AI시대의 통신 기술

여인동, Ciena SE
2024. 6

목차

1. 통신기술의 발전
2. AI시대의 통신 기술
3. What’s next?

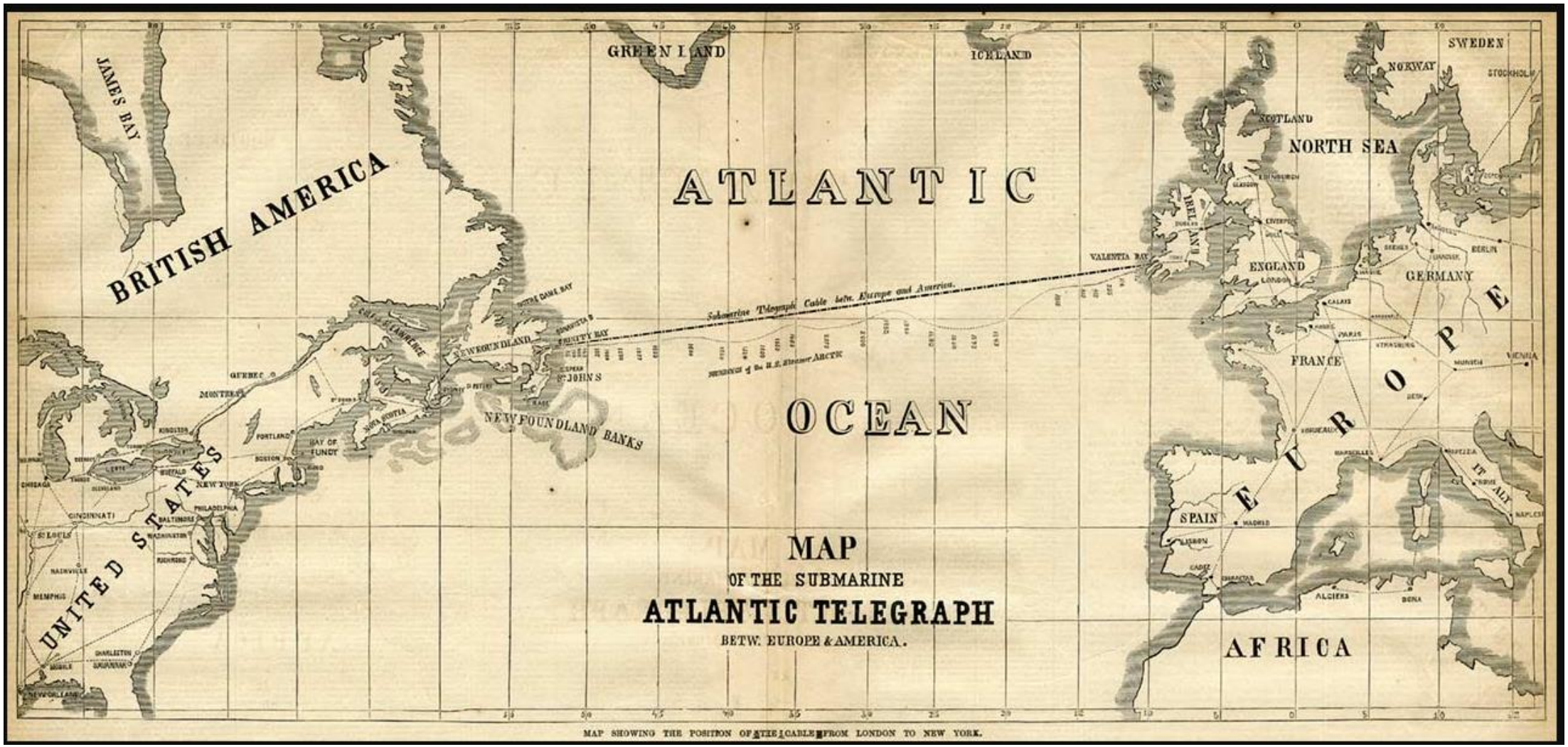
통신기술의 발전

통신기술 발전 방향

전신, 전화등의 매체를 사용하여 정보나 의사를 전달하는 것.

from Wiki

최초의 해저케이블 통신 – 1851년 대서양 횡단




ARCTIC SEA
Melville Sound
BAFFIN BAY
GREENLAND
SPITZBERGEN
Frans Josef Island
Liakhov Is. or New Siberia
Behring Sea
HUDSON BAY
NORTH AMERICA
PACIFIC OCEAN
BRITISH ISLES
RUSSIAN EMPIRE
Sea of Okhotsk
CHINESE EMPIRE
PERSIA
Tokyo
Gulf of Mexico
Caribbean Sea
Sahara or Great Desert
Philippine Is.
PACIFIC OCEAN
Ceylon
SOUTH AMERICA
ATLANTIC OCEAN
INDIAN OCEAN
Solomon Is.
Coral Sea
WESTERN AUSTRALIA
SOUTH AUSTRALIA
QUEENSLAND
NEW SOUTH WALES
SYDNEY
TASMANIA
NEW ZEALAND
Cape of Good Hope
PATAGONIA

어떻게 더 효율적으로 목소리를 전달할까? – 압축/신장(양자화, Quantizing)

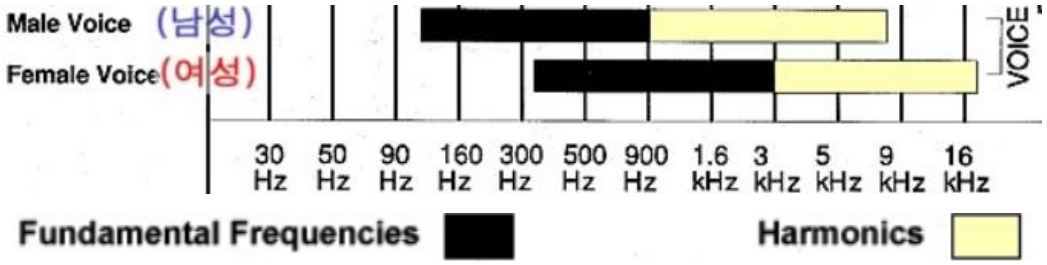
G.711 Companding algorithms

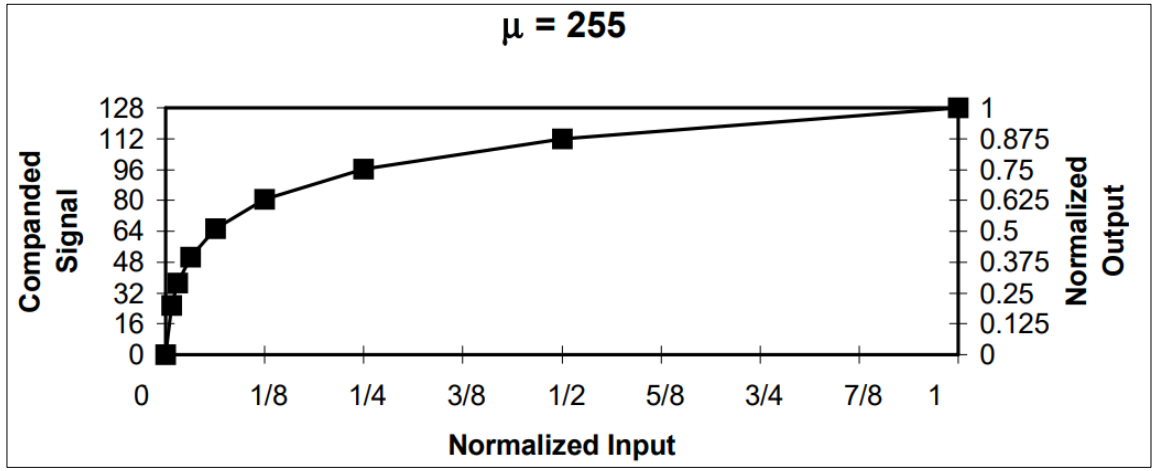
μ-law(used in T-Carrier)

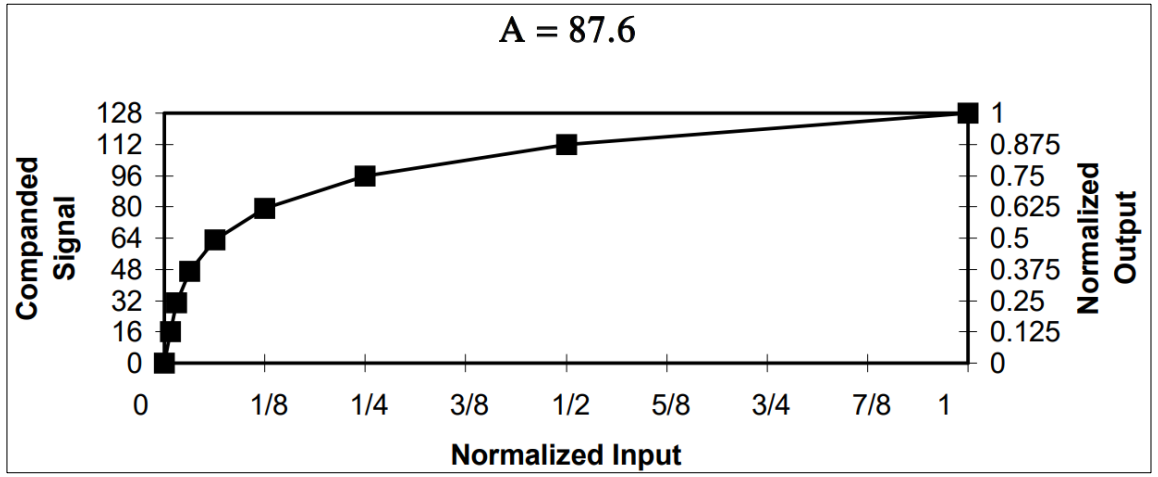
$$F(x) = \frac{\operatorname{sgn}(x) * \ln(1+\mu|x|))}{\ln(1+\mu)} \quad 0 \le |x| \le 1$$

$$F(x) = \begin{cases} \dfrac{A*|x|}{1+\ln(A)} & 0 \le |x| < \dfrac{1}{A} \\ \dfrac{\operatorname{sgn}(x)*(1+\ln(A|x|))}{1+\ln(A)} & \dfrac{1}{A} \le |x| \le 1 \end{cases}$$

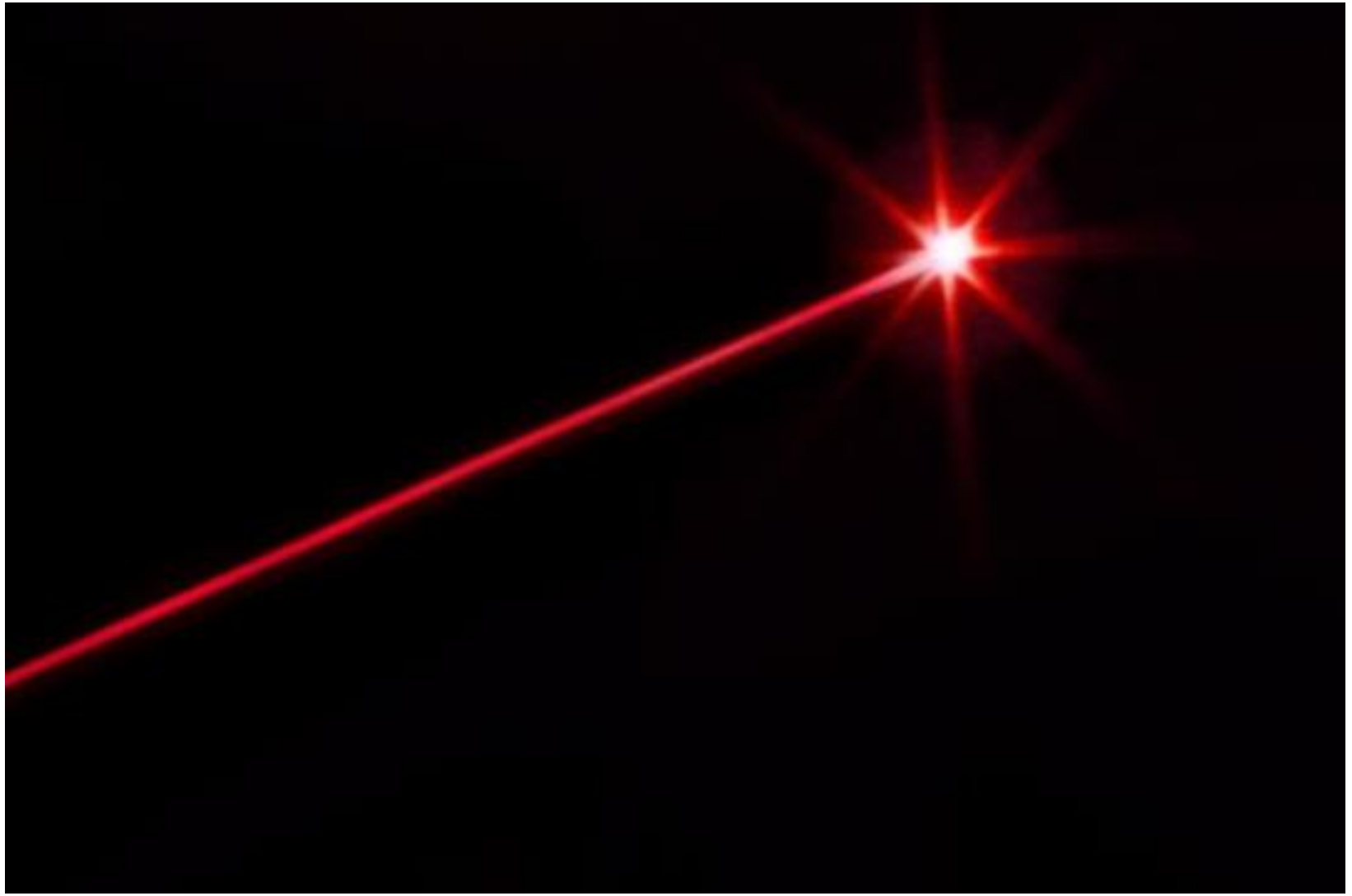
A-Law Equation



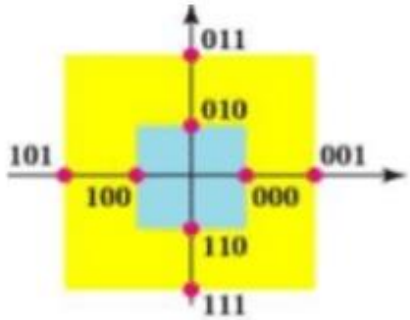




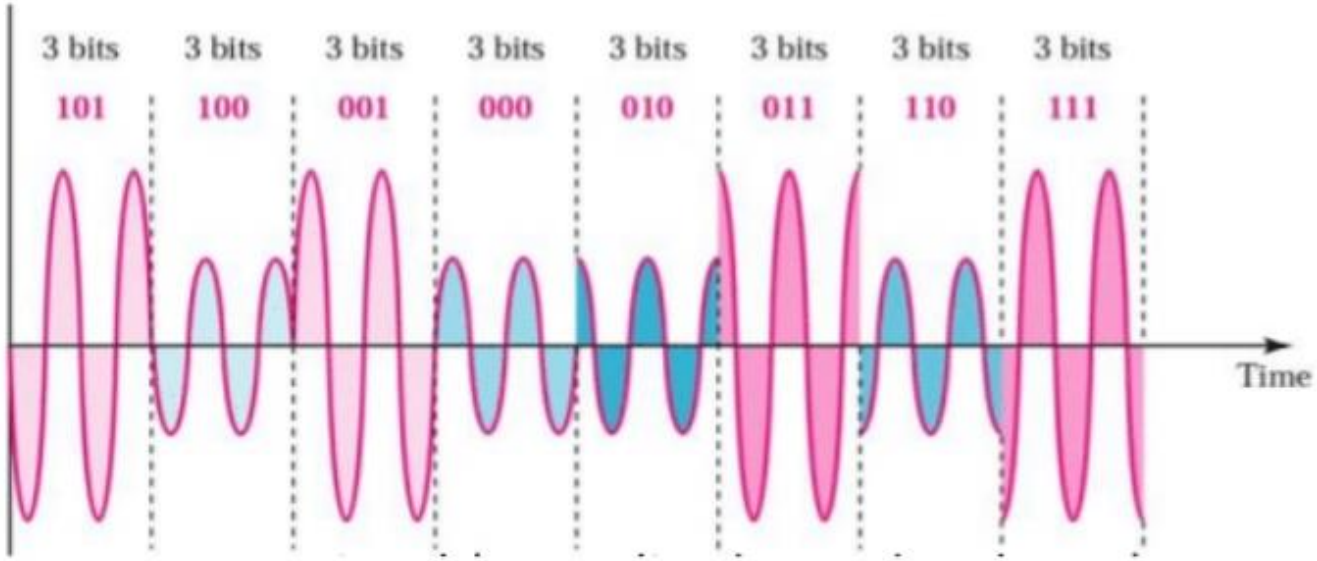
가장 효율적인 전송 방식 – 빛



어떻게 효율을 높일까?
ASK/FSK/PSK -> QAM -> PCS

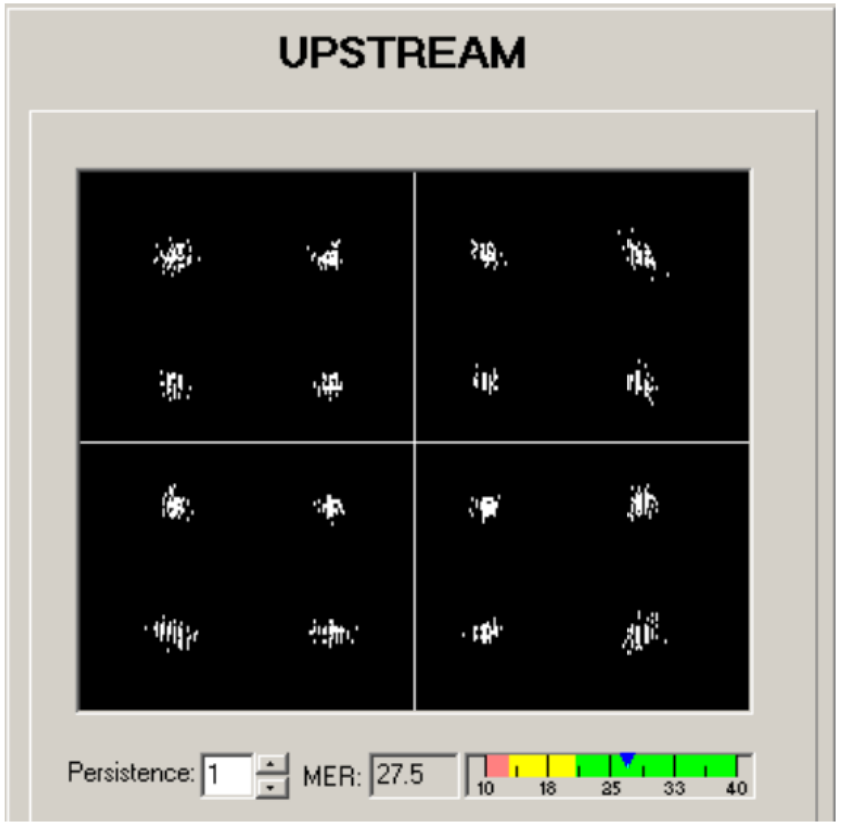


8-QAM
2 amplitudes, 4 phases

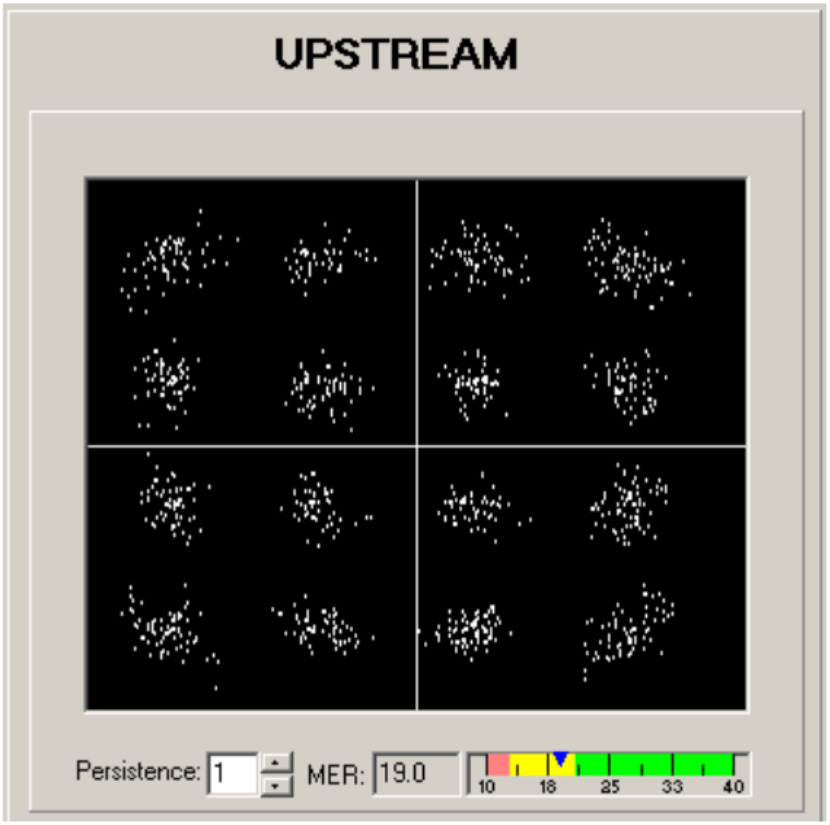


어떻게 효율을 높일까? ASK/FSK/PSK -> QAM -> PCS

Ideal



Noise

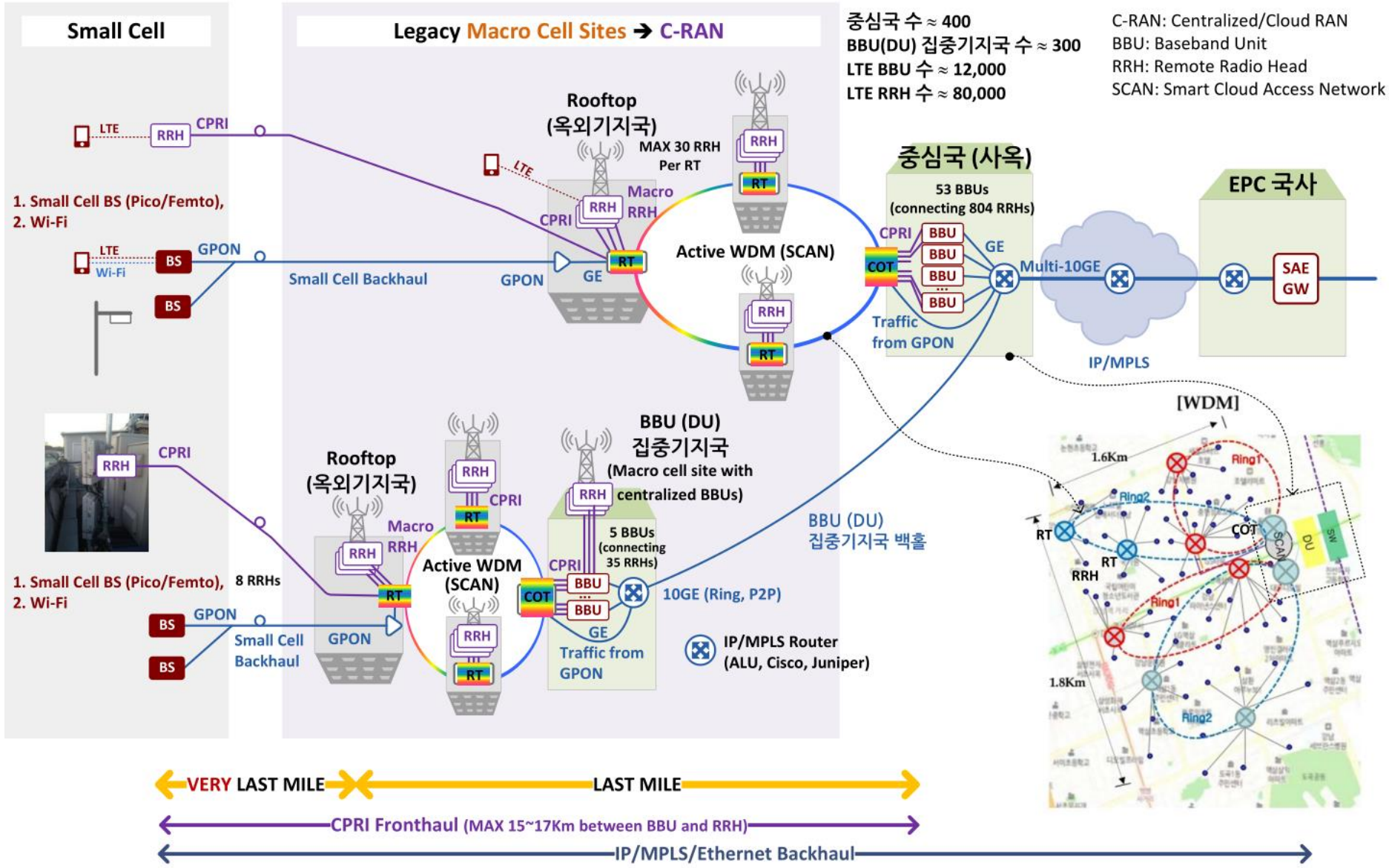


어떻게 더 많이 보내지? – 유선(WDM), 무선(MIMO)

https://www.edunet.net/nedu/contsvc/viewWkstContPost.do?contents_id=fs_a0000-2015-0702-0000-000000000180&head_div=s2015w

어떤 통신기술들이 쓰이고 있을까? Access

5G/LTE
CPRI
PON
MSPP
WDM
…



*넷마니아즈

어떤 통신기술들이 쓰이고 있을까?

Metro/Core

MPLS

Traffic Engineering

IP

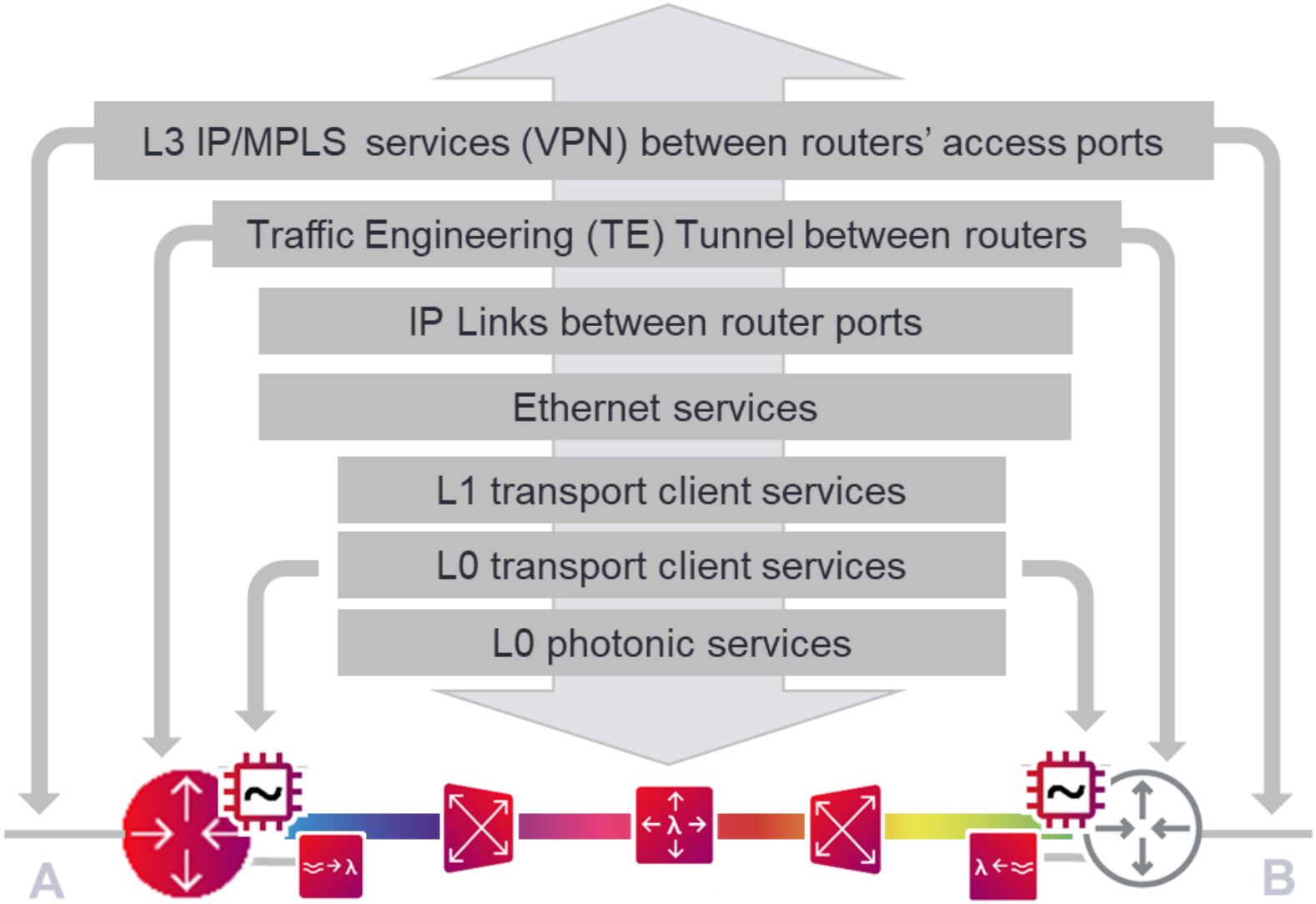
Ethernet

OTN

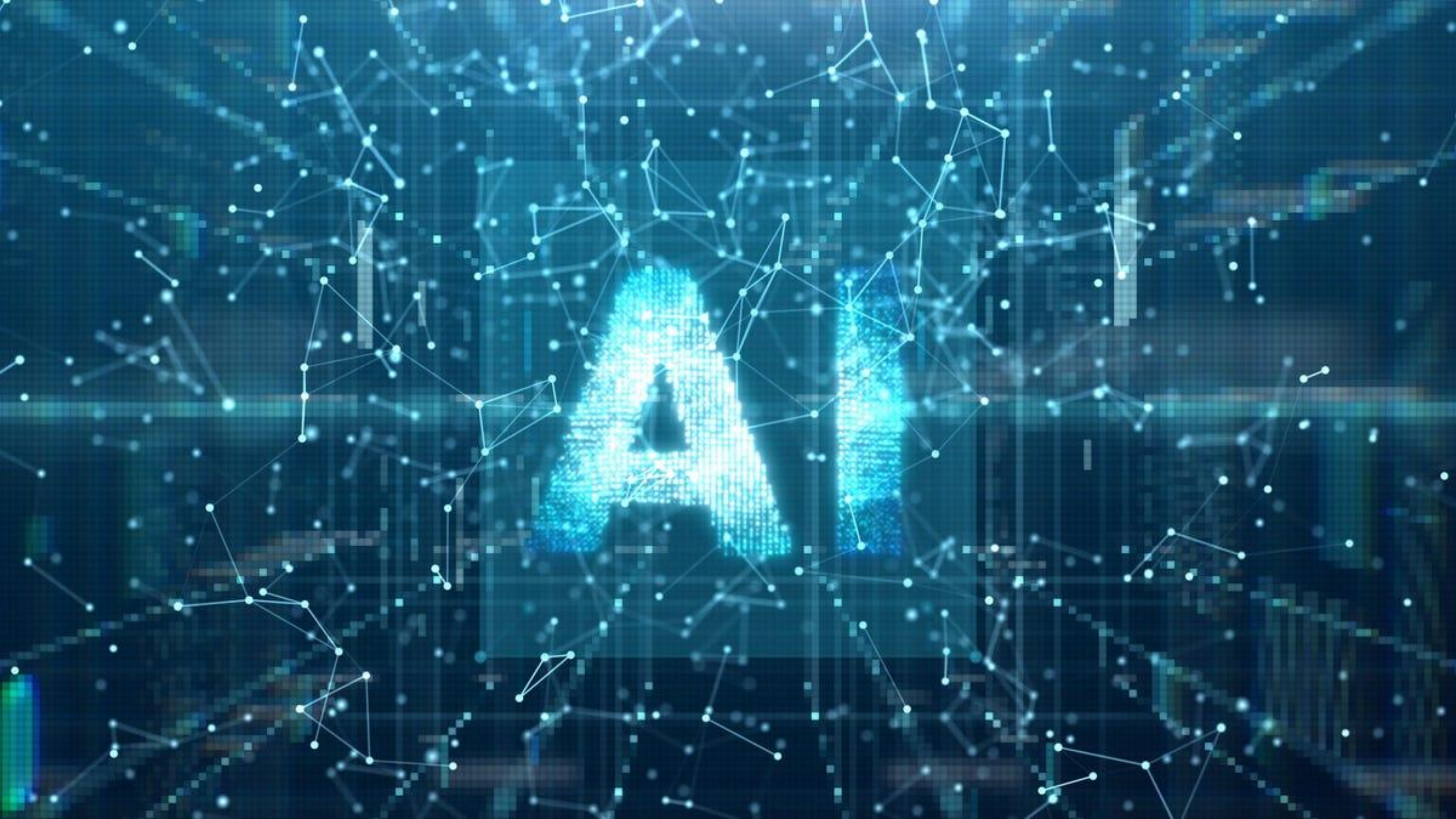
ROADM

Coherent

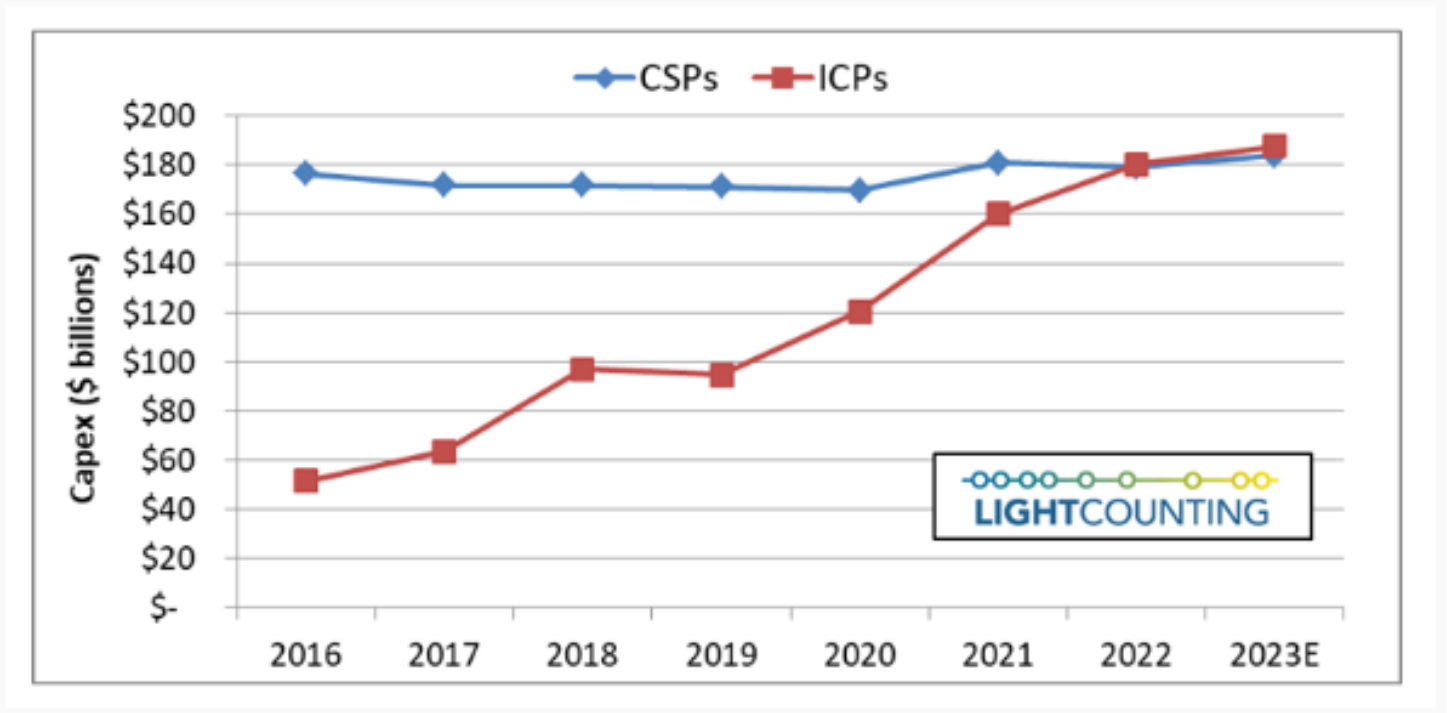
Fiber



AI시대의 통신기술


AI

CSP, ICP 투자예산 증감률 비교



AT&T, BT, DT, NTT, Softbank, Verizon, etc.

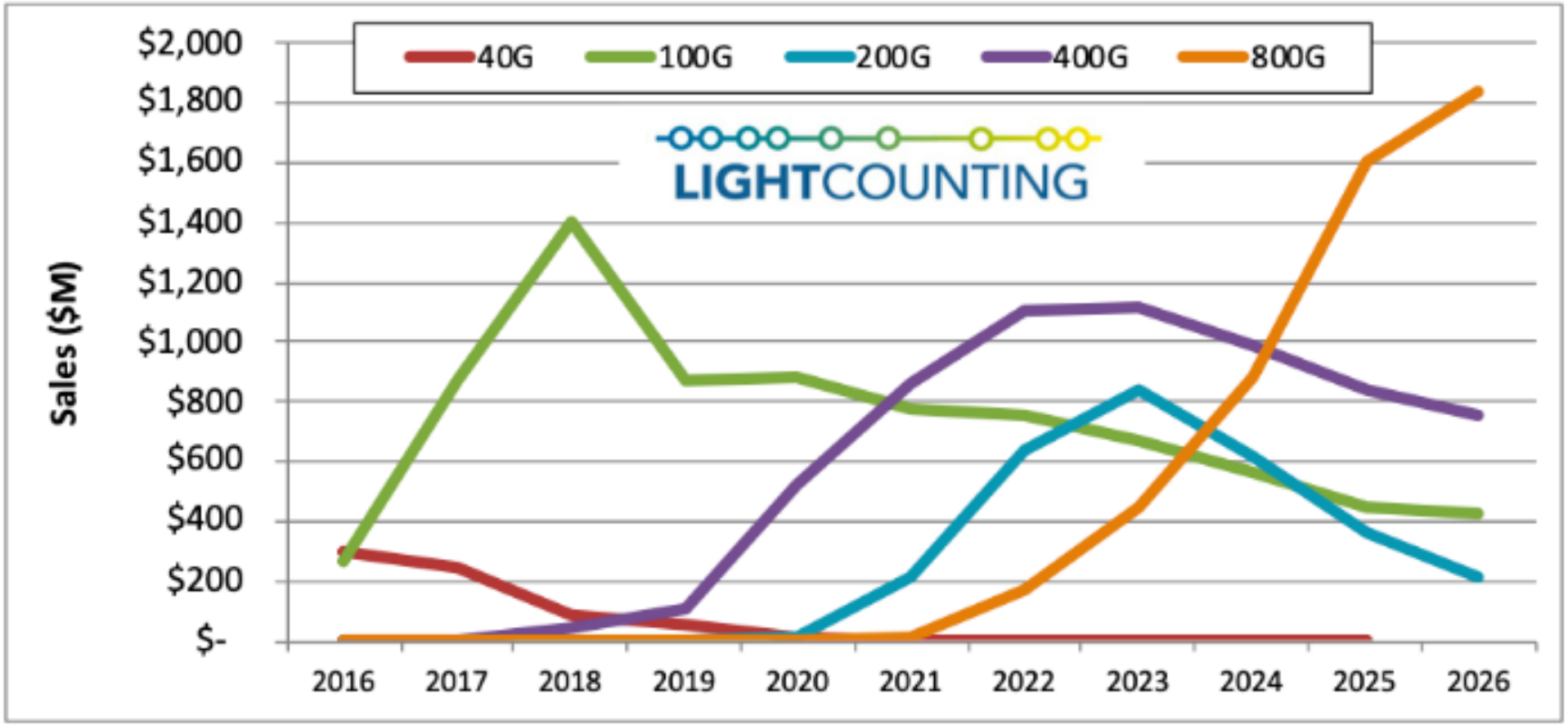
Comm Service Providers (CSPs)	Internet Content Providers (ICPs)

Alibaba, Alphabet, Apple, Baidu, eBay, Equinix, Meta, Tencent, etc.

*https://techblog.comsoc.org/category/optical-components/

Ethernet Transceiver 마켓 분석
Understanding the 400G+ market

Figure: Sales of Ethernet Transceivers to the Top 5 Cloud Companies



New supply chain shortages now
(100G VCSELs and 4x100G/8x100G DSP chips)

Gemini comes in three sizes

Our most capable and largest model for highly-complex tasks.

Our best model for scaling across a wide range of tasks.

Our most efficient model for on-device tasks.

Chat GPT 4 이상 AI 요구사양

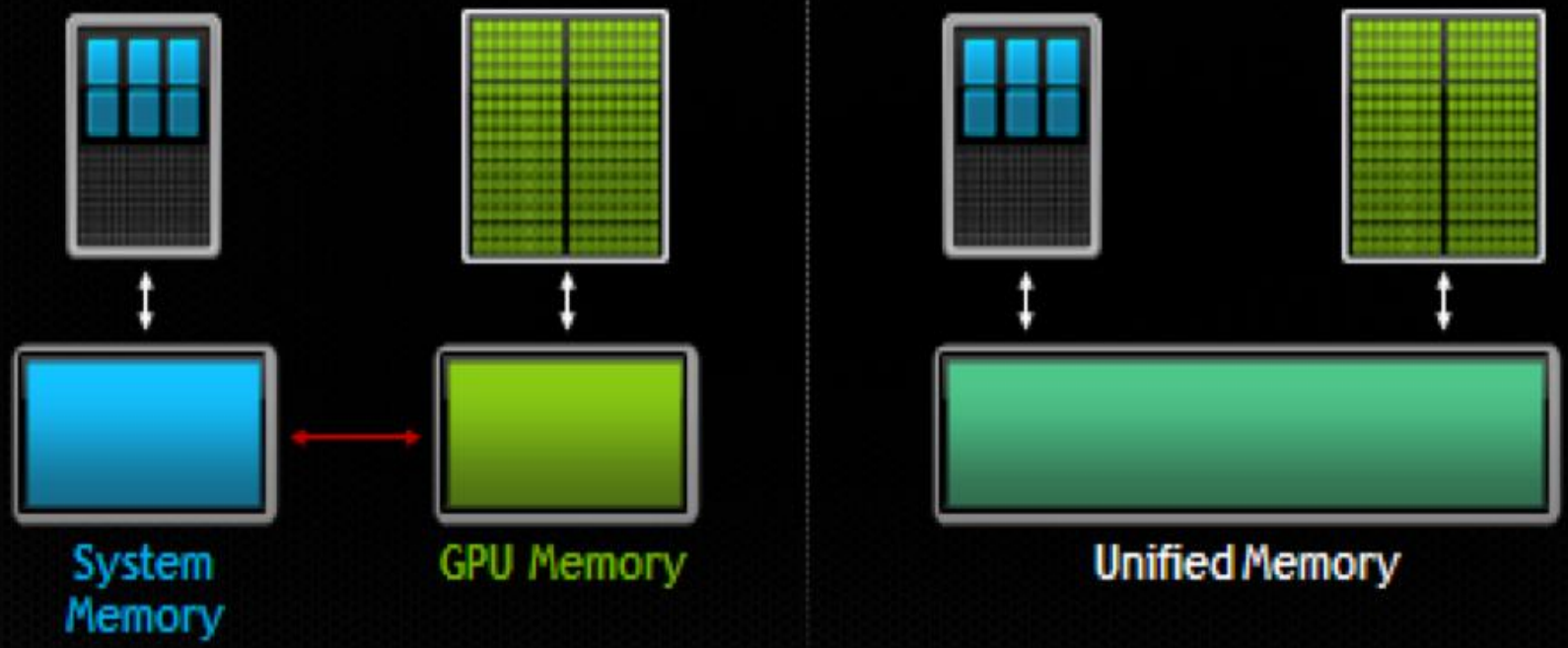
Mixtral 8×7B employs a **Mixture of Experts** (MoE) technique, incorporating eight expert models into one. This unique structure allows it to function at the speed of a **12B parameter-dense model** while containing four times the number of effective parameters. It's a breakthrough in efficiency but comes with a high VRAM requirement. The model needs a staggering **90GB of VRAM** in half-precision, posing a challenge for local setups and individual developers.

1,500W

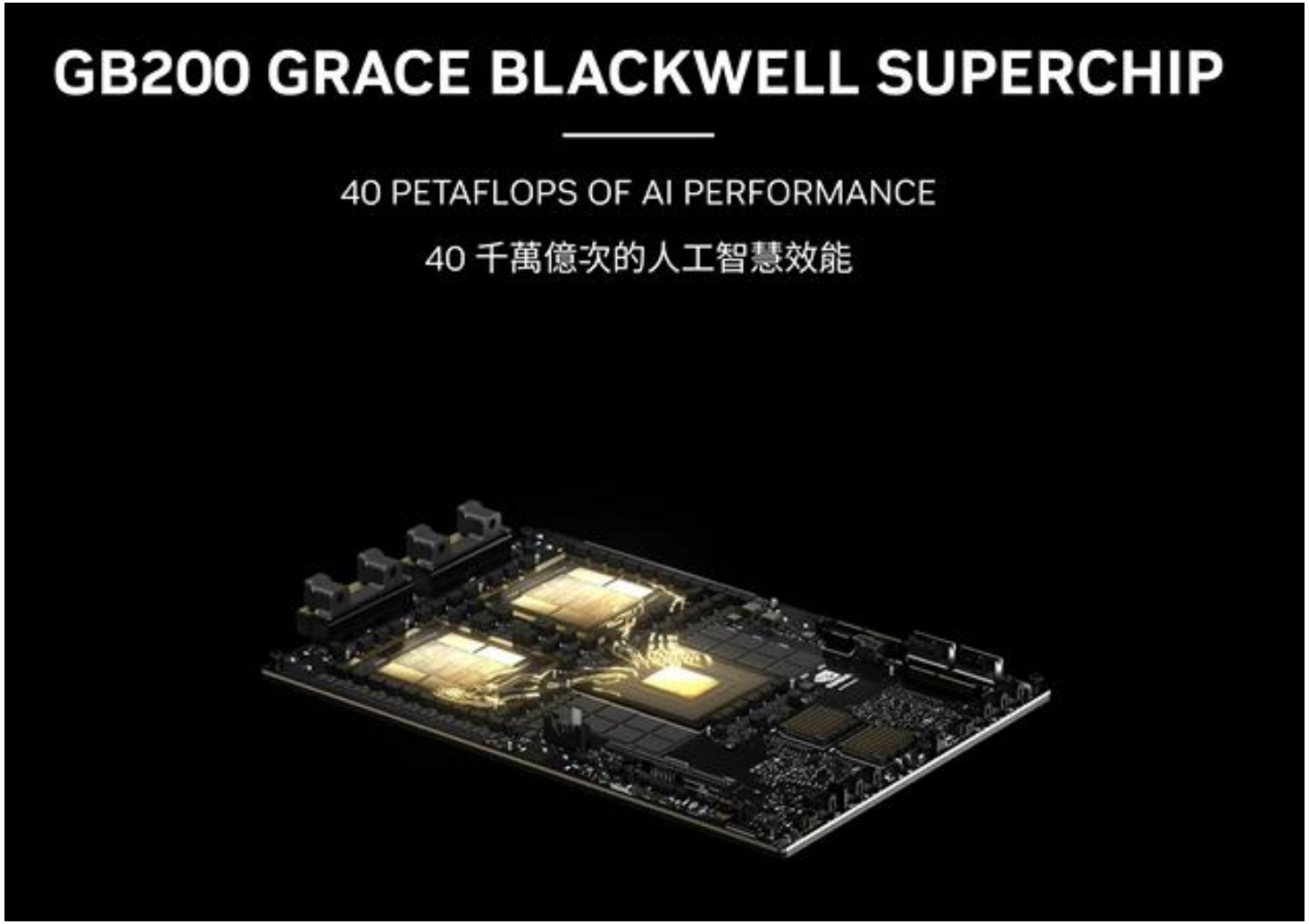
=

~145W

Unified Memory on Mac



COMPUTEX 2024 - NVDIA Blackwell 소개 중..

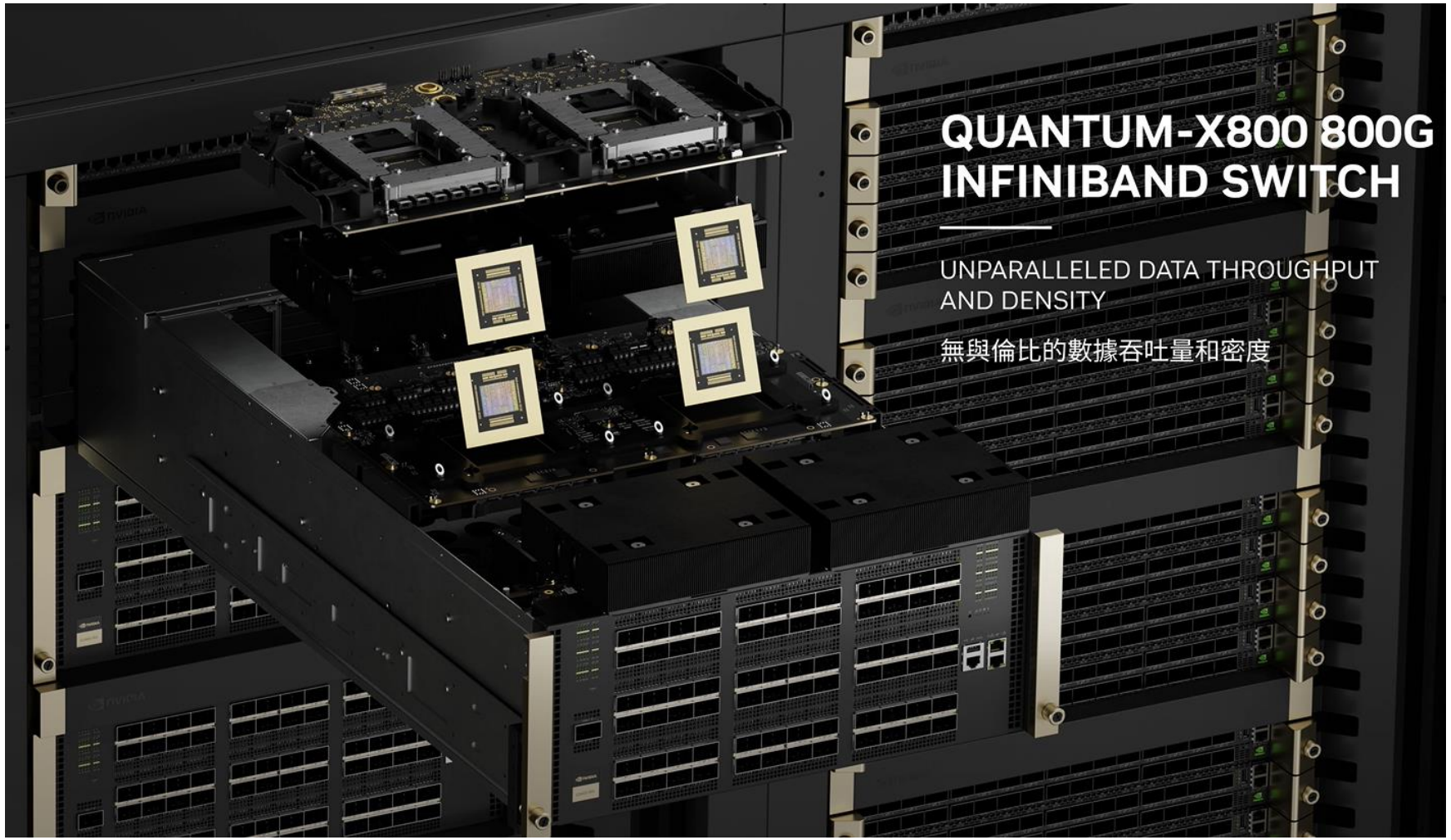


The AI Factory: How NVIDIA Builds the Engine for the New Industrial Revolution
https://www.youtube.com/watch?v=t_jiScDhf3M

COMPUTEX 2024 - NVDIA Blackwell 소개 중..

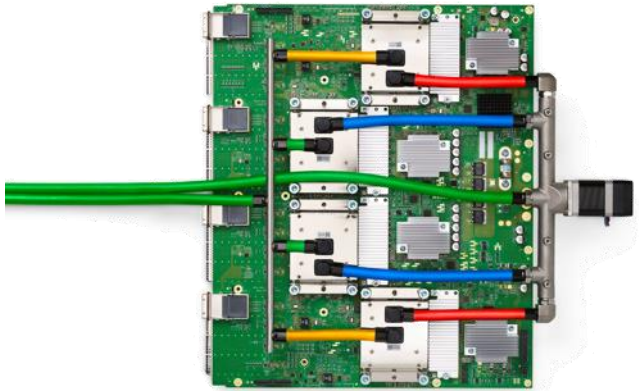
The AI Factory: How NVIDIA Builds the Engine for the New Industrial Revolution
https://www.youtube.com/watch?v=t_jiScDhf3M

COMPUTEX 2024 - NVDIA Blackwell 소개 중..



OCS Theme : Already Deployed in ML/AI Cluster

Google TPUv4



TPU v4: An Optically Reconfigurable Supercomputer for Machine Learning with Hardware Support for Embeddings

Industrial Product*

Norman P. Jouppi, George Kurian, Sheng Li, Peter Ma, Rahul Nagarajan, Lifeng Nai, Nishant Patil, Suvinay Subramanian, Andy Swing, Brian Towles, Cliff Young, Xiang Zhou, Zongwei Zhou, and David Patterson
Google, Mountain View, CA
{jouppi,gkurian,lsheng,pcma,rahulnagarajan,lnai,nishantpatil,suvinay,aswing,btowles,cliffy,zhoux,zongweiz}@google.com
and pattrsn@cs.berkeley.edu

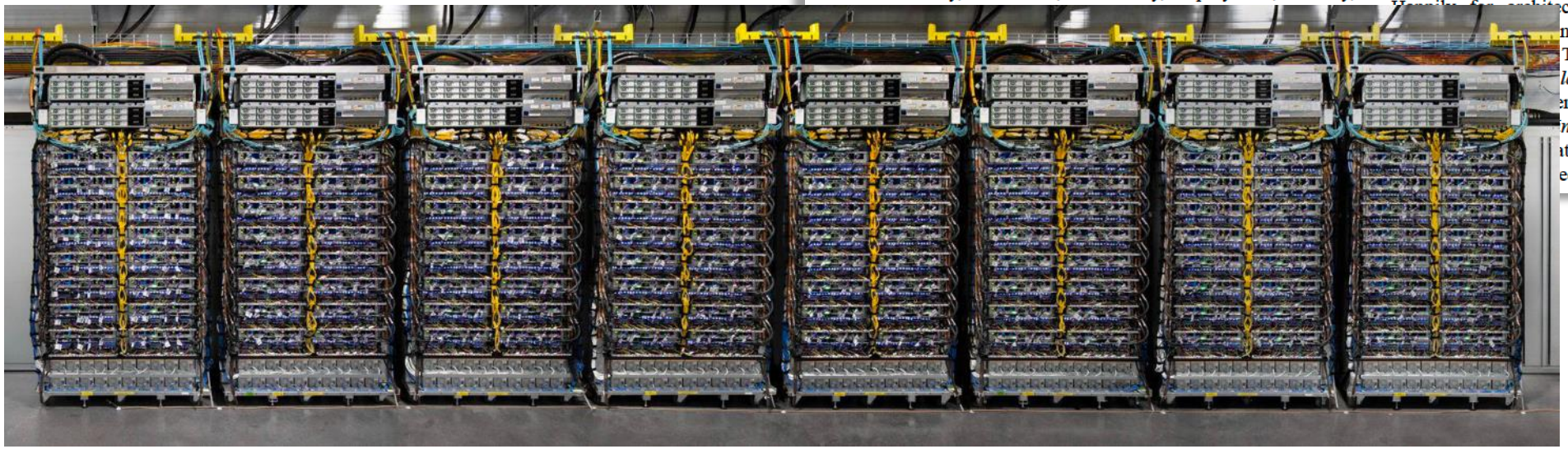
ABSTRACT

In response to innovations in machine learning (ML) models, production workloads changed radically and rapidly. TPU v4 is the fifth Google domain specific architecture (DSA) and its third supercomputer for such ML models. Optical circuit switches (OCSes) dynamically reconfigure its interconnect topology to improve scale, availability, utilization, modularity, deployment, security,

ACM Reference Format:
Norman P. Jouppi, George Kurian, Sheng Li, Peter Ma, Rahul Nagarajan, Lifeng Nai, Nishant Patil, Suvinay Subramanian, Andy Swing, Brian Towles, Cliff Young, Xiang Zhou, Zongwei Zhou, David Patterson. 2023. TPU v4: An Optically Reconfigurable Supercomputer for Machine Learning with Hardware Support for Embeddings: Industrial Product. In The 50th Annual International Symposium on Computer Architecture (ISCA '23), June 17–21, 2023, Orlando, FL, USA. ACM, New York, NY, USA, 15 pages. https://doi.org/10.1145/3579371.3589350.

1 INTRODUCTION

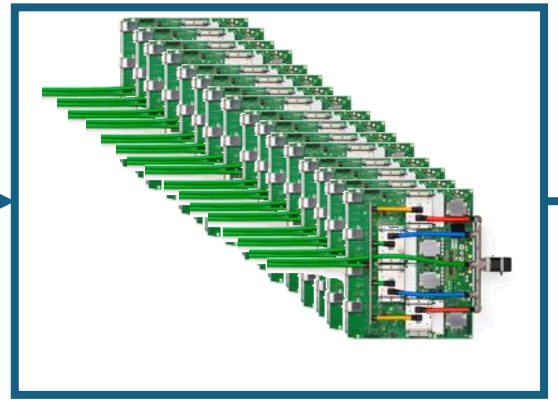
cts, machine learning (ML) models
n challenging ways, both in scale and
Table 1 and Section 7.7). Examples of
language models (*LLMs*) and examples
mbeddings necessary for recommender
ng recommendation models or *DLRMs*)
tions of Transformers and BERT. The
ecent LLMs [Bro20,Tho22,Nay21] has



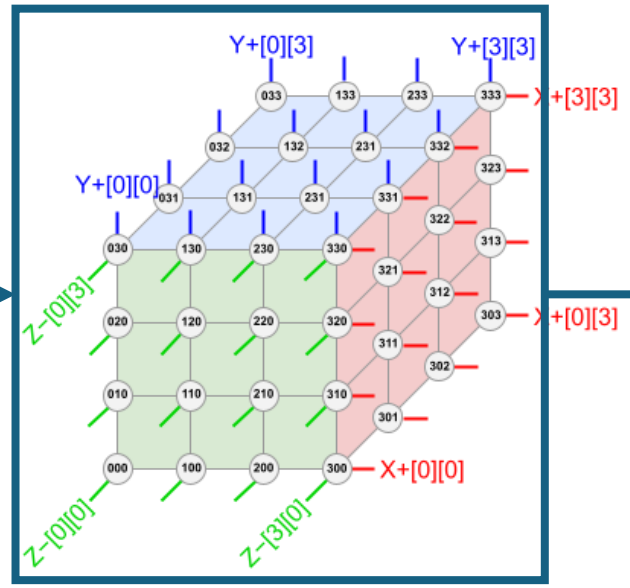
TPUv4 Architecture

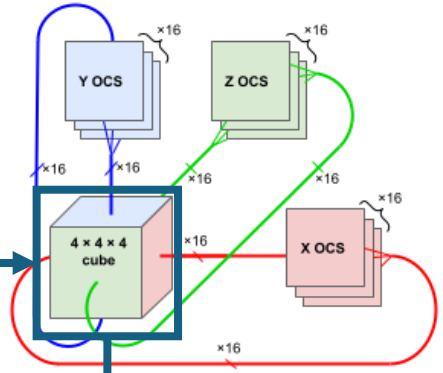
1 rack = 64 chips (chip = TPUv4)
1 row = 8 rack = 64x8 = 512 chips

8 row = 512 x8 = 4096 chips

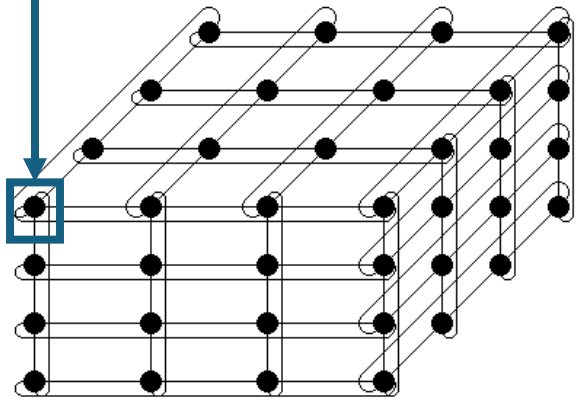


One 4^3 Cube
4x4x4 = 64-chips / Cube
6 x 16 i/f = 96 i/f





- Each rack contains 16 tray-host server pairs.
- 16 tray x 4 chips per tray = 64 chips
- Also 4 chip per CPU host.
- Each rack contains 64 chips and 16 CPU hosts (aka one 4^3 Cube)
- 4x4x4 3D electrical mesh per rack.



3D Torus

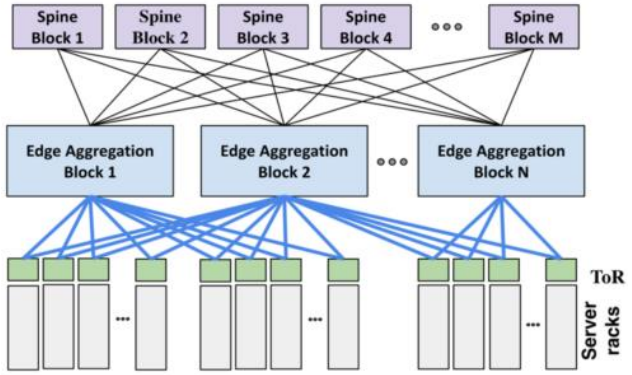
64 cubes = 64 x 64 chips per cube= 4096 chips

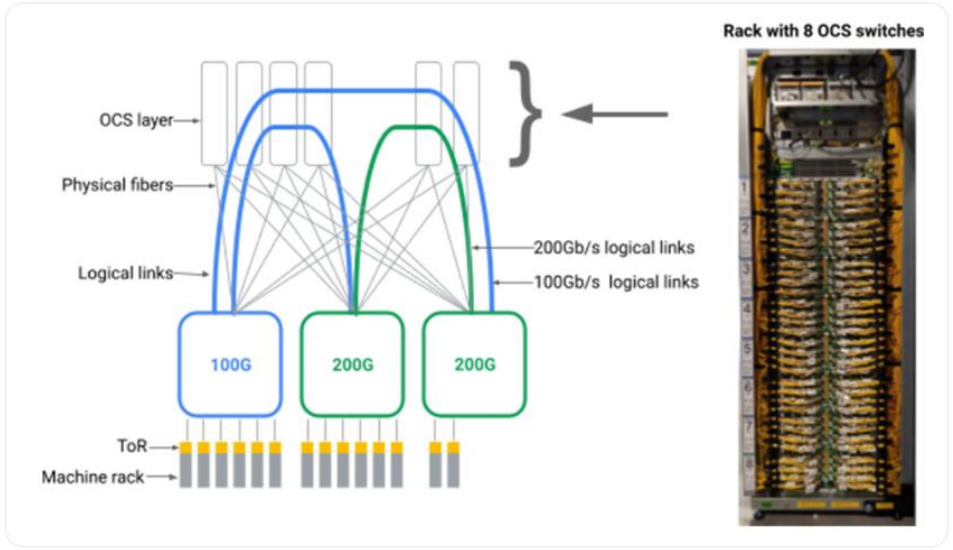
TPUv4 Architecture

다음 표는 v5e의 포드 사양 및 값을 보여줍니다.

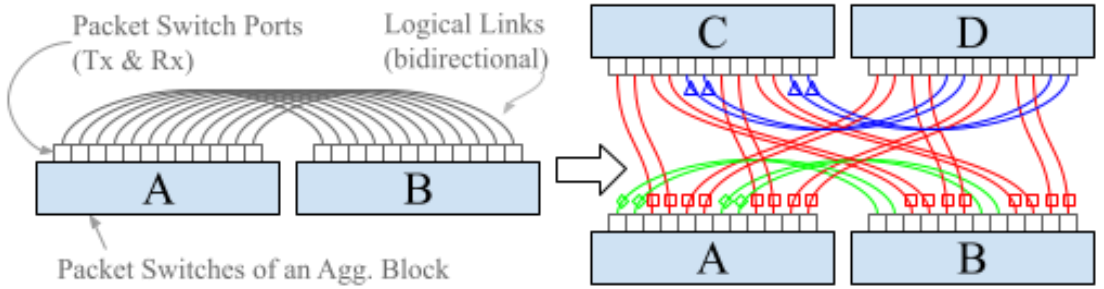
주요 포드 사양	v5e 값
TPU Pod 크기	칩 256개
상호 연결 토폴로지	2D 토러스
포드당 최고 컴퓨팅	100페타옵스(Int8)
포드당 올리듀스 대역폭	51.2TB/초
포드당 바이섹션 대역폭	1.6TB/초
포드당 데이터 센터 네트워크 대역폭	6.4Tbps

Google DC 내부 개선 사례 – “Mission Apollo”

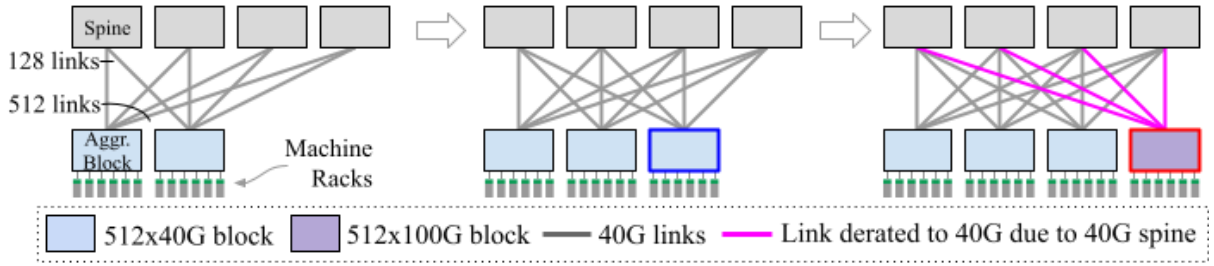




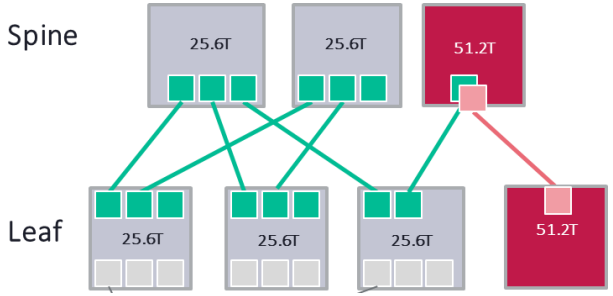
Pain #1: Rewiring



Pain #2: Spine Limit

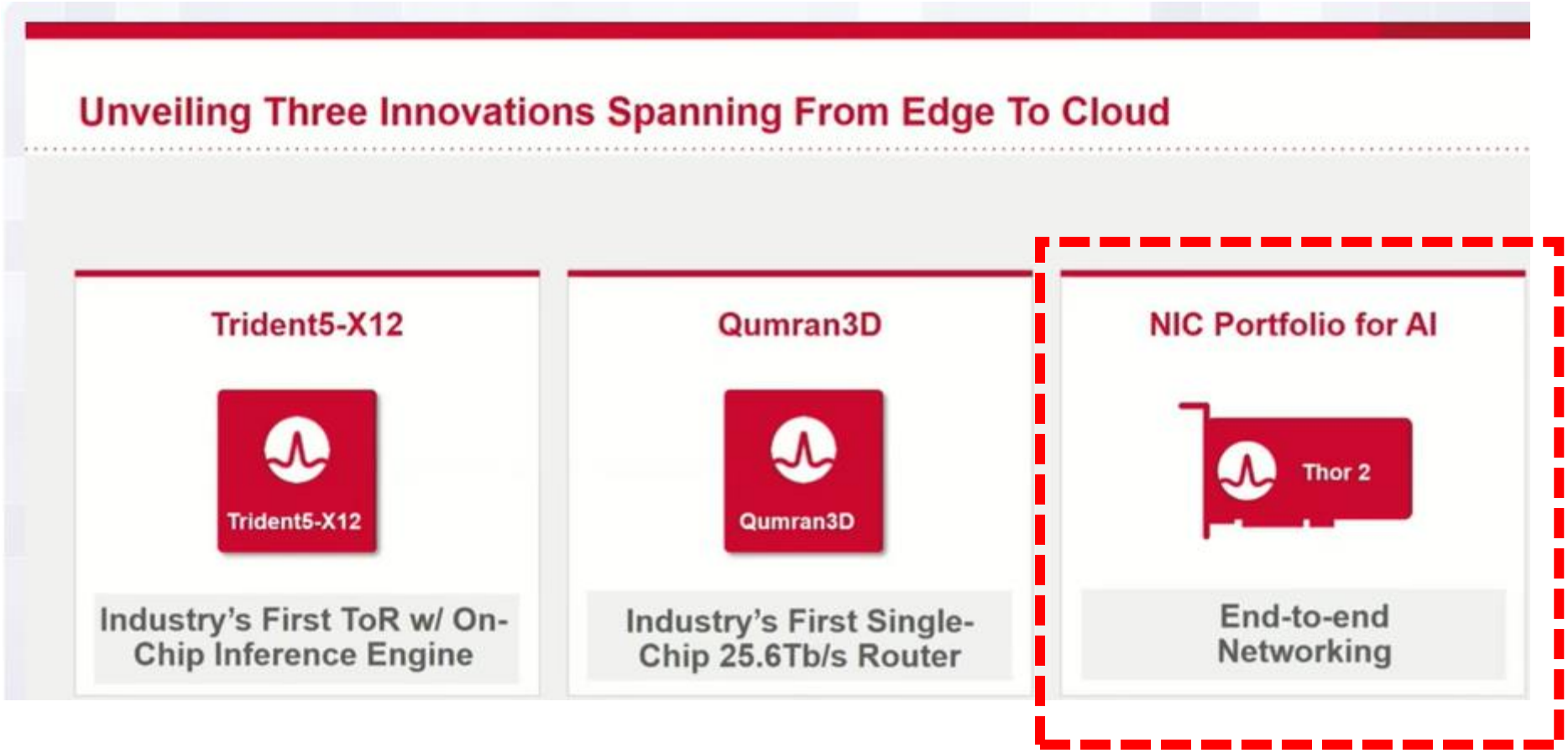


Pain #3: Retrofitting Optics

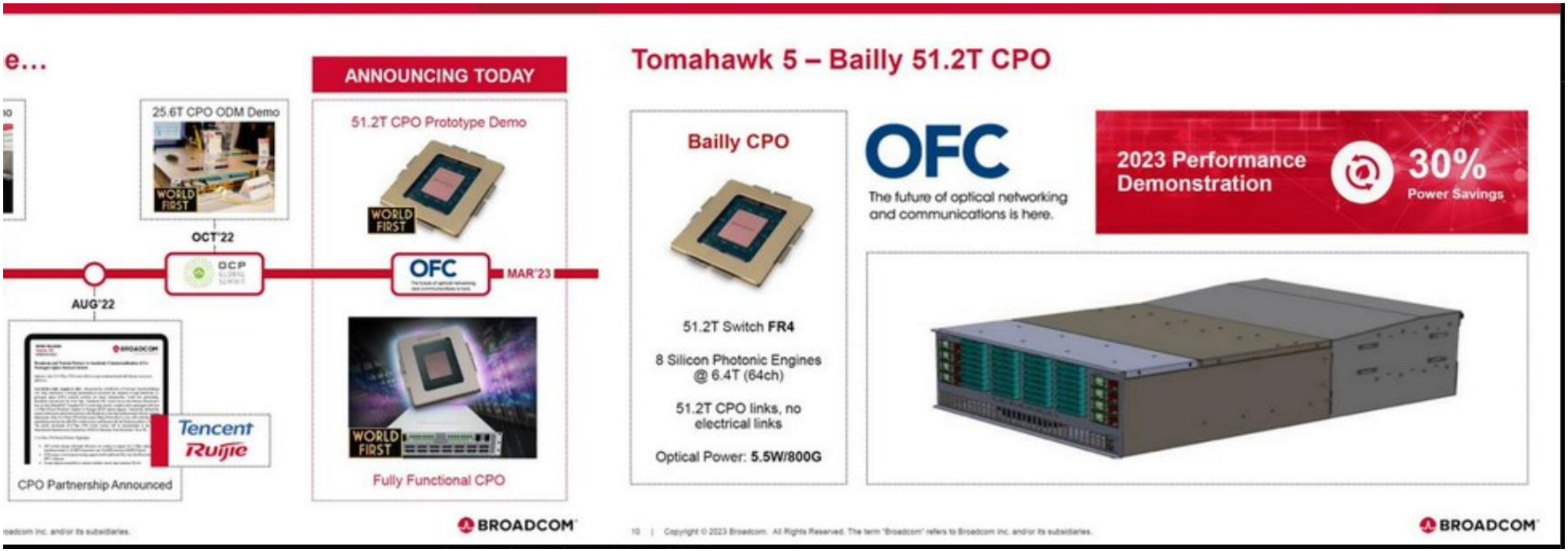


What’s NEXT?

Broadcom Roadmap 소개



Broadcom Roadmap 소개



*https://x.com/ogawa_tter/status/1768455026991895031/photo/4

400/800G Client pluggable은 왜 비쌀까?

Examples of optical front-end types and their subcomponents are outlined in the table below. There are many more varieties (e.g., LR1, SR4, etc.), but they follow the same patterns as shown.

Transceiver type	Laser Type	Number of Lasers	Fiber Type	Relative Cost
400G-SR8	50G VCSEL	8	8xMMF ribbon	$
400G-DR4	100G wideband laser (DML)	Usually, 1, split 4 ways	4xSMF ribbon	$$
400G-FR4	100G CWDM laser (EML)	4	1xSMF	$$$
400G-LR4	100G high-power CWDM laser (EML or SiPho)	4	1xSMF	$$$$
800G-SR8	100G VCSEL	8	8xMMF ribbon	$$
800G-DR8	100G wideband laser (DML or EML)	Usually, 1, split 4 ways	8xSMF ribbon	$$$
800G-2xFR4	100G CWDM laser (EML or SiPho)	Usually, 4, split 2 ways	2xSMF	$$$$

400/800G Client pluggable은 왜 비쌀까?

OOK

On/Off Keying

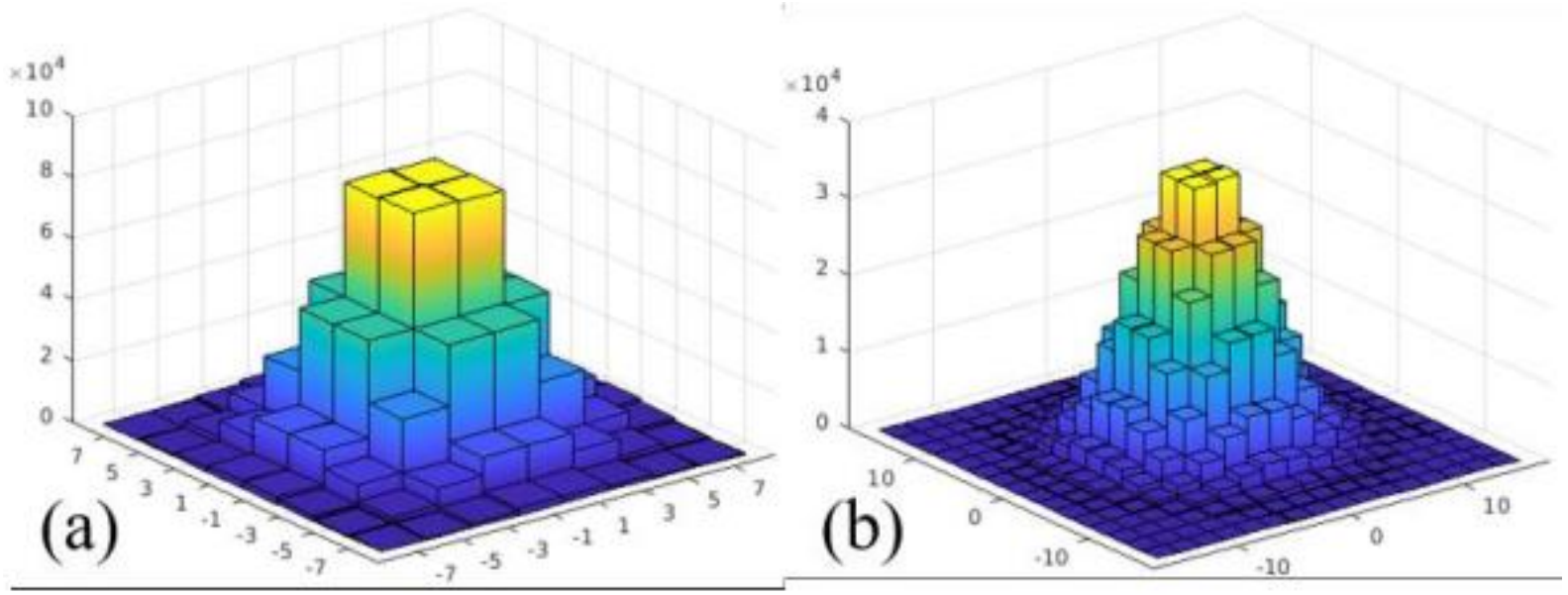
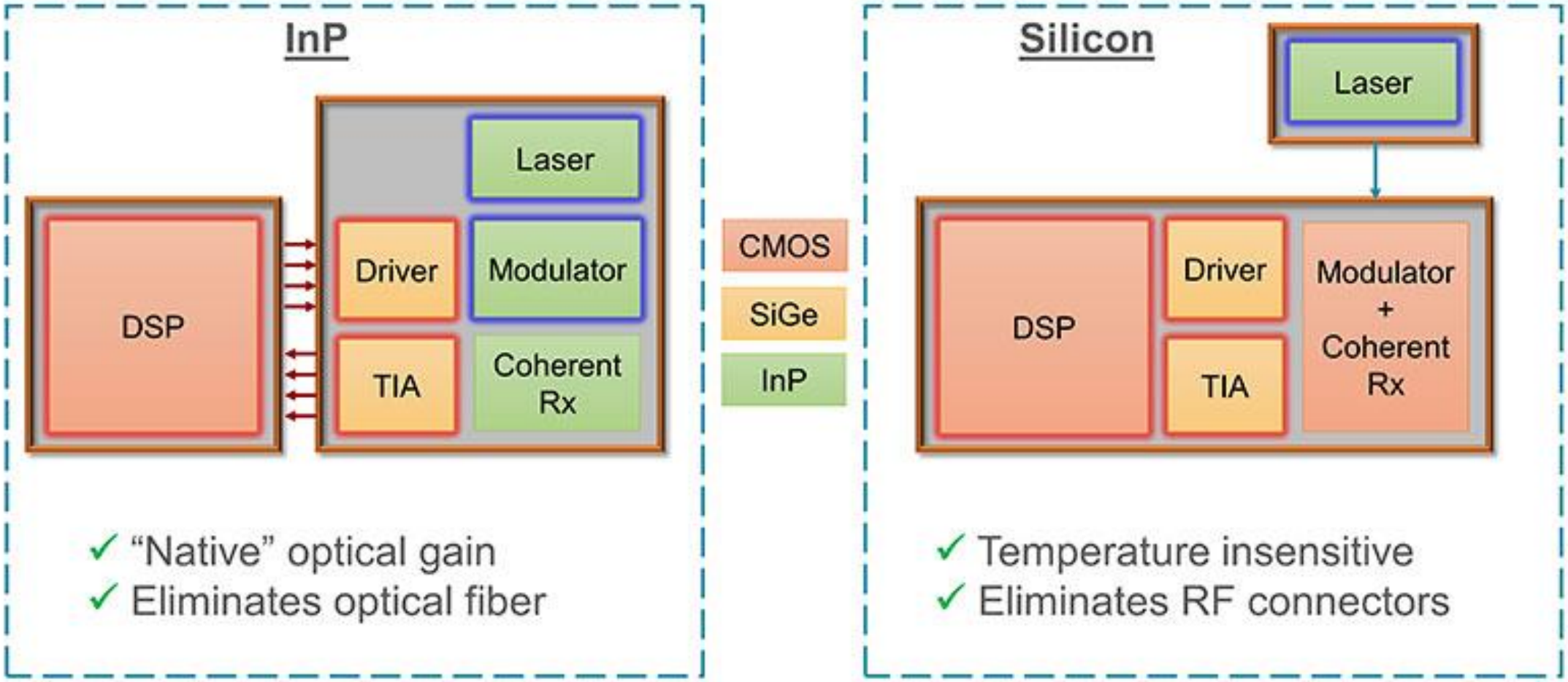


Fig. 1 Histogram of (a) PCS 64-QAM with $\nu = 0.0749$.

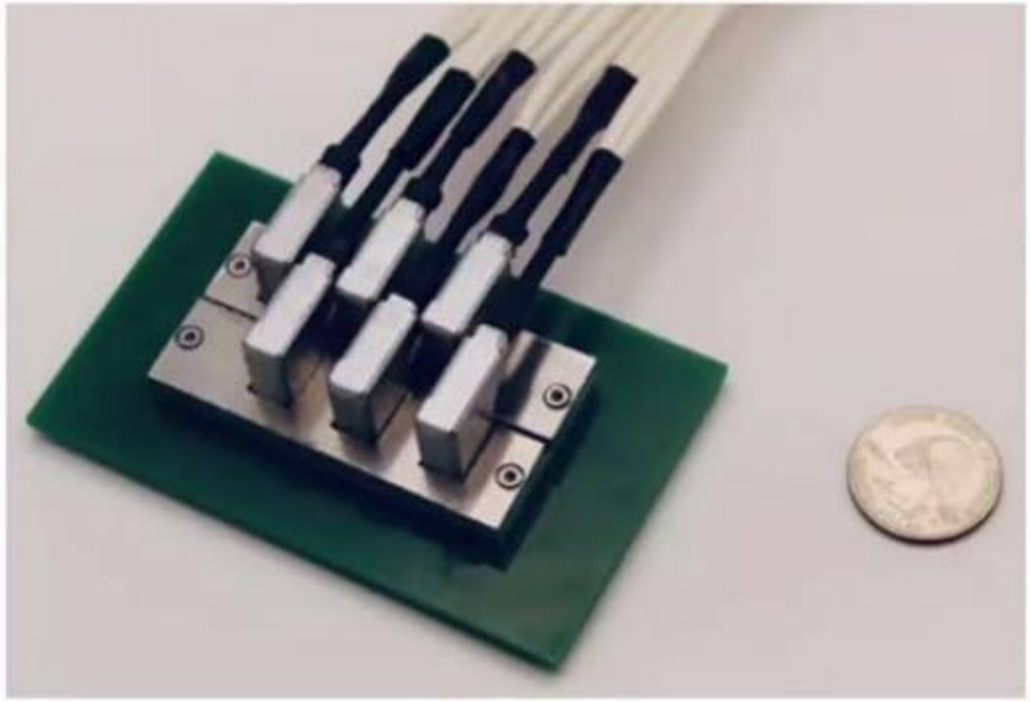
IMDD

DSP

400/800G Client pluggable은 왜 비쌀까?



400/800G Client pluggable은 왜 비쌀까?

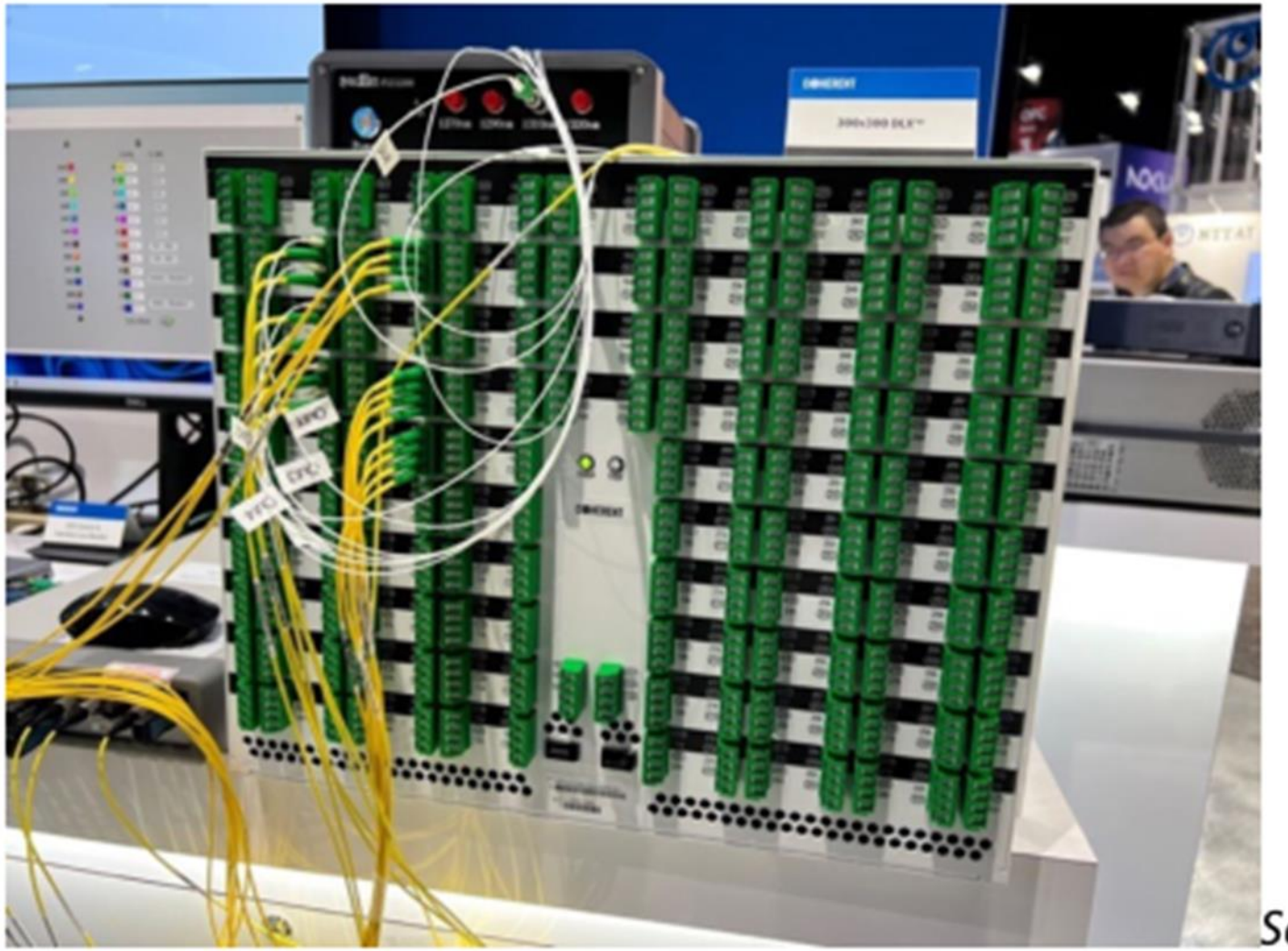


10 Tbps
(on half a business card)

8 Tbps
(ML/AI PCIe Card Interface)

Source: Nubis Communications

Coherent 社 광 스위치



Source: Coherent

노이즈 = 딜레이 = 투자비

감사합니다