Journal of Crystal Growth 630 (2024) 127598

Contents lists available at ScienceDirect

Journal of Crystal Growth

journal homepage: www.elsevier.com/locate/jcrysgro

Review of machine learning applications for crystal growth research

Kentaro Kutsukake *

Center for Advanced Intelligence Project, RIKEN, 1-4-1 Nihonbashi, Chuo-ku, Tokyo, Japan
Institute of Materials and Systems for Sustainability, Nagoya University, Furocho, Chikusa-ku, Nagoya, Japan

ARTICLE INFO

Keywords:
A1. Machine learning
A1. Optimization
A1. Materials informatics
A1. Process informatics
A1. Computer simulation
A1. Characterization

ABSTRACT

The application of information science and technology has led to a paradigm shift in scientific and technological research and crystal growth is no exception. Various types of application research have been conducted, and research methods that combine real experiments and simulations with information techniques are becoming increasingly complex. In this paper, I focus on the application of information science and technology to the field of crystal growth. In the first half, I discuss the characteristics of process informatics, including applications to crystal growth, from the perspective of how it differs from materials informatics. In the second half, by reviewing various application studies to crystal growth, I aim to highlight the characteristics and discuss future issues.

1. Introduction

Materials informatics (MI), which applies information science to materials science, has made significant progress. In a broad sense, MI covers all the research subjects of materials science including material design, material synthesis, and material characterization from atomic to macroscopic scale [1]. MI was activated by constructing materials databases mainly consisting of density functional theory (DFT) calculation results [2,3]. New and highly functional materials have been searched in a material exploration space consisting of nearly infinite combinations of elements and structures using the database. Faster and more efficient searches have been achieved using information science. In recent years, as the scope of MI has expanded, the MI research field has become more segmented. In a narrow sense, MI often refers to the material search mainly using the DFT calculation database.

In some material cases, material properties strongly depend on experimental process parameters such as temperature, pressure, and time. For these materials, information science has been applied to mainly optimization of process parameters. This is called "Process informatics" (PI) [4] to distinguish from the narrow sense MI. In the initial stages of PI research, simple optimization of process parameters has been the main subject; however, recently, many applications unique to PI have been studied, such as equipment design and parameter importance analysis, as will be reviewed in this paper.

In this paper, I first discuss the characteristics of and difficulties associated with PI from the perspective of its differences with MI. Next, I discuss the challenges and prospects of PI while reviewing research related to crystal growth.

2. Commonalities and differences between MI and PI

Fig. 1 illustrates the differences between MI and PI in terms of the size scale and simulation methods. To highlight the difference between MI and PI, in this paper let me consider MI and PI as follows: MI focuses on the search for materials, i.e., aiming at "what to make,'' and PI focuses on process development, i.e., aiming at "how to make.'' Corresponding to these purposes, in MI, the main interest is the physical properties that originate from structures at the atomic scale, and atomic and molecular simulations, such as DFT calculations, are mainly used for computational simulation to make the database. In contrast, PI targets the macroscopic physical properties depending on the process parameters, including the effects of impurities and crystal defects. Typical computational simulations involve coupling the calculations of heat, fluid, reactions, and electromagnetism to model the effect of process parameters in the fabrication equipment. In addition to simulation data, PI often uses the process data of real experiments. Note that the boundary between MI and PI is not clear, and they are sometimes collectively referred to as "materials informatics" in a broad sense.

Most MI and PI methods involve the creation of machine-learning models from data. Machine learning is a general term used for methods that derive useful laws (machine-learning models) from data. Among various types of machine-learning models, the types that output variable values according to the values of the input variable vectors are frequently used in MI and PI. Table 1 summarizes the typical inputs/

* Address: Center for Advanced Intelligence Project, RIKEN, 1-4-1 Nihonbashi, Chuo-ku, Tokyo, Japan.
E-mail address: kentaro.kutsukake@riken.jp.

https://doi.org/10.1016/j.jcrysgro.2024.127598
Received 9 November 2023; Received in revised form 19 December 2023; Accepted 23 January 2024
Available online 26 January 2024

outputs in MI and PI. In MI, the typical inputs are parameters that define materials, such as element species, composition, and crystal structure, and the typical outputs are material properties, such as electrical and thermal conductivity. On the other hand, in PI, in addition to process conditions, such as temperature, pressure, gas flow rate, and substrate position, the process time may also be an input variable. Equipment-specific parameters, such as configuration, structure, and the material types of equipment components, may also be included. In addition to variables representing the macroscopic material quality, such as electrical and optical properties, defect density, and uniformity, outputs may also include variables related to productivity and production costs, such as growth rate, growth time, and raw material yield.

Thus, the input and output variables used in MI and PI differ, which is directly linked to the difficulty of building a database. In principle, the atomic and molecular simulations used in MI provide the same results for the same parameters and calculation conditions. Therefore, sharing data (data from other groups can also be used) is of great significance, and there is active research on collecting data from the literature and creating databases [2,3,5]. However, PI has many parameters specific to the experimental system and equipment, and offsets due to equipment differences may need to be considered, even for universal parameters, such as temperature and pressure. Furthermore, although the number of parameters is enormous when the equipment configuration is included, the values for only a few parameters are usually described in the literature. Therefore, creating and using PI databases is difficult. However, in recent years, attempts have been made to build process databases for PI applications by limiting the target areas where the parameters can be standardized [3,6], and it is expected that the creation of common databases for PI will become more active in the future.

3. PI applications to crystal growth research

Fig. 2 summarizes the applications of PI in crystal growth grouped by the viewpoints of data, methods, and applications. In this paper, examples of PI research in the field of crystal growth for each group are reviewed.

Two main types of data are used in PI: actual experimental data and computational simulation data. For both data, a properly trained machine-learning model can predict the results of experiments or simulations in a short time and small computational cost. In addition to the prediction of the results, when using actual experimental data, the influence of the parameters is visualized by analyzing the trained machine-learning model, and the importance of the features is obtained. Authors have created a machine-learning model that predicts the interstitial oxygen impurity concentration from process data during single-crystal silicon growth using the Czochralski method [7]. The oxygen impurities dissolved in silicon melt from a quartz crucible and incorporated into a silicon crystal have positive effects on it, such as improved wafer strength and gettering of metal impurities by oxygen precipitates; however, the electrical properties degrade due to crystal defects caused by oxygen precipitates. Therefore, precise concentration control is required for the application of each device. Using a machine-learning model, we visualized the effects of process parameters by virtually changing the process conditions and predicting the results. Park et al. trained a machine-learning model using the borazine flow rate, temperature, and hydrogen gas flow rate as inputs and the crystal domain size as an output to synthesize hexagonal boron nitride using the CVD method [8]. By depicting changes in the domain size in a three-dimensional process parameter space using a machine-learning model, they visualized the influence of each parameter and obtained the process conditions that maximized the domain size. Chou et al. developed a machine-learning model using process conditions such as temperature, chamber pressure, oxygen gas flow rate, and argon gas flow rate as the input and growth rate as the output for the metalorganic chemical vapor phase epitaxy (MOVPE) of β-gallium oxide thin films [9]. Using random forest regression as a machine-learning method, they calculated the feature importance and evaluated the impact of each process condition on the growth rate.

Table 1
Typical inputs and outputs in MI and PI.

	Inputs	Outputs
MI	Element species, composition, crystal structure, etc.	Electrical conductivity, thermal conductivity, etc.
PI	Process conditions, process time, equipment-specific parameters, etc.	Macroscopic material properties, productivity, production cost, etc.

PI using computational simulation data enables the high-speed prediction of simulation results and optimization of growth conditions. In a computational fluid dynamics (CFD) simulation of the solution growth of SiC crystals, Tsunooka et al. developed a machine-learning model to predict the flow and carbon concentration distribution in the solution as parameters of temperature at the crucible corner, crystal rotation speed, crystal diameter, inner diameter of the crucible, solution height, and crystal position [10]. In this growth method, a carbon crucible acts as a carbon source material, and SiC crystals are grown by transporting carbon dissolved in a silicon-based solvent to a SiC seed crystal. It has been reported that prediction using a machine-learning model is 10^7 times faster than that using CFD simulation. Similarly, in a simulation of the solution growth of SiC crystals, Dang et al. developed a machine-learning model to predict the distribution of the crystal growth rate, crucible dissolution rate, and miscellaneous crystal grain growth rate using the heater power, crucible position, crystal rotation speed, and crucible rotation speed as parameters [11]. Furthermore, using the same model, they optimized the parameters using a genetic algorithm to determine the conditions under which the crystal growth, crucible dissolution, and miscellaneous crystal growth rates were large and uniform. By updating the geometry according to the obtained distribution of the crystal growth, crucible dissolution, and miscellaneous

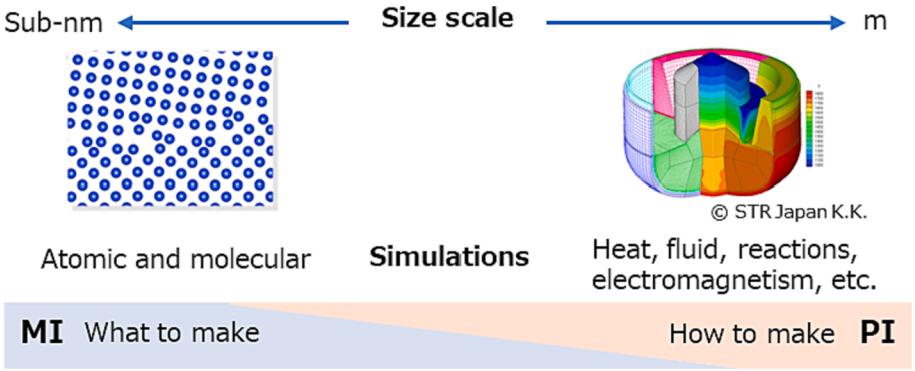

Fig. 1. Differences between MI and PI.

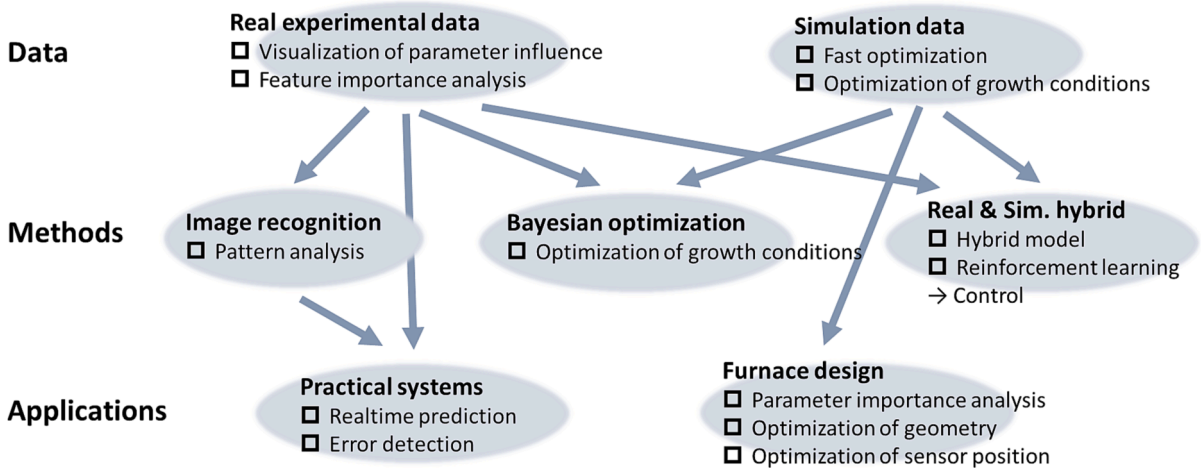

Fig. 2. PI applications to crystal growth research.

crystal growth rates, repeating the creation of a machine-learning model, and optimizing process conditions, process conditions compatible with long-term growth were established.

Bayesian optimization (BO) is a machine-learning method for sequential optimization that can decrease the number of experiments and simulations required to obtain an optimal solution [12]. BO determines the next condition of experiments or simulations by efficiently proceeding with a search based on a probabilistic model. BO is widely used in parameter optimization in various fields, such as computer science and materials science, and is also applied to both experiments and simulations in the field of crystal growth. In Si epitaxial growth, Osada et al. used BO to maximize the growth rate while maintaining crystal quality, such as defect density and film thickness uniformity [13]. In addition to the conditions determined using two types of BO with different crystal quality evaluation items, they adaptively incorporated the conditions proposed based on the expert knowledge of engineers. As a result, the growth rate achieved in short-term optimization was 1.8 times faster than that of conventional conditions.

Attempts have also been made to combine experiments and simulations. Ren et al. constructed a machine-learning model for the growth of single-crystal Si using the Czochralski method, which predicts the crystal diameter from process condition data, such as heater output, crucible moving rate, and crystal pulling rate [14]. They showed that a hybrid model combining a machine-learning model that predicts the crystal growth rate from process conditions and a mathematical model that calculates the crystal diameter from the crystal growth, crucible moving, and crystal pulling rates showed higher prediction accuracy than a model that directly predicts the crystal diameter. Reinforcement learning is a machine-learning method that determines the action to be taken depending on the current state. As reinforcement learning requires significant trial and error, applying a model learned in a simulator to a real environment (referred to as sim-to-real) is a common practice in the field of robot control [15]. Similar attempts have been made to control crystal growth. Ghritli et al. applied reinforcement learning to time-series solid–liquid interface shape control using the temperature gradient and rotation speed as parameters in the vertical Bridgman growth of InGaSb [16]. It was shown that reinforcement learning could improve the interface flatness more than Bayesian optimization for the Bridgman growth case. Tosa et al. applied reinforcement learning to control the diameter of crystals using the floating-zone melting (FZ) method [17]. A model was constructed to predict changes in the melt zone from a small amount of operational data, and an algorithm was proposed that used reinforcement learning to control the melt zone based on the prediction model.

Image recognition is a field in which machine learning is making great progress, and several applications related to crystal growth have also been proposed. Kwoen et al. used machine learning to analyze the reflection high-speed electron diffraction (RHEED) patterns in the molecular beam epitaxy (MBE) of GaAs [18]. A three-class classification model was constructed for the 2×4 and 4×4 surface structures of GaAs, and the quantum dot structure of InAs and used to identify the surface structure.

One of the targets of machine-learning applications is the implementation of practical systems. Anjum et al. used a machine-learning model to analyze the RHEED patterns in MBE using GaN substrates [19]. They constructed a model to predict whether or not the substrate was rotating, and built a system to detect anomalous substrate rotation. Authoers constructed a real-time prediction system using a machine-learning model that predicted the oxygen impurity concentration during single-crystal Si growth using the Czochralski method [20]. By inputting the process data obtained from the growth furnace into a machine-learning model in real time, it is possible to predict the oxygen impurity concentration at the current solid–liquid interface position.

Equipment design is an important issue in practical crystal growth, where the intuition and experience of designers are required owing to the high degree of freedom, which means a large number of parameters. Recently, several attempts have been made to utilize machine learning to support equipment design. In a simulation of single-crystal Ge growth using the Czochralski method, Dropka et al. created a machine-learning model that parameterized the internal structure and predicted the solid–liquid interface shape, crystal growth rate, and temperature gradient [21]. They aimed to provide suggestions to equipment designers by visualizing the influence of each parameter using a decision tree as a machine-learning method. Boucetta et al. used machine learning to optimize the position of a temperature sensor to efficiently evaluate the temperature distribution inside a crucible using simulation data from a Bridgman-type multicrystalline Si crystal growth furnace [22]. A dataset was developed by extracting the temperature distribution on the side of the crucible from the simulation results at various setting temperatures and crucible positions, assuming actual crystal growth, and machine-learning models were developed to predict the overall temperature distribution from those at the limited sensor positions on the side of the crucible. Machine-learning models were created for all combinations of sensor positions, and the sensor position with the highest prediction accuracy was determined to be the optimal position.

4. Conclusion

Finally, I discuss the challenges and prospects of PI in terms of database, optimization, and simulations. As mentioned in the section on the differences with MI above, one of the challenges of PI is the difficulty of creating a database. Here, I consider the difficulty in obtaining high-

quality data. In actual experiments, there is always the problem of noise. For example, in the case of temperature measurements using thermocouples, in addition to the temperature accuracy of the thermocouple, individual differences and positional deviations when replacing the thermocouples become noise factors when acquiring long-term data. Additionally, even replacing one equipment component changes the environment inside the furnace, which consequently changes the meaning of the previously measured temperature and the temperature to be measured. For example, replacing an old thermal insulator with a new one often leads to a temperature distribution change in a Bridgeman-type furnace. It is necessary to consider this as noise or change. In addition, in the case of simulation, it is necessary to be aware that the quality of the data changes depending on the settings of the calculation conditions, such as the type of mathematical model and tolerance for the convergence calculation. In the future, there is no doubt that the trend of database creation will expand in PI and MI; however, quantifying and guaranteeing the quality of the data remains an issue.

As mentioned above, Bayesian optimization can obtain appropriate experimental conditions to achieve high material properties with a small number of experiments. However, from the perspective of cost for real experiments, the number of experiments is still large. For example, in the case of Bayesian optimization of reference 13, approximately 300 experiments were required to optimize 11 process parameters. To further reduce the number of experiments, it will be necessary to develop advanced Bayesian optimization methods such as adaptive parameter reduction based on feature importance analysis and optimization combined with simulation. Hyperparameter tuning using the knowledge of scientists is also a promising solution [23].

As mentioned above, crystal growth simulations that model the conditions inside the furnace are important for PI. The problem with simulation is the discrepancy with actual experimental results. This difference is due to the simplification of the simulation model and differences in physical properties. The former is generally determined based on the compromise between accuracy and computational cost. The latter problem has been approached through data assimilation. In recent years, physics informed neural networks (PINNs), which applies the neural network framework directly to solve the governing equations instead of the simulation solver, has been attracting attention [24,25]. In addition, PINNs application for data assimilation is also being considered [26]. PINNs may solve the problem of simulation errors in PI for crystal growth.

In this paper, I highlighted the characteristics of PI, focusing on its application in the field of crystal growth, by discussing its differences from MI and examining various examples. The characteristics and application development of PI described here are common to many fields, not just crystal growth. These discussions will help in making use of PI.

CRediT authorship contribution statement

Kentaro Kutsukake: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Project administration, Visualization, Writing – original draft, Writing – review & editing.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Kentaro Kutsukake reports financial support was provided by New Energy and Industrial Technology Development Organization. Kentaro Kutsukake reports a relationship with RIKEN Center for Advanced Intelligence Project that includes: board membership, employment, and funding grants. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

Acknowledgements

Part of this study was based on the results obtained from projects JPNP20006 and JPNP14004 subsidized by the New Energy and Industrial Technology Development Organization (NEDO). The author acknowledges Toru Ujihara, Shunta Harada, Masaki Takaishi, Shota Seki, Ichiro Takeuchi, Kota Matsui, Noritaka Usami, Hiroaki Kudo, and Takuto Kojima of Nagoya University and Shion Takeno of RIKEN for their constructive discussions.

References

[1] J.M. Rickman, T. Lookman, S.V. Kalinin, Materials informatics: From the atomic-level to the continuum, Acta Mater. 168 (2019) 473–510, https://doi.org/10.1016/j.actamat.2019.01.051.
[2] Materials project. https://legacy.materialsproject.org/.
[3] NOMAD. https://nomad-lab.eu/nomad-lab/.
[4] M. Fujii, Significance of materials informatics and the development of new materials, JSAP Review 2022 (2022) 220416, https://doi.org/10.11470/jsaprev.220416.
[5] Starrydata. https://www.starrydata.org/.
[6] MIP Platform. https://unit.aist.go.jp/dmc/platform/MPI/.
[7] K. Kutsukake, Y. Nagai, H. Banba, Virtual experiments of Czochralski growth of silicon using machine learning: Influence of processing parameters on interstitial oxygen concentration, J. Cryst. Growth 584 (2022) 126580, https://doi.org/10.1016/j.jcrysgro.2022.126580.
[8] J.H. Park, A.Y. Lu, M.M. Tavakoli, N.Y. Kim, M.H. Chiu, H. Liu, T. Zhang, Z. Wang, J. Wang, L.G.P. Martins, Z. Luo, M. Chi, J. Miao, J. Kong, Revealing variable dependences in hexagonal boron nitride synthesis via machine learning, Nano Lett. 23 (2023) 4741–4748, https://doi.org/10.1021/acs.nanolett.2c04624.
[9] T.S. Chou, S.B. Anooz, R. Grüneberg, N. Dropka, W. Miller, T.T.V. Tran, J. Rehm, M. Albrecht, A. Popp, Machine learning supported analysis of MOVPE grown β-Ga_2O_3 thin films on sapphire, J. Cryst. Growth. 592 (2022) 126737, https://doi.org/10.1016/j.jcrysgro.2022.126737.
[10] Y. Tsunooka, N. Kokubo, G. Hatasa, S. Harada, M. Tagawa, T. Ujihara, High-speed prediction of computational fluid dynamics simulation in crystal growth, CrstEngComm 20 (2018) 6546–6550, https://doi.org/10.1039/C8CE00977E.
[11] Y. Dang, C. Zhu, M. Ikumi, M. Takaishi, W. Yu, W. Huang, X. Liu, K. Kutsukake, S. Harada, M. Tagawa, T. Ujihara, Adaptive process control for crystal growth using machine learning for high-speed prediction: Application to SiC solution growth, CrstEngComm 23 (2021) 1982–1990, https://doi.org/10.1039/D0CE01824D.
[12] B. Shahriari, K. Swersky, Z. Wang, R.P. Adams, N. De Freitas, Taking the human out of the loop: A review of Bayesian optimization, Proc. IEEE. 104 (2016) 148–175, https://doi.org/10.1109/JPROC.2015.2494218.
[13] K. Osada, K. Kutsukake, J. Yamamoto, S. Yamashita, T. Kodera, Y. Nagai, T. Horikawa, K. Matsui, I. Takeuchi, T. Ujihara, Adaptive Bayesian optimization for epitaxial growth of Si thin films under various constraints, Mater. Today Commun. 25 (2020) 101538, https://doi.org/10.1016/j.mtcomm.2020.101538.
[14] J.C. Ren, D. Liu, Y. Wan, Data-driven and mechanism-based hybrid model for semiconductor silicon monocrystalline quality prediction in the Czochralski process, IEEE Trans. Semicond. Manuf. 35 (2022) 658–669, https://doi.org/10.1109/TSM.2022.3202610.
[15] W. Zhao, J.P. Queralta, T. Westerlund, Sim-to-Real Transfer in Deep Reinforcement Learning for Robotics: A Survey IEEE Symposium Series on Computational Intelligence, ACT, Canberra, Australia, 2020, pp. 737–744. Doi: 10.1109/SSCI47803.2020.9308468.
[16] R. Ghritli, Y. Okano, Y. Inatomi, S. Dost, Control of growth interface shape during InGaSb growth by vertical gradient freezing under microgravity, and optimization using machine learning, Jpn. J. Appl. Phys. 61 (2022) 115502, https://doi.org/10.35848/1347-4065/ac99c2.
[17] Y. Tosa, R. Omae, R. Matsumoto, S. Sumitani, S. Harada, Data-driven automated control algorithm for floating-zone crystal growth derived by reinforcement learning, Sci. Rep. 13 (2023) 7517, https://doi.org/10.1038/s41598-023-34732-5.
[18] J. Kwoen, Y. Arakawa, Multiclass classification of reflection high-energy electron diffraction patterns using deep learning, J. Cryst. Growth. 593 (2022) 126780, https://doi.org/10.1016/j.jcrysgro.2022.126780.
[19] S. Anjum, H.Y. Lee, H.K. Noh, Rotation error detection of gallium nitride (GaN) substrate in MBE utilizing ensemble learning, Cryst. Growth Des. 23 (2023) 4138–4146, https://doi.org/10.1021/acs.cgd.2c01544.
[20] K. Kutsukake, Y. Nagai, T. Horikawa, H. Banba, Real-time prediction of interstitial oxygen concentration in Czochralski silicon using machine learning, Appl. Phys. Express. 13 (2020) 125502, https://doi.org/10.35848/1882-0786/abc6ec.
[21] N. Dropka, X. Tang, G.K. Chappa, M. Holena, Smart design of Cz-Ge crystal growth furnace and process, Crystals 12 (2022) 1764, https://doi.org/10.3390/cryst12121764.

[22] A. Boucetta, K. Kutsukake, T. Kojima, H. Kudo, T. Matsumoto, N. Usami, Application of artificial neural network to optimize sensor positions for accurate monitoring: An example with thermocouples in a crystal growth furnace, Appl. Phys. Express. 12 (2019) 125503, https://doi.org/10.7567/1882-0786/ab52a9.
[23] H. Xu, R. Nakayama, T. Kimura, R. Shimizu, Y. Ando, S. Kobayashi, N. Yasuo, M. Sekijima, T. Hitosugi, Tuning Bayesian optimization for materials synthesis: simulating two- and three-dimensional cases, Sci. Technol. Adv. Mater.: Methods 3 (2023) 1, https://doi.org/10.1080/27660400.2023.2210251.
[24] M. Raissi, P. Perdikaris, G.E. Karniadakis, Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations, J. Comput. Phys. 378 (2019) 686–707, https://doi.org/10.1016/j.jcp.2018.10.045.
[25] Y. Takehara, Y. Okano, S. Dost, Fast Prediction of Transport Structures in the Melt by Physics Informed Neural Networks during 'VMCz' Crystal Growth of Silicon, J. Chem. Eng. Jpn. 56 (2023) 1, https://doi.org/10.1080/00219592.2023.2236656.
[26] Q. He, D. Barajas-Solano, G. Tartakovsky, A.M. Tartakovsky, Physics-informed neural networks for multiphysics data assimilation with application to subsurface transport, Adv. Water Resour. 141 (2020) 103610, https://doi.org/10.1016/j.advwatres.2020.103610.