Google

# 200G per Lane for beyond 400GbE

- an update from the 7/27/2020 NEA presentation

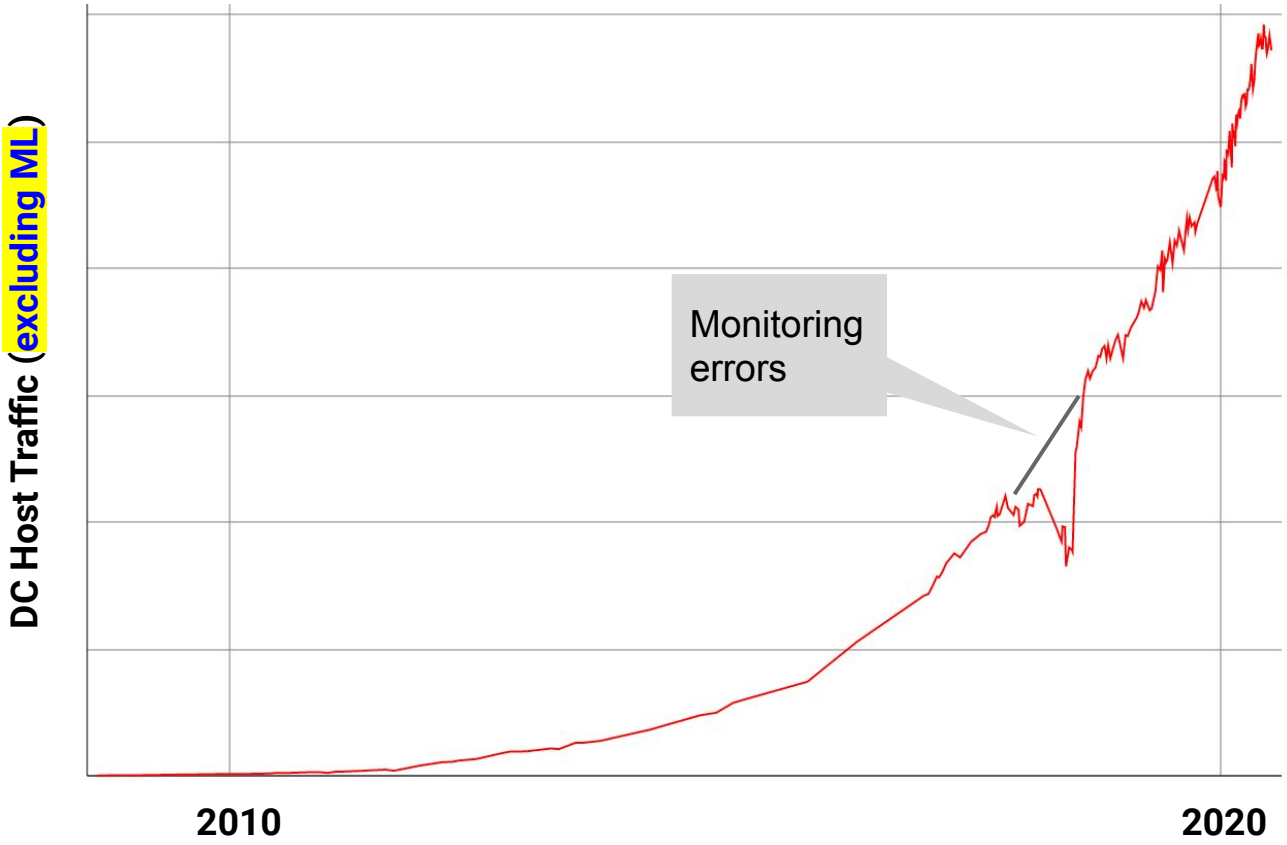
Cedric Lam, Xiang Zhou, and Hong Liu

03/29/2021
IEEE 802.3 B400G SG Meeting

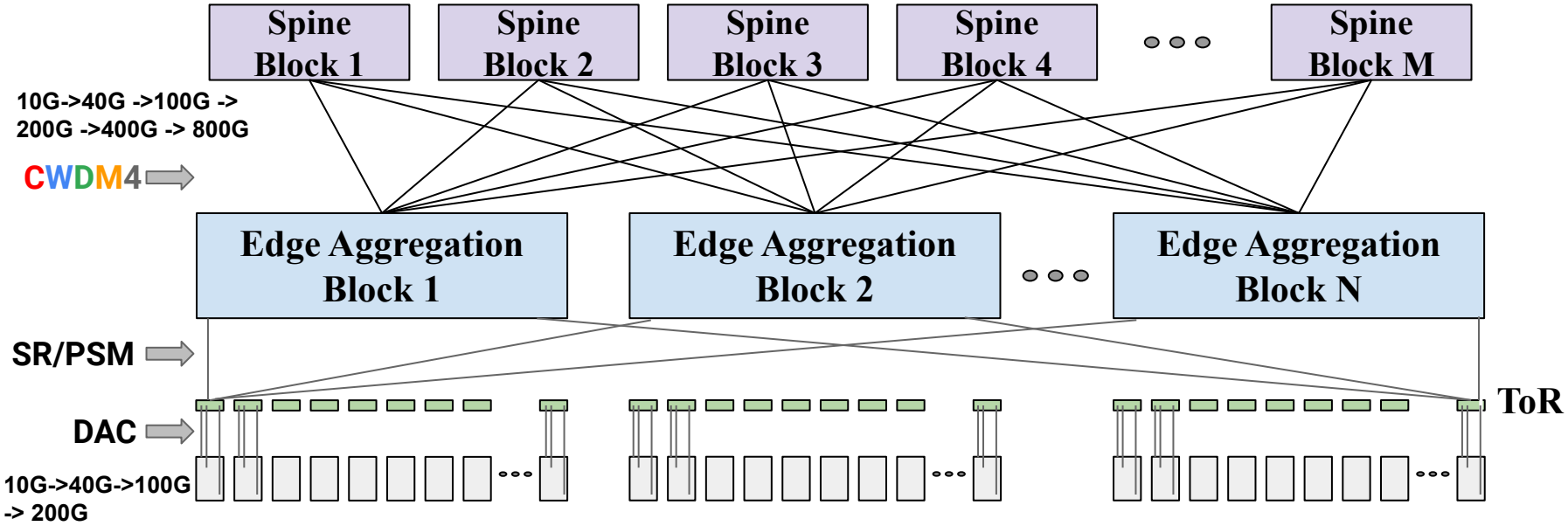
# Outline

- Driver and use cases for beyond 400G
- Justification for 200G per lane
  - Lower TCO
  - Scalability to 1.6T Ethernet
- 200G optical lane technical feasibilities
  - Baseline performance for different modulation format choices
  - Key component requirements
  - 200G per lane (optical) components readiness survey

# DC Traffic Continues to Grow Rapidly (Regular Servers)



# > 400GbE will be needed in DCN Fabrics

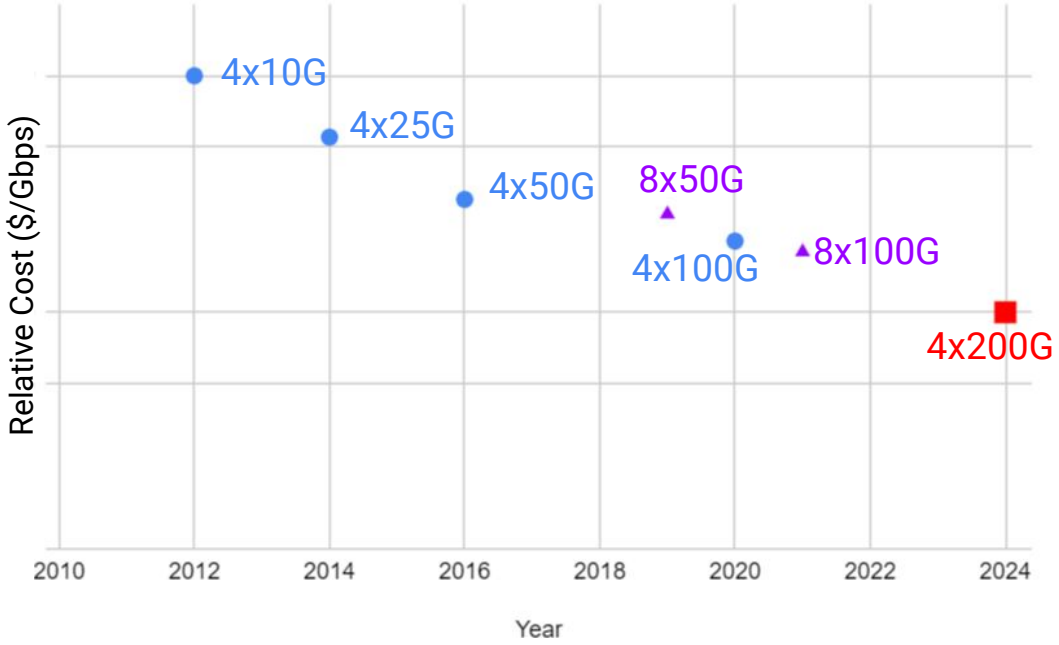



Backward compatibility between generations of interconnects enables smooth upgrade of datacenter networks.

# Why 200G per Lane?

- Cost efficiency for 800G
  - 4x200G
- Path to 1.6Tb & 3.2Tb per port
  - OSFP
  - OSFP-DD
  - CPO

# Cost/Gbps vs. Speed per Optical Lane




- **Faster optical lane speed is key to lower costs, but needs to align with electrical I/O speed for best cost & power efficiency**

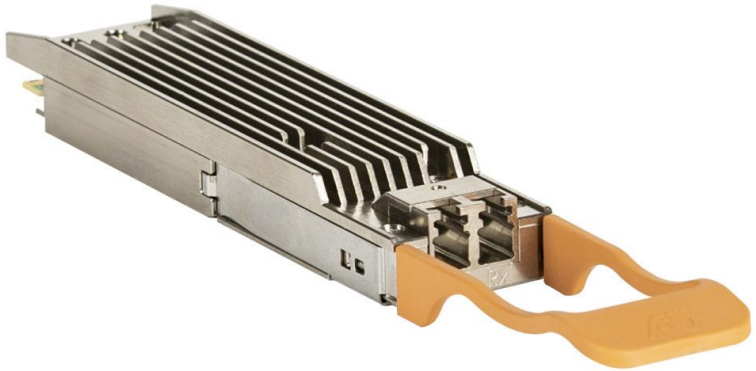
# Implementation Comparison of 800G

| | IM-DD PAM (8 lanes ) | IM-DD PAM (4 lanes) |
|---|---|---|
| **Baud Rate (Gbaud)** | 56G | ~112G |
| **Number of Lasers** | 8 | 4 |
| **MZMs and Drivers** | 8 | 4 |
| **PD/TIAs** | 8 | 4 |
| **Relative DSP power** | 1 | ~1.1 (stronger FEC and DSP ) |
| **Link distance** | Limited by dispersion (2km, CWDM8) | Limited by dispersion (< 1km*, CWDM4) |
| **Fan out granularity** | 100Gb/s | 200Gb/s |
| **Scale to 1.6Tb/s and beyond** | No | Yes |

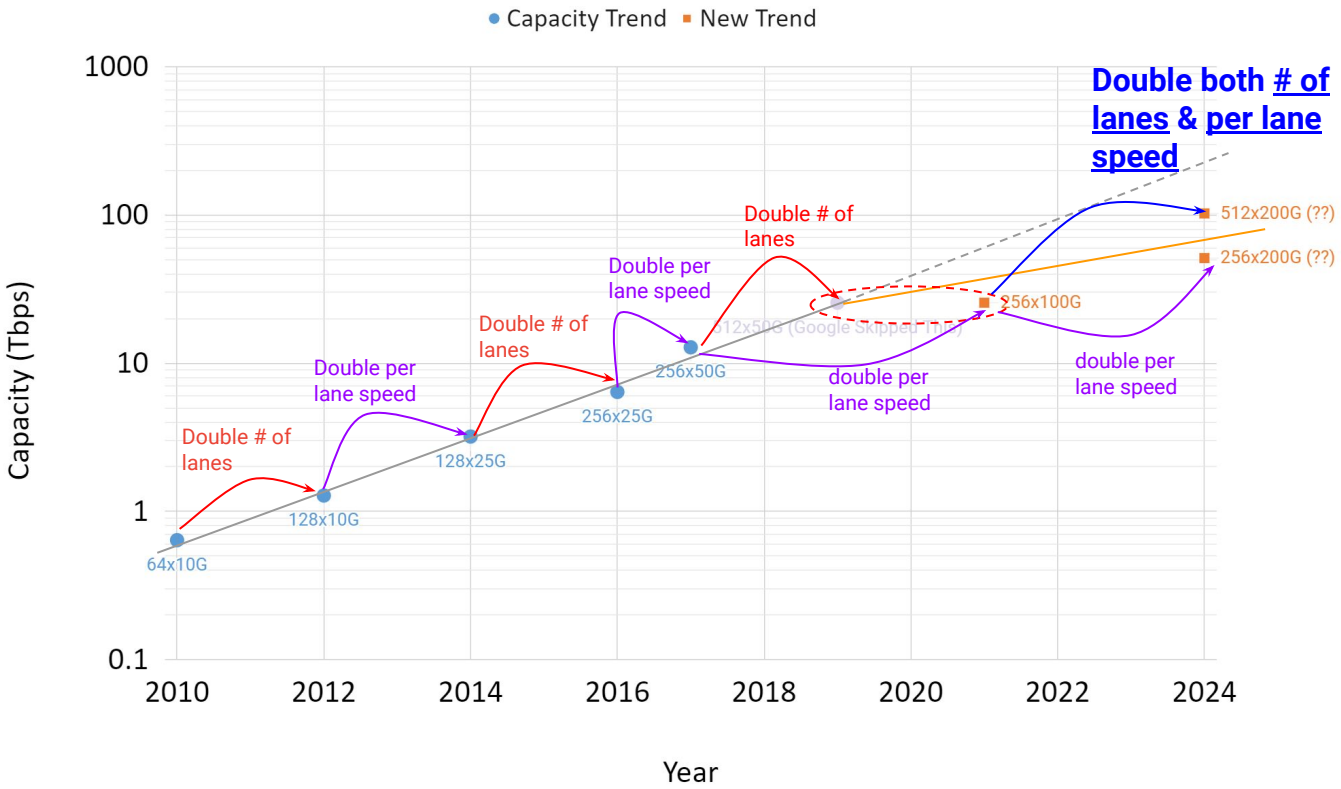
* Reach may be extended by more powerful DSP such as MLSE (Ilya Lyubomirsky, IEEE 2020 summer topical talk)

# Necessity of 200Gbps Electrical Lanes

- Scalability and visibility into 1.6T Ethernet
  - OSFP defined 8 electrical lanes
  - 8x 200G gives us 1.6Tb capacity
- Enable 100Tbps Switch ASIC
- Matching the electrical lane speed w/ optical lane speed
  - Simplifies module architectures
  - Reduces overall power consumption
  - Keeps the cost down in the long run
- Support for C2M, C2C and CR
  - Flexible, heterogeneous interconnects @lower cost
  - Better flexibility, serviceability and manufacturability in deployment



# 100Tbps Switch ASIC in 3 to 4 Years?




- Switch ASIC capacity growth slowed down but the demand is not.
  - New applications are emerging.
- It is harder to increase the number of lanes due to SI and # of packaging pins.
- Power consumption of switch ASIC is another concern

Google

# 200G Optical Lane Technical Feasibilities

# System Model

## Focus on the following Functions/Blocks

- Two candidate modulation formats: PAM4 and PAM6
- 2 types of transmitters
  - InP EML
  - SiP MZM
- PD + TIA: R=0.8A/W, IRN=16pA/sqrt(Hz), THD=3%
- Digital Electronics
  - 6-tap Tx FFE, 17-tap Rx-FFE, T-spaced
  - **FEC threshold 4e-3 assumed for 200Gb/s per lane***

* Ilya Lyubomirsky, “Coherent vs. Direct Detection for Next Generation Intra-Datacenter Optical Interconnects,” IEEE 2020 summer topical

# Overall comparison: PAM4 vs PAM6

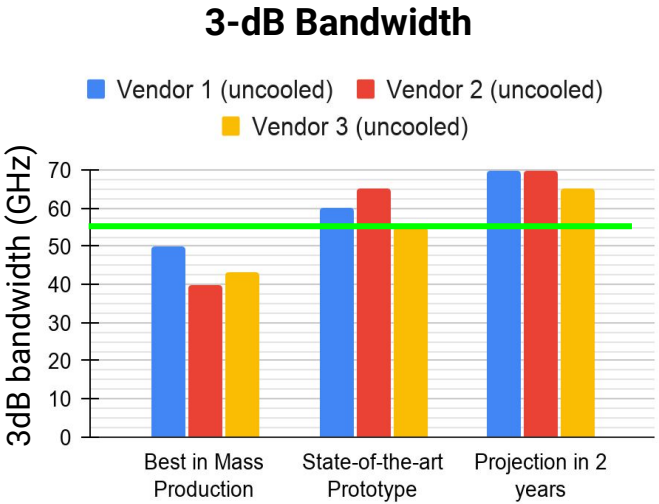
| | PAM4 | PAM6 |
|---|---|---|
| Baud rate | ~113Gbuad | ~90Gbaud |
| Rx sensitivity penalty [A] @45GHz BW | ~4.9dB | ~3.3dB |
| Rx sensitivity penalty [A] @50GHz BW | ~2.3dB | ~2.4dB |
| Rx sensitivity penalty [A] @55GHz BW | ~1.6dB | ~2.2dB |
| Support 1km O-CWDM4 CD with EML | Yes<br>CD penalty<1.5dB@55GHz | Yes<br>CD penalty<1dB@55GHz |
| DAC/ADC ENOB requirement | ~5.5 (stronger EQ) | ~5.5 (higher-order mod.) |
| Relative DSP power | **1** | **<1 ?** |

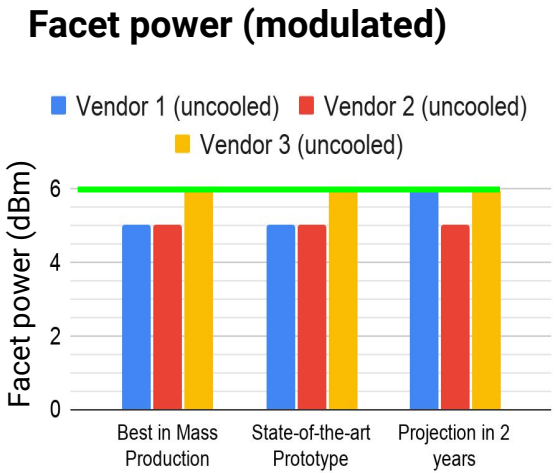
[A]: Compared to 106Gb/s per lane PAM4 with KP4 FEC

- **If PAM6 can achieve lower power, a dual-mode PAM4/PAM6 may be considered**
  - **PAM4 only for difficult links (higher link loss and/or MPI)**
  - **PAM6 for majority of the normal links to save overall network power**

# 200Gb/s per optical lane components survey

## Transmitter 1: InP EML




Preliminary requirements guideline to support 1km 800G CWDM4 reach

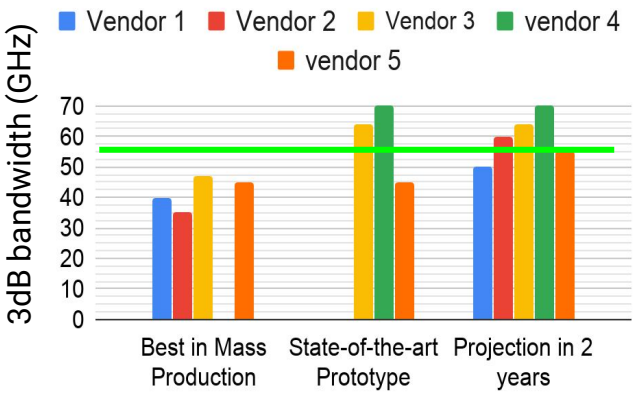
- Assume support both PAM4 and PAM6

- **Prototype**: 1 (**new**) vendor meets the preliminary guideline requirements for uncooled EML
  - July 2020, only 2 cooled prototypes met the preliminary requirements.
- **2-year projected**: ~~1~~ 2 vendors meet the preliminary guideline requirements for uncooled EML

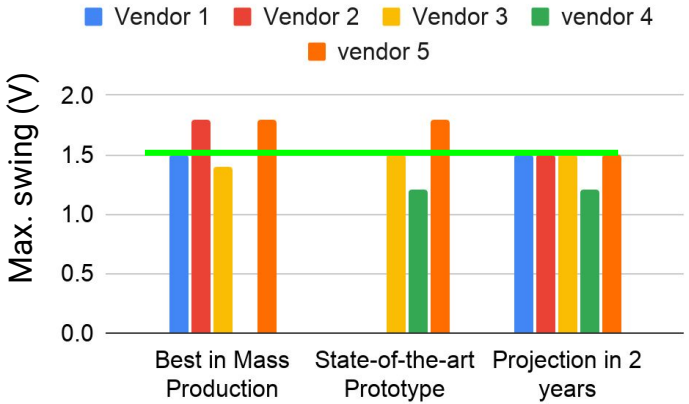
# 200Gb/s per optical lane components survey

## Transmitter 1 : EML Driver

**3-dB Bandwidth**




**Drive swing**

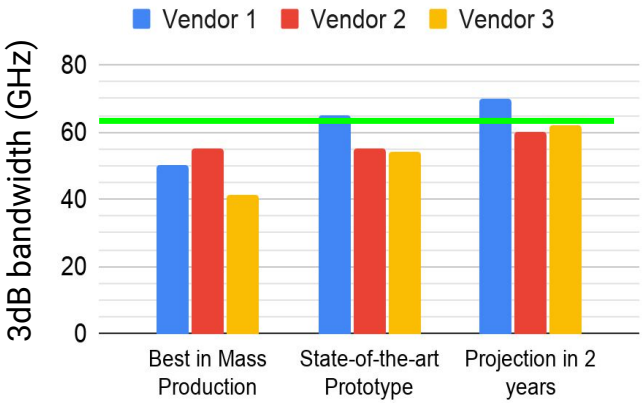



- **Prototype**: 1 vendor meets the preliminary guideline requirements
- **2-year projected**: 3 vendors meet the preliminary guideline requirements

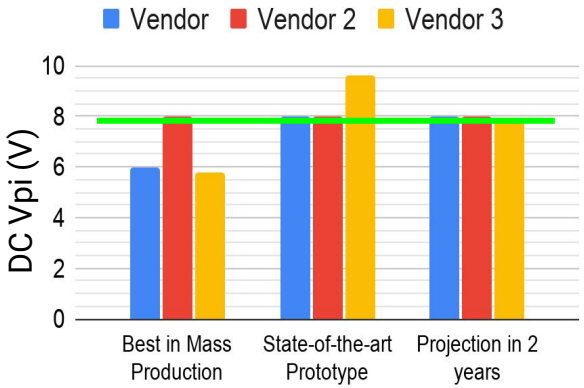
# 200Gb/s per optical lane components survey

## Transmitter 2: SiP-MZM

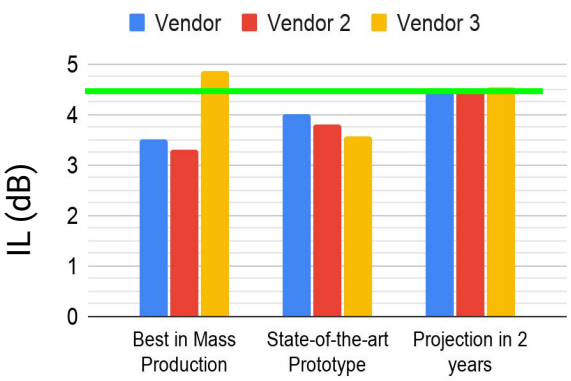
**6-dB Bandwidth**




**DC Vpi**




**Insertion loss (dB)**



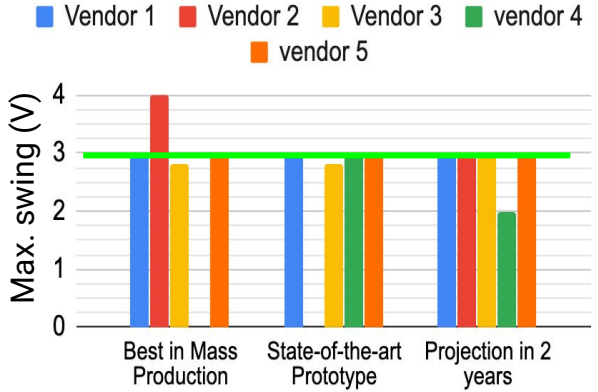

- **Prototype:** 1 (**new**) vendor meets the preliminary guideline requirements for DR reach
- **2-year projected:** ~~1~~ 2 vendor meets the preliminary guideline requirements for DR reach

# 200Gb/s per lane components survey

## Transmitter 2: SiP-MZM driver

**3-dB Bandwidth**

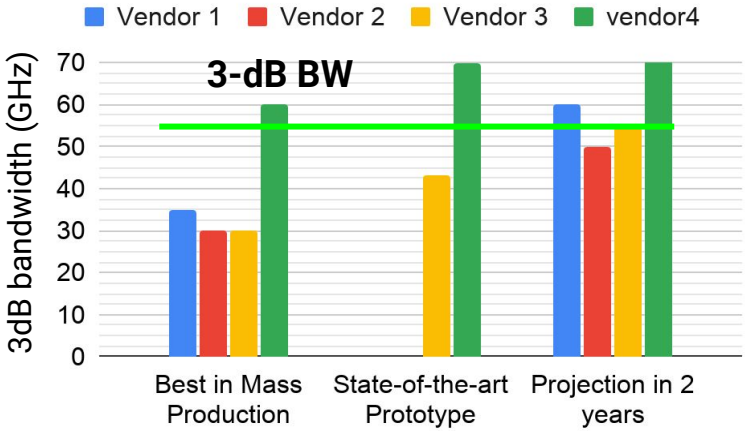

**Drive output swing**

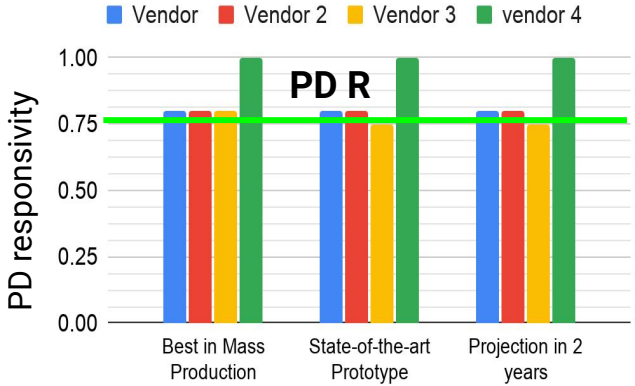



- **Prototype**: ~~2~~ 3 vendor meets the preliminary guideline requirements
- **2-year projected**: 4 vendors meet the preliminary guideline requirements

# 200Gb/s per optical lane components survey

## Receiver: PD+TIA



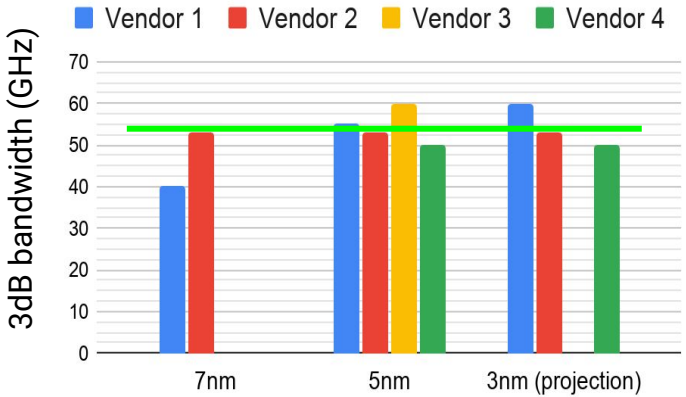




- **Prototype:** 1 (**new**) vendor meets the preliminary guideline requirements
- **2-year projected**: ~~2~~ 3 vendors meet the preliminary guideline requirements

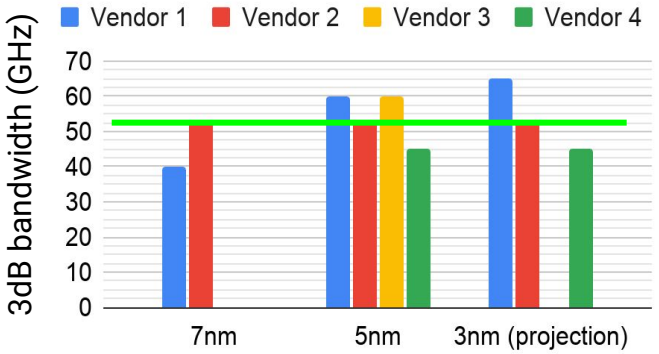
# 200Gb/s per optical lane components survey

## Digital Electronics: CMOS DAC and ADC

**DAC**




**ADC**




- **5nm CMOS**: ~~2~~ 3 vendors meet the preliminary BW guideline requirements
- **3nm CMOS**: main purpose to reduce power consumption

# 200G/s per Lane Technology Improvements

| | Vendor | Parameter | July 2020 Survey | | March 2021 Survey | |
|---|---|---|---|---|---|---|
| | | | Best Prototype | 2-year Projection | Best Protype | 2-year Projection |
| **Tx 1 (InP EML)** | 2 | EML 3dB BW(GHz), uncooled | **60** | 70 | **65** | 70 |
| | 3 | EML 3dB BW (GHz) | **43** | **47** | **55** | **65** |
| | | Cooled or uncooled EML | **Cooled** | **Cooled** | **Uncooled** | **Uncooled** |
| **Tx 2 (SiPh- MZM)** | New Vendor | 6dB-BW (GHz) | | | **54.1** | **62** |
| | | DC Vpi (V) | | | **9.6** | **7.9** |
| | | Intrinsic Insertion loss (dB) | | | **3.6** | **4.5** |
| **Tx 2 MZM Driver** | 5 | 3dB BW (GHz) | **45** | **55** | **60** | **65** |
| **Rx (PD+TIA)** | 3 | 3dB BW (GHz) | 43 | **50** | 43 | **55** |
| | 4 | 3dB BW (GHz) | **N.A** | **70** | **70** | **80** |
| **Digital** | 3 | 5nm DAC BW (GHz) | **~47.5** | N.A | **~60** | N.A |

# 200Gb/s per Optiacal Lane Components Readiness

For 500m DR4 (3dB) and 1km CWDM4 (4dB)

| | | Mass Production | Prototype | 2-year Projected |
|---|---|---|---|---|
| Transmitter 1<br>InP EML | InP EML | ✗ | ✓ (uncooled)<br>July 2020: only cooled<br>March 2021: uncooled | ✓ (uncooled) |
| | EML Driver | ✗ | ✓ | ✓ |
| Transmitter 2<br>SiP MZM | MZM (SiPh) | ✗ | ~~✗~~ Ready for DR-reach | Ready for DR-reach |
| | MZM Driver | ✗ | ✓ | ✓ |
| Receiver | PD / TIA | ✗ | ~~✗~~ ✓ | ✓ |
| Electronics | CMOS DSP | ✗ (7nm) | ✓ (5nm) | ✓ (5nm/3nm) |

# Conclusions

- Demands for datacenter bandwidths keep growing quickly.
- It is right time to develop the next higher-speed Ethernet beyond 400GbE
- For intra-datacenter applications, 200Gbps per lane IM-DD implementation provides:
  - Lower TCO
  - Pathway to 1.6Tbps Ethernet
- Technical feasibility of 200Gbps per optical lane is within the reach in the next two years
  - Well within the time frame to complete the next higher-speed Ethernet standard