FEC Issues – PCS Lock SMs

Mark Gustlin – Cisco

IEEE Dallas 802.3ba TF

November 2008

Supporters

- Jeff Maki – Juniper
- Magesh Valliappan – Broadcom
- Faisal Dada – JDSU
- Norbert Folkens – JDSU
- Pete Anslow – Nortel
- Gary Nicholl – Cisco
- Divya Vijayaraghavan – Altera
- Andy Weitzner - Marvell

Introduction

- The following slides look at the interactions between how KR FEC marks errors and how the PCS goes in and out of lock or High BER states

Background

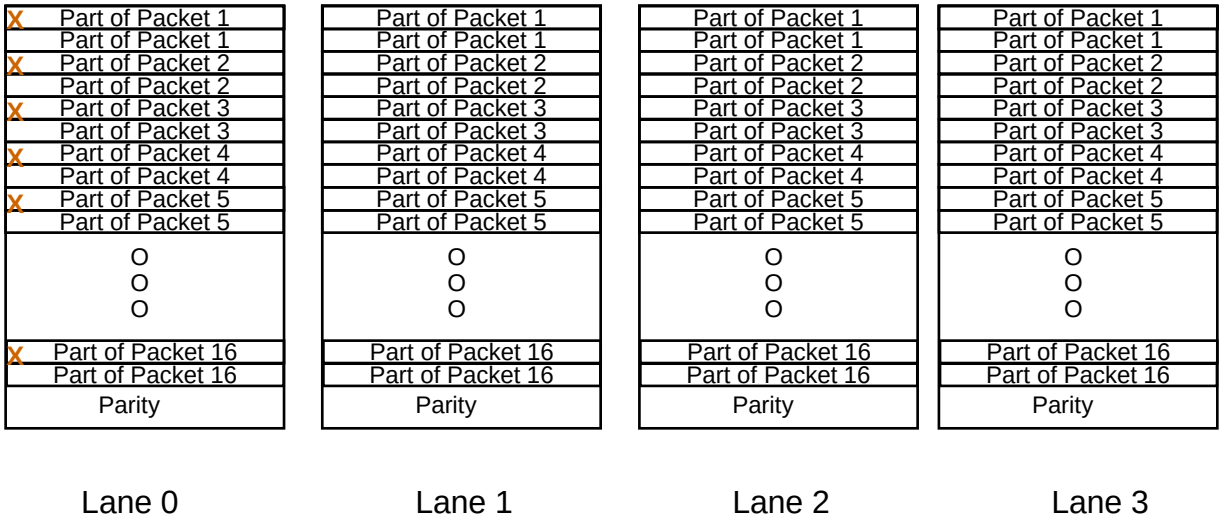
- In the original Clause 74 (FEC), when an uncorrectable error occurs, the FEC block marks the block bad by corrupting (to ’11’) the sync field of blocks 1, 9, 17, 25 and 32 (out of 32 66b blocks). This ensures all possible 64B packets that might be contained within the FEC block will be dropped by upper layers
- With 40GBASE-R/100GBASE-R PCS, packets are striped to multiple lanes, 8B at a time.
- FEC is run on each PCS lane, so a single FEC block has only slices of a packet
- With 40GBASE-R/100GBASE-R, to ensure all 64B packets are dropped, we have to mark at least 16 of 32 blocks bad for 40G, and all 32 blocks bad for 100G since packets are striped across many lanes

  In order to make the behavior of 40G and 100G consistent, I recommend that we mark 32 blocks bad for 40G also

- Once we do that though, it will cause havoc with the lock SMs, things will go out of lock (lane lock and high BER)

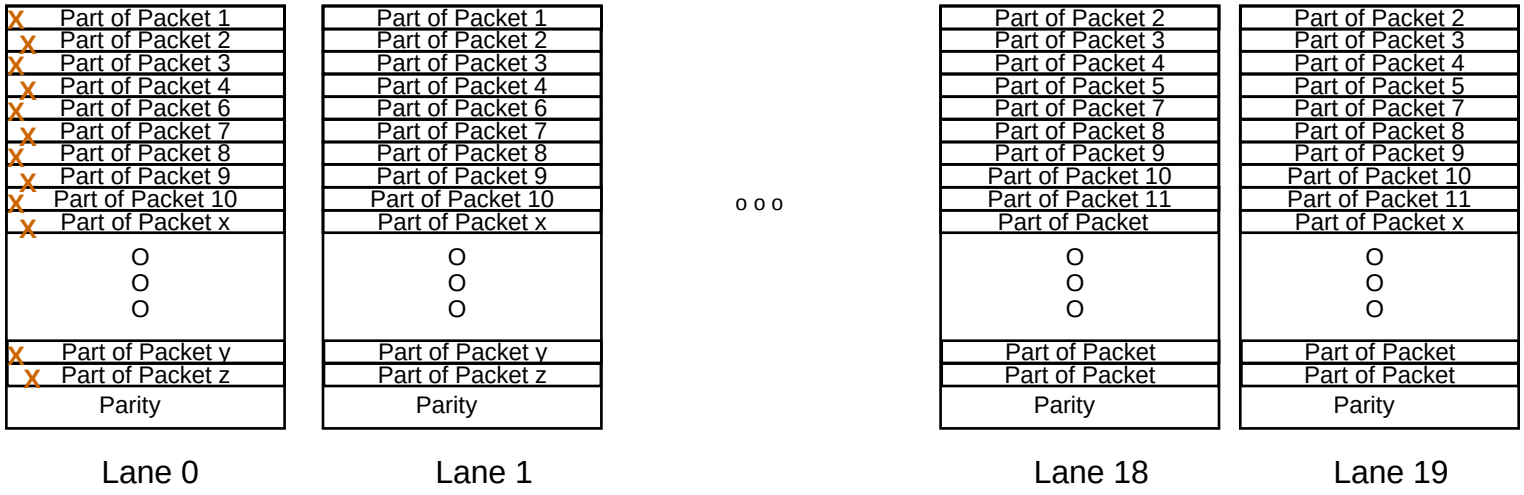
40G KR/CR FEC Block Marking

- When an uncorrectable error is encountered, you need to mark every other 66b block as bad to make sure all 64B or larger packets are dropped (this is in draft 1.0)
  - But for consistency, mark all 32 as bad



100G CR FEC Blocks

- When an uncorrectable error is encountered, you need to mark every 66b block as bad to make sure all 64B or larger packets are dropped (this is in draft 1.0)



10GBASE-R PCS SM Background

- Today with 10GBASE-R, there are two state machines operating on the receive stream and looking at the sync bits
  - Lock state machine
    - Looks for 64 non-errored sync blocks in a row to declare in lock
    - Looks for 16 errored sync blocks out of 64 to declare out of lock
  - BER state machine
    - Looks for 16 errored sync blocks out of a 125usec window to declare high BER
    - High BER ~ 10^{-4} (without FEC)

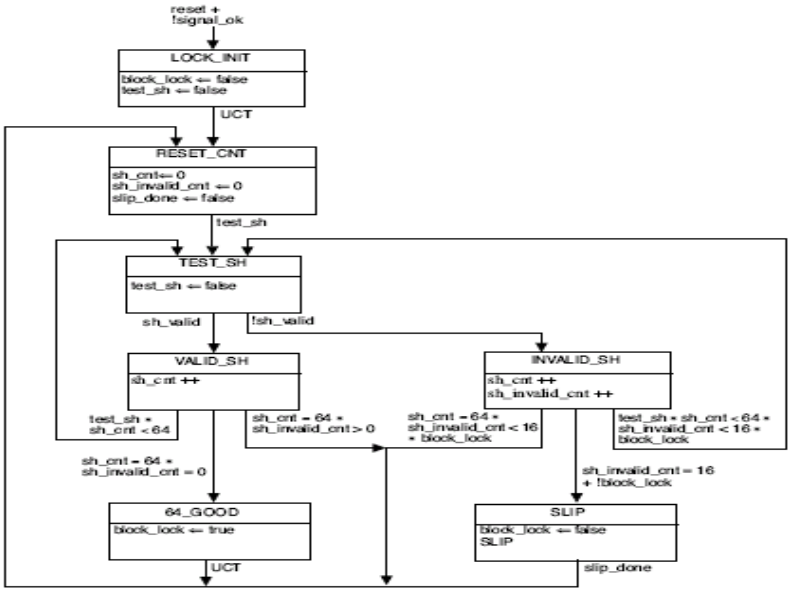


Figure 49–12—Lock state machine

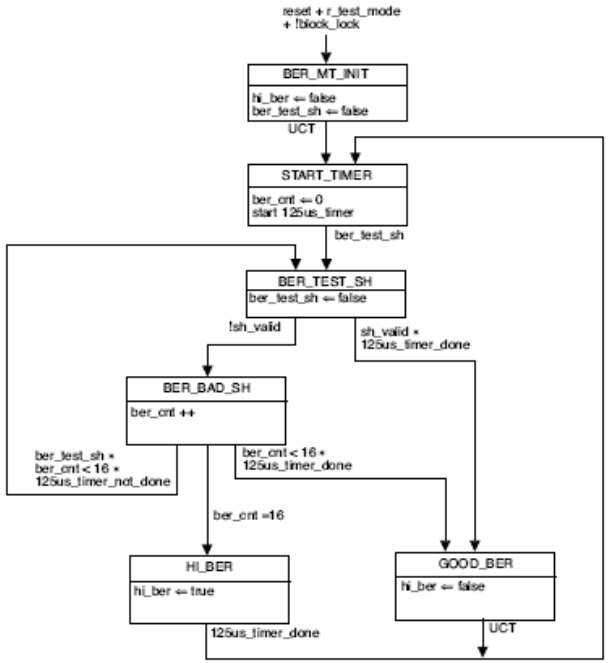


Figure 49–13— BER monitor state machine

Current 10GBASE-R SM Operation

- For an uncorrectable FEC block, the FEC sublayer corrupts 5 sync fields in one 32 sync field window

  If marking is enabled

- The Lock SM survives two uncorrectable FEC blocks in 64 66b blocks and stays in sync

  Lock SM looks for 16 out of 64 in error for declaring loss of lock, 2 uncorrectable FEC blocks results in 10 errors in 64 sync fields

  Each FEC block is 32 66b blocks, so the Lock SM will never go out of lock just from uncorrectable FEC blocks

- For the BER SM, it looks for 16 errored syncs in 125usec (19531 66b blocks), or up to 3 uncorrectable FEC blocks could occur in the 19k blocks and we would still keep lock.

  Three uncorrectable FEC blocks equals 15 marked bad 66b blocks

40/100GBASE-R PCS SM Background

- In Draft 1.0, the SMs are the similar to those for 10GBASE-R:

  Lane Lock state machine

  Looks for 64 non-errored sync blocks in a row to declare in lock on a per lane basis

  Looks for 16 errored sync blocks out of 64 to declare out of lock on a per lane basis

  BER state machine

  Looks for 16 errored sync blocks out in 31.25usec for 40GE (or 12.5usec for 100GE) window to declare high BER on the aggregate 40G or 100G signal

  High BER ~ 10^{-4}

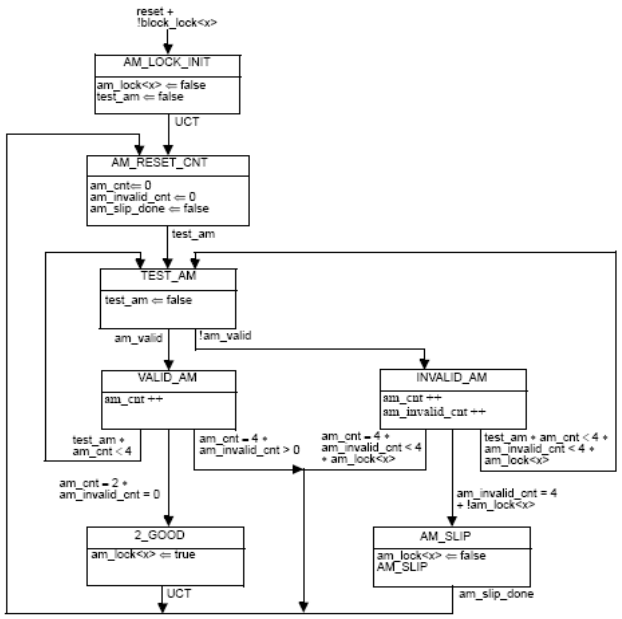


Figure 82–13—PCS alignment marker lock state diagram

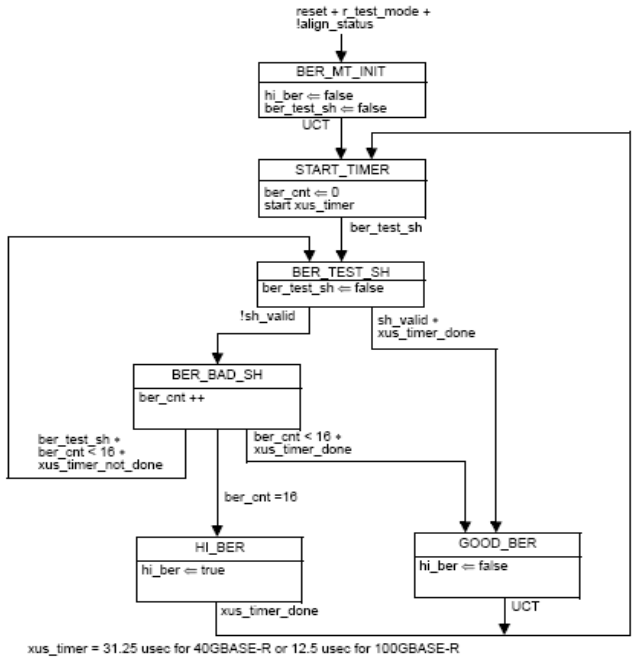


Figure 82–15—BER monitor state diagram

Current 40GBASE-R SM Operation

- The FEC sublayer corrupts 16 sync fields in one FEC block
  - Out of 32 66b blocks
  - Change to corrupting all 32 blocks for consistent behavior
- The Lane Lock SM will go out of lock with just one errored FEC block
  - Lock SM looks for 16 out of 64 in error for declaring loss of lock, 1 errored FEC blocks results in 32 errors in 32 sync fields
- For the BER SM, it looks for 16 errored syncs in 31.25usec (19531 blocks)
  - The BER SM goes into high BER state after a single uncorrectable FEC block due to 32 marked bad 66b blocks

Current 100GBASE-R SM Operation

- The FEC sublayer corrupts 32 sync fields in one FEC block
  - Out of 32 66b blocks
- The Lane Lock SM will go out of lock with just one errored FEC block
  - Lock SM looks for 16 out of 64 in error for declaring loss of lock, 1 uncorrectable FEC block results in 32 errors in 32 sync fields
- For the BER SM, it looks for 16 errored syncs in 12.5usec (19531 blocks)
  - The BER SM goes into high BER state after a single uncorrectable FEC block due to 32 marked bad 66b blocks

Solution Options

- How to allow for FEC marking without causing problems with lock times?
- Options:
  1. Corrupt the sync bits still, and modify the SMs accordingly

     1a. Should SMs be modified only if FEC is active?

     1b. Or have the SMs run the same even if FEC off?
  2. Out of band signal that informs the PCS of the errors
     - This won't easily allow the FEC to be implemented separately from the MAC/PCS

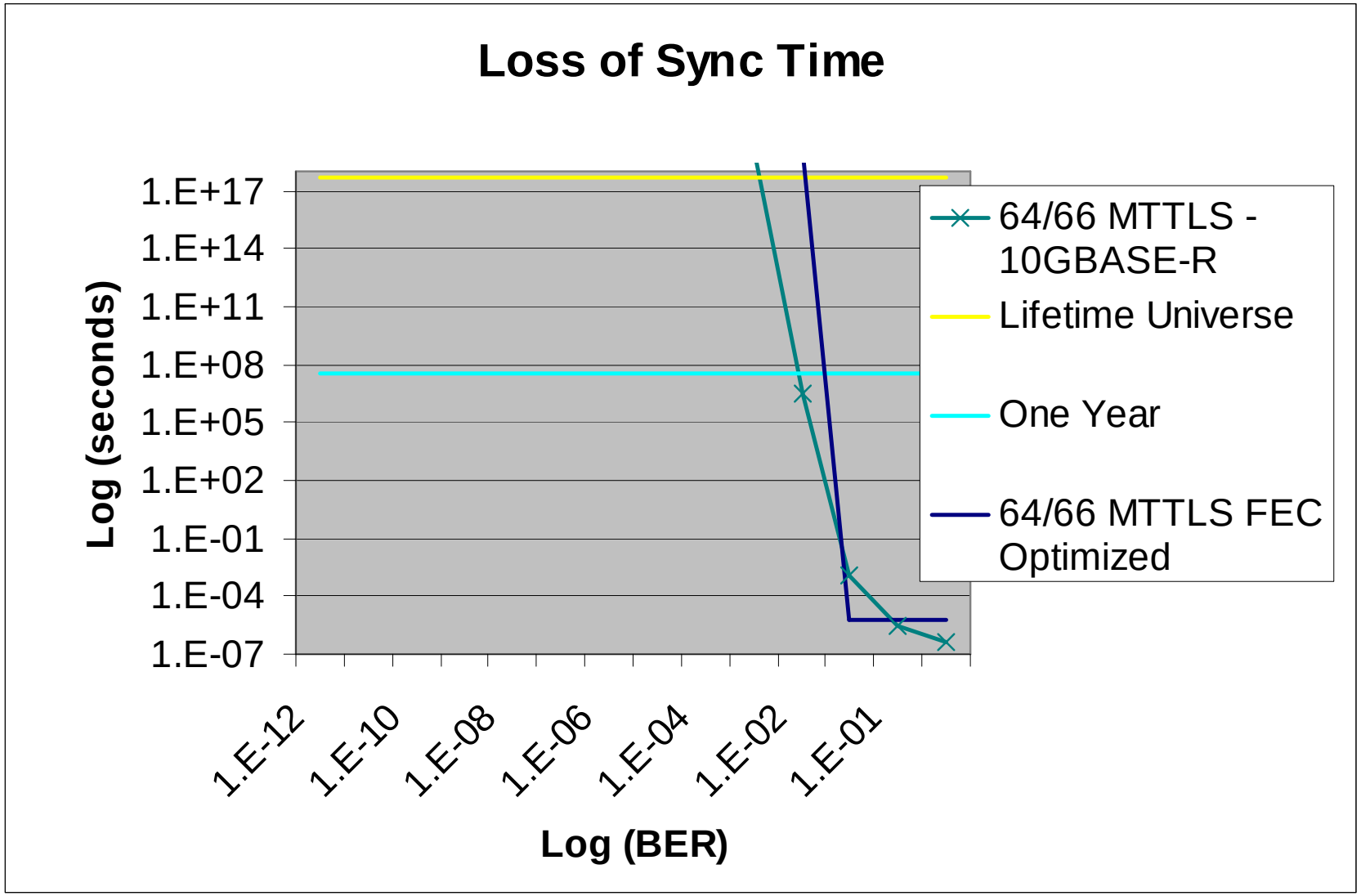
Best solution seems to be #1b

Note in 10GBASE-R PCS SMs run the same with or without FEC

Proposal – Lane Lock SM for 100/40G

- Lock SM operates on a per PCS lane basis
- Out of lane lock to in lane lock takes 64 non errored syncs in a row
  - Same as 10GBASE-R, same lock time
- In lock to out of lock takes 65 errors out of a 1024 sync window
  - For applications with or without FEC
- Performance is similar to the previous SM
- For 100GBASE-R and 40GBASE-R it allows 2 uncorrectable FEC blocks while staying in lock
  - Out of 32 FEC blocks (each with 32 66b blocks)

Proposal – Lane Lock SM for 100/40G



Performance without FEC

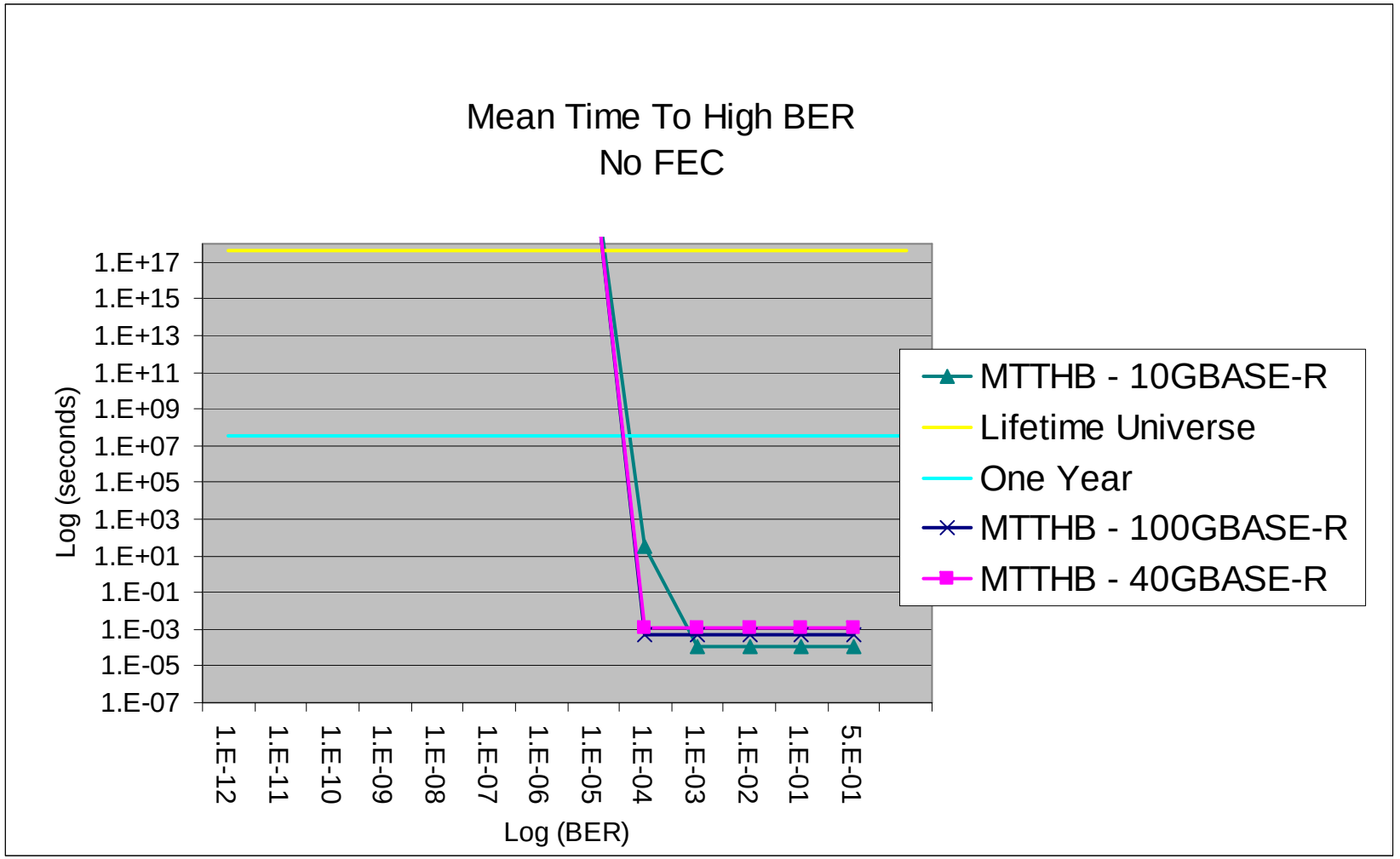
MTTLS = Mean Time To Lose Sync

Proposal – BER SM for 40/100G

- BER SM operates on a per interface basis
- In D1.0 high BER is declared when there are at least 16 sync errors in a 12.5usec (100G) or 31.25usec (40G) window
- Proposal is to change that to 97 sync errors in 500usec for 100G, 1.25msec for 40G
- Performance is similar to previous SM (without FEC)
- For 100GBASE-R and 40GBASE-R it allows 3 uncorrectable FEC blocks without going into the High BER state

  Out of 24413 FEC blocks (each with 32 66b blocks)

Proposal – BER SM for 40/100G



Performance without FEC

MTTHB – Mean Time To High BER

FEC Related Questions

- What is the lock and loss of lock FEC lock performance when compared to the PCS SM performance?

  Assume FEC is in the same chip as the PCS (so no extra errors added)

- If “4” consecutive blocks are received with good parity, report Block Sync
- If “8” consecutive blocks are received with bad parity, drop Block Sync

From clause 74: The FEC Synchronization process sets the sync_status flag to the PCS function to indicate whether FEC has obtained synchronization.

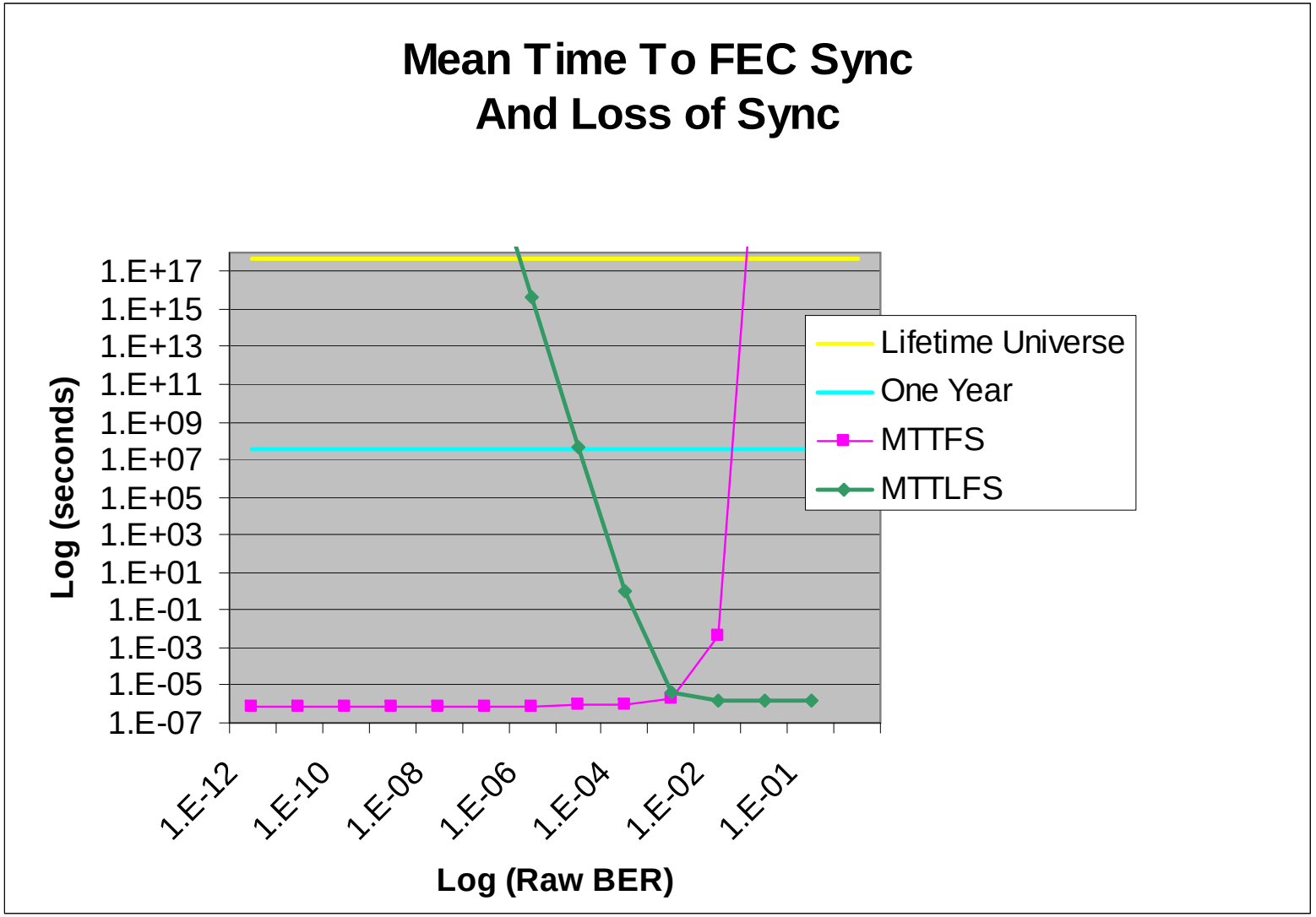
Q: What does the PCS do in the mean time, how is it notified??? What does FEC block send when it is not sync’d?

Seems that this is not specified yet, this would be needed if the PCS and FEC are in separate chips!

- A note about FEC marking, it essentially is multiplying errors after a certain point, that is why with FEC on Hi BER will trigger earlier

  It turns two single bit errors into 32 errored blocks!

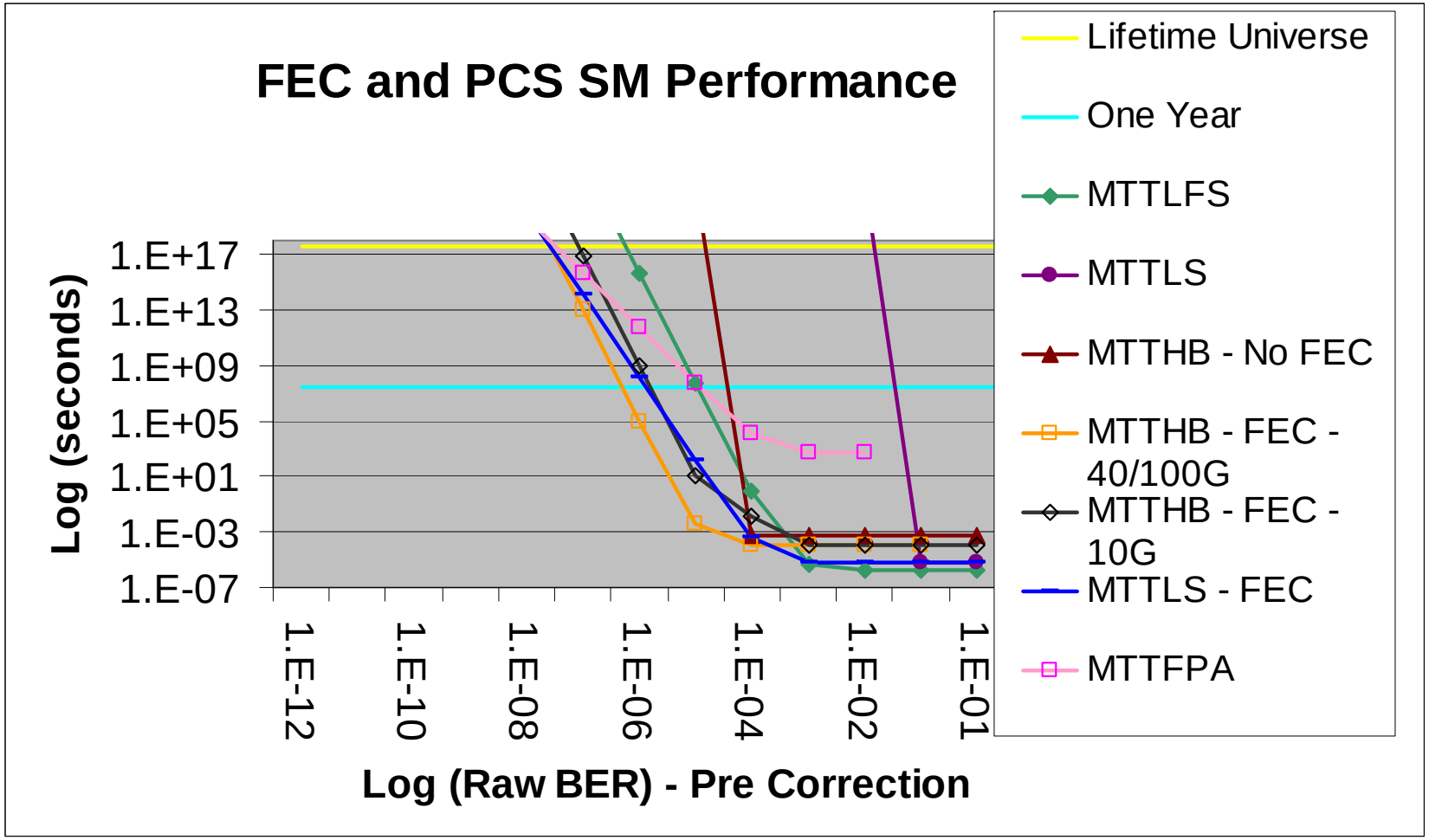
FEC Performance



MTTFS – Mean Time To FEC Sync

MTTLFS – Mean Time To Loss of FEC Sync

All Loss Curves on a Graph



MTTLFS - Mean Time To Lose FEC Sync
MTTLS - Mean Time To Lose Sync (66b)
MTTHB – Mean Time To High BER (no FEC)

MTTHB-FEC-100G - MTT High BER with FEC on
MTTHB-FEC-40G - MTT High BER with FEC on
MTTFPA – Mean Time To False Packet Acceptance (assumes uncorrelated errors)

Open Questions

- Is it ok that the behavior with FEC is different than the behavior without FEC for the High BER SM?

  This is also true for 10GBASE-R

- Should Error marking be mandatory for the copper interfaces?
- How should the FEC sublayer tell the PCS when it is out of lock?

  D1.0 Comment is to just pass data along to the PCS

- Is the FEC being used for improving on a BER of 10^{-12}, or making a link get up to 10^{-12}?

  Assume that it is making a 10^{-12} BER better

How to Indicate Out of FEC Lock?

- Today's clause 74 FEC only indicates out of FEC lock through a primitive which is part of the 16b wide bus structure of the current primitives
- We should have a way to indicate out of FEC lock via the serial interface that we now have for 40/100G
- Options:
  1. Do nothing, just pass through the data to the PCS block without recreating the 66b blocks
     - If it is FEC data coming in, and the FEC block can't sync up, if it is sent to the PCS as is, the PCS won't sync up and the interface is down
  2. Send all scrambled data (all bits)
  3. Send data but corrupt all sync fields
     - But if out of FEC lock, the FEC rx does not know where the sync field is!
     - If far end is not sending FEC encoded data, and it has valid sync fields we should make sure that does not go up to the PCS uncorrupted?
  4. Send a fixed pattern like all 1's?
     - But then the serdes go down?
- Option 1 seems the best way to go…