Development of a Digital Laboratory Integrating Modular Measurement Instruments

Akira Aiba*, †, Kazunori Nishio* and Taro Hitosugi*, **

Abstract

Recent advancements in digital technologies and machine-learning algorithms have contributed significantly to the development of digital laboratories. These systems autonomously investigate materials by integrating automated experimental setups. In this study, we developed a digital laboratory that connects a sputter deposition system, an X-ray diffraction (XRD) instrument (Rigaku SmartLab), and other measurement instruments. The key features of our system include (1) modularization of each experimental instrument to enable flexible adaptation to various experiments and (2) centralized cloud storage of measurement data in a unified format, allowing for data-driven materials science using machine learning. This article also presents a case study of autonomous experimentation to maximize the X-ray diffraction peak intensity ratio of $LiCoO_2$ thin films.

1. Introduction

Progress in machine learning and robotics has enabled automated material discovery[(1)–(3)] by delegating repetitive tasks, such as synthesis, property measurement, and data analysis, to robotic systems. Collecting large volumes of physical property data and associated metadata (e.g., synthesis conditions and environmental parameters) is essential for accelerating the discovery of novel materials.

In materials research, integrating diverse synthesis systems with structural and property characterization instruments necessitates standardizing the shape and size of samples and modularizing instruments. Research based on liquids has benefited from standardized tools, such as 96-well plates[(4)]. However, efforts to standardize and modularize equipment for solid-state materials, ranging from powders to thin films, are still under development[(5),(6)]. In particular, standardization of sample holder specifications and communication protocols is extremely important, and standardization of measurement data formats is also essential for improving interoperability and enabling efficient data analysis.

To address these challenges, we developed a digital laboratory (dLab) for thin-film solid-state samples that supports fully automated material synthesis and characterization via X-ray diffraction (XRD), scanning electron microscopy (SEM), Raman spectroscopy, ultraviolet-visible (UV-Vis) spectrophotometry, and electrical conductivity measurements. In dLab, modular instruments are interconnected through standardized connection methods and communicate using the Measurement analysis instrument Markup Language (MaiML) format, which has been officially registered as a Japanese Industrial Standard (JIS) standard. The integrated data are analyzed in the cloud, allowing machine-learning algorithms to determine subsequent synthesis parameters autonomously.

The dLab was applied to the synthesis of high-quality $LiCoO_2$(001) thin films, a representative positive-electrode material for lithium-ion batteries, by optimizing the intensity ratio of the 003 and 006 out-of-plane XRD peaks. This study, published in *Digital Discovery*[(7)], demonstrates the potential of integrating modular measurement instruments into an autonomous digital laboratory for advanced materials exploration.

2. Experiment

2.1. Overview

The dLab integrates modular thin-film synthesis and measurement instruments via a robotic transport system and unified control network (Figs. 1 and 2). Each module adheres to standardized specifications, including a sample holder with a sample size of 10 mm × 10 mm and 0.5 mm thickness, TCP/IP-based communication protocol, and data output in MaiML format. Specified CF flanges are used to interconnect the chambers and modules.

The dLab includes RF and DC sputtering systems, an X-ray diffractometer, a scanning electron microscope (SEM), a Raman spectrometer, a UV-Vis spectrophotometer, and an electrical conductivity measurement module, all compatible with robotic handling. The XRD module, based on the Rigaku SmartLab system, features a vacuum chuck for sample mounting. It is equipped with an integrated robotic arm (COBOTTA PRO shown in Fig. 1a) and an added automatic door support for automatic sample exchange (Fig. 1c).

"Bridge Server (Rigaku)" middleware is responsible for communication between the XRD system control

* School of Materials and Chemical Technology, Institute of Science Tokyo.
† Present address: Rigaku Corporation.
** Department of Chemistry, School of Science, The University of Tokyo.

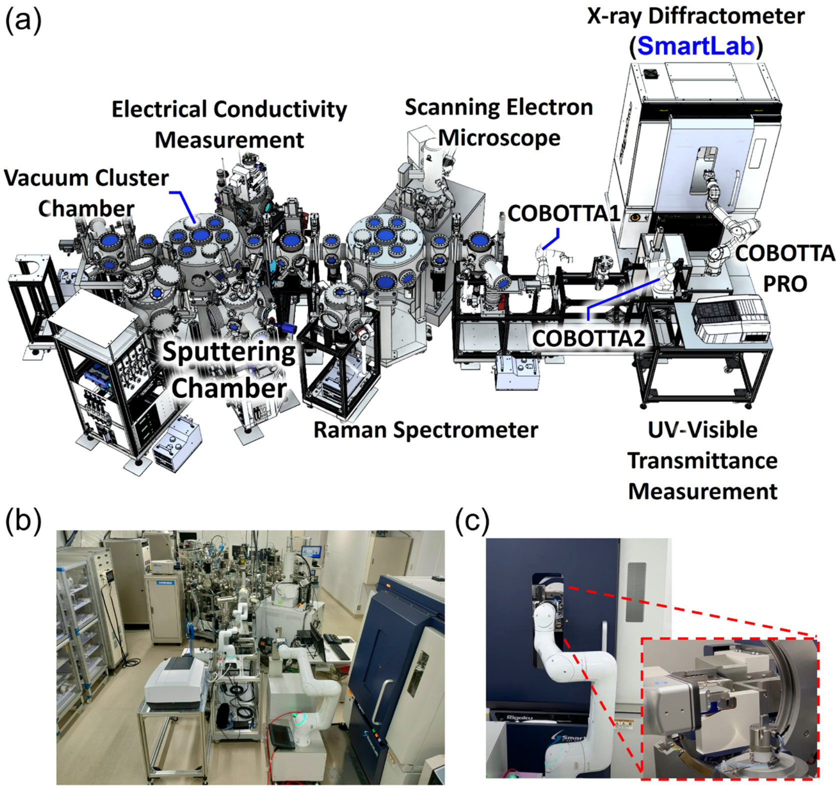

Fig. 1. Overview of the Automated and Autonomous Experimentation System

The dLab system integrates modular synthesis apparatuses (sputtering deposition system) with various property characterization and analysis instruments, including an X-ray diffractometer, a scanning electron microscope (SEM), a Raman spectrometer, a UV-Vis spectrophotometer, and an electrical conductivity measurement module. All components are interconnected for fully automated operation. Turbomolecular pumps continuously maintain a high-vacuum environment, achieving a base pressure of approximately 10^{-6} Pa. The system architecture adheres to the standards publicly disclosed at https://github.com/Hitosugi24/dLab regarding the shape and dimensions of the sample holders and the communication protocols used for interfacing between modules. Adapted from Ref. 7 with permission from the Royal Society of Chemistry.

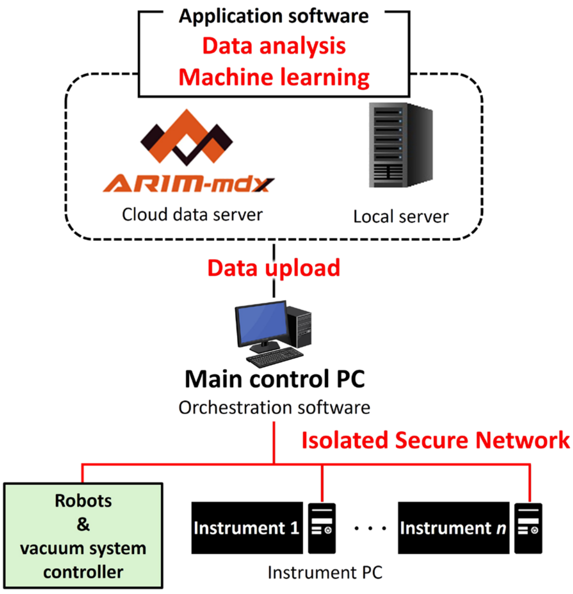

Fig. 2. Network Configuration Diagram of the System

The system's network is structured with a clear division between external and protected segments. Each synthesis and measurement/analysis module is connected to a main control PC via isolated and secured network. The main PC is also connected to a cloud-based data server, enabling seamless data upload and remote access to the experimental results.

software and the central control system, translating commands and converting output data into MaiML format. These data are uploaded to the cloud and a local server for analysis and machine learning. The XML-based MaiML format allows structured recording of measurement, pre-processing, and post-processing steps. It also makes use of reproducibility, traceability, and secure data reuse. An explanatory article of the MaiML format is available at https://doi.org/10.11470/oubutsu.92.3_142, and its guideline can be found at https://www.maiml.org/ (in Japanese).

2.2. Workflow of Autonomous Experiment

The Main Control PC (Fig. 2) orchestrates experiments, managing modules and sample transfer based on task files. Synthesis and measurement conditions are stored as text files; during optimization, selected parameters are varied to generate new tasks.

The dLab system autonomously synthesizes and analyzes thin films. Then, the data is uploaded to the cloud storage in MaiML format. Using cloud-stored data, a Bayesian optimization algorithm[(8)–(10)] suggests the next deposition parameters. We use the ARIM-mdx data system[(11)] for the cloud data server (note: this is different from the server used in the original study[(7)]). The system performs a closed-loop cycle of synthesis, measurement, analysis, and re-planning without human intervention, optimizing deposition conditions—here focused on substrate temperature as the deposition parameter.

2.3. Automatic XRD Peak Analysis of Thin Film Materials

An automated analysis program for out-of-plane XRD data analysis was developed using Python. Because Rietveld refinement is not applicable to oriented thin films, the program instead applies the BEADS algorithm[12] for baseline removal and uses find_peaks() function from SciPy[13] for peak detection.

The program analyzes the XRD pattern files in MaiML format using the workflow shown in Fig. 3. To distinguish the thin-film peaks from the substrate and K_β radiation, reference XRD data from a pristine Al_2O_3(0001) substrate was used. The Miller indices were then assigned to the peaks using the JCPDS data. $LiCoO_2$ peaks were assigned (not shown in the figure) to the 003n ($1 \leq n \leq 5$) reflections, matching results obtained manually by experts.

3. Result

The dLab system autonomously maximizes the crystallinity of $LiCoO_2$(001) thin films deposited on Al_2O_3(0001) substrates via RF magnetron sputtering. The degree of layered cationic ordering[14, 15], which enhances Li-ion diffusion[16–18], was evaluated by the intensity ratio of the 003 and 006 XRD peaks (I_{003}/I_{006}). This ratio, calculated automatically from the MaiML-formatted XRD data, served as the optimization target.

The substrate temperature (T_s) was varied from 200°C to 750°C in 10°C steps in the optimization. The initial measurements were performed at 200°C and 750°C. Subsequent deposition conditions were autonomously proposed using Bayesian optimization, forming a closed loop.

A total of 25 optimization iterations (27 data points) were conducted (Fig. 4), each taking approximately 2 hours. The highest I_{003}/I_{006} ratio, 35.1, was achieved at T_s = 660°C during the 14th iteration. The optimization curve revealed that the layered ordering significantly improved within the 600–700°C range. Owing to the automated system's ability to operate continuously, including overnight and on weekends, the process was significantly faster than manual experimentation, which would require additional time for sample handling, storage, and limited working hours.

This study demonstrated the ability of dLab to autonomously analyze XRD data and iteratively adjust the deposition parameters to achieve high-quality thin films without human intervention.

4. Outlook

The future development of the dLab system is envisioned along several key directions:

[System Expansion]

The dLab is not limited to thin films—it can also handle pellet samples, making it adaptable to bulk synthesis. By replacing the thin-film deposition module with a bulk synthesis module, the dLab could serve as a general platform for solid-state material development, enabling flexible system expansion to cover not only inorganic materials but also ceramics, polymers, and organic compounds.

The XRD system also supports automatic X-ray reflectivity (XRR) measurements. A future direction is to automate hardware adjustments using robotic systems, such as exchanging optical components such as monochromator and K_β filter. Enhancing robotic capabilities would broaden the range of accessible experiments.

Additional synthesis and characterization modules can be integrated *plug-and-play*, following publicly available standards, which are documented at https://github.com/Hitosugi24/dLab.

[Enhanced Data Utilization]

The dLab is the first full-scale implementation of the MaiML format, converting proprietary data into a standardized structure. MaiML can record the full experimental workflow—from sample preparation to data analysis—and can also serve as a command language for automated procedures. Currently, dedicated analysis software is used, but we plan to use third-party software to promote automation and standardization across different devices and systems.

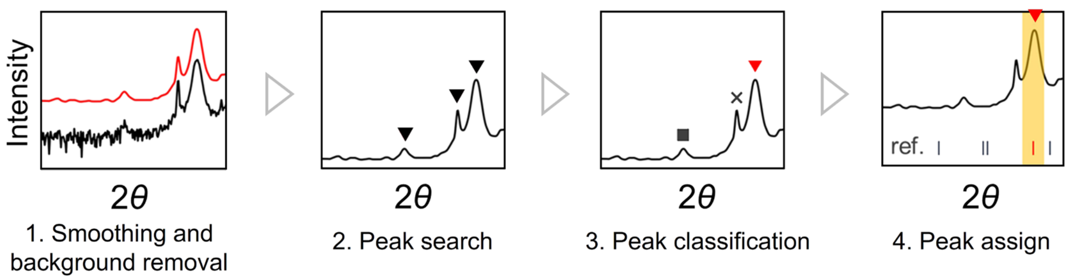

Fig. 3. Workflow of automatic XRD peak analysis

1. Smoothing of the data. The black line represents the raw data, and the red line shows the diffraction pattern after smoothing. 2. Peak detection. The identified peaks are marked with black ▼ symbols. 3. Peak classification. The peaks originating from the thin film (red ▼) and those from other sources (■: substrate-derived, ×: K_β line-derived) are distinguished. 4. Peak assignment. Peaks originating from the thin film are identified by comparing them with the powder XRD pattern. Reproduced from Ref. 7 with permission from the Royal Society of Chemistry.

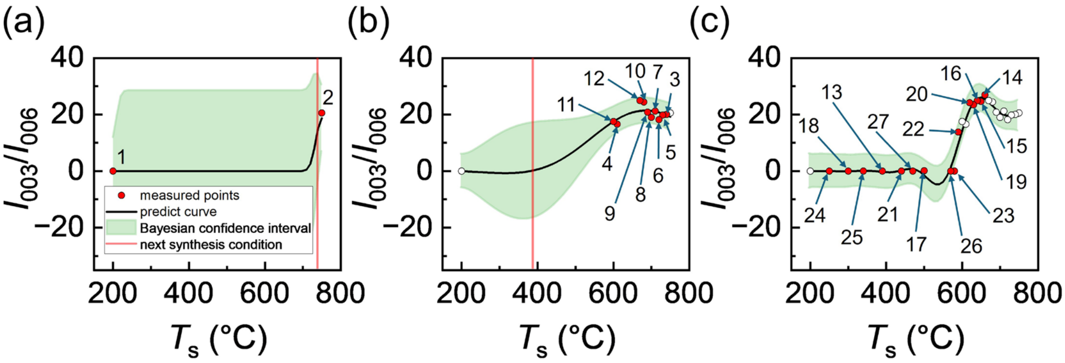

Fig. 4. Autonomous Exploration Process via Bayesian Optimization

(a) The exploration began with the initial dataset of substrate temperatures (T_s): minimum T_s of 200°C and maximum T_s of 750°C. The red circles represent the intensity ratios I_{003}/I_{006} of the $LiCoO_2$ 003 and 006 X-ray diffraction peaks for measured samples. The black line indicates the predicted curve, and the green shaded region represents the credible interval. The next synthesis condition is autonomously selected at the point where the acquisition function reaches its maximum value (red line in the figure). (b) After 11 cycles. (c) After 27 cycles. The white circles represent the previous experimental results, as shown in (a) and (b). Reproduced from Ref. 7 with permission from the Royal Society of Chemistry.

[Synthesis Process Logging and Learning]

Process data, such as temperature, gas flow, and pressure, are automatically logged. By linking this information with physical property measurements, the dLab generates rich datasets for machine learning. The next step is to uncover the hidden relationships between processes and properties.

[Adoption of Various Optimization Methods]

This study used Bayesian optimization with a Gaussian process regression model. The dLab system is designed to support other optimization methods. Software development is needed to allow flexible method selection based on specific tasks. The dLab can control multiple parameters, including temperature, gas pressure, and RF power. These capabilities enable high-dimensional exploration using machine learning techniques beyond Bayesian optimization.

[Scheduling]

The dLab operates robots and modules sequentially. If multiple operations could be carried out in parallel, the throughput of data collection could be further increased. It is important to plan and manage the order in which experiments are conducted using advanced mathematical science and simulations.

[Shared Access and Open Science]

We plan to share the dLab with external users via the Internet, enabling remote autonomous experimentation. The data from this autonomous experiment (SEM and XRD measurement data) are available at https://github.com/Hitosugi24/dLab.

[Data Visualization]

The dLab allows for the collection of vast amounts of data. To make effective use of this data, researchers must be able to grasp its essence. For that purpose, data visualization is essential to aid understanding.

[Digital twin: Integration of Simulation and Reality]

It is a great challenge to recreate real space, from materials synthesis to analytical measurements, in cyberspace. By combining quantum chemical calculations, first-principles calculations, and macroscopic theories, it should be possible to predict the composition and structure of materials with desired properties. Integrating the synthesis process data and the property data collected by the dLab with simulation data enables more advanced materials development.

5. Conclusion

We developed a modular platform that autonomously explores thin-film materials. By adopting the MaiML format, our system enables seamless data collection regardless of the manufacturer or equipment model. We aim to leverage this system to accelerate the development of new materials.

Acknowledgment

We would like to express our sincere gratitude to Professor Kanta Ono of Osaka University for his invaluable support in the development of dLab.

Fundings

This research was supported by the Japan Science and Technology Agency (JST) through the Moonshot R&D Program (MIRAI) [JPMJMI21G2] and the Core Research for Evolutional Science and Technology (CREST) Program [JPMJCR22O4], as well as by Grants-in-Aid for Scientific Research (KAKENHI) from the Japan Society for the Promotion of Science (JSPS) [24K01599]. Additional support was provided by the Open Facility Center at Tokyo Institute of Technology (currently the Core Facility Center at Institute of Science Tokyo) and by the Ministry of Education, Culture, Sports, Science and Technology (MEXT) through the Data Creation and Utilization-Type Materials Research and Development Project, The University of Tokyo "DX-GEM: Research Hub for Electrochemical Materials Toward Maximum Introduction of Renewable Energy" [JPMXP1122712807].

References

(1) N. Ishizuki, R. Shimizu, and T. Hitosugi: *Science and Technology of Advanced Materials: Methods*, **3** (2023), 2197519.

(2) M. Abolhasani and E. Kumacheva: *Nature Synthesis*, **2** (2023), 483–492.

(3) National Academies of Sciences, Engineering, and Medicine.: *Automated Research Workflows for Accelerated Discovery: Closing the Knowledge Discovery Loop*.

(4) B. Burger, P. M. Maffettone, V. V. Gusev, C. M. Aitchison, Y. Bai, X. Wang, X. Li, B. M. Alston, B. Li, R. Clowes, N. Rankin, B. Harris, R. S. Sprick, and A. I. Cooper: *Nature*, **583** (2020), 237–241.

(5) N. J. Szymanski, B. Rendy, Y. Fei, R. E. Kumar, T. He, D. Milsted, M. J. McDermott, M. Gallant, E. D. Cubuk, A. Merchant, H. Kim, A. Jain, C. J. Bartel, K. Persson, Y. Zeng, and G. Ceder: *Nature*, **624** (2023), 86–91.

(6) J. Chen, S. R. Cross, L. J. Miara, J.-J. Cho, Y. Wang, and W. Sun: *Nature Synthesis*, **3** (2024), 606–614.

(7) K. Nishio, A. Aiba, K. Takihara, Y. Suzuki, R. Nakayama, S. Kobayashi, A. Abe, H. Baba, S. Katagiri, K. Omoto, K. Ito, R. Shimizu, and T. Hitosugi: *Digital Discovery*, (2025), in press.

(8) R. Shimizu, S. Kobayashi, Y. Watanabe, Y. Ando, and T. Hitosugi: *APL Mater.*, **8** (2020), 111110.

(9) R. Nakayama, R. Shimizu, T. Haga, T. Kimura, Y. Ando, S. Kobayashi, N. Yasuo, M. Sekijima, and T. Hitosugi: *Science and Technology of Advanced Materials: Methods*, **2** (2022), 119–128.

(10) H. Xu, R. Nakayama, T. Kimura, R. Shimizu, Y. Ando, S. Kobayashi, N. Yasuo, M. Sekijima, and T. Hitosugi: *Science and Technology of Advanced Materials: Methods*, **3** (2023), 2210251.

(11) M. Hanai, R. Ishikawa, M. Kawamura, M. Ohnishi, N. Takenaka, K. Nakamura, D. Matsumura, S. Fujikawa, H. Sakamoto, Y. Ochiai, T. Okane, S. Kuroki, A. Yamada, T. Suzumura, J. Shiomi, K. Taura, Y. Mita, N. Shibata, and Y. Ikuhara: *ARIM-mdx Data System: Towards a Nationwide Data Platform for Materials Science*, 2024 IEEE International Conference on Big Data (BigData), 2024.

(12) X. Ning, I. W. Selesnick, and L. Duval, *Chemom. Intell. Lab. Syst.*, **139** (2014), 156–167.

(13) P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, and J. Bright: *Nat. Methods*, **17** (2020), 261–272.

(14) M. Antaya, K. Cearns, J.S. Preston, J.N. Reimers, and J.R. Dahn: *J. Appl. Phys.*, **76** (1994), 2799–2806.

(15) R. Huang, T. Hitosugi, C. A. J. Fisher, Y. H. Ikuhara, H. Moriwake, H. Oki, and Y. Ikuhara: *Mater. Chem. Phys.*, **133** (2012), 1101–1107.

(16) K. Mizushima, P. Jones, P. Wiseman, and J. B. Goodenough: *Mater. Res. Bull.*, 15 (1980), 783–789.

(17) H. J. Orman and P. J. Wiseman, *Acta Crystallogr. Sec. C*, **40** (1984), 12–14.

(18) T. Hewston and B. Chamberland: *J. Phys. Chem. Solids*, **48** (1987), 97–108.

