resilient to a loss of one or several wavelengths? If yes, should we allow synchronization on any wavelength?

Conclusion

- Multi-Lane MPCP operation is a major challenge.
- We cannot focus on RS and PCS functions until we resolve the multi-lane scheduling issues.
  - Don't know yet if these layers need to run at 25 Gb/s or 100 Gb/s
- Solving the Multi-Lane Reporting/Granting issue may require rethinking of the entire EPON architecture.
- We may need to consider drastic departures from the existing MPCP model.
- We may need to revisit some old assumptions/decisions:
  - Frame atomicity
  - Granting granularity
  - Signaling from MAC Control to PHY
  - etc.

Main conclusion

Everyone should drop what they are doing and work on the multi-lane scheduling issues.

Seriously.