

IX相互接続実証実験を通じて見えてきた400G導入で「変わること」「変わらないこと」

Shishio Tsuchiya

shtsuchi@arista.com

Agenda

- トランシーバーの電気信号のチューニングパラメーター
- インターフェースのリンクアップ時間

メーカーっぽい事言います

- サードパーティーオプティクスの使用は気をつけよう

サードパーティオプティクスとベンダーオプティクスの違い

電気信号IEEEで定義



光学インターフェース

フォーンファクターMSAで定義

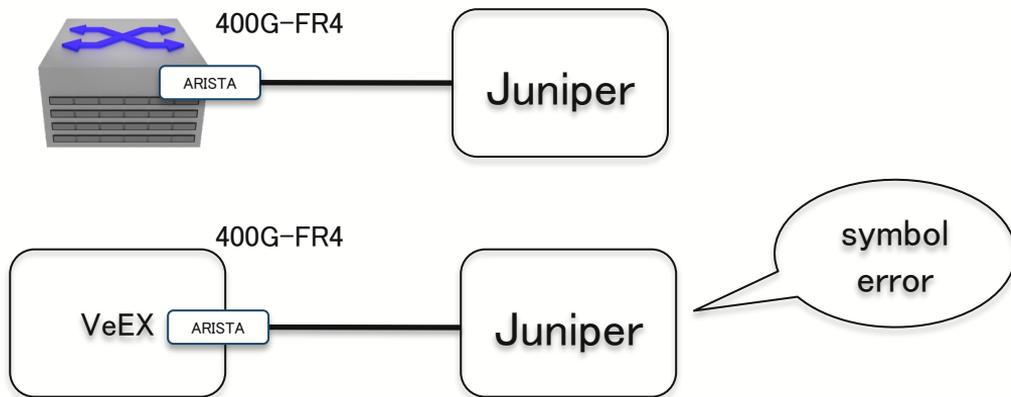


DSP/CDRIに保存されている
SerDesのチューニング値

ファームウェアは
マイクロプロセッサに保存

- 基本仕様は同じ
- Aristaで実機に接続し、繰り返しテストを実施
- 出荷ファームウェアを固定する
- EEPROMをプログラミングをする

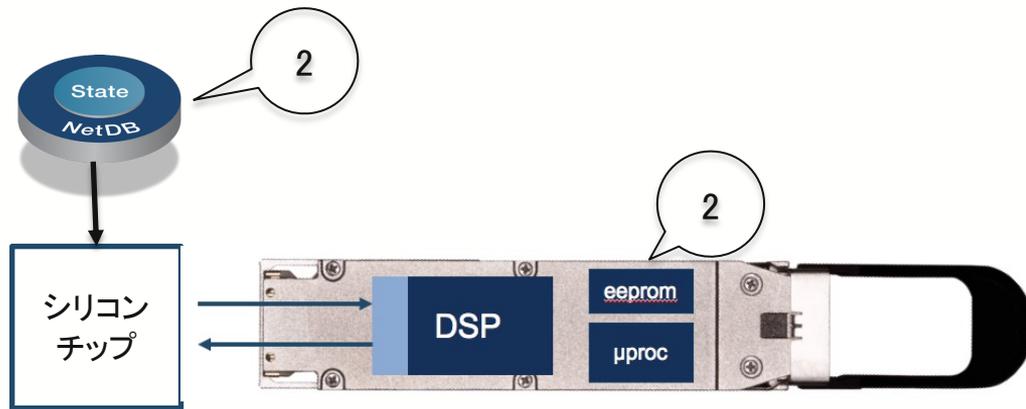
同じオプティクスであっても挙動が違う



- 測定器でJuniperに接続するとSymbol Errorを検出
- AristaでJuniperに接続すると特に問題が無い

トランシーバーの電気信号のチューニングパラメーター

<https://www.arista.com/en/support/toi/eos-4-22-0f/14299-transceiver-tuning-cli-commands>



- emphasis
 - 高周波数における減衰特性に応じて伝送信号の高域側を増幅する
- オプティクスの設定値はEEPROMに保存
- EOSではequalization(周波数最適化)/amplitude(振幅)/emphasis(増幅)が設定可能
- 測定器でも設定が可能ではあるがEEPROMの値は読み込まず、手動で設定を変更する
- 測定器で設定を変更する事で解決

トランシーバーの電気信号のチューニングパラメーター

<https://www.arista.com/en/support/toi/eos-4-22-0f/14299-transceiver-tuning-cli-commands>

```
Arista#show interfaces et14/1-8 transceiver tuning detail
```

```
* = Unsupported by module, value may be inaccurate
```

```
Rx Output Amplitude
```

Port	Module Default	Configured	Operational
Et14/1	2*	Module Default	2*
Et14/2	2*	Module Default	2*

```
Rx Output Pre-emphasis
```

Port	Module Default	Configured	Operational
Et14/1	1	3	3
Et14/2	1	3	3

```
Rx Output Post-emphasis
```

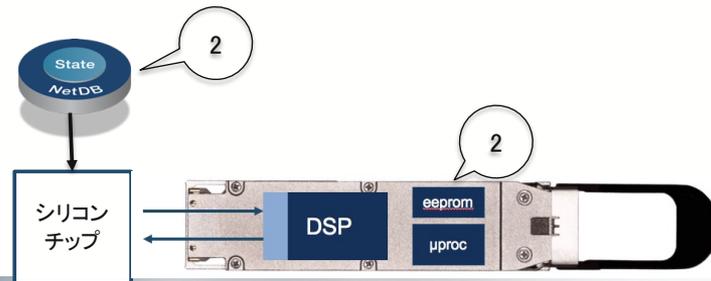
Port	Module Default	Configured	Operational
Et14/1	2	Slot Default	1
Et14/2	2	Slot Default	1

```
Tx Input Equalization
```

Port	Module Default	Configured	Operational
Et14/1	auto	Module Default	auto
Et14/2	auto	Module Default	auto
Et14/3	auto	Module Default	auto

```
--snip--
```

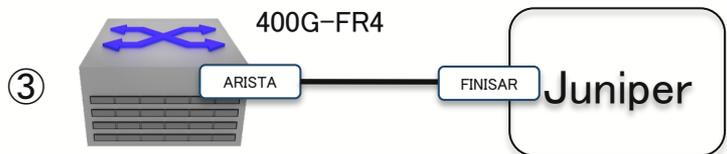
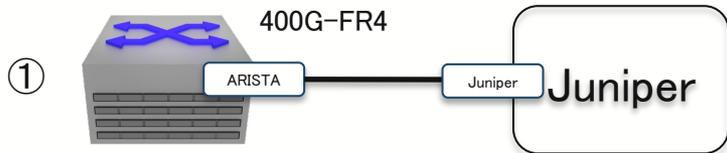
- Module Defaultの値はトランシーバーのEEPROMから読み込んだ値
- Configuredは設定した値
- Slot Defaultはシステムのデフォルト値
- *は読み込んだが設定不可の値



つまり

- モジュールのファームウェアのデフォルト値およびテストによって導き出された適正值が分かっている
- サードパーティのオプティクスを使用するときにはこれらの値および設定可能かなどの理解が必要

同じオプティクスであっても挙動が違う

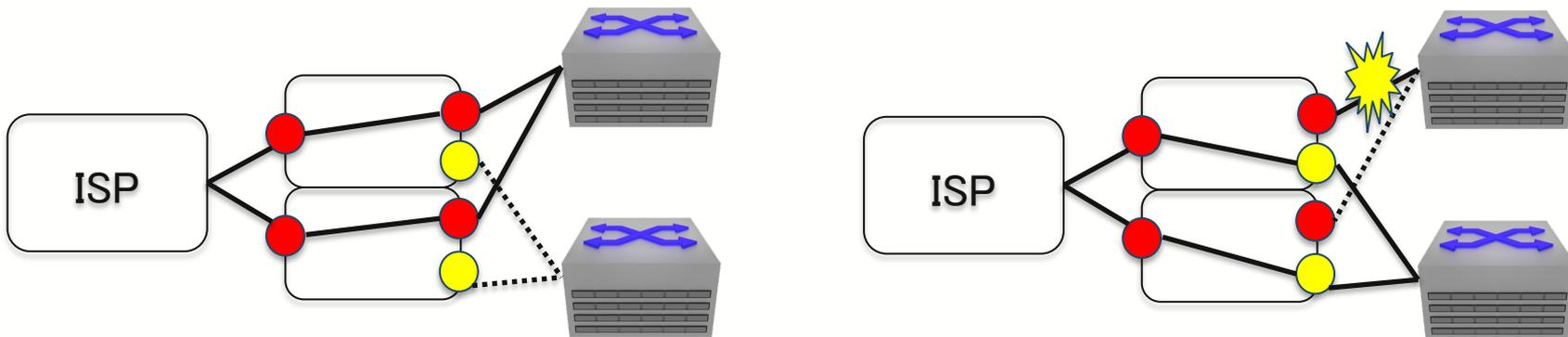


- ③の組み合わせの時のみ、立ち上がりにかかる時間がある
- トラブルの原因は**不明**だがAristaはEEPROMの値をデフォルトで使用する
- 下記はArista装置から見たEEPROMの値(実際に動作していた値)

Manufacture	Rx Output Amplitude	Rx Output Pre-emphasis	Rx Output Post-emphasis	Tx Input Equalization
FINISAR	3	3	0	auto
Arista Networks	2	3	0	auto

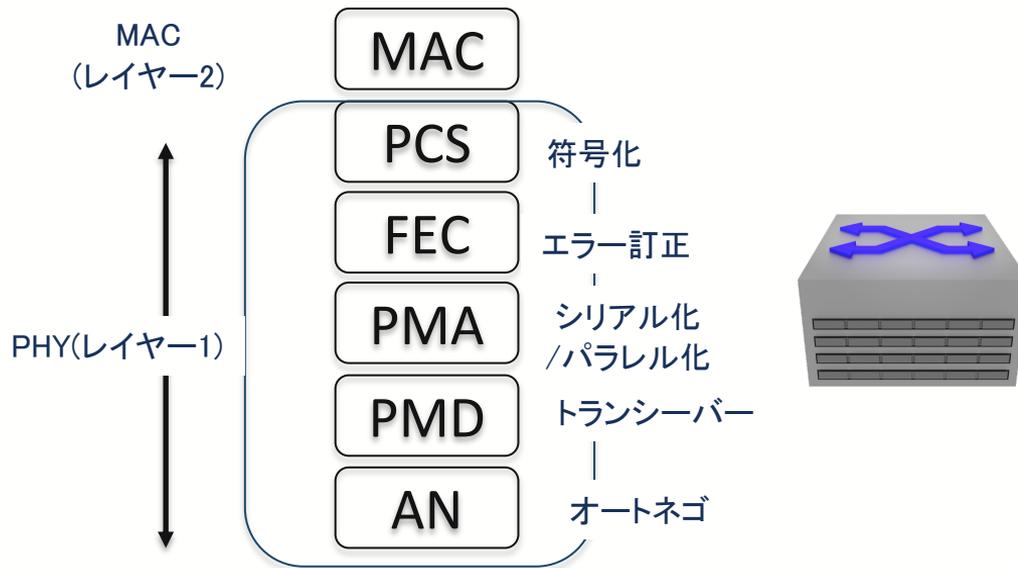
光スイッチのIXPでの利用

https://keytech.ntt-at.co.jp/network/prd_10221.html



- 日本のIXPでは光スイッチがよく使われている
- リンクやノード障害が起きた時には即座に切り替えを行う
- ISP側のルータが障害を気づかない為にはインターフェースが即座にアップする必要がある

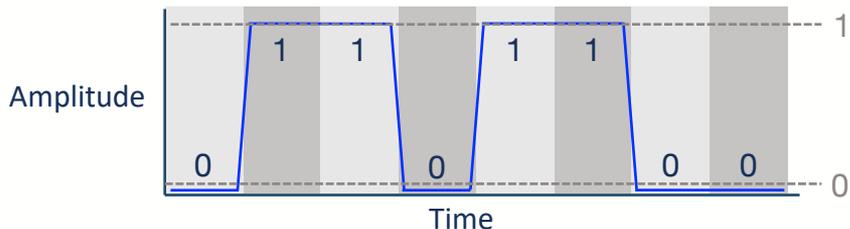
リンクはすぐにアップするのか？



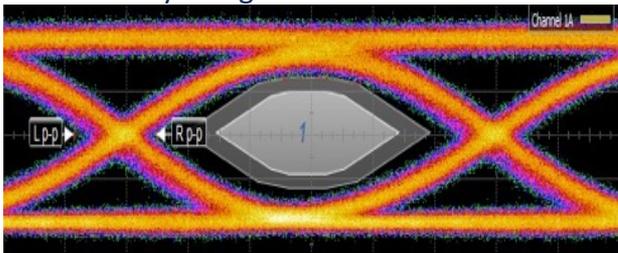
- 多くの実装はPHYで安定するのを確認してからアップする(2秒)
- phy link detection aggressiveでこれを50msecに
 - <https://www.arista.com/en/support/toi/eos-4-24-1f/14505-cli-command-for-fast-phy-link-up>
- 対抗ルータでほとんどのケースで障害に気づかずに切り替えが行われる

400Gだとどうなのか？

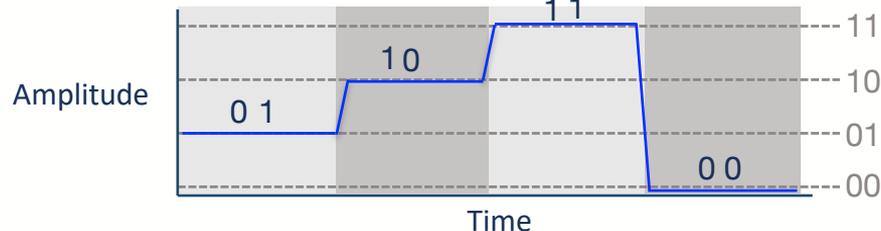
NRZ waveform for data: 0 1 1 0 1 1 0 0



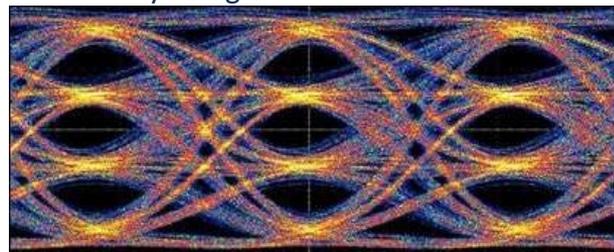
Eye diagram for NRZ data



PAM-4 waveform for data: 0 1 1 0 1 1 0 0



Eye diagram for PAM-4 data



- NRZよりPAM4は複雑になる
- 信号安定の為に2秒ほどの待ち時間が全体的に5-10秒へ
- 待ち時間をスキップしたところで2-3秒に変わる程度
- これは許容出来るのか？改修出来るのか？

まとめ

- 変わらない事
 - よりテスト/管理をされたトランシーバーを使う事が望ましい
 - 安定するまでは特に
- 変わる事
 - リンクのアップタイムなどは信号の変更の影響を受け長くなる傾向にある

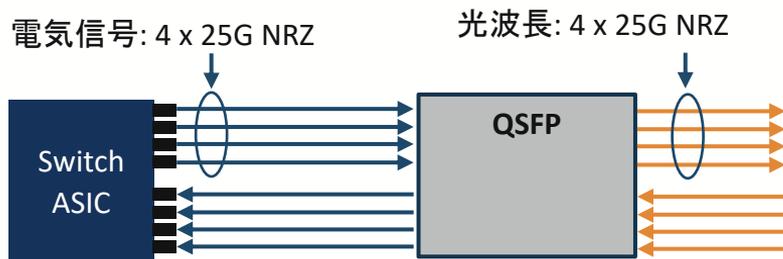


Thank You

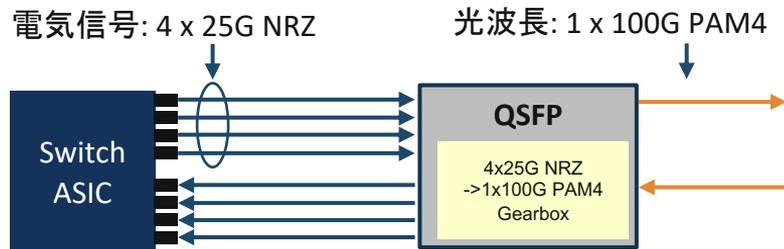
www.arista.com

100G DR/100G FRオプティクスモジュール

既存100Gオプティクス



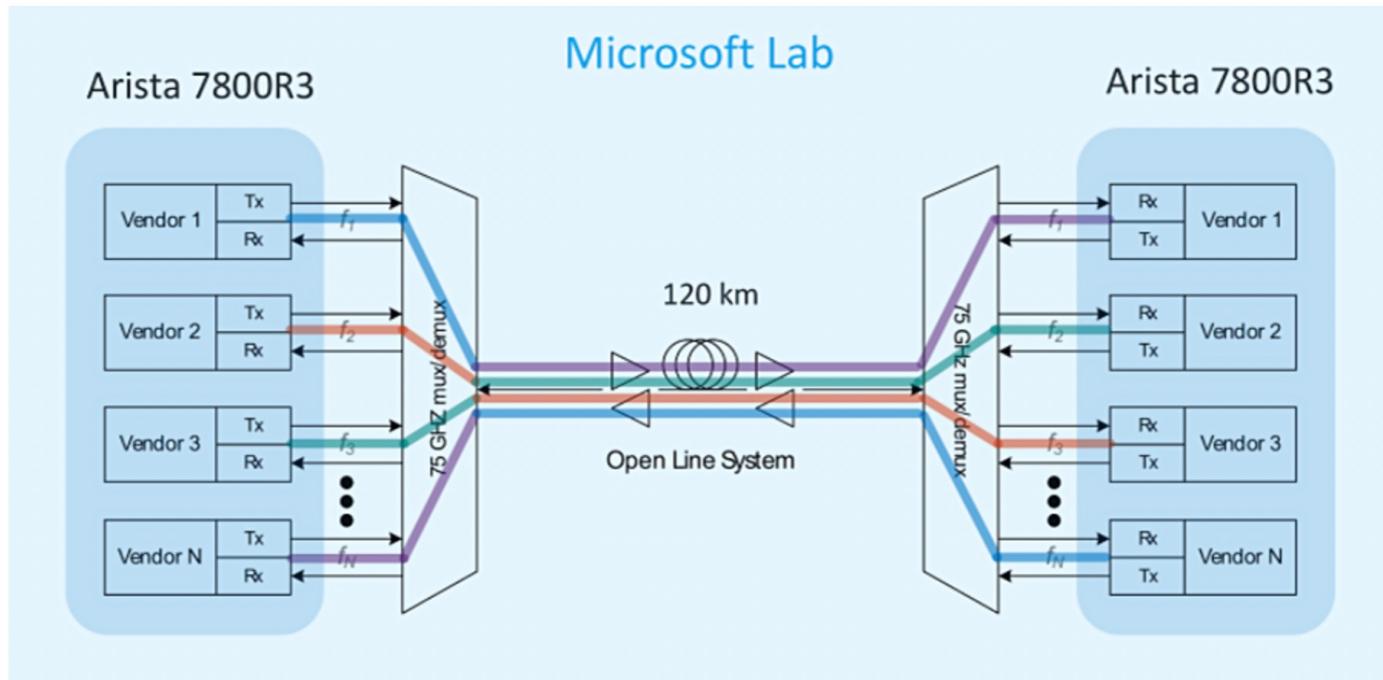
100G-DR/100G-FRモジュール



- 400Gの最大限活かし、既存の100Gシステムを収容可能な100G DR/FRモジュールがある
- ブレークアウトなどに活用
- QSFP28のフォームファクター

アリスタ、400G ZRの相互運用性テストを完了

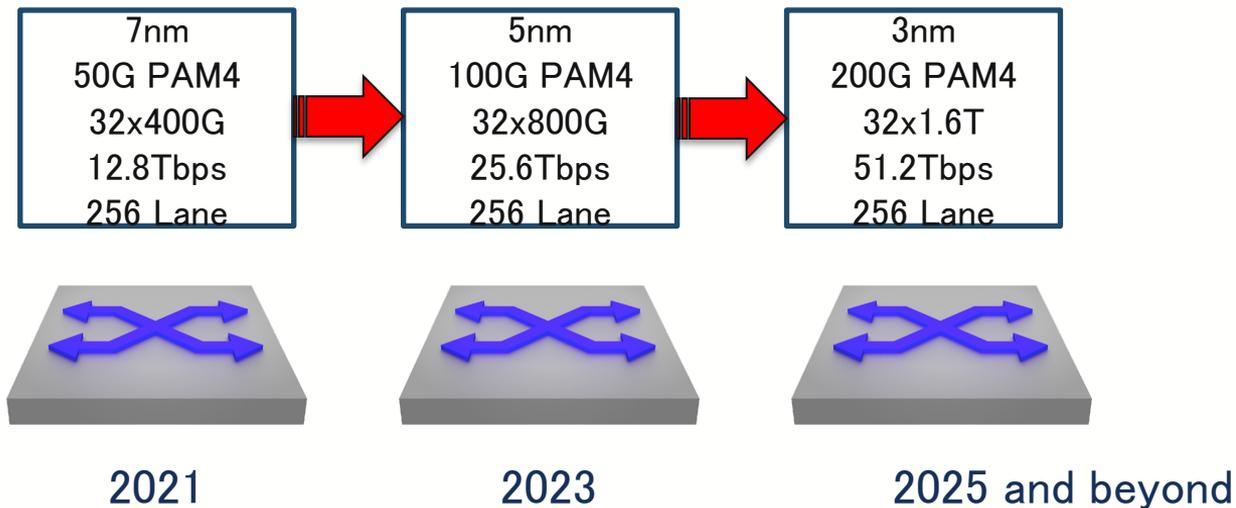
<https://www.arista.com/jp/company/news/press-release/12314-jp-pr-20210217>



Microsoftの120km DCIテストベッドでの400G ZR相互運用性テスト

- 各モジュールを実装し、相互接続試験を実施

商用シリコンロードマップ



- 5nmプロセッサにて100G PAM4 800Gbpsサポート
- 3nmプロセッサにて200G PAM4 1.6Tbpsへ

800Gbps/1.6Tbps Optics MSA

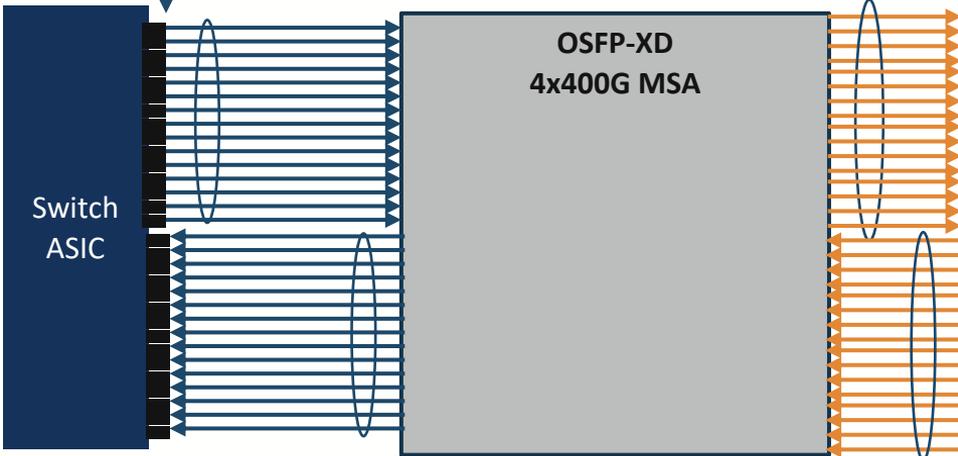
- OSFP MSA ANNOUNCES RELEASE OF OSFP 4.0 SPECIFICATION FOR 800G MODULES WITH
 - THE 800G SPEC COMPLETED, GROUP IS DEVELOPING SPECIFICATION FOR 1600G MODULES
 - <https://osfpmsa.org/press-releases/pr-20210603.html>
- QSFP-DD MSA GROUP ANNOUNCES NEW SPECIFICATIONS AND WHITEPAPER
 - <http://www.qsfp-dd.com/qsfp-dd-msa-group-announces-new-specifications-and-whitepaper/>
- Photonics Technology Leaders Join Forces to Form the 4x400G MSA
 - <https://4x400g.com/>

800Gオプティクスモジュール

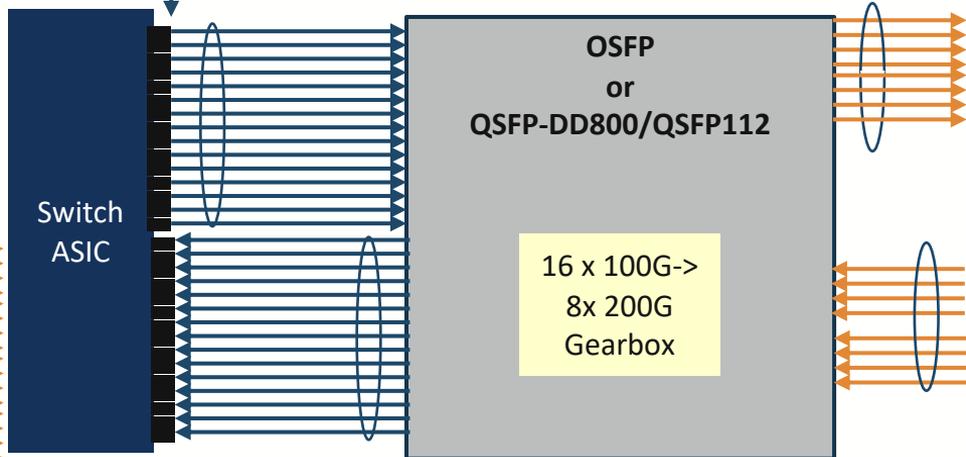
名称	モジュレーション	距離	コネクタ
800G-ZR	16QAM	100km	LC
800G-LR	16QAM	1-10km	LC
800G-FR4	200G-PAM4	1-2km	LC
800G-DR4	200G-PAM4	1-2km	QuadSN/MDC
800G-FR8/LR8	100G-PAM4	2km/10km	LC
800G-2FR4/2LR4	100G-PAM4	2km/10km	DualLC
800G-DR8	100G-PAM4	2km/10km	MPO/8xSN/MDC
800G-SR8	100G-PAM4	50m	MPO/8xSN/MDC

1.6Tbpsロードマップ

電気信号: 16 x 100G PAM-4



電気信号: 16 x 100G PAM-4



電気信号: 8 x 200G PAM-4

