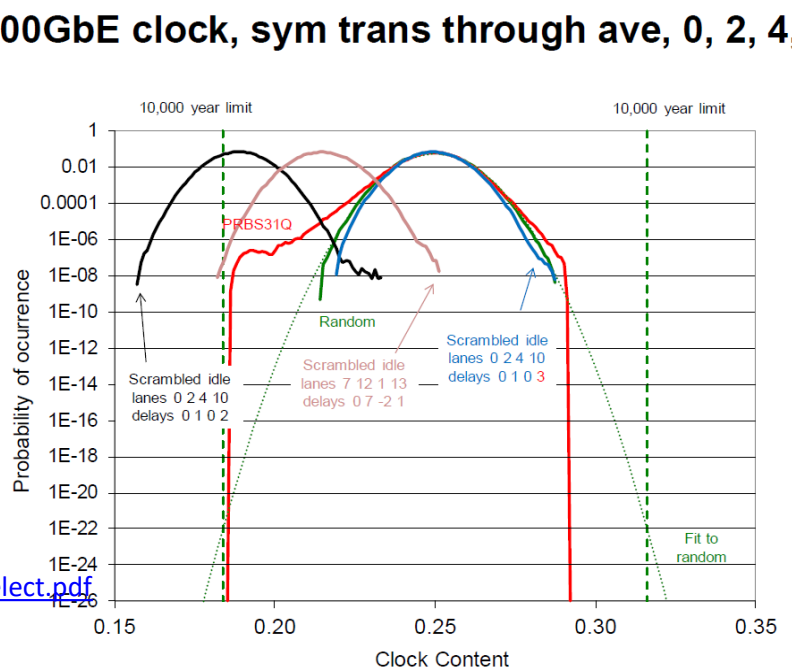
Multiplexing rules and impairments

Piers Dawe Nvidia
Zvi Rechtman Nvidia

Introduction

- With the 400G PCS/FEC and unrestricted bit multiplexing into 8 x 53 GBd PAM4 PMA lanes:
  - Usually, the content of each PMA lane is well scrambled and appears close to random (blue)
  - But for particular lane combinations and skews, when transmitting scrambled idle, a significant fraction of LSBs are the same as the preceding LSB (black, brown)
  - The transition density on the line ("clock content") is defective. See wertheim_3bs_01_0317* for explanation
  - *Example from anslow_01_121916_elect*^



*https://ieee802.org/3/bs/public/17_03/wertheim_3bs_01_0317.pdf

^https://ieee802.org/3/bs/public/adhoc/elect/19Dec_16/anslow_01_121916_elect.pdf

Introduction 2

- There are a huge number of ways to avoid this and a tiny fraction that cause it: see anslow_01_121916_elect
- wong_01_0117_logic* indicated that restricted bit muxing ("natural pairs") avoids the problem, and so does 1+D precoding, but switching off the Gray coding would not help.
- The damage done in dB of SNR, or risk of losing lock, or bad jitter transfer is hard to quantify; it very much depends on CDR implementation
  - The CDR used in compliance testing (jitter, TDECQ) may be very different to a particular product receiver
- This problem was discovered late in the development of 802.3bs; too late to require a changed transmitter in 802.3bs so only a warning was added, making it the receiver's problem

*https://ieee802.org/3/bs/public/adhoc/logic/jan26_17/wong_01_0117_logic.pdf

Round 2

- Five years later, everyone has had plenty of notice. With new designs to a new spec, as we debate restricted bit muxing (D1.0 173.4.2, ran_3df_01_2212*, comments 6, 166, 167, 169) we have the opportunity to repartition from an unpredictable burden on the receiver to a negligible burden on the transmitter
- Rules that address multiplexing from the FEC point of view, as in shrikhande_3df_01a_221004^ and D1.0 173.4.2) give a fraction of a dB SNR benefit (see opsasnick_3df_logic_220630a~) *at the allowed error ratio limit for a whole multi-section link*…
- But restricting the multiplexing for transmission over one or two AUIs alone has negligible FEC-related benefit
  - AUI max BER 1e-5 or 2e-5 max vs. 2.6e4 for 2 AUIs and one optical PMD, so most of the errors are unaffected by the fine details of FEC performance on Rx-side AUI

* https://ieee802.org/3/df/public/22_12/ran_3df_01_2212.pdf

^ https://ieee802.org/3/df/public/22_10/22_1004/shrikhande_3df_01a_221004.pdf

~ https://ieee802.org/3/df/public/adhoc/logic/22_0630/opsasnick_3df_logic_220630a.pdf

Mostly the same for 8 x 100G as for 4 x 100G

- The 800G PCS is substantially two 400G PCSs, and the transition density issue is the same: see wong_3df_logic_220630*
- Use of "natural pairs" or "equivalent pairs" fixes the transition density problem
  - In 800G we can have skew between the two scramblers as well as between lanes – does this change the conclusions?
- Are the rules in D1.0 173.4.2 enough to fix the transition density problem?
- Comment 6 and ran_3df_01_2212 show that the rules in D1.0 173.4.2 don't achieve their objective with regard to FEC performance. Do the changes proposed there fix the transition density problem?

* https://ieee802.org/3/df/public/adhoc/logic/22_0630/wong_3df_logic_220630.pdf

New correlation issue

- The two 400G flows in the PCS are the same apart from alignment markers, and they are synchronized. When fed by e.g. idle or RF or LF, their inputs are identical.
- So, if the scramblers start with the same seed at the same time, their outputs will be the same: PCS lane 1 will be the same as PCS lane 16 and so on, until something happens to break the symmetry
- If PCS lanes 1 and 16 are components of the same PMA lane (as they would be with "natural pairs" or "equivalent pairs", and the skews are very low as is natural, the result will be unsatisfactory
- But we want the link to come up when it starts!

Less of an issue

- That is the opposite to the situation in training, where there the state machines and scramblers for each lane are independent

What to do

- We should instruct implementers:
  - Not to transmit a signal from the PCS that a compliant PMA could multiplex to form a signal with defective clock content
    - There are a huge number of ways to get this right
  - Not to start the PCS scramblers so that the two PCS flows have the same content
  - To ensure that in training, adjacent PMA lanes do not have the same content; this can be done by a combination of PRBS seed and skew

What not to do

- We should not instruct implementers to do things that are ineffective and unnecessarily contradict other specifications
  - Instead, the default training seeds 4 to 7 should be the same as seeds 0 to 3, as in the ETC spec*
- No reason to require that the 8:32 PMA should obey multiplexing rules for the sake of FEC performance alone, as there are only a few PMA errors that would be poorly distributed
  - However - is the 8:32 PMA a thing that should be standardized in its own right, or is it an upside-down 32:8 PMA?
  - It's less confusing to describe it separately, but a there would not be a separate gearbox IC design for each

* https://ethernettechnologyconsortium.org/wp-content/uploads/2021/10/Ethernet-Technology-Consortium_800G-Specification_r1.1.pdf

Thanks!