Technology Development of High Efficiency and High Speed Processing Distributed Computing System using Heterogenous Material Integrated Optoelectronics

Weak Point of Distributed Processing on Current Networks

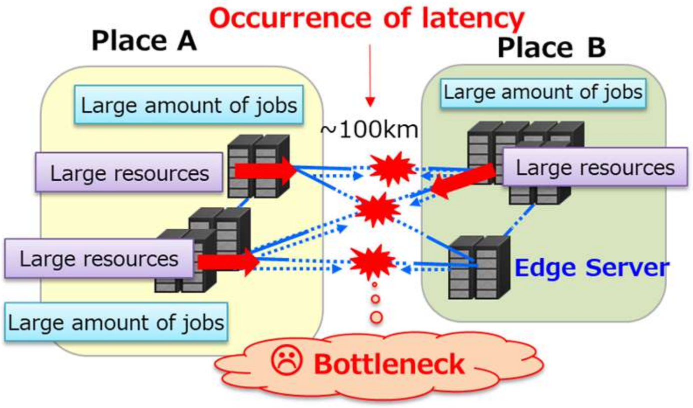


Fig.1 Weak point of current network

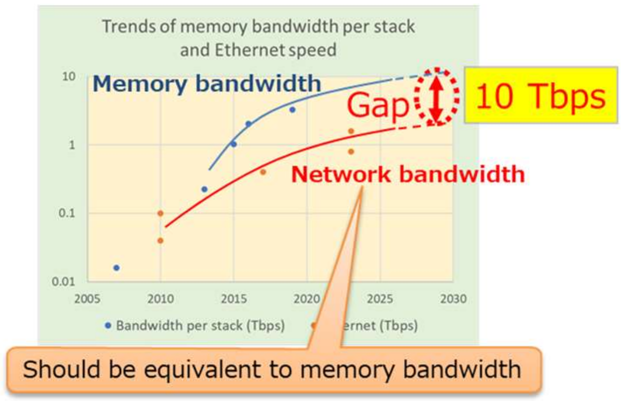


Fig.2 Memory and network bandwidth trends

- ◆ Network bottleneck between servers (Restrictions on communication speed and latency)
- ◆ Inefficient transaction due to non-uniform performances between servers

<The ways to solve weak points>

① 10 Tbps class network bandwidth

② Data transfer with low-power-consumption (<10 pJ/bit)

③ Middleware to enable latency management and data synchronization between servers

④ Elastic bandwidth tuning under consideration of connected server resources

Proposed Next-Generation Distributed Computing System

Proposal to resolve ①, ②

- ◆ Heterogeneous material integration technology (Realize photonic integrated circuits using heterogeneous material bonding)
- ◆ 10-Tbps class coherent transceiver technology with low-power-consumption (<10 pJ/bit)

Proposal to resolve ③, ④

- ◆ Multi-degree elastic optical network technology for distributed computing

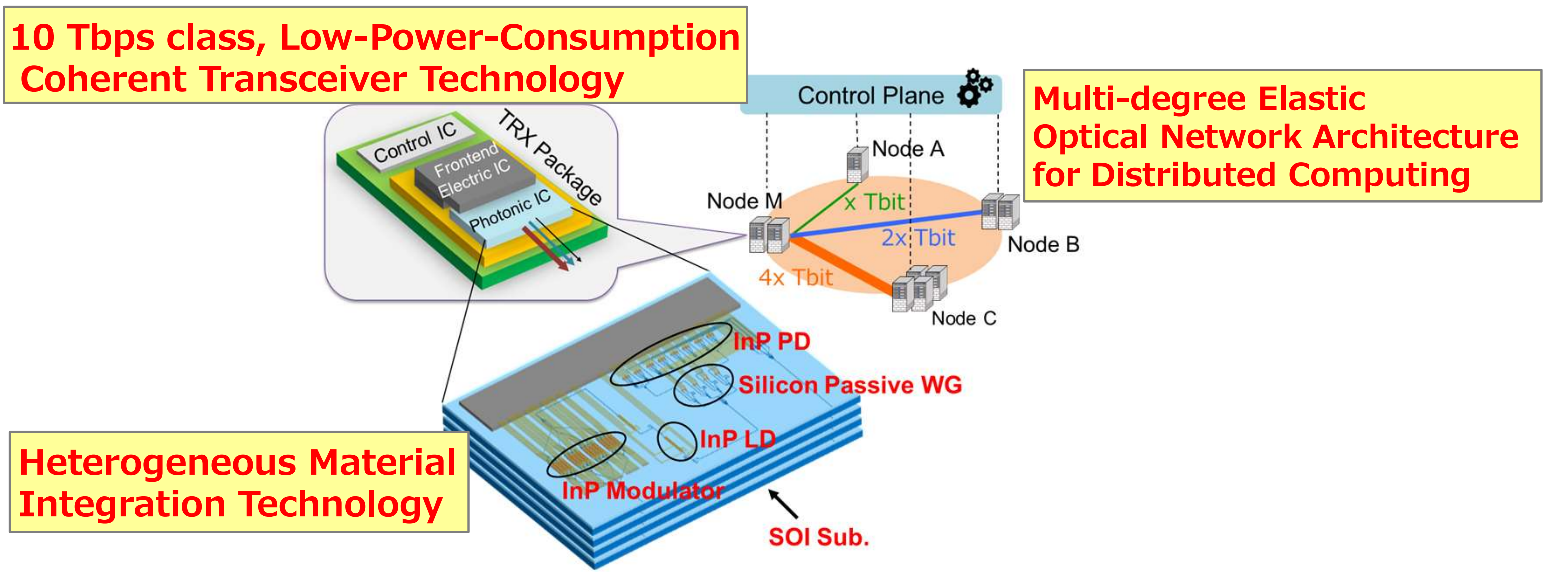


Fig.3 Overview of proposed next-generation distributed computing system

This Research and development is being conducted as a project, JPNP16007, commissioned by the New Energy and Industrial Technology Development Organization (NEDO).

PETRA Photonics Electronics Technology Research Association

Heterogeneous Material Integration Technology -1

Advantages of Heterogeneous Material Integration

Towards ultra-high-speed operation (10 Tbps) and low power consumption,
the III-V/Si heterogeneous material integration which deploys appropriate devices for each section is very attractive to overcome the limitation of single-material optical devices.

Advantages of heterogeneous material integration

Tab.1 Comparison of III-V, Silicon and Heterogeneous material integration

Items	III-V	Silicon	Heterogeneous material integration
High efficiency	Good	Fair	Good
High speed operation	Good	Fair	Good
Integration of light source	Yes	No	Yes
Integration of electronics	Poor	Good	Good
Miniaturization	Fair	Good	Good

III-V/Si photonic integrated circuit

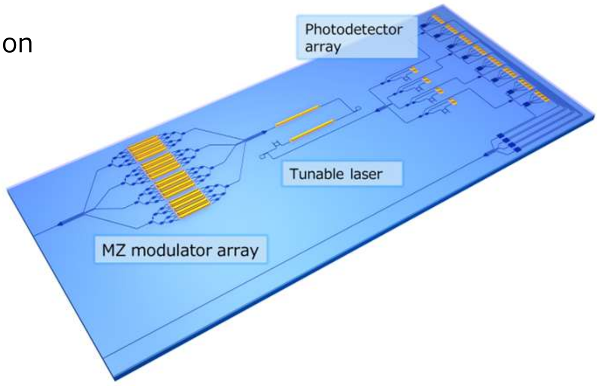


Fig.1 Perspective view of III-V/Si photonic integrated circuit

This PJ constructs the process line for the development of heterogeneous material integration devices and is aiming at the realization of III-V/Si large-scale photonic integrated circuits for 10 Tbps and low power dissipation optical transmission systems

InP chip on Si wafer bonding technology

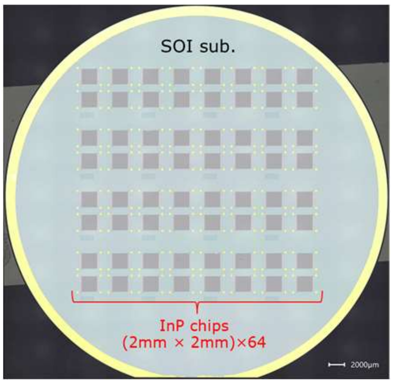


Fig.2 Wafer photograph of InP chips on SOI wafer

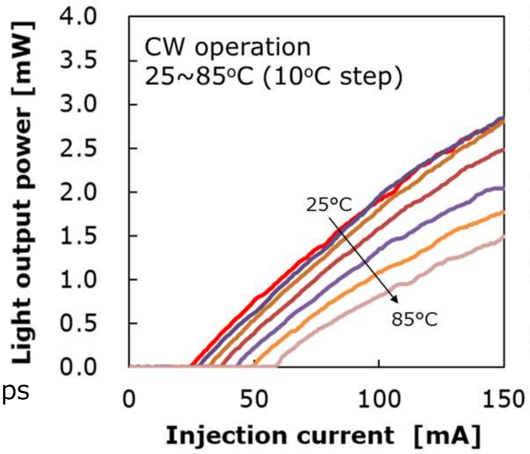


Fig.3 I-L characteristics of hybrid laser for various temperatures

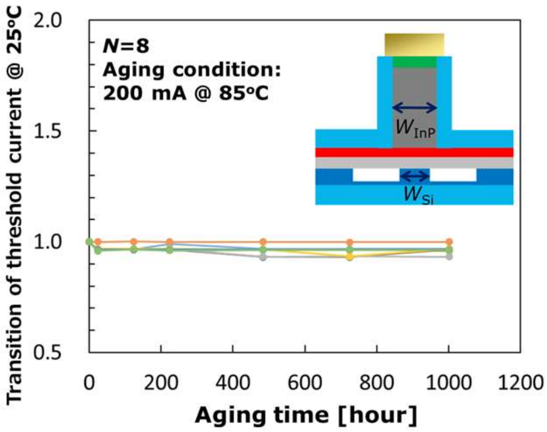


Fig4. Transition of threshold current after aging test (25°C)

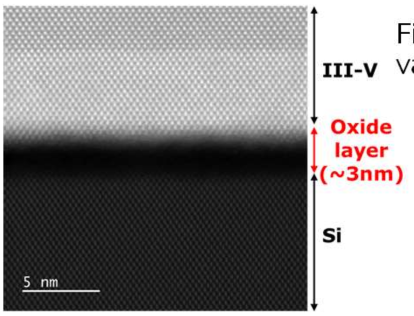


Fig.5 STEM image of III-V/Si bonding interface after bonding of InP chips on SOI wafe

- ◆ **High quality bonding interface by UV-ozone hydrophilization**
- ◆ **Achievement of CW operation up to 85°C in hybrid lasers**
- ◆ **Stable operation after aging test of 1,000h**

*hybrid: heterogeneous material integration

This Research and development is being conducted as a project, JPNP16007, commissioned by the New Energy and Industrial Technology Development Organization (NEDO).

Heterogeneous Material Integration Technology -2

Development of low power consumption wavelength tunable laser

III-V/Si hybrid wavelength tunable laser

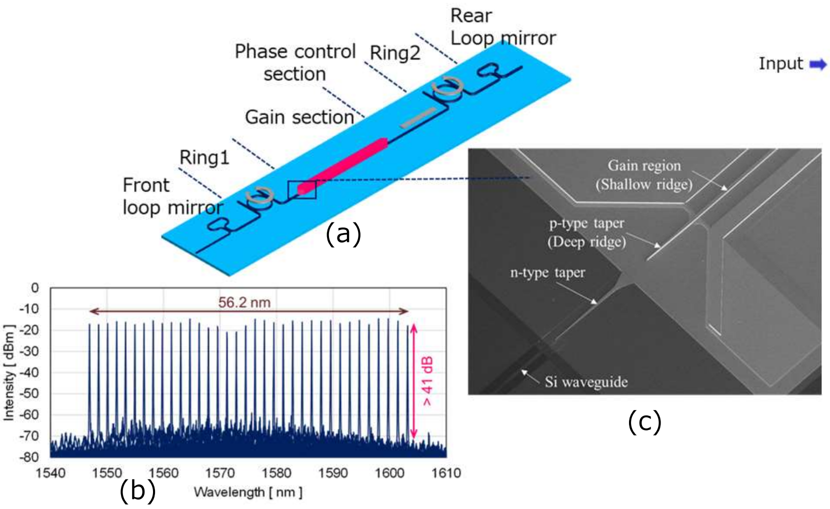


Fig.1 (a) Schematic and (b) spectra intensity of III-V/Si hybrid wavelength tunable laser and (c) SEM image of two-storied ridge structures

◆Wide wavelength tuning range (56nm) and single mode operation in entire tuning range
(Achieved by introducing two-storied ridge structures in III-V/Si optical coupling section)

◆Spectral linewidth: 187 kHz (@100 MHz)

*hybrid: heterogeneous material integration

Monitorable wavelength filter

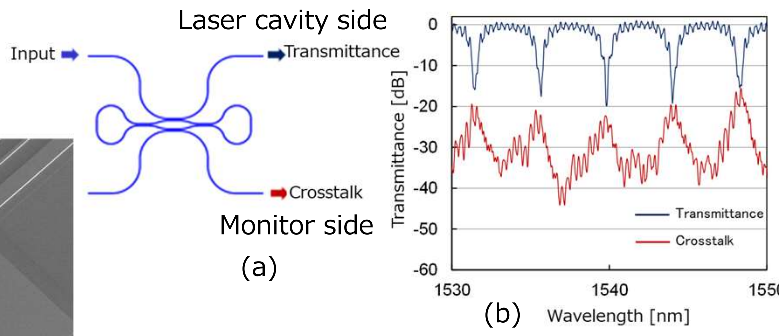


Fig.2 (a) Schematic and (b) spectra of monitorable wavelength filter

◆ Enable observation of spectrum independently from both monitor side and cavity sides independently

Development of wide bandwidth and high efficient modulator/photodetector

III-V/Si hybrid Mach-Zehnder modulator (MZM)

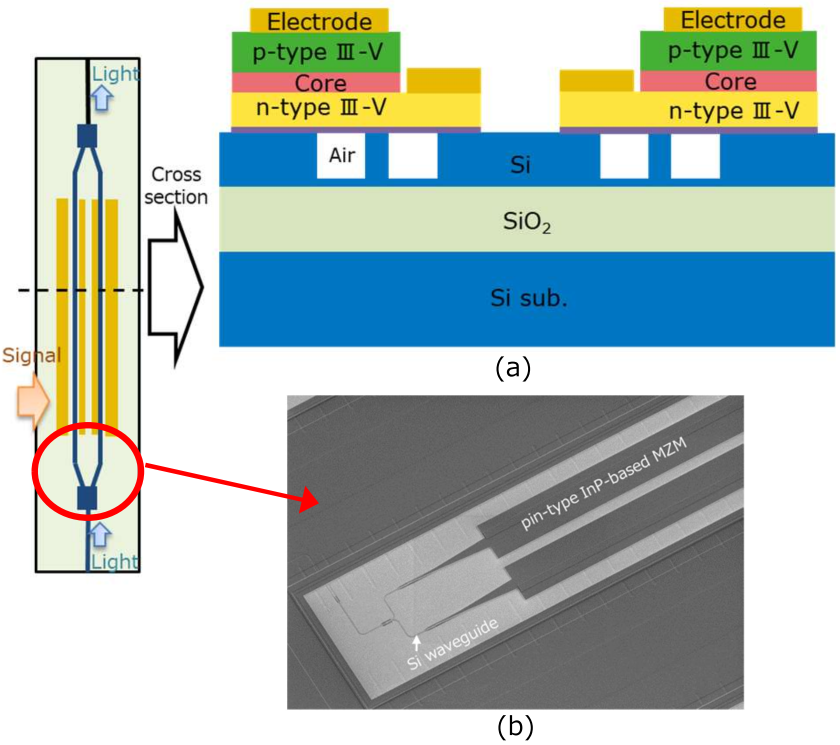
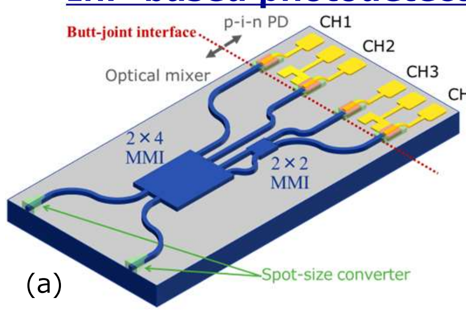


Fig.3 (a) Schematic of III-V/Si hybrid Mach-Zehnder modulator and (b) SEM image of InP MZM region

InP-based photodetector



Wide bandwidth (~80 GHz)
and high responsivity(>0.7 A/W)

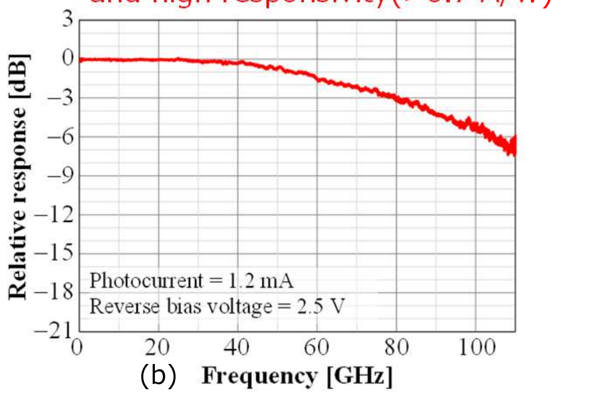


Fig.4 (a) Schematic and (b) Optical/electrical frequency response of InP-based photodetector

This Research and development is being conducted as a project, JPNP16007, commissioned by the New Energy and Industrial Technology Development Organization (NEDO).

10 Tbps class, Low-Power-Consumption Coherent Transceiver Technology -1

Novel Coherent Transceiver Architecture

Research on novel coherent transceiver architecture
to realize both larger throughput and lower power consumption

① Optical frontend technology
Partial offload of electronic function (DSP, DAC, ADC) to photonics
➡ Lower power consumption by reducing load to electronics

② Optical multiplexing technology
Realize large capacity signal by multiplexing plural optical sub-carriers
➡ Larger transceiver capacity without using high-performance electronics

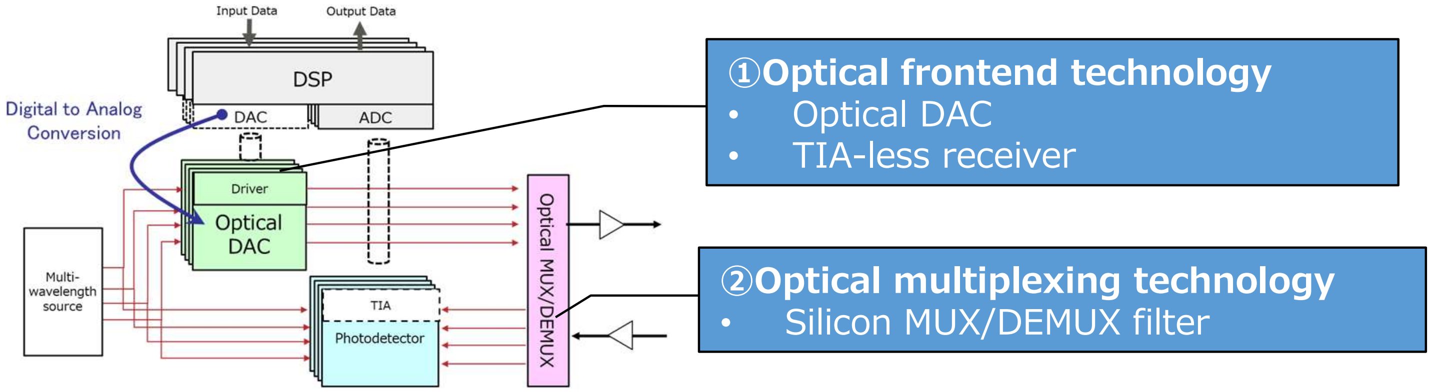


Fig.1 Overview of novel coherent transceiver architecture

Optical DAC transmitter

Realize Digital-to-Analog Conversion in optical domain
➡ Lower power consumption by eliminating e-DAC and linear amplifier

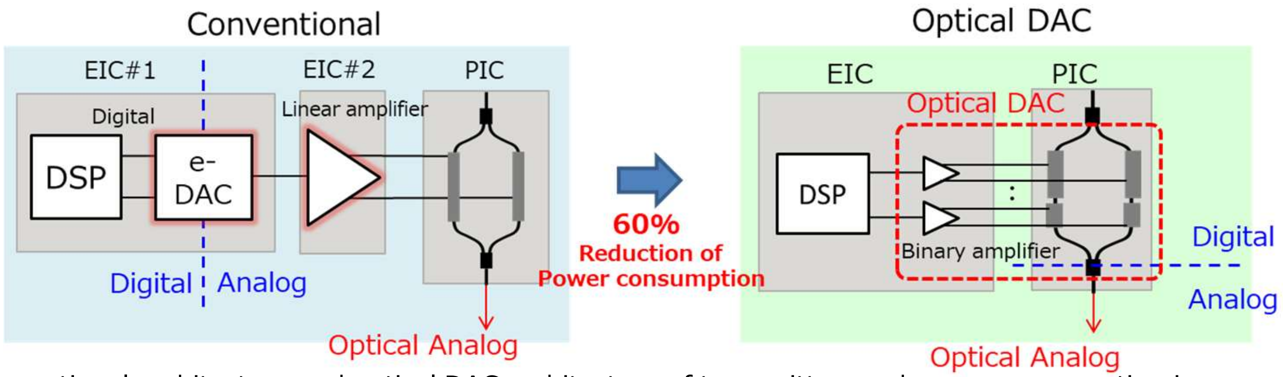
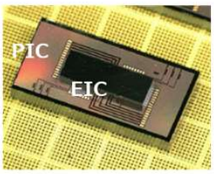


Fig.2 Conventional architecture and optical DAC architecture of transmitter, and power consumption improvement effect

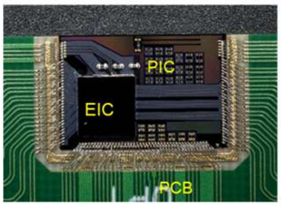
1st sample 2 bit 32 Gbaud



Constellation	
Power efficiency [pJ/bit]	**2.0** **about 1/10 from now on**

Fig.3 Photograph and characteristics of 1st sample

2nd sample 5 bit 64 Gbaud



EIC : Electronic Integrated Circuit
PIC : Photonic Integrated Circuit

Fig.4 Photograph of 2nd sample

This Research and development is being conducted as a project, JPNP16007, commissioned by the New Energy and Industrial Technology Development Organization (NEDO).

10 Tbps class, Low-Power-Consumption Coherent Transceiver Technology -2

TIA-less coherent receiver

Offloading a part of receiver functions from electrical to optical domain
➡ Reduce power consumption by eliminating Trans-Impedance Amplifier (TIA)

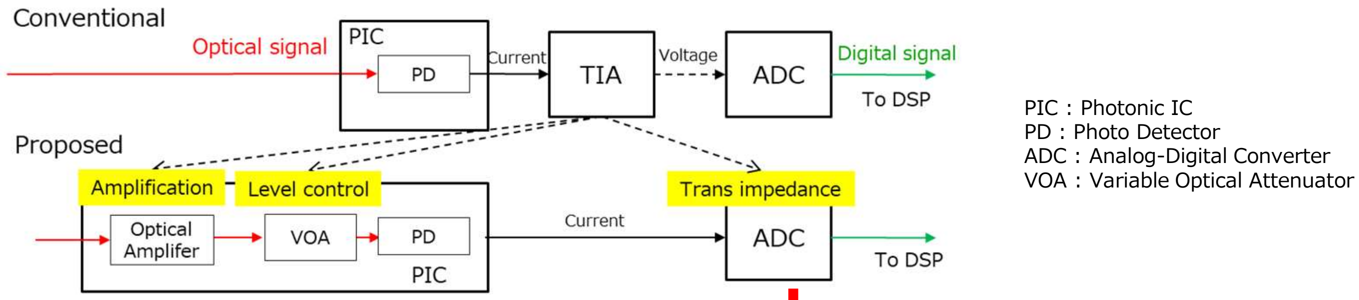


Fig.1 Architecture of conventional receiver and proposed TIA-less receiver

Proof of concept experiment using silicon PIC

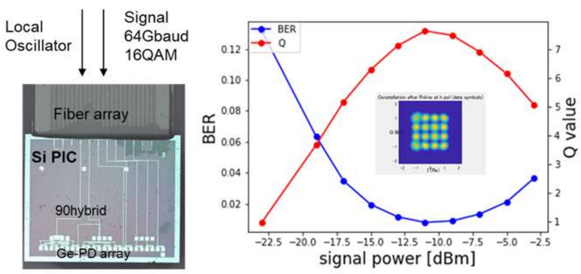


Fig.2 Photograph, BER and Q curve of prototype

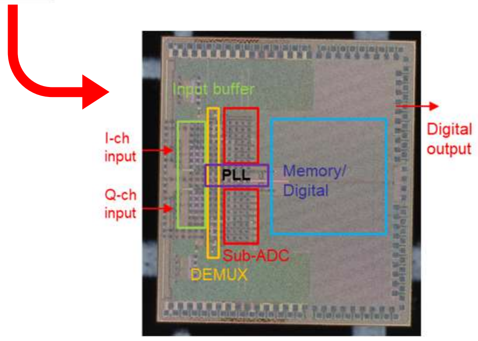


Fig.3 Photograph of Current-input ADC chip prototype

Silicon optical wavelength MUX/DEMUX

Realize multi-wavelength transceiver using high-performance wavelength MUX/DEMUX
➡ Increase transceiver capacity using parallelization. Reduce power consumption by offloading a part of digital signal processing to optical filter

MUX : Multiplexer
DEMUX : De Multiplexer

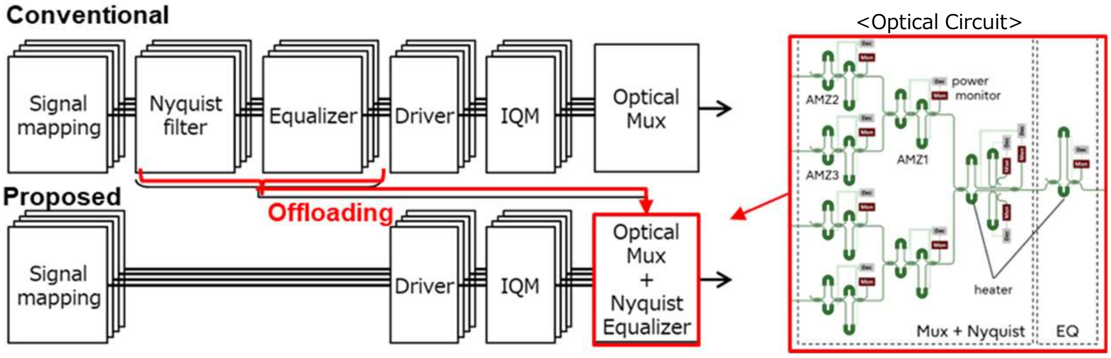


Fig.4 Block diagram of transmitter of conventional architecture and proposed architecture

Fig.5 Photograph of setups of concept experiment using prototyping system

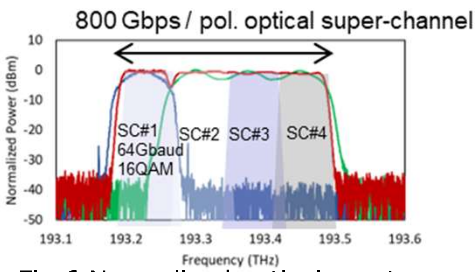


Fig.6 Normalized optical spectrum of proposed architecture

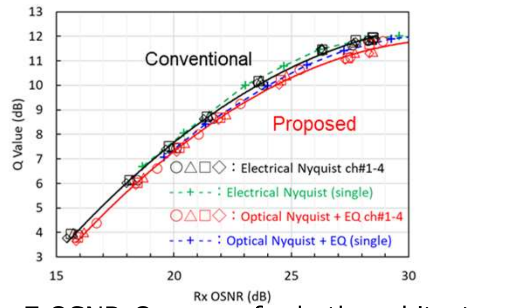


Fig.7 OSNR-Q curves for both architecture

This Research and development is being conducted as a project, JPNP16007, commissioned by the New Energy and Industrial Technology Development Organization (NEDO).

Multi-degree Elastic Optical Network Architecture for Distributed Computing -1

Research and Development Overview/Target

- ◆ Optimization of network resource assignment considering distributed computing applications
- ◆ Estimation and management of quality of transmission for dynamic optical path control
- ◆ Control and management mechanism for unifying computing and optical network resources
- ◆ Target:
  - Improve data transfer efficiency 4+ times
  - System demonstration integrating resource assignment, quality of transmission(QoT), and control and management mechanisms

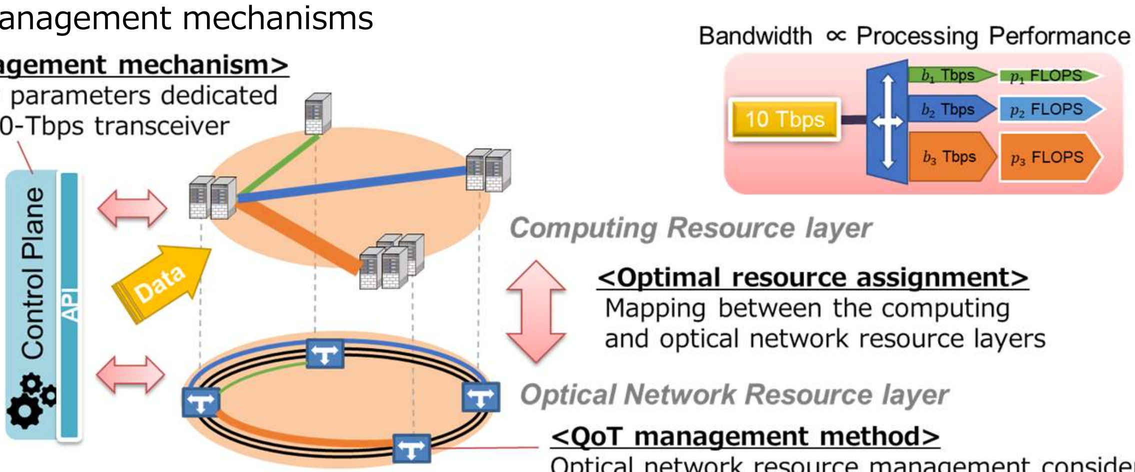


Fig.1 Research and development overview of multi-degree elastic optical network architecture

Setups

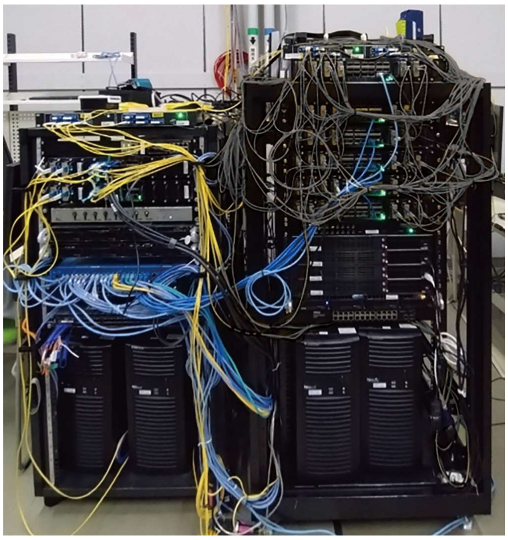


Fig.2 Photograph of setups

- ◆ Estimate minimal inter-datacenter communication where whitebox packet transponder equipment and servers are directly interconnected
- ◆ Up to 8.6 Tbps transmission capacity
- ◆ Run distributed computing benchmarks (cf. MPI benchmark) with 8 servers of each having a dual 100GbE NIC

Research and Development Progress

- ◆ Measure traffic patterns of MPI benchmarks
- ◆ Elaborate detailed parameters and conditions influencing the networking and computing performance
  - Server settings including NIC, OS, BIOS, driver, and software drivers
  - Whitebox packet transponder settings such as flow/traffic control

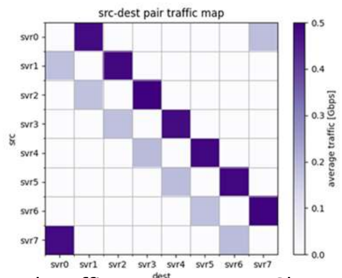


Fig.3 Measured traffic pattern : BT, ClassD

- ◆ Extend the Functional Block-based Disaggregation (FBD)model to introduce physical parameters such as optical insertion loss or gains into the topology description data
- ◆ Automated GSNR estimation coordinated with the opensource GNPy

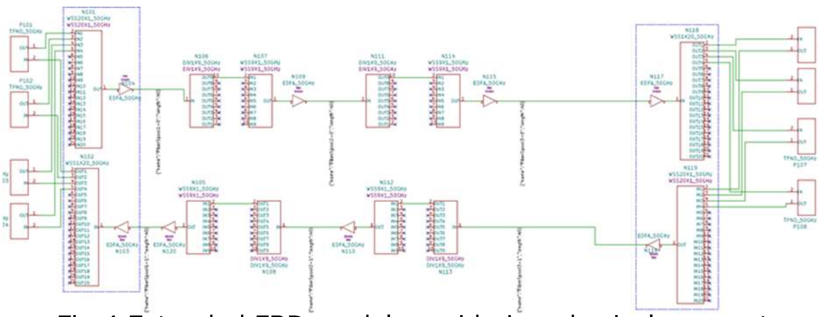


Fig.4 Extended FBD model considering physical parameters

This Research and development is being conducted as a project, JPNP16007, commissioned by the New Energy and Industrial Technology Development Organization (NEDO).

PETRA Photonics Electronics Technology Research Association

Multi-degree Elastic Optical Network Architecture for Distributed Computing -2

The world's first long-transaction capable RDBMS “Tsurugi".

What is Tsurugi, a Japanese RDB?

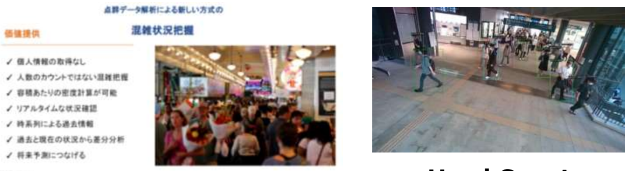
◇ **OSS RDBMS**
◇ **Ensures the most powerful consistency in RDB（serializable)**
◇ **Optimistic processing for low contention and high performance equivalent to KVS**
◇ **Extremely robust for write-heavy batch processing and can coexist with online processing**
⇒ Eliminates the need for “night batch processing.”
◇ **Postgres provided as front RDB ⇒ Reduction of introduction costs**
◇ **A prototype application is provided with it, making it a pragmatic RDB.**

RDB：Relational Database

Tsurugi's Demonstration Experiment

◆ **Tsurugi Usage Example**

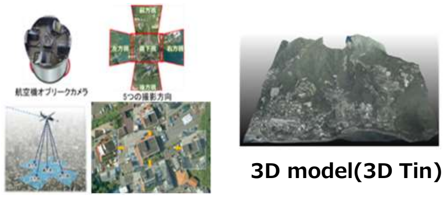
1. **Real-time DB of IoT data**
   ・Utilizing data collected by LiDAR to assess congestion
   ・Enables sequential processing of data to be collected （real time processing）



Head Count

Fig.1 Usage example for real-time DB of IoT data

2. **Application for e-Science**
   → Schema recommendation and large-scale data analysis using complex query workloads with modern, high-volume data
   ・Accelerate slow queries in current PostgreSQL clusters
   → Exploratory data analysis with large amounts of data
   ・Exploratory data analysis in the time direction using giant tables
3. **Utilization in the initial and emergency response phase after a disaster**
   → Reduction of 3DTin generation time from image data from aircraft opleak cameras
   ・Provide broad-based information and platforms during the initial and emergency phases in the initial and emergency phases
   ・Reduced 3DTin generation time using oblique camera images
   ・Automatic object extraction for additional information extraction



3D model(3D Tin)

Fig.2 Usage example for reduction of 3DTin generation time

4. **Application to Productivity Improvement**
   →Two types of business batch processing applications were prepared, and what used to be nighttime batch processing can now be executed at any time during the day without stopping online processing.

・BoM　Batch processing in the food manufacturing industry
・Billing Process　CalculationTelecom-based CDR and packet aggregation and rate calculation batches

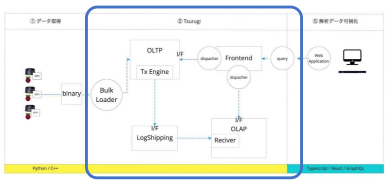
Tsurugi's Vision

◆ **Aiming for a high-speed database for distributed environments**

◇ **Configure the system with multiple RSA (Rack Scale Architecture)**
・Distributed processing across multiple RSAs in a single DC
・Distributed processing using multiple RSAs between DCs within a 100 km range

◇ **Implemented a proto-application using LiDAR on the premise of middleware that performs replication between nodes.**

Currently, applications are actually being released that use LiDAR to handle public system data and perform human flow analysis. This input processing is being done at Tsurugi on a trial basis, and we plan to expand this to distributed processing.



The point cloud data uploaded from multiple LiDAR terminals are processed by Master OLTP in micro-batch and the results are replicated to OLAP.

In the future, it will be possible to monitor the flow of people andther conditions at “public facilities and places” over long distances with low latency, independent of weather and other natural conditions, with the goal of becoming a social infrastructure such as automated driving in the near future.

◆ **Questions and inquiries about Tsurugi**　◆ **Details and explanation about "Tsurugi**

This Research and development is being conducted as a project, JPNP16007, commissioned by the New Energy and Industrial Technology Development Organization (NEDO).

PETRA Photonics Electronics Technology Research Association