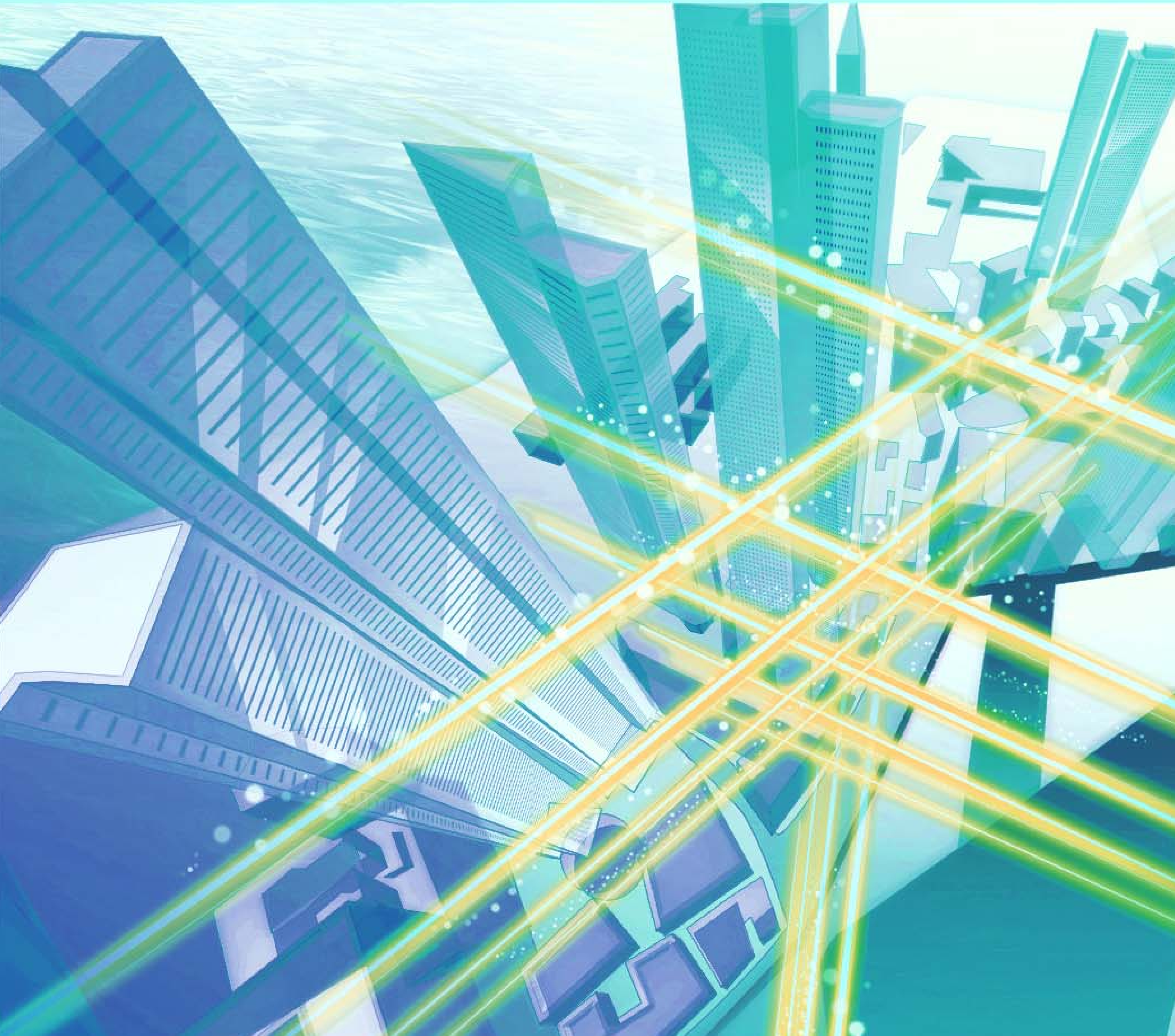
NTT Technical Review

5
2022

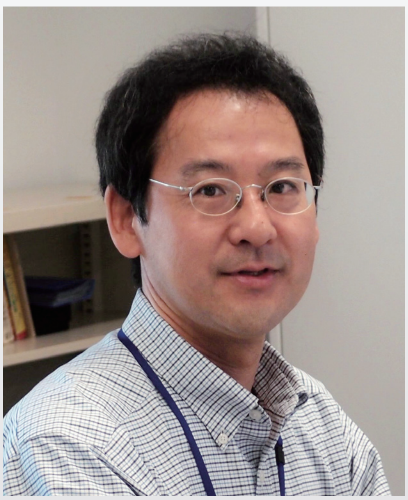


May 2022 Vol. 20 No. 5

NTT Technical Review

May 2022 Vol. 20 No. 5

Problem Solving Will Not Reduce the Number of Research Themes—It Will Open up New Research Areas



Hiroshi Sawada
Senior Distinguished Researcher, NTT Communication Science Laboratories

Abstract

From his extensive research in blind source separation, Senior Distinguished Researcher Hiroshi Sawada has received worldwide acclaim for his work with a domestic research partner on independent low-rank matrix analysis, namely, the development and integration of independent component analysis and non-negative matrix factorization. He has also ventured into the field of neural networks. We asked him about his research achievements and his attitude toward enhancing his research activities.

Keywords: blind source separation, independent low-rank matrix analysis, neural networks

The culmination of research on audio source separation: ILRMA showcases Japan's presence in the field

—Could you tell us about the research you are currently conducting?

I have been researching for a long time on *blind source separation*, which is a technology that separates mixed sound sources in situations in which a listener with eyes closed cannot tell under what conditions a recording is being made. Expanding a portion of this technology, I have also studied technologies for analyzing spatio-temporal data. More recently, I have also ventured into a new field for me, neural networks, which many people have been researching.

I developed blind source separation into a technology for estimating the structure of information sources and observation systems by combining non-negative matrix factorization (NMF), which captures the structure and features of information sources such as data and signals, and independent component analysis (ICA), which estimates how data and signals are observed with sensors through an observation system. NMF has developed into a method for analyzing spatio-temporal multidimensional data sets—by modelling spatio-temporal relationships between multidimensional data—to enable future prediction. It has recently evolved into a method for data assimilation and machine-learning-based crowd navigation (**Fig. 1**).

To reduce congestion and stabilize communication infrastructure at large-scale events, this data assimilation and machine-learning-based crowd navigation technology uses (i) real-time-observation data to

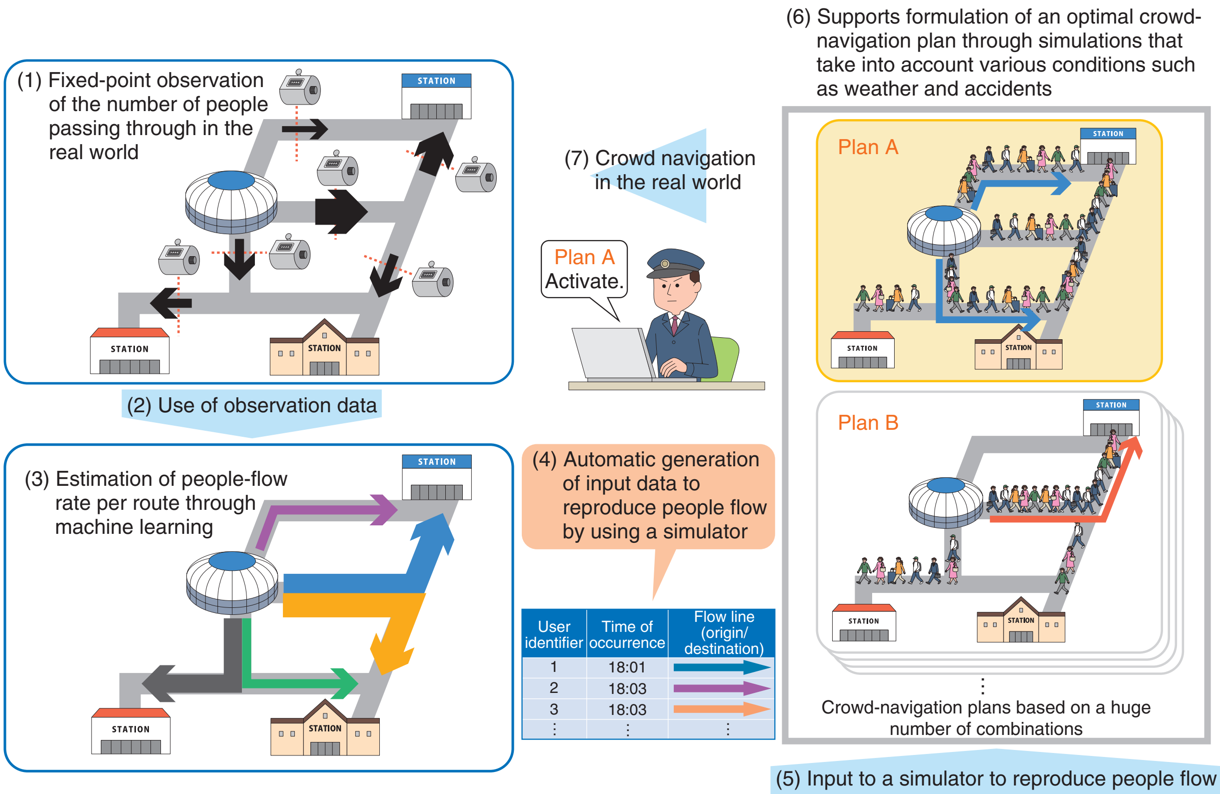


Fig. 1. Data assimilation and machine-learning-based crowd navigation.

detect incidents, such as congestion, that may occur in the near future through data assimilation and simulation and (ii) machine-learning-based crowd navigation to prevent congestion from occurring and ensure safety in advance.

Headed by NTT Fellow Naonori Ueda at NTT Communication Science Laboratories, many NTT researchers, including myself, are working on this technology. We had hoped to demonstrate the technology at the major international sporting event held in Tokyo in 2021; however, the event was closed to spectators to prevent the spread of COVID-19, so we were unable to demonstrate the technology. Led by Hitoshi Shimizu, a research scientist at NTT Communication Science Laboratories, we also work on research themes such as "a model for selection of attractions in a theme park and estimation of model parameters" and "theme-park simulation based on questionnaires for maximizing visitor surplus."

At the same time, by advancing blind source separation, we have been promoting *independent low-rank matrix analysis* (ILRMA) (**Fig. 2**), which integrates ICA and NMF. As the result of joint research with Tokyo Metropolitan University, National Institute of Technology, Kagawa College, and the University of Tokyo, the basic technology underpinning ILRMA was announced at an international conference in 2015 and described in a journal in 2016. After we further developed the technology, I gave a tutorial lecture on ILRMA at the International Conference on Acoustics, Speech and Signal Processing (ICASSP) 2018, the world's largest international conference in the field of audio and acoustic signal processing. After summarizing the content of that tutorial as a review paper in Asia-Pacific Signal and Information Processing Association (APSIPA) Transactions on Signal and Information Processing in 2019 [1], we received the APSIPA Sadaoki Furui Prize Paper Award. I also co-authored a tutorial lecture on further advanced topics for the European Signal Processing Conference (EUSIPCO) 2020.

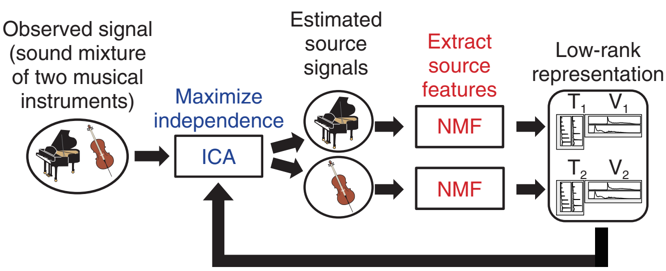


Fig. 2. ILRMA.

—It is a pity to hear about the impact of the COVID-19 pandemic; even so, you have achieved world-class results.

In recognition of these activities and past research results on blind source separation, ICA, and NMF, I was selected as a Fellow of the Institute of Electrical and Electronics Engineers (IEEE), the world's largest academic organization in the field of electrical and information engineering, in 2018, and as a Fellow of the Institute of Electronics, Information and Communication Engineers (IEICE) of Japan in 2019. In 2020, I received the Kenjiro Takayanagi Achievement Award, which honors individuals who contributed to the promotion and industrial development of the field of electronic science and technology in Japan through outstanding research. In awarding the prize, the committee stated, "ILRMA is the culmination of a series of studies on blind source separation that have been clarified as the original technology from Japan, and it is appealing to the world through an edited book, tutorials at international conferences, and a review paper." In January 2022, I was named an IEEE Signal Processing Society Distinguished Lecturer, an honor bestowed annually on five individuals in each field. I am proud that we could demonstrate the presence of NTT and Japan as a whole through ILRMA and other technologies.

Boost your organization and polish your skills

—Neural-network technology is a different field from ILRMA. How did you come up with it as a research theme?

While I have been focusing on research on blind source separation and analysis of spatio-temporal data, the third boom in artificial intelligence (AI) started around 2012, since then, neural networks based on deep learning have gradually become used in products and services in our daily lives. Although I had not been directly involved in neural networks, around 2017, like many other researchers, I began to think that neural networks was a very important technology and started learning it, even though I had to catch up with other researchers at first. The first step in taking on the challenge of such a new field is learning about it.

First, to boost research activity on neural networks from the perspective of NTT's laboratories as a whole, I started a colloquium on deep learning and a technical course on machine learning as part of my studies. The colloquium on deep learning is organized by members of NTT laboratories. For about 20 years, we have been organizing colloquiums on themes, such as speech and language, to bring experts together at least once a year who are based at the laboratories in different locations such as Musashino, Yokosuka, Atsugi, Keihanna, and Tsukuba.

The colloquium on deep learning is now five years old. Although it has been held online for the past two years due to the COVID-19 pandemic, it has allowed us to see who in the laboratories are pursuing technology in the field of deep learning and facilitated information exchange and discussion. Technical courses are also held for new employees and employees in their second to third year of employment. We have had technical courses on networks, information theory, etc. but not on machine learning, so we took this opportunity to launch one and augment the neural-network content from around 2019.

Through these efforts, I have deepened my own understanding of neural networks and am now able to use them in new research projects such as three related studies that I am currently conducting with collaborative researchers. One of these studies, an "acceleration method for learning fine-layered optical neural networks" (**Fig. 3**) [2], was developed with

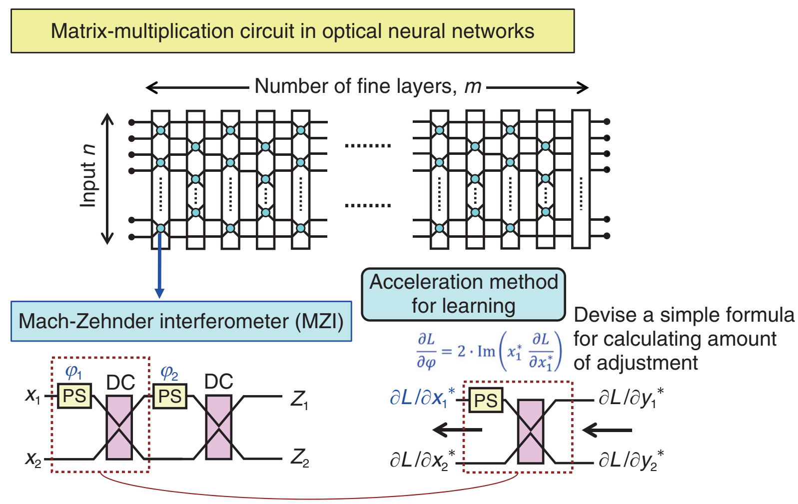


Fig. 3. Acceleration method for learning fine-layered optical neural networks.

Kazuo Aoyama, a researcher at NTT Communication Science Laboratories, and a paper on it was accepted to the International Conference on Computer-Aided Design (ICCAD) 2021. In addition to its novelty in the field of machine learning, the team headed by Dr. Masaya Notomi, a senior distinguished researcher at NTT Basic Research Laboratories, which has been collaborating with us in this study, uses the parameters learned with this method in experiments on actual optical devices.

In the Frontier Research and Development Committee of the Japan Society for the Promotion of Science "Strategic plan for industrial innovation platform by materials informatics," on which I served as a committee member for three years from 2016, we have been discussing how to use machine-learning techniques for developing new physical properties. In a joint research project led by Dr. Yuki K. Wakabayashi, a researcher at NTT Basic Research Laboratories, and Dr. Takuma Otsuka, a research scientist at NTT Communication Science Laboratories, we were able to use machine-learning techniques to derive the temperature and other conditions necessary to create a thin film with outstanding properties [3]. In collaboration with the University of Tokyo and Tokyo Institute of Technology, we have measured the electrical conduction of single-crystalline thin films of $SrRuO_3$ (Sr: strontium, Ru: ruthenium, O_3: ozone) prepared—at low temperature with a magnetic field—and were able to announce the first observation of quantum transport phenomena, which is peculiar to an exotic state called a magnetic Weyl semimetal [4]. I have been researching themes in the information field all my life and never dreamed that I would be involved in creating a material with new properties, so I'm very happy to have been able to be involved in this creation, even if only a little.

—What else is important to you to enhance your research activities using technology in your area of expertise?

Think about what technology is your specialty, how you can use it to play an active role, and what technology can contribute to the development of the field in which you are researching. For example, optical neural networks use complex numbers, and since complex numbers appear in research on ICA (namely, when an audio signal is converted into its frequency components using the Fourier transform), I thought I could make a contribution to that field. It is important to find areas that overlap with your field and areas in which your specialized skills can be used, rather than areas that are completely unknown to you. It might be that the more technologies you are specialized in, the better, but I think one is enough.

It is also important to learn and acquire new skills in the field you are new to and to find research colleagues and supporters. For example, the above-mentioned acceleration method for learning optical neural networks worked largely because my colleague, Kazuo Aoyama, was working with me on that project. I was also attracted to the fact that Dr. Masaya Notomi had expertise that I did not have, and I felt a sense of anticipation that synergy would occur between us and something new would be created by combining our experience and skills.

You can meet such people at the colloquiums I mentioned earlier and social gatherings, so it might be a good idea to take advantage of opportunities for casual chats at such social gatherings to get a feel for what different fields are like. I'm also a member of the Machine Learning and Data Science Center, which was established by NTT Fellow Naonori Ueda, and the information shared there has led to new research opportunities for me.

Incidentally, as research progresses and results are obtained over the years, the number of research themes naturally decreases as problems are solved. Instead of worrying about that fact, I want to approach my research activities with the attitude that there are more things we can do, and new research areas are being opened up. That attitude will perhaps lead you to search for collaborators and colleagues. Because what one person can do alone is limited, I have always valued collaboration with other researchers. Means of collaboration include effectively combining different fields of expertise or mutually understanding, confirming, and deepening each other's skills in the same field of expertise. However, as a premise, you must be an entity that can be interesting to and trusted by the other party. To gain trust in terms of the expertise that is expected of me, I want to do the things I am required of, such as provide something of value, take charge of my part in building experiments and systems, and present my ideas clearly when discussing issues.

Why don't you just do what you want to do without worrying too much?

—How do you plan to lead your life going forward as a researcher?

I want to continue my research as long as possible. One reason for that is the pleasure of learning from past research activities. Neural networks provide a good example of such learning. Such research has been continuing for more than 60 years, and looking at past research from new perspectives and techniques can reveal connections between old and new research.

Just as I have found a new field of activity in research on neural networks by learning it from scratch, I'm sure that the same thing will happen when I'm pursuing other themes in the future. As a researcher, I'm forever having to learn. It is fun for me to learn new things, so I hope to continue to learn well into the future.

I believe that a researcher is a person who creates novelty regardless of whether it is useful to anyone. New is not always good; that is, it is important to produce results that other researchers will find valuable and that they are willing to base new research on. There are many cutting-edge researchers, and many young people are reading new research papers. I want to challenge myself to find something that I'd like to try my hand at.

—Now that you have become a researcher with a major impact on the world, what words would you like to say to the person you were when you joined the company? Based on those words, what would you like to say to young researchers?

That is a bit difficult to answer, but I suppose I would say, "Why don't you just do what you want to do without worrying too much?" Looking back on myself, I was not very successful at first and was anxious about whether I would be able to produce results. I'd like to tell young researchers that their hard work will produce results in one way or another.

I feel pleasure in explaining difficult things in a simple and easy-to-understand way when I'm giving a tutorial lecture at IEEE, giving a lecture during a technical course, or writing a paper, and above all, when new research results are accepted for publication. In fact, a figure from my paper accepted by IEEE in 2013 appeared on the cover of the journal [5]. That made me very happy. I had worked hard with PowerPoint to create a diagram that would make it easy to understand concepts that would be difficult to understand if written in mathematical formulas. When the diagram was published, I felt that my hard work had paid off.

I think young researchers in our organization are working very hard. The field of machine learning, in particular, is currently booming, and many talented researchers are entering the field. It's very difficult to get your paper accepted by an international conference because you have to overcome high hurdles. Nevertheless, it's important to publicize your results, so let's continue improving the quality of our work, publishing our papers on arXiv (an open access repository of scientific research), and aiming for prestigious conferences. Keep in mind that doing so is not easy, and you may feel discouraged if your paper is not accepted; even so, at such times, look to the experience of your seniors: many of them had their results accepted only after years of hard work. Seniors, including myself, can be collaborators and help juniors write papers. It is also helpful to make steady efforts while taking a realistic approach.

References

[1] H. Sawada, N. Ono, H. Kameoka, D. Kitamura, and H. Saruwatari, "A Review of Blind Source Separation Methods: Two Converging Routes to ILRMA Originating from ICA and NMF," APSIPA Transactions on Signal and Information Processing, Vol. 8, No. 1, e12, 2019.

[2] K. Aoyama and H. Sawada, "Acceleration Method for Learning Fine-layered Optical Neural Networks," arXiv:2109.01731, 2021.

[3] Y. K. Wakabayashi, T. Otsuka, Y. Krockenberger, H. Sawada, Y. Taniyasu, and H. Yamamoto, "Machine-learning-assisted Thin-film Growth: Bayesian Optimization in Molecular Beam Epitaxy of $SrRuO_3$ Thin Films," APL Materials, Vol. 7, 101114, 2019.

[4] Press release issued by NTT, The University of Tokyo, and Tokyo Institute of Technology, "First Observation of Quantum Transport Phenomena Peculiar to an Exotic State," Oct. 9, 2020. https://group.ntt/en/newsrelease/2020/10/09/201009a.html

[5] Front cover of the IEEE Transactions on Audio, Speech, and Language Processing, Vol. 21, No. 5, 2013, https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6517958

■ Interviewee profile

Hiroshi Sawada received a B.E., M.E., and Ph.D. in information science from Kyoto University in 1991, 1993, and 2001. He joined NTT in 1993. His research interests include statistical signal processing, audio source separation, array signal processing, machine learning, latent variable models, graph-based data structures, and computer architecture. From 2006 to 2009, he served as an associate editor of the IEEE Transactions on Audio, Speech and Language Processing. He received the Best Paper Award of the IEEE Circuit and System Society in 2000, the SPIE ICA Unsupervised Learning Pioneer Award in 2013, the Best Paper Award of the IEEE Signal Processing Society in 2014, and the APSIPA Sadaoki Furui Prize Paper Award in 2021. He was selected to serve as an IEEE Signal Processing Society Distinguished Lecturer for the term 1 January 2022 through 31 December 2023. He is an IEEE Fellow, an IEICE Fellow, and a member of the Acoustical Society of Japan.

Research on Reducing CO_2 in the Ocean through the Application of Genome Editing Technology to the Carbon Cycle of Algae, Fish, and Shellfish

Sousuke Imamura
Distinguished Researcher, NTT Space Environment and Energy Laboratories

Abstract

According to the Intergovernmental Panel on Climate Change (IPCC) Sixth Assessment Report, 57.7% of carbon dioxide (CO_2) emissions into the atmosphere are absorbed by forests, and 34.6% is absorbed by the ocean. In this article, we interviewed Distinguished Researcher Dr. Sousuke Imamura about "CO_2 reduction technology in the ocean based on application of the genome editing technology."

Keywords: CO_2 reduction, genome editing, algae

Reducing CO_2 in the ocean through genome editing in algae, fish, and shellfish

—How is genome editing technology used to reduce the amount of CO_2 in the ocean?

According to the Intergovernmental Panel on Climate Change (IPCC) Sixth Assessment Report, only 4.8% of total carbon dioxide (CO_2) emissions into the atmosphere come from human activities. Of the remaining emissions, soil accounts for 61.3% and the ocean accounts for 33%. On the other hand, up to 57.7% is absorbed by forests, and the ocean also absorbs 34.6%.

NTT's communications services emit CO_2 as energy is consumed, so researching and developing technologies that reduce these emissions is vital both for the unimpeded development of the communications business and to reduce our carbon footprint. In addition to CO_2 reduction measures such as energy conservation, another feasible approach to absorbing and reducing CO_2 emissions is to harness the power of living organisms. Our focus is on the carbon cycle of the ocean, and we're conducting research into reducing CO_2 through the use of genome editing technology.

Figure 1 shows an overview of the technology used to reduce CO_2 in the ocean by applying genome editing technology to the carbon cycle of algae, fish, and shellfish. When atmospheric CO_2 dissolves into the

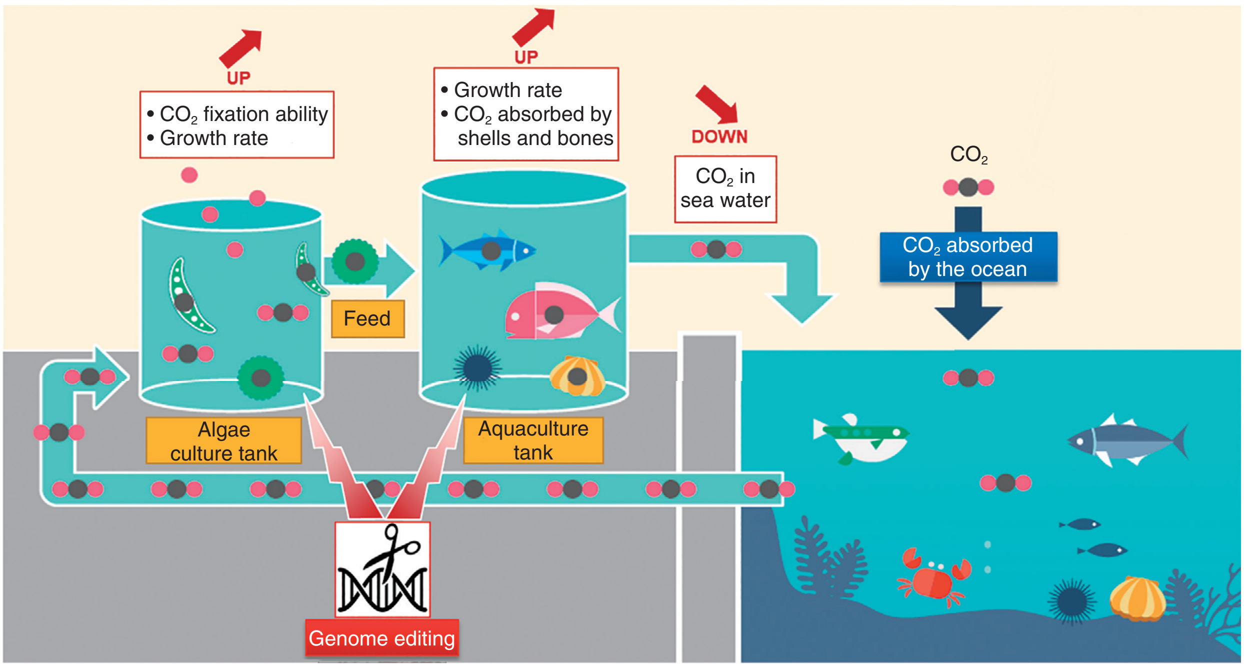


Fig. 1. Reducing CO_2 in the ocean by applying genome editing technology.

ocean, it is fixed by ocean-dwelling phytoplankton such as algae. The term "fixed" may sound strange, but it refers to the process of carbon fixation, in which inorganic carbon such as CO_2 is converted into organic carbon such as sugar and assimilated into the body. The carbon fixed by the algae is then taken in by fish and shellfish that eat the algae as food. This is the carbon cycle of the marine food chain.

We're trying to apply genome editing technology to key parts of the marine food chain, including algae, fish, and shellfish, which is a world first. As a result, we aim to encourage increased CO_2 fixation in algae and accelerate the growth of fish and shellfish, thereby increasing the amount of CO_2 absorbed by their bones and shells. As such, this research is aimed at synergistically increasing the total amount of carbon held in the carbon cycle by algae, fish, and shellfish, thereby reducing the amount of CO_2 present in seawater.

As the concentration of CO_2 in seawater decreases, the amount of CO_2 dissolved from the atmosphere increases accordingly, thus reducing the amount of CO_2 in the atmosphere. We've started a proof of concept with Regional Fish Institute, Ltd. (RFI), in which NTT will work on research and development of genome editing technologies that increase the amount of CO_2 fixation in algae, and RFI will work on research and development of genome editing technologies that increase the amount of carbon stored in the bodies of fish and shellfish.

—What drew your attention to algae?

Algae are identifiable as a phytoplankton that contribute to CO_2 fixation in the ocean, so they seemed like a natural area to focus on. They've been around for some three billion years, all the while producing oxygen through photosynthesis. No other organisms have a history comparable to that. We're breathing now thanks to the oxygen they've stored in the atmosphere, and the evolution of plants we're familiar with today can be traced back to algae. It's no exaggeration to say that algae have shaped the global environment as we know it. I felt algae would be a good route to understanding plants, so I've been studying them since my undergrad days.

In addition, since many of algae are single-cell organisms, they are easy subjects for genome editing.

—What exactly is genome editing technology?

The term "genome editing" is often confused with genetic modification. It's common to come across disclaimers about genetic modification on food packaging. For example, in Japan, you see messages like "does not contain genetically modified soybeans" on natto (fermented soybean) packages. Genetic modification involves taking genes derived from one thing and inserting them into another, like if you added photosynthesis genes from algae to the human body.

Genome editing technology, meanwhile, is about altering genetic sequences, or subtly editing the order of genes, to put it another way. It's considered relatively safe compared to genetic modification, because it doesn't involve inserting genes derived from an external organism into the body of another—it's just a case of changing gene sequences that are already there. It's the same basic principle as selective breeding of crops.

That said, it is vital that we carefully assess the environmental impact of genome-edited organisms and whether or not it could lead to any damage to biodiversity. That's why we're conducting our proof of concept in a confined space through land-based aquaculture (Fig. 1).

Our goal over the next five years is to put our first set of algae and sea creatures to practical use

—Can you tell us about where you're currently at with this research?

Right now we're conducting research on three fronts.

The first is research on feeding. We've started actually testing feeding the cultivated algae to the fish and shellfish, but as things stand, it seems that they have very particular preferences and won't really eat the algae. So we're looking into which types of algae they'll ingest more efficiently, and at what concentrations.

Our second area of research is genome editing. When editing genomes, you need to determine which sequences of which genes you want to edit. We're currently investigating the best conditions for editing genomes in order to produce the desired effect.

Our third point of research is reducing unwanted side effects caused by genome editing. As I mentioned earlier, algal mechanisms have been accumulated and refined over three billion years. If human beings start meddling with those mechanisms, there are bound to be consequences. For example, increasing CO_2 absorption or fixation ability may have a negative impact on growth. So we're researching the best ways of mitigating such trade-offs.

—What are your plans for future research?

The ocean is home to a remarkably wide variety of species. I began by outlining the percentages of CO_2 emissions and absorption, but these are just calculations at the end of the day. We can only make rough estimates derived from the assumption that all organisms would function in similar ways, based on existing data from a selection of algae and other species. We hope that by experimenting with land-based aquaculture in a confined space, we'll have a platform to generate quantitative data on the levels of CO_2 fixation in algae, to what extent the carbon is passed to fish and shellfish, and the overall effect on CO_2 levels. This approach will be an important point of focus for future practical applications. Our goal over the next five years is to determine the best combination of Alga A and Fish or Shellfish B, to evaluate the effectiveness and safety of this CO_2 reduction method, and to put our findings to practical use. We should then be able to explore how this technology will be used out in the actual ocean environment, not just within a confined space.

So far, I've talked about reducing CO_2 from an environmental perspective, but another important consideration is food. RFI has already successfully edited genomes to develop fish that could compensate for food shortages, such as red sea bream with more edible parts and tiger pufferfish that grow at an accelerated pace. By feeding algae that efficiently fix CO_2 in the ocean to genome-edited fish and shellfish, this technology should be capable of both reducing CO_2 levels and producing food at the same time. That said, genome editing is still a new and relatively untried technology, so it's essential to provide consumers with accurate information in order to dispel any concerns they might have about genome-edited food. In terms of shellfish, we're also conducting research into Akoya pearl oysters. Even people who are averse to eating genome-edited food may find it easier to accept decorative objects such as the pearls produced by these oysters.

Of course, our main objective is to increase CO_2 fixation, but there will likely be a growing need to develop research like this that adds value to fish and shellfish in the future.

We're also considering how to apply this technology to soil. Soil accounts for the largest proportion of CO_2 emissions, and crucially, it's the source of nutrition for plants. Any reduction in the amount of CO_2 generated from soil is important, no matter how small. Although there are many differences between soil and ocean, microorganisms play a huge role in both, so I have high hopes for some breakthrough findings in this area.

—Is there anything you would like to say to students, young researchers, and future business partners?

In order to accelerate the research I've described today and develop it into a more versatile technology, we're looking to work together with data scientists who can help us digitize and analyze data relating to genome editing, algae cultivation and feeding, as well as collaborating with people who specialize in entire ecosystems.

At NTT Space Environment and Energy Laboratories, we conduct research in an extensive range of fields. In addition to research like ours that explores and pursues a deeper understanding of life, we conduct research into the fundamentals of communications, networks and the engineering that supports these, as well as areas like energy and social sciences. We are working on new and challenging themes that have never been tackled before. I joined this research lab in March 2021 straight from the world of academia, and am excited to be working on such large-scale and well-balanced projects, from research through development.

At present, there remains a great deal we don't know about various biological mechanisms. In this field alone, I expect many new ideas to come to the fore, which could then develop into new areas of research as our understanding grows. NTT Space Environment and Energy Laboratories is an environment in which you can challenge yourself to try new things. If you're an innately curious person, we'd love to have you on board.

■ Interviewee profile

Before joining NTT in 2021, Sousuke Imamura completed a doctoral course in biotechnology at the United Graduate School of Agricultural Science at the Tokyo University of Agriculture and Technology, became a research fellow at the University of Tokyo, gained a Research Fellowship PD from the Japan Society for Promotion of Science at the University of Tokyo, became an Assistant Professor in the Department of Life Sciences at the Chuo University Faculty of Science and Engineering, and became an Associate Professor at the Laboratory for Chemistry and Life Science in the Institute of Innovative Research at Tokyo Institute of Technology. He is a member of the Sustainable Systems Group at NTT Space Environment and Energy Laboratories, as well as a lecturer at Meiji University's School of Agriculture and a visiting professor at the Laboratory for Chemistry and Life Science in the Institute of Innovative Research at Tokyo Institute of Technology. His specialty is plant molecular biology.

Progress in the Development of IOWN Technology

Masahisa Kawashima and Yosuke Aragane

Abstract

Technical improvements in full-stack layers are essential to develop a future infrastructure with higher capacity, lower latency, and lower power consumption toward a cyber-physical society. NTT announced its roadmap for this technology, the Innovative Optical and Wireless Network (IOWN), in April 2020 and has accelerated to implement the roadmap with global partners of the IOWN Global Forum (IOWN GF). IOWN GF developed and published a series of deliverables on Open All-Photonic Network (Open APN), IOWN for Mobile Network, Fiber Sensing for Open APN, Data-Centric Infrastructure, IOWN Data Hub, and a Reference Implementation Model. This article provides an overview of these deliverables and describes the growth of IOWN GF and its activities.

Keywords: All-Photonics Network, disaggregated computing, data hub, IOWN

1. Why the Innovative Optical and Wireless Network (IOWN)

A cyber-physical system is expected to provide a more productive and sustainable society. However, the capabilities of the current network and computing infrastructures cannot satisfy these requirements. Let us look at a use case for preventing collisions at road intersections by installing artificial intelligence (AI) network cameras into curve mirrors (**Fig. 1**) and the following four issues that need to be addressed.

The first issue is the processing capabilities of servers. Because there are many intersections in a city, it is necessary to upgrade server-processing capability to accommodate as many cameras as possible. Although the capabilities and efficiency of accelerators have improved, the current mechanism for receiving and sending data with AI accelerators forms a bottleneck.

The second issue is latency. The latency from capturing a dangerous situation to delivering the control command to the vehicle or person about to collide should be less than 10 ms. However, the current TCP/IP (Transmission Control Protocol/Internet Protocol) requires several round trips for the flow control to transfer image data.

The third issue is reliability. In other use cases, such as controlling drones and autonomous vehicles, communications should be connected continuously. Since high-capacity mobile communications, such as the fifth-generation mobile communication system (5G), use a high-frequency band, communication interruptions due to obstacles become more frequent.

The fourth issue is power consumption. Reducing power consumption is one of the most important social issues. We have to develop systems with as low power consumption as possible. However, according to our research, a computer infrastructure for AI analysis consumes more than 10 W of power per video stream, equivalent to the power consumed by an incandescent light bulb [1]. To reduce power consumption, it is necessary to conduct AI analysis only when there is something physically on the camera such as people and objects. This feature requires light-weight pre-processing to find the something on the camera before AI analysis. The current major technical approach tends to use edge computing for the light-weight pre-processing and reducing the response time. However, using edge computing for low-latency execution reduces the efficiency of work among datacenters and difficult to control on-demand information technology (IT) resource usage.

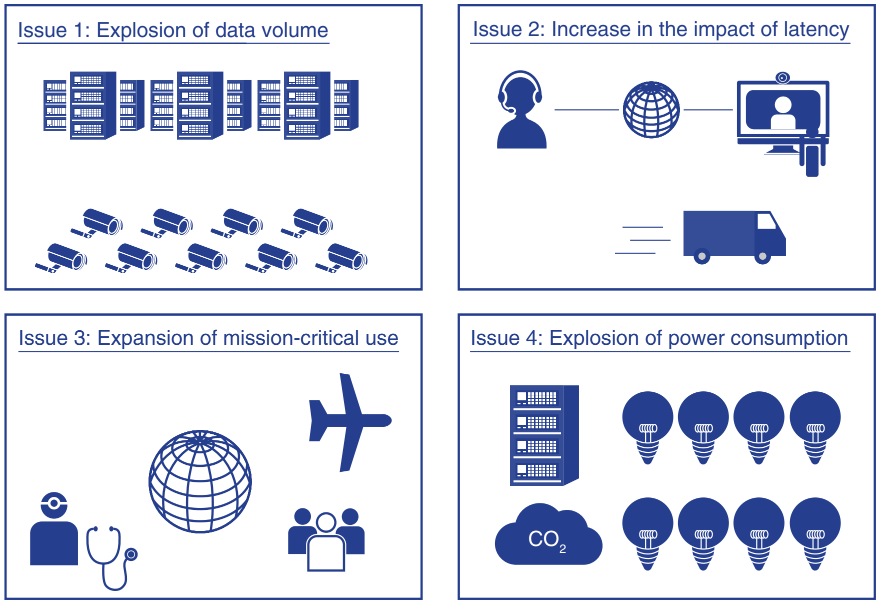


Fig. 1. Issues to be addressed.

These issues have been known in the communications and IT industries for several years, and high-speed mobile networks, such as 5G and edge computing, have been under consideration. However, in network services, packet-based transmission, which is a best-effort approach, has not been re-examined. Improvements have been made in individual layers such as communications and IT, but no matter how fast the network is, the best-effort approach requires delays for the flow control. Edge computing with the best-effort approach results in inefficient IT resource usage due to the split of work among datacenters; thus, full-stack technology is required instead of layer-dedicated improvements of technologies.

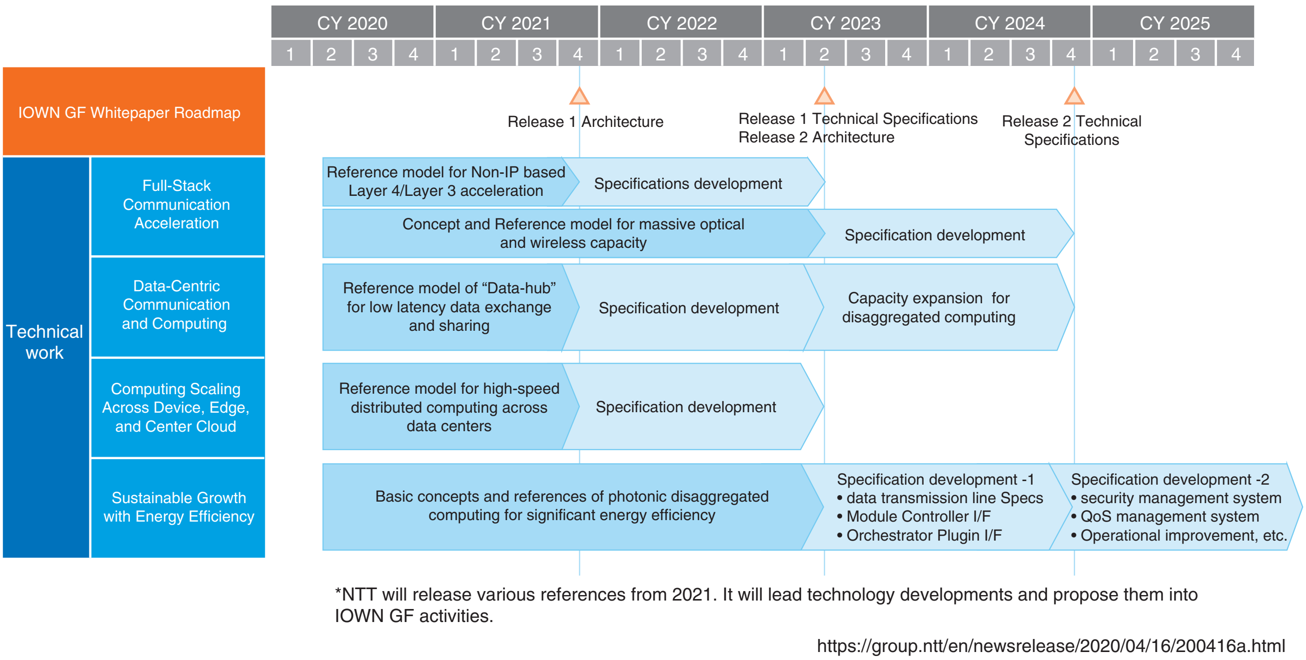
To advance full-stack technology in an open community, NTT, Intel, and Sony established the IOWN Global Forum (IOWN GF) in January 2020. In April 2020, IOWN GF published a white paper summarizing its expected innovations followed by NTT's announcement of NTT's technology development roadmap called "IOWN Technology Development Roadmap" (**Fig. 2**).

2. IOWN roadmap and IOWN GF

With the participation of many organizations worldwide, discussions at IOWN GF are accelerating. Key technologies defined in the roadmap are also being developed in collaboration with IOWN GF members. As a result of these efforts, six technical documents were published by IOWN GF [1]. The Feature Articles in this issue explain some of the key technologies described in these documents [2–4]. This section provides an overview of these documents, including their relationship to the IOWN Technology Development Roadmap. These documents are also linked to the functional structure image of IOWN, introduced in this journal in 2020 "Initiatives to Achieve the IOWN (Innovative Optical and Wireless Network) Concept" [5] (**Fig. 3**).

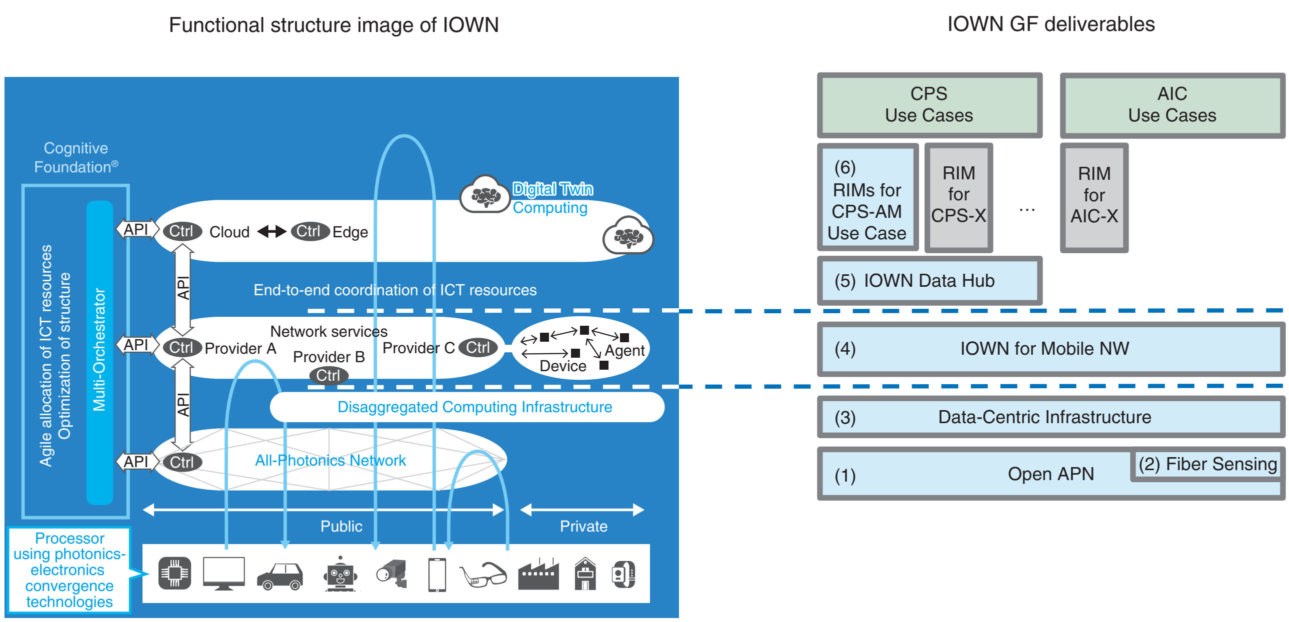
IOWN GF published two technical documents on the Open All-Photonic Network (Open APN) (Fig. 3(1)) and IOWN for Mobile Network (NW) (Fig. 3(4)). This is related to the "Concept and reference model for massive optical and wireless capacity" of the IOWN Technology Development Roadmap. The Open APN document defines an open architecture for building the All-Photonics Network (APN) with multiple vendors, which enables communication with a deterministic transmission rate and latency. The IOWN for Mobile NW document defines the building of a wireless network on IOWN infrastructure that achieves both high capacity and high reliability. It shows how the APN and disaggregated computing can be applied to the O-RAN (Open Radio Access Network) architecture.

The ability of optical fibers to propagate optical signals far away can be applied not only to data



I/F: interface
QoS: quality of service

Fig. 2. IOWN Technology Development Roadmap.



AIC: AI-integrated communication
AM: area management
API: application programming interface
CPS: cyber-physical system
Ctrl: controller
ICT: information and communication technology

Fig. 3. IOWN concept and IOWN GF activities.

communication but also to sensing. Businesses and public institutions that invest in fiber infrastructure for the APN will be able to use such fibers for sensing. The APN, which enables a large-capacity communication infrastructure, also enables the collection of sensing signals. For this technology, the architecture for adding the sensing function to the APN was also discussed in the document published as "Fiber Sensing for Open APN" (Fig. 3(2)).

IOWN GF also published a technical document on data-centric infrastructure (DCI) (Fig. 3(3)). It is related to the "Basic concepts and references of photonic disaggregated computing for significant energy efficiency" of the IOWN Technology Development Roadmap. DCI provides an efficient mechanism for sending data to the accelerators by equipping accelerators with a network input/output.

Data plane acceleration (DPA) in the DCI document presents the protocol stack and design for transferring data at high speed and low latency by using the APN between two remote DCI infrastructures. This is related to the "Reference model for non-IP based Layer 4/Layer 3 acceleration" of the IOWN Technology Development Roadmap. DPA allows multiple DCIs deployed on adjacent datacenters to share IT resources, reducing the loss due to work split among datacenters. IOWN can provide a large virtual datacenter (clustered datacenter) by combining multiple datacenters with the APN.

A clustered datacenter enables scalable and high-available database/storage with distributed database/storage among those in the cluster. IOWN GF developed this cloud-native database and storage architecture as the IOWN Data Hub (Fig. 3(5)). This is related to the "Reference model of 'Data Hub' for low-latency data exchange and sharing" of the IOWN Technology Development Roadmap. The IOWN Data Hub will be an infrastructure that connects countless data providers and data users in a cyber-physical society. In other words, the network is the database. In a cyber-physical society, we believe network infrastructure must provide not only data-transfer functions but also database functions.

IOWN GF also decided to develop reference implementation model (RIM) documents for specific use cases (Fig. 3(6)). This is related to the "Reference model for high-speed distributed computing across datacenters" of the IOWN Technology Development Roadmap. Full-stack technology should be developed focusing on the characteristics of each use case. IOWN GF has published an RIM for area management.

3. IOWN GF and its activities

3.1 Technologies and use cases

IOWN is taking on the ambitious challenge of creating an innovative infrastructure for both communications and computing. Therefore, IOWN GF needs to drive both technology development for the IOWN infrastructure and use cases of what future values those technologies can create. For developing innovative and valuable outcomes, IOWN GF established two types of Working Groups, i.e., Technology and Use Case, and Steering Committees to coordinate inter-working group activities (**Fig. 4**).

As can be seen from the member list on the IOWN GF website [6], not only companies and organizations that develop IOWN technologies but also many companies that use IOWN technologies to expand their businesses are participating. In use-case discussions, some members share the business challenges and future risks that they are facing and ask other members whether they can be addressed by IOWN GF activities or whether it is an attractive use case and something they should discuss together to address such challenges and risks.

To meet these business requirements, the Use Case Working Group tries to estimate numerical requirements such as response time, network bandwidth, and power consumption. The Use Case Working Group asks the Technology Working Group to develop innovative technologies to meet these requirements.

3.2 Native online activities

In the preparation phase for IOWN GF establishment, face-to-face meetings were considered essential to build relationships among the members and accelerate the forum activities. However, due to the COVID-19 pandemic, IOWN GF has been forced to operate fully online since its establishment. IOWN GF has been trying to determine if it can build relationships between members in this situation and have constructive discussions on how to create deliverables.

IOWN GF is a forum with diverse members from the US, Europe, and Asia, so it must consider several time zones. If an online meeting is held for 6 hours, like a normal international meeting, it will be late at night or early in the morning in some regions. IOWN GF has tried to limit the meeting time to no more than two hours to avoid late night and/or early morning scheduling for most members. The shorter meeting times resulted in more frequent meetings, so each task-force meeting has been arranged biweekly or

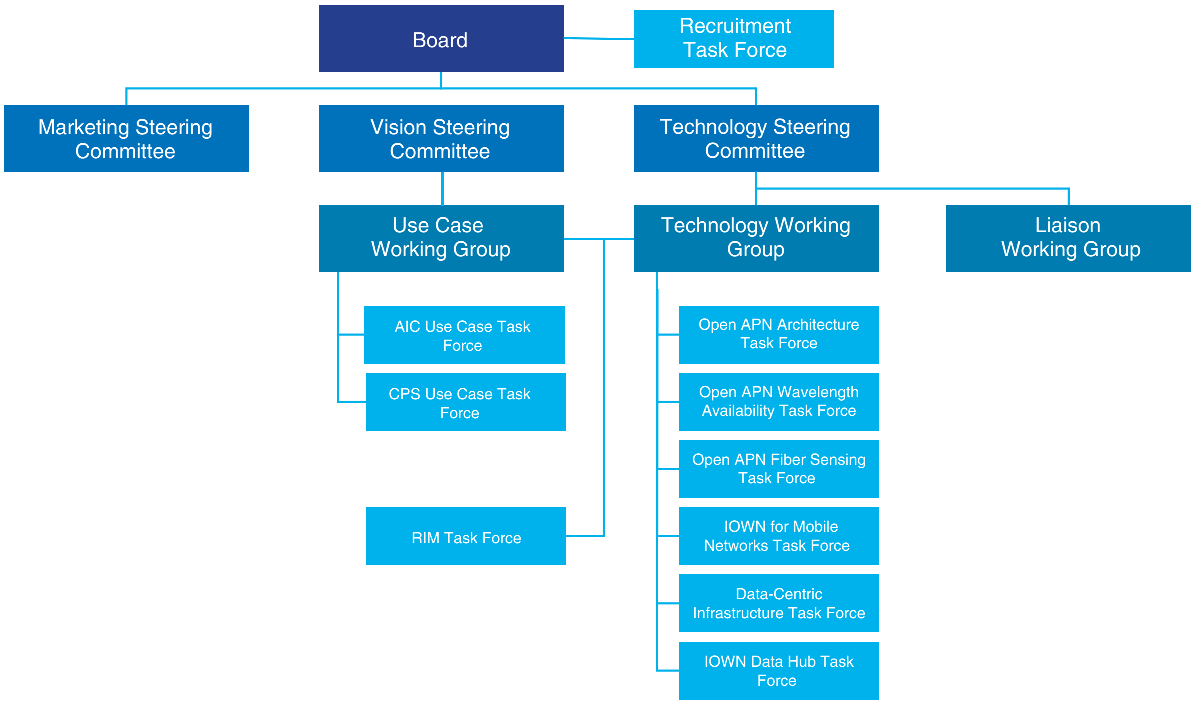


Fig. 4. IOWN GF organization.

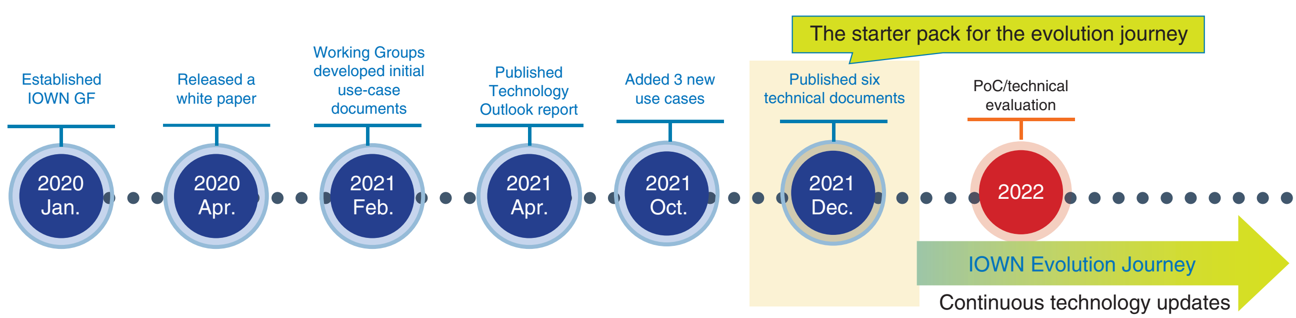


Fig. 5. IOWN GF achievements.

weekly. In these short meetings, members have been able to share their progress with each other, agree on the next action items, and provide feedback on the results of each member's activity on the action item.

This approach, however, means that it is difficult to follow the traditional document-contribution basis approach in major international organizations. It is difficult to implement a biweekly or weekly process in which each member submits a well-thought-out contribution document and takes much time to discuss and agree on each contribution. Therefore, IOWN GF has set up an online workspace; thus, members have been able to work together and develop deliverables more effectively and promptly. Despite the fact that all meetings have been online, IOWN GF has been able to frequently create and publish multiple results, including white papers, use-case documents, and technical documents, within a short period of only two years (**Fig. 5**). In addition to Working Group meetings and task-force meetings, IOWN GF has also regularly held online member meetings in which all members gather. The number of participants has grown to over 500. These results are due to the fact that IOWN GF is a native online

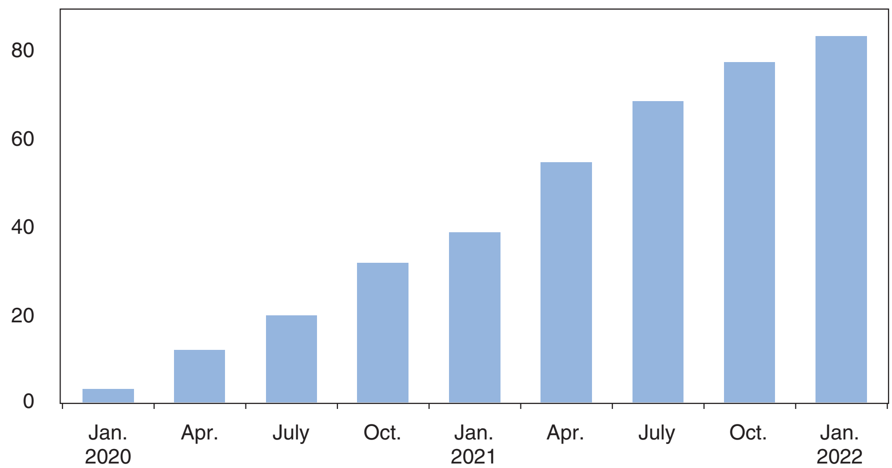


Fig. 6. Number of IOWN GF members.

organization.

3.3 Growth of activities

Through those activities and deliverables, various companies and organizations across industries have joined IOWN GF, and the number of members exceeded 80 companies/organizations at the end of 2021 (**Fig. 6**).

New members join almost every month. In December 2020, IOWN GF established a new membership for academic and research institutions (non-profits) to enable more organizations to participate in the development of IOWN technologies. Because IOWN GF uses online tools and has various task forces, it holds on-board sessions just before each member meeting to explain activities mainly to new members and supports members to smoothly participate in activities. As the number of members increases, the Board of Directors is also increasing. At the annual meeting in April 2021, Gilles Bourdon of Orange and Shingo Mizuno of Fujitsu were newly elected as Directors. Now a total of 9 Directors are responsible for decision-making for IOWN GF.

4. Plans in 2022

In less than two years since its establishment, IOWN GF has published technical documents. However, these deliverables are rough consensuses. Rather than taking the time to complete a document, it is more practical to put together a rough consensus in the short term, start a proof of concept (PoC) and technical evaluation early, and update the architectures and methodologies. Keeping agile thinking in mind, from 2022, IOWN GF will attempt to work on PoC activities and technical evaluation on the basis of published documents.

As indicated in the NTT Green Innovation toward 2040 [7] announced in 2021, NTT believes IOWN is an essential initiative for significantly reducing power consumption in its infrastructure. To achieve significant power savings, it is not enough to simply replace the infrastructure with an IOWN-based one but to redesign applications to take advantage of IOWN effectively and make operations more intelligent. In parallel with PoC activities, NTT will redesign applications and make operations more intelligent.

References

[1] IOWN GF, https://iowngf.org/technology/

[2] H. Nishizawa, J. Kani, T. Hamano, K. Takasugi, T. Yoshida, and S. Yasukawa, "Study on Open All-Photonic Network in IOWN Global Forum," NTT Technical Review, Vol. 20, No. 5, pp. 18–23, May 2022. https://ntt-review.jp/archive/ntttechnical.php?contents=ntr202205fa2.html

[3] H. Masutani, C. Schumacher, and K. Takasugi, "Data-centric-infrastructure Functional Architecture," NTT Technical Review, Vol. 20, No. 5, pp. 24–31, May 2022. https://ntt-review.jp/archive/ntttechnical.php?contents=ntr202205fa3.html

[4] T. Inoue, "Study of Storage Services at IOWN Global Forum," NTT Technical Review, Vol. 20, No. 5, pp. 32–36, May 2022. https://ntt-review.jp/archive/ntttechnical.php?contents=ntr202205fa4.html

[5] S. Iwashina, Y. Aragane, K. Minamihata, K. Shindo, and M. Fujiwara, "Initiatives to Achieve the IOWN (Innovative Optical and Wireless Network) Concept," NTT Technical Review, Vol. 18, No. 2, pp. 27–31, 2020. https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr202002iw1.html

[6] IOWN GF, Members, https://iowngf.org/members/
[7] Press release issued by NTT, "NTT Group's New Environment and Energy Vision 'NTT Green Innovation toward 2040,'" Sept. 28, 2021. https://group.ntt/en/newsrelease/2021/09/28/210928a.html

Masahisa Kawashima

Head of IOWN Development Office R&D Planning Department, NTT. Chair of the IOWN GF Technology Working Group.

He received a Ph.D. in electrical engineering from Waseda University, Tokyo, in 1994 and M.Sc. in technology management from MIT's Sloan School of Management, USA, in 2002. He is currently leading NTT's research and development (R&D) on IOWN as the head of IOWN Development Office. He is also serving as the chair of the Technology Working Group at IOWN Global Forum. He has been working as a bridge between technologies and businesses since he joined NTT in 1994. His expertise includes high-speed networking, software-defined networking, cloud/edge computing, AI, and data management. He is currently exploring innovative means of implementing digital-twin applications by leveraging the evolution of optical communication and optoelectronics integration technologies.

Yosuke Aragane

Vice President, IOWN Development Office R&D Planning Department, NTT. Alternate Director and Leader of Recruitment Taskforce, IOWN Global Forum.

He received an M.S. and Ph.D. from Tokyo Institute of Technology in 1997 and 2005. Since joining NTT in 1997, he has researched in the field of human factors in cyber security. He was a manager of NTT Secure Platform Laboratories to develop security R&D strategies and of NTT-CERT (Computer Security Incident Response and Readiness Coordination Team), the representative computer security incident response team (CSIRT) of NTT Group to establish global CSIRT service teams. Since 2016, he has been the chief security producer in NTT R&D planning department of NTT headquarters office, where he has been promoting security-related technologies and business across NTT Groups. Since 2019, he has been the vice president of IOWN Development Office to lead the establishment and operation of IOWN GF.

Study on Open All-Photonic Network in IOWN Global Forum

Hideki Nishizawa, Jun-ichi Kani, Takafumi Hamano, Koichi Takasugi, Tomoaki Yoshida, and Seisho Yasukawa

Abstract

The IOWN Global Forum is developing the Open All-Photonic Network (Open APN) to provide direct optical paths between any location on demand to implement use cases of the Innovative Optical and Wireless Network (IOWN). This article describes the Open APN design goals and technical challenges as well as its functional architecture, with reference to the technical document "Open All-Photonic Network Functional Architecture" published in early 2022.

Keywords: IOWN, Open All-Photonic Network, functional architecture

1. Introduction

The IOWN Global Forum (IOWN GF) has released documents for two use cases, Cyber-Physical Systems (CPS) and AI-Integrated Communications (AIC). These use cases have already started to be implemented worldwide at some level. However, the evolution of sensing/capture technologies suggests that the critical requirements of these use cases will be much higher than those achievable with current technologies. IOWN GF is developing the Open All-Photonic Network (Open APN) [1] to establish an open architecture for photonic networking so that service providers can integrate photonic network functions with their entire computing and networking infrastructure with more granularity.

2. Design goals of Open APN

Open APN is being developed for achieving the following concepts:

- **End-to-end lambda connection**: Users will have transceivers that connect directly to remote sites via a service provider's network (wavelength tunnels in **Fig. 1**).
- **Dynamic optical-path provisioning/control**: To provide optical transport services that directly connect users flexibly, a method of provisioning and controlling optical paths is required.
- **Energy efficiency**: Open APN based on an end-to-end direct optical connection will enable networking with less energy consumption by minimizing electrical processing. The architecture and specifications of Open APN should be defined in such a way that lower power consumption can be properly achieved in accordance with the policy.
- **Multi-operator's environment**: The network will be an environment that accommodates multiple federated network operators. Each network operator can deploy end-to-end lambda connections seamlessly without being annoyed by complicated resource-sharing procedures and any conflicts when isolating defects.
- **Computing-networking convergence**: To implement CPS and AIC use cases, new optical networking that is easily adaptable to distributed computing is needed. It will connect computing resources in distant locations with high-capacity optical paths on demand with target quality of transport definable by computers.
- **Automated resource reallocation**: The network will need to efficiently scale bandwidth up and down per endpoint as user demand shifts over time.
- **Format-free optical communication**: Open APN should allow a variety of optical modulation

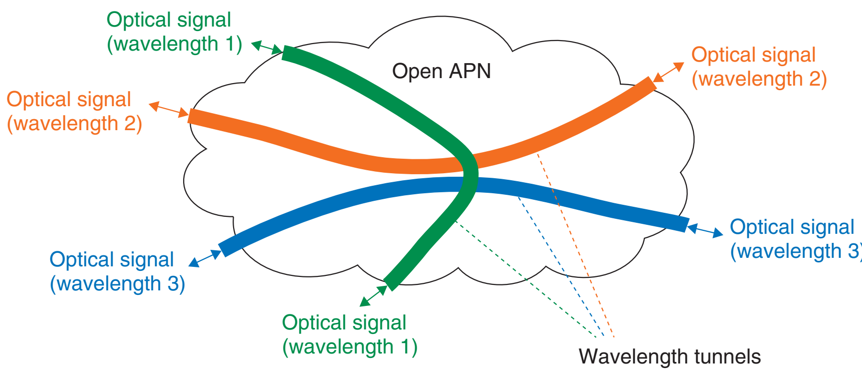


Fig. 1. Image of wavelength tunnels in the Open APN.

formats and upper-layer protocols. This will enable users to create new use cases with fiber infrastructures.

- **Intelligent monitoring**: To achieve more dynamic operations of Open APN, the network control-and-management systems must obtain sufficient information from Open APN devices. These systems should be low latency and highly secure for data collecting, storing, processing, analyzing, and sharing.

3. Evolution of optical transport technologies

Since the commercialization of digital coherent transmission systems started around 2010, the downsizing, power saving, and control-interface commonality of transmission systems have accelerated. In 2016, with the launch of Open ROADM [2], which defines interfaces and specifications to make reconfigurable optical add/drop multiplexer (ROADM) systems interoperable among vendors, and the Telecom Infra Project Open Optical & Packet Transport (TIP OOPT) [3], which defines open technologies, architectures, and interfaces in optical and Internet protocol networking, the openness of optical transmission technologies has accelerated. The application of digital technology provides advantages such as (1) being simpler and more flexible for the transmission line and its design, (2) enabling the independent evolution of hardware and software, (3) enabling the real-time measurement of transmission quality without affecting transmission quality, and (4) enabling the rapid estimation of transmission-line characteristics that determine transmission distance and capacity by using a Gaussian noise model. Over the next ten years, the convergence of computing and networking is expected to accelerate with the advent of co-packaged optics.

In fourth-generation (4G)/5G mobile networks, the bitrate of the mobile front haul (MFH) link per antenna has been enhanced to 10 or 25 Gbit/s, and there is also much discussion about standardization for applying wavelength division multiplexing (WDM) technologies to the MFH link [4, 5].

For enhancing fixed broadband access, Next-Generation Passive Optical Network Stage 2 (NG-PON2), which combines traditional time division multiplexing with dense WDM (DWDM) technology with 4 to 8 wavelengths, has been standardized and commercialized [6, 7]. Super-PON is under standardization to develop a long-distance (over 50-km) PON system through combining a 2.5G- to 10G-class PON system with DWDM technology (16 wavelengths or more) [8]. It is expected that common WDM networking will efficiently accommodate traffic in a metro-access area.

4. Technical challenges to implement Open APN

To implement Open APN, it is necessary to accelerate the further development of the technological trends described in the previous section and address the technical challenges in **Table 1**. The Open APN function architecture is defined with these solutions in mind, and the high-level reference architecture, control-and-management-plane architecture, and user-plane architecture are described separately.

5. Functional architecture of Open APN

5.1 High-level reference architecture

Figure 2 shows a high-level reference architecture

Table 1. Technical challenges to implement Open APN.

Items	Challenges
Direct optical paths between any locations	• Provisioning and management of the wavelength resources throughout the network, i.e., user, access, metro, and core networks
Dynamic online optical-path design	• Automatic selection of the optimum transmission mode of a coherent module based on the quality of the fiber-link system • Standardization of control signals to interconnect modules of different types and vendors • Fast route planning of an optical path that satisfies the reliability requirements • Support of the networking environment of multiple network operators and administrative domains, and/or devices from multiple vendors
Network attachment mechanisms for user-owned transceivers	• Admission control functions (authentication and filtering) and resolution of optical-path wavelength and endpoint addresses for a user-owned transceiver
Node architecture to support direct optical paths and dynamic path setup	• Direct optical connections between transceivers under the same ROADM • Remote deployment of transceivers and the support of their control and management channel in the conventional ROADM node • Advanced control of optical amplifier chains to support wavelength reconfiguration that will occur more frequently
Real-time performance monitoring	• Faster response within seconds or even milliseconds • Granular monitoring parameters for multiple domains of multiple operators • High compatibility with the current mechanism without affecting the transmission quality of the service traffic
Secure transport mechanisms for the optical paths originating from user premises	• Secure transport, such as Authentication, Authorization, and Accounting (AAA) without additional latency in connections between terminals installed in the customer's environment

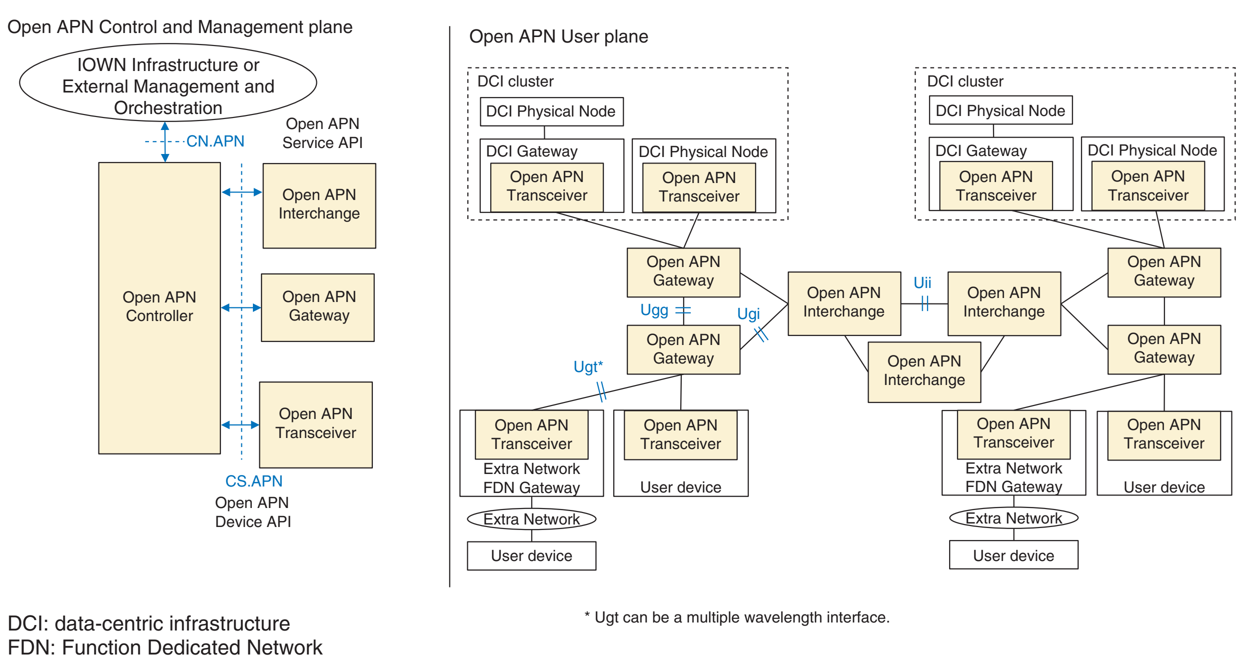


Fig. 2. Open APN high-level reference architecture.

of Open APN. The Open APN User plane consists of the Open APN Transceiver (APN-T), Open APN Gateway (APN-G), and Open APN Interchange (APN-I). The Open APN Control and Management plane consists of the Open APN Controller (APN-C) that communicates with the APN-T, APN-G, and

APN-I.

The APN-T is an endpoint for an optical path and transmits and receives optical signals. The APN-G is a gateway for an optical path and has the following five functions: (1) provision of control channels to communicate with the connected APN-Ts, (2) admission control in the user plane, (3) multiplexing/demultiplexing, (4) turn back, and (5) add/drop. The APN-I is an interchange for wavelength switching and has the following two functions; (1) wavelength cross-connect and (2) adaptation between the interfaces.

Compared with the current standard ROADM, the APN-G's functions (1), (2), and (4) and APN-I's function (2) can be regarded as the functions characteristic to Open APN.

5.2 Control-and-management-plane reference architecture

To implement the control-and-management plane of Open APN, as shown in Fig. 2, an Open APN Service application programming interface (API) is defined between the APN-C and external management-and-orchestration systems, and an Open APN Device API is defined between the APN-C and Open APN devices (APN-T, APN-G, and APN-I).

After authentication and activation of the APN-T, the APN-C creates, deletes, or resets an optical path between the APN-Ts on demand triggered by path-setup requests. The APN-C calculates the wavelength path route, wavelength to use, transmission/reception parameters for optical transmission devices, and configuration of optical transmission devices accordingly to satisfy user requirements regarding bandwidth, latency, and jitter.

The APN-C also monitors quality of transmission (QoT) information through the Open APN Device API in real time and determines whether user requirements about QoT are satisfied for each wavelength path. An example configuration of telemetry for the APN-C and Open APN devices and an example controller configuration for coordinating with the mobile system are shown in the document "Open All-Photonic Network Functional Architecture" [1].

5.3 User-plane reference architecture

The concept of Group of Optically Interconnectable Port (GOIP) has been introduced to implement scalable and interoperable Open APNs under physical constraints such as a limited number of wavelengths and transmission distance.

GOIP is defined as a group of optical ports for which a direct optical connection (Open APN optical path) can be established. The port means a connection interface between an APN-T and access link. The connection can be point-to-point and point-to-multipoint. **Figure 3** shows a configuration image of GOIP.

Because the total distance of the connection depends on bitrate and/or modulation methods, these methods of the optical transmission/reception between ports in GOIP are presented for each GOIP individually. In GOIP, there is at least one route that can establish a direct optical connection between ports. However, a direct optical connection may not be established when the shortest route is not available due to reasons such as shortage of wavelength resources, fiber cut, or equipment failure. GOIP design methods given these considerations are for further study.

In the Open APN architecture, wavelength tunnels should be managed appropriately in accordance with the specified bit rate and transmission distance, so a physical-layer reference architecture is defined and the physical reference of wavelength tunnels is described (**Table 2**).

6. Future outlooks

In IOWN GF, Open APN will evolve with a spiral approach that repeats the development of its specifications and demonstration. The Open APN functional architecture described in this article will be used as a base for proof-of-concept demonstration. The insights gained through the demonstration will be reflected in the future development of Open APN specifications and improvement in the Open APN functional architecture.

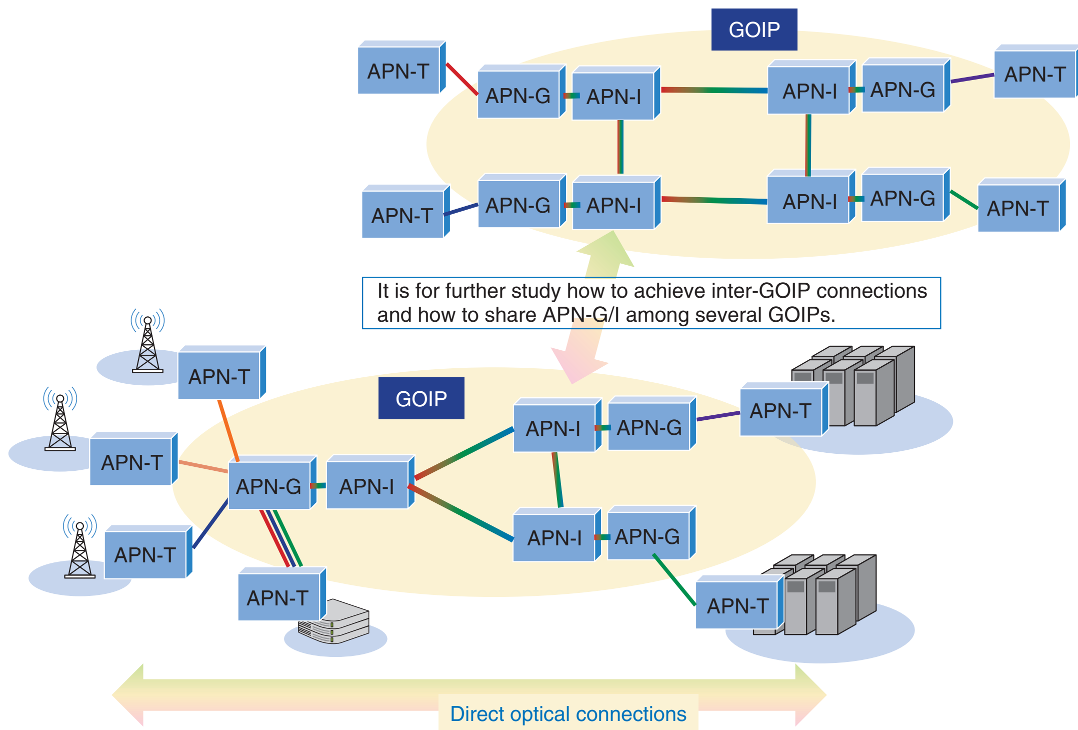


Fig. 3. Schematic diagram of GOIP.

Table 2. Physical reference of wavelength tunnels.

Category	Reference information
Requirements on the GOIP connection from the network	Required bandwidth
	Locations of the connected APN-T
Specifications of the connected APN-T connected by the wavelength tunnels	Information relative to the compatibility of APN-T for connection through wavelength tunnels (e.g., modulation format, symbol rate, used FEC, etc.)
	A bitrate of the APN-T connection
	Specification for the error-free transmission of APNT-T including tolerance to noise and impairments (e.g., OSNR tolerance or GSNR tolerance, tolerance to chromatic dispersion)
The physical reference of the wavelength tunnels	Physical information of the wavelength tunnels (e.g., used fibers, wavelengths, used bandwidth)

FEC: forward error correction
GSNR: generalized signal-to-noise ratio
OSNR: optical signal-to-noise ratio

References

[1] IOWN GF, "Open All-Photonic Network Functional Architecture," https://iowngf.org/technology/

[2] Open ROADM, http://openroadm.org/

[3] TIP OOPT, https://telecominfraproject.com/oopt/

[4] Mobile Optical Pluggable, https://www.ericsson.com/en/mobile-transport/mopa

[5] ITU-T Work Program, Recommendation ITU-T G.9802.1: Wavelength division multiplexed passive optical networks (WDM PON): General requirements, https://www.itu.int/itu-t/workprog/wp_item.aspx?isn=16803

[6] Recommendation ITU-T G.989: 40-Gigabit-capable passive optical networks (NG-PON2): Definitions, abbreviations and acronyms, https://www.itu.int/rec/T-REC-G.989/

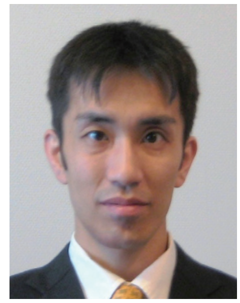
[7] C. Wilson (ed.), "Verizon, Calix Roll Out Commercial NG-PON 2," Light Reading, 2018. https://www.lightreading.com/gigabit/fttx/verizon-calix-roll-out-commercial-ng-pon-2/d/d-id/740093

[8] IEEE P802.3cs Increased-reach Ethernet optical subscriber access (Super-PON) Task Force, https://www.ieee802.org/3/cs/

Hideki Nishizawa
Senior Research Engineer, Supervisor, Frontier Communication Laboratory, NTT Network Innovation Laboratories.
He received an M.S. in physics from Chiba University in 1996 and joined NTT the same year, where he has been engaged in research on ubiquitous networking, photonic switching systems, and photonic transmission systems. He is a member of Institute of Electronics, Information and Communication Engineers (IEICE).

Jun-ichi Kani
Senior Distinguished Researcher, Group Leader, Optical Access Systems Project, NTT Access Network Service Systems Laboratories.
He received a B.E., M.E., and Ph.D. in applied physics from Waseda University, Tokyo, in 1994, 1996, and 2005. He joined NTT Optical Network Systems Laboratories in 1996, where he researched optical multiplexing and transmission technologies. He has been with NTT Access Network Service Systems Laboratories since 2003, where he is engaged in research and development (R&D) of optical communications systems for metropolitan and access network applications and currently heads the Access Systems Technology Group. He has been participating in the International Telecommunication Union - Telecommunication Standardization Sector (ITU-T) and the Full Service Access Network (FSAN) initiative since 2003.

Takafumi Hamano
Senior Research Engineer, Supervisor, Network Architecture Project, NTT Network Service Systems Laboratories.
He received a B.E. and M.E. in engineering from Kyoto University in 1994 and 1996. He joined NTT in 1996 and moved to NTT Network Service Systems Laboratories in 2002. His research interests include next-generation transport network architecture and systems. He is a member of IEICE.



Koichi Takasugi
Executive Research Engineer, Director of Frontier Communication Laboratory, NTT Network Innovation Laboratories.
He received a B.E. in computer science from Tokyo Institute of Technology, M.E. from Japan Advanced Institute of Science and Technology, and Ph.D. in engineering from Waseda University in 1995, 1998, and 2004. He was involved in the design and standardization of the Next-Generation Network (NGN) architecture. He has implemented and installed super high-density Wi-Fi systems in several football stadiums. He was also active in the artificial-intelligence field, such as diagnosing diabetes through machine learning. He is currently leading research on the network architecture and protocols in optical and wireless transport networks.

Tomoaki Yoshida
Project Manager, Optical Access Systems Project, NTT Access Network Service Systems Laboratories.
He received a B.E., M.E., and Ph.D. in communication engineering from Osaka University in 1996, 1998, and 2007. In 1998, he joined NTT Multimedia Systems Development Center. In 1999, he moved to NTT Access Network Service Systems Laboratories and has been engaged in research on next-generation optical access networks and systems. From 2013 to 2015, he was involved in the research of WDM/time division multiplexing-PON and worked on the standardization of optical access systems such as NG-PON2. He is a member of the Institute of Electrical and Electronics Engineers (IEEE) and IEICE.

Seisho Yasukawa
Project Manager, Network Architecture Project, NTT Network Service Systems Laboratories.
He received a B.E. and M.E. in applied physics from the University of Tokyo in 1993 and 1995. Since joining NTT in 1995, he has conducted R&D of asynchronous transfer-mode-based multimedia switching and NGN architecture. He has also been involved in standardization efforts for point-to-multipoint multiprotocol label switching. His current research includes the IOWN architecture.

Data-centric-infrastructure Functional Architecture

Hitoshi Masutani, Christoph Schumacher, and Koichi Takasugi

Abstract

Society is changing at a rapid pace due to technical innovations in Internet technologies and cloud computing. In the near future, ever-more refined artificial-intelligence applications are expected, working on ever-larger data sets. Implementing such applications requires increasing the computational capacity available to the providers of such newly emerging services. However, energy consumption of such services must be reduced to achieve ESG (environmental, social, and governance) and the sustainable development goals. To fulfill both of these objectives at the same time, a drastic rethink of computing infrastructure is required. With this in mind, the IOWN Global Forum is creating the *data-centric-infrastructure (DCI) functional architecture*. This article gives an overview of the first DCI functional architecture reference document about this next-generation computing foundation.

Keywords: IOWN, data-centric, DCI

1. The data-centric-infrastructure computing platform

The IOWN Global Forum (IOWN GF) is specifying a holistic architecture that comprises both networking and computing, as illustrated in **Fig. 1** and published the first data-centric-infrastructure (DCI) functional architecture reference document in January 2022 [1]. Within this overall architecture, the DCI subsystem provides computing and networking services to various types of applications. The DCI is specified with the assumption that a high-performance optical network, the Open All-Photonic Network (Open APN) [2], is available for mid- and long-range data transfers. The goal with the DCI is to provide a quality of service (QoS)-managed execution environment for applications making use of multiple types of computing resources, such as central processing units (CPUs), memory, field-programmable gate arrays (FPGAs), and graphics processing units (GPUs), that are placed in a distributed fashion. These computing resources that the DCI will include comprise both generic computation elements and accelerators for artificial intelligence (AI) processing. The DCI makes computing resources available to applications via application programming interfaces (APIs) accommodating the needs of service providers.

2. Technology gaps in current computing-platform architectures

IOWN GF conducted a gap analysis to identify the technological innovations required for future use cases. The following gaps that need to be addressed were identified.

2.1 Scalability issues

Different applications have different requirements regarding computing resources, memory, and input/output mechanisms. To appropriately serve a wide variety of applications, various computing resources meeting the requirements of each application need to be allocated without waste. However, classic, rack-oriented computing platforms are generally not considered able to efficiently combine resources located in different servers.

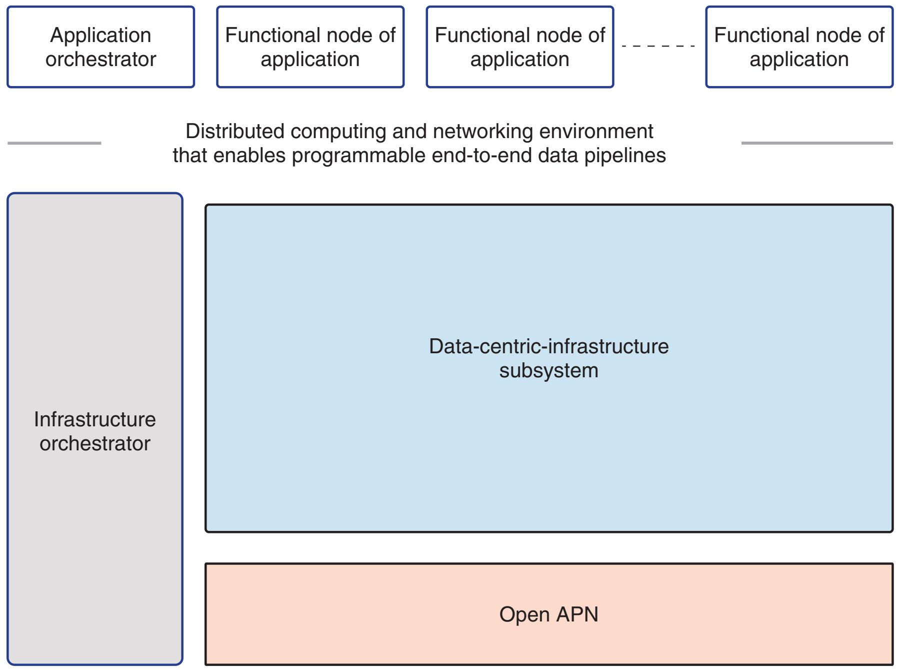


Fig. 1. Overview of the entire IOWN Global Forum architecture.

2.2 Performance issues

Different types of applications have different requirements. Some applications have stringent requirements toward latency and jitter. Classic servers are typically connected using networks designed to provide best-effort services. Such networks do not satisfy the requirements of many future applications. Therefore, the stringent requirements of future applications need to be taken into account from the beginning when designing future computing and network architectures.

2.3 Energy consumption issues

This issue is related to the scalability issue above. Multiple types of computing resources must be scheduled to execute work without idle time since idle resources will only consume energy without executing useful functions. Therefore, IOWN GF is developing an architecture that maximizes resource utilization while using energy more effectively.

3. Design goals

On the basis of the gap analysis outlined above, the following design goals were identified for the DCI:

(1) Provide scalability in an environment ranging from user devices over edge networks and clouds to other clouds.
(2) Enable the use of computing resources other than CPUs.
(3) Implement data pipelines on the basis of a high-speed optical network to enable efficient data transfers between applications of a use case.
(4) Enable sharing among accelerators, such as GPUs and FPGAs, without having to execute redundant data transfers.
(5) Simultaneously support different QoS objectives such as high bandwidth, bandwidth reservation, low latency, and low jitter.
(6) Provide a gateway function to exchange data between the classic Internet protocol (IP)-based network domain and non-IP-based network domain.

The DCI is being designed by taking into account these six goals. The relations between the DCI subsystem, Open APN, and pre-existing networks, as the resulting architecture, are illustrated in **Fig. 2**. End points, such as mobile network radio units or remote sensors, may connect to DCI clusters either directly via Open APN or indirectly using a network outside Open APN (non-Open APN extra network) as an intermediate step.

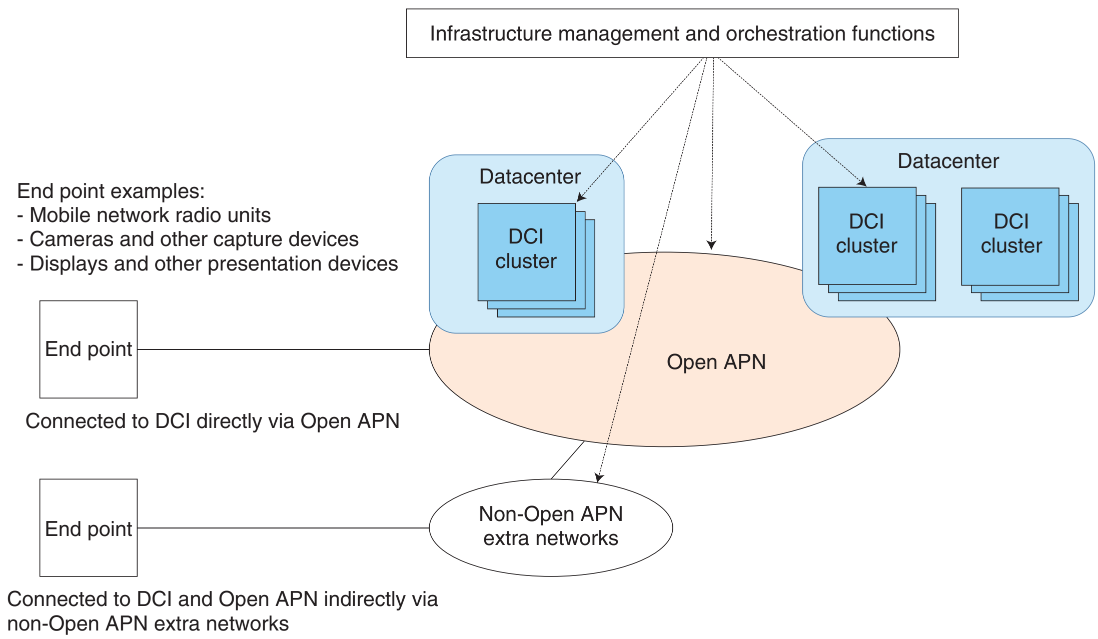


Fig. 2. Overview of Open APN and DCI clusters.

4. DCI cluster

The DCI architecture defines logical service nodes (LSNs). Following the DCI design goals, LSNs are a means to provide execution environments comprising, e.g., computing and networking resources to applications in appropriate units. LSNs provide users with resources that are logically separated at the hardware level. To provide such LSNs, multiple computing and networking resources must be appropriately selected and combined. The DCI architecture defines that DCI clusters are responsible for creating LSNs from the resources in their DCI cluster resource pools.

4.1 Structural elements of DCI clusters

The structural elements of DCI clusters are DCI physical nodes, the inter-node interconnect, and DCI gateway, as illustrated in **Fig. 3**.

(1) DCI physical node

The DCI physical node is the basic unit of computing nodes. The architecture of this node is designed to not only provide functionality of classical server mainboards but also provide access to many other computing resources such as FPGAs and GPUs.

The intra-node interconnect is meant to enable communication between these computing resources. It is used to share common data among various computing resources. When updating such shared data, synchronization is required. Therefore, in addition to the classic PCI (Peripheral Component Interconnect) express bus, the DCI functional architecture (FA) reference document mentions cache-coherent*1 interconnects, such as Compute Express Link (CXL), as alternatives for future designs.

(2) Inter-node interconnect

The inter-node interconnect corresponds to the top-of-rack switch of classical system designs. The DCI FA reference document mentions that this interconnect can support various QoS levels and that further details will be provided in future revisions.

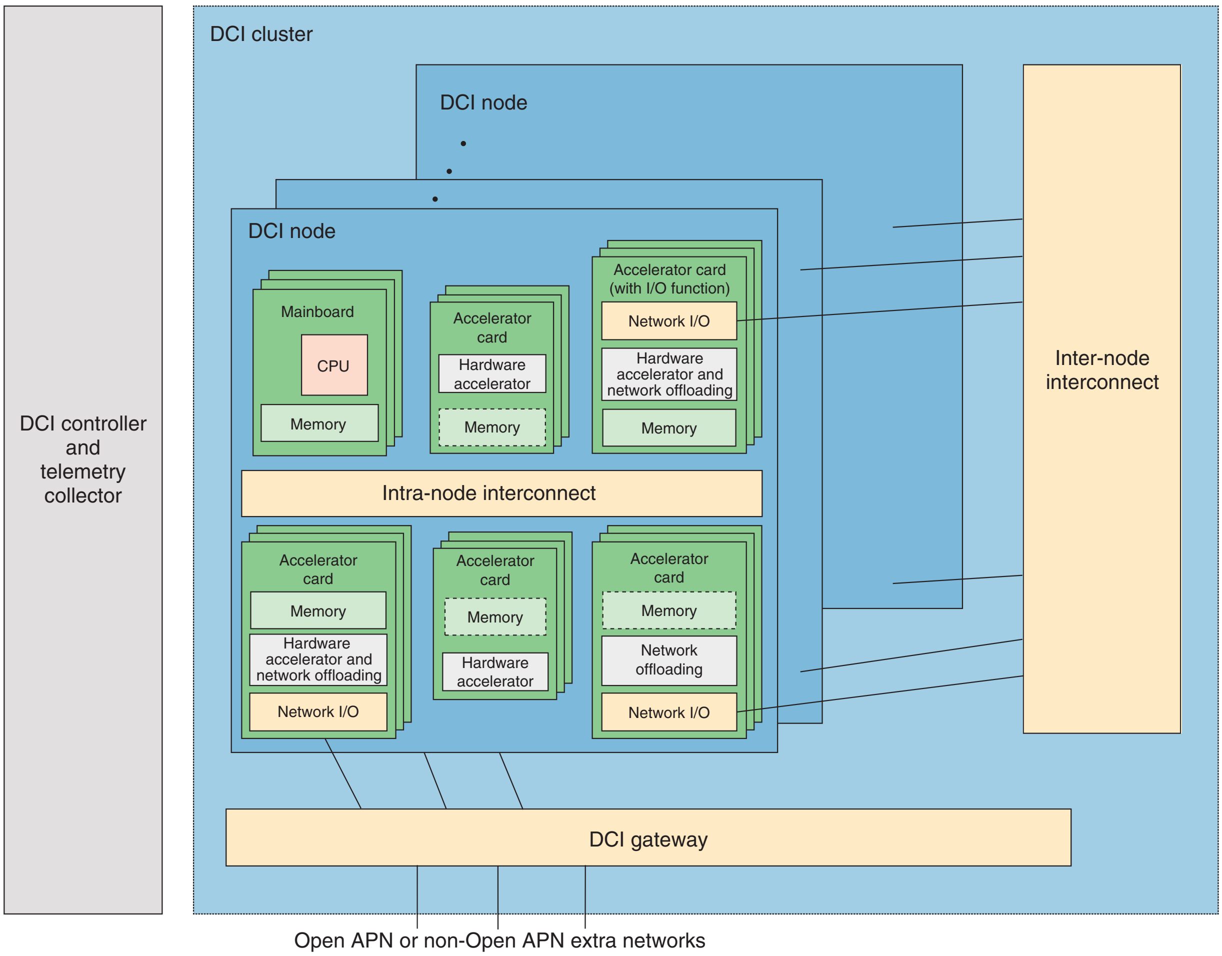
(3) DCI gateway

The DCI gateway connects the DCI cluster to Open APN. Like the inter-node interconnect, the DCI gateway must support various QoS levels, and further details are to be provided in future revisions of the DCI FA reference document.

4.2 DCI cluster controller

A DCI cluster controller manages the elements composing a DCI cluster. The DCI cluster controller receives requests from orchestrators, and on the basis

*1 Cache coherency: Cache coherency is a mechanism that allows multiple clients that are reading from and writing to memory or other shared resources to keep a consistent view on the data in their cache memories and the data in the main memory.



I/O: input/output

Fig. 3. Example of the DCI cluster architecture.

of these requests, conducts life-cycle management of LSNs, i.e., the DCI cluster controller is responsible for configuring, starting, and stopping LSNs. The DCI cluster controller is located outside the DCI cluster. There is no limit specified on how many DCI clusters a single DCI cluster controller may control.

5. DCI infrastructure as a service

The DCI infrastructure as a service (DCI IaaS) allows service providers to benefit from LSNs, as described by the DCI architecture, from optical networks as detailed by the Open APN architecture, and Function Dedicated Networks (FDNs)*2 described later, without having to maintain their own infrastructure. The relations between the DCI infrastructure provider, tenant platform-service providers, and end-user applications using such platforms are illustrated in **Fig. 4**. Tenant platform-service providers may request provisioning of LSNs and networks in-between these LSNs from DCI-infrastructure service providers. Then, tenant platform-service providers deploy middleware or applications on the LSNs and networks. Finally, these resources are offered as a service to end users and their applications. The APIs that tenant platform-service providers use to access DCI IaaS services are then defined. For example, regarding LSNs, which are the main concept of DCI, an API is defined to support actions such as creation, configuration, starting, and stopping.

*2 FDN: FDN is a concept defined by IOWN GF. The DCI FA reference document defines FDNs as logical networks created on top of physical networks. Such physical networks can be Open APN or other types of physical networks.

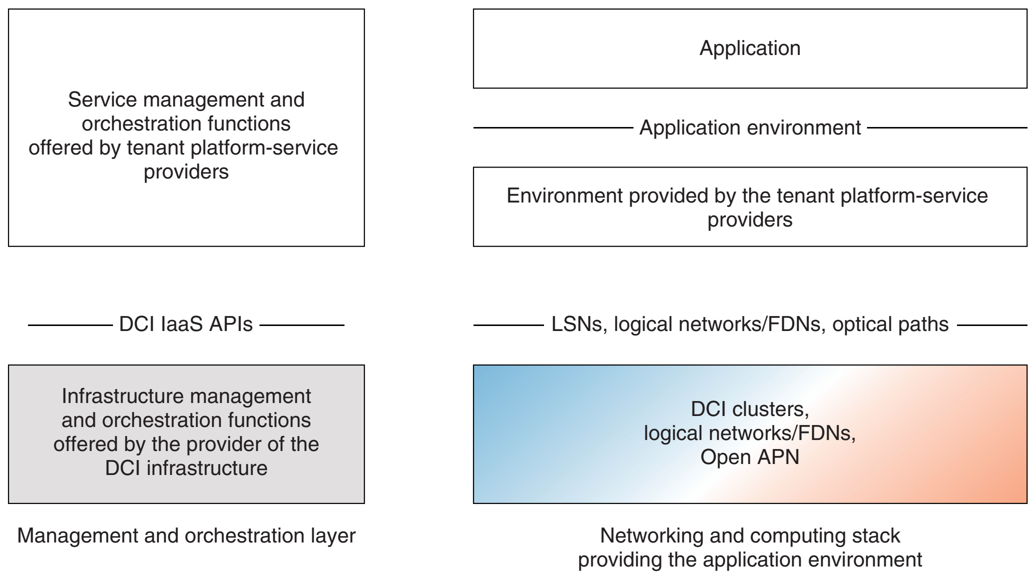


Fig. 4. Service model of DCI IaaS.

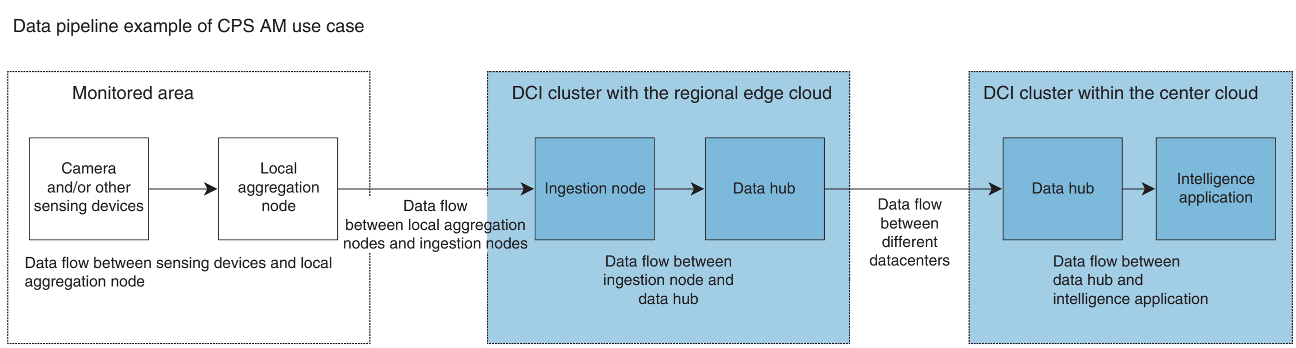


Fig. 5. CPS AM data pipeline example.

6. Analysis of the data plane to implement IOWN GF use cases

When analyzing various use cases, their parts can be classified into data flows and processing components. Therefore, use cases can be expressed as pipelines connecting these elements. One of the first use cases that IOWN GF is targeting is the cyber-physical systems (CPS) area management (AM) use case. A reference implementation model (RIM) as well as a data pipeline describing this use case are detailed in the IOWN GF RIM document published in January 2022 [3]. For example, this use case includes the scenario of first gathering the video streams of groups of 1000 cameras installed in a given area using local aggregation nodes, transferring these large amounts of real-time data to a regional edge cloud, then conducting continuous AI analysis, and finally alerting local security staff of events within the monitored area. The resulting data pipeline is illustrated in **Fig. 5**.

Each dataflow has different requirements toward forwarding. Furthermore, data will need to traverse different data planes. Therefore, the data planes that need to be accelerated need to be identified on the basis of this classification. For example, the intra-node interconnect may be used for communication within a given DCI physical node, and for DCI physical nodes located in different DCI clusters, a data plane using the DCI gateway and Open APN can be used. The DCI FA reference document classifies

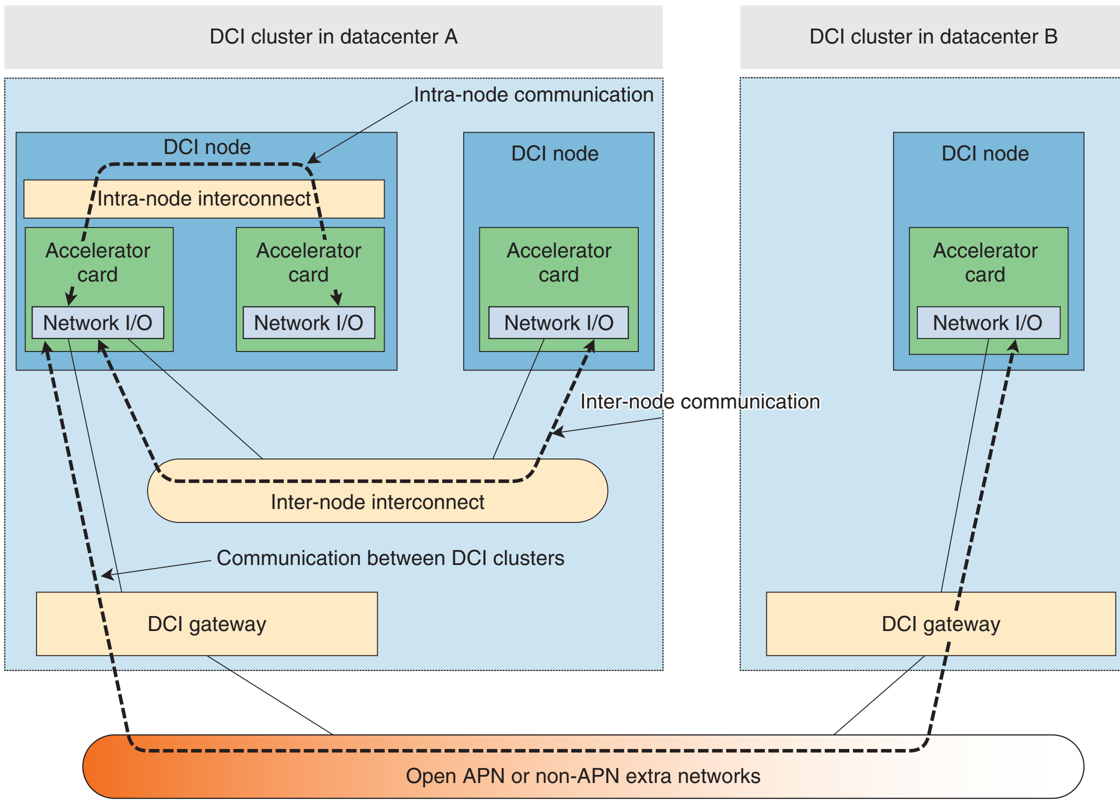


Fig. 6. The three data-plane types of the DCI.

these data-plane communication patterns, as highlighted in **Fig. 6**. The data pipeline shown in Fig. 5 is analyzed on the basis of this classification. From this analysis, the DCI FA reference document proposes a data-plane framework for those data-plane patterns that are difficult to support with current technology. The following section gives an outline of this framework.

7. Data forwarding acceleration by remote direct memory access

Remote direct memory access (RDMA) has been proposed as a method of accelerating data-plane long-range data transfers between computing resources that are distributed over a wide area. RDMA was originally developed for communication over relatively short distances up to around 10 meters within datacenters, targeting the communication of high-performance computing applications. However, since the resources of remote datacenters are also used within the overall system in an integrated fashion, RDMA adapted to long-range data transfers becomes necessary.

The data-plane framework proposed in the DCI FA reference document uses the widely used RDMA reliable connection (RC) as a transport mode. In this mode, an RDMA framework guarantees the completeness of data to provide a highly reliable connection service. The DCI FA reference document highlights the following points regarding how to avoid performance and reliability degradation when using the RDMA RC between locations separated by long distances.

7.1 Queue-depth optimization

With RDMA, send and receive operations are controlled by first creating send and receive requests in the form of work queue elements (WQEs). When sending data, a corresponding WQE is placed into a send queue to start the transfer. The amount of messages that can be sent without having to wait for an acknowledgment message from the receiver side can be increased by increasing the maximum length of the send queue of the RDMA network interface card (NIC). Therefore, long-range transmissions can maintain high throughput even when the round-trip time (RTT) is long. The DCI FA reference document gives the following formula to optimize the queue length for long-range RDMA transfers depending on the RTT:

$$\frac{\text{RTT} * \text{LineSpeed}}{\text{MessageSize}} = \text{Required QueueDepth}$$

7.2 Increasing the efficiency of data forwarding between RDMA-capable NICs and accelerator cards

To increase forwarding efficiency, the DCI FA reference document describes the possibility to reduce the number of times that data are copied to minimum by directly exchanging data between RDMA-capable NICs and accelerators and avoiding temporarily storing data in memory buffers as much as possible.

7.3 Reliability when adapting RDMA for long-range communication

By using a re-send mechanism, the RDMA RC can guarantee the completeness of data transfers even when faced with packet loss. However, especially in the case of long-range communication, re-sending lost data with the RDMA RC requires the RTT to complete, degrading throughput. To avoid such performance issues, the DCI FA reference document suggests that the underlying Open APN network layer should provide a function to provide high-quality optical paths to reduce the number of transmission errors and/or use more efficient re-transmission algorithms to mitigate the impact of errors in long-range transmissions.

7.4 Providing QoS in concert with the DCI control and management plane

Within a DCI cluster, multiple data flows exist. The DCI FA reference document states that classic cloud computing does not allow reserving network resources in advance, and with only the congestion-control algorithms in network appliances and servers, the required QoS cannot often be achieved. Therefore, the DCI FA reference document suggests to let services use resource reservation with different QoS classes.

8. Summary

The DCI FA reference document outlines the design goals that the computing architecture for the IOWN GF use cases must achieve. The key points are the definitions of the DCI cluster as well as the DCI IaaS service model. Furthermore, CPS AM is analyzed as an exemplary use case, and for transmission between physically separated DCI clusters, a data plane using long-range RDMA is proposed for acceleration. In the future, IOWN GF will proceed with experiments to verify the architectural and technological concepts as highlighted above, and the results as well as remaining issues will then be used to drive further innovation of IOWN GF technologies.

References

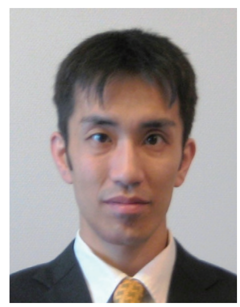
[1] IOWN GF, "Data-Centric Infrastructure Functional Architecture," Jan. 2022.
https://iowngf.org/wp-content/uploads/formidable/21/IOWN-GF-RD-DCI-Functional-Architecture-1.0-1.pdf
[2] IOWN GF, "Open All-Photonic Network Functional Architecture," Ver. 1.0, Jan. 2022.
https://iowngf.org/wp-content/uploads/formidable/21/IOWN-GF-RD-Open-APN-Functional-Architecture-1.0-1.pdf
[3] IOWN GF, "Reference Implementation Model (RIM) for the Area Management Security Use Case," Ver. 1.0, Jan. 2022.
https://iowngf.org/wp-content/uploads/formidable/21/IOWN-GF-RD-RIM-for-AM-Use-Case-1.0.pdf

Hitoshi Masutani
Senior Research Engineer, Network Service Platform Research Group, Core Network Technology Research Project, NTT Network Service Systems Laboratories.

He received a B.E. in communication engineering and M.E. in electrical, electronic and information engineering from Osaka University in 1999 and 2001. After joining NTT Network Innovation Laboratories in 2001, he studied multicast networking and Session-Initiation-Protocol-based home networking. In 2005, He moved to the Visual Communication Division of NTT-Bizlink, where he was responsible for developing and introducing visual communication services, including an IP-based high-quality large-scale video conferencing system and real-time content delivery system on IPv6 multicast. He also worked on developing their service order management system and network management system for video conferencing services. Since returning to NTT Network Innovation Laboratories in 2012, he has been engaged in the research and development (R&D) of programmable network nodes, including software-defined networking and network function virtualization, e.g., the high-performance software openflow switch called Lagopus. He is currently engaged in R&D of deterministic communication services' technologies.

Christoph Schumacher
Senior Research Engineer, Computing Systems Group, System Software Project, NTT Software Innovation Center.

He received a doctoral degree in engineering from RWTH Aachen University, Germany, in 2015. His current interests include emerging use cases with new types of requirements toward data-centers and future computing infrastructure.



Koichi Takasugi
Executive Research Engineer, Director of Frontier Communication Laboratory, NTT Network Innovation Laboratories.

He received a B.E. in computer science from Tokyo Institute of Technology in 1995, M.E. from Japan Advanced Institute of Science and Technology in 1997, and Ph.D. in engineering from Waseda University in 2004. He was involved in the design and standardization of the Next-Generation Network architecture. He was also active in the AI field, such as diagnosing diabetes through machine learning. He is currently leading research on the network architecture and protocols in optical and wireless transport networks.

Study of Storage Services at IOWN Global Forum

Tomohiro Inoue

Abstract

The IOWN Data Hub is a new data service that integrates database, storage, and other functions required in the Innovative Optical and Wireless Network (IOWN) era. The IOWN Data Hub is intended to be commonly used by other applications by leveraging the advanced IOWN infrastructure while maintaining compatibility with existing applications. This article introduces the status and the released document of the IOWN Data Hub discussed in the IOWN Global Forum.

Keywords: data hub, data-centric, storage

1. IOWN Data Hub

The IOWN Global Forum (IOWN GF) is a forum where participating companies collaborate to discuss common service-layer mechanisms that can leverage the new advanced communication and computing infrastructure created by the Innovative Optical and Wireless Network (IOWN). The IOWN Data Hub (IDH) is a data-hub service (integrated database, storage, and data-management services) for applications in the IOWN era. It is intended to be a data management and sharing infrastructure that enables fast and reliable data processing, usage, and exchange among multiple parties.

The IDH is an application functional node on top of the IOWN Data-Centric Infrastructure (DCI) architecture [1] and intended to be commonly used by other application functional nodes by maintaining compatibility with existing applications while leveraging the features of the IOWN Open All-Photonic Network (Open APN) [2] and DCI. IOWN GF defines a common infrastructure so that multiple service providers can develop and deploy IDH services.

In January 2022, the “Data Hub Functional Architecture” document (IDH document) [3] was released with the cooperation of forum members, including NTT, Oracle, and NEC. The document defines multiple models of the IDH. In the various use cases discussed in IOWN GF, there are numerous demand patterns expected of data hub services. IOWN GF typified these demands and organized them into several Service Types and Service Classes. The document also presents a reference implementation model for each Service Class using the IOWN infrastructure.

2. IDH Service Type and Service Class

2.1 Required data-hub services

IOWN GF has been investigating the system architecture and data flows for several use cases [4] towards full-stack optimization using the IOWN infrastructure, regardless of the layer structure of the existing system, which is a unique approach of IOWN GF that covers a wide range from communication layers to service layers. Nevertheless, data flows resulting from the analysis are conceptually close to those required by current Internet applications. It is assumed that the data flow used by applications in the early stages of IOWN deployment will not differ significantly from existing applications for compatibility reasons. This section introduces the Service Type and Service Class defined on the basis of these discussions.

2.2 Service Types

A Service Type is defined as a unit of functionality and behavior that will be implemented by one or

Table 1. IDH Service Types.

Service Type	Example of data flow	Supported APIs
Distributed Relational Database (DRD)	Structured data to be registered in RDBMS, labeled data of detected objects (output of AI inference), etc.	JDBC and ODBC that support SQL. REST API that accepts SQL-based data queries and responds in JSON format.
Key-Value Store (KVS)	Time-series data, sensor data, etc.	REST API (CRUD operations on single or multiple Key Values, support for simple queries)
Graph Store	User-related data, links to related event data, semantic-knowledge data, etc.	REST API that accepts queries based on SPARQL, PGQL, and other domain-specific languages (DSLs), and responds in JSON format
Message Broker	Notification messages, voice messages, etc.	General message broker API (e.g., Apache Kafka compatible API)
Object Storage	Video data, sensor data (e.g., point cloud), etc.	General Object Storage API (e.g., REST API that is compatible with AWS S3)

* Currently, only general requirements are listed; more detailed requirements will be discussed in the future.

AWS: Amazon Web Services
CRUD: Create, Read, Update and Delete
JDBC: Java Database Connectivity
JSON: JavaScript Object Notation
ODBC: Open Database Connectivity
PGQL: an SQL-based query language for the property graph data model.
RDBMS: relational database management system
REST: Representational State Transfer
SPARQL: Simple Protocol and RDF (Resource Description Framework) Query Language
SQL: Structured Query Language

several IDH services. Specifically, requirements such as supported application programming interfaces (APIs), queries, and data types are defined, and external applications can treat IDH services like databases or object storage by specifying a Service Type. Five Service Types are currently defined (**Table 1**).

(1) Distributed Relational Database

Distributed Relational Database (DRD) is a Service Type for storing structured data intended for data-analysis use cases such as online transaction processing in general databases and online analytical processing in data warehouses. Using this Service Type, various applications will store and use data without executing any extra data-conversion processing outside the IDH. Specific data that some structural queries can identify should be handled by DRD as long as the processing speed permits.

(2) Key-Value Store

Key-Value Store (KVS) is a Service Type used when storing a vast amount of data, although individual records are not that large. For example, time-series data and sensor data fall under this category.

(3) Graph Store

This Service Type stores data that are modeled as a graph. The stored data may follow a pre-defined schema called an ontology. The ontology describes how the data are structured and enables application developers to form semantic queries. Semantic queries can retrieve information on the basis of associations or a range of contexts.

(4) Message Broker

This Service Type is intended for use in the following situations where data cannot simply be added to existing data storage.

- There are so many messages that they cannot be accommodated by other Service Types and must be stored in several queues.
- The number of providers and users of data is enormous and dynamically changing.
- The data hub handles data of mixed sizes, types, and formats and requires a complex pipeline for format conversion and aggregation.
- The data hub only needs to relay messages and does not require data persistence.

(5) Object Storage

This Service Type stores relatively large object data, such as semi-structured data, and unstructured data, such as videos, images, and logs, at low cost. It is expected to be used for the long-term storage of big data and provide data-analysis applications such as artificial intelligence (AI) and data lakes for future data utilization.

2.3 Service Class

To implement IOWN use cases, an IDH service implementing one Service Type is insufficient, and multiple Service Types may be required simultaneously. The IDH defines the Service Classes shown in **Table 2** as implementations of services with one or more Service Types.

The Basic Service Class inherits only one of each Service Type. The Applied Service Class provides a

Table 2. IDH Service Classes.

Basic Service Class: inherits only one corresponding Service Type		
	DRD / KVS / Graph Store / Message Broker / Object Storage	
Applied Service Class: inherits multiple service types and/or integrates multiple service classes		
	Converged Database	• DRD • KVS • Graph Store • Message Broker
	Context Broker	• Graph Store • Message Broker • DRD
	Virtual Data Lake	• Object Storage
	Virtual Data Lake House	• DRD • KVS • Graph Store • Message Broker • Object Storage

package of functionalities that more complex use cases collectively require from the IDH, and it inherits from and integrates multiple Service Types. The IDH document defines the following four Applied Service Classes.

(1) Converged Database

This Service Class is designed for cases in which different types of data are to be processed simultaneously. For example, let us consider a scenario in which we analyze people's behavior in a building and provide appropriate support. We need to manage and process interactions of people with people, and people with things all at once, and the data structure that represents this needs to be highly flexible. However, if different IDH services are used for each data structure and data conversion is carried out at the application level, the real-time performance required by IOWN cannot be achieved. Converged Database provides a platform for high-speed simultaneous processing of various data to meet these needs.

(2) Context Broker

This Service Class functions as a sophisticated broker that enables applications to request data by flexibly specifying the context. Context can be expressed as a thing (e.g., a specific building), type of thing, geographic area, temporal area, or combination of these, allowing the specification of information for particular conditions.

(3) Virtual Data Lake (Federated Object Storage)

This Service Class extends the current concept of data lakes to provide a unified data lake by virtually integrating not only the IOWN infrastructure but also geographically dispersed data sources such as clouds and on-premises. When considering the implementation of various use cases of IOWN, not all data are necessarily managed within the IOWN infrastructure. Still, cases in which data in multiple cloud, edge, and on-premise environments are analyzed and processed across them are also anticipated. For these cases, Virtual Data Lake provides a mechanism to bundle multiple geographically dispersed storage units and present them as a single object storage unit.

(4) Virtual Data Lake House

IOWN envisions a future in which data held by multiple companies and organizations are shared openly and securely to solve social problems. For example, in a society where renewable energy sources account for most of the energy supply, it is expected that the entire system will be able to be stably controlled by sharing real-time data on the power demand of each company and household, in addition to data on the amount of power supplied. For this purpose, Virtual Data Lake House provides secure and transparent access to various data sources belonging to different owners.

3. Advantages of the IDH

As mentioned above, the IDH supports many of the features that future applications will require in databases and storage services. The IDH also takes advantage of the features of IOWN to achieve functions and performance that are insufficient for conventional databases due to infrastructure limitations. This section describes the current limitations and how they are expected to be solved in the IDH.

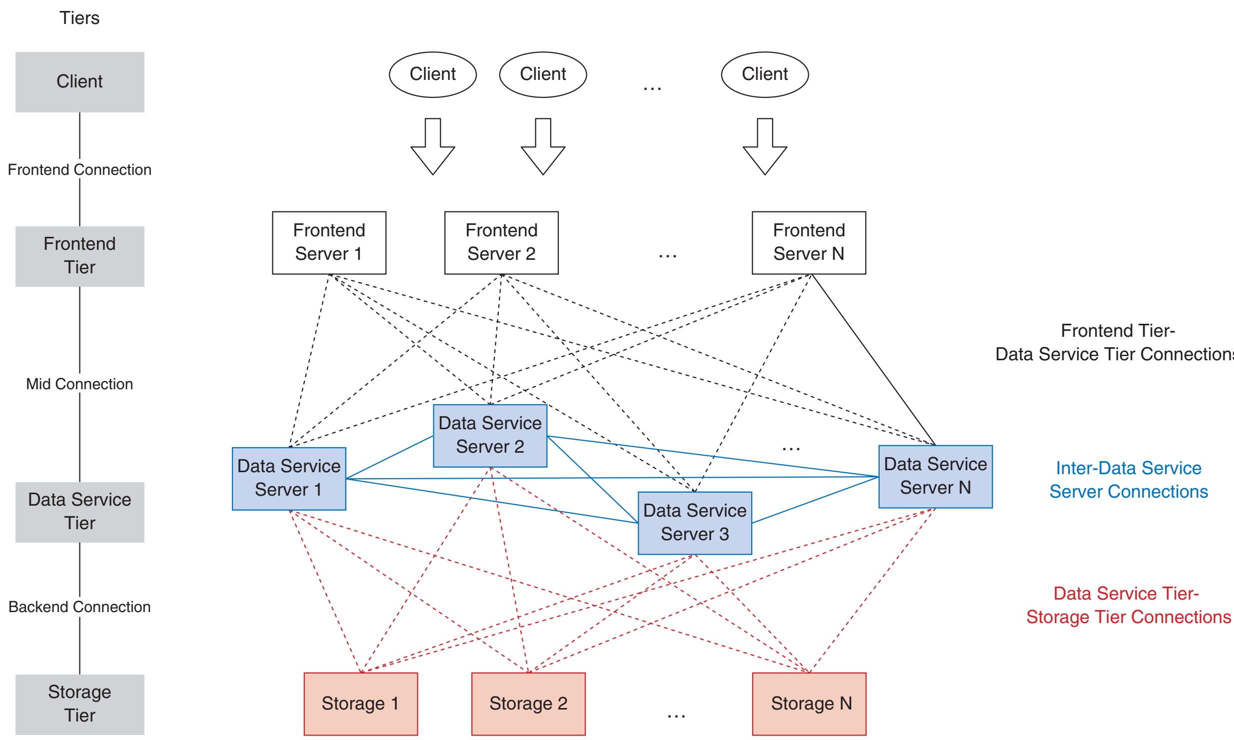


Fig. 1. IDH reference implementation model.

3.1 Limitations of the current situation

For example, suppose we use the Object Storage service implemented in the current cloud model to upload or download 10-GB-class data. The multiple copies of data between the client and storage where the data are stored will become a bottleneck, and stable performance will not be guaranteed. In addition, with the current technology, complex queries that require DRD to join petabytes of data across multiple servers are very slow due to the bottleneck of network connections between database servers. Similarly, for full-scan queries that read several exabytes of data from storage, data transfer from storage becomes a bottleneck. The IDH document also points out the limitations of current cloud-based implementations for other Service Classes.

3.2 Structure and improvements of the reference implementation model

The reference implementation model for the IDH is shown in **Fig. 1**. Although the details differ for each Service Class, this model is the common internal structure for all Service Classes. This structure consists of a Client that requests data from the IDH service, a Frontend Tier that provides load balancing, query routing, protocol conversion, etc., a Data Service Tier that provides data from the local cache or connected storage layer in response to data-access requests, and a Storage Tier that stores data. Each layer consists of multiple components (e.g., servers). The IDH uses IOWN infrastructure technology to accelerate inter-layer and inter-component connections and the operation of each component as follows.

(1) Optimization of communication between the Client, Frontend Tier, and Data Service Tier

The Client exchanges data from the Data Service Tier via the Frontend Tier. To achieve near real-time processing, the latency for these communications needs to be short, and the bandwidth needs to be high, especially when large amounts of data must be transferred repeatedly. In the IDH, Open APN provides direct optical communication between two endpoints to ensure low-latency high-bandwidth connection as needed.

(2) Optimization of communication within the Data Service Tier

When data processing occurs frequently and a large amount of data is exchanged between servers in the Data Service Tier, communication between servers becomes a bottleneck in the data service. Current implementations of inter-server communication using blocking protocols, such as TCP (Transmission Control Protocol), tend to waste server CPU (central processing unit) cycles due to input/output latency. In the IDH, the DCI mechanism enables a dynamic combination of multiple resources to build high-performance clusters. Furthermore, using protocol optimization in DCI (e.g., fast non-blocking communication using a relational database management system (RDMA)), the IDH is designed to process and manage vast amounts of data in near real time.

(3) Optimization of communication between the Data Service Tier and Storage Tier

In IOWN use cases, the total amount of data transferred in a single request can be several hundred terabytes. If there is not enough network bandwidth to communicate between the Data Service Tier and Storage Tier, the service will not respond for a long time. Increasing the network bandwidth with Open APN and optimizing the data transfer with DCI (RDMA) can also help to eliminate this bottleneck.

(4) Making each component smart

In addition to network acceleration, the IDH will actively offload data pre-processing to each component using smart network interface cards. For example, the Storage Tier component filters and aggregates data, reducing the amount of data transferred on the network to less than 1/100th. Also, by distributing to the Client and Frontend Tiers where the data are located in the Data Service Tier, data access requests can be sent directly to the most appropriate component, reducing excess data transfers between members and improving latency and bandwidth.

With these mechanisms, the IOWN reference implementation model is expected to run much faster than the current cloud-based implementation model. The IDH document describes more specific improvements in the implementation model of each Service Class.

4. The Future of the IDH

The IDH document summarizes the technical position of the IDH in the IOWN infrastructure and provides reference implementation models to solve existing cloud-based data processing bottlenecks. However, this should be called the Minimum Viable Version of the IDH features. Future discussions by IOWN GF will lead to more detailed APIs for each Service Type and a better implementation model. In addition, the advanced data processing flow required by future IOWN applications will require more advanced functionalities. For example, in use cases such as smart cities, data owned by multiple organizations will need to be handled across multiple data users. This will require the development of authentication and authorization mechanisms, as well as means to control and audit data usage policies. A feature such as secure computation that executes processing while keeping data secret should also be considered as a function of the IDH. The IOWN GF community will continue to discuss the ideal data-hub services required in the IOWN era.

References

[1] IOWN GF, "Data-Centric Infrastructure Functional Architecture," https://iowngf.org/technology/#Data-Centric-Infrastructure

[2] IOWN GF, "Open All-Photonic Network Functional Architecture," https://iowngf.org/technology/#Open-All-Photonic-Network

[3] IOWN GF, "Data Hub Functional Architecture," https://iowngf.org/technology/#Data-Hub

[4] IOWN GF, "Reference Implementation Model (RIM) for the Area Management Security Use Case," https://iowngf.org/technology/#Reference-Implementation-for-CPS-AM

Tomohiro Inoue

Senior Research Engineer, Supervisor, Social Information Sharing Research Project, NTT Social Informatics Laboratories.

He received a B.E. and M.E. from the University of Tokyo in 1999 and 2001. He joined NTT in 2001 and is currently engaged in the research and development of data hubs, trusted data-sharing infrastructure, and global data-space interconnection as the leader of Information Sharing Architecture Group.

The Challenge to Develop an Artificial Photosynthesis Device that Fixes CO_2 Using the Sun

Yuya Uzumaki, Sayumi Sato, Akihiro Kohno, and Takeshi Komatsu

Abstract

Interest in reducing greenhouse gases, especially carbon dioxide (CO_2), as a measure against climate change is increasing not only in countries but also in companies around the world. The NTT Group has formulated a new vision for zero environmental impact and declared its intention to achieve carbon neutrality by 2040. We expect artificial photosynthesis that converts CO_2 and water (H_2O) into hydrocarbons and molecular oxygen (O_2) using solar energy to be a technology that contributes to CO_2 reduction. To introduce this technology to the market, it is necessary to improve its efficiency and durability. We propose a gas-phase CO_2 reduction reaction system to improve solar-to-hydrocarbon conversion efficiency η_{STC} and conducting basic research on electrodes that make up this system. We studied a nickel oxide/indium gallium nitride (NiO/InGaN) photoanode to achieve both high efficiency and long lifetime. We also studied a copper (Cu)-fiber cathode to improve efficiency. In our system using these electrodes, formic acid (HCOOH) was produced in 140 hours of continuous light irradiation, resulting in an η_{STC} of 0.16%.

Keywords: CO_2 reduction, renewable energy, GaN

1. Introduction

Climate-change awareness is growing worldwide, even with the COVID-19 pandemic. In 2021, the World Economic Forum reported that both the likelihood and expected impact of environmental risks, for example, climate-action failure, extreme weather, and biodiversity loss, are greater than those of economic, geopolitical, societal, or technological risks [1]. Greenhouse gas reduction in particular is being promoted as a measure against climate change. Japan announced at COP26* that in 2030 it will have reduced its greenhouse gas emissions by 46% from those in 2013 and by 2050 will have become a carbon-neutral society. Therefore, in Japan, renewable energies, such as solar power, wind power, geothermal power, small and medium-sized hydropower, and biomass conversion, are attracting increasing interest as promising and diverse energy sources. The NTT Group has formulated a new vision called "NTT Green Innovation Toward 2040" for achieving zero environmental impact through the combination of increasing the use of renewable energy and decreasing energy consumption with IOWN (Innovative Optical and Wireless Network) technologies.

Our team is conducting basic research on *artificial photosynthesis* to contribute to carbon dioxide (CO_2) reduction. Artificial photosynthesis is a technology that converts CO_2 and water (H_2O) into hydrocarbons (formic acid (HCOOH), methane (CH_4), alcohols, etc.) and molecular oxygen (O_2) using solar energy. Currently, CO_2 capture and storage is expected to reduce CO_2, but it is a large-scale system and its

* COP26: The 26th session of the Conference of the Parties to the United Nations Framework Convention on Climate Change.

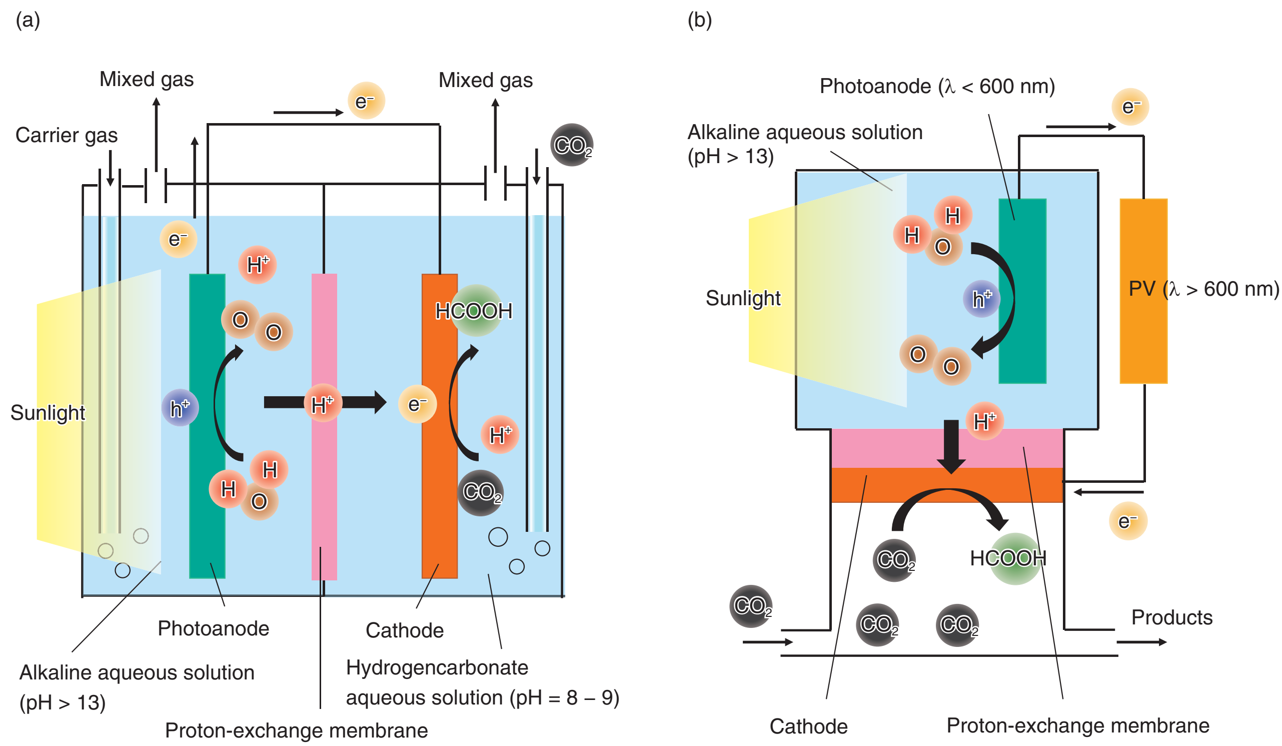


Fig. 1. Schematic illustrations of artificial photosynthesis systems: (a) conventional system using dissolved CO_2 as a reaction material, (b) our system using gas-phase CO_2 as a reaction material.

installation location is limited, especially in Japan. Artificial photosynthesis, which has scalability and is driven by sunlight, is expected to be a new method for CO_2 reduction. With the development of artificial photosynthesis technology, we envision that artificial photosynthesis systems can be attached to buildings, vehicles, etc. to fix CO_2 from the atmosphere, resulting in a more sustainable society.

2. Artificial photosynthesis for CO_2 reduction

An artificial photosynthesis system is mainly composed of a photoanode (negative electrode), cathode (positive electrode), and proton-exchange membrane (electrolyte), as shown in **Fig. 1(a)**. When the photoanode, which is a semiconductor called a photocatalyst, is irradiated with light having an energy larger than the band-gap energy of the semiconductor, electron-hole pairs are generated in the semiconductor, resulting in an oxidation-reduction (redox) reaction. The principle of this system is based on the Honda-Fujishima effect [2]. When high-concentration CO_2 gas is supplied to this system and light is applied to the photoanode, a water-oxidation reaction and CO_2-reduction reactions can proceed as follows:

Oxidation reaction: $2H_2O + 4h^+ \rightarrow O_2 + 4H^+$,
Reduction reactions:
$CO_2 + 2H^+ + 2e^- \rightarrow HCOOH$,
$CO_2 + 2H^+ + 2e^- \rightarrow CO + H_2O$,
$CO_2 + 6H^+ + 6e^- \rightarrow CH_3OH + H_2O$,
$CO_2 + 8H^+ + 8e^- \rightarrow CH_4 + 2H_2O$,
$2CO_2 + 12H^+ + 12e^- \rightarrow C_2H_4 + 4H_2O$.

The target performance for market introduction is 10% solar-to-hydrocarbon conversion efficiency η_{STC} and 10-year durability in the 2030s. The η_{STC} of CO_2-reduction reactions is written as follows:

$$\eta_{STC} = \{(n_{HCOOH} \times Q_{HCOOH}) + (n_{CO} \times Q_{CO}) + \cdots\} / (\text{solar energy}) \times 100\%,$$

where n is the amount of product and Q is the heat of formation. Some research institutes have achieved an efficiency of 10% [3], but most teams have achieved percentages less than that. In addition, no research institute has reported that both efficiency and durability are compatible. Achieving both high efficiency

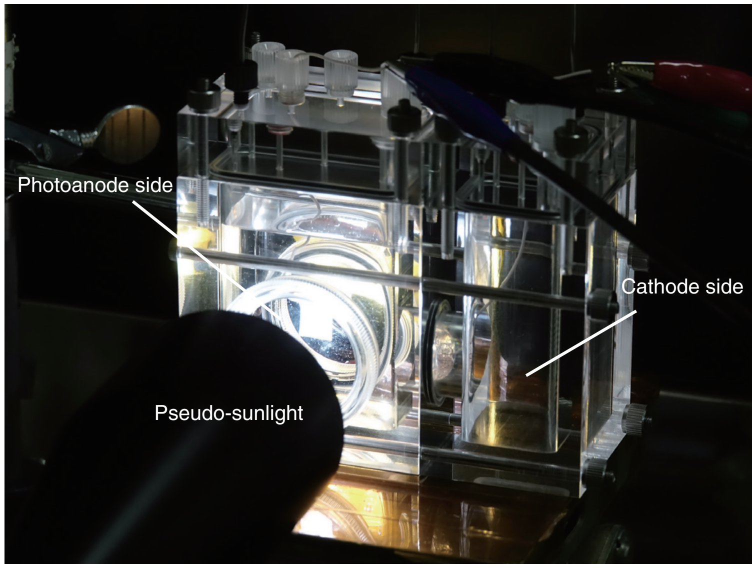


Fig. 2. Light irradiation experiment in our laboratory (photoanode area = 1 × 1 cm^2).

and long lifetime of artificial photosynthesis systems is therefore an important issue in research and development.

To improve the η_{STC}, it is important to increase the Faradaic efficiency of the CO_2-reduction reaction by increasing the CO_2 supply to the reaction area for suppressing the side-reaction ratio [$H^+ + e^- \rightarrow H_2$]. The Faradaic efficiency is the ratio of the number of electrons consumed in the objective reaction to that generated in the photoanode. Even if many electrons are generated in the photoanode, if the CO_2 supply is insufficient at the surface of the cathode, they will be consumed as the side reaction proceeds. In the conventional reaction system shown in Fig. 1(a), CO_2 is dissolved in an aqueous solution and supplied to the reaction area on the surface of the cathode. The amount of CO_2 supplied is limited by the solubility and diffusion coefficient of CO_2 in the solution, and it is difficult to suppress the side reaction.

We propose a gas-phase CO_2-reduction reaction system to solve this issue, as shown in **Fig. 1(b)**. It is easier to increase the concentration and diffusion coefficient of CO_2 in the gas phase than it is to increase them in the dissolved phase. Thus, this system can increase the CO_2 supply to the reaction area and improve the Faradaic efficiency of CO_2-reduction reactions. In addition, a photovoltaic (PV) power generator that uses the light transmitted through the photoanode is connected in series to increase the photocurrent by applying a voltage corresponding to the overvoltage (**Fig. 2**). We are conducting basic research to improve the η_{STC} and lifetime, focusing on the materials and structures of the photoanode and cathode.

2.1 NiO/GaN-based photoanode

The photoanode in an artificial photosynthesis system generates electrons for the proton and CO_2-reduction reactions and generates holes for the water-oxidation reaction. The surface of the photoanode is also a water-oxidation reaction area. We focus on a gallium-nitride (GaN)-based photoanode [4]. The band-gap energy of GaN is 3.4 eV, and the top of the valence band is lower than the oxidation potential of water, and the bottom of the conduction band is higher than the reduction potential of protons and CO_2 (**Fig. 3**). An aluminum gallium nitride (AlGaN)/silicon (Si)-doped GaN (n-GaN) heterostructure and indium gallium nitride (InGaN)/n-GaN heterostructure can improve the η_{STC} because of enhanced electron-hole separation due to the large polarization field in AlGaN and enhanced electron-hole generation due to the wide absorbable wavelength range in InGaN. Thus a GaN-based photoanode is expected to generate O_2 and hydrocarbons. However, there are still issues with improving the η_{STC} and durability toward the target performance.

To improve the η_{STC}, we investigated how to increase the light absorption of the photoanode material. Indium gallium nitride is a mixed crystal of GaN

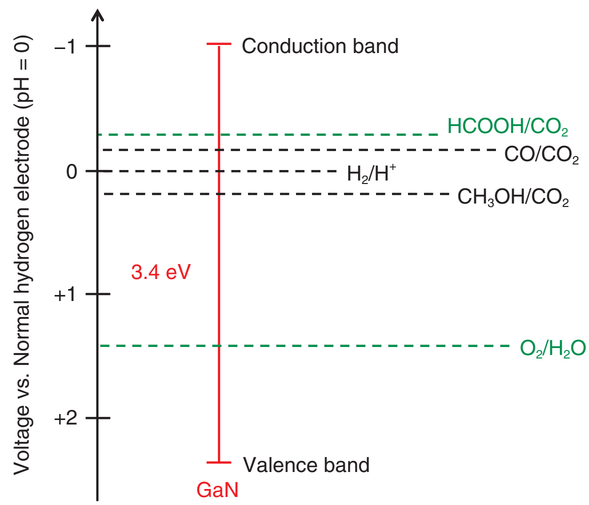


Fig. 3. Band-gap of GaN and redox-reaction level of CO_2-reduction reactions using water as an electron source.

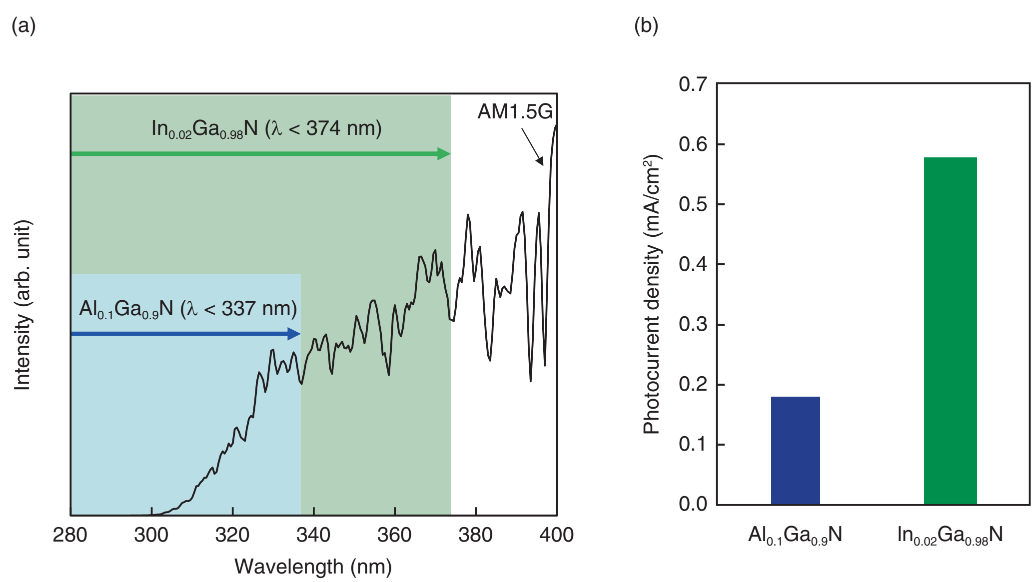


Fig. 4. (a) Theoretical absorption wavelength ranges of $In_{0.02}Ga_{0.98}N$ and $Al_{0.1}Ga_{0.9}N$ with the solar spectrum (AM1.5G) and (b) photocurrent densities measured using $In_{0.02}Ga_{0.98}N$ and $Al_{0.1}Ga_{0.9}N$ photoanodes 1 min after light irradiation.

and indium nitride (InN), and its band-gap energy can be adjusted by changing its In composition. Therefore, as shown in **Fig. 4(a)**, InGaN can use a wider wavelength range for the redox reaction than AlGaN. We prepared an $In_{0.02}Ga_{0.98}N$ photoanode by growing an $In_{0.02}Ga_{0.98}N$/n-GaN heterostructure on a sapphire substrate. We measured the photocurrent in the hydrogen (H_2)-production system, which is the simplest reduction-reaction system using platinum (Pt) as a cathode, under pseudo-sunlight irradiation. The photocurrent density, which is photocurrent per unit light-irradiation area, measured using a 100-nm-thick $In_{0.02}Ga_{0.98}N$ photoanode, was higher than that measured using a 100-nm-thick $Al_{0.1}Ga_{0.9}N$ photoanode (**Fig. 4(b)**) [5]. This suggests that expanding the wavelength range increases the photocurrent and that narrowing the band-gap energy of the photoanode material can further increase the photocurrent. We

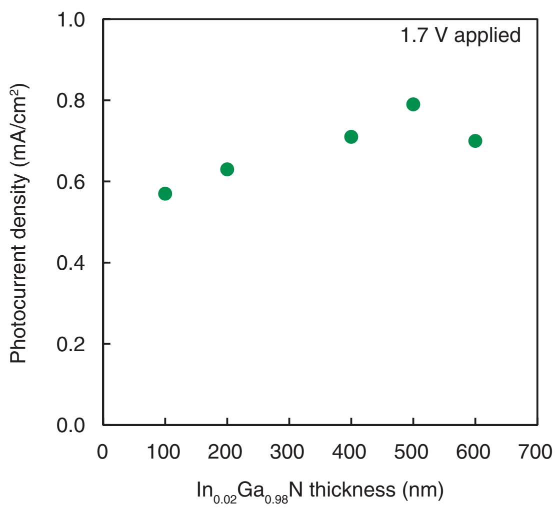


Fig. 5. Photocurrent-density dependence on $In_{0.02}Ga_{0.98}N$ layer thickness 1 min after light irradiation.

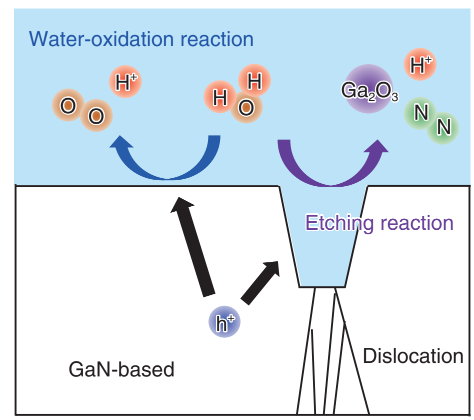


Fig. 6. Schematic illustration of etching process of GaN-based photoanode.

also studied the photoanode structure to further increase light absorption. The light absorption of an $In_{0.02}Ga_{0.98}N$ layer increases with increasing layer thickness because of decreasing light transmittance. **Figure 5** shows the dependence of photocurrent density on $In_{0.02}Ga_{0.98}N$-thickness in an H_2-production system to which 1.7 V was applied under pseudo-sunlight irradiation [6]. The photocurrent density increased with increasing $In_{0.02}Ga_{0.98}N$ thickness, and the maximum density was at 500 nm. Increasing the $In_{0.02}Ga_{0.98}N$ thickness increases the photocurrent density by increasing the amount of electron-hole pairs. It also promotes electron-hole pair recombination caused by the lattice defects in $In_{0.02}Ga_{0.98}N$, thus reducing the photocurrent density. It is thought that the photocurrent density maximized at 500 nm due to the balance between increasing and decreasing the amount of electron-hole pairs.

To improve durability, we investigated how to prevent the etching reaction [$2GaN + 3H_2O + 6h^+ \rightarrow Ga_2O_3 + 6H^+ + N_2$]. The etching reaction is driven by holes generated in GaN-based photoanodes and continues at the interface between the GaN and aqueous solution, as shown in **Fig. 6**. As this reaction progresses, electron-hole pair recombination is promoted by the decrease in the crystallinity of GaN, thus, decreasing the photocurrent. We therefore formed a protective layer on the GaN-based photoanode to eliminate the interface between the GaN and aqueous solution [7, 8]. Nickel oxide (NiO), which is known to transport holes, was used for the protective layer. The thin film of NiO, about 2 nm thick, was formed to sufficiently transmit light. The reaction using NiO/InGaN photoanodes continued for about 20 hours, even though that using bare-InGaN photoanodes had been inactivate for several hours. We also studied the photoanode structure to find how to further suppress the etching reaction [9, 10]. The etching reaction continues with dislocations near the InGaN surface as the starting point. Thus, we prepared a photoanode with reduced dislocations by growing an $In_{0.02}Ga_{0.98}N$/n-GaN heterostructure on a GaN substrate instead of a sapphire substrate, as shown in **Fig. 7(a)**. The dislocation density of the $In_{0.02}Ga_{0.98}N$/n-GaN heterostructure estimated from the full width at half maximum measured with an X-ray rocking curve decreased by an order of magnitude, and the surface roughness also improved. The stability of the photocurrent improved significantly, and the reaction continued for over 100 hours, as shown in **Fig. 7(b)**.

2.2 Metal-fiber cathode on proton-exchange membrane

The cathode catalyzes the CO_2-reduction reaction. Some metals have been reported to catalyze CO_2-reduction reactions, and it is known that the type of hydrocarbons produced differs depending on the type of metal. In our gas-phase CO_2-reduction system, the CO_2-reduction reaction continues in a triple-phase-boundary, as shown in **Fig. 8(a)** [11]. Therefore, it is also necessary to control the CO_2 diffusivity in the cathode to increase the amount of CO_2 supplied to the triple-phase boundary. When the cathode was formed on the proton-exchange membrane by using the plating method, the H_2-generation reaction was dominant

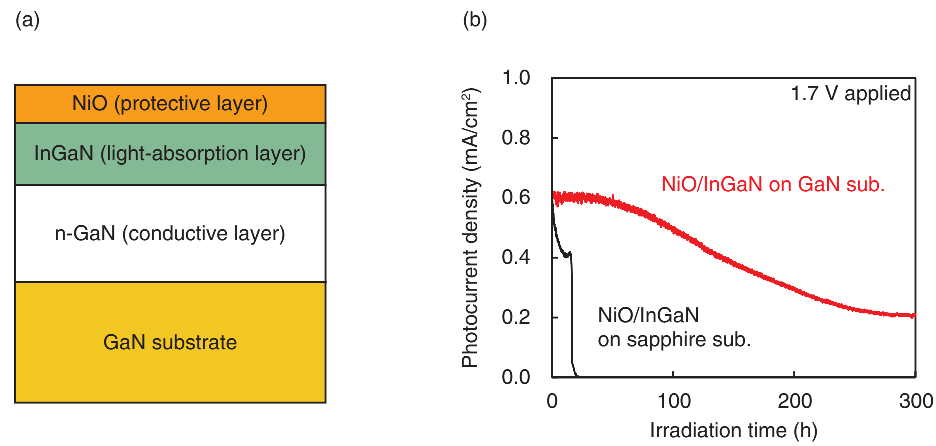


Fig. 7. (a) Structure of NiO/$In_{0.02}Ga_{0.98}N$ photoanode and (b) photocurrent densities in H_2-production system of NiO/$In_{0.02}Ga_{0.98}N$ photoanodes grown on GaN and sapphire substrates. Values obtained with 1.7 V applied under pseudo-sunlight irradiation.

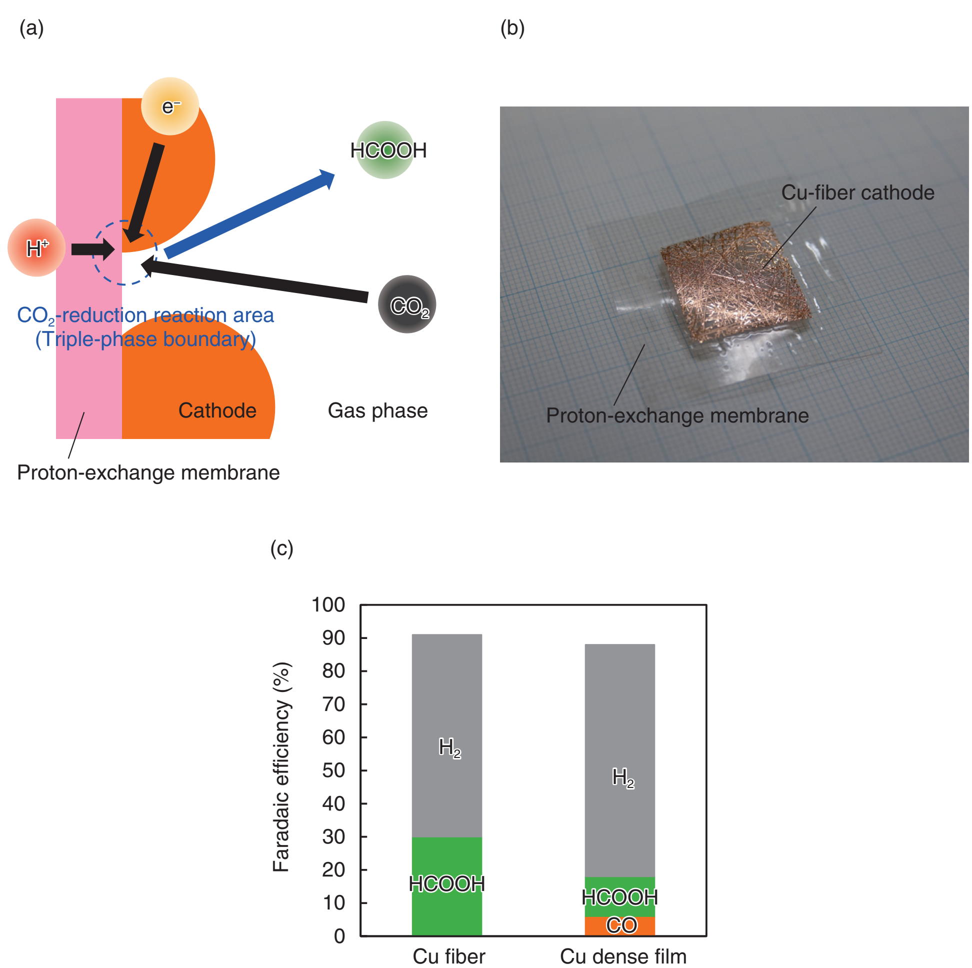


Fig. 8. (a) Schematic illustration of gas-phase CO_2-reduction reaction process at the triple-phase boundary as the reaction area, (b) appearance of Cu-fiber cathode on proton-exchange membrane, and (c) Faradaic efficiency of each product obtained using Cu-fiber and Cu-dense cathodes and a NiO/$Al_{0.1}Ga_{0.9}N$ photoanode.

and η_{STC} was low. This was because the CO_2 supply to the triple-phase boundary was insufficient due to the dense electrode.

To improve the η_{STC}, we studied the metal structure to increase CO_2 diffusivity in the cathode. Gas diffusivity generally depends on the porosity in the electrode. Thus, we focused on a metal-fiber structure with a large void easily controlled by changing the fiber diameter. The metallic fiber is formed on the proton-exchange membrane with a hot press under a condition that maintains the fiber structure, as shown in **Fig. 8(b)** [12]. **Figure 8(c)** shows the Faradaic efficiency of each product obtained using the copper (Cu) fibers and Cu dense film as the cathode in our gas-phase CO_2-reduction reaction system. In this system, HCOOH was mainly generated in the CO_2-reduction reaction and H_2 was also generated as a by-product. The Faradaic efficiency of HCOOH obtained using the Cu-fiber cathode improved compared with that of HCOOH and CO obtained using a Cu-dense-film cathode.

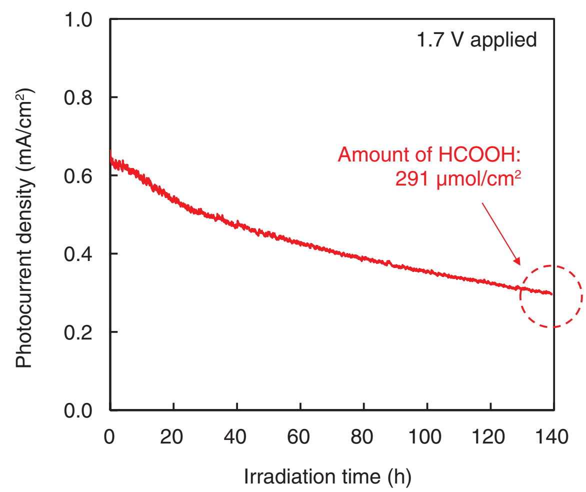


Fig. 9. Photocurrent density using NiO/$In_{0.02}Ga_{0.98}N$ photoanode and Cu-fiber cathode in our gas-phase CO_2-reduction system with 1.7 V applied under pseudo-sunlight irradiation.

2.3 Gas phase CO_2-reduction reaction using NiO/InGaN photoanode and Cu-fiber cathode

We conducted photoelectrochemical measurement using a NiO/$In_{0.02}Ga_{0.98}N$ photoanode grown on a GaN substrate and a Cu-fiber cathode on a proton-exchange membrane in our gas-phase CO_2-reduction reaction system. As shown in **Fig. 9**, the photocurrent density with 1.7 V applied decreased to about half the initial value after 140 hours of light irradiation. The reduction products were HCOOH and H_2. The amount of HCOOH produced 1 and 140 hours after beginning light irradiation were respectively 3.79 and 291 μmol. The η_{STC} and Faradaic efficiency of HCOOH 1 hour after beginning irradiation were respectively 0.25 and 28%, and the average η_{STC} and Faradaic efficiency of HCOOH 140 hours after beginning irradiation were respectively 0.16 and 24%. The amount of fixed carbon per unit light-irradiation area calculated using the chemical reaction formula for HCOOH production was 3.5 mg/cm^2.

3. Conclusion

We proposed the gas-phase CO_2 reduction system to improve the solar-to-hydrocarbon conversion efficiency η_{STC} of artificial photosynthesis. Focusing on the materials and structures of the photoanode and cathode in this system, we are conducting basic research to improve its characteristics. We studied an InGaN photoanode and its thickness to improve the η_{STC} by increasing the light absorption. We also studied a NiO protective layer and lower dislocations of InGaN to improve durability by suppressing the etching reaction of InGaN.

We studied a metal-fiber cathode for improving the η_{STC} by increasing CO_2 diffusivity in the cathode. Using our gas-phase CO_2-reduction reaction system with a NiO/InGaN photoanode grown on GaN substrate and Cu-fiber cathode, HCOOH was produced in 140 hours of continuous light irradiation, and the η_{STC} was 0.16%. For future work, we will study the combining PV as the visible-light response elements with our system for improving the η_{STC} and the system design for improving durability. We aim to achieve CO_2-fixation performance better than that of a plant.

References

[1] World Economic Forum, "The Global Risks Report 2021, 16th Edition," 2021.

[2] A. Fujishima and K. Honda, "Electrochemical Photolysis of Water at a Semiconductor Electrode," Nature, Vol. 238, pp. 37–38, 1972.

[3] N. Kato, Y. Takeda, Y. Kawai, N. Nojiri, M. Shiozawa, S. Mizuno, K. Yamanaka, T. Morikawa, and T. Hamaguchi, "Solar Fuel Production from CO_2 Using a 1 m-square-sized Reactor with a Solar-to-formate Conversion Efficiency of 10.5%," ACS Sustainable Chem. Eng., Vol. 9, pp. 16301–16037, 2021.

[4] M. Deguchi, S. Yotsuhashi, H. Hashiba, Y. Yamada, and K. Ohkawa, "Enhanced Capability of Photoelectrochemical CO_2 Conversion System Using an AlGaN/GaN Photoelectrode," Jpn. J. Appl. Phys., Vol. 52, 08JF07, 2013.

[5] Y. Uzumaki, Y. Ono, K. Kumakura, and T. Komatsu, "Electrochemical Properties of Water Splitting Reaction with NiO/$In_xGa_{1-x}N$/n-GaN

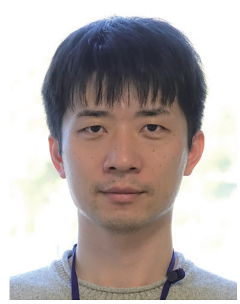
Photoanodes," The 85th Electrochemical Society of Japan Spring Meeting, 1Q22, 2018 (in Japanese).
[6] Y. Uzumaki, S. Sato, A. Kohno, and T. Komatsu, "Photoelectrochemical Properties in Water Splitting Reaction Using NiO/InGaN Photoanodes with Different InGaN Layer Thickness," The 89th Electrochemical Society of Japan Spring Meeting, 1B15, Mar. 2022 (in Japanese).
[7] Y. Ono, Y. Uzumaki, K. Kumakura, and T. Komatsu, "Photoelectrochemical Properties of AlGaN/n-GaN Photoanode with NiO Layer," Proc. of 2017 International Conference on Artificial Photosynthesis, pp. 4–17, 2017.
[8] Y. Uzumaki, Y. Ono, K. Kumakura, and T. Komatsu, "Impedance Analysis of Water Oxidation Reaction with NiO Thin Film/GaN-based Photoanode," The Electrochemical Society of Japan Fall Meeting, 1L32, 2017 (in Japanese).
[9] Y. Uzumaki, S. Sato, Y. Ono, K. Kumakura, and T. Komatsu, "Suppression of Degradation of Artificial Photosynthesis System Using GaN-based Thin Film with Different Dislocation Densities as Photoanodes," The Electrochemical Society of Japan Fall Meeting, 1B06, 2019 (in Japanese).
[10] Y. Uzumaki, S. Sato, Y. Ono, K. Kumakura, and T. Komatsu, "Stability of GaN-based Photoanode for Water Splitting Reaction under Light Irradiation for 300 h," International Conference on Artificial Photosynthesis, P1–61, 2019.
[11] S. Sato, Y. Uzumaki, Y. Ono, and T. Komatsu, "Photoelectrochemical Gas-phase CO_2 Reduction Reaction on Cu Cathode Formed on Proton Exchange Membrane," International Conference on Artificial Photosynthesis, P1–55, 2019.
[12] S. Sato, Y. Uzumaki, A. Kohno, and T. Komatsu, "Photoelectrochemical CO_2 Reduction Reaction with Cu Fiber Electrodes on Proton Exchange Membrane," The 89th Electrochemical Society of Japan Spring Meeting, 1B14, Mar. 2022 (in Japanese).

Yuya Uzumaki
Researcher, NTT Device Technology Laboratories.

He received a B.E. and M.E. in environmental studies from Tohoku University, Miyagi, in 2012 and 2014. He joined NTT Energy and Environment Systems Laboratories in 2014, where he started research and development of a backup system with a combination of power-generating units. He is currently with NTT Device Technology Laboratories, where he is engaged in the research and development of an artificial photosynthesis system for achieving zero environmental impact. His research interests include electrochemical analysis and electrode material and structural design. He is a member of the Electrochemical Society of Japan (ECSJ).

Sayumi Sato
Researcher, NTT Device Technology Laboratories.

She received a B.E. and M.E. in physics from Waseda University, Tokyo, in 2015 and 2017. She joined NTT Device Technology Laboratories in 2017, where she is engaged in the research and development of an artificial photosynthesis system. Her current interests are catalyst for redox reaction and electrode structural design. She is a member of ECSJ.



Akihiro Kohno
Research Engineer, NTT Device Technology Laboratories.

He received a B.E. and M.E. in applied chemistry from Tohoku University, Miyagi, in 2008 and 2010. He joined NTT Energy and Environment Systems Laboratories in 2010 and studied image-restoration algorithms for application to the growth of microalgae. He is currently engaged in the research and development of an artificial photosynthesis system at NTT Device Technology Laboratories. He is a member of the Institute of Electronics, Information and Communication Engineers (IEICE) and the Japan Society of Applied Physics (JSAP).

Takeshi Komatsu
Project Manager, Social Device Technology Laboratory, NTT Device Technology Laboratories.

He received a B.S. and M.S. in physics from Yokohama City University, Kanagawa, and Ph.D. in engineering from Nagaoka University of Technology, Niigata, in 1996, 1998, and 2007. He joined NTT in 1998 and until 2010 conducted research on fuel cells, in particular, a solid-oxide fuel cell. He was with the NTT Environmental Protection Office from 2010 to 2013. He was transferred to NTT Device Technology Laboratories in 2013, where he is currently working on the development of social device technologies that will revolutionize the user experience and overcome social challenges predicted from social trends, such as energy and environment, smart healthcare, and advanced infrastructure maintenance. He is a member of ECSJ.

Latest Research Results and ITU-T Standardization Activities on Soft Errors Caused by Cosmic Rays

Hidenori Iwashita

Abstract

Due to advances in digital transformation, it is important to take measures against soft errors caused by cosmic rays to maintain a safe and secure network. This article explains the latest research results on soft errors from NTT laboratories, commercialization of soft-error test technology, and standardization activities in the International Telecommunication Union - Telecommunication Standardization Sector (ITU-T).

Keywords: soft error, cosmic ray, ITU-T

1. Background and overview

As services become more varied and people seek greater convenience, modern social infrastructures are undergoing digital transformation. However, even as people enjoy greater convenience in everyday life, inexplicable problems with electronic devices are increasing due to cosmic phenomena. When cosmic rays from outer space collide with oxygen or nitrogen atoms in the atmosphere, neutrons are generated. When these neutrons collide with semiconductors within electronic devices, they can cause *soft errors*, which are errors caused by neutrons rewriting data stored in such devices. Soft errors can induce failures that critically impact social infrastructure (**Fig. 1**) [1, 2]. Various measures are taken to ensure stable operation of social infrastructure, such as error-prevention measures within electronic devices and redundancy in equipment and systems. However, as high integration and miniaturization of semiconductors advance, electronic devices will be increasingly affected by neutrons. A soft error is a temporary failure (memory bit inversion) caused by electrical noise, unlike a hard error in which a semiconductor device fails permanently, and the device is recovered by restarting or overwriting it. As shown in Fig. 1, the failure rate of soft errors has risen sharply compared with that of hard errors, which does not change due to the miniaturization of semiconductors. For example, 1 semiconductor device with 10,000 failures in time (FITs: indicates the number of failures per 1 billion hours) causes 0.09 failures per year. If a network is operated with 5000 units equipped with 6 of the same semiconductor devices as telecommunication equipment, it is expected that 262 failures will occur per year. It may be difficult to identify the cause of soft errors because such errors may cause malfunctions or system shutdown due to rewriting of saved data, and once the power is turned off, no trace is left. Since the probability of soft errors occurring per electronic device is extremely low and cannot be reproduced, it may be a heavy burden for the network operator to investigate the cause and take countermeasures. Against this background, soft-error countermeasures have become important for telecommunication systems that require high reliability.

Therefore, to enable countermeasures and evaluation of such soft errors, NTT laboratories have established a soft-error testing technology that can reproduce soft errors in a short time and calculate the soft-error occurrence rate in the natural world and various environments with high accuracy. We have also commercialized the technology and standardized it in the International Telecommunication

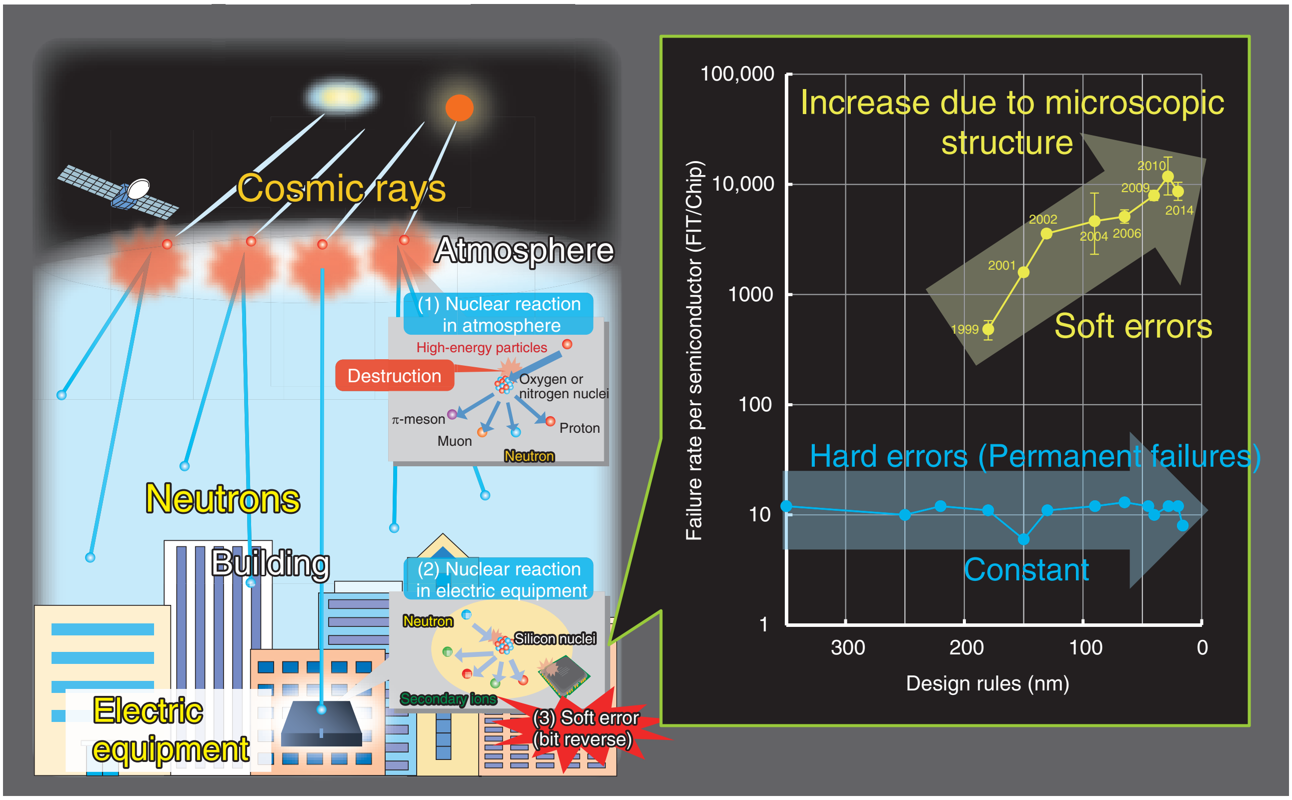


Fig. 1. Mechanism of soft error [1, 2].

Union - Telecommunication Standardization Sector (ITU-T) Study Group 5 (SG5).

2. Measurement of neutron energy characteristics that cause soft errors

To take measures against soft errors, it becomes important to pay attention to the number of failures due to soft errors arising per hour or per day in designing semiconductors and systems. To be able to calculate the number of soft-error-induced failures in a variety of environments, one needs to know the soft-error rate, i.e., the rate at which a soft error occurs at a certain level of neutron speed or energy.

The soft-error rate varies depending on the neutron energy level. The neutrons flying in an environment have an energy distribution that varies from place to place, e.g., on the Earth, in outer space, or on other planets. Therefore, to determine the number of failures caused by soft errors, it is necessary to take the number of neutrons at a given energy in each type of environment into consideration. This is calculated as follows:

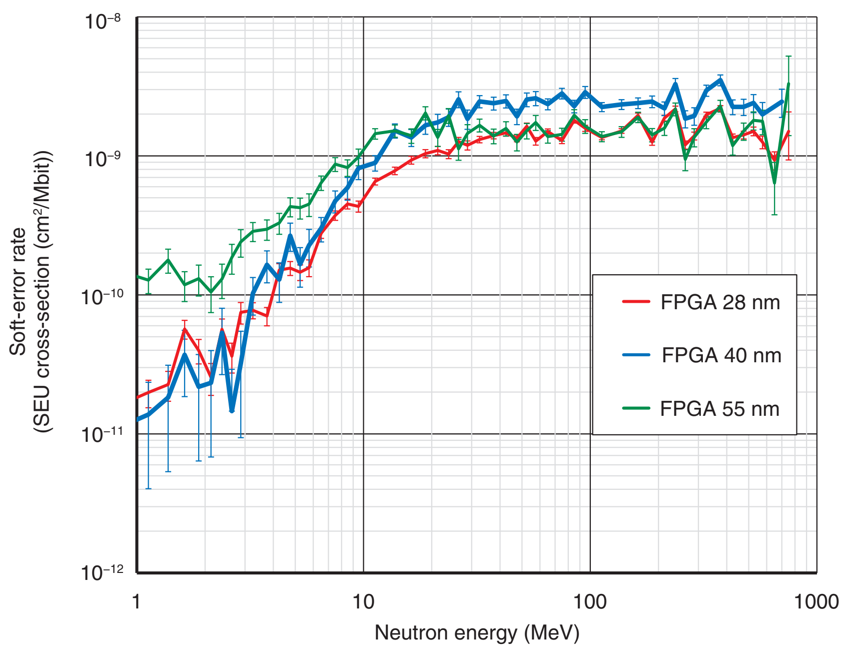
(i) Let the number of neutrons at energy E be φ(E).
(ii) Multiply this by the neutron-E-dependent soft-error rate σ(E).
(iii) The number of failures caused by neutrons at E can be calculated by φ(E) × σ(E).

The total number of failures caused by soft errors at a given place or environment can be obtained as in Eq. (1) by integrating over the number of failures ((iii) above) with all the E distributed in the environment.

Number of failures caused by soft errors

$$= \int_0^{\infty} \varphi(E)\sigma(E)dE. \quad (1)$$

Therefore, the data on the neutron-energy-dependent soft-error rate (in (ii) above) are essential for calculating the number of failures caused by soft errors. However, soft-error rates have been measured only for discrete energies using an accelerator. Consequently, soft-error rates could only be obtained at discrete energies. This has made accurate calculation



FPGA: field-programmable gate array
SEU: single event upset

Fig. 2. Soft-error rate dependent on neutron energy [3].

of the number of failures caused by soft errors extremely difficult. If we are to calculate the number of failures accurately, we need data on soft-error rates measured at continuously varying neutron energies, but such measurement has been considered impossible. NTT has developed an ultra-high-speed error-detection circuit that enables us to precisely measure the flight times of neutrons arriving at a semiconductor even if the velocities are close to the speed of light. From the flight time, we can deduce the speed of the neutrons causing the soft errors. The circuit makes it possible to measure soft errors caused by neutrons across an extremely wide range of energies up to 800 MeV (**Fig. 2**) [3].

3. Soft-error test using accelerator neutron-driven sources

Since the soft-error rate is extremely low per semiconductor device, it is difficult to reproduce soft errors at the development stage of telecommunication systems. Soft errors can be generated in a short time by irradiating the natural world with several orders of magnitude more neutrons. Conventionally, a single semiconductor device has been tested to reproduce soft errors using a high-energy accelerator (several 100 MeV), such as a powerful accelerator at the Los Alamos Neutron Science Center (LANSCE) in Los Alamos National Laboratory, USA. This is because the above-mentioned neutron energy-dependent soft-error rate has not been clarified, so a high-energy accelerator capable of generating a neutron spectrum having almost the same shape as the natural world shown in **Fig. 3** was used [4–6]. If the neutron-spectrum shape is the same, it is possible to easily calculate how many times the acceleration is relative to the natural world by the ratio of the number of neutrons from the accelerator to the number of neutrons in the natural world. However, it has been a very high hurdle in terms of securing machine time and cost since there are only a few accelerators with such specifications in the world. If there are the above-mentioned data on the soft-error rate depending on the neutron energy, even if the neutron spectrum shape is different, the data can be converted into the number of soft-error failures in the natural world, various environments not only at ground level but also at high altitudes, in space, or even on another planet, or other accelerator environments. As shown in Fig. 3, NTT has demonstrated that the neutron spectrum is different from that in the natural world, but soft errors can be reproduced even with the compact accelerator-driven

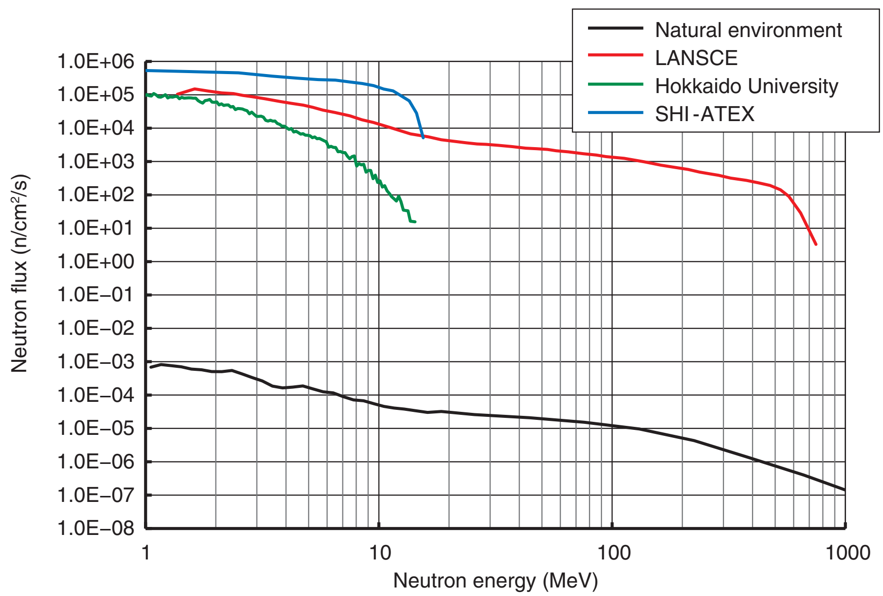
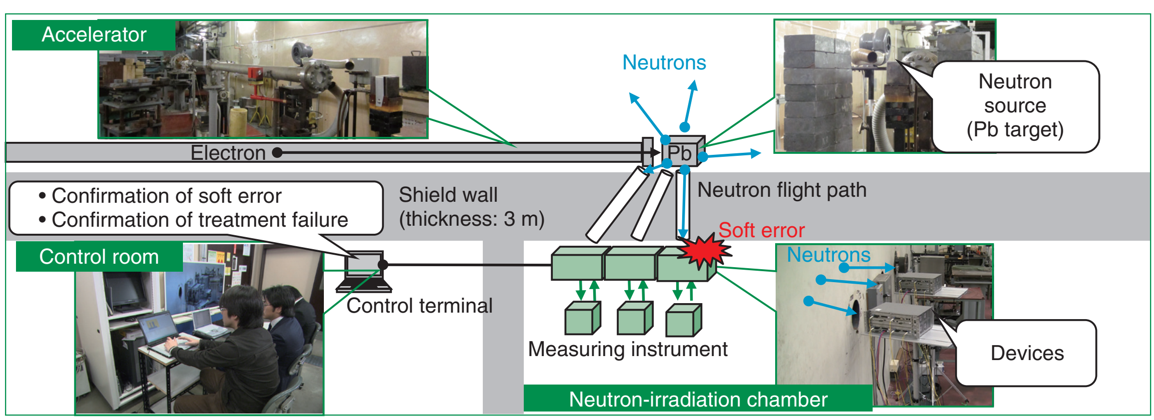


Fig. 3. Neutron spectra of natural environment and accelerators [4–6].



Pb: A chemical symbol of lead

Fig. 4. Soft-error test using a compact accelerator-driven neutron source at Hokkaido University.

neutron source using relatively low-energy electrons (33 MeV) owned by Hokkaido University. We also confirmed that soft errors could be evaluated at the development stage of telecommunication systems (**Fig. 4**). There are many accelerators with specifications of this level in Japan and can be fully used at the development stage from the viewpoint of securing machine time and cost.

Since the problem of such soft errors is related not only to telecommunication systems but also to all electronic devices that require high reliability used in infrastructure, etc., it is expected that there will be demand for this test. We conducted a joint experiment with Nagoya University and SHI-ATEX Co., Ltd. for commercialization, and in December 2016, NTT Advanced Technology started a commercial service for soft-error testing. Soft-error tests are currently being conducted on electronic devices other than for telecommunications.

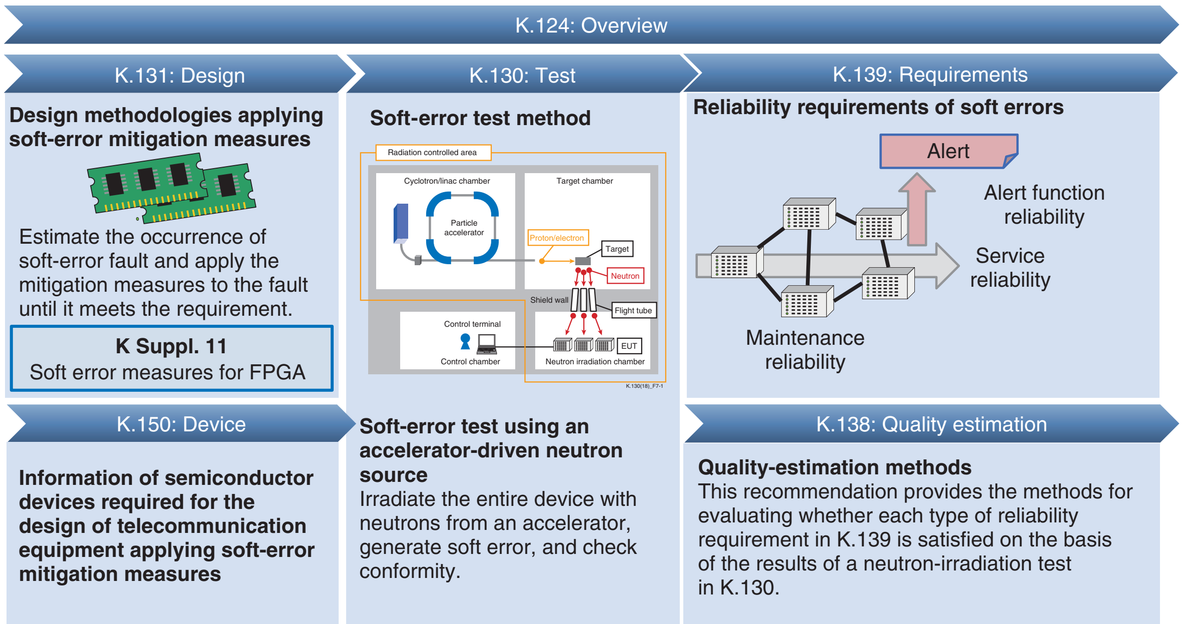


Fig. 5. Overview of Recommendations related to soft errors.

Table 1. List of Recommendations related to soft errors.

No.	Subject	Title
K.124	Overview	Overview of particle radiation effects on telecommunications systems
K.130	Test	Soft error test method for telecommunication equipment
K.131	Design	Design methodologies for telecommunication systems applying soft error measures
K Suppl.11	Supplement	Supplement to K.131 - Soft error measures for FPGA
K.139	Requirements	Reliability requirement of particle radiation effect for telecommunication systems
K.138	Quality estimation	Quality estimation methods and application guidelines for mitigation measures based on particle radiation tests
K.150	Device	Information of semiconductor devices required for the design of telecommunication equipment applying soft error mitigation measures

4. ITU-T's standardization

Against this background, at the October 2015 meeting of ITU-T SG5, commencement of a study on soft errors in telecommunication equipment was approved with the intention of defining requirements on measures to mitigate soft errors, ranging from design techniques to evaluation methods. The Ad Hoc Committee on Soft Error Testing (SOET Adhoc) member companies worked together and developed draft Recommendations. Six Recommendations for soft errors were then enacted at ITU-T in 2018. After that, revised versions that reflected the latest research mentioned above and how to address the issues of soft-error testing were enacted. These Recommendations provide design methodologies, test methods, quality-estimation methods, information on semiconductor devices, and reliability requirements for soft-error mitigation of telecommunication systems (**Fig. 5**, **Table 1**).

Specifically, K.124 provides an overview of the effects of particle radiation and design methods to mitigate the impact of soft errors, K.131 describes the principles and design methods for soft-error mitigation measures for the equipment that comprises carrier telecommunications networks, and K.150 defines characteristic parameters and functions of semiconductor devices that a telecommunication-equipment designer needs when implementing soft-error mitigation measures. In addition, K.139 defines the reliability requirements for telecommunication systems in relation to soft errors, K.130 shows the soft-error test methods using the accelerator, and K.138 describes the reliability-estimation methods on the basis of the results of soft-error testing taking into account the severity of the effect of soft errors. These Recommendations allow telecommunication-systems suppliers to understand soft-error tolerance before actual operation and clarify how much tolerance manufacturers should have.

5. Future perspective

We plan to proceed with research on new countermeasures and evaluation technologies in the space environment by using our technologies on the ground that we have thus far developed.

References

[1] K.Sup11: ITU-T K.131 - Soft error measures of field programmable gate arrays,
https://www.itu.int/rec/T-REC-K.Sup11-201711-S

[2] Xilinx, "Device Reliability Report," UG116 (v10.15), Nov. 2021.
https://www.xilinx.com/support/documentation/user_guides/ug116.pdf

[3] H. Iwashita, G. Funatsu, H. Sato, T. Kamiyama, M. Furusaka, S. A. Wender, E. Pitcher, and Y. Kiyanagi, "Energy-resolved Soft-error Rate Measurements for 1–800 MeV Neutrons by the Time-of-flight Technique at LANSCE," IEEE Trans. Nucl. Sci., Vol. 67, No. 11, pp. 2363–2369, Nov. 2020. https://doi.org/10.1109/TNS.2020.3025727

[4] JESD89B: Measurement and reporting of alpha particle and terrestrial cosmic ray induced soft errors in semiconductor devices,
https://www.jedec.org/standards-documents/docs/jesd-89a

[5] B. E. Takala, "THE ICE HOUSE: Neutron Testing Leads to More-reliable Electronics," Los Alamos Science, pp. 96–103, 2006.
https://permalink.lanl.gov/object/tr?what=info:lanl-repo/lareport/LA-UR-05-8767

[6] H. Iwashita, H. Sato, K. Arai, T. Kotanigawa, K. Kino, T. Kamiyama, F. Hiraga, K. Koda, M. Furusaka, and Y. Kiyanagi, "Accelerated Tests of Soft Errors in Network Systems Using a Compact Accelerator-driven Neutron Source," IEEE Trans. Nucl. Sci., Vol. 64, No. 1, pp. 689–696, Jan. 2017. https://doi.org/10.1109/TNS.2016.2626005

Hidenori Iwashita

Senior Research Engineer, NTT Space Environment and Energy Laboratories.

He received a B.S. and M.S. in nuclear engineering from Hokkaido University in 2006 and 2008. He has been researching soft errors since 2012 at NTT Network Service Systems Laboratories and has been engaged in research on a mitigation technique of soft errors in space at NTT Space Environment and Energy Laboratories since 2020.

External Awards

IPSJ/IEEE Computer Society Young Computer Researcher Award
Winner: Mitsuaki Akiyama, NTT Social Informatics Laboratories
Date: January 17, 2022
Organization: Information Processing Society of Japan (IPSJ)/Institute of Electrical and Electronics Engineers (IEEE) Computer Society

For research on offensive cybersecurity measurement and countermeasure.

DBSJ Kambayashi Young Researcher Award
Winner: Maya Okawa, NTT Human Informatics Laboratories
Date: February 27, 2022
Organization: The Database Society of Japan

For her international achievement in the areas of database, media content, information management, and social computing such as the papers "Deep Mixture Point Processes: Spatio-temporal Event Prediction with Rich Contextual Information" at the ACM SIGKDD Conference on Knowledge Discovery and Data Mining Knowledge Discovery and Data Mining (KDD) in 2019 and "Dynamic Hawkes Processes for Discovering Time-evolving Communities' States behind Diffusion Processes" at KDD in 2021.

37th Telecommunications Advancement Foundation Award
Winners: Maya Okawa, NTT Human Informatics Laboratories; Tomoharu Iwata, Yusuke Tanaka, NTT Communication Science Laboratories; Hiroyuki Toda, Takeshi Kurashima, NTT Human Informatics Laboratories
Date: March 4, 2022
Organization: Telecommunications Advancement Foundation

For "Dynamic Hawkes Processes for Discovering Time-evolving Communities' States behind Diffusion Processes."
Published as: M. Okawa, T. Iwata, Y. Tanaka, H. Toda, and T. Kurashima, "Dynamic Hawkes Processes for Discovering Time-evolving Communities' States behind Diffusion Processes," Proc. of KDD 2021, Aug. 2021.

The 3rd Fumiko Yonezawa Memorial Prize of the Physical Society of Japan
Winner: Keiko Takase, NTT Basic Research Laboratories
Date: March 10, 2022
Organization: The Physical Society of Japan

For research on quantum transport and control of spin-orbit interaction in novel semiconductor materials.

IPSJ SIG Software Engineering Contribution Award
Winner: Haruto Tanno, NTT Software Innovation Center
Date: March 15, 2022
Organization: IPSJ Special Interest Group on Software Engineering

For his significant contribution to activities of IPSJ SIG Software Engineering while serving as a secretary from fiscal 2017 to 2020.

Young Researcher's Award
Winner: Riku Ohmiya, NTT Access Network Service Systems Laboratories
Date: March 17, 2022
Organization: The Institute of Electronics, Information and Communication Engineers (IEICE)

For "Multiple RIS Control Method for Intelligent Radio-wave Design" and "Experimental Area Coverage Evaluation Based on Intelligent Reflector Control."
Published as: R. Ohmiya, T. Murakami, M. Iwabuchi, T. Ogawa, Y. Takatori, Y. Hama, D. Kitayama, and Y. Kishiyama, "Multiple RIS Control Method for Intelligent Radio-wave Design," Proc. of the 2021 IEICE General Conference, B-5-10, Online, Mar. 2021.
R. Ohmiya, R. Taniguchi, T. Murakami, K. Takahashi, R. Kudo, T. Ogawa, and Y. Takatori, "Experimental Area Coverage Evaluation Based on Intelligent Reflector Control," Proc. of the 2021 IEICE Society Conference, B-5-48, Online, Sept. 2021.

Young Researcher's Award
Winner: Yoshifumi Wakisaka, NTT Access Network Service Systems Laboratories
Date: March 17, 2022
Organization: IEICE

For "Frequency-division-multiplex Technique Enabling Both Sampling Rate Enhancement and Fading Noise Reduction in Phase-OTDR Vibration Sensing" and "Crosstalk Impact and Its Suppression in Frequency-division-multiplex Phase-OTDR Distributed Vibration Sensing."
Published as: Y. Wakisaka, D. Iida, Y. Koshikiya, and H. Nazuki, "Frequency-division-multiplex Technique Enabling Both Sampling Rate Enhancement and Fading Noise Reduction in Phase-OTDR Vibration Sensing," Proc. of the 2021 IEICE General Conference, B-13-1, Online, Mar. 2021.
Y. Wakisaka, D. Iida, Y. Koshikiya, and H. Nazuki, "Crosstalk Impact and Its Suppression in Frequency-division-multiplex Phase-OTDR Distributed Vibration Sensing," Proc. of the 2021 IEICE Society Conference, B-13-4, Online, Sept. 2021.

Young Researcher's Award
Winner: Yoko Yamashita, NTT Access Network Service Systems Laboratories
Date: March 17, 2022
Organization: IEICE

For "Design Guideline for Wideband 2-mode Long Period Grating with Arbitrary Coupling Efficiency" and "Study on an Optical Tap Composed of a Long-period Fiber Grating and an Asymmetric Waveguide."
Published as: Y. Yamashita, T. Mori, T. Matsui, and K. Nakajima, "Design Guideline for Wideband 2-mode Long Period Grating with Arbitrary Coupling Efficiency," Proc. of the 2021 IEICE General Conference, B-13-27, Online, Mar. 2021.
Y. Yamashita, T. Mori, T. Matsui, and K. Nakajima, "Study on an Optical Tap Composed of a Long-period Fiber Grating and an Asymmetric Waveguide," Proc. of the 2021 IEICE Society Conference, B-13-21, Online, Sept. 2021.

Young Researcher's Award
Winner: Masanori Koike, NTT Network Service Systems Laboratories

Date: March 17, 2022
Organization: IEICE

For "A Study on Extension of ITU-T P.1204.3 Model for VR Video Streaming" and "VR Video Quality Estimation Model Considering the Difference in Viewport Tile Quality."
Published as: M. Koike, Y. Urata, K. Yamagishi, "A Study on Extension of ITU-T P.1204.3 Model for VR Video Streaming," Proc. of the 2021 IEICE General Conference, B-11-28, Online, Mar. 2021.
M. Koike, Y. Urata, and K. Yamagishi, "VR Video Quality Estimation Model Considering the Difference in Viewport Tile Quality," Proc. of the 2021 IEICE Society Conference, B-11-8, Online, Sept. 2021.

Young Researcher's Award

Winner: Hiroki Iwasawa, NTT Network Service Systems Laboratories
Date: March 17, 2022
Organization: IEICE

For "Proposal of Dynamic TSN Scheduling in a Multitenancy Environment" and "Proposal of Inter-container TSN Arbitration in a Multitenancy Environment."
Published as: H. Iwasawa, N. Azuma, T. Kitsu, H. Masutani, and T. Kuwahara, "Proposal of Dynamic TSN Scheduling in a Multitenancy Environment," Proc. of the 2021 IEICE General Conference, B-6-39, Online, Mar. 2021.
H. Iwasawa, N. Azuma, T. Kitsu, H. Masutani, and T. Kuwahara, "Proposal of Inter-container TSN Arbitration in a Multitenancy Environment," Proc. of the 2021 IEICE Society Conference, B-6-18, Online, Sept. 2021.

ITU-AJ Accomplishment Award

Winner: Shinya Otsuki, NTT Access Network Service Systems Laboratories
Date: May 17, 2022
Organization: The ITU Association of Japan

For his contribution to the creation and revision of the International Telecommunication Union Radiocommunication Sector (ITU-R) Recommendations/Reports regarding fixed wireless access and broadband wireless access as well as the revision of the Radio Regulations regarding 5-GHz-band wireless local area networks through the activities of ITU-R Study Group 5 Working Parties 5A/5C.

Papers Published in Technical Journals and Conference Proceedings

Switching Independent Vector Analysis and Its Extension to Blind and Spatially Guided Convolutional Beamforming Algorithms

T. Nakatani, R. Ikeshita, K. Kinoshita, H. Sawada , N. Kamo, and S. Araki

IEEE/ACM Trans. Audio, Speech, Language Process., Vol. 30, pp. 1032–1047, Mar. 2022.

This paper develops a framework that can accurately perform denoising, dereverberation, and source separation using a relatively small number of microphones. It has been empirically confirmed that Independent Vector Analysis (IVA) can blindly separate N sources from their sound mixture even with diffuse noise when a sufficiently large number ($= M$) of microphones are available (i.e., $M \gg N$). However, the estimation accuracy is seriously degraded when the number of microphones, or more specifically $M - N$ (≥ 0), decreases. To overcome this IVA limitation, we propose switching IVA (swIVA) in this paper. With swIVA, the time frames of an observed signal with time-varying characteristics are clustered into several groups, each of which can be well handled by IVA with a small number of microphones, and thus accurate estimation can be achieved by individually applying IVA to each group. Conventionally, a switching mechanism was introduced into a Minimum-Variance Distortionless Response (MVDR) beamformer, and this paper extends the mechanism to work with a blind source separation algorithm. To incorporate dereverberation capability, we further extend swIVA to a blind convolutional beamforming algorithm (swCIVA) that integrates swIVA and switching Weighted Prediction Error-based dereverberation (swWPE) in a jointly optimal way. With swCIVA, two different time-varying characteristics of an observed signal are captured for dereverberation and source separation to achieve effective estimation. We show that both swIVA and swCIVA can be optimized effectively based on blind signal processing, and their performance can be further improved using a spatial guide for initialization. Experiments demonstrate that both the proposed methods largely outperformed conventional IVA and its convolutional beamforming extension (CIVA) in terms of objective signal quality and automatic speech recognition scores when using relatively few microphones.

Efficiently Generating Ground States Is Hard for Postselected Quantum Computation

Y. Takeuchi, Y. Takahashi, and S. Tani

The 25th Annual Conference on Quantum Information Processing (QIP 2022), Mar. 2022.

In this paper, we show that if ground states of any 3-local Hamiltonians can be approximately generated in quantum polynomial time with postselection, then PP = PSPACE. Our result is superior to the existing findings in the sense that we reduce the impossibility to an unlikely relation between classical complexity classes.