Laning approach for 802.3cy

George Zimmerman

CME Consulting, Inc / Marvell

Acknowledgement

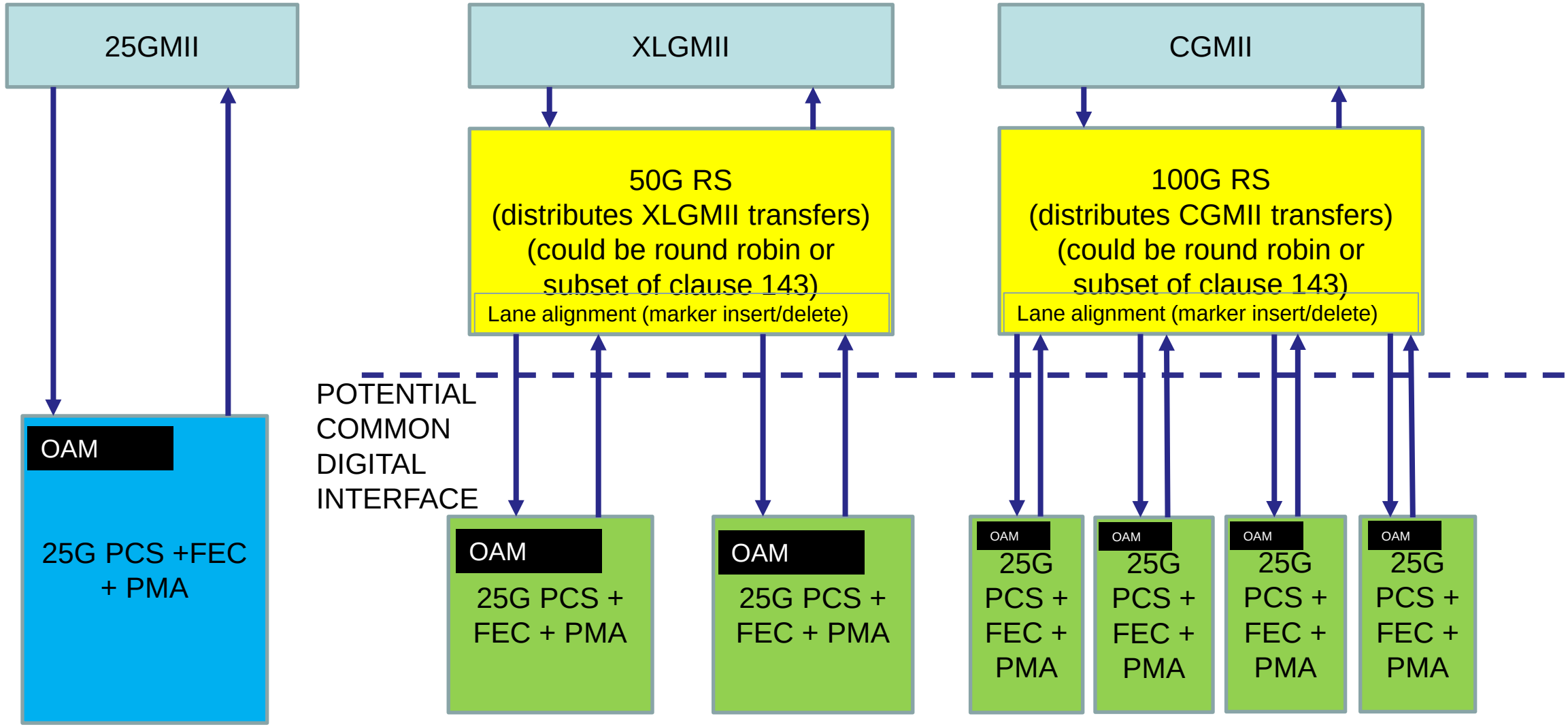
- A special thank you to:
  - Marek Hajduczenia
  - Glenn Kramer
  - Natalie Wienckowski

  For their patience, good advice, and insights on how we might adapt clause 143 for laning 25GBASE-T1 PHYs into a 50G or 100G link

Basic Laning Approach

- Several choices at first, one at last:
  - ~~Lane PMA only as a unit? (combine at FEC) (like Cl 55)~~
    - ~~Similar to BASE-T model, although FEC isn't separate sublayer~~
  - ~~Lane PMA & FEC as a unit? (combine at PCS) (like Cl 91 & 94)~~
    - ~~Allows integration and repetition of a PMA/FEC with independent BER~~
  - Lane PMA/FEC/PCS as a unit? (combine at RS) (Cl 143)
    - Allows independent PHY units to be bonded
    - PCS & FEC can still be internally laned if needed, independent of PMA

Lane PMA + FEC + PCS



Key Feature Decisions

- Architecture: PHYs joined at RS layer, PCS/PMA separate at 25G
- PHY-level Interface:
  - 2x or 4x 25 Gbps – is this 25GMII?
- How much smarts to put in the RS?
  - Need coordinated control of Resets and EEE modes modes
  - Need skew correction, alignment
  - Need lane swap correction
- Clause 143 has a lot of this, but more, and more complication
  - Designed for EPON
  - Uses MAC Control primitives specific to EPON

Clause 143 architecture

- Designed to work controlled by the MPMC
  - EPON MAC control sublayer provides MCRS_CTRL primitives not in Clause 4 MAC
- MCRS binds a MAC to multiple xMIIs
  - Each MAC is an “LLID” talking to multiple PCS/PMAs
  - Directions are independent
- MCRS converts serial MAC transmit data into parallel PHY streams
- MCRS maps xMII signals into individual MAC primitives
- Generates & expects continuous data & control characters

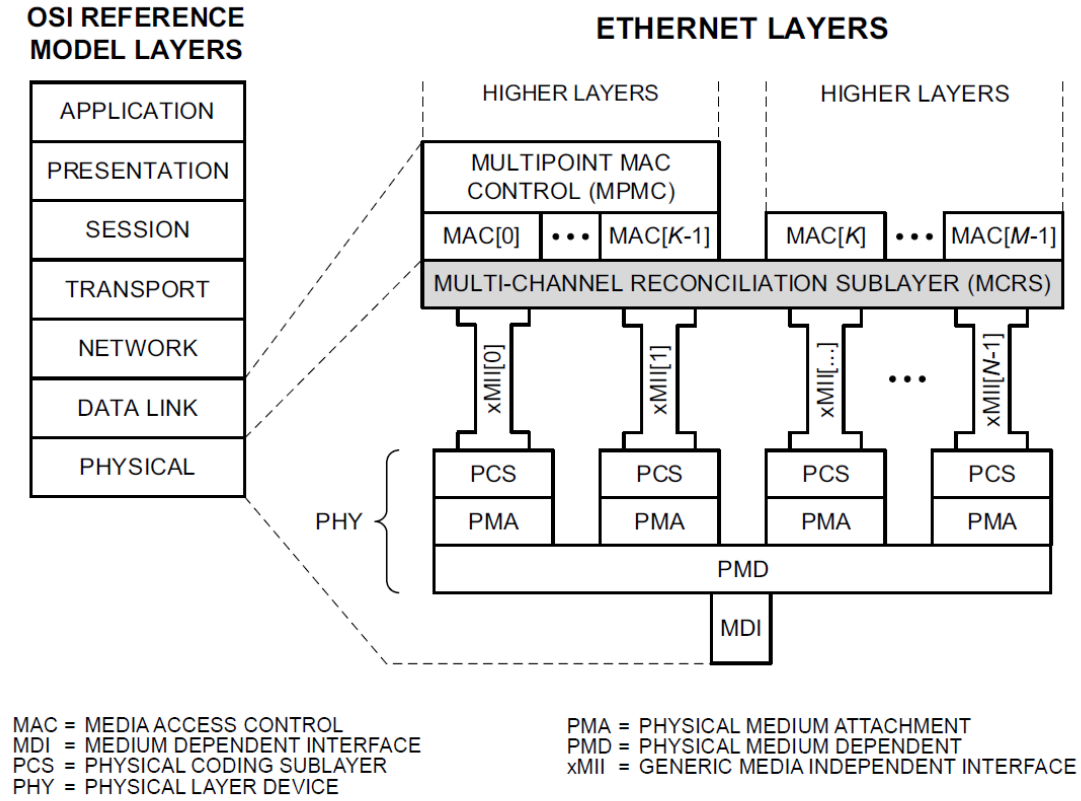


Figure 143–1—Relationship of MCRS to the OSI reference model

Source: IEEE P802.3dc D3.0, 143.1, 143.2

Clause 143 Envelope Quantum

- Clause 143 divides data into ‘Envelopes’
  - **Envelope** is a continuous transmission by a specific LLID (MAC iD) on a specific downstream / upstream channel
  - One EQ is 72 bits – 64 data and 8 control

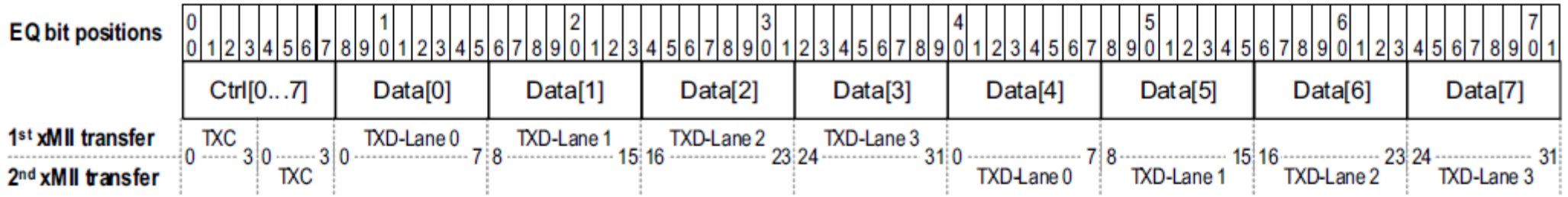
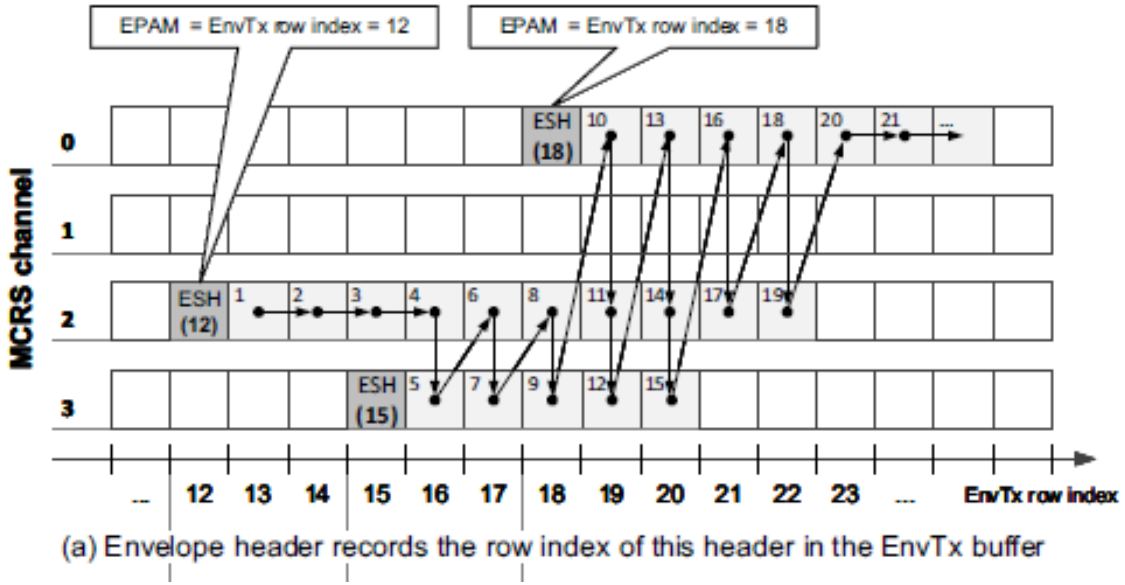


Figure 143–2—Envelope quantum (EQ) format

Source: IEEE P802.3dc D3.0

Envelope Headers Mitigate Skew

- Envelope Start Header records order of envelope at the transmitter
  - "EPAM" field stores index
- Controls position stored in receive buffer
  - Envelopes read out in order
  - (buffer depth dependent on skew to tolerate)



Source: IEEE P802.3dc D3.0

Envelope Header - detail

- Contains:
  - Start control code (0xFB)
  - EnvType flag bit
    - indicates start vs. continuation
  - EPAM (position alignment)
  - CRC8
  - Rest may not be needed by .3cy:
    - 22-bit Envelope Length field (suggest fixed-length)
    - two bits (E and K) reserved for encryption purposes
    - an LLID field (multiple MACs)

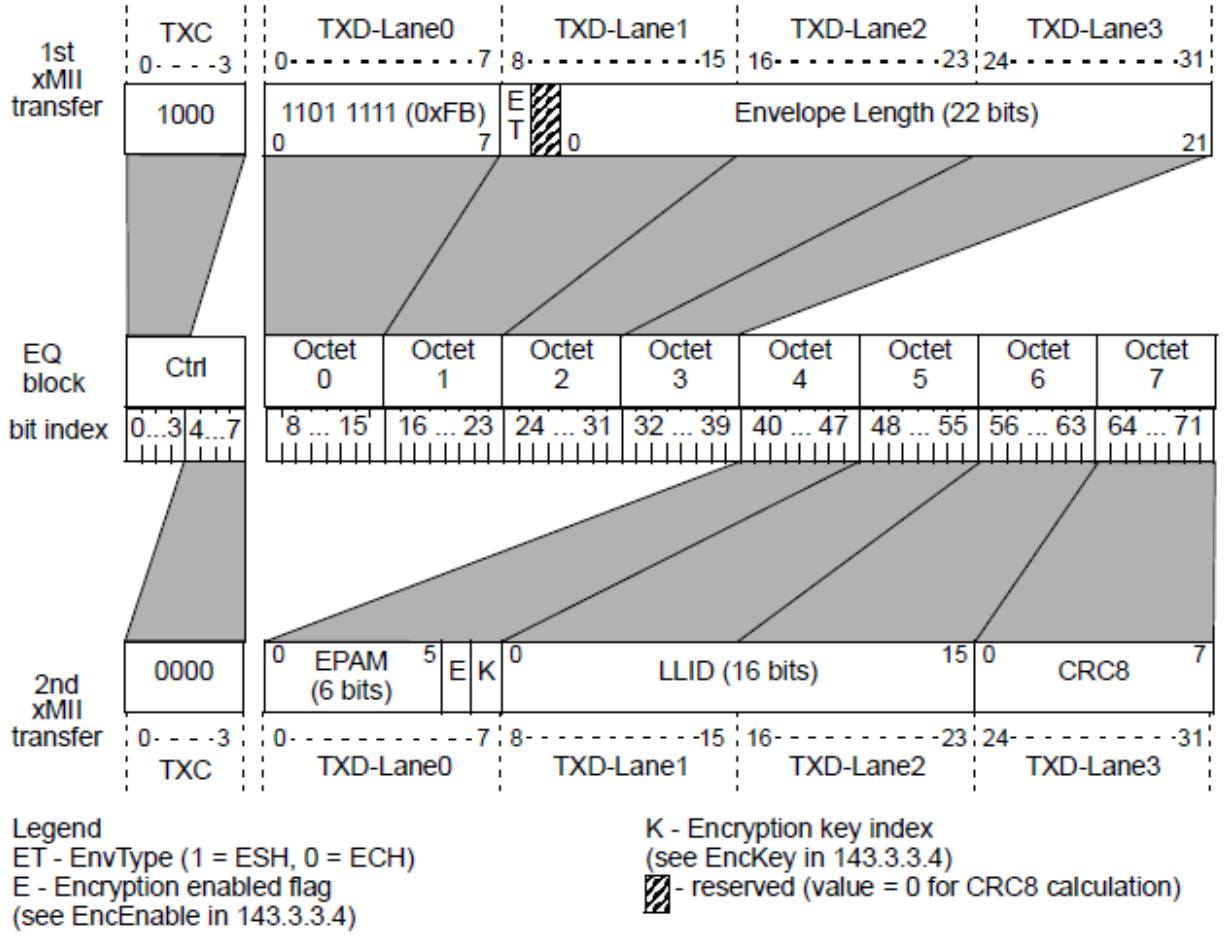
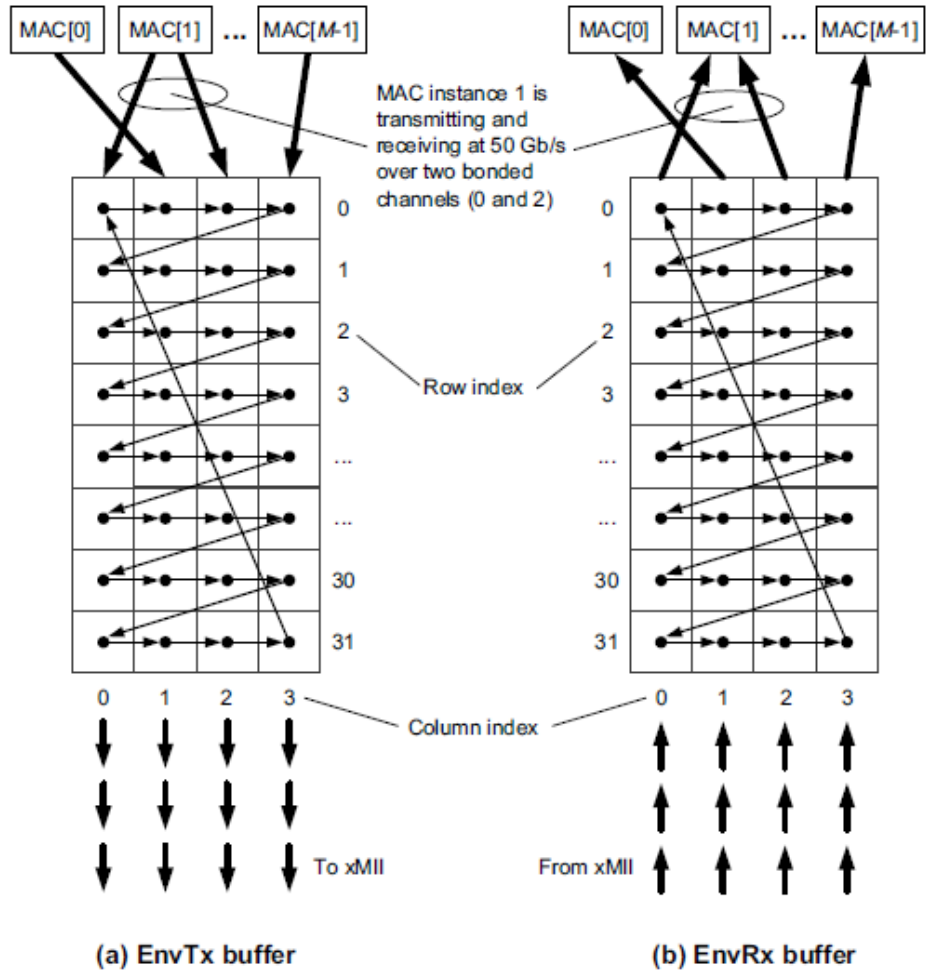


Figure 143–10—Mapping of envelope header fields into two xMII transfers

Source: IEEE P802.3dc D3.0

Clause 143 is more general than we need

- Allows multiple MACs to share multiple channels in a flexible way
- Allows dynamic rate bonding as traffic demands
- Enables buffering to accommodate FEC overhead
  - EPON PHYs pause data to accommodate FEC overhead



Source: IEEE P802.3dc D3.0

Figure 143–7—Internal structure of EnvTx and EnvRx buffers

Alternative

- Define a completely new laning
  - High speed laned PHY alignments are generally in the PCS, not the MII
  - Define alignment markers in the PHY frames
    - Unnecessary overhead in every 25Gb/s Ethernet link
  - Define new interfaces for the 25G PHYs when laned
    - Possibly different than 25GMII

Key parameters simplify

- EPAM records the relative position of an envelope
  - 6 bit field which controls the receive buffer positioning
  - EPON lanes could have a lot more skew than .3cy
    - But frequency lanes can't cross… so it is 'row' (skew) only
  - 802.3cy can divide the 6 bits of EPAM as row and column (lane)
- ADJ_BLOCK_SIZE, RATE_ADJ_SIZE can be zero
  - No FEC overhead to allow, no rate-matching needed
- LLID (for multiple MACs) is not needed – can be reserved

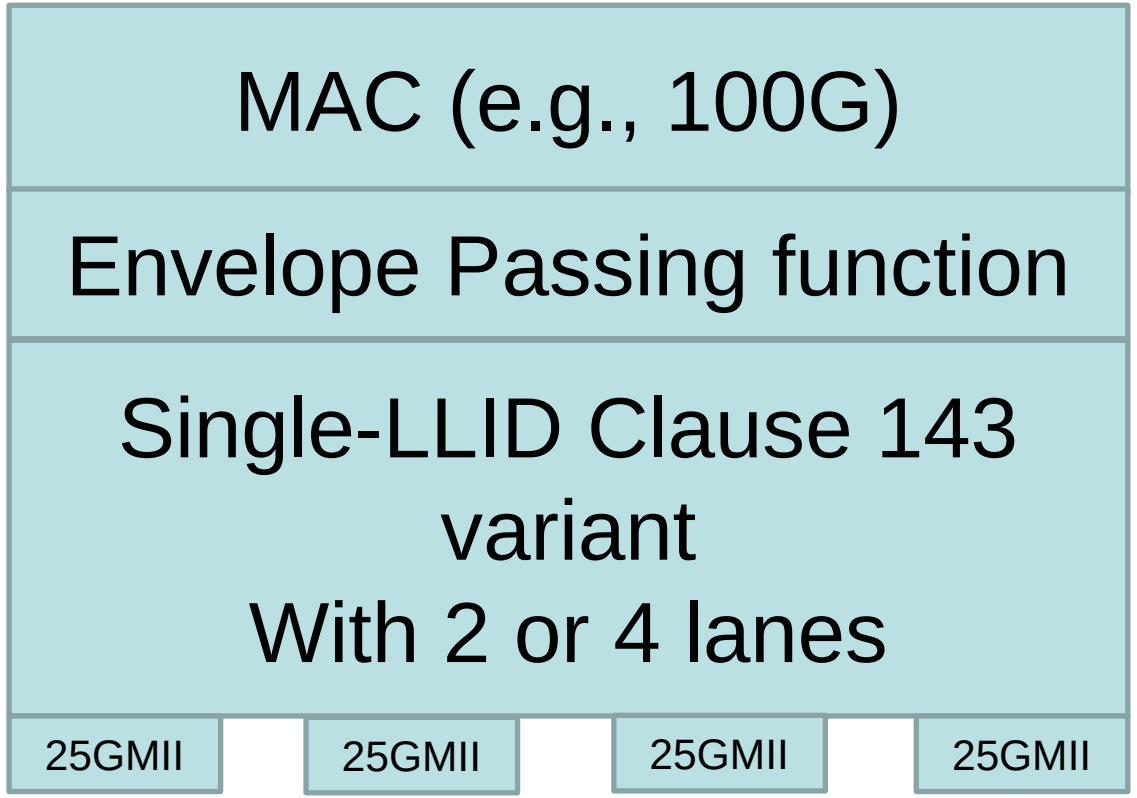
BUT… We have to add something

Requires MPMC (EPON MAC Control) or some other modification to generate MAC control primitives used in Cl 143 state diagrams

- **MCRS_CTRL[ch].request(link_id, epam, env_length) primitive**
  - Generated by the MPCP to request the next envelope transmission
- **MCRS_CTRL[ch].indication() primitive**
  - Generated by the MCRS to indicate that a given channel (lane) is ready for an envelope
- **MCRS_ECH[ch].indication(Llid) primitive**
  - Generated by the MCRS that an envelope has been received – generates the local timestamp

Possible Path forward

- Add a function to pass envelopes to the RS to have a simpler MCRS?
  - Only needed on the transmitter
  - Receive works normally – just needs TSSI interface
- Function can be in MAC/switch chip
- Enables reuse of interface and stand-alone PHY



Key simplifications & Path to Completion

- Fixed Envelope Size
- Fixed number of channels
  - No fallback to lower rate if one lane fails
- No need for rate adaption
  - FEC overhead accounted in PHY baud rate
- Result: No MCRS primitives?
  - Simplify MCRS transmit function to eliminate need for MCRS

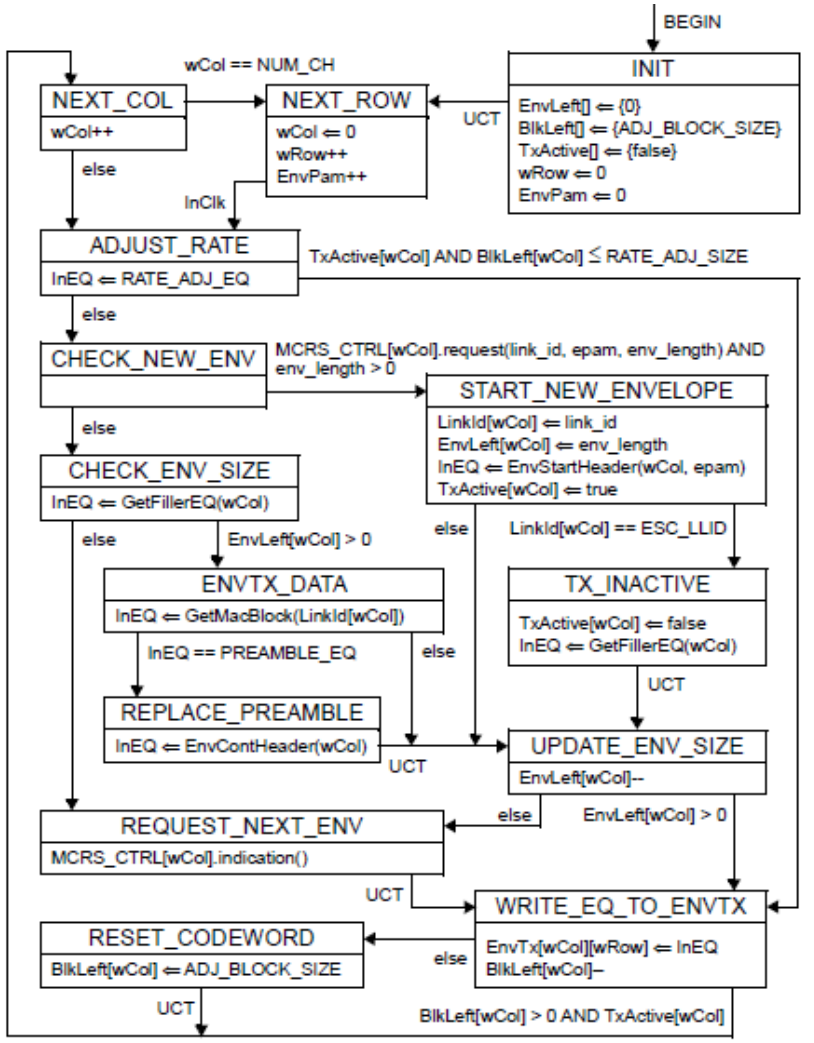


Figure 143–12—MCRS transmit function, Input process state diagram

Source: IEEE P802.3dc D3.0

Redrawing the Tx Function (concept – soliciting contributors)

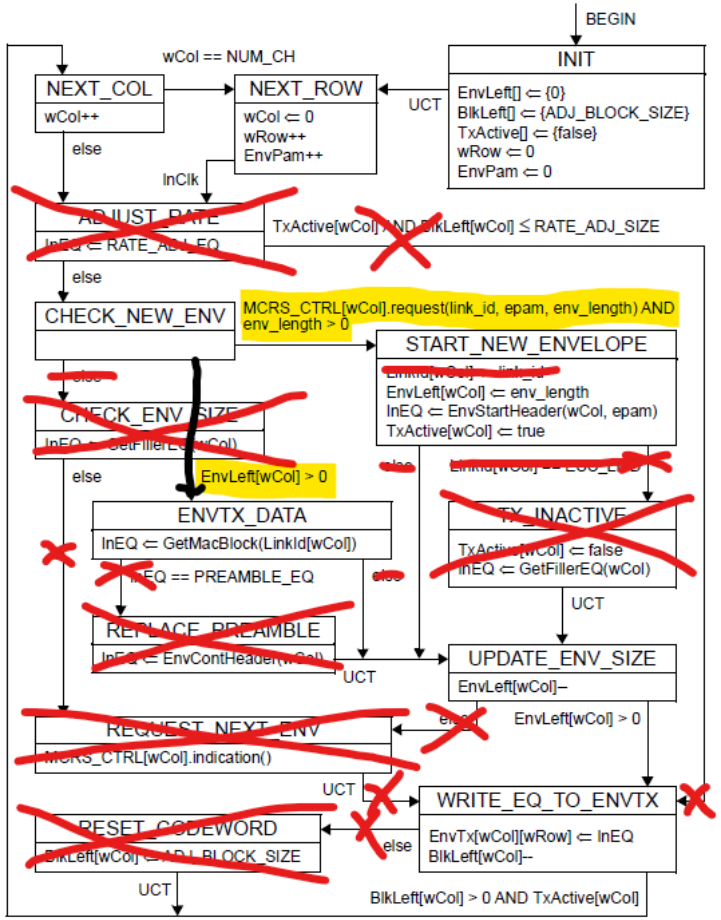


Figure 143–12—MCRS transmit function, Input process state diagram

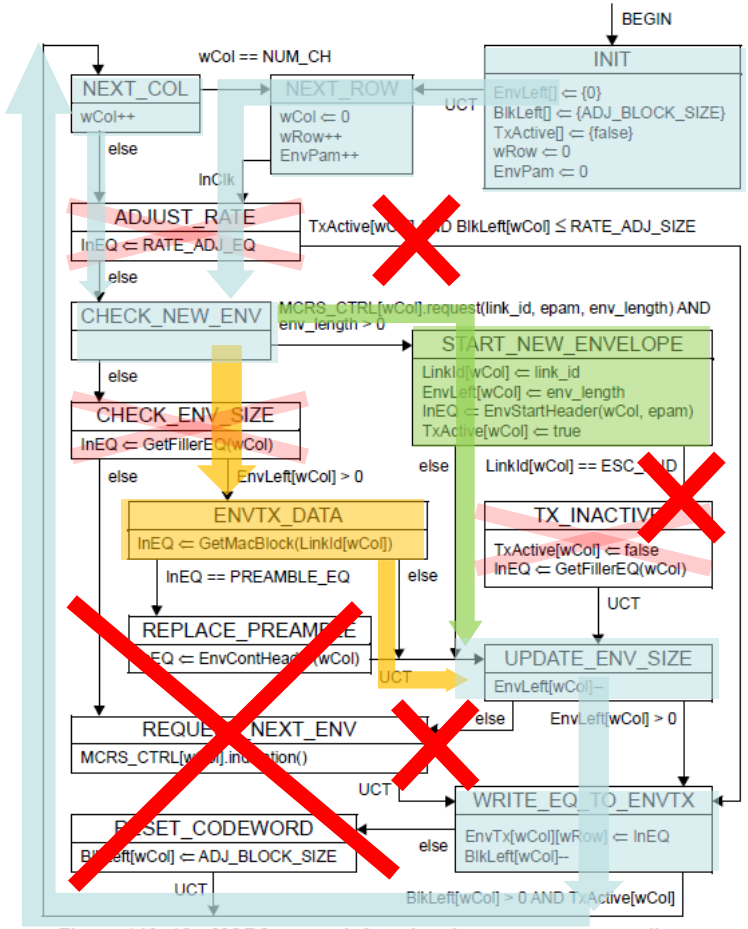


Figure 143–12—MCRS transmit function, Input process state diagram

Discussion?

THANK YOU!