www.pipelinepub.com
Volume 20, Issue 3

The All-Photonics Future *The Case(s) for Photonics-based Networking and Computing Infrastructure*

By: Vito Mabrucco

Data-driven economies are racing toward a cliff. Investments in digital transformation continue to grow, yet up ahead are limits to transport capacity and computing power. These constraints include the demise of Moore's Law and data infrastructure that is hampered by bottlenecks, scaling issues, and high energy consumption.

Fortunately, many service providers and technology vendors are aware of this situation. A large number agree that one way to bridge the looming chasm is with photonics. This article reviews one industry-driven approach and illustrates its capabilities through three reference models.

Questions and Answers

The dilemma is real. Over the past decade and more, enterprises around the world have invested hundreds of millions of dollars to digitally transform their operations and services. According to an IDC report, annual global expenditures in this category will reach nearly $3.9 trillion in 2027. But will the infrastructure be there to support this increased use of data?

Will the number of transistors in a silicon chip continue to double every two years, per Moore's Law? What about issues of network delay inherent to the protocols (TCP/IP) that prevail across the internet, or performance that cannot scale across diverse networks, or the unsustainable amounts of energythat data centers are expected to consume?

To answer the first question, transistors cannot get much smaller, meaning Moore's Law will need to be amended or repealed. As for network protocols, there are alternatives (such as UDP) that can be adapted for specific applications. A consensus is emerging, however, concerning one technology that can deliver exponential, across-the-board benefits commensurate with rising data-usage demands, namely: photonics.

The timing is right. Consider the growing acceptance of silicon as a medium for light, an idea that was proposed more than thirty years ago. One near-term application for today's booming silicon photonics market is eliminating optical-to-electrical/electrical-to-optical (OE/EO) conversions in network equipment. The benefits? Reduced latency, expanded data center performance, and increased energy efficiency.

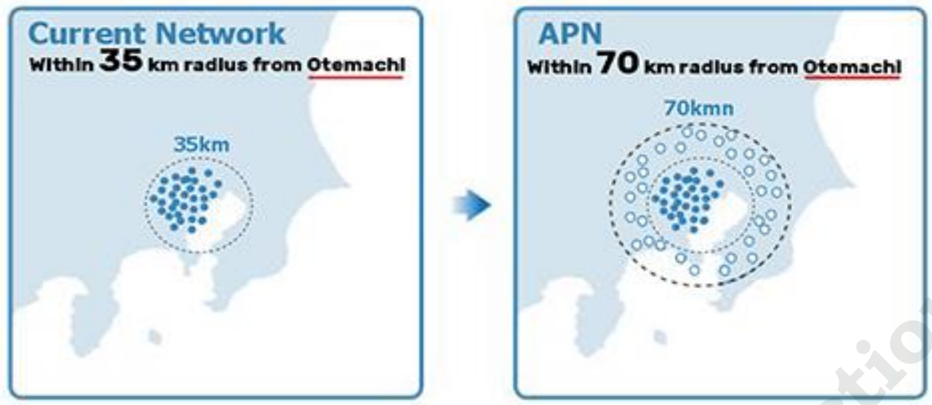

Figure 1: APN-Driven Data Center ExpansionSource: NTT R&D Forum Keynote Speech, 2023, slide 43

Going All Photonic

Leaning on silicon photonics, along with advanced edge computing, wireless, cryptographic, and other technologies, an industry group led by NTT, Intel, and Sony emerged several years ago with a roadmap for a new communications and computing platform. It is comprised primarily of an Open All Photonics Network (APN) and a related Data-Centric Infrastructure (DCI). Phase one (of four) has now been completed, with an extremely low-latency service having gone live in March 2023.

Built on an open and disaggregated architecture and comprised of transceiver, gate, and interchange devices, Open APN has achieved data transfers of up to 1.2 Tbps and communication lags that are one two hundredths of corresponding delays in conventional systems. Such a massive reduction in latency creates possibilities. It could enable the doubling of a data center network radius (See Figure 1), eliminating the lag in remote e-sports, and providing new services, such as performances of a geographically distributed orchestra. Other examples include the area of manufacturing and construction with the ability to do remote management of robotics, or in health care providing the true ability to do long-distance surgery. The industry group behind this approach, the Innovative Optical and Wireless Network Global Forum (IOWN GF), now includes more than 125 members. In its next phase, it will shift its focus from networking to computing. The goal over the next seven years is to incorporate APN-like technology first between circuit boards, then between chips, and finally inside chips, in accordance with the DCI architecture, which optimally subdivides computing resources based on specific purposes of data processing.

Use Case Overview

To get a better sense of how the Open APN/DCI platform and related technologies will unfold, let's see how it could be used in practice. The IOWN GF envisions use cases congregating around two categories:

Cyber-Physical System (CPS) use cases, such as physical security alerts, disaster notification, remote vehicle assistance, energy-efficient traffic routing, remote factory operation, process plant automation, network device failure prediction, network infrastructure management, disease outbreak prediction, renewable energy flow optimization, and social sustainability.

AI-integrated Communications (AIC) use cases, such as interactive live music, interactive live sports, cloud gaming, entertainment on-the-go (in vehicles), remote professional learning, extended reality (XR) navigation, augmented human communication, and instantiation of “another me.”

To accelerate the development and commercial availability of its architecture and technology, the IOWN GF has published several detailed reference implementation models, or blueprints for realizing attractive use cases. The models are also helpful for identifying potential technical issues and improving architecture and technology specifications. What follows are summaries of three models, two CPS (remote factory operation and area management security) and one AIC (interactive live music).

Three Reference Models

Remote Control Robotic Inspection. Geographically distributed manufacturing sites and skilled labor shortages make it difficult to position the right personnel in the right plant at the right time. This situation has created a compelling case for a maintenance expert who can remotely control on-site robots to carry out essential procedures, such as inspections, parts replacement, valve closure, etc.

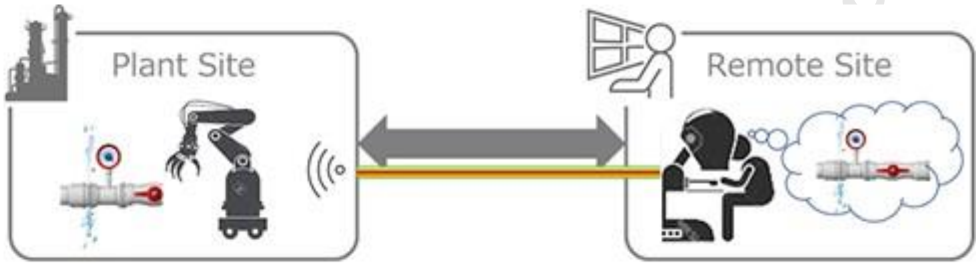

Figure 2: Overview of Remote Controlled Robot Inspection

Technology gaps and issues that would hinder this kind of operation include unpredictable network constraints affecting capacity, latency, and reliability of communication; the trade-off between high reliability and low latency in end-to-end communication; a lack of methods for streaming large volumes of high-definition video with low latency; instability of wireless coverage area; and operational limitations due to network boundary protection for security.

Addressing these gaps, the IOWN frameworks define a communication stack with low latency messaging using remote direct memory access (RDMA)-based protocols for hardware offloading, hardware-based data transfer with adaptive video encoding, and guaranteed wireless coverage through multipath utilization and wireless link quality prediction. This model also requires zero-trust protection, strong cryptography, and an infrastructure orchestrator.

Interactive Live Music. The next case involves live music in a virtual space. (See Figure 3 on next page.) Starting with a few “power users” and then scaling up to a massive number, this ultra-realistic service would harness consumer-grade devices and the power of distributed pooled computing resources over low-latency connections.

Among the specific requirements for this use case are low motion-to-photon latency between sensor devices and display devices (to reduce the chance of motion sickness); low end-to-end latency between capturing the audience member’s motion and reflecting it to others’ displays and vice versa (to enable interactive experiences, such as waving hands and shouting); and dynamic network and computing resource

allocation (to scale up to huge numbers). The IOWN APN platform can enable this future-looking case, delivering the benefits of high bandwidth and reduced power consumption while collecting massive data from capture devices. They can also set up ultra-low-latency connections directly between customer premise sites and the telco edge or core sites that are hosting the application computing resources. The IOWN DCI technologies can enable resource allocation flexibility, which in a disaggregated infrastructure helps reduce CPU costs and power consumption.

Area Management Security. There are many situations where security must be continuous and others where it needs to be tightened up, or applied on occasion, such as at airports upon the arrival of VIPs or on roadsides at marathon running events. This use case taps AI to analyze streamed surveillance video and Light Detection and Ranging (LiDAR) sensor data to detect criminal acts or accidents.

There are numerous gaps between today's centralized cloud-based implementation framework and the requirements in this model. Technical issues include excessive bandwidth cost, lack of deterministic service quality, virtualization overhead, unnecessary power consumption, insufficient resource utilization, data transfer burden on CPU, and costs of a data hub tier.

The application of IOWN APN and DCI technologies in this case can deliver 1) high bandwidth and reduced power consumption across core and access networks; 2) real-time monitoring services at low latency by reducing OE conversion and establishing a direct APN path between the customer and the telco edge/central clouds; 3) flexible resource management from heterogeneous and disaggregated device resources pool; and 4) an efficient RDMA-capable network across multiple sites.

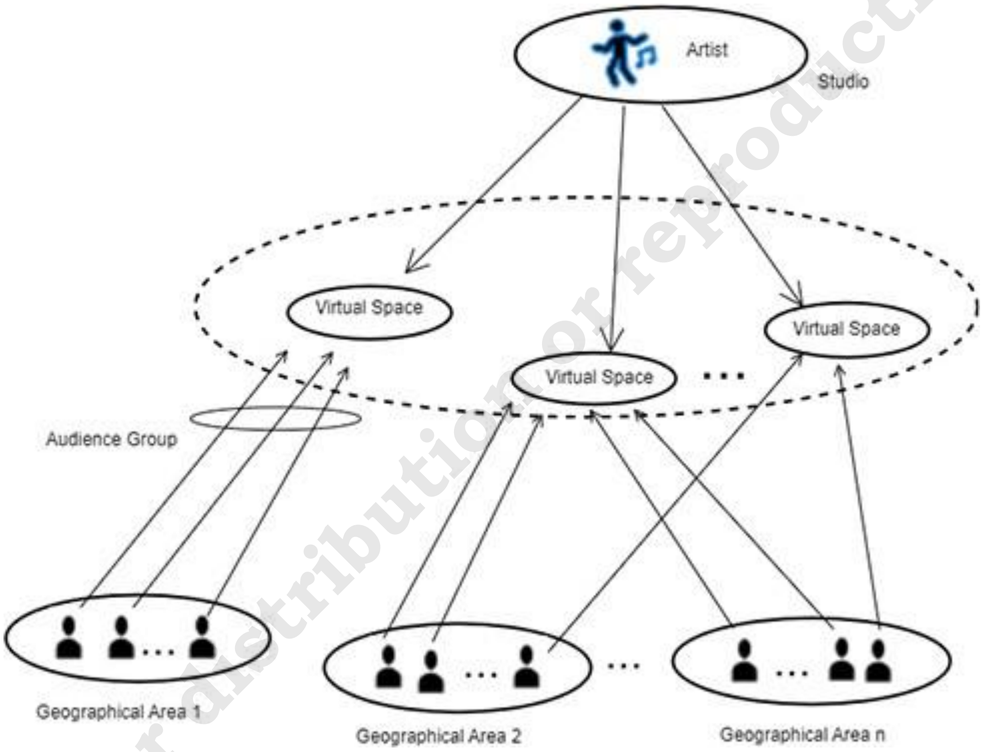

Figure 3: Virtual Spaces and Audience Members Source: RIM for the Live Music Entertainment Use Case, IOWN GF, 2023, p. 90

Lighting up the Future

There are other examples of what this platform can accomplish in terms of performance and services. As already noted, energy efficiency is baked into the technology, just to name one leading benefit. The end goal is a target of 100 times greater power efficiency for the photonics components, including optical fiber cables, transmission systems, and photonics-electronics convergence devices. But this initiative could also drive new energy-saving services, such as a smart-grid management system that automates the operation of a grid relying on renewable energy as its main power supply.

To become greener, more efficient, and more productive, enterprises are accelerating their use of data. Service and technology providers are striving to overcome technical gaps in transport and computing infrastructure that could inhibit new operational models and services. Photonics can provide a bridge to the future.