

100Gbps DCI技術解説

Broadband Tower Cloud&SDN研究所

加藤良輔 <rsk@lab.bbtower.net>

■ 加藤 良輔

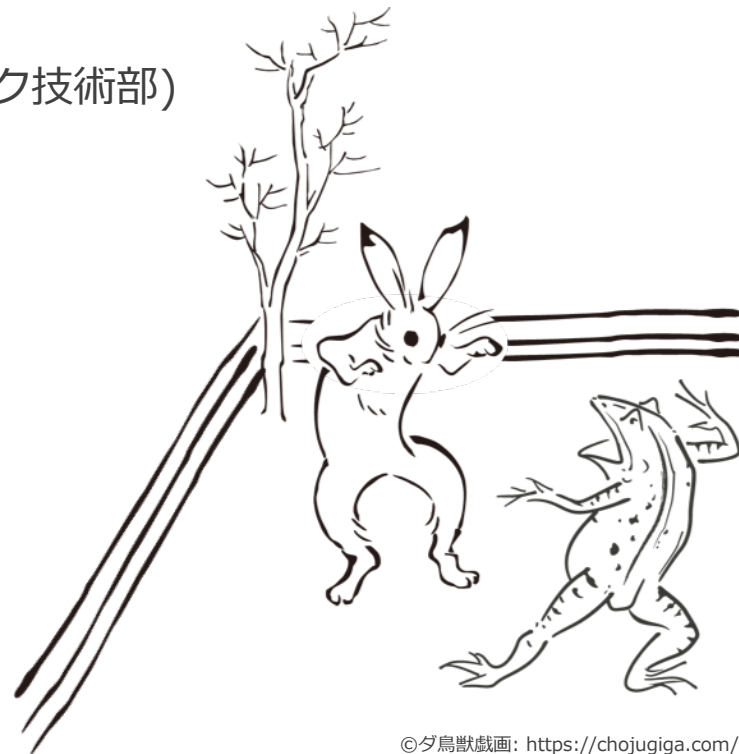
- ◆ Cloud&SDN研究所（なんでも屋） 所属
- ◆ 最近、真面目な肩書きが増えました（ネットワーク技術部）

■ 活動内容

- ◆ 社内サービスの研究開発・運用補助
- ◆ 研究所用AS (AS7530/AS59099) の設計運用
- ◆ 実験系Internet Exchange拠点運用
- ◆ 夫概 对外活動色々

■ 好物

- ◆ カメラとおもちゃの鉄砲
- ◆ ロス・インゴベルナブレス・デ・ハポン（プ男）



©ダ鳥獣戯画: <https://chojugiga.com/>

100Gbps伝送の現状

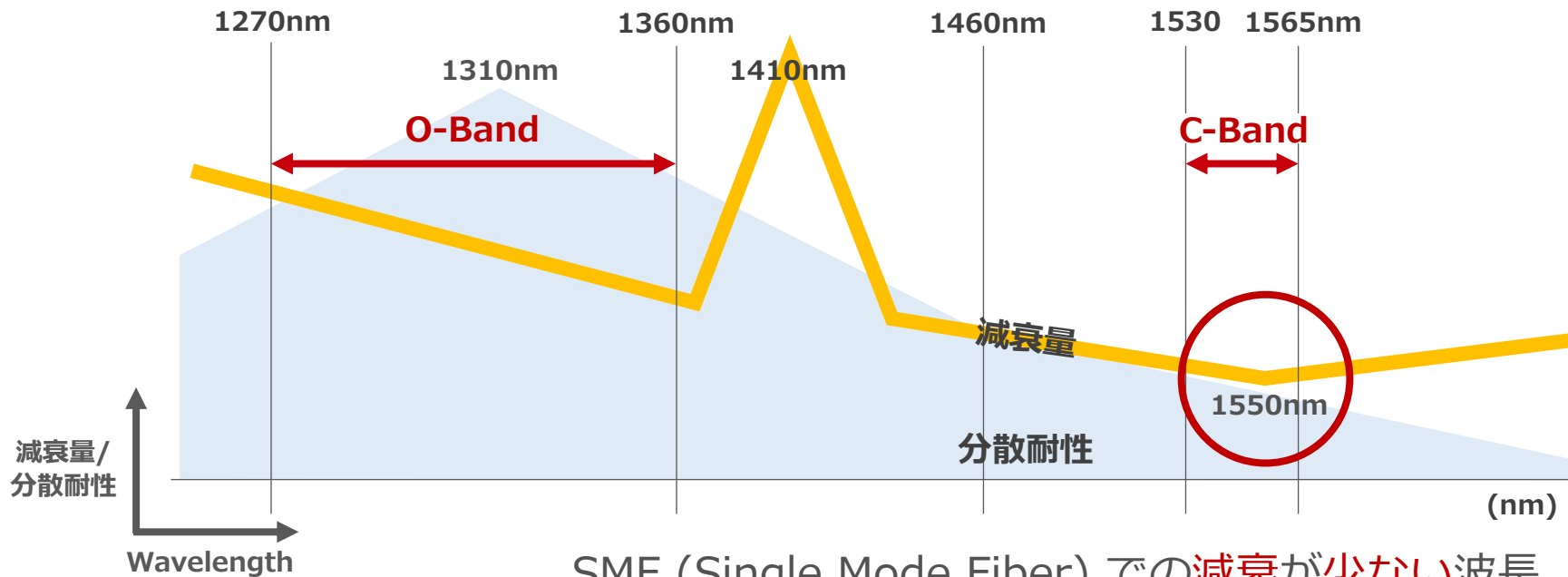
■ やっと100Gが当たり前になった時代

- ◆ ToRスイッチにまで100Gが付いている時代
 - SR4以外にも、CWDM4など短距離安価なトランシーバも増えた
- ◆ そろそろ長距離飛ばしたくなる頃合い

■ カジュアルに使っている100Gの光の正体

- ◆ ほとんどが O-Band (1310nm近辺) の製品
 - 長距離用ER/ZR もO-Band製品 (ER4-Lite=30km, ZR4=80km)
 - 10G-ER/ZRとかは1550nm帯を使っていたものだが・・・
- ◆ Ethernetで利用する100G製品は C-Band (1550nm近辺) のものがほとんど無い
 - C-Band は 長距離向けの波長帯
 - O-BandはC-Bandよりも損失が大きく、長距離が飛ばしづらい波長帯

JANOG44 「データセンタ間接続 (DCI) を一から作り直す話」
Wavesplitter 森川さん資料より抜粋+加工



SMF (Single Mode Fiber) での減衰が少ない波長

■ 長距離飛ばせる (≈C-Bandの) CFP2-ACO/DCO なかなか手が出せない実情

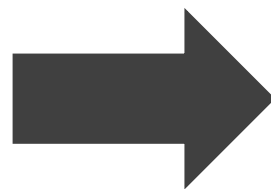
- ◆ 伝送装置と一部のメトロ向け装置でにしかインターフェイスがない
 - コスト感は東松さんのスライドにて

■ お手軽に都内ぐらいのセンチ間を100Gで繋ぎたい

- ◆ 本気の伝送装置・・・までは気合い入らない
- ◆ ちょうど良い製品がない
- ◆ ACO/DCOが数出て時間が経てばコストは下がるかもしれないが

■ 伝送の目的は

- ◆ 長距離飛ばす事
- ◆ 大容量に伝送する事
 - 一つの回線に光を多重する



今回は、
100G伝送を実現するための
要素技術を話していく



**100G の WDMが欲しい時は必ず話題に出てくる
「100G PAM4 DWDM QSFP28 module」**

100G PAM4 DWDM とは

- 100GbE装置のSerDes (Serializer/De-Serializer) は、“ほぼほぼ” 25Gbps ※
 - ◆ 100G-xxxx は だいたいLAN-WDM (O-Band) で 25Gの4波/レーンを使用して実現

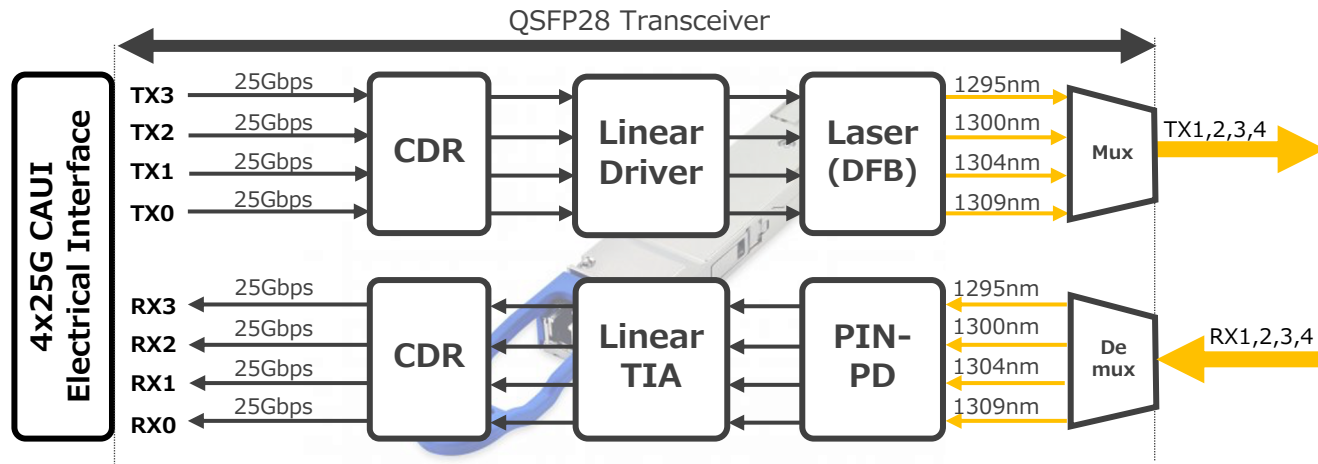


- 1GbE、10GbE のような1波の光を多重する伝送には「技術」が必要
 - ◆ 0/1の1値のシグナリングだけでは実現ができない

※ 除く 100GBase-FR, SR10

■ 100GbE装置のSerDes (Serializer/De-Serializer) は、“ほぼほぼ” 25Gbps ※

◆ 100G-xxxx は だいたいLAN-WDM (O-Band) で 25Gの4波/レーンを使用して実現



■ 1GbE、10GbE のような1波の光を多重する伝送には「技術」が必要

◆ 0/1の1値のシグナリングだけでは実現ができない

※ 除く 100GBase-FR, SR10

■ 100Gが使う4波をうまく伝送する必要がある

- ◆ 4波の光を効率化する
 - 1つの波長として伝送する
 - 「変調」をする
- ◆ 長距離伝送向けの光であるC-Bandを使用する

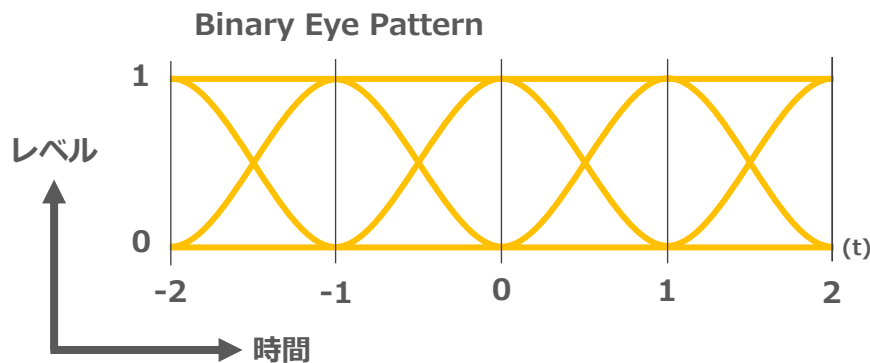
■ うまく伝送するために伝送装置は送受信機側で工夫する

- ◆ こういうところで「コヒーレント技術」が出る
 - 偏波、位相を変調して多値を持たせる
 - 専用のチップが必要になって高価になりがち
- ◆ 100G、200G、400G・・・ CFP2-ACO/DCO
 - 変調方式を変えて適切なデータ量を持たせる



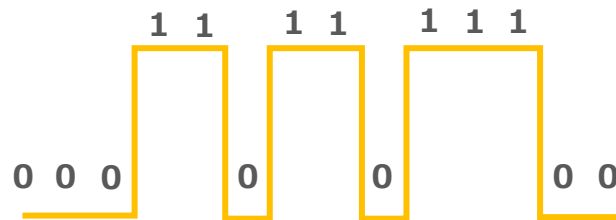
■ 非ゼロ復帰 (ノンゼロ)

- 光のオン/オフでシグナリング



光波形

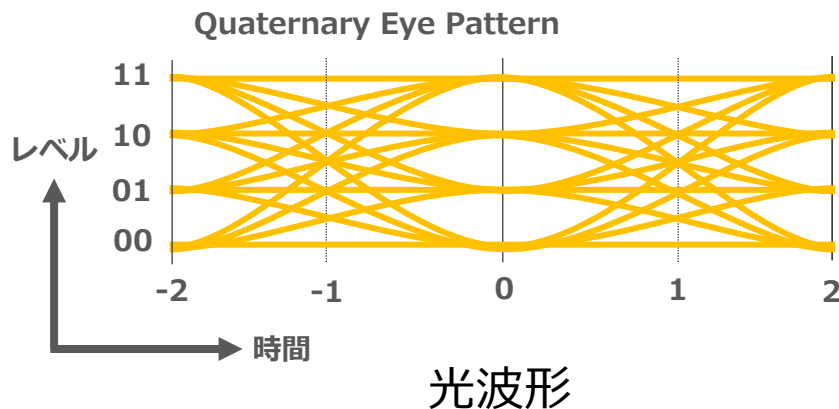
「0001101101110」



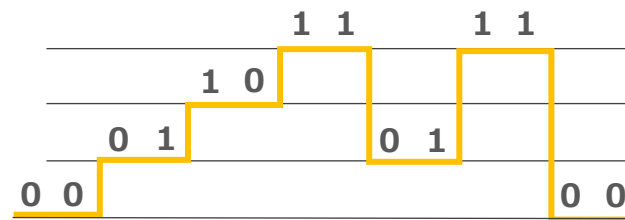
電気信号

4値のパルス振幅変調

光のオンオフ複合の振幅でシグナリング



「0001101101110」



■ DP-QPSK (偏波多重4値位相変調)

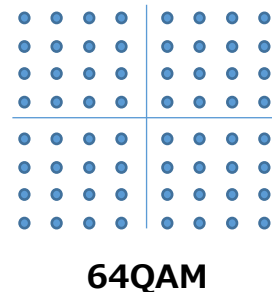
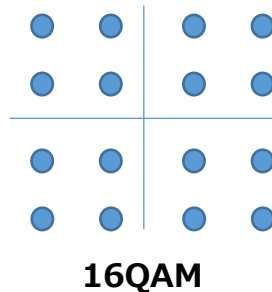
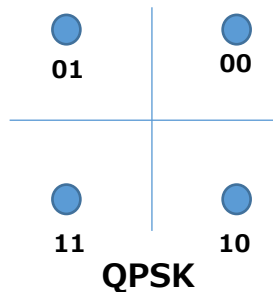
◆ 802.11b,g、デジタル無線で使われる

■ DP-16QAM (偏波多重16直角位相振幅変調)

■ たくさん多重できる変調

→ 高度なチップを積んで変調する必要がある

→ 距離が伸びないのでアンプする必要がある



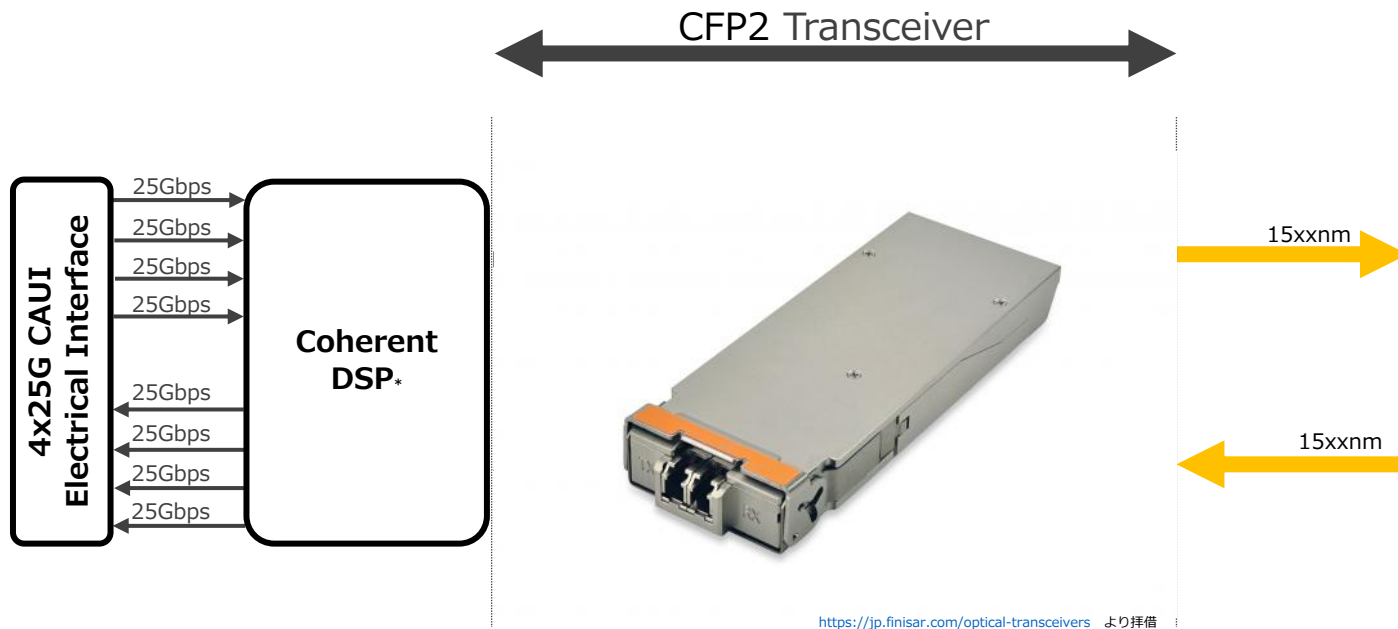
100G伝送の送受信機（トランスポンダ）

	100G CFP2-ACO	100G QSFP28 PAM4
IPoWDM	Yes	Yes
変調方式	1 λ x 28Gbaud DP-QPSK	2 λ x 28GBaud PAM4 (56G)
Grid Spacing	50GHz	100GHz
利用想定	80km以上	80km以内
RX下限	-23dB	-2dB
分散耐性	± 2400 ps/nm	± 100 ps/nm
ノイズ耐性	16dB/0.1nm	31dB/0.1nm
導入コスト	高い	低い

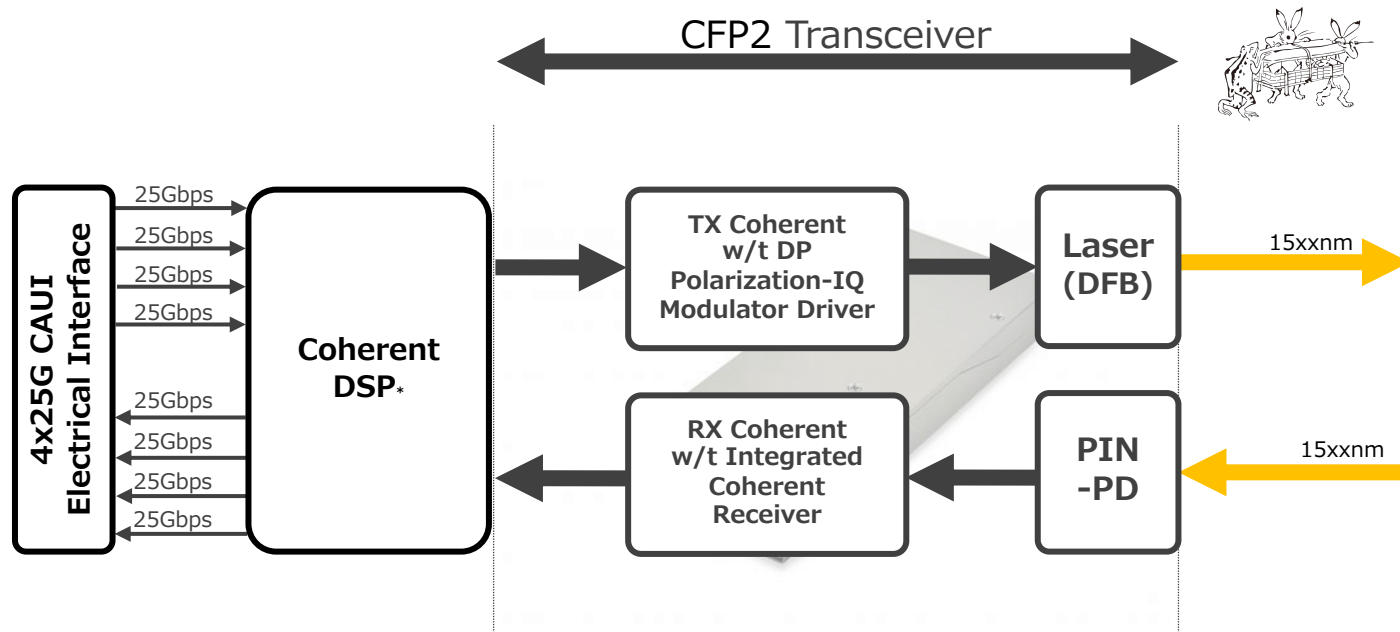
	100G CFP2-ACO	100G QSFP28 PAM4
IPoWDM	Yes	Yes
変調方式	1 λ x 28Gbaud DP-QPSK	2 λ x 28Gbaud PAM4 (56G)
Grid Spacing	50GHz	100GHz
利用想定	80km以上	80km以内
RX下限	-23dB	-2dB
分散耐性	± 2400 ps/nm	± 100 ps/nm
ノイズ耐性	16dB/0.1nm	31dB/0.1nm
導入コスト	高い	低い

今回はこっちの話

100Gトランシーバの構造 (ACO)



100Gトランシーバの構造 (ACO)

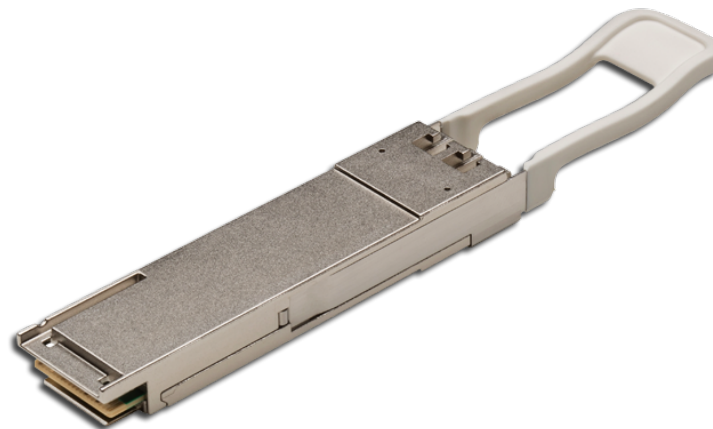


Coherent DSP (Digital Signal Processor) を
Transceiver側に積んだ DCO (Digital Coherent Optics) も存在

100G トランシーバの構造 (2λ DWDM PAM4)

QSFP28 Transceiver

4x25G CAUI
Electrical Interface

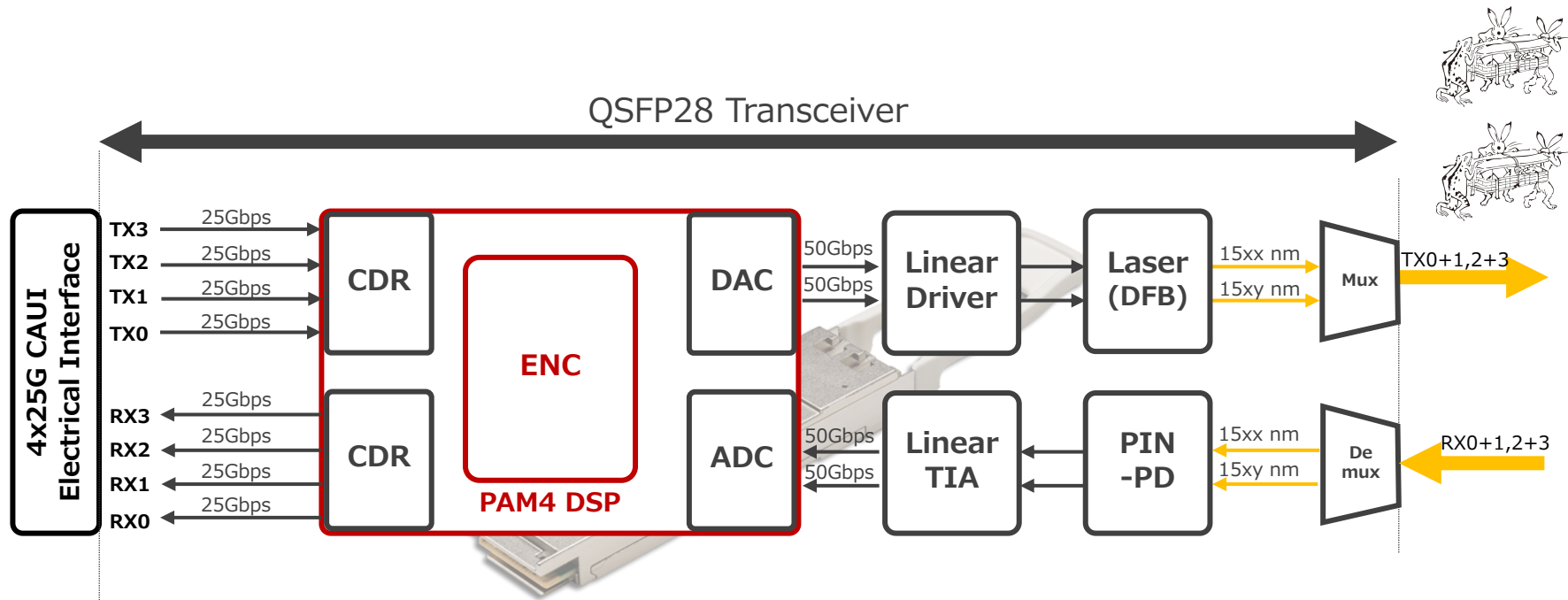


TX0+1,2+3

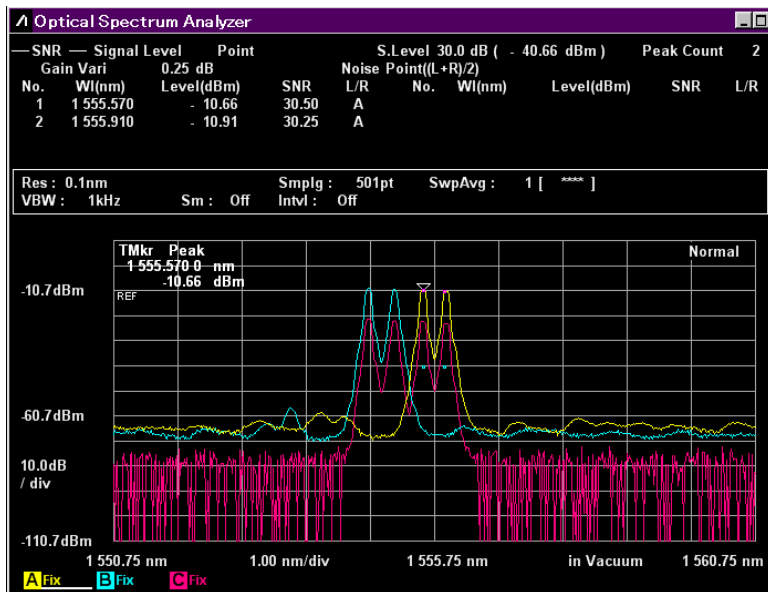
RX0+1,2+3

<https://www.inphi.com/products/>より拝借

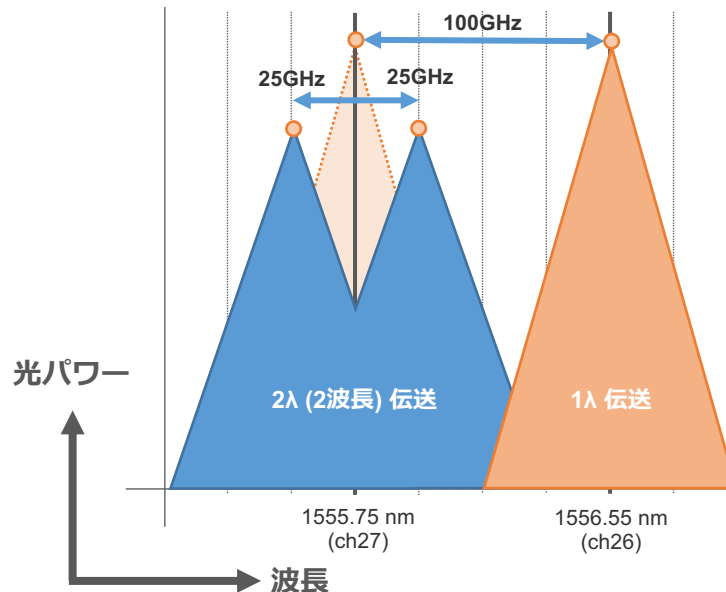
100Gトランシーバの構造 (2λ DWDM PAM4)



小型で低消費電力 (Coherent用に比べ) のPAM4用DSPをQSF28に搭載



2λ を多重した光をスペアナで見た図



100GDWDMの幅に2λが入っている図

PAM4変調用チップによる消費電力の増加

コヒーレントに比べて消費電力は少ないものの
動く箱、動かない箱、が出てくる

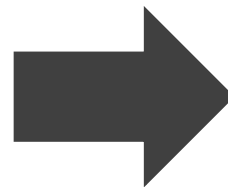


光のパワーの最大値は波長設計で決めた波長 $\pm 25\text{GHz}$

ITU-T Grid 100GHz Mux/Demux等の隙間を突く際どい事をしている
一波に比べて、波形がデリケートになる

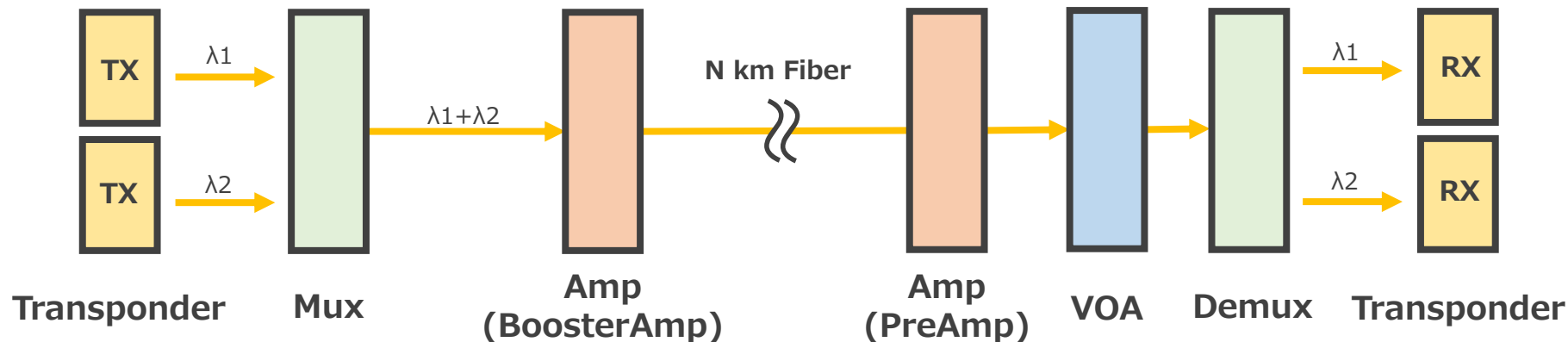
出力レベルが低い

消費電力の割に
隣同士ですら繋がらない光の強度

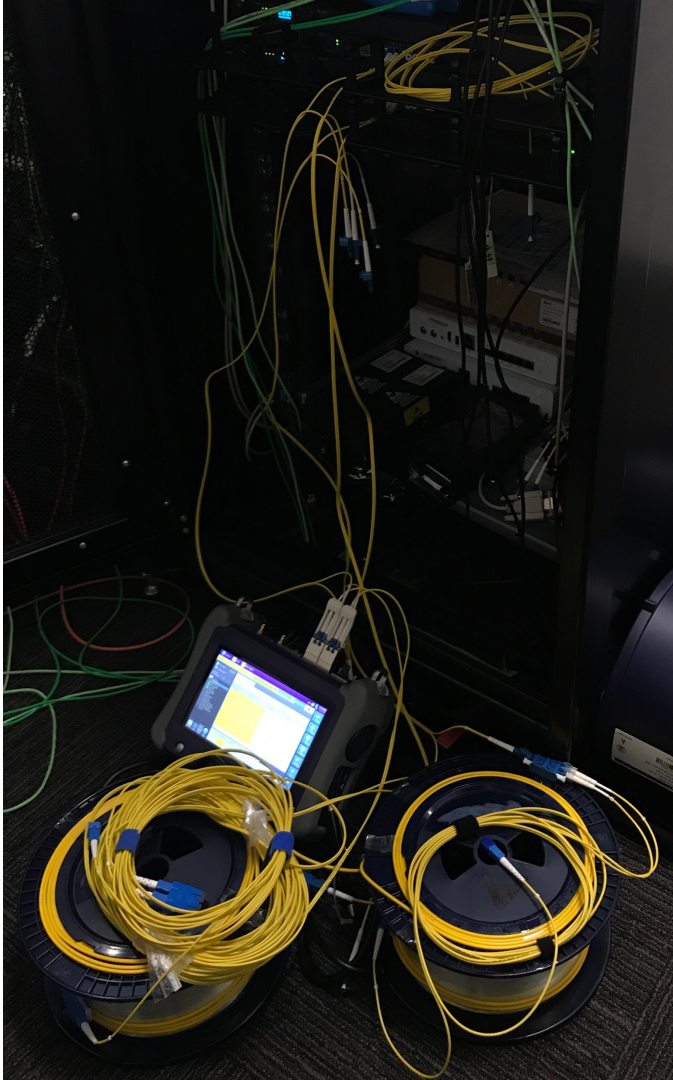


「100G PAM4 DWDM」は
この光を伝送することになる

伝送設計：入出力設計 (Gain Planning)



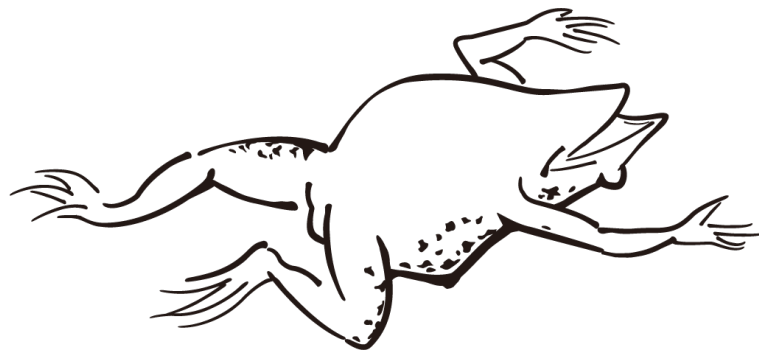
- **Transponder**
 - 異なる波長の光を発光・受光する装置
- **Mux/Demux**
 - 光の合波 (Multiplexer) する装置
 - 分波 (De-multiplexer) する装置
 - 異なる波長の光を一つに合波・分波する
- **VOA (Variable Optical Attenuator)**
 - 可変減衰器、光の強さを適切な値に落とす装置
- **AMP (EDF Amplifier = EDFA)**
 - ポンプ光を励起して信号を増幅する装置
 - BoosterAMP : 送信側に入れて「カツ」を入れる
 - PreAMP : 受信側に入れて弱った光を「拾う」



N km伝送実験の図



だが、長距離だとこのままでは繋がらない



テストに繋いでリンクが上がらない図

Port 1: 100GigE Layer 2 Traffic Term **Select Test** Port 2: 100GigE Layer 2 Traffic Term X Timing Source What's This?

Running 10m:44s
3 messages
Level (dBm) 1.5
Freq Dev (ppm) -2,195.4

Summary
Ethernet
Signal Present
Sync Acquired
Link Active
Marker Lock
Loss Of Align.
HI BER
Frame Detect
ATP Detect
Pattern Sync
VLAN Frame Detect
SVLAN Frame Detect
Local Fault Detect
Remote Fault Detect
History

Ethernet Payload
Traffic Jumbo 10000
Test Mode Frame Size

Summary Status Interface Signal

Rx Frequency (Hz)	102898600064	Signal Losses	548
Rx Freq Deviation (ppm)	(TOO LOW) -2,195.4	Signal Loss Seconds	226
Rx Freq Max Deviation (ppm)	(TOO LOW) -2,227.0	Sync Loss Seconds	464
Alignment Marker Lock	OFF	Link Loss Seconds	464
Sync Acquired	OFF	QSFP State	Ready
Link Active	OFF	QSFP Optical Rx Overload	OFF
Local Fault Detect	ON	QSFP Optical Rx Level (dBm)	1.5
Signal Loss Seconds	226	QSFP Optical Tx Level (dBm)	-6.5
Alignment Marker Loss Seconds	464	Rx Frequency (Hz)	102,898,600,064
Sync Loss Seconds	464	Rx Freq Deviation (ppm)	(TOO LOW) -2,195.4
Link Loss Seconds	464	Rx Freq Max Deviation (ppm)	(TOO LOW) -2,227.0
Local Fault Seconds	464	Tx Clock Source	Internal
Remote Fault Seconds	13	Tx Frequency (Hz)	103,125,000,000
Invalid Alignment Markers	29	Tx Freq Deviation (ppm)	0.0
		Tx Freq Max Deviation (ppm)	0.0
		Local Fault Seconds	464

Laser Actions Peak IFG Alarms Errors Faults
QSFP Connector Laser On Internal Clock Source
-1 +1 -10 +10 Freq Offset (ppm)

伝送設計: 分散補償設計 (Dispersion planning)

■ RX側の受光はアラートが出ない範囲であり、問題ない

QSFP State	Ready
QSFP Optical Rx Overload	OFF
QSFP Optical Rx Level (dBm)	1.5
QSFP Optical Tx Level (dBm)	-6.5

■ 103,125,000,000 Hz とは

- ◆ clock frequency
- ◆ 受光したシグナルが 103,125,000,000Hz に同期できず、常に走査している

⚠ Rx Frequency (Hz)	102,898,600,064
⚠ Rx Freq Deviation (ppm)	(TOO LOW) -2,195.4
⚠ Rx Freq Max Deviation (ppm)	(TOO LOW) -2,227.0
Tx Clock Source	Internal
Tx Frequency (Hz)	103,125,000,000
Tx Freq Deviation (ppm)	0.0
Tx Freq Max Deviation (ppm)	0.0

■ 本職ではない私が聞く「伝送あるある」問題

■ 「ノイズ」

■ 「分散」



ファイバ条件の良い(ノイズを考えない)
ダミーファイバ+最小の伝送構成でも同様の現象が起こる



同期できていないことや、状況から一番近いのは「**分散**※」では？

※実は、元々「分散に弱い」ことは知っていたりする

■ 「分散」 = 物質を伝わる光の波長ごとの遅延

◆ 「遅延」の結果、波形のパターンが「歪む」

■ 材料分散

◆ 光の波長とファイバ材質によって生じる遅延

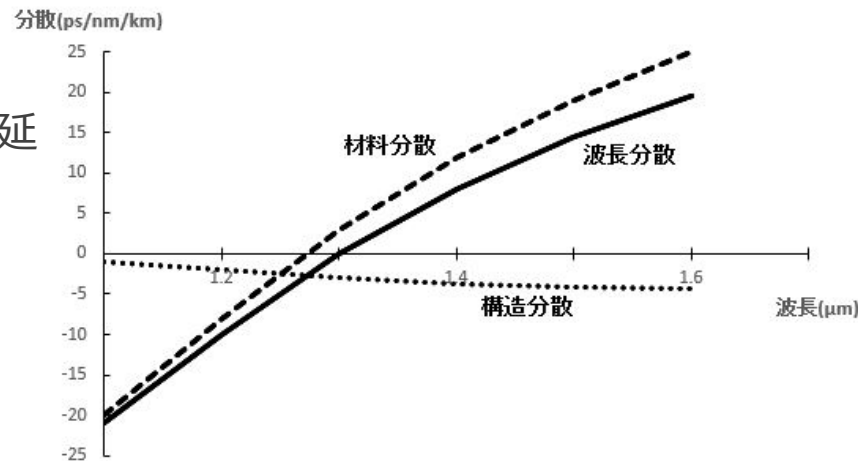
■ 構造分散

◆ ファイバのコアとクラッドの構造による遅延

「波長分散 = 材料分散+構造分散」

■ (偏波モード分散)

- ◆ ファイバのコア形状の歪みから屈折が発生する遅延
- ◆ 高速かつ長距離だと地味に効いてくる分散 ($0.1\text{ps/km}^{1/2}$)



<https://www.fiberlabs.co.jp/tech-explan/about-dispersion-compensating/> より拝借

■ 波長分散は「材料分散」の影響が大きい

■ 1310 nm は ゼロ分散

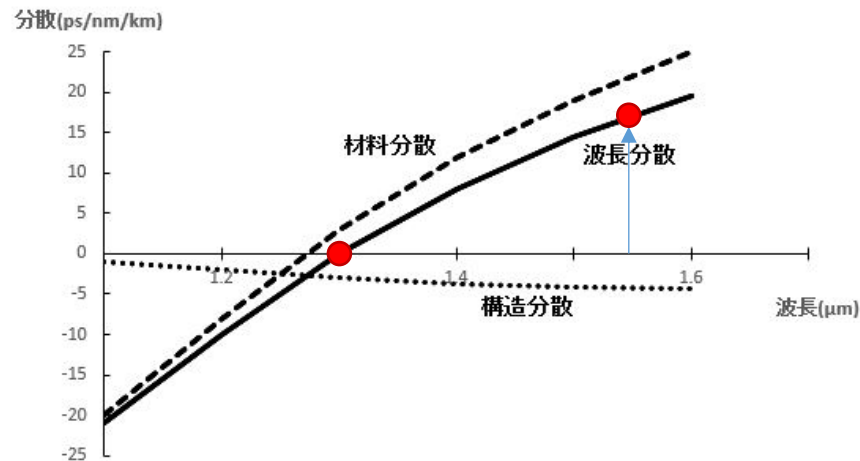
- ◆ SMFは O-Band を基準にした設計
- ◆ 長距離は飛ばないが、分散を考える必要がない

■ 15xx nm通すとなぜ分散が起こるか

- ◆ O-Bandの基準のファイバにC-Bandを
- ◆ 分散は距離に依存
 - 長距で伝送だととても効いてくる

■ C-Band用のファイバも存在

- DSF (Dispersion Shift Fiber)

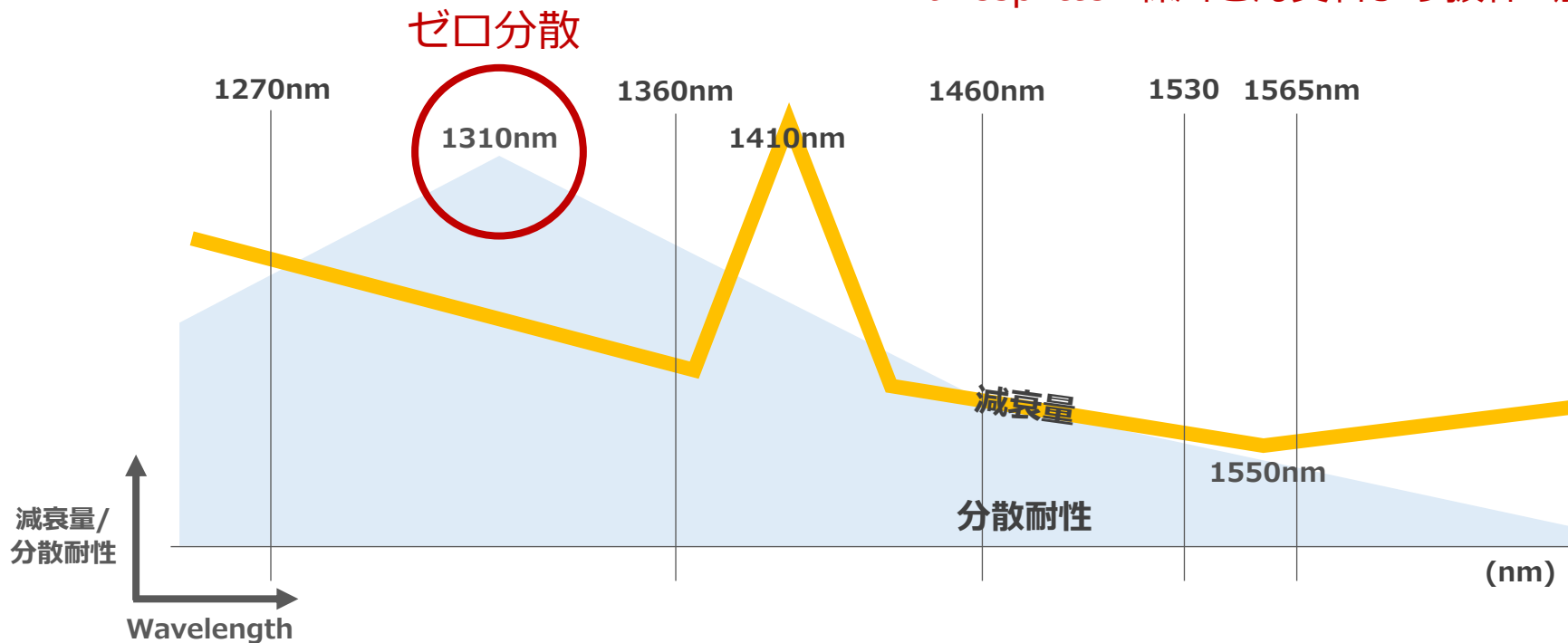


<https://www.fiberlabs.co.jp/tech-explan/about-dispersion-compensating/> より拝借

改めて O-Band & C-Band の違い

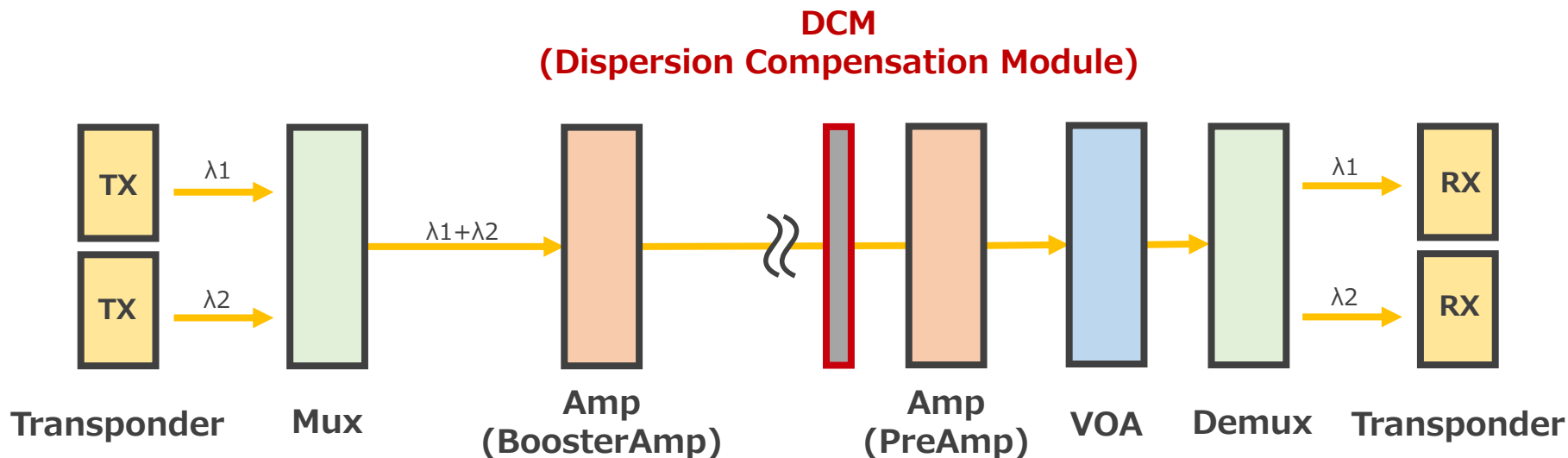
JANOG44 「データセンタ間接続 (DCI) を一から作り直す話」

Wavesplitter 森川さん資料より抜粋+加工



■ どう対策する

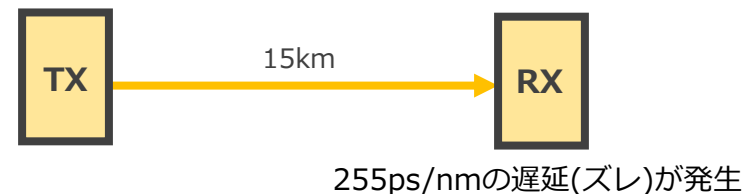
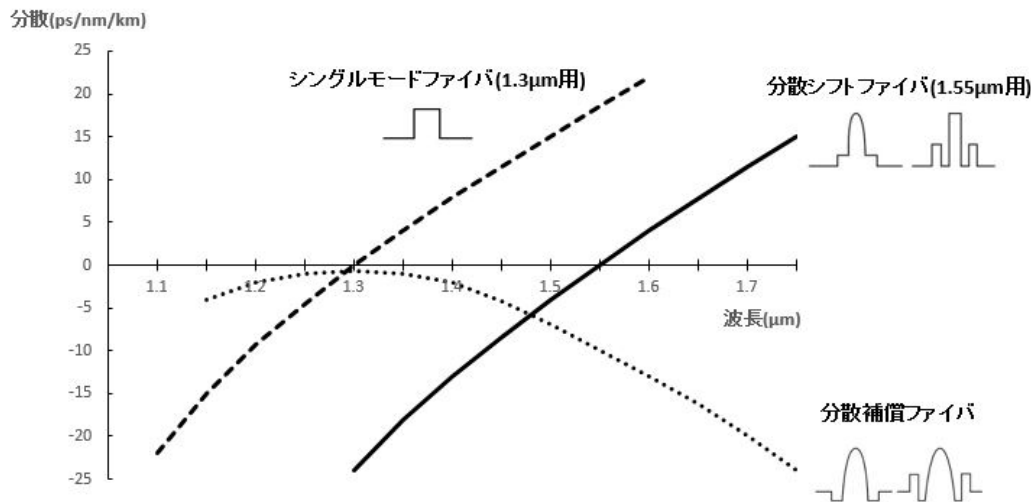
- 分散補償装置を入れる (DCM)



■ 「ずれるなら、逆方向にずらしてしまえ」

◆ ファイバ長さがさえわかれば算出は可能 (1km => +17ps/nm)

■ 特殊加工したファイバ (分散補償ファイバ) を通すことで遅延を軽減



<https://www.fiberlabs.co.jp/tech-explan/about-dispersion-compensating/> より拝借

オールグリーン

TX:5km+10km

RX:5km+10km

手元15kmのファイバでの伝送に成功

100G-CWDM4 <-> MC <-> 100G-PAM4 <-> AMP <-> (15km) <-> AMP <-> 100G-PAM4 <-> MC <-> 100G-CWDM4

導入事例

当日資料にて・・・

と思ったら画像許可が間に合わなかったのでタイトルと口頭で！

(言っても良いとは聞いていますので)

次回予告

- そもそも400Gの機器がコモディティ化するのはいつか
 - ◆ 早いに越したことはないが、もう少し後・・・な気が
- ZR/ZR+って本当に80/120km飛ぶの？
 - ◆ QSFP-DD/OSPFのサイズで？
 - DSPを積むのでわりとギリギリなサイズ
 - 消費電力から鑑みて、長距離飛ばせる光は出せるのか
 - ◆ 今回問題になった分散はDSPが頑張ってくれる
- 100G -> 400G伝送は革新的だが、長距離だと伝送ぢからが試される
 - ◆ 長距離400Gに備えて準備体操をしておきたい気持ち

まとめ

- PAM4による100G伝送は10G伝送以上に考えることが多い
 - ◆ 強度、反射、分散（今回の敵）
 - 気がつけば手元にAMP, VOA, DSM, スペアナ・・・
- 伝送以外でも多値化が進み、PAM4が当たり前になりつつある
 - ◆ 伝送以外にも 100G-[F,L,D]R や 400G-[L,F,D]R[4,8] など
- これが実現できると、そこそこ安価に
 - ◆ IPoWDM 構成でスイッチが直にWDMをやってもよい
 - ◆ メディコンを通しての100Gbpsでもよい



DIY伝送、気がつけばだんだん

「伝送沼」

に嵌っている気がする今日この頃・・・





5G Innovations