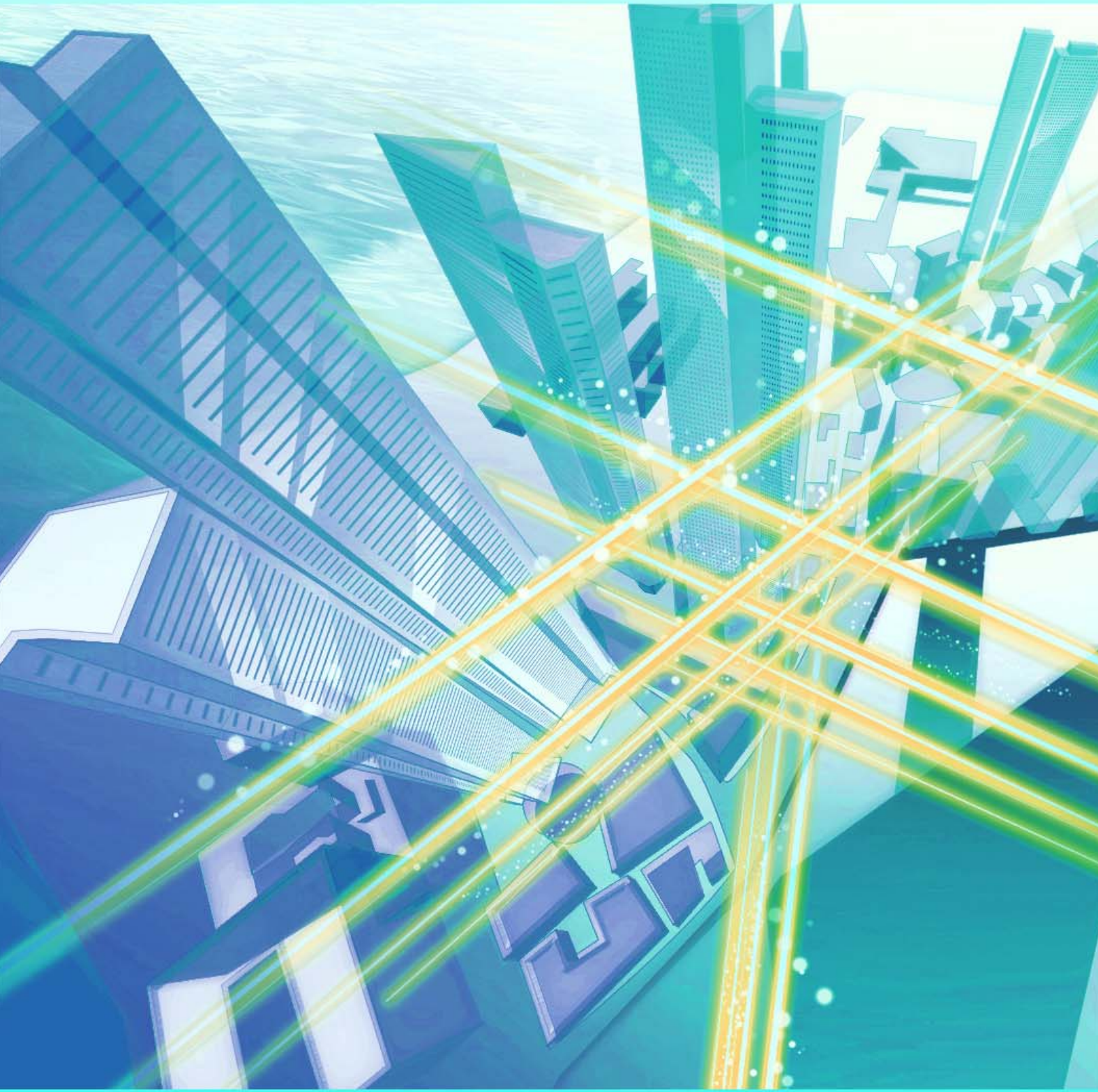
NTT Technical Review

5

2021



May 2021 Vol. 19 No. 5

NTT Technical Review

May 2021 Vol. 19 No. 5

Enthusiasm and Motivation Come First, Strategies and Tactics Follow—Addressing Social Issues in the Medical and Healthcare Field by Using Big Data

Koji Korekawa
President, NTT Life Science and Prime Research Institute for Medical RWD

Overview

In Japan, the average life expectancy and healthy life expectancy differ by 9 to 12 years. Since this gap is not only a social issue but also affects individual well-being, it is necessary to raise awareness of public health and promote the prevention, early diagnosis, and treatment of diseases. To address these social issues through the use of health and medical big data, the NTT Group has launched two healthcare companies; NTT Life Science and Prime Research Institute for Medical RWD. We interviewed Koji Korekawa, president of both companies about the services provided by these companies and his mindset as a top executive.

Keywords: healthcare, bio-digital twin, electronic medical record

Provide solutions that use big data related to medical and healthcare by using the NTT Group's technology

—You have become president of two companies, NTT Life Science and Prime Research Institute for Medical RWD. Could you tell us about the establishment and mission of these companies?

The two companies will develop platform businesses that use big data in the medical and healthcare field. NTT Life Science is a wholly owned subsidiary of NTT, and Prime Research Institute for Medical RWD (PRiME-R) is a joint venture between Kyoto University and NTT.

Social security expenditures account for 30% of Japan's national general-account budget for fiscal year 2021. Projections of the working-age population and medical benefits suggest that social security expenditures are expected to continue to increase as the birthrate declines and the population ages. The national medical expenditure totaled 42.3 trillion yen in fiscal year 2016, of which general medical treatment fees were 30.2 trillion yen (lifestyle-related diseases accounted for 10.4 trillion yen, or 35%). Moreover, life expectancy is 80.98 for men and 87.14 for women, but healthy life expectancy is 9 to 12 years lower than life expectancy.

These figures clearly show that Japan is facing social issues of declining birthrate and aging population, increase in medical expenses, decrease in working-age population, increase in lifestyle-related diseases, and the gap between life expectancy and healthy life expectancy. To address these issues, it is important to raise individual health awareness, extend healthy life expectancy through the prevention, early diagnosis, and treatment of diseases, and alleviate concerns about medical and nursing care costs in old age. Companies are also required to promote corporate health management, which is the strategic implementation of health management for employees from a management perspective to improve productivity and reduce medical costs.

However, it can be inferred from the aforementioned figures that few people fully understand their own health risks and think of them as their own problems. Even if they undergo annual health checkups, few people continuously take care of their health after receiving the results of those checkups. Although companies know the importance of corporate health management, they have not devised any concrete measures, and the data acquired from annual health checkups are not fully used. To rectify this situation, we will provide solutions that use big data related to medical and healthcare by using the NTT Group's information and communication technology (ICT).

—Tell us in detail the significance, value, and prospects of the entry of the NTT Group into the medical and healthcare field.

The strength of the NTT Group is its ICT, and the group is able to create new value by making use of a variety of data. For example, the NTT Group has adopted the slogan "Your Value Partner" as its group vision, and together with its partners, the group aims to address social issues through its business activities. By accumulating and applying a wide variety of data by leveraging ICT to improve existing systems and build and introduce new systems, technologies, and services, the NTT Group aims to create a Smart World within which we can solve the numerous problems facing society and create a better living environment. NTT's Innovative Optical and Wireless Network (IOWN) initiative aims to build an innovative network and information-processing infrastructure that overcomes the limitations of conventional infrastructures.

As one of those initiatives in fields related to healthcare, the NTT Group in November 2020 announced its Medical and Health Vision to contribute to the future of medical care through which people will become healthier and have hope for the future. Digital Twin Computing (DTC), which is a component of IOWN, is a key technology of making this vision a reality. By applying DTC, each person's mind and body can be precisely mapped (i.e., a bio-digital twin (BDT)), and the BDT is used to predict the future state of people's mind and body. To implement this vision and develop medical and

healthcare-related business, NTT established two strategic subsidiaries, which are our companies.

I want to meet the enthusiasm and expectations of people for creating a new market

—Tell us about the specific initiatives of the two companies.

From the perspective of preventive medicine, NTT Life Science will provide services to people who have not yet become ill, while PRiME-R will provide services to those who are outpatients and entered the treatment phase.

NTT Life Science is working on "Genovision," which is a service that supports corporate health management by proposing preventive measures and healthy behaviors by cross-checking genetic data from genetic tests against annual health-checkup data (test and medical-consultation results) acquired from companies. It is widely recognized that environmental factors such as lifestyles have a greater effect on the developing of many diseases than genetic factors. Accordingly, by understanding the physical constitution and the risk of diseases developing in the future on the basis of the genetic information, which is often referred to as the blueprint of the human body, we hope to be able to encourage people to review their lifestyles.

PRiME-R provides a service using CyberOncology™, which is an input support system for using various types of information (real world data (RWD)) in clinical settings such as handling electronic medical records (EMRs). The context behind this service is the current situation in Japan, where clinical information, such as EMRs, is not being fully used. RWD, including EMRs, is attracting attention for its use in medical research, clinical research, and development of pharmaceuticals and medical devices as well as daily diagnostic support for doctors. Current EMRs, however, are created as simple electronic recording media, and since a large amount of data is described and recorded freely by doctors on such media without being systematized, it is difficult to use such data for data analytics.

With this situation in mind, we built a system that supports the input of data to EMRs. The system uses a pull-down-menu format that allows users to select and enter standardized medical information from a list and store the input data in a systematized manner. The introduction of this system will also reduce labor for tasks such as registration of clinical information for the Center for Cancer Genomic and Advanced Therapeutics (C-CAT). We are also planning to launch a service to provide statistical data to medical institutions, pharmaceutical companies, and other parties by statistically processing the data accumulated in the system in a manner that individual patients cannot be identified and personal information can be protected.

All these initiatives are based on unprecedented ideas, and I believe that they will lead to a reduction in disparities in medical care provided, reduction in development costs for new drugs, deterrence of medical-system collapse, and development of new treatment methods.

—Is everything going well?

NTT Life Science began providing the Genovision service in April 2020, and PRiME-R started full-scale operations at the same time, so we have just reached the one-year mark. I am grateful that 20,000 people have agreed to take the Genovision test thus far, and I hope to make 2021 a year of great progress. PRiME-R has started several joint research projects with a pharmaceutical company and medical institutions, and we hope to establish a mechanism for collecting new data.

The start of the two companies and the novel-coronavirus (COVID-19) pandemic coincided, so the employees of both companies have not been able to

meet their colleagues face to face over the past year. Although both companies are still in the early stages in terms of sales, many people have shown much interest in our services. To prevent the spread of COVID-19, we are unable to visit the medical institutions, which are the customers of PRiME-R, so we are talking to them online. It is very difficult for doctors, who are users of our system, to coordinate their time with us; even so, we have been able to be in contact with doctors from all over Japan in a timely manner by taking advantage of online communications. It was a difficult situation, but it was also good for us because of the social climate in which online sales are accepted.

Enthusiasm and responsibility are required of top executives

—How do you feel as a top executive involved in various new initiatives at the new company?

I was previously working at the Research and Planning Department at NTT, where I was involved in the development of medical and healthcare businesses using the research results of NTT laboratories. At that time, I shared my awareness of issues in the medical field with executives of NTT Group companies, which was a starting point from which I could begin working toward addressing social issues in the medical field. One year has passed since we established these companies, and we are just beginning to establish their foundations; however, this is a new field, so we are confronted with various problems daily. By making decisions and establishing rules to deal with these unfolding problems, we are making steady progress toward our companies' goals, and I feel the weight of that responsibility very strongly.

Because both companies are small, we can share awareness of problems with our employees in detail and approach them with a sense of unity. However, regardless of the size of the companies and one's career, a top executive has a responsibility to make decisions, and I think that a vision is necessary for the success of the business. I want to develop the two companies into those that will become the foundation of the future medical and healthcare fields. I am working hard every day to meet the enthusiasm and expectations of people inside and outside the NTT Group to create new markets.

—How do you make daily judgments and decisions?

Basically, I make decisions by taking into account the effects and impact of the vision and goals we are pursuing. Since some initiatives will inevitably fail, and our business field is new, it is sometimes difficult to know what is a success and what is a failure. With this situation in mind, I place great importance on maintaining an unshakable attitude toward our vision and goals and having conversations to share our vision. The first partners in those conversations are medical professionals. We are professionals in ICT

but amateurs in the medical field, so it is essential for us to keep abreast of the most-advanced medical care, keep in touch with top doctors, and communicate with other professionals in the field. I am constantly striving to gain knowledge through dialogues with the professors from the University of Tokyo and Kyoto University with whom we are collaborating and to understand the style and culture of the field. There are many regulations in the medical field, and there are values that non-experts do not understand. The higher the hurdle, the more interesting it is to challenge, and the more rewarding it is to overcome.

I believe that it is impossible to move business development or anything forward without enthusiasm and motivation, and strategies and tactics should follow them. In the 27 years since I joined NTT, I have learned the importance of being independent and taking responsibility while doing various jobs in different fields. It was not until I was engaged in business development at NTT's Research and Development Planning Department that I became strongly motivated to address social issues in the medical and healthcare field. After I developed my own vision in this field, I started a challenge to fulfill the vision, and the establishment of the two companies has paved the way for that challenge.

Meetings with our collaborating professors are refreshing and always inspiring, and preparing for those meetings is a pleasure. Discussing solutions to social problems with executives from companies, doctors from medical institutions, and people from consulting companies is valuable for me, and I am grateful for their support and enjoying working with them. Both companies have just started to move from the starting line, and we will aim for our goals while further enhancing our resources, such as knowledge, experience, services, and personnel. I am not in a position to say anything outstanding because I have no track record yet; all the same, I will keep in mind my responsibility and the speediness of decisions while not forgetting the feeling of gratitude.

—Please give a few words to our researchers and engineers.

I want you to engage in research and development that benefits society. Our business includes data analytics, utilization, and collection as well as the trending fields of security, artificial intelligence, and data science. Without strengthening our research and technologies in each of these fields, we will not be able to win in the global market. Therefore, take the challenge of pursuing cutting-edge technology with enthusiasm and be confident in your research theme. We want to promptly implement your achievements in society, so let's work hard together toward that goal.

Interviewee profile

■ Career highlights

Koji Korekawa joined NTT in 1994. In his career, he joined the Personnel Department of NTT WEST in 2001, the First Division (currently, Corporate Strategy Planning Department) of NTT in 2004, and the Planning Department of the Corporate Business Headquarters of NTT WEST in 2007. He became general manager of Chugoku Regional Headquarters of NTT WEST in 2010, president of NTT Business Solutions MCS in 2011, chief producer of the Medical and Healthcare Group of the Research and Development Planning Department of NTT in 2014, section chief of the Medical Business Planning Office of the General Affairs Department of NTT in 2019 and president of NTT Life Science in 2019, and president of Prime Research Institute for Medical RWD in 2020.

Be Insensitive to Peer Pressure to Fight a Fierce Battle of Ideas

Shin'ya Nishida
Senior Distinguished Scientist, NTT Communication Science Laboratories

Overview

Many mechanisms by which the human brain recognizes the various complex properties of objects in the real world remain unknown. Understanding these mechanisms is critical for scientific understanding of human-sensory-information processing and advancements in information-engineering technology. Researchers at NTT Communication Science Laboratories are leading research on the human perception of specific properties of objects called *Shitsukan* (Japanese word for the sense of quality) from interdisciplinary perspectives, such as information science, neuroscience, and psychophysics, in conjunction with researchers from inside and outside the laboratories. We asked Shin'ya Nishida, a senior distinguished scientist at NTT Communication Science Laboratories, about his pioneering research on Shitsukan and attitude as a researcher.

Keywords: visual processing, Shitsukan, interdisciplinary research

Leading the world in research on Shitsukan perception

—Please tell us about the research you are currently working on.

I'm researching the human perception of specific properties of objects called *Shitsukan* (Japanese word for the sense of quality). Humans recognize various Shitsukan through their five senses, including vision, and instantly judge physical properties, material properties, conditions, and subjective value. Their ability to recognize Shitsukan plays an important role in human activity because humans are involved in the environment through recognizing and evaluating objects, making decisions, and controlling their body actions, all of which are deeply connected to Shitsukan.

The Shitsukan that I'm studying can be roughly divided into two types: (i) physical Shitsukan, such as physical reflectance properties (e.g., glossiness and transparency), materials (e.g., ceramics, metals), and conditions (e.g., dry and frozen), and (ii) subjective Shitsukan, such as beauty and preference. I believe that Shitsukan is the ability of the human brain to decode the nature of an object from sensory input. Many mechanisms with which the human brain recognizes the diverse Shitsukan of objects in the real world remain unknown. Explaining these mechanisms is critical not only for scientific understanding of human-sensory-information processing but also advancements in information-engineering technology

Fig. 1. Transforming image texture to make its surface look wet [3].

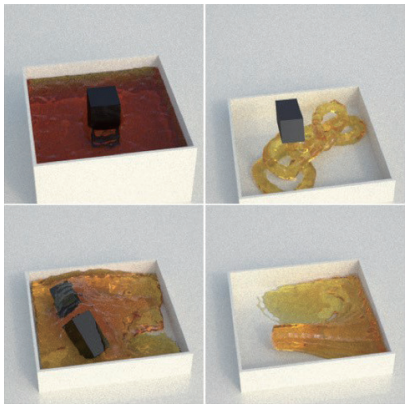


Fig. 2. Computer-simulated images used to analyze the perception of liquid Shitsukan (viscosity) [4].

for recognizing and generating the Shitsukan of real objects.

I'm one of pioneers of Shitsukan research. We started studying human material perception in the mid-1990s and made a few important contributions, including discovery of low-level image features of surface gloss, by 2010. Since then, using a joint-research grant called "Grant-in-Aid for Scientific Research on Innovative Areas" from the Ministry of Education, Culture, Sports, Science and Technology (MEXT), we have been promoting world-leading interdisciplinary research on Shitsukan. From fiscal year 2010 to 2014, I headed a research team on the theme "Human Shitsukan perception from visual, auditory and tactile information" in the interdisciplinary area research project "Brain and Information Science on Shitsukan (BISS) (Integrative studies of neural mechanisms and advanced information technologies for perception of material and surface qualities)" [1]. During that period, we studied color-luminance interaction related to gloss perception and clarified the mechanism for recognizing the Shitsukan (viscosity) of a liquid on the basis of motion information. By exploiting these human mechanisms, we developed a light-projection-mapping technology called *Hengento*, which creates the illusion that a still object appears to be moving.

From fiscal year 2015 to 2019, I headed the second interdisciplinary area research project "Innovative Shitsukan Science Technology (ISST) (Understanding human recognition of material properties for innovation in Shitsukan science and technology)" [2], and we pursued Shitsukan research from the perspectives of information science, neuroscience, and psychophysics and established the academic field of Shitsukan science and technology, which ranges from basic research to applied research in the fields of vision, touch, hearing, and language.

As a principal investigator of the research team "Visual, auditory and tactile Shitsukan recognition mechanisms based on signal modulations" within the above project, our team clarified the perception of (i) wetness on the surface of an object on the basis of luminance and color statistics [3] (**Fig. 1**) and (ii) ultrafine structures on the basis of a reduction in image contrast. At the same time, we analyzed the mechanism of perception of liquid texture by using an artificial neural network [4] (**Fig. 2**).

Extending the principle of Hengento, we developed a method of synthesizing ghost-free stereoscopic images called *Hidden Stereo* that does not cause blur when an image is viewed without stereo glasses [5]. The field of Shitsukan science and technology that we established through these two interdisciplinary area research projects (BISS and ISST) has been highly evaluated as internationally outstanding in terms of interdisciplinarity and breadth of vision.

—You have established a new academic field concerning Shitsukan. What is your research going forward?

I'm interested in human information processing, in particular, visual information processing. When considering politics and society, for example, I'm also interested in the reasons behind the actions of people. I'm always trying to understand humans not only using established disciplines such as psychology but

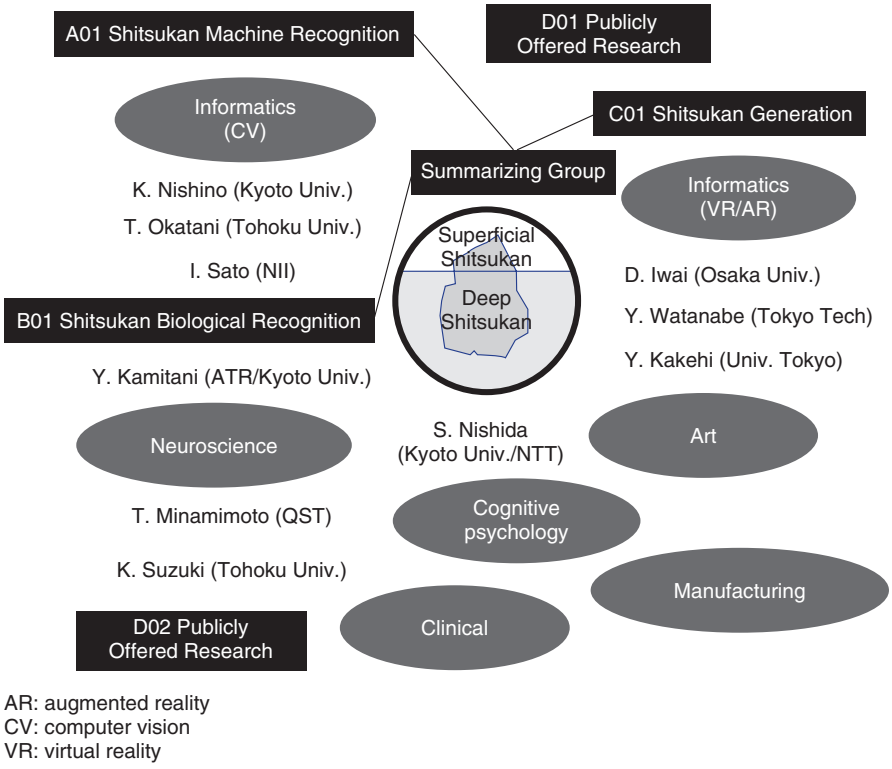


Fig. 3. Deep Shitsukan research project.

also looking for novel approaches that make sense to me.

In 2020, we started the third Shitsukan project, Transformative Research Areas research project "Deep Shitsukan (Analysis and synthesis of Deep Shitsukan information in the real world)" [6]. I again acted as head of the project. The key concept of the project is that to understand the essence of Shitsukan, it is necessary to understand (i) the superficial Shitsukan information processing that links the input information captured by the sensory organs to Shitsukan-attribute variables and linguistic labels of Shitsukan categories and (ii) the processing hierarchy that lies behind superficial Shitsukan, which we call *Deep Shitsukan*.

We currently assume four types of Deep-Shitsukan processes. The first process is calculating the multifaceted ecological meaning and value of things and events from Shitsukan information and often inducing an emotional response inside the body. The second process is predicting the outcome of behaviors and selecting the appropriate behavior by constructing an external model in the brain by integrating Shitsukan and other sensory attributes. The third process is determining how Shitsukan-information processing is influenced by the characteristics of the person doing the processing (for example, age, brain dysfunction, cultural background, and experience). The fourth process is discriminating between real and fake through the sensory information on real objects acquired from the five senses.

In this project, we aim to elucidate human Deep-Shitsukan processing from the perspective of cognitive neuroscience and develop innovative Deep-Shitsukan technology from the perspective of engineering. We will reveal the essence of sensory information that allows people to experience realistic Deep Shitsukan. We will also develop machine-vision technology for recognizing Deep Shitsukan and media technology for manipulating it at will, with which we expect to connect Shitsukan science to art. Our project consists of ten teams, each belonging to one of three research themes (**Fig. 3**). As the leader of one of the teams, I'll focus on a unified understanding of Deep-Shitsukan recognition from multiple angles, namely, vision, hearing, touch, and linguistic information, using psychophysics and sensory engineering.

A paradigm shift occurred as a result of machines catching up with humans

—We talked to you a while back in 2012. How has your research environment changed during that time?

Around 2012, human capabilities were still ahead of machines in a variety of intelligence-related tasks. However, over the past eight years, as artificial intelligence (AI) and machine learning advanced significantly, machines have been able to surpass humans, albeit to a limited extent, in terms of intellectual capabilities. This paradigm shift has had a considerable impact on me as a researcher of humans. By understanding the complex mechanisms of human cognition and behavior, we have been trying to obtain clues for developing machines with abilities comparable to humans. However, the exact opposite approach is now becoming commonplace. That is, attempting to understand human information processing by analyzing machines that have acquired human abilities by machine learning. This research revolution is taking place in neuroscience, and I want to fully use this new approach for advancing our research on Deep Shitsukan.

As these technologies and research are rapidly advancing, I'm studying new technologies with young researchers. A huge number of papers are published on the Internet, and it is difficult to choose which to read. I have a lot of work other than research on my hands, so it's not easy to find time to catch up on these advances. Although I'm excited about the new waves of discoveries that are coming one after another, I feel it difficult to stay on the cutting edge at all times. Therefore, I am trying to create my originality by fusing the newly acquired knowledge with the experience and knowledge I have cultivated thus far. I saw in the past that the basis of research on human senses had shifted from psychology to imaging of brain functions using functional magnetic resonance imaging, and I now see the field is moving toward fusion with AI research. I believe that having such a long research history will be an advantage over young researchers who only know the latest state-of-the-art approaches.

I think these are the fastest changes in research trends I've ever seen. I've heard from doctoral students who research machine learning and AI that the research they were working on in their first year had become outdated by their third year and no longer worth publishing. Today, they have rivals around the world, and when they publish their research results, another researcher will publish results superseding theirs within a year. They say that researchers' minds cannot rest. I think that researchers in AI are living in an exciting time but face fierce battles and a tough time mentally.

—What is the driving force behind consistent research activities in research areas that are showing tremendous development?

That's a difficult question. I tend to jump on anything new. Even though it may seem like I am sticking to one topic, my interests are constantly changing and evolving. This is my research attitude. People tend to paint a picture of researchers as pursuing a single theme, but I think I'm a little different in that I'm consistent with a sense of wanting to do something interesting.

Since the subject of my research, i.e., Shitsukan, is a very difficult problem, I have launched several projects on this topic to force me to work on it, so it may appear that I am sticking to one theme. However, I always enjoy listening to new ideas from many researchers participating in the project. Some ideas are very stimulating and interesting enough for me to change my research.

I don't accept popular research trends, even if it's a subject of interest, unless I really believe they are right. On the other hand, if I find something interesting, I want to adopt it even if it's new and unfamiliar to me. Although my interests vary, there are several reasons that I have remained a researcher. The first is because researchers don't take orders from others. Since my youth, I've been reluctant to follow instructions from others. I'm grateful to my bosses for allowing me to keep doing so. Fortunately for me, NTT Communication Science Laboratories has been a company that respects researchers and lets them do what they want to do. I think that if such an environment disappears, it would become less attractive to researchers.

Young researchers, your time has come!

—Researchers also need to select a research environment to gain that freedom.

I think that success in research results from a series of coincidences. To get the most out of such coincidences, it is important to be able to move freely. As long as you receive research funds, you need to make plans and conduct research in accordance with the

plan; however, as many researchers including Nobel laureates have said, it is important to leave such planning to researchers to a certain extent, give them more freedom and encourage them even if only 1% of them succeed and 99% fail. If we don't do that, we won't be able to open up routes to the future, and those failures, or "waste," are necessary to open the future. It is critical to be able to tolerate this waste to achieve good research results.

NTT has the strength to ensure such freedom. I am also teaching at a university. University faculty members have a tough time obtaining research funding, but NTT researchers do not have such a problem. Researchers may have unusual personalities, even a bit weird. That's why they can create something different. If that is the case, I think we can open up the path to the future by creating an environment that allows researchers to pursue research freely.

Such a free research environment is becoming ever less common in Japan, and the number of researchers is decreasing as well. Young researchers are having a difficult time trying to obtain one of the few research posts available. In the history of research activities in Japan, corporate laboratories have supported not only applied research but also basic research. Among those laboratories, NTT, with its mission and culture inherited from when it was a public corporation, still has an environment in which researchers can take their time and focus on basic research. I think the role played by NTT laboratories in research activities in Japan is significant.

I am also a member of the Science Council of Japan (Section I: Humanities and Social Sciences), and from that standpoint, I think that Japan should take a strategic approach to nurture the young generation of researchers that will create value in the future. In that regard, NTT has the ability to recruit and train many outstanding young researchers and give them the freedom and discretion to maintain their motivation. I hope that NTT will continue to demonstrate its power and contribute to the future of Japan.

—Please give a few words to our young researchers or students who want to become researchers.

As a humanities graduate with a science background, a corporate researcher, and a university faculty member, I have gained experience in looking at and thinking about things from multiple perspectives. I want to make use of this advantage to train young researchers while actively speaking my mind about the development of human resources at various opportunities.

I think that since the standpoints of young researchers at NTT laboratories and students differ, the method of guidance should also differ. Many NTT's researchers have earned their doctorate degrees and are already equipped and ready to be researchers, so I encourage them to improve themselves through research. On the contrary, students have not yet decided whether they will become researchers. Accordingly, I do not strongly encourage them to become researchers; instead, I give them research guidance with the hope that it will be useful to them in some way.

With that in mind, my message to young researchers and students is to become insensitive to peer pressure. If we consider the trend of researchers starting their own businesses, as in the USA, as one effect of peer pressure, it may be that peer pressure works well to raise the standard of research. In Japan, however, children are brought up to "be good girls or boys" or "act like everyone else" until they go to university, and I think a strong sense of self-regulation has developed as a result. I want you to discard that self-regulation and become insensitive to such peer pressure in regard to your research activities. In other words, it might be a good idea for you—as a researcher—to dare to be "the nail that sticks out."

It is an exciting time when the presentations of papers at international conferences on cutting-edge research are immediately published on the Internet, and new methods and technologies come out one after another. In such an era, researchers, especially those in the information field, may find it difficult to reap the benefits of belonging to a large research institute. Researchers in such fields do not necessarily require large-scale experimental equipment and technology in large laboratories for their research, and it has become possible for researchers outside major institutes to enter the ring of competition if they have an idea and a suitable computing environment. In other words, you are living in a time when you have to fight a fierce battle of ideas day and night. New ideas and directions can only come from research sites that are in constant contact with cutting-edge research. You can't expect your supervisors, who are busy with common tasks, to give you guidance. Therefore, young researchers, your time has come!

References

[1] Website of the Brain and Information Science on Shitsukan (material

perception) project, http://shitsukan.jp/BISS/content/137
[2] Website of the Innovative Shitsukan Science Technology project, http://shitsukan.jp/ISST/en/outline/index.html
[3] M. Sawayama, E. H. Adelson, and S. Nishida, "Visual Wetness Perception Based on Image Color Statistics," Journal of Vision, Vol. 17, No. 5, pp. 7–24, 2017.
[4] J. J. R. van Assen, S. Nishida, and R. W. Fleming, "Visual Perception of Liquids: Insights from Deep Neural Networks," PLoS Computational Biology, Vol. 16, No. 8, e1008018-29, 2020.
[5] T. Fukiage, T. Kawabe, and S. Nishida, "Synthesizing Ghost-free Stereoscopic Images for Viewers without 3D Glasses," NTT Technical Review, Vol. 15, No. 11, 2017. https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201711fa5.html
[6] Website of the Deep Shitsukan project (in Japanese), http://shitsukan.jp/deep/

■ **Interviewee profile**
Shin'ya Nishida

Senior Distinguished Scientist, Research Professor of Sensory Representation Group, Human Information Science Laboratory, NTT Communication Science Laboratories.

He received a B.S., M.S., and Ph.D. in psychology from Kyoto University in 1985, 1987, and 1996 and joined NTT in 1992. He is an expert in the psychophysical research on human visual processing, in particular, motion perception, cross-attribute/modality integration, time perception, and material perception. He served as president of the Vision Society of Japan and was an editorial board member of the Journal of Vision and Vision Research.

Activities toward New-principle Computers

Shiro Saito and Hideki Gotoh

Abstract

Non-von Neumann-type computers, such as quantum computers and Ising machines that operate on principles different from those of present-day computers, are attracting attention. They are particularly powerful when applied to specific types of problems such as combinatorial optimization problems, quantum chemical calculations, and prime factorization, and since solutions to problems such as these can have a significant impact on society, research of new-principle computers is moving forward at a vigorous pace. The following Feature Articles in this issue introduce theoretical and experimental activities regarding new-principle computers at NTT laboratories while describing recent advances in this field.

Keywords: quantum computer, coherent Ising machine, quantum annealer

1. New-principle computers

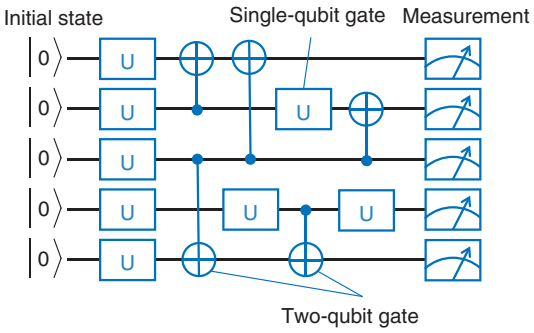
In today's semiconductor industry, there is an empirical rule called Moore's law that states, "the number of transistors on an integrated circuit doubles every 18 months." In truth, the miniaturization of semiconductor devices has been progressing in accordance with this law, and computer performance has been increasing yearly. If this empirical rule continues to hold as it has up to now, transistor size will become extremely small and approach its limit at the size of atoms. Before that, however, it is said that quantum effects will appear and that laws applicable to conventional electronic circuits will no longer hold constituting a *quantum mechanics barrier*. In the world of quantum mechanics, there are mysterious properties, such as the superposition state and wave-particle duality, not seen in classical mechanics. A computer that actively uses these properties and that operates on principles different from those of conventional computers is the quantum computer. In current information processing, calculations proceed using bits that can take on either of two values, 0 *or* 1. In contrast, a quantum bit (qubit), the basic device of a quantum computer, can achieve a superposition state of both 0 *and* 1. Given N qubits, it will be possible to achieve a superposition of 2^N states, which means that increasing the number of qubits will increase the number of superposition states in an exponential manner. This holds the possibility of massively parallel computation by executing computer operations against this huge number of superposition states. It is also a factor affecting the speed of quantum computers. A quantum computer, moreover, allows the wave property of a qubit to be used so that the desired solution can be found from the results of massively parallel computation in accordance with the phase-interference effects between qubits. In contrast to von Neumann-type computers based on conventional sequential computation, non-von Neumann-type computers, such as the quantum computer, that operate on new principles, have been recently attracting attention. The following Feature Articles in this issue introduce activities related to new-principle computers at NTT laboratories from both theoretical and experimental perspectives.

New-principle computers can be broadly divided into two types: gate-based quantum computers and quantum annealers. The former prepares multiple qubits in their ground states and executes calculations by repeating single-qubit-gate operations and two-qubit-gate operations, as shown in **Fig. 1(a)**. Quantum algorithms for prime factorization, large-scale searching, etc. have been discovered, and it has been

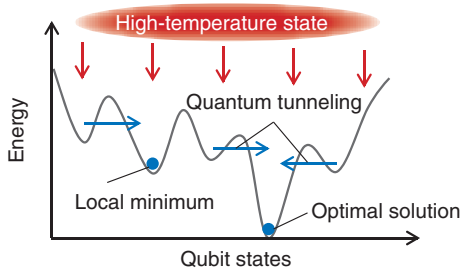
theoretically proven that they are faster than classical algorithms. However, qubits are fragile in the face of external noise and errors are frequent, so redundant qubits for error-correction purposes, i.e., a massive number of qubits, are needed. For example, a theoretical estimation has shown that 20,000,000 qubits would be needed to execute prime factorization of a 2048-bit number [1]. It is extremely difficult to manufacture qubits on such a scale with current technology. The latter type, i.e., quantum annealers, superposes all states of multiple qubits, in other words, prepares high-temperature states, and finds the state with the lowest energy (solution) by gradual cooling, as shown in **Fig. 1(b)**. It is said that the qubit states can escape from a local minima during this process due to a quantum tunnel effect, therefore reach a solution faster than classical annealing algorithms. Achieving higher processing speeds through quantum properties has not been rigorously proven, but approximate solutions to combinatorial optimization problems greatly needed by society can be obtained, so quantum annealers have been attracting attention. They execute calculations while slowly changing stable states, which makes them robust to noise. When implementing an optimization problem in an actual physical system with a quantum annealer, a physical model that expresses a mutually interactive spin system called an Ising model is used, so this type of computer is also called an *Ising machine*.

2. Ising machines

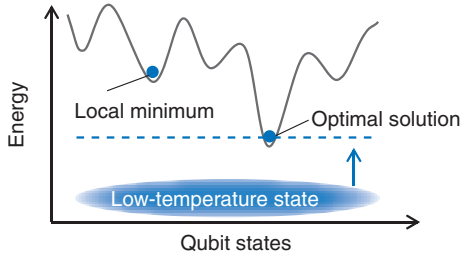
In the research on gate-based superconducting quantum computers, achievements such as observation of an entangled state among three qubits and the demonstration of double-qubit-gate operation were reported in 2010 and 2011, respectively. In 2011, however, exciting news came out of the Canadian venture company D-Wave Systems. They announced the world's first commercial quantum computer in the form of a quantum annealer (D-Wave One) implementing 128 qubits. This was an Ising machine that executed calculations using superconducting qubits. Compared with the aluminum qubits used in gate-based superconducting quantum computers, the D-Wave One qubits using superconducting niobium have an overwhelmingly shorter storage time for quantum information, so at the time, there was some doubt as to the quantum properties of this computer. However, subsequent demonstration experiments confirmed that the high-speed performance of the computer was due to quantum properties. Since then,



(a) Gate operations on a quantum computer



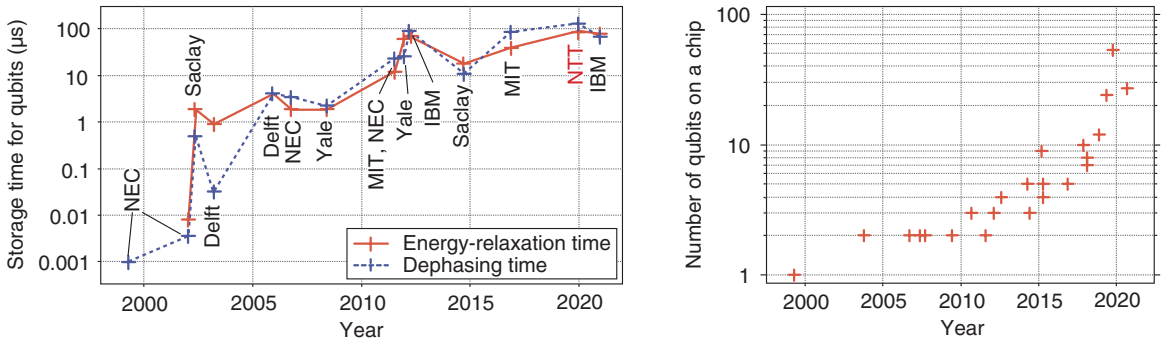
(b) Operating principle of a quantum annealer



(c) Operating principle of a coherent Ising machine

Fig. 1. New-principle computers.

the 512-qubit D-Wave Two was announced in 2013, 1000+ qubit D-Wave 2X in 2015, 2048-qubit D-Wave 2000Q in 2017, and 5000+ qubit D-Wave Advantage in 2020. Companies around the world are now using cloud services offered by D-Wave Systems, and research and development on the business use of quantum Ising machines is progressing. However, for actual business applications, Ising machines of even larger scale are needed. Furthermore, since superconducting qubits are solid-state devices on a chip, there is the constraint that coupling between qubits is limited to neighboring qubits. Consequently, when implementing an optimization problem in a quantum



(a) Storage time for superconducting qubits

(b) Integration of superconducting qubits

Fig. 2. Development of superconducting qubits.

annealer, many qubits beyond the size of the problem are needed.

NTT laboratories developed a coherent Ising machine (CIM) that can implement all-to-all spin couplings by using degenerate optical parametric oscillators (DOPOs) as artificial spins and confining multiple DOPO pulses within a fiber ring cavity. This scheme makes it possible to increase the number of artificial spins by extending the total length of the fiber ring cavity, so scaling up is relatively easy compared with solid-state devices. A CIM searches for a solution from the low-temperature side, as shown in **Fig. 1(c)**, by gradually increasing the intensity of DOPO pumping light. The solution is found when system energy reaches its optimal solution. Studies are still being conducted on how the quantum properties of DOPO contribute to computational performance, but in a classical combinatorial optimization problem called the maximum-cut (MAX-CUT) problem, it was shown that a CIM is faster than a computer using a classical annealing algorithm [2]. The Feature Article "Performance Comparison between Coherent Ising Machines and Quantum Annealer" [3] introduces a comparison experiment between CIMs and the quantum annealer from D-Wave Systems.

3. Gate-based quantum computers

How best to achieve a qubit—the basic device of a quantum computer—has been a problem for the last 20 years. A quantum two-level system is needed to achieve a 0 and 1 superposition state in a qubit, but what kind of physical system should be used for achieving this has been under discussion. Atoms or electrons for which the storage time of quantum information is long are difficult to integrate, but semiconductor or superconducting qubits, though having a short storage time, are solid-state devices, so semiconductor integration technology can be used. Optical qubits, meanwhile, have good affinity with optical communications, and a variety of candidates have been studied. This discussion continues to this day, and each of these physical systems has been developed while expanding its strengths. Superconducting quantum computers, ion-trap quantum computers, and semiconductor-dot quantum computers have achieved a level of integration of 50–70 qubits, 32 qubits, and 2–3 qubits, respectively. In particular, the announcement by Google in 2019 that the superiority of their 53-qubit superconducting quantum computer over existing supercomputers was a "proof of quantum supremacy" became a major topic [4]. Setting quantum computers to solve problems deemed to be their specialty does not prove that they are superior in all problems; however, this achievement was nevertheless a significant milestone. The development of superconducting qubits toward this milestone is shown in **Fig. 2**. In 1999 when operation of a superconducting qubit was demonstrated for the first time, storage time was short at 1 ns, making for extremely difficult experiments (Fig. 2(a)). Around 2012, however, this was extended to approximately 100 μs, an improvement of about five orders of magnitude. There has been no lengthening of storage time on that scale since then, but the number of qubits on a chip has increased dramatically (Fig. 2(b)). The reason for this is thought to be that the minimally required storage time toward higher integration had been reached at that time; as a result, the direction of research at many research institutions shifted from extending

storage time to increasing integration. Therefore, Google, IBM, and Intel developed 50-qubit superconducting quantum-computer chips that reflect an evolution to the point of quantum supremacy.

At NTT laboratories, research has been focused on a type of superconducting qubit different from the qubits integrated on current superconducting quantum computers. Making use of this difference, this research has involved a variety of activities from basic physical research, such as testing of macroscopic quantum superposition states [5], to applied research with an eye to local high-sensitivity magnetic field sensors [6]. The Feature Article "A Long-lived Tunable Qubit for Bosonic Quantum Computing" [7] introduces NTT's achievement of recording the world's longest storage time with this superconducting qubit (Fig. 2(a)) and discusses the possibility of applying this qubit to quantum computers.

4. Toward fault-tolerant quantum computers

On hearing the news that quantum supremacy had been demonstrated, one might think that the development of a practical quantum computer was soon at hand. However, the path to a working quantum computer is long. Current general-purpose quantum computers have a level of integration of about 50 qubits. In addition, quantum information is lost over time due to noise effects and gate operation is accompanied by errors, so the execution of complex calculations involving numerous repetitions of gate operations cannot be executed. Therefore, a quantum computer that is limited in terms of functions and scale is called a noisy intermediate-scale quantum (NISQ) device. This type of device, despite its limitations, can also exhibit quantum supremacy, so it can teach us much about the potential of quantum computers. However, executing complex quantum calculations on a practical scale will require fault-tolerant quantum computers. To meet this requirement, the approach is to configure a single qubit (logical qubit) that is error correctable and robust to noise by preparing multiple qubits (physical qubits) to provide redundancy. It will also be necessary to prepare a complex layered structure consisting of a layer for controlling physical qubits, one for controlling the logical qubit, one for executing algorithms above those layers, etc. The Feature Article "Designing Quantum Computers" [8] introduces the results of a theoretical study on what type of layered structure a fault-tolerant quantum computer should adopt. More specifically, constructing a logical qubit requires not only the integration of multiple physical qubits but also the development of software. For example, after evaluating the characteristics of individual physical qubits, a program would be needed to calibrate the control system with good efficiency. The design of error-correcting operations and a feedback circuit would also be required. To evaluate this design, a quantum circuit simulator would be essential. The Feature Article "Fault-tolerant Technology for Quantum Information Processing and Its Implementation Methods" [9] introduces the research and development of a software platform toward fault-tolerant quantum computation.

As explained at the beginning of this article, it is said that 20,000,000 physical qubits would be needed for a fault-tolerant quantum computer. Although qubit integration technology and control technology are evolving rapidly for this purpose, there is still a need for breakthroughs. As a near-future target, studies are being conducted on combining a high-performance NISQ device that does not execute error correction with a conventional computer and applying such a hybrid computer to quantum machine learning, quantum chemical calculations, etc. At NTT laboratories, research is moving forward on making maximum use of the capabilities of NISQ devices on the basis of knowledge of computational theory and information theory. The Feature Article "Theoretical Approach to Overcome Difficulties in Implementing Quantum Computers" [10] introduces methods of improving the performance of NISQ devices* such as eliminating noise by limiting the means of operation or making effective use of uninitialized qubits by using a high-speed algorithm. The above-mentioned article "Fault-tolerant Technology for Quantum Information Processing and Its Implementation Methods" [9] also describes quantum error mitigation that, while increasing computational cost, has no need for increasing the number of qubits as a technique for compensating for NISQ device noise.

5. Future outlook

Twenty years ago, research was focused on implementing a single qubit, and it was said that quantum

* In this context, an NISQ device refers to a quantum computer with limited functions that can be developed before the development of a large-scale quantum computer with an error-correction function is completed, and the technologies introduced in the Feature Article "Theoretical Approach to Overcome Difficulties in Implementing Quantum Computers" will not necessarily be available in the near future.

computer technology was still 100 years off. Then, ten years later, a double-qubit gate was achieved and research focused on how the number of qubits could be increased. At that time, some people were voicing the opinion that a quantum computer was now only 50 years away. Currently, quantum computers consisting of several tens of qubits are in operation and quantum supremacy is being demonstrated. Ising machines having several thousand bits have also been developed, some of which have been commercialized. Furthermore, the development of a fault-tolerant quantum computer 30 years from now has been proposed through a Japanese national project. Therefore, the research on new-principle computers has been accelerating, number of researchers has been increasing, and range of research has been expanding. That being said, the path to a large-scale quantum computer is undoubtedly long. For experimental researchers, a breakthrough is desperately needed to develop such a quantum computer. It may be necessary to apply high-yield process technology such as modern large-scale integrated circuits, integration technology including control systems, or network technology toward distributed quantum computing. Theoretical researchers, meanwhile, anticipate breakthroughs such as quantum error-correcting codes that do not consume resources and new quantum algorithms. There is also the possibility that an extraordinary scientist somewhere will propose a completely novel idea that can suddenly provide a breakthrough to problem solving. In any case, these are without doubt challenging themes for which no one knows what will happen, so from here on, we would like to keep a close watch on developments in this field.

References

[1] C. Gidney and M. Ekera, "How to Factor 2048 Bit RSA Integers in 8 Hours Using 20 Million Noisy Qubits," arXiv:1905.09749, 2019.

[2] H. Takesue, T. Inagaki, K. Inaba, and T. Honjo, "Quantum Neural Network for Solving Complex Combinatorial Optimization Problems," NTT Technical Review, Vol. 15, No. 7, 2017. https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201707fa2.html

[3] H. Takesue, T. Inagaki, K. Inaba, and T. Honjo, "Performance Comparison between Coherent Ising Machines and Quantum Annealer," NTT Technical Review, Vol. 19, No. 5, pp. 18–22, 2021. https://ntt-review.jp/archive/ntttechnical.php?contents=ntr202105fa2.html

[4] F. Arute, K. Arya, R. Babbush, D. Bacon, J. C. Bardin, R. Barends, R. Biswas, S. Boixo, F. G. S. L. Brandao, D. A. Buell, B. Burkett, Y. Chen, Z. Chen, B. Chiaro, R. Collins, W. Courtney, A. Dunsworth, E. Farhi, B. Foxen, A. Fowler, C. Gidney, M. Giustina, R. Graff, K. Guerin, S. Habegger, M. P. Harrigan, M. J. Hartmann, A. Ho, M. Hoffmann, T. Huang, T. S. Humble, S. V. Isakov, E. Jeffrey, Z. Jiang, D. Kafri, K. Kechedzhi, J. Kelly, P. V. Klimov, S. Knysh, A. Korotkov, F. Kostritsa, D. Landhuis, M. Lindmark, E. Lucero, D. Lyakh, S. Mandrà, J. R. McClean, M. McEwen, A. Megrant, X. Mi, K. Michielsen, M. Mohseni, J. Mutus, O. Naaman, M. Neeley, C. Neill, M. Yuezhen Niu, E. Ostby, A. Petukhov, J. C. Platt, C. Quintana, E. G. Rieffel, P. Roushan, N. C. Rubin, D. Sank, K. J. Satzinger, V. Smelyanskiy, K. J. Sung, M. D. Trevithick, A. Vainsencher, B. Villalonga, T. White, Z. J. Yao, P. Yeh, A. Zalcman, H. Neven, and J. M. Martinis, "Quantum Supremacy Using a Programmable Superconducting Processor," Nature, Vol. 574, pp. 505–510, Oct. 2019.

[5] K. Kakuyanagi, Y. Matsuzaki, H. Toida, H. Yamaguchi, S. Saito, and W. J. Munro, "Demonstration of Realism Violation on a Macroscopic Scale," NTT Technical Review, Vol. 15, No. 7, 2017. https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201707fa4.html

[6] H. Toida, K. Kakuyanagi, W. J. Munro, H. Yamaguchi, and S. Saito, "Electron Spin Resonance Spectroscopy Using a Superconducting Flux Qubit," NTT Technical Review, Vol. 17, No. 8, pp. 11–15, 2019. https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201908ra1.html

[7] L. V. Abdurakhimov, I. Mahboob, H. Toida, K. Kakuyanagi, and S. Saito, "A Long-lived Tunable Qubit for Bosonic Quantum Computing," NTT Technical Review, Vol. 19, No. 5, pp. 23–28, 2021. https://ntt-review.jp/archive/ntttechnical.php?contents=ntr202105fa3.html

[8] W. J. Munro, V. M. Bastidas, K. Azuma, and K. Nemoto, "Designing Quantum Computers," NTT Technical Review, Vol. 19, No. 5, pp. 29–33, 2021. https://ntt-review.jp/archive/ntttechnical.php?contents=ntr202105fa4.html

[9] Y. Tokunaga, Y. Suzuki, S. Endo, and R. Asaoka, "Fault-tolerant Technology for Quantum Information Processing and Its Implementation Methods," NTT Technical Review, Vol. 19, No. 5, pp. 40–44, 2021. https://ntt-review.jp/archive/ntttechnical.php?contents=ntr202105fa6.html

[10] S. Akibue, Y. Takeuchi, Y. Takahashi, G. Kato, and S. Tani, "Theoretical Approach to Overcome Difficulties in Implementing Quantum Computers," NTT Technical Review, Vol. 19, No. 5, pp. 34–39, 2021. https://ntt-review.jp/archive/ntttechnical.php?contents=ntr202105fa5.html

Shiro Saito
Senior Distinguished Researcher and Group Leader of 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 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 currently a guest professor. He was appointed as Distinguished Scientist of NTT in 2012 and Senior Distinguished Researcher in 2021. He is a member of the Physical Society of Japan (JPS) and the Japan Society of Applied Physics (JSAP).

Hideki Gotoh
Vice President, Head of NTT Basic Research Laboratories.

He received a B.E., M.E., and Ph.D. in engineering from Hiroshima University in 1991, 1993, and 2000. Since joining NTT Basic Research Laboratories in 1993, he has been working on optical physics and device applications of semiconductor nanostructures. He was a visiting researcher at University of Illinois at Urbana-Champaign in the United States in 2004. He was a guest associate professor at University of Tsukuba from 2010 to 2016 and a guest professor from 2016 to 2019. He is a member of JSAP and the Optical Society (OSA).

Performance Comparison between Coherent Ising Machines and Quantum Annealer

Hiroki Takesue, Takahiro Inagaki, Kensuke Inaba, and Toshimori Honjo

Abstract

NTT has been developing a coherent Ising machine (CIM), which efficiently finds solutions to ground-state-search problems of the Ising model using a network of optical parametric oscillators. We introduce our recent experiment that compared the computational performance of CIMs (the one developed by NTT and another developed by Stanford University) with a quantum annealer, which solves the Ising-model problems using superconducting devices.

Keywords: coherent Ising machine, quantum annealing, LASOLV

1. Introduction

A combinatorial optimization problem is to find the best combination of choices among a set of many choices and is generally considered a problem that cannot be efficiently solved with modern digital computers. It is known that any combinatorial optimization problem can be efficiently converted to a ground-state-search problem of the Ising model, which is a theoretical model describing the behavior of interacting spins. A recent trend is to solve such Ising model problems through experiments using physical systems that mimic the Ising spins, which are now called *Ising machines*. A pioneer of such Ising machines is the quantum annealer (QA) based on superconducting qubits as artificial spins. The Canadian company D-Wave Systems has developed QAs with as many as thousands of qubits [1]. A coherent Ising machine (CIM) is another Ising machine that uses degenerate optical parametric oscillators (DOPO) as spins [2, 3]. NTT has been developing a computation system based on the concept of a CIM called LASOLV™ [4]. In this article, we describe the experimental performance comparison between CIMs and a D-Wave QA undertaken by NTT, National Aeronautics and Space Administration (NASA), and Stanford University [5].

2. Coherent Ising machine

A DOPO is an optical oscillator that can be implemented by placing a phase sensitive-amplifier (PSA) in an optical cavity. A PSA is an optical amplifier based on optical parametric amplification and efficiently amplifies the 0 and π phase components relative to the pump phase [6]. As a result, a DOPO takes only the 0 or π phase above the oscillation threshold; thus, the discrete phase states can be used to represent the Ising spin states. The CIM developed by NTT generates thousands of time-multiplexed DOPO pulses by turning on and off the PSA placed in a 1-km optical fiber with a 1-GHz frequency (**Fig. 1**). The interaction between the DOPO pulses are implemented using a method called measurement-feedback. With this method, some of the pulse energies of N DOPO pulses are split using a beam splitter for each circulation in the cavity, and the amplitudes (with signs) of all N DOPO pulses are measured. The measurement results are input into a digital circuit for matrix computation, where the spin-spin interaction matrix (an $N \times N$ matrix) for a given Ising problem is

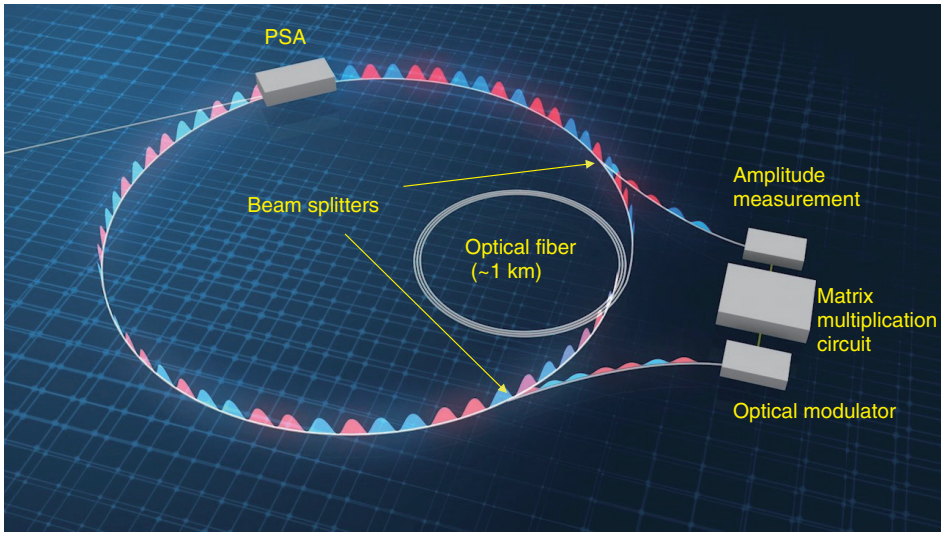


Fig. 1. Coherent Ising machine based on measurement-feedback.

installed in advance. The digital circuit multiplies the matrix and set of *N* measurement results to obtain the feedback signal for each pulse in the next circulation in the cavity. We then use the feedback signal to modulate an optical pulse, and the pulse is injected into the corresponding DOPO pulse inside the cavity through another beam splitter to complete measurement-feedback. By repeating this procedure typically during 100 to 1000 circulations in the cavity while increasing the pump amplitude from zero, the combination of the DOPO phases evolves into a phase that most stabilizes the whole system, which in many cases corresponds to the ground state of the given Ising problem. In 2016, we reported on a 2000-node CIM on the basis of measurement-feedback and demonstrated that the CIM could deliver a fine solution to a 2000-node optimization problem ~50 times faster than simulated annealing implemented on a modern digital computer [2].

3. Performance comparison with D-Wave QA

A quantum annealing algorithm starts with an initial state where all the spins are set at the superposition of spin up and down by applying transverse magnetic fields. Spin-spin interactions for a given Ising problem are then gradually implemented while decreasing the transverse field. With the help of quantum fluctuations, the whole system reaches ground states with high probability [7]. The D-Wave QA has been used in several proof-of-concept experiments such as optimization of traffic flow [8]. We report on an experiment of comparing the probabilities to obtain the exact solutions to the common Ising model problems obtained with the CIM developed by NTT and that developed by Stanford University and the 2000-spin D-Wave QA owned by NASA [5].

We denote the number of spins for a given problem, average number of couplings per spin, and coupling density as N_p, *d*, and *D* (= *d*/N_p), respectively. The success probabilities for the problems with $D = 50\%$ as a function of N_p are shown in **Fig. 2**(**a**). While the success probability of the CIMs did not decrease significantly with the problem size, that for the D-Wave QA decreased rapidly as we increased N_p and ended up at 0.001% with N_p = 50. **Figure 2**(**b**) shows the relationship between the success probabilities and *d*. When solving sparse problems with $d = 3$, the D-Wave QA slightly outperformed the CIMs. However, the performance of the D-Wave QA degraded rapidly when N_p increased. The success probabilities for $d = 3$ and $D = 50\%$ were comparable, indicating that success probability does not depend on *D*.

The performance difference between the CIMs and D-Wave QA probably originates from the difference in the implementation of the spin-spin couplings. The superconducting qubits of the D-Wave QA are physically connected in a graph structure called a Chimera graph, where each qubit has only six connections. Therefore, a given Ising problem needs to be converted to a problem on the Chimera graph structure before being input into the D-Wave QA, which often

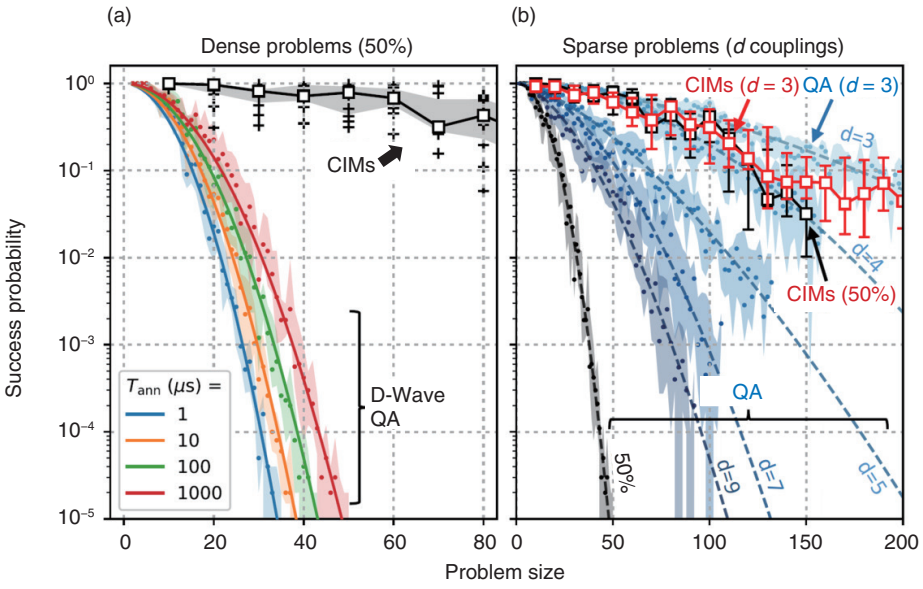


Fig. 2. Success probabilities of CIMs and D-Wave QA in solving Ising problems. (a) Dense problems (50% coupling density), (b) problems with various coupling densities.

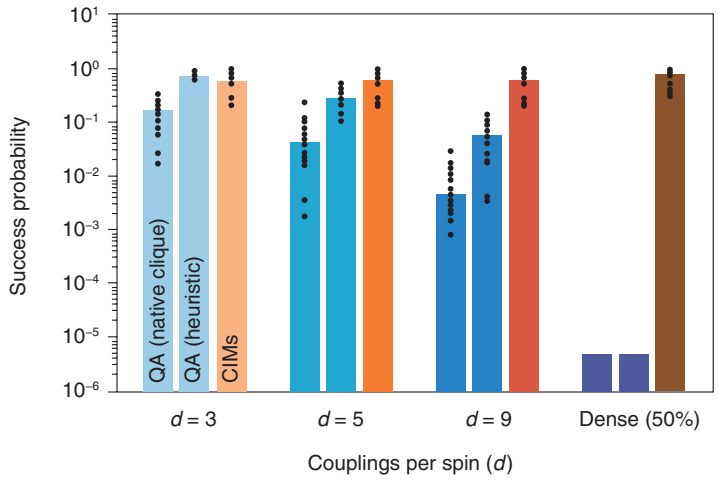


Fig. 3. Success probability dependence on coupling densities for 50-spin problems.

increases the number of required spins, especially when solving dense problems. On the other hand, all-to-all coupling among DOPO pulses is possible in the CIMs because of the use of measurement-feedback; thus, Ising problems with any coupling density can be input without any conversion.

In our experiment, we used two methods for converting Ising problems into the Chimera graph structure. One is the native clique embedding method, which embeds a problem into the Chimera graph in accordance with a pre-determined algorithm. The other is a heuristic method that undertakes optimization to minimize the required number of qubits in advance for each problem. The success probabilities of the D-Wave QA on the basis of these two embedding methods and that for the CIMs for 50-node Ising problems are shown in **Fig. 3**. The use of the heuristic method could decrease the effective problem size run

on the D-Wave QA, improving success probabilities compared with native clique embedding. Nevertheless, the success probabilities of the D-Wave QA did not reach that of the CIMs when d was 5 or larger, and the difference increased as d or D increased. This suggests that the manner of implementing the Ising model problems onto a physical system significantly affects the computational performance of Ising machines based on such systems.

4. Summary

We described the performance comparison between CIMs and the D-Wave QA (as of 2019). We expect that both will evolve and improve in performance. Important future work regarding these computers is to clarify if such new computers based on physical systems can have an advantage over modern digital computers and if such an advantage can lead to applications that benefit society.

References

[1] D-Wave Systems, www.dwavesys.com

[2] T. Inagaki, Y. Haribara, K. Igarashi, T. Sonobe, S. Tamate, T. Honjo, A. Marandi, P. L. McMahon, T. Umeki, K. Enbutsu, O. Tadanaga, H. Takenouchi, K. Aihara, K. Kawarabayashi, K. Inoue, S. Utsunomiya, and H. Takesue, "A Coherent Ising Machine for 2000-node Optimization Problems," Science, Vol. 354, No. 6312, pp. 603–606, 2016.

[3] P. McMachon, A. Marandi1, Y. Haribara, R. Hamerly, C. Langrock, S. Tamate, T. Inagaki, H. Takesue, S. Utsunomiya, K. Aihara, R. L. Byer, M. M. Fejer, H. Mabuchi, and Y. Yamamoto, "A Fully Programmable 100-spin Coherent Ising Machine with All-to-all Connections," Science, Vol. 354, No. 6312, pp. 614–617, 2016.

[4] J. Arai, S. Yagi, H. Uchiyama, K. Tomita, K. Miyahara, T. Tomoe, and K. Horikawa, "LASOLV™ Computing System: Hybrid Platform for Efficient Combinatorial Optimization," NTT Technical Review, Vol. 18, No. 1, pp. 35–40, 2020.
https://ntt-review.jp/archive/ntttechnical.php?contents=ntr202001fa5.html

[5] R. Hamerly, T. Inagaki, P. L. McMahon, D. Venturelli, A. Marandi, T. Onodera, E. Ng, C. Langrock, K. Inaba, T. Honjo, K. Enbutsu, T. Umeki, R. Kasahara, S. Utsunomiya, S. Kako, K. Kawarabayashi, R. L. Byer, M. M. Fejer, H. Mabuchi, D. Englund, E. Rieffel, H. Takesue, and Y. Yamamoto, "Experimental Investigation of Performance Differences between Coherent Ising Machines and a Quantum Annealer," Sci. Adv., Vol. 5, No. 5, eaau0823, 2019.

[6] T. Umeki, T. Kazama, T. Kobayashi, K. Enbutsu, R. Kasahara, and Y. Miyamoto, "Low-noise Amplification and Nonlinearity Mitigation Based on Parametric Repeater Technology," NTT Technical Review, Vol. 17, No. 5, pp. 20–26, 2019.
https://ntt-review.jp/archive/ntttechnical.php?contents=ntr201905fa3.html

[7] T. Kadowaki and H. Nishimori, "Quantum Annealing in the Transverse Ising Model," Phys. Rev. E, Vol. 58, No. 5, 5355, 1998.

[8] F. Neukart, G. Compostella, C. Seidel, D. von Dollen, S. Yarkoni, and B. Parney, "Traffic Flow Optimization Using a Quantum Annealer," Frontiers in ICT, Vol. 4, 29, 2017.

Hiroki Takesue
Senior Distinguished Scientist, Group Leader, Quantum Optical State Control Research Group, Quantum Science and Technology Laboratory, NTT Basic Research Laboratories.

He received a B.E., M.E., and Ph.D. in engineering science from Osaka University in 1994, 1996, and 2002. In 1996, he joined NTT Basic Research Laboratories, where he was engaged in research on lightwave frequency synthesis, optical access networks using wavelength division multiplexing, and quantum optics. He is currently pursuing research on communication and computation using quantum optics technologies. He is the recipient of several awards, including the ITU-T Kaleidoscope Conference 2nd Best Paper Award in 2008 and The Commendation for Science and Technology by the Minister of Education, Culture, Sports, Science and Technology of Japan (The Young Scientists' Prize) in 2010. He was a visiting scholar at Stanford University, California, USA, from 2004 to 2005, and a guest researcher at the National Institute of Standards and Technology (NIST), Colorado, USA, in 2014. He is a guest professor in the Graduate School of Engineering Science, Osaka University, and a member of the Institute of Electrical and Electronics Engineers (IEEE) and the Japan Society of Applied Physics (JSAP).

Takahiro Inagaki
Distinguished Scientist, Quantum Optical State Control Research Group, Quantum Science and Technology Laboratory, NTT Basic Research Laboratories.

He received a B.E., M.E., and Ph.D. in engineering science from Tohoku University, Miyagi, in 2007, 2009, and 2012. While at Tohoku University, he conducted research on quantum state transfer between photon polarization and an electron spin in a semiconductor. He joined the Quantum Optical State Control Research Group in NTT Basic Research Laboratories in 2012. He is currently researching quantum communication based on entanglement and a coherent Ising machine for combinatorial optimization problems. He is a member of the Physical Society of Japan (PSJ) and JSAP.

Kensuke Inaba
Senior Research Scientist, Quantum Optical State Control Research Group, Quantum Science and Technology Laboratory, NTT Basic Research Laboratories.

He received a B.E., M.E., and Dr.Eng. from the Department of Applied Physics, Osaka University, in 2003, 2005, and 2007. He joined NTT Basic Research Laboratories in 2008. Since 2004, he has been studying quantum many-body physics in correlated electron systems in condensed matter. His current interests are correlated cold atoms and photons and their application to quantum simulation and computation.

Toshimori Honjo
Senior Research Scientist, Supervisor, Quantum Optical State Control Research Group, Quantum Science and Technology Laboratory, NTT Basic Research Laboratories.

He received a B.S. and M.S. in information science from Tokyo Institute of Technology in 1996 and 1998, and a Ph.D. in engineering from Osaka University in 2007. In 1998, he joined NTT Software Laboratories, where he conducted research on secure communication systems. Since 2003, he has been researching quantum optics and quantum cryptography at NTT Basic Research Laboratories. His current research interests include physical computer systems and quantum communication. In 2009, he was a visiting researcher at the University of Vienna, Austria. He is a member of the Information Processing Society of Japan (IPSJ).

A Long-lived Tunable Qubit for Bosonic Quantum Computing

Leonid V. Abdurakhimov, Imran Mahboob, Hiraku Toida, Kosuke Kakuyanagi, and Shiro Saito

Abstract

This article reviews a recent experiment that demonstrated a new hybrid type of tunable superconducting qubit consisting of a two-dimensional (2D) superconducting capacitively shunted flux qubit coupled to a 3D microwave cavity. Such a 3D hybrid flux qubit has extremely long relaxation times, which are comparable to or exceed the best values reported for 2D flux qubits. This new design of frequency-tunable qubits is uniquely suited for applications in bosonic quantum computing, which exploits multiphoton states of 3D microwave cavities.

Keywords: superconducting flux qubit, 3D microwave cavity, quantum computation

1. Introduction

Quantum computation is a new computing paradigm based on well-studied, but not yet fully exploited, quantum phenomena such as superposition and entanglement. The key element of a quantum computer is a qubit: a two-level system that can be prepared in an arbitrary state $|\psi\rangle$ described by a linear combination (superposition) of qubit pure states $|0\rangle$ and $|1\rangle$ (i.e., $|\psi\rangle = \alpha\,|0\rangle + \beta\,|1\rangle$, where α and β are complex amplitudes). The number of computational basis states of a system of N qubits scales with the number of qubits as 2^N. For example, a state of a two-qubit system can be described by a superposition of $2^2 = 4$ states (i.e., $|\psi\rangle = \alpha_{00}\,|00\rangle + \alpha_{01}\,|01\rangle + \alpha_{10}\,|10\rangle + \alpha_{11}|11\rangle$). Some multi-qubit states can be represented as a product of individual qubit states: for example, the state of two qubits A and B $|\psi_{AB}\rangle = \frac{1}{2}(|0_A0_B\rangle + |0_A1_B\rangle + |1_A0_B\rangle + |1_A1_B\rangle)$ can be decomposed to the product of the states of A and B $|\psi_{AB}\rangle = |\psi_A\rangle|\psi_B\rangle = \frac{1}{\sqrt{2}}(|0\rangle_A + |1\rangle_A) \times \frac{1}{\sqrt{2}}(|0\rangle_B + |1\rangle_B)$. Other multi-qubit quantum states—so-called entangled states—cannot be obtained by multiplication of single qubit states, for example, the two-qubit state $\frac{1}{\sqrt{2}}|0_A0_B\rangle + |1_A1_B\rangle$. By using superposition and entanglement phenomena, information from many qubits can be simultaneously encoded in a single quantum state. Quantum algorithms, such as Shor's algorithm for integer factorization and Grover's search algorithm, use this property of quantum systems (quantum parallelism) to encode different function instances (different function realizations) in a single quantum state and compute their outcomes at once. A sufficiently large quantum computer will be able to solve a number of numerical problems, such as integer factorization and database search, much faster than conventional computers.

Several hardware platforms for quantum computing are being actively studied, including superconducting qubits, trapped ions, optical photonic systems, and spin qubits. Although the above hardware platforms have their own unique advantages and disadvantages, superconducting qubits is arguably the most advanced hardware platform for quantum computing, which is pursued by tech giants such as Google, IBM, and Microsoft, as well as new startups such as Rigetti Computing.

There are many types of superconducting qubits, but one particular variation (the so-called transmon) has become prevalent. Transmons can be roughly separated into two large groups: fixed-frequency and tunable qubits. Fixed-frequency qubits generally demonstrate better coherence times, while tunable qubits provide more freedom in terms of possible quantum protocols. In the most advanced superconducting

Table 1. The comparison table of currently the most advanced superconducting qubit platforms. [(a)] Values in the brackets are coherence times of the 2D transmon (ancilla) qubit. Data are taken from F. Arute et al. [1], E. J. Zhang et al. [2], M. A. Rol et al. [3], and N. Ofek et al. [4].

Company/ university	Qubit encoding	Qubit type	Tunable?	T_1 (μs)	T_2 (μs)	Number of qubits	Error correction?
Google	Two-level system	2D transmon	Yes	16	N/A	53	No
IBM	Two-level system	2D transmon	No	79	69	65	No
Delft University	Two-level system	2D transmon	Yes	32	34	5	No
Yale University	Cavity bosonic states	3D cavity + 2D transmon	No	N/A [35[(a)]]	320 [13[(a)]]	1	Yes

qubit platforms, both fixed-frequency qubits (IBM, Yale University) and tunable qubits (Google, Delft University) are used (**Table 1**).

Despite the recent impressive progress in building small-scale superconducting qubit systems, development of a universal quantum computer is still a very challenging task. The major obstacle is the loss of quantum information due to the interaction with noisy environments. For single qubit operations, the important figures of merit are energy-relaxation time T_1 and dephasing time T_2. In systems, in which quantum information is encoded directly into physical qubits (Google, IBM and Delft University systems), typical T_1 and T_2 values are in the range of 100 μs (Table 1), while characteristic qubit manipulation times are on the order of 100 ns. Therefore, error rates in state-of-the-art systems are on the order of $p \approx 10^{-3}$. To execute reliable quantum computations at such error rates, implementation of error-correction algorithms is necessary, which will require about 10^7 physical qubits. At the current level of technology, the brute-force approach to the scaling of superconducting qubit systems to such high qubit numbers is not straightforward. There is a limit to expanding the current qubit surface-mount technology, and next-generation mounting technologies such as distributed quantum computation are expected.

Recently, there has been significant interest in the idea of hardware-efficient quantum computing architectures, which use quantum systems with intrinsic protection against quantum errors. One of the promising approaches is bosonic quantum computing, which is based on the encoding of qubit states $|0\rangle_Q$ and $|1\rangle_Q$ into complex superposition states of a three-dimensional (3D) superconducting cavity. In the quantum regime, a 3D microwave cavity can be considered a harmonic oscillator with an infinite number of equidistant energy levels. Each level corresponds to a quantum state with a fixed number of microwave photons in the cavity (i.e., states $|0\rangle$, $|1\rangle$, …, $|n\rangle$ correspond to one, two, …, n photons in the cavity). The pure cavity states are equidistant and cannot be used for qubit encoding because it is difficult to address only two of them without exciting other states. In bosonic computing, this issue is solved by encoding qubit states $|0\rangle_Q$ and $|1\rangle_Q$ into superposition cavity states, such as $|0\rangle_Q = \frac{1}{\sqrt{2}}(|0\rangle + |4\rangle)$ and $|1\rangle_Q = |2\rangle$, by using an additional "ancilla" superconducting qubit coupled to the cavity. The advantage of bosonic quantum computing is that the main error mechanism is the cavity-photon loss, which is relatively easy to detect and correct.

In this article, a new type of tunable qubit, 3D hybrid flux qubit, is described that can be used for bosonic computing. Due to their tunability, such qubits can be used as ancillas for achieving dynamical coupling between photonic states of different cavities, which will be an important step towards building a large-scale bosonic quantum computer.

2. Qubit design and experimental results

Our 3D hybrid qubit consists of a capacitively shunted (c-shunt) 2D flux qubit embedded in a 3D copper cavity, as shown in **Figs. 1** and **2** [5]. C-shunt flux qubits were proposed theoretically in 2007 [6] and were initially fabricated in a 2D architecture [7]. The novelty of our 3D hybrid qubit design relies on the combination of a 2D c-shunt flux qubit with a 3D microwave resonator.

A schematic and images of the 2D c-shunt flux qubit used as a part of our 3D hybrid qubit are shown in Figs. 1(a)–(d). The 2D qubit consists of an aluminum superconducting loop interrupted by three Josephson junctions (Figs. 1(a), (d)). Two junctions are identical, while the third junction is smaller than the other two. The small junction is shunted by a large coplanar capacitor consisting of two large rectangular aluminum pads (Figs. 1(a)–(c)). The 2D qubit is fabricated on a sapphire substrate, and Josephson junctions

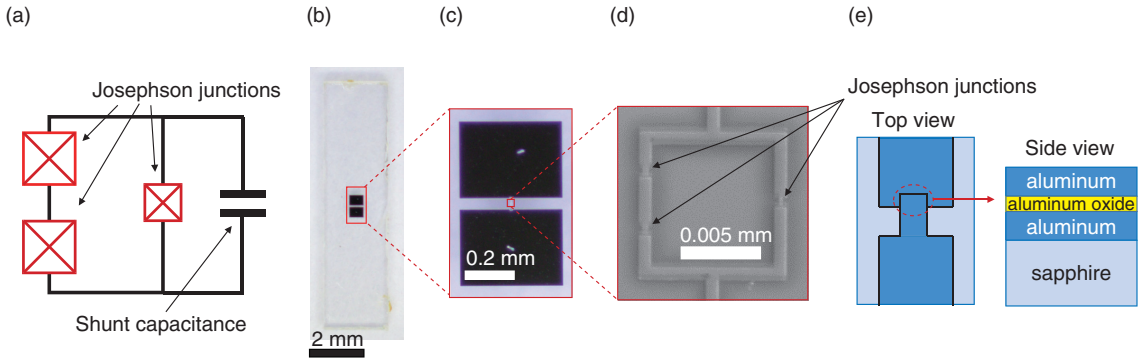


Fig. 1. (a) A schematic of a 2D superconducting flux qubit with a shunt capacitance. (b, c, d) Images of this qubit. (e) A schematic of a Josephson junction.

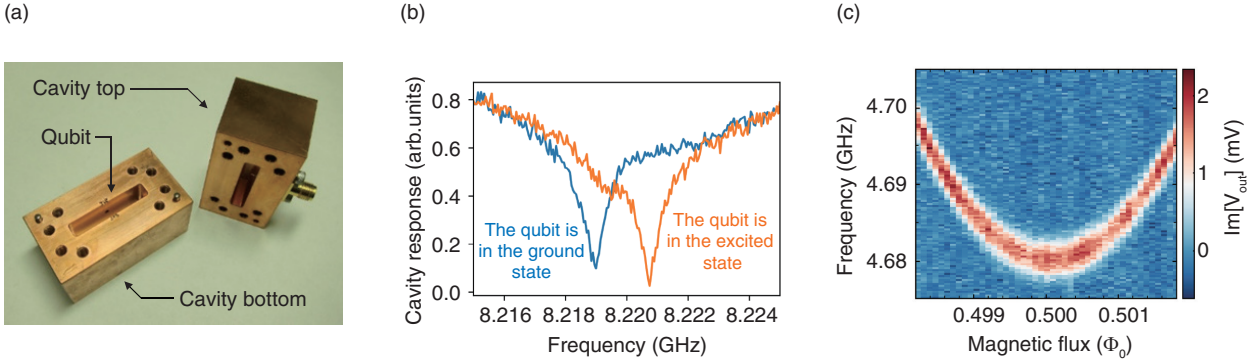


Fig. 2. (a) A photo of the 2D c-shunt flux qubit mounted inside a 3D microwave cavity. (b) Qubit state can be determined from the cavity resonance frequency. (c) The qubit frequency can be tuned by applying a magnetic flux through the qubit loop (Φ_0 is the magnetic flux quantum, $\Phi_0 \approx 2 \times 10^{-15}$ Wb).

are fabricated using double-angle shadow evaporation of aluminum. Each Josephson junction consists of two superconducting aluminum layers separated by a thin (about 2 nm) insulating layer of aluminum oxide formed by oxidizing the aluminum film in oxygen atmosphere (Fig. 1(e)).

The substrate with the 2D c-shunt flux qubit is embedded in a 3D microwave cavity (Fig. 2(a)), and the device is cooled to 10 mK (approximately –273°C) using a dilution refrigerator. Because of the capacitive coupling between the qubit and cavity, the cavity-resonance frequency shifts to higher frequencies when the qubit is excited from the ground state to the excited state (Fig. 2(b)); hence, the state of the qubit can be determined from the cavity-resonance frequency. The qubit-transition frequency can be tuned by applying an external magnetic flux $\Phi_e = B \times S$, where B is the applied magnetic field and S is the qubit loop area (Fig. 2(c)).

This 3D hybrid flux qubit demonstrates long T_1 in the range 60–90 μs [5] and T_2 up to 100 μs [8] (**Fig. 3(a)**). These coherence times exceed those reported for other types of flux qubits (**Fig. 3(b)**) and are comparable to relaxation times observed in fixed-frequency qubits such as transmon qubits.

3. Outlook

Our 3D hybrid flux qubit demonstrates relaxation times comparable to those reported for transmon qubits, which are typically used in superconducting platforms. This makes 3D hybrid flux qubits a possible alternative to transmons for applications in quantum computing. Due to its intrinsic 3D structure, this new design can be particularly useful for bosonic computing.

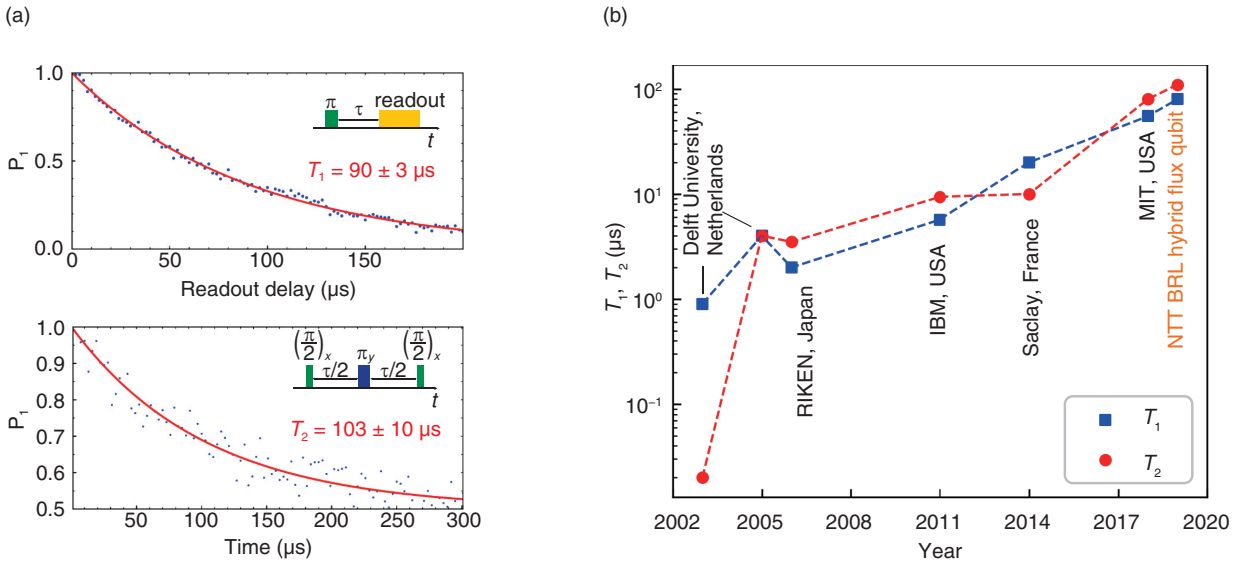


Fig. 3. (a) Results of energy-relaxation and dephasing measurements of our 3D hybrid flux qubit. (b) Evolution of relaxation times in superconducting flux qubits.

Quantum computing using bosonic states of 3D superconducting cavities is a promising approach for hardware-efficient quantum-information processing. The possibility of the error-correction of a single qubit encoded in cavity photon states has been recently demonstrated [9]. Next important steps will be improving the error-correction protocol to achieve full fault-tolerant protection and demonstrating logical operations (gates) between multiple error-corrected qubits. Two-qubit gates can be achieved by coupling qubits encoded in two cavities via ancilla superconducting qubits. For example, a controlled NOT (CNOT) gate between two cavity qubits using a fixed-frequency transmon qubit has been reported [10]. However, one of the problems of the transmon-based scheme is that the coupling between the cavities is fixed; therefore, there is an unwanted interaction between qubits encoded in cavity states even in the idle state. This issue can be solved using dynamical coupling schemes, in which the interaction between the cavities is switched on only for a short time. Our 3D hybrid flux qubit is uniquely suited for this purpose since its frequency can be adjusted by applying an external magnetic field. A possible concept of the dynamic coupling protocol using our 3D hybrid flux qubit is shown in **Fig. 4**. The coupling between cavity qubits is facilitated by two ancilla qubits capacitively coupled to each other (Fig. 4(a)). Ancillas are coupled to different cavities, and the interaction between the cavities can be switched on and off using the so-called iSwap gate [11] between ancilla qubits (Fig. 4(b)).

In conclusion, the new type of 3D superconducting qubit described in this article combines long coherence times with frequency tunability. Future work will explore possible applications of such high-coherence tunable qubits in hardware-efficient schemes of quantum computing, such as bosonic computing.

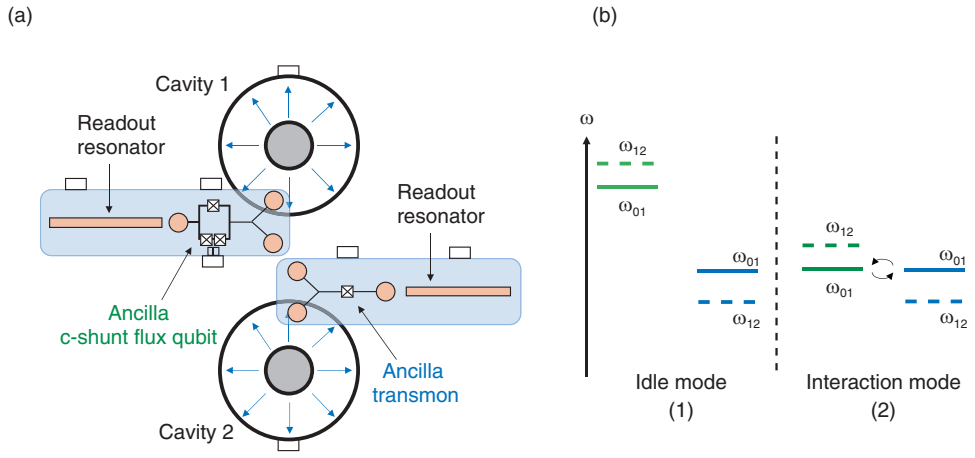


Fig. 4. (a) Schematic of a device to achieve two-qubit logical operation (gate) between logical qubits encoded in bosonic states of two coaxial 3D cavities. The dynamic coupling between bosonic qubits is based on the iSwap gate between ancilla qubits. (b) Schematic representation of the iSwap gate between ancilla qubits. (1) Ancilla qubits are in idle state separated by large detuning. (2) Ancilla qubits are tuned into resonance with each other.

References

[1] F. Arute, K. Arya, R. Babbush, D. Bacon, J. C. Bardin, R. Barends, R. Biswas, S. Boixo, F. G. S. L. Brandao, D. A. Buell, B. Burkett, Y. Chen, Z. Chen, B. Chiaro, R. Collins, W. Courtney, A. Dunsworth, E. Farhi, B. Foxen, A. Fowler, C. Gidney, M. Giustina, R. Graff, K. Guerin, S. Habegger, M. P. Harrigan, M. J. Hartmann, A. Ho, M. Hoffmann, T. Huang, T. S. Humble, S. V. Isakov, E. Jeffrey, Z. Jiang, D. Kafri, K. Kechedzhi, J. Kelly, P. V. Klimov, S. Knysh, A. Korotkov, F. Kostritsa, D. Landhuis, M. Lindmark, E. Lucero, D. Lyakh, S. Mandrà, J. R. McClean, M. McEwen, A. Megrant, X. Mi, K. Michielsen, M. Mohseni, J. Mutus, O. Naaman, M. Neeley, C. Neill, M. Yuezhen Niu, E. Ostby, A. Petukhov, J. C. Platt, C. Quintana, E. G. Rieffel, P. Roushan, N. C. Rubin, D. Sank, K. J. Satzinger, V. Smelyanskiy, K. J. Sung, M. D. Trevithick, A. Vainsencher, B. Villalonga, T. White, Z. J. Yao, P. Yeh, A. Zalcman, H. Neven, and J. M. Martinis, "Quantum Supremacy Using a Programmable Superconducting Processor," Nature, Vol. 574, pp. 505–510, 2019.

[2] E. J. Zhang, S. Srinivasan, N. Sundaresan, D. F. Bogorin, Y. Martin, J. B. Hertzberg, J. Timmerwilke, E. J. Pritchett, J. B. Yau, C. Wang, W. Landers, E. P. Lewandowski, A. Narasgond, S. Rosenblatt, G. A. Keefe, I. Lauer, M. B. Rothwell, D. T. McClure, O. E. Dial, J. S. Orcutt, M. Brink, and J. M. Chow, "High-fidelity Superconducting Quantum Processors via Laser-annealing of Transmon Qubits," arXiv:2012.08475, 2020.

[3] M. A. Rol, F. Battistel, F. K. Malinowski, C. C. Bultink, B. M. Tarasinski, R. Vollmer, N. Haider, N. Muthusubramanian, A. Bruno, B. M. Terhal, and L. DiCarlo, "Fast, High-fidelity Conditional-phase Gate Exploiting Leakage Interference in Weakly Anharmonic Superconducting Qubits," Phys. Rev. Lett., Vol. 123, No. 12, 120502, 2019.

[4] N. Ofek, A. Petrenko, R. Heeres, P. Reinhold, Z. Leghtas, B. Vlastakis, Y. Liu, L. Frunzio, S. M. Girvin, L. Jiang, M. Mirrahimi, M. H. Devoret, and R. J. Schoelkopf, "Extending the Lifetime of a Quantum Bit with Error Correction in Superconducting Circuits," Nature, Vol. 536, pp. 441–445, 2016.

[5] L. V. Abdurakhimov, I. Mahboob, H. Toida, K. Kakuyanagi, and S. Saito, "A Long-lived Capacitively Shunted Flux Qubit Embedded in a 3D Cavity," Appl. Phys. Lett., Vol. 115, No. 26, 262601, 2019.

[6] J. Q. You, X. Hu, S. Ashhab, and F. Nori, "Low-decoherence Flux Qubit," Phys. Rev. B, Vol. 75, No. 14, 140515(R), 2007.

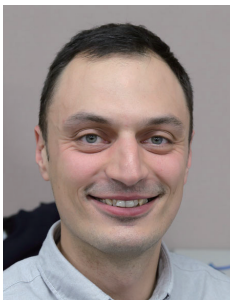
[7] F. Yan, S. Gustavsson, A. Kamal, J. Birenbaum, A. P. Sears, D. Hover, T. J. Gudmundsen, D. Rosenberg, G. Samach, S. Weber, J. L. Yoder, T. P. Orlando, J. Clarke, A. J. Kerman, and W. D. Oliver, "The Flux Qubit Revisited to Enhance Coherence and Reproducibility," Nat. Commun., Vol. 7, 12964, 2016.

[8] L. V. Abdurakhimov, I. Mahboob, H. Toida, K. Kakuyanagi, Y. Matsuzaki, and S. Saito, "Driven-state Relaxation of a Coupled Qubit-defect System in Spin-locking Measurements," Phys. Rev. B, Vol. 102, No. 10, 100502(R), 2020.

[9] N. Ofek, A. Petrenko, R. Heeres, P. Reinhold, Z. Leghtas, B. Vlastakis, Y. Liu, L. Frunzio, S. M. Girvin, L. Jiang, M. Mirrahimi, M. H. Devoret, and R. J. Schoelkopf, "Extending the Lifetime of a Quantum Bit with Error Correction in Superconducting Circuits," Nature, Vol. 536, 441, 2016.

[10] S. Rosenblum, Y. Y. Gao, P. Reinhold, C. Wang, C. J. Axline, L. Frunzio, S. M. Girvin, L. Jiang, M. Mirrahimi, M. H. Devoret, and R. J. Schoelkopf, "A CNOT Gate between Multiphoton Qubits Encoded in Two Cavities," Nat. Commun., Vol. 9, 652, 2018.

[11] R. Barends, C. M. Quintana, A. G. Petukhov, Y. Chen, D. Kafri, K. Kechedzhi, R. Collins, O. Naaman, S. Boixo, F. Arute, K. Arya, D. Buell, B. Burkett, Z. Chen, B. Chiaro, A. Dunsworth, B. Foxen, A. Fowler, C. Gidney, M. Giustina, R. Graff, T. Huang, E. Jeffrey, J. Kelly, P. V. Klimov, F. Kostritsa, D. Landhuis, E. Lucero, M. McEwen, A. Megrant, X. Mi, J. Mutus, M. Neeley, C. Neill, E. Ostby, P. Roushan, D. Sank, K. J. Satzinger, A. Vainsencher, T. White, J. Yao, P. Yeh, A. Zalcman, H. Neven, V. N. Smelyanskiy, and J. M. Martinis, "Diabatic Gates for Frequency-tunable Superconducting Qubits," Phys. Rev. Lett., Vol. 123, No. 21, 210501, 2019.



Leonid V. Abdurakhimov
Research Specialist, Superconducting Quantum Circuits Research Group, NTT Basic Research Laboratories.

He received a B.Sc. and M.Sc in applied physics and mathematics from Moscow Institute of Physics and Technology, Russia, and a Ph.D. in physics from the Institute of Solid State Physics, Russia, in 2005, 2007, and 2010. He was a research scientist at the Institute of Solid State Physics, Russia, from 2011 to 2012 then held postdoctoral positions at Okinawa Institute of Science and Technology, Japan (2013–2015), Cambridge University, UK (2016), and University College London, UK (2017–2018). He joined NTT Basic Research Laboratories in 2019 as a research specialist. His current research interests are superconducting quantum circuits and spin-based hybrid quantum systems. He is a member of the American Physical Society (APS).

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 a 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 Scientist of NTT in 2013. He is a member of the Institute of Physics (IoP) and APS.

Hiraku Toida
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 a 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).

Kosuke Kakuyanagi
Senior Research Scientist, 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.

Shiro Saito
Senior Distinguished Researcher and Group Leader of 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 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 currently a guest professor. He was appointed as Distinguished Scientist of NTT in 2012 and Senior Distinguished Researcher in 2021. He is a member of JPS and JSAP.

Designing Quantum Computers

William John Munro, Victor M. Bastidas, Koji Azuma, and Kae Nemoto

Abstract

We have reached the stage in our development of quantum technologies at which we are now able to construct the building blocks necessary to create small-scale quantum processors and networks. The challenge is how we scale up to fully fault-tolerant quantum computers involving extremely large numbers of interacting quantum bits.

Keywords: quantum design, quantum architecture, distributed quantum computation

1. Introduction

It has been known for almost half a century that the principles of quantum mechanics will allow technologies to be developed that provide new capabilities impossible with conventional technology or significant performance enhancements over our existing ones [1]. These benefits exploit 'quantum coherence' and in particular quantum entanglement to some degree for technological advantages in areas ranging from quantum sensing and imaging [2] to quantum communication [3] and computation [4]. While these fields are still in their infancy, we can already imagine a quantum internet connecting quantum computers together that take information from a number of inputs including quantum sensor arrays [5, 6]. Quantum clocks will synchronize all these devices if necessary [7]. How do we move from our few-qubit devices to noisy intermediate-scale quantum (NISQ) processors [8] and ultimately fault-tolerant quantum computers (FTQCs) [9], as depicted in **Fig. 1**? What is the path to achieve this?

The last decade has seen a paradigm shift in the capabilities of quantum technologies and what they can achieve. We have moved from the 'in-principle' few-qubit demonstrations in the laboratory to moderate-scale quantum processors that are available for commercial use. Superconducting circuits, ion traps, and photonics have enabled the development of devices with approximately 50 qubits operating together in a coherence fashion, and important quantum algorithms have already been performed on them. These NISQ processors have shown the capability to create complexity difficult for classical computers to calculate—the so-called quantum advantage [11, 12]. The noisy nature of the qubits, gates, measurements, and control systems used in these NISQ processors however severely limits the size of tasks that can be under taken on them (see Fig. 1). Noise-mitigation techniques may help a little to push the system size up, but fault-tolerant error-correction techniques are required if we want to scale up to large-scale universal quantum computers (machines with 10^6–10^8 physical qubits performing trillions of operations). The fundamental question is how we envision this given our current technological status.

2. NISQ processors and simulators

It is useful to begin our design consideration with an exploration of NISQ processors and simulators. They are fully programmable machines generally designed to undertake a specific range of tasks. In principle, they are universal in nature but noise limits them to being special-purpose machines. They have proved extremely useful in showcasing the potential that quantum physics offers. There are a number of physical systems from which these NISQ processors can be built despite superconducting circuits and ion traps being the most advanced [8]. A processor or simulator is however much more than just a collection of quantum bits (20–100 qubits) working together in

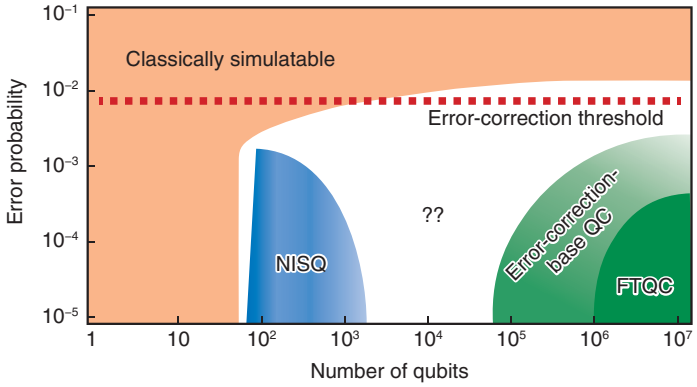


Fig. 1. The landscape for quantum computation in terms of resources (number of physical qubits) and error probabilities [10]. The orange-shaded area depicts the regime in which the tasks undertaken on it could also be classically simulated. The bluish area represents the NISQ regime in which tasks can be potentially performed with quantum advantage. Finally, the light-green area represents quantum computers (QCs) using error correction, while the dark-green area is fault-tolerant universal quantum computers.

some fashion. It is a highly integrated device involving many systems at different levels working seamlessly together, which we illustrate as a one-layer design in **Fig. 2**. At the top of the layer is the application one wants to perform on the NISQ processors, which could involve, e.g., simulation and sampling. The application can be quite an abstract object, so it is translated into an algorithm giving the instructions/rules that a computer needs to do to complete this task. These instructions are decomposed into a set of basic operations (gates, measurement, etc.) that the processor will run. Such operations are quite generic and not hardware specific. The operating system will take this sequence of gates and determine how they can be implemented on the given processor architecture. The layout and connectivity of the NISQ processor determines what algorithms can be performed. The operating system will turn them into a set of "physical instruction signals/pulses, etc." that the classical control system (CCS) will perform on the quantum computer unit (QCU). For many of the smaller size processors out there, the high-level aspects of the layer (operating system, architecture and above) are not integrated into the system and instead have been done offline—sometimes by hand. As the number of qubits in the processor grows, the integration of these higher-level aspects becomes critical and quite limiting if it is not done appropriately. Optimizations need to be done both with the algorithm and operating system to minimize the effects of the noisy physical system the program will run on. The coherence times of the qubits, quality of the gates, and measurements will ultimately limit the size of the computation one can perform. One will reach the stage where error always occurs—limiting the usefulness of the NISQ processor for real tasks. With a 100-qubit processor performing 100 gates on each qubit, an error as small 10^{-4} still makes it almost certain that the computation will have errors in it.

The NISQ processors and simulators are an important step in the development of large-scale universal quantum computers. They have shown us quantum advantage—where the quantum processor even using a modest number of qubits can do something faster than today's supercomputers using trillions of transistors. This has demonstrated the potential of the quantum approach. More importantly however, these NISQ processors have allowed us to focus on the overall system layer design and how it operates in practice.

3. Error-corrected quantum computers

While error-mitigation techniques can be used to suppress errors to a certain degree (maybe by several orders of magnitude), it seems unlikely that those techniques will scale to much larger processors. One has to establish the means to cope with the noisy nature of the processor. Quantum error correction is an essential tool to handle this but it comes with a large resource cost (due to the necessity of encoding quantum information at the logical level across many physical qubits). As such we immediately notice the blank area in Fig. 1 between the NISQ processor and

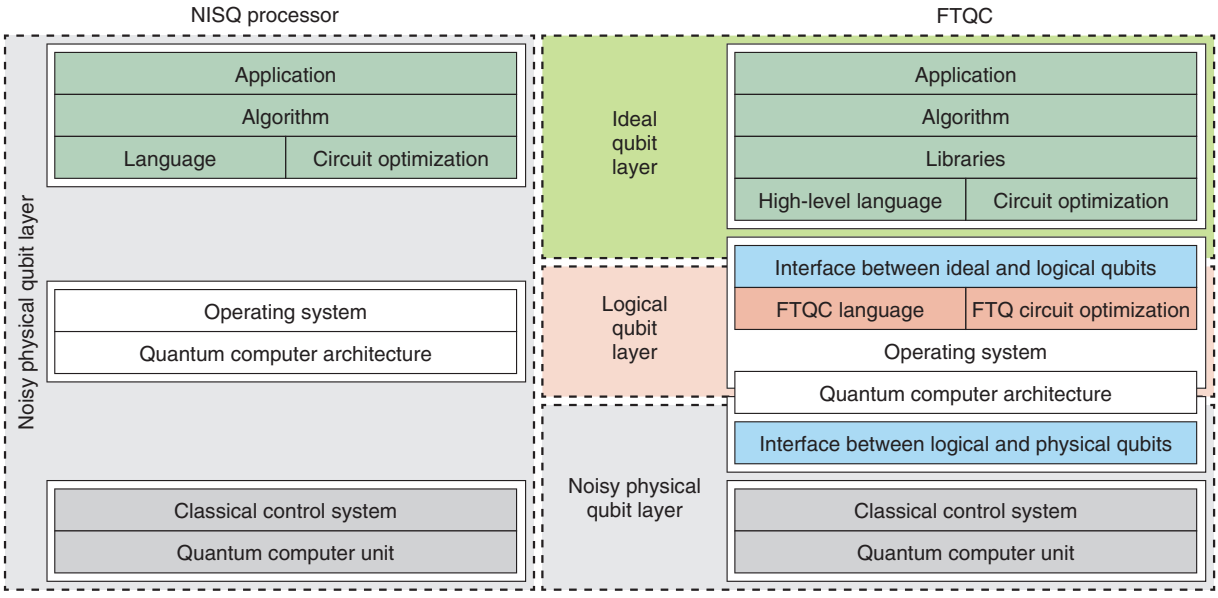


Fig. 2. Schematic diagram showing the quantum computational "system layers" for both a NISQ processor and future large-scale FTQC. These "layers" show the necessary components from the basic hardware to top-level application. The NISQ processor, the simpler of the two, begins at the lowest level with a QCU involving quantum bits and the necessary mechanism to manipulate them including initialization, controlled dynamics, and measurement and feedback. All such elements are in principle noisy, which directly leads to errors in the computation. The QCU is operated on through a CCS. Above this is the processor architecture and an operating system. At the top of the layer (highest level) we have our "application," which needs to be written as an algorithm using libraries written in a quantum language with optimization possible. Typically, noisy operations within the QCU limits the size of such processors. The FTQC, on the other hand, has a much more detailed and structured "system layer" arrangement due to the large size of the processor and the requirement to handle errors from the noisy QCU. This involves three layers: the noisy physical qubit layer, the logical qubit layer, and finally the ideal qubit layer. Each layer contains a number of components, and their integration is necessary for the FTQC to operate correctly. An interface is necessary between the various layers. Next the overall system design needs to be considered together as choices in one layer can have a profound effect on another one. Design changes in the upper-level logical qubit layer can force changes in the noisy physical qubit layer and vice versa.

error-corrected quantum computers. The gap could be many orders of magnitude if one needs to handle quite noisy physical qubits (e.g., 0.1% error probability). One should be operating in the regime in which the use of error correction (and the noisy qubits and gates) does not cause more error than it can fix. Therefore, there may be a number of applications in which a limited amount of error correction helps before full fault tolerance is needed (but with error propagation occurring). This is a largely unexplored but interesting regime. Error-corrected quantum computers should provide a natural bridge to FTQCs, which we consider next.

4. FTQCs

It is important to begin by defining what we mean by an FTQC. It is a more specific form of an error-corrected quantum computer that is able to run any form of computation of arbitrary size without having to change its design. It requires quite a large redesign of the NISQ layer structure. In fact, we need to split it into at least three distinct parts: the noisy physical qubit layer, logical qubit layer, and ideal qubit layer. These layers need to work in unison.

The highest-level layer (shown in green) is similar to that shown in the NISQ approach but is assumed to operate on ideal qubits but with libraries and a high-level language added to it. The purpose of the libraries is to provide useful subroutines the algorithm may need to use. The algorithm is then converted into a sequence of ideal gates and measurement operations (ideal quantum circuits). Circuit-optimization techniques can also be applied to reduce the number of ideal qubits and temporal resources required. It is important to mention that this layer is like a quantum

virtual machine and is agnostic to what hardware it is running on. Ideal qubits can have any quantum gate applied to them.

The middle layer is associated with logical qubits and their manipulation. Logical qubits are different from the ideal qubits mentioned above as only a restricted set of gate operations can be applied to them. This is a very important difference. Now, the middle layer acts as an interface to the ideal qubit layer and converts the quantum circuits (code) developed there into operations associated with logical qubits (with the restricted quantum gate sets etc.). These logical qubits are based on fault-tolerant quantum-error-correction codes operating well below threshold. This layer determines which quantum error codes will be used. Associated with these logical qubits are a fault-tolerant model of computation (e.g., braiding and lattice surgery) and a language to describe them [13]. The operation on those logical qubits is then decomposed into a set of physical operations that is passed to the noisy physical qubit layer.

The lowest layer (the noisy physical qubit layer) is similar to that seen in the NISQ processor but its operation throughout the computer is much more regular and uniform in an FTQC. It takes the physical qubit operations passed to it by the middle layer and establishes how they can be performed using the layout and connectivity of the hardware device. The operating system will turn them into a set of physical instruction signals/pulses etc. that the CCS will perform on the QCU.

While these layers have been presented separately, they must work seamlessly together in the FTQC [9]. One cannot assume that small changes within a layer will not significantly affect the other layers. The choice of the quantum-error-correction code within the middle layer for instance puts constraints on the computer architecture and quantum gates being performed within the QCU. Moving to a different code may require a completely different computing architecture.

5. Distributed quantum computers

A key aspect of the noisy physical qubit layer is the quantum computer architecture and the layout/connectivity of qubits and control lines. Like we have seen in the conventional computer world, the size of the monolithic processor becomes a bottleneck to performance. The solution was to go with a multicore approach. We expect a similar bottleneck to occur within our quantum hardware, so we can use a distributed approach with which small quantum processors (cores) are connected together to create larger ones. This modular approach has a number of distinct advantages including its ability to give long-range connectivity between physical qubits [14, 15]. Such modules could accommodate a few qubits through to thousands. This is a design choice, and the optimal size is currently unknown.

6. Discussion

In the design of FTQCs, one must not look only at what occurs within the layers individually. Optimizations in the high-level layer can have a significant effect on the resources required in the middle logical qubit layer, even decreasing the distance of the quantum-error correcting needed [13]. This in turn would mean fewer qubits and gates required at the physical qubit layer. Furthermore, one must be careful not to look at the computing system too abstractly, especially within the lowest layer. One must understand the properties of the physical systems from which our qubits are derived. The choice of physical system and how it is controlled will have a profound effect on the other layers. The fault-tolerant error-correction threshold heavily depends on the structure of the noise the qubits experience in reality.

References

[1] J. P. Dowling and G. J. Milburn, "Quantum Technology: the Second Quantum Revolution," Phil. Trans. R. Soc. Lond. A, Vol. 361, No. 1809, 2003.

[2] V. Giovannetti, S. Lloyd, and L. Maccone, "Quantum-enhanced Measurements: Beating the Standard Quantum Limit," Science, Vol. 306, No. 5700, pp. 1330–1336, 2004.

[3] N. Gisin and R. Thew, "Quantum Communication," Nature Photon., Vol. 1, pp. 165–171, 2007.

[4] R. P. Feynman, "Simulating Physics with Computers," Int. J. Theor. Phys., Vol. 21, pp. 467–488, 1982.

[5] W. J. Munro, K. Azuma, K. Tamaki, and K. Nemoto, "Inside Quantum Repeaters," IEEE J. Sel. Top. Quantum Electrons., Vol. 21, No. 3, 6400813, 2015.

[6] K. Azuma, S. Bäuml, T. Coopmans, D. Elkouss, and B. Li, "Tools for Quantum Network Design," AVS Quant. Sci., Vol. 3, No. 1, 014101, Feb. 2021.

[7] S. M. Brewer, J.-S. Chen, A. M. Hankin, E. R. Clements, C. W. Chou, D. J. Wineland, D. B. Hume, and D. R. Leibrandt, "^{27}Al+ Quantum-logic Clock with a Systematic Uncertainty below 10^{-18}," Phys. Rev. Lett., Vol. 123, 033201, 2019.

[8] J. Preskill, "Quantum Computing in the NISQ Era and Beyond," Quantum, Vol. 2, 79, 2018.

[9] K. Nemoto, S. Devitt, and W. J. Munro, "Noise Management to Achieve Superiority in Quantum Information Systems," Phil. Trans. R. Soc. A, Vol. 375, No. 2099, 20160236, 2017.

[10] J. Kelly, "A Preview of Bristlecone, Google's New Quantum Processor," ai.googleblog.com/2018/03/a-preview-of-bristlecone-googles-new.html

[11] F. Arute, K. Arya, R. Babbush, D. Bacon, J. C. Bardin, R. Barends, R. Biswas, S. Boixo, F. G. S. L. Brandao, D. A. Buell, B. Burkett, Y. Chen, Z. Chen, B. Chiaro, R. Collins, W. Courtney, A. Dunsworth, E. Farhi, B. Foxen, A. Fowler, C. Gidney, M. Giustina, R. Graff, K. Guerin, S. Habegger, M. P. Harrigan, M. J. Hartmann, A. Ho, M. Hoffmann, T. Huang, T. S. Humble, S. V. Isakov, E. Jeffrey, Z. Jiang, D. Kafri, K. Kechedzhi, J. Kelly, P. V. Klimov, S. Knysh, A. Korotkov, F. Kostritsa, D. Landhuis, M. Lindmark, E. Lucero, D. Lyakh, S. Mandrà, J. R. McClean, M. McEwen, A. Megrant, X. Mi, K. Michielsen, M. Mohseni, J. Mutus, O. Naaman, M. Neeley, C. Neill, M. Yuezhen Niu, E. Ostby, A. Petukhov, J. C. Platt, C. Quintana, E. G. Rieffel, P. Roushan, N. C. Rubin, D. Sank, K. J. Satzinger, V. Smelyanskiy, K. J. Sung, M. D. Trevithick, A. Vainsencher, B. Villalonga, T. White, Z. J. Yao, P. Yeh, A. Zalcman, H. Neven, and J. M. Martinis, "Quantum Supremacy Using a Programmable Superconducting Processor," Nature, Vol. 574, pp. 505–510, 2019.

[12] H. S. Zhong, H. Wang, Y. H. Deng, M. C. Chen, L. C. Peng, Y. H. Luo, J. Qin, D. Wu, X. Ding, Y. Hu, P. Hu, X. Y. Yang, W. J. Zhang, H. Li, Y. Li, X. Jiang, L. Gan, G. Yang, L. You, Z. Wang, L. Li, N. L. Liu, C. Y. Lu, and J. W. Pan, "Quantum Computational Advantage Using Photons," Science, Vol. 370, No. 6523, pp. 1460–1463, 2020.

[13] M. Hanks, M. P. Estarellas, W. J. Munro, and K. Nemoto, "Effective Compression of Quantum Braided Circuits Aided by ZX-Calculus," Phys. Rev. X, Vol. 10, No. 4, 041030, 2020.

[14] K. Azuma, H. Takeda, M. Koashi, and N. Imoto, "Quantum Repeaters and Computation by a Single Module: Remote nondestructive parity measurement," Phys. Rev. A, Vol. 85, No. 6, 062309, 2012.

[15] K. Nemoto, M. Trupke, S. J. Devitt, A. M. Stephens, B. Scharfenberger, K. Buczak, T. Nöbauer, M. S. Everitt, J. Schmiedmayer, and W. J. Munro, "Photonic Architecture for Scalable Quantum Information Processing in Diamond," Phys. Rev. X, Vol. 4, No. 3, 031022, 2014.

William John Munro
Senior Distinguished Scientist, Theoretical Quantum Physics Research Group, NTT Basic Research Laboratories.

Bill received a B.Sc in chemistry, M.Sc and D.Phil in physics from the University of Waikato, New Zealand, in 1989, 1991, and 1995. After several years in the computing industry, he returned to physics as a research fellow at the University of Queensland, Australia, from 1997 to 2000 before becoming a permanent staff scientist at Hewlett Packard Laboratories in Bristol, UK (2000–2010). He joined NTT Basic Research Laboratories in 2010 and was promoted to senior distinguished scientist in 2016. His research interests range from foundational issues in quantum science through to the design and development of quantum technology. He is a fellow of the Institute of Physics (IOP), American Physical Society (APS), and the Optical Society (OSA).

Victor M. Bastidas
Research Scientist, Theoretical Quantum Physics Research Group, NTT Basic Research Laboratories.

He received a BSc and MSc. in physics from Universidad del Valle, Colombia, in 2006 and 2009. In 2013 he received a Ph.D. (Dr. rer. nat.) from the Technical University of Berlin, Germany. He joined NTT Basic Research Laboratories in 2017 as a research specialist and in 2019 as a permanent research scientist. Since 2020, he has been a visiting associate professor in the group of Prof. Kae Nemoto at the National Institute of Informatics. He is a member of the German Physical Society.

Koji Azuma
Distinguished Researcher, Theoretical Quantum Physics Research Group, NTT Basic Research Laboratories.

He received a B.E., M.E., and Ph.D. in physics from Osaka University, the University of Tokyo, and Osaka University in 2005, 2007, and 2010, respectively. He joined NTT Basic Research Laboratories in 2010. He has been a joint appointment researcher of the PRESTO, Japan Science and Technology Agency since 2018, and a guest associate professor of Graduate School of Engineering Science, Osaka University, since 2019. He is a member of the Physical Society of Japan.

Kae Nemoto
Professor, National Institute of Informatics.

She is a full professor at the National Institute of Informatics (NII) and the Graduate University for Advanced Studies (SOKENDAI), Tokyo. Further she is also the director of the Global Research Center for Quantum Information Science, as well as being the co-director of the Japanese-French Laboratory for Informatics (JFLI). Her research is focused on applications for quantum computers, quantum computer architectures, quantum-error correction, quantum networks, and the quantum internet. She also leads a newly established academic education consortium "Mastering quantum technology" for the next generation of scientist and engineers in the field. Kae is one of the founders of the Quantum ICT forum in Japan, where she currently serves as the board vice director. Finally, she is a Fellow of both the IOP and APS.

Theoretical Approach to Overcome Difficulties in Implementing Quantum Computers

Seiseki Akibue, Yuki Takeuchi, Yasuhiro Takahashi, Go Kato, and Seiichiro Tani

Abstract

To develop large-scale fault-tolerant quantum computers, implementation technologies are required to meet extremely strict conditions. Basic research targeting such technologies is currently being conducted worldwide, and theoretical research can also make a relevant contribution. In this article, we introduce several theoretical research studies for deriving the maximum power of quantum computers with limited resources and/or restricted functions due to difficulties in physical implementation.

Keywords: quantum computers, quantum computing, quantum information processing

1. Quantum computers with limited resources and/or restricted functions

Expectations for the outstanding potential of quantum computers have surged recently; therefore, national projects, startups, and major companies are competing fiercely in developing quantum computers. In the near future, however, we can expect only *restricted* quantum computers in terms of the amount of computational resources and variety of available functions rather than full-fledged quantum computers. This is because large-scale fault-tolerant quantum computers require implementation technologies that meet extremely strict conditions. To develop such technologies as soon as possible, various studies on basic research are currently being conducted worldwide.

A relevant theoretical approach is to clarify with theoretical knowledge the limitation of the computational power of quantum computers with limited resources and/or restricted functions due to difficulties in physical implementation. In this article, we introduce several theoretical research studies for deriving the maximum power of such restricted quantum computers.

2. Overcoming difficulty in reducing noise

Current assumed quantum computers prepare and initialize a set of qubits then successively execute operations on one or two qubits. After measuring the resulting state, we obtain the output bit-sequence (**Fig. 1**). By designing these operations on the qubits

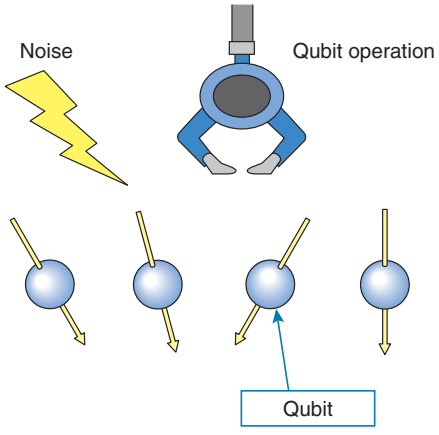


Fig. 1. The current model of quantum computers.

depending on the problem, it is possible to make the output provide information that solves the particular problem. For this purpose, the qubits must be unaffected by noise. A standard method for achieving this is using quantum error-correcting codes for large-scale quantum computers. This method involves encoding information on a single logical qubit with multiple physical qubits in a redundant manner, which can protect quantum information from noise. To use such a method, the noise level for physical qubits must be below a certain level. However, this requirement has not yet been met.

The fact that we can manipulate some qubits directly means that those qubits will be affected by the environment (e.g., noise). In other words, an abundance of controllability means that there is also a path along which noise may enter. Therefore, if an object that can handle quantum information has limited degrees of freedom to manipulate directly, the effect of noise will be small. However, quantum computers originate from ordinary computers; thus, we assume we can apply many types of operations directly for qubits in quantum computers. As a result, practical quantum information processing under restricted controllability has not been considered.

We theoretically investigated a situation in which only restricted operations can be applied [1], which would be useful when we limit controllability to reduce noise. This situation is as follows (indirect quantum control (**Fig. 2**)). We prepare two quantum systems. One is an internal system, which is directly incontrollable, and the other is an external system, which can be directly manipulated at will. Quantum information is transferred back and forth between the two systems through a fixed interaction. We found that the internal system can be divided into two parts. The first part affects the external system and the second part does not. Moreover, if the dimension of the external system is more than 3, the first part can be indirectly manipulated at will. In other words, any indirectly controlled quantum system has sufficient ability as a quantum information processor when the dimension of the directly controllable system (i.e., external system) is more than 3. However, this investigation only focused on the possibility or impossibility of achieving indirect quantum control and did not provide an answer to more in-depth questions such as how long it takes to implement the desired indirect quantum control. Since indirect quantum control has never been systematically analyzed, our investigation serves as a theoretical foundation for indirect quantum control, and further investigation is needed to answer more practical questions. In this sense, we have shown that indirect quantum control has the potential for denoising, but the specific process of denoising must be discovered through future investigations.

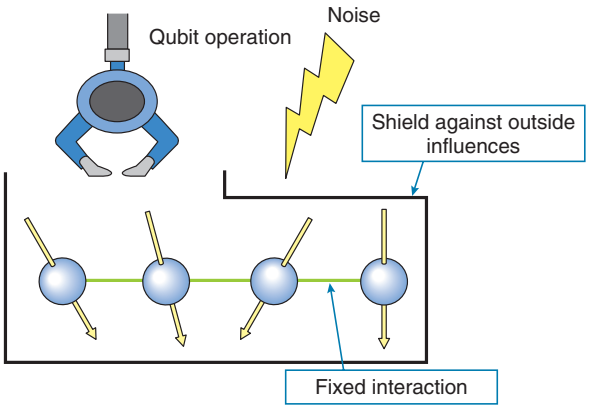


Fig. 2. The model of quantum information processor by indirect quantum control.

3. Overcoming difficulty in achieving quantum memory

A quantum computer sequentially repeats two procedures: measuring a highly controlled quantum system[*1] in an appropriate order and controlling the system on the basis of the previous computation results. Since quantum measurements inevitably disrupt the quantum system even if they are carried out accurately, the order of measurements is strictly restricted depending on the algorithm. Thus, the ability of changing the order of quantum measurements enables the design of many quantum algorithms. It also improves the key rate of the quantum key distribution, a highly secure cryptography.

A long (waiting) time until measurements are carried out, however, requires quantum memory, a mechanism to protect a quantum system from environmental noise. Although many methods of implementing quantum memory have been proposed, the number of measurements that can be executed within the memory time is much less than that required for a large-scale quantum computer.

We can carry out certain types of measurements that depend on previous computation results, i.e.,

*1 Quantum system: A physical system, such as a photon and electron, where quantum mechanical phenomena appear.

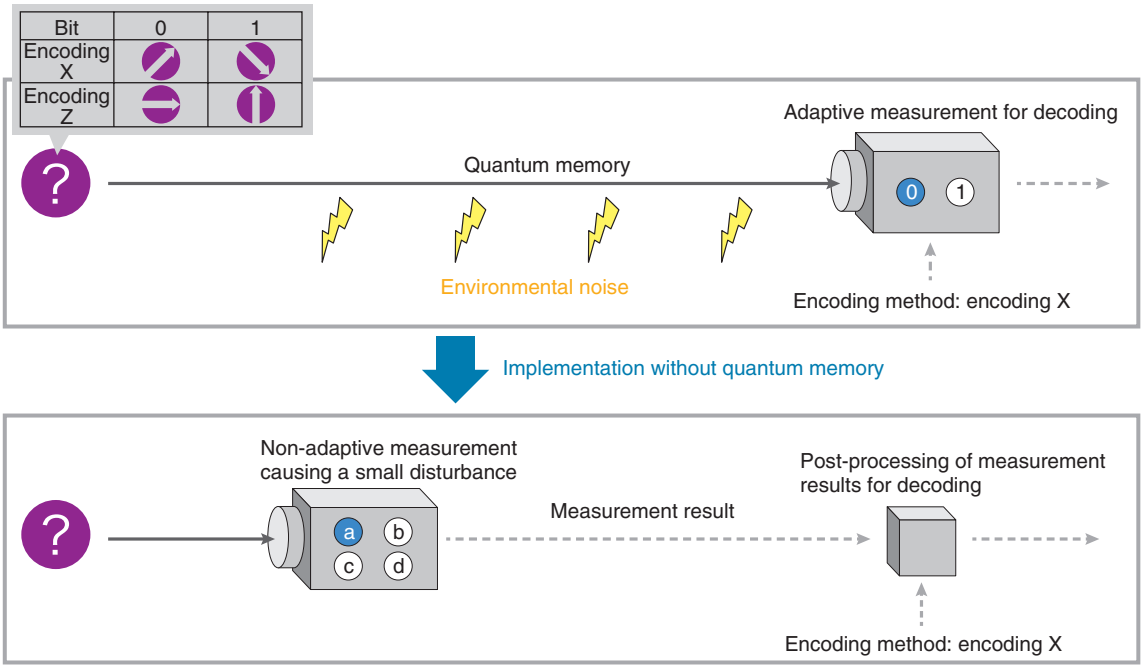


Fig. 3. Decoding one bit without using quantum memory.

adaptive measurements, before knowing the previous computation results and without disturbing future computation results, which makes quantum memory unnecessary. However, there is no known example of quantum-information-processing tasks benefitting from such adaptive measurements.

In our previous study [2], we found an application of adaptive measurements not requiring quantum memory in state discrimination, which is used as a subroutine in many quantum-information-processing tasks. Consider decoding one bit by measuring a quantum state that encodes the bit by using an encoding method randomly chosen from a pair of encoding methods. Intuitively, we may fail to decode a bit perfectly unless we carry out adaptive measurement depending on the encoding method. However, it is possible to decode the bit perfectly by using only non-adaptive measurement and simple post processing for certain pairs of encoding methods, as shown in **Fig. 3**. Since not all pairs of encoding methods enable perfect decoding, we derive succinct criteria for a pair of encoding methods to be perfectly decodable. Such perfectly decodable pairs can be used as a quantum-key distribution protocol with a higher key rate than a widely used encoding method if we can ignore communication-channel noise. However, since channel noise is not negligible in practical communication between distant parties, we need to conduct a detailed analysis of the key rate in practical channel noise.

4. Overcoming difficulty in initializing qubits

Fast quantum algorithms are required to achieve high-speed computation on quantum computers. Such algorithms are usually designed under the assumption that many qubits initialized to state 0 are available. These initialized qubits are useful for storing various intermediate results during a computation, which, for example, allows us to design highly parallel algorithms. Therefore, the availability of many initialized qubits significantly contributes to designing fast quantum algorithms.

Although initializing many qubits plays a key role in designing fast quantum algorithms, preparing such qubits is beyond the reach of current implementation technology. In fact, there is a limit on initialization accuracy; thus, when many qubits are initialized using current technology, some of the qubits can sometimes be in unintended states. Limiting the number of qubits to be initialized increases feasibility. However, with only a small number of initialized qubits, it is difficult to design fast quantum algorithms.

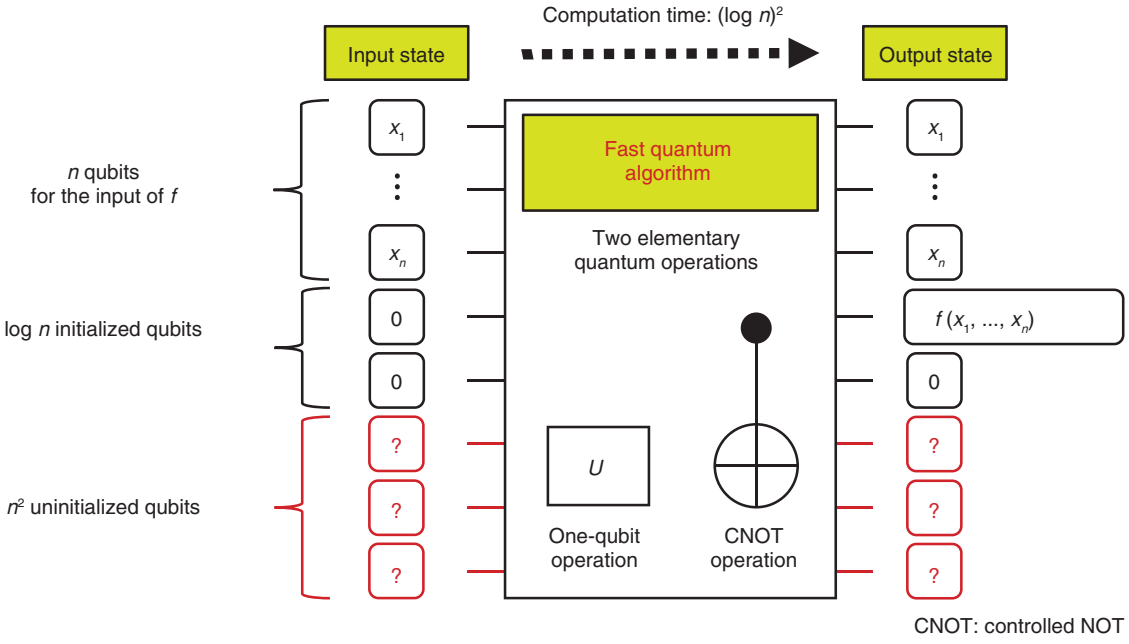


Fig. 4. Computing function f: $\{0,1\}^n \rightarrow \{0,1\}$ with many uninitialized qubits and a small number of initialized qubits.

We therefore focus on uninitialized qubits as computational resources for designing fast quantum algorithms together with a small number of initialized qubits. The states of uninitialized qubits are unknown, but, as with common qubits, their states can be changed by applying quantum operations. Since uninitialized qubits do not require initialization by definition, they are relatively easy to prepare.

Under the assumption that many uninitialized qubits and a small number of initialized qubits are available (**Fig. 4**), we designed fast quantum algorithms for computing symmetric Boolean functions (e.g., the logical OR function), which are key ingredients of more complicated quantum algorithms [3]. However, it seems difficult to design fast quantum algorithms for the same functions with only a small number of initialized qubits. Therefore, our algorithms indicate that the use of uninitialized qubits significantly contributes to designing fast quantum algorithms.

5. Overcoming restrictions on architectures

Several quantum computing models with functionality restricted to improve their feasibility have recently been proposed. In this section, by using the Fourier hierarchy, a hierarchy of quantum circuits, we introduce our previously proposed quantum computing model Hadamard-classical circuit with one-qubit (HC1Q) [4], as shown in **Fig. 5(a)**. In HC1Q, the basis transformations with Hadamard gates H[*2] are executed before and after the coherent classical computation. Note that no basis transformations are executed for the lowest qubit. As the level of the Fourier hierarchy becomes higher, the computational capability of quantum circuits also becomes higher. In particular, all quantum circuits contained in the first level can be efficiently simulated with classical computers. Therefore, to demonstrate the superiority of quantum computing over classical counterparts (i.e., quantum computational supremacy), we have to use quantum circuits in the second or higher level. Since HC1Q has quantum computational supremacy even though it is in the second level of the Fourier hierarchy, it can be considered as one of the most restricted quantum computing models with advantage over classical computers. More formally, we have shown that if HC1Q can be efficiently simulated with a classical computer, then the polynomial hierarchy, a concept in computational complexity theory[*3], collapses to its second level. Polynomial hierarchy is a hierarchy of decision problems that can be answered with

*2 Hadamard gate: A basic quantum gate applied on a single qubit.
*3 Computational complexity theory: A discipline systematically studying the hardness of problems. The P≠NP conjecture is the most common open problem in computational complexity theory.

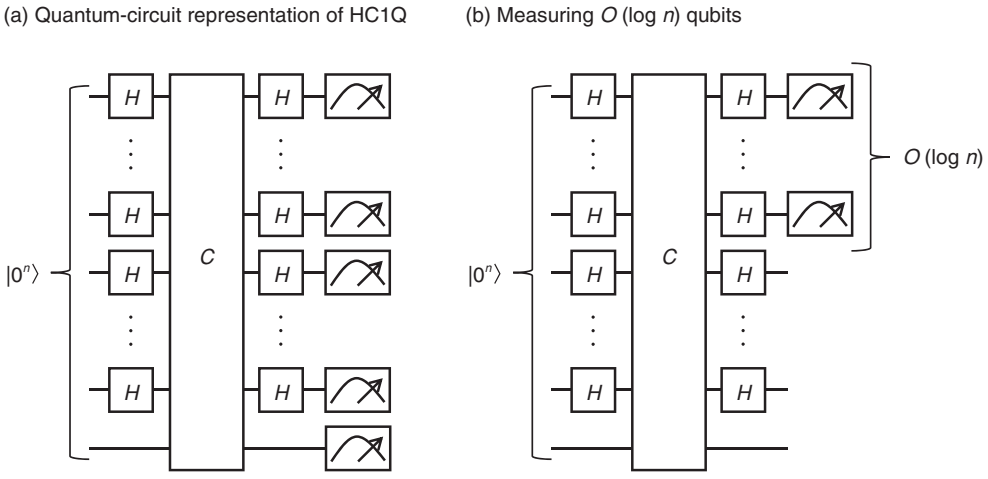


Fig. 5. Quantum computing model in the second level of the Fourier hierarchy. (a) Quantum-circuit representation of HC1Q. (b) Modified version, where the number of measured qubits is decreased.

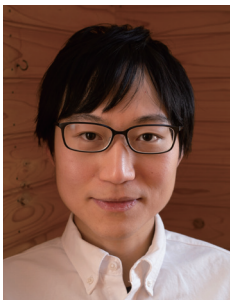
YES or NO and is strongly believed to not collapse. By associating two completely different hierarchies, i.e., the Fourier and polynomial hierarchies, we have shown the quantum computational supremacy of HC1Q. As shown in Fig. 5(a), all input qubits are measured in HC1Q. What happens if only a small number of qubits are measured? Interestingly (see **Fig. 5(b)**), the computational capability becomes equivalent to or strictly less than that of classical computers.

6. Outlook

Research on both software (i.e., algorithms) and hardware for quantum computers is essential for high-speed quantum computing. With our theoretical expertise, we will explore designing algorithms that solve important problems very fast on quantum computers as well as develop theoretical methods, such as those discussed in this article, to maximally extract computational power from quantum computer hardware.

References

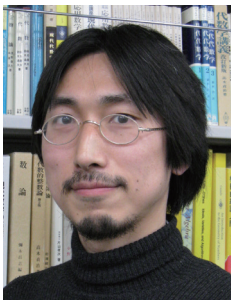
[1] G. Kato, M. Owari, and K. Maruyama, "Algebra and Hilbert Space Structures Induced by Quantum Probes," Ann. Phys., Vol. 412, 168046, 2020.
[2] S. Akibue and G. Kato, "Perfect Discrimination of Nonorthogonal Quantum States with Posterior Classical Partial Information," Phys. Rev. A, Vol. 99, No. 2, 020102, 2019.
[3] Y. Takahashi and S. Tani, "Power of Uninitialized Qubits in Shallow Quantum Circuits," Theor. Comput. Sci., Vol. 851, pp. 129–153, Jan. 2021.
[4] T. Morimae, Y. Takeuchi, and H. Nishimura, "Merlin-Arthur with Efficient Quantum Merlin and Quantum Supremacy for the Second Level of the Fourier Hierarchy," Quantum, Vol. 2, 106, 2018.



Seiseki Akibue
Research Scientist, Computing Theory Research Group, Media Information Laboratory, NTT Communication Science Laboratories.
He received a doctor of science in physics from the University of Tokyo in 2016. He joined NTT Communication Science Laboratories the same year and has been engaged in theoretical topics in quantum information and quantum computation. He is especially interested in asymptotic structures appearing in quantum mechanics.

Yuki Takeuchi
Researcher, Computing Theory Research Group, Media Information Laboratory, NTT Communication Science Laboratories.
He received a doctor of science from Osaka University in 2018 and joined NTT Communication Science Laboratories as a research associate the same year. He has been engaged in the theoretical investigation of quantum information and is especially interested in the verifiability of quantum computing. He began his current position in 2019. He is a member of the Physical Society of Japan.

Yasuhiro Takahashi
Senior Research Scientist, Computing Theory Research Group, Media Information Laboratory, NTT Communication Science Laboratories.
He received a Ph.D. in engineering from the University of Electro-Communications, Tokyo, in 2008. He joined NTT Communication Science Laboratories in 2000 and has been engaged in research on the design and optimization of quantum circuits. His research interests include quantum computing, computational complexity theory, and cryptography. He is a member of the Information Processing Society of Japan (IPSJ) and the Institute of Electronics, Information and Communication Engineers (IEICE).



Go Kato
Senior Research Scientist, Computing Theory Research Group, Media Information Laboratory, NTT Communication Science Laboratories.
He received a doctor of science in physics from the University of Tokyo in 2004 and joined NTT Communication Science Laboratories the same year, where he has been engaged in the theoretical investigation of quantum information. He is especially interested in mathematical structures emerging in the field of quantum information. He is a member of the Physical Society of Japan.

Seiichiro Tani
Distinguished Scientist, Computing Theory Research Group, Media Information Laboratory, NTT Communication Science Laboratories.
He received a B.E. from Kyoto University in 1993 and an M.S. and a Ph.D. in information science and technology from the University of Tokyo in 1995 and 2006. He joined NTT LSI Laboratories in 1995. Since 2003, he has been engaged in research on theory of quantum computing at NTT Communication Science Laboratories. He is a member of IEICE, IPSJ, IEEE (Institute of Electrical and Electronics Engineers), and Association for Computing Machinery.

Fault-tolerant Technology for Quantum Information Processing and Its Implementation Methods

Yuuki Tokunaga, Yasunari Suzuki, Suguru Endo, and Rui Asaoka

Abstract

To use quantum information processing in a wide range of applications, fault-tolerant processing is essential to cope with noise. Fault-tolerant quantum computing using quantum error-correcting codes is scalable, but the overhead of number of qubits or processes is large, so improving efficiency is an important research theme. Quantum error mitigation incurs computational cost, but it does not require the overhead of number of qubits, so it is expected to be used for near-future applications. In this article, we introduce our research activities on these topics as well as our studies toward the implementation of quantum information processing.

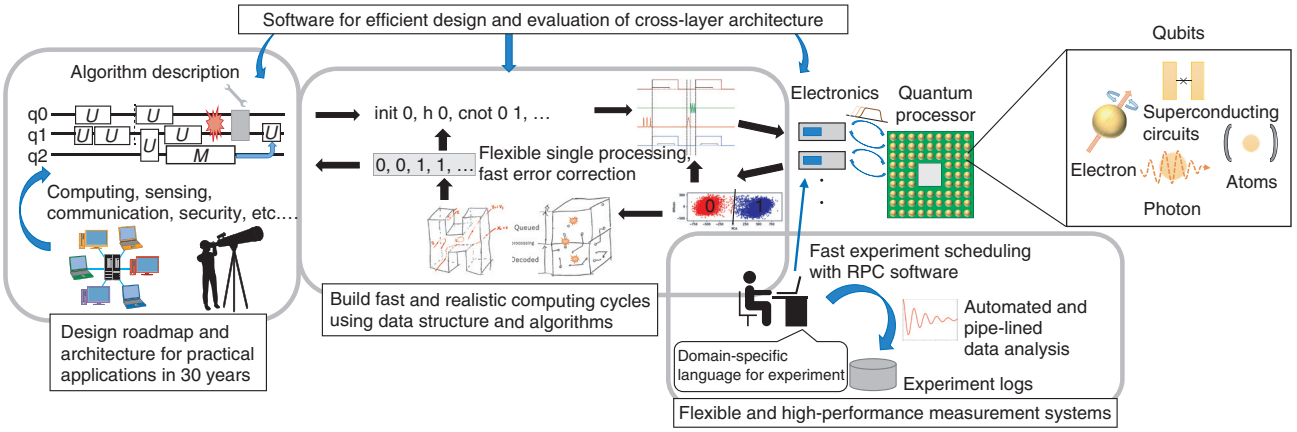
Keywords: quantum information processing, error correction, error mitigation

1. Introduction

The potential of quantum computers and quantum networks has been attracting attention recently. Fault-tolerant processing is the most essential technology for this because there is a significant difference in the amount of noise that can occur between quantum information processing and current information processing (classical information processing). In classical information processing, errors occur less frequently, so most information processing can be done without much concern for errors. However, in quantum information processing, errors occur so frequently that only small-scale computations can be executed without handling the errors. To make quantum information processing scalable, execute various useful quantum computations, and develop a large-scale quantum network, it is necessary to carry out fault-tolerant processing for quantum information and protect quantum information from noise. In quantum information processing, therefore, fault tolerance is expected to become a fundamental part of computer and network architecture technology. We are investigating fault-tolerant processing as an important fundamental technology to protect quantum information from errors and noise.

There are two major error-suppression methods for quantum information processing: using quantum error-correcting codes and quantum error mitigation. Quantum error-correcting codes redundantly encode one logical qubit into multiple physical qubits that can be corrected by decoding even if some errors occur. Fault-tolerant quantum computing using quantum error-correcting codes is the only method known to be capable of scalable quantum computing with fault tolerance and is expected to be essential for the practical application of large-scale quantum information processing in the future. However, the number of qubits and processes required for encoding and decoding will inevitably increase; therefore, it will take a long time to achieve large-scale fault-tolerant quantum computing.

Another method that has been attracting attention recently is the quantum error-mitigation method. This is a method of removing errors from the computational results by predicting the correct calculation



RPC: remote procedure call

Fig. 1. Concept of fault-tolerant quantum computing and its software infrastructure.

results without the overhead of number of qubits. Computational cost increases exponentially with the noise level for the prediction of the error-free results; hence, this is not scalable. However, it does not require the overhead of number of qubits, so it is considered an important method while the scale of quantum information processing is still small. It is also important that the elemental quantum gates be as accurate as possible and have high functionality for communication and distributed processing.

We introduce our research activities on software infrastructure for fault-tolerant quantum computing, quantum error-mitigation methods, and basic elements for the implementation of quantum information processing.

2. Research and development for fault-tolerant quantum computing

Quantum computers enable various types of novel information processing by using the superposition of quantum states but at the cost of being vulnerable to noise and slow throughput of unit instructions. In current memories, such as dynamic random access memory, a capacitor is charged or discharged to indicate 0 or 1, and under threshold operation with a sufficient margin, the memory is refreshed earlier than the information is lost by natural discharge. This mechanism enables long-time storage of information with a small overhead. In comparison, quantum memories using a superconducting circuit, which also manipulates electrons, are not only slow in executing basic operations but also cannot be directly refreshed due to the no-cloning theorem. This problem can be avoided by encoding information of qubits using quantum error-correcting codes and by repeatedly detecting and correcting errors indirectly. However, carrying out quantum computations with quantum error correction requires enormous resources: hundreds or thousands of times more in time and size than that without error correction [1]. Thus, in developing quantum computers, the loss of information from a volatile memory must be suppressed by fast feedback with error-correcting codes instead of a simple refresh. To develop a practical fault-tolerant quantum computer, it is essential to construct a system with not only excellent quantum devices but also scalability, broadband interfaces, and robust control. Our group, in collaboration with RIKEN and several universities, is developing a fault-tolerant quantum computer using superconducting qubits with both high performance and high reliability. In this section, we introduce three topics on the software infrastructure of quantum computers, an overview of which is illustrated in **Fig. 1**.

The first topic is a system to efficiently characterize and calibrate a large number of superconducting qubits on a chip. Currently, superconducting qubits must be individually controlled due to their inhomogeneity; thus, the measurement system becomes a huge distributed system consisting of a large number of field-programmable gate arrays and microwave sources. We constructed basic software for asynchronous control of such devices and automatic and fast calibration of a large number of qubits. We also proposed efficient methods for characterization and

calibration using machine learning and quantum randomness [2, 3].

The second topic is the design of peripheral circuits that execute the operations and feedback required for quantum error correction. The maximum likelihood estimation of errors in a high-performance quantum error-correcting code can be reduced to a graph problem called minimum-weight perfect matching. However, naive algorithms to solve this problem are not practical since they have large latency. Our group is working on a method for optimizing decoding circuits for small codes using machine learning [4] and on a fast decoder using single-flux-quantum circuits [5].

The third topic is a software toolchain to translate programs across multiple technology layers to evaluate the performance of architectures and simulate their behavior. In particular, our quantum circuit simulator, which is a part of the toolchain, became the world's fastest for several benchmarks [6]. As mentioned above, the development of a large-scale fault-tolerant quantum computer is a battle to control integrated quantum devices while ensuring high performance and reliability with optimized control devices and algorithms using a large-scale distributed system. This is a challenging task that is suitable for NTT laboratories, who have been studying these technologies.

3. Quantum error-mitigation method

In October 2019, Google announced that a quantum computer was able to solve a problem that would have taken 10,000 years to solve with a classical computer in 200 seconds using a small-scale quantum device with 53 qubits [7]. Small-scale, noisy quantum computers such as the one used in this demonstration are called noisy intermediate-scale quantum (NISQ) computers. There is a counterargument that the problem solved by Google can be solved in a few days using a supercomputer, not 10,000 years. However, as the size of quantum computers becomes larger, this gap is expected to widen further. Researchers worldwide are studying how to use such small-scale quantum devices practically. They are mainly expected to be applied to chemical calculations and machine learning.

However, the computational errors of such small-scale quantum computers cannot be ignored. Historically, quantum error *correction* has been studied as a method of suppressing errors. In quantum error correction, multiple qubits are used to encode one logical qubit, and the error is suppressed exponentially with respect to the number of qubits used for encoding. However, this method is not suitable for NISQ computers, which have a limited number of qubits. Therefore, quantum error *mitigation* has been extensively studied to suppress errors without increasing the burden on the hardware side [8]. Quantum error mitigation does not require the overhead in number of qubits but does require a larger number of measurements.

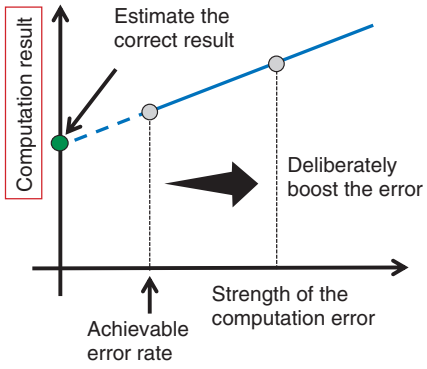


Fig. 2. Quantum error mitigation (extrapolation method).

A variety of error-mitigation methods have recently been proposed. In this section, we first explain the simplest and most straightforward extrapolation method. The computational error of a quantum computer can be increased by adding noisy operations, etc. The extrapolation method is to estimate the error-free result by extrapolating the original result and the result with the increased error. A conceptual diagram of the extrapolation method is shown in **Fig. 2**. There is also a method called the quasi-probability method, which identifies the error model of the computational error and executes an inverse transformation of the error to effectively cancel it [9]. There are also other methods such as the symmetry verification method and subspace expansion method. For further details, please refer to the review paper [8].

With the expectation on an increase in the number of qubits in the future, our group showed for the first time that quantum error mitigation is useful for further improving computational accuracy, even after fault-tolerant quantum computing using quantum error correction becomes possible to some extent [10], and that quantum error correction and quantum error mitigation should not be considered separately but should be considered as an integrated scheme.

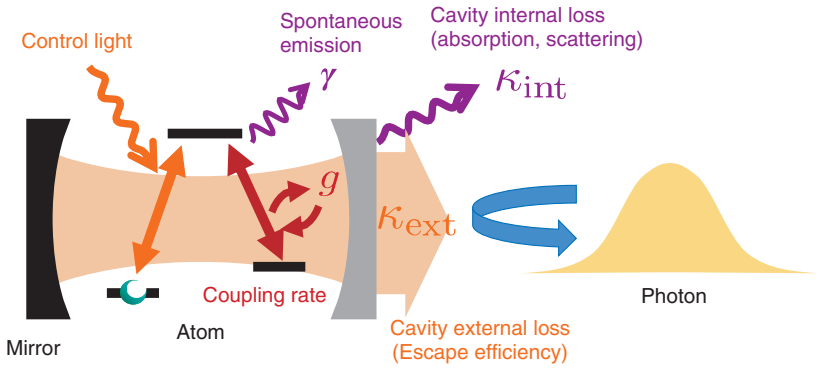


Fig. 3. Atom-photon interaction using cavity QED.

4. Implementation schemes for basic elemental technologies

Thus far, we have introduced how to effectively implement fault-tolerant processing from the software side for given hardware. It is also important for fault tolerance that the implementation schemes of the basic elements function with high accuracy and have scalability. It is also important that the functions become more sophisticated for communication and distributed processing. In particular, various elemental technologies based on the interaction between atoms and photons using cavity quantum electrodynamics (QED) (**Fig. 3**) are important for expanding the functions for quantum computing and quantum networks. NTT Secure Platform Laboratories is studying such implementation schemes for future quantum secure-network infrastructures including quantum repeaters.

It is first important to generate photons, which are used as qubits for quantum communication, with high efficiency and scalability. There are problems in scalability of current single-photon generation methods using a weak laser light or nonlinear crystals because of the unwanted multiple-photon generation events. Deterministic single-photon generation is possible with high efficiency using single-atom cavity QED systems [11]. We investigated the improvement in the repetition rate by shortening the pulse length, so that the overall performance of information processing can be improved [12]. When atoms are used as stationary qubits and photons are used as communication qubits, highly accurate quantum gates between atoms and photons are possible by using cavity QED systems. For these gates, we considered the trade-off in fault tolerance in terms of photon loss and gate fidelity and investigated implementation schemes that optimize them [13]. We are investigating circuit QED systems to efficiently execute quantum gates between superconducting qubits and microwave photons or the routing of the microwave photons for distributed processing of superconducting quantum computers [14, 15].

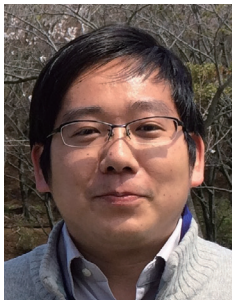
5. Conclusion

We are investigating fault-tolerant quantum computing using error-correcting codes and quantum error-mitigation methods from the viewpoint of software infrastructure. We are also engaged in research from the opposite side, that is, how to improve the accuracy and efficiency of the basic elements of the physical implementation and how to expand its functionality. We are attempting to create a new efficient architecture for a sufficiently large fault-tolerant quantum computer, which is expected to be put to practical use in the far future. In the meantime, we are conducting research on making quantum information processing as meaningful as possible in the near future by using quantum error mitigation and other methods.

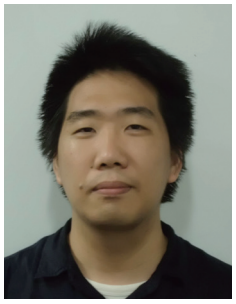
References

[1] A. G. Fowler and C. Gidney, "Low Overhead Quantum Computation Using Lattice Surgery," arXiv:1808.06709, 2018.
[2] K. Heya, Y. Suzuki, Y. Nakamura, and K. Fujii, "Variational Quantum Gate Optimization," arXiv:1810.12745, 2018.
[3] Y. Nakata, D. Zhao, T. Okuda, E. Bannai, Y. Suzuki, S. Tamiya, K. Heya, Z. Yan, Z. Kun, S. Tamate, Y. Tabuchi, and Y. Nakamura, "Quantum Circuits for Exact Unitary t-designs and Applications to Higher-order Randomized Benchmarking," arXiv:2102.12617, 2021.
[4] A. Davaasuren, Y. Suzuki, K. Fujii, and M. Koashi, "General Framework for Constructing Fast and Near-optimal Machine-learning-based Decoder of the Topological Stabilizer Codes," Phys. Rev. Research,

Vol. 2, No. 3, 033399, 2020.
[5] Y. Ueno, M. Kondo, M. Tanaka, Y. Suzuki, and Y. Tabuchi, "Quantum Error Correction with a Superconducting Decoder," 2nd Workshop on Quantum and Classical Cryogenic Devices, Circuits, and Systems (QCCC 2020), Dec. 2020.
[6] Y. Suzuki, Y. Kawase, Y. Masumura, Y. Hiraga, M. Nakadai, J. Chen, K. M. Nakanishi, K. Mitarai, R. Imai, S. Tamiya, T. Yamamoto, T. Yan, T. Kawakubo, Y. O. Nakagawa, Y. Ibe, Y. Zhang, H. Yamashita, H. Yoshimura, A. Hayashi, and K. Fujii, "Qulacs: a Fast and Versatile Quantum Circuit Simulator for Research Purpose," arXiv:2011.13524, 2020.
[7] F. Arute, K. Arya, R. Babbush, D. Bacon, J. C. Bardin, R. Barends, R. Biswas, S. Boixo, F. G. S. L. Brandao, D. A. Buell, B. Burkett, Y. Chen, Z. Chen, B. Chiaro, R. Collins, W. Courtney, A. Dunsworth, E. Farhi, B. Foxen, A. Fowler, C. Gidney, M. Giustina, R. Graff, K. Guerin, S. Habegger, M. P. Harrigan, M. J. Hartmann, A. Ho, M. Hoffmann, T. Huang, T. S. Humble, S. V. Isakov, E. Jeffrey, Z. Jiang, D. Kafri, K. Kechedzhi, J. Kelly, P. V. Klimov, S. Knysh, A. Korotkov, F. Kostritsa, D. Landhuis, M. Lindmark, E. Lucero, D. Lyakh, S. Mandrà, J. R. McClean, M. McEwen, A. Megrant, X. Mi, K. Michielsen, M. Mohseni, J. Mutus, O. Naaman, M. Neeley, C. Neill, M. Yuezhen Niu, E. Ostby, A. Petukhov, J. C. Platt, C. Quintana, E. G. Rieffel, P. Roushan, N. C. Rubin, D. Sank, K. J. Satzinger, V. Smelyanskiy, K. J. Sung, M. D. Trevithick, A. Vainsencher, B. Villalonga, T. White, Z. J. Yao, P. Yeh, A. Zalcman, H. Neven, and J. M. Martinis, "Quantum Supremacy Using a Programmable Superconducting Processor," Nature, Vol. 574, pp. 505–510, 2019.
[8] S. Endo, Z. Cai, S. C. Benjamin, and X. Yuan, "Hybrid Quantum-classical Algorithms and Quantum Error Mitigation," J. Phys. Soc. Jpn., Vol. 90, 032001, 2021.
[9] S. Endo, S. C. Benjamin, and Y. Li, "Practical Quantum Error Mitigation for Near-future Applications," Phys. Rev. X, Vol. 8, No. 3, 031027, 2018.
[10] Y. Suzuki, S. Endo, K. Fujii, and Y. Tokunaga, "Quantum Error Mitigation for Fault-tolerant Quantum Computing," arXiv:2010.03887, 2020.
[11] H. Goto, S. Mizukami, Y. Tokunaga, and T. Aoki, "Figure of Merit for Single-photon Generation Based on Cavity Quantum Electrodynamics," Phys. Rev. A, Vol. 99, No. 5, 053843, 2019.
[12] Y. Tokunaga, H. Goto, T. Utsugi, and T. Aoki, "Figure of Merit for the Efficiency of Single Photon Generation Using Cavity-QED Systems," 2nd International Symposium on Single Photon based Quantum Technologies, Berlin, Germany, May 2019.
[13] R. Asaoka, Y. Tokunaga, R. Kanamoto, H. Goto, K. Koshino, and T. Aoki, "Suitable Fault-tolerant Schemes for Cavity-QED-based Quantum Computation," 14th Pacific Rim Conference on Lasers and Electro-Optics (CLEO PR 2020), Aug. 2020.
[14] K. Koshino, K. Inomata, Z. R. Lin, Y. Tokunaga, T. Yamamoto, and Y. Nakamura, "Theory of Deterministic Entanglement Generation between Remote Superconducting Atoms," Phys. Rev. Applied., Vol. 7, No. 6, 064006, 2017.
[15] S. Masuda, S. Kono, K. Suzuki, Y. Tokunaga, Y. Nakamura, and K. Koshino, "Nonreciprocal Microwave Transmission Based on Gebhard-Ruckenstein Hopping," Phys. Rev. A, Vol. 99, No. 1, 013816, 2019.



Yuuki Tokunaga
Distinguished Researcher, NTT Secure Platform Laboratories.
He received a Ph.D. in science from Osaka University in 2007 and joined NTT in 2001, where he has been conducting research toward the realization of fault-tolerant universal quantum computing and long-distance secure quantum network.



Yasunari Suzuki
Researcher, NTT Secure Platform Laboratories.
He received a Ph.D. in engineering from the University of Tokyo in 2018. He joined NTT the same year, where he has been focusing on practical quantum error correction and fault-tolerant quantum computing for achieving reliable quantum systems. He is currently working on the development of fault-tolerant quantum computing at NTT Secure Platform Laboratories.

Suguru Endo
Researcher, NTT Secure Platform Laboratories.
He received a bachelor's and master's from Keio University, Kanagawa, in 2014 and 2016 and a Ph.D. from University of Oxford in 2019. He has been working as a researcher at NTT Secure Platform Laboratories since 2020. His research interests focus on hybrid quantum-classical algorithms, quantum error mitigation, and circuit QED.

Rui Asaoka
Posdoc, NTT Secure Platform Laboratories.
He received a Ph.D. in engineering from Tohoku University, Miyagi, in 2016. He joined NTT in 2020, where he has been engaged in research on quantum computing using light and atoms, especially focusing on its fault-tolerance.

Opening of the Art Exhibition "Digital × Hokusai [Middle Chapter]" for Providing New Ways to Experience Art during the Coronavirus Pandemic

Takehiro Suzuki

Abstract

NTT EAST has been holding an interactive art exhibition called "Digital × Hokusai [Middle Chapter]" since December 1, 2020 as a showcase for a "dispersed digital museum" that enables visitors to enjoy local culture and art beyond time and place, even during the coronavirus pandemic, by using the information and communication technology and assets owned by NTT EAST. The details of the company's initiatives concerning this exhibition are introduced in this article.

Keywords: local culture and art, during the coronavirus pandemic, regional revitalization

1. Introduction

From November 1, 2019 to February 28, 2020, NTT EAST held an interactive art exhibition called "Digital × Hokusai [Introduction]" [1] as an initiative to contribute to regional revitalization by sharing valuable local culture and art. The exhibition received positive feedback from many visitors about the advantage and potential of a digital museum that allows visitors to experience valuable local culture and art regardless of time or location.

Due to the spread of novel-coronavirus infections, museums and art galleries were required to shift their operations to adapt to new ways of enjoying culture and art. Under these circumstances, we planned to hold the next interactive art exhibition called "Digital × Hokusai [Middle Chapter]" in response to the many voices that wanted to share and enjoy the valuable culture and art of the local region and the growing need for a "new normal" concerning cultural appreciation through digital data.

2. Overview of "Digital × Hokusai [Middle Chapter]"

This art exhibition featuring works by Katsushika Hokusai and Utagawa Hiroshige, two of Japan's most famous *ukiyo-e* artists, is being held at the NTT Inter Communication Center (ICC) Gallery E in Nishi-Shinjuku, Tokyo. The exhibition is held in cooperation with Ars Techne Corporation, which has been officially certified by the Yamanashi Prefectural Museum of Art and the Musée d'Orsay in France as a replica master of paintings.

2.1 Exhibition of digital data and master replicas reproduced using high-definition technology and certified by the owners of artworks

The following paintings are being exhibited:

- Katsushika Hokusai, "Thirty-six Views of Mount Fuji" 47 works owned by Yamanashi Prefectural Museum of Art
- Utagawa Hiroshige, "Fifty-three Stations of the Tokaido" 55 works owned by Osaka Ukiyoe Museum
- Vincent van Gogh, "Starry Night on the Rhone"

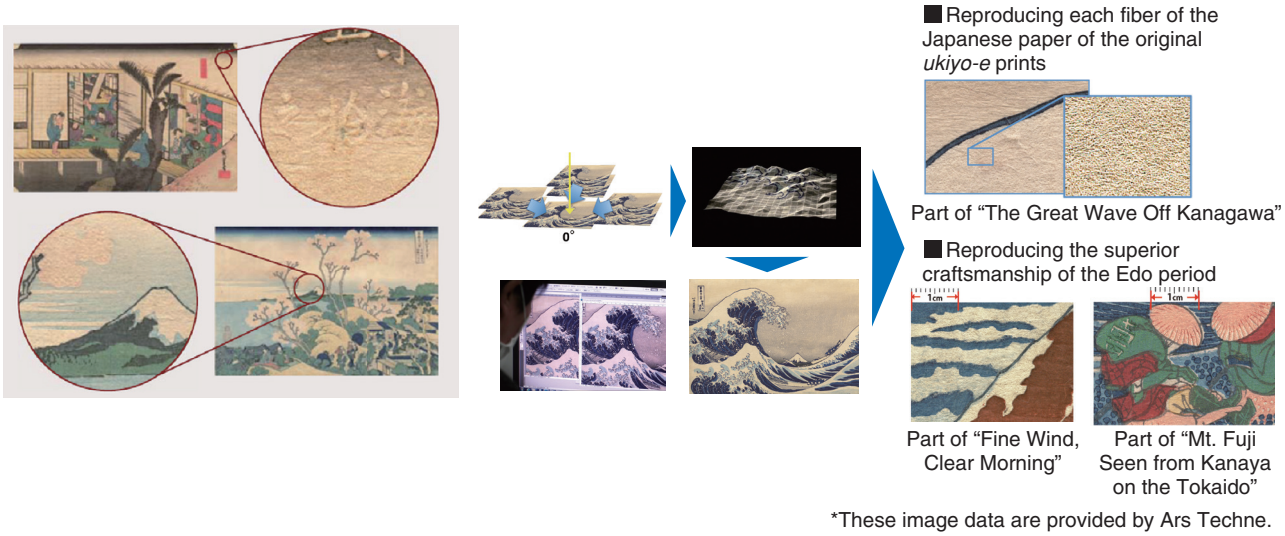


Fig. 1. DTIP: 3D-texture image-processing technology.

and 17 other works owned by Musée d'Orsay
- Gountei Sadahide, "Scene of Opening of a New Port in Yokohama, Kanagawa" owned by Kanagawa Prefectural Museum of Cultural History

Up until now, the exhibition of original artworks in museums has been restricted due to their susceptibility to damage, and it has not been possible to view them in proximity or under light. Even with high-resolution digitization of paintings, it has been difficult to reproduce the texture of materials and fine irregularities such as brush strokes.

Dynamic texture image processing (DTIP)[*1] is a three-dimensional (3D)-texture image-processing technology with features, such as high-definition multi-angle optical recording, developed by our collaboration partner, Ars Techne. We used DTIP for pseudo-stereoscopic image-composition conversion and 3D color-gamut separation and calibration with the original image, and were able to faithfully reproduce every fiber of the Japanese paper of the original *ukiyo-e* prints while conveying the artist's superior craftsmanship (**Fig. 1**).

2.2 Cooperation with local communities

Gallery E at the ICC, where the exhibition is being held, is linked to local facilities and hubs that have a connection to the works on display via a high-speed network so that as many people as possible can come into contact with Japanese culture and art. Therefore, the exhibition creates a world that can be enjoyed beyond time and place (**Fig. 2**).

2.3 Providing new experience using the latest digital applications

By fully using the latest digital applications, such as *naked-eye VR*, which allows viewers to experience virtual reality (VR) without goggles, a *3D dive theater*, in which viewers can experience the world of the artwork with their whole body as if being immersed in the painting, and *moving-art pictures*, which gives an illusion that paintings appear to be moving, we are providing new ways of appreciating cultural and artistic activities (**Fig. 3**).

2.4 High-quality distribution and secure accumulation of digital data

Taking advantage of the characteristics of NTT EAST telecommunications buildings and high-speed networks, namely, a closed network and secure environment, low latency, and disaster resistance, we will meet the needs required for the utilization of digital archives of cultural arts such as protection of cultural property rights, smooth content distribution, and disaster recovery[*2].

*1 DTIP: Original patented technology of Ars Techne Corporation.
*2 Disaster recovery: Measures to recover from damage caused by disasters or preventive measures to minimize damage.

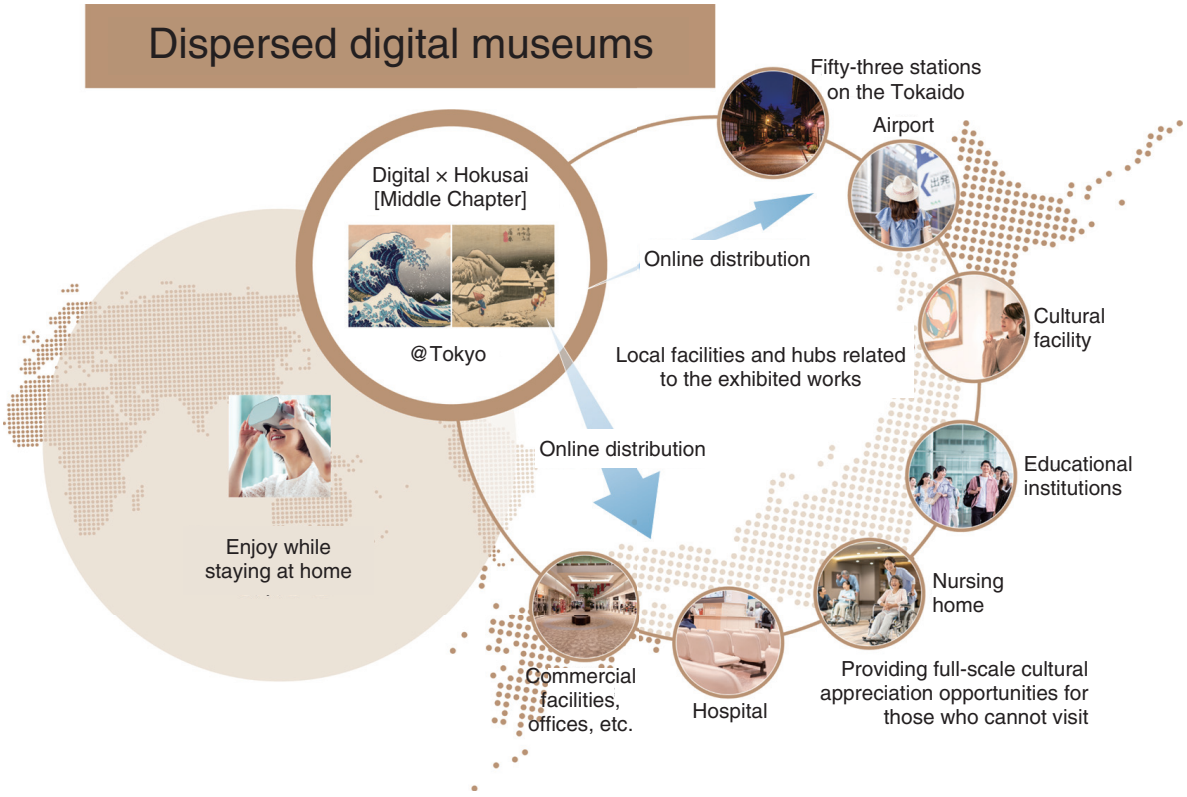


Fig. 2. Cooperation with local communities.

(a) Naked-eye VR

(b) 3D dive theater

Fig. 3. Providing new experience by using the latest digital applications.

2.5 Providing safe and secure non-contact museum solutions

Through mechanisms that allow visitors to enjoy all content without contact (such as naked-eye VR, 3D dive theater, and moving-art pictures), congestion detection with artificial-intelligence (AI) cameras, and guidance from AI robots, we will continue to provide museum solutions that allow visitors to appreciate exhibits with peace of mind during the coronavirus pandemic. On the homepage of this art exhibition [2], virtual viewing via a 360-degree viewer is provided to allow visitors to appreciate part

of the exhibition without having to physically be at the venue.

We will also plan to link this art exhibition with NTT's 3D spatial owned media called DOOR that embodies the *remote world* provided by NTT.

References

[1] NTT EAST, "'Digital × Hokusai [Introduction]' Exhibition Contents," https://www.ntt-east.co.jp/pr/hokusai_profile_en.html
[2] "Digital × Hokusai [Middle Chapter]" (in Japanese), https://www.ntt-east.co.jp/pr/hokusai-hasyo/
[3] DOOR (in Japanese), https://door.ntt/

Author: Takehiro Suzuki, Associate Manager, Corporate Strategy Planning Department, NTT EAST Corporation.

REALIVE360: Multi-angle Virtual-reality Video-streaming Service that Gives the Viewer a Realistic Feeling of Being in a Theater

Takahiko Sasahara, Masanori Emura, and Takafumi Fukatani

Abstract

While various events have been canceled or postponed due to the novel-coronavirus (COVID-19) pandemic, a new style of event called "online streaming of spectatorless events" is emerging. The NTT WEST Group has developed a multi-angle virtual-reality video-streaming service called REALIVE360 and began providing it as a service to enable people in remote rural areas to experience concerts and other events held in urban areas in a highly realistic manner. How to use REALIVE360, its main features, and its future development are introduced in this article.

Keywords: highly realistic, multi-angle viewing, 3D sound

1. Current status of live music and other events during the COVID-19 pandemic

The novel-coronavirus (COVID-19) pandemic has dramatically changed the event industry. Under the state of emergency imposed by the Japanese government to stop the spread of the virus in early 2020, almost all events and performances were canceled, postponed, or reduced in scale. Tokyo Olympic and Paralympic Games were postponed, and professional baseball and football leagues have been forced to postpone the opening of their seasons. Even after the state of emergency was lifted, music festivals, local festivals, and other events that were scheduled to be held in 2020 had been cancelled, postponed, or scaled back considerably. Clusters of COVID-19 infections linked to music-related events have been reported, so people's self-restraint generated under the state of emergency is continuing; thus, live-music venues across Japan have been closing one after another. The results of a computer simulation showed that if physical distance is maintained in a theater to prevent the spread of COVID-19, the room will be "full" at 15% capacity. These results have attracted significant attention [1]. Under these circumstances, the holding of events offline is not expected to return to "normal" anytime soon.

2. Online streaming of spectatorless events

Although it is difficult to hold events offline, a new style of event called "online streaming of spectatorless events" is emerging. The band Southern All Stars performed a live broadcast without spectators, and the number of people who purchased tickets reached about 180,000, which generated sales of an estimated 650 million yen (equivalent to about 6 million US dollars) [2]. That number of people was more than ten times the capacity of the venue being live streamed (17,000 people), and it has made the potential of online streaming known to a wider audience.

The advantage of online streaming is that the unit price of tickets is set at 2000 to 3000 yen, which is significantly lower than regular music concerts, so

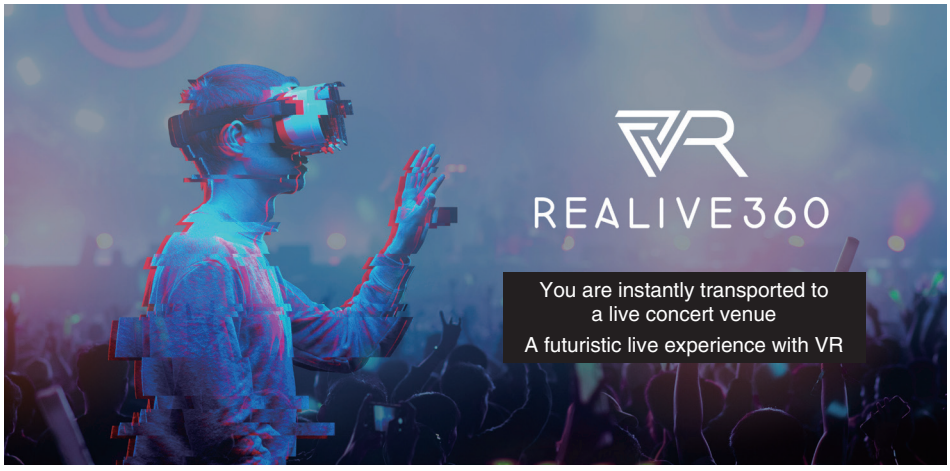


Fig. 1. Poster of REALIVE360.

people who are willing to pay that amount will purchase tickets. What's more, if the event is held on a weekday evening, fans from afar who are unable to get to the event for various reasons can simply decide to watch it online. Reducing the price of live-streaming tickets would be thought to lead to a decrease in sales; however, it made it possible to sell tickets to people who were reluctant to buy tickets and fans who had given up watching live shows.

3. REALIVE360

NTT WEST Group's Smart Life team began developing the multi-angle virtual-reality (VR) video-streaming service called REALIVE360 in 2018 (**Fig. 1**). Initially, our slogan for developing this service was to "Eliminate the disparity in experiences and opportunities between urban and rural areas through the power of ICT (information and communication technology)." The majority of entertainment-related events and people mobilized for such events, including live music, is concentrated in urban areas, and it is a hurdle for music fans living in rural areas to spend time and money on transportation to attend a concert in urban areas. Moreover, many people, such as the disabled, seniors, and families with small children, who love music, theater, and sports have difficulty getting to venues. We developed REALIVE360 as a service that enables such people to experience—with a high level of realism—concerts and other events held in urban areas, even if they live in remote rural areas (**Fig. 2**). In addition to the streaming technology of NTT WEST Group, we have established a collaborative system in partnership with companies such as Alpha Code Inc. that handles VR filming and editing. At the end of 2019, we streamed a live performance of Momoiro Clover Z, a Japanese idol group, which attracted a lot of attention as the first artist to be streamed. After that, as mentioned in the introduction, online streaming of events became the mainstream due to the COVID-19 pandemic. We collaborated with the radio station FM802 to implement REALIVE360 VR ZONE, a project to support the music industry, which is suffering from a decline in audience numbers and cancellation of events due to the pandemic (**Fig. 3**). We have accumulated achievements such as the streaming of the online event J-WAVE Innovation World Festa and the stage version of the animated movie "For Whom the Alchemist Exists."

3.1 Streaming 4K/8K high-quality 360° VR videos

The first feature of REALIVE360 is unprecedented streaming of 4K/8K high-quality 360° VR videos. Our unique technology eliminates the coarseness of image quality that was previously associated with 360° VR streaming. REALIVE360 uses the Panorama Super Engine, which was developed by NTT laboratories and commercialized by NTT TechnoCross, to stream 4K/8K high-quality VR video. 360° VR video requires 360° video data, including parts that are not visible at certain moments, so it requires a different means of thinking than what we apply to 16:9 video displayed on an ordinary flat-screen display. The 4K video data displayed on a flat screen are

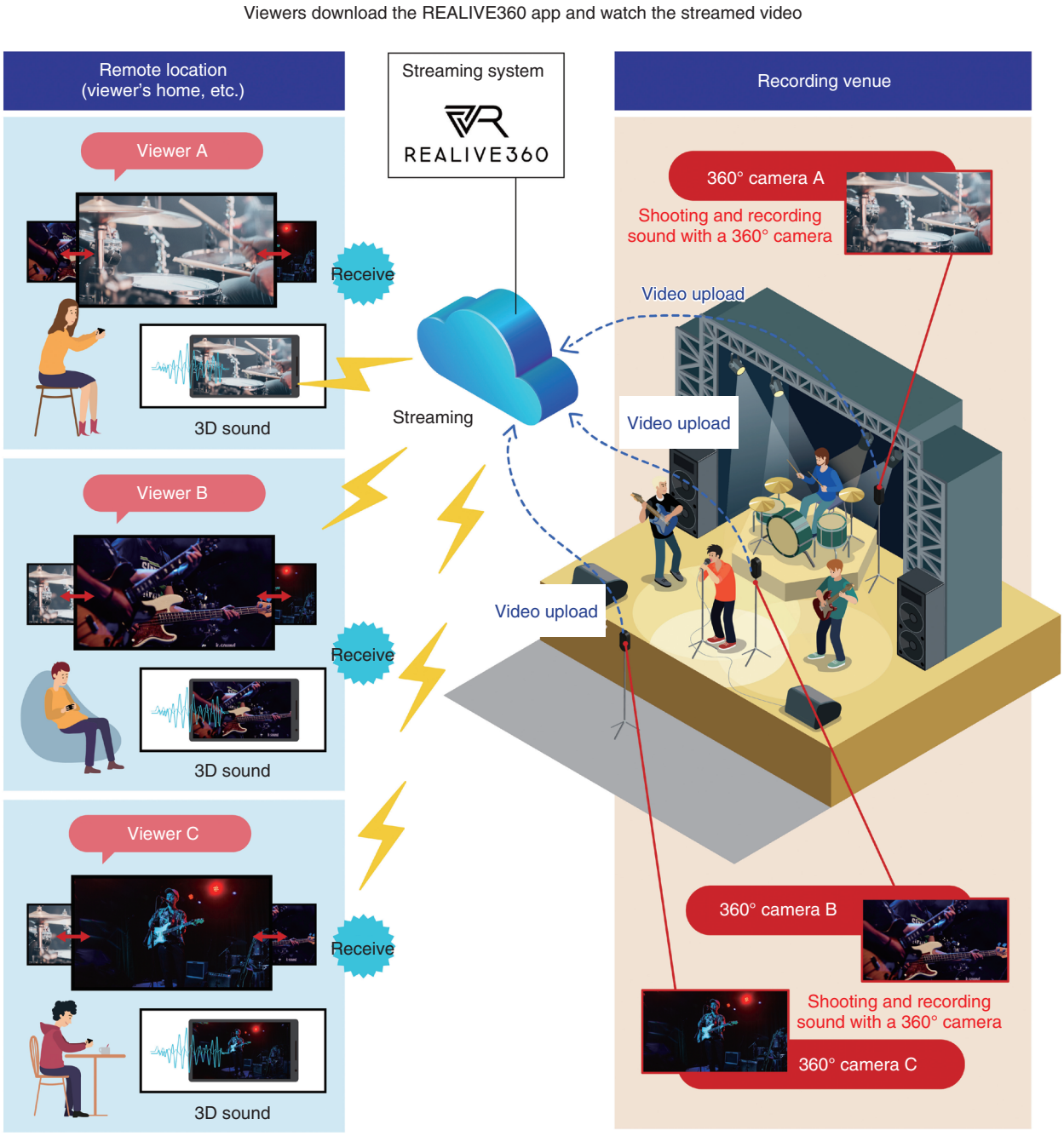


Fig. 2. Illustration of REALIVE360 service.

expressed in 4K resolution on that screen. For 4K 360° VR video, however, the video data (including the parts that are not visible at certain moments) are set to be in 4K resolution; therefore, the actual viewing area is expressed at a resolution less than 4K. As a result, when a 360° VR video is viewed, the screen will be blurrier than expected, despite the 4K resolution. If the resolution is increased to 8K to increase the image quality in the viewing range, the bandwidth required for the streaming network will increase, and the requirements for the viewing environment will be more difficult to meet. In other words, the viewer's network environment must be sufficiently refined, and that requirement might significantly affect the

Fig. 3. Performance by Kyuso Nekokami at REALIVE360 VR ZONE.

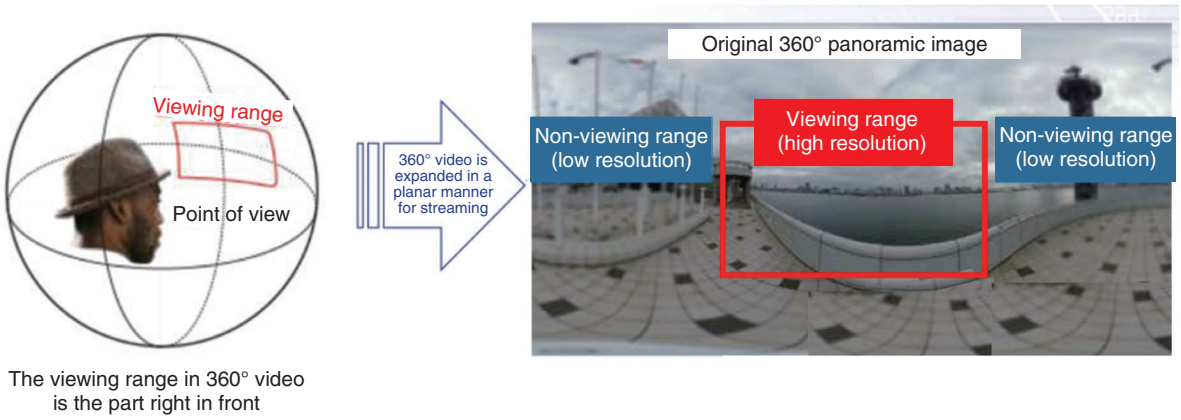


Fig. 4. Panorama Super Engine.

expansion of the number of viewers. Given these conditions, we decided to adopt the Panorama Super Engine, which can stream high-resolution video in the viewing area only and low-resolution video in the other unseen areas. When the Panorama Super Engine is used, the required network bandwidth can be reduced by about 75% for 4K resolution and 85% for 8K resolution, allowing viewers to enjoy 360° VR videos with 8K resolution even over a 4G (fourth-generation) mobile connection (**Fig. 4**).

3.2 Multi-angle viewing

The second feature of REALIVE360 is multi-angle viewing. Since viewers can freely select the angle from which they want to watch from multiple angles, they can proactively and subjectively enjoy the live performance from the angle they prefer. The viewer can also select angles that would not normally be possible, such as following favorite band members, looking across the audience from the stage as the performers would do, or looking down from the ceiling. Because this function is multi-angle, the viewer can

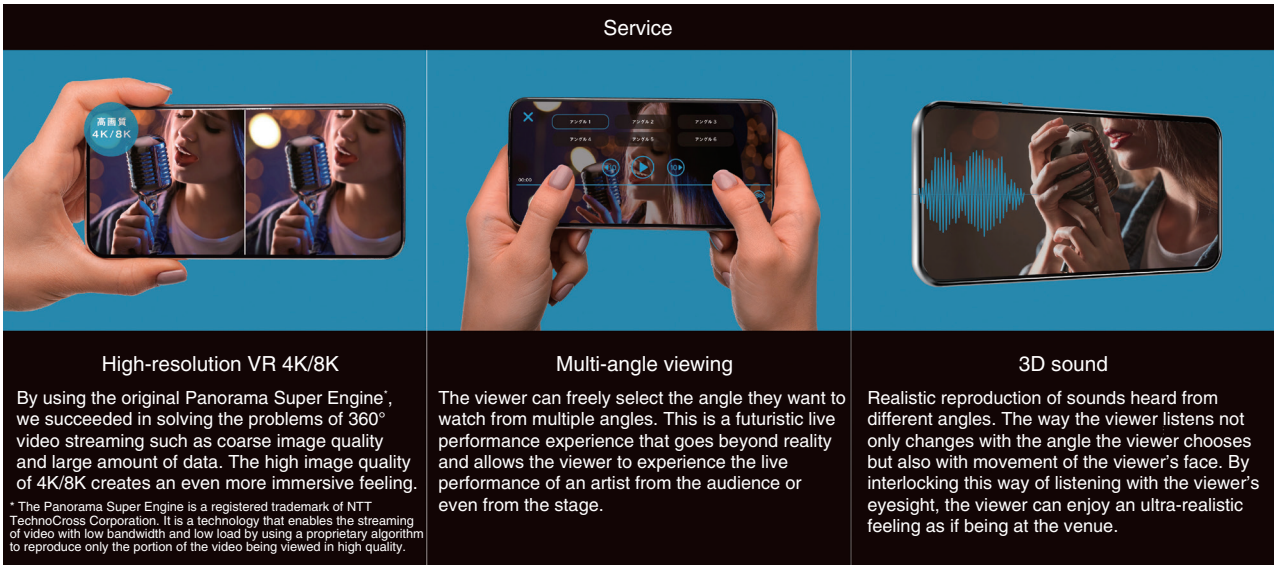


Fig. 5. Features of REALIVE360.

set an extraordinary sub-angle while setting the main angle that firmly holds the performer in a bust shot.

3.3 3D sound

The third feature of REALIVE360 is three-dimensional (3D) sound. For example, the singer's voice can be heard from the right side of the image, while the drums can be heard from the back of the image. The sound heard from each actual angle can be realistically reproduced in the VR video. The way the sound is heard changes not only when the viewing angle changes but also when the point being viewed in the VR video changes. Therefore, we are trying to increase the sense of reality not only in the image but also in the audio (**Fig. 5**).

REALIVE360 service is provided via an application, so it is not necessary to wear special goggles or headsets; in other words, it allows the viewer to enjoy the world of VR by simply using a smartphone or tablet. What's more, since REALIVE360 is linked with various ticket agencies, viewers can purchase tickets from one of those agencies, authenticate it with the issued serial code, and view the event via the REALIVE360 app.

4. Future developments

As explained above, REALIVE360 is currently mainly used for entertainment. We have received many requests for additional functions from event organizers who are service users, and we are currently developing the following functions; (i) those that enable the user to do online what used to be done when running events offline, such as a function that allows performers and viewers to communicate interactively and a function that displays advertisements such as sponsors; (ii) those that can be implemented only by online streaming, such as a money-transfer function that allows viewers to send money with a feeling of gratitude and support to the provider and a function for analyzing viewing data; and (iii) those that improve viewing extensibility, namely, a cast function allowing the VR videos to be viewed on a television or a web browser in addition to on the current app.

We plan to further expand REALIVE360 service into other fields such as education and tourism to cultivate the business-to-business market. We are also considering partnerships that transcend industry boundaries and hope to hear a wide range of opinions to help us solve problems and create new value in a variety of business situations.

References

[1] Okinawa Times, "It's almost empty, but with this it's full—A 'question' raised by a theater draws attention," May 2020 (in Japanese).
[2] Net IB News, "Southern All Stars' spectatorless live concert earned ticket sales of 650 million yen. Online streaming is a glimmer of hope," July 2020 (in Japanese), https://www.data-max.co.jp/article/36502

Authors (from left): Masanori Emura, Manager, Business Design Department; Takahiko Sasahara, Senior Manager, Business Design Department; and Takafumi Fukatani, Chief, Business Design Department; NTT WEST Corporation

Private 5G: A Key Solution for Driving Digital Transformation and Creating a Smart World

Ryo Maeda, Hiroaki Kakimoto, Yoshiaki Takeda, Yukinaka Matsuyama, Daisuke Nakamura, and Fukumasa Morifuji

Abstract

The wireless technology called private fifth-generation mobile communication system (5G) is currently attracting attention from various customers, including the manufacturing industry. Unlike Wi-Fi, which has been widely used as a wireless-communication technology to build private networks, private 5G (called "local 5G" in Japan), which uses licensed bands, requires special expertise for design, construction, and operation. NTT Communications has accumulated knowledge of private 5G through conducting demonstration experiments even before private 5G was institutionalized in Japan. This article introduces examples of joint demonstration experiments with customers using private 5G networks as well as technologies under development such as multi-access edge computing and network slicing.

Keywords: private 5G, digital transformation, network slicing

1. Data collection with private 5G, which supports digital transformation of customer business

As a DX Enabler™, NTT Communications (NTT Com) aims to join with customers in implementing digital transformation (DX), which will enable customers to create new businesses and strengthen their competitiveness. Data utilization is the key to implementing DX, and in 2019, NTT Com announced the Smart Data Platform for supporting it.

The Smart Data Platform has all the functions needed to collect, store, manage, and analyze data. Private fifth-generation mobile communication system (5G)—a wireless communication technology to build private networks using 5G enabling advanced data collection—is also included on the Smart Data Platform as an essential technology for implementing DX. By using technologies, such as private 5G, to create an environment for secure data collection and storage and provide a one-stop solution for data integration, analytics, and utilization, NTT Com is aiming to help its customers create business value.

2. Achieve communication with mobile devices and flexible production systems in a secure and high-quality wireless environment

Some of the features of private 5G are high-speed/high-capacity, low-latency communication, and simultaneous connection with many terminals. In particular, the ability to communicate at high speeds with moving objects is eagerly awaited by customers. Collecting large amounts of data from moving objects, such as trains, cars, drones, and robots, in a stable and real-time manner has been difficult with conventional wireless communication technologies such as Wi-Fi. Private 5G will make it possible to quickly collect large amounts of data, such as video, thus expand the scope of data utilization.

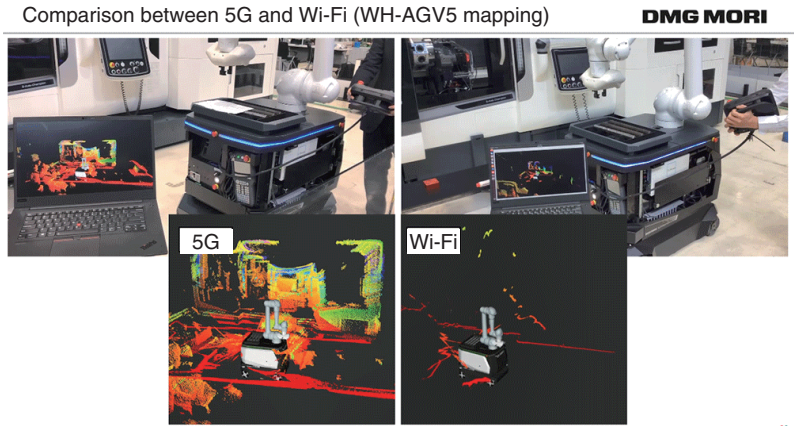


Fig. 1. Comparison between three-dimensional mapping speed of Wi-Fi and private 5G.

Security is another area in which private 5G is generating expectations. Similar to public 5G provided by mobile carriers, private 5G will use subscriber identity module (SIM)-based authentication and encryption to create a secure wireless communication environment.

For example, the input and output data of various types of manufacturing equipment are highly confidential, and security must be ensured. By using SIM-based authentication and encryption in a private 5G environment, risks such as spoofing and eavesdropping can be significantly reduced, and data that are difficult to exchange using conventional wireless communication technology can be sent and received safely.

The use of private 5G in situations in which robots work in collaboration with people is being considered. Applications of private 5G include collecting operational data of a robot in real time or assisting robot control from a remote location. NTT Com is conducting a demonstration experiment with DMG MORI CO., LTD. involving connection with robots via private 5G.

Private 5G also has the potential to expand its applications to fields other than manufacturing industries. For example, it is expected to be applied to facility security using robots or to equipment inspection using drones. NTT Com formed a consortium with SOHGO SECURITY SERVICES CO., LTD. (ALSOK) and Keikyu Corporation to conduct a field demonstration under Japan's Ministry of Internal Affairs and Communications' FY2020 project "Development Demonstrations for Implementing Private 5G to Solve Local Issues" [1].

2.1 Case study: Joint experiment with DMG MORI

In a joint experiment with DMG MORI, NTT Com is verifying the feasibility of using private 5G to remotely control an automatic guided vehicle (AGV) fitted with a collaborative robot. For example, by comparing the speed of collecting data about the state of the surrounding environment from the sensors installed on the AGV when using Wi-Fi or private 5G, we confirmed that private 5G can collect data faster and more stably than Wi-Fi (**Fig. 1**). By repeating such verifications, we are accumulating knowledge and expertise regarding the application of private 5G to machine tools.

Through these demonstration experiments, we expect to further improve the performance of AGVs such as improving the accuracy and safety of transport and reducing the weight of AGVs by using edge computing.

3. Provide *total coordination* not only with private 5G but also with network-construction technology cultivated over many years

NTT Com's strength lies in its ability to not only build a private 5G environment but also provide optimal communication quality and solutions through *total coordination* of the entire system, consisting of from networks connecting to the edge and cloud to applications. NTT Com's Smart Data Platform provides all the necessary functions for data utilization. Customers can collect and use data via private 5G by selecting and combining required functions such as network services (e.g., communication line, edge

Fig. 2. Demonstration experiment of private 5G at Arcs Urayasu Park.

Fig. 3. Verification of radio-wave propagation at Arcs Urayasu Park.

computing, and cloud computing) and various applications linked with private 5G.

Another strength of NTT Com is that it can flexibly propose solutions. Regarding wireless communications, NTT Com is not only focused on private 5G; it may also recommend private LTE (Long Term Evolution), Wi-Fi, or LPWA (low power wide area) networks depending on the customer's situation. NTT Com selects the most suitable equipment, services, and technologies for addressing customers' issues and proposes them flexibly in a well-coordinated manner.

4. Demonstration experiment of private 5G at Arcs Urayasu Park

NTT Com's efforts to develop private 5G technology is explained below. NTT Com obtained a license for private 5G experimental test using not only the 28-GHz band but also the 4.7-GHz band (which was institutionalized at the end of 2020) in June 2020 and started verifying radio wave-propagation characteristics at Arcs Urayasu Park, which is the training ground of NTT Com's rugby team (**Figs. 2** and **3**). The radio-wave characteristics in the 4.7- and 28-GHz bands differ significantly, although both bands are used for private 5G. The 28-GHz band can be fit when communication speed is crucial; however, transmission or diffraction in relation to obstacles cannot be expected, and its reach is short. It will thus be more advantageous in situations in which 5G communications are used within the line of sight, which is close to the ideal wireless environment (namely, no obstructions exist between the base-station antenna and receiving terminal).

The 4.7-GHz band, however, can send radio waves

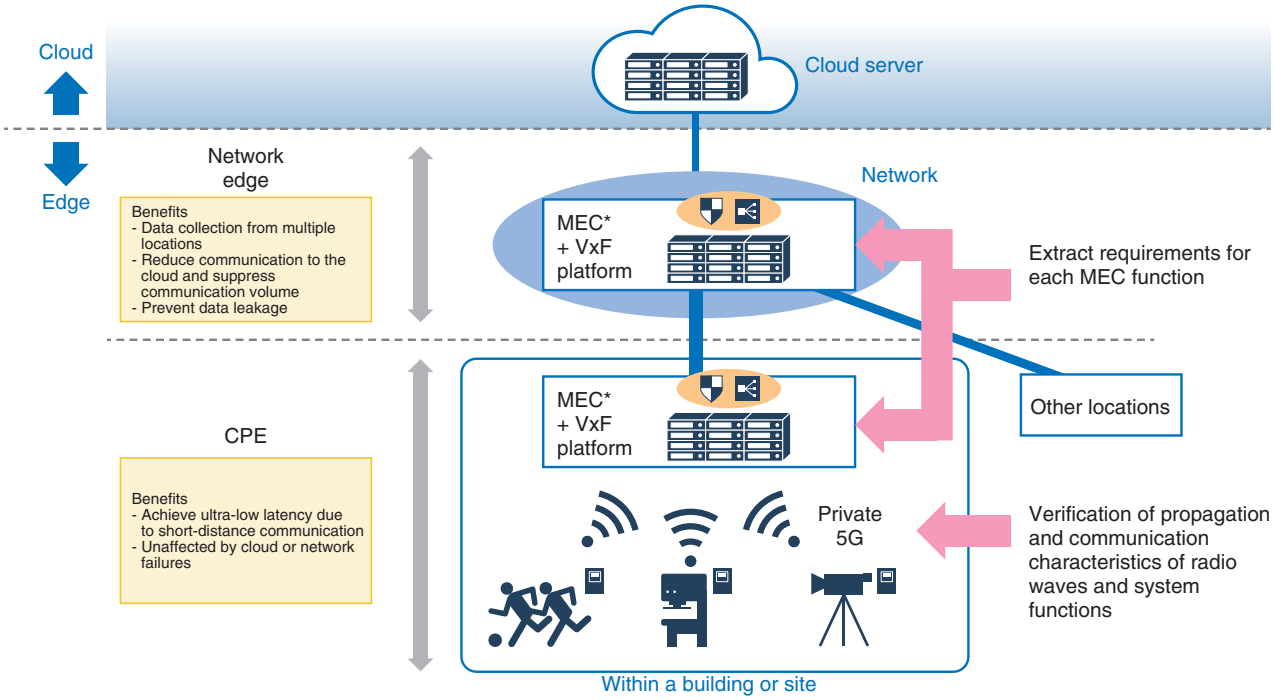


Fig. 4. Verification of private 5G combined with multi-stage MEC.

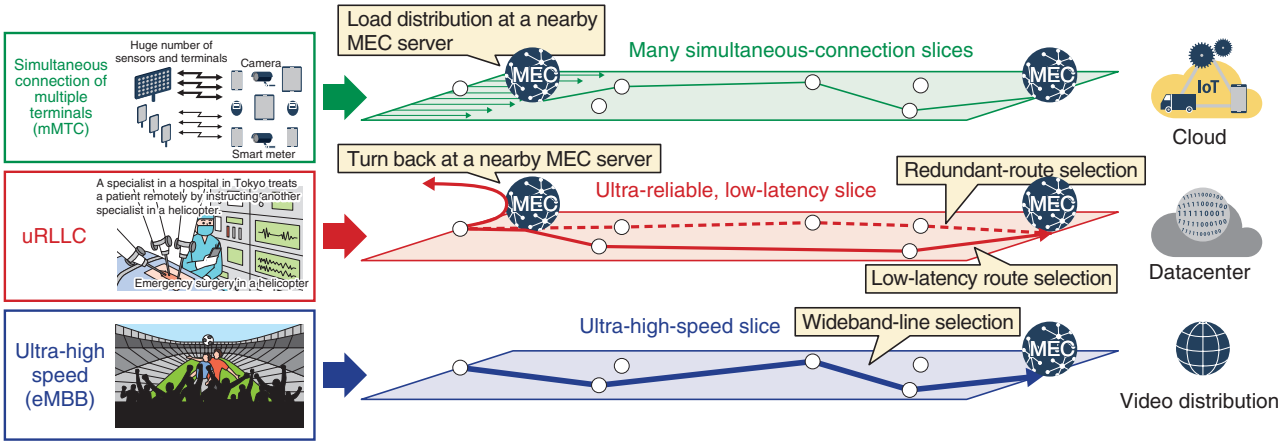
farther than possible with the 28-GHz band and allows radio waves to be transmitted and diffracted (to some extent) even if obstacles are present. It is therefore possible to cover a larger area than that with the 28-GHz band, such as a factory site, with as few base stations as possible. In consideration of the characteristics of these two frequency bands, it is important to select the appropriate frequency band in accordance with the customer's usage and communication needs.

NTT Com is also investigating various aspects of the linkage between private 5G and other network services. In particular, it is focusing on a technology called multi-access edge computing (MEC), which processes data collected by private 5G communication on a network closer to the user side instead of on the cloud. For customers who require low latency, MEC can be deployed on-premises (i.e., for in-house information systems) to enable real-time data collection and utilization. Moreover, when NTT Com's network services are used, it is possible to configure MEC at the network edge and use multiple levels of MEC depending on the application (**Fig. 4**). Furthermore, using NTT Com's next-generation interconnect service Flexible InterConnect makes it possible to securely connect not only to the edge but also to various cloud services.

By using MEC in this manner, it is possible to (i) reduce the amount of data sent to the cloud while achieving low latency and (ii) increase the possibility of continuing processing even if there is a problem with the cloud or Internet due to a disaster or other event. This type of support is one of the strengths of NTT Com, enabling it to offer a wide range of services.

5. Verification of end-to-end network slicing with a stand-alone design method

A feature of 5G is network slicing, which is used to build a virtual network using software. The 5G specifications developed by 3GPP (the Third Generation Partnership Project) specify three types of slices: high-speed, high-capacity communication, low latency, and simultaneous connection of multiple terminals. Regarding private 5G, slices can be created in



eMBB: enhanced mobile broadband
IoT: Internet of Things
mMTC: massive machine-type communication
uRLLC: ultra-reliable and low-latency communication

Fig. 5. Verification of end-to-end network slice.

accordance with the requirements of each application or terminal (**Fig. 5**).

NTT Com aims to create a variety of detailed network slices—not only for private 5G communication but also for the networks behind it—and control these slices end-to-end. This makes it possible to configure networks that are tailored to each customer's needs. For example, in highly confidential cases such as factory-operation data, the data can be placed on a slice on a closed network or on an on-premises network separated from slices connecting to information networks such as the Internet. In such a case, using Ericsson's private 5G solution Edge Gateway* and each vendor's products and technologies, NTT Com is conducting verification of sub-6 GHz bands and stand-alone design methods, end-to-end latency measurement and verification of slicing functions by connecting to backyard networks, and is expanding technological development and verification so that networks can be constructed flexibly. NTT Com will expand demonstrations by deploying private 5G in the actual field, such as installing various edge computing functions and linking with Smart Data Platform.

NTT Com successively demonstrated real-time video analysis with private 5G in a verification environment set up at Arcs Urayasu Park. As shown in the photo in **Fig. 6**, a video of a player kicking a rugby ball is transmitted via a low-latency slice, and the angle and initial velocity of the ball are immediately calculated by an application on the MEC server. From the calculation results, artificial intelligence (AI) predicts the highest point and distance of the ball, then the trajectory of the ball is plotted on the original video and output.

Therefore, NTT Com will propose how to use private 5G as a total network solution in consideration of the envisioned use cases.

6. Concluding remarks

Although private 5G is a new wireless communication technology, its foundation is none other than the network technologies that NTT Com have cultivated over the years. While using such technologies as well as the Smart City Lab (a co-creation environment) and considering services that can use private 5G at reasonable costs, it will accelerate co-creation with customers by creating private 5G services and solutions that promote DX for customers to create businesses and strengthen their competitiveness.

* Edge Gateway: Introducing Edge Gateway is expected to start with small-scale operations, and NTT Com has signed a contract with Ericsson as an early adopter—one of two companies in the world. And NTT Com has tuned it for use with its private 5G solutions and proceeding with demonstration tests.

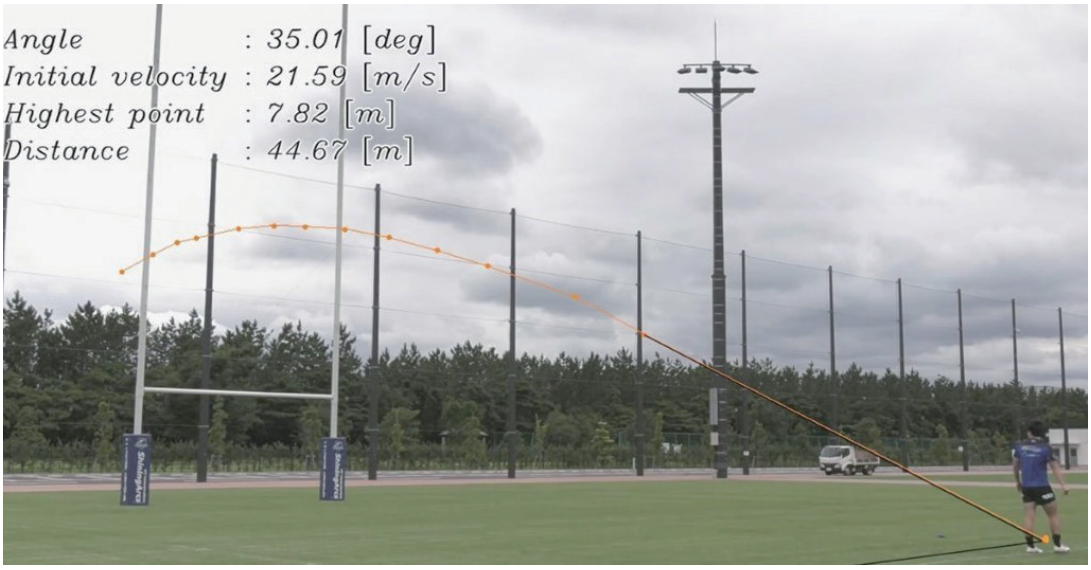


Fig. 6. Example of verifying network linkage: trajectory plotting and video analysis of a rugby ball.

Reference

[1] Press release issued by NTT Com on Nov. 20, 2020 (in Japanese). https://www.ntt.com/about-us/press-releases/news/article/2020/1120.html

Trademark notes

Authors (clockwise from top left): Yoshiaki Takeda, Manager, Solution Services; Ryo Maeda, Senior Manager, Solution Services; Hiroaki Kakimoto, Manager, Solution Services; Fukumasa Morifuji, Member, Innovation Center; Daisuke Nakamura, Manager, Innovation Center; and Yukinaka Matsuyama, Member, Innovation Center; NTT Communications Corporation

Speeding Up the Machine-learning Process with MLOps and Creating a Mechanism to Continuously Provide Service Value

Ei Yamaguchi

Abstract

Machine-learning operations (MLOps) is the machine-learning version of DevOps (development and operations) and represents the concept of how the people in charge of developing machine learning for a system and the people in charge of operating the system can collaborate to ensure smooth progress from implementation to operation of a commercial system. Recently, MLOps has been gaining in popularity; however, each vendor has a different definition of MLOps and there is no unified view. With that issue in mind, this article explains the background and basic concepts of MLOps as well as latest investigations and concrete means of implementing MLOps.

Keywords: machine learning, MLOps, AI

1. Background

The number of projects for launching new services or improving existing operations by using artificial intelligence (AI) is increasing yearly; however, there are many cases in which AI is not fully implemented in actual business.

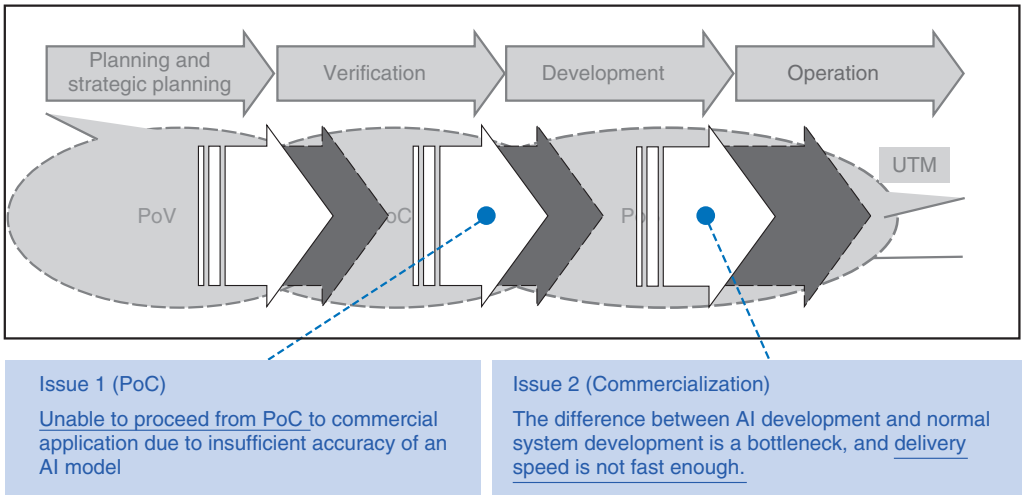
There are two main reasons for this state of affairs: (i) the accuracy of AI cannot be improved to a level that is sufficient for actual operations within a limited proof of concept (PoC) period and (ii) when the system developer takes over an AI model created by a machine-learning engineer and deploys it in a commercial system, it takes time to do it owing to the lack of communication and job splitting between them (**Fig. 1**).

Issues after service deployment include dealing with the phenomena of concept drift and data drift. For example, due to the spread of novel coronavirus (COVID-19) infections over the last year or so, people's behavior has been changing on a weekly, or even daily, basis. Consequently, the accuracy of AI models fine-tuned by data scientists before the change has deteriorated over time, making the models useless. Such a phenomenon is referred to as concept (or data) drift.

As a means of addressing the two issues shown in Fig. 1, a set of practices called machine-learning operations (MLOps) has recently been attracting attention. In this article, the approaches to use MLOps to address these issues are explained.

1.1 Issue 1: Unable to improve accuracy of an AI model to a level that can handle actual operations within the PoC period

To address the first issue, it is considered effective to increase the number of tunings of an AI model within a limited PoC period by streamlining the tuning process. It seems intuitively correct that increasing the number of tunings will improve the accuracy of the model because doing so will incorporate many measures that may contribute to improved accuracy. Two steps are considered effective to make the tuning process, which is a mundane task, more efficient:



PoV: proof of value
UTM: Unified Threat Management

Fig. 1. Issues concerning the machine-learning development process.

first, standardize the process and give it a common language; second, introduce tools that can streamline each step of the process.

For the first step, the analytical framework used in the data-analytics field is considered to be effective. For example, an analytical framework called CRISP-DM (cross-industry standard process for data mining) is available. This framework's process is different from the waterfall development process in that it permits movement between processes in an agile manner (**Fig. 2**). CRISP-DM is compatible with the machine-learning development process in which it is common to improve accuracy of an AI model through trial and error by examining questions such as "What happens if we use such feature values?" and "What happens if we change the algorithm?"

For the second step, open source software (OSS), cloud-service providers, and third-party vendors have released tools to improve the efficiency of machine-learning development. NTT DATA is providing the following services for introducing MLOps to help customers having problems with efficiency of developing machine-learning models introduce tools that can improve the efficiency of each process (**Fig. 3**).

A tool called AutoML is a typical example of a very effective tool that can directly contribute to improving accuracy of machine learning. AutoML selects the most-accurate machine-learning model by simultaneously running multiple algorithms necessary in the machine-learning development process, namely, feature design, model design, and model tuning.

However, even if these tools are used to improve the accuracy of AI, doing so will be meaningless unless the benefit to a customer's actual business can be clearly shown. For that purpose, it is necessary to translate indicators of AI-model accuracy, namely, accuracy and recall rate, into the words used in the customer's actual business so that the impact on their business is demonstrated.

1.2 Issue 2: Development of machine learning for commercial systems requires the cooperation of experts in various roles

An overview of the machine-learning development process and roles of the participating experts is given in **Fig. 4**. Even at the overview level, the participation of a variety of experts is necessary. Many projects in which AI engineers and other experts participate face the following issues:

- AI engineers tend to focus on improving model accuracy; however, the business and data-engineering side want AI engineers to make proposals concerning data generation and business processes required for reporting data quality and improving the data quality itself.
- AI engineers are not necessarily outstanding software developers, so the quality of the code they create may be poor, and the cost of rewriting that code by delivery-side developers to improve it to the high quality required for commercial use

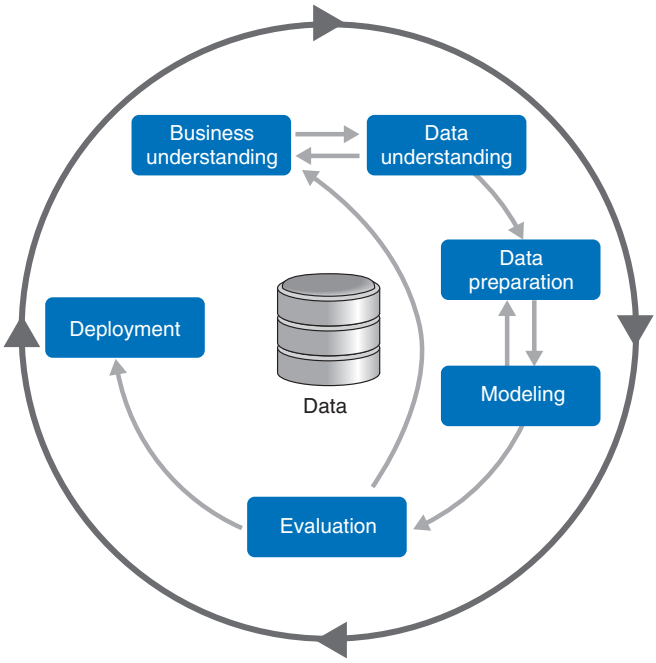


Fig. 2. CRISP-DM.

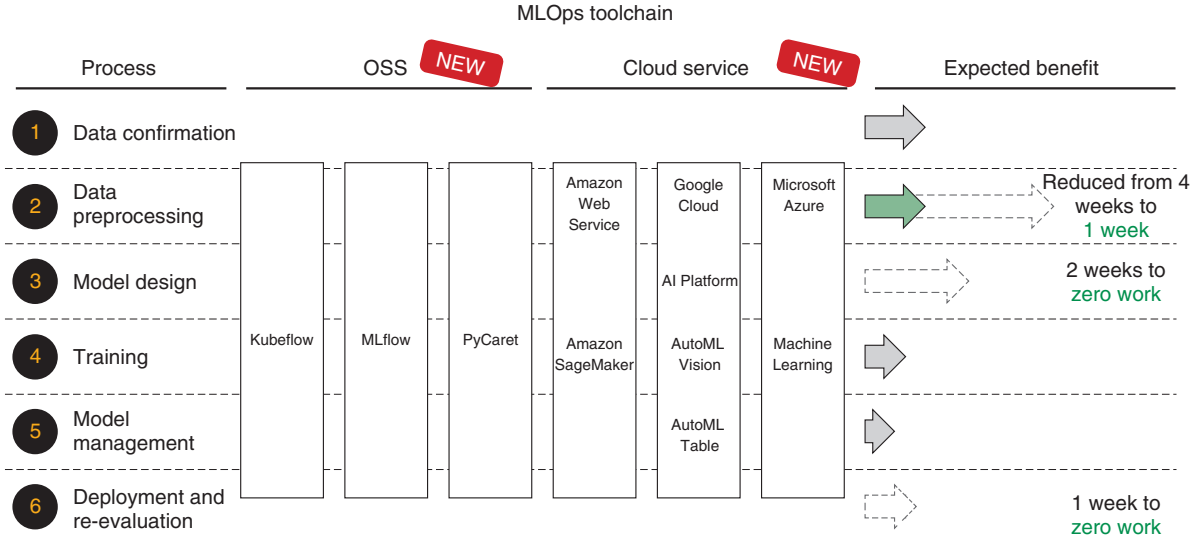


Fig. 3. MLOps toolchain of NTT DATA.

is very high.

In light of these issues, it is clear that the definition of the role of AI engineers needs to be broadened in regard to the commercial-development phase. In addition to technical knowledge to improve model accuracy, which is essential in the PoC phase, there is a need for AI engineers who can create high-quality code for implementing commercial systems, understand data and customer operations, and make proposals for improving data quality.

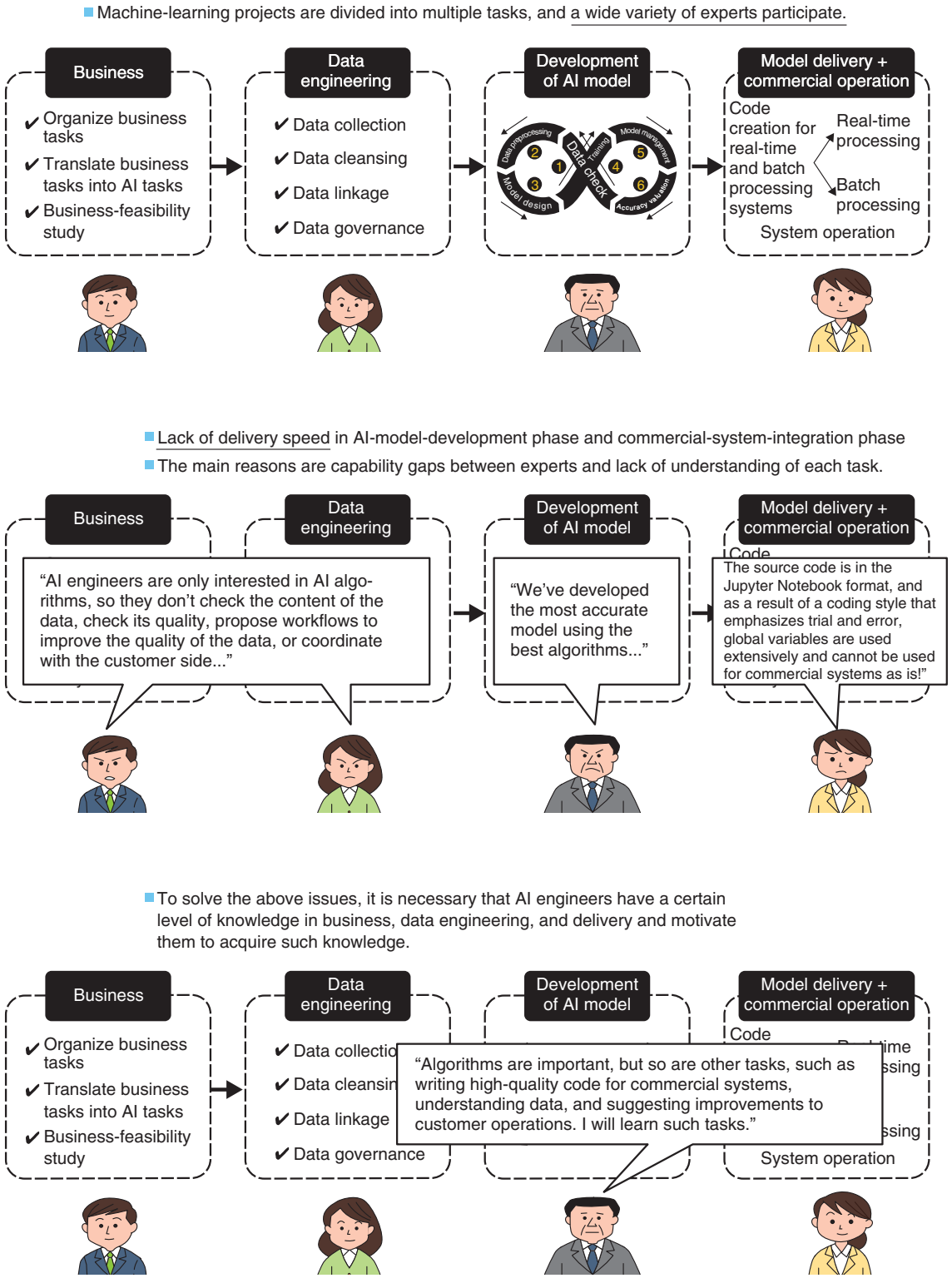


Fig. 4. Machine-learning development process and roles of experts.

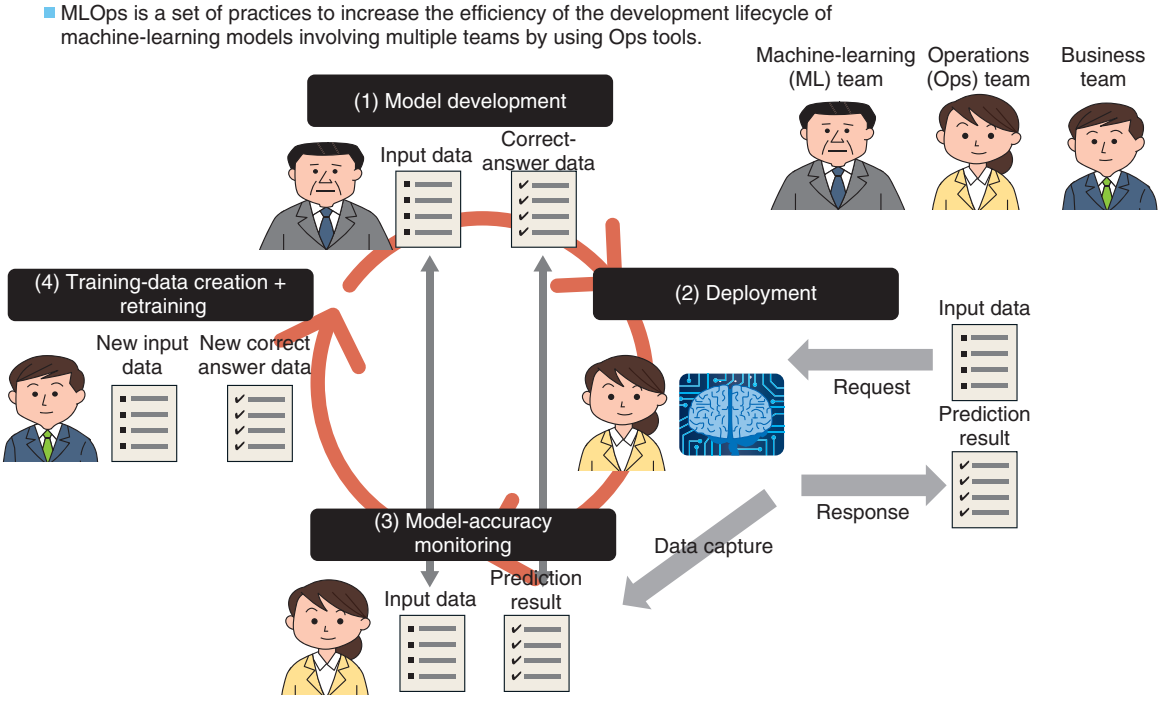


Fig. 5. Process flow of MLOps.

1.3 Issue 3: When an event that changes people's behavior occurs, such as a pandemic, machine-learning models that were tuned before the change become useless

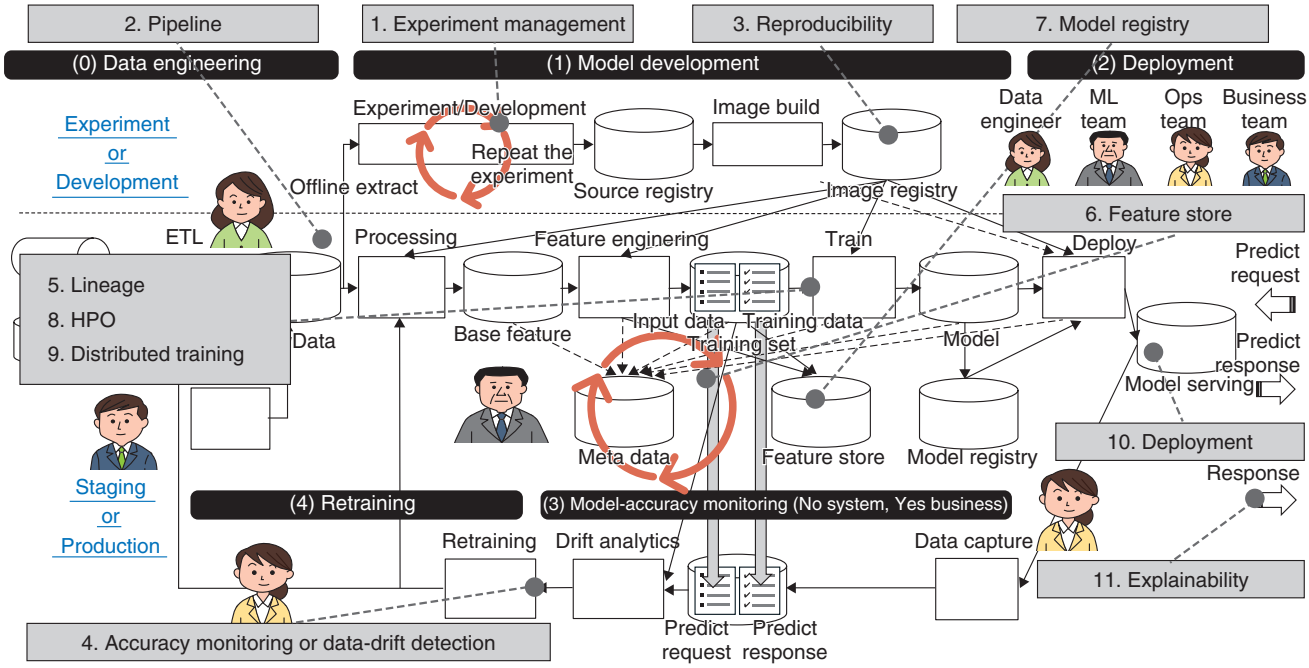
COVID-19 has spread rampantly around the world, and that situation has led to a change in people's purchasing habits and the need for fingerprint authentication instead of face recognition on smartphones because people wear masks. This situation is an example of the phenomena known as concept drift and data drift, and it represents the problem that AI models created thus far have become useless because the statistical properties of the data generated change due to changes in people's behavior. To deal with these phenomena, it is necessary to collect new data and rebuild the model with the collected data. However, it is inefficient to carry out that task manually every time this phenomenon arises; thus, it is necessary to automate the model-rebuilding process as a functional requirement and mechanism of the machine-learning system. In particular, for machine-learning models as well as general applications, the operation phase is crucial, and a mechanism for monitoring systematic error is not sufficient, that is, it is also necessary to monitor the accuracy of the machine-learning model.

The overall process of MLOps is shown in **Fig. 5**. Not only (1) model development and (2) deployment but also (3) model-accuracy monitoring and (4) training-data creation + retraining are required. If these processes are introduced together with the automation mechanism, it will be possible to adapt to concept drift or data drift as a system.

2. Future developments

MLOps is a recent technology trend, and the technology stack that underpins it, including definitions, is still immature. However, providers of OSS and cloud services as well as third-party vendors are working hard on developing MLOps tools and technologies, so it is necessary to continuously monitor these trends.

Although an industry-wide consensus definition of MLOps has not yet been established, the information provided by vendors, including those overseas, on MLOps can be organized into a system structure (**Fig. 6**) and 11 functional groups (**Table 1**). Due to space limitations, we cannot explain each function listed in the table, but it is expected that MLOps tools will continue to mature, and the social implementation of AI will accelerate in proportion to that



ETL: extract, transform, load
HPO: hyper parameter optimization

Fig. 6. Relationships between machine-learning system structure and MLOps tools.

maturing. As such a trend continues, various pieces of expertise about social implementation of MLOps and AI will continue to be accumulated, and by using that expertise, NTT DATA wants to contribute to the development of the machine-learning industry and the social implementation of AI.

Table 1. Eleven functional groups of MLOps.

Function name	Description	Expected benefits	
		Increase speed	Ease of use
1. Experiment management	After conducting a machine-learning experiment, the experiment can be traced later. Multiple experiments can be compared and visualized.	✓	
2. Automation of the pipeline	Multiple machine-learning processes can be stitched together into a job.	✓	
3. Reproducibility	It is possible to ensure packaging and portability with technologies such as containers so that machi-learning processing does not carried out in other environments due to the library on which machine-learning algorithms depend.		✓
4. Accuracy monitoring or data-drift detection	To be able to detect the deterioration in accuracy from the deviation in the distribution of the original data of the model, it is possible to compare the deviations between the baseline data and data acquired through Request.		✓
5. Lineage (tracking)	It is possible to trace the connection between multiple machine-learning processes (preprocessing, training, etc.) and the input and output of each process at a later phase by using an API (application programming interface), etc.		✓
6. Feature store	Feature-value data created by a data scientist in preprocessing can be shared with other data scientists and managed in a special storage area and portal.		✓
7. Model registry	Machine-learning models created by data scientists can be registered in a model-specific registry service and deployed to the environment later by “one click” of the approver.		✓
8. Hyper-parameter tuning	It is possible to execute a job dedicated to hyper-parameter tuning and select the job that generated the most accurate model.	✓	
9. Distributed training	Data partitioning or model decomposition can be used, and training of machine-learning models can be scaled out.	✓	
10. Multi-framework deployment	Generation of serving code for multiple machine-learning algorithms can be semi-automated, and advanced deployment methods (such as canary release) can be implemented.	✓	
11. Explainability	From the prediction results by machine learning, it is possible to determine what explanatory variables contribute to the results and explain the results.		✓

Trademark notes

Author: Ei Yamaguchi, NTT DATA Corporation

Business Application of BERT, a General-purpose Natural-language-processing Model

Tokuma Wachi

Abstract

BERT (Bidirectional Encoder Representations from Transformers) is gaining attention as an artificial intelligence (AI) technology that supports natural-language processing. To make BERT practical for business, NTT DATA is developing applications of BERT that can recognize the unique words and phrases used in various industries, making it possible to build an optimized AI model that meets the needs of individual customers. In this article, a financial version of BERT (Financial BERT) and domain-specific BERT framework that NTT DATA developed and is developing, respectively, are introduced as example applications.

Keywords: AI, natural-language processing, BERT

1. Introduction

Artificial intelligence (AI)-related technologies, such as deep learning, have made remarkable progress and are being introduced to fields that were previously impractical for application. Regarding image processing, it used to be difficult to distinguish between dogs and cats; however, by using deep learning, it has become possible to distinguish between dog breeds and even identify a dog's location at the pixel level. Regarding language processing, machine translation has improved dramatically over the last decade. In individual tasks, it is possible to automatically identify such matters as "The author of this text has a negative opinion." and "This word represents a person's name."

There has been a breakthrough with BERT (Bidirectional Encoder Representations from Transformers) [1], a technology that supports natural-language[*1] processing, which is attracting much attention. However, one of the challenges in applying BERT to business is that the expected accuracy in text classification cannot be achieved from documents that contain a large amount of domain-specific terminology of industries such as finance and healthcare. A financial version of BERT (Financial BERT) and domain-specific BERT framework that NTT DATA developed and is developing, respectively, to solve the above-described problem, are introduced in this article.

2. Natural-language-processing technology

Deep-learning technology has achieved high accuracy in a wide range of tasks, such as classification, detection, numerical prediction, and generation, and in certain tasks, it is even more accurate than humans. In 2015, in an image-classification benchmark task, it outperformed humans and gained much attention. The field of natural-language processing is also evolving. It is shifting from methods based on pattern recognition and frequency of occurrence to methods based on deep learning. As a result, it has become possible to automatically perform the following tasks:

- Analyze the text and decipher such sentiment as "The author of this sentence has a negative view." (positive-negative judgment)

*1 Natural language: A language that people use on a daily basis, such as Japanese and English.

- Quantify potential risks from the input textual data (scoring)
- Extract specific types of words (such as personal names and place names) from text (named-entity extraction)
- Retrieve information from the input documents and answer the related questions (question answering)

In natural-language processing, it is common to process input in units of words or sentences. It has recently become common to prepare a general-purpose model for processing words and sentences in general documents and then fine-tune the model for each specific task. This generic model is called a language model, and BERT, which is discussed below, is also a type of language model.

3. Overview of BERT

BERT is a general-purpose natural-language-processing model developed by Google. When BERT was released in 2018, it made headlines for breaking the previous records of various natural-language-processing benchmark tasks. For example, in the benchmark task described below, BERT outperformed human participants. In that task, participants were provided with sentences of about 140 words extracted from Wikipedia then asked to answer questions about the content.

The strength of BERT is that it makes it possible to solve problems in various domains and tasks by using a single model. Before BERT was developed, it was necessary to prepare a large amount of training[*2] data to develop a model for each task, because the model needs to learn the characteristics of the task from the training data from scratch. BERT can build a general-purpose model without training data by executing unsupervised pre-training with a large set of documents. Therefore, one significant achievement of BERT is that it makes it possible to construct a general-purpose language model simply by using a large number of sentences without any further processing such as annotation. In reality, however, only certain organizations with abundant technical capabilities and computing resources can construct such a general-purpose pre-training model.

To achieve high accuracy in a certain target task, only a small amount of training data is required to fine-tune the pre-trained generic BERT model. In the task of classifying the author's views into "positive" and "negative," for example, it is common to add a small weighting model in the latter part of the language-model layer, which can output the degree of "positive" and "negative" as numbers, and the final result will be output by comparing those numbers.

4. Japanese localization of BERT and NTT version of BERT

The target language of the original BERT released by Google is English. In Japan, institutions such as Kyoto University and the National Institute of Information and Communications Technology (NICT) have released Japanese pre-trained BERT models. The key to building a general-purpose pre-trained BERT model is to ensure the quality (diversity) and quantity of documents used for pre-training. Initially, the method of constructing a Japanese version of a pre-trained model was to use the full text of the Japanese version of Wikipedia (about 3 GB), which guarantees a certain level of quality and quantity and is easy to obtain. It has, however, become clear that this method does not perform well regarding spoken language, which appears less frequently than written language in Wikipedia.

NTT laboratories are constructing a Japanese pre-trained BERT model using a large-scale corpus (about 13 GB) they created, and the model outperforms the published pre-trained models in many tasks. Unless otherwise specified, the BERT described below refers to the BERT developed by NTT laboratories.

5. Additional training for domain-specific tasks

The original BERT and its Japanese version have performed better than conventional models. For example, in finance, they are expected to be used for automatic allocation of frequently answered question (FAQ) answers and risk information extraction from financial documents; and in the medical field, it is expected to be used in such cases as checking the content of electronic medical records and using information in medication package inserts. However, it has become a problem that the aforementioned method of fine-tuning a general-purpose model pre-trained with a large set of general documents cannot always achieve the expected level of accuracy in actual business applications.

*2 Training (data): Labeling data so that the AI model can interpret them in accordance with the task to be solved. For example, labelling a text reviewing a certain product as "favorable" or "unfavorable," and labelling "personal name" and "place name" appearing in a text. Adding information for training to data is called annotation.

Table 1. Difference in data scales.

*The concept of data scale varies with the target task and context. In this article, it is based on the categories listed in the following table.

Large scale	Medium scale	Small scale
A group of sentences the overall characteristics of which are difficult to capture manually	A group of sentences that is difficult to manually confirm completely, even by scanning	A group of sentences that can be manually confirmed and annotated completely
More than several gigabytes	Several megabytes to several hundred megabytes	Several kilobytes to several hundred kilobytes

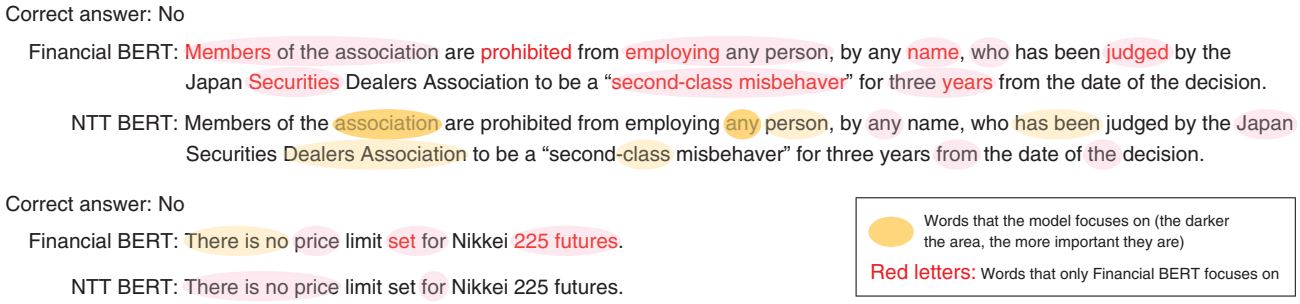


Fig. 1. Examples of questions answered correctly by Financial BERT.

This problem becomes apparent when the data of a business task are mostly specific to a particular domain (so-called domain-specific data). Examples of domain-specific data are listed as follows: data containing many technical terms in finance and the medical field, and driving-related data that contain specific knowledge of road-traffic laws and common practices. It is not realistic to prepare a large-scale domain-specific document collection for each domain; thus, to improve the accuracy of the general-purpose BERT model for domain-specific task, it is necessary to devise a means of handling domain-specific data.

A method for building a domain-adaptive pre-training model by training a pre-trained BERT model with additional small- to medium-scale groups of sentences (**Table 1**) has been proposed [2, 3]. In other words, a task-specific language model is created. Financial BERT and the domain-specific BERT framework adopt a similar additional pre-training approach.

6. Financial BERT

NTT DATA constructed a pre-trained model specialized for financial domains called Financial BERT by conducting additional pre-training with finance-related sentences collected from the Internet. To verify the performance of Financial BERT, we applied it to the qualification examination for the class-1 sales representative conducted by the Japan Securities Dealers Association [4], and compared its performance with other models. Financial BERT was the only model that achieved a score equivalent to the pass mark (308 points out of 440).

This qualification examination is taken by sales representatives who solicit securities transactions and derivative transactions. Between the two classes of the examination, class-1 is the higher-level qualification, and in 2019, 4633 examinees took the exam, and the pass rate was 67.6%. The content of the examination mainly consists of yes-or-no questions, where the examinee needs to tell whether the text is correct, and multiple-choice questions where the examinee is asked to select the correct one from five options. Many of the questions contain technical terms concerning financial products and knowledge about financial laws and regulations, so it is difficult to answer correctly by applying general knowledge only (**Fig. 1**).

The procedure of Financial BERT is explained as follows (**Fig. 2**).

(1) Collection of finance-related sentences from the Internet. (Web pages that can efficiently improve the performance of a BERT model are selected. The selection is based on the

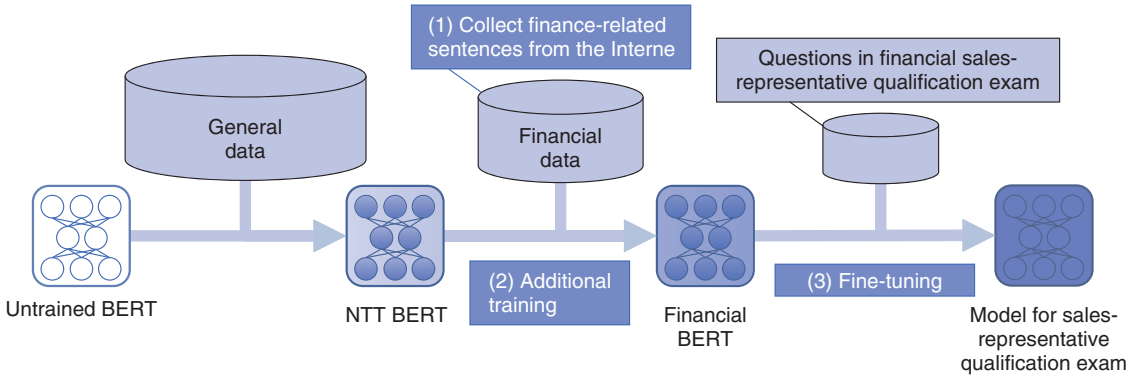


Fig. 2. Training and fine-tuning process of Financial BERT.

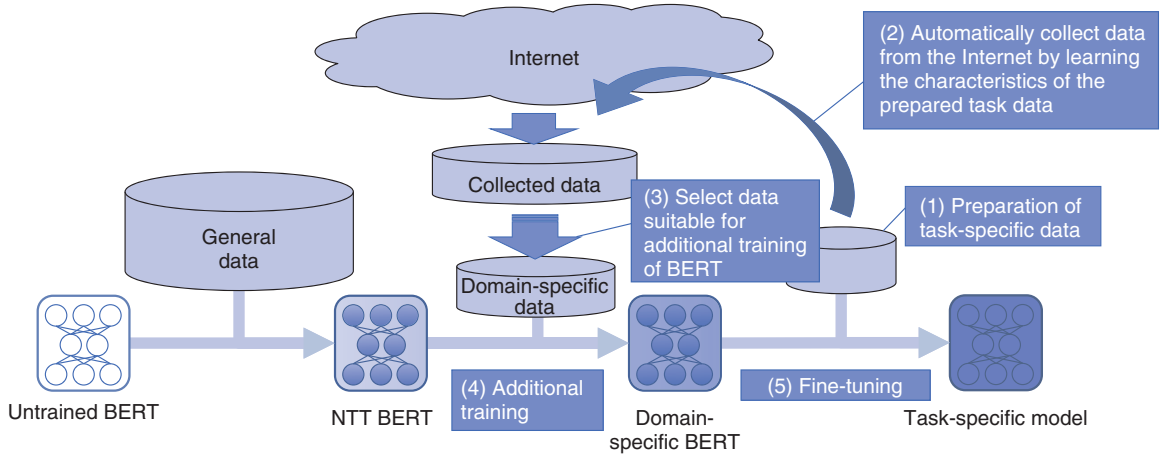


Fig. 3. Domain-specific BERT framework.

knowledge accumulated through the application of AI to the financial field in previous projects of NTT DATA.)

(2) Additional pre-training of the Japanese BERT developed by NTT using the sentences collected from the web pages using the approach mentioned above

(3) Fine-tuning of the pre-trained model with task-specific data

For (1) and (2), NTT DATA maintains pre-trained models; thus, when applying the model to projects of proof of concept (PoC) and system development, only the creation of training data and fine-tuning of the model for those projects are required.

7. Domain-specific BERT framework (under development)

The high accuracy of Financial BERT was achieved by additional training and fine-tuning the general-purpose BERT model. During the collection of financial sentences, web pages for experts were selected, and data were collected from those pages. Therefore, the cost for constructing the model is high and participation of experts is necessary. To solve these problems, NTT DATA is developing a domain-specific BERT framework for automating the collection of domain-specific data for additional training. The procedure of the domain-specific-BERT framework is described as follows (**Fig. 3**).

(1) Preparation of task-specific data (unsupervised training is sufficient at this point.)

(2) Automatic collection of sentences for additional pre-training from the Internet by learning the characteristics of the prepared task-specific data (i.e., extracting expressions that the general-purpose BERT cannot handle well from the task-specific data, generating queries from that information, and conducting Internet searches using those queries.)
(3) Selection (using an algorithm) of sentences that can be expected to improve accuracy for additional pre-training from the collected data. (i.e., sentences that can be expected to improve task accuracy are extracted from the sentence groups collected from the Internet, and a data set of domain-specific data is created.)
(4) Additional pre-training of the Japanese BERT developed by NTT using the previously selected sentence groups
(5) Fine-tuning the pre-trained model using the task-specific data (training data are required at this point.)

Similar to the application of Financial BERT, to apply the domain-specific BERT framework to actual PoC and system development, it is necessary to create task-specific training data and fine-tune the pre-trained model with those data. Higher accuracy is aimed for by executing automatic additional pre-training.

The strengths of the domain-specific BERT framework are (i) it is possible to construct an optimal model in accordance with customer data by automated data collection and selection and (ii) PoC and development periods can be shortened because there is no need to manually build a domain-specific BERT. Before BERT was developed, when handling domain-specific tasks that contain a large amount of technical terms, it was necessary to take specific measures for each task such as manually creating a dictionary. By using BERT, it is possible to build a general-purpose language model by using a large number of documents instead of building a dictionary. Using that model makes it possible to improve accuracy in various tasks. Moreover, it is expected that using the domain-specific BERT framework will further improve accuracy for domain-specific tasks.

8. Concluding remarks

In this article, BERT, a Japanese version of BERT, its application (Financial BERT), and a domain-specific BERT framework were introduced. The domain-specific BERT framework is currently being developed to further improve its accuracy and efficiency.

From 2021, we will provide the domain-specific BERT framework to customers in industries such as finance, healthcare, and manufacturing to help them create new businesses and improve the efficiency of existing businesses. Example use cases include automatic answering of FAQs, checking the content of electronic medical records, and detecting project risks from daily reports. We are also looking for PoC partners (in fields not limited to the aforementioned industries and use cases) to apply the domain-specific BERT framework to support our customers so that they can quickly apply BERT's advanced technology to their businesses.

Note: The technology described in this article has been tested only in Japanese. It can be applied to other languages, but customization is required.

References

[1] J. Devlin, M. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," arXiv:1810.04805, 2018.
[2] S. Gururangan, A. Marasović, S. Swayamdipta, K. Lo, I. Beltagy, D. Downey, and N. A. Smith, "Don't Stop Pretraining: Adapt Language Models to Domains and Tasks," arXiv:2004.10964, 2004.
[3] Y. Kawazoe, D. Shibata, E. Shinohara, E. Aramaki, and K. Ohe, "A Clinical Specific BERT Developed with Huge Size of Japanese Clinical Narrative," medRxiv, 2020. doi: https://doi.org/10.1101/2020.07.07.20148585
[4] Japan Securities Dealers Association, Sales Representative Qualification/Qualification Exam/Manual, https://www.jsda.or.jp/en/about/major-activities/html/examination-qualification.html

Author: Tokuma Wachi, Assistant Manager, Center for Digital Society Platform, R&D Headquarters, NTT DATA Corporation

Ultra-high-speed 300-GHz InP IC Technology for Beyond 5G

Hiroshi Hamada, Takuya Tsutsumi, Hideaki Matsuzaki, Hiroki Sugiyama, and Hideyuki Nosaka

Abstract

A 300-GHz-band 120 Gbit/s wireless transceiver (TRX) is presented using our in-house indium phosphide (InP) high-electron-mobility transistor (InP-HEMT) technology. A 300-GHz power amplifier (PA), which is the key component in the TRX, was developed using the backside DC line (BDCL) technique to increase its gain and output power. The measured maximum gain and saturated output power are respectively 20.5 dB and 12 dBm. The 300-GHz-band TRX was fabricated using this PA. The TRX achieves high data rates of 124 and 120 Gbit/s under back-to-back and 9.8-m-link-distance wireless data transmission conditions. To the best of our knowledge, these are the highest data rates among reported 300-GHz-band TRXs.

Keywords: 300 GHz, InP-HEMT, power amplifier

1. Introduction

A terahertz (THz) wave is an electromagnetic wave located in the boundary of a radio wave and light wave, as illustrated in **Fig. 1**. Its frequency range is from around 300 GHz to 10 THz. It had been very difficult to generate and control THz waves due to the lack of semiconductor devices working in this high frequency region. Due to the progress of miniaturization technique for semiconductor devices (transistors), the operation frequency range of cutting-edge transistors is above 1 THz [1]. By using these high-speed transistors, the applications of THz waves, such as wireless communication [2] and imaging/sensing for security [3], are being extensively investigated. High-speed wireless communication is one of the major applications of THz waves due to their broad bandwidth. In the next generation of the fifth-generation mobile communication system (5G), called beyond 5G or 6G, a data rate of more than 100 Gbit/s is considered necessary. The 300-GHz-band is considered suitable for beyond 5G due to its relatively low atmospheric attenuation (< 10 dB/km) in the THz region. In this article, we report on the recent achievements with a 300-GHz-band power amplifier (PA) we developed using our in-house indium phosphide (InP) high-electron-mobility transistor (InP-HEMT) integrated circuit (IC) technology and over-100 Gbit/s wireless transceiver front-end (TRX) fabricated using this PA.

2. InP-HEMT technologies

InP device technologies to fabricate 300-GHz-band ICs for TRXs are introduced in this section.

The mandatory device technology for 300-GHz-band ICs is the high-speed transistor InP-HEMT. There are two frequently used figures of merit for high-speed transistors, i.e., transition frequency (f_T) and maximum oscillation frequency (f_{MAX}). The former is an index to show the maximum switching speed and the latter shows the maximum power-amplification frequency of the transistor. Therefore, f_T is related to the operation frequency of a switching circuit such as a passive mixer and multiplier, whereas f_{MAX} is related to the operation frequency of active circuits such as an amplifier and oscillator. A cross-section schematic of our in-house InP-HEMT is shown in **Fig. 2(a)**. Generally, InP-HEMTs have superior high-speed characteristics by using the

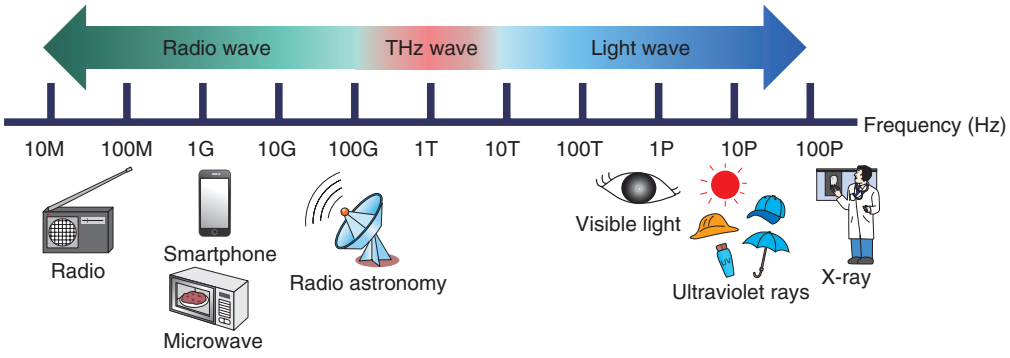


Fig. 1. THz wave.

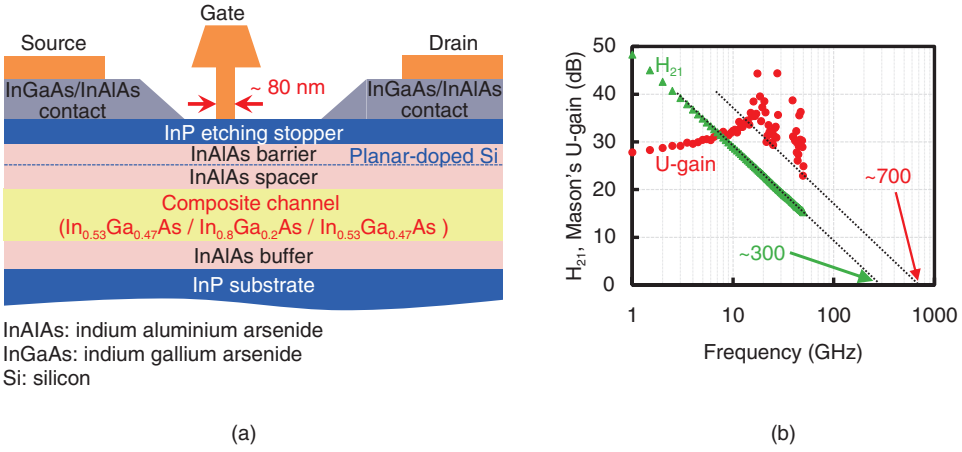


Fig. 2. (a) Schematic and (b) RF characteristics of in-house InP-HEMT.

high-electron-mobility material indium gallium arsenide ($In_{0.53}Ga_{0.47}As$), which can be epitaxially grown (i.e., lattice-matched with InP) on InP substrate as their channels. With our InP-HEMT technology, to attain higher speed than with normal InP-HEMTs, a composite channel composed of In-rich $In_{0.8}Ga_{0.2}As$ and lattice-matched $In_{0.53}Ga_{0.47}As$ [4] is applied using InGaAs because the ratio increase of In can enhance its electron mobility. The gate length also decreased to 80 nm to shorten the electron transit time in the channel and enhance both f_T and f_{MAX}. The measured current gain (H_{21}) and maximum unilateral gain (U-gain) of this InP-HEMT are plotted in **Fig. 2(b)**, showing high f_T and f_{MAX} of 300 and 700 GHz, respectively.

To fabricate 300-GHz-band ICs, it is not sufficient to have high-speed transistors. The problem specific to THz ICs caused by a substrate mode as described below should be managed. A substrate mode is an electromagnetic wave that propagates in substrate. When the substrate thickness is the same order of wavelength, the substrate mode can propagate. The commercially available InP substrate thickness is around 600 μm and the wavelength of a 300-GHz-band signal in InP substrate is below 500 μm. Therefore, a substrate mode can be guided, causing unwanted coupling between some ports of THz ICs, e.g., coupling of input and output ports of the amplifier by the substrate mode can cause the oscillation of that amplifier. To cut out the propagation of the substrate mode, substrate thinning and through substrate via (TSV) are applied in the back-end IC process [5, 6]. By using substrate thinning, the thickness of the InP substrate is reduced to 55 μm, as shown in **Fig. 3(a)**. TSVs are densely formed over the entire substrate area to reduce the substrate space that can

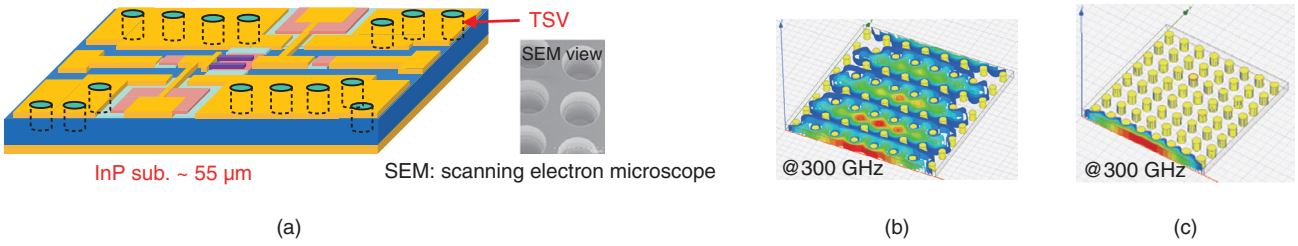


Fig. 3. (a) InP chip with substrate thinning and dense TSV formation, substrate-mode propagation for TSV pitches of (b) 100 μm and (c) 50 μm.

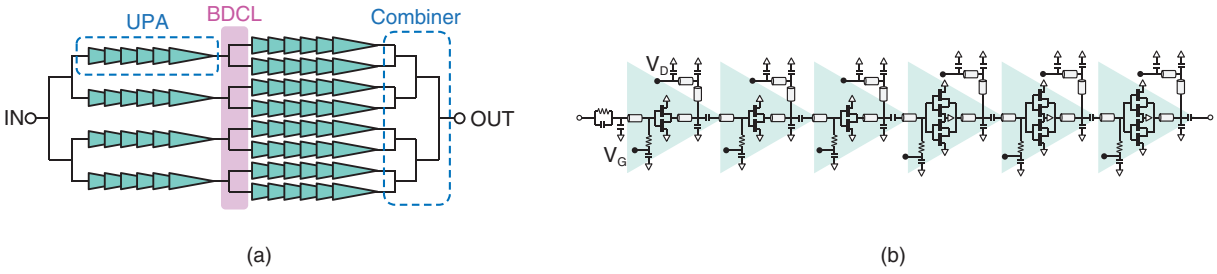


Fig. 4. Schematics of (a) 300-GHz PA and (b) UPA.

support the propagation of substrate modes, as also shown in Fig. 3(a). The TSV density is critical to sufficiently cut out the substrate mode. **Figures 3(b)** and **(c)** shows a 300-GHz substrate-mode propagation calculated using an electromagnetic simulator when the TSV pitch (TSV edge-to-edge distance) is set to 100 and 50 μm, respectively. By using the 50-μm pitch, the substrate mode is sufficiently cut out. Therefore, around the 50-μm pitch, the TSV layout is used in the 300-GHz-band PA, as described in the next section.

3. 300-GHz PA

The important characteristics of the PA for TRXs are gain and output power to achieve a sufficient signal-to-noise ratio (SNR) of 300-GHz-band wireless communication. A schematic of the PA is shown in **Fig. 4(a)**. It is composed of unit power amplifiers (UPAs) consisting of six-stage common-source amplifiers using the InP-HEMTs described in Section 2, as illustrated in **Fig. 4(b)**. This six-stage cascading design produces high gain of more than 10 dB for a UPA at 300 GHz. The inter-stage matching for each common-source amplifier stage in a UPA is designed to have small loss by using the low impedance matching technique [7, 8]. To achieve both high gain and high output power in a UPA, the first three stages are designed with high-gain two-fingered HEMTs, and the latter three stages are designed with four-fingered HEMTs that have higher power handling and slightly lower gain than two-fingered ones, as shown in Fig. 4(b). The PA has eight-paralleled UPAs in its output stage to achieve high output power by combining the output power of each UPA. In the middle part of the PA, the backside DC line (BDCL) is used to also achieve high gain and output power. The role of the BDCL technique is explained as follows.

The PA shown in Fig. 4 (a) uses many transistors due to the series-amplifier stage in a UPA and paralleled fashion in the PA output stage. In such a case, a very wide (several hundred microns) DC line should be used to support large total drain current (~ 1.8 A), which is necessary to operate many transistors. Therefore, the typical layout of this PA is similar to the one shown in **Fig. 5(a)**. Long radio frequency (RF) transmission lines (TLs) should overlap the wide DC line. These long RF TLs have high transmission loss in a high frequency range, such as the 300-GHz band, and reduce the gain and output power of the PA. To address this issue, the BDCL is introduced. With this introduction, the wide DC line is put on the

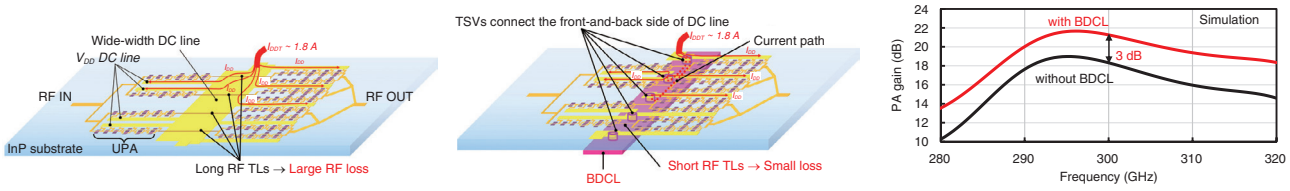


Fig. 5. Layout schematics of 300-GHz PA (a) without and (b) with BDCL and (c) PA gain comparison of these layouts.

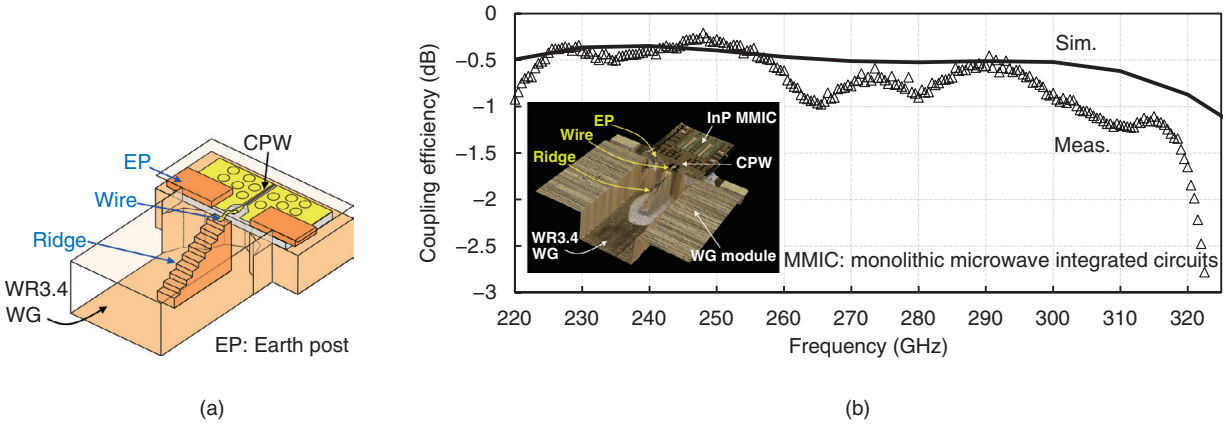


Fig. 6. (a) Schematic and (b) measured coupling efficiency of ridge coupler.

backside of the InP chip (**Fig. 5(b)**). Thus, the overlap between the RF TLs and wide DC lines in Fig. 5(a) is eliminated and the gain and output power of the PA increases. The simulation comparison of the PA gain is shown in **Fig. 5(c)**. By using the BDCL, the simulated gain increases by around 3 dB over 280–320 GHz.

The PA chip is difficult to handle for implementation in TRXs. To improve ease of handling, the IC packaging in a robust metal module is important. In the 300-GHz band, the widely used media is a rectangular waveguide (WG). Therefore, we developed a PA module with WG flanges for its input and output. The critical component for this PA module is the transition between the PA IC and WG. A ridge coupler [8, 9], shown in **Fig. 6(a)**, is used as the transition. It translates the guided mode of the WG to the coplanar waveguide (CPW) mode, which is used in the PA IC. A step-wise metal component formed on the center part of the WG, called a ridge, gradually transforms the impedance and forms of electromagnetic wave between the WG and CPW. A bondwire is used to connect the ridge and IC CPW pad. A three-dimensional photograph and simulated and measured coupling efficiencies of this ridge coupler are shown in **Fig. 6(b)**. The coupling loss is very small, less than 1 dB over 220–303 GHz.

The PA IC and module were fabricated using the techniques described above, as shown in **Figs. 7(a)** and **(b)**. The BDCL placed in the vicinity of the center part of the PA chip is electrically isolated from the DC ground by the mask and etching process in the InP back-end process, as shown in Fig. 7(a). The fabricated PA module has WR3.4-band (220–325 GHz) WG flanges for its input and output, as shown in Fig. 7(b).

The characteristics of the PA module were evaluated through small- and large-signal measurement. We first conducted a small-signal measurement using a vector network analyzer (VNA) and WR3.4-band frequency extenders. The measured S-parameters are shown in **Fig. 8**. It achieved a maximum gain (S_{21}) of 20.5 dB at 295 GHz and broad bandwidth. The reverse transmittance (S_{12}) is very small, less than

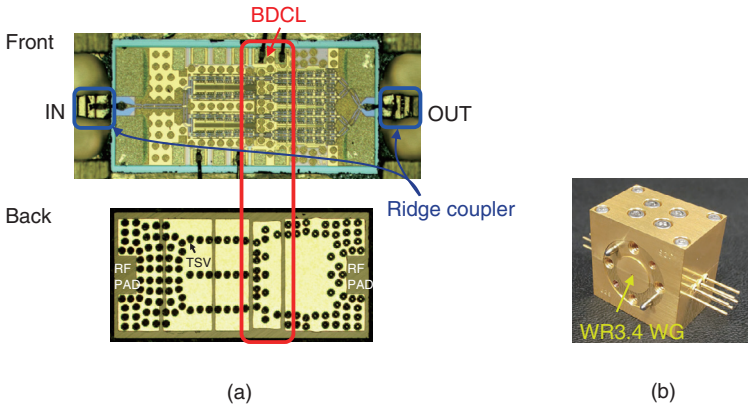


Fig. 7. Photographs of (a) 300-GHz PA chip and (b) module.

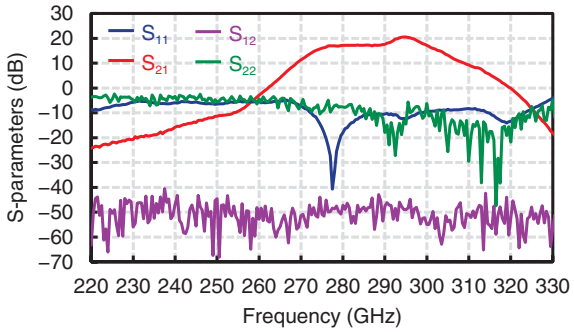


Fig. 8. Measured S-parameters of 300-GHz PA module.

–40 dB over the full WR3.4 band. Therefore, the PA module is very stable. This is because the InP back-end process shown in Fig. 3 successfully reduces substrate-mode propagation. The measured gain matches the simulated gain in Fig. 6(a).

Next, we measured large signal characteristics. The input-output characteristics are shown in **Fig. 9(a)**. The PA module achieved high saturated output power (Psat) and output 1-dB compression point (OP1dB) of 6 dBm. The measured frequency dependence of the Psat is shown in **Fig. 9(b)**. The Psat is larger than 10 dBm over 278–302 GHz. These high-power characteristics are derived from BDCL technique and the low-loss ridge coupler, as described above.

4. 300-GHz-band TRX

We fabricated a 300-GHz-band TRX using the PA module described in Section 3. The TRX has a heterodyne architecture. The transmitter (TX) consists of a frequency converter (mixer), local oscillation (LO) PAs to operate these mixers, and an RF PA (same as the one described in Section 3). The receiver (RX) is composed of a mixer, LO PA, and low noise amplifier (LNA). The intermediate frequency (IF) and LO frequency are set to 20 and 270 GHz, respectively. The fundamental mixer [8, 10] using our in-house InP-HEMT technology is used for both TX and RX. In the RX, the same PA discussed in Section 3 is used as the LNA. This TRX uses the upper-side band for the RF (290 GHz). High-gain (50 dBi) lensed horn antennas were used for both TX and RX in the wireless-transmission experiment discussed later. The lower-side band signal is cut by using the high pass filter (HPF), as shown in **Fig. 10**. Sixteen quadrature amplitude modulation (16QAM) is used as the modulation format for communication.

First, we conducted an experiment involving back-to-back data transmission. The TX and RX were directly connected through the attenuator (ATT) with

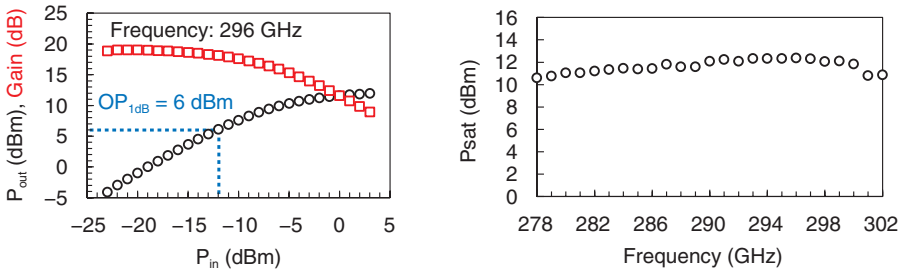


Fig. 9. Measured (a) input-output characteristics and (b) frequency dependence of Psat of 300-GHz PA module.

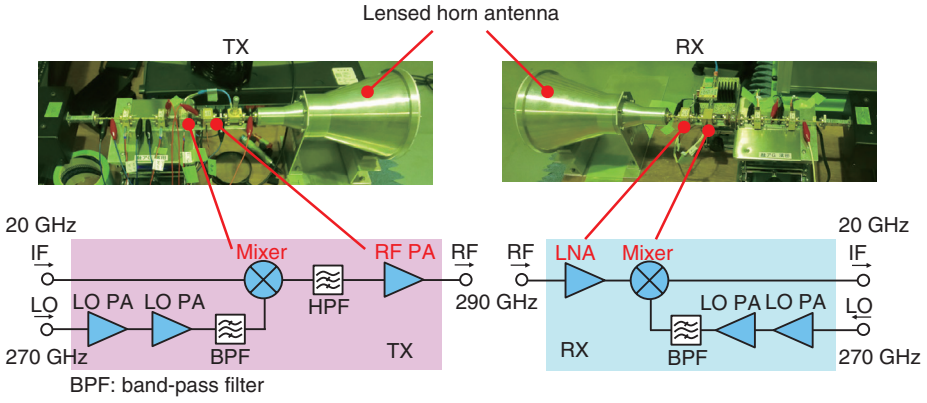


Fig. 10. Schematic and photograph of 300-GHz-band TRX.

attenuation of 9 dB. The role of this ATT is to protect the RX from the high output power of the PA. The SNR of the received IF signal was measured to evaluate the quality of communication. The SNR is related to the bit error rate (BER). The communication is judged successful when the measured SNR is larger than the required SNR (SNRreq), which is determined as the SNR where the corresponded BER is 10^{-3} in the following measurement. The SNRreq of 16QAM is 16.5 dB. The measured constellations and dependence of the baud rate with SNR of the received IF signal is shown in **Fig. 11**. As the frequency utilization efficiency of 16QAM is 4, the data rate is four times the baud rate. In the low baud rate region, the SNR is quite high, more than 25 dB. The SNR gradually degrades with the baud rate because the noise floor also increases with the increase in the baud rate. The maximum baud rate is 31 Gbaud, which corresponds to the very high data rate of 124 Gbit/s.

Next, we conducted an experiment involving wireless communication with the TRX with the 50-dBi lensed horn antennas described above. The link distance (TX antenna to RX antenna distance) was fixed to 9.8 m (**Fig. 12**). However, this link distance was virtually changed by changing the attenuation value of the variable ATT (VATT) inserted between the TX and TX antenna. This VATT equivalently adds the path loss between the TX and RX, and the path loss can be converted to the equivalent link distance. The measured SNR versus equivalent link distances of 15, 20, 25, and 30 Gbaud using the 16QAM signal are shown in **Fig. 13**. The high data rate of 120 Gbit/s (30 Gbaud) wireless transmission is successfully demonstrated. The received constellation of the 120 Gbit/s IF signal is also shown in Fig. 13. The maximum link distances, which are defined as the equivalent link distance where the SNR is the same as the 16QAM SNRreq, are 42, 29.5, 17.5, and 10.5 m for 15, 20, 25, and 30 Gbaud. The comparison of recently reported near-300 GHz TRX is shown in **Fig. 14**. The fabricated

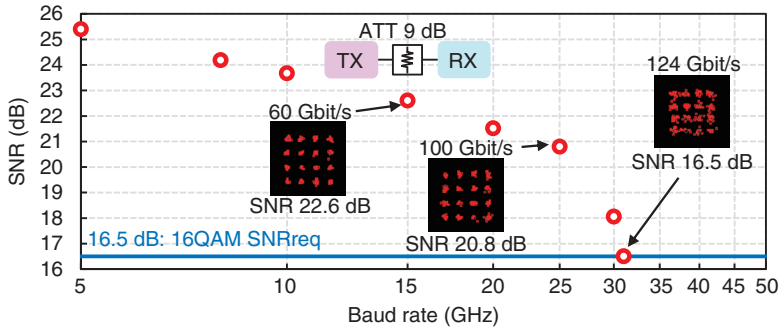


Fig. 11. Measured SNR vs. baud rate of 300-GHz-band TRX under back-to-back condition.

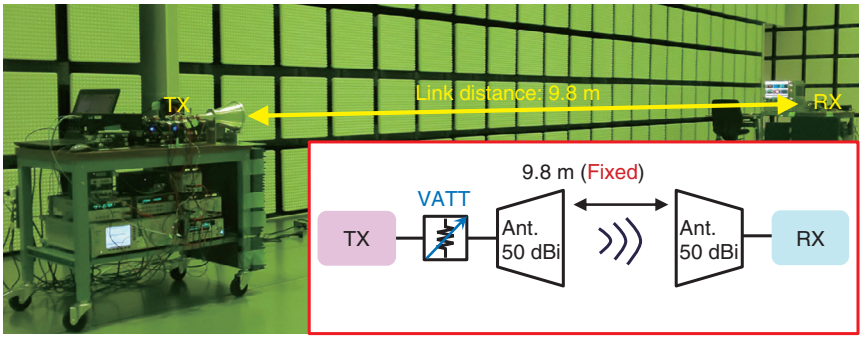


Fig. 12. Schematic and photograph of wireless-transmission experiment with 300-GHz-band TRX.

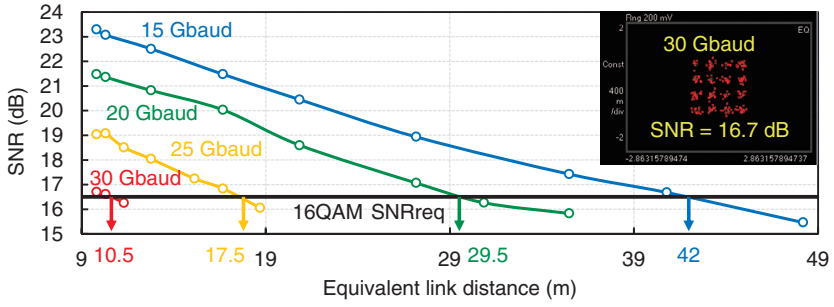
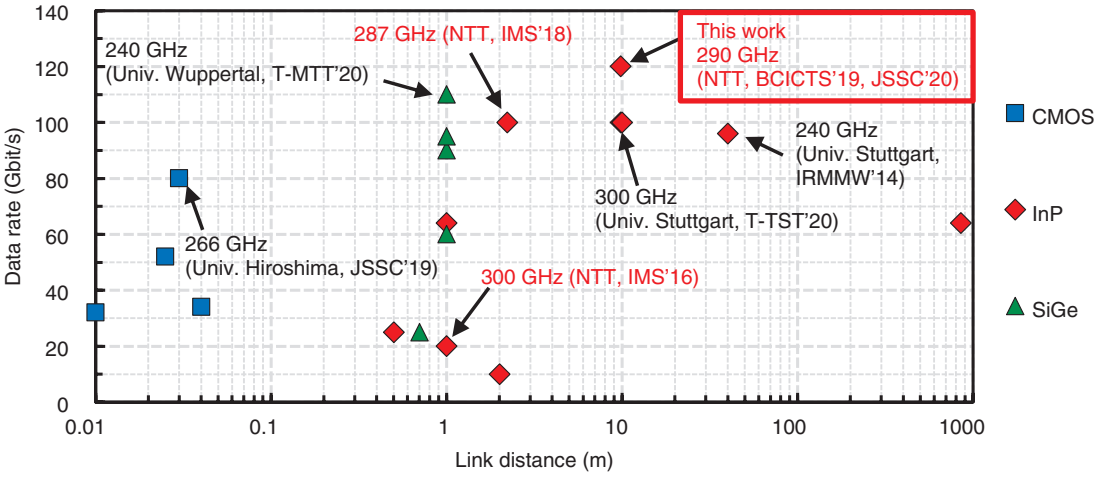


Fig. 13. Measured SNR vs. baud rate of 300-GHz-band TRX in 9.8-m wireless-communication experiment.

TRX shows the highest data rate among them.

5. Conclusion

We achieved 300-GHz-band 120-Gbit/s 9.8-m wireless transmission with a 300-GHz-band TRX using the high output power PA fabricated using our in-house InP-HEMT technology. The PA uses a special BDCL to lessen the RF loss and enhance its gain and output power. It achieves a maximum gain of 20.5 dB at 295 GHz, and Psat and OP1dB of 12 and 6 dBm at 296 GHz. The Psat is larger than 10 dBm for 278–302 GHz. The TRX was fabricated using the PA and our in-house 300-GHz fundamental mixer. It achieves a maximum data rate of 124 Gbit/s in 16QAM back-to-back data transmission. Wireless



CMOS: complementary metal-oxide semiconductor
SiGe: silicon-germanium

Fig. 14. Plot of data rate and link distance for reported 300-GHz-band TRXs.

data transmission was successfully demonstrated with a link distance of 9.8 m for data rates of 60, 80, 100, and 120 Gbit/s using 16QAM. To the best our knowledge, the fabricated TRX achieves the highest data rate among the reported 300-GHz-band TRXs.

Acknowledgments

The authors wish to acknowledge Kenichi Okada of Tokyo Institute of Technology and Ho-Jin Song of NTT Device Technology Labs (now working at Pohang University of Science and Technology) for their fruitful discussions. We also thank the continuous assistant of Yukio Yago at NTT Electronics Techno Corporation. This work was supported in part by the Ministry of Internal Affairs and Communications, Japan, through the Research and Development Program for Expansion of Radio Resources.

References

[1] W. R. Deal, K. Leong, W. Yoshida, A. Zamora, and X. B. Mei, "InP HEMT Integrated Circuits Operating above 1,000 GHz," Proc. of IEEE International Electron Device Meeting (IEDM), San Francisco, CA, USA, Dec. 2016.

[2] H. Elayan, O. Amin, R. M. Shubair, and M. Alouini, "Terahertz Communication: The Opportunities of Wireless Technology Beyond 5G," Proc. of International Conference on Advanced Communication Technologies and Networking (CommNet), Marrakech, Morocco, Apr. 2018.

[3] P. Hillger, J. Grzyb, R. Jain, and U. R. Pfeiffer, "Terahertz Imaging and Sensing Applications with Silicon-based Technologies," IEEE Trans. Terahertz Sci. Technol., Vol. 9, No. 1, pp. 1–19, Jan. 2019.

[4] H. Sugiyama, H. Matsuzaki, H. Yokoyama, and T. Enoki, "High-electron-mobility $In_{0.53}Ga_{0.47}As/In_{0.8}Ga_{0.2}As$ Composite-channel Modulation-doped Structures Grown by Metal-organic Vapor-phase Epitaxy," Proc. of International Conference on Indium Phosphide and Related Materials (IPRM), Takamatsu, Japan, June 2010.

[5] H. Hamada, T. Tsutsumi, H. Sugiyama, H. Matsuzaki, H.-J. Song, G. Itami, T. Fujimura, I. Abdo, K. Okada, and H. Nosaka, "Millimeter-wave InP Device Technologies for Ultra-high Speed Wireless Communications toward Beyond 5G," Proc. of IEEE International Electron Device Meeting (IEDM), San Francisco, CA, USA, Dec. 2019.

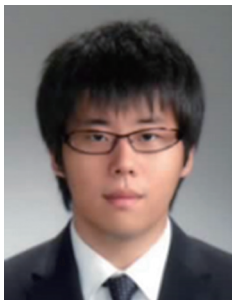
[6] T. Tsutsumi, H. Hamada, K. Sano, M. Ida, H. Matsuzaki, "Feasibility Study of Wafer-level Backside Process for InP-based ICs," IEEE Trans. Electron Devices, Vol. 66, No. 9, pp. 3771–3776, Sept. 2019.

[7] H. Hamada, T. Tsutsumi, G. Itami, H. Sugiyama, H. Matsuzaki, K. Okada, and H. Nosaka, "300-GHz 120-Gb/s Wireless Transceiver with High-output-power and High-gain Power Amplifier Based on 80-nm InP-HEMT Technology," Proc. of IEEE BiCMOS and Compound Semiconductor Integrated Circuits and Technology Symposium (BCICTS), Nashville, TN, USA, Nov. 2019.

[8] H. Hamada, T. Tsutsumi, H. Matsuzaki, T. Fujimura, I. Abdo, A. Shirane, K. Okada, G. Itami, H.-J. Song, H. Sugiyama, and H. Nosaka, "300-GHz-Band 120-Gb/s Wireless Front-End Based on InP-HEMT PAs and Mixers," IEEE J. Solid-State Circuits, Vol. 55, No. 9, pp. 2316–2335, Sept. 2020.

[9] T. Kosugi, H. Hamada, H. Takahashi, H.-J. Song, A. Hirata, H. Matsuzaki, and H. Nosaka, "250–300 GHz Waveguide Module with Ridge-coupler and InP-HEMTIC," Proc. of IEEE Asia-Pacific Microwave Conference (APMC), pp. 1133–1135, Sendai, Japan, Nov. 2014.

[10] H. Hamada, T. Fujimura, I. Abdo, K. Okada, H.-J. Song, H. Sugiyama, H. Matsuzaki, and H. Nosaka, "300-GHz, 100-Gb/s InP-HEMT Wireless Transceiver Using a 300-GHz Fundamental Mixer," Proc. of IEEE MTT-S International Microwave Symposium (IMS), pp. 1480–1483, Philadelphia, PA, USA, June 2018.



Hiroshi Hamada
Research Engineer, NTT Device Technology Laboratories.

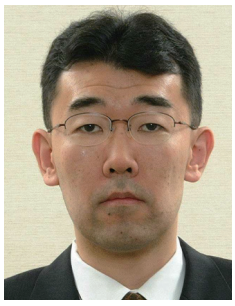
He received a B.E. and M.E. in electrical engineering from Tokyo Institute of Technology in 2009 and 2011. He joined NTT Photonics Laboratories in 2011, where he started the research and development of the millimeter-wave/terahertz (MMW/THz) monolithic microwave integrated circuits (MMICs) for wireless communications. He is currently with NTT Device Technology Laboratories, where he is engaged in the research and development of THz MMICs for ultrahigh-speed wireless communications and THz imaging and sensing technologies. His research interests include MMW/THz IC design, package design, and device modeling of III–V transistors. Mr. Hamada is a member of the Institute of Electrical and Electronics Engineers (IEEE) and Institute of Electronics, Information and Communication Engineers (IEICE). He has been serving as a member for the IEEE Microwave Theory & Techniques Society (MTT-S) Technical Committee on Microwave and Millimeter-Wave Solid State Devices (MTT-9). He was a recipient of the IEEE International Microwave Symposium Best Industry Paper Award in 2016, the URSI (Union Radio-Scientifique Internationale) Asia-Pacific Radio Science Conference Young Scientist Award in 2016, the APMC (Asia-Pacific Microwave Conference) Prize in 2018, and the IEICE Young Researcher's Award in 2019.

Takuya Tsutsumi
Senior Research Engineer, NTT Device Technology Laboratories.

He received a B.E. in electrical engineering from Osaka City University in 2006, and an M.E. and Ph.D. in electronic engineering and informatics from Kyoto University in 2008 and 2018. In 2008, he joined NTT Photonics Laboratories. From 2013 to 2016, he was with NTT Access Network Service Systems Laboratories, where he was engaged in the development of optical network systems. He is currently with NTT Device Technology Laboratories, where he is involved in the research of InP-HEMT devices and the development of backside fabrication processes for high-speed ICs. He is a member of IEEE.

Hideaki Matsuzaki
Senior Research Engineer, Supervisor, NTT Device Technology Laboratories.

He received a B.S. and M.S. in physics from Kyoto University in 1993 and 1995, and a Ph.D. in engineering from Toyama University in 2019. He joined NTT Atsugi Electrical Communications Laboratories in 1995. He is with NTT Device Technology Laboratories, where he engages in research and development of compound semiconductor devices such as InP-HEMTs, HBTs, photodiodes, and laser-diodes. He is a senior member of IEEE, senior member of IEICE, and a member of the Institute of Electrical Engineers of Japan.



Hiroki Sugiyama
Distinguished Laboratory Specialist, Senior Research Engineer, NTT Device Technology Laboratories.

He received a B.S. and M.S. in physics from Tokyo Institute of Technology in 1991 and 1993. He joined NTT in 1993, where he has been engaged in research and development of epitaxial growth and characterization technology of III–V compound semiconductors for ultrahigh-speed electron devices. He is a member of the Japan Society of Applied Physics and the Physical Society of Japan.

Hideyuki Nosaka
Senior Research Engineer, Supervisor, Group Leader of High-Speed Analog Circuit Research Group, NTT Device Technology Laboratories.

He received a B.S. and M.S. in physics from Keio University, Kanagawa, in 1993 and 1995, and a Dr. Eng. in electronics and electrical engineering from Tokyo Institute of Technology in 2003. He joined NTT Wireless System Laboratories in 1995, where he was engaged in research and development of MMICs and frequency synthesizers. Since 1999, he has been with NTT Photonics Laboratories, where he has been involved in research and development of ultrahigh-speed mixed-signal ICs for optical communications systems. He is a member of IEICE and served as a TPC member for the IEEE CSICS from 2011 to 2013 and IEEE ISSCC from 2013 to 2017. He has been serving as a member for the IEEE MTT-S Technical Committee on RF/Mixed-Signal Integrated Circuits and Signal Processing (MTT-15). He was a recipient of the 2001 Young Engineer Award, the 2012 Best Paper Award presented by IEICE, and the APMC 2018 Prize.

Standardization Trends on Cryptographic Algorithms and Protocols in ISO/IEC JTC 1 SC 27 WG 2

Keita Xagawa, Ryo Kikuchi, Atsunori Ichikawa, and Takayuki Miura

Abstract

International Organization for Standardization/International Electrotechnical Commission (ISO/IEC) Joint Technical Committee 1 Subcommittee 27 has developed and standardized methods, technologies, and guidelines for information security and privacy. Working Group 2 (WG 2) is responsible for the development and standardization of cryptographic and other security mechanisms. We introduce the latest standardization trends of cryptographic algorithms and protocols in WG 2.

Keywords: lightweight encryption, anonymous signature, elliptic curve cryptography, secret sharing, secure multiparty computation

1. Overall introduction to ISO/IEC JTC 1 SC 27 WG 2

International Organization for Standardization/International Electrotechnical Commission (ISO/IEC) Joint Technical Committee 1 (JTC 1) Subcommittee 27 (SC 27) is a standards body dealing with methods, technologies, and guidelines concerning information security and privacy protection. Working Group 2 (WG 2) deals with cryptographic and other security mechanisms. We discuss various methods, from basic encryption methods (e.g., blockciphers and hash functions) to advanced protocols (e.g., anonymous authentication and secure multiparty computation).

2. Lightweight encryption (ISO/IEC 29192 series and ISO/IEC 18033-7)

The performance of a cryptographic mechanism can be measured using various indicators such as computation time, latency, the power consumed, area of hardware implementation, and memory size used for computation. If a device runs on battery, low-power encryption is required, and if it runs in high-transmission environments, less computation time is required. If a device runs in environments where real-time performance is important, such as sensors in the human body, vehicles, and robots in factories, low-latency encryption is required. Lightweight cryptography refers to cryptography that is used in such environments and is "lighter" than existing standardized cryptography with certain indicators. Research on lightweight cryptography began in the early 2000s, and is still active in light of recent safety requirements for devices and trends in Internet of Things.

ISO/IEC has also established the ISO/IEC 29192 series, which summarize lightweight cryptography, and ISO/IEC 29167 series, which define cryptographic technology for radio frequency identification. The US National Institute of Standards and

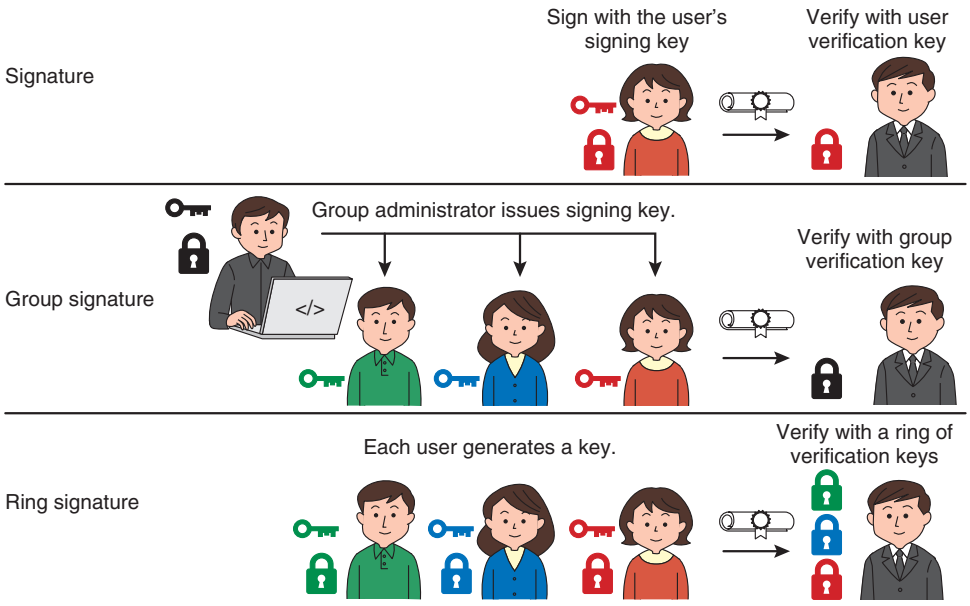


Fig. 1. Anonymous signature.

Technology (NIST) has been studying lightweight cryptography since 2014 and called for applications in 2018. The selection process (Round 2) is now in progress, which involves decreasing the number of candidates. Cryptography Research and Evaluation Committees (CRYPTREC) in Japan also studies lightweight cryptography and published the summary "CRYPTREC Cryptographic Technology Guidelines (lightweight cipher)" [1] in 2017.

WG 2 is responsible for the ISO/IEC 29192 series for various lightweight cryptographic mechanisms. The standard is divided into lightweight blockcipher (Part 2), lightweight stream cipher (Part 3), lightweight hash function (Part 5), lightweight message authentication code (MAC) (Part 6), and so on. ISO/IEC 29192-6 (Lightweight MAC) was published in 2019 and contains Chasky-12 and LightMAC, in which NTT is involved, proposed from Japan.

The standardization of tweakable blockcipher as ISO/IEC 18033-7 began in 2020. This standard will contain [2] Deoxys-BC in Deoxys, which is one of the lightweight ciphers selected by CAESAR (Competition for Authenticated Encryption: Security, Applicability, and Robustness), and SKINNY, in which NTT is involved.

3. Anonymous signature (ISO/IEC 20008 series)

Digital signatures are a basic technology used in various fields. The signer generates a signing key and verification key and publishes the verification key. The signer then generates a signature on the document using the signing key, and the verifier verifies the document and signature with the verification key.

When we look at a document and signature, we know which signer signed it. Thus, when a digital signature is used as an authentication, it is possible to know which verification-key owner carried out the authentication.

Some applications require that signers remain anonymous. For this purpose, a technique called anonymous signature has been studied. The ISO/IEC 20008 series deal with anonymous signatures (**Fig. 1**).

3.1 Group signature (ISO/IEC 20008-2)

In the group-signature scheme proposed in 1991, the group administrator issues a signing key to each user. A user signs the document with the signing key, and the verifier verifies the document and signature with the group's verification key. In this case, the verifier only knows that the signer belongs to the group. This protects the anonymity of the user. The standardization of group signatures began around 2010 as ISO/IEC 20008-2 and was published as a standard in 2013.

3.2 Ring signature (ISO/IEC 20008-3)

In the ring-signature scheme proposed in 2001,

each user has a signing key and verification key. The user signs the document with a signing key along with other verification keys. The verifier verifies the set of verification keys (called a ring), document, and signature.

In this case, the verifier only knows that the signer is the owner of one of the verification keys in the ring. This also protects the anonymity of the user. Compared with group signatures, it is less centralized because it does not require a group administrator.

In 2020, ISO/IEC 20008-3 started to standardize ring signatures in response to decentralization and certain blockchains incorporating them as a technology for achieving anonymity in electronic cash. NTT has also been conducting research and development (R&D) of ring signatures since the early 2000s, and actively providing suggestions and feedback.

4. Elliptic curve for pairing (ISO/IEC 15946-5)

Public-key cryptography is based on mathematical problems. For example, Diffie and Hellman's paper [3], which advocated public key cryptography, describes a key-exchange method based on the discrete logarithm problem. Then, around 1985, the construction of public key cryptography based on the discrete logarithm problem on elliptic curves was proposed. It was found that the sizes of the key and ciphertext could be made smaller than those of the usual discrete logarithm problem, and R&D advanced. Around 2000, it was discovered that by using the pairing function defined on an elliptic curve, one could construct novel cryptosystems (identity-based encryption, efficient threshold encryption, etc.) that were not possible before.

WG 2 has been standardizing elliptic curve cryptography, and from 1999 to 2004 this cryptography was compiled as the ISO/IEC 15946 series. Elliptic curve cryptography-based key establishment was later re-classified in ISO/IEC 15946-3 as another standard, and elliptic curve cryptography-based signatures were reclassified in 15946-2, -4 as signature standard. ISO/IEC 15946-1, which defines the terminology of elliptic curve cryptography, and ISO/IEC 15946-5, which describes how to construct elliptic curves and pairing functions, remain.

Kim (then at NTT) and Barbulescu in 2016 proposed an algorithm for solving the discrete logarithm problem on elliptic curves suitable for pairing functions [4]. This algorithm solves the discrete logarithm problem faster than other algorithms, which degrade the security levels of conventional elliptic curves. With the advent of this algorithm, the security evaluation and selection of new elliptic curves have been carried out among researchers and developers.

In WG 2, the security level of the elliptic curves in ISO/IEC 15946-5 has also declined, and in 2018, discussions began on revising the standards to improve the security levels of elliptic curves. The review, selection, and parameter setting of elliptic curves suitable for pairing are being discussed, and standardization is planned from 2021 to 2022. In response to the proposal of this analysis algorithm, NTT is also conducting research on a new elliptic curve [5] and actively contributing to discussions at WG 2.

5. Secret sharing (ISO/IEC 19592 series)

Secret sharing divides data to be kept secret into fragments (called 'shares') with proper encoding. Individual shares do not leak the original data, but if sufficient shares are gathered, then one can recover the original data even if some shares are lost (**Fig. 2**).

Since the original confidential information is not leaked from the shares, it can be used to prevent leakage of sensitive information. Usually, shares are kept by several people, and when a secret is needed, the shares are brought together for restoration of the secret. It is also a key technology for secure multiparty computation, which is described later. Because data can be recovered even if some shares are lost, it can also be used as a distributed data-storage technology or data-recovery technology in the event of data loss due to machine crash or disaster.

Since Shamir and Blakley's independent proposal of secret sharing in 1979, many schemes have been proposed. There are various differences such as safety, division method, and restoration method, and appropriate secrecy distribution must be selected in accordance with the usage scenario. If the same method is used, there may be differences depending on the implementation.

In 2014, the ISO/IEC 19592 series started the standardization of secret sharing. NTT has been active in the standardization of secrecy sharing in ISO/IEC, leading the development of the standard as an editor and contributing significantly to its publication in 2017. We contribute to the selection of easy-to-handle secret sharing by feeding back the knowledge obtained from NTT's research on secret sharing and secure multiparty computation and development of various products (secret sharing technology Trust-SS, distributed storage SHSS (Super High-speed Secret

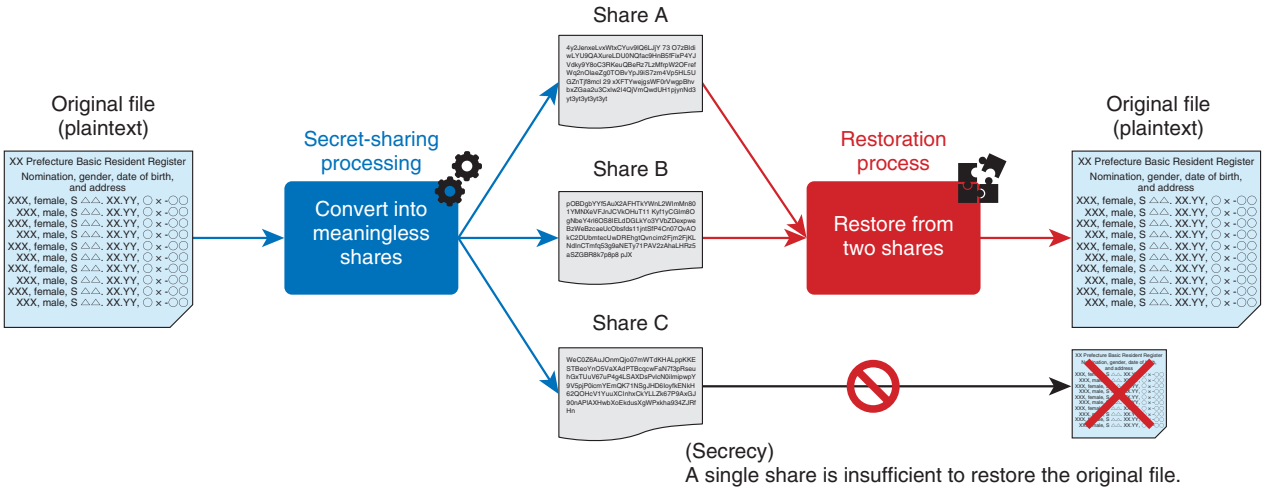


Fig. 2. Secret sharing.

	Range that can be protected	
Encryption	Communication and storage	—
Secure multiparty computation		Data processing

Fig. 3. Secret multiparty computation.

Sharing), secure multiparty computation technology "Sanshi®") using secret sharing.

NTT's products effectively use three secret-sharing schemes. ISO/IEC 19592-2, published in 2017, specifies five secret-sharing schemes, including these three [6]. In addition, standardization of secure multiparty computation based on secret sharing has recently started.

6. Secure multiparty computation (ISO/IEC 4922 series)

Secure multiparty computation is a technique to execute computations with encrypted data. General encryption protects data communication and storage. Secure multiparty computation can also protect the data-computation process. By using secure multiparty computation, analysis work using personal data of individuals and trade secrets of companies does not leak data and enables "not look inside" operation (**Fig. 3**). This will enable not only safer data processing but also new integrated analysis that transcends the boundaries of companies and industries by bringing together data that have been difficult to disclose to other organizations.

NTT is conducting R&D of secure multiparty computation using secret-sharing technology. That is, data are converted into shares by secret sharing then passed to the servers, which execute the computation without having to restore the original data from the

shares. In addition to NTT, various companies, universities, and research institutes are conducting and competing in R&D on secure multiparty computation.

In 2020, the ISO/IEC 4922 series started the standardization of secure multiparty computation. ISO/IEC 4922-1 will be the general standard for secure multiparty computation, and ISO/IEC 4922-2 will be the standard for secure multiparty computation based on secret sharing. NTT is actively leading the creation of standards as editors of both.

7. Future development

NTT will contribute to the development of international standards for cryptographic technology and protocols on the basis of our R&D expertise.

References

[1] CRYPTREC, "CRYPTREC Cryptographic Technology Guidelines (lightweight cipher)," Mar. 2017 (in Japanese). https://www.cryptrec.go.jp/report/cryptrec-gl-2003-2016jp.pdf

[2] The British Standards Institution, "ISO/IEC JTC 1/SC 27 N 20359, ISO/IEC NP 18033-7 Information technology - Security techniques - Encryption algorithms - Part 7: Tweakable block ciphers," https://standardsdevelopment.bsigroup.com/projects/9020-03695#/section

[3] W. Diffie and M. Hellman, "New Directions in Cryptography," IEEE Trans. Inf. Theory, Vol. 22, No. 6, pp. 644–654, Nov. 1976.

[4] T. Kim and R. Barbulescu, "Extended Tower Number Field Sieve: A New Complexity for the Medium Prime Case," Proc. of the 36th International Cryptology Conference (CRYPTO 2016), Part I, Vol. 9814, pp. 543–571, Santa Barbara, USA, Aug. 2016.

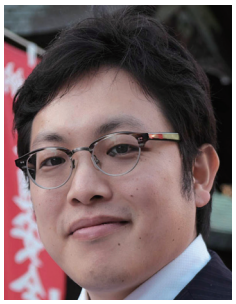
[5] Y. Kiyomura, A. Inoue, Y. Kawahara, M. Yasuda, T. Takagi, and T. Kobayashi, "Secure and Efficient Pairing at 256-bit Security Level," Proc. of the 15th International Conference on Applied Cryptography and Network Security (ACNS 2017), pp. 59–79, Kanazawa, Japan, July 2017.

[6] Press release issued by NTT on Nov. 23 (in Japanese). https://www.ntt.co.jp/news2017/1710/171023a.html

Keita Xagawa

Scientist, NTT Secure Platform Laboratories.

He received a B.S. from Kyoto University and an M.S. and D.S. from Tokyo Institute of Technology in 2005, 2007, and 2010. He joined NTT in 2010. He is presently engaged in research on cryptography and information security in NTT Secure Platform Laboratories. His research focus is on provable security and analysis in cryptography. He is a member of the Institute of Electronics, Information and Communication Engineers (IEICE) and International Association for Cryptologic Research (IACR). He has been an expert in ISO/IEC JTC 1/SC 27/WG 2 since 2018.



Ryo Kikuchi

Research Engineer, NTT Secure Platform Laboratories.

He received a B.E., M.E., and Ph.D. from Tokyo Institute of Technology in 2008, 2010, and 2015. Since 2010, he has been with NTT. His research focus is on cryptography and privacy-preserving data analysis. He has been an expert in ISO/IEC JTC 1/SC 27/WG 2 since 2013 and was a visiting researcher at the National Statistics Center from 2016 to 2019.

Atsunori Ichikawa

Research Engineer, NTT Secure Platform Laboratories.

He received a B.E. and M.E. from Tokyo Institute of Technology in 2015 and 2017. Since 2017, he has been with NTT. His research focus is on cryptography and oblivious data structures. He has been an expert in ISO/IEC JTC 1/SC 27/WG 2 since 2020.

Takayuki Miura

Research Engineer, NTT Secure Platform Laboratories.

He received a B.E. and M.E. from The University of Tokyo in 2017 and 2019. Since 2019, he has been with NTT. His research focus is on cryptography and machine learning security. He has been an expert in ISO/IEC JTC 1/SC 27/WG 2 since 2020.

External Awards

Research Encouragement Award
Winner: Shingo Omata, NTT Network Service Systems Laboratories
Date: January 22, 2021
Organization: The Institute of Electronics, Information and Communication Engineers (IEICE) Steering Committee on Network Software

For "A Study on Automatic Mapping of eTOM – OpenAPI Using COS Similarity Method."
Published as: S. Omata, "A Study on Automatic Mapping of eTOM – OpenAPI Using COS Similarity Method," 22nd Steering Committee on Network Software, NWS-22-3, June 2020.

Best Poster Award
Winners: Arinobu Niijima, Toki Takeda, Ryosuke Aoki, and Yukio Koike, NTT Service Evolution Laboratories
Date: February 25, 2021
Organization: The Augmented Humans International Conference 2021 (AHs 2021), Association for Computing Machinery (ACM)

For "Reducing Muscle Activity When Playing Tremolo by Using Electrical Muscle Stimulation."
Published as: A. Niijima, T. Takeda, R. Aoki, and Y. Koike, "Reducing Muscle Activity When Playing Tremolo by Using Electrical Muscle Stimulation," AHs 2021, Feb. 2021.

American Physical Society Outstanding Referee
Winner: Kenta Takata, NTT Basic Research Laboratories
Date: February 25, 2021
Organization: American Physical Society

For his outstanding refereeing to keep the standards of the journals of the American Physical Society at a high level and help authors improve the quality and readability of their articles.

American Physical Society Outstanding Referee
Winner: William John Munro, NTT Basic Research Laboratories
Date: February 25, 2021
Organization: American Physical Society

For his outstanding refereeing to keep the standards of the journals of the American Physical Society at a high level and help authors improve the quality and readability of their articles.

Best Technical Paper Award
Winners: Takashi Matsui, Yuto Sagae, Taiji Sakamoto, and Kazuhide Nakajima, NTT Access Network Service Systems Laboratories
Date: March 7, 2021
Organization: SPIE Photonics West 2021

For "Applicability of Standard 125 μm-cladding Multi-core Fiber for Wide-band and Long-haul Transmission."
Published as: T. Matsui, Y. Sagae, T. Sakamoto, and K. Nakajima, "Applicability of Standard 125 μm-cladding Multi-core Fiber for Wide-band and Long-haul Transmission," SPIE Photonics West 2021, 11713-2, Mar. 2021.

Young Researcher's Award
Winner: Takamitsu Tochino, NTT Access Network Service Systems Laboratories
Date: March 11, 2021
Organization: IEICE

For "Dynamic Bandwidth Allocation for Bandwidth Efficiency Improvement by Skipping Report Sequence."
Published as: T. Tochino, H. Ujikawa, Y. Sakai, and J. Terada, "Dynamic Bandwidth Allocation for Bandwidth Efficiency Improvement by Skipping Report Sequence," Proc. of the 2020 IEICE Society Conference, B-8-4, Online conference, Sept. 2020.

Young Researcher's Award
Winner: Keita Kuriyama, NTT Access Network Service Systems Laboratories
Date: March 11, 2021
Organization: IEICE

For "Experimental Study on Wide-band Single-carrier MU-MIMO System Using FIR-type Transmit Beamforming."
Published as: K. Kuriyama, H. Fukuzono, M. Yoshioka, and T. Hayashi, "Experimental Study on Wide-band Single-carrier MU-MIMO System Using FIR-type Transmit Beamforming," Proc. of the 2020 IEICE General Conference, B-5-164, Hiroshima, Japan, Mar. 2020.

Young Researcher's Award
Winner: Tatsuhiko Iwakuni, NTT Access Network Service Systems Laboratories
Date: March 11, 2021
Organization: IEICE

For "Interference Evaluation of Large-scale High-density Antenna Environments."
Published as: T. Iwakuni, D. Uchida, H. Kazui, S. Wai, C. Huan, N. Kita, and T. Onizawa, "Interference Evaluation of Large-scale High-density Antenna Environments," Proc. of the 2020 IEICE General Conference, B-5-88, Hiroshima, Japan, Mar. 2020.

Young Researcher's Award
Winner: Tomokazu Oda, NTT Access Network Service Systems Laboratories
Date: March 11, 2021
Organization: IEICE

For "Study on Measurement Accuracy of Splice Loss Measurement in FMF Based on BOTDA" and "Fundamental Study on Electric Field Distribution and BGS of LP11 Mode in FMF."
Published as: T. Oda, A. Nakamura, D. Iida, and H. Oshida, "Study on Measurement Accuracy of Splice Loss Measurement in FMF Based on BOTDA," Proc. of the 2020 IEICE General Conference, B-13-19, Hiroshima, Japan, Mar. 2020.
T. Oda, A. Nakamura, D. Iida, and H. Oshida, "Fundamental Study on Electric Field Distribution and BGS of LP11 Mode in FMF," Proc. of the 2020 IEICE Society Conference, B-13-15, Online conference, Sept. 2020.

Young Researcher's Award
Winner: Mizuto Nakamura, NTT Network Service Systems Laboratories
Date: March 11, 2021

Organization: IEICE

For "ACT Device Identification Method Using Time Fluctuation of Traffic Data" and "Time Correction Method of Time-series Data Using Waveform Similarity."
Published as: M. Nakamura, N. Hayashi, N. Tanji, A. Takada, T. Seki, and K. Yamagoe, "ACT Device Identification Method Using Time Fluctuation of Traffic Data," Proc. of the 2020 IEICE General Conference, B-14-6, Hiroshima, Japan, Mar. 2020.
M. Nakamura, N. Hayashi, A. Takada, T. Seki, and K. Yamagoe, "Time Correction Method of Time-series Data Using Waveform Similarity," Proc. of the 2020 IEICE Society Conference, B-14-5, Online conference, Sept. 2020.

Young Researcher's Award

Winner: Hiroki Ikeuchi, NTT Network Technology Laboratories
Date: March 11, 2021
Organization: IEICE

For "Root Cause Analysis Based on Massive Data Generated by Fault Injection."
Published as: H. Ikeuchi, G. E. Jiawen, Y. Matsuo, and K. Watanabe, "Root Cause Analysis Based on Massive Data Generated by Fault Injection," Proc. of the 2020 IEICE General Conference, B-7-32, Hiroshima, Japan, Mar. 2020.

MEF 3.0 Proof of Concept Showcase Innovation Award

Winners: Hiroki Baba, Shiku Hirai, Minoru Matsumoto, NTT Network Technology Laboratories; Mitsuo Amasaka, Kazuma Kamienoo, Takuya Satou, Ken Takahashi, Takayuki Nakamura, Takamitsu Narumi, Aki Fukuda, NTT Network Service Systems Laboratories
Date: March 11, 2021
Organization: MEF 3.0 Proof of Concept Showcase, MEF Forum

For "E2E Slicing for Extreme Services."

English Session Encouragement Award of Information and Communication Management

Winners: Shiku Hirai, Hiroki Baba, NTT Network Technology Laboratories; Saburo Seto, NTT Network Service Systems Laboratories
Date: March 18, 2021
Organization: IEICE Technical Committee on Information and Communication Management

For "Optimal Provisioning of Cloud-native Network Functions based on Performance Prediction."
Published as: S. Hirai, H. Baba, and S. Seto, "Optimal Provisioning of Cloud-native Network Functions based on Performance Prediction," Proc. of the 2020 IEICE Society Conference, BS-8-13, Online conference, Sept. 2020.

Papers Published in Technical Journals and Conference Proceedings

Digital to Natural - Innovation for Smart World

S. Yamamoto, A. Nakayama, and K. Kawazoe

International Journal of Informatics Society, Vol. 12, pp. 95–101, November 2020.

With the development of information and communication technologies, it is hoped that a world in which all people can live bountiful and happy lives can be achieved using innovative technologies. In other words, a Smart World. "Digital to Natural" is a transformation that is crucial to turning the concept of a Smart World into reality. It means not only pursuing the ultimate digital vision of high-speed, high-capacity, high-definition performance but also creating new value that can be achieved by naturally capturing and making the best use of a variety of information that previously could not be captured by humans. This will allow people to naturally and unconsciously benefit from technology. This paper describes what should be considered in order for technology to evolve into a more natural form and shows technologies that support it such as artificial intelligence, visual media, and information and communication technology infrastructures—IOWN (Innovative Optical and Wireless Network). This paper also presents concepts for several services that this technology can enable.

2D Position Estimation for Wireless LAN Terminals by the Access Point Using Distributed Antenna System

M. Hosoda, H. Sakamoto, T. Murakami, T. Mouri, A. Nakayama, T. Ogawa, and M. Miyamoto

IPSJ Journal, Vol. 62, No. 3, pp. 946–958, March 2021.

In this paper, we propose a two-dimensional (2D) positioning method for operators of MaaS (mobility as a service), events, facilities, etc. to obtain information on massive people flow by the positioning the terminals owned by the visitors. We have previously proposed a 1D positioning method for positioning widely used wireless local area network (LAN) terminals without needing to install any application in the terminals but by just connecting the terminal with an access point that uses a distributed antenna system. We extended this method to 2D positioning and introduce a method of calculating position estimation by integrating more antennas and multiple measurements. We conducted experiments to confirm the effectiveness of the proposed method and present the results in this paper. We also demonstrated the proposed method at an exhibition where many visitors gathered.

Riccati Equation as Topology-based Model of Computer Worms and Discrete SIR Model with Constant Infectious Period

D. Satoh and M. Uchida

Physica A: Statistical Mechanics and its Applications, Vol. 566, 125606, March 2021.

We propose discrete and continuous infection models of computer worms via e-mail or social networking site messengers that were previously classified as worms spreading through topological neighbors. The discrete model is made on the basis of a new classification of worms as *permanently* or *temporarily* infectious. A temporary infection means that only the most recently infected nodes are infectious according to a difference equation. The discrete model is reduced to a Riccati differential equation (the continuous model) at the limit of a zero difference interval for the difference equation. The discrete and continuous models well describe actual data and are superior to a linear model in terms of the Akaike information criterion. Both models overcome the overestimation that is generated by applying a scan-based model to topology-based infection, especially in the early stages. The discrete model gives a condition in which all nodes are infected because the vulnerable nodes of the Ricatti difference equation are finite and the solution of the Riccati difference equation plots discrete values on the exact solution of the Riccati differential equation. Also, the discrete model can also be understood as a model for the spread of infections of an epidemic virus with a constant infectious period and is described with a discrete susceptible-infected-recovered (SIR) model. The discrete SIR model has an exact solution. A control to reduce the infection is considered through the discrete SIR model.

Hardness of Efficiently Generating Ground States in Post-selected Quantum Computation

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

Physical Review Research, Vol. 3, 013213, March 2021.

Generating ground states of any local Hamiltonian seems to be impossible in quantum polynomial time. In this paper, we give evidence for the impossibility by applying an argument used in the quantum-computational-supremacy approach. More precisely, we show that if ground states of any 3-local Hamiltonian can be approximately generated in quantum polynomial time with postselection, then PP = PSPACE. Our result is superior to the existing findings in the sense that we reduce the impossibility to an unlikely relation between classical complexity classes. We also discuss what makes efficiently generating the ground states hard for postselected quantum computation.