Optical Network Technologies in the AI Era

Quoc Thai NGUYEN

Paris, 24 October 2025

HUAWEI

Powered by

Optical Transmission
OptiXtrans

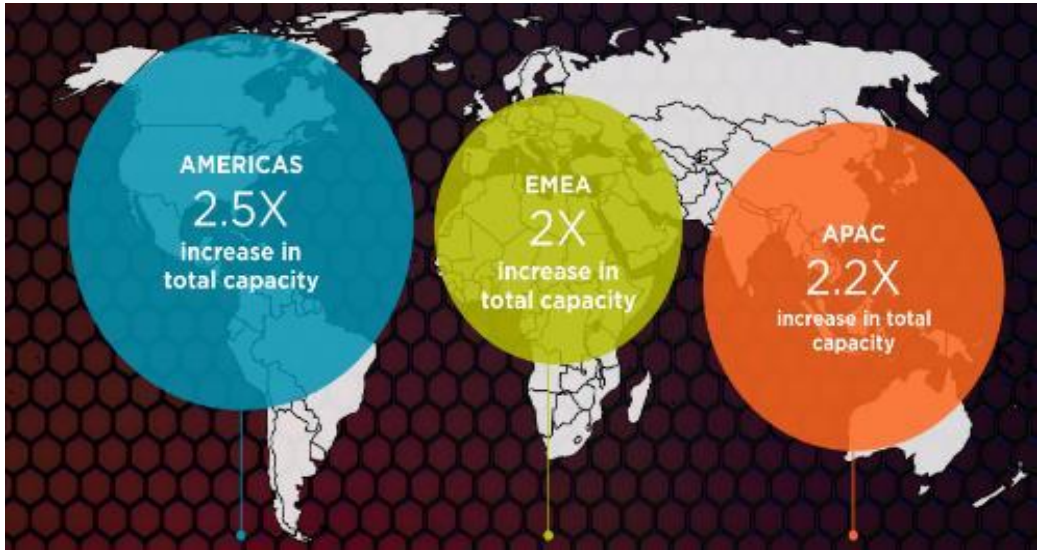
Optical Access
OptiXaccess

Optical Terminals
OptiXstar

Contents

The Acceleration of AI DC Construction is Driving Rapid Growth in DCI Traffic

Globally planned new DC capacity is more than double of the current installed base

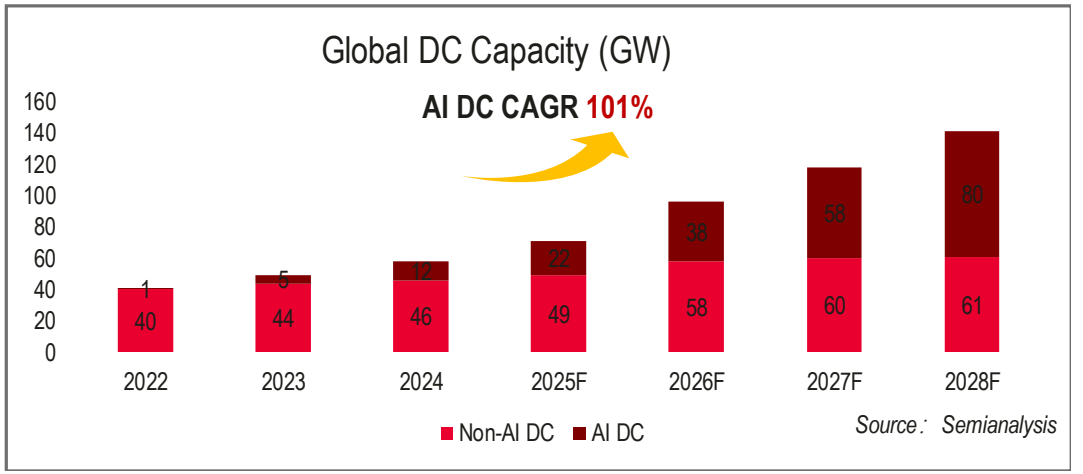


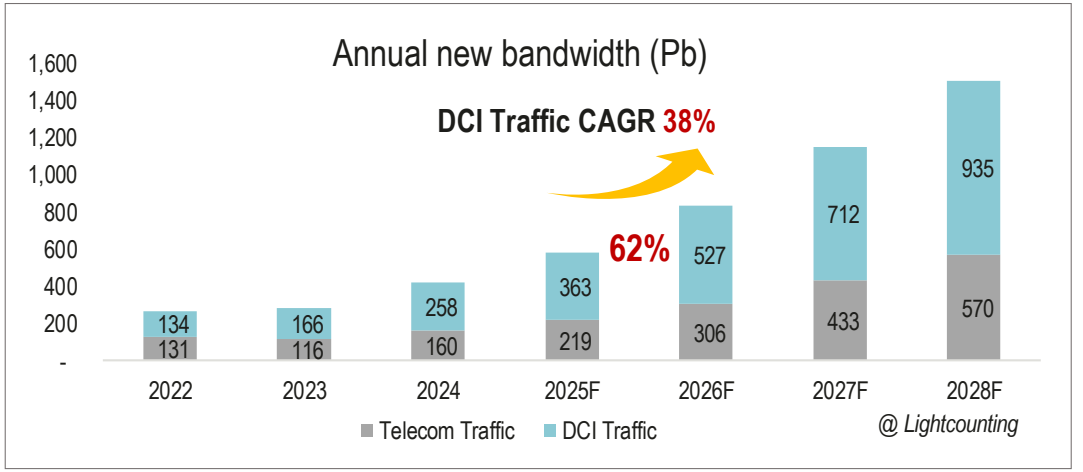
Operational Capacity	**16,820MW**	**6,218MW**	**10,584MW**
Vacancy	5%	9%	16%
Development Pipeline	**24,794MW**	**6,528MW**	**13,281MW**

Source: CUSHMAN & WAKEFIELD "2024 Global Data Center Market Comparison"

New DCs are predominantly (~83%) adopting the colocation mode, simultaneously driving demand for DCI bandwidth leasing.

Global AI DCs are growing at an annual rate exceeding 101%, driving a 38% CAGR in DCI traffic

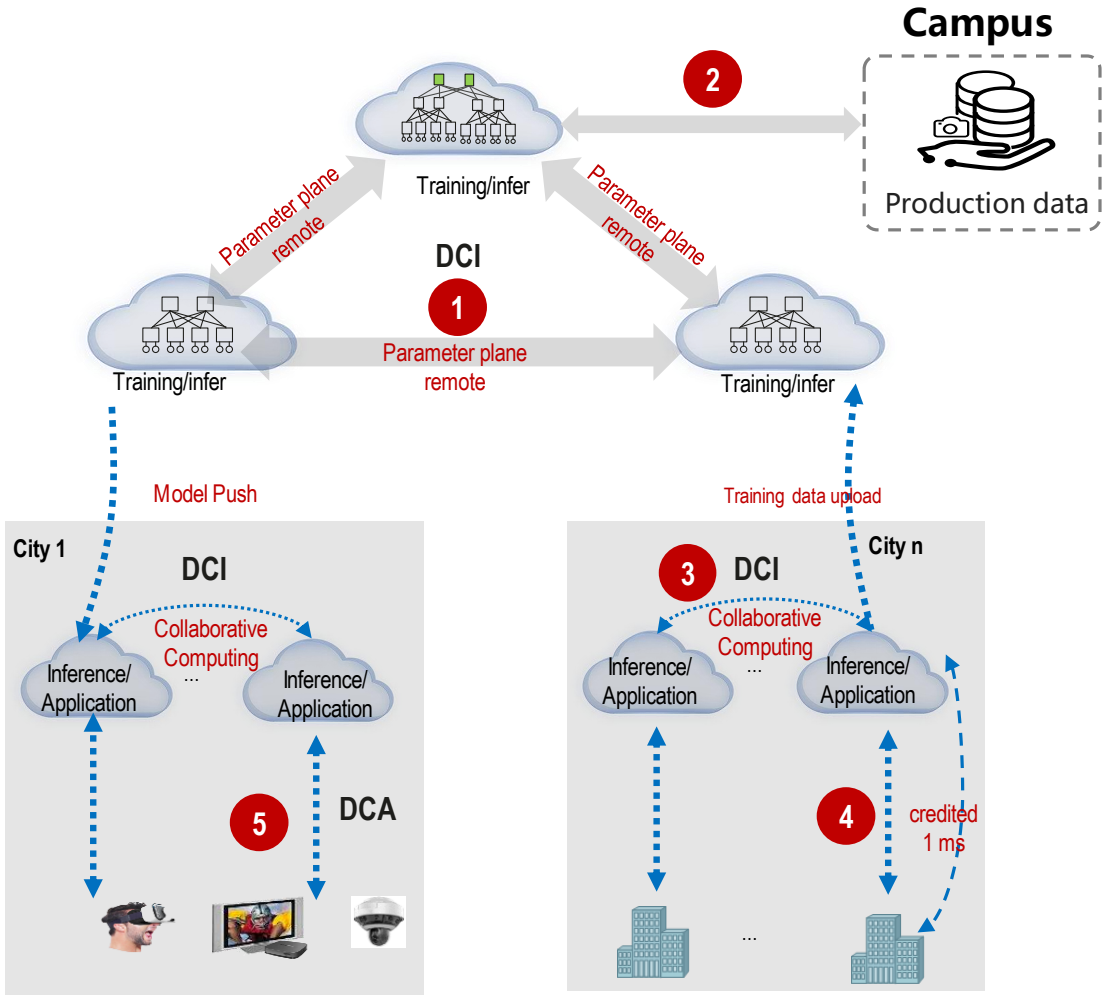




HUAWEI

Basic Protocol Change from Traditional TCP/IP to RDMA in AI Service

Network Changes in the AI Era: Efficient Cross-DC Collaboration and Strong Device-Edge-Cloud Interaction



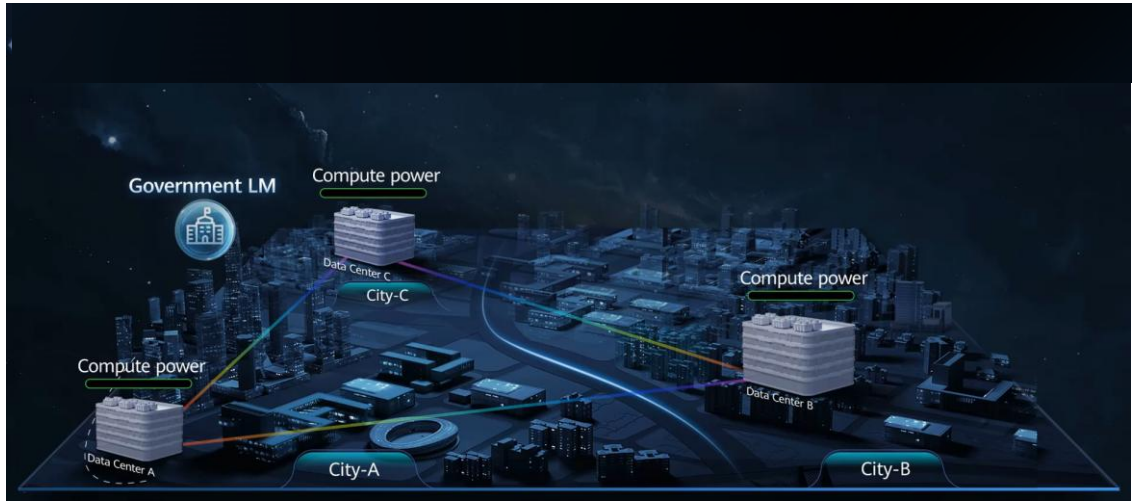
RDMA is Mandatory for AI Service, Guarantee Deterministic Experience

	Scenario	Protocol
1	**Distributed Computing** Parameter plane remote training	RDMA
2	**Storage computing separation** Sample & Parameter surface remote training	RDMA
3	**Data backup** Training data upload and model push	RDMA, TCP, UDP
4	**Inference private line** Internet, finance, government production	RDMA, TCP, UDP
5	**Terminal-cloud synergy computing** Cloud Internet cafe, cloud computer...	RDMA,TCP, UDP

HUAWEI

Remote Compute Scenarios: DCI Need Lossless and Ultra-Broadband

Remote Compute Scenario



Packet loss leads to Computing Loss

10 ^ (-3) → 50%

Packet loss → computing power loss

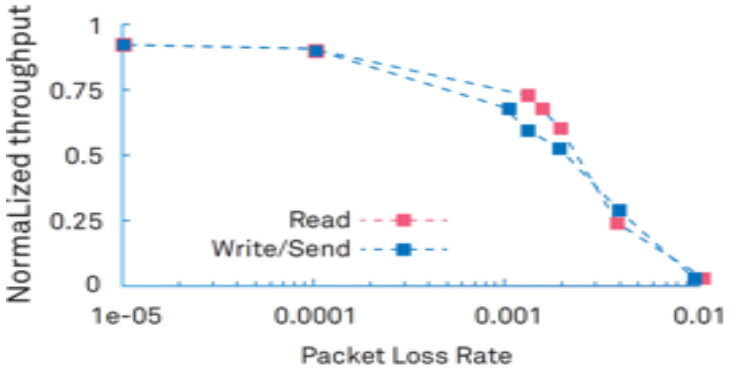
AI cross-DC bandwidth surges

200G / card → 1600Tbps

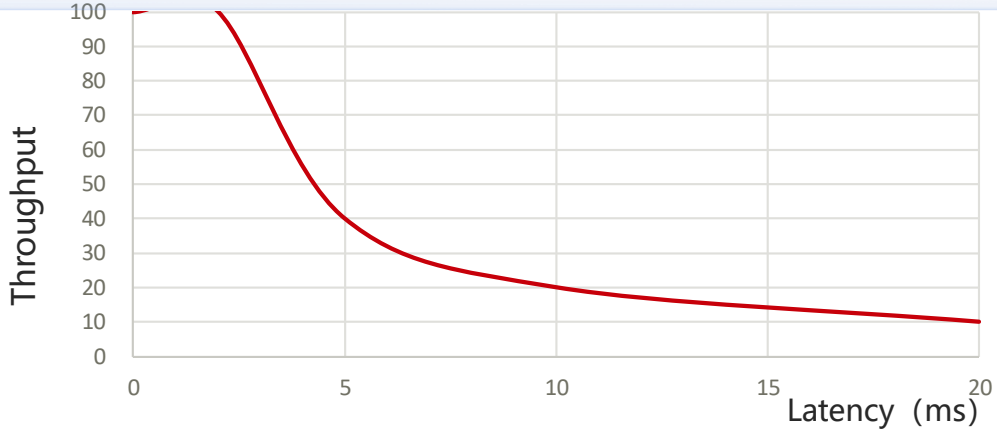
16K card model → non-convergence

Packet loss and latency cause throughput decrease of RDMA

Transmission packet loss leads to the decrease of throughput



Single-stream throughput decrease as latency increase



HUAWEI

Multi-DCs Collaboration Drive DCN Traffic “Overflows” to DCI

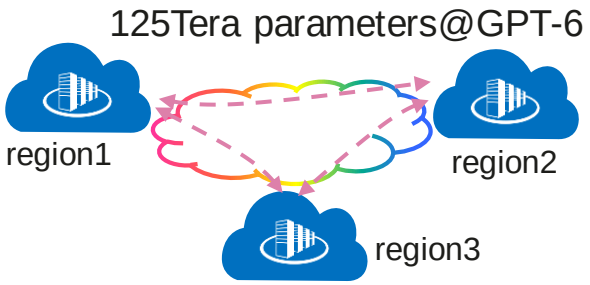
In Campus 1~10km-level
n*10K >> 100K training cards Cluster

- **Single building limitation:** power supply, space and heat dissipation etc.

In City 100km-level
n*1K >> 10K training cards Cluster

- **L1~L2 Small &Medium model training:** inner city computing power Collaboration

Cross-region 1000km-level
n*10K >> 100K training cards Cluster



- **L0~L1 Large Model training:** Collaboration of large clusters with 10K cards in multiple regions

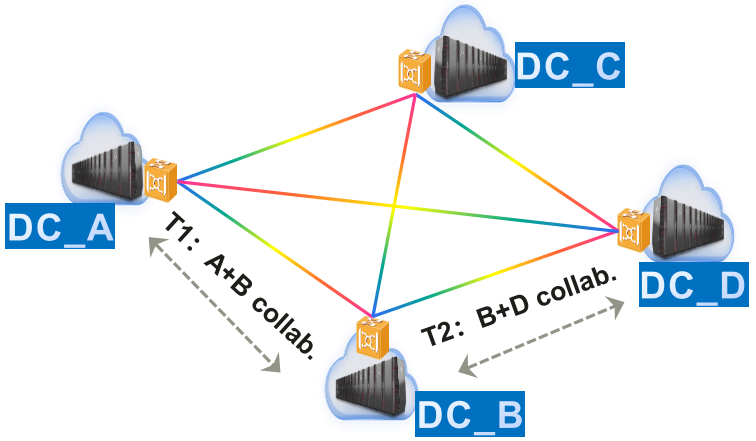
New Requirement to Transmission

“0” Package Loss

- No degradation of bandwidth , 99.999% availability
- 1/1000 package loss lead to 50% compute loss

Bandwidth

- Single-stream throughput=NIC bandwidth
- Collaborative typical BW: ~100T@10K Cards



Latency

- Distributed AI training is latency sensitive service
- PP & DP need stable low latency

Agility

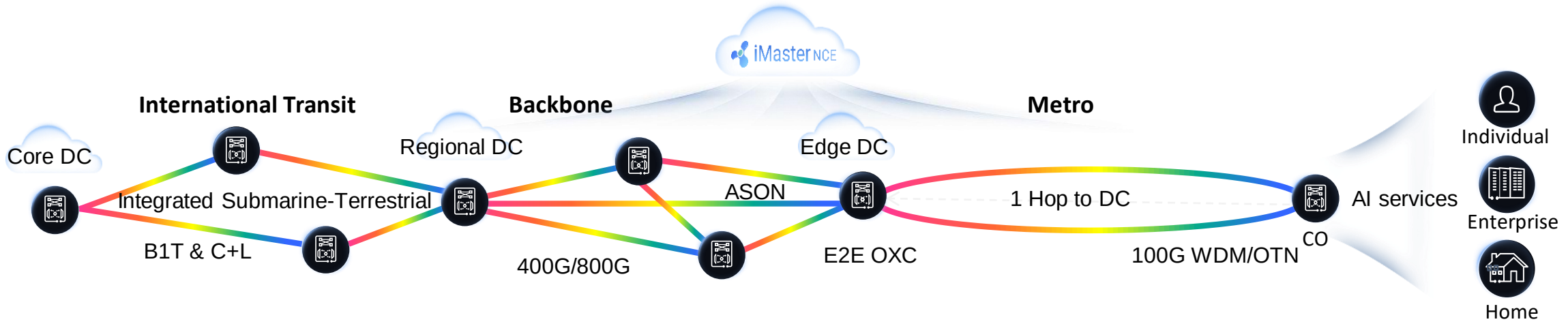
- Connection & BW are defined by training task
- Scheduling: Static => Dynamic / On-demand

HUAWEI

Contents

All-Optical Network Architecture for the Evolution into AI era

DCI traffic CAGR > 38%	18+ Carriers start GPUaaS	2x price by 99.99%->99.999%
B1T Ultra-broadband Backbone	**Optimal TCO & Ultra Elastic Connection**	**Intelligent OTN & AN**



Networking	Service Transmission	Network Mgmt & Ctrl
Ultra Capacity & Connectivity	**Ultra Elastic & Secure Connection**	**Intelligent OTN & AN**
B1T Ultra-broadband & OTN Cluster **Unified C+L**	**Quantum Encryption for 2B Access**	**Zero-Outage Network**

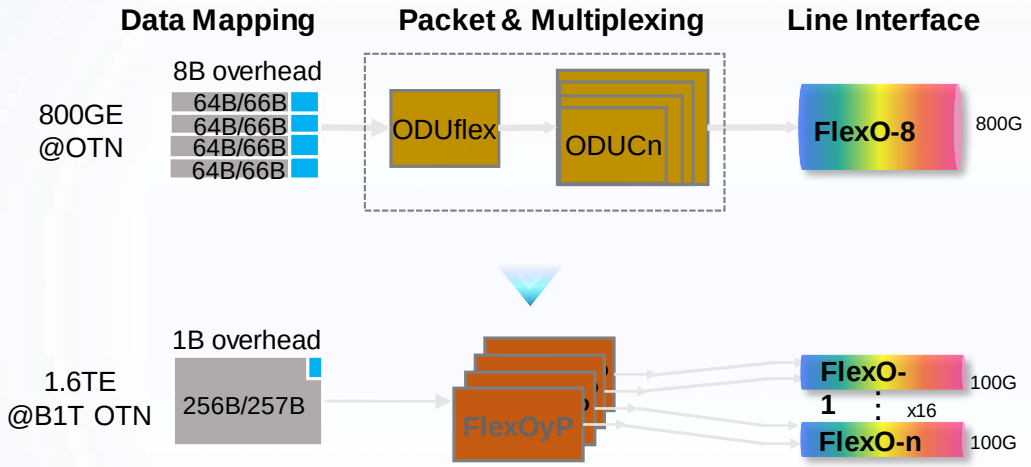
B1T *: Beyond 1Tb/s*
ASON*: Automatically Switched Optical Network*
OXC*: All Optical Cross-Connect*

HUAWEI

B1T OTN: Enabling the Next-Generation Optical Transport Network

B1T OTN New Protocol Powers 1.6TE Transport

1.6TE Ready, 20% ↑ Performance with Same FEC



Higher Efficiency

- 64B/66B → 256B/257B Mapping
- Lower Overhead, Efficiency ↑

Lower Latency

- ODUCn Single Frame → FlexOP Multi-Frame
- Simplified Layers, Site Latency 50% ↓

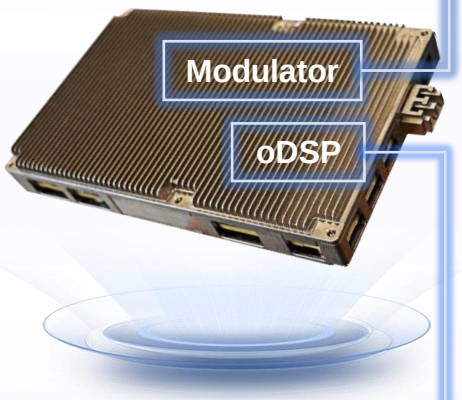
Better Flexibility

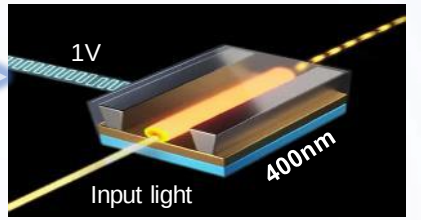
- Rigid Multiplexing → Elastic Connection
- Flexible N×100G Granularity

ITU-T G.709.b1t (Under development)

Materials & Algorithm Innovation Leads 2T/λ

New Modulator Materiel to Improve Baud Rate

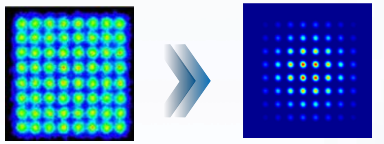




240G+ Baud, TFLN - Modulator

New Algorithm to Improve Distance

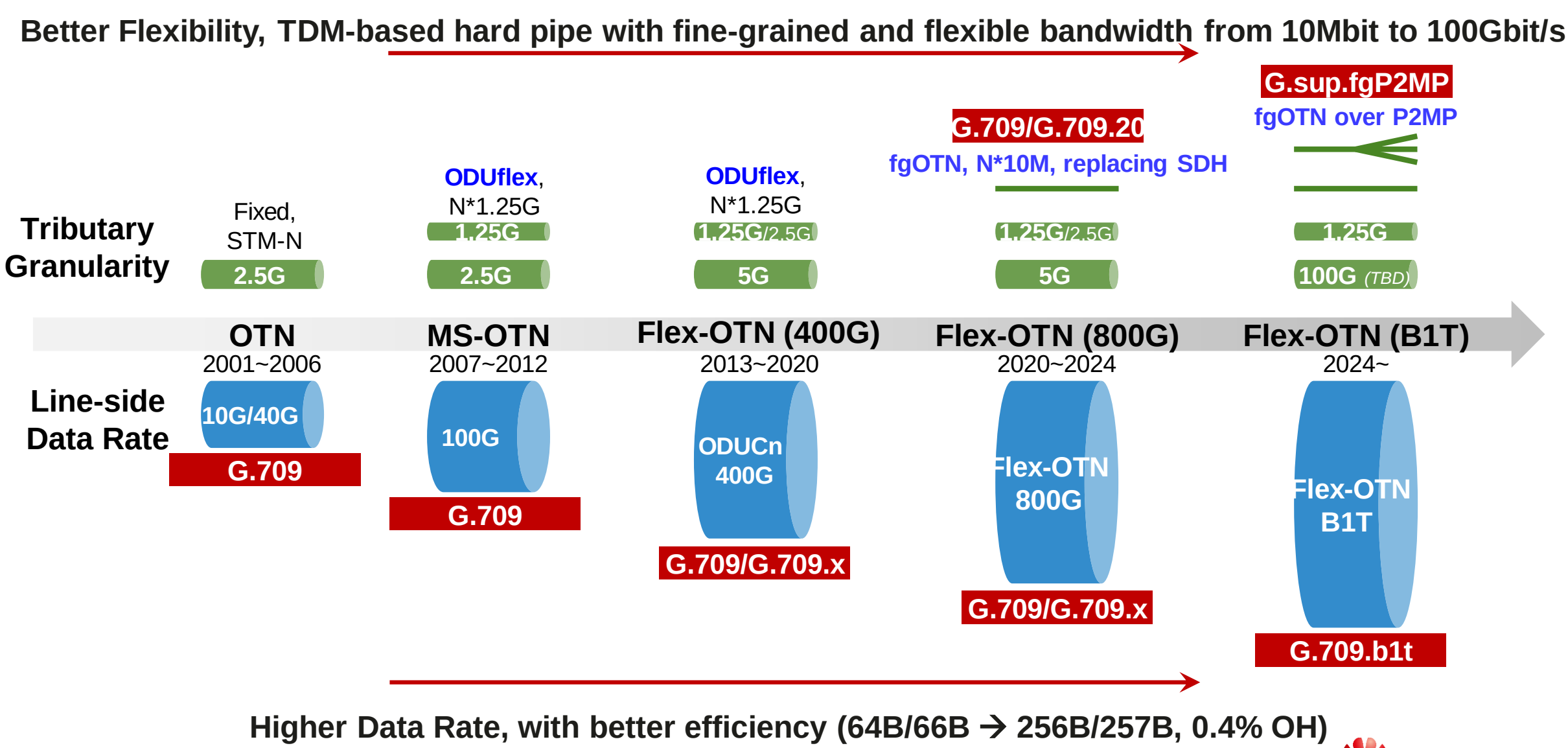
800G~2T Optical Module



PBC Non-linear Compensation

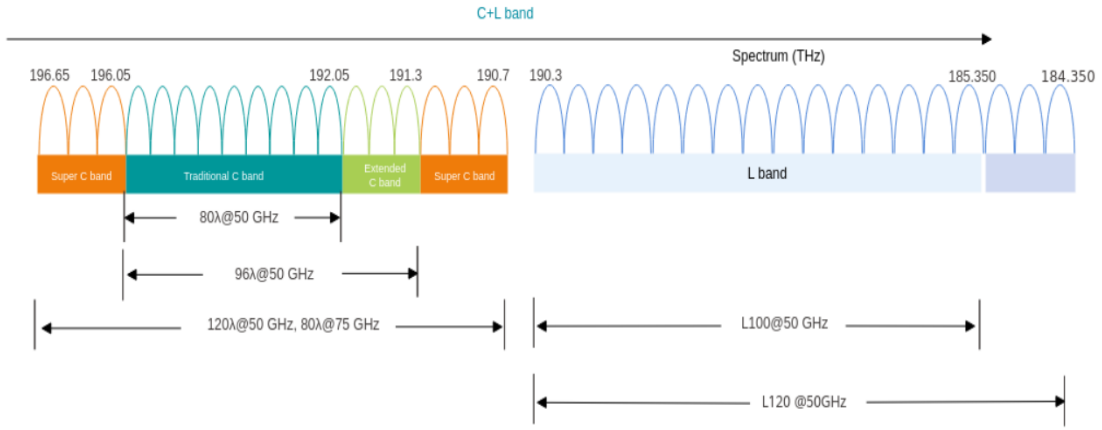
HUAWEI

OTN Standard Evolution: Higher Rate & Better Flexibility



C+L Optical Transmission System: Key Technologies and Challenges

Huawei C120+L120: capacity increases 25%



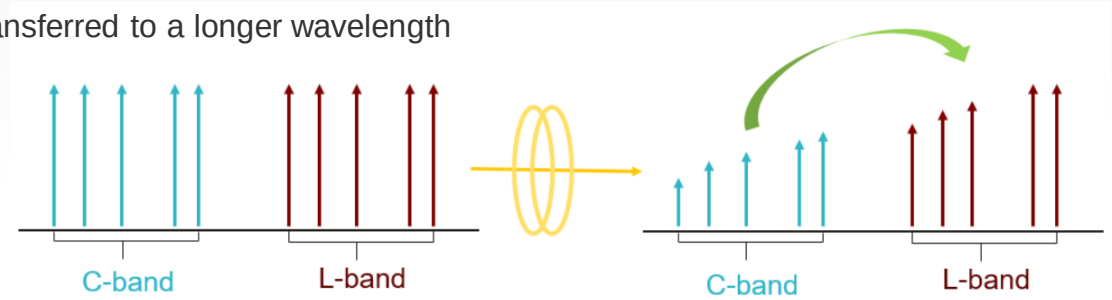
Key Technologies and Challenges

C+L band system, solution design have the following challenges

- **(1) Stimulated Raman scattering (SRS) control**
- **(2) Wide-spectrum low-noise amplification**
- **(3) Smooth L-band deployment on live networks**

SRS affects wavelength flatness

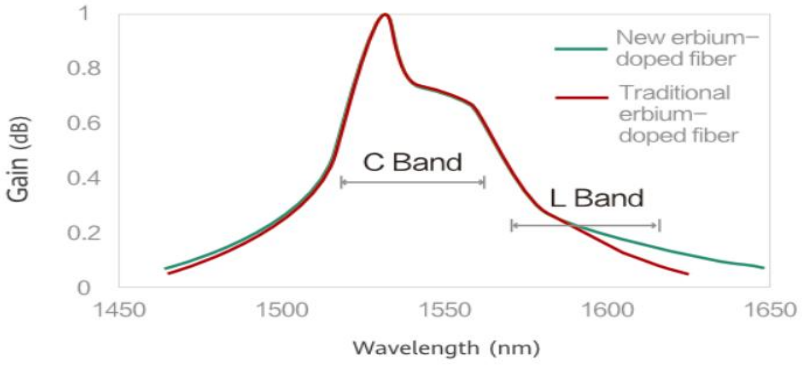
SRS is a non-linear effect of optical fibers, the energy of a shorter wavelength is transferred to a longer wavelength



How to achieving the optimal system performance ?

Noise figure of the L-band is larger than C band

Gain efficiency of L-band EDFA is relatively low, performance of L-band OAs deteriorates significantly compared with C-band

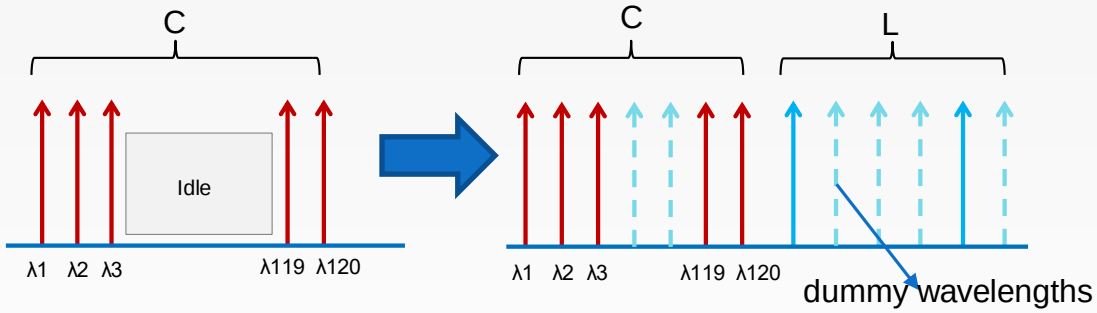


How to Improve the Performance of L-Band EDFA ?

HUAWEI

Dummy Light (DL) Filling and Ultra-wide L-band amplification technology

DL filling is a key technology of the C+L system



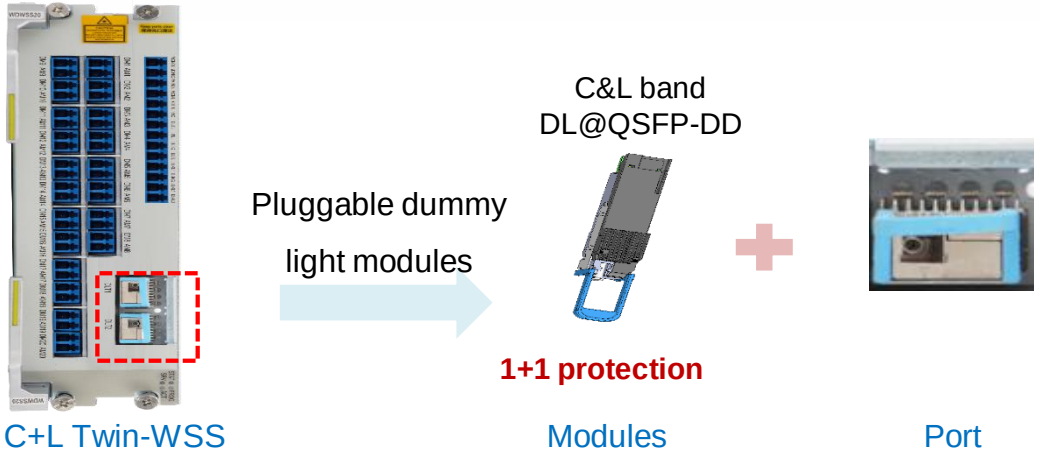
C Band wavelength system

Unused wavelengths can remain idle

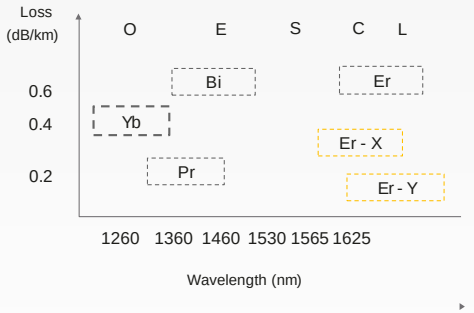
C+L Band wavelength system

Full-wave transmission, must be filled with dummy wavelengths

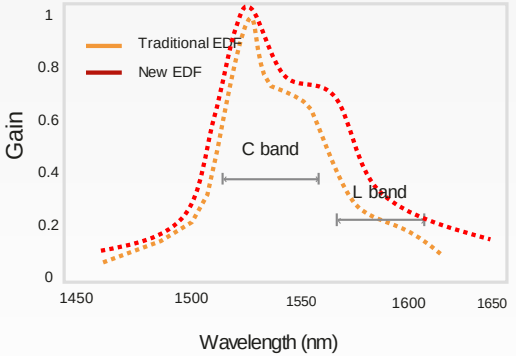
Huawei's pluggable DL solution



Innovative doping components, supporting ultra-wideband amplification



New elements are added to improve performance



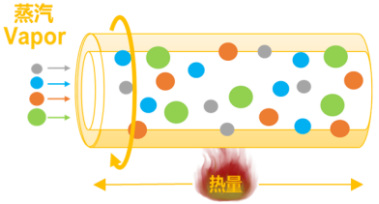
10,000+ tests, optimal ratio of doping elements

Innovative manufacturing process, optimizing amplification performance

Solution immersion

Low doping concentration
Uneven distribution

Vapor deposition process

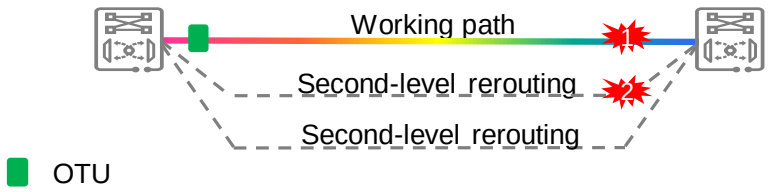


The **concentration and uniformity** of doping elements are optimized to improve amplification efficiency.

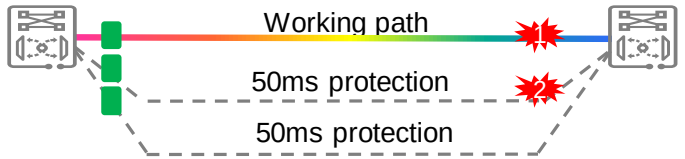
HUAWEI

Network Resilience: Enhanced WSON for 50ms Deterministic Optical Rerouting

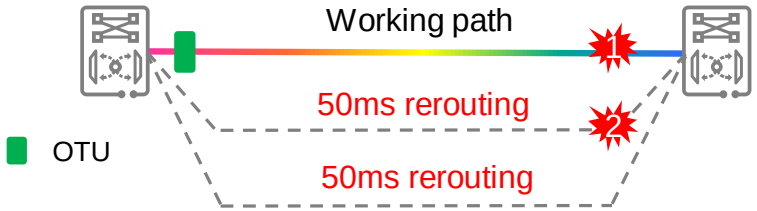
Traditional WSON: Insufficient reliability



OTN ASON Permanent 1+1: High cost



Enhanced WSON

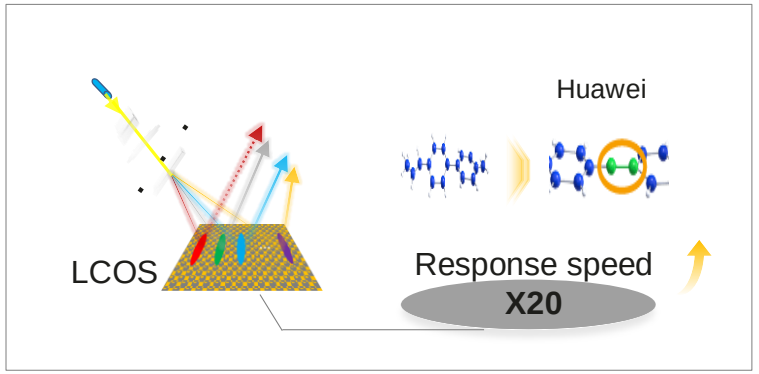


Highlight:

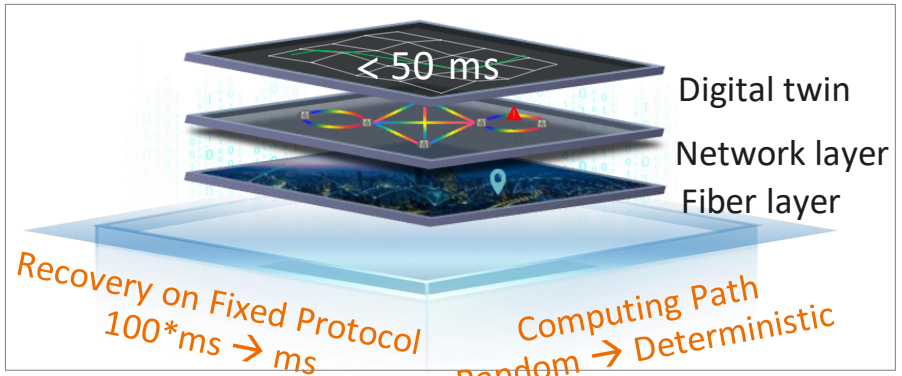
- Reliability is equal to OTN ASON Permanent 1+1
- Cost is equal to Traditional WSON

Key technological points

Fast-Response LCOS: Seconds → **< 50ms**



Intelligent Algorithm: Digital Twin Routing

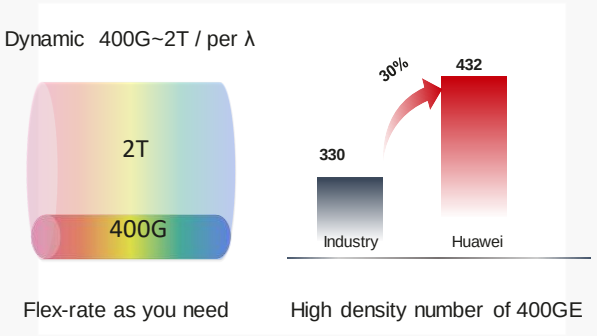


ETSI F5G(25)023024: Proposal to initiate a work item of 50-ms WSON-based wavelength rerouting to build a highly reliable optical transport network

WSON: Automatically Switched Optical Network

DC-OTN Builds All-Optical Lossless DCI for AI Computing

Ultra-high bandwidth



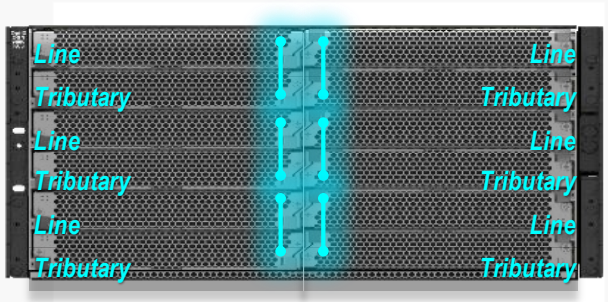
Flex-rate as you need

High density number of 400GE

All-optical lossless DCI

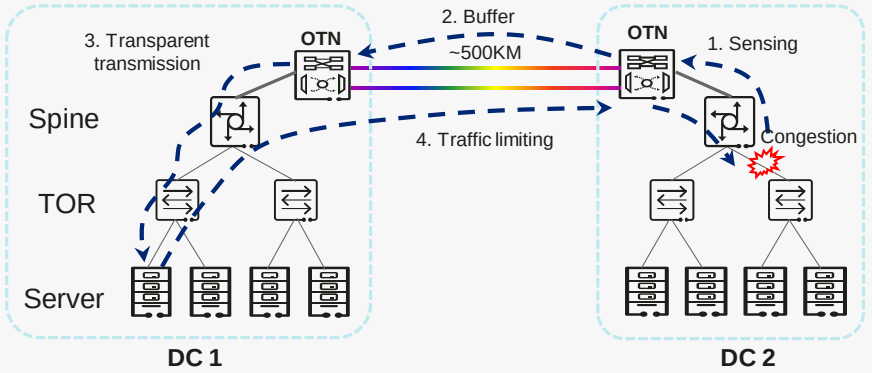
Low Latency

Simplified OTN architecture



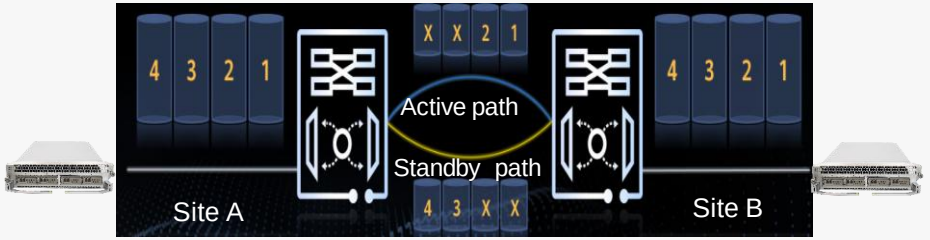
- Cross-connections supported by paired slots
- Power 30% ↓ latency 30% ↓

RDMA “0” Packet Loss



- ABS flow control algorithm to detect congestion and manage traffic
- Buffers packets to achieve “0” packet loss in the case of congestion

Hitless Switching "0" Bit error



Dual-fed and dual-receiving/Sorting and deduplication algorithm

- Site A: Dual-fed packets in sequence in different routes
- Site B: Selectively receives packets in sequence based on packets quality

HUAWEI

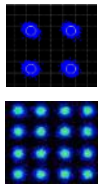
Increasing Fiber Capacity: 400G/800G & 12THz C+L Spectrum

ITU-T SG15

➢ **400G/800G port: Standards & technologies ready, mainstream for next 3~5 years**

BW per λ	10G/40G IMDD	100G/200G Coherent	**400G/800G Coherent**	1.6T+ Coherent
			Mature, massively deployed	**In early stage, under discussion**

	Format	Distance
400G	QPSK	2000km+
800G	16QAM	<1000km



- *ITU-T standard just kicked off*
- *Ready time will be ~2028*

G.694 G.698

➢ **C6T+L6T: A realistic way to increase the fiber capacity**

Spectrum	C4T 50GHz * 80CH	C6T 75GHz * 80CH	**C6T+L6T 150GHz * 80CH**	C6T+L6T+S 200+GHz * 80CH
				In early stage, under discussion

- The range definition of spectra C6T+L6T

	Wavelength (nm)	Frequency (THz)
C band	1524.30~1572.27	190.675~196.675
L band	1575.16~1626.43	184.325~190.325

Successful deployment of 400G LH & C6T+L6T in China (2500KM)

T22-SG15-C-0834!!MSW-E Considerations of wavelength range of B400G DWDM applications, ITU-T SG15 Plenary meeting, Montreal, 1-12 July 2024

Contents

Use Cases of AI-ready Network Infrastructure

US Lumen fiber infrastructure and performance

Building the backbone for AI

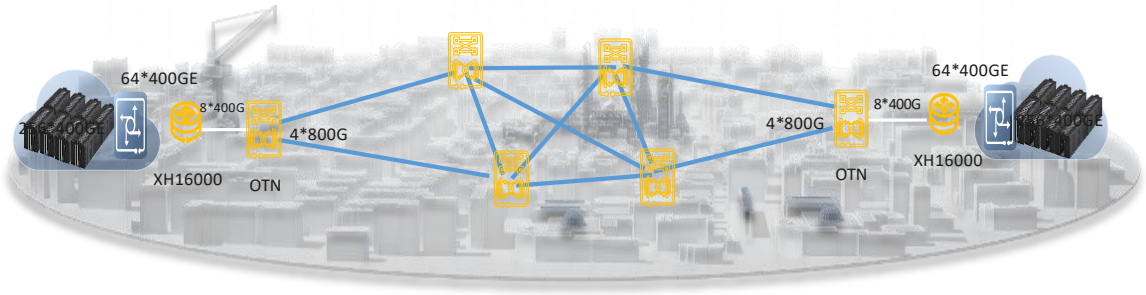
High-capacity connections primed for AI expansion.

- **Programmable Optical Switching** for on-demand connection
- **Flexible ethernet** (1/10/100/400) Gbps **and OTU** (1/2/2e/3/4)
- **Protected** and **unprotected** service paths
- **Low-latency routes** and **latency SLAs**

China Telecom: Innovative DCI packages to Enterprise

Packet Loss	High Cost	Physical Failure
Slow status feedback leads to congestion	non-convergence lead to large number port & device	Insertion loss, fiber cut and link fault

The industry 1st distributed ultra-broadband & lossless computing



The Baichuan3-177B model from Baichuan Technology has already been commercially deployed in this environment

<0.6% in 500km	Cost Saving 95%	Insensible 50ms
Computing Efficiency	32:1 Convergence	Fault Recovery

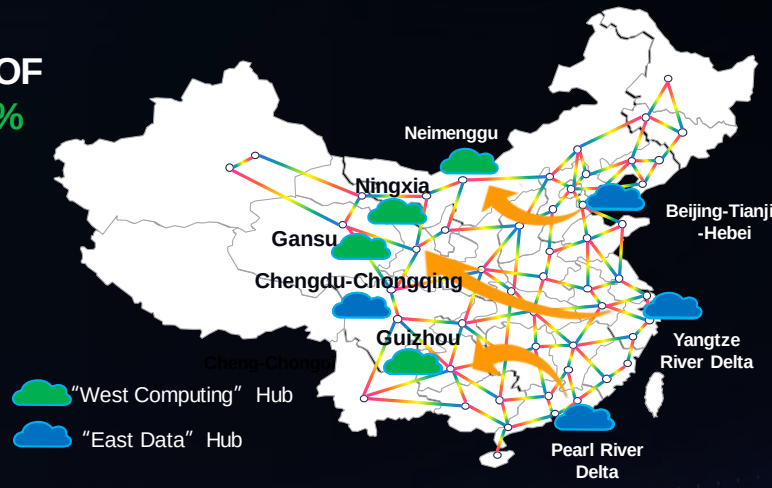
HUAWEI

China: “Accelerating the Construction of a National Integrated Computing Power Network"

In Dec. 2023, the National Development and Reform Commission, the Bureau of Data and other five departments jointly issued the "Implementation Opinions on Accelerating the Construction of a National Integrated Computing Power Network," proposing that by 2025, The 1 ms latency urban computing network, 5 ms regional computing network, and 20 ms cross-country hub node computing network are preliminarily **implemented in the demonstration area**.

Integrated Layout	Integrated Collaboration	Integrated Application	Integration Convergence	Integrated Advance
General + AI + Super computing center	East + Central +West computing power	Computing power + Data + Algorithms	Computing power + Green Energy	Computing power + Security Assurance

215 EFLOF
AI > 80%



Hyperscale All-Optical Backbone

- 400G port: 20k+
- Site: 3000+
- OXC: 100+
- Capacity: > 30k T
- Total distance: > 400k km

Low Latency

E2E OXC
Latency < 20ms

Energy Saving

Energy Cost
20% reduced / year

- Promote the integration of multi-heterogeneous computing power
- Strengthen the Scientific Layout of Various Computing Resources
- Enable the universal and easy-to-use computing services.

- Improving the efficiency of computing network
- Operation collaboration of computing and network.
- Cross-Regional Computing Power Scheduling System

- Promote the convergence of computing power, data, algorithms.
- Enhance collaboration of industry data and computing power
- Building a Trusted Computing Network Environment

- Drive energy conservation and consumption reduction in data centers
- Innovative Computing and Energy Collaboration Mechanism

- Improve the Security Assurance System of Computing Networks
- Build Balanced Mechanism for Developing Efficiency and Safeguarding Security

Contents

Advancing all intelligence, offering a new option for the world

All Intelligence

1. Building a solid computing foundation to support various models and applications
2. Prospering the Ascend AI ecosystem
3. Supporting diverse foundation models
4. Elevating Huawei products with AI
5. Optimizing internal operations and mgmt. with AI
6. Strengthening industries with AI

50+

Mainstream foundation models + industry models

Ascend C

openGauss

OpenEuler

CANN

[M]$ MindSpore

Kunpeng

Ascend

Full-Stack AI Capability，Provide Another Option for the World

Full-stack AI capability

Industry Applications

AI Model & Frameworks

Cloud Platform

Basic hardware & enablement

Openness & Open Source to Enrich Ascend Ecosystem

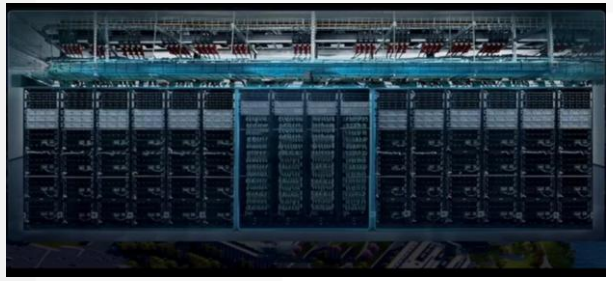
CANN*	MindSpore	MindIE	MindX	…
AI chip toolkit	AI framework	inference framework	Acceleration Kit	…

Finish Open Source Process by the end of 2025

**Except for：Compiler & instruction set*

SuperPod Solution: Build World No.1 Computing Power

Atlas 900 A3 SuperPoD*



- Max. **384** NPUs
- Max. **300** PFLOPS
- **300+** sets deployed
- **20+** customers served

Available at Q1 2025

1.7X

NVIDIA GB200 NVL72

Atlas 950 SuperPoD

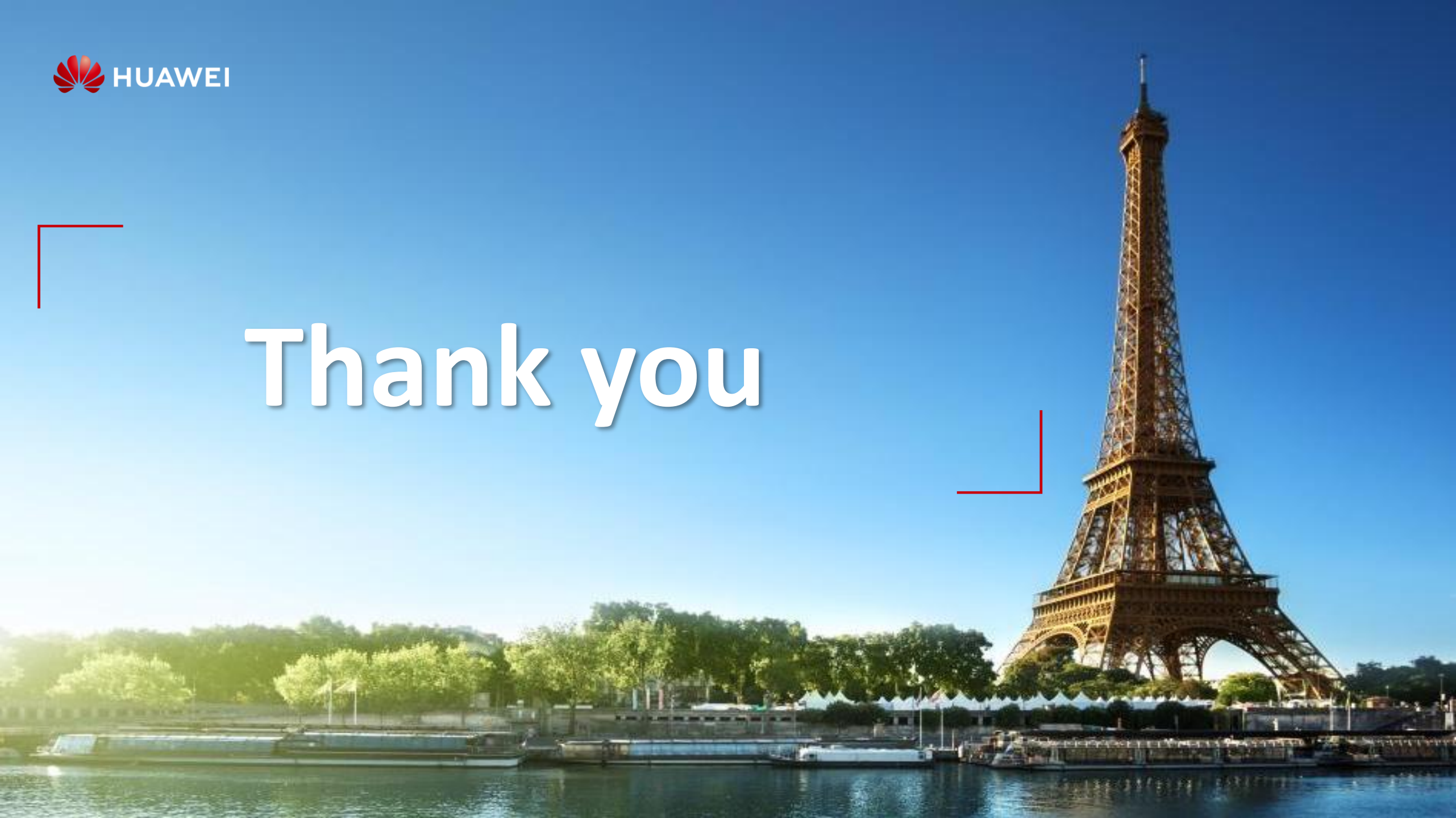
- Max. **8192** NPUs
- Max. **524** EPFLOPS

Available at Q4 2026

Atlas 960 SuperPoD

- Max. **15488** NPUs
- Max. **2** ZPFLOPS

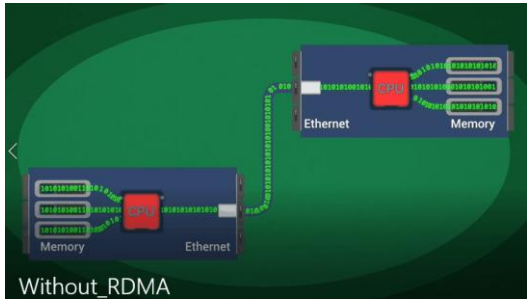
Available at Q4 2027


HUAWEI
Thank you

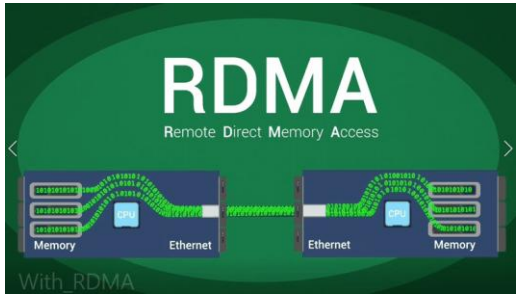
RDMA and RoCE Brief Introduction

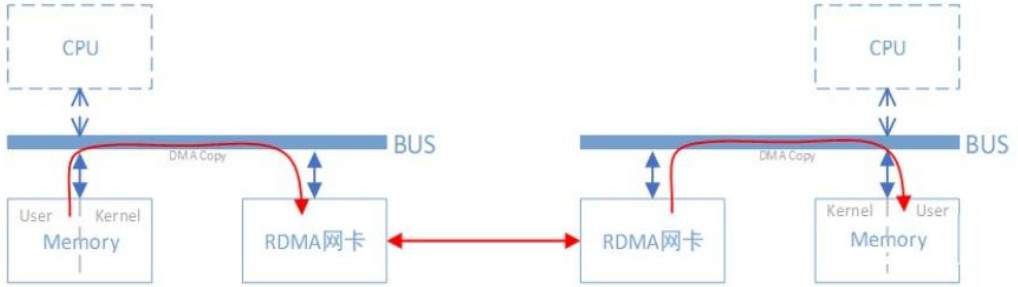
RDMA: Remote Direct Memory Access

Memory R/W with CPU

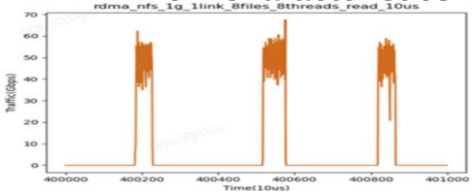


Memory R/W bypass CPU

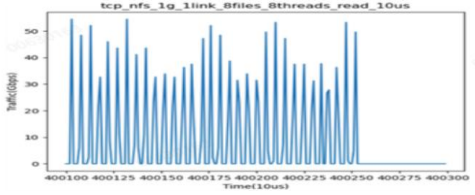




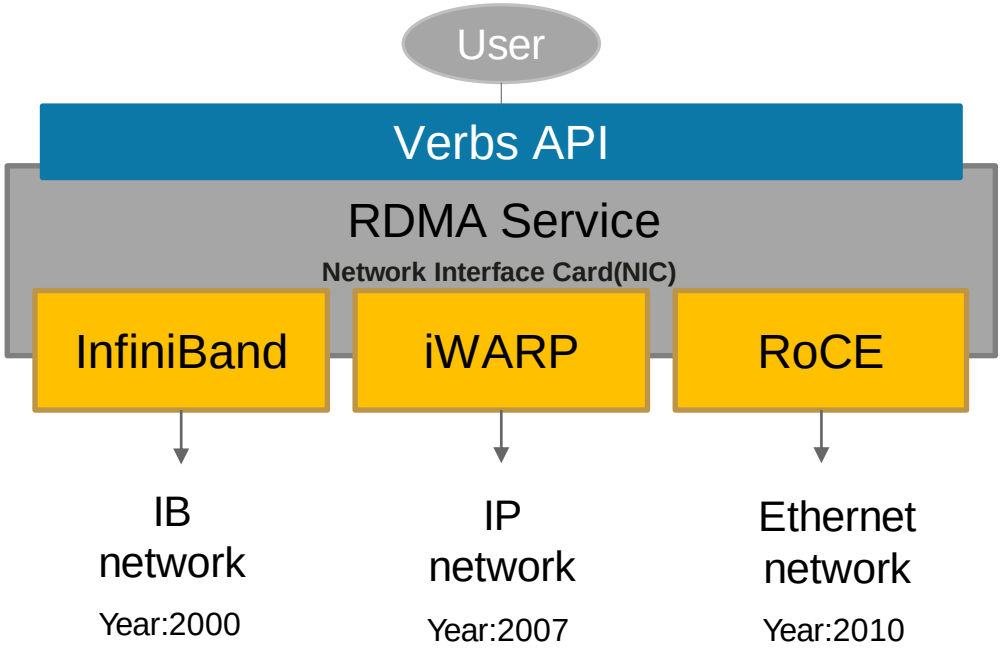
RDMA Flow Characteristics



TCP/IP Flow Characteristics



The burst bandwidth of the RDMA protocol is full, and the TCP/IP protocol is clipped to fill the peak and off-peak hours



RoCE: RDMA over Converged Ethernet
RoCE V2: 2nd generation of RoCE, network layer is UDP/IP, link layer is Ethernet

	InfiniBand (IB)	iWARP	RoCE
Organization	IBTA	IETF	IBTA
Performance	Best	Normal	Almost like IB
Cost	High	Middle	Low
Distance	Short	Long	Middle => Long

HUAWEI