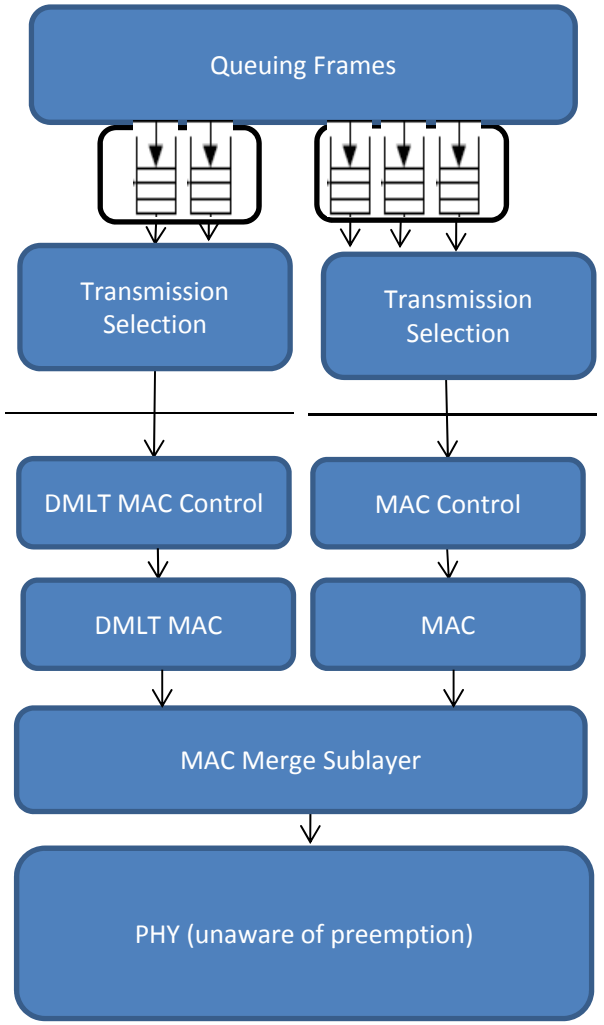
DMLT Negotiation
A Technical Feasibility Presentation

2013-05-09

Pat Thaler

MAC Merge Negotiation Goals

- Determine if the link partner supports MAC Merge
- Prevent MAC Client Frame Loss during the transition to MAC Merge Operation
- After enabling, check that MAC Merge is operational and fall back to normal operation if unsuccessful
- Independent of PHY



Existing Solution Space

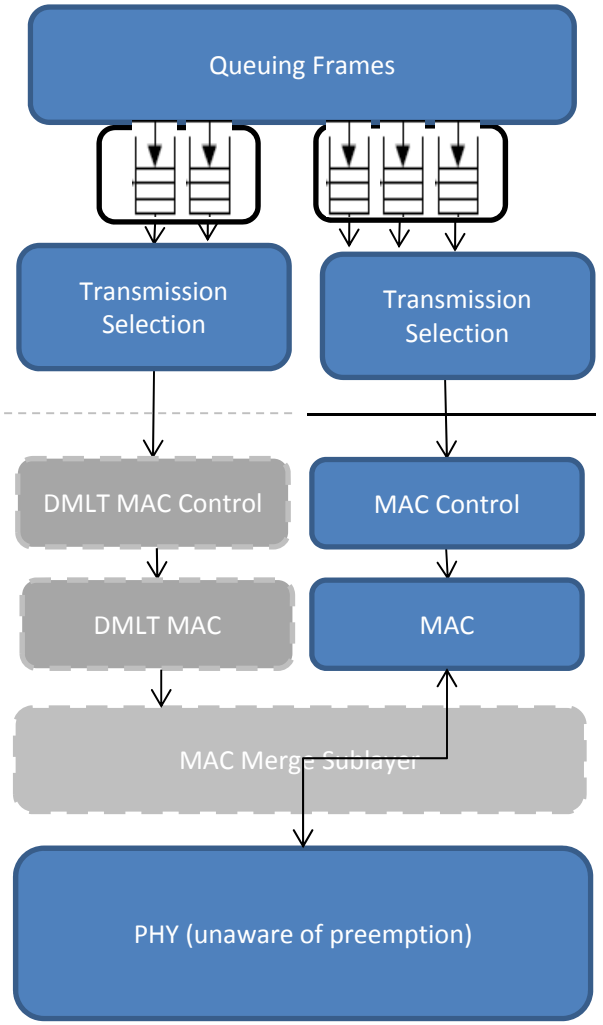
- LLDP
  - more suitable for capability exchange rather than negotiation,
  - not tightly coupled to MAC and PHY and
  - Doesn't provide for prevention of traffic loss during transition
- Auto-negotiation
  - Only supported on a subset of PHY types, e.g. not on optical PHYs except for 1 Gig.
- Slow protocols, e.g. similar to LACP
  - Slow limitation not needed and might delay detection of successful transition

MAC Control based negotiation

- Use MAC Control frames for negotiation
- If a transition to MAC Merge is enabled, MAC Control can block data frame transmission until the transition is successful

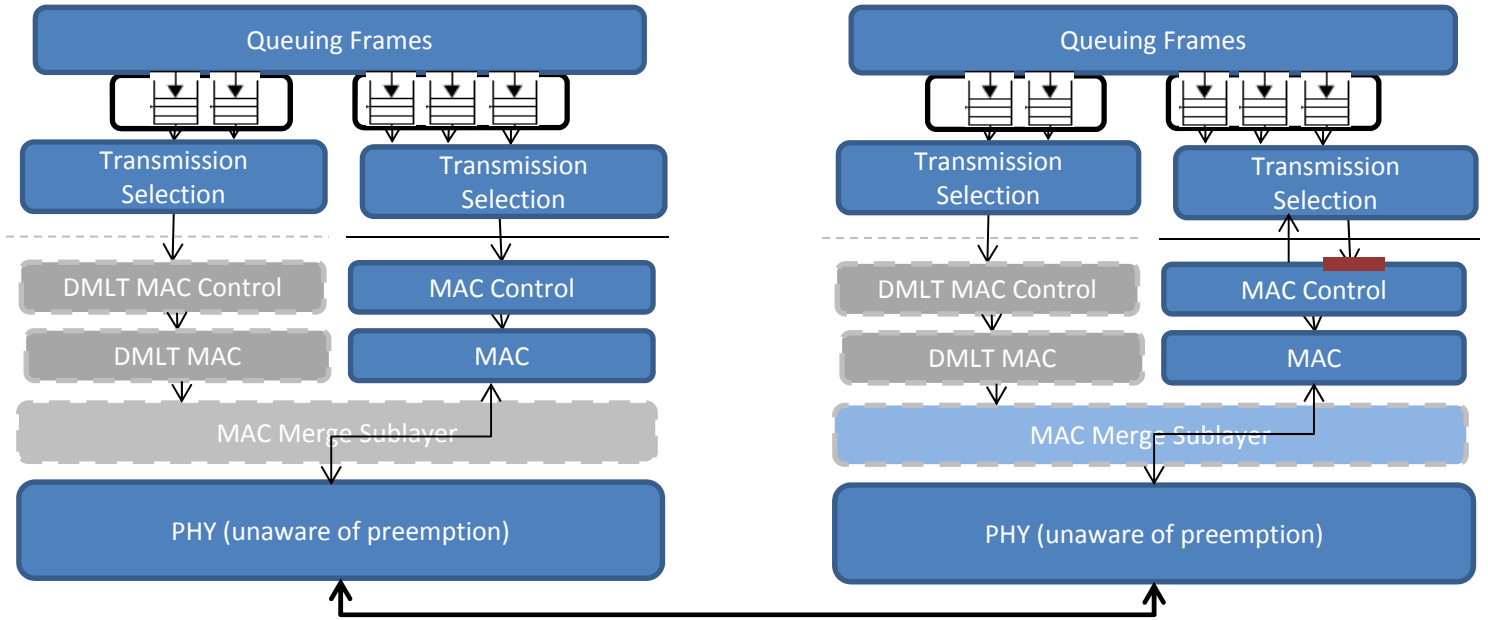
Initial Operation and Negotiation Start

- Before negotiation,
  - DMLT MAC and DMLT MAC Control are disabled
  - MAC Merge is bypassed so traffic passes through between MAC and PHY
- MA_Control.request from MAC Client initiates negotiation
  - MAC Control DMLT request frame sent to link partner



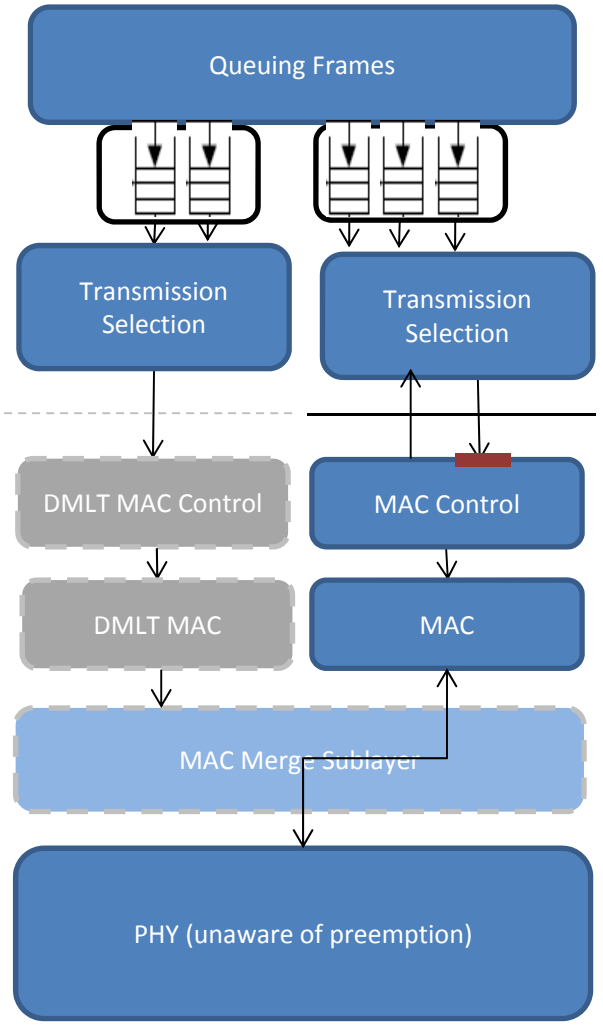
Link Partner receives DMLT request

- Link Partner passes MA_Control.indication of DMLT request to MAC Client
- MAC Client sends MA_Control.request to accept DMLT operation
- MAC Control
  - Disables further transmission of data frames
  - Sends MAC Control DMLT accept
  - Puts MAC Merge Sublayer in DMLT detect mode
  - Starts DMLT detect timer



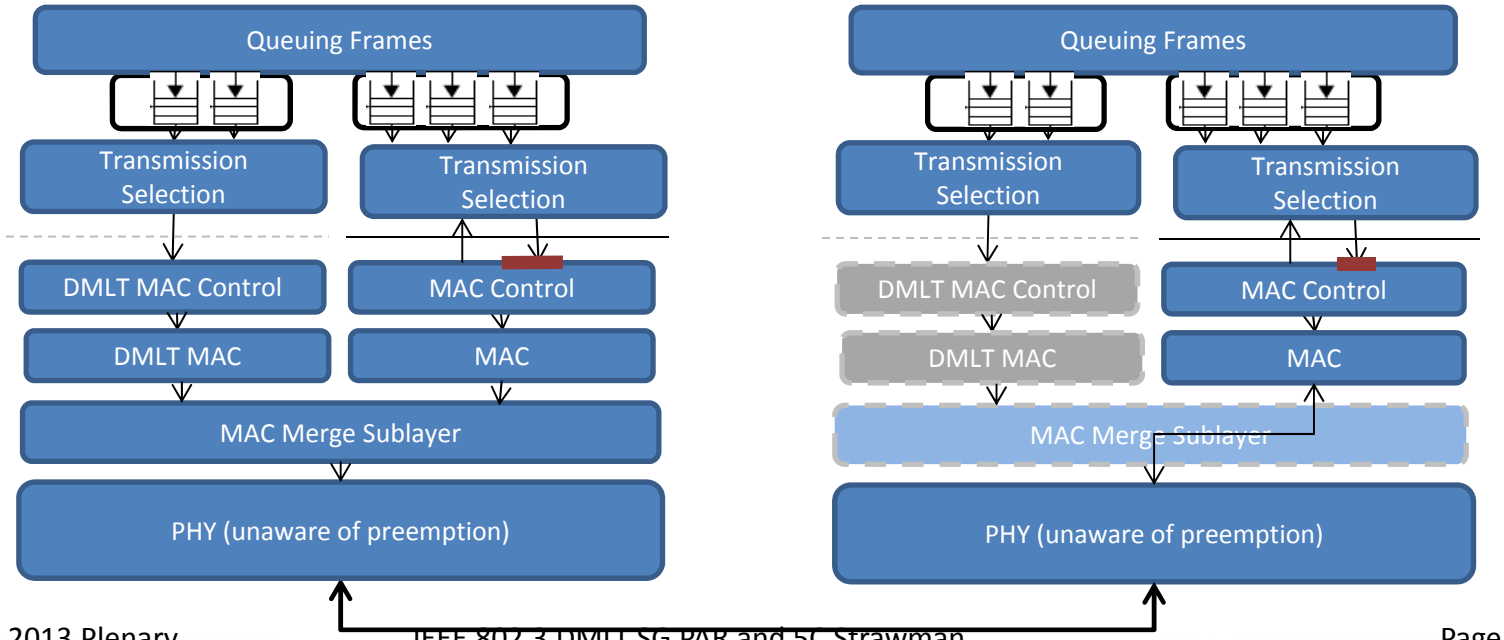
DMLT Detect Mode

- DMLT MAC and DMLT MAC Control are disabled
- MAC Merge does not alter frames transmitted from MAC to PHY
- MAC Merge watches for DMLT SFDs on received frames
  - Any frames received with normal SFD are passed to MAC unchanged
  - When DMLT SFD is received, receive side of MAC Merge Sublayer is activated



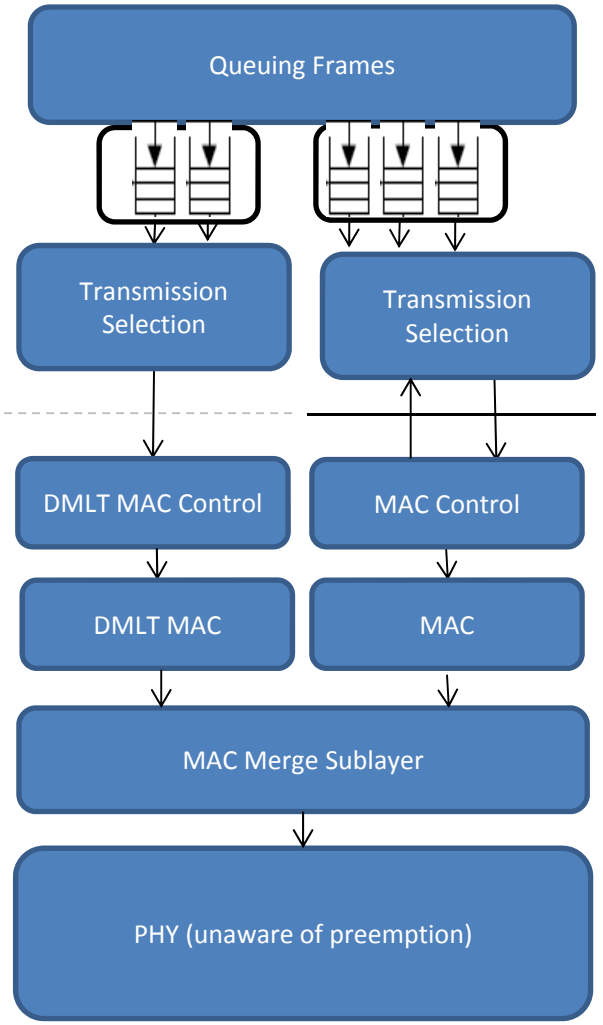
DMLT accept received

- Passes MA_Control.indication of DMLT accept to MAC Client
- MAC Control
  - Disables further transmission of data frames
  - Puts MAC Merge Sublayer in operational mode
  - Enables DMLT MAC Control and DMLT MAC
  - Starts DMLT Detect Timer
  - Periodically sends MAC Control DMLT detect



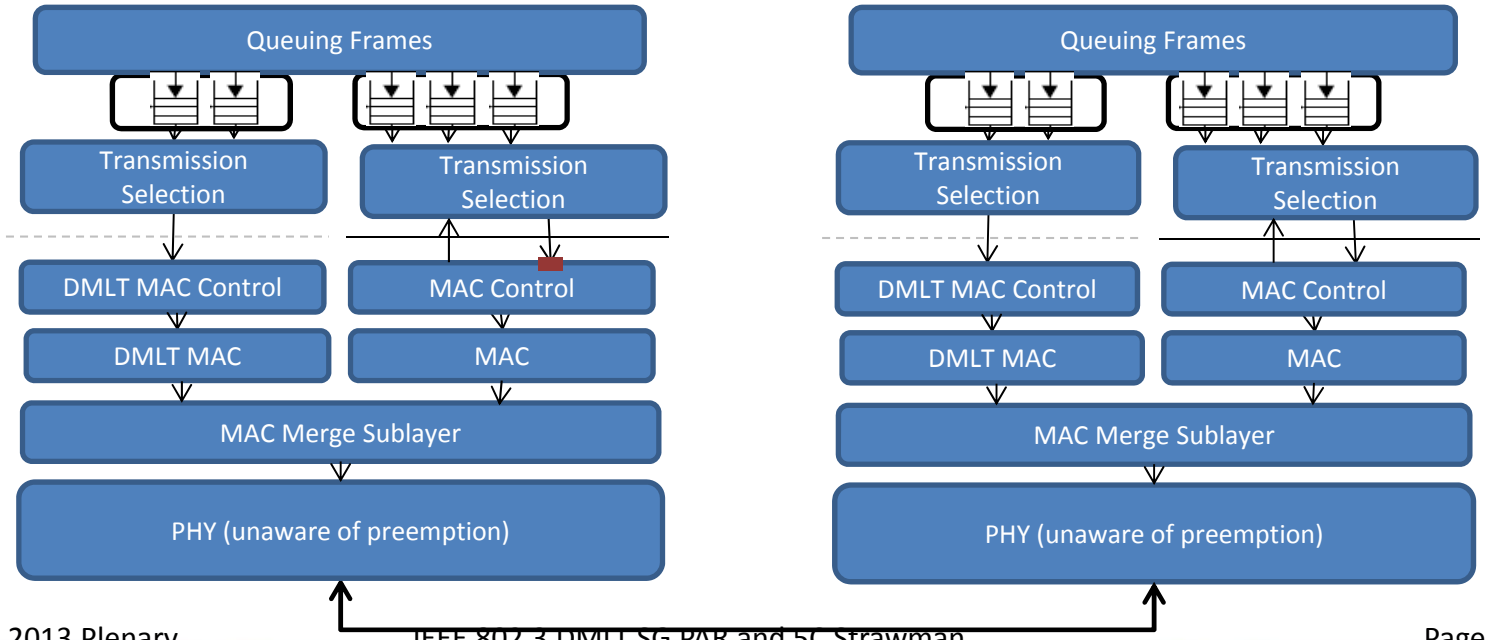
DMLT Operational Mode

- DMLT MAC and DMLT MAC Control are enabled
- MAC Merge operates normally
  - transmits frames with DMLT encapsulation
  - receives frames with DMLT encapsulation
  - frames with normal encapsulation are discarded



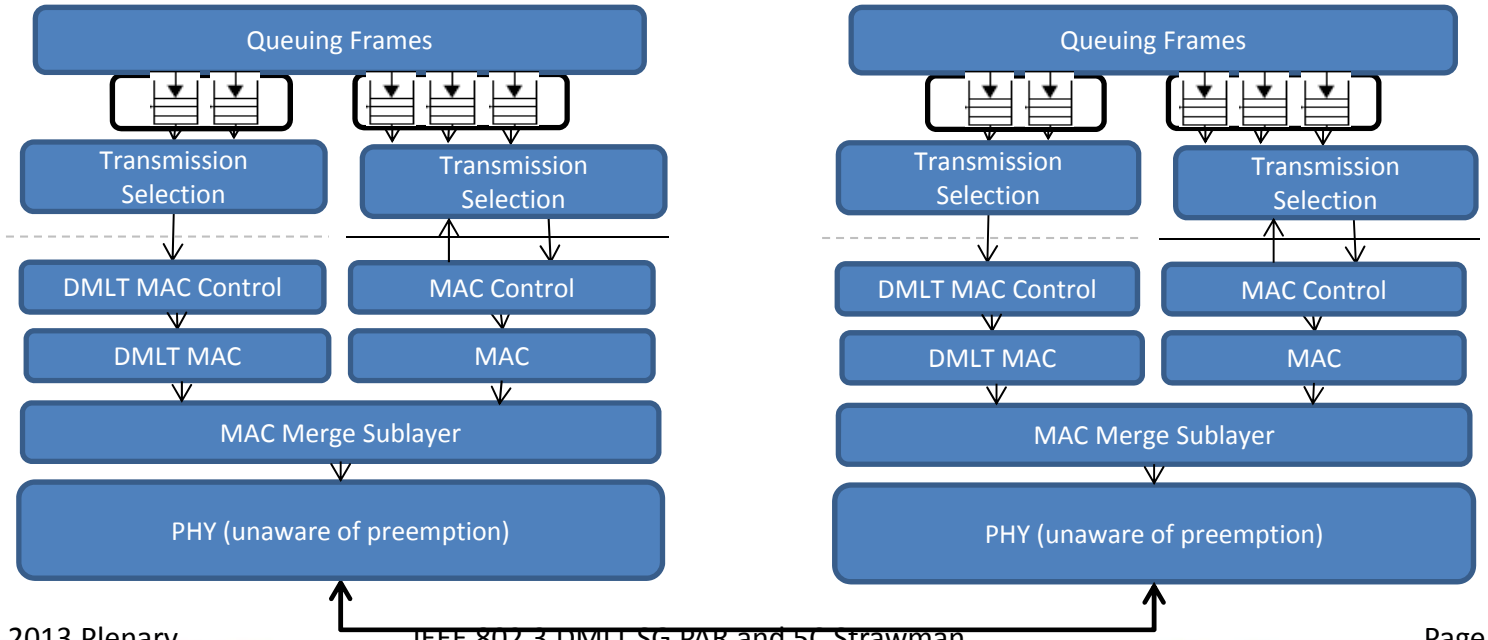
DMLT detect received by Link partner

- Link Partner passes MA_Control.indication of DMLT detect to MAC Client
- MAC Control
  - Puts MAC Merge Sublayer in operational mode
  - Enables DMLT MAC Control and DMLT MAC
  - Sends MAC Control DMLT detect
  - Enables data frame transmission



Entering fully operational DMLT mode

- Initiating MAC Control receives DMLT frames from Link Partner
- Passes MA_Control.indication of DMLT detect to MAC Client
- MAC Control
  - Enables data frame transmission



Recovery from time out

- If DMLT Detect Timeout expires before DMLT operation is established,
  - Transmit MAC Control DMLT fail frame
  - Start DMLT recovery timer
  - Disable DMLT MAC, DMLT MAC Control and MAC Merge sublayer
  - When DMLT recovery time expires, enable transmission of data frames

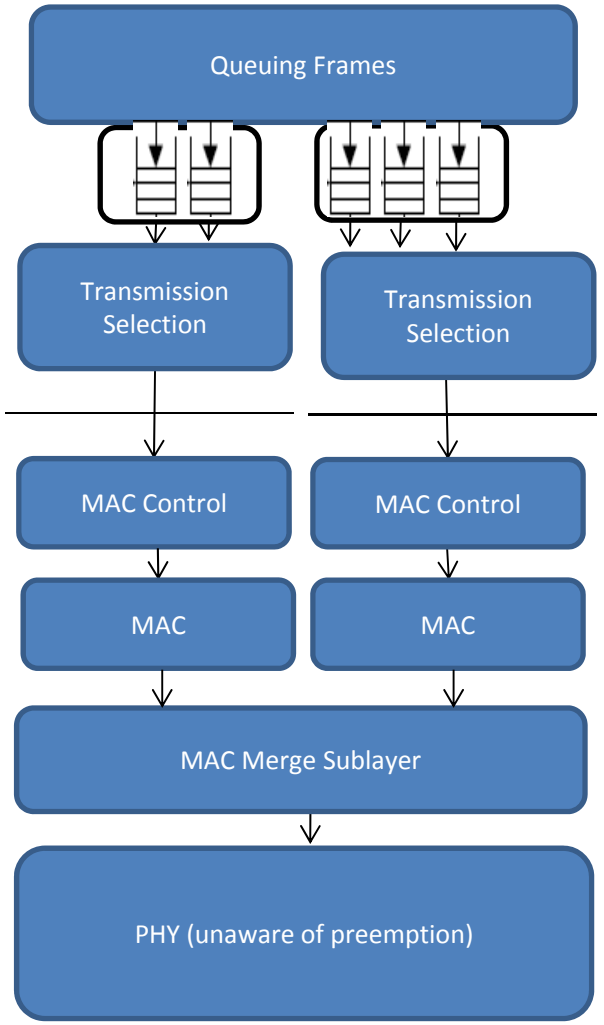
Thank You.

Questions or Comments?

Backup

DMLT negotiation goals

- Preserve frame integrity
  - No increase in undetected errors
- Indicate which MAC receive frame belongs to
- Minimize impact on throughput
- Transparent to existing non-deprecated PHYs above 10 Mb/s



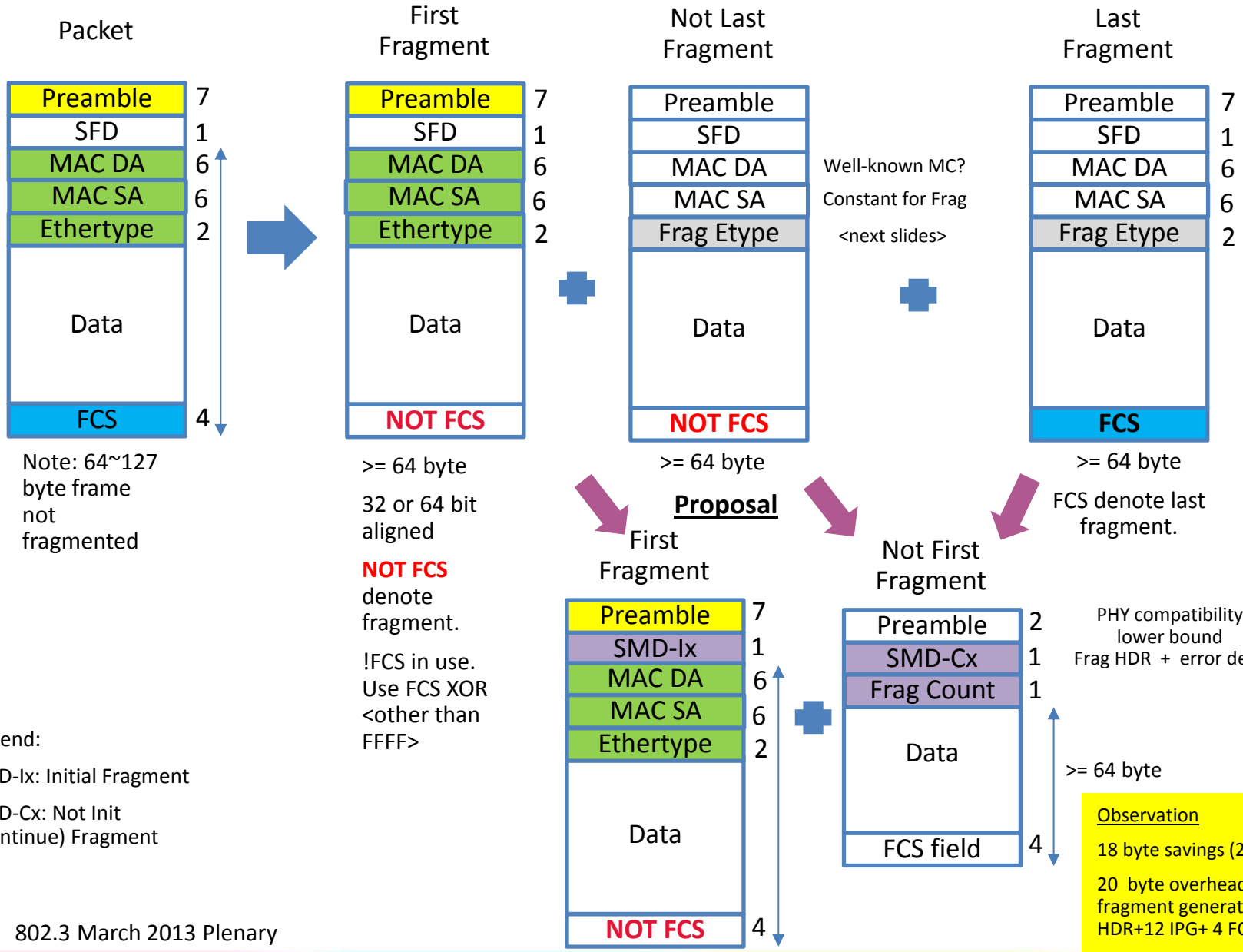
Terminology

- DMLT frame – frames with the lowest latency
- Preemptable frame – frames that are not distinguished to be low latency.
- Mframe -- A transmitted unit from MAC Merge that includes both whole frames and fragments of premptable frames – Mframe
    - Which is stands for MAC Merge frame – a unit that looks like a frame at the PHY layer but may contain a whole frame or a fragment of a MAC layer premptable frame. I'm leaning toward this one.

Assumptions

- MAC Merge does not have to be supported over deprecated PHY Clauses

Conceptual MAC Merge Frame Consideration



MAC encapsulation elements

- Preamble
  - Minimize use of preamble bytes for frame start
  - Provide at least 2 bytes of preamble for fragments
    - The 1000BASE-X PHY can drop up to 2 bytes of preamble and insert the SFD over another byte. In practice, implementations only drop 1 and many delay the start of preamble rather than drop and only overwrite 1 byte (similar to 10GBASE-X alignment)

- Identify Mframe as DMLT or start of a preemptable frame or a later fragment of a preemptable frame
- Protection for reassembly errors when an Mframe is lost
  - Frame number – circular count from 0 to 3
  - Fragment number – circular count from 0 to 3
- Identify last Mframe of a Frame
  - Mark end of preemptable Mframe

Mframe start

- For start of non-initial fragments
  - Insert 2 bytes of preamble followed by
  - SMD byte (Start Mframe Delimiter)
  - Frag byte (Fragment count)
- Preemptable Frame start and DMLT frame
  - Replace SFD with SMD
- SMD values have Hamming distance 4 from each other
- Frag values have Hamming distance 4 from each other

SMD and Count byte encodings

Mframe type	Frame #	SMD
DMLT	NA	0x33
Premptable frame start	0	0x66
	1	0xCC
	2	0xFF
	3	0xAA
Non-initial fragment	0	0xE1
	1	0xD2
	2	0x1E
	3	0x2D

Fragment #	Frag
0	0x66
1	0xCC
2	0xFF
3	0xAA

Frame CRC

- Frame CRC is generated by the MAC over the same frame bits as always. It is not altered by the MAC Merge sublayer

Indicating end of fragment

- Since one doesn't know that a frame will be preempted until transmission of the frame is in progress, a marking at the end transmission indicating that this is not the final fragment is needed.
- Two alternatives
  - All premptable frames have a fragment trailer that indicates whether this is the end of the frame.
  - Keep a running CRC calculation in the MAC Merge sublayer (identical to the CRC calculation done in the MAC sublayer) and append the intermediate result altered in some way to the frame.

- Note: Preempted fragment size should be no smaller than 64 bytes, and each fragment should be 32 or 64 bit aligned (except for the last one), and packet less than 127 bytes should not be preempted.

Fun with CRCs

- IEEE 802.3 inverts the calculated CRC and appends the result to the frame.
- Transmitting MAC Merge sublayer could invert the second two bytes of the intermediate CRC result (the CRC computed over the bytes of the MAC frame that have been transmitted so far) at the end of non-final fragments.
- Receiving MAC Merge sublayer runs a CRC calculation as the frame is received. When an Mframe ends, it compares the calculated value with the
  - If the difference between that and the last 4 bytes of the Mframe is 0xFFFF, it's the end of a MAC frame
  - If the difference between that and the last 4 bytes of the Mframe is 0x00FF, it's a non-final fragment.

MAC behavior for TransmitBit

- From 4.3.3 of IEEE Std 802.3-2012

During transmission, the contents of an outgoing frame are passed from the MAC sublayer to the Physical Layer by way of repeated use of the TransmitBit operation:

procedure TransmitBit (bitParam: PhysicalBit);

Each invocation of TransmitBit passes one new bit of the outgoing frame to the Physical Layer. The TransmitBit operation is synchronous. The duration of the operation is the entire transmission of the bit. The operation completes, when the Physical Layer is ready to accept the next bit and it transfers control to the MAC sublayer.

MAC behavior for ReceiveBit

- From 4.3.3 of IEEE Std 802.3-2012

During reception, the contents of an incoming frame are retrieved from the Physical Layer by the MAC sublayer via repeated use of the ReceiveBit operation:

function ReceiveBit: PhysicalBit;

Each invocation of ReceiveBit retrieves one new bit of the incoming frame from the Physical Layer. The ReceiveBit operation is synchronous. Its duration is the entire reception of a single bit. Upon receiving a bit, the MAC sublayer shall immediately request the next bit until all bits of the frame have been received.

Fragment trailer

- 1 byte – all zeros indicates frame incomplete, all ones indicates frame complete. (Or use 4 bits and leave 4 bits reserved).
  - Another alternative if we use shortened header for frame start would be to encode frame number for incomplete frames by use of 5 values with Hamming distance 4:
    - Frame complete
    - Frame incomplete, Frame #0
    - Frame incomplete, Frame #1
    - Frame incomplete, Frame #2
    - Frame incomplete, Frame #3