Cl 136.8.11 Link Training Issue (comment #1) Update

November 2020

Kent Lusted, Intel

Adee Ran, Intel

Zvi Rechtman, NVIDIA

Jeff Slavick, Broadcom

Summary

- There was an issue with the Cl 136.8.11 PMD Control Function (i.e. link training) described in https://www.ieee802.org/3/ck/public/20_07/lusted_3ck_02_0720.pdf
- Numerous offline discussions indicate support to fix the issue
- A two-step approach was taken:
  - 1. Identify candidate solution(s) to solve the problem for 3ck in Cl 162.8.11
  - 2. Explore *feasibility* of solution(s) as an optional feature for 3cd (in Cl 136.8.11) that is interoperable with existing devices
- Two solutions made the final cut: "B3" and "B4"
- Propose a TF straw poll to choose the solution

Three Directions Were Initially Considered

- A. State Indication via new Training Frame bits
- B. TX Squelch
- C. Only new timers

Option B had consensus to resolve the issue and provide a compatible-to-legacy solution

Solution Choices: Option B3 & B4

- Both use tx-based squelch mechanism to induce/detect an abort.
  - B3 at the end of the state machine
  - B4 at the beginning and end of the state machine
- Remote end monitors for tx squelch and aborts training process when necessary

Choice	Resolves LT deadlock	Legacy Cl 136 Compatibility	Non-AN73 "Reset during LT" Recovery Time*	# of Additional States
B3	Yes	Yes	80 msec + max_wait_timer (12 sec)	1
B4	Yes	Yes	80 msec	2

* Case of Legacy device connected to new device

B3

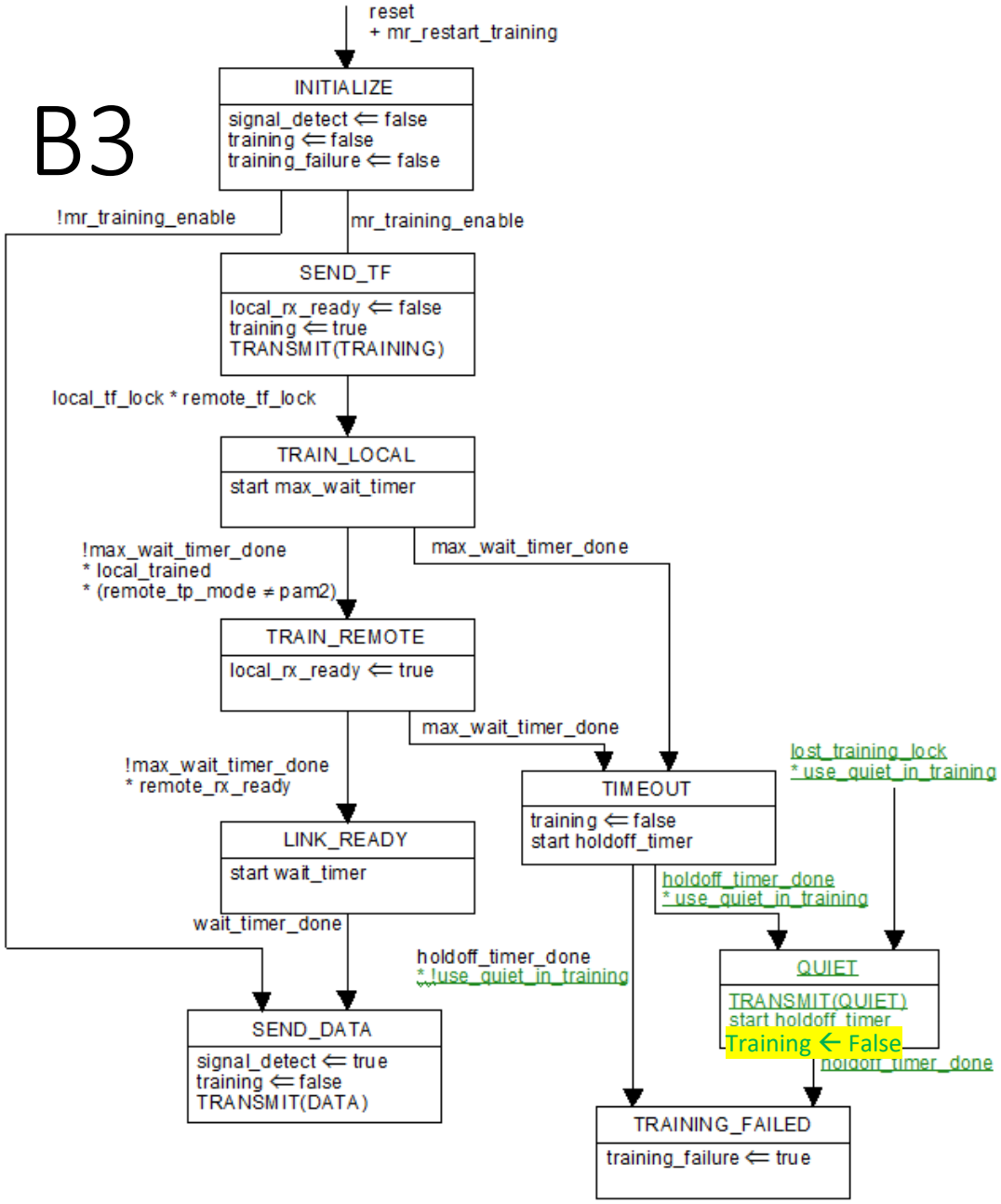

Figure 136–7—PMD control state diagram

B4

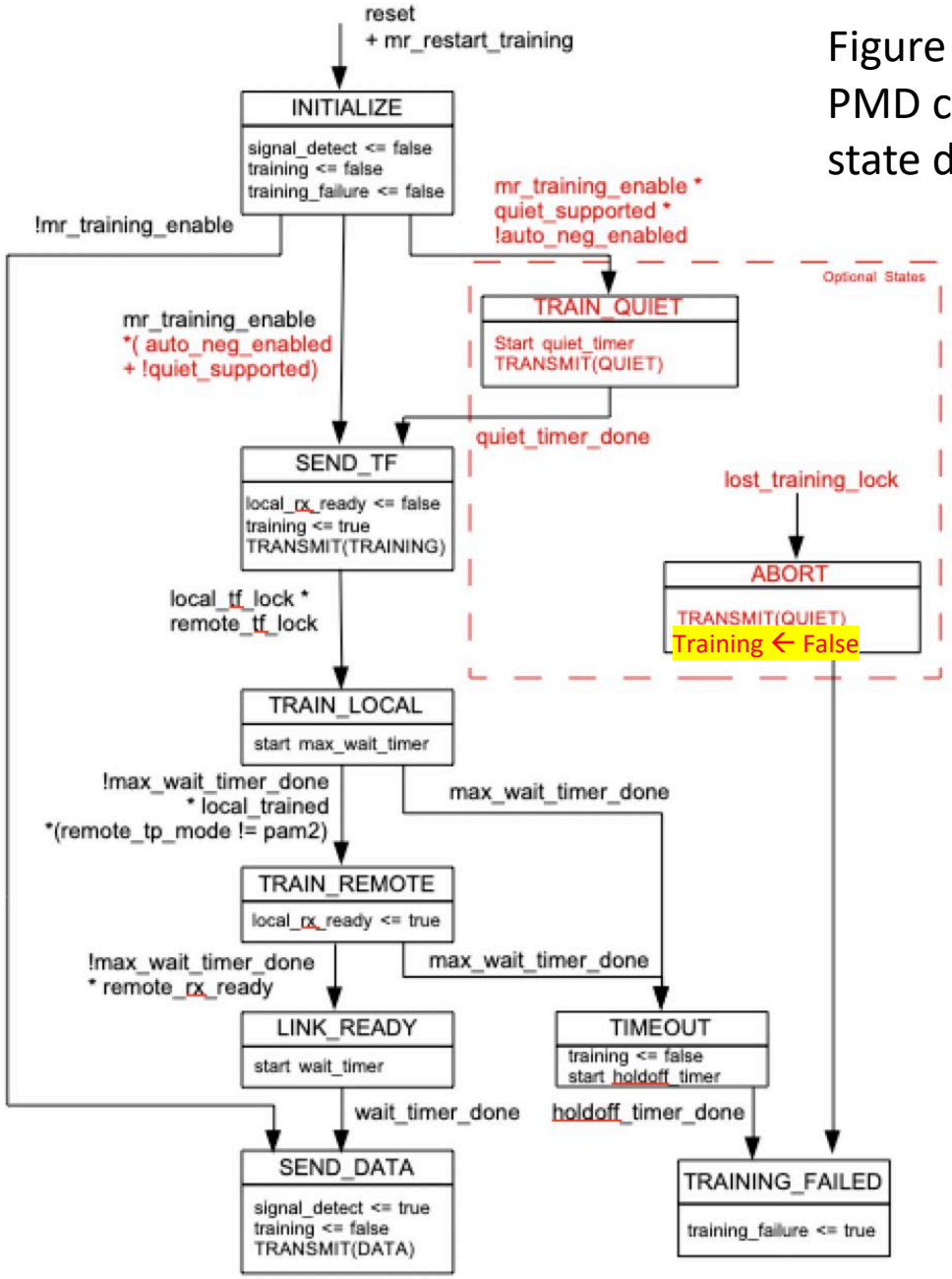

Figure 136-7 – PMD control state diagram

Common to both B3/B4

- Functions in Cl 136

TRANSMIT(tx_mode)

Controls the output of the PMD transmit function on the current lane. When tx_mode = DATA, the PMD transmit function output is the parameter of the PMD:IS_UNITDATA_*i*.request primitive (see 136.3). When tx_mode = TRAINING, the PMD transmit function output is the stream of symbols generated by the PMD control function. When tx_mode =QUIET, the PMD transmitter is turned off, equivalent to setting PMD_transmit_disable_*i* to one (see 136.8.7).

- **136.8.2 PMD transmit function**
  - *<modify the first paragraph as follows>*
  - The PMD transmit function has three operating modes DATA, TRAINING and QUIET. The operating mode is controlled by the PMD control state diagram (Figure 136–7).

  - *<add the following paragraph to the end of 138.8.2>*
  - When operating in QUIET mode the PMD transmit function shall turn off the transmitter such that the transmitter drives a constant level (i.e., no transitions) and does not exceed the differential peak-to-peak output voltage (max.) with Tx disabled in Table 136–11.

Link Training Issue Option B3

B3

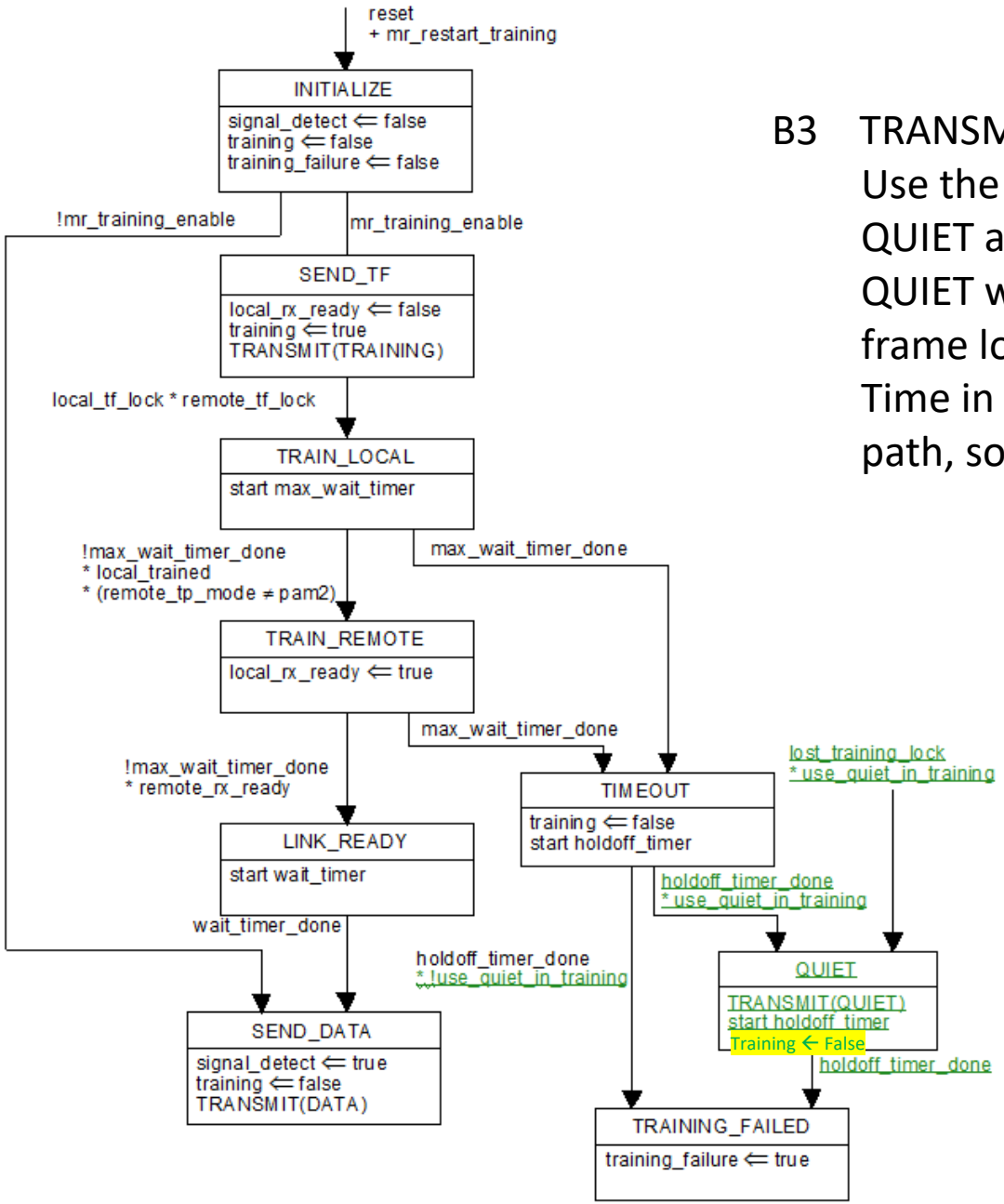

Figure 136–7—PMD control state diagram

B3 TRANSMIT(QUIET) in a new state QUIET after TIMEOUT

Use the holdoff timer twice

QUIET after “frame lock=0” frames indicates timeout

QUIET without “frame lock=0” frames indicates loss of frame lock

Time in the state diagram is extended, but only in failure path, so AN timeout is not affected

variable changes in C136

use_quiet_in_training

Boolean variable that is TRUE if the PMD disables the transmitter in some states of the PMD control state diagram (see Figure 136–7) and is set to FALSE otherwise.

lost_training_lock

Boolean variable that is set to TRUE when the PMD control function (see Figure 136–7) is in TRAIN_LOCAL or TRAIN_REMOTE states and local_tf_lock is FALSE continuously for a period of 20 ms or the detection of a non-compliant input signal has occurred continuously for 1ms. It is set to FALSE otherwise.

Link Training Issue Option B4

B4

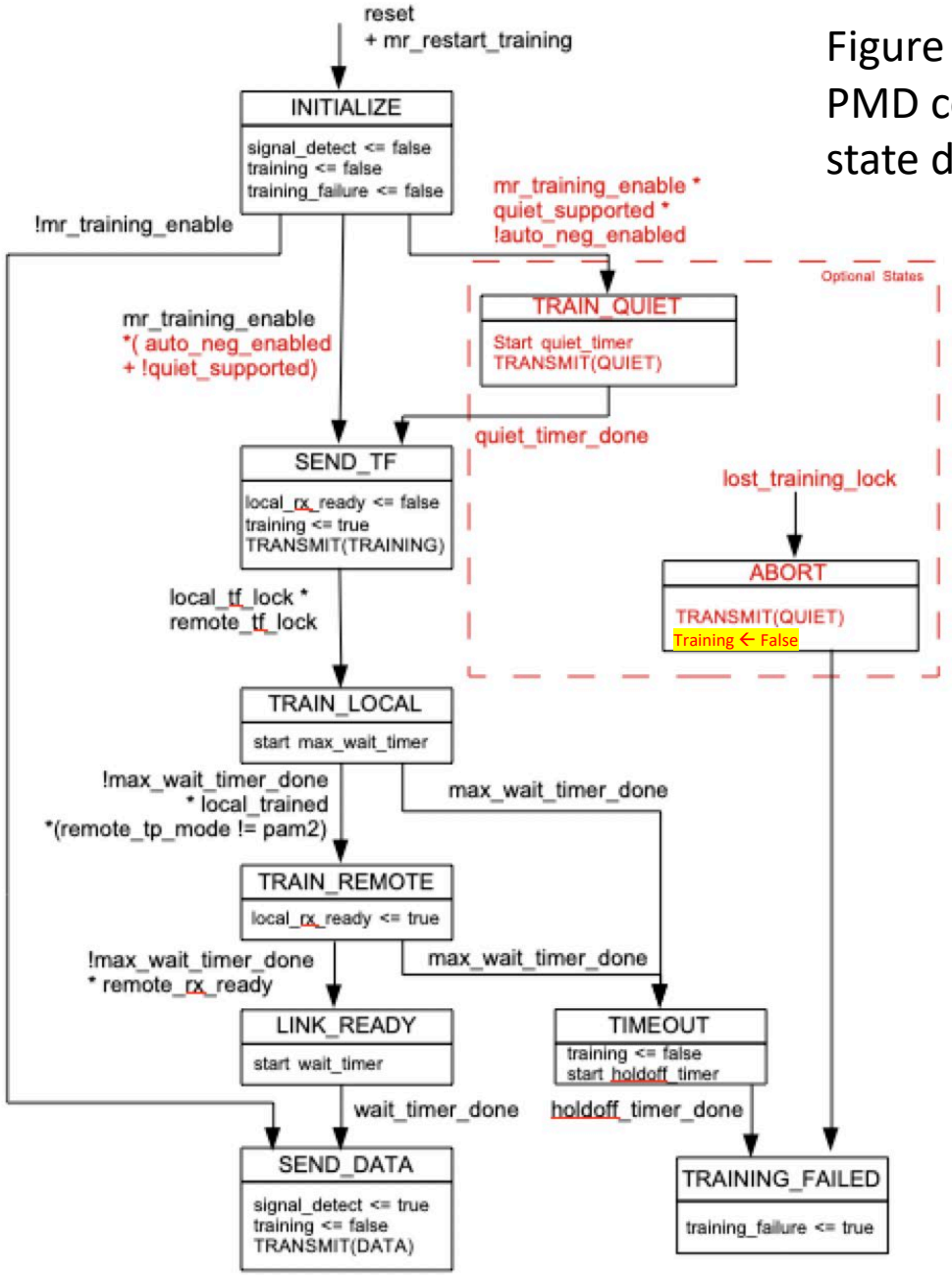

Figure 136-7 – PMD control state diagram

- **136.8.11.7.1 Variables**
  - auto_neg_enabled – Boolean variable equal to mr_autoneg_enable when auto-negotiation is supported (see Cl73) otherwise set to FALSE.
  - quiet_supported – Boolean variable set to TRUE if the optional QUIET state is supported, otherwise set to FALSE.
  - lost_training_lock – Boolean variable that is set to TRUE when the PMD control function is in TRAIN_LOCAL or TRAIN_REMOTE states (see Figure 136-7) and local_tf_lock is FALSE continuously for a period 20ms or the detection of a non-compliant input signal has occurred continuously for 1ms. When quiet_supported is FALSE, this variable is always FALSE.
- **136.8.11.7.3 Timers**
  - quiet_timer– This timer is started when the PMD control state diagram enters the TRAIN_QUIET state. The terminal count of quiet_timer is 80ms +/- 2%.

Additional Changes

- 162.8.11 PMD control function
  - G) When auto_neg_enabled is TRUE, a receiver is expected to assert local_tf_lock within 275 ms from entry into the AN_GOOD_CHECK state in Figure 73–11.

Fixed Rate bring-up thoughts

- Initial startup time increased by 80ms.
- Reset of one port mid-LT process ensures other port will reset prior to starting another LT max_wait_timer.
  - Port that is reset, can't zoom through startup and begin new LT process without forcing partner to reset as well. Thus resync of max_wait_timer is immediate not delayed by 1 max_wait_timer.
- Management sees a TRAIN_FAILED when "abort" occurs
- TIMEOUT and TRAINING_FAILED states continue to send LT FRAMES which can assist in debug of LT issues due to a max_wait_timer timeout.

AN bring-up thoughts

- AN startup of 20ms from AN_GOOD -> PMD frames is not affected.
- Failed AN startup (one side starts LT, other goes back to AN) causes side running LT to return quickly to AN.
- Reset of one side back to AN during LT. Side that didn’t get reset would quickly follow since AN begins in QUIET state for 50-75ms

Proposed Task Force Straw Poll

- I support updating the 100G/lane PMD Control function (i.e. Cl 162) in lusted_3ck_02_1020 using: (Choose one)
  - A. Option B3 (per slide 6, 8-9)
  - B. Option B4 (per slides 6, 11-13)

Backup

Solution Space Update

Some solutions include, but are not limited to:

Removed based on feedback in D1.2

- ~~Do nothing~~
- ~~Increase the duration of the holdoff_timer to exceed that of the max_wait_timer (>= 12 seconds)~~

Explored options

- Add monitoring of the local and received frame lock status (with TBD hysteresis) after the initial frame lock is achieved
- Implement an abort signaling mechanism

Consensus building is in progress to bring a detailed solution for TF consideration by the end of Oct 2020. Contact Kent for details

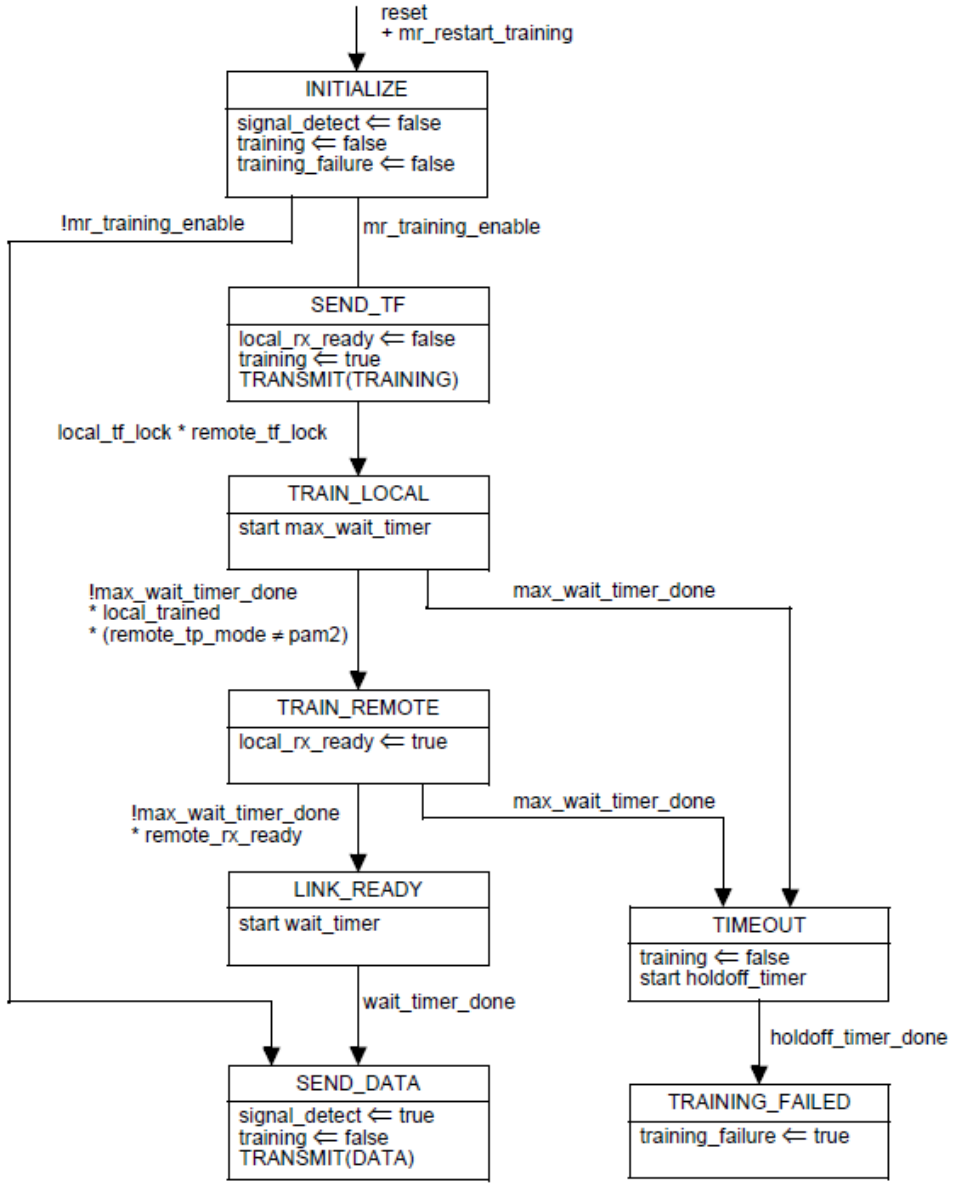

Figure 136–7—PMD control state diagram

Background

- In the IEEE 802.3cd-2018 project, an updated PMD Control Function (i.e. link training) was defined and specified in Cl 136.8.11
- Among other things, specific changes enabled the link training protocol to support link establishment between two devices without using Cl 73 Auto-Negotiation
  - For the customer use case of “forced PHY speed” on the link
  - See: http://grouper.ieee.org/groups/802/3/cd/public/adhoc/archive/slavick_101916_3cd_adhoc.pdf

Issue Summary

- The currently defined state machine in Clause 136.8.11 (Figure 136-7) does not autonomously recover from a partner restarting during link training in this “force PHY speed” case
  - Note: observed when the Clause 73 Auto-Negotiation state machine is not used.
- Unless a high-level management agent (i.e. SW or FW) detects the condition, the result could be either a persistent link down (i.e. link never comes up) or a link oscillation (up/down/up/down/etc).

Example Behavior

- Reset of Device #1 was not captured by Device #2
- The signals local_tf_lock and remote_tf_lock are only checked moving from the SEND_TF state to the TRAIN_LOCAL state.

  - Note: Cl 73 AN state machine (if it was used in this case) would expire link_fail_inhibit_timer and breakout out of the condition

Device #1	Device #2
INITIALIZE	INITIALIZE
SEND_TF	SEND_TF
TRAIN_LOCAL	TRAIN_LOCAL
__ERROR:_RESET	TRAIN_LOCAL
INITIALIZE	TRAIN_REMOTE
SEND_TF	TRAIN_REMOTE
TRAIN_LOCAL	TRAIN_REMOTE
TRAIN_LOCAL	TIMEOUT
TRAIN_LOCAL	TRAINING_FAILED
TRAIN_REMOTE	INITIALIZE
TRAIN_REMOTE	SEND_TF
TRAIN_REMOTE	TRAIN_LOCAL
TIMEOUT	TRAIN_LOCAL
TRAINING_FAILED	TRAIN_REMOTE
INITIALIZE	TRAIN_REMOTE
SEND_TF	TRAIN_REMOTE
TRAIN_LOCAL	TRAIN_REMOTE
TRAIN_LOCAL	TIMEOUT
TRAIN_LOCAL	TRAINING_FAILED
TRAIN_REMOTE	INITIALIZE
TRAIN_REMOTE	SEND_TF
TRAIN_REMOTE	TRAIN_LOCAL
TIMEOUT	TRAIN_LOCAL
TRAINING_FAILED	TRAIN_REMOTE
INITIALIZE	TRAIN_REMOTE
SEND_TF	TRAIN_REMOTE
TRAIN_LOCAL	TRAIN_REMOTE
TRAIN_LOCAL	TIMEOUT
TRAIN_LOCAL	TRAINING_FAILED

*** THE CYCLE CONTINUES ***