Handling loss of signal in the ILT/RTS functions

(addressing comment #I-109)

Adee Ran, Cisco
Leon Bruckman, NVIDIA
Luz Osorio, Nokia
Jeff Slavick, Broadcom
Matt Brown, Qualcomm
Tom Huber, Nokia
Mike Dudek, Marvell
Marco Mascitto, Nokia
Mike Sluyski, Cisco
Jeffery Maki, HPE
Gary Nicholl, Cisco
Sebastien Gareau, Ciena
Xianyun Meng, HPE
Arun Kumar Alapati, HPE

Background

- Previous ad hoc presentations:
  - Part 1 (ran_178b_01_260414) presented the problem – LOS is not addressed by the D3.0 specification.
  - Part 2 (ran_178b_01a_260422) presented possible solutions.
  - Straw polls indicated support for addressing the problem with a solution that enables choice between several behaviors ("Option Y").

Straw Poll #2 (Ran)

- When an LOS event is detected anywhere in a path that uses APSU, my preference is
  - A. Allow recovery from LOS without a full restart
  - B. Add a restart of training within the state diagram
  - C. Keep the current specifications (requiring restart by management)
  - D. Abstain

Results:

A: 5
B: 8
C: 1
D: 5

polls_178b_260414

Straw Poll #2 (Ran)

Per ran_178b_01a_260422, for a loss of signal (LOS) that occurs on an interface within a path after APSU has completed, I would support specifying the following:

- A. Option X - slide 8
- B. Option Y , Y1, or Y2 - slide 9
- C. Neither, make no changes
- D. Need more information

*note (further straw polls to follow on configurable options if appropriate)

- Results:
  A: 5
  B: 9
  C: 3
  D: 2

polls_178b_260422

Straw Poll #2a (Ran)

Per Slide #5 of ran_178b_01_260422, when an interface within a path is retaining or recovering, I would support the direction of the following:

- A. Option Y1
- B. Option Y2
- C. Option Y
- D. Need more information

- Results:
  A: 0
  B: 2
  C: 10
  D: 1

polls_178b_260422

Goals

- Add specifications in Annex 178B to:
  - Enable a receiver in PMD/AUI to handle detection of LOS by either:
    - Restarting training
    - Recovering when the signal reappears without re-training
  - Ensure restarting causes the peer to restart too
  - Define the behavior of the adjacent interface
- Address various application needs
- Maintain interoperability between devices even if they use different options

Observations

- When LOS is detected, **restarting training** is the safe way to recover
  - Provided that it causes the peer to restart too
  - If this is the only specified behavior, it requires restarting APSU on the whole path
  - **Proposed as default behavior**
- Some applications may benefit from **recovery without retraining**
  - Behavior needs to be specified
  - This should be **required to implement but optional to use - enabled by management**
- Transition to FAIL requires management to restart
  - This makes restarting required, but without a specified time
  - Consistent with previous training specifications
  - Enables implementation flexibility

Proposed changes

(details in the backup section)

Split the training control state diagram into two parts

This is “part 1” (Figure 178B-10a)

Minor changes in this part:

“C” pointers go to a new state ISL_FAIL (separate from training)

“B” pointers go to the existing PATH_READY

FAIL state split into two; TRAIN_FAIL only applies when mr_training_enable=true, then goes to ISL_FAIL

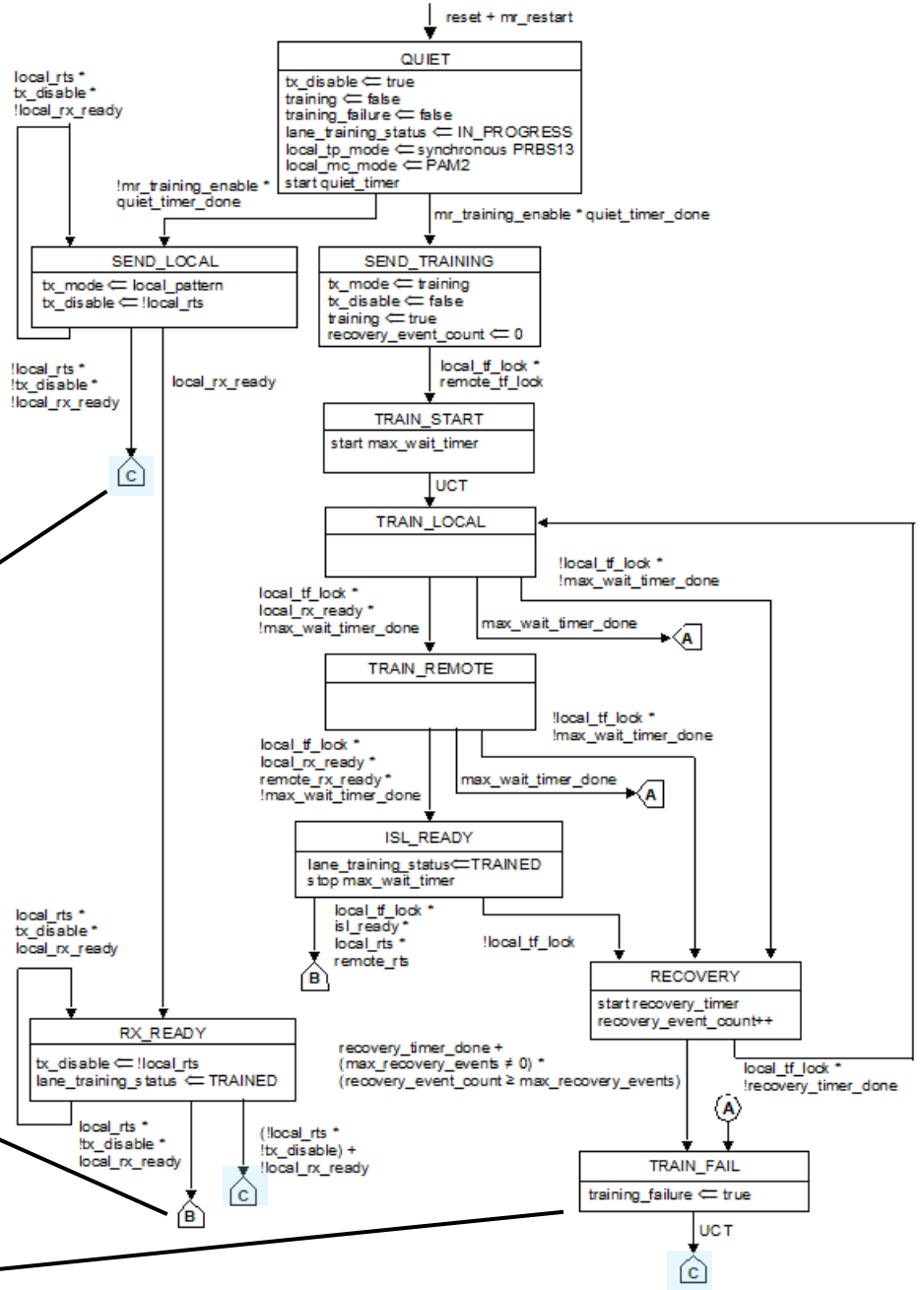


Figure 178B–10a—Training control state diagram, part 1

Additions to Figure 178B-10 to enable either restart or recovery from LOS

(using squelch or replacement signal on adjacent interface)

("Option Y")

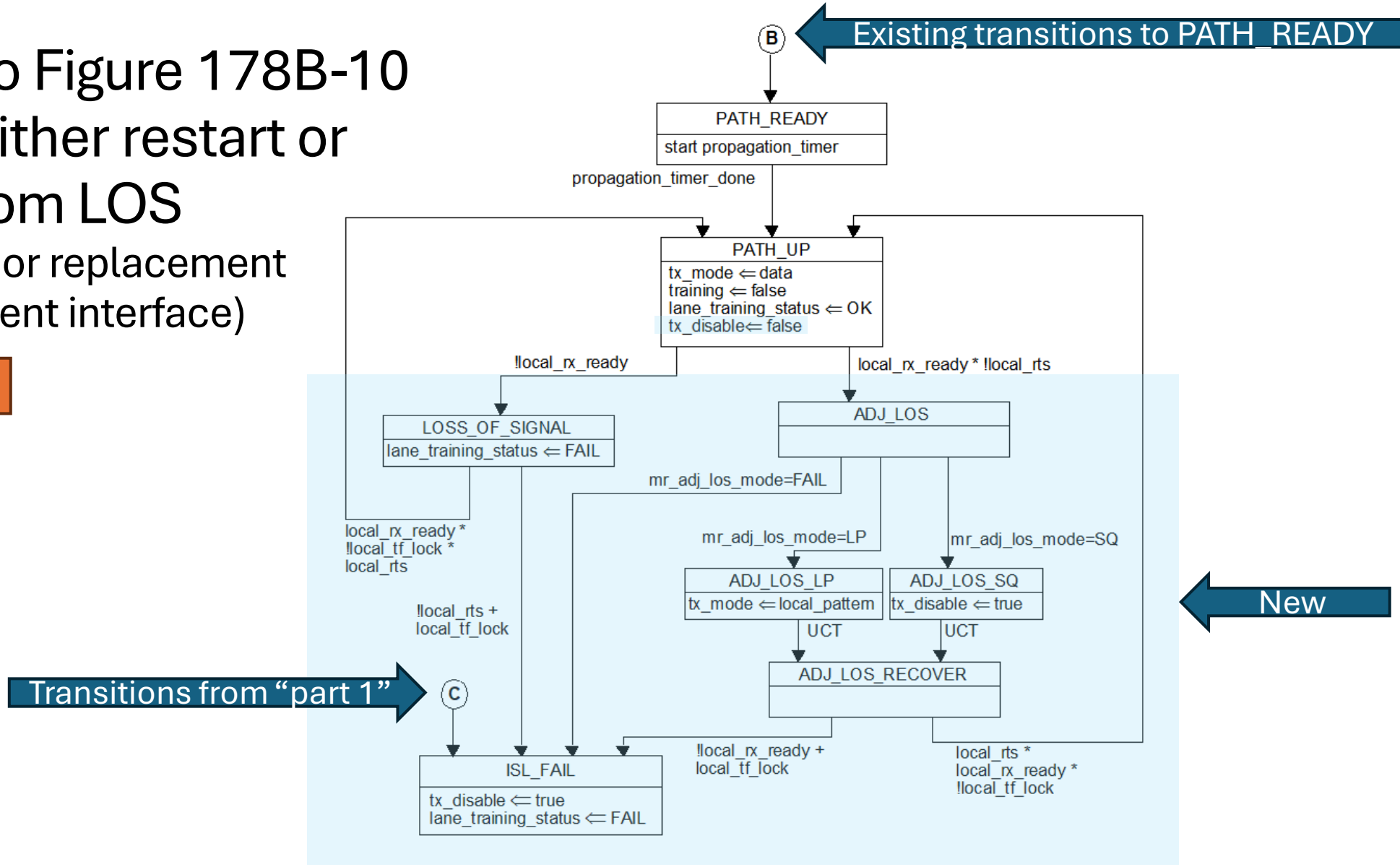


Figure 178B–10b—Training control state diagram, part 2

local_rx_ready

Needs to be set to false if the peer restarts – which creates a QUIET period of at least 100 ms

- In 178B.8.3.1, change the definition of local_rx_ready :

  Boolean variable that is set to true when the receiver on a lane of the interface has determined that the peer interface transmitter is transmitting a PAM4 signal, that the remote transmit has been optimized if mr_training_enable is true, the local receive equalizers have been optimized, and that no further adjustments are required for normal data transmission. The specific conditions for setting this variable to true are implementation dependent. It shall be set to false when an input signal corresponding to tx_disable=true in the peer is received for 100 ms (the minimum duration of the quiet_timer) or longer.

remote_rx_ready

Needs to be set to true when not in training

- Change the definition of remote_rx_ready in 178B.8.3.1:

  Boolean variable that indicates the value of local_rx_ready on a lane of the peer interface. If mr_training_enable is true and training is true, it is derived from the receiver ready bit of the status field of received training frames on the corresponding lane of the interface. Otherwise it is set to true.

In this proposal we address the “unrecoverable faults”, and we see no more “live lock” concern.

- In 178B.8.2.1, delete the second and third sentences in the note after the definition of mr_restart:

NOTE — There is no specified time limit for the APSU. ~~Restarting APSU might result in live-lock situation, thus APSU should only be restarted if there is an indication of an unrecoverable fault. The definition of unrecoverable fault is beyond the scope of this standard.~~

mr_adj_los_mode

New variable

- Add a new variable definition in 178B.8.3.1:

mr_adj_los_mode

Enumerated variable that controls the behavior of the training control state diagram (Figure 178B-10) in PATH_UP state when local_rts becomes false. This variable is assigned by management and can take one of the following values: FAIL, SQ, or LP. The default value is FAIL.

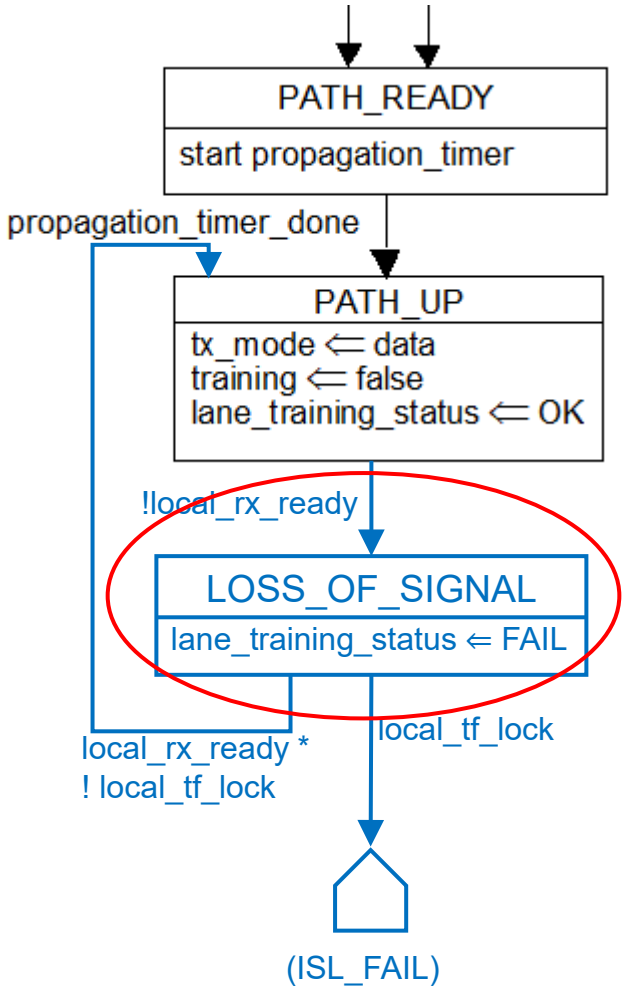
Backup

Interoperability

- The variable mr_adj_los_mode (choice between FAIL, SQ or LP) affects the behavior of the **adjacent** interface.
  - The value FAIL will also cause the interface that experienced LOS to restart.
- The values SQ and LP enable recovery without restart if other devices in the path also support it:
  - If an interface that always restarts (FAIL) is connected to a peer that supports recovery (SQ or LP), the peer will see the QUIET period followed by training frames – which will cause a restart.
  - If the adjacent interface is squelched and its peer has mr_adj_los_mode=FAIL, it will restart and cause the adjacent interface to restart too.

LOS at the input of this interface

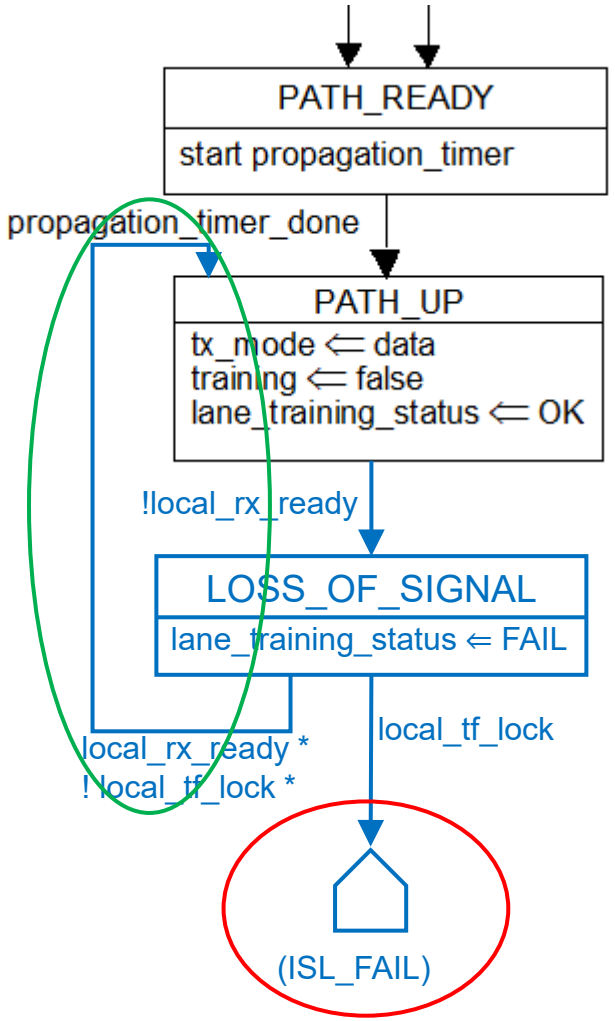
- LOS needs to be communicated to the adjacent interface to enable auto-squelch.
- In PATH_UP, when loss of signal event is detected, transition to a new state LOSS_OF_SIGNAL.
- In this state:
  - lane_training_status becomes FAIL
  - This is communicated to the adjacent interface via SIGNAL_OK
  - The output of this interface is not squelched, to enable Remote Fault signaling to be used (it is squelched only if ISL_FAIL is reached).



addition to Figure 178B–10

Recovery from LOS at the input of this interface

- Signal re-appearing is indicated by local_rx_ready
  - The timing and conditions to declare local_rx_ready is not specifie It should indicate “good data”.
- If the signal is not training frames (PAM4, mission data):
  - Return to PATH_UP
  - lane_training_status=OK propagates through service interface
- If it is training frames (PAM2):
  - This means positively that the peer has restarted training
  - We must restart training too (no other way to recover)
  - Transition to ISL_FAIL, which requires mr_restart to exit
  - Note: indicated in the diagram as local_tf_lock, but could be detected by other means, e.g. not detecting AMs in the PMA/PCS.
- Management can always restart training (mr_restart)
  - For example, if the signal does not re-appear within some time



addition to Figure 178B–10

Recovery from adjacent interface LOS

- When SIGNAL_OK on the service interface (from the adjacent interface) recovers to OK…
  - adjacent_intf_rx_ready becomes true
  - The RTS update state diagram advances from START
- The state diagram needs to set local_rts=true to make the local interface unsquelch (slide 20).
- If **mr_training_enable=false**, the transition is unconditional.
- But if **mr_training_enable=true**:
  - If a recovered clock is used (module/retimer) it needs isl_ready (local_rx_ready * remote_rx_ready on all lanes) to advance to SWITCH_CLOCK
  - In data mode, training frames are not received, so **remote_rx_ready** is irrelevant. It should be set to true in its definition.

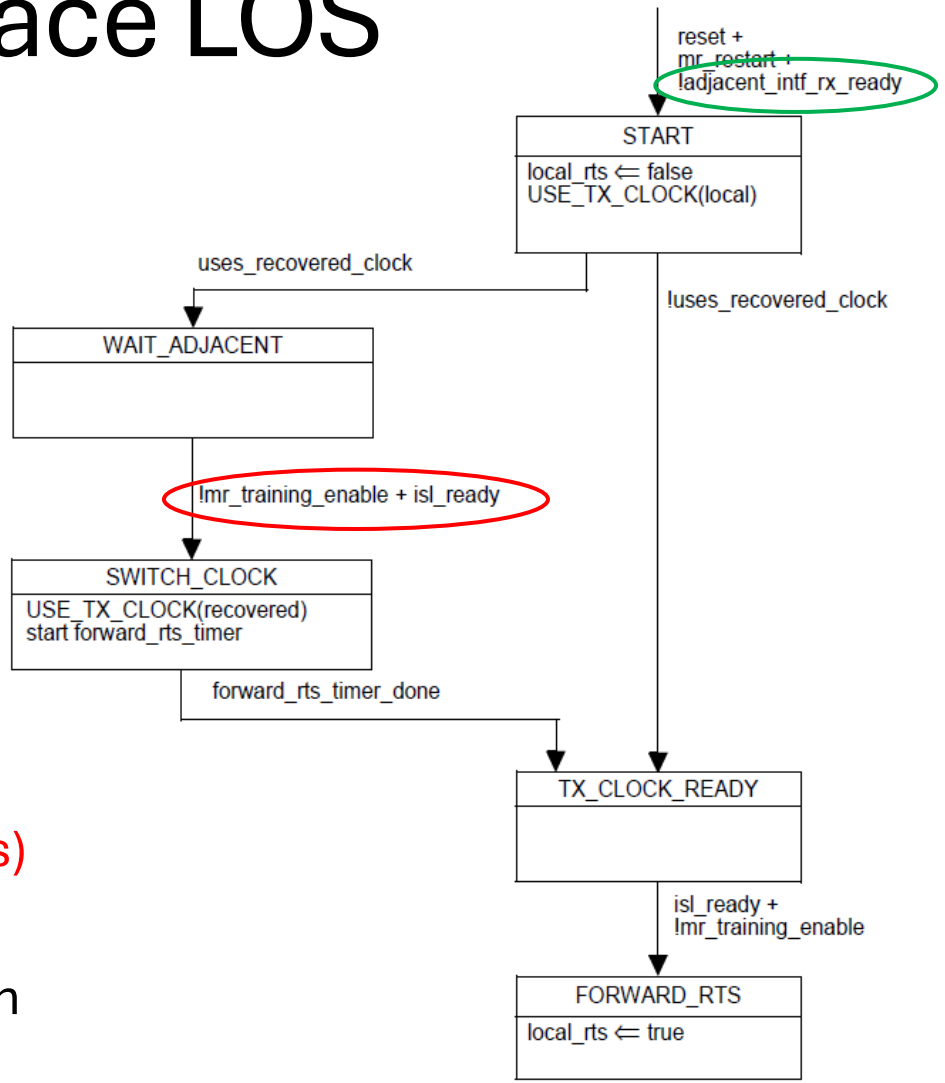
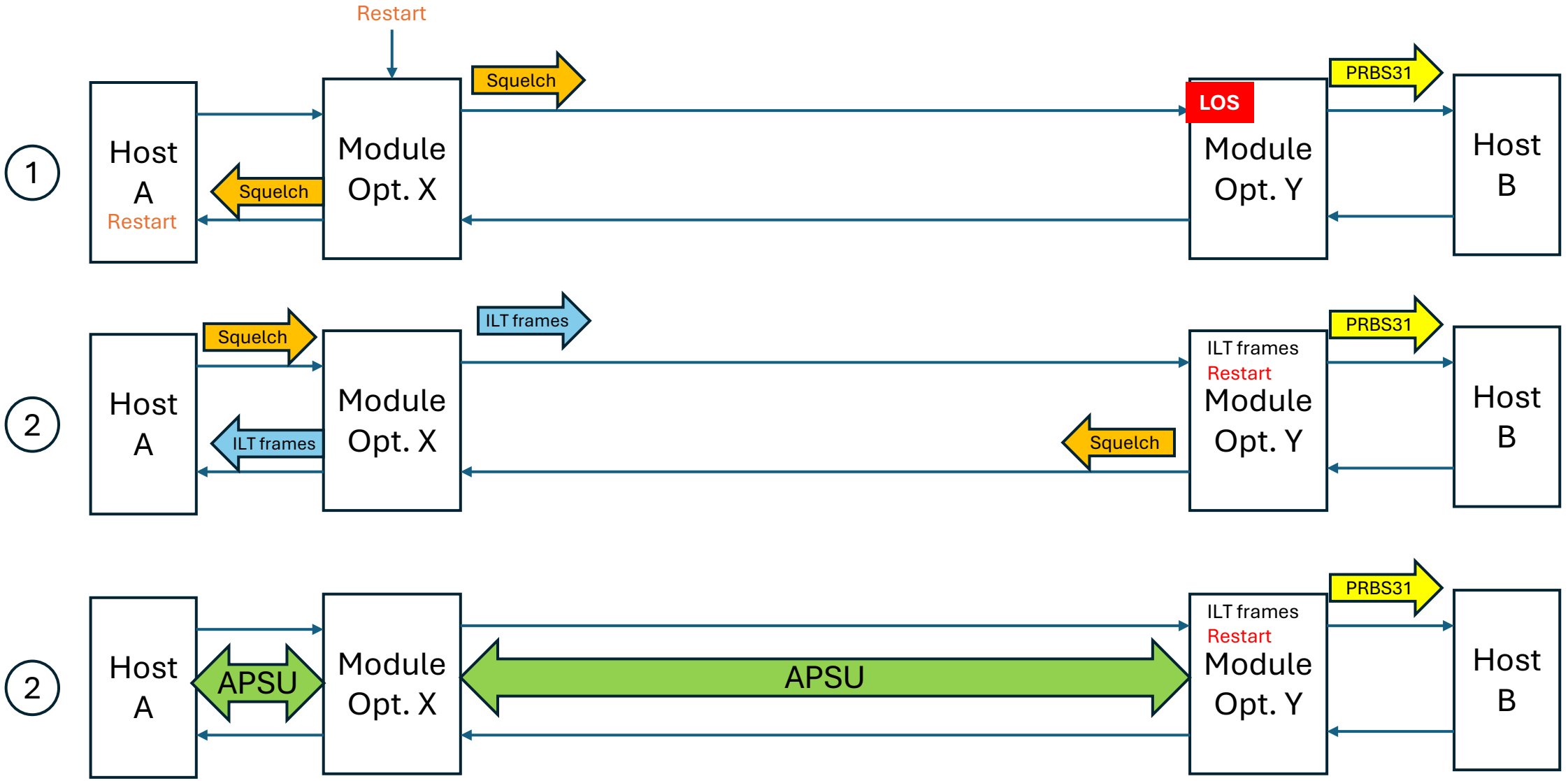
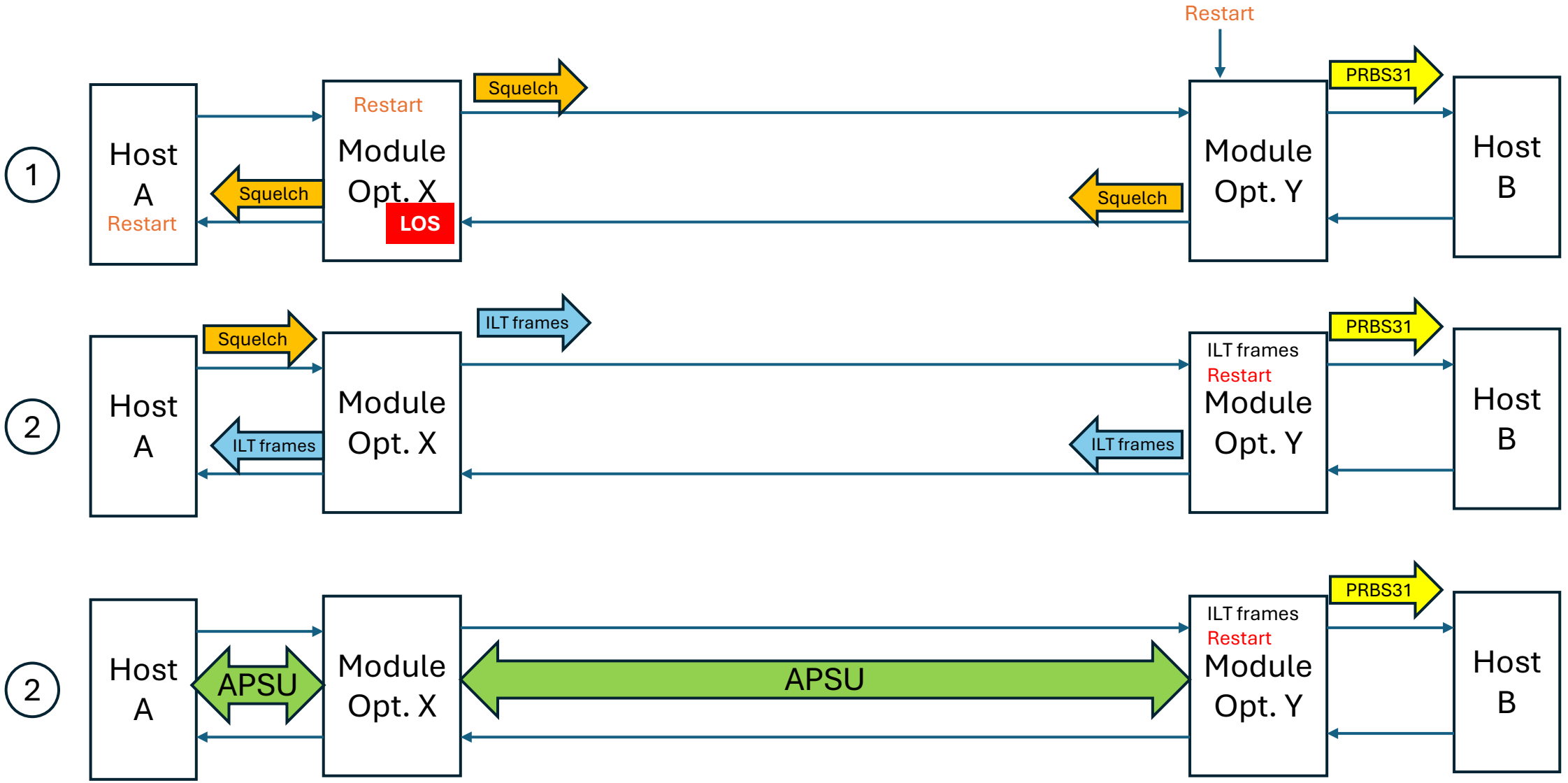


Figure 178B–9—RTS update state diagram

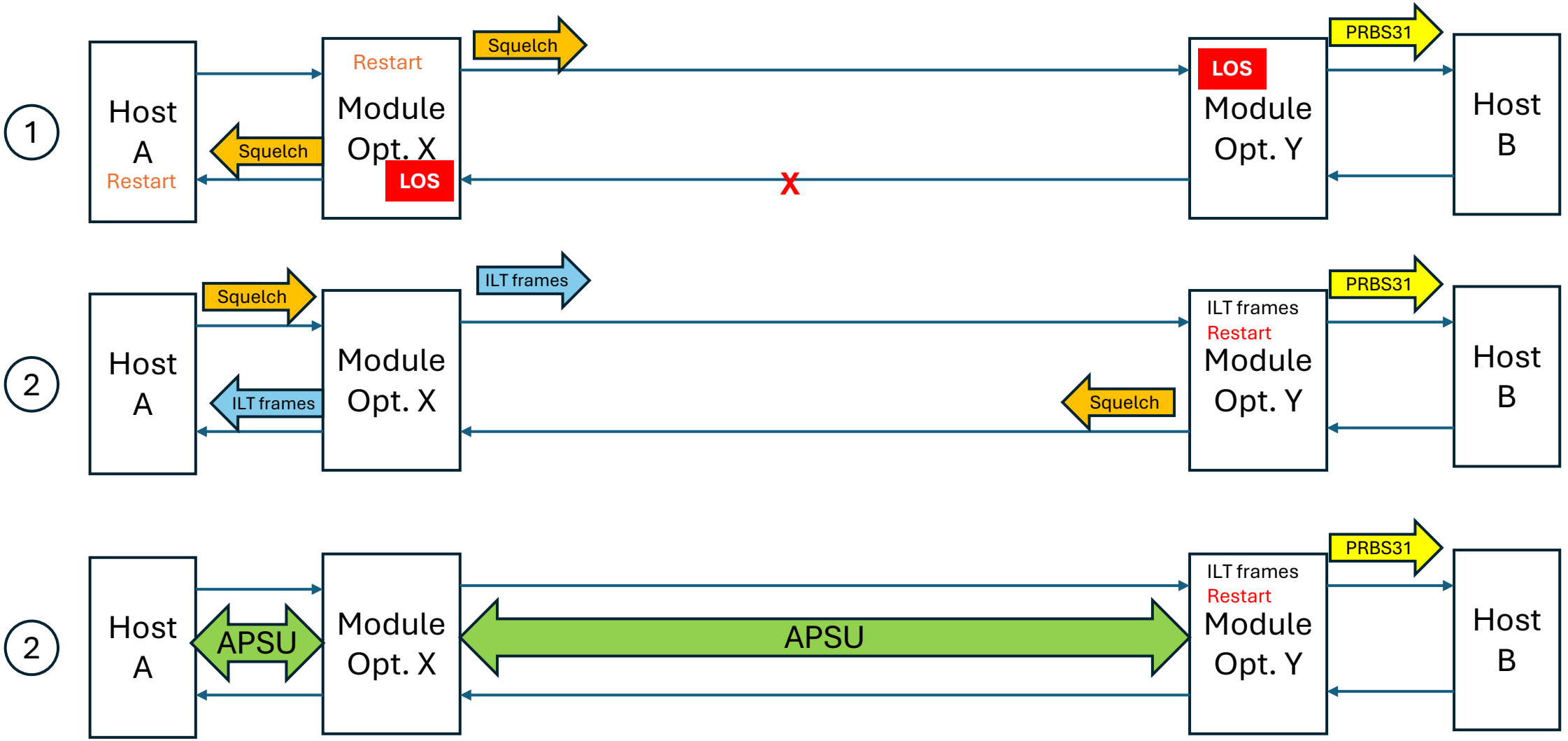
Option X – Option Y interoperability – Case A



Option X – Option Y interoperability – Case B



Option X – Option Y interoperability – Case C



Option X – Option Y interoperability – Case D