ISSN 2435-1288

Nikon Research Report

Vol.3 2021

Purpose of publication

This publication is being created to widely introduce the achievements of research and development activities conducted by Nikon Corporation. This is a result of R&D based on Nikon's core technologies of "opto-electronics" and "precision" technologies that have been incorporated in new products and/or often valued by external organizations such as academic societies. At the end of the report, the papers presented and published by our engineers in the previous year are listed.

Foreword

Representative Director
President

Toshikazu Umatate

The COVID-19 pandemic, which spread throughout 2020, is still having an enormous impact around the world. It is causing many aspects of society to be rapidly transformed. While our Nikon Group is also facing the same circumstances, I do not believe that the direction of business and technology required as we advance into the future will fundamentally alter, but the schedule has just suddenly been pushed forward.

During this critical transformation period, it is essential that the Nikon Group responds to new business models and business processes that emerge from the combination of sensing technology, communication technology, and big data analysis technology. It is also important to realize advanced and automatic manufacturing in response to the shortage of manpower, increased labor costs and diversification of individual values. I believe it is vital to accurately identify the needs of society and fully comprehend what society really desires with tools that Nikon can provide for it. To successfully meet these demands, we must continuously increase the sophistication of our unique tools and services.

This report shares some of our successful achievements, such as the data analysis technology that innovates business processes and technology that realizes highly functional robots that can contribute towards advanced manufacturing.

General Manager
Research & Development Division

Kazuhiro Kido

As significant changes and constraints continue around the world, we urgently need technologies to handle such difficult situations. Development processes also require a completely different level of flexibility and alterations.

Especially in the first half of 2020, many academic conferences were forced to cancel due to restrictions on business activities around the world. This also prevented Nikon from carrying out experiments that are essential for research and development, while limiting the opportunities for showing our development results and exchanging information. However, regardless of such circumstances, we were able to steadily move forward by reviewing the research and development approach itself. This resulted in reflecting our achievements in our new products, which continue to satisfy the demands of society. Thanks to that, we were also highly evaluated at academic societies that have become active again in the latter half of 2020.

We will share the Nikon Group's research and development results for this year through this report, which contains technical explanations of products released in FY2020, and papers valued by external organizations during the same year. I hope that this report provides you with a deeper understanding of the technologies we are working on and opportunities for you to employ Nikon's technologies for the benefit of society.

Nikon Research Report Vol.3
目次／ CONTENTS

技術解説/Technical Reports

研究開発論文/Research and Development Reports

技術解説

Technical Reports

顕微鏡用画像統合ソフトウェアにおける AI 技術紹介とライフサイエンス応用

門井宏平，畑口剛之，武居俊輔

Introduction of AI Technology in Imaging Software for Microscopes and Life Science Applications

Kohei KADOI, Takeshi HATAGUCHI and Shunsuke TAKEI

ライフサイエンスの分野で Artificial Intelligence による画像処理の応用が広がりつつある．我々は，顕微鏡システム用の画像統合ソフトウェアである NIS-Elements に，Deep Learning を用いた画像処理技術である NIS.ai を搭載した．NIS.ai は NIS-Elements に統合することで，ユーザーが容易に先進的な Deep Learning 技術を利用することができる特徴を持ち，画像の生成や領域分割の実施が可能である．本稿では NIS.ai の活用により，正確な解析結果の取得とユーザーの作業負荷低減が可能であることを紹介する．まず，非蛍光染色細胞の画像に対し NIS.ai を用いた解析が，蛍光染色細胞の画像を用いた従来の解析と同等の精度となることを確認した．次に，蛍光撮影において問題となる染色試薬による影響や蛍光撮影時に生じる光毒性を NIS.ai の活用により回避出来ることを実証した．最後に，NIS-Elements が制御する顕微鏡撮影手順に NIS.ai を組み込むことにより，従来の画像解析技術の利用では達成が困難であった自動撮影が可能となり，新たな価値を提供できることを示した．

The application of artificial intelligence in image processing is being investigated extensively in life sciences. We developed and installed NIS.ai, an image processing technology that uses deep learning, in NIS-Elements, which is an imaging software for microscope systems. By integrating NIS.ai with NIS-Elements, NIS.ai enables users to use advanced deep learning technology easily with tasks such as image conversion, segmentation and so on. In this study, we show that users can obtain accurate analysis results with minimal effort using NIS.ai. First, we confirm that the analysis using NIS.ai for images of unstained cells exhibits the same accuracy as that of the conventional analysis for images of stained cells. Second, we demonstrate that the effects of staining reagents and the phototoxicity that occurs when capturing fluorescence images can be avoided by utilizing NIS.ai. Finally, we demonstrate that by incorporating NIS.ai into NIS-Elements, automatic imaging, which is difficult to achieve using conventional image analysis, can be achieved, and a new value can be obtained.

Key words ライフサイエンス，顕微鏡，画像処理，人工知能，深層学習
life science, microscopy, image analysis, artificial intelligence, deep learning

1 Introduction

In the field of life science, digital image processing is being conducted during the observation of cells with biological microscopes to automate cell counting and classification. Recently, widespread studies have been conducted on image-processing technology using deep learning, which is a type of artificial intelligence (AI) that is applied to segmentation, classification, digital staining, sharpening, and super-resolution [1].

NIS-Elements is an imaging software for microscope systems with microscope and camera control, image processing, analysis, and reporting functions. Nikon previously equipped NIS-Elements with a microscope image-processing functions, applying several deep learning techniques. In this paper, we introduce NIS.ai behind the functions that may be adopted for digital staining (Convert.ai) and segmentation (Segment.ai), as well as their applications in life science.

2 Microscope image processing functions applying deep learning

Image processing technologies, such as morphological transformation, and non-deep machine learning technologies, such as random forest can be adopted for relatively simple segmentation and sharpening. However, proficient skills and individual adjustments, such as designing image filters, are required to improve the quality of processing results.

The two NIS.ai functions (Convert.ai, Segment.ai) introduced in this paper adopt convolutional neural networks, a type of deep learning, and employ supervised learning, which require training data. Because the design of image filters is automatically performed in its training process, deep learning does not require the traditional image processing or proficient skills required for machine learning. In addition, it is characterized by further improving accuracy by increasing the training dataset size or the number of training iterations.

When using deep learning, hyperparameter usually needs to be tuned, depending on the tasks to be applied. Tuning hyperparameters is a trial-and-error task, which is generally time- and labor intensive. However, since NIS.ai adopts optimized networked construction for microscope images, optimal results can be easily obtained with just a few settings even if the user is not familiar with deep learning.

Table 1　Settings for NIS.ai training

Convert.ai	Iterations
	Dynamic range adaptation (on/off)
Segment.ai	Iterations
	Dynamic range adaptation (on/off)
	Detect touching object (on/off)

Table 1 presents the settings for NIS.ai training. Iterations is an option that specifies the number of training repetitions. Although the optimal value depends on the amount of training data and variations, a value of approximately 1000 is usually acceptable. The training times required to use NVIDIA Quadro RTX 4000 to set 1000 iterations are approximately 3.5 h and 4.0 h with Convert.ai and with Segment.ai, respectively, and the inference time for images of 1600 x 1600 pixels in the same environment is 1 s or less per image. Although the training time is long, this poses no practical challenge because it is more important that a trained model, prepared at one time by the training, can be used several times with a short inference time. “dynamic range adaptation” is an option for specifying when there is little brightness variation in training data, while “detect touching object” is an option for specifying when the user wants a highly-precise isolation of objects in close proximity to each other in Segment.ai.

Processing by NIS.ai comprises two phases: a training phase for preparing a trained model using user-prepared training data, and an inference phase using the training results to output inferred images from the target data (Fig. 1).

In the training phase, the user needs to prepare an original image for conversion, as well as a ground truth image to serve as a teacher. The image acquisition function of NIS-Elements can acquire multi-channel images such as phase contrast images and fluorescent images. Users can use multi-channel images acquired with NIS-Elements as they are, or use images, which are processed with the NIS-Elements image processing functions, as training data.

In the inference phase, after the target data have been inferred with the trained model, the processing outputs are further processed by the NIS-Elements image processing functions, and analyzed by counting or tracking. Subsequently, the results can be output as graphs or exported to Excel.

In addition, by adopting a function that automates combinations of NIS-Elements standard image processing and analysis processing (GA3), and a function that automates the analysis processing steps from image acquisition with specified complex conditions (JOBS), it is possible to automate the series of processes from image acquisition to analysis. Furthermore, the analysis can be more efficient by changing the processes according to the analysis results. Since NIS.ai is integrated into NIS-Elements, it does not only provide AI-processing for microscope images, but also facilitates highly convenient, automated, and sophisticated analysis.

In the following sections, we introduce examples of NIS.ai applications with such features in life science.

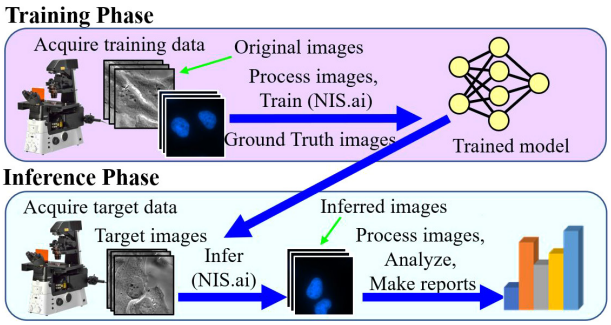


Fig. 1　Processing steps for training and inference

3 Examples of AI Applications in the Field of Life Science

This chapter introduces the following three cases using NIS.ai.

3.1. Highly accurate fluorescence image generation using unstained microscope images

3.2. Avoiding effects of stain reagents with digital staining

3.3. Automated workflow of acquiring microscope images

3.1. Highly accurate fluorescence image generation using unstained microscope images

Samples are stained in biology, medical, and drug discov-

ery studies to identify and detect specific structures and molecules in tissue and cells with a microscope. However, staining involves problems such as the cost of reagents, complexity of experiments, and varying staining results. Furthermore, since samples cannot be reused after staining, multiple samples must be prepared during clinical testing. This problem is addressed by adopting deep learning to output stained images of specific structures from unstained microscope images, such as bright field microscopy, which is also called digital staining. The Convert.ai of NIS.ai is a function that can be applied to this digital staining. In this section, we present an example of generating fluorescence images of cell nuclei from images acquired via phase contrast microscopy.

For the verification sample, we adopted a BS-C-1 cell line that constantly expresses the fluorescent-tagged protein localized on the cell nucleus surface. For observation, we used a Ti-E inverted microscope with a 20x objective lens (CFI S Plan Fluor ELWD ADM 20XC 0.45NA, Nikon, Japan), and acquired phase contrast images and fluorescence images with EMCCD (iXon3, Andor technology, Oxford Instruments, UK). Verification proceeded under conditions of 37°C, 5% CO_2, using a stage-top incubator (STX series, Tokai Hit, Japan) to maintain the culture environment.

Table 2 Summary of training conditions

		Section 3.1	Section 3.2	Section 3.3
Sample used		BS-C-1 cells	HeLa cells	Mouse kidney sections
NIS.ai		Convert.ai		Segment.ai
Training Image	Input	Phase contrast image		Bright field image
	Output	Fluorescence image of cellular nucleus		Glomerular region
Image size		512 x 512 pixels		2048 x 2048 pixels
Number of trained images		70	75	78
Iterations		1000		

For Convert.ai training, we prepared sample images in which the cell density condition was numbered from 20% to 120%, and after training under the conditions presented in Table 2 (Section 3.1 items), we applied the trained model to the time lapse images of a different field of view from the training data. The inference accuracy according to Convert.ai was evaluated based on the growth curve of the number of nuclei and F-score. The growth curve is a graph showing time lapsed alterations in the nuclei number of the same visual field, plotting the quantified number of cell nuclei using existing NIS-Elements image analysis functions. F-score is a common index that indicates the inference accuracy in machine learning. The closer it is to 1, the higher the accuracy. In addition, it was calculated based on whether or not the center of gravity of cell nuclei in ground truth fluorescence images and Convert.ai inferred the existence of images within 10 pixels.

The obtained results confirmed that localization of cell nuclei inferred by Convert.ai exhibits almost the same localization as ground truth fluorescent images (Fig. 2, white arrowhead). Furthermore, the growth curve also exhibited a curve very close to the results from ground truth, based on fluorescent images, and it verified that the inference accuracy was maintained, even when cell density changed (Fig. 3). It was demonstrated that F-score value for Convert.ai is higher than 0.90, if cell density is up to around 80%. It

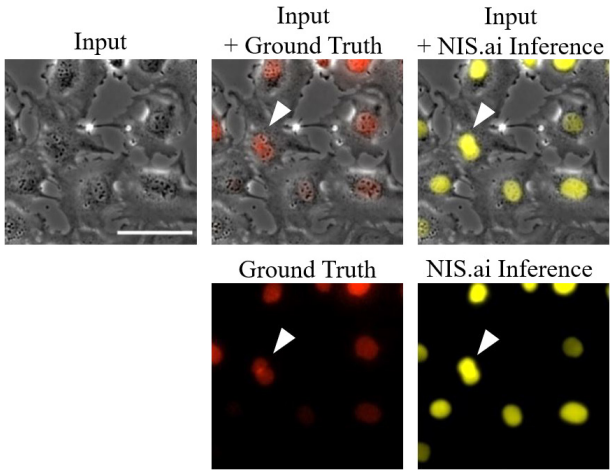


Fig. 2 Comparison of ground truth image of cell nuclei fluorescence and inference image from NIS.ai.
From top left, phase contrast image input into NIS.ai, and input image overlayed with ground truth (red), and input image overlayed with NIS.ai inference (yellow) fluorescence image. Bottom shows ground truth and NIS.ai inferred cell nuclei fluorescence image. White arrowhead indicates the same cell nucleus. Scale bar is 80 µm.

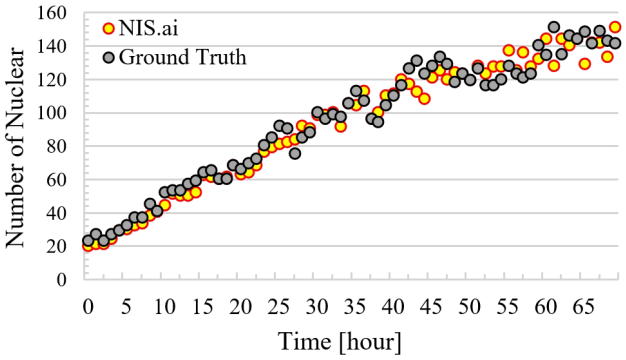


Fig. 3 Comparison of ground truth image of cell nuclei fluorescence, and growth curve of number of cell nuclei using inferred images from NIS.ai.

Table 3 Accuracy of cell nuclei inference by NIS.ai

Cell density (%)	Number of cell nuclei					F-score
	Ground truth	NIS.ai	True Positive	False Negative	False Positive	
30	24	22	22	22	0	0.96
50	65	63	62	3	1	0.97
80	73	69	67	6	2	0.94
100	101	99	89	12	10	0.89

was also shown that Convert.ai can infer with high accuracy, if cell density is around 100%, usually difficult to recognize border between cells.

From the above verification results, it can be deduced that by adopting the Convert.ai function of NIS.ai, cell nuclei can be inferred with high accuracy from unstained images with training, using realistic numbers of data. Utilizing NIS.ai can also save the fluorescence wavelength band used in fluorescence microscopy, making it a useful tool for multicolor observation in basic research fields. Furthermore, its use is also anticipated in clinical research in which valuable samples, such as those derived from patient disease tissues, are subject to analysis.

3.2. Avoiding effects of stain reagents with digital staining

Sample staining involves problems other than cost, labor, or variations in results between experimenters. Owing to the fact that toxicity caused by the addition of staining reagents and irradiated light for fluorescent images acquisition affects cellular dynamics, it is essential that conditions for reagent use and optical configuration of microscope are sufficiently examined. In addition, in the field of regenerative medicine where cells and tissues are returned to a living body, a problem exists, as the total sample examination with staining cannot be performed. It is possible to circumvent these challenges by adopting fluorescent digital staining with NIS.ai. In this section, we introduce verification results using Hoechst. Hoechst is a widely used staining reagent for cell nuclei detection, and staining with it is known to produce toxic and phototoxic substances [2].

We used HeLa cells for verification samples, and performed image acquisition with the same device configuration and environment for the verification as described in Section 3.1. We prepared two experimental groups to verify the effects of reagents. One was a sample of cell nuclei stained with Hoechst prior to image acquisition (“Stained with Hoechst” group), and the other was an unstained sample that applied the NIS.ai trained model (“Unstained (NIS.ai)” group). We acquired time lapse images of the two samples, and adopted them for quantitative analysis. Evaluate the inference accuracy of the unstained NIS.ai group, Hoechst staining was performed immediately before the last image acquisition, and ground truth images were acquired. For NIS.ai training data, we separately acquired images of samples with cell density conditions numbered from 20% to 120%, after staining with Hoechst. Subsequently, we prepared a model that trained these images under the conditions in Table 2 (Section 3.2 items) using Convert.ai, and output fluorescent images of cell nuclei from the input phase contrast images. Cell nuclei were detected and quantified for the Hoechst and unstained NIS.ai groups, respectively, and growth curves were prepared.

Based on the obtained results, we found that cells in the stained with Hoechst and unstained NIS.ai groups of comparable density at the start of time lapse were fewer in the stained with Hoechst group than in the unstained NIS.ai group after 60 h, owing to reagents and phototoxicity (Fig. 4). Growth curves also verified that cell growth was suppressed in the stained with Hoechst group (Fig. 5). This indicated that NIS.ai could achieve sufficiently higher accuracy than ground truth images at the final point of the time lapse (Fig. 5, red dot).

These results indicate that correct cell behavior could be observed by fluorescent digital staining using NIS.ai. It is

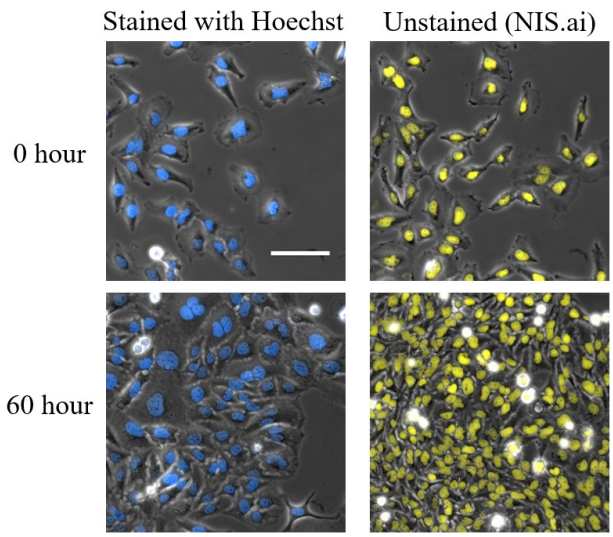


Fig. 4 Image comparison of cells stained with Hoechst prior to starting image acquisition (left) and unstained cells (right)

Left and right columns illustrate the cell nuclei regions stained with Hoechst (blue) and inferred by NIS.ai (yellow), respectively, with their respective phase contrast images superimposed over them. Same field of view images at 0 h and 60 h are arranged vertically. Scale bar is 100 μm.

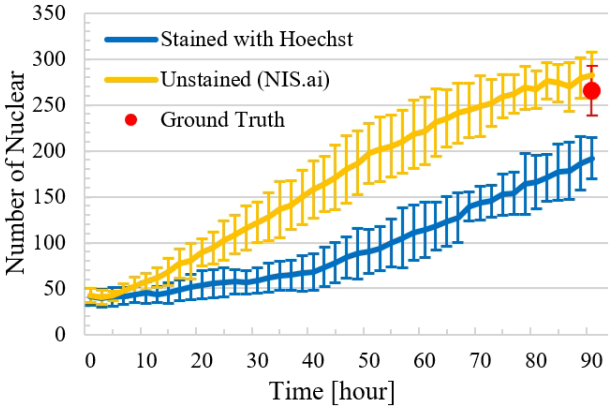


Fig. 5 Comparison of growth curves of number of cell nuclei using cells stained with Hoechst prior to starting image acquisition (blue) and inferred images by NIS.ai of unstained cells (yellow)

Shows the mean ± SD (n = 3) at each point in time. Unstained cells were stained with Hoechst at the final point, and were considered ground truth (red dot).

assumed that ascertaining the behavior of biological sample precisely, without incurring the effects of staining reagents and phototoxicity, will become increasingly important in the fields of biology, medical science, and drug discovery.

3.3. Automated workflow of acquiring microscope images

In studies using conventional microscope images, researchers present only representative microscope images of research subjects, which are often discussed in terms of qualitative results. However, in recent years, technical progress has made it possible to acquire and analyze a large number of images, and it has become necessary to quantitatively evaluate microscope images. Hence, manual image acquisition using a microscope is becoming a serious challenge in case of observing large samples, such as histological sections [3], as well as large-scale screening in drug discovery research [4]. For example, it is known that the specific structure of cells present in the glomerulus is lost in the kidney disease

Table 4 Number of glomerulus detected by NIS.ai and inference accuracy

Ground Truth	NIS.ai	False Negative	False Positive
97	103	1 (1.0%)	7 (7.2%)

nephrotic syndrome [5]. To detect such differences in characteristics, a microscope user needs to use a high-magnification objective lens and acquire images after visually confirming the position of specific structures successively, which places a huge burden on image acquisition and analysis tasks. By incorporating region segmentation by NIS.ai into the image acquisition steps with NIS-Elements, it is possible to narrow down image acquisition regions to specific structures, and thus reduce the burden required of users when conducting image acquisition and analysis. In this section, we present an example of automatically detecting and acquiring the glomerulus images from a kidney section sample.

For verification, we adopted a sample in which a mouse-derived kidney section was treated with Elastica-Masson stain, which tags connective tissues such as elastic and collagen fibers. A Ti2-E inverted microscope equipped with an A1R confocal microscope system was used for image acquisition. For image acquisition, we used a 20x objective lens (CFI Plan Apo Lambda 20X 0.75NA, Nikon, Japan) and CMOS camera (ORCA-Fusion, Hamamatsu Photonics, Japan) to detect the position of the glomerulus. For detailed structural observations, we used a 100x objective (CFI SR HP Apo TIRF 100XC Oil. 1.49NA, Nikon, Japan) and confocal microscope system. We prepared training data, showing a region of the glomerulus in 78 images cut out from full images of two kidney sections obtained by joining multiple images taken with a 20x objective lens. In addition, we constructed a model to output the glomerulus region, according

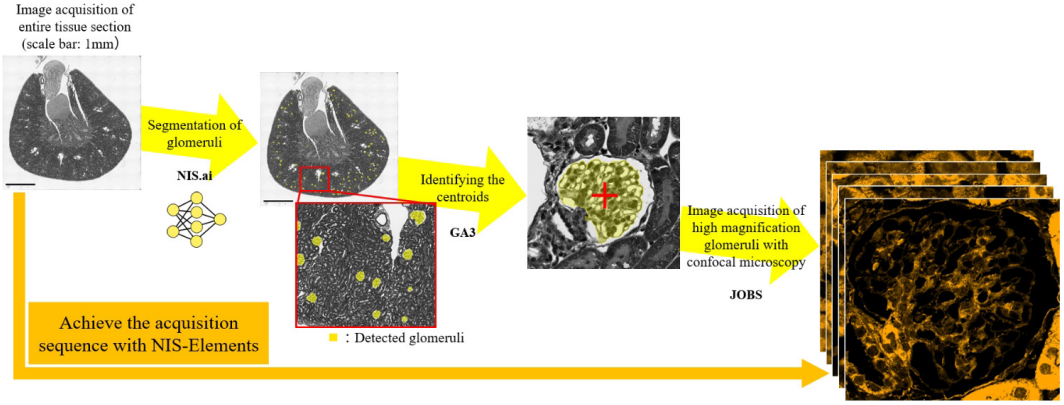


Fig. 6 Overall image of automated workflow from kidney glomerulus detection to high magnification image acquisition.

to the bright field image input from the Segment.ai function (Table 2, Section 3.3 items).

Fig. 6 illustrates a workflow for automating the steps from bright field image acquisition to glomerulus fluorescent image acquisition: 1) construction of kidney total image via bright field image acquisition and stiching, 2) inference of the glomerular region (Segment.ai), 3) calculation of glomerulus center-of-gravity coordinates, and 4) acquisition of glomerulus fluorescent images using a confocal microscope system.

Based on the obtained results, there were a few false positives of the glomerulus region with NIS.ai processing alone; however, only the glomerulus regions were accurately detected by incorporating filter processing, based on area and circularity in GA3 (Fig. 3, red frame). Furthermore, compared to the manual glomerulus detection results, false negatives or positives in NIS.ai processing remained at 1% or 7.2% for each.

From the above, it was demonstrated that regions that are difficult to detect via existing image processing can be accurately detected by NIS.ai. In particular, the lack of false negatives is crucial because in the case of false positives, captured images can be selected later; however, false negative regions need to be re-acquired images. In particular, missing regions that should be detected in clinical testing (for example, detection of cancerous areas) can be fatal. Furthermore, if NIS-Elements is used in combination with NIS.ai to set a sample in a microscope and commence image acquisition, a large number of images can be acquired by the experimenter (user) without being constrained in front of the microscope; hence, drastic labor saving scans can be expected. It is expected that in the future, the acquisition and analysis of large amounts of data for the purpose of quantification will increase in the fields of biology, medicine, and drug discovery. Therefore, the usefulness of automatic image acquisition and image processing workflows utilizing NIS.ai will increase.

4 Conclusion

NIS.ai is capable of high-precision digital staining and segmentation, and is effective in applied studies, such as in basic and clinical researches. Although not addressed in this paper, NIS.ai also provides functions for eliminating noise in microscope images and out-of-focus fluorescence leakage. Owing to the combination of these functions and future functional extensions, we further aim to contribute to improving user research efficiency and the discovery of novel findings by providing digital image processing and quantification results that cannot be realized by microscopes alone.

We are grateful to the assistant professor Matsui of the Department of Nephrology, Graduate School of Medicine, Faculty of Medicine, Osaka University, for providing us with mouse kidney-section samples.

References

[1] C. Belthangady and L. A. Royer, “Applications, promises, and pitfalls of deep learning for fluorescence image reconstruction,” *Nature Methods*, vol. 16, pp. 1215–1225, 2019.

[2] D. W. Siemann and P. C. Keng, “Cell cycle specific toxicity of the Hoechst 33342 stain in untreated or irradiated murine tumor cells,” *Cancer Research*, vol. 46, pp. 3556–3559, 1986.

[3] O. Zhanmu, X. Yang, H. Gong and X. Li, “Praffin-embedding for large volume bio-tissue,” *Scientific Reports*, vol. 10: 12639, 2020.

[4] J. R. Lin, M. F. Sichani and P. K. Sorger, “Highly multiplexed imaging of single cells using a high-throughput cyclic immunofluorescence method,” *Nature Communications*, vol. 6: 8390, 2015.

[5] B. Zhu, A. Cao, J. Li, J. Young, J. Wong, S. Ashraf, A. Bierzynska, M. C. Menon, S. Hou, C. Sawyers, K. N. Campbell, M. A. Saleem, J. C. He, F. Hildebrandt, V. D.Dagati, W. Peng and L. Kaufman, “Disruption of MAGI2-RapGEF2-Rap1 signaling contributes to podocyte dysfunction in congenital nephrotic syndrome caused by mutation in MAGI2,” *Kidney International*, vol. 96, pp. 642–655, 2019.

門井宏平　Kohei KADOI
ヘルスケア事業部 技術統括部 システム開発部
System Development Department
Technology Solutions Sector
Healthcare Business Unit

武居俊輔　Shunsuke TAKEI
ヘルスケア事業部 技術統括部 システム開発部
System Development Department
Technology Solutions Sector
Healthcare Business Unit

畑口剛之　Takeshi HATAGUCHI
ヘルスケア事業部 技術統括部 設計部
Designing Department
Technology Solutions Sector
Healthcare Business Unit

門井宏平
Kohei KADOI

大規模空間非接触計測器 APDIS MV4x0 レーザーレーダーの開発

Anthony Slotwinski，森本　樹，Ghassan Chamsine，Pawan Shirur，Eric Brandt，荒井正範

Development of the APDIS MV4x0 Laser Radar Large-Volume, Non-Contact Measuring System

Anthony SLOTWINSKI, Shigeru MORIMOTO, Ghassan CHAMSINE, Pawan SHIRUR, Eric BRANDT and Masanori ARAI

自動車業界の生産現場の要求は高度化し，信頼性はもとより高性能で高速な計測技術が求められている．それらの要求にこたえるべく，ニコンの光学コア技術とニコンメトロロジーのレーザー計測技術を融合した「APDIS シリーズ」を2020年10月にリリースした．本稿では本製品 APDIS の基礎原理となる部分に加え，APDIS シリーズとして実現した新たな開発要素について説明する．

The demands of production line metrology in the automotive industry are becoming greater and more challenging. High accuracy and high-speed dimensional part measurement as well as high reliability are key requirements for inspection. In order to meet these high expectations, we released the APDIS series in October 2020, which combines both Nikon's core optical technology and Nikon Metrology's laser radar technology. This paper describes the new technical features we developed and implemented in APDIS in addition to the product's basic principal of operation.

Key words 三次元計測機，非接触，周波数変調，レーザーレーダー，自動車部品検査
coordinate-measuring machine (CMM), non-contact, frequency modulated continuous wave (FMCW), Laser Radar, car body inspection

1 Introduction

In October 2020, Nikon released the APDIS MV4x0, the next generation of the Nikon Laser Radar. The MV430 and MV450 measurement systems are used for fast, automated and non-contact inspection of objects ranging from smaller components such as a car door to complete large assemblies such as commercial aircraft. It achieves this through a unique application of a non-contact, accurate laser-based measurement technology overcoming the limitations of traditional monolithic or portable metrology systems.

The ability to measure detail at a distance, without the need for handheld probes, targets or surface preparation means APDIS is ideally suited for repetitive, complex, hard to reach, delicate and labor-intensive inspection tasks, covering a huge range of manufacturing, industry and research applications. This paper describes various innovative features that are incorporated into the system.

2 Development Background

The heart of the APDIS system is a Frequency Modulated Continuous Wave (FMCW) coherent laser radar that uses a diode laser as its source. A waveform is used to change the frequency of the laser directly by modulating the laser's injection current, resulting in linear modulation. This type of modulation is often referred to as a chirp. The frequency can be expressed as a function of time in the following manner:

$$f(t) = f_0 + (\Delta f/\Delta t)t \quad (1)$$

Where f_0 is the center frequency of the laser. By using a lensing system, the modulated beam is focused at a target, where it is scattered and recollected by the optics after a round trip transit time t. The distance to the target, R, is calculated using the relationship:

$$\tau = 2R/c \quad (2)$$

where c is the velocity of light.

In an FMCW laser radar device, a portion of the transmitted beam is split from the incident light wave and forms the

local oscillator (LO), which is then mixed with the returned (signal path) energy. In a coherent laser radar, the beat frequency produced will be equal to:

$$Beat\ Frequency = f(t) - f(t+\tau) = (\Delta f/\Delta t)\tau \quad (3)$$

The beat frequency is measured electronically and used to calculate the distance to the target R:

$$R = c * Beat\ Frequency/(2(\Delta f/\Delta t)) \quad (4)$$

In early Frequency Modulated (FM) devices, the accuracy of range measurement was limited by the linearity of the frequency modulation over the counting interval. For example, if the target is one meter distant, a linearity of one part per thousand is necessary to ensure 1 mm accuracy. Advanced techniques employed in the Nikon Laser Radars enable a high degree of linearity. In addition, these techniques can detect and compensate for real time variances from linearity. This enables range measurements with single digit micron precision.

FMCW radars are largely immune to ambient lighting conditions and changes in surface reflectivity because they rely only on the beat frequency, which is not dependent upon signal amplitude, to calculate range. This enables the system to make reliable measurements with as little as one picowatt of returned laser energy. This corresponds to a nine order-of-magnitude dynamic range in sensitivity.

Fig. 1 depicts the linear frequency modulation, or "chirp", together with the corresponding "beat" frequency that results from combining the outgoing and incoming light signals. The laser base frequency is approximately 200 terahertz. The "beat" frequency is in the 10 MHz range. If the surface being measured is moving relative to the laser light source, the beat frequencies corresponding to laser upsweeps will be different from the beat frequencies corresponding to the downsweeps, due to Doppler frequency shifting. Measuring the frequency difference between these signals enables a determination of velocity to be made.

For precision measurements it is necessary to include a reference standard both for absolute ranging accuracy and

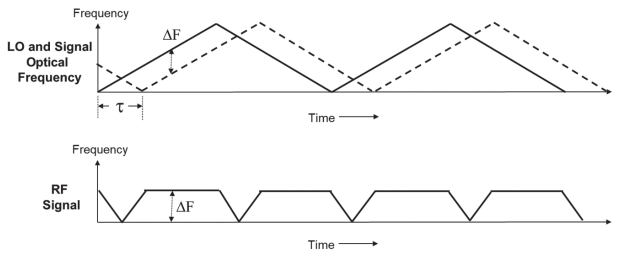


Fig. 1 Laser optical frequency and heterodyned RF signal of coherent laser radar.

to help linearize the laser's chirp waveform. For the internal reference standard, the light from an IR laser is split into two fibers by means of a fiber coupler (See Reference Arm in Fig. 2.). One path is sent to the mixing and focusing optics and used to measure range. The other path is directed to the reference standard that consists of an input fiber optic coupler, which splits the light into two fiber paths and an output fiber coupler, which recombines the light into a single fiber. The two paths of fiber between the two couplers are mismatched in length by several meters such that a laser radar signal is detected on the reference detector. This reference Mach Zehnder interferometer formed by the two couplers and the fiber between them is kept in a temperature-controlled container to prevent the fiber lengths from changing.

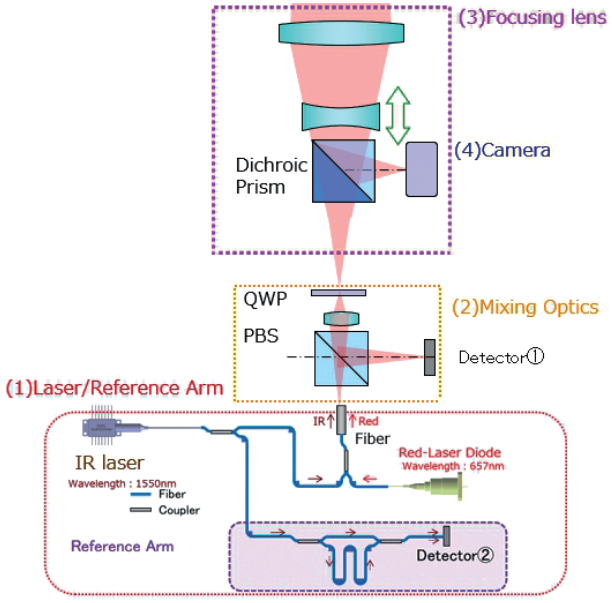


Fig. 2 Optics overview

The APDIS system has several improvements over previous versions of the laser radar. The first is the incorporation of a compact mixing optics assembly that is insensitive to environmental changes (temperature, etc.). Previous laser radars used a fiber optic circuit to perform the mixing function. This circuit was more costly and needed to be thermally controlled to maintain measurement stability. The second improvement is the inclusion of a high-resolution video camera in a boresighted, confocal relationship with the radar's measurement beam. The radar and the camera utilize a common focusing system allowing for a precise relationship between the two sensors. Another improvement is the physical sealing of the system to prevent water and dust egress. This is especially important in use in a factory environment. Finally, the capability to remotely measure velocity (vibrations) as well as range was incorporated to provide

additional data to the user. These improvements are described in more detail below.

3 Optical Design

The optical system of the APDIS laser radar consists of (1) the light source/reference arm unit, (2) the mixing optics or interferometer unit, (3) the focusing lens and (4) boresighted camera (Fig. 2).

The light source unit uses a 1550 nm diode laser to produce the frequency-modulated measurement light. This light is split between the reference interferometer discussed above and the measurement path. In addition, light from a red laser diode is combined with the IR laser light to provide the user with a visible indication of the output beam position. This red spot, when focused on a target, can be seen visually by the user both in the video display and on the actual target.

4 Mixing Optics (Interferometer Unit)

The mixing optics of the APDIS laser radar eliminates many of the fiber optic components used in previous systems. The miniature mixing optics are shown in Fig. 3. The laser light is transported to the mixing optics via a Polarization Maintaining (PM) fiber. The laser light exits the fiber and passes through a polarizing beamsplitter (PBS) which is fabricated to pass one linear polarization state and reflect the other polarization state. The light emitted from the fiber is in the linear polarization state that passes through the PBS. The light then passes through a focusing lens and then through a quarter wave plate retarder (QWP). The QWP is aligned such that the linearly polarized light is converted to circularly polarized light (for example right circularly polarized light.) In addition, the QWP has an antireflection coating on the first surface (the surface first encountered by the beam) and a partially reflective coating on second surface (the surface encountered by the beam after it has passed through the QWP). The QWP is also positioned such that the focused spot coincides with the partially reflective surface. The light that is reflected back becomes the LO path beam while the light that passes through the QWP becomes the signal path beam. This beam is then focused onto a target to be measured via the focusing optics. The LO path light, upon reflecting from the QWP surface, switches its circular polarization state to left circularly polarized light and then is converted by the second pass through the QWP to an orthogonal linear polarization state. This light is refocused by the focusing lens and is reflected by the PBS to a photodetector. Similarly the signal light that reflects off a target also undergoes a change in circular polarization state to left circularly polarized and, after passing back through the QWP, also is converted to the orthogonal linear polarization state and is reflected by the PBS to the sensor at which point it optically mixes with the LO light to generate the radar signal. The combination of the focusing lens and the QWP forms a Cats-eye retroreflector which ensures alignment between the LO and signal paths and provides for a stable optical configuration. Also, since the LO path is a subset of the signal path, any drift in the position of the mixing optics affect both paths equally, eliminating the temperature sensitive drift problem of the previous systems. In older systems, the mixing interferometer is fabricated from fiber optic components and housed in a temperature-controlled module, which hinders miniaturization. Therefore, this cat's eye arrangement is used to reduce the number of parts and realize miniaturization.

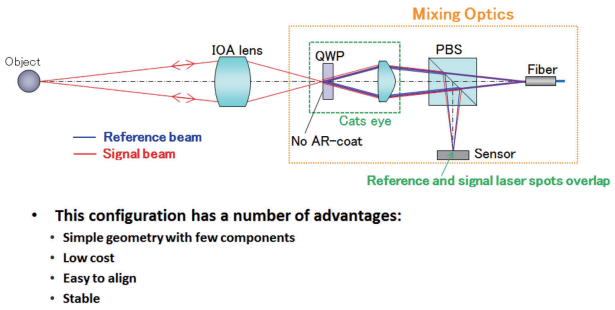


Fig. 3 Mixing Optics.

As to the stability of such an assembly, Fig. 4 shows the dependence of the laser spot position on the sensor as a function of QWP tilt and offset. As can be seen, if the QWP surface coincides with the image plane of the lens, the spot position is independent of QWP angle.

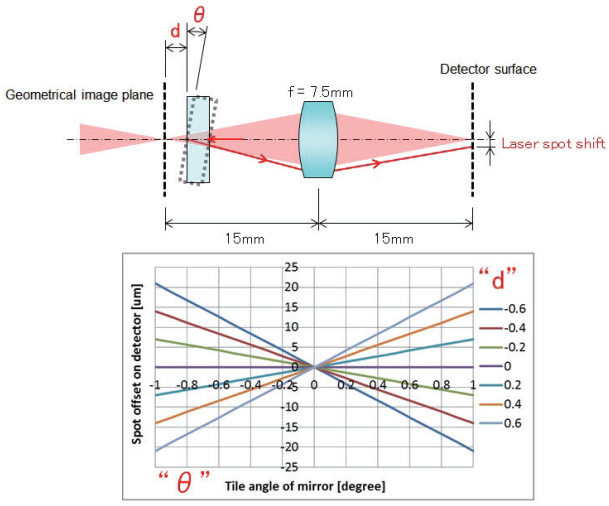


Fig. 4 Sensitivity of the mixing optics to misalignment.

Fig. 5 shows the effect of beam shift between the signal and the LO spots on the laser radar signal. As can be seen,

a 7um beam shift produces a 1.5 db loss in signal level. This is an acceptable loss for such measurement systems. This is the same loss that would be produced by a 0.4 mm QWP offset combined with a 0.6 degree QWP tilt. The manufacture and assembly of an optical mixing assembly to within these tolerances is easily accomplished.

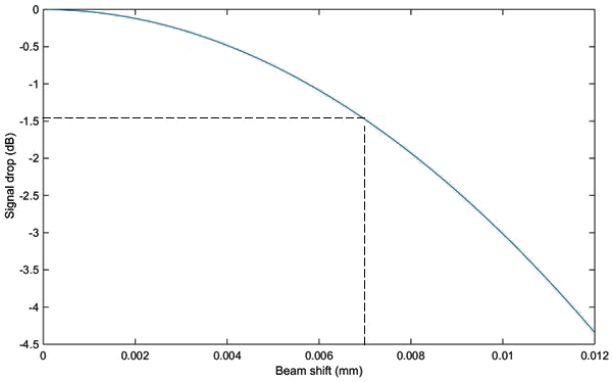


Fig. 5 Laser Radar signal loss due to beam shift.

5 Objective Lens and Boresight Camera

One of the improvements of the APDIS laser radar over previous models is the reduction of the minimum measurement range from 1 meter to 0.5 meters. This reduction in stand-off distance improves the usability of the instrument in factory assembly line settings where space is at a premium. The objective, or focusing, lens developed for APDIS is an optical system that supports this reduced range measurement. In addition, it also supports both the IR light for measurement and the visible camera wavelength range. In the past, the focusing optical system of the measurement light and the boresight camera were separate optical systems, and the optical paths were separated by a dichroic mirror. Therefore, since each optical system had its own focusing mechanism, it was impossible to match the measurement point with the camera field of view. In order to overcome this drawback, the APDIS lens has a dichroic prism placed between the focusing lens and the intermediate image plane (QWP surface) to separate the visible wavelength and the measurement wavelength. Since the focusing mechanism is shared by this configuration, the measurement point and the field of view of the boresight camera can be matched. Therefore, measurement support using object recognition based on camera images has become achievable.

It is possible to irradiate the measurement light and the red laser light for visual recognition from the fiber end, but if the axial chromatic aberration is completely corrected, the spot of the red light becomes too small and the visibility deteriorates, so by intentionally leaving the axial chromatic aberration, visibility is ensured. Another characteristic point is the dichroic prism. A general dichroic prism is a cube type with a slope of 45 degrees, but the APDIS dichroic prism has a slope at an angle of 30 degrees with respect to the incident optical axis (Fig. 6). As a result, the angle range of incident on the slope can be reduced, so that the polarization dependence of the transmittance/reflectance in the visible light region of the dichroic film can be reduced. If the polarization dependence of the dichroic film is large, the appearance of the camera image will change because the reflectance changes depending on the polarization direction due to the change in the angle of incidence on the deflection mirror when observing points with different elevation/azimuth angles. In order to suppress this problem, a prism with a special shape (Fig. 6) was adopted.

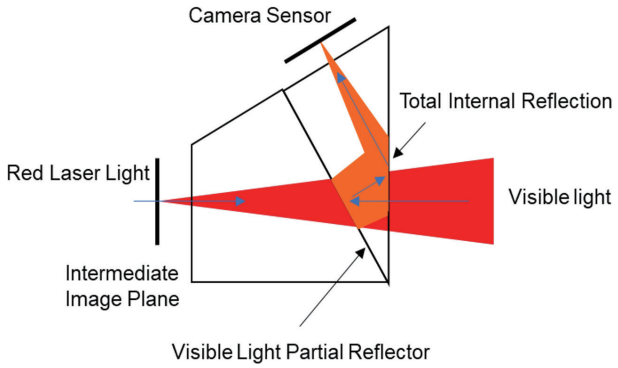


Fig. 6 Dichroic prism.

6 Camera Resolution

In past versions of the Laser Radar (LR) a collinear camera was in the unit with the following limitations: (a) the camera was an inexpensive surveillance camera whose calibration parameters changed with orientation to gravity and was environmentally sensitive (temperature), and (b) architecturally the camera and the LR data were brought together at too high a level to allow real time coordination between the camera and the LR data. The APDIS Laser Radar incorporates a metrology camera which (a) shares the same focusing optics with the LR measurements creating a confocal Laser Radar (cLR) and metrology relationship (6DOF) between the camera and the Laser Radar, (b) is high definition, and (c) is now tied to the LR data at a low level in the architecture minimizing the latency between the data allowing real time coordination of the LR and camera data.

The APDIS video camera is a 4.2 Megapixel (2048x2048) CMOS sensor in an RGB Bayer Matrix. As shown in Fig. 7, this sensor covers a 7degree field-of-view (FOV). However, this FOV is truncated to 5 degrees to remove the edges of

radar's circular output aperture so the effective camera resolution is about 1400x1400 pixels. Each pixel covers 63uradians of the image scene. The camera sensor mount was designed to maximize the stability between the Laser Radar and the video image. The APDIS system is vastly superior in this respect as compared to previous systems. APDIS retains sub-pixel stability in alignment regardless of the radar orientation. This is important as the system is often robot mounted in order to facilitate factory measurements. In addition, the alignment is stable to less than 2 pixels over the 5℃ to 40℃ operational temperature range of the radar.

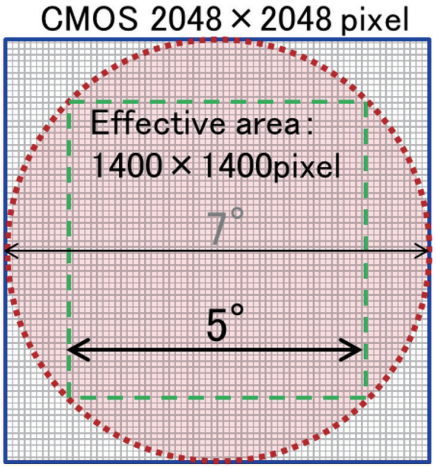


Fig. 7 Camera FOV.

Using the two measurement modes (LR and camera) in a confocal Laser Radar allows the LR to be pointed to optimally measure the feature of interest. Additionally, a low latency data interface allows real time algorithms and the tracking of features identifiable in the camera.

A calibrated camera can be viewed as an angle measurement device where the azimuth and elevation of every pixel in the picture can be determined. With the LR and the camera having a confocal relationship it simplifies the use of the range measurement to provide scale to the photo. This relationship allows the center pixel of the camera to be directly related to XYZ_{LR}. It cannot be guaranteed that the projection of the camera focal plane to the scene is perpendicular to the central ray of the Laser Radar. This relationship can be determined through a calibration process. With the calibrated camera planer features can be measured directly by the camera once the range is determined by the Laser Radar. The same can be said for features with known geometry, such as spheres.

7 IP54 Rating

The Protection against Ingress of fluids and solids into instrument packaging is a key feature of the APDIS system because of the harsh industrial environments required by automotive applications. The protection rating of 'IP54' was found adequate for these requirements. These ratings are determined as defined by industry standards defined by IEC 60529:1989+AMD1:1999+AMD2:2013 CSV - Degrees of protection provided by enclosures (IP Code) . The first number identifies protection against solids like dust and metal debris and the second number identifies protection against fluids like water. Fig. 8 describes rating derivation for IP54 rating guide.

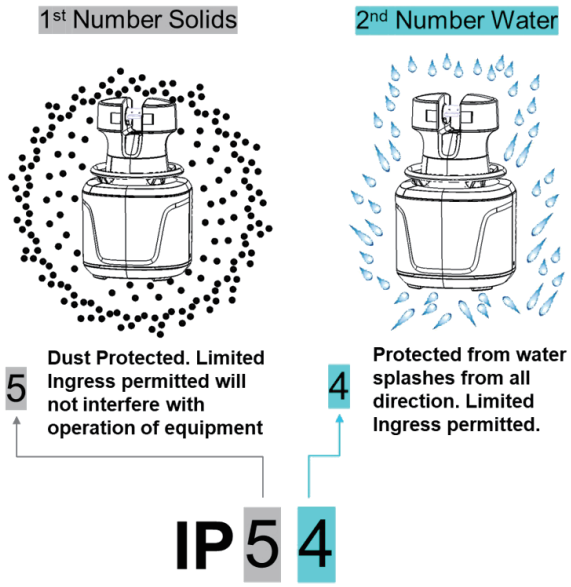


Fig. 8 IP54 Rating definition.

The design strategy for achieving the IP54 rating was developed based on type of interfaces across the APDIS design. (Fig. 9)

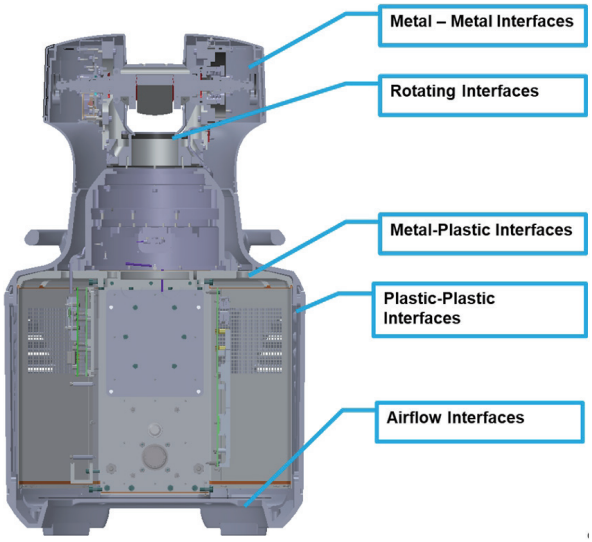


Fig. 9 APDIS cross-section showing various interfaces.

There were 3 main strategies for these interfaces.

a) Metal to metal interfaces used a technique to minimize

interface gaps to less than 1mm and application of hydrophobic coating to prevent ingress of water and dust. Labyrinth structures were also used to create a sufficient seal. Rotating interfaces on the APDIS pointing mirror assembly had very small interface gaps augmented by light weight lubricant grease to form a hydrophobic and dust barrier.

b) Plastic to metal and plastic to plastic interfaces used a variety of custom gaskets strategically applied to prevent or minimize ingress of water or dust. Some of the materials were open/closed cell foams or EPDM type material. Fasteners with O-ring were also used at critical locations (Fig. 10). The selection of gasket material was particularly impacted by interface geometry, fastening methods and location on APDIS.

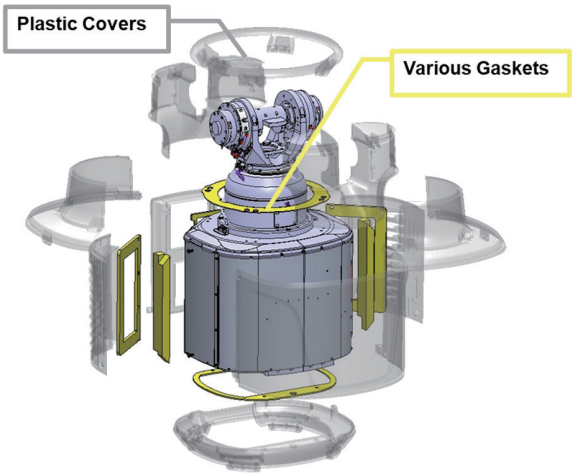


Fig. 10 Exploded view of external covers and various custom gaskets.

c) Interfaces where air flow for thermal management was require, water ingress was not completely avoidable. Here special vent panel materials and geometry which are hydrophobic were used. Strategically placed vent holes that facilitate purging of impressed fluids in only one direction.

Another strategy used was to incrementally design and test (in collaboration with the industrial design team responsible for external covers) using internal test methods that incorporated water detection paper and glycol-based smoke generator to detect ingress and leakages in interfaces.

The APDIS system tested showed no ingress dust ingress during IEC 60429 IP5X testing in locations which were sealed using techniques implemented above (Fig. 11).

APDIS system tested had minimum ingress into expected areas. The purge holes, and vent panels performed well to keep gasketed interfaces intact and prevent ingress of water during IEC 60529 testing for IPX4 standard (Fig. 12).

Fig. 11 APDIS after IP5X test.

Fig. 12 ADPIS after the IPX4 testing.

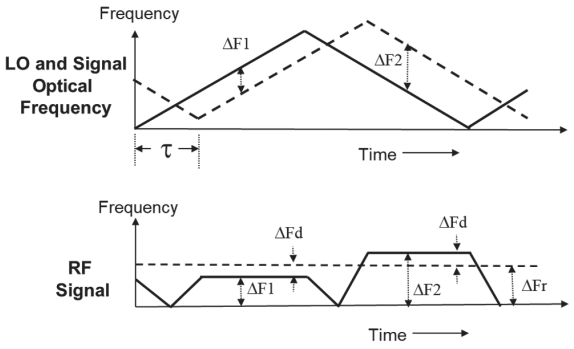
Ingress Protection is a key requirement for Automotive shop floor metrology where fluid splashes, welding and other process fumes and dust present hazards to measuring equipment and impact productivity and down time. The APDIS design strategies render the system very robust against such hazards by achieving the IEC 60529 IP54 rating.

8 Vibration Measurement Feature

As mentioned previously, if the surface being measured is moving relative to the laser light source, the beat frequencies corresponding to laser upsweeps will be different from

the beat frequencies corresponding to the downsweeps as shown in Fig. 13, due to Doppler frequency shifting. In a manufacturing environment vibrational noise is often present due to large motors, vehicles and other noise sources. By averaging the measured signal frequency from an upsweep with an adjacent down sweep, the Doppler noise can be removed from the range measurement. The APDIS Laser Radar is capable of making 2000 Doppler corrected range measurements per second. In addition, the Doppler frequency, and thus the velocity, can also be determined by taking the difference between the two frequencies. For velocities in the direction of the laser beam, the Doppler frequency is given by

$$F_d = (2*velocity)/1550\,nm. \quad (5)$$



Range Frequency = Fr = (ΔF1+ΔF2)/2
Doppler Frequency = Fd =|ΔF1-ΔF2|/2

Fig. 13 Effect of moving target.

Prior laser radar systems that performed this Doppler correction did not provide any velocity information. In the APDIS development, this capability was included. Thus, the APDIS system can be used to monitor the vibrational signature of milling machines and other tools in a non-contact manner. Such information can be used to provide feedback to the machining process and to monitor the general health of the machine.

Fig. 14 shows a comparison of the time series of the output of a conventional accelerometer attached to a milling machine and to the APDIS measured acceleration (derived from the measured velocity). The mill was cutting an aluminum block while the measurements were taken.

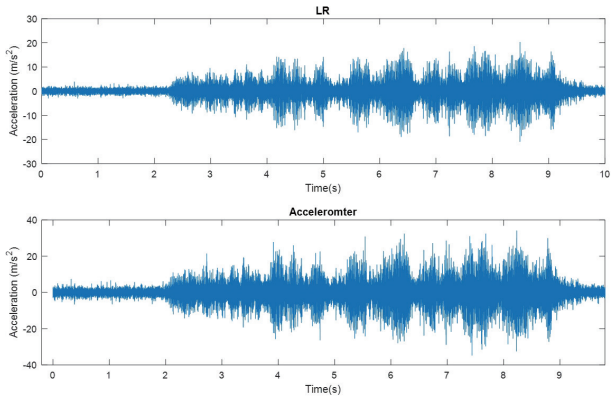


Fig. 14 Comparison of APDIS and conventional accelerometer.

9 Conclusion

The APDIS 4x0 Laser Radar was developed to operate in an industrial factory environment without loss of functionality or precision. It combines an FMCW laser radar with a high-resolution video camera in a stable, confocal configuration to provide for a unique metrology tool. Its sealing against dust and liquids provides for the ability to operate in a harsh manufacturing environment. It is also capable of operating as a laser vibrometer in order to monitor industrial machines.

Anthony SLOTWINSKI
Nikon Metrology, Inc.

森本　樹　Shigeru MORIMOTO
Nikon Metrology, Inc.

Ghassan CHAMSINE
Nikon Metrology, Inc.

Pawan SHIRUR
Nikon Metrology, Inc.

Eric BRANDT
Nikon Metrology, Inc.

荒井正範　Masanori ARAI
光学本部　開発戦略部
Strategic Technology Development Department
Optical Engineering Division


Anthony SLOTWINSKI


森本　樹
Shigeru MORIMOTO


Ghassan CHAMSINE


Pawan SHIRUR


Eric BRANDT


荒井正範
Masanori ARAI

「C3 eMotion」インテリジェントアクチュエータユニットの開発

渡邉昭宏，高橋喜二，湯本一樹，引地哲也

Development of the C3 eMotion, Intelligent Actuator Unit

Akihiro WATANABE, Yoshiji TAKAHASHI, Kazuki YUMOTO and Tetsuya HIKICHI

ロボットの多様化・高機能化ニーズにこたえるべく，モーター，減速機，駆動回路，ブレーキ，アブソリュートエンコーダをパッケージ化した，インテリジェントアクチュエータユニット「C3 eMotion」を2020年 4 月にリリースした．「C3 eMotion」は，2 つのアブソリュートエンコーダを組み込んだ「ダブルエンコーダ構造」を採用．独自の制御技術との組み合わせによる様々な機能を有する．本稿では，「C3 eMotion」が持つ技術的特徴について解説する．

In April 2020, Nikon released the intelligent actuator unit, “C3 eMotion,” that packages a motor, reducer, drive circuit, brake, and absolute encoders to meet the needs for diversification and high functionality of robots. The “C3 eMotion” includes Nikon’s original “double-encoder arrangement,” which incorporates two absolute encoders, and various functions that combine the original control technology with the double-encoder arrangement. This article describes the technical features of the “C3 eMotion.”

Key words ロボット用関節ユニット，ダブルエンコーダ構造，高精度位置決め，トルク検出，制御技術
robotic joint unit, double-encoder arrangement, high accuracy positioning, torque detection, control technology

1 Introduction

The modularization of robot components to strategically accelerate their development is being discussed in the robot industry [1]. It is believed that these efforts will facilitate freely conceived robot configurations, and help satisfy diversification and high-functionality requirements. In this paper, we elucidate the features and technical elements obtained by applying the double-encoder arrangement and original control technology of the “C3 eMotion [2]” (Fig. 1 C3 eMotion), an intelligent actuator unit, which was released in April 2020 to meet these needs.

Fig. 1　C3 eMotion

2 Hardware configuration

Conventionally, to configure the joints of an articulated robot, it is necessary to separately prepare each of the constituent element components, such as motors, encoders, reduction gears, and brakes, and then incorporate them into a robot base and robot arm. In addition, reduction gear and motor output characteristics, including encoder specifications, need to be designed according to the operating speed and inertia of the robot; consequently, a high level of technical skill is required to freely manufacture robots. Therefore, constructing a robot system still involves the purchase and use of robots manufactured by dedicated robot manufacturers. However, robot applications are becoming diversified; hence, there is a significant increase in the global needs for freely configurable robots, which are not exist.

C3 eMotion was released with the intention of realizing three objectives: “connect,” “control,” and “cooperate.” To facilitate freely-conceived robot configurations, an all-in-one package has been adopted by C3 eMotion, which combines a motor, reducer, brake, absolute encoder, and motor driver board in a single housing (Fig. 2). This packaging eliminates the need for the design and assembly of individual compo-

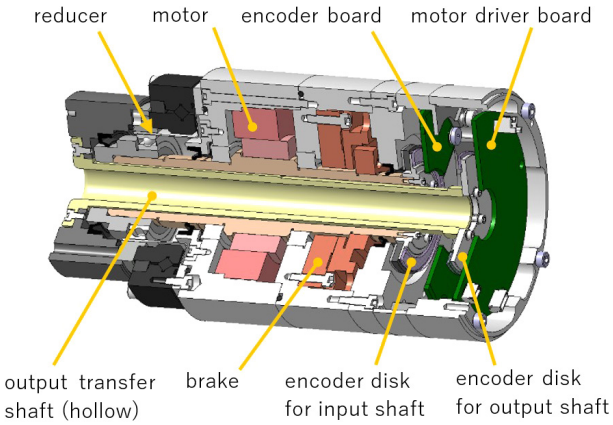


Fig. 2 Internal configuration of C3 eMotion

nents when configuring a robot, and allows the C3 eMotion to be treated as a single robot joint module. In addition, because this construction has the closest positioning of the motor driver board to the motor, it does not require power cable wiring to drive the motor within the mechanism of the robot, which reduces wiring, and is also advantageous in addressing noise emissions.

Furthermore, in addition to the absolute encoder for the input shaft (motor drive shaft), C3 eMotion also employs “double-encoder arrangement” with a high-resolution absolute encoder on the output shaft. Motor rotations are decelerated by the reducer and transmitted to the output shaft. The rotation information of the output shaft is transmitted through the hollow interior of the motor shaft by an output transmission shaft, and is read by an encoder mounted on the end face opposite the output shaft. The output shaft encoder and the encoder reading the rotation information of the input shaft are installed next to each other. Their rotation information is transmitted to the motor driver board via serial communication. The motor driver board is located in the immediate vicinity of the encoder, and it controls the motor rotation and brake based on the command values from a Master controller, as well as the information from the two encoders. The placement of the encoder and driver board minimizes the length of the encoder’s communication cable, and keeps electronic boards (e.g. the encoders) away from the grease-containing reducer that shock and vibration from the robot arm is applied directly.

Fig. 3 shows the system configuration of the C3 eMotion. Because C3 eMotion can be adopted as a robot joint module, it facilitates the configuration of a robot by connecting multiple C3 eMotion units with a robot arm. The communication between the motor driver board and Master controller is handled by the EtherCAT® [3] serial communication corresponding to the CiA402 drive profile. EtherCAT® is capable

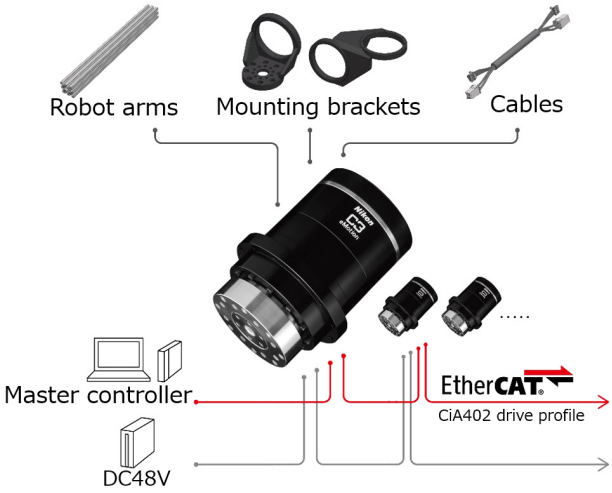


Fig. 3 System configuration

Table 1 Main Specifications

Model No.		IAU-15 (under development)	IAU-30 (under development)	IAU-60	IAU-200	IAU-300 (under development)
Power supply voltage	V	48				
Instantaneous maximum torque	N·m	4.8	30	55	200	400
Rated torque	N·m	2.4	10	30	95	130
Rated speed	min^{-1}	35	30	20	15	15
Maximum speed	min^{-1}	60	40	40	20	20
Reduction ratio	–	100	81		101	
Encoder *Input shaft only for IAU-15		Input shaft: 24 bits/rotation, 16 bits for multi-turn Output Shaft: 24 bits/rotation				
Output shaft encoder accuracy	arc-sec	–	± 15 or less			
Torque detection range (F.S.)	N·m	–	14	30	130	280
Torque detection accuracy	%	–	±5 F.S.			
Driver/Communication Method	–	Built into actuator/EtherCAT® CiA402				
Outer diameter	mm	ϕ50	ϕ70	ϕ80	ϕ110	ϕ142
Overall length	mm	73.4	161.1	164	179.7	185
Weight	kg	0.6	1.6	2.1	4.9	8.5

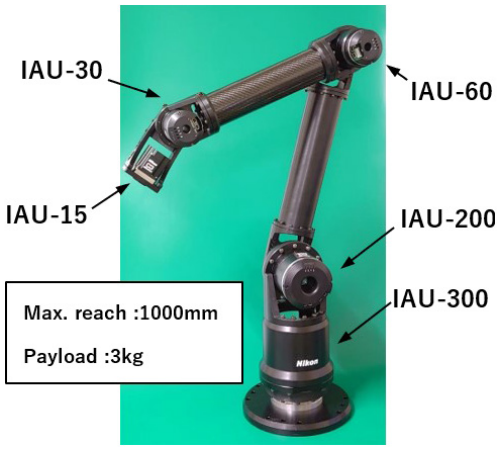


Fig. 4 Example of 5-axis robot using C3 eMotion

of daisy-chain connection, and is suitable for the multiaxial connection of C3 eMotion via a robot arm. Table 1 presents the main specifications of C3 eMotion. Presently, two versions of C3 eMotion, IAU-60 and IAU-200, have been released, and three other versions, IAU-15, IAU-30, and IAU-300 are under development. Fig. 4 shows an example of an arm robot configuration, using these five types.

This robot has a long-reach design of 1000 mm, although it has a low payload, as well as features not seen in commercially available robots, such as using plastic for the arm.

3 Double encoder structure

Angular transmission errors triggered by machining errors or friction are present in reducers that employ gears. C3 eMotion adopts the strain wave gear reducer widely employed in robots and other industrial equipment. In addition to angular transmission errors, the strain wave gear reducer has nonlinear friction and spring characteristics, including structurally generated hysteresis [4]. In conventional semi-closed control systems, these errors and spring characteristics appear as substantial position errors at the tips of robots. Manufacturers specializing in robots have improved the rigidity of robot arms compared to standardized robots, and have adopted techniques for estimating and correcting robot tip position errors to improve position accuracy [5]. Such high-precision technologies are difficult to apply, for example, when constructing freely-configured robots combining robot constituent elements in which system integrators have been modularized.

As aforementioned, C3 eMotion employs “double-encoder arrangement” with an encoder on the output shaft side (Fig. 5). An absolute encoder with a 24-bit resolution is employed on the output shaft encoder. An accuracy of ± 15 arc-sec or less is ensured via error compensation. Improved positioning accuracy is expected because the impacts of Angular transmission errors and nonlinear spring characteristics can be eliminated by adopting a control system, using output shaft rotation information.

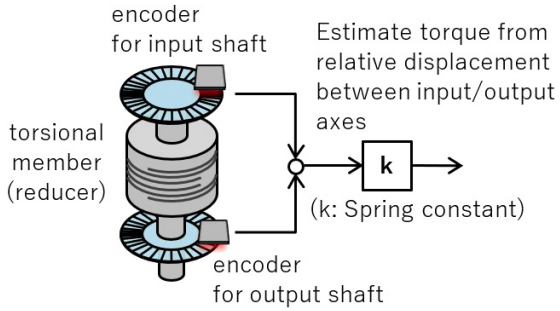


Fig. 5 Double-encoder arrangement

In addition, because the reducer exhibits nonlinear spring characteristics and can be regarded as a torsional elastic body, the torque can be easily measured by measuring and converting the relationship between the magnitude of torsion and torque beforehand.

4 Original control technology

In applications typified by laser welding, such as vehicle body welding and electrode connection of battery modules installed in electric and hybrid vehicles, there is an increasing demand for highly accurate TCP*1 positioning in robots; hence, the demand for high accuracy in the TCP*1 positioning of robots is steadily increasing.

Nikon has already cultivated highly advanced and precise control technologies, including the ultra-precise and accurate device stage-control typified by a semiconductor and flat panel display (FPD) lithography equipment. C3 eMotion performs motor control by providing position, speed, and torque commands from an upper level controller via EtherCAT® communication, and it has been equipped with an original control suited to it by applying the control technology developed by Nikon.

As demonstrated in the “double-encoder arrangement” optimally designed for C3 eMotion in the previous section, the original high-precision positioning control is characterized by adopting a control algorithm to properly combine two types of position information according to operating conditions: information from an encoder for detecting the motor drive shaft position (reducer input side encoder) and information from an absolute encoder for detecting the position of the C3 eMotion output shaft, which is the reducer output (reducer output shaft encoder; Fig. 6).

This proprietary Nikon mix control makes it possible to combine the advantages of both full-closed and semi-closed controls, thereby leading to improved positioning accuracy. In fully-closed control, the feeding back rotation information read by the reducer output shaft encoder facilitates the elimination of error factors triggered by the reducer, such as reducer angular transmission errors, as well as reducer twisting originating from load fluctuations owing to the weight of the robot arm and its change in posture. In con-

*1 Tool center point. Center when controlling an end effector.

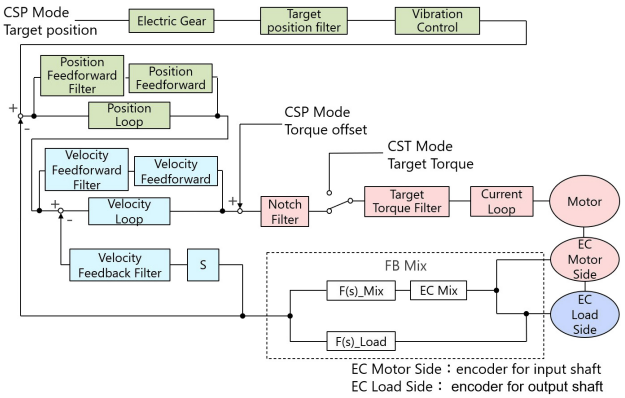


Fig. 6 Control block diagram

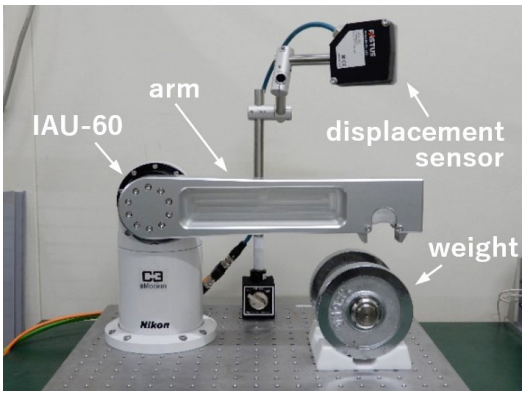


Fig. 7 Positioning performance verification device

trast, in the middle and high-speed ranges, the output shaft rotation information triggers a decrease in stability because the phase is delayed. Therefore, we adopt an original control system that achieves both improved positioning accuracy and stability by combining and feeding back input shaft rotation information.

5 High-accuracy positioning

Fig. 7 presents the positioning performance verification device for C3 eMotion. An arm was attached to the IAU-60, each of the weight lifting actions were performed with semi-closed and mixed controls, and the position of the arm tip during each scenario was compared. Fig. 8 presents a graph of the comparison results. From Fig. 8, deviations of 0.16 mm or more in the positioning command values from deflections can be observed owing to reducer transmission errors or applied load during semi-closed control operations without the output shaft encoder. However, in the case of mixed control with a double-encoder arrangement, reduced accuracy factors such as reducer torsion and angle transmission errors are cancelled, the deviation from the positioning command value is 0.01 mm or less, and a higher positioning performance than that of the semi-closed control case is demonstrated under load conditions.

When C3 eMotion is adopted as a robot joint, the reducer torsion and angle transmission error per the number of shafts accumulates in the robot tip. These accumulated errors are eliminated by adopting mixed control, using output shaft information; and the positioning accuracy at the robot tip can be expected to improve. An articulated robot with a 5 kg payload capacity incorporating C3 eMotion (Fig. 9) is adopted to repeatedly evaluate positioning performance with semi-closed and mixed controls when a 0.5-kg and 5-kg weight is attached to the robot tip. The reach of this robot is 830 mm, and it is evaluated at its maximum reach. When

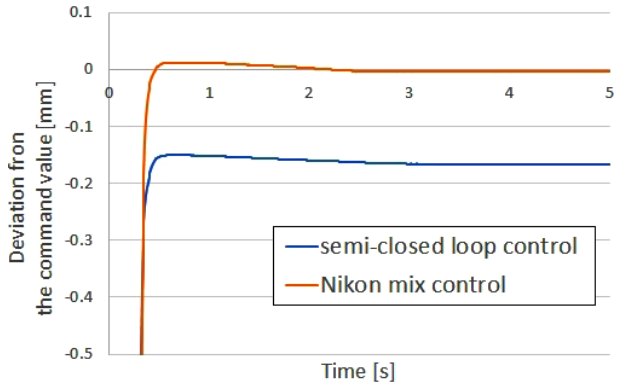


Fig. 8 Experimental results of positioning with each control loop

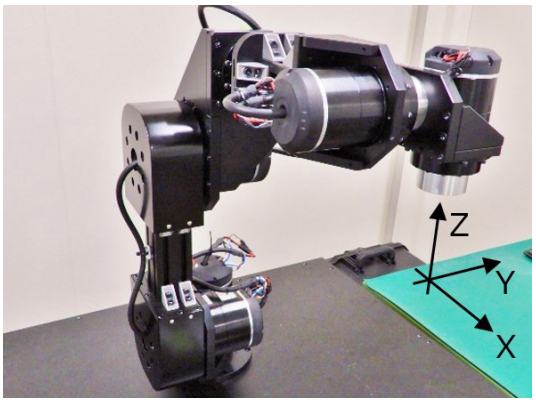


Fig. 9 Articulated robot (payload: 5 kg) for accuracy evaluation

semi-closed control is adopted, position changes are generated by reducer torsion from load. In contrast, during the application of mixed control, positioning is stable even if load conditions are altered.

Based on these results, the decrease in positioning accuracy is expected to be minimized in robot applications, such as work or end effectors, in which the tip mass changes when configuring arms of different lengths and mases with C3 eMotion.

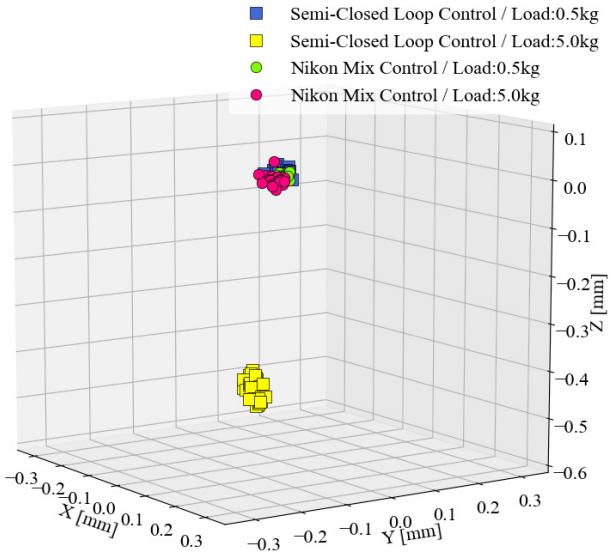


Fig. 10 Positioning evaluation results with robot

6 Torque detection function

As described above, in a double-encoder arrangement in which the reducer is regarded as a torsional elastic body, the torque can be detected by measuring and converting the relationship between the torsion amount and torque beforehand. However, errors owing to the eccentricity of the encoder disk are present in the encoders of the input and output shafts, and angular transmission errors are also present in the reducer. Therefore, accurate torque detection is impossible even if the output values of the two encoders are simply compared. In C3 eMotion, the output shaft encoder is corrected to an accuracy of ± 15 arc-sec or less using an external reference. Furthermore, an angle-correction value, including the angular error value obtained by comparing the output value of the corrected output shaft encoder with the output value of the input shaft encoder, is recorded beforehand, and the angle transmission error triggered by the reducer is also corrected. This facilitates the reduction of error in the entire encoder system, and improves torque detection accuracy.

Fig. 11 presents a typical example of torque detection values for the external input torque. Because the torsional characteristics of the strain wave gear are nonlinear, the spring constant adopted for torque conversion is converted using an approximate curve. In this example, the torque detection value is stable, regardless of the output shaft rotation speed, and errors in the torque detection value remain within the range of ± 5%.

In a typical servo motor system, torque is detected by the torque converting motor current values. Fig. 12 illustrates an example of the temperature characteristics of torque detection errors in terms of the motor current value conversion in IAU-60. Furthermore, the rotational speed when acquiring data is 10 min^{-1}. Owing to the temperature dependence of the motor winding and motor drive circuit element resistance, as well as the viscosity of the grease and lubricating oil in the mechanism, the motor current value varies depending on the temperature. Therefore, to accurately detect torque, complicated processes, such as correction by temperature, are required. In the double-encoder arrangement, the reducer torsion amount, which is less temperature

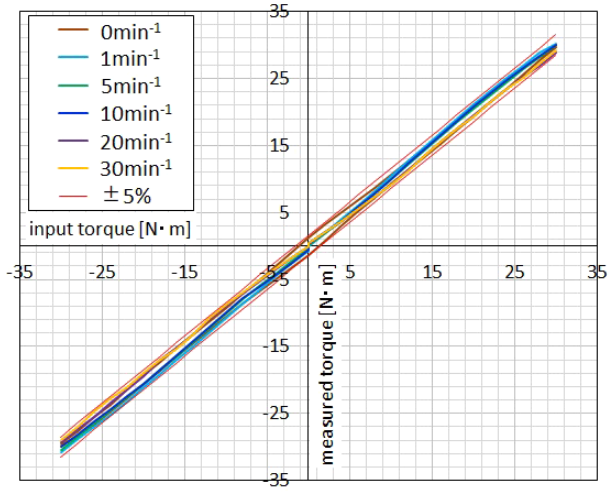


Fig. 11 Torque detection per revolution (IAU-60)

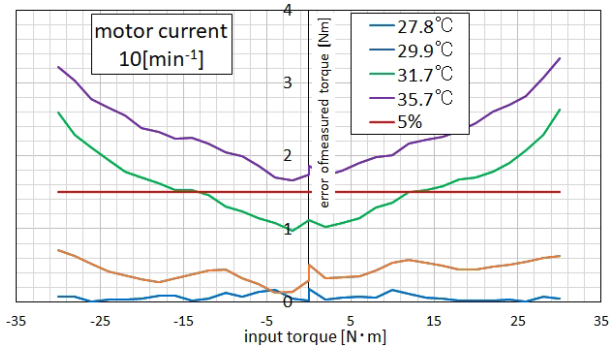


Fig. 12 Temperature–torque error characteristics (motor current detection)

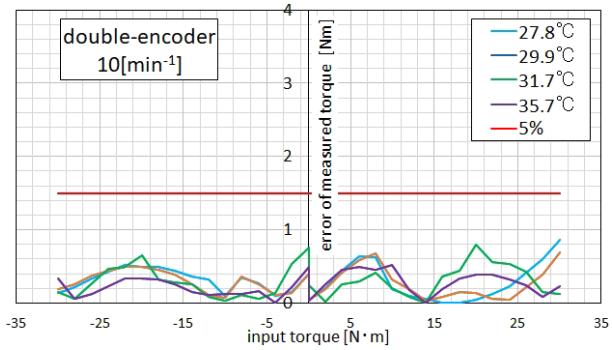


Fig. 13 Temperature-torque error characteristics (double encoder)

dependent, is directly measured by the encoder, thereby enabling stable torque detection independent of temperature (Fig. 13).

7 Conclusion

In C3 eMotion, various components, including a motor driver with proprietary Nikon built-in control technology, are packaged and daisy-chained by EtherCAT® to facilitate multi-axis connectivity. In addition, high-precision positioning and a torque detection function was realized by a double-encoder arrangement and original control technology. C3 eMotion adopts these technologies to enable freely-conceived configurations of articulated robots that could solely be realized in the past by manufacturers specializing in robots. Nikon intends to expand its C3 eMotion lineup in the future to satisfy diverse demands and launch peripheral devices such as special-purpose upper level controllers corresponding to typical robot configurations. In addition, we hope to contribute to the development and evolution of robot technology via the higher precision and further enhancement of precision.

References

[1] Headquarters for Japan's Economic Revitalization, "New Robot Strategy," Prime Minister's Official Residence, 2015.

[2] NIKON CORPORATION, "Intelligent actuator unit catalog," 24 11 2020. [Online]. Available: https://www.nikon.com/products/encoder/pdf/intelligent-actuator-unit.pdf.

[3] EtherCAT Technology Group, "EtherCAT Technology Group website," [Online]. Available: https://www.ethercat.org/default.htm.

[4] M. Yamamoto, M. Iwasaki, Y. Okitsu, K. Sasaki and T. Yajima, "Modeling and Compensation for Angular Transmission Error in Harmonic Drive Gearings (1st Report)," (in Japanese), *Journal of the Japan Society of Precision Engineering*, vol. 76, no. 10, pp. 1206–1211, 2010.

[5] J. Fujimori, R. Ienaka, Y. Horiuchi, T. Kubota, N. Takagi, K. Yama, A. Wakisaka and J. Kawabata, "Precision-Machining Robot System," *Kawasaki Technical Review*, no. 172, pp. 44–47, 2012.

渡邉昭宏　Akihiro WATANABE
デジタルソリューションズ事業部　開発・技術部
Development Department
Digital Solutions Business Unit

高橋喜二　Yoshiji TAKAHASHI
デジタルソリューションズ事業部　開発・技術部
Development Department
Digital Solutions Business Unit

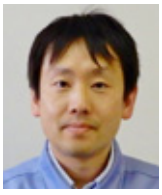
湯本一樹　Kazuki YUMOTO
デジタルソリューションズ事業部　開発・技術部
Development Department
Digital Solutions Business Unit

引地哲也　Tetsuya HIKICHI
デジタルソリューションズ事業部　開発・技術部
Development Department
Digital Solutions Business Unit

渡邉昭宏
Akihiro WATANABE

高橋喜二
Yoshiji TAKAHASHI

湯本一樹
Kazuki YUMOTO



引地哲也
Tetsuya HIKICHI

研究開発論文

Research and Development Reports

スペクトル狭帯化によるフーリエ変換法のダイナミックレンジ拡大†

中山　繁，鳥羽英光，藤原直樹，玄間隆志，武田光夫

Dynamic Range Expansion of Fourier-Transform Method by Spectrum-Narrowing

Shigeru NAKAYAMA, Hidemitsu TOBA, Naoki FUJIWARA, Takashi GEMMA and Mitsuo TAKEDA

フーリエ変換法において，位相分布の勾配が大きいフリンジパターンの解析を可能にするモデルベースのフリンジ解析技術を提案する．従来のフーリエ変換法では，大きな位相分布の勾配によって広がったフーリエスペクトルを空間キャリア周波数によって分離することができないため，測定可能な位相分布のダイナミックレンジが制限されていた．提案するモデルベースの反復手法は，広がったスペクトルを効果的に狭め，位相解析誤差を低減する．実験とシミュレーションにより，提案したスペクトル狭帯化技術の有効性を実証した．

In this study, a model-based fringe analysis technique is proposed for analyzing fringe patterns with large phase slopes using the Fourier-transform method. In the conventional Fourier-transform method, the dynamic range of the measurable phase distribution is limited because the Fourier spectra broadened by large phase slopes cannot be separated by the spatial carrier frequency. Our model-based iterative technique effectively narrows the broad spectrum and reduces phase analysis errors. Experiment and simulation results are presented to demonstrate the feasibility of the proposed spectrum-narrowing technique.

Key words シングルショットフリンジ解析，干渉法，フリンジ投影法，フーリエ変換法
single-shot fringe analysis, interferometry, fringe-projection profilometry, Fourier-transform method

1 はじめに

フーリエ変換法（FTM）[1]，[2] は，光学計測におけるシングルショットフリンジ解析の有用な手法であり，干渉法［3］やフリンジ投影法［4］などに広く使用されている．FTM は空間キャリアをフリンジパターンに導入し，そのフーリエスペクトルを信号スペクトル，共役スペクトル，および DC 成分に分離する．位相分布は，その分離された信号スペクトルのバンドパスフィルタリングとそれに続く逆フーリエ変換によって得られた複素解析信号から求められる．１枚のフリンジデータから位相分布が求まるので，振動などの外乱が存在する環境下での測定や高速現象の測定などに応用されている［5］．

その一方で，FTM では測定可能な位相分布のダイナミックレンジが制限されてしまう．空間キャリア周波数に比べて位相分布の勾配が大きい場合，信号スペクトルとその共役スペクトルの帯域幅が広くなり，３つのスペクトル成分が重なり合ってしまう．その結果，フィルタリングによって信号スペクトルを完全に分離・抽出することができなくなり，大きなリップル状の位相解析誤差が発生してしまう．例えば，計算機ホログラム（CGH）などの非球面波を球面波あるいは平面波に変換する素子を使用せずに，従来の球面測定用の干渉計で非球面テストを行うときに上記のような状況が発生する．

本稿では，大きな勾配を持つ位相分布に対処するための初期の研究成果［6］をさらに進め，位相変調フリンジのフーリエスペクトルの帯域幅を効果的に狭めることによって FTM のパフォーマンスを向上させる新しいモデルベースの手法を提案する．この手法は反復アルゴリズムに基づいており，従来の FTM よりも解析可能な位相分布のダイナミックレンジを広げることができる．不均一なフリンジが発生する状態で行った球面テストの干渉計実験により本提案の有効性を実証し，非球面テストを想定したシミュレーションによって極めて勾配が大きい位相分布への適用可能性を示す．

† 本稿は，引用文献［10］をもとに図や説明を追加したものである．実験検証において，データ解析途中のスペクトルの図とフィルター条件を変更したときの解析誤差を追記した．仮想インターフェログラム法の適用手順に関する説明を追記した．

2 スペクトル狭帯化 FTM

空間キャリアを導入したフリンジパターンの強度分布は次のように記述できる．

$$I(x,y)=a(x,y)+b(x,y)\cos[\phi(x,y)] \tag{1}$$

$$\phi(x,y)=2\pi(f_x x+f_y y)+\phi_S(x,y) \tag{2}$$

ここで，$\phi_S(x,y)$は位相分布であり，$a(x,y)$，$b(x,y)$はそれぞれバックグラウンドとエンベロープの強度分布を表し，f_xとf_yは導入した空間キャリアの周波数である．照明がほぼ均一で，均一な反射率の研磨された表面をテストするときのようなバックグラウンドとエンベロープがほぼ一定であると見なせる場合を想定する．この場合，位相分布$\phi_S(x,y)$がフリンジスペクトルの広がりの主な原因になる．本提案では，モデル化した位相分布$\phi_{S,MODEL}(x,y)$を反復アルゴリズムで使用して，広がったスペクトルを狭帯化することによって解析可能な位相分布のダイナミックレンジを拡大する．

問題設定としては，ブラインドテストと呼ばれる位相分布に関する事前情報がないケースと，設計データから理想的な位相分布が事前にわかっているケースに分けられる．Fig. 1 のフローチャートと Fig. 2 に模式的に示したデータ例を参照して，アルゴリズムの各ステップを説明する．

先ず，ステップ 1 ～ステップ 4 の反復処理に入る前に，フリンジパターンから DC 成分$a(x,y)$を削除する．

$$\begin{aligned} I_1(x,y)=&\frac{1}{2}b(x,y)exp\left[2\pi i(f_x x+f_y y)+i\phi_S(x,y)\right]\\ &+\frac{1}{2}b(x,y)exp\left[-2\pi i(f_x x+f_y y)-i\phi_S(x,y)\right]\end{aligned} \tag{3}$$

バックグラウンドがほぼ一定という条件から，$a(x,y)$のフーリエスペクトルはスペクトル領域の中心の狭い範囲に現れ，狭い DC 除去フィルターによって除去することができる．理想的にはこのフィルタリング操作は信号スペクトルをそのまま維持する必要があるが，位相分布の勾配が非常に大きく，信号スペクトル（および共役スペクトル）が除去帯域内へ広がる場合には，スペクトルのごく一部が DC 除去フィルターによって失われてしまう．

ステップ 1 ：モデル位相分布の生成

反復処理の初回において，位相分布に関する事前情報がないブラインドテストの場合は，Fig. 2 (a) に示すような取得されたフリンジパターンを通常の FTM で解析する．滑らかで連続的な位相分布でさえ，大きな勾配がある場合には Fig. 2 (b) のような広いスペクトルを持ち，フィルタリングによって信号スペクトルを完全に分離・抽出することができなくなり，大きなリップル状の位相解析誤差が発生してしまう．そこで，検出された位相分布を第36項までの Zernike 多項式［7］などの滑らかな 2 次元多項式でフィッティングすることによってモデル位相分布$\phi_{S,MODEL}(x,y)$を推定する．このフィッティングプロセスによってリップル状の大きな変動が排除され，モデル位相分布のより良い初期値として機能する．設計データから理想的な位相分布が事前にわかっている場合は，反復処理の初回ではその位相分布をモデル位相分布$\phi_{S,MODEL}(x,y)$に設定すればよい．位相分布の事前情報がある場合もない場合も，反復処理の 2 回目からは前のラウンドのステップ 4 で得られた位相分布の推定値$\phi_S(x,y)$に 2 次元多項式フィッティングを施し，新しいモデル位相分布$\phi_{S,MODEL}(x,y)$を生成する．

ステップ 2 ：共役スペクトル狭帯化・除去

次に，式 (3) のフリンジパターン$I_1(x,y)$に$\exp\left[i\phi_{MODEL}(x,y)\right]$を乗算する．ここで，$\phi_{MODEL}(x,y)$は，

$$\phi_{MODEL}(x,y)=2\pi(f_x x+f_y y)+\phi_{S,MODEL}(x,y) \tag{4}$$

であり，$\phi_{S,MODEL}(x,y)$は，真の位相分布$\phi_S(x,y)$の推定値としてステップ 1 で生成したモデル位相分布である．結果として得られる複素振幅分布$I_2(x,y)$は，おおよそ以下に示す第 1 項と第 2 項の和で表される．

第 1 項：

$$\frac{1}{2}b(x,y)\exp\left[4\pi i(f_x x+f_y y)+i\phi_S(x,y)+i\phi_{S,MODEL}(x,y)\right]$$

第 2 項：

$$\frac{1}{2}b(x,y)\exp\left[-i\phi_S(x,y)+i\phi_{S,MODEL}(x,y)\right]$$

モデル位相分布$\phi_{S,MODEL}(x,y)$が真の位相分布$\phi_S(x,y)$に近い場合，第 2 項（共役スペクトルに対応）の位相成分はほぼキャンセルされ，Fig. 2 (c) に示すように，そのスペクトルは狭くなり，孤立する．したがって，不要な共役スペクトルを狭い除去フィルターで除去でき，信号スペクトルへの漏れを減らすことができる．共役スペクトル除去後の複素振幅分布は，おおよそ第 1 項で表される．このプロセスは，設計値に基づいて製作した CGH を用いて非球面波を球面波あるいは平面波に変換することによって非球面形状を計測する手法に概念的に類似している［8］．

ステップ 3 ：信号スペクトル狭帯化・抽出

次に，残った位相分布（第 1 項）に$\exp\left[-2i\phi_{MODEL}(x,y)\right]$を乗算する．結果として得られる複素振幅分布$I_3(x,y)$は，おおよそ次式で表される．

$$\begin{aligned} I_3(x,y)&\approx\frac{1}{2}b(x,y)exp\left[i\phi_S(x,y)-i\phi_{S,MODEL}(x,y)\right]\\ &=\frac{1}{2}b(x,y)exp\left[i\Delta\phi_S(x,y)\right]\end{aligned} \tag{5}$$

$$\Delta\phi_S(x,y)=\phi_S(x,y)-\phi_{S,MODEL}(x,y) \tag{6}$$

Fig. 2 (d) に示すように$I_3(x,y)$のスペクトルが狭くなるた

め，適切な帯域幅でフィルタリングして信号スペクトルを抽出することができる．

ステップ 4 ：位相分布の算出

抽出した信号スペクトルに逆フーリエ変換を施すことで，$\Delta\phi_S(x,y)$を小さな誤差で求めることができる．求めた位相分布とモデル位相分布の差分形状を$\Delta\phi_S'(x,y)$とすると，位相分布の推定値$\phi_S'(x,y)$は式（7）で算出される．

$$\phi_S'(x,y)=\Delta\phi_S'(x,y)+\phi_{S,MODEL}(x,y) \quad (7)$$

このプロセスは，非球面テストのヌルテストに類似しており，モデル位相分布$\phi_{S,MODEL}(x,y)$の導入はCGHによる波面変換と同様な役割を果たしている．

繰り返し記載することになるが，次の反復ラウンドのステップ 1 では，ステップ 4 で得られた推定値$\phi_S'(x,y)$に 2 次元多項式フィッティングを施し，新しいモデル位相分布$\phi_{S,MODEL}(x,y)$を生成する．続けて，更新したモデル位相分布$\phi_{S,MODEL}(x,y)$を用いてステップ 2 ～ 4 を実行する．このような反復ラウンドを推定値$\phi_S'(x,y)$が収束するまで繰り返す．

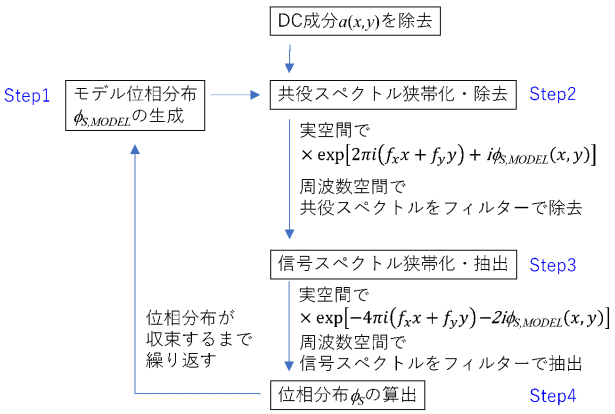


Fig. 1　アルゴリズム

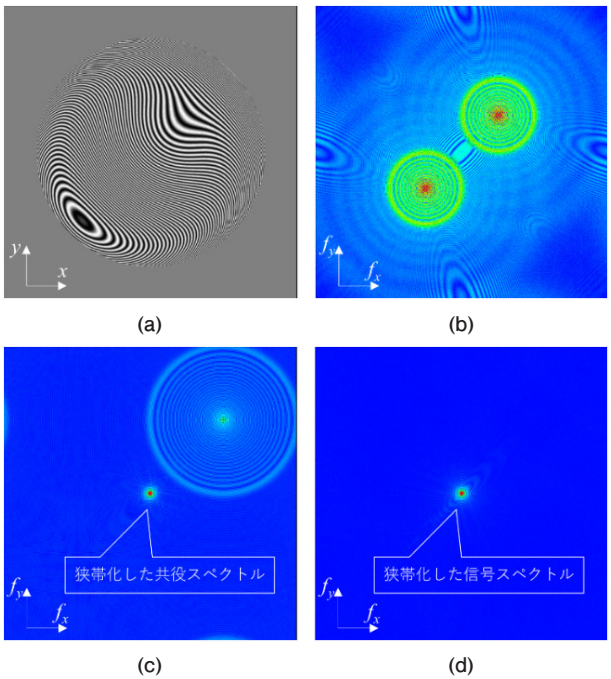


Fig. 2　（a）干渉縞の例 $I(x,y)$，（b）$I_1(x,y)$ のフーリエスペクトル，（c）ステップ 2 における $I_2(x,y)$ のフーリエスペクトル，（d）ステップ 3 における $I_3(x,y)$ のフーリエスペクトル

3 実験検証

本手法の有効性を確かめるため，フィゾー干渉計を用いた球面テストで実験評価を行った．実験系の配置を Fig. 3（a）に示す．被検面を光軸に沿ってシフトした上で傾斜させ，Fig. 3（b）に示した歪んだ干渉縞を形成した．データウィンドウは N＝512の正方形で，キャリア周波数f_x，f_yを約50サイクルに設定した．位相分布は主に 5.4 λPV のデフォーカス成分であった．この実験ではブラインドテストとして解析を行い，初期モデル位相分布には，通常の FTM で得られた位相分布に36項までの Zernike 多項式でフィッティングを施したものを用いた．DC 成分および共役スペクトルの除去フィルターは半径10サイクルの円形阻止フィルターとし，信号スペクトルの抽出フィルターは半径50サイクルの円形の通過フィルターとした．ステップ 2 で共役スペクトルを狭帯化した状態を Fig. 3（c）に，共役スペクトルを除去した後にステップ 3 で信号スペクトルを狭帯化した状態を Fig. 3（d）に示す．今回の実験ではステップ 1 ～ステップ 4 の反復は 3 回で収束した．

真の位相分布がわからないため，別途位相シフト法で同

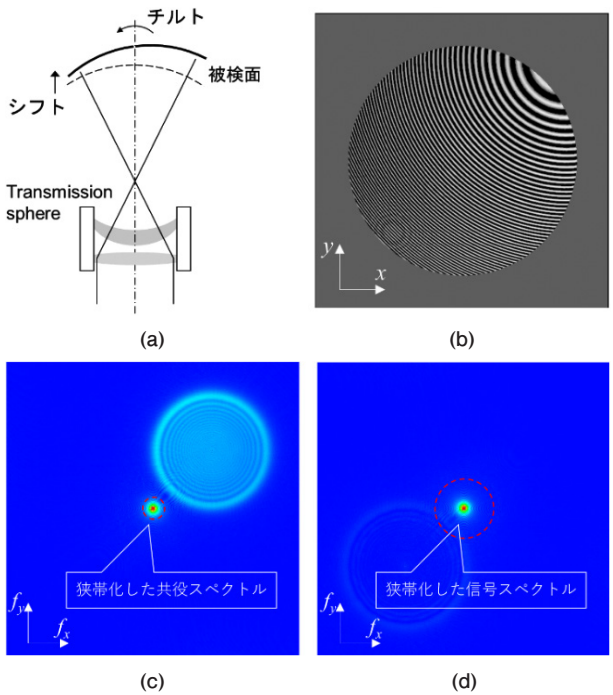
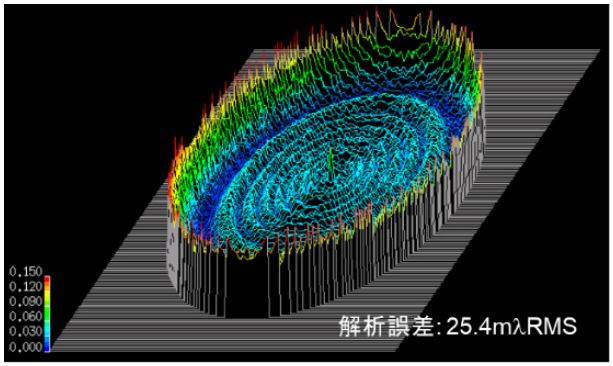
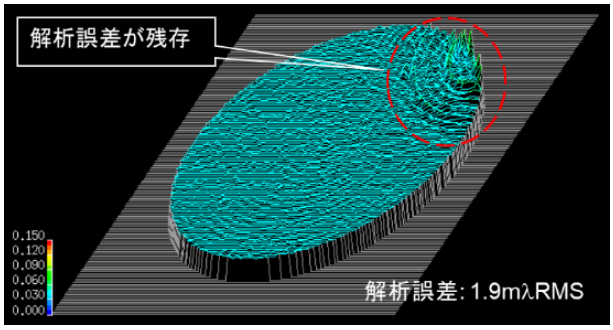


Fig. 3　（a）光学系模式図，（b）干渉縞，（c）ステップ 2 における $I_2(x,y)$ のフーリエスペクトル，（d）ステップ 3 における $I_3(x,y)$ のフーリエスペクトル

じ被検面の測定を行い，その測定結果を真の位相分布とみなして各手法の解析誤差を算出した．通常の FTM とスペクトル狭帯化手法の解析誤差をそれぞれ Fig. 4 (a) と Fig. 4 (b) に示した．通常の FTM の解析においても半径50サイクルの円形の通過フィルターを用いた．また，これらの FTM の帯域幅に適合させるために位相シフト法のデータに対して同じ帯域のローパスフィルターを施している．通常の FTM では 25.4 mλRMS の大きな誤差が見られるが，スペクトル狭帯化手法では誤差は 1.9 mλRMS まで減少した．スペクトル狭帯化手法によって大幅に解析誤差は改善したが，Fig. 4 (b) の右上に誤差が残っていることがわかる．なお，DC 成分と共役スペクトルの除去フィルターのサイズを半径20サイクルにすると解析誤差は 3.6 mλRMS まで増加した．適切なフィルターサイズを選定することが必要である．



(a)



(b)

Fig. 4　(a) 通常の FTM での解析誤差，(b) スペクトル狭帯化手法での解析誤差

次に，仮想インターフェログラム法（VIM）と呼ばれる手法［9］を組合わせて，この残留誤差を低減することを試みた．VIM は，位相分布，エンベロープおよびバックグラウンドの強度分布を設定し，式（1）および式（2）に従って数値的に作成した仮想インターフェログラムを解析することによって解析誤差を推定し，補正を施すものである．Fig. 5 を用いて VIM の適用手順を説明する．まず，スペクトル狭帯化手法を用いてデータを解析して位相分布を求める．次に，求めた位相分布を用いて仮想インターフェログラムを作成し，それをスペクトル狭帯化手法で解析し，スペクトル狭帯化手法で発生する解析誤差を推定する．最後に，データをスペクトル狭帯化手法で解析して得られていた位相分布から推定した解析誤差を減算補正する．

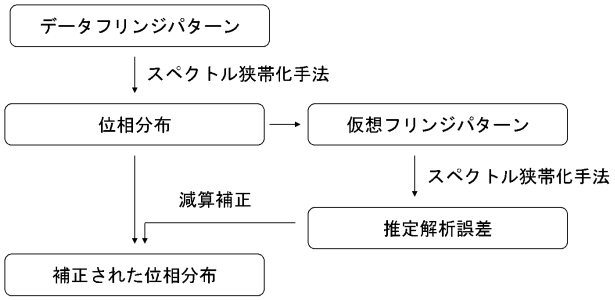
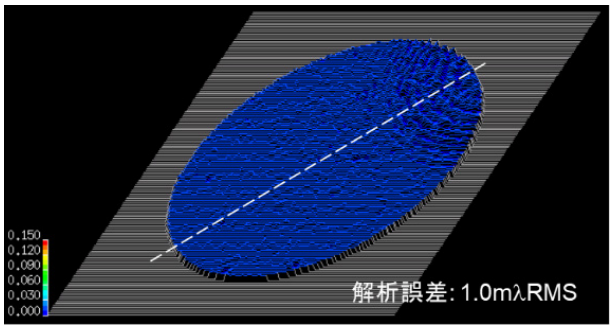
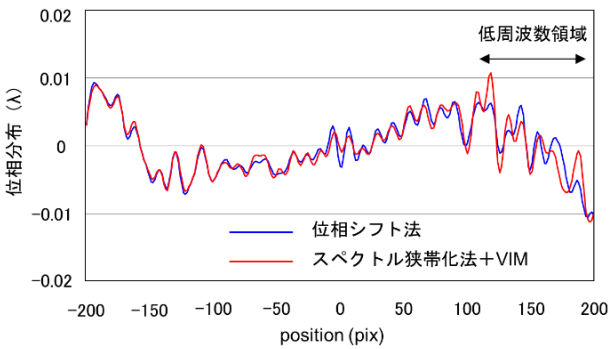


Fig. 5　スペクトル狭帯化手法と VIM の組合せ

Fig. 6 (a) に VIM を組合わせたときの解析誤差を示した．VIM を適用することで Fig. 4 (b) に残っていたリップル形状の誤差が半分に減少した．解析結果の様子を詳細に調べるため，Fig. 6 (a) の点線に沿った位相分布のプロファイルを Fig. 6 (b) に示した．スペクトル狭帯化手法と VIM の組合せで得られた位相分布が位相シフト法で得られた位相分布と概ね一致しており，この提案手法が高周波成分の形状を解析でき，製造誤差などを評価できることを示しているが，低周波数領域には小さな解析誤差が残っていることもわかる．



(a)



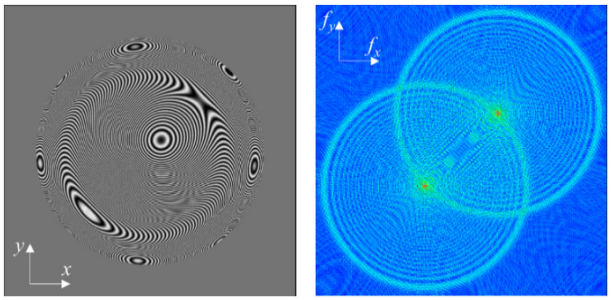
(b)

Fig. 6　(a) スペクトル狭帯化手法と VIM を組合せたときの解析誤差，(b) 各手法で得られた位相分布の断面図

4　非球面テストのシミュレーション

CGH などの専用の波面変換素子を用いずに非球面テストを行う場合を考え，そのようなケースへの適用可能性を調

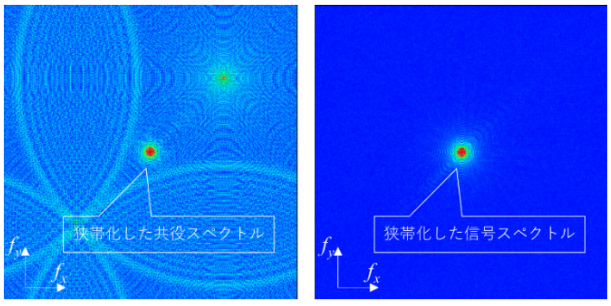
べるために数値シミュレーションを実施した．従来の球面をテストするための干渉計を想定し，設計形状を 15.0 λPV の大きさの Zernike 多項式の第 9 項で表される形状とし，製造された表面形状を 16.5 λPV の大きさの Zernike 多項式の第 9 項で表される形状とした．Fig. 7 (a) に，製造された表面形状に対してシミュレートした干渉縞を示す．45度方向の空間キャリア周波数を与えた高密度干渉縞であり，閉じた縞から予想されるように空間キャリア周波数はこの大きな位相分布に対して十分に高くはなく，干渉縞の信号スペクトルと共役スペクトルは Fig. 7 (b) に示すように大きくオーバーラップしている．



(a) (b)

Fig. 7 (a) シミュレートした干渉縞，(b) 干渉縞のフーリエスペクトル

このシミュレーションでは，モデル位相分布の初期値に設計データを採用し，ステップ 2 で共役スペクトルを狭め，ステップ 3 で信号スペクトルを狭めて所定の解析を行った．ステップ 2 で共役スペクトルを狭めた後のスペクトルを Fig. 8 (a) に，さらに共役スペクトルを除去した後にステップ 3 で狭めた信号スペクトルを Fig. 8 (b) に示す．



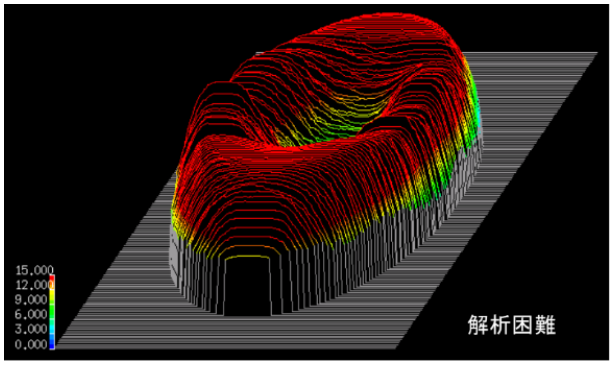
(a) (b)

Fig. 8 (a) ステップ 2 におけるフーリエスペクトル，(b) ステップ 3 におけるフーリエスペクトル

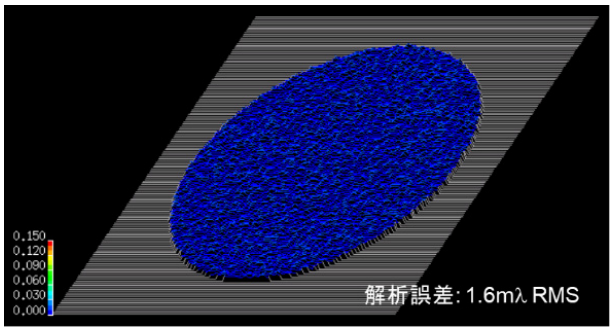
Fig. 9 (a)，(b) に示すように，通常の FTM では解析できなかったが，スペクトル狭帯化手法と VIM を組合わせることで 1.6 mλRMS の誤差で解析できた．

5 まとめ

大きな勾配を持つ位相分布に対する FTM の解析誤差を減らすために，スペクトル帯域幅を狭めるモデルベースの手法を提案した．フィゾー干渉計を用いた歪んだ干渉縞の解析実験では，従来の FTM に対して解析誤差が10分の 1 以下に低減し，提案した手法の有効性を確認できた．さらに，非球面テストを想定した数値シミュレーションにより，従来の FTM では解析できなかった極めて不均一なフリンジパターンを解析できることを示した．スペクトル狭帯化手法が，FTM で解析可能な位相分布のダイナミックレンジを拡大し，大きな振動や乱気流が存在する環境での測定，高速現象の瞬時測定などで新しい可能性を開くことを期待している．



(a)



(b)

Fig. 9 (a) 通常の FTM での解析誤差，(b) スペクトル狭帯化手法と VIM を組合せたときの解析誤差

引 用 文 献

[1] M. Takeda, H. Ina, and S. Kobayashi, “Fourier-transform method of fringe-pattern analysis for computer-based topography and interferometry,” *J. Opt. Soc. Am.*, vol. 72, no. 1, pp. 156–160, 1982.

[2] M. Takeda and K. Mutoh, “Fourier-transform profilometry for the automatic measurement of 3-D object shapes,” *Appl. Opt.*, vol. 22, no. 24, pp. 3977–3982, 1983.

[3] M. Takeda, “Spatial-carrier fringe pattern analysis and its applications to precision interferometry and profilometry: an overview,” *Ind. Metrol.*, vol. 1, no. 2, pp. 79–99, 1990.

[4] X. Su and W. Chen, “Fourier transform profilometry: a review,” *Opt. Lasers Eng.*, vol. 35, no. 5, pp. 263–284, 2001.

[5] M. Takeda, “Fourier fringe analysis and its application to

metrology of extreme physical phenomena: a review," *Appl. Opt.*, vol. 52, no. 1, pp. 20–29, 2013.

[6] S. Nakayama, H. Toba, N. Fujiwara, T. Gemma, and M. Takeda, "Fourier-transform method with high accuracy by use of iterative technique narrowing the spectra of a fringe pattern," in *Fringe 2009: 6th International Workshop on Advanced Optical Metrology*, W. Osten and M. Kujawinska, eds.: Springer, 2009, pp. 106–111.

[7] E. P. Goodwin and J. C. Wyant, *Field Guide to Interferometric Optical Testing*: SPIE Press, 2006, p. 24.

[8] D. Malacara, M. Servin, and Z. Malacara, *Interferogram Analysis for Optical Testing*, 2nd ed.: CRC Press, 2005, Chap. 8, pp. 416–418.

[9] H. Toba, Z. Liu, S. Udagawa, N. Fujiwara, S. Nakayama, T. Gemma, and M. Takeda, "Phase analysis error reduction in the Fourier transform method using a virtual interferogram," *Opt. Eng.*, vol. 58, no. 8, 084103, 2019.

[10] S. Nakayama, H. Toba, N. Fujiwara, T. Gemma, and M. Takeda, "Enhanced Fourier-transform method for high-density fringe analysis by iterative spectrum narrowing," *Appl. Opt.*, vol. 59, no. 29, pp. 9159–9164, 2020.

中山　繁　Shigeru NAKAYAMA
研究開発本部　光技術研究所
Optical Research Laboratory
Research & Development Division

鳥羽英光　Hidemitsu TOBA
研究開発本部　光技術研究所
Optical Research Laboratory
Research & Development Division

藤原直樹　Naoki FUJIWARA
生産本部　設備技術開発部
Equipment Engineering Department
Production Technology Division

玄間隆志　Takashi GEMMA
光学本部　開発戦略部
Strategic Technology Development
Department Optical Engineering Division

武田光夫　Mitsuo TAKEDA
宇都宮大学
Utsunomiya University

無容器法により作製した $LaGaO_3$ ガラスの構造，振動特性および電子状態の解析

吉本幸平，増野敦信，佐藤　至，江面嘉信，井上博之，上田　基，水口雅史，山本優也

Structural, Vibrational, and Electronic Analysis of $LaGaO_3$ Glass Prepared through Containerless Processing

Kohei YOSHIMOTO, Atsunobu MASUNO, Itaru SATO, Yoshinobu EZURA, Hiroyuki INOUE, Motoi UEDA, Masafumi MIZUGUCHI and Yuya YAMAMOTO

無容器法は，融液－容器界面で生じる不均一核生成を抑制することにより，融液のガラス形成を大きく促進することができる．近年，筆者らはこの手法を用いて，通常のるつぼ熔解では得られない $LaGaO_3$ ガラスの合成に成功した．$LaGaO_3$ ガラスは，高い屈折率や紫外～赤外域における広い光透過性など，典型的な網目形成酸化物を主体とするガラスとは異なる光学特性を有しており，新しい光機能材料としての応用が期待される．本研究では，$LaGaO_3$ ガラスにおける特異な光学特性の構造的起源を明らかにするため，ガラスの局所構，振動特性，電子状態の解析を行った．実験的に取得した回折データと分子動力学計算から $LaGaO_3$ ガラスの構造モデルを構築し，振動スペクトル計算や電子状態解析によって，ガラスの局所構造が振動特性や電子状態へ及ぼす影響について考察した．本研究で得られた知見は，無容器法を用いた新規光機能ガラスの組成設計に活用できるものと期待される．

Containerless processing is useful for vitrifying compositions with low glass-forming ability because it suppresses heterogeneous nucleation from the container wall. Using this method, we recently synthesized $LaGaO_3$ glass, which cannot be obtained using an ordinal melt-quenching process. $LaGaO_3$ glass exhibited excellent optical properties such as a high refractive index and wide optical transparency from the ultraviolet to infrared range, which are unusual in a typical oxide glass containing network-formers, and can be a new optical functional material for various applications. In this study, we investigated the local structures, vibrational characteristics, and electronic properties of $LaGaO_3$ glass to reveal the structural origins of unique optical properties. A structural model of $LaGaO_3$ glass was constructed from diffraction data and molecular dynamics simulations. In addition, the vibrational density of states and *ab initio* calculations were applied to evaluate the vibrational characteristics and electronic structures of $LaGaO_3$ glass.

Key words ガラス，光学特性，無容器法，分子動力学，第一原理計算
glass, optical property, containerless processing, molecular dynamics, *ab initio* calculations

1 Introduction

Ga_2O_3-based systems occasionally form glass even without the presence of network-former (NWF) oxides such as SiO_2, B_2O_3, and P_2O_5. Such unconventional gallate glasses have attracted interest owing to their superior optical properties, including long infrared cut-off wavelengths, high refractive indices, and large nonlinear optical properties [1], [2]. However, NWF-free systems generally show low glass-forming abilities, making it difficult to obtain bulk glass. Containerless processing is suitable for obtaining bulk glass with such unconventional NWF-free compositions because it suppresses heterogeneous nucleation and promotes a deep undercooling of the melt [3]–[5]. In recent years, we reported the vitrification of a $LaGaO_3$ composition using containerless processing [6]. Owing to its high oxygen polarizability and low maximum phonon energy, $LaGaO_3$ glass exhibits excellent optical properties, such as a high refractive index (1.954), relatively low dispersion (33.6 in Abbe number), wide optical transmittance range (280 nm to 8 μm), and efficient fluorescence emission by rare-earth activation [6], [7]. These results indicate the significant advantage of rare-earth-gallate glasses for various optical applications such as lenses, lasers, and infrared optics. Nevertheless, the local atomic and electronic structures, which are the physical origins of the optical properties, have not been investigated well in rare-earth-gallate glasses.

In this study, we evaluated three important aspects of

$LaGaO_3$ glass: local atomic structures, vibrational characteristics, and electronic structures. To achieve this, high-energy X-ray and neutron diffraction experiments, classical molecular dynamics (CMD), and *ab initio* molecular dynamics (AIMD) simulations were conducted. We constructed a detailed structural model of $LaGaO_3$ glass and investigated the origins of unique vibrational and optical properties by computing vibrational spectra and *ab initio* electronic structure calculations.

2 Experimental and Computational Details

2.1. Glass Synthesis

High-purity La_2O_3 (5N) and Ga_2O_3 (4N) powders were mixed in a stoichiometric ratio of $LaGaO_3$. The mixture was pressed into pellets and sintered at 1473 K for 12 h under ambient atmosphere. Using an aerodynamic levitation furnace, the pellets were placed on a conical nozzle, levitated by an O_2 gas flow, and melted using a CO_2 laser (Fig. 1). After melting for several tens of seconds, the CO_2 laser was turned off and the sample was cooled at a rate of several hundreds of kelvins per second, forming spherical glass with a diameter of 2–3 mm. The glass was annealed near the glass transition temperature (1026 K [6]) for 10 min to remove the internal strain.

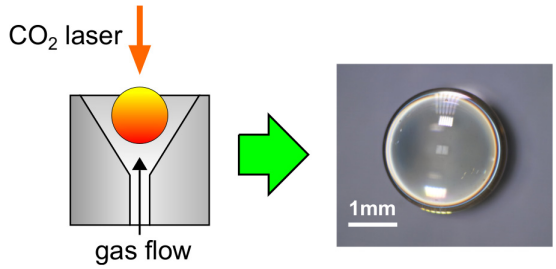


Fig. 1 Schematic illustration of glass synthesis using the aerodynamic levitation technique, and a photograph of $LaGaO_3$ glass.

2.2. Diffraction Experiments

High-energy X-ray and neutron diffraction experiments were conducted at the BL04B2 beamline of the SPring-8 synchrotron radiation facility [8] and at the high-intensity total diffractometer NOVA installed at BL21 of the Materials and Life Science Experimental Facility at the J-PARC spallation neutron source [9], respectively. The energy of X-ray was 113 keV and the wavelength range of the incident neutron beam was $0.12 < \lambda < 8.3$ Å. The observed scattering intensities for the samples were corrected for polarization, absorption, background, multiple, and incoherent scattering, and normalized based on the incident beam profile. The total correlation functions, $T(r)$s, were then obtained through a Fourier transformation of the structure factors with a Lorch modification function [10]. The scattering vector from 0.3 to 25 $Å^{-1}$ for the X-ray and 0.5 to 35 $Å^{-1}$ for the neutron beam was used for the Fourier transformation.

2.3. Classical Molecular Dynamics Simulation

CMD simulations were conducted using the LAMMPS code [11]. We used sets of interatomic potential parameters determined in our previous study [12]. Coulombic interactions were calculated using the Ewald summation method with a relative precision of 10^{-6}. The structural model of the $LaGaO_3$ glass contained 2500 atoms. The size of the cubic cells was set to reproduce the experimental density (5.91 g/cm^3 [6]) and periodic boundary conditions were applied. The initial atomic configuration was randomly generated with constraints between the atom pairs to avoid overlapping atoms. The simulation was executed at a time step of 1 fs, and an *NVT* (canonical) ensemble with a Nose-Hoover thermostat was applied. The model was first equilibrated at 5000 K for 50 ps and then melted at 3000 K for 50 ps. The melt was cooled to 300 K at a nominal cooling rate of 10 K/ps and equilibrated at 300 K for another 50 ps.

2.4. Vibrational Analysis

The vibrational density of states (VDOS) was computed from the CMD structural model based on a harmonic approximation [13], [14]. The dynamical matrix, *i.e.*, the matrix of the second derivatives of the total energy with respect to the Cartesian coordinates, was computed through a numerical evaluation of the first derivatives obtained from the minute atomic displacement. The mass-weighted dynamical matrix was diagonalized to obtain eigenvalues ω_j^2, where ω_j is the eigenfrequency at mode j and eigenvectors $\mathbf{e}^j$. The total VDOS, $g(\omega)$, is obtained as

$$g(\omega) = \frac{1}{3N}\sum_{j=1}^{3N}\delta(\omega-\omega_j),$$

where N is the number of atoms in the cell, and δ is convoluted by a Gaussian function with a full width at half-maximum of 15 cm^{-1}. A partial VDOS, $g_\alpha(\omega)$, is defined by

$$g_\alpha(\omega) = \frac{1}{3N}\sum_{j=1}^{3N}\sum_{i\in\alpha} w_\alpha^j\,\delta(\omega-\omega_j),$$

where w_α^j is a weight factor, and α represents the type of atom or part of the structure. For the partial VDOS of an atom α, the weight factor is defined as $w_\alpha^j = \left|\mathbf{e}_i^j\right|^2$. To investigate the atomic motion for different eigenmodes, the eigenvectors were projected onto several vibrational modes of

typical structural units. For instance, the weight factor of the stretching mode of the Ga–O bond, $w^j_{\mathrm{Ga-O,s}}$, is given as

$$w^j_{\mathrm{Ga-O,s}} = \frac{\sum_i \left| \tilde{\mathbf{u}}^j_i \cdot \frac{\mathbf{r}_{\mathrm{Ga-O}}}{|\mathbf{r}_{\mathrm{Ga-O}}|} \right|^2}{\sum_i \left| \tilde{\mathbf{u}}^j_i \right|^2},$$

where $\mathbf{r}_{\mathrm{Ga-O}}$ is the Ga–O bond vector and $\tilde{\mathbf{u}}^j_i = \mathbf{u}^j_i - \mathbf{u}^j_{\mathrm{Ga}}$ is the displacement of the O atom i relative to the displacement of its Ga neighbor. In addition, $\mathbf{u}^j_i$ is the Cartesian displacement vector, which is obtained as $\mathbf{u}^j_i = \mathbf{e}^j_i / m_i^{1/2}$, where m_i is the mass of the atom. By replacing $\mathbf{r}_{\mathrm{Ga-O}}$ with a specific vector, it is possible to calculate other vibration modes and structural units [12]. For the vibrational analysis above, a structural model containing 750 atoms was prepared. To prevent the occurrence of imaginary frequencies, the dynamical matrix was computed after further cooling the structural model from 300 K to below 1.0×10^{-5} K under the *NVT* ensemble to minimize the internal energy of the system.

2.5. *Ab Initio* Calculation

Ab initio calculations based on density functional theory (DFT) were conducted using a hybrid Gaussian and plane wave method as implemented in the CP2K code [15]. For each atom, we used molecularly optimized short-range double-zeta basis sets [16]. The cutoff of the plane wave basis set for the electron density was set to 500 Ry. The generalized gradient approximation (GGA) was adopted to treat exchange-correlation effects in the form of the Perdew–Burke–Ernzerhof energy functional [17]. Goedecker–Teter–Hutter norm-conserving pseudopotentials [18] were used with the following valence configurations: La ($5s^25p^65d^16s^2$), Ga ($3d^{10}4s^24p^1$), and O ($2s^22p^4$). Periodic boundary conditions were applied with a single Γ point in the Brillouin zone. A Born-Oppenheimer AIMD simulation was conducted at a time step of 2 fs, and an *NVT* ensemble with a Nose-Hoover thermostat was applied. The initial atomic configuration was created through the CMD simulation of 100 atoms in a cubic cell, which reproduced the experimental density. The structural model of $LaGaO_3$ glass was obtained by holding the initial atomic configuration at 2400 K for 10 ps, lowering the temperature to 300 K at a nominal cooling rate of 100 K/ps, and equilibration at 300 K for 10 ps.

3 Results and Discussion

3.1. Local Structures of $LaGaO_3$ Glass

Fig. 2 shows the total correlation functions of $LaGaO_3$ glass. The correlation peaks appearing at approximately 1.8 and 2.5 Å correspond to the Ga–O and La–O bonds, respectively, revealing that the average bond lengths of Ga–O and La–O in $LaGaO_3$ glass are 1.84 and 2.46 Å, respectively. The CMD model reproduces the experimental results well in the first nearest-neighbor correlations of Ga–O and La–O, whereas in the AIMD, the peak position of the Ga–O bond is located slightly on the long-distance side. By contrast, at 3–5 Å, the results of AIMD are more consistent with the experimental results in comparison with the CMD, suggesting that the AIMD can better reproduce the correlations of the second nearest-neighbor, such as O–O and cation–cation.

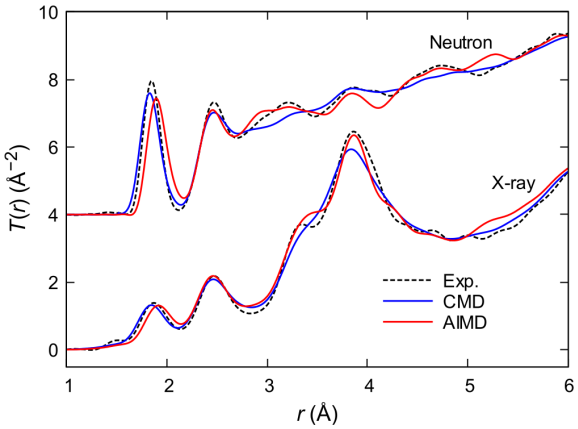


Fig. 2 Experimental and simulated total correlation functions of $LaGaO_3$ glass.

In addition, Fig. 3 shows the coordination number distribution of $LaGaO_3$ glass and the connectivity between the cation-oxygen polyhedra. It was confirmed that the Ga atoms formed a 4–6 coordination with O. The average coordination numbers of Ga–O and $N_{\mathrm{Ga-O}}$ in the CMD and AIMD show similar values of 4.3 and 4.4, respectively. By contrast, the average coordination number of La–O, *i.e.*, $N_{\mathrm{La-O}}$, is 7.4 and 7.8 in CMD and AIMD, respectively, the latter of which shows a larger value. From Fig. 3(e), it can be seen that the coordination number of O–Ga, $N_{\mathrm{O-Ga}}$, is 0–3. That is, in $LaGaO_3$ glass, there are four types of oxygen depending on the coordination environment: free oxygen, O_f ($N_{\mathrm{O-Ga}} = 0$); non-bridging oxygen, O_{nb} ($N_{\mathrm{O-Ga}} = 1$); bridging oxygen, O_b ($N_{\mathrm{O-Ga}} = 2$); and tricluster oxygen, O_t ($N_{\mathrm{O-Ga}} = 3$). The average coordination numbers of O–La, *i.e.*, $N_{\mathrm{La-O}}$, in O_f, O_{nb}, O_b, and O_t were 3.9, 2.8, 1.8, and 1.0 for CMD and 4.0, 2.9, 2.2, and 1.2 for AIMD, respectively. This means that each oxygen atom coordinates with approximately four cations on average, regardless of the $N_{\mathrm{O-Ga}}$ value.

Regarding the cation–oxygen polyhedral connectivity, the fractions of corner-sharing GaO_x–GaO_x and those of edge-

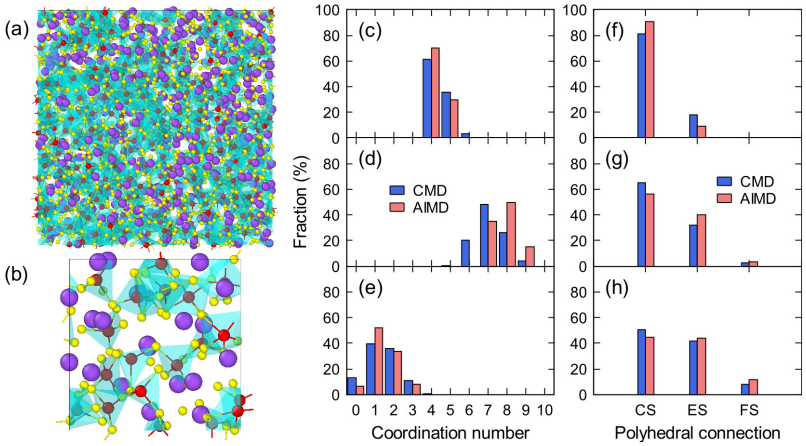


Fig. 3 Structural models for $LaGaO_3$ glass obtained through (a) CMD and (b) AIMD simulations. La, Ga, and O atoms are colored purple, red, and yellow, respectively. The coordination number distribution of (c) O around Ga, (d) O around La, and (e) Ga around O for the structural models of $LaGaO_3$ glass. The polyhedral connections of (f) GaO_x–GaO_x, (g) GaO_x–LaO_y, and (h) LaO_y–LaO_y linkages for the structural models of $LaGaO_3$ glass. CS, ES, and FS represent corner-, edge-, and face-sharing of oxygen, respectively.

and face-sharing GaO_x–LaO_y and LaO_y–LaO_y are slightly higher in AIMD than in CMD. The AIMD results show that the majority of GaO_x–GaO_x linkages form corner sharing, whereas approximately 10% form edge-sharing. Moreover, for the GaO_x–LaO_y and LaO_y–LaO_y linkages, the fractions of edge- and face-sharing connectivity were even higher. Some structural features confirmed in $LaGaO_3$ glass, such as edge-sharing cation–oxygen polyhedral linkages and oxygen triclusters, are unusual in NWF oxides such as SiO_2, B_2O_3, and P_2O_5, and are unique to glass systems that do not contain NWF oxides [14], [19], [20].

3.2. Correlation between Glass Structures and Vibration Characteristics

Fig. 4 shows the Raman scattering spectra [6] and calculated VDOS. The Raman scattering spectra showed distinct peaks at 300, 530, and 650 cm^{-1}. However, VDOS showed peaks at 100 and 600 cm^{-1} and a broad shoulder at 700–850 cm^{-1}. It is not always possible to compare the spectral shape of the total VDOS and the Raman spectra because the cross-section of Raman scattering differs depending on the structural units and modes. To compare the peak intensity and peak position of VDOS with Raman spectroscopy, it is necessary to evaluate the partial VDOS corresponding to certain structural units and modes. Fig. 4(b) also displays the partial VDOS of the La, Ga, and O atoms. The vibrational spectrum of the La atoms was symmetric and located at 100 cm^{-1}. The vibration of the Ga atoms has a broad peak at 180 cm^{-1} and a slight shoulder at approximately 630 cm^{-1}. The O atoms showed a large contribution within the entire range. These results indicate that the high-frequency range (above 400 cm^{-1}) is dominated by the vibration of O bonded with Ga.

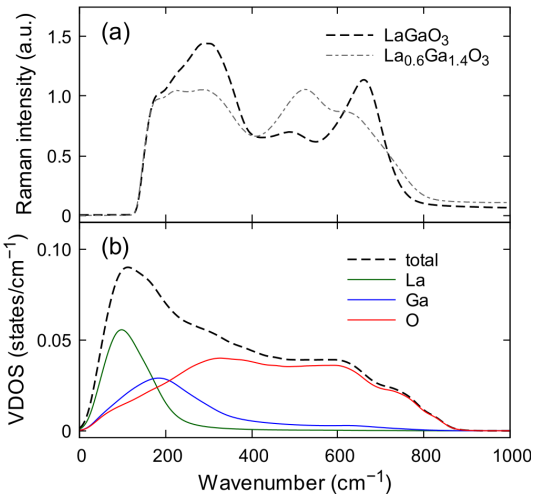


Fig. 4 (a) The Raman scattering spectra of $LaGaO_3$ and $La_{0.6}Ga_{1.4}O_3$ glasses [6]. The Raman spectra at below 180 cm^{-1} was cut off by a filter to remove Rayleigh scattering. (b) The calculated total VDOS and the partial VDOS of La, Ga, and O atoms for $LaGaO_3$ glass.

As shown in Fig. 5, the vibrational spectrum of the O atoms was further deconvoluted into the contributions of O_f, O_{nb}, O_b, and O_t. In addition, O_f shows a nearly symmetrical peak at 320 cm^{-1}. Because O_f corresponds to the oxygen atoms coordinated only with La, it is suggested that the peak at near 300 cm^{-1} within the Raman spectrum is caused by the vibration of the La–O bonds. It has also been reported that in La_2O_3-containing glass, a Raman signal at approximately 300 cm^{-1} was assigned to the stretching mode of the La–O bonds [21], supporting the results of this study.

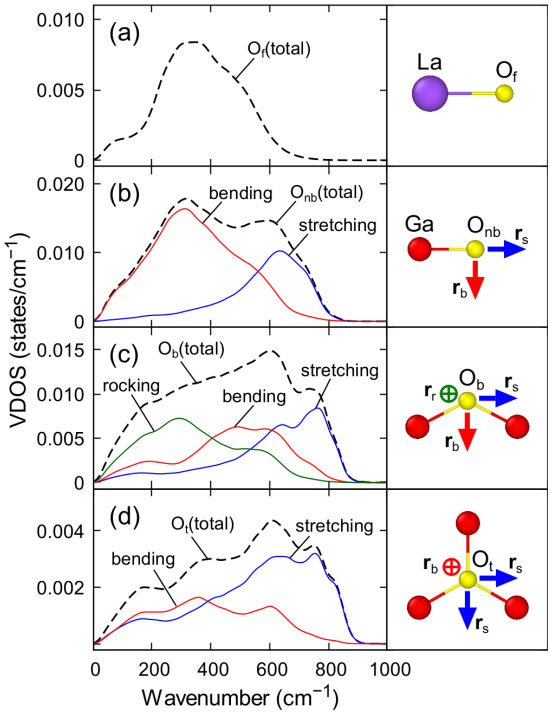


Fig. 5 Partial VDOS of $LaGaO_3$ glass for the projections onto the vibrations of (a) O_f, (b) O_{nb}, (c) O_b, and (d) O_t. The structural units and atomic displacements are schematically shown on the right side of the Figure. In addition, $\mathbf{r}_s$, $\mathbf{r}_b$, and $\mathbf{r}_r$ denote the displacement vectors representing the stretching, bending, and rocking modes, respectively.

According to a previous study conducted by Fukumi *et al.*, the Raman peak at 530 cm^{-1} was assigned to the Ga–O–Ga bending mode, and the peak at 650 cm^{-1} was assigned to the stretching vibration of GaO_x polyhedra, including non-bridging oxygen [22]. In the VDOS, the bending mode of the Ga–O_b–Ga linkage shows a large contribution to the region at 500 cm^{-1}, which is consistent with the assignment at 530 cm^{-1} by Fukumi *et al.* From Fig. 5, the stretching modes of the Ga–O_{nb} bonds and Ga–O_b–Ga linkages show a large contribution at 650 cm^{-1}, which supports the claim that the component at 650 cm^{-1} within the Raman spectra is attributed to the vibration of GaO_x polyhedra, including non-bridging oxygen.

In $LaGaO_3$ glass, the fraction at approximately 750 cm^{-1} within the Raman spectrum is weak, whereas in $La_{0.6}Ga_{1.4}O_3$ glass with a larger Ga_2O_3 content, a broad shoulder peak appears within this high-frequency region (Fig. 4). This high-frequency component was not assigned in the previous study by Fukumi *et al.* [22], and its structural origin is unknown. From the results of VDOS, the contribution of the stretching modes of Ga–O_b–Ga linkages and Ga_3–O_t tri-clusters are dominant at approximately 750 cm^{-1}, indicating that this high-frequency component originates from the bridging and tri-cluster oxygens. Because the vibration mode at 750 cm^{-1} is at the highest frequency in the La_2O_3–Ga_2O_3 glass system, it is strongly expected that the influence on the optical properties in the infrared region will be significant. We previously reported that the infrared absorption edge of La_2O_3–Ga_2O_3 glass shifted to a longer wavelength as the La_2O_3 content increased [6]. It is considered that this compositional dependence of the infrared absorption edge is due to the decrease in the bridging and tri-cluster oxygen with increasing La_2O_3 content, which results in a decrease in the vibrational component at 750 cm^{-1}.

3.3. Analysis of Electronic Structures

Fig. 6 shows the projected electronic density of states (DOS) of $LaGaO_3$ glass derived from the AIMD structural model. The simulated band gap energy of $LaGaO_3$ glass was 3.3 eV, which is smaller than the experimental value of 4.4 eV [6]. This is because of the well-known trend in which DFT with GGA inherently underestimates the band gap. Despite this limitation, DFT calculations are still useful for analyzing the band structures. The band between −19 and −12 eV is composed of La5p, Ga3d, and O2s orbitals, whereas the upper valence band between −7 eV and the Fermi level (0 eV) is dominated by the O2p orbital. In the upper valence band, the contribution of La and Ga by hybridization with the O2p orbital is weak, indicating the high ionic nature of oxygen. Further deconvolution of the electronic DOS into each type of oxygen atom (O_f, O_{nb}, O_b, and O_t) reveals that the top of the valence band is composed of O_f and O_{nb}. In addition, our simulation results suggest that the lower part of the conduction band is mainly composed of La5d and La4f orbitals.

We also determined the effective charge of each atom in $LaGaO_3$ glass using the Bader method [23]. The effective charges of La, Ga, and O atoms were derived as +2.08*e*, +1.76*e*, and −1.28*e*, respectively, which also supports the high ionicity of $LaGaO_3$ glass. The effective charges of O_f, O_{nb}, O_b, and O_t were derived as −1.30*e*, −1.28*e*, −1.27*e*, and −1.26*e*, respectively, indicating that oxygen with smaller $N_{O\text{–}Ga}$ shows a higher effective charge. This is because La has a lower electronegativity than Ga (1.10 and 1.81 for La and Ga, respectively) [24], and the La–O bond is more ionic than the Ga–O bond and shows a greater electron donation to the oxygen. However, the difference in the effective charge between each type of oxygen was not significantly large; thus, the electron density may be delocalized throughout the oxygen atoms. The experimentally determined average oxygen polarizabilities of $LaGaO_3$ glass were 2.36 Å^3 [6], which is significantly larger than that of other

oxide glasses containing NWF oxides [25]. From the results of our DFT calculations, the unique characteristics such as a large oxygen polarizability and high refractive index of $LaGaO_3$ glass can be understood by the high electron charge density of oxygen, which is significantly enhanced by the high electron donating ability of La.

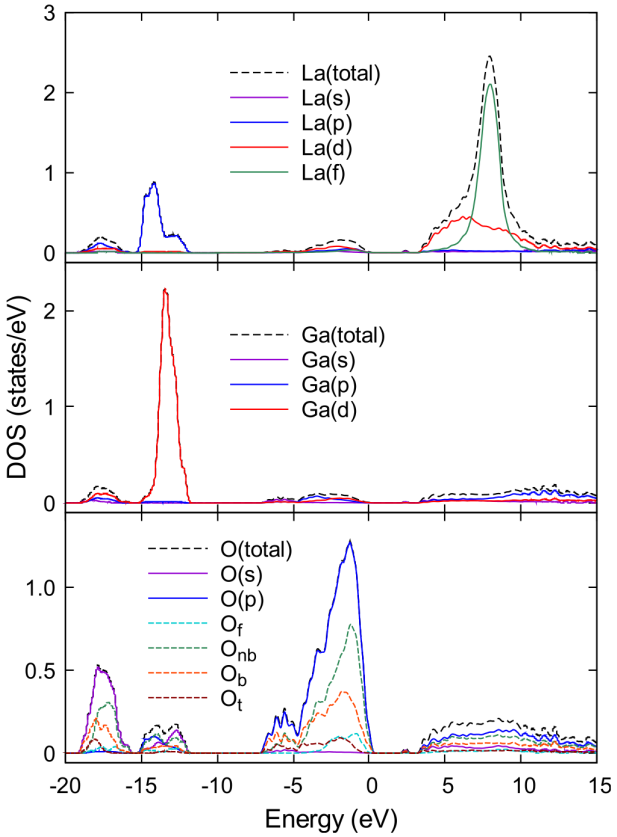


Fig. 6 The projected electronic DOS for $LaGaO_3$ glass.

4 Conclusion

We analyzed the local atomic structures, vibrational characteristics, and electronic structures of $LaGaO_3$ glasses prepared through a containerless processing. The structural models constructed by the diffraction experiments, CMD, and AIMD simulations revealed the average bond distances, average coordination numbers, and polyhedral connectivity of $LaGaO_3$ glass. Several unique structural features, such as oxygen tri-clusters and edge-sharing GaO_x–GaO_x polyhedral linkages, which are different from those of NWF oxides, were confirmed. The analysis of the calculated VDOS for $LaGaO_3$ glass showed good agreement with the assignments of the Raman spectrum. It was revealed that the broad component at the highest frequency (750 cm^{-1}) originated from the stretching vibrations of the bridging and tri-cluster oxygen, which explains the compositional dependence of the infrared absorption edge in the La_2O_3–Ga_2O_3 glass system. The projected electronic DOS provides clear insight into the band structures of $LaGaO_3$ glass, *i.e.*, the upper part of the valence band is dominated by the O2p orbital, and the conduction band is mainly composed of La5d and La4f orbitals. A Bader charge analysis indicates that La, with its high electron-donating ability, significantly enhances the electron density of the outer valence of oxygen, resulting in a high oxygen polarizability and high refractive index of $LaGaO_3$ glass.

Acknowledgements. The synchrotron radiation experiments were conducted at BL04B2 of SPring-8 with the approval of the Japan Synchrotron Radiation Research Institute (proposal nos. 2015A1366 and 2018A1251). The neutron diffraction experiment at the NOVA of J-PARC was approved by the Neutron Scattering Program Advisory Committee of IMSS, KEK (Proposal No. 2014S06). This study was supported in part by JSPS KAKENHI (grant nos. JP25410236, JP26249092, JP17H03121, JP18K18928, and 19H05163). We acknowledge Dr. Shinji Kohara (National Institute for Materials Science) and Dr. Koji Ohara (Japan Synchrotron Radiation Research Institute) for their help with the high-energy X-ray diffraction experiments and analysis at SPring-8, and Dr. Yohei Onodera (Kyoto University), Mr. Tomoki Oya (Hirosaki University), and Dr. Kazutaka Ikeda (High Energy Accelerator Research Organization) for their help with the neutron diffraction experiment and analysis at J-PARC. We also acknowledge Dr. Kensaku Nakamura (Nikon) and Dr. Shou Miyasaka (Nikon) for discussions on the structural modeling of the glass.

References

[1] W. H. Dumbaugh, "Heavy metal oxide glasses containing Bi_2O_3," *Phys. Chem. Glasses*, vol. 27, no. 3, pp. 119–123, 1986.

[2] S. Sakka, "Formation, structure, and properties of unconventional glasses containing gallium oxide," *Glass Phys. Chem.*, vol. 24, no. 3, pp. 257–267, 1998.

[3] J. Yu, Y. Arai, T. Masaki, T. Ishikawa, S. Yoda, S. Kohara, H. Taniguchi, M. Itoh, and Y. Kuroiwa, "Fabrication of $BaTi_2O_5$ glass-ceramics with unusual dielectric properties during crystallization," *Chem. Mater.*, vol. 18, no. 8, pp. 2169–2173, 2006.

[4] A. Masuno, H. Inoue, K. Yoshimoto, and Y. Watanabe, "Thermal and optical properties of La_2O_3–Nb_2O_5 high refractive index glasses," *Opt. Mater. Express*, vol. 4, no. 4, pp. 710–718, 2014.

[5] G. A. Rosales-Sosa, A. Masuno, Y. Higo, H. Inoue, Y. Yanaba, T. Mizoguchi, T. Umada, K. Okamura, K. Kato,

and Y. Watanabe, "High elastic moduli of a $54Al_2O_3$–$46Ta_2O_5$ glass fabricated via containerless processing," *Sci. Rep*., vol. 5, no, 15233, 2015.

[6] K. Yoshimoto, A. Masuno, M. Ueda, H. Inoue, H. Yamamoto, and T. Kawashima, "Low phonon energies and wideband optical windows of La_2O_3–Ga_2O_3 glasses prepared using an aerodynamic levitation technique," *Sci. Rep*., vol. 7, no. 45600, 2017.

[7] K. Yoshimoto, Y. Ezura, M. Ueda, A. Masuno, and H. Inoue, "2.7 μm mid-infrared emission in highly erbium-doped lanthanum gallate glasses prepared via an aerodynamic levitation technique," *Adv. Opt. Mater.*, vol. 6, no. 1701283, 2018.

[8] S. Kohara, M. Itou, K. Suzuya, Y. Inamura, Y. Sakurai, Y. Ohishi, and M. Takata, "Structural studies of disordered materials using high-energy X-ray diffraction from ambient to extreme conditions," *J. Phys.: Condens. Matter*, vol. 19, no. 506101, 2007.

[9] T. Otomo, K. Suzuya, M. Misawa, N. Kaneko, H. Ohshita, K. Ikeda, M. Tsubota, T. Seya, T. Fukunaga, K. Itoh, *et al.* "Fundamental research of hydrogen storage mechanism with high-intensity total diffractometer," *KENS Rep*. vol. 17, pp. 28–36, 2011.

[10] E. Lorch, "Neutron diffraction by germania, silica and radiation-damaged silica glasses," *J. Phys. C: Solid State Phys.*, vol. 2, no. 2, pp. 229–237, 1969.

[11] S. Plimpton, "Fast parallel algorithms for short-range molecular dynamics," *J. Comput. Phys.*, vol. 117, no. 1, pp. 1–19, 1995.

[12] K. Yoshimoto, A. Masuno, I. Sato, Y. Ezura, H. Inoue, M. Ueda, M. Mizuguchi, Y. Yanaba, T. Kawashima, T. Oya, *et al.* "Principal vibration modes of the La_2O_3–Ga_2O_3 binary glass originated from diverse coordination environments of oxygen atoms," vol. 124, no. 24, pp. 5056–5066, 2020.

[13] S. N. Taraskin and S. R. Elliott, "Nature of vibrational excitations in vitreous silica," *Phys. Rev. B*, vol. 56, no. 14, pp. 8605–8622, 1997.

[14] H. Inoue, A. Masuno, S. Kohara, and Y. Watanabe, "The local structure and vibrational properties of $BaTi_2O_5$ glass revealed by molecular dynamics simulation," *J. Phys. Chem. B*, vol. 117, no. 22, pp. 6823–6829, 2013.

[15] J. VandeVondele, M. Krack, F. Mohamed, M. Parrinello, T. Chassaing, and J. Hutter, "Quickstep: Fast and accurate density functional calculations using a mixed Gaussian and plane waves approach," *Comput. Phys. Commun.*, vol. 167, no. 2, pp. 103–128, 2005.

[16] J. VandeVondele and J. Hutter, "Gaussian basis sets for accurate calculations on molecular systems in gas and condensed phases," *J. Chem. Phys.*, vol. 127, no. 11, pp. 114105–114109, 2007.

[17] J. P. Perdew, K. Burke, and M. Ernzerhof, "Generalized gradient approximation made simple," *Phys. Rev. Lett.*, vol. 77, no. 18, pp. 3865–3868, 1996.

[18] S. Goedecker, M. Teter, and J. Hutter, "Separable dual-space Gaussian pseudopotentials," *Phys. Rev. B*, vol. 54, no. 3, pp. 1703–1710, 1996.

[19] J. Du, C. J. Benmore, R. Corrales, R. T. Hart, and J. K. R. Weber, "A molecular dynamics simulation interpretation of neutron and x-ray diffraction measurements on single phase Y_2O_3–Al_2O_3 glasses," *J. Phys.: Condens. Matter*, vol. 21, no. 205102, 2009.

[20] A. Masuno, S. Kohara, A. C. Hannon, E. Bychkov, and H. Inoue, "Drastic connectivity change in high refractive index lanthanum niobate glasses," *Chem. Mater.*, vol. 25, no. 15, pp. 3056–3061, 2013.

[21] L. G. Hwa, J. G. Shiau, and S. P. Szu, "Polarized Raman scattering in lanthanum gallogermanate glasses," *J. Non-Cryst. Solids*, vol. 249, no. 1, pp. 55–61, 1999.

[22] K. Fukumi and S. Sakka, "Raman spectra of binary alkali and alkaline earth gallate crystals and glasses," *Phys. Chem. Glasses*, vol. 29, no. 1, pp. 1–8, 1988.

[23] W. Tang, E. Sanville, and G. A. Henkelman, "A grid-based Bader analysis algorithm without lattice bias," *J. Phys.: Condens. Matter*, vol. 21, no. 084204, 2009.

[24] L. Pauling, "The nature of the chemical bond. IV. The energy of single bonds and the relative electronegativity of atoms," *J. Am. Chem. Soc.*, vol. 54, no. 9, pp. 3570–3582, 1932.

[25] V. Dimitrov and T. Komatsu, "Electronic polarizability, optical basicity and non-linear optical properties of oxide glasses," *J. Non-Cryst. Solids*, vol. 249, no. 2–3, pp. 160–179, 1999.

吉本幸平　Kohei YOSHIMOTO
研究開発本部　材料・要素技術研究所
Materials & Advanced Research Laboratory
Research & Development Division

増野敦信　Atsunobu MASUNO
弘前大学
Hirosaki University

佐藤　至　Itaru SATO
研究開発本部　材料・要素技術研究所
Materials & Advanced Research Laboratory
Research & Development Division

江面嘉信　Yoshinobu EZURA
研究開発本部　材料・要素技術研究所
Materials & Advanced Research Laboratory
Research & Development Division

井上博之　Hiroyuki INOUE
東京大学
The University of Tokyo

上田　基　Motoi UEDA
研究開発本部　材料・要素技術研究所
Materials & Advanced Research Laboratory
Research & Development Division

水口雅史　Masafumi MIZUGUCHI
研究開発本部　材料・要素技術研究所
Materials & Advanced Research Laboratory
Research & Development Division

山本優也　Yuya YAMAMOTO
研究開発本部　材料・要素技術研究所
Materials & Advanced Research Laboratory
Research & Development Division


吉本幸平
Kohei YOSHIMOTO


増野敦信
Atsunobu MASUNO


佐藤　至
Itaru SATO


江面嘉信
Yoshinobu EZURA


井上博之
Hiroyuki INOUE


上田　基
Motoi UEDA


水口雅史
Masafumi MIZUGUCHI


山本優也
Yuya YAMAMOTO

UCB 獲得関数と選定記述子を用いたガラス組成のベイズ最適化

中村健作，大谷直也，小池哲也

Bayesian Optimization of Glass Compositions with Upper Confidence Bound and Selected Descriptors

Kensaku NAKAMURA, Naoya OTANI and Tetsuya KOIKE

ニコンは100年以上にわたり光学ガラスの研究開発を実施しており，光学ガラスはカメラや顕微鏡など多数のニコン製品に搭載されている．光学ガラス開発において，“ガラスの組成設計”は，所望の物理特性を有するようにガラス構成元素の種類や量を調整する重要なプロセスであり，専門的な知識や経験に基づく試行錯誤が必要となる．一方，近年，機械学習を用いることで材料開発を加速する試みが注目されている．本研究では，機械学習による光学ガラス開発の高速化を目指す取組みの一環として，機械学習の一手法であるベイズ最適化を組成設計に適用する．具体的には，国際ガラスデータベース INTERGLAD のデータにおいて，ベイズ最適化を用いて高アッベ数組成を探索する．さらに，獲得関数 Upper Confidence Bound のパラメーター調整および，機械学習モデルの入力パラメーター，すなわち，記述子の選定が，ベイズ最適化の探索性能に与える効果を検証する．

Nikon has developed optical glass for over 100 years, and the optical glass has been installed in many Nikon products such as cameras and microscopes. In the development of the optical glass, the composition design of glass is important, in which the types and amounts of constituent elements are adjusted to obtain glass with desirable physical properties (e.g., refractive index and Abbe number), and numerous trials and errors based on the knowledge and experience of experts are required. By contrast, in material sciences, attempts to accelerate material developments using machine learning has been reported recently. In this study, we apply Bayesian optimization, a machine learning method, to the composition design of glass to accelerate the development of optical glass. It is demonstrated that compositions with high Abbe numbers can be identified using Bayesian optimization based on data from the International Glass Database, INTERGLAD. In addition, we discuss the effects of setting the parameter of an acquisition function, upper confidence bounds, and descriptors to the search performance of Bayesian optimization.

Key words 光学ガラス，組成設計，アッベ数，機械学習，ベイズ最適化
optical glass, composition design, Abbe number, machine learning, Bayesian optimization

1 Introduction

Glass is used in various applications, such as camera lenses, windows, and electronic displays. The specifications of glass properties differ depending on the application. For example, in the development of optical products such as camera lenses, the optical properties of glass, such as its refractive index and Abbe number ν_d, must be adjusted to satisfy the specifications [1],[2]. Composition design is a typical method for controlling the physical properties of glass because its properties depend significantly on its composition. Generally, composition design requires the knowledge and experience of experts; furthermore, it is time consuming because the number of element combinations is significant.

Recently, machine learning has garnered attention as an effective tool for accelerating the development of materials, including glass [3]–[10]. We have previously focused on one of the machine learning methods, i.e., Bayesian optimization (BO), which proposes the next experimental condition (e.g., chemical compositions) based on previous experimental data. BO has been applied to the development of various materials such as thermoelectric materials, shape-memory alloys, and oxide glass [7]–[10]. Unlike other optimization methods, BO can search for the next experimental condition in an extrapolated area because it employs acquisition functions that indicate the effectiveness of the experiment based on the predicted values and their uncertainties [11], [12].

Several types of acquisition functions are known, and we have focused on one of the acquisition functions, i.e., the upper confidence bound (UCB), which can achieve a balance between exploitation and exploration by setting a parameter for experiments [13]–[15]. For example, in glass development, if a glass composition that differs significantly from observed ones is required, then the exploration of BO should be enhanced by adjusting the parameter. Hence, BO with the UCB acquisition function is expected to benefit the composition design of glass.

In this study, we applied BO with the UCB acquisition function to optimize glass compositions to identify high ν_d compositions using the International Glass Database (INTERGLAD) [16]. We present the dependence of the search performance of BO on the balance between exploitation and exploration. Subsequently, we discuss the effect on the search performance with respect to the selection of input variables, i.e., descriptors, using random forest (RF) analysis. Generally, the selection of input variables is important to achieve good performances in machine learning [6]–[8], [10].

2 Methods

The composition and ν_d data were obtained from the INTERGLAD [16]. Some compositions that exhibited incorrect values were removed. We used the data of only the silicate system for which the amount of SiO_2 was more than 0 mol%. A total of 7181 compositions were used. The composition included the following 57 components: Al_2O_3, As_2O_3, B_2O_3, BaO, BeO, Bi_2O_3, CaO, CdO, Ce_2O_3, CeO_2, Co_2O_3, CoO, Cs_2O, CuO, Dy_2O_3, Er_2O_3, Fe_2O_3, Ga_2O_3, Gd_2O_3, GeO_2, HfO_2, In_2O_3, K_2O, La_2O_3, Li_2O, Lu_2O_3, MgO, MnO, MnO_2, MoO_2, MoO_3, Na_2O, Nb_2O_3, Nb_2O_5, Nd_2O_3, NiO, P_2O_5, PbO, Pr_2O_3, Rb_2O, SO_3, Sb_2O_3, Sb_2O_5, Sc_2O_3, SiO_2, Sm_2O_3, SnO, SnO_2, SrO, Ta_2O_5, TeO_2, TiO_2, Tl_2O, WO_3, Y_2O_3, ZnO, and ZrO_2. Fig. 1 shows the histogram of ν_d for the compositions. A histogram of the appearance of the components is shown in Fig. 2. We set the target value of ν_d to 70 and analyzed the speed in which BO identifies a composition with a ν_d exceeding 70. Approximately 1% of the total compositions indicated ν_d values exceeding 70.

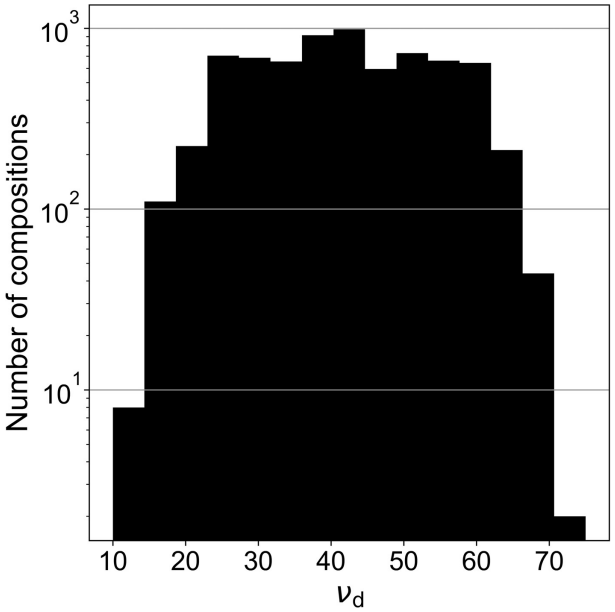


Fig. 1 Histogram of ν_d for collected compositions. Logarithmic scale is used for y-axis.

We used typical descriptors based on elemental physical properties [4], [6]–[10]. The descriptors were calculated from the numbers of elements in the compositions and the following 11 elemental properties: atomic number, Mendeleev number, column and row numbers in the periodic table, covalent radius, Ahrens ionic radius, electronegativity, first ionization energy, melting point, atomic weight, and density [17]–[21]. Specifically, two descriptors, mean x_{mean} and standard deviation x_{std}, were calculated for each property in each composition as follows:

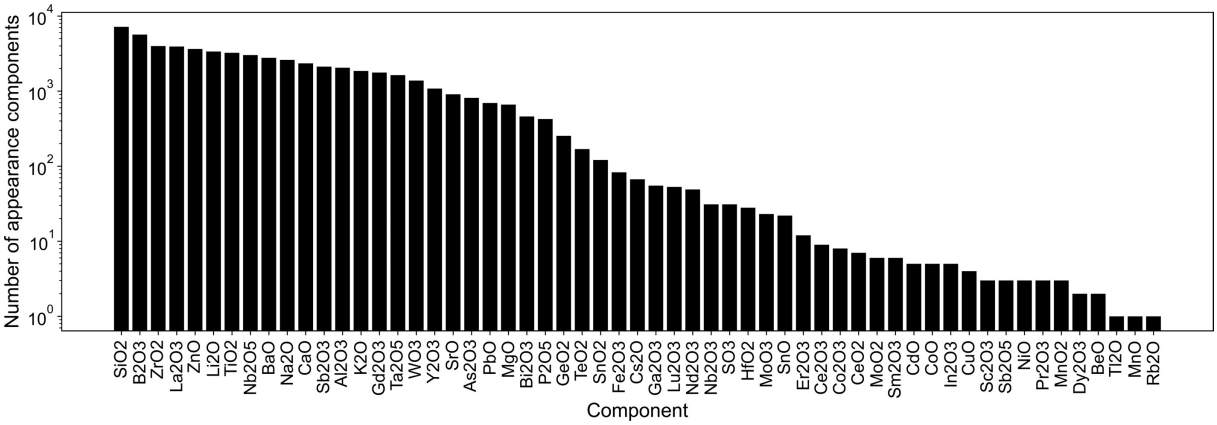


Fig. 2 Histogram of appearance components for collected compositions. Logarithmic scale is used for y-axis.

$$x_{\mathrm{mean}} = \frac{\sum_i y_i x_i}{\sum_i x_i}, \tag{1}$$

$$x_{\mathrm{std}} = \sqrt{\frac{\sum_i (y_i - x_{\mathrm{mean}})^2 x_i}{\sum_i x_i}}, \tag{2}$$

where i represents the element species, x_i the atomic fraction in the composition, and y_i the value of the physical property. In total, 22 descriptors were used for each composition. The set of descriptors was the same as in those in previous studies [7], [8]. The descriptors in the training data were normalized using the mean and standard deviation for each descriptor, i.e., the mean and variance of the values of the descriptors were set to zero and one, respectively. Descriptors are often normalized to equalize their scales [8]. Furthermore, the importance score for each descriptor in the prediction of v_d was calculated by fitting all the data (compositions and their v_d values) via RF regression. RF regression is a decision tree ensemble method that can output the importance of each descriptor [5], [17]. The descriptors with high importance contribute significantly to the prediction of v_d. We executed RF regression using the scikit-learn package [22]. Fig. 3 shows the importance of each descriptor. We analyzed the effect of descriptor selection on the BO search performance by comparing two cases. In the first case, all descriptors were used. In the second case, the following 11 descriptors with higher importance were used: density x_{mean} and x_{std}, Ahrens ionic radius x_{mean} and x_{std}, atomic weight x_{mean} and x_{std}, row numbers in the periodic table x_{mean}, column numbers in the periodic table x_{mean}, atomic number x_{mean} and x_{std}, and melting point x_{std}.

The procedure for BO in this study is as follows: five compositions were randomly selected as initial training data. The remaining compositions were composed as initial test data. The training data were fitted using Gaussian process regression. Gaussian process regression is a Bayesian inference method that outputs the uncertainty of prediction and the predicted value, and it is typically used in BO. We used a GPy library to implement Gaussian process regression [23]. We used a typical kernel function, i.e., the Gaussian kernel, for Gaussian process regression. Using Gaussian process regression, the predicted values and uncertainties (i.e., standard deviations) were obtained for each composition of the test data. Subsequently, the UCB acquisition functions a_{UCB} for the compositions were calculated as the criterion, as follows:

$$a_{\mathrm{UCB}} = \mu + \kappa\sigma, \tag{3}$$

where μ and σ are the predicted values and standard deviation, respectively; κ is a hyperparameter that controls the balance between exploitation and exploration. Although several expressions for the UCB have been proposed [13]–[15], we used a simple one, as shown in Eq. (3), which comprises only three parameters: μ, σ, and κ. By setting κ to a higher value, a composition different from that in the training data is proposed for BO. In this study, we performed BO with different values of κ to evaluate the dependence of BO performance on the balance between exploitation and exploration. We performed an experiment and observed the result under the condition with the highest values of acquisition functions in BO. Subsequently, we observed the v_d of a composition with the highest value of a_{UCB}, i.e., we added the composition and its v_d into the training data and removed them from the test data. This process was repeated until the v_d value of the highest a_{UCB} composition exceeded 70. In this study, when a composition with a high v_d was identified via a small number of observations, the search performance of BO was regarded as superior. We executed the BO search for 50 patterns of the initial training data at each κ value.

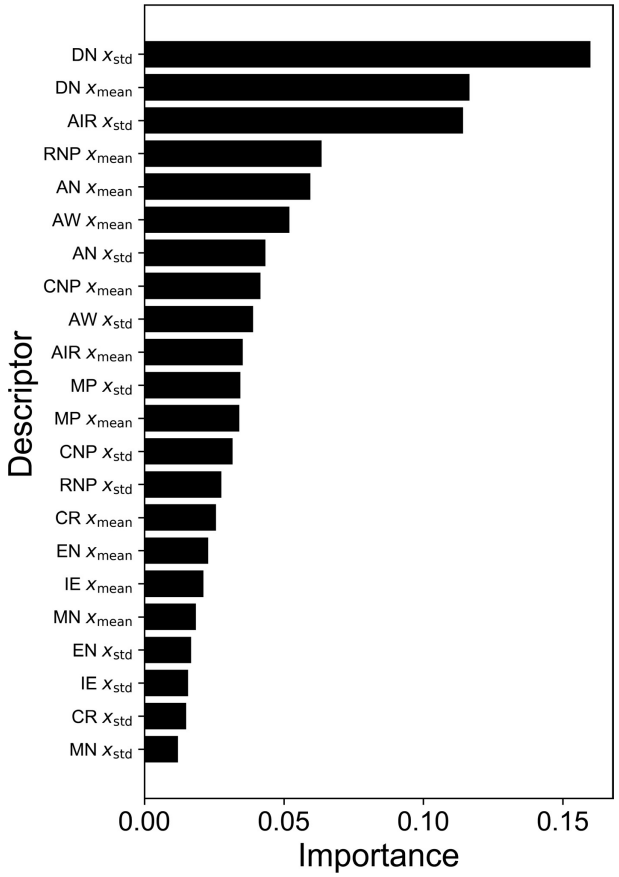


Fig. 3 Importance scores for descriptors calculated using RF. Abbreviations: DN, density; AIR, Ahrens ionic radius; RNP, row numbers in the periodic table; AN, atomic number: CNP, column numbers in the periodic table; AW, atomic weight; MP, melting point; CR, covalent radius; EN, electronegativity; IE, first ionization energy; MN, Mendeleev number.

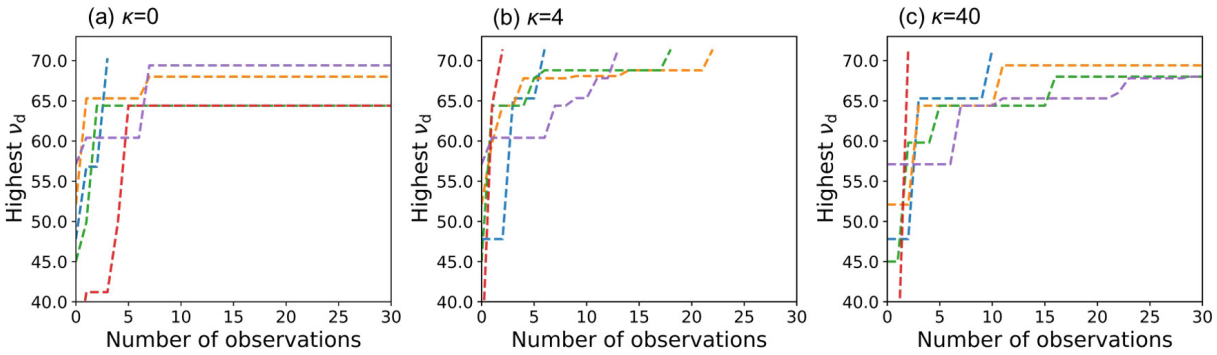


Fig. 4 The highest v_d values until 30th observation with all (non-selected) descriptors with different initial training data for five patterns at the different values of UCB parameter κ. Colors and dotted lines indicate the results for individual initial training data.

3 Results and Discussion

Fig. 4 shows the typical results of the BO in 30 observations for five patterns of the initial training data for different values of κ with 22 descriptors. As shown in Fig. 4, when κ = 4, compositions with v_d > 70 were identified until the 30th observation. However, for the other values of κ, compositions with a high v_d were not identified. BO with appropriate κ values enable compositions with high v_d to be identified rapidly. Fig. 5 shows the relationship between κ and the average number of BO observations required to identify compositions with a high v_d in 50 patterns of the initial training data using all the selected descriptors. In both descriptors, when κ is zero, the number of observations is high. In the case involving all descriptors, when the value of κ was less than 20, the average number of observations became the minimum. Subsequently, when the value of κ exceeded 20, the average number of observations increased. Because a large κ indicates that the uncertainty in the UCB (Eq. (3)) is significant, a vast composition region is searched during BO and compositions with a high v_d cannot be identified. By contrast, in the case involving selected descriptors, when κ is 20 or more, the average number of observations becomes the minimum and is similar for each κ. The average number of observations was smaller when the selected descriptors were used compared with when all descriptors were used. Therefore, these results suggest that tuning the UCB parameter and selecting descriptors can improve the search performance of BO. It is noteworthy that when using the selected descriptors, as κ increases, the average number of observations does not decrease, unlike the case for all descriptors. We speculate that the effect of uncertainty is less prominent when using the selected descriptors than when using all descriptors for a large value of κ in this study because the dimensions of the selected descriptors are smaller than those of all the descriptors.

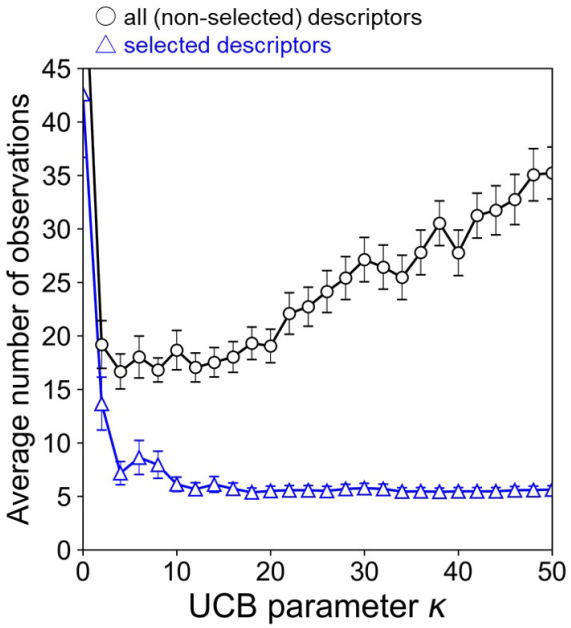


Fig. 5 Average number of observations required to identify v_d exceeding 70 for various values of κ using all (non-selected) and selected descriptors.

4 Conclusion

We demonstrated that BO with a UCB acquisition function enabled compositions with high v_d to be identified using data from the INTERGLAD. We demonstrated that the search performance of BO depended significantly on the UCB. Furthermore, BO with selected descriptors based on their importance scores obtained from RF was more effective in identifying compositions with high v_d than BO with all descriptors. Therefore, parameter tuning and the selection of appropriate descriptors are crucial for rapidly identifying compositions with desirable properties.

Acknowledgements. We would like to thank Dr. M. Mizuguchi, M. Ueda, K. Yoshimoto, and T. Kawashima of the Materials & Advanced Research Laboratory, Nikon Corpora-

tion, for providing advice regarding glass science.

References

[1] K. Yoshimoto, A. Masuno, M. Ueda, H. Inoue, H. Yamamoto, and T. Kawashima, "Thermal and optical properties of La_2O_3–Ga_2O_3–(Nb_2O_5 or Ta_2O_5) ternary glasses," *Journal of the American Ceramic Society*, vol. 101, pp. 3328–3336, 2018.

[2] K. Yoshimoto, A. Masuno, M. Ueda, H. Inoue, H. Yamamoto, and T. Kawashima, "Low phonon energies and wideband optical windows of La_2O_3–Ga_2O_3 glasses prepared using an aerodynamic levitation technique," *Scientific Reports*, vol. 7, p. 45600, 2017.

[3] D. R. Cassar, A. C. P. L. F. de Carvalho, and E. D. Zanotto, "Predicting glass transition temperatures using neural networks," *Acta Materialia*, vol. 159, pp. 249–256, 2018.

[4] D. R. Cassar, "ViscNet: Neural network for predicting the fragility index and the temperature-dependency of viscosity," *Acta Materialia*, vol. 206, p. 116602, 2021.

[5] E. Alcobaça, S. Mastelini, T. Botari, B. Pimentel, D. Cassar, A. Carvalho, and E. Zanotto, "Explainable machine learning algorithms to predict glass transition," *Acta Materialia*, vol. 188, pp. 92–100, 2020.

[6] K. Nakamura, N. Otani, and T. Koike, "PHYSICAL PROPERTY PREDICTING DEVICE, DATA GENERATING DEVICE, PHYSICAL PROPERTY PREDICTING METHOD, AND PROGRAM," (in Japanese), Japan Patent P2020–200213A, 2020.

[7] K. Nakamura, N. Otani, and T. Koike, "Search for oxide glass compositions using Bayesian optimization," *Journal of the Ceramic Society of Japan*, vol. 123, pp. 569–572, 2020.

[8] K. Nakamura, N. Otani, and T. Koike, "Multi-objective Bayesian optimization of optical glass compositions," *Ceramics International*, vol. 47, pp. 15819–15824, 2021.

[9] D. Xue, P. V. Balachandran, J. Hogden, J. Theiler, D. Xue, and T. Lookman, "Accelerated search for materials with targeted properties by adaptive design," *Nature Communications*, vol. 7, p. 11241, 2016.

[10] A. Seko, A. Togo, H. Hayashi, K. Tsuda, L. Chaput, and I. Tanaka, "Prediction of low-thermal-conductivity compounds with first-principles anharmonic lattice-dynamics calculations and Bayesian optimization," *Physical Review Letters*, vol. 115, p. 205901, 2015.

[11] J. Snoek, H. Larochelle, and R. P. Adams, "Practical Bayesian optimization of machine learning algorithms," *Proceedings of the NIPS 2012*, 2012.

[12] B. Shahriari, K. Swersky, Z. Wang, R. P. Adams, and N. de Freitas, "Taking the human out of the loop: A review of Bayesian optimization," *Proceedings of the IEEE*, vol. 104, pp. 148–175, 2016.

[13] P. Auer, N. Cesa-Bianchi, and P. Fischer, "Finite-time analysis of the multiarmed Bandit problem," *Machine Learning*, vol. 47, pp. 235–256, 2002.

[14] P. Auer, "Using confidence bounds for exploitation-exploration trade-offs," *Journal of Machine Learning Research*, vol. 3, pp. 397–422, 2002.

[15] D. D. Cox, and S. John, "A statical method for global optimization," *Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics*, vol. 2, 1992, pp. 1241–1246.

[16] "INTERGLAD Ver.7," International Glass Database System INTERGLAD Ver. 7, NEW GLASS FORUM. http://www.newglass.jp/interglad_n/. 2017.

[17] L. Breiman, "Random forest," *Machine Learning*, vol. 45, pp. 5–32, 2001.

[18] *Chronological Scientific Tables*, Tokyo: Maruzen Publishing Co., Ltd., 2016.

[19] B. Cordero, V. Gómez, A. E. Platero-Prats, M. Revés, J. Echeverría, E. Cremades, F. Barragán, and S. Alvarez, "Covalent radii revisited," *Dalton Transactions*, vol. 21, pp. 2832–2838, 2008.

[20] R. D. Shannon, and C. T. Prewitt, "Effective ionic radii in oxides and fluorides," *Acta Crystallographica Section B Structural Crystallography and Crystal Chemistry*, vol. 25, pp. 925–946, 1969.

[21] P. Villars, K. Cenzual, J. Daams, Y. Chen, and S. Iwata, "Data-driven atomic environment prediction for binaries using the Mendeleev number," *Journal of Alloys and Compounds*, vol. 367, pp. 167–175, 2004.

[22] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[23] "Sheffield ML/GPy - Gaussian process framework in python," GitHub. http://github.com/ShefieldML/GPy. 2020.

中村健作　Kensaku NAKAMURA
研究開発本部　数理技術研究所
Mathematical Sciences Research Laboratory
Research & Development Division

大谷直也　Naoya OTANI
研究開発本部　数理技術研究所
Mathematical Sciences Research Laboratory
Research & Development Division

小池哲也　Tetsuya KOIKE
研究開発本部　数理技術研究所
Mathematical Sciences Research Laboratory
Research & Development Division

中村健作
Kensaku NAKAMURA

大谷直也
Naoya OTANI

小池哲也
Tetsuya KOIKE

表面形状制御による水分散性 ITO ナノ粒子の作製

鈴木涼子，西　康孝，松原正樹，村松淳司，蟹江澄志

Water-Dispersible ITO Nanoparticles Prepared by Surface Shape Control†

Ryoko SUZUKI, Yasutaka NISHI, Masaki MATSUBARA, Atsushi MURAMATSU and Kiyoshi KANIE

近年フレキシブルディスプレーなどの有機基板を用いたデバイスが発展している．有機基板上への透明導電膜の作製においては低温成膜プロセスの確立が重要である．本論文では，このようなプロセスに好適な新形状の酸化インジウムスズ（Indium Tin Oxide, ITO）ナノ粒子の作製について報告する．従来報告されている立方体型 ITO ナノ粒子の作製方法を応用し，核生成・粒子成長プロセスを制御することによって表面に突起を有する ITO ナノ粒子を得た．この ITO ナノ粒子は立方体型 ITO ナノ粒子よりも比表面積が大きく多くの水分子が吸着可能であることから，親水性が高く，非常に高い水分散性を有していることが分かった．この ITO ナノ粒子水分散液はミスト成膜法，スピンコート法，ブラシ塗布法などの低温成膜プロセスへの適用が期待される．

Devices that use an organic substrate, such as flexible displays, have recently become widespread. A low-temperature deposition process is considered an important factor for preparing a transparent conductive film on an organic film. In this paper, we report uniquely shaped indium tin oxide (ITO) nanoparticles (NPs) that exhibit a uniform crystal orientation and are appropriate for use in a low-temperature deposition process. These ITO NPs have numerous protrusions on their surface. These protrusions were formed through control of the nucleation and particle growth of NPs prepared using a method previously developed for preparing cubic ITO NPs. The ITO NPs with protrusions had a larger specific surface area than the cubic ITO NPs because of their shape and could adsorb a greater number of water molecules. Therefore, the ITO NPs with protrusions exhibited greater hydrophilicity and greater dispersibility in water. We expect aqueous dispersions of the ITO NPs with protrusions to find applications in low-temperature deposition processes such as the mist deposition, spin coating, and brush coating methods.

Key words ナノ粒子，ITO，ソルボサーマル法，水分散性，透明導電性酸化物
nanoparticle, ITO, solvothermal synthesis, water dispersibility, transparent conductive oxide

1 Introduction

Indium tin oxide (ITO) is a widely used *n*-type semiconductor [1] that performs important roles in various electronic devices, including displays [2], because of its low resistivity, appropriate chemical stability, and transparency toward visible light [3]. Most ITO thin films have been prepared by a dry process such as sputtering [4]. ITO thin films prepared by such processes exhibit low resistivity; however, the equipment required for dry processing is expensive and the process itself requires high temperatures. Wet processes such as coating methods using inks containing ITO nanoparticles (NPs) also have problems, including high resistivity of the resultant films because of remaining dispersant and organic solvent [5].

The affinity between the surface of NPs and their dispersion medium is known to be a key factor governing the stability of NP dispersions [6]. Therefore, in the present work, we controlled the surface shape of ITO NPs to increase their stability in an aqueous dispersion without a dispersant. Specifically, the formation of numerous protrusions on ITO NP surfaces increased their specific surface area and the amount of water molecules they adsorbed, resulting in highly hydrophilic ITO NPs.

The surface shape of NPs has been controlled using various methods. Controlling particle growth *via* adsorption

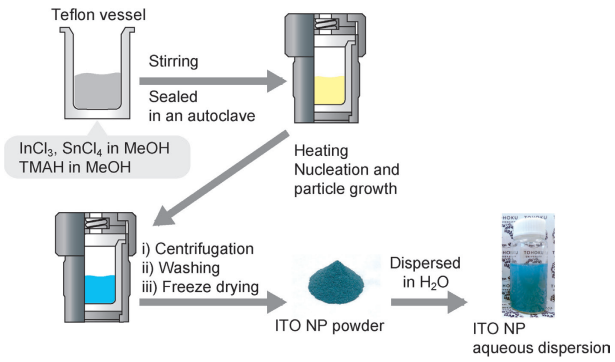
† Reprinted with permission from Suzuki,R.; Nishi, Y.; Matsubara, M.; Muramatsu, A.; Kanie, K. Single-Crystalline Protrusion-Rich Indium Tin Oxide Nanoparticles with Colloidal Stability in Water for Use in Sustainable Coatings *ACS Appl. Nano Mater.* **2020**, 3, 4870−4879. Copyright 2020 American Chemical Society.

additives [7] has been a focal point. In previous studies, amines and ions were found to be effective for controlling the shape of *a*-Fe_2O_3 NPs [8], [9]. Also, Kanie *et al.* developed a solvothermal one-pot method to prepare cubic ITO NPs, whose morphology reflects their crystal system, with a uniform crystal orientation [10], [11].

Inspired by these previous reports, we attempted to control the surface shape of ITO NPs by varying their synthesis conditions. We speculated that changing the nucleation and particle growth process by varying the concentration of the metal chloride used as a metal source could promote the heterogeneous nucleation of generated cubic ITO NPs, resulting in NPs with protrusions and a uniform crystal orientation. We also investigated the formation process and water dispersibility of the ITO NPs.

2 Experimental Procedure

The basic condition used to prepare the ITO NPs are described herein. First, a 1.8 M tetramethylammonium hydroxide (TMAH) methanol solution (7.5 mL) was added to a Teflon-made vessel. Then, a methanol solution of 0.36 M indium (III) chloride ($InCl_3$) and 0.04 M tin (IV) chloride ($SnCl_4$) (7.5 mL) was added, and the resultant mixture was stirred for 10 min at room temperature. The resultant solution was sealed in an autoclave and heated at 190°C for 24 h. The products were centrifuged and washed twice with ethanol and twice with water. The product (ITO NP powder) was then freeze-dried. In addition, an ITO NP aqueous dispersion was prepared by adding 40 mL of water to the ITO NP powder and dispersing the mixture by ultrasonication. (Scheme 1)



Scheme 1 Preparation of ITO NP powder and an ITO NP aqueous dispersion.

3 Results and Discussion

Effect of Heating Temperature

Fig. 1 shows XRD patterns of products prepared under the basic conditions at various temperatures from 100°C to 200°C. When the reaction temperature was 100°C (Fig. 1a), 130°C (Fig. 1b), and 150°C (Fig. 1c), no diffraction peaks assignable to In_2O_3 were observed. However, when the reaction temperature was 190°C (Fig. 1d) and 200°C (Fig. 1e), all of the observed peaks were assignable to In_2O_3. Thus, the XRD results indicate that a reaction temperature greater than 190°C was necessary to obtain a crystalline In_2O_3 phase.

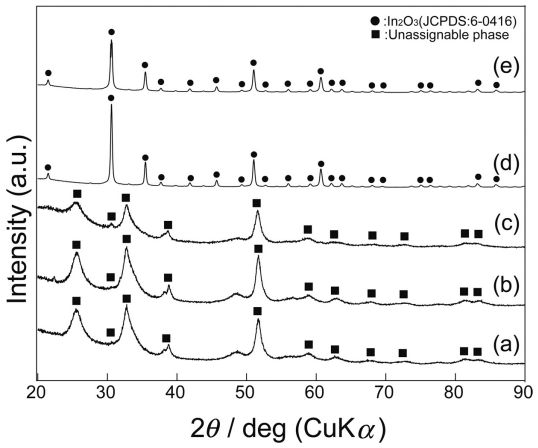


Fig. 1 XRD patterns of products obtained at various reaction temperatures: (a) 100°C, (b) 130°C, (c) 150°C, (d) 190°C, and (e) 200°C.

Effect of Metal Chloride Concentration

The concentration of the metal source was varied to investigate its effect on the surface shape of the ITO NPs. Under all of the investigated conditions, the Sn/In molar ratio for the sources was maintained at 11.1 mol%. Fig. 2a shows XRD patterns of the products obtained when the metal source concentration under the basic conditions was [$InCl_3$] = 0.18 M (i) and [$InCl_3$] = 0.36 M (ii). All of the peaks in both diffraction patterns were assignable to In_2O_3. Also, the Sn/In molar ratio, as determined from inductively coupled plasma (ICP) analysis results, was 10.3 mol% and 11.6 mol% for the products corresponding to [$InCl_3$] = 0.18 M and [$InCl_3$] = 0.36 M, respectively. These results show that ITO was obtained under both conditions. In both cases, the Sn/In ratio in the product matched that of the metal sources.

Fig. 2b and Fig. 2c show transmission electron microscopy (TEM) images of products prepared at [$InCl_3$] = 0.18 M and [$InCl_3$] = 0.36 M, respectively. Cubic-shaped ITO NPs were observed at [$InCl_3$] = 0.18 M, and ITO NPs with protrusions were observed at [$InCl_3$] = 0.36 M. The TEM images indicated that the sizes of the cubic NPs and the NPs with protrusions were 39 ± 12 nm and 38 ± 10 nm, respectively. These sizes are consistent with the crystallite size, 34 nm and 33 nm of the cubic NPs and the NPs with protrusions, respectively, calculated from the XRD results using

Scherrer's formula.

Fig. 2d shows a high-resolution TEM (HR-TEM) image and an electron diffraction (ED) pattern of one of the NPs shown in Fig. 2c. The NP shown in Fig. 2d appears to have a uniform crystal orientation, and a spot pattern assignable to the (200)

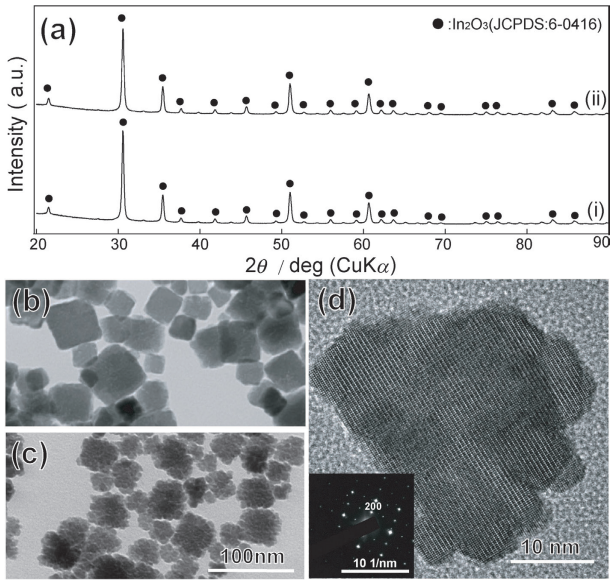


Fig. 2 (a) XRD patterns for products prepared under $[In^{3+}]$ = (i) 0.18 M and (ii) 0.36 M. TEM images of products prepared under $[In^{3+}]$ = (b) 0.18 M and (c) 0.36 M. (d) HR-TEM image of an ITO NP prepared under $[In^{3+}]$ = 0.36 M. The inset shows the ED pattern for the corresponding ITO NP.

crystal planes of In_2O_3 is observed. These results show that ITO NPs with protrusions that exhibit a uniform crystal orientation were obtained via control of the metal source concentration.

Effect of Heating Time on the Shape of ITO NPs

To investigate the formation process for the ITO NPs with protrusions, we varied the heating time. Fig. 3 shows TEM images of the products obtained as the heating time was varied from (a) 3 h to (b) 6 h , (c) 9 h , (d) 12 h, (e) 48 h, and (f) 168 h. For a heating time 3 h, NPs were not observed; an indeterminate form were obtained (Fig. 3a). In the XRD patterns, peaks assignable to In_2O_3 were not observed (results not shown). For a heating time 6 h, a few cubic particles were observed (Fig. 3b) and some of the XRD peaks were assignable to In_2O_3. When the heating time was 12 h or longer, all peaks in the XRD patterns were assignable to In_2O_3. In the TEM image for a heating time of 12 h (Fig. 3c), not only NPs with protrusions but also particleswith an indeterminate form were observed. When the heating time was 18 h and 48 h (Fig. 3d and 3e, respectively), only NPs with protrusions were observed. When the heating time was extended to 168 h (Fig. 3f), the protrusions on the NP

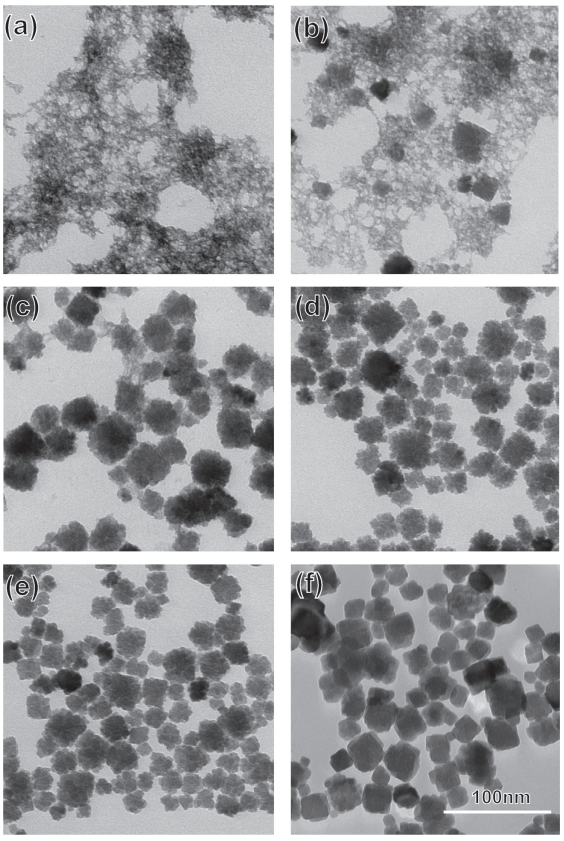


Fig. 3 TEM images of ITO NP products obtained at reaction times of (a) 3 h, (b) 6 h, (c) 9 h, (d) 12 h, (e) 48 h, and (f) 168 h.

surfaces disappeared. The Sn/In molar ratio for the products in Fig. 3a–3f are shown in Table 1. The Sn/In ratio decreased as the heating time was increased beyond 48 h.

Table 1 Sn/In molar ratio for ITO NPs prepared at various aging times

Aging time / h	Sn/In / mol%
3	12.7
6	12.7
9	12.0
12	12.1
48	9.3
168	8.4

Effect of Cl^- Concentration on Shape of ITO NPs

In a previous study [8], [9], the concentration of Cl^- was found to affect the shape of NPs. Therefore, the effect of Cl^- on the shape of the ITO NPs prepared in the present work was investigated. Tetramethylammonium chloride (TMACl) was added to the starting mixture used to prepare cubic ITO NPs (Fig. 2b). Fig. 4 shows TEM images of the obtained

products. In the TEM image corresponding to [TMACl] = 0.18 M (Fig. 4a), cubic NPs and NPs with slight protrusions on their surface were observed. When the concentration of TMACl was further increased in the range 0.27–0.54 M, the number of protrusions increased with increasing TMACl concentration (Fig. 4b–4d).

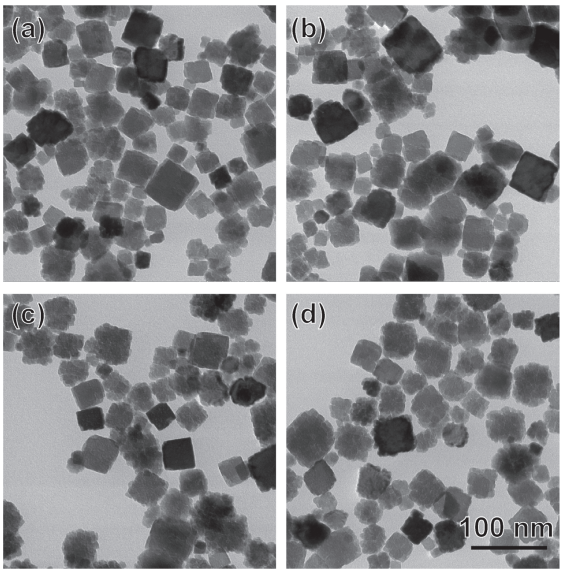


Fig. 4 TEM images of ITO NPs obtained when TMACl was added to the reaction mixture at a concentration of (a) 0.18 M, (b) 0.27 M, (c) 0.36 M, and (d) 0.54 M.

The results were used to evaluate the process by which protrusions form on ITO NPs (Scheme 2). Cubic ITO NPs were generated through homogeneous nucleation followed by particle growth. By contrast, ITO NPs with protrusions were generated through the growth of homogeneous particles, the adsorption of Cl^-, which inhibited particle growth, onto the cubic ITO NP surfaces, followed by heterogeneous and epitaxial particle growth. As a result, ITO NPs with protrusions and a uniform crystal orientation were obtained. The protrusions on the NP surfaces disappeared and the Sn/In molar ratio decreased with increasing heating time, indicating that Ostwald ripening occurred.

Dispersibility inWater

The water dispersibility of the ITO NPs with protrusions and that of the cubic ITO NPs were compared. Fig. 5a shows two ITO NP dispersions in the as-dispersed state. Both samples were homogeneously dispersed in water. Fig. 5b shows the two dispersions 2 weeks after they were prepared. The cubic ITO NPs were sedimented on the bottom of the vessel. By contrast, the ITO NPs with protrusions remained homogeneously dispersed in water, similar to the as-prepared dispersions shown in Fig. 5a. These results indicate that the ITO NPs with protrusions exhibited greater

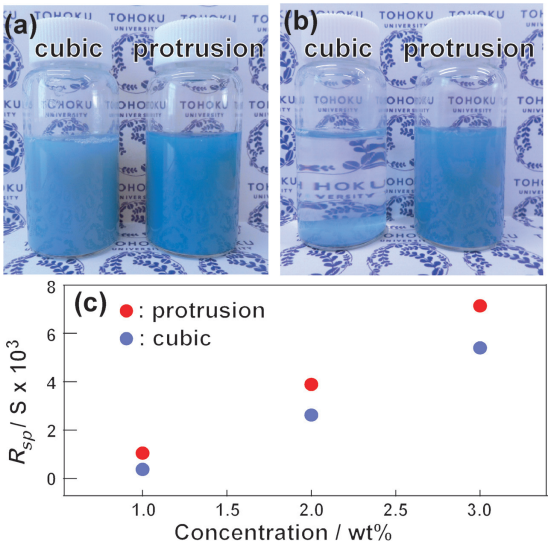
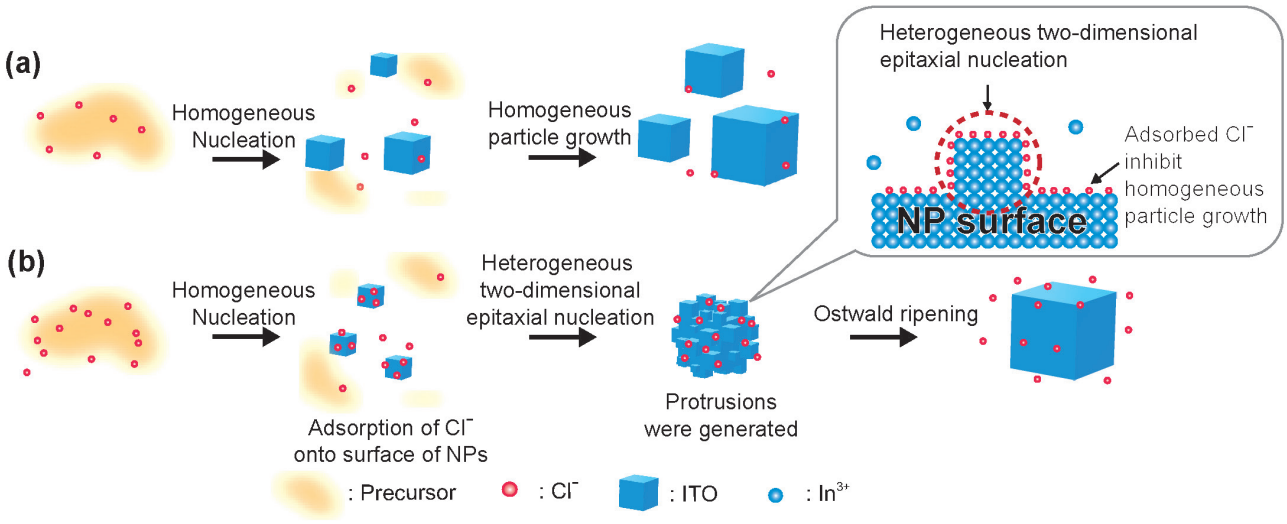


Fig. 5 Appearance of ITO NP dispersions (a) as dispersed and (b) after 2 weeks. (c) Relation between the concentration of the ITO NP dispersions and the *R*sp normalized specific surface area *S* of the ITO.



Scheme 2 Generation of (a) cubic ITO NPs and (b) ITO NPs with protrusions.

water dispersibility than the cubic ITO NPs.

Further analysis was conducted by measurement of the ITO NP relaxation times using ^{1}H NMR (Fig. 5c). The hydrophilicity parameter R_{sp} calculated from the results indicates the amount of water molecules adsorbed on the NP surfaces [12]. The relationship between the concentration of the ITO NP dispersions and the R_{sp} normalized specific surface area (S) of the ITO NPs contained in dispersions (R_{sp} / S) is shown in Fig. 5c. At each concentration, the ITO NPs with protrusions exhibited a larger R_{sp} / S value than the cubic ITO NPs. These results indicate that the ITO NPs with protrusions have more water molecules adsorbed on their surface than the cubic ITO NPs. Therefore, the cubic ITO NPs exhibited a more hydrophilic surface. The S of the ITO NPs with protrusions and that of the cubic ITO NPs, as determined by Brunauer–Emmett–Teller (BET) analysis of N_2 adsorption data, were 37.0 m^2/g and 26.3 m^2/g, respectively. The increase in specific surface area likely increased the hydrophilicity of the NPs.

4 Conclusion

ITO NPs with abundant protrusions on their surface and which could be dispersed in water for an extended period without a dispersant were successfully prepared *via* a solvothermal method. The ITO NPs with protrusions could be generated by heterogeneous particle growth induced by Cl^- adsorbed on cubic ITO NPs generated in the initial stage of the reaction. Therefore, the crystal orientation of ITO NPs with protrusions was uniform. Furthermore, the ITO NPs with protrusions were highly hydrophilic because of their shape and could form stable dispersions in water. The preparation of high-performance transparent conductive films using aqueous inks containing ITO NPs with protrusions without a high-temperature treatment may be feasible. We also expect our ITO NPs with protrusions will find applications in electrodes for flexible substrates.

References

[1] O. N. Mryasov and A. J. Freeman, "Electronic band structure of indium tin oxide and criteria for transparent conducting behavior," *Phys. Rev. B*, vol. 64, pp. 233111, 2001.

[2] Y. Gui, M. Miscuglio, Z. Ma, M. H. Tahersima, S. Sun, R. Amin, H. Dalir and V. J. Sorger, "Towards integrated metatronics: a holistic approach on precise optical and electrical properties of Indium Tin Oxide," *Sci. Rep*., vol 9, pp. 11279, 2019.

[3] M. Henry, P. M. Harrison and J. Wendland, "Laser direct write of active thin-films on glass for industrial flat panel display manufacture," *J. Laser Micro Nanoeng*., vol. 2, pp. 49–56, 2007.

[4] O. Tuna, Y. Selamet, G. Aygun and L. Ozyuzer, "High quality ITO thin films grown by dc and RF sputtering without oxygen," *J. Phys. D: Appl. Phys*., vol. 43, no. 5, pp. 055402, 2010.

[5] S. J. Hong, Y. H. Kim and J. I. Han, "Development of Ultrafine Indium Tin Oxide (ITO) Nanoparticle for Ink-Jet Printing by Low-Temperature Synthetic Method," *IEEE Trans. Nanotechnol*., vol. 7, pp. 172–176, 2008.

[6] S. Kangoa, S. Kalia, A. Celli, J. Njuguna, Y. Habibi and R. Kumar, "Surface modification of inorganic nanoparticles for development of organic-inorganic nanocomposites-A review," *Prog. Polym. Sci*., vol. 38, no. 8, pp. 1232–1261, 2013.

[7] Z. Wu, S. Yang and W. Wu, "Shape Control of Inorganic Nanoparticles from Solution," *Nanoscale*, vol. 8, pp. 1237–1259, 2016.

[8] T. Sugimoto, Y. Wang, H. Itoh and A. Muramatsu, "Systematic Control of Size, Shape and Internal Structure of Monodisperse α-Fe_2O_3 Particles," *Colloids Surf., A*, vol. 134, pp. 265–279, 1998.

[9] T. Sugimoto, M. M. Khan and A. Muramatsu, "Preparation of Monodisperse Peanut-type α-Fe_2O_3 Particles from Condensed Ferric Hydroxide Gel," *Colloids Surf., A*, vol. 70, pp. 167–169, 1993.

[10] (a) K. Kanie, T. Sasaki, M. Nakaya and A. Muramatsu, "Quaternary Ammonium Hydroxide-Assisted Solvothermal Synthesis of Monodispersed ITO Nanoparticles with a Cubic Shape," *Chem. Lett*., vol. 42, pp. 738–740, 2013.

[11] T. Sasaki, Y. Endo, M. Nakaya, K. Kanie, A. Nagatomi, K. Tanoue, R. Nakamura and A. Muramatsu, "One-step Solvothermal Synthesis of Cubic-Shaped ITO Nanoparticles Precisely Controlled in Size and Shape and their Electrical Resistivity," *J. Mater. Chem*., vol. 20, pp. 8153–8157, 2010.

[12] D. Fairhurst, T. Cosgrove and S. W. Prescott, "Relaxation NMR as a Tool to Study the Dispersion and Formulation Behavior of Nanostructured Carbon Materials," *Magn. Reson. Chem*., Vol. 54, pp. 521–526, 2016.

鈴木涼子　Ryoko SUZUKI
研究開発本部　材料・要素技術研究所
Materials & Advanced Research Laboratory
Research & Development Division

西　康孝　Yasutaka NISHI
FPD 装置事業部 開発統括部 先端技術開発部
Advanced Technology Development Department
Development Sector
FPD Lithography Business Unit

松原正樹　Masaki MATSUBARA
東北大学
Tohoku University

村松淳司　Atsushi MURAMATSU
東北大学
Tohoku University

蟹江澄志　Kiyoshi KANIE
東北大学
Tohoku University


鈴木涼子
Ryoko SUZUKI


西　康孝
Yasutaka NISHI


松原正樹
Masaki MATSUBARA


村松淳司
Atsushi MURAMATSU


蟹江澄志
Kiyoshi KANIE

曖昧な訓練データを用いた二値分類の適用

大谷直也，大坪洋介，小池哲也，杉山　将

An Application of Binary Classification using Ambiguous Training Data

Naoya OTANI, Yosuke OTSUBO, Tetsuya KOIKE and Masashi SUGIYAMA

教師あり学習において，しばしば専門家にとってもラベル付けが難しいデータ（曖昧なデータ）が存在する．本稿では，曖昧なデータが存在する状況下での二値分類問題を検討し，社内で取得した細胞培養データに適用する．曖昧なデータはラベル付けが難しいという情報を持っているため，半教師あり学習におけるラベルなしデータとは異なる扱いが必要である．また，曖昧なデータは訓練データに存在するものの，テスト時は正負の二値に分類するため，正負と曖昧なクラスによる三値分類問題とも異なる．我々の提案手法は，リジェクト付き分類を拡張する形で定式化した．具体的には，リジェクト付き分類は，リジェクトコスト c を有する0-1-c 損失に基づいて，分類器とリジェクタを同時に学習する方法であるが，我々は0-1-c-d 損失として曖昧なデータに対する誤分類ペナルティ d を導入し，分類器とリジェクタを同時に学習する方法を提案した．計算の容易性の観点から，我々は0-1-c-d 損失の凸の上界となる代理損失を用いて実装を行った．細胞培養データに対する数値実験を通じて，曖昧なデータから得られる情報を，二値分類問題に有効に活用できることを示した．

In supervised learning, ambiguous (A) samples that are difficult to label even by domain experts are often encountered. In this study, we consider a binary classification problem using such A samples and apply our in-house datasets of a cell culture process. This problem is substantially different from semi-supervised learning because unlabeled samples are not necessarily difficult samples. Furthermore, it is different from the three-class classification involving positive (P), negative (N), and A classes because the test samples are not to be classified as the A class. Our proposed method extends binary classification with a reject option, which trains a classifier and a rejector simultaneously using P and N samples based on the 0–1-c loss with a rejection cost, c. More specifically, we propose to train a classifier and a rejector based on the 0–1-c-d loss using P, N, and A samples, where d is the misclassification penalty for A samples. In our practical implementation, we use the convex upper bound of the 0–1-c-d loss to achieve computational tractability. Numerical experiments using the in-house datasets demonstrate that our method can successfully utilize the additional information resulting from such A training data.

Key words 曖昧なサンプル，リジェクト付き分類，二値分類
ambiguous samples, classification with reject option, binary classification

1 Introduction

Supervised learning has been successfully deployed in various real-world applications, such as medical diagnosis [1] and manufacturing systems [2]. However, when the amount of labeled data is limited, current supervised learning methods become unreliable [3].

To efficiently obtain labeled data, domain knowledge has been used in many applications [2], [4]. However, as indicated in some studies [5], [6], ambiguous (A) samples that are substantially difficult to label even by domain experts are often encountered.

The goal of this study is to propose a novel classification method that can manage A samples. Specifically, we consider a binary classification problem where, in addition to positive (P) and negative (N) samples, A samples are available for training a classifier. Because of the characteristics of A samples, they are assumed to be located near the boundary between P and N classes.

We may consider employing three-class classification methods for the P, N, and A classes. However, because we intend to classify test samples only in the P or N class, not in the A class, naive three-class methods cannot be directly used in our problem. Moreover, they cannot utilize the information that the A class exists between the P and N classes. Another related approach is classification with a reject option [7], [8], where A test samples are not classified into P or N classes but as rejected (R). However, classification methods

with a reject option do not consider A samples in the training phase; hence, they cannot be employed in our problem.

Semi-supervised learning may be related to the current problem, where unlabeled (U) data, in addition to P and N data, are used to train a classifier [9]. In semi-supervised learning, U samples are P and N samples that have not yet been labeled, and they are not necessarily difficult samples to be labeled. By contrast, A samples in our target problem are typically distributed at the intersection of P and N classes. Thus, as the problem setups are intrinsically different, merely using semi-supervised learning methods in the current problem may not be optimal. Our problem and related methods are summarized in Table 1.

Table 1 Problem settings of related and our methods.

Methods	Labels in training data	Labels predicted in test phase	Relationship among classes
Binary classification	P / N	P / N	None
Three-class classification	Class 1 Class 2 Class 3	Class 1 Class 2 Class 3	None
Classification with reject option	P / N	P / R / N	R samples are in P/N mixed regions
Semi-supervised learning	P / U / N	P / N	U samples belong to P or N
Our proposal	P / A / N	P / N	A samples are in P/N mixed regions

To effectively solve the classification problem involving A data, we propose to extend classification with a reject option that trains a classifier and a rejector simultaneously using P and N samples based on the 0-1-*c* loss with a rejection cost, *c* [8]. The proposed method trains a classifier and a rejector based on the 0-1-*c*-*d* loss using P, N, and A samples, where *d* is the misclassification penalty for A samples. Then, in the test phase, we use the trained classifier to assign P or N labels to the test samples. Through experiments using an in-house cell culture dataset, we demonstrate that the proposed method can improve the test classification accuracy by using A samples in the training phase.

2 Formulation

In this section, we formulate our target problem, named classification with ambiguous data (CAD), and propose a new method for solving CAD.

2.1. Preliminary

We consider three class labels, namely, P, A, and N: $y \in \mathcal{Y}_0 = \{1, 0, -1\}$. We assume that we are assigned a set of P, A, and N samples $\{(x_i, y_i)\}_{i=1}^{N}$ drawn independently from a probability distribution with density $p_0(x, y)$ defined on $\mathcal{X} \times \mathcal{Y}_0$. Let $h : \mathcal{X} \to \mathbb{R}$ denote a discriminant function, with which a class label is predicted to be P or N (not predicted to be A) for a test input point, *x*, as $\hat{y} = \mathrm{sign}(h(x))$. Our goal is to learn a discriminant function that accurately classifies the test samples (not in the A class). Our key question in this scenario is whether we can utilize A training data to improve the classification accuracy of the discriminant function.

Hence, we develop a new method based on classification with a reject option (CRO) [8]. We first review the CRO method before deriving the new method.

2.2. Classification with Reject Option using Support Vector Machine (CRO-SVM)

Cortes *et. al.* [8] introduced a rejection function, $r : \mathcal{X} \to \mathbb{R}$, in addition to the discriminant function, to identify regions with a high risk for misclassification. When the rejection function yields a positive value, the corresponding sample is classified into the P or N class by using classifier *h*; otherwise, the sample is rejected and not classified. When a sample is rejected, a rejection cost, *c*, is incurred, which trades off the risk of misclassification. To realize this idea, the *0-1-c loss* is introduced:

$$L_{01c}(h, r, x, y) = 1_{yh(x) \le 0} 1_{r(x) > 0} + c 1_{r(x) \le 0}, \tag{1}$$

where 1_A is the indicator function that yields 1 if statement *A* is true and 0 otherwise. When $c = 0$, all samples are rejected because the loss function does not incur any cost. By contrast, when $c \ge 0.5$, no samples are rejected because the expectation of the 0-1 loss, $1_{yh(x) \le 0}$, is less than 0.5; thus, the 0-1-*c* loss is reduced to the 0-1 loss. Therefore, we only consider *c* such that $0 < c < 0.5$.

Based on the 0-1-*c* loss, the problem is expressed as

$$\begin{aligned} (h^*, r^*) &= \operatorname*{argmin}_{(h,r)} R(h, r), \\ R(h, r) &= \mathbb{E}_{p_0(x,y)} \left[L_{01c}(h, r, x, y) \right], \end{aligned} \tag{2}$$

where h^* and r^* denote the optimal discriminant function and rejection function, respectively, and $\mathbb{E}_{p_0(x,y)}$ denotes the expectation over $p_0(x, y)$. In practice, because the true density, $p_0(x, y)$, is unknown, we typically use the empirical dis-

tribution to approximate the expectation:

$$\widehat{R}(h)=\frac{1}{N}\sum_{i=1}^{N}L_{01c}(h,r,x_i,y_i). \tag{3}$$

Because of the discrete nature of the 0–1-*c* loss, its direct optimization is computationally intractable. To avoid discontinuity, the following surrogate loss, known as the max-hinge (MH) loss, is introduced:

$$L_{\mathrm{MH}}(h,r,x,y)= \max\left(1+\frac{\alpha}{2}\big(r(x)-yh(x)\big),c\big(1-\beta r(x)\big),0\right), \tag{4}$$

where $\alpha, \beta > 0$ are the hyperparameters used to control the shape of the surrogate loss. The surrogate loss is an extension of the hinge loss, which is employed in a support vector machine (SVM) [10].

Further, introducing L2 regularization, basis functions $\phi_1(x),\ldots,\phi_N(x)$, and slack variables $\xi=(\xi_1,\ldots,\xi_N)^\top$ with $^\top$ being the transpose yields the following quadratic program:

$$\left(\hat{w},\hat{u},\hat{\xi}\right)=\underset{(w,u,\xi)}{\operatorname{argmin}}\left[\frac{\lambda}{2}\|w\|^2+\frac{\lambda'}{2}\|u\|^2+\frac{1}{N}\sum_{i=1}^{N}\xi_i\right] \quad \text{s.t.}\begin{pmatrix}\xi_i\geq 1+\frac{\alpha}{2}\big(r_i-y_ih_i\big)\\ \xi_i\geq c\big(1-\beta r_i\big)\\ \xi_i\geq 0\end{pmatrix}\text{for } i=1,\ldots,N, \tag{5}$$

where $w=(w_1,\ldots,w_N)^\top$ are the coefficients of the discriminant function; $u=(u_1,\ldots,u_N)^\top$ are the coefficients of the rejection function; $\lambda,\lambda'>0$ are the L2 regularization parameters; h_i and r_i denote the values of the discriminant function and rejection function at sample point x_i expressed as $h_i=\sum_{j=1}^{N}w_j\phi_j(x_i)$ and $r_i=\sum_{j=1}^{N}u_j\phi_j(x_i)$, respectively. The resulting discriminant and rejection functions are expressed as $h(x;\hat{w})=\sum_{j=1}^{N}\hat{w}_j\phi_j(x)$ and $r(x;\hat{u})=\sum_{j=1}^{N}\hat{u}_j\phi_j(x)$, respectively.

We refer to this method as CRO-SVM.

2.3. Proposed Method: Classification with A Data using SVM (CAD-SVM)

To manage A training data in the SVM formulation, we extend the 0–1-*c* loss to the *0–1-c-d loss*, as Eq. (6):

Table 2 0–1-*c* loss function.

Judgment (*h*, *r*) / Label *y*	P *h* > 0 *r* > 0	R *r* ≤ 0	N *h* ≤ 0 *r* > 0
P: *y* = 1	0	*c*	1
N: *y* = −1	1	*c*	0

Table 3 0–1-*c*-*d* loss function.

Judgment (*h*, *r*) / Label *y*	P *h* > 0 *r* > 0	R *r* ≤ 0	N *h* ≤ 0 *r* > 0
P: *y* = 1	0	*c*	1
A: *y* = 0	*d*	0	*d*
N: *y* = −1	1	*c*	0

$$L_{01\mathrm{cd}}(h,r,x,y)= 1_{y^2=1}\left(1_{yh(x)\leq 0}1_{r(x)>0}+c1_{r(x)\leq 0}\right)+d1_{y=0}1_{r(x)>0}. \tag{6}$$

Tables 2 and 3 present comparisons of the behaviors of the 0–1-*c* and 0–1-*c*-*d* losses, respectively. For the P and N samples, the 0–1-*c*-*d* loss behaves the same as the 0–1-*c* loss. In contrast, for the A samples, the 0–1-*c*-*d* loss incurs penalty *d* when they are classified as the P or N class. Therefore, A samples tend to be classified into the A class if we employ the 0–1-*c*-*d* loss. Unlike the CRO formulation, CAD utilizes A samples to learn a rejector explicitly.

This discussion may mislead us as if we are just solving a three-class problem involving P, N, and A classes. However, we do not classify the test samples into the A class, but only into the P and N classes. To solve the CAD problem, we utilize a binary discriminant function, *h*, and a rejection function, *r*, as in the CRO formulation reviewed earlier. We train *h* and *r* based on the 0–1-*c*-*d* loss, and we use only *h* in the test phase to classify the test samples into P and N classes. Owing to the interplay between *h* and *r* in the 0–1-*c*-*d* loss, we can utilize A samples to train *h* through *r*.

Similar to the 0–1-*c* loss, we consider the following convex upper bound of the 0–1-*c*-*d* loss, named max-hinge-ambiguous (MHA) loss, as a surrogate to avoid its discrete nature:

$$\begin{aligned}L_{01\mathrm{cd}}(h,r,x,y)&\leq 1_{y^2=1}L_{\mathrm{MH}}(h,r,x,y)+d1_{y=0}\max\big(1+\beta r(x),0\big)\\ &=y^2\max\left(1+\frac{\alpha}{2}\big(r(x)-yh(x)\big),c\big(1-\beta r(x)\big),0\right)+\big(1-y^2\big)\max\big(d\big(1+\beta r(x)\big),0\big)\\ &\leq y^2\max\left(1+\frac{\alpha}{2}\big(r(x)-yh(x)\big),c\big(1-\beta r(x)\big),0\right)+\big(1-y^2\big)\max\big(\eta d\big(1+\beta r(x)\big),0\big)\\ &\equiv L_{\mathrm{MHA}}(h,r,x,y),\end{aligned} \tag{7}$$

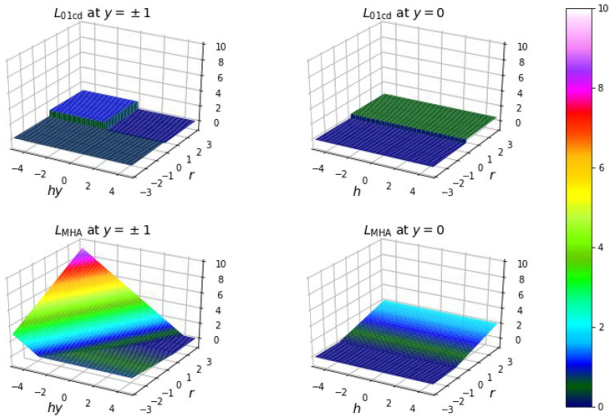


Fig. 1 0–1-*c*-*d* loss, L_{01cd}, and its surrogate loss, L_{MHA}.

where $\eta \geq 1$ is a hyperparameter that controls the shape of the surrogate loss (see Fig. 1 for the visualization).

Next, similar to CRO-SVM, we have the following quadratic program:

$$\left(\hat{w},\hat{u},\hat{\xi}\right) = \underset{(w,u,\xi)}{\operatorname{argmin}} \left[\frac{\lambda}{2}\|w\|^2 + \frac{\lambda'}{2}\|u\|^2 + \frac{1}{N}\sum_{i=1}^{N}\xi_i\right]$$

$$\text{s.t.} \begin{pmatrix} \xi_i \geq y_i^2\left(1+\frac{\alpha}{2}\left(r_i - y_i h_i\right)\right) \\ \xi_i \geq y_i^2 \eta c\left(1-\beta r_i\right) \\ \xi_i \geq \left(1-y_i^2\right)\eta d\left(1+\beta r_i\right) \end{pmatrix} \tag{8}$$

$$\text{for } i = 1,\ldots,N.$$

This formula expresses our proposed method, CAD-SVM.

To select hyperparameters (α,β,η), we can apply the following theorem (its proof is available in [11]):

Theorem 1 *For each* $x \in \mathcal{X}$, *let*

$$\left(h_{01cd}^*, r_{01cd}^*\right) = \underset{(h,r)}{\operatorname{argmin}} \mathbb{E}_{p_0(y|x)}\left[L_{01cd}\left(h,r,x,y\right)\right], \tag{9}$$

and

$$\left(h_{MHA}^*, r_{MHA}^*\right) = \underset{(h,r)}{\operatorname{argmin}} \mathbb{E}_{p_0(y|x)}\left[L_{MHA}\left(h,r,x,y\right)\right]. \tag{10}$$

Then, for

$$\alpha^* = 2\left(1-2c\right), \quad \beta^* = 1+2c, \quad \eta^* = \frac{2}{1+2c}, \tag{11}$$

the signs of $\left(h_{MHA}^*, r_{MHA}^*\right)$ *match those of* $\left(h_{01cd}^*, r_{01cd}^*\right)$.

In the next section, we demonstrate that this method is feasible. It is noteworthy that Eq. (11) does not include *d*.

3 Numerical Experiments

In this section, we report the experimental results obtained using an in-house dataset from a cell culture process. A detailed performance evaluation and a comparison with baseline methods on other datasets have been reported in [11].

3.1. Dataset

For real-world applications, we prepared an in-house cell-culture dataset. This dataset contains 124 fields of view (FOVs). For each FOV, images were acquired three times at T = 99, 279, and 459 h. All images for each FOV were analyzed using an image processing software, CL-Quant [12], and converted to eight morphological features, such as the average brightness and average area of cells. Based on the final image for each FOV (T = 459 h), each FOV was annotated by experts. If the cells in the image appeared healthy/damaged, the image was labeled as P/N. Otherwise, the experts assign A labels to samples that cannot be confidently classified as healthy or damaged. The numbers of samples for the P, N, and A classes were 41, 59, and 24, respectively. Our goal was to predict the final state of each FOV (annotated by the experts in this simulation) using morphological features obtained from each time point of the culturing process. In total, we trained and evaluated three types of datasets (Datasets 1, 2, and 3), corresponding to the time point of the input images, T. For Datasets 1 and 2, the images from which we extracted the input features and those from which we annotated the output labels were different; this is illustrated in Fig. 2.

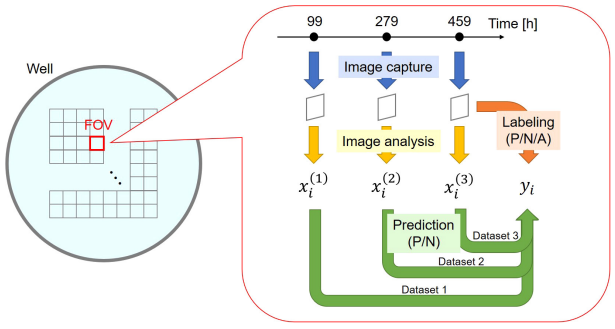


Fig. 2 Schematic image of datasets. We created three types of datasets and evaluated corresponding models.

3.2. Experimental Settings

Using the aforementioned datasets, we compared the classification performance of the SVM, SVM-RL (random label), LapSVM [13], two-step SVM, CRO-SVM, CRO-SVM-RL, and CAD-SVM.

For each method, 1500 test runs were performed by changing the training and test datasets, which were randomly selected from the original dataset. The ratio of the

training and test datasets was 4:1. For each test run, five-fold cross-validation was performed to determine the relevant parameters. For validation and in the test phase, only P and N samples were applied to the discriminant function; hence, we were able to evaluate the binary classification accuracy.

We determined 10 hyperparameters $(\lambda,\lambda',\sigma,\sigma',\tau,c,d,\alpha,\beta,\eta)$, where σ is the width of the Gaussian radial basis function in the basis function $\phi_i(x)=\exp\left(-\frac{\|x-x_i\|^2}{2\sigma^2}\right)$, σ' is the hyperparameter of the weight matrix, W, of the graph Laplacian expressed as $W_{ij}=\exp\left(-\frac{\|x_i-x_j\|^2}{2\sigma'^2}\right)$ (only in the LapSVM), and τ is the coefficient of the graph Laplacian regularization (only in the LapSVM). The hyperparameters (α,β,η) were determined by using Eq. (11), and other hyperparameters were selected via five-fold cross-validation (see [11] for details). The experimental procedure applied for each dataset and method is summarized in Algorithm 1.

Input: $\mathcal{D}=\{(x_i,y_i)\}_{i=1}^{N}$, $\mathcal{H}=\{(\lambda_\xi,\lambda'_\xi,\sigma_\xi,\sigma'_\xi,\tau_\xi,c_\xi,d_\xi)\}_{\xi=1}^{\Xi}$
For $j=1,\ldots,1500$ **do**:
  $\mathcal{D}_{\text{train}} \leftarrow$ (random set of $0.8N$ records in $\mathcal{D}$)
  $\mathcal{D}_{\text{test}} \leftarrow \mathcal{D} \setminus \mathcal{D}_{\text{train}}$
  $\left(\mathcal{D}_{\text{train}}^{(1)},\ldots,\mathcal{D}_{\text{train}}^{(5)}\right) \leftarrow$ (random set that equally divide $\mathcal{D}$ into 5)
  For $k=1,\ldots,5$ **do**:
    $\mathcal{D}_{\text{valid}} \leftarrow \mathcal{D}_{\text{train}}^{(k)}$
    $\mathcal{D}_{\text{subtrain}} \leftarrow \mathcal{D}_{\text{train}} \setminus \mathcal{D}_{\text{valid}}$
    For $\xi=1,\ldots,\Xi$ **do**:
      Create a model using $\mathcal{D}_{\text{subtrain}}$ under hyperparameter setting $\mathcal{H}_\xi$
      $a_\xi^{(k)} \leftarrow$ (binary classification accuracy of $\mathcal{D}_{\text{valid}}$)
  $\overline{a_\xi} \leftarrow \frac{1}{5}\sum_{k=1}^{5} a_\xi^{(k)}$
  $\xi^* \leftarrow \operatorname{argmax}_\xi \overline{a_\xi}$
  Create a model using $\mathcal{D}_{\text{train}}$ under hyperparameter setting $\mathcal{H}_{\xi^*}$
  $a_j \leftarrow$ (binary classification accuracy of $\mathcal{D}_{\text{test}}$)
Calculate the mean and standard deviation of a_j

Algorithm 1 Experimental procedure for each dataset and method.

3.3. Results

Table 4 summarizes the test accuracy of each method. The CAD-SVM showed statistically significant improvements over the other methods, particularly in the earlier stages of the culturing process. In the earlier stages, the input data contained few or inaccurate information; therefore, utilizing A samples would be beneficial. However, because the input data contained almost complete information during the final state, the information of A samples need not be utilized. If the information of A samples is intrinsically meaningless, then the SVM would be a better solution as it utilizes the hinge loss directly based on the 0–1 loss (*i.e.*, the binary classification accuracy). Overall, the CAD-SVM is a promising method for utilizing A samples.

Table 4 Test accuracy for each timepoint, where ± denotes standard deviation. Boldfaced numbers represent the best and comparable results with 5% t-test.

	Dataset 1 (T = 99)	Dataset 2 (T = 279)	Dataset 3 (T = 459)
SVM	0.732 ± 0.092	0.799 ± 0.088	**0.941 ± 0.049**
SVM-RL	0.730 ± 0.096	0.805 ± 0.088	0.929 ± 0.058
LapSVM	0.731 ± 0.091	0.801 ± 0.089	0.931 ± 0.055
Two-step SVM	0.733 ± 0.097	0.788 ± 0.090	0.931 ± 0.054
CRO-SVM	0.747 ± 0.095	**0.814 ± 0.087**	**0.939 ± 0.050**
CRO-SVM-RL	0.740 ± 0.097	**0.818 ± 0.085**	0.920 ± 0.063
CAD-SVM	**0.755 ± 0.094**	**0.819 ± 0.087**	0.937 ± 0.051

4 Conclusion

In this study, we aimed to reduce labeling cost and improve classification accuracy by allowing labelers to provide A labels for difficult samples. We extended a classification method with a reject option and proposed a novel classification method, named CAD-SVM, which uses the 0–1-*c*-*d* loss. We derived a surrogate loss for the 0–1-*c*-*d* loss, thereby allowing us to convert the optimization problem into a convex quadratic program. We conducted numerical experiments and demonstrated that A labels can be effectively used to improve the classification accuracy.

Although our proposed method was based on the SVM, it would be more useful if it is applicable to other models, particularly to deep neural networks. In future studies, we will conduct a theoretical analysis of the proposed method in terms of the statistical consistency and convergence rate. Extending the proposed loss function to semi-supervised, imperfect labeling, or multiclass problems is also a promising direction for future research.

References

[1] A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau and S. Thrun, "Dermatologist-level classification of skin cancer with deep neural networks," *Nature,* vol. 542, no. 7639, pp. 115–118, 2017.

[2] R. Ren, T. Hung and K. C. Tan, "A generic deep-learning-based approach for automated surface inspection," *IEEE transactions on cybernetics,* vol. 48, no. 3, pp. 929–940, 2017.

[3] F. Pesapane, M. Codari and F. Sardanelli, "Artificial intelligence in medical imaging: threat or opportunity? Radiologists again at the forefront of innovation in medicine," *European radiology experimental,* vol. 2, no. 1, p. 35, 2018.

[4] K. Konishi, M. Mimura, T. Nonaka, I. Sase, H. Nishioka and M. Suga, "Practical method of cell segmentation in electron microscope image stack using deep convolutional

neural network," *Microscopy*, vol. 68, no. 4, pp. 338–341, 2019.

[5] Y. Li, B. Wu, B. Ghanem, Y. Zhao, H. Yao and Q. Ji, "Facial action unit recognition under incomplete data based on multi-label learning with missing labels," *Pattern Recognition,* vol. 60, pp. 890–900, 2016.

[6] S. A. Shahriyar, K. M. R. Alam, S. S. Roy and Y. Morimoto, "An approach for multi label image classification using single label convolutional neural network," in *2018 21st international conference of computer and information technology (ICCIT)*, 2018.

[7] P. L. Bartlett and M. H. Wegkamp, "Classification with a reject option using a hinge loss," *Journal of Machine Learning Research*, vol. 9, no. Aug, pp. 1823–1840, 2008.

[8] C. Cortes, G. DeSalvo and M. Mohri, "Learning with rejection," in *International Conference on Algorithmic Learning Theory*, 2016.

[9] T. Sakai, M. C. du Plessis, G. Niu and M. Sugiyama, "Semi-supervised classification based on classification from positive and unlabeled data," in *Proceedings of the 34th International Conference on Machine Learning*, 2017.

[10] V. N. Vapnik, *The Nature of Statistical Learning Theory*, Berlin, Germany: Springer-Verlag, 1995.

[11] N. Otani, Y. Otsubo, T. Koike and M. Sugiyama, "Binary classification with ambiguous training data," *Machine Learning*, vol. 109, pp. 2369–2388, 2020.

[12] S. V. Alworth, H. Watanabe and J. S. J. Lee, "Teachable, high-content analytics for live-cell, phase contrast movies," *Journal of biomolecular screening*, vol. 15, no. 8, pp. 968–977, 2010.

[13] M. Belkin, P. Niyogi and V. Sindhwani, "Manifold regularization: A geometric framework for learning from labeled and unlabeled examples," *Journal of machine learning research*, vol. 7, no. Nov, pp. 2399–2434, 2006.

大谷直也　Naoya OTANI
研究開発本部　数理技術研究所
Mathematical Sciences Research Laboratory
Research & Development Division

大坪洋介　Yosuke OTSUBO
研究開発本部　数理技術研究所
Mathematical Sciences Research Laboratory
Research & Development Division

小池哲也　Tetsuya KOIKE
研究開発本部　数理技術研究所
Mathematical Sciences Research Laboratory
Research & Development Division

杉山　将　Masashi SUGIYAMA
理化学研究所
RIKEN
東京大学
The University of Tokyo


大谷直也
Naoya OTANI


大坪洋介
Yosuke OTSUBO


小池哲也
Tetsuya KOIKE


杉山　将
Masashi SUGIYAMA

生産工程における不良要因検知とエミュレーターの推定†

大坪洋介，大谷直也，近末恵美，杉山　将

Defect Factor Detection in Production Processes and Emulator Estimation

Yosuke OTSUBO, Naoya OTANI, Megumi CHIKASUE and Masashi SUGIYAMA

精密機器は大量の部品や部組から構成されている．実際の量産工程においては多数の製品を管理しているため，検査工程に対応する下流のデータは取得できるが，上流にある部品パラメーターを取得することはできない．したがって，工程中に仕様を満たさない不良品が発生した場合，人的に不良要因を調査し特定する必要がある．本研究では，近似ベイズ推定に基づいた生産工程における不良要因検知手法を提案する．この手法は，次の2ステップで構成される．1）生産工程シミュレーターを活用し，データから逆問題を解く，2）不良が発生したデータと正常データを用いて，各部品パラメーターに対する不良要因スコアを計算する．実際のシミュレーターを用いて検証し，合理的な結果が得られることを示す．さらに，スパースモデリングを組み合わせることで，シミュレーターを模倣する疑似的な数理モデルを構成できることを示す．

Precision mechanical products consist of lots of parts and assemblies. Data accumulation systems are usually installed to monitor production processes, but only downstream data can be acquired with respect to the processes of precision equipment. So, when a defect occurs, mechanical experts must manually analyze a physical product to identify the cause (factor). In this study, we propose a novel detection method for the defect factor based on the approximate Bayesian computation (ABC) with a design simulator. Our method consists of two stages: 1) solve the inverse problem from the data; 2) derive defect factor scores from two datasets corresponding to normal data and data with defects. Numerical experiments and an application to actual data yield consistent results with design information given as domain knowledge. Furthermore, it is shown that a surrogate model, i.e., an emulator, that imitates a simulation can be estimated by combining a sparse modeling.

Key words 生産工程，不良要因検知，エミュレーター推定，近似ベイズ計算，スパースモデリング
production processes, defect factor detection, emulator, approximate Bayesian computation, sparse modeling

1 はじめに

近年，生産工程に設置したセンサーをネットワークでつないで構築したフィジカル空間のデータを，データ分析技術を使ってサイバー空間で分析・モニタリングし，分析結果をフィードバックして生産工程を高度化するというサイバーフィジカルシステムが注目されている［1］．精密機器の生産工程においても，少量多品種生産が主流となり開発期間が短縮されたため，工程や測定機などのデータを収集・分析し，生産工程の自律化や不良要因特定などが望まれている．特に，加工装置データや中間の製品出来栄えデータなどを使い，測定されていない部品寸法，装置の設定などを予測する技術は仮想計測（Virtual Metrology）と呼ばれ，主に半導体製造工程で活用されている［2］，［3］．仮想計測は量産工程において不良要因等を特定する際に有効であるが，予測したいデータの項目数に対し測定データの項目数が少ない場合には，そのまま適用することができない．

精密機器の生産工程は，大量の部品を部組から総組状態まで組立てる組立工程と，部品・部組の位置や姿勢などを調整する調整工程，製品が最終性能を満たすことを検査する検査工程など多数の工程から構成され非常に複雑である．部品寸法は最終性能に大きな影響を与えるが，厳しい公差を設定しすぎると低コストでの量産が難しくなる．そこで，光学製品の設計では機械部品と光学部品に適切に公差を配分できるよう，設計段階で量産時の光学性能・量産性・公差感度を評価可能な生産シミュレーター［4］などが利用される．量産中も部品寸法をモニタリングできれば不良要因

† 本稿は，引用文献［11］に対してエミュレーター推定の考察を加えたものである．

特定に活用できるが，部品数が多いためすべての部品を測定することはコストの面で現実的ではなく，抜取検査などによって品質を確認している．

本研究の第一の目的は，生産工程から取得される限られたデータから，不良要因となる部品パラメーターを検知することである．ここでの不良とは，量産された製品群の性能が設計仕様を満たさないことを指す．具体的には，設計仕様を満たす正常のデータセットと不良品を含むデータセットが与えられた際に，両者の差分から不良要因を検知することを目指す．一般に，取得できるデータは検査工程を経た総組の性能データであり，大部分の部品パラメーターのデータは取得できないため，データ駆動型の解析だけでは不良要因は検知できない．そこで，部品パラメーターから最終性能が出力される生産シミュレーターと生産工程データを組み合せることで不良要因検知を実現する．また，本研究の第二の目的は，エミュレーター（疑似モデル）を構成することである．エミュレーターとは，一般に複雑なシミュレーターを模倣する疑似的な数理モデルである．本研究では，スパースモデリング［5］によってシミュレーターと同等の結果を与えるエミュレーターを構成することができ，最終性能に寄与する変数を自動で選択できることを示す．

2 問題設定とアプローチ

Fig. 1 は生産される製品の生産工程の概略を示している．

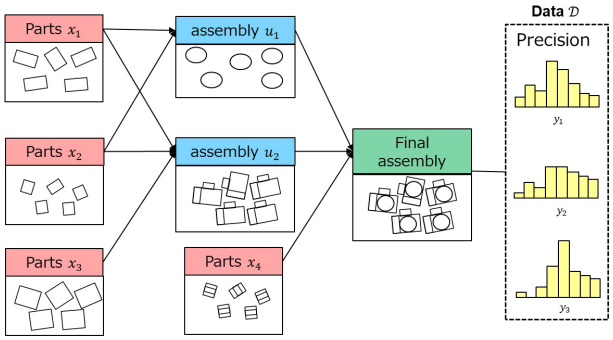


Fig. 1　生産工程の概略図．部品（parts）x から部組（assembly）u が組み立てられ，さらにそれらを組み合わせて総組（final assembly）が完成する．総組の性能が検査され N 個分のデータ y が得られる．

部品パラメーター $\boldsymbol{x}$ は，部品を特徴づける量を表す．例えば，穴径などの部品寸法や設計値からの位置ずれであり，設計仕様に基づいて予め定義域が決められている．これらが組み合わされ，部組といわれる中間的なモジュールが構成される．この特徴量を $\boldsymbol{u}$ と表す．さらにこれらが組み合わされて総組が完成する．設計指針に基づき，総組が検査され性能 $\boldsymbol{y}$ が得られる．

本研究では，総組の生産数 N 台を 1 ロットとし，ロット単位でデータが管理されていることを想定し，取得されるデータ $\mathcal{D}$ は 1 ロットあたり $\left\{\boldsymbol{y}^{(i)}\right\}_{i=1}^{N}$ と書く．不良要因検知の目標は，基準となる正常ロットから取得された製品データセット $\mathcal{D}_A$ と，不良が生じたロットから取得された製品データセット $\mathcal{D}_B$ が与えられた場合に，不良要因となる部品パラメーターを推定することである．しかし，特殊な場合を除いて，工程の上流のデータは得られないため，完全なデータ駆動型の変化要因検知は不可能である．そのため本研究では，生産工程シミュレーターと密度比推定を用いた不良要因検知手法を提案する．提案手法は，次の 2 つのパートからなる．

(i)．正常データセット $\mathcal{D}_A$ から，生産工程シミュレーターを用いて部品パラメーター $\boldsymbol{x}$ の分布を推定する．

(ii)．上記 (i) で推定された部品パラメーターセットの分布に基づいて，検知対象データ $\mathcal{D}_B$ から，各部品パラメーターの不良要因スコアを計算する．

特に上記 (i)．は，生産工程シミュレーターを $\boldsymbol{f}$ として

$$\boldsymbol{y} = \boldsymbol{f}_0(\boldsymbol{x}) + \boldsymbol{\epsilon} \equiv \boldsymbol{f}(\boldsymbol{x}) \tag{1}$$

において，製品性能 $\boldsymbol{y}$ から部品パラメーター $\boldsymbol{x}$ を求める問題に対応する．ここで $\boldsymbol{x}$ と $\boldsymbol{y}$ はそれぞれ D_x，D_y 次元のベクトルとする．また $\boldsymbol{\epsilon}$ は工程中に含まれるノイズであり，確率的因子として生産工程シミュレーター $\boldsymbol{f}$ に含まれている．また $\boldsymbol{f}_0$ はベースラインとなる決定的モデルであり，式 (1) は単体の製品の性能と部品パラメーターの関係を表していることに注意する．仮に，工程に含まれるノイズが適切にモデリングされており，取得されるデータが単体のものであれば，式 (1) に何らかの逆問題解法（正則化やベイズ推定）を用いることで，部品パラメーターを求めることは可能かもしれない（Fig. 2 (a)）．しかし本研究では量産データを扱うため，単体の製品性能が出力であるような式

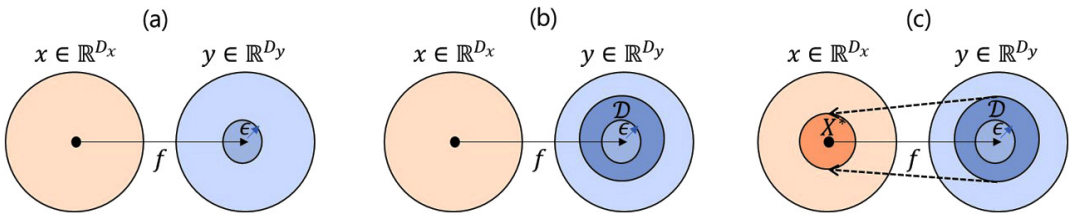


Fig. 2　(a)．生産工程シミュレーターの概念図．(b)．製品集合の性能データ $\mathcal{D} = \left\{\boldsymbol{y}_i\right\}_{i=1}^{N}$ とシミュレーターの関係．(c)．問題 (i) はデータ集合 $\mathcal{D}$ を表現する部品パラメーター集合 X^* を見つけることである．

(1) の逆問題として定式化することは難しい（Fig. 2 (b)）．本研究では，製品集合の性能から，これを上手く表現するような部品パラメーターの集合 X^* を推定するアルゴリズムを提案する．次に，正常データセットと不良品が含まれるデータセット $\mathcal{D}_B$ に対して，上記アルゴリズムを適用し，差分を調べることで不良の要因候補を見つけることができるはずである．

さらに（i）の過程でデータセット $\mathcal{D}$ から推定された X^* を用いることで $\boldsymbol{y}=\boldsymbol{f}(\boldsymbol{x})$ なる統計モデルを推定することが可能である．これは，シミュレーターと同等の入出力関係を持つエミュレーターと見做すことができる．

3 提案手法

本節では，提案法の詳細を述べる．

3.1. 近似ベイズ計算を用いた部品パラメーターの推定：Approximate Bayesian Computation（ABC）inspired inverse

部品パラメーター $\boldsymbol{x}$ の事後分布は，ベイズの定理によって次のように書ける．

$$p(\boldsymbol{x}|\boldsymbol{y})=\frac{p(\boldsymbol{y}|\boldsymbol{x})p(\boldsymbol{x})}{p(\boldsymbol{y})} \tag{2}$$

ここで $p(\boldsymbol{y}|\boldsymbol{x})$ は $\boldsymbol{x}$ が与えられた下での $\boldsymbol{y}$ の尤度であり，式（1）におけるシミュレーターに対応する $p(\boldsymbol{x})$ は事前確率に対応し，部品パラメーターが予め持っている情報を記述することができる．近似ベイズ計算は，尤度が複雑な関数である場合でも，データ $\mathcal{D}=\{\boldsymbol{y}_i\}_{i=1}^N$ が与えられた下で，次のようなプロセスで近似的に事後確率を近似的に推定することができる方法論である［6］，［7］．

1. 事前分布 $\pi(\boldsymbol{x})$ から N 個のパラメーター $\boldsymbol{x}^{(j)}, j=1,\ldots,N$ をサンプリングする．
2. シミュレーションを行う：$\boldsymbol{y}^{(j)}=\boldsymbol{f}\left(\boldsymbol{x}^{(j)}\right), j=1,\ldots,N.$
3. データとシミュレーションの出力の距離

$$\rho\left(S(\mathcal{D}),S\left(\boldsymbol{y}^{(i)}\right)\right)\leq\epsilon \tag{3}$$

を計算し，ϵ が小さい $\boldsymbol{x}^{(i)}$ を受理する．ここで，S は十分統計量，ρ は距離尺度である．

上記の通り，通常の近似ベイズ計算では，シミュレーターの出力がデータの分布全てを表現することを仮定しているが，本研究で使用するシミュレーターは前節で述べたように製品単体の性能を出力するため，製品集合全体の分布を表現できない．そこで，事前分布 $\pi(\boldsymbol{x})$ の範囲を広く取っておき，N_{sim} 回サンプリングを行い，それぞれの部品パラメーター $\boldsymbol{x}^{(j)}$ に対してシミュレーションを行うことを考える．すなわち，量産データを表現するために広範囲からサンプリングしたパラメーター $\boldsymbol{x}$ に対してシミュレートして，データ $\boldsymbol{y}$ を包含するような分布を作っておく．シミュレーションによって得られる分布は，データを含む大きな分布になるが，ここから密度比

$$r(\boldsymbol{y})=\frac{p^{\mathrm{data}}(\boldsymbol{y})}{p^{\mathrm{sim}}(\boldsymbol{y})} \tag{4}$$

に比例する受理確率に従って $\boldsymbol{x}$ を受理することにする．設計指針に基づき，総組が検査され性能 $\boldsymbol{y}$ が得られる．

ここで，$\mathcal{D}=\{\boldsymbol{y}_i\}_{i=1}^N\sim p^{\mathrm{data}}(\boldsymbol{y})$ と $\mathcal{D}_{\mathrm{sim}}=\left\{\boldsymbol{y}^{(i)}\right\}\sim p^{\mathrm{sim}}(\boldsymbol{y})$ はそれぞれデータとシミュレーター出力の密度である．このような受理基準は，次のような性質がある．

- $r(\boldsymbol{y}^{(j)})$ が小さければ，データ密度が小さく，このような場合は $\boldsymbol{y}^{(j)}$ を出力する $\boldsymbol{x}^{(j)}$ は棄却される．
- $r(\boldsymbol{y}^{(j)})$ が大きければ，データ密度が大きく，このような場合は $\boldsymbol{y}^{(j)}$ を出力する $\boldsymbol{x}^{(j)}$ を受理する．

このアルゴリズムを ABC inspired inverse と呼び，詳細を Algorithm1にまとめる．また，出力されるサンプルに対して，次の定理が導ける．

Lemma. Algorithm 1 において受理されたサンプル $Y^*=\left\{\boldsymbol{y}^{*(j)}\right\}_{j=1}^{n'}$ の分布は $p^{\mathrm{data}}(\boldsymbol{y})$ 一致する．

Theorem. Algorithm 1 において受理されたサンプル $X^*=\left\{\boldsymbol{x}^{*(j)}\right\}_{j=1}^{n'}$ の分布は事後分布 $p(\boldsymbol{x}|\mathcal{D})$ と一致する．

証明は Appendix に掲げる．実際には，密度比 r をデータ集合 $\mathcal{D}$ とシミュレーターの出力集合 $\mathcal{D}_{\mathrm{sim}}$ から推定する必要があり，この意味で推定誤差が加わる．密度比は Kullback-Leibler importance estimation procedure（KLIEP）と呼ばれる手法によって p^{sim} と p^{data} を各々推定することなく，直接推定することができる［8］．

Algorithm 1 ABC inspired inverse

Input: $\mathcal{D}=\{\boldsymbol{y}_i\}_{i=1}^N$, $\boldsymbol{f}(\cdot)$, $\pi(\cdot)$, and N_{sim}

Output: $\boldsymbol{X}^*=\left\{\boldsymbol{x}^{*(j)}\right\}_{j=1}^{n'}$

1: **for all** $j=1,\ldots,N_{\mathrm{sim}}$ **do**
2: Draw a sample from a prior: $\boldsymbol{x}^{(j)}\sim\pi(\boldsymbol{x})$.
3: Perform simulation: $\boldsymbol{y}^{(j)}=\boldsymbol{f}(\boldsymbol{x}^{(j)})$.
4: **end for**
5: Compute the density ratio by applying KLIEP: $r(\boldsymbol{y})=p^{\mathrm{data}}(\boldsymbol{y})/p^{\mathrm{sim}}(\boldsymbol{y})$.
6: Accept each sample $\boldsymbol{x}^{(j)}$ with the probability proportional to the density ratio.

3.2. 不良要因スコアの計算

ここでは，正常データ $\mathcal{D}_A$ と，別ロットから取得された検知対象データ $\mathcal{D}_B$ が取得された場合を考える．$\mathcal{D}_B$ は性能 $\boldsymbol{y}$ に不良があることを想定しているため，これを実現する部品パラメーターセットの分布は，$\mathcal{D}_A$ に対して Algorithm 1 を適用して得られるサンプル集合 X_A^* の分布とは異なっていると考えられる．ここで，不良ロットのデータ $\mathcal{D}_B$ は不良要因となる部品パラメーターの影響を強く受けて変化しているとすると考える．逆に，不良要因ではないパラメーターは，どのような値をとっても $\mathcal{D}_B$ を再現できないと考えられる．このような考えに基づいて，正常時の部品パラ

メーターの分布を用いて不良要因スコアを導出する．

まず正常データ $\mathcal{D}_B$ に対して Algorithm 1 を適用し，サンプル $\boldsymbol{X}_A^*$ を得る．次に，不良要因が独立に $\mathcal{D}_B$ の性能 $\boldsymbol{y}$ に影響を与えるとして，x_d を事前分布からサンプリングし，それ以外のパラメーター $\boldsymbol{x}_{\setminus d}$ を $\boldsymbol{X}_A^*$ からサンプリングする．このようにサンプリングされた部品パラメーター集合は，特定の変数 x_d 以外は正常データを出力するような集合であることに注意する．このようにサンプリングされた部品パラメーター集合と検知対象データ $\mathcal{D}_B$ に対し，Algorithm 1 を適用し，受理数 a_d を計算する．a_d が大きければ不良データを再現しており，a_d が小さければ x_d をどのようにサンプリングしても不良データを再現できない，と考えることができる．従って，受理数 a_d は，x_d が不良データをどの程度再現するかの指標であり，この値が大きければ x_d が不良要因である可能性が高いと考えられる．ただし，各部品パラメーターの最終的な不良要因スコアを算出する際には，次のことに注意する．

- ベースライン：設計仕様や調整プロセスが異なるため，総組の性能 $\boldsymbol{y}$ への感度は部品パラメーターに依存してまちまちである．
- 乱数系列依存性：確率的なアルゴリズムのため，乱数系列依存性がある．

上記を踏まえて，部品パラメーター x_d の最終的な不良要因スコア $\{c_d\}_{d=1}^{D_x}$ を次のように定義する．

$$c'_d = \frac{1}{M\bar{a}_d^0}\sum_{m=1}^{M} a_d^m, \left(1 \le d \le D_x\right) \tag{5}$$

$$c_d = \frac{c'_d}{\max\left(\boldsymbol{c}'\right)} \tag{6}$$

ここで，a_d^m は m 番目の乱数シードにおいて，$\mathcal{D}_B$ における x_d の受理数である，また，$\bar{a}_d^0 \equiv 1/M\Sigma_m^M a_d^{m,0}$ で定義され，$\mathcal{D}_A$ における受理数のシード平均を表す．また，$\max\left(\boldsymbol{c}'\right)$ は $\boldsymbol{c}' = \left[c'_1,\ldots,c'_{D_x}\right]$ の最大値を表し，不良要因スコア c_d は 0 から 1 の範囲を取る．不良要因スコア計算のためのアルゴリズムを Algorithm 2 にまとめる．

Algorithm 2 Defect factor detection via the ABC inspired inverse

Input: $\mathcal{D}_A$, $\mathcal{D}_B$, $\boldsymbol{f}(\cdot)$, $\pi(\cdot)$, N_{sim}, and M
Output: $\boldsymbol{c} = \{c_d\}_{d=1}^{D_x}$

Apply the ABC inspired inverse to $\mathcal{D}_A$, and obtain $\boldsymbol{X}_A^* = \left\{\boldsymbol{x}^{*(j)}\right\}_{j=1}^{n'}$.
for all $m = 1,\ldots,M$ **do**
  for all $\xi = \mathcal{D}_A$ and $\mathcal{D}_B$ **do**
    for all $d = 1,\ldots,D_x$ **do**
      for all $j = 1,\ldots,n'$ **do**
        Generate samples: $x_d^{(j)} \sim \pi(x_d)$, $\boldsymbol{x}_{\setminus d}^{(j)} = \boldsymbol{x}^{*(j)}$.
        Perform simulator: $\boldsymbol{y}^{(j)} = \boldsymbol{f}(\boldsymbol{x}^{(j)})$.
      end for
      Compute the density ratio by using KLIEP: $r(\boldsymbol{y}) = p^{\mathrm{data}}(\boldsymbol{y})/p^{\mathrm{sim}}(\boldsymbol{y})$.
      Accept each sample $\boldsymbol{x}^{(j)}$ with the probability proportional to the density ratio.
      Assign the total number of accepted samples $a_d^{m,\xi}$.
    end for
  end for
end for
Calculate the defect score (6).

3.3. エミュレーターの推定

本節では Algorithm 1 によって推定された部品パラメーターのサンプル X^* とデータ $\mathcal{D}$ を用いて，エミュレーター（疑似モデル）を推定する．生産工程中では特定の部品パラメーターの値が調整されることがあるため，最終的な性能に対して感度がない部品パラメーターが含まれる．

L1正則化付き線形モデルを用いることで，性能を出力する統計モデルを構成することができると共に，最終的に感度がある部品パラメーターを自動的に抽出することができる．

総組の性能 $\boldsymbol{y}$ と部品パラメーター $\boldsymbol{x}$ の関係が

$$y_k = f_k\left(\boldsymbol{x}\right) \simeq \boldsymbol{w}_k^T\boldsymbol{x}, 1 \le k \le D_y \tag{7}$$

と表すことができ，係数 $\boldsymbol{w}_k \in \mathbb{R}^{D_x}$ はスパースなベクトルであると仮定する．すなわち，総組の精度と部品パラメーターの関係を完全な線形モデルで近似し，なおかつ影響を及ぼす部品パラメーターは限られているとする．データ $\mathcal{D}_A$ から Algorithm 1 によって推定された部品データの集合 $X_0 = \left\{\boldsymbol{x}_0^{(i)}\right\}_{i=1}^{N}$ から $\left\{\boldsymbol{w}_k\right\}_{k=1}^{D_y}$ を次のように推定する．

$$\boldsymbol{w}_k^* = \arg\min_{\boldsymbol{w}_k}\left[\sum_i^N\left(y_k^{(i)} - \boldsymbol{w}_k^T\boldsymbol{x}_0^{(i)}\right)^2 + \lambda\left\|\boldsymbol{w}_k\right\|_1\right], \quad 1 \le k \le D_y \tag{8}$$

ここで $\|\cdot\|_1$ は L1ノルムを表し，λ は L1正則化の強さを決めるパラメーターである．この最適化は Lasso と呼ばれ Least Angle Regression（LARS）や Alternating Direction Method of Multipliers（ADMM）を用いて実行できる［9］，［10］．

4 シミュレーターの構成

本研究では，カメラ内部のモジュールの生産を扱う．このモジュールの生産工程は，Fig. 3 のような工程モデルによって表される．16の部品（x_1, ..., x_{16}）から部組が構成され，そのうちの 3 つの部品（x_1, x_5, x_{12}）に関して調整を行い，さらに 4 つの部品（x_{17}, ..., x_{20}）を組付け，総組状態で最終的に 4 次元の精度指標（y_1, ..., y_4）が計測される．すなわち $D_x = 20$，$D_y = 4$ である．なお，調整工程では u_i, u_i', u_i'', u_i''' などの中間出力が一定の規格内に収まるよう部品（x_1, x_5, x_{12}）の位置を Δx_1，Δx_5，Δx_{12} だけずらす．量産工程では500の総組 1 ロットとして生産されるとする．

部品パラメーター x_d の事前分布は，一様分布 $\mathcal{U}\left(\lambda_d^1, \lambda_d^2\right)$ で与えられることとし，分布のパラメーターは設計仕様か

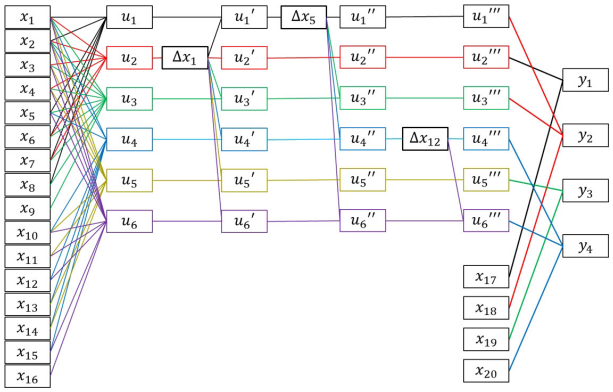


Fig. 3　カメラ内モジュールの工程モデル．20の部品パラメーターから総組が構成され，4つの精度指標で検査が行われる．本工程では，x_1，x_5とx_{12}に対して調整が行われる．

ら次のように設定する．

$$\lambda_d^1 = \begin{cases} -0.1, 1 \le d \le 16 \\ -0.2, 17 \le d \le 20 \end{cases}, \lambda_d^2 = \begin{cases} 0.1, 1 \le d \le 16, \\ 0.2, 17 \le d \le 20 \end{cases} \tag{9}$$

5 結　果

5.1. 不良要因検知

5.1.1. 数値シミュレーション

本節では，不良要因検知のシミュレーションを行い，提案手法の有効性と限界を示す．これは次のような手続きで行う．

1. 部品パラメーターx_dを正規分布$\mathcal{N}\left(\mu_d, \sigma_d^2\right)$に従って$N$サンプル生成し，$\boldsymbol{y}^{(i)} = \boldsymbol{f}\left(\boldsymbol{x}^{(i)}\right), i = 1, ..., N$を計算する．このセット$\mathcal{D}_A = \left\{\boldsymbol{y}^{(i)}\right\}_{i=1}^N$を正常データとする．ここで，製品仕様から

$$\mu_d = 0, 1 \le d \le 20 \ , \sigma_d^2 = \begin{cases} 0.01, 1 \le d \le 16, \\ 0.02, 17 \le d \le 20 \end{cases} \tag{10}$$

と決めた．

2. 不良要因は正常の分布からずれていると仮定し，正規分布$\mathcal{N}\left(\mu_f + \delta\mu_f, \left(\sigma_f + \delta\sigma_f\right)^2\right)$から$N$サンプル生成する．その他の部品パラメーター$x_d$は正常時の分布$\mathcal{N}\left(\mu_d, \sigma_d^2\right)$から$N$サンプル生成する．このサンプルセットを用いて$\boldsymbol{y} = \boldsymbol{f}(\boldsymbol{x})$から$\mathcal{D}_B = \left\{\boldsymbol{y}^{(i)}\right\}_{i=1}^N$を検知対象データとする．
3. $\mathcal{D}_A$と$\mathcal{D}_B$を用いてAlgorithm 1とAlgorithm 2を適用し，最終的な不良要因スコア$\left\{c_d\right\}_{d=1}^{D_x}$を得る．

まず，下流部品パラメーターx_{19}を不良要因だと想定し，生成分の平均と分散を$\delta\mu_{19} = 0.1$，$\delta\sigma_{19} = 0.1$だけ変化させた場合をFig. 4 (a) 左に示すx_{19}が変化したことによってy_3が変化している．Fig. 4 (a) 右から，真の不良要因パラメーターx_{19}のスコアc_{19}が高い値になっていることが分かる．部品パラメーターx_{13}, x_{14}とx_2も高いスコアを示しているが，これはy_3と関連しているパラメーターであることがFig. 3からわかるため，合理的な結果であると考えられる．Fig. 4 (b) は，x_1を変化させた場合の結果である．この場合，性能パラメーター$\boldsymbol{y}$と不良要因スコアに顕著な差異が見られないが，これはx_1が調整工程を経ているため$\boldsymbol{y}$に対して感度がほとんどないことが原因であると考えられる．このように，結果系に対して感度がない変数は不良要因として特定することは不可能である．一方，Fig. 4 (c) はx_7を真の不良要因だと想定した場合である．上流のパラメーターであってもc_7が顕著に高い値を示し，不良要因として特定で

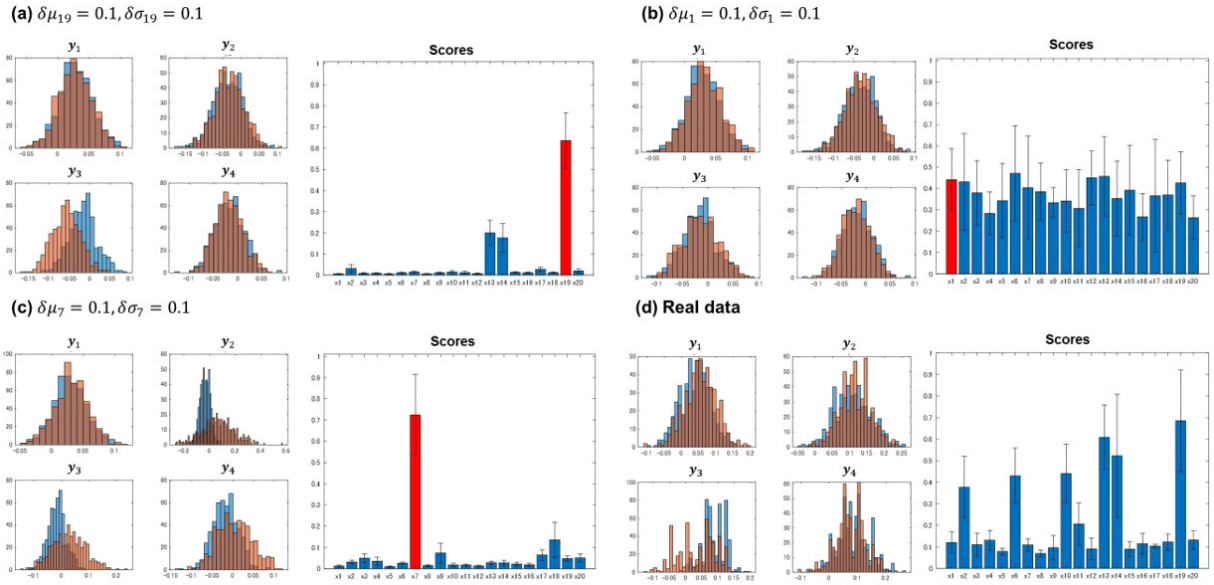


Fig. 4　(a)–(c)．シミュレーションの結果．左列のヒストグラムは$\mathcal{D}_A$の分布（青）との分布（赤）を表す．右列の棒グラフは不良要因スコアcを表し，エラーバーは標準偏差を表す．右列は不良要因スコアを表し，真の不良要因を赤で表している．(d)．左列のヒストグラムは実際に取得された正常ロットのデータ$\mathcal{D}_A$の分布（青）と不良品を含むロットのデータ$\mathcal{D}_B$の分布（赤）を表す．

きていることが分かる．

5.1.2. 実データの検証

本節では，実際の生産現場で取得されたデータの解析結果を述べる．取得された約140ロットの生産データから，正常に生産されているロットと不良が確認されたロットを1つずつ選定した．Fig. 4 (d) にデータと不良要因スコア示す．このときは，特に y_3 が顕著に変化しており，不良要因スコア c_{19} が顕著に高い値を取っていることが分かる．実際の現場では x_{19} の部品に不具合があったことが確認されていることから，これは得られた結果と整合する．また，その他の高い不良要因スコアを持つ変数についても，Fig. 3 から y_3 に紐づく変数が検出されていることから，製品設計の観点から自然な結果であると考えられる．

5.2. エミュレーターの推定

Fig. 5 (a) は，設計の感度解析から得られる $\boldsymbol{y}$ と $\boldsymbol{x}$ の設計知識による感度解析の結果である． 0 は感度が低いことを表し， 1 は感度が高いことを表す．Fig. 5 (b) は，$\boldsymbol{y}$ の各々の成分に対して推定された $\boldsymbol{w}^*$ である．横軸は正則化パラメーター λ を表し Fig. 中の垂線はクロスバリデーションの結果を表す．$\boldsymbol{w}^*$ は20次元のベクトルであり，Fig. 5 (a) のドメイン知識に基づいて，感度があるものは実線，ないものは破線で表現している．この図から，感度がない部品パラメーターに対応する成分は 0 になり，感度があるパラメーターは有限の値を取る傾向があることが分かる．したがって，完全な線形モデルで近似された式 (7) で表されるエミュレーターは，性能と部品パラメーター間の感度を表すモデルとして有用であると考えられる．また $\boldsymbol{w}^*$ が 0 になりやすい部品パラメーターは，性能への感度が小さいため，不良要因スコアとして抽出しづらいといえる．

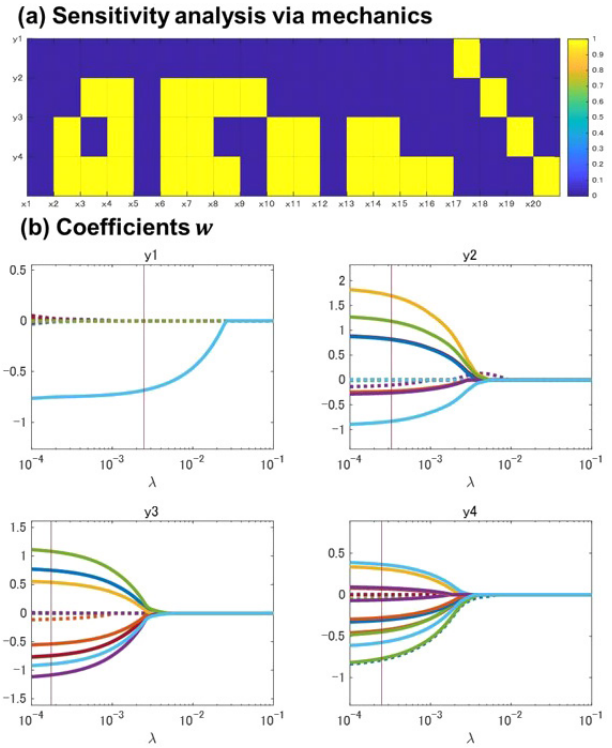


Fig. 5　(a)．設計の感度解析からドメイン知識として得られる結果．精度 y_k に感度が高ければ 1，低ければ 0 を表している．(b)：Lasso によって得られた正則化パラメーター λ に対する係数 w の結果．垂線はクロスバリデーションで得られた λ の値である．

6　まとめ

本論稿では，限られたデータしか取得できないような量産工程において，シミュレーションを用いた逆問題解法 (ABC inspired inverse) と不良要因検知手法を提案した．シミュレーションを通じて提案手法の適用可能性と限界について論じ，検知可能な変数の特性を明らかにした．さらに，実際の生産工程で得られたデータを用いて提案手法を適用し，現場での検証結果と整合する結果を得ることができた．また，複雑な非線形モデルを代替するエミュレーターの構成方法を提案した．提案手法はスパースモデリングを活用することで，感度のある部品パラメーターを自動で選択することも可能である．本研究の成果は，生産工程の効率化・自律化の観点で重要な役割を果たすと考えられる．

Appendix A. Lemma の証明

次の表記を用いると

$$I=\begin{cases}1 \text{ if } \boldsymbol{y} \text{ is accepted}\\ 0 \text{ otherwise,}\end{cases} \tag{A. 1}$$

受理された $\boldsymbol{y}$ の分布は

$$p(\boldsymbol{y}^*)=p(\boldsymbol{y}|I=1)=\frac{p^{\mathrm{sim}}(\boldsymbol{y})\Pr(I=1|\boldsymbol{y})}{\Pr(I=1)}$$

$$=\frac{cp^{\mathrm{sim}}(\boldsymbol{y})\dfrac{p^{\mathrm{data}}(\boldsymbol{y})}{p^{\mathrm{sim}}(\boldsymbol{y})}}{\Pr(I=1)}=p^{\mathrm{data}}(\boldsymbol{y}), \tag{A. 2}$$

と計算できる．ここで $\Pr(I=1|\boldsymbol{y})$ はサンプルの受理確率を表し，式 (4) よりデータの出力とシミュレーターの出力の密度比に比例する．比例定数は $c \equiv \Pr(I=1)$ としている．□

Appendix B. Theorem の証明

$$p(\boldsymbol{x}|I=1)=\int p(\boldsymbol{x}|\boldsymbol{y})p(\boldsymbol{y}|I=1)d\boldsymbol{y}$$

$$=\pi(\boldsymbol{x})\int\frac{p(\boldsymbol{y}|\boldsymbol{x})p^{\mathrm{data}}(\boldsymbol{y})}{p^{\mathrm{sim}}(\boldsymbol{y})}d\boldsymbol{y} \tag{A. 3}$$

ここで Lemma を用いた．一方，ベイズの定理より $\boldsymbol{x}$ の事後確率は下記のように表すことができる．

$$p(\boldsymbol{x}|\mathcal{D}) = \frac{\pi(\boldsymbol{x})}{p(\mathcal{D})}\int p(\boldsymbol{y}|\boldsymbol{x})p(\mathcal{D}|\boldsymbol{y})d\boldsymbol{y} \qquad (A.4)$$

よって，式（$A.3$）と式（$A.4$），$p(\mathcal{D}|\boldsymbol{y}) \propto \frac{p^{\mathrm{data}}(\boldsymbol{y})}{p^{\mathrm{sim}}(\boldsymbol{y})}$ により $p(\boldsymbol{x}|I=1) = p(\boldsymbol{x}\,|\,\mathcal{D})$ となり，受理された $\boldsymbol{x}$ の分布と事後分布が一致することが示された．□

引　用　文　献

［1］B. B. a. H. K. J. Lee, “A Cyber Physical Systems architecture for Industry 4.0-based manufacturing systems,” *Manufacturing Letters*, vol. 3, pp. 18–23, 2015.

［2］岡崎隼也，“半導体製造工程における Virtual Metrology 技術，”*応用数理*，第29巻，第１，pp. 31–34, 2019.

［3］J. M. a. J. Iskandar, “Big Data Analytics for Smart Manufacturing: Case Studies in Semiconductor Manufacturing,” *Processes*, vol. 5, no. 39, 2017.

［4］佐々木豊春，新海雅彦，東山孝一郎，田中文基，岸浪健史，“鏡筒光学製品における統計的公差設計システム，”*精密工学会誌*，第64巻，第７，pp. 1090–1095, 1998.

［5］I. R. a. G. Grabarnik, Sparse Modeling: Theory, Algorithms, and Applications, CRC Press, 2015.

［6］Y. F. a. M. A. B. S. A. Sisson., Handbook of approximate Bayesian computation, Chapman & Hall/CRC., 2016.

［7］D. J. B. R. C. G. a. P. D. S. Tavare, “Missing data: Our view of the state of the art,” *Psychological Methods*, vol. 7, no. 2, pp. 147–177, 1997.

［8］T. S. a. T. K. M. Sugiyama, Density Ratio Estimation in Machine Learning, Cambridge Press, 2012.

［9］R. Tibshirani, “Regression Shrinkage and Selection via Lasso,” *J. R. Statist. Soc. B*, vol. 58, no. 1, pp. 267–288, 1996.

［10］S. Boyd, “Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers,” *Foundations and Trends in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2010.

［11］大坪洋介，大谷直也，近末恵美，杉山将，“産工程における不良要因検知，”*第34回人工知能学会全国大会論文集*, Vols. 214-GS-2-04, pp. 1–4, 2020.

大坪洋介　Yosuke OTSUBO
研究開発本部　数理技術研究所
Mathematical Science Research Laboratory
Research & Development Division

近末恵美　Megumi CHIKASUE
研究開発本部　数理技術研究所
Mathematical Science Research Laboratory
Research & Development Division

大谷直也　Naoya OTANI
研究開発本部　数理技術研究所
Mathematical Science Research Laboratory
Research & Development Division

杉山　将　Masashi SUGIYAMA
理化学研究所
RIKEN
東京大学
The University of Tokyo


大坪洋介
Yosuke OTSUBO


大谷直也
Naoya OTANI


近末恵美
Megumi CHIKASUE


杉山　将
Masashi SUGIYAMA

外部発表一覧（2020年）
List of papers presented/published externally (2020)

題名 Article Title	著者名 Author	掲載誌名，発表学会・協会 Journal title	巻，号，頁，年，論文番号 等 Publication information
Enhanced Fourier-transform method for high-density fringe analysis by iterative spectrum narrowing	Shigeru Nakayama, Hidemitsu Toba, Naoki Fujiwara, Takashi Gemma, Mitsuo Takeda*	Applied Optics	Vol. 59, Iss. 29, pp. 9159–9164, 10 October 2020
Association between hypertension and retinal vascular features in ultra-widefield fundus imaging	Gavin Robertson, Alan Fleming, Michelle Claire Williams*, Emanuele Trucco*, Nicola Quinn*, Ruth Hogg*, Gareth J McKay*, Frank Kee*, Ian Young*, Enrico Pellegrini, David E Newby*, Edwin J R van Beek*, Tunde Peto*, Baljean Dhillon*, Jano van Hemert, Thomas J MacGillivray*	Open Heart	Vol. 7, Iss. 1, Article number e001124, 8 January 2020
Optic Disc and Fovea Localisation in Ultra-widefield Scanning Laser Ophthalmoscope Images Captured in Multiple Modalities	Peter R. Wakeford, Enrico Pellegrini, Gavin Robertson, Michael Verhoek, Alan D. Fleming, Jano van Hemert, Ik Siong Heng*	Annual Conference on Medical Image Understanding and Analysis	MIUA 2019: Medical Image Understanding and Analysis pp. 399–410 Conference paper First Online: 24 January 2020
CT performance of compact X-ray source with small focal spot using a 950 keV linear accelerator	Norihito Matsunaga, Tomoya Sato, Atsushi Yamada, Masayuki Zaike, Daisuke Nikai, Tadashi Hatano*, Masashi Yamamoto*	10th Conference on Industrial Computed Tomography（iCT2020）	doi: 10.1117/12.2550780（March 2020）
Pixelated Mask optimization on Quantum Computers	Yosuke Okudaira, Satoshi Yashiki	SPIE Advanced Lithography	SPIE 11327, Optical Microlithography XXXIII, 1132705（23 March 2020）
Novel overlay correction using inline alignment station（iAS）for scanner	Takahisa Kikuchi, Ayako Sugimoto, Shigeru Eto, Akira Okutomi, Naoki Morita	SPIE Advanced Lithography	SPIE 11327, Optical Microlithography XXXIII, 113270W（23 March 2020）
High-order field distortion correction using standalone alignment technology with modeling and sampling optimization	Takehisa Yahiro, Katsushi Makino, Haruki Saito, Steven Tottewitz*, Boris Habets*, Patrick Lomtscher*, Jiro Hanaue*	Proceedings of SPIE, Metrology, Inspection, and Process Control for Microlithography XXXIV	Vol. 11325, 2020, Article number 113251Y
Single-Crystalline Protrusion-Rich Indium Tin Oxide Nanoparticles with Colloidal Stability in Water for Use in Sustainable Coatings	Ryoko Suzuki, Yasutaka Nishi, Masaki Matsubara*, Atsushi Muramatsu*, Kiyoshi Kanie*	ACS Applied Nano Materials	Vol. 3, Iss. 5, pp. 4870–4879, 2020
A general theory of far-field optical microscopy image formation and resolution limit using double-sided Feynman diagrams	Naoki Fukutake	Scientific Reports	Vol. 10, 2020, Article number 17644
Principal Vibration Modes of the La2O3–Ga2O3 Binary Glass Originated from Diverse Coordination Environments of Oxygen Atoms	Kohei Yoshimoto, Atsunobu Masuno*, Itaru Sato, Yoshinobu Ezura, Hiroyuki Inoue*, Motoi Ueda, Masafumi Mizuguchi, Yutaka Yanaba*, Tatsunori Kawashima, Tomoki Oya, Yohei Onodera*, Shinji Kohara*, Koji Ohara*	Journal of Physical Chemistry B	Vol. 124, Iss. 24, pp. 5056–5066, 2020
画像処理による蛍光顕微鏡Zスタックデータの DOF 調整と3D 観察	萩原恒幸	光学シンポジウム講演予稿集	Vol. 45th（CD-ROM），pp. 15–19（2020.06.25）
生産工程における不良要因検知	大坪洋介，大谷直也，近末恵美，杉山　将*	人工知能学会全国大会（Web）	Vol. 34th，Page.ROMBUNNO.2I4-GS-2-04（WEB ONLY）（2020）
光と物質の相互作用を利用した超解像蛍光顕微鏡	嶽　文宏	精密工学会誌（Web）	Vol. 86, No. 7, pp. 524–528（J-STAGE）（2020）
Search for oxide glass compositions using Bayesian optimization with elemental-property-based descriptors	Kensaku Nakamura, Naoya Otani, Tetsuya Koike	Journal of the Ceramic Society of Japan	Vol. 128, Iss. 8, pp. 569–5721, August 2020
OCTを含む全てのレーザー顕微鏡を扱う統一結像理論	福武直樹，安野嘉晃*	第81回　応用物理学会秋季学術講演会	応用物理学会秋季学術講演会講演予稿集（CD-ROM）Vol. 81st Page. ROMBUNNO.9p-Z17-2（2020.08.26）
Binary classification with ambiguous training data	Naoya Otani, Yosuke Otsubo, Tetsuya Koike, Masashi Sugiyama*	Machine Learning	Vol. 109, Iss. 12, pp. 2369–2388, December 2020

*印は，株式会社ニコン及びニコングループ会社外の共同研究者
*Co-researcher outside NIKON CORPORATION and Nikon Group Companies

Nikon Research Report Vol. 3
Published September 2021

NIKON CORPORATION