# Supplementary Information

## Digital laboratory with modular measurement system and standardized data format

Kazunori Nishio[1*], Akira Aiba[1], Kei Takihara[2], Yota Suzuki[1], Ryo Nakayama[3], Shigeru Kobayashi[3], Akira Abe[4], Haruki Baba[4], Shinichi Katagiri[4], Kazuki Omoto[5], Kazuki Ito[6], Ryota Shimizu[3] and Taro Hitosugi[1,3**]

1 School of Materials and Chemical Technology, Institute of SCIENCE TOKYO, Kanagawa 226-0026, Japan

2. School of Materials and Chemical Technology, Tokyo Institute of Technology, Tokyo 152-8552, Japan

3 Department of Chemistry, The University of Tokyo, Tokyo 113-0033, Japan

4 JEOL Ltd., Tokyo 196-8558, Japan

5 Product Division, Rigaku Corporation, Tokyo 196-8666, Japan

6 X-ray Research Laboratory, Rigaku Corporation, Tokyo 196-8666, Japan

*nishio.k.f0ca@m.isct.ac.jp

**hitosugi@g.ecc.u-tokyo.ac.jp

## Supplemental Materials

### 1. Modularization of Sample Synthesis and Analysis Equipment

The digital laboratory (dLab) is a revolutionary system that interconnects experimental equipment for material synthesis and measurement (Fig. 1). This system offers several benefits alongside a robust data collection and analysis system. We have standardized the sample holders to transport samples between experimental modules using a robot. We use a substrate with a size of 10 mm × 10 mm and 0.5 mm in thickness. Thin-film deposition and various physical property measurements are performed using the sample holder shown in Fig. S1.

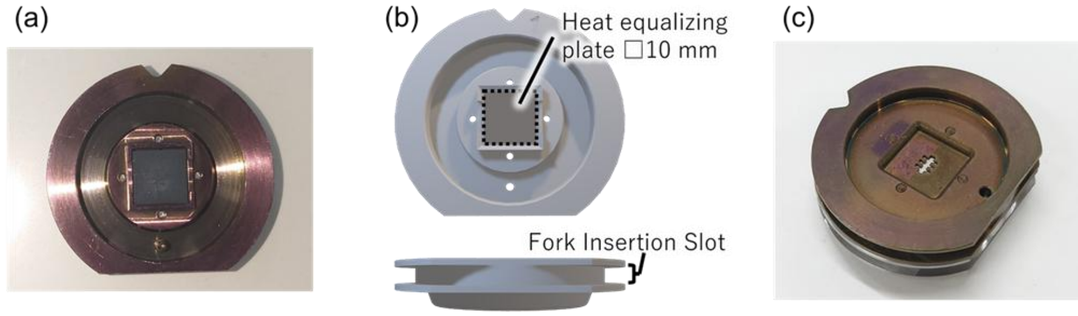


Figure S1: (a) Photograph and (b) schematic of a sample holder for dLab. It can accommodate a substrate up to 10 mm × 10 mm in size in the center. A robot arm with a fork picks up the sample holder and transports a sample between various modular instruments. (c) Photograph shows the installation of a metal mask, which is transported by a robot and placed over the sample holder.

Additionally, in the atmosphere, the robot directly handles the substrate. A video demonstrating the complete automated experimental process is provided in the Supplementary Materials, which includes:

1. Sample holder transportation by the robot.
2. Deposition in the sputtering module.
3. SEM observation and EDS analysis in the SEM module.
4. Raman spectroscopy in the Raman measurement module.
5. Transfer from the vacuum chamber to the atmosphere and sample pickup.
6. XRD measurement in the XRD module.

In the sputtering module, it is possible to pattern thin films using a metal mask. A mask holder equipped with a metal mask can be placed over the sample holder, as shown in Fig. S1(c). The mask and sample holders are manufactured to the same specifications, enabling them to be attached and detached by the JEL robot.

The vacuum chambers comprising the cluster system are continuously evacuated by turbomolecular pumps, leading to vibrations propagating to the connected modules. Therefore, two vibration isolation measures were implemented when integrating the SEM module into the cluster system. Vibration dampers were installed at the connection points between the chamber and the floor to mitigate vibration propagation. Observations of the same sample under identical measurement conditions are presented in Fig. S2(c) and (d) before and after damper installation, respectively. The introduction of vibration dampers effectively reduced noise and enhanced resolution.

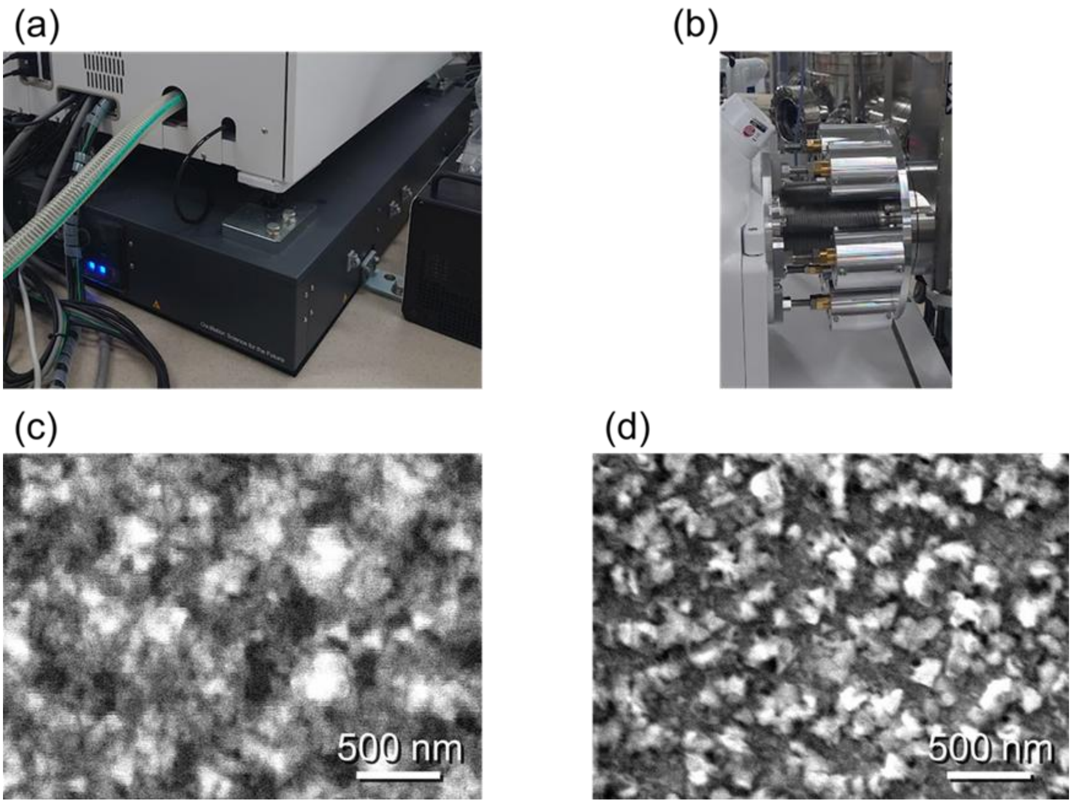


Figure S2: (a) An active damper is installed under the SEM module to reduce vibrations from the floor. (b) A damper was installed between the SEM module and the vacuum chamber to reduce vibrations transmitted from the vacuum chamber. (c, d) SEM images before and after the installation of dampers depict the same sample ($LiCoO_2$(001) thin film on $Al_2O_3$(0001) substrate) observed under identical conditions.

Figure S3 illustrates the time-dependent variation in substrate temperature during synthesis. In our sputtering apparatus, the substrate, placed on an Inconel heat spreader, is heated by an infrared laser, and its temperature is monitored using a radiation thermometer, which serves as the substrate temperature. The emissivity was calibrated based on aluminum (Al) and indium (In) melting points. A sample with Al and In sheets on a glass substrate was introduced into the sputter module, and the emissivity was adjusted to align the displayed temperature with the melting points of Al (660.3°C) and In (156.6°C) when these sheets melted upon heating. The laser output is regulated via PID control to maintain a consistent substrate temperature throughout the synthesis process. Additionally, detailed process logs capture disruptions during synthesis, enabling subsequent analysis and considerations.

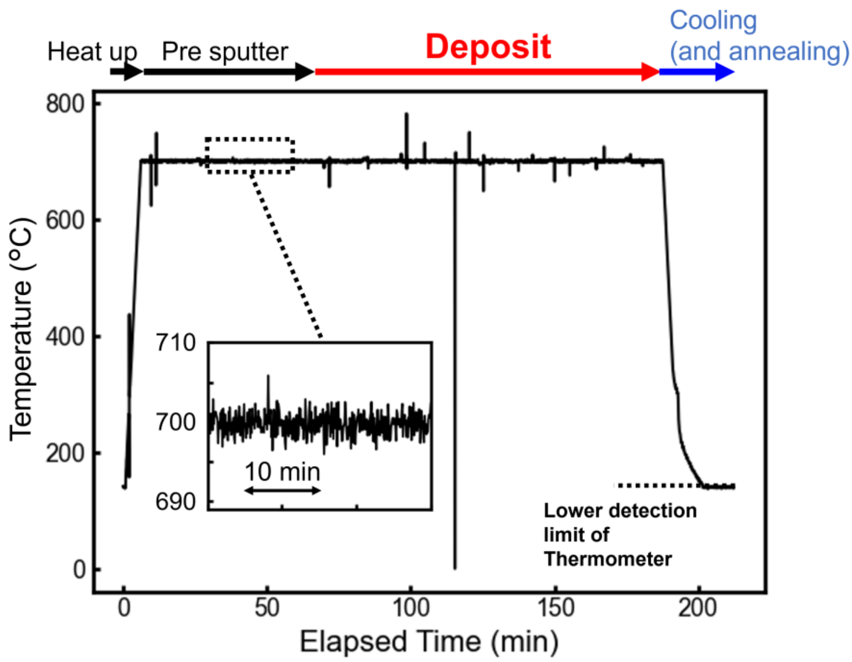


Figure S3: Temporal variation of substrate temperature during automated synthesis. In the sputter module, temperature control is managed by PLC, ensuring temperature fluctuations remain within approximately ±1.4°C of the set value of 700°C. Occasional abrupt changes in temperature readings, attributed to disturbances or electrical noise from the equipment, underscore areas for future improvement.

We introduce an advanced experimental setup integrating an X-ray diffraction (XRD) measurement instrument with a fully automated robotic system. With its high-precision handling capabilities, the robot plays a crucial role in loading samples, positioning them accurately within the XRD instrument, and autonomously initiating data collection sequences. This integration significantly enhances the reproducibility and throughput of XRD measurements by minimizing human error, making the system more reliable and reducing the time required for each analysis. As a result, this setup allows for autonomous material exploration without human intervention.

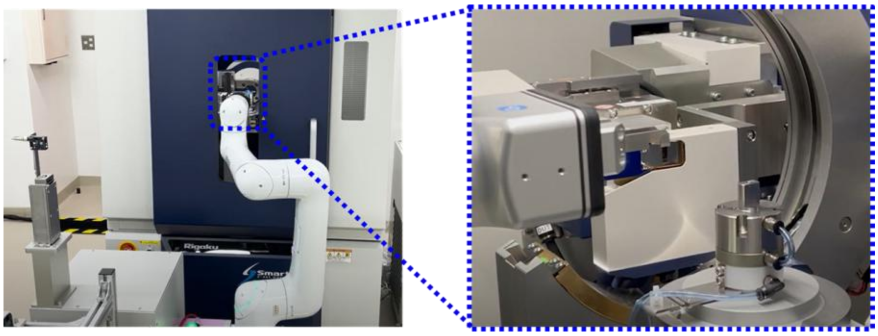

Figure S4: The automated robotic system combined with the X-ray diffraction (XRD) measurement instrument. The robot (COBOTTA PRO) is programmed to handle sample loading and positioning, and the XRD measurement instrument is programmed to automatically start data acquisition, enhancing the precision and efficiency of the XRD analysis. The integrated system allows for high-throughput measurements with minimal human intervention.

We can expand the functionality of our system by integrating new modules for material synthesis and property evaluation. Our vision is to enable these modules to connect in a Plug and Play fashion, allowing for the construction of a digital laboratory by introducing and interconnecting various modules. The specifications for sample holder shapes and communication protocols required for connectivity are available on the Hitosugi Laboratory website [1]. Experimental instruments built to these specifications can be physically integrated into our system and seamlessly operate as part of it. In the future, our goal is to expand the scope of material synthesis and develop automated processes for ceramics, polymers, organic materials, and beyond.

## 2. Software Architecture

In the dLab, experiments are automated by sending tasks to different modules from a central computer. Experimentation proceeds automatically through sequential command transmission based on a standardized communication protocol. Modularization of analytical instruments with protocol-compliant control functions enhances the system's overall expandability, facilitating adaptation to diverse experimental needs.

## 3. Data Collection and Utilization Using a Common Data Format

Data collected from diverse measurement and analysis instruments is uploaded to Amazon AWS via the campus network in MaiML format [2] and aggregated (Fig. 2). This allows cloud-based software to efficiently integrate and analyze the accumulated data, enhancing the capability to leverage experimental results comprehensively.

Previously, researchers had to download data stored on cloud servers to their computers for processing with dedicated software. However, dLab enables direct data uploading from measurement devices to the cloud, facilitating universal access without time or location constraints.

The MaiML format, described in XML, employs tags and attributes within the <maiml> element to define data semantics. As Supplementary Material, we provide an example of observation data for $LiCoO_2$ thin films in MaiML format using SEM. The files are technically described in a predecessor format called xmail but conform closely to MaiML standards. For detailed specifications, refer to references [2] and MaiML format guidelines [3].

MaiML-formatted measurement data files use XML to structure information into four main sections under <maiml> element: <document>, <protocol>, <data>, and <eventlog>. The <document> section includes file metadata like creator, equipment manufacturer (identified through the <vendor> tag, confirming JEOL as the product vendor), and creation timestamp.

The <protocol> section outlines measurement and analysis steps along with default conditions, depicted via Petri net diagrams under the <pnml> element. This section uses <transition> tags for operations, <place> tags for samples and conditions, and <arc> tags for flow connections. In our example, <place> tags detail SEM observation conditions and their connections to <transition> tags representing measurement operations, with <place> tags for results linked as outputs.

Default conditions are defined using templates under <protocol> element, specifying key attributes and values for each element. These templates indicate corresponding <place> tags within the Petri net, ensuring consistent settings across different measurements.

The <data> section details actual measurement operations and collected data. Within <results> under <data> element, specifics such as sample used and conditions employed are recorded with numerical values corresponding to key attributes. These values override defaults set in <protocol> element, ensuring an accurate representation of actual measurements.

The <eventlog> section logs measurement operations, including end times. In our example, a series of measurements concluded at 16:42:34 on June 1, 2023 (Japan time).

MaiML format's flexible data definitions accommodate terms from various equipment manufacturers or users, capturing numerical results and the entire measurement process, including sample preparation and synthesis. This ensures process reproducibility. Additionally, MaiML supports data integrity and encryption, standardized by JIS (Japanese Industrial Standards) in 2024 (JIS K0200).

## 4. Automatic and Autonomous Material Discovery

Machine learning and robotics enable automatic and autonomous experiments. "Automatic experiments" are experiments prescribed by humans, which robots execute repetitively and accurately. On the other hand, "Autonomous experiments" involve computers directing the next experiment, which robots carry out according to instructions. The results are then fed back to the computer to determine subsequent experimental steps. This iterative process, depicted in a closed-loop cycle (Fig. S5), explores the material discovery space until the desired material is synthesized entirely without human intervention.

We previously achieved automatic and autonomous experiments in 2019, reporting on the minimization of electrical resistance of $TiO_2$ thin films [4]. This technology is integrated into dLab system.

We have tailored Bayesian optimization for synthetic experiments. Specifically, we define the process window based on the experimenter's intuition, insights, and experience, then determine suitable hyperparameters through prior simulation [5,6].

In this study, we performed experiments to maximize the intensity ratio of the $LiCoO_2$ 003 and 006 reflections ($I_{003}/I_{006}$), with the substrate temperature ($T_s$) as the parameter. Initially, two samples were prepared at $T_s$ of 200 °C and 750 °C, and the corresponding $I_{003}/I_{006}$ ratios were automatically calculated. These $T_s$ values and respective $I_{003}/I_{006}$ ratios, were registered in the dataset, after which new synthesis conditions were proposed using Bayesian optimization. The dLab automatically conducted the deposition and XRD measurements in a closed-loop manner. The results of the autonomous experiments, performed 25 times by dLab (resulting in a total of 27 data points), are presented in Table S1.

Table S1: The intensity ratio of the $LiCoO_2(001)$ thin-film samples fabricated in the autonomous experiments. If the peaks for the 003 or 006 reflections were not detected, $I_{003}/I_{006}$ was set to 0.

| Data Number | $T_s$ (°C) | $I_{003}/I_{006}$ |
|---|---|---|
| 1 (initial) | 200 | 0 |
| 2 (initial) | 750 | 20.582 |
| 3 | 740 | 20.024 |
| 4 | 610 | 16.569 |
| 5 | 730 | 19.796 |
| 6 | 720 | 18.187 |
| 7 | 710 | 21.228 |
| 8 | 700 | 18.987 |
| 9 | 690 | 20.826 |
| 10 | 680 | 24.428 |
| 11 | 600 | 17.551 |
| 12 | 670 | 24.959 |
| 13 | 390 | 0 |
| 14 | 660 | 26.956 |
| 15 | 650 | 24.625 |
| 16 | 640 | 24.895 |
| 17 | 500 | 0.11337 |
| 18 | 300 | 0 |
| 19 | 630 | 23.447 |
| 20 | 620 | 24.223 |
| 21 | 440 | 0 |
| 22 | 590 | 13.781 |
| 23 | 580 | 0 |
| 24 | 250 | 0 |
| 25 | 340 | 0 |
| 26 | 570 | 0 |
| 27 | 470 | 0 |

It is possible to explore vast exploration spaces and discover new materials through automated and autonomous experimental technology combined with comprehensive materials characterization. Multifaceted evaluation is expected to yield unexpected discoveries, providing researchers with opportunities for serendipitous results—moments of unexpected insight. For instance, a breakthrough was achieved in electrode materials while conducting automated and autonomous experiments to search for solid electrolytes [7]. This approach is anticipated to expand research opportunities that challenge conventional wisdom, facilitating efficient innovation creation.

Additionally, this system can collect numerical data during the synthesis process, such as the time variation of oxygen partial pressure and substrate temperature changes. This data lets us elucidate the correlation between material properties and the synthesis process.

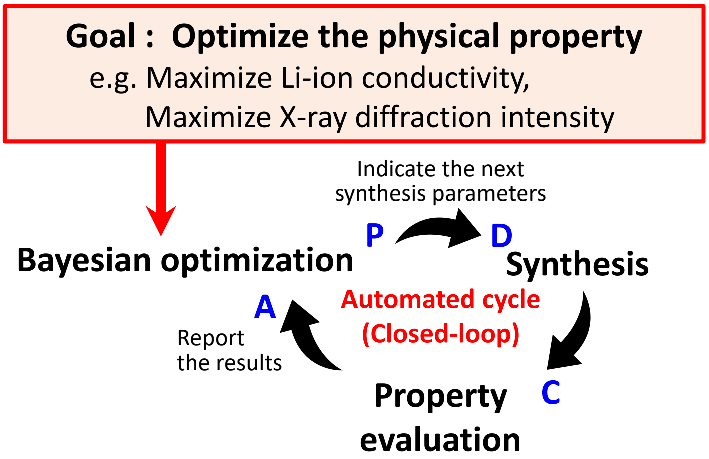


Figure S5: Closed loop. We conduct the synthesis of samples and optimize the physical property measurements fully automatically, feeding back results to determine new synthesis conditions through machine learning. This enables autonomous cycling of the PDCA (Plan, Do, Check, Act) cycle for planning (P), execution (D), property evaluation (C), and estimation of optimal synthesis conditions (A).

## References


[1] Standardization / Collaborations of Researchers, AI and Robots 2023. Available from: https://solid-state-chemistry.jp/ai_robotics/en/standardization.html

[2] Shingo I, Tomo S, takuo Y, *et al*. Standardization of output data format for measurement and analysis instruments. Oyo Buturi. 2023; 92(3): 142-146.

[3] Unified data format support guidelines: Japan Analytical Instruments Manufacture's Association; Available from: https://www.maiml.org/

[4] Shimizu R, Kobayashi S, Watanabe Y, *et al*. Autonomous materials synthesis by machine learning and robotics. APL Materials. 2020; 8 (11).

[5] Xu H, Nakayama R, Kimura T, *et al*. Tuning Bayesian optimization for materials synthesis: simulating two- and three-dimensional cases. Science and Technology of Advanced Materials: Methods. 2023; 3 (1).

[6] Nakayama R, Shimizu R, Haga T, *et al*. Tuning of Bayesian optimization for materials synthesis: simulation of the one-dimensional case. Science and Technology of Advanced Materials: Methods. 2022 2022/12/31;2(1):119-128.

[7] Kobayashi S, Shimizu R, Ando Y, et al. Autonomous Exploration of an Unexpected Electrode Material for Lithium Batteries. ACS Materials Letters. 2023; 5 (10): 2711-2717.