OTN Compatibility for 40 Gb Ethernet

Stephen J. Trowbridge
Alcatel-Lucent

Supporters

- Gary Nicholl (Cisco)
- Pete Anslow (Nortel)
- Mark Gustlin (Cisco)
- Med Belhadj (Cortina)
- Frank Chang (Vitesse)
- Martin Carroll (Verizon)
- Ralf-Peter Braun (Deutsche Telekom)
- George Young (AT&T)
- Thomas Fischer (Nokia Siemens Networks)

IEEE 802.3 HSSG – OTN Compatibility Objective

- OTN support is written as a speed-independent objective
  - OTN compatibility is important for both 100 GbE and 40 GbE
  - For 100 GbE, OTN compatibility is a given, since we have a “blank sheet of paper” for 100G transport and ITU-T can design OTU4/ODU4 so that 100 GbE will fit
  - For 40 GbE, need to ensure that a mapping exists into the payload area of a standard OPU3 (40.149716 Gb/s with worst case -20ppm clock)

How do we get 40 GbE to fit standard ODU3?

Agreement needed between IEEE P802.3ba and ITU-T SG15

- Option 1 – Reduce MAC rate to ~38.9 Gbit/s or less and use 64B/66B coding
- Option 2 – Use a 40G MAC and a more efficient linecode than 64B/66B (e.g., proposed 512B/513B coding)
- Option 3 – The LAN interface uses a 40G MAC and 64B/66B, but the mapping into OTN transcodes into a more efficient linecode (e.g., proposed 512B/513B coding)

Advantages & Disadvantages of different approaches

Option	Advantages	Disadvantages
Option 1 – 38.9G MAC rate	• Simple and straightforward • Complete bit transparency • Any proprietary extension that fits in the 66-bit codewords can be transported	• Not a “round” number • No 4x10G wire-speed bridge without packet loss • Might not be able to “underclock” 10G components for reuse
Option 2 – more efficient Ethernet line code	• Complete bit transparency • Any proprietary extension that fits in the 513-bit codewords can be transported	• No reuse of 10G PCS • Reduced Hamming distance between control block codes
Option 3 - Transcode LAN PCS to more efficient code in OTN mapper	• Minimum constraints on LAN implementation • Maximum reuse of 10G components for LAN • Reduced Hamming distance for control block codes protected by FEC	• Codeword rather than complete bit transparency • Risk of proprietary extensions to 64B/66B coding that break transcoding • Errors not always propagated transparently

Option 3 considered most likely

How can the disadvantages be addressed?

- Codeword transparent mapping reproduces the original bitstream PROVIDED:
  - There are no proprietary extensions to the 64B/66B control block set
  - There are no errors

How can proprietary Extensions to the 64B/66B control block set be prevented?

- VERY strong language already exists in the standard, for example:

49.2.4.4 Control codes

The same set of control characters are supported by the XGMII and the 10GBASE-R PCS. The representations of the control characters are the control codes. XGMII encodes a control character into an octet (an eight bit value). The 10GBASE-R PCS encodes the start and terminate control characters implicitly by the block type field. The 10GBASE-R PCS encodes the ordered_set control codes using a combination of the block type field and a 4-bit O code for each ordered_set. The 10GBASE-R PCS encodes each of the other control characters into a 7-bit C code).

The control characters and their mappings to 10GBASE-R control codes and XGMII control codes are specified in Table 49–1. All XGMII and 10GBASE-R control code values that do not appear in the table shall not be transmitted and shall be treated as an error if received.

- The text associated with the block structure is too weak:

49.2.4.3 Block structure

• • •

All unused values of block type field[7] are reserved.

- Better would be: "Any block type value not appearing in figure 49-7 shall not be transmitted and shall be considered an error if received"
- Needs to be specified for 40 GbE PCS. 10 GbE PCS doesn't have to change

What types of errors cannot be mapped with 64B/66B to 512B/513 transcoding

- Any 66B codeword with SYNC= “00” or SYNC=“11”
- Any 66B codeword with SYNC=“10” and a block type field that is NOT one of the 15 shown in IEEE 802.3(2005) Figure 49-7
- Since these are errors, they can be mapped (before transcoding) to:

SYNC	D_0 Block type	D_1	D_2	D_3	D_4	D_5	D_6	D_7
10	0x1E	/E/	/E/	/E/	/E/	/E/	/E/	/E/

- Any legal, error free 66B codeword can be reproduced bit for bit at the egress
- Illegal or errored codewords that cannot be transcoded can still be propagated as an error, preserving the accuracy of any end-to-end performance monitoring information
- Any error in the SYNC bits or control block type introduced in the OTN network (extremely unlikely if FEC used) will be converted to an error 66B codeword on egress.

Review of 512B/513B coding – more economical coding with <=0.36425% overhead re-using concept from 64B/65B GFP-T coding of 1GbE

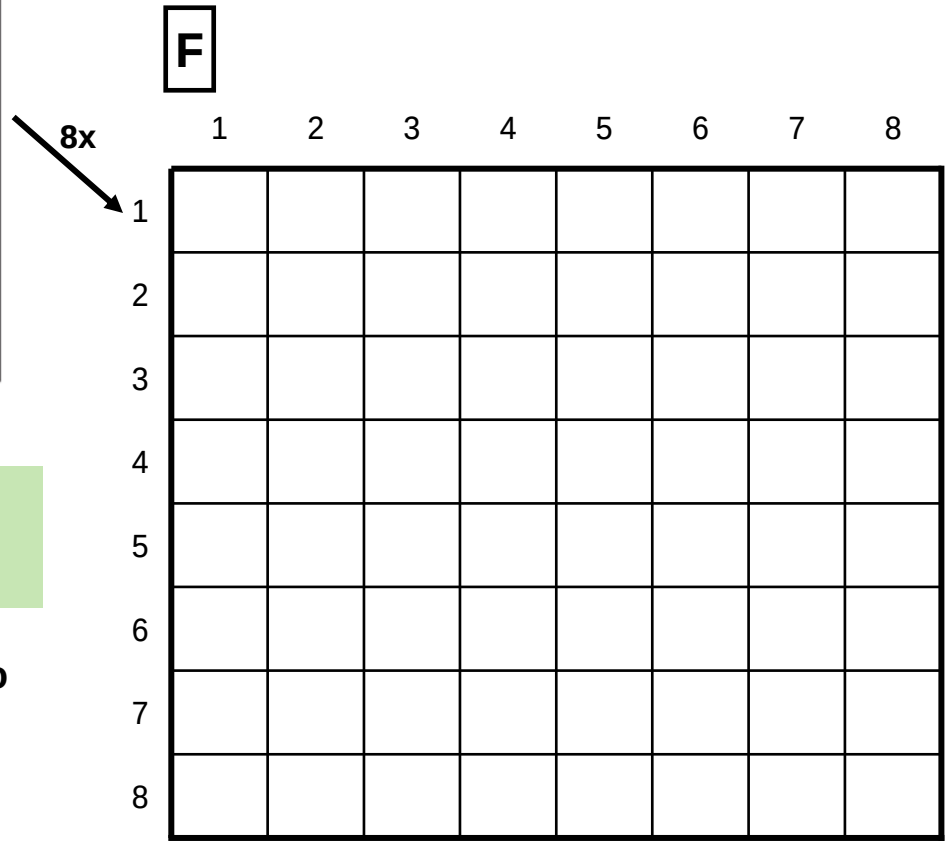
Input Data	Sync	Block Payload							
Bit Position: / Data Block Format:	0 1	2							65
$D_0\,D_1\,D_2\,D_3/D_4\,D_5\,D_6\,D_7$	01	D_0	D_1	D_2	D_3	D_4	D_5	D_6	D_7
Control Block Formats:		Block Type Field							
$C_0\,C_1\,C_2\,C_3/C_4\,C_5\,C_6\,C_7$	10	0x1e	C_0	C_1	C_2	C_3	C_4	C_5	C_6	C_7
$C_0\,C_1\,C_2\,C_3/O_4\,D_5\,D_6\,D_7$	10	0x2d	C_0	C_1	C_2	C_3	O_4	D_5	D_6	D_7
$C_0\,C_1\,C_2\,C_3/S_4\,D_5\,D_6\,D_7$	10	0x33	C_0	C_1	C_2	C_3		D_5	D_6	D_7
$O_0\,D_1\,D_2\,D_3/S_4\,D_5\,D_6\,D_7$	10	0x66	D_1	D_2	D_3	O_0		D_5	D_6	D_7
$O_0\,D_1\,D_2\,D_3/O_4\,D_5\,D_6\,D_7$	10	0x55	D_1	D_2	D_3	O_0	O_4	D_5	D_6	D_7
$S_0\,D_1\,D_2\,D_3/D_4\,D_5\,D_6\,D_7$	10	0x78	D_1	D_2	D_3	D_4		D_5	D_6	D_7
$O_0\,D_1\,D_2\,D_3/C_4\,C_5\,C_6\,C_7$	10	0x4b	D_1	D_2	D_3	O_0	C_4	C_5	C_6	C_7
$T_0\,C_1\,C_2\,C_3/C_4\,C_5\,C_6\,C_7$	10	0x87		C_1	C_2	C_3	C_4	C_5	C_6	C_7
$D_0\,T_1\,C_2\,C_3/C_4\,C_5\,C_6\,C_7$	10	0x99	D_0		C_2	C_3	C_4	C_5	C_6	C_7
$D_0\,D_1\,T_2\,C_3/C_4\,C_5\,C_6\,C_7$	10	0xaa	D_0	D_1		C_3	C_4	C_5	C_6	C_7
$D_0\,D_1\,D_2\,T_3/C_4\,C_5\,C_6\,C_7$	10	0xb4	D_0	D_1	D_2		C_4	C_5	C_6	C_7
$D_0\,D_1\,D_2\,D_3/T_4\,C_5\,C_6\,C_7$	10	0xcc	D_0	D_1	D_2	D_3		C_5	C_6	C_7
$D_0\,D_1\,D_2\,D_3/D_4\,T_5\,C_6\,C_7$	10	0xd2	D_0	D_1	D_2	D_3	D_4		C_6	C_7
$D_0\,D_1\,D_2\,D_3/D_4\,D_5\,T_6\,C_7$	10	0xe1	D_0	D_1	D_2	D_3	D_4	D_5		C_7
$D_0\,D_1\,D_2\,D_3/D_4\,D_5\,D_6\,T_7$	10	0xff	D_0	D_1	D_2	D_3	D_4	D_5	D_6	

Figure 49–7—64B/66B block formats

Current 64B/66B coding used in 10GBase-R (IEEE 802.3-2005, Clause 49)

- **66B control blocks (without sync bits) are sorted to the top of the 8-row block**
- **66B data blocks (without sync bits) are sorted to the bottom of the 8-row block**

8 x 8 octet frame with one "Flag" bit indicating the presence of control blocks

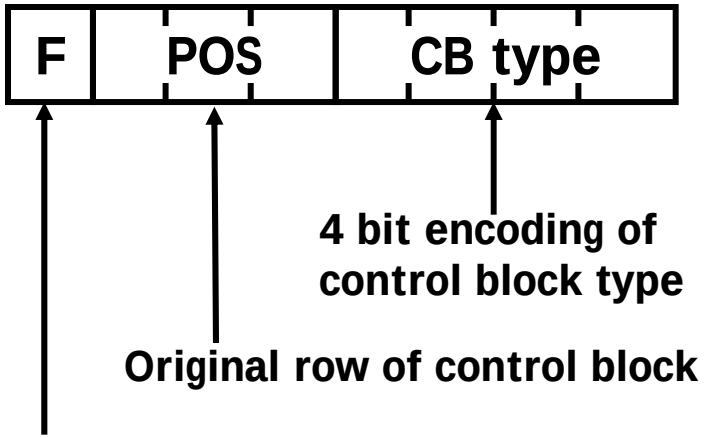


Alcatel·Lucent

4 bit encoding of 64B/66B control block type

Type	4 bit code
0x1e	0001
0x2d	0010
0x33	0011
0x66	0100
0x55	0101
0x78	0110
0x4b	0111
0x87	1000

Type	4 bit code
0x99	1001
0xaa	1010
0xb4	1011
0xcc	1100
0xd2	1101
0xe1	1110
0xff	1111



Encoding of control block types releases enough bits to allow control blocks to be sorted back to their original position at the egress

512B/513B coding Examples

0

	1	2	3	4	5	6	7	8
1	D	D	D	D	D	D	D	D
2	D	D	D	D	D	D	D	D
3	D	D	D	D	D	D	D	D
4	D	D	D	D	D	D	D	D
5	D	D	D	D	D	D	D	D
6	D	D	D	D	D	D	D	D
7	D	D	D	D	D	D	D	D
8	D	D	D	D	D	D	D	D

All Data Blocks

1

	1	2	3	4	5	6	7	8
1	0aaacccc	X	X	X	X	X	X	X
2	D	D	D	D	D	D	D	D
3	D	D	D	D	D	D	D	D
4	D	D	D	D	D	D	D	D
5	D	D	D	D	D	D	D	D
6	D	D	D	D	D	D	D	D
7	D	D	D	D	D	D	D	D
8	D	D	D	D	D	D	D	D

One Control, Seven Data Blocks

aaa, bbb, ddd = original row(s) of control blocks

cccc = 4 bit encodings of control block type(s)

X X X X X X X = per format of 64B/66B control block type(s)

1

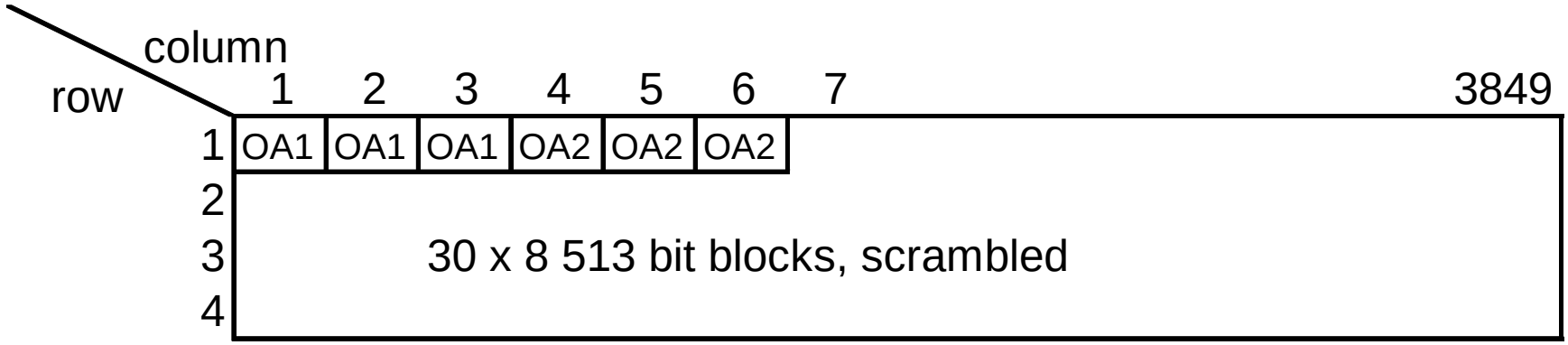
	1	2	3	4	5	6	7	8
1	1aaacccc	X	X	X	X	X	X	X
2	1bbbcccc	X	X	X	X	X	X	X
3	0dddcccc	X	X	X	X	X	X	X
4	D	D	D	D	D	D	D	D
5	D	D	D	D	D	D	D	D
6	D	D	D	D	D	D	D	D
7	D	D	D	D	D	D	D	D
8	D	D	D	D	D	D	D	D

Three Control, Five Data Blocks

Economical Linecode – Finishing touches

- The 512-bit/513-bit coding uses 0.1953125% of the allowable 0.36425% overhead
- Combine 8 (or 8n) 513-bit blocks into a 513 (or 513n) byte super-block to have an integral number of bytes
- Scrambling to ensure sufficient transitions and timing recovery (e.g., the $1 + x + x^3 + x^{12} + x^{16}$ OTN scrambler)
- Some sort of framing (at the super-super-block level) to recover the start of frame. Could use an OTN-like frame with 7 byte FAS out of a 4x4080 byte frame (an additional 0.0429% of overhead)

Possible Framing to find start of 513 bit blocks



OA1	1111 0110
OA2	0010 1000

OTN scrambler: $1 + x + x^3 + x^{12} + x^{16}$

Overhead added: 0.039%

Conclusions

- A promising compromise for mapping of 40 GbE into standard ODU3 is to transcode the 64B/66B PCS into 512B/513B
- The specification of 40 GbE PCS in 802.3 needs to be iron-clad to prevent the use of non-standard 66B codewords that cannot be transcoded. The standard needs to be extremely clear that these are not to be transmitted and will be treated as errors by receiving equipment.
- In concert with IEEE 802.3, ITU-T should amend/revise G.709 to include mapping 40 GbE into standard OPU3 using transcoding
- Assuming that an agreement is reached between IEEE and ITU-T around a transcoding approach, notes should appear in both the IEEE 802.3 and ITU-T G.709 standards with the 40 GbE equivalent of Figure 49-7 warning of the relationship between these standards and that the codeword set should not be expanded or changed without coordination between IEEE and ITU-T.