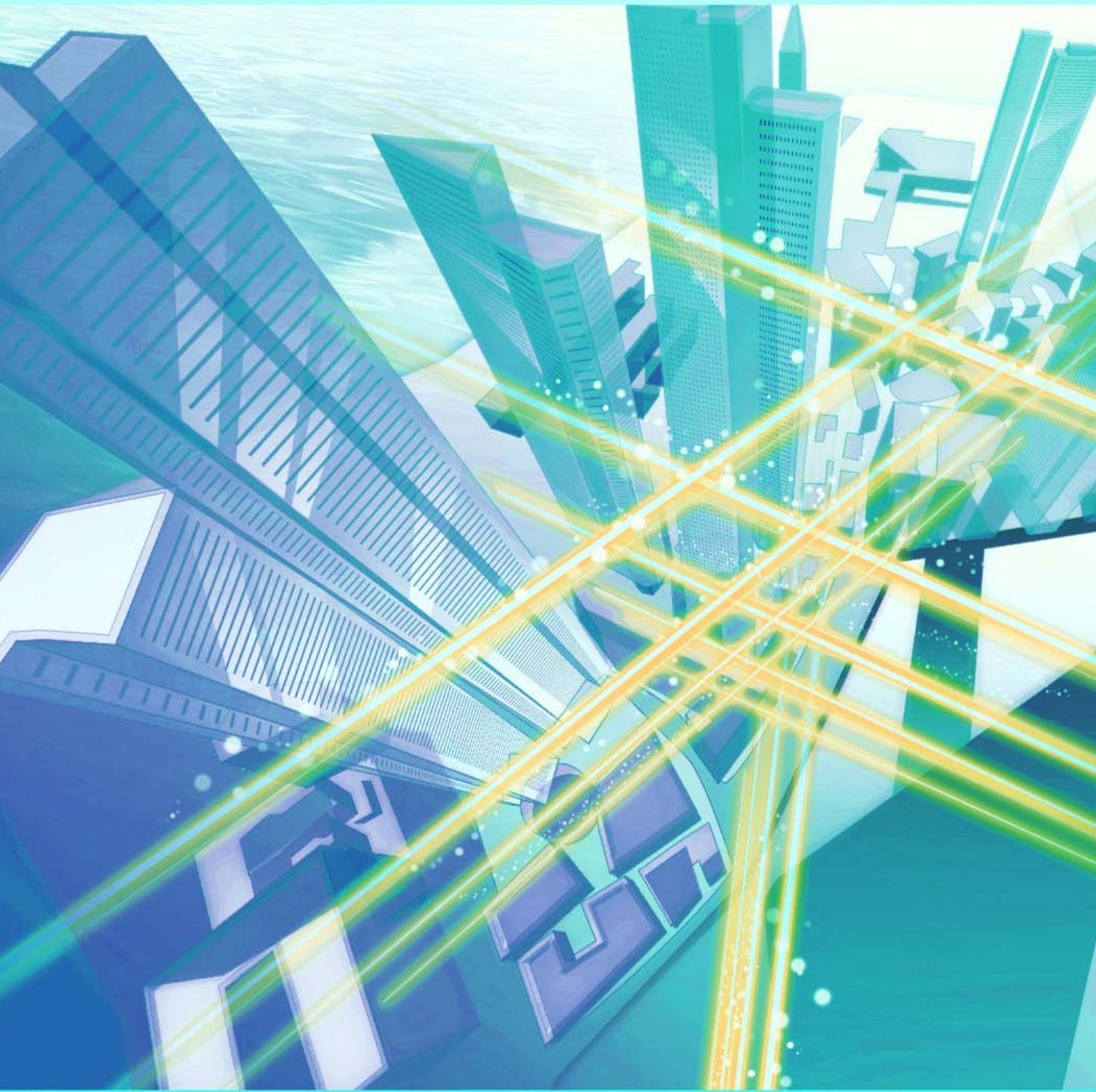
NTT Technical Review

5

2024



May 2024 Vol. 22 No. 5

NTT Technical Review
May 2024 Vol. 22 No. 5

Pursuing Triple Win for Us, Partners, and Society and Trying Our Best to Live in the Present

Yukari Tsuji
Senior Vice President, Head of NTT Information Network Laboratory Group

Abstract

NTT Information Network Laboratory Group promotes research and development (R&D) of communication networks and innovative environmental and energy technologies to create a sustainable and prosperous society. Focusing on three pillars of the Innovative Optical and Wireless Network (IOWN), robust networks, and environment and energy, the laboratory group aims to contribute to building a sustainable infrastructure of an information society and provide new value. We asked Yukari Tsuji, senior vice president, head of NTT Information Network Laboratory Group, about the laboratory group's R&D strategy and her mindset as a top executive.

Keywords: IOWN, sustainability, R&D management

Creating a sustainable infrastructure for an information society

—What is the mission of NTT Information Network Laboratory Group?

Our mission is to provide new value by building a sustainable infrastructure of an information society that connects everything. When people hear the word "sustainable" or "sustainability," many may think of environmental measures such as decarbonization, but we think "sustainability" has a much deeper meaning than that. I will explain the sustainability that we are aiming for in terms of our three pillars of research and development (R&D), i.e., the Innovative Optical and Wireless Network (IOWN), robust networks, and environment and energy.

Regarding the first pillar, IOWN, we aim to develop ultra-high-speed, low-latency, low-power-consumption networks. It is obvious that the amount of communication traffic and information handled by datacenters will increase exponentially with the advancement of the data-driven society and evolution and popularization of artificial intelligence (AI) as represented by generative AI. As this trend continues, power consumption will become a serious social problem. IOWN will be able to reduce power consumption. It will also enable local production for local consumption of data and power by connecting datacenters and edges via the All-Photonics Network (APN). If many users can use a variety of services and solutions simultaneously within IOWN, additional social costs will be kept as low as possible, even as the numbers of use cases and users increase.

The second pillar, robust networks, is a key element in achieving sustainability in the sense that a communication network can continue to be used under any circumstances. To minimize damage and impact

due to disasters and failures, it is necessary to increase network resilience. To this end, we are steadily advancing R&D on visualization of network status and the impact of disasters and failures on services. Additionally, technologies for zero-touch operation minimize operator involvement in network operations, which will also play a key role as countermeasures against the future decline in the workforce.

The third pillar, environment and energy, is not hard to imagine. Aiming to achieve both economic growth and zero environmental impact, we are researching proactive sustainability technologies, such as energy creation, and developing technologies for global environmental observation and forecasting that will enable a society to adapt to the changing environment. For example, if extreme weather events such as typhoons and linear rain bands could be predicted promptly and with high accuracy, it would be possible to take countermeasures proactively, which would also contribute to building robust networks.

We thus view "sustainability" from multiple perspectives and aim to create a truly sustainable infrastructure of an information society through our above-described R&D.

—I understand that R&D from various perspectives is necessary to create a sustainable infrastructure of an information society. What kind of lineup does NTT Information Network Laboratory Group have in place to conduct such R&D?

NTT Information Network Laboratory Group consists of three laboratories: NTT Network Service Systems Laboratories, NTT Access Network Service Systems Laboratories, and NTT Space Environment and Energy Laboratories.

NTT Network Service Systems Laboratories is researching and developing infrastructure technologies that support network architectures and network systems for future network services as well as technologies for communication-traffic control, quality management, and network operations. NTT Access Network Service Systems Laboratories is researching and developing wireless access, optical-fiber access, access infrastructure, and related operation technologies to create a smart society. NTT Space Environment and Energy Laboratories is researching technologies for creating overwhelmingly clean next-generation energy and reducing environmental impact by predicting the future of the global environment and society and helping society to adapt to the environment.

I consider it very important that these three laboratories belong to NTT Information Network Laboratory

Group. As I mentioned earlier, to build a sustainable infrastructure of an information society and make it useful to society, it is necessary to consistently consider network as well as environmental and energy aspects and obtain their synergistic effects. To break through the limitations of conventional technologies and create a prosperous society, we take pride in our daily R&D work while collaborating with researchers and business partners in Japan and abroad in a manner that enables us to chart a course for practical implementation of our research results in society.

Creating an environment in which researchers with different capabilities can interact with each other

—Would you tell us about the R&D that you are focusing on?

Our current major goal is to establish technologies and commercialize business to make IOWN easier for more users to use. NTT Information Network Laboratory Group is particularly involved in developing the APN, a component of IOWN. Our goal is to provide an environment in which users can use optical paths for a variety of use cases, whenever they want to use them, in the amount they need, and with the quality appropriate for their usage. To reach this goal, we are developing technologies to provide end-to-end optical paths. For example, we developed the Photonic Gateway, which aggregates various user equipment with low latency and the Photonic Exchange, which can connect networks with different types of optical fiber. We are also researching and developing controllers for controlling the APN according to user requirements. We are also investigating next-generation network architecture and operations. For example, we intend to create flexible service experiences that are not limited by user environment, terminal equipment, or service. To achieve this, the key technology of networks is to coordinate and speed-up information processing between terminal equipment and the cloud anytime and anywhere. Regarding wireless communications, in addition to increasing capacity and reducing latency, we are researching and developing proactive wireless-control technologies that adapt to changes in the wireless environment and technologies to extend wireless coverage to unexplored areas by using non-terrestrial networks.

—How is your R&D progressing? Could you tell us about some creative ways to promote R&D?

We are making good progress despite some difficulties. IOWN was initially targeted for implementation in 2030, but IOWN APN1.0, a leased line service, was launched in March 2023. After unveiling our latest technology at Expo 2025 Osaka, Kansai, Japan, we will gradually roll out full-scale and on-demand IOWN services. Since the concept of IOWN was announced in 2019, NTT Information Network Laboratory Group has been studying the specifics of the concept and preparing various technologies, and the fiscal year 2024 is an important year for creating a path toward the deployment of these technologies in business. To clarify when and in what form each technology should be deployed to achieve our goals in accordance with our roadmap, it is necessary to bring together the knowledge of various players and combine our collective strengths.

We therefore called on key members from multiple related departments and research laboratories and set up a venue for cross-technical and cross-disciplinary discussions, in which I also participate. Rather than just responding to reports, I listen to the multifaceted

opinions of a wide range of participating members, which adds depth to discussions and increases a sense of reality. Through these activities, I have seen a change in mindset among members to one in which everyone is willing to work together across organizations and feel that the speed and depth of our R&D have increased.

To foster such a climate of open-minded cooperation, as the head of NTT Information Network Laboratory Group, I intend to create a work environment in which our researchers can make the most of their individual talents. We have researchers who use cutting-edge technology to produce "world's firsts" or "record-high performance," researchers who think about network architecture, and researchers who are very good at engineering ingenuity. I believe that having such researchers with various specialized skills is an important asset. An open work environment is a key factor in enabling each one of us to make use of their unique characteristics and demonstrate their abilities with confidence. With that in mind, I opened a chat room called "Yukari's Room" on our internal information-sharing tool so I could start to get members of our laboratory group to know me. Every day, I post short messages about how I'm feeling and what I'm thinking, and when I shared about having this interview, I received a series of supportive marks from the members. Therefore, an atmosphere of mutual interaction is being created.

What looks like failure is a process to success in which you confirmed that the method does not work

—It has been more than a year since you assumed current position. How are you feeling?

Every day is much more enjoyable than I imagined. I suppose that it is partly because of my own way of looking at things. I think it is also because our members are enthusiastically working on a very wide range of technologies, and I'm able to exchange ideas with them on a variety of issues on a daily basis. To be honest, before I assumed my current position, I was a little worried about whether the research laboratories that I was returning to after a long time away would be full of energy and whether the researchers would be friendly toward me, but my fears turned out to be completely unfounded.

However, I sometimes think that because we are a group of very talented researchers, we might focus too much on fulfilling our own and our team's responsibilities and respecting the activities of others and other teams, thus confine ourselves to the areas of our responsibilities. To dispel this mindset, as I mentioned above, we are working to create an open

atmosphere so that all parties can step in and discuss things a little at a time and to work in a cross-organizational manner. We are working to fill gaps not only within our laboratory group but also with other NTT research laboratories and operating companies, and I think things have started to turn in the right direction over the past year. Fiscal year 2024 will be a year of major shifts in the implementation phase of IOWN, so I want us to proceed with thorough consideration so that IOWN can be used by many customers.

—Finally, could you give a message to our researchers and engineers as well as our customers and partners?

First, to our researchers and engineers. Many things that "look like failure" will happen in the course of your research. I've also experienced such "failures." However, you might think of it as a "failure" because you just have decided so yourself. What looks like "failure" is the result of having confirmed that the method you used doesn't work for a certain purpose, but a different method might work if you try it. Therefore, what seemed like a "failure" is in fact a "process to success." With this way of thinking in mind, you should not be afraid of taking on challenges. I want you to approach R&D with the mindset of trying things out, saying, "If this method doesn't work, I'll just try another method."

As a message to our customers and partners, we want to create a sustainable infrastructure of an information society. To that end, we want to establish an ecosystem in which our R&D results are incorporated and strive for a situation in which we, our customers and partners, and society all win. Let's build a great future together.

My job as top management is to provide direction and to make the right decisions on a case-by-case basis. For that reason, I strive to keep my mind and body always in a stable state. Behind this attitude are the lessons I learned through team sports, which I have been involved in since childhood. I have been playing basketball since my school days, and I noticed players who each had their own special skills honed their individual skills then combined their skills with those of others to gain even more power for the team. When I took up hot-air ballooning at university, I learned firsthand that there are things, such as the weather and other balloons, I cannot control. There's no point in worrying much about things I cannot control, therefore, I started thinking that I should stop worrying about things I cannot control and focus on what I can control.

That is how I arrived at the word *jikon* (only now), which is a Buddhist term. I interpret this term to mean that the only things that I can change are those concerning "now" and "me" and that since the past cannot be changed, we should try our best to live in the present. The term *jikon* is my motto. I think it would be ideal if we could be of service to society while enjoying the present to the fullest.

Interviewee profile

■ Career highlights

Yukari Tsuji joined NTT in 1989. She became director of the R&D Center at NTT WEST in 2014, vice president, head of NTT Network Technology Laboratories in 2016, board director, deputy general manager of Network Innovation Business Headquarters, and director of IOWN Business Promotion Office at NTT Advanced Technology in 2019, and general manager of IOWN Innovation Business Headquarters at NTT Advanced Technology in 2022. She has been in her current position since June 2023.

Achieving the World's Highest Wireless Transmission Rate by Delving into Theories That Appeared about 100 Years Ago

Doohwan Lee
Senior Distinguished Researcher,
NTT Network Innovation Laboratories

Abstract

With the advent of smartphones, mobile communications have rapidly become popular. Fifth-generation mobile communications system (5G) services, which feature high speed, large capacity, low latency, and simultaneous connection to multiple terminals, are spreading, and the 3rd Generation Partnership Project (3GPP), an international standardization organization, has begun initial discussions on 6G. The technological foundation for the advancement of mobile communications is wireless transmission technology. Doohwan Lee, a senior distinguished researcher at NTT Network Innovation Laboratories, and his colleagues achieved the world's first high-capacity wireless transmission of 1.44 Tbit/s using the sub-terahertz band. We interviewed him about the technology that enabled this achievement, his research themes aiming at exploiting all the properties of radio waves, and his thoughts on the importance of researcher well-being for enabling them to make people happy through research.

Keywords: wireless transmission, orbital angular momentum, sub-terahertz band

Break through the limits of wireless transmission technology in terms of high speed and high capacity through OAM multiplexing technology

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

I have been researching wireless communications since I was a student and continued after joining NTT. Research on wireless communications covers a wide range of fields—including research on radio-wave propagation, suppression and control of interference, and methods for putting signals on radio waves. In these research fields, I have been pursuing high-speed, high-capacity wireless transmission technology.

There have been three directions for achieving high-speed and large-capacity transmission with wireless technology: (i) increase the spatial-multiplexing order through methods such as multiple-input multiple-output (MIMO), (ii) broaden the transmission bandwidth, and (iii) increase the modulation

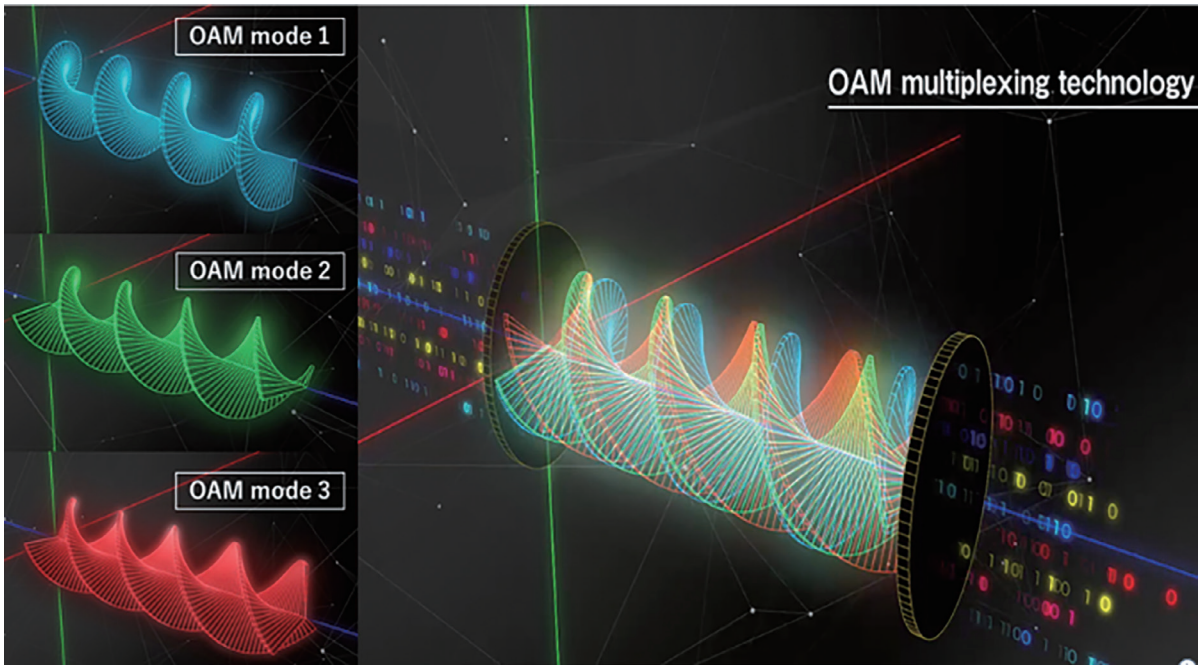


Fig. 1. Principle of OAM multiplexing transmission technology.

level. However, the technology for increasing the modulation level has almost reached its limit, and although the modulation level can be further increased by 20–30%, it is unlikely that it can be increased 10 or 100 times. Under those circumstances, to increase the spatial-multiplexing order, I thought that a new approach was needed and focused on orbital angular momentum (OAM) multiplexing transmission, which involves applying the principal of OAM to wireless communications.

OAM is a physical property of radio waves characterized by the phase of a radio wave rotating on a plane perpendicular to the direction of propagation, and the trajectory of the same phase takes on a helical structure in the direction of propagation (**Fig. 1**). The number of phase rotations is called an OAM mode. For example, the left of Fig. 1 shows, from top to bottom, the trajectory of the same phase in OAM mode 1, OAM mode 2, and OAM mode 3. Radio waves having the OAM property (hereafter, OAM waves) cannot be received without a receiver having the same OAM mode at the time of transmission. As in the relationship between a nut and bolt, an OAM wave can only be received by receivers with the same helical structure. Accordingly, when multiple radio waves having different OAM modes are transmitted simultaneously, they can be separated without interfering with each other by using a receiver with the same number of phase rotations corresponding to each OAM mode of the transmitted wave. Therefore, by putting a data signal in each OAM mode, multiple different signals can be transmitted in a multiplexing manner [1] (right of Fig. 1). In theory, the OAM mode can be increased infinitely, so the multiplexing order can be increased infinitely. However, an OAM wave tends to spread spatially as the wave progresses, and this tendency becomes more pronounced as the OAM mode increases (i.e., number of phase rotations increases). Owing to this spreading nature of OAM waves, for a fixed-sized receiver antenna, as transmission distance increases, the power of the radio waves that can be received by that antenna decreases. Consequently, the higher the OAM mode (more spreading), the lower the ability to extend transmission distance. In other words, the transmission distance is limited.

Although OAM appeared as a theoretical theory in the early 20th century, it has never been tested in a real environment because technology to implement it had not been established. As wireless communications technology in the high-frequency band has matured, OAM multiplexing transmission using a wide bandwidth in the millimeter-wave band has been demonstrated. Laboratory-level experiments

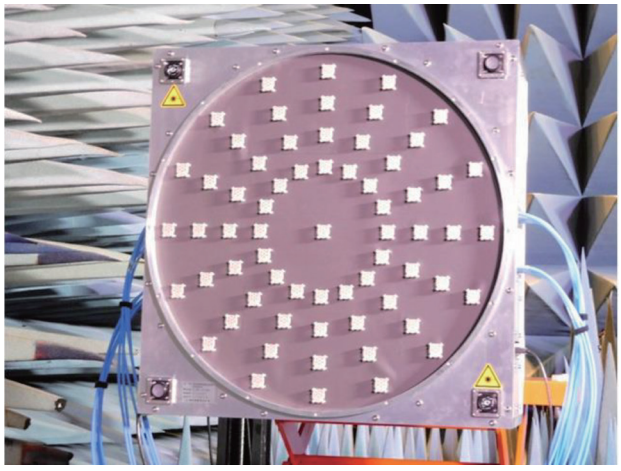


Fig. 2. Prototype transceiver.

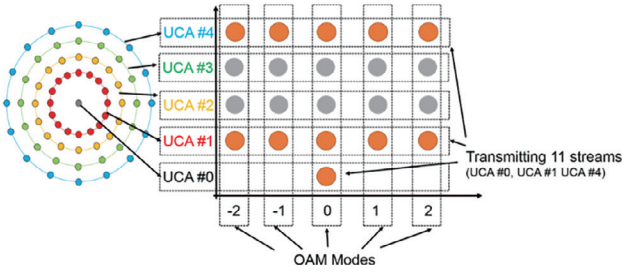


Fig. 3. Example of configuring OAM-MIMO multiplexing transmission.

have been reported by the University of Southern California in the USA: they demonstrated 32-Gbit/s OAM multiplexing transmission using the 28- and 60-GHz bands in 2014 and 2016, respectively. In 2018, NTT achieved the world's first OAM multiplexing transmission with a capacity of 100 Gbit/s over a distance of 10 m using the millimeter-wave band, and in 2020, NTT demonstrated 100-Gbit/s transmission over a distance of 100 m.

—You achieved the world's first 1.44-Tbit/s wireless transmission in the sub-terahertz band through OAM multiplexing transmission technology, right?

To extend transmission distance without using higher OAM modes and increase capacity by simultaneously carrying out OAM multiplexing transmission of multiple sets while maintaining the property that radio waves in different OAM modes do not interfere with each other, we devised OAM-MIMO multiplexing transmission technology that combines OAM technology and widely used MIMO technology. We then developed a prototype transceiver for wireless transmission with OAM-MIMO multiplexing transmission technology in the 28-GHz band (**Fig. 2**).

The prototype transceiver is equipped with four uniform circular array (UCA) antenna elements, which can transmit and receive radio waves in five OAM modes, and one antenna element in the center. With the exception of UCA #0 in the center, UCAs #1 to 4 can generate five OAM modes (–2, –1, 0, 1, and 2) and multiplex radio waves of those OAM modes. Signals multiplexed in the same mode are separated at the receiver by using MIMO technology (**Fig. 3**). With the configuration of antennas shown in Fig. 3, a total of 21 data signals can be transmitted simultaneously.

Signal processing technology to achieve 100-Gbit/s-class wireless transmission using the prototype transceiver consists of three technologies: (i) adaptive modulation and coding (AMC) that adaptively determines the modulation level and channel-coding rate in consideration of the reception quality of each signal, (ii) transmission-power control technology, and (iii) signal-separation technology on the receiver side. Using these technologies, we achieved error-free 100-Gbit/s wireless transmission of more than 10 multiplexed signals over a distance of 10 m (in 2018) and 100 m (in 2020).

We adopted an approach for increasing the spatial-multiplexing order by multiplexing multiple OAM waves using an analog circuit called a Butler matrix. This approach makes it possible to reduce the enormous amount of digital signal processing required to eliminate interference between radio waves in different OAM modes in high-capacity transmission exceeding 1 Tbit/s. We developed an antenna-integrated Butler matrix that operates over a wide bandwidth and with low loss (**Fig. 4**). The antenna-integrated Butler matrix is designed to simultaneously generate and separate eight different OAM waves over a very wide bandwidth (135 to 170 GHz) in a manner that enables multiplexing and transmission of eight data signals.

To transmit eight OAM waves simultaneously using the Butler matrix, the phase of the radio waves must be controlled with extremely high precision. Since the phase of radio waves varies with frequency, it is extremely difficult to uniformly control the phase over a wide bandwidth by using analog circuits. We

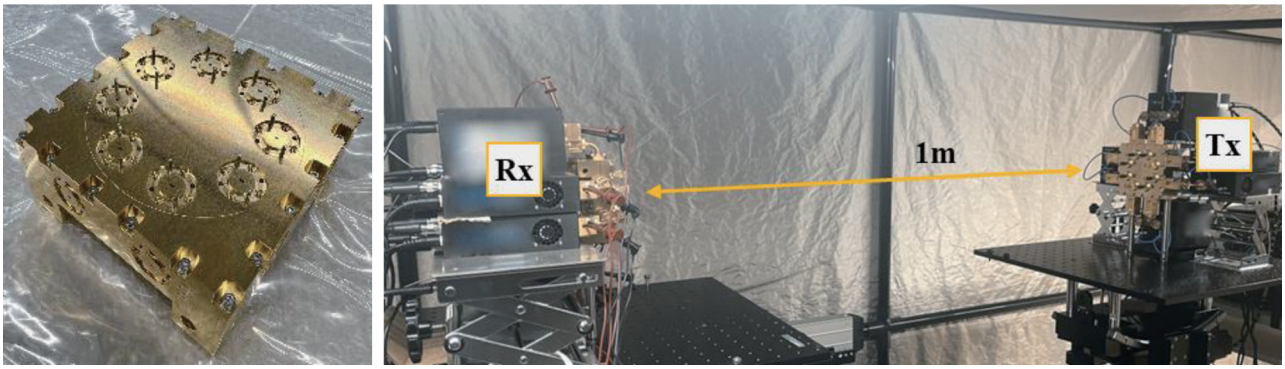


Fig. 4. Antenna-integrated Butler matrix developed for the sub-terahertz band (left) and transmission experiment (right).

therefore analyzed the unique radio-wave propagation in a waveguide, which differs from free space, and devised a phase shifter that can theoretically align the progression of the phase uniformly over a wide bandwidth. By designing a multilayer three-dimensional path (**Fig. 5**) that includes the aforementioned phase shifter so that all paths are electrically equal in length and planar intersections in the circuit (a factor that degrades performance) are eliminated, we developed a Butler matrix that can provide the necessary phase for each OAM mode over a 35-GHz bandwidth and can transmit eight multiplexed data signals. By transmitting the signals by OAM multiplexing in each of two different polarizations, it is also possible to multiplex and transmit twice as many data signals simultaneously without them interfering.

In March 2023, we used this antenna-integrated Butler matrix for transmission tests that demonstrated the world's first large-capacity wireless transmission at a total data rate of 1.44 Tbit/s using the sub-terahertz bands of 135.5 to 151.5 GHz and 152.5 to 168.5 GHz. Our goal is to achieve (i) broadband and high-speed wireless transmission comparable to optical transmission systems and (ii) seamless connection between wireless and optical transmission systems without the need for a complex digital-signal-processing system for multiplexing the OAM waves because the analog circuit (Butler matrix) handles the multiplexing of those waves.

Going forward, we intend to conduct demonstration tests of OAM multiplexing transmission over longer distances exceeding 100 m while envisioning various applications such as wireless backhaul/fronthaul between base stations and relay transmission. We expect our technologies will support the creation and spread of a variety of future services, such as virtual reality, augmented reality, high-definition video transmission, connected cars, and telemedicine, as a wireless communication technology in the era of the Innovative Optical and Wireless Network (IOWN) and sixth-generation mobile communications system (6G).

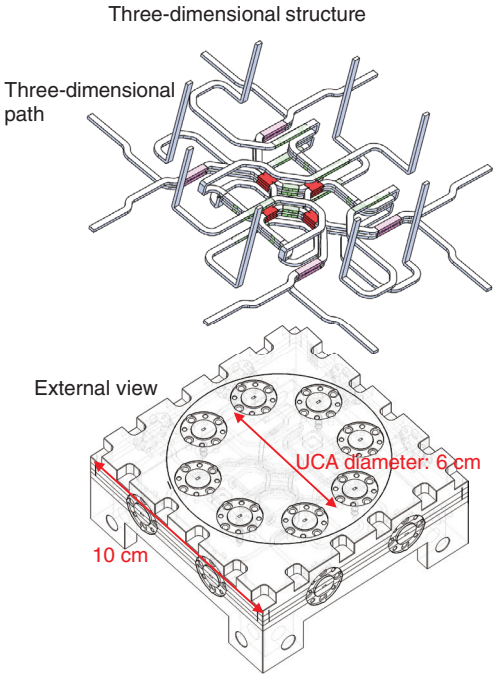


Fig. 5. Three-dimensional structure and external view of Butler matrix.

Research for exploiting all the properties of radio waves

—It is an interesting approach to delve into past theories and record the world's highest data rate. What kind of research will you focus on going forward?

Regarding OAM, it has become possible to demonstrate its theory, proposed about 100 years ago, and we are looking forward to its development toward the practical implementation. Radio waves have many other physical properties that have not been exploited or focused on, and I want to make a theme of using those properties to increase the degree of freedom in the use and expand the range of wireless technology.

For example, since radio waves can travel only in a straight line, if an obstacle is placed in their direction of travel, it will stop them at that point. Diffraction, which allows a beam of a radio wave to somehow avoid the obstacle and travel beyond it, also occurs. However, a portion of the beam leaks as it wraps around the obstacle, so a technology that bends the beam to avoid the obstacle is being investigated. Radio waves also spread out as they travel, which reduces the received power of the waves, but if the radio wave beam can be narrowed, the received power can be increased. Light is like an extremely high-frequency version of radio waves; however, it has a much narrower beam, such as in a laser, which has not yet been achieved with radio waves. Although interference is often treated as a nuisance in regard to wireless communications, I wonder if we can improve communication efficiency by exploiting it.

I believe that by exploring these themes, we will be able to develop wireless communications that have never been conceived of before. As we repeat this research process, I believe we will be able to achieve the ultimate world of fully using all the properties of radio waves and controlling radio waves. I want to play the role of fully using the last remaining property of radio waves.

Good ideas are born when you have plenty of energy, which is increased by the state of well-being

—Please tell us what you keep in mind as a researcher.

I think some people have the impression that to become a leading researcher, you must constantly think about research and immerse yourself in it and be conscious of research in your daily life. Sometimes you need to be immersed in your research, but if that is all you do, when you hit a wall, everything will end there. Rather, I believe that research is born when you have plenty of energy. That is to say, when you are in an unbalanced situation in which you don't have much time and physical strength, you won't come up with any ideas for research. Research ideas will only come to fruition if you have the energy and leeway to think of them, both physically and psychologically, in your private and work life. I've heard stories of people coming up with good ideas while taking a bath. Many people relax while taking a bath, so it's a time when they have some free time to think. I once came up with an idea that led to a patent while brushing my teeth!

Therefore, I try to live a life that allows me to have plenty of energy. I try to keep my working hours to maintain a good work-life balance, I try to get plenty of exercise and respect for my private and family life. I try not to push myself too hard because I believe that if I can make my situation calm and save my strength, I'll be happier in my personal life and will be able to come up with research ideas. On a similar note, research on well-being as a use case for 6G is ongoing; however, I think that if you are not in a state of well-being, you cannot carry out research on well-being. I believe that the results of my research can ultimately make society and people happy, and I cannot do research that makes people happy unless I am happy too. It is important to ensure my well-being and enjoy happiness first; then I can do research to make others happy.

—Please give a message to younger researchers.

As I mentioned above, as you experience the state of well-being, a fulfilling life, and happiness, your research will become unique and fulfilling; as a result, you will also become satisfied, creative, and happy in a manner that creates a virtuous cycle. Therefore, you will be able to enjoy your research. Nevertheless, time is limited, and there are deadlines you must meet. By smartly devising the order of work and the way to proceed, you can create more time, which can be used to do things for the sake of your well-being, thereby leading to a virtuous cycle.

That said, what well-being means to you and the rhythm of your life differ from person to person, so the virtual cycle that I've talked about does not work for everyone. To create a virtual cycle for yourself, you need to think about the essence of your own well-being and the rhythm of your life. Only you

understand yourself; you are the expert on yourself.

References

[1] D. Lee, H. Sasaki, Y. Yagi, T. Yamada, T. Kaho, and H. Shimizu, "Transmission Technology toward the Next Generation of 5G Envisioned by NTT," Radio Engineering & Electronics Association Bulletin FORN, Vol. 10, No. 328, pp. 22–25, 2019 (in Japanese).

[2] Press release issued by NTT, "NTT successfully demonstrates 100 Gbps wireless transmission using a new principle (OAM multiplexing) as a world's first—New groundbreaking technology for achieving next generation of 5G systems—," May 15, 2018. https://group.ntt/en/newsrelease/2018/05/15/180515a.html

[3] Press release issued by NTT, "World's First Successful 1.4-Tbit/s Wireless Transmission in the Sub-THz Band—Contributing to the Creation of New Wireless Services Enabled by IOWN and 6G—," Mar. 30, 2023. https://group.ntt/en/newsrelease/2023/03/30/230330a.html

■ **Interviewee profile**

Doohwan Lee 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.

Optical Cryptographic Circuit Technology Consisting of Optical Logic Gates for a Secure Information-processing Platform

Junko Takahashi
Distinguished Researcher, NTT Social Informatics Laboratories

Abstract

The next-generation communication network driven by optical technology will be achieved in the sixth-generation mobile communication system (6G) of the Innovative Optical and Wireless Network (IOWN) era promoted by NTT. However, a problem of concern here is cryptographic technology. For example, even if optical circuits can be achieved, cryptographic operations in their present form would generate delays or excessive power consumption preventing the advantages of optical technology from being fully used. To solve this problem, NTT is researching technology for replacing the complex processing of cryptographic operations traditionally performed by electronic circuits with optical cryptographic operations. In this article, we talk with NTT Distinguished Researcher Junko Takahashi who devised an "optical-operations technique" supporting the IOWN era and implemented "optical cryptographic circuits." We hear about the present state of research in optical cryptographic technology and the future outlook for this research.

Keywords: optical cryptographic circuit, All-Photonics Network, photonic disaggregated computing platform

Achieving a secure APN through innovative cryptographic technology

—Dr. Takahashi, can you first explain "optical cryptographic circuit technology consisting of optical logic gates" for us?

"Optical cryptographic circuit technology consisting of optical logic gates" that I am researching is technology that achieves cryptographic operations traditionally performed electronically by optical circuits. The catalyst for beginning this research was NTT's major promotion of its Innovative Optical and Wireless Network (IOWN) vision in the years 2019–2020. At that time, I thought "Couldn't innovative security technologies including cryptographic technology be created by applying optical technology, the key to IOWN, in the research of cryptography and security technology that I had so far been involved in?" This was the origin of my research theme.

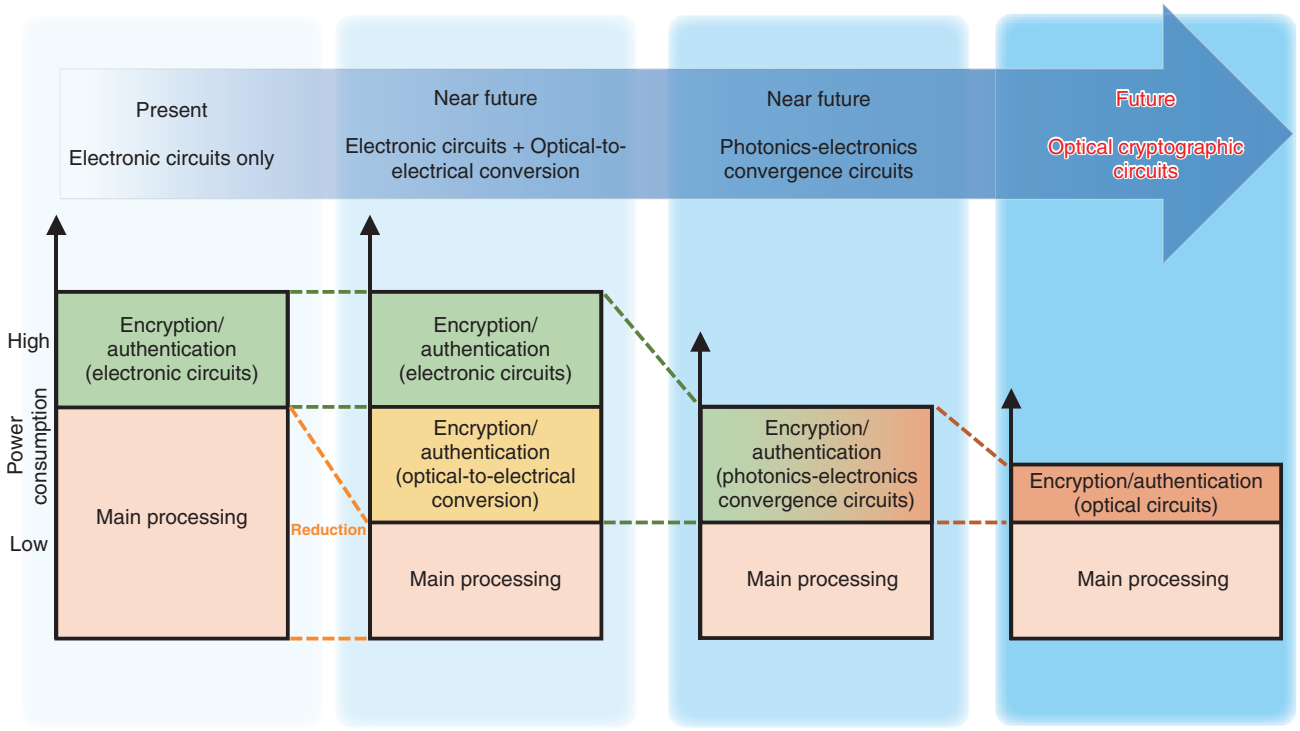


Fig. 1. Research overview of optical cryptographic circuits.

In the past, it was common sense to execute complex processing like cryptographic operations electronically. In IOWN, however, the aim is to achieve low latency and low power consumption by replacing the computing environment including servers and all sorts of electronic devices with devices based on optical technology and interconnecting them by an optical network. Achieving the processing of operations by optical technology, however, is not without its problems. Specifically, if only cryptographic operations were to be implemented on electronic circuits, the result would be high overhead (load) due to optical-to-electrical conversion causing a drop in operation performance throughout the computing environment. My new research therefore began with "optical cryptographic circuit technology consisting of optical logic gates." Thinking that operations with even lower latency and lower power consumption could be achieved and the load on the environment decreased by implementing even cryptographic operations using light in the All-Photonics Network (APN) of the IOWN vision, I began to fabricate and research cryptographic circuits using optical circuits (**Fig. 1**).

—What kinds of difficulties have you encountered in actual research?

On starting out with my research, I thought that I could implement optical cryptographic circuits simply by combining existing optical-operations devices. I assumed that this could be easily achieved in about the same period as the design and fabrication of electronic circuits. In actuality, however, I found in the course of my research that there were a variety of physical problems different from theory. To begin with, optical-operations devices necessary for cryptographic operations were still in the development stage throughout the world, and even the devices that were ready for use had limitations in how they could be used.

Achieving fundamental cryptographic techniques on optical circuits requires that cryptographic processing be configured with multiple logical operations. However, optical-operations devices that can presently be used have limitations such as the inability of performing multiple logical operations. In addition, there are devices that are still in the development stage in terms of performance requirements such as latency and power consumption. So the present state

of affairs is that current devices cannot at all meet the numerical targets set for APN. Therefore, one challenging point in my research was how to configure existing complex cryptographic functions using optical-operations devices that are limited in their operation. Additionally, peripheral circuits such as memory and registers required for executing cryptographic operations have not yet been developed with optical technology, so it was necessary for me to implement them on electronic circuits. Therefore, another challenging point in terms of achieving low latency and low power consumption was how to demonstrate maximum performance by combining optical circuits and peripheral circuits in some way.

Furthermore, in terms of fabricating optical devices, all processing including the drafting of designs and fabrication of circuit chips and peripheral circuits would take a very long time in units of years, so that also was a major obstacle that I had not expected until I began this research. For this reason, I had to search out devices that could be manufactured in a relatively short period of time and from which I could extract good performance. Circuit chips, moreover, could not be modified once fabricated, so there was no going back in this regard, and as a result, I had to carefully consider how best to configure the circuits and what method to use in manufacturing them. Finally, using optical-operations devices with good performance had a tendency to expand the overall circuit scale, so determining how to minimize this also became a major issue.

—Please tell us about some of your research achievements to date.

In research up to the present, I have devised an operations technique that achieves a nonlinear operation—an important type of function in cryptographic operations—using a technique called "one-hot encoding." Specifically, I have fabricated an optical cryptographic circuit based on silicon photonics technology targeting a function that performs 4-bit input/output table conversion. Basically speaking, this one-hot encoding technique performs operations based on optical wiring. Unlike electronic circuits, optical wiring has no wiring resistance making low-latency operations possible. In addition, while ordinary cryptographic operations use bit representations ('0' or '1'), such bit representations must be converted to hexadecimal (base 16) representations to use the one-hot encoding technique. Additionally, by using Mach-Zehnder optical modulators here, which among optical-operations devices are relatively easy to implement, and by combining these optical modulators with optical wiring, I have implemented an optical circuit chip having a cryptographic nonlinear function as a world's first.

To test the operations performed by this optical circuit chip, I also fabricated an optical device that combines the optical circuit chip with an electrical section and confirmed that the nonlinear function could indeed operate correctly. This optical device features the ability of performing an operation with extremely low latency on the order or several tens of picoseconds after the electrical signals controlling the optical modulators (input signals to the nonlinear function) are input, which I believe will make a big contribution to meeting APN targets.

"Light" shines exactly on places that appear to be unrelated to research

—Please tell us your outlook for this research going forward.

In circuits that have so far been researched and developed, input is made up of electrical signals while output consists of an optical signal, which means a photonics-electronics convergence circuit. Today, however, with the aim of achieving low-latency and low-power operations, I am in the process of fabricating a new type of optical circuit chip achieving a nonlinear function having optical signals as both input and output using optical-operations devices called Y-gates and Ψ-gates that enable logical operations. I confirmed through experiment that this optical circuit as well could perform operations correctly, so I feel that we are making steady progress toward APN of the future.

As for future developments, considering that a "photonic disaggregated computing platform" featuring distributed hardware resources is scheduled to be established in the 2030s, my aim is to achieve a computing environment that implements optical cryptographic operations circuits so that anyone can safely use this new communications platform (**Fig. 2**). In terms of numerical targets, I am working daily in my research to fabricate circuits that can operate at "1/1000 the latency" and "100 times more energy efficiency" of conventional electrical cryptographic operations circuits. Further into the future, moreover, I would like to fabricate optical cryptographic circuits that can be used not only in a terrestrial computing environment but also in a computing environment

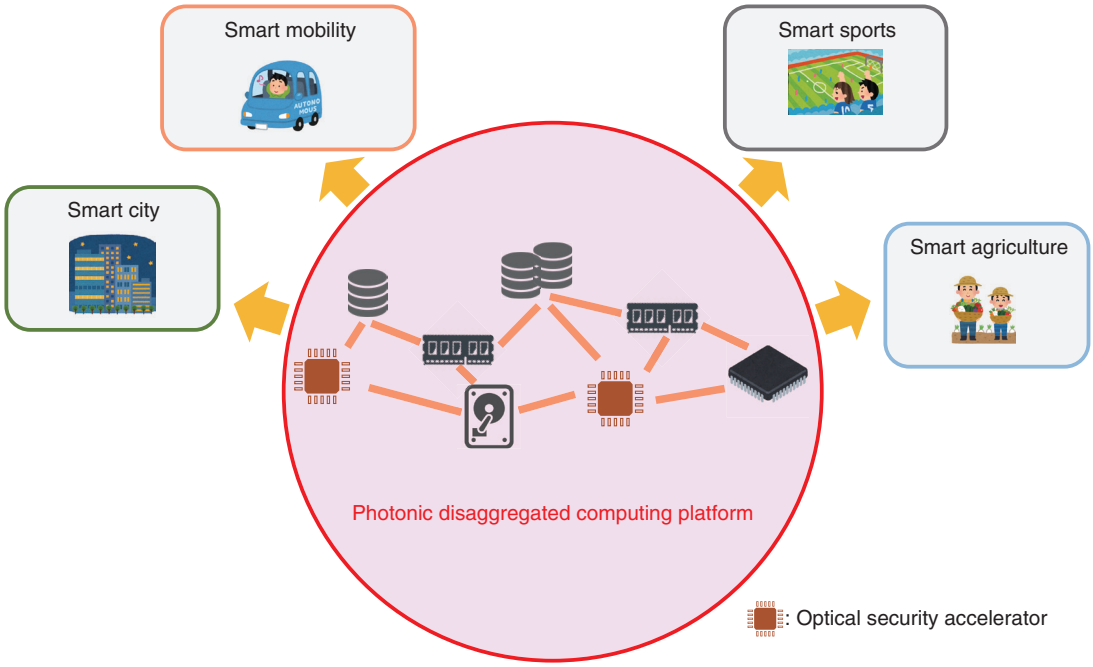


Fig. 2. The future envisioned: contributing to a safe-and-secure, low-latency, and low-power computing platform.

in space (Space Integrated Computing Network) and to develop them into a technology that will make a major contribution to new next-generation communications platforms.

The research that I pursued was unprecedented, and reactions from the outside when I started out were all skeptical in the manner of "This is like magic—is it even possible?" For sure, in 2019 when I began this research, people who thought that only the network portion of the communications platform consisted of optical communications were in the majority, and devices that could implement some operations with light were still in the development stage. Under these conditions, it was natural to have doubts about replacing cryptographic circuits originally designed with digital circuits with light. However, when demonstrating for the first time that cryptographic operations could be performed with light and announcing that achievement, I received favorable responses from many people, and I remember how happy I was that "I was able to change even if only a little the common belief that such a thing was impossible." At present, it is still not possible to operate all cryptographic circuits by light, but I would like to take on research that even more people agree with by gradually demonstrating the possibilities of optical circuits.

—What is your impression of NTT laboratories?

NTT Social Informatics Laboratories that I belong to deals not only with basic cryptographic technology but also with a wide range of research themes related to social systems. These include security-related research of cyber-attack countermeasures technology targeting attacks made against the network and Internet-of-Things systems and of data protection technologies such as secure computation technology and artificial intelligence security, research aiming for the well-being of everyone, and research related to legal systems and ethics. I myself was engaged in the research of cyber-attack countermeasures technology separate from the research of cryptographic technology for about six years before beginning my present research. In this way, I can get involved in a variety of research themes related to social systems at NTT Social Informatics Laboratories. I can even get involved in cyber-security research after completing my work in cryptographic-related research themes and then return once again to cryptography. In short, NTT Social Informatics Laboratories creates an unconstrained, relaxed atmosphere in which a researcher can move in and out of different research themes. For researchers having interests in many fields, I feel that the Laboratories provides a very

attractive environment.

I also feel that a strong point of NTT Social Informatics Laboratories that deals with a variety of research themes is the ability to easily consult with nearby researchers whenever encountering a problem outside of one's specialty during research. I once encountered a problem related to legal matters and ethics in my research and I was faced with a situation in which I had to interpret this problem. Since this was naturally outside of my specialty, I was very distressed, but since a researcher who dealt with research themes related to the law and ethics was nearby at the laboratory that I belonged to at that time, I was able to consult with that person immediately on how to solve my problem and to continue with my research smoothly. In this way, providing a means of quickly solving problems outside of one's range of research when a problem arises is truly an NTT strong point. I am very grateful for this environment that allows me to get on with my research with a sense of speed.

—Dr. Takahashi, please leave us with a message for researchers, students, and business partners.

Generally speaking, we can think of the work of research as mainly consisting of writing papers, making presentations at academic societies, and submitting patent applications, as well as the development of research results into business applications. However, ever since I entered the company, I've also placed importance on everything I can do on my own, that is, processes outside of research that naturally arise in the course of conducting research. For example, while working on cyber-attack countermeasures technology for cars several years ago, we had to evaluate the characteristic functions of cars from the viewpoint of security, and that required that we get our hands on an automobile manufactured in a specific period having certain functions related to driving. Since such an automobile was of a model no longer sold as a new car, we had to contact used-car dealers all over Japan and eventually found one that had what we needed, so we went out there and conducted negotiations and procedures ourselves. And after that, we of course had to maintain the car ourselves on a daily basis. We also selected various types of equipment that we needed to carry out this research, made a field trip to the only place in Japan where such equipment was currently installed to gather information, and did everything ourselves that was needed for installing the equipment at NTT including negotiations with installers and in-house procedures related to installation work.

Such a process, to be sure, does not appear at first glance to be directly related to the work of research. However, I think that taking on everything by the sweat of one's brow can reveal the backside of an industry and bring the overall picture into view. In particular, when jumping into a research field of an industry different from one's own research field, one must learn the implicit features of that industry such as what is taken to be "common sense" and what "manners" are practiced. Armed with this knowledge, it often becomes easier to interact with people in another field. For this reason, I always "do everything myself to the degree possible" in my research work.

Research of optical cryptographic circuits using optical logic gates that I am now working on requires many processes. In addition to theoretical studies of cryptography, it involves the fabrication of optical circuit chips and optical devices based on electrical circuits, the fabrication of peripheral devices, and construction of environments for performing proof-of-principle experiments on circuit chips. Since I cannot do all of this on my own, I greatly appreciate the involvement of many people who help me to move my research forward. It is very reassuring for me that NTT laboratories, which bring together many specialists in fields ranging from theory to devices, will continue to provide an environment in which I can give shape to my own ideas and thoughts. In such an environment, I would like to work with as many researchers, students, and business partners as possible in providing base technologies for configuring a secure information infrastructure of the IOWN era. To any readers who are interested in this endeavor, I would say, "Let's work together in taking on the challenges of the future."

■ **Interviewee profile**

Junko Takahashi received her B.S. in physics in 2004 and M.S. in physics and applied physics in 2006 from Waseda University. She entered NTT in 2006 and received her Ph.D. in engineering from the Graduate School of Informatics and Engineering, University of Electro-Communications in 2012. Her specialty is hardware security. She has previously been engaged in side-channel-attack countermeasures technology, automobile security technology, and embedded security technology. She is presently engaged in the research of optical cryptographic circuits at NTT Social Informatics Laboratories. She is an Invited Researcher at the Cyber Physical Security Research Center, National Institute of Advanced Industrial Science and Technology (AIST). She is a member of the Institute of Electronics, Information and Communication Engineers (IEICE) and Information Processing Society of Japan (IPSJ). She has been an NTT Distinguished Researcher since 2023.

Materials and Sensing Technologies for Constructing On-chip Biological Models

Masumi Yamaguchi, Aya Tanaka, and Kazuhide Kumakura

Abstract

By creating unique process technologies of bio-friendly soft materials, such as hydrogels, and combining them with biomaterials, such as cells, the breadth of research and the potential for application of these technologies have dramatically increased. Under NTT's Medical and Health Vision, new basic technologies are being developed to contribute to medical care and medicine. This article introduces the latest developments in bio-soft materials research at NTT Basic Research Laboratories.

Keywords: soft materials, cells, on-chip biological models

1. Bio-soft materials research at NTT Basic Research Laboratories

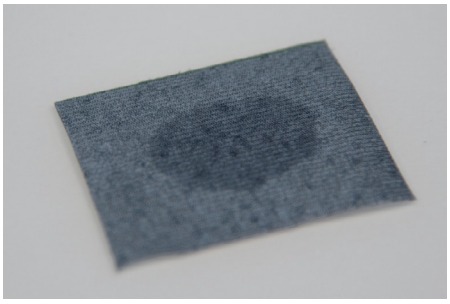
NTT Basic Research Laboratories (NTT BRL) has been conducting research on biointerfaces targeting information communication with the brain since the 1980s. Research on nano-bio devices combined with microfabrication technologies, such as semiconductors, and live imaging of biomolecules using a high-speed scanning probe microscope has been carried out [1]. NTT announced its Medical and Health Vision "Realization of Bio-Digital Twin" in November 2020 and has started developing medical and health support technology [2]. A bio-digital twin is a technology that creates a map of the body and mind in a virtual space to predict and simulate diseases. In addition to data science, NTT's bio-soft materials research has been incorporated into this initiative to conduct basic research on devices such as bio-micro robots for diagnosis and treatment of the human body and organ-on-a-chips that reproduce the functions of the human body and organs on a device chip. This is an initiative that involves engineers and researchers in a wide range of fields, including devices, bioscience, and materials science, and is not limited to researchers in medicine and information science. For those who have been engaged in basic research on bio-soft materials at NTT BRL, this activity has broadened their visions from a narrow scope of communication with the brain to a wide range of device technologies that acquire biological information and technologies that artificially mimic biological functions.

The development of the conductive fabric called hitoe™*1 is one of the first studies that led to the expansion of our vison and has its origins in the technique of coating surgical silk threads with the conductive polymer poly(3,4-ethylenedioxythiophene)-poly(styrenesulfonate) (PEDOT:PSS) in an experiment to detect neural activity in the mouse brain, which was originally conducted at NTT BRL. This fabric was developed jointly with Toray Industries, Inc. as a hydrophilic electrode material for detecting

*1 hitoe™: A conductive fabric developed jointly by Toray Industry and NTT. The fabric is made conductive by coating a nanofiber with a conductive polymer PEDOT:PSS. Since it does not use metal fibers and has a high affinity for moisture such as sweat, it has a soft texture and high adhesion to the skin. It has features such as the ability to measure biological signals in sweaty situations such as sports.

(a) Medical belt electrode

(b) Telemetric electrocardiogram (ECG) transmitter



(c) hitoe fabric that absorbs water

Fig. 1. Hydrophilic electrode materials developed from bio-soft materials research at NTT BRL are now being applied to medical device services.

electrical potentials from the human skin such as electrocardiogram and electromyogram. In 2014, it was launched as sportswear for measuring heart rate [3]. More recently, services as medical devices have been developed and are expected to be powerful tools for measuring biological data for medical care, health, and well-being [4] (**Fig. 1**).

In bio-soft materials research at NTT BRL, not only the extension of traditional technologies and research themes but also new techniques and themes are being developed and promoted. The Feature Articles "Recent Updates on Bio-soft Materials Research" in this issue highlights research results obtained at NTT BRL related to on-chip biological models that reproduce cellular and biological functions on a chip and mimic biological movements, as well as results produced in combination with state-of-the-art sensing technology.

2. On-chip biological model

The technology of materials processing has been becoming more sophisticated and finer with the progress in precision machinery and electronic devices, and at NTT laboratories, glass and semiconductor microfabrication technologies have evolved in line with the research and development of electrical and optical communication devices. In conventional bio-soft materials research, we tended to think in terms of making good use of current fabrication technologies, such as microfabrication of silicon semiconductors, to integrate with biotechnology. However, the combination of so-called hard and dry materials with advanced processing technologies is only one part of the expanding the possibilities of bio-soft materials research.

Since 2016, when the Feature Articles "Forefront Research on Bio-soft Materials" was published in this journal [1], NTT BRL has been developing original processing technologies for bio-friendly materials (materials that can be safely placed in living bodies, grow cultured cells, and serve as scaffolds for cell culture), focusing on soft materials, rather than simply current processing technology. The pioneering

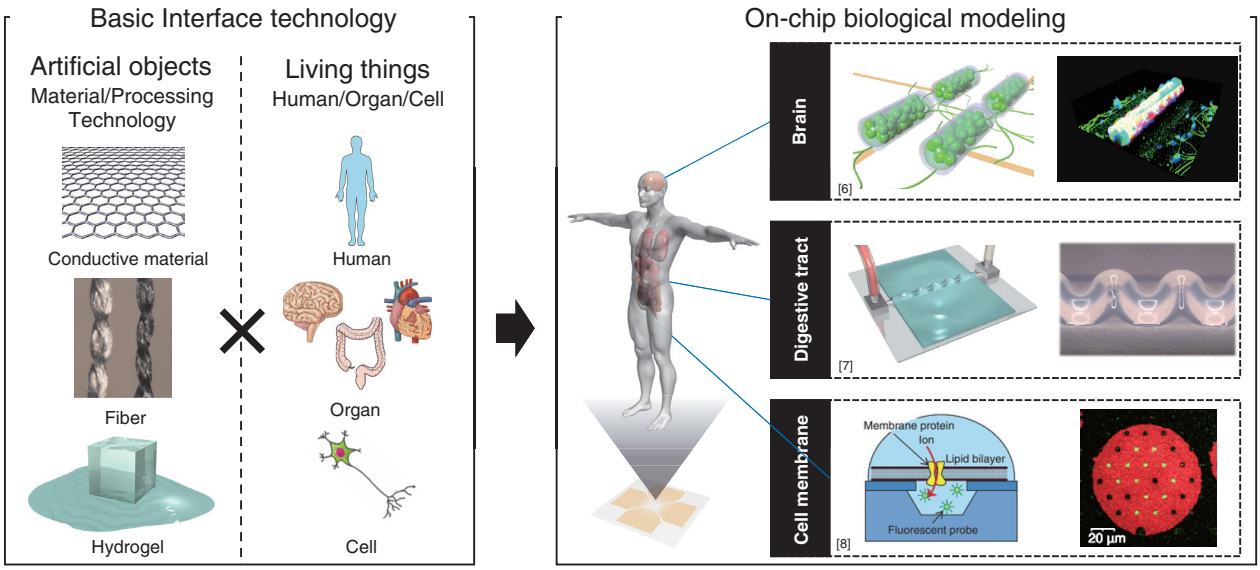


Fig. 2. On-chip biomedical model investigated at NTT BRL.

research is the three-dimensional (3D) self-assembly of the silk-gel thin-film bilayer system [5]. This research has led to the creation of the on-chip cultured brain model using the graphene-parylene bilayer system, which is introduced in this issue [6]. The controlled buckling-delamination technique using a hydrogel thin film on a glass substrate was also established [7]. The principle of these two processing techniques is unique: they spontaneously form rolls and process them into channels by successfully controlling stresses resulting from differences in physical properties at the interface of foreign materials, which are rather undesirable in semiconductor substrate growth and processing. These unique processing technologies have led to the development of basic technologies such as the fabrication of 3D electrodes for detecting biological signals and the formation of flow channels with a wide variation in thin films. In 2023, dynamic control technology was developed to mimic the movement of living organisms by light irradiation. Since all these technologies involve using highly biocompatible materials, research has progressed rapidly, such as for fabricating a cell-culture scaffold and observing the movement of three-dimensionally formed cell tissues and biological signals (**Fig. 2**).

More details can be found in each article of this issue. With the self-assembly technology of graphene/parylene, which is introduced in the article "Brain-on-a-chip Model Using Deformable Graphene-based Electrode Array" [6], cells can be wrapped without a large load by spontaneously forming a 3D roll structure using a sacrificial layer, and a large number of cell masses can be cultured and a neural network can be formed on the chip between them. We succeeded in acquiring biological signals from each cell mass using this graphene as an electrode. With this self-assembly technology of graphene, it is also possible to control the roll diameter and 3D shape by controlling the number of graphene layers and thickness of the base material.

With the buckling-delamination technology of hydrogel introduced in the article "Creation of Hydrogel Actuator toward Construction of On-chip Biological Models" [7], the flow path structure with variations is formed by controlling the adhesion pattern and hardness of hydrogel on a glass substrate, and the biomimetic motion is reproduced using photo-to-heat by synthesizing the photo-heat reactive hydrogel, thus enabling the action of blood vessel and intestine on the chip. Culturing cells on this hydrogel actuator is expected to lead to new developments in research combining the movement of the device and the flow in a channel with cultured cells.

The article "Functional Evaluation of Bilayer Lipid for the Development of Artificial Cell-membrane Structures" [8] introduces the research progress on artificial cell membranes. An artificial cell membrane is a technology for reproducing on a chip the function of membrane proteins that exist in or on the surface

of a cell membrane by covering microfabricated wells on a silicon semiconductor substrate with a lipid bilayer membrane, which is the substrate of the cell membrane. Regarding artificial-cell-membrane devices, we are studying methods to control phase-separated membranes caused by differences in lipid components on substrates and the introduction of recombinant membrane proteins by insect-cell-derived viruses. The development of a method for suppressing the electrical leakage caused by the flow of ions through the gap between the substrate and lipid film has been progressing and will be soon established as a sealing technology for electrophysiological measurement.

3. Contribution to precision medicine

A device for reproducing organ function on a chip in vitro is called an organ-on-a-chip. It is attracting attention for being useful for drug-discovery research, personalized medicine, disease modeling, etc. (**Fig. 3**). Although simple organ-on-a-chips are commercially available, they are expected not only to reproduce individual organ functions but also replace animal experiments by mimicking more complex biological systems. However, this requires various factors that have not yet been developed, such as the shape of 3D organs, movement of living organisms, transport mechanisms of blood and body fluids, and communication mechanisms of nerves.

Since some organ-on-a-chips are available commercially, they are being used for medical research. In developing technology for practical devices, it is extremely important to clarify the purpose. In other words, without a clear purpose and what measurements you want to make, design guidelines will wobble, and eventually you will end up with a device that nobody can use. We should be aware that when we link research with the creation of devices for an application, the focus of research will shift to the application or combination of existing technologies and enhancing their performance to fulfill the requirements. The purpose of our research on on-chip biological models at NTT BRL is not the development for such practical devices, rather the creation of new functions by discovering new possibilities such as structure, motion, and sensing through synthesis and processing of bio-friendly materials. This has led to the development of technology that reproduces complex biological and biomimetic functions on a chip that cannot be achieved by extending conventional technologies.

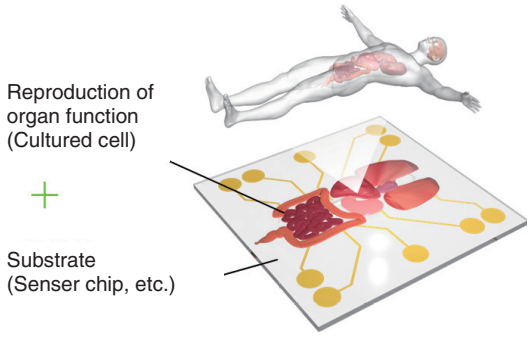


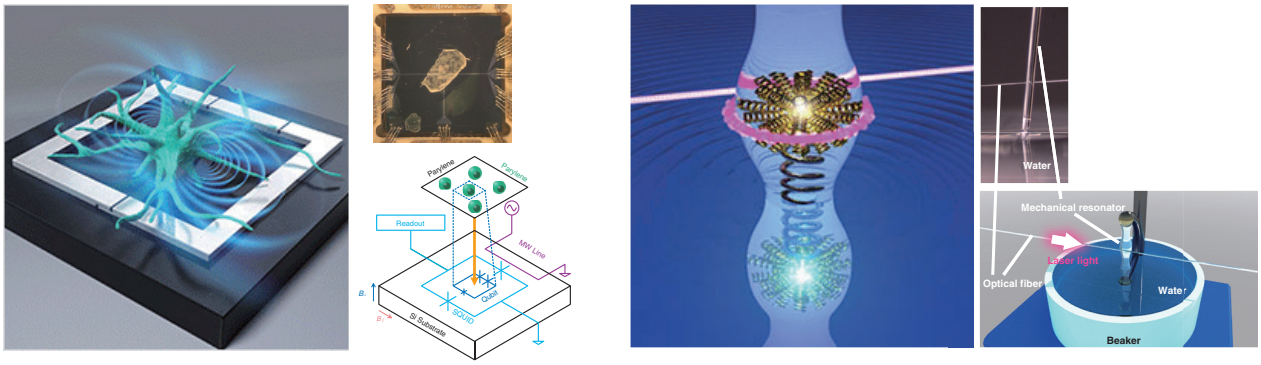
Fig. 3. Conceptual diagram of an organ-on-a-chip.

4. Collaboration with highly sensitive sensing technologies

To achieve more sensitive biosensing, NTT BRL is researching the combination of quantum technology and nano-mechanics technology (**Fig. 4**). The article "Characterization of Metal Ions in Neurons Using a Superconducting Flux Qubit" [9] in this issue introduces the sensing of iron ions in nerve cells by using a superconducting qubit. Superconducting qubits are not only the computational elements of quantum computers but are also very sensitive magnetic sensors that operate at ultra-low temperature. A superconducting flux qubit sensor can detect 20 electron spins (iron ions) and has a spatial resolution determined by its size (typically 10 to several micrometer square). This article introduces the measurement of the magnetization of nerve cells attached to a substrate. As this technology evolves, it is expected that trace metal elements can be analyzed with spatial resolution on a cell-by-cell basis. Although not introduced in this issue, research is underway on micromechanical vibrator sensors to detect the liquid properties and particles in liquids with ultra-high sensitivity [10]. High-sensitivity sensor technology, which has been used to target dry and hard materials such as metals and semiconductors, is expected to be applied to wet and soft bio areas.

5. Cherish the opportunity to encounter research partners

Even if our research topics are created from the bottom up, by connecting them with the issues of a research topic that we do not know much about, we may be able to find a new direction of using the research results and open up new applications. In



(a) Sensing of iron ions in nerve cells by using a superconducting qubit [9]

(b) Measuring particles and viscosity in liquids with ultrahigh sensitivity using micromechanical vibrator sensor [10]

Fig. 4. High-sensitivity sensing technologies are being developed as biosensors.

fact, when the hydrogel fluidics device developed using our buckling-delamination technology introduced in this issue was exhibited at NTT R&D Forum, which is held every year by NTT laboratories, a professor from a medical institute suggested that the device could be used in experiments that would not be possible with conventional device chips. This is not to say that bottom-up basic research has nothing to do with what is actually useful, only that its orientation is not toward the development for specific practical devices. We believe that the on-chip biological models and sensing technologies introduced in this article can be used in various ways. It is important not to lose sight of the essence of research by pursuing the needs that have emerged, but we would like to conduct research so that we do not miss the progress made from new encounters.

Toward effective coordination of such research, bio-soft materials research at NTT BRL promotes joint development and research in both Japan and overseas, mainly at the NTT Bio-Medical Informatics Research Center. Tetsuhiko F. Teshima, a senior research scientist at the Medical & Health Informatics Laboratories at NTT Research, Inc., who worked on 3D self-assembly of silk gel thin films and the graphene parylene bilayer system at NTT BRL, is currently conducting research based at the Technical University of Munich. In Japan, we also participate in World Premier International Premium Research Institute for Human Metaverse Medicine (WPI-PRIMe)*2 based at Osaka University. We will continue to carry out our research to open the frontiers of bio-soft materials by valuing new encounters in Japan and overseas.

References

[1] Feature Articles: "Forefront Research on Bio-soft Materials," NTT Technical Review, Vol. 14, No. 8, pp. 7–43, 2016. https://ntt-review.jp/archive/2016/201608.html
[2] Feature Articles: "NTT's Medical and Health Vision toward Creation of Bio-digital Twin," NTT Technical Review, Vol. 19, No. 7, pp. 17–51, 2021. https://ntt-review.jp/archive/2021/202107.html
[3] T. Teshima, S. Tsukada, and H. Nakashima, "Conductive Composite Material for Vital Data Measurement," NTT Technical Review, Vol. 14, No. 8, pp. 38–43, 2016. https://doi.org/10.53829/ntr201608fa7
[4] Special Feature: "Expansion of Business in the Healthcare Field and Full-scale Entry into the Medical Field by NTT TechnoCross," Monthly Business Communication, Vol. 60, No. 8, pp. 57–64, 2023 (in Japanese).
[5] T. Teshima, H. Nakashima, Y. Ueno, S. Sasaki, C. S. Henderson, and S. Tsukada, "Cell Encapsulation and 3D Self-assembly Using Multi-layered Polymeric Thin Films," NTT Technical Review, Vol. 16, No. 8, pp. 53–61, 2018. https://doi.org/10.53829/ntr201808ra1
[6] K. Sakai, T. F. Teshima, and T. Goto, "Brain-on-a-chip Model Using Deformable Graphene-based Electrode Array," NTT Technical Review, Vol. 22, No. 5, pp. 31–38, May 2024. https://ntt-review.jp/archive/ntttechnical.php?contents=ntr202405fa3.html
[7] R. Takahashi, "Creation of Hydrogel Actuator toward Construction of On-chip Biological Models," NTT Technical Review, Vol. 22, No. 5, pp. 24–30, May 2024. https://ntt-review.jp/archive/ntttechnical.php?contents=ntr202405fa2.html
[8] A. Oshima and Y. Kashimura, "Functional Evaluation of Bilayer Lipid for the Development of Artificial Cell-membrane Structures," NTT Technical Review, Vol. 22, No. 5, pp. 39–45, May 2024. https://ntt-review.jp/archive/ntttechnical.php?contents=ntr202405fa4.html
[9] H. Toida, K. Sakai, T. F. Teshima, K. Kakuyanagi, I. Mahboob, and S. Saito, "Characterization of Metal Ions in Neurons Using a Superconducting Flux Qubit," NTT Technical Review, Vol. 22, No. 5, pp. 46–51, May 2024. https://ntt-review.jp/archive/ntttechnical.php?contents=ntr202405fa5.html
[10] NTT press release, "Using light to measure particles and viscosity in liquids with ultrahigh sensitivity - Toward application as sensors in

*2 WPI PRIMe: A research center at Osaka University as part of the World Premier International Research Center Initiative (WPI) sponsored by the Ministry of Education, Culture, Sports, Science and Technology.

the chemical, biology, and rheology fields that can measure at desired locations in liquids -," Nov. 3, 2022. https://group.ntt/en/newsrelease/2022/11/03/221103a.html

Masumi Yamaguchi
Senior Research Scientist, Supervisor, Group Leader of Molecular and Bio Science Research Group of NTT Basic Research Laboratories, Bio-Medical Informatics Research Center and Kashino Diverse Brain Research Laboratory of NTT Communication Science Laboratories.

He received a B.S. in physics from Kanazawa University in 1998 and M.S. and Ph.D. in physics from Kyoto University in 2000 and 2003. He joined NTT Basic Research Laboratories in 2003. His research interests range from basic devices that enrich people's health, well-being, and communication to their social application. He is a member of the Physical Society of Japan, the Japan Society of Applied Physics (JSAP), and the Institute of Electronics, Information and Communication Engineers (IEICE).

Aya Tanaka
Senior Research Scientist, Multidisciplinary Materials Design and Science Laboratory, Molecular and Bio Science Research Group, NTT Basic Research Laboratories and Bio-Medical Informatics Research Center.

She received a B.S., M.S., and Ph.D. in chemistry from Hokkaido University in 2004, 2005, and 2008. From 2008 to 2009, she was a postdoctoral fellow at Newcastle University, UK. In 2009, she joined NTT Basic Research Laboratories. Her research interests are to understand the mechanical and topographical interactions of cells and the extracellular matrix. A major focus of her current research is the study and development of soft materials with well-defined microstructures for biological applications. She is a member of the Society of Polymer Science, Japan, the Chemical Society of Japan, and JSAP.

Kazuhide Kumakura
Vice President, Head of NTT Basic Research Laboratories.

He received a B.E., M.E., and Ph.D. in electrical engineering from Hokkaido University in 1993, 1995, and 1998. He joined NTT Basic Research Laboratories in 1998. His principal research fields are thin film growth and device applications of semiconductors. From 2007 to 2008, he was a visiting scientist at Paul Drude Institute for Solid State Electronics, Germany. He is a member of JSAP, the Electrochemical Society of Japan, and the Materials Research Society.

Creation of Hydrogel Actuator toward Construction of On-chip Biological Models

Riku Takahashi

Abstract

To express and maintain organ-like higher-order functions in cultured cells, it is important to create an environment that is close to that of a living organism. NTT Basic Research Laboratories has established a technology to control the shape of hydrogel, a material with properties similar to those of living tissue, on a chip into forming thin-film and tubular architectures that resemble living organs. On the basis of the above-mentioned technology, this article introduces the development of an actuator that can move large, fast, and arbitrary areas by optical stimulation and a technology to control its motion as if it were a living organ.

Keywords: hydrogel, actuator, organ-on-a-chip

1. Reproduction of in vivo-like environment on a chip

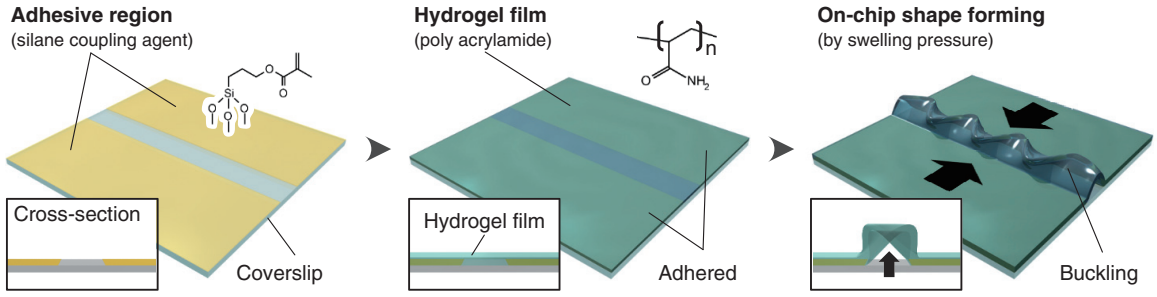
Assay platforms[*1] using cultured cells have attracted attention as an alternative to animal testing in the development of cosmetics, pharmaceuticals, agrochemicals, and other products. The use of human cells is expected to be an attractive tool that does not require consideration of species-specific differences in response, not to mention the elimination of animal ethics issues. However, to conduct appropriate evaluation without the use of living organisms, it is necessary to reproduce in vivo functions that cannot be achieved with conventional culture methods. Therefore, microphysiological systems have been extensively studied such as organ-on-a-chip technology for artificially reproducing conditions similar to those in vivo and allows cultured cells to express and maintain advanced organ functions [1]. If we can construct an on-chip biological model that reproduces organ functions on a sensor substrate based on these technologies, it is expected to be possible to acquire a variety of information on various organs as detailed data with cellular-level resolution. In addition, the use of induced pluripotent stem (iPS) cells[*2] has unlimited potential in acquiring and analyzing biological data reflecting racial differences and individual genome diversity[*3] at the molecular, tissue, and organ levels. To construct this on-chip biological model, the challenge is how to make the environment in which cells are cultured as close to that of in vivo as possible. For this purpose, technologies that can simultaneously help achieve bio-friendly materials, a shape close to that of the living tissue, and an in vivo stimulation environment are required.

Therefore, NTT Basic Research Laboratories has focused on hydrogel, a material with high biocompatibility, and has been developing a technology to control the material's three-dimensional (3D) shape to mimic that of a living tissue. Hydrogel is a jelly-like material in which a large amount of water is trapped in a polymer network and used in contact lenses, diapers, and other products around us. In our bodies,

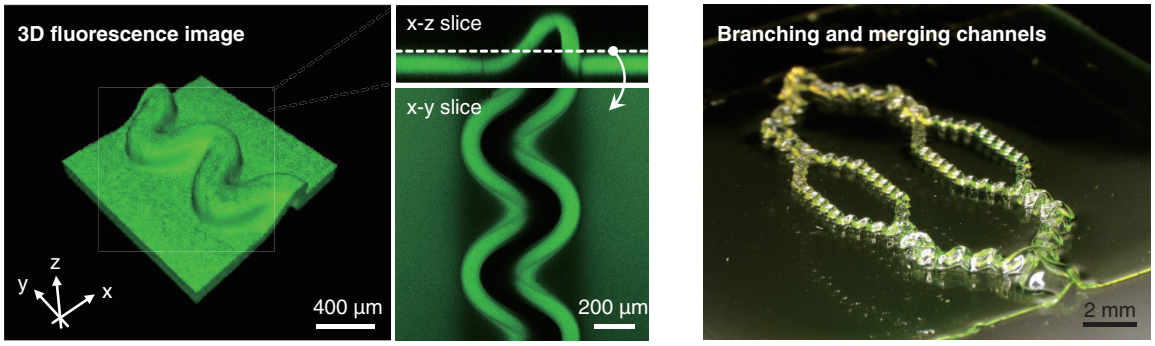
*1 Assay platform: A principle or method that enables multifaceted evaluation of compounds in drug-discovery research.

*2 iPS cells: Cells with the potential to change into all types of cells and the ability to grow almost indefinitely, which are created by introducing specific genes into human somatic cells.

*3 Genome diversity: The genome, which is all the genetic information expressed in a string of DNA, differs in sequence by about 0.1% among individuals, resulting in diversity.



(a) On-chip shape formation by swelling pressure of hydrogel film



(b) Observation of 3D architecture by fluorescence microscopy

(c) Shape control by adhesion pattern

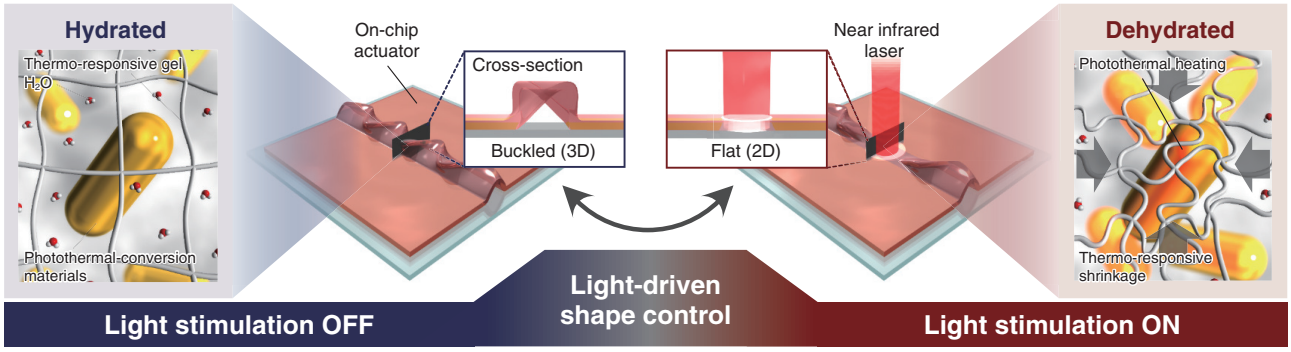
Fig. 1. On-chip shape-control technology for hydrogel films.

there are also network structures of biological macromolecules and proteins and they retain water. In other words, our body is also a hydrogel and one of the most suitable materials as a scaffold for cell growth. However, the volume of the hydrogel changes depending on the amount of water in the network, and the material is very soft and brittle due to high water content, making it difficult to control its shape on the substrate. Therefore, we reported an on-chip shape-control technology that uses the force of hydrogel to swell by absorbing water to form a thin and hollow tubular shape like a living organ[*4] such as a blood vessel or intestinal tube [2]. An overview of this technology is shown in **Fig. 1**. According to the fabrication scheme shown in Fig. 1(a), the surface of a glass substrate is first treated with a silane coupling agent that chemically bonds to the hydrogel, and a treated/untreated pattern is formed using general lithography techniques[*5] (a rectangular pattern is shown in the example). The hydrogel thin film is then synthesized on the substrate, resulting in a hydrogel thin film bonded to the glass substrate in a patterned manner. Subsequently, the hydrogel is immersed in water, which absorbs water and generates pressure due to swelling. As a result, a physical phenomenon called buckle-delamination[*6] occurs in the non-adhered area to release pressure from the surroundings, forming a thin-film/tubular channel structure consisting of a hydrogel thin film and glass substrate. Figure 1(b) shows the result from fluorescent observation of a 3D structure. The obtained structure can be controlled with physical parameters, such as hydrogel hardness, thickness, water content, and geometric pattern of the non-adhered area [3], and can reproduce smooth 3D structures such as folds and wrinkles observed in

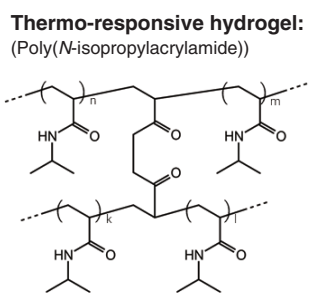
*4 Organ: A unit that constitutes the body of a multicellular organism and is a collection of tissues that cooperate to perform a certain function.

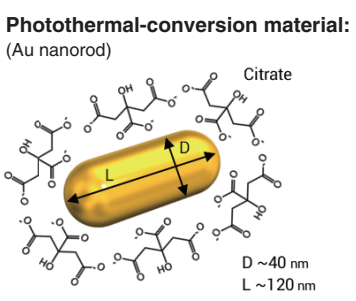
*5 Lithography technique: A technique to form a pattern by applying a photosensitive material, exposing it to light and developing it. It is used in the manufacture of semiconductor integrated circuits and various other products.

*6 Buckle-delamination: When a thin film is compressed from both ends, the center of the film is lifted while curving, and the film is deformed into an arch-like structure. This is the buckle-delamination phenomenon in which the film peels off while bending (buckling).

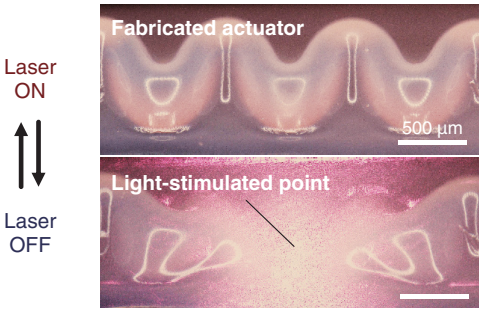


(a) Actuation mechanism of light-driven actuator by composite of thermo-responsive hydrogel and photothermal-conversion materials





(b) Materials comprising photoresponsive hydrogels



(c) Actuation behavior during light stimulation

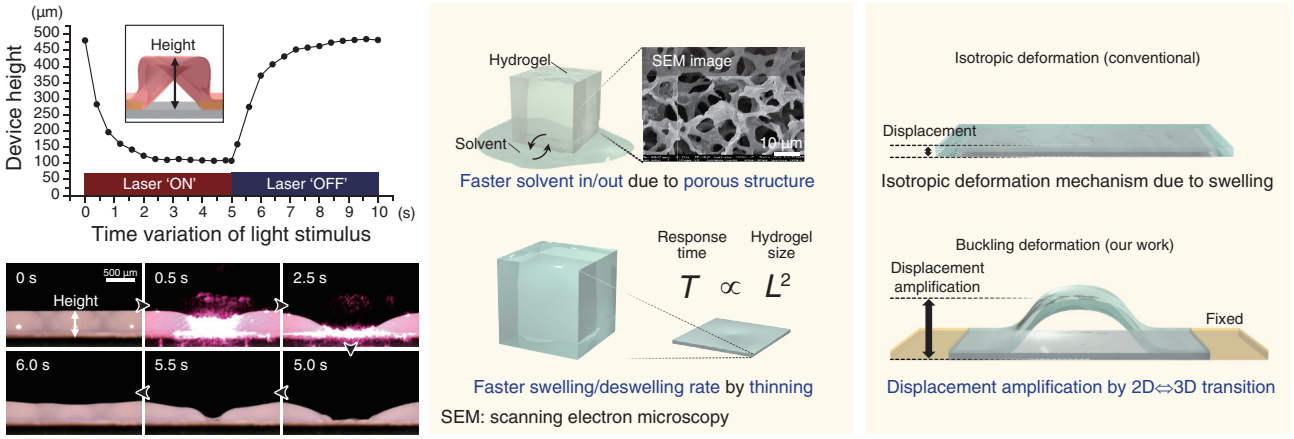
Fig. 2. Concept of light-driven on-chip hydrogel actuator.

vivo. Characteristic channel-like structures can also be freely designed by controlling the adhesion pattern, making it possible to fabricate branching and merging channels that mimic a vascular network (Fig. 1(c)).

2. Hydrogel actuator with both large deformation and fast response

The technologies introduced thus far have enabled us to fabricate a shape similar to that of a living organism on a chip using bio-friendly materials. However, to reproduce the dynamic stimulation environment similar to in vivo, it is necessary to have a technology that can "move" the shape of the hydrogel thin film. A simple method is to connect the channel structure made of the hydrogel film to an external pump to control the flow pressure. The extremely soft structure of this channel can deform significantly in response to pressure, making it possible to reproduce the pulsating contraction motion of blood vessels [4]. However, there are problems with this method, such as uniform deformation of the entire structure and the need to connect an external device, so a method for easily reproducing complex movements similar to those of a living organ is required. Therefore, we have designed and fabricated a hydrogel thin film that can be deformed by light stimulation, then proposed a light-driven actuator*7 that adapts to on-chip shape-control technology [5]. The device fabricated in this study combines a temperature-responsive hydrogel that expels water and contracts in response to temperature and a photothermal-conversion material that generates heat upon light irradiation, enabling the actuator to switch between a hydrated and dehydrated state by turning light stimulation on and off (**Fig. 2(a)**). The temperature-responsive hydrogel is a Polyisopropylacrylamide gel*8, which is also used as a base material for cell culture, and the photothermal-conversion material is a gold nanorod*9 that efficiently

*7 Actuator: A device that converts energy into motion.
*8 Polyisopropylacrylamide gel: A hydrogel that can shrink in response to temperature while expelling retained water. It is biocompatible and used as a cell culture substrate.



(a) High-speed and large deformation actuation (b) Key points regarding fast response (c) Key points regarding large deformation

Fig. 3. High-performance actuator with both high-speed response and large deformation.

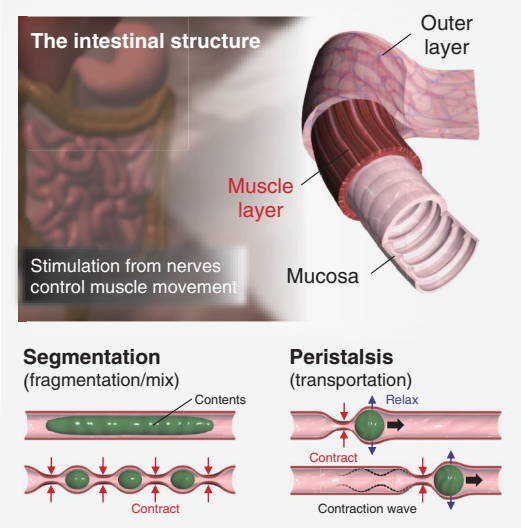
generates heat when irradiated with near-infrared light (**Fig. 2(b)**). Therefore, by irradiating the buckling structure of the hydrogel thin film on the chip with a near-infrared laser, it is possible to control the shape of the local buckling state and flat state. The actuation behavior is shown in **Fig. 2(c)**. When the 3D tubular structure in the buckled state at the center of the image was irradiated with a near-infrared laser, it immediately deformed to a flat state with a completely collapsed tube, and rapidly recovered to its original buckled state again when the optical stimulation was stopped. To investigate this deformation behavior in more detail, we observed the actuator from the side and plotted the change in height of the actuator over time during light stimulation, as shown in **Fig. 3(a)**. Surprisingly, the actuator shifted to a flat state within 2 seconds after the start of light stimulation and recovered to a buckled shape within about 2 seconds after the end of light stimulation. In addition, the actuation-deformation ratio (the ratio of the amount of deformation estimated from the maximum and minimum heights) of the actuator reached approximately 360%. This high-speed response and large deformation actuation is much higher than that of conventional light-driven hydrogel actuators.

The following is a summary of the key points of fast response and large deformation. **Figure 3(b)** shows the structure that enabled fast response. Conventional hydrogels have a network structure with a mesh size of several 10 nm, and water molecules move through these holes. The hydrogel used in this study has a special fabrication method that produces a network of pores several micrometers in diameter (porous structure). In other words, the hydrogel has a structure with pores nearly 1000 times larger than those of conventional hydrogels, allowing water to flow in and out easily like a sponge, and the response to volume change accompanying solvent transfer is faster. It is also known that the response time increases in proportion to the square of the hydrogel size. We used the hydrogel thin film of about 100 μm as the actuator, which enabled swelling and de-swelling in a shorter time than thicker hydrogels.

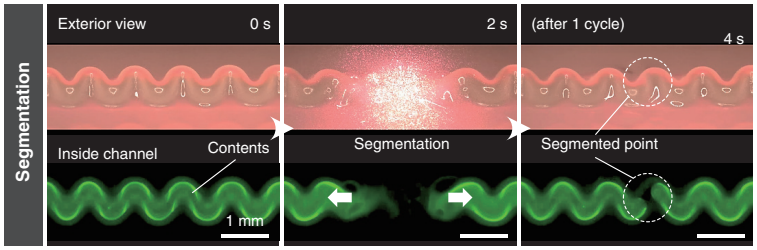
Figure 3(c) shows the point of deformation that enables large deformation. When hydrogel swells by absorbing water or shrinks by expelling water, the deformation is typically isotropic, with all directions changing at the same ratio. We applied on-chip shape-control technology with which a portion of the hydrogel is fixed (restrained) to the substrate during swelling, resulting in 3D deformation due to buckling in the film-thickness direction. This deformation is large in the film thickness direction, acting as a displacement-amplification mechanism. This allows displacement that is approximately 10 times larger than the deformation ratio for each direction of normal swelling deformation, enabling large deformation.

This actuator is able to operate stably in air. Normally,

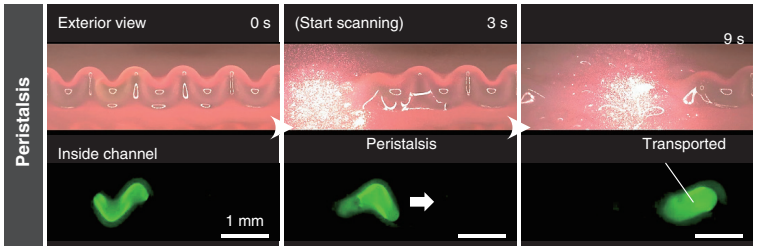
*9 Gold nanorods: Gold-derived nanomaterials shaped like an elongated capsule. When irradiated with light of an appropriate wavelength, the light energy is converted into heat (photothermal conversion material). By adjusting the length of the rod, the light-absorption peak can be controlled from 550 to 1400 nm.



(a) Characteristic motion behavior in the intestine



(b) Reproduction of segmentation by cyclic application of light stimulation



(c) Reproduction of peristalsis by scanning control of light stimulation

Fig. 4. Biomimetic actuation by controlling light stimulation.

actuators based on the swelling/de-swelling of hydrogel could only operate in the presence of water, making it difficult to use them in air for long periods. However, our actuator with a channel structure can hold water in the tube, allowing stable operation in air without drying out. It is also known that the actuation performance can be finely tuned by changing the hardness of the gel and height of the actuator. Therefore, as an on-chip actuator that can operate in various environments, it is expected to be applied in a wide range of fields, such as microfluidics*10 and soft robotics*11.

3. Demonstration of biomimetic actuation

Our actuator is superior not only in performance but also in behavior control. Because the stimulus source is light, it can be remotely irradiated to the targeted position, and there is no need to connect any external devices, such as wiring, to the actuator. The combination of materials we used shrinks rapidly when heated to 35ºC or higher. By adjusting the concentration of gold nanorods, a photothermal-conversion material, the actuator is designed to heat quickly to about 35ºC during light stimulation without excessive heating. Therefore, only the irradiated area is deformed during light stimulation, and the effect of thermal diffusion on the surrounding area is suppressed. As described above, this actuator enables precise positional and local control of deformation by light stimulation, and control of complex behaviors that mimic those of biological organs is also possible.

We focused on the characteristic movement of the intestinal tract, an overview of which is shown in **Fig. 4(a)**. The intestinal tract has a three-layered structure, consisting of an outer layer that protects organs from friction, a muscular layer that generates characteristic movements, and a mucosa that handles digestion and absorption. The two characteristic movements are controlled by the fine control of the contractile movements of the muscular layer by stimulation from nerves. Specifically, the cyclic movements of contraction and relaxation are repeated at specific points to generate segmental movements that break up and agitate the intestinal contents. The contraction and relaxation propagate as a continuous wave (contraction wave) at neighboring points,

*10 Microfluidics: A research field that deals with fluid systems in which the fluid behavior is different from the conventional scale with respect to the fluid flowing in a channel with a width of about μm.

*11 Soft robotics: A research field in robotics that deals with robots that use flexible materials and have delicate and supple movements.

which generates peristalsis that transports the intestinal contents as if whipped cream is being squeezed out.

We attempted to reproduce these two types of movements by temporally and spatially controlling light stimulation (**Fig. 4(b)**). We first injected a highly viscous oil (mineral oil), which mimics the contents of the intestine, into the actuator. The oil was colored with a green fluorescent substance so that the contents could be observed with a fluorescence microscope. The irradiation was then repeated at specific points on the actuator in 2-second cycles to switch between buckling and flat states, i.e., open/close motion of the channel structure. It was confirmed that the continuous contents were crushed and could be separated. This is because the channel structure can be completely closed, demonstrating the possibility of reproducing segmental movement by taking advantage of the hollow structure and large deformation capability of this actuator.

We then injected a small mass of the intestinal contents model described above inside the actuator. A point slightly behind the mass was then optically stimulated to collapse the channel structure, and the irradiated point was slowly shifted (scanned). We were able to reproduce the contraction wave of the collapsed channel as it moved and confirm the movement of the contents along with it. This is due to the large deformation that can close the channel structure and the positional and local control of the deformation, demonstrating that peristaltic motion can be reproduced.

Thus, it was found that complex movements of a living organ can be reproduced by using an actuator as a muscle layer and optical stimulation control as a neural control in a living organism. The motion reproduced by the actuator is comparable to that of the intestinal tract of a rat, which is a living organ of the same scale.

4. Outlook for the future

We introduced an actuator that reproduces the in vivo stimulation environment by controlling biofriendly materials to a shape similar to that of a living tissue as a fundamental technology for the fabricating on-chip biological models. It has become clear that physical stimuli (mechanical stress) inside and outside cells are involved in the regulation of various functions such as cell growth, differentiation, proliferation, and morphogenesis. Therefore, this technology, which can reproduce complex and dynamic mechanical stimuli of living organs on a chip, is expected to provide an in vivo-like environment and, in combination with cell culture, to provide a more sophisticated biological model. The actuator is designed to be efficiently driven at 35°C in air, which is different from the driving conditions suitable for cell culture (37°C in culture medium), but we confirmed that the operating temperature can be finely tuned by adjusting the material composition. In the future, we will select cultured cells on which we wish to examine the effects of mechanical stress and construct an experimental system in which the cells can be cultured during the application of stimuli.

Integration with biosensors for acquiring biological data is also a future challenge. The actuator consists of a thin hydrogel film that can mimic the in vivo environment and a support substrate made of a solid, such as glass, in a heterogeneous (two different types) structure. The support-substrate part, from which various material types can be selected, is easy to process unlike the hydrogel thin film part. Therefore, we believe that sensors can be fabricated and integrated using semiconductor processing technology, which is NTT laboratories' strength.

If these technologies mature and can reproduce various organ functions of an individual on a high-performance biosensor substrate, it will be possible to acquire personalized biological data with high accuracy and from multiple angles. This is thought to lead to the construction of a bio-digital twin*12, a model that reproduces oneself in a digital space. In particular, the ability to acquire data under conditions that are difficult to test in actual living organisms (extreme environments, drug administration, etc.) is expected to contribute to complementing the model. We will continue to accumulate fundamental technologies to construct such on-chip biological models.

References

[1] D. E. Ingber, "Reverse Engineering Human Pathophysiology with Organ-on-chips," Cell, Vol. 164, No. 6, pp. 1105–1109, 2016. https://doi.org/10.1016/j.cell.2016.02.049

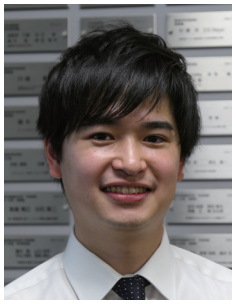
[2] R. Takahahsi, A. Tanaka, M. Yamaguchi, and Y. Koike, "Bionics Technology for the Future of Medicine and Health," NTT Technical Review, Vol. 19, No. 7, pp. 46–51, 2021. https://doi.org/10.53829/ntr202107fa6

[3] R. Takahashi, H. Miyazako, A. Tanaka, and Y. Ueno, "Dynamic Creation of 3D Hydrogel Architectures via Selective Swelling Programmed by Interfacial Bonding," ACS Appl. Mater. Interfaces, Vol. 11, No. 31, pp. 28267–28277, 2019. https://doi.org/10.1021/acsami.9b05552

*12 Bio-digital twin: A precise mapping of each person's body and mind in cyberspace and is the technology that supports the basis of NTT Medical and Health Vision.

[4] R. Takahashi, H. Miyazako, A. Tanaka, Y. Ueno, and M. Yamaguchi, "Tough, Permeable and Biocompatible Microfluidic Devices Formed through the Buckling Delamination of Soft Hydrogel Films," Lab on a Chip, Vol. 21, No. 7, pp. 1307–1317, 2021. https://doi.org/10.1039/D0LC01275K

[5] R. Takahashi, A. Tanaka, and M. Yamaguchi, "Biomorphic Actuation Driven via On-chip Buckling of Photoresponsive Hydrogel Films," Adv. Funct. Mater., Vol. 33, No. 24, p. 2300184, 2023. https://doi.org/10.1002/adfm.202300184



Riku Takahashi

Researcher, Multidisciplinary Materials Design and Science Laboratory, Molecular and Bio Science Research Group, NTT Basic Research Laboratories and Bio-Medical Informatics Research Center.

He received a B.E, M.E., and Ph.D. in polymer (hydrogel) chemistry from Hokkaido University in 2013, 2015, and 2018. From 2015 to 2018, he was awarded a Japan Society for the Promotion of Science Research Fellowship for Young Scientists (DC1) to investigate tough hydrogels. He joined NTT Basic Research Laboratories in 2018 and started his research on the development and application of on-chip shape-control methods for hydrogels. He received the certificate of award for encouragement of research at the Material Research Society of Japan (MRS-J) symposium 2019 and at the Japan Society of Applied Physics (JSAP) symposium 2023. He is a member of the MRS-J, JSAP, Society for Chemistry and Micro-Nano Systems (CHEMINAS), and the Society of Polymer Science, Japan (SPSJ).

Brain-on-a-chip Model Using Deformable Graphene-based Electrode Array

Koji Sakai, Tetsuhiko F. Teshima, and Toichiro Goto

Abstract

Since the brain transmits electrical signals between neurons in three-dimensional (3D) space, we developed a 3D deformable electrode array to measure electrical signals from a 3D network of cultured neurons using a graphene self-folding technique. By recording the electrical signals spatiotemporally from the electrode array, we demonstrated that the reconstruction of a 3D neuronal network on the electrode array (brain-on-a-chip model) enables us to investigate how neurons communicate with each other and helps us understand how the structure of the neuronal network affects the neuron's communication.

Keywords: organ-on-a-chip, neural interface, neuronal network

1. Brain-on-a-chip model

The function of living tissues is highly dependent on their structure. The tubular structure of blood vessels and intestines allows for the flow of substances, and the fibrous structure of skeletal muscles generates traction force. The brain also has structural features, including three-dimensionality and modularity. For example, the cerebrum is divided into motor, somatosensory, and visual cortices, which have different roles. Within each area, neurons are not uniformly distributed but maintained in dense clusters of cells and weakly connected clusters of cells. These weak connections between cell clusters with specific roles are called modular structures, and the brain functions require a balance of connectivity between modules that work synchronously or individually.

We investigated how these brain structural features affect neuronal synchronization at the cellular level by growing cultured neurons on an electrode substrate. Cultured neurons obtained from brain tissue can still generate electrical signals. Furthermore, cells growing on the substrate spontaneously form a neuronal network so that the communication between cells can be observed from electrodes placed at multiple points. However, cells on a flat substrate grow uniformly and in a flat plane in a structure completely different from that of the brain. Therefore, it is necessary to artificially manipulate the structure by constructing scaffolds for cultured neurons to grow and by localizing the cells to reproduce the brain structural features, including three-dimensionality or modularity. We propose a strategy in which the electrodes are used as three-dimensional (3D) scaffolds to use the electrodes to measure the electrical signals of the cells and arrange the cells to form the desired 3D structure. In this article, we introduce a 3D deformation technique of electrodes and the growth of cells in 3D structures using the electrodes as scaffolds. We then reproduce both 3D and modular neuronal networks on an electrode array by connecting 3D cell clusters (**Fig. 1**). The development of a brain-mimicking neuronal network on the chip is called a brain-on-a-chip model. After introducing the results of recorded neural signals from the brain model, we provide an overview of further development of the technology.

2. Cell encapsulation within the 3D deformable electrode

To use electrodes as 3D cell scaffolds, we developed

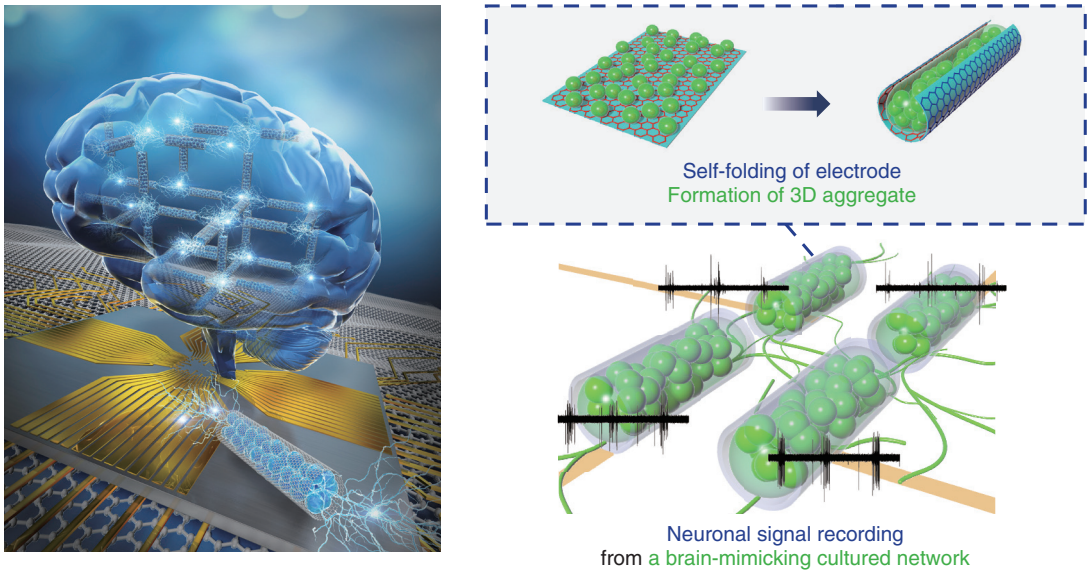
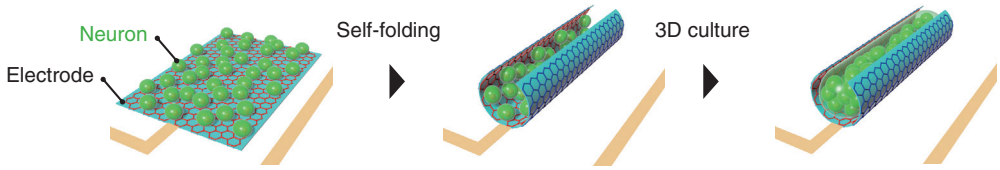


Fig. 1. Images of brain-on-a-chip model. The figure is a modification of the original paper [6] (Creative Commons License (Attribution 4.0 International)).

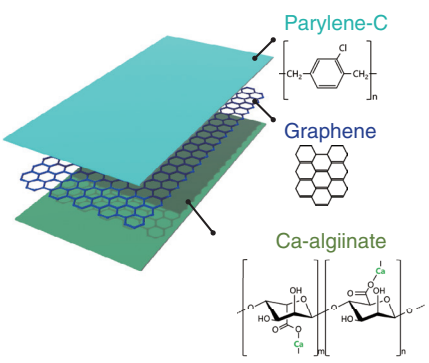
a method of folding electrodes three-dimensionally and encapsulating cells within them [1–3]. The procedure of this method is shown at the top of **Fig. 2**. The process of growing cultured cells three-dimensionally has been actively studied in tissue engineering and organ-on-a-chip technology. However, it is challenging to apply 3D culture methods to recording techniques with a flat electrode substrate (microelectrode array) because poor contact between the flat electrode and cultured 3D tissue decreasing signal amplitude. This is one of the fundamental reasons for the difficulty of developing sensing technology in organ-on-a-chip. Therefore, electrodes must have a 3D curved structure to be compliant with the surface of the 3D tissue. Current planar electrodes can be fabricated by photolithography with a spatial resolution corresponding to the size of cells (10–100 μm scale). The processing accuracy of photolithography supports biosensing technology for measuring cellular signals. However, it is challenging to form 3D electrodes with precise geometries due to the technical difficulty of manually bending precisely fabricated electrodes.

We incorporated a technique called thin-film self-assembly into an electrode array substrate to bend a precisely fabricated electrode. In a bilayer thin film of two different materials, the film folds when one of the layers is subjected to shrinkage or stretching forces. Self-assembly is a method of inducing spontaneous folding of a thin film when detached from a substrate by preparing a bilayer thin film on a substrate and applying internal stress to one of the layers in advance. Typical examples of deformation include transforming a thin rectangular film into a cylindrical shape and transforming a dice net into a cube shape. We developed a self-folding thin film composed of biocompatible materials [4]. By using these techniques and knowledge, we have further explored materials that can be used as electrodes.

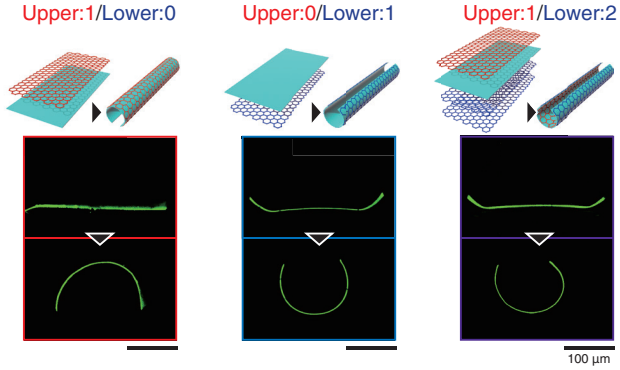
As a material for a self-folding electrode, we selected graphene consisting of one layer of carbon atoms because it is highly biocompatible, electroconductive, and flexible enough to be bent (Fig. 2(b)). Graphene has also been reported as an electrode material of brain-implantable devices. After screening several candidate materials for the self-folding thin film, we found that a polymer material called parylene-C can induce self-folding by attaching graphene to a parylene-C sheet [2, 3, 5]. As shown in Fig. 2(c), the graphene/parylene-C bilayer is flat while attached to the substrate, but once it is removed from the substrate, it folds into a cylindrical shape within a few seconds. The critical point is the high adhesion of the graphene to perylene-C. Because of the high adhesion, graphene does not slip from the perylene-C sheet during the folding, and the residual stress generated inside the film is the driving force for the folding. Parylene-C is another material widely used for



(a) Formation of 3D neuronal aggregate using self-folding graphene



(b) Material composition of self-folding electrode



(c) Dependence of folding direction on graphene position

Fig. 2. Self-folding of graphene-based thin electrode. The figure is a modification of the original paper [6] (Creative Commons License (Attribution 4.0 International)).

the insulating layer of neural electronic devices, and its excellent biocompatibility and transparency also make it suitable for microscopic cell observation. Material candidates for a sacrificial layer, which supports the bilayer thin film on the substrate, were also screened. Calcium alginate gel was selected as the material for the sacrificial layer because it does not affect the survival of cultured cells during dissolution. The calcium alginate gel can be dissolved with a reagent that removes calcium ions (Ca^{2+}) without affecting the survival of cultured cells. After adding the Ca^{2+}-chelating reagent, the thin film immediately detaches from the substrate and completes the self-folding. After stacking three layers of thin films containing calcium alginate gel, parylene-C, and graphene, the tri-layers are etched with oxygen plasma through a photolithographically patterned mask, allowing the thin films to be shaped with micrometer-scale precision. Rectangular films from 100 to 1000 µm are typically formed to obtain a cylindrical shape with a curvature radius of 10–200 µm after self-folding. The curvature radius depends on the thickness of the parylene-C film. For the bilayer of graphene and parylene-C film, the graphene becomes the outer surface of the cylindrical structure (Fig. 2(c)). When cultured cells are wrapped, they can only contact the insulating parylene-C thin film. We fabricated a sandwich structure of graphene and parylene-C thin film, as shown on the right side of Fig. 2(c), and found that this tri-layer film can also spontaneously fold. Thus, we developed a self-folding electrode with an inner conductive graphene layer [6].

Since graphene is not only biocompatible but also highly transparent, cells can be fluorescently stained and observed under a microscope to visualize the process of cell growth inside the folded electrode and the formation of 3D neuronal aggregate (**Fig. 3**). A culture medium containing neurons obtained from the rat hippocampus is dropped onto an electrode. Self-folding is then induced, resulting in the encapsulation of neurons in the folded electrode (Fig. 3(a)). Figure 3(b) shows neurons are attached sparsely inside the folded electrode on the first day of culture. They grow and elongate their process, increasing the tissue volume formed by cell aggregation. After 1–2 weeks of culture, the tissue volume increases to fill almost the entire inside of the tube, forming a 3D neuronal aggregate along the shape of the tube. After 3–4 weeks of culture, most neurons are still alive, and the 3D tissue is maintained inside the cylindrical

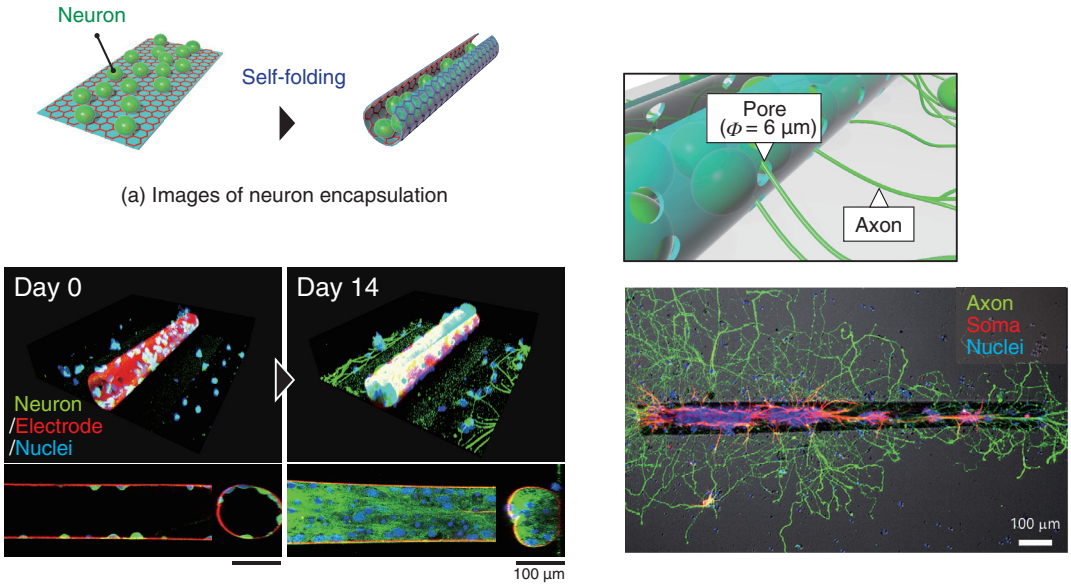


Fig. 3. Formation of 3D neuronal aggregate by neuron encapsulation. The figure is a modification of the original paper [1, 6] (Creative Commons License (Attribution 4.0 International)).

electrode. The electrode structure is modified to maintain the culture for such a long period. If the culture medium is simply decreased, cells cannot survive due to a lack of nutrients and oxygen. Long-term culture of 3D tissue requires a continuous supply of culture medium to the inside of the tissue. It is challenging to supply culture medium from the outside culture using the folded structure. Therefore, as shown in Fig. 3(c), we incorporated pores into the electrode surface to improve the material permeability [1, 2]. By using the resolution of photolithography to form tiny pores (3–6 μm), the permeability of nutrients and oxygen can be improved while keeping the neurons confined inside the tube.

The role of the pores is not only to improve permeability. Pores also function as a pathway for neurons to connect structurally to the outside of the tube. Neurons connect to other neurons by extending thin projections called axons to form a neuronal network. To achieve the modular structure described at the beginning of this article, it is essential to provide tiny pores so that 3D neuronal aggregates can be connected through axons. The bottom of Fig. 3(c) shows a stained image of an axon (green) extending outside the tube through tiny pores. The soma (cell body; red) remains inside the tube, and only the axons can be extended outside the tube and connected to other 3D neuronal aggregates. This method solves the technical problem of structural mismatch between 3D neuronal aggregates and electrodes and enables the connection of 3D neuronal aggregates by providing tiny pores for axons to pass through to form a modular neuronal network.

3. Brain-on-a-chip model using self-folding electrode array

In addition to three-dimensionality, the modular structure is another feature of the brain. We developed an electrode array chip in which multiple self-folding electrodes are aligned to form the modular structure by connecting multiple 3D culture tissues encapsulated within folded electrodes [6]. As shown in **Fig. 4**, a neuronal network is formed by arranging self-folding electrodes in an 8 × 8 multi-array of substrates and forming 3D neuronal aggregates inside each electrode, with each 3D neuronal aggregate as a module. Neurons spontaneously extend their axons, and the 3D neuronal aggregates are connected. Weak connections are formed between the neuronal aggregates compared with the dense connections between the neurons in the 3D neuronal aggregates, forming a modular structure with distinctly solid and weak connections. By independently wiring each electrode to

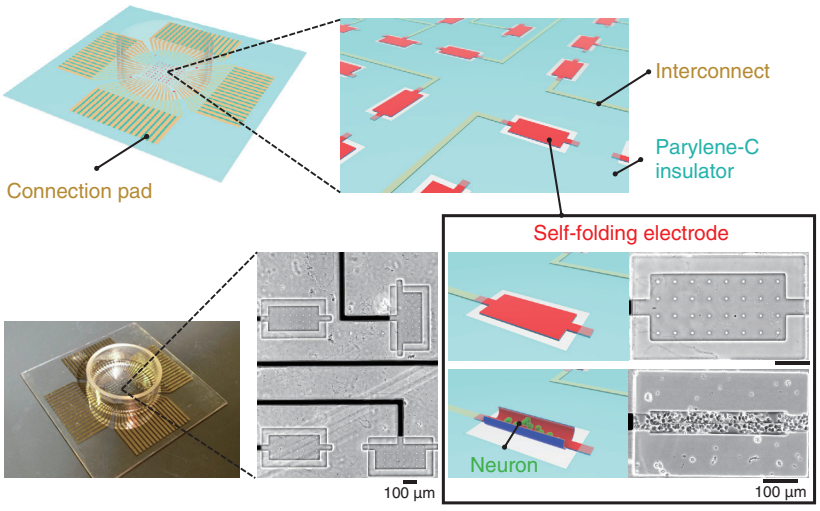


Fig. 4. Self-folding electrode array. The figure is a modification of the original paper [6] (Creative Commons License (Attribution 4.0 International)).

an external recording device (Fig. 4), neural signals exhibited from cultured 3D tissues can be recorded separately. By analyzing the synchronization and time delay of the recorded neural signals, it is possible to investigate how the multiple 3D neuronal aggregates are connected and how the neuronal network is formed. This engineered neuronal network on the chip device functions to reproduce the structural features of the brain while electrically visualizing the interactions within the neuronal network. We conducted experiments for testing two points: long-term measurement performance and the effect of the structural features of the network on the synchronization of neural firing.

We investigated long-term measurement performance by comparing the recorded signals between the self-folded 3D electrode and the existing flat electrode. The deformation of the electrode can solve the structural mismatch between the flat electrode and 3D neuronal aggregate. We evaluated how long neural signals could be stably recorded from individual 3D neuronal aggregates. **Figure 5** summarizes the recording results. As a comparison, the recorded signal decreased and disappeared as the culturing days progressed when a 3D neuronal aggregate was formed on a conventional planar electrode. In addition to the structural mismatch, the cause of decreased signals is the ability of cultured cells to migrate. As shown at the top of Fig. 5, the movement of the 3D neuronal aggregate from the flat electrode decreased the contact area, and the recorded signal gradually weakened. However, the three-dimensionally deformed tubular electrode maintained good contact between the tissue and electrode because the 3D neuronal aggregate was fixed inside the tubular electrode. The number of electrodes that could be recorded and the recorded neural firing frequency as a function of the culturing period is shown at the bottom of Fig. 5.

We then examined the relationship between the structure of the neuronal network and synchronization of neural firing. In neuronal networks, neurons fire synchronously and generate electrical signals. In modular structures, neurons are tightly connected to each other within a module that is a 3D neuronal aggregate and tend to fire synchronously, while neurons are loosely connected between different modules and are relatively less synchronized. To investigate the effect of such structural features on synchronization, we recorded spontaneous activity from networks with modular and 3D structures (3D/modular networks). We compared the recorded signals with those recorded from neuronal networks in which cells are uniformly grown on the planar substrate (2D/uniform networks). **Figure 6** shows the results of spontaneous activity recorded at each electrode as a plot of the firing time at each electrode for visualizing the synchronization between the 3D neuronal aggregates. In the 2D/uniform network, the firing pattern is dominated by the repetition of synchronized firing due to the strong connection.

In the 3D/modular network, a variety of firing

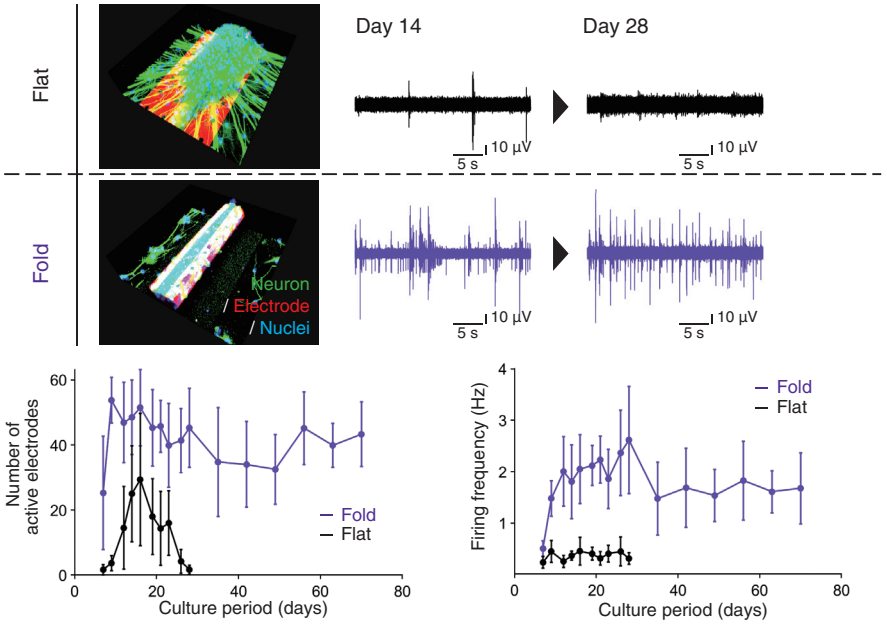


Fig. 5. Long-term recording from 3D neuronal aggregate by folded electrode. The figure is a modification of the original paper [6] (Creative Commons License (Attribution 4.0 International)).

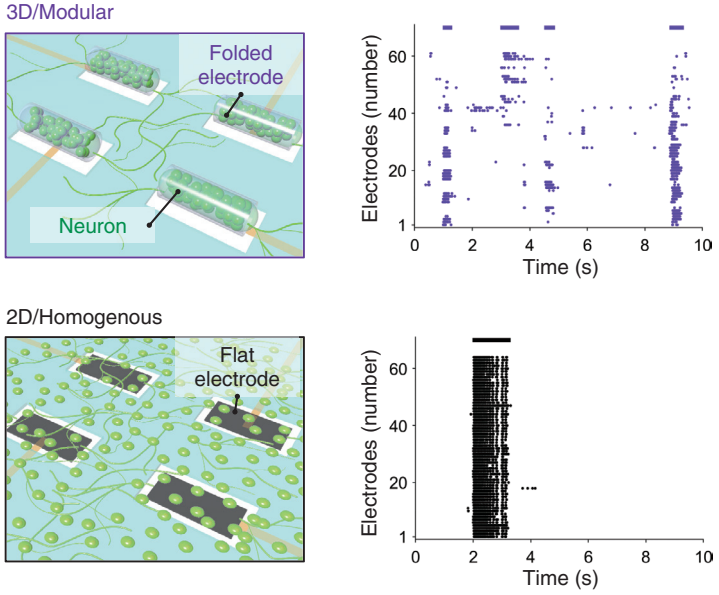


Fig. 6. Effect of network structure on synchronous firing patterns. The figure is a modification of the original paper [6] (Creative Commons License (Attribution 4.0 International)).

patterns in which several 3D neuronal aggregates synchronize with each other was increased. The firing patterns included neural firings that do not participate in the synchronization. This result indicates various firing patterns in which several modules fire synchronously while some fire asynchronously. In the brain, it has been theoretically shown that such a modular structure that generates diverse firing

patterns with a mixture of synchronous and asynchronous firing is advantageous for information processing. In this study, we experimentally demonstrated that the balance between synchronous and asynchronous firing fluctuates by reproducing modular and 3D structures in the cell culture model. The changes in the synchronous firing can be evaluated for an extended period. For example, after 7–9 days of culture, when axons bridge the 3D neuronal aggregates, asynchronous firing gradually becomes synchronous. After 2–3 weeks of culture, the overall synchronous firing becomes dominant, and the interval between repeated synchronous firings becomes shorter. The electrophysiological recordings help us better understand the formation of neuronal networks related to the brain's developmental process. Thus, the brain-on-a-chip model constructed in this study enables us to investigate brain systems related to the structure and maturation of neuronal networks from their firing patterns by changing the culture time and shape parameters such as electrode size and spacing.

4. Summary and future perspective

In this article, we demonstrated modeling neuronal networks in the brain by culturing cells on the electrode array while reproducing 3D and modular structures characteristic of the brain. Even single neurons have more exciting properties than other cells, but their collective behavior changes depending on the culture conditions, including the structure and materials of the substrate. Although the environment is fundamentally different from that of the actual brain, there is a possibility that new findings related to brain function and development will be revealed by exploring the preserved characteristics of cultured neurons. The closer the structure and properties of the neuronal network are to those of the brain, the more we can uncover relationships between the behavior of the neurons and brain function. In addition to physiological phenomena, this technology also has the potential to be a pathological model. For example, it would be used to construct a neuronal network that simulates autism, which is known to be associated with abnormalities in the structure of neuronal networks. Another disease model could be constructed by mixing neurons from induced pluripotent stem (iPS) cells derived from patients with neurological diseases. This model can be used to investigate drug responses to neuronal networks with brain-like functions without using the actual brain, contributing to medical research and drug discovery by reducing the number of experimental animals and accelerating drug testing. The performance of this model will be improved by increasing the number of electrodes, adding multimode stimulations (light, heat, vibration, etc.), and increasing the variety of 3D shapes of the electrodes. Targeting events related to brain function and pathology will enable us to develop advanced brain-on-a-chip models that better reflect physiological phenomena and diseases.

References

[1] K. Sakai, T. F. Teshima, H. Nakashima, and Y. Ueno, "Graphene-based Neuron Encapsulation with Controlled Axonal Outgrowth," Nanoscale, Vol. 11, No. 28, pp. 13249–13259, 2019. https://doi.org/10.1039/C9NR04165F

[2] T. F. Teshima, K. Sakai, H. Nakashima, and Y. Ueno, "Self-folded Three-dimensional Graphene for Biointerfaces," NTT Technical Review, Vol. 18, No. 2, pp. 32–39, 2020. https://doi.org/10.53829/ntr202002ra1

[3] T. F. Teshima, C. S. Henderson, M. Takamura, Y. Ogawa, S. Wang, Y. Kashimura, S. Sasaki, T. Goto, H. Nakashima, and Y. Ueno, "Self-folded Three-dimensional Graphene with a Tunable Shape and Conductivity," Nano Lett., Vol. 19, No. 1, pp. 461–470, 2019. https://doi.org/10.1021/acs.nanolett.8b04279

[4] T. F. Teshima, H. Nakashima, Y. Ueno, S. Sasaki, C. S. Henderson, and S. Tsukada, "Cell Encapsulation and 3D Self-assembly Using Multi-layered Polymeric Thin Films," NTT Technical Review, Vol. 16, No. 8, pp. 53–61, 2018. https://doi.org/10.53829/ntr201808ra1

[5] T. Goto, T. F. Teshima, K. Sakai, and M. Yamaguchi, "Three-dimensional Self-folding Assembly of Multi-layer Graphene at the Interface with a Polymeric Film," AIP Adv., Vol. 12, No. 7, 075002, 2022. https://doi.org/10.1063/5.0096473

[6] K. Sakai, T. F. Teshima, T. Goto, H. Nakashima, and M. Yamaguchi, "Self-folding Graphene-based Interface for Brain-like Modular 3D Tissue," Adv. Funct. Mater., Vol. 33, 2301836, 2023. https://doi.org/10.1002/adfm.202301836

Koji Sakai

Research Scientist, Bio-medical Informatics Research Center (BMC), Molecular and Bio Science Research Group, NTT Basic Research Laboratories.

He received a B.E. in engineering, M.E. in environment, and Ph.D. in engineering from the University of Tokyo in 2013, 2015, and 2018. From 2016 to 2018, he was a research fellow of the Japan Society for the Promotion of Science (JSPS). He joined NTT Basic Research Laboratories in 2018 and has been researching neuroengineering, electrophysiology, tissue engineering, and bioelectronics. He is a member of the Japan Society of Applied Physics (JSAP), the Institute of Electrical Engineers of Japan (IEEJ), and the Society for Chemistry and Micro-Nano Systems (CHEMINAS).

Tetsuhiko F. Teshima

Senior Research Scientist, Medical & Health Informatics Laboratories, NTT Research, Inc.

He received a B.S. in veterinary medicine, M.S. in cell biology, and Ph.D. in information science and technology from the University of Tokyo in 2009, 2011, and 2014. In 2014, he joined NTT Basic Research Laboratories. In 2020, he started to work as a senior research scientist at NTT Research, Inc., USA, a visiting scholar at Technische Universität München, Germany, and the guest international associate professor at Keio University, Japan. His current interests are in self-assembly of nanomaterials and soft materials with a particular focus on the bioelectronics and biointerfaces. He is a member of the Materials Research Society of the USA, the Institute of Electrical and Electronics Engineers (IEEE), and European Organ-on-Chip Society (EUROoCS).

Toichiro Goto

Research Scientist, Bio-medical Informatics Research Center, Molecular and Bio Science Research Group, NTT Basic Research Laboratories.

He received a B.S and M.S. from Waseda University, Tokyo, in 1996 and 1998, and Ph.D. in engineering from Shizuoka University in 2012. Originally, he was engaged in NTT Microsystem Integration Laboratories. He joined NTT Basic Research Laboratories (NTT BRL) in 2004. He then joined NTT Intellectual Property Center and returned to NTT BRL in 2018. He is currently working on graphene-based implantable electrodes for electrical stimulations and patent analysis for flexible devices. He is a member of JSAP.

Functional Evaluation of Bilayer Lipid for the Development of Artificial Cell-membrane Structures

Azusa Oshima and Yoshiaki Kashimura

Abstract

Lipids and membrane proteins are related in many reactions such as intracellular vesicular transport and intercellular communication. We are studying artificial cell membranes as a bottom-up approach to understand cellular reactions at the molecular level. Various elemental technologies are required to introduce membrane proteins into bilayer lipid membranes, which are the basic structure of cell membranes, and evaluate them. In this article, we introduce our efforts in developing an artificial cell membrane.

Keywords: bilayer lipid membrane, membrane protein, cell

1. Reconstruction of bilayer lipid membrane

Biological cells are covered with a lipid membrane that separates the inside and outside of the cell. The cell membrane contains a variety of membrane proteins, which adhere to or penetrate lipid membranes and account for almost 50% of the cell-membrane weight. Membrane proteins include receptor proteins that receive signal molecules such as neurotransmitters and transporters that transport ions using energy. Ion-channel-type receptor proteins are used for signal transduction by inducing a conformational change in the receptor due to ligand binding, which opens an ion channel, allowing ions to enter the cell and induce a change in electrical potential. Information transmission in the body is essential for vital activities, and abnormalities in information transmission can cause a variety of diseases. Thus, membrane proteins and lipid membranes, which are the basis for maintaining their functions, are closely related to biological functions. Understanding their dynamics is important not only in the life sciences but also in medicine and drug discovery. One method of evaluating lipids and proteins involves using bilayer-lipid-membrane reconstruction systems. Such systems assemble target lipids and membrane proteins under controlled solution and temperature conditions to reproduce phenomena on specific cell membranes and evaluate them at the molecular level.

The lipid molecules that compose the membrane have a hydrophilic head and hydrophobic hydrocarbon chains and are insoluble in water. In solution, lipid molecules form a lipid bilayer structure, in which a monolayer of lipid molecules overlaps two monolayers to confine hydrophobic groups inside by self-accumulation. We previously proposed a simple cell model for functional evaluation of lipids and proteins by combining bilayer lipid membranes and well-shaped substrates fabricated using nano- and micro-scale processing techniques (**Fig. 1**) [1]. This artificial cell membrane consists of a freestanding bilayer lipid membrane (freestanding membrane) area at the well and a support bilayer lipid membrane (supported membrane) area on the substrate, and membrane proteins are introduced into the freestanding membrane area. The inside and outside of the well are separated by the freestanding membrane, which mimics the inside and outside of a cell. The well structure is approximately the same size as a cell. The bilayer lipid membrane covering the well is supported by a solid substrate, but the lipid bilayer and substrate are not physically adsorbed and exist through a water layer a few nanometers thick. It is also known that each lipid molecule that constitutes

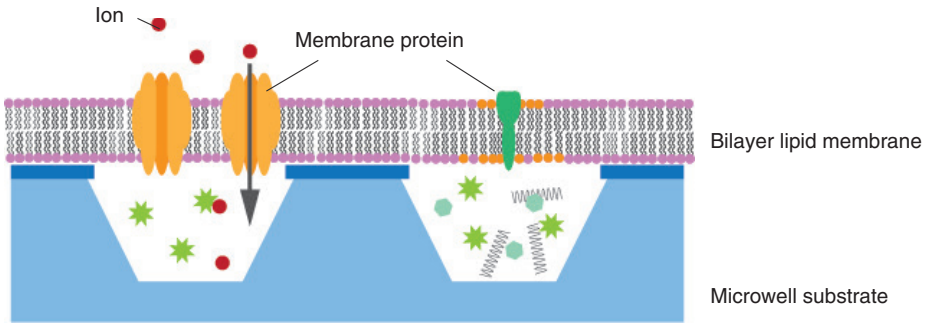


Fig. 1. Conceptual illustration of artificial cell membrane.

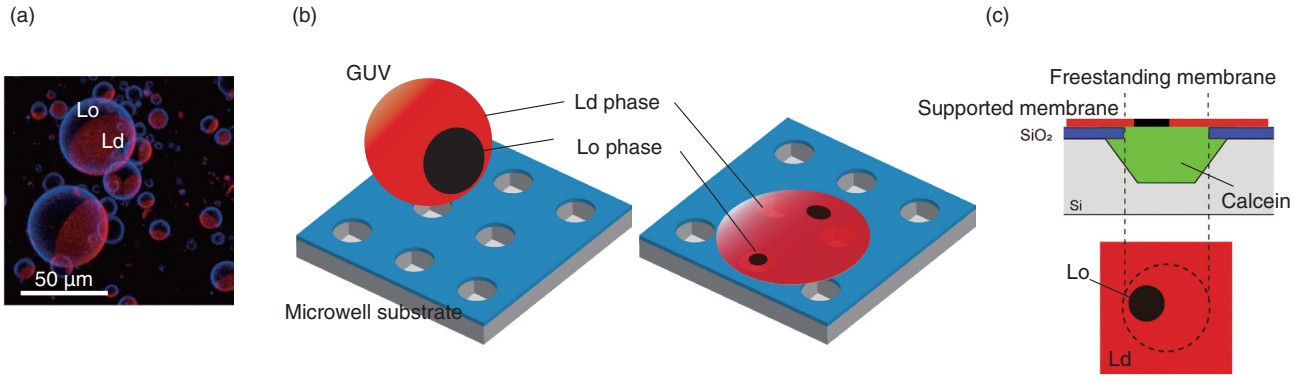
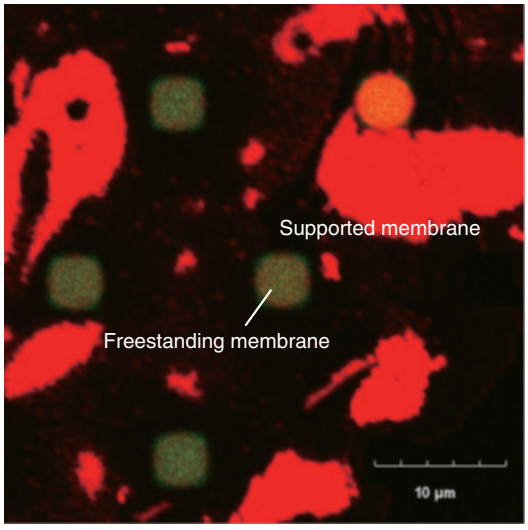


Fig. 2. Fluorescence image of GUVs and formation of bilayer lipid membrane on well-shaped substrate.

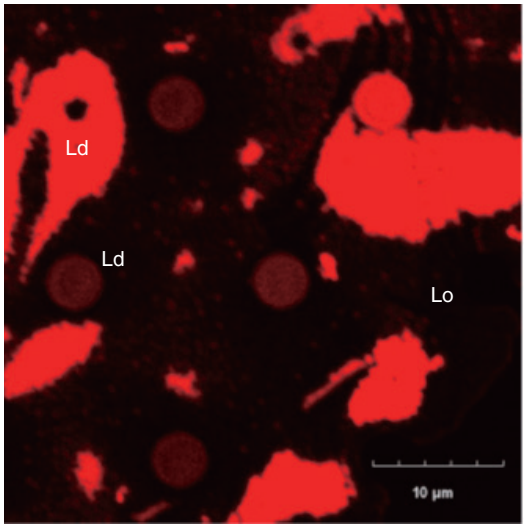
the bilayer lipid membrane is constantly moving, and the state of the bilayer lipid membrane can be evaluated using this lipid movement as an indicator. We discuss our evaluation of lipid-phase separation and changes in the bilayer-lipid-membrane shape using artificial cell-membrane substrates, the introduction of membrane proteins into bilayer lipid membranes, and the steps required to evaluate the ions that pass through the introduced membrane proteins.

2. Lipid-phase separation membrane

Biological cell membranes are composed of a variety of lipid molecules, which are roughly categorized into saturated and unsaturated lipids. These lipid molecules do not mix uniformly in the membrane, but saturated lipids are thought to form local domains with cholesterol present in the cell membrane. These lipid domains, called rafts, are in a different phase state than the surrounding membrane, and proteins involved in signal transduction are localized there. The roles and functions of lipid rafts need to be elucidated. However, since it is difficult to observe them directly in cell membranes, bilayer-lipid-membrane reconstruction systems have been used to understand the movement of lipid molecules as a physical phenomenon. A ternary mixture of saturated lipids, unsaturated lipids, and cholesterol produces a lipid-phase separation of a liquid ordered (Lo) phase/liquid disordered (Ld) phase. The Lo phase is less fluid and has a higher molecular density than the Ld liquid phase. We observed how such a phase-separation membrane behaves when formed on the well. To form phase-separated membranes on well structures, giant unilamellar vesicles (GUVs) were prepared from the ternary mixture of lipid solution. To distinguish each phase, the Lo and Ld phases were fluorescently labeled with laudan and with lipid conjugate with rhodamine, respectively. **Figure 2(a)** shows a fluorescent image of the prepared GUV. The blue areas indicate the Lo phase and red areas the Ld phase. When this GUV was ruptured on the well



(a) Merged fluorescence image of calcein (green) confined in the wells and bilayer lipid membrane labeled with rhodamine lipid (red).



(b) Fluorescence image of only bilayer lipid membrane labeled with rhodamine lipid. Bright red region shows Ld phase and dark region shows Lo phase.

Fig. 3. Fluorescence image of phase-separation membrane on the well substrate.

substrate, as shown in **Fig. 2(b)**, a bilayer lipid membrane sealing the well was formed. To determine whether a lipid bilayer had formed, the fluorescent dye calcein was sealed inside the well (**Fig. 2(c)**). **Figure 3** shows a fluorescent image of the phase-separation membrane on the well substrate. Red indicates an Ld phase, while the dark region indicates an Lo phase. The green fluorescence of calcein confirms whether the bilayer lipid membrane had successfully sealed the wells (Fig. 3(a)). The fluorescence image filtered for calcein-derived green fluorescence is shown in Fig. 3(b), and red fluorescence was observed in all the wells that were successfully sealed. This indicates that the Ld phase is preferentially present in a freestanding membrane, and the Lo phase is present only in the supported membrane [2]. Immediately after GUV rupturing, the Lo-phase areas were also observed in the freestanding membrane but decreased with time as they moved in the freestanding membrane. The Ld-phase area increased and the Lo-phase area disappeared from the freestanding membrane. This indicates that the lipids that form the low-fluidity Lo phase (saturated lipids and cholesterol) moved from the freestanding to the supported membrane, while the lipids that form the high-fluidity Ld phase (unsaturated lipids) moved into the freestanding membrane. Since the freestanding membrane is not in contact with the substrate, there is no substrate-lipid interaction as in the supported membranes, and the lipid molecules can move easily. The freestanding membrane is not supported, so it should be flexible in shape. These results indicate that even in a single continuous bilayer lipid membrane, each lipid has its own environment in which it is likely to exist. If we can control the movement of lipids by clarifying the conditions under which the Lo and Ld phases are likely to exist, we will be able to reproduce a wider variety of cell-membrane reactions and construct a cell-membrane model that is closer to living organisms.

3. Shape changes in bilayer lipid membrane

In our artificial cell membrane, the Ld phase is preferentially present in the freestanding membrane. We observed how the phase separation changes when the shape of the freestanding membrane is changed. The inside and outside of the well are separated by a bilayer lipid membrane, which allows the solution to be exchanged during measurement. It is difficult for ions and large, uncharged polar molecules (amino acids and glucose) to permeate through the lipid bilayer, but small molecules, such as H_2O and O_2, permeate. As shown in the schematic in **Fig. 4(a)**, when the solution outside the well is exchanged to cause an osmotic pressure difference between the

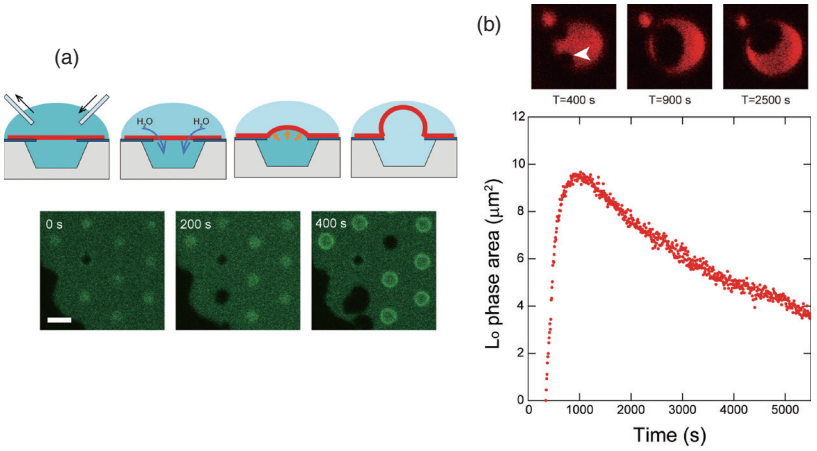


Fig. 4. The change in freestanding membrane shape induced by osmotic pressure difference. (a) Schematic of freestanding membrane-expansion mechanism and fluorescence images after solution exchange outside the well. (b) Time-lapse image of Lo phase and variation in Lo-phase area with time.

inside and outside of the well, H_2O flows into the inside of the well through the bilayer lipid membrane to cancel the osmotic pressure difference, and the solution volume inside the well increases [3]. Because of the fluidity of the bilayer lipid membrane, the lipid moved from the support membrane area and the freestanding membrane expanded like a balloon (Fig. 4(a), fluorescence image). When there was an osmotic pressure difference caused by the ternary mixture lipid bilayer of saturated lipid, unsaturated lipid, and cholesterol, the Lo-phase area, indicated with the arrow in **Fig. 4(b)**, increased as the freestanding membrane expanded, although the Lo-phase area decreased in the isobaric state without osmotic pressure difference. This is because the lipids (saturated lipids and cholesterol) that form the Lo phase in the supported membrane surrounding the freestanding membrane moved into the freestanding membrane as the solution volume inside the well increased. After the freestanding membrane had finished expanding and most of the freestanding membrane was converted to the Lo phase, the Lo phase was successfully maintained in the freestanding membrane for more than one hour [4] (Fig. 4(b)). When the shape of the freestanding membrane was changed, a change in the movement of the lipids was observed. This method of physically controlling the movement of bilayer lipid membranes is important for studying the factors that affect their state.

4. Membrane-protein introduction using membrane-fusion function of budding viruses from insect cells

It is an important in reconstruction systems to efficiently introduce membrane proteins into bilayer lipid membranes. If membrane proteins can be placed in bilayer lipid membranes in a controlled number and orientation while maintaining their function, many cell-membrane reactions can be reproduced. Baculoviruses use insect cells as hosts. Since baculoviruses are covered by the insect cell membrane when they bud out of insect cells, the expressed membrane protein is added to the insect-cell-derived budding virus (BV). Insect-cell-derived BVs have the ability to fuse with cells to invade the next host.

We focused on this function and applied it to the introduction of membrane proteins into bilayers lipid membranes [5]. We labeled the BV from insect cells with fluorescence, added it to the lipid bilayer formed on a silicon substrate, and examined changes in the fluorescence intensity of the labeled BV on the bilayer lipid membranes in solutions of different pH. **Figure 5(a)** shows the change in fluorescence intensity after one hour. The horizontal axis indicates the pH of the solution, and the fluorescence intensity of the dye labeled with the BV increases when the pH falls below 5.5, indicating an increase in the amount of BV reaching the bilayer lipid membranes. This indicates an increase in the amount of BV reaching the bilayer lipid membranes. A pH between 6.0 and 5.5 seems to be the threshold for activating the fusion

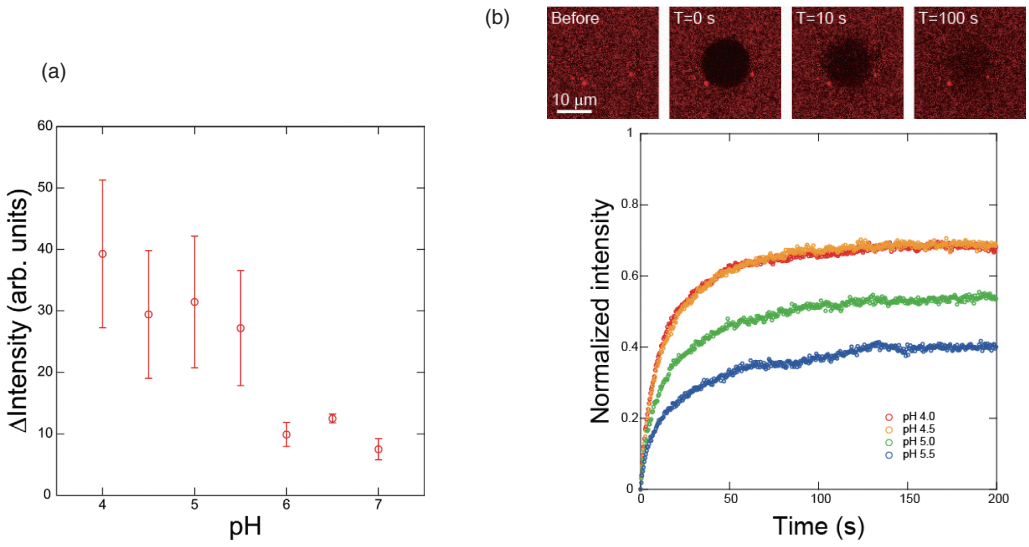


Fig. 5. BV fused to bilayer lipid membrane depending on solution pH. (a) Average of intensity 1 hour after adding BV. (b) Fluorescence recovery curve for each pH.

ability of the BV.

However, there is no significant difference in fluorescence intensity under pH 5.5. When bilayer lipid membranes fuse together, there are three stages of fusion: adherent, semi-fused, and fully fused. In some cases, the fusion proceeds no further than the adhered state. Since our goal is to introduce membrane proteins into bilayer lipid membranes, it is preferable that the membrane components derived from the BV are fully fused to them, but we cannot determine at what pH this is achieved from the results of fluorescence-intensity changes alone. Therefore, we evaluated the transfer of the components of the BV to the lipid bilayer by analyzing the lateral diffusion of lipids. Fluorescence recovery after photobleaching is a general method to evaluate the movement of lipids and components in a bilayer lipid membrane by analyzing the fluorescence-recovery curve obtained from the process of fading a fluorescent dye at a specific location by intense light and replacing the surrounding fluorescent dye with the faded dye by lateral diffusion. **Figure 5(b)** shows the fluorescence image of the movement of the quenched fluorescent dye by lateral diffusion. The fluorescence-recovery curve for each pH was obtained from the change in brightness of the fluorescence image, and the fluorescence-recovery rate increased at lower pH. This indicates that the amount of total fusion increases at lower pH due to the increase in the ratio of the membrane components of the BV that are transferred and diffused into the bilayer lipid membrane on the silicon substrate. The relationship between the pH of the solution and state of membrane fusion is shown in **Fig. 6**, which indicates that the ratio of membrane attachment is high around pH 5.5, and the ratio of total fusion increases as the pH decreases. In our artificial cell membrane, a solution pH under 4.5 is the optimum condition for membrane-protein incorporation. Under these solution conditions, we succeeded in introducing membrane proteins into the freestanding membrane area at the well.

5. Toward electrophysiological detection of ion-channel activity on device

The next step is to evaluate the function of the introduced membrane proteins. We measured ion flow through α-hemolysin, a membrane protein that forms pores in lipid bilayers and enables ion permeation, by using an artificial cell-membrane substrate. However, there is ion leakage from the external solution into the inside well, and this leakage must be reduced to detect signals of picoampere-level ion influx, as is the case with receptor membrane proteins. From theoretical and experimental investigations, it was found that the source of ion leakage is the inflow of ions from the interfacial water layer as small as 2 nm between the lipid bilayer and well substrate. We previously proposed a method of reducing ion leakage by surface modification of the substrate

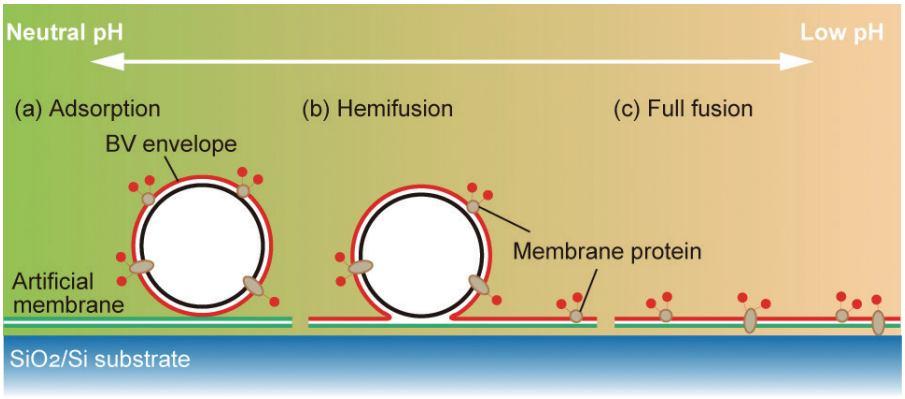


Fig. 6. pH dependence of BV fused to bilayer lipid membranes.

surface with bovine serum albumin to block the interface between the interfacial water layer and external solution, which is the path of ion leakage [6]. We have also found a simple and easy method of using dilute ionic liquid solutions as the aqueous solution. Ionic liquids are liquid salts composed only of ions, which are generally bulky organic ions. We succeeded in significantly reducing ion leakage through the interfacial water layer by using ionic liquids as electrolytes instead of ordinary inorganic salts and confining the bulky ions within the interfacial water layer. Using the lipid bilayer on the well substrate formed with these techniques, we successfully measured the ion-permeation function of α-hemolysin for more than a few hours and with high sensitivity.

6. Future development

Lipids and membrane proteins are not only related in intercellular communication but also in many reactions such as intracellular vesicular trafficking. We aim to develop a method of evaluating and controlling the movement of lipids and membrane proteins and reproducing the reactions that occur in cells to quantitatively evaluate the role of each biomolecule, which is difficult to visualize in cells. By controlling the movement of biomolecules, such as lipids and membrane proteins, with NTT's substrate fabrication technology that can be adapted to biomaterials, we expect to develop a bilayer-lipid-membrane reconstruction system that enables the understanding of complex transmission pathways in the body through a bottom-up process.

References

[1] K. Sumitomo, A. McAllister, Y. Tamba, Y. Kashimura, A. Tanaka, Y. Shinozaki, and K. Torimitsu, "Ca^{2+} Ion Transport through Channels Formed by α-Hemolysin Analyzed Using a Microwell Array on a Si Substrate," Biosens. Bioelectron., Vol. 31, No. 1, pp. 445–450, 2012. https://doi.org/10.1016/j.bios.2011.11.010

[2] K. Sumitomo and A. Oshima, "Liquid-ordered/Liquid-crystalline Phase Separation at a Lipid Bilayer Suspended over Microwells," Langmuir, Vol. 33, No. 46, pp. 13277–13283, 2017. https://doi.org/10.1021/acs.langmuir.7b02156

[3] A. Oshima, N. Inui, Y. Ueno, and K. Sumitomo, "Water Permeability through the Lipid Bilayers Suspended over Microwells on Si Substrates," IEEJ Trans. Electron. Inf. Syst., Vol. 140, No. 4, pp. 421–425, 2020 (in Japanese). https://doi.org/10.1541/ieejeiss.140.421

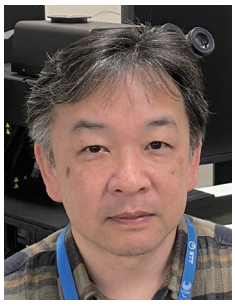
[4] A. Oshima, H. Nakashima, and K. Sumitomo, "Phase Separation in Freestanding Bilayer Lipid Membrane Induced by Osmotic Pressure Difference," Jpn. J. Appl. Phys., Vol. 59, No. 2, 027001, 2020. https://doi.org/10.7567/1347-4065/ab65a7

[5] A. Oshima, K. Nakanishi, N. Kasai, H. Nakashima, K. Tsumoto, and K. Sumitomo, "Mechanism of Budded Virus Envelope Fusion into a Planar Bilayer Lipid Membrane on a SiO_2 Substrate," Langmuir, Vol. 38, No. 18, pp. 5464–5471, 2022. https://doi.org/10.1021/acs.langmuir.1c03456

[6] Y. Kashimura, K. Sumitomo, and H. Nakashima, "Biodevices Using Microwells Sealed with Artificial Lipid Bilayers: Improvement of Sealing Performance by Protein Coating," Electr. Commun. Jpn., Vol. 103, No .9, pp. 15–22, 2020.

Azusa Oshima
Research Scientist, NTT Basic Research Laboratories.

She received a B.S. in physics from Nihon University, Tokyo, in 2008, and an M.E. and Ph.D. in biomedical engineering from Tohoku University, Miyagi, in 2010 and 2013. She joined NTT Basic Research Laboratories in 2013. She has been engaged in research on artificial bilayer lipid membranes for nanobioscience. She is a member of the Japan Society of Applied Physics (JSAP) and the Japan Society for Analytical Chemistry.



Yoshiaki Kashimura
Research Scientist, NTT Basic Research Laboratories.

He received a B.S., M.S., and Ph.D. in chemistry from the University of Tokyo in 1993, 1995, and 1999. He joined NTT Basic Research Laboratories in 1999. From 2005 to 2006, he was a visiting professor at the Chinese Academy of Sciences, Beijing, China. He has been researching transport properties at the single-molecule level ranging from small molecules to macrobiomolecules. His is currently conducting research on nanobiodevices. He is a member of JSAP, the Surface Science Society of Japan, the Institute of Electronics, Information and Communication Engineers, and the Material Research Society.

Characterization of Metal Ions in Neurons Using a Superconducting Flux Qubit

Hiraku Toida, Koji Sakai, Tetsuhiko F. Teshima, Kosuke Kakuyanagi, Imran Mahboob, and Shiro Saito

Abstract

In the race to develop superconducting quantum computers, superconducting quantum circuit technology has made remarkable progress. The technology has also been applied to quantum sensing, especially highly sensitive magnetometers. The target of on-chip measurement by superconducting quantum sensors has thus far been limited to impurities in semiconductors. However, the range of applications is not limited to such hard materials. This article describes the detection of ferric ions in neurons with a high-sensitivity, high-spatial-resolution magnetometer using a superconducting flux qubit.

Keywords: superconducting quantum circuits, quantum sensing, biosensing

1. Trace elements in living organisms

Iron is a trace element that plays a significant role in the body. Physiologically, it is involved in oxygen transport via hemoglobin and energy production via adenosine triphosphate*1. Pathologically, it is involved in many neurodegenerative diseases such as Alzheimer's disease. To understand these phenomena, it is necessary to study the transport and oxidation-state changes of ferric ions.

2. Methods for studying metal ions in biological samples

Mass spectrometry is a typical method for studying metal ions in biological samples. This method is quantitative and enables cell-by-cell analysis. However, in-situ observation is difficult because the analytical process involves ionization, i.e., cell disruption by inductively coupled plasma or laser irradiation. Optical methods, such as Raman spectroscopy, are also used to analyze metallic elements. They are suitable for in-situ observation and imaging of intracellular structures. However, these methods cannot detect changes in the valence of ions due to redox reactions or protein binding, so other methods are required to obtain this information.

Electron spin resonance (ESR) is widely used in various fields such as physics, chemistry, biology, geosciences, pharmaceuticals, and medicine. It is also useful for analyzing metal ions in biological samples, including ion valences. The difference between metal ions can be distinguished by determining the g-factor parameter. Typically, the analysis is conducted by filling small glass tubes with several milliliters of samples. The information obtained is therefore an average of the sample. In addition, spatial resolution and sensitivity limit cell-by-cell imaging.

A magnetometer based on a superconducting quantum circuit*2 has micrometer spatial resolution and

*1 Adenosine triphosphate: A source of energy for cells.

*2 Superconducting quantum circuit: A superconducting circuit containing a circuit element that behaves in a quantum mechanical way. Superconducting quantum circuits consist of a Josephson junction with nonlinear current and voltage characteristics in addition to the inductors and capacitors of ordinary electronic circuits.

can analyze materials with unpaired electrons[*3] such as metal ions. The method can therefore be a useful tool for investigating metal ions in biological samples at the cellular level.

3. Quantum sensing using superconducting quantum circuits

Quantum sensing with superconducting quantum circuits is mainly aimed at ultra-sensitive measurements of magnetic fields or their source electron spins. An electron spin in materials acts as a probe for the properties of the surrounding elements. It can therefore be used as a tool for material analysis. Research on quantum sensing using superconducting quantum circuits is progressing with two main approaches. In the first approach, ultra-high sensitivity is achieved by improving the amplifier noise and other sensitivity limiting factors of ESR spectrometers using superconducting quantum circuit technology. In the second approach, small numbers of electron spins can be detected using magnetometers based on superconducting qubits, which are highly sensitive to magnetic fields. Both approaches can detect the samples in the vicinity of superconducting quantum circuit chips. NTT is researching the latter approach, using superconducting flux qubits (FQs). We developed an ultra-sensitive sensor capable of detecting 20 electron spins in 1 second of signal integration. With this sensor, we successfully analyzed solid-state samples with micrometer spatial resolution by ESR [1, 2].

The sensors based on superconducting quantum circuits can also achieve high spatial resolution because the size of the superconducting quantum sensor can be reduced. For example, an FQ consists of a rectangular loop. The size of the loop is a few micrometers on each side. It selectively detects only the magnetic field penetrating through the loop, enabling measurements with micrometer-scale spatial resolution.

The high sensitivity and high spatial resolution are particularly powerful when applied to biological samples. The size of an FQ of a few micrometers is comparable to the size of a typical cell, giving us the spatial resolution to sufficiently distinguish individual cells. As the integration of superconducting qubits is progressing for quantum computing applications, it is possible to conduct cell-by-cell imaging with single-cell spatial resolution by fabricating two-dimensional arrays of FQs.

As the first step toward cellular imaging, we conducted a proof-of-principle experiment to measure trace elements in cells using an FQ. We could measure ferric ions in neurons with single-cell-level spatial resolution [3].

4. Method for detecting electron spins using an FQ

Figure 1 shows a schematic of the electron spin measurement system using an FQ. The FQ and neurons were placed in a dilution refrigerator, and experiments were conducted at temperatures below 200 mK (below –272.95°C). Biological samples are rarely measured at such cryogenic temperatures. Such a combination of biological samples and low temperature raises the concern that the low temperature may alter the behavior of neurons and prevent biological information from being obtained. However, since the target of this study is the metal ions in the sample, measurements can be made even at low temperatures. In addition, the sensitivity of electron spin detection generally improves with decreasing temperature, so a low-temperature environment is preferable for measuring small amounts of metal ions.

The FQ was fabricated on a silicon substrate. As shown in Fig. 1, an FQ is a loop-shaped device containing three Josephson junctions[*4] (indicated by the X symbols in the figure). The clockwise and counterclockwise currents flowing through the loop correspond to two quantum states, forming a two-level quantum system. The FQ is surrounded by a superconducting quantum interference device (SQUID)[*5]. The SQUID is connected to an amplifier, analog-to-digital converter, personal computer, etc. at room temperature to read out the quantum state of the FQ. The transition frequency between two levels of the FQ is controlled by changing the magnetic flux through the loop. This magnetic flux is provided by a superconducting magnet placed near the FQ ($B_{\perp}$ shown in Fig. 1).

The principle of electron spin detection using an FQ is shown in **Fig. 2**. As mentioned above, FQs are usually controlled by an external superconducting magnet. However, the resonance frequency of the FQ

*3 Unpaired electron: An electron in an atom or molecule that has not formed a pair. These electrons are a source of magnetization. ESR can measure the properties of materials containing unpaired electrons.

*4 Josephson junction: A circuit element with a very thin insulating film sandwiched between two superconductors, used as a nonlinear element in superconducting quantum circuits.

*5 SQUID: A loop-shaped device with two Josephson junctions used as a highly sensitive magnetometer.

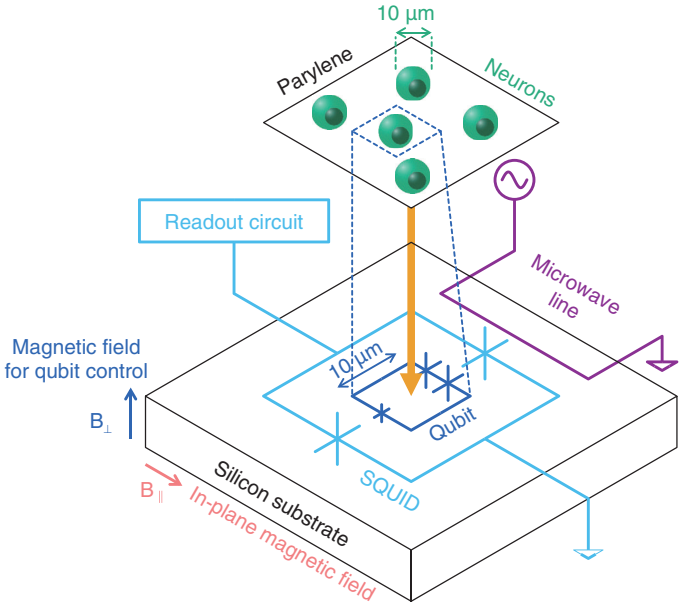


Fig. 1. Experimental setup for electron spin detection using an FQ. The figure is a modification of the original paper [3] (Creative Commons License (Attribution 4.0 International)).

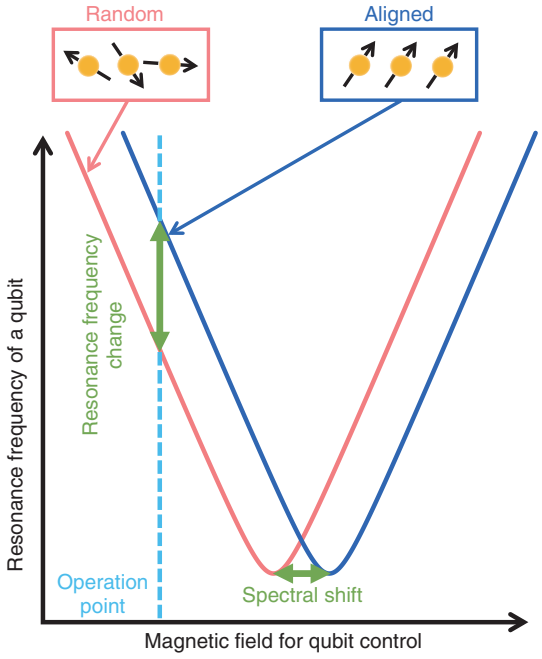


Fig. 2. Principle of electron spin detection using an FQ. The figure is a modification of the original paper [3] (Creative Commons License (Attribution 4.0 International)).

is also changed when a sample acts as a magnetic field source in the vicinity of the FQ. If the sample is paramagnetic*6, the magnetization of the sample, i.e., the alignment of the electron spins, can be controlled by the temperature and magnetic field applied to the sample. At low temperatures and high magnetic fields, the effect of aligning electron spins along the direction of the magnetic field is greater than thermal fluctuation. Therefore, the sample shows a large magnetization (blue box in Fig. 2). In contrast, at high temperatures and low magnetic fields, the electron spins are oriented in different directions due to thermal fluctuations. Therefore, the sample shows small magnetization (red box in Fig. 2). The resonance frequency of the FQ changes because the magnetic flux through the qubit loop changes as the magnetization of the sample changes. Therefore, the magnetic properties of the sample can be studied by measuring the change in the resonance frequency of the FQ as a function of the temperature and magnetic field ($B_{\parallel}$ in Fig. 1). It is important to note that the direction of the magnetic field for controlling the electron spins ($B_{\parallel}$ in Fig. 1) is different from that of the magnetic field for controlling the qubit ($B_{\perp}$ in Fig. 1).

*6 Paramagnetism: One of the magnetic properties of materials. When no external magnetic field is applied, the direction of electron spins is random. However, when a magnetic field is applied, the electron spins align, and magnetization appears.

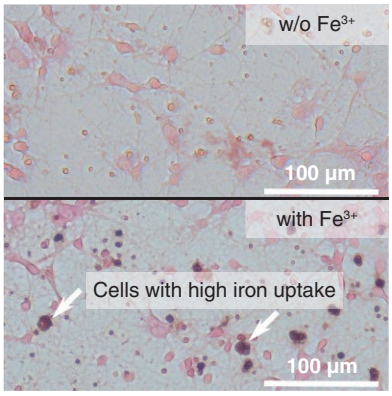


Fig. 3. Optical micrographs of neurons cultured in normal medium and in medium supplemented with iron. The figure is a modification of the original paper [3] (Creative Commons License (Attribution 4.0 International)).

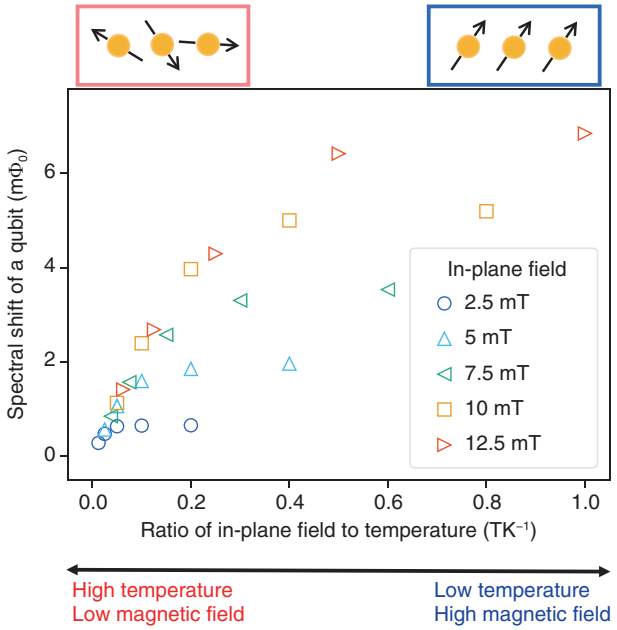


Fig. 4. Magnetization of neurons measured using an FQ. The vertical axis is expressed in units of magnetic flux quanta (Φ_0). The figure is a modification of the original paper [3] (Creative Commons License (Attribution 4.0 International)).

The sensitivity of electron spin detection improves when the magnetic interaction between the sample and the FQ is large. This is usually achieved by reducing the distance between them. However, since biological samples may contain liquid, insulating properties cannot be expected. Therefore, for the FQ to function properly, an insulating film must be placed between the biological sample and the FQ. We have solved this problem by using a biocompatible polymer film (parylene) with a thickness of a few micrometers for the insulation.

Rat hippocampal neurons were used as biological samples. These neurons were cultured on the parylene film. To simulate neurological diseases where cells contain elevated levels of iron and other metals, the neurons were cultured in a medium supplemented with ferric ions. As shown in **Fig. 3**, we prepared neurons that took up more ferric ions than those cultured in a standard culture medium. The structure of neurons can be maintained at low temperatures by cross-linking proteins with paraformaldehyde and freeze-drying them after the fixation.

The parylene film with cultured neurons was placed on a silicon substrate on which the FQ had been fabricated. The loop size of the FQ is 24 × 6 µm^2, about the size of a neuron. Therefore, the data obtained in the experiment can be considered to reflect the properties of a single cell.

5. Measuring neurons using an FQ

We first conducted an experiment to detect the magnetization of neurons. We measured the spectral shift corresponding to the change in the resonance frequency of the FQ. The in-plane magnetic field of several milli Tesla and the temperature were controlled. **Figure 4** shows that at the lower temperature or higher magnetic field, the spectrum of the FQ shifted significantly: the sample showed a large magnetization. This result is consistent with the expected behavior when the sample contains paramagnetic material. Thus, the neurons and/or the parylene may contain paramagnetic components.

To determine whether the neurons or the parylene is paramagnetic, we measured only the parylene under the same experimental conditions (**Fig. 5**). Parylene is a polymer film composed of carbon and hydrogen that contains no unpaired electrons. In principle, it would not be paramagnetic. However, a small paramagnetic response can be observed due to damage such as oxidization. Such a control experiment is therefore necessary. Figure 5 shows that the magnetization of parylene was significantly smaller than that of neurons. This result indicates that the magnetization detected in Fig. 4 is almost entirely from the neurons.

The results thus far suggest that the neurons contain

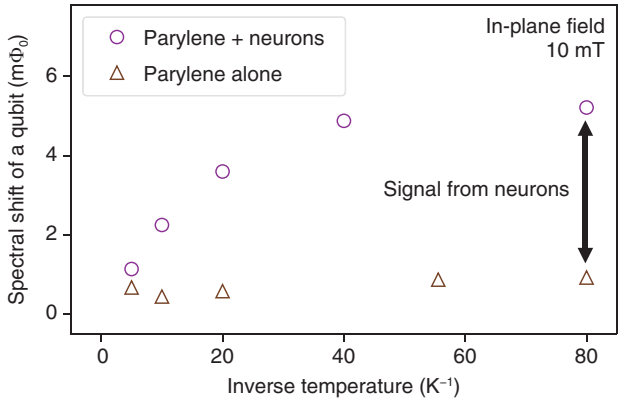


Fig. 5. Comparison of magnetization between neurons cultured on parylene and parylene alone. The vertical axis is expressed in units of magnetic flux quanta (Φ_0). The figure is a modification of the original paper [3] (Creative Commons License (Attribution 4.0 International)).

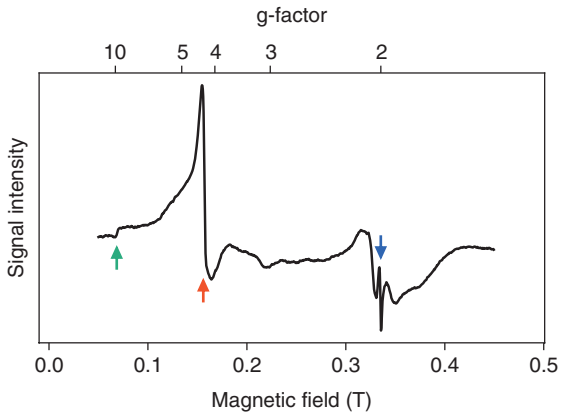


Fig. 6. ESR spectrum of neurons. The figure is a modification of the original paper [3] (Creative Commons License (Attribution 4.0 International)).

some paramagnetic material. However, the origin of this paramagnetism is unknown. We collaborated with Ono and Hori laboratory at Shizuoka University, which specializes in low-temperature ESR, to further analyze the samples. The g-factor can be obtained from the ESR spectrum. By comparing this value with known material-specific g-factors, the source of the magnetization can be identified.

The ESR spectrum of neurons is shown in **Fig. 6**. The peaks corresponding to g-factors of 9.8, 4.3, and 2.0 can be observed in this spectrum. The g-factors 9.8 and 4.3 are known to correspond to ferric ions in the cell [4]. The magnitude of the peaks indicates that this is the main source of the magnetization in the sample. Combining this result with the magnetometry result of the FQ, we can say that we have detected ferric ions in neurons with spatial resolution at the single-cell level. The peak with a g-factor of 2.0 is due to copper ions, which may also contribute to the magnetization in smaller amounts.

We conducted a qualitative analysis of metal ions in neurons using an FQ. The results of magnetometry using an FQ enable the quantification of metal ions contained in neurons. The magnitude of the change in the resonance frequency of the FQ corresponds to that of the magnetization. Therefore, this change in resonance frequency can be converted to the number of metal ions by comparing the results of a separate experiment on a reference sample with a known electron spin concentration. The results in Fig. 4 indicate that the measured rat hippocampal neurons contained 8 μg of iron per gram in the dry state. This result is consistent with 2 to 34 μg/g obtained in previous studies on human brain cells [4, 5].

6. Prospects

Using quantum sensing technology with FQs, we have successfully detected and analyzed metal ions in neurons quantitatively and qualitatively with single-cell-level spatial resolution. Future work includes single-cell resolution ESR spectroscopy of neurons using an FQ and imaging of metal ions in tissues by fabricating two-dimensional arrays of FQs.

References

[1] R. P. Budoyo, K. Kakuyanagi, H. Toida, Y. Matsuzaki, and S. Saito, "Electron Spin Resonance with Up to 20 Spin Sensitivity Measured Using a Superconducting Flux Qubit," Appl. Phys. Lett., Vol. 116, No. 19, 194001, May 2020. https://doi.org/10.1063/1.5144722

[2] S. Saito, K. Mizuno, T. Takenaka, H. Toida, and K. Kakuyanagi, "Quantum Information Technology Based on Superconducting Quantum Circuits," NTT Technical Review, Vol. 21, No. 6, pp. 29–35, June 2023. https://doi.org/10.53829/ntr202306fa3

[3] H. Toida, K. Sakai, T. F. Teshima, M. Hori, K. Kakuyanagi, I. Mahboob, Y. Ono, and S. Saito, "Magnetometry of Neurons Using a Superconducting Qubit," Commun. Phys., Vol. 6, 19, Feb. 2023. https://doi.org/10.1038/s42005-023-01133-z

[4] P. Kumar, M. Bulk, A. Webb, L. van der Weerd, T. H. Oosterkamp, M. Huber, and L. Bossoni, "A Novel Approach to Quantify Different Iron Forms in *ex-vivo* Human Brain Tissue," Sci. Rep., Vol. 6, 38916, Dec. 2016. https://doi.org/10.1038/srep38916

[5] L. H. P. Vroegindeweij, L. Bossoni, A. J. W. Boon, J. H. P. Wilson, M. Bulk, J. Labra-Muñoz, M. Huber, A. Webb, L. van der Weerd, and J. G. Langendonk, "Quantification of Different Iron Forms in the Aceruloplasminemia Brain to Explore Iron-related Neurodegeneration," Neuroimage Clin., Vol. 30, 102657, Apr. 2021. https://doi.org/10.1016/j.nicl.2021.102657

Hiraku Toida
Senior Research Scientist, Superconducting Quantum Circuits Research Group, NTT Basic Research Laboratories.

He received a B.E., M.E., and Ph.D. in arts and sciences from the University of Tokyo in 2008, 2010, and 2013. He joined NTT Basic Research Laboratories in 2013, where he studied coherent coupling between an erbium-doped crystal and superconducting flux qubit. From April to June of 2014, he was with NTT Microsystem Integration Laboratories, where he studied numerical simulation of plasmonic waveguides. He is currently studying highly sensitive spin sensing using superconducting quantum circuits. He is a member of the Physical Society of Japan (JPS) and the Japan Society of Applied Physics (JSAP).

Koji Sakai
Research Scientist, Bio-medical Informatics Research Center (BMC), Molecular and Bio Science Research Group, NTT Basic Research Laboratories.

He received a B.E. in engineering, M.E. in environment, and Ph.D. in engineering from the University of Tokyo in 2013, 2015, and 2018. From 2016 to 2018, he was a research fellow of the Japan Society for the Promotion of Science (JSPS). He joined NTT Basic Research Laboratories in 2018 and has been researching neuroengineering, electrophysiology, tissue engineering, and bioelectronics. He is a member of JSAP, the Institute of Electrical Engineers of Japan (IEEJ), and the Society for Chemistry and Micro-Nano Systems (CHEMINAS).

Tetsuhiko F. Teshima
Senior Research Scientist, Medical & Health Informatics Laboratories, NTT Research, Inc.

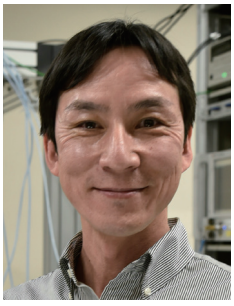
He received a B.S. in veterinary medicine, M.S. in cell biology, and Ph.D. in information science and technology from the University of Tokyo in 2009, 2011, and 2014. In 2014, he joined NTT Basic Research Laboratories. In 2020, he started to work as a senior research scientist at NTT Research, Inc., USA, a visiting scholar at Technische Universität München, Germany, and guest international associate professor at Keio University, Japan. His current interests are in self-assembly of nanomaterials and soft materials with a particular focus on the bioelectronics and biointerfaces. He is a member of the Materials Research Society of the USA, the Institute of Electrical and Electronics Engineers (IEEE), and European Organ-on-Chip Society (EUROoCS).

Kosuke Kakuyanagi
Senior Research Scientist, Supervisor and Group Leader of Superconducting Quantum Circuits Research Group, NTT Basic Research Laboratories.

He received a B.S., M.S., and Ph.D. in science from Hokkaido University in 2000, 2002, and 2005. He joined NTT Basic Research Laboratories in 2005 and has been studying superconducting qubits. He is currently engaged in an experimental study of physics on superconducting quantum circuits. He is a member of JPS and JSAP.

Imran Mahboob
Distinguished Researcher, Superconducting Quantum Circuits Research Group, NTT Basic Research Laboratories.

He received an MPhys. in theoretical physics from the University of Sheffield, UK, in 2001 and Ph.D. in physics from the University of Warwick, UK, in 2004. He joined NTT Basic Research Laboratories in 2005 and had been researching electromechanical systems to harness their nonlinear dynamics to develop phonon-based information technologies. Since 2017 he has been engaged in the development of microwave quantum optics for quantum technology applications. He was appointed as distinguished researcher of NTT in 2013.



Shiro Saito
Senior Distinguished Researcher, Superconducting Quantum Circuits Research Group, NTT Basic Research Laboratories.

He received a B.E., M.E., and Dr. Eng. in applied physics from the University of Tokyo in 1995, 1997, and 2000. He joined NTT Basic Research Laboratories in 2000. Since then he has been engaged in quantum information processing using superconducting circuits. He was appointed as distinguished researcher of NTT in 2012 and senior distinguished researcher in 2021. He was a guest researcher at Delft University of Technology from 2005 to 2006. He was a guest associate professor at Tokyo University of Science from 2012 to 2020 and promoted to a guest professor in 2020. He is a member of JPS and JSAP.

Creating Smart Cities through Digital Twins

Ippei Shake and Chihiro Yamamoto

Abstract

We describe our initiatives in creating smart cities using digital twin (DT) technology. In these Feature Articles on Urban DTC for Creating Optimized Smart Cities Attentive to the Individual, we first outline the concept of a totally optimized smart city. We then describe a variety of services that use digitalized data and artificial intelligence (AI), an integrated platform that achieves total optimization by interlinking those services, examples of services that link areas over a wide region, and the AI value platform that makes it easy to provide such cross-domain services.

Keywords: digital twin, AI, smart city

1. Introduction

There are many problems in modern society that need to be solved in a variety of areas including health, education, disaster management, poverty, population, and climate change. To this end, the Sustainable Development Goals were proposed by the United Nations as goals that mankind should strive to achieve. In Japan, the future form of society that should be targeted has been defined as "Society 5.0." There is also demand for the provision of new services as lifestyles change in the wake of the COVID-19 pandemic and user needs diversify. Under these conditions, the value provided by each industry is changing greatly from what has been considered common sense up to now and is taking on the form of cross-field services that go beyond individual industries. In short, services are becoming increasingly complex.

Against this background, NTT Smart Data Science Center (Smart Data Science Research Project, NTT Computer and Data Science Laboratories) is working to transform society through the use of digital twins (DTs). Our aim is to develop industry and solve social problems through the creation of new services that use data and artificial intelligence (AI). In these Feature Articles on Urban DTC for Creating Optimized Smart Cities Attentive to the Individual, we introduce our initiatives toward smart cities. The direction of our study flows from (a) optimization in units of residential and building districts to (b) urban optimization spanning multiple districts and (c) optimization that includes supply and demand on a national scale (**Fig. 1**).

2. Initiatives toward smart cities

We wish to provide the following two types of value as our ultimate goal.

(1) Totally optimized smart cities integrating society and people: Mutual optimization of value among individuals and communities and among a variety of industries extending as far as the global environment. In the future, smart cities will take on wide-area features in both spatial and virtual terms, and a sense of well-being and prosperity will be sustained by capitalizing on the features of each district (Social Well-Being and Flourishing Society).

(2) Smart cities that respond to diversifying needs by optimizing value for the individual: Making people's lives even better in a natural way that is attentive to latent and constantly changing individual needs (Personalized Well-Being City).

We are moving forward on initiatives for presenting

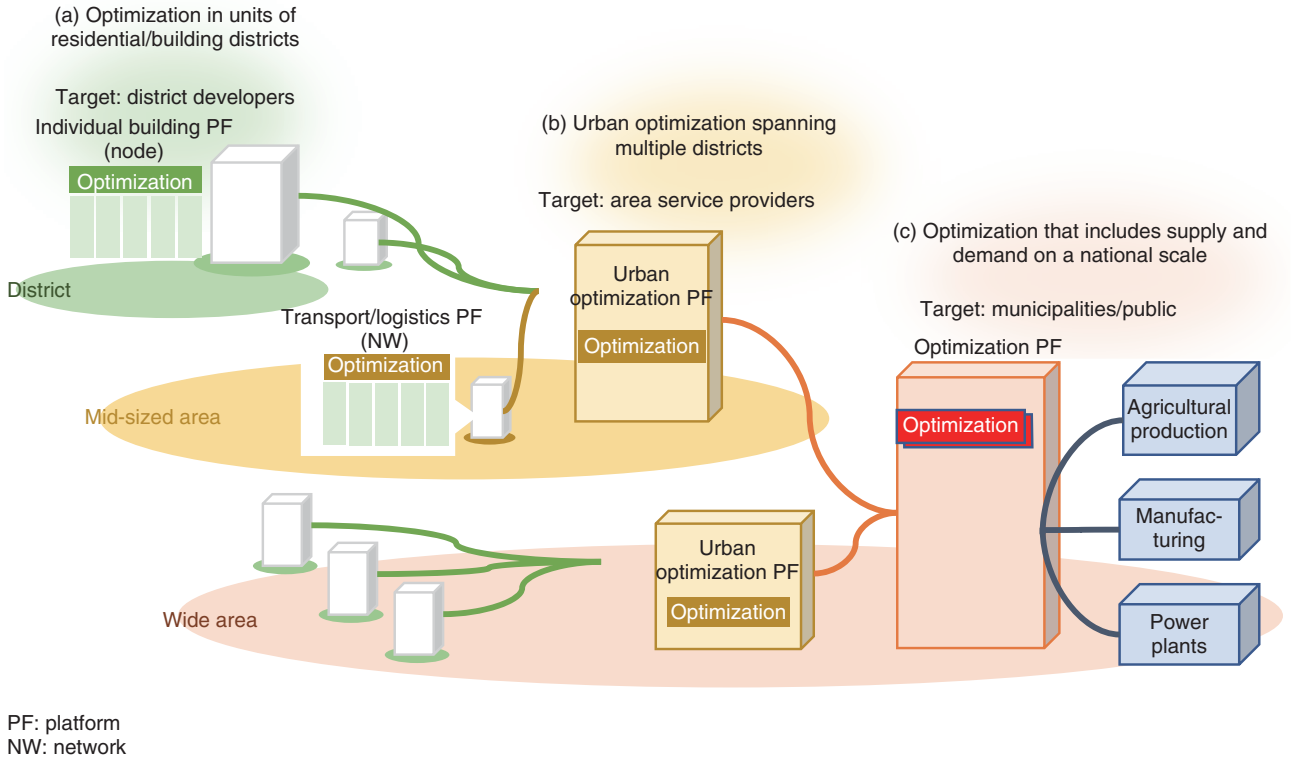


Fig. 1. Embodying optimization through inter-AI linking.

mechanisms and concrete services for the smart cities described above. All these mechanisms and services use DT technology that models and reproduces states in the real world in cyber space and that feeds back value to the real world by simulating a variety of state changes.

Building districts and areas covering multiple districts as in study items (a) and (b) above constitute a field that provides value across multiple and overlapping industry domains such as energy, retail (stores and eating/drinking establishments), real estate (building management, offices, and residences), and mobility. Therefore, configuring an entire city/area/district as a single DT is not very realistic, so the approach that we take is to configure a DT for each industry domain and link the simulations of each DT to configure DTs in units of districts. Thus, various types of value can be provided simultaneously to shop owners and managers within a district and social problems can be solved, such as by optimizing energy consumption and facility usage and reducing food loss while providing services that make the day-to-day life and activities of individuals more comfortable. When setting out to accomplish this, however, there are difficulties that must be faced since a variety of elements, such as the behavior and state of each individual within a district, state inside of each building (stores, offices, etc.), and social environment (energy, etc.), are all intertwined.

For example, focusing on energy savings can reduce personal comfort, and aiming to reduce food loss at a store, such as by limiting the amount of stocked goods to reduce the risk of unsold items, may prevent customers from purchasing desired items. In other words, identifying and optimizing a single need can lead to conditions in which other needs are not satisfied. A district is a space in which services for a variety of stakeholders exist and multiple needs become intertwined in a complex manner. There is a separate optimization index, or key performance indicator, for each service, but the individual behavior of people can influence each service and the external environment, which can influence other people and other services. If we could take a bird's-eye view of this entire situation and optimize it, we might be able to provide a pleasant environment and make life

easier, ideally for anyone.

Our goal is to achieve total optimization of multiple services in an urban district through Digital Twin Computing (DTC). We call this initiative "urban DTC." To take into account the mutual effects (linking) among multiple services, urban DTC uses predictions of behavior in units of individuals, predictions of changes in the environment and things brought about by that behavior, and predictions of the mutual influence among things that have also been influenced. We can make a broad concept of DT of a district more specific by first envisioning various types of future services and defining in detail the value that they can provide in a district. This value includes not only the user experience (UX) of visitors to the district, the people who work there, and the people who live there but also support services that enable building owners, district managers, and tenants of stores and offices to provide an even better experience for their customers (district visitors) and employees at all times. If we broadly define all such provided value as UX, we can think of "data-driven urban development" more precisely as "UX-driven urban development," which makes it important to begin with the design of future UX.

Turning now to optimization over a wide area that expands to the entire nation as described in study item (c) above, it is necessary to focus on industry domains that influence each other's supply and demand over a wide area. For example, we can consider that domains that are in close contact with people's lives and behavior (energy use, retail, real estate, mobility, etc.) would benefit from optimization executed from a demand standpoint. When such domains execute optimization with a focus on convenience and comfort, it is not so important to think from a perspective of a wide area on a national scale. However, there are also industries that would execute optimization by supply-and-demand matching by considering supply and logistics (quantity and type, timing, production location, etc.), that is, industries such as manufacturing, agriculture, and electric power and energy that must take into account the entire supply chain. In these industries, problems arise in the optimization of quantities, types, delivery routes, etc. between consumers and supply sources such as factories, production regions, and power plants. For this reason, we are developing a supply-chain optimization platform using DTs for each industry.

3. Conclusion

In our work on (a) optimization in units of residential and building districts, we have thus far achieved individual services such as air-conditioning optimization for common areas in buildings and zero-food-loss stores using DTs. In these Feature Articles, we outline a personalized air-conditioning service [1], store-management optimization and one-to-one marketing stores [2], a robot delivery service based on mobile orders [3], and hospitality services using a DT integrated platform and integrated app that achieve total optimization by linking such services [4]. Regarding our work on (c) optimization that includes supply and demand on a national scale, we introduce a virtual market for agricultural logistics [5] and optimization of power supply and demand [6]. We have successfully developed multiple AI models and an inter-AI linking mechanism for diverse domains and extracted and solved associated problems toward their provision as services. On the basis of this achievement, we outline our development of the AI value platform embodying the knowledge gained from a bird's eye view of cross-domain services [7].

References

[1] S. Mieda, S. Sakai, and T. Hata, "Personalized Air Conditioning Achieving Both Personal Comfort and Energy Savings," NTT Technical Review, Vol. 22, No. 5, pp. 64–70, May 2024. https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr202405fa8.html

[2] S. Kim, K. Nishimoto, M. Mizushima, S. Yamada, and J. Tomita, "Using Smart-store Behavior Data to Optimize Sales Promotion," NTT Technical Review, Vol. 22, No. 5, pp. 71–77, May 2024. https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr202405fa9.html

[3] S. Ouji, Y. Fukumoto, and N. Matsuura, "Robot Delivery Service with Mobile Ordering," NTT Technical Review, Vol. 22, No. 5, pp. 78–84, May 2024. https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr202405fa10.html

[4] J. Ito, H. Sato, and A. Hoshino, "Urban District Experiences for the Individual Generated by the Linking of Digital Twins—Digital-twin-integrated Platform and Integrated App," NTT Technical Review, Vol. 22, No. 5, pp. 56–63, May 2024. https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr202405fa7.html

[5] T. Baba, T. Hikichi, S. Toyoshima, Y. Hirano, Y. Hanakago, A. Isoda, and J. Tomita, "Improving Efficiency of Agricultural-product Distribution by Using a Virtual Market," NTT Technical Review, Vol. 22, No. 5, pp. 94–101, May 2024. https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr202405fa12.html

[6] T. Maki, K. Matsui, T. Fujinami, H. Kurasawa, and J. Tomita, "Refining Solar-power-generation Plans to Achieve Stable Power Supply by Predicting Total Solar Irradiance," NTT Technical Review, Vol. 22, No. 5, pp. 102–109, May 2024. https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr202405fa13.html

[7] T. Hata, M. Kodama, M. Fujishima, Y. Adachi, K. Fukuda, and Y. Yokohata, "AI Value Platform Accelerates Data Valorization by Consolidating SDSC's Elemental Technologies," NTT Technical Review, Vol. 22, No. 5, pp. 85–93, May 2024. https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr202405fa11.html

Ippei Shake
Vice President, NTT Smart Data Science Center, NTT Computer and Data Science Laboratories.

He received a B.S. and M.S. in physics and Ph.D. in informatics from Kyoto University in 1994, 1996, and 2006. In 1996, he joined NTT Optical Network Systems Laboratories and was engaged in the research of high-speed optical transmission systems. He was also engaged in the development of the NGN (Next Generation Network) and in human resources and has managed several collaborative projects with Dwango, Shochiku, Toyota and so on. Since 2016, he has been managing the data-driven digital-transformation projects for smart cities, mobility, healthcare, and agriculture. In this field, he has been also engaged in research and development of digital twins and Digital Twin Computing. He is a member of the Institute of Electrical and Electronics Engineers (IEEE) and the Institute of Electronics, Information and Communication Engineers (IEICE).

Chihiro Yamamoto
Senior Research Engineer, NTT Smart Data Science Center, NTT Computer and Data Science Laboratories.

She received an M.E. in global information and telecommunication studies from Waseda University, Tokyo, in 2008 and joined NTT the same year. She has been engaged in digital transformation projects for smart cities. She is a member of IEICE.

Urban District Experiences for the Individual Generated by the Linking of Digital Twins—Digital-twin-integrated Platform and Integrated App

Jun Ito, Hiroyuki Sato, and Ai Hoshino

Abstract

An urban district includes a variety of stakeholders each with their own interests with trade-offs existing among those interests. The digital-twin-integrated platform was developed to achieve social well-being by calculating a state that obtains a balance among those trade-offs on the basis of values predicted by multiple digital twins. An integrated app was also developed that provides services for the individual based on pre- and post-behavior context by determining behavior spanning multiple services using a single identifier.

Keywords: urban DTC, digital-twin-integrated platform, integrated app

1. Total district optimization by Digital Twin Computing

An urban district consists of people, things, and environmental factors that influence each other in complex ways, which is why representing a district by a single digital twin (DT) is difficult. We therefore take the following alternative approach. First, we model a DT for each industry domain making up the district, next, we model the DT of a person mediating among multiple DTs, and finally, we apply linking*1 to the DTs and model their mutual interaction by Digital Twin Computing (DTC) (**Fig. 1**).

Since each DT has a different objective, running a DT independently may give rise to competition among different DTs. For example, the objective of a DT for a restaurant (zero-food-loss DT) is to minimize food loss on the premise that food loss occurs in accordance with the difference between the predicted values and actual values of the number of customers, sales figures for each menu item, etc. However, forcing people to visit other restaurants for the sake of controlling food loss competes with the DT of a person whose objective is to maximize personal satisfaction (personal-service DT). There is therefore a trade-off between food-loss control and personal satisfaction—minimizing food loss and maximizing satisfaction cannot be achieved simultaneously. This makes it necessary to find an optimal solution that achieves a balance between food loss and satisfaction.

Given multiple objective functions as in the case of minimizing food loss and maximizing satisfaction, the task of finding an optimal solution that achieves a balance among those functions is called a multi-objective optimization problem*2. In this type of problem, an optimal solution is not uniquely

*1 Linking: The interconnecting of a group of functions each of which provides value on its own.

*2 Multi-objective optimization problem: Since optimization becomes increasingly difficult for four or more objective functions, such a case is distinguished as a many-objective optimization problem.

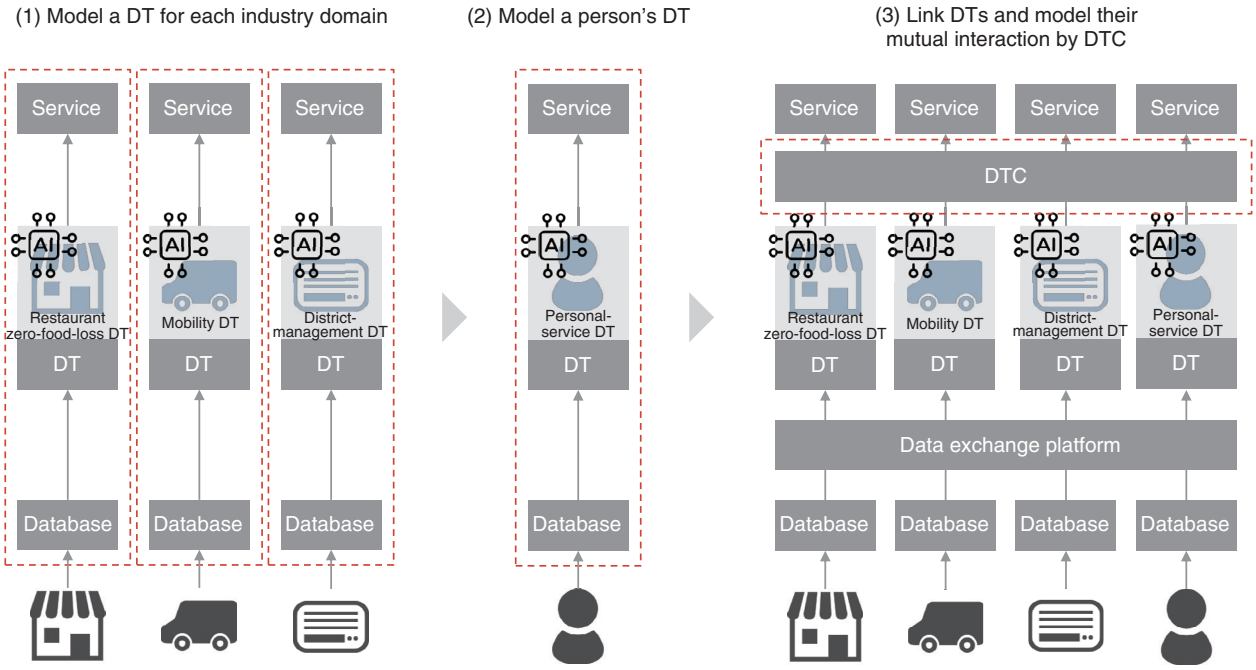


Fig. 1. Approach to district modeling in urban DTC.

determined. Rather, multiple optimal solutions form a curved line (in the case of two objective functions) or curved surface (in the case of three or more objective functions). This curved line or curved surface is called the Pareto front (**Fig. 2**). The process of linking DTs and modeling their mutual interaction by DTC means formulating a multi-objective optimization problem in relation to multiple linked DTs and determining the Pareto front.

The objective with DTC is to provide new value by achieving total optimization of a district beyond optimization achieved independently by a single DT. It is used to determine a state of social well-being that achieves balanced trade-offs among multiple DTs by first solving the multi-objective optimization problem then feeding back those results to the real world to achieve that state by automatically controlling the environment, influencing people's behavior, etc. There is therefore a need for base functions that connect DTs to conduct total optimization calculations and provide the results of those calculations to external systems via an application programming interface (API). There is also a need for an application that encourages behavior modification by providing a person with optimal content by an optimal means of communication at an optimal time.

2. Digital-twin-integrated platform

The digital-twin-integrated platform (DT-integrated platform) achieves the base functions described above. It corresponds to the DTC layer in urban DTC architecture [1] and consists of a DT-connection function, total-optimization function, API, etc. (**Fig. 3**). The DT-connection function is an input/output interface to the DTs. It takes the predicted values of each DT as input to the total-optimization function and outputs the results of total-optimization calculations for environmental control (such as air-conditioning control) to specific DTs. The total-optimization function executes DTC as the core function of the DT-integrated platform. The API, meanwhile, is an input/output interface to external systems. It takes the log obtained from the customer-contact app and outputs the results of total-optimization calculations for influencing people's behavior to the customer-contact app.

The total-optimization function presents two main issues. The first is formulating the multi-objective optimization problem for multiple linked DTs, and the second is determining the Pareto front in a realistic period.

First, problem formulation means determining how to provide the results of striking a balance among

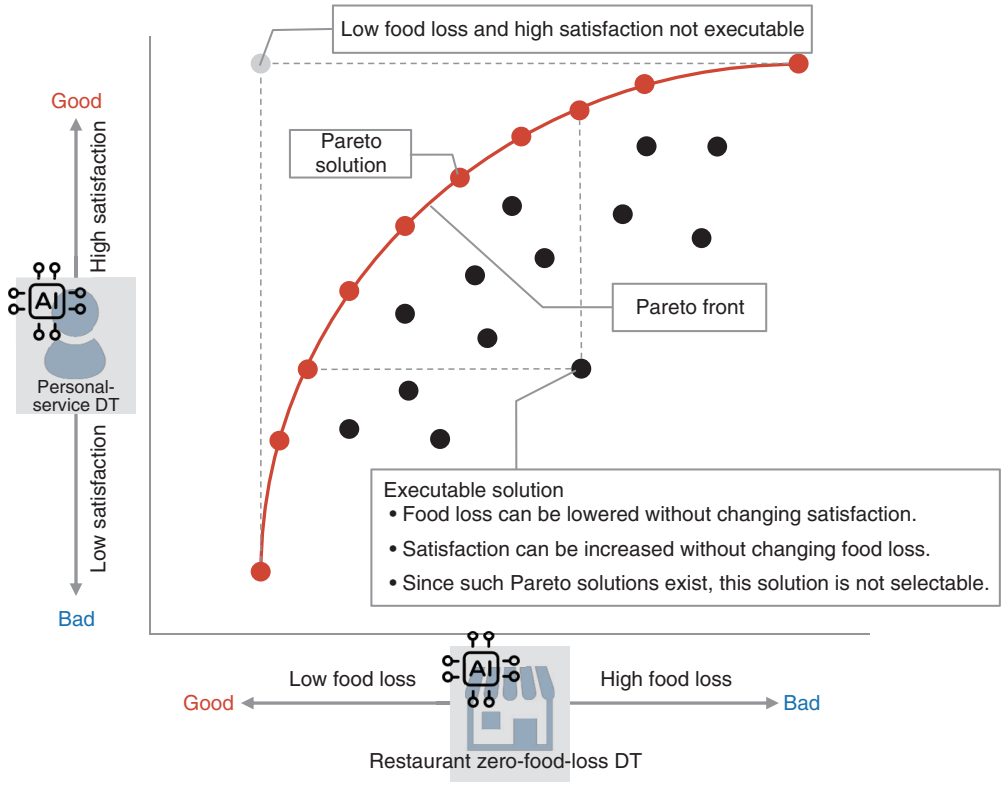


Fig. 2. Pareto front of food loss and satisfaction.

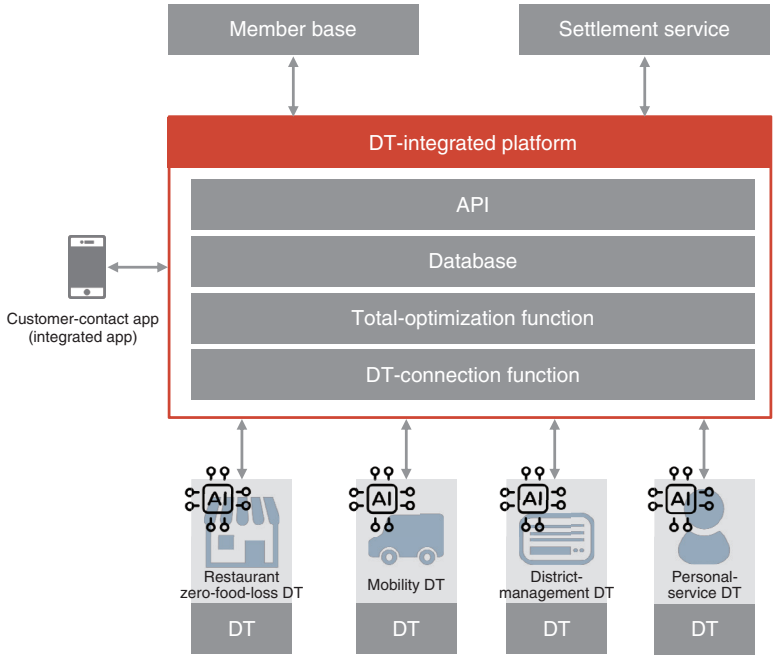


Fig. 3. Configuration of DT-integrated platform.

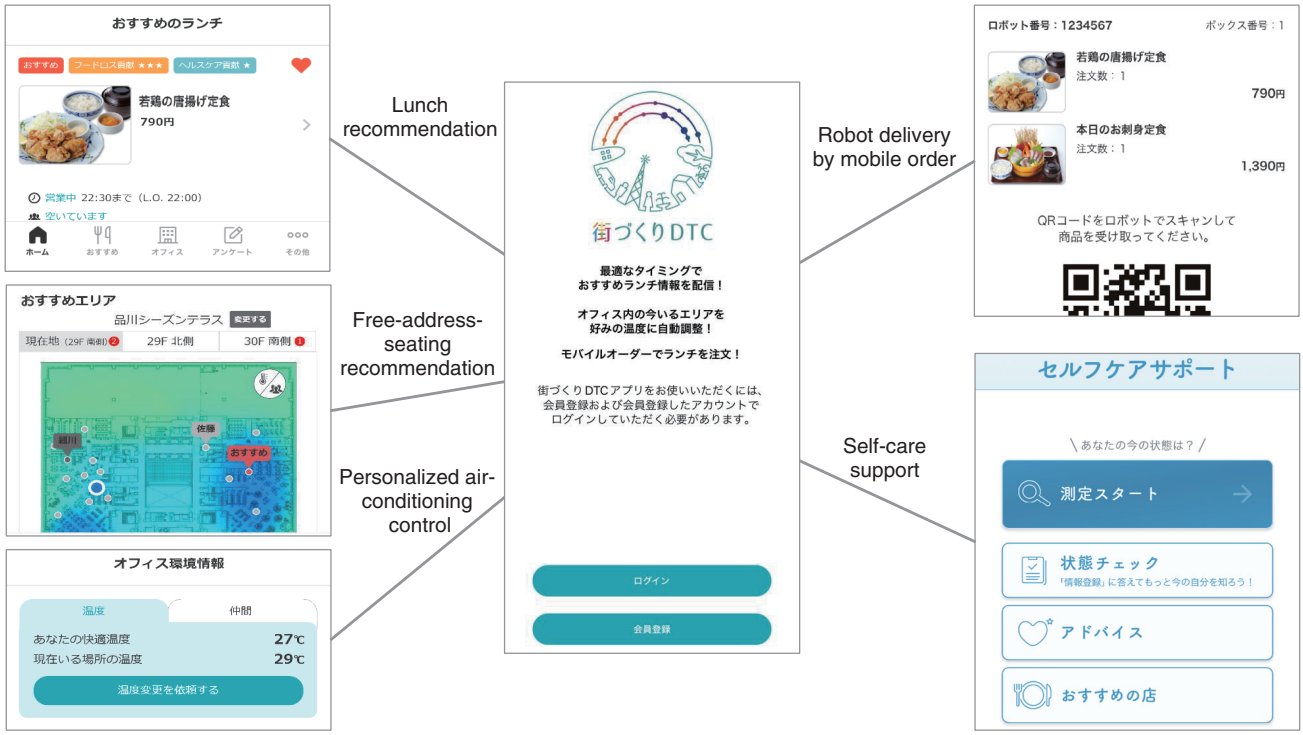


Fig. 4. Integrated app screen.

multiple DTs as value in the form of concrete services given certain types and a certain number of DTs, which means that such formulation can take on different forms. Ideally, total optimization would be executed immediately after connecting DTs to the DT-integrated platform, but automatically interpreting the complex interaction among people, things, and the environment is difficult. We therefore envisioned a variety of specific services, such as lunch recommendation and free-address-seating recommendation, and formulated the problem for each. Details of those two services are described later as case studies.

Next, we solved the problem of determining the Pareto front in a realistic period by dividing total optimization into two parts: overall adjustment and partial modification. Overall adjustment is executed in a state in which real-world objects (restaurants, people, etc.) corresponding to each DT are inactive such as during the night or early morning. It is a computational process that sets up an all-day action plan based on the predicted value of each DT. Partial modification, on the other hand, is executed in response to a previously set trigger condition (e.g., food loss surpasses a certain threshold) in a state in which real-world objects corresponding to each DT are active. It is a computational process that compensates for the differences in the action plan created in the overall adjustment process. Overall adjustment and partial modification can be used to partially change the action plan depending on the situation even while setting up the overall action plan. This approach prevents a situation in which overall adjustment is repeatedly executed whenever a trigger occurs so that the next computational process begins before the previous computational process completes, which would prevent computational processing from being completed in a realistic period.

3. Integrated app

The integrated app encourages behavior modification as described earlier. It corresponds to urban DTC architecture [1] and the customer-contact app in Fig. 3. This app provides services oriented to the district user such as lunch recommendation, free-address-seating recommendation, personalized air-conditioning control, robot delivery by mobile order, and self-care support (**Fig. 4**).

To achieve a state of social well-being envisioned

with the DT-integrated platform, it will be necessary to encourage people to behave as envisioned. A commonly used approach to this end had been to determine user preferences based on past service-usage logs and to encourage behavior through recommendations based on those preferences. However, in a district that provides multiple services, making such recommendations on a service-by-service basis is insufficient from the viewpoint of improving the user experience of value within an urban district. We believe that this can be achieved by connecting user experiences of value spanning multiple services as a series of behavioral paths (journey), determining user preferences across multiple services, and predicting user behavior on the basis of transitions in service usage. We call this process "value-journey-type behavior prediction."

To achieve value-journey-type behavior prediction, a single identifier (ID) must be used to continuously keep track of the behavior of a user using multiple services. The integrated app achieves this by executing authentication with the member base connected to the DT-integrated platform, enabling one-stop use of the services provided by DTs and DTC with a single ID. The usage logs of multiple services are accumulated in the DT-integrated platform as training data, enabling look-ahead behavior prediction based on pre- and post-behavior context.

4. Case studies

The following introduces lunch recommendation and free-address-seating recommendation as specific case studies of using the DT-integrated platform and integrated app.

4.1 Case study 1: Lunch recommendation

The integrated app notifies the user of a lunch recommendation when arriving at the office or heading out to lunch. It determines a recommended lunch (time, menu, means) for each user through total optimization of three DTs: the personal-service DT, restaurant zero-food-loss DT, and mobility DT[*3]. The district-management DT controls air conditioning considering that the number of customers visiting a restaurant changes in accordance with recommendations. If all members take lunch as recommended, a totally optimized state can be achieved that strikes a balance among satisfaction, food loss (difference between the predicted and actual values of the sales figures for each menu item), and the usage rate of robot delivery. By controlling air conditioning in a proactive manner in accordance with predictions of an increase or decrease in people flow, a totally optimized state can also be achieved that strikes a balance between thermal comfort and energy cost.

In making a lunch recommendation, food loss and usage rate of robot delivery are treated as a penalty with regard to user satisfaction, so we formulated this recommendation by converting the multi-objective optimization problem for three DTs to a single-objective optimization problem using the weighting method[*4]. In single objective optimization, no Pareto front is formed since an optimal solution is uniquely determined, but if the solution space is a convex set, the set of optimal solutions obtained by varying the weights becomes a Pareto front. Since we can imagine the owners of district businesses thinking of a district in the manner of "I would like to create a district with high user satisfaction" or "I would like to create a district with little food loss," we adopted a design that approaches the actualization of various district concepts by varying the strengths of the weights.

The concept of recommending any one of three menu items to each of three users is shown in **Fig. 5**. All members like ramen but differ in the degree to which they like sushi and bread. The situation is such that food loss is likely to occur for sushi that includes raw fish and that there is leeway in the usage rate of robot delivery. The results of total optimization for recommending lunch are shown at the right of the figure. Compared with a satisfaction-prioritized district, the balance-prioritized district results in no significant drop in satisfaction while lowering food loss and improving the usage rate of robot delivery. Business owners in the district may even desire to lower food loss even more and increase the usage rate of robot delivery even if satisfaction drops further. They can make adjustments in this regard by varying the strengths of the weights. Since the flow of people headed to restaurants will decrease through the use of robot delivery, this situation can be reflected in air-conditioning control such as by turning down air conditioning on the restaurant floor.

In a demonstration experiment involving lunch recommendations conducted at the URBANNET

*3 Mobility DT: The DT of a mobile object created using the data from a sensor installed on the mobile object. The mobility DT used in lunch recommendation is achieved by making the delivery robot into a DT.

*4 Weighting method: A method for converting a multi-objective optimization problem into a single-objective optimization problem by multiplying each objective function in the multi-objective optimization problem by a weight and summing the results to obtain a new objective function.

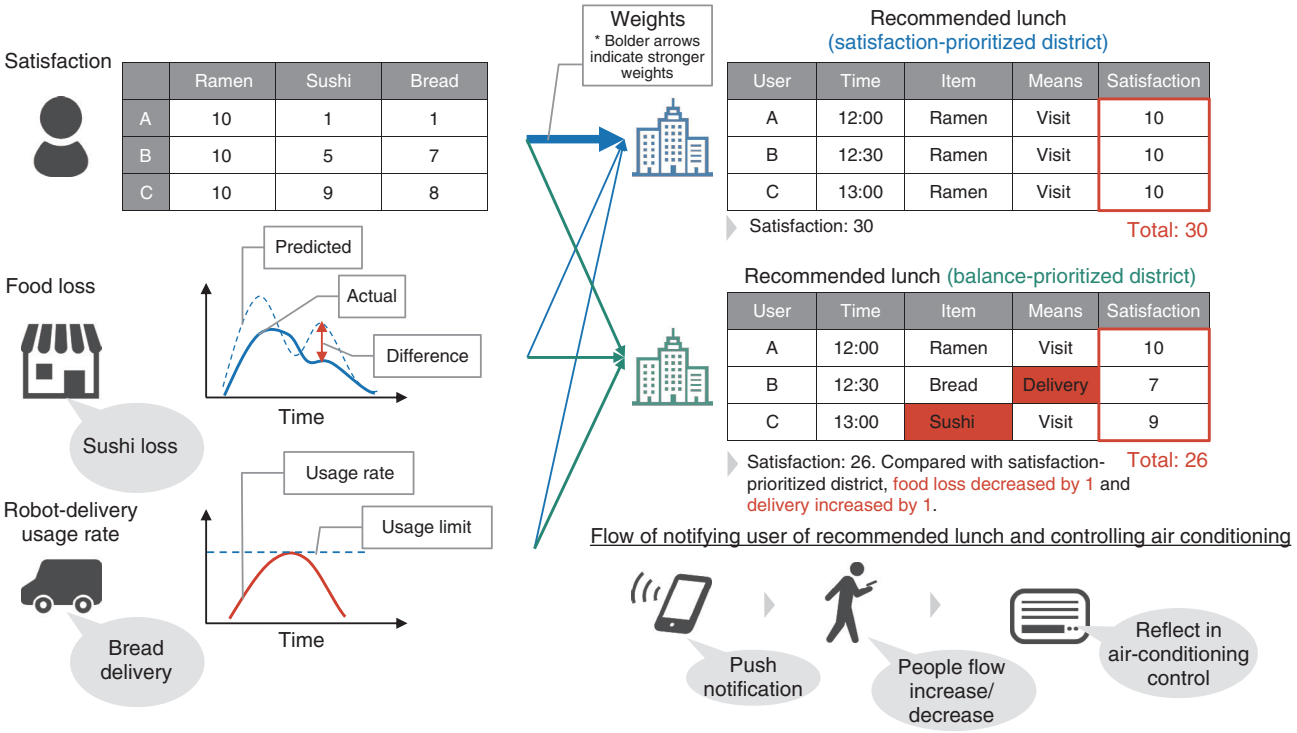


Fig. 5. Concept of making lunch recommendations.

NAGOYA nexta BUILDING from February to March 2023 [2], it was confirmed that the linking of four DTs could operate as expected.

4.2 Case study 2: Free-address-seating recommendation

In this case study, the integrated app recommends an available seat at which work can be efficiently done to users arriving at work. It determines which seat to recommend through total optimization of two DTs, personal-service DT and district-management DT. If all members take their seat as recommended, a totally optimized state can be achieved that strikes a balance between thermal comfort (temperature at which work can be comfortably done) and distance from people that the user wants to work with.

To make free-address-seating recommendations, we formulated the problem by treating thermal comfort as a constraint on the distance between the user and people that the user wants to work with and converting the multi-objective optimization problem for two DTs to a single-objective optimization problem using the constraint method[*5]. However, there are problems that must be addressed when using the constraint method. For example, the Pareto front cannot be understood unless the constraint is varied multiple times, and a solution may not be found if the constraint is excessively severe. If thermal comfort is expressed in terms of the difference in temperature at the user's desired seat and recommended seat, however, the range of acceptable temperature difference can be estimated to some extent beforehand, so we judged that there were no problems in this use case. Considering that taking a seat or changing seats as recommended may not necessarily occur, we decided to conduct optimization calculations assuming that the rate of moving to a recommended seat depends on the user and recommend a section consisting of users with a low moving rate to a user with a high moving rate.

The concept of arranging seating for ten groups of people each consisting of ten individuals is shown in **Fig. 6**. It is assumed that people have temperature preferences in terms of heat or cold sensitivity even within the same group and that some groups have a relationship in terms of wanting to work together, as

*5 Constraint method: A method for treating one of the objective functions in a multi-objective optimization problem as the objective function and the other functions as constraints to convert the problem into a single-objective one.

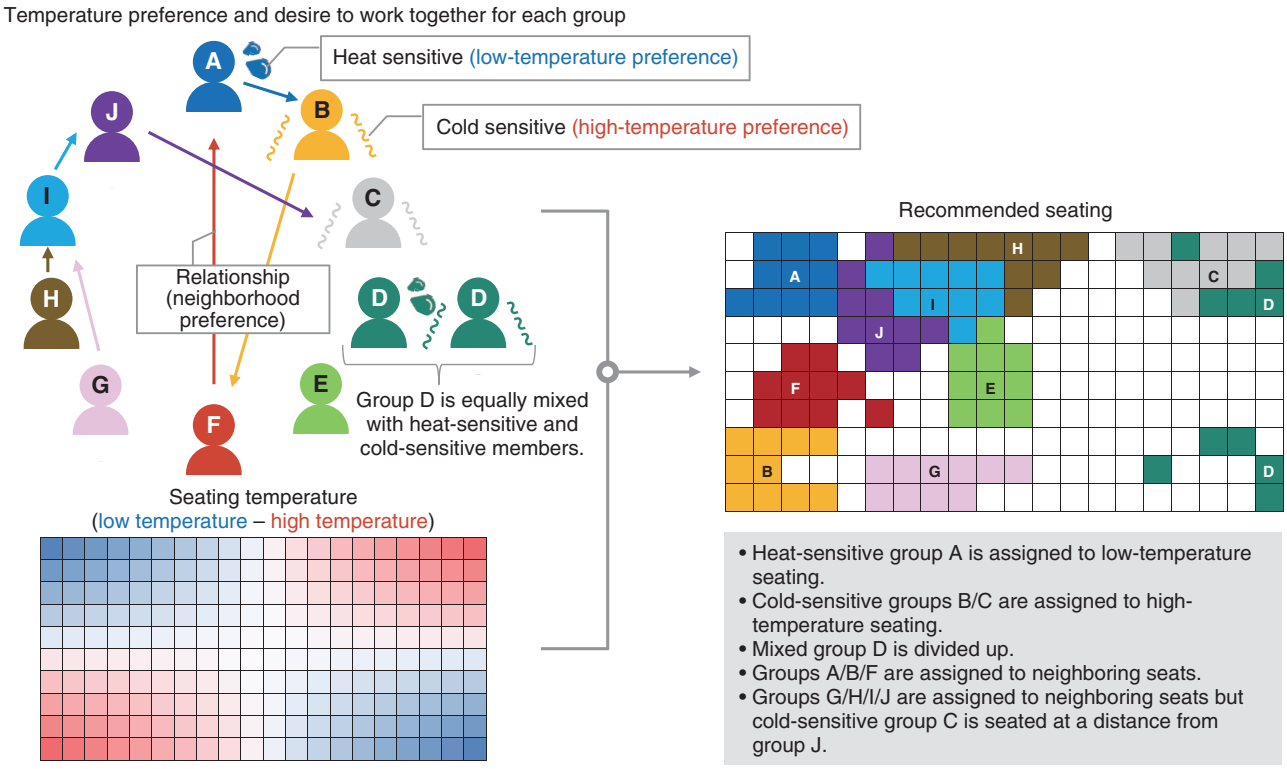


Fig. 6. Concept of making free-address-seating recommendations.

indicated with the arrows in the figure. It is also assumed that we have a state that allows for conditions that produce uneven temperatures instead of uniform cooling or heating, thus reducing energy costs. The recommended seating shown at the right of Fig. 6 is the result of total optimization. Heat- and cold-sensitive groups are seated in accordance with seat temperatures and that groups who wish to work together are seated near each other.

We have been conducting a demonstration experiment on making free-address-seating recommendations since November 2023 at an office building owned by an NTT Group company.

5. Future developments

Along with an increase in DTs for each industry domain, we can expect an increase in use cases that will require trade-off solutions. We can also expect specific initiatives in modeling by urban DTC to accumulate from here on and progress to be made in achieving a general-purpose platform by identifying and consolidating common functions. To achieve a state of social well-being envisioned by the DT-integrated platform, it will be necessary to encourage people to behave as envisioned. There are limits, however, to the information-recommendation approach, so we plan to implement a more effective approach in the integrated app that drives behavior modification through, for example, nudges and incentives.

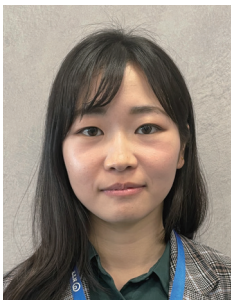
Going forward, we will continue to promote research and development to achieve smart cities in which an entire district takes shape in cyber space by harmonizing multiple DTs and district experiences for the individual are obtained in physical space.

References

[1] C. Yamamoto, I. Shake, S. Fukada, and S. Ueno, "Data-driven and Optimized Smart Cities Using Urban DTC," NTT Technical Review, Vol. 19, No. 1, pp. 45–51, 2021. https://doi.org/10.53829/ntr202101fa7
[2] M. Kamiya, T. Hata, D. Asahi, and A. Hoshino, "Demonstration of 'Urban DTC®' at URBANNET NAGOYA nexta BUILDING," Business Communication, Vol. 60, No. 2, pp. 18–20, 2023 (in Japanese).

Jun Ito
Senior Research Engineer, NTT Smart Data Science Center, NTT Computer and Data Science Laboratories.
He received a B.E. and M.E. from Waseda University, Tokyo, in 2009 and 2011. He joined NTT in 2011, where he worked on research related to social media and location-based analytics. In 2015, he moved to NTT Resonant, Inc., where he worked as a project manager for system development and sales engineer of AI services. He returned to NTT in 2020 and has been engaged in research and development related to smart cities.

Hiroyuki Sato
Research Engineer, NTT Smart Data Science Center, NTT Computer and Data Science Laboratories.
He received a B.E. and M.E. from the University of Tokyo in 2012 and 2014. He joined NTT DOCOMO, INC. in 2015, where he was engaged in research and development related to data science. He moved to NTT in 2022 and has been engaged in research and development related to smart cities.



Ai Hoshino
Researcher, NTT Smart Data Science Center, NTT Computer and Data Science Laboratories.
She received a B.E. and M.E. from Tokyo University of Marine Science and Technology in 2019 and 2021. She joined NTT in 2021 and has been engaged in research and development related to smart cities.

Personalized Air Conditioning Achieving Both Personal Comfort and Energy Savings

Satoshi Mieda, Sota Sakai, and Takahiro Hata

Abstract

Modern offices use uniform air-conditioning control, which means that office space cannot provide a comfortable environment for everyone since some people will feel cold and others will feel hot. Our aim is to achieve comfortable offices that make it easier for office users to do their work and improve their performance by providing air conditioning at temperatures tailored to personal preferences. In this article, we introduce our work in space control for achieving both energy saving and personal comfort.

Keywords: air-conditioning control, AI, personalization

1. Introduction

At NTT Smart Data Science Center (SDSC), we have developed an air-conditioning-control-scenario calculation technology, called predicted mean vote (PMV)[*1] air-conditioning technology [1], for reducing energy consumption while maintaining a comfortable environment for visitors in the common sections of buildings by predicting an increase/decrease in the number of visitors and executing optimal control of air conditioning. We succeeded in reducing energy consumption by up to 50% in a building's common sections in which people move about compared with exclusive sections by maintaining a temperature at which many people feel comfortable without taking personal preferences into account. This was accomplished by predicting an increase/decrease in the number of people so that air conditioning could be controlled efficiently. However, there can be considerable dissatisfaction in exclusive sections in which people tend to stay long if personal preferences are not met, so there is a need for air-conditioning control oriented to temperatures tailored to personal preferences. To improve energy efficiency throughout an entire building, therefore, we investigated personalized air-conditioning technology for reducing energy consumption while satisfying individual preferences.

We describe our approach to developing our personalized air-conditioning technology that satisfies personal preferences in a single space, such as an office, in contrast to conventional air-conditioning control for an even and uniform environment.

2. Goal with personalized air-conditioning technology

The usage format of office buildings can be broadly divided into common sections and exclusive sections. Corridors and lobbies, for example, are generally called common sections in which people rarely stay a long time and are usually moving about. In a common section, the time that one person stays is relatively short with many people come and go all the time. This makes it important to create an environment that feels comfortable to many people. However, in exclusive sections occupied by a variety of tenants in an office building, users stay for a relatively long time. Therefore, individual differences in what constitutes comfortable temperature and dissatisfaction with the temperature can significantly affect user comfort compared with common sections. Thinking about

*1 PMV: An index of a person's thermal comfort.

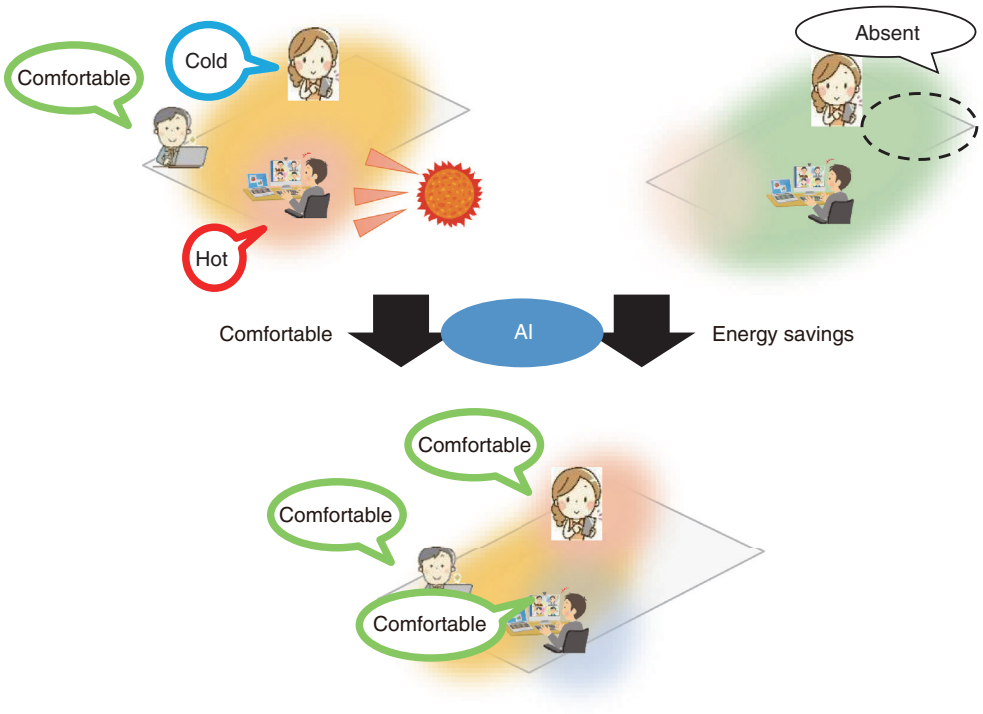


Fig. 1. Air-conditioning control with personalized air-conditioning technology.

specific action, user behavior in an exclusive section can vary from individual work to meetings with several people, and places that are used by people within an office can likewise vary. Since there are individual differences in temperature and humidity preferences that are commonly expressed as sensitivity to heat or cold, and given that people tend to stay for a relatively long time in an exclusive section, we consider it necessary to provide comfortable air-conditioning control in a flexible manner and an environment that satisfies every individual.

A comfortable environment is important in an exclusive section, but air conditioning consumes nearly half of a building's energy consumption [2]. In seasons in which air conditioning is used more than usual (summers and winter), the pursuit of comfort generally requires much energy, and executing air-conditioning control simply for comfort's sake can lead to an increase in energy consumption. At SDSC, we aim to achieve both personal comfort and energy saving for exclusive sections, the same as with our PMV air-conditioning technology for common sections described at the beginning of this article.

With the spread of remote working, it can no longer be assumed that all seats in an office will be used, and seats and areas that are hardly used is increasing. It would therefore not be desirable to maintain a uniform temperature in all places within an exclusive section from the viewpoint of energy consumption. For this reason, our plan for personalized air conditioning is to create a variation in temperature not in the vertical direction but in the horizontal direction by creating locations with strong air conditioning and others with weak air conditioning by taking into account where individuals are located. We thus seek to achieve both personal comfort and energy saving, as mentioned above (**Fig. 1**).

3. Comparison with conventional building air-conditioning control

3.1 Conventional air-conditioning control

In office air conditioning, the typical approach is to set a temperature for an entire room or section and control air conditioning to achieve a uniform temperature. Thus, it would not seem possible to provide everyone with a comfortable environment since the type of work that people do and personal preferences with respect to temperature differ. If both hot and cold areas exist within an office, the blowout

temperature of air-conditioning vents setting for either a hot or cold location may be changed on the basis of a request from a user. However, when changing temperature, the temperature in the vicinity of the requester will also change, affecting the comfort of surrounding users. The spread of remote working and the resulting phenomenon of sparsely occupied offices have made the energy consumed by applying uniform air-conditioning control at a fixed temperature an issue that needs to be addressed.

3.2 Air-conditioning-control-scenario calculation technology (PMV air-conditioning technology)

Our PMV air-conditioning technology achieves both common comfort and energy saving in the common sections of a building. PMV is an index of personal comfort determined, for example, from temperature, humidity, heat radiation, air currents, amount of activity, and amount of clothing. For a common section, such as a corridor or lobby, in which many people are coming and going, our PMV air-conditioning technology has been successful in reducing energy consumption by up to 50% by predicting an increase or decrease in the number of people while keeping the PMV within a recommended comfort range [3]. This technology can be used to keep many people comfortable, but it cannot achieve a level of comfort supporting personal temperature preferences as required in an exclusive section of a building. We therefore investigated air-conditioning technology tailored to personal temperature preferences (personalized air-conditioning technology) based on air-conditioning control by using reinforcement learning we studied for our PMV air-conditioning technology.

3.3 Personalized air-conditioning technology

In contrast to conventional air-conditioning control that operates air conditioning in a uniform manner at all locations whether or not users are present, the aim for exclusive sections of a building is air-conditioning control for efficiently achieving both personal comfort and energy saving by creating variation in temperature. This technology calculates temperature variation that should reflect the personal temperature preferences of everyone in an exclusive section as an ideal temperature distribution and determines an air-conditioning-control scenario by using artificial intelligence (AI) and setting this ideal temperature distribution as the objective. This air-conditioning-control scenario presents an operation plan that describes how the airflow and temperature of multiple air-conditioning vents should be set and how air-conditioning control is executed. Our aim is to achieve both personal comfort and energy saving by automatically controlling air conditioning in accordance with this scenario. We next describe the configuration of our personalized air-conditioning technology.

4. Configuration of personalized air-conditioning technology

To reproduce an ideal temperature distribution with temperature variation in an actual office on the basis of the personal preference of each user, our personalized air-conditioning technology consists of three major components: temperature-distribution generation, air-conditioning agent, and seating recommendation (**Fig. 2**).

The temperature-distribution-generation component calculates an ideal temperature distribution that should be achieved within an office according to each user's temperature preference and office-attendance conditions. To determine beforehand whether a user plans to be at the office, this component makes a prediction using the office-attendance-prediction function. By taking into account each user's temperature preference may change in accordance with outside temperature and other factors, it estimates the temperature then calculates an ideal temperature distribution that should be achieved within that office by calculating the ideal temperature distribution on the basis of the temperature-preference and office-attendance information determined beforehand.

Targeting the ideal temperature distribution calculated described above, the air-conditioning-agent component sets up an air-conditioning-control policy in a simulation environment, calculates an optimal air-conditioning-operation scenario, and reproduces a temperature distribution with temperature variation.

Finally, the seating-recommendation component recommends work locations corresponding to user comfort from a future temperature distribution on the basis of air-conditioning-control prediction.

The following sections describe these components in more detail.

4.1 Temperature-distribution generation

The temperature-distribution-generation component generates the target temperature distribution (ideal temperature distribution) required by the air-conditioning agent to control air conditioning. It first

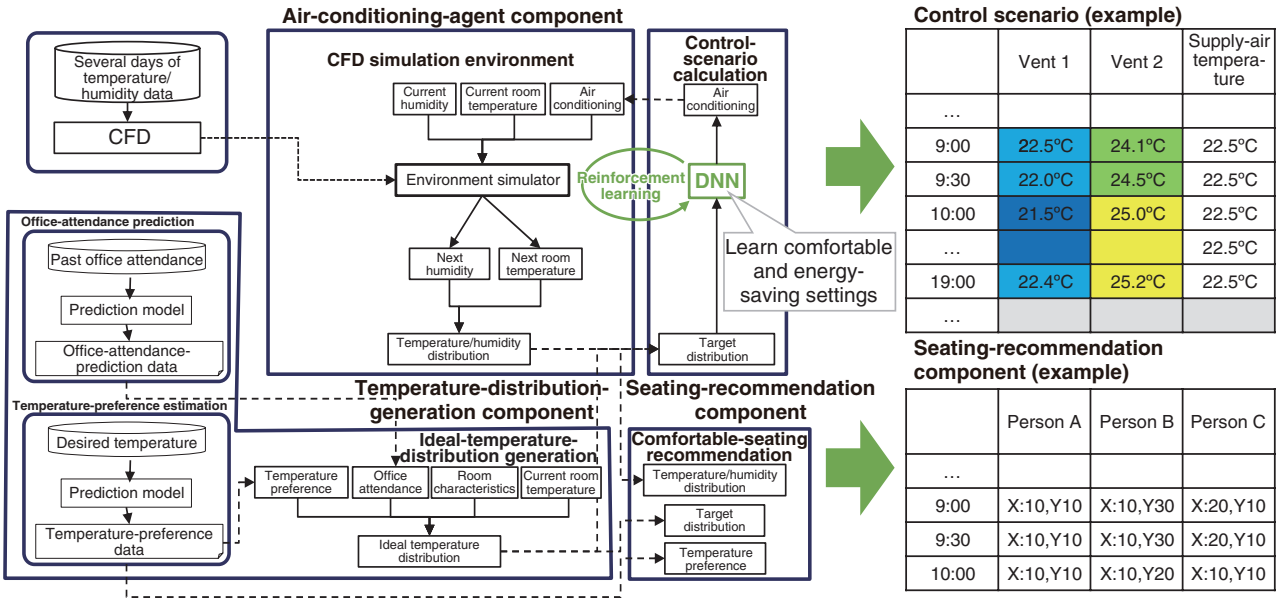


Fig. 2. Configuration of personalized air-conditioning technology.

estimates user-temperature preferences using a history of user orders for desired temperature to create locations reflecting those user-temperature preferences. It then predicts office attendance for each user to control air conditioning in accordance with user presence in the office. Finally, it calculates an ideal temperature distribution from the above information and air-conditioning-load characteristics within the office (room characteristics) for use by the air-conditioning agent as its target distribution.

(1) Temperature-preference estimation

Although it is possible to use temperature preferences based on self-reports from users to generate an ideal temperature distribution and achieve temperature variation by controlling air conditioning accordingly, it is thought that accurately understanding one's own temperature preference would likely be difficult for a user. To understand users' preferred temperatures, we obtain orders from users for desired locations or temperatures within an office via an integrated app, for example. In short, this function estimates the temperature preference of each user from historical information on temperatures at locations where users have been present within the office and on desired temperatures such as higher temperatures (when it is cold) or lower temperatures (when it is warm).

(2) Office-attendance prediction

It is necessary to determine whether a particular user will be coming to the office to generate a temperature distribution that reflects the user's preferred temperature after arriving at the office. This function accumulates a history of attendance for each user as training data via an integrated app or other means and predicts whether the user will be at the office so that an ideal temperature distribution can be generated.

(3) Ideal-temperature-distribution generation

This function generates an ideal temperature distribution using temperature preferences and attendance prediction described above, information on room characteristics obtained by analyzing the trend of warming and cooling at each office location, and the temperature distribution at that time. Even within an office, there are locations where temperature can easily rise or fall due to factors such as outside air temperature, sunlight, and placement of air-conditioning vents, which can affect air-conditioning load. To achieve air-conditioning control with good energy efficiency, room characteristics need to be determined by analyzing temperature distributions within the office during times of air-conditioning control over several days. These room characteristics, information on user temperature preferences, and information on office attendance can be used to generate an ideal temperature distribution that can save energy while satisfying user comfort.

4.2 Air-conditioning agent

The air-conditioning-agent component uses an environment simulator that reproduces an office environment to output an air-conditioning-control scenario for reproducing an ideal temperature distribution on the basis of reinforcement learning. An air conditioner that automatically executes air-conditioning control according to this air-conditioning-control scenario changes the settings of vents, supply-air temperature, etc. at different points in time and performs energy-saving air-conditioning control that also satisfies personal comfort.

Our PMV air-conditioning technology also calculates and controls an air-conditioning-control scenario using an environment simulator and reinforcement learning. Our personalized air-conditioning technology achieves energy-saving air-conditioning control that satisfies personal comfort by constructing a detailed simulation environment with finely arranged temperature sensors, preparing a target temperature distribution that sets a temperature for each user seat as a control target, and applying reinforcement learning in combination with multiple control targets such as supply-air temperature, blowout temperature (air volume) of vents, and airflow devices.

Both our PMV air-conditioning and personalized air-conditioning technologies calculate a control scenario through reinforcement learning of air-conditioning control. In evaluating rewards such as energy efficiency through air-conditioning control, the air-conditioning load may drop at a certain instant but may also increase later resulting in an increase in energy consumption, so such an evaluation cannot be conducted only at that instant. It is therefore necessary to evaluate whether energy consumption can be minimized while achieving target temperatures by air-conditioning control over the course of a day. To execute complicated control of control points such as air-volume output from multiple vents and temperature of air supplied to each vent and to conduct long-term evaluations, we have been calculating air-conditioning-control scenarios using reinforcement learning.

(1) Computational-fluid-dynamics simulation environment

Executing air-conditioning control by reinforcement learning requires searching for optimal air-conditioning control by changing and controlling a variety of settings such as supply-air temperature and blowout temperature in accordance with temperature conditions within the environment and acquiring an optimal control method toward the target temperature distribution. This requires tens of thousands of trials using many conditions and control settings, so conducting actual trials in an actual space is difficult. For this reason, we reproduce environments using computational fluid dynamics (CFD)[*2] that can simulate temperature and airflow. Applying ordinary CDF requires the setting of complex parameters using specialized knowledge in relation to fluid dynamics, but we estimate CFD parameters by placing multiple temperature sensors within an exclusive section. We simulate air-conditioning control using this environment.

(2) Control-scenario calculation by reinforcement learning

We use reinforcement learning in a simulation environment to calculate an air-conditioning-control scenario. Using this simulation environment, we learn how air-conditioning supply-air temperature and air volume at a certain temperature distribution can be set to achieve the target temperature distribution while minimizing energy consumption. Using such a simulation environment enables a trial-and-error process with many variations, but searching for optimal air-conditioning control for a variety of target temperature distributions results in high computational complexity. To solve this problem, we efficiently search for optimal air-conditioning control by sampling locations such as seats where users are assumed to be present from the temperature distribution reproduced in the simulation environment and the target temperature distribution and by using the difference between the reproduced temperature and target temperature at each sampling point as training input (**Fig. 3**).

4.3 Seating recommendation

The seating-recommendation component recommends a location within a room where time can be passed in a comfortable manner to each user on the basis of the user's temperature preference, orders placed by the user requesting office temperature to be changed, and the current-temperature distribution and future-temperature distribution to be reproduced in the room.

(1) Current-temperature distribution

Proposing an optimal seat or location within an office when the user wishes to move requires the temperature distribution at that time. However,

*2 CFD: A method for simulating fluid movement by solving the equations governing the movement of a fluid. For our personalized air-conditioning technology, we use CFD to reproduce temperature changes within a room.

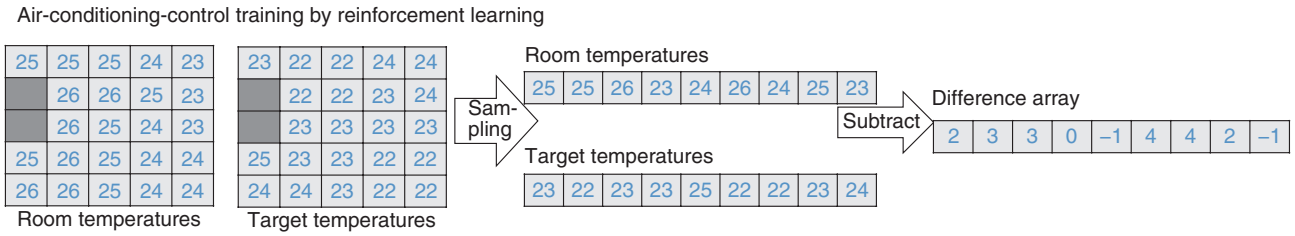


Fig. 3. Temperature-distribution sampling.

installing temperature sensors at all seats or locations within an exclusive section to measure the temperature at each location is difficult in terms of cost and securing those installation locations. Therefore, we use temperature information obtained from a number of discretely placed temperature sensors to predict the temperature distribution throughout the office and calculate the current-temperature distribution. We construct a prediction model for office temperature from the history of temperature-sensor information at each point to predict temperature even at locations where no temperature sensor is installed.

(2) Future-temperature distribution

When a user comes to work or decides to move seats, the temperature distribution to be reproduced in the future is needed. The temperature distribution at different time points in the future is determined on the basis of the operation scenario computed by the air-conditioning agent and of air-conditioning control in the simulation environment.

(3) Comfortable-seating recommendation

When a user arrives at the office or places an order requesting a change in temperature, this function recommends a location that satisfies user comfort or a location that will satisfy it in the future. When the comfort of other users would be affected or the energy efficiency of air conditioning would be degraded, there are times when it would be more appropriate to suggest people to change seats rather than make major changes to air conditioning to satisfy the comfort of many users and reduce energy consumption. This function calculates the location that would satisfy the user's temperature preference over the long term from the temperature distribution in the office at that time and temperature distribution planned for the future and recommends a location that matches the user's temperature preference.

5. Issues and future outlook

The following issues must be addressed in implementing our personalized air-conditioning technology.

(1) Prediction of user movement within a room

We can predict whether a user will be coming to work, but to control air conditioning tailored to user movement, it will be necessary to predict where users will be located within the room and reflect that in the ideal temperature distribution.

(2) Improving the resolution of reproduced temperature distributions

To control the desired temperatures of many users with pinpoint accuracy, it will be necessary to achieve high resolution in both the spatial and temporal directions such as by increasing the control targets that act on air conditioning.

Going forward, we plan to verify the effectiveness of seating recommendations in an office tailored to users' preferred temperatures and air-conditioning control tailored to actual user behavior.

References

[1] Website of NTT R&D, "Air Conditioning Control Scenario Calculation Technology," https://www.rd.ntt/e/research/sds0003.html

[2] The Energy Consumption Center, Japan, "Energy Consumption at Offices" (in Japanese), https://www.eccj.or.jp/office_bldg/01.html

[3] Press release issued by NTT Corporation, East Japan Railway Company, NTT FACILITIES, Inc., and NTT DATA Corporation, "Demonstrating both energy conservation and comfortable environments through feed-forward AI air-conditioning control based on small-volume learning," Nov. 1, 2021. https://group.ntt/en/newsrelease/2021/11/01/211101a.html

Satoshi Mieda
Research Engineer, NTT Smart Data Science Center, NTT Computer and Data Science Laboratories.

He received an M.I.S.T. from Osaka University in 2011 and joined NTT the same year. He studied collaborative work and video communication at NTT Media Intelligence Laboratories. From 2013 to 2017, he was involved in service development at NTT Resonant. Since 2017, he has been active in the development of smart-city applications and data analysis at NTT Smart Data Science Center.

Sota Sakai
Researcher, NTT Smart Data Science Center, NTT Computer and Data Science Laboratories.

He received an M.E. in information science from Osaka Prefecture University in 2019 and joined NTT COMWARE the same year. From 2019 to 2023, he was involved in the technology verification of image-recognition AI and edge AI. In July 2023, he moved to NTT and has been active in the development of smart-city applications and analysis of data at NTT Smart Data Science Center.

Takahiro Hata
Senior Research Engineer, NTT Smart Data Science Center, NTT Computer and Data Science Laboratories.

He received an M.Inf. from Kyoto University in 2006 and joined NTT the same year. He studied ubiquitous services at NTT Network Innovation Laboratories. From 2010 to 2013, he was involved in the development of public wireless local area network services at NTT WEST. Since 2013, he has been active in the development of smart-city applications and analysis of pedestrian flow and driving data of vehicles at NTT Smart Data Science Center.

Using Smart-store Behavior Data to Optimize Sales Promotion

Sunyeong Kim, Keita Nishimoto, Masahide Mizushima, Setsuo Yamada, and Junji Tomita

Abstract

Implementation of information and communication technology in retail stores is advancing, with self-scan shops using smart carts and others that enable "empty-handed" purchasing without even going to a register. In such shops, we can obtain point-of-sale data including customer identifiers and behavior data such as movement flows within the store, and that enables us to understand a customer's behavior leading to a purchase. This was not possible with conventional shops. This article introduces an initiative that uses such data to create a digital twin of a store, which is used for sales promotion or more efficient operations.

Keywords: smart store, marketing, flow classification

1. Smart store expansion and store digital twins

Information and communication technology (ICT) is being deployed in brick-and-mortar retail establishments, such as convenience stores and supermarkets, to improve the shopping experience for customers, reduce personnel costs, or increase earnings. For example, in December 2016, Amazon created a stir when it opened the first Amazon Go store in Seattle, Washington. The store uses cameras and sensors to recognize customers and products automatically, enabling customers to complete purchases by simply picking up products and leaving the store. Following in the footsteps of Amazon Go, Shanghai company, Cloudpick, developed a similar retail system, which is being used to develop more than 200 stores in 11 countries around the world, including Japan, USA, Germany, France, Singapore, and South Korea. San Francisco, California company, Zippin, is also developing about ten stores, mainly in the USA. Another Seattle company, AiFi, is launching a system that does not use sensors and determines purchases using only image recognition. In Japan, major convenience store chains established trial stores between 2020 and 2021. In particular, Family Mart quickly established their system in collaboration with TOUCH TO GO Co. Inc., and is steadily adding stores using the system.

In the supermarket domain, Trial Holdings began deploying a smart cart (RegiCart) system that is able to scan products in February 2018. Aeon also started its "RegiGo" service in March 2020, implementing smart carts with a dedicated terminal, and as of 2023, it is expanding this system to approximately 200 stores throughout Japan.

We refer to retail stores that are implementing ICT in these ways as smart stores. In addition to collecting customer identifier (ID) and purchase data (ID-point-of-sales (POS) data), these smart stores are able to gather data on the position of each customer from when they enter to when they leave the store ("flow data"), and depending on the store, how much time they spent at each product shelf and what products they pick up and perhaps replace on the shelves ("reaching data"). Thus, data on both what was purchased and the process of making the purchase can be obtained. These data can be used to create a store digital twin (DT) to estimate each customer's needs and preferences, and their behavior and purchases in the store can be predicted, to target sales promotion more accurately to each customer or improve efficiency in store operation.

This article introduces a store DT initiative within a smart store. We first describe the types and characteristics of data that can be obtained in a smart store. We then discuss an analysis we conducted of problems and solutions when proposing measures more targeted to individual customers and other store improvements. We then introduce a method for classifying customer behavior using ID-POS and flow data and an initiative to propose measures that combine it with customer information for automatically selecting customer and purchase groups (segments) that have potential for increased sales.

2. Types and characteristics of data obtainable in a smart store

There are various types of customer-purchasing-behavior data that can be obtained in a smart store. The three main types are the following.

(1) ID-POS data: ID-POS data refer to ordinary purchase (POS) data that are collected using a cash register and linked with a customer ID. The customer ID can be obtained from a smartphone app when entering the store, through face recognition or other means and associated with a purchase, or by presenting a membership card or using points when making a purchase. The POS data generally include the type and quantity of a purchased product, the price, any discount or coupon used, and payment information. ID-POS data follow the sequence of a particular customer's purchases through time, so they can be used to analyze their purchase history and preferences. For example, the data can show what type of products they buy often, when they buy them, how much they buy, and what combinations of products they buy. This can be used to present products and services suited to that particular customer or facilitate cross-selling or up-selling measures.

(2) Flow data: Flow data refer to the record of where and how the customer moved through the store, expressed for example as a time sequence of coordinates within the store. Flow data can be captured using infrared or RGB cameras and used to analyze customers' purchasing process and movement patterns. The data can show, for example, the order in which a customer navigates the store or which product shelves they stop in front of and for how long. This can enable understanding of customers' interests and needs, optimization of store layout and product displays, or real-time product and service suggestions.

(3) Reaching-behavior data: Reaching-behavior data refer to recording of behaviors when customers pick up a product and possibly return it to the shelf. Most reaching-behavior data are obtained using data from weight-sensors on the shelves and can be used to analyze customers' desire to purchase or their decisiveness. For example, the data can indicate which products the customer reaches for, the probability that they will purchase it if they have picked it up, or what products they are comparing it with. This can be used to improve product pricing or placement or provide coupons or points to promote purchases.

3. Issues when suggesting measures targeting individual customers

There has been no way to record purchasing behavior of customers in a brick-and-mortar store, and analysis could not go beyond understanding the overall buyer demographic for each product from POS data using artificial intelligence (AI) cameras to understand which areas are crowded or the age or gender distribution of customers in each time block. By combining these with the ID-POS and flow data available in a smart store, it is possible to understand the process leading to a purchase for each customer and each purchase.

The following are issues that could arise with such systems.

(1) Too many patterns in purchase-behavior data: There are many variations in the processes by which customers purchase a particular product. Even considering a single path taken, there are many non-specific elements, such as exploring to find the product, sudden path changes, and time spent at the shelves deciding on a purchase. A mechanism able to extract and process the main patterns from this complex data is needed.

(2) Segment granularity considering measures that can be taken in a physical store: Compared with online stores, ways to intervene with a customer in a physical store are limited. There are measures that can be taken on an individual basis, such as using coupons or notifying an app, but there are many measures that affect the entire store, such as changing

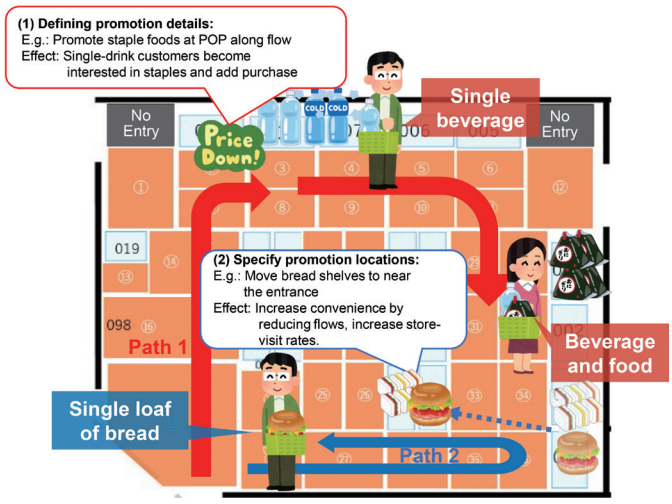


Fig. 1. Example of segments in a store and per-segment measures.

the product or location or promoting products using signage. As such, even for one-to-one marketing, it is important to separate segments with granularity that is suited to planning some physical measures.

(3) Discovering segments where intervention can be effective and estimating the effect of the measure: Data obtainable in a smart store span many branches, so a huge amount of work is required to manually select segments that can be addressed with a measure, with all their various combinations of attributes. However, very little data have been recorded regarding the effects of sales-promotion measures taken in physical stores, and an approach using machine learning to find sets of measures that are effective would require large amounts of training data and be extremely difficult.

We are conducting research and development on the following technologies to address the above issues. For issue (1), we are developing an in-store behavior-classification technology that executes clustering on complex customer behavior and purchasing patterns, accounting for constraints in a physical store, and presents them with a suitable granularity. For issue (2), we are developing a technology to combine customer, purchase, and behavior data available in a smart store, automatically find segments with potential for intervention, estimate the effect of measures taken on them, and propose such measures.

4. Technology to classify movement in retail stores to promote sales

In-store sales promotions to prompt customers to make purchases in the store are important for increasing sales. Various options are available, such as optimizing the store layout and product displays and suggesting products and services using digital signage or point-of-purchase (POP) materials. Store managers must understand the behavior of customers in the store to select and implement the best from among these in-store promotion options. Segmenting the flows of customers and their purchasing tendencies in the store and presenting them can enable the store manager to plan methods and scenarios for sales promotions targeted to each segment.

Figure 1 shows a hypothetical example in which there are three segments of customers in the store. It is assumed that in this store, there are many customers that either buy a single beverage or a single beverage and some food staples, following the path shown in red, and other customers that purchase a single loaf of bread, following the path shown in blue.

Without executing such segment classification, it is difficult to decide the content and placement of promotions in the store objectively, but if such segmentation is done, it is possible to propose specific measures, such as installing POP or signage along the route shown in red or suggesting purchase of food staples to customers purchasing a single beverage, to increase sale amounts.

We developed a technology that automates the

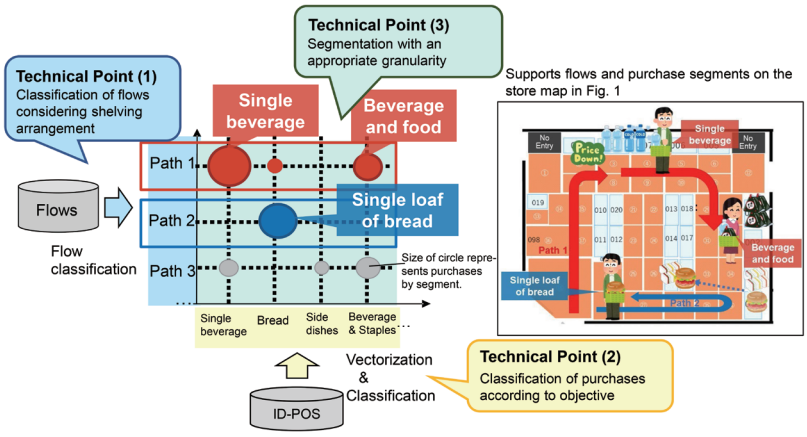


Fig. 2. Overview of in-store behavior-classification technology.

segmentation of the combinations of flows and purchases and presents the results to the store manager. To visualize behavior in a store, there are solutions that show customer-location tendencies in the store as a heat map and displaying these tendencies for each product shelf, but they are not segmented in accordance with flows or purchases.

The technology we developed (in-store behavior-classification technology) has three components: (1) a flow classifier, (2) a purchase classifier, and (3) segmentation technology for combined flows and purchases (**Fig. 2**).

(1) Flow classifier: This component takes in-store flow data as input, computes similarity among flows, groups and classifies similar flows and selects the main flow patterns. Flow data are given as a sequence of areas that each customer passes through, with sequences of varying length, so it is not possible to compute similarity using a simple vector-similarity calculation. There are also constraints such as shelving and aisles in a store, so it is possible that similar flows could be confused with others that are not actually similar if similarity is computed on the basis of only physical distances. To resolve these issues, we applied a method called dynamic time warping (DTW) [1], and used a method based on distance moved within the store to calculate similarity considering the arrangement of shelves and aisles rather than just physical distance. With these methods, we have shown improved accuracy of classification, particularly for short flows that tend to be misclassified easily.

(2) Purchase classifier: This component takes purchase data from each customer as input, creates vectors representing the purchases of each customer, groups and classifies similar purchases, and selects the main purchase patterns. When classifying purchases, product-category information already associated with products is frequently used. However, the granularity of such categories varies greatly by managing entity and categories are often focused on the ingredients rather than the customer-purchasing factors needed for sales promotion (e.g.: food staples, side dishes, beverages), so they cannot be used as-is for purchase classification. We found that a characteristic of product categories is whether products tend to be purchased individually or with other products and if with other products, what sort of products. We call this characteristic "how-purchased." On the basis of this concept, we defined the tendency for product categories to be purchased individually or with other products in terms of a vector that expresses the characteristics of the category and used it in the classification of each purchase. We showed that doing so results in classification results that are close to the results from subjective classification.

(3) Segmentation technology for combinations of flows and purchases: This technology combines the results of flow and purchase classification and executes segmentation with an appropriate granularity. For this research, we defined appropriate granularity to mean a

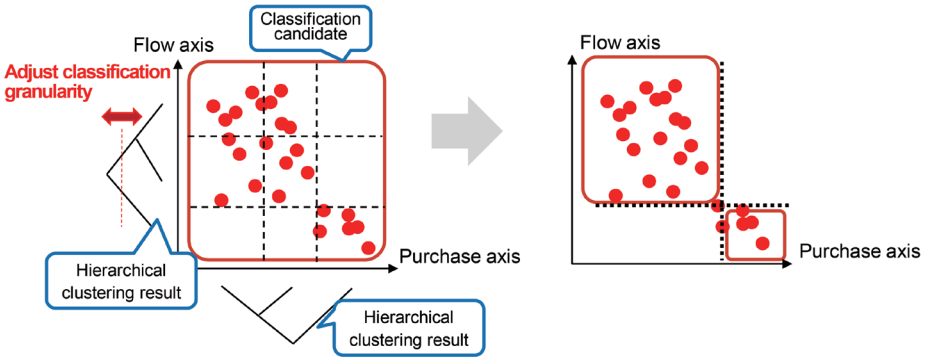


Fig. 3. Segmentation technology for combining flows and purchases.

segment granularity that results in a strong connection between flows and purchases. To implement this, we carried out hierarchical clustering on the classifications from (1) and (2) then indexed the strength of the connection between flows and purchases using the mutual-information content to identify an appropriate granularity (**Fig. 3**).

With the technologies introduced thus far, we were able to classify into segments by flow and purchase, as shown on the left in Fig. 2, enabling store managers to consider and implement sales promotions within a store in accordance with each segment.

5. Technology to automatically suggest promotion measures to optimize retail store operation

When implementing measures to increase sales or improve a store, it is generally easy to identify "who" needs improvement from the results of data analysis, but it is more difficult to get suggestions on "how" improvements should be made. In practice, decisions on measures to take are made based on the store manager's knowledge and expertise.

To address this issue, we proposed an approach in which, for a given targeted segment (intervention group), we identify another group with similar purchasing and behavior tendencies but better characteristics such as more frequent store visits or higher-priced purchases (the preferred group). We then identify measures that will move the intervention group toward the preferred group. For example, suppose, as shown in **Fig. 4**, that the purchase amount for the flow going directly to the beverages then returns to the entrance (red) is 120 yen, but the amount for the flow going past the snacks in the center (blue) is 150 yen. In this case, if customers on the red flow are guided to the next aisle, we can assume that their purchase amount could also increase. The expected effect of this measure can also be calculated as (no. of intervention group purchases) × (rate of change in the intervention group) × 30 yen (the difference in purchase amount between the intervention and preferred groups).

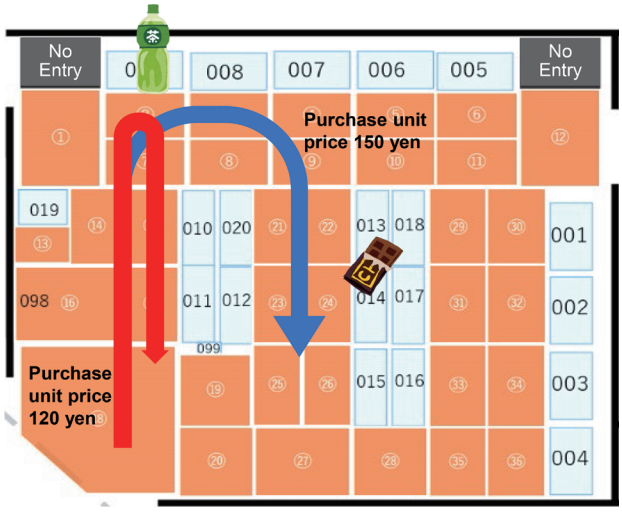


Fig. 4. Example of intervention group (red) and preferred group (blue).

5.1 Data required for this approach

The technologies needed to implement this approach are (1) automatic customer segmentation that selects an appropriate intervention group and preferred group from among the wide range of data

Measure candidate list

Rank	Expected result	Measure example	Main classification attribute	Secondary classification attribute
1	250,000	Product-set sales	Purchased together	Timeslot
2	100,000	Product-set sales	Purchased together	Day of the week
3	80,000	Signage advertising	Flow pattern	Purchased together
4	50,000	Product-set sales	Purchased together	Age group
5	40,000	Signage advertising	Flow pattern	Product category
	...			

List of segments for intervention

Rank	Expected result	Intervention group	Preferred group	Measure example
1	100,000	Single-beverage, noon	Food and beverage, noon	Onigiri and tea, set purchase
2	80,000	Single sweets, evening	Sweets and coffee, evening	Chocolate and coffee, set purchase
3	70,000	Food and beverage, noon	Meal, beverage and sweets, noon	Meal, desert, set purchase

Fig. 5. Measure candidate list and list of segments for intervention.

available in a smart store, and (2) a technology to compute the expected effect of the measure, which computes the effect that can be expected from the selected intervention group considering characteristics of the preferred group. An outline of these technologies is given below.

(1) Automatic customer segmentation: The objective with this technology is to select an intervention group and preferred group that have similar purchasing and behavior tendencies for which a statistically significant and sufficient effect of a measure can be expected. Even if customers and purchases are each partitioned randomly and differences in store-visit frequency or purchase amounts are compared, we would not know whether the differences arose by chance or represent a stable tendency. We used a method for testing the significance of differences between various distributions and computing the degree of difference in distributions between groups to select combinations of segments for which measures can be expected to have a sufficient effect.

(2) Expected-effect computation technology: This technology computes the effect that can be expected for a measure when applying a concrete, rule-based measure to the intervention group selected in (1), influencing it toward the preferred group behavior. A feature of this technology is that it defines rules for the attributes needed for various sales-promotion measures then searches for meaningful candidate measures by narrowing down the combinations that include those attributes.

5.2 Procedure for this approach

We explain the procedure for this approach using the example illustrated in **Fig. 5** in which possible measures are taken to increase the number of items purchased by customers and the expected results of those measures computed from a store's purchase and behavior history.

(1) STEP 1: Register rules for measures

Measures that can be taken in the store and rules for them are first defined. For example, in a measure to sell products in sets, information on which products can be combined is important, so we create a required attribute for "may be sold together." For a measure using signage, its location is important, so we create a required attribute for "flow pattern."

(2) STEP 2: Select segment combinations that are significant and effective

For the measures and rules registered in STEP 1, execute the following.

1) Generate sets of attributes including those defined as required.
2) For the sets of attributes from 1), use the segmentation technology described above to select intervention and preferred groups that are meaningful and for which a measure will be effective.

(3) STEP 3: Compute a specific draft measure and its expected effect

For the intervention groups obtained in STEP 2, compute the expected effect as described above and from the results, compute the overall expected effect from all the measures. In Fig. 5, implementing sale of sets of products focusing on the timeslot and the purchased together measures had the highest expected effect, showing that selling sets of products from three specific intervention groups is a possible measure

to take.

For various candidate measures, specific intervention segments and expected effects are displayed, enabling store managers to select measures with high expected effect. This is the value provided with this technology.

We plan to verify the usefulness of this technology in real stores, particularly the appropriateness of the measures suggested and its effect in reducing the amount of work to implement such measures. We will also conduct further studies to implement it practically.

6. Further issues and future prospects

We described an operation-optimization technology that uses purchase and behavior data obtained in a smart store to efficiently guide the introduction of measures for sales promotion and store improvements. However, looking at in-store behavior is not sufficient for increasing customer visits or purchases. For example, it is difficult to understand information motivating a visit or a purchase from in-store data alone due to factors such as a product becoming fashionable on social media or customers being lured away to a competing store. Thus, it will be necessary to analyze these data together with data from various other sources. We intend to combine in-store purchase and behavior data with external data, analyze these large-scale data efficiently, and study technology to estimate customer incentives for visiting and making purchases in a store.

Reference

[1] H. Sakoe and S. Chiba, "Dynamic Programming Algorithm Optimization for Spoken Word Recognition," IEEE TASSP, Vol. 26, No. 1, pp. 43–49, 1978.

Sunyeong Kim
Senior Research Engineer, NTT Smart Data Science Center, NTT Computer and Data Science Laboratories.
She received an M.S. in information science from the University of Tokyo and joined NTT in 2004. Her research interests include data science, applied statistics, especially marketing science and customer behavior analysis on offline shopping.

Keita Nishimoto
Former Research Engineer, NTT Smart Data Science Center, NTT Computer and Data Science Laboratories*.
He received an M.S. in information science from Nagoya University in 2014. His research interests include complex systems science and data analysis, especially customer-behavior analysis. He is a member of the Information Processing Society of Japan (IPSJ).
*He is currently a freelance researcher.

Masahide Mizushima
Senior Research Engineer, NTT Smart Data Science Center, NTT Computer and Data Science Laboratories.
He joined NTT in 1992. His research interests include software engineering and data science.

Setsuo Yamada
Research Engineer, NTT Smart Data Science Center, NTT Computer and Data Science Laboratories.
He received an M.S. in information science from Tokyo Denki University in 1992 and Ph.D. in engineering from Tottori University in 2006. He joined NTT in 1992. From 1997 to 2000, he was with ATR Spoken Language Translation Research Laboratories. He joined NTT in 2000. His research interests include natural language processing, artificial intelligence, software engineering, and data science. He is a member of IPSJ and the Association for Natural Language Processing.

Junji Tomita
Senior Research Engineer, Supervisor, NTT Smart Data Science Center, NTT Computer and Data Science Laboratories.
He received an M.S. in computer science from Keio University, Kanagawa and joined NTT in 1997. He was a visiting scholar at the University of Washington in 2005 and received a Ph.D. from Keio University in 2012. He is engaged in the practical application of AI technology such as renewable energy, construction, distribution, and marketing. He is a member of IPSJ and was a bord member of the Database Society of Japan from 2016 to 2020.

Robot Delivery Service with Mobile Ordering

Saori Ouji, Yoshifumi Fukumoto, and Nobuhiko Matsuura

Abstract

In urban office districts, there is a movement toward using indoor mobility solutions, such as using robots, to provide necessary services, such as cleaning, security, and delivery, for tenants and building owners to reduce operational costs. Anticipating the use of multiple robots for various services within buildings, efforts are underway to implement systems involving multiple heterogeneous robots. This article discusses efforts toward achieving robot delivery services via mobile ordering, enabling busy office workers to make retail purchases while in the office. It focuses on the development of optimal robot-control technology that accurately predicts robot delivery times and explores optimal travel routes to ensure timely delivery.

Keywords: AI, robot, digital twin

1. What are robot delivery services?

A robot delivery service is a service in which users can freely select the destination, delivery point, and time for their ordered items, which are then delivered by robots. Users place orders and make payments to the store using a dedicated mobile ordering service on their smartphones or similar devices. The order information is transmitted to the robots, which then collect the items from the store and deliver them to the specified locations according to the order information. After completing the delivery, the robots return to a designated location, where they wait for the next order while charging the battery as needed.

As shown in **Fig. 1**, if there are multiple orders at the same time, they will be consolidated and delivered in a single robot delivery. If a new order comes in after the robot has departed, another waiting robot will be used for the delivery.

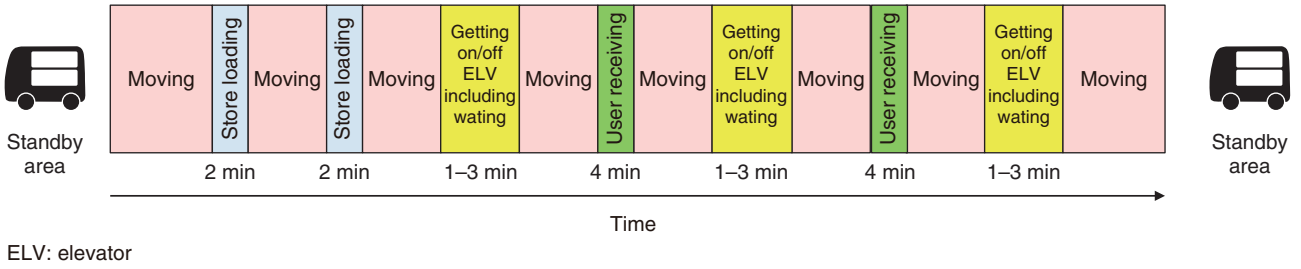
Introducing robot delivery services to office buildings provides a new value experience by delivering goods to busy office workers at their desired locations and times when they do not have time to go shopping. This contributes to creating new sales opportunities for understaffed stores by providing delivery services without the need for additional personnel.

To provide robot delivery services, it is necessary to eliminate the task of manually calculating and setting the robot's travel routes while considering the order details. Therefore, we have developed optimal robot-control technology, enabling automatic and efficient route calculation without human intervention.

2. Challenges with robot delivery services

To implement a robot delivery service, there are requirements to minimize the number of robots and travel time for successive orders and predict and provide the delivery time in advance so that users can efficiently receive the products. Three main factors affect travel time.

1) Robots attempt to avoid collisions by detecting obstacles during travel and slowing down or temporarily stopping if an approach is detected. This avoidance action increases the robot's travel time, so it is necessary to consider congestion on the route and the use of building facilities (such as elevators).
2) It is necessary to consider the operating hours



ELV: elevator

Fig. 1. Robot-delivery-service flow.

of the store. Store staff gather the targeted items while checking the order details, and when the robot arrives at the store, they load the items onto it. Additional tasks, such as warming up certain items, may also occur. The time required for this series of tasks is not always constant and varies greatly depending on the characteristics of the ordered items and congestion of the store, so consideration is necessary.

3) It is necessary to consider competition between robots. For example, if two robots are running side by side, they may perceive each other as moving obstacles and repeatedly decelerate or pause to avoid collision. When two robots pass each other in a corridor, there is a risk of them getting stuck as they fail to avoid each other properly, potentially resulting in the halt of delivery. The simultaneous movement or passing of robots could disrupt pedestrians, making it crucial to avoid such situations altogether.

3. Optimal robot-control technology

Our optimal robot-control technology is a solution to the above-mentioned challenges. It involves graphing the routes of spaces targeted for robot delivery services, such as city blocks and buildings, and searches for the optimal route by predicting the travel time for each route and point as well as the competition between multiple robots. This technology encompasses four other technologies: 1) route-time prediction, 2) store-loading-time prediction, 3) optimal-route search on the graph, and 4) priority control for resolving robot competition. In this context, nodes in the graph represent waiting points for robots, loading points for stores, reception points for users, and branching points for routes (including waiting points for elevators), while edges represent passages between nodes.

3.1 Route-time-prediction technology

Figure 2 gives an overview of the route-time-prediction technology. It involves collecting data on pedestrian flow and elevator-operation status to quantify the congestion levels at each point. Using machine learning, future congestion levels are predicted on the basis of past performance data. By predicting the travel time for each route on the basis of the forecasted congestion levels, the travel time for each route (edge) at the time of delivery initiation can be obtained. The obtained travel times are used as costs for the corresponding edges in the graph structure.

The travel time for each route varies significantly depending on the congestion levels along the route. This is due to the presence of many people on the route (moving randomly and requiring the robot to avoid collisions), resulting in repeated abrupt stops by the robot and requiring more irregular travel time compared with uninterrupted travel. This phenomenon applies similarly to optimal route exploration on the graph, where route selection based on congestion levels is essential to achieve optimal routes and accurate estimates of travel time. This technology enables accurate planning of operations by estimating delivery times considering environmental and time changes, minimizing the deviation between actual and estimated values and enabling the use of accurate travel times for operational planning.

3.2 Store-loading-time-prediction technology

Figure 3 gives an overview of the store-loading-time-prediction technology. The time required for the series of tasks from preparing the items to loading them onto the robot varies significantly depending on the characteristics of the ordered items and congestion levels in the store. Therefore, the ordered items are classified on the basis of characteristics that affect loading, such as whether they belong to categories

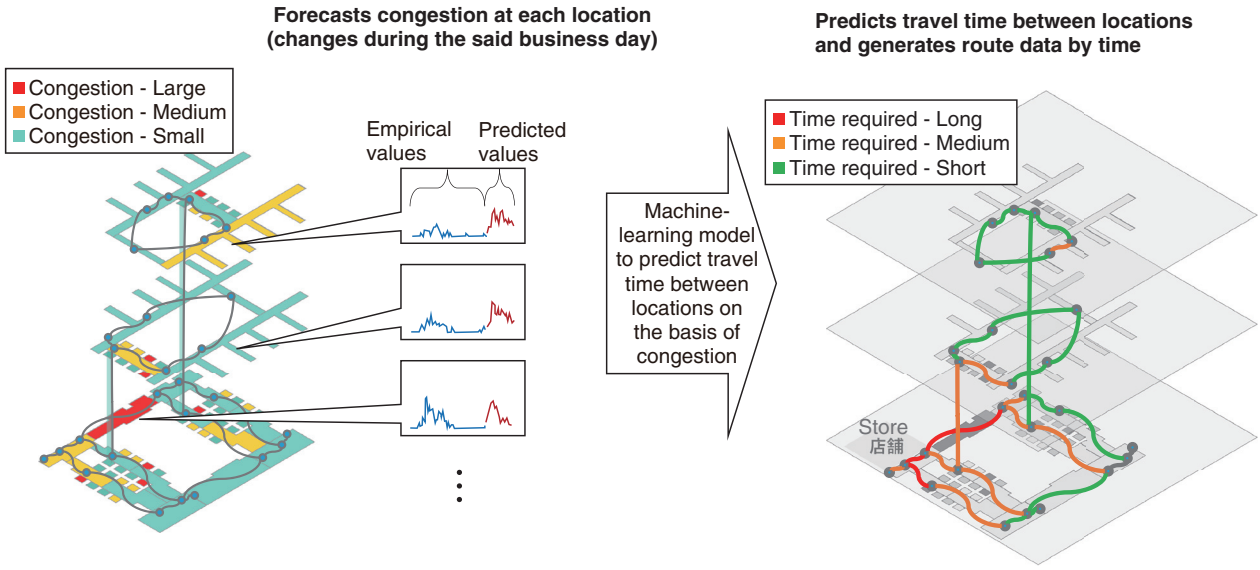


Fig. 2. Route-time-prediction technology.

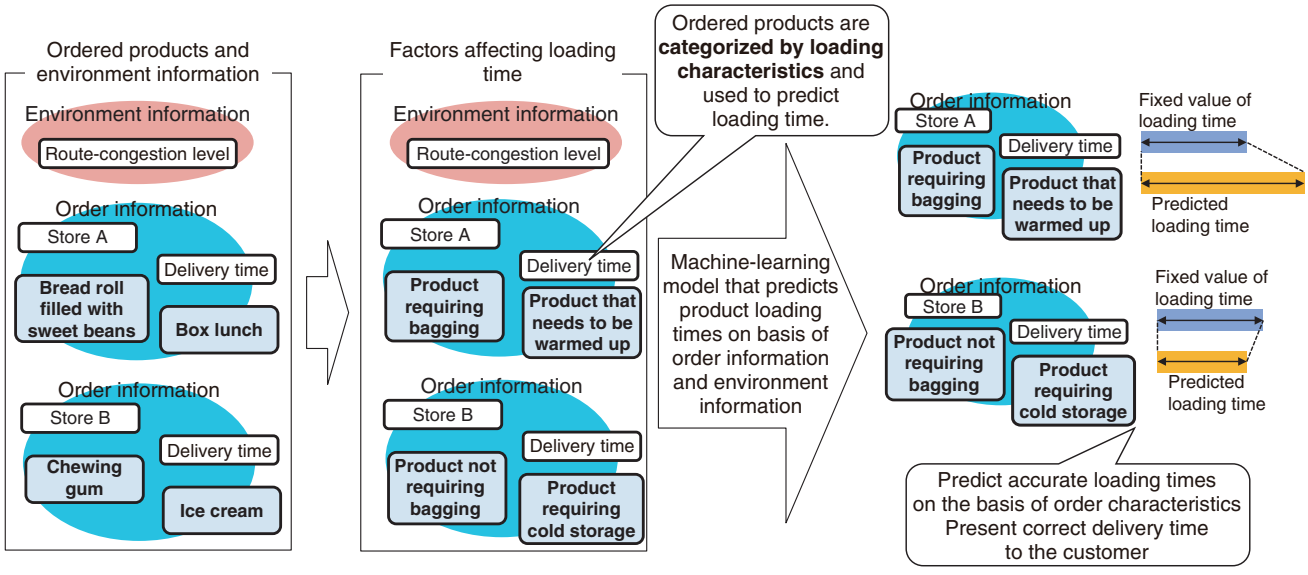


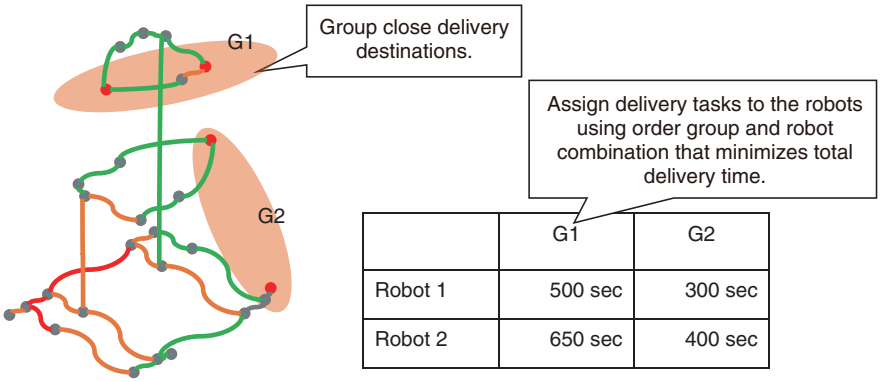
Fig. 3. Store-loading-time-prediction technology.

that require warming or need to be fixed in a drink holder, using basic information about the ordered items (name, description, etc.). On the basis of these characteristics and the congestion levels in the store, the loading time is then predicted. The predicted loading time is used as costs for nodes corresponding to each store in the graph structure.

This technology enables the calculation of loading times on the basis of the characteristics of the ordered items, thereby minimizing the deviation between actual and estimated values.

3.3 Technology for optimal route search on the graph

The technology for optimal route search on the graph calculates the route with the minimum travel



	G1	G2
Robot 1	500 sec	300 sec
Robot 2	650 sec	400 sec

Fig. 4. Grouping destinations for multiple orders.

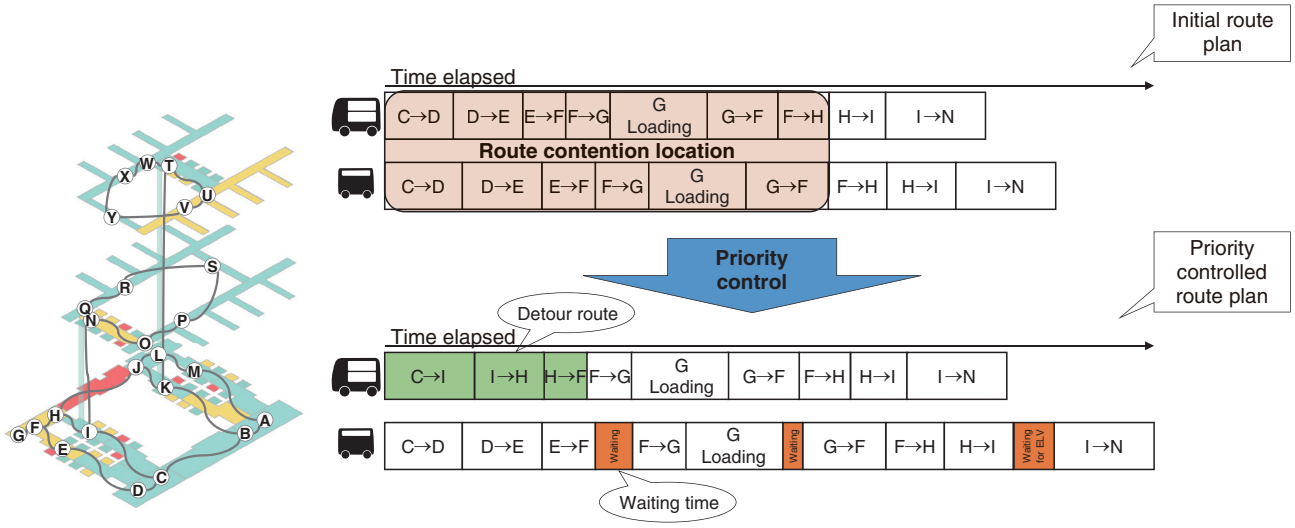


Fig. 5. Priority-control technology for resolving robot competition.

time using the graph with the cost calculated with the above two technologies. A graph is predicted and created at the time of delivery on the basis of order information, and the shortest route between each point (waiting points, loading points, reception points, waiting points) to be passed during delivery is calculated from the graph. If there are multiple orders simultaneously, orders with nearby destinations are grouped sequentially, and the shortest routes and travel times are calculated for each robot assignment, optimizing the allocation to minimize the total time. **Figure 4** outlines this process.

This technology takes into account environmental and time changes, selects the optimal route among multiple routes, and devises an efficient operational plan.

3.4 Priority-control technology for resolving robot competition

Figure 5 gives an overview of the priority-control technology for resolving robot competition. To operate multiple robots within the same area, it is necessary to avoid overlap and competition of the planned travel routes of each robot. For example, in the figure, there are multiple overlapping points as two robots move from points C to N. To avoid route conflicts, diversion routes or waiting times are set to ensure that each robot does not pass through the same route at the same time. This resolves route competition by avoiding collisions and being stuck between robots with

the minimum additional time. The same procedure is followed to resolve competition when there are three or more robots.

With this technology, it is possible to mitigate the delay factors in operational planning, such as route overlap and competition, and the resulting travel time delays, which cannot be suppressed by optimal route search on the graph, when multiple robots are operated simultaneously.

4. Evaluation of optimal robot-control technology

We evaluated robot delivery services using our optimal robot-control technology through a demonstration experiment targeting an office building [1]. Among the four elements comprising our optimal robot-control technology, we evaluated the route-time-prediction technology and the technology for optimal route search on the graph.

We developed a system for the experiment in which workers in the building can order items from the building's restaurants using a dedicated mobile ordering service. The robots were capable of traversing floors by collaborating with elevators and autonomously travel from the loading point in front of the restaurant to the reception point on the floor where the user worked. If the robot arrived at the reception point and nobody receives the items within 10 minutes, it was determined that the user was unavailable to receive them, and the robot returned to the restaurant. The items returned to the restaurant were held there, and the user would then need to retrieve them directly from the restaurant.

We provided two patterns for order placement: scheduled delivery and immediate delivery. Scheduled delivery allows users to select a desired time slot from 30-minute intervals throughout the day and place their orders accordingly. Immediate delivery automatically selects the earliest available time slot for delivery and places the order. After receiving orders in both of the above formats, the robot delivery service uses our optimal robot-control technology to execute optimal route searches. The robots are then scheduled to follow these routes accordingly. The estimated delivery time is also calculated on the basis of the predicted travel time and presented to the user in advance.

We used thermopile-type sensors to acquire congestion information. These sensors contain non-contact temperature sensors capable of detecting and measuring human body-heat radiation. Installed on the ceiling, the thermopile-type sensors measure the number of people within their detection range directly below them, which we used as a measure of congestion at that location.

The demonstration was conducted from October 2022 to March 2023. The period was divided into two halves: the first half focused mainly on collecting basic data, confirming operational flows of the restaurants, and conducting interviews with users. During the second half, evaluations were conducted while actual users placed orders.

During the evaluation, we compared the predicted arrival time presented to the users with the actual arrival time. Three comparison patterns were prepared as follows.

- Pattern 1: A comparison between the arrival time calculated on the basis of the travel time designated during the initial setup by the robot vendor and the actual arrival time. This pattern served as a reference for evaluation, indicating the performance without the application of our technology.
- Pattern 2: A comparison between the predicted arrival time using a model tuned simply to minimize errors (error-suppression model) and the actual arrival time.
- Pattern 3: A comparison between the predicted arrival time using a model tuned to ensure that the predicted value does not become smaller than the actual value (downward-swing-suppression model) and the actual arrival time.

We constructed a downward-swing-suppression model of Pattern 3 in response to the user's viewpoint that there is no problem if the robot arrives before the scheduled time because the user plans to pick up the product at the indicated arrival time. Pattern 3 is the model in which the robot is more likely to have arrived by the expected arrival time for the user to be able to receive the product, although the error is larger than in the error-suppression model of Pattern 2.

The results are listed in **Table 1**.

Comparing Patterns 1 and 2, the error (mean absolute error (MAE)) was reduced from 215 to 25.8 seconds. The number of downward-swing orders that arrived later than scheduled, causing users to wait, decreased from 190 to 99, demonstrating a substantial improvement by using our optimal robot-control technology.

A comparison of Patterns 2 and 3 revealed that the number of downward-swing orders can be reduced to about one-fourth by using the downward-swing-suppression model, although the error is larger. It is

Table 1. Results of comparison between predicted and measured values for each pattern.

Comparison of predicted and measured values	Pattern 1	Pattern 2	Pattern 3
	Robot setting	Error suppression	Downward-swing suppression
MAE	215.0	25.8	49
RMSE	226.3	43.1	58.6
Number of downward-swing orders (number of orders that arrive later than scheduled)	190	99	25
Number of downward-swing orders of 1 min or more	177	19	4

RMSE: root mean squared error

important to choose between the error-suppression and downward-swing-suppression models depending on the situation. However, from the user's viewpoint, the downward-swing-suppression model is often considered more effective with less waiting time, and from the operator's viewpoint, the error-suppression model can plan robot operations more efficiently.

5. Future outlook

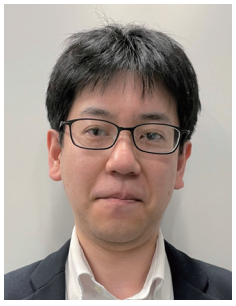
By using our optimal robot-control technology, it is possible to accurately predict travel times and conduct optimal route searches. Moving forward, we will conduct evaluations and improvements through demonstration experiments for the store-loading-time-prediction technology and the priority-control technology for resolving robot competition. This will further optimize our optimal robot-control technology.

By controlling not only delivery robots but also various other robots, such as those for cleaning and security, on an integrated platform, optimization and labor saving can be achieved for an entire city block.

Reference

[1] Press release issued by NTT Urban Solutions, NTT Urban Development, NTT Communications, and NTT Corporation, "NTT Group Companies Begin Demonstration Experiment of Autonomous and Efficient Food Delivery Using Robots Creating Valuable New Experiences for Office Workers & Sales Opportunities for Restaurants," Oct. 3, 2022. https://group.ntt/en/newsrelease/2022/10/03/221003a.html

Saori Ouji
Researcher, NTT Smart Data Science Center.
She received an M.E. in engineering from Tokyo Institute of Technology in 2018 and joined NTT Software Innovation Center the same year. She has engaged in the research on software engineering. In 2021 she moved to NTT Smart Data Science Center and has been engaged in research and development of software technology related to smart cities.



Yoshifumi Fukumoto
Research Engineer, NTT Smart Data Science Center.
He joined NTT in 2009, where he engaged in research related to distributed systems. In 2016, he moved to NTT WEST, where he engaged in research related to low power wide area networks and a data exchange platform. He returned to NTT in 2019 and has been engaged in the research and development of software technology related to digital transformation and smart cities.

Nobuhiko Matsuura
Senior Research Engineer, NTT Smart Data Science Center.
He received an M.S. in computer science from Shizuoka University in 2012 and joined NTT the same year, where he engaged in research on distributed systems. In 2020, he moved to NTT Urban Solutions, Inc., where he worked as a project manager for urban development using NTT Group technologies and solutions. He returned to NTT in 2023 and has been engaged in research and development of software technology related to smart cities.

AI Value Platform Accelerates Data Valorization by Consolidating SDSC's Elemental Technologies

Takahiro Hata, Midori Kodama, Miho Fujishima, Yu Adachi, Kenichi Fukuda, and Yuki Yokohata

Abstract

We at NTT Smart Data Science Center (SDSC) are developing an artificial intelligence (AI) value platform to accelerate data valorization by consolidating the results from various industry domains that we have been working on. By cataloging analytical case studies, we aim to support problem design and accelerate valorization by automatically experimenting with processing techniques for aggregated real data, thus shortening the initial trial-and-error phase of data analysis. In this article, we introduce our efforts toward implementing the AI value platform.

Keywords: AI value platform, data analysis, mathematical optimization

1. Introduction

At NTT Smart Data Science Center (SDSC), we have been addressing customer challenges in various areas such as lifestyle, consumption, mobility, logistics, and resources, particularly to enable urban development through Digital Twin Computing (DTC) (**Fig. 1**).

Looking back at these achievements, it is apparent that the source of value in the real world (physical space) is based on real data acquired under the constraints imposed by reality. We have come to understand that this can be explained by the mechanism of a cyber physical system, where optimal states are calculated in the digital twin (DT) in cyber space and the results are returned as controls or recommendations for the real world (**Fig. 2**). Optimization, playing the most crucial role here, has been studied extensively, even with complex and vast models. However, with advancements in artificial intelligence (AI) and machine-learning technologies, it has become feasible to solve optimization problems by learning from rich and well-structured data within reasonable computation time.

To achieve valorization in the real world, however, various constraints specific to each field such as measurement conditions, operational conditions, and costs must be met. Often, raw data as measured cannot be directly used without preprocessing.

In light of these circumstances, SDSC has been developing an AI value platform that consolidates core technologies and insights from various domains of urban DTC, which SDSC has been addressing, to accelerate data valorization to bridge the gap between the physical and cyber realms and return value to the real world on the basis of the data.

The AI value platform defines and embodies new value in actual industry domains through AI, making each value tangible as services for business operations and customer value within those domains. It encompasses common analysis functions such as real-data processing, prediction, and optimization, as well as functions required for actual data integration and interfaces that enable users to experience the value in their business operations or as individual customers.

In the AI value platform, core AI functionalities are integrated into parts that can be organically used

Fig. 1. Examples of domains we have worked on.

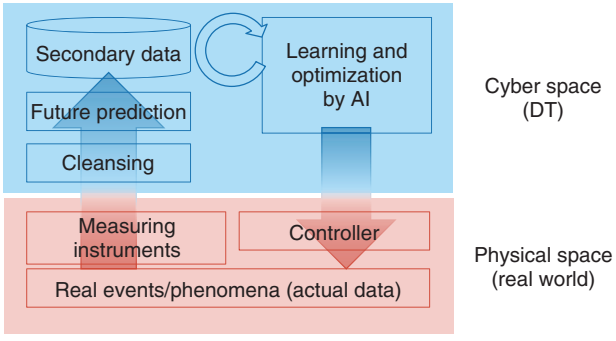


Fig. 2. Conceptual diagram of cyber physical system.

within services offered across multiple industry domains. By combining these parts in accordance with the characteristics of the obtained value, it becomes possible to accelerate valorization by facilitating the selection of necessary functionalities for new services, automating initial data processing, and minimizing trial and error.

We introduce two main innovations among various efforts to accelerate valorization within the functionalities of the AI value platform:

(a) aggregating and providing refined problem designs as references, derived from past achievements, to ensure operability in society
(b) extracting preprocessing tasks, such as data cleansing and future forecasting, which precede optimization, as common analysis functions of the AI value platform, making them applicable to other datasets.

2. How the AI value platform accelerates data valorization

We explain how the aforementioned functionalities (a) and (b) contribute to the valorization process.

The valorization process at SDSC mostly follows the typical steps of data analysis shown in **Fig. 3**. In the problem-design phase (1), customer problems and surrounding knowledge are decomposed, leading to the formulation of optimization problems. In the data-analysis phase (2), data are measured in a real environment, and optimization methods are implemented in the DT. If these steps are deemed effective, the process moves to the validation phase (3), where actual equipment is operated for field validation.

Functionalities (a) and (b) in particular accelerate the problem-design and data-analysis phases. This not only speeds up valorization but also helps customers realize the significance of the initiatives by accelerating steps (1) and (2) to demonstrate results promptly, especially in SDSC's efforts, which require the cooperation of customers and using real data. While the AI value platform is also being developed to be used as a DT, we omit discussion on this aspect as it pertains to different areas than those covered by (a) and (b).

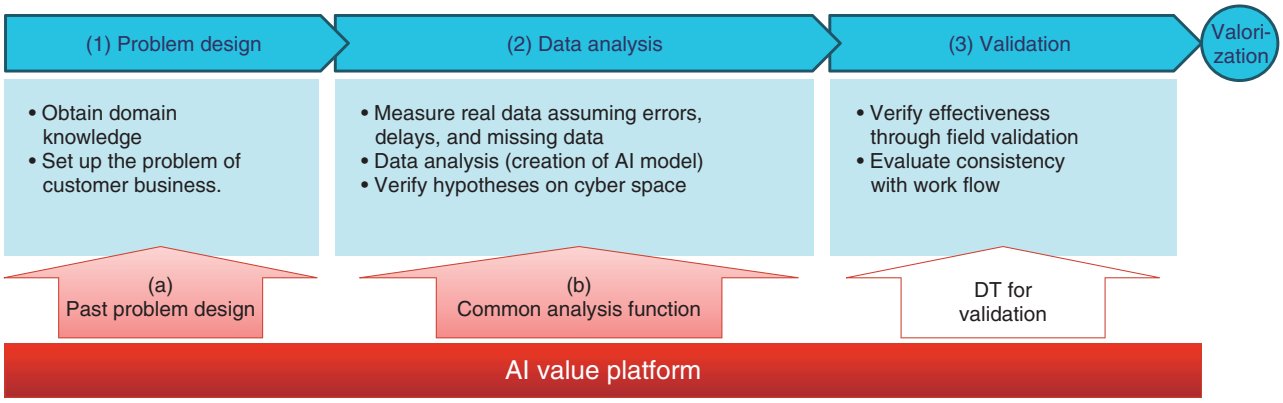


Fig. 3. Valorization flow and AI value platform's contribution.

(a) Provide examples of optimization problem design as references

In the real world, addressing challenges can often be translated into mathematical optimization problems that minimize (or maximize) specific metrics while meeting various real-world constraints. Example challenges that SDSC has addressed include:

- to reduce energy consumption in air conditioning, minimize the total energy consumption of air conditioning operations for the day while maintaining visitor comfort
- to deliver products to as many customers as possible on time, plan routes for multiple robots for product delivery to maximize the number of deliveries in the shortest time
- to respond promptly to emergency calls, minimize the time required for ambulances to arrive at the scene by changing the fire stations where ambulances wait.

These examples represent mathematical optimization problems with constraints. However, it is important to note that translating real-world problems into mathematical problems can be challenging and is often considered the most crucial and difficult process in valorization.

Ensuring the practicality of this problem design requires the ability to correctly understand and decompose the customer's problems and surrounding environment then effectively align and redesign methods for solving them within the allotted time. For this problem design to be feasible, all the following issues must be resolved:

- tolerating errors, delays, and missing data in measurements
- developing analytical logic to derive optimal solutions from measurement data
- controlling within an acceptable time frame from the analysis results to execution
- incorporating the execution procedure into operational workflows.

It is rare to anticipate and resolve all these issues completely in a single attempt. In reality, it often involves iterating through multiple problem designs, operationalizing them, and gathering various feedback to revise and refine the necessary areas. This iterative process is essential for refining problem designs and ensuring their practical applicability.

Experienced data scientists, who have accumulated a wealth of experience in data analysis, have encountered various problems that can arise in each stage of measurement, analysis, control, and operation in various situations. They are familiar with numerous cases of both success and failure, enabling them to reach realistic solutions in fewer cycles. To become such experienced data scientists, it is essential not only to gain personal experience but also to seek advice from experienced professionals and review past deliberations and materials to relive the trial and error of predecessors. This is the aim of accumulating optimization problem design cases as references.

The aim with the AI value platform is to accumulate problem designs that data scientists have solved in the past so that they can be referenced as guides when facing new problems. This provides a means for data analysts facing new challenges to relive past experiences and examples. Even if the domains are different, being able to reference the process of problem decomposition and reduction to mathematical optimization problems in one's valorization process for the problems they encounter, it is expected that they

will be able to quickly carry out practical problem designs.

(b) Make optimization preprocessing a common analysis function that can be applied to other data

Temperature, brightness, air-conditioning settings, and the number of people in the area are quantified through sensors and become data. Using these data, we can understand or predict real-world phenomena and optimize various situations by controlling equipment or recommending actions. However, various factors, such as sensor malfunctions, movements, obstructions, or sudden increases in the number of people due to events, can introduce errors, missing data, or anomalies in the data. These errors, gaps, and anomalies hinder our understanding of phenomena and decrease prediction accuracy. Therefore, before optimization, data cleansing is necessary to process the data into a suitable form for optimization.

The data that can be measured are past events. Therefore, even if equipment control or behavioral recommendations are based on past data, it may not be optimal when equipment or people move. For example, in building air conditioning, changes in set temperature and airflow may take several dozens of minutes to several hours to be reflected in the space as changes in temperature and humidity, due to the large area managed by the air-conditioning system. During this time, it is common for crowded periods to end, resulting in excessive air conditioning. To prevent this, predicting when the optimization results are reflected can also be considered as a form of preprocessing.

At SDSC, we sometimes use AI, which requires a large amount of data, as an optimization method. However, in the real world, it is often not possible to conduct parallel or accelerated experiments, making it difficult to collect large amounts of data. On the basis of the examples from SDSC, it has been found that in such cases, specific data-processing techniques unique to machine learning are used, including:

- various data-augmentation techniques to supplement small amounts of data
- rebalancing to address extreme frequency differences between positive and negative examples
- typification of phenomena when there is a shortage of data due to a multitude of rare events (clustering).

These processing techniques can also be considered preprocessing, as they involve generating secondary data on the basis of the data measured before optimization.

The preprocessing tasks depend not only on the problem design introduced in (a) but also on the characteristics of the measuring instrument, measurement environment, and the measured data (phenomenon). This means that even if the problem designs are different, the preprocessing can still be effective. Therefore, in the AI value platform, the decision was made to extract the preprocessing part as a common analysis function. By modularizing the common analysis function in the same format, it is possible to try various methods simultaneously when analyzing new data and compare the results, enabling the quick selection of suitable preprocessing methods. Therefore, (a) and (b) support and accelerate the valorization process at SDSC.

3. Difference from conventional technology

In today's digital age where various real-world phenomena can be digitized, data analysis has become a challenge tackled by many individuals and companies. For this reason, numerous analysis technologies, libraries, and platforms that aggregate them are being provided. We describe the differences between common data-analysis platforms, various libraries, and the AI value platform.

We first explain the characteristics of representative analysis environments, such as R [1] and Python [2].

The R language is a language and software environment known for statistical analysis, developed with reference to the S language environment originally created at AT&T. It provides an interactive interpreter environment and is capable of processing large amounts of data rapidly. It also includes implementations of numerous statistical methods, enabling users to begin data analysis immediately upon installation. Additionally, advanced processes not implemented by default can be achieved by importing external packages.

Python is a general-purpose programming language with procedural aspects. While the programming language aspect of Python is not our main focus, like R, it includes many statistical processing, vector, and matrix calculation functions in its standard library. Despite being an interpreter, Python is known for its fast numerical computation library. Being a general-purpose programming language, it is well-suited for implementing complex processes. It also has robust support for graphics processing units

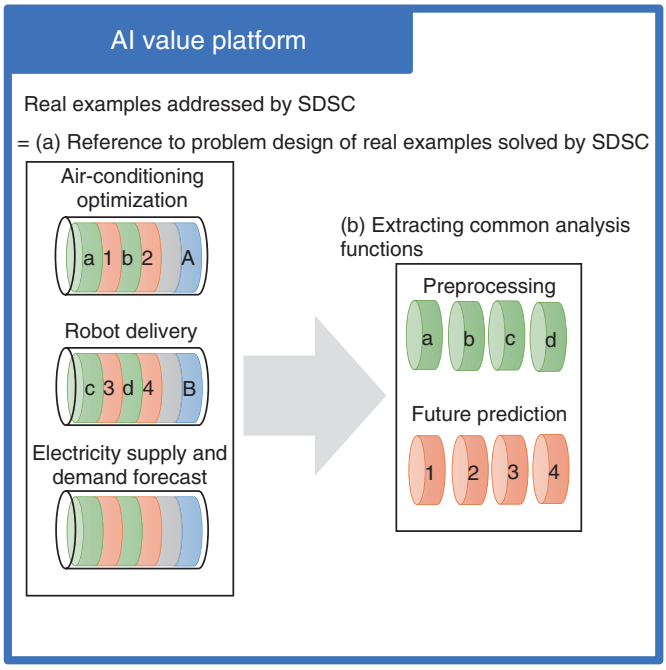


Fig. 4. Image of AI value platform implementation.

(GPUs), making it highly usable for advanced machine-learning tasks such as deep learning using neural networks. Python is widely chosen as an option for data-analysis implementation due to its capability to use cutting-edge algorithms by importing external libraries.

In addition to the two environments mentioned as examples, there are numerous programming languages and analysis platforms that aggregate algorithms implemented in these languages. In a broader sense, business intelligence tools for visualizing data from various aspects and relational databases with data accumulation and aggregation functions can also be considered as part of the analysis environment.

The common feature worth highlighting among these environments is their provision of versatile analytical functions that do not limit data formats or domains. This is, of course, expected in the context of providing analytical solutions. However, when seeking to solve real-world challenges using such versatile analytical environments, data analysts with expertise in problem design and preprocessing must spend time and effort selecting the appropriate method from a multitude of analytical techniques. Bridging the gap between cyber and physical spaces and optimizing them in a practical way still requires problem design and preprocessing, necessary in the valorization process, and which remain significant areas involving considerable time and effort through trial and error.

To address this, in the AI value platform, preprocessing techniques proven to solve real-world challenges are modularized to accelerate this trial-and-error process. Using this approach aims to provide support for these processes, which are difficult to support in general analytical environments, thus significantly differing from conventional technologies.

4. Design policy for the AI value platform

The following is a design policy (**Fig. 4**) for the AI value platform based on the ideas introduced thus far.

(a) Providing references to past problem design cases

As we have mentioned, to reduce a real-world problem to a mathematical optimization problem, it is necessary to have experience in solving many constraints and designing problems. The AI value platform supports analysts by allowing them to relive past problem designs, thus assisting in problem design through experience. In other words, the knowledge we want to provide with this functionality is how to resolve real-world constraints in a way that

does not distort the mathematical optimization problem. Specifically, within the AI value platform, past examples are analyzed as mathematical optimization problems, where

- goals → key performance indicators (KPIs) set for those goals → formulation as optimization problems
- measurement data → preprocessing executed
- optimization problems → optimization algorithms used.

We aim to list these real-world constraints and adopted methods as sets, enabling reference and search functionalities.

In this process, we consider successful cases to be more important than failure cases. While learning from failures is valuable, when it comes to accelerating the value realization with the AI value platform, the best approach is often to emulate successful cases. Although SDSC's initiatives appear to solve problems across various domains, fundamentally, the real world is driven by human social activities. This commonality forms the characteristics of measurement data and the operational conditions that solution methods must meet. Therefore, we believe that leveraging past successful cases is possible across different domains, as they share this common foundation.

(b) Standardization of preprocessing as common analysis function

The aforementioned commonalities also apply to preprocessing, which aims to eliminate constraints in the physical space.

SDSC's approach to data valorization involves handling real-world phenomena as time-series numerical data, capturing them in human social cycles such as time, day, day of the week, and week. Using periodicity and similarity, techniques like smoothing, interpolation, and data augmentation are implemented. The methods used to measure physical phenomena, such as temperature, brightness, power consumption, and population, are similar regardless of the domain.

Preprocessing methods leveraging these commonalities can be extracted by decomposing the cases SDSC has solved. The AI value platform was developed to standardize effective preprocessing in the same format, enabling uniform application to new data and accelerating the trial-and-error process of preprocessing.

5. Commercialization initiative

As mentioned earlier, SDSC has been providing new value by carrying out data preprocessing, future prediction, optimization, visualization, and optimal control tailored to the specific challenges faced by customers. The AI value platform provides an environment like the one shown in **Fig. 5**, where users can analyze sample data or their own data to confirm their usefulness. For each service, a dedicated DT is provided, offering independent execution environments for each customer (tenant). If the results of data valorization or services meet the customer's satisfaction, they can choose to use the environment on the AI value platform as it is, integrate it with their own systems and facilities, or customize it in accordance with the connected systems and facilities.

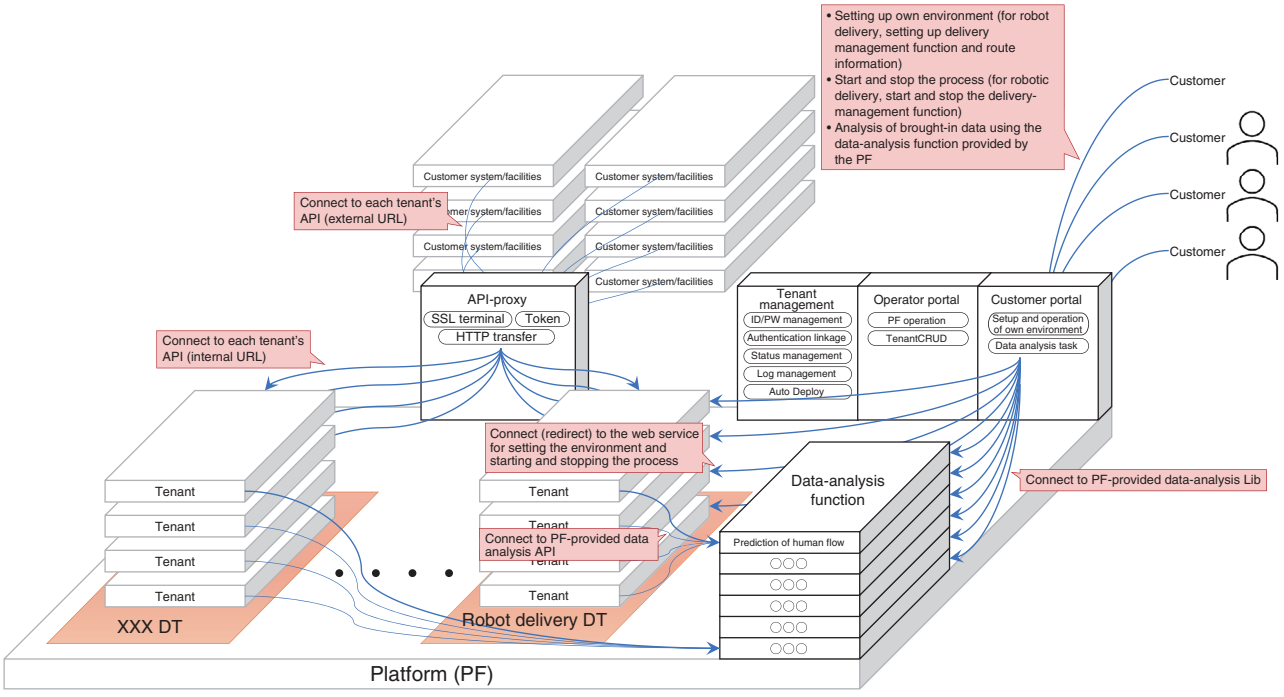
By using the aforementioned common analysis functions, it becomes possible to try and compare various data-analysis algorithms, enabling the selection of the optimal method.

Specific services are introduced below.

For instance, in efforts aimed at reducing food loss in stores, using customer data, we conduct future predictions such as forecasting the number of visitors and sales of each menu item. By visualizing these data, we can reduce excessive purchasing and preparation, thus reducing food loss (see **Fig. 6**).

When implementing this initiative at the customer's facility, quick results can be obtained using cameras, sensors to measure visitor numbers, and point-of-sales (POS) data (sales per menu item), as depicted in **Fig. 7**, which summarizes the necessary data and analysis functions for the service.

SDSC has been working on data-value-creation initiatives such as energy-efficient and comfortable air conditioning control, personal air conditioning in offices, achieving zero food loss in restaurants, optimizing store operations using customer-behavior analysis, efficient robot delivery to achieve effective distribution, optimal placement of emergency teams, high-precision power-generation forecasting and demand matching, detection and prediction of traffic congestion, nationwide optimization of agricultural product distribution, hospitality services in smart homes and urban areas based on predictive behavior, among others, as shown in Fig. 1. With the support of the AI value platform facilitating value creation, we will contribute to a future society where people's comfortable living standards are significantly improved.



API: application programming interface
HTTP: Hypertext Transfer Protocol
ID: identifier
Lib: library
PW: password
SSL: secure sockets layer
URL: uniform resource locator

Fig. 5. Grand design of AI value platform services.

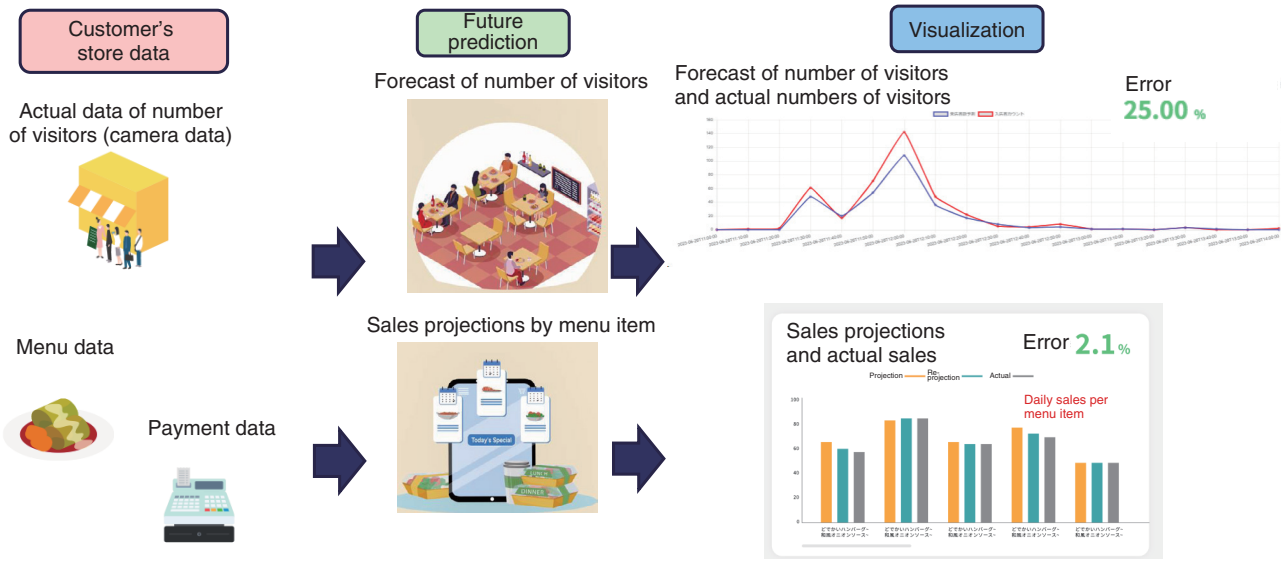
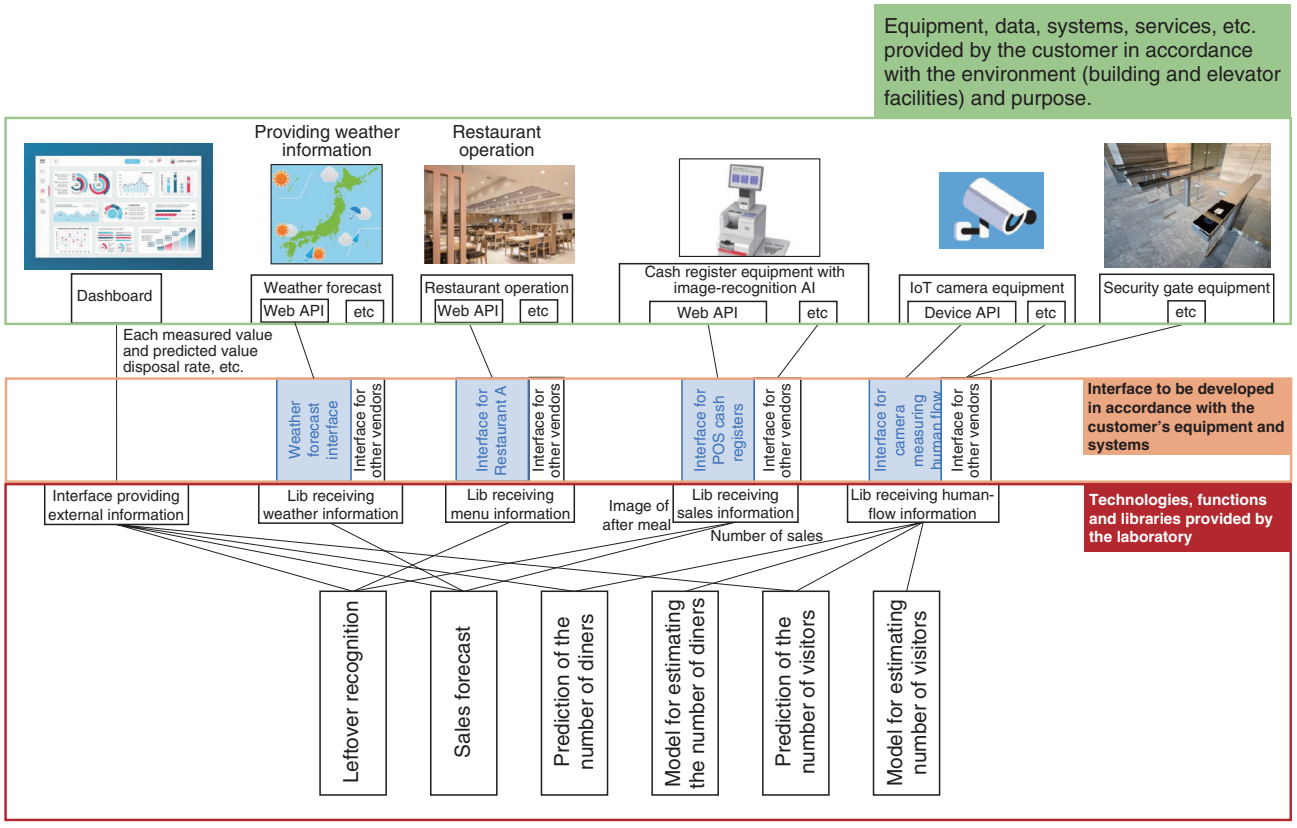


Fig. 6. Zero food loss initiatives: Forecasting the number of store visitors and number of sales per menu item.



IoT: Internet of Things

Fig. 7. Image of the system for implementing zero food loss policies.

6. Challenges and prospects for the AI value platform

We introduced our attempt to support new data valorization by consolidating the knowledge of the actual cases we have solved into the AI value platform. However, there are several challenges in aggregating insights into this platform.

One challenge is that the functionalities used in actual cases are implemented in the most efficient way within those cases. To extract them as common functionalities, each functionality needs to be redesigned to be applicable in other domains, assuming the same execution capability.

Another challenge is that as the common analysis functions increase, there is a need to select the appropriate analysis functionalities for application. For this purpose, metadata resembling semantics capable of understanding the characteristics and meanings of the data is necessary. These metadata may need to be reviewed whenever a new domain is added.

In fiscal year 2023, we completed the minimum implementation of the AI value platform. We will improve upon it on the basis of the feedback from actual users to address these challenges.

References

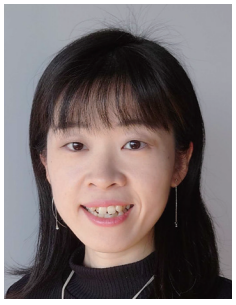
[1] Website of R, https://www.r-project.org
[2] Website of Python, https://www.python.org

Takahiro Hata
Senior Research Engineer, NTT Smart Data Science Center, NTT Computer and Data Science Laboratories.

He received an M.Inf. from Kyoto University in 2006 and joined NTT the same year. He studied ubiquitous services at NTT Network Innovation Laboratories. From 2010 to 2013, he was involved in the development of public wireless local area network services at NTT WEST. Since 2013, he has been active in the development of smart-city applications and analysis of pedestrian flow and driving data of vehicles at NTT Smart Data Science Center.

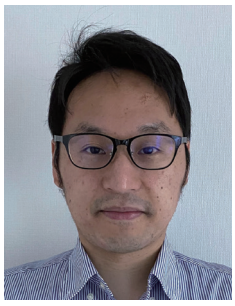
Midori Kodama
Research Engineer, NTT Smart Data Science Center, NTT Computer and Data Science Laboratories.

She received an M.S. in biosciences and informatics from Keio University, Kanagawa, in 2015 and joined NTT the same year. She studied ubiquitous services and analysis of educational data at NTT Network Innovation Laboratories. Since 2017, she has been active in the development of smart-city applications and data analysis related to air-conditioning systems at NTT Smart Data Science Center.



Miho Fujishima
Researcher, NTT Smart Data Science Center, NTT Computer and Data Science Laboratories.

She received an M.S. in earth and planetary science from the University of Tokyo in 2017 and joined NTT WEST the same year. From 2018 to 2021, she was involved in the development of home routers and technology in the Internet-of-Things and AI fields. In 2022, she moved to NTT laboratories and studied multi-agent mobility simulation as part of DTC. Since October of 2023, she has been active in the development of smart-city applications and analysis of data at NTT Smart Data Science Center.



Yu Adachi
Senior Research Engineer, NTT Smart Data Science Center, NTT Computer and Data Science Laboratories.

He received an M.E. from the University of Electro-Communications, Tokyo, in 2009 and joined NTT the same year, where he studied distributed data management. From 2012 to 2014, he was involved in service development at NTT WEST. From 2015 to 2020, he was involved in software engineering at NTT Software Innovation Center. Since 2021, he has been active in the development of smart-city applications and analysis of data at NTT Smart Data Science Center.

Kenichi Fukuda
Researcher, NTT Smart Data Science Center, NTT Computer and Data Science Laboratories.

He received an M.E. in control engineering from Osaka University in 2018 and joined NTT the same year. He studied optimization of ambulance allocation at NTT Network Innovation Laboratories. He has been active in the development of smart-city applications and data analysis at NTT Smart Data Science Center.

Yuki Yokohata
Senior Research Engineer, Supervisor, NTT Smart Data Science Center, NTT Computer and Data Science Laboratories.

She received an M.S. in earth and planetary science from Nagoya University, Aichi, in 2004 and joined NTT the same year. She studied ubiquitous services at NTT Network Service Systems Laboratories. From 2008 to 2015, she was involved in the development of public wireless local area network services at NTT Communications. Since 2016, she has been active in the development of smart-city applications and analysis of data at NTT Smart Data Science Center.

Improving Efficiency of Agricultural-product Distribution by Using a Virtual Market

Toshihiro Baba, Takafumi Hikichi, Shiori Toyoshima, Yasuhiro Hirano, Yasushi Hanakago, Akira Isoda, and Junji Tomita

Abstract

To improve the efficiency of agricultural-product distribution, NTT is creating a virtual market that can optimize commercial distribution and logistics on a nationwide scale by consolidating information on the supply and demand of agricultural products. This virtual market will help solve the so-called "2024 problem" facing the logistics industry due to the enforcement of new overtime regulations and ease labor shortages in fields related to the distribution of agricultural products, improve the economic effects of distribution by reducing time and costs, and reduce the environmental burden of distribution by reducing greenhouse-gas emissions and food loss.

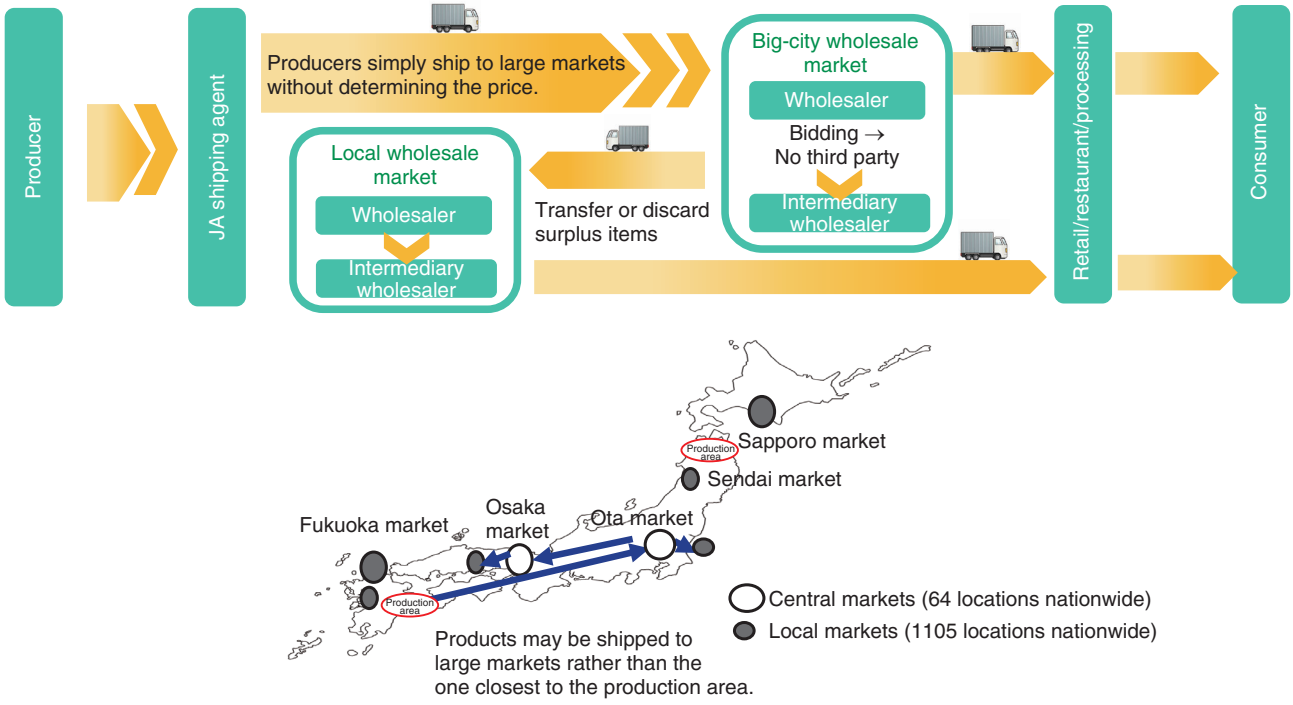
Keywords: agricultural-product distribution, virtual market, 2024 problem

1. Current status of agricultural-product distribution

The current status of agricultural-product distribution through markets in Japan is shown in **Fig. 1**. About 80% of fruits and vegetables are currently traded through wholesale markets [1]. Most of that produce is shipped to markets in large cities such as Tokyo and Osaka. However, from the perspective of market principles, if an excess of agricultural products occurs, prices will fall. The problem is that the shipments concentrate in the big-city markets; thus, local markets are unable to gather products. The resulting surpluses in the big-city markets are transferred to the local markets, but some products are discarded because they are no longer fresh. Producers are unable to determine the demand at each market, and that inability results in wasted delivery due to shipping to distant markets even though demand is stronger at nearby markets.

In fact, vegetables produced in Kyushu are shipped to Tokyo, while the same vegetables produced in the northern Kanto region (nearer to Tokyo than Kyushu) are shipped to Osaka (nearer to Kyushu). Since retailers do not know the shipment volume that will arrive at the market, they are unable to decide what amounts of goods to buy until just before they arrive; as a result, the retailers may have to change their processing and sales plans in accordance with the incoming volume that arrive at the market.

In addition to the above issues, the digitalization of operations has not progressed in various areas related to the agricultural-product distribution, which makes it difficult to comprehend the total volume of supply and demand, delivery routes, and other aspects of the overall distribution process. In these situations, the "2024 problem" is also having an impact on the distribution of agricultural products. The "2024 logistics problem" (2024 issue) is the collective term for issues arising from the imposition of limits on the working hours of drivers in the logistics sector. As the daily mileage per truck driver decreases and it becomes



JA: Japan Agricultural Co-operative

Fig. 1. Current status of distribution of agricultural products via markets.

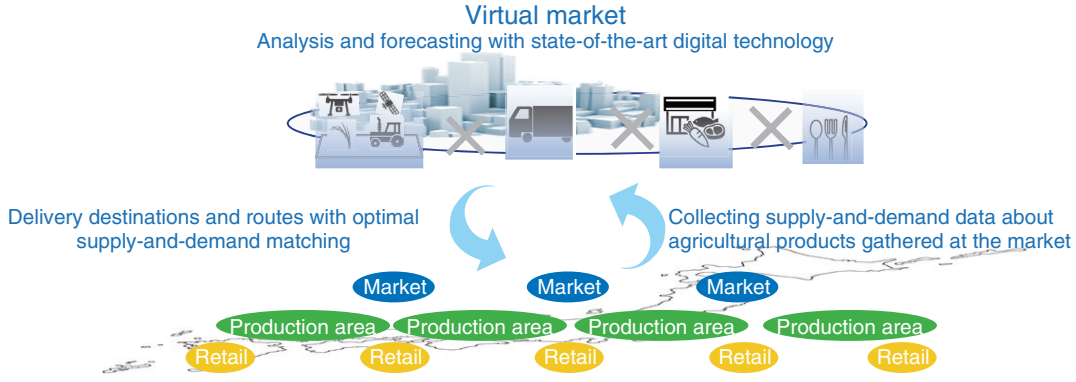


Fig. 2. Illustration of the virtual market.

more difficult to transport products over long distances, distribution over long distances will become more difficult in some production areas. The increasing cost of transportation in production areas near large cities will also require more-efficient transportation in terms of shorter delivery distances and times.

2. The future we want: the virtual market

To solve the above problems concerning agricultural-product distribution, we are building a virtual market, as shown in **Fig. 2**. This virtual market uses Digital Twin Computing (DTC)* to (i) predict future supply and demand from nationwide data on agricultural-product distribution and (ii) complete buying

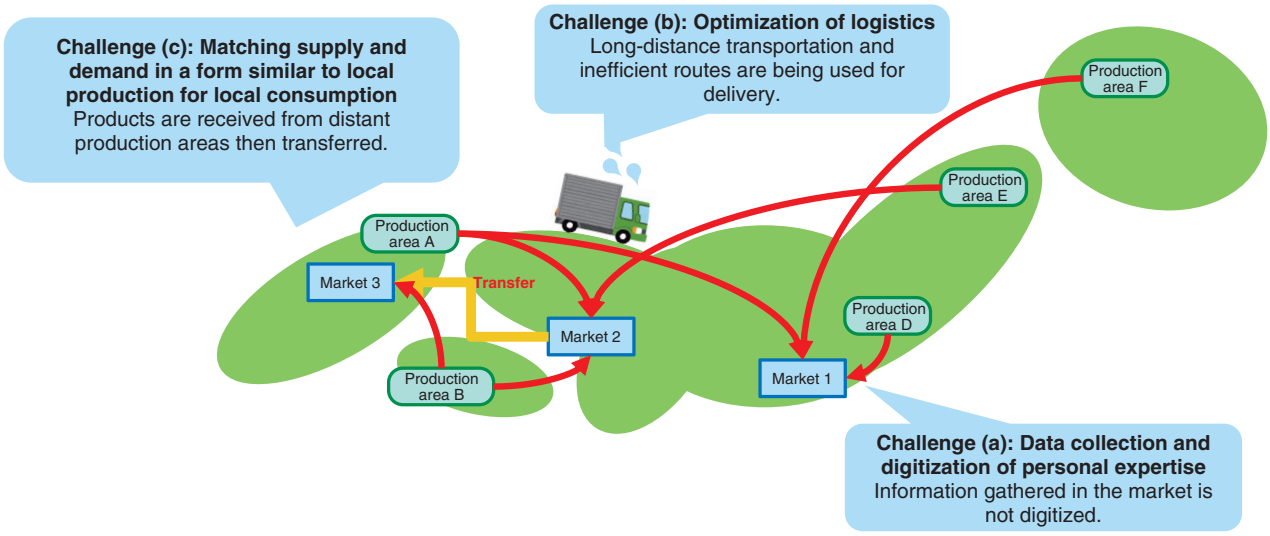


Fig. 3. Challenges to be addressed.

and selling in a virtual space before agricultural products are transported to the market.

In the virtual market, the commercial flow, including operations such as supply-and-demand matching, is completed while logistics delivers products via optimal routes on the basis of the results of matchings in the virtual market. By considering the most-suitable place to send agricultural products before shipping them from the production area, the virtual market makes it possible to (i) prevent transfers of products due to imbalances in supply and demand between markets and (ii) consolidate agricultural products from nearby production areas and other items in a manner that will enable delivery with no waste. The following list contains example values that the virtual market can provide to various stakeholders.

(1) Producers: By adjusting the supply-and-demand balance through forecasting, it becomes possible to avoid price decreases due to oversupply. By shipping products to closer markets, shipping costs and wasted agricultural produce can be reduced and income will thus increase by providing fresh, high-value vegetables.
(2) Wholesale market: By adjusting the supply-and-demand balance through forecasting, costs, such as transferring unsold items and supplying from neighboring markets due to product shortages, can be reduced. Digitizing analog transactions will also help mitigate labor shortages.
(3) Logistics providers: The delivery distance can be reduced by optimizing the route. Forecasting of shipping volume enables more-accurate delivery planning and efficient arrangement of personnel, trucks, etc.
(4) Intermediary wholesalers and retailers: By being able to predict the volume of incoming stocks, it will be possible to arrange in advance for the materials and personnel necessary for processing agricultural products.
(5) Consumers: The time required for distribution is shortened, and consumers will have access to fresher produce than ever before.

3. Challenges and approaches concerning the virtual market

Implementation of the virtual market faces three major challenges, each of which—along with our strategies—is introduced below and illustrated in **Fig. 3**.

3.1 Challenge (a): Data collection and digitization of expertise

Many organizations involved in the distribution of agricultural products have not yet digitized their operations. Among the stakeholders involved in the distribution of agricultural products, such as producers,

* DTC: A technology that represents various digital copies of real space in cyberspace and uses those copies for simulations such as data analysis and future prediction.

wholesale markets, and intermediary wholesalers, information is fragmented due to the predominance of analog transactions based on telephone calls, fax, and paper, and those types of transactions involve a great deal of wasted time and labor. The current lack of digitization of data on transactions makes it difficult to collect and analyze information. Furthermore, many of the operations involved in agricultural-products assignment have become individualized and the expertise has become tacit knowledge of individuals. In preparation for future transactions in the virtual market, it is necessary to (i) establish a data-collection mechanism while digitizing business operations and (ii) accumulate and analyze the collected data so that future forecasts and business expertise, which is tacit knowledge, can be reproduced digitally.

3.2 Challenge (b): Optimization of logistics

Agricultural products produced in local areas are currently transported over long distances to wholesale markets in large cities. However, in response to the 2024 problem, it will be necessary to shorten the distance that those products can be transported at one time and the time it takes to load and unload them. Therefore, each production area needs to transport agricultural products gathered at collection points to their destinations in the optimal manner. The optimum solution to the transport problem must account for a variety of factors such as travel routes, transit points, mixable items that can be transported in the same truck, and time required for loading and unloading produce.

3.3 Challenge (c): Matching supply and demand in a form similar to local production for local consumption

Even if information is digitized and collected in the virtual market, optimizing delivery without changing the conventional shipping destination and digitizing supply-and-demand matching will not alleviate the concentration of shipments to major urban markets. The concentration of shipments to big-city markets will continue, and waste will remain because surplus agricultural products will still be transferred between markets or shipped far away. The optimal supply-and-demand matching that we want to achieve using the virtual market is to match supply and demand in consumption areas close to the production area in a manner similar to local production for local consumption. To achieve this goal, it is necessary to (i) collect information not only from one market but also from multiple markets on the virtual market, and (ii) achieve optimal supply-and-demand matching by taking into account information on supply-and-demand volumes and delivery costs for the multiple markets.

4. Current efforts in implementing the virtual market

To implement the virtual market, we are working with the Shinmei Group, a major wholesaler of agricultural products with multiple fruit-and-vegetable wholesale markets within its group, to examine and address each of the above challenges step by step. We have been conducting a demonstration experiment on challenge (a), namely, data collection and digitization of expertise. In this experiment, we are working on the digitization of information on orders (demand) and supply gathered in the market and operations relating to the matching of supply and demand. The latter issue is referred to as product allocation.

4.1 Overall picture of digital distribution

The current and future work flows of the distribution system are shown in **Fig. 4**. Regarding the current work flow, information about incoming products and orders from intermediary wholesalers are exchanged by phone, fax, etc. The information about supply and demand, which comes in various forms, is compiled by the persons in charge of product allocation in the form of paper memos, Excel spreadsheets, or other methods to which they are accustomed. At that time, supply and demand is not adjusted in advance, so not only the supply and demand in accordance with production area and class/grade but also the total amounts of supply and demand do not match exactly. Therefore, for each individual order or intermediary wholesaler, the extent to which differences in production area, grade/class, and quantity are acceptable is determined on the basis of the expertise of the persons in charge of product allocation (which is not written on order forms), and products are distributed in accordance with the situation at the time. Wholesale markets have a principle of "full acceptance," namely, all agricultural products that arrive must be accepted and sold due to regulations. Therefore, they look for intermediary wholesalers who are willing to accept additional quantity, even if the quantity is larger than ordered. All incoming products are manually allocated to orders.

After the digitalization of the distribution system, the supply-and-demand information is consolidated

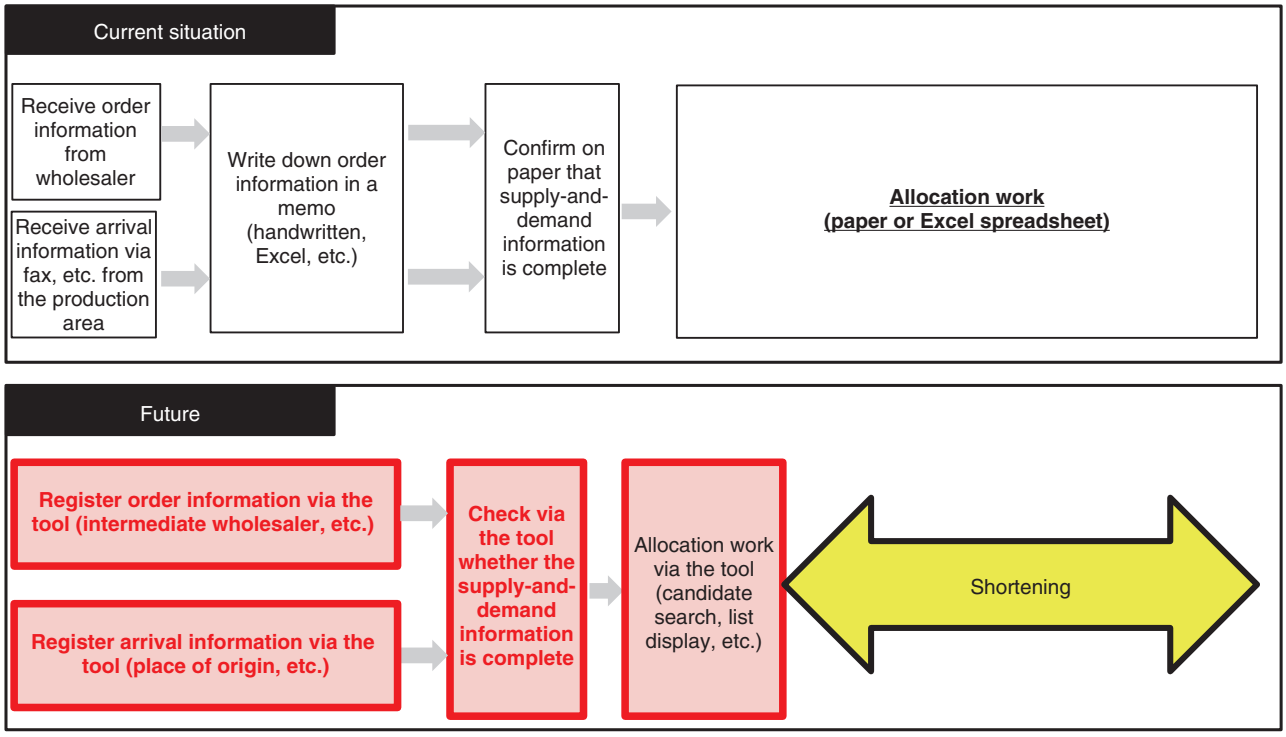


Fig. 4. Comparison of work flows of allocation operations before and after digitalization.

on a tool, which shows the status of supply and demand clearly. Instead of having to manually search for orders for product allocation one by one, the tool supports users searching for candidates and creating an allocation plan. The tool enables users to link orders and products and search for candidates with various conditions via a simple operation. We are aiming for automatic allocation; that is, an allocation plan is created at the touch of a button on the basis of aggregated supply-and-demand information. This capability will significantly reduce work time, even if a part of the allocation plan must be manually corrected.

4.2 Functional blocks of the allocation tool

The product allocation tool consists of the functional blocks shown in **Fig. 5**. The flow of using this tool consists of the following four steps.

(1) Supply information from production areas is registered in the supply database, and order information from intermediary wholesalers is registered in the order database.
(2) The allocation-calculation block invokes the rule-based and history-based allocation functions to create an allocation plan in accordance with the registered supply information and order information.
(3) The created allocation plan is checked on the user interface (UI), and the person in charge of allocation makes corrections manually as necessary.
(4) Finally, confirmed distribution results are collected and analyzed using the history-based allocation function, and the learned content is reflected in the next generation of automatic allocation plans.

4.3 Requirements for the allocation-calculation block

A requirement for automatic allocation is to reflect the various judgments made by the person in charge of allocation during actual work. Examples judgments made by the person in charge of allocation include the following:

- Carry out matching with the highest priority on securing quantity.
- Carry out matching as specified in the order regarding production area, grade, and class.
- Carry out matching not necessarily according to the order but to the assumptions of the person in

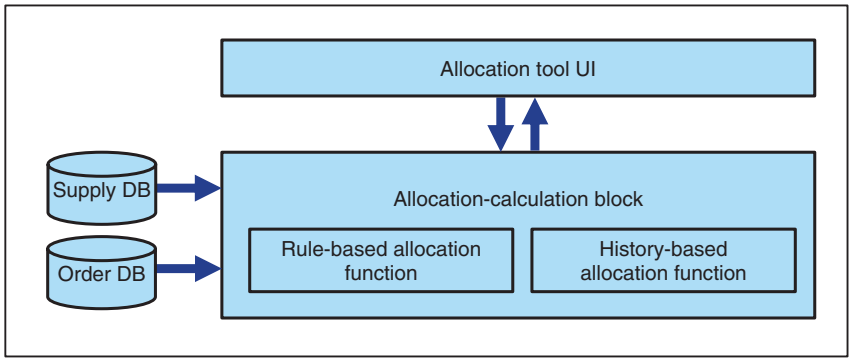


Fig. 5. Block diagram of functions of allocation tool.

Table 1. Supply information.

Share	Production area	Grade	Class	Quantity
Supply 1	Tochigi Prefecture	A	L	6
Supply 2	Kumamoto Prefecture	A	M	10

Table 2. Order information.

Share	Production area	Grade	Class	Quantity
Order 1	Tochigi Prefecture	Excellent	L	10
Order 2	Kumamoto Prefecture	A	L	10

charge of allocation.

- Names of grades and classes depend on the shipping agent and may differ from those stated in the order. Handle these differences correctly and match them properly.

These matters should be properly handled with the tool. The person in charge of allocation makes decisions flexibly in accordance with various circumstances, such as variations in supply due to the time of year or weather conditions or orders with different conditions such as bargain sales or food/gourmet fairs. The conditions that must be taken into account depend on the intermediary wholesalers. We are investigating an automatic-allocation method that can appropriately reflect the various conditions and assumptions taken into account by the person in charge of allocation.

4.4 Overview of algorithm for automatic allocation

The order and supply are first compared by assigning similar attributes to the order and supply. An example of this allocation procedure is shown in **Tables 1** and **2**, respectively. Orders that must be fulfilled according to quantity have a "quantity required" flag attached to them to indicate that condition. For example, if "order 1" has that flag, for "order 1," the algorithm will output an allocation result that satisfies the order quantity of "10." Next, the example of "order 1" is used to illustrate that the person in charge of allocation can consider different decision criteria, such as whether the matching is prioritizing production areas or satisfying quantity.

(1) When products are allocated with priority given to the production area: "order 1" is assigned to "supply 1," which does not meet the quantity condition but meets the production-area condition, and the quantity is set to "6."

(2) When products are allocated with priority given to meeting quantity: "order 1" is assigned to "supply 2," which is from a different

production area but satisfies the quantity condition, and the quantity is set to "10."

(3) When products are allocated according to an order with a required-quantity flag and priority is given to the production area: "order 1" is assigned to "supply 1," which satisfies the production-area condition, and the quantity is set to "6." If this does not satisfy the required quantity condition, "supply 2" is further assigned even if the production is in a different area from the condition. By assigning a quantity of "4" to "supply 2," the quantity is set to "10."

Therefore, automatic allocation makes it possible to output allocation results according to various criteria (tacit knowledge) considered by the person in charge of allocation. To create the algorithm for supply-and-demand matching with such data, in addition to expressing the conditions as rules as described above, we are investigating a method of expressing the correctness of matching as a parameter and using this method to optimize the matching.

The history-based allocation function is also used to digitize the expertise possessed by persons in charge of allocation. This function accumulates and learns whether the person in charge of allocation used the allocation plan created by the tool as is or modified it and if they modified it, how they modified it. We are trying to use this function to learn the expertise of the person in charge of allocation, which cannot be expressed in a rule-based manner, and to output the results of allocation in consideration of that expertise.

5. Future developments

The digitalization of the allocation system is only the first step toward implementing the virtual market. This initiative will change the conventional agricultural-product distribution system and create a new system that separates the products and commercial flow. To create the virtual market, therefore, the understanding and cooperation of all stakeholders will be required. In addition to technology, other issues concerning the virtual market remain to be resolved; regardless, to provide better value to the many people involved, from producers to consumers, we are committed to making significant changes in the distribution of agricultural products in Japan through the virtual market.

Reference

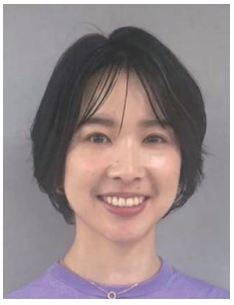
[1] Ministry of Agriculture, Forestry and Fisheries, "Survey on Agricultural Products Distribution," 2018 (in Japanese). https://www8.cao.go.jp/kisei-kaikaku/suishin/meeting/wg/nourin/20181003/181003nourin04.pdf

Toshihiro Baba
Senior Research Engineer, NTT Smart Data Science Center, NTT Computer and Data Science Laboratories.

He graduated from Kyoto Institute of Technology and joined NTT WEST in 2007. Since 2023, he has been investigating the efficiency improvement of agricultural-product distribution in his current position.

Takafumi Hikichi
Senior Research Engineer, NTT Smart Data Science Center, NTT Computer and Data Science Laboratories.

He received an M.E. from Nagoya University and joined NTT in 1995. He is engaged in research and development in information technologies such as speech, language, and data processing. He received a Ph.D. in computer science from Nagoya University in 2007. He is currently investigating the data science for agricultural-product distribution. He is a member of the Institute of Electronics, Information and Communication Engineers (IEICE).



Shiori Toyoshima
Senior Research Engineer, NTT Smart Data Science Center, NTT Computer and Data Science Laboratories.

She received an M.E. in computer science from Ochanomizu University, Tokyo, in 2011 and joined NTT the same year. From 2014 to 2022, she was a solution sales representative at NTT Communications, providing consulting services about private cloud and data analytics solutions to corporate clients. She is currently investigating the optimization of agricultural distribution using data science.

Yasuhiro Hirano
Senior Research Engineer, NTT Smart Data Science Center, NTT Computer and Data Science Laboratories.

He received an M.E. from Kanazawa University and joined NTT in 1989. He is engaged in research and development in information technologies such as databases, multimedia, IC cards, media search, and voice mining. He is currently investigating the data science for agricultural-product distribution. He is a member of IEICE and the Institute of Electrical and Electronics Engineers (IEEE) Computer Society.

Yasushi Hanakago
Senior Research Engineer, NTT Smart Data Science Center, NTT Computer and Data Science Laboratories.

He received an M.S. from Hokkaido University and joined NTT in 1990. He is currently engaged in investigating the data science for agricultural-product distribution.

Akira Isoda
Research Engineer, NTT Smart Data Science Center, NTT Computer and Data Science Laboratories.

He received an M.E. from Tohoku University, Miyagi, in 2014. He joined NTT Network Innovation Laboratories in 2014, where he was involved in R&D of space division multiplexing (SDM) optical transmission network systems, including crosstalk monitoring in multi-core fibers and SDM nodes. He was with NTT Service Evolution Laboratories, where he engaged in R&D of agricultural-damage-reduction technologies and crop-management systems using network-monitoring techniques, such as bird and beast detection systems with complex stimulation and yield estimation and stock simulation tools. He is currently working at NTT Computer and Data Science Laboratories focusing on demand-and-supply matching technologies for an efficient agricultural delivery network as part of the virtual market concept. He is a member of IEICE.

Junji Tomita
Senior Research Engineer, Supervisor, NTT Smart Data Science Center, NTT Computer and Data Science Laboratories.

He received an M.S. in computer science from Keio University, Kanagawa and joined NTT in 1997. He was a visiting scholar at the University of Washington in 2005 and received a Ph.D. from Keio University in 2012. He is engaged in the application of AI technology to real fields including renewable energy, construction, distribution, and marketing. He is a member of the Information Processing Society of Japan and was a board member of the Database Society of Japan from 2016 to 2020.

Refining Solar-power-generation Plans to Achieve Stable Power Supply by Predicting Total Solar Irradiance

Toshitaka Maki, Kazuma Matsui, Takashi Fujinami, Hisashi Kurasawa, and Junji Tomita

Abstract

To reduce the environmental burden, it is necessary to adjust the supply and demand of electricity so that the proportion of renewable energy is increased. For solar-power generation, which is widely used, total solar irradiance is the main variable used in power-generation planning because it is strongly correlated with the amount of electricity generated. However, measurement and prediction of total solar irradiance produce values with large errors, leading to large errors in power-generation plans, which in turn cause problems in terms of stable power supply and the cost of adjusting supply and demand. In light of these issues, we developed a technology that combines weather information and power-generation results to predict total solar irradiance with high accuracy in a manner that makes it possible to refine solar-power-generation plans.

Keywords: photovoltaics, power-generation forecast, total solar irradiance

1. Introduction

As social and economic systems have been developed, technologies and services that respond to the diverse needs of individuals have been created, making it possible for people to lead rich social lives. However, the problems of global warming and climate change have become more apparent, and carbon neutrality has become a major theme.

In September 2021, NTT Group formulated a new environment and energy vision, called "NTT Green Innovation toward 2040," which aims to simultaneously achieve zero environmental impact and economic growth by "reduction of environmental impact through business activities" and "creation of breakthrough innovation."

Our major efforts to achieve carbon neutrality include "increased use of renewable energy" and "lower energy consumption with the Innovative Optical and Wireless Network (IOWN)" [1]. To expand the use of renewable energy, we will not only procure renewable-energy power sources from other companies but also develop and promote renewable energy such as photovoltaics (PV) within our own group and collaborate with partners and NTT Group companies [2].

To expand the use of renewable energy, it is essential to provide a stable supply of electricity. Renewable-energy sources are difficult to plan, that is, it is difficult to predict when, where, and how much power will be generated because the amount of power generated depends on natural conditions.

Electric-power producers had been able to generate stable income from the sale of electricity under the feed-in tariff (FIT) system[*1] without having to create a renewable-energy-generation plan, so they have been able to promote the development and use of renewable energy. However, with the introduction of the feed-in premium (FIP) system[*2], electric-power

*1 FIT system: A system in which power companies purchase electricity generated from renewable energy at a fixed price.

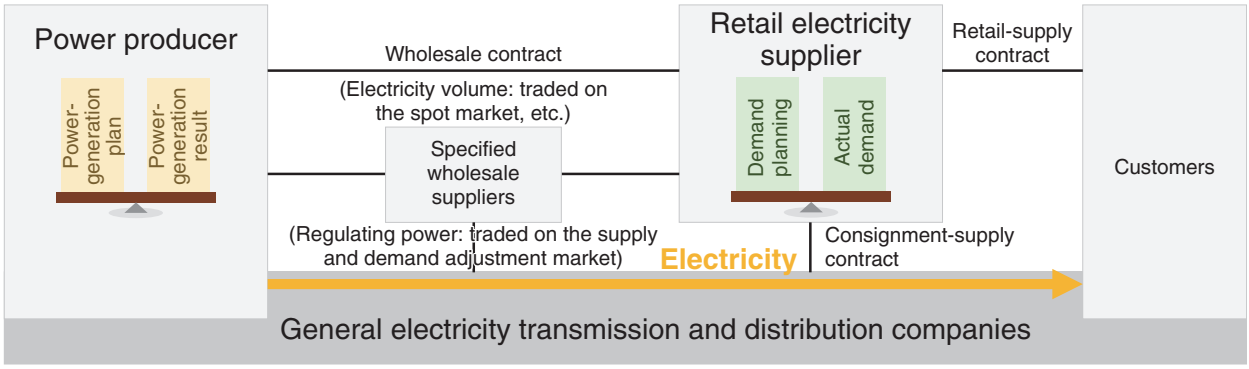


Fig. 1. Overview of an electric-power system.

producers are now required to create a renewable-energy-generation plan and trade their power in the electricity market and other markets on the basis of that plan.

A variety of renewable-energy sources, such as hydropower, wind, and solar power, are available; however, in Japan, the use of PV is expanding, especially from the perspective of economic efficiency and environmental suitability (suitable locations). With PV, power generation (output) and total solar irradiance are strongly correlated; therefore, to draw up accurate power-generation plans, it is necessary to accurately measure and predict total solar irradiance.

We first broadly discuss stable power supply, which is fundamental in regard to expanding the use of renewable energy. We then discuss the measurement and correction technology for total solar irradiance we developed to improve PV-output forecasting.

2. The necessity for a stable supply of electricity

Simply put, to achieve a stable supply of electricity, the supply and demand of electricity must be perfectly matched to prevent power outages. If supply exceeds (or, conversely, subceeds) demand, various electrical products will malfunction and, in the worst case, power outages will occur. Accordingly, electricity utilities that support the electric-power system are obligated to make utmost efforts to ensure a stable supply of electric power. Japan has introduced a planned price coincidence system, under which power producers and retail electricity suppliers are obligated to match planned and actual electricity consumption in 30-minute increments per time block, and the system is operated so that the generation plan and demand plan coincide as a whole.

The electric-power system shown in **Fig. 1** is composed of a power producer, retail electricity supplier, and general electricity transmission and distribution company. The obligation to match plans and actual results is imposed on the power producer and retail electricity supplier, and if a discrepancy (imbalance) between their power-generation and demand plans and actual results unavoidably occurs, the general electricity transmission and distribution company will use its regulatory power to eliminate the imbalance. In this case, however, depending on the amount of imbalance, settlement is required at a later date, and the cost tends to be higher than when supply and demand are adjusted in an electricity market such as the spot (one-day-ahead) market or the hourly (same-day) market. It is therefore essential that the power producer and retail electricity supplier formulate accurate plans to avoid imbalances to reduce business risks and maximize profits.

The imbalance caused by the power producer is generally called the generation imbalance, while the imbalance caused by the retail electricity supplier is called the demand imbalance. As shown in **Table 1**, there are two types of imbalances, and the type depends on one of two imbalance situations. For example, if the actual power generated falls short of the generation plan, this situation is called a shortage imbalance, and the general electricity transmission and distribution company will make up the shortage and bill the customer for the amount of electricity generated at a later date. Conversely, if the actual power generated exceeds the generation plan, this

*2 FIP system: A system by which businesses sell electricity generated from renewable energy on the electricity market, etc. and add a premium (subsidy amount) to the selling price.

Table 1. Types of imbalance.

	Plan > Actual	Plan < Actual
Power producer	Shortage imbalance Amount of power generated is lower than planned, so there is a power shortage.	Surplus imbalance Amount of power generated exceeds the plan, so there is a surplus of electricity.
Retail electricity supplier	Surplus imbalance Amount of power generated exceeds the plan, so there is a surplus of electricity.	Shortage imbalance Amount of power generated is lower than planned, so there is a power shortage.

situation is called a surplus imbalance, and the general electricity transmission and distribution company will purchase the surplus amount of electricity. In a surplus imbalance, the power producer appears to incur no loss; however, they do not know the purchase price until a later date and, in any case, they must pay compensation to the general electricity transmission and distribution company in accordance with the adjusted purchase amount. As mentioned above, it is thus essential for power producers and retail electricity suppliers to formulate accurate plans to avoid imbalances, reduce business risks, and maximize profits.

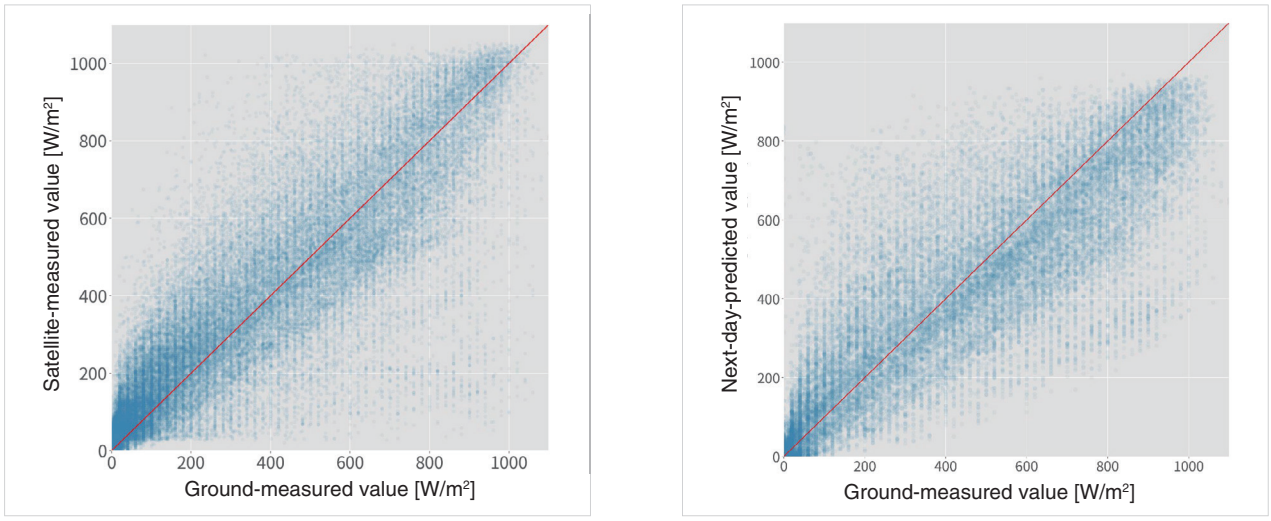
3. Challenges in predicting PV output

To draw up accurate power-generation plans, it is necessary to accurately predict the amount of power generated. Renewable-energy sources, also known as naturally variable power sources, are subject to large fluctuations in power output due to factors such as weather conditions, seasons, and time of day. As for one of those power sources, PV output varies greatly with the intensity of the total solar irradiance falling on the surface of the PV panels. Basically, PV output can be accurately predicted if, in addition to the intensity of total solar irradiance, the effects of shadows caused by topography and buildings, the characteristics of the PV panels and power-conditioning system (PCS)[*3], and any water droplets and dirt on the surface of the PV panels are known. However, current technologies and services for PV output forecasting have large errors, and each company is working to reduce these errors by incorporating various innovations in forecasting technology, establishing power-generation balancing groups, and reducing errors through renewable-energy aggregation. Efforts to control the generation imbalance by introducing grid storage batteries are also being made [3]. However, the smoothing effect of these measures is limited, and even if grid storage batteries are used, it is necessary to ensure that they have enough capacity to absorb the imbalance. From the standpoint of a power producer, it is a prerequisite that the generation imbalance be minimized to achieve cost benefits; therefore, it is essential to reduce errors in forecasted PV output, thus promote the introduction and expansion of renewable-energy sources.

Errors in forecasted PV output are mainly due to errors in measured and forecasted total solar irradiance, so it is important to reduce those errors. Total solar irradiance is generally measured using meteorological satellites, and as shown in **Fig. 2(a)**, there is a significant error between ground-based and satellite measurements. If the ground and satellite measurements agreed perfectly, all the points would be plotted on the red line; however, in reality there is a considerable error. Since total solar irradiance measured at ground level is the true value, it is desirable to be able to use the ground-level measurements. However, only about 50 meteorological observatories across Japan are equipped to measure ground-level total solar irradiance, which makes it impossible to measure total solar irradiance uniformly over a wide area. Technologies and services that use sunshine-duration data obtained from AMeDAS (Automated Meteorological Data Acquisition System) of the Japan Meteorological Agency to pseudo-measure total solar irradiance are available; however, total solar irradiance and sunshine duration are different parameters, and there are only 690 AMeDAS stations nationwide [4] that can measure sunshine duration. It is thus difficult to obtain accurate total solar irradiance for any given location.

Two main models are used to predict total solar irradiance: numerical and statistical. Numerical models predict future total solar irradiance by calculating elements such as solar and atmospheric irradiation, cloud coverage, and aerosol concentration on the

*3 PCS: A system that converts direct current (DC) power generated by solar panels into alternating current (AC) power.



(a) Comparison of ground-measured values and satellite-measured values

(b) Comparison of ground-measured values and next-day-predicted values

Fig. 2. Example of measurement errors and forecast errors of total solar irradiance.

basis of physical laws. Statistical models predict future total solar irradiance by analyzing the relationship between total solar irradiance and other variables on the basis of historical measurement data and satellite images. Ensemble prediction models—which combine numerical and statistical models—and, more recently, deep-learning prediction models have been proposed [5]. However, total solar irradiance predicted by the currently widely used numerical forecasting models is generally calculated after predicting various meteorological factors such as cloud cover and atmospheric pressure, and as shown in **Fig. 2(b)**, prediction errors in those factors lead to large errors in predicted total solar irradiance. The figure shows the relationship between the ground-measured total solar irradiance and the predicted (next day) values provided by weather companies. The errors shown here appear directly as errors in predicted PV output. To reduce errors in predicted PV output, we therefore constructed a high-quality PV-output prediction model by reducing measurement errors in total solar irradiance. The model reduces the prediction error in total solar irradiance and enables highly accurate prediction of PV output.

4. Measurement and correction technology for total solar irradiance by using PV output

As shown in **Fig. 3**, this measurement and correction technology consists of two methods: estimation of measured total solar irradiance and correction of predicted total solar irradiance. These methods reduce the error between measured and predicted total solar irradiance.

4.1 Estimation of measured total solar irradiance

To reduce measurement errors in total solar irradiance, it would be ideal to install pyranometers at high density on the ground; however, that method is not realistic owing to high installation and maintenance costs. Therefore, we use PV output—which can be monitored remotely—to measure total solar irradiance with high precision. Although technologies for measuring total solar irradiance by using PV output have been available for some time, they have not yet been developed into services because they only use PV-panel output and do not take into account the effects of neighboring buildings and topography as well as water droplets and dirt on the PV-panel surfaces.

With the above-mentioned issues in mind, we proposed a method for measuring total solar irradiance with high accuracy by using meteorological information obtained at multiple locations, information about PV facilities, and power-generation records while indirectly taking into account the effects of neighboring buildings and topography. In regard to social implementation, we have started to provide the

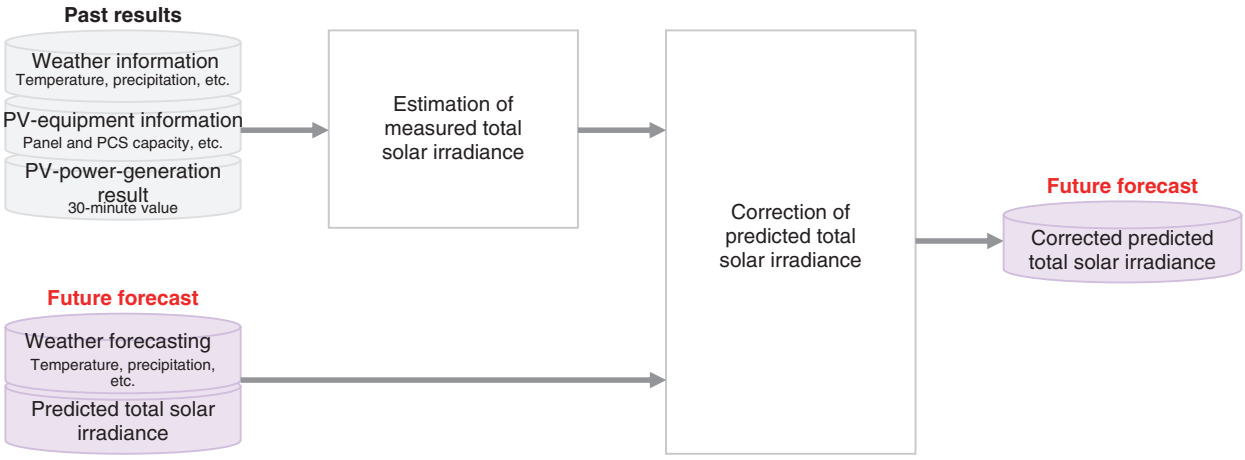
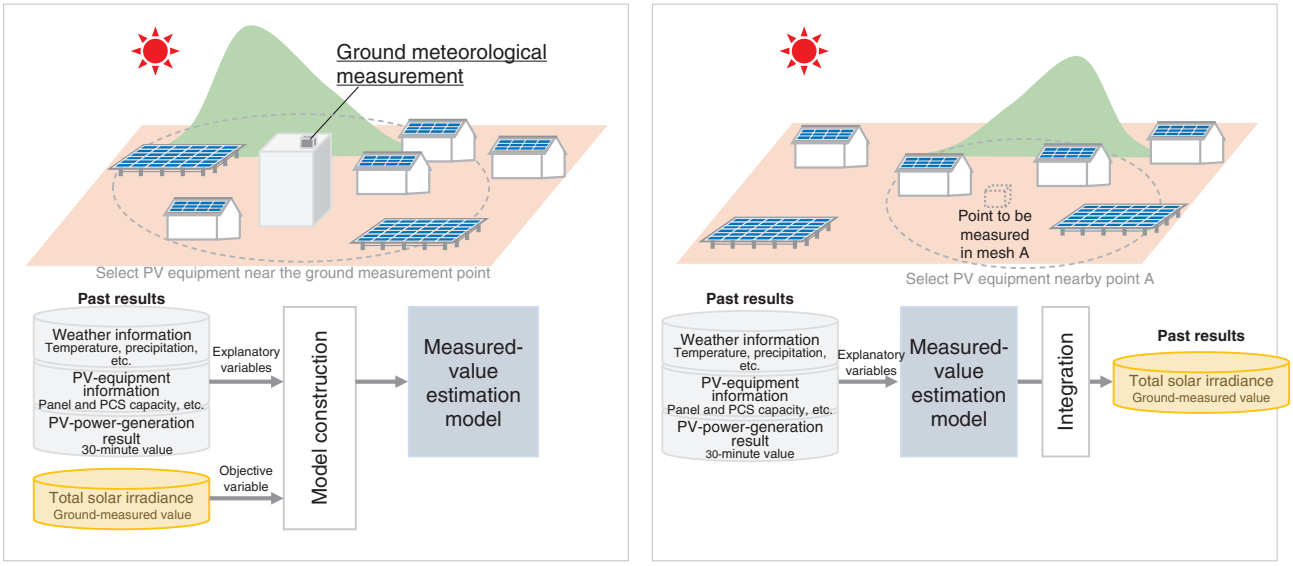


Fig. 3. Overview of our measurement and correction technology for total solar irradiance.



(a) Construction of measured-value estimation model

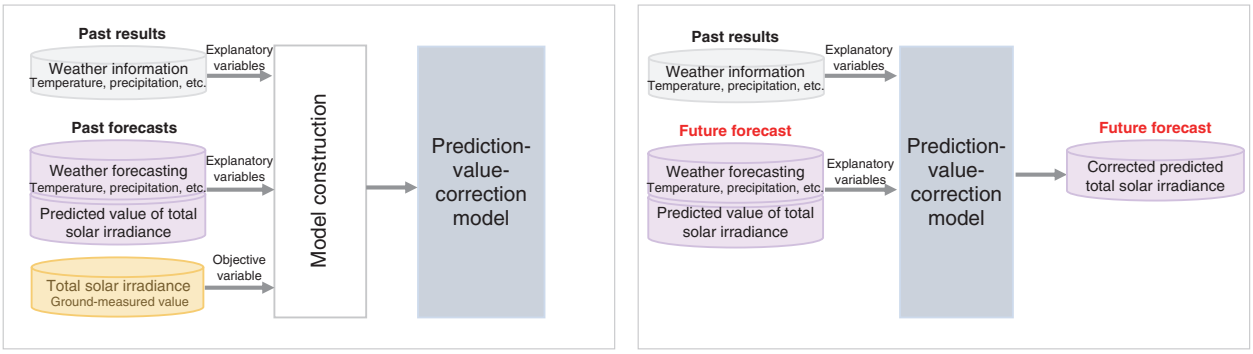
(b) Measurement of total solar irradiance by using measured-value estimation model

Fig. 4. Estimation of measured total solar irradiance.

proposed method as a service to power-generation planners and others.

As shown in **Fig. 4**, the proposed method has two phases: (a) construction of a measured-value estimation model and (b) measurement of total solar irradiance by using the measured-value estimation model.

In phase (a), weather stations that measure ground-level total solar irradiance and PV equipment around our own measurement points are selected, weather information, PV-equipment information, and power-generation results are used as explanatory variables, the ground-level measurement of total solar irradiance is used as an objective variable, and a measured-value estimation model is constructed. For example, to improve the accuracy of the ground-level total-solar-irradiance measurements by taking into account the condition of water droplets and dirt on the surface of the PV panels, the amount of precipitation on the



(a) Construction of predicted-value-correction model

(b) Correction of predicted total solar irradiance by using predicted-value-correction model

Fig. 5. Correction of predicted total solar irradiance.

day of the measurements and the trend of PV output from several days before the measurement are incorporated as explanatory variables. Since PV output may not exceed the PCS capacity*4 and it may be uncorrelated with total solar irradiance, total solar irradiance can be accurately estimated—even when the PV output is at its peak—by incorporating PV output before and after the measurement period in the model as explanatory variables.

In phase (b), the same explanatory variables as in phase (a) are used to incorporate various data at different locations and meshes*5 in the model as explanatory variables. Therefore, it is possible to estimate total solar irradiance equivalent to ground-based measurements at high accuracy even for locations where total solar irradiance is not measured at ground level.

The performance of the proposed method was evaluated at ten sites around Japan (Hokkaido, Fukushima, Gunma, etc.), and the evaluation results confirmed that measurements made with this method have a 50% lower error rate than those made with conventional meteorological satellites.

4.2 Correction of predicted total solar irradiance

To reduce the prediction error of total solar irradiance, the error between the measured and predicted values of total solar irradiance in the past is analyzed, and the predicted total solar irradiance is corrected using the latest total solar irradiance at the forecast time.

Predicted values of total solar irradiance, even when they are predicted one hour ahead of time, have large prediction errors. To reduce such errors, available technologies and services use the latest total solar irradiance measured using meteorological satellites to correct predicted values of total solar irradiance; however, the measured values of total solar irradiance contain large errors, and the corrections to those errors are not fully effective.

We use the ground-level equivalent of total solar irradiance obtained with our measured-value estimation model to correct the predicted values of total solar irradiance. As shown in **Fig. 5**, this method includes two phases: (a) construction of a predicted-value-correction model and (b) correction of predicted total solar irradiance by using the model. In phase (a), the predicted-value-correction model is constructed from historical weather information, weather forecasts, and historical measurements and predictions of total solar irradiance. The correction of predicted values takes into account the following three factors: (i) the continuity of weather conditions, (ii) existence of region-specific weather characteristics, and (iii) tendency for predicted values to be lower than actual values.

For factor (i), for example, if the weather is currently cloudy, it is often cloudy for the next several hours. Accordingly, the error in predicted total solar irradiance can be reduced by correcting the total solar irradiance several hours in advance on the basis of the current total solar irradiance.

For factor (ii), weather characteristics are regionally specific; for example, Kanagawa Prefecture has many sunny hours, while Akita Prefecture has few

*4 PCS capacity: Maximum capacity of DC/AC conversion.
*5 Mesh: A region that is divided into a mesh.

sunny hours. From another viewpoint, for example, the effect of altitude and topography can cause regional differences (of about 30 minutes) in the number of hours the sun is out, even within the same Tokyo metropolitan area. Corrections to predicted total solar irradiance that take into account such region-specific weather characteristics can reduce errors in the predictions.

For factor (iii), predicted values may be lower overall than actual values due to the difference between mesh sizes used in the measurements and predictions. For example, for reasons related to the specifications of the numerical prediction model, a 1-km mesh is used for measurements, while a 5-km mesh is used for predictions. As a result, the predicted values may be lower than the actual values. The error can be reduced by taking into account and correcting for such deviations in values due to differences in mesh size. By considering the above three factors, it is possible to correct the predicted values of total solar irradiance so they become close to the actual values.

In phase (b), the corrected future total solar irradiance can be obtained by inputting the latest (future) weather forecast into the correction model.

This method has succeeded in correcting not only the predicted total solar irradiance several hours ahead on the same day but also the predicted total solar irradiance for the next day and the day after that. For these predictions, compared with conventional methods, this method reduced prediction error by about 25%, although regional differences in the effectiveness of the correction were observed.

5. Future developments

To make renewable energy a main power source, it is necessary to ensure a stable supply of electricity and reduce business risks. It is thus essential to reduce the power-generation imbalance that affects renewable energy. To reduce the error in predicted PV output, we proposed a method for estimating measured total solar irradiance and a method for correcting predicted total solar irradiance to reduce the error between the measured and predicted values of total solar irradiance, which is the main cause of error in predicted PV output. We confirmed the effectiveness of these methods in the field. However, we found that when the errors in predicted total solar-irradiance values provided by weather companies are large, correction with the proposed method was less effective in reducing the errors.

With that issue in mind, we plan to investigate new prediction technologies that do not rely on predicted total-solar-irradiance values provided by weather companies. As well as researching forecasting of power generation, we are also pursuing research and development on using digital twins to optimize the entire supply-demand matching process by taking into account the supply and demand of electricity to enable a stable supply of electricity in a decarbonized society. Thus, we are contributing to the creation of a sustainable society.

References

[1] T. Yoshimatsu, Y. Yagi, K. Matsuo, and K. Asakura, "NTT Group's New Environmental and Energy Vision 'NTT Green Innovation toward 2040,'" NTT Technical Journal, Vol. 33, No. 12, pp. 44–47, 2021 (in Japanese).

[2] Y. Sumida and T. Kirimoto, "NTT Anode Energy's Smart Energy Business," NTT Technical Journal, Vol. 33, No. 12, pp. 56–59, 2021 (in Japanese).

[3] Agency for Natural Resources and Energy, "Energy White Paper 2021," pp. 172–190, 2022 (in Japanese). https://www.enecho.meti.go.jp/about/whitepaper/2022/pdf/3_3.pdf

[4] Ministry of Land, Infrastructure, Transport and Tourism, "The Role of AMeDAS" (in Japanese). https://www.mlit.go.jp/common/001043841.pdf

[5] A. M. Assaf, H. Haron, H. N. A. Hamed, F. A. Ghaleb, S. N. Qasem, and A. Albarrak, "A Review on Neural Network Based Models for Short Term Solar Irradiance Forecasting," MDPI, Appl. Sci., Vol. 13, No. 14, p. 8332, 2023. https://doi.org/10.3390/app13148332

Toshitaka Maki
Research Engineer, Scientist, NTT Smart Data Science Center, NTT Computer and Data Science Laboratories.
He received an M.E. and Ph.D. in engineering from Fukuoka Institute of Technology in 2015 and 2018. In the doctoral program, he was a fellow researcher (DC2) of Japan Society for the Promotion of Science (JSPS). He is engaged in the application of AI technology related to electricity such as renewable energy, flow control, and market price prediction.

Kazuma Matsui
NTT Smart Data Science Center, NTT Computer and Data Science Laboratories.
He received an M.E. from Keio University, Kanagawa, in 2022. He joined NTT in 2022. From 2022, he has been investigating AI technology and data science.

Takashi Fujinami
Senior Research Engineer, NTT Smart Data Science Center, NTT Computer and Data Science Laboratories.
He received an M.E. from Nara Institute of Science and Technology in 1996. He is engaged in the research and development of data-science technologies.

Hisashi Kurasawa
Senior Research Engineer, NTT Smart Data Science Center, NTT Computer and Data Science Laboratories.
He received a B.E., M.E., and Ph.D. from the University of Tokyo in 2006, 2008, and 2011. His research interests are data mining and medical and health informatics. He is a member of the Institute of Electronics, Information and Communication Engineers (IEICE), Information Processing Society of Japan (IPSJ), and Japan Association for Medical Informatics (JAMI).

Junji Tomita
Senior Research Engineer, Supervisor, NTT Smart Data Science Center, NTT Computer and Data Science Laboratories.
He received an M.S. in computer science from Keio University, Kanagawa and joined NTT in 1997. He was a visiting scholar at the University of Washington in 2005 and received a Ph.D. from Keio University in 2012. He is engaged in the application of AI technology to real fields including renewable energy, construction, distribution, and marketing. He is a member of IPSJ and was a board member of the Database Society of Japan from 2016 to 2020.

Sub-surface-hydrogen-measurement Method for Estimating Hydrogen-embrittlement Risk in Concrete Poles

Ryuta Ishii, Takuya Kamisho, and Akira Sugiyama

Abstract

Concrete poles play an important role in the infrastructure providing communication services to customers. If the risk of hydrogen embrittlement of rebar embedded in concrete can be determined from information about the environment in which the equipment is installed, the safety of concrete poles can be further improved. Our research group has developed a method for measuring the sub-surface hydrogen concentration of rebar in concrete poles, which is one of the most important parameters for measuring hydrogen embrittlement.

Keywords: deterioration risk evaluation, concrete poles, hydrogen embrittlement

1. Maintenance of concrete poles

Various infrastructure structures, such as bridges and steel towers, support daily life. Improving the safety of such infrastructures can make our lives sustainable and better. Our group is researching the deterioration of materials used in infrastructure to contribute to its maintenance and management from the viewpoint of materials science.

Concrete poles are one of the most important components of the infrastructure providing communication and power to customers. Many are located very close to us. Therefore, if one breaks, it may lead to a serious accident or communication outages. High safety is required for concrete poles to prevent serious accidents. A concrete pole is a structure consisting of concrete and rebar, and the safety of the pole is secured by the rebar inside. The rebar is protected from corrosion by the alkaline environment of the concrete. Since rebar is embedded in concrete under tensile stress, it applies compressive stress on the concrete. This prevents the concrete from cracking.

As described above, concrete and rebar complement each other in strength, so the rebar in a concrete pole does not break under normal conditions. However, when various negative conditions combine, a degradation phenomenon known as hydrogen embrittlement may develop, leading to a reduction in the strength of concrete poles [1–3]. It is difficult to detect the deterioration of rebar because it is embedded in the concrete and cannot be inspected visually. For such invisible deterioration, it is useful to use a method for evaluating the risk of deterioration from environmental information surrounding the location of the equipment, such as temperature and humidity [4]. If it is possible to evaluate the hydrogen-embrittlement risk of rebar in concrete poles from environmental information, it would be possible to preferentially update equipment that is at high risk. This could improve the safety of entire facilities.

2. Hydrogen embrittlement of rebar in concrete poles

Hydrogen embrittlement is a phenomenon in which the strength of a metal decreases due to hydrogen atoms entering the metal, leading to premature cracking and fracture. Whereas normal metals fracture in a ductile manner, metals degraded by hydrogen embrittlement have a glass-like fracture surface morphology (**Fig. 1**). This phenomenon, also known as delayed fracture, occurs decades after equipment is installed,

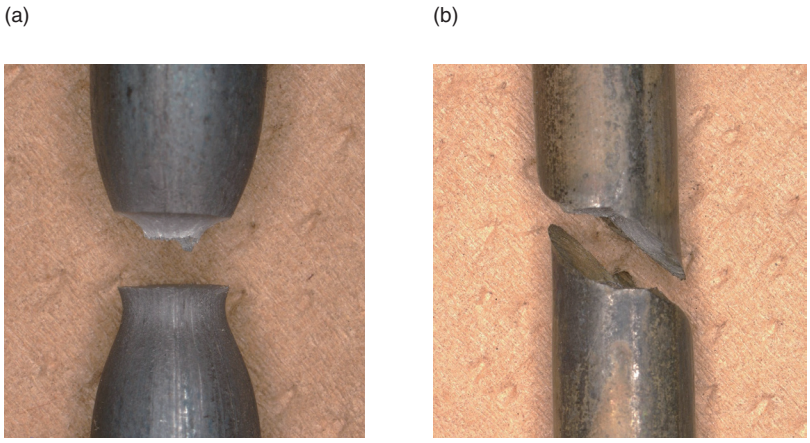


Fig. 1. (a) Fracture of metal without hydrogen penetration (normal condition). (b) Fracture of metal with hydrogen penetration (hydrogen embrittlement).

and its progression is difficult to evaluate. The mechanism is not well understood and is being actively discussed by many researchers worldwide.

Hydrogen embrittlement is considered to develop in concrete poles as follows. Cracks first appear in concrete due to unfortunate circumstances. If a crack is large enough to reach the rebar, the alkaline concrete adhering to the rebar is locally lost around where the crack reached, i.e., the ability to inhibit corrosion is lost, so corrosion of the steel begins. Hydrogen is generated by the reduction reaction of water (cathodic reaction) that proceeds together with the dissolution reaction of iron (anodic reaction). Some of the generated hydrogen enters the rebar as hydrogen atoms, causing the rebar to undergo hydrogen embrittlement [5].

$$Fe \rightarrow Fe^{2+} + 2e^{-} \text{ (anodic reaction)}$$
$$2H_2O + 2e^{-} \rightarrow H_2 + 2OH^{-} \text{ (cathodic reaction)}$$

One of the most important parameters in evaluating the risk of hydrogen embrittlement is the sub-surface hydrogen concentration of rebar. The higher the amount of sub-surface hydrogen in the rebar generated by corrosion, the greater the risk of hydrogen-embrittlement fracture. It is therefore vital to know how the sub-surface hydrogen concentration of rebar in a concrete pole is affected by environmental factors, but there is currently no way to measure this.

3. Hydrogen-permeation test

Electrochemical hydrogen-permeation tests are typically conducted to measure the sub-surface hydrogen concentration [6]. Hydrogen atoms entering from one side of the test object are ionized on the nickel (Ni) or palladium (Pd) passivation layer formed on the opposite side of the penetration side and detected as a hydrogen-permeation current (**Fig. 2**). Assuming that hydrogen diffuses one-dimensionally over a sample length, the following relationship is known between the hydrogen permeation current and sub-surface hydrogen concentration [7].

$$i = \mathrm{F}D\frac{C}{L},$$

where i is the hydrogen permeation current, C is the sub-surface hydrogen concentration, F is Faraday's constant, and L is the sample length. Therefore, the sub-surface hydrogen concentration of a sample can be determined by measuring the hydrogen-permeation current. This is a relatively simple method for measuring the sub-surface hydrogen concentration and is the only method that can measure it continuously and nondestructively.

However, the general electrochemical hydrogen-permeation test is limited in its test geometry to plates only. Therefore, it is not possible to measure the sub-surface hydrogen concentration of rebar in concrete poles with such a method. We developed a hydrogen permeation test method for concrete samples cut from concrete poles, and the sub-surface hydrogen concentration of rebar in concrete poles was measured.

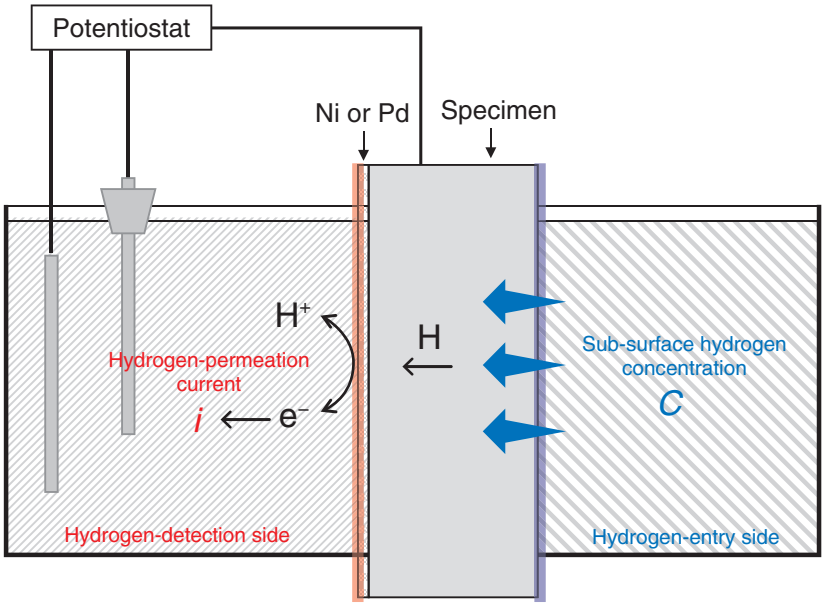


Fig. 2. Schematic showing hydrogen-detection principle of hydrogen-permeation test.

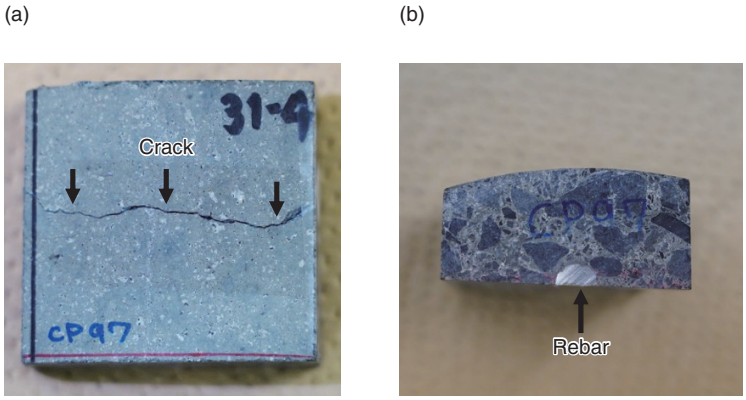


Fig. 3. Concrete specimen for our hydrogen-permeation test: (a) surface side (crack side) and (b) top side.

4. Hydrogen-permeation test extended to concrete specimens

The concrete sample we used for this study was about 4-cm square and included rebar cut from a concrete pole (**Fig. 3**). A crack was introduced into the concrete sample by using the three-point bending method. The sample was set in a self-made electrochemical test cell (L-shaped cell) and was in contact with a sodium hydroxide (NaOH) solution packed in the cell on the hydrogen-detection side (**Fig. 4(a)**). The hydrogen-detection area was set as a 3-mm circle to prevent the solution from contacting the concrete and allow only the rebar to contact the solution, thereby reducing the noise current (**Fig. 4(b)**). Hydrogen was generated by dropping 100 µL of 0.05M sodium chloride (NaCl) solution (equivalent to rainwater) into the concrete crack to cause the rebar to artificially corrode. The solution traveled through the crack and contacted the rebar, causing the corrosion reaction to proceed and start hydrogen generation. The corrosion conditions of the rebar were assumed to be those for the corrosion of rebar in concrete poles in outdoor environments. To obtain

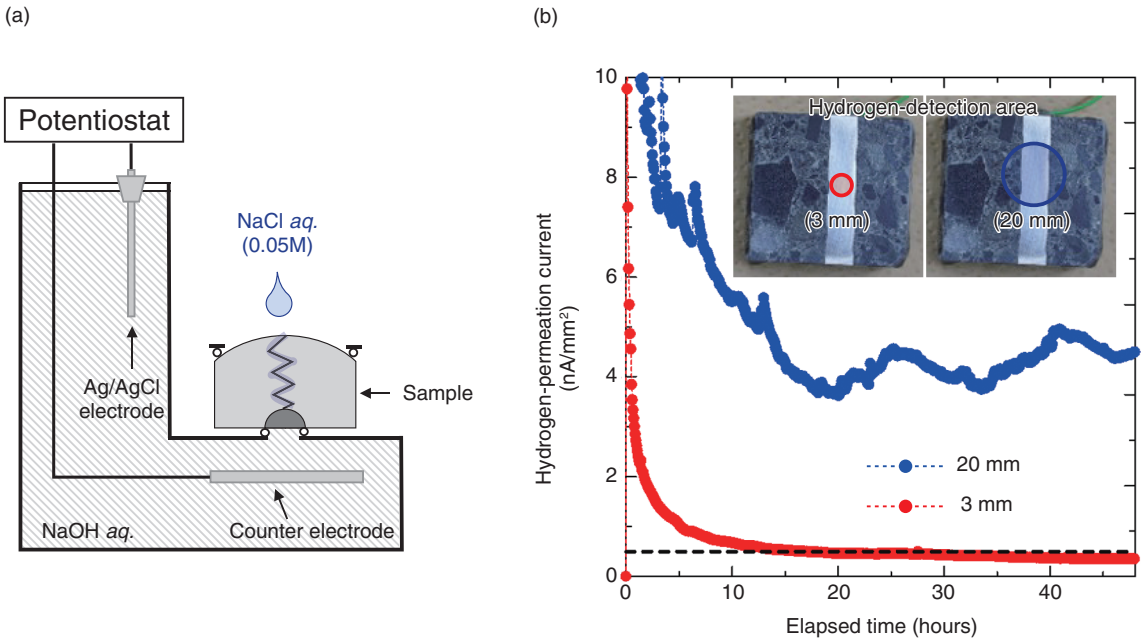


Fig. 4. (a) Schematic of our hydrogen-permeation test. (b) Background current when hydrogen detection area is 3 and 20 mm.

the sub-surface hydrogen concentration of rebar from the obtained experimental data, a simulation of hydrogen diffusing through rebar was carried out. The calculation was carried out assuming that atomic hydrogen diffuses in a round rod according to the following two-dimensional diffusion equation.

$$\frac{dH(t,x,y)}{dt} = D\left(\frac{\partial^2 H(t,x,y)}{\partial x^2} + \frac{\partial^2 H(t,x,y)}{\partial y^2}\right)$$

The corrosion reaction, i.e., hydrogen penetration, was assumed to occur on the surface where rebar was located just under the crack. A boundary condition with a constant sub-surface hydrogen concentration was set along the rebar surface, as shown in **Fig. 5(a)**. As shown in **Fig. 5(b)**, hydrogen penetrates from the steel surface and diffuses through the steel to reach the hydrogen detection surface. To reduce computational complexity, the simulation was conducted with a four-quarter circle in consideration of symmetry [8].

5. Hydrogen-penetration behavior of rebar in concrete poles

Figure 6 shows the results for the penetration current obtained experimentally and the calculated penetration current obtained by simulation at various temperatures (23, 30, and 40°C). The results of the theoretical simulation were generally consistent with the experimental values. The change in sub-surface hydrogen concentration over time obtained from the simulation at each temperature (~0.08 ppm) was consistent with the amount of hydrogen that was supposed to enter the steel by atmospheric corrosion (0.1 ppm) [9]. Due to the crack in the concrete, oxygen diffusion was suppressed and slightly low. These results indicate that the hydrogen generated in the concrete and diffused in the round rod was correctly captured with our method.

At all temperatures, the hydrogen-permeation current began to increase a certain period (about 2 hours at 23°C) after the NaCl solution was dropped and peaked several hours later (about 10 hours at 23°C). The higher the temperature, the faster the hydrogen-permeation current increased and reached maximum. These results are reasonable considering the low diffusion coefficient of high-strength steel and the temperature dependence of the diffusion coefficient. Thereafter, a gradual downward trend was observed over 48 hours. Compared with the test results for atmospheric corrosion of bare pure iron, the decrease in the hydrogen-permeation current was extremely gradual, meaning that the corrosion continued for a long period. This long-lasting corrosion is attributed

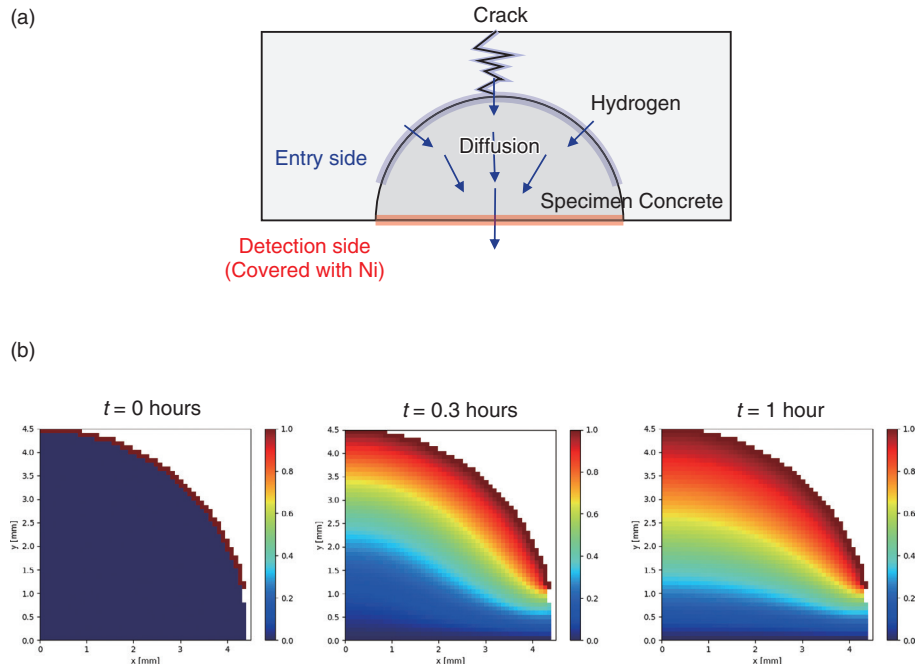


Fig. 5. (a) Hydrogen-diffusion behavior in rebar assumed in simulation. (b) Hydrogen diffusion in rebar (quarter circle) obtained by simulation.

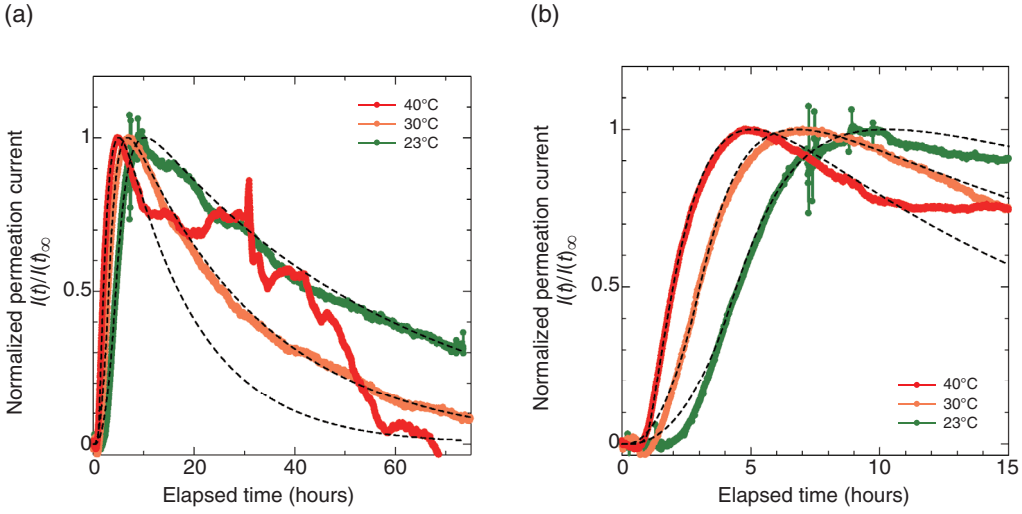


Fig. 6. (a) Hydrogen-permeation currents of concrete specimens at 23, 30, and 40°C. Current values are normalized to maximum. Dashed lines are results of each simulation. (b) Enlarged view shows up to 10 hours of elapsed time.

to the fact that cracks in concrete strongly retain water. It is assumed that a considerable amount of moisture is retained inside concrete cracks for a long period, and a condition such as solution corrosion is assumed to continue. The difference between the experimental and simulated results at 40°C is considered due to the rapid evaporation rate of the solution. Evaporation of the solution in the crack led to the concentration of chloride ions and accelerated corrosion.

By using our hydrogen-permeation test method, it was possible to clarify the hydrogen-penetration behavior of rebar in concrete. This knowledge can be used in technology for evaluating the risk of hydrogen embrittlement of rebar in concrete poles and will be important for examining methods for repairing concrete poles.

6. Conclusion

To further improve the safety of concrete poles, it is useful to estimate the risk of hydrogen embrittlement of rebar embedded in concrete from environmental information on installed equipment. We developed a method for measuring the sub-surface hydrogen concentration of rebar in concrete poles, which is important for evaluating the risk of hydrogen embrittlement of rebar. Our method is considered important for understanding how hydrogen embrittlement of concrete poles proceeds.

References

[1] L. Vehovar, V. Kuhar, and A. Vehovar, "Hydrogen-assisted Stress-corrosion of Prestressing Wires in a Motorway Viaduct," Engineering Failure Analysis, Vol. 5, No. 1, pp. 21–27, 1998. https://doi.org/10.1016/S1350-6307(97)00034-4

[2] J. Woodtli and R. Kieselbach, "Damage Due to Hydrogen Embrittlement and Stress Corrosion Cracking," Engineering Failure Analysis, Vol. 7, No. 6, pp. 427–450, 2000. https://doi.org/10.1016/S1350-6307(99)00033-3

[3] M. Elices, A. Valiente, L. Caballero, M. Iordachescu, J. Fullea, J. Sánchez-Montero, and V. López-Serrano, "Failure Analysis of Prestressed Anchor Bars," Engineering Failure Analysis, Vol. 24, pp. 57–66, 2012. https://doi.org/10.1016/j.engfailanal.2012.03.007

[4] T. Kamisho, R. Ishii, and M. Tsuda, "Prediction of Hydrogen Embrittlement of Reinforcing Steel Bars in Concrete Poles," NTT Technical Review, Vol. 18, No. 11, pp. 43–47, 2020. https://doi.org/10.53829/ntr202011ra1

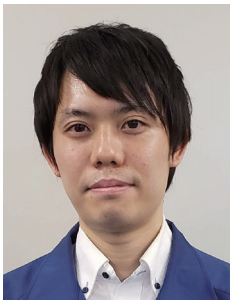
[5] T. Tsuru, Y. Huang, M. R. Ali, and A. Nishikata, "Hydrogen Entry into Steel during Atmospheric Corrosion Process," Corrosion Science, Vol. 47, No. 10, pp. 2431–2440, 2005. https://doi.org/10.1016/j.corsci.2004.10.006

[6] M. A. V. Devanathan and Z. Stachurski, "The Mechanism of Hydrogen Evolution on Iron in Acid Solutions by Determination of Permeation Rates," Journal of the Electrochemical Society, Vol. 111, pp. 619–623, 1964. https://doi.org/10.1149/1.2426195

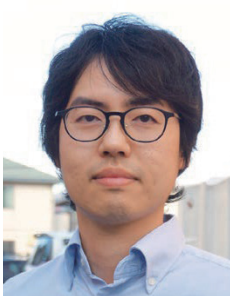
[7] J. Crank, "The Mathematics of Diffusion," Calrendon Press, Oxford, 1975.

[8] R. Ishii, T. Kamisho, N. Fujimoto, and M. Tsuda, "Measurement of Surface Hydrogen Concentration of Steel Bar with Electrochemical Hydrogen Permeation Method," Zairyo-to-Kankyo, Vol. 71, pp. 116–120, 2022 (in Japanese).

[9] T. Kushida, "Hydrogen Entry into Steel by Atmospheric Corrosion," ISIJ International, Vol. 43, No. 4, pp. 470–474, 2003. https://doi.org/10.2355/isijinternational.43.470



Ryuta Ishii
Researcher, Sustainable Material Research Group, NTT Device Technology Laboratories.
He received a B.S. and M.S. in condensed matter physics from Osaka University in 2012 and 2018. He joined NTT Device Technology Laboratories in 2018 and is currently researching hydrogen embrittlement of concrete poles.



Takuya Kamisho
Research Engineer, NTT Device Innovation Center*.
He received a B.S. and M.S. in meteorology from Hokkaido University in 2008 and 2010 and joined NTT Environment and Energy Research Laboratories in 2010. He was engaged in the development of prediction technology for hydrogen embrittlement at NTT Device Technology Laboratories from 2018 to 2023.
* He is currently with NTT EAST Corporation.

Akira Sugiyama
Group Leader, Sustainable Material Research Group, NTT Device Technology Laboratories.
He received a B.S. and M.S. in computer science from Tokyo Institute of Technology in 1994 and 1996. He joined NTT Basic Research Laboratories in 1996 and engaged in research on natural language processing. He moved to NTT Energy and Environment Systems Laboratories in 2004. From 2016 to 2018, he was with the Technical Assistance and Support Center, NTT EAST. From 2018 to 2021, he was with the NTT Environment Promotion Office. He was transferred to NTT Device Technology Laboratories in 2021, where he is currently engaged in the development of technology to visualize infrastructure deterioration. He is a member of the Information Processing Society of Japan.

External Awards

FIT Encouragement Award
Winner: Yuukou Iinuma, NTT Device Innovation Center
Date: September 11, 2023
Organization: Information Processing Society of Japan (IPSJ)

For "Object Detection in 8K Ultra-high-definition Video on Edge Devices."
Published as: Y. Iinuma, S. Hatta, H. Uzawa, S. Yoshida, and K. Yamazaki, "Object Detection in 8K Ultra-high-definition Video on Edge Devices," Proc. of the 22nd Forum on Information Technology (FIT2023), Osaka, Japan, Sept. 2023.

Best Paper Award
Winners: Motohiro Makiguchi, NTT Human Informatics Laboratories; Ayaka Sano, NTT Human Informatics Laboratories; Takahiro Matsumoto, NTT Human Informatics Laboratories; Hiroshi Chigira, NTT Human Informatics Laboratories; Takayoshi Mochizuki, NTT Human Informatics Laboratories
Date: October 15, 2023
Organization: Association for Computing Machinery (ACM)

For "Implementation of Interactive Mirror-Transcending Aerial Imaging System."
Published as: M. Makiguchi, A. Sano, T. Matsumoto, H. Chigira, and T. Mochizuki, "Implementation of Interactive Mirror-Transcending Aerial Imaging System," Proc. of the 2023 ACM Symposium on Spatial User Interaction (SUI 2023), Article no. 12, Sydney, Australia, Oct. 2023.

ICCE 2024 Best Session Presentation Award
Winners: Masahiro Fukui, NTT Computer and Data Science Laboratories; Kazunori Kobayashi, NTT sonority, Inc.; Noriyoshi Kamado, NTT Computer and Data Science Laboratories
Date: January 8, 2024
Organization: The 42nd IEEE International Conference on Consumer Electronics (ICCE 2024)

For "A Seat Headrest Loudspeaker System with Personalized Sound Zone Capabilities."
Published as: M. Fukui, K. Kobayashi, and N. Kamado, "A Seat Headrest Loudspeaker System with Personalized Sound Zone Capabilities," Proc. of ICCE 2024, Las Vegas, NV, USA, Jan. 2024.

Best Paper Award
Winners: Tomohiro Korikawa, NTT Network Service Systems Laboratories; Chikako Takasaki, NTT Network Service Systems Laboratories; Kyota Hattori, NTT Network Service Systems Laboratories; Hidenari Ohwada, NTT Network Service Systems Laboratories
Date: February 21, 2024
Organization: 2024 IEEE International Conference on Computing, Networking and Communications (ICNC 2024)

For "A Routing Method with Link Information-based Rule Selection in Non-Terrestrial Networks."
Published as: T. Korikawa, C. Takasaki, K. Hattori, and H. Ohwada, "A Routing Method with Link Information-based Rule Selection in Non-Terrestrial Networks," Proc. of ICNC 2024, pp. 807–812, Big Island, HI, USA, Feb. 2024.

Silver Contribution Award
Winner: Tomohiro Korikawa, NTT Network Service Systems Laboratories
Date: February 21, 2024
Organization: ICNC 2024

For contribution to the IEEE Communications Society through the continued submission and presentation of papers at ICNC.

Young Researcher's Award
Winner: Shuhei Yoshida, NTT Device Innovation Center
Date: March 5, 2024
Organization: The Institute of Electronics, Information and Communication Engineers (IEICE)

For "Pose Estimation Technology Based on AI Inference for High-definition Image."
Published as: S. Yoshida, H. Uzawa, Y. Iinuma, S. Hatta, Y. Omori, Y. Horishita, D. Kobayashi, K. Nakamura, and K. Yamazaki, "Pose Estimation Technology Based on AI Inference for High-definition Image," Proc. of the 2023 IEICE General Conference, D-12-24, Omiya, Saitama, Japan, Mar. 2023.

Young Researcher's Award
Winner: Yasunori Yagi, NTT Network Innovation Laboratories
Date: March 5, 2024
Organization: IEICE

For "Experimental Evaluation on Sub-THz OAM Multiplexing Transmission with Scaling Reflector Antenna."
Published as: Y. Yagi, H. Sasaki, D. Lee, and R. Kudo, "Experimental Evaluation on Sub-THz OAM Multiplexing Transmission with Scaling Reflector Antenna," Proc. of the 2023 IEICE Society Conference, B-17-3, Nagoya, Aichi, Japan, Sept. 2023.

Telecom System Technology Award
Winner: Motoharu Sasaki, NTT Access Network Service Systems Laboratories
Date: March 21, 2024
Organization: The Telecommunications Advancement Foundation

For "Extension of ITU-R Site-General Path Loss Model in Urban Areas Based on Measurements from 2 to 66 GHz Bands."
Published as: M. Sasaki, M. Nakamura, N. Kuno, W. Yamada, N. Kita, T. Onizawa, Y. Takatori, H. Nakamura, M. Inomata, K. Kitao, and T. Imai, "Extension of ITU-R Site-General Path Loss Model in Urban Areas Based on Measurements from 2 to 66 GHz Bands," IEICE Transactions on Communications, Vol. E104.B, No. 7, pp. 849–857, 2021.