

# 材料プロセスインフォマティクスの課題と展望

## — 結晶成長応用を中心に —

Future Issues of Materials Process Informatics

**Key-words** : Process informatics, Crystal growth, Machine learning, Optimizations

### 沓掛 健太郎

Kentaro KUTSUKAKE

(Center for Advanced Intelligence Project, RIKEN)

#### 1. はじめに

材料科学に情報科学を応用したマテリアルズインフォマティクス (MI) が大きく発展し、物質探索の高速化・効率化により、さまざまな新物質・高機能物質が開発されている。加えて近年は、材料プロセスインフォマティクス (PI) も着実な発展を遂げ、材料の作製・加工プロセスに対しても情報科学を用いたさまざまな応用がなされている。本稿では、材料プロセスのうち結晶成長への応用を中心に、まずPIの概要として、MIとの違いの観点からPIの特徴や難しさを議論する。次いで、結晶成長関連の研究事例をオーバービューしながら、PIの課題と展望を述べる。

#### 2. MIとPIの共通点と相違点

図1にMIとPIの相違を、サイズスケール、シミュレーション手法、実験手法の観点でまとめる。MIは「何を作るか」という物質・素材探索に、PIは「どう作るか」というプロセス開発に主眼が置かれる。これらの目的に対応してMIでは、原子スケールでの構造に

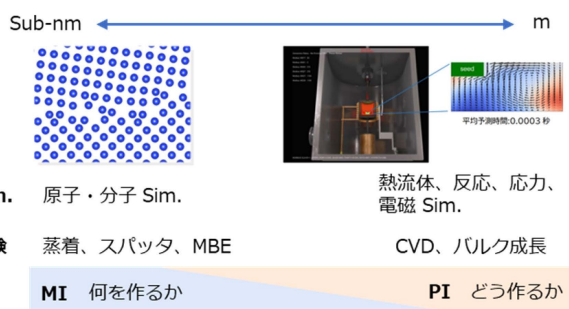


図1 MIとPIの相違

起源をもつ物性が興味の対象となり、シミュレーションでは原子・分子シミュレーションが主に用いられる。また実験では、蒸着、スパッタ、分子線エピタキシー (MBE) など、元素種・組成が主要なパラメータで、プロセス条件の影響が小さい手法が用いられることが多い。一方、PIでは、不純物や結晶欠陥の影響を含むマクロな物性が対象となり、シミュレーションでは作製装置全体をモデル化するような熱、流体、反応、電磁などを連成したシミュレーションが用いられる。また実験では、気体や液体の流れの制御が重要となる化学気相堆積法 (CVD) やバルク結晶成長が対象となることが多い。ただし、MIとPIの境界は明確ではなく、またMIとPIをまとめて広義で“マテリアルズインフォマティクス”と呼ばれることもある。

多くのMI/PIでは、データから機械学習モデルを作成する。機械学習は、データから有用な法則 (機械学習モデル) を求める手法の総称であり、その法則にはさまざまなタイプがあるが、入力変数ベクトルの値に応じた変数値を出力する形式がMI/PIでは頻繁に用いられる。表1にMIおよびPIでの入出力の典型をまとめる。MIでは元素種、組成、結晶構造といった物質を規定するパラメータが入力となり、バンドギャップなどの材料物性が出力となる。一方PIでは、温度、圧力、ガス流量、基板位置などのプロセス条件に加えて、時間も入力変数となる場合がある。また装置の種類・構成・構造、部材種類といった装置固有のパラメータが加わる場合もある。出力には、欠陥密度、均一性などのマクロな材料品質を表す変数に加えて、成長速度・成長時間、原料歩留りといった生産性・コストに関わる変数も用いられる場合がある。

このように、MIとPIでは用いられる入力・出力変数が異なるが、この違いはデータベース作成の難しさに直結する。MIで用いられる原子・分子シミュレーションは、計算条件の違いによる微小な結果の違いはあるが、原理的には同じ入力に対しては同じ出力が得られるはずである。実験も同様に、制御パラメータが少なく、普遍性が高い手法が用いられる。したがって、データの共有に大きな意味があり (他人のデータも使える)、文献からデータを収集してデータベースを作

表1 MI/PIの機械学習モデルの典型的な入出力

	入力	出力
MI	元素種, 組成, 結晶構造	材料物性
PI	プロセス条件 時間 装置変数	材料品質 生産性

成する研究も活発に行われている<sup>1)</sup>。一方PIは、実験系や装置固有のパラメータも多く、また温度や圧力といった普遍的なパラメータに対しても装置の違いによるオフセットを考慮しなければならない場合もある。また装置構成まで含めるとパラメータ数は膨大であるが、文献に値が記載されているのは通常一部のパラメータのみである。したがってPIでは、データベース作成とその利用には困難が伴う。しかし近年、パラメータの共通化が可能な範囲を対象を絞ってPI向けにプロセスデータベースの構築を進める試みも始まっており<sup>2),3)</sup>、今後はPI用の共通データベース作成もますます活発化すると予想される。

### 3. 結晶成長への応用

図2に、Web of Science<sup>4)</sup>にて、「machine learning」と「crystal growth」のキーワードでヒットした論文数の年ごとの推移を示す。2012年以前は、該当する論文は1件もなく、2013年に初めて論文が登場した。その後、2017年ごろから急速に論文数が増加し、2023年現在まで右肩上がりの増加が続いている。このような傾向は結晶成長分野に限らず、多くの分野で同じであると予想されるが、材料科学の中でも、結晶成長というプロセス寄りの分野で論文数が急増していることは、PI研究の広がりを示す証左であろう。

図3に結晶成長分野へのPI展開をまとめる。PIに用いられるデータには、実際の実験データとシミュレーションデータの大きく2種類がある。以下では、図3のまとめりにごとに、結晶成長分野におけるPI研究事例をレビューする。

実際の実験データを用いたPIでは、パラメータ影響の可視化、特徴量重要度の解析が行われている。著者らは、チョクラルスキー法による単結晶シリコンの育成において、プロセスデータから酸素不純物濃度を

予測する機械学習モデルを作成した。石英ルツボからシリコン融液に溶解し結晶に取り込まれる酸素不純物は、シリコン結晶中において、強度向上や酸素析出物による金属不純物のゲッタリングといったプラスの効果と、酸素析出物に起因した結晶欠陥による電気的特性の低下といったマイナスの効果があり、用途ごとに精密な濃度制御が要求される。機械学習モデルを用いて、仮想的にプロセス条件を変化させて結果を予測することで、変化させたプロセスパラメータの影響を可視化した<sup>5)</sup>。Parkらは、CVD法による六方晶窒化ホウ素の合成において、ボラジン流量、温度、水素ガス流量を入力、結晶ドメインサイズを出力とする機械学習モデルを作成した。機械学習モデルを用いて3次元のプロセスパラメータ空間におけるドメインサイズの変化を描くことで、各パラメータの影響を可視化し、またドメインサイズを最大化するプロセス条件を取得した<sup>6)</sup>。Chouらは、 $\beta$ -酸化ガリウム薄膜の有機金属化学気相エピタキシー(MOVPE)において、温度、チャンバ圧力、酸素ガス流量、アルゴンガス流量などのプロセス条件を入力、成長速度を出力とした機械学習モデルを作成した。ここで機械学習手法としてランダムフォレスト回帰を用いることで、特徴量重要度を算出し、各プロセス条件の成長速度への影響の大きさを評価した<sup>7)</sup>。

シミュレーションデータを用いたPIでは、シミュレーション結果の高速予測、成長条件の最適化が行われている。角岡らは、SiC結晶の溶液成長の熱流体シミュレーションにおいて、境界条件の温度、結晶回転速度、結晶径、炭素ルツボ内径、溶液高さ、結晶位置をパラメータとして、炭素ルツボ内の炭素過飽和度および流れの分布を予測する機械学習モデルを作成した。この成長法では、炭素ルツボは炭素原料を兼ね、シリコン系溶媒に溶解した炭素が種結晶に輸送されることでSiC結晶が成長する。機械学習モデルによる予測により、シミュレーションと比較して $10^7$ 倍の高速化が報告され

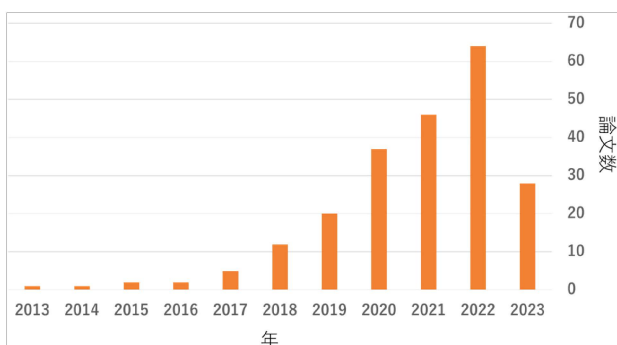


図2 機械学習 & 結晶成長に関連した論文数の推移。2023年は7月までの積算。

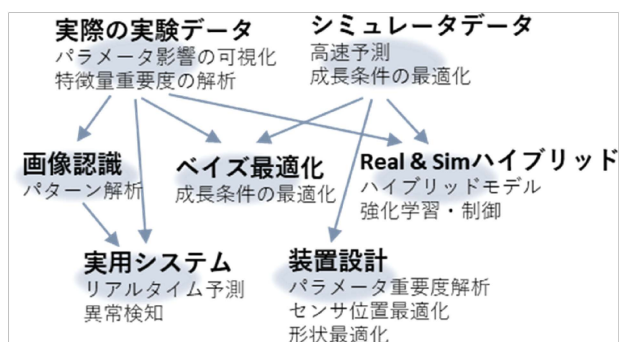


図3 結晶成長分野へのPI展開

ている<sup>8)</sup>。Dangらは、同じくSiC結晶の溶液成長のシミュレーションにおいて、ヒータ出力、炭素ルツボ位置、結晶回転速度、炭素ルツボ回転速度を入力として、結晶成長速度、ルツボ溶解速度および雑晶成長速度の分布を予測する機械学習モデルを作成した。さらに、同モデルを用いて、結晶成長速度が大きくかつ均一、坩堝溶解速度が均一、雑晶成長速度が均一となる条件を遺伝的アルゴリズムによって求めた。得られた結晶成長速度、ルツボ溶解速度、雑晶成長速度の分布にしたがって、ジオメトリを更新し、再び機械学習モデルの作成とプロセス条件の最適化を繰り返すことで、長時間成長に対応したプロセス条件シーケンスを得た<sup>9)</sup>。

ベイズ最適化は、逐次最適化のための機械学習手法であり、確率モデルに基づいて効率的に探索を進めることで最適解を得るのに必要な実験・シミュレーション回数を少なくすることができる<sup>10)</sup>。結晶成長分野においても、実験、シミュレーションの両方に適用されている。長田らは実際のSiエピタキシャル成長において、ベイズ最適化を用いて、欠陥密度や膜厚均一性といった結晶品質を維持しつつ、成長速度を最大化した。ここでは、結晶品質評価項目を変えた2種類のベイズ最適化による条件提案に加えて、エンジニアの専門知識に基づく条件提案を適応的に取り入れる実践的な最適化を提案し、これを用いて短期間で1.8倍の成長速度を達成した<sup>11)</sup>。

実際の実験とシミュレーションを組み合わせる試みも行われている。Renらは、チョクラルスキー法による単結晶Siの育成において、ヒータ出力、石英ルツボ上昇速度、結晶引き上げ速度のプロセス条件データから結晶直径を予測する機械学習モデルを構築した。このとき、結晶直径を直接予測するモデルよりも、プロセス条件データから結晶成長速度を予測する機械学習モデルと、結晶成長速度、石英ルツボ上昇速度、結晶引き上げ速度から結晶径を計算する数理モデルとを組み合わせるハイブリッドモデルの方が、予測精度が高いことを示した<sup>12)</sup>。また強化学習は、現在の状態に応じてどのような行動をとるべきかを決定する機械学習手法の一つである。学習には多くの試行錯誤が必要であるため、シミュレータで学習した内容を現実環境に適用すること（Sim-to-Realと言われる）がロボット制御の分野ではよく行われている<sup>13)</sup>。同様の試みが結晶成長制御に対しても検討されている。GhritliらはInGaSbの垂直ブリッジマン成長において、温度勾配と回転速度をパラメータとした時系列の固液界面形状制御に強化学習を応用した。この事例では、ベイズ最適化よりも強化学習の方がより界面平坦性を向上

できることが示された<sup>14)</sup>。土佐らは、浮遊帯域溶融（FZ）法による結晶育成において、口径制御に強化学習を応用した。限られた操業データから溶融帯の変化を推定するモデルを構築し、このモデルに基づいて強化学習を用いて制御するアルゴリズムが提案された<sup>15)</sup>。

画像認識は、機械学習が大きく発展している分野の一つであるが、結晶成長に関連した応用もいくつか提案されている。Kwoenらは反射高速電子線回折法（RHEED）パターンの解析に、機械学習を用いた。GaAsの2×4表面構造、GaAsの4×4表面構造、InAsの量子ドット構造の3クラス分類モデルを構築し、表面構造の特定に利用した<sup>16)</sup>。

実用システムとして利用する報告もいくつかなされている。Anjumらは、GaN基板を用いたMBEにおいて、同様にRHEEDパターンの解析に機械学習モデルを用いた。このとき、基板が回転しているかどうかを判別するモデルを作成し、基板回転の異常を検知するシステムを構築した<sup>17)</sup>。また著者らは、上で述べたチョクラルスキー法による単結晶Siの育成において酸素不純物濃度を予測する機械学習モデルを用いて、リアルタイム予測システムを構築した。育成装置から取得したプロセスデータをリアルタイムで機械学習モデルに入力することで、現在の固液界面位置における酸素不純物濃度を予測できる<sup>18)</sup>。

装置デザインは、自由度が高く、すなわちパラメータが膨大であり、熟練設計者の勘と経験が大きくものを言う技術分野であるが、機械学習を活用する試みもなされている。Dropkaらは、チョクラルスキー法による単結晶Ge育成のシミュレーションにおいて、結晶引き上げ速度、結晶回転速度などのプロセスパラメータに加えて、反射板の角度と位置、ヒータの位置と長さなどの形式で炉内構造をパラメータ化し、固液界面形状、結晶成長速度、温度勾配を予測する機械学習モデルを作成した。このとき、機械学習手法として決定木を用いて各パラメータの影響を可視化することで、装置設計者に示唆を与えることを企図した<sup>19)</sup>。Boucettaらは、ブリッジマン型の多結晶Si結晶成長炉のシミュレーションにおいて、石英ルツボ内温度分布を効率よく測定する温度センサ位置の最適化を、機械学習を用いて行った。実際の結晶成長を想定したさまざまな温度、石英ルツボ位置でのシミュレーション結果から、石英ルツボ側面の温度分布を抽出し、石英ルツボ側面数か所のセンサ位置の温度から全体の温度分布を予測する機械学習モデルを、すべてのセンサ位置の組合せについて作成し、最も高い予測精度を与えるセンサ位置が最適な位置であるとした<sup>20)</sup>。



## 4. 課題と展望

最後にまとめて代えて、PIの課題と展望を述べる。PIの課題の一つはデータベース作成の難しさである。上記のMIとの相違の項ではデータ共有の難しさを議論したが、ここでは良質なデータを取得することの難しさを考えたい。実際の実験では、ノイズの問題が常に付きまとう。たとえば、熱電対による温度計測の場合も、熱電対自身の温度精度に加えて、付け替えた際の個体差や位置ずれも、長期間データを取得する際のノイズ要因となる。加えて、装置の内の部材を一つ入れ替えただけでも装置内環境が変わってしまい、それまで計測していた温度とこれから計測する温度の意味が変わってしまうことがあり、このような影響もノイズもしくは変化として、考慮する必要がある。またシミュレーションの場合も、数理モデルの種類や収束計算の許容誤差など、計算条件の設定によってデータの質が変わってしまうことには注意が必要である。今後、PIにおいてもデータベース化の流れが拡大していくことは間違いないと予想されるが、データの質をどのように数値化して保証・担保するかは今後の課題である。

本稿では、結晶成長分野への応用を中心に、PIの特徴を、MIとの相違の観点から議論すること、さまざまな事例を見ることを通して、浮かび上がらせた。ここで述べたPIの特徴や応用展開は、結晶成長分野に限らず多くの分野に共通する。本稿が読者諸兄のPI活用の一助になれば幸いである。

**謝辞** 本稿を執筆するにあたってお世話になった方々に深く感謝致します。

### 文 献

- 1) Starrydata, <https://www.starrydata.org/>
- 2) NOMAD, <https://nomad-lab.eu/nomad-lab/>
- 3) MIP Platform, <https://unit.aist.go.jp/dmc/platform/MPI/>
- 4) Web of Science, <https://clarivate.com/ja/solutions/web-of-science/>
- 5) K. Kutsukake, Y. Nagai and H. Banba, "Virtual experiments of Czochralski growth of silicon using machine learning: Influence of processing parameters on interstitial oxygen concentration", *J Cryst Growth*, **584**, 126580 (2022).
- 6) 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 and J. Kong, "Revealing Variable Dependences in Hexagonal Boron Nitride Synthesis via Machine Learning", *Nano Lett*, **23**, 4741-4748 (2023).
- 7) T. S. Chou, S. Bin Anooz, R. Grüneberg, N. Dropka, W. Miller, T. T. V. Tran, J. Rehm, M. Albrecht and A. Popp, "Machine learning supported analysis of MOVPE grown  $\beta$ -Ga<sub>2</sub>O<sub>3</sub> thin films on sapphire", *J Cryst Growth*, **592**, 126737 (2022).
- 8) Y. Tsunooka, N. Kokubo, G. Hatasa, S. Harada, M. Tagawa and T. Ujihara, "High-speed prediction of computational fluid dynamics simulation in crystal growth", *CrystEngComm*, **20**[41], 6546-6550 (2018).
- 9) Y. Dang, C. Zhu, M. Ikumi, M. Takaishi, W. Yu, W. Huang, X. Liu, K. Kutsukake, S. Harada, M. Tagawa and T. Ujihara, "Adaptive process control for crystal growth using machine learning for high-speed prediction: application to SiC solution growth", *CrystEngComm*, **23**[9], 1982-1990 (2021).
- 10) 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*, **104**[1], 148-175 (2016).
- 11) K. Osada, K. Kutsukake, J. Yamamoto, S. Yamashita, T. Kodaera, Y. Nagai, T. Horikawa, K. Matsui, I. Takeuchi and T. Ujihara, "Adaptive Bayesian optimization for epitaxial growth of Si thin films under various constraints", *Mater Today Commun*, **25**, 101538 (2020).
- 12) J. C. Ren, D. Liu and Y. Wan, "Data-Driven and Mechanism-Based Hybrid Model for Semiconductor Silicon Monocrystalline Quality Prediction in the Czochralski Process", *IEEE Transactions on Semiconductor Manufacturing*, **35**[4], 658-669 (2022).
- 13) W. Zhao, J. P. Queralta and T. Westerlund, "Sim-to-Real Transfer in Deep Reinforcement Learning for Robotics: a Survey", 2020 IEEE Symposium Series on Computational Intelligence (SSCI), Canberra, ACT, Australia, 2020, pp. 737-744.
- 14) R. Ghritli, Y. Okano, Y. Inatomi and 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**[11], 115502 (2022).
- 15) Y. Tosa, R. Omae, R. Matsumoto, S. Sumitani and S. Harada, "Data-driven automated control algorithm for floating-zone crystal growth derived by reinforcement learning", *Sci Rep*, **13**[1], 7517 (2023).
- 16) J. Kwoen and Y. Arakawa, "Multiclass classification of reflection high-energy electron diffraction patterns using deep learning", *J Cryst Growth*, **593**, 126780 (2022).
- 17) S. Anjum, H. Y. Lee and H. K. Noh, "Rotation Error Detection of Gallium Nitride (GaN) Substrate in MBE Utilizing Ensemble Learning", *Cryst Growth Des*, **23**[6], 4138-4146 (2023).
- 18) K. Kutsukake, Y. Nagai, T. Horikawa and H. Banba, "Real-time prediction of interstitial oxygen concentration in Czochralski silicon using machine learning", *Applied Physics Express*, **13**[12], 125502 (2020).
- 19) N. Dropka, X. Tang, G. K. Chappa and M. Holena, "Smart Design of Cz-Ge Crystal Growth Furnace and Process", *Crystals (Basel)*, **12**[12], 1764 (2022).
- 20) A. Boucetta, K. Kutsukake, T. Kojima, H. Kudo, T. Matsumoto and N. Usami, *Appl. Phys. Express*, **12**[12], 125503-1-5 (2019).

### 筆者紹介

沓掛 健太郎 (くつかけ けんたろう)



国立研究開発法人理化学研究所革新知能統合研究センター研究員。名古屋大学未来材料・システム研究所客員准教授。東北大学特任教授(客員)。応用物理学会インフォマティクス応用研究グループ代表。東北大学金属材料研究所助教。名古屋大学未来社会創造機構特任講師などを経て2018年より現職。専門は結晶工学と機械学習応用。  
[連絡先] 〒464-8601 名古屋市千種区不老町名古屋大学工学部2号館(竹内研究室) 理化学研究所革新知能統合研究センター(竹内チーム)  
E-mail: Kentaro.kutsukake@riken.jp