

---

# イーサネットの最新動向

-100Gbイーサネット 他-

---

日立製作所 中央研究所

西村信治

# 目次

1. IEEE802.3標準
2. 次世代イーサネットのマーケット
3. 従来のイーサネット標準
4. 100Gbイーサネットへの提案技術
5. 今後の焦点

---

# 1. IEEE802.3標準

1.1 歴史

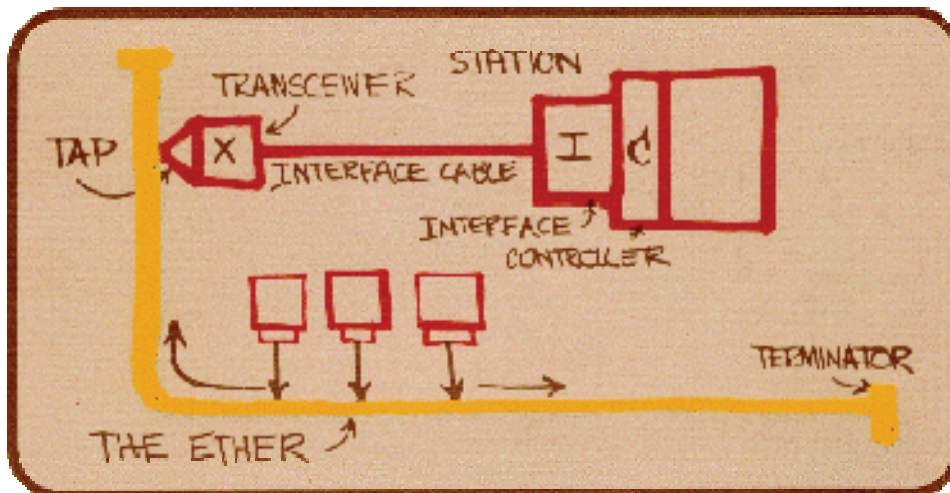
1.2 全体像

1.3 標準化ポイント(Criteria)

1.4 進行ルール

1.5 投票権

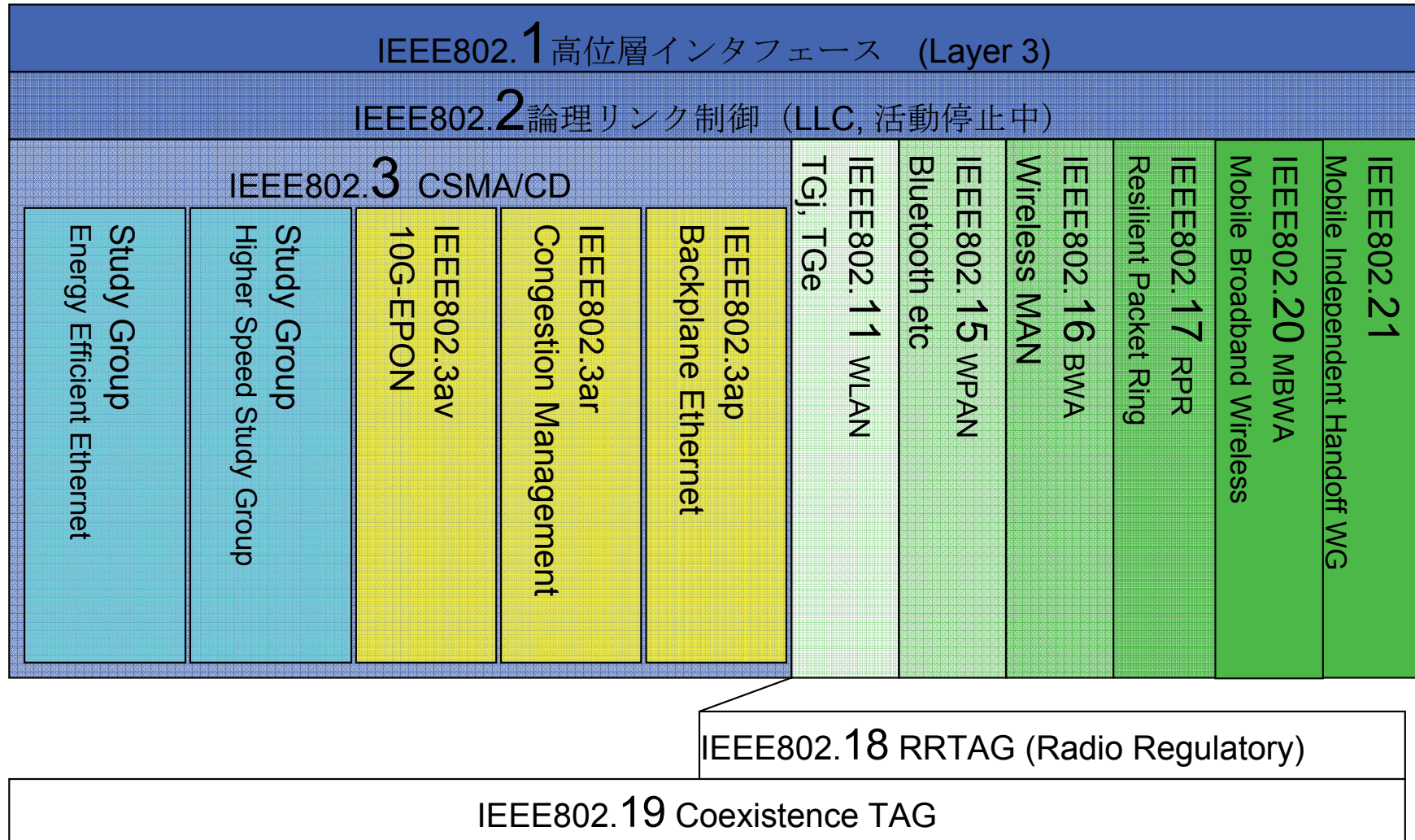
# 1.1 Ethernetの歴史



- 1976年6月の学会発表スライド
- 論文: **Ethernet: Distributed Packet Switching for Local Computer Networks**, *Robert M. Metcalfe and David R. Boggs*, Xerox Palo Alto Research Center

- 1973年 Robert M. Metcalfe(メトカフ)が発明  
Xerox Palo Alto Research Center
- 1979年 DIX規格製品化  
DEC: コンピュータ、Intel: 半導体、Xerox: 特許
- 1980年2月(802) IEEE802委員会発足、LAN標準化着手
- 1983年 IEEE802.3 **CSMA/CD**規格化
- 1992年 Fast Ethernet(100Mbps)標準化着手
- 以下、ほぼx10/4年でより高速なイーサネットが提案され続けている

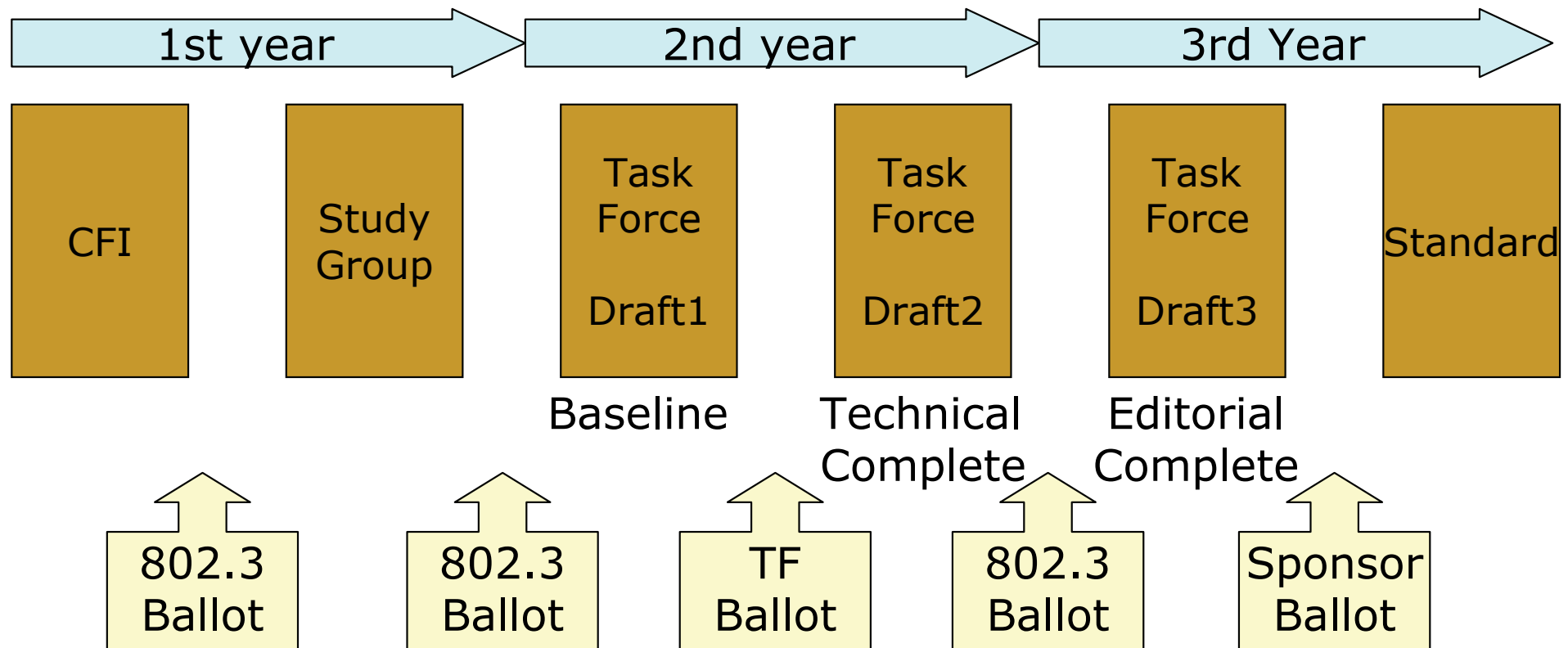
# 1.2 802階層 (<http://www.802.org>)



TAG: Technical Architecture Group

# 1.3 標準化プロセス

- およそ3年で標準化を完成させる
- 技術は、Draft2決定までの2年で決まる(製品化着手)
- 規格となるには >75%の賛成が必要



# 1.4 IEEE802.3標準のポイント

## ■ 5 Criteria

### □ Broad market potential

- Broad sets of applications
- Multiple vendors and numerous users
- Balanced cost (LAN versus attached stations)

### □ Compatibility

### □ Distinct Identity

### □ Technical feasibility

- Demonstrated system feasibility
- Proven technology, reasonable testing
- Confidence in reliability

### □ Economical feasibility

- Known cost factors, reliable data
- Reasonable cost for performance
- Consideration of installation costs

# 1.5 標準化参加資格(802.3Voter)

- 802.3Voterとは
  - 個人資格(会議参加は自由)
  - 最近4回の会議中2回に75%の出席が条件
    - 各会議冒頭のセッションで、チェアに認定を受ける
    - 出席率が未達となった時点で、資格剥奪
  - 合計200名
- 802.3Voterの権利
  - 802.3 WG Ballotでの投票資格



---

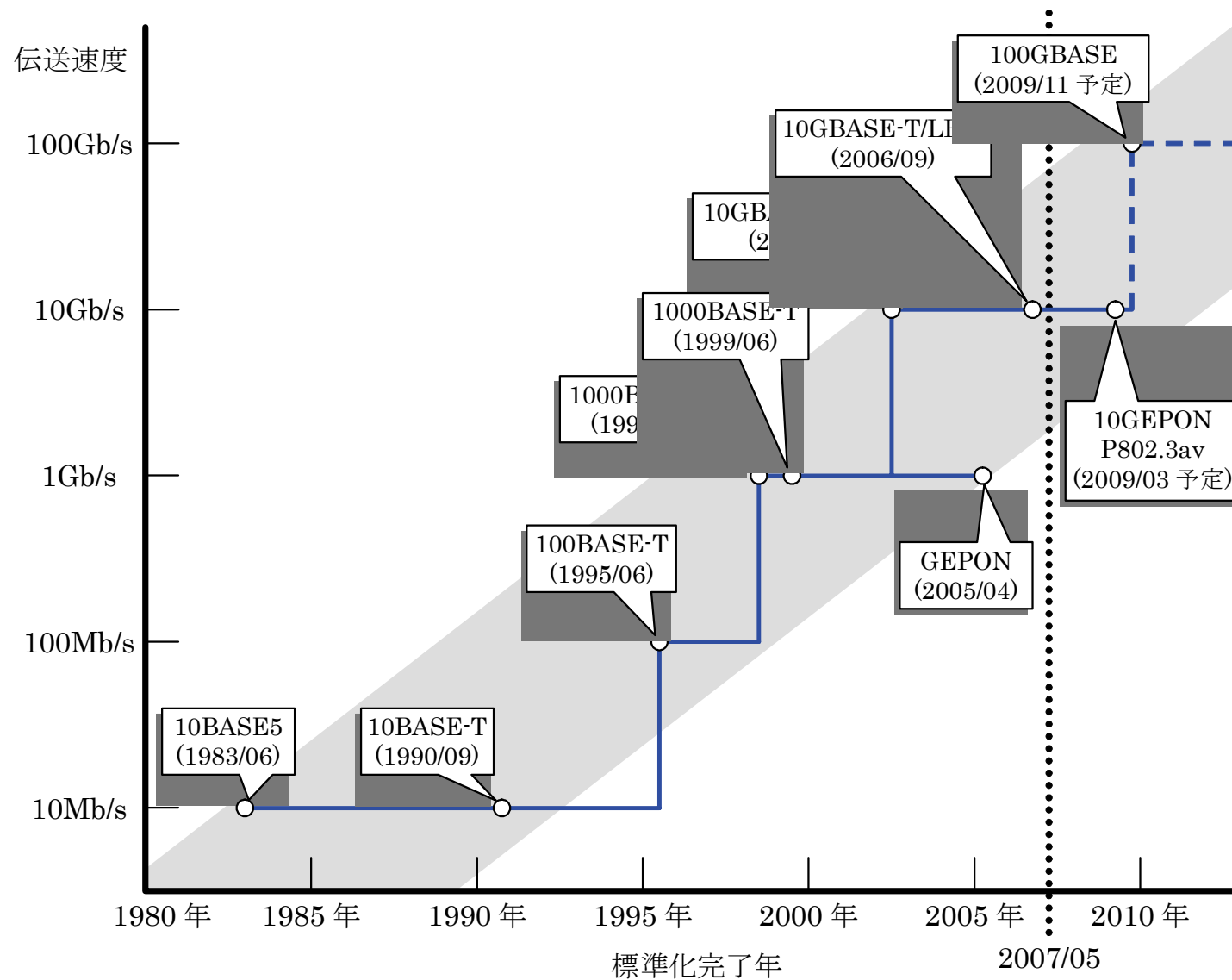
## 2. 次世代イーサネットのマーケット

### 2.1 イーサネットトレンド

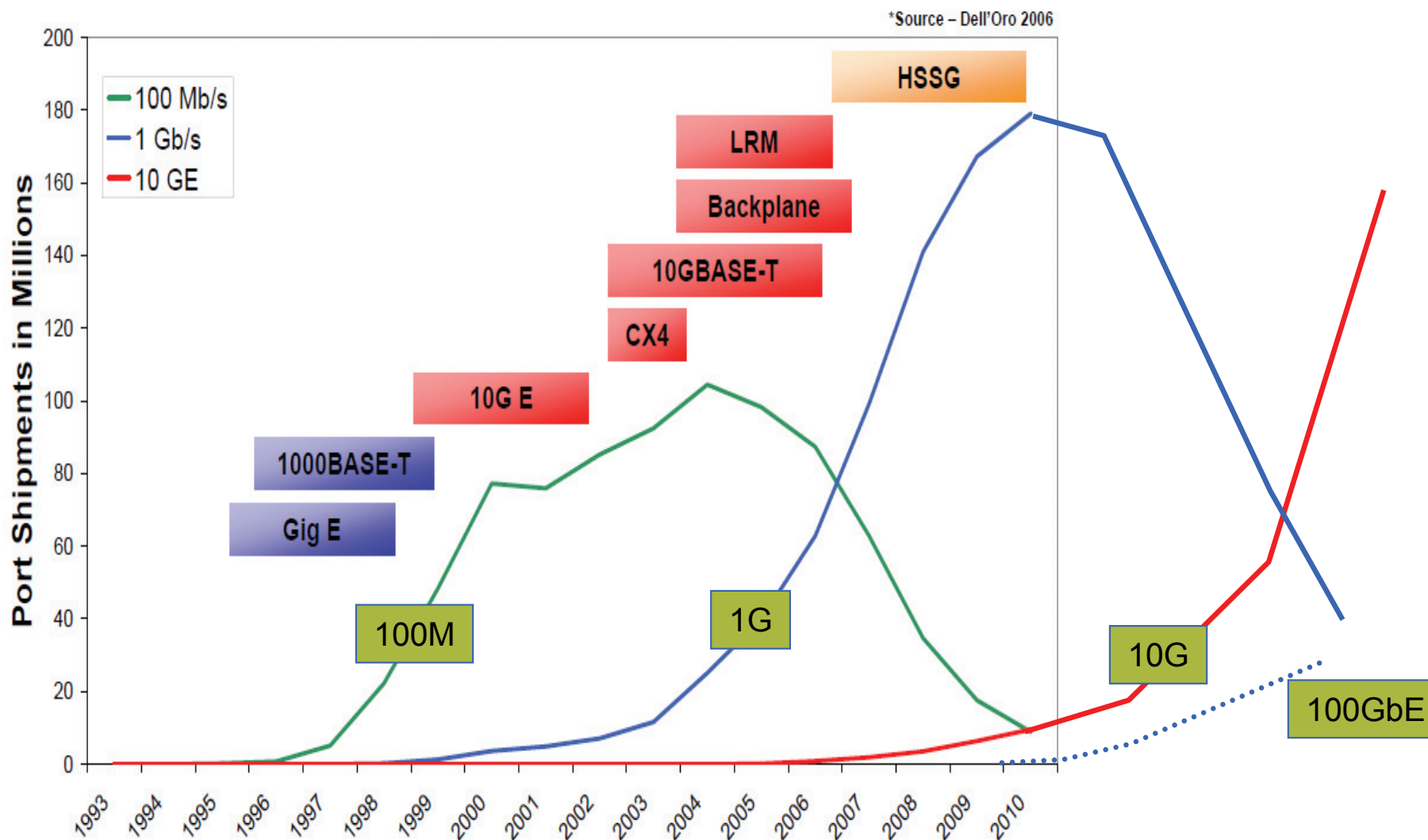
### 2.2 ニーズと想定使用方法

- (1) バックボーンネットワーク
- (2) ビデオ配信
- (3) データセンタ/サーバセンタ
- (4) High-performance Computing

## 2.1 トレンド(速度)



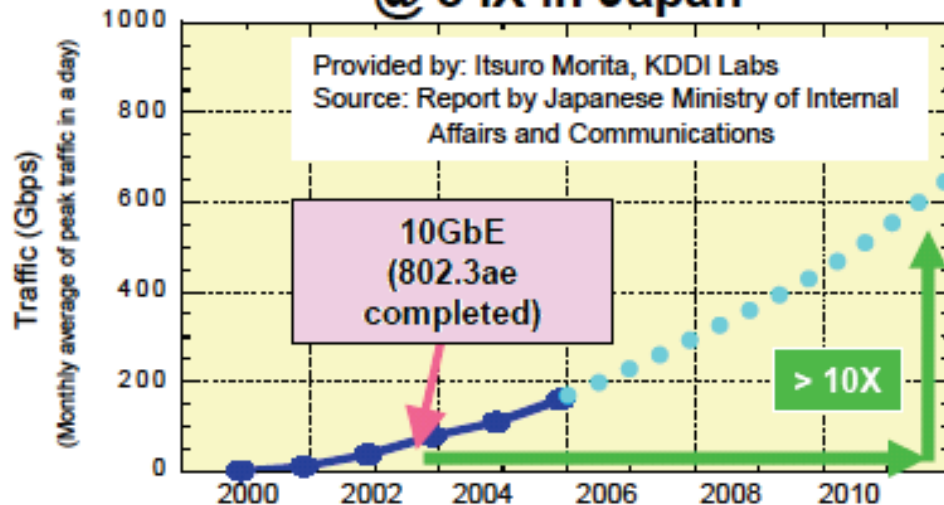
## 2.1 トレンド(出荷ポート数)



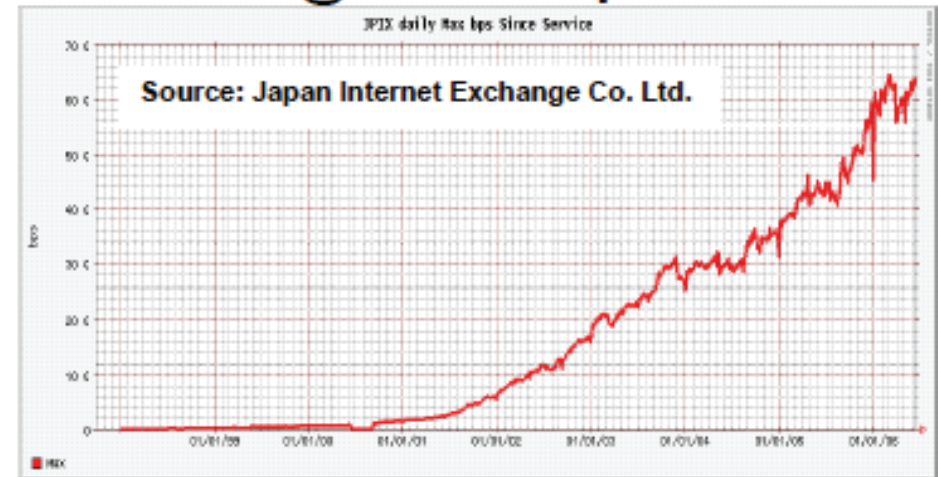
[http://www.ieee802.org/3/cfi/0706\\_1/CFI\\_01\\_0706.pdf](http://www.ieee802.org/3/cfi/0706_1/CFI_01_0706.pdf)

## 2.2 バックボーンネットワーク(トラフィック)

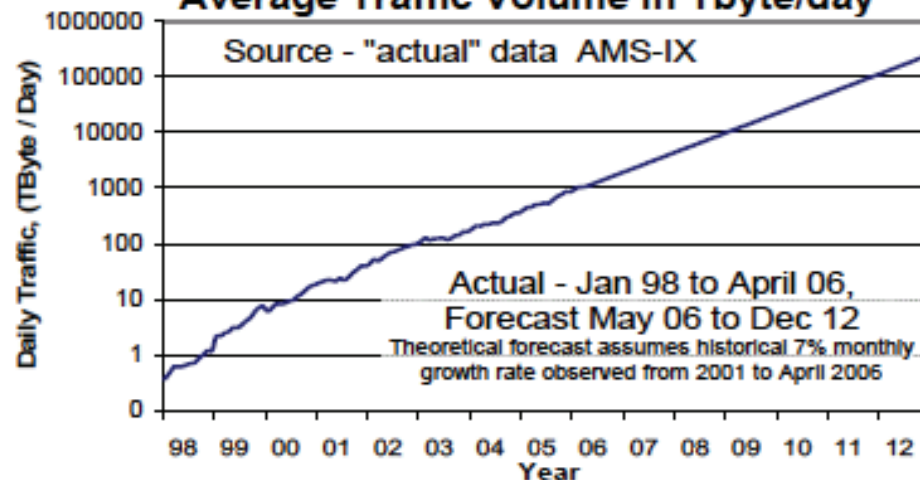
### Monthly Average Peak Traffic @ 3 IX in Japan



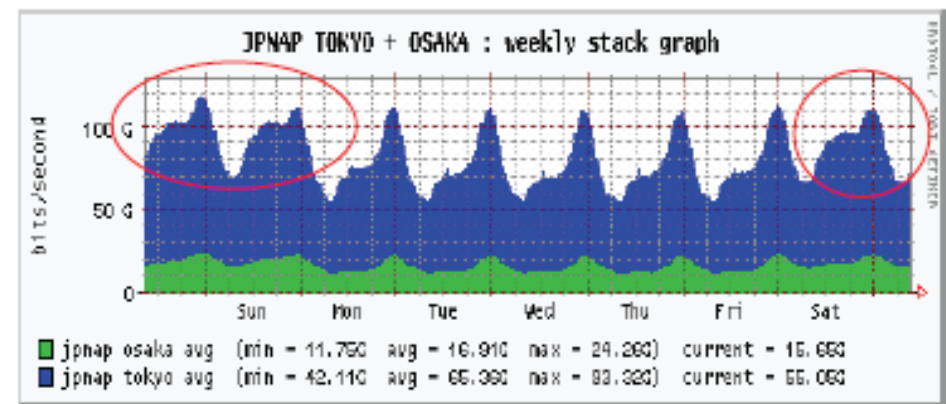
### Daily Maximum Gbps @ JPIX in Japan



### AMS-IX in Amsterdam Average Traffic Volume in Tbyte/day

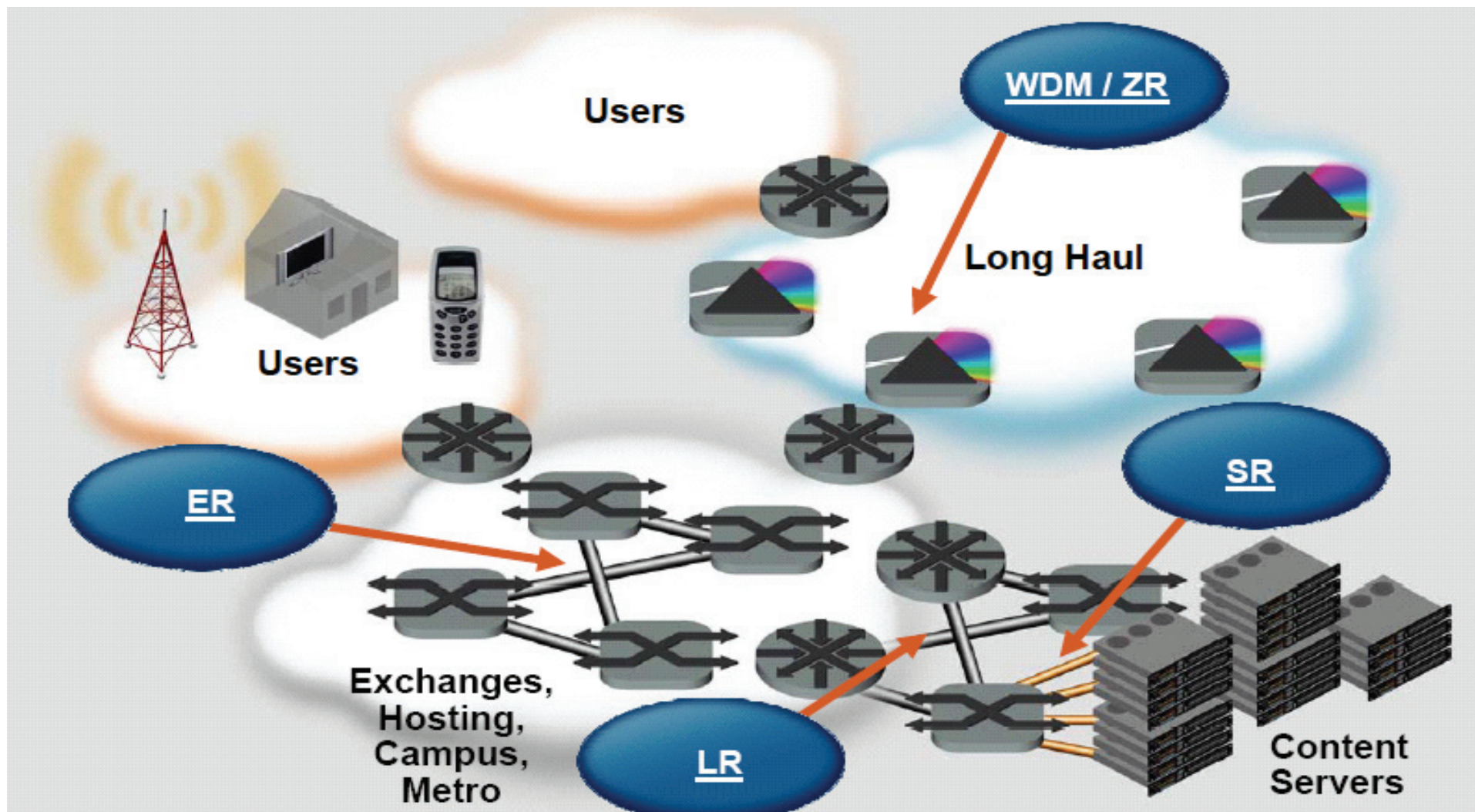


### Daily Traffic @ JPNAP in Japan



Average traffic growth rate since 2003 6% per month  
Used with permission from Internet Multifed Co.

## 2.2 バックボーンネットワーク(アーキテクチャ)



[http://www.ieee802.org/3/hssg/public/mar07/goergen\\_02\\_0307.pdf](http://www.ieee802.org/3/hssg/public/mar07/goergen_02_0307.pdf)

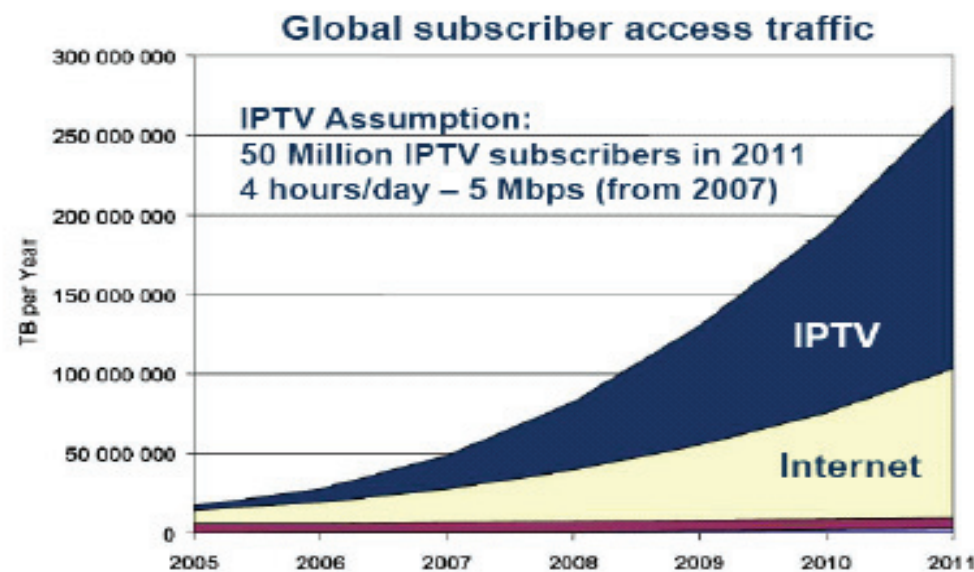
## 2.2 アクセス網高速化

- **Applications need 50+ Mbps per home**
  - File Sharing / Peer to Peer
  - IPTV / Video on Demand / High Definition TV
  - Gaming
- **Broadband access**
  - 50 / 100 Mbps typical in Japan and Korea
  - US / Europe - 1 to 30 Mbps
    - Verizon fiber deployment for FIOS
    - Comcast to trial 100 Mbps service in 2006 (DOCSIS 3.0)
    - Centillium Communications announces VDSL2 OFFERING 100 Mbps symmetric data rates
    - Utah's Utopia in June's Spectrum
      - 100M point-to-point network to a potential 2.5M customers
      - “easily” upgradeable to 1G
    - IEEE 10GEPON Study Group

## 2.2 ビデオ配信

### Broadband access

Fixed access traffic growth



Access traffic growth through increased:

- broadband penetration
- bandwidth demanding services

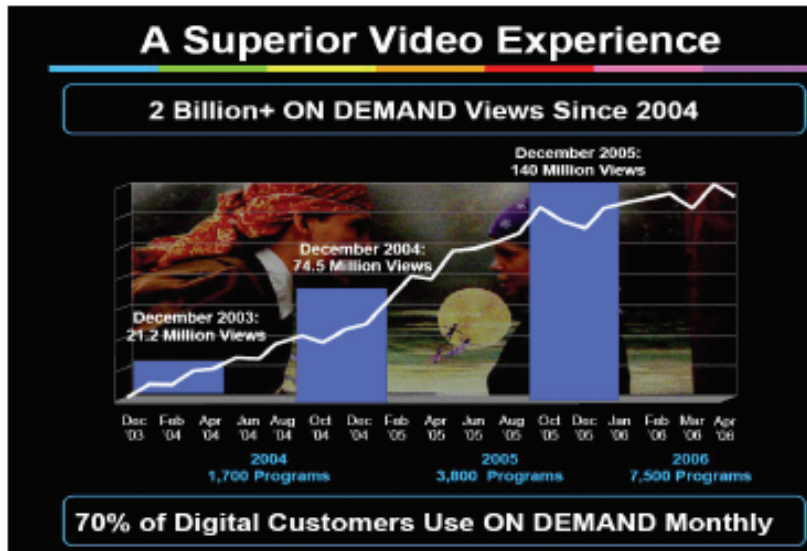
Service	Bandwidth
VoIP	100 kbps
IMS/Video conf	0.7 - 1.5 Mbps
Internet	0.2 - 5.0 Mbps
Gaming	0.2 - 0.5 Mbps
SDTV (MPEG-2)	6 Mbps
SDTV (MPEG-4)	3 Mbps
HDTV (MPEG-2)	20 Mbps
HDTV (MPEG-4)	10 Mbps

Typical service portfolio (2010):

1 HDTV, 2 SDTV, gaming, voice, high-speed internet → 25 – 30 Mbps (MPEG-4)

Reference: Alping\_01\_1106.pdf

## 2.2 ビデオセンタ



Used with permission from Comcast

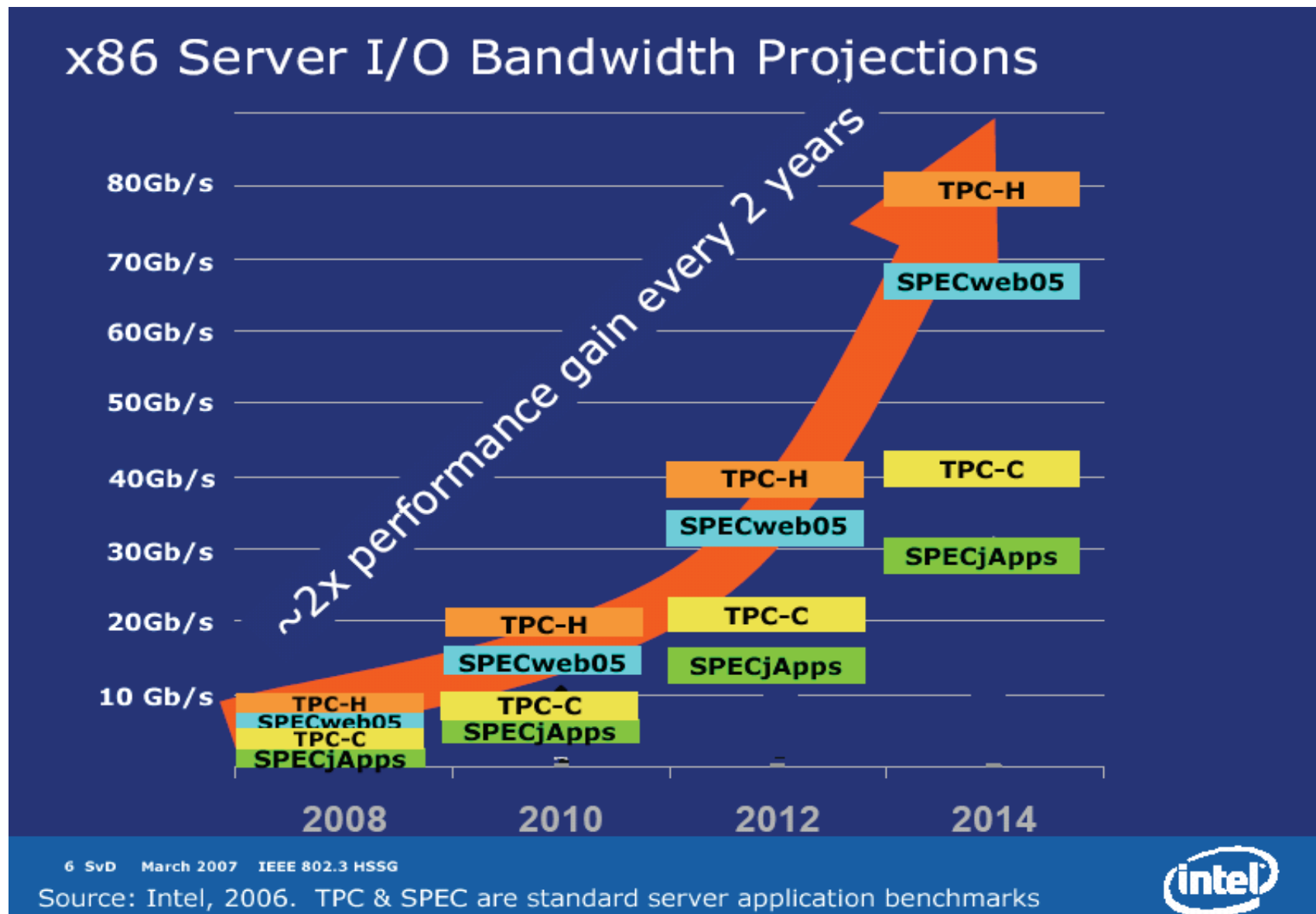
- **Comcast – customer profile**
  - 10M Digital subscribers - 47% of customer base
  - 70% Digital customers use Video on Demand
  - Only 28% are getting High Definition programming

Source: Ian Cox, Cisco Systems / Comcast

- **Network profile**
  - Driving bandwidth factor is Video on Demand unicast traffic
    - Standard Definition (SD): 3.5Mbps
    - High Definition (HD): 19 Mbps
  - Regional networks service a metro area
  - One 10 GbE link can only deliver 2500 SD streams
  - Sections of the regional network use up to 7 x 10 GbE links today
    - 1 to 2 bi-directional links for Voice, Broadcast TV, Internet
    - 2 to 5 uni-directional links Video on Demand
- **Personalized Content – killer app!**
  - Growth potential for digital subscribers
  - Growth potential for HD programming
  - Bandwidth requirements of HD



## 2.2 サーバインタフェース



[http://www.ieee802.org/3/hssg/public/mar07/vandoorn\\_01\\_0307.pdf](http://www.ieee802.org/3/hssg/public/mar07/vandoorn_01_0307.pdf)

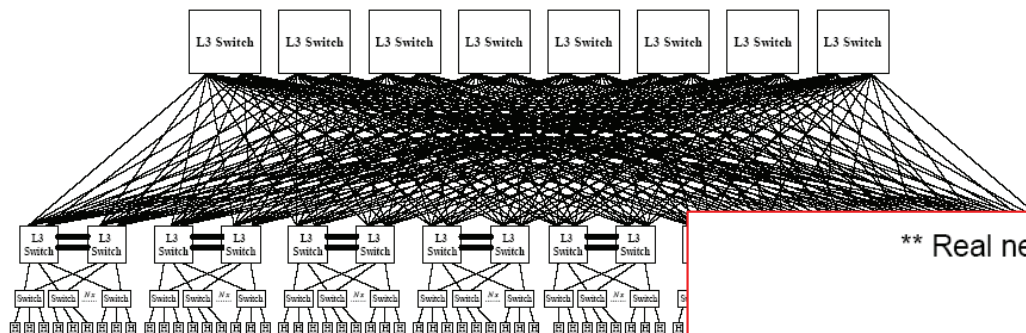
# 2.2 データセンタ/サーバセンタ

I'm currently building this...

\*\* 8 way ECMP w/ 2x10GE LAGs\*\*

\*\* Way too many paths \*\*

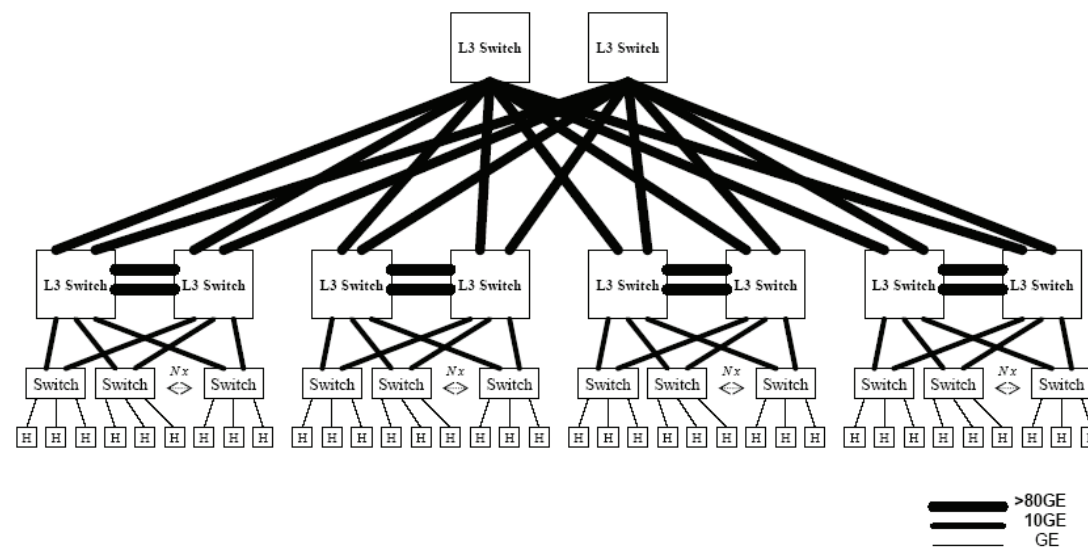
\*\* Way too many cables \*\*



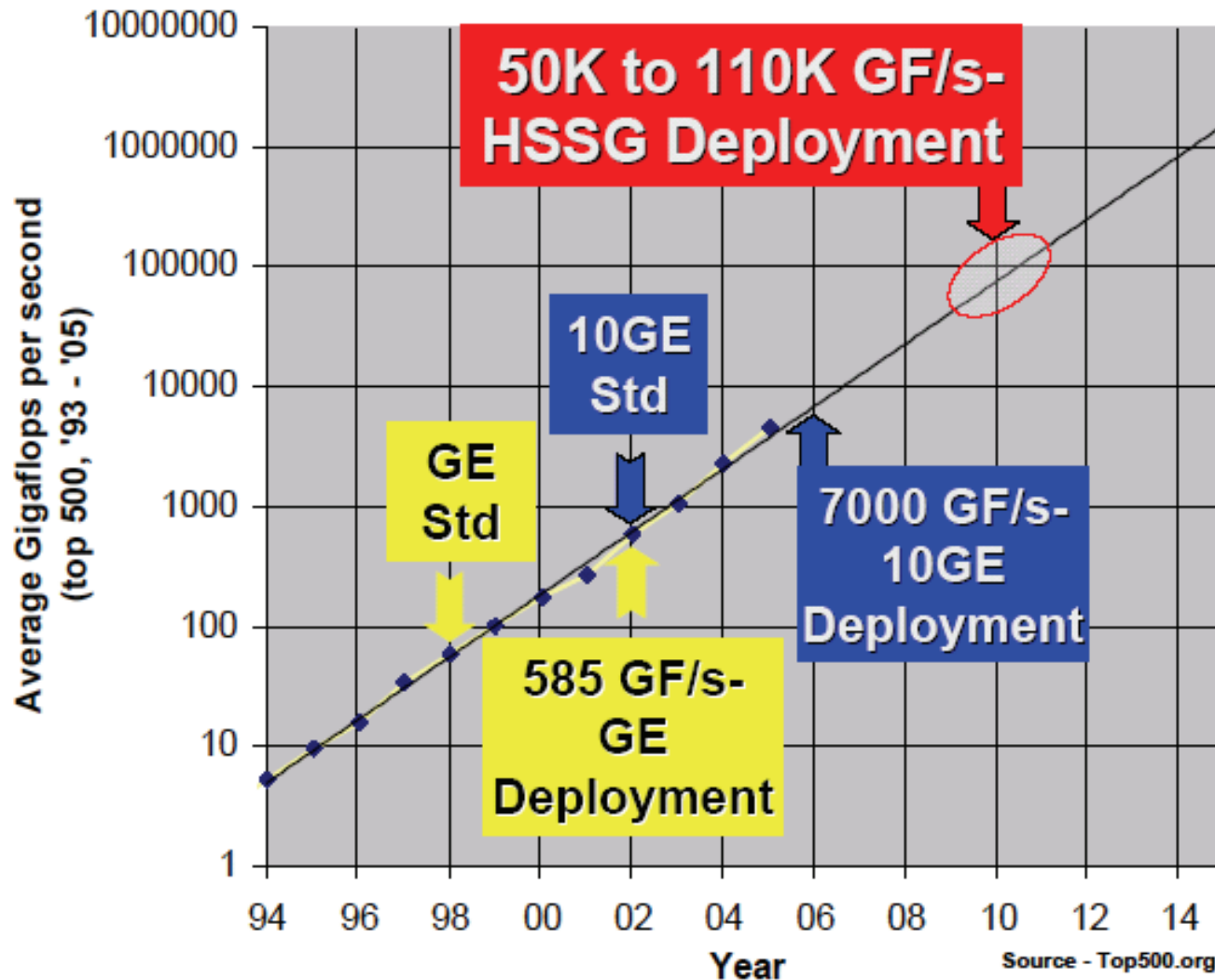
\*\* Real need for at least 80GE today (ideally more) \*\*

\*\* Better mix of link speeds \*\*

\*\* Keep ECMP to 2 paths \*\*



## 2.2 HPC



### Historical Observations

- 4 years between standard and deployment
- GE captures 50% of market in 3 years (2002 - 2005)
- 10GE enters top500 in 2006 (included in Gigabit Ethernet #'s)
- 12x increase in Avg Gigaflops: 10x increase in Ethernet Interconnect

### Forecast

- 10x increase needed at 84,000 Gigaflops (2010)
- There will always be early adopters.

## 3. 従来/進行中の10Gイーサネット標準

3.1 10Gbイーサネット(xR,xW)

3.2 10GBASE-T

3.3 10GBASE-LRM

3.4 バックプレーンイーサ

3.5 10G-EPON

# 3.1 10Gb Ethernet規格

LAN	WAN	Media	Link Speed (Gbps)	Power (W, max)	Transmission Length			Standard
					Copper	MMF	SMF	
ZR		1.5um Serial	10.3125	10~3.5	-	-	80km	業界規格
ER	EW	1.5um serial	10.3125	10~3.5	-	-	40km	Finish
LR	LW	1.3um serial	10.3125	6~2.5	-	-	10km	Finish
LX4	-	1.3um 4-WWDM	3.125	6~4	-	300m (50um)	10km	Finish
LRM		1.3um serial	10.3125	6~2.5	-	220m (62.5um)	-	Finish
SR	SW	0.85um serial	10.3125	6~1.5	-	28m (62.5um)	-	Finish
<b>10GE-PON</b>		<b>TBD</b>	<b>10</b>	<b>TBD</b>		-	<b>10km 20km</b>	<b>2006/Mar. START</b>
<b>T</b>		<b>Twist pair</b>	<b>(0.625)</b>	<b>10?</b>	<b>100m</b>	-	-	<b>Finish</b>
<b>CX4</b>		<b>Shielded</b>	<b>3.125</b>	<b>?</b>	<b>15m</b>	-	-	<b>Finish</b>
<b>Backplane</b>		<b>FR4-board</b>	<b>10</b>	<b>3?</b>	<b>1m</b>			<b>D3.2</b>

WWDM: Wide wavelength Multiplexing, MMF: Multi-mode Fiber, SMF: Single-mode Fiber

赤字:新規格

## 3.2 10GBASE-T (802.3an)

- ターゲット：  
UTPケーブル(Cat6,7)を用いた100m低コスト10Gbps伝送

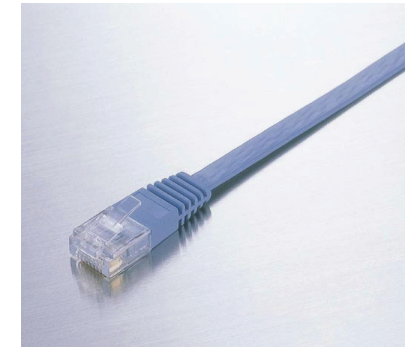
- 規格内容

	名称	伝送容量	伝送速度	並列数	距離
1	10GBASE-T	10Gbps	1Gbps(PAM12)	4	CAT7 100m
2		10Gbps	1Gbps(PAM12)	4	CAT6 65m

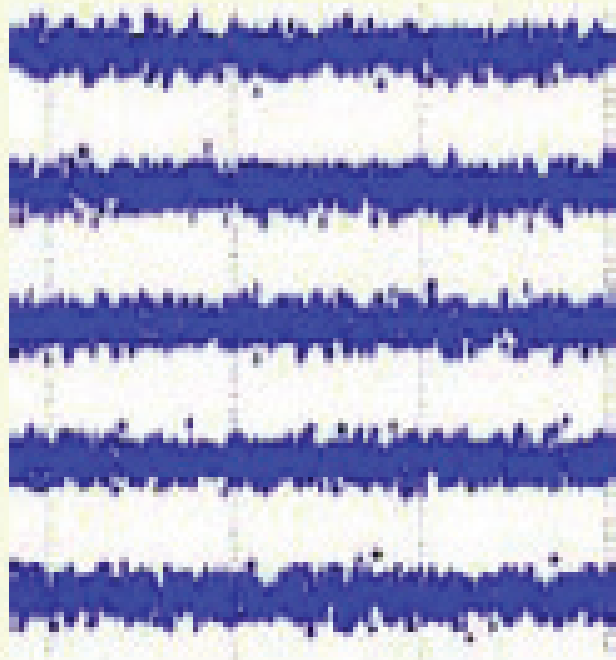
- スケジュール：

- 開始： 2003年1月
- D1.0： 2004年9月
- D2.0： 2005年4月
- 終了： 2006年7月

- 現在の焦点： ほぼ収束(PCS,PMAの細かい修正中)消費電力が課題

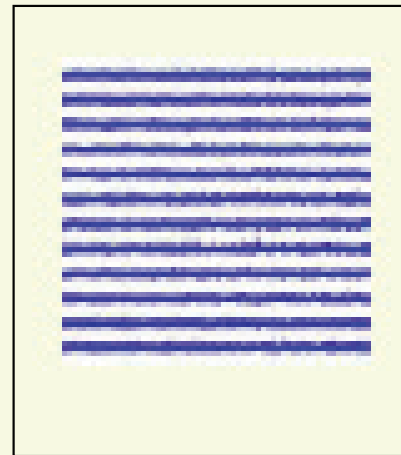


## 3.2 10GBASE-T: 多値伝送方式



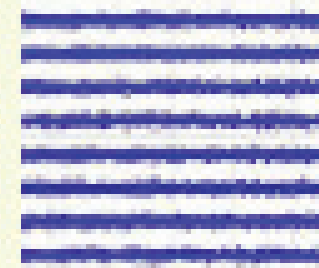
PAM-5 eye diagram at the receiver for 1000BASE-T with 6dB coding gain

日立, Teranetics  
NEC-Ele, Broadcom  
Key-Eye



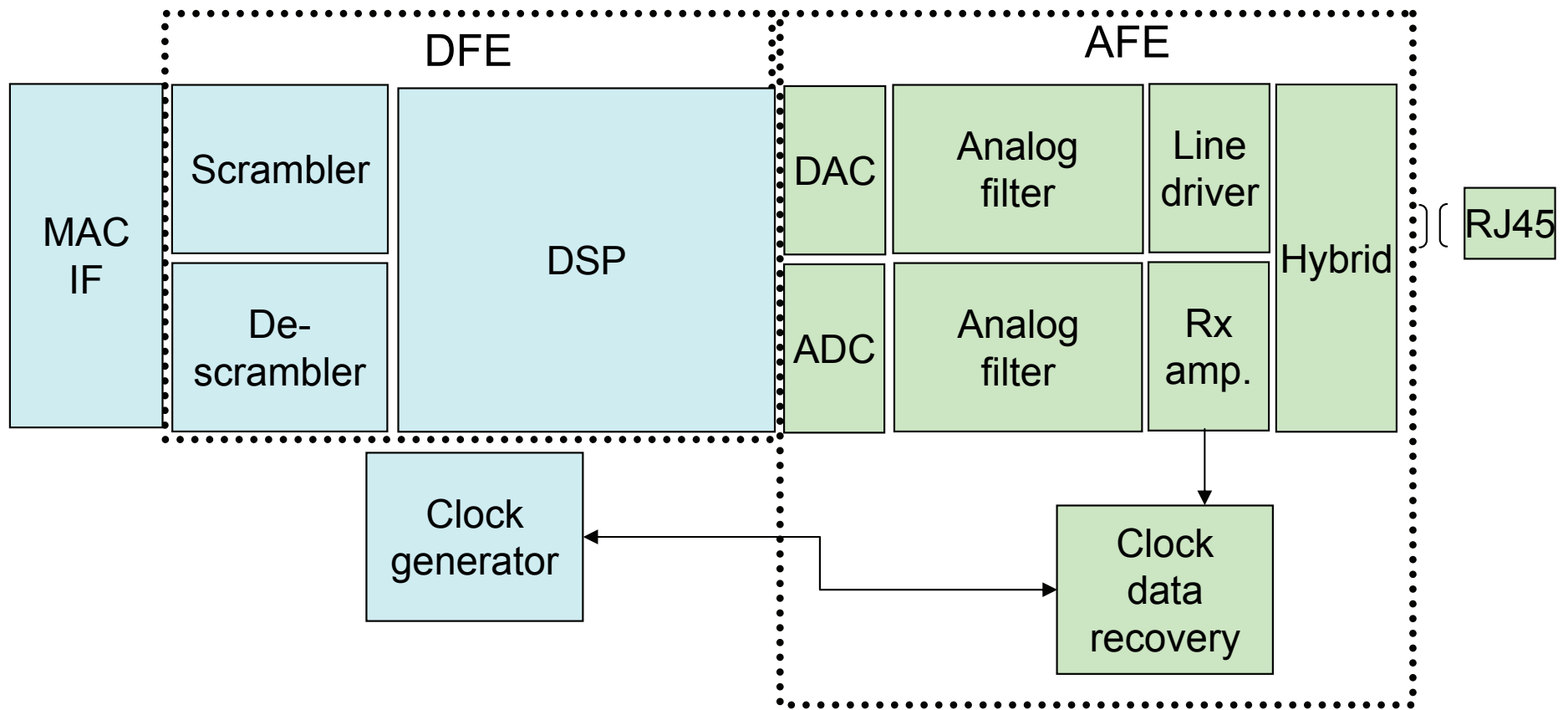
PAM-12 eye diagram at the receiver for 10GBASE-T with 12dB coding gain

Phy Ten



PAM-8 eye diagram at the receiver for 10GBASE-T with 12dB coding gain

# 3.2 10GBASE-T 実用化の焦点



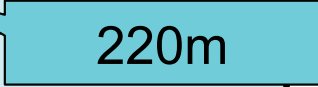
	技術課題	現状スペック	要因
1	ゲート規模	7-8 Mゲート	FEC(LDPC):4M, エコーキャンセル:2M
2	遅延時間 (MAC-I/O起点)	500 ns	FEC(LDPC)処理
3	消費電力	7~8W (90nmプロセス)	FEC(LDPC)回路



## 3.3 10GBASE-LRM(802.3aq)

- ターゲット: FDDIグレードのMMFを用いた300m伝送, DFEを活用  
DFE: Decision Feedback Equalizer
- 規格内容

	使用ファイバ	伝送容量	距離
1	50umコア	10 Gbps	<del>300m</del>
2	62.5umコア	10 Gbps	220m
3	62.5umコア(10GBASE-SR)	10 Gbps	30m



- スケジュール:
  - 開始 2003年11月
  - D1.0 2004年11月
  - D3.1 2006年3月
  - デモ: 2006年5月GW@Interop LasVegas
- 課題: 各社間相互接続と価格(SR以下がターゲット)

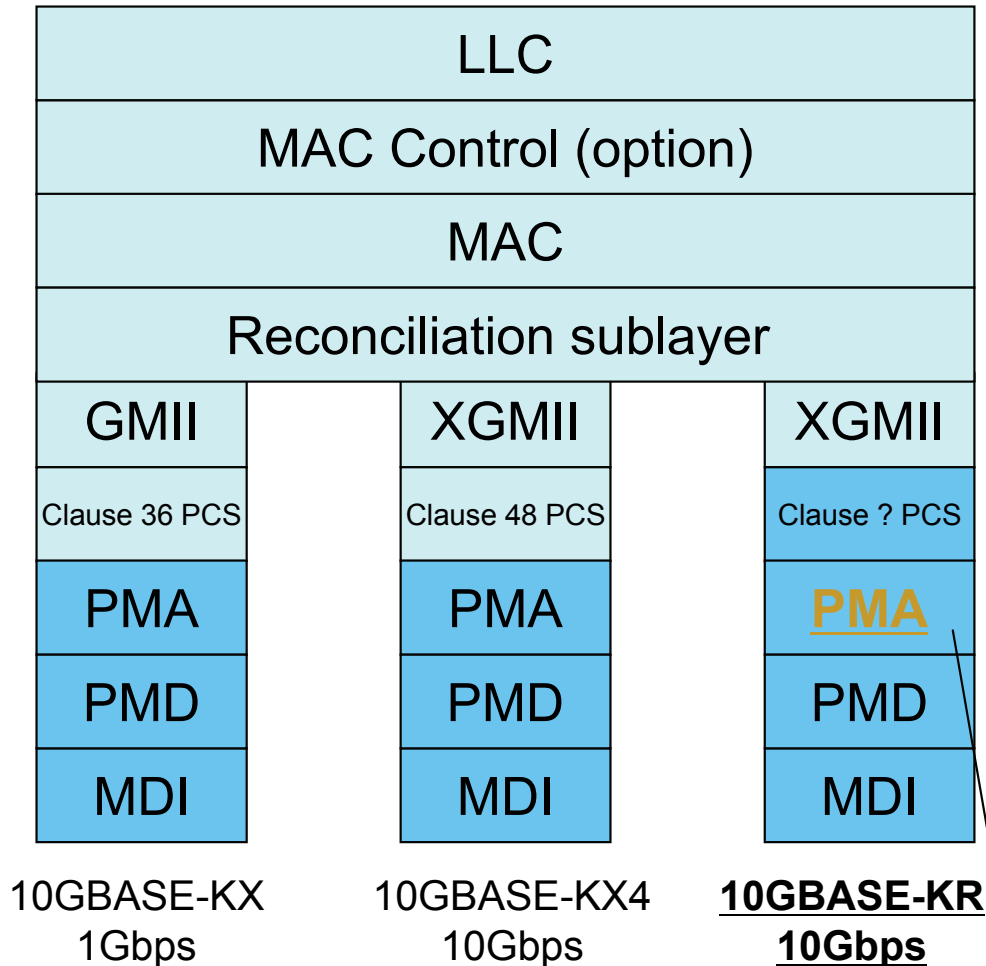
# 3.4 Backplane Ethernet(802.3ap)

- ターゲット:ブレード型装置のバックプレーン10ギガシリアル接続  
(post-PCI-Express)
- 規格内容

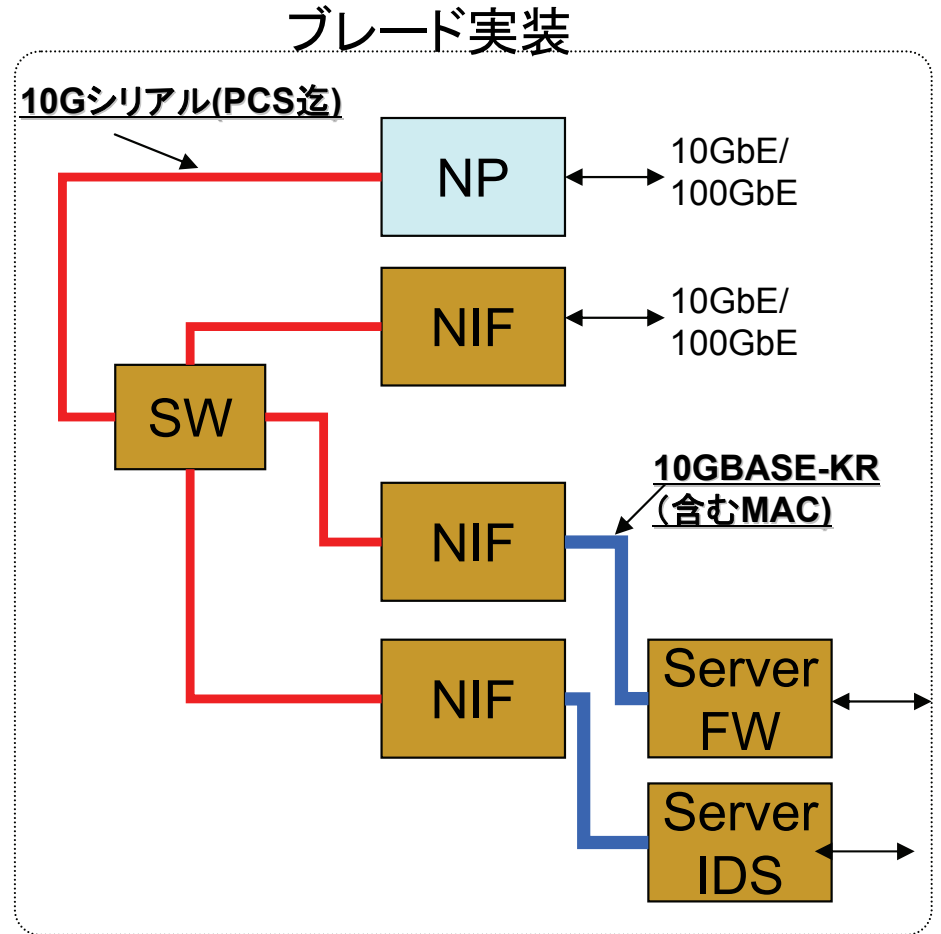
	名称	伝送容量	伝送速度	並列数	距離
1	10GBASE-KX	1 Gbps	1Gbps	1	1m
2	10GBASE-KX4	10 Gbps	3.125 Gbps	4	1m
3	10GBASE-KR	10 Gbps	10 Gbps	1	1m

- スケジュール:
  - 開始 2003年11月
  - D0.7 2004年11月
  - D0.8 2004年2月(投票2/16開始)
  - D1.0 2005年3月
  - D2.0 2005年7月
  - 終了 2007年7月

# 3.4 10Gb Backplane レイヤ構造



規格化箇所



ルータ/スイッチでの  
活用方法案

焦点

## 3.5 10G-EPON

- ターゲット: G-EPONの10Gbps版規格

	名称	上り容量	下り容量	並列数	距離
1	10G/1G	1 Gbps	10Gbps	1	10km/20km
2	10G/10G	10 Gbps	10 Gbps	1/4-CWDM	10km/20km

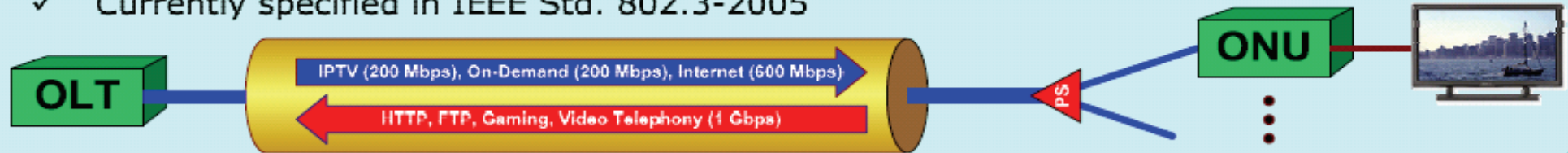
検討事項: G-EPON, VIDEO信号(1550-1560nm)とファイバ中で共存

- スケジュール:
  - 開始 2006年3月
  - Task Force設立 2006年7月
  - D1.0 2007年3月(予定)
  - D2.0 2008年3月(予定)
  - 終了 2009年3月(予定)
- 焦点: 10Gバースト信号検出、信号共存時の波長配分

# 3.5 10GE-PON 構成案

## 802.3ah: 1 Gbps downstream / 1 Gbps upstream

- ✓ The first commercial FTTH technology with Gigabit bandwidth deployed in the world
- ✓ Currently specified in IEEE Std. 802.3-2005



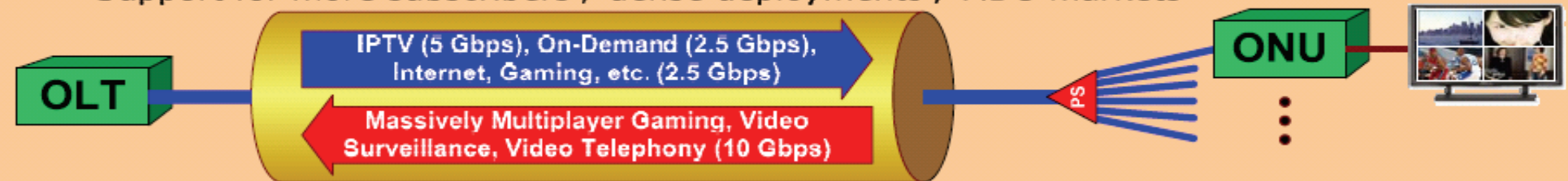
## Option 1: 10 Gbps downstream / 1 Gbps upstream

- ✓ Providing more downstream bandwidth to support advanced digital TV services
- ✓ CATV replacement



## Option 2: 10 Gbps downstream / 10 Gbps upstream

- ✓ Support for advanced, bandwidth-intensive upstream and downstream services
- ✓ Support for more subscribers / dense deployments / MDU markets



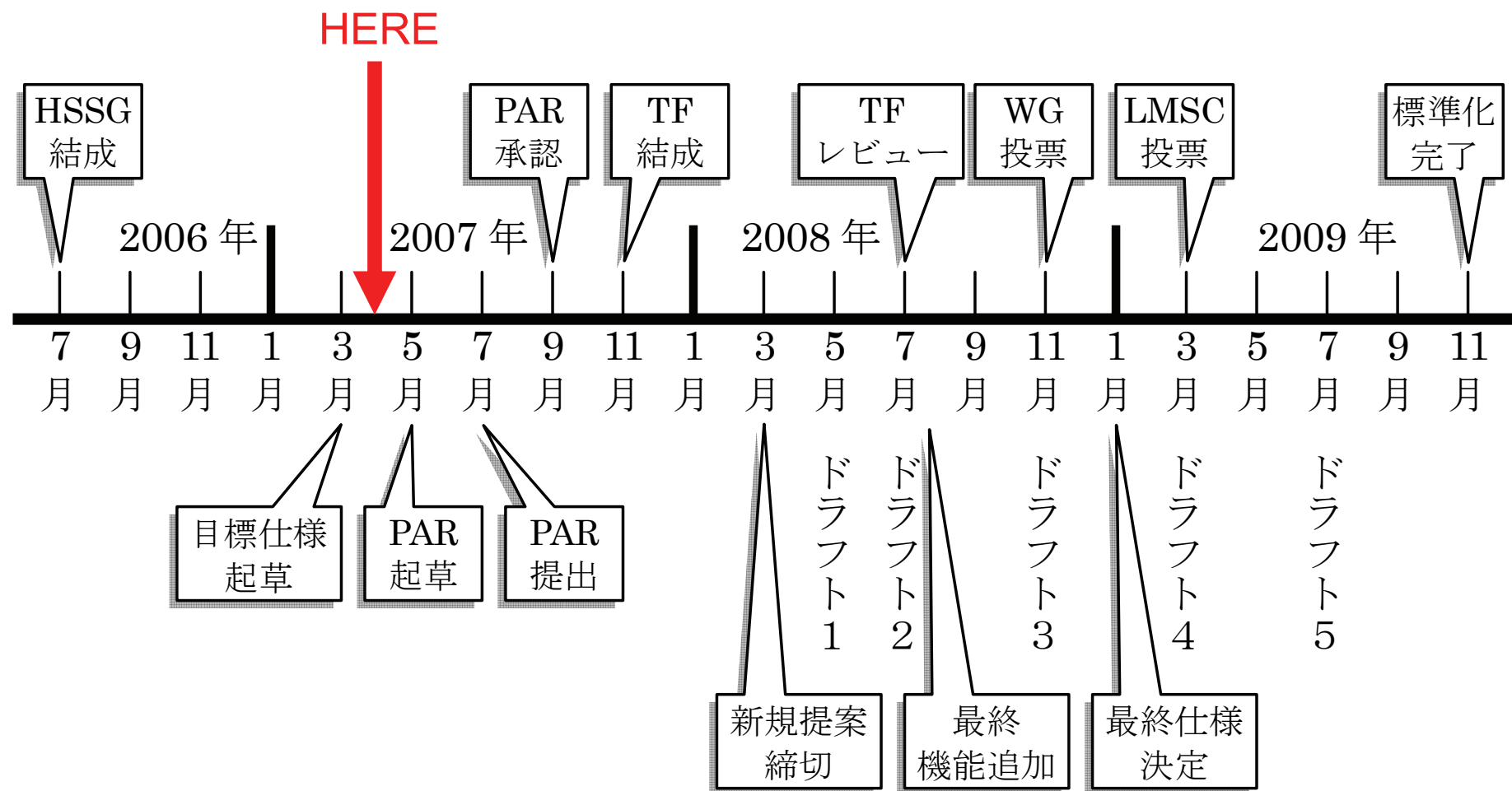
## 4. 100Gbイーサネットへの提案技術

- 4.1 規格のカバー範囲(Objectives),スケジュール
- 4.2 全体アーキテクチャ
- 4.3 光インタフェース (パラレル、シリアル)
- 4.4 電気インタフェース
- 4.5 論理レイヤ

## 4.1 100Gイーサ規格範囲(Objectives)

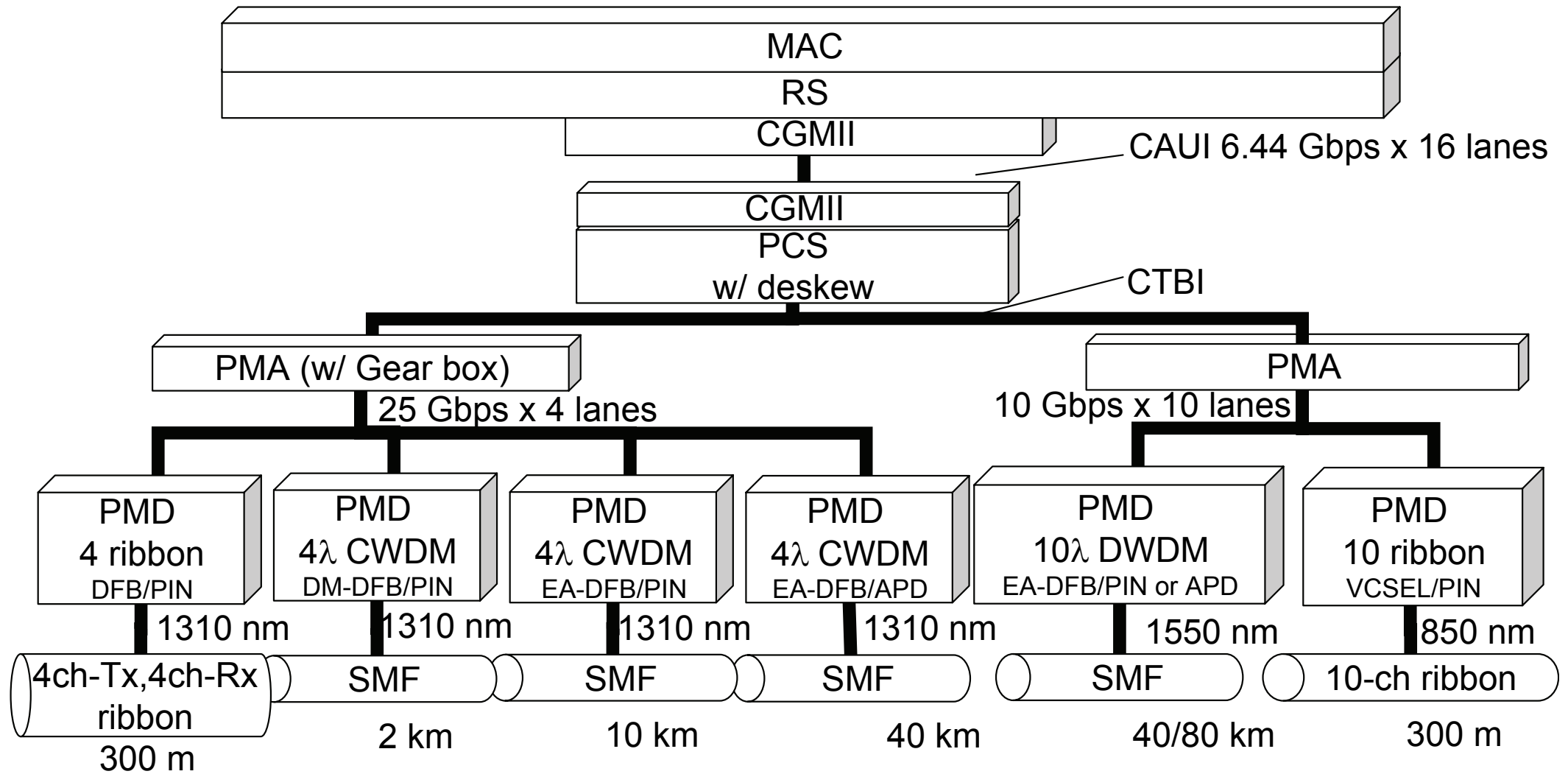
- Support full-duplex operation only
- Preserve the 802.3 / Ethernet frame format at the MAC Client service interface
- Preserve minimum and maximum **Frame Size** of current 802.3 Std
- Support a speed of **100 Gb/s** at the MAC/PLS interface
- Support at least **10km on SMF**.
- Support at least **100 meters on OM3 MMF**.
- Support a **BER better than or equal to  $10^{-12}$**  at the MAC /PLS service interface.

# 4.1 想定スケジュール





## 4.2 全体構造 (日立提案)



## 4.3 光インタフェース

- 要求されるユーザメリット
  1. 技術的実現性
  2. 既存規格との互換性
  3. 将来性
  4. 部品コスト(部品数)
  5. 消費電力
  6. パッケージサイズ

## 4.3 光インタフェース (提案マップ)

### Reach (Technical) Feasibility of 100GE alternatives

SMF	10km 1310nm	40km 1310nm	10km 1550nm	40km 1550nm
10x10G DML	yes (10λ span needs semi-cooling)	yes (need new DML & RX APD/SOA)	yes (need new DML)	maybe (need new DML)
10x10G ML	yes	yes (need RX APD/ SOA)	yes	yes
5x20G / 4x25G DML	yes (need new DML)	maybe (need new DML & RX SOA)	maybe (need new DML)	no
5x20G / 4x25G ML	yes (need new EML)	yes (need new EML & RX SOA)	yes	yes (need RX DC)
2x50G DQPSK ML	yes (need I/Q ML)	yes (need I/Q ML & RX DC & SOA)	yes (need I/Q ML & RX DC)	yes (need I/Q ML & RX DC)
1x100G TDM ML	yes (need new ML & maybe RX DC)	yes (need new ML & RX DC & SOA)	yes (need new ML & RX DC & SOA?)	yes (need I/Q ML & RX DC & SOA?)

Green shading designates alternatives under detailed study by Fiber Optic Ad Hoc contributors.

## 4.3 光インタフェース (提案マップ)

### Cost (1/Economic Feasibility) of 100GE alternatives

SMF	10km 1310nm	40km 1310nm	10km 1550nm	40km 1550nm
10x10G DML	mid	mid	low	mid
10x10G ML	mid	mid	mid	mid
5x20G / 4x25G DML	low	mid	low	not feasible
5x20G / 4x25G ML	mid	mid	mid	not economically feasible (RX DC)
2x50G DQPSK ML	high	not economically feasible (RX DC)	not economically feasible (RX DC)	not economically feasible (RX DC)
1x100G TDM ML	high (RX DC?)	not economically feasible (RX DC)	not economically feasible (RX DC)	not economically feasible (RX DC)

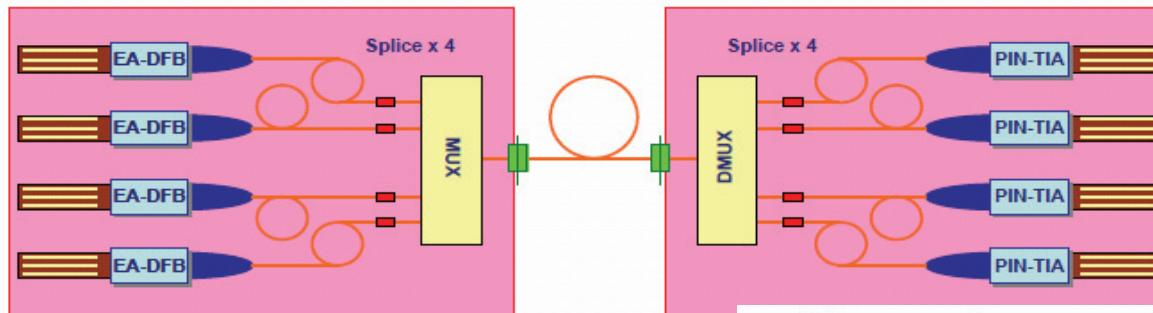
Green shading designates alternatives under detailed study by Fiber Optic Ad Hoc contributors.

[http://www.ieee802.org/3/hssg/public/mar07/dove\\_01\\_0307.pdf](http://www.ieee802.org/3/hssg/public/mar07/dove_01_0307.pdf)

## 4.3 光インタフェース (WDM)

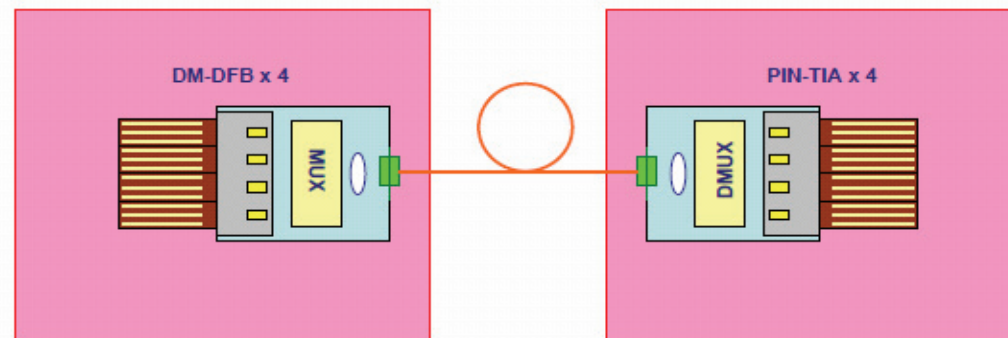
- *The most commercially available solution today and maybe 2009*

- Use of discrete (separate) 1310nm EA-DFB components
- Use of discrete (separate) PIN-TIA components
- Use of Optical MUX/DMUX connected using fiber splicing to optical components.



- *The most aggressive scheme is multi-channel uncooled integrated optical components*

- Challenge (Breakthrough): high power uncooled 1310nm DM-DFB.
- Challenge (Breakthrough): Low cross-talk multi-channel packaging



## 4.3 光インタフェース (25G vs 20G)

		25G	20G
Optical link configuration		25 Gbps x 4 $\lambda$	20 Gbps x 5 $\lambda$
Wavelength pitch		20 nm (*)	
Transmission length		10 km SMF (CWDM, without cooler)	
EA-DFB		Loss limit: 28 km	Loss limit: 27 km
DM-DFB		Dispersion limit: 10 km	Dispersion limit: 9.6 km
Gearbox 10 ch v.s. n-ch		Slightly complex	Easy
Power Consumption	Tx / Rx AC (*1)	1	1
	Tx / Rx DC (*2)	1	1.25
Cost	Total OSA (*3)	1	1.25

(\*1) Assume 10 Gbps AC power = 1.0

→ 4 x 25 Gbps AC power = 2.5: Total AC power 2.5 x 4 = 10.0

→ 5 x 20 Gbps AC power = 2.0: Total AC power 2.0 x 5 = 10.0

(\*2) Laser bias current, EA bias voltage, etc.

(\*3) 20-Gbps laser and 25-Gbps laser are almost same cost.

## 4.3 光インタフェース(リボン)

### Technical Feasibility - 10x10GVCSEL Reliability

#### Wearout Time (Depends on uniformity)

- Perfect uniformity → 12x same as 1x
- Exp. reports: 12x wearout time ~1/2 as for 1x

#### Random Failures

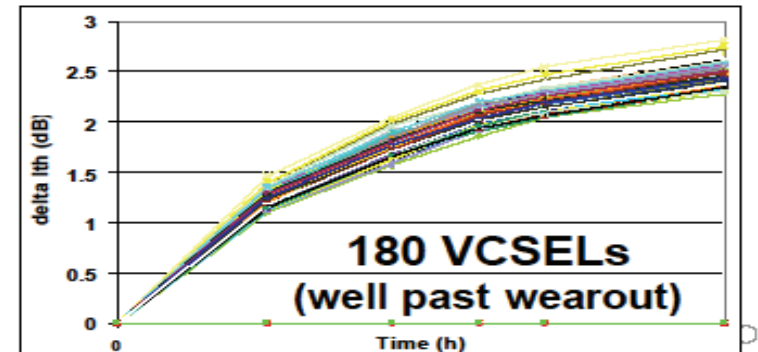
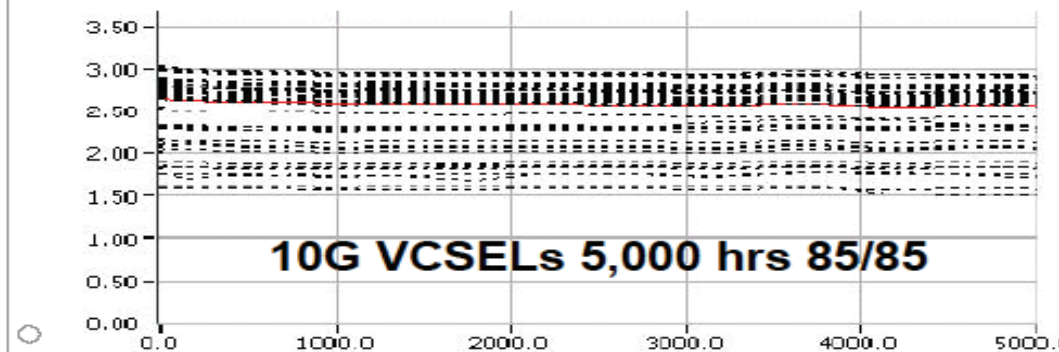
- 12x array failure rate nearly 12 times the 1x failure rate
- Virtually all “random” failures are eliminated through burn-in

#### ESD-Related Failures

- Above-threshold ESD events damage 1x and 12x about equally

#### Non-Hermetic Packaging

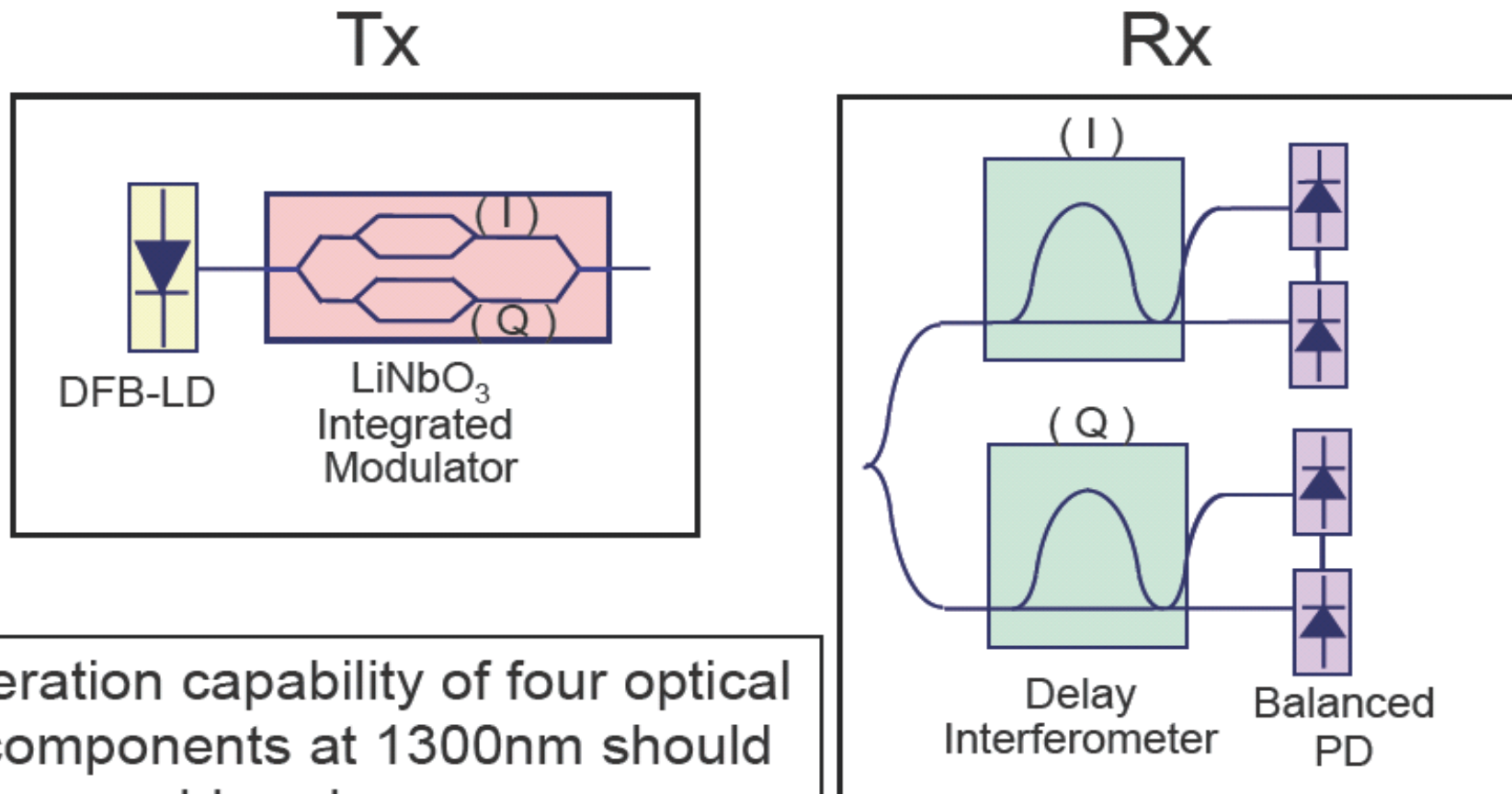
- 12x3G VCSELs robust to harsh environments



4 of 5

Picolight -IEEE HSSG March 2007

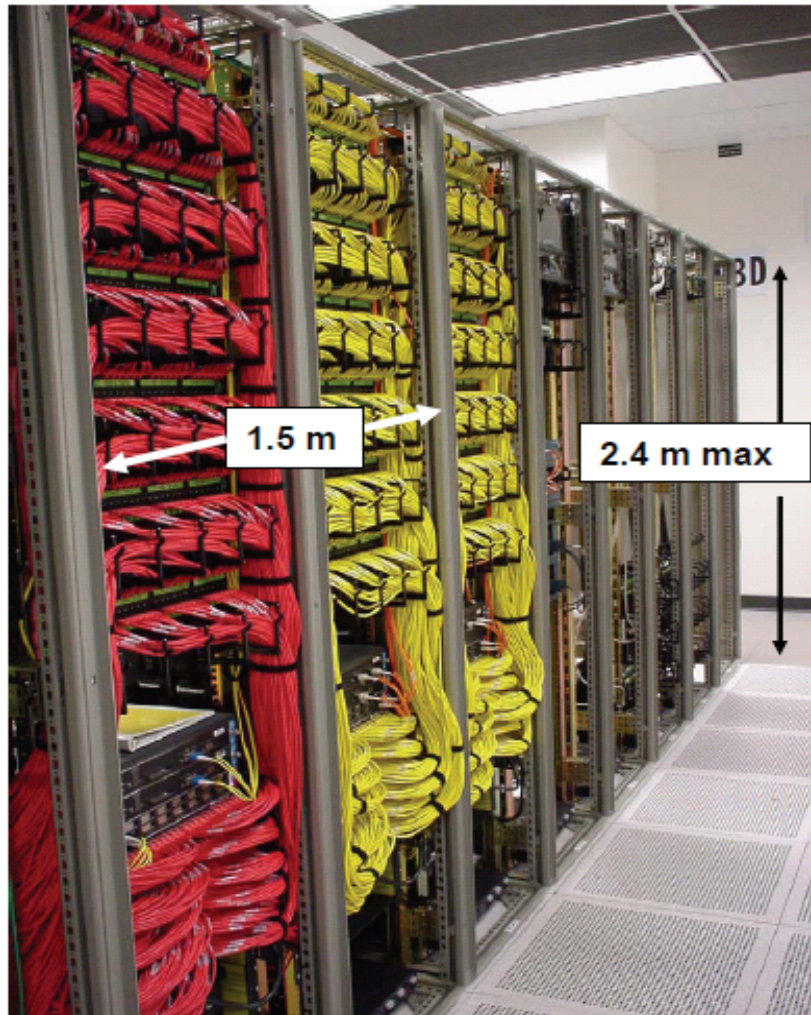
## 4.3 光インタフェース(シリアル)



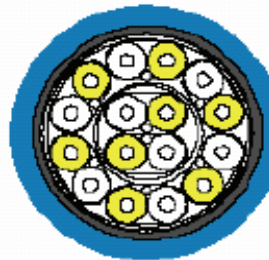
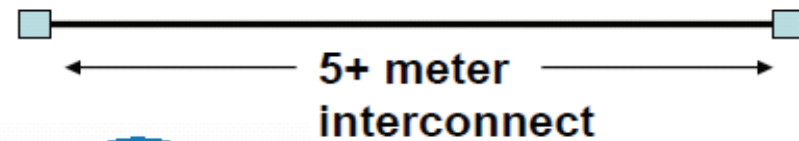
Operation capability of four optical components at 1300nm should be considered.



## 4.4 電気インタフェース



### Intra/Inter rack/cabinet applications



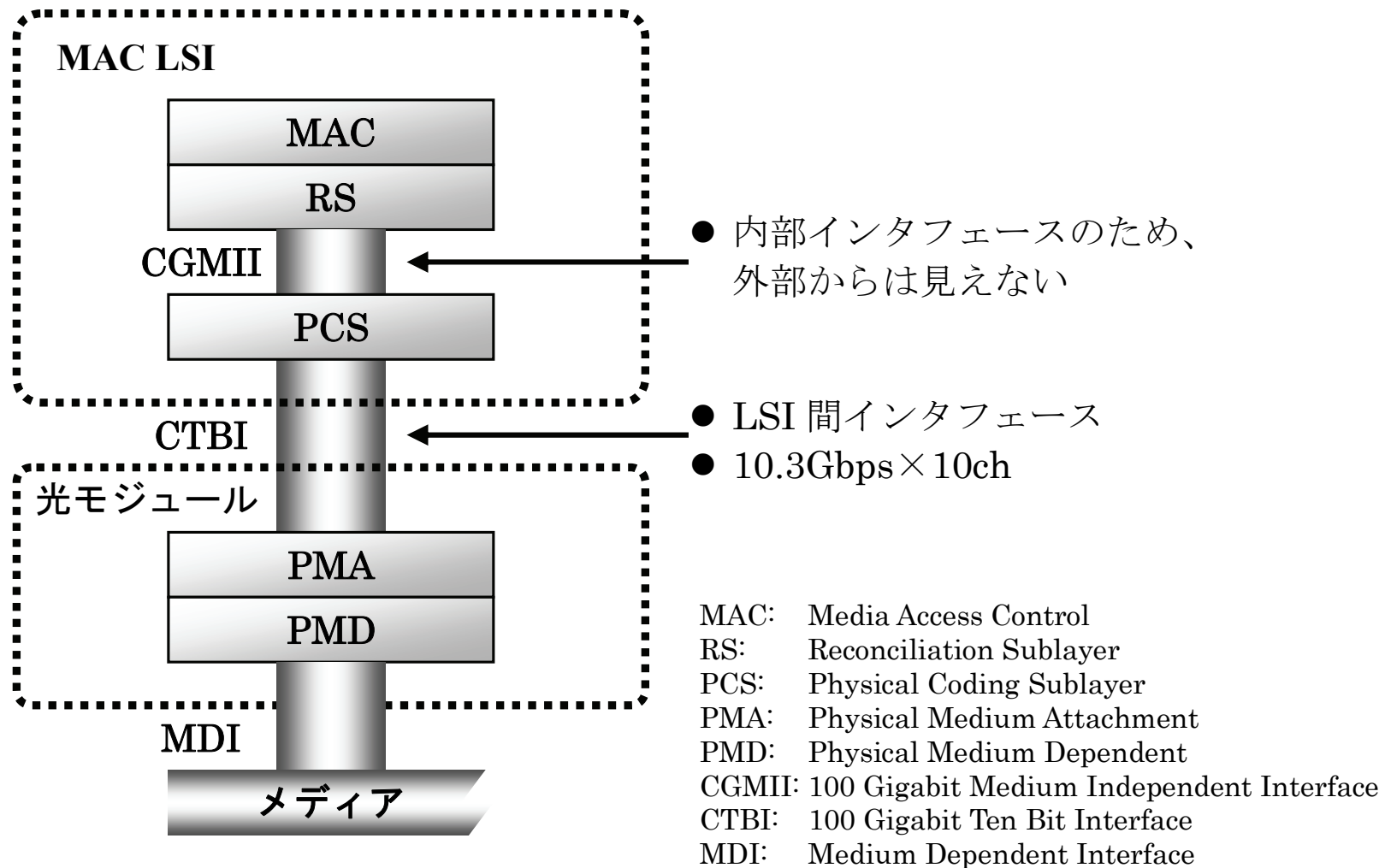
100 ohm 8 pairs  
28 AWG - 5.6mm (0.220 in)

### TIA-942 - Cabinet and rack height

- The maximum rack and cabinet height shall be 2.4 m (8 ft).
- Preferably no taller than 2.1 m (7 ft) for easier access to the equipment or connecting hardware installed at the top.

[http://www.ieee802.org/3/hssg/public/jan07/diminico\\_01\\_0107.pdf](http://www.ieee802.org/3/hssg/public/jan07/diminico_01_0107.pdf)

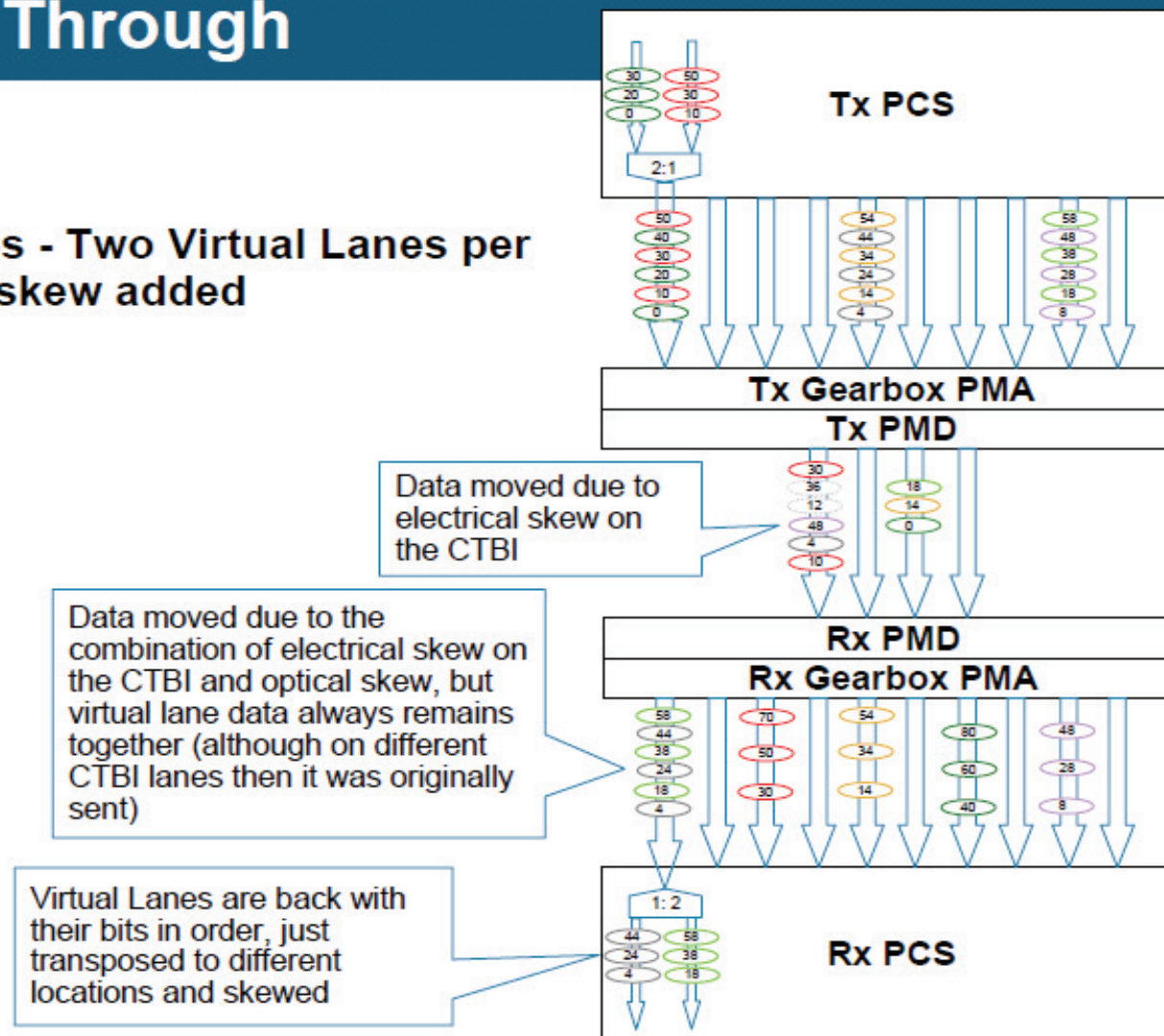
## 4.5 論理レイヤ (PMA,PCS) CTBI



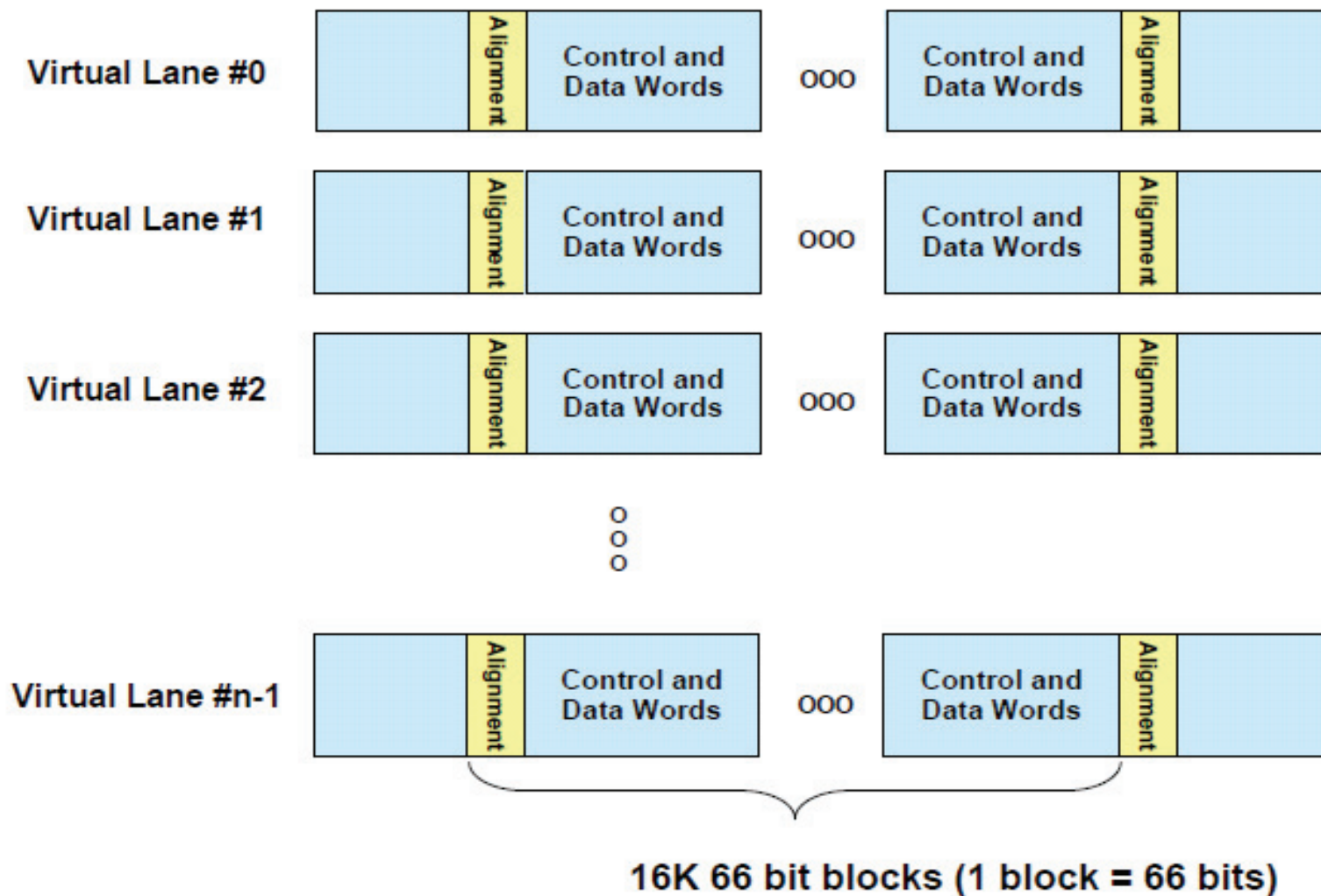
# 4.5 論理レイヤ (Virtual Lane)

## Bit Flow Through

4 PMD Lanes - Two Virtual Lanes per CTBI Lane, skew added

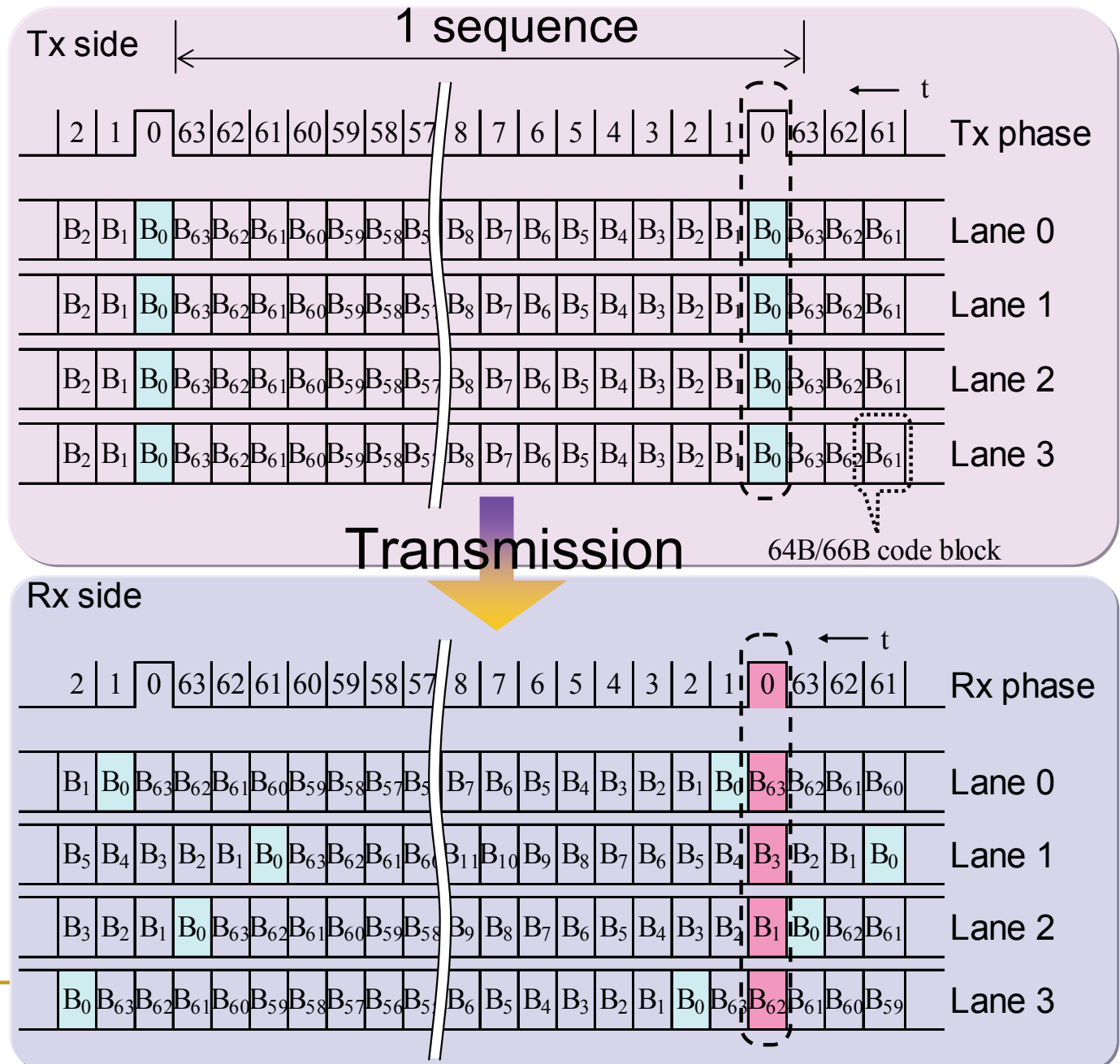


## 4.5 論理レイヤ (スキュー補正・シスコ)



# 4.5 論理レイヤ(スキュー補正・日立)

- Uses Idle sequence like the XAUI, but uses 64B/66B code
- TX side: Same regular idle sequence is output to all lanes at the same time
- RX side: Skew between lanes is detected based on phase difference between received sequences and internal phase in Rx



---

# 5. 今後の焦点

5.1 今後の会議の争点

5.2 製品における課題

# 5.1 今後の会議の争点

## 4月/5月会議の焦点

1. 距離のターゲット      ○      10km, 100m  
   議論中      40km, (2km)
2. MAC通信容量      ○      100Gbps  
   議論中      40Gbps
3. 物理インタフェース      ○      光  
   議論中      銅線
4. 光インタフェース (コスト/電力/サイズ依存)
  - ① 光チャネル構成 10 ch / 5 ch / 4 ch
  - ② 電気インタフェース 10 ch / 5 ch / 4 ch
  - ③ 伝送方式 Dense WDM, Coarse WDM, Ribbon, DQPSK
5. ロジック回路 (電力/サイズ/遅延依存)
  - ① 符号化方式 (64B/66B)
  - ② スキュー補正
  - ③ 100G-MAC

DQPSK: Differential quadrature phase-shift keying

## 5.2 部品における課題

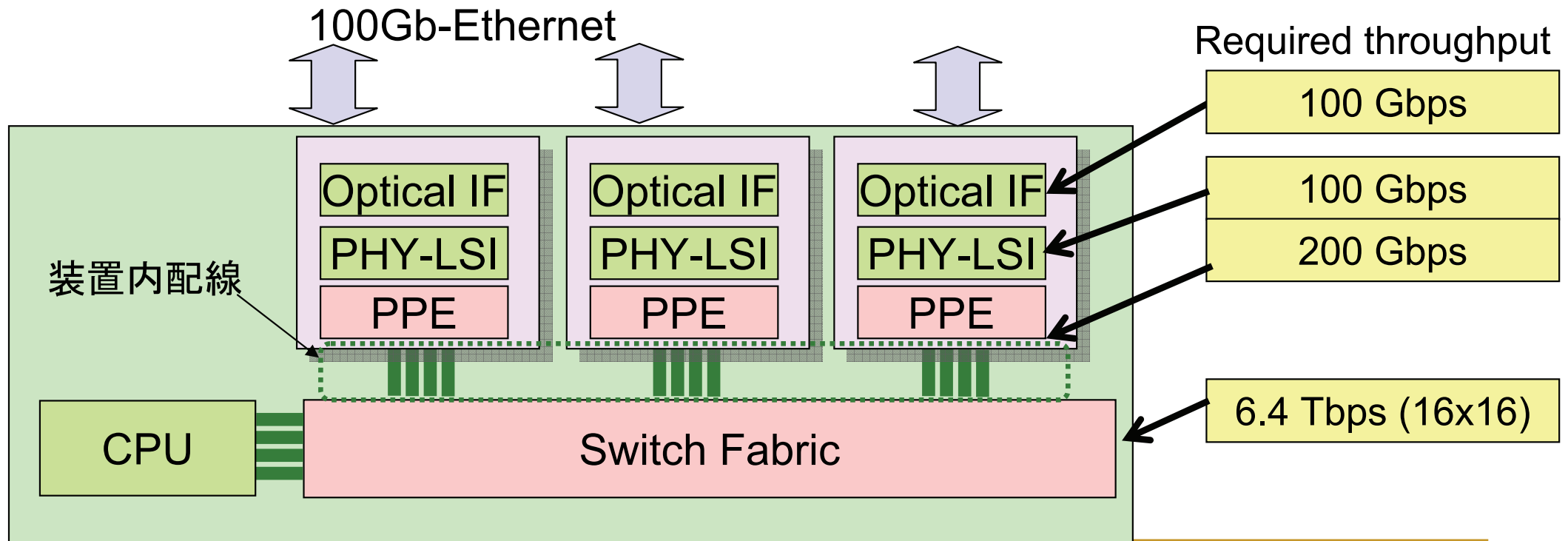
- 光インタフェース
  - 小型化 (XENPAK以下, 現状弁当箱)
  - 省電力 (10W以下, 現状30W)
  - 低コスト (集積化/省部品化, CH毎個別部品)
  - 長距離伝送 (>40km, 規格以上の伝送距離)
- ロジック回路 (PHY、MACフレイマ)
  - 省電力 (特にLSIインタフェースの消費電力低減)
  - 小型 (=コスト、65nm-CMOS使用)
  - 集積度 (ポート数、PHYとMACの一体集積)

PHY: Physical , MAC: Media Access Control  
OTN: Optical Transport Network



## 5.2 装置における課題

- 高速光インターフェース: 100Gbps長距離低コスト伝送
- 高速PHY-LSI: レイヤ1処理(フレーム抽出等), CMOS
- 高速パケット処理: 200Gbps・省電力パケット解析処理
- 大容量スイッチング: 5~10Tbps・省電力経路交換
- 装置内配線: 5~10Tbps省電力LSI間接続



Block Diagram of routers

# まとめ

- 100Gbイーサネット
  - 規格成立は2009年夏
  - 主要用途は メトロネットワークとデータセンタ
  - 並列伝送の利用によりまず実現(現実的)
  - 普及の鍵は コスト、消費電力、サイズ
  - 技術的には、伝送と共に、信号処理部の難易度が高い