512B/513B Transcoding and FEC for 100 Gb/s Backplane and Copper Links

IEEE 802.3 100 Gb/s Backplane and Copper Cable Study Group

Chicago, September 13-14, 2011

Roy Cideciyan - IBM

Motivation for 512B/513B transcoding

- Transcoding (TC) is performed prior to forward error correction (FEC). Transcoding compresses data and therefore reduces total overhead.
- 100 Gb/s backplane and copper cable task force is considering transcoding from 64B/66B encoded data into 64B/65B encoded data by dropping a header bit
- 64B/65B transcoding has

$$(65/64) / (513/512) - 1 = 1.36\%$$

  higher line rate than 512B/513B transcoding with TC Rate = 512/513
- Main motivation for 512B/513B transcoding in conjunction with (N,K) Reed-Solomon (RS) coding with m-bit symbols is **lower overclocking than 64B/65B transcoding**

$$\text{Overclocking} = (64/66) / (\text{TC Rate} * K/N) - 1$$

- 512B/513B transcoding results in **reduced channel loss**
- **Lower power consumption**

64B/66B block formats in 100GBASE-R

- 64B/66B coding used in 100GBASE-R (IEEE 802.3ba-2010, Clause 82)
  - 1 type of data block with 2-bit header 01
  - **11** types of control blocks (CB) with 2-bit header 10 where the **8-bit block type field** indicates the type of control block format

	Input Data	Sync	Block Payload							
	Bit Position: / Data Block Format:	0 1	2							65
	$D_0 D_1 D_2 D_3 D_4 D_5 D_6 D_7$	01	D_0	D_1	D_2	D_3	D_4	D_5	D_6	D_7
	Control Block Formats:		Block Type Field							
1	$C_0 C_1 C_2 C_3 C_4 C_5 C_6 C_7$	10	0x1E	C_0	C_1	C_2	C_3	C_4	C_5	C_6	C_7
2	$S_0 D_1 D_2 D_3 D_4 D_5 D_6 D_7$	10	0x78	D_1	D_2	D_3	D_4	D_5	D_6	D_7
3	$O_0 D_1 D_2 D_3 Z_4 Z_5 Z_6 Z_7$	10	0x4B	D_1	D_2	D_3	O_0	0x000_0000		
4	$T_0 C_1 C_2 C_3 C_4 C_5 C_6 C_7$	10	0x87		C_1	C_2	C_3	C_4	C_5	C_6	C_7
5	$D_0 T_1 C_2 C_3 C_4 C_5 C_6 C_7$	10	0x99	D_0		C_2	C_3	C_4	C_5	C_6	C_7
6	$D_0 D_1 T_2 C_3 C_4 C_5 C_6 C_7$	10	0xAA	D_0	D_1		C_3	C_4	C_5	C_6	C_7
7	$D_0 D_1 D_2 T_3 C_4 C_5 C_6 C_7$	10	0xB4	D_0	D_1	D_2		C_4	C_5	C_6	C_7
8	$D_0 D_1 D_2 D_3 T_4 C_5 C_6 C_7$	10	0xCC	D_0	D_1	D_2	D_3		C_5	C_6	C_7
9	$D_0 D_1 D_2 D_3 D_4 T_5 C_6 C_7$	10	0xD2	D_0	D_1	D_2	D_3	D_4		C_6	C_7
10	$D_0 D_1 D_2 D_3 D_4 D_5 T_6 C_7$	10	0xE1	D_0	D_1	D_2	D_3	D_4	D_5		C_7
11	$D_0 D_1 D_2 D_3 D_4 D_5 D_6 T_7$	10	0xFF	D_0	D_1	D_2	D_3	D_4	D_5	D_6

Figure 82–5—64B/66B block formats

Alignment markers in 100GBASE-R

- 66-bit alignment markers are inserted into n=20 PCS virtual lanes
- In each virtual lane there are 2^{14}-1 = 16383 66-bit blocks between two alignment markers
- For 100 Gb/s transmission over 4 physical lanes, each physical lane contains 5 virtual lanes and therefore **5** types of alignment markers (AM).

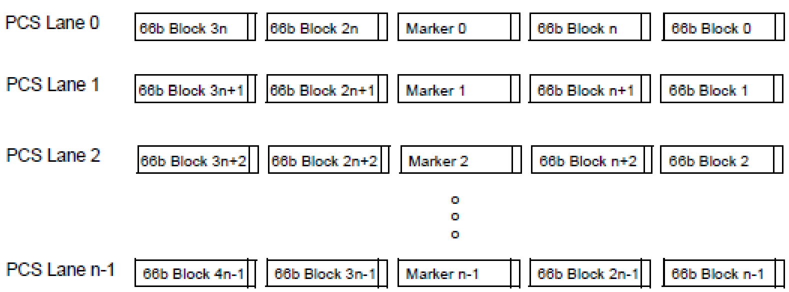


Figure 82–7—Alignment marker insertion

Alignment markers are special types of control blocks

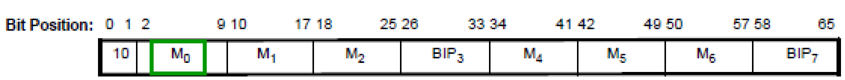


Figure 82–9—Alignment marker format

Table 82–2—100GBASE-R Alignment marker encodings

PCS lane number	Encoding[a] {M_0, M_1, M_2, BIP_3, M_4, M_5, M_6, BIP_7}	PCS lane number	Encoding[a] {M_0, M_1, M_2, BIP_3, M_4, M_5, M_6, BIP_7}
0	0xC1, 0x68, 0x21, BIP_3, 0x3E, 0x97, 0xDE, BIP_7	10	0xFD, 0x6C, 0x99, BIP_3, 0x02, 0x93, 0x66, BIP_7
1	0x9D, 0x71, 0x8E, BIP_3, 0x62, 0x8E, 0x71, BIP_7	11	0xB9, 0x91, 0x55, BIP_3, 0x46, 0x6E, 0xAA, BIP_7
2	0x59, 0x4B, 0xE8, BIP_3, 0xA6, 0xB4, 0x17, BIP_7	12	0x5C, 0x B9, 0xB2, BIP_3, 0xA3, 0x46, 0x4D, BIP_7
3	0x4D, 0x95, 0x7B, BIP_3, 0xB2, 0x6A, 0x84, BIP_7	13	0x1A, 0xF8, 0xBD, BIP_3, 0xE5, 0x07, 0x42, BIP_7
4	0xF5, 0x07, 0x09, BIP_3, 0x0A, 0xF8, 0xF6, BIP_7	14	0x83, 0xC7, 0xCA, BIP_3, 0x7C, 0x38, 0x35, BIP_7
5	0xDD, 0x14, 0xC2, BIP_3, 0x22, 0xEB, 0x3D, BIP_7	15	0x35, 0x36, 0xCD, BIP_3, 0xCA, 0xC9, 0x32, BIP_7
6	0x9A, 0x4A, 0x26, BIP_3, 0x65, 0xB5, 0xD9, BIP_7	16	0xC4, 0x31, 0x4C, BIP_3, 0x3B, 0xCE, 0xB3, BIP_7
7	0x7B, 0x45, 0x66, BIP_3, 0x84, 0xBA, 0x99, BIP_7	17	0xAD, 0xD6, 0xB7, BIP_3, 0x52, 0x29, 0x48, BIP_7
8	0xA0, 0x24, 0x76, BIP_3, 0x5F, 0xDB, 0x89, BIP_7	18	0x5F, 0x66, 0x2A, BIP_3, 0xA0, 0x99, 0xD5, BIP_7
9	0x68, 0xC9, 0xFB, BIP_3, 0x97, 0x36, 0x04, BIP_7	19	0xC0, 0xF0, 0xE5, BIP_3, 0x3F, 0x0F, 0x1A, BIP_7

[a]Each octet is transmitted LSB to MSB.

In each physical lane carrying data at 25 Gb/s there are 11 CB + **5** AM = 16 types of 66-bit control blocks

512B/513B transcoding latency

- Transcoding maps eight 66-bit blocks with a payload of 512 bits into one 513-bit block

- 512B/513B transcoding can be done
  - across all four physical lanes: low transcoding latency of 5.1 ns
  - across a single physical lane: medium transcoding latency of 20.4 ns
  - across a single virtual lane: high transcoding latency of 102 ns (unacceptable)

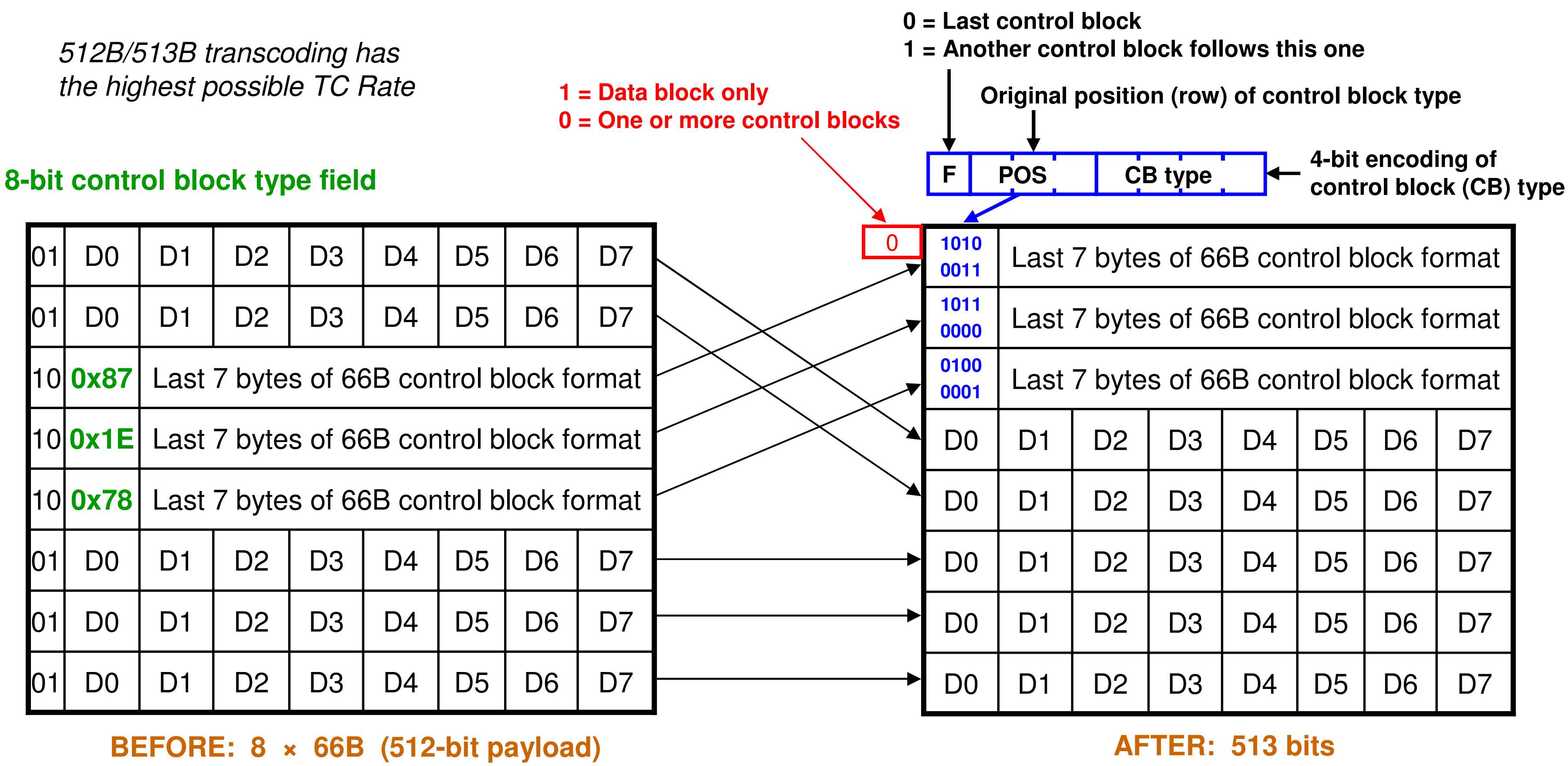
- Negligible latency for inverse 512B/513B transcoding (TC) at the receiver as inverse TC can be combined with FEC decoding. We assume that there are an integer number of 512B/513B transcoded blocks per FEC block.

Total transcoding latency: 5.1 ns for TC across all 4 physical lanes
20.4 ns for TC across a single physical lane

512B/513B transcoding

- A 512B/513B transcoding scheme was proposed for 40 Gb/s Ethernet in trowbridge_01_0707.pdf
- We can use 512B/513B TC for 100GBASE-R if there are not more than 16 types of control blocks
- All control blocks in a 513-bit transcoded block are at the top after reshuffling

512B/513B transcoding has the highest possible TC Rate



4-bit encoding of control block type

- For **transcoding across a single physical lane** there are 11 + **5** = 16 control block types per physical lane which can be encoded by 4 bits. In the table below, we assume that the five alignment markers in a physical lane are CB types 0xC1, 0xF5, 0xA0, 0x5C and 0xC4 from PCS virtual lanes 0, 4, 8, 12 and 16. Each physical lane has a separate 4-bit encoding table. However, all four 4-bit encoding tables agree in their first 11 entries.

CB type	0x1E	0x78	0x4B	0x87	0x99	0xAA	0xB4	0xCC	0xD2	0xE1	0xFF	0xC1	0xF5	0xA0	0x5C	0xC4
4-bit code	0000	0001	0010	0011	0100	0101	0110	0111	1000	1001	1010	1011	1100	1101	1110	1111

11 control blocks

5 alignment markers

- For **transcoding across all four physical lanes (PL)** a 4-bit encoding table can be constructed. **In each PL lane there is 1 primary alignment marker p-AM** (e.g. Marker 0 in PL0) and 4 secondary alignment markers (e.g. Markers 4, 8, 12, 16 in PL0) following p-AM. Two header bits of primary alignment markers are dropped. We can **map** the first byte of all 20 – 4 = **16 secondary alignment markers s-AM into** a byte such that it indicates **the last unused control block type**. Therefore, we have a total of 11 CB + **4** * **1 p-AM** + **1 s-AM** = 16 control block types and we can use 512B/513B transcoding.

Total latency of transcoding and FEC

- Total latency of transcoding and FEC has two contributors
  - <u>At the transmitter</u>

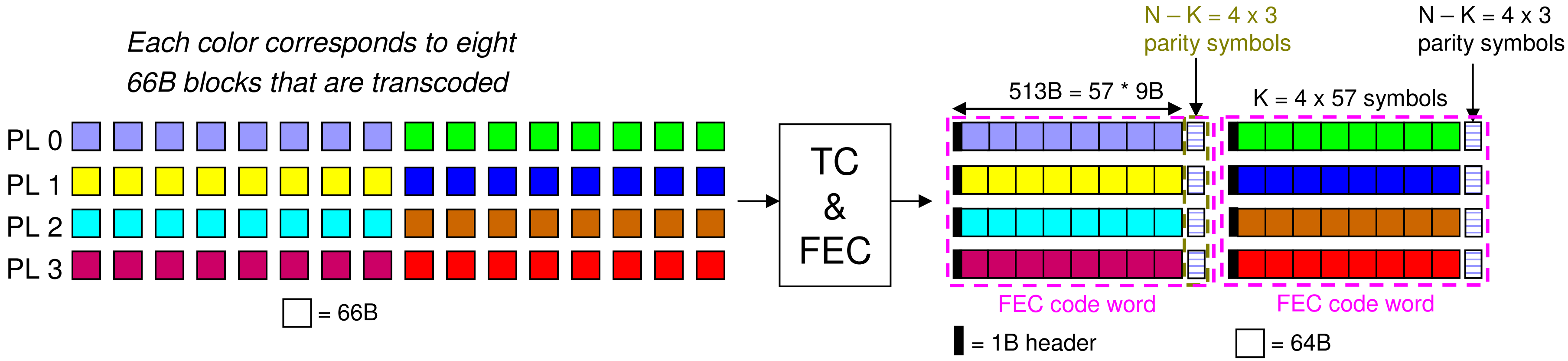
    Encoding latency = Transcoding latency + FEC encoding latency
  - <u>At the receiver</u>

    Decoding latency = Inverse transcoding latency + FEC decoding latency
- We assume FEC code is (N,K) block code over $GF(2^m)$ and all K m-bit symbols are available in a buffer. Minimum achievable latency for one-shot FEC encoding is 1 multiplication over $GF(2^m)$ + 4 * four-input XOR gate delay for $K < 258$ and is therefore negligible.
- FEC across all four physical lanes as proposed in wang_01_0511

$$\text{Total latency} = \text{Transcoding} + \underbrace{\text{FEC encoding}}_{\text{Negligible}} + \underbrace{\text{FEC decoding}}_{\substack{\text{2x to 3x} \\ \text{FEC block latency}}} + \underbrace{\text{Inverse Transcoding}}_{\text{Negligible}}$$

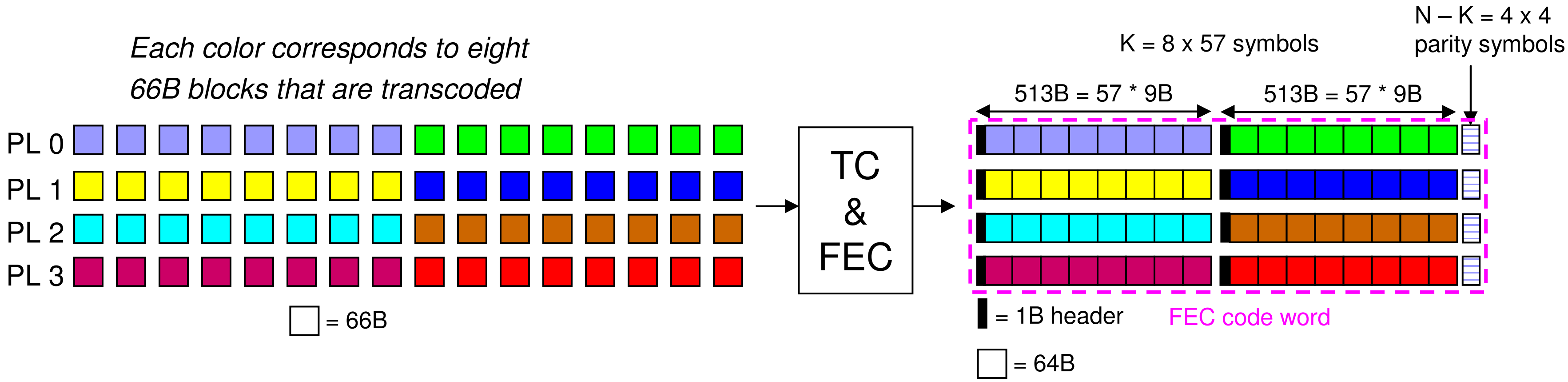
Total latency ~ Transcoding latency + 2x to 3x FEC block latency

Transcoding across a physical lane: Part 1



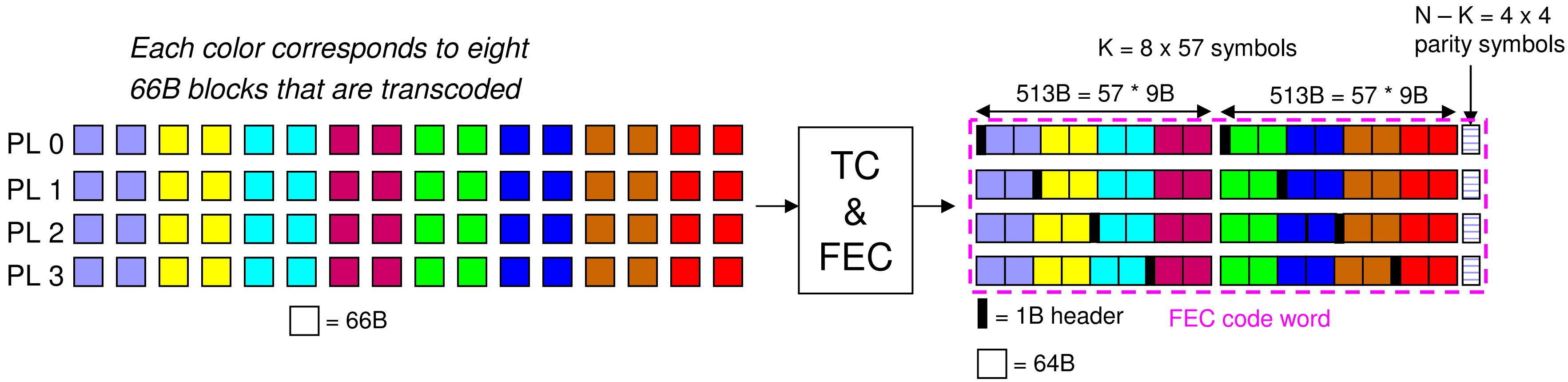
- RS(240,228) code with t=6 error correction capability and m=9 symbol size
- Properties
  - 10240 FEC code words within 1 alignment period satisfying alignment proposal in gustlin_02a_0511
  - 4 transcoded blocks within 1 FEC code word
  - 2.27% overclocking and 156.25 MHz clock multiplier is 168 ¾
  - total latency of 61 to 82 ns
  - 513-bit FEC striping
- Other FEC options
  - RS(244,228) FEC code with t=8, m=9
  - RS(248,228) FEC code with t=10, m=9

Transcoding across a physical lane: Part 2



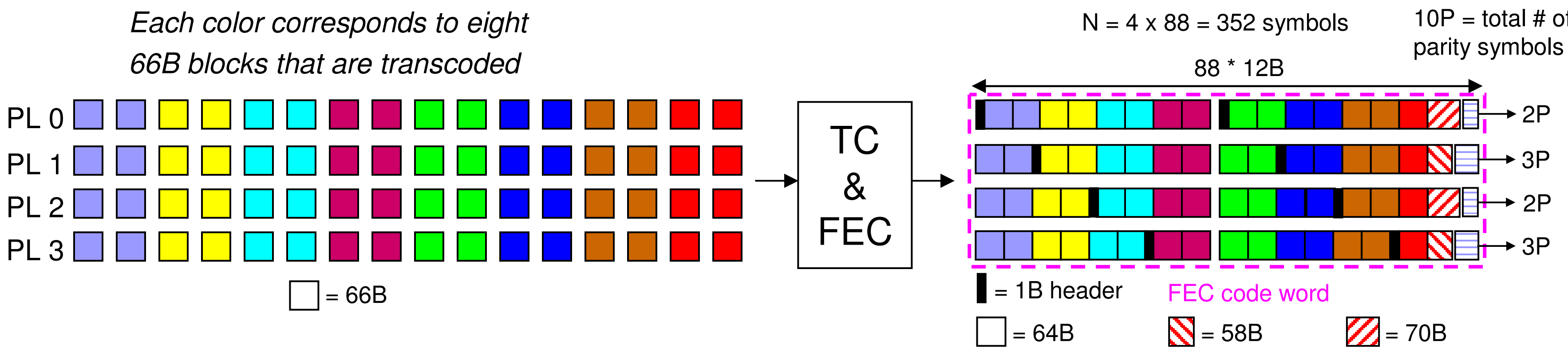
- RS(472,456) code with t=8 error correction capability and m=9 symbol size
- Properties
  - 5120 FEC code words within 1 alignment period satisfying alignment criterion
  - 8 transcoded blocks within 1 FEC code word
  - 0.6% overclocking and 156.25 MHz clock multiplier is 165 15/16
  - total latency of 102 to 143 ns
  - 513-bit FEC striping
- Other FEC options
  - RS(476,456) FEC code with t=10, m=9

Transcoding across all physical lanes: Part 1



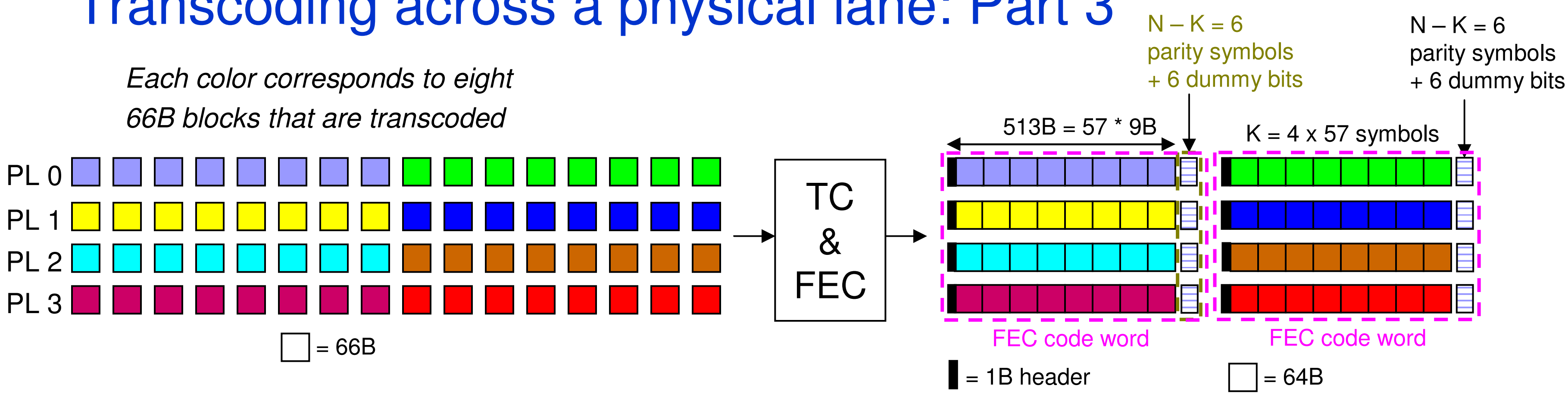
- RS(472,456) code with t=8 error correction capability and m=9 symbol size
- Properties
  - 5120 FEC code words within 1 alignment period satisfying alignment criterion
  - 8 transcoded blocks within 1 FEC code word
  - 0.6% overclocking and 156.25 MHz clock multiplier is 165 15/16
  - total latency of 87 to 128 ns
  - header bit of 513B transcoded block rotated across all 4 physical lanes

Transcoding across all physical lanes: Part 2



- RS(352,342) code with t=5 error correction capability and m=12 symbol size
- Properties
  - 5120 FEC code words within 1 alignment period satisfying alignment criterion
  - 8 transcoded blocks within 1 FEC code word
  - 0% overclocking and 156.25 MHz clock multiplier is 165
  - total latency of 87 to 128 ns
  - header bit of 513B transcoded block rotated across all 4 physical lanes
  - 2*t = 10 parity symbols not divisible among 4 physical lanes
  - In this example PL0 and PL2 have 86 information symbols + 2 parity symbols whereas PL1 and PL3 have 85 information symbols + 3 parity symbols
  - 13-bit error burst at DFE output always corrupts at most two symbols in one FEC code word

Transcoding across a physical lane: Part 3



- RS(234,228) code with t=3 error correction capability and m=9 symbol size
- Properties
  - 10240 FEC code words within 1 alignment period satisfying alignment criterion
  - 4 transcoded blocks within 1 FEC code word
  - 0% overclocking and 156.25 MHz clock multiplier is 165
  - total latency of 61 to 82 ns
  - 513-bit FEC data striping, 15-bit FEC parity striping
  - 2*t = 6 parity symbols not divisible among 4 physical lanes
  - end of FEC code word contains 2 parity symbols that are split into 4 physical lanes together with 6 dummy bits
    - e.g., PL0 and PL2 can have one 9-bit parity symbol + 4-bit split parity + 2 dummy bits = 15 bits
    - e.g., PL1 and PL3 can have one 9-bit parity symbol + 5-bit split parity + 1 dummy bit = 15 bits
  - 10-bit error burst at DFE output always corrupts at most 2 symbols in one FEC code word

Code Comparison

(1): TC across all 4 physical lanes
(2): TC across a single physical lane
(3): 10GBASE-KR
(4): bhoja_01_0911
(5): proposed by John Ewen, IBM
10 GHz / 64 = 156.25 MHz

TC	FEC	K	N	t	m	Line Rate [Gb/s]	Latency [ns]	Overclocking [%]	Multiplier of 156.25 MHz
512B/513B	RS	228	240	6	9	26.36718	46 – 67 (1) 61 – 82 (2)	2.3	168 3/4
512B/513B	RS	228	244	8	9	26.80664	46 – 67 (1) 61 – 82 (2)	4	171 9/16
512B/513B	RS	228	248	10	9	27.24609	46 – 67 (1) 61 – 82 (2)	5.7	174 3/8
512B/513B	RS	456	472	8	9	25.92773	87 – 128 (1) 102 – 143 (2)	0.6	165 15/16
512B/513B	RS	456	476	10	9	26.14746	87 – 128 (1) 102 – 143 (2)	1.4	167 11/32
512B/513B	RS	228	234	3	9	25.78125	46 – 67 (1) 61 – 82 (2)	0	165
512B/513B	RS	342	352	5	12	25.78125	87 – 128 (1) 102 – 143 (2)	0	165
64B/65B	Fire (3)	2080	2112	11B burst	1	25.78125	20	0	165
64B/65B	RS (4)	208	224	8	10	27.34375	41 – 61	6.1	175
64B/65B	RS (4)	416	448	16	10	27.34375	82 – 123	6.1	175
64B/65B	RS (4)	104	112	4	10	27.34375	20 – 31	6.1	175
64B/65B	RS (5)	260	272	6	10	26.56250	51 – 77	3	170
64B/65B	RS (5)	260	280	10	10	27.34375	51 – 77	6.1	175

Summary

- 512B/513B transcoding reduces line rate, overclocking, channel loss and power consumption when compared to 64B/65B transcoding

- Several FEC options for 512B/513B transcoding were proposed and compared to each other and to other FEC proposals in terms of total latency, line rate, overclocking, capability to correct errors and multiplier of 156.25 MHz clock
  - RS(352,342) code with m=12 correcting t=5 symbols has 0% overclocking where multiplier of 156.25 MHz clock is 165. Transcoding can be done across all four physical lanes or across a single physical lane.
  - Low-latency RS(234,228) code with m=9 correcting t=3 symbols has 0% overclocking where multiplier of 156.25 MHz clock is 165. Transcoding can be done across all four physical lanes or across a single physical lane.

- Performance evaluation/comparison of proposed coding schemes in terms of bit error rate at the output of FEC decoder for various types of channels will guide the selection of a suitable TC/FEC scheme for 100 Gb/s backplane and copper links