

ANRITSU
TECHNICAL
REVIEW

Sep.
2026
No. **34**

Anritsu Technical Review No.34
Issue on September 2026

© 2026 ALL RIGHTS RESERVED
by ANRITSU CORPORATION
5-1-1 Onna, Atsugi-shi, Kanagawa
243-8555, Japan

<https://www.anritsu.com>

Articles

Development of Test Solution for Next-Generation eCall by Shoichi Nakamura, Kenji Fujita, Kenji Takemoto, Wakana Inoue, Ryoji Yamamoto, Shinji Nara, Yuto Ichikawa	1
Development of 200G/Lane 60 GHz Optical Sampling Oscilloscope by Yuki Yokoyama, Shuichi Matsuda, Keisuke Kishida, Ryota Takasu	7
Optical Properties of Multilayer Graphene by Masahiro Kamada, Tomohiro Matsui	13
Trends and Overview of Evaluation Techniques in Image Recognition AI by Ryu Kidokoro, Lawrence Earl Carandang	22
Overview of Public-key Cryptography by Takashi Mori	30
Development of Near-Infrared Spectroscopy Module for Process Analytical Technology by Daisuke Oida, Noriaki Yamasaki, Masaru Koshihara, Takanori Saitoh	39
Development of X-ray Inspection System for Saving Labor and Effort by Tatsuki Komatsubara, Naoya Saito, Ryotaro Kitabayashi, Takashi Kanai, Kazunori Nakayama	48
Technical Demonstration of EV Powertrain Evaluation Combining Simulation and Real Power Supply by Takuya Yoda, Yuta Shinchi	54
Prism: An Open-Source Framework for Building XAML Applications in WPF for Instrumentation by Tom Roberts	62
Contribution to other Publications	68

Development of Test Solution for Next-Generation eCall

Shoichi Nakamura, Kenji Fujita, Kenji Takemoto, Wakana Inoue, Ryoji Yamamoto, Shinji Nara, Yuto Ichikawa

[Summary]

In Europe and other countries, eCall (emergency call) systems have long been operated to speed up life-saving assistance by automatically alerting emergency services when a car accident occurs. As mobile networks transition from W-CDMA/GSM to LTE/NR, a new NG eCall (Next Generation eCall) standard has been developed and type approval based on NG eCall has become mandatory from January 2026 in Europe. To obtain type approval, in-vehicle modules must pass the NG eCall conformance test (EN 17240:2024 standard). Since eCall relies on emergency call functionality, pre-verification using real mobile network connections is impractical, making a simulator-based test environment essential. Anritsu offers an NG eCall test solution in compliance with the NG eCall conformance test. Meanwhile, Europe has growing demand for Hybrid eCall—a combination of NG eCall and legacy eCall—aimed at expanding network coverage. Alongside this, demand for simulator-based test environments is also growing. Anritsu was the first company in the industry to provide a test solution supporting the Hybrid eCall conformance test (EN18052:2025 standard).

1 Introduction

The original eCall system was developed in Europe and various countries subsequently built their own systems based on the European standard. While there are multiple standards for eCall, the basic process is that when eCall is triggered, an emergency call is made between the in-vehicle system (IVS) and the Public Safety Answering Point (PSAP).

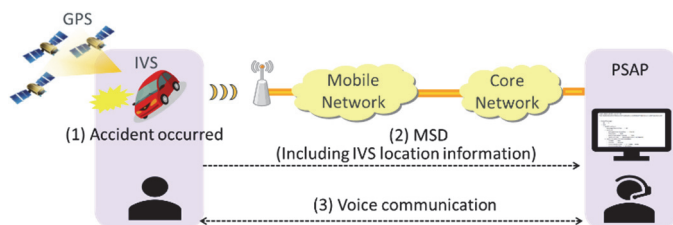


Figure 1 eCall Outline

(1) Accident Occurrence

When an accident occurs, an eCall-equipped vehicle automatically initiates an emergency call to the PSAP, triggered by the strong impact or sudden stop.

(2) MSD Message Sent

Accident data known as the MSD (Minimum Set of Data), which includes the vehicle's location, is transmitted to the PSAP, which then uses the received information to support rescue assistance.

(3) Voice Call

Once the MSD transmission is complete, a voice call is established with the PSAP using the microphone and speaker in the vehicle.

European eCall uses W-CDMA/GSM as its communication standard, transmitting MSD and voice data over PSTN (Public Switched Telephone Network). eCall has been mandatory in all newly registered vehicles in Europe since January 2018.*

NG eCall (Next Generation eCall) is being developed as a replacement for the conventional eCall system, with type approval under NG eCall becoming mandatory from January 2026. It uses LTE/NR (New Radio) as its communication standard and uses SIP (Session Initiation Protocol) and RTP (Real-time Transport Protocol) for communication with PSAPs. These upgrades offer improved voice quality and greater data reliability. Meanwhile, due to incomplete LTE network coverage in parts of Europe, as well as the extended continued operation of W-CDMA/GSM networks, there is growing interest in Hybrid eCall — a technology that combines NG eCall with conventional eCall to bridge the gap between the two systems.

Since eCall handles emergency calls, pre-deployment testing over live mobile networks is difficult. As a result, there is strong demand from automobile makers, in-vehicle device vendors, and module vendors for a simulator environment that can be used for testing purposes. To address this need, the following eCall test solutions have been developed. While eCall, NG eCall, and Hybrid eCall are all systems standardized in Europe, other countries have also developed their own variants based on the European standards. One example is South Korean eCall.

Table 1 eCall Test Solution Support Status

Region	Name	Support Status
Europe	eCall	Supported
Europe	NG eCall	Supported
Europe	Hybrid eCall	Supported
S. Korea	South Korean eCall	Supported

This paper describes the details of NG eCall and Hybrid eCall development in Europe which have been the focus of recent attention.

* Existing eCall provisions were abolished in January 2026 when EU NG eCall became mandatory.

2 Standards

The standards for eCall, NG eCall, and Hybrid eCall in Europe are developed by CEN (European Committee for Standardization) and ETSI (European Telecommunications Standards Institute). In-vehicle module vendors must pass the conformance tests defined by CEN/ETSI to demonstrate that the eCall functionality implemented in their modules complies with these standards. The conformance test standards are divided into test items for in-vehicle systems and test items for PSAPs. Our developed NG eCall and Hybrid eCall test solutions satisfy both sets of test items.

2.1 EN17240:2024 Specification

EN17240:2024 is the conformance test standard for European NG eCall. Since NG eCall must be fitted to all newly registered vehicles from January 2026, in-vehicle systems must pass the EN17240:2024 in-vehicle system test items prior to vehicle type approval testing.

2.2 EN18052:2025 Specification

EN18052:2025 is the conformance test standard for European Hybrid eCall. Although fitting of Hybrid eCall to all newly registered vehicles has yet to be mandated (as of December 2025), in-vehicle systems manufacturers are very interested due to its effectiveness in expanding coverage, and demand for testing is increasing. While EN17240:2024 testing is conducted in an environment with only LTE cells, EN18052:2025 testing is conducted in an environment with both LTE cells and either W-CDMA or GSM cells.

3 Market and Customers

The eCall test solution is primarily used for in-vehicle system conformance testing (e.g. EN17240, EN18052).

Users: Automotive manufacturers, in-vehicle device vendors, module vendors

Use Case 1: In-vehicle system conformance testing (EN17240, EN18052)

Use Case 2: In-vehicle system eCall functional testing (mobility, semi-normal testing)

Use Case 3: Vehicle type approval testing

4 eCall Test Solution System Configuration

The eCall test solution consists of the MX703330E simulating the PSAP functions, the MD8475A/B, MT8000A, and MX800070A simulating mobile communication network base stations, and the MX847503A test automation tool.

The MX703330E test functions are compatible with different eCall standards, so users can test based on their specific standard. The MD8475B and MT8000A pseudo-base stations support W-CDMA/GSM/LTE and LTE/NR, respectively.

The MX847503A tool automates manual test operations and PASS/FAIL evaluation; test execution is automated by running test packages incorporating various test sequences. We are developing test packages supporting in-vehicle system conformance tests, including NG eCall and Hybrid eCall.

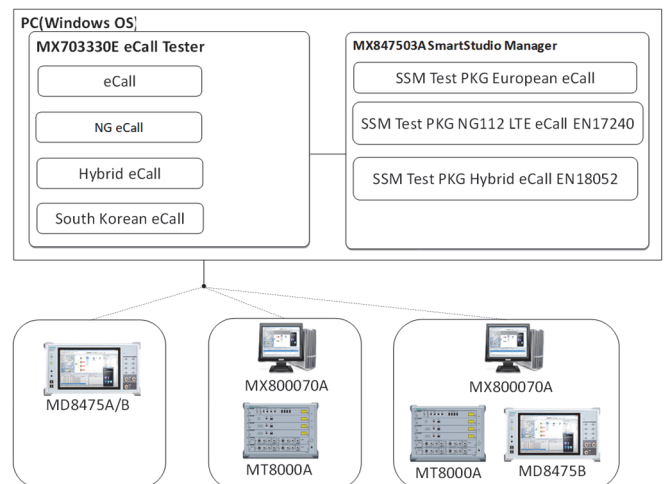


Figure 2 eCall Test Solution System

5 NG eCall Test Solution

The following three functions are the key technical requirements for an NG eCall test solution meeting the NG eCall conformance test standard (EN17240:2024).

- Pseudo-base station function
- PSAP NG eCall Semi-Normal Function
- PSAP NG eCall In-band Function

5.1 Pseudo-base Station Function

When an in-vehicle system connects to an LTE cell, the base station transmits control signals necessary for connecting the in-vehicle system to the LTE cell. These control signals are called Broadcast Information. It includes the “eCall over IMS Supported (ECL)” parameter defined in 3GPP TS36.331 Release 14 and later. The ECL parameter indicates whether an LTE cell supports NG eCall connectivity. NG eCall conformance test items verify that the behavior of the in-vehicle system changes according to the value of this ECL parameter. To address this, we implemented a function in the pseudo-base station equipment that enables transmission of Broadcast Information with configured ECL parameter.

When an in-vehicle system requests connection to an LTE cell, the LTE cell may reject the connection sometimes. NG eCall conformance test items verify the behavior of the in-vehicle system when connection is rejected. To support this test function, we implemented a connection rejection function on the pseudo-base station side. Figure 3 shows a reference example. The pseudo-base station responds to a connection request from the in-vehicle system (NAS: PDN Connectivity Request message) by returning a connection rejection (NAS: PDN Connectivity Reject message), thereby rejecting the connection.

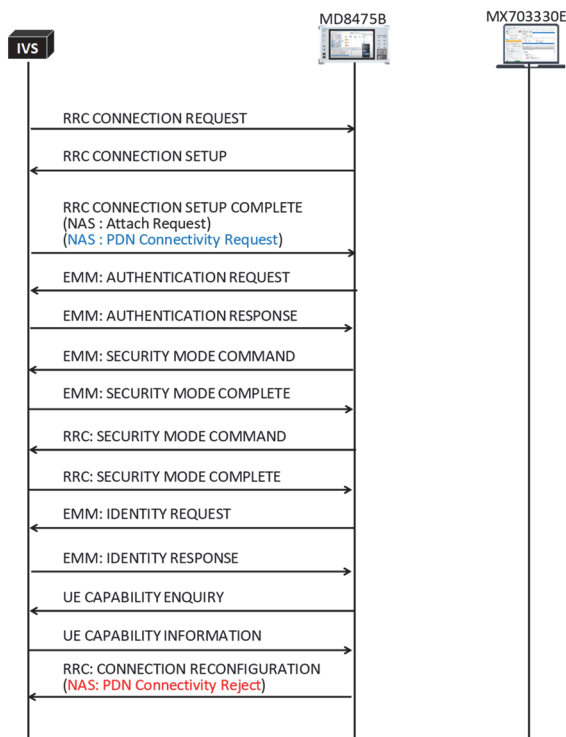


Figure 3 Connection Rejection Function

5.2 PSAP: NG eCall Semi-Normal Function

When an in-vehicle system requests an NG eCall connection to a PSAP, the PSAP may reject the connection sometimes. NG eCall conformance test items verify the behavior of the in-vehicle system when connection is rejected; this testing is called “NG eCall semi-normal testing”. The NG eCall connection request from the in-vehicle system to PSAP is made via SIP. To support this test function, the NG eCall semi-normal test function has been implemented in the MX703330E. Figure 4 shows the sequence when NG eCall connects normally.

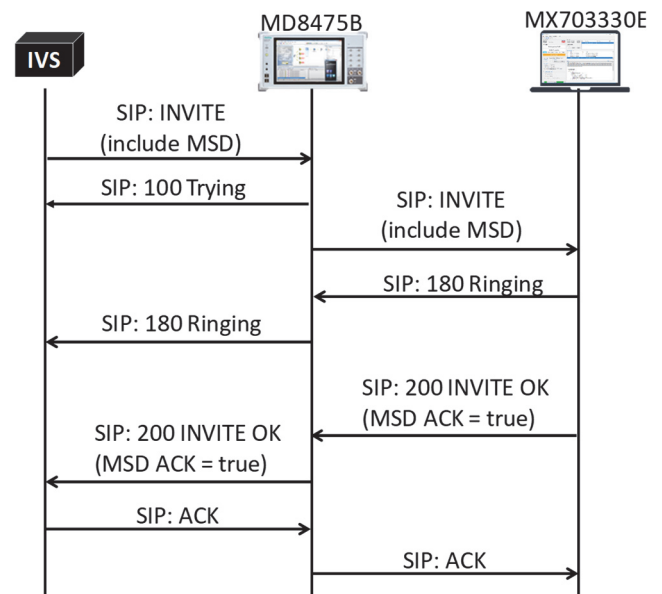


Figure 4 Normal NG eCall Sequence

Figure 5 shows an example of the sequence when the NG eCall semi-normal function is used. The PSAP returns SIP: 486 Busy in response to an NG eCall connection request (SIP: INVITE) from the in-vehicle system, thereby rejecting the connection.

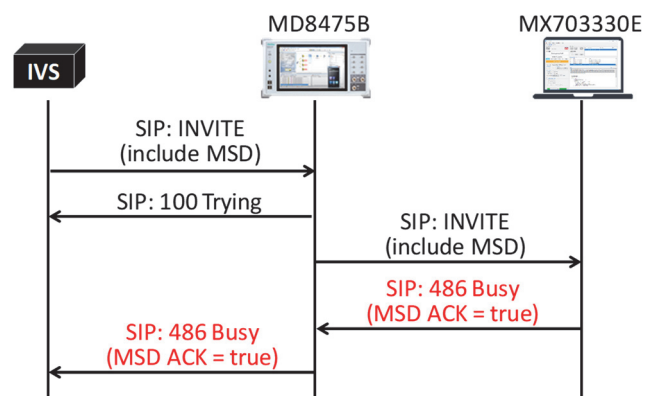


Figure 5 NG eCall Semi-Normal Sequence

5.3 PSAP: NG eCall In-band Function

While the in-vehicle system supports both legacy eCall and NG eCall, the PSAP may only support legacy eCall. The in-vehicle system first attempts to connect to the PSAP using NG eCall, but upon receiving a specific connection rejection message from the PSAP, it switches to legacy eCall while remaining on the LTE network and attempts to continue communication. In this case, both the in-vehicle system and the PSAP modulate the MSD into audio data using the “in-band modem” used by legacy eCall, map it to RTP, and transmit it. This behavior is called “NG eCall In-band.”

The NG eCall conformance tests include test items to verify that the NG eCall In-band functionality of the in-vehicle system operates correctly. To support this test item, the MX703330E implements the NG eCall In-band functionality. Figure 6 shows the behavior of the in-vehicle system and the MX703330E during the transition from NG eCall to NG eCall In-band.

When using Hybrid eCall, both eCall connection requests and NG eCall connection requests are sent from the in-vehicle system to the PSAP. Therefore, the PSAP must simultaneously listen for both types of connection requests. To meet this requirement, we implemented a simultaneous standby function in the MX703330E. Additionally, to allow users to easily check the connection status of the PSAP, we implemented a function to display the connection status on the MX703330E PSAP Status screen.



Figure 7 MX703330E PSAP Status Screens

6 Hybrid eCall Test Solution

Testing Hybrid eCall conformance (EN18052:2025 standard) in a simulated environment, requires a pseudo-base station supporting LTE, W-CDMA/GSM, and NR. A simulation environment satisfying the standard requirements was constructed by combining the MD8475B supporting W-CDMA/GSM/LTE with the MT8000A supporting NR/LTE.

7 NG eCall/Hybrid eCall Conformance Test Package

The NG eCall conformance tests (EN17240:2024 standard) and Hybrid eCall conformance tests (EN18052:2025 standard) contain a large number of test items, and each test procedure is complex. The test automation tool MX847503A handles external control of the MX703330E.

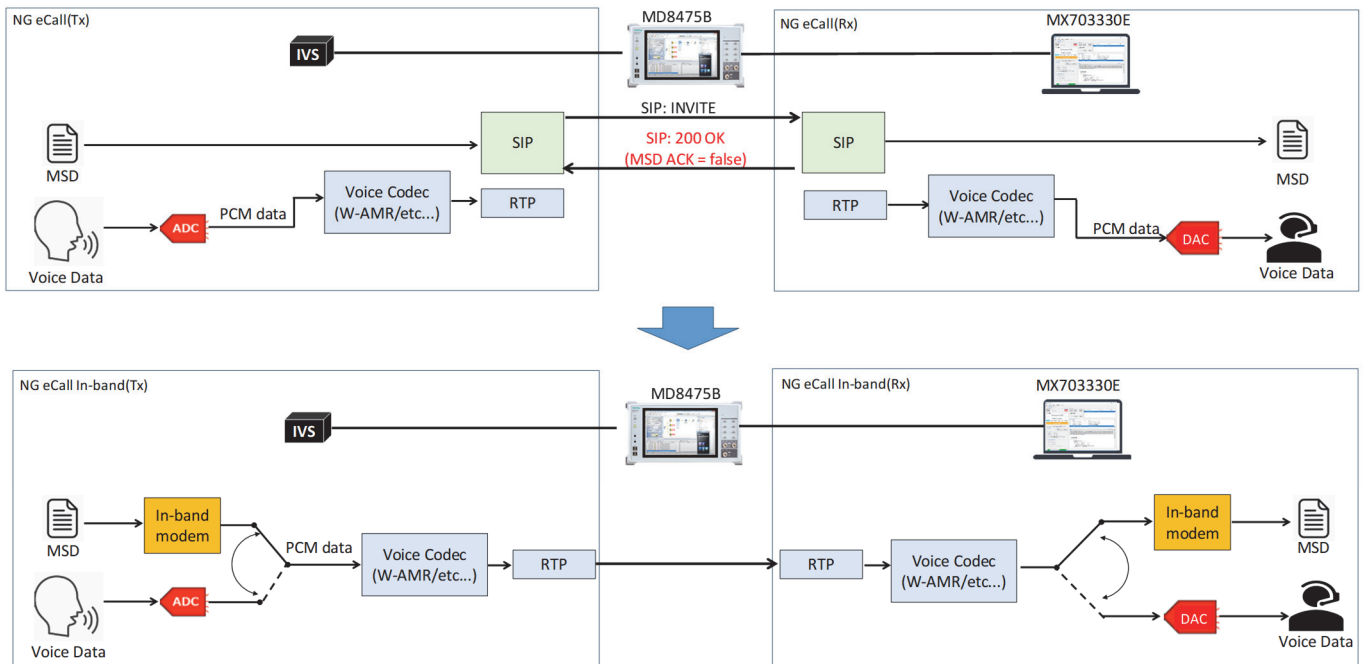


Figure 6 NG eCall In-band Operation

Automated eCall testing is achieved by creating the MX703330E GUI operation and test verification procedures as test cases that run on the MX847503A. This collection of MX847503A test cases is called a test package. We have developed a test package product supporting both the NG eCall conformance tests (EN17240:2024 standard) and the Hybrid eCall conformance tests (EN18052:2025 standard).

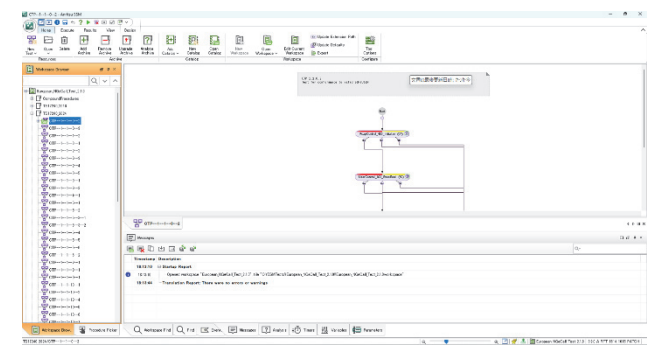


Figure 8 Test Case Execution Screen (EN17240:2024)

The test package saves test results as log data and can generate test reports from that log data. Reports can be output and saved in PDF format.

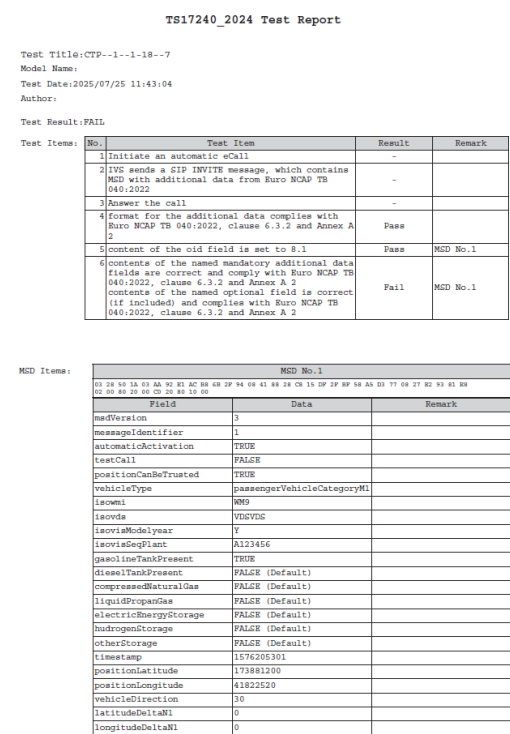


Figure 9 Test Report (EN17240:2024)

8 Conclusions

To meet growing demand for NG eCall and Hybrid eCall testing, which are becoming the mainstream in Europe, Anritsu has developed an eCall test solution supporting both

the NG eCall conformance tests (EN17240:2024 standard) and the Hybrid eCall conformance tests (EN18052:2025 standard). Furthermore, a test package has been developed as a test automation solution that enables users to perform conformance tests easily.

eCall is an important system for enabling rapid assistance in the event of a traffic accident. Development and adoption of new standards is expected to accelerate as transportation infrastructure continues to advance across countries worldwide. Anritsu is continuing efforts to build a safe mobile society by anticipating global safety requirements and standards trends and providing test solutions.

References

- 1) EN17240:2024 Intelligent transport systems - ESafety - ECall end to end conformance testing for IMS packet switched based systems
- 2) EN18052:2025 Intelligent transport systems - ESafety - ECall end to end conformance testing for eCall H LAP in hybrid circuit switched/packet switched network environments

Authors



Shoichi Nakamura
1st Team
Product Development Dept.
IoT Test Solutions Div.
Test & Measurement Company



Kenji Fujita
1st Team
Product Development Dept.
IoT Test Solutions Div.
Test & Measurement Company



Kenji Takemoto
1st Team
Product Development Dept.
IoT Test Solutions Div.
Test & Measurement Company



Wakana Inoue
Software Engineering Dept.
Engineering Div.



Ryoji Yamamoto
Software Engineering Dept.
Engineering Div.



Shinji Nara
Software Engineering Dept.
Engineering Div.



Yuto Ichikawa
Software Engineering Dept.
Engineering Div.

Development of 200G/Lane 60 GHz Optical Sampling Oscilloscope

Yuki Yokoyama, Shuichi Matsuda, Keisuke Kishida, Ryota Takasu

[Summary]

Demand is growing rapidly for testing optical transceivers meeting high-speed Ethernet standards such as 800GbE and 1.6TbE in order to support high-speed transmissions required by data centers faced with rising data volumes. To solve this challenge in accordance with the IEEE 802.3dj standard, we developed the MP2110A-080 60 GHz optical sampling oscilloscope to simultaneously measure PAM4 signals of up to 120 Gbaud across four lanes. Evaluation demonstrated an amplitude deviation of 0.22 dB from the optical reference receiver required frequency response. Furthermore, the maximum variation in key measurement items over an input level range of -1 to -4 dBm was 0.25 dB for TDECQ, 0.07 dB for Outer ER, and 0.5% for Linearity, achieving high conformance with ideal characteristics and high-accuracy testing.

1 Introduction

The widespread rollout of cloud computing services using AI and machine learning is rapidly increasing the amount of information handled by data centers^{1), 2)}. Expanding the transmission capacity of servers and network equipment in data centers is an urgent priority in addressing this challenge, and the transition to 800GbE and 1.6TbE high-speed Ethernet—speeds that far surpass conventional rates—is progressing. In particular, demand for optical transceivers using the OSFP (Octal Small Form Factor Pluggable) form factor operating at 106.25 Gbaud PAM4 per lane (106 Gbaud in this article) is forecast to surge from 2025. The total bit rate at the physical layer is approximately 212.5 Gbit/s, and Ethernet standards treat this as a 200 Gbit/s class per lane (200G/lane) accounting for overhead such as FEC.

Development and manufacturing of high-speed optical transceivers uses optical sampling oscilloscopes (scopes hereafter) to evaluate signal waveform quality and for production-line testing. In 2017, Anritsu developed³⁾ the MP2110A BERTWave™ integrating a 25 Gbit/s 4-channel BERT (Bit Error Rate Tester) with a scope for optical transceiver development and manufacturing. Subsequently, Anritsu has offered new solutions addressing the demand for high-speed optical transceiver testing primarily in the manufacturing market, commercial PAM4 (4 Pulse Amplitude Modulation) signal analysis, and a 53 Gbaud-compatible CRU (Clock Recovery Unit) option⁴⁾.

However, conventional scopes are limited to measurements up to 60 Gbaud and are unable to support 800GbE

and 1.6TbE optical transceivers outputting signals at 106 Gbaud and requiring a wider bandwidth. Therefore, we have extended the bandwidth of the MP2110A scope and developed a 60 GHz bandwidth 4-channel model (MP2110A-080 hereafter) that can measure PAM4 signals up to 120 Gbaud simultaneously across four lanes.

Developing this option to measure optical waveforms output from 106 Gbaud transceivers simultaneously across four lanes required new designs for an O/E (Optical to Electrical) conversion module with an optical bandwidth of 60 GHz and a sampler module with an electrical bandwidth of 80 GHz. This article explains the key points in designing these component technologies and presents methods for analyzing PAM4 signals used in high-speed optical transceivers. It also reports the validity of measurement results obtained using this option.

2 Oscilloscope Outline

2.1 Overview and Structure

Figure 1 shows the front of the MP2110A-080. The new cabinet design is based on the MP2110A and accommodates the MP2110A-080 scope board in the top slot and the MP2110A-055 53 Gbaud CRU board in the bottom slot. While the cabinet structure follows that of the MP2110A, cooling has been enhanced to suppress heat generated by the FPGA (Field-Programmable Gate Array) mounted on the scope board, ensuring stable measurement while accommodating the thermal load associated with high-speed signal processing.



Figure 1 External View of MP2110A with 080 Option Installed

Figure 2 shows the MP2110A-080 block diagram. The scope has four input terminals (Ch A to Ch D) for the 4-channel data signals, as well as a trigger clock input terminal (Trigger Clock Input) for synchronizing measurements. The CRU has a Data Input terminal for the clock recovery signal, as well as a Data Output terminal for capturing the internally branched signal. It also has a Recovered Clock Out terminal to output the clock signal recovered from the input data signal, which is supplied as a trigger to the scope. The CPU controls the Scope and CRU systems via Ethernet, managing the overall system operation by configuring measurement conditions and transferring data.

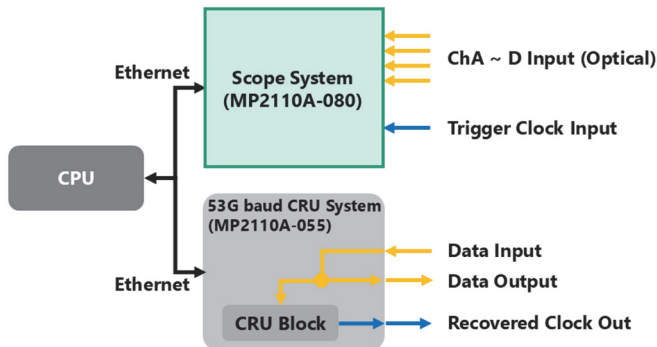


Figure 2 MP2110A-080 Block Diagram

2.2 Scope Configuration

Figure 3 shows the detailed MP2110A-080 scope block diagram. First, each signal input to optical channels Ch A to Ch D is converted to an electrical signal by an O/E conversion module (O/E module hereafter) and is then fed to a sampler module that samples each channel input signal based on the timing signal from the trigger system. The signal output from the sampler module is amplified by the IF (Intermediate Frequency) circuit and then converted to digital data by the ADC (Analog to Digital Converter). The converted data is analyzed by software processing at the FPGA and CPU, and finally displayed as Eye pattern analysis results.

The trigger system receives a clock signal synchronized

with the input signals of Ch A to Ch D from the Trigger Clock Input terminal. The first-stage Pre-Scaler 1 adjusts the supplied trigger clock frequency to fall within the input range of the DDS (Direct Digital Synthesizer). If the trigger clock frequency is lower than the DDS input range, a PLL (Phase Locked Loop) is used to up-multiply it. The DDS adjusts the output frequency to control the sampling timing. Unwanted components in the DDS output, such as image and harmonic components, are removed by the subsequent Band Pass Filter. The signal is then converted by Pre-Scaler 2 to a frequency at which the sampler module can operate. Finally, high-speed pulses are generated by the SRD (Step Recovery Diode) driver for output as timing signals driving the sampler module. The MP2110A-080 samples high-speed signals at 106 Gbaud using this series configuration.

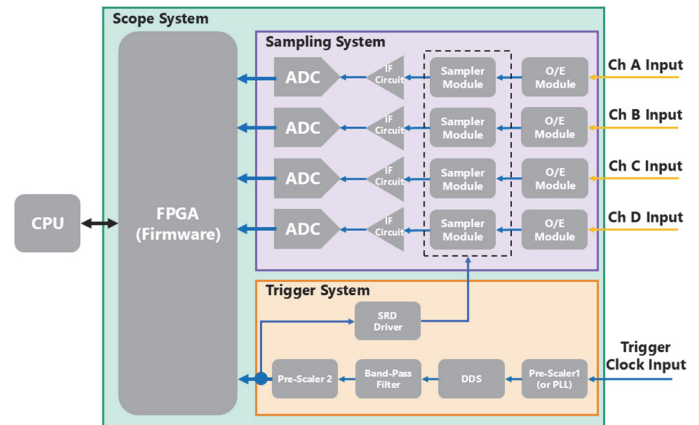


Figure 3 MP2110A-080 Scope Block Diagram Details

3 Design Requirements

IEEE 802.3dj⁵⁾ specifies a 4th-order Bessel-Thomson response with a 3-dB bandwidth of 53.125 GHz as the characteristic of the reference receiver, which serves as the measurement standard for 106 Gbaud optical signals. However, in practice it is difficult to accurately approximate a wide-band characteristic of 60 GHz in hardware, and the standard states that deviations from the ideal response can be compensated for in software. In order to account for compensation accuracy, the scope hardware requires a wider input bandwidth than the 3-dB bandwidth. To achieve this wider bandwidth, the MP2110A-080 incorporates a new in-house developed O/E module with a 3-dB bandwidth of 60 GHz, and a sampler module with a 3-dB bandwidth of 80 GHz, assuring a system-wide bandwidth of 60 GHz. The developed O/E and sampler modules are described in the next two sections.

3.1 60 GHz O/E Module Development

Figure 4 shows the internal block diagram of the 60 GHz O/E module. The bandwidth of the existing MP2110A O/E module is 35 GHz, which is insufficient for measuring 106 Gbaud signals. To measure signals up to 120 Gbaud, the bandwidth must be extended to 60 GHz, taking the Nyquist frequency into account. To achieve a wider bandwidth, a new high-frequency module was developed that integrates the optical system from the optical fiber, a PD (Photo Diode), and a TIA (Transimpedance Amplifier) into one unit. Using a lens structure in the optical system made it possible to use a PD with a light-receiving diameter of 12 μm . Reducing the PD light-receiving diameter reduces the PD capacitance and shortens the charge travel distance, resulting in a faster response time. As a result, the PD bandwidth was extended by approximately 1.9 times compared to a conventional 20 μm PD.

A new MMIC (Monolithic Microwave Integrated Circuit) dedicated to this module was designed for the subsequent TIA using the company's in-house ultra-high-speed InP process. Furthermore, parasitic elements were reduced by positioning the PD bare chip and TIA MMIC chip as close as possible to each other and mounting using the shortest possible wire bonding interconnects. The TIA chip output is routed using a quartz thin-film substrate with a V-connector, achieving wideband electrical signal output. These design choices achieved a bandwidth exceeding 60 GHz with noise characteristics of 8.4 μW (typ.) at 53 GHz and distortion characteristics of 0.7 dB (typ.) at an optical input of 4 mW.

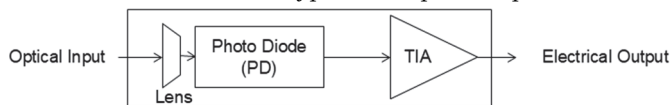


Figure 4 60 GHz O/E Module Block Diagram

3.2 80 GHz Wideband Sampler Module Development

Figure 5 shows the equivalent circuit diagram for the developed sampler module. The module structure is suited to high-frequency operation and consists primarily of a thin-film substrate and MMIC. The input converted to an electrical signal by the preceding O/E module is supplied to the RF input of the sampler module. A strobe pulse corresponding to the sampling period supplied from the external trigger circuit is applied to the LO input, and track-and-hold operation is performed by the internal sampling MMIC. The resulting IF output voltage is then amplified in the next stage, digitally converted by the

ADC, and finally processed as a waveform. The following technologies extended the bandwidth of the conventional sampler module from 40 GHz to 80 GHz.

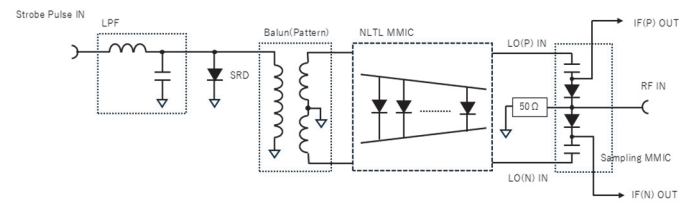


Figure 5 80 GHz Wideband Sampler Module Block Diagram

(1) New NLTL MMIC Design

An NLTL (Nonlinear Transmission Line) is a circuit with a structure in which nonlinear elements are arranged periodically along a transmission line, and as the input waveform propagates, the phase velocity varies due to voltage dependence. At high potential, the capacitance increases and the phase velocity decreases, while at low potential, the capacitance decreases and the phase velocity increases. As a result, the trailing low-potential side catches up with the leading high-potential side, and the edge is compressed in time, causing the falling edge of the waveform to steepen.

This nonlinear effect generates harmonic components, making it possible to obtain high-frequency pulses and wideband signals. In this development, the company's in-house NLTL technology was applied to newly design an MMIC for the sampler module. Previously, pulse shortening of the strobe input pulse was achieved using only an SRD, but the addition of an NLTL MMIC enabled further pulse shortening and extension of harmonic components, contributing to the wider bandwidth.

(2) New Design of Low-Parasitic-Capacitance Sampling Circuit MMIC

The sampling circuit MMIC consists of track diodes operating as high-speed switches and a hold capacitor that retains the sampled charge. Reducing parasitic capacitance is key to extending the sampling bandwidth. To this end, the diode element cell was miniaturized using Anritsu's GaAs process, minimizing parasitic capacitance.

(3) Flip-Chip Mounting of MMIC on Thin-Film Substrate

The MMIC in the conventional sampler module was connected to the thin-film substrate by wire bonding, but this method causes parasitic inductance due to the wire length, which was a limiting factor in bandwidth extension. Therefore, the MMIC connection was changed to flip-chip mounting,

significantly reducing the parasitic inductance component. Figure 6 shows the analysis results of the NLTL output pulse waveform and frequency spectrum. The NLTL output pulse has a steep fall time of 10 ps, and the corresponding frequency spectrum extends beyond 100 GHz. These efforts achieved an 80 GHz bandwidth while improving the conversion efficiency of the sampling output, minimizing parasitic capacitance, and without degrading noise characteristics.

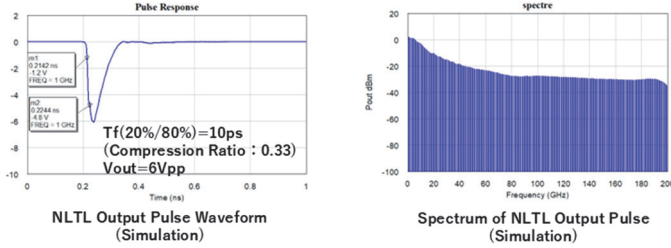


Figure 6 NLTL Output Pulse Waveform and Frequency Spectrum

3.3 PAM4 Analysis

IEEE 802.3dj specifies the measurement requirements for 106 Gbaud PAM4 signals. As with 53 Gbaud PAM4 signals, the primary evaluation parameter is TDECQ (Transmitter and Dispersion Eye Closure Quaternary)⁴. Table 1 shows the parameter differences for 106 Gbaud signals as specified in IEEE 802.3dj Draft 2.2 compared to 53 Gbaud signals specified in IEEE 802.3bs⁶. As speeds increase, simply extending the bandwidth is no longer sufficient to ensure communication quality, and channel degradation and reduced system margin are more pronounced. For this reason, the IEEE 802.3dj measurement requirements are more stringent in order to guarantee signal quality across the entire system. It should be noted that since IEEE 802.3dj is currently in draft form, its content is subject to future revision.

Table 1 Differences in Measurement Requirements Defined by IEEE 802.3dj Draft 2.2 and IEEE 802.3bs

Requirement	IEEE 802.3dj	IEEE 802.3bs
3-dB Bandwidth	53.125 GHz	26.5625 GHz
Target SER ^{*1}	4.56E-4	4.8E-4
Equalizer Specification	15 Tap FFE ^{*2} + DFE ^{*3}	5 Tap FFE
Tap Coefficient Constraints	Added Stricter Numerical Upper Bound Constraints	Basic Constraints Only

*1 SER: Symbol Error Ratio

*2 FFE: Feed Forward Equalizer

*3 DFE: Decision Feedback Equalizer

The IEEE 802.3 standard specifies multiple evaluation metrics to ensure interoperability of optical transceivers

used in high-speed Ethernet communications. TDECQ is one typical metric for evaluating the transmission performance of optical transceivers using PAM4 signals. Figure 7 shows a PAM4 waveform and a TDECQ measurement image. The amount of noise added to the transmitted signal is evaluated until the target SER is reached using an ideal receiver model. A smaller TDECQ value indicates a higher quality optical transmission waveform, and is defined by Eq. (1).

$$TDECQ = 10 \log_{10} \left(\frac{\text{Outer OMA}}{6} \times \frac{1}{Q_t R} \right) \text{ (dB)} \quad \dots (1)$$

Q_t : SER (Symbol Error Rate) specified by IEEE802.3dj

$$Q_t = 3.428 \text{ at } 4.56 \times 10^{-4}$$

R : Additional noise required to achieve SER = 4.56×10^{-4}

Measurement requires applying the reference receiver characteristics defined in the standard (4th-order Bessel-Thomson response) and the reference equalizer (FFE and DFE), as well as measurement under conditions simulating actual system operation. This allows fair evaluation of transmitter quality from the receiver's perspective by normalizing to the receiver level, while suppressing discrepancies in evaluation caused by measurement instrument characteristics. Moreover, this ensures that identical standard-based results are obtained from different vendors, thereby guaranteeing interoperability.

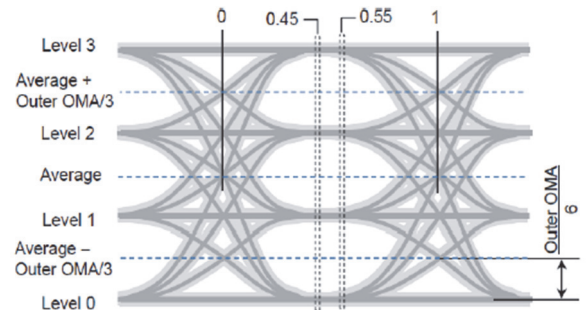


Figure 7 PAM4 Waveform and TDECQ Measurement

Figure 8 outlines the TDECQ measurement process. First, one period of the signal output from the transmitter is captured. The pattern used for measurement is SSPRQ (Short Stress Pattern Random Quaternary), as defined in IEEE 802.3. This pattern is designed to reproduce the effects of ISI (Inter-Symbol Interference) and nonlinear distortion within a short period, and the standard explicitly specifies use of SSPRQ for TDECQ measurement. Next, a 4th-order Bessel-Thomson filter as defined in the standard is applied to the acquired signal waveform, after which the tap coefficients of the FFE are calculated. The tap coefficients must be optimized to minimize the TDECQ value within the constraints

specified by the standard. However, since the tap resolution is extremely high and an exhaustive search of all combinations would result in enormous computational cost, IEEE 802.3 specifies optimization using the MMSE (Minimum Mean Square Error) algorithm. The MMSE method minimizes the mean square error (MSE) between the received signal and the ideal signal, and MSE is defined by Eq. (2).

$$MSE = E[|y(n) - d(n)|^2] \dots \quad (2)$$

$y(n)$: Output signal after FFE

$d(n)$: Ideal signal

In actual calculations, tap coefficients that minimize the MMSE are determined by matrix operations as expressed by Eq. (3).

$$w = R^{-1}p \dots \quad (3)$$

w : Vector of determined tap coefficients

R : Autocorrelation matrix ($X^T X$)

p : Cross – correlation vector ($X^T d$)

X : Data matrix

d : Ideal signal corresponding to the data matrix

After obtaining the tap coefficients, FFE is applied to the signal waveform. Subsequently, random noise following a Gaussian distribution is added. As shown in Figure 7, the threshold values corresponding to the four levels of the PAM4 signal are determined based on the center value of each Eye opening. Using these threshold values, the amount of noise required to reach the target SER is searched, and TDECQ is finally calculated based on that noise amount.

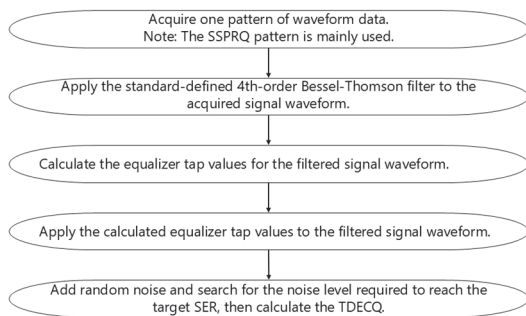


Figure 8 TDECQ Measurement Process

4 Performance Test

4.1 System Frequency Characteristics

Figure 9 compares the measured frequency response of the entire MP2110A-080 system and the ideal 4th-order Bessel-Thomson response. Here, the system response has been

software-corrected to approximate a 4th-order Bessel-Thomson response with a 3-dB bandwidth of 53.125 GHz. The system exhibits a flat response close to the ideal characteristics in the low-frequency region, and follows the ideal up to 60 GHz with a gradual roll-off in the high-frequency region. The standard deviation of the amplitude error of the system response relative to the ideal characteristics at each frequency point is 0.22 dB.

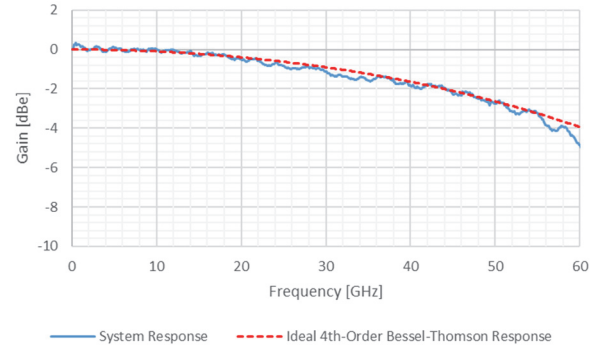


Figure 9 System Frequency Characteristics with Applied Digital Filter (53.125 GHz)

4.2 Optical Signal Waveform and Sensitivity Performance

The validity of the optical signal waveform and sensitivity performance were verified using a scope with the developed 080 option. Figure 10 shows the measurement setup for verification. The measurement conditions are as follows: 106 Gbaud rate, SSPRQ input pattern, and only a 15-tap FFE applied as the equalizer.

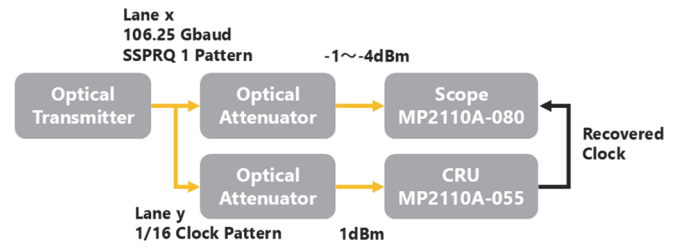


Figure 10 System Evaluation Setup

Figure 11 shows an example of the measured waveform when the optical input level to the scope is -2 dBm. It exhibits little waveform distortion, with an excellent TDECQ value of 2.11 dB. Figure 12 shows the results of evaluating the optical input level dependence for the main measurement items: TDECQ, Outer ER (Extinction Ratio), and Linearity. The scope optical input level was adjusted using an external variable optical attenuator. For a scope with ideal characteristics, the measurement results would be independent of the optical input level, and the graph plots would be flat. The measurement results show the average values

obtained from 10 repeat measurements under each condition. The variation in each measurement item over input levels from -1 to -4 dBm confirms the stable results regardless of input level, with peak-to-peak values of 0.25 dB for TDECQ, 0.07 dB for Outer ER, and 0.5% for Linearity.

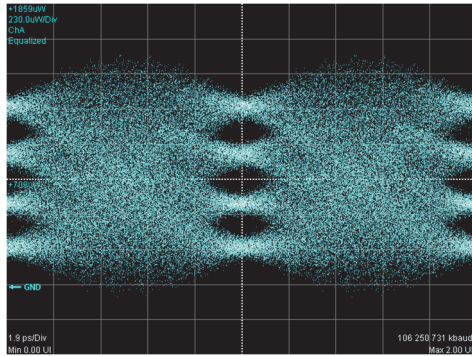


Figure 11 Example of observed optical waveform (106 Gbaud)

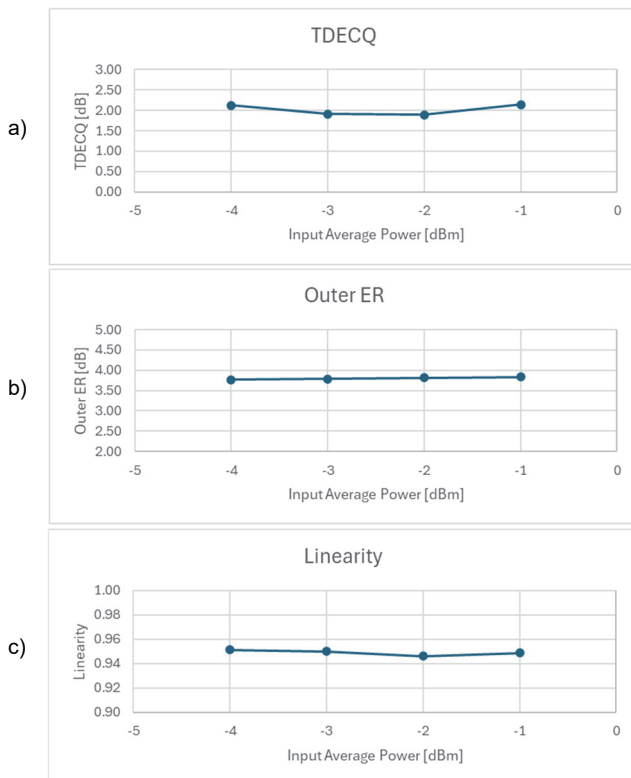


Figure 12 Optical input level dependence of each measurement result
a) TDECQ, b) Outer ER, c) Linearity

5 Conclusions

The rapid increase in data volumes handled by data centers is driving demand for 800GbE- and 1.6TbE-compatible optical transceivers. To meet these test requirements, we have extended the bandwidth of the MP2110A scope by developing the MP2110A-080 option for simultaneously

measuring signals of up to 120 Gbaud across four lanes. Stable performance with little input level dependence over a range of -1 to -4 dBm is achieved for the key TDECQ, Outer ER, and Linearity measurements, supporting provision of an 800GbE- and 1.6TbE-compatible 4-channel oscilloscope for high-accuracy measurements.

Going forward, we will continue to contribute to improving the production efficiency and evaluation quality of optical transceivers that support high-speed, large-capacity communication infrastructure by providing optimal solutions for testing optical transceivers.

References

- 1) Ethernet Alliance 2025 Ethernet Roadmap-Digital Version
- 2) IEEE Transmitter™ AI Is Changing The Landscape of Data Centers
- 3) T. Murakami, Y. Yokoyama, T. Kamizono, F. Hirabayashi, "Development of Sampling Oscilloscope MP2110A," Anritsu Technical Review No.26, pp.29-34 (Sep. 2018).
- 4) Y. Yokoyama, T. Murakami, K. Kishida, H. Yoshino, "400GbE Multichannel High-Accuracy TDECQ Measurement," Anritsu Technical Review No.29, pp.15-20 (Sep. 2021).
- 5) IEEE Std 802.3dj™ Draft 2.2
- 6) IEEE Std 802.3bs™

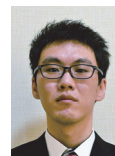
Authors



Yuki Yokoyama
Product Development Dept.
Service Infrastructure Solutions
Test & Measurement Company



Shuichi Matsuda
Product Development Dept.
Service Infrastructure Solutions
Test & Measurement Company



Keisuke Kishida
Product Development Dept.
Service Infrastructure Solutions
Test & Measurement Company



Ryota Takasu
Product Development Dept.
Service Infrastructure Solutions
Test & Measurement Company

Publicly available

Optical Properties of Multilayer Graphene

Masahiro Kamada, Tomohiro Matsui

[Summary]

Multilayer graphene is a promising candidate for optical device applications because its excellent optical properties can be controlled by the number of layers. This study measured the optical contrast, reflectance, and Raman scattering of multilayer graphene ranging in thickness from monolayer up to 107 nm, and comprehensively evaluated the optical properties, including correlations between these characteristics. The relationship between the number of graphene layers and color, which had previously only been understood qualitatively, was quantified for the first time by optical contrast, and demonstrated that the thickness can be determined from the contrast with an accuracy of ± 1 nm for thicknesses up to approximately 40 nm. We also demonstrated that the reflectance can be explained by the Fresnel equations by treating multilayer graphene as a uniform medium that ignores its layered structure. In addition to multiple reflections of the Raman scattered light, Raman spectroscopy showed the added importance of interference effects. Relationships between each of the characteristics were clarified by performing these measurements on the same sample.

1 Introduction

Graphene is a sheet-like material one carbon atom thick, possessing various outstanding physical properties, and is expected to find applications across a wide range of fields. Thanks to its atomic-scale thinness combined with excellent electrical and thermal conductivity, and that it can be exposed to the external environment, monolayer graphene is anticipated for use in applications such as electronic devices and molecular sensors. On the other hand, interlayer interactions when stacking graphene in multiple layers can enhance its original properties or give rise to entirely new ones, so applied research into multilayer graphene is also advancing. Of particular interest among these applications is its use in optical devices. Monolayer graphene exhibits a remarkably high absorbance of 2.3%, independent of wavelength¹⁾, and this absorbance can be controlled by varying the number of layers or the stacking structure^{2),3)} which has led to advanced research into applications such as antennas for electromagnetic fields and light.

Practical use of multilayer graphene requires knowing the number of layers (thickness) precisely. The thickness of a single graphene layer is approximately 0.334 nm, and accurate measurement uses instruments such as atomic force microscopes (AFMs) and scanning tunneling microscopes (STMs). However, while these instruments offer high resolution, considerable time is required to perform measurements. There is a simpler method for determining thickness that makes use of the “color” of graphene. On its own, graphene is transparent, but it can be “seen” under an optical microscope⁴⁾ by placing it on a Si substrate with a thermally grown oxide (SiO_2) film of appropriate thickness. By using the contrast relative to the substrate extracted as RGB or grayscale intensities from optical microscope images, it is possible to distinguish differences in thickness down to a single atomic layer for up to approximately six layers⁵⁾. The fact that the number of layers can be identified easily without using expensive instruments such as AFM may be regarded as one of the factors behind the explosive spread of graphene research and its applications.

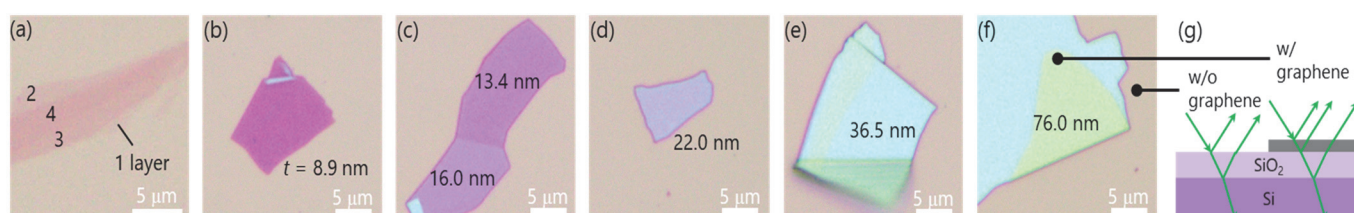


Figure 1 (a) to (f) Optical images of multilayer graphene of various thicknesses¹⁰⁾.

Numbers in figures indicate (a) number of graphene layers; (b)–(f) thickness measured by AFM.

(g) Schematic diagram showing difference in optical path of light incident on substrate with and without graphene.

On the other hand, the change in the color of graphene at thicknesses exceeding about ten layers remains understood only qualitatively. Specifically, on a Si substrate with the widely used 300-nm-thick SiO₂ surface layer, the color of graphene changes from reddish tones to bluish tones as the thickness increases, and eventually to metallic gray, the color of graphite. Thus, the fact that monolayer graphene, whose thickness is only 0.334 nm, can be observed under an optical microscope and moreover exhibits various colors depending on its thickness is not only practically useful for layer-number determination, but is also intrinsically fascinating.

According to reflectance spectrum measurements that yield a more fundamental physical quantity than RGB intensities, the spectra up to nine layers are accounted for by the Fresnel equations, and the number of layers can be determined from the reflectance at a wavelength of approximately 550 nm⁶⁾. In addition, Raman spectroscopy is also a known method for determining the number of layers by making use of the optical properties of graphene. In the very thin regime of approximately one to three layers, the number of layers can be determined with good accuracy from the intensity ratio of the 2D band, which arises from a second-order lattice vibrational process, to the G-band, which originates from in-plane vibrations of the carbon atom honeycomb lattice, as they appear in the Raman scattering spectrum. However, this method for determining the number of layers is not well suited to thicker regimes. The layer-number dependence of the G-band intensity has been investigated up to a thickness of about 30 nm (about 90 layers)⁷⁻⁹⁾, and it is known that the intensity first increases with thickness, then turns to decrease, and ultimately converges to a constant value. However, the thickness at which the intensity transitions from increasing to decreasing differs among research groups and has yet to be quantified.

While the layer-number dependence of optical properties such as RGB contrast, reflectance, and Raman scattering of graphene has been investigated, much of this understanding remains at a qualitative level, and each study has been carried out independently by different research groups. Consequently, there is still no unified understanding that includes the correlations between these individual properties. We therefore measured RGB intensities, reflectance, and Raman scattering spectra on the same samples of multilayer

graphene ranging in thickness from monolayer up to 107 nm (approx. 319 layers), along with an investigation of the relationship between thickness and color. This enabled direct comparison of the thickness dependence of each property, permitting a comprehensive study that also encompasses the relationships between them¹⁰⁾. This article presents the findings of this work.

This study used graphene prepared by exfoliating graphite onto a Si substrate with a 268-nm-thick SiO₂ surface layer. The thickness of the SiO₂ was determined from reflectance spectra as described later. The thickness (t) of the graphene was determined by AFM (Dimension XR, Bruker) measurements. The imaging conditions of optical microscopy (BX53M, Olympus Corporation) such as exposure time and aperture were fixed across all measurements unless otherwise noted. A spectroscopic reflectometer (TohoSpec3100, Toho Technology Corporation) was used for reflectance measurements. A laser Raman microscope (inVia Raman microscope, Renishaw plc) with an excitation wavelength of 532 nm was used for Raman spectroscopy. This article is structured as follows: Section 1 overviews the intriguing fact that graphene, a material of atomic-scale thinness, can be observed under an optical microscope, along with the current state of understanding regarding the layer-number dependence of the optical properties of graphene. Section 2 presents our research into the optical properties of multilayer graphene. Section 2.1 shows the results for RGB contrast, Section 2.2 for reflectance, and Section 2.3 for Raman scattering measurements, while Section 2.4 discusses the relationships between these properties.

2 Optical Properties of Multilayer Graphene

2.1 RGB Contrast

As illustrated by the optical images in Figures 1(a)–(f), graphene displays a rich variety of colors depending on its thickness. This is explained, as described later, by the reflection of light at each interface shown in Figure 1(g) and the resulting interference effects¹⁰⁾. The numerical values in the figures represent the number of graphene layers in Figure 1(a), and the thickness measured by AFM in Figures 1(b)–(f). Figure 2(a) shows the 256-level RGB intensities ($X_{R,G,B}$) extracted from the optical image of the one-to-four-layer graphene in Figure 1(a) against the number of graphene layers. The inset shows a histogram of X_G . From the inset, it can be seen that X_G takes

discrete values corresponding to the number of graphene layers, demonstrating that single-atomic-layer resolution is achievable. As shown in Figure 2(a), $X_{R,G,B}$ varies linearly with the number of layers, demonstrating that the number of graphene layers can be determined from $X_{R,G,B}$.

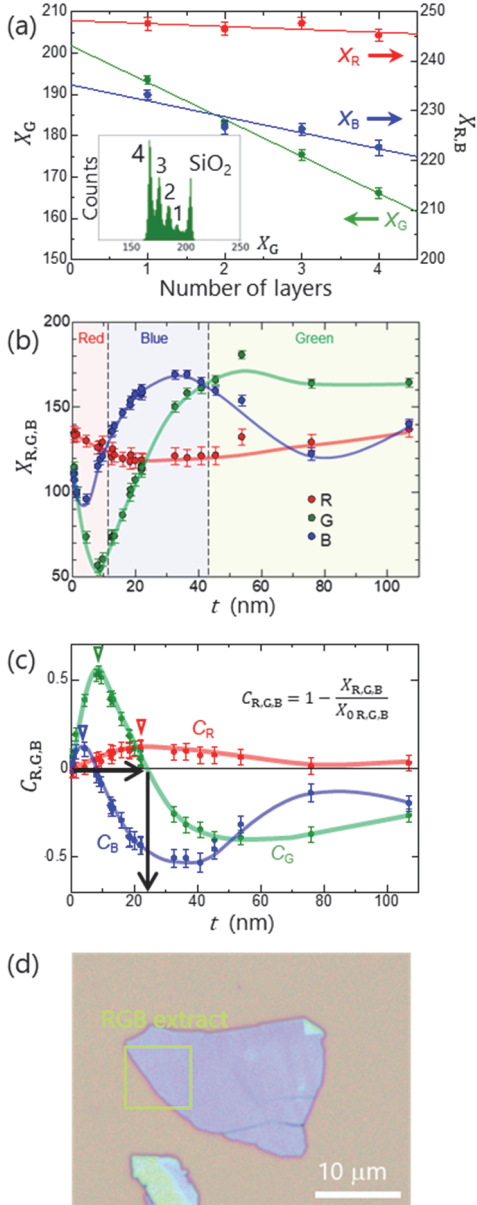


Figure 2 (a) RGB intensities $X_{R,G,B}$ of the one-to-four-layer graphene shown in Figure 1(a)¹⁰. The inset shows a histogram of X_G . (b) $X_{R,G,B}$ of multilayer graphene extracted from optical images¹⁰. (c) Optical contrast $C_{R,G,B}$ ¹⁰. (d) Optical image of multilayer graphene. $C_{R,G,B}$ was obtained from the region enclosed by the green frame, and the thickness was estimated from (c). Note that the $X_{R,G,B}$ values in Figure 2(a) were obtained from optical images acquired with a longer exposure time than the measurements described later in order to improve resolution in the very thin regime, and therefore the numerical values do not correspond to those in Figure 2(b).

Figure 2(b) shows $X_{R,G,B}$ plotted over the entire range of measured thickness t . As t increases, the dominant component shifts from red (X_R) to blue (X_B) at $t \sim 10$ nm, and then to green (X_G) at $t \sim 40$ nm. This study is the first to quantitatively evaluate the color changes of graphene, which had previously been understood only at a qualitative level.

The fact that graphene, an atomic-scale material, can be observed under an optical microscope and moreover displays a variety of colors depending on its thickness is explained by light interference effects. As shown schematically in Figure 1(g), incident light is reflected at each of the interfaces between air, graphene, SiO_2 , and Si, and these reflected lights are emitted externally. In practice, due to multiple reflections, there are countless optical paths of greater complexity than those depicted in the figure, and the light traveling along these paths interferes with one another. Whether the interference results in constructive or destructive superposition is determined by the ratio of the optical path difference to the wavelength. As a result, the intensity of the interfering light varies periodically with t , and this period becomes longer for light of longer wavelengths. In this way, since the t -dependence differs according to the wavelength of light, the color of graphene changes with increasing thickness.

However, there is individual variation of about ± 20 nm even among products with the same nominal SiO_2 thickness, and this variation changes the optical path length of light reflected at the SiO_2/Si interface, meaning that $X_{R,G,B}$ values differ from substrate to substrate even for graphene of the same thickness. While all data in Figure 2(b) were measured on the same substrate, the intensity ratio relative to the substrate—namely the contrast $C_{R,G,B}$ —is less susceptible to differences in SiO_2 thickness than the absolute values of $X_{R,G,B}$ and is therefore more practical as an index for the unified evaluation of graphene on different substrates.

In this study, the RGB contrast $C_{R,G,B}$ between graphene and the substrate was defined by the following equation:

$$C_{R,G,B} = 1 - \frac{X_{R,G,B}}{X_{0,R,G,B}}, \quad (1)$$

where $X_{R,G,B}$ is the RGB intensity of the graphene and $X_{0,R,G,B}$ is the RGB intensity of the substrate. Figure 2(c) shows the t -dependence of $C_{R,G,B}$ obtained from the $X_{R,G,B}$ values in Figure 2(b). It shows that $C_{R,G,B}$ increases with increasing t , but begins to decrease beyond a certain t , and the variation becomes small for $t > 40$ nm. The thickness at which $C_{R,G,B}$

transitions from increasing to decreasing is indicated by open triangles in Figure 2(c), and takes a different value for each color. This difference can be understood from the relationship between light interference and wavelength discussed earlier. That is, since the wavelengths of the RGB colors are ordered $B < G < R$ from shortest to longest, the maxima appear in the order $C_B < C_G < C_R$ from smaller to larger values of t .

Using Figure 2(c), we examined whether the layer-number determination method based on $C_{R,G,B}$ can be applied to regimes thicker than the six layers demonstrated in the prior study⁵⁾. Values of t can be determined uniquely by combining the values of C_R , C_G , and C_B extracted from optical images. Since the determination accuracy depends on the slope of the variation of $C_{R,G,B}$ with respect to t , it is effectively determined by C_G , which has the steepest slope. As a result, an accuracy of approximately ± 1 nm (roughly three atomic layers) can be achieved from this slope and the uncertainty in C_G up to about $t = 40$ nm. For example, the multilayer graphene shown in the frame in Figure 2(d) has $C_G = -0.01$, from which t is estimated to be 24 ± 1 nm. This value is in good agreement with the value measured by AFM of $t = 23.6$ nm.

2.2 Reflectance

As we have seen, $C_{R,G,B}$ is a very practical index that estimates layer number with high accuracy, despite being readily obtainable from an optical microscope. However, $C_{R,G,B}$ is not a direct physical quantity, because it represents the reflected light intensity multiplied by the spectral transmittance of the lenses and filters as well as the spectral sensitivity characteristics of the color sensor. It is also heavily dependent on the imaging conditions. Therefore, to evaluate the physical properties in detail, the t -dependence of the reflectance (R) across the visible light region was measured.

Figure 3(a) shows the R spectra of multilayer graphene ranging from monolayer up to $t = 107$ nm. The thick black line represents the R of the SiO_2/Si substrate. This figure shows that R varies substantially with wavelength particularly for thin graphene, but as the thickness increases this variation becomes smaller, ultimately converging to a value almost independent of wavelength.

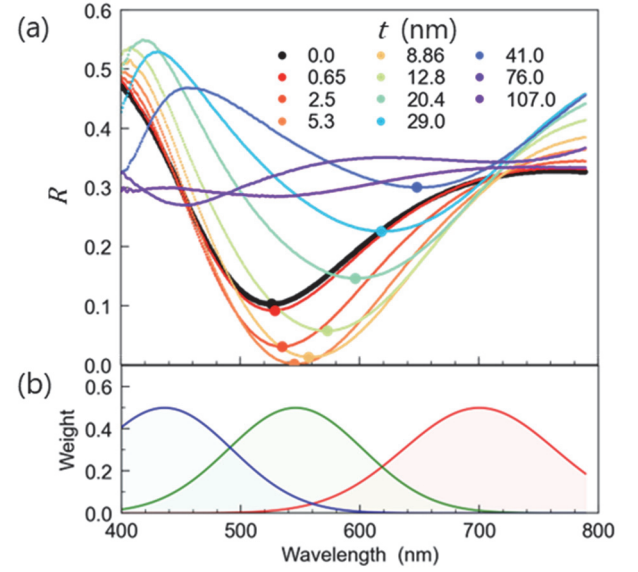


Figure 3 (a) Reflectance R of multilayer graphene of various thicknesses in the visible light region¹⁰⁾. Circles indicate local minimum values of R at each t . (b) Weighting functions for R, G, and B used in calculating virtual $X'_{R,G,B}$ from $R^{(0)}$.

As indicated by the circles in Figure 3(a) representing the local minimum values of R at each t , $R = 0$ at $t \sim 5$ nm and wavelength $\lambda \sim 550$ nm. This indicates that when monochromatic light of a specific wavelength ($\lambda \sim 550$ nm) is incident on graphene of a certain thickness ($t \sim 5$ nm), the graphene appears completely black and is effectively “invisible.” At this point, however, the contrast with the substrate is at its maximum, so graphene remains visible through its contrast with the substrate. In general, reflected light contains, in addition to light reflected at the material surface, components such as emitted light from each graphene layer including Raman scattered light¹¹⁾. However, this zero reflectance demonstrates that the influence of such emitted light components is negligibly small relative to the magnitude of R .

Next, virtual RGB intensities $X'_{R,G,B}$ were calculated from the experimentally obtained R , and compared with $X_{R,G,B}$ extracted from the optical images. Specifically, $X'_{R,G,B}$ were computed by integrating R over the wavelength range corresponding to each color, weighted by a Gaussian distribution as shown in Figure 3(b). The center wavelengths and full widths at half maximum (FWHM) of the Gaussian distributions in Figure 3(b) were set to 700 nm and 90 nm for R, 546.1 nm and 80 nm for G, and 435.8 nm and 80 nm for B. Note that, since the measured wavelength range of R is 400 to 800 nm, wavelength components outside this range are

discarded. Figure 4 shows $X_{R,G,B}$ and $X'_{R,G,B}$ calculated from R . Since the actual spectral sensitivity of each color sensor is more complex than a simple Gaussian distribution, the measured and calculated values do not agree perfectly. However, it is clear that $X_{R,G,B}$ is qualitatively reproduced by $X'_{R,G,B}$.

Subsequently, the experimentally obtained R was compared with theoretical calculations based on the Fresnel equations. For simplicity, as shown in Figure 5(a), we used a model in which multilayer graphene is assumed to be a homogeneous medium, ignoring its layered structure, and R was calculated from the following equations.

$$R = \frac{[r_1 e^{i(\varphi_1 + \varphi_2)} + r_2 e^{-i(\varphi_1 - \varphi_2)} + r_3 e^{-i(\varphi_1 + \varphi_2)} + r_1 r_2 r_3 e^{i(\varphi_1 - \varphi_2)}]^2}{e^{i(\varphi_1 + \varphi_2)} + r_1 r_2 e^{-i(\varphi_1 - \varphi_2)} + r_2 r_3 e^{-i(\varphi_1 + \varphi_2)} + r_1 r_3 e^{i(\varphi_1 - \varphi_2)}} \quad (2)$$

$$r_1 = \frac{n_0 - n_1}{n_0 + n_1}, \quad r_2 = \frac{n_1 - n_2}{n_1 + n_2}, \quad r_3 = \frac{n_2 - n_3}{n_2 + n_3} \quad (3)$$

$$\varphi_1 = \frac{2\pi n_1 t}{\lambda}, \quad \varphi_2 = \frac{2\pi n_2 d_{\text{SiO}_2}}{\lambda}, \quad (4)$$

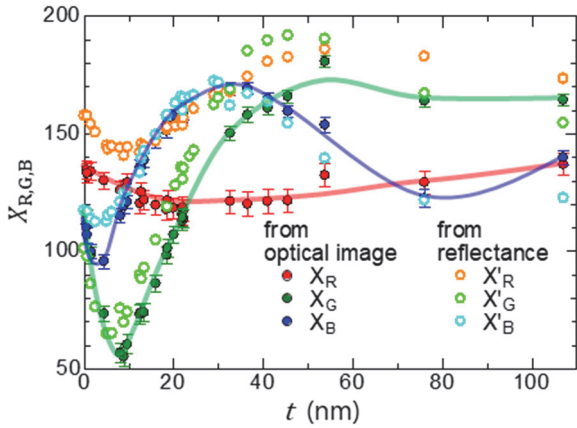


Figure 4 Comparison between $X_{R,G,B}$ extracted from optical images and $X'_{R,G,B}$ calculated from reflectance $R^{(0)}$. Solid lines are guides for the eyes.

Here, n_i ($i = 0, 1, 2, 3$) are the complex refractive indices of air, graphene, SiO_2 , and Si, respectively, λ is the wavelength of light, and d_{SiO_2} is the thickness of SiO_2 . In this calculation, the refractive index of air was set to $n_0 = 1$, the refractive index of multilayer graphene n_1 was taken from the literature value for monolayer graphene¹²⁾, and the refractive indices of SiO_2 and Si, n_2 and n_3 , were taken from literature values^{13),14)}. n_1 , n_2 , and n_3 are wavelength-dependent and cannot be expressed by simple formulas. For reference, $n_1 = 2.64 - 1.38i$, $n_2 = 1.46$, and $n_3 = 4.07 - 0.03i$ at $\lambda = 550$ nm. d_{SiO_2} was determined to be 268 ± 1 nm by fitting the

experimentally measured R of the SiO_2/Si substrate shown in Figure 5(b).

Figures 5(c)–(f) compare the experimentally measured R and the theoretical calculations based on the Fresnel equations for multilayer graphene of representative thicknesses. The measured values and calculated results are in very good agreement across the entire range of measured thicknesses and wavelengths. These results indicate that, for the R in the visible light region, multilayer graphene can be treated as a homogeneous medium with the layered structure ignored, and that evaluation of multilayer graphene using the refractive index of monolayer graphene remains valid for thicknesses up to 107 nm. This is also consistent with the result that the influence of emitted light, such as Raman scattering, on R can be ignored.

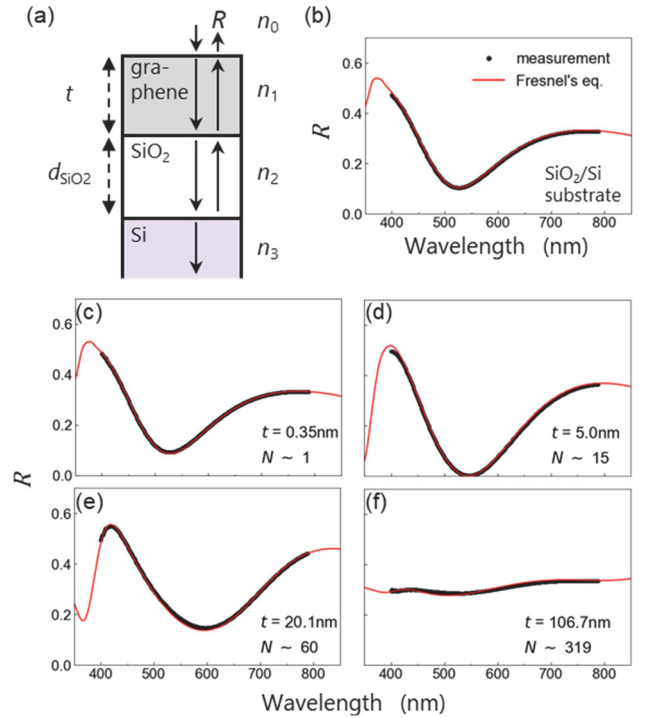


Figure 5 (a) Schematic diagram of model used for theoretical calculation of reflectance R . Comparison between experimentally measured R and calculated results based on Fresnel's equations for (b) SiO_2/Si substrate and (c)–(f) multilayer graphene of various thicknesses¹⁰⁾. Black circles are experimentally measured values obtained at a wavelength resolution of 1 nm, and red lines are calculated values.

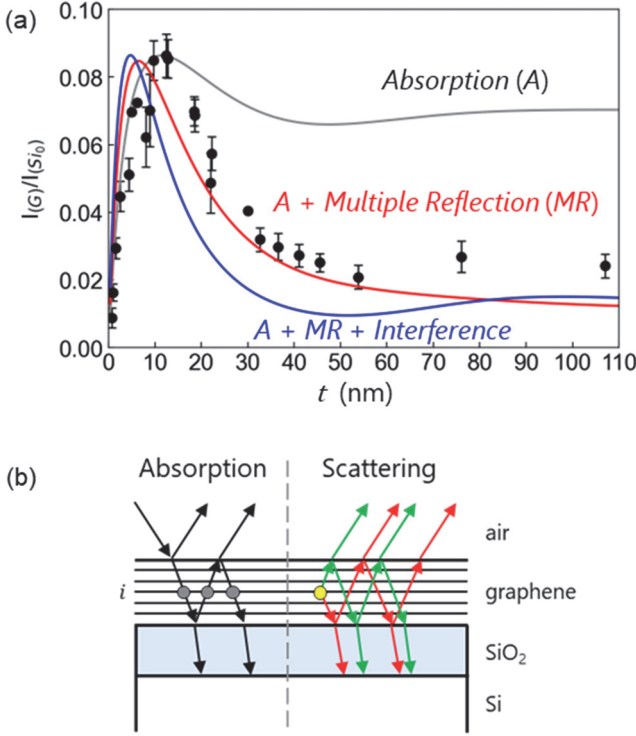


Figure 6 (a) Graphene thickness t dependence of G-band intensity ratio I_G/I_{Si0} ¹⁰. Black circles represent measured values, and solid lines show theoretical calculation results normalized in height direction. The gray line shows the result considering only graphene Absorption (A); the red line adds Multiple Reflection (MR) of the Raman scattered light to that; and the blue line further adds Interference of the Raman scattered light. (b) Schematic diagram of optical paths for light absorbed (left side of dashed line) and scattered (right side of dashed line) in graphene.

2.3 Raman Scattering Intensity

So far, we have focused on the reflected light component of the incident light on graphene, but there are also transmitted light and components absorbed by each graphene layer. A portion of the absorbed light is emitted as Raman scattered light. In this section, we evaluated the light absorption by multilayer graphene using Raman spectroscopy measurements. Among three major peaks known in the Raman scattering spectrum of graphene, namely the D, G, and 2D bands in order of increasing wavenumber, we focused on the layer number dependence of the G-band intensity (I_G), because its peak structure does not change with the number of layers. In Raman spectroscopy measurements, variation in signal intensity arises due to fluctuations in laser output between measurements. For this reason, each measurement was carried out by mapping a region containing both the graphene flake and the

SiO₂ surface, and evaluated using the intensity (I_G/I_{Si0}) normalized by the scattering peak intensity (I_{Si0}) originating from the Si beneath the SiO₂, as measured at the SiO₂ surface.

Figure 6(a) shows the dependence of I_G/I_{Si0} on t . As t increases, I_G/I_{Si0} increases with thickness up to $t \sim 12$ nm, but beyond that it begins to decrease, eventually settling at an approximately constant value. This behavior has been reported in previous studies⁷⁻⁹), but by extending measurements to even thicker regions, this study found that at around $t \sim 30$ nm the intensity does not simply converge monotonically to a constant value, but instead undergoes slight fluctuations.

To understand this behavior, we compared experimental results and theoretical calculations based on several models. First, the t -dependence of the Absorption (A) derived from the Fresnel equations was calculated. On the left side of the dashed line in Figure 6(b), the path of incident light into graphene is illustrated by black arrows. A portion of the incident light passes through the graphene multiple times while being partially reflected at each interface, and a portion of the light is absorbed by each graphene layer each time. In the figure, the absorption in the i -th layer counted from the top is indicated by gray circles. The total amount of light absorbed by graphene is calculated as the sum of the absorption in each i -th layer expressed as:

$$A = \sum_{i=1}^N |a_i|^2 \quad (5)$$

$$a_i = \frac{t_1 (e^{-2\pi k_1 z_i / \lambda_0} e^{-i2\pi n_1 z_i / \lambda_0} + r' e^{-2\pi k_1 z'_i / \lambda_0} e^{-i2\pi n_1 z'_i / \lambda_0})}{1 + r_1 r' e^{-i4\pi n_1 t / \lambda_0} e^{-4\pi k_1 t / \lambda_0}} \quad (6)$$

$$t_1 = \frac{2n_0}{n_0 + n_1} \quad (7)$$

$$r' = \frac{r_2 + r_3 e^{-i2\varphi_2}}{1 + r_2 r_3 e^{-i2\varphi_2}} \quad (8)$$

Here, N is the number of graphene layers, k_1 is the extinction coefficient of graphene (the imaginary part of the complex refractive index), $\lambda_0 = 532$ nm is the wavelength of the incident light, $z_i = 0.334 i$ (nm) is the depth of the i -th layer from the surface, and $z'_i = 2t - z_i$. The calculation result of Eq. (5) is shown as the gray line (A) in Figure 6(a). While the behavior of I_G/I_{Si0} increasing with t in the thin region is reproduced, for $t > 12$ nm, I_G/I_{Si0} decreases whereas A remains nearly constant, indicating that absorption alone is not sufficient to explain the results.

Therefore, a model that includes the Raman scattered light was considered. The right side of the dashed line in Figure 6(b) shows the path of Raman scattered light emitted from a given graphene layer. When Raman scattered light is emitted from the yellow circle in the figure, there is both a component directed outward (up in figure), shown by the green arrows, as well as a component directed toward the substrate side (down in figure), shown by the red arrows. Since these light components undergo repeated multiple reflections at each interface, there are countless complex paths, but only components that are ultimately emitted outward from the graphene surface are detected by Raman spectroscopy. Since not all the light scattered in this way is detected, the experimental values converge to a value smaller than $A^{15)}$. When multiple reflections of the Raman scattered light are considered, the G-band intensity can be expressed by the following equations, where b_i denotes the fraction of Raman scattered light emitted from the i -th layer counted from the top that is released to the outside¹⁵⁾.

$$I_G \propto \sum_{i=1}^N |a_i b_i|^2 \quad (9)$$

$$b_i = \frac{t'_1 (e^{-2\pi k_1 z_i / \lambda'} + r'_1 e^{-2\pi k_1 z'_i / \lambda'})}{1 + r'_1 r'_1 e^{-2\pi k_1 t / \lambda'}} \quad (10)$$

$$t'_0 = \frac{2n_1}{n_0 - n_1} \quad (11)$$

Here, λ' is the wavelength of the G-band Raman scattered light corresponding to the excitation wavelength of 532 nm, and is approximately 580.8 nm. This model assumes that the Raman scattered light is emitted with independent phases and that no interference occurs. Accordingly, Eq. (10) uses k_1 alone rather than the complex refractive index n_1 and excludes the phase term. The calculation result of Eq. (9) is shown as the red line ($A+MR$) in Figure 6(a). The behavior of I_G/I_{Si0} transitioning from increasing to decreasing is reproduced by considering the multiple reflections of the Raman scattered light, but the fluctuation of I_G/I_{Si0} observed at $t > 30$ nm is not reproduced. This is attributed to the assumption that interference between Raman scattered light that has traveled via different paths is neglected by using k_1 rather than n_1 in Eq. (10).

Therefore, a calculation incorporating the interference of Raman scattered light was performed. Specifically, k_1 in Eq. (10) was replaced by the complex refractive index n_1 , and the

G-band intensity was calculated from:

$$I_G \propto \sum_{i=1}^N |a_i b'_i|^2 \quad (12)$$

$$b'_i = \frac{t'_1 (e^{-2\pi n_1 z_i / \lambda'} + r'_1 e^{-2\pi n_1 z'_i / \lambda'})}{1 + r'_1 r'_1 e^{-2\pi n_1 t / \lambda'}} \quad (13)$$

This corresponds to considering only the interference between light that is scattered from the same graphene layer and emitted to the outside via different paths. The calculation result is shown as the blue line ($A+MR+Interference$) in Figure 6(a). Both the behavior of I_G/I_{Si0} transitioning to a decrease and the intensity fluctuations in the thicker region are reproduced. This result demonstrates that in order to understand Raman scattering in multilayer graphene, not only multiple reflections of the Raman scattered light but also interference effects are important.

Note that to focus on discussion of the thickness dependence, each calculation result in Figure 6(a) has been normalized so that the maximum values coincide. Further consideration is needed regarding absolute values of scattering intensity.

Finally, in the experimental results, the scatter in I_G/I_{Si0} becomes larger around $t = 10$ nm. A similar feature has been reported in several previous studies^{8),9)}, but the cause has not been clarified yet and is a subject for future investigation.

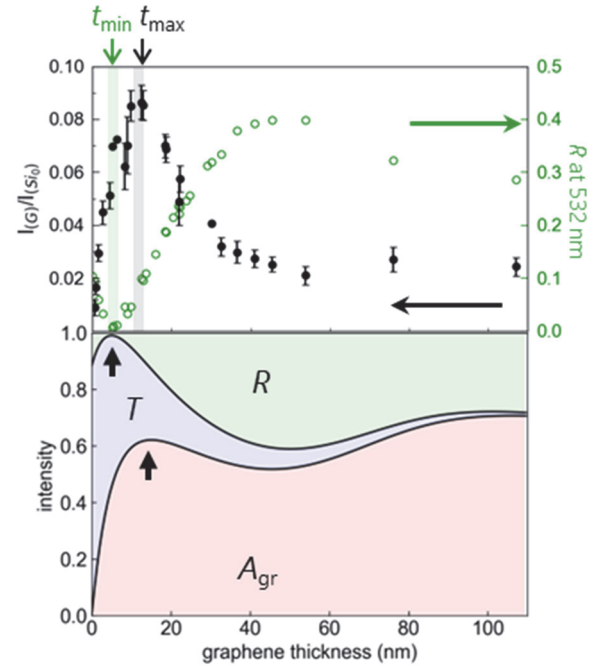


Figure 7 Top panel compares I_G/I_{Si0} and reflectance R at Raman excitation wavelength of 532 nm¹⁰⁾. Bottom panel shows fractions of reflectance R , transmittance T , and absorption in graphene A_{gr} relative to incident light, calculated by Fresnel's equations¹⁰⁾.

2.4 Comparison of Reflectance and Raman Scattering Intensity

Finally, we compared I_G/I_{SiO} , which corresponds quasi-equivalently to absorbed light, with reflectance R for reflected light at 532 nm—the same wavelength used in the Raman spectroscopy. The t -dependence of both quantities is shown in the top panel of Figure 7. As t increases, I_G/I_{SiO} first rises and then begins to decrease from $t_{\max} \sim 12$ nm, while R first decreases and then begins to increase from $t_{\min} \sim 5$ nm. Intuitively, it might be expected that when R is small, the amount of absorption increases and therefore I_G/I_{SiO} increases; but while the experimental values qualitatively follow this trend, the observed $t_{\min}$ and $t_{\max}$ are not equal. This is thought to be because not all the non-reflected light is absorbed by graphene. Instead, there is a component T that transmits through the graphene and penetrates into the SiO₂/Si substrate. The bottom panel of Figure 7 shows the t -dependence of the reflectance R , transmittance T , and the absorption into graphene A_{gr} obtained from $A_{gr} = 1 - R - T$ of multilayer graphene calculated using the Fresnel equations. T was calculated using the following equation.

$$T = \left[\frac{t_1 t_2 t_3}{e^{i(\varphi_1 + \varphi_2)} + r_1 r_2 e^{-i(\varphi_1 - \varphi_2)} + r_2 r_3 e^{-i(\varphi_1 + \varphi_2)} + r_1 r_3 e^{i(\varphi_1 - \varphi_2)}} \right]^2 \quad (14)$$

$$t_2 = \frac{2n_1}{n_1 + n_2}, \quad t_3 = \frac{2n_2}{n_2 + n_3} \quad (15)$$

Here, from Eqs. (7) and (15), the denominator of Eq. (14) is a constant, and since φ_1 is proportional to t from Eq. (4), T decreases approximately exponentially with respect to t . The presence of this component explains why the t value at which R reaches its minimum and the t value at which A_{gr} reaches its maximum are offset from each other. This offset is understood as the observed difference between $t_{\min}$ and $t_{\max}$.

3 Conclusions

Focusing on the fact that the color of multilayer graphene changes significantly with thickness, we measured the optical properties, including optical contrast, reflectance, and Raman scattering, of multilayer graphene ranging in thickness from monolayer up to 107 nm. Furthermore, we conducted a comprehensive study that also includes the correlations between each of these properties by performing these measurements on the same sample.

First, we quantitatively evaluated for the first time the change in color of multilayer graphene with thickness using the RGB contrast of optical images that had until now only been known qualitatively. We also demonstrated that the simple method of determining the number of graphene layers from RGB contrast remains valid with an accuracy of approximately ± 1 nm (three atomic layers) even for thick graphene up to $t \sim 40$ nm.

Second, we measured reflectance as a more fundamental physical quantity than RGB, and showed that the spectra can be almost completely reproduced by the Fresnel equations for thicknesses ranging from a monolayer up to 107 nm (about 319 atomic layers). This means that multilayer graphene can be treated as a uniform medium, ignoring its layered structure, from the perspective of reflectance. Furthermore, we demonstrated that the refractive index of monolayer graphene remains applicable up to a thickness of 107 nm and clarified that the reflectance of graphene of a specific thickness can reach zero for incident light of a specific wavelength. These results suggest that light emitted from each graphene layer, such as Raman scattered light, is negligible compared to the reflected light intensity.

Raman spectroscopy was performed to investigate light absorption by graphene. We observed for the first time that, in addition to the Raman scattering intensity first increasing and then transitioning to a decrease with increasing graphene thickness, the intensity also fluctuates at thicknesses of 30 nm and above. This indicates that not only multiple reflections but also interference effects are important for Raman scattering.

Lastly, we directly compared the thickness dependence of these optical properties and clarified the relationships between them. Although the thickness at which reflectance reaches its minimum does not match the thickness at which Raman scattering intensity reaches its maximum, this discrepancy can be understood by considering the respective contributions of reflection, transmission, and absorption.

This study demonstrated a practical method for determining the number of layers in multilayer graphene, while also clarifying its fundamental optical properties. These findings are expected to contribute to future applied research aimed at fabricating multilayer graphene optical devices.

Acknowledgements

This research was conducted as a collaborative study with Dr. Ken-ichi Sasaki of NTT Basic Research Laboratories. We would like to take this opportunity to express our sincere gratitude for his help. The reflectance measurements of multilayer graphene were supported by “Nanotechnology Platform Japan” of the Ministry of Education, Culture, Sports, Science and Technology (MEXT), Japan Grand Number JPMXP09F22UT1064, and conducted in Takeda Cleanroom with help of Nanofabrication Platform Center of School of Engineering, the University of Tokyo.

References

- 1) R. R. Nair *et al.*, Science **320**, 1308 (2008).
- 2) M. Shimatani *et al.*, Appl. Phys. Express. **12**, 122010 (2019).
- 3) K. Sasaki and K. Hitachi, Commun. Phys. **3**, 90 (2020).
- 4) P. Blake *et al.*, Appl. Phys. Lett **91**, 063124 (2009).
- 5) B. Ma *et al.*, Carbon **109**, 384 (2016).
- 6) Z. H. Ni *et al.*, Nano Lett. **7**, 2758 (2007).
- 7) D. Yoon *et al.*, J. Korean Phys. Soc. **55**, 1299 (2009).
- 8) X.-L. Li *et al.*, Nanoscale **7**, 8135 (2015).
- 9) Y.-S. No *et al.*, Scientific Report **8**, 571 (2018).
- 10) M. Kamada *et al.*, Carbon Trends **16**, 100389 (2024).
- 11) K. Sasaki *et al.*, C **10**, 18 (2024).
- 12) M. A. El-Sayed *et al.*, Nanomaterials **11**, 1230 (2021).
- 13) I. H. Malitson, J. Opt. Soc. Am. **55**, 1205 (1965).
- 14) D. E. Aspnes and A. A. Studna, Phys. Rev. B **27**, 985 (1983).
- 15) Y. Y. Wang *et al.*, Appl. Phys. Lett. **92**, 043121 (2008).

Authors



Masahiro Kamada
MATSUI Laboratory
Advanced Research Laboratory



Tomohiro Matsui
MATSUI Laboratory
Advanced Research Laboratory

Publicly available

Trends and Overview of Evaluation Techniques in Image Recognition AI

Ryu Kidokoro, Lawrence Earl Carandang

[Summary]

With the rapid evolution of artificial intelligence (AI) and its expanding application in fields where accuracy is critical, such as healthcare and quality inspection, comprehensive evaluation of AI models and their development processes is essential. This article overviews practical methods for evaluating image recognition AI and demonstrates the role of eXplainable AI (XAI) within an evaluation framework that encompasses not only accuracy but also interpretability and reliability. We demonstrate the need for evaluation at the pipeline level and introduce application of representative XAI methods, including Local Interpretable Model-Agnostic Explanations (LIME) and Class Activation Maps (CAMs), to inference interpretation and model improvement. We also discuss robustness evaluation through uncertainty quantification using Test-Time Augmentation and Bayesian networks. Furthermore, we discuss the outlook for challenges related to applying these evaluation methods to diverse tasks and real-world data, as well as integration of evaluation metrics with development objectives.

1 Introduction

Recent advances in deep learning have led to rapid proliferation of AI across a wide range of fields, including business and healthcare. In particular, generative AI and large language models (LLMs) have become central to corporate IT strategies. At the same time, the opacity of AI decision-making poses a significant challenge for ensuring reliability, making it urgent to establish evaluation techniques that guarantee fairness and accountability. Against this backdrop, the importance of a comprehensive evaluation framework centered on eXplainable AI (XAI) is growing.

XAI has drawn attention as a technology that resolves the black-box problem of AI and makes the basis for its decisions understandable. In critical sectors such as healthcare and finance, visualizing the rationale behind AI decisions is indispensable from both a regulatory and ethical standpoint. However, evaluation of current AI models tends to focus on accuracy and speed, making it difficult to identify the causes of misjudgments and bias. XAI methods such as LIME (Local Interpretable Model-agnostic Explanations)¹⁾ and SHAP (SHapley Additive exPlanations)²⁾ have enabled contribution analysis, but the evaluation of explanation fidelity and stability remains in its early development stages. Trustworthy AI requires not only performance evaluation but also comprehensive assessment techniques that integrate explanation quality and uncertainty.

This article introduces novel evaluation metrics and

methods for quantifying interpretability, fairness, and explanation reliability through XAI.

The Microsoft COCO dataset (© 2014 Microsoft. Licensed under the Creative Commons Attribution 4.0 International License.) was used as training data for the detection model³⁾.

2 AI Model Evaluation

Building on the AI reliability challenges outlined in Section 1, this section introduces evaluation techniques for AI models.

There are various quantitative and qualitative perspectives in evaluation; for example, accuracy quantitatively assesses model performance on tasks such as classification. Training time indicates how quickly an AI model can be created and is an important factor in time-constrained operations. Furthermore, the validity of the decision rationale is a qualitative perspective that reflects the reliability and explainability of the results produced by an AI. These perspectives are indispensable for making a comprehensive assessment of AI performance.

2.1 AI Development Pipeline Evaluation

While Section 2 addresses the evaluation of “AI models,” the evaluation scope is not limited to the model itself. This section introduces the “AI development pipeline” as an additional subject of evaluation. A development pipeline (or simply an “AI pipeline”) refers to the entire sequence of steps from AI model development to just before deployment, and consists of four components: developers; datasets; AI models;

and inference results, along with six processes: data collection; pre- and post-processing; training; inference; and performance calculation. Figure 1 shows how the components and processes are interconnected.

A reactive development approach is common in traditional machine learning where a model is built first, its accuracy is then evaluated, and improvements are made based on the results. However, this approach can lead to unnecessary data collection and preprocessing, especially when feature selection during the data preparation stage is insufficient, resulting in reduced development efficiency. To address this challenge, it is important to identify features and noisy data that either do not contribute to or negatively impact performance through model evaluation. This enables optimization of data design, including feature selection, and makes it possible to build high-accuracy models even with limited data.

In this way, model evaluation serves not merely as a performance check, but as an important guide for data design and model improvement.

2.2 Explainable AI (XAI)

Recent worldwide research has aimed at interpreting black-box AI models to improve reliability and performance. One outcome of such research is eXplainable Artificial Intelligence (XAI). XAI refers to techniques and methods that present the basis for AI decisions and predictions in a form that people can understand. It serves as one approach to AI evaluation by comparing features that AI focuses on with human perspectives. This article introduces several widely used XAI techniques.

2.2.1 Global Explanation and Local Explanation

XAI methods are generally classified into two categories based on their scope of explanation: global explanation and local explanation¹⁾. Global explanation analyzes the overall behavior and dominant internal tendencies of an AI model, evaluating model behavior using multiple test data points. A representative example is the decision tree. Conversely, local explanation aims to understand the reasoning behind individual prediction results and to identify which components of the input contributed most. Unlike global explanation, local explanation evaluates model behavior using only a single test data point. A representative example is the saliency map¹⁾. This article focuses on local explanation and provides a detailed discussion of the saliency map as a representative method in Section 3.1 What Is a Saliency Map?

2.2.2 Model-Specific and Model-Agnostic Methods

XAI methods can also be classified according to whether they have access to a model's internal parameters (such as weights and gradients).

Model-specific methods use internal parameters to analyze outputs between network layers, and can therefore only be applied to specific model architectures. For example, XAI techniques for recurrent neural networks (RNN) cannot be applied to convolutional neural networks (CNN) without adjustment, as their architectures and outputs are different.

On the other hand, model-agnostic methods only analyze changes in input and output, allowing them to be applied to a wide range of models, regardless of their internal structure.

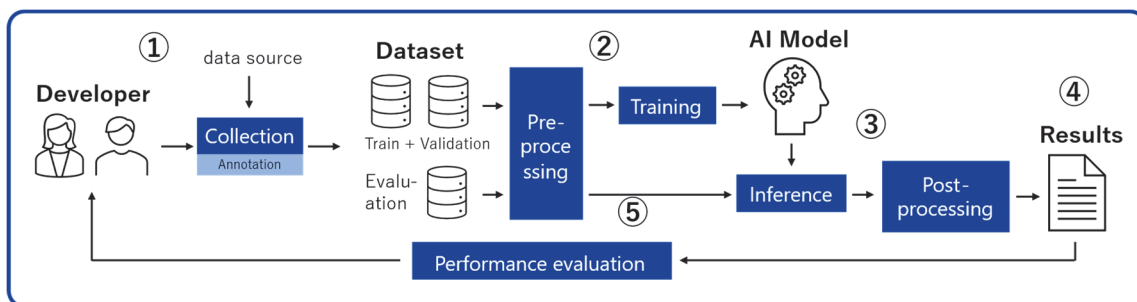


Figure 1 AI Pipeline. ① Developers acquire data from various sources and, where necessary, label (annotate) data to create the dataset. The dataset is then divided into three subsets: training data, validation data, and evaluation/test data. ② Training and validation data are fed through preprocessing to train the AI model (AI training). ③ Evaluation data is fed through preprocessing and is used for inference by model trained in ②. ④ Where necessary, results from ③ are fed through postprocessing to obtain the final results and measure model performance. ⑤ Where necessary, the experiment is repeated with different conditions based on model performance.

¹ Heat maps are often used to visualize saliency maps. In fact, a saliency map is often referred to as a “heat map” rather than a “saliency map.”

Since model-specific methods make use of internal parameters, they enable more fine-grained explanation and evaluation by examining specific parts of the AI. For example, by examining a specific layer of an image recognition neural network, one can verify whether the AI has appropriately learned features such as the shape, color, and size of objects. Model-agnostic methods, on the other hand, offer the advantage of being applicable to any model regardless of internal structure. It is therefore necessary to select the appropriate method depending on whether the model's internal mechanisms and parameters are accessible. Specific quantification methods for uncertainty are discussed in detail in Section 4 Evaluation of Model Robustness Using Uncertainty Calculation.

2.3 Evaluation Metrics

Accuracy is the most important and primary metric for measuring performance, but it is not the sole evaluation criterion. AI models and their development pipelines can also be assessed from the perspectives of transparency, consistency, data quality, and computational efficiency. These evaluation perspectives are worth considering because they are directly or indirectly related to model accuracy.

Interpretation of reasoning logic (transparency) evaluates the possibility that even a correct prediction may be based on erroneous information. A typical example is the case where an image classification model predicts “polar bear” based on snow in the background—a phenomenon called Clever Hans effect⁵⁾. As noted earlier, methods such as XAI can be used to identify the features a model is focusing on and verify whether they align with the features a human would notice. This assesses the effectiveness of the training method while also enhancing user trust.

Quantifying model robustness (consistency) evaluates variability caused by unavoidable factors, such as lighting and angle in input data such as camera images. While humans readily adapt to such changes, AI models' decisions are susceptible to even minor differences in pixel values. A good model must be capable of stable detection regardless of changes in conditions. Robustness is a metric for measuring this stability, and serves as a basis for evaluating under what conditions a model can be considered reliable.

Training Data Attribution (TDA) analyzes the impact of mislabeled or redundant data to evaluate data quality. For example, in an image classification task, if a photo of a cat is

incorrectly labeled as ‘dog’ and the model is trained on that image, it may cause degraded performance. To address such issues, TDA is used to calculate the importance (weight) of each training data point when updating model parameters. Based on this importance, it is possible to identify any data that either does not contribute to or may hinder model training and exclude it from the dataset.

Measurement of a model's computational efficiency (efficiency) evaluates whether the model's size can be optimized while maintaining accuracy. The higher the model's complexity and performance, the larger its file size and the greater the computational resources required to run it. Techniques such as quantization⁶⁾ and distillation⁷⁾ are means of simplifying a model while preserving its accuracy; computational efficiency can then be evaluated by comparing model size before and after applying these methods.

2.4 Application of XAI Techniques to Performance Improvement

In addition to providing developers with ideas for hyperparameter tuning, this article introduces several ways to directly leverage evaluation results as a form of feedback loop within the AI pipeline. For example, during development of an object detection model, a saliency map (discussed in detail in Section 3) can be computed for specific training images and compared against the ground truth (reference data) (Figure 2). This comparison result can be used as an additional loss value for use in backpropagation, serving as an example of loss/feature/gradient augmentation through XAI methods. Saliency maps can also be used to refine inference results. For example, it is possible to apply processing such as lowering the confidence score for images wherein the detected bounding box does not align with the high-saliency region on the saliency map.

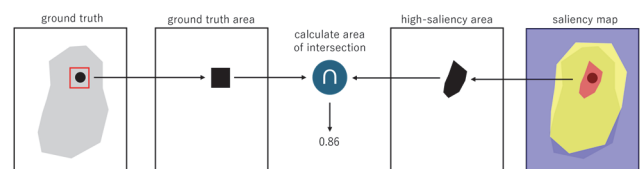


Figure 2 Example of a method for comparing ground truth with a saliency map. The intersection area (Intersection over Union, IoU) is computed between the bounding box area of the ground truth and the high-saliency area extracted by the saliency map.

Meanwhile, as discussed in the previous section, TDA can be applied to dataset design. It is widely known that selection of training data significantly influences model

performance, and dataset design using TDA is a means of quantitatively evaluating the quality and informativeness of training samples. This makes it possible to clearly identify the data that is most beneficial for the model.

3 Interpretation of Reasoning Logic Using Saliency Maps

This section introduces saliency maps for evaluating AI in object detection.

3.1 What Is a Saliency Map?

A saliency map is an image showing the data features that are most attended to during inference. ‘Attended’ here refers to the features that contribute to the model’s output of an inference result. Figure 3 shows a saliency map visualized using Eigen-CAM (Eigenvalue-based Class Activation Mapping)⁸⁾ applied to an image of a dog. As shown, the basis for the ‘dog’ classification was determined from the features at the center of the dog.

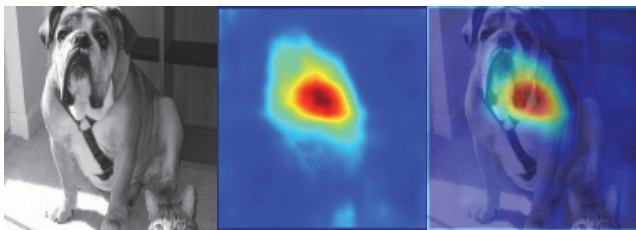


Figure 3 (Left to right) Input image, a saliency map, and input image overlaid with a saliency map

3.2 Saliency Map Generation Techniques

While there are many methods for generating a saliency map, this article introduces three: the model-agnostic LIME, the model-specific Eigen-CAM, and D-CLOSE (Detector-Cascading multiple Levels of Segments to Explain)⁴⁾.

3.2.1 LIME

LIME is a method that generates a set of samples by randomly perturbing parts of the input data and obtaining the model’s confidence scores for those samples, thereby constructing a local explanation model. Specifically, it performs linear regression using the perturbed data features as explanatory variables and the model’s predicted confidence scores as the objective variable in order to estimate the importance of each feature. If the perturbed region is important for the prediction, the confidence score will decrease; otherwise, we can expect that the confidence score will be the same. For object detection models, a method that masks the target region by replacing all pixel values with 0 can be

adopted as the perturbation when generating images visualized with LIME. An example of the result is shown in Figure 4(b), which suggests that the model determines the ‘dog’ object based on features such as the face, head, and legs.

3.2.2 CAM/Grad-CAM/Eigen-CAM

CAM (Class Activation Mapping) is a technique that leverages the structure of CNNs (Convolutional Neural Networks)⁹⁾ to visualize which regions in an image contributed to the classification. The algorithm generates a map (activation map) that highlights important regions in the image by multiplying the output of the final convolutional layer of the CNN by the weights of the classification layer.

A weakness of CAM is that it is constrained by model architecture. Specifically, it can only be applied to CNNs that have a classification layer that uses Global Average Pooling (GAP), which limits its flexible application to many existing models. This led to the proposal of Grad-CAM (Gradient-weighted Class Activation Mapping). Grad-CAM is a method that uses gradients with respect to the model’s output to weigh the feature map of the final convolutional layer, thereby visualizing the regions of attention within the image¹⁰⁾. This makes it applicable to a wider range of CNN-based models without dependence on a specific model architecture.

For tasks such as object detection, Eigen-CAM is particularly effective. It is a method that performs principal component analysis (PCA) or singular value decomposition (SVD) on a feature map to extract and visualize eigenvectors. This enables extraction of regions to which the model pays attention in a class-independent manner.

In addition to Eigen-CAM, there have been various derivatives and improvements proposed for CAM, such as Cluster-CAM¹¹⁾ and Expected Grad-CAM¹²⁾. These methods aim to improve processing speed and accuracy by clustering feature maps and refining gradient calculations. CAM-based methods are an important means of presenting the basis for a model’s decisions in a form that people can understand, serving as a technology within eXplainable AI (XAI) that continues to attract significant attention.

Figure 4(c) shows the image visualized by Eigen-CAM, showing that the model determines the ‘dog’ object based on features around the face.

3.2.3 D-CLOSE

The D-CLOSE method achieves visualization by

repeatedly obtaining model predictions for a target image using multiple perturbation regions of different sizes, and then integrating those results. While LIME applies perturbations at a single segmentation granularity and generates explanations using a local linear model, D-CLOSE applies perturbations at multiple scales ranging from coarse to fine segmentation, making it possible to simultaneously capture the influence of both broad and detailed features. Figure 4(d) shows the image visualized by D-CLOSE, suggesting that the model determines the ‘dog’ object based on the face and head features.

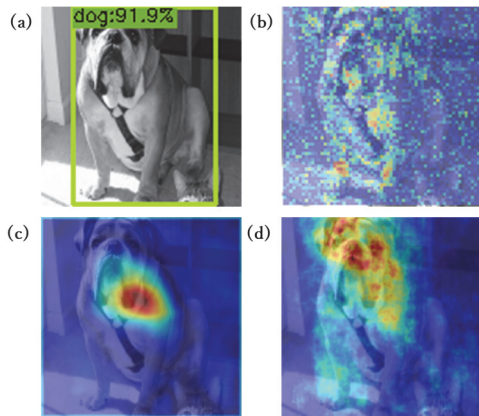


Figure 4 (a) Input image with inference result (b) Saliency map generated by LIME (c) Saliency map generated by Eigen-CAM (d) Saliency map generated by D-CLOSE

4 Evaluation of Model Robustness Using Uncertainty Calculation

This section introduces a method for quantifying uncertainty, a metric that indicates the reliability of model predictions.

Unlike humans, AI models cannot answer ‘I don’t know’ when presented with inputs they have not been trained on, and will always return some prediction. For example, if a lion image is given to an image classification model that has only been trained on images of cats and dogs, there is a high possibility that it will misclassify the image as ‘cat’ with high confidence, despite never having been trained on lions.

One metric for evaluating the reliability of predictions is ‘uncertainty.’ In object detection, while confidence indicates the probability of an object’s existence, uncertainty represents the reliability of the prediction itself. Following the earlier example, an inference without a proper basis (i.e., a prediction of ‘cat’ for a lion image) may have high uncertainty, even if the model’s confidence score is high.

The simplest method for measuring uncertainty is to have the model make multiple predictions on the same input under conditions that do not alter the essential nature of the input data. For example, uncertainty can be quantified by observing the variance in predictions obtained by modifying the input image’s color or size. Common metrics for uncertainty include confidence intervals, range of variation, variance, and standard deviation.

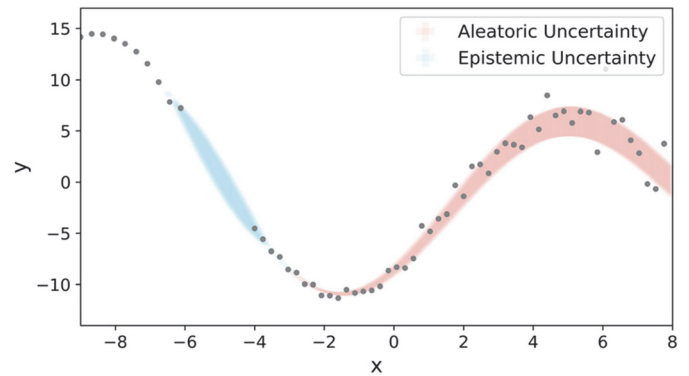


Figure 5 Epistemic uncertainty and aleatoric uncertainty¹³⁾. Due to insufficient data, the behavior between $x=-6$ and $x=-4$ cannot be modeled, giving rise to epistemic uncertainty. Meanwhile, due to variance in $x=0$ to $x=8$, the actual behavior becomes unknown. As a result, the model’s prediction becomes a range rather than a fixed value, giving rise to aleatoric uncertainty.

Uncertainty can be broadly divided into two types: epistemic uncertainty and aleatoric uncertainty¹⁴⁾. The former arises from data outside the training distribution or from inappropriate model selection, and can be mitigated by adding data or improving the model. The latter stems from noise and randomness inherent in the environment, and is difficult to mitigate unless the environment itself is changed. Figure 5 visualizes both types in the context of a regression task.

As mentioned previously, a common method for quantifying uncertainty is to obtain multiple inference results for the same test input. The following sections introduce several methods classified according to whether or not modification of the network’s internals is required (model-specific or model-agnostic).

4.1 Model-Agnostic Methods

A common method for estimating uncertainty is to train multiple models on the same dataset with different initial values, hyperparameters, or architectures—a technique known as Deep Ensembles¹⁵⁾. By varying the training conditions, each model follows a different training trajectory, and

if the models converge to different points, their inference results may also differ. Assuming high-quality training data was used, high uncertainty with respect to unknown data indicates that the model has failed to correctly capture the important features—that is, it reflects epistemic uncertainty.

A related technique is bootstrap aggregating (bagging), which involves training multiple models using different subsets drawn by sampling with replacement from the training data (bootstrap sets), without changing the model architecture¹⁶⁾.

However, creating multiple models is not always practical. As a single-model approach, Test-Time Augmentation (TTA) allows evaluation of an existing model without requiring any additional training processes¹⁷⁾. In TTA, multiple transformations, such as affine transformations (e.g., translation and rotation) or recoloring operations (e.g., contrast enhancement), are applied to the input being evaluated, and each transformed version is passed through the model to obtain multiple inference results. Figure 6 shows several examples of such transformations.

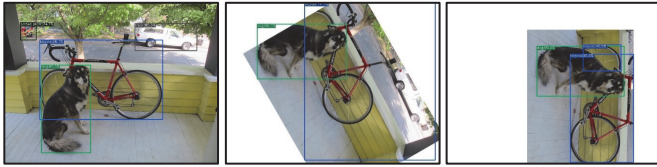


Figure 6 (Left to right) Original image, Transformed image #1, Transformed image #2

By exploring a range of transformation patterns and recording detection confidence with respect to the transformation, the degree of model uncertainty can be visualized as a graph. This makes it possible to pinpoint the model's weaknesses and appropriately adjust the training methodology.

Figure 7 shows how the detection confidence for the 'dog' object in the left image of Figure 6 varies with image rotation angle (-150° to 150°). At 0° (original image), the confidence is 0.95, but drops to 0.64 at a rotation of 78° , and detection fails entirely at approximately 100° . This demonstrates that the model is not invariant to rotation and that such transformations may affect its performance.

While this graph can itself be treated as a quantification of uncertainty, when a single metric is required, 'flatness' measures such as standard deviation (σ) or maximum slope are effective. This article uses σ . Since confidence is bounded within the interval $[0,1]$, its σ is constrained to the interval

$[0,0.5]$. Accordingly, Eq. (1) for expressing uncertainty as a percentage is defined as:

$$\text{Uncertainty} = 2\sigma \times 100\% \quad (1)$$

Computing the σ of the confidence values in Figure 7 yields $\sigma = 0.2689$, and from Eq. (1), the uncertainty is 53.78%. Therefore, with respect to the detection confidence of 0.95 for 'dog' in the left image of Figure 6, we can say that this prediction is approximately 53.8% uncertain.

The uncertainty metric reflects a model's stability and consistency rather than its performance. Accurate evaluation requires consideration of both. A model that achieves high performance but exhibits high uncertainty may be failing to capture essential features and instead relying on background cues; possible causes include insufficient diversity in the training data, among various other factors. Conversely, when both performance and uncertainty are low, this suggests that the model is overly simplistic, making a review of its architecture important. The ideal model achieves both high performance and low uncertainty.

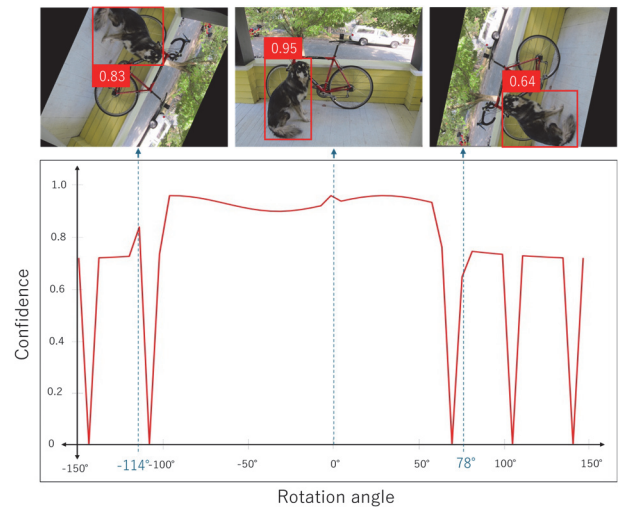


Figure 7 Image Rotation Angle vs. Detection Confidence for 'Dog'

4.2 Model-specific Methods

Compared to model-agnostic methods, model-specific methods require access to the model's source code, making their implementation more sophisticated and challenging. However, these methods enable more refined evaluation and allow developers to identify specific areas of the model in need of improvement. The following outlines two such methods.

In conventional neural networks, precise weights are computed for each neuron or node and adjusted during training.

By contrast, in Bayesian networks¹⁸⁾ the network weights are assumed to be random variables, and the goal of training is to estimate the mean and variance of the probability distributions (typically Gaussian distributions) that describe these variables. During the forward pass when inputting data, sampling is performed from these probability distributions to obtain specific weight values. When computing uncertainty for a given input, multiple forward passes are executed, resulting in a different combination of weight values within the network for each pass. Equation (2) shows the mathematical formulation of a Bayesian neuron.

$$y = f(\sum_i w_i x_i + b), \text{ where } w_i \sim N(\mu_i, \sigma_i^2) \quad (2)$$

y : Neuron output

$f(\cdot)$: Activation function

w_i : Weight

x_i : Input features

b : Bias value

μ_i, σ_i^2 : mean and variance of distribution N

When a model includes Dropout layers²⁾, Monte Carlo Dropout¹⁹⁾ is a technique that leverages them to compute uncertainty. While Dropout is originally only used during training to prevent overfitting, this technique keeps the Dropout layers active at inference time as well, assigning zeros to random weights in order to vary the weight patterns. This effectively disables the corresponding weights, producing an effect analogous to that of a Bayesian network. As a result, multiple inference outputs can be obtained.

5 Conclusions

In AI model development, it is important to evaluate accuracy after building a model and to make improvements based on the results. However, conventionally, improving a model requires addition of large amounts of data, entailing considerable cost. One reason for this is that the basis for AI decision-making is a black box, making it difficult for people to understand what data characteristics are effective. This article addresses these challenges by introducing XAI techniques and presenting implementation examples, thereby demonstrating the potential for a more effective approach to selecting additional training data. Furthermore, we propose that by incorporating these XAI techniques into the performance evaluation of an AI pipeline, it is possible to visualize

the presence of insufficient or excess data and facilitate better understanding. It should be noted that since this study evaluated an object detection model, different evaluation methods will be required for other tasks.

Going forward, we plan to pursue application of these methods to different AI models and real-world data, as well as further development and evaluation of XAI techniques using uncertainty-based functions. In addition, when incorporating these evaluation techniques into an AI pipeline, it will be necessary to design evaluation metrics suited to the specific objectives, and these too remain subjects for future research.

References

- 1) M. T. Ribeiro *et al.*, ACM SIGKDD **22**, 1135–1144 (2016).
- 2) S. Lundberg *et al.*, Neural Inf. Process. Syst. **31**, 4768–4777 (2017).
- 3) T.-Y. Lin *et al.*, Eur. Conf. Comput. Vis. **13**, 740–755 (2014).
- 4) V. B. Truong *et al.*, Proc. Mach. Learn. Res. **222**, 1385–1400 (2023).
- 5) J. Kauffmann *et al.*, arXiv:2006.10609 (2020).
- 6) N. Wang *et al.*, Adv. Neural Inf. Process. Syst. **32**, 7686–7695 (2018).
- 7) G. Hinton *et al.*, NIPS Workshop (2014).
<https://doi.org/10.48550/arXiv.1503.02531>
- 8) M. B. Muhammad *et al.*, arXiv:2008.00299 (2020).
- 9) B. Zhou *et al.*, IEEE Conf. Comput. Vis. Pattern Recognit., 2921–2929 (2016).
- 10) R. R. Selvaraju *et al.*, Int. J. Comput. Vis. **128**, 336–359 (2019).
- 11) Z. Feng *et al.*, arXiv:2302.01642 (2023).
- 12) V. Buono *et al.*, arXiv:2406.01274 (2024).
- 13) C. Yang *et al.*, J. Cheminform. **15**, 13 (2023).
- 14) A. Kendall *et al.*, Adv. Neural Inf. Process. Syst. **30**, 5574–5584 (2017).
- 15) B. Lakshminarayanan *et al.*, Adv. Neural Inf. Process. Syst. **30**, 6402–6413 (2017).
- 16) L. Breiman, Mach. Learn. **24**, 123–140 (1996).
- 17) G. Wang *et al.*, Med. Image Anal. **54**, 154–162 (2019).
- 18) D.J.C. MacKay, Neural Computation, **4**, 448–472 (1992).
- 19) Y. Gal *et al.*, Proc. Mach. Learn. Res. **48**, 1050–1059 (2016).

² Dropout layer: A layer that randomly disables some neurons (computational units) during training.

Authors



Ryu Kidokoro
3rd laboratory
Advanced Research Laboratory



Lawrence Earl Carandang
3rd laboratory
Advanced Research Laboratory

Publicly available

Overview of Public-key Cryptography

Takashi Mori

[Summary]

Cryptography for securely exchanging confidential information over communication channels that may be subject to eavesdropping has a long history, and was also used for military and diplomatic applications during World War I. In encrypted communication, encryption and decryption are performed based on a key. However, with symmetric-key cryptography, where the encryption key and decryption key are identical, the challenge lies in how to securely share the common key between the sender and receiver. In other words, since it is not possible to transmit a common key over a communication channel that may be subject to eavesdropping, it is necessary to distribute the common key through a separate secure means. Public-key cryptography solves this challenge by preparing different keys for encryption and decryption, making the encryption key public as a public key, and keeping the decryption key secret as a private key. There are key exchange methods for securely sharing a secret key over a communication channel that may be subject to eavesdropping, and these too fall under the broader definition of public-key cryptography. Public-key cryptography is widely used in encrypted communications on the internet and has become an indispensable technology in our modern information society. This article explains the basic principles of RSA cryptography, Diffie-Hellman key exchange, and elliptic curve cryptography as current public-key cryptographic systems, and introduces post-quantum cryptography, which is difficult to break even using quantum computers, as next-generation public-key cryptography.

1 Introduction

Human society needs to exchange confidential information through secure communications. However, there is a risk of eavesdropping by third parties during actual communication. Wireless communication in particular poses the problem that interception by third parties is easy, but it is difficult for the legitimate sender and receiver to detect that interception has occurred. Consequently, encryption is needed to prevent exposure of secret information even if communications are intercepted by third parties. Encryption has a long history, and was also used in World War I for military and diplomatic purposes¹⁾. In other words, the purpose of cryptography is to securely transmit secret information over a communication channel that may be subject to eavesdropping.

For example, in classical cryptography (Figure 1), the sender (Alice) shifts each letter of the plaintext “hello” by 4 characters to produce the ciphertext “lipps,” which is then sent to the receiver (Bob). Even if the eavesdropper (Eve) intercepts the ciphertext “lipps,” she cannot make sense of it, whereas the receiver (Bob) can obtain the plaintext “hello” by shifting each letter of the received ciphertext “lipps” back by 4 characters. If we break this cipher down into “scheme” and “key,” the scheme is to shift each letter of the alphabet

by n characters, and the key is $n = 4$. Both the scheme and the key are required to break a cipher. However, in classical cryptography, knowing the scheme made it easy to break the cipher, so both the scheme and the key were kept secret. In the example above, if the scheme of shifting the alphabet by n characters is known, the cipher can always be broken by checking all 26 possible values of n .

On the other hand, in modern cryptography, the system is designed so that security is maintained even if the scheme is known, as long as the key remains unknown. Therefore, only the key is kept secret while the scheme is made public. It is believed that making the scheme public, rather than keeping it secret, allows it to be more widely studied and actually results in higher security.



Figure 1 Example of Classical Cryptography

In the classical cipher described above, the number of characters shifted during encryption, n , must equal the number of characters shifted back during decryption, n' (i.e., $n = n'$). A cipher in which the encryption key and the decryption key

are the same is called a symmetric-key cipher (or secret-key cipher). In symmetric-key cryptography, the challenge lies in how the sender and receiver can securely share the common key. In other words, since the common key cannot be sent over a communication channel that may be subject to eavesdropping, it is necessary to arrange a separate secure means of delivering the common key. Public-key cryptography solves this problem. As shown in Figure 2, the receiver (Bob) prepares two different keys for encryption and decryption: the encryption key is made public as the public key, while the decryption key is kept secret as the private key. The sender (Alice) encrypts the plaintext using the public key and sends the ciphertext, and the receiver (Bob) decrypts the ciphertext using the private key to obtain the plaintext. Since the public key used for encryption cannot decrypt the ciphertext, an eavesdropper (Eve) cannot obtain the plaintext even if she acquires both the public key and the ciphertext. The invention of public-key cryptography made it possible to securely transmit secret information using only a communication channel that may be subject to eavesdropping. This article explains the basics of the main public-key cryptographic schemes currently in widespread use, and also presents an overview of post-quantum cryptography (PQC), which represents the next generation of public-key cryptography.

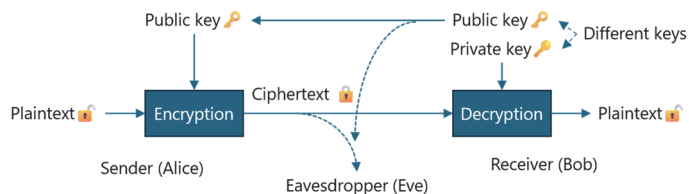


Figure 2 Public-key Cryptography Procedure

2 Features of Public-key Cryptography

The first challenge is the number of keys required for many-to-many encrypted communication. In symmetric-key cryptography, since sharing a key allows the ciphertext to be decrypted with that shared key, a different key is required for every possible combination of communicating parties. In contrast, with public-key cryptography, even if multiple senders share the same public key for encryption, security is maintained because the public key cannot be used to decrypt the ciphertext. For this reason, even when conducting many-to-many encrypted communication, the number of key pairs required is only equal to the number of receivers. For

example, in an 8-to-8 communication scenario (Figure 3), symmetric-key cryptography requires 28 keys, while public-key cryptography requires only 8 key pairs.

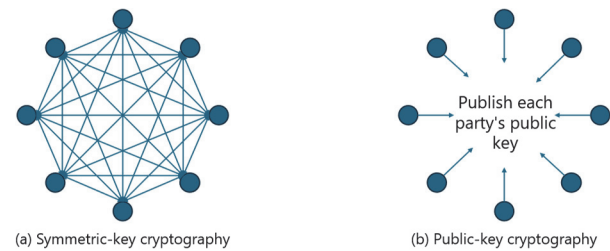


Figure 3 Number of Keys Required by 8-to-8 Cryptographic Communications

The second challenge is the encryption of large volumes of data. Although public-key cryptography has the advantage of not requiring sharing of a secret key and enabling many-to-many encrypted communication with ease, it has the drawback of requiring a large amount of computation for both encryption and decryption. Therefore, transmitting large volumes of data using public-key cryptography poses the challenge of placing a heavy computational load on the system. On the other hand, symmetric-key cryptography has the drawbacks of requiring secure sharing of a common key and needing a large number of common keys for many-to-many encrypted communication, but it has the advantage of requiring only a small amount of computation for both encryption and decryption. For this reason, encrypted communication over the internet commonly uses a method in which public-key cryptography is first used to share a key between the sender and receiver, and that key is then used to transmit data via symmetric-key cryptography. Since the amount of data involved in key sharing using public-key cryptography is small, the computational load does not pose a significant problem, and large volumes of data are transmitted efficiently via symmetric-key cryptography. At the same time, the sharing of the common key is carried out securely through public-key cryptography. This method thus cleverly avoids the drawbacks of both types of cryptography.

3 Current Public-key Cryptography²⁾

3.1 RSA Cryptography

RSA cryptography is a public-key cryptosystem published in 1978 by three researchers—Rivest, Shamir, and Adleman (RSA being their initials)—and is based on the difficulty of factoring large numbers into their prime factors. That is,

given two prime numbers p and q , it is easy to compute their product $n = p \times q$, but finding p and q from a given product n through prime factorization requires an enormous amount of computation.

Figure 4 shows the RSA cryptography procedure.

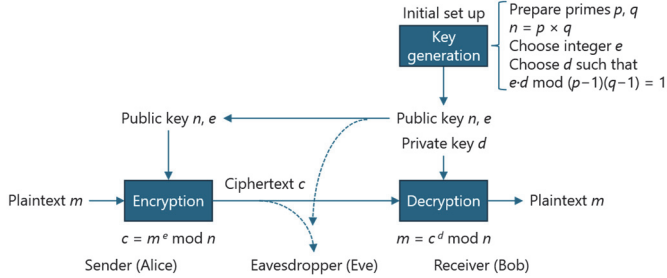


Figure 4 RSA Cryptography Procedure

First, the following initial setup is performed on the receiver side.

- Prepare two prime numbers p and q
- Compute the product of the primes $n = p \times q$
- Choose a positive integer e that is less than $(p-1)(q-1)$ and coprime to $(p-1)(q-1)$
- Choose d such that $e \cdot d \bmod (p-1)(q-1) = 1$ (where mod denotes the remainder of division)
- Make n and e public as the public key
- Keep d secret as the private key

This numerical example uses small numbers:

- Prepare two prime numbers $p = 3$, $q = 5$
- Compute the product of the primes $n = 3 \times 5 = 15$
- Choose a positive integer $e = 3$ that is less than $(3-1)(5-1) = 8$ and coprime to 8
- Choose $d = 3$ such that $e \cdot d \bmod (3-1)(5-1) = 1$ ($e \cdot d = 3 \times 3 = 9$, $9 \bmod 8 = 1$)
- Make $n = 15$, $e = 3$ public as the public key
- Keep $d = 3$ secret as the private key (In this numerical example e and d are equal, but in practice they would have different values.)

RSA cryptography makes use of the properties of exponentiation in mod n , specifically that raising an integer m to the power of $(p-1)(q-1) + 1$ modulo n returns the original m . For example, as shown in the numerical example in Table 1, raising any value of m to the power of $(p-1)(q-1) + 1 = 9$ modulo 15 returns the original m .

Table 1 Exponentiation Calculations for mod 15

m	Exponentiation								
	1	2	3	4	5	6	7	8	9
1	1	1	1	1	1	1	1	1	1
2	2	4	8	1	2	4	8	1	2
3	3	9	12	6	3	9	12	6	3
4	4	1	4	1	4	1	4	1	4
5	5	10	5	10	5	10	5	10	5
6	6	6	6	6	6	6	6	6	6
7	7	4	13	1	7	4	13	1	7
8	8	4	2	1	8	4	2	1	8
9	9	6	9	6	9	6	9	6	9
10	10	10	10	10	10	10	10	10	10
11	11	1	11	1	11	1	11	1	11
12	12	9	3	6	12	9	3	6	12
13	13	4	7	1	13	4	7	1	13
14	14	1	14	1	14	1	14	1	14

Let m denote the message to be transmitted. On the sender side, the message m is encrypted using the public key n and e according to the following equation, and the ciphertext c is transmitted.

$$c = m^e \bmod n \quad (1)$$

On the receiver side, the received ciphertext c is decrypted using the private key d according to the following equation.

$$m = c^d \bmod n \quad (2)$$

Through encryption and decryption, the message m is raised to the power of $e \cdot d$ modulo n . Since e and d are chosen during the initial setup such that $e \cdot d \bmod (p-1)(q-1) = 1$, and raising m to the power of $(p-1)(q-1) + 1$ modulo n returns the original m , the original message m is recovered through decryption. For example, as shown in the numerical example in Table 2, encrypting m by raising it to the power of $e = 3$ and decrypting the ciphertext by raising it to the power of $d = 3$ returns the original m .

Table 2 Encryption and Decryption Using Exponentiation

m	Encryption Exponentiation			Decryption Exponentiation		
	1	2	3	1	2	3
1	1	1	1	1	1	1
2	2	4	8	8	4	2
3	3	9	12	12	9	3
4	4	1	4	4	1	4
5	5	10	5	5	10	5
6	6	6	6	6	6	6
7	7	4	13	13	4	7
8	8	4	2	2	4	8
9	9	6	9	9	6	9
10	10	10	10	10	10	10
11	11	1	11	11	1	11
12	12	9	3	3	9	12
13	13	4	7	7	4	13
14	14	1	14	14	1	14

In RSA cryptography, n and e are made public, but decryption requires d . If p and q were known, d could be derived from $e \cdot d \bmod (p-1)(q-1) = 1$, and the cipher would be broken. Finding the prime numbers p and q such that $n = p \times q$ from a known n is a problem of prime factorization, and since prime factorization becomes extremely difficult when p and q are very large numbers, RSA cryptography is considered computationally unfeasible to break. However, as computing power improves, the speed of prime factorization also increases year by year, making it necessary to periodically increase the size of p and q . Note that in the initial setup on the receiver side, since the product $n = p \times q$ is computed after preparing the two prime numbers p and q , there is no need to perform prime factorization of large numbers.

The practical problems with RSA cryptography include issues arising from increasing key length and operational issues related to key generation³⁾. For increasing key length, while it is necessary to lengthen the key in response to improvements in computing performance in order to maintain security, processing long keys is difficult in environments with limited processing capacity, such as IC cards and embedded devices. The key generation issue is that RSA cryptography uses two large prime numbers, and if one of the primes is the same for different users, the private key can be derived relatively easily. The probability of different users randomly selecting the same prime number is negligibly small, but there is a possibility of using the same prime due to faults in random number generators or administrator errors. Even if one prime number is the same, the public key, which is the product of the two prime numbers, will differ, making it difficult to detect the duplication, and this is a problem.

3.2 Diffie-Hellman (DH) Key Exchange

DH key exchange is a key sharing method published in 1976 by Diffie and Hellman and is based on the difficulty of the discrete logarithm problem for large numbers. That is, given integers g and a and a prime number p , it is easy to compute the power $g^a \bmod p$. However, finding the integer a given the integer g , prime p , and the power $g^a \bmod p$ is a discrete logarithm problem that requires an enormous amount of computation.

Computation of $g^a \bmod p$ can be performed with a small

number of multiplications instead of performing a multiplications, as follows. Compute $x_2 = g^2 \bmod p$, $x_4 = x_2^2 \bmod p = g^4 \bmod p$, $x_8 = x_4^2 \bmod p = g^8 \bmod p$, $x_{16} = x_8^2 \bmod p = g^{16} \bmod p$, $\dots$ and then, for example, when seeking the 22nd power, compute $x_{16} \cdot x_4 \cdot x_2 = g^{16} \cdot g^4 \cdot g^2 \bmod p = g^{22} \bmod p$, which allows large powers to be computed efficiently.

Similarly, the logarithm problem of finding a when y_a and g are known in $y_a = g^a$ can be solved easily as follows. Compute $x_2 = g^2$, $x_4 = x_2^2 = g^4$, $x_8 = x_4^2 = g^8$, $x_{16} = x_8^2 = g^{16}$, $\dots$ and then efficiently find a by progressively narrowing down the range, for example as $x_{16} \cdot x_4 > y_a \geq x_{16}$. While the logarithm problem without $\bmod p$ can be solved efficiently in this way, the same approach cannot be applied to the discrete logarithm problem with $\bmod p$, making it difficult to find a .

Figure 5 shows the procedure for DH key exchange and is described below.

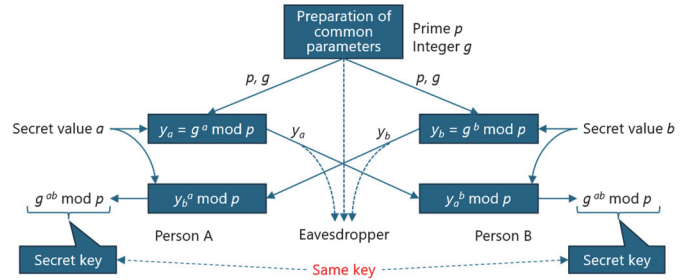


Figure 5 DH Key Exchange Procedure

- Prime p and integer g are made public as common parameters.
- Person A chooses a secret value a , computes $y_a = g^a \bmod p$, and makes it public.
- Person B chooses a secret value b , computes $y_b = g^b \bmod p$, and makes it public.
- Person A uses their own secret value a and Person B's public information y_b to compute the secret key $y_b^a \bmod p$.
- Person B uses their own secret value b and Person A's public information y_a to compute the secret key $y_a^b \bmod p$.
- Both parties' secret keys are equal to $g^{ab} \bmod p$.

To illustrate with a small numerical example:

- Prime $p = 23$ and integer $g = 9$ are made public.
- Person A's secret value is $a = 17$, and $y_a = 9^{17} \bmod 23 = 3$ is made public.
- Person B's secret value is $b = 5$, and $y_b = 9^5 \bmod 23 = 8$ is made public.

- Person A uses secret value $a = 17$ and Person B's public information $y_b = 8$ to compute the secret key $8^{17} \bmod 23 = 13$.
- Person B uses secret value $b = 5$ and Person A's public information $y_a = 3$ to compute the secret key $3^5 \bmod 23 = 13$.
- Both parties' secret key is 13.

In this way, both parties can share a secret key without ever transmitting it over a potentially eavesdropped channel. For an eavesdropper, since prime p , integer g , $y_a = g^a \bmod p$, and $y_b = g^b \bmod p$ are all public, if the value of a were known, the secret key could be derived via $y_b^a \bmod p$ and if the value of b were known, the secret key could be derived via $y_a^b \bmod p$. However, finding the value of a from $y_a = g^a \bmod p$, or finding the value of b from $y_b = g^b \bmod p$, is a discrete logarithm problem, and since making p a very large prime makes the discrete logarithm problem difficult to solve, breaking DH key exchange becomes computationally unfeasible. However, as computing power improves, the speed at which the discrete logarithm problem can be solved also increases year by year, so it is necessary to periodically increase p . Note that in the DH key exchange procedure, since the exponentiations $y_a = g^a \bmod p$ and $y_b = g^b \bmod p$ are computed after preparing the secret values a and b , there is no need to solve the discrete logarithm problem. RSA encryption described in the previous section can transmit arbitrary messages, so key sharing is also possible by sending a secret key as a message, making it usable for both data encryption and key sharing. In contrast, DH key exchange allows both parties to share some secret key, but cannot transmit arbitrary messages, making it a method dedicated solely to key sharing.

3.3 Elliptic Curve Cryptography (ECC)

ECC is a public-key cryptosystem independently proposed by Miller and Koblitz around 1985, and is based on the difficulty of the discrete logarithm problem on elliptic curves. That is, given a point G on an elliptic curve and a natural number s , it is easy to compute the elliptic curve multiplication $T = s \times G$. However, the elliptic curve discrete logarithm problem, which asks for the natural number s given a point G on an elliptic curve and the elliptic curve multiplication $T = s \times G$, requires an enormous amount of computation.

An elliptic curve is a curve expressed as $y^2 = x^3 + ax + b$,

and for example, the elliptic curve with $a = -1$ and $b = 1$ appears as shown in Figure 6. The concept of arithmetic operations on the elliptic curve is as follows: Let $G = (0.5, 0.790569)$ be a point on the elliptic curve. As shown in Figure 7, find the intersection point A of the tangent line at point G and the elliptic curve, then let B be the point obtained by inverting the sign of y at point A . This is defined as the elliptic curve addition and multiplication $B = G + G = 2 \times G$. Similarly, as shown in Figure 8, find the intersection point C of the line connecting point G and point B with the elliptic curve, then let D be the point obtained by inverting the sign of y at point C . This is defined as the elliptic curve addition and multiplication $D = B + G = 3 \times G$. Also, as shown in Figure 9, when there is no intersection point between the line connecting point A and point B and the elliptic curve, we set $A + B = O$, and define the subtraction on the elliptic curve as $B = O - A = -A$. The computation of the elliptic curve multiplication $T = s \times G$ can be performed with a smaller number of additions, as described below, rather than performing s additions. Similarly to the exponentiation calculation in the previous section, we compute $A_2 = G + G = 2 \times G$, $A_4 = A_2 + A_2 = 4 \times G$, $A_8 = A_4 + A_4 = 8 \times G$, $A_{16} = A_8 + A_8 = 16 \times G$, $\dots$, and when, for example, we wish to find the 22-fold multiple, we can efficiently compute multiplication by large numbers as $A_{16} + A_4 + A_2 = 16 \times G + 4 \times G + 2 \times G = 22 \times G$.

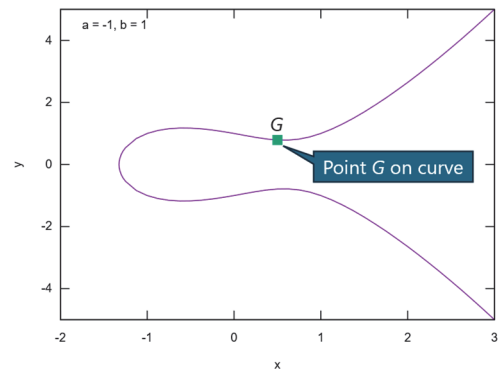


Figure 6 Example of Elliptic Curve

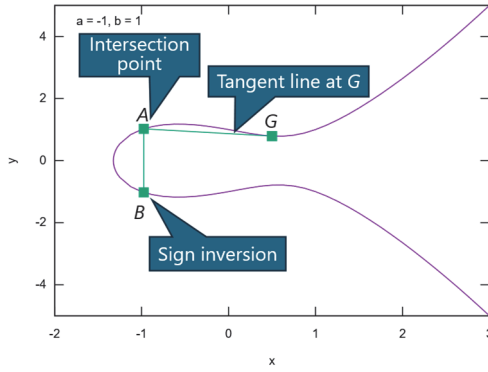


Figure 7 Example 1 of Addition on Elliptic Curve

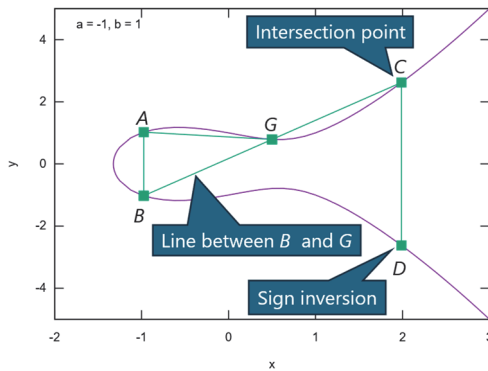


Figure 8 Example 2 of Addition on Elliptic Curve

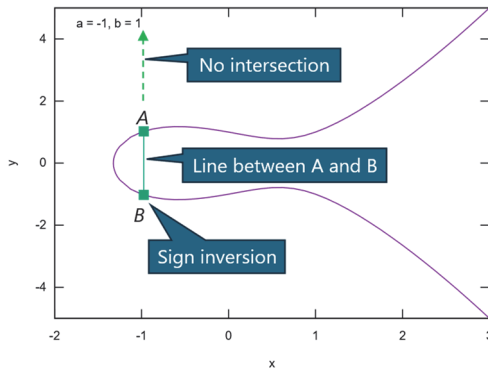


Figure 9 Example of Subtraction on Elliptic Curve

The discrete elliptic curve used in public-key cryptography is an integerized version of the elliptic curve described above. The concept is to let p be a prime number, and let x , y , a , and b be integers from 0 to $p-1$. The discrete elliptic curve is a discrete integer value expressed as $y^2 = x^3 + ax + b \pmod{p}$. Once a point G on the discrete elliptic curve is determined, if the natural number s is known, the multiplication $T = s \times G$ on the discrete elliptic curve can be computed efficiently in the same manner as described above. Conversely, the discrete logarithm problem of finding the natural number s from $T = s \times G$ is difficult to solve.

The procedure for ECC is explained under Figure 10.

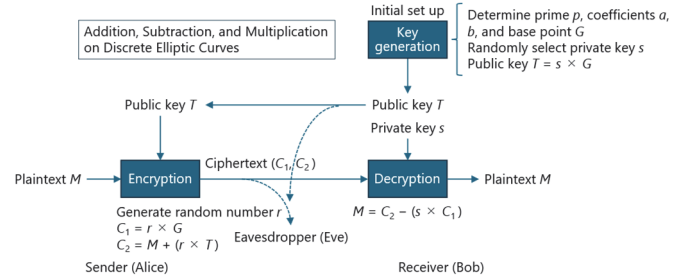


Figure 10 ECC Procedure

Key Generation:

- Publish the prime p , coefficients a , b , and base point G as common parameters.
- The receiver side randomly selects a private key s and publishes $T = s \times G$ as the public key.

Encryption:

- The sender side generates a random number r and computes $C_1 = r \times G$.
- To send the message M , compute $C_2 = M + (r \times T)$ and transmit (C_1, C_2) as the ciphertext.

Decryption:

- The receiver side computes $C_2 - (s \times C_1)$ using the ciphertext (C_1, C_2) and the private key s .
- As a result, $C_2 - (s \times C_1) = M + (r \times s \times G) - (s \times r \times G) = M$, and the message M is obtained.

A numerical example with small numbers is shown below:

Key Generation:

- Publish the prime $p = 263$, coefficients $a = 2$, $b = 3$, and base point $G = (200, 39)$ as common parameters.
- The receiver side selects the private key $s = 13$ and publishes $T = s \times G = (114, 104)$ as the public key.

Encryption:

- The sender side generates a random number $r = 35$ and computes $C_1 = r \times G = (219, 39)$.
- To send the message $M = (88, 84)$, compute $C_2 = M + (r \times T) = (250, 161)$ and send $(C_1, C_2) = ((219, 39), (250, 161))$ as the ciphertext.

Decryption:

- The receiver side computes $C_2 - (s \times C_1)$ using the ciphertext (C_1, C_2) and the private key $s = 13$.
- As a result, $(250, 161) - (13 \times (219, 39)) = (250, 161) - (201, 155) = (88, 84)$, and the message M is obtained.

The practical advantages of ECC include the benefits of short key lengths and operational aspects of key generation³⁾.

ECC can achieve equivalent security to RSA cryptography with a key length of approximately one-tenth of that required, making it suitable for use in environments with limited processing performance, such as IC cards and embedded devices. In ECC, a different natural number s should be selected for each user as the private key. If the same natural number s is selected by different users, the public key $T = s \times G$ will also be the same, making it easy to detect duplicates. This provides an operational advantage in key generation, because the problem of issuing weak keys is less likely.

3.4 Elliptic Curve Diffie-Hellman (ECDH) Key Exchange

ECDH key exchange is a method that performs DH key exchange using elliptic curves. Like ECC, it is based on the difficulty of the discrete logarithm problem on elliptic curves.

The procedure for ECDH key exchange is explained under Figure 11.

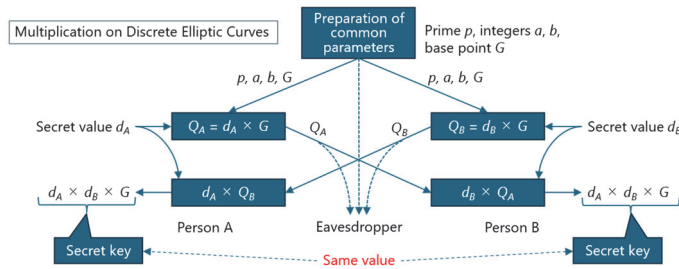


Figure 11 ECDH Key Exchange Procedure

- Publish the prime p , coefficients a , b , and base point G as common parameters.
- Person A decides on a secret value d_A and computes $Q_A = d_A \times G$ to publish.
- Person B decides on a secret value d_B and computes $Q_B = d_B \times G$ to publish.
- Person A computes the secret key $d_A \times Q_B$ using their own secret value d_A and Person B's public information Q_B .
- Person B computes the secret key $d_B \times Q_A$ using their own secret value d_B and Person A's public information Q_A .
- Both parties' secret keys are equal to $d_A \times d_B \times G$.

A numerical example with small numbers is shown below:

- Publish the prime $p = 263$, coefficients $a = 2$, $b = 3$, and base point $G = (200, 39)$ as common parameters.
- Person A decides on a secret value $d_A = 23$ and publishes $Q_A = d_A \times G = (198, 95)$.
- Person B decides on a secret value $d_B = 79$ and publishes $Q_B = d_B \times G = (129, 9)$.

- Person A computes the secret key $d_A \times Q_B = 23 \times (129, 9) = (133, 138)$ using their own secret value $d_A = 23$ and Person B's public information $Q_B = (129, 9)$.
- Person B computes the secret key $d_B \times Q_A = 79 \times (198, 95) = (133, 138)$ using their own secret value $d_B = 79$ and Person A's public information $Q_A = (198, 95)$.
- Both parties' secret keys are equal to $d_A \times d_B \times G = (133, 138)$.

In this way, both parties can share a secret key without transmitting it over a communication channel that may be subject to eavesdropping.

The ECC described in the previous section can transmit arbitrary messages, so key sharing is also possible by sending the secret key as a message, making it usable for both data encryption and key sharing. In contrast, ECDH key exchange allows both parties to share some secret key, but cannot transmit arbitrary messages, and is therefore a method dedicated exclusively to key sharing.

3.5 Strength of Public-key Cryptography

The security of the aforementioned public-key cryptography is based on the enormous computational effort required to solve the prime factorization and the discrete logarithm problem of large numbers. A comparison between the computational performance of the world's top supercomputer and the computational performance required to solve prime factorization and the discrete logarithm problem on elliptic curves within one year has been reported by CRYPTREC⁴⁾, and it would take an unrealistic period of over 1000 years to break the currently used 2048-bit RSA or 256-bit ECDH. However, since the computational performance of supercomputers improves year by year in accordance with Moore's Law, it is necessary to periodically increase the key length of cryptographic systems to raise the computational effort required for decryption. Conversely, by periodically increasing the key length, it is possible to maintain the security of public-key cryptography against the improvement in supercomputer computational performance that follows Moore's Law.

On the other hand, there is Shor's algorithm, which uses a quantum computer to efficiently solve prime factorization and the discrete logarithm problem. For this reason, although RSA cryptography and ECC are theoretically considered broken, breaking actual cryptographic systems requires a large-scale quantum computer, and current quantum

computers are nowhere near sufficient in scale, meaning that RSA cryptography and ECC have not actually been broken in practice. However, there is a technique known as a “Harvest Now, Decrypt Later attack” by which an attacker saves intercepted ciphertext and decrypts it at a future time when quantum computer performance has improved. This poses a threat for confidential data that does not become obsolete over time, and therefore countermeasures are needed in advance.

4 PQC

4.1 CRYSTALS-Kyber

Research and development of PQC that is difficult to break even using large-scale quantum computers is being actively pursued as successors to the current RSA cryptography and ECC. The US National Institute of Standards and Technology (NIST) launched a public solicitation for PQC standardization in 2016, and CRYSTALS-Kyber, based on lattice problems, was adopted as a key exchange algorithm and has been published as the standardization document FIPS 203⁵⁾. The lattice problem used in PQC is

$$\begin{cases} a_{11}s_1 + a_{12}s_2 + \cdots + a_{1n}s_n \approx t_1 \pmod{q} \\ a_{21}s_1 + a_{22}s_2 + \cdots + a_{2n}s_n \approx t_2 \pmod{q} \\ \vdots \\ a_{m1}s_1 + a_{m2}s_2 + \cdots + a_{mn}s_n \approx t_m \pmod{q} \end{cases} \quad (3)$$

the problem of finding $\mathbf{s} = (s_1, s_2, \dots, s_n)^T$ from $\mathbf{A} = (a_{ij})$ and $\mathbf{t} = (t_i)^T$ in the above equations of simultaneous linear approximation equations modulo prime q . When simultaneous equations use “=” without error, they can be solved efficiently. However, when the simultaneous equations use “ $\approx$ ” with error, they are difficult to solve.

The CRYSTALS-Kyber procedure⁶⁾ is shown under Figure 12.

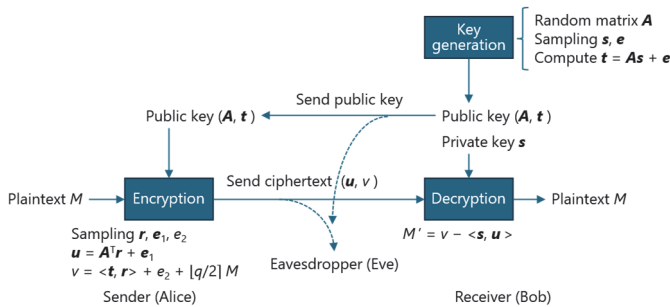


Figure 12 CRYSTALS-Kyber Procedure

Key Generation:

- Generate a matrix $\mathbf{A}$ randomly.
- Sample $\mathbf{s}, \mathbf{e}$ from a binomial distribution such that they are sufficiently small compared to q .
- Compute $\mathbf{t} = \mathbf{A}\mathbf{s} + \mathbf{e}$, and set $(\mathbf{A}, \mathbf{t})$ as the public key and $\mathbf{s}$ as the private key.

Encryption:

- Sample $\mathbf{r}, \mathbf{e}_1, \mathbf{e}_2$ such that they are sufficiently small compared to q .
- For plaintext M , compute $\mathbf{u} = \mathbf{A}^T \mathbf{r} + \mathbf{e}_1$, $v = \langle \mathbf{t}, \mathbf{r} \rangle + \mathbf{e}_2 + \lfloor q/2 \rfloor M$, and set $(\mathbf{u}, v)$ as the ciphertext. ($\lfloor a \rfloor$ denotes rounding to an integer, and $\langle \mathbf{t}, \mathbf{r} \rangle$ denotes the inner product of $\mathbf{t}$ and $\mathbf{r}$)

Decryption:

- Using the ciphertext $(\mathbf{u}, v)$ and private key $\mathbf{s}$, compute $M' = v - \langle \mathbf{s}, \mathbf{u} \rangle$; if $|M'|$ is less than $q/4$, the decrypted value is 0, otherwise it is 1.

4.2 Transition to PQC

In the transition to PQC, if the data protection period (period during which cryptographic protection is expected) is X years, the infrastructure migration period (period required to replace cryptographic processing) is Y years, and the period until cryptographic decryption by quantum technology is Z years, then as shown in Figure 13, the cryptography will be broken when $X + Y > Z$. Modern society uses public-key cryptography widely to protect various types of information, and the transition to PQC requires a long period (Y years) and considerable effort. While the likelihood of a quantum computer capable of breaking cryptography emerging in the near future is considered low, revolutionary advances in quantum technology could occur, making it difficult to estimate the period until cryptographic decryption by quantum technology (Z years). For this reason, the following responses are required: starting the transition to PQC as early as possible; implementing measures to shorten the infrastructure migration period (Y years) and the data protection period (X years); and monitoring developments in quantum technology⁷⁾.

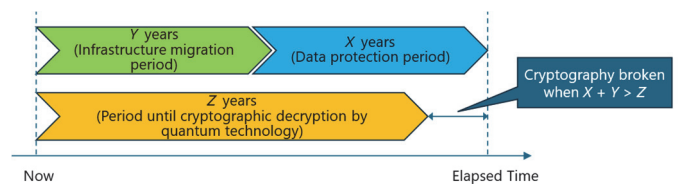


Figure 13 Mosca's Theorem

Social implementation of PQC has progressed in recent years. For example, Apple has announced the introduction of PQC into its iMessage app, which uses conventional ECC and PQC in combination to maintain security even if either one of the cryptographic systems is broken⁸⁾.

5 Conclusions

Cryptography is necessary to securely exchange confidential information over communication channels that may be subject to eavesdropping. However, symmetric-key cryptography requires sharing a common key between the sender and receiver, and the challenge lies in how to securely distribute this common key. Public-key cryptography solved this challenge by using different keys for encryption and decryption, making it possible to distribute the public key for encryption over a communication channel possibly subject to eavesdropping. Since public-key cryptography is also well-suited to many-to-many communication, it is widely used in encrypted communication on the current internet. This article explained the basic principles of RSA cryptography, DH key exchange, and ECC as current public-key cryptographic systems. However, it is known that these cryptographic systems can be efficiently broken when large-scale quantum computers are developed in the future, posing a threat to confidential data that does not become obsolete over time. For this reason, R&D into PQC is being conducted as a successor to the current RSA cryptography and ECC, and a method based on lattice problems has been adopted as a key exchange algorithm in the NIST PQC standard. Public-key cryptography is widely used in our modern information society, and while the infrastructure migration requires considerable effort, social implementation of PQC has already begun, and it is expected that the transition to PQC will continue going forward.

References

- 1) Shigeo Tsujii, "Development of Cryptography from for Military and Diplomacy to for Social Use," *Journal of IEICE*, vol. 106, no. 11, pp. 971–976, Nov. 2023 (in Japanese).
- 2) D.R. Stinson, *Cryptography Theory and Practice*, second edition, Chapman & Hall/CRC, 2002.
- 3) Takenobu Seito and Junji Shikata, "Recent Developments in Public-Key Cryptography: From RSA to Elliptic Curve Cryptography," *Kin'yu Kenkyu*, Institute for Monetary and Economic Studies, Bank of Japan, July 2013 (in Japanese).
<https://www.imes.boj.or.jp/research/papers/japanese/kk32-3-2.pdf>
- 4) CRYPTREC Report 2024, "Cryptographic Technology Evaluation Committee Report," Mar. 2025 (in Japanese).
- 5) National Institute of Standards and Technology, "Module lattice-based key-encapsulation mechanism standard," Aug. 2024.
<https://doi.org/10.6028/NIST.FIPS.203>
- 6) Masaya Yasuda, "Introduction to NIST-standardized Cryptography Based on Lattices," *Journal of IEICE*, vol. 106, no. 11, pp. 982–987, Nov. 2023 (in Japanese).
- 7) Tadahiko Ito, "Issues for the Transition to Post-quantum Cryptography," *Journal of IEICE*, vol. 106, no. 11, pp. 1026–1030, Nov. 2023 (in Japanese).
- 8) Apple Security Engineering and Architecture, "iMessage with PQ3: The new state of the art in quantum-secure messaging at scale," Feb. 2024.
<https://security.apple.com/blog/imessage-pq3/>

Author



Takashi Mori
Technology Planning Dept.
Advanced Research Laboratory

Publicly available

Development of Near-Infrared Spectroscopy Module for Process Analytical Technology

Daisuke Oida, Noriaki Yamasaki, Masaru Koshihara, Takanori Saitoh

[Summary]

Due to chronic pharmaceutical supply shortages, continuous manufacturing is being explored as a means of achieving stable, high-quality drug production. We developed a near-infrared (NIR) spectroscopy module for process analytical technology (PAT) applications supporting continuous manufacturing. This article first describes the module's fundamental architecture, key features, and data analysis methods. We then evaluate module's performance by using pharmaceutical excipients to demonstrate that the module achieves high-accuracy estimates of concentration and is suitable for deployment in continuous manufacturing processes.

1 Introduction

Pharmaceutical supply shortages have persisted since around 2020, triggered by a contamination scandal in which numerous patients suffered adverse health effects. Investigations revealed systemic failures in quality control, including non-compliance with approved manufacturing procedures and falsification of quality test records. In response, Good Manufacturing Practice (GMP) regulations governing pharmaceutical production and quality management were revised. As a result, quality control deficiencies were identified at other pharmaceutical manufacturers, and subsequent business suspension orders led to widespread shipment halts across many drug products. Compounding this, the COVID-19 pandemic that emerged the same year drove a sharp surge in demand for certain medications, including cold remedies, further exacerbating the supply shortages. Supply shortages have continued during resurgences of COVID-19 and influenza. Against this backdrop, establishing a manufacturing framework capable of stably supplying high-quality pharmaceuticals has become an urgent priority.

Pharmaceuticals are manufactured by the following sequential processes. First, the active ingredient is precisely weighed (weighing step). It is then mixed with a liquid such as water, granulated, and dried (granulation step). Next, excipients and other additives are blended in (mixing step), and finally the mixture is compressed into tablet form (tableting step). At each step, all material is removed from the manufacturing equipment and inspected for quality before proceeding to the next stage. If the material passes inspection, it is transferred to the equipment for the

subsequent step. This currently predominant approach is known as batch manufacturing. This approach has notable drawbacks because material transfer and inspection between steps are performed manually, making the process labor-intensive and introducing the risk of contamination.

As a solution, continuous manufacturing is being explored whereby each process step is linked directly to enable uninterrupted production. Performing all steps from weighing through to tableting in one continuous flow, is expected to cut costs, shorten production time, and prevent contamination. However, continuous manufacturing requires monitoring product quality inline rather than removing material for offline inspection. For example, moisture content is a critical parameter at the granulation step, while blend uniformity is key at the mixing step. The technologies used to measure these parameters inline are collectively called Process Analytical Technology (PAT)¹⁾. The most widely used PAT is near-infrared (NIR) spectroscopy²⁾⁻⁴⁾.

Near-infrared spectroscopy is an analytical technique that determines sample composition by analyzing the transmission or reflectance spectra of near-infrared light (wavelength range: 800 to 2500 nm) irradiated onto a sample. When NIR radiation is incident on a sample, molecular bonds predominantly found in water and organic compounds, such as O–H, N–H, and C–H groups, undergo vibrational excitation, resulting in absorption of the incident radiation. Since the degree of light absorption is dependent on the chemical composition of the sample, qualitative and quantitative compositional information can be extracted through spectral analysis.

In NIR spectroscopy, broadband white light is directed onto

the sample, and the transmitted or reflected light is dispersed before being detected by a photodetector to capture the spectrum. The white light sources used in NIR spectroscopy typically operate at relatively low irradiance, enabling non-destructive measurement of samples without causing photodegradation or structural alteration. Furthermore, the use of high-speed photodetectors allows rapid spectral acquisition, thereby facilitating real-time compositional analysis. Based on these characteristics, NIR spectroscopy can be regarded as an optimal inline inspection technology for PAT applications.

NIR spectroscopic instruments used in PAT applications are integrated directly into existing pharmaceutical manufacturing lines. Consequently, a compact form factor is required to accommodate line space constraints. In addition, lightweight construction is necessary to facilitate easy handling during maintenance. Furthermore, at blending processes where the instrument itself rotates along with the blending vessel, high vibration resistance is an essential requirement.

We have developed an NIR spectroscopy module that satisfies the requirements of PAT applications and can be readily integrated into pharmaceutical manufacturing lines. This article describes the configuration and key features of the developed NIR spectroscopy module. Furthermore, the spectral acquisition methodology and data analysis procedures are discussed in detail. Finally, by presenting the results of compositional analysis on samples prepared by blending pharmaceutical excipients used in actual drug formulations, we demonstrate the applicability of the developed module to quantitative analysis of pharmaceutical ingredients.

2 Near-Infrared Spectroscopy Module

2.1 External and Internal Structure

Figure 1 shows the external appearance of the developed NIR spectroscopy module. The module is cylindrical with a diameter of 44.6 mm and a length of 111 mm, a mass of about 550 g, and is constructed from stainless steel. The upper surface has an M25 threaded fitting for direct mounting onto pharmaceutical manufacturing equipment.

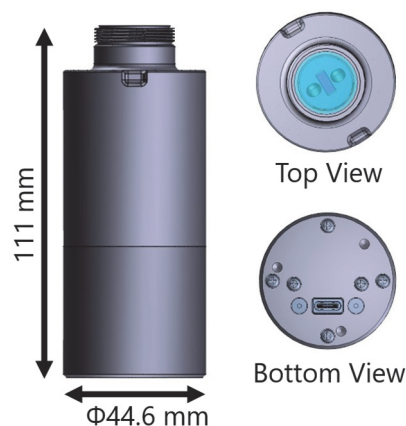


Figure 1 External Appearance of the NIR Spectroscopy Module

Figure 2 and Table 1 present the internal configuration diagram and specifications of the NIR spectroscopy module, respectively. The module primarily consists of two tungsten lamps for NIR illumination of the sample, a Linear Variable Filter (LVF) for dispersing the reflected light from the sample, and a line sensor that converts the dispersed light into electrical signals to capture the spectrum. Tungsten lamps are used in NIR spectroscopy due to their broad output wavelength coverage across the NIR region, low cost, and stable output power.

The LVF is an optical filter used for light dispersion, and Figure 3 shows its characteristics. The wavelength of light transmitted through the LVF is linearly related to the position along the filter. Consequently, the spectral intensity distribution, or the light intensity as a function of wavelength, can be measured by positioning a line sensor directly beneath the LVF. The module design can be simplified and made compact by exploiting the spectral characteristics of the LVF. Furthermore, unlike FT-NIR (Fourier Transform Near-Infrared) spectroscopy, which requires moving components to acquire spectra via an interferometer, the LVF-based design contains no moving parts, thereby providing superior vibration resistance.

The wavelength resolution of the LVF is defined as the FWHM (Full Width at Half Maximum) of the transmitted intensity profile when monochromatic light is incident on the filter. The wavelength resolution of the LVF in the present module is 2% of the central wavelength. For example, when monochromatic light at 1500 nm is incident on the filter, the FWHM of the transmitted spectrum is 30 nm. An InGaAs sensor, which exhibits high sensitivity in the NIR region, is used as the line sensor, with a pixel count of 256.

The module can be powered and controlled by connecting it to a PC via the USB port on the underside. The user can adjust the spectral acquisition timing and configure parameters such as the integration time (6 μ s or greater) and the on/off state of the tungsten lamp by using the bundled spectral capture software.

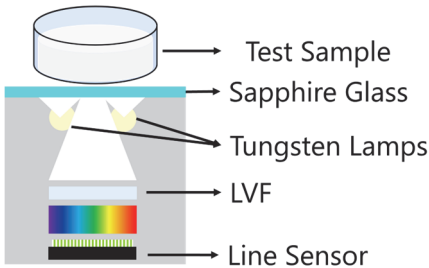


Figure 2 Near-Infrared Spectroscopy Module Internal Structure

Table 1 Near-Infrared Spectroscopy Module Specifications

Model	Standard Wavelength	Long Wavelength
Wavelength Range	1000 to 1700 nm	1250 to 2100 nm
Wavelength Resolution	2% of wavelength (30 nm @ 1500 nm)	
Pixel Count	256	
Integration Time	>6 μ s	
Max. Connections	8 per PC (limited by PC power supply capacity)	
Measurement Speed	>1 kHz (with 1 connection, 100 μ s integration time)	
Mass	<550 g	
Dimensions	$\phi 44.6 \times 111$ mm	
Material	SUS316	
Protection Class	IP67-equivalent	
I/F and Power Supply	USB3.0	
Power Consumption	2.5 W (5 V, 500 mA)	
Ambient Temperature Range	0 to 40°C	
Supported OS	Windows 10, 11	

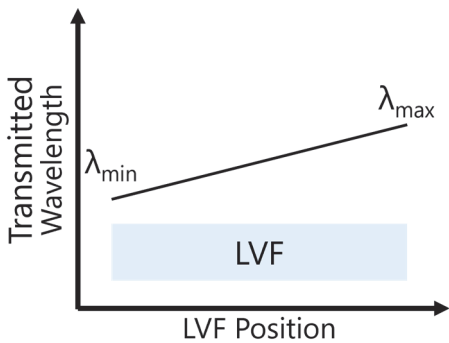


Figure 3 LVF Characteristics

2.2 Features

There are two NIR spectroscopy modules with different wavelength ranges. The first standard wavelength model covers a wavelength range of 1000 to 1700 nm, which is commonly used in NIR spectroscopy. The second long-wavelength model operates over an extended wavelength range of 1250 to 2100 nm, shifted toward longer wavelengths relative to the standard model. The long-wavelength model is better suited for moisture content measurement compared to the standard wavelength model. The primary absorption peaks of water molecules in the NIR region are at about 1450 and 1940 nm. However, the peak at about 1940 nm exhibits significantly greater absorbance. Consequently, the long-wavelength model, which encompasses this higher-absorbance peak, determines moisture content with higher accuracy.

In addition to the different wavelength ranges, both modules are designed to be compact and lightweight for easy integration into pharmaceutical manufacturing lines, while also offering superior vibration resistance. Furthermore, the spectral acquisition rate of 1 kHz enables real-time compositional estimation. Moreover, since up to eight modules can be controlled from one PC, multiple manufacturing lines can be monitored simultaneously from one workstation.

3 Applications in Pharmaceutical Production Lines

This section explains how the NIR spectroscopy module with the described features is used in pharmaceutical manufacturing lines. Figure 4 shows typical application examples in granulation and blending processes.

At granulation, a liquid such as water is mixed with the active pharmaceutical ingredient to form uniform granules. To this end, the moisture content is measured using the NIR spectroscopy module, and the mixing duration and liquid discharge volume are controlled based on the measured values, maintaining a constant moisture content in the granules.

Similarly, in the subsequent blending process, the NIR spectroscopy module is used to measure the concentration of the active pharmaceutical ingredient and excipients, and the mixing duration and excipient quantities are controlled based on the measured values, thereby enabling production

of pharmaceuticals with uniform composition. In this manner, use of the NIR spectroscopy module makes it possible to maintain quality control at each manufacturing stage.

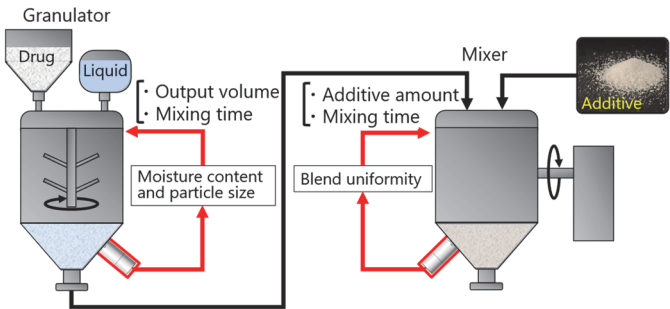


Figure 4 Module Use in Granulation and Mixing

4 Spectral Measurement and Analysis

Figure 5 shows the workflow for spectral measurement and analysis. First, three types of spectra are captured using the near-infrared spectroscopy module: the sample spectra (Figure 5(a)), the reference spectrum (Figure 5(b)), and the offset spectrum (Figure 5(c)). Subsequently, the spectral analysis stage performs two processing steps on the captured spectra: preprocessing (Figure 5(d)) and component estimation (Figure 5(e)). The preprocessing step reduces spectral noise and enhances peaks, thereby improving the accuracy of the downstream component estimation. Component estimation quantifies sample constituents by

constructing a calibration model using Partial Least Squares (PLS)⁵⁾ regression.

4.1 Spectral Measurement

The sample spectra are captured by measuring multiple samples with varying compositions. The reference spectrum is captured by measuring a standard reflector characterized by high reflectance and low wavelength dependence in the near-infrared region. The offset spectrum is captured without a sample.

The integration time must be carefully considered during spectral measurement. An excessively short integration time causes spectral peaks to be buried in noise, while an excessively long integration time results in sensor saturation. Since spectral intensity also depends on the sample type and distance between the near-infrared spectroscopy module and the sample, the integration time must be configured appropriately for the given measurement conditions.

4.2 Spectrum Analysis

4.2.1 Preprocessing

Preprocessing is performed using the spectra captured as described in the preceding section. The preprocessing pipeline applies four sequential operations to the spectra: temporal averaging, reflectance conversion, smoothing and differentiation, and normalization. Figure 6 shows the resulting spectra after each processing step.

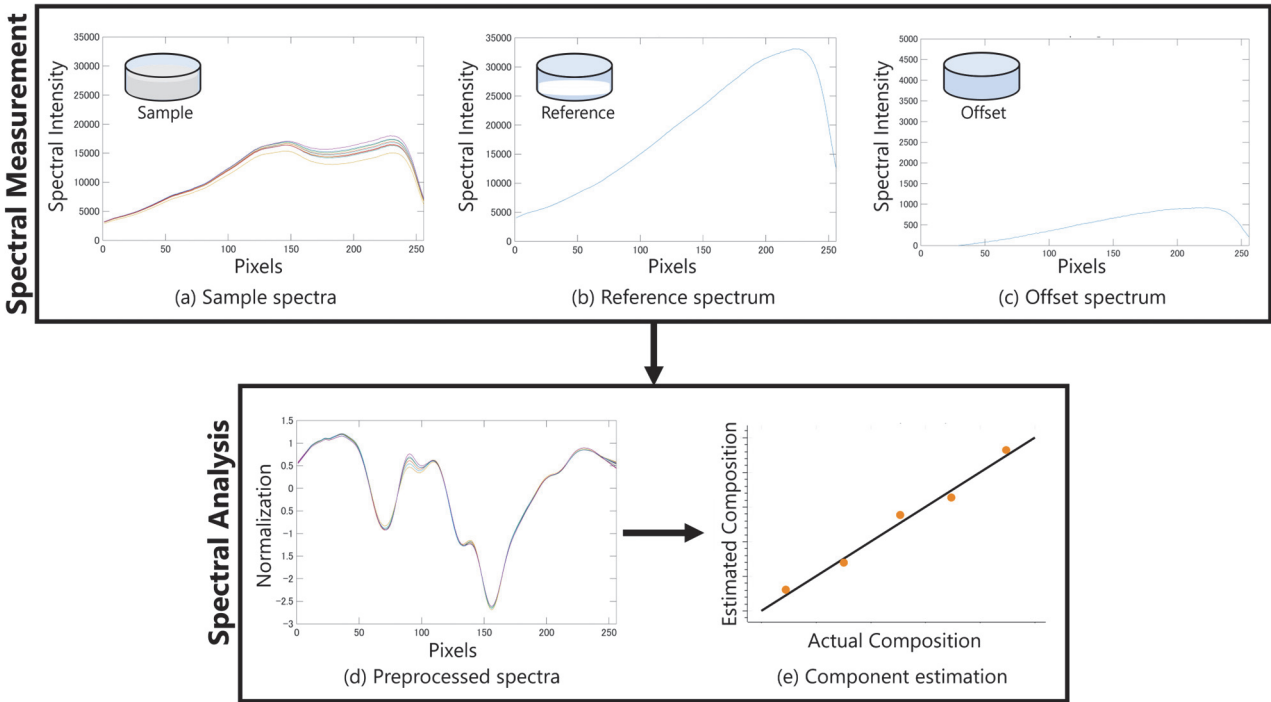


Figure 5 Spectral Measurement and Analysis Flow Sequence

First, multiple captured spectra are averaged (Figure 6(a)), which reduces random noise attributable to factors such as sensor sensitivity.

Next, conversion to reflectance R (Figure 6(b)) is performed using Eq. (1),

$$R = \frac{I_S - I_O}{I_R - I_O} \quad (1)$$

where, I_S , I_R , and I_O are the sample spectrum, reference spectrum, and offset spectrum, respectively. Converting the spectra to reflectance compensates for spectral variability arising from unit-to-unit variation in the light source and sensor.

Next, smoothing and differentiation are applied (Figure 6(c)). Smoothing reduces noise, while differentiation extracts spectral peaks. Both operations can be performed simultaneously using the Savitzky–Golay method⁽⁶⁾, in which a polynomial is fitted within a defined window and the central value of the fitted polynomial replaces the data within that window. Smoothing is achieved over the full spectrum by applying this operation repeatedly while sliding the window across the entire spectral range. Furthermore, differentiation can also be performed by analytically differentiating the fitted polynomial.

Finally, normalization is applied (Figure 6(d)), which compensates for spectral drift caused by factors such as temperature variation. The normalized spectrum X is computed using Eq. (2),

$$X = \frac{I - \mu}{\sigma} \quad (2)$$

where, I is the differentiated spectral data, μ is the mean, and σ is the standard deviation. All spectral data are standardized to zero mean and unit standard deviation using normalization. As shown in the enlarged insert in Figure 6(d), preprocessing enhances peak variations attributable to differences in sample composition.

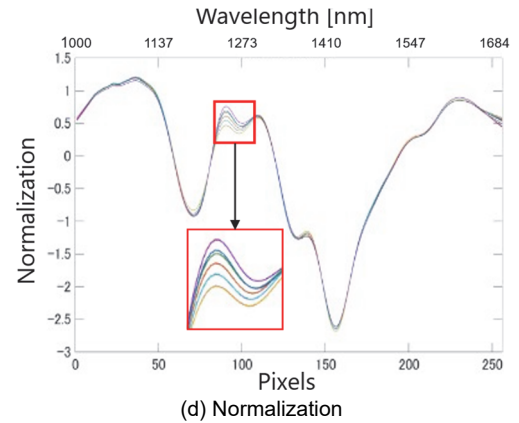
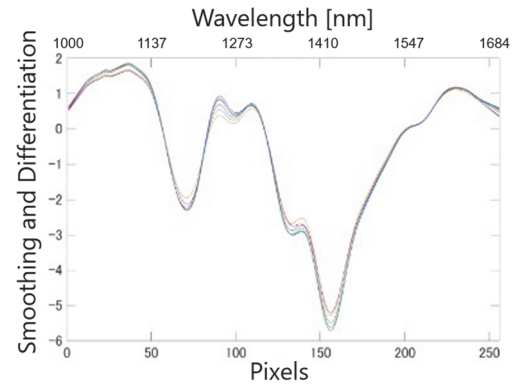
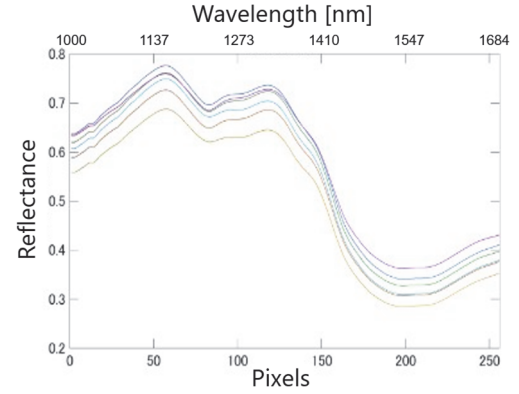
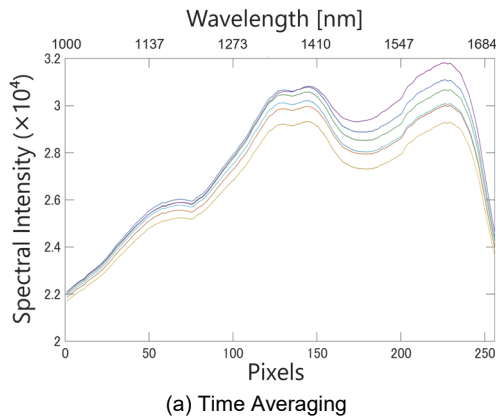


Figure 6 Preprocessing Spectral Conversion

4.2.2 Composition Estimation

At component estimation, calibration data is created from the preprocessed spectral data. The calibration data B satisfies the following multiple regression equation,

$$y = XB + D \quad (3)$$

where, y is the component matrix, X is the standardized spectral data obtained from Eq. (2), and D is the residual. The calibration data B can be analytically determined under the condition that the residual D is minimized. However, since the spectra captured by near-infrared spectroscopy exhibit multicollinearity with high correlation between wavelengths, component estimation becomes unstable with calibration data based on Eq. (3).

Therefore, we used PLS, which creates new variables (latent variables) from the preprocessed data that contribute to component estimation. This method stabilizes component estimation.

Using PLS, $\mathbf{X}$ and $\mathbf{y}$ in Eq. (3) are expressed by the following equations,

$$\mathbf{X} = \sum_{n=1}^N \mathbf{t}_n \mathbf{P}_n^T + \mathbf{E} = \mathbf{T} \mathbf{P}^T + \mathbf{E} \quad (4)$$

$$\mathbf{y} = \sum_{n=1}^N \mathbf{t}_n \mathbf{q}_n + \mathbf{f} = \mathbf{T} \mathbf{q} + \mathbf{f} \quad (5)$$

where, N is the number of principal components, $\mathbf{t}_n$ is the n -th principal component, $\mathbf{T}$ is the latent variable, $\mathbf{P}$ and $\mathbf{q}$ are the loading matrices of $\mathbf{X}$ and $\mathbf{y}$, respectively, and $\mathbf{E}$ and $\mathbf{f}$ are the residuals of $\mathbf{X}$ and $\mathbf{y}$, respectively. Also, the superscript T denotes transpose of a matrix. The latent variable $\mathbf{T}$ is determined so that the covariance with the component matrix $\mathbf{y}$ is maximized as,

$$\mathbf{T} = \frac{\mathbf{X} \mathbf{X}^T \mathbf{y}}{|\mathbf{X}^T \mathbf{y}|} = \mathbf{X} \mathbf{W} \quad (6)$$

where, $\mathbf{W}$ is the weight vector. $\mathbf{P}$ and $\mathbf{q}$ are analytically determined under the condition that $\mathbf{E}$ and $\mathbf{f}$ in Eq. (4) and (5) are minimized. Using these, the calibration data $\mathbf{B}$ when using latent variables is expressed as,

$$\mathbf{B} = \mathbf{W} (\mathbf{P}^T \mathbf{W})^{-1} \mathbf{q} \quad (7)$$

Figure 7 shows the calibration data created using the preprocessed spectra from Figure 6. The component can be estimated by calculating the inner product of this calibration data and the preprocessed spectra.

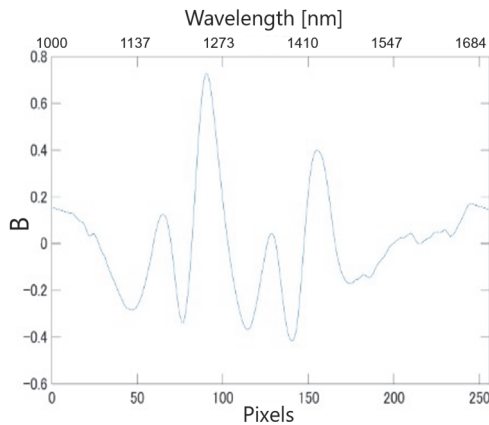


Figure 7 Calibration Data

4.2.3 Evaluation of Calibration Data

Before applying the created calibration data to an actual pharmaceutical line, it is necessary to evaluate the component estimation accuracy. Therefore, the preprocessed

spectra are split into training data and test data. Next, calibration data is created using the training data, and the component estimation accuracy is evaluated. Finally, the test data component estimation accuracy is evaluated using the same calibration data.

There are two indicators for evaluating the component estimation accuracy of the calibration data. The first is the coefficient of determination R^2 , which is calculated by the following equation,

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - y'_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (8)$$

where, n is the number of samples, y_i is the actual component, y'_i is the estimated component, and $\bar{y}$ is the mean value of the actual components. The closer the coefficient of determination is to 1, the higher the estimation accuracy. Figure 8 shows the change in the coefficient of determination with respect to the number of principal components. It is evident that increasing the number of principal components brings the coefficient of determination closer to 1, but caution is needed because too many principal components cause overfitting, where accuracy improves for training data while accuracy decreases for test data. In practice, parameters (such as the number of principal components and the order of differentiation) are adjusted so that the coefficient of determination approaches 1 for both the training data and the test data.

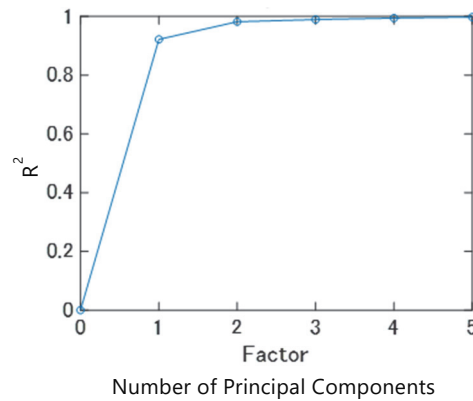


Figure 8 Change in Coefficient of Determination R^2 with Number of Principal Components

Another indicator is the Root Mean Square Error (RMSE), which is calculated by the following equation,

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - y'_i)^2} \quad (9)$$

where, n , y_i , and y'_i are the same as defined above. RMSE

is an indicator showing how much error the estimated concentration has relative to the actual concentration.

5 Evaluation of Component Estimation Accuracy

To evaluate the component estimation accuracy of the developed module, samples were prepared by mixing three types of additives, and concentrations were estimated. In addition, an automated verification device was fabricated to efficiently measure all samples.

5.1 Samples

The three additives were lactose, Ceolus (Asahi Kasei Corporation), and calcium stearate, which are widely used in pharmaceuticals. In this evaluation, the calcium stearate concentration was estimated. For this purpose, ten samples with different concentrations (0%, 1%, 2%, 3%, 3.5%, 4%, 5%, 6%, 8%, and 10%) of calcium stearate were prepared. The concentrations of the other two additives were kept constant across all samples so as not to affect the measurements. The samples were transferred to Petri dishes at a thickness of approximately 1 cm.

5.2 Automated Verification Device

Figure 9 shows the appearance of the automated verification device used to measure the samples. The rotating table at the top of the device has 12 equally spaced holes. A total of 12 Petri dishes is placed above these holes: 10 containing samples, one containing a standard reflector for reference spectrum measurement, and one empty dish for offset spectrum measurement. The developed near-infrared spectroscopy module is mounted under the rotating table to measure spectra from the bottom surface of the Petri dishes. Since the automated verification device has a built-in rotary stage, the 12 samples were measured continuously.

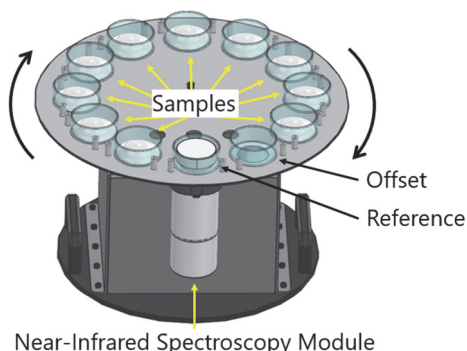


Figure 9 External View of Automated Verification Device

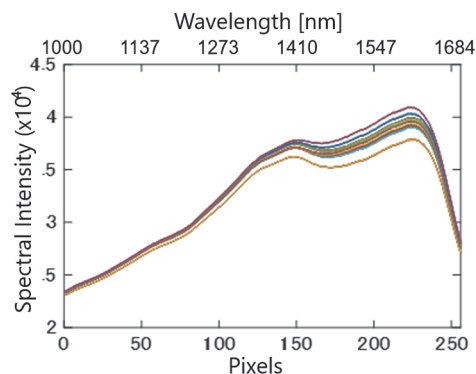
5.3 Measurement Conditions

Both the standard wavelength and long-wavelength

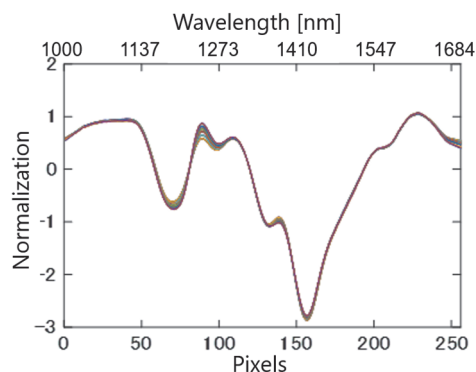
models were used for the measurements. The number of spectra measurements was set to 800 per sample. The integration time was set to 1280 μs for the standard wavelength model and 1270 μs for the long-wavelength model. These values were set so that the maximum value of the reference spectrum would be 80% of the maximum intensity of the sensor.

In addition, to determine any spectral variation due to sample position, spectra were measured over a wide sample area while rotating the stage. The number of time averages per sample was set to 80. As a result, the number of data points after time averaging was 10 per sample. Furthermore, the smoothing window size was set to 13 and the order of differentiation was set to 1.

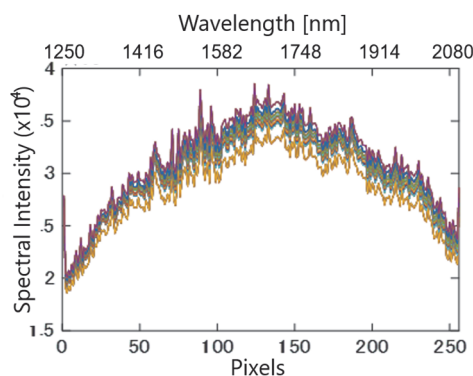
Six samples of 0%, 2%, 4%, 6%, 8%, and 10% were used as training data, and four samples of 1%, 3%, 3.5%, and 5% were used as test data. Figure 10 shows raw and preprocessed spectra measured by both models. The raw spectra of the long-wavelength model in Figure 10(c) show large intensity variations between pixels due to the sensitivity characteristics of the sensor, but the influence of these sensitivity characteristics is removed by preprocessing, as shown in Figure 10(d).



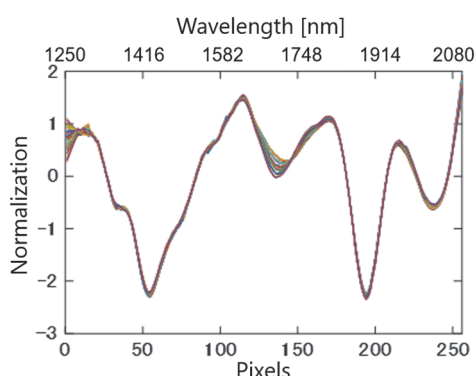
(a) Raw Spectra (Standard Wavelength Model)



(b) Preprocessed Spectra (Standard Wavelength Model)



(c) Raw Spectra (Long-Wavelength Model)



(d) Preprocessed Spectra (Long-Wavelength Model)

Figure 10 Raw and Preprocessed Spectra Measured by Standard and Long-Wavelength Models

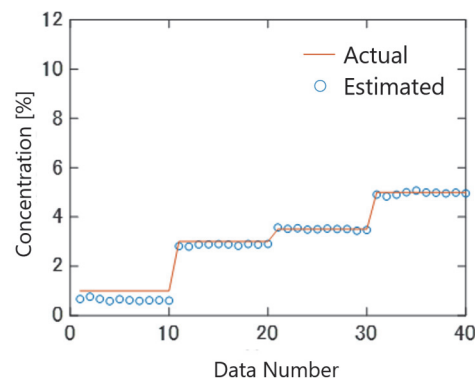
5.4 Concentration Estimation Results

Figure 11 shows the estimated calcium stearate concentration results. Figures 11(a) and 11(b) show the results for the standard and long-wavelength models, respectively. The y-axis is concentration, the x-axis is the data number after time averaging; the orange line indicates actual concentration, and the blue circles indicate estimated concentration. The concentration errors (RMSE) for the standard wavelength model and the long-wavelength model were 0.20% and 0.24%, respectively, demonstrating high estimated concentration accuracy. In addition, the estimation speed was several times per second, facilitating real-time concentration estimation.

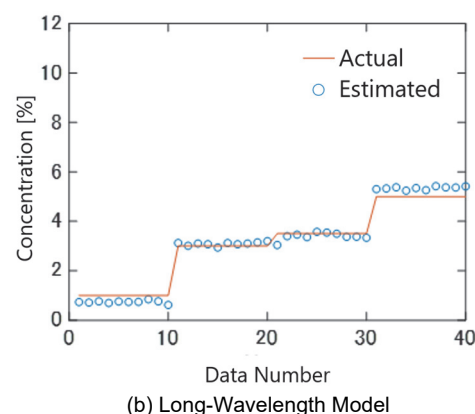
6 Conclusions

This study focused on near-infrared spectroscopy among PAT technologies to develop a near-infrared spectroscopy module that can be easily integrated into pharmaceutical manufacturing lines. A compact and lightweight module with high resistance to vibration was achieved by using an LVF for spectroscopy. Evaluation using common pharmaceutical additives demonstrated that the developed module has high

estimated concentration accuracy and can estimate concentration in real-time. These results show the developed module has practical applications in pharmaceutical component analysis on continuous pharmaceutical lines.



(a) Standard Wavelength Model



(b) Long-Wavelength Model

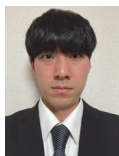
Figure 11 Estimated Calcium Stearate Concentration Results

References

- 1) Wu S, Liu Y, Fan X, Shen Y, Qu H. Trends and new process analytical technologies in pharmaceutical manufacturing. *Int J Pharm.* 2025 Sep 15;682:125957.
- 2) Yukihiro Ozaki, "Near-Infrared Spectroscopy," Spectroscopy Series, vol. 2, Kodansha, 2015 (in Japanese).
- 3) Sierra-Vega NO, Karry KM, Románach RJ, Méndez R. Monitoring of high-load dose formulations based on co-processed and non co-processed excipients. *Int J Pharm.* 2021 Sep 5;606:120910.
- 4) Kiricenko K, Klinken S, Kleinebudde P. Proof of a LOD prediction model with orthogonal PAT methods in continuous wet granulation and drying. *J Pharm Sci.* 2025 Jan;114(1):176-184.
- 5) Rosipal, R., Krämer, N. (2006). Overview and Recent Advances in Partial Least Squares. In: Saunders, C., Grobelnik, M., Gunn, S., Shawe-Taylor, J. (eds) Subspace, Latent Structure and Feature Selection. SLSFS 2005. Lecture Notes in Computer Science, vol 3940. Springer, Berlin, Heidelberg.

- 6) Savitzky, A., & Golay, M. J. E. (1964). Smoothing and differentiation of data by simplified least squares procedures. *Analytical Chemistry*, 36(8), 1627–1639.

Authors



Daisuke Oida
1st Development Dept.
Development Div.
Sensing & Devices Company



Noriaki Yamasaki
1st Development Dept.
Development Div.
Sensing & Devices Company



Masaru Koshihara
1st Development Dept.
Development Div.
Sensing & Devices Company



Takanori Saitoh
1st Development Dept.
Development Div.
Sensing & Devices Company

Development of X-ray Inspection System for Saving Labor and Effort

Tatsuki Komatsubara, Naoya Saito, Ryotaro Kitabayashi, Takashi Kanai, Kazunori Nakayama

[Summary]

By using the latest detector and image processing technologies, we developed an X-ray inspection system designed to minimize labor and operational effort. A high-resolution imaging-based inspection algorithm reduces the false detection rate to one-seventh of conventional systems, helping improve yields. High-quality inspection reduces the need for re-inspection and rework, significantly lowering the operational burden on users. In addition, by increasing automatic configuration, we successfully reduced the man-hours required for inspection setup. 1-Pass Product Setup enables faster foreign-object inspection configuration in one test run, minimizing required preparation and more efficient inspection start-up.

1 Introduction

It has been 10 years since Anritsu first developed the XR75 series X-ray inspection systems currently in commercial use¹⁾. Over this period, Anritsu has developed a diverse range of systems to address a broad spectrum of inspection targets and market needs, including systems specialized for sealing checks and those designed primarily for the meat-processing industry.

The regulatory landscape surrounding the food industry has also undergone significant changes over the past decade. The partial amendment to the Food Sanitation Act, promulgated in 2018, made HACCP-based sanitation management mandatory under the Food Sanitation Act in 2021. Furthermore, the transitional period for the positive list system governing food-use utensils and container packaging expired in May 2025, reflecting the steady year-on-year advancement of regulatory frameworks aimed at ensuring food safety. As these regulatory developments illustrate, food safety is a theme of enduring importance that transcends decades, and interest in the topic has been growing in recent years.

At the same time, labor shortages have long been a serious challenge in the food manufacturing industry, and the situation has grown increasingly severe in recent years with the added pressures of rising raw material and labor costs. In this environment, manufacturers must simultaneously maintain rigorous quality control and stable production operations, placing a significant burden on the shop floor. In fact, labor productivity in the food manufacturing industry remains at approximately 60% of the manufacturing sector average, making it one of the lowest-performing industries

in the sector. A key contributing factor is the high degree of dependence on manual labor due to insufficient automation²⁾.

In response to these challenges, Anritsu has developed a system that enables highly accurate inspection with ease, regardless of the operator's skill level.

Advances in X-ray detector performance and the evolution of our proprietary image processing technology have achieved up to about 40% improvement in inspection capability for stainless steel materials, enabling the detection of previously difficult-to-detect contaminants, such as small metallic foreign objects and low-density substances like stones. Significant cost savings in complaint handling are expected by reducing foreign-object contamination complaints. The false detection rate has also been cut to one-seventh that of conventional systems, helping reduce food waste and improving yields. Furthermore, the system still delivers stable inspection quality for foreign objects that were already detectable by conventional systems.

The improved inspection performance has created greater margin for adjusting inspection parameters. Leveraging this advantage, the system is designed to maintain high inspection quality independent of operator skill level by automating the majority of these adjustments. As a result, the system achieves both reduced operational burden on the production floor and improved yields.

Specific features include 1-Pass Product Setup and detection limit configuration for reducing operator workload, as well as improved operability through an enhanced touchscreen interface. This article overviews the system.



Figure 1 External View of X-ray Inspection System

2 Principles of X-ray Inspection

An X-ray inspection system is a device that uses X-rays to non-destructively examine the internal condition of products. Anritsu X-ray inspection systems are used primarily on food and other manufacturing lines to detect foreign-object contamination and product defects in inspected items. These systems are installed inline on food production lines, enabling continuous sequential inspection without interrupting the production flow. The X-ray inspection system is configured so that products under inspection pass through the interior of the unit while X-ray-based inspection is performed.

X-rays are a part of the electromagnetic spectrum known for their high directionality and penetrating power. X-rays emitted from the source pass through the product under inspection and are detected by an X-ray detector on the opposite side. Since the amount of X-rays passing through the product depends on the product internal structure, the intensity of the detected transmitted X-rays is represented as gradations of light and dark, enabling capture of information such as the shape and internal structure of the inspected product. The amount of passing X-rays varies with the type and density of the material. As a result, areas with a foreign object show a clear difference in image contrast, appearing as shadows. The presence or absence of foreign objects is determined based on these contrast variations.

In X-ray inspection, the detector is a critical component directly involved in output of inspection images, and its

performance is directly linked to the accuracy of inspection results. An X-ray detector consists of a scintillator, a photodiode array, and a readout circuit, which detect X-rays as follows. First, X-rays incident on the detector are absorbed by the scintillator. Electrons in the scintillator are excited by the absorbed energy and transition to a higher energy state. As the electrons subsequently return to their ground state, they emit visible light corresponding to the energy difference. Visible light converted from X-rays in this manner can then be detected by the photodiode array. This visible light is then converted into electrical charge by the photodiodes and accumulated in internal capacitors. Finally, the accumulated charge is sequentially read by the readout circuit, enabling X-ray detection and outputting differences in intensity as an image.

The method of detecting X-rays by first converting them into visible light using a scintillator and subsequently converting that light into electrical charge is called the indirect conversion method. This approach is the predominant technology for X-ray detectors used in X-ray inspection systems, owing to factors such as the availability of materials. However, since this method involves two conversion stages, signal loss occurs during the process, which tends to reduce the clarity of the output image. To achieve higher sensitivity and improved stability in X-ray inspection systems, careful selection of components such as the scintillator, as well as optimized system design are critical, and these technical considerations determine the direction of system development.

3 Boosting Production Efficiency with High-Speed Image Processing

We have successfully developed an X-ray inspection system capable of performing inspections using low-noise, high-resolution images by the comprehensive redesign of the X-ray detector's scintillator, photodiode array, readout circuit, and X-ray optical system. Use of higher-definition images enables identification of minute foreign objects that were previously difficult to detect. In practice, verification of detection performance at a conveyor speed of 40 m/min for a specific food product confirmed that while the conventional model can detect stainless steel wire down to $\phi 0.6$ mm, the XR76 can detect stainless steel wire down to $\phi 0.2$ mm.

At the same time, challenges exist in achieving higher

resolution in X-ray detectors. As the pixel count increases, the volume of image data grows larger, resulting in an increased computational load for image processing. Consequently, processing time is extended, making it more difficult to complete foreign-object determination while the product is passing through the system. One approach to this challenge is to increase the system length to secure additional processing time. However, this results in a larger overall system footprint, requiring greater installation space. Furthermore, as the conveyor speed of the product under inspection increases, the volume of image data to be processed grows accordingly, imposing increasingly stringent constraints on processing time.

The conventional XR75 series intentionally reduces acquired data size to lighten the processing load when the conveyor speed exceeds 40 m/min before inspection is performed. While this approach enables inspection at high conveyor speeds without extending the machine length, it causes image compression that significantly reduces foreign-object detection accuracy compared to a lower conveyor speed.

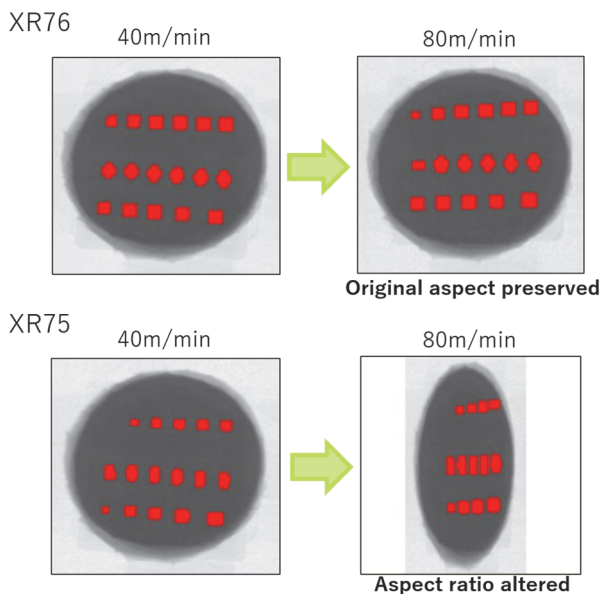


Figure 2 Image Comparison

In contrast, the XR76 achieves high-speed processing of high-pixel-count images by improvements to both the software and the foreign-object detection algorithm, as well as to the enhanced X-ray detector. This enables inspection without image compression even under high-speed inspection conditions, achieving stable foreign-object inspection using high-quality images on high-speed production lines.

Figure 2 shows test pieces placed on a ham sample and

conveyed through the X-ray inspection system. The foreign objects, from top to bottom, are stainless steel (SUS) balls, SUS wire, and ceramic balls. Using the conventional XR75 at a conveyor speed of 80 m/min, the image is compressed in the conveyor direction, and the number of detected SUS balls decreases from five to four. In contrast, the new XR76 exhibits no image compression even at a conveyor speed of 80 m/min, and detects all six SUS balls at high-speed conveying without any drop in detection accuracy. The system is capable of high-sensitivity inspection on production lines with high conveyor speeds, contributing to improved production efficiency and fewer customer complaints.

4 Cutting Labor and Saving Effort by Automation and Improved Operability

4.1 Cutting Operator Dependency at Inspection Setup

In line with recent trends toward labor reduction, the system incorporates many features designed to minimize operator dependency, enabling even users unfamiliar with X-ray inspection systems to configure foreign-object inspection settings with ease.

As shown in Figure 3, conventional inspection systems require several preparatory steps before foreign-object inspection can start, including determining the X-ray output level, feeding uncontaminated products (OK products) about ten times, and fine-tuning parameters to reduce false detections. These tasks are cumbersome, and many aspects, including the degree of parameter adjustment, depend heavily on user operation.

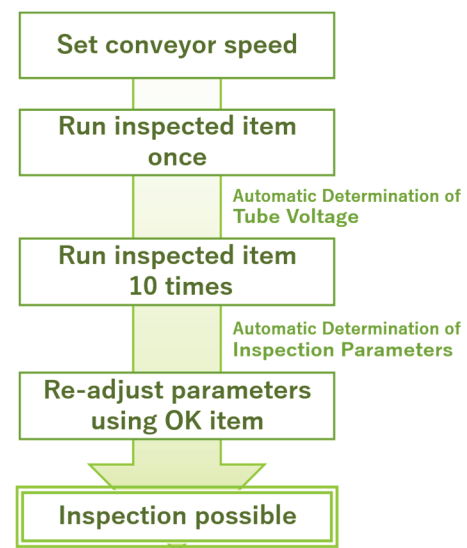


Figure 3 Preparations Before Starting Inspection

The improved inspection stability of the XR76 resulting from fewer false detections through enhanced sensitivity creates a greater margin in configuring inspection parameters, supporting more automated settings. The system has several dedicated features (Table 1) to minimize operator-dependent factors and allow anyone to begin inspection with ease.

Table 1 Comparison of Inspection Modes

	Conventional: Sensitivity- Focused	Detection Limit Setting Mode: Balance	1-Pass Product Setup
Screen Transitions	12	12	5
Tube Voltage Setting	Yes	Yes	No
Passing Test Product	10 times recommended	10 times recommended	Once
Readjusting Detection Parameters	Comparable to conventional models	Few adjustments	Unnecessary
Performance	Maximum device sensitivity	Sensitivity $\geq$ conventional model	Foreign-object inspection only*

* Requires confirmation that it meets management sensitivity requirements

4.1.1 Fewer Parameter Adjustments

To reduce false positives, it is typically necessary to register a product variant using actual inspection targets and then re-adjust the inspection parameters accordingly. At this stage, applying parameters with sufficient margins can reduce false detections caused by variations in the shape, condition, and overlap of the inspection targets. However, this makes it more difficult to discriminate between the inspection target and foreign objects, resulting in lower detection accuracy. Current practice requires accumulated experience and domain expertise in order to fine-tune the parameters appropriately to achieve accurate inspection while suppressing false detections.

The XR76 “Detection Limit Setting Mode: Balance” function cuts the number of manual parameter re-adjustments required after automatic configuration by applying parameter settings with sufficient margins that minimize false detections. However, under normal circumstances, such a configuration approach would contribute to lower inspection accuracy. However, advances in high-resolution detectors and image processing technology can capture the presence of foreign items with greater clarity to achieve inspection accuracy equal to or exceeding that obtained through fine-tuned

adjustments on conventional machines. Furthermore, the XR76 system displays a One-Touch Adjustment button on the parameter re-adjustment screen. Whereas it was previously necessary to manually re-adjust multiple parameters individually, pressing this button allows simultaneous re-adjustment of multiple parameters, enabling any operator to commence inspection with settings that have sufficient margins.

4.1.2 Simplified Inspection Setup

In conventional inspection machines, parameters such as noise reduction values used to detect foreign objects are determined by passing the product to be inspected through the system about ten times while the product X-ray images are analyzed internally using multi-faceted image processing algorithms to extract product characteristics.

In contrast, the new 1-Pass Product Setup, which is limited to use for detecting foreign objects, automatically configures parameters to minimize false detections by passing an inspection target through the system just once. Since inspection accuracy is affected by the inspection target characteristics, separate verification that the required control sensitivity is satisfied remains necessary. However, in addition to detection limit adjustment, no X-ray output configuration is required, and since nearly all settings can be configured automatically with a single pass of the inspection target, preparations can be completed with minimum necessary effort.

4.2 Improved Operability

Since smartphones have become an indispensable part of everyday life, the design of any user interface (UI) must consider the operational feel and functionality of smartphones that users are accustomed to. The XR76 achieves operability and touch sensitivity comparable to that of smartphones by overhauling both the touchscreen panel integrated into the display and the associated software. Specifically, intuitive gesture-based operations are now supported, including scroll functionality via swiping and pinch-in/pinch-out zooming of X-ray transmission images, providing users with a more comfortable and accessible operating environment.

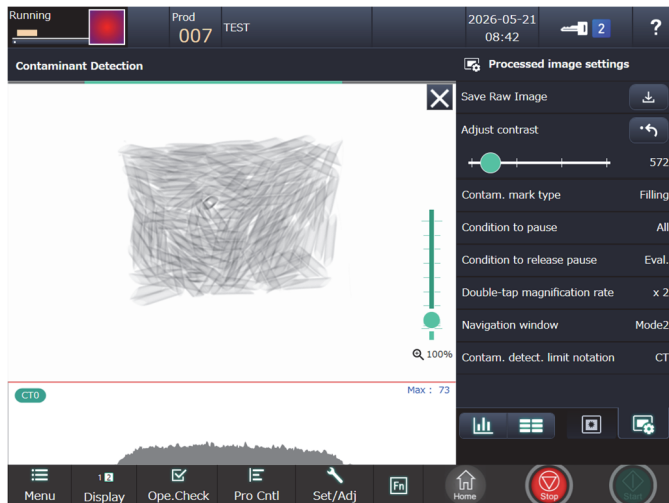


Figure 4 Configuration Screen for X-ray Transmission Image Parameters at Inspection Interface

Figure 4 shows the screen on the inspection display for configuring transmission image conditions. Efficient adjustment of display settings is achieved with the fewest taps by consolidating the most frequently used parameters for transmission image verification at a single screen. Furthermore, minimizing screen transitions and structuring the interface helps users access the desired settings without becoming disoriented by navigation changes. At the same time, the fundamental operational framework, including button layout, inherits the UI of existing systems, ensuring that users are not required to relearn operating procedures from scratch. This allows users who have previously operated conventional machines to transition to the new system with ease.

4.3 Long-Life Design with Less Maintenance

Optimized internal airflow through advanced thermal design, along with technology to extend the X-ray tank service life via low-power operation, which are both inherited from previous developments, are carried over into the XR76 series. Three primary components contribute to extended service life: the X-ray detector, X-ray source, and specially sealed heat dissipation unit.

Using high-sensitivity X-ray detectors reduces the output of the X-ray tank, successfully minimizing wear on both the X-ray tank and the X-ray detector. A cooler-free, specially sealed enclosure developed using fluid dynamics simulation reduces power consumption. Furthermore, the failure rate of key electrical components is also reduced by the structural design preventing ingress of outside air.

These technologies have achieved an XR76 design service

life of 30,000 hours, creating a system with low replacement frequency due to end-of-life or failure, reduced running costs, and compatibility with labor-saving operations. As a demonstrated track record, the XR75 series X-ray tank and X-ray detector service life using this technology has been extended to more than 2.5 times that of preceding series. Since this figure includes units replaced before end-of-life as part of preventive maintenance, the actual service life is estimated to be even longer.

5 Main Specifications

Figure 5 shows the XR76 dimensional drawings, and Table 2 lists the main specifications.

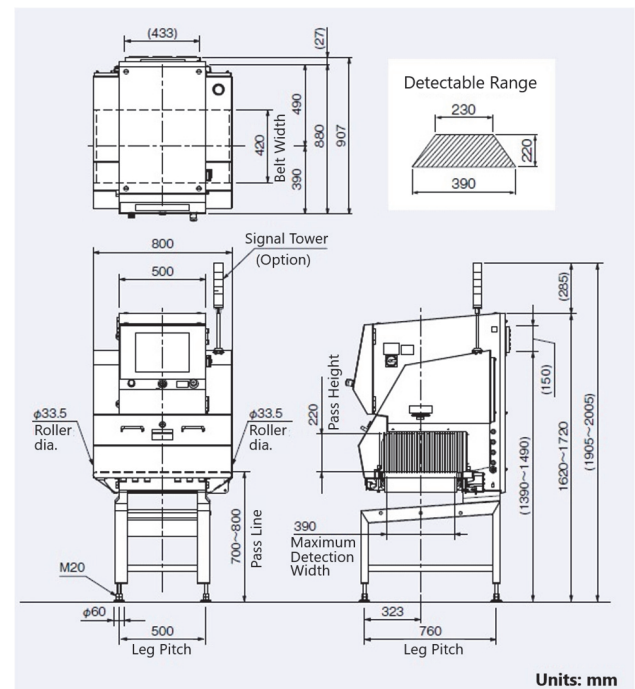


Figure 5 External Dimensions

6 Conclusions

Amid growing public concern over food safety, food processing facilities are facing an increasingly serious array of challenges, including labor shortages and rising material and labor costs. To address these challenges, we developed a new X-ray inspection machine with a focus on personnel reduction and labor savings. Significant labor savings are anticipated through fewer customer complaints and re-inspections by enabling high-sensitivity, low false-detection inspection using the latest X-ray detector and image processing technologies. For customers in high-mix production, features such as 1-Pass Product Setup greatly reduce the

configuration effort required to detect foreign objects, thereby cutting the operator burden and facilitating both personnel reduction and labor savings.

This XR76 model is designed for personnel reduction and labor savings on the production floor while simultaneously ensuring food quality and delivering safe and reliable food products to consumers. Anritsu is continuing to develop inspection machines that bring operational ease and peace of mind to production environments through enhanced inspection efficiency, thereby helping build a sustainable society.

References

- 1) M. Yagi, H. Koba, T. Kikuchi, N. Saito, M. Inoue, K. Nagatsuka, S. Mitani, "Development of XR75 X-ray Inspection System," Anritsu Technical Review No. 24, pp.100-106 (Sep. 2016).
- 2) Ministry of Agriculture, Forestry and Fisheries (MAFF), Japan, "Vision for Overcoming Labor Shortages in the Food Manufacturing Industry," 2019, p. 8 (in Japanese).

Authors



Tatsuki Komatsubara
Product Development Dept.
Development Div.
Infivis Company



Naoya Saito
Product Development Dept.
Development Div.
Infivis Company



Ryotaro Kitabayashi
Product Development Dept.
Development Div.
Infivis Company



Takashi Kanai
Product Development Dept.
Development Div.
Infivis Company



Kazunori Nakayama
Product Development Dept.
Development Div.
Infivis Company

Table 2 Main Specifications

Model	KXS7622AVCTV	KXS7634AVCTV
X-ray Output	Tube Voltage 30 to 80 kV, Tube Current 0.4 to 3.3 mA, 100 W max.	
Safety	1.0 μSv/h max. X-ray leakage prevention by safety device	
Display	15-inch capacitance-type LCD with touch panel	
Detection Area	240 mm max. width, 120 mm max. height	390 mm max. width, 220 mm max height
Belt Width	270 mm	420 mm
Preset memory	200 or 999 products (switchable)	
Conveyor Speed/ Performance	10 to 60 m/min, 5 kg max.	10 to 60 m/min, 5 kg max.
	61 to 90 m/min, 2 kg max.	
Power requirements	100 to 240 Vac, Single Phase, 50/60 Hz, 700 VA max., ±10% permissible power fluctuation	
Mass	230 kg	290 kg
Environmental conditions	Temperature 0° to 35°C, Humidity 30% to 85% (no condensation)	
Protection class	IP66-equivalent	
Exterior	Stainless Steel (SUS304)	

Publicly available

Technical Demonstration of EV Powertrain Evaluation Combining Simulation and Real Power Supply

Takuya Yoda, Yuta Shinchi

[Summary]

As a countermeasure to global warming, the automotive sector is pursuing decarbonization through an xEV shift. To accelerate widespread adoption of Battery Electric Vehicles (BEVs), which are expected to become the future mainstream powertrain, automakers are stepping up their transition from conventional powertrains to motor-based systems. To help achieve a BEV society, we conducted a technical demonstration of Power HIL — a novel approach that integrates Hardware-in-the-Loop Simulation (HILS), a methodology widely used in automotive development, with a bidirectional DC power supply. This enables powertrain performance evaluation equivalent to that of a real vehicle as part of system-level assessment, and is expected to contribute to improved development efficiency, reduced cost, and a decrease in rework during physical vehicle evaluation. Going forward, we will advance validation using actual vehicle components with the aim of enhancing BEV development efficiency, improving quality, and expanding application to other domains.

1 Introduction

Recent climate change driven by global warming has become increasingly pronounced with Japan experiencing more frequent sudden heavy rainfall causing floods and landslips, and heightening the urgency for countermeasures. To help achieve a sustainable society, reduction of greenhouse gas emissions, especially carbon dioxide (CO₂) emissions, represents a critical challenge.

To address this challenge and promote greater efficiency in EV development, we have undertaken a technical demonstration of EV powertrain evaluation that integrates simulation with a real power supply. This article describes the background and context of the challenge in relative detail through to section 3, while section 4 and beyond present the results and status of the technical demonstration.

2 Social Challenges Driving xEV Adoption

According to a 2023 survey by Japan's Ministry of Land, Infrastructure, Transport and Tourism, the automotive sector as a whole accounts for approximately 16.5% of total domestic CO₂ emissions, underscoring the need to advance decarbonization. In the automotive market, a shift toward xEVs (x Electric Vehicles), which use electricity as a power source, is progressing with the aim of reducing CO₂ emissions. Representative vehicle types include BEVs (Battery Electric Vehicles), HEVs (Hybrid Electric Vehicles), PHEVs (Plug-in Hybrid Electric Vehicles), and FCEVs (Fuel Cell

Electric Vehicles). The government has established targets to promote widespread adoption of xEVs, and individual automakers are actively engaged in technology development, establishment of production methodologies, and building of necessary organizational frameworks. While significant future growth in BEV adoption is anticipated, this trajectory is accompanied by a range of social and technical challenges, including development of charging infrastructure, reduction of battery charging time, and extension of driving range.

3 BEV Development Challenges

3.1 Model-Based Development and Simulation Technology

In automotive development, MBD (Model-Based Development) — a development methodology that leverages simulation technology for the purpose of improving efficiency and quality — has been widely adopted. Behavior can be verified at the design stage by modeling control specifications and the structural characteristics of automotive components and subjecting them to simulation.

As shown in Figure 1, the MBD process can be represented by a V-model spanning requirements definition to physical vehicle evaluation. The primary simulation methodologies at each respective stage are MILS (Model-in-the-Loop Simulation), SILS (Software-in-the-Loop Simulation), and HILS (Hardware-in-the-Loop Simulation).

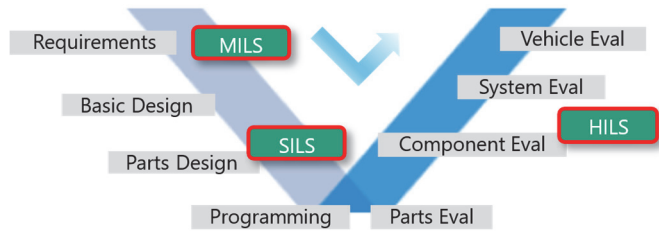


Figure 1 MBD V-type Model

Table 1 lists each method.

Table 1 Simulation Methods

Name	Outline
MILS Model-in-the-Loop Simulation	Verify control logic without physical hardware, using a model-based approach
SILS Software-in-the-Loop Simulation	Verify control algorithms using the implemented code in a PC virtual environment
HILS Hardware-in-the-Loop Simulation	Verify operation using hardware (ECU) and dedicated platform device

Among these methods, HILS plays a particularly critical role. A dedicated platform device serving as the simulation execution environment (HIL simulator hereafter) is connected via wiring harnesses to multiple ECUs (Electronic Control Units) mounted in the vehicle, enabling exchange of control signals, including I/O, CAN, and Ethernet, based on simulation results. This supports verification of complex cooperative behavior between ECUs, as well as repeated testing of abnormal conditions and failure modes that would be difficult to reproduce in a real-world environment.

However, verifying and evaluating the performance of a powertrain that handles electrical power requires continuous refinement of the simulation model.

3.2 Powertrain Evaluation

The powertrain of BEVs is transitioning from the engine and fuel (such as gasoline) used in conventional internal combustion engine vehicles to a motor system centered on motors, inverters, and batteries (Figure 2). This transition necessitates new initiatives, including revisions to performance evaluation criteria for newly introduced components.



Figure 2 Transition to Motor System

Meanwhile, competition in the automotive market, both domestically and internationally, is intensifying, creating pressure to further streamline development processes and bring new vehicle models to market earlier. However, when issues arise during powertrain evaluation after completion of a physical vehicle, the resulting rework is substantial, leading to delays in the development schedule and increased costs. To mitigate and avoid such risks, it is more effective to conduct system-level evaluation equivalent to that of a physical vehicle using real electrical power prior to actual vehicle evaluation.

3.3 Integrating Simulation and Real Power Sources

To streamline BEV development, we are working to achieve Power HIL, a powertrain performance evaluation method that integrates HILS with a bidirectional DC power supply (power supply unit hereafter). This method features a HIL simulator combining environment, vehicle, component, and control models to simulate operations such as charging and driving (acceleration, deceleration, hill climbing, rough-road driving, etc.) to control the power supply unit, to evaluate performance with power transfer equivalent to an actual vehicle applied to the motor/inverter and battery. Figure 3 shows an example of a Power HIL system configuration using a real battery.

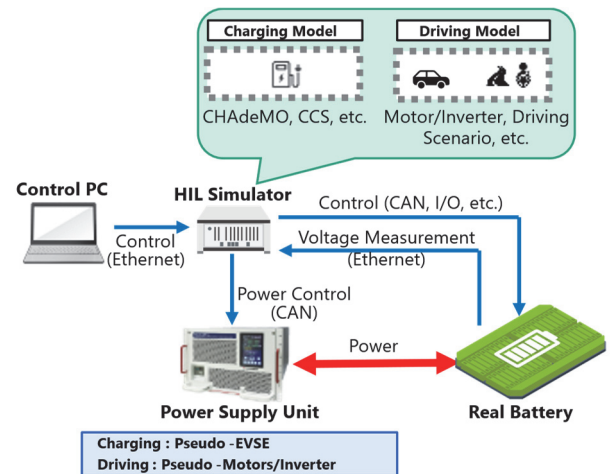


Figure 3 Power HIL System Configuration using Real Battery

Furthermore, as shown in Figure 4, the verification accuracy of MILS and SILS is improved to refine the model by feeding back evaluation data to the upstream process.

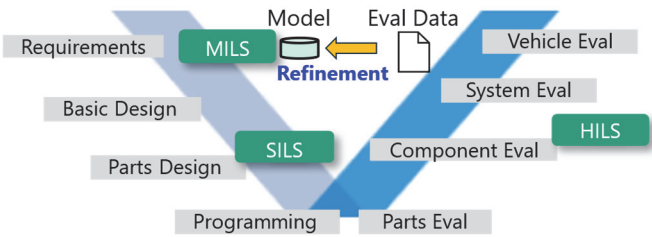


Figure 4 Refinement of the Model Using Evaluation Data

4 Technical Demonstration

In fiscal 2024, we collaborated with a domestic automobile manufacturer to demonstrate Power HIL utilizing existing equipment. The following describes this initiative.

4.1 System Outline

The system was constructed by combining a control PC (Table 4), a HIL simulator (Table 5), and a power supply unit (Table 6). Since using actual vehicle components during simulation model development carries safety risks, a power supply unit was adopted as a pseudo-battery.

Figure 5 shows the basic system configuration.

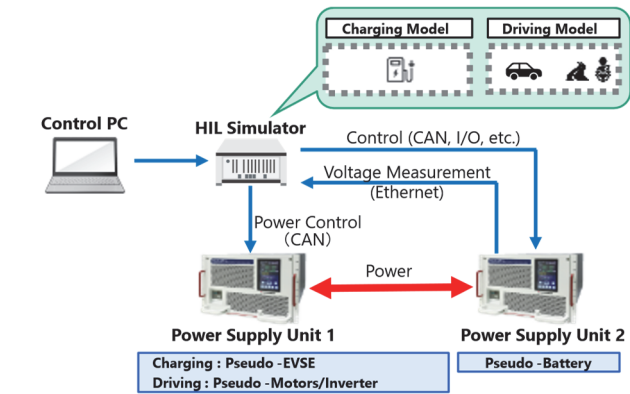


Figure 5 Basic System Configuration

The primary objectives of this demonstration were to validate the model executed on the HIL simulator and to confirm the feasibility of power supply control. With safety considerations in mind, power supply control was conducted at approximately 1/20 scale of the actual power level (i.e., the power handled in a real vehicle). Table 2 lists the components constituting the system and their respective roles.

Table 2 Equipment and Roles

Name	Function
Control PC	Controls operation of HIL simulator
HIL Simulator	Executes simulation models replicating charging and driving, and controls power supply unit
Power Supply Unit 1	Charging: Simulates charging by supplying power as EVSE Driving: Simulates motoring and regeneration through power supply and recovery
Power Supply Unit 2	Simulates battery by supplying and recovering power

The Takasago Seisakusho RZ-X2-10K was used as the power supply unit because it provides sufficient power capacity and charge/discharge functionality for the demonstration, offers high efficiency and low noise performance, excels in responsiveness and stability to rapid changes in voltage and current, and has an extensive track record among domestic automobile manufacturers.

A CAN (Controller Area Network) and Ethernet were used as the interface between the HIL simulator and the power supply unit. This demonstration conducted two evaluations: “charging simulation” (section 4.2), which evaluates battery charging performance by simulating EVSE (Electric Vehicle Supply Equipment: charging station) operation, and “driving simulation” (section 4.3), which evaluates powertrain performance by simulating motor/inverter operation during vehicle driving.

Additionally, prototype software (Power HIL Platform Software hereafter) was developed incorporating integrated control and data monitoring functions for each piece of equipment, including the HIL simulator and power supply unit. It creates simulation conditions, execution control, and status display via a GUI.

4.2 Charging Simulation

BEV charging methods include normal charging and fast charging, each with various charging standards. Table 3 lists the major fast-charging standards and regions.

Evaluating charging performance requires equipping a facility with EVSE units compatible with each standard at substantial capital investment and installation space. Therefore, charging simulation aims to reproduce each standard within a single evaluation system and assess charging performance including the battery control ECU. This demonstration selected CHAdeMO, which is widely adopted in Japan, as the target standard.

Table 3 Major Fast-Charging Standards and Regions

Charge Standard	CHAdEMO	CCS	NACS	GB/T
Region	Japan	Type 1: N. America Type 2: Europe	N. America	China

The charging was simulated by assigning one HIL simulator and one power supply unit to each of the pseudo-EVSE and pseudo-EV, following the sequence outlined in steps (1) through (4) below (Figure 6).

- (1) Start the charging sequence between the pseudo-EVSE and pseudo-EV.
- (2) Control the output of power supply unit 1 from the pseudo-EVSE in accordance with the current value commanded by the pseudo-EV in step (1).
- (3) Measure the voltage of power supply unit 2 within the pseudo-EV, and calculate the SoC (State of Charge: battery charge level) and next current command value, and notify the pseudo-EVSE via the communication in step (1).
- (4) Repeat step (3) until SoC reaches 100%.

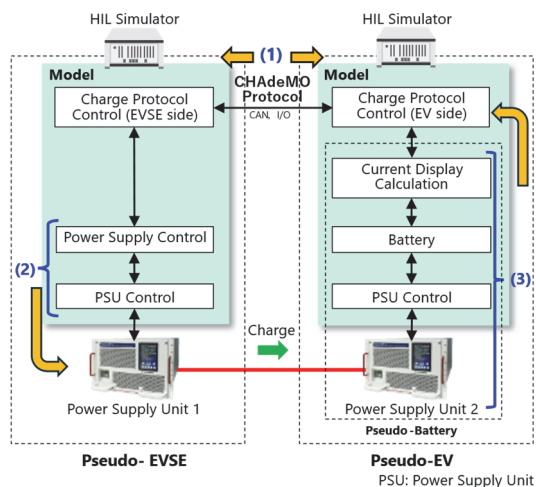


Figure 6 Demonstration System Configuration for Charging Simulation

Figure 7 shows the measurement results for the pseudo-battery SoC during the charging simulation.

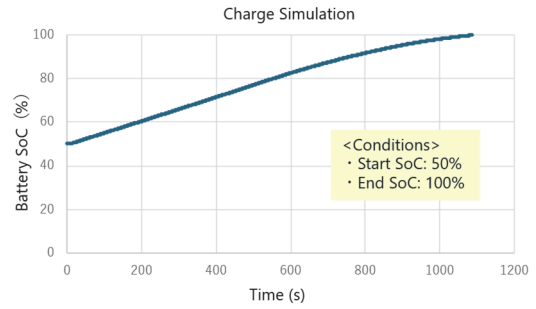


Figure 7 Measurement Results for Pseudo-Battery SoC During Charging Simulation

The results show that the pseudo-battery SoC increases as charging time elapses, demonstrating the model validity, including charging protocol control, and confirming that evaluation of battery charging performance is feasible. Furthermore, the charging monitor screen of the Power HIL Platform Software displays measurement data, such as energy and battery SoC during charging, enabling identification of normal and abnormal conditions on one screen and facilitating appropriate judgment and response (Figure 8).



Figure 8 Power HIL Platform Software Charging Monitor Screen

Although this demonstration targeted CHAdEMO, charging simulation for other charging standards can also be achieved simply by modifying the charging protocol control model and the I/O of the HIL simulator, without altering the fundamental system architecture.

4.3 Driving Simulation

During vehicle operation, a BEV transfers power between the traction motors and battery in accordance with motoring and regeneration, as governed by accelerator/brake inputs. Motoring refers to the state in which electrical energy is drawn from the battery to drive the motors during acceleration via accelerator input. Regeneration refers to the state in which the motors operate as a generator during deceleration via brake input, thereby charging the battery. Driving simulation reproduces this sequence of operations with the objective of evaluating powertrain performance under conditions closely approximating actual vehicle operation.

This demonstration adopted the WLTC (Worldwide-harmonized Light vehicles Test Cycle), an internationally recognized standard for fuel and energy consumption evaluation, as the driving pattern. The WLTC consists of four phases: low speed, medium speed, high speed, and extra-high speed, encompassing driving conditions representative of urban roads, suburban roads, and motorways.

The simulation vehicle model used was a standard model provided by the HILS vendor. Power consumption was calculated based on the motoring/regeneration of the motors/inverter model. During motoring operation, the power supply unit was controlled to output a negative value (–) to simulate discharge of the pseudo-battery.

Conversely, during regenerative operation, charging was simulated by controlling the power supply unit to output a positive value (+). The driving simulation was carried out using one HIL simulator and two power supply units, following the procedure outlined in steps (1) through (3) (Figure 9).

- (1) A standard vehicle model within the HIL simulator reproduces, via simulation, driving state under the WLTC pattern.
- (2) Calculates motoring/regeneration (–/+) of motors/inverter and controls power supply unit 1 accordingly.
- (3) Measures terminal voltage of power supply unit 2, serving as pseudo-battery, and computes SoC.

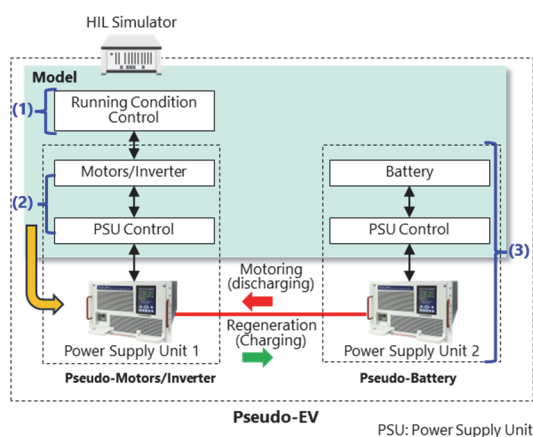


Figure 9 Driving Simulation Demonstration System Configuration

Figure 10 shows the measured vehicle speed, motoring/regeneration, and pseudo-battery SoC for each WLTC phase during driving simulation.

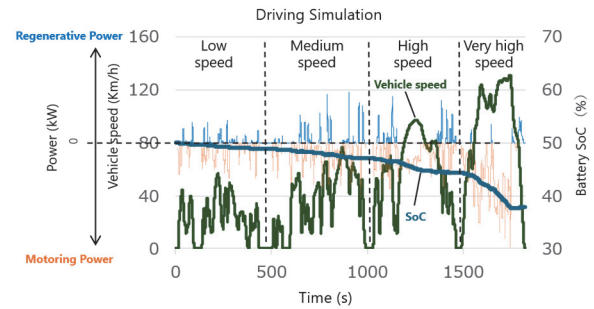


Figure 10 Measured Pseudo-Battery SoC During Driving Simulation

The results confirm that power fluctuations equivalent to those of an actual vehicle — corresponding to power consumption by motors driving in each phase and power generation by motors during deceleration — are successfully reproduced, demonstrating that powertrain performance under driving conditions can be evaluated without using a physical vehicle. This driving simulation also enables the calculation of energy consumption efficiency, a key evaluation metric. The driving monitor screen of the Power HIL Platform Software displays measured data including vehicle speed, accelerator/brake operation, power consumption associated with motoring/regeneration state of the motors/inverter, and SoC, helping the operator assess on one screen the normality or abnormality of charging and discharging during driving, and facilitating prompt and appropriate judgment and corrective action (Figure 11).

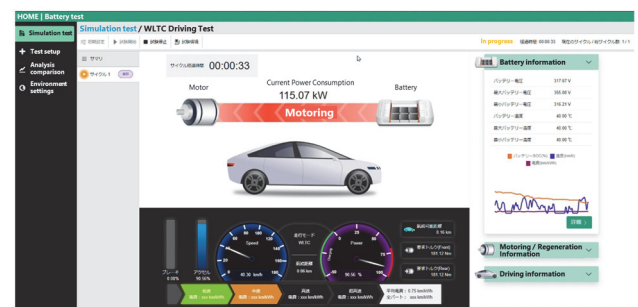


Figure 11 Power HIL Platform Software Driving Monitor Screen

This demonstration used a standard model provided by the HILS vendor as the vehicle for simulating driving. However, evaluation of powertrain performance under conditions more closely approximating actual vehicle operation is achievable by substituting a model customized to a vehicle from each automotive manufacturer.

4.4 Power System Switching

In this powertrain performance evaluation, charging simulation and driving simulation are switched alternately to

monitor the SoC transition and calculate energy consumption. However, the power transfer circuits are configured separately in each simulation, resulting in different connection points to the battery. To eliminate cable reconnection work, a junction box was prototyped (Table 7) and inserted between the power supply unit and the battery (Figure 12) to achieve seamless power-circuit switching.

Furthermore, since the power supply unit and battery are not connected directly, the junction box enables isolation even in the event of unintended current flow, thereby improving safety.

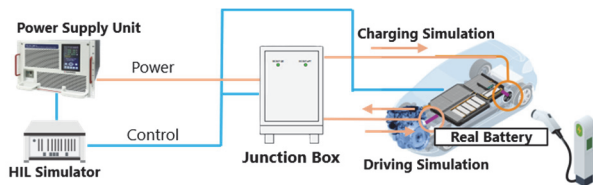


Figure 12 Power System Switching

The prototyped junction box was designed primarily to verify feasibility, including component selection and structural design needed for power circuit switching, and was about 1/20 scale of the actual power level, matching the demonstration system.

Figure 13 shows the internal structure of the junction box.

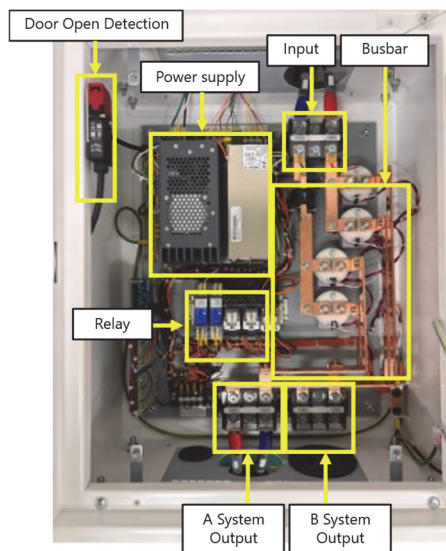


Figure 13 Internal Structure of Junction Box

Validation in the demonstration environment confirmed the adequacy of the components and circuits, as well as the system control sequence, considering safety. Adaptation to full-scale power applications is possible by replacing components with those meeting the specifications for the required applied voltage and current.

5 Future Challenges and Outlook

This demonstration confirmed the Power HIL feasibility using a motors/inverter model and pseudo-battery. The next step will be a similar demonstration using actual vehicle components to verify the validity of the models and methodology.

Going forward, the automotive market is expected to see further growth driven by widespread adoption of xEVs and fully autonomous driving. Furthermore, advances in technologies such as SDV (Software Defined Vehicle) and OTA (Over-The-Air) updates enable modification of vehicle ride quality and functionality through software, expanding efforts to sustain long-term user satisfaction even after market introduction.

At development, there is growing emphasis on improving the efficiency of design and evaluation processes, with an increasing tendency to expand use of simulation. Nevertheless, powertrain systems still require efficient performance evaluation using actual electrical power. Since this methodology is applicable both to the automotive market and to a broad range of industries, including agricultural machinery, construction machinery, marine vessels, aviation, and aerospace, we intend to expand its application scope.

6 Conclusions

EV powertrain technology was evaluated to improve the efficiency of BEV development by integrating simulation with real electrical power sources. The expected battery SoC transition was successfully confirmed by controlling the power supply unit from a HIL simulator to simulate EVSE power control, as well as power control via motor drive and regenerative power generation to simulate the powertrain during driving. These results demonstrate that evaluation equivalent to that of an actual vehicle is achievable.

Based on these demonstration results, we aim to help accelerate development of automobiles toward reaching a BEV-based society through the early market introduction of Power HIL.

References

- 1) Ministry of Land, Infrastructure, Transport and Tourism (MLIT), Japan, "Carbon Dioxide Emissions in the Transport Sector," (in Japanese), https://www.mlit.go.jp/sogoseisaku/environment/sosei_environment_tk_000007.html

- 2) Ministry of Economy, Trade and Industry (METI), Japan, “Automobile and Storage Battery Industry: Main Future Initiatives,” (in Japanese),
https://www.meti.go.jp/policy/energy_environment/global_warming/ggs/05_jidosha.html
- 3) TAKASAGO, LTD., “Regenerative Hybrid power supply RZ-X2-100K,”
<https://www.takasago-ss.co.jp/en/products/power-supply/rz-x2-100k/>

Authors



Takuya Yoda
 Solution Development Dept.
 Environmental Measurement
 Company



Yuta Shinchi
 Solution Development Dept.
 Environmental Measurement
 Company

Table 4 Control PC Main Specifications

Item	Specification
OS	Windows 10 Pro for Workstations Plus (64 bit)
CPU	Intel® Xeon® Gold 5222 Processor × 2 (Dual-processor)
Graphics	NVIDIA® TITAN RTX (GDDR6 24 GB)
Main Memory/Storage	96 GB (DDR4-2933)/1 TB (SATA 6 Gb/s)
Network (LAN)	1 × RJ45 (10GBASE-T, 1000BASE-T, 100BASE-TX)
Power	1125 W/100 Vac (50/60 Hz)
Dimensions/Mass	Approx. 216 (W) × 556 (H) × 445 (D) mm/ approx. 23.7 kg (standard configuration)
Operating Environment	Temperature: 10° to 35°C; Humidity: 10% to 90% (no condensation)

Table 5 HIL Simulator Main Specifications

Item	Specification
OS	RTOS
CPU	Intel® Core™ i7 Processor
Main Memory/Storage	4 GB (DDR4)/8 GB flash memory
Interface	Gigabit Ethernet
	RS232 interface with standard UART
I/O Board	1 × RJ45 (10BASE-T, 100BASE-TX, 1000BASE-T)
	1 × Dsub-9P
	4 × RJ45 (10BASE-T, 100BASE-TX, 1000BASE-T)
I/O Board	Ethernet
	CAN
	8 (CAN, CAN-FD)
I/O Board	Multi-I/O
	Analog, Digital
Power	350 W/100 to 240 Vac (50/60 Hz)
Dimensions/Mass	Approx. 250 (W) × 400 (H) × 200 (D) mm/ approx. 7.5 kg (excluding board)
Operating Environment	Temperature: 0° to 50°C; Humidity: 5% to 95% (no condensation)

Table 6 Power Supply (RZ-X2-10K) Main Specifications

Item			Specification
Output	Rated output voltage/rated output current/ rated output power		+100 V/±300 A/±10 kW
Input	Operating power supply		180 to 250 Vac, 3-phase, 45 to 65 Hz
Constant Voltage (CV)	Set range	Hi/Lo	+0.00 to +102.00 V/+0.000 to +30.600 V
Constant Current (CC)	Set range	Hi/Lo	−306.00 to +306.00 A/−30.600 to +30.600 A
Constant Power (CP)	Set range	Voltage: Hi/Current: Hi	−10.200 to +10.200 kW
		Voltage: Hi/Current: Lo	−3.060 to +3.060 kW
		Voltage: Lo/Current: Hi	−9.180 to +9.180 kW
		Voltage: Lo/Current: Lo	−918.0 to +918.0 W
Series/Parallel Operation	Expandability	Max. connected units	Series: 5 units/Parallel: 10 units
	Max. rated output voltage/Max. rated output current		+500 V/±3000 A (10 units in parallel, without series)
	Max. rated output power		±200 kW
Other Functions	External communication interface		LAN/CAN CAN FD (option)
Operating Environment	Ambient temperature/humidity		0° to 40°C/20% to 85% (no condensation)
Dimensions/Mass			430 (W) × 355 (H) × 650 (D) mm/approx. 70.0 kg

Table 7 Junction Box (Power Line Switching Unit) Main Specifications

Item	Specification	
Inputs/Outputs	1/2 (A, B)	
Rated Voltage/Rated Current	Rated impressed voltage: 100 Vdc/Rated carrying current: ±100 Adc	
Voltage Monitor Output	Scale/Accuracy	+10 V/+100 V (Internal circuit insulation voltage: 500 Vdc)/within 100 V ±3%
Current Monitor Output	Scale/Accuracy	±10 V/±100 V (Internal circuit insulation voltage: 500 Vdc)/within 100 V ±3%
Output Switching	Signal specification	Voltage-free contact input, 24 Vdc, 20 mA min.
Input/Output Lines	1500 V, MLFC 60 mm ² , 2-conductor twisted, 10 m (crimp terminals M8, M12) × 3	
Input Operating Voltage	Single-phase 90 to 110 Vac, 50/60 Hz, approx. 10 A max.	
Dimensions/Mass	Approx. 600 (W) × 500 (H) × 300 (D) mm/approx. 40 kg	

Publicly available

Prism: An Open-Source Framework for Building XAML Applications in WPF for Instrumentation

Tom Roberts

[Summary]

This paper examines the Prism framework as a foundational open-source solution for building large-scale, modular WPF applications in instrumentation and scientific software. It discusses the architectural challenges inherent in WPF and MVVM development and explains how Prism addresses these challenges through standardized infrastructure such as dependency injection, modularity, region-based composition, navigation, and loosely coupled communication. Key Prism features—including the ViewModel Locator, Region Manager, Event Aggregator, and Dialog Service—are explored in the context of scalability, testability, and long-term maintainability. Finally, the paper considers Prism's evolution as an open-source project and its continued relevance for enterprise-grade instrumentation applications.

1 Introduction

Open-source frameworks are essential for instrumentation and scientific software as they provide pre-built, community-vetted foundations that allow developers to skip complex infrastructure tasks and focus on specialized features. These frameworks accelerate development cycles by offering modular, reusable components for large-scale enterprise applications, ensuring they can scale efficiently to meet the demands of timely feature additions without breaking existing ones. Furthermore, the inherent transparency of open source fosters a culture of collaborative problem-solving and rigorous peer review, which enhances software reliability.

Over the past 10–15 years, there has been a shift from Windows Forms (WinForms) to Windows Presentation Foundation (WPF) for enterprise line-of-business applications. WPF offers a more modern User Interface (UI) framework with powerful data binding, Model-View-ViewModel (MVVM) support, and styling/templating, which makes large applications easier to maintain and scale. WPF has a significant learning curve because it requires mastering several advanced concepts simultaneously, including XAML^(1,2),

a declarative markup language for UI design that is distinct from traditional C# code. WPF's core architectural pattern, MVVM^(3–5), demands a major shift in thinking from simpler event-driven models like WinForms to a more decoupled approach. Additionally, understanding complex internal mechanics such as dependency properties, routed events, and data binding^(6–9) can be daunting for beginners. Unlike older frameworks, WPF also introduces sophisticated and powerful layout systems, templates, and styles but adds significant verbosity and initial confusion. Finally, debugging can be difficult because runtime errors in the XAML object tree often produce abstract or unhelpful messages.

Implementing the MVVM pattern is essential for professional WPF development because it decouples the UI from the application's core logic, enabling applications to scale without becoming unmanageable. By moving state and behavior into a separate ViewModel, developers create logic that is unaware of the View and communicates through data binding, commanding and notifications. As illustrated in Figure 1, the View and ViewModel interact indirectly through well-defined mechanisms, while the Model remains

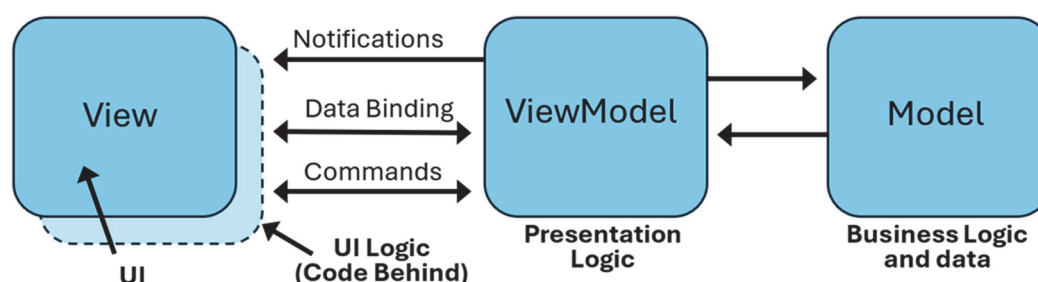


Figure 1 Relationships and interactions between MV

isolated from UI concerns. This separation allows designers to modify the visual experience in XAML without risk of breaking underlying functionality, while also facilitating comprehensive unit testing of business logic in complete isolation from the UI thread.

While the core WPF framework provides the raw building blocks for MVVM, such as data binding and commands, it lacks the high-level infrastructure required to handle complex application-wide concerns. Professional MVVM frameworks fill these gaps by providing standardized “plumbing” for common tasks like message-based communication between disconnected components, integrated dependency injection (DI) for managing object lifetimes, and automated property change notification. Without these frameworks, developers often resort to writing repetitive “boilerplate” code for every new screen, which increases the likelihood of human error and significantly slows development cycles.

Scaling a WPF application to an enterprise level requires a robust modular architecture that frameworks simplify through advanced navigation and composition systems. These frameworks allow large teams to build and maintain independent features as separate plug-ins that can be dynamically loaded at runtime. This modular approach ensures that a change in one functional area of the application does not inadvertently cause a failure in another, making the entire system more resilient and easier to maintain.

Ultimately, leveraging an established MVVM framework standardizes the development process and enforces architectural best practices across a project. By providing a common language and set of conventions, these tools help teams avoid the “spaghetti code” that frequently plagues long-lived desktop applications. As applications evolve, built-in support for sophisticated features—such as automatic View-to-View-Model discovery and cross-component event systems—enables a level of agility and maintainability that is difficult to achieve through manual implementation alone.

2 Prism Library Evolution

The Prism MVVM Framework was established in 2008 by Microsoft’s patterns & practices team, originally under the name Composite Application Library. It was developed to provide official guidance and a set of libraries for building

“composite applications” - complex, large-scale WPF software that requires a modular and loosely coupled architecture^{10,11}. By implementing the MVVM pattern along with other architectural principles like DI and Event Aggregation, Prism was designed to solve the challenges of maintaining and testing sophisticated XAML-based applications. It allows multi-disciplinary teams to develop, test, and evolve individual application components independently, ensuring the UI remains separate from business logic to improve long-term flexibility and code reuse. In 2015, Microsoft transitioned the project to the open-source community^{12,13}.

As of 2026, the Prism Library remains a dominant framework for building modular XAML applications. Its core package, Prism.Core, has surpassed 19.5 million total downloads on NuGet, with a daily average of approximately 8,600 downloads. The user base is extensive and well-established, supported by a mature ecosystem of community-driven documentation and active development targeting the latest .NET environments.

Prism¹⁴ fundamentally simplifies the MVVM pattern by offering built-in components which ensure a clean separation between UI and business logic. Prism’s integration with common DI containers allows developers to manage object lifetimes and resolve dependencies effortlessly, promoting a loosely coupled system. This separation of concerns enhances testability, allowing ViewModels and services to be verified independently using design-time mocks or automated tests without a live UI. Modularity is a core strength, enabling large applications to be decomposed into independent modules that can be developed, tested, and updated by separate teams in parallel. The framework further supports a composite UI approach through regions and navigation services, allowing dynamic content loading into specific placeholders. Advanced communication between these decoupled parts is facilitated by the Event Aggregator, which provides a thread-safe, loosely coupled publisher/subscriber mechanism. By standardizing these architectural patterns, Prism reduces technical debt and ensures that applications remain scalable and maintainable as requirements evolve.

A trend among open-source maintainers is the shift toward tiered licensing models to address the long-term sustainability of open-source projects and prevent maintainer burnout.

In 2025, Prism migrated from a community to a dual community/commercial license model. These licensing structures are beneficial because they create a more sustainable financial model by allowing open-source maintainers to earn directly from commercial use while keeping the code accessible for community use. By balancing commercial interests with open-source principles, these frameworks can secure the funding necessary for professional-grade support, dedicated maintenance, and faster feature development. Ultimately, this trend helps stabilize the software supply chain, ensuring that the essential infrastructure upon which 96% of organizations¹⁵⁾ now rely remains actively supported into the future.

3 Integration of Dependency Injection (DI) containers in Prism

DI is a critical design pattern that decouples the creation of an object from its usage, facilitating a separation of concerns that is vital for scalable software. By centralizing the management of instances, developers can avoid hard-coded dependencies, making applications more modular and significantly easier to unit test through mock injections. The Prism MVVM framework streamlines this process by integrating with popular DI containers like Unity or DryIoC, providing a dedicated ViewModel Locator to automatically resolve and wire up dependencies at runtime. Prism further enhances instance management by handling object lifetimes, ensuring that services and components are properly registered as singletons or transient instances based on application needs. Without DI, developers often resort to the Singleton pattern or manual instantiation, which can lead to severe pitfalls such as tight coupling and rigid codebases that are difficult to modify. A lack of centralized management frequently causes memory leaks, as long-lived objects may hold onto references of short-lived components, preventing the garbage collector from reclaiming memory. These leaks are notoriously difficult to track because they often manifest as gradual performance degradation rather than immediate crashes. Identifying these hidden references can be extremely difficult and time-consuming to track down and resolve. Prism's structured approach ensures that dependencies are explicit and lifecycle-managed, mitigating the risk of memory leaks and the extreme debugging challenges associated with manual instance control.

4 Loose coupling benefits

In software design, loose coupling refers to components interacting through well-defined interfaces rather than direct references. Building features with loose coupling in Prism offers several critical benefits:

- **Independent Evolution:** Features can be developed, updated, or replaced in isolation, ensuring that changes in one module do not cause cascading failures or require re-deploying the entire application.
- **Enhanced Testability:** Because components are not tightly bound, they can be easily mocked or stubbed, allowing for more reliable unit testing in isolation.
- **Improved Maintainability:** With a clear separation of concerns, bugs are easier to isolate, and the codebase remains more manageable as the application scales.
- **Scalability:** Different teams can work on separate functional areas simultaneously without blocking each other, speeding up the overall development lifecycle.

5 Modules

A Prism module is a self-contained, feature-focused unit in a Prism application that bundles related views, ViewModels, services, and registrations. It allows developers to build and load application functionality independently, making large WPF apps easier to scale, test, maintain, and evolve over time. When an application has many modules, it is best to locate them outside the main Prism (shell) project. Typically, the shell project contains only the application startup, shared infrastructure, and layout, while each module lives in its own separate class library project. This keeps features isolated, reduces coupling, and allows teams to work on modules independently.

Figure 2 shows sections of a typical instrument application window: MenuBar, ToolBar, SideBars, StatusBar, Chart area, etc. The MenuBar may have many defined Regions such as: FileMenuRegion, TraceMenuRegion, DisplayMenuRegion, etc. The ToolBar could have similar or related regions to the MenuBar regions: FileToolBarRegion, TraceToolBarRegion, DisplayToolBarRegion, etc. with shared functionality. For example, views, ViewModels, services, and registrations serving the TraceMenu and TraceToolBar regions would most likely be co-located in the same Prism Module.

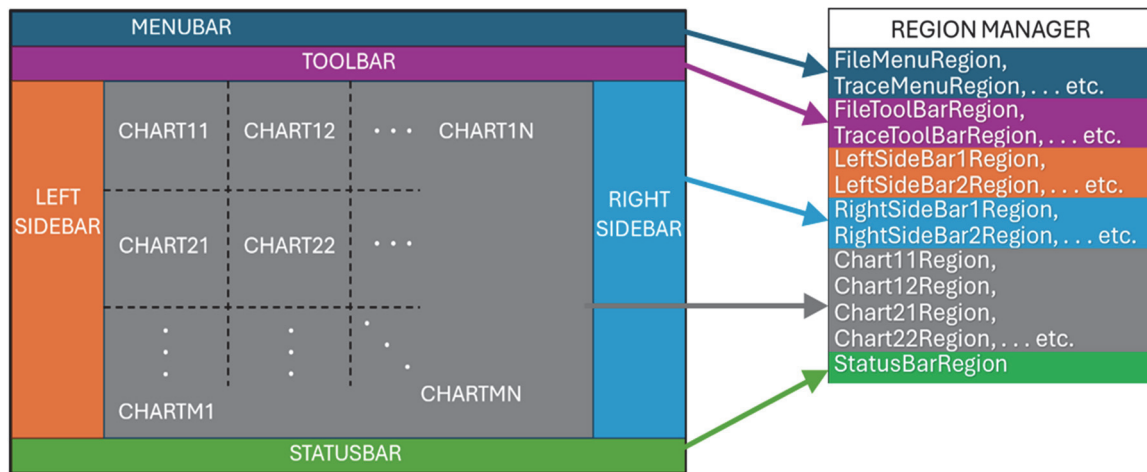


Figure 2 Application Main Window sections and Region names

6 Custom Controls for Instrumentation

Standard controls provided by WPF are usually insufficient for applications developed for instrumentation products. Interactive custom chart controls (e.g., Tabular, Rectangular, Smith, Polar, Gage, Vector Constellation, Eye Diagram, 3D surface¹⁶, etc.) with their own properties are required. Many of these charts are available from third-party vendors but rarely satisfy all customer UI requirements so custom controls are often developed internally. Custom controls are another modular component that can be developed independently and evolve with customer requirements. These interactive custom controls can be treated like any other standard control and injected into regions as an instance or instances using Prism. The development of custom controls is not discussed in this paper, but resources are available at Pluralsight¹⁷. From a Prism perspective, custom controls function as modular components and are often packaged within feature modules.

7 The ViewModel Locator

When using the MVVM pattern, a way to connect to the Views DataContext to the instance of the View's ViewModel is necessary. Prism's `ViewModelLocator` simplifies making this connection. While several approaches exist, Prism's `ViewModelLocator` avoids manual wiring by automatically creating and assigning the ViewModel to the view's DataContext. Prism's implementation is lightweight and container-driven. It works by setting a single attached property on the view, which triggers Prism to resolve the ViewModel using dependency injection. Prism relies on naming conventions to locate the correct ViewModel, expecting Views to live

in a Views namespace and ViewModels to live in a ViewModels namespace. The default rule appends "ViewModel" to the view name to find the corresponding ViewModel. Developers can override the naming convention or register ViewModels explicitly to fully control the mapping.

8 Regions and Region Manager

Regions are the primary mechanism to inject or display views to users throughout the application. A Region is simply a placeholder for dynamic content and can be created in XAML or in code. Regions have no knowledge of the Views that will be moved in and out of the Region. The Region Manager maintains the collection of the regions and allows access to them. A Region is defined in a Region Adapter, which is a host control for the Region. Prism supports various Region Adapters: `ContentControl`, `ItemsControl`, and `Selector` controls such as `ComboBox`, `ListBox`, `Ribbon` and `TabControl`. Prism's supported region adapters cover almost all scenarios needed; however, custom region adapters can be constructed. For example, defining a Region in a `StackPanel` would require a custom `StackPanel Region Adapter`¹⁴. The chart area in Figure 2 has $m \times n$ chart regions with grids hosted in a `TabControl`, for example. Any chart type instance could be moved in or out of any $m \times n$ chart regions using Region Navigation.

9 Region Navigation

Region (or View) Navigation in Prism provides a loosely coupled way to move views into and out of regions, making it ideal for MVVM scenarios where ViewModels should not reference view types directly. Region Navigation manages

the full navigation lifecycle, including activating and deactivating views and maintaining back and forward navigation stacks. Navigation is performed using URIs, which act as unique keys mapped to view types, eliminating direct dependencies on views. To enable navigation, views must be registered with Prism using the container registry, typically by calling `RegisterForNavigation`, which by default uses the view's name as the key. Once registered, navigation can be triggered through the `IRegionManager` by calling `RequestNavigate` with a region name and navigation key. Alternatively, navigation can be requested directly on an `IRegion` instance retrieved from the Region Manager. The Region Manager is usually accessed through constructor injection. This approach allows ViewModels to participate cleanly in navigation while preserving MVVM principles. Additionally, Region Navigation supports passing parameters and tracking where a view is within the navigation process.

10 Communicating between different areas of the application

Prism helps WPF applications become more decoupled, making them easier to build, maintain, extend, and test, but this decoupling can make communication between application areas more complex. One common challenge is enabling Views and ViewModels to communicate with different areas of the application without tight dependencies. Prism addresses this problem through the `IEventAggregator` interface which provides a loosely coupled, event-based communication mechanism. The Event Aggregator works using publishers and subscribers, where one part of the application publishes an event and other parts subscribe to receive it. Multiple publishers and subscribers can exist for the same event as shown in Figure 3. Subscribers must register for an event before publishers send it, and once published, the Event Aggregator routes the event to all subscribers. Events can also carry data objects, known as a payload, allowing information to be passed between application areas. Subscribers can filter events based on payload values to ensure they only handle relevant messages. The Event Aggregator uses weak references by default to prevent memory leaks, but subscribers that opt into strong references must be manually unsubscribed.

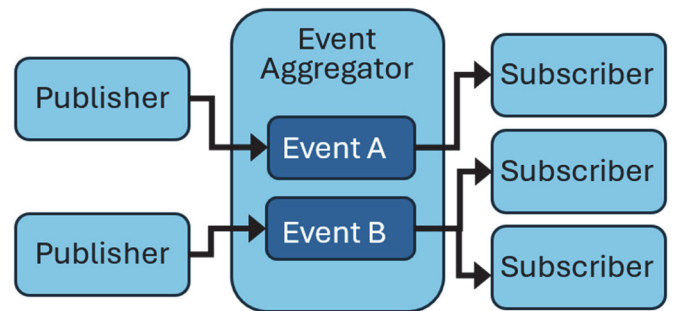


Figure 3 Event Aggregator¹¹⁾

11 Dialogs and the Dialog Service

Applications frequently need to prompt users for input or display notifications, which is typically done through dialogs. A dialog in Prism is a `UserControl` designed to display information, request confirmation, or collect user input. In Prism, dialogs are shown without directly tying ViewModels to UI elements. Prism uses the `IDialogService`, which allows dialogs to be invoked from ViewModels without referencing views. Dialogs can be shown in either modal mode, which blocks interaction with the rest of the application, or non-modal mode, which allows continued interaction and even multiple dialog instances.

Dialogs in Prism have a ViewModel that implements the `IDialogAware` interface. This interface provides key features such as setting the dialog title, controlling whether it can close, responding when it opens or closes, and receiving parameters for initialization. Dialogs are triggered by commands in the ViewModel, keeping business logic decoupled from the UI. Dialogs must be registered with the Prism DI container. Once registered, the `IDialogService` can be used to display the dialog in a clean, testable, and maintainable way. The dialog service can also be used to show multiple shells¹⁸⁾ or what is sometimes called detached windows.

12 Conclusions

This paper highlights the main features of Prism and its benefits in providing a standardized infrastructure to accelerate front-end software development for instrumentation. The Prism MVVM framework is indispensable for large-scale applications because it provides a robust, modular architecture that allows independent teams to develop, test, and deploy features simultaneously without conflict. Its core features—such as Dependency Injection, Event Aggregation, and region management—are essential for managing the

high complexity of modern enterprise software by ensuring components remain loosely coupled. Prism eliminates the need for developers to write repetitive boilerplate code or re-design foundational infrastructure. This pre-built functionality significantly reduces overall development time, allowing teams to focus exclusively on business logic rather than boilerplate plumbing. Furthermore, Prism’s navigation and modularity features provide a standardized “common shell” that ensures consistent UI behavior across large codebases. Ultimately, using this established framework future-proofs applications by making them easier to maintain and evolve as business requirements grow over time.

13 References

1) Huber, T. C. (2024, June). XAML layout in depth [Video]. Pluralsight. www.pluralsight.com

2) Nathan, A. (2014). XAML unleashed. Sams Publishing.

3) Noyes, B. (2021, October). WPF MVVM in depth [Video]. Pluralsight. <https://www.pluralsight.com/courses/wpf-mvvm-in-depth>

4) Huber, T. C. (2022, October). WPF and MVVM: Advanced model treatment [Video]. Pluralsight. www.pluralsight.com

5) Huber, T. C. (2022, October). WPF and MVVM: Test driven development of ViewModels [Video]. Pluralsight. <https://www.pluralsight.com/courses/wpf-mvvm-test-driven-development-viewmodels>

6) MacDonald, M. (2012). Pro WPF 4.5 in C#: Windows Presentation Foundation in .NET 4.5 (4th ed.). Apress. doi.org

7) Noyes, B. (2022, October). WPF Data Binding in Depth [Video]. Pluralsight. <https://www.pluralsight.com/courses/wpf-data-binding-in-depth>

8) Noyes, B. (2022, October). Extending XAML applications with behaviors [Video]. Pluralsight. <https://www.pluralsight.com/courses/extending-xaml-applications-behaviors>

9) Noyes, B. (2022, October). WPF Productivity Playbook [Video]. Pluralsight. <https://www.pluralsight.com/courses/wpf-productivity-playbook>

10) Brumfield, B., Cox, G., Hill, D., Noyes, B., Puleio, M., & Shifflett, K. (2011). Developer’s guide to Microsoft Prism 4: Building modular MVVM applications with Windows Presentation Foundation and Microsoft Silverlight. Microsoft Press.

11) Microsoft Patterns & Practices. (2014, April). Developer’s guide to Microsoft Prism Library 5.0 for WPF, Microsoft Press. <http://download.microsoft.com/download/5/2/6/52627B2B-AE05-4781-AA5D-7EBCC4C86D9B/Prism5forWPF.pdf>

12) Prism Library. (2024). Prism (Version 9.0) [Computer software]. GitHub. <https://github.com/PrismLibrary/Prism>

13) Prism Library. (2024, February). Prism Library [Website]. <https://prismlibrary.com/>

14) Lagunas, B. (2022, May). Introduction to Prism 7 for WPF [Video]. Pluralsight. <https://www.pluralsight.com/courses/prism-wpf-introduction>

15) OpenLogic by Perforce, Eclipse Foundation, & Open Source Initiative. (2025). 2025 State of Open Source Report. <https://www.openlogic.com/resources/state-of-open-source-report>

16) Xu, J. (2016). Practical .NET chart development and applications: Advanced chart programming for real-world .NET 4.5 applications using C#, WPF, and MVVM. UniCAD Publishing.

17) Lagunas, B. (2024). Introduction to WPF custom controls [Video]. Pluralsight. <https://www.pluralsight.com/courses/wpf-custom-controls-introduction>

18) Lagunas, B. (2021, July). Prism problems & solutions: Showing multiple shells [Video]. Pluralsight. <https://www.pluralsight.com/courses/prism-showing-multiple-shells>

Author



Tom Roberts
Service Infrastructure Solutions
US Division
Test & Measurement Company

Publicly available

Contribution to other Publications

Contribution to other Publications (April 2025 to March 2026)

Author	Affiliation	Title	Publication
Ryota Kaji Kodai Ishida Kazuyoshi Inoue Takuya Oda Katsuhiro Takenaga Kentaro Ichii	*1 M M *1 *1 *1	Comparison of Crosstalk Measurement Methods for Weakly-coupled Multicore Fibers	OECC
K. Komatsu K. Ishida S. Beppu D. Soma S. Sumita T. Fukushima K. Inoue H. Takahashi Y. Wakayama T. Tsuritani	*2 M *2 *2 *2 M M *2 *2 *2	Impact of Inter-Core Crosstalk on Coherent Optical Time-Domain Reflectometry in Repeated Multicore Fibre Systems	ECOC 2025 (The 51st European Conference on Optical Communication, 28 Sep. - 2 Oct. 2025)
Takamoto Yokosawa Tomohiro Matsui	R R	Unconventional thermal conductivity of suspended zigzag graphene nanomesh	arXiv:2601.10935
Hiroshi Shibata Kohei Watanabe Hiroshi Yamamoto Ken Ito Motohiro Banno Shinji Morimoto Yosho Kanda	*3 *3 *3 *3 M M M	Evaluation of anomaly detection in optical transport networks using GSNR estimation method	The Institute of Electronics, Information and Communication Engineers Society Conference, B-6-65
Hiroshi Shibata Kohei Watanabe Hiroshi Yamamoto Ken Ito Motohiro Banno Shinji Morimoto Yosho Kanda	*3 *3 *3 *3 M M M	Practicality Evaluation of GSNR Estimation Methods Using DCO Transceivers	To be published in IEICE Technical Report NS2025 (in Japanese)
Takashi Kawamura	M	Current Status and Future Prospects of Terahertz Wave Technology (Supervised by Osamu Hashimoto)	Joho Kiko Co., Ltd., Chapter 1, Section 4, Subsection 1, pp. 59–69, July 2025. (in Japanese)
Eiji Taniguchi Jyunichi Sano Yuma Kinno Kohei Suzuki	I I I I	Comprehensive quality assessment of all solid dosage units using Real-Time Near-Infrared Transmission Spectroscopy System (Ariphas)	PHARMA TECH JAPAN, Vo.42, No.2 (2026)
Motoki Inoue Eiji Taniguchi Jyunichi Sano	*4 I I	Real-Time Total Content Uniformity Assessment of Solid Dosage Forms Using Near-Infrared Transmission Spectroscopy	Analytical Chemistry 2025, 97, 11, 6111-6117

Author	Affiliation	Title	Publication
K. Komatsu K. Ishida S. Beppu D. Soma S. Sumita T. Fukushima H. Takahashi Y. Wakayama T. Tsuritani	*2 M *2 *2 *2 M *2 *2 *2	Repeater Fault Detection and Localization Using Coherent OTDR for Multicore Fiber Submarine Cable Systems	OFC2026 (Optical Fiber Communication Conference & Exhibition, 15 - 19 March 2026)
Haruto Horino Takahiro Matsuda Hirofumi Suganuma Shinichi Segawa	*5 *5 M M	Validation of an IRS Deployment Method Based on Single-Reflection Point Distribution	IEICE Technical Report, CQ2025-119, Mar. 2026.
Akiko Tada Masaru Koshihara Takanori Saitoh	D D D	High-accuracy OFDR-type dynamic distributed strain measurement using phase integration coherence compensation method and interferometer base point shift	29th International Conference on Optical Fiber Sensors, 136392V (22 May 2025)
Akiko Tada Masaru Koshihara Takanori Saitoh	D D D	High accuracy OFDR system using phase integration coherence compensation method and its application to shape measurement	IEICE Tech. Rep., vol. 125, no. 199, OFT2025-57, pp. 118-123, Oct. 2025.
Akiko Tada Masaru Koshihara Takanori Saitoh	D D D	High accuracy OFDR system with phase integration coherence compensation	73rd Meeting on Lightwave Sensing Technology LST73-14 (2025)

R : Advanced Research Laboratory

M : Test&Measurement Company

I : Infivis Company

D : Sensing&Devices Company

E : Environmental Measurement Company

H : Engineering Div.

*1 : Fujikura Ltd

*2 : KDDI Research, Inc.

*3 : NTT, Inc.

*4 : Hoshi University

*5 : Tokyo Metropolitan University



ANRITSU CORPORATION

<https://www.anritsu.com>
5-1-1 Onna, Atsugi-shi, Kanagawa 243-8555, Japan