IOWN for digital twin enabled societies

Masahisa Kawashima
Nippon Telegraph and Telephone Corporation, 1-5-1 Otemachi Chiyoda-ku, Tokyo 100-8116, Japan
masahisa.kawashima@ntt.com

Abstract: Achieving the extreme data volume and velocity requirements of digital twin applications energy-efficiently is challenging. IOWN will address this challenge by making architectural shifts in computing and networking with the evolution of optical technologies.

1. Threats and Opportunities For Societies

The United Nations reports[1] that the gap between the actual fossil fuel production and the target for global warming limitation will continue to expand despite the increasing severity of the sustainability issue. The report is not surprising because people cannot quickly abandon their growth opportunities even if they are aware of the problem. Hence, societies need to establish new ways of life and business to achieve both growth and sustainability. Digital twin computing described below will potentially pave the way for sustainable growth.

Digital twin computing cognizes the actual situation of people, things, and areas in the real world and updates their data objects in the cloud. These data objects are called digital twins. As the analyzable up-to-date mirror of the real world, digital twins enable us to make short-term predictions and optimize planning in many fields of our business and life. For example, we may be able to achieve carbon-neutral transportation by deploying electric vehicles based on demand forecasts and charging the vehicles with the nearest renewable energy[2]. In addition, societies may stabilize energy distribution by dynamically controlling energy production and consumption[2]. In such ways, digital twin computing will raise our productivity while reducing carbon emissions.

2. Challenges for Digital Twin Computing

As we deploy digital twin applications to reduce our carbon emission, people will naturally demand that the energy consumption by digital twin systems should not exceed the energy-saving enabled by them. Unfortunately, this is challenging. This section explains why with a case analysis in AI-based area monitoring for smart cities.

As shown in figure 1, Innovative Optical and Wireless Network (IOWN) Global Forum highlights data volume and velocity requirements for AI-based area monitoring in [3]. According to the report, data captured by cameras/sensors will amount to 48 Gbps just for one medium-sized building. Those data are cognized to generate instant feedback to actuators for optimization or safety control purposes. For many such applications, the time between sensing and actuation should be less than 100 msec.

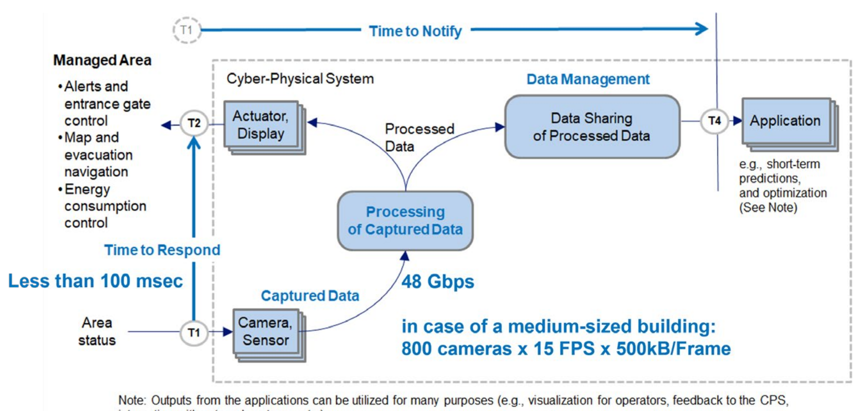

Fig. 1 Data Volume and Velocity Requirements in AI-based Area Monitoring[2]

These data volume and velocity requirements are very challenging for today's computing and networking infrastructure. Summarized below are key reasons:

Issue 1: In today's IP networking, optical wavelength paths are used only to transport aggregated traffic between large data centers. Outside of the optical wavelength paths are best-effort packet switching/routing networks.

Hence, data should be buffered at many layers and many forwarding nodes for flow control and retransmission. This increases the latency and overall energy consumption.

Issue 2: Although the energy efficiency in AI inference computation has been soaring with the progress of AI accelerators, as the data volume increases, the host CPU gets more overwhelmed with the workload of receiving data because the host OS, i.e., software running on the host CPU, usually handles TCP/IP.

Issue 3: We now rely on accelerators but the elastic operation of accelerators is challenging. This is because sharing accelerators across physical computers spoils the performance gain and creates additional data transfer workload on the host CPU. Edge computing further worsens this issue, fragmenting the IT resource pools.

3. Innovative Optical and Wireless Network (IOWN)

NTT publicized its vision to build Innovative Optical and Wireless Network (IOWN), a new infrastructure that evolves with optical communication and optoelectronic integration technologies[4]. Through the IOWN Global Forum establishment, this vision has become shared globally by more than 75 memberships. IOWN will solve the above challenges by making architectural shifts in computing and networking as described below:

All Photonic Network

The ongoing disaggregation of optical network nodes, e.g., ROADM, enables us to distribute the optical networking functions. Taking advantage of this trend, IOWN All Photonic Network (APN) allows user nodes to have optical transceivers and provides optical wavelength paths directly connecting user nodes' transceivers. The direct optical paths will enable user nodes to transfer/stream data with less buffering, e.g., RDMA over APN.

Initially, there would be a limit on the size, i.e. the number of ports and wavelength paths, of one APN. This is because the number of wavelength channels is limited, and efficient wavelength conversion methods are still under study. However, even with the size constraint, APN will embrace many emerging use cases such as digital twin computing because these use cases require point-to-point connectivity rather than any-to-any connectivity. In addition, the evolution of ultra-wideband optical transport and efficient wavelength conversion technologies will raise the size limit and ultimately realize a scale-free network.

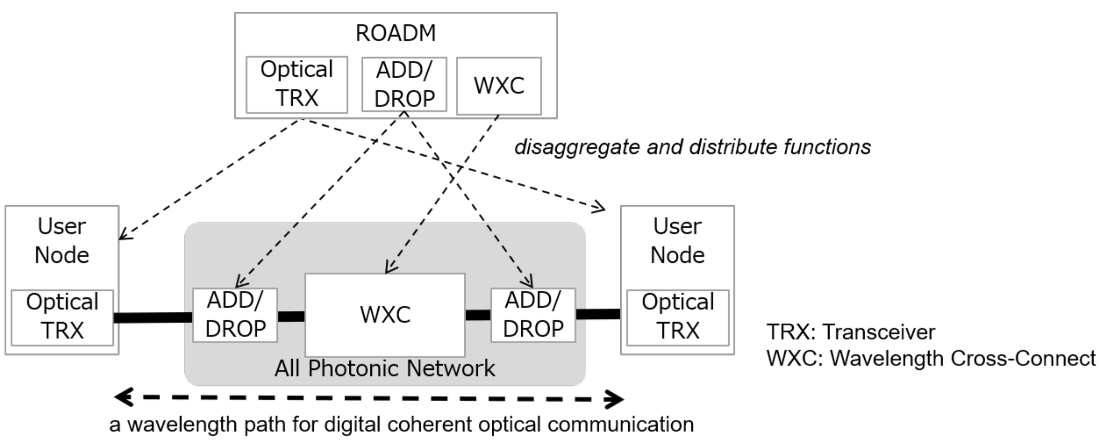

Fig.2. Disaggregation of Optical Network Nodes and All Photonic Network (APN)

Photonic Disaggregated Computing

Several suppliers have recently released processor cards with network interface units, e.g., Data Processing Unit (DPU) and Infrastructure Processing Unit (IPU). These networkable processors realize a data plane that spans across physical computers without increasing the workload on the host CPUs. Embracing this trend, IOWN will build a Photonic Disaggregated Computing infrastructure by connecting networkable processors with a high-speed and energy-efficient optical interconnect. In addition, with high-speed inter-DC connections by APN, Photonic Disaggregated Computing will create a big IT resource pool across data centers. This WAN-wide disaggregated computing will facilitate the elastic operation of accelerators.

Currently, we have only a limited variety of networkable processors. However, the evolution of optoelectronic integration technologies will lower the hurdle for processors and other modules to have optical network I/O.

Moreover, nano electro-optic modulators[5] will help us reduce the energy consumption of optical I/Os significantly. Future optoelectronic integration will also enable interconnecting chiplets inside a package optically.

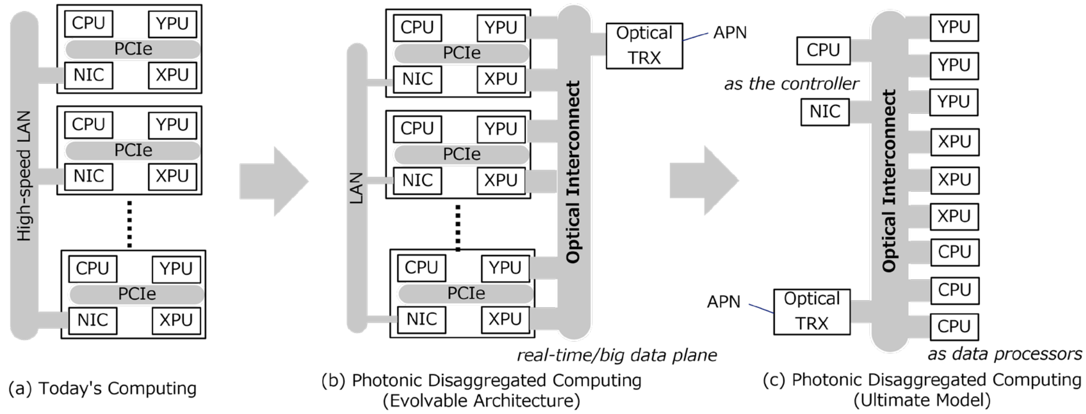

Fig. 3. Photonic Disaggregated Computing

Use Case Example: AI-based Area Monitoring for a Digital Twin City

Fig. 4 illustrates an AI-based area monitoring system for a digital twin city as an example use case of IOWN. This implementation solves Issues 1-3 as follows:

- APN connects a sensor/camera gateway and the edge computing data center with a wavelength path. The wavelength path enables RDMA-based data copying between the sensor/camera gateway's memory and the memory of XPU, e.g., AI accelerator, in the edge data center (DC). In this way, captured data are input to data processors, with much less buffering and without overwhelming the host CPUs in the edge DC.
- Photonic disaggregated computing enables resource sharing across physical servers. Furthermore, APN enables us to create a big IT resource pool across data centers, resulting in a higher IT utilization rate and reduced energy consumption.

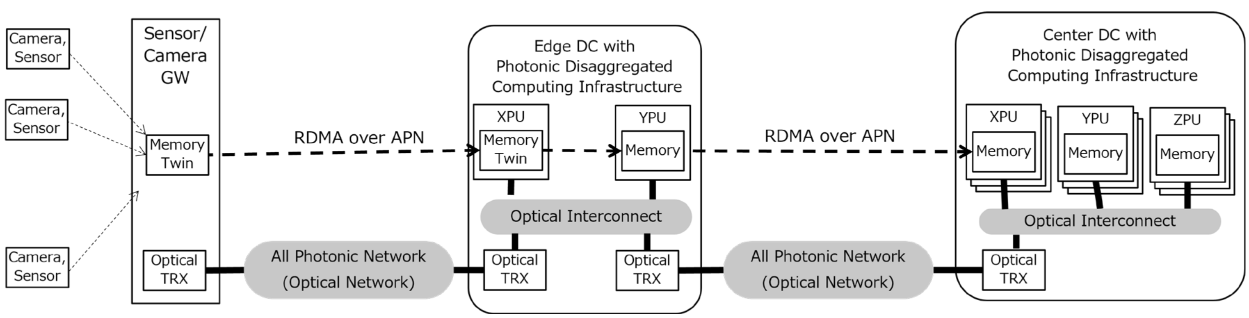

Fig.4. AI-based Area Monitoring with IOWN

4. Conclusion

IOWN APN and Photonic Disaggregated Computing reduce data buffering and copying to achieve the extreme data volume and velocity requirements energy-efficiently. Their initial versions will have a few constraints, e.g., the size of subnets. However, even the initial versions will embrace many emerging use cases such as digital twin computing. More importantly, APN and Photonic Disaggregated Computing will keep evolving, raising their capacity and energy efficiency, with the progress of optical communication and optoelectronic integration technologies, in other words, thanks to OFC researchers' efforts.

References
[1] United Nations, 2021 Report The Production Gap, https://productiongap.org/wp-content/uploads/2021/10/PGR2021_web_rev.pdf
[2] IOWN Global Forum, Cyber-Physical Systems Use Case Release 1, Version 1.0, IOWN Global Forum, October 2021
[3] IOWN Global Forum, System and Technology Outlook Report , IOWN Global Forum, April 2021
[4] NTT, What's IOWN? The Social Background and Purpose, https://www.rd.ntt/e/iown/0001.html, NTT, 2019
[5] K. Nozaki, S. Matsuo, T. Fujii, K. Takeda, A. Shinya, E. Kuramochi, and M. Notomi, Femtofarad optoelectroic integration demonstrating energy-saving signal conversion and nonlinear functions, Nature Photonics, 13, 454-459, April 2019