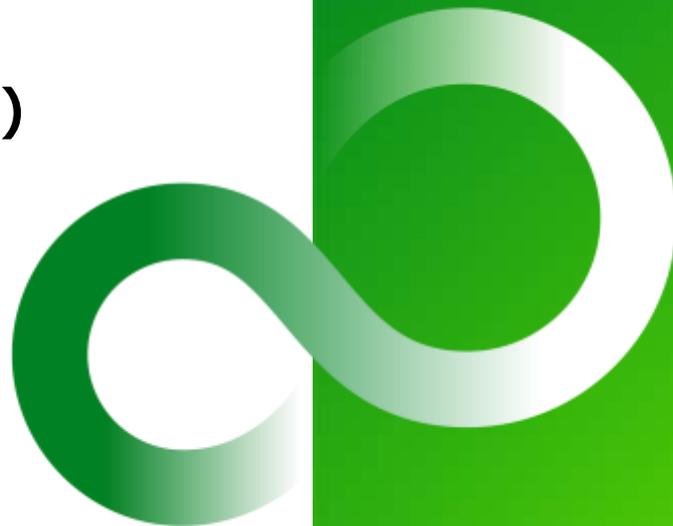


**MPLS Japan 2023 (Optical セッション)
B400G最新動向 ならびに
将来の光伝送に向けた課題と期待**

2023/10/26

栃尾 祐治 (富士通)



- 過去数年(2019, 2022年)、B400G動向としてお話ししたことの更新になります
 - https://mpls.jp/2019/presentations/MPLS_Japan_2019_tochio_web-1.pdf
 - https://mpls.jp/2022/presentations/mpls2022-tochio_web.pdf
 - 2019年では、各標準化団体の動向を中心にB100Gに関わる動向を振り返り、B400Gの課題について紹介
 - 2022年では、B400G, 特に 800G をめぐる動向について、課題を含め紹介
- 今回は、以下の2点について紹介
 - **800G**イーサネット(IEEE P802.3df)が大詰めを迎え、連動して進めてきたITU-T, OIFにおける800G関連の標準化、特に、**ITU-TにおけるB400G関連**の最新動向を中心に紹介
 - 将来の光伝送に向けた課題と期待 (B400Gに限らない)
 - **1.6T**に向けた状況と予測
 - **WDM(Media, L0)**における運用面の課題
 - 光伝送における**新領域への適用**における期待

● B400G光伝送(特に800G)標準化動向

● IEEE802.3

- P802.3df, P802.3df, P802.3cw

● OIF 800ZR/800LR

● ITU-T

- FlexO勧告大更新
- 800GbE収容とG.709(OTN)多重階梯
- (補足) fgOTN (fine grain OTN)

● 将来の光伝送に向けた課題と期待

- 1.6Tに向けた状況と予測 ↑
- WDM(Media, L0)における運用面の課題
- 光伝送における新領域への適用における期待
 - Data Center向けネットワーク

● まとめ

● お断り

- 基本的に標準化動向に基づく個人の見解が中心であり一部を除き、弊社の見解ではありません
- 標準化に関しては原則公開情報をもとに資料を紹介しますが、一部審議中の資料が含まれます。取り扱いにご注意ください(審議中資料は公開(web)には含めない予定)
- 限られた時間の発表のため、スキップするスライドがあるかもしれません(後日ご確認を)

● 参考: 弊社に関するフォトニクス関連最新情報

- ECOC 2023 – Fujitsu Blog <https://thecinict.com/ecoc-2023/>
 - Fujitsu All Photonic Network Vision
 - 1FINITY Ultra Optical System



B400G光伝送(特に800G)標準化動向

● 現在の400G光伝送の世界

- IEEE802.3
- ITU-T (Q11/15)
- OIF 400ZR
- OpenZR+
- *Open ROADM*

● B400G光伝送(特に800G)の動向

- IEEE802.3
- ITU-T work (FlexO & B400G検討)
- OIF 800ZR/LR
- *800G MSA etc*

IEEE

400G, 800G, 1.6T, …のB400Gの基本的な伝送規格は、**まず Ethernet (IEEE802.3) が起点になって検討**

ITU-T (OTN)

Ethernet (MAC or Code(64B/66B))を、**光ネットワークで転送することを目的**に、基本となる**デジタルインタフェース規格**(フレーム、マッピング、FEC、電気光IF)をWDM規格*も考慮しながら検討

*WDM規格はQ6/15で担当

OIF, OpenZR+など

IEEE, ITU-Tを念頭に、コンポーネント(DSP, フォームファクターなど)規格も考慮しながら光伝送システムとして検討

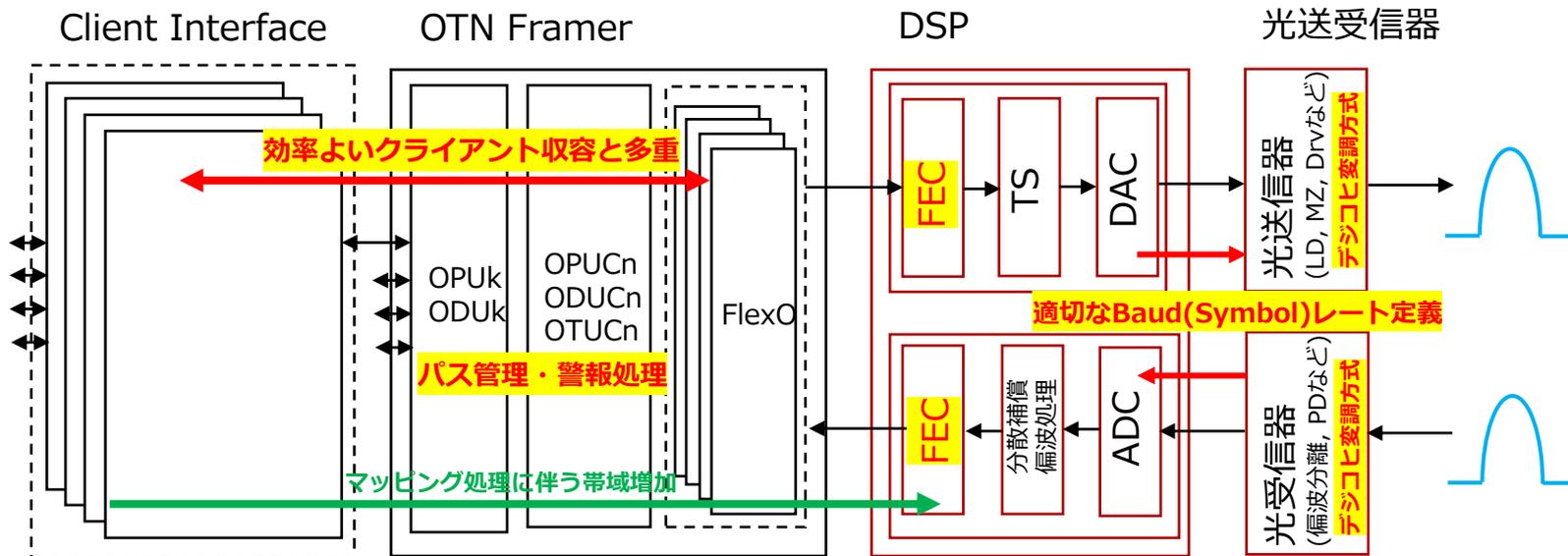
注: ここで紹介するものは、(光)伝送特性・フォーマットなどが中心であり実装観点については深く掘り下げていない点ご了承ください

B400Gを実現するための課題など

昨年(MPLS Japan) 資料

FUJITSU

- 以下のトレードオフとどう対処していくかが、B400G、当面は800G光伝送技術実現に向けた課題
 - 効率よいクライアント收容と多重: 従来のOTNマッピング規定に従うか、一部を割愛して帯域増加を抑えるか
 - パス管理・警報処理: 帯域増加を抑えOH処理を割愛すると、従来のOTN技術で培われたパス管理・警報処理を犠牲に…
 - 適切なBaudレート定義: 実際の光伝送速度を決める大事なファクター。800Gで120-130Gbaudあたりが予測されるが、高baudレートは、波長帯域幅増加(多重効率低下)にもつながるので、高baudレートを目指せばいい、ということでもない



● IEEE802.3

- IEEE802.3において、主要な400Gイーサネットの規格は標準化完了
 - IEEE802.3-2022として発行済み
- 現在継続中の(光伝送による) 400Gイーサネットは
 - P802.3cw (400GBASE-ZR) ならびに、
 - 800GE/1.6TEも視野にTask force化されたP802.3df, P802.3djのサブセット定義
- P802.3cw (400GBASE-ZR)
 - **Physical Layers and Management Parameters for 400 Gb/s Operation over DWDM (dense wavelength division multiplexing) systems**
 - ターゲット：80km
 - 2024/9承認予定
- 800G/1.6TイーサネットはP802.3df, djで作業中
 - 詳細次スライド

IEEE802.3 (P802.3df & P802.3dj)

- 2021/11: 800G/1.6T向けイーサネットとしてP802.3dfがスタート

Ethernet Rate	Assumed Signaling Rate	AUI	BP	Cu Cable	MMF 50m	MMF 100m	SMF 500m	SMF 2km	SMF 10km	SMF 40km
200 Gb/s	200 Gb/s	Over 1 lane		Over 1 pair			Over 1 Pair	Over 1 Pair		
400 Gb/s	200 Gb/s	Over 2 lanes		Over 2 pairs			Over 2 Pair			
800 Gb/s	100 Gb/s	Over 8 lanes	Over 8 lanes	Over 8 pairs						
	200 Gb/s	Over 4 lanes		Over 4 pairs			Over 4 pairs	1) Over 4 pairs 2) Over 4 λ's		
	TBD								Over single SMF in each direction	Over single SMF in each direction
1.6 Tb/s	100 Gb/s	Over 16 lanes								
	200 Gb/s	Over 8 lanes		Over 8 pairs			Over 8 pairs	Over 8 pairs		

https://www.ieee802.org/3/B400G/public/21_1028/B400G_overview_c_211028.pdf

- 後(2022/11)に、1.6Tならびに200G signaling rate (200G lane)については、完成の時期を考慮して、**P802.3dj**として分離独立。2026/3承認を目指す
 - **P802.3dj: 200Gb/s, 400 Gb/s, 800 Gb/s, and 1.6 Tb/s Ethernet Task Force**
- スコープを分割されたP802.3dfは、800G, 400Gを対象に2024/7承認を目指し活動中
 - ただし、**今年中に Standards Association ballot が発行**されるため、技術的には今年中に固まる見込み

● P802.3df (400G/800G関連)

- **Non rate specific:**
 - Support a BER of better than or equal to 10^{-13} at the MAC/PLS service interface (or the frame loss ratio equivalent)
 - Provide support to enable mapping over OTN
- **400 Gb/s Related:**
 - Support a MAC data rate of 400 Gb/s
 - Support Physical layer specification of 400 Gb/s over 4 pairs of SMF with lengths up to at least 2 km
- **800 Gb/s Related:**
 - Support a MAC data rate of 800 Gb/s
 - Support **optional eight-lane 800 Gb/s attachment unit** interfaces for chip-to-module and chip-to-chip applications
 - Support Physical layer specification of 800 Gb/s operation:
 - over 8 pairs of MMF with lengths up to at least 50 m
 - over 8 pairs of MMF with lengths up to at least 100 m
 - over 8 pairs of SMF with lengths up to at least 500 m
 - over 8 pairs of SMF with lengths up to at least 2 km(電気系IFは割愛)

● P802.3dj (800G/1.6T関連)

- **Non rate specific:**
 - Same as p802.3df
- **800 Gb/s Related:**
 - Support a MAC data rate of 800 Gb/s
 - Support optional **four-lane 800 Gb/s** attachment unit interfaces for chip-to-module and chip-to-chip applications
 - Support Physical layer specification of 800 Gb/s operation:
 - over 4 pairs of SMF with lengths up to at least 500 m & 2km
 - over 4 wavelengths over a single SMF in each direction with lengths up to at least 2 km & at least 10 km
 - over 1 wavelength over a single SMF in each direction with lengths up to at least 10 km
 - over a single SMF in each direction with lengths up to at least 40 km
- **1.6 Tb/s Related:**
 - Support a MAC data rate of 1.6 Tb/s
 - Support optional **sixteen-lane or eight-lane 1.6 Tb/s** attachment unit interfaces for chip-to-module and chip-to-chip applications
 - physical layer specification of 1.6 Tb/s operation:
 - over 8 pairs of SMF with lengths up to at least 500 m
 - over 8 pairs of SMF with lengths up to at least 2 km(電気系IFは割愛; 1.6T関連で2km以上の規定は現時点では未検討)

800ZR/800LRともに2023/4Q完了予定であるが2024年にずれ込む予定 **昨年(MPLS Japan)の資料に一部加筆**

- **800ZR:** 400ZRをそのまま800G対応にしたもの

- Single channel 800G coherent interface, 150GHz spacing
- Inter-DC applications over 80km reach
- 100GE/200GE/400GE/800G-ETC/**800GE** client types
- DP-QAM16 & OFEC, **118.2Gbaud**
- ITU-T FlexO-8ベース(FlexO-8e)のフレーム定義

- **800LR:**

- Unamplified link (6-8dB loss)
- Single channel 800G coherent interface
- Intra-DC applications up to 10km reach
- **DP-QAM16 123.6Gbaud**
- **FEC: RS(544, 514)/outer & BCH(126.110)/inner**
- 400GE and 800G-ETC/**800GbE**
- Form factor agonistic
- 低遅延、低消費電力なモジュール実現を目指し新たなコヒーレントIF

参考文献:

<https://www.oiforum.com/wp-content/uploads/ECOC-2022-Market-Focus-Gass-final1.pdf>

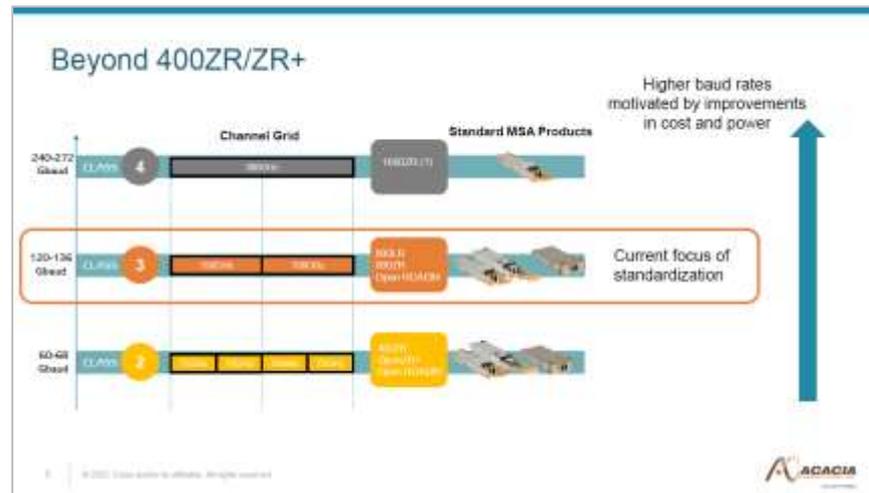


https://www.oiforum.com/wp-content/uploads/OFC2023_OIF_800ZR_800LR_Josef-Berger.pdf



- 1.6T ZR(1600ZR)の作業開始合意
 - 2023 3Q会合で合意
 - <https://www.oiforum.com/oifs-q3-technical-and-mae-committees-meeting-spurs-four-new-innovative-projects-link-training-for-224g-electrical-interfaces-1600zr-data-center-optical-network-coordination-energy-effi/>
 - Microsoft, Marvell, Cisco, Infinera & 多数サポーター
 - DCI向けであることは共通
 - 800ZRも進行中であることは承知の上で、1.6Tについても検討すべきという考えのもと新規プロジェクトとして承認
 - 2026年完成予定
- なお、800G/1.6T FlexEも合意(済み)
 - 2023 2Q
 - <https://www.oiforum.com/oif-launches-four-innovative-projects-covering-a-broad-range-of-technologies-at-q2-technical-and-mae-committees-meeting/>

- 参考: 1600ZRのイメージ
 - 以下の最上段



<https://www.oiforum.com/wp-content/uploads/OIF-Panel-Tom-Williams-Final.pdf>

2023/4会合で以下の文書をコンセント(のち承認)

- **G.709.1(FlexO-SR)改正4: まもなく発行予定**
 - FlexO-8の基本を定義
 - FlexO-8-RSなど詳細はこれから
 - FlexO-xe(FlexO-8e, FlexO-4e)定義
 - OIF 400ZR/800ZR, OpenZR+(400G)で採用しているFlexO(相当)のイーサネットマッピング(直収)を可能に
 - GMPについても記載
 - FlexO Regen(eration) を定義
 - FlexO AIS, FlexO LCK定義
- **G.798(OTN装置処理)改訂**
 - 既存のFlexO関連更新
- **G.sup58改訂**
 - FOIC 1.1 (100G lane)等記載

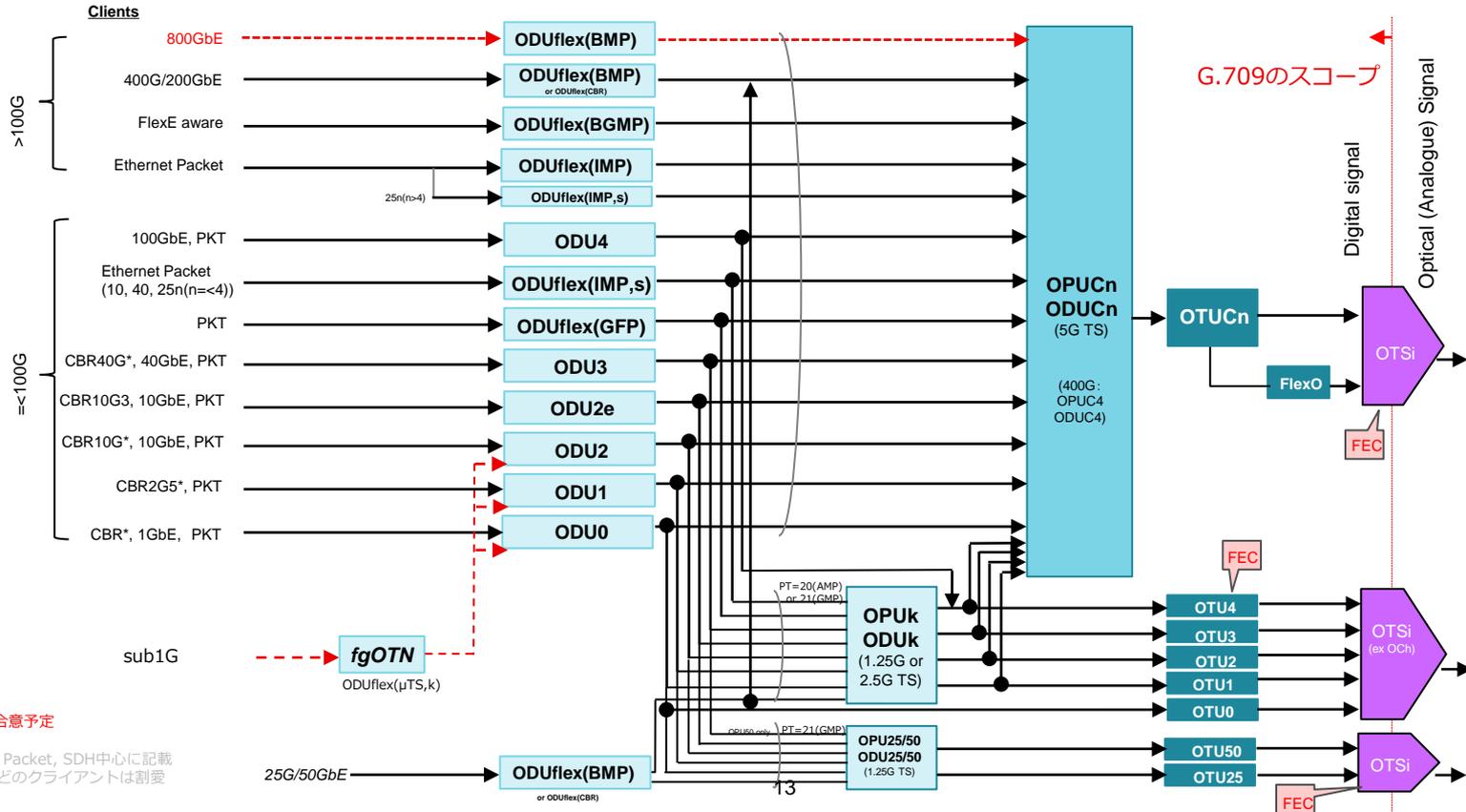
本日は、過渡期ではあるものの、大幅に更新されるFlexOの最新動向について、改正4を中心に、(再編されても、混乱を伴わないよう)紹介

2023/11会合で以下の文書をコンセント予定

- **G.709(OTN)関連**
 - 改正として以下を追記または更新
 - 800GbEマッピング記載
 - GMP(Generic mapping procedure)一般化
 - fgOTN定義と記載
- **FlexO勧告再編**
 - 以下4勧告(新規または改訂)をコンセント
 - **G.709.1**をFlexO共通関連(フレーム, FlexO Secなど)を集約
 - **G.709.3**を **B100G向け** FlexO LR(Long reach)に改編
 - 100G, 200G, 400Gが対象 (追加はなし)
 - G.709.1の RS FEC関連を、新勧告**G.709.5**として独立
 - 800G も含む
 - **B400G LR**相当FlexO を新勧告**G.709.6**として規定
 - 400G, 800G が対象ざっくり記すと130Gbaud相当のFlexOとも
- FlexO, fgOTN アーキを反映した**G.872**更新(Q12)
- fgOTN関連勧告(G.709.20), **G.sup58**更新

G.709: OTN 多重階梯(2023)

● G.709: OTN 多重階梯(2023)

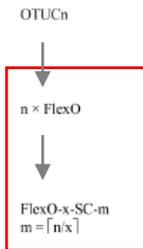


赤点線は年内合意予定

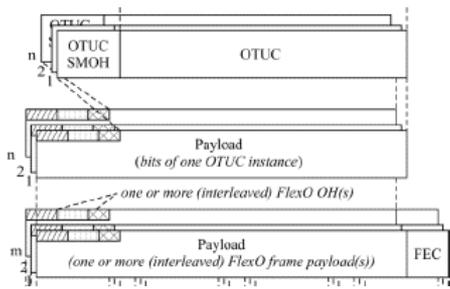
主にEthernet, Packet, SDH中心に記載
FC, IB, SDIなどのクライアントは割愛

● おさらい: FlexO とは

- G.709で既定の、B100G向けOTN (OTUCn)を転送するためフレーム定義
 - 現在は100G, 200G, 400Gの光モジュールでの定義 (FlexO-1,2,4)
 - FEC定義により複数の勧告で定義
 - G.709.1 (FlexO-SR): RS FEC
 - G.709.3 (FlexO-LR): SC FEC, CFEC, OFEC



この処理が実は
難解かつ重要



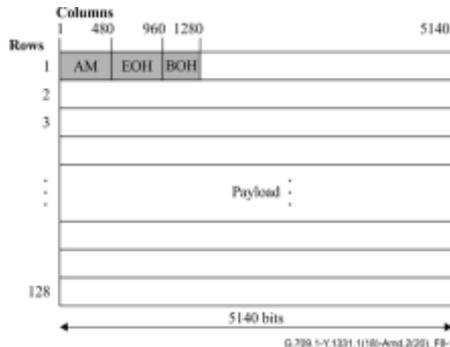
(ITU-T G.709.1から抜粋)

例
n=8 i.e., 800G(OTUC8)を
400G FlexO (m=4), i.e.,
FlexO-4で伝送する場合

OTUC ベースに8(n)個の
FlexOフレームを作成

FlexO-4フレーム x 2
に再編。FECをつけて
光伝送を行う

● FlexO Frame structure (100G基本フレーム)



FlexO-xはこのフレーム
を再編・拡張

- FlexO-2: 10280 x 128
- FlexO-4: 10280 x 256
- FlexO-8は今回決定



- EOH/BOH: Extended/basic Overhead
 - 次以降に紹介

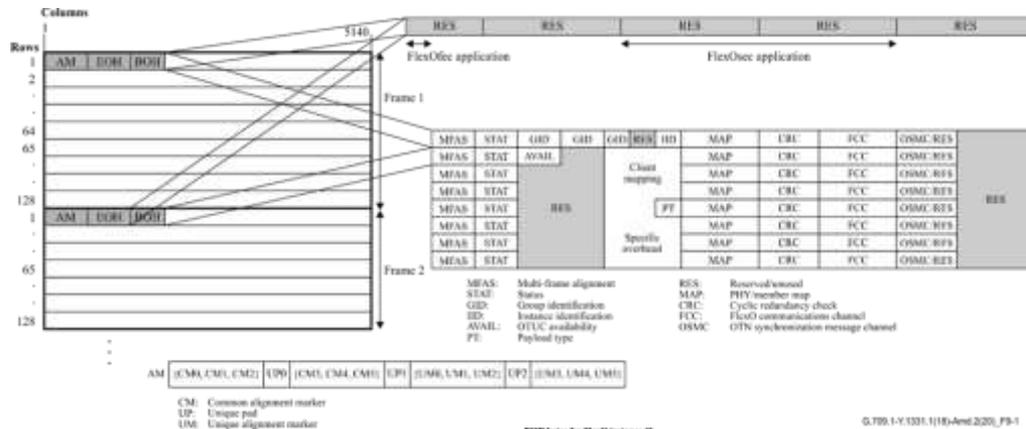
● FlexO-4における実伝送速度(10~25%の帯域増加)

- FlexO-4-RS (RS FEC): 447.237898 Gbps
- FlexO-4-SC (SC FEC): 450.937476 Gbps
- FlexO-4-DSH (CFEC): 502.639686 Gbps
- FlexO-4-DO (OFEC): 505.115743 Gbps

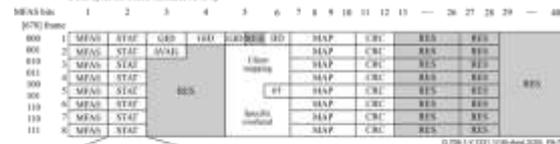
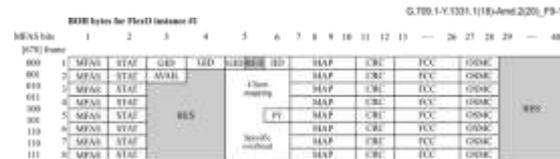
FlexO関連 (BOH) ※これまで

- OTU同様にマルチフレーム構成
 - FlexO と FlexO-x とで微妙に異なる点注意
- 現在、定義されているフィールドと目的
 - **STAT** (Status)
 - 現在は RPF (Remote Physical Layer Fault) 1ビットのみ、FlexO-x の instance#1 に定義
 - **GID** (group identification)
 - FlexO-x-<fec> interface が属する Group に対する ID。受信側で指定された FlexO Groupに属することを確認。instance ごとに設定される
 - 20 bit で定義され双方向で定義される
 - **IID** (instance identification)
 - Group 配下に属する 各 instance 番号。8 bit
 - **MAP** (PHY/member map)
 - 256ビット(1~254が有効)で定義され、各 instance ID (IID) の Group への属し方を示す。IID 相当のビット位置に 1 のフラグを立てることで、各 instance の存在を示す
 - **AVAIL** (OTUC availability)
 - FlexO-1-RS 100G interfacesに対し1をセット(するが、実質使われない)
 - **PT** (payload type) は後述
 - **CRC**
 - CRC-16/MFAS, OSMC and FCC を除くOHに対して実行

$$x^{16} + x^6 + x^5 + x^3 + 1$$

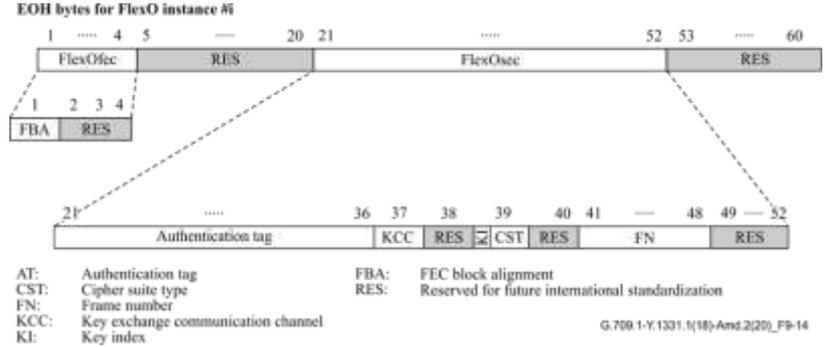


- **FCC**
 - FlexO management communications channel
 - 896-bit
- **OSMC**
 - OTN synchronization message channel
 - SSM ないしは PTP messagesがGFPを介して実装される



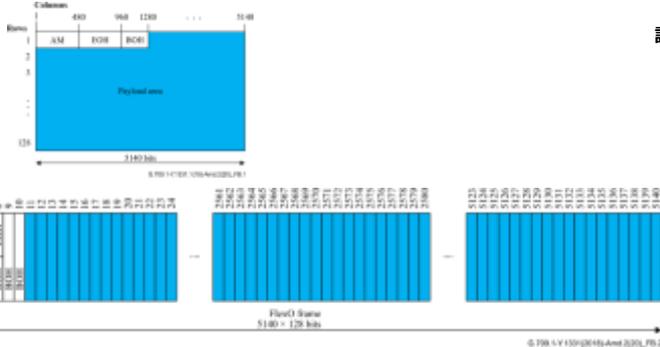
FlexO (EOH) ※これまで

- 現在のFlexO OH (EOH)で定義されているものは、FlexOfec と FlexOsecだけ
 - FlexOfec
 - Staircase FEC block alignment (FBA)を定義 (G.709.3)
 - FlexOsec
 - 以下を定義
 - Authentication tag (AT)
 - Frame number (FN)
 - Key index (KI)
 - Cipher suite type (CST)
 - Key exchange communication channel (KCC)
 - GCM-AES-256 を前提定義 (詳細はAnnex A参照)
 - FlexO 100G インスタンスごとに実行

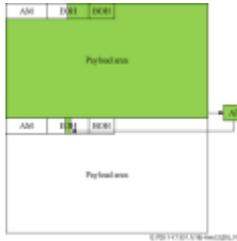
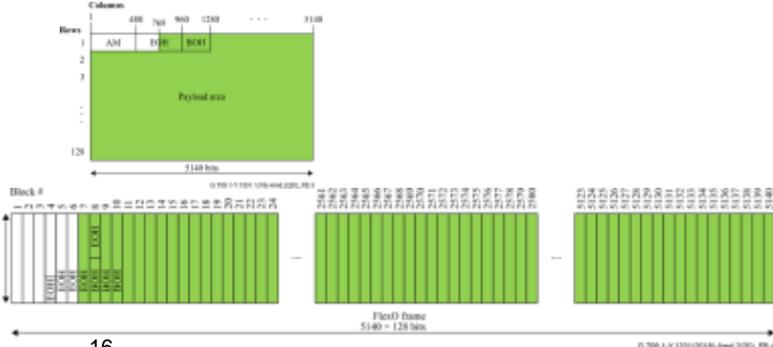


G.709.1-Y.1331.1(18)-Amd.2(20)_F9-14

暗号化

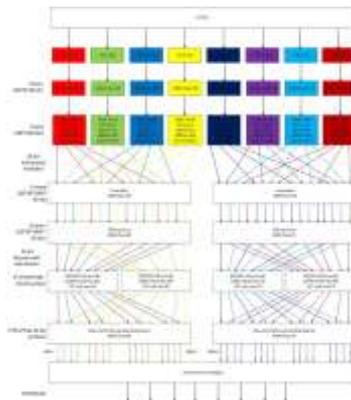
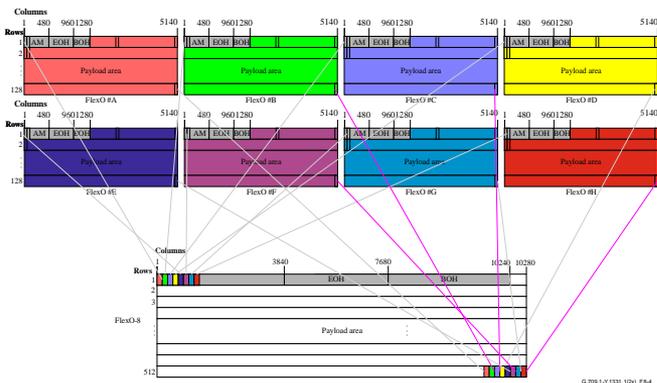


認証



● FlexO-8の基本を定義

- 800G FlexO-8(e) frame structure
 - 改正4では、LR系FlexO-8(e) (16QAM) interfaceにのみ言及し、Interleaving $z=128$ と記されているが、SR系 FlexO-8 ($z=10$)も許容
 - 右下の図> G.709.5に



● FlexO-xe(FlexO-8e, FlexO-4e)定義

- OPU/ODU/OTUを用いず、Ethernetを直収可能な FlexOフレームを定義
 - ZRフレーム(FEC付加前)と等価

Table 8-1 – FlexO bit rates

Interface Type	FlexO nominal bit rate	FlexO bit-rate tolerance
FlexO-x	$491384/462961 \times 99\ 532\ 800$ kbit/s	± 20 ppm
FlexO-xe	$21845/25984 \times 766 \times 156\ 250$ kbit/s	± 20 ppm

NOTE 1 – The nominal FlexO bit rate used with FlexO-x is approximately: $105\ 643\ 510.782$ kbit/s. The FlexO bit rate can be derived from the OTUC bit rate as follows: $4112/4097 \times \text{OTUC bit rate} = 4112/4097 \times 239/226 \times 99\ 532\ 800$ kbit/s.

NOTE 2 – The nominal FlexO bit rate used with FlexO-xe is approximately: $100\ 622\ 438.327$ kbit/s. The FlexO bit rate can be derived from 156M Ethernet clock multiple as follows: $514\ 544 \times 1445/1624 \times 766 \times 156\ 250$ kbit/s.

Table 8-2 – FlexO frame and multi-frame periods

Interface Type	Frame period (Note)	Multi-frame period (Note)
FlexO-x	~ 6.228 μ s	49.822 μ s
FlexO-xe	~ 6.539 μ s	52.310 μ s

NOTE – The period is an approximated value, rounded to 3 decimal places.

● PT拡張

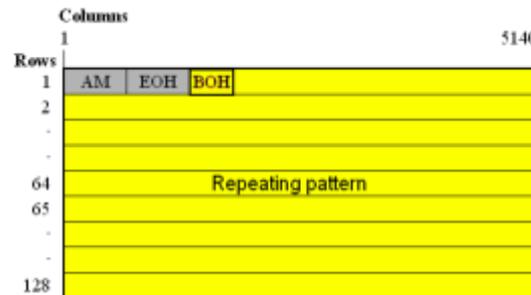
- 以下を追加定義
 - 0x05: FlexO-xe (GMP based Ethernet mapping)
 - 0xFE: PRBS test (PRBS31)
 - 0x55/0xFE: LCK/AIS向けに確保 (パターンデータで上書き)
 - なお、OIF向けの確保(0x40, 0x41)は改正3で実施済み

Table 9-4 – FlexO payload type code points

MSB 1 2 3 4	LSB 5 6 7 8	Hex code (Note 1)	Interpretation
0 0 0 0	0 0 0 0	00	Bit synchronous OTUC mapping, see clause 10
0 0 0 0	0 0 0 1	01	Experimental mapping (Note 2)
0 0 0 0	0 0 1 0	02	GMP based OTUC mapping for OTUCn multiplexing for FlexO-x- fec interfaces, see [ITU-T G.709.3]
0 0 0 0	0 1 0 0	04	Reserved for FlexOsec SquelchText value
0 0 0 0	0 1 0 1	05	GMP based Ethernet mapping for FlexO-xe- fec interfaces
0 1 0 0	0 0 0 0	40	Reserved OIF 800ZR mapping and multiplexing of Ethernet clients
0 1 0 0	0 0 0 1	41	Reserved OIF 800ZR PRBS test pattern
0 1 0 1	0 1 0 1	55	Reserved for STAT LCK maintenance signal
1 0 0 0	x x x x	80-8F	Reserved codes for proprietary use (Note 3)
1 1 1 1	1 1 1 0	FE	PRBS test pattern, see clause 10.3
1 1 1 1	1 1 1 1	FF	Reserved for future STAT AIS maintenance signals

● FlexO AIS/LCK

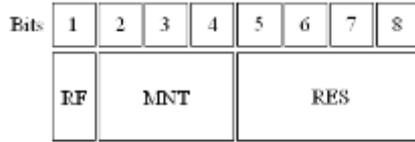
- ODU AIS/LCKと同じパターンデータを規定
 - AIS: 1111 1111
 - LCK: 0101 0101
- BOHも上書きの対象に



- 詳細動作はG.798(OTN装置処理)で今後記載予定

● FlexO OHの拡張と警報転送機能追加

● STAT(BOH)



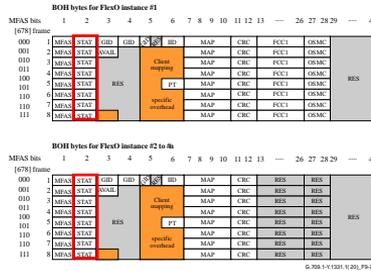
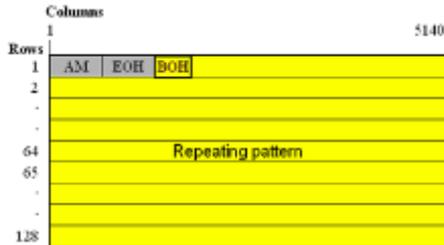
G.709.1-Y.1331.1(1E)-Amd.4(20)_F9-10

- RF: RPF改め

● MNT (Maintenance) :

- 000 – Normal operation
- 111 – AIS
- 101 – LCK

AIS も LCK も動作的にはODU-AIS / ODU-LCKと同じになり、パターンデータ(スケルチングデータ)を生成AM, EOHを残し送信する



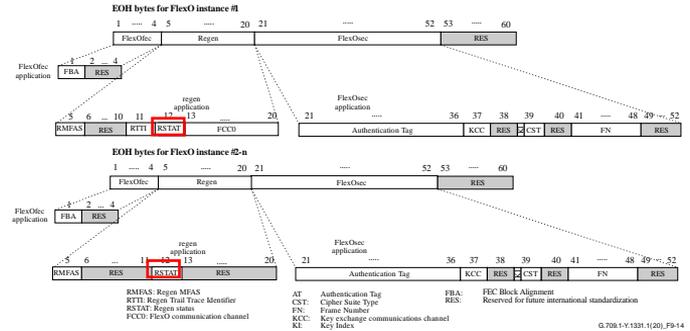
G.709-1-Y.1331.1(1E)-Amd.4(20)_F9-14

● RSTAT(EOH)

- Regen(FlexO-x)のための定義



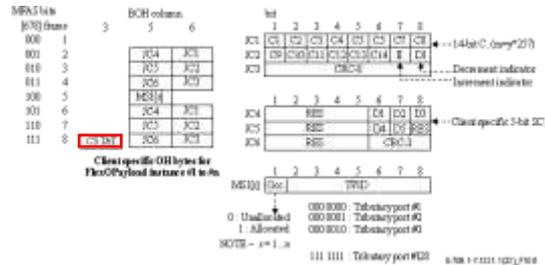
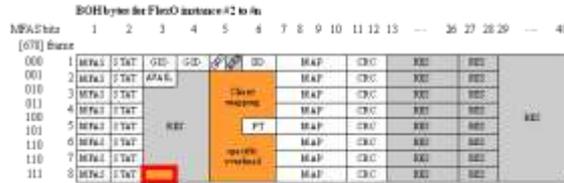
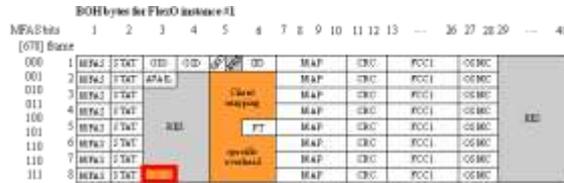
G.709.1-Y.1331.1(1E)-Amd.4(20)_F9-14



- RF: Remote fault
- LD/RD: Local and remote degrade. FlexO-x-<fec>, つまりFlexO向け光モジュールのdegrade通知

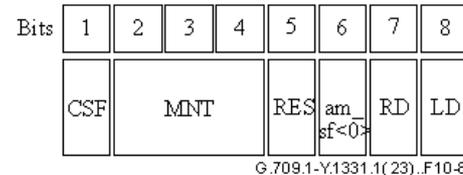
● FlexO OHの拡張と警報転送機能追加(続き)

- CSTAT (BOHの一部)
- FlexO-xe (イーサネット直収)に対応



● CSTAT定義

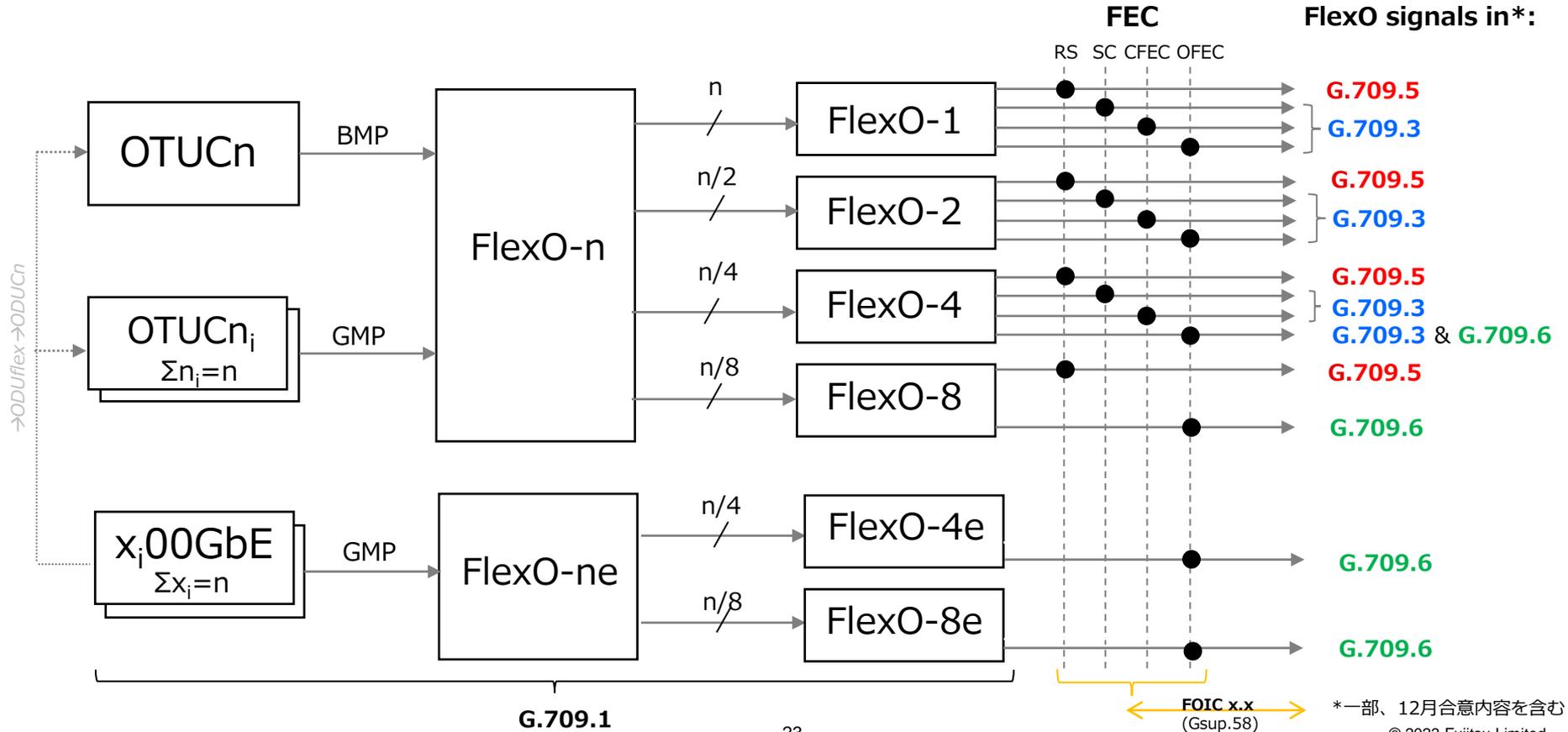
- クライアント向け
- Client signal fail (CSF)
 - CSFを受信したら、LFをoverwrite
- Maintenance (MNT)
 - 000 – Normal operation
 - 101 – LCK
 - LCK投入したらETC(クライアントイーサネット)は折り返し
- Client local and remote degrades (LD, RD)
 - am, RD, LD は400ZR/800ZRに順番
- Reserved (RES)



G.709.1-Y1331.1(23)..F10-8

詳細動作はG.798(OTN装置処理)で今後定義の方向

G.709.x: FlexO 多重階梯(2023)



*一部、12月合意内容を含む
© 2023 Fujitsu Limited

● 勧告体系更新

- 以下の通り合意
- 黄マーカ一部は次回コンセント予定
- 今回から fgは**fine grain**として定め、fgOTN, fgMTNそれぞれ別レイヤとして定義する
 - common sub1g layerは考えない

Topic	fgOTN	fgMTN
Support of sub1G clients overview	G.709.20	G.8312.20
fg Layer Architecture	G.872 Annex	G.8310 Annex
fg Interface	G.709 Annex	G.8312 Annex
Server/fg adaptation	G.709 Annex PT in main body	G.8312 Annex PT in main body
Protection	G.808.4	
Equipment	G.798 Annex	G.8321 Annex
Synchronization	G.8251 Annex	G.mtn-sync Annex
Management Requirements	G.874 Annex	G.8350 Annex
Protocol-neutral Information Model	G.875 Annex	G.8350 Annex

将来の光伝送に向けた課題と期待(1)

WDM(Media, L0)における運用面の課題

標準化の求められる(期待)されること

● 運用管理として求められる姿

- つまりユースケース、要求事項
- (期待する)担当団体: OIF, ITU-T, MEF

● ネットワーク特性を考慮したアーキテクチャ定義(主にData plane)

- 担当団体: ITU-T

● ネットワークモデル定義

- UML, YANGなどの Information Model, Data Model
 - ひいてはAPIにも関連
- 担当団体: ITU-T, IEEE, IETF (ccamp)
 - ONFは…
 - 実は、OSCなど含めると多数あり
 - 当方のMPLS Japan 2017が参考になるかも◎

● 運用管理の視点からの要件定義とインタフェース定義

- FCAPSがベースになるがAPI定義も関連
 - この要件定義、という言葉が意外とキーワード(マジックワード)かも
 - ネットワーク性能要件、顧客ならでの管理要件の具体化をどう定義するべきか
 - 担当団体: ITU-T, IETF(TEAS, ccamp), OIF, TIP, ONF
- ### ● 他、Intelligentな運用に向けた枠組み
- 自動化、自律化、AI/MLとの連携
 - ユースケース・要求・要件定義との  ? ?
 - 担当団体: ITU-T, IETF, OIF, ONAP

出典 <https://blog.nic.ad.jp/2022/8417/>

光ネットワーク運用管理における標準化動向 ~新しいアーキテクチャ勉強会(光の巻)振り返り その1~

標準化の求められる(期待)されること

● 運用管理として求められる姿

- つまりユースケース、要求事項
- (期待する)担当団体: OIF, ITU-T, MEF

● ネットワーク特性を考慮したアーキテクチャ定義(主にData plane)

- 担当団体: ITU-T

● ネットワークモデル定義

- UML, YANGなどの Information Model, Data Model
 - ひいてはAPIにも関連
- 担当団体: ITU-T, IEEE, IETF (ccamp)
 - ONFは…
 - 実は、OSCなど含めると多数あり
 - 当方のMPLS Japan 2017が参考になるかも◎

現場・現物・現状に忠実に理想を追い求める件

● 運用管理の視点からの要件定義とインタフェース定義

- FCAPSがベースになるがAPI定義も関連
 - この要件定義、という言葉が意外とキーワード(マジックワード)かも
 - ネットワーク性能要件、顧客ならでの管理要件の具体化をどう定義するべきか
 - 担当団体: ITU-T, IETF(TEAS, ccamp), OIF, TIP, ONF
 - 他、Intelligentな運用に向けた枠組み
 - 自動化、自律化、AI/MLとの連携
 - ユースケース・要求・要件定義との   ? ?
 - 担当団体: ITU-T, IETF, OIF, ONAP
- ここのあるべき・ありたい姿の飽くなき探求、これが将来への架け橋になる件

将来の光伝送に向けた課題と期待(2)

伝送における新領域(DC向けネットワーク)への適用における期待

● 背景

- AI/MLによるデータ処理量の爆発的な増大とそれを支えるサーバー構成・技術の革新への期待
- 脱酸素社会の実現に向けた電力削減への取り組みとコミットの必要性
 - ネットワーク系では、半数を占めるDCネットワーク系の消費電力削減が急務。特にサーバー系
- 光電融合・高速伝送など光関連技術の進化

● 要求

- テラビット級DC/DCIの実現(提供)
- リソース最適を図るサーバー構成技術
- スケールしかつ低遅延伝送方式の提供
 - DC内に対してCPU-メモリ間の遅延削減による処理速度の向上・リソース最適化が求められるだけでなく
 - 高度なデータ処理(i.e. スケール)のためにはDC間において、伝送速度向上と伝送距離延伸化がもとめられる
- これらを実現しつつ、省スペース・高度実装化・低消費電力化を提供できるDC/DCI構成・運用技術

● 課題

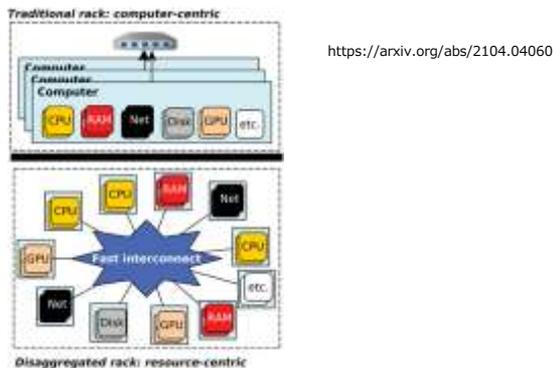
- 現在のプラガブルでの限界(が、やがて来る)
 - 800GE/1.6TE標準化(IEEE802.3)も200G laneは2026年まで延伸。2T以上では、従来の延伸では2030年までもつかどうか疑問
- 224G以上の電気レーンの実現性
- 既存のラック構成ではCPU-MEM間リソースの最適配置ができない
- 拡張にはCPU-MEM間低遅延化が望まれるが、メモリー伝送速度が伸びないことで、スケールしない

● 起きていること(起ころうとしていること)

- DC内のDisaggregation(Composable)化
 - リソースの最適配置
- Spine-Leaf に捉われないDCの拡張と構成
- 光技術(コンポーネント)のDC/DCIへの適用
- CXLなど低遅延かつスケール可能なでのリソース間接続技術の出現

● DC内のDisaggregation(Composable)化 ● CXLなどのリソース間接続技術の出現

- Brice Ekane et al, *Network in Disaggregated Datacenters*



- R Lin, et al, *Disaggregated Data Centers: Challenges and Trade-offs*

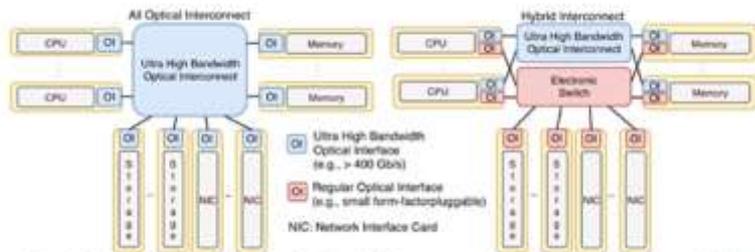
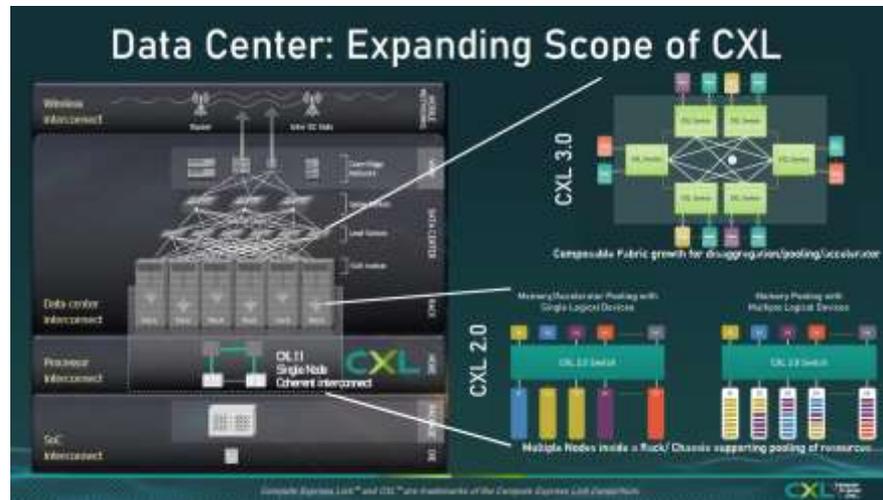


Figure 2. Fully disaggregated rack with all-optical interconnect (left) and hybrid interconnect (right)

- **Compute Express Link (CXL):** PCIeベースのプロセッサからメモリなど周辺機器・アクセラレーターへの新しいリンクプロトコル
- CPU-Memory の Desegregation を実現しつつ低遅延かつスケールするDC、かつフレキシブルなリソース提供を実現

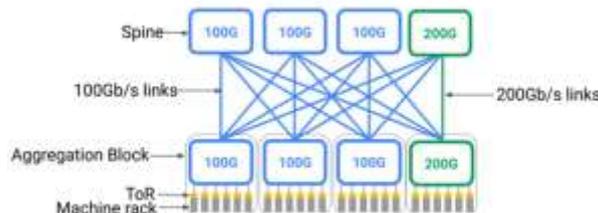


https://0c141887-fbe4-4ec3-be17-
 adc8d70d3922.usrfiles.com/ugd/0c1418_037d4ba31f4b44cf9fcb37f5b36ae4d6.pdf

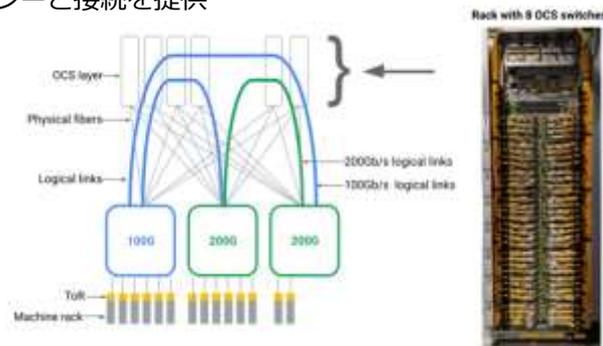
- Spine-Leafに捉われないDCの拡張と構成 & 光技術(コンポーネント)のDC/DCIへの適用
- Googleの光技術のデータセンターへの適用の紹介
 - 光スイッチ(OCS; MEMS光スイッチ)のDC内導入ならびにスイッチ制御手法を提案
 - DC内における従来と異なるAggregatorの提案
 - 提案手法により電気スイッチおよび光トランシーバーの電力を41%削減、Capexを30%削減、伝送帯域を5倍以上
 - DC間だけでなくDC内への光化への期待
- 主な文献
- Leon Poutievski et al., *Jupiter Evolving: Transforming Google's Datacenter Network via Optical Circuit Switches and Software-Defined Networking*
 - <https://conferences.sigcomm.org/sigcomm/2015/pdf/papers/p183.pdf>
- R. urara et al, *Mission Apollo: Landing Optical Circuit Switching at Datacenter Scale*
 - <https://arxiv.org/abs/2208.10041>

● 従来の Clos トポロジーの課題

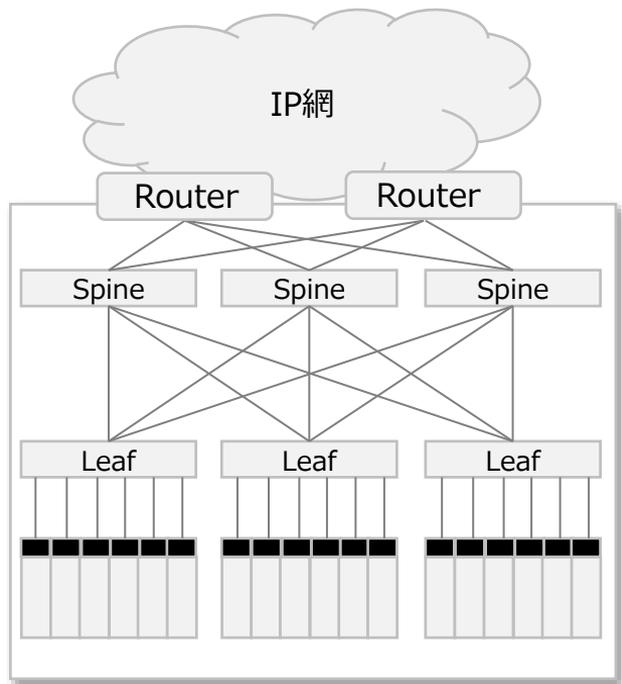
- 接続速度が均一である必要があるため、リンクスピードの高速化を伴ったアップグレードが困難(パフォーマンスが出ない)



- 光スイッチを伴った柔軟な接続を提供することでフレキシブルなトポロジーと接続を提供



Data Center向けネットワーク: 今後に向けて

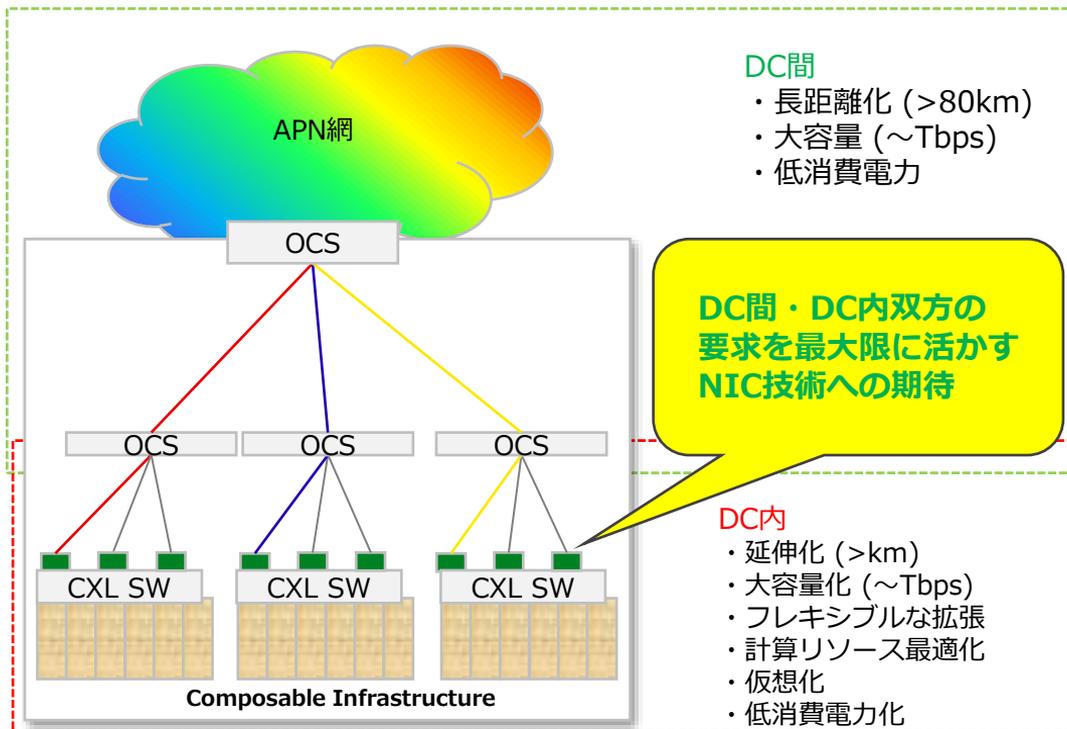


FY2023

■ Top of Rack

■ Photonic Smart NIC

OSC: Optical Circuit Switch



DC間

- ・長距離化 (>80km)
- ・大容量 (~Tbps)
- ・低消費電力

DC間・DC内双方の
要求を最大限に活かす
NIC技術への期待

DC内

- ・延伸化 (>km)
- ・大容量化 (~Tbps)
- ・フレキシブルな拡張
- ・計算リソース最適化
- ・仮想化
- ・低消費電力化

FY2030~

IOWN GF文書(DCI FA, APN FA)ならびに、
弊社資料(Fujitsu All Photonic Network Vision)もご参照くださいませ

まとめ

- 以下の内容について紹介

- **B400G光伝送(特に800G)標準化動向**

- IEEE802.3 (P803.2df, djなど), OIF 800ZR/800LR, ITU-T G.709シリーズ(FlexO, 800GbE収容) の最新標準化動向
 - 三者間で密接に関わった作業状況の紹介
- 800G関連標準化は今年度でほぼ完成
 - 今後(2024～)は1.6Tに向けた活動に本格シフト?

- **将来の光伝送に向けた課題と期待**

- WDM(Media, L0)における運用面の課題
 - 詳細はこちらを→ <https://blog.nic.ad.jp/2022/8417/>
光ネットワーク運用管理における標準化動向 ～新しいアーキテクチャ勉強会(光の巻)振り返り その1～
- 光伝送における新領域への適用における期待
 - 今後のDC/DCIのあるべき姿と、光技術(伝送・デバイスなど)の進展への期待
 - 一方、DC(コンピューティング)の世界とどう付き合う(or 距離を取る)のか、相手を知りつつ連携が必要かも

Thank you

