100GBASE-KP4 Training Consensus Building Call

Kent Lusted, Intel
Adee Ran, Intel
Matt Brown, AppliedMicro

(Regarding Comment #38)

Supporters

- Andre Szczepanek, Inphi
- Arash Farhood, Cortina Systems
- Brad Booth, Dell
- Dave Chalupsky, Intel
- Rich Mellitz, Intel
- Vasu Parthasarathy, Broadcom
- Ilango Ganga, Intel
- Arthur Marris, Cadence
- Will Bliss, Broadcom
- Stephen Bates, PMC Sierra

Assumptions

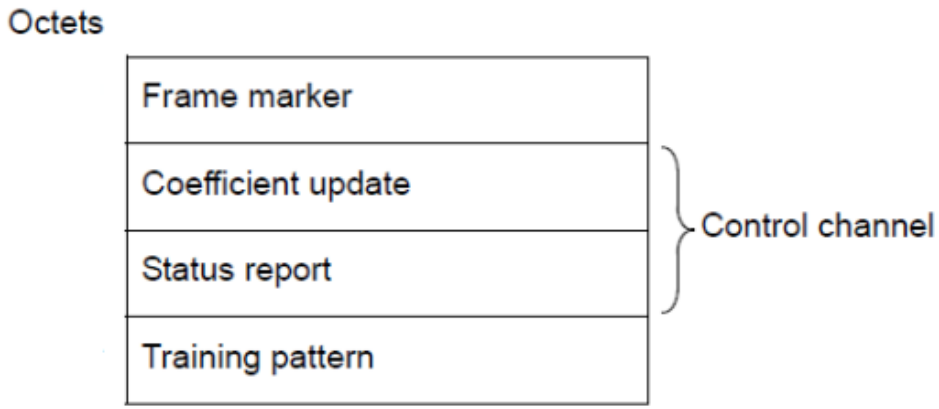
- Based on P802.3bj Draft 1.0 and P802.3bh Draft 3.1
- Maximize compatibility with existing training mechanism in Clause 72.6.10
- Leverage the tight integration and inter-dependence of the PMA and PMD functions, specific to Clause 94.
  - Similar to PCS/PMA dependency in 10GBASE-T link training (Clause 55)

PMD Control Function

- Take Clause 72.6.10 (10GBASE-KR PMD Control) as the baseline
  - Keep as much in common as possible, (i.e. training state diagram, coefficient update process, etc.)
- Modify as given in the following slides

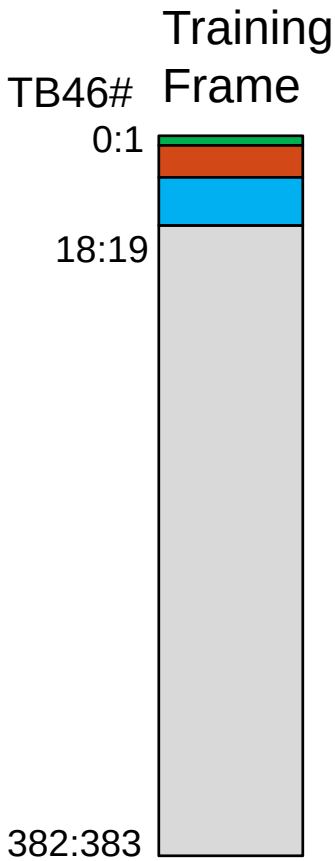
Training Frame Structure

- Keep same basic format as Clause 72
  - Frame marker
  - Control channel
    - Coef update field
    - Status report field
  - Training pattern



100GBASE-KP4 Change Preface

- Pack the Frame Marker, Control Channel and Training Pattern into 46 UI training frame words (TFW)
  - Each TFW corresponds to two full 46-bit terminated blocks (TB46).
  - Simplifies design (i.e. no gearbox)
  - Enables early data alignment during training period
  - Enables regular data recovery and fast switching to data mode
- Use PAM4 signaling in the training pattern
- At end of training process, make it easy to lock to the correct offset in PMA frame
- Add parity check to both control channel fields



384 x 46 = 17764 PAM4 symbols in 1 training frame
1 Training Frame = 384 Termination Blocks (TB46)
1 Training Frame = 192 Training Frame Words (TFW)

Easy Lock to PMA Frame

- Training Frame alignment to the Overhead Frame logic is important for rapid transition from the training mode to data mode
  - 100GBASE-KP4 PMA frame consists of 696 TB46
  - Proposed 100GBASE-KP4 training frame is 384 TB46
  - Gcd(696, 384) = 24
  - The offset is periodical with a cycle of 696/24=29 training frames
- Training frames shall always be transmitted with offset which is $24 \cdot k,\ k \in \{0 \ldots 28\}$ words relative to the 40-bit overhead
  - Other offsets are illegal
  - Assuming the first training frame starts at offset 0 (k=0), then the second training frame starts at offset 384 (k=16), and subsequent training frames start at offsets with k equal to 3, 19, 6, 22 …

Frame Marker and Control Channel

- To simplify link training in the link establishment process when TRANSMIT(TRAINING), the PAM4 multi-level signaling shall not be used for the Frame Marker, Coefficient Update, and Status Report fields.
  - Make it full-swing, i.e. “NRZ” like
- During the transmission of the Frame Marker, Coefficient Update, and Status Report fields, the PAM4 transmitter *shall bypass the overhead frame, termination block, gray coding, and 1/(1+D) mod 4 precoding stages of the PMA transmit and receive functional specifications*.
  - *Therefore, the output levels shall be restricted to level -1 for a 0 and level +1 for a 1 to enable easy receiver lock to the training pattern over poor quality and non-equalized channels.*

Frame Marker – TFW #1

- Frames are delimited by the 46 PAM4 symbol pattern, 23 +1 symbols followed by 23 -1 symbols, +1s first, as expressed in 13.59375 Gbd symbols.
- This pattern does not appear in the control channel or the training pattern and therefore serves as a unique indicator of the start of a training frame.

		UI						
TFW	**T-block**	45:36	35:26	25:16	15:6	5:0		
0	0:1	1111111111	1111111111	1110000000	0000000000	000000		frame marker

Control Channel Cells

- Signaled using differential Manchester encoding (DME) like in Clause 72.6.10.2.2
- Each DME cell contains 2 DME transition positions.
  - Each transition position is the mid-point of the cell.
- The data cell length is 10 100GBASE-KP4 PAM4 symbols. (~736ps)
  - Approximately the same duration at 10GBASE-KR cells
  - Transition position is the mid-point of the cell.

Control Channel Encoding

- Control channel uses 9 TFWs
  - TFW #1 to #9
- Pack 4 data cells of control channel into 1 TFW
  - Cells take definitions and ordering as shown in slides 13-14.
    - Coef Update Field before Status Update Field
    - Highest # cell first, in descending order
  - Uses 40 of 46 PAM4 symbols in 1 TFW

What about the Remaining 6 UI?

- Define the last 6 PAM4 symbols in each of TFW #2-11 as overhead cell
  - Transition position is 3 PAM4 symbols
- Set overhead cell to a DME logic ‘1’ of 6 PAM4 symbols width
  - Preserve the DC balance on the line
  - 000111 or 111000, depending on previous cell value
    - Same DME coding rules as before

Coef Update Field

11:7 Reserved
6 Parity Check

Cell ordering not finalized

Table 72–4—Coefficient update field

Cell(s)	Name	Description
15:14	Reserved	Transmitted as 0, ignored on reception.
13	Preset	1 = Preset coefficients 0 = Normal operation
12	Initialize	1 = Initialize coefficients 0 = Normal operation
11:6	Reserved	Transmitted as 0, ignored on reception.
5:4	Coefficient (+1) update	5 4 1 1 = reserved 0 1 = increment 1 0 = decrement 0 0 = hold
3:2	Coefficient (0) update	3 2 1 1 = reserved 0 1 = increment 1 0 = decrement 0 0 = hold
1:0	Coefficient (–1) update	1 0 1 1 = reserved 0 1 = increment 1 0 = decrement 0 0 = hold

Status Report Field

- 20 cells
  - 5 TFWs
  - 10 TB46
- Add new features
- Keep coef status and move Receiver Ready
- Cell ordering not finalized

Cell(s)	Name	Description
19	Parity Check	Parity calculation for Status Report Field
18:14	EEE State	Current EEE state of local transmitter, if EEE is implemented.
13:12	Training Frame Countdow	Number of training frames remaining before link training process transitions to data mode
11:7	PMA Alignment Offset	Relative location of the next training frame within the PMA frame
6	Receiver ready	1 = The local receiver has determined that training is complete and is prepared to receive data. 0 = The local receiver is requesting that training continue.
5:4	coefficient (+1) status	5 4 1 1 = maximum 1 0 = minimum 0 1 = updated 0 0 = not_updated
3:2	coefficient (0) status	3 2 1 1 = maximum 1 0 = minimum 0 1 = updated 0 0 = not_updated
1:0	Coefficient (-1) status	1 0 1 1 = maximum 1 0 = minimum 0 1 = updated 0 0 = not_updated

Two New Parity Check Fields

- This is an improvement over the original clause 72 rules
  - Guarantees DC balance of DME cells during training
    - Coef update and status report fields always starts with +1 PAM4 symbols
  - Increases protection against false acceptance of sensitive messages, e.g. preset, init, receiver ready
- Use cell 6 of the coef update field and cell 19 of status report field to encode a parity check for each respective field
  - The two parity bits are calculated for each field (coef update, status report) separately, such that each field *as a whole, including the parity bit* has even parity
  - Parity is the number of logical-one cells in the field, modulo 2 (not including the overhead bits)
- If a parity violation is detected within the bounds of the respective field in a given training frame, the contents of that field for that frame shall be ignored.
  - i.e. parity error in status report only affects status report, not coef update

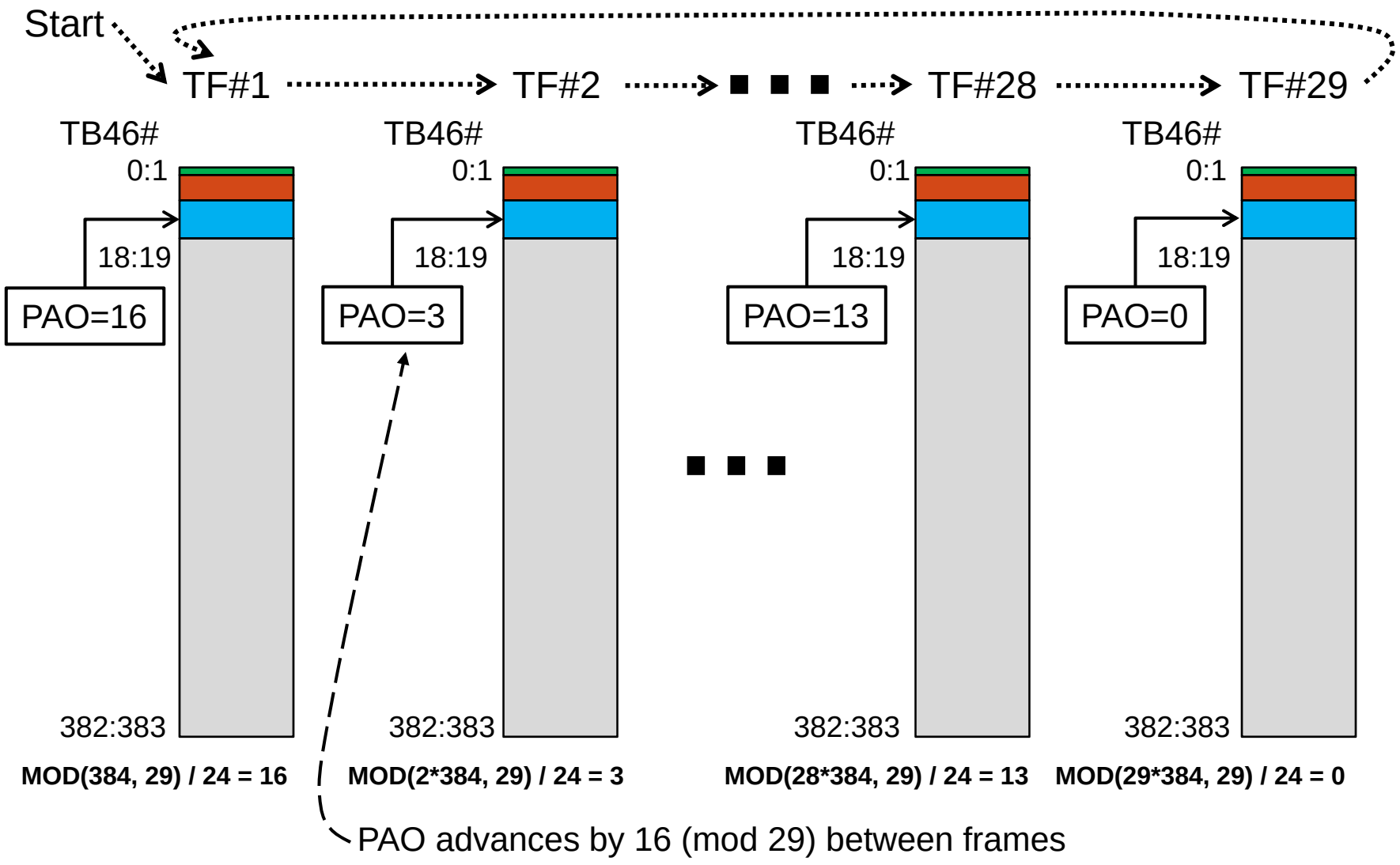
New Status Report Cells

- EEE State (Cells 18:14)
  - Current EEE state of local transmitter, if EEE is implemented
  - Otherwise, reserved and set to 0.
  - See brown_0x_xxxx.pdf for more details
- Training frame countdown counter (Cells 13:12)
  - Used to signal the transition from training to data mode.
  - Start at 3, decrement toward 0 during the last 3 frames sent (2 -> 1 -> 0)
    - 3 indicates 3 or more frames remaining
  - When a frame is sent with this value = 0, after the last TFW of training pattern is sent, transmission immediately switches to the PMA frame

New Status Report Cells (2)

- PMA Alignment Offset -- PAO (Cells 11:6)
  - To shift instantly to data mode after the last training frame ends, RX needs to know the relative offset of the first data block (TB46) from the 40-bit overhead within the PMA frame
  - PAO encodes the relative location of the TB46 after the end of the training frame (mod 696) as a 5-bit integer
  - The start of the next training frame is 24 * PAO
    - 0: marker aligned with 40-bit overhead
    - 1: marker is at offset of 24*1=24 termination blocks from 40-bit overhead
    - …
    - 28: marker is at offset 24*28=672 termination blocks from 40-bit overhead
    - 29 to 31: invalid, never transmitted, ignored on reception

PMA Alignment Offset Example



1 Training Frame = 384 Termination Blocks (TB46)

Frame Marker and Control Channel Mapping Ordering

Symbol TX order
(send row before advancing block)

Block TX order

TFW	T-block	UI 45:36	35:26	25:16	15:6	5:0	
0	0:1	1111111111	1111111111	1110000000	0000000000	000000	frame marker
1	2:3	cell 15	cell 14	cell 13	cell 12	overhead	coef update
2	4:5	cell 11	cell 10	cell 9	cell 8	overhead	coef update
3	6:7	cell 7	cell 6	cell 5	cell 4	overhead	coef update
4	8:9	cell 3	cell 2	cell 1	cell 0	overhead	coef update
5	10:11	cell 19	cell 18	cell 17	cell 16	overhead	status report
6	12:13	cell 15	cell 14	cell 13	cell 12	overhead	status report
7	14:15	cell 11	cell 10	cell 9	cell 8	overhead	status report
8	16:17	cell 7	cell 6	cell 5	cell 4	overhead	status report
9	18:19	cell 3	cell 2	cell 1	cell 0	overhead	status report

Training Pattern Motivation

- Use the PMA transmit and receive functional specifications as currently defined in P802.3bj Draft 1.0 to enable the transmitter and receiver to exercise termination block, gray coding, and 1/(1+D) mod 4 precoding stages.
  - Overhead framer does not have to be exercised, but alignment is tracked through PMA Alignment Offset (PAO) cell
- Generate multi-level PAM4 signaling for receiver calibration
- Choose a pattern that is PMA termination block friendly

Training Pattern Details

- Apply termination, Gray coding, 1/(1+D) mod 4 precoding and PAM4 mapping
- PRBS13 is used for training pattern generation, followed by its inverse
  - 8191 bits are generated from the LFSR (a full PRBS13 cycle)
  - Additional 8189 bits are generated and sent inverted (a full PRBS13 cycle minus 2 bits)
  - Last 2 bits of the inverse PRBS13 are discarded, so PRBS uses 16380/45=364 TB46
  - 4 initial seeds selected to create desired properties
- Pack each 45 bits of PRBS13 into TB46
- Training frame (frame marker, control channel and training pattern) contains 384 TB46

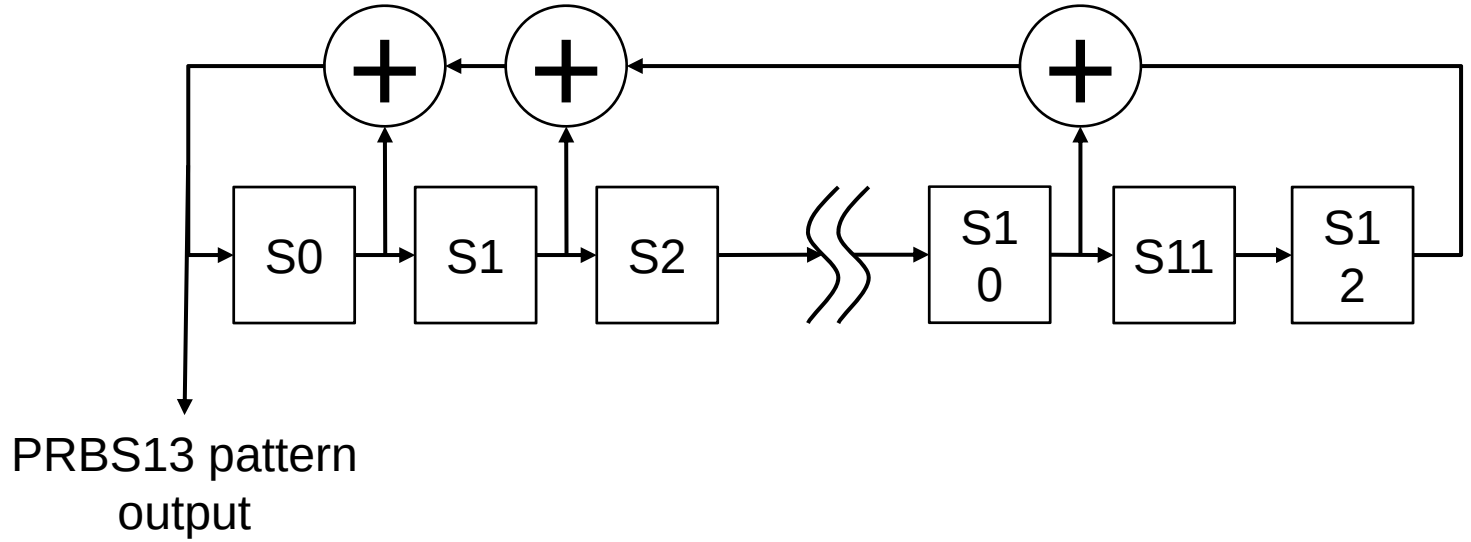
Training Pattern Initial State

- Use four unique states to create a distinct DC balanced sequence for each lane
  - Helps with lane order identification later on.
  - After generating 8191+8189=16380 bits, the next training frame starts (marker and control channel); LFSR state is re-initialized so that the training pattern starts with the same bits for all training frames.
- PRBS state and precoder state shall not advance during frame marker and control channel transmission and is reset for each training frame.

PRBS13 pattern generator

- The PRBS13 generator shall produce the same result as the implementation shown below, which implements the Fibonacci generator polynomial:

$$G(x) = 1 + x + x^2 + x^{11} + x^{13}$$



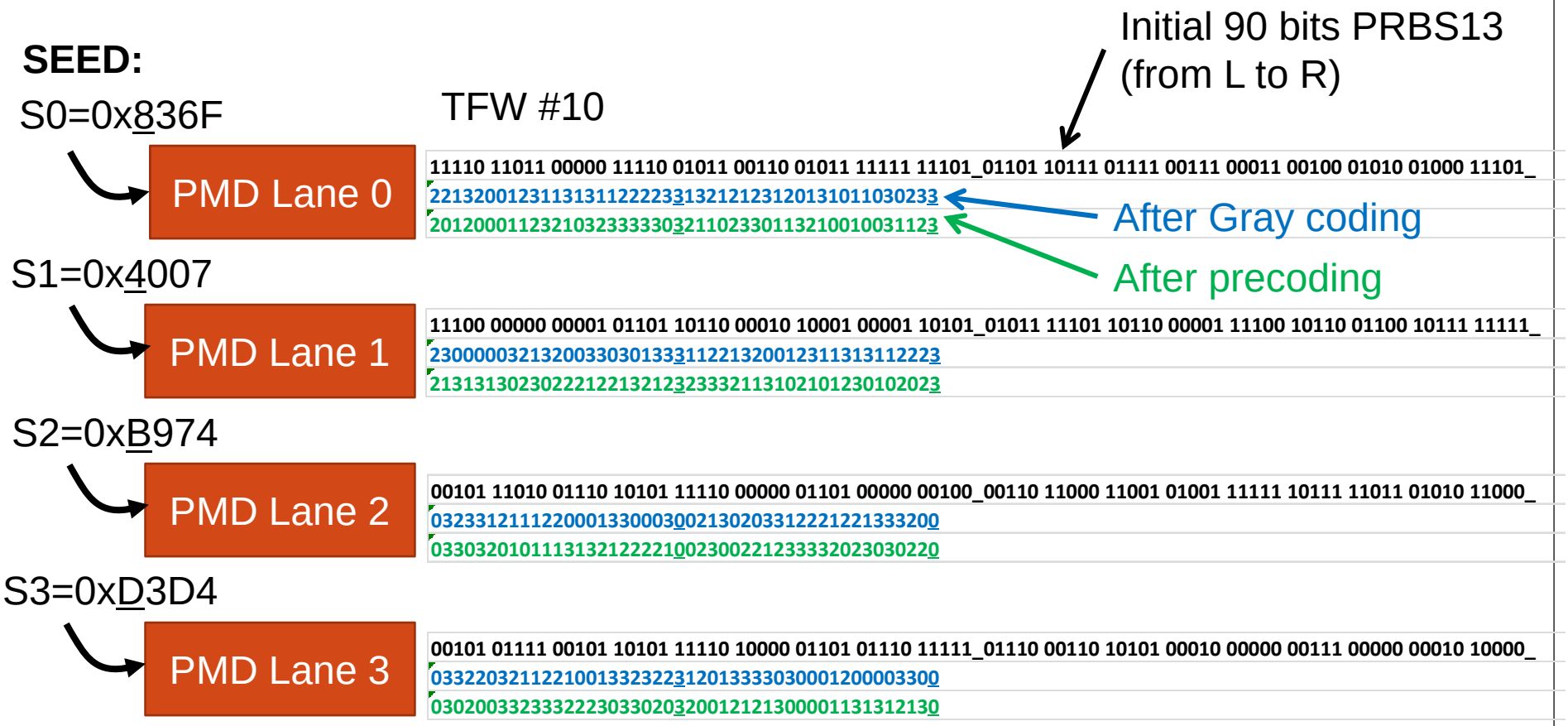
Training Pattern Seeds

- The *initial outputs* of the PRBS13 generator, right after the control channel transmission, shall be different for each of the PMD lanes, as follows (LSB transmitted first)
  - PMD lane 0: 0x836F
  - PMD lane 1: 0x4007
  - PMD lane 2: 0xB974
  - PMD lane 3: 0xD3D4

Note: for the underlined characters, only the LSB is part of the seed; the 3 MSBs are calculated by the polynomial

- Note: in this implementation, to generate these initial outputs, the *initial state* of the shift register should be set as follows (LSB in S0, MSB in S12)
  - PMD lane 0: 0x0355
  - PMD lane 1: 0x06FF
  - PMD lane 2: 0x0E16
  - PMD lane 3: 0x1C87

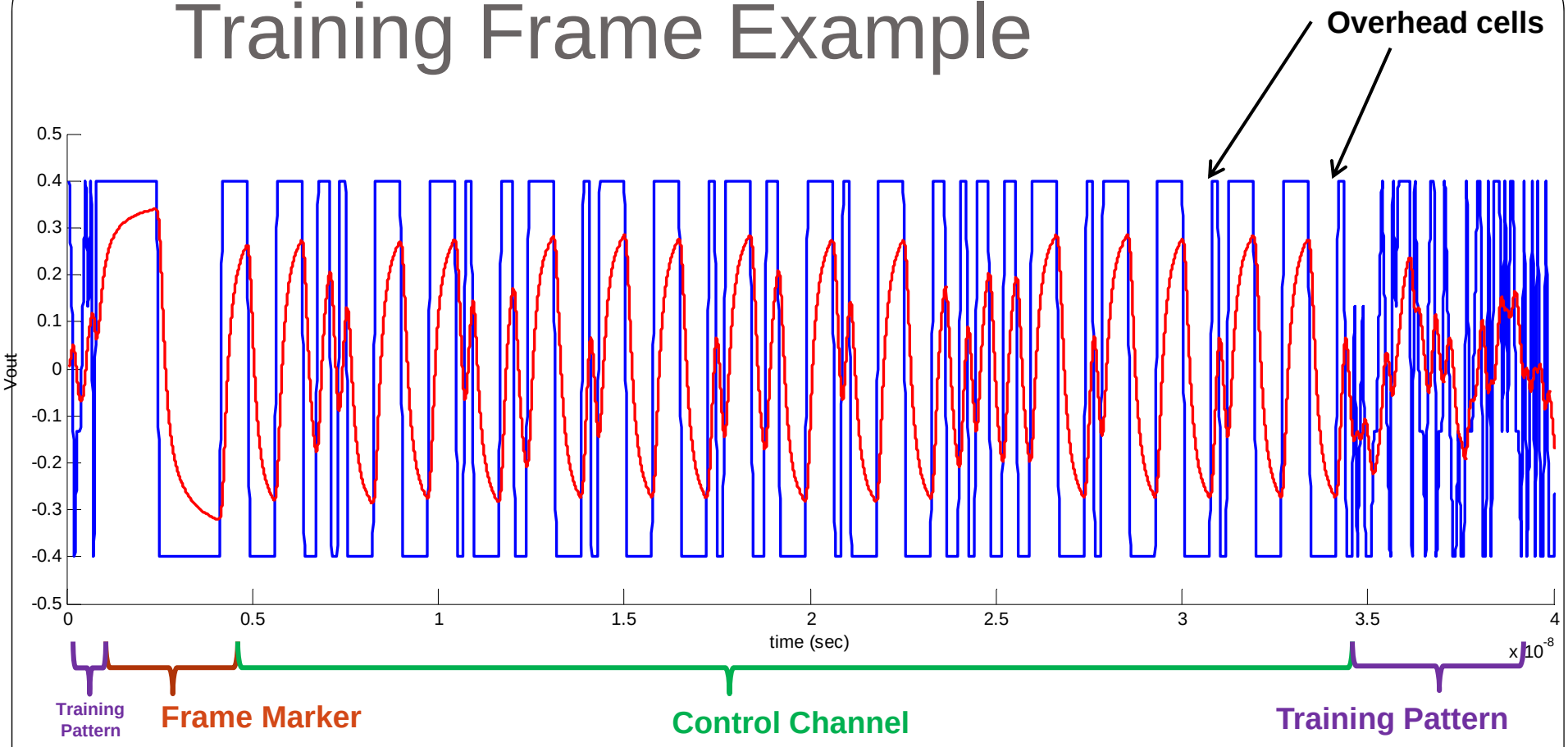
Training Pattern Seed Mapping



Notes:

- For the underlined characters, only the LSB is part of the seed; the 3 MSBs are calculated by the LFSR equation
- Precoder starts from state 0

Training Frame Example



- Blue = no channel
- Red = after IL = ~30dB @ 7GHz channel (without TXFFE equalization)
- Training pattern shown is incomplete. Figure is zoomed to frame marker and control channel

Channel used "TEC_Whisper42p8in_Nelco6_THRU_C8C9" from
http://www.ieee802.org/3/100GCU/public/ChannelData/TEC_11_0428/shanbhag_03_0411.pdf

Training Frame Time Breakdown

- 1 Training frame contains 384 TB46
  - 8832 PAM4 symbols or 649.7 ns;
  - 95% of the duration is rich-content signal
- For reference, 1 training frame for 10GBASE-KR is ~425 nsec
- Link_fail_inhibit_timer (KR) = 500-510 msec
  - ~770k frames in 100GBASE-KP4
  - ~1.2E6 frames in 10GBASE-KR

Field	# of TB46	# of TFWs	TB46 start index	TB46 end index	TFW start index	TFW end index
Marker	2	1	0	1	0	0
coefficient request	8	4	2	9	1	4
status update	10	5	10	19	5	9
PRBS13+PRBS13i, truncated to 16380 bits	364	182	20	383	10	191

PMA frame length [TB46]	TF length [TB46]	GCD	Possible offsets	PAO width	PRBS length [TB46]	PRBS % of TF	TF length(ns)
696	384	24	29	5	364	94.8%	649.7

Conclusion

- 100GBASE-KP4 Training frame in this presentation has the following qualities:
  - Reuses most of the existing 10GBASE-KR PMD training mechanism
  - Encompasses the PMA and PMD architecture unique to 100GBASE-KP4
  - Provides PAM2 for ease of alignment and PAM4 signaling for receiver calibration
  - Supplies DC-balanced, lane-specific seeded training pattern
  - Offers parity check for coef update and status report fields
  - Enables fast and efficient transition to data mode
  - Facilitates EEE signaling, if required