ISSN 1349-3671

2009 April No. 31

Science & Technology Trends

Quarterly Review

Science & Technology Foresight Center, NISTEP

Information and Communication Technologies

Nanotechnology and Materials

Frontier

Social Infrastructure

Science and Technology Trends —Quarterly Review

No. 31 April 2009

Science & Technology Foresight Center of NISTEP

Foreword

This is the latest issue of "Science and Technology Trends __ Quarterly Review".

National Institute of Science and Technology Policy (NISTEP) established Science and Technology Foresight Center (STFC) in January 2001 to deepen analysis with inputting state-of-the-art science and technology trends. The mission of the center is to support national science and technology policy by providing policy makers with timely and comprehensive knowledge of important science and technology in Japan and in the world.

STFC has conducted regular surveys with support of around 2000 experts in the industrial, academic and public sectors who provide us with their information and opinions through STFC's expert network system. STFC has been publishing "Science and Technology Trends" (Japanese version) every month since April 2001. The first part of this monthly report introduces the latest topics in life science, ICT, environment, nanotechnology, materials science etc. that are collected through the expert network. The second part carries insight analysis by STFC researchers, which covers not only technological trends in specific areas but also other issues including government R&D budget and foreign countries' S&T policy. STFC also conducts foresight surveys periodically.

This quarterly review is the English version of insight analysis derived from recent three issues of "Science and Technology Trends" written in Japanese, and will be published every three month in principle. You can also see them on the NISTEP website.

We hope this could be useful to you and appreciate your comments and advices.

Dr. Kumi OKUWADA
Director, Science and Technology Foresight Center
National Institute of Science and Technology Policy

NISTEP has moved to a new office

Contact information

Science and Technology Foresight Center
National Institute of Science and Technology Policy
Ministry of Education, Culture Sports, Science and Technology (MEXT)

3-2-2, Kasumigaseki, Chiyoda-ku, Tokyo 100-0013, Japan
Telephone +81-3-3581-0605 Facsimile +81-3-3503-3996
URL http://www.nistep.go.jp/index-e.html
E-mail stfc@nistep.go.jp

Executive Summary

Information and Communication Technologies

1 Trends in the Commercialization and R&D of New Information Network Infrastructure

p.8

The telephone was invented in the late 19th century. Telephone networks, created based on a technology called circuit switching, have made the telephone a vital piece of social infrastructure. Meanwhile, television and other broadcast networks arose following the invention of radio communication around 100 years ago. These networks have continued in existence to the present dayusing video-transmission technologies as their cores. While telephone networks use circuit switching, a method called packet switching was invented in the early 1960s, and gave birth to the Internet. Although packet switching was originally developed for the communication of text and other types of data, as the Internet has developed it has become possible to offer a wide range of communication via this method. In particular, through packet switching, it has become possible to implement voice communication, traditionally the realm of the telephone network, and video transmission, previously the bailiwick of television broadcasting. This has caused a massive change in market demand for electronic communications, away from telephone calls and toward data; the telephone network, originally created as social infrastructure for the telephone, is being transformed at a fundamental level.

Telecommunications carriers and telecommunications policymakers in industrialized countries are highly interested in next-generation networks (NGNs) as information and communications infrastructure to replace the declining voice-communication market. An NGN uses a technology called the Internet protocol (IP) to operate communications networks as an integrated network. IP is the core technology used in Internet packet switching. Industrialized countries are operating a large number of NGNs on experimental bases. In March 2008, Japan became the first country to begin operating an NGN as a commercial service.

Meanwhile, research is advancing into new types of networks. Called new generation networks (NWGNs) to differentiate them from NGNs, research into NWGNs is investigating new types of social infrastructure that are not constrained by the developmental path that broadcast and communications networks have followed. NWGNs take a longer-term view of network infrastructure. Most research focuses on extensions of IP, additionally investigating ways to support demand for low-volume, high-frequency communication in an assumed environment of ubiquitous networking.

This paper starts by describing a global trend toward standardization, and goes on to describe various initiatives relating to NGNs around the world. It then provides an overview of projects researching NWGN architectures being advanced with a long-term perspective in the United States, Europe, and Japan.

Rather than relying exclusively on the knowledge of engineering experts, future discussion of information networks will require the knowledge and views of a

wide range of experts. It is necessary to study the impact of networks on research and development in other fields, and devise ways to create fertile ground for innovation in a wide range of fields.

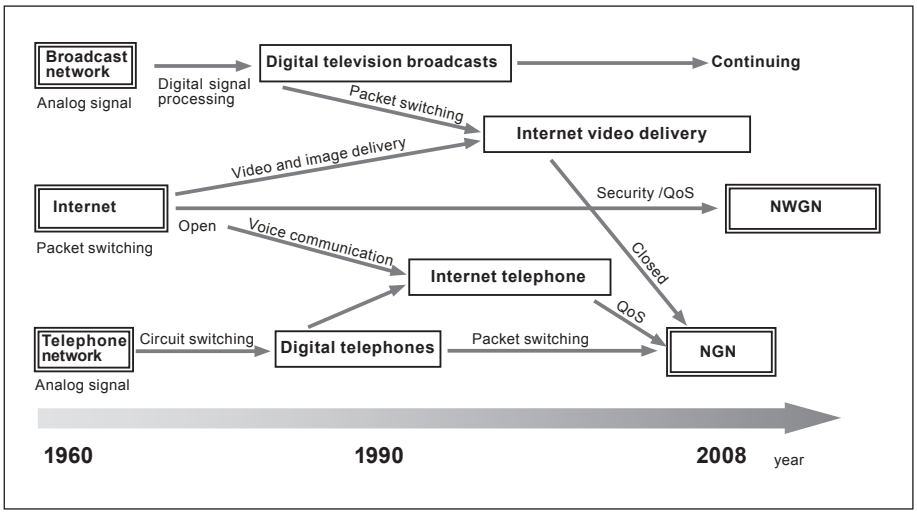


Figure : Technical Advances and the Relationship between NGNs and NWGNs

Prepared by the STFC

(Original Japanese version: published in November 2008)

Information and Communication Technologies

2 Developments in Television Band Frequency Sharing Technology

p.24

The digitalization of television broadcasting, scheduled for completion in July 2011, is underway. The channel plan for digital television broadcasting is specified in the ordinances and notices of the Ministry of Internal Affairs and Communications, which include the criteria for interference protection ratios. If there are unused channels meeting the criteria, the frequencies assigned to those channels may be shared with other uses. The shared used of television bands will soon be realized.

Two experimental projects for sharing unused channels are being conducted. The Yokosuka Research Park R&D Promotion Committee operates a new narrow-area digital community broadcasting which transmits community programs to a limited narrow area within a radius of about 300 meters through One Segment Broadcasting with low antenna power. This allows people in downtown areas to watch information programs on department stores and shopping districts on a cell phone. The Shimane Ubiquitous Project, on the other hand, covers part of Matsue City using MediaFLO wireless technology. Integrating multimedia broadcasting and IC card technology, the project enables IC card holders to use points and coupons received while watching service information transmitted to stores. The circulation patterns of the card holders can be observed from their card usage histories.

Future research and development should focus on advanced frequency sharing technology to increase the shared use of unused channels. Unlike the conventional development concept of broadcasting, this new broadcasting service oriented toward “responding to the needs of the smallest possible number of people in the narrowest possible area” aims at a kind of community television. It is a step

closer to the integration of communications and broadcasting. Digitalization is also proceeding in other countries, therefore, if the "webification of broadcasting" spreads worldwide, new business opportunities will be created for Japanese industry.

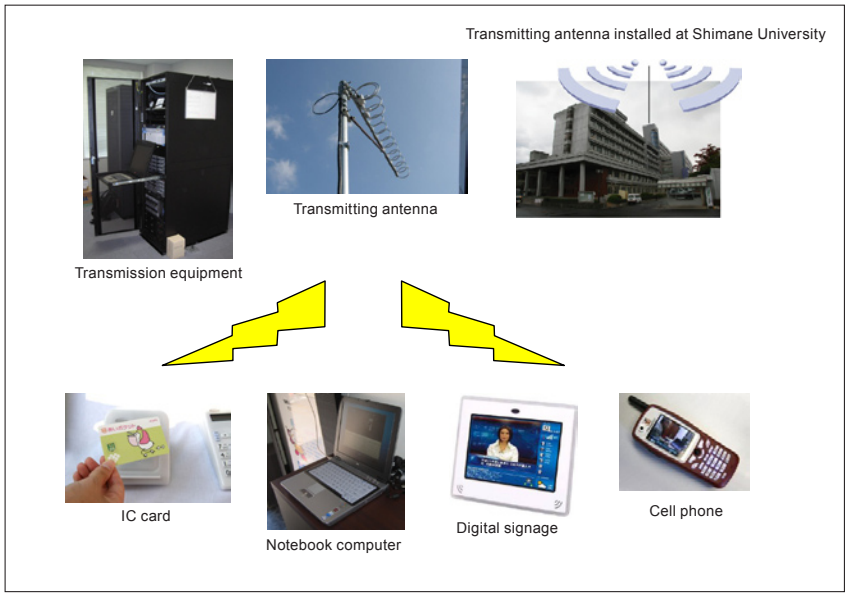


Figure : Types of information services provided by the Shimane Ubiquitous Project

Source: Shimane Ubiquitous Project

(Original Japanese version: published in December 2008)

Information and Communication Technologies

3 Overcoming the Language Barrier with Speech Translation Technology

p.36

Communication across the barrier of multiple national languages is currently a major problem that can sometimes be even more serious. Establishing automatic speech translation technology, which could instantly translate spoken words into the listener's language, would greatly contribute to Japan's globalization through economic activities, multilingual tourism business, and improved service to foreign residents.

Speech translation is a technology integrating speech recognition, text translation, and speech-synthesis technologies. Over the past few years, the accuracy of speech translation has improved dramatically thanks to advances in each of these component technologies as well as expanding databases. Since the beginning of this century, rapid advances have been made in technical research based on "corpora (corpuses)" of speech and text databases supplemented with various types of linguistic information. The technology has now progressed to the stage where sentence-by-sentence consecutive translation of ordinary spoken travel-related conversation is feasible between Japanese, English, and Chinese. There are still many research challenges to overcome, however, particularly with multilingual speech translation: speech translation is strongly affected by speaker-variability: dependent; there is great diversity of expression and new words and concepts are constantly being created in accordance with changes in society.

Although the research and development of basic technologies is vital for speech translation technology, the collection of real-world corpora and machine learning are also essential. There are some aspects of this technology whose performance can only truly improve through use. Moving forward, it will be vital to continue to introduce this technology wherever possible, while conducting repeated field experiments. It is also necessary to advance a scheme for linking multiple national languages, in order to break the linguistic barriers between many different countries. Speech translation in Japan is advanced, and the technology in this country can lead standardization and the like in other countries.

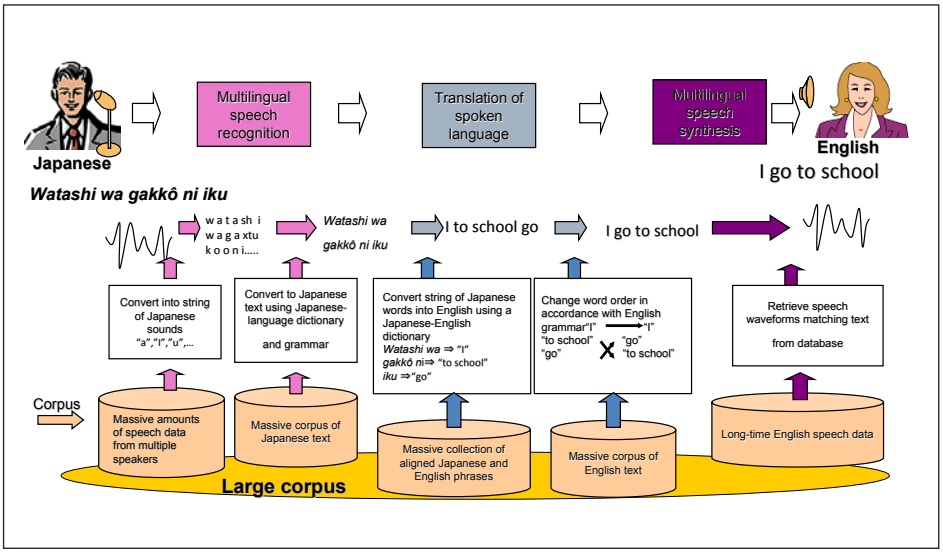


Figure : The mechanism of the speech translation system

Prepared by the STFC

(Original Japanese version: published in Augaust 2008)

Nanotechnology and Materials

4 Necessity of True Bulk GaN Single Crystal and Trends in Research and Development

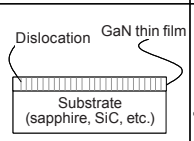
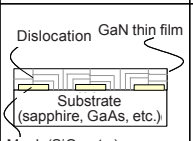
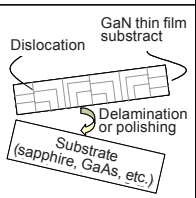
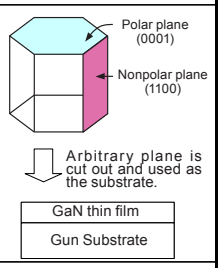
p.50

Gallium nitride (GaN) semiconductors are currently used in practical applications such as light-emitting diodes (LED), laser diodes, etc., and thus are closely related to our daily lives. LEDs are used as backlight devices for liquid crystal displays, beginning with cells phones, while laser diodes are used in recording and reproduction of high definition images as a light source for Blu-ray Discs. As higher performance is achieved in GaN devices in the future, applications will expand to general lighting and transistor applications such as inverters. Trial calculations have shown that these devices will also make an important contribution to energy saving.

GaN semiconductor crystals are obtained by growing crystals on a substrate of a different material such as sapphire or silicon carbide (SiC). However, because the material used in crystal growth and the substrate material are different, numerous crystal defects occur, with an adverse effect on device performance, including output, life, etc. Furthermore, spontaneous polarization occurs due to the relationship between the substrate of this heterogeneous material and the orientation of the crystal planes in GaN grown in this process, and this has become an impediment to improvement of light emission efficiency.

When attempting to achieve substantially higher performance in the future, the existing crystal growth technology using a heterogeneous substrate will reach its limits. To overcome these limits, a "true bulk GaN single crystal" is strongly desired. Such a material would contain minimal crystallographic defects and could be cut to arbitrary crystal planes. However, the necessary crystal growth technology has not been established. Whether Japan can seize the initiative in research and development of crystal growth technology for GaN, which can be considered one of the few remaining substrate materials in the competition for leadership, will have an important meaning for the future of GaN semiconductor device technology as a whole.

Table : Epitaxial crystal growth of GaN and its applications

	Heteroepitaxial			Homoepitaxial
	Type A	**Type B**	**Type C**	**Type D**
				
Merits/ Demerits	· Large area : To inch class · Crystal defects (dislocations) : To $10^9 cm^{-2}$ · Polarity:Spontaneous polarization by C(0001) plane	· Large area : To inch class · Crystaldefects (dislocations) : To $10^9 cm^{-2}$ · Polarity : Spontaneous polarization by C(0001) plane	·Large area: To inch class ·Crystal defects (dislocations): To $10^9 cm^{-2}$ ·Polarity:Spontaneous polarization by C(0001) plane. ·High cost	· High quality · Arbitrary polar plane · Crystal size: mm order
Applica-tions	Near UV to green LED White LED (pseudowhite)	Blue-violet LD (BV-LD; low power)	Blue-violet to blue LD	Not realized

Prepared by the STFC

(Original Japanese version: published in November 2008)

Frontier

5 Trends and Problems in Earthquake Prediction Research

p.65

Effective advance information was not available for the series of inland earthquakes which struck Japan in recent years or for the massive Sichuan Earthquake in China in 2008. For this reason, the expression "earthquakes can occur anywhere and anytime" has taken root in the mass media. Underlying this, however, a mistrust of the current earthquake prediction research can be seen.

In Japan, the earthquake prediction research system based on cooperation between governmental agencies and universities began with "Earthquake Prediction – Current Status and Plans for Its Promotion," which was compiled in 1962, as a basic guideline. In the 1990s, while involved in a worldwide debate on the possibility of earthquake prediction, the Great Hanshin-Awaji Earthquake (1995) suddenly struck this country, calling into question the approach to research for earthquake prediction up to that time. Occasioned by this event, the government began efforts directed toward "earthquake forecasting," and the earthquake prediction researchers' group changed directions toward a physical elucidation of the process by which earthquakes occur. Although the 10 years since that time have seen the completion of "National Seismic Hazard Maps for

Japan," the actual occurrence of earthquakes does not necessarily follow the Seismic Hazard Maps. This is attributed to the fact that "smaller active faults," including those in coastal areas, were not fully identified. To remedy this problem, a plan for a new project to identify these small-scale active faults was established.

On the other hand, following the change toward a physical approach, a series of new discoveries have been made in earthquake prediction research in the last 10 years, resulting in a dramatically deeper understanding of the process in earthquake occurrence. It can be said that "earthquake prediction," in the sense of adding additional information to "earthquake forecasting" which is based on statistical analysis, already exists. However, because its social significance is inadequate, a large gap remains between "earthquake prediction" as understood by researchers and the "earthquake prediction" demanded by society.

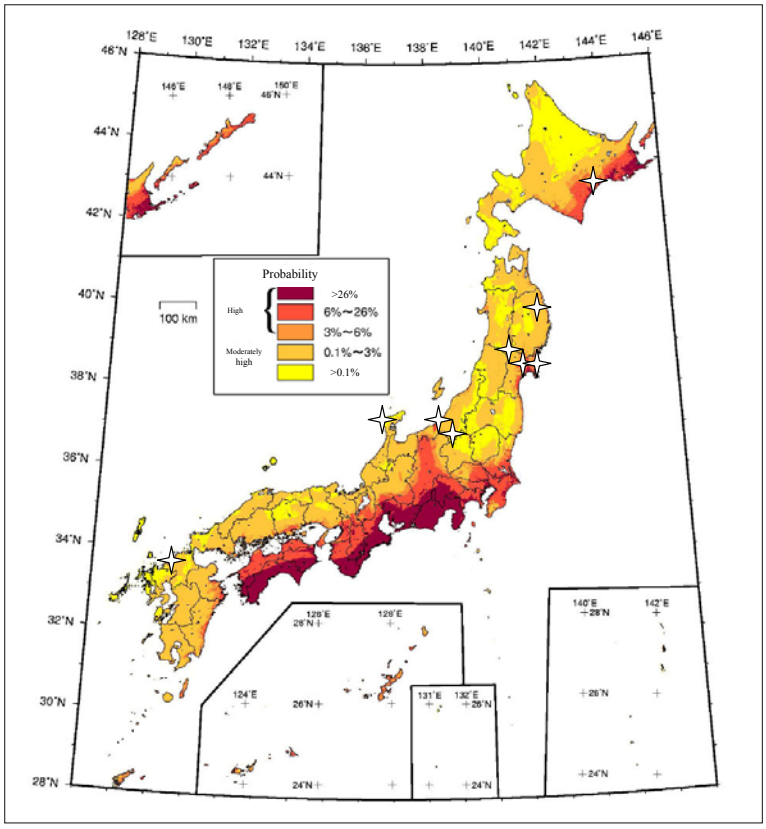


Figure : Seismic Hazard Map

Source : Reference [15]

(Original Japanese version: published in October 2008)

Social Infrastructure

6 The Role of Operations Research towards Advanced Logistics

p.85

Product supply chains of recent Japanese industries are now at a major turning point as they attempt to cope with structural changes due to the globalization of industrial operations and changes in the relationships between related firms, and more advanced functional requirements, such as adaptability to volatile markets, energy conservation, response to environmental issues. Naturally, supply chain logistics of the products and services is also under pressure to reform and adopt more advanced operation technologies suited to these circumstances. In the planning and operation of complex and widely spread activities in the recent

logistics, adaptation of mathematical analysis and optimization techniques by using operations research (OR) methodology is strongly required. Moreover, together with promoting application of OR, the development of more advanced OR techniques is also required.

Not only through the conventional improvements in the hardware aspect, such as transportation facilities and equipment, applying OR to logistics network design and optimization problems makes it possible to realize more efficient logistics operation through optimized warehouse location and inventory allocation in the large-scale distribution and transportation networks. It would also helps prevention of global warming by efficient transportation. However, in Japan, application of OR to logistics is not as popular as in Europe and the United States, and practical research in the universities is not active.

In order to respond to the issues in today's logistics, and to solve the problems of the future, research and development of advanced OR techniques together with research for problem solving should be essential. As a recent challenge, logistics would be an important field of the research in service science that is a hot issue in Japan to increase productivity and improve quality in the service industry, which accounts for a major part of the industrial structure. For enhancement of the research in logistics, both prioritizing science and engineering, and promotion of interdisciplinary research with fields of social science such as economics are required. It is also necessary to promote joint interagency and interregional research across the different institutions in various government offices.

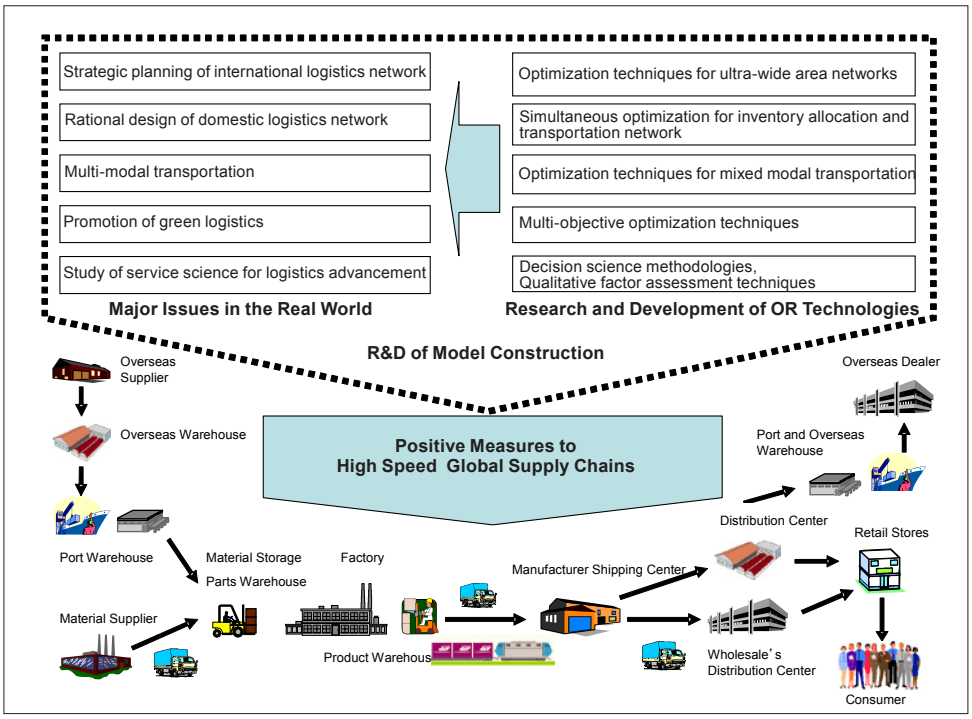


Figure : Real problems requiring solution R&D on model construction methods

Prepared by the STFC

(Original Japanese version: published in October 2008)

1

Trends in the Commercialization and R&D of New Information Network Infrastructure

Akihiro Fujii
Hajime Yamada
Affiliated Fellow

1 Introduction

1-1 The trend toward integration of communications infrastructure into the internet protocol

The telephone was invented in the late 19th century. Telephone networks, created based on a technology called circuit switching, have made the telephone a vital piece of social infrastructure. Meanwhile, television and other broadcast networks arose following the invention of radio communication some 100 years ago. These networks have continued in existence to the present day, forming the core of video-transmission technologies. While telephone networks use circuit switching, a method called packet switching was invented in the early 1960s, and gave birth to the Internet. This technology was originally developed for the transmission of text and other data. Using this method, information is split into small units called "packets," each of which is sent with the destination address. This allows the communication network to be utilized flexibly and efficiently.

Advances in Internet-related technologies have made it possible to offer a wide range of services using packet switching, including voice and high-definition video. Until now, however, different functionality has been provided using completely different infrastructure: voice calls using the telephone network; data communication using the Internet; and video transmission via television broadcasting. Additionally, the telephone network itself includes wired fixed phones and wireless mobile phones. In other words, information networks are part of our social infrastructure, but they are currently not integrated into a single social infrastructure. These services are instead a hotchpotch, each developing according to its own design goals and technology system.

Some argue that we should utilize the packet-switching method used by the Internet to uniformly operate these various communications infrastructures, each with its own history and provenance. The four types of communication above can be summarized as follows.

(1) Data communication, such as email and the transmission of information via the Web (provided via the Internet using packet switching)
(2) Calls using the fixed-phone network (The majority of these communications are currently operated based on circuit-switching technologies. Some are implemented using packet switching, as "Internet phone.")
(3) Distribution of video and audio via broadcast networks (This is generally provided using broadcast-network infrastructure; some new services, such as video distribution, are provided over the Internet.)
(4) Mobile communications using the mobile-phone network (Voice calls use telephone-network technology. Email, Web browsing, and other services use packet-switching technology.)

The Next Generation Network (NGN) is attempting to provide these different types of communications on a uniform communications infrastructure, using packet switching.

The social infrastructure originally built for the telephone is currently undergoing a major

transformation. The core of the packet-switching system used on the Internet is a method called the Internet Protocol (IP). IP is one of the fundamental underlying technologies of Internet packet communication. It prescribes things like packet structure and the use of network addresses on the Internet. In other words, IP is at the heart of the NGN's uniform network operation.

Integrating existing communication infrastructure, such as fixed-phone and mobile-communications networks, to use IP (i.e., packet switching) will make it possible to offer a wide range of communications-service functionality, including video distribution, in a uniform way. This will expand the possibilities of creating completely new services. It is also argued that integrating communications networks to IP will reduce the cost of maintaining and managing the equipment used to provide voice-call services.

Telecommunications carriers and telecommunications policymakers in industrialized countries are highly interested in next-generation networks (NGNs) as information and communications infrastructure operated uniformly using IP, in order to replace the declining voice-communication market. Industrialized countries are operating a large number of NGNs on experimental bases. In March 2008, Japan became the first country to begin operating an NGN as a commercial service. Some telecommunications carriers also use the term "NGN" in the marketing materials for their own new services, giving the impression that "NGN" refers to a particular service package. The conversion of communication networks to IP, however, is a trend that no carrier will be able to avoid–even the ones that do not offer physical transmission lines. A wide range of responses to this trend have begun [9,10].

1-2 The start of NGN services

In March 2008, Japan became the first country in the world to begin offering commercial NGN services. The trend leaders in related businesses are large-scale telecom carriers such as Nippon Telegraph and Telephone Corporation (NTT) in Japan and British Telecom (BT) in the United Kingdom.

Table 1 shows some examples of NGN services that NTT has begun. These services all require the company or home to have optical fiber installed. The benefit of these commercial services to the user is that they provide high bandwidth, enabling the transmission of high-definition video data. Quality guarantees are also vital for communication of video data. This depends on whether a function called "bandwidth guarantee"is offered; this function guarantees that a certain volume of packets will be available within a specific period of time.

The fundamental structure of the Internet is designed based what is called a "best effort"network. The Internet has thus traditionally not been very good at guaranteeing communications-bandwidth availability. As an example, when video data like a television broadcast is distributed via the Internet, if the bandwidth of this data is not guaranteed then it may not be possible to transmit a fixed volume of data continuously. If this happens, it causes the quality of the video to degrade, resulting in a jumpy/scrambled picture or dropped frames. This

Table1 : Sample NGN services (NTT)

Service type		Contents of services
	Quality	
Optical broadband service	Bandwidth guarantee and best effort	To detached houses (100 Mbps) To apartments (100 Mbps) To businesses (1 Gbps)
VoIP	Bandwidth guarantee	Hikari (optical-fiber) telephone Videophone
VPN (virtual private networks)	Bandwidth guarantee	(Scheduled to be offered in future)
	Best effort	VPN
Content-distribution services	Bandwidth guarantee	Unicast Multicast
	Best effort	Unicast Multicast
Ethernet service	Best effort	Ethernet

Prepared by the STFC based on reference [9]

is what people experience when they access video over a poor-quality Internet connection.

The implementation of bandwidth guarantees is a point requiring attention in relation to Japan's move to digital broadcasts in 2011. Although digital streaming (distribution) of broadcast content has already been achieved, if bandwidth guarantees can be achieved, then it could become possible for the Internet infrastructure to replace the broadcast network. The broadcasters of television programs, however, are cautious about supporting the move to the Internet, because it could cause them to lose their existing rights. In fact, it is possible that for the near future, the social infrastructure of the communication network will continue to develop while serving to supplement the role played by the broadcast network to date. Telecom carriers, broadcasters, and content producers in Japan and other countries are currently competing for business dominance of these emerging application fields.

1-3 Efforts to further develop communication infrastructure

There has recently been a surge in research and development of new networks taking a longer-term perspective. In other words, a debate is beginning over the very core concepts of communication networks as social infrastructure, as we look toward a future beyond the NGNs that are available commercially today. These new networks are called New Generation Networks (NWGN), in order to distinguish them from NGNs. Although the NWGNs being considered essentially have IP at their cores, they are a completely new concept in communication infrastructure that is not necessarily confined by IP.

The study of NWGNs is one of the more important challenges for the research and development of information and communications technologies in each country. The Global Environment for Network Innovations (GENI) in the United States and Euro-NGI in Europe are two examples; these will be described below. An example in Japan is AKARI, a project to design next-generation architecture being implemented by Japan's Ministry of Internal Affairs and Communications (MIC) and National Institute of Information and Communications Technology (NiCT).

As alluded to above, one challenge for the research and development of new information networks like NGNs and NWGNs is building infrastructure that can provide uniform support for a communication networks that have developed along different trajectories, and the wide range of services offered over these networks. Figure 1 provides a conceptual overview of the process of

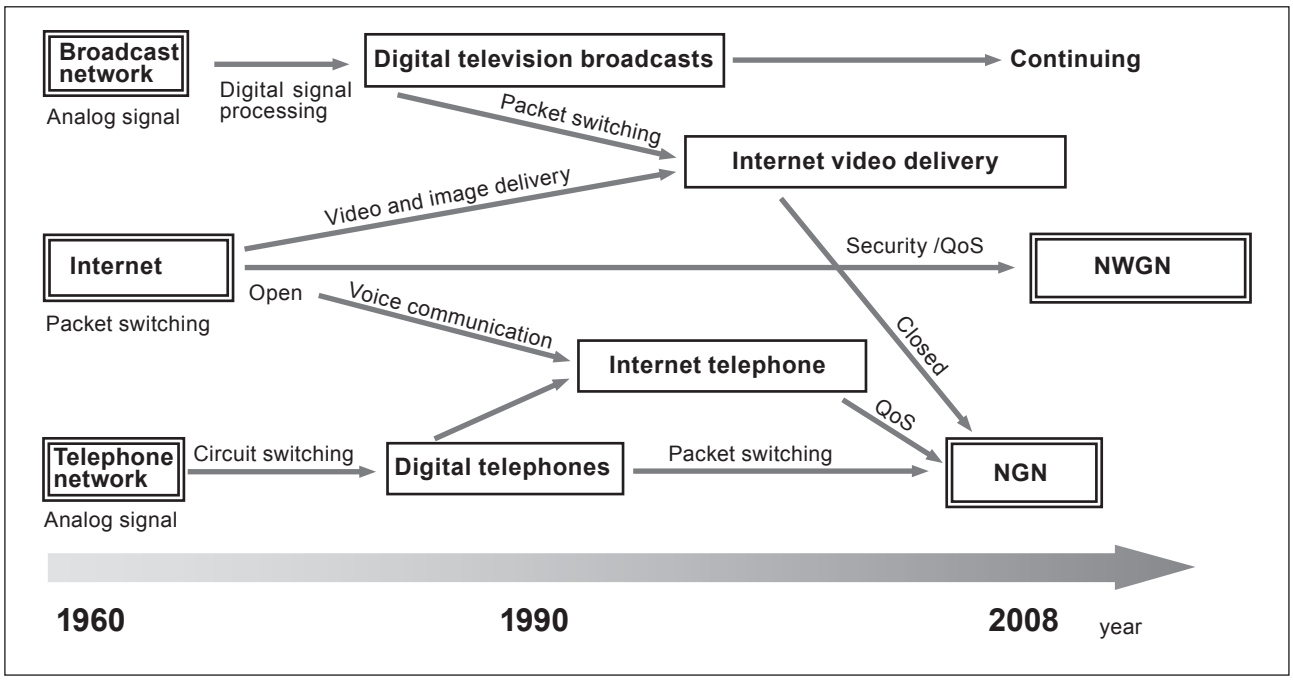


Figure 1 : Technical Advances and the Relationship between NGNs and NWGNs

Prepared by the STFC

merging these services.

Figure 2 provides a conceptual view of the relationship from the broadcast networks, Internet, and telephone networks to NGNs and NWGNs. There are two main elements connecting the broadcast networks, telephone networks, and the Internet: VoIP (voice over IP; also called "Internet telephone") and video distribution over the Internet. These services bring to the Internet functionality conventionally provided by the telephone network and broadcast network, respectively. It is expected that the functionality currently being introduced under the label "NGN services" will be advanced by existing telecommunications carriers. In other words, the trio of telephone, data communication, and broadcasting will progress within the NGN framework through the introduction of quality of service (QoS) guarantees and enhanced security functionality into the IP world.

Research into future NWGNs is likely to progress in parallel to the commercialization of NGNs described above. NWGNs take a longer-term view of network infrastructure. Most research focuses on advances to IP, additionally investigating ways to support low-volume, high-frequency communication on the assumption that ubiquitous networking will have been achieved.

The Internet that uses packet switching has developed as an open network, where anyone can join the communication network. Meanwhile, telephone networks have developed in a closed environment, as strategic national enterprises. NGNs could be called closed services, offered by specific telecommunications carriers. In contrast, NWGNs are attempting to utilize the open environment that the Internet has always had, and which has served as an engine of its growth.

The following and subsequent chapters describe technology trends that are transforming the nature of communication infrastructure, with IP technology at the core. Chapter 2 outlines the commercialization of NGNs being advanced around the world with telecoms-vendor participation (with a focus on ITU-T). Chapter 3 describes research and development of NWGNs, being advanced with a longer-term perspective.

2 Trends in commercial NGNs

2-1 Changing communication needs and the response by telecom carriers

The market needs for communications in industrialized nations began to change greatly in the late 1990s, from voice to data. According to the MIC's Information and Communications Statistics Database, both the number and time of calls are falling significantly for total fixed, mobile, and VoIP telephone calls. In 2000, there were a total of 144.8 billion total calls, with a total calling time of 7.03 billion hours. In 2005, there were 121.1 billion calls, with a total of 4.36 billion hours. These are reductions of 16% and 38%, respectively. As a result, telephone service, once assumed to be a cash cow, is becoming a millstone around the necks of telecom carriers.

According to this database, the total broadband traffic (estimated from the number of subscribers) went from 269 Gigabits/second in September 2004 to 880 Gigabits/second in May 2008-a three-fold increase. Services delivering text, images, music, video, and other content via broadband Internet have also grown at an extremely high pace. A technology called voice over Internet protocol (VoIP) has also emerged. This technology has made it possible to send and receive voice telephone calls over the Internet, sending audio signals using IP. This has made the inevitability of providing telephone and Internet services separately less apparent. Functionality to guarantee bandwidth is being created in the NGN frameworks; if this functionality is added to VoIP, it will become possible to control the audio quality and delays of the conversation according to the call rates.

Telecom carriers in Japan have also started to change their business strategies, in response to new market demands relating to communications. They have reduced spending on networks for telephone services to the bare minimum required for maintenance and management, and are refraining from investing in equipment updates. As a reflection of this, the size of the switch market is shrinking at a tremendous pace. According to a study by the Communications and Information Network Association of Japan (CIAJ), the total

value of the Japanese switch market was 142.7 billion yen in 2007, a 29% decrease from the level of 200.9 billion yen two years earlier, in 2005. Despite this decline, however, there is still huge demand for voice telephone calls. This gave rise to the idea that calls should be conducted over the Internet using VoIP, integrating the telephone and Internet networks. The result was the NGN.

2-2 Source of the NGN

In the late 1990s, many telecom carriers in the industrialized world realized the potential of NGNs, and began researching them. The group that advanced the standardization of third-generation mobile phones were the ones that began moving actively to actually utilize NGNs. They first considered the possibility of introducing IP in mobile communications. In particular, they studied such methods as managing addresses according to use movement. These studies came together in a new standard for mobile communications. The international standard for third-generation mobile phones became available right in the very early 2000s. But the worldwide adoption of third-generation technology has been slow, and second-generation mobile phones continue to gain greater penetration worldwide. According to figures in a flash report released by the GSM Association on 22 September 2008, there are 3,804,060,000 mobile-phone users worldwide, of whom 3,059,130,000 – 80% of the total – use the older second-generation GSM (Global System for Mobile). On top of this, the number of these users increased by 800 million over a year and a half.

The numbers of users are growing especially in Asia, South America, Africa, and Eastern Europe. There are few fixed phones in these regions, and when installing means of communication there for the first time, it is cheaper than laying phone lines. This is why GSMs are chosen, and why they are growing so rapidly.

A second-generation phone costs only a third as much as a third-generation phone. This gap will need to be closed in order to for third-generation mobile communication to be popularized. One idea that has been proposed as a trump card for achieving this is saving on costs by building the core network using Internet technology. The idea is to use Internet communication devices for mobile voice communication, because they can be supplied relatively cheaply. This is called an "all-IP network." The international standard IP Multimedia System (IMS) was created from a background of active initiatives of this sort. This could be called the source of the NGN.

IMS adopted the Session Initiation Protocol (SIP), which is a protocol for controlling multimedia sessions, including voice, using IP. When viewing a Website, the user sends a request consisting of a URL (Uniform Resource Locator) that is about 30 characters long, and large amounts of text, images, and other data are sent back in response. In contrast, in a telephone call, the caller and callee are in an equal relation in terms of the amount of data sent. SIP was developed in order to facilitate the sending and receiving of data between two parties in this type of equal relation. This technology is being used to convert the mobile-phone communication network to use IP as well.

2-3 Creation of an international standard for NGNs

The international standardization body ITU-T plays a leading role in communications protocols and other efforts to achieve NGNs. Telecom carriers around the world are members of the ITU-T, and it takes commercial perspectives into account for offering these services. Table 2 shows the most important and distinctive features demanded of standards for NGNs.

The European Telecommunications Standards Institute (ETSI) is one of the leaders in the standardization of third-generation mobile phones. The ETSI has begun working to extend the all-IP network from mobile to fixed communications. It advocates a communication protocol called Telecommunications and Internet converged Services and Protocols for Advanced Networking (TISPAN) for achieving this.

The ETSI has started to get results from its standardization efforts relating to TISPAN. It has brought these results to the International Telecommunication Union Telecommunication Standardization Sector (ITU-T), and begun international standardization as an NGN. At the general assembly of the ITU-T held in October 2004, the standardization of NGNs was identified as the number-one challenge, and this has

Table 2 : Main features of NGNs

	Feature
1	All-packet network using IP
2	Support for broadband capabilities and wide range of last-mile technologies
3	Separation of transport (network) and services
4	Full separation of control functionality and services
5	Provides a wide range of multimedia services, including voice as well as video and data, in a modular fashion
6	The framework enables a service to provide the same experience no matter how the user is connected to the network
7	Guarantees end-to-end quality of service (QoS), in accordance with the network quality and the user's device
8	The framework identifies users by assigning them IP addresses and IP-network routing
9	Ensures interoperability with existing networks
10	It is possible to access a wide range of service providers, freely and without restriction
11	Provides a high level of mobility, including ubiquitous access
12	Provides seamless communication over fixed and mobile networks
13	Complies with restrictions, including emergency response, and ensuring privacy and security

Prepared by the STFC based on ITU-T recommendations and other documents

subsequently sparked international discussion.

The first ITU-T recommendation was titled "General Overview of NGN," commonly known as "Y.2001." This recommendation, which was published in December 2004, defines the framework and architecture model for NGNs. The definition in the recommendation is as follows:
"Next Generation Network (NGN): A packet-based network able to provide telecommunication services and able to make use of multiple broadband, QoS-enabled transport technologies. In this network, service-related functions are not only independent from underlying transport related technologies, but also closely related with each other for supplying services. It enables unfettered access for users to networks and to competing service providers and/or services of their choice. It supports generalized mobility which will allow consistent and ubiquitous provision of services to users."

This definition is somewhat abstract. Although the integration on IP is still the fundamental nature of NGNs, this definition by the ITU-T may be leaving some room for interpretation, giving consideration to the different situations of telecom carriers in various countries.

2-4 Experiments relating to NGN and current situation

NGNs have been operated on a test basis in many industrialized nations. Although the status of these tests has been reported widely, Table 3 lists data compiled by the Organization for Economic Co-operation and Development (OECD)[1]. NGN experiments have been run in many countries: Austria, Canada, Finland, France, Germany, Italy, Japan, Korea, Poland, the United Kingdom, and the United States. This shows that NGNs are a challenge common to the industrialized world, and that Japan has not particularly taken the lead in this testing. Table 3 also suggests that manufacturers of telecommunications devices are cooperating in this testing as partners.

There are two types of NGN experiment: The first type of experiment is building the communications network itself. Verification of service quality and the like fall into this category. The second type of experiment tests NGN services. Elisa Communications in Finland, KT in Korea, and Qwest in the US are conducting this type of experiment.

2-5 Status of NGN adoption in the West

In 2001, the United Kingdom published "UK Online: the broadband future." This document defines the country's vision for domestic communications infrastructure. In September 2004, then-prime minister Tony Blair declared that they would "bring broadband technology to every home in Britain that wants it by 2008." It would be fair to say that the UK is placing more focus on broadband convergence than other European nations. The telecom carrier BT has announced that it would convert its telephone network completely

Table 3 : Status of NGN Testing in Various Countries

Telecommunication network operators (Country)	Research and test activities	Partners
Telekom Austria (Austria)	Preliminary tests over last 2 years	
Bell Canada (Canada)	Operating 'Innovation Center' and other R&D laboratories	Nortel
TeliaSonera (Finland)	Developed NGN and platforms	Tellabs
Elisa Communications (Finland)	Research on NGN architecture and protocols	
	Launched the 'NGN and new communication service R&D project	
France Telecom (France)	Carried out experiments for NGN	
	Developing 'QoS monitoring systems'	
	Developing NGN architecture for mobile/fixed networks interoperability	Siemens
	Developed network architecture for NGN	Alcatel
Deutsche Telekom (Germany)	Implementing a pilot project	Alcatel, Siemens
Telecom Italia (Italy)	Established 'Telecom Italia Lab'	
NTT (Japan)	Runs a R&D program focused on RENA	
KT (Korea)	Establising test-bed for NGN	Lucent
	Launched a softswitch development project	LG
	Started to develop an application service test platform	Intel Korea
TP (Telekomunikacia Polska) (Poland)	Started NGN R&D project	France Telecom
BT (UK)	Align technology and sales	Cisco
Qwest (US)	Developing broadband business multimedia services	Cisco

Prepared by the STFC based on reference [1].

to IP, and it is also focused on converting to broadband. Until now, BT has built separate networks for each service. As a result, it now has 16 of them. In the future, the company plans to merge all of these into a broadband IP network. The next-generation network that BT is building is called "21CN" [3]. The company is investing £10 billion (about ¥2.5 trillion) on this project. BT has begun converting to an IP network starting in southern Wales, based on the 21CN plan. BT plans to switch over 115,000 users per week over the four-year period from 2007 to 2011. A total of 5,500 subscriber switches will be connected to 100 metro nodes, which will in turn be connected to the core network in 14 locations.

Although the same move is underway in Finland, Germany, and Italy as well, what is going on in the Netherlands in particular is symbolic insofar as considering what the NGN should be like. Figure 2 describes the situation in the Netherlands, which is relatively advanced. The Dutch telecom carrier KPN has already started to convert its network to all IP. This strategy includes connecting connection points called cabinets installed along roadsides using optical fiber. There are a total of 28,000 cabinets, and the total investment will be €900 million. The company calculates that it can save a total of €850 million in operating costs by switching to an NGN, so in terms of investment it makes economic sense. KPN also plans to offer optical fiber as much as possible to new customers starting in fiscal 2007 [2]. Creating an NGN requires new capital investment. This increased cost can be recovered over the long term through reduced operating expenses. KPN is advancing its plan to create an NGN under this logic.

The United States has been slower than Europe overall. Some telecom carriers, however, have launched IPTV service, which offers television broadcasts via broadband. The selling points of IPTV include more channels than cable television; digital recording features; and the ability to control digital recording via the Web.

The various countries' telecom carriers are split over whether to offer services like video distribution, or to just offer an NGN as a

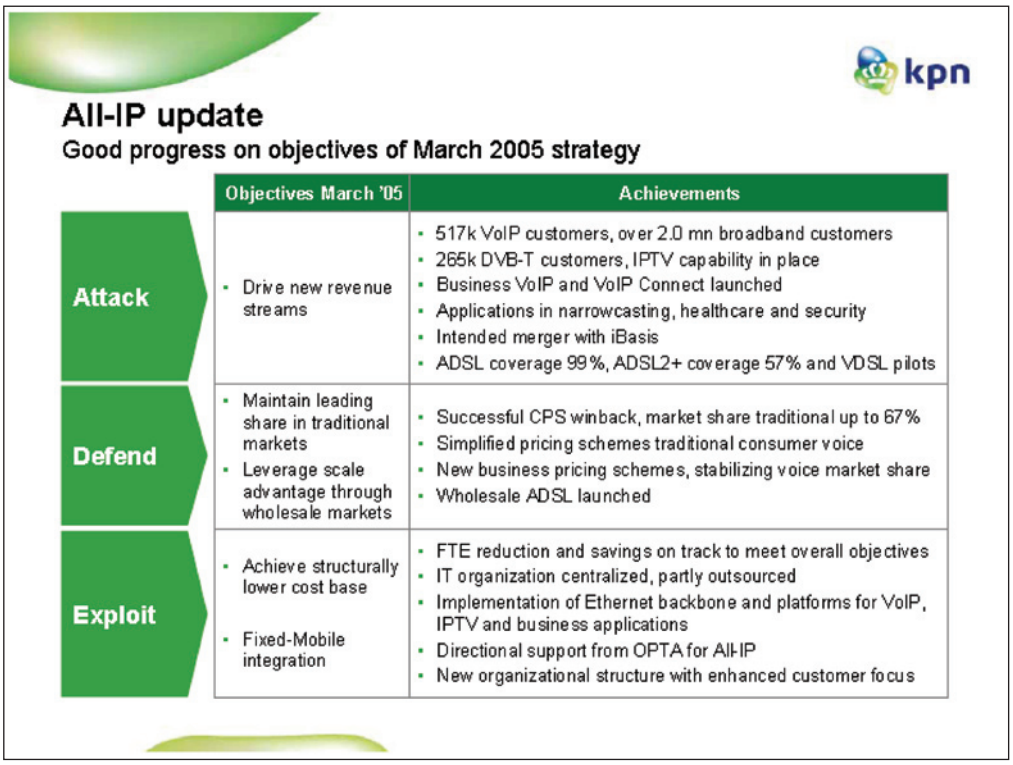


Figure 2 : KPN's All-IP target and progress

Source: References [4,5] (according to performance announcement at end-2006)

communication environment. Companies are currently monitoring the reactions of users, while searching for the optimum business model.

2-6 Status of NGN adoption in asia

The Asian country with the most impassioned commitment to NGNs is Korea. In Korea, NGNs are called "Broadband convergence Networks"(BcN). In 2004, the Korean government announced the u-Korea Promotion Strategy as its strategy for creating broadband infrastructure. It will build a wide-area integrated BcN as a seamless infrastructure merging communications, broadcasting, and the Internet. The target of the strategy is to enable connection by 20 million subscribers by 2010.

The targets set out in the u-Korea Promotion Strategy are highly ambitious. The target speed is 1,000 times the speed of ADSL (10 megabits/second to 10 gigabits/second), and they are attempting to accomplish this using optical fiber. Korea's commitment to NGNs is part of its strategy to achieve dominance in the information and telecommunications markets by being the first to attain leading-edge Internet technologies. Korea estimates that the total size of the worldwide market for NGN-related technologies is $180 billion/year. In other words, it expects the NGN market to grow to over half the size of the total network market.

Korea has also contributed many documents for the standardization of NGNs to the ITU-T. When one of the authors was visiting the Telecommunications Technology Association in the fall of 2007, he met with one of the people involved with this effort, and was told the following.

"The Korean government funds many projects, and one of the indicators for evaluating the performance of these projects is the degree to which it contributes to international standardization. Because of this evaluation indicator, Korean companies contribute many documents to the ITU-T. They sent 114 documents in 2005, 102 of which were accepted. In 2006, they sent 150 documents, 146 of which were accepted. This number is sure to increase in 2007."

In other words, the government actively uses the level of contribution to international standards as a measure of the success of the projects it funds. Europe is playing a leading role in the creation of international standards for NGNs. Given this background, the comments above by a person involved with the Korean efforts are extremely interesting.

Next, Chinese fixed-phone carriers invested $25.4 billion in 2005. This investment is being used to build NGNs. In parallel, carriers are converting their mobile-communications networks to use IP. Foreign-owned communications-device manufacturers and others are actively starting up businesses in China. China is also laying international undersea cable in partnership with foreign-owned firms.

The comments of Hong Kong's Telecommunications Standards Advisory Committee also actively promote NGNs: A report published by the committee states that NGNs will connect to the conventional public phone network, and gradually replace its functionality, and that over the long term, conventional public-phone networks will be phased out completely[6].

Singapore calls its next-generation network for mobile communications the Next Generation National Infocomm Infrastructure (Next Gen NII). In 2006, Singaporean government began a call for proposals of a conceptual design [7]. Next Gen NII has ambitious targets, aiming for use connections of 100 megabits/second initially, and 100 gigabits/second in the future. These high-speed connections are naturally expected to support such video communications as high-definition television broadcasts. The total investment is calculated at US $1.2 billion.

The Indian government realized the need for NGNs in 2005, and began policy initiatives to encourage them. The government predicts that NGNs will have a major impact on telecom carriers as well as on consumers, especially in sparsely populated areas. In 2006, the Telecom Regulatory Authority of India (TRAI) published a report emphasizing the following points (paraphrased)[8]:

The government wishes to create a single nationwide license for NGN services (voice, data, and video). It will also make it possible for smaller telecom carriers in sparsely populated regions to obtain licenses for NGN services in those regions. The TRAI believes that both national-scale service and service in sparsely populated regions should be achieved by promoting interconnectivity and unbundling. It also recommends a policy of easing the conditions for obtaining radio licenses in order to promote the dissemination of NGNs in sparsely populated regions.

3 Trends in NWGN research

3-1 Basic concept of NWGNs

While the concept of NGNs described above has been depicted as an extension of existing telecommunications, the research into NWGNs going on around the world deals with completely new network architectures. Researchers are creating a new networking concept, breaking away from the development path that broadcast and communication networks have followed to date, and starting from a blank slate. Their position is that after we have created such networks, we will think about how to migrate to them from our current networks.

The greatest challenge for researching new network architectures is network security. The fundamental design concept of the current Internet contains security vulnerabilities. This is due to the fact that the networks developed in an open state. NWGNs will create safe and secure social infrastructure by minimizing these vulnerabilities to the greatest extent possible. Doing this will require such actions as a revision of the technical system relating to network addresses, and more advanced support for movement by communicators. The management of these networks must thus include closed elements that have been absent from the Internet to date. The Internet has developed in many directions in the current open environment. NWGNs must be flexible enough to encompass the diverse applications that have grown up to date, and support further progress as well. In order to support these demands, which at first glance appear to be at odds, the NWGNs must make it possible to flexibly select the technical infrastructure components, and integrate them within a simple architecture. Below, the authors describe trends in the efforts to research and develop NWGNs.

3-2 Japan's AKARI project

The Architecture Design Project that Illuminates the Path to the New Generation Network (AKARI) is a project that the NiCT has been implementing since 2007[16]. The AKARI project "aims to build technologies for new generation networks by 2015, developing a network architecture and creating a network design based on that architecture." Below

is an overview of the design concept, referring to the AKARI project documentation.

The AKARI project first sets out to "pursue an ideal solution for network design from a clean slate without being impeded by existing constraints." The project is thus studying a grand design for a network to serve as social infrastructure, without being shackled by the historical developments of telephone, data communications, and broadcast technologies described at the beginning of this document. Table 4 lists the research targets of the AKARI project.

These research goals all come back to the design of network architecture. The term "architecture"of course comes originally from the world of construction, and is used to indicate the fundamental structure of the network. The AKARI project is building the ideal architecture based on the following three principles:

(1) KISS (Keep It Simple, Stupid)
(2) Sustainable and Evolutionary
(3) Reality Connection

NWGN research projects outside Japan (described below) also emphasize these concepts, and are taking similar directions in their research. Taking the existing state of information networks as a given, and extending to future development, these are probably essential factors. Below, each of these is described in more detail.

(1) The KISS principle (Keep It Simple, Stupid)

It is said that in the infancy of the development of the Internet, the leader of a development project told his subordinates to "Keep It Simple, Stupid." If the specifications for the routers and other devices making up the network become too complex, then vendors may capture the market based on specific technologies. The idea behind KISS is to make the technical specifications as simple as possible, in order to guarantee that it is possible to build a network environment through open participation.

KISS must be the underlying principle of the entire network, in order to increase diversity, extensibility, and reliability. If one's aim is to offer more advanced network functionality, this could make the architecture more complex. This principle is an admonition about making things complicated without due consideration. In order to adhere to this principle, the AKARI project has introduced principled approaches to research and development: "end-to-end," "crystal synthesis," and "common layer."

(2) Sustainable and evolutionary

The diversity and richness of applications that have appeared since the Internet came to be used commercially in the early 1990s gives us an indication of the importance of this principle. Taking mobile phones as an example, applications like browsing Web pages were not envisioned when mobile phones first appeared.

According to the AKARI project documentation, "The new generation network architecture must be designed as a sustainable network that can evolve and develop in response to changing requirements. It is important for the network to have a simple structure and for service diversity to be ensured in end or edge nodes." It further states, "The Internet to date has also maintained this principle. Having a shared addressing system, while enabling the networks forming subparts of the Internet to manage themselves, has helped make it possible to provide a flexible environment that has engendered diverse growth and development. The new network environment should also have an extensible architecture that encourages the development of innovative new applications."

The AKARI project is advancing development based on this principle, aiming to achieve self-organizing network design; robust, large-scale networks; controls for a topologically fluctuating network; scalable, distributed controls; and openness.

(3) Reality connection

Many of the technical issues occurring on the Internet today are due to the fact that entities (communicators) in the address space of the network are disassociated from the space of real-world society. To give an extreme example, it is possible to fake the sender of email spam, thus making it difficult to identify the perpetrator. Moreover, sensor networks are being built, and in the near future their use is expected to increase. From an operational and management perspective, there is a need to separate physical and logical

Table 4 : Research targets of the AKARI project

(1) Peta-bps class backbone network, 10Gbps FTTH, e-Science
(2) 100 billion devices, machine to machine (M2M), 1 million broadcasting stations
(3) Principles of competition and user-orientation
(4) Essential services (medical care, transportation, emergency services), 99.99% reliability
(5) Safety, peace of mind (privacy, monetary and credit services, food supply traceability, disaster services)
(6) Affluent society, disabled persons, aged society, long-tail applications
(7) Monitoring of global environment and human society
(8) Integration of communication and broadcasting, Web 2.0
(9) Economic incentives (business-cost models)
(10) Ecology and sustainable society
(11) Human potential, universal communication

Source : Reference [16]

addressing in these networks. There are expected to be scenarios in which a specific location could use a huge number of addresses during a short period of time. At this time, it will be necessary to meet the demands for authentication and traceability for a large number of communication devices. The separation of physical and logical addressing, bi-directional authentication, and traceability have thus been identified as critical research topics.

3-3 America's GENI project

The Global Environment for Network Innovations (GENI) project conducts large-scale research into next-generation networks and their applications, funded by a research grant from the US National Science Foundation (NSF) [14]. A huge research grant of $400 million will be invested into research on next-generation information networks, over a five-year period starting in the fall of 2007.

On 25 April 2007, the project concentrated a year and a half of discussion over the implementation plan, and published a Project Execution Plan, comprising a Research Plan, System Requirements Document, Functional Design, and Construction Plan. In September 2007, the five-year research project began. The GENI project is researching the future of information networks through research and development in such fields as sensors, opto-electronics, systems on a chip, fast, large-scale computation, large-scale databases, and new algorithms, unfettered by the existing Internet framework. GENI's physical layer also consists of a diverse range of network devices, including wireless networks. Meanwhile, on the applications front, experiments using the network are interoperable by means of a software management facility. In order to make this possible, the following four key ideas are critical for the subprojects conducting experiments and R&D.

(1) Sub-project components must be programmable, to support any network environment. (2) It must be possible to virtualize the implementation of multiple components. This permits the operation of experiments designed for continuity over a broad scope. (3) It must be possible for terminal devices and users to seamlessly participate in experiments. Creating implementations at the stage of actual operation enables evolutionary improvement. (4) Components must have modular structures, enabling new technologies to be flexibly added and removed. This is to hold up to dynamic operation.

The NSF's Computer & Information Science and Engineering Directorate (CISE) had a central role in the proposal of the GENI project. It aims to promote computer science, telecommunication technology, information science and technology. Through the CISE, information-network and distribute-systems researchers and academics have discussed this project through several workshops.

3-4 Europe's Euro-NGI project

In Europe, a project called Euro-NGI (Next Generation Internet) is one of the main projects of a basic science and technology program called the 6th Framework Program (FP6). Euro-NGI has been implemented since 2003, and in the subsequent FP7 framework, it was continued under the name

Table 5 : Objectives set out in Euro-NGI plan

1) Integrating and rationalizing the European research efforts - Establishing and Updating a Knowledge Roadmap.
2) Implementing a Strong Program of Jointly Executed Research Activities.
3) Sharing Methods, Tools and Platform Developments.
4) Organization of internal Workshops for a thorough exchange of ideas and knowledge.
5) Facilitating the Information Exchange and the usage of Advanced Communication Tools.
6) Facilitating the Mobility of Researchers and PhD Students.
7) Creation of a European graduate courses program for PhD students to maintain and the high expertise level.
8) Creation of a Summer School as a tool for communication and for training of young researchers.
9) Spreading Excellence.
10) Consolidate collaboration with industry.

Source : Reference [15]

Euro-FGI (Future Generation Internet)[15].

Major research institutes in Europe participate in this project, across national boundaries. The Euro-NGI project is being implemented in accordance with the ten objectives in the Euro-NGI Plan, which are listed in Table 5.

From the standpoint of managing a research project, the greatest challenge for Europe is to holistically expand the research activities within its domain without waste, and integrate the results of this research. The Virtual Center of Excellence was built for this purpose: to integrate the knowledge gained from research throughout Europe. The center uses such means as remote interregional videoconferencing systems and Web-mediated chat rooms for experts, in order to create networks for researchers and share knowledge effectively. Of course, the need for actual meeting and exchange is not forgotten either; there is also mutual exchange of research at the graduate-school level, including by having PhD students advised by professors in different institutions.

4 Conclusions

In order to consider the challenges for research and development of next-generation networks, let us have another look at the research objectives for Japan's AKARI project, listed in Table 4. At most, the only research objectives that are purely technological are (1) and (2); most of them assume regulatory or systemic reforms, or refinement via market competition.

In this document, the authors have argued that from a technological perspective, IP convergence of the networks will drive the next technological evolution in the communications field. The ones with the strongest need for IP convergence, however, are the telecom carriers that want to optimize investment in infrastructure. Although users have heard promises about increased bandwidth, bandwidth guarantees, and the like, the benefits of IP convergence are actually hard for ordinary users to understand.

The objectives in Table 4 that are influenced by non-technical factors (3 to 11) have strong relationships with applications in modern society. Approaches to such matters as policy decision-making and considerations of business marketability will likely prove essential to achieving these research objectives.

Take for example "integration of broadcasting & communication" in item 8. If communication services with guaranteed bandwidth are offered at appropriate prices, then it will be possible for the Internet to take the place of current television functionality, as well as to create new services and businesses. The integration of broadcasting and communication is more a policy issue than a technology issue.

As social infrastructure, the Internet has already given birth to many global enterprises. Moving forward, adding the perspectives of bandwidth guarantees and ubiquitous networking will increase the convenience to users, as well as make possible the creation of new industries that we cannot even imagine today.

Rather than relying exclusively on the views of engineering experts, discussion of the future of information networks will require the knowledge and views of experts in a wide range of fields. Innovation in many fields would be impossible to

imagine without the social infrastructure of the network. It is thus necessary to study the impact of NWGNs on research and development in other fields, and discuss measures for them to serve as a foundation for creating innovation in a wide range of fields.

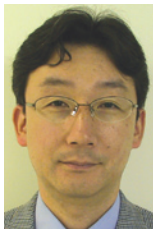
Acknowledgements

The authors gratefully acknowledge the valuable discussion and relevant materials provided by Dr. Masato Shinagawa, affiliated fellow at the Science and Technology Foresight Center of the National Institute of Science and Technology Policy.

References

[1] Working Party on Telecommunication and Information Services Policies, "Next Generation Network Development inOECD Countries," OECD (2005) : http://www.oecd.org/dataoecd/58/11/34696726.pdf
[2] UK Cabinet Office, "UK Online: the broadband future," : http://www.connectingsw.net/uploads/ukonline_1.pdf
[3] "What is 21CN?," : http://www.btplc.com/21CN/Whatis21CN/index.htm
[4] KPN, "4th Quarter Results, 2006," : http://www.kpn.com/upload/1863237_9475_1170923304042-KPNQ406a.pdf
[5] KPN, "Annual Report 2006," : http://www.kpn.com/upload/1786687_9475_1173767749534-KPN_Annual_Report_and_Form_20-F_2006.pdf
[6] Telecommunications Standards Advisory Committee, "Overview of Next Generation Network," : http://www.ofta.gov.hk/en/ad-comm/tsac/ts-paper/ts2005p13.pdf
[7] Computer World, "Next Gen NII," : http://www.computerworld.com.my/ShowPage.aspx?pagetype=2&articleid=3614&pubid=3&issueid=86
[8] Telecom Regulatory Authority of India: http://www.trai.gov.in/
[9] "Future of NGNs in a Maze," Weekly Diamond (1 March 2008) (Japanese)
[10] NTT Technology Journal : http://www.ngs-forum.jp/library/journal.html (Japanese)
[11] T.Kiyosada and H. Yamada, "Trends in the Research and Development of Mobile Communication Systems," Trends in Science and Technology, No. 1, April 2001 (Japanese)
[12] A. Ogasawara, "Next Generation Content Distribution Technologies in the Broadband Era" (Japanese), Trends in Science and Technology, No. 31, October 2003 (Japanese)
[13] A.Fujii, "Technology Trends in Internet Routers : Toward the Creation of Next Generation Communication Infrastructure", Trends in Science and Technology, No. 33, December 2003 (Japanese)
[14] Global Environment Network Innovations (GENI) website : http://www.geni.net/
[15] Euro-NGI website : http://www.eurongi.org/
[16] AKARI Project website : http://akari-project.nict.go.jp/

Profile



Akihiro FUJII
Affiliated Fellow,NISTEP

Associate Professor, Department of Applied Informatics, Faculty of Science and Engineering, Hosei University
Dr. Eng.. After work in the research of distributed computing and communications protocols, Dr. Fujii carried out a project to build an e-commerce system. He is currently interested in the influence that innovations in information and communications technologies have on business and policy.

Hajime YAMADA
Affiliated Fellow,NISTEP

Professor, Department of Social Economic Systems, Faculty of Economics, Toyo University
Dr. Yamada holds a Master's Degree and Dr. Eng. from Keio University. He also holds a Master's Degree in Management of Technology from the Massachusetts Institute of Technology. After direct involvement in research at NTT, Dr. Yamada began working in management, including creating and proposing research strategies. He has been a professor at the Toyo University Faculty of Economics since 2002.

(Original Japanese version: published in November 2008)

2

Developments in Television Band Frequency Sharing Technology

HAJIME YAMADA
Affiliated Fellow

1 Introduction

The digitalization of television broadcasting, scheduled for completion on July 24, 2011, is underway. All broadcast stations are building digital broadcast relay networks. The statistics compiled by the Japan Electronics and Information Technology Industries Association (JEITA) show that, as of the end of August 2008, the number of receivers for terrestrial digital television broadcasting shipped to domestic markets reached 24,380,000, which is attributable to the rapid penetration of liquid crystal and plasma flat-panel televisions.

The ultimate aim of the digitalization of television broadcasting is the effective use of radio waves. Analog broadcasting uses Channels 1 to 62, whereas digital broadcasting only uses Channels 13 to 52. The unused 22 channels, which correspond to a frequency bandwidth of 130 megahertz (MHz), are converted to other uses. Given the current pressure on available radio wave resources due to the spread and development of various information and communications technologies such as cell phones and intelligent transport systems (ITS), the availability of such a wide bandwidth for use raises expectations for new applications.

The Committee on the Effective Use of Radio Waves, a subsidiary organization of the Telecommunications Council under the Ministry of Internal Affairs and Communications, studied the effective use of the 22 channels and submitted a report to the Ministry in July 2007[1]. Figure 1 summarizes the report. The recommendations include a variety of uses for non-television broadcasting, private communications, ITS, and telecommunications (for cell phones). Digitalization projects are also in progress in other countries.

If it is possible for multiple applications to share the same frequency, radio waves can be used

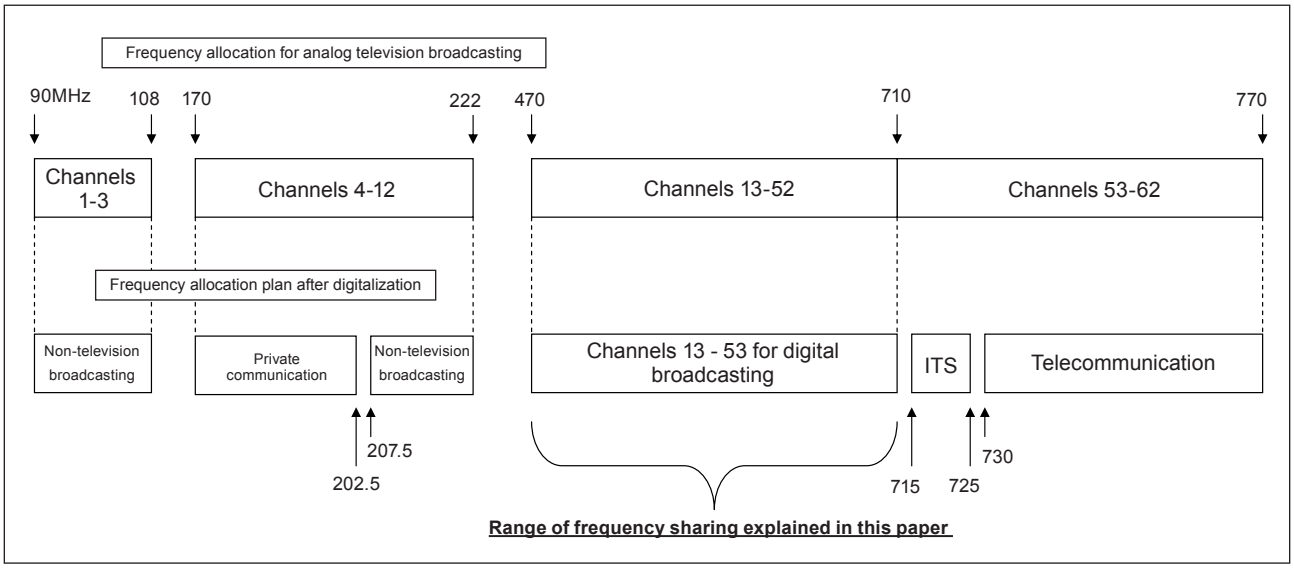


Figure 1 : Conversion plan for frequency bands made available by the digitalization of television broadcasting

Prepared by the STFC based on reference [1]

more effectively. The status of the basic research on frequency sharing technology was published previously in the November 2004 issue of Science and Technology Trends[2]. Advances in technology have made television band frequency sharing after the completion of digitalization a realistic possibility. Besides the newly available 22 channels described above, the shared use of the frequency range between Channels 13 and 52 for other applications will have a significant effect on our society.

Community FM stations, which started in 1992, soon spread throughout Japan. According to the statistics compiled by the Ministry of Internal Affairs and Communications, their number had reached 218 by the end of 2007[3]. These stations are positioned as media which meet community needs not covered by the conventional prefecture-based radio stations. This paper introduces two experimental projects on frequency sharing, which aim at a kind of community television broadcasting with a possibility of evolving into a new medium in the same way as community FM broadcasting.

2 Interference protection ratios and channel plans

Let us assume that broadcast stations A and B are located at Points A and B, respectively, as shown in Figure 2. The intensity of the radio waves (received field intensity) from Station A decreases with distance, reaching the sensitivity limit of the receiver at Point C. In other words, broadcasts from Station A can be received between Points A and C. In the same way, broadcasts from Station B can be received between Points B and D.

Even if Stations A and B broadcast in the same channel, no interference occurs if Points C and D are sufficiently distant from each other. The distance between the two stations using the same channel can, in principle, be narrowed until Points C and D approach each other infinitely. In actual cases, however, the field intensity of the radio waves from Station B received at Point C (in the same way, the field intensity of the radio waves from Station A received at Point D) is limited to a lower level than the sensitivity limit of the receiver in order to avoid interference with a good margin under a technical condition called the "interference protection ratio."

Based on experimental results obtained by NHK (Japan Broadcasting Corporation), the Telecommunications Technology Council submitted a report on the interference protection ratio in digital broadcasting in 1999[4]. Then the Ministry of Posts and Telecommunications issued an ordinance and a notice on the basis of the findings of that report in the same year. The revised versions of these documents were published in 2003 as Ministry of Internal Affairs and Communications Ordinance No. 26 of 2003 and Ministry of Internal Affairs and Communications Notice No. 37 of 2003, respectively.

Table 1 shows the interference protection ratios extracted from the above ordinance and notice.

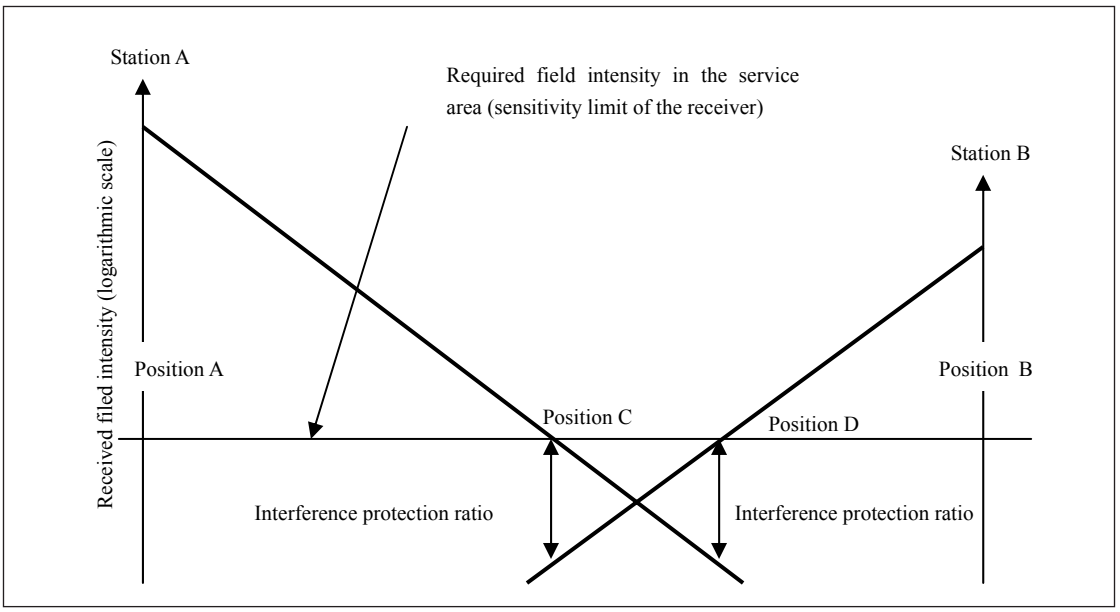


Figure 2 : Field intensity condition for enabling shared use of the same channel (explanation of principle)
Prepared by the STFC

Table 1 : Interference protection ratio

Station A (Desired wave)	Station B (Interference wave)	Frequency difference	Interference protection ratio
Analog	Analog	Same channel	45dB
		Lower adjacent (Interference with lower adjacent channel)	0dB
		Upper adjacent (Interference with upper adjacent channel)	10dB
	Digital	Same channel	45dB
		Lower adjacent (Interference with lower adjacent channel	0dB
		Upper adjacent (Interference with upper adjacent channel)	10dB
Digital	Analog	Same channel	30dB[Note]
		Lower adjacent (Interference with lower adjacent channel)	-21dB
		Upper adjacent (Interference with upper adjacent channel)	-24dB
	Digital	Same channel	28dB
		Lower adjacent (Interference with lower adjacent channel)	-26dB
		Upper adjacent (Interference with upper adjacent channel)	-29dB

Note : Changed to 20dB in the channel allocation plan with the expectation of improved receiver characteristics

Prepared by the STFC based on the report of the Telecommunications Council and other references

When the desired wave to be received at Point C is the analog broadcast wave from the antenna of Station A, the digital broadcast wave from the antenna of Station B interferes with the desired wave at Point C. The technical condition for this case specified in Table 1 requires that the field intensity of the radio wave from Station B must be 45dB smaller than that from Station A. When the desired wave is analog, the same interference protection ratio value is required for both analog and digital interference waves. On the other hand, the interference protection ratio when both the desired and interference waves are digital is specified as 28dB, which allows the field intensity of the interference wave to be 17dB larger than that in the case of analog versus digital waves.

The "upper adjacent" and "lower adjacent" mean the cases where the interference wave uses an upper channel and a lower channel, respectively, immediately adjacent to the channel used by the station transmitting the desired wave. When Station A uses Channel 30 and Station B uses its upper adjacent, i.e., Channel 31, the required interference protection ratios for the cases of analog versus analog waves and digital versus digital waves are 10dB and −29dB, respectively. A negative value of the interference protection ratio means that the desired wave can be properly received without interference even if the field intensity of the interference wave is larger than that of the desired wave.

As described above, digital broadcasting has a stronger tolerance for interference than analog broadcasting, making it possible to narrow the distances between broadcast stations using the same channel and facilitating the effective use of adjacent channels. Thus, digital broadcasting can make more effective use of radio waves than analog broadcasting. This is the technical justification of the bold policy of digitalizing television broadcasting, which reduces the frequency range by using the range for only 40 channels instead of that for 62 channels in order to make the remaining range available for new applications.

The interference protection ratios and the calculation formula[NOTE 1]for field intensities are specified in the technical conditions given by the ordinance and notice.

[NOTE1]
The formula calculates the received field intensity decreasing with the distance from a broadcast station.

Actual channel plans are determined considering the perturbations caused by urban areas, mountains, sea surfaces, etc., in addition to the calculated results. Simulation software programs for this purpose have been developed and are being used[5]. The specified channel plans of digital broadcasting are published by the Ministry of Internal Affairs and Communications[6].

3 Principle of frequency sharing

In the Kanto Region, which includes Tokyo, NHK and key commercial broadcast stations are using Channels 21-27 to broadcast in digital from antennas installed on Tokyo Tower. The number of channels used is 10 at most, including one other local station in Tokyo and the Open University of Japan. It therefore appears that the remaining 30 channels are unused.

In actual broadcasting, the radio waves from Tokyo Tower cannot cover the whole Kanto area due to the disturbing effects of intervening mountains. Relay stations have been built to retransmit television broadcast signals to areas with poor reception. NHK General TV uses Channel 27 for transmissions from Tokyo Tower, but its relay stations in Odawara, Kiryu, and Chichibu use Channels 19, 37, and 13, respectively, for retransmission. Such relay stations consume a considerable part of the remaining 30 channels.

The system in which relay stations use a different channel to retransmit transmitted signals is called a Multi Frequency Network (MFN). Integrated Services Digital Broadcasting–Terrestrial (ISDB-T), the Japanese digital broadcasting system, can in fact be constructed as a Single Frequency Network (SFN). The Ministry of Internal Affairs and Communications and the industry are running a publicity campaign to market the ISDB-T in South American countries, explaining that the ISDB-T can be constructed as an SFN[NOTE 2].By nature, the frequency use efficiency of the SFN is higher than that of the MFN.

Bound by the design concept in the age of analog broadcasting, the Japanese domestic system uses an MFN. The number of unused channels (hereafter called empty channels) is small, but a few exist in the Kanto Region. For example, the residents of Minato Ward directly under Tokyo Tower do not need to receive NHK General TV transmissions from Chichibu, and in any case, it is virtually meaningless for radio waves from Chichibu to reach Minato Ward. As a result, Channel 13 is empty in the ward. On the other hand, Channel 27 is empty in Chichibu City, Saitama Prefecture, because NHK General TV over that channel cannot reach Chichibu.

In regions with a small number of television stations, the number of empty channels is larger. A sample survey of Okayama Prefecture based on the channel plan published by the Ministry of Internal Affairs and Communications shows that 32 relay stations use Channel 29 and 29 relay stations use Channel 13, whereas Channels 23, 25, and 35 are not used at all. Channel 23, for example, is used at Hachibuse in Tottori Prefecture, Hamada in Shimane Prefecture, Hiroshima, Kure, and Miyoshi in Hiroshima Prefecture, and Iwaki in Ehime Prefecture. All of these areas are distant from Okayama City in Okayama Prefecture. Hence, Channel 23 is practically empty within Okayama City.

Digital broadcast relay networks are under construction. The first-step construction of the systems for relaying broadcasts from prefectural capitals has been finished. The second- and third-step construction of the systems for relaying broadcasts to mountainous areas is underway. The second and third steps also use the MFN. Meanwhile, the Ministry of Internal Affairs and Communications has already succeeded in developing a relay technology for SFN using a simple technology called a gap filler[7]. Using this technology in the second- and third-step systems to reduce the number of channels to be used will be effective for increasing the number of empty channels.

The absolute requirement for using an empty channel is not to cause interference with the digital braodcasting. This means that using Channel 23 in Okayama City must not disturb broadcast reception in Hiroshima City. For this, the user of the empty

[NOTE2]
Mr. Takada from NHK, for example, emphasized the SFN in his lecture “ISDB-T Technologías de transmisión y sistemas de alertas de emergencia,” which was presented at a seminar held in Columbia on March 12, 2008. (http://www.dibeg.org/seminar/0803colombia_ISDB-T_seminar/2_ISBT_ews_Spanish.pdf)

channel should control the antenna power (power provided from the transmitter to the antenna). As shown in Figure 2, a small difference in distance may reduce the field intensity of the radio waves from Station B to a level lower than the sensitivity limit of the receiver if the antenna power is small.

As described so far, television band frequency sharing is a technology for using different empty channels in different areas, and emits radio waves with low antenna power.

4 Experimental projects for the shared use of empty channels

Two experimental projects for sharing unused channels are being conducted in Japan. One is a project being conducted by the Yokosuka Research Park (YRP) R&D Promotion Committee and Area Portal Company Limited. The other is an experimental project being conducted at the Specified Ubiquitous District in Matsue City, Shimane Prefecture. Both projects intend to develop a new broadcasting system, which can be properly called a community television system, with a possibility of evolving into a new medium in local communities.

4-1 Experiments in Yokosuka research Park

The YRP R&D Promotion Committee has been conducting experiments in which it offers community programs to a limited area using One Segment Broadcasting. The ISDB-T divides a channel into 13 segments and bundles several segments to transmit moving images and sounds. One Segment Broadcasting broadcasts information to mobile devices such as cell phones using only one segment.

Based on the experimental results, the first general meeting of the Preparatory Committee for New Narrow-area Digital Community Broadcasting was held on September 29, 2008. Aiming at the practical use of One Segment Broadcasting, the Committee will play a central role in conducting demonstration experiments at the Yokosuka Specified Ubiquitous District[8].

The basic technology for new narrow-area community broadcasting was transferred to the Area Portal Company, which has conducted several verification experiments[9]. The main point of this narrow-area broadcasting is to offer community programs with a very low antenna power of 10 milliwatts (mW). This output power is used to meet

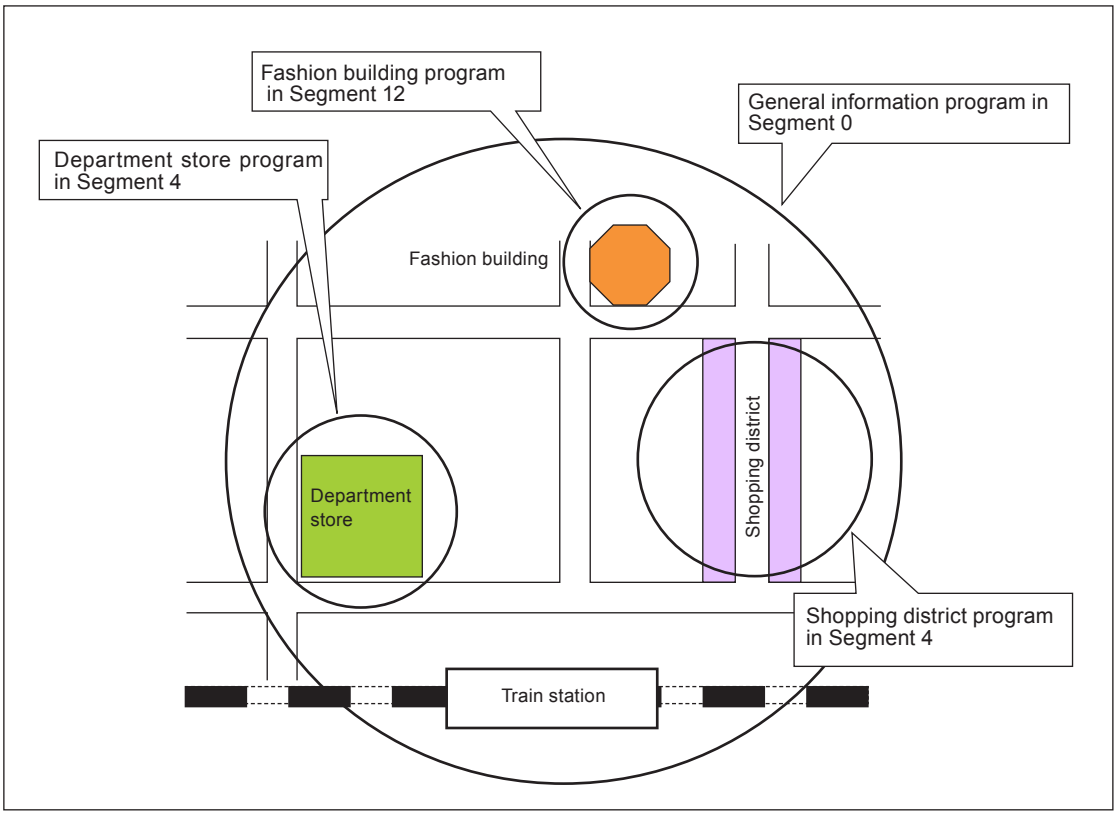


Figure 3 : Image of broadcasting in front of a train station where a department store and shopping district broadcast various programs

Prepared by the STFC based on the references provided by Area Portal Company

the regulation specifying that a wireless station license is not needed to run a station with an antenna power of less than 10mW in the frequency ranges used for wireless LANs and other applications.

In the case of an antenna power of 10mW, the calculation formula for field intensities specified by the ordinance mentioned above gives a receivable area radius of about 300 meters. This is why the broadcasting is called "narrow-area." There is no possibility of disturbing distant relay stations using the same channel. One Segment Broadcasting is part of the ISDB-T technology, and its interference protection ratios can be determined using Table 1.

Cell phones can receive One Segment Broadcasting, but with a minor software change, can also receive new narrow-area digital community broadcasting. Each television channel is divided into 13 segments, and hence provides 13 independent programs at the same time. Even if alternate segments are used in order to improve tolerance against interference, it can provide 7 programs.

The new narrow-area digital community broadcasting technology enables independent broadcasting using the same channel at points several kilometers distant from each other. In the Tokyo metropolitan area, for example, broadcasting is possible in front of the large train stations at Shibuya, Shinjuku, and Ikebukuro, even though these stations are located only minutes apart.

Figure 3 shows an image of broadcasting in front of a train station. The youth-oriented fashion building and the department store use Segment 12 and Segment 4, respectively, and the shopping district, which is somewhat distant from the department store reuses and Segment to broadcast their original programs. A general information program is broadcast to the whole area using Segment 0.

In 2006, the Area Portal Company conducted an experiment at the National Museum of Western Art, in which it broadcasts a museum information program and individual programs for each artist in a batch. In 2007, the company conducted an experiment in which it broadcasts university life programs at Hiyoshi Campus, Keio University. The company also carried out several experiments in which it broadcasts car racing at Fuji International Speedway and Suzuka International Racing Course.

4-2 Experiments at the Matsue specified ubiquitous district, shimane prefecture

Using the system of Specified Ubiquitous Districts established by the Ministry of Internal Affairs and Communications, Matsue City in Shimane Prefecture started an experimental Shimane Ubiquitous Project on the shared use of television bands. The district was approved in January 2008. It is planning to continue verification experiments on the next-generation broadcast medium based at Shimane University until March 2011.

Shimane Prefecture, Matsue City, and Shimane University, and MediaScope, a local company playing a central role, are participating in the Shimane Ubiquitous Project Promotion Council[10]. The project uses MediaFLO wireless technology. MediaFLO is a multimedia broadcasting standard for mobile terminals specified in ITU-R BT.1833 which was standardized by the ITU (International Telecommunication Union)[11]. The standard also includes One Segment Broadcasting.

Using an antenna installed on the roof of a building on the campus of Shimane University, the project broadcasts programs to Matsue City over Channel 62 with an antenna power of 100W. The antenna power is as large as 10,000 times that for the new narrow-area digital community broadcasting described above. Since the analog Channel 62 is an empty channel in this area and no interference is expected, the Ministry of Internal Affairs and Communications issued a license for the project.

One Segment Broadcasting is designed so that all individual segments carry the same amount of information. On the other hand, MediaFLO dynamically allocates appropriate amounts of bit to individual program on an instant-by-instant basis. For example, when the instant-by-instant image movement is slight and contains much written information, as in news programs, a small amount of bit is allocated to the program. On the other hand, in sports programs, where movement is rapid, a large amount of bit is allocated to the program. Using the statistical characteristic that programs requiring large and small amounts of bit appear randomly at each moment, MediaFLO can transmit more programs in a channel than One Segment Broadcasting.

MediaFLO does not need to broadcast all programs to cell phones. As shown in Figure 4, it is also possible to transmit some programming to electric display

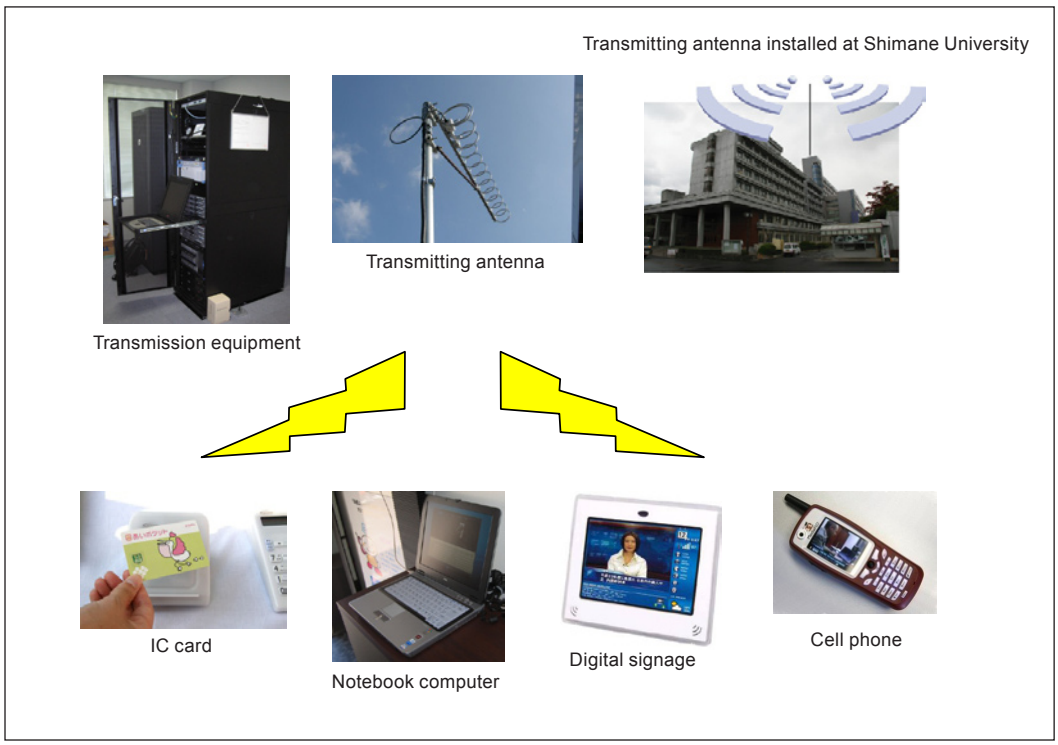


Figure 4 : Types of information services provided by the Shimane Ubiquitous Project

Source: Shimane Ubiquitous Project

boards, store display boards, and giant street screens (collectively called "digital signage"). Shimane Prefecture actually plans to broadcast advertising and area information to digital signage.

Sending disaster prevention information to digital signage can also be an important application. Since disasters due to sudden heavy downpours are geographically localized, broadcast stations can provide individual communities with unique information using digital signage, rather than using a digital television broadcast covering the whole area of a prefecture. Digital signage broadcasting can also be used in night-time storage of information for reproducing high-resolution moving images, which requires much time, and reproduction of those images on giant screens during the day.

One of the features of the Shimane Ubiquitous Project is the integration of IC card technology and multimedia broadcasting. A broadcast station sends service information to digital signage equipment placed at storefronts through MediaFLO. Customers use the points and coupons taken from the digital signage equipment and stored in their IC cards (Felica Pockets). The circulation patterns of the customers can be observed from their card usage histories. The project is planning to conduct experiments on such area-oriented information services.

4-3 New possibilities found through the two experiments

The new narrow-area ubiquitous community broadcasting and the Shimane Ubiquitous Project described above have commonalities and differences.

One of the commonalities is that they intend to realize a new type of broadcasting directed to narrow areas such as station squares and particular areas within a city. Conventional television broadcasting covers a whole prefecture in the smallest case and a wider area like Kanto Region in the largest case. Information programs such as those on small events in front of Shibuya Station and on local restaurants for the tourists visiting Matsue Castle have never been broadcast daily. Only companies like those in automobile, food, beverage, and cosmetic industries, which have customers in a wide area, advertise by conventional commercial broadcasting. In contrast, the two experimental projects target small sponsors in towns.

In other words, their business models are clearly different from that of the conventional television broadcasting in that they are intended to "respond to the needs of the smallest possible number of people in the narrowest possible area," whereas the conventional model is oriented toward "the largest possible number of people in the widest possible area." Commercial broadcast stations are participating in new narrow-

area digital community broadcasting, possibly because they expect that the new type of broadcasting will not cannibalize the conventional type.

The new type of broadcasting differs from the conventional one in its interoperability with the Internet. The new broadcasting systems are more active in using the Internet, for example, to link receivers to the websites of local restaurants. This kind of service is a step closer to the integration of communications and broadcasting.

On the other hand, the following differences may be mentioned.

With a minor software change, the new narrow-area digital community broadcasting provides programs to One Segment Broadcasting-supporting cell phones, whose domestic sales volume is larger than 30 million units. Since this type of broadcasting already has a large number of potential users, immediate penetration can be expected, provided it interests potential users. For the time being, the Shimane Ubiquitous Project using MediaFLO requires dedicated receivers. Thus, this system must develop services for which users are willing to use a dedicated receiver. As described above, MediaFLO can send more programs than One Segment Broadcasting at the same time and can provide digital signage with information programs. It also enables high-resolution broadcasting using stored program data. It is thought that the two new types of broadcasting systems will be commercialized through a process of mutual competition.

The interference protection ratio for One Segment Broadcasting is already specified in existing regulations, whereas it will be necessary to determine the ratio for MediaFLO experimentally before it can be put into practical use. This is a serious problem for MediaFLO applications. The Specified Ubiquitous Districts were implemented to evaluate the experimental results in the districts as a basis for determining whether nationwide application is feasible and appropriate. Therefore, if the Shimane Ubiquitous Project succeeds, moves to establish the interference protection ratio for MediaFLO are likely. Since both MediaFLO and ISDB-T use a transmission method of orthogonal frequency division multiplexing (OFDM), the radio wave characteristics of MediaFLO are considered to be approximately the same as those of ISDB-T. From this viewpoint, there is a high possibility that the same interference protection ratio as that for ISDB-T will be established for MediaFLO.

5 Aiming at further shared use of empty channels

The two experiments described previously are pioneering in demonstrating the feasibility of television band frequency sharing to the world. Their further development will contribute to the more effective use of radio waves. What direction will sharing technology take in the future? This will be explained using Figure 5.

Figure 5 explains the degrees of difficulty in frequency sharing. Low antenna power technically means that short distance movement of a receiver will reduce the received field intensity to below the receiver sensitivity limit. This means that frequency sharing is easier with low antenna power than with high antenna power. Increasing the amount of information transmitted per second requires using a wider frequency bandwidth. Whether multiple channels can be used simultaneously depends on whether empty channels are available, which makes the realization of higher bit-rate frequency sharing more difficult.

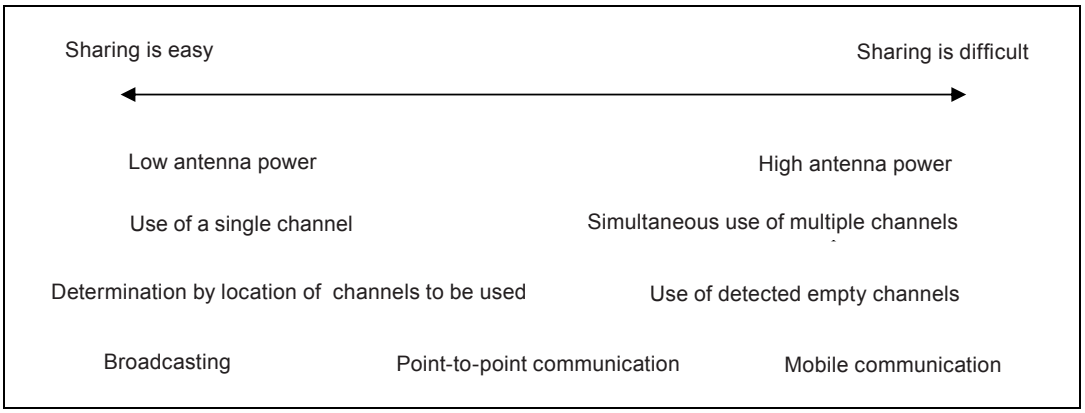


Figure 5 : Conceptual diagram showing the degrees of difficulty in frequency sharing

Prepared by the STFC

The method currently being tested, which determines empty channels based on locations, such as Minato Ward in Tokyo and Okayama City in Okayama Prefecture, is the closest to practical use. The next development target would be a system, called Geo Location System, which memorizes data on the relationship between locations and empty channels in the transmitter, uses GPS (Global Positioning System) to identify a location, and selects empty channels at the location. Practical application of "cognitive wireless systems," which scan frequencies just before transmission from the transmitter in order to find and use empty channels, is more difficult. Without the establishment of a method of avoiding mistakes, such as misidentifying a channel as an unused channel due to blockage by a building, practical application will be difficult. Research on cognitive radio technology is active in Japan, as shown by the fact that National Institute for Information and Communications Technology has established a New Generation Wireless Communications Research Center. Interest in cognitive radio technology has also been increasing in other countries. For example, the in-house magazine of the IEEE (Institute of Electrical and Electronics Engineers) published a report on cognitive radio in 2004[12]. Cognitive radio will be a focal point of the future research and development on frequency sharing.

Frequency sharing for broadcasting is easiest from the viewpoint of applications. Since the number of transmitting broadcast stations is limited, measures against possible interference are relatively simple. Communications between fixed stations are also a relatively easy application. One of its possible uses is a high-speed wireless LAN for home use. Application to mobile communications is the most difficult. To ensure that empty channels can be found, responding to every movement of mobile terminal, will require adoption of the previously described technologies such as Geo Location and cognitive radio, which have many technical problems.

The explanation presented above indicates that the two experiments discussed here, namely, new narrow-area digital community broadcasting and the Shimane Ubiquitous Project, are the frequency sharing methods with the highest possibility of application. Future research and development will be more advanced and challenging.

6 Trends in the United States and Europe

Both the United States and Europe are also working on television band frequency sharing.

In the United States, empty channels in television bands are called "white spaces." The American information and communications industries have proposed that white spaces should be used for applications other than broadcasting. The proponent companies organized the Wireless Innovation Alliance[13] to put pressure on the FCC (Federal Communications Commission). Consumers formed the Public Interest Spectrum Coalition[14] to support the movement[15]. Under this pressure, the FCC has been conducting experiments to verify the feasibility of this proposal since about 2006. Before meeting to vote on whether the FCC will authorize the use of white spaces, FCC Chairman Kevin Martin showed a positive attitude[16], and the FCC meeting on November 4, 2008 unanimously approved the use of white spaces[17]. The new use includes unlicensed communication devices with small antenna power. This decision will encourage many American companies to begin developing the use of white spaces.

In Europe, radio wave resources which become available by digitalizing television broadcasting are called "digital dividends." The European countries are interested in how to use these digital dividends.

The most active among them is the United Kingdom. OFCOM (Office of Communications) published a report supporting the active use of digital dividends[18]. Europe adopted a system of Digital Video Broadcasting for Terrestrial (DVB-T) for digital broadcasting. The UK is standardizing a technology to more effectively use frequencies called DVB-T2 (transmitting highly compressed moving image information), and has also begun the development of broadcast equipment[19]. DVB-T2 assumes about 45% higher capacity than that of DVB-T[20]. The concept is that, since the system produces some room in frequency space, licensed broadcasters may use the remaining frequencies for other applications.

Among the European countries which are building DVB-T broadcast networks, Germany, France, and Spain have adopted the SFN in order to conserve radio wave resources. Thus, the European countries

are developing policies and measures to actively use digital dividends.

7 Conclusion and suggestions

Television band frequency sharing will be put to practical use in the near future. The new narrow-area digital community broadcasting and the Shimane Ubiquitous Project are pioneering activities from a global perspective. Accumulating such experimental experience before other countries will secure a position of technical leadership for Japan. From this viewpoint, the policy of the Ministry of Internal Affairs and Communications which established the Specified Ubiquitous Districts in order to conduct experiments on new radio wave applications can be highly praised.

For effective use of radio waves, further promotion of these moves is necessary. Revisions of the regulations specified in the ordinances and notices concerned should be accelerated so that official licenses can be issued to qualified applicants rather than temporary and experimental ones, under the condition that they do not cause interference with existing television broadcasting. Particularly for MediaFLO, experiments should be conducted to determine the interference protection ratio.

In future research and development on advanced frequency sharing technologies, efforts should be made in the direction outlined in Chapter 5.

Moreover, it is not inevitable that frequency sharing will be limited in television bands. Allowing other frequency bands to share frequencies, so long as they do not disturb existing applications, will bring our society closer to a real ubiquitous society which is capable of using radio waves more effectively.

The new broadcasting service oriented toward "responding to the needs of the smallest possible number of people in the narrowest possible area" was proposed based on an idea completely different from the concept of conventional television broadcasting. Internet websites have given an unprecedentedly number of people the opportunity to transmit information. Likewise, some say that this new type of broadcasting represents the "webification of broadcasting."[21] It can also be considered a service which is closer to the integration of communications and broadcasting. The culture of using cell phones for email appeared in Japan and spread to the world. If the "webification of broadcasting" spreads to the world in a similar manner, it will create new business opportunities for Japanese industry.

Acknowledgments

The author interviewed Yoshiyuki Takeda of NTT DOCOMO, Junichi Kato of Area Portal, Mamoru Yano of MediaScope, Jun Yamada of Qualcomm Japan, and many other persons concerned. He would like to express his deep gratitude to them.

References

[1] Ministry of Internal Affairs and Communications, Partial Report from Telecommunications Council and the Results of the Call for Opinions on the Technical Conditions for the Effective Use of Radio Waves in VHF/UHF Bands, July 2007. (Japanese) : http://www.soumu.go.jp/s-news/2007/070627_4.html

[2] Hajime Yamada and Akihiro Fujii, "Technology and Policy Trends in Frequency Sharing," Science and Technology Trends, No. 44, November 2004. (Japanese) which was published in English in Science & Technology Trends Quarterly Review, No.15, April 2005

[3] Information and Communications Statistics Database compiled by the Ministry of Internal Affairs and Communications. (Japanese) : http://www.johotsusintokei.soumu.go.jp/field/index.html

[4] (Former) Ministry of Posts and Telecommunications, Reports from Telecommunications Council on Terrestrial Digital Television Broadcasting, May 24, 1999. (Japanese) : http://www.soumu.go.jp/joho_tsusin/pressrelease/japanese/housou/990524j701.html

[5] For example, Hiroshi Murakami, Taiji Murakami, and Tomoaki Kaneko, "Transmission Network Systems for Terrestrial Digital Broadcasting," Toshiba Review, Vol. 62, No. 2 (2007), 58. (Japanese)

[6] Ministry of Internal Affairs and Communications, Channel Plan for Terrestrial Television Broadcast Stations (as of June 30, 2008). (Japanese) :
http://www.soumu.go.jp/s-news/2008/pdf/080331_11_bt1.pdf

[7] Ministry of Internal Affairs and Communications, "On the Reports on 'Survey Study on Terrestrial Digital Broadcast Transmission Equipment in Disadvantaged Areas,'" March 2007. (Japanese) :
http://www.ktab.go.jp/new/2006/0315-4.pdf

[8] Yokosuka Research Park, "Report on the First General Assembly on 'Preparatory Committee for New Narrow-area Digital Community Broadcasting." (Japanese) :
http://www.yrp.co.jp/yrprdc/nb_digital-bc/n_digil-bc.html

[9] Website of Area Portal (Japanese) : http://www.areaportal.co.jp/

[10] Website of MediaScope. (Japanese) : http://www.ms-ltd.co.jp/

[11] Ministry of Internal Affairs and Communications, Comparison of Terrestrial Digital Broadcasting Systems for Mobile Terminals. (Japanese) :
http://www.soumu.go.jp/joho_tsusin/policyreports/chousa/mobile_media/pdf/080218_2_si1.pdf

[12] Gregory Staple and Kevin Werbach, "The End of Spectrum Scarcity," IEEE Spectrum, March 2004. :
http://www.spectrum.ieee.org/mar04/3811

[13] Website of Wireless Innovation Alliance. : http://www.wirelessinnovationalliance.com/

[14] Website of Public Interest Spectrum Coalition (PISC). :
http://www.newamerica.net/people/public_interest_spectrum_coalition_pisc/recent_work

[15] Kevin Werbach, "From Broadcasting to Network Age: Lessons from the Digital TV Transition in the United States," the First 2008 Symposium of Information and Communications Policy Forum, July 25, 2008. The video of Mr. Werbach's lecture can be viewed on the following websites :
http://jp.youtube.com/watch?v=Z25E-VupMj4&fmt=18
http://jp.youtube.com/watch?v=jn5yD5s8Occ&fmt=18
http://jp.youtube.com/watch?v=-Do4_HgCPaw&fmt=18

[16] "FCC's Martin backs 'white space' use, eyes Nov. vote," CTIA Smartbrief, October 2008. :
http://www.smartbrief.com/news/ctia/storyDetails.jsp?issueid=1F427BDB-88EE-42C4-AA81-DF0ED0563ED5&copyid=2C430C09-EF44-476C-96DD-6865039640E6&brief=ctia&sb_code=rss&&campaign=rss

[17] Federal Communications Commission, "FCC adopts rules for unlicensed use of television white spaces," November 2008. :
http://hraunfoss.fcc.gov/edocs_public/attachmatch/DOC-286566A1.doc

[18] OFCOM, Digital Dividend Review, December 2007. :
http://www.ofcom.org.uk/consult/condocs/ddr/statement/

[19] OFCOM, "Broadcasters awarded licenses for HD programmes on Digital Terrestrial TV," October 2008. :
http://www.ofcom.org.uk/media/news/2008/10/nr_20081017

[20] DVB Project, "2nd Generation Terrestrial: The world's most advanced Digital Terrestrial TV System," August 2008. :
http://www.dvb.org/technology/fact_sheets/DVB-T2%20Fact%20Sheet.0808.pdf

[21] Hisashi Kamio, "MediaFLO aims at the Webification of Broadcasting–An Interview with Qualcomm Japan Chairman Jun Yamada," ITmedia, September 2008. (Japanese) :
http://plusd.itmedia.co.jp/mobile/articles/0809/17/news038.html

Profile

Hajime YAMADA
Affiliated Fellow,NISTEP

Professor, Department of Social Economic Systems, Faculty of Economics, Toyo University
Dr. Yamada holds a Master's Degree and Dr. Eng. from Keio University. He also holds a Master's Degree in Management of Technology from the Massachusetts Institute of Technology. After direct involvement in research at NTT, Dr. Yamada began working in management, including creating and proposing research strategies. He has been a professor at the Toyo University Faculty of Economics since 2002.

(Original Japanese version: published in December 2008)

3

Overcoming the Language Barrier with Speech Translation Technology

SATOSHI NAKAMURA
Affiliated Fellow

1 Introduction

The global, borderless economy has made it critically important for speakers of different languages to be able to communicate. Speech translation technology – being able to speak and have one's words translated automatically into the other person's language – has long been a dream of humankind. Speech translation has been selected as one of the ten technologies that will change the world. There are especially high hopes in Japan for a speech-translation system that can automatically translate one's everyday speech; the Japanese find the acquisition of foreign languages extremely difficult due to such factors as Japan's geographical conditions and the isolation of the Japanese language. As Japan and the Japanese become increasingly international, such speech-translation technology would be a great boon to the nation.

Automatic speech translation technology consists of three separate technologies: technology to recognize speech (speech recognition); technology to translate the recognized words (language translation); and technology to synthesize speech in the other person's language (speech synthesis). Recent technological advances have made automatic translation of conversational spoken Japanese, English, and Chinese for travelers practical, and consecutive translation of short, simple conversational sentences spoken one at a time has become possible (Japanese-to-English translation is equivalent to a TOEIC score of 600 or more).

There are still many challenges to overcome, however, before this technology can become viable. Some examples are the need to support more languages, and for automatic acquisition of place names, personal names, and other proper nouns. Technology should also be established for simultaneous interpretation, where a continuous stream of speech is translated. The individual technologies used in speech translation also have a wide range of applications, including speech information retrieval, interactive navigation, dictation, and summarization and archiving, and new uses are expected to emerge as well.

This report starts by affirming the significance of speech-translation technology, and providing an overview of the state of research and development to date, and the history of automatic translation technology. It goes on to describe the architecture and current performance of speech translation systems. It also touches on worldwide trends in research and development, describes the practical application of speech translation technology, its extension to Asian languages, and describes efforts to standardize interfaces. It concludes by summarizing the challenges and future prospects for speech-translation technology, and proposes a way forward to resolve these challenges in the promotion of this technology.

2 The history of speech translation technology

2-1 The significance of speech translation research and history to date

Speech translation is a technology that translates spoken language into speech in another language. Speech-translation technology is significant because it enables speakers of different languages from around the world to communicate, erasing the language divide in global business and cross-cultural exchange. Achieving speech translation would have tremendous scientific, cultural, and economic value for humankind. The article "10 Emerging Technologies

That Will Change Your World" in the February 2004 issue of An MIT Enterprise Technology Review lists "Universal Translation" as one of these ten technologies. The article showcases a number of translation technologies, but focuses on speech-translation technology.

Speech translation first grabbed attention at the 1983 ITU Telecom World (Telecom '83), when NEC Corporation performed a demonstration of speech translation as a concept exhibit. Recognizing that many years of basic research would be required to implement speech translation, the Advanced Telecommunications Research Institute International (ATR) was subsequently founded in 1986, and began a project to research speech translation. Researchers from a wide range of research institutes both in Japan and internationally joined this project[1]. In 1993, an experiment in speech translation was conducted linking three sites around the world: the ATR, Carnegie Melon University (CMU), and Siemens. After the start of ATR's project, speech translation projects were started around the world. Germany launched the Verbmobil project; the European Union the Nespole! and TC-Star projects; and the United States launched the TransTac and GALE projects. The GALE project was started in 2006 to translate automatically Arabic and Chinese into English. The goal of this project is to automate the extraction of vital multilingual information that up until then had been performed by humans; the project architecture consists of a batch text-output system. In contrast, the objectives of the ATR and NEC are speech translation enabling face-to-face and non-face-to-face cross-language communication in real time. Online speech-to-speech translation is thus an integral component of this research, and immediacy of processing is a key factor.

Speech translation integrates three components: speech recognition, language translation, and speech synthesis. Each of these technologies presents its own difficulties. In particular, a requirement of this technology is recognizing and translating spoken language; this is much more difficult than translating text, because spoken language contains ungrammatical, colloquial expressions, and because it does not include punctuation like question marks, exclamation marks, or quotation marks. Mistakes in speech recognition also cause major translation errors. Consequently, researchers have chosen a development technique of increasing accuracy to a usable level by initially restricting the system to relatively simple conversation, rather than supporting all forms of conversation from the start. Table 1 shows the history of speech-translation technology. Research and development has gradually progressed from relatively simple to more advanced translation, progressing from scheduling meetings, to hotel reservations, to travel conversation. Moving forward, however, there is a need to further expand the supported fields to include a wide range of everyday conversation and sophisticated business conversation.

Table 1 : Trends in the Research and Development of Speech Translation

Research Phase	1980s Confirmation of Feasibility	1990s Extension of Technology	2000s Attempts at Practical Systems
Fields	Simple reservations (ATR-phase 1)	Reservations and scheduling (ATR-phase 2, Verbmobil)	*Everyday travel conversation (ATR-phase 3) *Translation of keynote speeches (TC-Star) *Conversation for military use (TranTac) *Intelligence collection (Gale)
Linguistic features	Grammatically correct expressions	Everyday expressions that may be context-dependent or ungrammatical	Expressions including a wide range of topics and proper nouns
Phonological features	Clear pronunciation	Unclear pronunciation	Audio including background noise
Translation method	Rule-based translation Translation using artificial intermediate language	Example-based translation Translation using English as intermediate language	Statistically-based translation Direct translation of multiple languages

Note: ATR-phase 1: 1986 to 1992; ATR-phase 2: 1993 to 1999; ATR-phase 3: 2000 to 2005. For other projects, refer to the text.

Prepared by the STFC

2-2 The history of automatic translation

Of the three components of speech translation, recent advances in text translation technology have made a major contribution to the realization of automatic speech translation technology. Research into text translation technology has a long history going back more than half a century.

Warren Weaver of the Rockefeller Foundation advocated research into automatic translation technology for text in 1946, shortly after the birth of the first computer. At the time, the Rockefeller Foundation had a huge influence on the United States' science and technology policy. Then in 1953, Georgetown University and IBM began joint research of automatic translation using the 701 (the first commercial computer, developed by IBM). In 1954, the world's first automatic translation system was built on this computer, demonstrating the possibility of translation from Russian to English. Consisting of a dictionary of 250 terms and 6 rules, the translation capabilities of this system were extremely limited, but the demonstration had a huge impact on society. People at the time felt that the language barrier would soon be knocked down. Subsequently, as part of its response to the shock of the Sputnik launch, a whopping $20 million were invested into research on automatic translation by US government.

In 1965, however, the Automatic Language Processing Advisory Committee (ALPAC) presented a grave report to the US National Academy of Sciences. The report stated that because automatic translation would not be practical for the foreseeable future, research efforts should instead be directed at language theory and understanding to serve as the underpinnings of this technology. In the US, budgets for automatic translation were subsequently cut, and the focus turned to basic research, with the key concepts being meaning and understanding. One famous result from this time is language understanding using world knowledge by Winograd in 1970. The base knowledge base in this kind of research, however, was insufficient, and it cannot be said to have tied directly into improved performance of automatic translation in a general or practical sense.

Three great waves of technological innovation hit Japan in the 1980s: rule-based translation, example-based translation, and statistically-based translation. In Japan, a project to translate abstracts of the science and technology literature of the Science and Technology Agency (dubbed the Mu project) was successful. As a result, research and development into rule-based automatic translation, based on dictionaries and rules (analytic grammar rules, conversion rules, and generative grammar rules), began to gain popularity. A venture named Bravis launched sales of a commercial translation program. This spurred the commercialization of automatic-translation software by such big-name IT companies as Fujitsu, Toshiba, NEC, and Oki Electric Industry. All of the commercial software packages in the world today, and nearly all of the Web-based software, have this rule-based technology as their cores. Because better and more complete specialized dictionaries were an effective way to improve translation quality, slow but steady efforts have built up to increase dictionary sizes from a few tens of thousands of entries to millions of entries.

Meanwhile, in 1981 professor Makoto Nagao of Kyoto University took a hint from the translation process carried out by humans to propose an example-based translation method using sentences similar to the input sentence and their translations (together called "example-based translations"). This example-based translation, combined with further research at Kyoto University and ATR around 1990, created a second wave that spread from Japan to the rest of the world. This method has been incorporated into some commercial rule-based systems; it is also currently being used as the core method for a Japanese-to-Chinese translation project for scientific and technical publications being led by the National Institute of Information and Communications Technology (NICT).

Then in 1988, IBM proposed a method called statistical machine translation, combining pure statistical processing that excludes grammatical and other knowledge with a bilingual corpus. This method did not get attention for some time, however, for a number of reasons: the paper was difficult to understand, computer performance was lacking, the translation corpora were too small, the method of execution was only published in patent specifications, and it was not effective for languages other than related languages like English and French. Around 2000, however, a new method called phrase-based statistical machine translation was proposed, and buoyed by more complete bilingual corpora and more powerful computers, this created the third major wave.

Today, nine out of ten research papers in the field are on statistically-based translation. It is difficult to tell at this time whether this research domain will continue to grow.

Today, the three waves above are just now overlapping. We have gradually come to learn the strengths and weaknesses of the rule-based, example-based, and statistically-based approaches to automatic translation. The current opinion is that the best performance can be achieved by fusing these three approaches in some way, rather than by using any one of them in isolation. The three methods, however, have a common problem: they all translate at the sentence level. They cannot use contextual information. In other words, they do not make use of the relationships with the surrounding text, and thus cannot ensure cohesion. Statistical machine translation in particular performs automatic translation without analyzing the meaning of the input sentence, and so sometimes generates nonsensical translations.

The method of using example-based and statistically-based methods is called "corpus-based translation,"and this paper primarily presents methods using statistical machine translation. A corpus is a database of text with supplementary linguistic information added, such as pronunciations, part-of-speech information, and dependency information. The next and subsequent chapters primarily describe corpus-based translation methods.

3 Overview of speech translation technology and performance

3-1 Multilingual speech translation processing architecture

Figure 1 shows the overall architecture of the speech-translation system. Figure 1 illustrates an example where a spoken Japanese utterance is recognized and converted into Japanese text; this is then translated into English text, which is synthesized into English speech. The multilingual speech-recognition module compares the input speech with a phonological model consisting of a large quantity of speech data from many speakers (the model consists of the individual phonemes making up the speech utterances), and then converts the input speech into a string of phonemes represented in the Japanese katakana syllabary. Next, this string of phonemes is converted into a string of words written in the Japanese writing system (mixed kana and kanji characters), so that the probability of the string of words is maximized. In this conversion, string of words appropriate as a Japanese utterance is generated based on the occurrence probability of a string of three words using an engine trained on large quantities of Japanese text. These words are then translated by a conversational-language translation module, replacing each Japanese word in the string with the appropriately

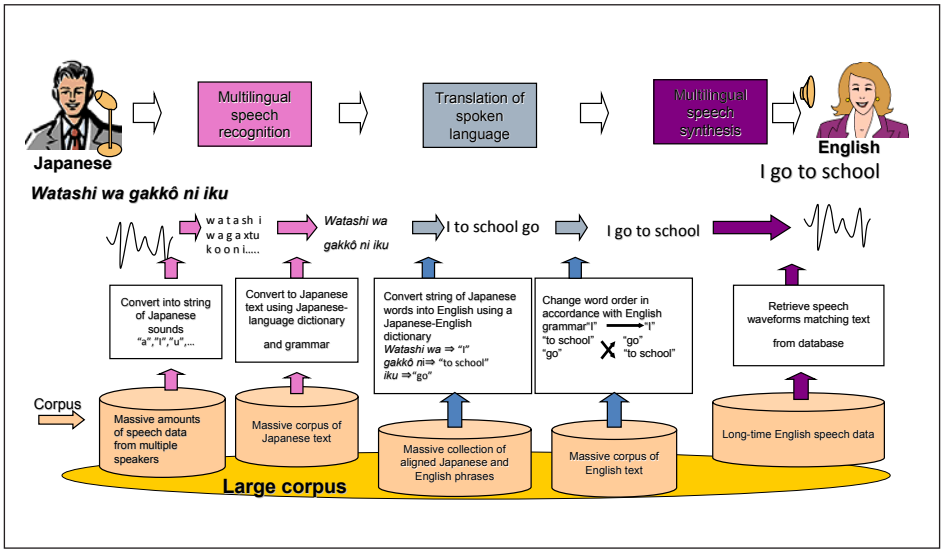


Figure 1 : The mechanism of the speech translation system

Prepared by the STFC

corresponding English word. The order of the English words is then changed. In this procedure, the Japanese words in the string are replaced by English words using a translation model trained on pairs of Japanese-English translations. In order to rearrange the words into a proper English utterance, a string of words appropriate as an English utterance is generated based on the occurrence probability of a string of three words using an engine trained on large quantities of English text. This is then sent to the speech synthesis module. The speech synthesis module estimates the pronunciation and intonation matching the string of English words, selects matching waveforms from a database of long-time speech data, connects them, and performs high-quality speech synthesis. The method of speech recognition and synthesis using statistical modeling and machine learning based on massive speech corpora is called "corpus-based speech recognition and synthesis."

ATR developed its speech-translation system[1,2] by collecting a corpus of general spoken travel conversation, in order to implement speech translation of travel conversation. To date, the project has created a Basic Travel Expression Corpus (BTEC) consisting of 1,000,000 matched pairs of Japanese and English sentences, and 500,000 each of matched Japanese-Chinese and Japanese-Korean pairs. This is the world's largest translation corpus of multilingual travel conversation. The English sentences in the corpus are an average of seven words long and cover such everyday travel-conversation topics as greetings, problems, shopping, transportation, lodging, sightseeing, dining, communication, airports, and business. Below is an example of spoken English translations of a Japanese sentence.

The Japanese is "mado o akete mo ii desu ka"
Here are the corresponding English sentences:

1. may i open the window
2. ok if i open the window
3. can i open the window
4. could we crack the window
5. is it okay if i open the window
6. would you mind if i opened the window
7. is it okay to open the window
8. do you mind if i open the window
9. would it be all right to open the window
10. i'd like to open the window

As can be seen by these examples, the utterances used in speech translation are not complete sentences – they often lack subjects, and no capitalization is used in subjects and proper nouns – and even questions do not have question marks. It is also necessary to handle extremely colloquial expressions.

In addition to BTEC, data collected from a corpus of about 10,000 utterances of dialog recorded under real-life conditions mediated by a speech translation system called Machine Aided Data (MAD) was evaluated; also evaluated was a dataset called Field Experiment Data (FED). This data was collected via a field experiment performed at Kansai International Airport, with the cooperation of the prefecture of Osaka, over a total of five days between December 2004 and January 2005. The data consists of a total of about 2,000 utterances of conversation between foreign-language speakers (39 English speakers and 36 Chinese speakers) and guides at a tourist center mediated by the speech-translation system.

3-2 Comparative study with human speech-translation capability

It is extremely difficult theoretically to evaluate the accuracy of speech translation. If the evaluation of the speech synthesis module is not included, evaluation is made by feeding a number of test sentences into the system, and evaluating the quality of the output. In this sense, the method for evaluating speech translation is essentially the same as that for evaluating automatic text translation. For speech translation, however, the utterances that are evaluated are not strings of text but speech.

Two methods are used to evaluate translation quality: one method where the translations are manually given subjective ratings on a five-point scale, and another that compares the similarity between the output of the system and previously prepared reference translations. A number of rating scales have been proposed for the latter, including BLEU, NIST, and word error rate (WER). Recently, these scales have come to be widely used[4]. Since these results are simple numerical values, it is possible to use them to compare two different systems. What these scores cannot answer, however, is how the system with the higher score will perform in the real world.

A method has been proposed to resolve this issue, by estimating system performance in human terms, estimating the system's corresponding Test of English

for International Communication (TOEIC) score. First, native speakers of Japanese with known TOEIC scores ("TOEIC takers") listen to test Japanese sentences, and are asked to translate them into spoken English. Next, the translations by the TOEIC takers are compared against the output of the speech-translation system by Japanese-English bilingual evaluators. The human win rate is then calculated as the proportion of tests sentences for which the humans' translations are better. After the human win rate has been completely calculated for all TOEIC takers, regression analysis is used to calculate the TOEIC score of the speech-translation system. Figure 2 shows system performance converted into TOEIC scores. When using relatively short utterances like those in basic travel conversation (BTEC), the speech-translation system is nearly always accurate. The performance of the speech-translation system on conversational speech (MAD and FED) is, however, equivalent to the score of 600 (TOEIC) by the Japanese speakers.

Furthermore, performance drops significantly when dealing with long, rare, or complex utterances. There is thus still room for improvement in performance.

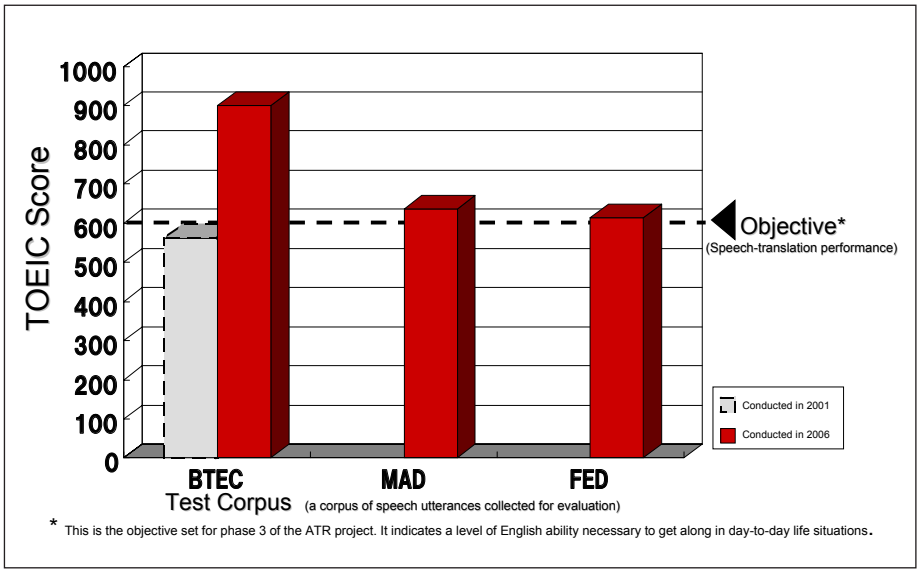


Figure 2 : Example of evaluating the accuracy of speech translation using TOEIC scores

Source : Reference [1]

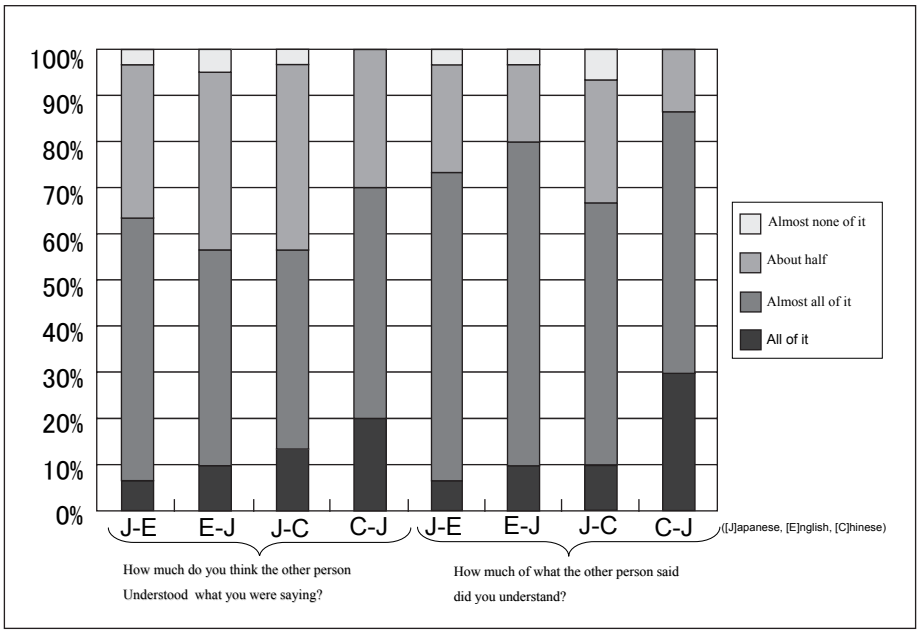


Figure 3 : Evaluation of level of comprehension based on questionnaires

Source : Reference [6]

3-3 Field experiments using speech translation device

A field experiment was conducted in Downtown Kyoto from 30 July to 24 August 2007[6], with the objective of evaluating the characteristics of communication mediated by a stand-alone speech translation device about the size of a personal organizer, as well as evaluate the usability of this device. The field experiment was set up as follows, in order to minimize the restrictions on the test subjects: (1) The people with whom the subjects conversed were not selected ahead of time, in order to collect a diverse range of expressions while using the speech-translation device in realistic travel situations, such as transportation, shopping, and dining. (2) Although the subjects were told the purpose of the dialog ahead of time, no restrictions were placed on the exact destination or proper names of items to purchase. (3) Subjects were allowed to change the topic freely depending on the flow of the conversation. (4) Subjects were allowed to move to different locations as appropriate, in accordance with the task. (5) No time limit was placed on single dialogs.

In the case of transportation, the objective was considered to have been met if the subject was able to obtain information about the destination or to actually travel there. For shopping and dining, the objective was met if the subject completed the purchase of the article or the meal and received a receipt.

In addition to quantitative evaluations of speech recognition rates, dialog response rates, and translation rates, the experiment also evaluated the level of understanding based on questionnaires. As shown in Figure 3, in the evaluation of the level of understanding of 50 native English speakers, about 80% said that the other person understood nearly everything that they said, and over 80% said they understood at least half of what the other person said. This result suggests that the performance of speech-translation devices could be sufficient for communication.

4 Worldwide trends in research and development

International evaluation workshops give a strong boost to the development of speech-translation technologies. An international evaluation workshop is a kind of contest: the organizers provide a common dataset, and the research institutes participating in the workshop complete, creating systems that are quantitatively evaluated. The strengths and weaknesses of the various proposed algorithms are rated from the results of the evaluation, and the top algorithms are then widely used in subsequent research and development. This allows research institutes to perform research both competitively and cooperatively, promoting efficient research. Some representative examples of international evaluation workshops are presented here, describing automatic evaluation technologies that support competitive research styles via evaluation workshops.

(a) The International Workshop on Spoken Language Translation (IWSLT)[7] is organized by C-STAR, an international consortium for research on speech translation including ATR in Japan, CMU in the United States, the Institute for Research in Science and Technology (IRST) in Italy, the Chinese Academy of Sciences (CAS), and the Electronics and Telecommunications Research Institute (ETRI) in Korea. The workshop has been held since 2004. Every year, the number of participating institutes increases, and it has become a core event for speech translation research. The subject of the workshop is speech translation of travel conversation from Japanese, Chinese, Spanish, Italian, and other languages into English. Two distinguishing features of the IWSLT are that it is for peaceful uses (travel conversation) and that the accuracy of the translation is fairly good, because it is a compact task.

(b) Global Autonomous Language Exploitation (GALE) [8] is a project of the US Defense Advanced Research Projects Agency (DARPA). It is closed and non-public. US $50 million are invested into the project per year. The purpose of the project is to translate Arabic and Chinese text and speech into English and extract intelligence from them. A large number of institutions are divided into three teams and compete over performance. The teams are run in units of the fiscal year in which the targets are assigned, and every year the performance is evaluated by outside institutions. In the United States, research on automatic translation is currently strongly dependent on DARPA budgets, and the inclinations of the US Department of Defense are strongly reflected.

Methods for evaluating translation quality have become a major point of debate at these workshops.

There are various perspectives on translation quality, such as fluency and adequacy, and it has been considered a highly knowledge-intensive task. A recently proposed evaluation method called BLEU is able to automatically calculate evaluation scores with a high degree of correlation to subjective evaluations by humans. This makes it possible to develop and evaluate systems repeatedly in short cycles, without costing time or money, which has made translation research and development much more efficient[4].

5 Practical applications of speech translation technology

The improved processing power and larger memories of computers and more widespread networks are beginning to make it possible to implement portable speech translation devices. Advances are being made in the development of standalone implementations in compact hardware, and distributed implementations connecting mobile phones and other devices to high-performance servers over a network.

It is not feasible to implement the standalone method on a computer that is carried around, due to such issues as size, weight, and battery lifetime. There is also expected to be demand in situations where wireless and other infrastructure is not available. In light of these issues, efforts are being directed toward the commercialization of dedicated mobile devices with built-in speech-translation functionality. In 2006, NEC developed the world's first commercial mobile device (with hardware specifications of a 400-MHz MPU and 64 MB of RAM) with onboard Japanese-to-English speech translation.

Meanwhile, in November 2007 ATR developed a speech translation system for the DoCoMo 905i series of mobile phones as a distributed implementation using mobile phones and network servers. The system, called "shabette honyaku" (see Figure 4), was released by ATR-Trek, and is the world's first speech translation service using a mobile phone. Then in May 2008, a Japanese-to-Chinese speech-translation service was begun on the DoCoMo 906i series. Figure 5 shows the architecture of the speech recognition module used in the distributed speech translation. The mobile phone (front end) performs background noise suppression, acoustic analysis, and ETSIES 202 050-compliant encoding[9], and sends only the bit-stream data to the speech recognition server. The speech recognition server (back end) then expands the received bit-stream, performs speech recognition, and calculates word reliability. One of the benefits of using this system architecture is that it is not bound by the information-processing limitations of the mobile phone, making large-scale, highly precise phonological and linguistic models to be used. Since these models are on the server and not the mobile phone, they are easy to update, making it possible to keep them up to date at all times. The system is already in wide use: as of June 2008, there have been a cumulative total of over five million accesses.

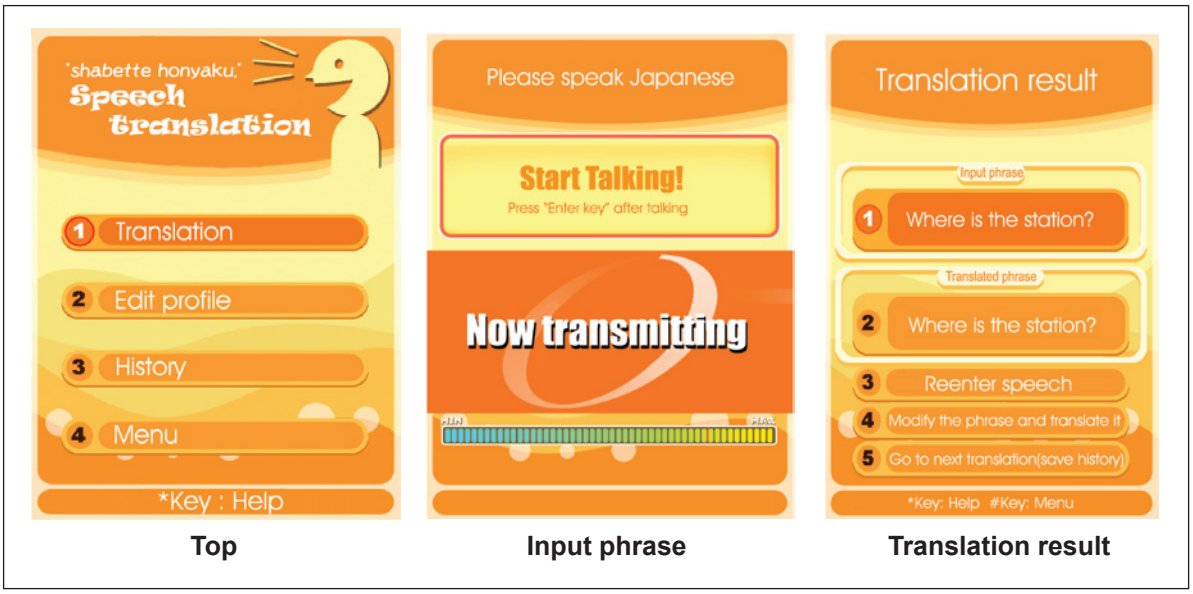


Figure 4 : The world's first speech translation service using a mobile phone

Source : ATR-Trek Co., Ltd.

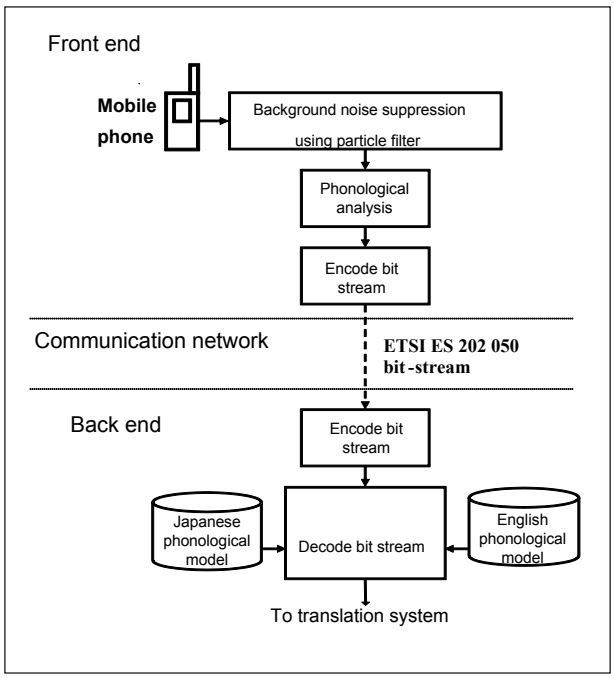


Figure 5 : Architecture of speech recognition module used in distributed speech translation

Prepared by the STFC based on the references [9,12]

6 Standardization for support of multiple languages in speech translation

As speech translation technology overcomes linguistic barriers, it would be preferable for researchers and research institutions from many different countries to research it jointly. The C-STAR international consortium for joint research of speech translation, in which ATR and CMU play a central role, has been quite active in international joint research.

Meanwhile, the foreign travel destinations of Japanese people – whether for tourism, emigration, or study abroad – are becoming more diverse, and people from a large number of countries are coming to Japan in increasing numbers for tourism, study, and employment. These and other changes are heightening the need for means of interaction with people from non-English speaking countries.

In particular, Japan is strengthening its social and economic ties in the Asian region including Russia, and enhancing mutual understanding and economic relations at the grassroots level has become a key challenge. Relations with the rest of Asia are more vital to Japan than ever before. Consequently, rather than English, Japan needs to be able to get along in the languages of its neighbors, such as Chinese, Korean, Indonesia, Thai, Vietnamese, and Russian – languages that until now have not been widely taught or spoken in this country.

Against this backdrop, A-STAR was founded as a speech translation consortium for creating the basic infrastructure for spoken language communication overcoming the language barriers in the Asia-Pacific region. Rather than the research and development of technology proper, however, the consortium's objective is to establish an international joint-research organization to design formats of bilingual corpora that are essential to advance the research and development of this technology, to design and compile

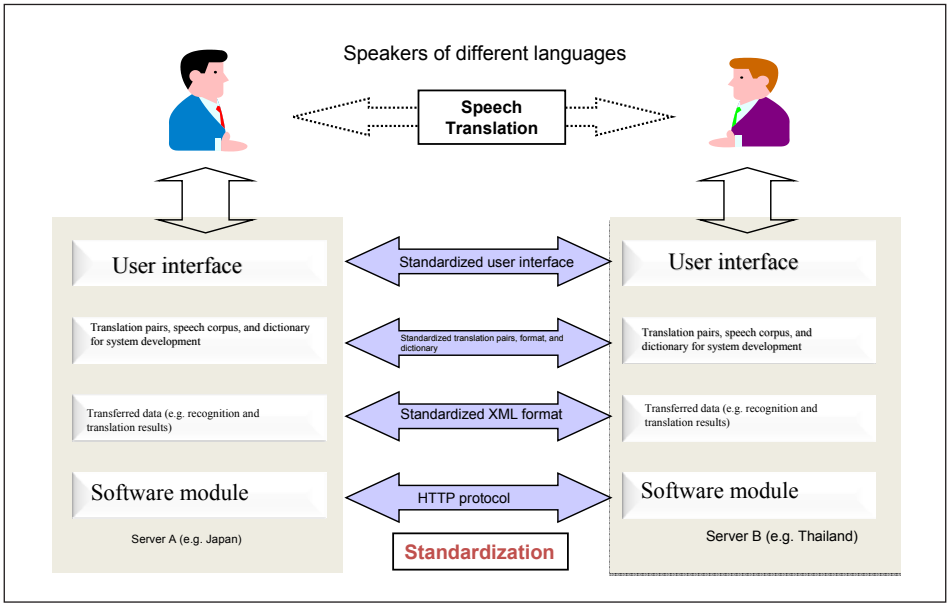


Figure 6 : Illustration of speech translation standardization

Source : Reference [12]

basic bilingual corpora between Asian languages, and to standardize interfaces and data formats to connect speech translation modules internationally, jointly with research institutions working in this field in the Asia-Pacific region. The consortium's activities are contracted as research by the Asia Science and Technology Cooperation Promotion Strategy, which is a project of the special coordination funds for promoting science and technology. This project has further been proposed and adopted as APEC TEL (Telecommunications and Information)[10] project. It is also moving to create an expert group in the APT ASTAP (Asia-Pacific Telecommunity Standardization Program) in order to create a draft of the standardized interface and data formats for connecting speech-translation modules[11]. Figure 6 illustrates the standardized connections being considered in this project. This will standardize the interfaces and data formats of the modules making up the speech translation architecture, in order to enable their connection over the Internet. It is also necessary to create common speech-recognition and translation dictionaries, and compile standardized bilingual corpora. The basic communication interface will be Web-based HTTP 1.1 communication, and a markup language called STML (speech translation markup language) is currently being developed as the data format for connecting applications[12].

7 Challenges and future prospects of speech translation technology

7-1 Challenges for the development of speech translation

As described above, speech translation is a technology enabling communication between speakers of different languages. There are still many research challenges to overcome, however: in particular, there is great speaker dependency and diversity of expression; additionally, new words and concepts are constantly being created in accordance with changes in society. Speech translation technology is currently at the level of simple utterances of about seven words in length, such as travel conversation. Consequently, there are still many unsolved challenges before speech translation will be capable of handling long, complex speech such as a newspaper or lecture. Below are listed some of the immediate technical challenges.

1) Evaluating and Improving Usability in Practical Applications

Human speakers have many inherent differences. People have many differences in speaking style, accent, and form of expression. Speech translation must aim to suppress variations in performance due to these differences, and provide the same high level of performance for all users. Additionally, acoustic noise, reverberation, and speech by other speakers have a huge impact during real world use. Measures to remedy these external factors are also extremely vital. Meanwhile, from the standpoint of usability as a communication tool, it is essential to further reduce the time from speech recognition to translation to speech synthesis. When speech translation is used, the user does not understand the translation language. For this reason, there are no techniques to check whether a translation is correct. A method must thus be provided for the user to check whether the translation is correct, by such means as translating it again back into the user's language, or back-translating it. When considering it as a tool for gathering information while traveling, it is also essential to at the same time provide a means to gather information via the Internet in multiple languages, not only by asking people.

These challenges require field testing and technology development to be performed in parallel, as well as a growth loop of data collection, improving performance, improving usability, and providing trial service.

2) Support for multiple languages

Although English is becoming the de facto worldwide lingua franca, what is needed is not a system that will translate into only English, but one that will translate directly into the 6,000 languages said to exist on our planet today. Multilingual speech translation requires a system of speech recognition, translation, and speech synthesis for each of these languages. In other words, massive speech corpora, bilingual corpora, and text corpora are required for each of these languages. The collection of speech corpora in particular is extremely expensive. This type of technology could also have great value in the sense of preserving languages in a process of decline and extinction.

3) Standardization for the connection of speech translation worldwide via the network

Module connections are also being standardized in the Asia-Pacific region. Moving forward, it will

be necessary to advance standardization for wide international connectivity, and the development of a joint-research structure.

4) Relaxing of copyright to enable example translations to be used via the web

The development of speech translation technology requires a text corpus of the source language, a text corpus of the translation language, a bilingual corpus of translations between the two languages, and speech corpora. It is extremely expensive to create and collect these corpora using conventional methods. One method that is currently gaining attention is collecting data from the Web via the Internet, which continues its explosive growth. For example, the secondary use of news and other media published in multiple languages would be an effective way to improve the performance of speech translation. As of this time, however, copyright issues have not been resolved.

5) Using the latest proper nouns based on the user's current location

There are huge numbers of proper names of people, places, and things. Incorporating all of these proper nouns into the speech-translation system at the same time would be nearly impossible, both in terms of performance and time. It would therefore be efficient to automatically acquire proper nouns corresponding to the user's location using GPS or the like, and perform speech recognition, translation, and speech synthesis tailored to that location.

7-2 Research and development roadmap

Figure 7 shows the history of speech translation to date, and indicates future directions of research and development. In 2010, an international research consortium on Asian languages plans to prototype speech translation via the Internet.

It is conjectured that the international research consortium will come out with a prototype including Western European languages and with greater standardization of interfaces by around 2015. Japan's Project to Accelerate Benefits to Society (described in the next section) plans to establish technology for networked speech translation by 2012, after various field testing. In the mid to long term, speech translation capable of continuous simultaneous interpreting of business and lectures is expected to be available by around 2015, and by 2025, multilingual simultaneous interpretation is expected to be available that is capable of contextual awareness and summarization, gradually bringing us closer to the dream of simultaneous interpretation.

8 Japanese policy regarding speech translation

8-1 The cabinet office project to accelerate benefits to society

"Creating spoken-communication technology to overcome the language barrier"

The speech translation project of the Japanese Ministry of Internal Affairs and Communications has been set as the Cabinet Office Project to Accelerate Benefits to Society. This project has been active since fiscal 2008. Amidst increasing globalization, the goal of the project is to deepen mutual understanding between nations through direct international communication at the level of individual citizens. The project aims to create an automatic speech translation system enabling Japanese people to break the language barrier and speak and interact directly with people from other countries in the Asia-Pacific region and elsewhere. Giving consideration to current user needs and the technological advances that can be expected over the next five years or so, the project is planning and advancing field testing of such applications as tourism, shopping, and international-exchange events. After the project ends, it will quickly be tied into commercial services in industry, accelerating the benefit to society derived from the results of the project.

The project is developing technology to distribute proper nouns like place names and personal names, as well as translation knowledge corresponding to a wide range of topics, over the network. It is working to establish network-based speech-translation technology combining this network with translation devices. It is also working to make innovations in speech-translation communication more visible, in order to help popularize it and facilitate the advent of practical/commercial systems. This effort to make the technology more visible is based on the awareness that the mismatch between developers and users inhibits practical/commercial applications. It thus aims to publicize progress in technology development as appropriate, in order to enable users to envision themselves using it (e.g. what kind of situations they

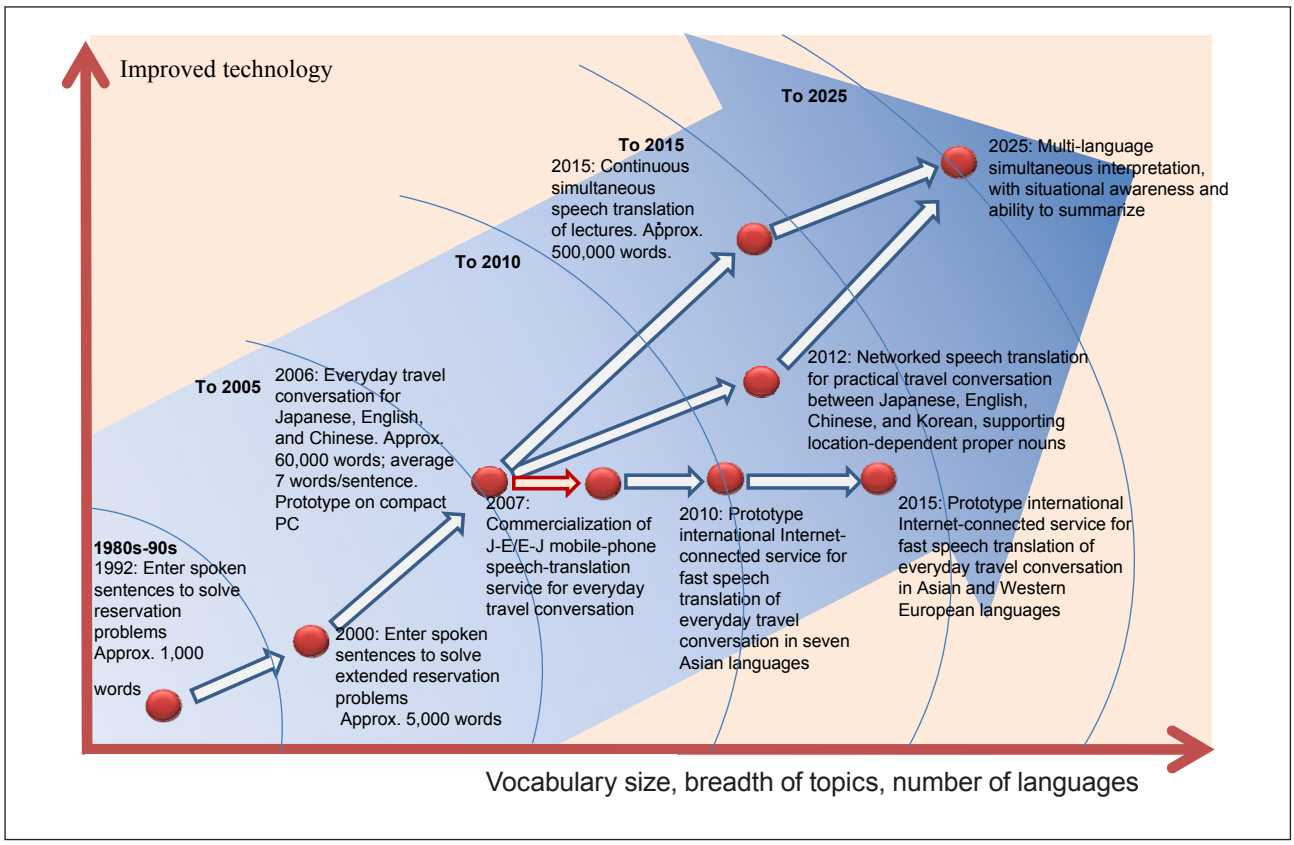


Figure 7 : Trend forecast for research and development of speech translation technology

Prepared by the STFC

could use it in). It will then build functionality and user interfaces suited to these types of use, and repeatedly field test it in concrete situations. By repeatedly validating the technology through close collaboration with both developers and users, the project will play a vital role in a grassroots international exchange by enabling ordinary travelers to enjoy overseas travel with almost no inconveniences in Japan and the English and Chinese-speaking worlds. The project is also expected to create new business and promote local industry and development.

8-2 Special ubiquitous zone

The Special Ubiquitous Zone Project was created based on the Ministry of Internal Affairs and Communications' "ICT Innovation Promotion Program" and "ICT International Competitiveness Enhancement Program." Planned to run for three years starting in fiscal 2008, the project's aim is to support the development and field testing of innovative services. The Special Ubiquitous Zone project was adopted from a proposal by a consortium of eight corporations (Intage Inc.; Toei Kyoto Studio Co., Ltd.; NICT; ATR; JTB Business World Tokyo Corp.; Willcom, Inc.; and NEC), represented by Kyoto Industrial Support Organization 21. This project is being conducted in collaboration with the prefecture of Kyoto, to develop a mobile-phone service to provide market surveys, multilingual translation, and tourist information targeting foreign visitors to Kyoto, and to field test ubiquitous multifunctional servers supporting next generation PHS, achieving sophisticated mobile communication to popularize this widely in tourist areas. Another aim is to improve the satisfaction of foreign travelers and promote the tourism industry. In addition to multilingual translation, the project is developing services that are easy for souvenir shops and restaurants in tourist areas to introduce and use, by utilizing such leading-edge technologies as wearable video delivery servers supporting next generation PHS. The project can also take advantage of the fact that it will be used in a special zone to eliminate copyright and other issues in corpus development.

9 Conclusion

Advances in speech and language research have brought speech translation close to the practical level for simple topics where there is a relatively

clear value of use. At the current level, however, speech translation has only reached the stage of creating the core technologies. In order to achieve more sophisticated speech translation, research and development should be further accelerated. Below are some points that should be the subject of focus moving forwards.

Firstly, one distinctive feature of corpus-based technologies is that they improve with use. It is consequently vital to secure opportunities for field and social testing, and to actively use developed technologies. Events like the Olympics and World Fairs, where speakers of many different languages can be expected to attend, are an ideal opportunity to field test multilingual speech-translation technology. It is thus vital to utilize these opportunities to advance the technology. NICT conducted a monitor experiment at the Beijing Olympics, chiefly targeting travelers from Japan. It developed a speech translation system supporting proper nouns in the city of Beijing, and had monitors use speech translation devices to communicate in the city, using the devices for such purposes as transportation, sightseeing, and shopping. A questionnaire was used to survey users' satisfaction with the service, helping to bring speech translation technology a step closer to viability in practical use.

If Japan wishes to become a major tourist destination, then an effective measure could be to provide continuous tourism information services to foreign tourists, using spoken language translation. Meanwhile, the numbers of foreign residents and workers in Japan are increasing, and multilingual speech translation could be an essential means of communication for local governments, medical facilities, police, and education. Even if there are interpreters present, it should help reduce interpreting costs. If all of these systems are used separately, however, then only fragmentary knowledge will be accumulated, making it inefficient in terms of feedback for research and development. It will probably be necessary for national and local government and the private sector to form a cooperative framework in order to improve efficiency. For example, it could be effective to distribute compact translation devices to public institutions where their need is foreseen, and loan these devices to foreign workers and tourists free of charge.

Secondly, speech translation is a technology that translates spoken words in different languages. Although translation into English is of course vital, it will also be highly significant if speech translation can work directly between Japanese and the native languages of many different countries. For this reason, it is vital to increase the numbers of languages supported. There are limits to how far this research and development can progress in Japan alone, especially when it comes to collecting corpora. A scheme of collaboration between countries with many different languages is needed; in other words, a mechanism is needed to enable various countries to work in partnership to research speech translation, speech, and language. Creating international spoken language technology research centers and the like as a scheme for collaborative R&D should render feedback from a wide range of research into the collection of speech and dialect data, language structure, and the like.

Thirdly, when many countries begin to actually research and develop speech translation, it will be necessary to standardize the interfaces to connect these various language processing modules. The development of connection methods, data formats, dictionaries, and the like must maintain an eye toward standardization. We must avoid a situation in which each country develops its own system, and the systems are not mutually compatible. Speech translation technology is advanced in Japan, and this country can thus lead other countries with relation to standardization.

Finally, attention must be given to copyright. Speech and language processing require speech and text corpora, and the performance of speech translation depends heavily on the quantity and quality of these corpora. Consequently, the use of corpora of news broadcasts, newspapers, and the Internet is extremely effective. Current copyright law does not take secondary uses such as these types of corpora into account. In order to research and develop new technologies, it will be necessary to revise and administer the law so that it is more flexible. This topic is currently being debated by the Copyright Working Group of the Cultural Council, and a conclusion will be published in the near future. It will be necessary to systematically reorganize the topics for future full-scale speech translation services, and reconsider the response, including service models, after the results of this study are released.

References

[1] S.Nakamura et al,“ATR Multi-lingual Speech-To-Speech Translation System”,IEEE Trans. ASLP, vol.14, no. 2 (2006)

[2] A. Finch et al, “The NICT/ATR Speech Translation System for IWSLT 2007”, IWSLT (2007)

[3] T. Takezawa et al, “Toward a Broad-coverage Bilingual Corpus for Speech Translation of Travel Conversations in the Real World” Proc. LREC (2002)

[4] Yasuda et al,“Automatic Translation Evaluation Technologies and Their Application for the Research and Development of Automatic Translation”, Journal of Japanese Society for Artificial Intelligence, Vol. 23, Issue 1 (2008) (Japanese)

[5] Sugenoya et al,“Proposal for a Method to Evaluate the Speech Translation Capability of Speech Translation Systems and Humans, and Comparative Experiment”, IEICE Transactions J84-D-II, 11 (2001) (Japanese)

[6] Itoh et al, “Field Study and Evaluation of Japanese-English-Chinese Speech Translation Device”, 1-Q-33, Proceedings of the Acoustical Society of Japan Spring Meeting (2008) (Japanese)

[7] IWSLT : http://www.slt.atr.jp/IWSLT2004/

[8] SLTC e-Newsletter,“DARPA’s GALE Program to Get More Challenging in 2007” (Example from GALE Project) :
http://ewh.ieee.org/soc/sps/stc/News/NL0701/NL0701-GALE.htm

[9] ETSIES 202 050 ETSI ES 202 050 v1.1.1 Speech Processing, Transmission and Quality aspects (STQ); Distributed Speech Recognition; Advanced Front-end Feature Extraction Algorithm; Compression Algorithms, ETSI, April 2002.

[10] APEC TEL WORKING GROUP :
http://www.apectelwg.org/

[11] ASTAP :
http://www.aptsec.org/Program/ASTAP/

[12] Kimuraet al,“Considerations for a Communication Interface for a Multilingual Speech Translation Platform”, 3-Q-17, Proceedings of the Acoustical Society of Japan Autumn Meeting (2007) (Japanese)

Profile

Satoshi NAKAMURA
Affiliated Fellow,NISTEP

Executive Researcher, National Institute of Information and Communications Technology, Japan
Project Leader, MASTAR Project
Director, ATR Spoken Language Communication Research Laboratories
ATR Fellow
http://www2.nict.go.jp/x/x162/mastar

Doctor of Engineering. Currently involved in speech and language processing, with a focus on speech translation and speech recognition. Visiting Professor at University of Karlsruhe, Germany and Professor, Keihanna Joint Graduate School.

(Original Japanese version: published in August 2008)

4

Necessity of True Bulk GaN Single Crystal and Trends in Research and Development

SEIJI SARAYAMA
Affiliated Fellow

1 Introduction

Today, silicon (Si) semiconductor devices are used in virtually all electronic equipment, spanning a wide range of applications from consumer products, including various types of home electrical appliances such as personal computers (PC), televisions (TV), and cell phones, automobiles, and others, to industry, where examples include electric trains, plant control equipment, etc. Si semiconductors are used mainly in electronic devices such as memories and CPUs (central processing unit) in the form of MOS type transistors, bipolar transistors, and others.

On the other hand, as direct transition-type semiconductors, compound semiconductors, represented by gallium arsenide (GaAs)[NOTE 1] and indium phosphide (InP),NOTE 1) are suited to use in light-emitting devices, and are applied in optical devices such as laser diodes (LD) and light-emitting diodes (LED). LDs are used in optical communications and optical devices such as compact disk (CD) and DVD players, while LEDs are used in a variety of familiar products such as remote controls for home electrical appliances, which operate in the near infrared waveband, and red LEDs in displays. As other features of the compound semiconductors, these devices have high carrier mobility and low leakage current/low capacity, and are widely applied as high frequency transistors in cell phones, satellite broadcast receivers, etc. Silicon carbide (SiC) and other types are also applied in high output devices.

Among the compound semiconductors, GaN-based semiconductor materials,[NOTE 2] centering on gallium nitride (GaN), are wide bandgap semiconductors of the direct transition type, and are therefore semiconductors which are suitable for use in high speed, high power transistors and light-emitting devices in the ultraviolet-blue-green region, which cannot be realized with Si, GaAs, and similar semiconductor materials. Si, GaAs, and similar semiconductors were an object of intense R&D activity from the 1940s into the 1970s, contributing to practical application of this type. In contrast, virtually no practical applications of GaN-based semiconductors have been realized, as R&D requires much time due to the difficulty of growing crystals. A blue LED (pn junction type) was commercialized using a GaN-based semiconductor for the first time in 1993, which was followed by practical application of a white LED, a blue-violet laser diode (BV-LD), and other devices. White LEDs are currently used in practical applications including backlights for various types of liquid crystal displays, beginning with cell phones, and in flashlights and automobile headlights, and efforts are being made to expand this

[NOTE1]
In this paper, GaAs-based semiconductors include AlGaAs, which is a mixed crystal semiconductor of GaAs and AlAs, as well as InGaAs and others. Similarly, InP-based semiconductors include mixed crystal semiconductors containing InP.

[NOTE2]
GaN is a group III nitride semiconductor. In this paper, the group III nitride semiconductors AlN and InN and their mixed crystal semiconductors are referred to as GaN-based semiconductors, and devices using these materials are referred to as GaN-based semiconductor devices.

type into the large market for general lighting in the future. Ultraviolet-blue LEDs are used in deodorizing devices as a light source for the photocatalyst, while blue-green LEDs are used in traffic lights, various types of indicators, and the like. Among LDs, BV-LDs have been applied practically as a light source for the Blu-ray Disc technology, and as such, are used in recording and playback of high definition video images. In electronic devices, active R&D is also underway aiming at practical application of high speed, high output transistors for use in future cell phone ground stations, switching devices for hybrid car inverters, etc. Thus, GaN-based semiconductor devices are becoming a necessary and indispensable part of our daily lives.

According to the "International Comparison of Science and Technology/Research and Development: Electronic Information and Communications – 2008"[1] prepared by the Japan Science and Technology Agency (JST), Japan's GaN-based semiconductor devices were evaluated as either "extremely advanced" or "advanced" in comparison with those in other countries in all of the categories of research, technology, and industry. However, "maintain the status quo" has been a trend in recent years, whereas there is an "upward tendency" in other countries, particularly in China, Korea, and Taiwan. Thus, these countries are closing the gap with Japan.

In substrate materials, Japan holds the top share in Si semiconductors and compound semiconductors such as GaAs/InP, etc., and possesses international competitiveness in this area. The United States holds the lead only in SiC. On the other hand, as will be discussed in detail in this paper, a manufacturing technology for GaN substrate materials has not been established. Therefore, seizing the initiative in crystal growth technology for GaN, which is one of the few remaining materials for which a substrate manufacturing technology has not been established, is considered to be of crucial significance for future GaN-based semiconductor device technology as a whole.

This paper describes the necessity of research and development of a "true bulk single crystal" of GaN in GaN-based semiconductor devices, and the current status and issues for R&D in this area. As used in this paper, "true bulk single crystal" means a bulk crystal which contains minimal crystallographic defects (dislocations) and enables cutting of arbitrary crystal planes, as can already be obtained with Si, GaAs, and InP.

2 Limits of heteroepitaxial technology in GaN-based semiconductors and necessity of bulk crystal

2-1 Range of application of GaN-based semiconductors

Research and development on GaN-based semiconductor materials has a long history. Like GaAs and other compound semiconductors, R&D on crystal growth was carried out using the vapor deposition method beginning in the 1960s. Improvement of crystallinity by using a low temperature buffer layer in 1986[2] and the discovery of p-type conduction in 1989[3], which can be considered two major topics among the scientific breakthroughs in this field, led to the appearance of a commercial blue LED in 1993. Both of these two scientific breakthroughs were achievements of Prof. Isamu Akasaki of Nagoya University (at the time). Subsequently, great progress was made in the development of crystal growth technologies using the vapor deposition method, and practical applications were realized, centering on LEDs and LDs. These devices are now a familiar part of our daily lives. A comparison of the physical properties of GaN-based semiconductors and other semiconductors, and the devices which utilize those properties, is shown in Table 1.

In the area of light-emitting devices, light-emitting devices in the ultraviolet (UV) to visible light region have been realized, taking advantage of the fact that the GaN-based semiconductors are direct transition-type semiconductors and have a wide bandgap. By combining mixed crystals of GaN, AlN, and InN, it may be possible to realize light-emitting devices from the UV to the infrared (IR) region. This is shown concretely in Figure 1. The potential of GaN-based semiconductors as light-emitting devices covers the wavelength region from 200nm, which is the UV region, to near 1500nm, which is used in optical fibers. Of this, at present, LEDs in the wavelength region of 210-550nm and LDs with wavelengths of 342-488nm have been realized as light-emitting devices at the R&D level. Commercial devices include LEDs in the 365-520nm wavelength region and LDs in the

Table 1 : Physical properties of semiconductor materials and features of GaN-based semiconductors

Semiconductor materia	Si	SiC (4H)	Diam-ond	GaAs	Group III nitride (GaN-based) semiconductors			
					GaN	AlN	InN	Devices utilizing features
Transition	Indirect			Direct				▪ UV-visible light-infrared light-emitting devices ▪ Power transistors ▪ High frequency
Bandgap (eV)	1.1	3.3	5.5	1.4	3.4	6.2	0.6～0.7	
Electron mobility ($cm^2 V^{-1} s^{-1}$)	1500	1000	1800	8500	1200	–	4000	
Insulation breakdown field ($MV\ cm^{-1}$)	0.3	3.0	4.0	0.4	3.3	–	2.0	
Saturation electron velocity ($10^7 cm\ s^{-1}$)	1.0	2.0	2.5	2.0	2.5	2.0	4.2	
Thermal conductivity ($W\ cm^{-1} K^{-1}$)	1.5	4.9	20.9	0.5	2.1	2.9	0.8	

Prepared by the STFC based on Reference [3]

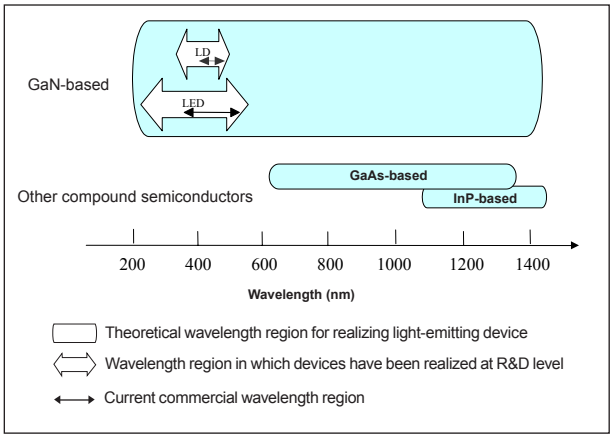


Figure 1 : Suitable wavelengths for light-emitting devices
Prepared by the STFC

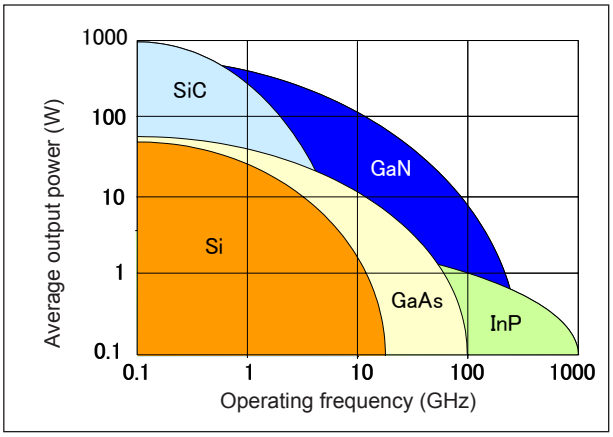


Figure 2 : Applicable regions for high frequency/high output electronic devices
Prepared by the STFC

400-450nm region. However, in spite of the potential of the remaining wavelength region, this region is still unexplored from the practical viewpoint. Future technical innovations will be necessary in order to utilize the remaining wavelength region.

On the other hand, as shown in Table 1, GaN-based semiconductors display excellent electron mobility, dielectric breakdown field, electron saturation velocity, and thermal conductivity properties, and thus have large possibilities as high frequency, high output transistor materials. Figure 2 shows the regions of application of the various types of semiconductors in electronic devices. GaN-based semiconductors have potential in the high frequency, high output region which cannot be realized with other semiconductor materials.

Furthermore, because GaN-based semiconductors do not contain harmful substances like the As in GaAs, they offer high environmental compatibility. Depletion of resources of the elements Ga and N is not a concern, which is also a promising point for future device materials.

2-2 Applications of heteroepitaxial method and its limits

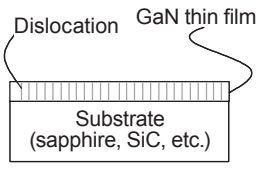
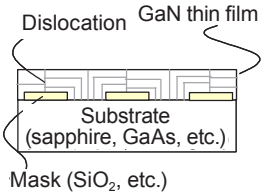
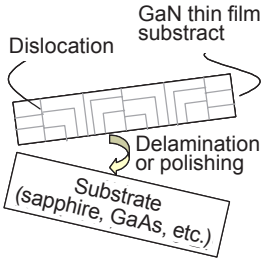
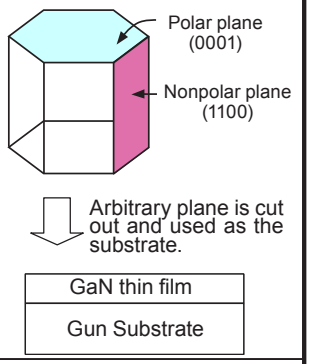
Because Si, SiC, GaAs, and InP all exist in bulk crystal form, substrates of Si, SiC, GaAs, and InP are used in epitaxial growth[NOTE 3] of their crystals, respectively. This type of crystal growth using a substrate of the same material is termed "homoepitaxial growth." However, because it is not possible to obtain bulk crystals of GaN, a substrate of a different material must be used with GaN-based semiconductors when growing crystals for device production. This is a major barrier which makes it impossible for GaN-based semiconductors to demonstrate their outstanding potential. At present, GaN-based semiconductor devices are produced by crystal growth of a GaN based semiconductor thin film on a substrate of sapphire (Al2O3) or SiC. This

type of crystal growth, in which epitaxial growth is performed on a substrate of a heterogeneous material, is termed "heteroepitaxial growth." Table 2 shows the differences between the various types of heteroepitaxial growth and homoepitaxial growth in the methods of forming GaN-based semiconductor crystals.

In crystals obtained by heteroexpitaxial growth, there are problems related to crystallographic defects (dislocations) and the polar plane. First, because the material being grown and the substrate material are different, numerous dislocations occur due to differences in the thermal expansion coefficient and lattice constant.[4] This has an adverse effect on device performance, particularly output and life.[5,6] Moreover, spontaneous polarization occurs due to the relationship of the orientation of the crystal plane with the heterogeneous substrate, making it difficult to improve light emission efficiency.

In obtaining the GaN crystals used practically in LEDs, a GaN-based thin film with a thickness of several μm is grown on a sapphire or SiC substrate by the Metal Organic Vapor Phase Epitaxy (MOVPE) method at a temperature in excess of 1000°C (Type A in Table 2). However, due to the difference in the lattice constants and thermal expansion coefficients of the substrate and GaN-based material, formation of dislocations with a high density of $10^9 cm^{-2}$ or more is unavoidable. This meant that development of Type A crystals to an expanded wavelength region including UV, higher outputs, LD applications, and the like was not possible. Therefore, research and development were carried out with the aim of reducing the dislocation density by heteroepitaxial growth, and it was possible to reduce the dislocation density to the $10^7 cm^{-2}$ level by growing GaN crystals on a sapphire substrate by Epitaxial Lateral Overgrowth (ELO) using Vapor Phase Epitaxy[7] (Type B in Table 2). Subsequently, as a result of further development of this Type B method, it was possible to obtain a GaN thick film substrate with a dislocation density of around $10^5 cm^{-2}$ by first growing a GaN thick film with

Table 2 : Epitaxial crystal growth of GaN and its applications

	Heteroepitaxial			Homoepitaxial
	Type A	**Type B**	**Type C**	**Type D**
				
Merits/ Demerits	· Large area : To inch class · Crystal defects (dislocations) : To $10^9 cm^{-2}$ · Polarity:Spontaneous polarization by C(0001) plane	· Large area : To inch class · Crystaldefects (dislocations) : To $10^9 cm^{-2}$ · Polarity : Spontaneous polarization by C(0001) plane	· Large area: To inch class · Crystal defects (dislocations): To $10^9 cm^{-2}$ · Polarity:Spontaneous polarization by C(0001) plane. · High cost	· High quality · Arbitrary polar plane · Crystal size: mm order
Applica-tions	Near UV to green LED White LED (pseudowhite)	Blue-violet LD (BV-LD; low power)	Blue-violet to blue LD	Not realized

Prepared by the STFC

[NOTE3
"Epitaxial" refers to crystal growth in an upward direction with the crystallographic axes aligned. The term is derived from the Greek epi ("above") and taxy ("ordered"). Epitaxial is also abbreviated "epi."

a thickness of several 100μm on a GaAs substrate[8] or sapphire substrate, followed by delamination of the GaN crystal substrate.[8,9] (Type C in Table 2). Results in which a dislocation density of less than 10^5cm^{-2} was realized locally have also been achieved with this Type C method.[10] Although the width of this low dislocation region is limited to around 500μm, a practical BV-LD (405nm) for Blu-ray Discs has been realized by using Type C GaN as the substrate. Manufacture of these LDs is possible because the active layer (light-emitting region) of the LD is narrow in this application, having a width of several μm, and this can be aligned with the low dislocation region of the Type C substrate. However, this is inadequate for applications which require a size on the mm order, such as LEDs and transistors. It may be noted that Type B and Type C by the VPE method are technologies which were developed in Japan, and at present, Japanese companies hold the top share in substrates for BV-LDs.

On the other hand, because of the difference in lattice constants with the substrate, orientation of the normal heteroepitaxial growth plane is the (0001) plane (C plane). Because the C plane is a crystal plane which possesses polarity, spontaneous polarization occurs. This is a problem because the carrier injected into the device will not contribute efficiently to light emission or transistor operation if spontaneous polarization occurs. From this viewpoint, there are fundamentally limits to improvement in the performance of crystals grown by heteroepitaxial growth.

Sumitomo Electric Industries, Ltd. and Hitachi Cable, Ltd., which are Japanese makers of compound semiconductor substrates, have mass-produced or shipped samples of Type C thick film GaN substrates.[8,9] Of these, Sumitomo Electric's GaN substrate is applied commercially as a substrate for BV-LDs. This maker succeeded in producing a low dislocation density region with a width of several 100μm by a method in which the dislocations are concentrated while forming pits on a GaAs substrate,[10] and realized a substrate size of 2 inches.

A distinctive feature of the Hitachi Cable method is the substrate delamination method. A GaN thin film is grown on a sapphire substrate by the MOVPE method, a titanium (Ti) film is formed on this film by vapor deposition, and a GaN thick film is then grown to a thickness of several 100μm by vapor phase deposition. This causes voids to form in the vicinity of the Ti film, and it is possible to form a GaN thick film substrate by thermal delamination of these voids during the removal operation. Although the average dislocation density is around 10^6cm^{-2}, it is possible to obtain a substrate size of φ3" by this method. However, in both cases, the product displays polarity because the orientation of the principal plane obtained is the C plane (0001), and issues of crystal quality and cost still remain. In particular, the cost of Type C is high, at several \100,000 to 1,000,000 per piece for the φ2" size, because a sacrificial substrate is necessary.

Thus, at present, both of these substrate makers are employing the vapor deposition method, which is a highly realistic method, but this does not mean that a "true bulk single crystal" has been realized. As mentioned previously, because perfect, dislocation-free crystals have been obtained with Si and GaAs and it is possible to cut out and use an arbitrary plane as the substrate, various applications can be realized by homoepitaxial growth on this substrate. Accordingly, for the future, the realization of a dislocation-free "true bulk single crystal"[NOTE 4] (Type D in Table 2) of GaN is awaited.

2-3 Extension of applications by realization of "true bulk single crystal" and its social impact

The improved device properties and expansion of applications shown in Table 3 can be expected by realizing a "true bulk single crystal" with GaN-based semiconductors.

If a "true bulk single crystal" is realized with GaN-based semiconductors, high output, high efficiency, and an expanded wavelength region can be expected in light-emitting devices such as LEDs and LDs by taking advantage of the reduction in defect density

[NOTE4]
In GaN, there are cases in which Type C is called "bulk crystal." In this paper, the expression "true bulk single crystal" is used to refer to bulk crystals which are dislocation-free or have a low dislocation density ($\leqq 10^3cm^{-2}$) that cannot be realized with heteroepitaxial technology, and also enable cutting of any arbitrary crystal plane.

and the non-polar plane. In lighting using LEDs, in addition to popularization as general lighting devices taking advantage of the high efficiency and high color-rendering property of LEDs, this technology may open the way for new types of technical lighting. Assuming high efficiency and low cost in the white LEDs which are now used in some lighting applications, such as backlights for liquid crystal displays, full-scale replacement of incandescent light bulbs, fluorescent tubes, halogen and other general lighting, and technical lighting by LED lighting will be possible. A trial calculation has shown that an energy saving effect (reduction in annual power consumption) of approximately 20% could be achieved by popularization of LED lighting.[NOTE 5]

As LDs, devices with output power from several 10mW to around 200mW in BV wavelength range (450 nm) are marketed as light sources for Blu-ray Discs. Laser life is secured by alignment of the active layer with the region of a Type C polar substrate having a locally small dislocation density, as discussed previously. In the future, if a Type D substrate can be realized, it will be possible to satisfy both improved performance and reduced cost because a high quality LD structure can be manufactured without considering the distribution or deviations of dislocations. Furthermore, use of the non-polar plane will make it possible to produce longer wavelength devices, contributing to the realization of green (>500nm) LDs. At present, green lasers are generally wavelength conversion devices using Second Harmonic Generation (SHG), but problems related to miniaturization and cost reduction remain. Here, as well, it will be possible to reduce the size and cost of portable projectors, laser TVs, etc. by realizing green LDs with GaN-based devices. With the ongoing trend toward larger screens in liquid crystal TVs and plasma TVs, which are continuing to become the leading television technologies, power consumption is also increasing. If an inexpensive laser TV can be realized, energy consumption can be reduced by 1/2 to 1/3 that of liquid crystal TVs and plasma TVs of the same size, even in larger screen televisions,[13] and it will be possible to satisfy simultaneously large screen size and energy saving. (For details, see the following Column.)

In electronic devices, HFET (Hetero-junction Field Effect Transistor), which are a lateral device,

Table 3 : Expansions of applications by realizing true bulk single crystal (Type D)

Device	Improved features	Expanded applications	Social impact: Energy saving ( ) shows CO_2 reduction
LED	High out put Expanded wavelength region · Shorter wavelengths · Longer wavelengths	· High efficiency and high color-rendering in lighting (post fluorescent, halogen, etc. from general lighting to technical lighting) · Photocatalysts · Medical (sterilization) applications	Reduced power consumption by LED lighting · 2020:12.4billion kWh/yr (6.86 million tons/yr) · 2030:18.7billion kWh/yr (10.39 million tons/yr)
LD	Same as above.	· High speed writing in optical devices (next-generation DVDs) · Laser displays → Portable projectorsLaser television	Reduced power consumption by laser TV · 2012: 14.5 billion kWh/yr (8.02 million tons/yr)
Electronic device	Lateral type transistors (FET) · High output (to 100W) · High frequency Vertical type devices (IGBT, thyristors, etc.) · Low resistance · Normally-off operation	· High speed/high output transistors →High speed and high capacity in mobilecommunications, downsizing of land stations, low power consumption · Power transistors → High efficiency in industrial machinery, hybrid cars, and electric vehicles by high efficiency in inverters	Reduced power consumption by GaN devices[11] Transmitting/receiving amplifiers for cell phone ground stations + general purpose inverters · 2020: 9.8 billion kWh/yr (5.41 million tons/yr) · 2030: 20.2 billion kWh/yr (11.21 million tons/yr)

Prepared by the STFC

[NOTE5]
Assuming penetration of LEDs in indoor lighting is approximately 30% in 2030, as an energy saving effect (reduction of annual power consumption), it is predicted that power consumption can be reduced by 20 billion kWh, corresponding to approximately 20% of that used in lighting in 2005.

are currently manufactured using Type A, and are applied practically in base stations for mobile communications, such as cell phones, as high frequency devices operating at up to the GHz region. In the future, however, high output transistors which operate in the millimeter waveband (30-100GHz) will be necessary in order to achieve higher speed and large capacity in mobile communications. As shown in Figure 2 GaN should be the most suitable material from the viewpoint of material properties,[11] and will contribute to system miniaturization and reduced power consumption. On the other hand, as vertical devices,[NOTE 6] inverters with Si devices are used in industrial applications and hybrid cars. However, here again, higher system efficiency can be achieved by replacing these Si devices with GaN-based devices in the future, and as a result, energy savings and improved fuel economy in automobiles are considered possible. According to a trial calculation, the reduction in power consumption which can be realized by energy saving if GaN-based electronic devices are applied in amplifiers for transmitters in cell phone base stations and industrial inverters is approximately 10 billion kWh/year (2020).[11]

As discussed above, GaN-based semiconductors are materials that can provide the key to contributing to energy saving and reduction of emissions of greenhouse gases (GHG) as light-emitting devices and electronic devices (transistors).

3 Current status and problems of true bulk single crystal growth technology

3-1 Technical trends in true bulk GaN single crystals

The "true bulk GaN single crystal" described here, as mentioned previously, indicates a substrate material of a GaN bulk crystal which is dislocation-free or has a low dislocation density ($\leqq 10^3 cm^{-2}$) that cannot be realized with heteroepitaxial technology, and also makes it possible to cut out any arbitrary crystal plane.

While saying aiming at a "true bulk GaN single crystal," there are many cases in which the substrate or a seed crystal is used as the seed that forms the basis for crystal growth (excluding cases of self-nucleation growth, as will be discussed later). Methods of crystal growth are broadly classified as the epitaxial bulk (abbreviated "epi-bulk") method and the bulk method, depending on whether a substrate or a seed crystal is used as the basis of crystal growth. The vapor phase deposition method and liquid phase growth method are used in each of these methods.

To date, research and development has been carried out on both the gas phase and the liquid phase methods as growth methods for bulk GaN single crystals, but a "true bulk GaN single crystal" has not been realized with either method. Difficulties arise in research and development because existence in a melt, as with Si, GaAs, and InP, is difficult due to the high dissociation pressure of nitrogen in GaN.[17] However,

Column: "Satisfying both larger screen size and energy saving by laser TV"

The TVs which are currently available in the market are generally the plasma and liquid crystal types, and larger screen sizes are becoming increasingly popular. In line with this trend, a trial calculation shows that Japanese users will own approximately 12.6 million plasma sets and 20.4 million liquid crystal sets with screens of the 50 inch size or larger in the year 2012.[14] Because the power consumption of plasma and liquid crystal TVs of the 50-55 inch class is around 500W (Panasonic TH50PZ800: 585W, Sony KDL-55XR1: 480W), assuming average TV viewing time is 4 hours per day,[15] the annual power consumption of TVs of 50 inch size and larger is 25.1 billion kWh. If these types of TVs are replaced by TVs with low power consumption of 200W, such as laser TVs, an annual energy saving of 15.4 billion kWh can be realized. Converted by the annual power consumption of a general household in Japan, which is 5,650kWh,[16] this is equivalent to the total power consumption of approximately 2.73 million households. In addition to this energy saving benefit, a high color-rendering property is also a feature of laser TVs.

[NOTE6]

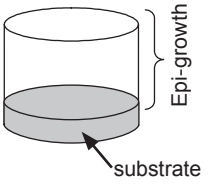
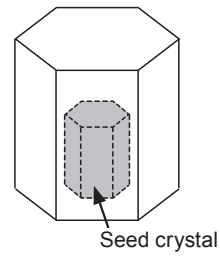
The terms "lateral type" and "vertical type" used when describing transistors mean that the carrier (electron/hole) is transferred and acts laterally or vertically, respectively.

as shown in Table 4, active R&D is being conducted in Japan and other countries with the aim of solving this problem and producing better substrate materials.

In gas phase growth in the epi-bulk method, the VPE method is used. This is a method in which crystal growth of GaN is performed by a vapor phase from gallium chloride gas and ammonia gas on a Type C substrate, and the film thickness is increased to a degree that makes it possible to cut out an arbitrary crystallographic plane in the thickness direction. With this method, uniform growth of high quality crystals is difficult, and although a non-polar plane (m plane) with a size of □10mm in the thickness direction has reportedly been obtained,[18] it was not possible to produce a high quality substrate with a large area. Moreover, because the original substrate was produced by the heteroepitaxial technique, the dislocation density was on the order of $10^5 cm^{-2}$.

Liquid phase growth techniques in the epi-bulk method include the high pressure solution method, ammonothermal (also called "ammothermal") method, and flux method. The high pressure solution method is a method in which GaN is grown by dissolving nitrogen at an ultra-high pressure (10,000-20,000atm) in a high temperature (1600°C) melt of Ga.[19,20] Growth to a thickness of several 100μm on a substrate has been reported,[21] but a crystal size that can be cut out in the thickness direction was not realized. In the ammonothermal method, a GaN raw material is dissolved and recrystallized in ammonia from supercritical to subcritical (400-500°C, 1000-4000 atm).[22] It has been reported that a crystal with a thickness of several 10μm was grown on a Type C substrate using an acidic mineralizer,[23] and crystals on the order of 5mm was grown using an alkali mineralizer.[24] In the flux method, GaN crystals are grown by dissolving nitrogen at several 10atm in a mixed melt of Ga and an alkali metal such as lithium (Li), potassium (K), or sodium (Na) at around 800°C.[25,26] Because Na is mainly used in this method, it is also called the Na flux method. Using the flux method, R&D has been conducted in an attempt to obtain an epi-bulk crystal on a φ2" substrate,[27] and crystal growth to a thickness of approximately 3mm has been achieved. Figure 3(a) shows an example of a 2" size crystal which was realized by the Na flux method as an epi-bulk method. The number of crystallographic defects is small in comparison to that with other methods, as the dislocation density is on the order of $10^5 cm^{-2}$. This is considered to be

Table 4 : Development trends in bulk GaN substrates

Method	Growthmethod			Features	Condition
Epi-bulkmethod (growth of thick film using Type C substrate) 	Vapor phas	VPE (Vapor Phase Epitaxy)		Substrate used as basis for growth is Type C (selective growth to thick film by vapor phase on substrate of sapphire, GaAs, etc.; dislocation density is on $10^8 cm^{-2}$ level). GaN substrate on the order of □10mm has been realized by cutting in the thickness direction.	Japanese companies hold the top share of Type C GaN substrates for next-generation DVD;epi-bulk is under development.
	Liquid phase	Liquid phase growth	High pressure solution method	*1 Dissolution of nitrogen in metallic Ga under high temperature, high pressure conditions of 10,000-20,000atm and 1600°C, and epitaxial growth on a substrate. Several 100μm in thickness direction.[18]	*2 Original Polish technology; at present, Poland is still most advanced.
			Ammonoth-ermal method	*3 Dissolution and recrystallization of GaN raw material in supercritical or subcritical NH_3 (temperature and pressure of 400-500°C and several 1000atm). Several 10μm to several mm in thickness direction.	Research in US using an alkali mineralizer and in Japan using an acidic mineralizer.
			Flux method	*4 Dissolution of nitrogen (pressure <100atm) from vapor phase in a metallic melt of Ga and an alkali metal such as N or K, and growth of GaN crystal at a temperature of around 800°C. Several mm in thickness direction.	Original Japanese technology; Japan is most advanced.
Bulk method (use of seed crystal, as shown below; use of self-nucleation growth 	Vapor phase	Vapor phase deposition method		Growth of GaN from vapor of metallic Ga and NH_3 or N_2 under high temperature of 1200°C or higher. Crystal size >1mm	In research phase, but little work in recent years.
		Sublimation method		Sublimation of GaN powder at around 1500°C and crystal growth in seed crystal in low temperature part. Crystal size <1mm.	
	Liquid phase	Solution growth		Growth of GaN crystal on the order of 100μm by slow cooling under high partial pressure of nitrogen and high temperature of 60,000atm and 2200°C or higher.	Research phase; in stage where existence of melt has been confirmed.
		Liquid phase growth	High pressure solution method	Same method as *1. Crystal of mm order is grown by self-nucleation growth without using a substrate.	Same as *2.
			Ammonoth-ermal method	Same method as *3. GaN crystal of 100μm to 1 inch size is grown by self-nucleation growth or seed crystal growth without using a substrate.	Polish company has realized a high quality 1" crystal.
			Flux method	Same method as *4. GaN crystal of several mm size is grown by self-nucleation growth or seed crystal growth without using a substrate.	R&D is being carried out by Japanese universities and companies.

Prepared by the STFC

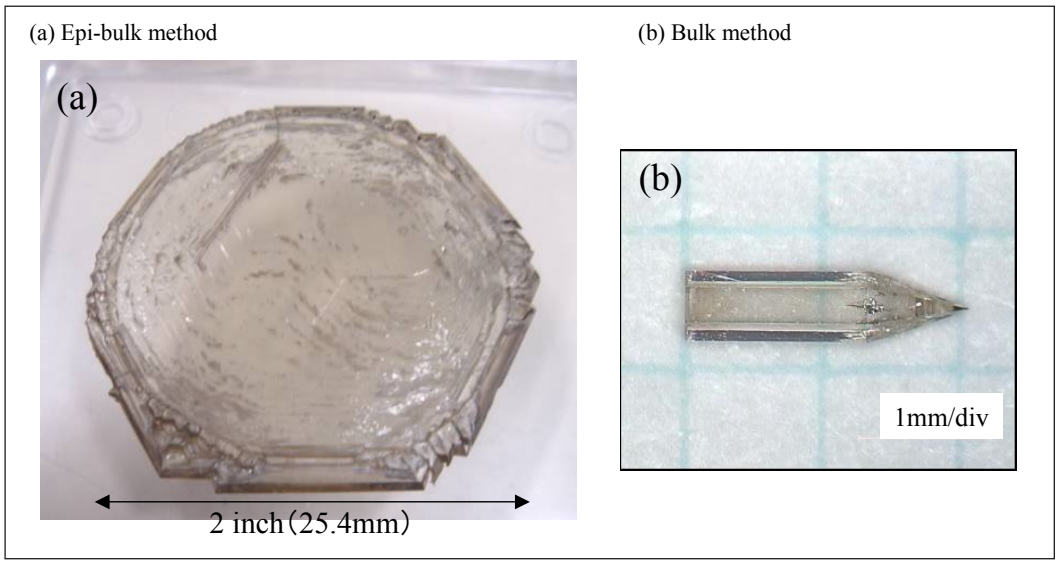


Figure 3 : Photographs of GaN crystals (Na flux method)

Source: (a) Mori Laboratory, Osaka University, (b) Ricoh

because flexure and consolidation of the dislocations in the horizontal direction occurred in the thick film epitaxial growth process.

Vapor phase growth and liquid phase growth are also used in the bulk method. The vapor phase methods are the vapor phase deposition method and the sublimation method. The vapor phase deposition method is a method in which GaN is synthesized directly from Ga vapor and ammonia or nitrogen gas under a high temperature exceeding 1200°C.[28] The sublimation method is a method in which the GaN raw material is sublimated and grown to seed crystals in the low temperature region.[29] However, the size of the crystals obtained is small. Due to the difficulty of crystal growth with the vapor phase deposition and sublimation methods, progress in research has been sluggish in recent years.

Liquid phase growth in the bulk method comprises melt growth and solution growth. Due to the high dissociation pressure of nitrogen in GaN, it has been confirmed that GaN exists as a melt only under high temperature, high nitrogen pressure conditions of 60,000atm and 2200°C or more.[30] Because these are not practical temperature and pressure conditions, R&D is not active. Accordingly, the focus of R&D on liquid phase growth in the bulk method is mainly on solution growth. The above-mentioned high pressure solution method, flux method, and ammonothermal method are also used in solution growth in the bulk method. These three solution growth techniques are basically the same as in the epi-bulk method, but they differ from those used in the epi-bulk method in that crystal growth is performed from self-nucleation growth[NOTE 7]or seed crystal growth. Figure 3(b) shows an example of a crystal obtained by self-nucleation growth. With the high pressure solution method, it has only been possible to obtain platelet-shaped crystals (thickness: approximately 10μm) on the mm order by self-nucleation growth; thus, this method has not reached practical application. With the ammonothermal method, needle-like crystals of several 100μm have been grown by self-nucleation growth using an acidic mineralizer, while wafer-like crystals 1 inch in size have been grown using an alkali mineralizer.[31] With the flux method, high quality columnar crystals of the mm order have been grown by self-nucleation growth and seed crystal growth,[26] and the possibility of increasing the size of the crystal increased by increasing the size of the crucible has been confirmed.

In any case, with the bulk method, it has not been possible to obtain a practical crystal size, namely, several 100μm to several mm. Although quality is extremely high, in that no dislocations can be observed, it is necessary to increase the size of the crystal to a degree which makes it possible to cut out crystals in the radial and axial directions of hexagonal columnar crystals, while maintaining high quality in order to obtain a non-polar plane.

[NOTE7]
Self-nucleation growth means nucleation and crystal growth from a condition in which no seed crystal or substrate exists

4 Efforts by various countries to solve technical problems

4-1 Industry-university-government R&D projects in Japan

(1) High efficiency photoelectric conversion compound semiconductor development project

The High Efficiency Photoelectric Conversion Compound Semiconductor Development Project (commonly known as the 21st Century Lighting Project) was carried out with the support of the New Energy and Industrial Technology Development Organization (NEDO) from fiscal year 1998 to 2003.[32] In this project, research and development on high efficiency near-UV LEDs, phosphors, substrate crystal growth technologies, and other topics were carried out with the aim of realizing LEDs for lighting. As part of the project, R&D on the above-mentioned high pressure solution method was carried out by Yamaguchi University and the Japan Energy Corporation. As a result, high quality platelet crystals with a size of 10mm and a dislocation density of less than $10^3 cm^{-2}$ were obtained by self-nucleation growth. The orientation of the principal plane of the crystal was the C plane (0001), and this was a polar plane.

However, industrial handling was difficult because the thickness of the crystal was several 10μm, and practical application was not achieved due to the large scale of the equipment.

(2) Semiconductor substrate crystal manufacturing technology for next-generation lighting

Using matching funds from the Special Coordination Funds for Promoting Science and Technology, research and development on GaN bulk crystals, centering on the epi-bulk ammonothermal method, was carried out by Tohoku University and the Mitsubishi Chemical Corporation during fiscal years 2004 to 2006.[23] In this project, epitaxial growth of GaN crystals with a film thickness of several 20μm was possible using a φ2" Type C GaN crystal as the substrate. The growth rate was approximately 1μm/h.

However, the orientation of the principal plane of the crystal was the C plane (0001), which is a polar plane. Moreover, because the dislocation density was the same or greater than that of the substrate ($10^{6\text{-}8} cm^{-2}$), practical application was not achieved. Thus, with the ammonothermal method to date, it has not been possible to satisfy both high quality (reduction of dislocations) and a size that makes it possible to cut out an arbitrary crystallographic plane, these being the necessary conditions for a true bulk crystal by the epi-bulk method.

(3) Development of semiconductor for high efficiency UV light-emitting devices

Development of a semiconductor for use in high efficiency UV light-emitting devices was promoted as a NEDO project during the period FY2004-2006. Osaka University, Toyoda Gosei Co., Ltd., and NGK Insulators, Ltd. conducted research and development on the epi-bulk method by the Na flux method.[33] As a result, a φ2" size with a dislocation density of $10^5 cm^{-2}$ was obtained (Figure 3(a)). The substrate used here, as in the above section (2), was a φ2" GaN substrate (Type C) grown by the VPE method. It was possible to grow GaN crystals with film thicknesses up to 3mm on this substrate by the Na flux method. With this method, the dislocation density of the obtained crystal was $10^5 cm^{-2}$, which is smaller than that of the substrate ($10^8 cm^{-2}$). This reduction is attributed to the fact that flexure and consolidation of dislocations like that in ELO of Type B in Table 2 occurs even without a mask in the crystal growth process by the flux method. The growth rate was also the highest among the liquid phase growth methods, at approximately 30μm/h. The orientation of the principal plane of the crystal was the C plane (0001), which is a polar plane. At present, sample crystals produced using this technology are being supplied by a joint venture called Frontier Alliance, which consists mainly of researchers from Osaka University. For further promotion of practical application, it will be necessary to develop large-scale equipment capable of accelerating the growth rate and maintaining stable long-term growth.

4-2 Other R&D trends in Japan

(4) R&D in private companies

At private companies in Japan, in addition to Japan Energy, Mitsubishi Chemical, Toyoda Gosei, and NGK Insulators, which were mentioned in sections (1) through (3), Ricoh Co., Ltd. is also involved in research and development using the Na flux method.[26] Japan Energy and Ricoh have been able to grow high quality (low dislocation density)

crystals by self-nucleation growth, and in particular, it was possible to obtain the m plane (side plane of columnar shape) in a colorless, transparent columnar crystal with Ricoh's Na flux method (Figure 3(b)). However, in all cases, the size of the crystal is still inadequate. In particular, it is not possible to manufacture devices with crystals produced by the high pressure solution method, as the crystal thickness is thin. On the other hand, because the crystals obtained by the Na flux method are of mm order, a further increase in the size of the crystals is desired in order to mass-produce devices.

4-3 R&D trends in other countries

(5) MURI (Multi university Research Initiative) project in United States

This project was carried out during the period 2001-2003 for the purpose of "Growth of Bulk Wide Bandgap Nitrides and Wafering" and centered on the University of North Carolina. In the project, research and development were conducted, respectively, on the vapor phase deposition method, ammonothermal method, and flux method as bulk GaN crystal growth methods. There have been no reports that these efforts have reached practical application.

(6) Research at University of California at Santa Barbara (UCSB) in United States

In the Solid State Lighting and Display Center (SSLDC) at UCSB, 5mm crystals were obtained in 82 days with the ammonothermal method.[24] As in the previous sections (2) and (3), these were obtained by the epi-bulk method using a Type C VPE GaN substrate. The crystals were judged not to be of high quality based on the results of X-ray diffraction and coloration of the crystals. However, the fact that a bulk crystal was obtained for the first time by the ammonothermal method can be considered a significant achievement. In this method, an alkali substance called a mineralizer was used in the material in order to enhance its solubility. When an alkali mineralizer is used, GaN displays negative solubility (i.e., its solubility increases as temperature decreases). On the other hand, in the research in Japan described in (2), an acid mineralizer was used; in this case, GaN displays positive solubility.

(7) R&D in Poland

The high pressure solution method is a method which was originally developed by the High Pressure Research Center (HPRC) in Poland. Although high quality GaN crystals are obtained, the size of the crystal and the equipment required are problems, and practical application has not been achieved.

The company Ammono Sp.zo.o in Poland is involved in R&D on the ammonothermal method, and has announced that it obtained high quality φ1" crystals using alkali mineralizer.[31] However, the growth rate, reproducibility, and industrial potential of this process are all unknown. Although research on the ammonothermal method is underway in Japan, the United States, and other European countries, it can be said that the research in Poland is the most advanced.

5 R&D issues for "true bulk GaN single crystal growth"

As R&D issues for "true bulk GaN single crystal growth," the requirements identified from device needs are summarized in Table 5 by the main crystal growth methods. The principal requirements are a substrate size $\geqq$ φ2 inches, high quality (dislocation density $<10^3 cm^{-2}$), and an arbitrary polar plane (possible to cut out planes either with or without polarity).

Research and development at various institutions is continuing to clarify the technical merits and demerits of each method. The vapor phase epitaxy (VPE) method has a high growth rate (sever 100μm/h), and it is considered probable that improvements in this technology will be made in the future under private-sector initiative. In contrast to this, with liquid phase growth (solution growth), it is expected to be possible to obtain high quality crystals, but the growth rate is slow (several μm/h to several 10μm/h), and both time and breakthroughs are still required for the establishment of device technology and crystal growth conditions.

In the realization of a "true bulk GaN single crystal," the most important requirements are, firstly, the achievement of high quality, meaning a dislocation density $<10^3 cm^{-2}$, and the possibility of cutting out arbitrary crystallographic planes. In achieving these conditions, a key target is to realize a size of 2" or larger. Looking at Table 5, the methods which either satisfy or have the potential to satisfy the essential

Table 5 : Summary of bulk GaN single crystal growth methods

Method		Requirements (needs for practical application from device		
		Size ≧φ 2	Quality (dislocation density) <10^3cm^{-2}	Arbitrary polar plane
Epi-bulk	VPE	△ (approx.10mm square)	× (low defect density has been achieved locally, but limit of heteroepitaxial method)	○(m plane substrate has been obtained by cutting in the cross sectional direction, but thickness is a problem)
	High pressure solution method	× (thickness to 100μm)	? (same order as substrate)	× (thickness is problem)
	Ammonothermal method	△ (thickness to 5mm)	× (larger than substrate, ≧ 10^3cm^{-2})	△ (thickness is problem)
	Flux method	△ (thickness to 3mm)	△ (to 105cm^{-2}; lower locally)	△ (thickness is problem)
Bulk	High pressure solution method	× (thin platelet shape, thickness of several 10μm)	○	× (polar C plane grows as principal plane)
	Ammonothermal method	△ (φ 1" has been realized using alkali mineralizer)	○	? (thickness is unknown)
	Flux method	△ (to several mm)	○	○

Prepared by the STFC

requirements of quality and an arbitrary polar plane are the flux method in the epi-bulk type and the ammonothermal method and flux method in the bulk type. Among these, the epi-bulk flux method differs from the other epi-bulk methods in that it is possible to produce crystals with a lower dislocation density than the substrate, suggesting the prospect of achieving higher quality. Furthermore, as mentioned in (3), this method was investigated in an industry-university-government project and is already in the stage where samples are being shipped. R&D is currently underway, with the aims of further reducing the dislocation density and achieving a large-scale size for maintaining the arbitrary polar plane. On the other hand, neither the ammonothermal method nor the flux method in bulk approach has been studied in an industry-university-government project up to the present time. With the ammonothermal method, research on high quality and large scale is progressing, particularly in Poland, and foreign countries also hold the lead in intellectual property. In contrast, the flux method is a technology which was discovered by Prof. Hisanori Yamane of Tohoku University in 1997, and active research and development are continuing in Japan even today. Moreover, the patent which is the basis for the flux method is common to the epi-bulk and bulk approaches and is owned by a Japanese company, and peripheral patents have been granted to Japanese companies and universities.

Accordingly, for the time being, it is considered that Japan should aim at realization of a practical "true bulk GaN single crystal" by developing and utilizing crystal growth methods by the flux method in both the epi-bulk method and the bulk method.

The key to practical application of both of these methods is ultimately realizing large scale in the size of the crystal. In starting from high quality crystals and upscaling these to the size of φ2 to 4", which is necessary for device manufacture, efforts in combination with device development are demanded. As the division of roles in industry-university-government collaboration, it is considered desirable that universities work to elucidate the mechanism of crystal growth, private companies develop the crystal growth process from the design of crystal growth devices, while government serve as a coordinator.

6 Conclusion

Because two of the key scientific breakthroughs in GaN-based semiconductor devices were achieved in Japan in the past, and these have led to the practical applications of today, it is no exaggeration to say that these are Japanese-originated semiconductor devices. The development of crystal growth techniques not only realizes devices with higher functions, but can also make an important contribution to preventing global warming through energy saving.

High quality bulk crystals of Si, GaAs, and InP all exist. Substrates are produced using these bulk

crystals, and devices can be manufactured on the respective substrates. However, in the case of GaN, a high quality bulk crystal, that is, a "true bulk single crystal," does not exist, and for this reason, it is not possible to take full advantage of the intrinsic potential of GaN.

Among GaN crystal growth technologies, the vapor phase epitaxy (VPE) method offers the fastest growth rate, and it is therefore thought that technical improvements will be made in this technology under private-sector initiative. The liquid phase growth method, which should make it possible to obtain an even higher quality "true bulk single crystal," is currently in the basic research stage, and further breakthroughs and time will be required. Accordingly, in the future, it is desirable to promote research and development of a "true bulk GaN single crystal" by industry-university-government collaboration, based on a division of roles in which universities conduct research to elucidate the mechanism of crystal growth, companies develop the crystal growth process from the design of crystal growth devices, while government coordinates these activities. If Japan realizes a "true bulk GaN single crystal" at an early date, in advance of other countries, it is considered possible to maintain and enhance this country's international competitiveness in the field of GaN-based semiconductor devices in the future.

Acknowledgements

The author discussed the early period of research with Prof. Hiroshi Amano of Meijo University, who was a pioneer in GaN-based devices and carried out research and development on the low temperature buffer layer and p-type devices. Photographs of GaN crystals were provided by Prof. Yusuke Mori of Osaka University and Ricoh Co., Ltd. Explanations of survey data were provided by the Research and Development Association for Future Electron Devices and the Japan Research and Development Center for Metals. The author wishes to express his deep appreciation to all concerned for their generous cooperation.

References

[1] "International Comparison of Science and Technology/Research and Development: Electronic Information and Communications – 2008," Japan Science and Technology Agency (JST), Center for Research and Development Strategy (February 2008).(Japanese)

[2] H. Amano, N. Sawaki, I. Akasaki and Y. Toyoda: App. Phys. Lett. 48, 353 (1986)

[3] H. Amano, M. Kito, K. Hiramatsu and I. Akasaki: Jpn. J. App. Phys. 28, L2112 (1989)

[4] S. D. Lester et al.: High dislocation densities in high efficiency GaN based light-emitting diodes, App. Phys Lett.,66, 1294 (1995)

[5] H. Amano: Development of Short Wavelength Visible/Ultraviolet Light-Emitting Devices and Semiconductor Heteroepitaxy," Oyo Butsuri, 71, 1329 (2002).(Japanese)

[6] M. Takeya, et al. : "Relationship between Dislocation Density and Reliability in GaN-based Semiconductor Lasers," Proceedings of the 63rd Autumn Joint Meeting of the Japan Society of Applied Physics, 27a-YH-11 (2002).(Japanese)

[7] A.Usui et al.: Thick GaN Epitaxial Growth with Low Dislocation Density by Hydride Vapor Phase Epitaxy,Jpn. J. Appl. Phys., 36, L899 (1997)

[8] K. Motoki et al : Preparation of Large Freestanding GaN Substrates by Hydride Vapor Phase Epitaxy Using GaAs as a Starting Substrate, Jpn. J. Appl. Phys., 40, L140 (2001)

[9] Y. Ohshima et al. : Preparation of Freestanding GaN Wafers by Hydride Vapor Phase Epitaxy with Void-Assisted Separation Jpn. J. Appl. Phys., 42, L1 (2003)

[10] Motoki et al. : "Low Dislocation GaN Crystal by Advanced-DEEP Method," Japan Society for the Promotion of Science, No. 161 Committee, Materials of the 54th Workshop 5(2007). (Japanese)

[11] Report on "Technology Strategy Roadmap for Nitride-based Compound Semiconductors – FY2006," Japan Machinery Federation, Japan Research and Development Center for Metals (2007). (Japanese)

[12] 72nd Council for Science and Technology Policy, Item 3, Recent Trends in Science and Technology: "A Bright Future Illuminated by Advanced LEDs" (December 2007); "Report on the Technology Strategy Roadmap for Nitride-based Compound Semiconductors – FY2006," Japan Machinery Federation, Japan Research and Development Center for Metals (March 2007). (Japanese)
[13] Laser Focus World Japan, 2008. 8, P. 34
[14] SID 2008, BC-4, iSuppli, The TV Market and its Impact on the Display Business (2008)
[15] NHK Broadcasting Culture Research Institute : "National Time Use Survey 2005 Report (February 2006). (Japanese)
[16] Central Research Institute of Electric Power Industry, (Rev.) "Questions regarding "Do all electrified homes contribute to preventing global warming?" : http://creipi.denken.or.jp/pieceeco/data/071001.pdf .(Japanese)
[17] J. Karpinski J. Jun and S. Porowski: Equilibrium pressure of N2 over GaN and high pressure solution growth of GaN, J. Cryst. Growth, 66, 1 (1984)
[18] K. Fujito et al.: High-Quality Nonpolar m-Plane GaN Substrate Grown by HVPE, ICNS07 I1 (2007)
[19] S. Porowski and I. Grzegory: Thermodynamical properties of III-V nitrides and crystal growth of GaN at high N2 pressure J. Cryst. Growth, 178, 174 (1997)
[20] T. Inoue Y. Seki O. Oda: Growth of Bulk GaN Single Crystals by the Pressure-Controlled Solution Growth Method The 1st Asian Conference on Crystal Growth and Crystal Technology, T-B-12, 341 (2000)
[21] I. Grzegory et al.: Seeded growth of GaN at high N2 pressure on (0001) polar surfaces of GaN single crystalline substrates, Materials Science in Semiconductor Processing, 4, 535 (2001)
[22] D. R. Ketchum and J. W. Kolis : Crystal growth of gallium nitride in supercritical ammonia, J. Cryst. Growth,222, 431 (2001)
[23] Report on "Semiconductor Substrate Crystal Manufacturing Technology for Next-Generation Lighting," Tohoku University Institute of Multidisciplinary Research for Advanced Materials, Mitsubishi Chemical Co., Ltd. (February 2008). (Japanese)
[24] T. Hashimoto et al.: Growth of Bulk GaN Crystals by the Basic Ammonothermal Method, Jpn. J. Appl. Phys., 46,L889 (2007)
[25] H. Yamane et al.: Preparation of GaN Single Crystals Using a Na Flux Chem. Mater. 9 (1997) 413.
[26] S. Sarayama and H. Iwata: High Quality Crystal Growth of Gallium Nitride by Flux Method, Ricoh Technical Report, 30, 9(2004)
[27] F. Kawamura et al.:Growth of a Two-Inch GaN Single Crystal Substrate Using the Na Flux Method, Jpn. J.Appl. Phys., 45, L1136 (2006)
[28] H. Shin et al.: Growth and decomposition of bulk GaN: role of the ammonia/nitrogen ratio, J. Cryst. Growth,236, 529 (2002)
[29] S. Kurai et al. : Homoepitaxial Growth of GaN on Thick GaN Substrates Prepared by Sublimation Method, Proceedings of TWN'95, Nagoya, B-6 (1995)
[30] W. Utumi et al. : Congruent melting of gallium nitride at 6 GPa and its application to single-crystal growth Nature Materials, 2, 735 (2003)
[31] R. Doradzinski et al.: Excellent crystallinity of truly bulk ammonothermal GaN 5th International Workshop on Bulk Nitride Semiconductors (2007)
[32] Report on "High Efficiency Photoelectric Conversion Compound Semiconductor Development Project (21st Century Lighting Project)," Japan Research and Development Center for Metals (March 2003). (Japanese)
[33] Report on "Semiconductor Development for High Efficiency UV Light-Emitting Devices Project," New Energy and Industrial Technology Development Organization, Nanotechnology and Materials Technology Development Department, (2008). (Japanese)

Profile

Seiji SARAYAMA
Affiliated Fellow,NISTEP

Senior Researcher, Tohoku Laboratory, Research and Development Center, Ricoh Co., Ltd.
At Ricoh, Dr. Sarayama is engaged in research and development of GaAs, GaN, and other compound semiconductor materials and devices. His goal is to contribute to environmental technology by new material and device technologies. He is interested in MOT/technology strategy and their practice. His efforts in R&D are based on the idea that satisfying both pleasure and stress is key.

(Original Japanese version: published in November 2008)

5

Trends and Problems in Earthquake Prediction Research

SHOZO MATSUMURA
Affiliated Fellow

1 Introduction

The M8.0 earthquake (M: magnitude) which struck Sichuan Province in China on May 12, 2008 was a major disaster of historic proportions, resulting in a total of nearly 100,000 dead and missing. According to newspaper reports, dissatisfaction over the fact that no information predicting this earthquake was provided erupted in the affected area. In particular, the large number of victims claimed by the collapse of elementary and middle school buildings is viewed as a problem, and there has also been criticism to the fact that China's earthquake countermeasures have devoted energy to observation of well water and the earth's crust, but seismic countermeasures for buildings have been delayed.[1] This tragedy may become the occasion for a review of policies in connection with earthquake disaster prevention in China and will perhaps lead to seismic retrofitting of buildings. Even though the shock of the Sichuan earthquake had still not subsided, Japan was struck by the Iwate-Miyagi Nairiku Earthquake (M7.2) on June 14, which caused more than 20 deaths. This was followed by the Iwate-Chubu Earthquake (6.8) on July 24. All of these events have strengthened the impression that earthquakes strike unexpectedly, at any time and place. Expressions of this sort have become a way of disposing of the matter and have taken root in the mass media. Considered from another direction, however, it can also be said that this is a clear manifestation of distrust toward earthquake prediction research as it now exists. What, then, is the current state of earthquake prediction research?

Although prediction had been considered a central challenge for earthquake research, not limited to China, but worldwide, this trend changed around the mid-1990s. In Japan, this was occasioned by the Great Hanshin-Awaji Earthquake (M7.3), which struck Kobe on January 17 causing more than 6,400 deaths, 1995, and subsequently led to a review of earthquake countermeasures as a matter of national policy. Accompanying this, a complete change in earthquake research was unavoidable, and the previous single-minded dedication to earthquake prediction was replaced by specialization either in more practical disaster prevention research or in more fundamental scientific research. However, this does not mean that earthquake prediction as such was neglected. The proper course in science is "prediction of future events based on an analysis of past events." It should perhaps be said that this attitude was an attempt to take a more scientific stance toward the word and concept of "prediction," which had been used easily or somewhat carelessly up to that time. Today, more than 10 years since that change in direction, what is the relationship of the reformed earthquake research to prediction, and what contribution has it made? Have we not reached the time when we should stop and look back on how close we have come to achieving the specified target, and whether we have failed to grasp the target itself?

This paper examines the content of related research, divided into "earthquake prediction" in the narrow sense and "earthquake forecasting." Although there are no large differences in the meanings of the words "prediction" and "forecasting" themselves, here, they will be distinguished as follows: "Earthquake forecasting" means estimation of the magnitude and probability of the occurrence of an earthquake which may occur in a certain location, when that location is designated. The probability of occurrence is obtained by statistical operations premised on the fact that earthquakes occur repeatedly in the same location, and is based on information obtained by excavation of past evidence, that is, the scale, cycle, and deviations in the cycle of earthquakes, and the time which has

elapsed since the most recent earthquake. In this case, the information which provides the grounds for the probability value does not change with increasing time in spite of the fact that the probability value increases with time. In contrast, in "earthquake prediction," how close the accumulated stress in the source area is to the limit is estimated from the transition in measured data, when a designated earthquake is the object. In other words, premised on detection of some type of precursory phenomenon, this approach attempts to increase the amount of information itself prior to the occurrence of an earthquake, and in particular, to dramatically increase information immediately before the earthquake. While reviewing the history and current status of both "earthquake forecasting" and "earthquake prediction," the author would like to analyze the problems involved in both approaches, including his personal impressions.

2 Evolution of earthquake prediction research

2-1 Frequency of destructive earthquakes in Japan

First, let us confirm the frequency with which earthquakes actually occur in Japan. This is simpler if earthquakes are classified by magnitude. The average annual frequency of earthquakes in Japan and the surrounding oceans is 0.1 for M8 earthquakes, 1 for M7, and 10 for M6. Thus, frequency increases by approximately 1 order (10 times) for each 1 order decrease in magnitude. The frequency of earthquakes worldwide is approximately 10 times these numbers, which means, conversely, that approximately 10% of the world's earthquakes occur in Japan and the oceans that surround it.[2] For this reason, Japan is known as an earthquake-prone country. What essentially concerns ordinary people is not the size of an earthquake itself, but the extent of damage that it causes. However, the frequency of earthquakes which caused considerable damage in the past is larger than most people recognize.

Table 1 presents a list of earthquakes since 1900 which caused 10 or more deaths. Figure 1 is a graph showing the secular change in their cumulative frequency. There have been a total of 36 of these destructive earthquakes during the past 109 years, and their average interval is 3.1 years. From the graph, it appears that the frequency of destructive earthquakes has decreased since 1950, but this seems to be due to slightly decreased seismic activities during the last 50 years, rather than the effects of promoting seismic countermeasures. However, even while saying that the frequency of earthquakes is low, the average interval in the latter part of the period is 4.5 years, and a revival of activity can also be seen in recent years.

Table 1 : List of earthquakes causing 10 or more deaths/missing persons in Japan since 1900 (Year, location, magnitude, number of dead/missing)

Major destructive earthquakes (10 or more dead/missing)

1900 Miyagi Hokubu (M7.0)	17	1944 Tonankai (M7.9)	1,223
1901 Hachinohe-oki (M7.2)	18	1945 Mikawa (M6.8)	2,306
1905 Geiyo (M7.3)	11	1946 Nankai (M8.0)	1,330
1909 Anegawa (Eno) (M6.8)	41	1948 Fukui (M7.1)	3,769
1911 Kikaijima (M8.0)	12	1949 Imaichi (M6.4)	10
1914 Akita Semboku ((M7.1)	94	1952 Tokachi-oki (M8.2)	28
1914 Sakurajima (M7.1)	35	1964 Niigata (M7.5)	26
1922 Chijiwa Bay (M6.9)	26	1968 Tokachi-oki (M7.9)	52
1923 Kanto (M7.9)	142,807	1974 Izu Hanto-oki (M6.9)	30
1924 Tanzawa (M7.3)	19	1978 Izu-Oshima-kinkai (M7.0)	25
1925 Kitatajima (M6.8)	428	1978 Miyagi-oki (M7.4)	28
1927 Kitatango (M7.3)	2,925	1983 Nihon-kai Chubu (M7.7)	104
1930 Kita Izu (M7.3)	272	1984 Nagano Seibu (M6.8)	29
1931 Nishi Saitama (M6.9)	16	1993 Hokkaido Nansei-oki (M7.8)	202
1933 Sanriku (M8.1)	3,064	1995 Hyogo Nanbu (M7.3)	6,437
1939 Oga (M6.8)	27	2004 Chuestsu (M6.8)	68
1940 Kamui Misaki-oki (M7.5)	10	2007 Chuetsu-oki (M6.8)	15
1943 Tottori (M7.2)	1,083	2008 Iwate-Miyagi Nairiku (M7.2)	23

Prepared by the STFC based on Reference [3]

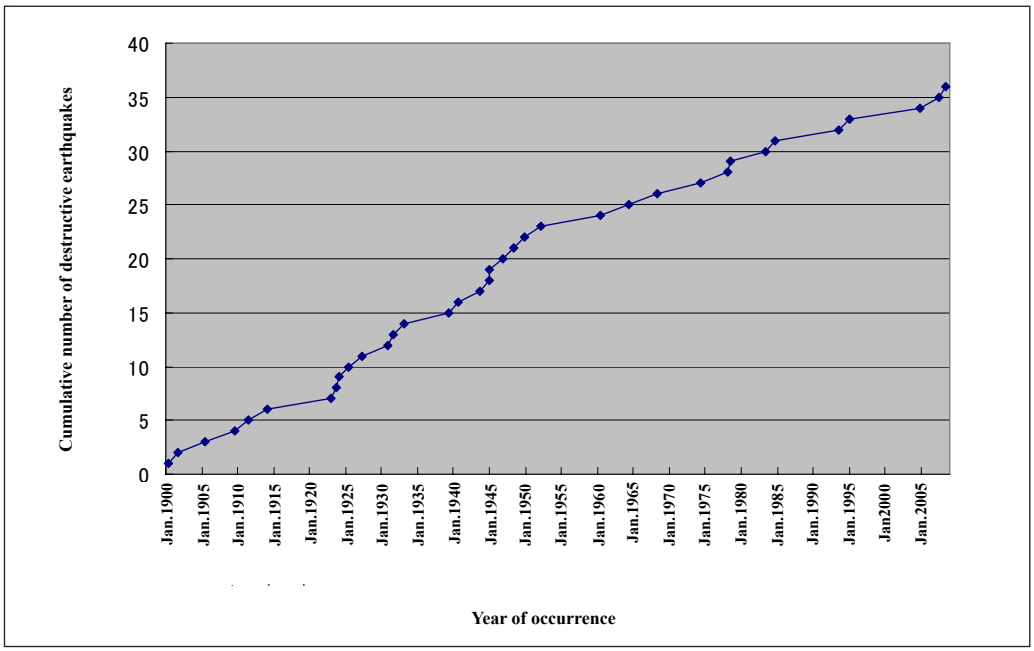


Figure 1 : Distribution of cumulative number of destructive earthquakes (data at points in Table1)

Prepared by the STFC

The feeling that earthquakes occur at long intervals and have no particular relationship with oneself, in spite of the actual situation, can be attributed to the fact that people do not know where a given earthquake will strike. The objective of researching earthquakes should be to give some type of answer to this question, namely, where and what kind of earthquakes will occur in the future. In this sense, earthquake prediction, while being a challenge for scientific research, was at the same time a "dream" of humankind. However, prediction is still a "dream" today. This is because there are still no examples which are widely recognized as successful earthquake prediction in the strict sense. Therefore, let us begin by tracing the history of earthquake prediction research up to the present.

2-2 History of earthquake prediction research in various countries

Saying that there have been absolutely no successful attempts in earthquake prediction may be somewhat misleading. First, some examples which are generally considered to be successful predictions should be mentioned. The most famous example is the Haicheng earthquake (M7.3) in Heibei Province, China in 1975.[4] Seismic activity had become active in the vicinity several years before the earthquake, and various anomalous phenomena, including microtremors, crustal movements, and changes in the groundwater, occurred frequently beginning several days before the event. The earthquake warning which was issued based on these precursory phenomena was effective, and a large number of people were able to take refuge. In spite of this, it is difficult to recognize this success as the result of earthquake prediction in the scientific sense because no warning was issued before the Tangshan Earthquake (M7.8) which struck the same Hebei Province in 1976 and caused an historically-unprecedented 240,000 deaths. Although it seems that China subsequently continued with attempts at earthquake prediction, there have been no reports of success. The 2008 Great Sichuan Earthquake mentioned in the Introduction is an example of failure. Thus, in view of the inability to demonstrate reproducibility and universality, the above-mentioned example cannot be regarded as a scientific achievement.

Another well-known example of research is the VAN method in Greece. Greece is extremely seismically active because it is located at a plate convergence zone, and as a result, it has suffered much damage due to earthquakes exceeding M5. A research group led by Prof. Varotsos of the University of Athens proposed a method of earthquake prediction based on anomalous changes in monitoring signals from a seismic geoelectric potential observation network installed at the southern edge of the Balkan peninsula. A successful example which resulted in an actual evacuation in 1993 was reported.[5] However, some have also expressed doubts about the objectivity

of this method because the mechanism responsible for producing the anomalous subterranean electrical signals as a precursory phenomenon is unknown. On the other hand, another report supported the reliability of the VAN method based on a statistical study.[6] Although no conclusion has been reached, the VAN method has had a large effect on this area of research, and in Japan as well, earthquake prediction by electromagnetic techniques now occupy the mainstream position in sessions on earthquake prediction at conferences, taking precedence over mechanical methods based on seismic and crustal movements, groundwater, and the like.

The fact that the VAN method has become the main topic in conferences stirred a worldwide debate as to whether earthquake prediction is possible in principle. This debate was ignited by Prof. R.J. Geller (then Assistant Professor), who had recently joined the University of Tokyo from Stanford University. Prof. Geller contributed papers to Nature and other journals asserting that prediction is impossible in principle because the time of occurrence and magnitude of earthquakes is controlled by contingency.[7] This claim was countered by a group led by Prof. M. Wyss of the University of Alaska (now at WAPMERR (World Agency of Planetary Monitoring and Earthquake Risk Reduction), who held an international symposium on earthquake prediction and argued that precursory phenomena unmistakably exist. Thus, in the 1990s, an unprecedented controversy developed between a faction claiming that earthquake prediction was possible and a faction claiming it was impossible. Thereafter, the controversy was gradually forgotten without reaching any definitive conclusion, and earthquake prediction research as a whole showed a declining tendency.

The United States was no exception to this trend. Interest in earthquake prediction research in the United States is concentrated on the San Andreas Fault on the West Coast. This can be attributed to the fact that the Fort Tejon Earthquake (M8.0) which struck Los Angeles in 1857 and the 1906 San Francisco Earthquake (M7.8) occurred along this fault, and a recurrence of these earthquakes is feared. Because this fault undergoes steady slip motion in the intermediate area between these two great earthquakes, great earthquakes do not occur in the intermediate region. M6 class earthquakes had occurred at intervals of somewhat more than 20 years near the town of Parkfield, which is located in this region, and the next earthquake was forecast to occur by 1993. As there was no concern about injury due to an earthquake of this size in that area, this was perceived to be a favorable opportunity for earthquake prediction. Numerous observation devices and human monitors were concentrated at the site for an earthquake prediction experiment call the "Parkfield Experiment," but the expected earthquake failed to materialize. An M6 earthquake actually occurred in 2004, which was more than 10 years late and was probabilistically low, but in this case, the expected precursory phenomena were not detected.[8] This marked a turning point, after which earthquake prediction research also waned in the United States, following the trends in other countries.

2-3 History of earthquake prediction research in Japan

Japan has always played a leading role in earthquake prediction research. In 1962, the authorities in the scientific world of the time compiled recommendations on the direction of earthquake prediction research in a document called "Earthquake Prediction-Current Status and Plans for Its Promotion" (so-called "Blueprint"), and based on this, an earthquake prediction project was launched in 1965 with a budget from the government.[9] The Coordinating Committee for Earthquake Prediction (hereinafter, Coordinating Committee) was established in 1969, and as an advisory body to the President of the Geographical Survey Institute, its activities have continued to the present. In 1970, the Coordinating Committee identified seismic hazard regions nationwide and issued recommendations that observation and monitoring should be strengthened in these areas (subsequently revised in 1978). In designating these areas, areas were selected not based only natural conditions, but also considering social conditions. According to Yoshimitsu Okada (now President of the National Research Institute for Earth Science and Disaster Prevention), many major earthquakes have occurred in the designated area in the 29 years since the revision, including the Great Hanshin-Awaji Earthquake, and a prediction success rate of 80% was achieved. (Figure 2).[10] On the other hand, Katsuhiko Ishibashi, who was an assistant in the Earthquake Research Institute at the University of Tokyo at the time (and later became a Professor of Kobe University), presented his theory of a

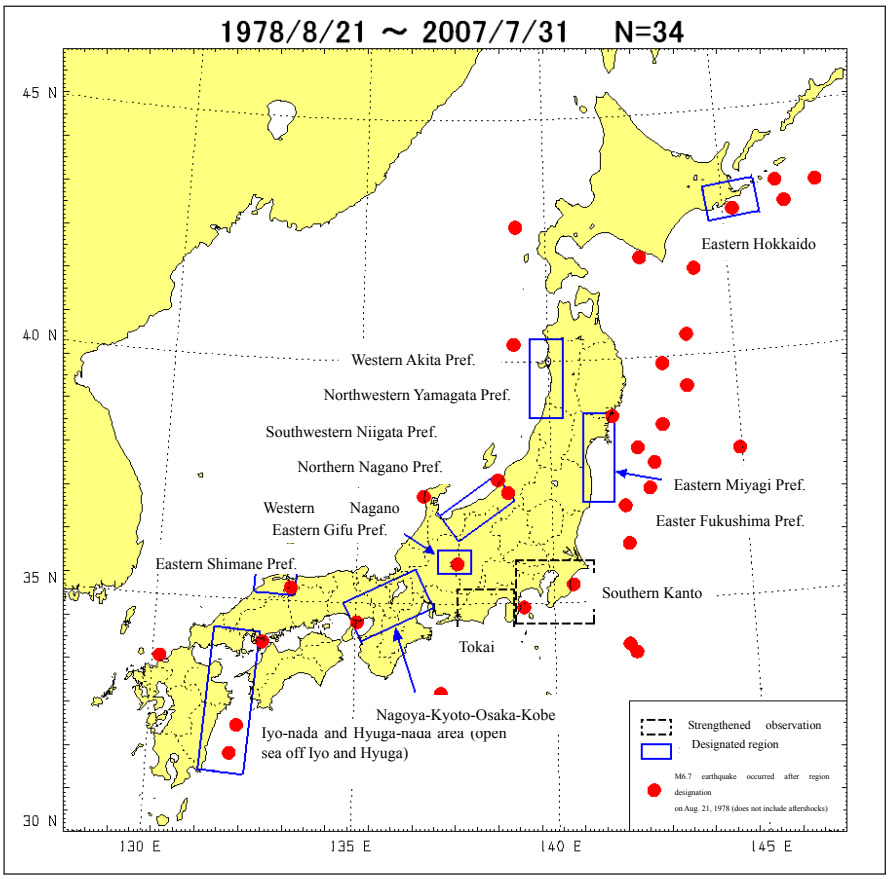


Figure 2 : Regions designated by the coordinating committee for earthquake prediction and actual earthquakes M6.7 or larger

Source : Reference [12]

Great Suruga Bay Earthquake at the 1976 Autumn Conference of the Seismological Society of Japan.[11] At Suruga Bay, strain has continued to accumulate for more than 120 years since the Ansei-Tokai Earthquake of 1854, and it has been pointed out that this area is now at an extremely volatile crisis point. The government attached considerable importance to Ishibashi's theory, and in 1978, enacted the first law, which was called the "Special Measures Law for Countermeasures against Large Earthquake,"[12] for earthquake countermeasures for a M8 class Suruga Bay Earthquake, i.e., a Tokai Earthquake. Based on this law, the area where a seismic intensity of 6 or higher was assumed, centering on Shizuoka Prefecture, was demarcated as an "Area under Intensified Measures against Earthquake Disaster," and routine monitoring of this area was assigned to the Japan Meteorological Agency. In order to fulfill its mission of predicting Tokai earthquakes, the Japan Meteorological Agency organized the "Earthquake Assessment Committee for Areas under Intensified Measures against Earthquake Disaster (EAC)," which thereafter held monthly meetings to discuss arrangements. During this period, in 1978, the Izu-Oshima-Kinkai Earthquake (M7.0) occurred on the active fault between the Izu Peninsula and Izu-Oshima Island, causing 25 deaths. In this earthquake, anomalous phenomena were discovered before the actual occurrence in a diverse range of observation items, including seismic activity, crustal movement, groundwater level, radon gas concentration, and others. Thus, it was evaluated that there still remains one step away to actually predicting earthquake.[13] In 1977, a group headed by Masakazu Otake (now Professor Emeritus of Tohoku University and Chair of the above-mentioned Coordinating Committee) discovered a quiescence in seismic activity around Oaxaca in Mexico and published a paper warning of an impending major earthquake. The following year, in 1978, the M7.7 Oaxaca Earthquake occurred, giving a strong impression that earthquake prediction is possible.[14] From the 1970s through the 1980s, Japan developed a nationwide microtremor observation network, and earthquake prediction research centers were established in universities and national research institutes. In view of these and similar examples, it

can be said that this was the period when earthquake prediction aroused the highest expectations in Japan, and it was believed that earthquake prediction would be realized in the near future.

2-4 Impact of the Great Hanshin-Awaji Earthquake

On January 17, 1995, when the worldwide debate on the possibility of earthquake prediction was at its peak, the Great Hanshin-Awaji Earthquake (M7.3) occurred unexpectedly, rupturing an active fault from the city of Kobe to Awaji Island, and more than 6,400 persons died in the ensuing disaster. Although some reports had forecast this earthquake, the residents of the region gave hardly any thought to the danger of a massive earthquake. Therefore, there was strong criticism of earthquake research, which had not issued effective advance warnings in spite of the unprecedented numbers of victims in recent years. This became the occasion for a sweeping review of the position of prediction research. The impact of the Hanshin-Awaji Earthquake was also sufficient to cause a major change in national policy. While this did not led to a rejection of the concept of earthquake prediction, dependence on prediction in earthquake countermeasures was fundamentally eliminated. Concretely, a new Earthquake Research Division was established in the former Science and Technology Agency (later moved to the Earthquake and Disaster-Reduction Division, MEXT). The "Headquarters for Earthquake Prediction Research Promotion" which had existed up to that time was renamed the "Headquarters for Earthquake Research Promotion" (hereinafter, "HERP"), and was also reorganized.[15] A Policy Committee and an Earthquake Research Committee were established under HERP, and a large number of sectional meetings and subcommittees were established. These committees conduct short-term and long-term assessments of Japan's earthquake activities and communicate the results to the larger society by way of the mass media. The name "prediction" in administrative organizations was removed or changed. However, this does not mean that "prediction" was eliminated completely. At one time there was a view that overlapping of the Earthquake Prediction Coordinating Committee and the Earthquake Research Committee was a problem and the Coordinating Committee should be abolished, but ultimately the Coordinating Committee has continued to exist up to the present. In spite of the fact that discussions in the two Committees partially overlap, their purposes and the nature of their evaluations are slightly different. The Earthquake Assessment Committee in the Japan Meteorological Agency also survived. Although this is due to the formal provisions of the existing law, this body has continued to exist because hopes for prediction have not been abandoned, at least in the limited case of Tokai earthquakes. In this connection, it may be noted that the Central Disaster Management Council conducted a review of the assumed source area of Tokai earthquakes in 2001.[16] Based on new observational information, the source area that had existed until that time was greatly revised, and the object area for disaster prevention countermeasures was expanded. However, the basic framework and approach to prediction remained unchanged.

3 Change to earthquake forecasting

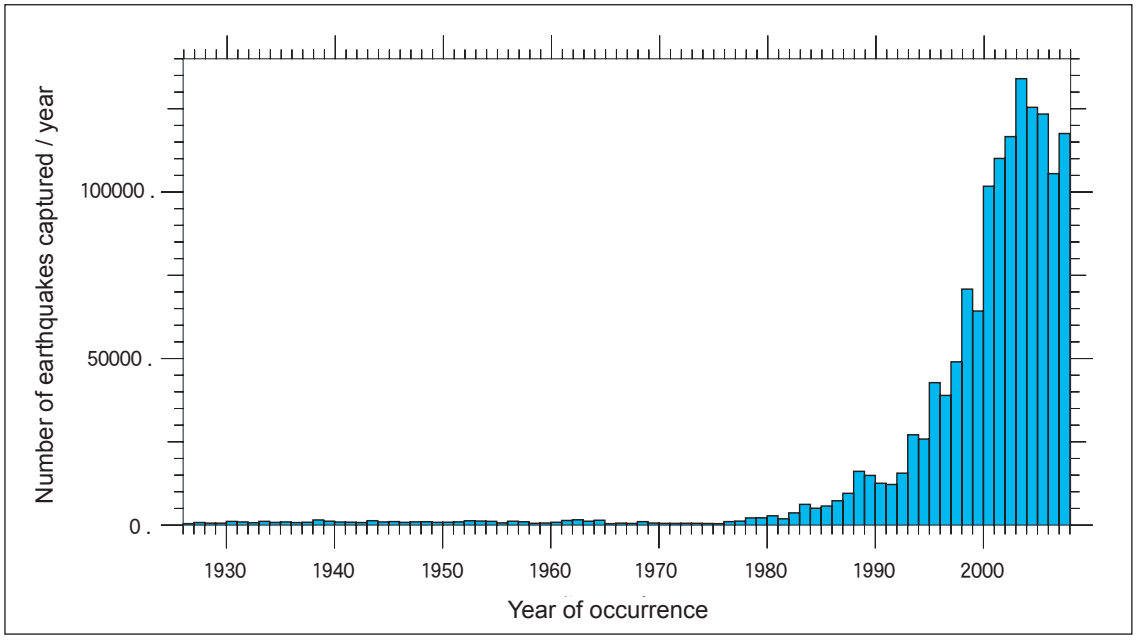
As a change in policy accompanying the establishment of HERP, administrative policies changed from the former orientation toward prediction to forecasting the occurrence of earthquakes. As mentioned previously, earthquake forecasting does not pursue precursory phenomena, but rather, assesses the probability of the occurrence of a major earthquake statistically, based on an assessment of active faults and events which have occurred in the past.

3-1 Construction of the Kiban Network (basic nationwide seismic network)

One aspect which underwent a major transformation was the nationwide observation network.[15] The basis of observation of crustal movement is observation of earthquakes and crustal movement (expansion/contraction, rising/falling of the ground surface). Up to that time, however, observation of earthquakes had been the responsibility of the Japan Meteorological Agency, universities, and national research institutes, while the Geographical Survey Institute had been responsible for observation of crustal movement. It can be said that this period was an "age of rival warlords." While these various organizations provided coverage for the entire country, the observational density was coarse. Moreover, specifications were not unified. More than anything else, this was a high barrier to mutual use of data between organizations. An enormous length of time, on the order of several years,

and a corresponding amount of work were required in making geodetic measurements of the entire country. With the world in the midst of the IT revolution and rapid technical progress in the fields of observation, communications, and data processing, an outdated observation network which failed to change was fated to be left behind. Therefore, following the Great Hanshin-Awaji Earthquake of 1995, HERP decided to budget new funds and promote the construction of a nationwide observation network based on unified specifications. Broadband, high accuracy technology was adopted for seismic observation, and quasi-real time measurement using GPS was applied to crustal movement. As a result, in less than 10 years, Japan succeeded in constructing a high performance, high density observation network unprecedented in the world. The content of what came to be called the "Kiban Network" (basic nationwide seismic network) comprises a crustal movement observation network using GPS (approximately 1400 points nationwide, centering on GEONET by the Geographical Survey Institute), a high sensitivity seismograph observation network (approximately 1000 points, centering on Hi-net of the National Research Institute for Earth Science and Disaster Prevention (NIED)), a strong ground motion observation network (approximately 7000 points including the NIED's K-net), and a broadband seismograph observation network (approximately 100 points, centering on NEID's F-net). It is particularly noteworthy that all the data obtained by this Kiban Network are publicly available via the internet. As a result, researchers throughout the country are placed in virtually the same research environment concerning the availability of data, irrespective of the university or institute to which they belong, and this at a stroke has accelerated free competition in research. Now, when a destructive earthquake or other earthquake which should be considered a problem occurs, the basic information characterizing the earthquake, such as the geometry of the fault system, rupture process, etc., is analyzed and reported extremely quickly. Figure 3 shows the transition in the number of earthquakes captured in and around Japan from the earthquake catalog of JMA (the Japan Meteorological Agency). This does not mean that there was a particularly great change in seismic activity as such during this period, but the number of earthquakes captured has increased steadily as the observation network was expanded. In particular, it can be understood that the increase was very rapid from around 2000, when the Kiban Network came on-stream.

Furthermore, in recent years, the reach of research and observation has expanded from land to ocean areas. JAMSTEC (Japan Agency for Marine-Earth Science and Technology) constructed a deep sea drilling vessel called the "Chikyu" and has begun research on the deep crustal structure of the earth in ocean areas, and is also promoting the development



* The rapid increase from around the year 2000 was due to the implementation of the Kiban Network.

Figure 3 : Number of earthquakes captured in and around Japan by the Japan Meteorological Agency

Source : Reference [12]

of a sea-bottom seismograph network to cover Japan's coastal seas.[17] The Japan Coast Guard and universities are engaged in the development of a crustal movement observation network in ocean areas by combining GPS and acoustic ranging devices installed on the sea bottom.[18]

It can be said that the construction of a nationwide observation network, beginning with the Kiban Network, has had a great effect on the promotion of earthquake research as a whole, including the disaster prevention aspect, but this does not mean that there are no problems in these efforts. The following two points may be mentioned as misgivings of the author, mainly in connection with the operational aspect. In the "age of rival warlords" before the construction of the Kiban Network, the former Imperial universities had substantial jurisdiction over their respective regions, separately from the nationwide network administered by JMA. For example, Hokkaido University had jurisdiction in Hokkaido, Tohoku University in Tohoku, Tokyo University in Kanto and Shinetsu, Nagoya University in Chubu, Kyoto University in Kinki, Chugoku, and Shikoku, and Kyushu University in Kyushu. This method of dividing jurisdictions led to partitioning of the data, creating barriers to research activities, and became a starting point for reflection after the Great Hanshin-Awaji Earthquake. Conversely, however, this system also was a factor that produced a consciousness and sense of responsibility as a "home doctor," because each university was responsible for monitoring and evaluating the crustal movements in its own area. In contrast, the present system has produced a situation of excessive competition, with connotations of a certain kind of waste, in that researchers throughout the country are engaged in exactly the same analysis using identical data. This can perhaps be called progress, in the sense of speeding up processing and unifying information management. However, from the standpoint that one wishes to recommend monitoring and research from a long-term perspective by a "home doctor," who has a thorough knowledge of the tectonics (structural motion, represented by plate tectonics) and condition of activity of that researcher's own region, the feeling of a kind of dilemma is unavoidable.

Secondly, there are also differences in the time scale in the cycle of earthquake occurrence and in the speed of technical innovation. Considering the fact that one cycle of an earthquake is at least several decades to several centuries, continuation of observation over the long term under stable conditions is an essential condition. However, it is difficult to avoid changes in observation conditions due to technical innovation and the evolution of the system. This is also a dilemma. Technical innovation does not necessarily lead to good results. For example, the Earthquake Catalog (Figure 3) prepared by the Japan Meteorological Agency, which covers more than 80 years, is an invaluable data resource of which Japan can boast to the world, but due to technical innovations and changes in the system, the magnitude shown in the Catalog is not uniform. This has greatly reduced the value of the data so painstakingly collected. Long-term maintenance and operation of the Kiban Network, which is deployed on a large scale, is also accompanied by greater difficulties than construction and maintenance of the system. It is necessary to be aware that preserving the consistency of observations unaffected by changes in the larger environment is an important challenge for the future.

3-2 Preparation of Seismic Hazard Maps

HERP is engaged in the preparation of a "National Seismic Hazard Map of Japan" as a fundamental part of its mission.[15] As mentioned previously, "forecasting," as the term is used here, means an assessment of the probability of the occurrence of the next earthquake preconditioned on an assumption that earthquakes occur with quasi-regularity in a given location. Although the existence of precursory phenomena is still a matter of debate, the majority of researchers support the quasi-regularity of earthquakes. This is also a reason for the change in the policies of administrative authorities from "earthquake prediction" to "earthquake forecasting." The information necessary for an assessment comprises three parameters, namely, the earthquake occurrence period and its deviation and the date of the most recent earthquake for the earthquake in question. In addition, a statistical model which expresses the deviation from regularity is also necessary. HERP has adopted a BPT (Brownian Passage Time) model in which the rate of accumulation of a certain amount of stress is affected by random disturbances.

Based on these arrangements, the procedures which can actually be promoted are as follows. Although the earthquakes which occur in ocean areas, for example, along ocean trenches, have a scale on the order of

M8, their cycle is short, at several decades to several 100 years. Accordingly, evidence of many such earthquakes can be found in the historical record. As a result, assessment of the probability of occurrence is comparatively easy, and reliability is high. On the other hand, earthquakes which occur at shallow active faults in inland areas have long time scales of at minimum 1000 years or more, and their histories are virtually unknown. HERP specified 98 active faults with lengths exceeding 20km as major active faults (Figure 4; total of 110 as a result of subsequent additions), and conducted a survey of these faults. It then attempted to determine the values of the above-mentioned parameters for each of the active faults by performing trench excavation surveys (survey by excavating a shallow trench) and boring surveys (survey by boring a deep hole). In actuality, cases in which the values of parameters are determined conclusively are rare, and considerable deviations and indeterminacy cannot be avoided. Nevertheless, the provisional results were compiled over a period of 10 years. The numerical distributions of the magnitudes of the earthquakes which are assumed to occur here are as shown by the black bars in Figure 5, and their average magnitude is M7.3. (The total number is 136 because it is considered that earthquakes will occur in segments of the long and large faults among the 98 major active faults.) An assessment of active faults smaller than the major faults was also made, resulting in an average of M6.8 (shown by the white bars in Figure 5, total number of 178). In addition to these, as "earthquakes having a sources which are difficult to designate the seismic source," the probability of occurrence of "problem earthquakes" was calculated from the actual measured distribution of the magnitude of the earthquakes and largest earthquake set in each region. The result of combining all of the above assessments is considered to be the earthquake occurrence probability at each location.

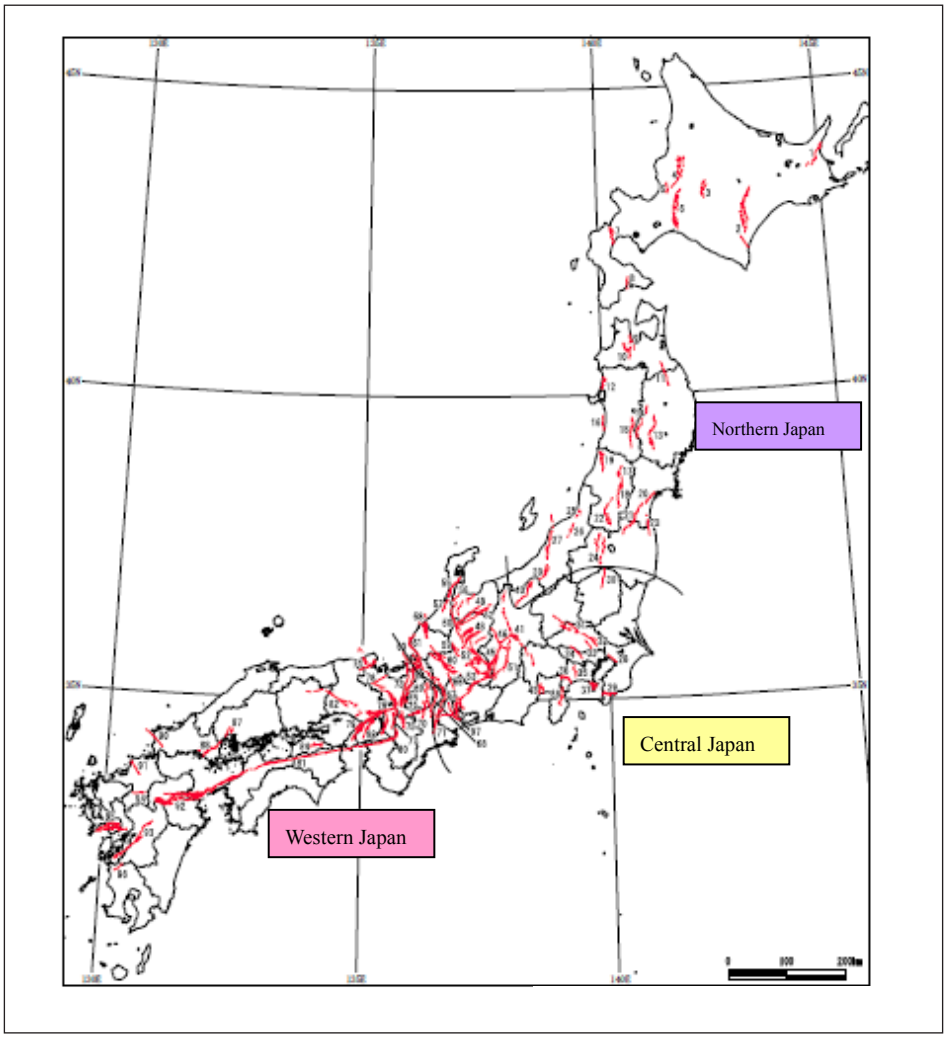


Figure 4 : Major active faults listed by the Headquarters for Earthquake Research Promotion

Source : Reference [15]

If the occurrence probability of earthquakes can be obtained, the procedure moves next to assessment of the motion at each location, that is, ground motion. Here, the final surface ground motion is calculated using a combination of various assessment techniques, including evaluation equations for motion based on the magnitude of the earthquake and the distance from the source, or a seismic wave synthesis method based on a fault model, and assessment of the amplitude of the seismic wave due to the subsurface structure and basement structure, etc. Local governments determine damage assumptions and disaster prevention countermeasures based on these results.

The detailed description of these procedures will be omitted here. However, the first trial edition of

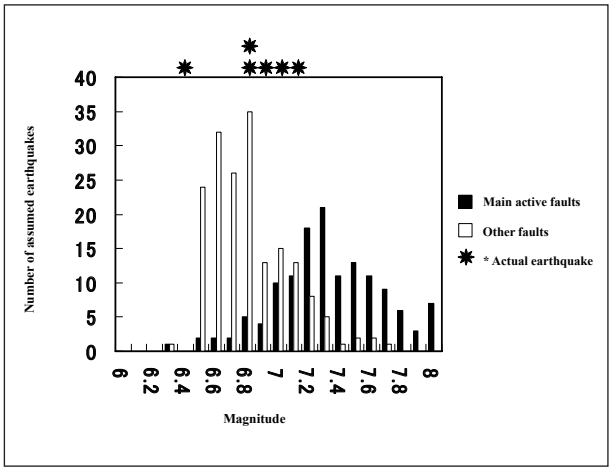
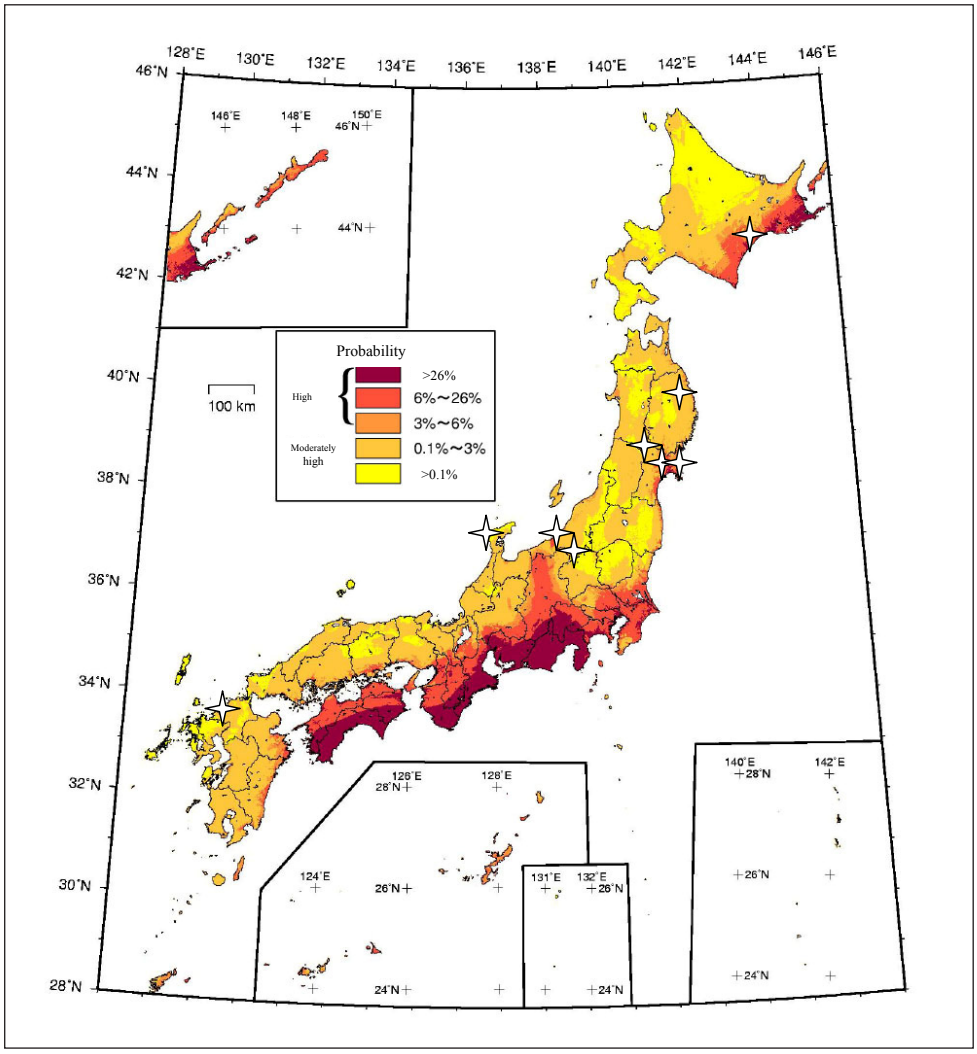


Figure 5 : Numerical distribution of magnitudes of assumed earthquakes occurring at inland active faults

Source : Reference [15]



Distribution of the probability of seismic motion of seismic intensity 6-Lower or higher within the next 30 years. White stars show the locations of actual 6-Lower and higher earthquakes in the most recent 5 year period. For the original map, see:
http://www.j-shis.bosai.go.jp

Figure 6 : Seismic Hazard Map

Source : Reference [15]

the "National Seismic Hazard Map" of Japan was published in May 2002, and the complete edition was published in March 2005, 10 years after the Great Hanshin-Awaji Earthquake (Figure 6).[19] Although the Hazard Map has been revised annually since that time, there have been no major changes. This map shows the probability of seismic intensity 6 Lower and higher earthquakes within the next 30 years in each region by color. The probability of occurrence of a 6 Lower earthquake, in other words, the risk of disaster, can be understood in 5 steps from the lowest probability of less than 0.1% to the highest probability of 26% or more. Almost all of the areas with the highest probability are distributed in a region lying along the Pacific Ocean from Shizuoka Prefecture to Kochi Prefecture. This is due to an impending ocean trench-type earthquake (Tokai-Tonankai-Nankai earthquake) along the Nankai Trough. In the region with the 2nd highest risk, with the exception of the Fossa Magna in central Japan, all lie along the Pacific coast. This is testimony to the fact that the frequency of inland active fault earthquakes differs by one order from that of ocean trench-type earthquakes.

3-3 Results of Seismic Hazard Map

As of 2008, only 3 years had passed since the publication of the complete edition of the Seismic Hazard Map. Accordingly, the accumulation of events is still inadequate for an evaluation. However, because problem earthquakes have also occurred during this period, it is necessary to evaluate the results by some method as a basis for future revisions. In other words, has "forecasting" really been successful?

Since the publication of the Seismic Hazard Map, including the trial edition, a total of 10 seismic intensity 6 Lower and higher earthquakes (shown by the white stars in Figure 6) actually occurred during the 5 year period up to October 8, 2008. Table 2 shows how the occurrence probability of these earthquakes was assessed in the Hazard Map using numerical values (30 year probability) which the author studied from a published report.[19]

First, because high occurrence probabilities of 60% for the Tokachi-oki Earthquake of September 2003 and 98% for the Miyagi-Oki Earthquake of August 2005 were shown in advance, it can be said that these interplate earthquakes occurred basically as forecast (for the Tokachi-Oki Earthquake, the value was immediately before occurrence of the earthquake; Miyagi-Oki is regarded as a fragment of the assumed earthquake). Among the remaining 8 earthquakes, the Miyagi-Oki Earthquake of May 2003 and the Iwate-Chubu Earthquake of July 2008 can be regarded as earthquakes within the Pacific Plate, and the Miyagi -Hokubu earthquake of July 2003 and the Niigata-Chuetsu Earthquake of October 2004 can be regarded as earthquakes at active faults other than the main active faults. The remaining four earthquakes were "unexpected" active fault earthquakes which occurred at an active fault in a coastal region or at an inland concealed fault (blind thrust fault). However, while these were also categorized as "earthquakes having sources that are difficult to designate the seismic source," this does not mean that there was no awareness of the occurrence of the earthquake as such.

Table 2 : Earthquakes of seismic intensity 6-Lower and higher in most recent 5 year period

		Assumed M		Probability of occurrence (%/30 years)	
May, 2003	Miyagi-oki Earthquake	M7.1	—	0.2～0.5	(intraplate earthquake)
Jul, 2003	Miyagi Hokubu Earthquake	M6.4	M7.1	0082	(Asahiyama flexure belt)
Sep, 2003	Tokachi-oki Earthquake	M8.0	M8.1	60	(interplate earthquake)
Oct, 2004	Chuetsu Earthquake	M6.8	M7.1	0.79	(Muikamachi fault zone)
Mar, 2005	Fukuoka Hokusei-oki Earthquake	M7.0	—	0.05	(coastal active fault)
Aug, 2005	Miyagi-oki Earthquake	M7.2	Mw7.4	98	(interplate earthquake)
Mar, 2007	Noto Hanto-oki Earthquake	M6.9	—	0.02～0.05	(coastal active fault)
July, 2007	Chuetsu-oki Earthquake	M6.8	—	0.2	(coastal active fault)
Jun, 2008	Iwate-Miyagi Nairiku Earthquake	M7.2	—	0.1～0.2	(geological fault)
Jul, 2008	Iwate Chubu Earthquake	M6.8	—	0.1～0.2	(intraplate fault)

Prepared by the STFC based on Reference [3]

The first two earthquakes mentioned above were ocean-trench type interplate earthquakes, and their occurrence probabilities were high. On the other hand, even assuming that the remaining 8 were supposed to be possible, their probabilities were considered to be extremely low. As a result, in a certain sense, it may be natural that the forecasts of the former were "hits" and the latter were "misses." Nevertheless, there are perhaps problems in this kind of simplistic summary.

Essentially, in order to assess "hits" and "misses" correctly, a complex statistical analysis is necessary. Here, however, only a simple trial calculation was performed. From Table 2 the average of the 30 year probability of the 8 earthquakes is 0.226%. Assuming the extent of the source of an average magnitude M7.0 earthquake is 200km2, the entire country of Japan would be divided into 1850 zones. If the probability of earthquakes occurring in 8 or more zones in a 5 year period is calculated for this case, the value is extremely small, at 0.0001%. In other words, in effect, the occurrence of these 8 earthquakes should be virtually impossible based on the assumptions. This indicates that there was some kind of problem in the assumptions.

3-4 Problems in seismic hazard map

In the preparation of the Seismic Hazard Map, "earthquakes at smaller active faults" and "earthquakes having sources (active faults) which are difficult to designate" were also considered. However, as discussed in the previous section, the results give an undeniable impression that earthquakes of around M7 occurred unexpectedly. Beginning immediately after the Noto Hanto-Oki Earthquake of March 2007, a number of newspapers published editorials to the effect that, "In Japan, it is impossible to escape earthquakes no matter where you live. The fact that the Seismic Hazard Map shows a low hazard level does not mean you can feel secure." Considering the level of the existing forecasting technology, the evaluation of the Seismic Hazard Map is extremely severe. It would also be hasty to evaluate the Seismic Hazard Map, which is essentially intended to be long-term information, based on the results from only 5 years. On the other hand, the Seismic Hazard Map is not a research paper, but rather, is information which the administrative authorities and researchers have presented to society. For this reason, how this information is received by society must not be ignored. At minimum, it is necessary to listen to criticisms that the forecasting of earthquakes at inland active faults was unsuccessful. Therefore, let us analyze the location of the problems in the current stage.

The first problem is that the active faults which cause earthquakes were not fully adequately captured. In Figure 5, the star marks indicate the magnitudes of six active fault earthquakes which actually occurred, overlaid on an assumed magnitude distribution map which was used in the Seismic Hazard Map. Comparing the form of distribution, it can be understood that these six earthquakes did not occur at the major active faults, but correspond to earthquakes occurring at smaller active faults other than the main faults. Among these, it seems possible to say that two earthquakes were assumed in advance, these being the July 2003 Miyagi-Hokubu Earthquake (Asahiyama flexure belt) and the October 2004 Chuetsu Earthquake (Muikamachi fault zone), but the remaining four were clearly missed. In Figure 5, the average magnitude of the earthquakes occurring from the major active faults was M7.3, while that of the earthquakes at active faults other than the major faults was M6.8, or a difference of 0.5 by magnitude. From this, it can be estimated that the earthquake occurrence frequency of the latter is $\sqrt{10}$ times that of the former. Furthermore, from the fact that the average interval between the occurrence of the former is 4200 years (harmonic average value), and that of the latter is 5300 years, if the number of active faults which are sources of these earthquakes is calculated, the number of active faults other than the major faults should be approximately 4 times that of the major faults (in this case, 136 major active faults), in other words, more than 500. In actuality, the number listed (178) is only about 1/3 of this number. Accordingly, it can be inferred that the majority of smaller active faults were not captured.

From the list of active faults in Chart 8, of the four earthquakes which were missed, it is considered that three were due to active faults in coastal areas, and one was due to a concealed fault which was not identified from the surface geological survey. Based on this result, On July 13, 2008, HERP announced a policy of identifying a total of 2000 active faults nationwide over the next 10 years, focusing on active faults in coastal areas and identification of concealed faults that seem to exist along the extensions of the main active faults.[21] However, it is estimated

that many of the object active faults are less than 10km in length, and it is highly possible that small-scale active faults of this type will not appear on the surface. Therefore, there have been suggestions that, in addition to surface surveys based on tectonic geomorphology, comprehensive use of information on the subsurface structure may be necessary, including reflection seismic surveys, electromagnetic seismic surveys, seismic wave velocity structure, and gravity distribution. Although improvement of the Seismic Hazard Map by future surveys is awaited, even in this case, it will be necessary to resign ourselves to the possibility of earthquakes which are omitted from those assumed.

As a further problem, simultaneously with the difficulty of identifying active faults, it is also difficult to evaluate activity. Activity evaluation means estimation of the frequency of occurrence of characteristic earthquakes at individual active faults. In the case of the major active faults, information can be obtained directly by trench surveys (investigation by excavating a shallow trench), but it is extremely difficult to obtain adequate information on small-scale active faults whose positions are difficult to designate. As mentioned in the previous section, the forecast value of earthquakes of around M7 as now published is overwhelmingly smaller than their actual frequency. This may not be caused only by a failure to identify faults. Systemic problems in the earthquake occurrence probability evaluated from the activity of active faults are also a possibility. In the past, Wesnousky et al. pointed out that there is a discrepancy between the crustal movement rate at the geological time scale based on activity evaluation of active faults, and the movement rate at the scale of decades to centuries based on observations of crustal movement and seismic activity.[22] This problem still has not been solved. Although it goes without saying that information on active faults is fundamental to long-term earthquake forecasting, research from the viewpoint of explaining, without contradiction, the results of observations of seismic activity and crustal movement in recent years must not be undervalued. It would appear necessary to compensate for omissions in evaluations of active faults by unifying this type of research.

4 New research on earthquake prediction

4-1 New trends

The impact of the Great Hanshin-Awaji Earthquake was not limited to administration, but also spurred new action in the earthquake prediction researchers' group. The desire of the researchers' group to explore new approaches in prediction research coalesced in the Coordination Council for Earthquake Prediction Research in Universities, which is an organization that coordinates activities between universities and subsequently played a core role in promoting earthquake prediction research, including liaison with the administrative authorities. The direction of research which was recommended through this organization abandoned the phenomenological approach of exclusively searching for precursory phenomena followed up to that time, and returned to an approach based on physical science, which attempts to elucidate the process by which earthquakes actually occur. When the site of occurrence of one earthquake is imaged, its temporal evolution process is divided into three stages, these being a "process of stress release by the occurrence of the earthquake and recovery of a locked state," "preparatory process for occurrence of the earthquake," and "process immediately before occurrence of the earthquake." A step-by-step research strategy was delineated for each of these stages, in which the phenomena are reproduced and predicted by observation, analysis, interpretation, and simulation, and this was adopted as a roadmap for achieving prediction. A number of research groups were established by stage or by item, and a large number of research meetings and symposiums were held. Although the discoveries and results during the past 10 years have exceeded the original expectations, these are supported by the existence of the Kiban Network.

4-2 Discoveries which are key to prediction

Among the new discoveries of the past 10 years, two which are particularly important from the viewpoint of earthquake prediction are the discovery of the phenomenon of "slow slip" and recognition of "asperities." Slow slip, as the words imply, refers to a slip which occurs slowly around a fault. In contrast to true earthquakes (in the conventional sense), in

which slip between plates or slip of a locked zone on a fault plane occurs suddenly in a time of several seconds to at most 1 minute, in a slow slip event, slippage continues gradually for several days or as much as several years. Slips include kinds with a variety of time constants. The existence of phenomena called "slow earthquakes" and "silent earthquakes" had be predicted previously.[23] However, this type of long-period movement is difficult to detect with conventional seismographs. Therefore, in order to capture these phenomena as facts, it was necessary to construct the Kiban Network with the GEONET, F-net, and Hi-net systems. Recently, a series of discoveries have been made, beginning with long-period slow slip events having a time constant of several years, and also including slips distributed over a wide range of frequencies, such as short period slow slips, ultra-low frequency earthquakes, low frequency earthquakes, and low frequency tremors, and it has been found that these phenomena appear in locations determined by their mutual linkage.[24]

At the same time, the existence of asperities was also demonstrated. Asperity is a key concept for understanding the slip process of locked zones. Asperities are the parts of the total locked zone on a fault plane or between plates which are locked particularly strongly. Although this concept itself had been argued since an early date,[25] the actual existence of asperities could not be detected until higher accuracy was achieved in observation techniques.[26,27] This research demonstrated that asperities are not extinguished, but rather, undergo a process of rupture and recovery of the locked state which is repeated any number of times. It has also been possible to arrive at an interpretation that, in large earthquakes exceeding M7, multiple asperities exist around the source, and differences may occur in the rupture mode in the same series of earthquakes, depending on how these asperities combine. Another idea which is continuing to gain acceptance is that a weakly locked zone exists between pairs of asperities, and that part, which had been locked in the stage of the preparatory process, slips in a quasi-static manner prior to the occurrence of an earthquake. Furthermore, actual observation has confirmed that there is a hierarchy in the sizes of asperities, and "repeating earthquakes" of M3 to M5 scale occur with high regularity at very small asperities.[28]

4-3 Promotion of simulation research

In parallel with observation, great strides have been made in computer simulations based on numerical equations which describe frictional sliding.[29] It has become possible to simulate the complex and diverse phenomena which actually occur by creating repeatedly-occurring earthquakes hypothetically on a computer and adjusting the boundary conditions. For example, it is possible to simulate repeated rupture of the same asperity, create an earthquake series with various aspects due to ruptures involving different combinations of multiple asperities, reproduce slow slip events, etc. In addition, a project aimed at creating models of the tectonics and crustal movement of the entire Japanese archipelago[30] was carried out using the supercomputer "Earth Simulator".

The purposes of simulation research are not limited to interpretation of observed phenomena. First, a forecast of near future events is calculated by inputting observed values showing the current conditions as initial values. The various parameters defining the simulation are then adjusted by comparing the results of this forecast and the actually-observed conditions (this process is called "data assimilation"). This process is repeated and the accuracy of the forecast is improved until a practical earthquake prediction is achieved. In this manner, simulation is incorporated in the roadmap for earthquake prediction research as an indispensable technique.

4-4 Feasibility of earthquake prediction

By following a roadmap which includes the four stages of future forecasting by observation, analysis, interpretation, and simulation, earthquake prediction research has achieved steady development. Thus, it can correctly be said that the new direction of research, namely, attempting to elucidate the process by which earthquakes occur based on physical science, is bearing fruit. For example, the cyclical process of the characteristic earthquake which was discovered in the offshore area at Sanriku Kamaishi was analyzed in detail, arriving at a point where it was basically possible to predict the next occurrence.[31] Tokai-Tonankai-Nankai earthquakes occur cyclically with a period of somewhat longer than 100 years, but the pattern of occurrence of these three earthquakes is not the same each time. It has also become possible to reproduce this irregular occurrence pattern by simulation.[32]

Regarding the current status of prediction research, can it be said that research is approaching its target by following the roadmap laid out earlier? In other words, are we near the day when practical earthquake prediction will be possible? The answer is no. It is not a fact that researchers have actually succeeded in earthquake prediction. The road to universal prediction still cannot be seen.

For example, one concrete problem is "preslip." Preslip refers to a type of slip that occurs immediately before slip of the entire source area, including the rupture of the asperities at its nucleus, and starts slowly from the beginning and then gradually accelerates. Detection of this phenomenon is considered the leading candidate for advance prediction. It can be said that the expectations placed on preslip detection are one result produced by simulation research, but in fact, preslip has not been confirmed even once in actual situations. Preslip was not discovered in any of the 10 earthquakes mentioned in Table 2. Moreover, this is not limited to Japan; as of this writing, there have been no reports of the capture of preslip anywhere in the world. However, it is still early to issue a conclusion. High sensitivity instruments which are actually capable of detecting preslip have only been installed in a very small number of locations. The conditions were most favorable in the September 2003 Tokachi-oki Earthquake (M8.0), but in this case as well, the nearest tiltmeter to the source was installed more than 100km away. In the case of inland active fault earthquakes, the observation points were closer, but it is thought that preslip could not be detected due to the small size of the earthquakes themselves. As will be discussed below, prediction of a Tokai earthquake is premised on detection of preslip. However, in this case, the fact that an M8 class source will exist under the special condition of being inland contributes to the expectations that preslip detection will be possible, although with much difficulty.

In addition to the above, there is also the problem of seismic quiescence. In the verification by Prof. Wyss introduced in section 2-2, quiescence of seismic activity was recognized as the most probable precursory phenomenon.[33] Actual quiescence is related to a large number of derivative conditions, including the scale of the object background earthquake, the area and statistical significance of the quiescence, the homogeneity of the database, etc.. Therefore, the resultant quiescence information exists in a mixed state. The possibility that quiescence may be one of the few meaningful precursory phenomena is a point that many researchers acknowledge, but there is still no established theory explaining the mechanism by which it occurs. Although there are moves to interpret quiescence using the new concepts of asperity and quasi-static slip, given the current conditions, it cannot necessarily be said the research in this field is progressing.[34]

Looking at trends in scientific societies, it can be understood that a large number of researchers are not actively involved in earthquake prediction research. Figure 7 shows author's classification of approximately 600 titles presented at the 2007 Fall Conference of the Seismological Society of Japan. According to this, in excess of 50% of the titles were classified as either "Structural analysis" or "Fault model/rupture process/source mechanism." When "Geological structure/tectonics" is included, analysis on so-called spatial information totals 64%. In contrast to this, "Change in seismic activity/crustal movement/earthquake prediction," which should be called temporal information, account for only 10%. "Structural analysis" is the foundation of earthquake research, and is also the starting point of research on earthquake prediction. However, when aware of the new research strategy which has begun to investigate the various processes up to the occurrence of earthquakes by the new approach based on physical science, the current division of topics is biased, and appears to show a lack of balance. In earthquake prediction research, analysis/research on data based on long term observation and monitoring is fundamental due to the nature of the research. The recent climate that demands quick results affects researchers in their choice of topics, and there may also be structural factors that cause researchers to avoid research on subjects like earthquake prediction, for which results cannot be promised in advance.

Phenomenological prediction research has been inactive worldwide since the 1990s. Today, however, this area is continuing to produce new trends, particularly in the United States. As the object of this work is prediction of earthquakes in California, this may represent a mutual competition among individual researchers, in which researchers propose respective forecasts under the same set of conditions.[35] However, this trend has also had a ripple effect in Europe, and moves in response to these trends have begun to appear in Japan. Thus,

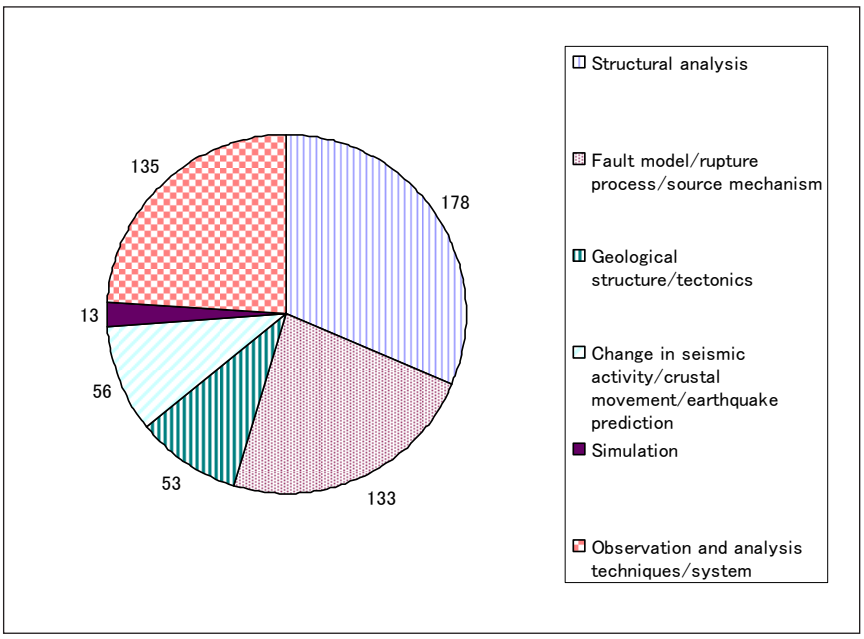


Figure 7 : Percentages of research fields judged from program of 2007 fall meeting of the seismological society of Japan

Prepared by the STFC

it may be possible to say that this is a revival of the research technique of at least attempting to extract precursory phenomena from observed events, which was rejected once in the past, against the background of new knowledge and technologies.

4-5 Prediction of Tokai earthquake

In Japan, the sole case in which the administrative authorities have acknowledged the possibility of prediction is the Tokai earthquake. In actuality, even assuming the appearance of preslip, the Tokai earthquake is the only earthquake with observation conditions that make it possible to capture this phenomenon. In March 1998, the Earthquake Assessment Committee for Tokai earthquakes revised its standard for convening, which is invoked when anomalous activity is detected.[16] If preslip is detected, the Assessment Committee will convene, based on an assumption that an earthquake will occur within 72 hours. It can be said that this is one result of progress in simulation techniques. Of course, this is ultimately only a result of simulation and is not backed by fact. Therefore, even assuming actual preslip is detected in the Tokai region, there is a remaining element of anxiety, in that the prediction based on this will itself be an "unrehearsed performance," i.e., a prediction of a major disaster based on an unproven assumption. There are several problems, including this, which stand in the way of prediction of a Tokai earthquake.

The fact that expectations for prediction of a Tokai earthquake are placed on the appearance of preslip is based on the prior example of the 1944 Tonankai Earthquake. The day before this earthquake, an unexpected change in inclination was observed in measurements of the water level around Kakegawa.[36] However, questions were raised as to whether this observation was a true crustal movement or not.[37] Moreover, even assuming this change in inclination was actual, no convincing reason was given for the fact that the observation was made at the Kakegawa, which is located at a considerable distance from the source.

According to the forecast by HERP, the 30 year probability of a Tokai earthquake is 87% (reference value).[19] This result is based on the fact that the average interval between Tokai earthquakes, which have occurred four times since the 1498 Meio Earthquake during the Muromachi period, is 119 years. Actually, however, apart from the previous Ansei Tokai Earthquake of 1854 and the Hoei Earthquake of 1707 which preceded it, there are questions about the existence of earlier Tokai earthquakes than this, and the interval between occurrences may be longer. In other words, there is a possibility that the current probability is much smaller than that in the HERP forecast. This is also related to the suggestion that the relative velocity

of the plates in the assumed source area of a Tokai earthquake may be smaller than originally assumed (4cm/yr). For example, there is a theory that part of the Philippine Sea Plate on which the Izu Peninsula rides is a microplate that has separated from the main body.[38] According to this theory, the relative velocity is assumed to be 2cm/yr, and if this is true, the interval between the occurrence of earthquakes would be two times the issued value.

Furthermore, many researchers think that there is a high possibility of linkage between Tokai earthquakes and Tonankai and Nankai earthquakes. This is because, historically, there are no examples in which a Tokai earthquake occurred independently. In the current condition, the probability of the next Tonankai earthquake is 60%. From this forecast, the crisis point should come around 2030. Assuming linkage between these earthquakes, the next Tokai earthquake will be triggered by a Tonankai earthquake, which means that its occurrence should be delayed until another Tonankai earthquake occurs.

As this discussion suggests, many difficult questions regarding the feasibility of predicting a Tokai earthquake remain unanswered. However, this notwithstanding, it can be said that the Tokai earthquake is the closest to prediction. Beginning around the year 2000, the largest slow slip in the history of observation occurred directly under Hamana Lake, which is adjacent to the assumed source, and simultaneously with this, significant changes were seen in the condition of locking in the source area. Using the dense GPS network, which is prominent even in Japan's high density observation network, it has become possible to grasp such movements in detail.[39] In other words, changes in the current condition are being captured on a moment-by-moment basis only in the case of the Tokai earthquake. Moreover, simulation research on the Tokai earthquake is gradually approaching a realistic level. Conversely, if prediction of a Tokai earthquake is thought impossible even in light of these conditions, this is equivalent to denying the possibility of earthquake prediction as such. In this sense, the Tokai earthquake must be considered a touchstone for earthquake prediction in general.

4-6 Prediction of Ibaraki-oki earthquake

On May 8, 2008, an M7.0 earthquake occurred off the coast of Ibaraki Prefecture. Here, it is known that six earthquakes have occurred virtually periodically at intervals of more than 20 years from the first historical earthquake in 1896 up to the present. Furthermore, from an analysis of the seismic waveform, it is also known that the same asperity ruptured in the most recent earthquake and the previous M7.0 earthquake which occurred in 1982.[27] It has been reported that this asperity appears to have formed with a relationship to a subducting seamount.[40] Our understanding of the Ibaraki-oki earthquakes has increased dramatically based on these facts and discoveries and the new recognition of the process by which earthquakes occur developed up to the present. Changes in seismic activity were detected before the event in this earthquake,[41,42] and there is ample reason to expect that some type of advance information will be possible when the next Ibaraki-oki earthquake arrives in about 20 years.

5 Conclusion

In "earthquake forecasting," for which the administrative authorities are responsible, a Seismic Hazard Map of Japan has been produced as the final result of a series of analytical techniques prescribed basically as manual-like procedure. If problems arise in the forecast results, this is due to inadequacy of the information used, and further study is judged to be necessary. In reality, however, the problems are not limited to an inadequate amount of information. Issues at the research level remain in the analytical techniques used in the preparation of the Hazard Map.

On the other hand, in "earthquake prediction," which has been promoted during the past 10 years based on a new direction, great progress has been made in elucidating the process by which earthquakes occur based on a physical science approach, focusing particularly on the fault rupture process. Even assuming that generalization of pinpoint prediction, as in a Tokai earthquake, is still not possible, in comparison with simple "earthquake forecasting," it may be permissible to say that "earthquake prediction" exists even today, in the sense of providing supplementary information based on analysis of stress conditions. However, this information still lacks effectiveness, and in this sense, there is a large gap between the reality of "earthquake prediction" on which researchers have fixed their gaze, and the "prediction" expected by society.

Recently, an article called "Earthquake Prediction – A Complete Change in Assumptions," which was published by the Asahi Shimbun newspaper, reported this fact and explained in simple terms the difficulty of earthquake prediction research.[43] The author did not interpret this as a negative statement on "prediction," but rather, saw it as an article which gave a feeling again of the strong interest of society in earthquake prediction. In other words, the motivation for prediction research is not simply the scientific interest of researchers.

The example of the Ibaraki-oki earthquake introduced in section 4-6 truly shows that, in prediction research, the level of research as a whole is frequently raised at once by the occurrence of a target earthquake. Conversely, under ordinary conditions, progress in research appears to advance slowly. If the difficulty of achieving earthquake prediction can be seen, the true nature of that difficulty lies in the fact that the results cannot be seen in advance, as is true in all creative research.

References

[1] Asahi Shimbun, May 13, 2008. (Japanese)

[2] K. Abe and S. Noguchi, "Determination of magnitude for large shallow earthquakes, 1898-1917, " Phys. Earth Planet. Inter., 32, 45-59 (1983). (Japanese)

[3] Y. Okada, "Seismological Maps of Japan,"Tokyo Shoseki Co., Ltd. (2004). (Japanese)

[4] T. Tsukuda, "Recent Science of Earthquake Prediction," Science-eye, SIS-039 (2007), (Japanese)

[5] Observation by the VAN method : http://contest.thinkquest.jp/tqj2000/30295/prediction/van.html

[6] Hamada, K., Statistical evaluation of the SES predictions issued in Greece: alarm and success rates, Tectonophysics, 224, 203-210 (1993).

[7] R. J. Geller, "Shake-up for earthquake prediction", Nature, 352, 275-276 (1991).

[8] The Parkfield,California,Earthquake Experiment : http://earthquake.usgs.gov/research/parkfield/index.php

[9] "The Coordinating Committee for Earthquake Prediction Japan: 30 Years of Progress," Ministry of Construction, Geographical Survey Institute (2000). (Japanese)

[10] Y.Okada,"Regions Designated by the Coordinating Committee for Earthquake Prediction Japan and Earthquakes in Recent Years," Report of the Coordinating Committee for Earthquake Prediction Japan, 79, 626-628 (2008). (Japanese)

[11] K. Ishibashi, "Study of Major Earthquakes Expected in the Tokai Region – Regarding the Suruga Bay Great Earthquake – ," Proceedings of the Conference of the Seismological Society of Japan, No. 2, 30-34 (1976). (Japanese)

[12] Special Measures Law for Countermeasures against Large Earthquake : http://www.bousai.go.jp/jishin/tokai/houritsu/taishin/contents.htm

[13] K.Hamada,"Earthquake Prediction,"Morikita Publishing Co., Ltd. (1986). (Japanese)

[14] M.Ohtake,T.Matsumoto,and G.V.Latham, "Seismicity Gap near Oaxaca, southern Mexico as a probable precursor to a large earthquake," Pure and Appl. Geophys., 115, 375-385 (1977).

[15] Headquarters for Earthquake Research Promotion : http://www.jishin.go.jp/main/index.html (Japanese)

[16] Central Disaster Management Council, Expert Research Committee on Tokai Earthquake : http://www.bousai.go.jp/jishin/chubou/tokai/index.html (Japanese)
Standard for convening the Earthquake Assessment Committee of the above : http://www.bousai.go.jp/jishin/chubou/tokai/3/4-01.html (Japanese)

[17] Japan Agency for Marine-Earth Science and Technology : http://www.jamstec.go.jp/j/ (Japanese)

[18] Hydrographic and Oceanographic Department, Japan Coast Guard : http://www1.kaiho.mlit.go.jp/KOHO/chikaku/kaitei/sgs/detail.htm (Japanese)

[19] Trial edition (2003 Ed.) of "National Seismic Hazard Map of Japan" :
http://www.jishin.go.jp/main/choukihyoka/03mar_kakuritsu/setsumei.pdf
Complete edition of the above (2005 Ed.) :
http://www.jishin.go.jp/main/chousa/05mar_yosokuchizu/bunsatsu1.pdf
[20] Japan Seismic Hazard Information Station :
http://www.j-shis.bosai.go.jp/ (with link to English)
[21] Yomiuri Shimbun, July 13, 2008. (Japanese)
[22] S.G.Wesnousky, C.H. Scholz, and K. Shimazaki, "Deformation of an island arc: Rates of moment release and crustal shortening in intraplate Japan determined from seismicity and quaternary fault data," J. Geophys. Res., 87, 6829-6852 (1982).
[23] I.Kawasaki,"What is a Slow Earthquake? " NHK Books, 1055 (2006). (Japanese)
[24] I.Kohara,"Detection of Short-Period Slow Slip Event Simultaneously with Deep Low Frequency Tremor – Result by Inclination Observation with National Research Institute for Earth Science and Disaster Prevention Hi-net – ," Journal of the Geodetic Society of Japan, 53, 25-34 (2007). (Japanese)
[25] H.Kanamori,"The nature of seismicity patterns before large earthquakes, Earthquake Prediction," Maurice Ewing Series 4, 1-19 (1981).
[26] "Designation of 4 Asperities Considered Sources of Tokai Earthquake," Science and Technology Trends: Topics, December 2007, No. 81,p.9 (2007). (Japanese)
[27] Y. Yamanaka, NGY Seismology Note :
http://www.seis.nagoya-u.ac.jp/sanchu/Seismo_Note/ (Japanese)
[28] T. Matsuzawa, "Strategies and Scope of Earthquake Prediction," Journal of Geography, 110,771-783 (2001). (Japanese)
[29] S. T. Tse and J. R. Rice, "Crustal earthquake instability in relationship to the depth variation of frictional slip properties," J. Geophys. Res., 91, 9452-9472 (1986).
[30] Development of simulation model for forecasting crustal movement of Japanese archipelago :
http://www.mext.go.jp/b_menu/shingi/gijyutu/gijyutu6/sonota/05053002/h16houkoku/1502.pdf (Japanese)
[31] N. Uchida et al., "Interplate Quasi Static Slip off Miyagi and Fukushima Prefectures Estimated from Small Repeating Earthquake Data," Journal of the Seismological Society of Japan, 2, 59, 287-295 (2007). (Japanese)
[32] T. Hori et al., "A numerical simulation of earthquake cycles along the Nankai trough, southwest Japan: Lateral variation in frictional property due to the slab geometry controls the nucleation position," Earth Planet. Sci. Lett., 228, 215-226 (2004).
[33] R.E.,Habermann,"Precursory Seismic Quiescence : Past, Present, and Future," Pure and Appl. Geophys., 126, 279-332 (1988).
[34] S. Matsumura, "What can be read from changes in seismic activity? – Designation of asperities in interplate earthquakes – ," Special Issue of the Journal of the Seismological Society of Japan, to be published. (Japanese)
[35] D. Schorlemmer et al., Earthquake Likelihood Model Testing, Seism. Res. Let., 78, 17-29 (2007)
[36] K. Mogi,"Temporal Variation of Crustal Deformation during the Days preceding a Thrust-type Great Earthquake-The 1944 Tonankai Earthquake of Magnitude 8.1, Japan," Pure and Appl. Geophys., 122, 765-780 (1984).
[37] T. Sagiya, "Reconsideration of the Precursory Anomalous Tilt Change before the 1944 Tonankai Earthquake Revisited," 2004 Fall Meeting of the Seismological Society of Japan, A45 (2004). (Japanese)
[38] K. Heki, S. Miyazaki, "The Central Japan Plate Boundary and Tokai Earthquake," Chikyu Monthly, extra edition, 41,146-150 (2003). (Japanese)
[39] S.Ozawa et al., "Detection and monitoring of ongoing seismic slip in the Tokai region, central Japan, "Science, 298, 1009-1012 (2002).
[40] K.Mochizuki et al.,"Weak Interplate Coupling by Seamounts and Repeating M-7 Earthquakes," Science, 321, 1194-1197 (2008).
[41] S.Imoto,"Short-term Probability of the May 2008 Ibaraki-ken-oki Earthquake," Report of the Coordinating Committee for Earthquake Prediction, 81, scheduled contribution. (Japanese)

[42] S.Matsumura,"Common Changes in Patterns of Prior Seismic Activity in 1982 and 2008 Ibaraki-oki Earthquakes," Report of the Coordinating Committee for Earthquake Prediction, 81, scheduled contribution. (Japanese)
[43] Asahi Shimbun, August 27, 2008. (Japanese)

Profile

Shozo Matsumura
Affiliated Fellow,NISTEP

Senior Expert Researcher, National Institute for Earth Science and Disaster Prevention
Doctor of Science (Dr.Sc.). Dr. Matsumura's specialty is seismology. His aim is to detect precursory phenomena preceding major earthquakes by observing microearthquakes. He is particularly interested in understanding the condition of stress redistribution generated by slow slip and quasi-static slip from changes in patterns of seismic activity, focusing on Tokai earthquakes. He is an Expert Member of the Headquarters for Earthquake Research Promotion.

(Original Japanese version: published in October 2008)

6

The Role of Operations Research towards Advanced Logistics

EIZO TAKAI
Affiliated Fellow

1 Introduction

Product supply chains of recent Japanese industries are now at a major turning point in two major aspects, namely, structural changes in the nature of supply chains and more advanced functional requirements. Structural changes have been brought about by various circumstances, including the more complex and geographically wider international logistics associated with the globalization of corporate activities, a change in the relationships between companies from the simple serial procurement/supply chains of the past to open network-type supply chains, and changes in modes of logistics due to expansion of e-business accompanying the development of information and communications technology (ICT), among others. On the other hand, more advanced functional requirement levels are the result of shortening of product model change cycles and the need for adaptability to rapidly-changing markets, rationalization and energy conservation in response to high energy costs, and the like. The importance of the more advanced supply chains resulting from the progress of globalization is emphasized in this year's "White Paper on Monodzukuri (Promotion Policy for Basic Technologies in Monodzukuri in FY2007)."[1]

These requirements mean, in other words, that innovation and more advanced techniques are demanded in logistics in supply chains.

In realizing innovation and more advanced techniques in logistics, use of more advanced and sophisticated information technology is necessary. In this meaning promotion of the practical application of Operations Research (OR) is strongly required, as well as the development of more sophisticated OR theories and methodologies. Although improvement of hardware, such as transportation facilities and the infrastructure, is also important, problems will not be solved without more advanced planning and operation management systems using advanced software technologies and optimization methodologies. In rational planning of the highly complex distribution network as a whole, optimization of factory and warehouse location, and optimization of transportation routes by using OR is indispensable. In order to respond to recent business environment with rapidly-changing markets, and continual entry of new products an efficient operating system with flexibility and speed achieved by OR is demanded. The evolution of logistics is not limited simply to progress in hardware in the form of automation of warehouses and transportation facilities; progress and adaptation of OR techniques are also extremely important.

Furthermore, as recent challenges, high expectations are placed on innovation by service science as a means of increasing productivity and improving quality in the service industry, which now accounts for a major part of the industrial structure. As an important object of research in this field, logistics is considered to be an extremely promising area.

This report focuses on research and development of logistics design for realizing more advanced supply chains, the importance of which has been pointed out in the above-mentioned "White Paper on Monodzukuri." The concepts and current problems of supply chains and logistics are introduced, and secondly the proper roles of OR in solving the current and future logistics problems are discussed. Finally, the primary issues for the promotion of research on logistics and OR in Japan in the future are proposed.

2 Challenges and problems for logistics in Japan

2-1 Development of logistics in supply chains and related issues

The concept of "physical distribution," meaning the activities of distribution and inventory of the raw materials and products which support the real economy, has a long history. The movement of improving and rationalizing such activities through economic and technical research has been accelerated particularly in the United States since the end of World War II. The concept of "physical distribution management" appeared following World War II, and evolved into the concept of "logistics" as a new business model representing a fusion of physical distribution and information systems. In Japan, these two terms are used virtually synonymously, but properly speaking the former is a concept that focuses on the individual work of movement (transportation) and storage of goods, while the focus of the logistics is the rational design, management, and operation of distribution as a whole utilizing information. "Logistics" is originally a military term meaning the activities of supplying required troops, weapons and ammunition, provisions, etc. in accordance with tactical plans. Thus, the analogous activities in the field of economics were sometimes termed "business logistics." Quoting the definition of the Council of Logistics Management (CLM; now the Council of Supply Chain Management Professionals, CSCMP),[2] "Logistics management is the part of supply chain management that plans, implements, and controls the efficient, effective forward and reverses flow and storage of goods, services and related information between the point of origin and the point of consumption in order to meet customers' requirements." [2]

The activities in logistics, include inventory management, transportation management, warehouse management, material handling, packaging, returns/recovery in general, In some cases, such activities as demand forecasting, order management are also included. In addition to these, Lambert [3] mentions some other activities as plants and/or warehouse location design, customer service, procurement, parts/service support, and others. While top management function of logistics, CLO (Chief Logistics Officer), is just started to be recognized in Japan, this management function is already established in the United States. From this, one may say that the function of logistics is considered to be much larger and important in the United States than in Japan.

The concept of "supply chain" appears in the above-mentioned definition of logistics, and indicates the series of activities for supplying goods and services from procurement of raw materials to the final customer, through the chain of activities of production, sales, and distribution. The basic concept of traditional transportation business development is "correct delivery of products" for a single company or product. Differs from the concept of former simple transportation of single goods, the concept of "supply chain" is to create and provide final value to consumers considering wider range of integrated corporate activities starting from raw materials to final products, R&D and CRM. The concept of a "supply chain" was further expanded by the concept of Supply Chain Management (SCM) starting from 1990s that intend to realize new, rational business process through the integration of business activities clearing the boundaries of different business divisions and/or companies.

Figure 1 shows the general concept of a supply chain.[4] The totality in which the activities of central companies and the various related companies in the process from raw materials to the final consumption are combined by flows of business, information, and products/services (raw materials, parts, products, goods) forms entire supply chain. The solid line arrows in the figure show the flows of products from upstream to downstream, while the broken line arrows show the major flows of information such as order issuance, demand, etc., which run from downstream to upstream. It also shows that the supply chain is constrained by the capabilities of the member organizations in various factors such as the production or processing capacity, specialties, information, capital, human resources, etc. The aim of SCM is to realize more advanced supply chain by increasing the productivity, efficiency, and effectiveness of this system as a whole through establishing appropriate memberships, well balanced capabilities, and information connection.

SCM is a management methodology developed in the 1990s. Its aim is to create an efficient operation system by eliminating overproduction and shortages

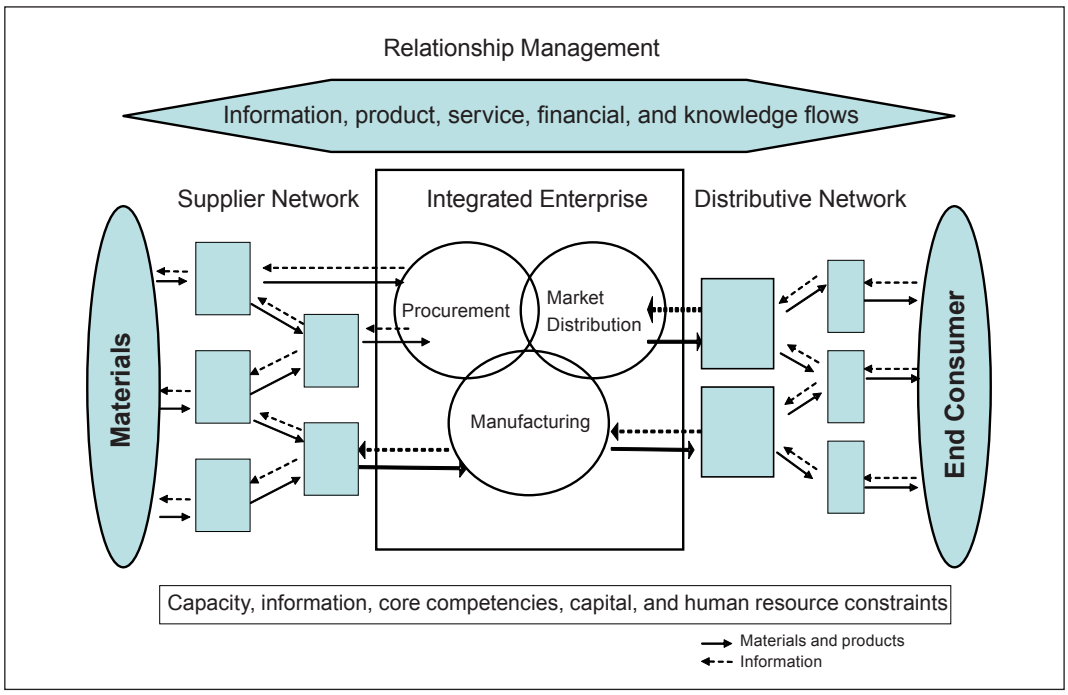


Figure1 : Generalized supply chain model

Prepared by the STFC based on reference [4]

to cope with rapidly changing volatile market situation , where conventional production planning based on traditional demand forecasting is not appropriate This is to be achieved by rational combination of production and logistics management at the same level. Its basis is reconsidering the whole business process across the barriers between organizations and integrate it as a single stream. In a sense, SCM has accelerated such changes as the shortening of product life cycles and the evolution of the internationalization of product manufacturing. In other words, in the era of physical distribution, supply chains were consisted of simple transportation connecting the inventories distributed all of the stages from source suppliers, production plants, warehouses, and retail stores. SCM is a methodology of transforming supply chains by connecting final consumer to the suppliers directly eliminating intermediate inventories as far as possible and provide efficiency with shortening of lead time.

The success achieved by the integrated model of the business process called QR (Quick Response), which enabled the rebirth of the apparel industry in the United States, is considered to be the beginning of the concept of SCM.[5] That success made it popular and many models of various names using various techniques were then conceived by the pioneers as Dell Corporation, Wal-Mart, and others. Although a variety of methodologies were proposed, their aims can be summarized as shortening of supply lead time, avoidance of defects, reduction of wasteful overproduction and excess inventories by sharing downstream information (demand side) with the upstream process, and controlling the upstream process on this basis. With the above-mentioned QR, the American domestic apparel industry succeeded in shortening the total process, which originally had required 56 weeks from thread for weaving to the finished product, to 12 weeks by realizing flexible production scheduling coupled to market information, shortened production lead time and shortened inventory turn around time of materials and semi products, etc. This represented the establishment of a business model which made it possible to supply products quickly in line with market trends, while greatly reducing total stocks, and introducing new products within a short time. This made it possible to compete successfully with imported cloths, which only had the advantage of low production cost. The implementation of this model was the occasion for the birth of the new type of apparel companies with worldwide retail networks. The concept that "downstream controls upstream" is also the basis for the world-renowned model called the "Toyota kanban system." and JIT (Just-In-Time) procurement.

Today, SCM has developed into a concept covering wide range of operation design and management including product design to customer relationship management. However, its basis is still on the

establishment of closer coordination of the information flow and operational workflow in production and logistics. Accordingly, the heart of the system lies in logistics.

As mentioned previously, the focus of this report is research and development related to designing logistics. Figure 2 shows the major issues in logistics existing among the major stakeholders in a supply chain. In this figure, although the problems in actual logistics are more congested, one may recognize that the importance of establishing plans in all operational flows and management levels, from strategic decision-making to routine operational control. As classified very broadly in Figure 2, the main issues in the planning can be divided into (1) issues related to international logistics, (2) issues in the location problems and network design for production plants and warehouses, (3) issues in the inventory management and inventory allocation in the network, (4) issues on the decision of transportation route, modal selection, vehicle assignment., (5) other related issues such as demand forecasting, cargo loading, labor management, etc. In addition, recently, as a combination of the above mentioned problems, issues of green logistics, including greenhouse gas reduction, conservation and recycling of materials, have also become important

In establishing the various types of plans shown in Figure 2, it is necessary to pursue optimality by using diverse data in many cases under various external conditions. In order to obtain rational optimum plans for these diverse and complex problems, application of mathematical techniques using OR is demanded. Moreover, research and development of more advanced techniques corresponding to the complexity and larger scale of problems is also necessary. These problems are discussed in Chapter 3.

2-2 *Recent problems in logistics in Japan*

As discussed at the outset, supply chain of Japanese industry is now facing major turning points, and rationalization and advancement of logistics is a crucial issue for strengthening the foundations of monodzukuri (art of manufacturing) in Japan to cope with the various problems such as the ever widening international logistics, requirements for higher market adaptability, economic pressure for rationalization, and satisfying energy conservation and environment-friendliness needs. All of these requirements are the keys to the procurement of raw materials and parts

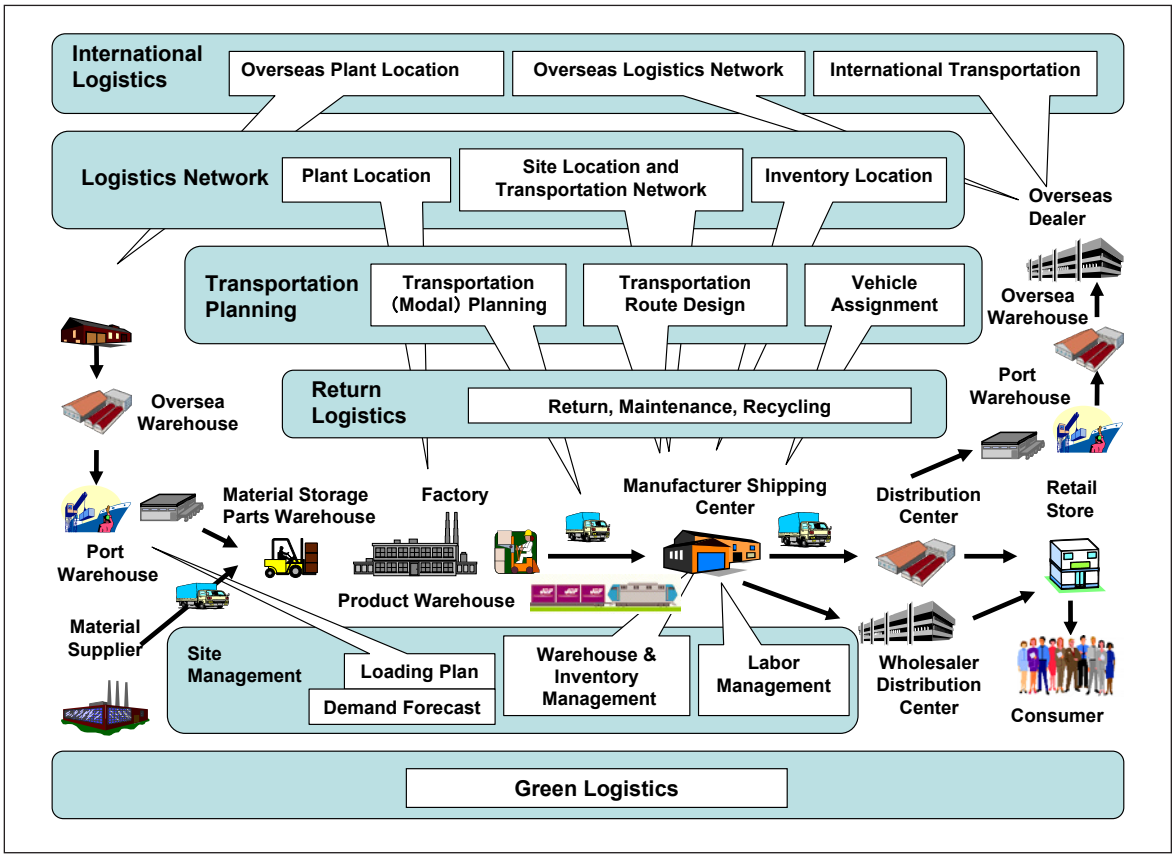


Figure2 : Major issues in logistics planning and management

Prepared by the STFC

and product supply to the market. Internationally, the development of a large-scale land transportation network extending from northeastern Asia to India, the Middle East, and Europe ("Eurasian Land Bridge") is considered likely in the future. Therefore, research on policies for strengthening Japan's international competitiveness, such as cooperation and contribution toward the realization of international transportation systems including these regions, is also an important issue.

The total amount of logistics costs in Japan's domestic industries is approximately ¥42 trillion/year (2005), amounting to about 8% of this country's gross domestic product (GDP).[6] Technical progress can be seen in this field, as exemplified by the use of advanced material handling tools, transportation facilities and RFID, and improvements have been realized in the software aspect, such as popularization of inventory optimization software and transportation management systems. However, the labor productivity in this sector is still low. According to the "International Comparison of Labor Productivity (2006)" published by the Japan Productivity Center for Socio-Economic Development.[7] Japan's labor productivity in the logistics sector is approximately 50% of that of the United States. Thus, further rationalization is strongly demanded in this sector.

It has been pointed out that the possible causes of this low productivity may include various factors, the as-delivered pricing system, which is a unique business custom in the logistics and sales industries in Japan (system in which the cost of transportation to the customer is included in the price of the product; because this makes the cost of distribution a latent factor, it is considered to be a cause of low consciousness of rationalization in distribution), delivery with small lot and high frequency due to the small storage space of stores, strong preference of consumers on freshness, requirements for large variety of goods, and other cultural factors. However, there is also an opinion that one big factor is strict quality requirements on delivery in Japan (on-time delivery rate, order fulfillment rate, less damage rate, etc.), which are without parallel in other countries. For example, according to the recent research, the on-time delivery rate in the United States is approximately 90%, but in contrast, the rate in Japan is 99.99%.[8] This undoubtedly supports highly efficient production in Japanese manufacturing factories, but on the other hand, the possibility that this increases delivery costs by requiring the same accuracy in areas where strict on-time delivery is not necessarily essential has also been pointed out.

In the distribution costs of ¥42 trillion mentioned previously, nearly 70% is costs associated with transportation. Due to heightened interest in so-called green logistics, the main purpose of which is reduction of CO_2 emission demanded by global warming countermeasures in recent years, and the necessity of energy and fuel saving on transportation due to the spike in the oil price, interest in rational transportation technology and delivery systems is extremely high. Research on the solutions involving a combination of transportation methods, such as land transport by truck and rail, sea transport, air transport, etc., which is termed multi-modal transportation, is a promising area on this problem. In order to realize this rationally, R&D activities should not be limited on hardware, but also should include optimization planning using advanced mathematical techniques by OR to realize complex and large scale logistics.

Needless to say, the effects of changes in the supply chain will extend to industry as a whole. Rationalization in logistics is not a problem only for the logistics industry (transportation and shipping industry, forwarders, warehousing industry, 3PL (third-party logistics)) in the service industry sector; its necessity in manufacturing industries and the energy industry is also increasing.

In the logistics planning's shown in the Figure 2, there are areas where advanced techniques have already been accepted. However, considering the effect of changes in the supply chain as a whole, solutions to problems which are difficult for the private sector alone, such as optimum design of large-scale international transportation networks, including their operation, development of advanced technologies for realizing green logistics, and the like, are expected in specialized research institutions such as universities. Chapter 3 describes problem-solving techniques using OR and the necessary research and development to solve the important focused problems for which effects are expected.

3 Advancement of logistics by OR and expected research fields

In a complex activity like logistics, problem solving by assumptions of a simple cause-and-effect relationship alone is not sufficient. Optimum plan should be derived through the analysis and comparison of many alternative case studies simulating the behavior of logistics organization or processing system under the premises based on present situations or future planning. This means that application of the OR methodology based on mathematical process is strongly required.

3-1 Development of OR and logistics

OR (operations research) was born as operational research, which is a technique corresponding to technical research, in order to make effective use of new military technologies such as radar, night fighters, underwater mines, etc. in the United Kingdom in the final stage of World War II. Subsequently, the concept migrated to the United States, where it developed under the name of operations research (in this case, tactical operation research), and provide a wide range of achievement, from frontline military operations as such to logistical support in behind. After the war, this technology was made available to private business sectors and was widely applied in various industries, evolving into the OR of today.

Many definitions have been given for OR. In general, however, as shown in Figure 3, the term indicates a methodology in which corporate and social activities producing goods and services are understood as a system which outputs results when inputs are given, the system of those basic activities is constructed as a mathematical model, and problems regarding the operational methods of the system are analyzed using mathematical tools conforming to the purpose in order to produce the optimum solution. In practice, the problem finding and derivation of solutions are carried out by using a variety of techniques, such as mathematical and statistical analyses, system simulations, optimization algorithms, etc.

From the very start of its development, OR had a strong relationship with problems related to logistics. Many of the basic technologies in the initial stage have also become basic technologies of logistics planning, even today. That is, the basic technologies of demand forecasting, inventory theory, determination of optimum order quantity (economic order quantity), shortest path planning, mathematical programming methods (resource allocation problems, optimal location problems), queuing theory, discrete

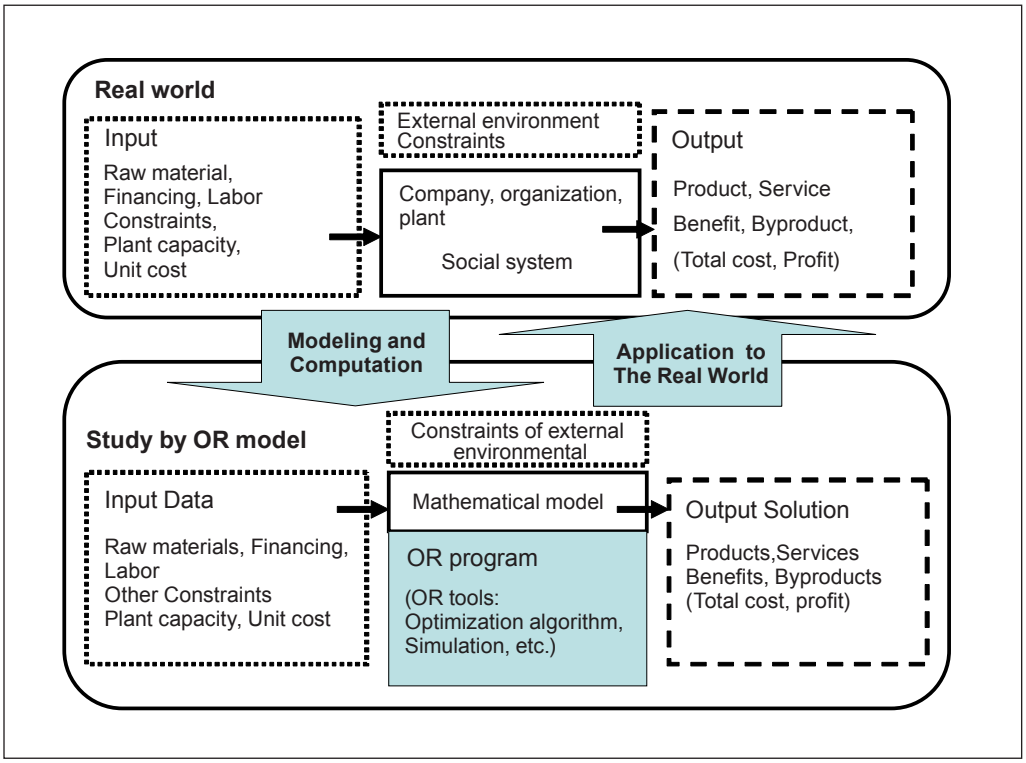


Figure3 : Concept of study by OR model

Prepared by the STFC

simulation, etc. were researched as problem-solving techniques.

OR is an academic discipline which has already produced successful results in the real world, but like other fields of research, it is also one where evolution is constantly demanded. Since its birth, OR has continued to develop and evolve by developing new solution methods as a science for problem-solving in management and planning, which is always unified with practical business in dealing with new problems. As a result, the cycle of opening new situations for application by the development of new theories has continued to bring about further progress.

In an example related to logistics, the remarkable drop of airline fares in recent years is a result of adopting rational fleet operation scheduling using state of art planning techniques. It had long been understood that large-scale mathematical models for this type of problems could be solved by integer programming method, there was no method of solving large-scale integer programming problems with a model size exceeding one million variables at a practical speed and the application of the technique was limited to small partial planning models. The breakthrough was made by the algorithm develops by Narendra Karmarkar of AT&T's Bell Laboratories (at the time). (AT&T was granted a patent for the Karmarkar algorithm. As this was the world's first patent for a mathematical equation, it caused a controversy argument on the patentability of equations over the world).[9] Today, many practical methods to solve this type of problems have been proposed and commercially available in the form of optimization software, and are widely used in the planning for semiconductor production, optimization of the logistics network, and similar problems.

Penetration of OR in the logistics field in Japan is still in behind with Europe and the United States. One reason for this is the fact that virtually not much practical research has been conducted in universities. In Europe and the United States, research on and the establishment of logistics strategies using OR is carried out widely as joint research by universities and private companies and/or public organizations, and the results of the research are often made public in journals and presentations at academic conferences, etc., including the results of research and implementation by private companies. However, in the Japan, there is little awareness of the importance of this. Not only collaboration between companies and universities are rare, but companies also tend to avoid publication of data and results to a greater extent than is actually necessary. As a result, there have been delays in a variety of aspects, including practice, R&D, and the training of skillful people with advanced knowledge and application capabilities in companies. Although Japan has many outstanding researchers, it is unfortunate that their work has been limited to theoretical study in the laboratory. The development of policies that overcome this condition is demanded. The author hopes to see the establishment of policies which activate research and development in connection with practical problems and research issues desirable to promote from a policy standpoint, as will be discussed in the following and sharing of the results by industry and universities. Logistics in Japan has achieved a unique development suited to the culture of consumers and corporate management in this country. This is also an area where more advanced techniques and contributions can be expected from research and development using Japanese approaches. In addition to practical benefits, promotion of public research and publication of the results can be expected to improve the current situation in which research results are not made public, and thus will also be useful in promoting theoretical research.

3-2 Challenges toward advanced logistics and the role of OR

(1) Response to wider area, more complex supply chain networks

(a) Optimization of facility location and transportation route in large-scale logistics networks

One issue with particular attention in recent logistics is the problem of optimization of the network structure as a whole, as these networks are spreading and also more complex.[10,11]

A transportation network in logistics comprises connecting nodes such as resource suppliers, production plants, warehouses, and the final consumer, and the transportation links joining those nodes. The mathematical model of entire network could be an extremely large-scale network model, often having equations and variables more than 10,000. A number of reports have presented examples in which it was possible to realize nearly 30% reduction of operation

and transportation cost by optimal rationalization of the site location arrangement and selection, and transportation network. In particular, reduction of transportation costs, which accounts for the main part of cost reduction, makes a direct contribution to reducing transportation distance, and therefore also contributes to reducing emission of greenhouse gases.

Because problems of optimum routes and site locations in large-scale, multi-stage networks treat variables which must be expressed by numerical values, i.e., the number of site, number of cars or transportation facilities, and combinations of delivery routes and sites, these are fundamentally optimization computation by using extremely large-scale integer programming models (several 10,000 to 1 million variables or more). Until recently, this type of problem could only be handled by research institutions with very limited computational capabilities. In recent years, with progress in computational algorithms and higher speed and larger capacity of computer hardware, it has now become possible to solve these problems at a practical speed using a general personal computer. As a result, it has become relatively easy to perform optimal design and planning for large-scale, wide area supply chains.

Figure 4 shows an example of the scale of that type of computational model. An optimization model for route selection which includes combinations of all the alternatives, such as the number of nodes, product models, transportation method, and the like, will be enormous. In Japan, application of mathematical models for rationalization of this type of supply chain network had attracted a little attention in the past due to the lack of knowledge of and interest in mathematical optimization. However, due to the increased transportation costs caused by the recent jump in fuel prices, ongoing closure/consolidation of bases in response to corporate mergers and other factors, etc., application is continuing to advance in a wide range of areas in Japan. Figure 5 shows an example of optimization planning for distribution network in North America for a consumer electronics maker. [12] In Europe and the United States, where transportation network is spread in broad geographical fields, many variations of research including multi-modal transportation are being carried out in both the theoretical and practical aspects by private corporations, universities and other research institutions.

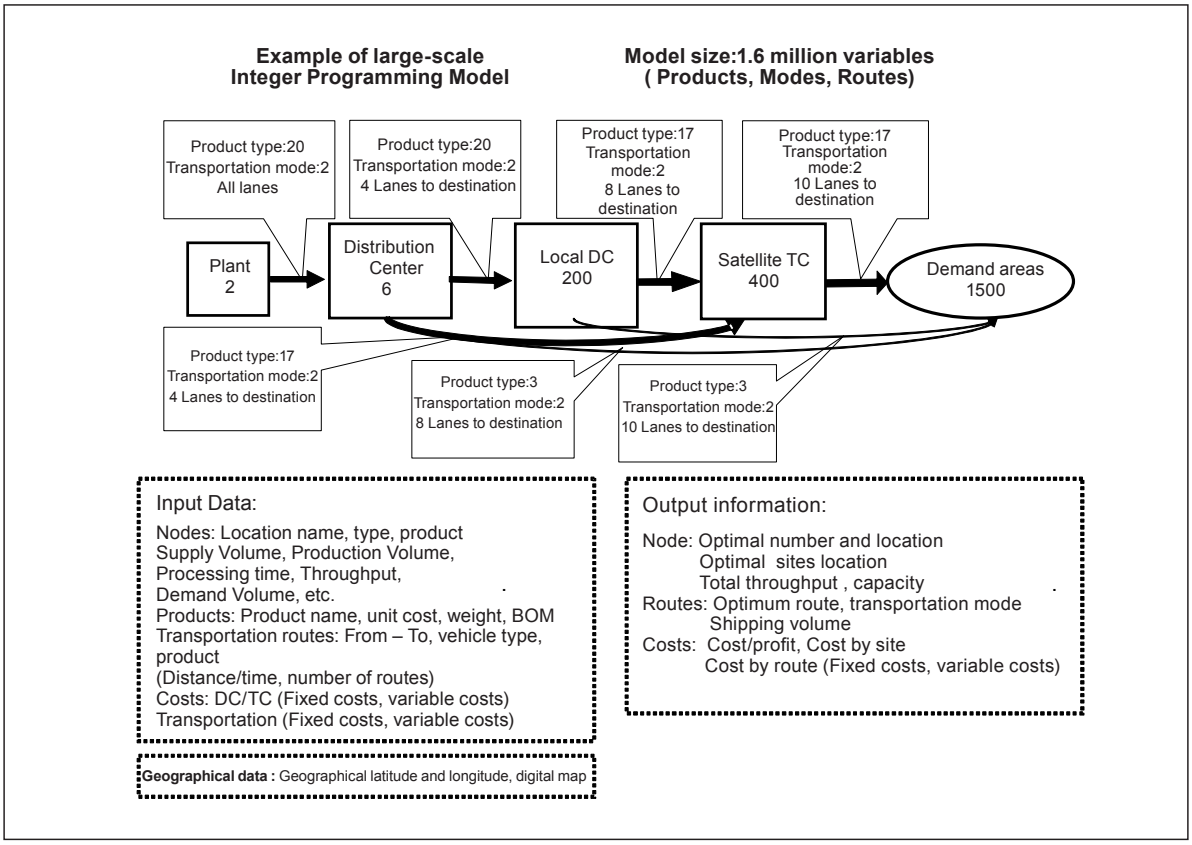
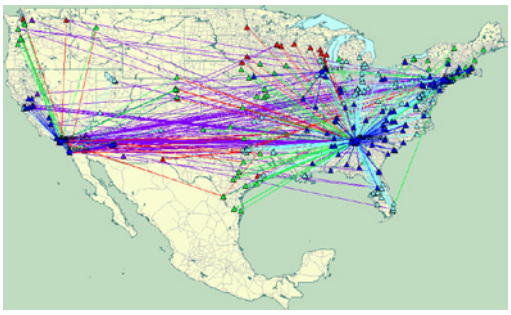
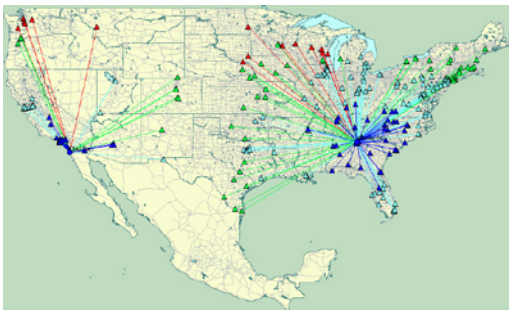


Figure4 : Example of large-scale logistics network model

Prepared by the STFC based on materials of Frameworx Inc.



Original delivery routes before optimization



Optimization result with 2 distribution centers
(Total distance and delivery time were shortened by more than 30%)

Figure5 : Optimization results of North American distribution network of a consumer electronics maker.
Prepared by the STFC based on reference[12]

In particular, in the European nations, for example, the Netherlands and Belgium, government policy promotion measures has been studied by using mathematical models and proposed,[13-15] and logistics centers have been developed targeting to service American and Japanese companies, with the country's port or airport as the point of origin, and a multi-modal integrated transportation network combining sea transportation (short sea shipping by Ro-Ro ship), truck, railway, and river transportation for the entire EU region.

In Japan, the opening of the FY2008 "White Paper on Monodzukuri" mentioned the importance of supply chains and the necessity of promotion of strategic logistics networks. In this aspect, cooperation in logistics with the Asian countries and Russia will be particularly important for Japan in the future. For example, that report states that several companies are already started using the Trans-Siberian Railway to cut transportation time from Japan to Europe by half, from more than 40 days to around 20 days, and the government has also announced that it will support modernization under the "Eurasian Industrial Investment Bridge" concept.[16] While hardware such as measures for vibration of railways and transportation facilities will be necessary in the construction and operation of an Asia-wide transportation network, what will be particularly important in the future will be rational operation of the entire network of transportation routes, combining land and sea routes. Therefore, it is desirable to carry out a strategic research, including cooperation with the other East Asian nations, and to implement construction and improvement of port and harbor facilities and inland terminal warehouses in a policy-oriented manner, based on horizontal cooperation among the ministries and agencies concerned. In the future modernization of logistics in Asia, Japan must take the lead in optimization research on the international logistics network cooperating other neighboring nations.[17]

(b) Inventory reduction by optimization of safety stocks

In wide area networks, placement of large quantities of inventories close to the final demand is not the optimal solution, economically speaking, for responding to fluctuations in final demand and shortening lead time in supplying customers. Where in the network to hold the optimum safety stock, and what amount to hold, considering supply and transportation lead time, are major problems from the viewpoints of effective utilization of resources and economic optimality. Obtaining practical solutions had been considered a difficult problem, but this is now approaching a solution by recently-developed techniques using dynamic programming and complex system simulation technology, and others.[10,18] There are examples in which the cost reduction achieved by this type of optimization reaches 30%. Likewise, in order to respond to globalization, which was mentioned previously, a large economic effect by further development of optimization technology is also expected in the future. In particular, in manufacturing industries which are engaged in international market development, the profit from optimization of the inventory location and quantity of service parts, based on the relationship between the service level and

inventory costs, is large. A leading northern European furniture maker[19] and a European tire maker, PC printer makers, automobile makers, and others who are expanding internationally are actively engaged in research. Theoretically, optimization of site locations from the viewpoint of transportation costs and inventory handling in network should essentially be performed simultaneously; however, no solution method which enables simultaneous optimization of these two problems has been discovered as of the present. This is a field where further research is expected.

(c) Modal shifting and multi-modal transportation

Accompanying the recent globalization of production and market, optimization programming of multi-modal, internationally integrated transportation combining land (truck, container, rail), sea (container ships and trailer ships), and air has become necessary. Because this is accompanied by greater complexity in transportation routes and the difficulty of time planning, here, new mathematical method is expected for rational selection of transportation routes and operational planning. Furthermore, as mentioned previously, in the establishment of plans for large geographical areas, and particularly the Asian region and transportation between Asia and Europe, research on multi-modal transportation will be indispensable.

Similar problems have arisen in CO_2 reduction plans by modal shift in Japan. Here, planning which combines trunk-line transportation and local transportation ("last one mile") has become necessary.[20-23] Recently, examples have been published in which a CO_2 reduction of 20% was realized by large-scale milk run-type collection/delivery to an electronic part assembly plant, and trunk-line transportation to complete-product demand areas and combined delivery to final customers implementing a large-scale vehicle allocation and control system.[24-26]

It has become difficult for research institutions such as universities to do research in the field represented by the above case due to the difficulty of obtaining practical data in Japan. In Europe and the United States, it is possible to obtain transportation cost tariffs openly. Moreover, as mentioned previously, several alternative means of transportation exist for the geographically extensive networks (truck, trailer, rail, river transportation, air routes). Therefore, research on this area is being carried from a variety of directions. In Japan, joint research with private companies is of course being carried out for effective utilization of local ports and harbors and activation of regional industrial clusters, but more active research on optimization of transportation extending beyond administrative divisions is also desired.

(2) Issues in which development of new methodologies for model construction and optimization are expected

The following are the topics important in the logistics of any company. However, further research on solution methods, conducted through industry-university collaboration, is expected. If the development of new techniques can be realized by cooperation between practical businesspeople and Japan's outstanding researchers in the mathematical sciences, this will have the potential to contribute to logistics and OR at the world level, although there are still problems which are left to individual companies in practical responses.

(a) Dynamic delivery planning optimization for agile response to demand

In the future, high expectations will be placed on the development of new mathematical solution techniques for realizing advanced production and transportation systems which are capable of responding quickly to changing demand. Technologies for real-time rescheduling and vehicle assignment and route planning by using real-time data collection through GPS and RFID are expected by breakthroughs such as practical application of the aforementioned large-scale optimization algorithms, etc.

Rationalization of truck dispatching using tools involving mathematical algorithms is continuing to gain acceptance in Japan. However, because energy saving and cost reduction by alleviation of urban traffic and improvement of the current 50% truck load rate will contribute directly to reduction of greenhouse gases, more sophisticated vehicle allocation planning systems and their dissemination are desired. Wide implementation of detailed operation plans by these technologies, including higher efficiency by joint loading and joint delivery of cargos, reduction of operating costs by proper time windows for deliveries, combined collection and delivery, etc. will also

contribute to promoting green logistics, which will be discussed later in this paper.

(b) Procurement planning for risk management and economic optimality

Accompanying globalization and open sourcing in the procurement of raw materials and parts, solutions using combinatorial theory can be expected in responding to risks such as natural disasters and terrorism, procurement planning with stable supply and economy, and optimum selection of sources involving complex factors.

In order to secure alternate means of transportation in response to traffic disruptions due to natural disasters such as earthquakes, etc., the necessary public financing must be decided rationally. Simulation of social costs by mathematical study is considered useful in this kind of decision making.

(c) Concurrent design of products and supply chain

Recently, one area where research is also continuing to advance in other countries is a methodology of performing design of the supply chain simultaneously and in parallel with product design, in which the product design itself is changed considering optimum procurement and product delivery. In one example from Japan, a copying machine maker succeeded in reducing costs by broadly shortening delivery time and reducing stocks, by restudying where the location of final product assembly should be located in the supply chain and performing design for modularization of parts to enable production of final products easily at intermediate locations in response to market requirement, however this practice was not led from theoretical optimization. In generalizing this kind of design technique, including part procurement routes, research and development of new mathematical techniques is required, including simulation techniques for complex systems, etc.

Supply chain optimization, including the product design, will have a positive influence in the implementation of carbon footprint, which will be applied in the near future. The carbon footprint of products will be shown on the product package, etc. Under this concept, the greenhouse gases emitted over the entire life cycle of the product, from raw materials to production, sale, distribution and consumption, are expressed in terms of the amount of CO_2 emissions. This is being promoted actively in the United Kingdom, and study in Japan has also begun. By encouraging consumers to select products with low environmental impacts, manufacturers can be encouraged to produce products with lower impacts.[27] In addition to generation of greenhouse gases by energy sources in the production plant, environmental loads can also increase as a result of transportation, depending on the site location of the plant.

(d) Research on mathematical solution measures for new transportation methods and their effective utilization

In the future, it is thought that the development of new means of transportation will be promoted with the aim of reducing energy costs and environmental impacts. Innovations in the hardware aspect are continuing to be realized. These include international trailer transportation using Ro-Ro ships, responding to the needs of medium- and short-distance marine transportation. Ro-Ro is an abbreviation for Roll-on/Roll-off ship, and is a type of cargo vessel with a structure that enables loading and unloading of vehicles such as trailers under their own power. This method is more suitable than the container method for medium- and short-distance transportation. Other examples include practical application of multi-modal transportation using new rail containers and dedicated trailers, which have already been implemented by an automobile maker, transportation of 40 foot marine containers by rail and consolidation of cargos and arrangement of transshipment centers for its effective utilization, etc. New optimization systems for operation planning suited to these innovations are demanded. In addition to this, in order to realize circulatory-type logistics for resource saving, further research must be carried out on international container management, pallet management, route planning techniques for circulatory transportation, and similar topics.

(e) Design of business work flow structure in supply chains and research for optimization

Research on application of a standardized process operation reference model (SCOR model)[29] and business process modeling (IDEF, etc.) for optimization of the structure and work flow of supply chains combining multiple companies and different business entities is

increasingly necessary. In supply chain rationalization in Japan, which tends to begin and end with only a response to individual issues, there is little familiarity with top-down type model structures and design concepts that intend this kind of total optimality from the outset. Recently, these have at last begun to attract attention. Further research for future supply chain rationalization from a total viewpoint is demanded. On the other hand, methodologies such as SCOR are fundamentally premised on top-down organizational management. Accordingly, in order to demonstrate effectiveness in Japan's bottom-up type decision function, unique theories for modeling and implementation which are not simply copies of those in Europe and the United States must be researched.

(f) Research on supply chain structures in cooperation with social science approach

Supply chain networks comprise numerous companies in the same or different industries (and in cases, companies with different nationalities). The decision-making in each company must be conscious of the existence of other decision-making units (other companies). However, conventional theory and management solutions are materialized on the precondition of a single decision-making unit. For scientific treatment of strategic decision-making premised on this type of interaction, it is necessary to borrow the assistance of game theory and microeconomics, or agent model approach, etc.[29] In order to analyze the behavior of a supply chain network, which has a complex structure, including environmental problems, and measure and improve their efficiency, interdisciplinary cooperative work between research by scientific and mathematical approaches and research by social science approaches is indispensable. In other countries, research on supply chain and logistics problems by this kind of cooperative work are made widely, and a number of papers have been published in management science journals. It is thought that researching the ideal form of management from an interdisciplinary standpoint will not be only useful in corporate decision-making, but will also provide valuable knowledge for making industrial policy in Japan.

3-3 Role of OR in promotion of green logistics

With the start of the 1st commitment period under the Kyoto Protocol, which is from 2008 to 2012, Japan is targeting a 6% reduction in emissions of greenhouse gases during this period from the baseline years (fiscal 1990 for CO_2, CH_3, and N_2O and fiscal 1995 for the 3 gases classified as chlorofluorocarbon-replacing material, etc.). However, in actuality, Japan's emissions of greenhouse gases in fiscal 2006 were 1.34 billion tons (CO_2 conversion), which was an increase of 6.2% against the baseline year.

The main cause of this increase was a large increase of 12% in fiscal 2006 against the baseline year in CO_2 emissions originating from fossil fuels, which account for approximately 90% of Japan's emissions of greenhouse gases. Emissions from the transportation sector, which is responsible for approximately 20% of CO_2 emissions originating from fossil fuels, increased by about 20% from fiscal 1990.

In addition to these circumstances, transportation costs have also increased due to the recent sharp rise in the price of fuel oil. Against this background, energy conservation in the logistics sector has become extremely important for protecting the foundations of Japanese industry.

As touched on in the connection with the application of OR to various problems discussed in Chapter 3, rationalization of logistics can make many direct contributions to the realization of green logistics. However, distinctive problems or challenges also exist in green logistics.

In application of OR to green logistics, one topic which particularly requires research is a technique and model structure for simultaneous optimization of multiple objective functions, i.e., energy efficiency, greenhouse gas emissions, cost, and the like. In optimization of mathematical models, research on the establishment of solutions which simultaneously satisfy multiple objective functions and so-called multi-objective programming is necessary. Several proposals have been advanced as solution methods for multi-objective function models for green logistics. These proposals can be broadly classified into (1) methods in which multiple objective functions are solved individually, assigned certain weights, and summarized in a single function, which is then either minimized (total environmental impacts + cost, etc.) or maximized (effect of countermeasures), (2)

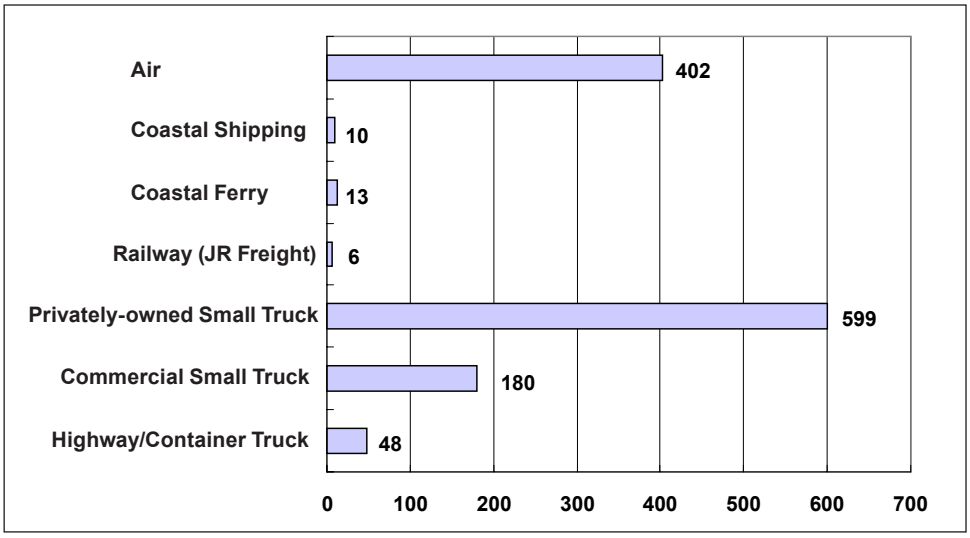


Figure6 : CO_2 emissions by means of transportation (g/ton-km)

Prepared by the STFC based on reference[30]

methods in which minimization or maximization of a certain objective function is solved using the target values of other objective functions as constraints, and (3) methods in which the point of convergence of the values of multiple objective functions is obtained using, for example, Pareto optimization, etc. However, none of these proposals has become a definitive method. Therefore, research on both methods of modeling actual problems and mathematical solution methods is expected.[20]

Although already touched on in Chapter 3, one other major challenge is research in connection with modal shift and multi-modal transportation. Emissions of CO_2 differ greatly depending on the means of transportation, as shown in Figure 6. Accordingly, research on means of transportation will demonstrate effectiveness in the prevention of global warming. Among concrete countermeasures, an example in which a CO_2 reduction of 20% was achieved by appropriate vehicle assignment and deliver planning was mentioned previously as one method that has already begun to be implemented. Modal shift is expected to have an even larger effect. Already, an example in which a CO_2 reduction of nearly 80% was realized by switching to rail transport for trunk-line transportation and future reduction plans by joint use of short sea shipping have been reported.[25] As also noted previously, it is considered that implementation of the concept of carbon footprint and its importance will become increasingly great.

In Japan, however, until now study has been limited to the range of individual companies, and almost no research has been done by OR on inter-company problems or toward national policy. In Europe, much research is being done on the EU as a whole, particularly in universities, and proposals are being made. In promoting rationalization of logistics as a national policy, a cooperative system which extends beyond the boundaries not only of companies, but also those of government ministries and agencies and local governments is indispensable.

Universities and public research institutes should play a large role in drafting composite and comprehensive plans, not limited to structuring individual infrastructure such as port and harbor or road construction and improvement, but also in the effective distribution of cargo collection and delivery bases and improvement of their functions, combination of trunk-line transportation, including railways, and deliveries to individual customers, which is called “last one mile,” etc., as well as in the development of models for these problems, research on optimization algorithms, and the like. Prioritized financial support policies for this type of research may also be effective.

As mentioned earlier, it is also necessary to study modal selection, including land, sea, and air, in research in connection with construction of a transportation network which includes the entire Asian region and Russia, and policies for reducing international environmental impacts accompanying the realization of such a network.

3-4 Expectations on service science research towards advanced logistics

In the final report of the United States’ Council of

Competitiveness, which is entitled "Innovate America: Thriving in a World of Challenge and Change" (April 12, 2004; generally abbreviated as the "Palmisano Report"),[31] "service science" was introduced as a new academic field with the potential to induce innovation. Service science is important as one direction for the development of science and technology responding to new industrial structures. Needless to say, it is considered possible to obtain substantial benefits by advances in service science in the logistics industry, which occupies a large position in the service industry.

The Palmisano Report also pointed out the following with regard to the relationship between service science and OR: "There is no field in which a new interdisciplinary approach is so clearly necessary as in the new field of 'service science'. Service science is a fusion of existing fields, including computer science, operations research (OR), production engineering, mathematics, management science, decision science, the social sciences, forensic sciences, and others. This causes reform in corporate activities as a whole, and encourages innovation in the region where the specialized knowledge of technology intersects with business."

As good examples of improvement of service productivity by OR, bank ATM windows and JR railway ticket counters can be mentioned. By changing the form of queuing from the previous queuing at each window to a forked form of queuing, waiting time as a whole was shortened and the service rate per unit of time was increased. This change also had a positive effect on customer psychology. This is an example of application of queuing theory, which is one typical OR technique. Although popularization in Japan was comparatively recent, it is generally said that a demonstration experiment video by (then) Prof. Morimura of Tokyo Institute of Technology on the 1991 NHK television program "Try and Try" had an very large influence on its rapid acceptance.[32] Queuing theory is used in the design of the number of service windows and their operation, and in addition to ticket counters, has also been applied to the design of toll booths on expressways, the number of passport inspection gates in airports, etc.

One of the purposes of service science is improvement of productivity, reduction of costs, and improvement of quality by application of engineering techniques to services. In other words, it is a science which researches a total design by quantifying services, modularizing and standardizing service solutions, and combining modules, and through these practices, applying engineering techniques to the service industry and thereby improving service productivity.

Logistics is considered to be an area where important effects can be expected by promoting interdisciplinary research by the service science approach. In particular, in the aspect of service quality, quantitative concepts such as KPI (Key Performance Indicator), metrics, benchmarking, and similar techniques applied to service quality have already achieved penetration. This field also has conditions which facilitate study of objects of research, in that data collection is easy, the relationship between the service provider and beneficiary (customer) can be defined relatively clearly, etc.

For example, one quality-related concept in logistics is "Perfect Order Fulfillment Rate." Although several different definitions of this concept exist, for instance, in the definition in the above-mentioned SCOR, this is defined as the numerical value obtained by multiplying the four percentages of the quantitative completion rate, on-time delivery rate, documentation completeness, and non-damage rate.[28] This has become one index of quality assurance for logistics contractors in Europe and the United States. However, in Japan, contracts which attach importance to numerical values other than cost are still rare. In the future, it is thought that this may be suitable material for research on the relationship between quality and cost or productivity.

On the other hand, in research on service science, it is considered necessary to avoid over-reliance on a reductionism methodology that reduces the object to its basic elements and understands it in numerical values that can be quantified physically. In performing analyses of service quality and customer satisfaction and preparing the optimum design, techniques in the field called decision science within the larger discipline of OR could be very effective. Examples include AHP[34] (Analytical Hierarchy Process; recently used in narrowing the field of candidates for relocation of the capital) for assessing the value of non-quantitative factors, conjoint analysis (technique used in analysis in marketing for product design, etc.), DEA[35] (Data Envelopment Analysis; in actual examples, has been used in analysis of public service institutions such as libraries and hospitals, etc.) for evaluating the

efficiency of organizations while simultaneously handling multiple inputs and outputs, and others.

In service science, areas of research where Japan can particularly contribute are research in connection with evaluations of qualitative/sensible quality, which tends to be slighted in the Western European approach, and research based on holistic ideas, which consider the whole and its parts simultaneously and are a distinctive quality of the Japanese people's way of understanding the world

Japan's unique, and it can be said, excessive insistence on quality can also be seen in the field of logistics. While this also has an aspect of increasing social costs, at the same time, its effect in bringing about a more pleasant life for the country's citizens must not be overlooked. If the Japanese approach can compensate for the weaknesses of other evaluation standards that insist on quantifiable numerical values by promoting interdisciplinary and boundary-region research, this will be important as part of this country's contribution to the development of service science. In the past, in quality control of industrial products, Japan demonstrated that it is possible to satisfy both quality improvement and cost reduction simultaneously, which had been considered mutually contradictory until that time. Similar examples can be seen in logistics. For instance, in deliveries of goods to convenience stores, vehicle allocation planning by OR and transfer-type delivery centers have been used, and delivery schedules by detailed temperature range have been combined in dedicated vehicles. When the first convenience stores were opened in 1974 in Tokyo, the number of deliveries was 70 vehicles per store per day and 42 two years later. As a result of these efforts, the number of deliveries was reduced to 15 in 1987 and to only 9 in 2005, while the freshness of boxed lunches and bread has actually improved. This is an example in which both service and efficiency issues were successfully solved.[36] From this viewpoint, it is thought that Japan can contribute to service science as a whole if it can analyze this country's strengths and weaknesses in logistics and, with this as a reference point, conduct research on quality in logistics using an interdisciplinary approach.

4 Conclusions and recommendations

Supply chains play a critical role as the foundation for industry, but are now facing major turning points. This report has described the effectiveness of support for rational planning and operation of supply chains by optimization methods and simulations using OR techniques, and in particular, the challenges which will require research and development efforts in the future.

In particular, Figure 7 summarizes five key themes among problems requiring solutions, as described in Chapter 3, and the corresponding R&D themes for OR techniques. As mentioned previously, in order to respond to the challenges of greater complexity, large geographic extent, and internationalization which today's logistics confronts, and to solve future problems when they arise, research must be promoted on issues having a strategic viewpoint, as shown on the left in Figure 7. It is also necessary to simultaneously promote research and development of more advanced OR techniques which are capable of modeling complex realities, as shown on the right, and providing practical solutions.

In addition to promoting research on these key challenges, when promoting developmental research on advanced logistics and OR techniques for this purpose in the future, the policy responses outlined below will also be necessary.

(1) Logistics as an engineering subject and development of human resources

In comparison with Europe and the United States, rationalization and optimization of logistics by OR techniques has not been well promoted in Japan in the past. One reason for this is that logistics has been understood as basically an engineering subject in Europe and the U.S. As a result, research has been actively promoted in science and technology-related educational and research institutions, and independent departments or faculties were established in higher education from an early date. In contrast, in Japan, there is a strong tendency to understand logistics as a business in commercial science and as an ancillary to sales and marketing. For this reason, its position has not been well established; either in companies or in the science and engineering community.[37] In recent years, science and engineering universities have become central to the promotion of advanced logistics in Europe, the U.S., and some Asian countries. Likewise, it is desirable to create a logistics research promotion system which prioritizes science and engineering in Japan. In Europe and the U.S., training

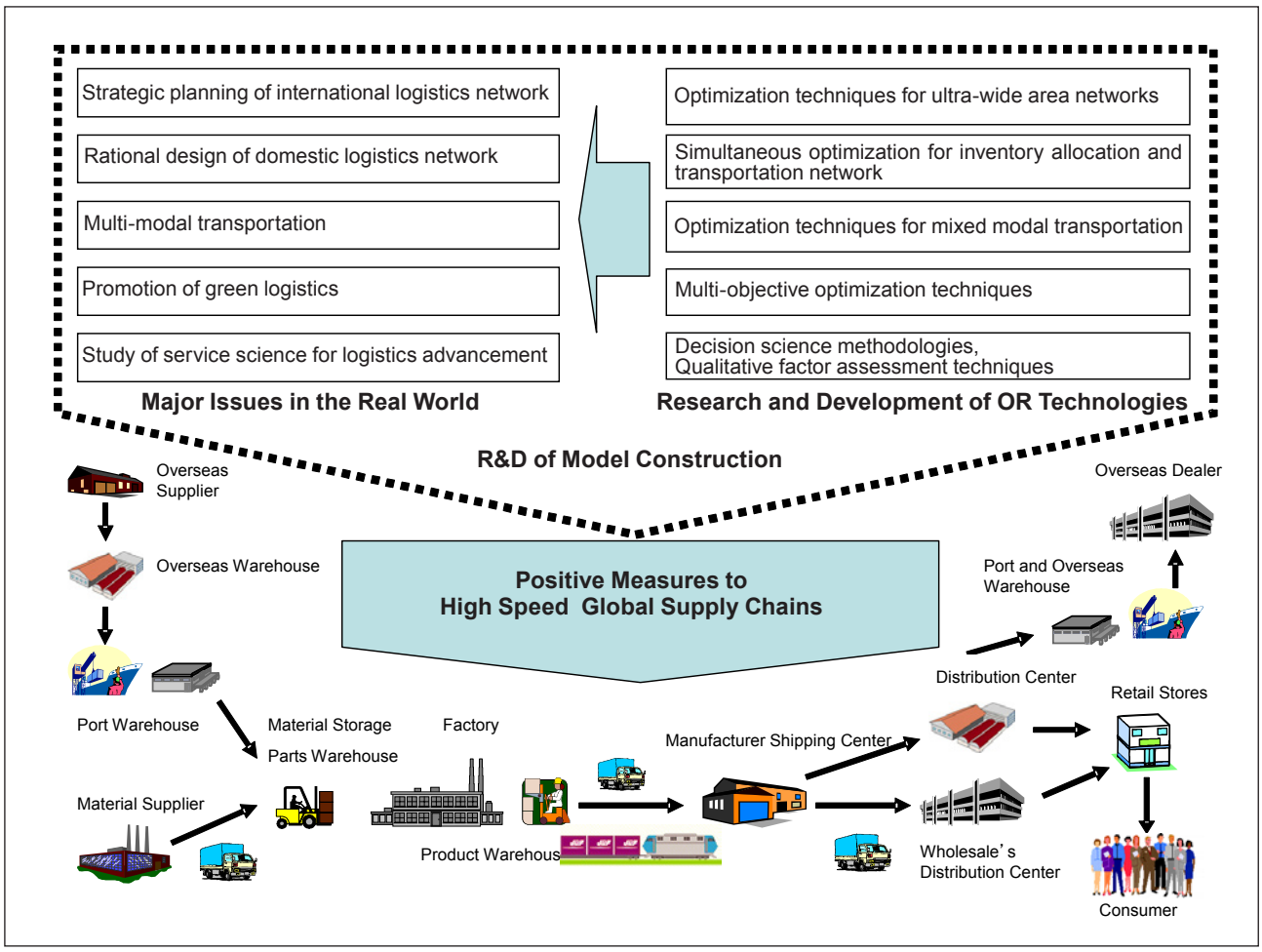


Figure7 : Real problems requiring solution R&D on model construction methods

Prepared by the STFC

of specialists with advanced capabilities, for example, who can make full use of OR techniques, has been mentioned in policy recommendations in connection with supply chains, and joint research to establish practical academic system and curriculum to educate SCM specialist has begun in industry, universities, and government agencies.[38,39] Universities in the Asian countries have also embarked on training of specialists in SCM and logistics specialists with advanced capabilities by promoting cooperation with European and American universities.[40-42] In order to realize more advanced logistics in Japan, training of human resources on the same level as those in other countries is an urgent matter for promoting research and development, and does not permit further delay. By promoting logistics education and research in institutions of higher education, Japan can contribute to improvement of the international level in Asia in this area, and can become a promoter of international cooperation. This is considered one important policy issue for the future.

(2) Recognition of logistics, SCM, and OR as research fields

At present, the items of supply chain management and logistics are not included in the systems, fields, sub-divisions, or detailed items of the Grant-in-Aid for Scientific Research (Kakenhi) program of the Japan Society for the Promotion of Science (JSPS). In order to promote research and development in these fields, it is considered necessary to recognize these as research fields by adding them to the JSPS program. These items are also not included in the classification of technical fields by the Japan Science and Technology Agency (JST), and it is desirable that they be added to this classification. Furthermore, the items of management science, OR, and management engineering, which are considered important areas for promoting research in Europe and the United States, are not independently included in the technical fields of both programs. It is desirable that addition of these items also be realized quickly. A number of academic societies which already have histories and scales exist. Hence, if this type of action can be taken, it will

make a large contribution to activating research and development in the related academic societies.

(3) Interdisciplinary and interagency research system

If the development of OR to green logistics and services science is considered, promotion of interdisciplinary research combining science and engineering related fields, fields of social science such as economics, and the human sciences is demanded. In this kind of boundary-area research and development, there are latent possibilities for Japan to contribute to the development of OR at the world level by researching the strengths and weaknesses of the Japanese social system and realizing problem-solving from a more total viewpoint, without being bound by the reductionism approaches favored in Europe and the United States. A research support system for realizing these possibilities is required.

Acknowledgements

In writing this paper, the author interviewed more than 30 persons including university researchers and businesspeople. I wish to express my profound gratitude to all those concerned for the information and materials provided, which brought invaluable reference for this work. In particular, I wish to acknowledge the valuable suggestions received through discussions with the members of the supply chain-related research chapter of Operations Research Society of Japan, Japan Industrial Management Association, Japan Logistics Society, CSCMP Japan Roundtable, and Japan Chapter of Supply Chain Council.

References

[1] Ministry of Economy, Trade and Industry, Ministry of Health, Welfare and Labor, Ministry of Education, Culture, Sports, Science and Technology, "White Paper on Monodzukuri 2008 (Promotion Policy for Basic Technologies in Monodzukuri in FY2007)." (2008). (Japanese)

[2] Council of Supply Chain Management Professionals : http://cscmp.org/default.asp

[3] D. M. Lambert, J.R. Stock, "Strategic Logistics Management," Mcgraw Hill/Irwin (2000).

[4] Donald J. Bowersox, et al. (trans. by Eji Abo), "Supply Chain Logistics," Asakura Publishing Co., Ltd. (2004). (Japanese)

[5] Mitsuru Kuroda, ed.,"Supply Chain Management : Theory and Reality of Intercompany Cooperation," Asakura Publishing Co., Ltd. (2004). (Japanese)

[6] Japan Institute of Logistics Systems, "Report on Survey of Logistics Costs FY2007," (2008). (Japanese)

[7] Japan Productivity Center for Socio-Economic Development,"International Comparison of Labor Productivity (2006)." (Japanese)

[8] Takao Enkawa, "Strengths of Japan's Site Capabilities and Weaknesses of Management Seen from an International Comparison by SCM Performance Score Card," National Institute of Science and Technology Policy, Materials of an internal lecture (2008). (Japanese)

[9] Hiroshi Konno, "Useful Linear Equation – 50 Years of the Integer Programming 'Capricious Princess,'" Nippon-Hyoron-sha Co., Ltd. (2006). (Japanese)

[10] David Simchi-Levi, et al. (trans. supervised by Mikio Kubo), "Designing and Managing the Supply Chain – Concepts, Strategies, and Examples," Asakura Publishing Co., Ltd. (2006). (Japanese)

[11] Mikio Kubo, "Introduction to Supply Chain Optimization," Asakura Publishing Co., Ltd. (2004). (Japanese)

[12] Kazuyuki Izutsu,"Basic Concepts and Functional Requirements of International Logistics," Report of the Supply Chain Workshop, Operations Research Society of Japan, (2005). (Japanese)

[13] "Supply Chain 2020 Project Working Paper, A Review of the Leading Opinions on the Future of Supply Chains," Dec. 2004.

[14] Allan Woodburn, "An Investigation of Container Train Service Provision and Load Factors In Great Britain," University of Westminster (2006).

[15] Toshihisa Fujiwara, "Survey of Total Land, Sea, and Air Logistics in Belgium – As an Advanced Example for East Asia/Northern Kyushu – ," Preprints of the 25th National Conference of the Japan Logistics Society (2008). (Japanese)
[16] Japan Association for Trade with Russia and NIS, Materials of the 16th RBC International Business Conference, Noriaki Kojima, Director, Russia Office, Europe-Middle East-Africa Division, Economic and Industrial Policy Bureau, Ministry of Economy, Trade and Industry (METI), "The Japanese-Russian Economic Relationship and the Railway Field" : http://www.rotobo.or.jp/activities/rbc/rbc9_2.pdf (Japanese)
[17] Masaharu Iwama, "Globalization of Asia Considered from the Viewpoint of Logistics," Shizuoka Academic Press (2008). (Japanese)
[18] David Simchi-Levi, "A New Approach for Inventory Planning," Presentation, CLM2004 (2004).
[19] J. Shapiro, T. Dudas, "Building the Optimal Logistics & Distribution Network Strategy," Presentation, CLM2004 (2004).
[20] Irina Harris, Mohamed Naim and Dr. Christine Mumford," A Review of Infrastructure Modeling for Green Logistics," Cardiff University (2007).
[21] J. Quariguasi Frota Neto, J.M. Bloemhof-Ruwaard, J.A.E.E. van Nunen and H.W.G.M. van Heck, "Designing and Evaluating Sustainable Logistics Networks," Erasmus Research Institute of Management (ERIM) (2006).
[22] A. Sbihi, R.W. Eglese, "The Relationship between Vehicle Routing & Scheduling and Green Logistics- , A Literature Survey," Lancaster University Management School (2007).
[23] Julie Paquett, "The Supply Chain Response to Environmental Pressures," MIT (2005).
[24] Green Logistics Partnership Conference, "Guide to Green Logistics," Commerce and Information Policy Bureau, METI (2008). (Japanese)
[25] Japan Institute of Logistics Systems, "2nd Period Conference on Green Logistics in Japan, Report on Activities of the Green Logistics Study Group FY2007," Japan Institute of Logistics Systems (2008). (Japanese)
[26] Yuji Yano, Katsuhiko Hayashi, "Typology of Efforts by Japanese Companies to Address Logistics-related Environmental Issues," Ryutsu Keizai University (2007)
[27] Hiroki Kudo (Institute of Energy Economics, Japan), "Recent Trends and Future Outlook for Standardization of Greenhouse Gas Management," Japan Environmental Management Association for Industry (2008). (Japanese)
[28] Supply Chain Council, "Supply Chain Operation Reference Model V.9.0" (2008).
[29] M. Nagarajan, G. Sosic, "Game-theoric Analysis of Cooperation Among Supply Chain Agents: Review and Extensions," European Journal of Operations Research (2006).
[30] Japan Federation of Freight Industries "Logistics Data Book" (2007). (Japanese)
[31] National Innovation Initiative and Summit and Report, "Innovate America-Thriving in a World of Challenge and Change," Council of Competitiveness (2005).
[32] Hidenori Morimura et al., "Try & Try: Smart Waiting and Making Wait," NHK Archives (January 29, 1991). (Japanese)
[33] Seiichi Fujita et al., "Actuality of the Japanese Logistics Site and the Quest for Best Practice," Material Flow (August 2008). (Japanese)
[34] Eizo Kinoshita and Eiji Ono eds., "AHP and Conjoint Analysis," Gendai-Sugakusha (2004). (Japanese)
[35] Kaoru Tone, "Measurement and Improvement of Management Efficiency," Union of Japanese Scientists and Engineers (1993). (Japanese)
[36] Yoh Tanaka, "Seven-Eleven – Secrets of the Winner," Nikkei Publishing, Inc. (2006). (Japanese)
[37] Takao Enkawa, Ryuichi Ichikawa, "Managers Born from Logisticians," Management and Logistics Vol. 1 (2008). (Japanese)
[38] Jake Barr John Dischinger, "The Emerging Need for Supply Chain Talent: Working to fill the gap . . . , " Supply-Chain World 2008, Supply Chain Council (2008).
[39] D. Aquino and L. Draper, "Supply Chain Talent : State of the Discipline," AMR Research
[40] National University of Singapore : http://www.tliap.nus.edu.sg/
[41] The Hong Kong University of Science and Technology : http://www.ielm.ust.hk/

[42] Supply Chain & Logistics Institute : http://www.tli.gatech.edu/research/china/
"TLI-Asia Pacific Program Sets Goals for Next Five Years" :
http://www.tli.gatech.edu/news-events/release.php?id=188

Profile

Eizo TAKAI
Affiliated Fellow,NISTEP

Chief Technical Advisor, Frameworx Inc.
http://www.frame-wx.com
Mr. Takai's research field is OR/MS. He is also attracted by the interesting possibilities of application of these techniques to SCM and logistics in various industries. Regardless of age, he enjoys thinking about and realizing new ideas with young people. He feels that collaboration beyond the boundaries of generation and specialization is a source of new vitality and innovation.

(Original Japanese version: published in October 2008)

About SCIENCE AND TECHNOLOGY FORESIGHT CENTER

It is essential to enhance survey functions that underpin policy formulation in order for the science and technology administrative organiztions, with MEXT and other ministries under the general supervision of the Council for Science and Technology Policy, Cabinet office (CSTP), to develop strategic science and technology policy.

NISTEP has established the Science and Technology Foresight Center (STFC) with the aim to strengthen survey functions about trends of important science and technology field. The mission is to provied timely and detailed information about the latest science and technology trends both in Japan and overseas, comprehensive analysis of these trends, and reliable predictions of future science and technology directions to policy makers.

Beneath the Director are six units, each of which conducts surveys of trends in their respective science and technology fields. STFC conducts surveys and analyses from a broad range of perspectives, including the future outlook for society.

The research results will form a basic reference database for MEXT, CSTP, and other ministries. STFC makes them widely available to private companies, organizations outside the administrative departments, mass media, etc. on NISTEP website.

The following are major activities:

1. Collection and analysis of information on science and technology trends through expert network

— STFC builds an information network linking about 2000 experts of various science and technology fields in the industrial, academic and government sectors. They are in the front line or have advanced knowledge in their fields.

— Through the networks, STFC collects information in various science and technology fields via the Internet, analyzes trends both in Japan and overseas, identifies important R&D activities, and prospects the future directions. STFC also collects information on its own terms from vast resources.

— Collected information is regularly reported to MEXT and CSTP. Furthermore, STFC compiles the chief points of this information as topics for “Science and Technology Trends” (monthly report).

2. Reserch into trends in major science and technology fields

— Targeting the vital subjects for science and technology progress, STFC analyzes its trends deeply, and helps administrative departments formulate science and technology policies.

— The research results are published as articles for "Science Technology Trends" (monthly report).

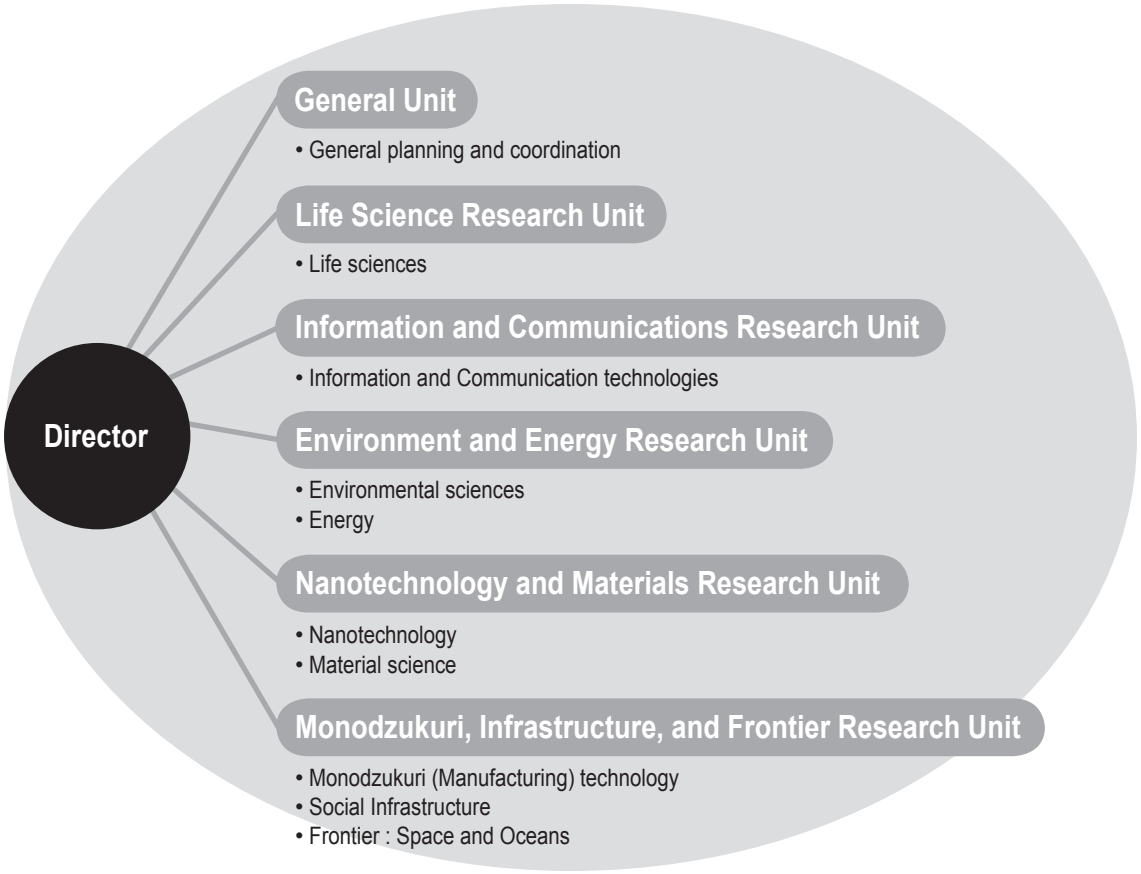
3. S&T foresight and benchmarking

— S&T foresight is conducted every five years to grasp the direction of technological development in coming 30 years with the cooperation of experts in various fields.

— International Benchmarking of Japan's science and engineering research also implemented periodically.

— The research results are published as NISTEP report.

Organization of the Science and Technology Foresight Center



＊Units comprise permanent staff and affliated fellows
＊The Center's organization and responsible are reviewed as required

NISTEP

Science & Technology Foresight Center

National Institute of Science and Technology Policy (NISTEP)
Ministry of Education, Culture Sports, Science and Technology, JAPAN

Science and Technology Trends —Quarterly Review 2009. 4