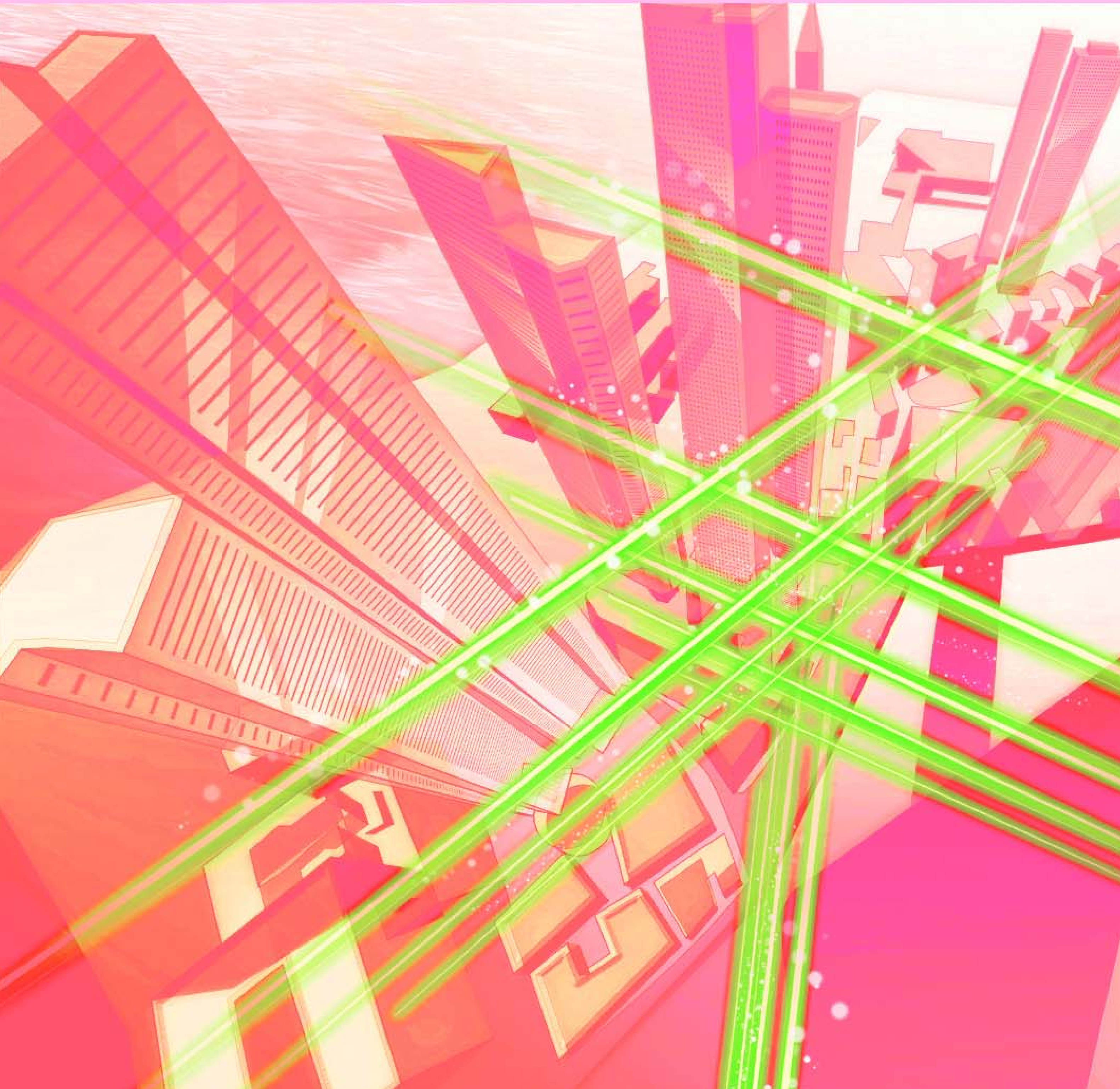
NTT Technical Review

3
2022



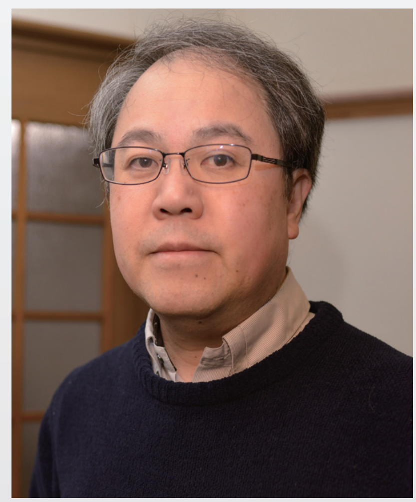
March 2022 Vol. 20 No. 3

NTT Technical Review

March 2022 Vol. 20 No. 3

I Want to Create a World Like That of Astro Boy, Where People and Computers Share the Same Sound Space and Can Freely Cooperate with Each Other

Tomohiro Nakatani
Senior Distinguished Researcher, NTT Communication Science Laboratories



Abstract

Automatic-speech-recognition technology has developed rapidly and is now commonly used in voice interfaces such as those of smartphones and smart speakers; however, the technology must be further improved to enable smooth interaction between computers and humans. Tomohiro Nakatani, a senior distinguished researcher at NTT Communication Science Laboratories, has been at the forefront of research regarding speech enhancement, which removes ambient noise and reverberation from various sounds and accurately extracts only the sound that the person wants to hear. We asked him about the progress of his research and attitude as a leading researcher.

Keywords: voice-user interface, speech enhancement, convolutional beamformer, selective listening

Creating a voice-user interface that understands human speech in any environment

—Would you tell us about your current research?

My research goal is to create a technology that enables computers to distinguish the desired sound from a variety of sounds in an environment and recognize conversational speech. I talked about the same goal during my previous interview in 2016. Compared to back then, operating smartphones and other devices using a voice-user interface has become commonplace. However, even with the current speech-recognition technology based on artificial intelligence (AI), when a person speaks to a computer equipped with such technology, he or she needs to follow a special procedure and change the way of speaking so that the AI can understand. In the future, for robots and other devices to be able to integrate more deeply into our lives, it is necessary to create a natural voice-user interface that can properly recognize natural conversations of people to make users

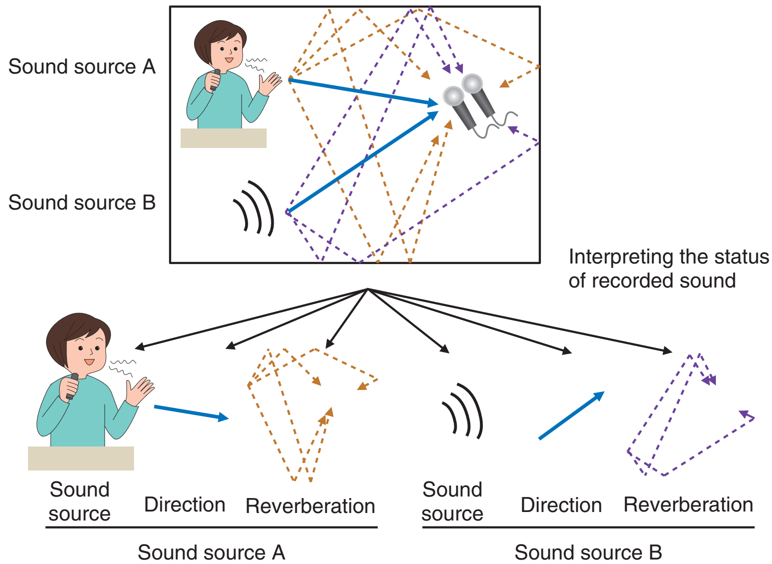


Fig. 1. Elemental decomposition of a recorded sound using multiple microphones.

feel as if they are talking to people in daily conversation.

To make such a natural voice-user interface a reality, I'm currently researching *speech enhancement*, which is a technology to distinguish speech of a target speaker from a recorded sound. In our daily lives, we don't always talk near a microphone. Speech data recorded with a microphone at a distance from the person speaking will include sound reflected off walls (reverberation), the voices of multiple people (speech sources), and background noise. By applying dereverberation, source separation, and denoising, speech enhancement produces speech with quality equivalent to that recorded with a microphone near the mouth of a particular speaker.

We take a two-pronged approach to researching speech enhancement. The first approach is *elemental decomposition* of a recorded sound (**Fig. 1**) with which the sound is decomposed into individual sound elements on the basis of the different physical and statistical properties of the elements. This elemental decomposition is a technological area suited to computers; in other words, it is something that humans cannot do, but computers excel at. One of our recent achievements is the implementation of the world's first convolutional beamformer (BF) that can simultaneously and optimally execute dereverberation, source separation, and denoising. It had only been possible to execute each process individually; thus, it was not possible to get the best performance out of each process when the processes were combined. Our convolutional BF makes it possible to further improve the quality of the enhanced speech from a recorded sound.

The second approach to speech enhancement is *selective listening* (**Fig. 2**) with which only the speech of the person one wants to hear is extracted. This is a technique that people apply in their daily lives; however, until recently, it has been impossible for computers. This approach is represented by our technology called SpeakerBeam, which uses deep learning to extract only the speech that matches the voice characteristics of a particular speaker (as described in articles in the November 2018 [1] and September 2021 [2] issues of this journal). It was the world's first demonstration of human-like selective listening. We have also developed a technology called Recurrent Selective Attention Network, which can accurately estimate who said what and when from ever-changing conversations (conversation analysis).

I believe that a natural voice-user interface can be developed by taking advantage of the technological aspects that computers excel at while simultaneously stepping into the processes that humans excel at.

Fig. 2. Selective listening.

It is a good idea to listen to your seniors and others even if you feel that their way differs from your own

—Was there a reason for starting to work on this topic? What do you keep in mind when searching for a research problem and topic?

Since I joined NTT, I have been working on source separation. When I began my career at NTT, research on source separation was still in its infancy, and there was an atmosphere in which people said, "There's no way we can do that; there are other problems that need to be solved first." Nevertheless, I continued to pursue source separation, believing that if humans can do it, computers should be able to do it, too. Currently, it may not be easy for me to continue research that does not always get the approval of others; however, at that time, I believed in my feeling and pursued research on a topic that I wish I could pursue rather than something I could do right away. As a result, I have been able to continue my research in this field for 30 years, so I think it must have been a really good topic.

I think that, in a sense, searching for research topics and problems is like solving a puzzle, that is, exploring different possibilities until the puzzle is solved. For example, a technology for distinguishing between multiple voices did not exist. Accordingly, after wondering how such a task could be achieved, I made a hypothesis about the characteristics of a voice from the frequency distribution of the speech signals, repeated this step for other speech signals, and identified a core element that is common to many hypotheses, setting it as a research topic.

This process is a matter of knowledge and intuition. Regarding knowledge, you have to study hard, which means reading the latest technical papers and studying textbooks for traditional techniques. The more you do that, the more your skills will improve. There is no other way to improve your knowledge. Regarding intuition, I think it depends on whether you have the vision to think "If it were me, I would do this." I think that is an important ability to have both as a researcher and as a person.

—Even if you become a senior distinguished researcher, you are studying every day.

As the world develops technologically, we cannot afford not to study. In the field of neural networks, the speed of computers has increased dramatically, and the number of researchers has increased so much that research is progressing at a rapid pace, so if you don't study hard, you will never keep up. However, just studying haphazardly will not contribute to actual research. You need to use your intuition and knowledge at the same time, and always think about how you can apply them to what you want to do.

It is also important to have opportunities for discussions with fellow researchers at international conferences. You can only collect a limited number of ideas with your eyes and ears; therefore, it is undoubtedly important to broaden your horizons through discussions with other researchers. In the past, I was instructed by my seniors to "meet people" and "visit research institutions" when I went to international conferences and other events, and I was almost forced to visit researchers halfway around the world. At that time, I had little experience as a researcher and no confidence in communicating in English, so meeting those demands was a very high hurdle for me. Regardless, through such efforts, I was able to come into contact with matters that I had no idea about while getting to know the researchers involved in those matters. Although getting to know someone at a conference is important, visiting their research institutions is the only way to learn many things and discuss them in depth. Even during the current COVID-19 pandemic, I continue to keep up with the human relationships that I nurtured through email and teleconferences.

It may seem like a hurdle to contact a researcher you don't know to ask for a meeting, but the first step is to send an email. In fact, researchers are generally very kind and basically accept meeting with you because they themselves want to get to know other researchers. Sometimes other researchers contact me for the first time, and I never turn them down. By learning the importance of and actually contacting people, I was able to connect with leading researchers in the field of dereverberation and hold workshops with like-minded researchers. Researchers are caught in a world of self-satisfaction if they are not recognized by others; therefore, if you want to be recognized worldwide, you should listen to and refer to the opinions of your seniors and others around you, even if you feel their opinions differ from your own.

Researchers may be like boxes

—Would you tell us what type of attitude should researchers have?

First, I think it is vital to let society know what kind of research you are working on and what problems you want to solve. When you declare your intentions and engage in activities, you can gather a variety of information, and sometimes, you are given solutions that you would not have found on your own.

I believe that researchers are like boxes. A label written on the box states the research agenda. I collect various data myself for the box, and other people send me data and information to add to the box. This action facilitates a project to develop a certain technology, and how much the technology can be developed may depend on the ability of the researcher. Moreover, to collect beneficial information, the way the label is written is also important, so you have to think carefully about how to express it.

The biggest concern of researchers is whether the problem you are working on is something you should really be working on or something you really need to solve, and you must constantly verify it and make the best of it. Fortunately, I found a very good research topic at the beginning of my career and have been working on it for 30 years with almost the same basic *label* and goal. However, as the world has been rapidly developing technologically, I am beginning to think that some parts of my research have become outdated. It may be necessary to make a major upgrade to my research topic and face a new challenge. Assuming that I can continue my research until I am 65 years old, I have about 10 years left as a researcher. I want to think about how I can make those 10 years worthwhile and contribute to society and my research field.

As a researcher, I continue to be interested in enabling a computer to understand sound in the same manner as humans. I don't think that interest will change in the future. A computer should also be able to do what a human can do. In reality; however, it is quite difficult to do that, which makes it all the more interesting, and it has become my life's work. Many problems remain to be solved; even so, in cooperation with many researchers, I want to create a world like that of Astro Boy in which people and computers can share the same *sound space* in any situation, and talk appropriately and cooperate freely while understanding the surrounding circumstances.

—Do you have any words of advice for junior researchers?

My basic advice is to find fellow researchers and create labels for yourself. You can have a cohort of fellow researchers in many situations. For example, in the workplace, there are people you can trust and discuss things with. If you attend a domestic conference, there are fellow researchers who are your rivals. If you participate in an international conference outside Japan, you can link up with fellow researchers who are doing cutting-edge research. As I mentioned before, the technological-development speed of the world and research activities is ever increasing, so we need to work in unison with our fellow researchers to keep up. I have been fortunate to have had many opportunities to collaborate with overseas research institutions and build mutually beneficial relationships.

If you feel stuck in your life in general, you should ask for help and enjoy your life. Because researchers sometimes challenge themselves too seriously when confronting a problem, they tend to shut themselves away if left alone. To prevent that situation from occurring, it may be necessary to ask your seniors to give you support.

Research activities mostly end in failure. At the time of my previous interview, I had a young child, so I compared research to raising a child and said that even if you fail, a seed of research will always be found in that failure. That feeling has not changed. If the result was not what you expected, a different law of nature from what you believed must exist there. By thoroughly examining the situation, you may get the chance to find the next seed of research. When you are inexperienced, you may have a tough time making decisions due to the lack of materials to verify. Nevertheless, you should face the consequences and accumulate experience through repeated failures. If you don't fail, you won't succeed, so don't let failure discourage you from finding the seeds of success.

References

[1] M. Delcroix, K. Zmolikova, K. Kinoshita, S. Araki, A. Ogawa, and T. Nakatani, "SpeakerBeam: A New Deep Learning Technology for Extracting Speech of a Target Speaker Based on the Speaker's Voice Characteristics," NTT Technical Review, Vol. 16, No. 11, pp. 19–24, 2018.
https://ntt-review.jp/archive/ntttechnical.php?contents=ntr201811fa2.html

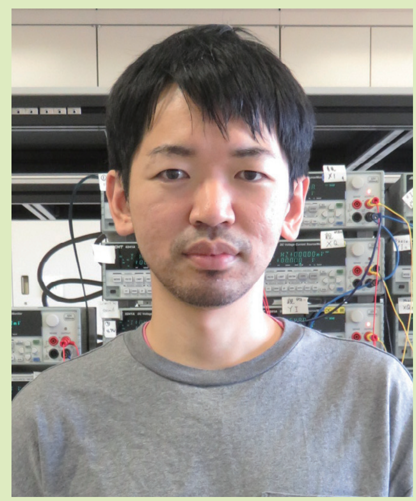
[2] M. Delcroix, T. Ochiai, H. Sato, Y. Ohishi, K. Kinoshita, T. Nakatani, and S. Araki, "Developing AI that Pays Attention to Who You Want to Listen to: Deep-learning-based Selective Hearing with SpeakerBeam," NTT Technical Review, Vol. 19, No. 9, pp. 39–45, 2021.
https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr202109fa4.html

■ Interviewee profile

Tomohiro Nakatani received a B.E., M.E., and Ph.D. from Kyoto University in 1989, 1991, and 2002. Since joining NTT as a researcher in 1991, he has been investigating speech-enhancement technologies for developing intelligent human-machine interfaces. He was a visiting scholar at Georgia Institute of Technology, USA, in 2005 and a visiting assistant professor in the Department of Media Science, Nagoya University, from 2008 to 2018. He received the 2005 Institute of Electronics, Information and Communication Engineers (IEICE) Best Paper Award, the 2009 Acoustical Society of Japan (ASJ) Technical Development Award, the 2012 Japan Audio Society Award, the 2015 Institute of Electrical and Electronics Engineers (IEEE) Automatic Speech Recognition and Understanding Workshop (ASRU) Best Paper Award Honorable Mention, and the 2017 Maejima Hisoka Award. He was a member of the IEEE Signal Processing Society (SPS) Audio and Acoustic Signal Processing Technical Committee (AASP-TC) from 2009 to 2014 and a member of the IEEE SPS Speech and Language Processing Technical Committee (SL-TC) from 2016 to 2021. He served as an associate editor of the IEEE/ACM Transactions on Audio, Speech and Language Processing from 2008 to 2010, chair of the IEEE Kansai Section Technical Program Committee from 2011 to 2012, chair of the IEEE SPS Kansai Chapter from 2019 to 2020, a workshop co-chair of the 2014 REVERB Challenge Workshop, and general co-chair of the 2017 IEEE ASRU. He is a fellow of IEEE, and member of IEICE and ASJ.

Research on Ultra-high-speed Mach-Zehnder Electro-optic Modulators for Long-distance Optical Transmission of 1 Terabit per Second

Yoshihiro Ogiso
Distinguished Researcher, NTT Device Innovation Center/NTT Device Technology Laboratories



Abstract

With the proliferation of the Internet of Things and the launch of fifth-generation mobile communication system services, there is a demand for the optical communication networks that support these services to be even faster. In this article, we speak to Distinguished Researcher Yoshihiro Ogiso, who is researching Mach-Zehnder electro-optic modulators: the end points of fiber-optic communication that contribute significantly to higher speed.

Keywords: electro-optic modulator, Mach-Zehnder modulator, signal-to-noise ratio

Electro-optic modulators convert electrical signals into optical signals

—How does an electro-optic modulator work?

An easy way to visualize optical communication is to picture a flashlight turning on and off to send Morse code messages from the top of a cliff to a ship. By turning the flashlight on for “1” and off for “0,” it is possible to send a digital signal. Of course, such a signal could not reach the other side of the Earth, so actual optical communication is carried out by passing optical signals through optical fibers.

The personal computers and smartphones we use every day are powered by electricity and use electrical signals. Therefore, in order to transmit data via optical communication, electrical signals must be converted to optical signals at some point. This is what electro-optic modulators do. They may sound fairly niche, but they are actually important optical devices that are key to fiber-optic communication.

—What are the Mach-Zehnder electro-optic modulators that you are researching?

I mentioned earlier about turning a flashlight on and off to send a signal, and that really is the principle of an electro-optic modulator.

The simplest method is turning the laser emitter, which is the source of light, on and off. This method is called direct modulation, and is used in optical network units for home use because of its low cost.

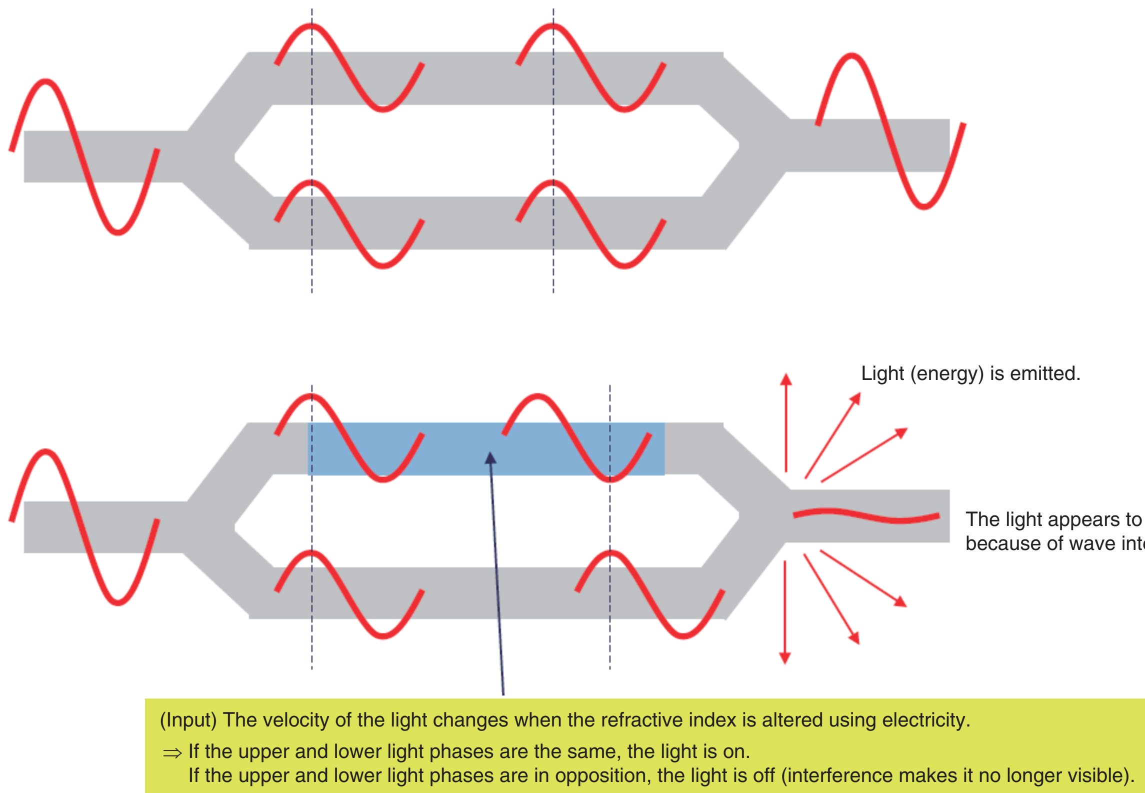


Fig. 1. Principle of Mach-Zehnder electro-optic modulators.

Another method is to shut off the light by placing something that absorbs light in front of a continuously emitting laser. This is called an electro-absorption modulator.

Finally, there is the Mach-Zehnder electro-optic modulator. The Mach-Zehnder electro-optic modulator is constructed so that the light is divided into two at a branch point and then combined, as shown in **Fig. 1**. Light is a wave, so when the upper and lower light waves overlap as shown in the top image, the phase (the position of the peaks and valleys of the wave) is the same, and so the light is on. On the other hand, if the speed of the light is changed by altering the refractive index in the middle using electricity, as shown in the bottom image, this puts the light waves in opposition, with the peaks and valleys canceling each other out, and so the light goes off. The wave interference means light (energy) is emitted, disappears, and is no longer visible.

While direct modulation and electro-absorption modulator need to consider electron carriers that move at a finite velocity, the Mach-Zehnder electro-optic modulator has the advantage of being able to handle high-speed signals, as it mainly uses electric fields that are less sensitive to carrier movement.

—What performance is required from electro-optic modulators?

The most important factor for an electro-optic modulator is the signal-to-noise ratio (SNR). The key requirement is converting electrical signals with a high SNR, which means high-quality signals, into optical signals without them degrading.

To meet this requirement, electro-optic modulators require three key performance features. The first is a low-voltage drive, which is how sensitive the reaction of light is to weak electrical signals. The second is a high bandwidth, which is also related to sensitivity. This indicates how well high sensitivity can be maintained even at high speeds when switching on and off. And the third is low optical insertion loss. The intensity of the light coming out of the electro-optic modulator corresponds directly to the strength of the signal. No matter how high the sensitivity is, the SNR will drop if the overall output drops, so we need a factor that prevents these losses. These three are trade-offs against each other, and broadening them overall means a competitive advantage for the company.

At university I was studying electro-optic modulators made from $LiNbO_3$ (lithium niobate), which was the main material used at the time, but since joining NTT, I have mainly been engaged in research on electro-optic modulators using InP (indium phosphide), a semiconductor material that has low drive voltage, high bandwidth, and can be miniaturized.

Enabling long-distance transmission of 1 terabit per second per wavelength

—In 2019, the world's first long-distance transmission of 1 terabit per second per wavelength was successfully demonstrated.

Increasing speed requires a number of factors. I worked on the part that converts high-speed electrical signals into high-speed optical signals, NTT Device Technology Laboratories was responsible for the part that produces high-speed electrical signals, and NTT Network Innovation Laboratories was responsible for the processing that restores degraded signals at the destination. The whole of NTT achieved this 1 terabit speed together.

There are three ways to speed up an electro-optic modulator. The first is increasing the symbol rate, which simply increases the speed at which light is turned on and off. The second is with higher-order modulation. Only 1 bit of information can be transmitted by turning the light on or off, but if the brightness is set to 0, 0.25, 0.5, 0.75, and 1, for example, it can send a signal like "00001" or "10011." In addition to the brightness and intensity of the light, quadrature amplitude modulation is currently being developed to transmit a 6-bit or 8-bit signal at one time by dividing the phase of the light as a wave. The third is multiplexing. Light has multiple wavelengths, and even if different colors of light are mixed together and transmitted, they can be divided back up by the receiver. In other words, if you create different colors of light and put an optical signal on each, you can transmit as many signals as there are colors with one fiber.

Of these, higher-order modulation and high symbol rate have achieved high speeds, and the electro-optic modulator I created still holds the fastest record in the world. In the past, the speed of electro-optic modulators was plateauing and becoming a bottleneck, so having the most advanced technology in this area is a big advantage for us.

—What are the prospects for practical use?

At present, demand for the field of communication is very high and competition is fierce. As a result, the cutting-edge research level is almost at the actual development level in this field. NTT Device Innovation Center, where I work, is particularly focused on research that can be put into practical use. I believe our mission is not only to come up with theoretical ideas, but to propose and launch technologies that can be applied to tangible things and reach mass production quickly.

The term "valley of death" is often used to refer to a situation in which there are ideas that work well in research, but are difficult to get into the development phase and put into practical use. For example, fabricating InP devices is particularly challenging. InP is a dense material that has alternating indium and phosphorus atoms, so when InP is etched to create a waveguide through which the light of a Mach-Zehnder electro-optic modulator passes, the surface ends up not being very clean. Phosphorus atoms are pulled out, and those parts become a metal-rich indium wall. When another metal is patterned close to those parts, it absorbs the electricity (electric field), which causes the electrical signal to stop passing through.

You therefore need specialist knowledge to reduce damage when cutting and make it the way you pictured it, which we call craftsmanship. That said, if it can only be made by a craftsman, it is not suitable for mass sale, so we are currently working on developing wide-margin design documents and specifications that convey a variety of expertise in the process sheets. These will enable mass-production of electro-optic modulators with similar characteristics even when made by external manufacturers. Other production lines are also being considered, and our final goal is to put them into practical use within a few years.

—What kinds of possibilities will this technology unlock?

In direct terms, it will allow high performance from fifth-generation mobile communication system (5G) services and could be applied to even higher capacity 6G services. It will be effective in data transfer for video streaming, communication between datacenters, and more. It is also expected to be applicable to radio-over-fiber technology, which receives radio waves, transmits them to remote locations via optical communication, then sends them again as radio waves.

Electro-optic modulators could also be applied to fields other than communication. In fact, there are many places where electrical signals are converted into optical signals. In the field of sensing, electro-optic modulators are used to generate optical frequency combs (used for measuring the spectra of the frequency components of light) and for gas sensing. In radio observatories, electro-optic modulators are used to process observation data from distant celestial objects. In other cases, light is converted to electricity. For example, in automotive sensors, information sensed by lasers is converted into electrical signals and shown on displays.

I suspect there may be a lot more scenarios where a need to convert electrical signals into high-speed, high-quality optical signals arises. I have been conducting research and development in the field of communication, but I would love to develop technologies that can be used in other fields too.

■ **Interviewee profile**

Yoshihiro Ogiso joined NTT in 2010, where he was assigned to NTT Photonics Laboratories. He was reassigned to NTT Device Innovation Center following reorganization in 2014. Since July 2021, he has also worked for NTT Device Technology Laboratories. Since university, he has been consistently engaged in the research and development of electro-optic modulators. He received a Ph.D. in engineering.

The World Made Possible by IOWN Digital Twin Computing

Ichibe Naito, Katsuo Inaya, Shigeru Fujimura, Takao Nakamura, Ryo Kitahara, and Koya Mori

Abstract

Digital Twin Computing (DTC) to achieve future forecasting and optimization by linking the real and digital worlds is now being researched and developed as one of the main pillars of the Innovative Optical and Wireless Network (IOWN) vision targeted by NTT. This article introduces the world under DTC and four grand challenges in making DTC a reality.

Keywords: IOWN, Digital Twin Computing, grand challenges

1. IOWN Digital Twin Computing

The Digital Twin Computing (DTC) initiative will be achieved on an innovative network and information processing platform having ultra-large capacity, ultra-low-latency, and ultra-low-power features on NTT's Innovative Optical and Wireless Network (IOWN). This initiative aims to achieve large-scale and high-accuracy future forecasting and advanced means of communication with new value beyond the limits of conventional information and communications technology (ICT) (**Fig. 1**).

2. The world targeted by the DTC initiative

The three key features of DTC are as follows.

(1) Provide a common means for analyzing large-scale and complex interactions among digital twins to enable a wide variety of digital twins to be freely recombined for analysis, testing, and forecasting.

(2) Enable the construction of a virtual world and extension of real-world functions and interactions by making full use of digital properties to combine derived digital twins achieved through replication and processing.

(3) Enable interactions of social aspects such as human behavior and communication by reproducing and representing the inner state of humans including the thoughts and decisions of individuals in digital space.

An explosive increase in the volume of distributed data is expected. In the next industrial revolution called "Society 5.0," there will be an increase in cooperation between humans and machines in the real world, but there should also be a real-time exchange of information between humans and machines in cyberspace through the use of digital twins. The purpose of digital twins introduced in many industries is to copy and simulate things and spaces existing in the real world. However, the DTC initiative to be achieved in IOWN is a new computing paradigm that will enable reproduction of the real world on a scale and with high accuracy heretofore unseen by freely recombining and performing diverse operations on all types of digital twins. It will also achieve mutual interactions in digital space that include the inner state of humans beyond the mere physical reproduction of the real world. The aim is to achieve an environment that can analyze and test a wide variety of social issues. By selecting a better future on the basis of the results of such testing and implementing a feedback cycle to the present, DTC aims to support the ever-changing real world while creating worlds that present an abundance of options tailored to individual desires.

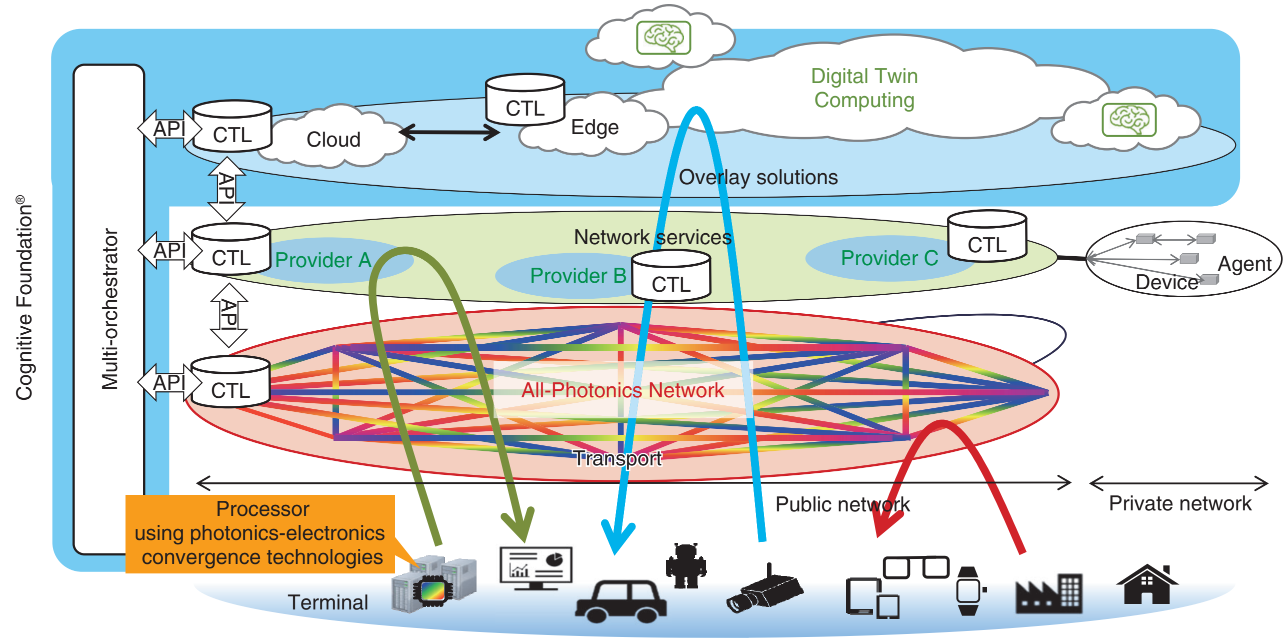


Fig. 1. Conceptual diagram of IOWN.

3. Activities closely related to the DTC initiative

The DTC initiative will contribute to the provision of new value. This will include the 4D digital platform™, which will execute future forecasting and optimization through detailed integrating space and time to solve social problems, as well as the creation of new value and improvement in well-being, which will lead to an enriching future for living a happy life through the introduction of smart technologies in the medical field.

The 4D digital platform™ matches and integrates with high-accuracy four-dimensional information consisting of *latitude*, *longitude*, *altitude*, and *time* enabling data fusion and future forecasting in conjunction with diverse industrial platforms. To achieve a practical platform, NTT is promoting a variety of activities together with various partners in diverse industrial fields. Our goal is to contribute to a behavioral transformation toward a better future by analyzing diverse social activities in the present and forecasting the future, by applying technology that can make position-and-time data obtained from sensors highly accurate, and integrating these data in real time on the Advanced Geospatial Information Database with highly precise and abundant semantic information. Through such future forecasting, we aim to provide new value such as, for example, increasing the smoothness of road traffic flow; maximizing the use of urban assets, such as energy, distribution, and emergency vehicles; enabling cooperative maintenance of social infrastructures; and promoting the understanding of the Earth toward a sustainable environment and disaster prevention.

A bio digital twin (BDT) is one way of contributing to well-being. A BDT is achieved through technologies for estimating unknown functions in the living body through mathematical modeling based on various types of information related to the living body and on the mechanism behind the information network within the body. A BDT constructed with these technologies can be used to derive multiple options in prevention, treatment, and care and conduct detailed simulations of the effects of those options. The aim is to contribute to a future of ongoing medical care that will enable people to face their future health with less worry (**Fig. 2**).

4. Future challenges for making DTC a reality

The above DTC initiative draws up a society that should be aimed for. To achieve such a society, it will

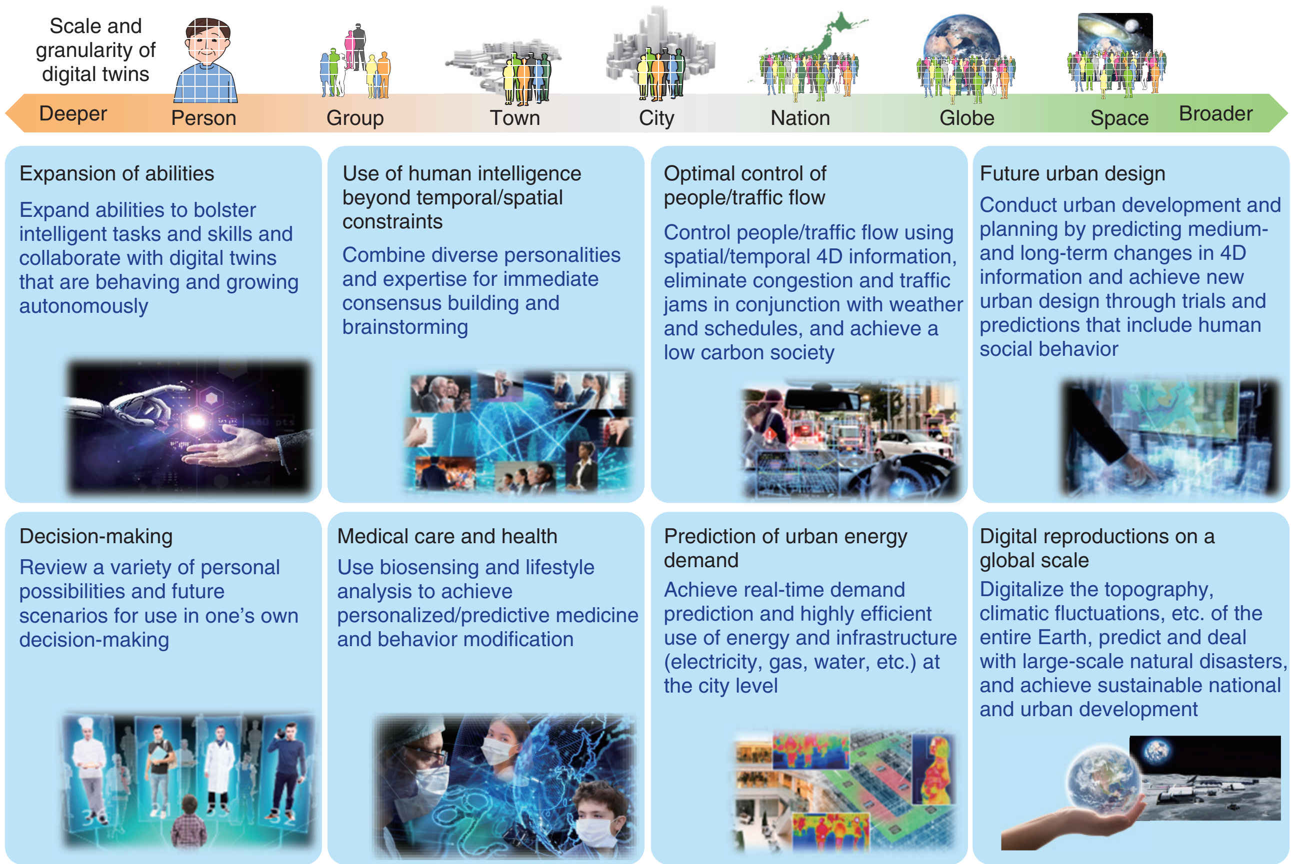


Fig. 2. Application areas of DTC.

be necessary to take a giant leap beyond current technical standards in a variety of technical fields. While adopting an approach that steadily advances technology through research and development (R&D) on the basis of the addition of new ideas through *forecasting*, there will be a need to take on challenges that drive major technical innovations without being bound by conventional approaches.

With this in mind, we decided to pursue R&D through a vision-driven approach. In this approach, we first offered up a bold future vision even if it should appear to be difficult to achieve at first glance. Then, by carrying out *backcasting* from the future vision, we established targets and identified problems to be solved in the most direct way and aimed for our targets by solving the problems (**Fig. 3**).

We had many discussions on setting targets in R&D to actualize our vision for a future society that many people would think attractive to live in. We believe that in a future society having a high diversity of people, expanded opportunities and possibilities, an increasingly complex social structure, and increasing uncertainty on a global scale, harmonious relationships among the Earth, society, and individuals should be fostered while enhancing a person's reason for living and emotional richness. Finally, we narrowed down the directions of this wide-ranging DTC initiative and arrived at the following four major R&D targets to be achieved as grand challenges.

(1) Mind-to-Mind Communications: A new form of communication that transcends not only differences in language and culture but also differences in individual characteristics such as personal experiences and sensitivities to enable direct understanding of how other people perceive and feel the world around them.

(2) Another Me: Not your spouse, your child, or your best friend but rather a digital alter ego as a new and irreplaceable partner that can expand your opportunities in life many times over and live and grow with you.

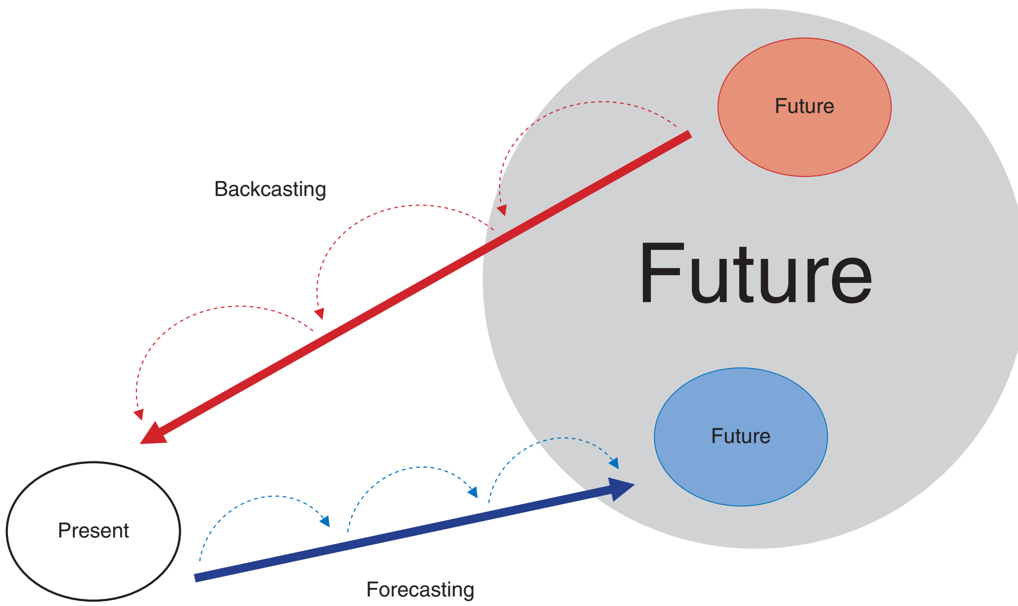


Fig. 3. Forecasting and backcasting.

(3) Exploring Engine for the Future Society: An engine for exploring the form of future societies so that individuals can select the actions and behavior that they desire toward such a society.
(4) Inducing Inclusive Equilibrium Solutions for the Earth and its Social and Economic Systems: Solutions that present an inclusive equilibrium that achieves harmony between the autonomy of the global environment and that of the social and economic systems of the global environment while presenting multiple options for transforming the social system toward that equilibrium.

The Feature Articles in this issue introduce the world targeted by each of these grand challenges as well as current R&D activities and technologies now under study [1–4].

5. Partnering strategy

Each of these grand challenges presents a great vision, and to make these visions a reality, it will be necessary to combine the abilities of many stakeholders having a wide range of knowledge, technologies, and resources. In short, these four grand challenges are not technology-driven targets that can be achieved by the development of a single epoch-making technology. Rather, they can be viewed as solutions to fundamental social problems that require the transformation of social and human consciousness.

For this reason, the strategy to be adopted in collaborating with needed partners will have to be different than any conventional strategy. Partners had been chosen with a focus on similarity and complementarity from the viewpoint of technological elements, and many cases took on the form of joint development with respect to a specific technology. However, achieving the targets of the DTC grand challenges will require searching for partners that can collaborate and combine abilities in finding solutions all from the viewpoint of achieving a stated vision. This will require tie-ups with partners in different fields that can share a common vision such as partners having completely different technologies than us, having expertise for solving social problems as opposed to technical problems, and having the abilities and resources for transforming society such as government and non-governmental organization. It will also require the combining of strengths possessed by NTT and each partner.

We can give the fourth grand challenge, "Inducing Inclusive Equilibrium Solutions for the Earth and its Social and Economic Systems," as a specific example of this strategy. In this challenge, we proposed a grand vision of presenting an inclusive equilibrium that achieves harmony between the autonomy of the global environment and that of the social and economic systems of the global environment while presenting multiple options for transforming the social system toward that equilibrium. We are now pursuing R&D toward the embodiment of that vision together with specialists in climate models and those in circulation models including financial models that support this vision, with us at NTT laboratories working together as a system to interconnect all the people involved. As part of this process, we are searching out new forms of collaboration with partners that are not limited to the development of individual technologies but that combine in some way the technologies and knowledge needed to achieve an overall vision.

Standardization activities can also be viewed as a partnering strategy. In the past, members having strengths in the same technical area within an industry would come together to formulate common specifications. In DTC, however, the scope of application is expansive with a number of industries included, and there are elements of DTC in which acceptance by society is an important factor such as human digital twins. As a result, DTC standardization activities are not just a matter of holding discussions on technical issues. It is also necessary to promote discussions with diverse stakeholders with a view to implementing DTC in society.

6. Conclusion

The DTC initiative holds the possibility of greatly changing the way that society should be 10 and 20 years from now. To this end, there will be a need for R&D that takes an approach different from the conventional one by proposing future visions as typified by grand challenges. To achieve IOWN that includes the DTC initiative, it will be essential to bring together a variety of research and technology fields such as the social sciences, humanities, natural sciences, applied sciences, as well as interdisciplinary fields. At NTT laboratories, we will continue in our efforts to implement the IOWN vision while collaborating with experts and global partners from this wide array of research and technology fields.

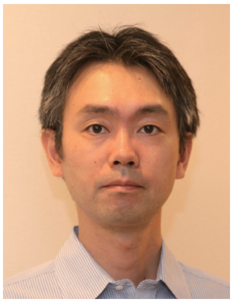
References

[1] R. Saijo, Y. Tokunaga, D. Yamaguchi, L. Andarini, S. Matsuo, I. Toshima, T. Kurahashi, and S. Ozawa, "Studies on Skill Level and Dialogue Satisfaction for Achieving Mind-to-Mind Communications Technology," NTT Technical Review, Vol. 20, No. 3, pp. 16–20, Mar. 2022.
https://ntt-review.jp/archive/ntttechnical.php?contents=ntr202203fa2.html

[2] A. Ohtsuka, C. Takayama, F. Nihei, R. Ishii, and T. Nishimura, "Technologies for Achieving Another Me," NTT Technical Review, Vol. 20, No. 3, pp. 21–25, Mar. 2022.
https://ntt-review.jp/archive/ntttechnical.php?contents=ntr202203fa3.html

[3] N. Shigematsu, A. Isomura, I. Ueno, N. Oki, Y. Arakawa, and K. Yoshida, "Initiatives toward Development of the Exploring Engine for the Future Society," NTT Technical Review, Vol. 20, No. 3, pp. 26–30, Mar. 2022.
https://ntt-review.jp/archive/ntttechnical.php?contents=ntr202203fa4.html

[4] M. Maruyoshi, Y. Muto, and D. Tokunaga, "Coupled Simulation Technology for Visualizing Environmental and Economic Social Cycles," NTT Technical Review, Vol. 20, No. 3, pp. 31–35, Mar. 2022.
https://ntt-review.jp/archive/ntttechnical.php?contents=ntr202203fa5.html



Ichibe Naito

Senior Research Engineer, Research and Development Planning Department, NTT.

He joined NTT in 2006 and has been engaged in R&D of distributed data stream management system and Personal Information Management system. In 2010, he moved to NTT Communications where he designed the database of the operation support system and developed an Infrastructure as a Service service. In 2013, he joined NTT Software Innovation Center and developed an operation support system for distributed object storage. He is currently engaged in the research and development of DTC, security, and data processing.

Katsuo Inaya

Director, R&D Produce Group Medical and Health Care Team, Research and Development Planning Department, NTT.

He joined NTT in 1995. He is an experienced engineer with a long history of working in the information technology and services industry. He has experience in the areas of enterprise software, business development, strategy, strategic partnerships, and mobile devices. His current business interests include deep learning technology and the medical industry.

Shigeru Fujimura

Senior Manager, Strategic Business Creation Team, Research and Development Planning Department, NTT.

He joined NTT in 2005. Since then, he has been involved in the research and development of Web mining, Web technology, and blockchain technology. He is the W3C (World Wide Web Consortium) Advisory Committee Representative of NTT and has been in charge of promoting the 4D digital platform™ since 2021.

Takao Nakamura

Senior Research Engineer, Supervisor, NTT Digital Twin Computing Research Center.

He joined NTT Human Interface Laboratories in 1994 and studied media processing technologies, content distribution systems, artificial intelligence, and their applications. He was seconded to the Ministry of Internal Affairs and Communications of Japan from 2007 to 2009, where he was responsible for policy making and promoting national information security policies. He is working on research and development of the concepts and technologies for DTC.

Ryo Kitahara

Senior Research Engineer, NTT Digital Twin Computing Research Center.

He joined NTT in 2003. His research interests include image-processing technologies. Since 2020, he has been in charge of research on DTC at NTT Digital Twin Computing Research Center.

Koya Mori

Senior Research Engineer, NTT Digital Twin Computing Research Center.

Since joining NTT in 2004, he has been involved in the research and development of an internet-of-things application based on software technologies, such as OSGi, OpenStack, and Edge Computing. Since 2020, he has been in charge of research on DTC at NTT Digital Twin Computing Research Center.

Studies on Skill Level and Dialogue Satisfaction for Achieving Mind-to-Mind Communications Technology

Ryohei Saijo, Yoko Tokunaga, Daichi Yamaguchi, Lidwina Andarini, Shohei Matsuo, Iwaki Toshima, Takao Kurahashi, and Shiro Ozawa

Abstract

We have undertaken the development of mind-to-mind communications technology to enable a uniform means of communication that can be mutually understood by all humans. As the first step in this effort, we focused on (1) how to convey information tailored to the other person by studying technologies for presenting information tailored to work skills and (2) how to evaluate communication by evaluating and estimating satisfaction on the basis of positivity and impact of a dialogue participant. As the next step, we will refine the technologies presented in this article to improve understanding of utterance intention through expressions that match one another's sensitivities.

Keywords: mind-to-mind communications, skill level, satisfaction level

1. Activities toward mind-to-mind communications

In human communication, information uttered by a person will generally not be conveyed to the other person with 100% accuracy. It is not unusual for some misunderstanding (miscommunication) to occur, and real dialogue may never become established (discommunication). As a means of communication to avoid these problems, we are developing mind-to-mind communications technology to help convey the feelings (sensitivities) that one wants to communicate to another person. By conveying sensitivities, mind-to-mind communications technology conveys what a person wants to convey accurately, which makes it possible to find a new solution or reach a better agreement while maximizing communication results and the satisfaction level of the parties concerned. This requires that sensitivity be considered from two viewpoints, i.e., sensitivity of the sender and sensitivity of the receiver. In other words, the task is to determine how the sensitivity that a person wants to convey will be expressed depending on the sensitivity of the sender and predict how that will be interpreted depending on the sensitivity of the receiver. Studies are needed to determine what is necessary to satisfy the parties concerned in actual communication. The reason for this is as follows. In a situation in which communication aims to solve a problem or build a consensus, it is not enough to simply reach a better consensus from on objective point of view—it is also essential that the parties concerned be satisfied and reach an agreement in a subjective manner (on the basis of the sensitivities of each party).

We present the following two example studies related to these two problems:

(1) How can the sensitivity that a person wants to convey be conveyed in a manner tailored to the other person? We investigated an interface that can determine a person's level of skill from biological signals and switch the

information to be presented accordingly.

(2) What effect does the attitude of the dialogue participants have on the degree of satisfaction with respect to the results of that dialogue? We examined the results of a basic experiment on the degree to which positive dialogue participation can contribute to satisfaction.

We hope that these studies will serve to stimulate discussion on mind-to-mind communications technology.

2. Interface technology for presenting information tailored to work skills

Given a scenario of collaborative work in which a variety of people come together to accomplish the same task while engaging in communication, it is extremely important that information be conveyed in a manner tailored to the other person. We have been aiming for some time to make collaborative work even smoother by achieving information communication tailored to the other person through the use of human digital twins. In terms of technologies and services in support of collaborative work, there have been many proposals on how best to communicate work instructions or supplementary information from one user to another. Most of these technologies and services provide such support in a uniform manner to each user (for example, displaying helpful marks at specific locations within an online collaborative work space).

However, the support needed during work differs according to the user's skill level, and the cognitive load may also differ in accordance with the information being given (or not being given). For example, we can consider a game of pair *shogi* (Japanese chess played in pairs) in which one pair exchanges opinions on their next move. For expert and intermediate players, a simple information display, such as displaying marks at important locations on the *shogi* board, is considered sufficient. However, for beginners, there may be a need for a more detailed explanation such as showing how pieces are moved using a drawing or illustration and explaining the merits of each move. When targeting users of various levels, there is a need for technology that would enable a system to understand those differences and provide appropriate support tailored to each type of user.

This technology would estimate a user's skill level from biometric information obtained from the user while working and from information related to past work experience and reproduce those aspects in a human digital twin corresponding to that user. That human digital twin could then be used to infer the information needed by the user and switch the information display accordingly without obstructing work (**Fig. 1**). This technology will enable collaborative work support promoting smooth communication even among users having different sensitivities, such as intuition and ability of understanding, that affect work skills.

Referring to a previous study [1] stating that eye movement (gaze) during work differs according to skill level with respect to that work, we have undertaken the use of sensing data centered around eye tracking*. Our aim is to conduct machine learning that combines the amount of eye movement, movement patterns, user work experience, difficulty of work, etc. to estimate the user's skill level. We are pursuing the design of an information-presentation method tailored to the user through wide-ranging studies on information displays including the modal to be used (graphics, voice, etc.), granularity of information, and placement of information in space.

As an example of an activity related to skill level, we introduced interface technology for estimating skill level using gaze data and for displaying information on the basis of that estimation. Going forward, we will deepen our studies on information-presentation methods that incorporate user-interface/user-experience knowledge [2] accumulated at NTT Digital Twin Computing Research Center while expanding our work toward the development of a more human-centric system.

3. Technology for estimating satisfaction level based on positivity and impact of a dialogue participant

When building a consensus through dialogue, whether the dialogue that took place was satisfactory or the results achieved are acceptable may differ from one participant to another. For example, there are people who place importance on having a positive attitude as in "I was able to talk a lot and enjoyed the discussion" and who evaluate their level of satisfaction with the dialogue on the basis of the number of times they got to speak. However, there are also people who place importance on the degree to which they were able to contribute as in "My opinion helped to build a consensus" and who evaluate their level of

* Eye tracking: A biometric technique that analyzes the motion of a person's eye to clarify visual attention and other characteristics.

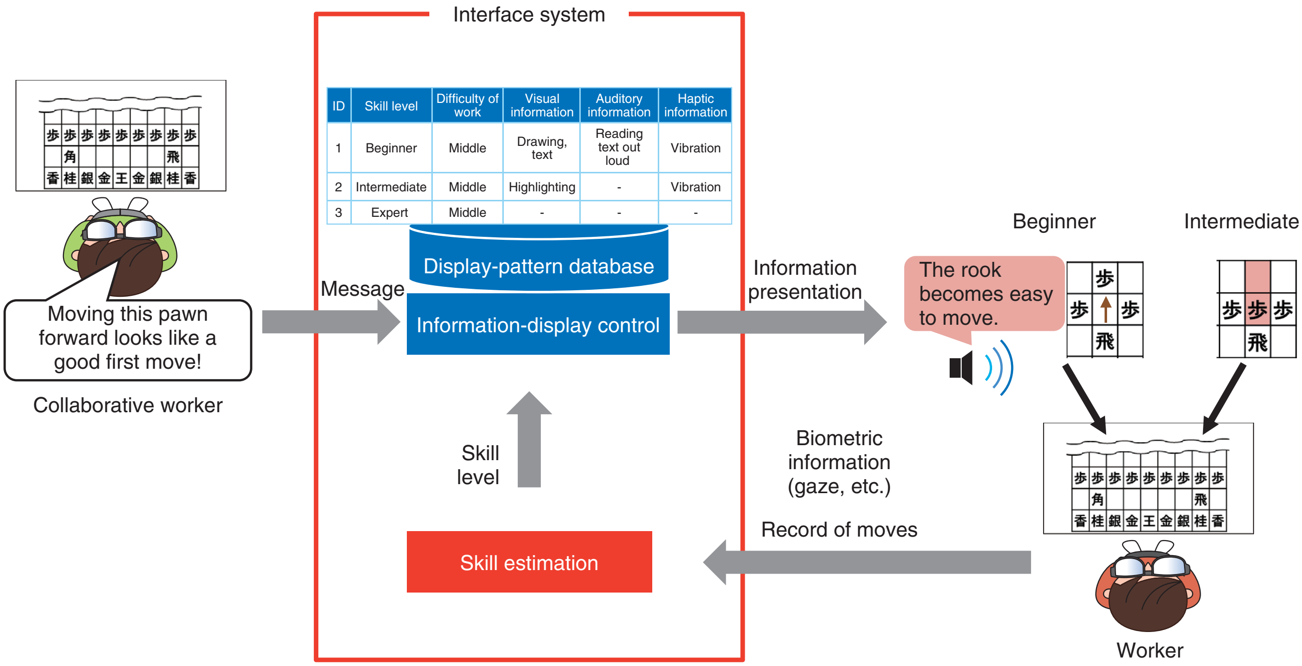


Fig. 1. Overview of interface system for displaying information tailored to skill level.

satisfaction on the basis of whether their opinion is reflected in the results achieved. It is thought that such differences are influenced by the values that participants hold with respect to dialogue. That is to say, differences arise according to what individual participants consider essential for a dialogue and to what they consider to be the factors changing the levels of satisfaction and acceptance [3]. We are testing the reproduction of such values in human digital twins as the inner state of humans to estimate satisfaction level for individual dialogue participants. With this technology, we should be able to estimate the levels of satisfaction and acceptance of each dialogue participant and the reasons for those estimations, follow-up on each participant using expressions tailored to the sensitivities of that person, and provide assistance for improving creativity in consensus building such as through teaming techniques or dialogue intervention to raise the level of satisfaction of all participants. Among the various factors that can be considered to affect the degree of dialogue satisfaction, we describe an estimation technique that focuses on participant positivity and the effect of participant remarks (**Fig. 2**).

First, we considered the extent to which a participant was able to positively participate in a dialogue to be one factor affecting that participant's level of satisfaction and quantified that as a positivity score. We first classified remarks made by a participant as either content-rich or content-poor. Content-rich remarks are those that include content related to the topic of the dialogue. They are relatively long utterances that include many nouns, verbs, and adjectives. Content-poor remarks, on the other hand, are those consisting of backchannel. They are relatively short utterances that include hardly any nouns, verbs, or adjectives. We then defined the positivity score by dividing dialogue data into specific time frames and using the amount of content-rich and content-poor remarks made by a specific participant in each time interval.

Next, considering that the degree to which remarks made by a participant has stimulated the dialogue and helped build consensus can contribute to the satisfaction level of that participant, we quantified that as an impact score. This score is based on how long the dialogue continued on the topic initially proposed by the target participant. We first prepared a list of important words on the basis of a document compiling the results of the consensus reached by the group conducting the dialogue. In the event that such important words appeared in the participant's utterances, we then determined the number of turns that the participant took in uttering those words including related words within the group. We also added weights

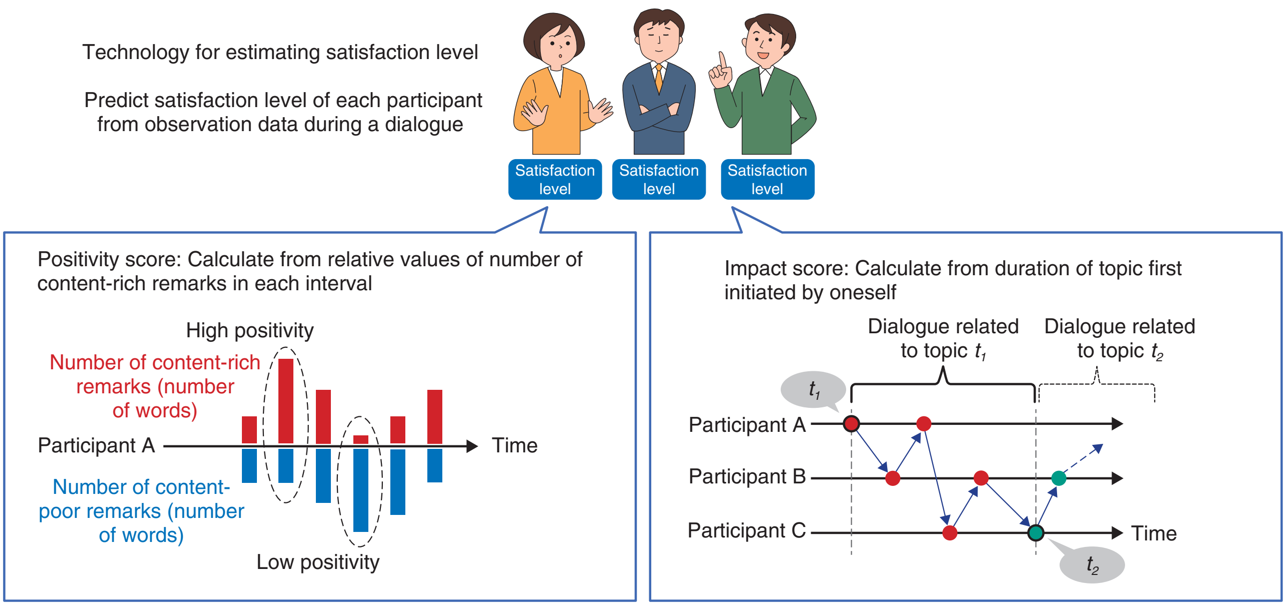


Fig. 2. Overview of dialogue-satisfaction estimation using positivity toward dialogue and impact of remarks.

depending on whether the person who first used those important words during the dialogue was that participant or another participant and calculated the impact score on the basis of the above characteristics.

We conducted a simulated dialogue among four participants meeting for the first time and estimated the level of satisfaction using the two scores described above. For comparison, we treated the answers given in a questionnaire on satisfaction level as correct. Results revealed that estimation accuracy improved using the positivity score when each topic was created during the beginning portion of a dialogue. It was also found that the effects of the positivity score were minimal in brainstorming and consensus building during the middle and end of the dialogue but that estimation accuracy improved using the impact score. These results indicate that the factors affecting satisfaction level may differ with the time elapsed from the beginning of the dialogue and with the stage of consensus building. In future research, we plan to study the factors affecting satisfaction level for not only dialogue among strangers but also among participants who have already built a personal relationship.

4. Future outlook

We discussed how to achieve mind-to-mind communications from two important viewpoints. From the viewpoint of what should be conveyed in what manner to convey sensitivity, we described our study on measuring skill level with respect to the receiver's subject matter, on switching to information deemed useful to the receiver depending on skill level, and making it easy to convey the sensitivity that a person wants to convey. To measure skill level, we examined a mechanism that uses sensing data centered around eye tracking and used measurement results to switch between skill levels and the information needed instead of relying on indeterminate elements such as the person's self-reports. From the viewpoint of determining (predicting) how a person feels as an outcome to mind-to-mind communications, we analyzed how a positivity score of dialogue participants could affect the degree of satisfaction with dialogue results. Our initial hypothesis was that dialogue participants with a positive attitude would be satisfied with dialogue results, but the experimental results fell short in terms of clearly indicating the relationship between a positive attitude and satisfaction. However, we are just at the beginning of achieving mind-to-mind communications technology and feel that these studies are a necessary part of the trial-and-error process.

We will continue with the studies presented in this article to achieve mind-to-mind communications technology through expressions tailored to others' sensitivities to demonstrate the true value of diversity and improve well-being. We look forward to more discussions and studies on mind-to-mind communications with all concerned.

References

[1] H. Matsubara, "Toward a System That Learns Something Forever," Journal of the Japanese Society for Artificial Intelligence, Vol. 18, No. 5, pp. 564–567, 2003 (in Japanese).

[2] R. Saijo, T. Sato, S. Eitoku, and M. Watanabe, "Method for Subtly Directing User's Gaze toward Information Breaking into Ongoing Activity," The Transactions of Human Interface Society, Vol. 23, No. 1, pp. 51–64, 2021 (in Japanese).

[3] M. Nakatani, Y. Ishii, A. Nakane, C. Takayama, and T. Hayashi, "Improving Participant Satisfaction and Quality of Output in Group Dialogue," Correspondences on Human Interface, Vol. 22, pp. 67–74, 2020 (in Japanese).

Ryohei Saijo

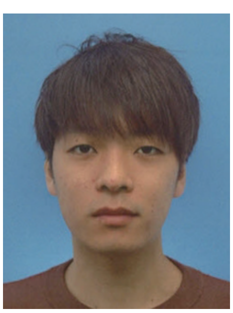
Researcher, NTT Digital Twin Computing Research Center.

He received an M.E. from University of Tsukuba, Ibaraki, in 2018 and joined NTT the same year. His research interests include human-computer interaction, information presentation methods, and human digital twin computing.

Yoko Tokunaga

Research Engineer, NTT Digital Twin Computing Research Center.

She received an M.E. from the graduate school of Kyoto University in 2010 and joined NTT the same year. Her research interests include conversation analysis and digital twin computing.



Daichi Yamaguchi

Internship student, NTT Digital Twin Computing Research Center.

He is currently studying at the graduate school of Osaka University. His research interests include multimodal interaction and sensing.

Lidwina Andarini

Researcher, NTT Digital Twin Computing Research Center.

She received an M.E. from Nara Institute of Technology in 2016 and joined NTT the same year. Her research interests include the human digital twin platform and computer networking.

Shohei Matsuo

Senior Research Engineer, NTT Digital Twin Computing Research Center.

He received an M.E. from the graduate school of Global Information and Telecommunication Studies, Waseda University, Tokyo, in 2006. He joined NTT in the same year and engaged in research and development of video coding and its standardization activities. His current research interests include video coding technologies, image processing, human visual systems, and digital twin computing.

Iwaki Toshima

Senior Research Engineer, NTT Digital Twin Computing Research Center.

He received an M.E. from the University of Tokyo in 2002 and joined NTT the same year. He received a Ph.D. in engineering from Tokyo Institute of Technology in 2008. His research interests include human-computer interaction, robot audition, and digital twin.

Takao Kurahashi

Senior Research Engineer, NTT Digital Twin Computing Research Center.

He received an M.E. in industrial engineering from Hosei University, Tokyo, in 2002 and joined NTT the same year. His research interests include human-computer interaction and robotics.

Shiro Ozawa

Senior Research Engineer, Supervisor, NTT Digital Twin Computing Research Center.

He received a B.E. and M.E. from Tokyo University of Mercantile Marine (now, Tokyo University of Marine Science and Technology) in 1997 and 1999. After joining NTT as a researcher, he worked for NTT Cyber Space Laboratories (now, NTT Human Informatics Laboratories). His fields of interest are computer vision, video communication, user interfaces, human-computer interaction, and 3D displays. He is a member of the Institute of Electronics, Information and Communication Engineers, the Institute of Image Information and Television Engineers, and the Institute of Image Electronics Engineers of Japan.

Technologies for Achieving Another Me

Atsushi Ohtsuka, Chihiro Takayama, Fumio Nihei, Ryo Ishii, and Toru Nishimura

Abstract

"Another Me," a grand challenge of Digital Twin Computing announced by NTT in 2020, aims to achieve the existence as a person's alter ego that appears to have the same intelligence and personality as an actual human and is recognized and active in society as that person. As the first step in achieving this, we constructed a digital twin that acts like the represented person and is capable of asking questions in line with the viewpoints held by that person. This article describes in detail the main technologies behind Another Me, namely, question-generation technology, body-motion-generation technology, and dialogue-video-summarization technology.

Keywords: Digital Twin Computing, grand challenges, Another Me

1. Introduction

The loss of various opportunities in one's life is becoming a social problem, such as the inability to participate in society despite an interest or desire to do so, possibly due to the difficulty of balancing childcare or caregiving of elderly family members with work. To dramatically increase the opportunities for a person to become more active and grow while expanding and merging the range of activities from the real world to the virtual world, we are taking up the challenge of achieving Another Me as a digital version of oneself [1, 2] (**Fig. 1**). Another Me will be able to participate actively as oneself in a way that transcends the limitations of the real world and share the results of such activities as experiences of oneself. There are three key technical issues surrounding this challenge: the ability to think and act autonomously as a human, the ability to have a personality like the represented person, and the ability to give feedback on the experiences obtained by Another Me. In this article, we describe in detail the main technologies for addressing these issues: question-generation technology, body-motion-generation technology, and dialogue-video-summarization technology.

2. Question-generation technology

To get Another Me to act autonomously, the digital twin must be able to make decisions about its next action. A means of collecting information as decision-making material is essential for this to be possible. We developed question-generation technology focusing on questions as a means of collecting information. Given input in the form of documents or text of conversations, this technology can automatically generate questions evoked by that input text. Generating questions and obtaining the answers to them in this manner enables a digital twin to autonomously collect information that it lacks.

Question-generation technology differs from current question-generation technology in two ways. First, the content of questions to be generated can be controlled by *viewpoint*. A person's values and position in life are greatly reflected by the questions that the person asks. Taking as an example an internal memo circulated within a company for decision-making purposes, we can expect the sales department to ask many questions about prices and costs, while a review by the legal department would likely include many questions related to legal compliance. By inputting a viewpoint label simultaneously with text,

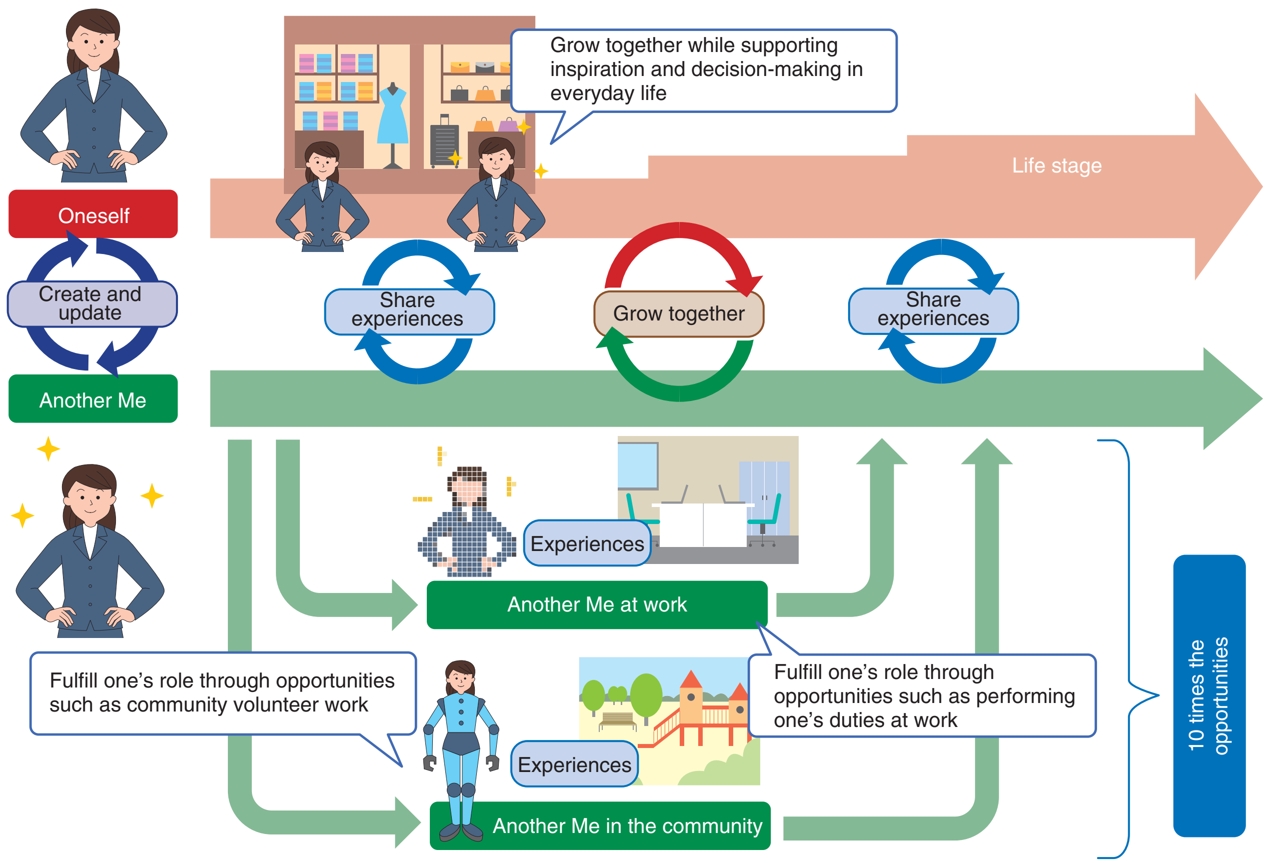


Fig. 1. Another Me.

the question-generation technology can generate questions corresponding to the input viewpoint. For example, inputting "money" as a viewpoint label will generate questions related to an amount of money as in costs, while inputting "law" as a viewpoint label will generate questions related to regulations and compliance. The viewpoint label to be input can be changed depending on the values, affiliated organization, etc. of a digital twin incorporating question-generation technology. This means that the digital twin can ask optimal questions according to its current thoughts and situation enabling it to collect the information needed to make decisions.

The question-generation technology can also recognize the input viewpoint and text content to automatically determine whether to generate questions. If content related to the input viewpoint has already been entered, or if an answer can be understood by simply reading the input text, there is a function for blocking the generation of questions. For example, when inputting the viewpoint label "money," questions related to money will be generated if there are no entries for price, cost, etc., but if there is a sufficient number of entries related to money such as price and cost, no questions will be generated. Therefore, when a digital twin is making decisions, it will ask questions only when having an insufficient amount of information and will quit asking questions and move on to processing for the next decision once it has collected a sufficient amount of information.

3. Body-motion-generation technology

To obtain the feeling that Another Me has the same personality as an actual person, its appearance is, of course, important, but making voice, way of speaking, and body motion like that person is also important. To date, we have clarified that differences in body motion, such as facial expressions, face and eye movements, body language, and hand gestures, can be felt as differences in personality traits [3] and that those differences can be a significant clue to

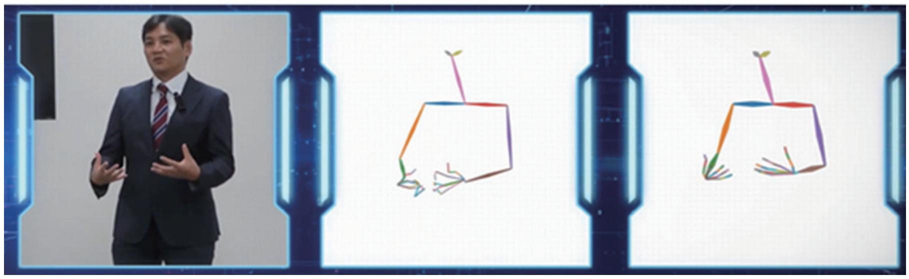


Fig. 2. Example input video of target person, results of generating body motion, and actual (correct) body motion from the input video.

distinguishing individuals [4].

Giving such body motions to an autonomous system in the manner of Another Me (for example, an interactive agent) and manipulating them accordingly is a difficult technical problem from an engineering standpoint. This problem has so far been addressed in technologies for generating human-like body motion and body motion corresponding to personality traits from utterance text [5, 6]. However, the generation of the same movements as those of a specific, actual person has not yet been achieved.

Against this background, we developed technology for automatically generating body motion like that of a real person on the basis of uttered-speech information in the Japanese language. We construct a model for automatically generating body motion like that of an actual person by preparing only video data (time-series data consisting of speech and images) of that person. We then use this model to automatically generate movements like those that person would make when speaking by simply inputting uttered-speech information. The first step for constructing the generation model involves using automatic speech-recognition (ASR) technology to extract utterance text from the voice data included in the video data of the target person when speaking. The positions of joint points on the person's body are also extracted automatically from the video data. The next step involves training the generation model by using a deep learning method called a generative adversarial network that can generate the positions of body joint points from voice and utterance text. To construct a model during this learning process that can capture even a person's subtle habits and generate a wide range of movements, we devised a mechanism for achieving high-level resampling of the data during training. As a result, we have achieved the highest level of performance in subjective evaluations with regard to a person's likeness and naturalness (as of October 2021) [7]. With this technology, we constructed a model for generating body motion using speech in the Japanese language as input. **Figure 2** shows an example input video of the target person, results of generating body motion, and actual (correct) body motion from the input video.

This body-motion-generation technology can automatically generate body motion of a particular person in Another Me, computer-graphics characters, humanoid robots, etc. It can also easily generate avatar body motions of a particular person in a web meeting from only uttered-speech information.

We plan to construct a model that can train a body-motion-generation model from a small amount of data and develop generation technology that can achieve a better likeness of an existing person.

4. Dialogue-video-summarization technology

Dialogue-video-summarization technology summarizes recorded dialogue over a length of time shorter than the actual meeting and generates video not only of the content of that dialogue but also of the atmosphere of the place where the dialogue occurred.

Our aim is to achieve a society in which humans and Another Me can grow together. It will not be sufficient to simply use Another Me as a surrogate for oneself—it will also be necessary to efficiently feed the experiences obtained by Another Me back to oneself. It will also be necessary to convey not only the content of what Another Me does but also the emotions likely to be felt at that time and at that place so that the represented person can treat what Another Me does as one's actions. We are researching dialogue-video-summarization technology targeting dialogue as one technology for feeding back the experiences of Another Me to the represented person.

As the first step in this research, we are taking up dialogue-condition-estimation technology and video-summarizing technology to enable efficient reviews of meetings. These combined technologies analyze and reconfigure the video of a meeting taken with a small camera or obtained from a web meeting and generate a video summary.

(1) Dialogue-condition-estimation technology

This technology uses a collection of information in various forms (multimodal information) such as the speech and behavior of the speaker including video of the dialogue as a clue to estimating diverse dialogue conditions. These include the importance and persuasiveness of each utterance, the intention and motivation of utterances, the personality traits and skills of individual participants, and the role of participants within the conversation [8–13]. The technology constructs a highly accurate estimation model by using a variety of machine-learning techniques. These include a multimodal deep-learning technique, which collects a variety of information such as the temporal change in the behavior of each participant, synchronization of movements among participants, change in the speaker's voice, and content of utterances for learning purposes, and a multitask-learning technique that simultaneously estimates multiple dialogue conditions.

(2) Video-summarizing technology

This technology uses the results of estimating various dialogue conditions obtained with the technology described above to extract important comments, questions asked of other participants, and comments made in reaction to other participants' opinions and reconfigure the meeting video to a shorter video of about one fourth the actual time. Subtle nuances of a participant's comments can be conveyed by the facial expressions or the tone of voice of that participant included in the summarizing video.

This dialogue-video-summarization technology enables a user to efficiently grasp in a short period the flow of a discussion, state of participants (e.g., an approving or opposing attitude to an opinion), and atmosphere (energy level) of the meeting that one was unable to attend. Our future goal is to summarize and convey not only dialogue among humans but also between a human and digital twin and even among digital twins. In the area of digital-twin behavior beyond dialogue, we aim to research techniques for feeding back information to humans in a more efficient manner with a higher sense of presence.

References

[1] Press release issued by NTT, "NTT Announces New R&D Projects of Digital Twin Computing," Nov. 13, 2020. https://group.ntt/en/newsrelease/2020/11/13/201113c.html

[2] R. Kitahara, T. Kurahashi, T. Nishimura, I. Naito, D. Tokunaga, and K. Mori, "Research and Development of Digital Twin Computing for Creating a Digitalized World," NTT Technical Review, Vol. 19, No. 12, pp. 16–22, 2021. https://doi.org/10.53829/ntr202112fa1

[3] Y. Nakano, M. Ooyama, F. Nihei, R. Higashinaka, and R. Ishii, "Generating Agent's Gestures that Express Personality Traits," The Transactions of Human Interface Society, Vol. 23, No. 2, pp. 143–154, 2021 (in Japanese).

[4] C. Takayama, M. Goto, S. Eitoku, R. Ishii, H. Noto, S. Ozawa, and T. Nakamura, "How People Distinguish Individuals from Their Movements: Toward the Realization of Personalized Agents," Proc. of the 9th International Conference on Human-Agent Interaction (HAI 2021), pp. 66–74, Nov. 2021.

[5] R. Ishii, R. Higashinaka, K. Mitsuda, T. Katayama, M. Mizukami, J. Tomita, H. Kawabata, E. Yamaguchi, N. Adachi, and Y. Aono, "Methods of Efficiently Constructing Text-dialogue-agent System using Existing Anime Characters," Journal of Information Processing, Vol. 29, pp. 30–44, Jan. 2021.

[6] R. Ishii, C. Ahuja, Y. Nakano, and L. P. Morency, "Impact of Personality on Nonverbal Behavior Generation," Proc. of the 20th ACM International Conference on Intelligent Virtual Agents (IVA 2020), Article no. 29, pp.1–8, Oct. 2020.

[7] C. Ahuja, D. W. Lee, R. Ishii, and L. P. Morency, "No Gestures Left Behind: Learning Relationships between Spoken Language and Freeform Gestures," Proc. of Findings of the Association for Computational Linguistics: EMNLP 2020, pp. 1884–1895, Nov. 2020.

[8] F. Nihei and Y. Nakano, "Validating the Effectiveness of a Meeting Summarizing Browser that Implements an Important Utterance Detection Model Based on Multimodal Information," The Transactions of Human Interface Society, Vol. 22, No. 2, pp. 137–150, 2020 (in Japanese).

[9] R. Ishii, K. Otsuka, S. Kumano, R. Higashinaka, and Y. Aono, "Estimation of Personal Empathy Skill Level Using Dialogue Act and Eye-gaze during Turn-keeping/changing," IPSJ Journal, Vol. 62, No. 1, pp. 100–114, 2021 (in Japanese).

[10] R. Ishii, S. Kumano, and K. Otsuka, "Estimating Empathy Skill Level Using Gaze Behavior Depending on Participant-role during Turn-changing/keeping," The Transactions of Human Interface Society, Vol. 20, No. 4, pp. 447–456, 2018 (in Japanese).

[11] T. Onishi, A. Yamauchi, A. Ogushi, R. Ishii, Y. Aono, and A. Miyata, "Analyzing Head and Face Behaviors along with Praising," IPSJ Journal, Vol. 62, No. 9, pp. 1620–1628, 2021 (in Japanese).

[12] R. Ishii, X. Ren, M. Muszynski, and L. P. Morency, "Multimodal and Multitask Approach to Listener's Backchannel Prediction: Can Prediction of Turn-changing and Turn-management Willingness Improve Backchannel Modeling?", Proc. of IVA 2021, pp. 131–138, Sept. 2021.

[13] R. Ishii, X. Ren, M. Muszynski, and L. P. Morency, "Can Prediction of Turn-management Willingness Improve Turn-changing Modeling?", Proc. of IVA 2020, No. 28, pp. 1–8, Oct. 2020.

Atsushi Ohtsuka

Research Engineer, NTT Digital Twin Computing Research Center.

He received an M.S. in informatics from the University of Tsukuba, Ibaraki, and joined NTT in 2013. He received a Ph.D. in informatics from the University of Tsukuba in 2020. His research interests are natural language processing and information retrieval. He has been in charge of research for Digital Twin Computing at NTT Digital Twin Computing Research Center.

Chihiro Takayama

Research Engineer, NTT Digital Twin Computing Research Center.

He received an M.S. in information engineering and M.A. in business from Waseda University, Tokyo, and joined NTT in 2009. His research interests include human-computer interaction. Since 2020, he has been in charge of research for Digital Twin Computing at NTT Digital Twin Computing Research Center.

Fumio Nihei

Researcher, NTT Digital Twin Computing Research Center.

He received a Ph.D. in science and technology from Seikei University, Tokyo and entered the university as a postdoc researcher in 2019. After working as a postdoc, he joined NTT in 2021. His research interests are multimodal interaction and social signal processing. Since 2021, he has been in charge of research for Digital Twin Computing at NTT Digital Twin Computing Research Center.

Ryo Ishii

Distinguished Researcher, NTT Digital Twin Computing Research Center.

He received an M.S. in engineering from the Tokyo University of Agriculture and Technology in 2008 and joined NTT the same year. He received a Ph.D. in informatics from Kyoto University in 2013. He was a visiting scholar at Carnegie Mellon University from 2019 to 2020. He is currently a distinguished researcher at NTT Human Informatics Laboratories. His research interests are multimodal interaction and social signal processing. He is a member of the Institute of Electronics, Information and Communication Engineers (IEICE), the Japanese Society for Artificial Intelligence (JSAI), and Human Interface Society.

Toru Nishimura

Senior Research Engineer, NTT Digital Twin Computing Research Center.

He received an M.S. in information engineering from Nagoya University in 1998 and joined NTT the same year. He has been involved in the research and development of an operation support system for IT systems including networks. Since 2020, he has been in charge of research for Digital Twin Computing at NTT Digital Twin Computing Research Center.

Initiatives toward Development of the Exploring Engine for the Future Society

Naoko Shigematsu, Atsushi Isomura, Isoo Ueno, Nobuhiro Oki, Yutaka Arakawa, and Kazuhiro Yoshida

Abstract

This article introduces the "Exploring Engine for the Future Society," one of the grand challenges toward actualizing the Digital Twin Computing concept and outlines its architecture and three key technologies.

Keywords: Exploring Engine for the Future Society, grand challenges, essential layer selection

1. Goals of the Exploring Engine for the Future Society

The sophistication of social structures and increasing number of natural disasters and other uncertain events have made it difficult to determine the impact of individual actions and cooperation on social and natural environments, as well as the resulting benefits to individuals. Therefore, to enable the exploration of social and lifestyle approaches for coexisting with nature, we are conducting research and development on the Exploring Engine for the Future Society. We are studying frameworks for using digital twins to accurately represent a society where people carry out their activities, as well as for exploring the future by enabling interactions between digital twins while making repeated changes in people's behavior.

To develop the Exploring Engine for the Future Society, we are aiming to incorporate not only the known interactions between objects, such as interactions in traffic flow, but also complex and unknown human interactions, such as those involving emotions, into digital twins. We are also conducting modeling that incorporates information propagation, crowd psychology, and collective dynamics. Using high-definition geospatial and sensing information, we are also developing simulation platform technologies for constructing digital twins of the real world in real time and simulations that are specific to individual models to visualize, for example, the future resulting from interactions between human flow and information propagation. These technologies drive the digital twins of society, enabling the presentation of various future scenarios as well as possible measures that should be taken now through backcasting from the ideal vision of society. We believe that various collaborations can be reinforced if residents of the local community, as well as government and corporate management, see the same future image through the visualization of the future society in different ways. We also aim to achieve a world that encourages voluntary activities and in which individuals can proactively think of and discover actions and collaborations for achieving the type of society they wish to build (**Fig. 1**).

2. Architecture of the Exploring Engine for the Future Society

As shown in **Fig. 2**, the Exploring Engine for the Future Society is linked with external systems, such as the 4D digital platform™ (high-definition digital platform), to obtain high-definition geospatial information and real-world sensing information. The

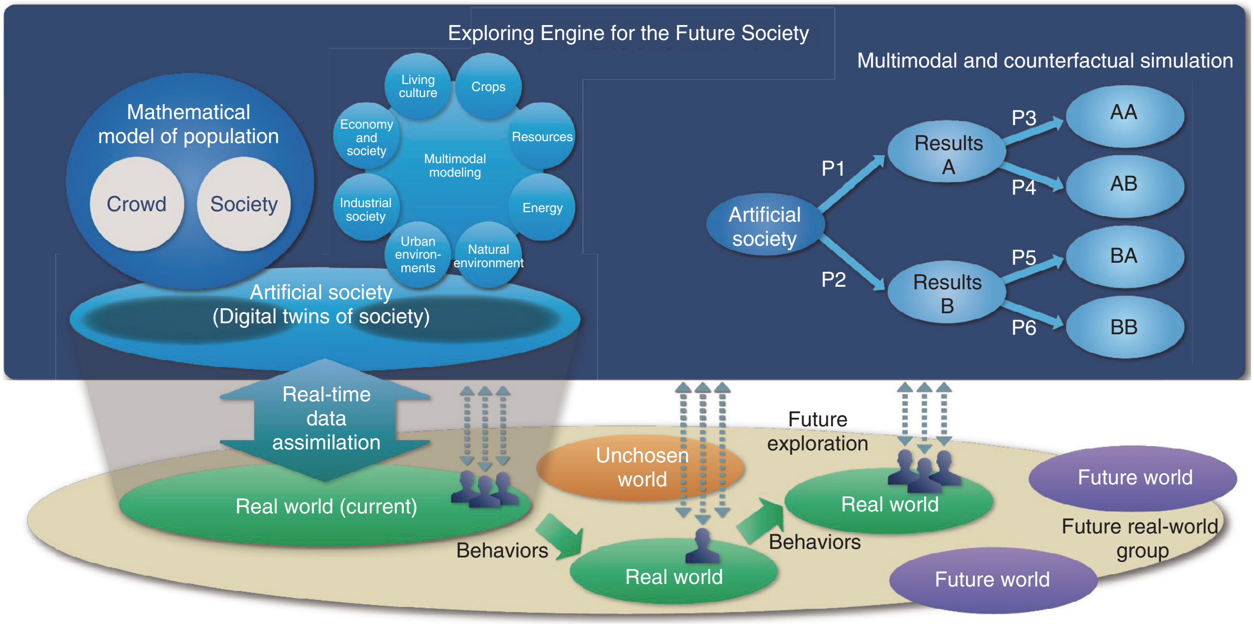


Fig. 1. Exploring Engine for the Future Society.

Exploring Engine for the Future Society is composed of two major fundamental technologies. The first is the digital-twin platform that provides a variety of data as components of a digital twin for building a virtual society. The second is the simulation platform that couples multi-layer simulations, such as for human flow, traffic flow, and information propagation, in accordance with the keywords that users input into the application search criteria for the future they envision. The platform, therefore, enables accurate capture of the changes in society where people carry out their activities, towards driving and visualizing a virtual society.

The digital-twin platform can be connected as needed to external high-definition digital platforms using interfaces such as the City Geography Markup Language (CityGML) and the Building Information Model (BIM). Digital twins for people, things, etc. are then constructed in real time on the basis of the data collected from the high-definition digital platforms. The data are made available as components of the digital twins, the information of which is updated synchronously with reality. We are considering using current metaverse software[*1], such as Autodesk products and Unity/Unreal Engine, to configure digital twins.

The simulation platform has a multi-layer simulation-execution unit containing individual simulation-execution units specific to individual models, such as human flow and traffic flow. The multi-layer simulation-execution unit of the simulation platform first obtains the necessary digital-twin components from the digital-twin platform depending on the keywords supplied by the application. The digital-twin components and several simulation-execution units are then used to couple simulations and calculate the interactions between various layers, such as human flow, traffic flow, logistics, weather, rivers, as well as information propagation through social networking services (SNS), resident behavior, and resident consciousness. The results of the linking and calculations are then returned to the application. Considering the interaction of various layers enables the calculation of, for example, the effects of human flow due to information propagation from a few minutes to a few years into the future.

3. Core technologies

Developing the Exploring Engine for the Future Society, which requires high-precision simulation of the future, is difficult to achieve with current technologies due to three major issues. First, although

*1 Refers to software for building virtual spaces in three dimensions.

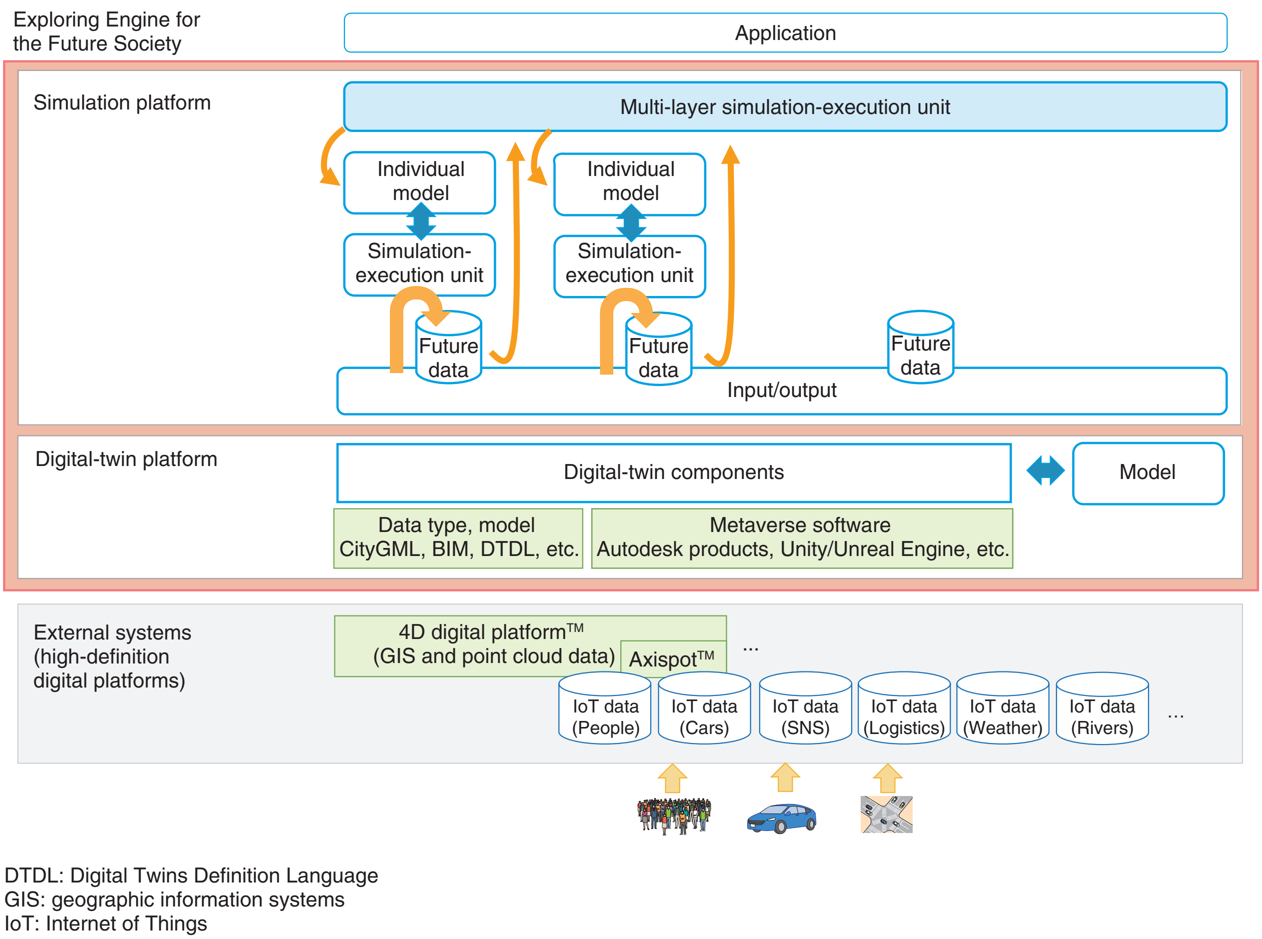


Fig. 2. Architecture of the Exploring Engine for the Future Society.

there is a wide variety of simulation technologies available, there is no technology that can select the best layer (or layers) from among the large number of simulation layers when carrying out flexible searches. Second, because the configurable parameters for current simulations are fixed before running simulations, it would not be possible to handle new parameters that must be considered if a sudden event occurs at some point in the future (catastrophes, major inventions, etc.). The third issue is that since typical simulations produce results by setting parameters and models first, they cannot carry out reverse simulation by setting the desired future society first then calculating and solving the intermediate process to arrive at that society. To address these three technical issues, NTT undertook the development of three technologies; essential-layer-selection technology, what-if simulation technology, and inverse simulation technology.

(1) Essential-layer-selection technology

This technology extracts only those layers that are highly relevant to the search terms from among the large number of layers that exist, depending on the search terms entered by the user. First, attributes such as “event,”*2 “location,”*3 “time,”*4 and “numerical value”*5 are extracted from the search terms. Next, “event,” “location,” and “time” are entered for each of the large number of layers provided by the Exploring Engine for the Future Society, and the impact of the output on “numerical value” is calculated. On the

*2 Refers to events (e.g., lockdown, change of prime minister, typhoon) subjected to simulation.
*3 Refers to the spatial range (e.g., Japan, Tokyo, 5 km^2 around Mitaka Station) for performing the simulation.
*4 Refers to the time range (e.g., after 3 minutes, after 1 month, after 10 years) for conducting the simulation.
*5 Refers to the number (e.g., infected persons, birth rate, demand for blankets) outputted as the result of the simulation.

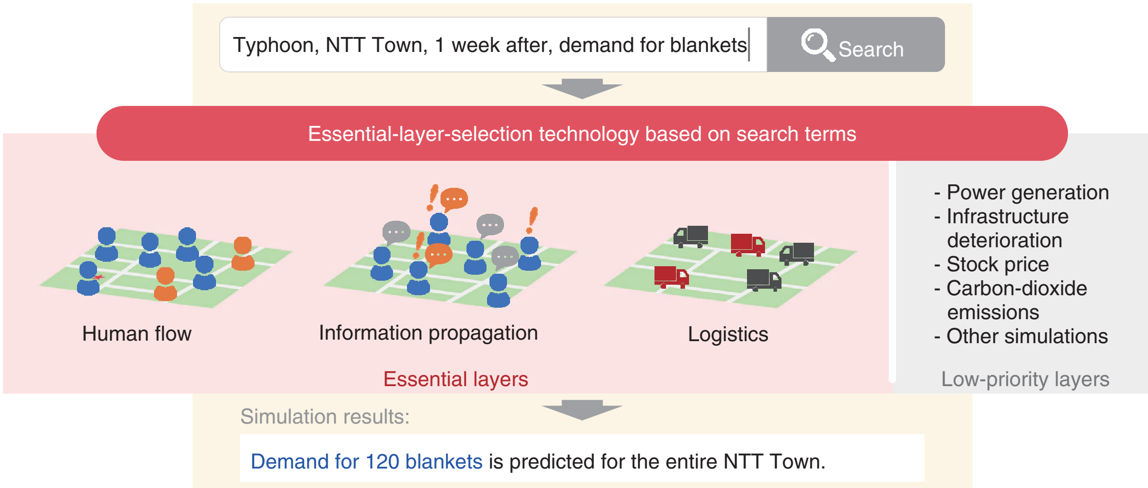


Fig. 3. Flow of presenting search results.

basis of this impact, the essential layers that should be used first are then selected and used to calculate the final simulation results.

For example, if the government wants to determine the "demand for blankets (numerical value)" in evacuation centers upon the occurrence of a large-scale typhoon. As shown in **Fig. 3**, when "typhoon," "NTT Town," "demand for blankets," "after one week" are entered as search terms, the human-flow layer for calculating human movement, the information-propagation layer for calculating information propagation through people's conversations and SNS, and the logistics layer for understanding the distribution status of blankets are selected. Layers other than these three are not selected because of their low relevance on "demand for blankets." Finally, the user is presented with the predicted results by using the selected layers.

(2) What-if simulation technology

This technology enables what-if calculations about "what happens if xx occurs?" for more complex events than what conventional simulation can handle. For example, if there is a breakthrough invention that allows humans to fly freely, new parameters and models need to be added to allow humans to move freely along the altitude direction (on the Z axis). NTT, therefore, aims to implement what-if simulations for sudden events that previously could not be handled, by automating i) the generation of parameters and models that otherwise could not be estimated before simulation and ii) the setting of the range of possible parameter values.

(3) Inverse simulation technology

This technology enables inverse simulations about "what actions should be carried out to create a world like xx?" for more complex conditions than what conventional simulation can handle. Exploring the ways to build the envisioned future society requires a consideration of the comparative impact of different areas, such as environment, economy, culture, healthcare, human flow, transportation, and events. For example, creating a society that is capable of achieving zero deaths from COVID-19 requires calculating interactions of multiple areas outside healthcare, such as the economy, human flow, and the status of operations of business establishments. NTT, therefore, aims to carry out inverse simulation that can take into account interactions among these different areas.

4. Future developments

Going forward, we will develop modeling technology by incorporating crowd psychology and collective dynamics in addition to layers with known interactions such as human flow and traffic flow. We will develop technologies for driving and predicting a future society through the coupling of multi-layer simulations.

We will pursue further developments to visualize a future society by considering the application of these technologies in the simulation of evacuation experiences for local governments and residents. We will also consider enabling the selection of the best measures to achieve regional reconstruction and infrastructure rehabilitation for local governments and infrastructure operators engaged in the reconstruction of disaster-hit areas.

Naoko Shigematsu
Research Engineer, NTT Digital Twin Computing Research Center.

She received a B.S. and M.S. in geophysics from Tohoku University, Miyagi, in 1993 and 1995. In 1995, she joined NTT Telecommunication Networks Laboratories, and in 1999 she moved to NTT EAST R&D Center, where she engaged in research of asynchronous transfer mode network operation systems. In 2000, she moved to NTT Information Sharing Platform Laboratories, where she engaged in research of storage area networks. Since 2020, she has been in charge of research for Digital Twin Computing at NTT Digital Twin Computing Research Center.

Atsushi Isomura
Researcher, Digital Twin Computing Project, NTT Digital Twin Computing Research Center.

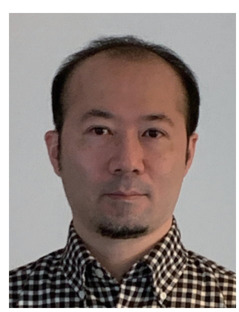
He received a Master's degree in information science and technology from Aichi-Prefectural University in 2016. He joined NTT Software Innovation Center the same year and developed Axispot™, a distributed spatio-temporal database management technology. He is currently working on digital-twin data-management technology at Digital Twin Computing Research Center. He received the Institute of Electronics, Information and Communication Engineers (IEICE) Tokai Section Student Award in 2015 and the best poster award in 12th International Conference on Ubiquitous Healthcare. He presented at RedisConf19 held in San Francisco.

Isoo Ueno
Research Engineer, NTT Digital Twin Computing Research Center.

He received a B.S. and M.S. in earth science from Kobe University, Hyogo, in 1990 and 1992. He joined NTT in 1992 and engaged in basic research on complex systems. In 1998, he launched Internet businesses in Hong Kong and UK. He also managed the development projects of OCN web hosting, Biz authentication, and the global billing system of NTT Communications. In 2021, he developed the polygon-based data aggregation technology of Axispot™. He is currently working on Digital Twin Computing.

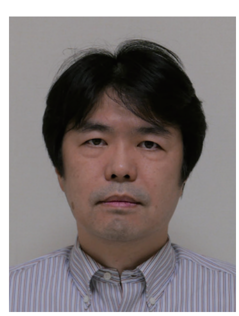
Nobuhiro Oki
Senior Research Engineer, NTT Digital Twin Computing Research Center.

He received a B.E. and M.E. in chemistry from Tokyo Institute of Technology in 1994 and 1996. He joined NTT in 1996 and launched the Internet telegram system (D-Mail). In 2008, he joined NTT Communications and launched various cloud services, including Smart Data Platform (formerly called Enterprise Cloud). In 2018, he joined NTT laboratories and developed Axispot™. He is currently working on digital-twin data-management technology.



Yutaka Arakawa
Senior Research Engineer, NTT Digital Twin Computing Research Center.

He received a B.E. and M.E. in electronic engineering from the University of Tokyo in 2001 and 2003 and Ph.D. in engineering from the University of Electro-Communications, Tokyo, in 2014. He joined NTT in 2003. His current research interests include distributed systems and data structure design. He is a member of IEICE.



Kazuhiro Yoshida
Senior Research Engineer, NTT Digital Twin Computing Research Center.

He received a B.S. and M.S. in engineering from Tokyo Institute of Technology in 1994 and 1996 and joined NTT in 1996. Since 2021, he has been in charge of research on Digital Twin Computing at NTT Digital Twin Computing Research Center.

Coupled Simulation Technology for Visualizing Environmental and Economic Social Cycles

Masahiro Maruyoshi, Yuichi Muto, and Daisuke Tokunaga

Abstract

Maintaining sustainability of human society in harmony with the environment into the future is regarded as the most-important issue on Earth. Accordingly, the NTT Digital Twin Computing Research Center is taking up the grand challenge of inducing inclusive equilibrium solutions for the Earth and its social and economic systems by making the interrelationships between them as computable as possible. Our efforts to meet this challenge are described in this article.

Keywords: climate and economic society, coupled simulation, circulation system

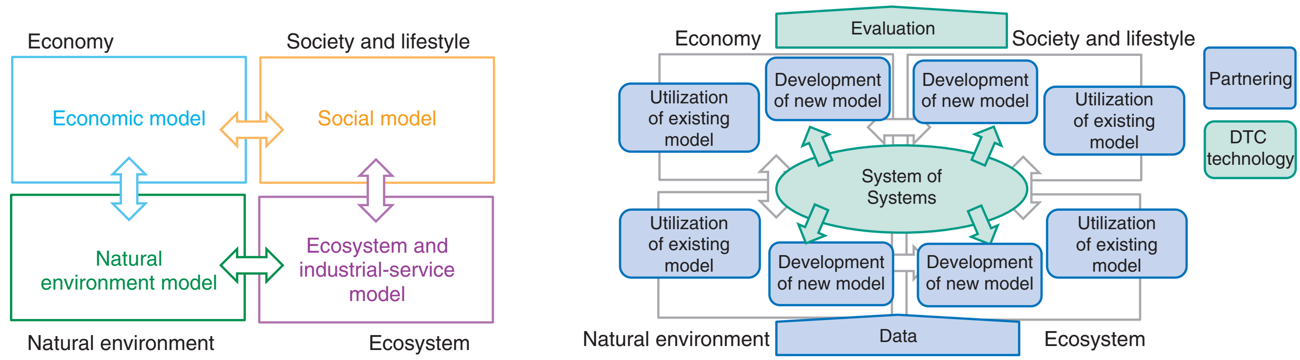
1. Inclusive sustainability of the environment and socioeconomics

Against the background of the social impact of recent global climatic and environmental changes (such as the increase in global heat waves, sudden torrential rain, droughts and wildfires, and changes in ecosystems symbolized by increases in numbers of endangered species and changes in crop production), discussions on the reconstruction of social systems to maintain the Earth's environment have begun.*1 The Earth's environment has its own autonomy. Problematic environmental changes are a result of the influence of human economic and social activities on this autonomy. This complex chain reaction can be viewed as an inclusive *circulation system* that includes the environment, economy, and society on a global scale.

Discussions on impact assessment of climate change*2 and reconstruction of social systems are supported by science and technology, which experts in various fields have been improving over the years. Scientific approaches in various fields have made it possible to analyze in detail phenomena that are highly uncertain and difficult to observe. Integrated assessment modeling (IAM) provides direct impact assessments spanning various fields (such as the environment and economy) by using macroscopic and statistical indicators. It is necessary to take an inclusive view of the Earth as a circulation system rather than considering each problem individually to solve various problems related to the environment and people's lives on a global scale toward the future. We believe that the development of computational technologies that enable the construction of models with high-accuracy and cross-field collaboration will accelerate. At the NTT Digital Twin Computing Research Center, we set the goal of contributing to achieving inclusive sustainability of the environment

*1 Example discussion on reconstruction of social systems: The World Economic Forum called for a "Great Reset" as its theme for 2021. Although this theme was set forth in the wake of the COVID-19 pandemic, discussions have started with an awareness of environmental and ecological issues such as global climate change.

*2 Example assessment of impact of climate change: The Intergovernmental Panel on Climate Change (IPCC) released the Working Group I (WGI) report of the Sixth Assessment Report stating that "There is no doubt about the impact of human activities on climate change." WGII and WGIII are scheduled to report on impact assessment and adaptation measures.



(a) Simple model of inclusive circulation system

(b) Partnering and research subjects

DTC: Digital Twin Computing

Fig. 1. Partnering strategies for realizing the grand challenge.

and socioeconomics by promoting the development of computational technologies and making it possible to model an inclusive global circulation system.

Our strategy for achieving this goal is conceptualized in **Fig. 1**. It is assumed that the inclusive circulation system is that shown in Fig. 1(a). Economic impact on the environment affects industry and ecosystem services (e.g., food production and resource circulation), and changes in those conditions affect social aspects (e.g., health, inequality, and well-being). These changes in social aspects create the power to accelerate and decelerate the economy and change the state of the economy, and that economic change further propagates the impact on the environment. These changes can be viewed as a system that circulates inclusively in this manner. As mentioned above, experts in each field comprising the system have spent many years developing models and simulators that analyze and predict phenomenon using a scientific approach. We aim to construct a circulation system by connecting these models and simulators, which are becoming ever more accurate. Therefore, we believe that computational technologies that can determine the causality of these models, adjust the scale differences, and evaluate the stability of the circulation system are important (Fig. 1(b)). In this article, the challenges related to the system of systems (SoS), which is an important computational technology, and the status of developing a proof of concept (PoC), which is our initial effort to evaluate the circulation system, are discussed.

2. Heterogeneous-simulator coupling technology for reproducing the chain between the environment and society

The environment and socioeconomic activities have each been modeled on the basis of past-to-present data obtained from many years of observations, and those models have evolved as simulation technologies for predicting the future from the present. However, precisely predicting the relationship between the actual global environment and people's economic activities is difficult by only simulating a single domain because various domains are intricately connected.

A similar problem exists in the field of industrial production, and is in the process of being solved with an SoS by combining several single-domain simulation techniques. For example, coupled simulation technology has been applied to the design of automobiles and space satellites, which consist of multiple systems (such as engines and transmissions) supplied by different vendors. An *emergent property* is a property that an entity gains when it becomes part of an SoS. Such properties are also true for visualizing environmental and economic social cycles to discover emergent properties through simulations as an SoS. However, it has not yet been put to practical use because of the huge amount of computation and problems related to the coupling between simulators when the global environment and people's economic activities are involved.

3. Three challenges concerning coupling of heterogeneous simulators and NTT's efforts

To develop a heterogeneous simulator that can reproduce the chain linking the environment and society, we are researching and developing ways to address the following three major challenges.

3.1 Challenge 1: Coupling of heterogeneous simulators

As technologies for coupling multiple types of simulators, under the guidance of the US Department of Defense, High-Level Architecture (HLA) was standardized by the Institute of Electrical and Electronics Engineers (IEEE) as IEEE1516 [1] and Functional Mock-up Interface (FMI) [2] was standardized by Daimler AG. HLA is often used to simulate discrete events such as traffic congestion, evacuation guidance, and development of space satellites, while FMI is often used to simulate continuous events such as coupled control of various devices in a running car. To reproduce the chain linking the environment and society, it is necessary to (i) incorporate discrete events into continuous events and (ii) develop execution-control technology to couple both events. Moreover, as the number of simulators to be coupled increases, it becomes more difficult to analyze problems related to processing speed and simulation accuracy. In addition, HLA and FMI do not have functions for executing analysis, so users need to monitor and analyze performance information themselves.

3.2 Challenge 2: Resolving differences in resolution of processing data between different simulators

When simulating phenomena occurring on the Earth's surface, for ease of computation, it is common to process them on a unit-area and unit-time basis. To do this processing, we build a mathematical model that can be solved with the data obtained in the past. However, the spatial resolution (unit area), temporal resolution (unit time), and time period to be covered will vary in accordance with the data to be handled. It is therefore necessary to unify the differences in the spatial and temporal resolutions of the simulators when they are coupled. To unify these differences, it is easy to convert the fine-resolution data into coarse-resolution data by averaging, etc. However, simulation results with the desired resolution cannot be obtained with the coarse data. When only single-domain results are available from simulations based on coarse data, it is necessary to downscale to finer resolution on the basis of physical (mechanical) laws or statistical and empirical relationships.

3.3 Challenge 3: Reducing the overall computational complexity of the simulation

Simply coupling simulations in areas such as the atmosphere, climate, and oceans, for which highly accurate predictions are becoming possible, would result in an explosion of computational complexity that cannot be handled in a practical amount of time. To solve this problem, it is necessary to develop a technology for model transformation involving (i) creating a surrogate model by applying machine learning, deep learning, etc. using observation data and prediction data obtained from the original simulation as learning data and (ii) reducing the amount of calculation when executing a simulation.

4. Developing a PoC as an initial approach to evaluating circulation systems

Using the above-mentioned technologies, we are also developing a PoC for simulation of environmental and socioeconomic linkages. Through these efforts, we will communicate our values, i.e., achieving inclusive sustainability of the environment and socioeconomics, while exploring new research topics by implementing and evaluating our technologies.

A large number of models are involved in reproducing global-scale chains. To start formulating the PoC, we first focused on the water cycle, which is closely related to both the natural environment and social economy. The water cycle is a phenomenon, the value of which remains the same regardless of scale and appears on a wide range of scales, from cycles limited to countries and basins to global cycles such as the entire globe.

We implemented a circulation simulation of a water cycle and water use by implementing a prototype execution control of the following three simulators using the coupled technology of heterogeneous simulators described in Challenge 1 (**Fig. 2**).

- Meteorological simulator (simulating simulator output using rainfall data in AMeDAS[*3])
- River simulator (MIKE)
- Economic-water-use simulator (simple uniquely implemented model)

By coupling meteorological and river simulations,

*3 AMeDAS: Automatic Meteorological Data Acquisition System. A weather observation system developed by the Japan Meteorological Agency.

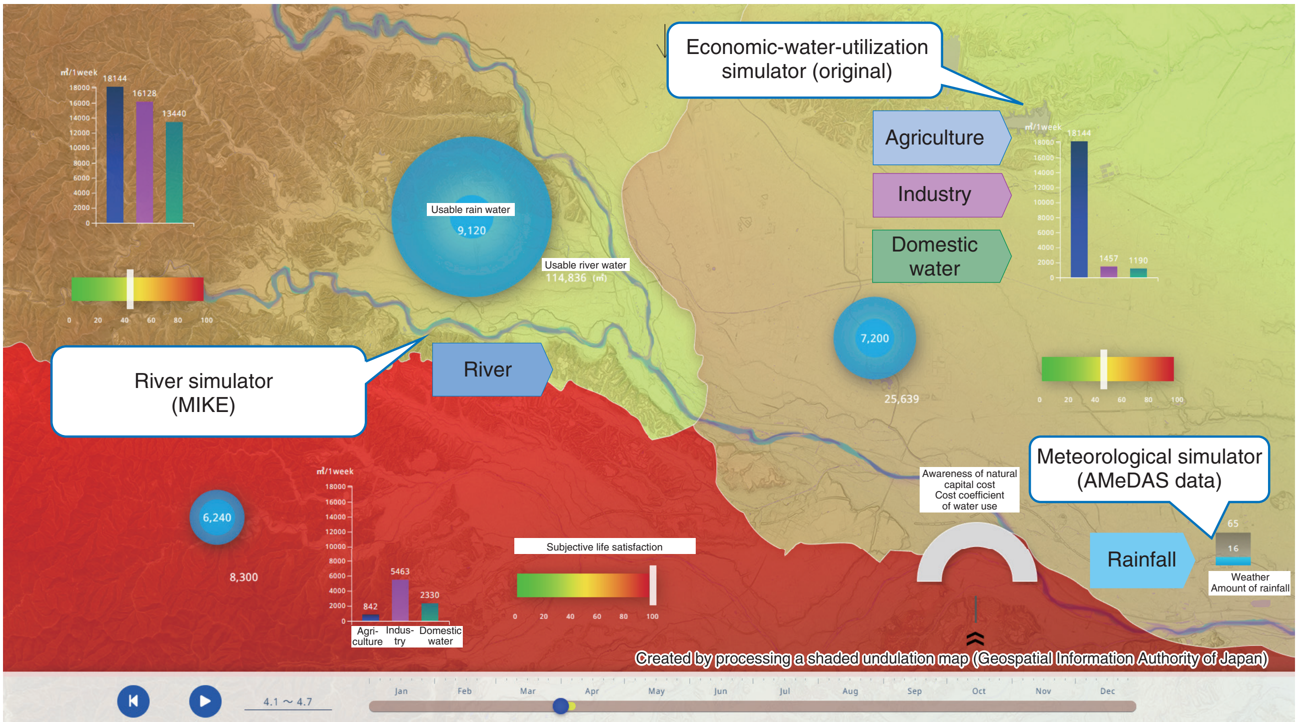


Fig. 2.   Example circulation simulation of water cycle and water use.

we can reproduce the amount of river water on the basis of the amount of rainfall that flows into the river. By coupling economic water-utilization simulations, we were able to reproduce a water cycle that takes into account water taken from rivers for economic activities (agriculture, industry, and domestic water), consumed, and returned as reusable water to the rivers.

To reproduce rainfall, change in the volume of river water, water consumed by humans, and temporal water environmental load, it is necessary to resolve the difference in resolution of the data handled by each simulator. With this PoC, the spatial and temporal differences described in Challenge 2 appear as differences in the scales and unit time of water-volume data for water cycle and water use. We handled such differences by implementing the conversion process and optimizing the simulation settings.

For example, rainfall data output by meteorological simulations are given in units of amount of water per mesh, while river inflow data required for input to river simulations is given in units of amount of water inflow per watershed area. Although both units represent water volume, they have different spatial resolutions. For our PoC, we are trying to convert rainfall data into inflow-water volume on the basis of the watershed area and inflow efficiency of the region.

In terms of temporal resolution, it is also necessary to match the temporal units of water volume exchanged by rainfall, rivers, and water use. For this PoC, we designed and coupled all the simulations to operate at the temporal resolution of water volume per hour. We are planning to design downscaling technology that will make it possible to handle such differences in resolution.

There is currently no limit placed on the consumption of water in our daily activities, and that unlimited use puts a burden on the water cycle. We developed a model that assumes a society that adopts the concept of *natural capital cost*, which considers the consumption of natural resources such as water as a reasonable cost in the resource cycle. With this model, we confirmed that it is possible to simulate the effects of changes in water use and the load on the water cycle by changing the intensity of awareness of natural capital costs.

The PoC we are currently developing has implemented a specific basin-type water-cycle model. It focuses on water resources as a natural environmental system and water use as a socioeconomic activity. To reproduce the global cycle, it is necessary to incorporate and integrate various models of the natural environment and socio-economy. Accordingly, we will continue our research and development by forming partnerships with various experts and external organizations.

References

[1] IEEE 1516-2010: IEEE Standard for Modeling and Simulation (M&S) High Level Architecture (HLA)-- Framework and Rules.

[2] FMI, https://fmi-standard.org/

Masahiro Maruyoshi

Senior Research Engineer, Supervisor, NTT Digital Twin Computing Research Center.

He received a B.E. and M.S. in information engineering from Nagoya University, Aichi, in 2000 and 2002. He joined NTT in 2002 and was engaged in research and development of software-defined network technologies. Since 2020, he has been in charge of research for Digital Twin Computing at NTT Digital Twin Computing Research Center.

Daisuke Tokunaga

Senior Research Engineer, NTT Digital Twin Computing Research Center.

He received an M.S. in information science from Kyushu Institute of Technology, Fukuoka, in 1999 and joined NTT the same year. His research interests include large-scale application system technologies. Since 2020, he has been in charge of research on Digital Twin Computing at NTT Digital Twin Computing Research Center.

Yuichi Muto

Senior Research Engineer, NTT Digital Twin Computing Research Center.

He received a B.E. and M.S. in information science from Tohoku University, Miyagi, in 1998 and 2000. He joined NTT in 2000 and was engaged in research and development concerning the Communication Traffic & Service Quality Project at Service Integration Laboratories. Since 2021, he has been in charge of research for Digital Twin Computing at NTT Digital Twin Computing Research Center.

Experimental Evaluation of High-capacity Wireless Transmission Using Orbital Angular Momentum Multiplexing Technology

Doohwan Lee, Yasunori Yagi, Hirofumi Sasaki, and Hiroyuki Shiba

Abstract

We are researching and developing technology to achieve terabit-class wireless transmission toward six-generation wireless systems. In this article, we present experimental evaluations of high-capacity wireless transmission using orbital angular momentum (OAM) multiplexing on a 40-GHz frequency band with a bandwidth of 1.5 GHz at distances of 100 and 200 m. Our OAM antennas have two uniform circular arrays (UCAs) with the same diameter using different linear polarizations and Butler matrices, which are analog devices for generating and separating OAM modes. OAM beams are generated using a Butler matrix, which is an analog radio-frequency circuit, and radiated from the UCAs. We implemented an OAM multiplexing transmitter (Tx) and receiver (Rx) on the 40-GHz band. The Tx and Rx can generate and separate seven OAM modes (0, ±1, ±2, ±3) and two polarizations (vertical and horizontal), which is a total of 15 streams including the center antenna element to transmit OAM mode 0. Each channel carries a 1.5-Gbaud signal. We experimentally evaluated the transmission capacity at distances of 100 and 200 m. The results indicate successful wireless data transmission of 137 Gbit/s at a distance of 100 m and 117 Gbit/s at a distance of 200 m.

Keywords: orbital angular momentum (OAM), point-to-point, line-of-sight, uniform circular array, 40-GHz band

1. Introduction

The demand for achieving high-capacity wireless transmission is increasing due to exponentially increasing data traffic toward six-generation (6G) wireless systems [1]. The use of wireless transmission is expected to increase in all fields including connected cars, virtual reality/augmented reality, three-dimensional cameras, and high-definition video transmission. Considering such an increase, it is anticipated that the hundreds-of-gigabit level to terabit level wireless transmission will be necessary in the 2030s [2–4].

There have been many reports on high-capacity wireless transmission in orbital angular momentum (OAM) and multiple-input multiple-output (MIMO)/single-input single-output (SISO) multiplexing [5–17], as shown in **Fig. 1**. We provide a summary of these reports below. There have been reports on experiments over 100 Gbit/s in short distances (<10 m) in laboratory environments, and approximately 100 Gbit/s in outdoor environments (≥100 m). In terms of transmission capacity, 1.056-Tbit/s wireless transmission over 3.6 m was reported by GTEC, Fudan University using 4×4 MIMO on D-band [5]. In terms of transmission distance, Northrop Grumman,

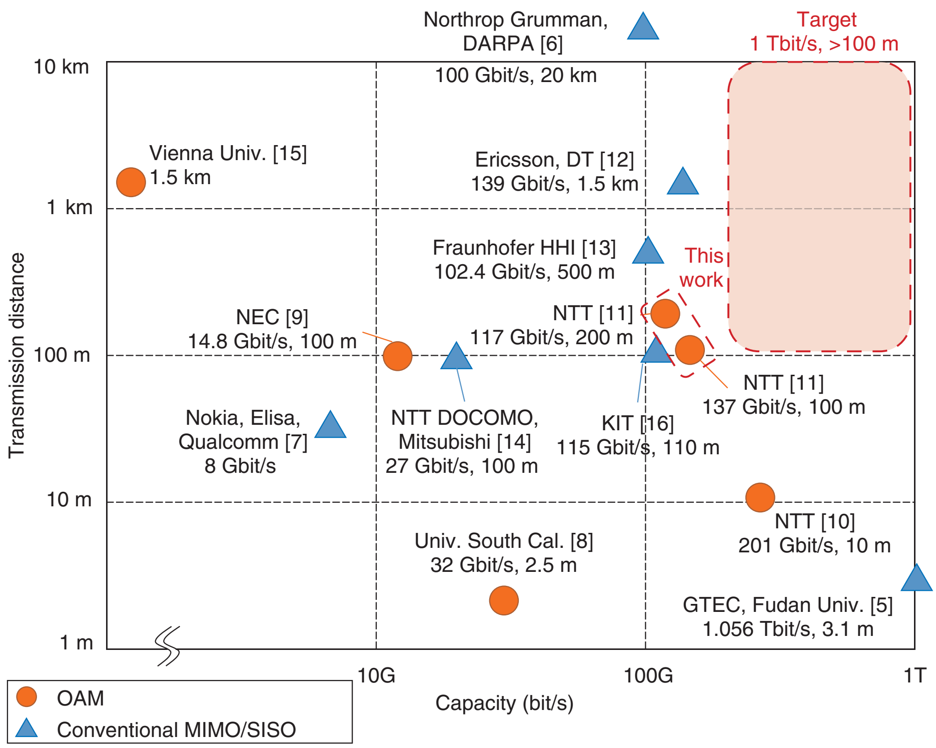


Fig. 1. Reports of experimental demonstration on high-capacity wireless transmission using conventional MIMO/SISO and wireless OAM transmission.

DARPA [6] reported 100-Gbit/s wireless transmission (total rate of transmission and reception) at a distance of 20 km. The result from 5G commercial networks was the 8-Gbit/s transmissions reported by Nokia, Qualcomm, and Elisa [7]. Regarding OAM multiplexing, 32-Gbit/s transmission on the 60-GHz band over 2.5-m distance was reported by University of Southern California [8], and 14.8-Gbit/s transmission on the 150-GHz band over 100 m using real-time signal processing was reported by NEC [9]. Considering beyond 5G and 6G wireless communication, we set our target to 1 Tbit/s and over 100-m transmission.

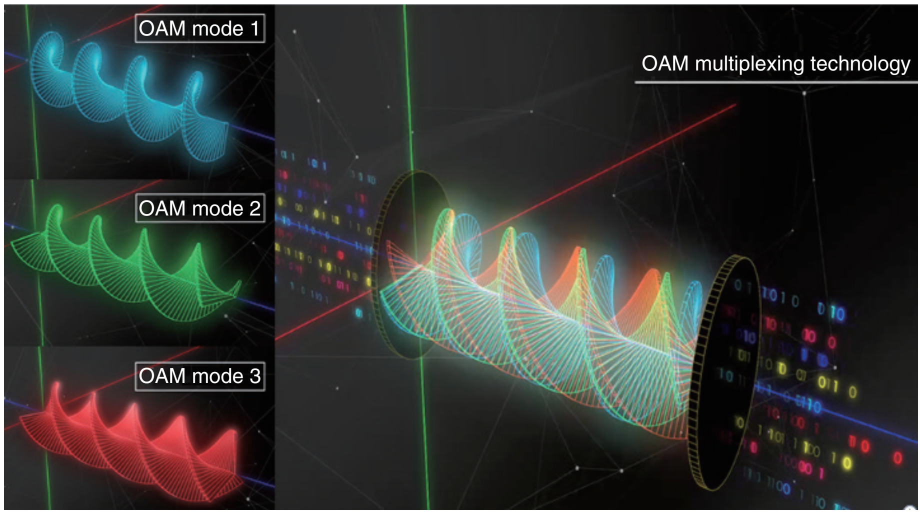
To achieve such terabit-class wireless transmission, we are researching OAM multiplexing transmission technology as a new spatial multiplexing technology that enables high-capacity wireless transmission. We previously reported on experiments of over 200-Gbit/s transmission using OAM-MIMO multiplexing on the 28-GHz band at a distance of 10 m in a shielded room [10]. We further achieved wireless data transmission of 137 Gbit/s at a distance of 100 m and 117 Gbit/s at a distance of 200 m [11] using OAM multiplexing on the 40-GHz band. The remainder of this article reports on the principle of OAM multiplexing and our recent research.

2. OAM multiplexing technology

2.1 Principle

OAM is a physical property of electro-magnetic waves characterized by a helical phase front in the propagation direction. Since these characteristics can be used to create multiple orthogonal channels, wireless communication using OAM can increase radio-spectrum efficiency and expand the capacity of the scarce radio spectrum. The number of phase rotations is called an OAM mode. Different data signals can be transmitted at the same time by transmitting different signals using radio waves having different OAM modes [8].

Given an electromagnetic wave having this OAM property, the trace of the same phase takes on a helical shape in the direction of propagation (**Fig. 2**). For example, Figure 2(a) illustrates the trance of the same phase in OAM modes 1, 2, and 3. Radio waves having this OAM property cannot be received without a receiver (Rx) having the same number of phase rotations at the time of transmission. For this reason, if



(a) Trace of the same phase (b) Concurrent transmission

Fig. 2. Principle of OAM multiplexing transmission technology.

multiple radio waves having different OAM modes are superposed, they can eventually be separated without mutual interference as long as an Rx is prepared that can receive radio beams with the same phase rotations corresponding to each OAM mode. Figure 2(b) shows an example in which different signals are put on OAM modes 1, 2, and 3 and transmitted simultaneously. Technology for transmitting multiple data signals using this feature is called OAM multiplexing transmission technology.

2.2 Generation and separation

To generate a beam carrying the OAM mode n (L = n), antenna elements are connected with phase shifters that make n × 360 degrees of rotations. **Figure 3(a)** shows an example of beam generations of OAM modes 0, 1, and 2 using a uniform circular array (UCA) consisting of 8 antenna elements. Note that concurrent transmission of multiple OAM modes can be achieved by superposing multiple OAM signals, as in **Fig. 3(c)**.

The separation of beams carrying OAM modes can be done in a similar manner as generation using antenna elements connected with phase shifters that make opposite direction of rotations. As far as the number of antenna elements are larger than 2n, n × 360 degrees of rotations are orthogonal one another. Therefore, the separation of each OAM mode from a mixed OAM mode signal can be done without aliasing. **Figure 3(b)** shows an example of the phase of each antenna element corresponding to the example above.

2.3 Transmission and reception

The beams with different OAM modes can be approximately generated or separated simultaneously using a UCA with a discrete Fourier transform (DFT). They are orthogonal as long as they propagate coaxially. Mathematically, the channel between UCAs is a circulant matrix, so the channel can be diagonalized using a DFT matrix as the singular value decomposition (SVD) method, meaning oppositely arranged transmitting and receiving UCAs can create several isolated channels regardless of the propagation distance. Theoretically, the OAM has an infinite depth of field, but the discretization of space sampling limits the maximum number of available OAM modes when using a UCA and DFT processor. The maximum number is equal to the number of antenna elements in a UCA. When using UCAs for transmitting and receiving antennas, the beam propagation with OAM modes can be analyzed using the standard narrow band MIMO theory.

We now discuss the theoretical background of an OAM multiplexing system with UCAs. **Figure 4** shows a schematic chart of OAM multiplexing transmission. When UCAs are placed opposite each other, the antenna-array configuration is axisymmetric so that the channel matrix becomes circulant. When the number of antenna elements in a UCA is defined as *n*,

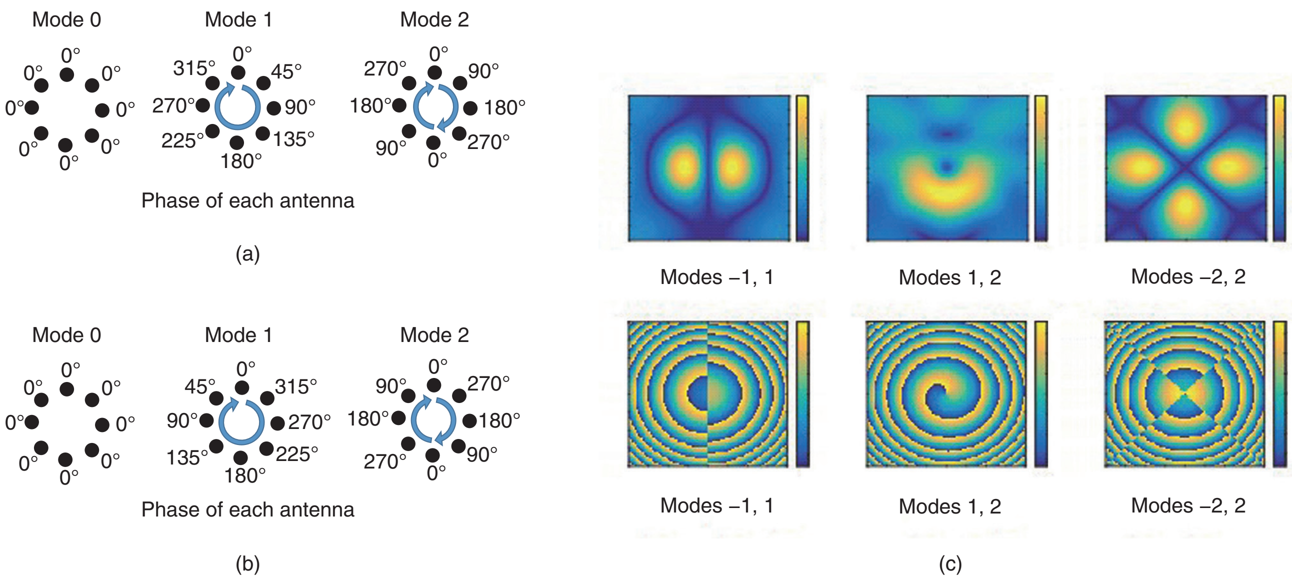


Fig. 3. (a) Generation of OAM modes with a UCA, (b) separation of OAM modes with a UCA, (c) intensity distribution (above) and phase distribution (below) of combined OAM modes (L is the number of OAM modes).

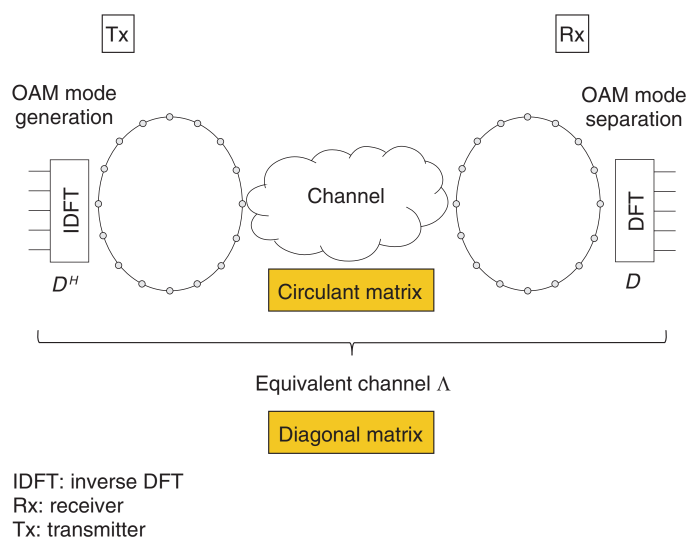


Fig. 4. Schematic of OAM multiplexing system.

the circulant matrix is diagonalized using the $n \times n$ DFT matrix

$$D_n\,(x, y) = \exp\left(-j\left(\frac{2\pi(x-1)}{n}\right) l(y)\right) / \sqrt{n}\,, \qquad (1)$$

as

$$H = D_n \Sigma D_n{}^H, \qquad (2)$$

$$\Lambda = \Sigma = \begin{Bmatrix} \lambda_{l(1)} & & & 0 \\ & \lambda_{l(2)} & & \\ & & \ddots & \\ 0 & & & \lambda_{l(n)} \end{Bmatrix}, \qquad (3)$$

where l denotes OAM mode corresponding to the yth

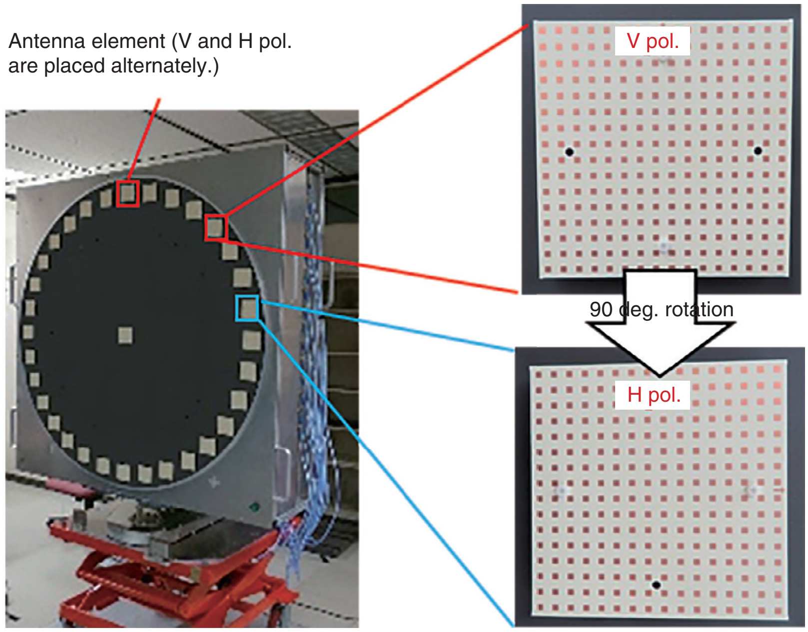


Fig. 5. Prototype of OAM transmission equipment.

eigenvector, H and Σ denote the channel matrix between the receiving and transmitting UCAs and diagonal matrix, respectively, and Λ denotes the equivalent channel matrix including DFT processing, as shown in Fig. 4.

One of the strengths of OAM multiplexing is that it only requires static processing and DFT, so we can use analog devices for OAM-mode generation and separation without digital signal processing. A Butler matrix is an analog circuit that can process DFT calculations with hybrid circuits and phase shifters, so we used it for OAM-mode generation and separation in our system [11, 17].

3. Transceiver design

We designed a 40-GHz-band OAM multiplexing transceiver to conduct outdoor experiments. **Figure 5** shows this transceiver and its antenna elements. The diameter of the UCA is 120 cm, and a total of 33 antenna elements (16 elements each for vertical (V) and horizontal (H) polarization UCA and the center) are placed on the circumference. The antenna size depends on the frequency, as shown in **Table 1**. The size can be reduced when higher frequency bands are used. The gain of the antenna element is 25 dBi. This antenna element can radiate V or H polarization by rotating 90 degrees. One antenna element is placed on the center of a UCA to transmit mode 0 at the singular points of the OAM mode.

Table 1. Equivalent antenna sizes.

Frequency band [GHz]	40	90	150	300
Antenna size [mm]	1200	800	620	440

Figure 6 shows a block diagram of the OAM-transmission experiment. Baseband (BB) signal waveforms generated by Matrix Laboratory were input to the transmitter (Tx) using arbitrary waveform generators (AWGs). Analog BB signals were up-converted to radio frequencies (RFs) of 39.5 to 41 GHz signals in the Tx. Finally, 7 OAM modes signals were generated and multiplexed using Butler matrices in each polarization and transmitted from the 16 antenna elements. The mode isolation of these Butler matrices was over 15 dB [17]. We designed 7×16 Butler matrix working on the 40-GHz band for DFT processing to generate and separate seven OAM modes (OAM modes 0, ±1, ±2, ±3). We first designed 8×16 Butler matrices and terminated a port for OAM mode 4. By using these Butler matrices, we can reduce RF chains because we just prepare the RF chains for the OAM modes used to multiplex rather than for all antenna elements.

The OAM beams were propagated to the Rx. The

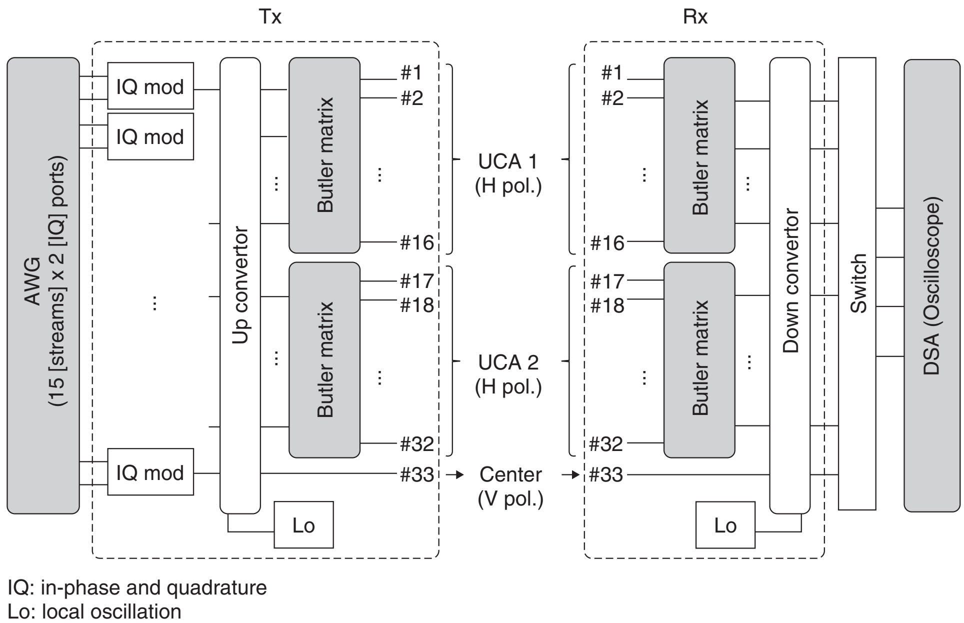


Fig. 6. Block diagram of our experiment.

Table 2. System specifications.

Parameters	Values	Parameters	Values
OAM modes	–3, –2, –1, 0, 1, 2, 3	Baud rate	1.5 Gbaud
Polarization	V/H	FFT size/CP size	1024/16 symbols
Diameter of UCA	120 cm	FEC (outer/inner)	BCH/LDPC
Num. of antennas	16 elements/pol. & center	Equalization	SC-FDE/MMSE
Antenna gain	25 dBi/element	Modulation (prepared)	QPSK–1024QAM
Frequency	39.5–41 GHz	Coding rate (prepared)	1/2, 3/5, 2/3, 3/4, 4/5, 5/6, 8/9, 9/10

CP: cyclic prefix
FFT: fast Fourier transform
MMSE: minimum mean square error
QAM: quadrature amplitude modulation
QPSK: quadrature phase-shift keying
SC-FDE: single-carrier frequency domain equalization

Rx was embedded with a variable attenuator, which made it possible to adjust the received signal level appropriately even if the received power varies with the transmission distance. The received signals were separated for each OAM mode signal by using the Butler matrices and down-converted to intermediate frequency (IF) signals. Finally, the digital serial analyzer (DSA) stored and saved the IF signals as waveform data. The saved waveforms were demodulated by MATLAB off-line signal processing, and we evaluated the transmission capacity. We coded the multiplexed data streams with a low-density parity-check (LDPC) code and Bose-Chaudhuri-Hocquenghem (BCH) code with a frame length of 64,800 bits on the basis of the Digital Video Broadcasting - Satellite - Second Generation (DVB-S2) standard used as forward error collection (FEC). We applied an adaptive modulation and coding algorithm with which the modulation order and adaptively determined the coding rate depending on the signal-to-noise ratios (SNRs). We prepared error-free SNR ranges in advance as a table for each modulation order and coding rate. We summarize the system specifications of the OAM transmission in **Table 2**.

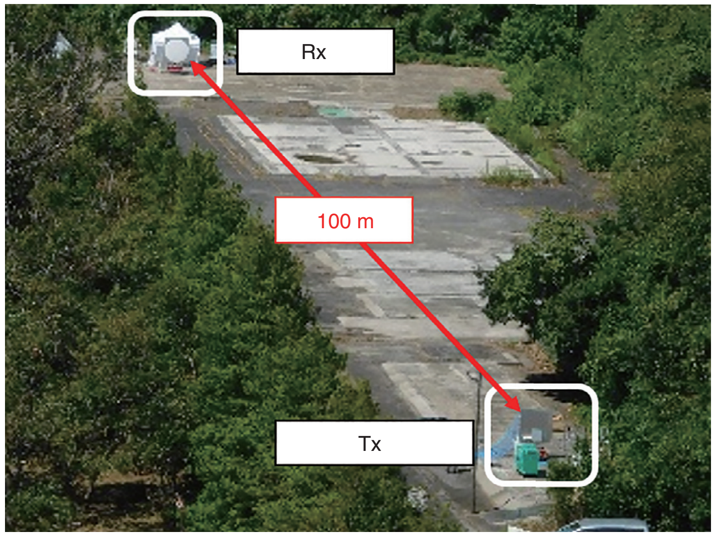


Fig. 7. Experimental environment for wireless OAM multiplexing transmission at a distance of 100 m in the field.

4. Experimental evaluations

4.1 Experimental setup

Table 2 lists the specific parameters of our experimental setup. We multiplexed 15 data streams using 7 OAM modes and 2 UCAs with different linear polarizations in both transmitting and receiving antennas. We then designed antenna elements in the UCAs, which were patch antenna arrays with an antenna gain of 25 dBi, as shown in Fig. 5. The isolation between different linear polarizations of the antenna elements was more than 30 dB on average across the whole bandwidth. In accordance with the relationship between effective bits in a symbol and SNR of our digital signal processing and the symbol rate of 1.5 GHz, roughly a 20-dB SNR on average of 15 data streams is enough to obtain more than 100 Gbit/s. The amount of transmitting power and antenna gains was roughly 53 dB. The channel gain was then roughly –90 dB, and the noise floor including the receiver noise figure was –67.7 dBm. We thus determined that our experimental system was able to obtain more than a 20-dB SNR on average across all of OAM modes.

4.2 Experimental results

We conducted outdoor wireless transmission experiments at distances of 100 and 200 m. Each distance was a line-of-sight environment, as shown in **Figs. 7** and **8**, respectively. The alignment between the Tx and Rx was conducted using a laser rangefinder placed at the corner of the antenna. Error-free transmission with channel coding was 137.3 Gbit/s with 15 streams at a distance of 100 m. The spectral efficiency was 91.5 bit/s/Hz. The 200-m error-free

Fig. 8. Experimental environment for wireless OAM multiplexing transmission at a distance of 200 m in the field.

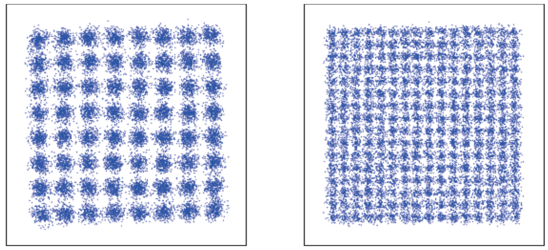


Fig. 9. Representative constellation diagrams of 64 QAM and 256 QAM.

Table 3. The receiving SNRs, modulation, and coding rate during transmission experiment at distances of 100 and 200 m.

	100-m transmission			200-m transmission		
Order	Transmission capacity [Gbit/s]	Modulation	Coding rate	Transmission capacity [Gbit/s]	Modulation	Coding rate
#1	11.25	1024QAM	3/4	11.25	1024QAM	3/4
#2	10.80	256QAM	9/10	10.80	256QAM	9/10
#3	10.80	256QAM	9/10	10.80	256QAM	9/10
#4	10.80	256QAM	9/10	10.80	256QAM	9/10
#5	10.80	256QAM	9/10	10.80	256QAM	9/10
#6	10.80	256QAM	9/10	10.80	256QAM	9/10
#7	10.80	256QAM	9/10	10.67	256QAM	8/9
#8	10.67	256QAM	8/9	9.60	256QAM	4/5
#9	10.00	256QAM	5/6	8.10	64QAM	9/10
#10	9.60	256QAM	4/5	7.20	64QAM	4/5
#11	9.60	256QAM	4/5	6.00	64QAM	2/3
#12	8.10	64QAM	9/10	4.80	16QAM	4/5
#13	6.75	64QAM	3/4	3.60	16QAM	2/3
#14	4.00	16QAM	2/3	2.25	QPSK	3/4
#15	2.50	QPSK	5/6			
Total	137.27 Gbit/s			117.47 Gbit/s		

transmission with channel coding was 117.5 Gbit/s with 14 streams. **Figure 9** shows the representative measured constellations of 64 quadrature amplitude modulation (QAM) and 256 QAM. The SNRs were 21.03 and 24.55 dB, and the bit error rates (BERs) without FEC were 0.428 and 1.622%. We transmitted a few data frames and confirmed reproducibility. The spectral efficiency was 78.3 bit/s/Hz. The multiplex number of 14 at 200 m was due to the modified water-filling-power allocation and providing a margin for determining modulation and coding rates. **Table 3** shows the transmission capacity, modulation, and coding rate of each stream. They are listed in descending order of transmission capacity, that is, the order of the receiving SNR. The modulations used were from quadrature phase-shift keying (QPSK) to 1024 QAM for each stream, and transmission rates were up to 11.25 Gbit/s. We previously reported the achievement of 119.7 Gbit/s transmission at a distance of 100 m and showed that similar capacity can be achieved at a distance of 200 m.

5. Discussion and conclusion

We designed transmitting and receiving antennas of OAM multiplexing for over 100-Gbit/s wireless transmission at a distance of 100 m. Our antennas have two UCAs with 16 antenna elements for different linear polarizations, and Butler matrices for OAM-mode generation and separation. Therefore, we could use two orthogonal bases of seven OAM modes (OAM mode 0, ±1, ±2, ±3) and dual polarization. An antenna element of V-pol. was arranged at the center of each antenna. The radio wave transmitted from the center antenna element was OAM mode 0, which provided an additional data stream for multiplexing. Note that there were two data streams in OAM mode 0 of V-pol., which were not orthogonal to each other. However, they had different eigenvectors for the radius direction, so they could be isolated by traditional MIMO-based digital signal processing. We conducted a transmission experiment using OAM multiplexing in the field at distances of 100 and 200 m. We confirmed that the transmission capacity of over 100 Gbit/s can be achieved even at 200 m. We implemented offline signal processing including transmit precoding based on SVD and the OAM multiplexing Tx and Rx on the 40-GHz band. This Tx and Rx can generate and separate seven OAM modes (−3, −2, −1, 0, 1, 2, 3) by using Butler matrices for each polarization. Each channel carries a 1.5-Gbaud signal. The results indicate that wireless data transmissions of 137.3 Gbit/s at a distance of 100 m and 117.5 Gbit/s at a distance of 200 m were successful. The spectral efficiency was respectively 91.5 and 78.3 bit/s/Hz at 100 and 200 m.

References

[1] DOCOMO 6G White Paper, Feb. 2021. https://www.nttdocomo.co.jp/english/corporate/technology/whitepaper_6g/

[2] The 3rd Generation Partnership Project (3GPP), https://www.3gpp.org/

[3] Y. Corre, G. Gougeon, J. Doré, S. Bicaïs, B. Miscopein, E. Faussurier, M. Saad, J. Palicot, and F. Bader, "Sub-THz Spectrum as Enabler for 6G Wireless Communications up to 1 Tbps," Presented at 6G Wireless Summit, Mar. 2019. http://www.6gsummit.com/wp-content/uploads/2019/04/Day3_Session7_Corre_SIRADEL.pdf

[4] SAMSUNG's 6G White Paper, Jul. 2020.

[5] X. Li, J. Yu, L. Zhao, K. Wang, C. Wang, M. Zhao, W. Zhou, and J. Xiao, "1-Tb/s Millimeter-wave Signal Wireless Delivery at D-band," J. Lightw. Technol., Vol. 37, No. 1, pp. 196–204, Jan. 2019.

[6] Press release issued by Northrop Grumman, "Northrop Grumman, DARPA Set New Standard for Wireless Transmission Speed," Aug. 2018. https://news.northropgrumman.com/news/releases/northrop-grumman-darpa-set-new-standard-for-wireless-transmission-speed

[7] Press release issued by Nokia, "Nokia, Elisa and Qualcomm Achieve 5G Speed Record in Finland," Nov. 2020. https://www.nokia.com/about-us/news/releases/2020/11/18/nokia-elisa-and-qualcomm-achieve-5g-speed-record-in-finland/

[8] Y. Yan, L. Li, Z. Zhao, G. Xie, Z. Wang, Y. Ren, N. Ahmed, S. Sajuyigbe, S. Talwar, M. Tur, N. Ashrafi, S. Ashrafi, A. F. Molisch, and A. E. Willner, "32-Gbit/s 60-GHz Millimeter-wave Wireless Communication Using Orbital Angular Momentum and Polarization Multiplexing," Proc. of IEEE International Conference on Communications (ICC) 2016, May 2016.

[9] Press release issued by NEC, "NEC Successfully Demonstrates Real-time Digital OAM Mode Multiplexing Transmission over 100m in the 150GHz-band for the First Time," Mar. 2020. https://www.nec.com/en/press/202003/global_20200310_01.html

[10] Y. Yagi, H. Sasaki, T. Yamada, and D. Lee, "200 Gbit/s Wireless Transmission Using Dual-polarized OAM-MIMO Multiplexing on 28 GHz Band," Proc. of IEEE GLOBECOM Workshop on HCPtP, Dec. 2019.

[11] Y. Yagi, H. Sasaki, T. Semoto, T. Kageyama, T. Yamada, J. Mashino, and D. Lee, "Field Experiment of 117 Gbit/s Wireless Transmission Using OAM Multiplexing at a Distance of 200 m on a 40 GHz Band," Proc. of IEEE ICC Workshop on OAMT, June 2021.

[12] C. B. Czegledi, M. Hörberg, M. Sjödin, P. Ligander, J. Hansryd, J. Sandberg, J. Gustavsson, D. Sjöberg, D. Polydorou, and D. Siomos, "Demonstrating 139 Gbps and 55.6 bps/Hz Spectrum Efficiency Using 8×8 MIMO over a 1.5-km Link at 73.5 GHz," Proc. of 2020 IEEE/MTT-S International Microwave Symposium (IMS), Aug. 2020.

[13] C. Castro, R. Elschner, T. Merkle, C. Schubert, and R. Freund, "Experimental Demonstrations of High-capacity THz-wireless Transmission System for Beyond 5G," IEEE Commun. Mag, Vol. 58, No. 11, pp. 41–47, Nov. 2020.

[14] Press release issued by NTT DOCOMO, "Mitsubishi Electric and NTT DOCOMO Achieve World's First 27Gbps Throughput in 5G Outdoor Trials," Nov. 2018. https://www.nttdocomo.co.jp/english/info/media_center/pr/2018/1122_00.html

[15] M. Krenn, R. Fickler, M. Fink, J. Handsteiner, M. Malik, T. Scheidl, R. Ursin, and A. Zeilinger, "Communication with Spatially Modulated Light through Turbulent Air across Vienna," New J. Phys., Vol. 16, No. 11, Nov. 2014.

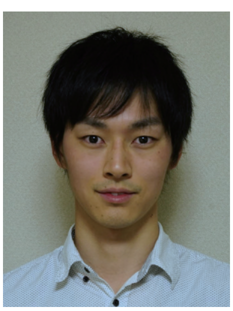
[16] T. Harter, C. Füllner, J. N. Kemal, S. Ummethala, J. L. Steinmann, M. Brosi, J. L. Hesler, E. Bründermann, A.-S. Müller, W. Freude, S. Randel, and C. Koos, "Generalized Kramers-Kronig Receiver for Coherent Terahertz Communications," Nat. Photonics, Vol. 14, pp. 601–606, Sept. 2020.

[17] H. Sasaki, Y. Yagi, T. Yamada, T. Semoto, and D. Lee, "An Experimental Demonstration of over 100 Gbit/s OAM Multiplexing Transmission at a Distance of 100 m on 40 GHz Band," Proc. of IEEE ICC Workshop on OAMT, June 2020.

Doohwan Lee

Distinguished Researcher, Wave Propagation Laboratory, NTT Network Innovation Laboratories.

He received a B.S. (Hons.) and M.S. in electrical engineering from Seoul National University, South Korea, in 2004 and 2006, and Ph.D. in electrical engineering and information systems from the University of Tokyo in 2009. He has been with NTT Network Innovation Laboratories since 2009. From 2012 to 2014, he was an assistant professor at the Research Center for Advanced Science and Technology, the University of Tokyo. Since 2016, he has been a part-time lecturer at Kanagawa University. His research interests include compressed sensing, software/cognitive radio, signal processing, and OAM multiplexing. He received the Best Paper Award and Best Technical Exhibition Award from the Institute of Electronics, Information and Communication Engineers (IEICE) Software Radio in 2011, the IEICE Communications Society Excellent Paper Award in 2012, the IEICE SRW (Short Range Wireless Communications) Young Researcher's Award in 2016, the Best Technical Exhibition Award from SmartCom 2014, the Best Paper Award from SmartCom 2016, the Best Paper Award and Special Technical Award in Smart Radio from IEICE Smart Radio in 2017, and the Distinguished Contributions Award from IEICE Communications Society in 2021.



Yasunori Yagi

Researcher, Wave Propagation Laboratory, NTT Network Innovation Laboratories.

He received a B.E. and M.E. in electrical engineering from Tokyo University of Science in 2015 and 2017. He joined NTT Network Innovation Laboratories in 2017 and has been engaged in research on high-capacity wireless system.

Hirofumi Sasaki

Researcher, Wave Propagation Laboratory, NTT Network Innovation Laboratories.

He received a B.E. and M.E. in engineering from Osaka University in 2011 and 2013. He joined NTT Network Innovation Laboratories in 2013, where he was engaged in research on heterogeneous wireless network system in spectrum sharing scenarios. Since 2015, he has been engaged in research on super high-speed wireless communication in millimeter-wave regions. He received the Best Paper Award from the IEICE SmartCom 2016, the Best Paper Award and Special Technical Award in Smart Radio from the IEICE Smart Radio in 2017, Young Researcher Award from IEICE in 2017. He served as the secretary of IEEE GLOBECOM 2019 Workshop on High Capacity Point-to-Point Wireless Communications (HCPtP 2019) and GLOBECOM 2020 Workshop on High Capacity Wireless Communications (HCWC 2020).

Hiroyuki Shiba

Senior Research Engineer, Supervisor, Wave Propagation Laboratory, NTT Network Innovation Laboratories.

He received a B.E. and M.E. from Gunma University in 1995 and 1997 and Ph.D. from Waseda University, Tokyo, in 2014. Since joining NTT Wireless Systems Laboratories in 1997, he has been engaged in research on software-defined radio, cognitive radio technologies, and flexible wireless systems. He is currently engaged in research and development of high-capacity wireless systems. He is a senior member of IEICE.

Recent Standardization Activities in ITU-R SG 3

Wataru Yamada

Abstract

When considering new frequency allocation and transmission technology, it is essential to standardize the radio-wave propagation model. This article reports on the latest trends in ITU-R (International Telecommunication Union - Radiocommunication Sector) Study Group 3, an international standardization organization for radio-wave propagation models led by the author and his colleagues as the representative of the NTT Group.

Keywords: WRC, IMT, HIBS

1. Configuration and role of ITU-R SG 3

An interference-evaluation model for conducting frequency-sharing studies is developed when considering a new wireless system to be discussed in each Study Group (SG) of the International Telecommunication Union - Radiocommunication Sector (ITU-R). SG 3 [1], the group of radio-wave propagation experts, receives and responds to inquiries about propagation models from other SGs in the form of liaisons. It plays the role of advancing the discussion of each SG. Therefore, the primary purpose of SG 3's activities is to support other SGs. To provide this support in a timely manner, it is essential to model all propagation characteristics before system examination, and these propagation models are recommended and reported as the ITU-R P Series. These Recommendations and Reports of the P Series are used not only for the use of ITU-R but also for the institutionalization of each country, including Japan.

At the June 2021 meeting, deliberations were centered on the World Radiocommunication Conference (WRC)-23 agenda, which was decided at WRC-19 under the structure shown in **Table 1**, resulting in 24 revisions to the P Series Recommendations and Reports.

This article describes the latest trends in the hot topics of the June 2021 meeting, i.e., "New Frequency Allocation to IMT (International Mobile Telecommunications)" and "Propagation Model for High-altitude Platform Stations."

2. New Frequency Allocation to IMT

The WRC-23 agenda 1.2 to consider identification of the frequency bands 3300–3400 MHz, 3600–3800 MHz, 6425–7025 MHz, 7025–7125 MHz, and 10.0–10.5 GHz for IMT was decided. The target frequency range in the Asian region is only 7025–7125 MHz. In response to this decision, SG 3 is in the process of revising the relevant Recommendations. **Table 2** shows SG 3-related Recommendations related to this agenda.

From Working Party (WP) 5D, which is in charge of this agenda, to SG 3, inquiries about radio-wave propagation models for frequency-sharing and system-compatibility evaluation for the four scenarios shown in **Fig. 1** have been issued.

To respond to this inquiry, SG 3 is considering introducing the utilization method for the element that can be applied by the current Recommendation and revising the Recommendation by incorporating the corresponding technology for the element that cannot be applied by the current Recommendation. Since each Recommendation has differences in the target environment and frequency, it is necessary to evaluate frequency sharing and system compatibility by combining various Recommendations to correspond

Table 1. SG 3's June 2021 meeting structure.

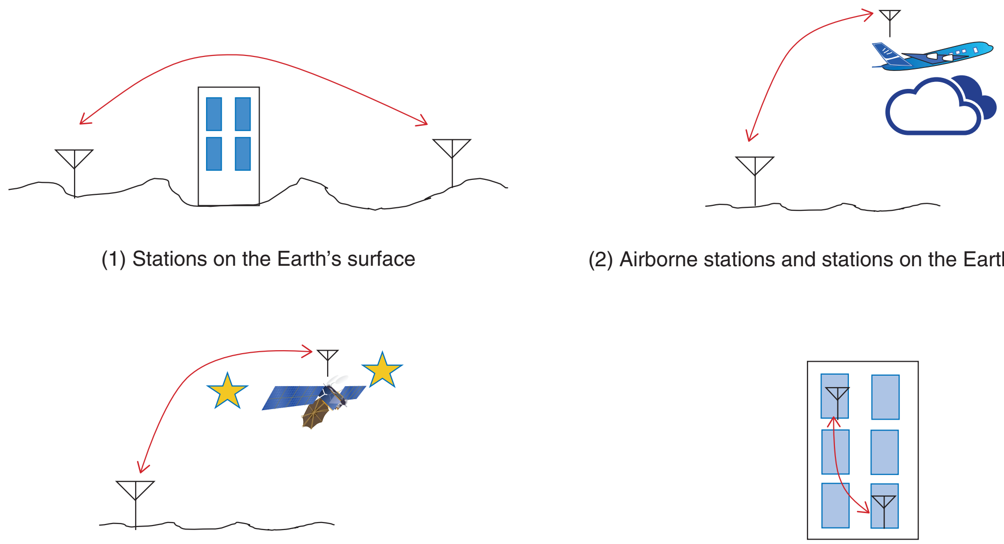
ITU-R SG 3: Radiowave propagation Chair: C. Wilson (Australia) Vice-Chair: G. A.-A. Aws Majeed (Iraq), C. Allen (UK), S.-H. Bae (Korea), M. Pattanaik (India), Y. R. M. Dhossa (Togolese), A. Belkhadir (Morocco), L. Castanet (France), T. Al-Saif (Kuwait), O. Iastrebtsova (Russia), Z. Zhao (China) Counsellor: M. D. Botha			
WP 3J: Propagation fundamentals	Chair: C. Riva (Italy) Vice-Chair: L. Castanet (France)	Sub-G 3J-1: Effects of the clear atmosphere	Chair: C. Allen (UK)
		Sub-G 3J-2: Effects of clouds and precipitation	Chair: A. Martellucci (ESA)
		Sub-G 3J-3: Global mapping and statistical aspects	Chair: L. Castanet (France)
		Sub-G 3J-4: Vegetation and obstacle diffraction	Chair: S. Salamon (Australia)
WP 3K: Point-to-area propagation	Chair: P. Mckenna (USA) Vice-Chair: H. Suzuki (Australia), W. Yamada (Japan)	Sub-G 3K-1: Path-specific prediction methods	Chair: I. Stevanovic (Swiss)
		Sub-G 3K-2: Path-general prediction methods	Chair: F. Lewicki (Poland)
		Sub-G 3K-3: Short-range propagation	Chair: W. Yamada (Japan)
WP 3L: Ionospheric propagation and radio noise	Chair: C. Behm (USA) Vice-Chair: A. Canavitsas (Brazil)	Sub-G 3L-1: MF, LF and lower-frequency propagation	Chair: A. Canavitsas (Brazil)
		Sub-G 3L-2: Trans-ionospheric propagation	Chair: R. Orus-Perez (ESA)
		Sub-G 3L-3: Radio noise	Chair: E. Hill (USA)
WP 3M: Point-to-point and Earth-space propagation	Chair: G. Feldhake (USA) Vice-Chair: R. Rudd (UK) L. Lin (China)	Sub-G 3M-1: Terrestrial paths	Chair: S. Salamon (Australia)
		Sub-G 3M-2: Earth-space paths	Chair: L. Castanet (France)
		Sub-G 3M-3: Interference paths	Chair: I. Stevanovic (Swiss)
		Sub-G 3M-4: Digital products	Chair: A. Martellucci (ESA)
		Clutter and building-entry loss	Chair: R. Arefi (Intel) and R. Rudd (UK)

LF: low frequency
MF: medium frequency

Table 2. P Series Recommendations related to Agenda 1.2.

Recommendation	Title of Recommendation
ITU-R P.452	Prediction procedure for the evaluation of interference between stations on the surface of the Earth at frequencies above about 0.1 GHz
ITU-R P.528	A propagation prediction method for aeronautical mobile and radionavigation services using the VHF, UHF and SHF bands
ITU-R P.619	Propagation data required for the evaluation of interference between stations in space and those on the surface of the Earth
ITU-R P.1238	Propagation data and prediction methods for the planning of indoor radiocommunication systems and radio local area networks in the frequency range 300 MHz to 450 GHz
ITU-R P.1411	Propagation data and prediction methods for the planning of short-range outdoor radiocommunication systems and radio local area networks in the frequency range 300 MHz to 100 GHz
ITU-R P.1546	Method for point-to-area predictions for terrestrial services in the frequency range 30 MHz to 4000 MHz
ITU-R P.1812	A path-specific propagation prediction method for point-to-area terrestrial services in the frequency range 30 MHz to 6000 MHz
ITU-R P.2001	A general purpose wide-range terrestrial propagation model in the frequency range 30 MHz to 50 GHz
ITU-R P.2108	Prediction of clutter loss
ITU-R P.2109	Prediction of building entry loss

SHF: super high frequency
UHF: ultra high frequency
VHF: very high frequency



(1) Stations on the Earth's surface

(2) Airborne stations and stations on the Earth's surface

(3) Stations in space and stations on the Earth's surface

(4) Stations within the same building

Fig. 1. Scenarios required for studying new frequency allocation to IMT for WRC-23.

to the four scenarios in Fig. 1. The following describes the use of each Recommendation for frequency-sharing and system-compatibility evaluation.

2.1 Stations on the Earth's surface

(1) Recommendation P.452: Prediction procedure for the evaluation of interference between stations on the surface of the Earth at frequencies above about 0.1 GHz

This Recommendation aims to calculate the interference between ground stations at frequencies of about 0.1 to 50 GHz. When using the clutter correction of this Recommendation as an interference evaluation, information on the distance between the radio station and nearby obstacles and the height of those obstacles is essential. If there is no information about nearby obstacles, the additional loss should be calculated using the method in Recommendation P.2108. In this Recommendation, the basic propagation-loss calculation of free space was revised at the June 2021 meeting so that the actual distance can be used instead of the great circle distance[*1] used thus far.

(2) Recommendation P.2001: A general purpose wide-range terrestrial propagation model in the frequency range 30 MHz to 50 GHz

This Recommendation provides a time-percentage calculation method for the fundamental propagation loss in the range of 30 MHz to 50 GHz. This calculation method is suitable when using Monte Carlo simulations because it can provide a full-time percentage range from 0 to 100%. This Recommendation was updated at the June 2021 meeting to revise the calculation of the total propagation loss in free space so that the actual distance can be used instead of the great circle distance.

(3) Recommendation P.1812: A path-specific propagation prediction method for point-to-area terrestrial services in the frequency range 30 MHz to 6000 MHz

This Recommendation is suitable for predicting wireless-communication systems with a frequency range of 30 MHz to 6 GHz, a distance of 0.25 km to about 3000 km, and a ground clearance of about 3 km. Power fluctuations can predict time percentages in the range of 1 to 50% and location probability in the range of 1 to 99%. This method makes detailed estimation possible on the basis of the terrain profile in which the propagation path contains clutter. This Recommendation was revised at the June 2021 meeting to extend the upper-frequency limit from 3 to 6 GHz and address the double count of clutter losses.

(4) Recommendation P.1546: Method for point-to-area predictions for terrestrial services in the frequency range 30 MHz to 4000 MHz

*1 Great circle distance: The distance of the route connecting two points on the Earth with an arc along the ground.

This Recommendation is a terrestrial-service-prediction method in the frequency range of 30 MHz to 4 GHz, intended for use on land, sea, and mixed land-sea routes up to 1000 km in length, with an effective transmit antenna height of up to 3000 m. This method builds an estimation formula on the basis of interpolation/extrapolation from the field-strength curve derived from actual measurements such as distance, antenna height, frequency, and time percentage. This method has clutter correction for the receiving terminal but cannot be combined with Recommendation P.2108.

(5) Recommendation P.1411: Propagation data and prediction methods for the planning of short-range outdoor radiocommunication systems and radio local area networks in the frequency range 300 MHz to 100 GHz

This Recommendation provides a method of predicting outdoor propagation losses on routes less than 1 km. The environments covered in this Recommendation are very high-rise/high-rise/low-rise cities as well as suburban, residential, and rural. While many prediction methods are based on specific frequencies within the target range, site-general prediction methods can be used to calculate a basic transmission loss characteristics by arbitrarily specifying the frequency, environment, and distance. It is inappropriate to add clutter loss since it inherently involves the effects of clutter. However, building-entry losses needs to be considered separately, and Recommendation P.2109 can be used.

2.2 Airborne stations and stations on the Earth's surface

(1) Recommendation P.528: A propagation prediction method for aeronautical mobile and radionavigation services using the VHF, UHF and SHF bands

This Recommendation describes how to predict basic transmission losses in the 100 MHz to 30 GHz frequency range for air-to-air, ground-to-air, and air-to-ground propagation paths. The parameters required for the calculation are the distance between the antennas, height of the antenna from the average altitude above sea level, frequency, polarization, and time percentage. This Recommendation does not include the effects of hydrometeor*2, scintillation*3, and diffraction due to irregular topography. However, it can model propagation paths beyond the horizon that may be useful for interference analysis.

2.3 Stations in space and stations on the Earth's surface

(1) Recommendation P.619: Propagation data required for the evaluation of interference between stations in space and those on the surface of the Earth

This Recommendation provides a method for calculating propagation effects, such as tropospheric refraction and beam diffusion loss, and a combination of calculations for single-entry or multi-entry interference analysis and is valid up to 100 GHz. If the terrain radio station explicitly has information on the location and dimensions of the terrain or scatter obstacles on the local path, it describes how to calculate the associated loss. This Recommendation also refers to Recommendation P.2108 on clutter loss. When using the entry loss of a building, it is necessary to calculate an additional loss using Recommendation P.2109 in accordance with the environment of the ground terminal. At the June 2021 meeting, explanations were added to help more accurate understanding of this Recommendation.

2.4 Stations within the same building

(1) Recommendation P.1238: Propagation data and prediction methods for the planning of indoor radiocommunication systems and radio local area networks in the frequency range 300 MHz to 450 GHz

This Recommendation covers the frequency range from 300 MHz to 450 GHz and provides various estimation methods for propagation paths contained in a single building. Therefore, it helps in the study of interference scenarios in which terminals are in the same building. At the June 2021 meeting, a new basic propagation-loss prediction model was established that covers the frequency range of 100 GHz or less for office, factory, and corridor environments, and the coefficients required for propagation-loss prediction at 310 and 410 GHz were added.

2.5 Propagation model related to all scenarios

(1) Recommendation P.2108: Prediction of clutter loss

This Recommendation calculates the additional loss of propagation due to clutter, such as buildings around the radio station. The height-gain model that

*2 Hydrometeor: A phenomenon in which raindrops and ice particles fall or float in the atmosphere.

*3 Scintillation: A phenomenon that causes short-period fluctuations in amplitude when radio waves pass through the atmosphere or ionosphere.

corrects the loss for height described in Section 3.1 of this Recommendation can derive the frequency range from 30 MHz to 3 GHz of the additional loss caused by the clutter around the radio station. Diffraction is assumed to be the primary phenomenon in forest and urban environments, and reflection and scattering are considered to be the main phenomena in other environments such as water or sea area, countryside, suburban, and dense urban. When deriving the additional loss with this model, it is necessary to define a typical clutter height. At the June 2021 meeting, the applicable frequencies were expanded on the basis of the measured data to accommodate a broader range of frequencies. The additional loss in cities and suburban environments, which is essential from the viewpoint of interference evaluation, corresponds to the tilt angle of the propagation path. However, the frequency range from 10 to 100 GHz is the estimation range of the current Recommendation. For this reason, studies are underway to extend this estimation method to 10 GHz or less for the next meeting.

(2) Recommendation P.2109: Prediction of building entry loss

This Recommendation derives the additional loss when one radio station is outdoors and the other is inside a building. Coefficients are given to calculate the further loss for two types: typical buildings and thermal efficiency. Since the building-entry loss calculated in this Recommendation was examined independently of other propagation phenomena, the estimation when clutter exists around the radio station is possible by adding the building-entry loss derived in the Recommendation in units of decibels to the additional loss derived in Recommendation P.2108. However, when finding a specific probability for the loss, a mathematical convolution operation is required to find the combined probability distribution.

3. Propagation Model for High-altitude Platform Stations

WRC-23 agenda 1.4 decided to "consider, in accordance with Resolution 247 (WRC-19), the use of high-altitude platform stations as IMT base stations (HIBS) [2] in the mobile service in certain frequency bands below 2.7 GHz already identified for IMT, on a global or regional level." In response to this decision, SG 3 is in the process of revising the relevant Recommendations. The main WP in charge of this agenda is WP 5D. However, they inquired about the radio-wave propagation model that can be used for frequency-sharing and compatibility evaluation between HIBS and terrestrial/aviation/satellite services to SG 3.

Recommendation P.1409 recommends a propagation model related to high-altitude platform stations. However, it is raised in Agenda 1.4, such as the fact that a radio-wave propagation model of 10 GHz or more from the sky to urban and suburban environments has not been defined. The radio-wave propagation model for assessing the impact of HIBS was not sufficient. Therefore, the development of a radio-wave propagation model that enables various evaluations of HIBS is underway. In the previous Recommendation P.1409, the model was formed from the six elements of free-space loss, attenuation due to the atmosphere, attenuation due to clouds, backward scatterings such as raindrops, rainfall attenuation, and convection-sphere scintillation. While cloud attenuation is merged into rainfall attenuation, new elements, such as convection scattering, earth diffraction, topography and obstacle diffraction, clutter attenuation, vegetation attenuation, and indoor penetration loss have been added. A model consisting of 11 elements (**Fig. 2**) has been recommended.

However, since the radio-wave propagation model required for frequency-sharing and compatibility evaluation could not be developed, an offline meeting was held at Correspondence Group 3J-3K-3M-14 to conclude at the next meeting. Discussions will continue until the next SG 3 meeting.

4. Conclusion

ITU-R SG 3 is essential for the effective use of frequencies, which are rare resources shared worldwide. Since wireless systems are an area where technological progress is rapid, ITU-R SG 3 must produce results in a shorter period while deepening cooperation with other SGs. In addition, it is expected that discussions related to Beyond 5G (fifth-generation mobile communication systems) and 6G will proceed after WRC-23. While observing these latest trends, we will continue to actively promote standardization activities to contribute to the effective use of frequencies for the NTT Group, Japan, and the world.

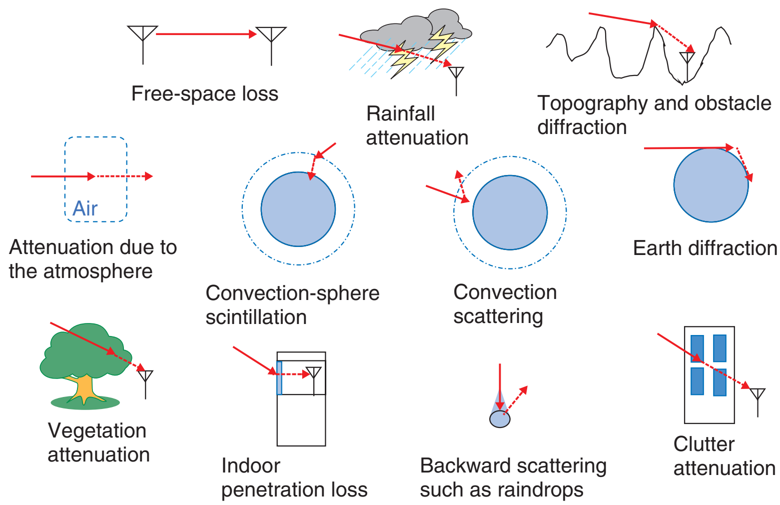


Fig. 2. Eleven propagation elements required for HIBS study.

References

[1] ITU-R SG 3, https://www.itu.int/en/ITU-R/study-groups/rsg3/Pages/default.aspx

[2] Press release issued by NTT DOCOMO, "Zephyr High Altitude Platform Station (HAPS) achieves connectivity in trial conducted by Airbus and NTT DOCOMO," Nov. 15, 2021. https://www.nttdocomo.co.jp/english/info/media_center/pr/2021/1115_00.html

Wataru Yamada

Distinguished Researcher, Wireless Access Project, NTT Access Network Service Systems Laboratories.

He received a B.E., M.E., and Ph.D. from Hokkaido University in 2000, 2002, and 2010. Since joining NTT in 2002, he has been engaged in the research of propagation characteristics for wide band access systems. From 2013 to 2014, he was a visiting research associate at the Centre for Telecommunications Research in King's College, University of London, UK. He has been serving as a vice-chair of Working Party 3K in ITU-R Study Group 3 since 2016. He received the Young Researcher's Award in 2006, the Communications Society Best Paper Award in 2011, the Best Paper Award in 2014 and 2019 from the Institute of Electronics, Information and Communication Engineers (IEICE), and the Best Paper Award in International Symposium on Antennas and Propagation (ISAP) in 2016. He is a member of the Institute of Electrical and Electronics Engineers (IEEE).

External Awards

Industrial Distinguished Leader
Winner: Shingo Tsukada, NTT Basic Research Laboratories
Date: September 25, 2021
Organization: Asia-Pacific Signal and Information Processing Association (APSIPA)

For nanofiber-electroconductive textile “hitoe” for electrocardiogram monitoring and cardiac pacemaker electrodes.

Distinguished Lecturer
Winner: Hiroshi Sawada, NTT Communication Science Laboratories
Date: November 21, 2021
Organization: The Institute of Electrical and Electronics Engineers (IEEE) Signal Processing Society

He has been selected to serve as an IEEE Signal Processing Society Distinguished Lecturer for the term 1 January 2022 through 31 December 2023.

Best Paper Award
Winners: Kazuki Yamamura, NTT Social Informatics Laboratories; Yuntao Wang, Eiichiro Fujisaki, JAIST
Date: December 3, 2021
Organization: The 24th Annual International Conference on Information Security and Cryptology (ICISC 2021)

For “Improved Lattice Enumeration Algorithms by Primal and Dual Reordering Methods.”
Published as: K. Yamamura, Y. Wang, and E. Fujisaki, “Improved Lattice Enumeration Algorithms by Primal and Dual Reordering Methods,” Proc. of ICISC 2021, Springer LNCS, to appear.

APSIPA Sadaoki Furui Prize Paper Award
Winners: Hiroshi Sawada, NTT Communication Science Laboratories; Nobutaka Ono, Tokyo Metropolitan University; Hirokazu Kameoka, NTT Communication Science Laboratories; Daichi Kitamura, National Institute of Technology, Kagawa College; Hiroshi Saruwatari, The University of Tokyo
Date: December 15, 2021
Organization: APSIPA

For “A Review of Blind Source Separation Methods: Two Converging Routes to ILRMA Originating from ICA and NMF.”
Published as: 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.

IEEE Fellow
Winner: Shoko Araki, NTT Communication Science Laboratories
Date: January 1, 2022
Organization: IEEE

For contributions to blind source separation of noisy and reverberant speech signals.

Papers Published in Technical Journals and Conference Proceedings

Color Saturation Control by Modulating Spectral Power Distribution of Illumination Using Color Enhancement Factors
M. Tsuchida, A. Kimura, and N. Harada
Journal of Electronic Imaging, Vo. 30, No. 6, 063022, Dec. 2021.
https://doi.org/10.1117/1.JEI.30.6.063022

Color enhancement factors are spectral components that modulate the spectral power distribution (SPD) of illumination for interactive color saturation control. They can enhance more than one target color simultaneously or independently while maintaining the current color appearance. However, the color enhancement factors obtained in our previous work depended on the lighting system. Moreover, the effects of color muting, such as decreasing saturation of the target color, were limited because the color enhancement factors were optimized to increase color saturation. We overcame these two issues and succeeded in implementing color enhancement factors that can both enhance and mute color saturation. First, we conducted detailed simulations of color saturation enhancement and muting to determine parameters for calculating the color enhancement factors (e.g., the center wavelength and full-width at half-maximum) and assessed their potentiality. The results showed that maximum variations of color muting became almost the same as those of color enhancement. We also found that SPDs of illumination for increasing and decreasing color saturation of red, green, and blue had three peaks, respectively, and that they appeared for enhancement and muting alternately. In addition, the color enhancement factors for enhancing and muting color saturation were roughly symmetric. We then performed experiments using a newly designed 12-color LED lighting system, and the results showed color enhancing and muting effects. Finally, we demonstrated color restoration of discolored cultural heritage and coloring simulation of industrial products.

Multi-source Domain Generalization Using Domain Attributes for Recurrent Neural Network Language Models

N. Tawara, A. Ogawa, T. Iwata, H. Ashikawa, T. Kobayashi, and T. Ogawa

IEICE Transactions on Information and Systems, Vol. E105-D, No. 1, pp. 150–160, Jan. 2022.

Most conventional multi-source domain adaptation techniques for recurrent neural network language models (RNNLMs) are domain-centric. In these approaches, each domain is considered independently and this makes it difficult to apply the models to completely unseen target domains that are unobservable during training. Instead, our study exploits domain attributes, which represent common knowledge among such different domains as dialects, types of wordings, styles, and topics, to achieve domain generalization that can robustly represent unseen target domains by combining the domain attributes. To achieve attribute-based domain generalization system in language modeling, we introduce domain attribute-based experts to a multi-stream RNNLM called recurrent adaptive mixture model (RADMM) instead of domain-based experts. In the proposed system, a long short-term memory is independently trained on each domain attribute as an expert model. Then by integrating the outputs from all the experts in response to the context-dependent weight of the domain attributes of the current input, we predict the subsequent words in the unseen target domain and exploit the specific knowledge of each domain attribute. To demonstrate the effectiveness of our proposed domain attributes-centric language model, we experimentally compared the proposed model with conventional domain-centric language model by using texts taken from multiple domains including different writing styles, topics, dialects, and types of wordings. The experimental results demonstrated that lower perplexity can be achieved using domain attributes.

Physical Topology Optimization for Highly Reliable and Efficiently Wavelength-assignable Optical Networks

K. Higashimori, F. Inuzuka, and T. Ohara

Journal of Optical Communications and Networking, Vol. 14, No. 3, pp. 16–24, Mar. 2022.

Optical networks, such as wavelength-division multiplexing networks and elastic optical networks, require high reliability and efficient wavelength allocation. We propose a new physical topology optimization method that achieves both high reliability and efficient wavelength allocation. We compare the results of conventional algebraic connectivity optimization, which discusses physical topology optimization from the viewpoint of spectral graph theory, with those based on reliability optimization, which is important from the viewpoint of network operations. From the comparison analysis, it is shown that our proposed two-step degree bounded reliability-optimization model, which optimizes reliability with appropriate node degree constraints, leads to a physical topology that has high wavelength capacity without compromising reliability.
